

THESE DE DOCTORAT DE

Par

Typhaine VIOLO

Développement de méthodes de biologie synthétique appliquées à la génération de vaccins glycoconjugués contrôlés et à la modification des capacités de reconnaissance de protéines.

Thèse présentée et soutenue à Nantes le 3 Mars 2020

Unité de recherche : UFIP, UMR CNRS 6286

Thèse N° : à définir

Rapporteurs avant soutenance :

Gilles Truan Directeur de recherche CNRS - TBI UMR INRA 792 et UMR CNRS 5504
Laurence Mulard Directrice de recherche CNRS - Institut Pasteur – UMR CNRS 3523

Composition du Jury :

Président : Gilles Truan Directeur de recherche CNRS - TBI UMR INRA 792 et UMR CNRS 5504

Examineurs :
Barbara Mouratou Maître de conférences - CRCINA ISERM U1232
Patrice Soumillon Professeur des Universités - Université Catholique de Louvain
Elisabeth Ficko-Blean Chargée de recherche CNRS - Station Biologique de Roscoff UMR CNRS 8227

Dir. de thèse : Cyrille GrandJean Chargé de recherche - UFIP UMR CNRS 6286
Co-dir. de thèse : Jacques Lebreton Professeur des Universités - CEISAM UMR 6230
Co-enc. de thèse : Emilie Camberlein Maître de conférence - UFIP UMR CNRS 6286

*"La vie n'est pas d'attendre que les orages passent,
c'est d'apprendre comment danser sous la pluie"*

Sénèque

*"On a intérêt au bonheur des autres,
on a intérêt à ce que les autres ne soient pas malades,
on a intérêt au succès des autres
parce que nous sommes une équipe collective."*

Jacques Attali

"Aucun de nous ne sait ce que nous savons tous, ensemble."

Euripide

*"Entre ce que je pense, ce que je veux dire,
ce que je crois dire, ce que je dis,
ce que vous avez envie d'entendre, ce que vous comprenez...
il y a dix possibilités qu'on ait des difficultés à communiquer.
Mais essayons quand même..."*

Bernard Werber

Remerciements

Tout d'abord je remercie les membres du jury pour avoir accepté de lire et d'évaluer mon travail. Je voudrais donc remercier particulièrement les Docteurs Gilles Truan et Laurence Mullard pour avoir accepté d'être rapporteur de cette thèse, ainsi que les Docteurs Barbara Mouratou, Elizabeth Ficko-Blean et le Professeur Patrice Soumillion pour avoir accepté d'être membres de ce jury.

Je remercie la Région Pays de la Loire pour avoir soutenu financièrement ce projet grâce à la bourse "Paris scientifique" sans qui ce travail n'aurait pas pu avoir lieu.

Cette thèse a été réalisée au sein de l'UFIP (Unité Fonctionnalité et Ingénierie des Protéines) dont le Professeur Bernard Offmann est le directeur. Je te remercie pour ton accueil chaleureux au sein du laboratoire, merci également pour ton humanité malgré ton agenda débordé, de m'avoir 'désignée d'office' pour participer au comité de vie de l'unité au détour d'un couloir et n'oublions pas pour cette superbe danse de la chimie click.

Je voudrais chaleureusement remercier mon directeur de thèse Docteur Cyrille Grandjean pour sa forte implication dans le projet, sa franchise, ses conseils précieux, notamment en chimie, nos discussions scientifiques intéressantes, même tard à l'heure de partir et un merci pour tes petites blagues et jeux de mots perpétuels (même si Annie va m'en vouloir pour cet encouragement).

Je remercie aussi le Professeur Jacques Lebreton, co-directeur de cette thèse, pour son implication et la conduite du projet au sein du laboratoire de chimie CEISAM. Ce projet étant très dépendant de cet axe de travail, le travail effectué au CEISAM était d'une importance cruciale.

Je remercie tout particulièrement la Docteure Emilie Camberlein, co-encadrante de cette thèse. Je te remercie pour ton encadrement si juste et plein de bienveillance qui m'a permis d'aller au bout. Merci pour ta disponibilité au quotidien, pour ton aide précieuse et efficace, pour m'avoir écouté me plaindre et m'avoir remise d'attaque quand j'en avais besoin grâce à ton optimisme. Merci d'avoir été là pour moi scientifiquement et humainement partageant avec moi, échec et succès (petite danse de joie :). Merci également de m'avoir pleinement intégrée à tes projets ambitieux lors de la création de la première équipe iGEM de Nantes, ça a été une belle expérience. J'en profite également pour remercier ici cette équipe iGEM 2019 pour leur enthousiasme débordant et leur implication. Enfin Emilie, je voudrais également te remercier pour nous avoir fait goûter aux résultats de tes talents de brasseur ainsi que pour mon initiation de danse indienne.

Je n'oublie bien sûr pas mes nombreux collaborateurs en stage (Fanny, Bastien, Mehdi, Adrien, Denis, Morgane et Anthony), en tant que collègues au sein de l'UFIP ainsi qu'au sein de la Plateforme (Christophe, Johann, Amina, Annie, Cathy, Pierre, Stéphane), ou en tant que collaborateurs extérieurs au sein du CEISAM (Jacques, Arnaud, Corentin et Antoine), ou au sein de la plateforme Therassay (Dorian Caudale). Merci pour vos soutiens, le temps et l'aide que vous avez apportés à ce projet qui est le produit d'un grand travail commun. Merci également à ceux qui auraient contribué à ce projet de thèse en prenant du temps pour répondre à une question au détour d'un couloir du labo. Un merci particulier au Docteur Gilles Truan qui a accepté de me recevoir dans son laboratoire du LISBP (nouvellement TBI) afin de débloquent nos problématiques d'incorporation d'acides aminés non naturels ainsi qu'au Post-doctorant Robert Quast qui m'a coaché sur place et à leur équipe pour cette intégration éclair.

Un grand merci à mes collègues de l'UFIP qui ont fait de ce laboratoire un endroit où il fait bon vivre. Un merci particulier aux membres de l'équipe de glycobio. Merci donc à Emilie, Cyrille, Charles, Fabienne, Amélie, Annie, Laura, Franck, Corinne, Amina, Claude, Andy, Marine, Aline. Merci pour ces pauses cafés, des vrais moments d'échanges scientifiques (ou sur la vie) qui me sont devenus indispensables. Merci également pour les nombreux gâteaux du labo, y compris les chouquettes days (Merci Amélie !), qui m'ont souvent permis de pallier l'absence d'un petit déjeuner. Plus sérieusement, merci pour votre soutien moral, scientifique, ou technique. Merci également à l'équipe de Leila, nos nouveaux colocataires du bâtiment 9 qui partagent avec nous la vie toute particulière au sein de ce bâtiment. Merci aussi, aux équipes du bâtiment 25 pour avoir partagé mes pauses du midi ainsi que mes moments de solitude face au BLitz®.

Merci particulièrement aux doctorants Surbhi, Ennys, Johann, Thomas, Alex, aux anciens doctorants Maesh et Yahanar et aux nouveaux doctorants Xue, Udita, Yue, Fang. Merci pour ces sorties plages, autour d'une bière, d'un cocktail ou d'un kebab improvisé à la place de la gym tonic! Merci pour cette superbe tradition du vendredi pizza-film sur grand écran! Merci particulièrement à toi Surbhi pour ta gentillesse et ton énergie débordante, nos bavardages autour d'un café, pour tes massages de tête gratuits et pour ta présence tout du long.

Merci aux soirées entre filles de l'UFIP qui brise les barrières inter-équipes et donne une occasion de plus de boire un verre! Je remercie également le comité d'organisation de la vie du labo (Damien, Bernard, Houda, Cathy, Fabienne, Pascal) pour m'avoir intégré à l'organisation de ces sorties annuelles et différents repas qui permettent de tisser des liens autrement.

Un merci également à mes amis et ma famille qui sont là malgré la distance. Un merci particulier à Florelle qui a partagé avec moi ces instants de thésard à distance et à mes amies de l'ENSTBB toujours là ! Un merci particulier à ma sœur, Laurianne, qui est toujours présente malgré tout. Et enfin, un grand merci à mon mari, sans qui je ne serais pas vraiment moi-même. Merci pour ton soutien, ta présence, ta patience et bien entendu pour ton amour.

Merci donc à vous tous (collègues, famille, amis) qui avez partagé avec moi ces moments de vie.

Sommaire

Remerciements	2
Sommaire	4
Liste des Figures	10
Liste des Tableaux	14
Liste des Abréviations.....	15
Avant-Propos.....	18
Laboratoires d'accueil	18
Présentation succincte des deux axes constituant cette thèse	18
Organisation du manuscrit.....	19
Chapitre 1 : Méthodes d'incorporation d'acides aminés non naturels.....	20
I. Généralités	21
I.1 Généralités sur la synthèse protéique	21
I.2 Composition et structure des protéines	22
I.3 Mécanisme de la traduction chez les procaryotes	25
II. Incorporation d'acides aminés non naturels par modifications <i>in situ</i>	27
III. Incorporation d'acides aminés non naturels <i>in vitro</i>	28
III.1. Incorporation par synthèse chimique.....	28
III.2. Incorporation par pré-couplage de l'ARNt.....	29
IV. Incorporation d'acides aminés non naturels <i>in cellulo</i>	32
IV.1 Généralités sur les aminoacyl ARNt synthétase	32
IV.1.1 Mécanisme réactionnel catalysé par les aminoacyl ARNt synthétases.....	32
IV.1.2 Structures des aminoacyl ARNt synthétases	32
IV.2 Incorporation des acides aminés non naturels par méthode résidu spécifique	37
IV.2.1. Principe de la méthode d'incorporation résidu spécifique	37
IV.2.2. Méthode permettant l'incorporation résidu spécifique <i>in vivo</i>	37
IV.3 Incorporation des acides aminés non naturels de manière site spécifique	39
IV.3.1. Principe de la méthode d'incorporation site spécifique.....	39
IV.3.2. Aminoacyl ARNt synthétases orthogonales.....	40
IV.3.3. Mutagenèse des aminoacyl ARNt synthétases.....	46
IV.4 Optimisation et nouvelles stratégies	54
IV.4.1. Entrée de l'acide aminé non naturel : transport et concentration.....	54
IV.4.2. Conditions liées à la machinerie de la traduction	55
IV.4.3. Conditions liées à l'expression du couple aaRS / ARNt et de la protéine cible	60
IV.4.4. Incorporation de deux acides aminés non naturels différents au sein d'une même séquence	61
IV.5 Incorporation d'acides aminés non naturels chez les eucaryotes.....	62

V.	Applications.....	63
V.1.	Etude des modifications post traductionnelles	63
V.2.	Bioconjugaison orthogonales	64
V.3.	Incorporation de NAA au sein de sites actifs ou de reconnaissance	65
Chapitre 2 : Stratégies de lutte contre <i>Streptococcus pneumoniae</i>		66
I.	Généralités sur <i>Streptococcus pneumoniae</i>	67
I.1	Physiologie	67
I.1.1	Composition de la capsule.....	68
I.1.2	Autres composants de surface de <i>Streptococcus pneumoniae</i>	70
I.2	Réservoir	71
I.3	Pathologies associées à <i>S. pneumoniae</i>	72
I.3.1	Population à risque.....	72
I.3.2	Infections non invasives liées au pneumocoque	73
I.3.3	Infections invasives liées au pneumocoque	73
I.4	Facteurs de virulence et mécanismes liés à la pathogénicité.....	74
I.4.1	Mécanisme de transmission	74
I.4.2	Etapas de colonisation et d'invasion : les différents facteurs de virulence	75
I.4.2.1	Capsule polysaccharidique	75
I.4.2.2	Autres facteurs de virulence des premiers stades d'infection (mucus)	77
I.4.2.3	Les adhésines (adhérence à la couche épithéliale)	78
I.4.2.4	La pneumolysine et autres facteurs (passage de la couche épithéliale).....	79
I.4.2.5	Autres facteurs de virulence aux stades infectieux avancés	80
II.	Moyens de lutttes contre <i>Streptococcus pneumoniae</i>	81
II.1	Traitement à l'aide d'antibiotiques.....	81
II.2	Prévention à l'aide de Vaccins	84
II.2.1	Rappels sur le fonctionnement des vaccins.....	84
II.2.2	Bref historique des vaccins pour <i>Streptococcus pneumoniae</i>	86
II.2.3	Les vaccins polysaccharidiques : Pneumovax®	87
II.2.4	Les vaccins glycoconjugués.....	91
II.2.5	Vaccins: principales stratégies émergentes.....	99
Chapitre 3 : Génération de Vaccins glycoconjugués contre <i>Streptococcus pneumoniae</i>		105
Avant-propos / Présentation des objectifs		106
I.	Génération de conjugués monovalents site-spécifiques entre la mPsaA et le tétrasaccharide PnTS14 à l'aide de la Propargyl-Lysine.	109
I.1	Méthode d'incorporation de l'acide aminé non naturel	109
I.2	Synthèse de la Propargyl-Lysine	109
I.3	Choix des sites d'incorporation de la Propargyl-Lysine et mutagénèse de la mPsaA	112
I.4	Vérification de l'entrée de la Propargyl-Lysine dans la bactérie	114

I.5	Production des mutants mPsaA-PrK au sein du plasmide pET11a	115
I.6	Clonage des mutants mPsaA-TAG au sein du plasmide pET24, production et purification des protéines recombinantes	116
I.7	Conjugaisons par chimie click (CuAAC)	117
I.8	Développement d'une méthode de purification par chromatographie d'affinité à l'aide de la Galectine-3	121
I.9	Production des conjugués monovalents site spécifiques	123
I.10	Conclusion sur la production d'un vaccin monovalent site spécifique.....	124
II.	Optimisation de l'incorporation de la Propargyl-Lysine	124
II.1	Surexpression d'un transporteur de lysine	125
II.2	Mutagenèse du domaine N-terminal de l'aminocyl ARNt synthétase	126
II.3	Optimisation par plans d'expérience	127
II.3.1	Définition du problème	127
II.3.2	Choix du modèle et réalisation des expériences	130
II.3.3	Analyse des résultats	131
III.	Génération de conjugués multivalents site-spécifiques	135
III.1	Triple-incorporation de la Propargyl-Lysine au sein de la mPsaA.....	135
III.2	Conjugaison du Triple mutant.....	137
IV.	Analyse de la réponse immunitaire des glycoconjugués	138
IV.1.	Réponse immunitaire dirigée contre la mPsaA	138
IV.2.	Réponse immunitaire contre le polysaccharide capsulaire du sérotype 14.....	140
IV.3.	Bilan de la réponse immunitaire.....	141
	Conclusion	144
	Chapitre 4 : Matériel et Méthode pour la génération de vaccins glycoconjugués	146
I.	Synthèse de la PropargylLysine (PrK) : (échelle 500 mg)	147
II.	Test de pénétration de la Propargyl-Lysine dans les bactéries.....	147
III.	Production et analyse des protéines recombinantes et vaccins glycoconjugués.....	148
III.1	Transformation au sein des souches d'expression	148
III.2	Production des protéines sauvages (mPsaA WT)	148
III.3	Production des protéines mutées contenant une Propargyl-Lysine (mPsaA-K24PrK, mPsaA-K213PrK, et mPsaA-K229PrK).....	149
III.4	Vérification de l'incorporation de l'acide aminé non naturel par SDS-PAGE/WesternBlot	149
III.5	Purification des protéines recombinantes.....	149
III.6	Dichroïsme Circulaire	150
III.7	Conjugaison des protéines avec le tétrasaccharide-N ₃ ou avec la fluorescéine-N ₃	150
III.8	Purification sur colonne d'exclusion stérique et lyophilisation	150
III.9	Purification des glycoconjugués sur Galectine 3 fixée sur colonne NHS	151
III.10	Coupage de l'étiquette histidine par la protéase TEV.....	151
III.11	Production d'un vaccin glycoconjugué à l'aide de la méthode aléatoire	152

IV.	Biologie moléculaire – Préparation des gènes de la protéine cible mPsaA.....	152
II.1	Clonage du gène de la mPsaA dans le plasmide pET24d	152
II.2	Compétence des cellules	153
II.3	Transformation et vérification de séquences	153
II.4	Mutation du gène de la mPsaA par amplification PCR	153
V.	Optimisation de l'incorporation de l'acide aminé non naturel.....	154
IV.1	Optimisation de l'incorporation de la Propargyl-Lysine par la méthode de plan d'expérience.....	154
IV.2	Optimisation de l'incorporation de la Propargyl-Lysine par méthode d'ajout d'un transporteur	155
IV.3	Optimisation de l'incorporation de la Propargyl-Lysine par mutagenèse ciblée de l'aaRS	156
VI.	Analyse de la réponse immunitaire	157
V.1	Immunisation des souris pour évaluer la réponse immunitaire induite par les vaccins produits	157
V.2	Analyse de la réponse immunitaire par ELISA	157
	Chapitre 5 : Interactions Lectines - Saccharides.....	158
I.	Introduction générale	159
I.1	Lectines	159
I.1.1	Définition et Classification des lectines	159
I.1.2	Valence des lectines	160
I.2	Particularité des Carbohydrate Binding Modules.....	161
I.3	Génération de lectines artificielles	163
I.3.1	Les lectines organiques	163
I.3.2	Les aptamères.....	165
I.3.3	Inactivation de glycoenzymes.....	166
II.	Reconnaissance moléculaire protéine - sucre	166
II.1	La diversité des saccharides ligands.....	166
II.2	Types de liaisons permettant la reconnaissance protéine - sucre.....	168
II.2.1	Les liaisons Hydrogènes	168
II.2.2	Interactions cycle aromatique et saccharide.....	169
II.2.3	Autres types d'interactions.....	170
III.	Méthodes classiques d'ingénierie des lectines.....	172
III.1	Applications des lectines.....	172
III.1.1	Utilisation des lectines dans le domaine analytique et diagnostic.....	172
III.1.2	Utilisation des lectines en thérapeutique.....	173
III.2	Ingénierie des lectines - Modification de la spécificité.....	174
III.2.1	Production de lectines recombinantes.....	174
III.2.2	Méthodes d'ingénierie protéique.....	174

III.3	Ingénierie des lectines - Modification de la valence des lectines.....	176
IV.	Ingénierie protéique avec des acides aminés non naturels.....	176
IV.1	Utilisation d'acides aminés non naturels au sein de site de reconnaissance ou site catalytique.....	176
IV.1.1	Sondes biophysiques	177
IV.1.2	Etude des caractéristiques d'un acide aminé clé.....	178
IV.1.3	Liaison covalente de cofacteurs (notamment ions métalliques).....	179
IV.1.4	Contrôle optiques des propriétés catalytiques	179
IV.1.4	Modification de sa régiosélectivité ou énantiosélectivité.....	179
IV.1.5	Acide aminé non naturel en tant qu'acide aminé catalytique	180
IV.2	Utilisation des acides aminés non naturels chez les lectines	180
IV.2.1	Méthodes de criblage avec des acides aminés non naturels	180
IV.2.2	Valence	181
IV.2.3	Liaison boronate	181
IV.2.4	Interaction tryptophane - carbohydrate	182
Chapitre 6 : Stratégies de modification des propriétés de reconnaissance de lectines et CBM à l'aide d'acides aminés non naturels.....		183
Objectif et Organisation de la modification des lectines		184
I.	Incorporation d'une fonction aldéhyde	184
II.	Incorporation d'une fonction boronate	185
I.	Incorporation d'une fonction aldéhyde photoactivable	186
I.1	Synthèse d'un acide aminé possédant une fonction aldéhyde photoactivable	186
I.2	Choix des protéines cibles.....	188
I.2.1	Protéine cible n°1 : AAL-S2	189
I.2.2	Protéine cible n°2 : CBM32-DD2.....	191
I.2.3	Protéine cible n°3 : CBM40.....	191
I.3	Mise au point d'une méthode de sélection	192
I.3.1	Test de pénétration de l'ONpGlu au sein des cellules BL21(DE3)	192
I.3.2	Pré-sélection des mutants de la synthétase par une méthode de bioinformatique	193
I.3.3	Sélection expérimentale	194
II.	Incorporation d'une fonction aldéhyde protégée	198
II.1	Synthèse d'un acide aminé non naturel portant une fonction aldéhyde protégée	199
II.2	Incorporation de l'AcidK au sein des protéines cibles	199
II.2.1	Production des protéines sauvages et vérification de leur résistance à la lyophilisation et au traitement de déprotection	199
II.2.2	Incorporation de l'AcidK au sein des protéines cibles et preuve de la déprotection.	200
II.3	Test d'affinité des lectines cibles	202
II.3.1	Biotinylation des ligands (sialyl-lactose pour CBM40 et chitosane pour CBM32).....	202
II.3.2	Mesures d'affinité.....	204

III.	Incorporation d'une fonction boronate	210
III.1.	Choix des protéines cibles et tests d'expression	210
III.2.	Test d'expression des protéines cibles	212
III.3.	Incorporation de la phénylalanine boronate au sein de NagBb.....	213
	Conclusion	214
	Chapitre 7 : Matériels et Méthodes pour la modification de lectines	217
I.	ONpGlu : Mise au point d'une méthode de sélection	218
I.1	Synthèse de l'ONpGlu	218
I.2	Biologie moléculaire.....	220
I.2.1	Clonages EGFP en aval de mPsaA-K193TAG	220
I.2.2	Modification du plasmide pEVOL : suppression d'une des deux copies de l'aaRS....	220
I.3	Génération des mutants <i>in silico</i>	221
I.3.1	Préparation de la protéine.....	221
I.3.2	Préparation des acides aminés ONpGlu substrats	221
I.3.3	Génération des mutations et sélections des séquences les plus probables	222
I.4	Méthodes de sélection	222
II.	Utilisation de l'AcDK	224
II.1	Synthèse de l'acide aminé non naturel AcDK.....	225
II.2	Choix des protéines cibles à modifier avec l'AcDK.....	226
II.3	Incorporation de l'AcDK	226
II.3.1	Protéines cibles : génération des mutants	226
II.3.2	Mutagenèse pour obtenir la synthétase spécifique à l'AcDK	227
II.3.3	Production des protéines sauvages(WT) et mutants Glutamate (Glu)	227
II.3.4	Production des protéines mutées avec l'AcDK	227
II.3.5	Vérification de l'incorporation de l'AcDK dans une protéine	227
II.3.6	Purification des protéines.....	228
II.3.7	Analyse des protéines par dichroïsme circulaire	228
II.3.8	Mesures d'affinités par méthode BLItz	228
III.	Utilisation de la boronate phénylalanine.....	229
III.1	Choix des cibles.....	230
III.2	Incorporation du boronate au sein de protéines.....	230
III.2.1	Souches et plasmides utilisés	230
III.2.2	Production des protéines	231
	Conclusion Générale.....	233
	Références.....	236
	Annexes	267
	Annexe 1 : Tableau de la composition des principaux sérotypes de <i>S. pneumoniae</i>	267
	Annexe 2 : Présentation des plasmides.....	270

Annexe 3 : Liste des souches et des plasmides	275
Annexe 4 : Liste des amorces oligonucléotidiques	277
Annexe 5 : Packages de Rstudio pour réaliser des plans d'expériences	283
Annexe 6 : Résultats d'une analyse de plan d'expérience Rstudio	284
Annexe 7 : Liste des publications et communications	285
Annexe 8 : Publication JoVE (soumise).....	286

Liste des Figures

Figure 1: Présentation de la synthèse protéique	21
Figure 2 : Niveaux de structures des protéines	22
Figure 3: Mécanisme de protéase à sérine (exemple de la Trypsine).....	23
Figure 4: Représentation en diagramme de Venn des principales propriétés des acides aminés naturels.....	24
Figure 5: Mécanisme d'élongation lors de la traduction.....	26
Figure 6: Mécanisme de la terminaison de la traduction.....	27
Figure 7: Mécanisme de ligation par l'utilisation d'une intéine (A); Utilisation de ce mécanisme pour produire une protéine semi-synthétique (B)	29
Figure 8: Ligation d'un acide aminé non naturel sur un ARN de transfert (ARNt) à l'aide d'une T4 RNA ligase.....	30
Figure 9: Utilisation de flexizymes en fonction de l'acide aminé à charger et de son groupe partant.	31
Figure 10: Mécanisme réactionnel catalysé par l'aminocyl ARNt synthétase	32
Figure 11: Structure de l'ARNt.....	33
Figure 12: Structure d'une Thréonine ARNt synthetase en complexe avec son ARNt.....	34
Figure 13: Présentation de la signification d'une voie orthogonale	39
Figure 14: Incorporation d'un acide aminé non naturel (NAA) par la méthode site spécifique	40
Figure 15: Vue d'ensemble de la pyrrolysyl ARNt synthétase (PylRS).....	42
Figure 16: Site de reconnaissance de la pyrrolysyl ARNt synthétase (PylRS).....	43
Figure 17: Reconnaissance de l'ARNt par la pyrrolysyl ARNt synthétase (PylRS)	43
Figure 18: Structure de la Tyrosyl ARNt synthétase (TyrRS) issue de <i>M.jannaschii</i>	44
Figure 19 : Poche de reconnaissance de la tyrosine par la tyrosyl ARNt synthétase de <i>M. jannaschii</i>	45
Figure 20: Modes de reconnaissance de l'ARNt selon l'organisme d'origine de la Tyrosyl ARNt synthétase.	46
Figure 21: Cycle de sélection permettant de changer la spécificité d'une aminocyl ARNt synthétase (aaRS).....	47
Figure 22: Sélection positive et négative d'un couple aminocyl ARNt synthétase et ARNt.....	48
Figure 23: Méthode de sélection PACE appliquée à l'évolution d'aminocyl ARNt synthétase	50
Figure 24: Site actif de la pyrrolysine ARNt synthétase (PylRS)	51
Figure 25: Site actif de la Tyrosyl ARNt synthétase de <i>M. jannaschii</i> (MjTyrRS).....	52
Figure 26: Interaction entre la TyrRS de <i>M. jannaschii</i> et son ARNt : effet de la mutation D286R.	53
Figure 27: Facteurs influant la production d'une protéine contenant un acide aminé non naturel	54
Figure 28: Méthodologie pour la réaffectation d'un codon sens (ici exemple du codon AGG).....	58
Figure 29: Aspect morphologique de <i>Streptococcus pneumoniae</i>	67
Figure 30: Différences structurales entre le sérotype 19A et 19F	69
Figure 31: Identification de sérotype par la réaction de Quellung	70

Figure 32: Schéma explicatif de la fonction d'un pilus.....	71
Figure 33: Pourcentage de porteurs de Streptococcus pneumoniae en fonction de l'âge chez une population à haut taux de maladies pneumococcales (Population aborigènes d'Australie).....	72
Figure 34: Incidence des maladies invasives liées à Streptococcus pneumoniae entre 2004 et 2005 en Angleterre et Ecosse image reproduite à partir de la publication (Ramassubramanian 2015)	72
Figure 35: Transmission et propagation de Streptococcus pneumoniae.....	75
Figure 36: Histologie de l'épithélium respiratoire.....	78
Figure 37: Colonisation de l'épithélium par Streptococcus pneumoniae	79
Figure 38: Formation d'un pore par la pneumolysine.....	80
Figure 39: Schéma présentant les mécanismes d'action des principaux antibiotiques.....	82
Figure 40: Profils de résistance et de multirésistance aux antibiotiques selon le sérotype	84
Figure 41: Points essentiels du fonctionnement biologique d'un vaccin.....	85
Figure 42: Frise chronologique présentant les différents types de vaccins développés contre Streptococcus pneumoniae.....	86
Figure 43: Mécanismes de biosynthèse de la capsule.	87
Figure 44: Mécanismes de biosynthèse de la capsule.	88
Figure 45: Méthode de synthèse automatique d'oligosaccharides	89
Figure 46: Mécanisme de conversion de polysaccharides capsulaires en antigène T-dépendant.	93
Figure 47: Effet de l'introduction des vaccins conjugués sur le type de sérotypes impliqués dans les pathologies pneumococcal invasives chez les enfants de moins de 5 ans	94
Figure 48: Effet de l'utilisation des vaccins glycoconjugués sur la proportion de souches résistantes aux antibiotiques.....	95
Figure 49 : Modification des polysaccharides afin de les conjuguer aux protéines.....	96
Figure 50: Modification sur les protéines afin de greffer des polysaccharides capsulaires.	96
Figure 51: Principales réactions bioorthogonales	98
Figure 52: Procédé de production classique d'un glycoconjugué.	104
Figure 53: Présentation des partenaires du vaccin glycoconjugué développé.....	106
Figure 54 : Synthèse de l'acide aminé non naturel : la Propargyl-Lysine.....	110
Figure 55: RMN et CCM obtenue lors de la synthèse de la Propargyl-Lysine	111
Figure 56: Choix des positions d'incorporation de la PrK au sein de la mPsaA (PDB: 1PSZ).....	113
Figure 57: Formes tronquées de la Pneumococcal surface adhésine A mature (mPsaA).....	114
Figure 58: Vérification de l'entrée de la Propargyl-Lysine dans les bactéries par Chromatographie sur Couche mince (CCM)	115
Figure 59: Vérification par WesternBlot de la production des différents mutants mPsaA contenant la Propargyl-Lysine	116
Figure 60 : Vérification par SDS-PAGE de la production des différents mutants mPsaA contenant la Propargyl-Lysine	117
Figure 61 : Purification par chromatographie d'affinité sur colonne de nickel du mutant mPsaA-K213.	117
Figure 62 : Analyse par Dichroïsme circulaire de la résistance de la protéine mPsaA aux réactifs de conjugaison.	118
Figure 63 : Analyse par Dichroïsme circulaire de la résistance de la protéine mPsaA aux réactifs de conjugaison sur la nuit.	118
Figure 64 : Schéma représentant la réaction de conjugaison CuAAC entre la protéine et la fluorescéine.....	119
Figure 65 : Conjugaison de la mPsaA WT et K24PrK avec la fluorescéine.	119
Figure 66: Conjugaison par chimie click de la mPsaA (WT et K24) avec le tétrasaccharide.	120
Figure 67 : Structure du tétrasaccharide-N3.....	121

Figure 68 : Conjugaison du mutant mPsaAK24PrK en présence d'un gradient de concentration en cuivre.....	122
Figure 69: Méthode de purification des conjugués mPsaA-tétrасaccharide à l'aide de galectine-3..	123
Figure 70 : Présentation des différents procédés testés pour la production de vaccins glycoconjugués.	124
Figure 71 : Comparaison de l'efficacité de l'incorporation de la PrK avec ou sans surexpression de LysP.....	125
Figure 72 : Evaluation de l'impact des mutations R19H, H29R et T122S sur l'incorporation de la Propargyl-Lysine	126
Figure 73: Schéma représentant un domaine d'étude pour un plan d'expérience à deux facteurs (à gauche) ou trois facteurs (à droite).....	128
Figure 74 : Répartition des points expérimentaux selon un plan d'expérience de type Box-Behnken	130
Figure 75: Représentation graphique des résidus en fonction des valeurs ajustées grâce au modèle	132
Figure 76: Représentation de la fluorescence en fonction de la température et de la concentration en arabinose (à gauche) et en fonction de la température et du pH (à droite)	133
Figure 77 : Analyse par WesternBlot d'une incorporation triple ou une mono-incorporation de la Propargyl-Lysine au sein des souches BL21(DE3) ou B95ΔA.....	137
Figure 78: Conjugaison de la mPsaA sauvage, simple mutant (K24PrK) ou triple mutant (K24PrK-K213PrK-K229PrK) avec la fluorescéine.	137
Figure 79: Conjugaison du Triple mutant mPsaA-K24PrK-K213PrK-K229PrK avec le tétrasaccharide	138
Figure 80 : Analyse de la réponse immunitaire dirigée contre la mPsaA (IgG) à J35	140
Figure 81: Exemples comparatif de structure 3D de lectines riche en structure β (A) et riche en structure α (B)	160
Figure 82: La multivalence des lectines peut se traduire par une multimérisation (exemple 2A) ou par une répétition en tandem de séquences peptidiques (exemple 2B).	161
Figure 83: Schéma représentant la clusterisation de lectines à la surface d'une membrane	161
Figure 84: Structures 3D issus de trois types de Carbohydrate Binding Module (CBM).....	162
Figure 85: Réactivité d'un boronate permettant l'échange d'une fonction diol avec une forme tétraédrique (sp ³) transitoire.	163
Figure 86: Principe de l'impression moléculaire	164
Figure 87: Exemple de lectine organique (Ferrand, Crump, and Davis 2007).....	165
Figure 88: Présentation des angles (φ, ψ, et ω) au niveau de la liaison osidique (ici au niveau du α-D-glcp-(1→6)-α-D-glcp (iso-maltose)).....	166
Figure 89: 13 unités monosaccharidiques forment la majeure partie (80%) des oligosaccharides chez les mammifères.....	167
Figure 90: Schéma du site actif de CBM40 (issus de NanA de Streptococcus pneumoniae), (PDB 4C1W).....	169
Figure 91: Orientation les plus fréquentes entre le cycle d'un sucre et un cycle aromatique permettant l'établissement d'interactions CH- π.....	170
Figure 92: Représentation du site actif de la lectine LecA de Pseudomonas aeruginosa complexé avec le galactose.....	171
Figure 93: Principe des méthodes de sélections	176
Figure 94: Schéma représentant la réaction entre une fonction aldéhyde incorporée au sein d'une protéine et une fonction amine (A) ou des fonctions alcools (B) présentes sur le substrat.....	184
Figure 95: Modification de l'AcdK réalisée par Wang et al. afin de générer une diméthyl-lysine (Z. A. Wang et al. 2017).	185

Figure 96: Bilan synthèse de l'aldéhyde intermédiaire	186
Figure 97 : Ajout du groupement photolabile sur la fonction aldéhyde.....	187
Figure 98 : Dernière étape de synthèse afin de déprotéger les groupements amines et carboxyliques de la chaîne principale de l'acide aminé.	187
Figure 99: Démasquage de la fonction aldéhyde des composés 7 et 9 par irradiation UV (longueur d'onde 330nm).....	188
Figure 100: Comparaison des structures de l'ONpGlu et de l'acide glutamique	189
Figure 101: Site de reconnaissance de AAL-S2 (PDB: 1OFZ),	190
Figure 102: Extrait de résultat d'un alanine scanning sur la protéine AAL-S2	190
Figure 103: Site reconnaissance du CBM32-DD2 issu de Paenibacillus sp. IK-5 (PDB 4ZZ8).....	191
Figure 104: Site actif du CBM40 issu de NanA de Streptococcus pneumoniae, (PDB:4C1W).....	192
Figure 105: Vérification de l'entrée de l'ONpGlu au sein des bactéries.....	193
Figure 106: Présentation des différents mutants de l'aaRS issus de l'analyse bioinformatique	194
Figure 107: Sélection test sur milieu solide autoinductif.	196
Figure 108: Test de différentes concentrations de la Propargyl-Lysine afin de déterminer des conditions de sélection en plaque 96 puits.....	197
Figure 109: Extrait de données de sélections des différents mutants prédits par l'analyse bioinformatique pour l'incorporation de l'ONpGlu.....	198
Figure 110: Comparaison des structures de l'ONpGlu et de l'AcidK avant et après déprotection par réduction grâce aux UV pour ONpGlu ou au TCEP pour AcidK.	199
Figure 111: Profils de dichroïsme circulaire montrant la résistance des protéines CBM40 et CBM32(DD2) au traitement par le TCEP	200
Figure 112: Profils de dichroïsme circulaire montrant la résistance des protéines CBM40 et CBM32(DD2) à la conservation par lyophilisation.....	200
Figure 113: Test de conditions pour l'incorporation de l'AcidK	201
Figure 114: Vérification de l'incorporation de l'AcidK au sein des protéines (A) CBM40 par western blot et (B) CBM32-DD2 par SDS PAGE ;	202
Figure 115: Principe de la mesure d'affinité par méthode Biolayer Light Interferometry (BLItz®)	203
Figure 116: Vérification de la biotinylation des ligands (A) 3'Sialyl-lactose (3'SL), et 6'Sialyl-lactose (6'SL) et (B) du chitosan par chromatographie sur couche mince.	204
Figure 117: Courbes d'association et de dissociation des différents mutants CBM32-DD2 avant et après traitement TCEP.....	205
Figure 118: Effet du traitement TCEP sur les courbes de dissociation des différents mutants étudiés	206
Figure 119: Stratégie de correction de la concentration protéique des mutants CBM40 comparativement à la protéine sauvage.....	207
Figure 120: Profils de liaison des différents mutants de CBM40 avant et après traitement au TCEP	208
Figure 121: Effet du traitement TCEP sur les courbes d'affinité des différents mutants.....	209
Figure 122: Site de reconnaissance de la protéine NagBb (PDB: 5WZN)	210
Figure 123: Site de reconnaissance de la protéine GH98Sp (PDB: 2WMI).....	211
Figure 124 : Site de reconnaissance de la protéine CLC1b, (PDB: 3WSR)	212
Figure 125: Analyse SDS-PAGE d'extraits bruts de cultures montrant l'expression des différentes protéines cibles en milieu LB ou autoinductif.	213
Figure 126: Analyse SDS-PAGE montrant l'incorporation de la Boronate phénylalanine au sein de la protéine NagBb-Y329TAG.	213
Figure 127 : Méthode de synthèse de l'aldéhyde intermédiaire (composé 6).	218
Figure 128 : Ajout du groupement permettant le masquage de la fonction aldéhyde.	219

Figure 129 : Etape finale de la synthèse permettant de produire l'acide aminé non naturel déprotégé	219
Figure 130 : Schéma générale de la synthèse de l'AcidK	225
Figure 131 : Schéma représentant la déprotection de la fonction aldéhyde par action du TCEP sur le groupement 4-azidobenzoyloxycarbonyl	225

Liste des Tableaux

Tableau 1: Quelques exemples de déterminants au niveau de l'ARNt	33
Tableau 2: Classification des aminoacyl ARNt synthétases.....	35
Tableau 3: Différences caractéristiques des aaRS selon leur classe	36
Tableau 4: Exemples d'analogues d'acides aminés incorporés par la méthode résidu spécifique.....	38
Tableau 5: AaRS et ARNt utilisés pour la modification du code génétique in vivo	41
Tableau 6 : Exemples d'acides aminés non naturels pour l'étude de modifications post-traductionnelles	64
Tableau 7 : Exemple de 3 structures de polysaccharides capsulaires.....	68
Tableau 8: Exemple de composition de capsules non chargées négativement.....	76
Tableau 9: Présentation des principaux facteurs de virulence	81
Tableau 10: Mécanismes de résistance aux antibiotiques développés par <i>S. pneumoniae</i>	83
Tableau 11: Exemples d'oligosaccharides représentatifs de différents sérotypes synthétisés	90
Tableau 12: Schéma vaccinal recommandé en France à partir de 2017.....	92
Tableau 13: Différentes stratégies vaccinales basées sur des antigènes protéiques.	100
Tableau 14: Stratégies de vaccins glycoconjugués avec des protéines porteuses immunogènes.....	102
Tableau 15: Résumé de précédents travaux vaccinaux réalisés au sein du laboratoire.....	108
Tableau 16 : Facteurs pouvant influencer le taux d'incorporation de la Propargyl-Lysine.....	128
Tableau 17 :Détermination du domaine d'étude (niveaux hauts et bas des facteurs).....	129
Tableau 18 : Vérification du modèle prédit.....	134
Tableau 19 : Titres des anticorps IgG dirigés contre la mPsaA (à J35)	139
Tableau 20 : Comparaison des résultats pour des immunisations avec les précédentes études.....	142
Tableau 21: Exemples de lectines selon leurs origines, leur principale fonction et leur type de reconnaissance.....	159
Tableau 22: Lectines reconnaissant les 6 principaux monosaccharides.....	167
Tableau 23: Présentation des différentes utilisations de lectines en analyse et diagnostique	172
Tableau 24: Différents exemples de lectines ayant des potentiels antitumoraux selon différents mécanismes.....	173
Tableau 25: Résultats possibles lors de la sélection d'aaRS avec notre méthode de sélection.....	195
Tableau 26 : Exemples récapitulant les possibilités liées à la méthode de sélection employée	220

Liste des Abréviations

- Aa : Acide aminé (naturel)
- AAL : *Aleuria Aurantia* Lectine
- AAL-S2 : Forme monosite (site 2) de la lectine AAL
- aaRS : Aminoacyl ARNt synthétase (aa sera éventuellement remplacé par l'abréviation en trois lettres de l'acide aminé correspondant)
- AcdK : N ϵ -(4-azidobenzoxycarbonyl)- δ , ϵ -dehydrolysine
- ADN : Acide désoxyribonucléique
- ARN : Acide Ribonucléique
- ARNm : ARN messenger
- ARNt : ARN de transfert
- ATP : Adénosine Triphosphate
- Bpa : Boronate phénylalanine
- BSA : Bovine sérum albumine
- CBM : Carbohydrate Binding Module
- CCM : Chromatographie sur Couche Mince
- CLC1b : lectine de type C humaine
- CMHII : Complexe majeur d'histocompatibilité de type II
- CMI : Concentration minimale inhibitrice
- CP14 : Polysaccharide Capsulaire du sérotype 14 (de *Streptococcus pneumoniae*)
- CPA : Cellule Présentatrice d'Antigène
- CRM197 : Cross Reacting Materials 197 (mutant non toxique de la toxine diphtérique)
- CTLA-4 : antigène 4 du lymphocyte T cytotoxique
- CuAAC : Copper(I)-Catalyzed alkyne-azide cycloaddition
- DO : densité optique
- dPly : Pneumolysine détoxifiée
- EDTA : Acide éthylènediaminetétraacétique

-EF-Tu : Facteur d'élongation

-EGFP : Enhanced Green Fluorescent Protein

-ELISA : Enzyme linked immunosorbent assay

-FACS : Cytométrie en flux

-Fc : Fragment constant

-FRET : Fluorescence Resonance Energy Transfert

-GH98Sp: Glycoside Hydrolase 98 (issue de *Streptococcus pneumoniae*)

-GTP : Guanosine Triphosphate

-HPLC : Chromatographie en phase liquide à haute performance

-IEDDA : Inverse Electron Demand Diels Alder

-IgA1 : Immunoglobuline de type A et de sous-classe 1

-IgG : Immunoglobuline de type G

-IgM : Immunoglobuline de type M

-IPTG : isopropyl β -D-1-thiogalactopyranoside

-ITC : Titration Calorimétrique Isotherme

-Kd : Constante de dissociation

-LPS : Lipopolysaccharide

-MjTyrRS : Tyrosyl ARNt synthétase issue de *Methanocaldococcus jannaschii*

-mPsaA: mature Pneumococcal surface adhésine A

-NAA : Acide aminé non naturel

-NagBb: Alpha-N-acetylgalactosaminidase issue de *Bifidobacterium bifidum*

-NanA : Neuraminidase A

-ONpGlu : Ortho-Nitro-para-Glutamate

-PACE : *Phage Assisted Continuous Evolution*

-PCR : Polymerase Chain Reaction

-PDB : Protein Data Bank

-pH : potentiel Hydrogène

- PMSF : fluorure de phénylméthylsulfonyle
- PnTS14 : Tétracaccharide du sérotype 14 (de *Streptococcus pneumoniae*)
- PrK : N-ε-propargyloxycarbonyl-L-Lysine (communément appelée Propargyl-Lysine)
- PyIRS : Pyrrolysyl ARNt Synthétase
- R² : coefficient de détermination
- RF : *Releasing Factor*, soit Facteur d'arrêt de la traduction
- RMN : Résonance Magnétique Nucléaire
- RSL : Lectine issue de *Ralstonia solanacearum*
- RT-PCR : Reverse Transcriptase- Polymérase Chain Réaction
- SDSL : Site directed Spin Labeling
- SDS-PAGE : sodium dodecyl sulfate polyacrylamide gel electrophoresis
- SIDA : Syndrome d'Immunodéficience Acquis
- SPAAC : Strain Promoted Alkyne-Azide cycloaddition
- SPR : Surface Plasmon Resonance
- TBS : Milieu Tris Salin
- TCEP : tris(2-carboxyethyl)phosphine
- TEV : Tobacco Etch Virus
- TFA : Acide Trifluoroacétique
- UMP : Uridine Monophosphate
- UPRT : Uracile phosphoribosyltransférase
- UV : Ultra Violet
- VGEF : Vascular endothelial Growth Factor
- VIH : Virus d'Immunodéficience humaine
- WT : sauvage

Avant-Propos

Laboratoires d'accueil

Ce travail de thèse a été réalisé au sein du laboratoire Unité Fonctionnalité et Ingénierie des Protéines, (UFIP-UMR CNRS 6286). Cette unité est une unité mixte CNRS et Université de Nantes.

L'Unité est composée de 5 équipes centrées sur l'étude de fonction de protéines impliquées dans différents contextes biologiques :

Equipe 1 : Conception de protéines *in silico* (dirigée par Yves Henri Sanejouand)

Equipe 2 : Ingénierie moléculaire et glycobiologie (dirigée par Cyrille GrandJean)

Equipe 3 : Mécanisme et régulation de la réparation de l'ADN (dirigée par Fabrice Fleury)

Equipe 4 : Epigénétique: prolifération et différenciation (dirigée par Christophe Thiriet)

Equipe 5 : Epigénomiques des microalgues et interactions avec l'environnement (dirigée par Leila Tirichine)

Grâce à la multidisciplinarité de cette unité, des recherches à l'interface entre la biologie, la chimie et la bioinformatique sont réalisées. Ce cadre pluridisciplinaire est favorable à la réalisation de travaux de biologie de synthèse. Le travail de thèse présenté ici est d'ailleurs la concrétisation de ce nouvel axe au sein de l'équipe 2 : « Ingénierie moléculaire et glycobiologie », équipe intrinsèquement pluridisciplinaire puisque composée de chimistes et de biologistes. Les travaux de cette équipe sont variés allant de thématiques plutôt fondamentales (évolution d'enzymes pour la synthèse d'oligosaccharides, ingénierie d'enzymes pour la production de glycosylations régiosélectives, regroupement d'enzymes pour un fonctionnement optimisé appliqué à la production de biocarburant) ou thérapeutiques (génération d'inhibiteur multivalent, production de vaccins glycoconjugués et inhibiteurs de Galectines). L'UFIP possède également une plateforme de service (la plateforme IMPACT) qui a participé à ce projet pour l'analyse des protéines modifiées.

Ce travail de thèse a également été développé en collaboration avec le CEISAM UMR CNRS 6230 (Chimie Et Interdisciplinarité Synthèse Analyse et Modélisation) à Nantes, et plus particulièrement avec l'équipe SYMBIOSE qui comprend une plateforme de service de synthèses à façon. Ce laboratoire a en effet réalisé la synthèse des acides aminés non naturels originaux de ce projet.

Enfin, les tests biologiques sur les modèles murins ont été réalisés au sein de la Plateforme Therassay à Nantes.

Présentation succincte des deux axes constituant cette thèse

Ce projet de thèse s'inscrit dans une volonté de développer une nouvelle thématique au sein du laboratoire : la biologie synthétique. Il s'agit d'une discipline qui regroupe des travaux d'ingénierie du vivant très variés allant de l'ingénierie métabolique à la construction d'organismes minimaux artificiels. Il n'y a pas de définition officielle commune à tous (Osbourn et al. 2012), mais voici deux définitions assez représentatives du domaine :

« *La biologie de synthèse est le design et l'ingénierie de domaines biologiques, de nouvelles machines et systèmes biologiques ainsi que le redesign de systèmes biologiques existants.* » (Clarke 2012).

« *La biologie de synthèse est l'ingénierie de la biologie : la synthèse de systèmes complexes, basés ou inspirés du vivant, avec des fonctions qui n'existent pas dans la nature. L'ingénierie peut être appliquée à tous les niveaux de la hiérarchie des structures biologiques de molécules aux cellules*

entières, tissus et organismes. Par essence, la biologie de synthèse va permettre le design de systèmes biologiques par une approche rationnelle et systématique. » (Krassnig 2007).

Ce domaine s'inscrit en continuité de la biologie moléculaire qui permet la modification du vivant mais avec une nouvelle approche pluridisciplinaire liée aux sciences de l'ingénieur. Ainsi, la biologie de synthèse consiste en une modification rationnelle à partir de la caractérisation de « briques » biologiques que l'on peut ensuite modifier, assembler et transférer d'un organisme à un autre afin d'atteindre une nouvelle fonction.

De plus, d'un point de vue fondamental, l'ingénierie des systèmes biologiques devrait permettre d'améliorer nos connaissances du vivant. En effet, comme l'a écrit Richard P. Feynman (Prix Nobel physique 1965) « Ce que je ne peux pas créer, je ne le comprends pas ». L'ingénierie du vivant permet alors d'expérimenter des modifications du vivant et de les étudier afin de mieux comprendre celui-ci.

Au cours de cette thèse nous nous sommes intéressés à un pan précis de la biologie synthétique qui est **l'incorporation d'acides aminés non naturels au sein de protéines** afin d'apporter de nouvelles propriétés à des protéines impliquées en glycobiochimie. Cette technologie a alors été appréhendée selon deux axes indépendants :

-la première partie consiste à mettre au point l'incorporation d'acides aminés non naturels au sein du laboratoire en l'appliquant à la génération d'un vaccin glycoconjugué. Cette première partie consiste donc à s'approprier cette technologie en l'appliquant à une thématique déjà exploitée au laboratoire et d'y apporter une approche complémentaire.

-la deuxième partie consiste à utiliser des acides aminés non naturels au sein de sites de reconnaissance de protéines interagissant avec des sucres afin d'en améliorer les propriétés. Pour cela, plusieurs stratégies ont été employées comprenant l'utilisation de différents acides aminés non naturels.

Organisation du manuscrit

Dans un premier temps, un état de l'art de la technologie d'incorporation des acides aminés non naturels sera présenté (Chapitre 1). Par la suite, le manuscrit sera divisé en deux parties selon les deux principaux axes de ce travail: la génération d'un vaccin glycoconjugué à l'aide d'acides aminés non naturels (Partie I) et la modification des propriétés de reconnaissance de protéines (Partie II).

La partie I sur la génération de vaccin glycoconjugué commencera par une introduction sur les vaccins contre *Streptococcus pneumoniae* et les différentes méthodes permettant la génération de glycoconjugués (Chapitre 2) puis par le contexte biologique et scientifique de ce projet ainsi que les objectifs associés, suivis des résultats obtenus ainsi que de leur discussion (Chapitre 3) et enfin les méthodologies appliquées (Chapitre 4).

La partie II sur la modification des propriétés de reconnaissance des protéines commencera par une introduction générale sur les protéines reconnaissant les sucres (Chapitre 5) puis par le contexte scientifique de ce projet ainsi que les objectifs associés, suivis des résultats obtenus ainsi que de leur discussion (Chapitre 6) et enfin les méthodologies appliquées (Chapitre 7).

Une conclusion générale permettra ensuite de finaliser ce manuscrit de thèse.

***Chapitre 1 : Méthodes d'incorporation
d'acides aminés non naturels***

Les protéines représentent environ 50% du poids sec des cellules et constituent une des plus grandes classes de biomolécules qui composent tous les êtres vivants. De plus, elles remplissent un champ très large de fonctions, allant de fonctions catalytiques à des fonctions structurales. Pourtant, ces protéines ne sont généralement formées qu'à partir d'un enchaînement de 20 acides aminés naturels différents. Nous allons dans un premier temps revoir leur synthèse de façon succincte, ainsi que leur structure de manière générale avant de s'intéresser plus particulièrement aux acides aminés qui les composent et aux différentes méthodes qui permettent d'incorporer des acides aminés non naturels.

I. Généralités

I.1 Généralités sur la synthèse protéique

L'ADN contient l'information génétique qui permet la synthèse des différents constituants cellulaires, y compris les protéines. A la manière d'un livre de stockage, celui-ci permet de conserver l'information de la manière la plus stricte possible à l'aide de ces deux brins qui constituent deux copies de l'information. Ensuite lors du processus qui mènera à la synthèse des protéines, une partie de cet ADN est transcrit en ARN messager (ARNm) simple brin. Il s'agit alors en quelque sorte d'une note qui permet de donner succinctement les informations essentielles pour la formation de protéines. Cette information génétique est alors toujours constituée d'un enchaînement de nucléotides (A, T/U, G, C) (Figure 1)

Par la suite, cet ARN messager (ARNm) sera lu par un organite, le ribosome, qui fera correspondre chaque codon (triplet de nucléotides) avec un ARN de transfert (ARNt). Cet ARN de transfert aura été préalablement couplé à un acide aminé grâce à l'action d'une enzyme, l'aminocyl ARNt synthetase (aaRS). Pour reprendre la comparaison de l'ADN en tant que livre, et l'ARNm en tant que message court issu de ce même livre, le ribosome serait en fait un lecteur qui lirait ce message (ARNm) grâce à un langage, sorte de codex. Ce codex serait en fait constitué d'aaRS qui fait le lien entre chaque codon (triplet de nucléotides) porté alors par l'ARNt et chaque acide aminé. Le ribosome utilise donc l'ARNt couplé à l'acide aminé, issu de ce codex pour traduire ce message d'un langage génétique (polymère de nucléotides, à 4 bases) à un langage protéique (polymère d'acides aminés, à 20 acides aminés). Le polymère d'acide aminé alors produit constitue la base d'une protéine. En effet, à la manière d'un texte qui est formé d'un enchaînement de mots, les protéines sont formées d'un enchaînement d'acides aminés (Figure 1).

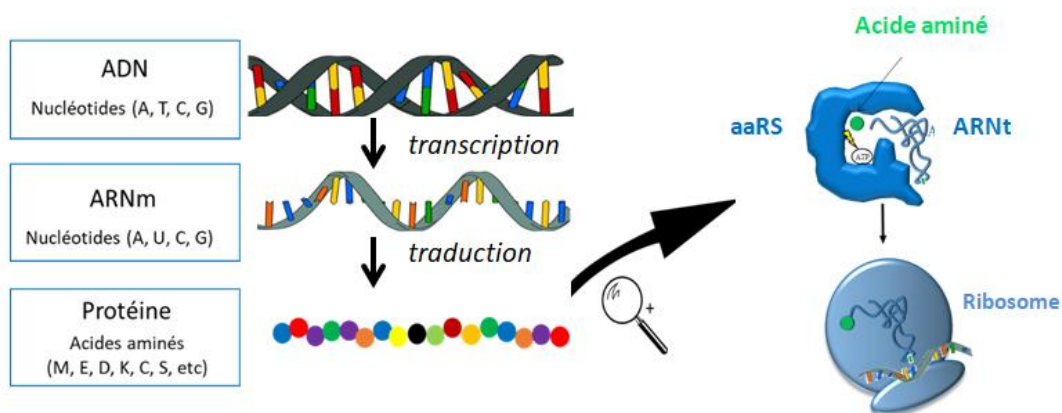


Figure 1: Présentation de la synthèse protéique

L'ADN porte l'information génétique dont une partie est ensuite retranscrite en ARN messager. Celui-ci est lu par le ribosome à l'aide d'ARNt préalablement chargés d'un acide aminé grâce à l'action d'une aaRS (aminocyl ARNt synthetase) afin de produire le polypeptide qui sera la structure primaire de la protéine produite.

1.2 Composition et structure des protéines

Les protéines sont donc des polymères d'acides aminés. Ceux-ci sont composés d'une chaîne principale qui permet leur polymérisation et d'une chaîne latérale qui porte des fonctions chimiques différentes. Il en existe 20 naturels différents et chacun possède des caractéristiques physicochimiques uniques (encombrement, charge, polarité, réactivité). L'enchaînement de ces acides aminés constitue la structure primaire d'une protéine. Le polymère libéré par le ribosome (section 1.3) va se replier selon des structures secondaires (feuillet β , hélice α , boucles), qui vont elles-mêmes s'agencer de façons variées dans l'espace et former ce qu'on appelle la structure tertiaire des protéines. Enfin, plusieurs unités de protéines peuvent s'agencer entre elles pour former une structure quaternaire (Figure 2).

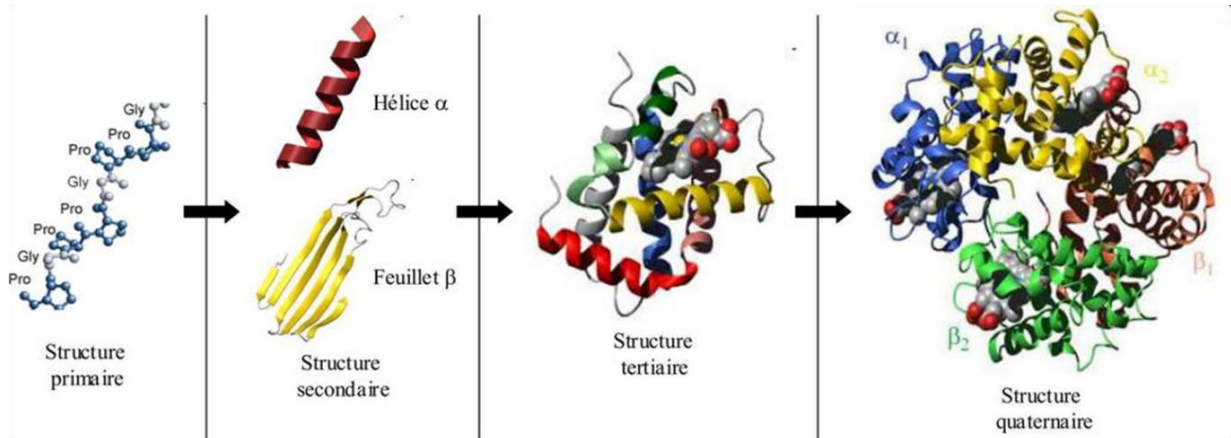


Figure 2 : Niveaux de structures des protéines

Structures des protéines du niveau le plus simple (à gauche) au niveau le plus complexe (à droite). Image extraite de la thèse (Barret 2014).

Les protéines tirent leurs propriétés de leur structure, mais aussi des chaînes latérales portées par les acides aminés qui les composent. Ainsi, au sein du site actif d'une enzyme la nature des chaînes latérales des acides aminés a un rôle essentiel. Prenons comme exemple la trypsine dont l'action est d'hydrolyser une liaison peptidique suivant une lysine ou une arginine. Au sein de son site actif, elle possède trois acides aminés ayant des propriétés clés pour son mécanisme d'action : une sérine, une histidine et un acide aspartique. En effet, ces acides aminés interagissent entre eux et permettent de rendre une fonction alcool de la chaîne latérale de la sérine nucléophile ce qui permet l'attaque de la liaison peptidique à hydrolyser selon le mécanisme présenté Figure 3.

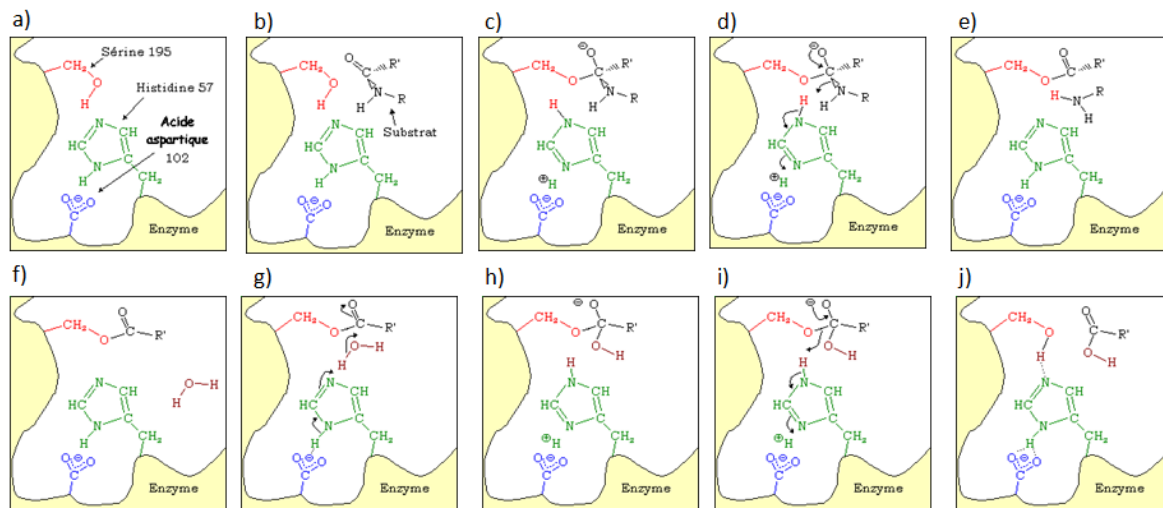


Figure 3: Mécanisme de protéase à sérine (exemple de la Trypsine)

La fonction carbonyle du substrat va interagir avec cette fonction alcool nucléophile de la sérine (b-c), puis par une délocalisation d'électrons jusqu'à l'histidine (d) il y a rupture de la liaison peptidique et relargage d'une partie du substrat clivé (e), puis grâce à une molécule d'eau au sein du site actif, le transfert inverse d'électrons va entraîner le relargage du substrat restant. Séries d'images créées à partir du site web: http://www.sciences-en-ligne.com/DIST/Data/Ressources/lic2/chimie/chi_orga/bioorga/enzymes/enzymes.htm.

Naturellement, il existe essentiellement 20 acides aminés qui constituent la majorité des protéines. Stéphane Freeland et ses collaborateurs travaillent sur les origines et les évolutions du code génétique en Astrobiologie. Ils ont classé ces acides aminés selon 3 propriétés : la taille, la charge et l'hydrophobicité. Cette classification a permis de montrer que ces 20 acides aminés naturels couvrent l'essentiel de l'espace chimique et que cette répartition ne semble donc pas aléatoire (Ilardo and Freeland 2014). La notion de couverture de l'espace chimique vient d'une méthode de découverte de nouveaux médicaments en chimie qui consiste à essayer de couvrir les différentes possibilités de molécules selon un descripteur choisi (encombrement, propriétés). La représentation des caractéristiques étudiées va alors définir un espace au sein duquel les molécules pourront être représentées et classées. Cette représentation va être utile afin de classer les molécules en différentes catégories permettant de définir des éléments clés pour l'activité recherchée. Lorsqu'une molécule avec une caractéristique recherchée est identifiée, étendre progressivement la zone de recherche permet de tester différentes modifications pour optimiser les propriétés ciblées. De plus, identifier une zone de l'espace chimique non exploitée permet d'appréhender le développement de molécules innovantes.

Les questions fondamentales ayant découlé de ces travaux de classification des acides aminés sont : "Pourquoi ces 20 acides aminés là et pourquoi uniquement 20 ?"

Les acides aminés n'ont pas été intégrés au sein de la vie biologique en même temps. Dix acides aminés les plus simples (A, D, E, G, I, L, P, S, T, V) sont supposés être les premiers à avoir été utilisés. Ils présentent alors déjà une couverture minimale qui permet de réaliser le repliement de protéines (Ilardo and Freeland 2014). Expérimentalement, des protéines formées uniquement à partir de 13 acides aminés ont été produites par mutagenèse des acides aminés C, F, I, M, Q, T, W qui ont été remplacés au sein de la séquence. Ces protéines sont correctement repliées et actives bien qu'étant moins efficaces et moins stables (Shibue et al. 2018). La composition de ces protéines est de 80% des supposés 10 premiers acides aminés, et cela a ainsi pu démontrer la fonctionnalité déjà importante apportée par ceux-ci. Puis les 10 suivants (R, N, C, Q, H, K, M, F, W, Y) auraient été progressivement intégrés au sein du vivant. Ils auraient alors permis de couvrir les zones de l'espace

chimique sous-représentées et ainsi maximiser les possibilités apportées (Ilardo and Freeland 2014). Mais ils auraient également permis d'augmenter la continuité de la couverture de cet espace chimique (Ilardo and Freeland 2014). De manière générale, les 20 acides aminés naturels permettraient donc de fournir un éventail assez large couvrant ainsi efficacement l'espace chimique. D'ailleurs, leur intégration dans la vie biologique a permis le développement de fonctions protéiques très variées permettant l'adaptation de la vie dans des conditions très diverses.

Cependant, bien que n'ayant pas été intégré au sein de la vie biologique, un plus large éventail d'acides aminés aurait été originellement disponible. Le code génétique permet de fournir 64 codons différents, l'évolution aurait donc pu permettre l'incorporation d'une plus grande variété d'acides aminés. Néanmoins, la redondance du code génétique permettrait de limiter l'impact de mutations en les rendant silencieuses. De plus, les acides aminés possédant des propriétés similaires sont codés par des codons proches ce qui permettrait à la protéine produite de conserver sa structure et sa fonction lors de l'apparition de mutations ponctuelles (Tze-Fei Wong 1980). Ces caractéristiques du code génétique limitent ainsi la perte de fonctions biologiques (Freeland et al. 2000). De plus, des théories soutiennent que la limitation serait inscrite dans la reconnaissance entre l'aaRS et l'ARNt (Saint-Léger et al. 2016). L'ajout de nouveaux acides aminés impliquerait l'adaptation du système de traduction et serait alors assez inefficace par rapport au bénéfice qu'il apporterait à l'organisme. En effet, la reconnaissance de l'ARNt par l'aaRS doit être suffisamment spécifique pour ne pas le confondre avec un ARNt codant pour un autre acide aminé. Mais la structure de l'ARNt doit pouvoir s'adapter à sa prise en charge par le ribosome ainsi que par le facteur d'élongation lors du processus de la traduction, commun à tous les ARNt. Malgré tout, cette limitation a été ponctuellement outrepassée naturellement par l'utilisation biologique de quelques acides aminés naturels rares (sélénocystéine, pyrrolysine). Ceux-ci sont généralement incorporés au sein de protéines ayant des rôles très spécifiques. Par exemple, la pyrrolysine est un acide aminé naturel rare qui est incorporé au sein de méthyltransférases intervenant dans la synthèse du méthane, chez des archées méthanogènes. De plus, d'autres mécanismes ont permis d'élargir les fonctionnalités disponibles au sein des protéines comme les modifications post-traductionnelles ainsi que par l'utilisation de co-facteurs. Par exemple, certaines enzymes utilisent des cofacteurs métalliques permettant des réactions d'oxydo-réduction.

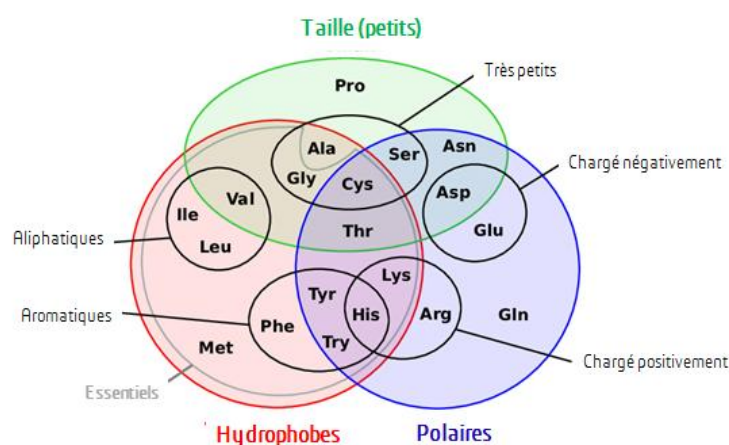


Figure 4: Représentation en diagramme de Venn des principales propriétés des acides aminés naturels. Image traduite à partir de la communication de (Jean-Baptiste, Berthelot, and Favre 2016).

Malgré les difficultés liées à l'ajout de nouveaux acides aminés non naturels (NAA) dans les séquences des protéines, de nombreux scientifiques ont travaillé sur leur incorporation. Ces

incorporations ont alors permis d'apporter de nouveaux outils pour l'ingénierie de protéines ou pour répondre à des questions scientifiques d'ordre divers (applications section V). Tout d'abord, nous allons voir rapidement les mécanismes de la traduction chez les procaryotes (section 1.3), puis les différentes techniques qui permettent l'incorporation des NAA au sein des protéines en nous focalisant particulièrement sur les techniques développées *in vivo* chez les bactéries étant donné qu'il s'agit des méthodes utilisées au cours de cette thèse (sections II, III et IV). Enfin, nous présenteront les différentes catégories d'applications ayant découlé de l'incorporation des NAA dans des protéines (section V).

1.3 Mécanisme de la traduction chez les procaryotes

Comme nous le verrons ultérieurement, l'incorporation d'acides aminés non naturels est principalement réalisée par le détournement de la machinerie de la traduction. Ce travail de thèse ayant été exclusivement réalisé chez *E. coli*, nous allons alors rappeler ici les mécanismes de la traduction chez les procaryotes.

La traduction s'effectue au niveau d'un complexe ribonucléotidique appelé ribosome et formé de deux sous-unités : une petite (sous-unité 30S) et une grande (sous-unité 50S). Ces sous-unités sont libres dans le cytoplasme et s'associent au moment de l'initiation de la traduction. La petite sous-unité joue alors un rôle crucial au niveau de l'association entre les différents codons de l'ARN messager (ARNm) et les ARN de transferts (ARNt) couplés aux différents acides aminés, tandis que la grande sous-unité est le siège des réactions de polymérisation permettant la production de la chaîne peptidique. Au niveau du ribosome, on va alors distinguer trois sites : le site A, le site P et le site E. Le site A est le siège de la liaison à un ARNt en attente, tandis que le site P est le siège de la réaction de polymérisation et le site E est un site de sortie du complexe ARNt acide aminé.

La traduction va alors s'effectuer selon trois étapes : l'initiation, l'élongation et la terminaison.

L'initiation de la traduction passe par des facteurs d'initiations (IF1, IF2 et IF3) qui favorisent la reconnaissance de l'ADN par la petite sous-unité du ribosome au niveau d'une séquence conservée (Shine Dalgarno). L'ARNt codant pour la méthionine (codon initiateur ATG) va alors se fixer, les facteurs d'initiation vont être relargués et la grande sous-unité du ribosome va s'assembler à la petite. Puis le processus d'élongation va se mettre en place (Steitz 2008).

Lors de l'élongation (Figure 5), les différents ARNt préalablement liés aux acides aminés pour lesquels ils codent sont délivrés au ribosome grâce au facteur d'élongation EF-Tu couplé au GTP. Lorsque l'ARNt correspond au codon de l'ARNm, le GTP est hydrolysé, le facteur d'élongation est alors relargué et l'ARNt se place au niveau du site A du ribosome. La formation de la liaison peptidique est alors réalisée entre les acides aminés des sites P et A par la rupture de la liaison entre l'acide aminé du site P et son ARNt au profit de la liaison entre les deux acides aminés. L'ARNt libre au niveau du site E sera relargué et pourra être recyclé pour une future utilisation. Puis, une translocation itérative du ribosome permet la lecture de chaque codon de l'ARNm et la prolongation de la chaîne polypeptidique produite (Steitz 2008).

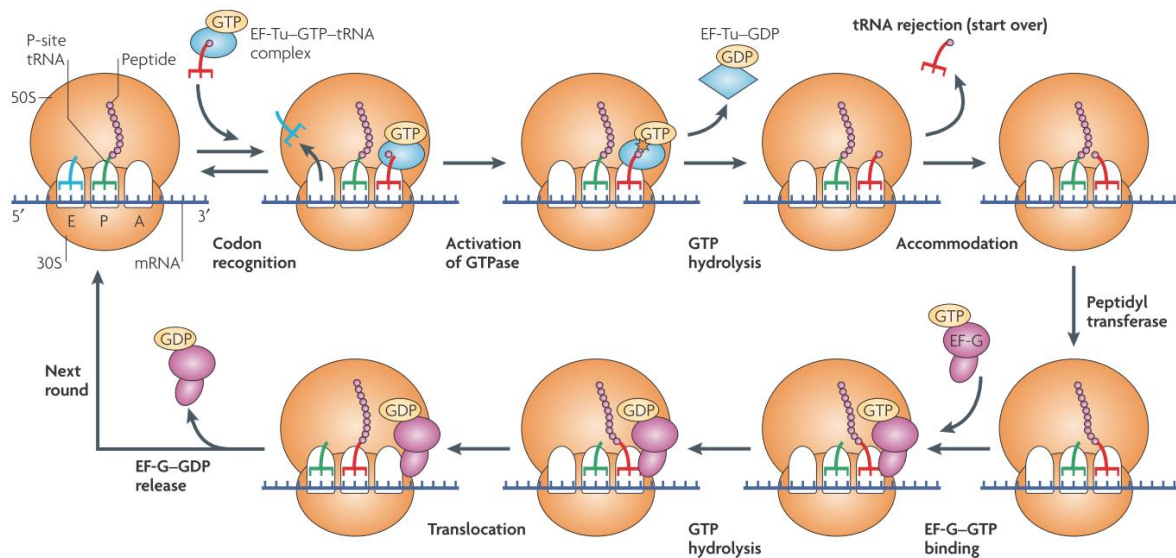


Figure 5: Mécanisme d'élongation lors de la traduction

Lors de l'élongation, le facteur d'élongation (EF-Tu) délivre l'ARNt aminoacylé au ribosome. Si l'anticodon de l'ARNt correspond, le GTP est hydrolysé et l'EF-Tu est relargué. La réaction entre les acides aminés portés par les ARNt du site A et du site P est catalysée ce qui libère l'ARNt du site P. Celui-ci sera ensuite relargué au niveau du site E. A chaque lecture d'un codon, un acide aminé est ajouté au peptide produit et il y a translocation du ribosome au niveau du codon suivant. Image extraite de la publication (Steitz 2008)

La terminaison de la traduction aura lieu au niveau du codon stop (Figure 6). Des facteurs de terminaison (appelés *Releasing Factors* ou *RF*) vont favoriser l'arrêt de la traduction au niveau de codon stop. Chez les procaryotes, il en existe 3 : RF1, RF2 et RF3. Ces facteurs ont une forme similaire à un ARNt et vont alors occuper le site A du ribosome. Le facteur RF1 reconnaît les codons stops TAG et TAA, alors que le facteur RF2 reconnaît les codons stops TGA et TAA. Leur liaison au niveau du codon stop va alors entraîner la libération du polypeptide produit par l'activité peptidyl-transférase du ribosome qui cette fois va ajouter une molécule d'eau. Le facteur RF3 couplé au GTP va ensuite favoriser la libération du facteur d'arrêt de la traduction (RF1 ou RF2). La dissociation du ribosome et de l'ARNt restant sont ensuite favorisée par les facteurs RRF et EF-G ainsi que IF3 (Steitz 2008).

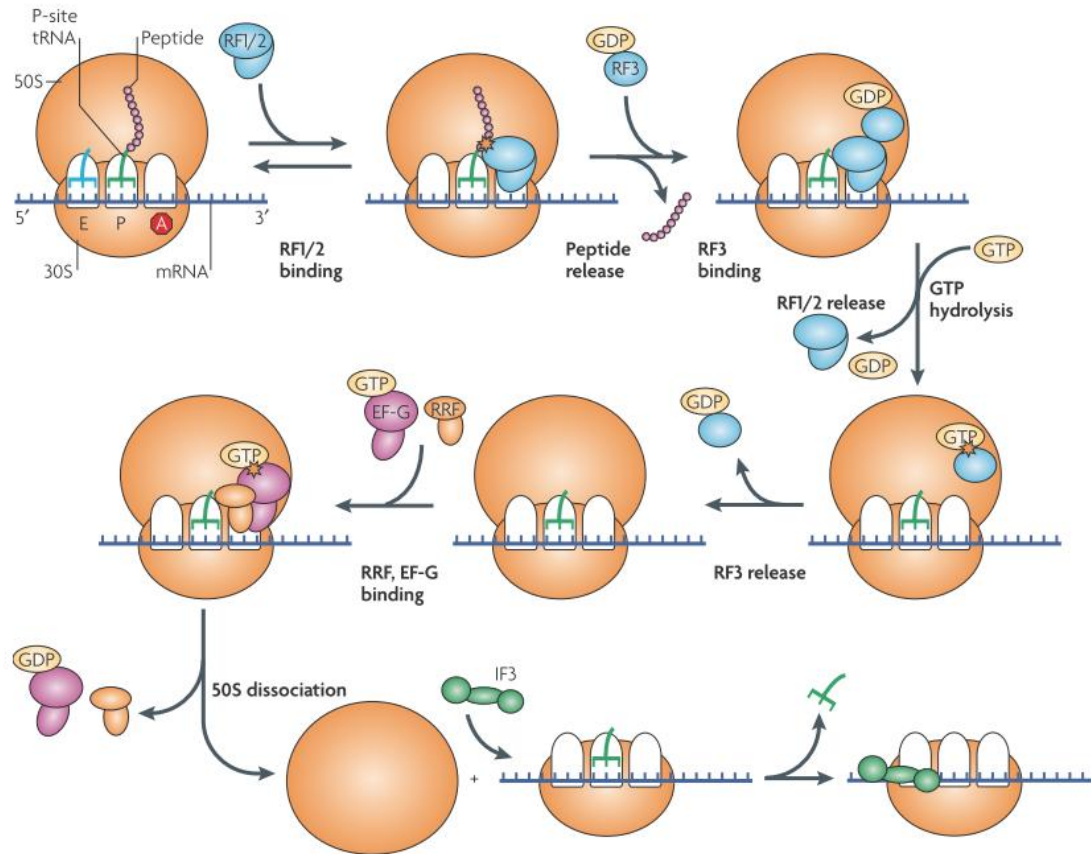


Figure 6: Mécanisme de la terminaison de la traduction

Lors de la terminaison, un facteur de terminaison (RF1 ou RF2) va spécifiquement reconnaître le codon stop. Leur liaison va permettre la libération du peptide, et le facteur de terminaison RF3 va favoriser la libération des facteurs RF1 ou RF2 puis la dissociation du ribosome. Image extraite de la publication (Steitz 2008)

II. Incorporation d'acides aminés non naturels par modifications *in situ*

Les acides aminés sont discriminés entre eux par leurs chaînes latérales qui portent différentes fonctions. Les modifications post-traductionnelles naturelles génèrent des motifs comme la phosphorylation ou la glycosylation. La modification chimique de résidus peut permettre de générer des sites de fixations que l'on peut coupler à une variété de motifs que l'on pourra alors considérer comme des acides aminés non naturels (Sakamoto and Hamachi 2019). Toutefois, ces méthodes ne permettent généralement pas une réaction sélective, donc précise et ainsi conduisent à un mélange de produits différents. En effet, les acides aminés naturels portent des fonctions de réactivités proches, il est donc difficile de cibler spécifiquement un acide aminé donné par des modifications chimiques. L'abondance relative de l'acide aminé à modifier dans la protéine sera alors critique pour la sélectivité. Par exemple, l'utilisation de la modification chimique de lysines, très abondantes à la surface, aboutit à un mélange de composés selon le nombre et la position des lysines ayant réagi. Les lysines ont cependant une réactivité plus ou moins forte selon leur accessibilité au solvant et leur pKa de leurs chaînes latérales qui seront affectées par le microenvironnement. Ainsi, les lysines proches d'acides aminés hydrophobes auront un pKa plus bas à cause d'un effet de désolvatation. Afin de contrôler plus spécifiquement les modifications apportées, le groupe de Gonçalo J.L. Bernades a donc exploité ces caractéristiques afin de cibler la lysine la plus réactive de 5 protéines grâce au sulfonyl-acrylate (Matos et al. 2018). L'utilisation de réactions chimiques impliquant des acides aminés moins abondants et ayant une réactivité particulière, comme les cystéines, permet également de générer des produits plus homogènes. Ainsi, il est par exemple

possible de modifier chimiquement les cystéines en déhydroalanine par une dialkylation du groupement thiol suivie d'une élimination en conditions acides (Chalker et al. 2011). Lorsque la protéine ne contient pas de cystéine, celle-ci peut alors être générée par mutagenèse dirigée à la position ciblée. Cependant, si la protéine contient une cystéine impliquée dans un rôle structural, d'autres méthodes impliquant des tyrosines peuvent être appliquées (Sakamoto and Hamachi 2019). Globalement, ces méthodes chimiques sont donc limitées par la séquence de la protéine native.

Théoriquement, les méthodes permettant d'incorporer des acides aminés non naturels au cours de la synthèse protéique permettent de s'affranchir de ces problématiques liées à la séquence en acides aminés naturels de la protéine à modifier.

III. Incorporation d'acides aminés non naturels *in vitro*

III.1. Incorporation par synthèse chimique

L'insertion de NAA peut être réalisée *in vitro* au cours de la synthèse chimique du polypeptide. Bien que maîtrisée, la synthèse de macromolécules comme les protéines rencontre toujours des barrières techniques. Une revue bibliographique réalisée par Kimmerlin et ses collaborateurs fait le bilan de l'évolution de la synthèse chimique des protéines (Kimmerlin and Seebach 2005). Voici brièvement les deux méthodes principales:

i) synthèse peptidique en « phase solide » : cette méthode découverte par Merrifield en 1963 a engendré un véritable essor de la synthèse peptidique par voie chimique et a été l'objet d'un prix Nobel (Merrifield 1963). La synthèse peptidique se déroule par ajouts successifs d'acides aminés dont la fonction amine primaire et les fonctions de la chaîne radicale sont protégées puis déprotégées afin d'ajouter sélectivement l'acide aminé suivant dont la fonction carboxylique primaire aura été préalablement activée pour réaliser le couplage. La synthèse sur un support fixe facilite cette succession d'étapes grâce à la simplification de la purification entre chacune de ces étapes de synthèse (par simples lavages et filtrations) et à la possibilité d'utiliser un très large excès de réactifs conduisant à des réactions de couplages quasi-totales. Cette méthode a actuellement abouti à l'automatisation de la synthèse peptidique. Néanmoins cette technique est plutôt réservée à la synthèse de peptides ou de petites protéines (50-100 acides aminés maximum) (Hermodson 1996). Pour pallier cette limitation et avoir accès à des protéines de plus grandes tailles il a été proposé de synthétiser des fragments de ces protéines puis de les assembler par des réactions qualifiées de ligation chimiosélectives afin d'obtenir les cibles. Plusieurs stratégies ainsi que de nombreuses méthodes de ligation comme « la ligation chimique native » ont été développées.

ii) synthèse peptidique par méthode semi-synthétique : Cette méthode consiste à synthétiser des peptides et à engendrer des réactions de ligation entre un peptide avec un C-term α -thioester et un autre peptide avec une cystéine en N-term et ainsi former une liaison peptidique par une réaction de ligation (appelée aussi acyl-réarrangement) (Dawson and Kent 2000). Les fonctions nécessaires (α -thioester) sont préalablement produites de manière chimique ou enzymatique (Kimmerlin and Seebach 2005). Le procédé enzymatique comprend l'utilisation d'une intéine. En effet, une intéine est une séquence qui s'excise elle-même en reliant les séquences présentes à ces extrémités par une liaison peptidique et la génération d'un thioester (Figure 7A). Un thioester généré à partir d'une protéine recombinante contenant une intéine pourra alors former une liaison peptidique avec un peptide synthétique pouvant contenir un NAA (Figure 7B) (Debelouchina and Muir 2017).

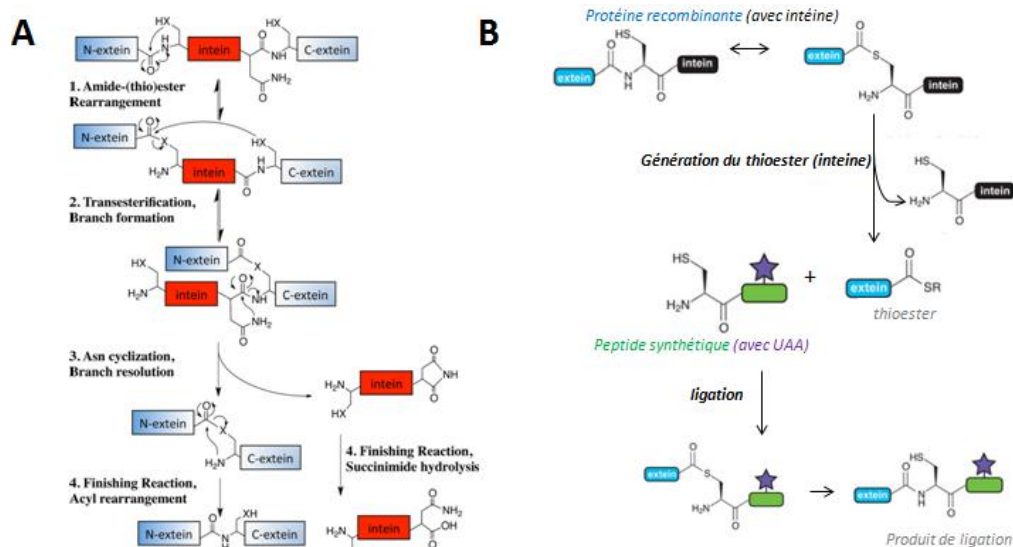


Figure 7: Mécanisme de ligation par l'utilisation d'une intéine (A); Utilisation de ce mécanisme pour produire une protéine semi-synthétique (B)

A) Le processus est initié par l'attaque nucléophile d'une sérine, thréonine ou cystéine (X est donc un atome d'oxygène ou de soufre) de la liaison peptidique ce qui génère un thioester (1). Cette réaction est suivie d'une trans-estérification entre le premier résidu de la séquence en C-terminal de l'intéine et le thioester généré ce qui libère la partie N-terminal de l'intéine (2). Le dernier résidu de l'intéine est une asparagine et par délocalisation des électrons de l'azote de cette chaîne latérale, la liaison entre l'intéine et la partie C-terminale est alors coupée (3); L'asparagine en C-terminal se cyclise et par une hydrolyse au succinimide produit une iso-asparagine tandis qu'un réarrangement conduit à la formation de la liaison peptidique entre la séquence en N-terminal et en C-terminal. Afin de relier deux fragments, l'arginine située entre l'intéine et la séquence en C-terminale peut être mutée en Alanine pour stopper la réaction à l'étape 2. Figure extraite de la publication de (Mills, Johnson, and Perler 2014).

B) L'utilisation du mécanisme de génération du thioester par une intéine peut être utilisé pour réaliser de la synthèse semi-synthétique par le biais de protéine recombinante contenant l'intéine et un peptide de synthèse. Image adaptée à partir de la publication (Debelouchina and Muir 2017)

L'incorporation d'un acide aminé non naturel au sein d'une protéine peut donc se faire pendant sa synthèse chimique totale ou partielle (Nilsson, Soellner, and Raines 2005). De manière générale, ces méthodes de synthèse sont difficiles à mettre en place pour la synthèse totale de protéines et doivent notamment faire face aux problématiques du repliement correct des protéines (Chang C Liu and Schultz 2010). Néanmoins, des protéines complètes ont pu être produites par cette méthode comme cela a pu être démontré par exemple par la production de glycoprotéines telles que l'erythroprotéine (environ 30-37 kDa), l'interleukine IL6 (environ 30 kDa) ou encore l'interleukine IL-8 (dimère de 59kDa et de 67kDa) (Carlo and Yasuhiro 2018).

III.2. Incorporation par pré-couplage de l'ARNt

D'autres approches biosynthétiques ont aussi été développées pour de l'expression *in vitro* de protéines contenant des acides aminés non naturels (NAA). Comme nous l'avons vu précédemment (section I.1), les ARN de transfert (ARNt) permettent la correspondance entre le code génétique et la séquence primaire de la protéine en associant un acide aminé au codon correspondant. Des ARNt ayant été chargés artificiellement avec un acide aminé non naturel peuvent être utilisés lors de la traduction par le ribosome. La production de la protéine modifiée peut alors se réaliser *in vitro* grâce à des mélanges réactionnels contenant les différents composants nécessaires à la traduction, auxquels on aurait ajouté cet ARNt relié à l'acide aminé non naturel (Hecht 2009). Ces mélanges peuvent être indéfinis lorsqu'ils sont dérivés de lysats cellulaires, ou définis lorsqu'il s'agit d'un mélange des différents éléments purifiés (Rolf, Rosenthal, and Lütz 2019). L'expression de

protéines *in vitro* peut être avantageux lors de la production de protéines toxiques ou très hydrophobes (Rolf, Rosenthal, and Lütz 2019). Cependant, comparativement à des méthodes de production *in vivo*, cette méthode de production a des limitations dont principalement une efficacité de production plus faible, mais aussi des difficultés pour réaliser la montée en échelle de production dues au coût élevé des mélanges rationnels ou à l'irrégularité des lysats cellulaires (Rolf, Rosenthal, and Lütz 2019). Toutefois, pour des petites échelles de production, l'utilisation de mélanges définis permet de s'affranchir de certaines limitations en éliminant des composés cellulaires comme des facteurs d'arrêts de la transcription ou des paires spécifiques d'aaRS-ARNt afin de détourner n'importe quel codon pour l'incorporation d'un acide aminé non naturel (Goto, Katoh, and Suga 2011).

Pour charger ces acides aminés non naturels (NAA) sur l'ARNt, des nucléotides sont liés chimiquement au NAA puis chargés enzymatiquement à des ARNt tronqués en 3' (Figure 8) (Robertson, Ellman, and Schultz 1991). Cependant, les étapes de synthèse sont techniquement difficiles et comportent différentes étapes utilisant des acides aminés protégés avec un rendement global faible (généralement moins de 10%) (Giel-Pietraszuk and Barciszewski 2006).

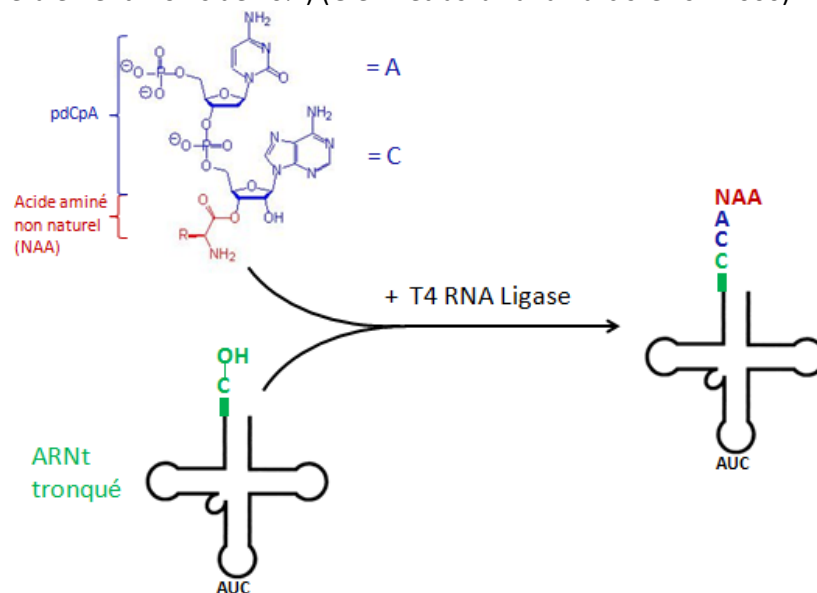


Figure 8: Ligation d'un acide aminé non naturel sur un ARN de transfert (ARNt) à l'aide d'une T4 RNA ligase
L'ARNt est produit par une transcription d'une version tronquée. En effet, il lui manque deux bases phosphatées (dC et A) au niveau de son bras accepteur. L'acide aminé non naturel à incorporer est chimiquement synthétisé relié à ces nucléotides phosphatés (pdCpA). Ces deux partenaires sont ensuite couplés à l'aide de la T4 RNA ligase.

Face à ces difficultés techniques, les propriétés de certains ARN catalytiques ont également été exploitées. En effet, en 1982 lors de travaux réalisés sur la transcription, Thomas Cech et Sidney Altman ont découvert que certains ARN (ultérieurement appelés ribozymes) étaient dotés d'activités catalytiques ce qui leur a valu le prix Nobel de chimie en 1989 (Cech and Altman 1989). A la suite de cette découverte, de nombreux ARN catalytiques ont été découverts et des méthodes de sélection ont été développées afin de générer des ARN catalytiques possédant de nouvelles activités (Walter and Engelke 2002). Ainsi, en 2003, Suga et ses collaborateurs ont mis au point des ribozymes artificiels appelés flexizymes capables de réaliser l'aminocyclation des ARNt (Murakami, Saito, and Suga 2003). Ils reconnaissent l'extrémité de l'ARNt en 3' (RCCA-3') avec en position 73 (désigné ici par R) une Guanidine, une Adénine ou un Uracile et un groupe partant lié à l'acide aminé à charger sur l'ARNt (Figure 9) (Ohuchi, Murakami, and Suga 2007). Afin d'élargir les outils disponibles pour incorporer un large panel d'acides aminés, plusieurs types de flexizymes ont été développés et sont

spécifiques envers leur groupe partant. Ainsi des flexizymes (eFx) ont été conçus pour le transfert d'acides aminés aromatiques sont spécifiques pour des acides aminés sous forme d'esters cyanométhylés, alors que ceux (dFx) plus spécialisés pour des acides aminés non aromatiques sont spécifiques pour des acides aminés sous forme d'esters 3,5-dinitrobenzyles. Cette dernière classe de flexizymes (dFx) permet l'incorporation d'un très large panel d'acides aminés. Cependant, dans le cas d'acides aminés très encombrants, une classe (eFx) reconnaissant des acides aminés sous forme d'ester 4-chlorobenzyle est généralement efficace. Enfin dans le cas d'une baisse de solubilité de l'acide aminé à cause du groupe partant, une dernière classe de flexizyme (aFx) reconnaît des acides aminés sous forme de thioesters 4-[(2-aminoéthyl)carbamoyle]benzyles (Figure 9) (Goto, Katoh, and Suga 2011). Bien que variable selon l'acide aminé à charger, le rendement de l'acylation de l'ARNt est généralement assez fort (10% à > 90%). De plus, comparativement à l'aminoacylation semi-enzymatique des ARNt, l'utilisation des flexizymes est techniquement simple car réalisée par la simple incubation dans la glace du flexizyme, de l'ARNt et de l'acide aminé à charger (Goto, Katoh, and Suga 2011).

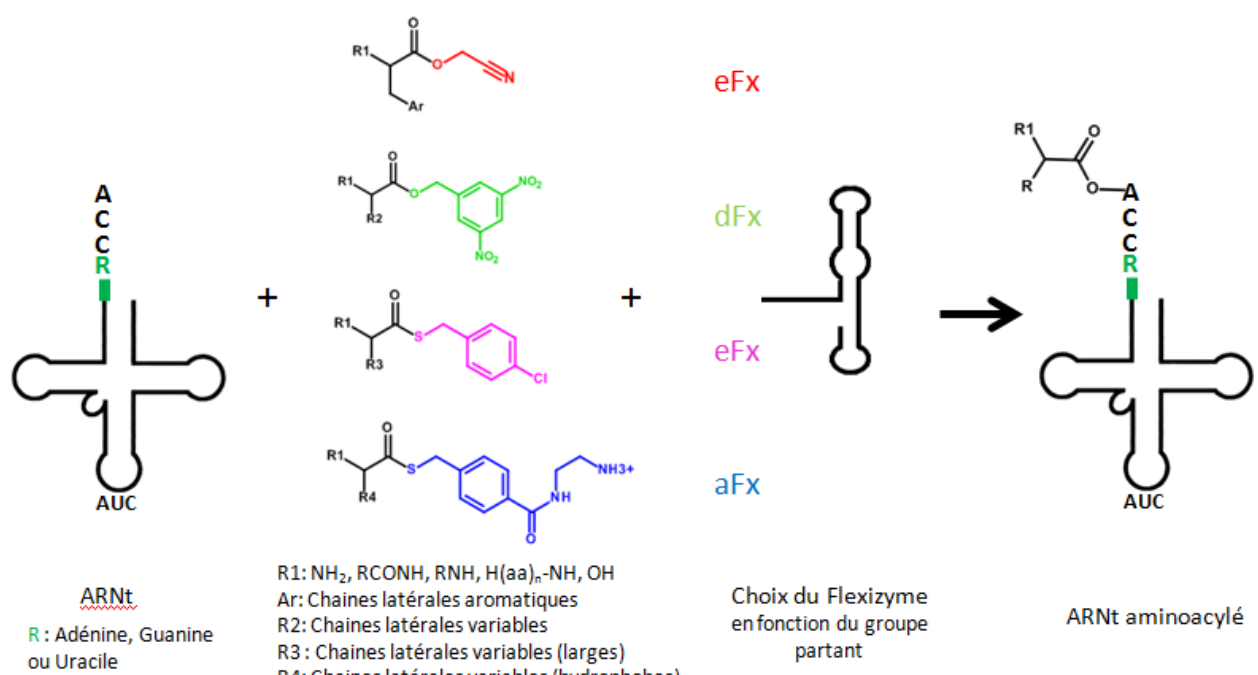


Figure 9: Utilisation de flexizymes en fonction de l'acide aminé à charger et de son groupe partant schéma réalisé à partir des informations de la publication (Goto, Katoh, and Suga 2011).

L'utilisation d'ARNt pré-couplé à l'acide aminé non naturel est également possible *in vivo* par la micro-injection de celui-ci au sein d'ovocytes (Monahan, Lester, and Dougherty 2003). Néanmoins cette technique est limitée par la délivrance des ARNt aminoacylés qui ne peut pas être réalisée de manière continue et la nature invasive des techniques de microinjections ou transfections (Q. Wang, Parrish, and Wang 2009). Cependant, les ovocytes de grenouilles *Xenopus laevis* sont particulièrement utilisés lors de recherches sur des mécanismes cellulaires fondamentaux car ces mécanismes sont très conservés par rapport à l'Homme et ce matériel biologique est peu coûteux (Nikos 2012). L'utilisation d'acides aminés non naturels fluorescents a notamment permis d'étudier le fonctionnement de canaux ioniques membranaires (Kalstrup and Blunck 2017).

IV. Incorporation d'acides aminés non naturels *in cellulo*

L'incorporation d'acides aminés non naturels *in vivo* est également essentiellement basée sur la *misacylation* d'un ARNt, afin qu'au cours de la traduction, le ribosome réalise son incorporation dans la protéine synthétisée. Au sein de la cellule, le chargement de l'ARNt est réalisé par les aminoacyl ARNt synthétases (aaRS). Nous allons donc tout d'abord décrire le fonctionnement de ces enzymes avant de voir les techniques détournant leurs propriétés *in vivo*.

IV.1 Généralités sur les aminoacyl ARNt synthétase

IV.1.1 Mécanisme réactionnel catalysé par les aminoacyl ARNt synthétases

Une aminoacyl ARNt synthétase est spécifique d'un acide aminé donné et du ou des ARNt correspondants. Le transfert de l'acide aminé sur l'ARNt va être catalysé en deux étapes (Figure 10). Tout d'abord, l'acide aminé (Aa) à charger est activé (sous la forme AMP-Aa) grâce à une molécule d'ATP et à la formation de pyrophosphate (PPi). Puis une attaque nucléophile à partir de la base 76 de l'ARNt engendre la liaison de celui-ci avec l'acide aminé et la libération d'AMP. La stéréochimie de cette dernière étape varie selon le type d'aaRS. En effet, les aaRS peuvent être phylogénétiquement classées selon deux catégories (classe I ou classe II) ((Perona and Gruic-Sovulj 2014), Tableau 2).

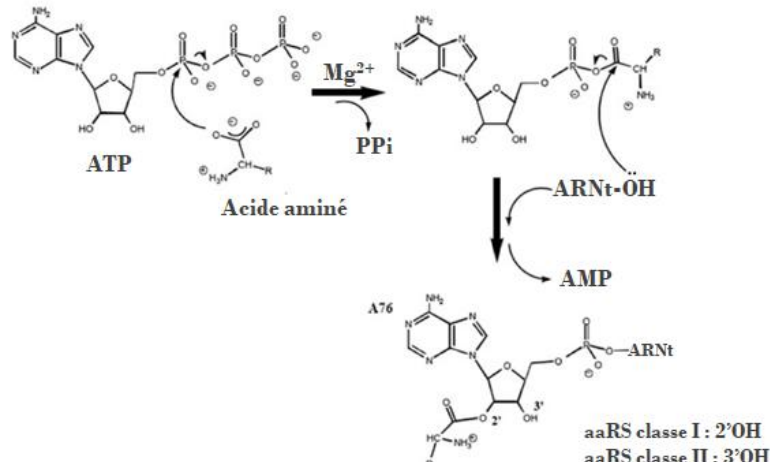


Figure 10: Mécanisme réactionnel catalysé par l'acide aminé sur l'ARNt synthétase

L'acide aminé sur l'ARNt synthétase catalyse la liaison de l'acide aminé sur l'ARNt selon ces deux étapes: tout d'abord l'hydrolyse de l'ATP permet l'activation de l'acide aminé (1), ensuite une attaque nucléophile de l'ARNt sur l'acide aminé permet le chargement de celui-ci et la libération de l'AMP (2) ; image reproduite à partir de la publication (Perona and Gruic-Sovulj 2014).

IV.1.2 Structures des aminoacyl ARNt synthétases

IV.1.2.1 Reconnaissance

En dehors de la reconnaissance de l'ATP, les aaRS sont spécifiques pour deux types de substrats : l'acide aminé et une série d'ARNt isoaccepteurs. Les ARNt possèdent des structures globalement proches en forme de "L". Des motifs appelés déterminants permettent la reconnaissance spécifique de ces ARNt par l'aaRS (

Tableau 1) (Giege, Sissler, and Florentz 1998). Ils sont constitués soit de modifications post-traductionnelles sur des bases de l'ARNt, soit de nucléotides identifiés. Les déterminants sont généralement présents au niveau du bras accepteur ou de l'anticodon (Figure 11), bien qu'il existe quelques exemples au niveau du corps de l'ARNt. Par exemple, la base discriminante en position 73 ainsi que l'appariement au niveau des bases 1 et 72 sont des déterminants fréquents (Giegé and Eriani 2014). Concrètement, ces déterminants peuvent être positifs c'est-à-dire favoriser la reconnaissance de l'ARNt par son aaRS correspondante ou négatifs et réduire la reconnaissance de

l'ARNt par une aaRS donnée. Leurs effets se traduisent par des interactions avec des acides aminés de l'aaRS ou par un rôle indirect en modifiant la structure de l'ARNt et donc son positionnement dans le site actif de l'aaRS (Giegé and Eriani 2014). En effet, le bon positionnement de l'extrémité 3'-A76 de l'ARNt dans le site actif serait essentiel pour l'efficacité du transfert (Perona and Guic-Sovulj 2014).

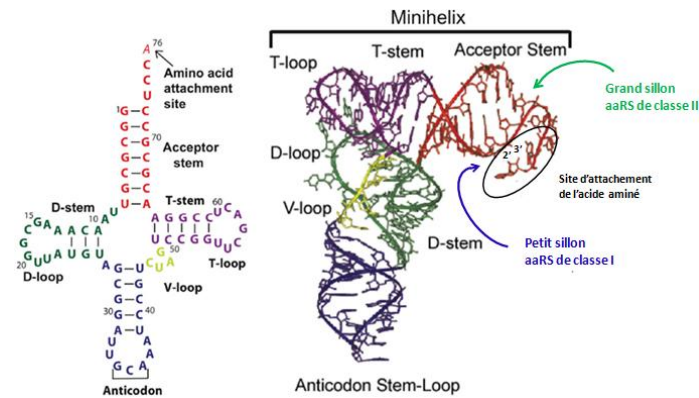


Figure 11: Structure de l'ARNt

L'ARNt forme une structure secondaire en forme de trèfle (à gauche sur la figure) qui structurellement se caractérise par une forme générale en L (à droite sur la figure). Les différentes parties de l'ARNt ont été identifiées selon des boucles (loop sur la figure). Le bras accepteur (acceptor stem) correspond à celui sur lequel l'acide aminé sera greffé. Nous pouvons noter que deux types d'aaRS existent (section IV.1.2.2) et reconnaissent l'ARNt selon le petit ou le grand sillon (indiqué à droite sur la figure). Image modifiée à partir de la publication (Hou 2010)

Tableau 1: Quelques exemples de déterminants au niveau de l'ARNt

La reconnaissance de l'ARNt par l'aaRS est liée à quelques éléments clés présents au niveau de celui-ci. Ces éléments sont appelés déterminants. Ils peuvent avoir un effet positif, c'est-à-dire favoriser leur reconnaissance par une aaRS ou négatif, c'est-à-dire limiter ou même empêcher leur reconnaissance par une autre aaRS. Ces informations sont tirées de la publication (Giegé and Eriani 2014)

Déterminant	ARNt/organisme	Effet
Anticodon entier	ARNt ^{Cys}	Positif : CysRS, GlnRS, IleRS, MetRS, TrpRS, AsnRS, AspRS, HisRS, LysRS, PheRS, ThrRS
G₇₃	ARNt ^{Arg} , ARNt ^{Gln} , ARNt ^{Trp} , ARNt ^{Asp} , ARNt ^{Asn} , ARNt ^{Ser} / Tous	Positif : ArgRS, GlnRS, TrpRS, AspRS, AsnRS, SerRS
Appariement U1-A72	ARNt ^{Gln} / Tous	Positif : GlnRS
C₃₅	ARNt ^{Arg} , ARNt ^{Cys} , ARNt ^{Trp} , ARNt ^{Gly} / Tous	Positif: ArgRS, CysRS, TrpRS, GlyRS
A₉	ARNt ^{Trp} / Tous	Positif : TrpRS
Lysidine₃₄ (base C modifiée) position 34	ARNt ^{Ile} / <i>E.coli</i>	Négatif sur la reconnaissance par MetRS
m¹G₃₇ (base G méthylée) position 37	ARNt ^{Asp} / levure	Négatif sur la reconnaissance par ArgRS
A₃₆	ARNt ^{Arg} / <i>E.coli</i>	Négatif sur la reconnaissance par TrpRS
Misappariement G₃-U₇₀	ARNt ^{Ala} / levure	Négatif sur la reconnaissance par ThrRS

La reconnaissance des acides aminés doit également être très fine. En effet, les erreurs de chargement auraient des conséquences importantes sur le fonctionnement de la cellule en conduisant à la production de protéines modifiées ce qui pourrait entraîner la mort cellulaire (Perona

and Gruic-Sovulj 2014). La spécificité des aaRS envers un acide aminé est en général très forte. Ainsi, même des acides aminés ayant un encombrement similaire comme l'asparagine et l'acide aspartique ou la tyrosine et la phénylalanine sont spécifiquement reconnus par leur aaRS respectives. Certains acides aminés clés de l'aaRS permettent alors cette haute spécificité. Par exemple, au niveau de l'aspartyl ARNt synthétase (AspRS), une lysine en position 198 interagit avec le groupement carboxylate de la chaîne latérale de l'acide aspartique, favorisant ainsi la reconnaissance de ce substrat. Des travaux ont d'ailleurs pu montrer que la mutation de cette lysine en leucine ou glycine, entraîne une perte de la spécificité pour l'acide aspartique avec une différence d'énergie de liaison de l'AspRS avec l'acide aspartique ou l'asparagine qui devient nulle (Archontis et al. 1998). De plus, certaines aaRS ont également une précision renforcée par des fonctions de relecture qui leur permettent d'hydrolyser des acides aminés activés ou transférés par erreur sur un ARNt (Figure 12). Ces fonctions sont présentes au niveau de résidus très conservés au sein du site actif ou par la présence d'un site d'édition distinct du site actif permettant l'hydrolyse d'acides aminés déjà transférés sur l'ARNt (Ling, Reynolds, and Ibba 2009). Ainsi, au niveau des aaRS, l'aminocyclation des ARNt peut ainsi être maîtrisée à l'aide de deux cribles : le site actif qui ne prendra en charge qu'une gamme très restreinte d'acides aminés, suivi d'une vérification plus stricte par un site d'édition. Par exemple, la Phenylalanyl ARNt synthétase (PheRS) peut transférer à la fois la phénylalanine et la tyrosine sur l'ARNt^{Phe}. Cependant, le site d'édition permettra l'hydrolyse spécifique de la tyrosine grâce à une fonction carboxylate de l'acide aminé E334 formant spécifiquement une liaison hydrogène avec la tyrosine qui sera donc stabilisée au sein du site d'édition (Sasaki et al. 2006).

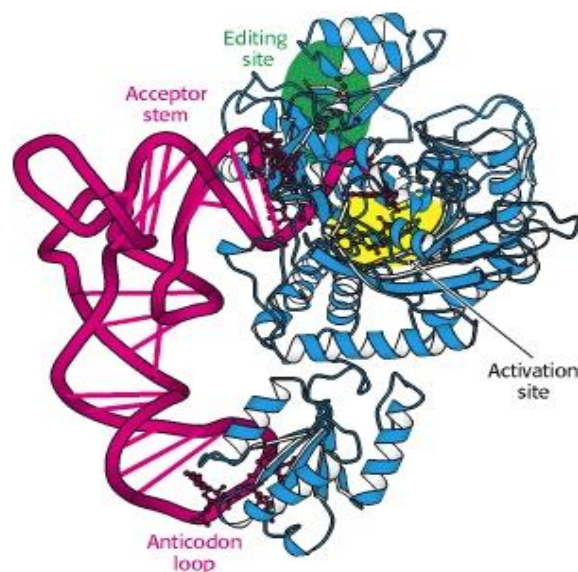


Figure 12: Structure d'une Thréonine ARNt synthétase en complexe avec son ARNt.

Cette aaRS présente un site actif ainsi qu'un site d'édition, ce qui permet de garantir la correcte aminocyclation de l'ARNt. Cette image est extraite de la publication (Berg JM, Tymoczko JL, and Stryer L 2002)

Après l'aminocyclation de l'ARNt, celui-ci sera délivré au ribosome par l'intermédiaire du facteur d'élongation (EF-Tu). Ce dernier reconnaît à la fois l'acide aminé et l'ARNt et possède une affinité équivalente pour tous les ARNt correctement couplés. Ainsi, si le facteur d'élongation reconnaît un acide aminé avec une faible affinité, il aura une affinité plus forte pour l'ARNt correspondant afin de contrebalancer cet écart et avoir une affinité équivalente pour chaque ARNt couplé. De plus, il a pu être montré que le facteur d'élongation avait une affinité faible pour certains ARNt misacylés (incorrectement couplés, par exemple : Glu-ARNt^{Gln}, Asp-ARNt^{Asn}) ce qui pourrait permettre de renforcer la fiabilité du décodage du code génétique (Ling, Reynolds, and Ibba 2009).

IV.1.2.2 Caractéristiques structurales des Aminoacyl ARNt synthétases selon leur classe

Les aminoacyl ARNt synthétases ont été phylogénétiquement classées selon deux principales classes contenant chacune 10 aaRS (Tableau 2). Seule la Lysyl-ARNt synthétase (LysRS) est plus difficilement classée car elle possède une caractéristique structurale de classe I chez les eucaryotes et chez la plupart des bactéries mais de type II chez les archées et quelques bactéries (Alexandre Ambrogelly 2000). L'origine phylogénétique commune à ces deux classes n'a pas été déterminée, mais les gènes des aaRS de classe I et II pourraient être la conséquence d'une évolution à partir de gènes bidirectionnels ou de différents cadres de lectures (Pak, Kim, and Burton 2018).

Tableau 2: Classification des aminoacyl ARNt synthétases

Les aminoacyl ARNt synthétases (aaRS) peuvent être classées phylogénétiquement selon deux classes. Les aaRS de classe I sont généralement monomériques ou dimériques alors que les aaRS de classes II sont dimériques ou tétramériques. De plus, certaines aaRS des deux classes possèdent une fonction d'édition, ce qui leur permet d'avoir une meilleure fiabilité lors de l'aminocyclation de l'ARNt. Tableau reproduit à partir de la publication de (Perona and Gruic-Sovulj 2014)

AaRS de classe I	Structure quaternaire	Fonction d'édition	AaRS de classe II	Structure quaternaire	Fonction d'édition
Sous classe IA			Sous classe IIA		
MetRS	α, α_2	Oui	SerRS	α_2	Oui
LeuRS	α	Oui	ProRS	α_2	Oui
IleRS	α	Oui	ThrRS	α_2	Oui
ValRS	α	Oui	GlyRS	α_2	Non
Sous classe IB			HisRS	α_2	Non
CysRS	α, α_2	Non	Sous classe IIB		
GlnRS	α	Non	AspRS	α_2	Non
GluRS	α	Non	AsnRS	α_2	Non
Sous classe IC			LysRS	α_2	Oui
TyrRS	α	Non	Sous classe IIC		
TrpRS	α_2	Non	PheRS	$(\alpha\beta)_2, \alpha$	Oui
Sous classe ID			GlyRS	$(\alpha\beta)_2$	Non
ArgRS	α_2	Non	AlaRS	α_2, α	Oui
Sous classe IE			SepRS	α_4	Non
LysRS	α	Non	PyIRS	α_2	Non

Les principales différences structurales entre ces deux classes sont répertoriées dans le

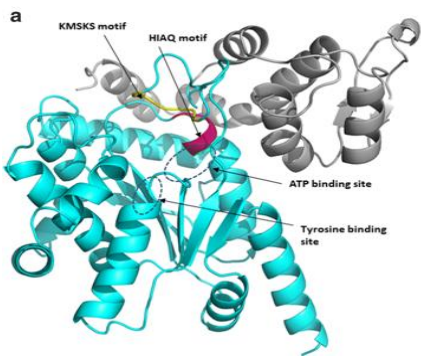
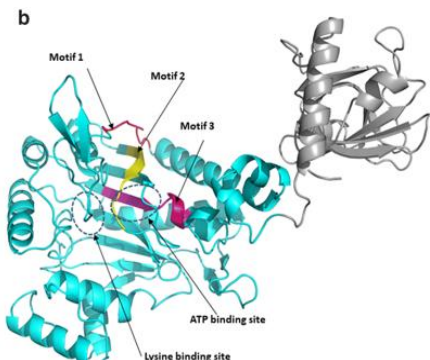
Tableau 3.

Tout d'abord, au niveau de leur structure globale, les aaRS de classe I sont principalement monomériques avec quelques exceptions homodimériques, alors que les aaRS de classe II sont principalement des dimères ou des tétramères. De plus, les aaRS de classe I possèdent un site de type "Rossmann" (c'est-à-dire un feuillet β formé par 6 brins en parallèle liés à deux paires d'hélices α , dans l'ordre β - α - β) permettant de lier l'ATP et comprenant deux séquences consensus très conservées (HIGH et KMSKS) (Perona and Gruic-Sovulj 2014). Leur site actif est plutôt ouvert et ces aaRS lient donc généralement des acides aminés encombrants (Rajendran et al. 2018). La structure des aaRS de classe II correspond à des feuillets β anti-parallèles, avec trois motifs principaux conservés (G ϕ X ϕ X ϕ XP ϕ ϕ , (F/Y/H)RX(E/D)(4 à 12 résidus)(R/H)XXXFXXX(D/E), et G ϕ G ϕ G ϕ (D/E)R ϕ ϕ ϕ ϕ avec ϕ pour un acide aminé hydrophobe et x n'importe quel acide aminé). Le motif 1 est impliqué dans la dimérisation de l'enzyme tandis que le motif 3 se lie à l'ATP et que le motif 2 est impliqué dans le couplage de l'ATP, l'acide aminé et l'ARNt (Giegé and Eriani 2014). Au niveau du site actif, l'ATP est notamment reconnu sous une forme coudée caractéristique de cette classe.

Comme nous l'avons vu précédemment (section IV.1.1), les aaRS de classe I effectuent l'aminocyclation de l'ARNt au niveau de l'extrémité 2'OH, alors que les aaRS de classe II l'effectuent au niveau de l'extrémité 3'OH. Il s'agit d'une conséquence du mode d'interaction de celles-ci avec l'ARNt qui s'effectue principalement par le biais du petit sillon de ce dernier pour les aaRS de classe I et par le grand sillon pour les aaRS de classe II (Figure 11)(Rajendran et al. 2018).

Au niveau cinétique, l'étape limitante pour la classe II serait plutôt l'activation de l'acide aminé par sa liaison à l'AMP (Perona and Gruic-Sovulj 2014) tandis que pour la classe I, il s'agit de l'étape du relargage de l'ARNt aminocyclé. Cette étape pourrait alors être facilitée par la formation d'un complexe avec le facteur d'élongation EF-Tu (C. M. Zhang et al. 2006). Ainsi, par exemple, l'affinité de la cystéinyl ARNt synthétase (CysRS, aaRS de classe I) est 10 fois supérieure pour l'ARNt aminocyclé que pour l'ARNt seul. Cette forte reconnaissance entraîne un relargage lent du produit. Cependant, le facteur d'élongation EF-Tu a une affinité 300 fois supérieure pour l'ARNt aminocyclé avec la cystéine, ce qui permet alors de faciliter ce relargage de l'ARNt aminocyclé par l'aaRS (C. M. Zhang et al. 2006).

Tableau 3: Différences caractéristiques des aaRS selon leur classe

Caractéristiques	aaRS de classe I	aaRS de classe II
Structure quaternaire	Monomères, parfois dimères	Dimères ou plus
Organisation globale du site	Rossmann (Feuillets β parallèles)	Feuillets β anti-parallèles
Motifs conservés	HIGH KMSKS	-G ϕ XX ϕ XXP $\phi\phi$ -(F/Y/H)RX(E/D)(4 à 12 résidus)(R/H)XXXFXXX(D/E) -G ϕ G ϕ G ϕ (D/E)R $\phi\phi\phi\phi\phi$ (avec ϕ acide aminé hydrophobe et X n'importe quel acide aminé)
Conformation de l'ATP au sein du site	Etendue	Coudée
Liaison à l'ATP	Principalement par le petit sillon	Principalement par le grand sillon
Aminoacylation de l'ARNt	En 2'-OH	En 3'-OH
Etape limitante	Relargage de l'ARNt aminocyclé	Activation de l'acide aminé
Exemple de structures, images extraites de la publication (Nyamai and Tastan Bishop 2019)	TyrRS, issue du parasite de la malaria (<i>Plasmodium</i>)	LysRS, issue du parasite de la malaria (<i>Plasmodium</i>)
Les domaines représentés en gris correspondent au domaine de fixation à l'anticodon.		

IV.2 Incorporation des acides aminés non naturels par méthode résidu spécifique

IV.2.1. Principe de la méthode d'incorporation résidu spécifique

L'incorporation des acides aminés non naturels de manière résidu-spécifique permet de remplacer un type d'acide aminé (par exemple toutes les Méthionines) par des acides aminés non naturels dans une protéine (Singh-Blom, Hughes, and Ellington 2013). Cette incorporation utilise les propriétés de reconnaissance des aaRS. En effet, comme nous l'avons vu les aaRS ont une reconnaissance spécifique des acides aminés naturels les uns par rapport aux autres. Ainsi, les aaRS ont spécifiquement développé des fonctions d'édition contre des acides aminés naturels proches de leurs substrats (section IV.1.2.1). Cependant, si un acide aminé non naturel a une structure proche d'un acide aminé naturel, il pourrait être pris en charge par une aaRS. Ainsi, "par erreur" un NAA peut être inséré dans la séquence protéique à la place d'un acide aminé naturel duquel il est structuralement proche.

IV.2.2. Méthode permettant l'incorporation résidu spécifique *in vivo*

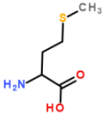
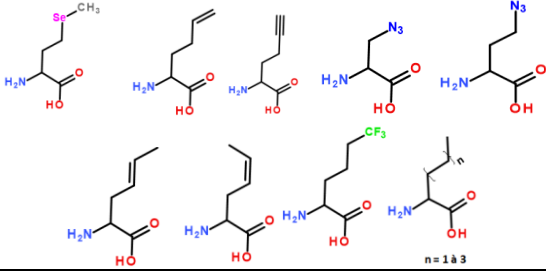
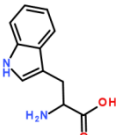
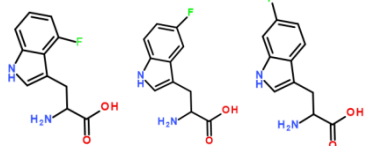
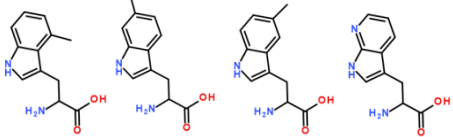
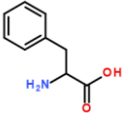
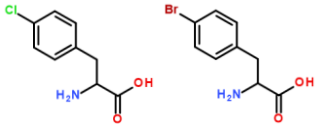
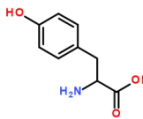
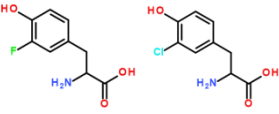
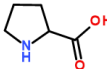


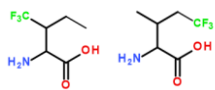
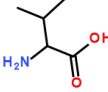
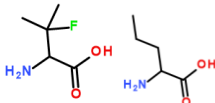
Malgré une structure analogue à l'acide aminé naturel, l'aaRS a généralement une affinité plus faible pour l'acide aminé non naturel. Un mélange de protéine sauvage et de la protéine modifiée avec l'acide aminé non naturel sera alors produit avec un rapport plus ou moins important selon cette différence d'affinité. Afin d'avoir une protéine modifiée pure, la production de celle-ci doit alors s'effectuer en l'absence de l'acide aminé naturel remplacé. Une souche auxotrophique pour celui-ci doit donc être utilisée pour garantir son absence. Cependant, l'acide aminé non naturel ne sera pas uniquement incorporé au sein de la protéine cible mais également au sein de nombreuses protéines endogènes. Ces modifications majeures peuvent donc altérer la croissance bactérienne. La souche est donc préalablement amplifiée afin de fournir le matériel biologique nécessaire. Celui-ci peut ensuite être cultivé dans un milieu où l'acide aminé naturel a été remplacé par l'acide aminé non naturel et où la production de la protéine cible sera induite (Singh-Blom, Hughes, and Ellington 2013).

Le taux d'incorporation du NAA et donc le taux d'expression de la protéine modifiée peuvent être augmentés par la surexpression de l'aaRS (Kiick, van Hest, and Tirrell 2000; Kiick and Tirrell 2000) ou la modification de son site actif par mutagenèse dirigée (Datta et al. 2009; Kast and Hennecke 1991) afin de compenser une éventuelle faible reconnaissance de l'acide aminé non naturel. De manière générale, cette technique est surtout limitée par la structure du NAA étant donné qu'il doit avoir une structure analogue à celle de l'acide aminé naturel remplacé. Cependant la diversité des NAA utilisables a été étendue par l'ingénierie des aaRS (Hadd and Perona 2014). Actuellement, différents analogues ont pu être incorporés par cette méthode dont des analogues de méthionine, d'isoleucine, de phénylalanine, de proline, de tryptophane, de tyrosine, et de valine (Tableau 4).

Au niveau des autres limitations de cette technique, la nécessité de la déplétion de l'acide aminé naturel au cours du procédé le complique notamment la production en fermenteur, ce qui rend la production à grande échelle difficile (Chang C. Liu and Schultz 2010). Par ailleurs, l'impact de l'acide aminé à incorporer sur la croissance bactérienne ainsi que l'utilisation de souches auxotrophes peut limiter le rendement de production de la protéine d'intérêt (Chang C. Liu and Schultz 2010). Globalement, cette méthode permet d'incorporer efficacement plusieurs acides aminés non naturels au sein d'une protéine. Elle est donc particulièrement appropriée pour l'incorporation multiple d'un acide aminé non naturel à plusieurs positions. Cependant, sauf s'il est

possible de muter considérablement la protéine d'intérêt, il n'y a pas de contrôle sur le site que l'on veut modifier. Les approches d'incorporation des acides aminés non naturels de manière site spécifique permettent alors de répondre à ces besoins complémentaires (section suivante IV.3).

Tableau 4: Exemples d'analogues d'acides aminés incorporés par la méthode résidu spécifique

Acide aminé remplacé	Analogues	Références bibliographiques
<p>Méthionine (Met)</p> 	<p>Sélenométhionine, analogues alcène, alcyne, azoture, fluoré et autre :</p> 	<p>(Cowie and Cohen 1957) (Kiick and Tirrell 2000; van Hest and Tirrell 1998; Kiick et al. 2002)</p>
<p>Tryptophane</p> 	<p>Analogues Fluorés :</p>  <p>Autres analogues:</p> 	<p>(Singh-Blom, Hughes, and Ellington 2013)</p>
<p>Phénylalanine</p> 	<p>Dérivés halogénés :</p> 	<p>(Ibba and Hennecke 1995)</p>
<p>Tyrosine</p> 	<p>Analogues halogénés :</p> 	<p>(Minks et al. 2000)</p>
<p>Proline</p> 	<p>Cycle différent :</p> 	<p>(Pine 1978)</p>
<p>Isoleucine</p> 	<p>Analogues Fluorés :</p> 	<p>(P. Wang, Tang, and Tirrell 2003)</p>
<p>Valine</p> 	<p>Dérivé Fluoré et autre dérivé</p> 	<p>(Hartman et al. 2007)</p>

IV.3 Incorporation des acides aminés non naturels de manière site spécifique

IV.3.1. Principe de la méthode d'incorporation site spécifique

L'incorporation de façon site spécifique consiste à remplacer un acide aminé naturel par un acide aminé non naturel (NAA) à une position précise sans changer le reste de la séquence protéique. Pour cela, il faut un couple aaRS/ARNt qui permette cette introduction spécifique du NAA. Tout d'abord, l'ARNt ne doit pas posséder un anticodon déjà utilisé pour l'incorporation d'un autre acide aminé afin d'éviter toute incorporation à une position non voulue. Un codon sera détourné et il s'agit très souvent d'un codon stop, principalement le codon ambre TAG dont le taux d'utilisation est bas avec 93% des gènes terminés par TAA ou TGA chez *E. coli* (Chang C Liu and Schultz 2010). De plus, l'aaRS et l'ARNt doivent être orthogonaux au système de traduction endogène (Figure 13). C'est-à-dire que l'ARNt ne doit pas être pris en charge par des aaRS endogènes et que l'aaRS orthogonale ne doit pas prendre en charge ni des ARNt endogènes, ni des acides aminés naturels. L'absence de réactivité croisée permet alors d'ajouter une voie indépendante qui permet de coder spécifiquement l'acide aminé non naturel (Chang C Liu and Schultz 2010). L'aaRS va spécifiquement greffer l'acide aminé non naturel au niveau de l'ARNt orthogonal qui sera ensuite pris en charge lors de la traduction pour permettre l'incorporation du NAA au sein de la protéine. La position à laquelle il sera incorporé aura alors été préalablement déterminée par un codon stop généré par mutagenèse dirigée au niveau de la séquence de la protéine d'intérêt (Figure 14).

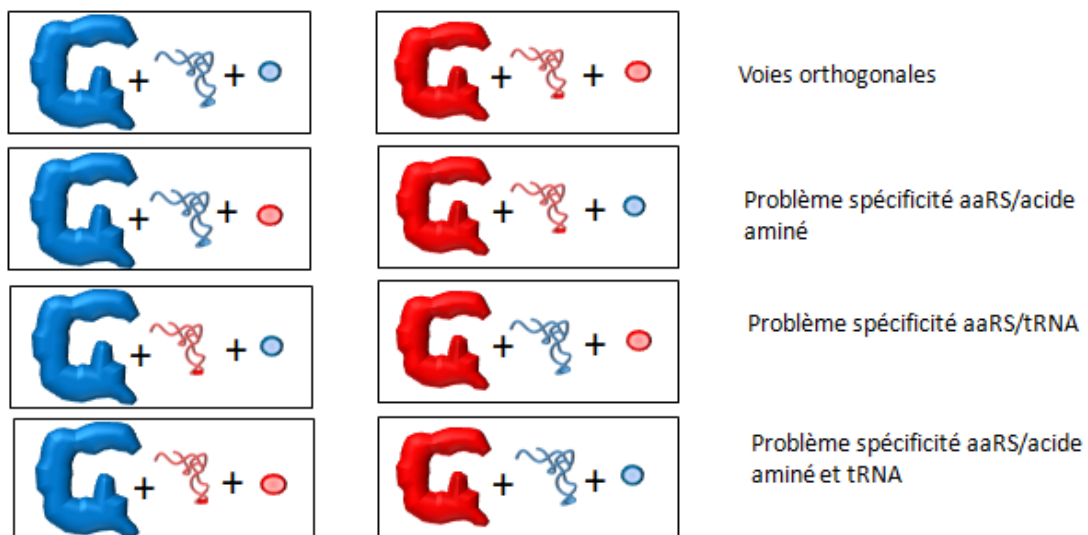


Figure 13: Présentation de la signification d'une voie orthogonale

Les voies endogènes sont représentées en bleu et la voie ajoutée en rouge. Afin d'être orthogonale au système endogène, celle-ci ne doit pas interagir avec les voies déjà en place. C'est-à-dire que l'ARNt ne doit pas être pris en charge par des aaRS endogènes et que l'aaRS orthogonale ne doit pas prendre en charge ni des ARNt endogènes, ni des acides aminés non naturels

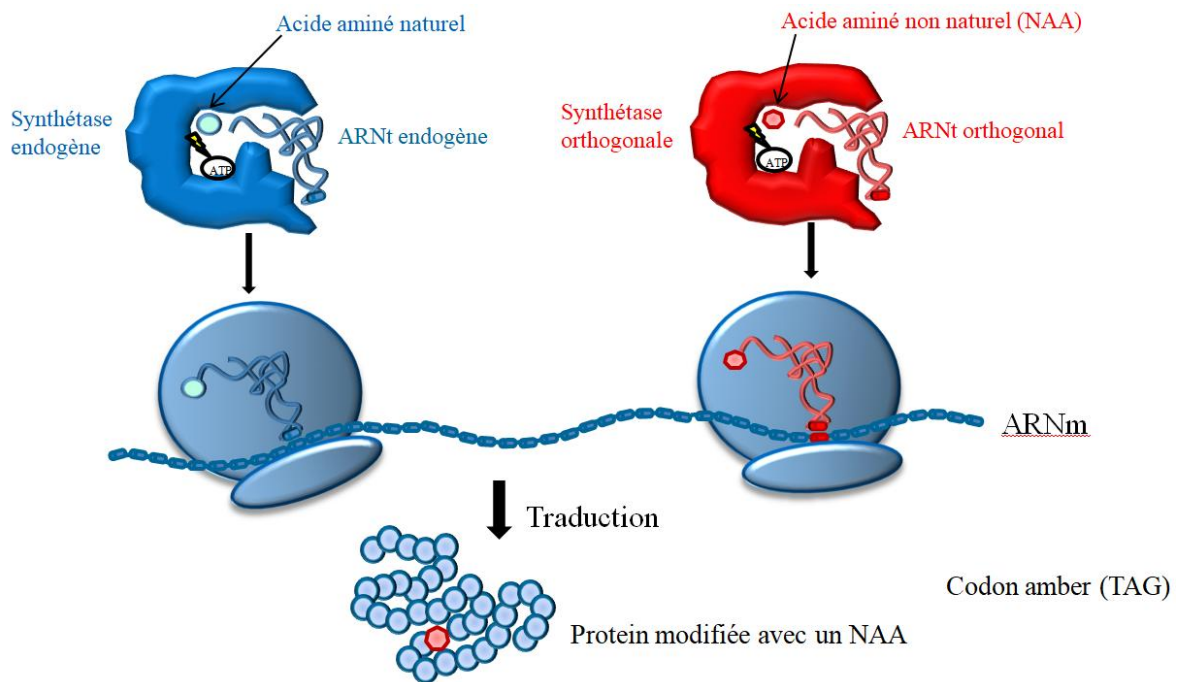


Figure 14: Incorporation d'un acide aminé non naturel (NAA) par la méthode site spécifique
 Lors de la traduction, l'aaRS orthogonale va charger le NAA sur l'ARNt orthogonal qui va alors être pris en charge par le ribosome lors de la traduction pour incorporer le NAA à la position du codon stop TAG. Figure reproduite à partir de la publication (Q. Wang, Parrish, and Wang 2009)

IV.3.2. Aminoacyl ARNt synthétases orthogonales

IV.3.2.1 Modification du code génétique porté par les aaRS in vivo

Comme nous l'avons vu précédemment (section IV.1.2.1), les aaRS reconnaissent principalement des éléments appelés déterminant au niveau de l'ARNt. En transposant ces éléments d'un ARNt à un autre, on peut alors modifier le code génétique. Ainsi, un ARNt^{Met} issu d'*E. coli* codant pour la méthionine peut être chargé par une thréonine par une modification de son anticodon de CAU à GGU (Schulman and Pelka 1990). Au niveau protéique, la modification de motifs de reconnaissance de déterminants au sein d'une glutamine ARNt synthétase a entraîné son utilisation de l'ARNt^{Glu} codant pour le glutamate en tant que substrat (Agou et al. 1998). Ainsi, la reconnaissance aaRS / ARNt est principalement portée par ces déterminants. Cependant tous les déterminants pour les couples aaRS / ARNt n'ont pas été identifiés. Afin de privilégier une absence de réaction croisée avec les aaRS / ARNt endogènes, un couple aaRS / ARNt phylogénétiquement éloigné de l'organisme de production est généralement choisi. Ainsi, une dizaine de couple aaRS / ARNt orthogonaux ont pu être transférés au sein d'organismes hétérologues (Tableau 5 ci-dessous) (Melnikov and Söll 2019).

Tableau 5: AaRS et ARNt utilisés pour la modification du code génétique *in vivo*

Tableau construit à partir des informations issues de la publication (Melnikov and Söll 2019). Les paires les plus utilisées sont indiquées en vert et seront développées section IV.3.2.2.

Couple Aminoacyl ARNt synthétase / ARNt	Organisme d'origine	Organisme de production	Différences de Déterminants critiques au niveau de l'ARNt
Histidine HisRS / ARNt^{His}	<i>C. crescentus</i>	<i>E. coli</i>	- G-1 et C73 pour <i>E. coli</i> - A73 pour <i>C. crescentus</i>
Leucine LeuRS / ARNt^{Leu}	<i>M. thermoautotrophicus</i> <i>H. sapiens</i>	<i>E. coli</i>	Différentes structures tertiaires
Lysine LysRS / ARNt^{Lys}	<i>P. horikoshi</i>	<i>E. coli</i>	-A73 pour les bactéries -G73 pour les archées
Phénylalanine PheRS / ARNt^{Phe}	<i>S. cerevisiae</i>	<i>E. coli</i>	-U20 pour les bactéries -G20 chez les archées
Proline ProRS / ARNt^{Pro}	<i>P. horikoshi</i>	<i>E. coli</i>	Modification de l'anticodon
Pyrrolysine PylRS / ARNt^{Pyl}	Différentes souches d'archées (<i>M. mazei</i> , <i>M. barkeri</i>)	Différentes bactéries et eucaryotes	Structure ARNt unique (boucle variable), interface ARNt aaRS unique
Tryptophane TrpRS / ARNt^{Trp}	<i>Saccharomyces cerevisiae</i>	<i>E. coli</i>	Séquence du bras accepteur différente
Tyrosine TyrRS / ARNt^{Tyr}	<i>Saccharomyces cerevisiae</i> <i>Methanocaldococcus jannaschii</i>	<i>E. coli</i>	-G1-C72 chez les bactéries -C1-G72 chez les eucaryotes et les archées
Acide aspartique AspRS / ARNt^{Asp}	<i>Saccharomyces cerevisiae</i>	<i>E. coli</i>	Séquence du bras accepteur différentes chez les bactéries et les eucaryotes
Serine Phosphate SepRS/ARNt^{Sep}	<i>M. maripaludis</i> <i>M. jannaschii</i>	<i>E. coli</i>	Interface unique entre l'aaRS et l'ARNt
Glutamine GlnRS / ARNt^{Gln}	<i>Saccharomyces cerevisiae</i>	<i>E. coli</i>	Modification du bras accepteur et de la D-stem

IV.3.2.2 Principales aaRS utilisées : la pyrrolysyl ARNt synthétase (*Methanosarcina mazei* et *Methanosarcina barkeri*) et la tyrosyl-ARNt synthétase (*Methanocaldococcus jannaschii*)

Nous allons maintenant détailler les structures des deux principales aminoacyl ARNt synthétases utilisées lors de l'incorporation d'acides aminés non naturels de manière site spécifique.

La pyrrolysyl ARNt synthétase et son ARNt (*M. mazei* et *M. barkeri*) :

La pyrrolysyl ARNt synthétase (PylRS) est une aaRS de type II qui se présente sous forme d'homodimères (Figure 15). Elle est composée de deux domaines : un domaine N-terminal (120 acides aminés environ) et un domaine C-terminal catalytique (270 acides aminés environ) (Yanagisawa et al. 2008a). Ce dernier, lorsqu'il est seul, possède une capacité catalytique légèrement amoindrie par rapport à la protéine complète (Yanagisawa et al. 2008a). Pendant longtemps le domaine N-terminal n'a pas pu être cristallisé, car il induisait une agrégation de la protéine. De plus, comme ce domaine ne présentait pas de similarité avec des domaines présents chez d'autres aaRS, aucune structure n'avait été prédite. Les domaines catalytiques ont donc été caractérisés de façon indépendante. Il existe une très forte conservation entre les domaines catalytiques de la PylRS issue de *Methanosarcina mazei* et de *Methanosarcina barkeri*. En effet, des mutations permettant de changer la spécificité de ces aaRS ont ainsi pu être transférées entre les PylRS de *M. mazei* et *M. barkeri* (Wan, Tharp, and Liu 2014). Néanmoins, le domaine N-terminal de la PylRS de *M. mazei* a pu

récemment être co-cristallisé avec l'ARNt^{Pyl}. (Suzuki et al. 2017). Cette partie N-terminale interagit avec l'ARNt et est de longueur différente selon l'organisme d'origine. Ainsi, elle est plus courte chez *M. barkeri* que chez *M. mazei*. Une protéine chimère composée du domaine N-terminal de *M. mazei* et du domaine catalytique de *M. barkeri* a été construite par Suzuki *et al.* et possède ainsi une solubilité accrue *in vitro* (Suzuki et al. 2017). De manière globale, cette enzyme possède une homologie de structure très importante avec la phénylalananyl ARNt synthétase avec qui elle partage pourtant moins de 30% d'identité. En effet, il est possible que ces protéines soient issues d'une divergence évolutive à partir d'une même protéine (Iwabe et al. 1989).

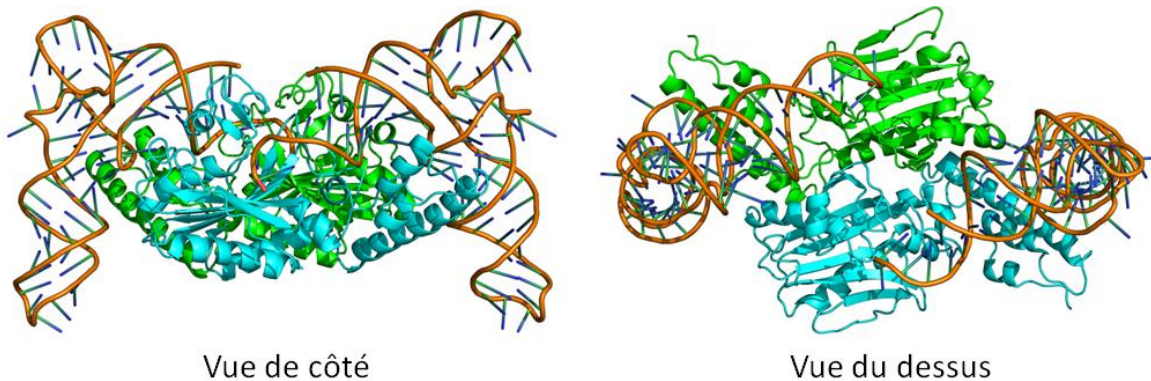


Figure 15: Vue d'ensemble de la pyrrolysyl ARNt synthétase (PylRS)

La PylRS forme des dimères qui reconnaissent deux ARNt. Nous pouvons également remarquer sur cette figure que les anticodons ne sont pas reconnus par la protéine. La structure est une Pyrrolysyl ARNt synthétase issue de *Desulfitobacterium halniense* car les structures des PylRS issues de *M.mazei* et *M.barkeri* n'ont pas été encore obtenues avec l'ARNt. Cependant, cette enzyme présente plus de 50% d'homologie avec les PylRS issues des archéobactéries. Image réalisée sur PyMol à partir de la structure PDB: 2ZNI.

Au niveau du site actif, la reconnaissance de la pyrrolysine s'effectue principalement par des interactions hydrophobes peu spécifiques et des liaisons hydrogènes impliquant la tyrosine Y384 et l'asparagine N346 (Figure 16). L'acide aminé Y384 est présent sur une boucle qui se repliera alors sur le site catalytique lors de la liaison de l'ATP et de la pyrrolysine tandis que l'acide aminé N346 est considéré comme charnière dans la reconnaissance des dérivés de lysines (Figure 16 et Figure 24). Cette reconnaissance est assez peu spécifique mais la pyrrolysine est un acide aminé d'une taille importante ce qui limite la compétition et donc la pression de sélection (Wan, Tharp, and Liu 2014). En effet, la protéine ne présente pas de reconnaissance très spécifique, ni de fonction d'édition permettant d'assurer la correcte aminoacylation de l'ARNt. Ainsi, la protéine sauvage a pu être utilisée pour incorporer de nombreux acides aminés non naturels de structure différentes dérivées de celle de la lysines (Dumas et al. 2015).

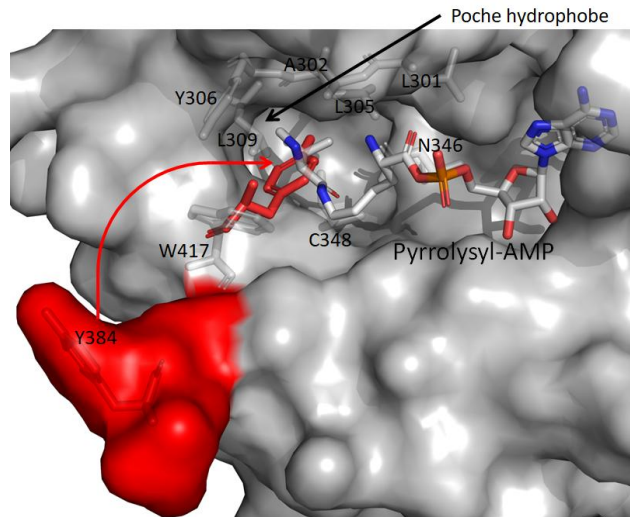


Figure 16: Site de reconnaissance de la pyrrolysyl ARNt synthétase (PylRS)

La reconnaissance de la pyrrolysine par la PylRS est principalement régie par des interactions hydrophobes peu spécifiques au niveau d'une poche. L'acide aminé Y384 (en rouge sur la figure) est présent sur une boucle qui se refermera en présence du substrat (voir flèche sur la figure et représentation en sticks de l'acide aminé en position fermée) et formera alors une liaison hydrogène avec la pyrrolysine. Les acides aminés formant la poche de reconnaissance sont : Y384, W417, L309, Y306, A302, L305, L301, N346, et C348. (image réalisée avec PyMol à partir de la superposition des structures ouverte (2E3C) et fermée (2Z1M) du site catalytique de la PylRS)

Au niveau de la reconnaissance de l'ARNt par la PylRS, celle-ci s'effectue à la fois par le biais du domaine C-terminal et du domaine N-terminal qui reconnaissent chacun une face de l'ARNt (Figure 17). Cette reconnaissance coordonnée est donc très spécifique malgré l'absence de reconnaissance de l'anticodon. Le domaine N-terminal reconnaît la forme concave très caractéristique de la boucle T de l'ARNt^{Pyl} ainsi que son bras variable très court, ce qui permet de garantir l'orthogonalité de cette reconnaissance (Suzuki et al. 2017). En effet, les autres ARNt sont ainsi exclus par gêne stérique. De plus, cette reconnaissance est complétée par une boucle au niveau du domaine C-terminal qui reconnaît le bras accepteur et le bras D de l'ARNt par des interactions grâce à de nombreux acides aminés basiques (Yanagisawa et al. 2008a).

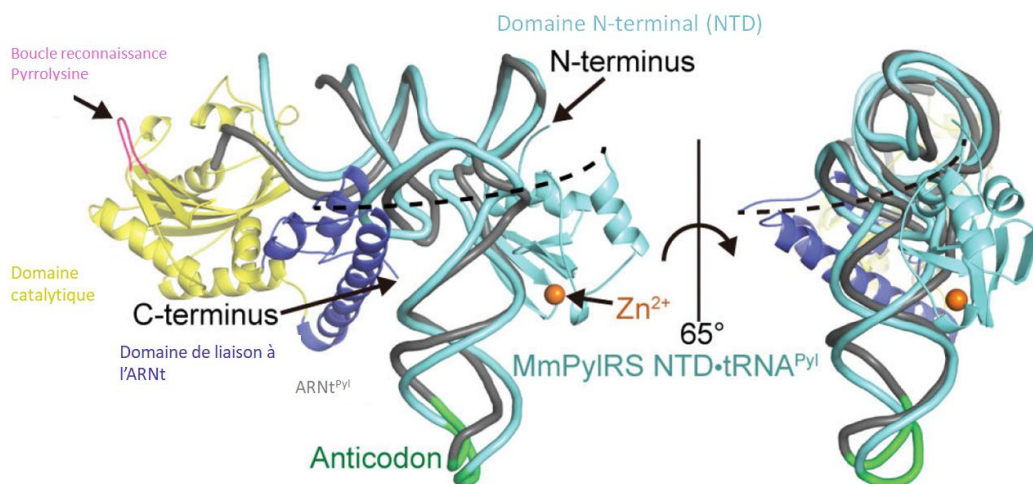


Figure 17: Reconnaissance de l'ARNt par la pyrrolysyl ARNt synthétase (PylRS)

L'ARNt est reconnu à la fois par le domaine C-terminal (au niveau du domaine en bleu foncé sur la figure), et par le domaine N-terminal (en bleu clair sur la figure). Le domaine C-terminal reconnaît le bras accepteur et le bras D de l'ARNt tandis que le domaine N-terminal reconnaît la boucle T et le bras variable de l'ARNt. Cette reconnaissance coopérative permet alors une grande spécificité de reconnaissance. Image extraite de la publication (Suzuki et al. 2017).

Ainsi, ce couple PylRS/ ARNt^{Pyl} a été largement utilisé pour l'incorporation d'acides aminés non naturels grâce à ces caractéristiques clés : i) une reconnaissance indépendante de l'anticodon ce qui permet de l'altérer, ii) une polyspécificité importante grâce à une reconnaissance de l'acide aminé assez peu spécifique, iii) une orthogonalité de ce couple à la fois au sein d'organismes procaryotes et eucaryotes, ainsi que par rapport au couple TyrRS / ARNt issu de *M. jannaschii* ce qui permet de réaliser des incorporations multiples (section IV.4.4) (Suzuki et al. 2017).

La tyrosyl ARNt synthétase et son ARNt (*M. jannaschii*) :

Le couple tyrosyl ARNt synthétase (TyrRS) / ARNt^{Tyr} issu de *M. jannaschii* fut le premier détourné pour incorporer un acide aminé non naturel (méthyltyrosine) de manière site spécifique (L. Wang et al. 2001). Cette TyrRS est une aminoacyl ARNt synthétase de classe I qui forme des homodimères (Figure 18). Sa structure est très semblable aux aaRS de type I. Elle est formée d'un domaine N-terminal d'environ 117 acides aminés qui reconnaît l'ARNt et d'un domaine C-terminal catalytique de type Rossmann d'environ 232 acides aminés (section IV.1.2.2) (Tsunoda et al. 2007). Ces deux domaines sont reliés par une boucle possédant la séquence KMSKS typique des aaRS de type I. Globalement sa structure est assez similaire à la TyrRS humaine et ce couple aaRS / ARNt n'est d'ailleurs pas orthogonal chez les eucaryotes (Y. A. N. Zhang et al. 2005). Cependant, le couple TyrRS / ARNt issu de *M. jannaschii* est orthogonal au sein des bactéries (L. Wang et al. 2001).

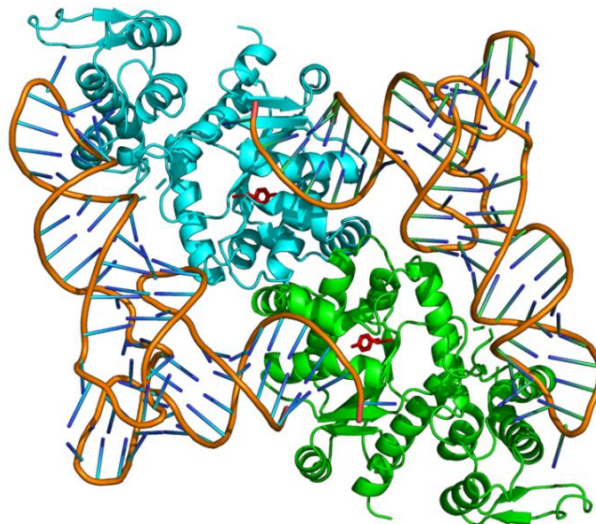


Figure 18: Structure de la Tyrosyl ARNt synthétase (TyrRS) issue de *M. jannaschii*

La TyrRS forme des dimères qui reconnaissent alors deux ARNt. Nous pouvons également remarquer que la TyrRS reconnaît les extrémités de l'ARNt au niveau de l'anticodon et du bras accepteur. (Image réalisée sur PyMol à partir de l'assemblage issu du code PDB 1J1U).

La reconnaissance de la tyrosine par la TyrRS de *M. jannaschii* (MjTyrRS) s'effectue au niveau d'une petite poche profonde chargée négativement grâce notamment aux acides aminés D158 situé au fond et E36, E172 situés à l'entrée de la poche (Figure 19) (Y. A. N. Zhang et al. 2005). Ainsi, les résidus Y32, Y151, Q155, D158 et Q173 sont conservés parmi les TyrRS issues de différentes espèces (Y. A. N. Zhang et al. 2005). Cette hydrophilicité limite la reconnaissance de la phénylalanine qui est très proche stériquement du substrat de l'enzyme (Y. A. N. Zhang et al. 2005). Par ailleurs, la phénylalananyl-ARNt synthétase possède une poche plutôt hydrophobe. Cette poche est également capable d'utiliser la tyrosine en tant que substrat et les PheRS possèdent donc une fonction d'édition afin d'assurer la correcte aminoacylation de l'ARNt (section IV.1.2.2) (Sasaki et al. 2006). La TyrRS ne possède pas de fonction d'édition ce qui facilite son détournement pour l'incorporation d'acides

aminés non naturels. La paire TyrRS / ARNt^{Tyr} de *M. Jannaschii* a donc été utilisée pour incorporer de nombreux dérivés de tyrosines (Dumas et al. 2015).

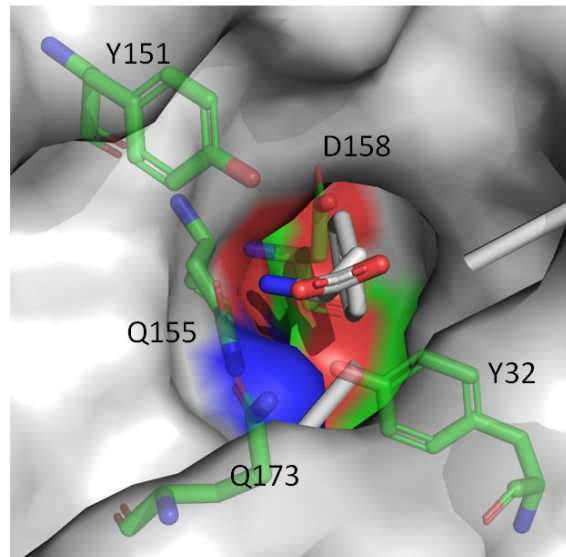


Figure 19 : Poche de reconnaissance de la tyrosine par la tyrosyl ARNt synthétase de *M. jannaschii*
La reconnaissance de la tyrosine s'effectue au niveau d'une poche étroite et très hydrophile grâce à la présence des acides aminés D158, Y151, Q155, Q173 et Y32. (Image réalisée sur PyMol à partir de la structure PDB:1J1U).

Contrairement à la PylRS, la TyrRS reconnaît l'anticodon de l'ARNt (Figure 18 et Figure 20). Le changement de celui-ci par le codon TAG a d'ailleurs entraîné une baisse de l'efficacité catalytique d'un facteur 300. Cependant, cette baisse de reconnaissance a pu être restaurée à l'aide d'une mutation clé (section IV.3.3.2) (Kobayashi et al. 2003). Les trois éléments déterminants principaux reconnus au niveau de l'ARNt sont : i) l'anticodon, ii) la paire de base C1-G72 et iii) la base discriminante A73 (Tsunoda et al. 2007). La reconnaissance de l'ARNt par la TyrRS est variable en fonction de l'organisme d'origine ce qui permet d'expliquer son orthogonalité chez les bactéries (Figure 20). Tout d'abord, l'anticodon est reconnu par le domaine C-terminal des TyrRS qui reconnaît également le bras variable qui est très long dans le cas des TyrRS bactériennes. Par ailleurs, malgré leur appartenance aux aaRS de classe I, le domaine C-terminal des TyrRS issues des bactéries et des archées reconnaissent l'ARNt par le grand sillon contrairement aux TyrRS d'origines eucaryotes (Tsunoda et al. 2007). Quant à la paire C1-G72, sa reconnaissance est particulièrement conservée entre les TyrRS issues des archées et des eucaryotes alors qu'il s'agit d'une paire G1-C72 chez les TyrRS bactériennes (Tsunoda et al. 2007). Par ailleurs, la base discriminante A73 est reconnue à l'aide d'une arginine très conservée entre les eucaryotes et les archées (Tsunoda et al. 2007). Ainsi, des différences structurales importantes existent entre les TyrRS issues des bactéries et des archées ce qui pourrait expliquer cette orthogonalité spécifique.

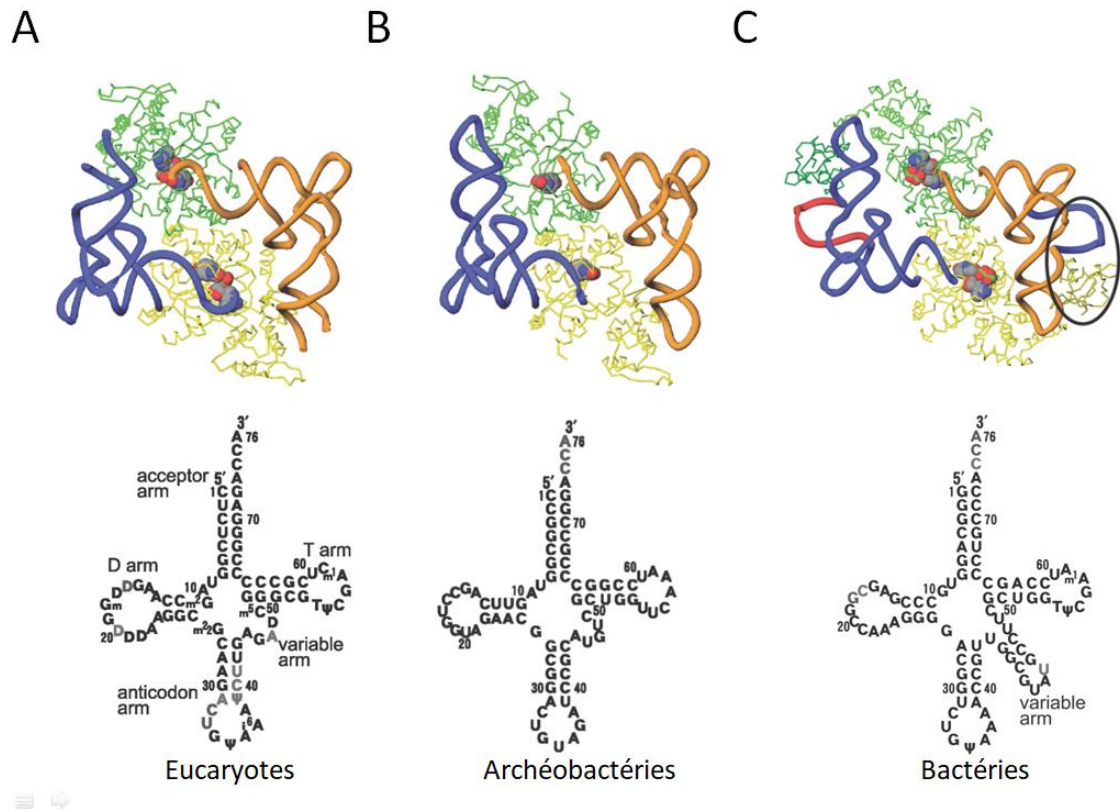


Figure 20: Modes de reconnaissance de l'ARNt selon l'organisme d'origine de la Tyrosyl ARNt synthétase. La reconnaissance de l'ARNt varie selon l'organisme d'origine. Tout d'abord, l'anticodon est reconnu par le domaine C-terminal de la TyrRS qui reconnaît également le bras variable qui est très long dans le cas de la TyrRS bactérienne. Par ailleurs, malgré leur appartenance aux aaRS de classe I, le domaine C-terminal des TyrRS issues des bactéries et des archées reconnaissent l'ARNt par le grand sillon contrairement à la TyrRS d'origine eucaryote. Quant à la paire C1-G72, sa reconnaissance est particulièrement conservée entre les TyrRS issues des archées et des eucaryotes alors qu'il s'agit d'une paire G1-C72 chez les TyrRS bactériennes. Par ailleurs, la base discriminante A73 est reconnue à l'aide d'une arginine très conservée entre les eucaryotes et les archées. (Image extraite de la publication (Tsunoda et al. 2007)).

IV.3.3. Mutagénèse des aminoacyl ARNt synthétases

IV.3.3.1 Méthodes de sélection

Les aaRS possèdent une polyspécificité qui permet l'incorporation de différents acides aminés non naturels par ces enzymes sous leur forme sauvage (Dumas et al. 2015). Toutefois, afin d'élargir leurs capacités de reconnaissance, différents mutants ont été générés. Il est complexe de prédire précisément des mutations permettant la reconnaissance d'un nouvel acide aminé non naturel par une aaRS. Ainsi, des méthodes de mutagénèse couplées à des méthodes de sélection sont généralement utilisées (Chang C. Liu and Schultz 2010). Les différentes mutations sont le plus souvent générées par une méthode de mutagénèse dirigée saturante par PCR incluant un codon dégénéré NNK, correspondant à 32 codons différents, qui permet de tester tous les acides aminés naturels à une position donnée (Zheng, Baumann, and Reymond 2004). La librairie doit ensuite être transformée au sein de *E. coli* ou de *S. cerevisiae* et la sélection est réalisée ensuite selon des cycles de sélections positives et négatives permettant respectivement de sélectionner les mutants actifs et de contre-sélectionner les mutants non spécifiques (Figure 21).

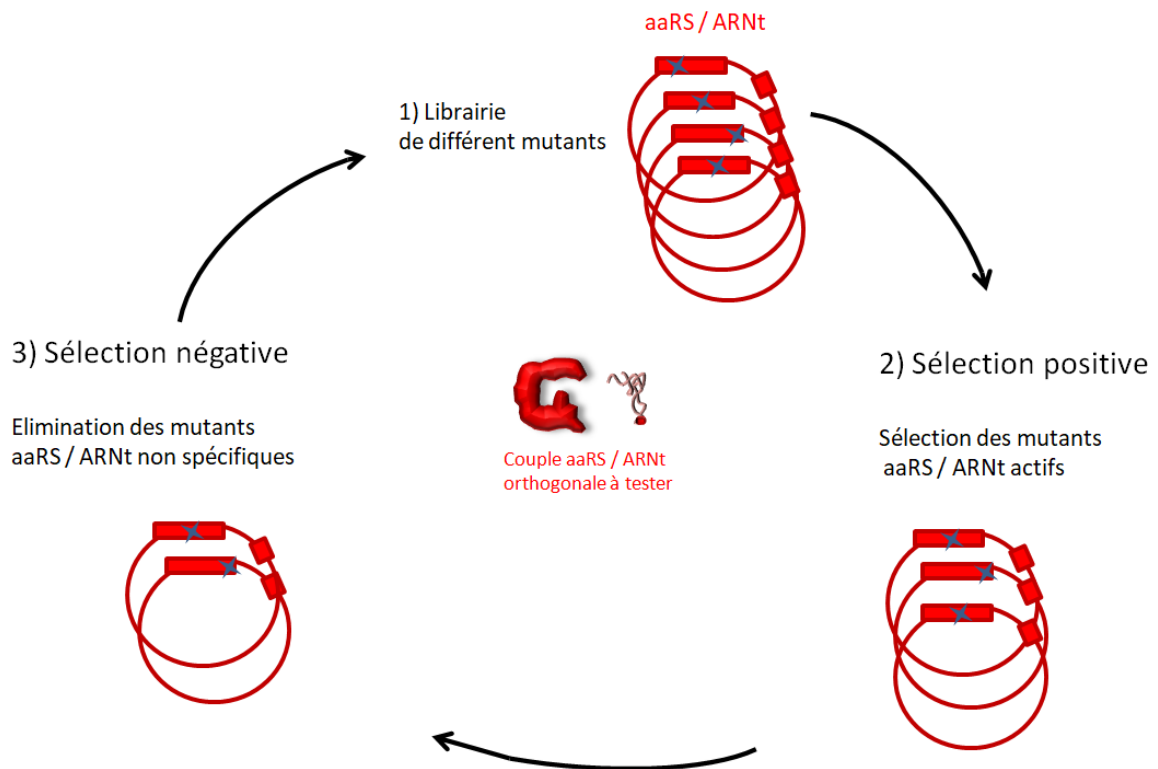


Figure 21: Cycle de sélection permettant de changer la spécificité d'une aminoacyl ARNt synthétase (aaRS)
 L'élargissement des capacités de reconnaissance des aaRS s'effectue généralement par la réalisation de mutagenèse dirigée suivi de cycles de sélections positives et négatives. La sélection positive permet alors de sélectionner les mutants permettant la suppression du codon TAG et donc supposés actifs. Tandis que la sélection négative élimine les mutants non spécifiques en s'assurant que la suppression du codon TAG est bien spécifique de l'incorporation d'un acide aminé non naturel.

La méthode de sélection la plus répandue consiste (Figure 22)(Chang C. Liu and Schultz 2010) :

-pour la sélection positive : à utiliser l'expression d'un gène essentiel ou de résistance à un antibiotique (le plus souvent chloramphénicol, ou ampicilline) au sein duquel un codon stop TAG a été inséré au niveau d'un site permissif. L'expression est alors réalisée en présence de l'acide aminé non naturel. Si la bactérie survit en présence de l'acide aminé non naturel, cela signifie que le couple aaRS / ARNt est capable d'exprimer le gène de résistance et donc probablement d'incorporer le NAA. La pression de cette sélection pourra alors être plus ou moins forte selon la concentration en antibiotique utilisée.

-pour la sélection négative : à utiliser l'expression d'un gène toxique au sein duquel un codon stop TAG a été inséré au niveau d'un site permissif. L'expression est alors réalisée en l'absence de l'acide aminé non naturel. Si la bactérie meurt, cela signifie que le couple aaRS / ARNt n'est pas complètement orthogonal et qu'il permet l'incorporation d'un acide aminé naturel en réponse au codon TAG.

Entre chaque cycle, les plasmides correspondant aux différents mutants de l'aaRS couplée à l'ARNt doivent être retransformés avec le plasmide de sélection positive ou négative (Figure 22).

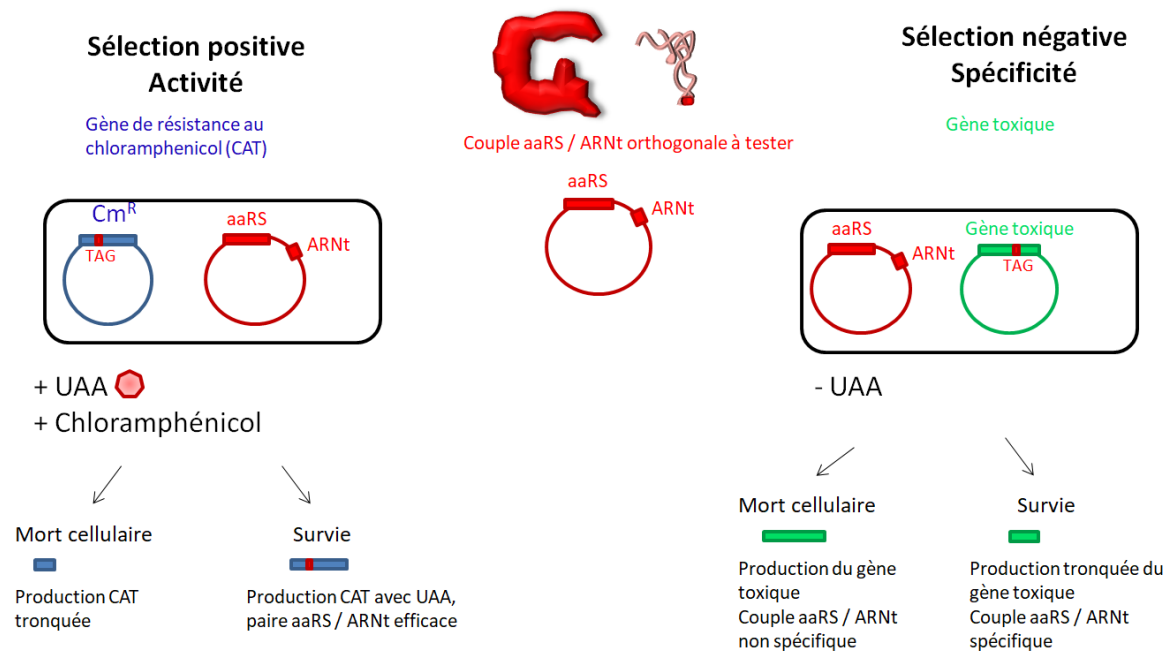


Figure 22: Sélection positive et négative d'un couple aminoacyl ARnt synthétase et ARnt

La méthode de sélection ci-dessus consiste:

-pour la sélection positive : à utiliser l'expression d'un gène de résistance au chloramphénicol au sein duquel un codon stop TAG a été inséré au niveau d'un site permissif. L'expression est alors réalisée en présence de l'acide aminé non naturel et de chloramphénicol. Si la bactérie survit, cela signifie que le couple aaRS / ARnt est capable d'exprimer le gène de résistance et donc probablement d'incorporer l'acide aminé non naturel.

-pour la sélection négative : à utiliser l'expression d'un gène toxique au sein duquel un codon stop TAG a été inséré au niveau d'un site permissif. L'expression est alors réalisée en l'absence de l'acide aminé non naturel. Si la bactérie meurt, cela signifie que le couple aaRS / ARnt n'est pas complètement orthogonal et qu'il permet l'incorporation d'un acide aminé naturel en réponse au codon TAG.

Ainsi, alors que la sélection positive permet d'enrichir en mutants efficaces pour la suppression d'un codon TAG, la sélection négative s'assure que cette suppression est bien spécifique de l'incorporation d'un acide aminé non naturel. En effet, les différentes mutations au niveau du site actif peuvent affaiblir l'orthogonalité de l'aminocyl ARnt synthétase, il est donc important de s'assurer que celle-ci ne prenne pas en charge d'acide aminé naturel.

Afin de limiter les étapes de transformation et réaliser la sélection positive et négative à partir du même transformant, un autre plasmide de sélection a été développé par Schultz et ses collaborateurs. La sélection est réalisée grâce à un gène de fusion entre le gène de résistance au chloramphénicol (chloramphénicol acétyltransférase, CAT) et l'uracile phosphoribosyltransférase (UPRT), au sein duquel deux codons stop TAG ont été ajoutés. La sélection positive est ensuite réalisée grâce à l'utilisation du chloramphénicol alors que la sélection négative est effectuée en présence du 5-fluoro-uracile qui est convertie en 5-fluoro-dUMP par l'UPRT, ce qui inhibe alors la voie de synthèse de la thymidine (base nucléotidique) causant ainsi la mort de la cellule (Melançon and Schultz 2009). On peut cependant noter que l'extraction plasmidique entre chaque sélection permet d'éliminer la potentielle dérive génétique au niveau du génome de la souche bactérienne. Néanmoins, l'utilisation de deux codons stops TAG permet de rendre la sélection positive plus stringente.

L'utilisation de l'expression d'une protéine fluorescente comme la GFP peut également être utilisée. Ainsi, un codon stop TAG est inséré à un site permissif au sein du gène de la GFP. Lorsque les bactéries sont mises en culture avec l'acide aminé non naturel, les bactéries fluorescentes contiennent un couple aaRS / ARnt efficace pour la suppression du codon TAG. En l'absence de

l'acide aminé non naturel, les bactéries fluorescentes correspondent alors à un couple aaRS / ARNt qui réalise cette suppression de manière non spécifique, avec un acide aminé endogène. Cette méthode couplée à une analyse par FACS et tri cellulaire peut alors permettre un débit de sélection important et très sensible (A. E. Lin and Lin 2019).

Récemment, l'utilisation de phages a permis de développer des nouvelles méthodes de sélections pour les aaRS (Suzuki et al. 2017; Bryson et al. 2017). Pour cela, l'adaptation de la méthode de sélection PACE (*Phage Assisted Continuous Evolution*) a été effectuée par Liu et ses collaborateurs afin de réaliser la sélection de façon continue et très rapide d'aminoacyl ARNt synthétases (Figure 23)(Bryson et al. 2017). La sélection PACE consiste en l'utilisation du bactériophage M13, de sa dépendance à l'expression de la protéine pIII, et de la génération de mutations aléatoires directement au sein du phage. Les phages sont des virus infectant les bactéries afin de reproduire leur matériel génétique. Le phage M13 exprime à sa surface la protéine pIII qui lui permet de se lier au pilus de la bactérie qu'il va infecter. Son cycle de reproduction est d'environ 25 minutes. Des techniques classiques de phage display utilisent l'expression d'une protéine de fusion avec la protéine pIII à la surface du phage afin de sélectionner différents mutants de la protéine sur ses caractéristiques de liaison à une cible. Lors de ces méthodes, la librairie des différents mutants est générée par des méthodes de biologie moléculaire (Kushwaha, Schäfermeyer, and Downie 2014). La méthode PACE a été mise au point à partir de ce principe par Liu et ses collaborateurs en 2011 (Esvelt, Carlson, and Liu 2011). Cette méthode consiste à lier l'activité ciblée au cycle de reproduction et donc d'infection du phage. Pour cela, le gène pIII est supprimé du plasmide qui contient l'ADN du phage et transféré sur un autre plasmide. Le gène pIII sera alors produit si le gène muté effectue la fonction recherchée et les phages contenant les mutants efficaces vont alors se multiplier tandis que les phages contenant les mutants inefficaces ne vont pas se multiplier ou beaucoup moins. De plus, afin de ne pas générer une librairie de mutants aléatoires manuellement entre chaque cycle, un autre plasmide inductible à l'arabinose est également utilisé. Ce dernier contient un ADN polymérase sans fonction de relecture (dnaQ926) et un domaine de réparation d'ADN polymérase qui favorise les erreurs (pol V), ce qui permet d'augmenter le taux d'erreur d'un facteur 100 lors de la réplication. Cette méthode est donc facile et très rapide car elle permet ainsi de générer des cycles de sélection au fur et à mesure des cycles d'infection du phage sans intervention humaine (Esvelt, Carlson, and Liu 2011). Afin d'évoluer des aminoacyl ARNt synthétase pour l'incorporation d'un acide aminé non naturel, deux stratégies ont été développées : i) un codon stop TAG a été généré au niveau du gène pIII ou ii) le gène pIII a été mis sous le contrôle d'un promoteur inductible par une T7 ARN polymérase, et un codon stop TAG a été généré au niveau du gène de cette polymérase. Le gène pIII n'est donc exprimé que lorsque la suppression du codon TAG est efficace (Bryson et al. 2017). La sélection négative a été développée selon le même type de méthode mais par l'expression d'un gène pIII-neg inductible qui baisse l'efficacité de l'infection (Bryson et al. 2017). Cette méthode a alors permis l'évolution d'une PylRS 45 fois plus efficace pour l'incorporation de la Boc-Lysine et l'évolution d'une TyrRS pour l'incorporation de la p-iodo-phénylalanine en 48h (Bryson et al. 2017).

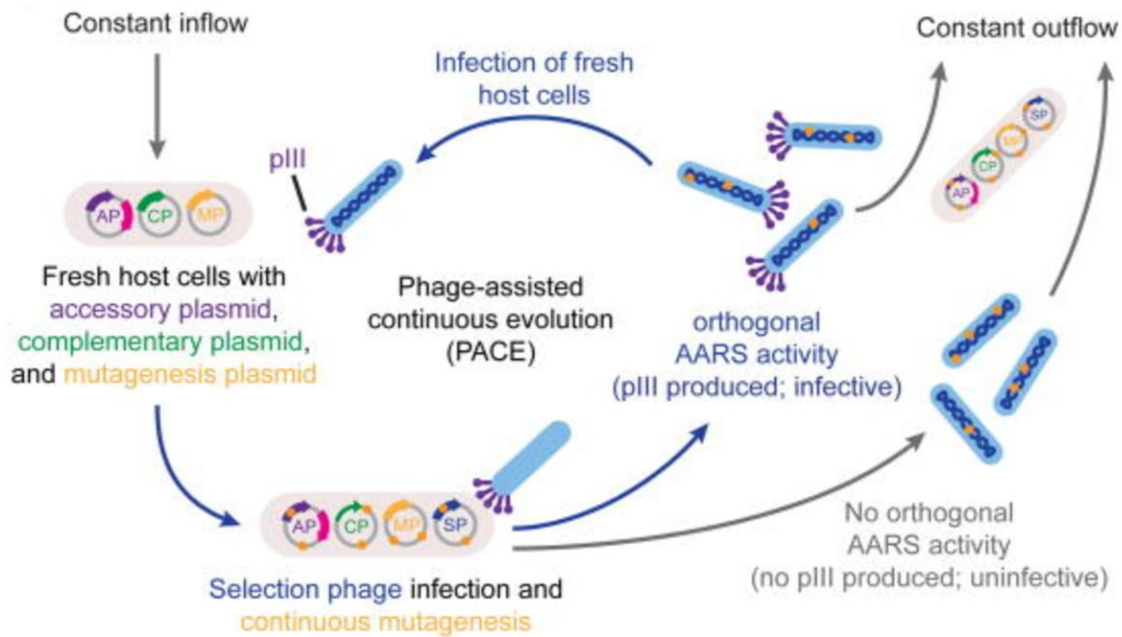


Figure 23: Méthode de sélection PACE appliquée à l'évolution d'aminocyl ARNt synthétase

Le plasmide accessoire (AP) contient le gène pIII et l'ARNt avec l'anticodon ambre, Le plasmide complémentaire (CP) contient alors l'ARN polymérase T7 et le plasmide de mutagénèse (MP) favorise les erreurs de réplication et donc le taux de mutations générées. Quant au phage (SP), il exprime tout ses gènes sauf le gène pIII qui est remplacé par le gène codant pour l'aminocyl ARNt synthétase à évoluer. Le phage qui possède une copie efficace pour l'incorporation de l'acide aminé non naturel sera alors capable de se répliquer. Ayant un avantage sélectif, au cours des différents cycles d'infections, les phages comprenant une aaRS efficace seront alors plus nombreux. De plus, le cycle d'infection d'un phage étant très court (environ 20-30min), plusieurs cycles de sélection peuvent être générés très rapidement. (Image extraite de la publication(Bryson et al. 2017)).

Bien que généralement effectuée chez *E. coli*, les étapes d'évolution dirigée peuvent parfois être réalisées chez la levure. En effet, afin d'incorporer des acides aminés non naturels chez les eucaryotes des paires aaRS / ARNt issues d'*E. coli* sont parfois détournées. Cette paire ne sera alors pas orthogonale à *E. coli* et les étapes de sélection ne peuvent donc pas se dérouler chez la bactérie. Des méthodes de sélection ont alors été mises au point au sein de la levure. De façon analogue à la sélection classique chez *E. coli*, la sélection chez la levure est réalisée par une sélection positive et négative qui pourra être effectuée à l'aide d'un unique plasmide. Le codon stop est en effet inséré au sein d'un activateur de la transcription (GAL4) qui contrôle l'expression d'un gène essentiel et d'un gène générant un produit toxique. Le gène essentiel consiste alors en la restauration d'une auxotrophie (URA3, LEU2, HIS3, LYS2). L'utilisation du gène URA3 peut alors permettre de réaliser également la sélection négative par la mise en culture de la souche en présence d'acide 5-fluoro-orotique (5'FOA) qui sera alors converti par Ura3 en un produit toxique (Chin et al. 2003).

IV.3.3.2 Mutations généralement observées

Afin d'élargir les capacités de reconnaissance des aaRS, les mutations sont générées au niveau de la poche de reconnaissance de l'acide aminé à greffer sur l'ARNt. Ces mutations sont spécifiques de l'acide aminé à incorporer. Cependant, à partir de la revue bibliographique réalisée par Dumas et ses collaborateurs et la publication de Wan *et al.*, nous avons comptabilisé la fréquence de chaque mutation et ainsi pu déterminer les mutations les plus fréquemment retrouvées que nous avons ensuite représentées sur les Figure 24 et Figure 25 (Dumas et al. 2015; Wan, Sharp, and Liu 2014).

Ainsi pour la Pyrrolysyl ARNt synthétase, nous pouvons voir sur la Figure 24 les points de mutations les plus fréquemment retrouvés (Y384, N346, C348, Y306 et L309), et ceux assez fréquemment observés (A302, L305, W417). Ces numérotations font référence à la Pyrrolysyl ARNt synthétase de *M. mazei* mais nous pouvons noter que ce sont des points de mutations homologues à *M. barkeri* puisque leurs sites actifs sont superposables. Au niveau de la reconnaissance de la pyrrolysine, deux liaisons hydrogènes sont formées avec l'acide aminé N346 et l'acide aminé Y384 (Wan, Tharp, and Liu 2014). L'acide aminé Y384 est présent au niveau d'une boucle qui referme le site actif en présence de l'ATP et de l'acide aminé substrat (Yanagisawa et al. 2008a). Y384 est fréquemment mutée en phénylalanine, ce qui permettrait d'augmenter la vitesse d'aminocyclation indépendamment de l'acide aminé substrat (Yanagisawa et al. 2008b). Quant à l'acide aminé N346 de l'aaRS, il est considéré comme un "gatekeeper" limitant la reconnaissance aux dérivés de lysines et est essentiel de par son interaction avec l'amine en alpha de cette famille d'acides aminés (Wan, Tharp, and Liu 2014). La mutation N346A diminuerait la capacité catalytique de l'enzyme par la perte de la liaison hydrogène qui pourrait être compensée par la mutation A302T (Takimoto et al. 2011). Cependant, la mutation conjointe de N346A et C348A a été décrite comme permettant d'élargir la poche de reconnaissance, favorisant notamment l'incorporation de nombreux dérivés de phénylalanine (Tharp et al. 2014). La mutation de la position N346 est d'ailleurs fréquemment retrouvée (Wan, Tharp, and Liu 2014). Quant aux positions Y306 et L309, elle sont situées au fond de la poche hydrophobe de reconnaissance, leurs mutations peuvent donc permettre de l'agrandir (Wan, Tharp, and Liu 2014).

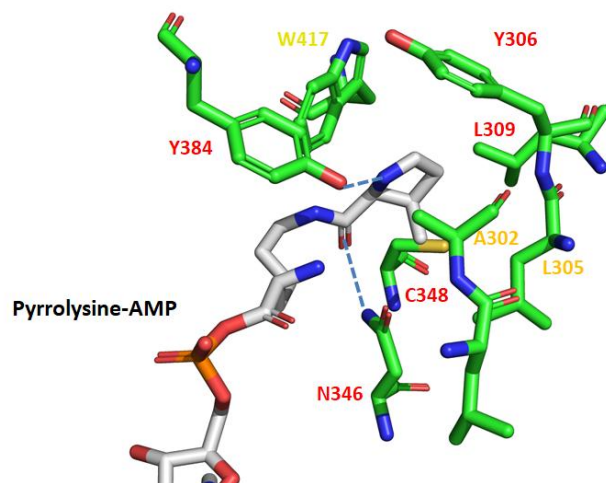


Figure 24: Site actif de la pyrrolysine ARNt synthétase (PylRS)

La reconnaissance de la pyrrolysine par la PylRS est principalement réalisée au niveau des acides aminés suivants: Y384F, W417, Y306, L309, A302, L305, C348 et N346. D'après une analyse que nous avons réalisé à partir de la revue bibliographique réalisée par (Dumas et al. 2015), nous pouvons noter que ces acides aminés sont plus ou moins fréquemment l'objet de mutations pour changer la spécificité de la PylRS. Ainsi, les acides aminés indiqués en rouges sont très fréquemment mutés et les acides aminés indiqués en orange sont fréquemment mutés. (Image PyMol réalisée à partir de la publication de Dumas et al. et de la structure PDB 2ZIM).

En ce qui concerne la Tyrosyl ARNt synthétase de *M. jannaschii*, nous pouvons voir sur la Figure 25, les points de mutations les plus fréquemment retrouvés (très fréquemment (Y32 et D158), et assez fréquemment (L162 et I159, L65, H70, E107, F108 et Q109)). Les acides aminés Y32 et D158 forment des liaisons hydrogènes avec la tyrosine. Ces points de mutations peuvent alors permettre d'augmenter la taille de la cavité de liaison et favoriser un changement de spécificité. Ainsi, par exemple, pour l'incorporation de la *p*-acétylphénylalanine, les mutations Y32L et D158G se sont montrées efficaces. La mutation Y32L supprime une liaison hydrogène entre les groupements

hydroxyles ce qui diminue la reconnaissance du substrat naturel et favorise une interaction de type Van der Waals avec un acide aminé hydrophobe (ici le groupement acétyle), tandis que la mutation D158G agrandit la poche de reconnaissance (R. Chen, Greer, and Dean 1996). Les acides aminés G34, L65, Q155 et H70 ont des contacts hydrophobes avec la tyrosine (substrat naturel) et peuvent également faire l'objet de mutations (Turner et al. 2006). Des acides aminés beaucoup plus volumineux que la tyrosine ont ainsi pu être incorporés, tels qu'un dérivé fluor photoprotégé de la tyrosine ou l'acide aminé p-benzoyl-phénylalanine permettant de réaliser de la photoconjugaison (Dumas et al. 2015). Les points de mutation L65, H70, F108, I159 et L162 permettent d'élargir la poche de reconnaissance et de changer la spécificité de l'aaRS (Dumas et al. 2015). Un mutant de la Tyrosyl ARNt synthétase (Y32L, L65V, F108W, N109M, D158G, I159A) a démontré une polyspécificité importante pour l'incorporation de différents dérivés de tyrosine en para (D. D. Young et al. 2011). Ces mutations ont permis de générer une poche plus large et malgré la perte des liaisons hydrogènes formées par Y32 et D158, l'acide aminé substrat est stabilisé par des liaisons de type Van der Waals (D. D. Young et al. 2011).

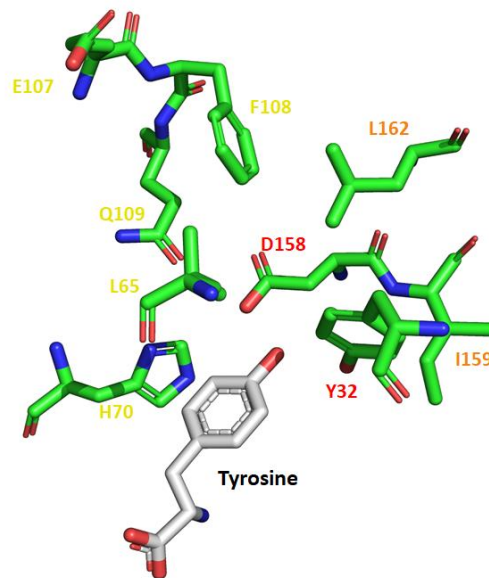


Figure 25: Site actif de la Tyrosyl ARNt synthétase de *M. jannaschii* (MjTyrRS).

La tyrosine est principalement reconnue au niveau d'une poche formée par les acides aminés Y32, D158, L162, F108, E107, Q109, L65, H70. Grâce à une analyse que nous avons réalisée à partir de la revue bibliographique réalisée par (Dumas et al. 2015), nous avons pu déterminer les acides aminés les plus fréquemment mutés, indiqués ici en rouge, et les acides aminés fréquemment mutés, indiqués ici en orange. Ces différentes mutations permettent des changements de spécificité au niveau de la Tyrosyl ARNt synthétase.

La découverte des mutations permettant l'incorporation d'un acide aminé non naturel est généralement réalisée à partir de méthodes aléatoires ou semi-rationnelles. L'effet des mutations n'est que rarement étayé par des structures cristallographiques. Cependant, les mutations observées sont généralement des mutations qui permettent de nouvelles interactions afin de favoriser l'acide aminé non naturel par rapport au substrat naturel ou de répondre à une gêne stérique. Toutefois, la sélection de mutations non sens a pu être observée lors de l'utilisation de la méthode de sélection PACE décrite précédemment (section IV.3.3.1). Ainsi, un codon stop a été généré entraînant la séparation des deux domaines de la pyrrolysyl ARNt synthétase. La traduction s'arrête générant un premier domaine d'environ 93-102 acides aminés selon le site de troncation sélectionné et un second domaine commençant par une méthionine à la position 107. Cette PylRS formée de deux unités indépendantes a une activité catalytique similaire à la protéine initiale uniquement en présence des

deux domaines. Sa sélection est cependant surprenante et son avantage évolutif n'a pas été encore identifié (Bryson et al. 2017). Néanmoins, nous pouvons noter qu'une pyrrolysyl ARNt synthétase (PylRS) bactérienne issue de *Desulfitobacterium hafniense* avec une homologie de séquence de 39% avec la PylRS issue de l'archée *M. mazei* est naturellement encodée par deux gènes distincts (Nozawa et al. 2009).

Par ailleurs, les méthodes de sélection décrites précédemment (section IV.3.3.1) peuvent également permettre de renforcer les interactions entre l'aminocyl ARNt synthétase et l'ARNt. Ainsi, la mutagenèse n'est pas générée au niveau du site actif mais au niveau de l'interface de l'interaction entre l'aaRS et l'ARNt. La sélection sera alors plus stricte afin de sélectionner les mutants les plus efficaces. Selon le modèle de sélection, cela peut se traduire par l'utilisation d'une plus forte quantité d'antibiotique, la suppression de deux codons stop TAG, ou la sélection des bactéries les plus fluorescentes (Vangmayee Sharma et al. 2018; Vargas-Rodriguez et al. 2018; Lei Wang, Xie, and Schultz 2006). Ainsi, les mutations F261G/L, H283L, M285V/G et plus particulièrement la mutation D286G/R ont ainsi été identifiées comme augmentant la reconnaissance entre la Tyrolysyl ARNt synthétase de *M. jannaschii* et son ARNt^{Tyr} (Nanxi Wang et al. 2015; Gan et al. 2017). En effet, la mutation D286R a été répertoriée comme baissant l'affinité de la TyrRS de *M. jannaschii* pour l'anticodon GUA au profit de l'anticodon CUA de l'ARNt^{Tyr} suppresseur du codon TAG (Figure 26) (Kobayashi et al. 2003). Au niveau de la Pyrrolysyl ARNt synthétase, son domaine N-terminal a été évolué et les mutations R19H, H29R et T122S ont été identifiées comme augmentant d'un facteur 3 à 4 l'incorporation d'acide aminé non naturel. Bien que n'ayant pas été démontré structurellement, l'apport de ces mutations est supposé améliorer la reconnaissance de l'ARNt^{Pyl} par la PylRS (Vangmayee Sharma et al. 2018).

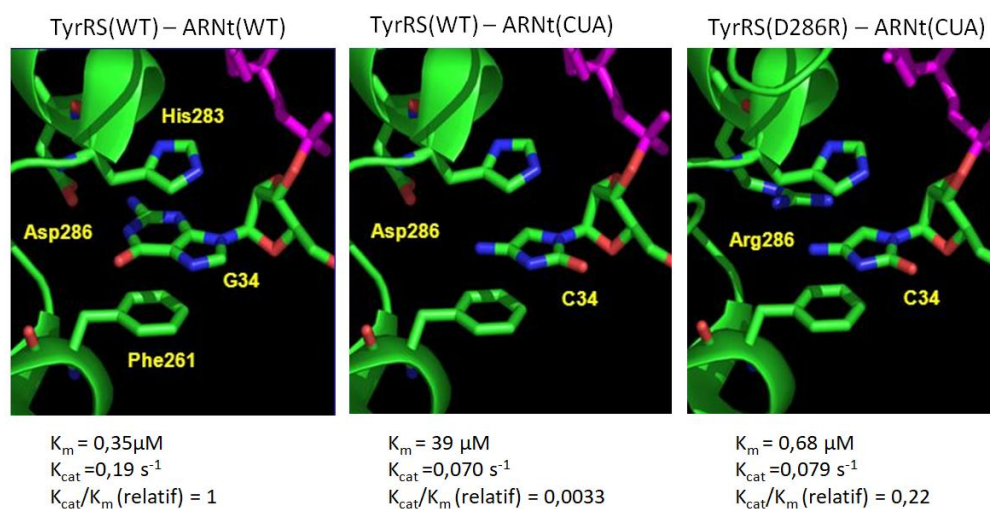


Figure 26: Interaction entre la TyrRS de *M. jannaschii* et son ARNt : effet de la mutation D286R. La mutation D286R entraîne une meilleure reconnaissance de l'anticodon CUA qui permet la suppression du codon TAG. En effet, l'arginine 286 interagit alors avec la troisième base C. Image obtenue à partir de (Ryu 2015).

De la même manière, l'ARNt peut également être évolué afin d'augmenter cette interaction ou sa prise en charge par la machinerie endogène de la traduction (section IV.4.2.1). Au niveau de l'ARNt^{Pyl} de la pyrrolysyl ARNt synthétase, lors de travaux cherchant à augmenter son affinité pour le facteur d'élongation, Schultz et ses collaborateurs ont identifié les mutations U49, A65/G50, C64 (au lieu de C49, G65/C50, G64) comme permettant d'augmenter l'efficacité d'incorporation d'un facteur 3 lors de l'incorporation d'acétyl-lysine. Des expériences de cinétique n'ont pas pu lier leur effet à

une meilleure interaction avec l'EF-Tu et il a alors été supposé qu'elle permettait une meilleure interaction avec la PylRS (Fan et al. 2015).

IV.4 Optimisation et nouvelles stratégies

Nous allons voir les principaux paramètres qui pourraient influencer sur le rendement de production d'une protéine contenant un acide aminé non naturel *in vivo* chez *E. coli*, les différentes méthodes développées afin d'optimiser celui-ci ainsi que les stratégies émergentes permettant l'incorporation d'acides aminés non naturels.

In vivo, le procédé de production d'une protéine modifiée avec un acide aminé non naturel sera influencé par trois types de facteurs (Figure 27) : i) disponibilité de l'acide aminé non naturel (transport et concentration intracellulaire), ii) conditions liées à la machinerie de la traduction, iii) conditions d'expression du couple aaRS / ARNt orthogonal ainsi que de la protéine cible.

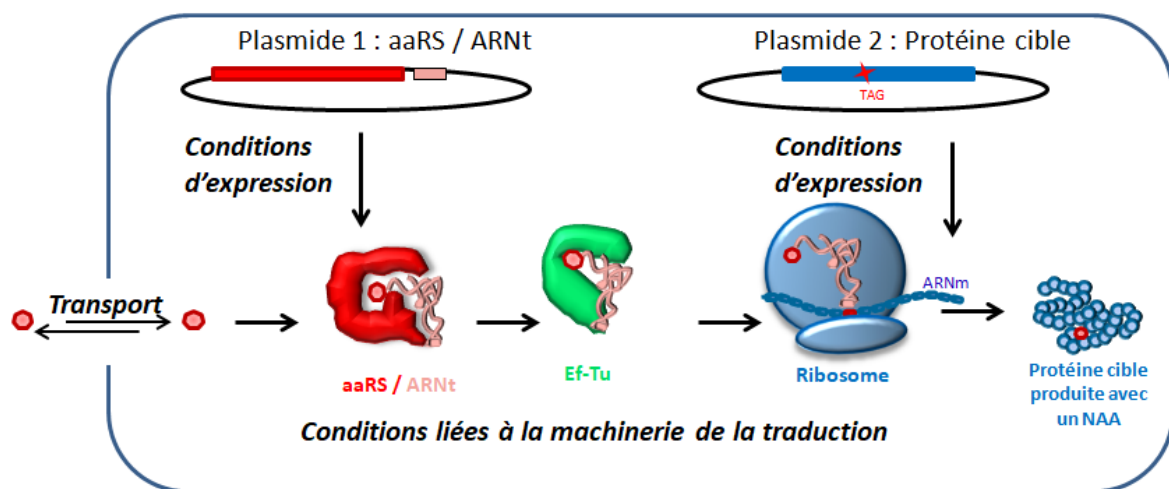


Figure 27: Facteurs influant la production d'une protéine contenant un acide aminé non naturel
L'efficacité d'incorporation d'un acide aminé non naturel pourra être affectée par : i) la disponibilité de l'acide aminé non naturel (transport et concentration intracellulaire), ii) les conditions d'expression du couple aaRS / ARNt orthogonal, iii) l'intégration de cette nouvelle voie à la machinerie endogène de la traduction, iv) les conditions d'expression de la protéine cible et conditions de croissance bactérienne.

IV.4.1. Entrée de l'acide aminé non naturel : transport et concentration

Afin d'être incorporé au sein de protéine, l'acide aminé non naturel doit pouvoir être biodisponible et donc entrer au sein de la cellule. Les acides aminés non naturels sont ainsi simplement ajoutés dans le milieu de culture. Ils sont alors supposés entrer par diffusion passive à travers la membrane ou lorsqu'ils ont une structure proche d'un acide aminé naturel par transport actif. Les études qui déterminent la voie permettant le transport d'un acide aminé non naturel sont rares. Ainsi, Harrison *et al.* ont montré que le transport de la L-4-azaleucine s'effectue probablement par le biais de transporteurs d'acides aminés aromatiques (*aroP*) et de transporteurs spécialisés des acides aminés à chaîne ramifiées (*LIV-I*) (Harrison et al. 1975). De la même façon, la surexpression de transporteurs de la proline grâce à des chocs osmotiques a permis d'augmenter le transport de l'hydroxyproline d'un facteur 4 (Buechter et al. 2003). Une étude à plus grande échelle a été réalisée par Liu et Schultz sur l'entrée de 138 acides aminés non naturels chez *E. coli*. Ils ont alors pu remarquer que les acides aminés non naturels les plus proches structurellement d'acides aminés naturels possédaient les plus forts niveaux de toxicité, probablement dû au dérèglement de voies essentielles, telles que les voies de biosynthèse ou de stockage d'acides aminés naturels, ou par leur interaction inhibitrice avec diverses enzymes (D. R. Liu and Schultz 1999). Ils ont alors particulièrement étudié le transport d'acides aminés toxiques en complétant le milieu avec un

large excès de l'acide aminé naturel avec lequel le transport était supposé compétitif. Cette complémentation du milieu permettait alors de contrecarrer cette toxicité par une compétition supposée au niveau du transport. L'apport en grande quantité de l'acide aminé naturel permettrait alors de limiter les effets toxiques dus à une inhibition de voies de synthèse ou à des interactions non spécifiques avec des enzymes (D. R. Liu and Schultz 1999). Cette méthode a permis de suggérer une large tolérance des transporteurs de l'acide glutamique et de la glutamine à des modifications de la longueur de la chaîne latérale, de la présence de groupements cétone ou méthylène (en γ), et de différents groupements en fin de chaîne latérale (D. R. Liu and Schultz 1999).

Cependant, le développement de transporteurs optimisés pour le transport d'acides aminés non naturels est encore plus rare. Très récemment, Ko et ses collaborateurs ont, pour la première fois, modifié le transporteur de la leucine chez *E. coli* afin de faciliter le transport de différents acides aminés non naturels dont essentiellement des dérivés de phénylalanine ou de lysine (Ko et al. 2019). Trois mutations (W18A, M21E, Y276L) de ce transporteur ont ainsi permis d'améliorer les concentrations cellulaires d'un facteur 2 à 3 pour l'azido-phénylalanine et d'un facteur 5 pour l'azido-lysine entraînant une augmentation de 20 à 50% de la production de la protéine cible (eGFP) avec 3 à 5 fois moins d'acide aminé non naturel (Ko et al. 2019).

En effet, l'apport de l'acide aminé non naturel peut être assez critique. Certains acides aminés ont une très faible biodisponibilité du fait de leur faible transport comme la O-sulfo-tyrosine qui est chargée négativement et qui est alors utilisée à des concentrations finales très fortes pour espérer qu'une quantité suffisante entre dans la cellule (10mM au sein du milieu de culture) (Chang C. Liu et al. 2009; Ko et al. 2019). Les difficultés ou le coût de synthèse d'un acide aminé non naturel peut alors être une limitation non négligeable. Une stratégie en développement est alors d'identifier des voies métaboliques afin de réaliser leur synthèse *in vivo* au sein de l'organisme de production. Ainsi, les voies complètes permettant la synthèse de la p-aminophénylalanine, d'une propargyl-pyrrolysine et de 4 acides aminés non naturels portant un groupement cétone ont pu être incorporées chez *E. coli* (Mehl et al. 2003; Jung et al. 2014; Ehrlich et al. 2015). Cependant, bien que cela puisse apporter des avantages *in fine* en terme de coûts, cela nécessite un travail considérable afin de déterminer les voies de synthèses nécessaires et également de mettre au point leur expression sans trop perturber le fonctionnement basal de l'organisme producteur.

IV.4.2. Conditions liées à la machinerie de la traduction

IV.4.2.1 Adaptation à la machinerie de la traduction endogène

La paire aaRS / ARNt doit être orthogonale au système de traduction endogène (section IV.3.2) mais l'interaction de l'ARNt avec les différents éléments de la traduction doit rester suffisamment efficace pour une bonne incorporation de l'acide aminé non naturel.

Tout d'abord, le facteur d'élongation (EF-Tu) permet de délivrer au ribosome les différents ARNt aminoacylés (section I.3). Ce facteur reconnaît tous les ARNt correctement aminoacylés avec une affinité similaire par son interaction à la fois au niveau de l'ARNt et de l'acide aminé. Un ARNt couplé avec le mauvais acide aminé ou avec un acide aminé non naturel peut donc avoir une affinité plus faible avec le facteur d'élongation EF-Tu, ce qui limitera sa présentation au ribosome et donc son incorporation. Ainsi, pour les acides aminés possédant une chaîne latérale large, des mutations ont permis l'élargissement de la poche de reconnaissance des NAA par l'EF-Tu d'*E. coli* (S65A, D216A, V274A) multipliant ainsi le taux d'incorporation de l'azido-phénylalanine par 4 (Gan et al. 2017). D'autre part, les acides aminés non naturels chargés, tels que les dérivés phosphatés, ont généralement un faible taux d'incorporation à la fois à cause de leur faible transport mais également à cause de leur interaction faible avec le facteur EF-Tu (Fan, Ip, and Söll 2016). Les mutations E216V,

D217G et F219G ont alors permis d'observer pour la première fois une incorporation *in vivo* de l'O-phosphotyrosine de l'ordre de 5% par rapport à l'expression de la protéine sauvage (Fan, Ip, and Söll 2016). De la même façon, les mutations H67Y, E216D, D217R, N274R ont permis de doubler la production de protéines contenant la sélénocystéine (Haruna et al. 2014). Les différentes mutations identifiées sont donc très dépendantes de l'acide aminé incorporé. De plus, le facteur EF-Tu possède un rôle de contrôle de la correcte aminoacylation de l'ARNt, sa reconnaissance doit donc à la fois conserver une polyspécificité mais aussi rester assez stricte pour ne pas favoriser des erreurs d'incorporation (section IV.1.2.1). Récemment, lors d'une étude réalisée par Deley *et al.*, la comparaison des séquences du facteur EF-Tu issues de différents organismes et de son paralogue SelB a permis d'identifier des points de mutations clés au sein de l'EF-Tu d'*E. coli* (N63A/D216A/K263A/N273A avec une prédominance pour la mutation D216A). Ces mutations permettraient de rendre le facteur d'élongation EF-Tu plus polyspécifique tout en gardant une reconnaissance suffisante pour limiter les erreurs de la traduction (Deley Cox, Cole, and Gaucher 2019)

Le ribosome a un rôle central dans la traduction (section I.3). Afin d'améliorer les taux d'incorporation d'acides aminés non naturels, des ribosomes orthogonaux ont été mis au point avec des modifications au niveau du site A ou P ou encore au niveau de la reconnaissance de l'ARNt orthogonal. L'orthogonalité du nouveau ribosome modifié est assurée par l'utilisation d'une séquence modifiée de Shine Dalgarno au niveau de l'ARNm de la protéine cible (K. Wang et al. 2007). Le facteur d'arrêt de la traduction RF1 se lie au niveau du site A (section I.3). Les mutations U531G et U534A au niveau de l'ARN ribosomal 16S déterminées par Chin et ses collaborateurs ont permis de réduire cette interaction (K. Wang et al. 2007). Le ribosome produit possédait alors une fidélité similaire aux ribosomes endogènes tout en augmentant le taux de suppression du codon TAG de 20 à plus de 60% pour une simple incorporation et de moins de 1% à plus de 20% pour une double incorporation (K. Wang et al. 2007). Le site P est le site de catalyse de la liaison peptidique (section I.3). Il a donc également été modifié afin de réaliser la formation de chaînes polypeptidiques contenant des acides aminés de la série D (Dedkova et al. 2003), ou des acides aminés de type β (Maini et al. 2013; Dedkova et al. 2012). Lors d'une étude réalisée par Terasaka *et al.* ce site a également été modifié au niveau de la reconnaissance entre le ribosome et l'ARNt. Le bras accepteur de l'ARNt finit traditionnellement par les bases CCA-3'. Les modifications ont alors permis de générer un ribosome capable de reconnaître des ARNt aminoacylés qui possédaient une extrémité CGA-3'. Ces ARNt avaient préalablement été aminoacylés par des ribozymes (section III.2). *In vitro*, la présence conjointe de ces ribosomes orthogonaux et des ribosomes sauvages permettait alors de produire deux protéines très distinctes à partir du même ARNm (Terasaka et al. 2014).

IV.4.2.2 Développement de souches spécifiques

Au cours de la traduction, des facteurs de terminaison (RF) vont favoriser l'arrêt de la traduction au niveau de codon stop (section I.3). Le facteur RF1 reconnaît les codons stops TAG et TAA, alors que le facteur RF2 reconnaît les codons stops TGA et TAA. Cependant, l'incorporation d'acides aminés non naturels (NAA) de façon site spécifique utilise généralement le détournement du codon stop TAG. L'ARNt permettant l'incorporation de l'acide aminé non naturel va alors être en compétition avec le facteur d'arrêt de la traduction RF1.

Afin d'augmenter les niveaux d'incorporation de NAA, une souche bactérienne (C321 Δ A) au sein de laquelle le facteur RF1 a été supprimé et les codons TAG endogènes remplacés par des codons stops TAA a été construite (Lajoie et al. 2013). Cependant, cette souche a un défaut de croissance et ne possède pas la T7 ARN polymérase qui permet la production de protéines recombinantes par le système d'induction T7 très communément utilisé. Par la suite, une autre

souche bactérienne a été construite (B95ΔA), à partir de la souche d'expression BL21, par la suppression du RF1 et le remplacement de 95 codons TAG de son génome (parmi les 273) par des codons TAA ou TGA afin d'avoir un impact minimum par l'apparition de mutations (Mukai, Hoshi, et al. 2015). En effet, au sein du génome d'*E. coli*, 18% des cadres de lectures en recouvre un autre (Mukai, Hoshi, et al. 2015). Aussi, le remplacement d'un codon stop TAG par un autre codon stop peut avoir des conséquences au sein de gènes portés par l'autre brin d'ADN d'*E. coli*. Ainsi, grâce à ces modifications appliquées préférentiellement aux gènes essentiels pour la croissance, la souche B95ΔA ne présente alors pas de défaut de croissance et permet également la production de protéines sous le contrôle du promoteur T7. *In fine*, la construction de ces souches a alors permis une augmentation du rendement de l'incorporation site spécifique d'acides aminés non naturels, à une ou plusieurs positions (Smolskaya and Andreev 2019). Cependant, les niveaux de production restent généralement inférieurs à la production d'une protéine sauvage. Il est possible que d'autres mécanismes encore inconnus expliquent ce défaut de production.

IV.4.2.3 Génération de nouveaux codons

Lors de l'incorporation d'acides aminés non naturels au niveau d'une protéine, un ARNt doit être chargé avec celui-ci puis utilisé en réponse au codon correspondant lors de la traduction par le ribosome pour l'incorporer au sein de la chaîne peptidique néosynthétisée. Afin d'éviter une erreur d'incorporation au sein de la protéine, le codon ne doit donc pas déjà être attribué à un autre acide aminé. Pour répondre à cette problématique, un ARN de transfert qui détourne le codon stop TAG est généralement utilisé. Ces ARNt viennent d'organismes qui incorporent des acides aminés naturels rares en réponse au codon ambre. Cependant, lors de l'incorporation de plusieurs acides aminés non naturels, il peut être nécessaire de détourner plusieurs codons. Différentes stratégies ont été développées dans ce sens.

Réaffectation de codons sens :

Le code génétique est redondant avec plusieurs codons et ARNts codants pour le même acide aminé. Cependant, certains codons sont rares ou même absents dans certains organismes. Ainsi, les codons AGG et AGA codant pour l'arginine chez *E. coli* ne sont que très rarement utilisés dans cette bactérie (respectivement 0,14% et 0,21%) (Kane 1995). Quant au codon de l'arginine CGC, celui-ci est absent du génome de certains mycoplasmes (*Mycoplasma capricolum*) (Oba et al. 1991). Il faut également noter que certains ARNt vont pouvoir reconnaître différents codons. En effet, la reconnaissance d'un codon par l'anticodon de l'ARNt peut être moins fiable au niveau de la troisième base que l'on appelle base flottante (*wobble base*). Cette reconnaissance moins stricte peut s'expliquer par un mauvais appariement induit par une méthylation (^mG/U) ou par la présence d'une base modifiée, comme l'inosine qui permet un appariement avec les bases uracile, adénine et cytosine. Les ARNts peuvent en effet être modifiés par des enzymes ce qui va influencer sur leur reconnaissance du codon (Agris et al. 2018).

La réaffectation d'un codon sens doit donc faire face à plusieurs problématiques : la reconnaissance de ce codon sens par un ARNt endogène (modifié ou non) et l'orthogonalité du couple ARNt / aaRS (Figure 28).

Il faut donc choisir une paire aaRS/ARNt orthogonale et la modifier pour la reconnaissance du codon détourné. Afin de faciliter le processus, un ARNt dont l'anticodon n'est pas reconnu par l'aaRS est généralement choisi. Il peut alors s'agir de PylRS, SerRS, LeuRS, AlaRS, HisRS (Krishnakumar and Ling 2014). Le couple PylRS / ARNt^{Pyl} est connu pour avoir une reconnaissance très spécifique et constitue un choix préférentiel. Cependant, il a été montré que la modification de l'anticodon de l'ARNt^{Pyl} entraînait un faible risque d'aminocyclation par des aaRS endogènes chez *E. coli* ou des mycoplasmes (Krishnakumar and Ling 2014). Bien que cette erreur d'incorporation soit faible, le

couple ARNt / aaRS ne serait pas alors aussi orthogonale. Ainsi, les éléments déterminants (section IV.1.2.1) qui définissent les différents modes de reconnaissance entre les ARNt et les aaRS doivent être connus afin de pouvoir assurer l'absence d'une interaction avec les aaRS endogènes au profit d'une interaction avec l'aaRS orthogonale. L'utilisation d'une paire orthogonale phylogénétiquement éloignée de l'organisme producteur permet généralement de s'assurer de cette orthogonalité (Krishnakumar and Ling 2014). L'ajout de déterminants négatifs (section IV.1.2.1) peut aussi permettre de limiter la reconnaissance du nouvel ARNt par les aaRS endogènes (Krishnakumar and Ling 2014).

Concernant la reconnaissance du codon détourné par des ARNt endogènes, il est possible de réaliser une réaffectation complète ou incomplète.

La réaffectation incomplète sera une méthode essentiellement statistique. En effet, pour cela il est possible de fournir un ARNt avec un anticodon très affiné (appariement complet) pour un codon qui est reconnu avec une affinité faible au niveau de la base flottante par les ARNt endogènes. Cette stratégie statistique peut être facilitée lors du détournement d'un codon rare ou la compétition sera plus faible (Link and Tirrell 2005). Ainsi, lors d'une étude réalisée par Tirrell et ses collaborateurs, la réaffectation du codon UUU de la phénylalanine vers la 2-naphthylalanine a conduit à une réaffectation efficace à 80% (Link and Tirrell 2005). Cependant, la PheRS reconnaît l'anti-codon de l'ARNt ce qui pourrait expliquer cette efficacité incomplète.

Afin d'effectuer une réaffectation complète, deux possibilités sont envisageables : i) supprimer l'ARNt endogène qui reconnaît le codon détourné et remplacer tous les codons génomiques lus par cet ARNt, ii) modifier l'anticodon des ARNt endogènes pour empêcher l'action des enzymes modifiant celui-ci (Krishnakumar and Ling 2014).

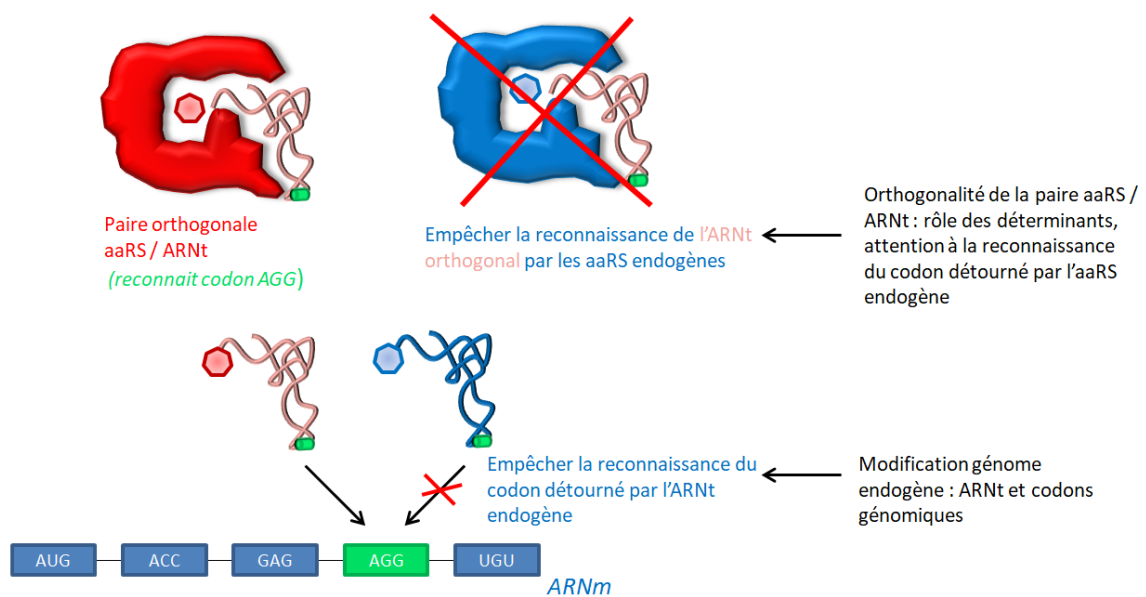


Figure 28: Méthodologie pour la réaffectation d'un codon sens (ici exemple du codon AGG).

La réaffectation de codon sens doit faire face à deux problématiques. i) La reconnaissance de l'ARNt par une aaRS endogène. En générale cette problématique est assurée par l'utilisation d'une paire aaRS / ARNt phylogénétiquement éloignée pour favoriser l'absence de réactivité croisée. ii) la reconnaissance du codon détourné par des ARNt endogènes. Pour éviter cette reconnaissance, il faut supprimer ou modifier l'ARNt endogène et remplacer tous les codons détournés (ici AGG) au niveau du génome.

Ainsi, la réaffectation de codon sens nécessite des manipulations génétiques importantes. L'utilisation de génomes synthétiques ou de méthodes de modifications génétiques à haut débit est nécessaire pour la réalisation de ce type de réaffectation (Krishnakumar and Ling 2014). De plus, contrairement à la réaffectation d'un codon stop, l'analyse de l'incorporation d'un acide aminé non

naturel est plus complexe et fait appel à une analyse en spectrométrie de masse afin de déterminer la fiabilité de cette incorporation. En effet, l'acide aminé non naturel est alors en compétition avec des acides aminés naturels plutôt qu'avec une terminaison de la traduction. Ces méthodes ont été utilisées chez *E. coli*, chez la levure, ou au niveau de mycoplasmes dont le génome est assez réduit (Krishnakumar and Ling 2014).

Par exemple, une étude réalisée par Mukai *et al.*, a permis la réaffectation du codon AGG à l'aide d'un mutant de la PylRS et du remplacement de 27 codons AGG génomiques par des codons synonymes au niveau de gènes essentiels. Cependant, les ARNt endogènes reconnaissent à la fois le codon AGG et AGA, grâce à la troisième base flottante. Afin d'éviter la modification du codon AGA, les anti-codons des ARNt endogènes ont donc été modifiés et l'ARNt reconnaissant le codon AGG a été codé au sein d'un plasmide qui n'est pas maintenu à haute température. La capacité d'*E. coli* à réaliser la traduction d'AGG par une arginine a donc été rendue thermosensible. Ainsi, grâce à l'ajout de la paire orthogonale PylRS / ARNt orthogonale, à haute température les arginines codées par AGG étaient remplacées par l'acide aminé non naturel (ici une homoarginine). (Mukai, Yamaguchi, et al. 2015).

Utilisation d'un code génétique à 4 bases :

Le cadre de lecture d'un gène est défini par la succession des triplets qui codent pour le gène. Il commence par un codon initiateur ATG et se termine par un codon stop. Naturellement, des décalages du cadre de lecture peuvent être observés. Le décalage du cadre de lecture fait généralement référence à l'insertion (+1) ou à la délétion (-1) d'une base au niveau du gène ce qui décale la lecture réalisée par le ribosome. Cependant, il existe également des ARNt supprimeurs de ces décalages du cadre de lectures (+1) (Farabaugh 2000). Ces ARNt possèdent des anti-codons à 4 bases appelés quadruplets. Leur utilisation afin d'incorporer des acides aminés non naturels a d'abord essentiellement été réalisée par le biais d'ARNt pré-couplé *in vitro* ou par microinjections (sections III.2) (K. Wang, Schmied, and Chin 2012). Cependant, l'efficacité de la traduction par ces ARNt ayant un anti-codon à 4 bases est très faible, à cause d'une mauvaise adaptation au ribosome. *In vivo*, l'augmentation de cette efficacité induit une toxicité importante, probablement par la modification du protéome. Néanmoins, Chin et ses collaborateurs ont créé une voie orthogonale de la traduction pour favoriser cette prise en charge comportant un ARNm orthogonal, un ribosome orthogonal et l'ajout des ARNt portants des quadruplets. L'orthogonalité de l'ARNm a été réalisée par la modification de la séquence Shine Dalgarno qui permet l'assemblage du ribosome (section I.3). Le ribosome a alors été modifié pour reconnaître cette séquence et pour permettre une meilleure accommodation des ARNt quadruplets. Cette stratégie a alors permis l'incorporation de deux acides aminés non naturels par les paires aaRS / ARNt issues de *M. jannaschii* et *M. barkeri* en réponse au codon stop TAG et à un codon quadruplet AGGA (K. Wang, Schmied, and Chin 2012).

Utilisation de bases nucléotidiques non naturelles :

La mise au point de nouvelles bases nucléotidiques (donc différentes de A, T, U, G, C) pourrait fournir des codons n'ayant aucune attribution naturelle. Ces codons pourront alors être attribués à l'acide aminé non naturel à incorporer. L'inefficacité de la prise en charge des bases non naturelles lors des processus de réplication, transcription, et traduction a longtemps limité leur utilisation au sein de production *in vitro* (Appella 2009). Cependant, récemment, Romesberg et ses collaborateurs de l'institut de Craig Venter ont réussi à générer une bactérie capable de croître avec deux bases supplémentaires, ce qui permettrait alors la génération de 216 codons. De plus, au cours de ces travaux, ils ont démontré leur utilisation *in vivo* pour l'incorporation d'acides aminés non naturels permettant de réaliser des conjugaisons spécifiques (propargyl-lysine et azido-phénylalanine

indépendamment) au sein de la protéine modèle GFP (Y. Zhang et al. 2017). Bien que l'utilisation de bases non naturelles soit encore très liée à des questions de recherche fondamentale, ces résultats démontrent le potentiel applicatif de leur utilisation.

IV.4.3. Conditions liées à l'expression du couple aaRS / ARNt et de la protéine cible

De manière générale, l'incorporation d'un acide aminé non naturel au sein d'*E. coli* se réalise grâce à sa co-transformation par deux plasmides (Figure 27). Un plasmide porte alors le couple aaRS / ARNt orthogonal à la traduction endogène et le second porte la séquence codante pour la protéine cible au sein de laquelle un codon stop TAG a été généré au niveau du site d'incorporation ciblé.

Tout d'abord, afin de favoriser une co-transformation correcte de ces deux plasmides, non seulement ils doivent posséder des gènes de sélections différents (résistance aux antibiotiques) mais ils doivent également posséder des origines de réplication différentes. En effet, lors de la division cellulaire, les plasmides vont être répliqués et les différentes copies plasmidiques vont alors être réparties entre la cellule mère et la cellule fille. Si deux plasmides ont des origines de réplication semblables, leur répartition au cours de la division risque d'être faussée limitant leur maintien au cours des différentes générations (Austin and Nordström 1990). La combinaison des origines de réplifications la plus utilisée pour l'incorporation d'acide aminé non naturel est p15A et pBR322 (Smolskaya and Andreev 2019).

Au niveau du plasmide portant les éléments orthogonaux, à notre connaissance, il n'existe pas d'étude rationnelle et exhaustive ayant testé les nombreux promoteurs, les différents inducteurs (à différentes concentrations et temps d'ajout), les conditions de cultures, ainsi que le nombre de copies du couple aaRS / ARNt et leur proportion. Cependant, quelques combinaisons ont été testées et sont détaillées ci-dessous.

Au niveau du couple TyrRS / ARNt^{Tyr} issu de *M. jannaschii*, le plasmide pSup-MJ^{Tyr}TyrRS-6TRN a montré des taux d'incorporation améliorés chez *E. coli*. Ce plasmide possède une origine de réplication p15A avec 6 copies d'ARNt^{Tyr} au sein de deux cassettes sous le contrôle d'un promoteur issu d'*E. coli* (*proK*) et une copie de la TyrRS sous le contrôle du promoteur *glnS'* au lieu de *glnS*. Cette modification au niveau du promoteur le rend plus efficace car insensible à la méthylation du site dam (5'GATC^{3'}) (Plumbridge and Söll 1987). Enfin la synthétase TyrRS possède la mutation D286R qui favorise son interaction avec l'ARNt^{Tyr} suppresseur (Smolskaya and Andreev 2019).

Par la suite, le plasmide pEVOL a montré des meilleurs taux d'incorporation avec des pyrrolysyl ARNt synthétases (*M. mazei* et *M. barkeri*) ainsi qu'avec des tyrosyl ARNt synthétases (*M. jannaschii*) (Smolskaya and Andreev 2019). Ce plasmide possède deux copies de l'aaRS, une sous le contrôle d'un promoteur constitutif (*glnS'*) et l'autre sous le contrôle d'un promoteur inductible à l'arabinose ainsi qu'une copie de l'ARNt toujours sous le contrôle du promoteur *proK*. Cependant, le bras T de l'ARNt a été modifié afin favoriser son interaction avec le facteur d'élongation (EF-Tu) (T. S. Young et al. 2010).

Un troisième type de plasmide (pUltra) a également été développé afin d'optimiser l'incorporation d'acides aminés non naturels (Chatterjee et al. 2013). Contrairement aux plasmides pSup et pEVOL précédents, celui-ci comporte une origine de réplication *cdf* qui permet de maintenir plus de copies (par rapport à l'origine p15A) au sein de la bactérie tout en restant toujours compatible avec la plupart des plasmides d'expression classiques. Il possède le même type d'expression de l'ARNt que pour le plasmide pEVOL mais une seule copie de l'aaRS sous le contrôle d'un promoteur inductible à l'IPTG. Ce plasmide a alors pu être utilisé en combinaison avec le plasmide pEVOL et un plasmide d'expression lors de l'incorporation concomitante de deux types d'acides aminés non naturels différents (section IV.4.4).

Au niveau du plasmide portant la séquence de la protéine cible, les conditions optimales d'expression sont très dépendantes de la cible protéique. Cependant, l'association entre les deux plasmides d'expression joue probablement un rôle important. Ainsi, l'expression du couple aaRS / ARNt et de la protéine cible se réalise généralement de façon indépendante (Q. Wang, Parrish, and Wang 2009). Cette stratégie nécessite donc deux systèmes d'induction indépendants qui sont généralement des promoteurs très communément utilisés comme les promoteurs inductibles à l'arabinose et à l'IPTG. Malheureusement, la description de la mise au point de ces différents paramètres est rarement développée au sein des différentes publications. Néanmoins, Wandrey *et al.* ont réalisé un plan d'expérience afin de déterminer une combinaison des conditions optimales lors de l'incorporation de la propargyl-lysine ou de l'acide (S)-2-amino-6-((2-azidoéthoxy)carbonylamino) hexanoïque (ALK) au sein de la GFP (Wandrey *et al.* 2016). Au sein de leur système d'expression, l'ARNt est exprimé constitutivement et l'aaRS ainsi que la protéine cible sont exprimées sous le contrôle d'un promoteur inductible à l'IPTG. Les conditions optimales qu'ils ont alors déterminées consistaient en une induction avec 300 à 400 μM d'IPTG entre 0 et 0,5 h et un ajout de l'acide aminé dès le départ à une concentration très élevée pouvant atteindre 10 μM pour ALK et 38 μM pour la PrK (Wandrey *et al.* 2016).

IV.4.4. Incorporation de deux acides aminés non naturels différents au sein d'une même séquence

L'incorporation de deux acides aminés non naturels différents peut se réaliser par une utilisation conjointe de la méthode d'incorporation résidu spécifique et site spécifique. Au sein d'une souche auxotrophe, un codon sens peut alors servir à l'insertion d'un acide aminé non naturel (NAA) tandis que l'expression d'une paire aaRS/ARNt orthogonale peut servir à l'incorporation d'un autre NAA à la place d'un codon stop (Ayyadurai *et al.* 2011). Cependant, si le contrôle de la position d'incorporation est important pour les deux partenaires, il sera alors nécessaire d'utiliser deux paires aaRS/ARNt orthogonales. En plus d'être orthogonales au système de traduction endogène, ces paires devront également être orthogonales entre elles afin de rester indépendantes. Les stratégies mises au point utilisent généralement le détournement du codon stop TAG et TGA. Au niveau des paires orthogonales entre elles, le couple PylRS/ARNt (*M. mazei* ou *M. barkeri*) et le couple TyrRS/ARNt (*M. jannaschii*) peuvent être utilisés ensemble au sein d'*E. coli* (K. Wang *et al.* 2014). Généralement, les niveaux d'incorporation sont très faibles, cette double incorporation a donc principalement été utilisée pour des applications nécessitant peu de matériel comme le FRET (*Florescence Resonance Energy Transfert*) (Quast *et al.* 2019; T. C. Lee *et al.* 2016). Très récemment, une troisième voie orthogonale a été également couplée aux voies orthogonales précédentes (PylRS / ARNt^{Py} et TyrRS / ARNt^{Tyr}) par l'utilisation d'un couple TrpRS / ARNt^{Trp} dérivé d'*E. coli* (Italia *et al.* 2019). La paire TrpRS / ARNt^{Trp} a été générée par évolution au sein d'*E. coli* à la suite du remplacement de la paire TrpRS endogène par une paire issue de la levure. La nouvelle paire résultante a alors pu montrer une orthogonalité à la fois chez *E. coli* et chez les eucaryotes (Italia *et al.* 2017). La suppression des trois codons stops a été réalisée par l'utilisation conjointe de ces aaRS / ARNt et a alors permis l'incorporation de trois acides aminés non naturels portant des fonctions de conjugaison orthogonales donc très spécifiques et indépendantes. L'efficacité d'incorporation étant très limitée (rendement de production de 2% par rapport à la protéine sauvage), la terminaison de la traduction de la protéine cible a alors été réalisée grâce à la multiplication de codons stops à la fin du gène d'intérêt (Italia *et al.* 2019).

IV.5 Incorporation d'acides aminés non naturels chez les eucaryotes

Jusqu'à ce point d'introduction, nous n'avons que très peu évoqué la méthodologie d'incorporation d'un acide aminé non naturel au sein d'un organisme eucaryote. Cependant, bien que les processus de traduction diffèrent, la méthodologie d'incorporation des acides aminés non naturels est très similaire et est basée sur l'utilisation de paires aaRS / ARNt orthogonales. Nous allons cependant voir ci-dessous les principaux éléments spécifiques à l'incorporation d'acides aminés non naturels au sein d'eucaryotes.

Tout d'abord, comme nous venons de le soulever précédemment (section IV.3.2), le couple aaRS/ARNt utilisé doit être orthogonal au système de traduction de l'organisme producteur. La pyrrolysyl ARNt synthétase (PylRS) des archées ayant une reconnaissance de l'ARNt très spécifique, elle est utilisable à la fois chez les procaryotes et les eucaryotes. Des couples aaRS / ARNt issus d'*E. coli* (codant pour la leucine ou la tyrosine) sont également très utilisés au sein de cellules eucaryotes pour réaliser l'incorporation d'acides aminés non naturels. Lorsque ces ARNt bactériens sont utilisés (généralement ARNt^{Leu} ou ARNt^{Tyr}), leur anticodon doit être muté en CUA pour être complémentaire au codon ambre. Cette mutation peut baisser leur affinité envers l'aaRS, et des mutations supplémentaires peuvent être appliquées afin de maintenir une bonne reconnaissance entre ces partenaires (Lei Wang 2017). L'utilisation de paires orthogonales à la fois au sein des eucaryotes et des procaryotes peut faciliter le développement de nouveaux acides aminés non naturels. En effet, l'évolution d'un couple aaRS / ARNt est généralement réalisée chez *E. coli*, ou à défaut chez la levure (section IV.3.3) (Lei Wang 2017).

Par ailleurs, la paire TyrRS/ARNt^{Tyr} issue de *M. jannaschii* n'est pas utilisable telle quelle chez les eucaryotes. En effet, les TyrRS issues des eucaryotes et archées ont un mode de reconnaissance de l'ARNt proche avec notamment la même paire de nucléotides discriminante (C1:G72) (section IV.3.2.2). Cependant, des travaux réalisés par Thibodeaux *et al.* en 2010 ont permis de rendre cette paire orthogonale chez les eucaryotes en remplaçant les éléments de reconnaissance par leur équivalent bactérien. Étant donné que de nombreux mutants avec différentes spécificités ont été développés chez *E. coli*, cela permet d'apporter un nouveau lot d'aaRS / ARNt utilisables chez les eucaryotes (Thibodeaux *et al.* 2010).

L'expression de l'ARNt provenant des archées au sein de la cellule eucaryote s'est également avérée problématique. En effet, l'induction de l'expression des ARNt procaryotes est physiologiquement induite par un promoteur en amont de leur séquence. Cependant, l'expression des ARNt eucaryotes est induite grâce à la reconnaissance de séquences conservées au sein de leur gène (boîte A et boîte B) par des facteurs de transcription. Toutefois, dans les années 80, un promoteur externe de l'ARN polymérase III de type 3 a été identifié au sein de cellules mammifères (Schramm and Hernandez 2002). Celui-ci est présent en amont du gène et est composé d'une boîte TATA ainsi que de deux séquences conservées "Proximal Sequence Element" (PSE) et "Distal Sequence Element" (DSE) (Schramm and Hernandez 2002). Son utilisation couplée à une séquence additionnelle en 3' ainsi que la présence de cassette de plusieurs copies d'ARNt a permis l'expression d'ARNt procaryotes ou issus d'archées avec un niveau suffisant pour réaliser l'incorporation d'acides aminés non naturels (Lei Wang 2017). Ainsi, cette stratégie développée pour la première fois par Wang *et al.* en 2007 a également été efficacement utilisée au sein d'une variété d'organismes eucaryotes y compris des animaux et des plantes (Lei Wang 2017). Un autre promoteur de la polymérase III externe au gène de l'ARNt (issus de la levure), a également permis une expression plus forte des ARNt^{Tyr} et ARNt^{Leu} bactériens (Lei Wang 2017).

Tout comme pour les procaryotes (section IV.4.1), l'acide aminé non naturel doit pouvoir entrer dans la cellule afin de pouvoir être incorporé *in vivo*. Le masquage de leur fonction carboxylique par la formation d'un ester peut permettre de faciliter leur apport. En effet, l'acide aminé alors plus neutre et lipophile peut traverser la membrane plus facilement. La présence d'estérases non spécifiques intracellulaires chez des cellules eucaryotes permet alors de rendre l'acide aminé disponible pour une incorporation. Le rendement d'incorporation a pu être ainsi multiplié par 4 avec une réduction significative de la quantité d'acide aminé non naturel nécessaire (Lei Wang 2017). Des transporteurs de peptides peuvent être présents chez les cellules eucaryotes et l'utilisation de dipeptides permet donc également d'augmenter considérablement leur biodisponibilité (Lei Wang 2017).

Enfin, au sein des cellules eucaryotes, un mécanisme appelé *nonsense-mediated mRNA decay* (NMD) cible les ARN messager (ARNm) avec un codon stop précoce et stimule sa dégradation. En effet, les mécanismes d'épissage de l'ARNm peuvent entraîner des erreurs ce qui serait délétère pour la cellule. Ainsi, après une première traduction de l'ARNm, si un codon stop est situé trop en amont, la protéine UPF1 va alors se fixer à celui-ci et recruter différents partenaires protéiques dont des enzymes entraînant la dégradation de l'ARNm (Lykke-Andersen and Jensen 2015). Etant donné que l'incorporation d'acides aminés non naturels de manière site spécifique se réalise généralement par la suppression du codon stop ambre, ce mécanisme réduit la quantité de protéine modifiée produite. Ainsi, la suppression du gène UPF1 au sein d'une souche de levure a doublé l'incorporation d'acide aminé non naturel. Cet effet a également été démontré au sein de *Caenorhabditis elegans* par l'utilisation d'ARN interférent augmentant les rendements par 5,6 (Lei Wang 2017).

Au cours de cette thèse, nous nous concentrons essentiellement sur l'incorporation d'acides aminés non naturels chez *E. coli* car l'ensemble du travail réalisé ici s'est déroulé au sein de cet organisme. Cependant, différents acides aminés non naturels ont efficacement été incorporés chez les eucaryotes dont principalement la levure, des cellules mammifères, des cellules souches, ou encore des animaux comme des vers (*C. elegans*), drosophiles, poissons zèbres et souris (Lei Wang 2017). L'utilisation de ces modèles permet alors l'étude de mécanismes cellulaires ou physiologiques plus proche de l'Homme.

V. Applications

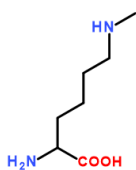
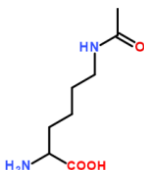
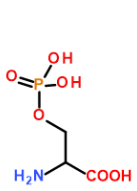
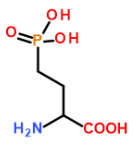
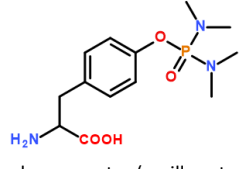
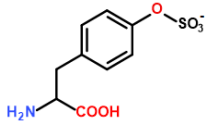
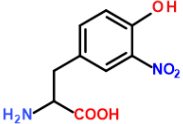
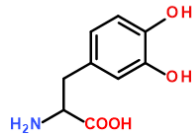
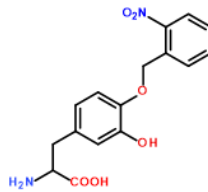
L'utilisation d'acides aminés non naturels se prête à l'étude de nombreux domaines biologiques. Les différentes utilisations peuvent être classées selon 3 grandes catégories : i) l'étude de modifications post traductionnelles, ii) la bioconjugaison orthogonale, iii) l'utilisation au sein d'un site de reconnaissance ou site actif.

V.1. Etude des modifications post traductionnelles

En général les modifications post-traductionnelles sont des processus dynamiques et plusieurs types de modifications peuvent être appliqués sur une même protéine selon les conditions physiologiques (H. Chen et al. 2018). Leur étude est donc complexe et l'utilisation d'acides aminés non naturels permet de simplifier les paramètres tout en permettant l'étude de ces modifications au sein de l'environnement cellulaire natif. Ainsi, de nombreux acides aminés non naturels ont été utilisés pour ce type d'étude et ont été répertoriés au sein de la revue bibliographique réalisée par Chen *et al.* (H. Chen et al. 2018). Afin d'incorporer ces acides aminés porteurs de modifications post-traductionnelles ou de mimes, la méthode d'incorporation site-spécifique est privilégiée. En effet, la modification concerne généralement des acides aminés abondants au sein des protéines et cibler un

site particulier permet d'avoir une étude plus précise (H. Chen et al. 2018). Les acides aminés développés sont donc porteurs des modifications post-traductionnelles qui peuvent être modifiées au sein de la cellule, ou consistent en des analogues (par exemple non clivables *in vivo*) ou précurseurs de ces mutations (H. Chen et al. 2018). Ceux-ci ont alors permis d'étudier la méthylation et l'acétylation des lysines et arginines, la phosphorylation des sérines, tyrosines et thréonines, ainsi que la sulfatation, la nitration et l'hydroxylation des tyrosines (H. Chen et al. 2018) (Tableau 6).

Tableau 6 : Exemples d'acides aminés non naturels pour l'étude de modifications post-traductionnelles

Modification post-traductionnelle étudiée	Acide aminé non naturel		Référence	
Méthylation et acétylation de la lysine			(Neumann, Peak-Chew, and Chin 2008) (Lammers 2018)	
Phosphorylation de sérine et de la tyrosine		 Analogue non hydrolysable de la phosphosérine	 Analogue neutre (meilleur taux d'incorporation que pour la Phospho-tyrosine)	(Rogerson et al. 2015) (Hoppmann et al. 2017)
Sulfatation de la tyrosine			(Chang C. Liu et al. 2009)	
Nitration de la tyrosine			(Gerding et al. 2019)	
Hydroxylation des tyrosines		 Hydroxylation masquée par un groupement photolabile	(Alfonta et al. 2003) (Hauf et al. 2017)	

V.2. Bioconjugaison orthogonales

Les acides aminés non naturels peuvent donc permettre d'apporter une réactivité dite orthogonale, c'est à dire absente des molécules du vivant. Ces caractéristiques ont été très exploitées afin de réaliser des conjugaisons ciblées entre différents partenaires. Ainsi, différents types de réactivité ont pu être incorporés au sein des protéines et ont été plus particulièrement détaillés au sein du chapitre 2. La revue bibliographique récente réalisée par Lee *et al.* répertorie les différents acides aminés permettant ces bioconjugaisons orthogonales (K. J. Lee, Kang, and Park 2019). Ces nouvelles méthodes de conjugaisons orthogonales ont pu être utilisées dans le domaine des biomatériaux, de génération d'anticorps conjugués, d'imagerie cellulaire, de génération de protéines artificielles multivalentes, de délivrance spécifique de médicaments, ou encore d'études structurales par FRET (E. Kim and Koo 2019; Spicer, Pashuck, and Stevens 2018). Nous pouvons ainsi noter que ces méthodes de conjugaison sont très utilisées par la communauté scientifique.

Cependant, l'utilisation d'acides aminés non naturels afin de produire des glycoconjugués à des fins vaccinales comme il sera développé au sein du chapitre 3 est très rare. A notre connaissance, uniquement un brevet de la société SutroVax et une publication réalisée par un groupe partenaire ont pu être répertoriés (Fairman, Heinrichs, and Chan 2018; Kapoor et al. 2018). Cette publication fait état de l'incorporation de l'azido-phénylalanine au sein d'une protéine antigénique de la malaria par une méthode d'incorporation *in vitro*. Cette protéine modifiée a ensuite été spécifiquement couplée avec un glycosylphosphatidylinositol générant ainsi un vaccin glycoconjugué (Kapoor et al. 2018). Au cours de la présente thèse, la technologie d'incorporation d'acides aminés non naturels a tout d'abord dû être implémentée au sein du laboratoire. Ensuite, cette technologie a été appliquée à la génération de glycoconjugué à des fins vaccinales, ce qui constitue l'originalité de nos travaux.

V.3. Incorporation de NAA au sein de sites actifs ou de reconnaissance

L'étude de sites de reconnaissance ou sites actifs par l'utilisation d'un acide aminé non naturel nécessite l'utilisation d'une méthode d'incorporation ciblée. Différentes applications ont pu être développées telles que l'étude des caractéristiques des protéines (au niveau biophysique ou encore des caractéristiques clés d'un acide aminé), la modification des caractéristiques de reconnaissance (renforcement de la liaison de co-facteurs, contrôle optique des propriétés, modification régiosélectivité / énantiosélectivité). Ces applications seront plus particulièrement détaillées lors du chapitre 5 traitant de la modification des lectines. Globalement, nous pourrions noter que l'utilisation des NAA dans le but de modifier les caractéristiques de reconnaissances des protéines sont assez rares et que les travaux présentés lors du chapitre 6 sont donc particulièrement originaux.

***Chapitre 2 : Stratégies de lutte contre
Streptococcus pneumoniae***

I. Généralités sur *Streptococcus pneumoniae*

I.1 Physiologie

Streptococcus pneumoniae est communément appelée "pneumococque" et cause des pathologies chez l'Homme principalement au niveau du système ORL (Oreille – Nez – Larynx) (section I.3). Au niveau physiologique, c'est une bactérie aéro-anaérobie facultative qui peut donc croître à la fois en milieu pauvre et riche en oxygène. Cette bactérie est Gram positive et possède donc une large paroi formée de peptidoglycane (composé d'unités répétées de N-acétylglucosamine et d'acide N-acétylmuramique reliées entre elles par des chaînes peptidiques (Figure 29A). Des protéines de surface ainsi que la capsule polysaccharidique γ sont ancrées. Au niveau morphologique, elle se présente sous la forme de *cocci* le plus souvent associés en paires (appelés diplocoques) entourées d'une capsule polysaccharidique (Figure 29B et C).

La production de la capsule est régulée lors des différents stades de l'infection (section I.4) par des facteurs environnementaux comme la source de carbone (dominance du glucose dans le sang et du galactose dans le mucus respiratoire) (Troxler et al. 2019) ou la teneur en oxygène (J N Weiser et al. 2001). Au niveau macroscopique, deux morphologies peuvent donc être observées : des colonies opaques et des colonies transparentes (Figure 29D)(Manso et al. 2014). Les bactéries ayant une capsule épaisse (d'apparence opaque) seraient plutôt invasives et donc retrouvées par exemple dans le sang tandis que celles possédant une capsule plus fine correspondraient à une phase physiologique permettant l'expression de protéines de surface lors la colonisation afin de promouvoir l'adhérence aux cellules épithéliales cibles (Chai et al. 2017). La capsule joue donc un rôle important au cours des différentes étapes d'infection, ce qui est détaillé section I.4.

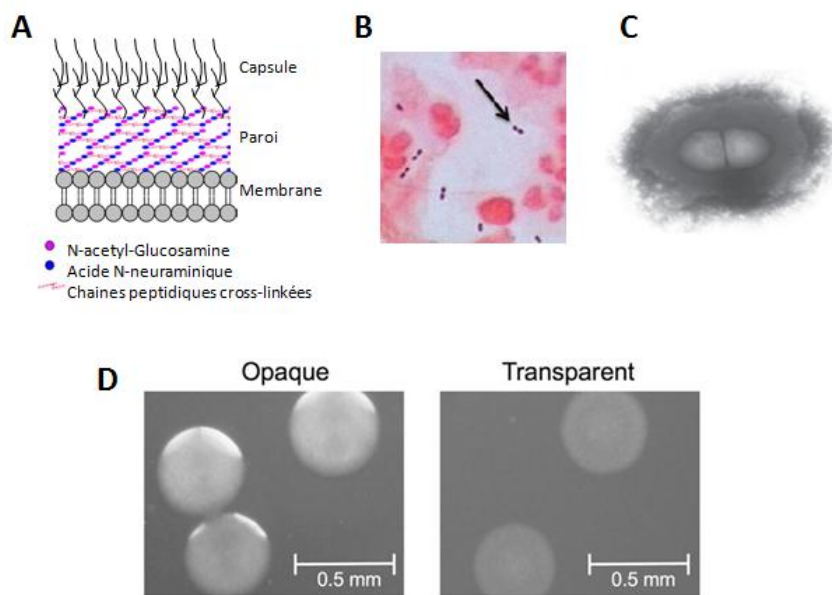


Figure 29: Aspect morphologique de *Streptococcus pneumoniae*

A) Représentation schématique de la structure externe des bactéries de type gram +, Image modifiée à partir du site https://wikieducator.org/Bacterial_Structure

B) Coloration gram issue d'un échantillon de cellules lymphocytaires en présence de *S. pneumoniae*, image extraite du site CDC : <https://www.cdc.gov/meningitis/lab-manual/chpt06-culture-id.html>

C) Représentation schématique de *S. pneumoniae* sous la forme de diplocoques encapsulés (Tomos, Perros, and Dimopoulos 2017)

D) Image issue de colonies du sérotype 3 opaques et transparentes observées au microscope (Chai et al. 2017)

1.1.1 Composition de la capsule

La nature de la capsule a été identifiée en 1925 par Avery *et al.* qui l'ont alors désignée comme le premier antigène non protéique (O T Avery and Heidelberger 1925). Cette capsule constitue la couche la plus externe de *S. pneumoniae*. Elle est composée d'unités saccharidiques répétées (homopolysidiques) ou diverses (hétéropolysidiques) formant des polymères branchés ou linéaires de haut poids moléculaire (500 à 2000 kDa) (Rieux 2002). Sa composition varie d'une souche à l'autre et ces souches sont alors classées en sérotypes. En effet, elles peuvent être catégorisées selon la reconnaissance de la capsule par des anticorps issus d'un sérum. Les anticorps sont alors dirigés contre des motifs saccharidiques spécifiques. Actuellement, 97 sérotypes ont été identifiés et la composition de leur capsule a notamment été répertoriée au sein de la publication de Geno *et al.* (Geno et al. 2015). La diversité potentielle des polysaccharides est très importante car elle porte à la fois sur les différents sucres les constituant et leurs différents substituants (phosphate, groupement acétyle) ainsi que sur leur enchainement et leur type de lien osidique (α ou β). Par exemple, si l'on considère le nombre de combinaisons de polymères potentiels à partir de 20 unités pour un polymère de longueur de 6 unités, la diversité potentielle de peptides est très inférieure à la diversité de polysaccharides potentiellement générés avec respectivement environ 64 millions de combinaisons pour des peptides et 192 780 millions pour les polysaccharides (Werz et al. 2007). Les sucres constituant les polysaccharides capsulaires sont généralement chargés négativement à cause de la présence de groupements phosphates ou d'acide glucuronique (donc sous leurs formes oxydées) (Geno et al. 2015). La composition de la capsule des principaux sérotypes impliqués dans les pathologies liées à *S. pneumoniae* est répertoriées dans le tableau présenté en Annexe n° 1. Trois exemples représentatifs de composition polysaccharidique de sérotypes sont présentés ci-dessous Tableau 7.

Tableau 7 : Exemple de 3 structures de polysaccharides capsulaires

Glcp signifie Glucose sous la forme pyranosyle (*p*), *GlcpA* signifie Glucose sous sa forme d'acide glucuronique, *GlcNAc* signifie N-acétyl-glucosamine, *Galp* signifie Galactose sous sa forme pyranosyle, *Rha* signifie Rhamnose, *Rib-ol* signifie Ribitol, *P* représente un groupement phosphate.

Sérotipe	Composition polysaccharidique
3	$\rightarrow 3$ - β -D-GlcpA-(1 \rightarrow 4)- β -D-Glcp-(1 \rightarrow)
6B	$\rightarrow 2$ - α -D-Galp-(1 \rightarrow 3)- α -D-Glcp-(1 \rightarrow 3)- α -L-Rhap-(1 \rightarrow 4)-D-Rib-ol-(5 \rightarrow P \rightarrow)
14	$\rightarrow 6$ - β -D-GlcpNAc-(1 \rightarrow 3)- β -D-Glcp-(1 \rightarrow 4 ↑ 1 β -D-Galp

1.1.1.1 Méthodes d'identification de sérotypes

Les sérotypes sont identifiés selon deux systèmes d'annotations : un Américain et un Danois, bien que le système Danois soit plus largement utilisé par la communauté scientifique. En effet, celui-ci tient compte de la réactivité croisée, c'est à dire de la reconnaissance par un même sérum de différents sérotypes. Ces sérotypes sont alors regroupés au sein de sérogroupes et leurs annotations sont différenciées par une lettre (exemple les sérotypes 19A et 19F du séro groupe 19) (Henrichsen 1979). Ceux-ci sont généralement structurellement très proches. Par exemple le sérotipe 19A est différencié du sérotipe 19F par une seule liaison (Figure 30). Cependant, bien que d'apparence minime, cette modification a une influence importante sur sa flexibilité et le nombre de conformations adoptées par le polysaccharide (Figure 30)(Kuttel et al. 2015). En effet, comme nous

pouvons le voir sur la Figure 30, les expériences de dynamique moléculaire couplées à des analyses RMN réalisées par Kuttel *et al.* présentent différentes conformations structurales en milieu aqueux, et les profils ne sont pas équivalents pour les deux sérotypes. Notamment, une conformation stabilisée par des liaisons hydrogènes intramoléculaires apparaît clairement (entourée en pointillés à gauche de la Figure 30, et présentation de sa structure à droite de la Figure 30). De plus, ces modifications structurales semblent avoir une importance biologique puisque les anticorps produits contre le sérotype 19F ne présentent qu'une efficacité limitée contre une infection par le sérotype 19A (Hausdorff, Hoet, and Schuerman 2010; R. Reinert, Jacobs, and Kaplan 2010). La caractérisation précise de la composition des différents sérotypes a donc un sens structural mais également biologique très important, d'autant qu'ils permettent le développement de stratégies vaccinales basées sur leurs compositions (section II.2).

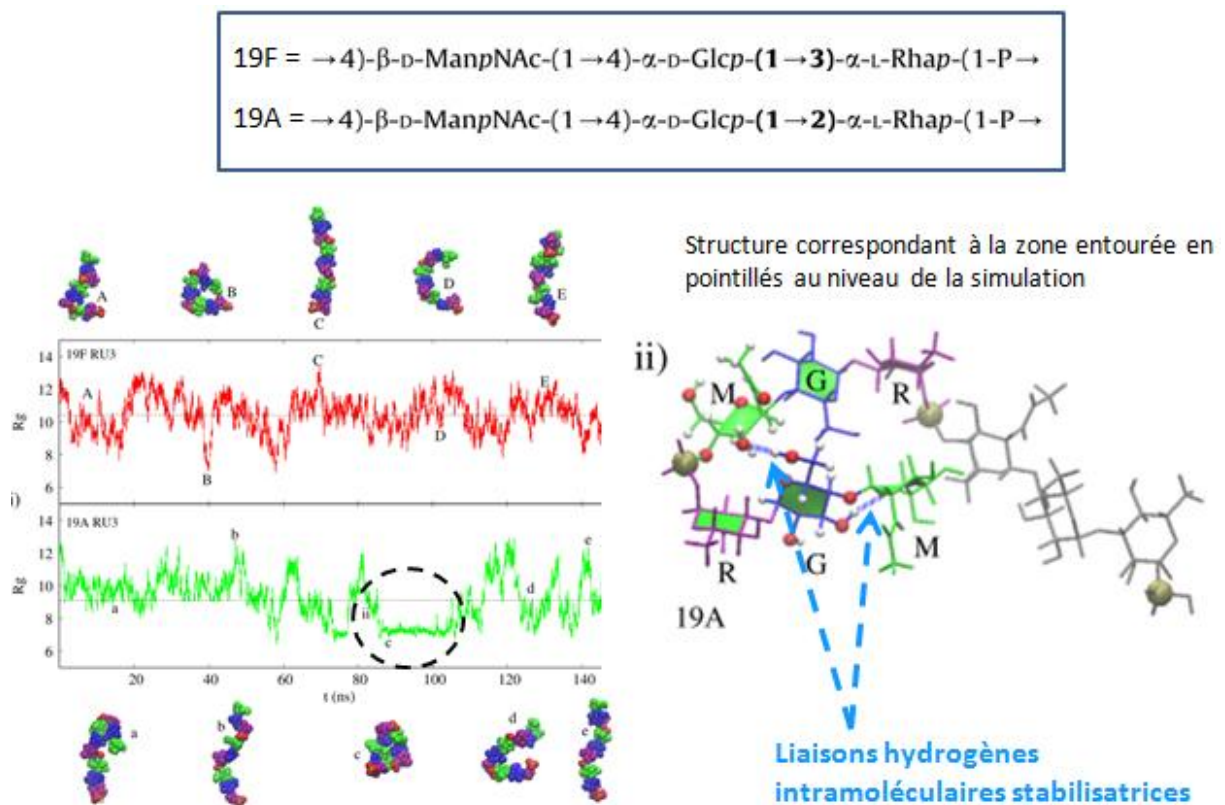
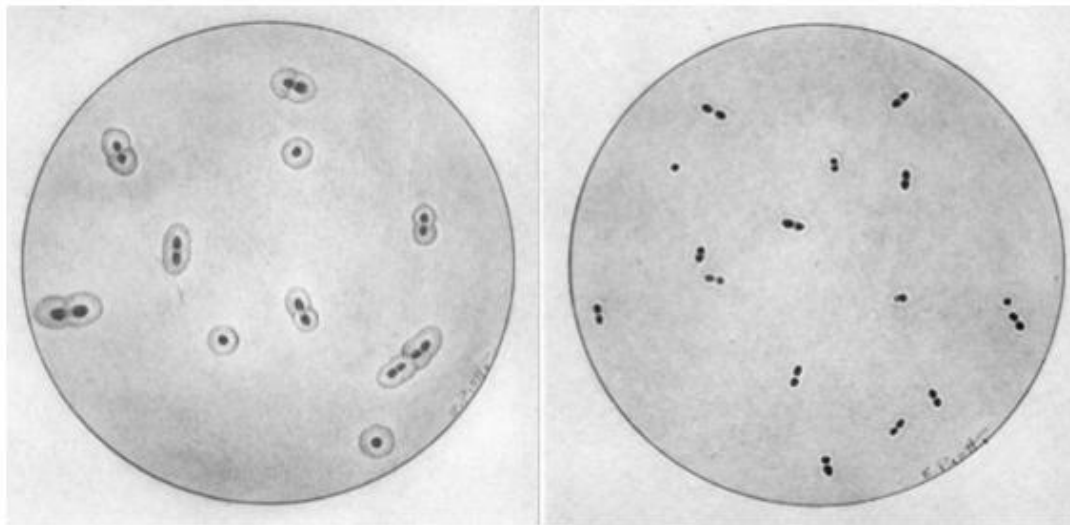


Figure 30: Différences structurales entre le sérotype 19A et 19F
 Analyse de dynamique moléculaire du Rayon de giration du polysaccharide capsulaire (courbe rouge pour le sérotype 19F et courbe verte pour le sérotype 19A) couplée à des analyses RMN de certaines conformations (schématisées de part et d'autres des courbes). La structure présentée à droite de l'image correspond à la conformation stabilisée du sérotype 19A entourée en noir sur la figure à gauche. Image issue de la publication de (Kuttel et al. 2015).

Plusieurs méthodes ont été développées afin de permettre le classement en différents sérotypes notamment par des méthodes immunologiques telles que le test de précipitation, d'agglutination ou la réaction de Quellung. La réaction de Quellung est une méthode simple et rapide et est donc très utilisée (Habib, Porter, and Satzke 2014). Cette réaction consiste à mélanger une suspension de la souche de *S. pneumoniae* à identifier avec un sérum. L'opacification de la capsule par la liaison à des anticorps spécifiques de celle-ci lui donne une apparence plus gonflée, et ainsi plus visible au microscope (Figure 31). Dans le cas d'une réponse négative, la capsule reste mince (Figure 31). Cependant, ces méthodes peuvent être présentées comme subjectives à cause de

l'interprétation non quantitative des résultats. De plus, une capsule endommagée ou faiblement exprimée ne permet alors pas un sérotypage par cette technique (Geno et al. 2015).



Réaction positive (capsules gonflés)

Réaction négative

Figure 31: Identification de sérotype par la réaction de Quellung

Les bactéries dont le sérotype est à identifier sont mises au contact d'un sérum. Lorsque le sérum contient des anticorps qui reconnaissent des motifs polysaccharidiques de la capsule, celle-ci apparaît opacifiée et plus large (à gauche sur la figure). Dans le cas d'une réponse négative, les anticorps ne reconnaissent pas la capsule et celle-ci reste fine en apparence (à droite sur la figure). Image adaptée à partir de (White, Robinson, and Barnes 2008)

Le développement de méthodes analytiques telles que la chromatographie en phase gazeuse (CPG), la résonance magnétique nucléaire ou la spectrométrie de masse (MS) ont permis de faciliter l'analyse structurale précise des polysaccharides. Mais ces méthodes nécessitent une pureté assez importante de l'échantillon et celui-ci est fréquemment contaminé avec des polysaccharides issus de la paroi cellulaire (Geno et al. 2015). De plus, les groupes labiles tels que les acétylations sont difficilement étudiés notamment à cause de leurs migrations à d'autres positions au niveau d'une même unité saccharidique mais également au sein du polysaccharide (Lassfolk et al. 2019).

D'autres méthodes sont basées sur les séquences des gènes impliqués dans la synthèse de la capsule. En effet, en 2006, l'Institut de Sanger a séquencé les *loci* des gènes permettant la synthèse de la capsule chez 90 sérotypes de *S. pneumoniae* permettant ainsi leur détermination par PCR (S. D. Bentley et al. 2006). De plus, grâce à des données de séquençage, l'agence de la santé Anglaise (Public Health England, PHE) a développé un outil informatique automatisé permettant d'identifier un sérotype à partir d'une séquence d'ADN (Kapatai et al. 2016). Ces méthodes peuvent également permettre d'identifier des souches difficilement identifiables à cause d'une régulation négative de leurs capsules, ou de l'absence de celle-ci. Le projet de grande ampleur "Global Pneumococcal Sequencing" qui vise à séquencer 20 000 souches de *S. pneumoniae* autour du globe devrait permettre d'en apprendre plus sur les différentes caractéristiques génétiques et leurs relations avec les différents sérotypes (S. Bentley, McGee, and Breiman 2016).

1.1.2 Autres composants de surface de *Streptococcus pneumoniae*

A la surface de *S. pneumoniae*, plusieurs protéines sont également présentes de façon plus ou moins accessible selon l'épaisseur de la capsule. Ces protéines vont généralement jouer un rôle important dans les étapes de l'infection et particulièrement au niveau de l'étape de colonisation.

Certaines protéines ainsi que leurs différents rôles sont plus particulièrement développés au niveau de la description des facteurs de virulence décrits section I.4.

En tant que bactérie immobile, *S. pneumoniae* ne possède pas de flagelle. Par contre, elle possède des pili : les pili communs et les pili sexuels (Henriques-Normark and Tuomanen 2013). Les pili communs sont des filaments protéiques de plusieurs micromètres de long, présents à intervalle régulier au niveau de la surface bactérienne. Ils facilitent l'adhésion bactérienne à des surfaces. En effet, des protéines ayant des rôles d'adhésines peuvent être présentes à leurs extrémités (Figure 32, section I.4.2.3). Quant aux pili sexuels, ceux-ci sont plus longs et vont avoir un rôle de transmission du matériel génétique entre différentes souches bactériennes.

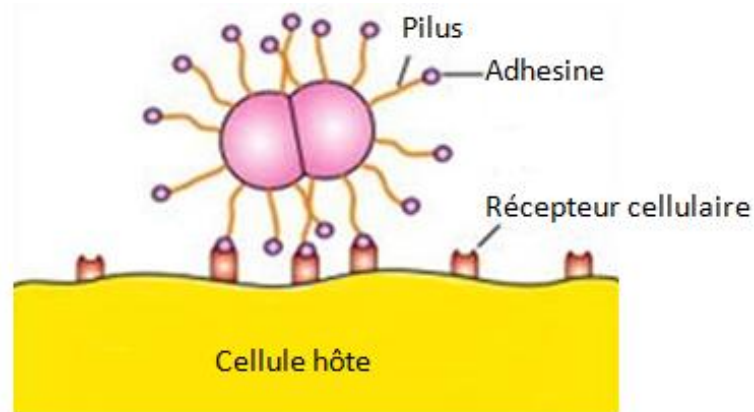


Figure 32: Schéma explicatif de la fonction d'un pilus

Un pilus est un filament protéique de plusieurs micromètres de long. Il peut être présent à intervalle régulier au niveau de la surface bactérienne ou au niveau de pôles. Des adhésines peuvent être présentes aux extrémités de ces pili et favorisent ainsi l'adhésion bactérienne.

I.2 Réservoir

S. pneumoniae a été isolée pour la première fois de manière simultanée par Pasteur et Sternberg en 1881 à partir de salive humaine ((Pasteur, Chamberland, and Roux 1881; Sternburg 1881)). En effet, *S. pneumoniae* est une bactérie commensale de l'Homme qui colonise les voies respiratoires supérieures (nasopharynx) (Bogaert, De Groot, and Hermans 2004). La flore bactérienne du nasopharynx se constitue dans les premiers mois de la vie de l'enfant et est composée d'une variété de souches dont *S. pneumoniae* (Donkor 2013). On la retrouve ainsi chez de nombreux jeunes enfants sans apparition de symptômes. Par exemple, au cours d'une étude réalisée en Papouasie-Nouvelle-Guinée, *S. pneumoniae* était retrouvée chez 60% des enfants en période néonatale et tous les enfants testés avaient été colonisés durant les 3 premiers mois de leur vie (Gratten et al. 1986). Ces taux sont cependant plus faibles au niveau des pays développés avec par exemple 12% des enfants colonisés après 3 mois en Suède (Aniansson et al. 1992). Le taux de portage est également généralement plus faible chez les adultes. Par exemple, comme nous pouvons voir sur la Figure 33, une étude épidémiologique réalisée par Mackenzie *et al.* sur une population possédant un fort taux de maladies pneumococcales a montré que le taux de portage est très fort chez les jeunes enfants de moins de 5 ans (80%), avec en général un taux de portage de 67,1% chez les enfants et adolescents de 2 à 15 ans contre 26% chez les adolescents et adultes de plus de 16 ans (Mackenzie et al. 2010). La diminution du taux au cours du temps pourrait s'expliquer par le développement d'un système immunitaire plus mature.

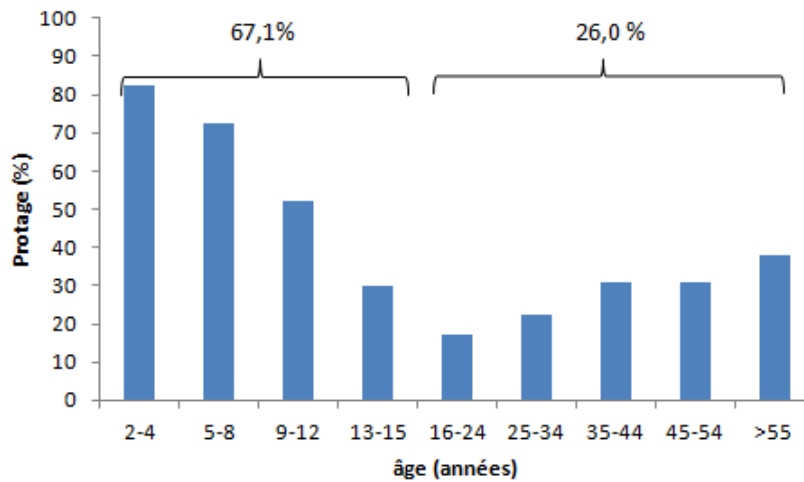


Figure 33: Pourcentage de porteurs de *Streptococcus pneumoniae* en fonction de l'âge chez une population à haut taux de maladies pneumococcales (Population aborigènes d'Australie).

Le portage de *S. pneumoniae* a été déterminé grâce à des analyses microbiologiques (morphologie, sérotypage) à partir de prélèvements naso-pharyngés. Cette étude a été réalisée sur 324 adultes (>16ans) et 218 enfants (<2-15ans). Au sein de cette population, les vaccins PCV7 et 23PPV y ont été introduit fin 2001 (section vaccins II.2). Les prélèvements ont été réalisés en 2002 et 2004. Image produite à partir des données issues de la publication (Mackenzie et al. 2010)

La colonisation du nasopharynx est une étape préalable à l'apparition des pathologies liées à *S. pneumoniae* (Bogaert, De Groot, and Hermans 2004) qui sont détaillées dans le paragraphe suivant (mécanisme détaillé section I.4).

I.3 Pathologies associées à *S. pneumoniae*

I.3.1 Population à risque

S. pneumoniae est responsable de pathologies graves telles que la pneumonie, la méningite ou la septicémie ou plus bénignes comme des otites ou des sinusites (sections suivantes I.3.2 et I.3.3). Tout comme le portage, la population la plus concernée est principalement celle des jeunes enfants de moins de 2 ans mais aussi les personnes âgées de plus de 65 ans (Figure 34).

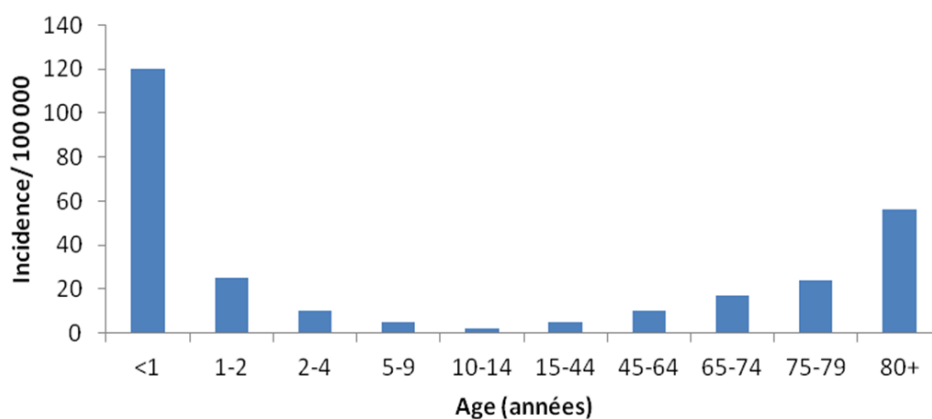


Figure 34: Incidence des maladies invasives liées à *Streptococcus pneumoniae* entre 2004 et 2005 en Angleterre et Ecosse image reproduite à partir de la publication (Ramasubramanian 2015)

De manière générale, un système immunitaire immature ou affaibli est un facteur de risque important. En dehors de l'âge, les autres facteurs de risques sont l'immunodépression congénitale ou acquise (agammaglobulinémie, myélome, hémopathie, VIH, déficit en complément), les maladies

pulmonaires chroniques, les infections virales (principalement la grippe), les terrains particuliers (insuffisance respiratoire, rénale, hépatique chronique, le diabète, la fistule cérébro-spinale, les implants cochléaires), la vie en grande promiscuité (prisons, camps militaires, foyers, hôpitaux de jour, crèche) ou encore une mauvaise hygiène de vie (alcoolisme, tabagisme) (Frottier and Ambroise-Thomas 2005; Delmont et al. 2012; Örtqvist, Hedlund, and Kalin 2005).

1.3.2 Infections non invasives liées au pneumocoque

Les pathologies dues à *S. pneumoniae* sont classées en deux catégories : les infections invasives et les infections non invasives. Lors des infections non invasives, *S. pneumoniae* ne pénètre pas dans des zones normalement stériles comme par exemple le sang ou le liquide céphalorachidien. Ces pathologies sont donc généralement moins graves et incluent les sinusites ou les otites mais également les pneumonies qui peuvent avoir de lourdes conséquences.

Sinusites :

La sinusite bactérienne est en général précédée par une sinusite virale d'une durée supérieure à 10 jours. Cette infection virale va alors faciliter l'infection bactérienne, notamment par un affaiblissement du système immunitaire ainsi que par des altérations de la couche cellulaire épithéliale au niveau des sinus, ce qui va favoriser l'adhérence bactérienne. Une seconde infection, bactérienne cette fois-ci, peut alors avoir lieu, le plus souvent par des bactéries aérobies facultatives telles que *S. pneumoniae* (Brook 2011). La principale complication d'une sinusite bactérienne est l'extension de l'infection à l'endocrâne et aux cavités orbitaires. Cependant, ces complications sont plutôt favorisées par l'immunodépression ou par des anomalies anatomiques (Régloix et al. 2017).

Otites :

Les otites moyennes (ou *otitis media*) sont une inflammation de la muqueuse qui tapisse la cavité du tympan. C'est en général une complication d'une infection virale des voies respiratoires supérieures (Nokso-Koivisto, Marom, and Chonmaitree 2015). Cette maladie est très commune chez les jeunes enfants. En effet, on estime que 60% des enfants ont déjà vécu au moins une otite à l'âge d'un an (Moore and Pilishvilli 2010). Cependant, les complications peuvent entraîner des infections plus graves telles qu'une mastoïdite (inflammation d'un os temporal derrière l'oreille qui peut entraîner une perte d'audition) ou des complications endocrâniennes graves comme une méningite.

Pneumonie :

La pneumonie est une infection des voies respiratoires au niveau du parenchyme qui entraîne des difficultés respiratoires dont la principale cause infectieuse est *S. pneumoniae*. Cette infection est souvent précédée d'une infection virale qui altère l'efficacité du système immunitaire contre des infections virales (Robinson, Kolls, and Alcorn 2015). L'infection bactérienne est alors issue d'une contamination par aérosol du nasopharynx vers les alvéoles pulmonaires où *S. pneumoniae* se multiplie alors dans le tissu épithélial entraînant une inflammation (Randle, Ninis, and Inwald 2011). Malgré son caractère non invasif, la pneumonie peut avoir des conséquences très graves. Chaque année, environ 400 000 hospitalisations seraient dues à une pneumonie aux Etats-Unis (Moore and Pilishvilli 2010). De plus, en 2017, l'agence mondiale de la santé a reporté 808 694 morts d'enfants de moins de 5 ans à cause d'une pneumonie au niveau mondial, soit 15% des décès (Who 2017).

1.3.3 Infections invasives liées au pneumocoque

Les infections invasives liées au pneumocoque telles que la méningite ou la septicémie, ont généralement des conséquences très graves allant jusqu'au décès. En effet, *S. pneumoniae* est l'une des principales causes de décès d'enfants de moins de 5 ans dans le monde (Who 2017).

Méningite :

La méningite est une infection des membranes (appelées méninges) entourant le cerveau ou la moelle épinière ce qui entraîne une inflammation. Les méningites virales sont bénignes et le patient se remet généralement sans traitement, contrairement aux méningites bactériennes qui peuvent avoir des conséquences très graves. Les principaux symptômes sont une nuque rigide, une fièvre importante, une sensibilité à la lumière, un mal de tête, de la confusion et des vomissements (WHO 2018). *S. pneumoniae* est responsable d'environ 50% de l'ensemble des méningites bactériennes (Auburtin and Timsit 2001). Chez les adultes, elle constitue la cause la plus fréquente de méningites bactériennes avec 70% des cas, et est associée avec un fort taux de mortalité avec 18 à 30% de mort de patients et des séquelles neurologiques pour la moitié des survivants avec des pertes de capacités auditives ou neurologiques (Engelen-Lee et al. 2016).

Septicémie :

La septicémie est le stade d'une infection au cours duquel les bactéries se propagent par voie sanguine dans l'organisme. Elle peut partir d'un foyer infectieux comme une pneumonie. Si elle n'est pas détectée et traitée rapidement, elle peut engendrer un choc septique avec d'importantes lésions d'organes et entraîner la mort du patient (WHO 2019; Gaieski et al. 2013; Yuste et al. 2005). Il a été estimé par l'organisation mondiale de la santé que la combinaison de la septicémie et des méningites cause plus de mort d'enfants de moins de 5 ans que la malaria (WHO 2018). Ainsi, la septicémie est la maladie pneumococcale avec le plus haut taux de mortalité, soit environ 20% (Backhaus et al. 2016).

1.4 Facteurs de virulence et mécanismes liés à la pathogénicité

Comme nous l'avons vu précédemment, *S. pneumoniae* est présent chez de nombreux individus sans pour autant déclarer de pathologie (section I). Les mécanismes de transmission et de virulence de cette bactérie sont décrits en détails dans la revue réalisée par Weiser et al. (Jeffrey N Weiser, Ferreira, and Paton 2018). Dans les sections suivantes, nous allons reprendre les principaux éléments permettant la transmission et mettre en lumière les principaux facteurs de virulence lors des étapes de colonisation et d'invasion qui seront également répertoriés au sein du Tableau 9, (présenté à la fin de cette section). Notons que malgré les nombreux mécanismes de virulence impliqués, cette bactérie conduit au développement de pathologies principalement dans le cas de systèmes immunitaires faibles ou affaiblis et donc semble être plutôt un pathogène opportuniste.

1.4.1 Mécanisme de transmission

La transmission de *S. pneumoniae* se réalise par son inhalation au sein de gouttelettes sous forme d'aérosols ou par des contacts rapprochés avec des sécrétions ou des surfaces contaminées par des individus porteurs (Jeffrey N Weiser, Ferreira, and Paton 2018). En effet, cette bactérie peut survivre et même croître au sein de sécrétions salivaires (Verhagen et al. 2014) mais aussi survivre à la dessiccation, et former des biofilms, ce qui lui permet de résister à des conditions environnementales hostiles tout en conservant ses capacités de colonisation (Marks, Reddinger, and Hakansson 2014; Walsh and Camilli 2011). Par ailleurs, au cours de l'année cette transmission est plus fréquente pendant les mois les plus secs et froids (automne et hiver) où une transmission sera également plus fréquemment associée à une infection virale (Hartzell, Oster, and Gaydos 2003; Numminen et al. 2015). En effet, l'infection par le virus de la grippe ainsi que l'inflammation sont des facteurs qui favorisent la transmission (Diavatopoulos et al. 2010; Short et al. 2012). De plus, l'expression d'une toxine inflammatoire de *S. pneumoniae*, la pneumolysine (Ply), a été montrée comme indispensable pour permettre cette transmission (Zafar et al. 2017). Une fois le nouvel hôte colonisé, la bactérie va alors pouvoir se développer localement ou de façon plus invasive générant ainsi les maladies pneumococcales décrites précédemment (Figure 35, section I.3).

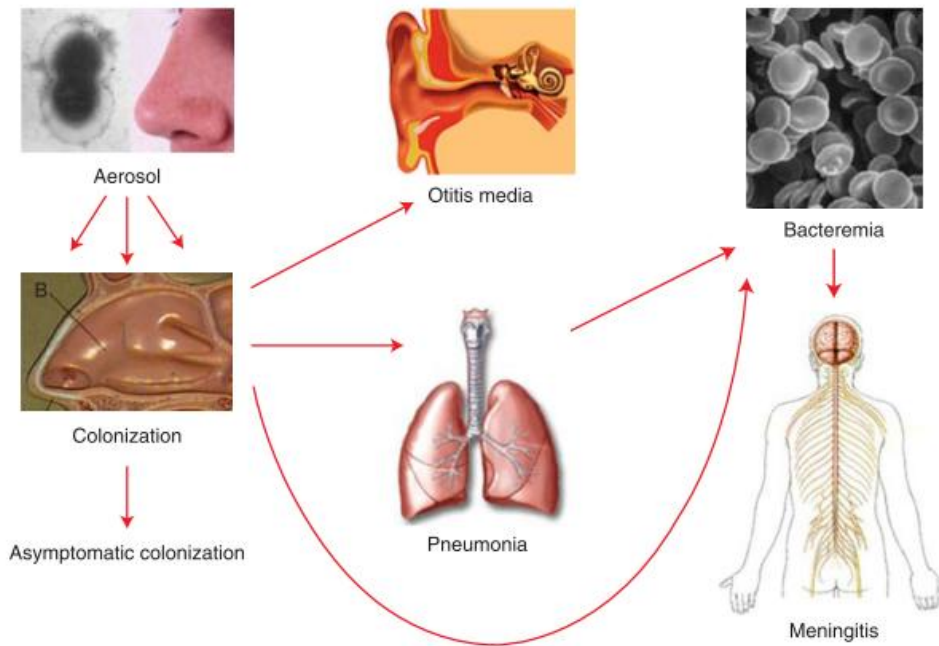


Figure 35: Transmission et propagation de *Streptococcus pneumoniae*.

La transmission de *S. pneumoniae* va s'effectuer à partir d'aérosols qui vont conduire à une colonisation du nasopharynx, qui peut rester asymptomatique, ou se propager vers d'autres organes dont notamment l'oreille, générant ainsi une otite, ou les poumons entraînant une pneumonie. L'infection du sang (septicémie) peut être issue d'une infection directe par le nasopharynx ou à partir d'une infection pulmonaire. Quant à l'infection des méninges du cerveau, elle s'effectue généralement à partir d'une infection du sang. Image extraite de la publication de (Henriques-Normark and Tuomanen 2013)

1.4.2 Etapes de colonisation et d'invasion : les différents facteurs de virulence

Nous allons détailler ici les principaux facteurs de virulence selon le stade d'infection (Tableau 9). Ces facteurs sont des composants exprimés par *S. pneumoniae* qui lui permettent de coloniser le nasopharynx mais également d'échapper ou d'inhiber le système immunitaire.

1.4.2.1 Capsule polysaccharidique

Comme nous l'avons vu précédemment (section 1.1.1), la capsule constitue la couche la plus externe de la bactérie et est de composition très variable selon les sérotypes. Il s'agit du principal facteur de virulence de *S. pneumoniae*. En effet, il a par exemple été montré que la digestion enzymatique de la capsule du sérotype 3 diminuait sa virulence d'un facteur de plus d'un million (Oswald T. Avery and Dubos 1931). Cette capsule va avoir un rôle important dès les premiers stades de l'infection.

En effet, un mucus tapisse la face externe du système respiratoire. Il est composé essentiellement d'eau, de polysaccharides et de protéines fortement glycosylées (les mucines). Il a pour rôle principal l'hydratation du système respiratoire mais il constitue également la première barrière lors de l'étape de colonisation (Bakshani et al. 2018). Ce mucus va empêcher les bactéries de se lier aux cellules épithéliales en les piégeant dans un réseau visqueux élastique non-Newtonien (avec notamment une viscosité variable en fonction des battements ciliaires) (Bakshani et al. 2018). Cependant, certains sérotypes de *S. pneumoniae* sont chargés négativement à pH physiologique (de par les groupements phosphates plus ou moins accessibles ou la présence de sucres sous la forme d'acide uronique). La capsule va alors repousser électrostatiquement le mucus ce qui va lui éviter d'être piégée dedans et donc évacuée. De plus, la présence de protéines de liaison aux mucines chez la bactérie va favoriser son maintien dans le nasopharynx (Jeffrey N Weiser, Ferreira, and Paton 2018).

Par ailleurs, la charge de la capsule va aussi lui permettre de repousser électrostatiquement les phagocytes également chargés négativement (Y. Li et al. 2013). Les phagocytes interviennent dans la réponse immunitaire innée. Ce sont des cellules du système immunitaire qui permettent d'éliminer des corps étrangers comme des bactéries. Il s'agit des macrophages, des cellules dendritiques et des neutrophiles. Le corps étranger est d'abord internalisé à l'aide de pseudopodes (prolongements cytoplasmiques) au sein d'une vacuole de phagocytose, puis il sera dégradé à l'aide d'enzymes de dégradation issues du lysosome. Les corps étrangers peuvent être reconnus selon deux mécanismes : i) directement par une interaction grâce aux récepteurs PRR (Pathern Recognition Recepteurs) des phagocytes qui reconnaissent les PAMP (Pathogènes Associated Molecular Patern) comme des lipopolysaccharides bactériens, ou ii) indirectement grâce aux opsonines (on parle alors d'opsonophagocytose), tels que des anticorps ou le facteur du complément C3b (Gregoriadis et al. 1992). Cependant, la phagocytose est également limitée par la présence de la capsule qui crée une barrière physique empêchant la fixation des anticorps spécifiques à la paroi bactérienne ainsi que la fixation des phagocytes au complément C3b qui se serait éventuellement fixé (Rieux 2002; Hyams, Camberlein, et al. 2010). De plus, la capsule forme également un écran protecteur contre les éventuels anticorps dirigés contre des protéines de surface ou contre des peptides antimicrobiens présents dans le nasopharynx (Rieux 2002; Habets, Rozen, and Brockhurst 2012).

Bien que le rôle de la charge semble important, certaines capsules ne sont pas chargées négativement. Par exemple les capsules des sérotypes 7A, 7F, 33F, 33A et 37 sont neutres et la capsule du sérotype 1 est zwitterionique (porte à la fois des charges négatives par des acides glycuroniques (GalpA) et des charges positives par les groupements amines du 2-acetamido-4-amino-2,4,6-trideoxygalactose (AATGalp)) (Tableau 8). Quant au sérotype 14, sa structure est moins soluble et formerait plutôt un gel plus imperméable, ce qui pourrait expliquer sa forte capacité d'invasion (Geno et al. 2015). La charge n'est donc probablement pas le seul mécanisme d'interaction entre la capsule et l'hôte (Geno et al. 2015).

Tableau 8: Exemple de composition de capsules non chargées négativement
AATGal, 2-acetamido-4-amino-2,4,6-trideoxy-d-galactose; Ac, acetate; Gal, galactose; GalA, galacturonic acid; GalN, Galactosamine; GalNAc, N-acetylgalactosamine; Glc, glucose; GlcA, glucuronic acid; GlcN, glucosamine; GlcNAc, N-acetylglucosamine; P, phosphate; Rha, rhamnose; Rib, ribose; Rib-ol, ribitol; f, furanose; p, pyranose.

Sérotype	Composition polysaccharidique
1	$\rightarrow 3$ - α -AATGalp-(1 \rightarrow 4)- α -D-GalpA _{2,0,3,3} Ac ₂ -(1 \rightarrow 3)- α -D-GalpA-(1 \rightarrow
7A	$\rightarrow 6$ - α -D-Galp-(1 \rightarrow 3)- β -L-Rhap2Ac-(1 \rightarrow 4)- β -D-Glcp-(1 \rightarrow 3)- β -D-GalpNAc-(1 \rightarrow 4 ↑ 1 α -D-GlcpNAc-(1 \rightarrow 2)- α -L-Rhap
7F	$\rightarrow 6$ - α -D-Galp-(1 \rightarrow 3)- β -L-Rhap2Ac-(1 \rightarrow 4)- β -D-Glcp-(1 \rightarrow 3)- β -D-GalpNAc-(1 \rightarrow 2 4 ↑↑ 1 1 β -D-Galp α -D-GlcpNAc-(1 \rightarrow 2)- α -L-Rhap

14	→6)-β-D-GlcpNAc-(1→3)-β-D-Galp-(1→4)-β-D-Glcp-(1→ 4 ↑ 1 β-D-Galp
33F	→3)-β-D-Galp-(1→3)-α-D-Galp-(1→3)-β-D-Galp-(1→3)-β-D-Glcp-(1→5)-β-D-Galp2Ac0.5-(1→ 2 ↑ 1 α-D-Galp
33A	→3)-β-D-Galp-(1→3)-α-D-Galp-(1→3)-β-D-Galp5,6Ac2-(1→3)-β-D-Glcp-(1→5)-β-D-Galp2Ac-(1→ 2 ↑ 1 α-D-Galp
37	→3)-β-D-Glcp-(1→ 2 ↑ 1 β-D-Glcp

Par ailleurs, la composition très variable de la capsule lui permet de limiter la sensibilité de *S. pneumoniae* face au système immunitaire. En effet, cette variabilité impose à l'organisme hôte de produire des anticorps spécifiques pour chaque sérotype mais impacte également l'efficacité du dépôt du complément (Rieux 2002; Hyams, Yuste, et al. 2010). De plus, des phénomènes de séroconversions ont été observés *in vivo*, c'est à dire le changement d'un sérotype en un autre au niveau d'une souche (Nesin, Ramirez, and Tomasz 1998). Cela pourrait s'expliquer par des mutations simples qui différencient certains sérotypes ou par la compétence à la transformation naturelle de *Streptococcus pneumoniae*, c'est à dire sa capacité à échanger des fragments d'ADN avec des bactéries voisines issues d'autres sérotypes ou même de bactéries phylogénétiquement voisines telles que *Streptococcus mitis* et *Streptococcus oralis* (Chaguz, Cornick, and Everett 2015).

1.4.2.2 Autres facteurs de virulence des premiers stades d'infection (mucus)

Au niveau du mucus, d'autres composants sont présents pour faciliter l'élimination de pathogènes tels que les anticorps (essentiellement des IgA1), le lysozyme ou la lactoferrine (Bakshani et al. 2018). Cependant, *S. pneumoniae* peut échapper aux IgA1 grâce à l'expression de ZmpA, une protéase spécifique qui induit leur dégradation (Janoff et al. 2014). L'expression d'IgG grâce à des stratégies vaccinales permet de contourner ce problème car les IgG ne sont pas sensibles à cette dégradation (section II.2.4). Par ailleurs, une déacétylase (PgdA) ainsi qu'une O-acétyltransférase (Adr), deux enzymes produites par *S. pneumoniae*, permettraient de réduire la sensibilité de la paroi bactérienne à la dégradation par le lysozyme en réduisant son degré d'acétylation au niveau des glycosamines et en augmentant celui de l'acide N-neuraminique (Davis et al. 2008). Quant à la lactoferrine, c'est une glycoprotéine qui se lie au fer, le rendant moins disponible pour les pathogènes. Cependant, les protéines PspA et PspC (aussi appelé CbpA) exprimées à la surface par *S. pneumoniae* peuvent se lier à la lactoferrine pour récupérer ce fer (Shaper et al. 2004). De plus, ces protéines ont également d'autres rôles favorisant la virulence. Ainsi, PspC agit en tant qu'adhésine pour se lier aux cellules hôtes par sa capacité de liaison à un récepteur d'immunoglobuline (R.

Anderson and Feldman 2011). De plus, PspA inhibe le dépôt du complément, limitant l'élimination de la bactérie par opsonophagocytose (Rieux 2002).

Par ailleurs, la neuraminidase (NanA) ainsi que d'autres enzymes de dégradation des glycanes (par exemple BgaA, StrH, EndoD, SpGH92, et NgtS) ont également un rôle important lors des premiers stades de colonisation (Robb et al. 2017). En effet, celles-ci vont à la fois permettre la progression de la bactérie au sein du mucus, grâce à la dégradation des mucines et à l'utilisation des glycanes disponibles comme source de carbone, mais également permettre la libération de sites de fixations des adhésines bactériennes sur la cellule hôte par dégradation des glycanes du glycocalyx, qui pourront également servir de source de carbone (Jeffrey N Weiser, Ferreira, and Paton 2018). De plus, les domaines de fixations aux sucres (CBM) des neuraminidases peuvent aussi agir à la manière d'adhésines afin de favoriser la fixation aux cellules de l'hôte (Jeffrey N Weiser, Ferreira, and Paton 2018).

1.4.2.3 Les adhésines (adhérence à la couche épithéliale)

Derrière cette première barrière composée du mucus se trouvent les cellules épithéliales qui forment la couche externe de l'appareil respiratoire de l'hôte. Cet épithélium est formé de cellules productrices du mucus ainsi que de cellules ciliées (Figure 36). Les cils vont effectuer des mouvements afin d'éliminer les différentes poussières ou corps à éliminer au sein du mucus. C'est également cet épithélium qui sécrète les molécules défensives tels que les IgA, le lysozyme, ainsi que les peptides antimicrobiens permettant de lutter contre des pathogènes. De plus, des jonctions serrées sont présentes entre les cellules épithéliales afin de rendre cette couche imperméable à l'entrée de pathogènes extérieurs (Coraux et al. 2005). L'endommagement de cette couche épithéliale par d'autres agents infectieux tels que le virus de la grippe peut ainsi faciliter la pénétration de *S. pneumoniae* dans les tissus (Wren et al. 2014).

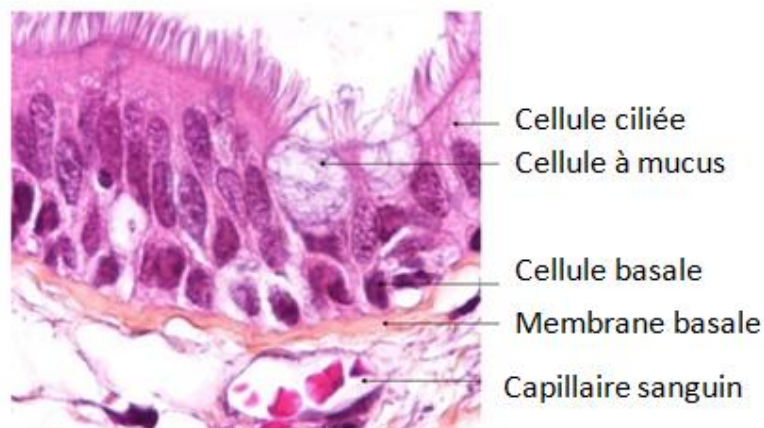


Figure 36: Histologie de l'épithélium respiratoire

Il est composé de cellules ciliées qui vont permettre l'évacuation de poussières ou d'éléments étrangers à éliminer, et de cellules productrices de mucus ainsi que d'autres éléments défensifs (IgA, lysozyme, peptides anti-microbiens). Les cellules sont maintenues entre elles grâce à des jonctions serrées. La membrane basale définit la limite de l'épithélium. Image extraite de la thèse de Leman Raphaël (Leman 2014).

La bactérie pathogène peut adhérer aux cellules hôtes lors d'une phase de l'infection appelée colonisation. Cette adhérence va se réaliser par le biais de protéines bactériennes membranaires : les adhésines. De nombreux facteurs vont alors être exprimés par la bactérie lors de l'adhésion. De plus, au cours de ce stade, la capsule sera moins exprimée et aura donc une épaisseur plus faible, ce qui devrait permettre d'augmenter l'accessibilité des adhésines (Figure 37). Ces adhésines vont alors interagir avec des cibles protéiques ou glycosydiques au niveau des cellules épithéliales. Ainsi, la PsaA

(Pneumococcal surface adhesin A) se lie aux cadhérines humaines, la PspC (Pneumococcal Surface protéine C) à un récepteur à la surface des cellules épithéliales (pIgR), la protéine pCho (phosphatidyl Choline binding) se lie au facteur PaFr. Certaines adhésines sont présentes à l'extrémité de pili (section 1.1.2), c'est le cas notamment des protéines du pili RrgA, RrgB, RrgC qui reconnaissent différentes structures saccharidiques du glycocalyx à la surface des cellules cibles. Ces protéines peuvent également avoir des fonctions additionnelles. Par exemple la PsaA est un transporteur de la famille ABC qui intervient dans l'absorption du manganèse. Les protéines RrgA, RrgB, RrgC sont également impliquées dans la formation de biofilm ce qui favorise la résistance de *S. pneumoniae* (Jeffrey N Weiser, Ferreira, and Paton 2018; Rieux 2002).

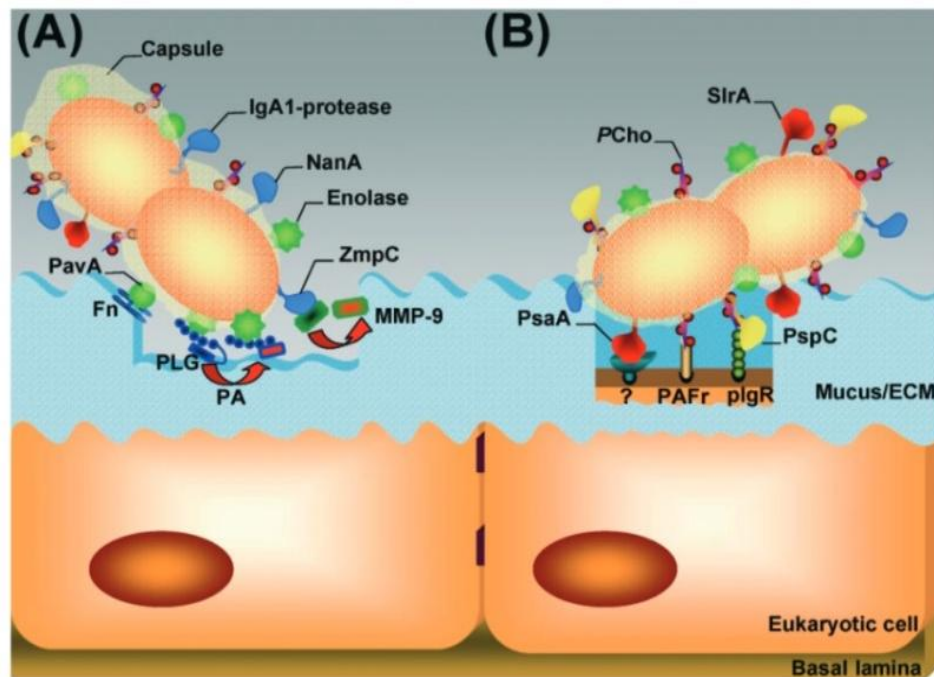


Figure 37: Colonisation de l'épithélium par *Streptococcus pneumoniae*

A) phase de passage au sein du mucus B) Phase de colonisation par adhérence cellulaire. Nous pouvons remarquer que la capsule polysaccharidique est moins abondante, ce qui permet une présentation des adhésines à la surface afin de favoriser l'adhérence à la cellule cible. Image extraite de la publication (Bergmann and Hammerschmidt 2006).

1.4.2.4 La pneumolysine et autres facteurs (passage de la couche épithéliale)

La pneumolysine est un facteur de virulence important de *S. pneumoniae*. C'est une protéine cytotoxique synthétisée par la bactérie qui est libérée lorsque celle-ci est détruite. Elle va alors interagir avec le cholestérol et la membrane plasmique et induire la formation de pores (Figure 38). Leur nombre trop important va alors entraîner la lyse de la cellule eucaryote (Hirst et al. 2004). De plus, elle peut également endommager l'épithélium par la séparation des jonctions serrées, rendant ainsi l'épithélium perméable au passage de bactéries entre les cellules (Hirst et al. 2004).

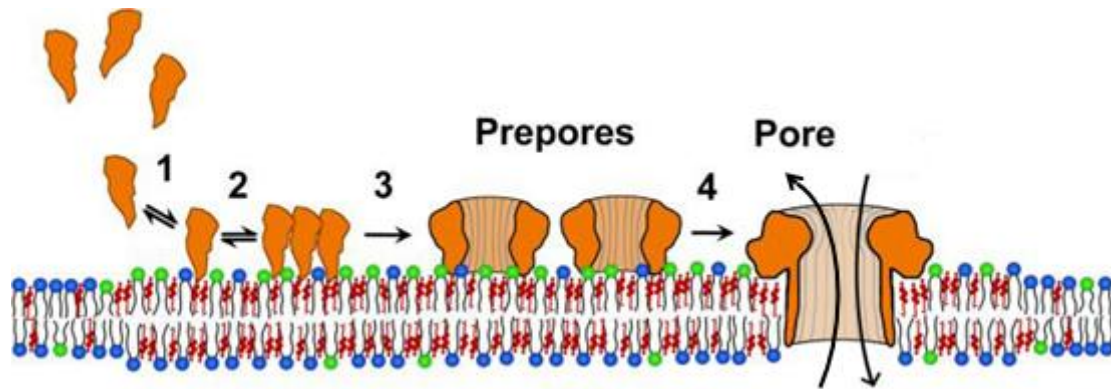


Figure 38: Formation d'un pore par la pneumolysine

Etape 1: la pneumolysine soluble se lie à des zones de la membrane riches en cholestérols, et SM, Etape 2: différentes unités de pneumolysine s'assemblent entre elles, Etape 3: Formation d'arc ou de cercles, Etape 4: Transition en un pore actif, Etape 5: Pore. Image adaptée à partir de la publication de (Drücker et al. 2018).

Cette protéine a également d'autres rôles dans la virulence de *S. pneumoniae*. Elle va notamment inhiber les mouvements des cellules ciliées de l'hôte qui permettent l'élimination de corps étrangers au niveau du mucus. Elle va aussi stimuler la production de cytokines inflammatoires, inhiber la prolifération des cellules lymphocytaires T, et diminuer l'activité bactéricide des neutrophiles. De plus, elle va activer le complément en lui permettant de se fixer au fragment constant (Fc) des IgG ce qui va donc réduire l'activité du complément au sein du sérum (Hirst et al. 2004).

Des dommages de la couche épithéliale peuvent également être générés par la production de peroxyde d'hydrogène (H_2O_2) par de protéines pneumococcales comme SpxB et GlpO (Jeffrey N Weiser, Ferreira, and Paton 2018). En effet, l' H_2O_2 va générer un stress oxydatif par la production d'espèces réactives de l'oxygène ce qui va fragiliser les cellules eucaryotes, voire causer la mort cellulaire. De plus, ces espèces réactives vont limiter la compétition de *S. pneumoniae* avec d'autres bactéries ne possédant pas de catalases qui permettent de dégrader l' H_2O_2 .

1.4.2.5 Autres facteurs de virulence aux stades infectieux avancés

Le passage de la couche épithéliale peut se réaliser par passage entre des cellules endommagées mais également par la liaison entre la protéine ChoP exprimée par la bactérie et deux récepteurs de l'hôte, le PAFR (Platelet Activating Factor Receptor) ou le PIGR (Polymeric Immunoglobulin Receptor), qui permettent à la bactérie d'utiliser le système d'endocytose de la cellule eucaryote pour se créer un passage à travers les différents tissus cellulaires (Jeffrey N Weiser, Ferreira, and Paton 2018).

Une fois que *S. pneumoniae* aura passé les barrières défensives de l'organisme, la bactérie doit s'adapter à un nouveau microenvironnement par l'expression de différents facteurs. Par exemple, le manque d'ions Manganèse ou Zinc impliquera la plus forte expression de protéines du groupe des transporteurs ABC comme PsaA ou PiuA. La production de la capsule est également fortement induite, ce qui permettrait de limiter l'élimination de la bactérie par le système immunitaire d'après les mêmes mécanismes que ceux présentés précédemment.

Les principaux facteurs de virulence que nous avons présentés ci-dessus sont répertoriés au sein du Tableau 9 suivant.

Tableau 9: Présentation des principaux facteurs de virulence

Facteur de virulence	Modes d'action
Capsule	-Limite l'empêchement au sein du mucus -Limite la phagocytose -Limite le dépôt du complément et donc l'opsonophagocytose -Barrière physique contre des anticorps ou des peptides antimicrobiens -Variabilité et possible changement de sérotype pour échapper à la réponse immunitaire
Protéase ZmpA	-dégrade les IgA1
Déacétylase PgdA	-déacétyle la paroi (au niveau des glucosamines) entraînant une baisse de la reconnaissance de la paroi par le lysozyme
O-acétyltransférase	-acétyle l'acide N-neuraminique entraînant une baisse de la reconnaissance de la paroi par le lysozyme
Protéine PspA	-liaison à la lactoferrine (qui piège le fer) -inhibition du dépôt du complément et donc de l'opsonophagocytose
Protéine PspC	-Liaison à la lactoferrine (qui piège le fer) -Adhésine : liaison à la cellule hôte (récepteur d'immunoglobuline)
Neuraminidases et autres glycosylhydrolases <i>(ex. NanA, BgaA, EndoD)</i>	-Dégradation des mucines (motifs glycosylés) et donc facilitation de la progression au sein du mucus -Dégradation du glycocalyx des cellules hôtes, favorisant ainsi la liaison des adhésines aux cellules épithéliales cibles -Fixation à l'aide de domaines lectiniques qui agissent comme des adhésines
Adhésines <i>(ex. PsaA, PspA, PspC, RrgA, RrgB, RrgC)</i>	-Fixation aux cellules hôtes par le biais de différentes cibles (protéiques, glycanes)
Pneumolysine	-formation de pore entraînant la lyse des cellules épithéliales -séparation des jonctions serrées et donc perméabilité de l'épithélium -inhibition mouvements ciliaires, donc baisse de l'élimination par le mucus -stimulation de production de cytokines -inhibition de la prolifération des lymphocytes T -baisse de l'activité des neutrophiles -activation du complément par sa liaison au Fc des IgG (donc baisse de son activité au sein du sérum)
SpxB et GlpO	-production de H ₂ O ₂ , entraînant des dommages cellulaires
ChoP	-utilisation du système d'endocytose pour pénétrer au sein des tissus

II. Moyens de luttés contre *Streptococcus pneumoniae*

II.1 Traitement à l'aide d'antibiotiques

Actuellement pour lutter contre une infection due à *S. pneumoniae*, le traitement consiste en l'utilisation d'antibiotiques. Il s'agit de molécules issues de champignons ou de plantes qui leur permettent de lutter contre les organismes environnants. Ils agissent en bloquant des systèmes biologiques essentiels à la survie ou au développement cellulaire comme : i) inhibition de la synthèse de la paroi bactérienne, ii) inhibition de la synthèse de la membrane cytoplasmique, iii) inhibition de la synthèse protéique, iv) inhibition de la synthèse de l'ADN (Figure 39). Les antibiotiques ont ainsi un

effet toxique qui entraîne l'arrêt de la prolifération cellulaire dans le cas d'antibiotiques dits "bactériostatiques" ou la mort cellulaire dans le cas d'antibiotiques dits "bactéricides" (Prasad Chakraborty, Pramanik, and Roy 2012).

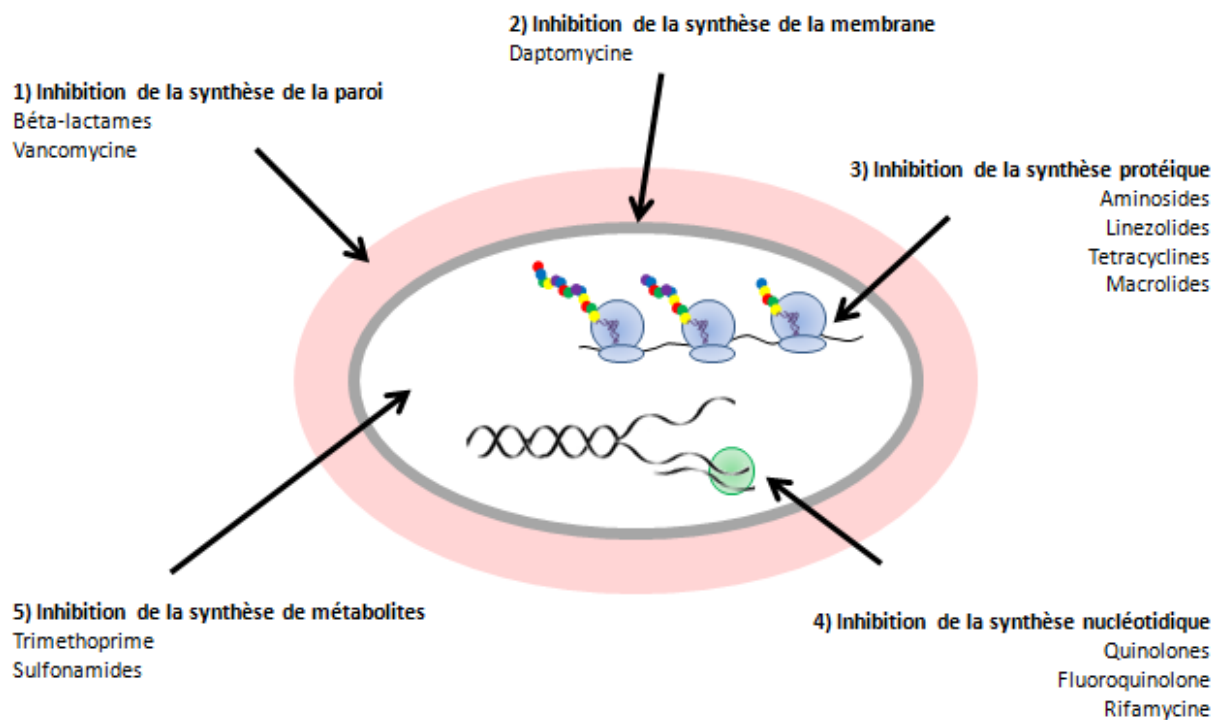


Figure 39: Schéma présentant les mécanismes d'action des principaux antibiotiques.

Les antibiotiques bloquent des mécanismes biologiques essentiels pour la survie ou la croissance de la bactérie. Ainsi, des antibiotiques bloquent la voie de synthèse de la paroi, de la membrane, de la synthèse protéique ou nucléotidique ou encore la synthèse de métabolites clés. Figure adaptée à partir de la publication (Prasad Chakraborty, Pramanik, and Roy 2012).

Cependant, *S. pneumoniae* développe de plus en plus de résistances aux antibiotiques. Ainsi la pénicilline a été initialement très utilisée en tant que traitement depuis les années 1940, mais une première souche résistante de *S. pneumoniae* a été isolée en 1967 (Tomasz 1997; Appelbaum 2002). Par la suite, des cas de multirésistance ont été répertoriés à partir de 1977 (R. R. Reinert et al. 2005). Une résistance aux antibiotiques est définie par une concentration minimale inhibitrice (CMI). Plus cette concentration est élevée, plus la souche est résistante. Ainsi, on différencie des souches ayant une résistance de niveau bas, de niveau intermédiaire, ou de niveau élevé. Plusieurs mécanismes sont impliqués dans la résistance de *S. pneumoniae* aux antibiotiques et sont développés dans le Tableau 10 ci-dessous (Cherazard et al. 2017; Varon and Gutmann 2004). Il s'agit le plus souvent de mutations génétiques altérant l'affinité de l'antibiotique à sa cible, ou à des modifications de son transport par augmentation de l'imperméabilité de la membrane ou par son transport actif vers l'extérieur. De plus, le développement de résistance chez *S. pneumoniae* est favorisé grâce à sa capacité à réaliser du transfert de gènes horizontal avec d'autres sérotypes ou avec d'autres souches commensales du microbiote du rhinopharynx telles que *Streptococcus mitis* et *Streptococcus oralis* (Chaguza, Cornick, and Everett 2015). De plus, cette capacité est amplifiée sous l'influence d'un stress tel que l'utilisation d'antibiotiques (Prudhomme et al. 2006; Hiller et al. 2010). Ainsi, en 2017 *S. pneumoniae* a été classée par l'agence mondiale de la santé comme un des 12 pathogènes contre lesquels il est nécessaire de développer rapidement des nouveaux antibiotiques afin de pouvoir lutter efficacement contre les infections causées par cette bactérie (WHO 2017).

Tableau 10: Mécanismes de résistance aux antibiotiques développés par *S. pneumoniae*.
Le pourcentage de résistance correspond à la prévalence aux Etats-Unis jusqu'en 2017 (Cherazard et al. 2017).

Antibiotique	Mécanisme d'action	% Résistance	Mécanisme de résistance
Béta-lactames (ex. <i>Pénicilline</i> , <i>Cefuroxime</i>)	Inhibition de la synthèse de la paroi bactérienne par interaction avec les Protéines de Liaison à la Pénicilline (PLP)	Pénicillines -PCN V: 41,8% -PCN G: 13,8% Cephalosporines <1% -29,9% Imipénème 23,3%	Mutations génétiques (par recombinaison ou par mutations ponctuelles) des cibles (PLP) qui entraînent une baisse de l'affinité de l'antibiotique
Macrolides (ex. <i>Erythromycine</i> , <i>Azithromycine</i>)	Inhibition de la synthèse protéique par fixation au ribosome (unité 50S)	20-40%	Modification génétique de la cible (gène <i>ermB</i>), modification du transport de l'antibiotique (par efflux), et modification de l'antibiotique
Lincosamides (ex. <i>Clindamycine</i>)	Inhibition de la synthèse protéique par fixation au ribosome (unité 50S)	21,8%	Modification génétique de la cible (gène <i>ermB</i>)
Fluoroquinolones (ex. <i>Cyprofloxacin</i>)	Inhibition de la réplication d'ADN par fixation aux extrémités de l'ADN empêchant leur assemblage	<1-2%	Mutagenèse de la cible (topo-isomérase IV ou gyrase), ou augmentation de l'efflux de l'antibiotique
Tétracyclines (ex: <i>Doxycycline</i>)	Inhibition de la synthèse protéique, empêchement de la fixation de l'ARNt amino acylé dans le site A du ribosome (étape d'élongation)	25,9%	Détachement de l'antibiotique lié au ribosome par des protéines dites "Protéines protectrices du ribosome", principalement Tet(M) et Tet(O)
Trimethoprim et Sulfamethoxazoles (TMP-SMX)	Inhibition de la synthèse nucléotidique par le blocage de la voie de synthèse des folates par effet synergique de TMP et SMX. SMX inhibe l'enzyme Dihydropteroate synthetase (DHPS) et TMP inhibe l'étape métabolique suivante réalisée par la dihydrofolate reductase (DHFR)	25-45%	Mutagenèse au niveau de la cible réduisant l'affinité de l'antibiotique. TMP : simple substitution au sein du gène DHFR. SMX : mutation dans le gène chromosomique de DHPS.

Tous les sérotypes de *S. pneumoniae* n'ont pas développé le même degré de résistance aux antibiotiques. Cependant, la distribution des différents sérotypes varie selon le portage et l'invasivité des souches. Cette différence dans les profils de résistance selon les sérotypes peut donc être due à différents facteurs. Par exemple, la colonisation est plus importante et plus longue chez les enfants qui sont également plus susceptibles d'être exposés à des antibiotiques et donc de développer des résistances. Ainsi, les sérotypes les plus courants chez les enfants et ayant le plus fort taux de portage sont ceux pour lesquels on retrouve le plus de résistances aux antibiotiques (Song et al. 2012). Il s'agit des sérotypes 6B, 9V, 14, 19F, 23F, 6A et 19A. Nous pouvons voir sur la Figure 40 un graphique présentant les pourcentages de résistance des souches selon l'antibiotique et selon le sérotype. Comme nous pouvons le voir sur cette figure, les sérotypes 14 et 19F sont ceux qui présentent le plus de résistance, avec 46% de souches multirésistantes pour le sérotype 14 et 17% pour le sérotype 19F (Stacevičienė et al. 2016). Cependant, ces données concernaient des enfants n'ayant pas été vaccinés et les stratégies vaccinales que nous allons voir dans les sections suivantes (II.2) ont eu tendance à faire baisser la circulation des sérotypes portant des résistances aux antibiotiques, et donc à modifier ces profils.

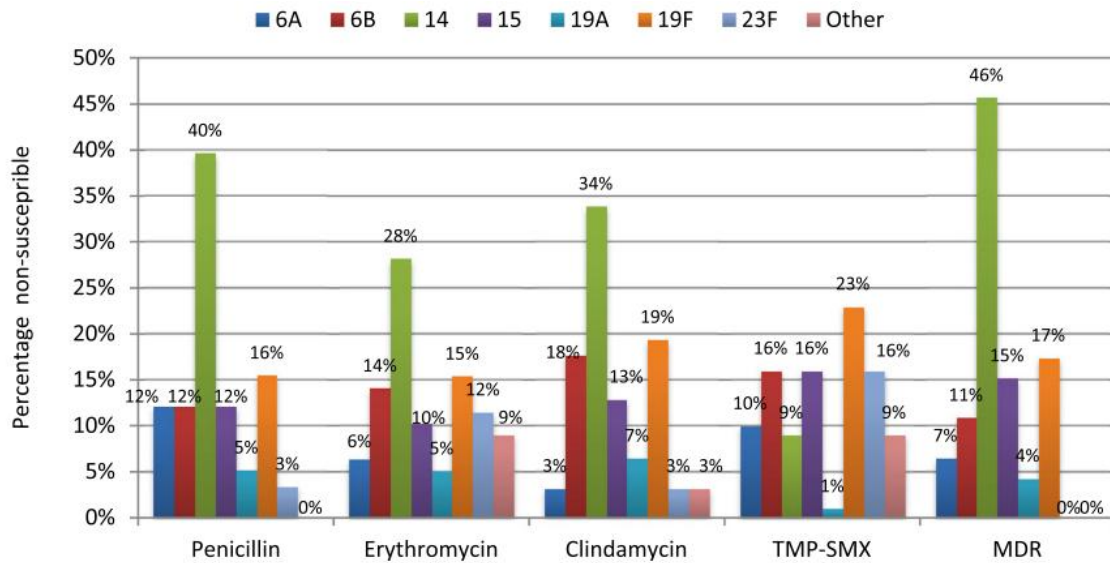


Figure 40: Profils de résistance et de multirésistance aux antibiotiques selon le sérotype
Ce graphique est extrait de la publication de (Stacevičienė et al. 2016) qui traite de données épidémiologiques sur des enfants (avant scolarisation et avant l'utilisation du vaccin chez les enfants) en Lituanie sur la période Fév-2012/Mars-2013.

Développer des stratégies alternatives aux antibiotiques contre *S. pneumoniae* est donc essentiel, c'est dans ce contexte de développement de résistance que les vaccins contre *S. pneumoniae* ont leur importance. Plusieurs types de vaccins ont été développés et sont détaillées dans les sections suivantes.

II.2 Prévention à l'aide de Vaccins

II.2.1 Rappels sur le fonctionnement des vaccins

Globalement, les vaccins consistent à exposer l'organisme à un antigène qui procèdera à une simulation d'infection, ce qui permettra à l'organisme de mieux combattre le pathogène lors d'expositions ultérieures. L'antigène peut être d'origines diverses (protéique, saccharidique, organismes entiers atténués) et sera reconnu par le système immunitaire qui en conservera alors une mémoire permettant une réaction secondaire plus rapide, plus importante et plus spécifique lors d'une infection future. La réponse sera donc spécifique de l'antigène contenu dans le vaccin. Il doit donc être choisi avec précaution pour permettre une élimination efficace du pathogène par la réponse secondaire.

Au niveau biologique, plusieurs cellules du système immunitaire vont être impliquées dans le processus de vaccination (Figure 41). Au niveau de l'antigène, des structures appelées épitopes vont être reconnues par des récepteurs présents au niveau du système immunitaire. Ces épitopes peuvent être linéaires ou structuraux. Dans le cas d'épitopes structuraux, leur conformation spatiale sera importante pour la stimulation du système immunitaire. Deux types d'épitopes vont alors avoir un rôle très spécifique : les épitopes B et les épitopes T. Les épitopes B sont des épitopes généralement structuraux et exposés au solvant, qui vont être reconnus par les lymphocytes B et contre lesquels des anticorps seront *in fine* générés. De plus, la reconnaissance de ces épitopes antigéniques par des cellules présentatrices d'antigène (CPA) (macrophage, cellule dendritique ou lymphocyte B) va entraîner l'internalisation de l'antigène. Des fragments de cet antigène seront présentés à des lymphocytes T auxiliaires au niveau d'un récepteur membranaire (le complexe majeur d'histocompatibilité de type II, CMHII). Ces fragments présentés au niveau du CMHII sont appelés épitopes T. Différents outils de prédictions ont été développés pour les identifier (Sanchez-Trincado, Gomez-Perosanz, and Reche 2017). Ce sont des épitopes généralement linéaires, de 9 à 22 acides

aminés de long, avec un cœur de 9 acides aminés dont des acides aminés clés, appelés ancras, qui se lient au niveau du CMHII. Les outils prédictifs doivent également prendre en compte la variabilité du CMHII d'une personne à l'autre selon différents types HLA (*Human Leukocytes Antigène*) (Sanchez-Trincado, Gomez-Perosanz, and Reche 2017). Une fois l'épitope T reconnu par un lymphocyte T auxiliaire, celui-ci a un rôle d'amplification de la réponse immunitaire. En effet, une fois stimulés par une CPA, ils vont entraîner la différenciation des lymphocytes B en plasmocytes afin que ces derniers produisent des anticorps pour neutraliser l'antigène. De plus, les CPA vont également amplifier la réponse immunitaire par la libération de cytokines et instaurer une réponse mémoire par la génération de cellules mémoires de durée de vie plus longue (généralement 5 à 10 ans pour les lymphocytes B ou T mémoires).

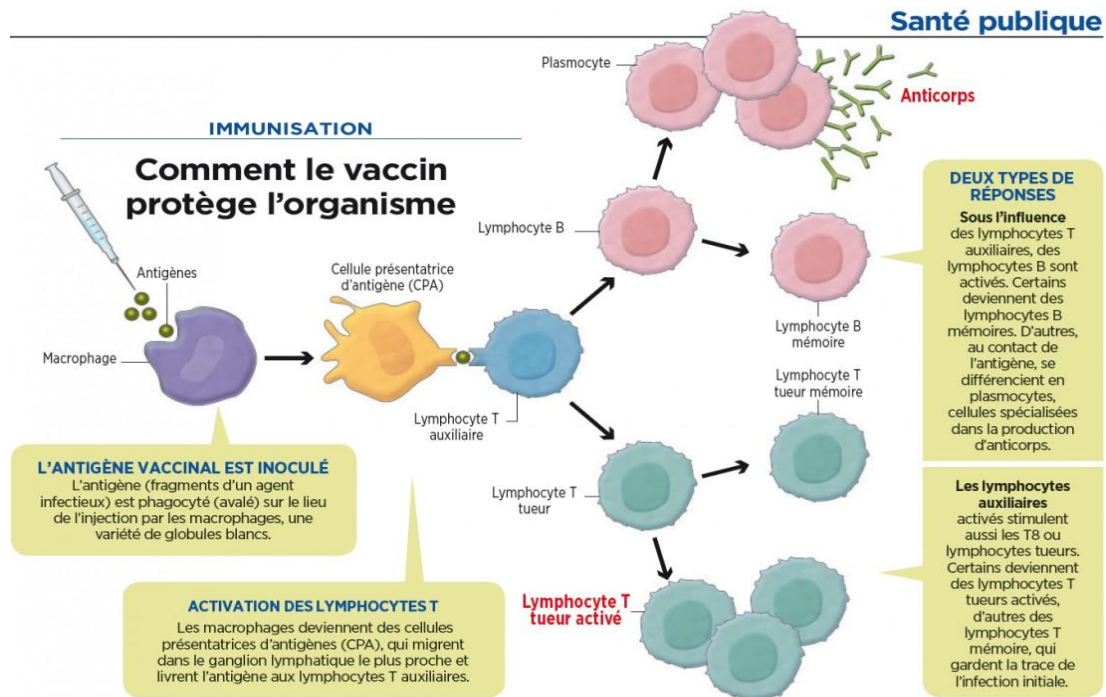


Figure 41: Points essentiels du fonctionnement biologique d'un vaccin

Un antigène est reconnu par une cellule présentatrice d'antigène, qui va l'intégrer et en présenter une partie (épitope T) à un lymphocyte T auxiliaire. Celui-ci va déclencher la différenciation des lymphocytes B en plasmocytes, qui vont produire des anticorps, mais aussi amplifier la réponse par la libération de cytokines et induire l'installation d'une réponse mémoire par la différenciation des lymphocytes B en lymphocytes B mémoire. Image extraite de la Revue «Sciences et Avenir» n° 826, parue en décembre 2015, page 37 (Lafond 2015).

Lors de la différenciation des lymphocytes B en lymphocytes B mémoires, des phénomènes d'épissage alternatif vont conduire à un changement isotopique des anticorps produits. Le lymphocyte va alors produire des IgG à la place d'IgM. Les IgM sont caractéristiques d'une réponse primaire, ils sont moins spécifiques mais ont une bonne liaison à la cible reconnue par un effet d'avidité grâce à leur pentavalence. Les IgG sont donc produits lors d'une réponse secondaire et seront plus spécifiques car le lymphocyte stimulé aura subi des hypermutations somatiques (essentiellement des substitutions) suivies de sélections sur la cible qui permettent ainsi d'augmenter l'affinité contre l'antigène.

Par ailleurs, la réponse immunitaire au niveau des lymphocytes cytotoxiques, également appelés "lymphocytes tueurs", permet d'éliminer les cellules infectées ou modifiées. Cependant, cette immunité ne sera pas abordée ici, car elle n'intervient pas dans le cadre de la réponse à *S. pneumoniae* ou des vaccins contre cette bactérie.

Plusieurs caractéristiques des vaccins sont importantes, notamment la stabilité globale, la nature et la dose de l'antigène, la voie d'introduction du vaccin (sous-cutanée, intra-nasale, intradermique), la présence d'adjuvants et leur nature, ainsi que le programme vaccinal (nombre d'injection, espacement entre les injections). Ici nous allons uniquement traiter des caractéristiques liées à l'antigène.

II.2.2 Bref historique des vaccins pour *Streptococcus pneumoniae*

Après la découverte de *S. pneumoniae* en 1880 par Pasteur et Stenberg, différentes études ont mené en 1911 au développement du premier vaccin anti-pneumococcique basé sur des bactéries de pneumocoque entières tuées. Par la suite, la découverte du premier antibiotique par Fleming en 1928 et le développement de cette thérapie a ralenti l'intérêt des recherches sur ces vaccins. Cependant, avec l'apparition grandissante de souches résistantes, l'importance des vaccins a été reconsidérée. En 1916, Avery et Dochez avaient isolé pour la première fois des polysaccharides capsulaires de *S. pneumoniae*. Puis, dans les années 1930, les capsules saccharidiques ont été identifiées comme un facteur de virulence important et leur spécificité selon la souche a été également mise en lumière par l'absence de réactivité croisée de sérums selon le sérotype ciblé. La capsule ayant un rôle critique pour la réponse immunitaire, diverses recherches ont *in fine* mené en 1977 au développement du premier vaccin composé de polysaccharides capsulaires, qui était alors composé de polysaccharides issus de 14 sérotypes différents. Ce nombre a par la suite pu être augmenté à 17 en 1980, et enfin 23 en 1983. Ce vaccin polysaccharidique de valence 23 est encore utilisé de nos jours. Cependant, à cause de ses limitations d'application, des vaccins composés de polysaccharides capsulaires conjugués à des protéines ont été développés, avec tout d'abord un vaccin conjugué contenant 7 sérotypes mis sur le marché en 2000. Par la suite, la valence de ces vaccins conjugués a également été augmentée avec un vaccin contenant 10 sérotypes sorti en 2009 et un autre contenant 13 sérotypes sorti en 2010 (Frise chronologique Figure 42) (Butler, Shapiro, and Carlone 1999; Grabenstein and Klugman 2012).

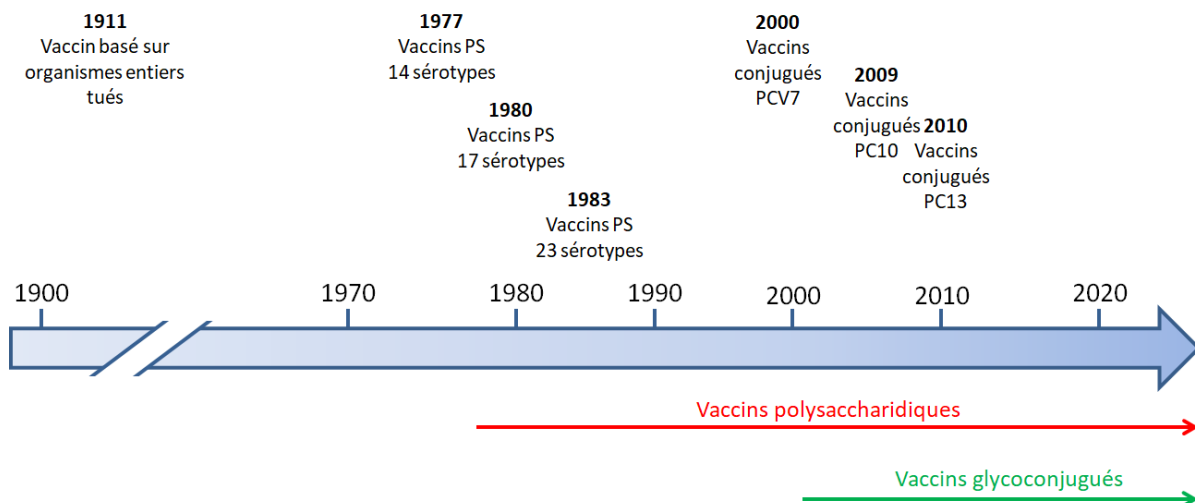


Figure 42: Frise chronologique présentant les différents types de vaccins développés contre *Streptococcus pneumoniae*

Dans les sections suivantes, nous allons essentiellement nous focaliser sur les vaccins encore actuellement utilisés et recommandés en France : le vaccin polysaccharidique de valence 23 (Pneumovax®) et le vaccin glycoconjugué de valence 13 (PCV13, Prevenar13®).

II.2.3 Les vaccins polysaccharidiques : Pneumovax®

II.2.1.1 Composition

Le vaccin Pneumovax® est commercialisé par le laboratoire MSD Vaccins et est composé de 25µg de différents polysaccharides capsulaires issus de 23 sérotypes de *S. pneumoniae* (25µg/pour chaque sérotype). Il s'agit des sérotypes 1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19F, 19A, 20, 22F, 23F et 33F. L'administration de ce vaccin s'effectue par injection sous-cutanée ou intramusculaire (HCSP 2018).

II.2.1.2 Méthodes pour produire les polysaccharides capsulaires

Comme nous l'avons vu précédemment (section I.1.1), la capsule est un polysaccharide de haut poids moléculaire, constituée d'unités saccharidiques répétées. *In vivo*, la capsule est rapidement et fortement produite durant la phase exponentielle de croissance de la bactérie. Les gènes de production des capsules *cps* sont organisés en cassette (Geno et al. 2015). Deux mécanismes de synthèse de la capsule ont été identifiés chez *S. pneumoniae* : i) la voie dépendante à Wzy et ii) la voie dépendante à une synthase (Aanensen et al. 2007; Geno et al. 2015). La voie dépendante à Wzy est la plus représentée parmi les différents sérotypes, et la voie dépendante à une synthase est présente chez les sérotypes 3 et 37. De manière générale, ces mécanismes consistent en la production des unités répétées d'oligosaccharides reliées à un phospholipide sur la face intracellulaire puis à leur exportation sur la face externe de la membrane et à leur transfert directement au niveau de la paroi ou par l'intermédiaire de protéine ou de lipides. La capsule y est reliée par une liaison entre l'extrémité réductrice du glucose de la capsule et les résidus N-acétylglucosamine pour la paroi (Larson and Yother 2017). Le mécanisme dépendant de la synthase se réalise à l'aide d'une seule enzyme tandis que le mécanisme dépendant de Wzy est réalisé à l'aide d'une série d'enzymes (Figure 43 et Figure 44).

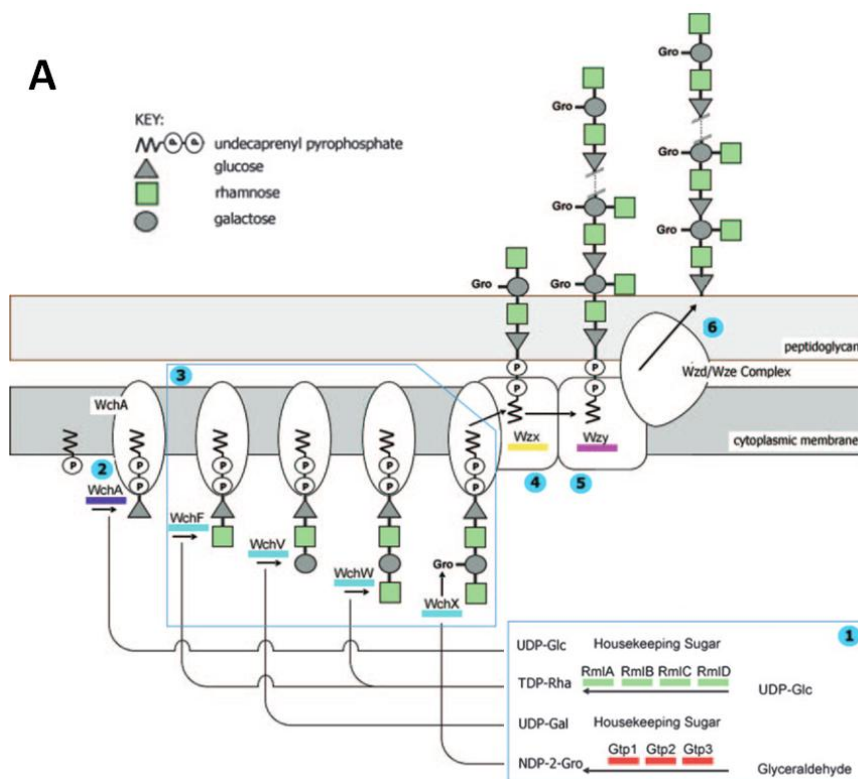


Figure 43: Mécanismes de biosynthèse de la capsule.

A) Mécanisme Wzy dépendant : (1-2) la biosynthèse commence par le transfert de l'UDP-glucose sur un lipide-phosphate de la membrane interne, (3) les autres unités saccharidiques sont ajoutées séquentiellement par des glycosyltransférases, (4) le complexe est ensuite transféré de l'autre côté de la membrane par une flippase Wzx puis (5-6) le complexe est transféré sur la paroi formée de peptidoglycane. (image extraite de la publication (Aanensen et al. 2007))

B

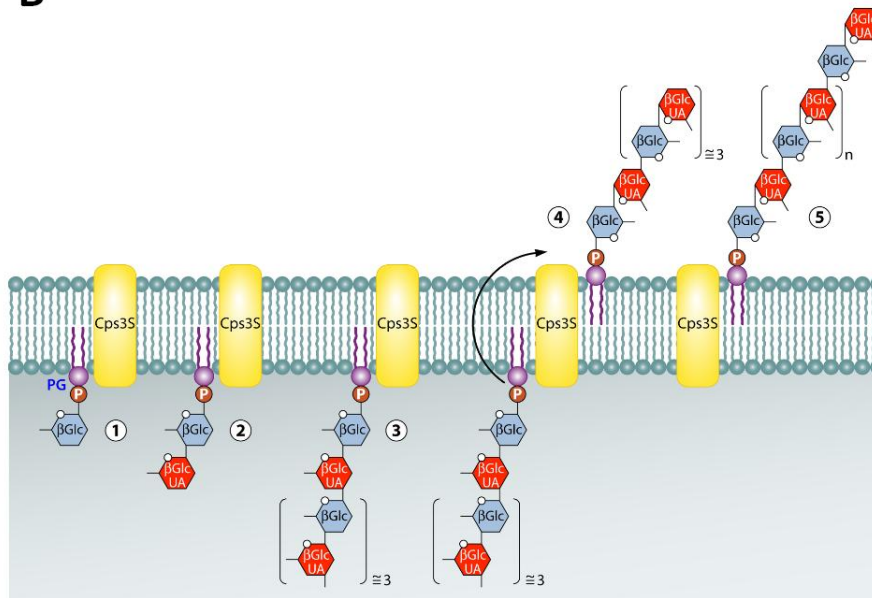


Figure 44: Mécanismes de biosynthèse de la capsule.

B) mécanisme synthase dépendant : (1) la synthase Cps3S commence la synthèse par le transfert de glucose à partir d'UDP-glucose sur un lipide phosphate de la membrane interne, (2-3) les autres unités sacchariques dont des acides glucuroniques sont ajoutés séquentiellement à partir d'UDP-acide-glucuronique, jusqu'à la formation de l'octosaccharide, (4) celui-ci est ensuite transféré par la Cps3S sur la face externe de la membrane où la longueur du polysaccharide est encore augmentée (image extraite de la publication (Geno et al. 2015)).

Différentes méthodes de purification de ces capsules ont été développées en se basant sur des précipitations fractionnées, des traitements par des protéases et nucléases, différentes techniques chromatographiques (ionique, exclusion stérique, affinité), ainsi que des techniques d'ultrafiltration (Morais, Dee, and Suárez 2018). Les contaminants principaux sont des protéines, des nucléotides, et des sucres composant la paroi. Plusieurs brevets ont été déposés par différentes compagnies dont les entreprises Merck et l'Institut Mérieux afin de protéger les méthodes utilisées en production (Morais, Dee, and Suárez 2018). Les polysaccharides capsulaires ayant été purifiés à partir de cultures bactériennes ne sont cependant pas définis, et des variations peuvent être observées entre les différents lots de production (Morais, Dee, and Suárez 2018).

La synthèse chimique de polysaccharides capsulaires permet d'obtenir des composés plus purs. Elle consiste essentiellement en la formation de liaisons glycosidiques stéréosélectives et l'utilisation de groupements protecteurs. Cependant, ces méthodes restent difficiles et toutes les structures ne sont pas facilement accessibles (Geno et al. 2015). Plusieurs oligosaccharides de sérotypes différents ont néanmoins été produits avec succès, ils ont été répertoriés dans la revue réalisée par Colombo *et al.* et quelques exemples sont présentés dans le Tableau 11 (Colombo, Pitirollo, and Lay 2018). De plus, le développement d'une méthode pour assembler automatiquement des sucres entre eux par l'équipe de Seeberger (Automated glycan assembly, AGA) permettrait de faciliter cette synthèse jusqu'alors complexe (Guberman and Seeberger 2019). Cette méthode consiste en la réalisation d'une synthèse d'oligosaccharides sur un support solide de manière analogue à la synthèse peptidique (Figure 45). La première unité saccharidique est reliée à un support par un lien, et les assemblages suivants sont réalisés à partir de sucres ayant des fonctions protégées pour éviter les réactions non spécifiques en termes de régiosélectivité et stéréosélectivité. Ils sont ensuite déprotégés sur le support et l'opération est répétée jusqu'à l'obtention de l'oligosaccharide ciblé. L'oligosaccharide est ensuite récupéré par clivage de son lien

avec la résine. Cette méthode a d'ailleurs déjà été utilisée pour la synthèse de polysaccharides capsulaires à destination vaccinale contre *S. pneumoniae*, notamment par la synthèse de quatre tétrasaccharides différents du sérotype 8 et de trisaccharides du sérotype 3 (Weishaupt et al. 2016; Schumann et al. 2017). Cette méthode n'est cependant pas encore mise au point pour la production à très grande échelle lors de procédés de production commerciaux.

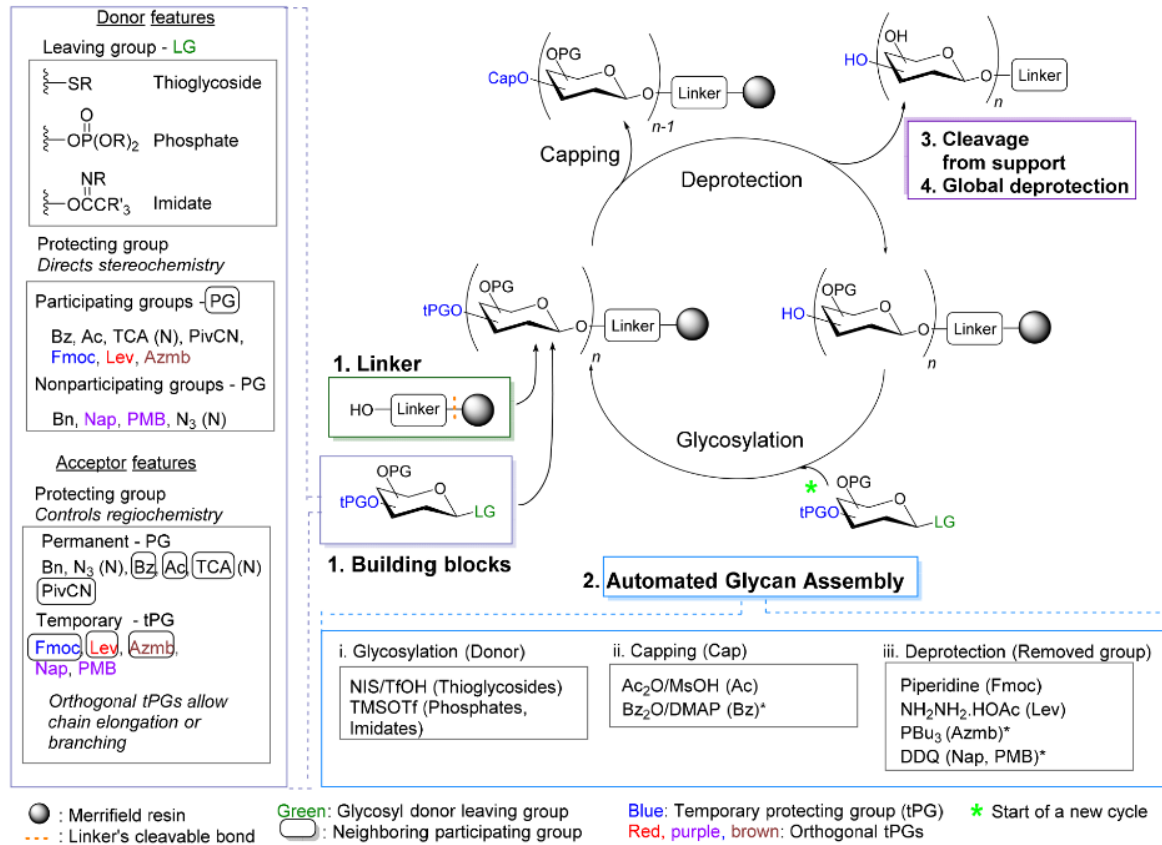


Figure 45: Méthode de synthèse automatique d'oligosaccharides

La première unité saccharidique est reliée à un support par un lien, et les assemblages suivants sont réalisés à partir de sucres ayant des fonctions protégées pour éviter les réactions non spécifiques en termes de régiosélectivité et stéréosélectivité. Ils sont ensuite déprotégés sur le support et l'opération est répétée jusqu'à l'obtention de l'oligosaccharide ciblé. Image extraite de la publication (Guberman and Seeberger 2019)

La production de ces oligosaccharides doit être représentative d'épitopes qui seront reconnus par le système immunitaire. Le choix de ces unités représentatives est donc crucial pour leur efficacité biologique. La composition des unités répétées et leurs modifications (purvate, O-acétylation, phosphate) sont caractéristiques de chaque sérotype. Mais ce ne sont cependant pas les seuls éléments impliqués dans l'épitope reconnu par le système immunitaire. En effet, la capacité immunogène d'une seule unité saccharidique n'est donc généralement pas la même qu'une molécule constituée de plusieurs unités répétées (Khatun, Stephenson, and Toth 2017). Ainsi, plusieurs études se sont ainsi intéressées à la longueur du polysaccharide nécessaire pour conserver la structure tridimensionnelle de l'épitope et donc ses caractéristiques immunogènes. Cette longueur optimale est ainsi très variable d'un sérotype à l'autre. Par exemple, pour le sérotype 14, un tétrasaccharide (<1kDa) serait suffisant pour générer des anticorps permettant l'opsonophagocytose alors que pour le sérotype 6B un polysaccharide de minimum 10kDa est nécessaire (Khatun, Stephenson, and Toth 2017). L'utilisation de polysaccharides synthétiques permet d'étudier ces caractéristiques mais apporte également une meilleure stabilité dans la composition des préparations antigéniques.

De nombreuses structures sont donc généralement synthétisées pour un même sérotype et testées afin de déterminer celle qui génèrera la meilleure réponse immunitaire (Colombo, Pitirollo, and Lay 2018). Les différents oligosaccharides sont ensuite testés lors de vaccination sur des modèles murins, puis les titres d'anticorps efficaces contre le polysaccharide natif sont enfin évalués par ELISA, SPR ou STD RMN afin de les classer selon leur force d'immunogénicité.

Par ailleurs, lorsque l'oligosaccharide est instable en solution ou difficilement synthétisable, il est possible d'avoir recours à des mimes structuraux (Ronchi et al. 2013). Ces analogues structuraux peuvent également être utilisés pour déterminer les caractéristiques de l'oligosaccharide permettant de générer une réponse immunitaire. Par exemple, les polysaccharides zwitterioniques (portant à la fois une charge positive et négative) ont des capacités immunogéniques particulières de par leur capacité à activer directement les lymphocytes T après leur présentation au niveau du CMHII (Q. Zhang et al. 2017). Ainsi, afin de mieux comprendre les caractéristiques de la position de ces charges, des analogues du tétrasaccharide issu du sérotype 14 ont été produits (Gragnani et al. 2019). Cependant, les différents analogues produits n'ont pas montré de différences significatives (Gragnani et al. 2019).

Tableau 11: Exemples d'oligosaccharides représentatifs de différents sérotypes synthétisés

Sérotype	Partie synthétisée	Référence
3	Hexasaccharide β -D-GlcpA-(1→4)- β -D-Glcp-(1→3)- β -D-GlcpA-(1→4)- β -D-Glcp-(1→3)- β -D-GlcpA-(1→4)- β -D-Glcp-(1→O-(CH ₂) ₃ NH ₂)	(Lefeber et al. 2002)
9 (9A, 9L, 9V, 9N)	Trisaccharides α -D-Galp-(1→3)- β -D-ManpNAc-(1→4)- β -D-Glcp et α -D-Glcp-(1→3)- β -D-ManpNAc-(1→4)- β -D-Glcp	(Alpe and Oscarson 2002)
4	Trisaccharide β -D-ManpNAc-(1→3)- α -L-FucpNAc-(1→3)- α -D-GalpNAc β -D-ManpNAc-(1→3)- α -L-FucpNAc-(1→3)- α -D-GalpNAc- Pyruvate	(Pereira et al., 2015)
23F	Tétrasaccharide [glycerol-(2-P→3)][α -L-Rhap-(1→2)]- β -D-Galp-(1→4)- β -L-Rhap-(1→4)- β -D-Glcp-3-aminopropyl	(K. Yu et al. 2016)
12F	Hexasaccharide β -D-GalpNAc-(1→4)-[α -D-Glcp-(1→3)]- β -D-ManpNAcA	(Seeberger et al., 2017)
19A	Trisaccharide β -D-ManpNAc-(1→4)- α -D-Glcp(1→3)-Rha	(Morelli et al. 2018)
14	Tétrasaccharide β -D-Galp-(1→4)- β -D-Glcp-(1→6)-[β -D-Galp-(1→4)]- β -D-GlcpNAc	(Safari et al. 2008)

II.2.1.3 Mécanisme d'action des vaccins polysaccharidiques

Ces vaccins sont uniquement composés d'antigènes polysaccharidiques qui sont des antigènes T-indépendants. La structure formée d'unités répétées les polysaccharides va permettre une activation importante des lymphocytes B grâce à un effet d'avidité qui va entraîner une cascade de signaux dont l'activation de la Bruton's tyrosine kinase (Btk) et donc la différenciation des lymphocytes B en plasmocytes producteurs d'anticorps (Vinuesa, De Lucas, and Cook 2001).

Par ailleurs, de part leur nature T-indépendante, les polysaccharides ne seront pas présentés au niveau du complexe d'histocompatibilité de type II (CMHII, section II.2.1) par les CPA. La réponse immunitaire ne fera donc pas intervenir les lymphocytes T auxiliaires et il n'y aura pas de production de cellules mémoires classiques, ni de maturation de l'affinité par hypermutations somatiques (section II.2.1). Des cellules immunitaires spécifiques contre des polysaccharides capsulaires ayant une longue durée de vie ont pourtant pu être observées dans la moelle osseuse à la suite d'une immunisation par des polysaccharides capsulaires (Taillardet et al. 2009). Ainsi, contrairement à ce qui est généralement admis, il s'agit d'un mécanisme attribué à une réponse immunitaire T dépendante (Taillardet et al. 2009). Les polysaccharides capsulaires pourraient donc induire une réponse non conventionnelle dont le mécanisme n'est pas clairement identifié.

De manière générale, la protection générée par un vaccin polysaccharidique n'est pas aussi forte que pour un vaccin classique et baisse très fortement après 5 à 7 ans avant d'atteindre un niveau nul après 10 ans (Musher et al. 2010; Hammitt et al. 2011). Actuellement, les antigènes des polysaccharides capsulaires sont généralement considérés comme ne générant pas de mémoire immunogène, et la réponse immunitaire est principalement composée d'IgM. Ces anticorps sont spécifiques pour la couche la plus externe de la bactérie et ils sont très efficaces pour son élimination en favorisant l'opsonophagocytose.

II.2.1.4 Efficacité et limites des vaccins polysaccharidiques

Les vaccins polysaccharidiques ont pu démontrer leur efficacité dès le premier vaccin polysaccharidique en 1977 lors de tests réalisés chez des travailleurs dans des mines d'or en Afrique (Westerink, Schroeder, and Nahm 2012). Il n'a pas d'effet sur le portage mais il est cependant efficace contre les maladies invasives, et son taux d'efficacité sur la population globale a été évalué à 57% aux Etats-Unis, avec des taux supérieurs pour certaines populations à risque (par exemple 75% pour les personnes âgées immunocompétentes de plus de 65 ans, 84% pour les personnes ayant un diabète de type II, 73% pour les personnes ayant des maladies cardiovasculaires) (Butler et al. 1993; Boelsen et al. 2015). En France, il est d'ailleurs recommandé pour les personnes à risque et âgées de plus de 2 ans. En effet, il n'est pas efficace chez les jeunes enfants (moins de 2 ans), qui constituent malheureusement une part majeure de la population affectée par le pneumocoque (Westerink, Schroeder, and Nahm 2012). Il a été supposé que ce manque de réponse chez les jeunes enfants envers les antigènes T-indépendants pouvait être lié à une immaturité de la zone splénique marginale (lieu de stockage des lymphocytes B mémoires ayant été en contact avec un antigène) ou à l'expression plus faible des récepteurs CD21 qui sont impliqués dans la reconnaissance des polysaccharides capsulaires par les lymphocytes B (Licciardi et al. 2012). La stratégie la plus courante est de convertir les polysaccharides en antigènes T-dépendants par leur greffage sur une protéine porteuse, ce qui a entraîné le développement de vaccins glycoconjugués (section II.2.4 suivante).

II.2.4 Les vaccins glycoconjugués

II.2.2.1 Composition et schéma vaccinal

Le vaccin Prevenar13® est commercialisé par l'entreprise Pfizer et est composé de différents polysaccharides capsulaires issus de 13 sérotypes de *S. pneumoniae* conjugués à une protéine porteuse : un mutant non toxique de la toxine diphtérique, la CRM197. Les polysaccharides des sérotypes 1, 3, 4, 5, 6A, 6B, 7F, 9V, 14, 18C, 19F, 19A et 23F sont contenus dans ce vaccin. Son administration s'effectue par injection sous-cutanée ou intramusculaire. En France, il est recommandé pour tous les enfants de moins de 2 ans dont l'administration est préconisée en trois temps: à l'âge de deux mois, de 4 mois et de 11 mois. Pour les personnes de plus de 2 ans présentant un risque élevé, un schéma vaccinal utilisant une combinaison des vaccins Prevenar13® et Pneumovax® est actuellement également recommandée (Tableau 12). Lors de leur utilisation

combinée, le vaccin Prevenar13[®] est toujours administré avant le vaccin Pneumovax[®] car sinon le vaccin Pneumovax[®] peut baisser l'efficacité de la réponse apportée par le Prevenar13[®] (HCSP 2018).

Tableau 12: Schéma vaccinal recommandé en France à partir de 2017 (PV23 = Pneumovax[®]; PCV13 = Prevenar13[®])(HCSP 2018)

Enfants de moins de 2 ans	Enfants de 2 à 5 ans à risque	Enfant de plus de 5 ans et adultes à risques	
Enfants de moins de 2 ans sauf cas ci-dessous : PCV13 : à 2 mois, 4 mois et 11 mois	PV23 : à 2 ans	Immunodéprimés, syndrome néphrotique, brèche ostéo-méningée, implants cochléaires ou candidat à l'implantation :	Personnes à risque élevé (sauf cas précédemment cités) : PV23
Prématurés et nourrissons à risque : PCV13 à 2 mois, 3 mois, 4 mois et 11 mois	Si non vaccinés précédemment : 2 doses de PCV13 (S0, S8) puis PV23 (S16)		

II.2.2.1 Mécanisme d'action des vaccins glycoconjugués

Comme nous l'avons vu précédemment (section II.2.1.3), les polysaccharides capsulaires sont des antigènes T-indépendants. Ils ne génèrent donc pas de réponse mémoire classique car ils ne sont pas présentés aux lymphocytes T au niveau du CMHII. Cependant, en les greffant à des protéines porteuses, ceux-ci sont alors convertis en antigènes T-dépendant par le mécanisme décrit Figure 46 (Avcı et al. 2011). Le lymphocyte B reconnaissant le polysaccharide capsulaire greffé va alors internaliser l'antigène et des fragments de la protéine porteuse vont être présentés à la surface au niveau du complexe majeur histocompatibilité de type II (CMHII), ce qui va entraîner une stimulation des lymphocytes T auxiliaires. Ceux-ci vont donc entraîner l'amplification de la réponse immunitaire par libération de cytokines, ainsi que la différenciation des lymphocytes B activés en plasmocytes et cellules B mémoires. Les cellules B mémoires seront alors spécifiques du polysaccharide capsulaire ayant été reconnu initialement par le lymphocyte B. Ainsi, la réponse immunitaire anti-polysaccharide est similaire à une réponse classique induite lors d'une vaccination avec un antigène T-dépendant (section II.2.1). Il y aura donc une réponse secondaire plus forte, plus spécifique et durable. Grâce aux phénomènes de changements isotopiques et d'hypermutations somatiques (section II.2.1), les anticorps produits seront donc principalement des IgG ayant une affinité plus forte que les IgM produits eux lors d'une réponse initiale.

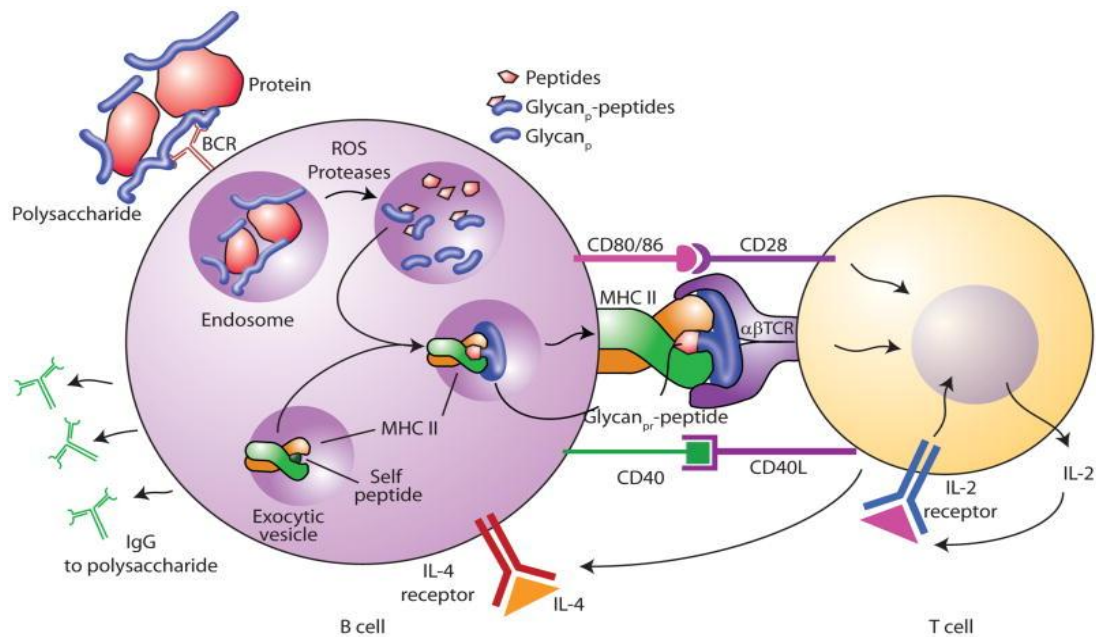


Figure 46: Mécanisme de conversion de polysaccharides capsulaires en antigène T-dépendant.

Lorsque des polysaccharides greffés sur une protéine sont reconnus par un lymphocyte B, il y a une internalisation et ensuite une présentation des parties de la protéine porteuse (potentiellement liées au polysaccharide) au niveau du MHCII à des lymphocytes T auxiliaires ce qui entraîne la différenciation du lymphocyte en plasmocyte mais également en cellules mémoires ayant pour spécificité le polysaccharide initialement reconnu. La réponse immunitaire est alors plus pérenne qu'une réponse T-indépendante et la réponse secondaire sera également plus rapide et plus spécifique. Image extraite de la publication (Avci et al. 2011)

Afin que cette conversion soit efficace, la protéine porteuse doit donc posséder des épitopes T auxiliaires qui pourront être présentés au niveau du CMHII et posséder des groupements fonctionnels accessibles pour réaliser la conjugaison (section II.2.2.2). Actuellement, la protéine CRM197 est celle utilisée pour les vaccins pneumocoques conjugués commerciaux, et seulement 5 protéines porteuses sont utilisées sur le marché à ce jour pour les vaccins glycoconjugués (Toxine du Tétanos (TT), Toxine Diphtérique (DT), Cross Reacting Material 197 (CRM197), "Non typeable *H. influenzae* protein D" (NTHIPD), "Outer membrane protein complex" (OMPC)) (Micoli, Adamo, and Costantino 2018). Ces protéines porteuses ont été sélectionnées parmi des protéines déjà utilisées au sein de vaccins, dont la non toxicité avait déjà été bien établie et dont les procédés de production à grande échelle était maîtrisés (Micoli, Adamo, and Costantino 2018). De nouvelles stratégies sont étudiées afin d'utiliser d'autres types de protéines porteuses. En effet, la réponse à un vaccin glycoconjugué pourrait être réduite à cause d'une immunisation préalable par un vaccin ayant la même protéine porteuse (Pichichero 2013). Par ailleurs, des protéines porteuses ayant également un rôle immunologique spécifique contre l'antigène sont développées (section II.2.5.2).

La conversion de l'antigène polysaccharidique capsulaire peut être vérifiée par détection des anticorps anti-polysaccharides de type IgG au sein de sérums. De plus, après plusieurs injections leur niveau est normalement augmenté grâce à l'effet d'amplification de la réponse secondaire (Nishat and Andreana 2016).

II.2.2.1 Efficacité et limites du vaccin

Lors de la mise au point du premier vaccin glycoconjugué (PCV7), moins de sérotypes ont été inclus comparativement au vaccin non conjugué pour des raisons de coût et de problèmes techniques. Ce vaccin comportait alors les 7 sérotypes responsables de 85% des infections invasives pneumococcales chez les enfants au sein des pays développés (Rennels et al. 1998). Après sa mise sur le marché en 2000, le taux de maladies invasives à pneumocoques a alors rapidement chuté avec

une baisse de 69%, chez les enfants de moins de 2 ans, déjà observée en 2001 aux Etats-Unis (Figure 47) (Whitney et al. 2003). Cependant, malgré une baisse efficace du nombre total de cas de pathologies pneumococcales invasives, un phénomène de remplacement de sérotypes a été observé, ce qui semble limiter l'efficacité du vaccin (Weinberger, Malley, and Lipsitch 2011). Ainsi, la proportion des maladies invasives à pneumocoques dues aux sérotypes non vaccinaux est passée de 17% à 98% aux Etats-Unis en 2007 par rapport à 1998-1999, avec une forte part imputable au sérotype 19A avec une proportion de 48% contre 3% en 1998-1999 (Kellner 2011). Une cause identifiée est le changement de sérotype par transfert de gène horizontal qui permet à des sérotypes invasifs d'échapper à la pression vaccinale (Brueggemann et al. 2007). En effet, le sérotype 19A ne diffère du sérotype 19F (intégré dans PCV7) que par la configuration d'une seule liaison. Cependant, l'utilisation du vaccin Prevenar13® a permis de palier temporairement ce problème de remplacement de sérotypes en intégrant 6 nouveaux sérotypes à sa composition (Figure 47). Malgré tout, l'émergence de nouvelles souches de remplacement commence déjà à se mettre en place, notamment les souches 8 (en Afrique du Sud et en Angleterre), 12F (en Israël, en France, en Angleterre et en Allemagne), et 33F (aux Etats-Unis et au Canada) (Lo et al. 2019). De plus, étant donné que le traitement consiste en l'utilisation d'antibiotiques, ces souches montrent également des profils de résistance accrues et la prise en charge des pathologies dues à ces sérotypes peut donc s'avérer difficile (Lo et al. 2019).

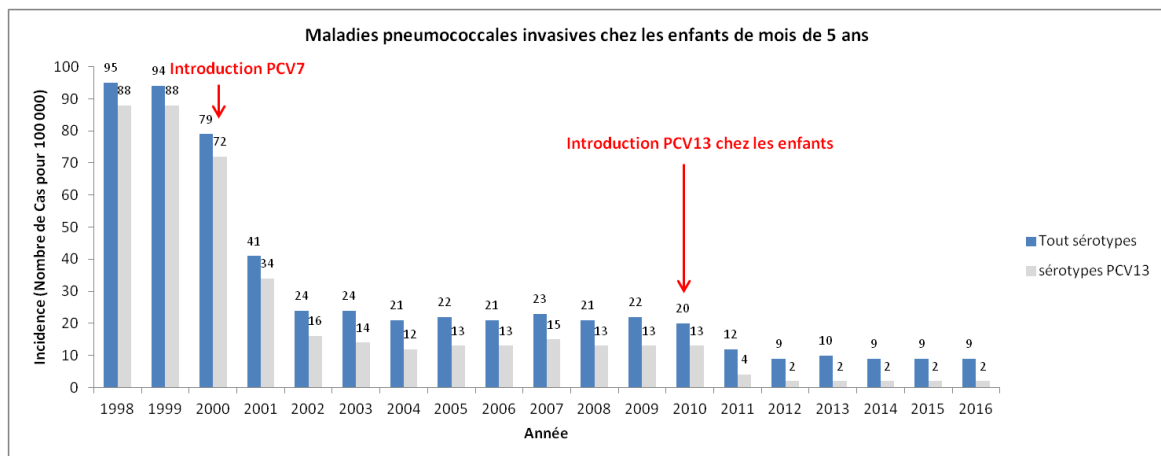


Figure 47: Effet de l'introduction des vaccins conjugués sur le type de sérotypes impliqués dans les pathologies pneumococcales invasives chez les enfants de moins de 5 ans
Image reproduite à partir du site (CDC 2017).

Cependant, mis à part les souches de remplacement, l'utilisation du vaccin Prevenar13® a également permis de réduire ce phénomène de résistance aux antibiotiques (Figure 48). Cette baisse peut s'expliquer par deux effets : grâce i) à la baisse de la circulation des souches résistantes contenues dans les vaccins, ii) ainsi qu'à la baisse du nombre de pathologies nécessitant l'utilisation d'antibiotiques chez les jeunes enfants (Klugman and Black 2018).

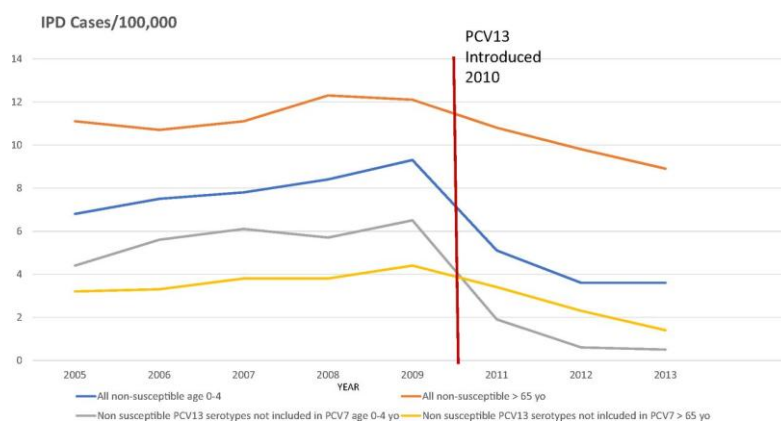


Figure 48: Effet de l'utilisation des vaccins glycoconjugués sur la proportion de souches résistantes aux antibiotiques
Image extraite de la publication

Globalement, le vaccin Prevenar13® a clairement démontré son efficacité pour limiter le développement de pathologies pneumococcales invasives, mais il a également permis de réduire le portage (Esposito and Principi 2015). En effet, le portage est une étape indispensable au développement des maladies pneumococcales et une étude a pu montrer une réduction du portage chez les enfants de moins de 5 ans de 74% des sérotypes couverts par le vaccin chez les enfants vaccinés et une baisse de 50% de ces sérotypes parmi les enfants non vaccinés (ayant un entourage d'enfants vaccinés) (Esposito and Principi 2015). En effet, en baissant le portage, la transmission de ces sérotypes est également réduite et la population non vaccinée peut alors bénéficier d'une immunité collective.

Cependant, malgré ces chiffres encourageants, le phénomène de remplacement de sérotypes observé pose question sur l'avenir de l'efficacité des vaccins actuels. Différentes stratégies sont donc en développement afin d'améliorer la couverture sérotypique (section II.2.5).

II.2.2.2 Méthodes de conjugaison des polysaccharides sur la protéine porteuse

La conjugaison des polysaccharides capsulaires aux protéines se fait généralement par une approche aléatoire, le nombre et la position des polysaccharides sur la protéine ne sont donc pas maîtrisés. Pourtant, ces paramètres influent sur la réponse immunitaire. Des approches permettant de mieux les contrôler sont donc développées. Nous allons tout d'abord voir les méthodes aléatoires avant de présenter le développement de ces nouvelles approches.

II.2.2.2.1 Méthodes de conjugaison aléatoires

Afin de greffer les polysaccharides capsulaires sur les protéines, l'un des deux partenaires est préalablement modifié chimiquement afin de pouvoir réaliser la conjugaison. Cette conjugaison est généralement basée sur la réactivité des fonctions amines des lysines ou des fonctions acides des acides glutamiques ou aspartiques présents sur la protéine porteuse. En effet, les lysines et acides aspartiques ou glutamiques sont plus susceptibles de présenter les fonctions de leurs chaînes latérales à la surface de la protéine et exposées au solvant. De plus, les fonctions amines sont nucléophiles et les fonctions acides carboxyliques sont électrophiles (au niveau de la fonction C=O). Ces fonctions ont donc des propriétés chimiques permettant de réaliser des réactions de conjugaison. Il s'agit d'une conjugaison aléatoire car toutes les fonctions amines ou acides carboxyliques présentes à la surface ne vont pas réagir. Un mélange sera donc obtenu avec une moyenne de couplage selon les proportions de réactants utilisés, la réactivité et l'accessibilité des fonctions susceptibles d'être modifiées.

Modification des sucres :

Les polysaccharides capsulaires peuvent réagir avec les fonctions amines des lysines par amination réductrice après génération d'une fonction aldéhyde par oxydation du saccharide ou après

activation par l'ajout de groupements réactifs : i) ester succinimide, ii) cyano-ester, iii) ester squarique (Figure 49A). Quant aux fonctions carboxyliques, elles peuvent réagir avec un sucre dérivé par une fonction amine ou hydrazide par une réaction de condensation (Figure 49B) (Micoli, Adamo, and Costantino 2018).

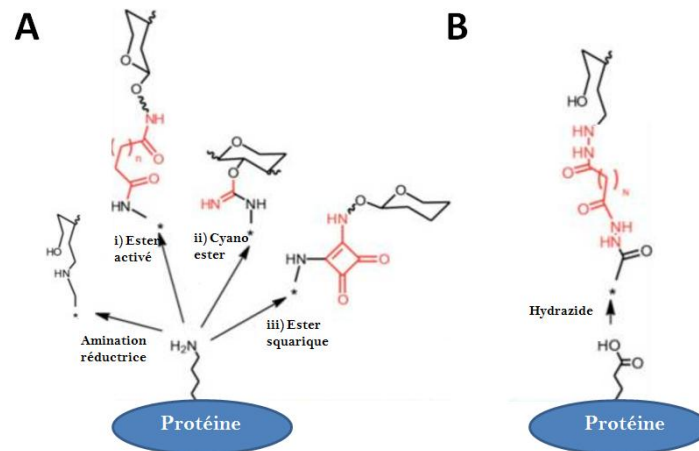


Figure 49 : Modification des polysaccharides afin de les conjuguer aux protéines. Le greffage des polysaccharides peut s'effectuer grâce à la modification des polysaccharides pour générer des groupements réactifs afin de pouvoir réagir avec des lysines (A) ou des acides glutamiques ou aspartiques (B). Image adaptée à partir de la publication (Micoli, Adamo, and Costantino 2018)

Modification des protéines :

La modification préalable des fonctions amines ou acides carboxyliques sur les protéines ajoute un lien portant des fonctions permettant la conjugaison tout en limitant l'encombrement stérique et ainsi augmente l'efficacité de cette réaction. Les groupes ajoutés peuvent être i) des groupes maléimides ou thiols, ii) des groupes bromoacétyle ou cyano-ester, ou iii) des groupes dihydrazines, ainsi que iv) des fonctions orthogonales permettant des conjugaisons spécifiques et rapides (Figure 50) (Micoli, Adamo, and Costantino 2018).

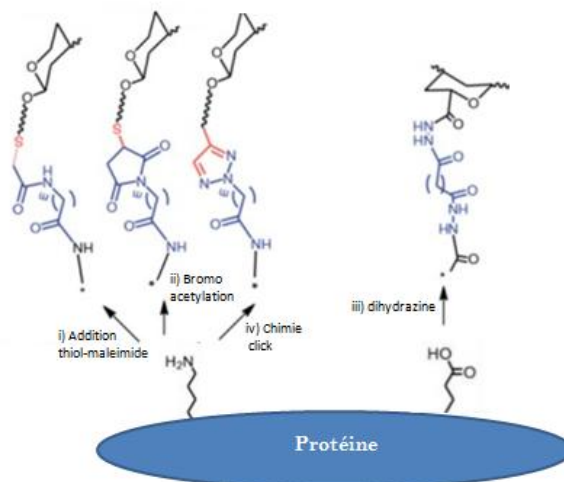


Figure 50: Modification sur les protéines afin de greffer des polysaccharides capsulaires. Le greffage des polysaccharides sur la protéine porteuse peut également s'effectuer grâce à des modifications de protéines au niveau des amines ou des fonctions acides carboxyliques. Image adaptée à partir de la publication de (Micoli, Adamo, and Costantino 2018).

La méthode de conjugaison utilisée dans le vaccin Prevenar13® consiste en une amination réductrice (Figure 49) par une oxydation des fonctions diols au niveau des polysaccharides avec de

l'acide périodique, suivie d'une réaction avec les fonctions amines de la protéine en présence de borohydrure de sodium (P. W. Anderson, Eby, and Rochester 1994). Il s'agit d'une conjugaison aléatoire, les sites de conjugaison ne sont donc pas maîtrisés. D'autres méthodes sont développées dans le but de maîtriser les sites de fixation (section suivante II.2.2.2.2).

II.2.2.2.2 Méthodes de conjugaison spécifiques

Méthodes basées sur la réactivité de la protéine native :

La réactivité d'acides aminés moins fréquemment présents au sein des protéines tels que les cystéines ou les tyrosines permet de réaliser des conjugaisons plus ciblées. En effet, les cystéines portent des fonctions thiols nucléophiles qui permettent de réaliser des conjugaisons de type maléimides-thiols. De plus, des stratégies ont également utilisé la réduction de ponts thiols couplés à l'ajout d'un lien portant une fonction réactive permettant la conjugaison (Sakamoto and Hamachi 2019). Une cystéine peut être ajoutée en N ou C-terminal de la protéine afin de réaliser une conjugaison plus ciblée par la génération chimique d'une liaison peptidique (procédé dit de "ligation chimique native") (Sakamoto and Hamachi 2019). Quant aux tyrosines, la fonction hydroxyle portée par le cycle aromatique peut permettre de réaliser des conjugaisons par l'utilisation de ses propriétés acides induites par la délocalisation des électrons sur le groupe benzyle (Sakamoto and Hamachi 2019).

In vivo, des enzymes peuvent transférer des polysaccharides sur des protéines après la reconnaissance des séquences signales. Ces enzymes ont été exploitées dans le cadre d'une technologie permettant de générer des protéines glycosylées (Protein Glycan Coupling Technology) dans le cadre d'une application vaccinale (procédé développé section II.2.5.2) (Kay, Cuccui, and Wren 2019; Giuseppe Stefanetti et al. 2015; Z. Ma et al. 2018).

Toutes ces méthodes sont cependant limitées par la séquence protéique initiale. En effet, l'utilisation de la thiol-addition sous-entend que la protéine ne possède pas de cystéines impliquées dans un rôle structural important, et l'utilisation d'enzymes nécessite l'introduction de la séquence signal dans la protéine cible, ce qui peut affecter les épitopes de la protéine porteuse. Une méthode permet de s'affranchir de ces limitations par l'ajout de fonctions orthogonales permettant la conjugaison à des positions choisies. Il s'agit de l'utilisation d'acides aminés non naturels.

Méthodes basées sur les acides aminés non naturels :

L'incorporation d'acide aminé non naturel permet d'ajouter des fonctions bioorthogonales pour réaliser de la "chimie click", c'est à dire avec une réactivité hautement spécifique et de manière très efficace (Patterson, Nazarova, and Prescher 2014). De plus, ces réactions peuvent s'effectuer dans des conditions biocompatibles, c'est-à-dire à température et pH physiologiques et en présence de composants biologiques. Mais malgré un fort potentiel, leur utilisation pour la génération de vaccins glycoconjugués est rare. En effet, à notre connaissance l'exploitation d'acides aminés non naturels pour les vaccins glycoconjugués est citée dans plusieurs revues présentant les différentes méthodes de conjugaison disponibles mais n'est publiée qu'au sein d'un brevet couvrant largement les possibilités de leur utilisation par incorporation de manière site spécifique ou résidu spécifique (Fairman, Heinrichs, and Chan 2018) et au sein d'une publication réalisée par la même entreprise sur la génération d'un vaccin glycoconjugué contre la malaria. Dans cette étude, la protéine porteuse a été produite *in vitro* avec un acide aminé non naturel portant une fonction azoture permettant de réaliser la conjugaison avec l'haptène sucre par chimie click (CuAAC) (Kapoor et al. 2018).

Néanmoins, plusieurs types de fonctions pourraient être ajoutés par l'intermédiaire d'acides aminés non naturels pour réaliser différentes réactions de conjugaison bioorthogonale (Figure 51). La première réaction bioorthogonale ayant été décrite est la réaction de Staudinger entre une fonction azoture et des phosphines (Schilling et al. 2011). Cette réaction est cependant lente et génère des

sous-produits. Des réactions de cycloaddition ont alors été mises au point dont notamment la réaction "*Copper(I) catalyzed alkyne azide cycloaddition*" (CuAAC) qui consiste en une réaction entre une fonction azoture et une fonction alcyne en présence d'ions cuivre^I (Rostovtsev et al. 2002). Les fonctions impliquées dans cette réaction sont petites, ce qui limite les risques d'encombrement stérique. De plus, cette réaction est rapide et ne génère pas de sous-produits. Par contre le Cuivre^I est un composé cytotoxique qui peut nuire à certaines applications *in vivo*. Une variante de cette réaction consiste à utiliser un partenaire ayant sa fonction alcyne déformée au sein d'un cycle ("*Strain promoted alkyne azide cycloaddition*", SPAAC) (Agard, Prescher, and Bertozzi 2004). Cette déformation due à la tension apportée par le cycle permet de rendre cette fonction efficace en absence de cuivre^I, mais également d'accélérer la réaction. Cependant, les composés cycliques impliqués sont gros et présentent des profils d'hydrophobicité élevés, ce qui peut aboutir à des interactions non spécifiques. Malgré tout, aujourd'hui, plus de 10 partenaires différents ont été développés ce qui fournit un large éventail d'outils permettant cette réaction (Ramil and Lin 2013). Enfin, les réactions de Diels-Alder à demande inverse (IEDDA) présentent une autre réactivité bioorthogonale avec la vitesse de réaction la plus rapide à ce jour (Rossin et al. 2013). Des travaux ont couplé deux types de réactions (CuAAC et SPAAC) permettant ainsi des couplages orthogonaux multiples (Sachdeva et al. 2014). Ces réactions chimiques orthogonales permettent donc de réaliser sélectivement des réactions de couplage en présence de mélanges réactionnels complexes tels que la cellule. La recherche future se dirige vers la mise au point de toujours plus d'outils de couplage avec des efforts importants pour fournir des réactions orthogonales plus rapides, ayant une réactivité photoactivable ou encore ayant des motifs moins encombrants (Devaraj 2018).

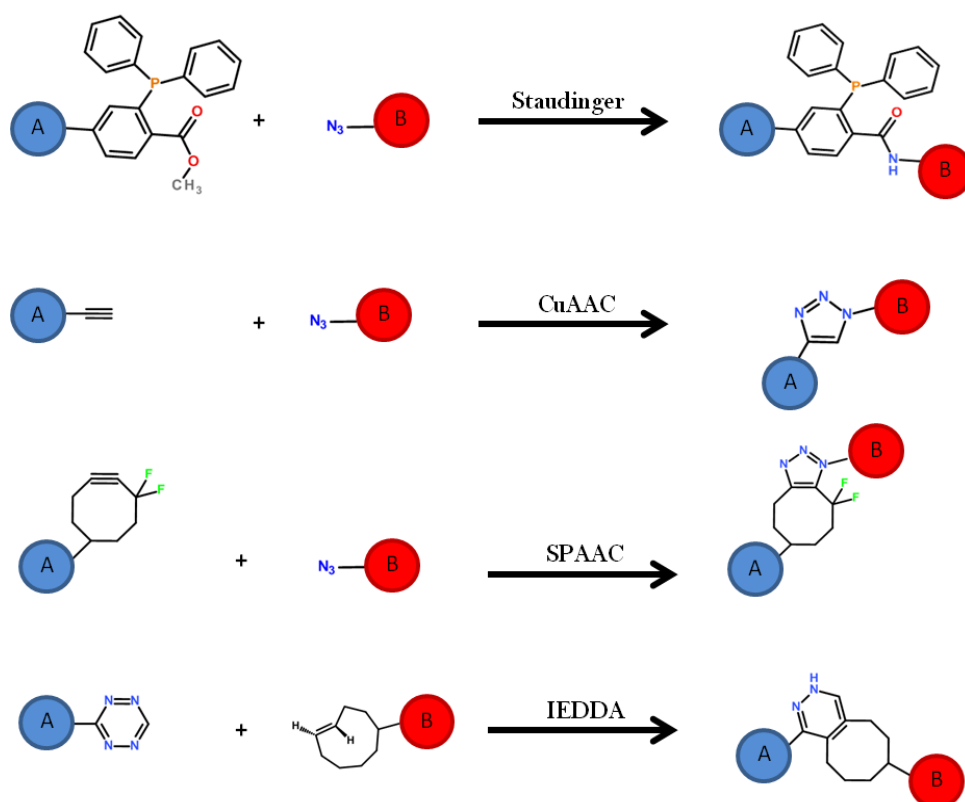


Figure 51: Principales réactions bioorthogonales
Image reproduite à partir de la publication (Grammel and Hang 2013)

Des aminoacyl-ARNt synthétases (Chapitre 1) ont été développées afin d'incorporer différents acides aminés non naturels, portant des fonctions permettant les réactions de conjugaison décrites ci-dessus, au sein de protéines de manière site-spécifique (Dumas et al. 2015). Bien que cette technologie ne soit pas encore très exploitée dans le domaine des vaccins, les réactivités CuAAC, SPAAC, IEDDA sont donc utilisables afin de réaliser des conjugaisons entre une protéine porteuse et un polysaccharide capsulaire.

Ces réactivités orthogonales ont d'ailleurs pu être exploitées sans l'utilisation d'acides aminés non naturels dans le cadre de dérivation de fonctions permettant la réalisation de réaction de type CuAAC ou SPAAC (Berti and Adamo 2018; Grandjean et al. 2005). Cependant afin d'obtenir une conjugaison ciblée, les dérivation doivent être réalisées sur des acides aminés naturellement en faible proportion au sein de la protéine (cystéines ou tyrosines par exemple, voir Chapitre 1 section II).

II.2.5 Vaccins: principales stratégies émergentes

II.2.5.1 Vaccins en phases cliniques

Afin de faire face aux problématiques de remplacement de sérotypes, plusieurs stratégies sont développées.

Tout d'abord, un vaccin conjugué PCV15 développé par l'entreprise Merck (appelé V114 sur la base de donnée Clinicaltrials.gov) intégrant deux sérotypes additionnels (22F et 33F) par rapport au Prevenar13® de l'entreprise Pfizer a été testé en essai clinique et est actuellement en phase 3 jusqu'en 2020 (NCT03480802 2018). C'est à dire que l'innocuité ainsi que l'efficacité du vaccin ont déjà été prouvés et que celui-ci entre maintenant dans une phase permettant de démontrer son utilité par rapport aux vaccins existants avant sa mise sur le marché.

De plus, différentes stratégies basées sur l'utilisation de nouvelles protéines antigéniques au sein de vaccins sont également en développement (Tableau 13). Comme nous l'avons vu précédemment, (section I.4.2.3) des protéines telles que les adhésines peuvent être accessibles et surexprimées au cours de la phase de colonisation. Ces protéines sont généralement plus conservées entre les sérotypes que les polysaccharides capsulaires. La présentation d'épitopes protéiques permettrait alors d'avoir une meilleure couverture sérotypique. Notons qu'en se basant sur le profil d'expression des protéines de surface de *S. pneumoniae*, un vaccin pourra alors avoir une efficacité plus ciblée sur une phase spécifique de l'infection (colonisation, infections locales, ou infections invasives) (Pichichero, Khan, and Xu 2016a). Ainsi, les anticorps dirigés contre des adhésines sont supposés réduire l'étape de colonisation et donc le portage. La pneumolysine a également été intégrée à certaines formulations vaccinales. Cette protéine étant cytotoxique, deux stratégies ont été utilisées afin de la neutraliser : par méthode chimique (traitement au formaldéhyde) chez GlaxoSmithKline, ou par mutagenèse (substitution) chez Sanofi Pasteur. Enfin lorsqu'un rôle protecteur a pu être démontré, d'autres protéines ayant un rôle dans les fonctions nécessaires à la croissance ou à la virulence de *S. pneumoniae* telles que des protéines liants les métaux (PhpA, PsaA) ou des protéines impliquées dans la division cellulaire (hydrolase du peptidoglycane PcsB, kinase impliquée dans la régulation de la division cellulaire StkP) ont également été intégrées à des stratégies vaccinales actuellement en essais cliniques. Comme de nombreuses protéines sont exprimées à la surface de *S. pneumoniae*, les stratégies utilisant des mélanges de protéines antigéniques sont prépondérantes (Pichichero, Khan, and Xu 2016a). Par ailleurs, un vaccin couplant l'utilisation de protéines antigéniques avec un vaccin conjugué traditionnel (sur CRM197) est également exploré par GSK. En effet, les polysaccharides capsulaires couvrent une surface importante du pathogène, et des protéines antigéniques conservées entre les sérotypes peuvent

apporter une réponse plus large. Coupler les polysaccharides capsulaires à des protéines antigéniques pourrait alors permettre de potentialiser l'efficacité du vaccin. Enfin, l'utilisation de bactéries non encapsulées de *S. pneumoniae* ayant été préalablement tuées a permis d'efficacement générer des anticorps de type IgG contre des protéines de surface telles que la PspA et la pneumolysine. De plus, comparativement à un vaccin glycoconjugué ou protéique bien défini, celui-ci sera produit plus simplement et à bas coût et sera donc plus facilement distribuable dans les pays en développement (Lagousi et al. 2019). Cependant, il est difficilement standardisable et le vaccin a plus de risque de présenter des effets indésirables non maîtrisés à cause de sa composition (Morais, Teixeira, and Suarez 2019). Globalement, les données pré-cliniques de ces différents vaccins (protéiques mais aussi cellules entières) semblent indiquer que ceux-ci pourraient contrebalancer les risques dus aux remplacements de sérotypes (L. H. Lee, Gu, and Nahm 2014). Cependant, les niveaux d'expression de ces protéines antigéniques ainsi que leurs séquences primaires varient selon les sérotypes. L'utilisation combinée de différentes protéines cibles permet de limiter cette problématique mais augmente le coût de production du vaccin. L'identification de séquences immunogènes et leurs caractérisations a permis le développement de nouvelles stratégies actuellement en essais pré-cliniques basées sur l'utilisation de combinaisons de peptides antigéniques (Lagousi et al. 2019).

Tableau 13: Différentes stratégies vaccinales basées sur des antigènes protéiques.

La phase I permet de vérifier l'innocuité du vaccin, alors que la phase II détermine les conditions d'utilisation (dose, schéma vaccinal). Enfin la phase III permet de valider l'intérêt du vaccin par rapport à des vaccins existants, si la balance bénéfice - risque est favorable le vaccin est alors mis sur le marché après une autorisation préalable (AMM) et entre alors dans une phase de pharmacovigilance (phase IV) qui inventorie les symptômes éventuels ainsi que leur fréquence d'apparition. (données obtenues grâce à la base de données clinicaltrials.gov ainsi qu'à la publication de (Pichichero, Khan, and Xu 2016b))

Antigène	Type de protéine	Laboratoire ou Compagnie	Phase clinique
PspA	Adhésine (liaison à la choline)	Sanofi-Pasteur	Phase I
PspA PsaA	Adhésines -liaison à la choline -liaison aux cadhérines	Sanofi-Pasteur/CDC	Phase I
PhpA PhtB	-Métal binding protéine -Adhésine	GSK	Phase II
PsaA PcsB StkP	-Adhésine, transporteur ABC -hydrolase du peptidoglycane -kinase impliquée dans la division cellulaire	InterCell AG/Novartis/PATH	Phase I
PhtD PcpA PlyD1	-Adhésine -Adhésine (Liaison à la choline) -Toxine (lyse des cellules hôtes)	Sanofi-Pasteur	Phase I
PhtD +/- dPly +/-PCV10	-Adhésine -Toxine (lyse des cellules hôtes) -Vaccin conjugué de valence 10	GSK	Phase I
PhtD dPly PCV10	-Adhésine -Toxine (lyse des cellules hôtes) -Vaccin conjugué de valence 10	GSK	Phase II
SP0148 SP1912 SP2108	-transporteur ABC (<i>Prédiction</i>) -thioredoxine (<i>Prédiction</i>) -transporteur ABC de sucre (<i>Prédiction</i>)	Genocea	Phase II a
Cellules entières non-encapsulées tuées	NA	Boston Children's Hospital- Instituto Butantan-PATH	Phase I

II.2.5.2 Nouvelles stratégies pour la génération de vaccins glycoconjugués

Ce travail de thèse porte sur la génération d'un vaccin glycoconjugué ciblé afin d'en étudier des caractéristiques clés (Chapitre 3). Nous allons donc voir ci-dessous les principales stratégies émergentes concernant le développement de vaccins glycoconjugués contre *S. pneumoniae* à travers : i) le choix de la protéine porteuse, ii) la maîtrise des paramètres de conjugaison et iii) un mode de production innovant.

i) Choix de la protéine porteuse :

Comme nous l'avons vu précédemment, 5 protéines porteuses sont actuellement utilisées pour générer des vaccins glycoconjugués (section II.2.2.1). Ces protéines porteuses vont permettre de convertir le polysaccharide capsulaire en antigène T dépendant. Cependant, plusieurs stratégies évaluent l'apport d'une protéine qui posséderait un double rôle de protéine porteuse et de protéine générant une réponse immunitaire spécifique contre le pathogène ciblé. En effet, des protéines issues de *S. pneumoniae* peuvent générer une réponse immunitaire. De plus, celles-ci peuvent être conservées entre les différents sérotypes et donc générer une réponse élargie à plusieurs sérotypes (Micoli, Adamo, and Costantino 2018). Les principales protéines immunogènes de *S. pneumoniae* étudiées en tant que protéines porteuses sont répertoriées dans le Tableau 14.

La protéine PspA est une protéine très immunogène qui est impliquée dans l'adhésion de *S. pneumoniae* et dans l'inhibition de sa liaison au complément. Elle est présente chez tous les sérotypes et possède des régions conservées. Elle a été évaluée comme potentielle protéine porteuse en combinaison avec un polysaccharide issu des sérotypes 14, 23 ou 6B chez la souris et a pu montrer une conversion de la réponse anti-polysaccharide ainsi que la production d'anticorps contre la PspA (Tableau 14). Elle a également été évaluée chez la souris en combinaison avec des polysaccharides issus de *Salmonelle* ce qui permettrait une double protection contre la fièvre typhoïde et les maladies liées au pneumocoque (Kothari et al. 2014). Néanmoins, une faible homologie avec une myosine cardiaque humaine a soulevé des inquiétudes sur une réactivité croisée auto-immune qui n'ont cependant pas été confirmées expérimentalement (Lagousi et al. 2019).

La pneumolysine est une protéine toxique impliquée dans la formation de pores lors de l'infection (section I.4.2.4). Une pneumolysine mutante détoxifiée a été conjuguée à des polysaccharides issus des sérotypes 6B, 14, 19F et 23F et son immunisation chez la souris a pu générer un niveau d'anticorps anti-polysaccharides capsulaires comparable à un conjugué avec une protéine porteuse classique (TT) ainsi que des anticorps anti-pneumolysine (Michon et al. 1998). La Pneumolysine détoxifiée (dPly) a également été étudiée ainsi que la protéine PhtD (Pneumococcal histidine triad D (adhésine)) en tant que protéines porteuses d'un vaccin combiné contenant 8 polysaccharides capsulaires en essais cliniques (Tableau 14) (Pauksens et al. 2014). La complexité due à la multiplication des antigènes a rendu difficile les conclusions de l'étude mais les protéines dPly et PhtD ont été considérées comme des protéines porteuses efficaces permettant à la fois la conversion d'antigènes polysaccharidiques spécifiques et la génération d'anticorps anti-protéiques.

La protéine PsaA est une protéine impliquée dans l'adhésion et le transport d'ions manganèse (Rajam et al. 2008). Elle est très conservée parmi les différents sérotypes et les anticorps anti-PsaA permettent une protection contre la colonisation (Rajam et al. 2008). De plus, les épitopes B et T de la PsaA ont pu être identifiés (Chapitre 3, section I.3). Cette protéine a été évaluée en tant que protéine porteuse lors de travaux de conjugaison avec du polysaccharide purifié issu du sérotype 23F ou un tétrasaccharide issu du sérotype 14 (Tableau 14), ainsi que lors d'une étude couplant la PsaA à des polysaccharides capsulaires issus de *Hemophilus influenzae b* (Z. Chen 2016). Le couplage de la PsaA avec ces polysaccharides capsulaires a alors permis la génération d'anticorps contre la PsaA ainsi que la conversion des polysaccharides en antigènes T dépendants.

Tableau 14: Stratégies de vaccins glycoconjugués avec des protéines porteuses immunogènes

Protéine porteuse	Sérotypes des polysaccharides capsulaires associés	Etape d'infection ciblée par les anticorps produits	Références
PspA	14 23 6B Polysaccharides origine <i>Salmonella Typhi</i>	Contre les stades invasifs	Sérotype 14 : (Da Silva et al. 2017; Santamaria et al. 2011) Sérotype 23 : (Csordas et al. 2008) Sérotype 6B : (Perciani et al. 2013; H. Lin et al. 2015) Hybride avec polysaccharides issus de <i>Salmonella Typhi</i> (Kothari et al. 2014)
dPly	6B, 14, 19F, 23F 19A	Contre les stades invasifs	Sérotypes 6B, 14, 19F, 23F : (Michon et al. 1998) Sérotype 19A : en utilisation combinée avec 22F conjugué à PhtD, 4, 5, 7F, 14 conjugué à HiD, 18C conjugué à TT et 19F à DT (Pauksens et al. 2014)
PhtD	22F	Inhibe la colonisation	utilisation combinée avec 19A conjugué à dPly; 4, 5, 7F, 14 conjugué à HiD; 18C conjugué à TT et 19F à DT (Pauksens et al. 2014)
PsaA	23F 14 Polysaccharide origine <i>Hemophilus influenzae b</i>	Inhibe la colonisation	Sérotype 23F: (H. Lin et al. 2010) Sérotype 14: (Prasanna et al. 2019; Pillot et al. 2019) Hybride avec polysaccharides issus de <i>Hib</i> : (Z. Chen 2016)

ii) Paramètres de conjugaison :

L'utilisation de méthodes de conjugaison plus ciblées (section II.2.2.2.2) permet d'étudier les caractéristiques d'un vaccin glycoconjugué mieux défini et ainsi de comprendre l'influence de paramètres tels que le nombre d'unités polysaccharidiques sur une protéine, leur position au sein de la séquence, ainsi que la nature du lien qui relie la protéine et le polysaccharide ainsi que la composition précise du polysaccharide capsulaire. Jusqu'à présent, l'utilisation de méthodes de conjugaison aléatoires était prépondérante lors de la génération de vaccins glycoconjugués. L'étude des facteurs indiqués ci-dessus est donc rare. Pourtant, ils sont reconnus comme influençant la réponse immunitaire (Khatun, Stephenson, and Toth 2017).

Tout d'abord, concernant, les méthodes de conjugaison, celles-ci ont une influence sur la présentation des épitopes. Ainsi, l'utilisation d'un lien espaceur limite l'encombrement stérique et favoriserait donc l'exposition des épitopes polysaccharidiques comme cela a notamment pu être montré dans une étude sur un vaccin contre le sérotype 14 de *S. pneumoniae* (Mawas et al. 2002). Cependant, la reconnaissance de certains liens reliant le polysaccharide à la protéine comme des groupes triazoles peuvent limiter la réponse protectrice en générant des anticorps non spécifiques (Adamo et al. 2014).

Par ailleurs, bien que très peu étudié, le ratio entre la protéine et le nombre d'unités polysaccharidiques, ainsi que leur position au sein de la séquence joueraient également un rôle significatif. En effet, la position de greffage pourrait influencer sur l'efficacité de la reconnaissance des polysaccharides par les lymphocytes ainsi que sur le type de peptides présentés au niveau du CMHII (fonctionnement biologique du vaccin détaillé section II.2.2.1). Ainsi, dès 1992, Ishioka *et al.* ont étudiés la présentation de saccharides au niveau du CMHII et ont pu montrer que les polysaccharides ne semblaient pas capables d'être présentés au niveau du CMHII. Cependant, ils ont également

montré que selon la position de motifs glycosidiques sur des épitopes T, l'immunogénicité de ces peptides pouvait en être altérée ou non (Ishioka et al. 1992). Par la suite, Pozsgay *et al.* travaillant sur un vaccin contre *Shigella dysenteriae* ont testés plusieurs vaccins glycoconjugués réalisés à partir de différents types de saccharides (purifiés ou synthétiques de tailles variables allant d'un tétramère à un hexadécamère) greffés sur une protéine porteuse avec des ratio variables selon le conjugué. Toutes les souris ont été immunisées avec une dose constante de polysaccharide de 2,5µg et ils ont alors pu montrer des différences significatives entre les différents vaccins, avec une meilleure réponse au niveau d'un conjugué avec un hexadécamère en proportion 9/1 par rapport à la protéine alors que le conjugué réalisé avec le polysaccharide purifié a entraîné une réponse dirigée contre le polysaccharide très faible (Pozsgay et al. 1999). Puis en 2011, une étude réalisée par Avci *et al.* a suggéré que des fragments de polysaccharides pourraient être présentés avec l'épitope T lors de sa présentation au niveau du complexe CMHII (section II.2.2.1) (Avci et al. 2011). En effet, lors de leurs travaux sur des peptides conjugués, les chercheurs ont pu montrer que les lymphocytes reconnaissent efficacement des peptides antigéniques glycoconjugués (Avci et al. 2011). De plus, les niveaux d'anticorps produits pour le glycoconjugué peptidique étaient supérieurs aux niveaux obtenus pour le conjugué protéique aléatoire malgré une dose antigénique 100 fois plus forte pour ce dernier (Avci et al. 2011). Ces résultats suggèrent qu'une présentation différente des antigènes polysaccharidiques peut clairement optimiser la réponse immunitaire contre ces derniers. Par la suite, deux études réalisées par Novartis en 2014 et 2015 sur un vaccin contre la salmonelle ont également montré que l'influence du ratio protéine/polysaccharide ainsi que du type de lien entre le polysaccharide antigénique et la protéine porteuse a également une influence significative. Par exemple, malgré un ratio protéine / polysaccharide antigénique équivalent, l'utilisation de différentes méthodes de conjugaison influençait la réponse immunitaire d'un facteur 10 (Giuseppe Stefanetti et al. 2015; G Stefanetti et al. 2014). Plus récemment, Carboni *et al.* ont généré différents vaccins glycoconjugués à partir de la plus petite unité polysaccharidique représentative du sérotype III de Streptocoque de type B (pentamère) greffée en quantité variable sur la protéine porteuse CRM197 (de 3 à 11 pentamères par protéine). Ils ont alors pu montrer qu'une multiprésentation du pentamère sur la protéine avait un effet positif sur la réponse anti-polysaccharidique, probablement grâce à un effet d'avidité au niveau de la reconnaissance de l'antigène par les lymphocytes B (Carboni et al. 2019).

Globalement, les paramètres de greffage sont encore peu étudiés et les méthodes aléatoires présentent des structures non homogènes. Cependant, l'émergence de techniques permettant le couplage spécifique entre les partenaires du vaccin, comme l'utilisation d'acides aminés non naturels, devrait permettre la multiplication de ces études et ainsi la mise en évidence de caractéristiques générales permettant la conception d'un vaccin glycoconjugué optimal. En effet, ces méthodes permettent également la génération d'un vaccin de composition constante et très définie.

iii) Production recombinante de vaccins glycoconjugués :

Comme nous l'avons vu précédemment (sections II.2.1.2 et II.2.2.2), le procédé classique de production d'un vaccin glycoconjugué consiste en la conjugaison entre des polysaccharides capsulaires purifiés à partir du pathogène *S. pneumoniae* puis greffés à une protéine porteuse qui aura été préalablement produite de façon recombinante (Figure 52).

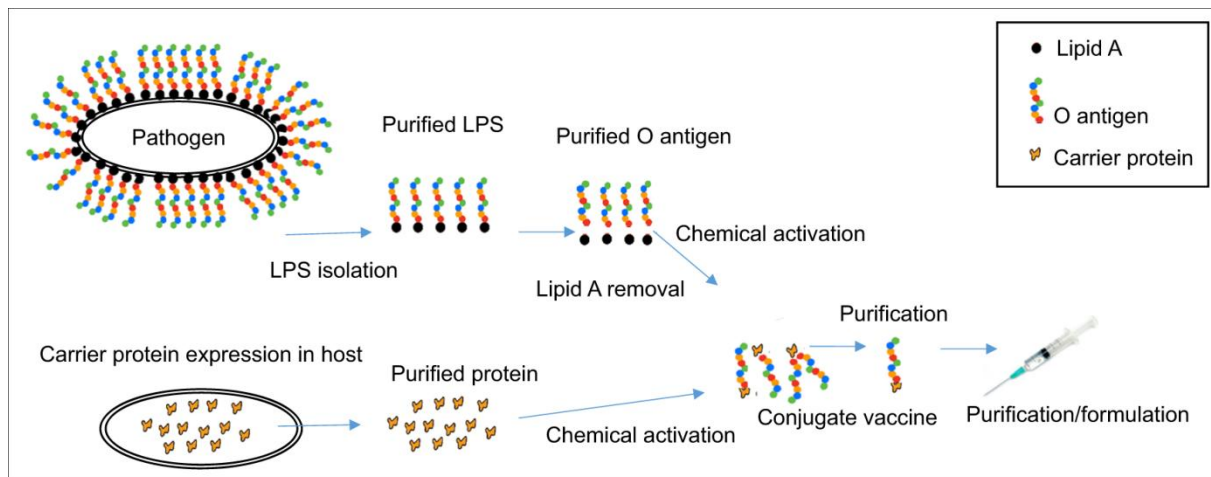


Figure 52: Procédé de production classique d'un glycoconjugué.

En général, un vaccin glycoconjugué est produit à partir de polysaccharides capsulaires purifiés qui seront greffés à une protéine porteuse préalablement produite au sein d'un hôte d'expression (le plus souvent *E. coli*). Image extraite de la publication de (Kay, Cuccui, and Wren 2019).

Une méthode alternative consiste à produire le glycoconjugué de façon complètement recombinante par le biais de l'utilisation de la technologie de couplage de saccharides sur des protéines au sein de bactéries (Protein Glycan Coupling Technology, PGCT) (Kay, Cuccui, and Wren 2019). Cette technologie est issue de la découverte de PglB, une oligosaccharyltransférase (OST), au sein de *Campylobacter jejuni* qui permet de transférer une variété de polysaccharides (possédant un groupe acétamide à l'extrémité réductrice donc pour les sérotypes 1, 4, 5, 12F, 12A, 12B, 25F, 38, 44, 45 et 46) sur une protéine portant la séquence signal D/EXNYS/T (Kay, Cuccui, and Wren 2019). Cette séquence peut être ajoutée en N- ou C-terminal d'une protéine de manière unique ou répétée afin d'augmenter la proportion de polysaccharides greffés. Toutes les étapes de production du glycoconjugué peuvent alors être réalisées chez *E. coli*, ce qui permettrait de réduire drastiquement les coûts de production. La production peut alors être effectuée par l'utilisation de plasmides codant pour les gènes d'expression permettant la synthèse du polysaccharide, le gène d'expression de la protéine porteuse possédant la séquence signal D/EXNYS/T, et le gène codant pour l'oligosaccharyltransférase (Kay, Cuccui, and Wren 2019). La recherche de paralogues de PglB tente d'étendre les types de polysaccharides transférables par cette technique (sans groupe acétamide à leur extrémité), d'augmenter l'efficacité de transfert, ainsi que d'identifier d'autres séquences utilisables pour des transferts ciblés d'oligosaccharides (Kay, Cuccui, and Wren 2019; Harding et al. 2019). Outre ces développements, les principales problématiques de cette technologie sont la génération de souches d'*E. coli* sans système de synthèse de polysaccharide qui interfère avec la synthèse ciblée, et la réduction de la séquence signal (Kay, Cuccui, and Wren 2019). Dans le cadre de vaccins contre *S. pneumoniae*, cette technologie a été utilisée pour générer un vaccin glycoconjugué entre plusieurs polysaccharides capsulaires et la protéine porteuse rEPA (forme recombinante de l'endotoxine A issue de *Pseudomonas aeruginosa*) qui est actuellement en étude clinique (phase I) (Kay, Cuccui, and Wren 2019). Cette technologie a également été utilisée pour générer un vaccin glycoconjugué contenant le polysaccharide capsulaire du sérotype 4 greffé sur des protéines immunogènes (neuraminidase NanA, ABC transporteurs PiuA et Sp0148), testé chez la souris (Reglinski et al. 2018). Au cours de cette étude, l'intensité de la réponse anti-capsulaire s'est avérée différente selon la protéine porteuse (avec un niveau plus fort pour PiuA qui était alors comparable au Prevenar13®) ce qui souligne l'importance de ce facteur.

***Chapitre 3 : Génération de Vaccins
glycoconjugués contre Streptococcus
pneumoniae***

Avant-propos / Présentation des objectifs

Comme nous l'avons vu lors du chapitre précédent, les vaccins glycoconjugués sont un moyen efficace de générer une protection contre des maladies bactériennes invasives y compris chez les jeunes enfants (Chapitre 2, section II.2.4). Cependant, leur efficacité est impactée par le phénomène de remplacement des sérotypes. Les vaccins actuellement disponibles sur le marché doivent donc évoluer. De plus, ils sont produits par une méthode de conjugaison aléatoire et ne sont donc pas homogènes, ce qui compromet l'étude de l'influence du site de greffage des sucres ou du ratio sucres greffés / protéine porteuse. Ces deux paramètres ont pourtant été montrés comme influents sur la réponse immunitaire (Chapitre 2, section II.2.5.2). Au cours de ce travail de thèse, nous avons donc développé une nouvelle stratégie permettant de répondre aux trois objectifs suivants :

- Objectif 1 : Fournir un vaccin contre *Streptococcus pneumoniae* ayant une couverture sérotypique plus large.
- Objectif 2 : Fournir un glycoconjugué site-spécifique et homogène afin de permettre l'étude de l'influence de la position de greffage du saccharide sur la réponse immunitaire.
- Objectif 3 : Fournir un glycoconjugué multivalent homogène afin de permettre l'étude de l'influence du ratio sucre greffé/protéine porteuse sur la réponse immunitaire.

Stratégie pour fournir un vaccin ayant une couverture sérotypique plus large :

Comme nous l'avons vu lors du chapitre précédent (section II.2.5), différentes pistes sont actuellement envisagées afin de permettre une couverture sérotypique plus large du vaccin contre *S. pneumoniae*. Lors de cette thèse, le modèle d'étude choisi est un conjugué déjà utilisé au laboratoire pour d'autres études sur la génération de vaccin contre *S. pneumoniae*. Il est composé de la protéine Pneumococcal Surface Adhesin A sous sa forme mature, c'est-à-dire sans son peptide d'adressage à la membrane (notée mPsaA) conjuguée à un tétrasaccharide immunogène issu du sérotype 14 (PnTS14) (Figure 53).

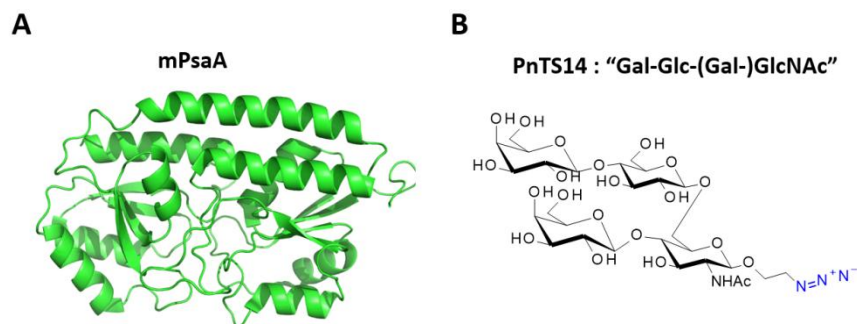


Figure 53: Présentation des partenaires du vaccin glycoconjugué développé.

Le vaccin glycoconjugué sera formé de : A) une protéine très conservée de *Streptococcus pneumoniae* (la mPsaA), B) cette protéine sera couplée à un tétrasaccharide issu du sérotype 14 (PnTS14).

La PsaA est une protéine très conservée au sein des 90 sérotypes de *S. pneumoniae*. De plus, elle est immunogène et permet la génération d'anticorps protecteurs (Chapitre 2, II.2.5). L'utilisation de cette protéine en tant que protéine porteuse permet donc de l'utiliser sous un double rôle de protéine porteuse et immunogène. Etant très conservée, son utilisation pourrait permettre le développement d'anticorps présentant une réactivité croisée contre les différents types de sérotypes

de *S. pneumoniae*. Quant au tétrasaccharide PnTS14 utilisé dans ce modèle d'étude, il correspond à la plus petite unité de polysaccharide capsulaire identifiée permettant la génération d'anticorps fonctionnels contre le sérotype 14 de *S. pneumoniae* (Safari et al. 2008). Il a pour avantage d'être plus facilement synthétisable comparativement à d'autres unités saccharidiques capsulaires plus complexes. De plus, le sérotype 14 est l'un des plus virulents et l'un de ceux ayant développé le plus de résistance aux antibiotiques (Stacevičiene et al. 2016). L'association de la PsaA et du PnTS14 devrait alors permettre de répondre au premier objectif. Cette association a d'ailleurs déjà été mise en place lors d'une thèse développée en parallèle de celle-ci (thèse de Maruthi Prasanna, (Prasanna et al. 2019)).

Contexte et Stratégie pour fournir un glycoconjugué homogène permettant l'étude de la position de greffage sur du sucre PnTS14 sur la protéine porteuse mPsaA :

Ce conjugué entre la mPsaA (forme mature de la PsaA) et le tétrasaccharide PnTS14 a donc déjà été utilisé lors d'une étude précédente au laboratoire portant sur le mode d'administration de ce vaccin (thèse Maruthi Prasana (Prasanna et al. 2019)). Le conjugué entre la mPsaA et le tétrasaccharide PnTS14 avait alors été généré à l'aide d'une conjugaison aléatoire grâce à la dérivatisation des lysines par des groupements S-acétylthio permettant la conjugaison (Chapitre 2, section II.2.2.2.1). Le conjugué était composé d'une moyenne de 5,4 PnTS14 pour 1 protéine mPsaA. Ce vaccin glycoconjugué ainsi que des contrôles tels que la protéine mPsaA seule, ou le conjugué analogue BSA-PnTS14 ont alors été utilisés afin d'immuniser des souris, dont la réponse a été analysée par ELISA. Un des résultats intéressants de cette étude est que la réponse induite par le conjugué mPsaA-PnTS14 entraînait une réponse moins forte contre la mPsaA en comparaison de la réponse induite grâce à l'immunisation avec la mPsaA seule. Parmi les nombreuses hypothèses pouvant expliquer ce résultat, il est possible que :

- Hypothèse 1 : La réponse anti-sucre est plus importante que la réponse anti-protéine à cause d'une compétition entre les différents épitopes.

En effet, la réponse immunitaire peut se développer contre une partie d'un antigène que l'on appelle épitope dominant. Peu de réponse immunitaire sera alors déployée contre un épitope secondaire. Il est donc possible que la présence du sucre capte l'essentiel de la réponse immunitaire limitant ainsi la réponse anti-mPsaA.

- Hypothèse 2 : Le greffage des tétrasaccharides masque des épitopes de la mPsaA

En effet, les sites de greffage des tétrasaccharides n'étant pas maîtrisés, ceux-ci peuvent cacher des séquences immunogènes de la protéine mPsaA.

Comme nous pouvons le voir ci-dessus, au cours de cette étude la question de l'influence du site de greffage du tétrasaccharide s'est posée. Les différents épitopes B et T de la PsaA ont été identifiés ce qui permet d'avoir une approche rationnelle sur la position du greffage du tétrasaccharide par rapport aux séquences immunogènes. Cependant, pour étudier ce paramètre, nous devons générer des glycoconjugués site-spécifiques. Pour cela, il est possible d'utiliser la réactivité des cystéines, ce qui est testé dans le cadre d'un projet réalisé en parallèle de cette thèse (Thèse Aline Pillot, (Pillot et al. 2019)). Cependant, pour certaines protéines les cystéines peuvent être impliquées dans leur structure 3D, il n'est donc pas toujours possible de réaliser des mutants cystéines afin de réaliser de la conjugaison ciblée. Nous cherchons donc à implémenter une nouvelle technologie de conjugaison ciblée par l'utilisation d'acides aminés non naturels qui serait applicable à un plus grand nombre de protéines. En effet, ceux-ci peuvent en théorie être incorporés à toute position au sein de la protéine. Pour cela, nous allons utiliser la réactivité de la N-ε-

propargyloxycarbonyl-L-Lysine communément appelée Propargyl-Lysine (PrK). En effet, le groupement propargyle de la PrK peut réagir spécifiquement par une réaction bioorthogonale de type chimie click (CuAAC) avec un groupement azoture présent sur le tétrasaccharide synthétique. Afin de contrôler sa position d'incorporation, la méthode d'incorporation site-spécifique utilisant des aminoacyl-ARNt synthétases a été utilisée lors de cette thèse (Chapitre 1, section IV.3). Trois positions ont été testées afin de déterminer quel type de site de greffage permet une réponse immunitaire optimale. Ce travail est présenté section I suivante. Un travail d'optimisation de l'incorporation de l'acide aminé non naturel a également été réalisé et sera présenté section II suivante.

Stratégie pour fournir un glycoconjugué multivalent homogène afin de permettre l'étude de l'influence du ratio sucre greffé/protéine porteuse sur la réponse immunitaire :

L'incorporation d'acides aminés non naturels est une technologie que nous voulions implémenter au sein du laboratoire et la rareté de leur utilisation lors de la mise au point de vaccins glycoconjugués malgré leur fort potentiel a permis l'élaboration d'un axe original (prochainement publié au sein du journal JoVE). Cependant, la génération de vaccins site-spécifiques homogènes était une question déjà en cours d'étude au laboratoire. Pour cela des mutants cystéines ont été générés et les conjugués ont été réalisés par des réactions de thiol addition (Thèse Aline Pillot, (Pillot et al. 2019)). Ces conjugués ont ensuite été utilisés pour l'immunisation de souris, dont la réponse immunitaire a été analysée par ELISA. Ces conjugués sont monovalents, c'est à dire qu'il y a un seul tétrasaccharide lié à une protéine. La réponse immunitaire anti-mPsaA s'est révélée variable en fonction de la position de greffage ce qui souligne l'importance de l'objectif précédent.

Ce travail a également soulevé un second problème. En effet, la réponse anti-PnTS14 s'est avérée très faible par rapport aux réponses obtenues lors de l'étude des conjugués générés de façon aléatoire au cours de la thèse de Maruthi Prasanna. Ces résultats obtenus lors des travaux des thèses de Maruthi Prasanna et Aline Pillot, sont résumés dans le Tableau 15 ci-dessous (Prasanna et al. 2019; Pillot et al. 2019). Bien que nous ne puissions pas négliger l'impact de la position de greffage et des différents paramètres composant le vaccin (dose injectée, nombre des injections, adjuvant), nous pouvons remarquer que la répartition de l'intensité de la réponse immunitaire entre le tétrasaccharide PnTS14 et la mPsaA semble être corrélée avec le rapport saccharide/protéine. En effet, la réponse immunitaire semble plus favorablement dirigée contre la mPsaA lorsque le conjugué comprend environ un tétrasaccharide PnTS14 pour une protéine. A l'inverse, la réponse est plus forte contre le tétrasaccharide lorsque le rapport est de 5,4 tétrasaccharides pour une protéine.

Nous avons alors développé une méthode permettant la génération d'un glycoconjugué site-spécifique trivalent entre la mPsaA et le tétrasaccharide PnTS14 à l'aide de l'incorporation multiple de Propargyl-Lysine. Ce travail est détaillé dans la section III.

Tableau 15: Résumé de précédents travaux vaccinaux réalisés au sein du laboratoire

Type de conjugué (ratio PnTS14/mPsaA)	Réponse anti-PnTS14	Réponse anti-mPsaA	Référence
Aléatoire (5,4/1)	Forte	Faible	(Prasanna et al. 2019)
Cystéine (1/1)	Faible	Forte	(Pillot et al. 2019)
Propargyl-Lysine (3/1)	Objet de ce projet	Objet de ce projet	En cours

I. Génération de conjugués monovalents site-spécifiques entre la mPsaA et le tétrasaccharide PnTS14 à l'aide de la Propargyl-Lysine.

I.1 Méthode d'incorporation de l'acide aminé non naturel

Afin d'incorporer la PrK au sein de la mPsaA, la méthode d'incorporation d'acides aminés non naturels de manière site spécifique (décrite Chapitre 1, section IV.3) est employée. Cette méthode consiste à utiliser la machinerie de la traduction. Pour rappel, lors de la traduction classique, une aminoacyl ARNt synthetase (aaRS) catalyse la liaison entre un acide aminé et l'ARNt correspondant. Puis cet ARNt chargé est utilisé par le ribosome lors de la traduction afin de synthétiser le polypeptide. Afin d'incorporer un acide aminé non naturel (NAA), une aaRS et un ARNt orthogonaux sont utilisés. Cette aaRS est spécifique pour l'acide aminé non naturel ainsi que pour l'ARNt orthogonal qui possède un anti-codon permettant la suppression du codon stop TAG (Figure 14, Chapitre 1, section IV.3). Afin d'incorporer un acide aminé non naturel de manière site-spécifique, il faut une souche bactérienne d'expression co-transformée avec deux plasmides : un plasmide portant le gène d'expression de la protéine cible avec un codon stop TAG à la position d'incorporation choisie, ainsi qu'un plasmide portant les éléments orthogonaux pour l'incorporation de NAA (l'aaRS et l'ARNt). L'acide aminé non naturel sera alors ajouté au milieu de culture bactérienne et sera intégré à la protéine recombinante durant sa traduction dans la bactérie.

Pour incorporer la Propargyl-lysine, nous avons utilisé le couple pyrrolysyl ARNt synthétase et ARNt issu de l'archée *Methanosarcina mazei*. En effet, cette aaRS sauvage reconnaît spécifiquement la PrK sans mutation supplémentaire au niveau de son site actif. Comme cette aaRS et cet ARNt sont issus d'une archéobactérie, ils sont orthogonaux au système de traduction de l'organisme de production (ici *E.coli*). Cette propriété est importante car elle évite les erreurs d'incorporation dues à une réactivité croisée entre l'aaRS orthogonale et des ARNt endogènes ou entre l'ARNt orthogonal et des aaRS endogènes. De plus l'ARNt possède un anticodon complémentaire du codon stop TAG, un codon non utilisé chez la bactérie *E. coli*.

I.2 Synthèse de la Propargyl-Lysine

L'acide aminé non naturel utilisé au cours de ce projet est la Propargyl-Lysine (PrK). Cet acide aminé porte une fonction propargyle pouvant réagir spécifiquement dans un milieu complexe avec une fonction azoture par chimie click. Afin de l'incorporer au sein de la protéine mPsaA, la PrK a été ajoutée au milieu de culture lors de la production protéique et doit donc être synthétisée en grande quantité. Au cours de cette thèse, environ 10 grammes ont été produits selon une synthèse chimique en deux étapes à partir de la Boc-L-Lysine commerciale (Chapitre 4, section I et Figure 54). La première étape consiste en une attaque nucléophile du groupement amine de la chaîne latérale de la Boc-Lysine sur le chloroformate de propargyle en milieu aqueux basique afin d'y greffer la fonction alcyne. Cette méthode est apparentée à la méthode de Schotten-Baumann qui conduit à la formation d'une fonction à partir d'un chlorure d'acyle et d'une amine primaire ou secondaire. La deuxième étape consiste à déprotéger la fonction amine de la chaîne principale à l'aide de l'acide trifluoroacétique (TFA). Cette déprotection a été suivie par Chromatographie sur Couche Mince (CCM), puis l'identité du produit obtenu par précipitation dans le diéthyléther a été vérifiée par RMN (Figure 55). La déprotection est visible sur CCM car l'acide aminé non naturel déprotégé est plus polaire grâce à la déprotection de la fonction amine et donc son coefficient de migration est plus faible au sein du solvant. En effet, le coefficient passe de 0,90 à 0,38 par rapport à l'acide aminé protégé dans le mélange Dichlorométhane/Méthanol (en proportion 9/1). L'acide aminé non naturel

est ensuite dissout à 100mM dans de l'eau distillée et la solution obtenue est stockée en aliquots de 1mL en vue de son utilisation ultérieure.

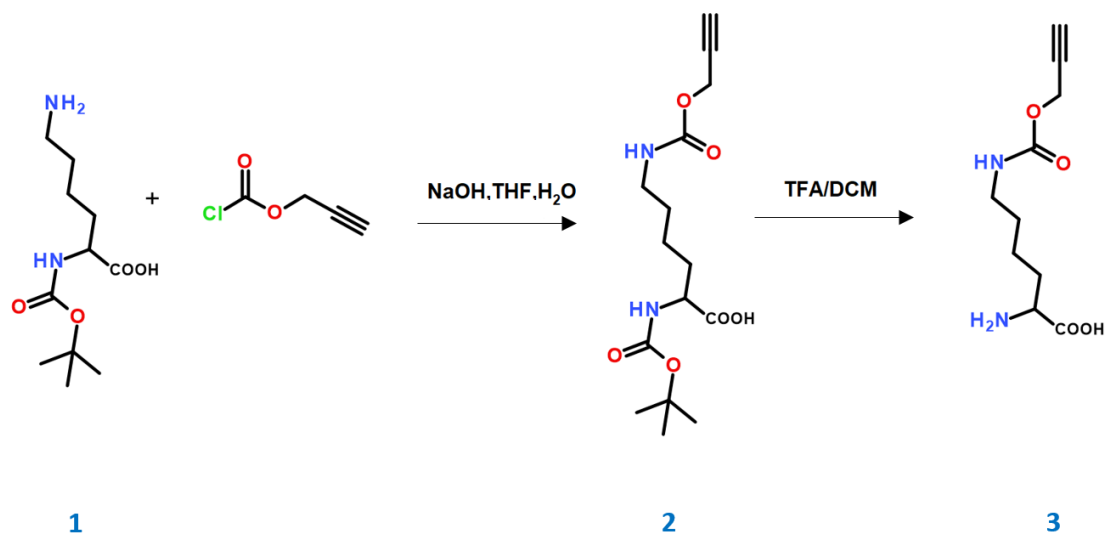


Figure 54 : Synthèse de l'acide aminé non naturel : la Propargyl-Lysine

500mg (2.03mmol) de Boc-Lysine commerciale a été dissout dans 5mL de NaOH (1M) et 5mL de THF, puis la solution a été refroidie. 158 μ L (1.62mmol) de propargyl-chloroformate est ajouté goutte à goutte à la solution, puis la solution est incubée à température ambiante avec agitation pendant environ 10h. Après lavage avec de l'éther froid, le produit intermédiaire 2 est extrait avec de l'acétate d'éthyle. Puis les solvants sont évaporés. Le rendement de cette première étape est d'environ 80%. Le composé intermédiaire 2 est ensuite dissout dans 4mL de Dichlorométhane et 4mL de TFA est ajouté goutte à goutte. La solution est incubée environ 2h à température ambiante, puis les solvants sont évaporés et le produit final (composé 3) est précipité à l'aide de diéthyl-ether. Le rendement de cette deuxième étape est d'environ 90%.

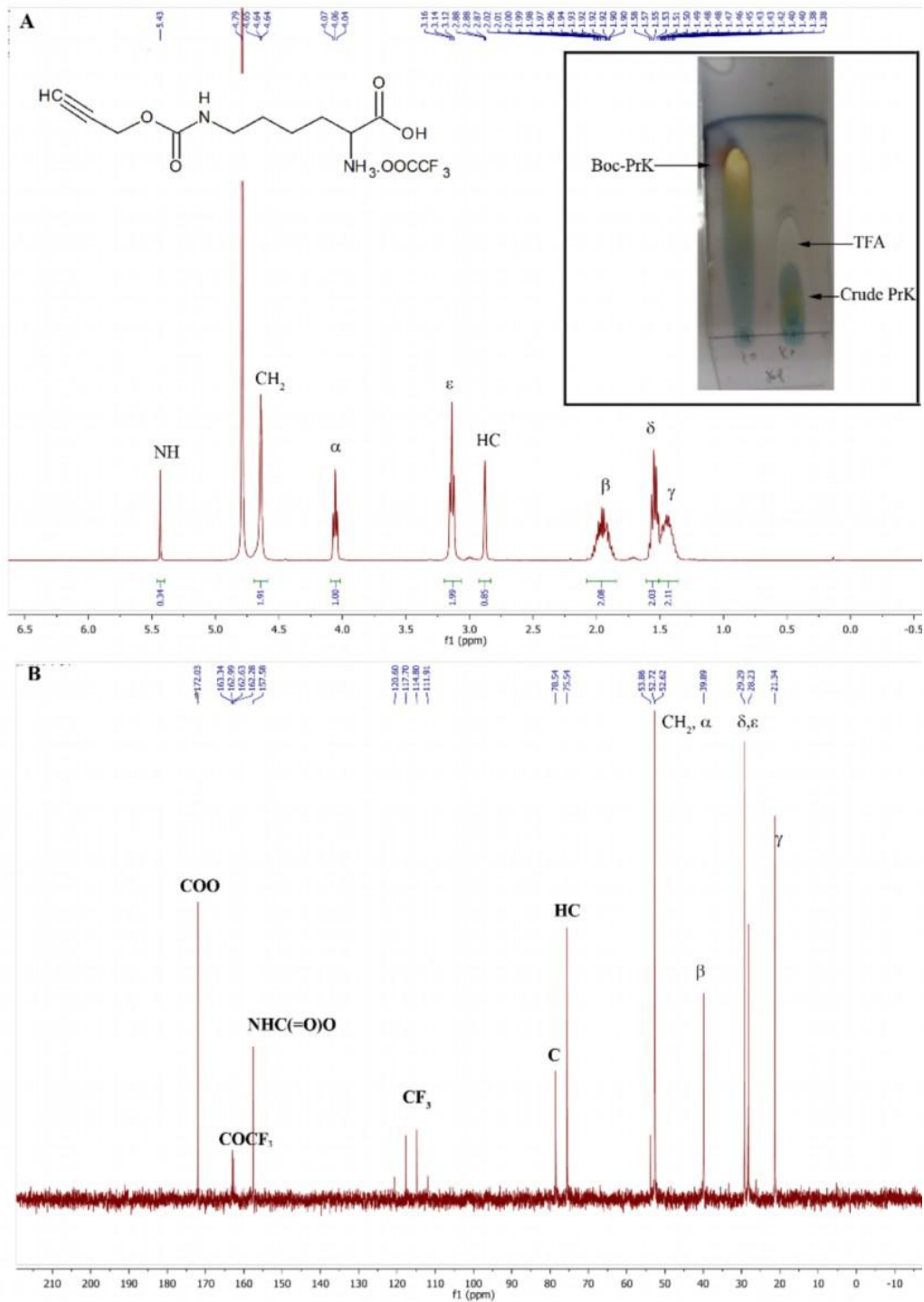


Figure 55: RMN et CCM obtenue lors de la synthèse de la Propargyl-Lysine

-Insert: Suivi de la déprotection de la Boc-L-Lysine (prop-2-nyloxycarbonyl)-OH par Chromatographie sur couche mince (sur plaque de silicium de 0,25mm, dont la migration a été réalisée dans un mélange Dichlorométhane/Méthanol (9:1) et révélée par de la vaniline (dans un mélange acide sulfurique/éthanol 1.5:95 v/v)). Les spots sont les suivants : à gauche la Boc-L-Lys(prop-2-nyloxycarbonyl)-OH (Rf 0.90), et à droite la propargyl-lysine (Rf 0.38).

-Spectres RMN 400 MHz 1H (A) et 13C (B) de la propargyl-lysine dans du D₂O.

I.3 Choix des sites d'incorporation de la Propargyl-Lysine et mutagénèse de la mPsaA

Malgré l'importance de ce paramètre, l'impact de la position du greffage de l'haptène sur une protéine porteuse immunogène n'a que très rarement été étudié. Au cours de cette thèse, nous allons créer un vaccin glycoconjugué homogène. Le saccharide peut alors être conjugué au sein, à proximité ou éloigné d'un épitope B ou T. Une présentation du tétrasaccharide à proximité d'un épitope pourrait soit potentialiser la réponse immunitaire, soit masquer des épitopes importants.

Les épitopes B sont les épitopes linéaires ou structuraux contre lesquels des anticorps sont développés. Tandis que les épitopes T sont des portions de séquences présentés au niveau du CMHII aux lymphocytes T qui vont induire une amplification de la réponse. Ces deux types d'épitopes ont été identifiés sur la protéine mPsaA (Figure 56). Les épitopes B ont été déterminés grâce à des études réalisées à l'aide d'anticorps monoclonaux ou par phage display en utilisant des bibliothèques de peptides (Srivastava et al. 2000; Johnson et al. 2002). Quant aux épitopes T, ils ont été prédits *in silico* puis vérifiés expérimentalement chez la souris (Singh et al. 2013). Les épitopes B sont représentés en rose et les épitopes T en rouge sur la Figure 56.

Le choix des positions à tester afin de générer des vaccins glycoconjugués monovalents site-spécifique s'est porté sur trois lysines : K24, K213 et K229 (Figure 56). La position K24 est située loin des épitopes de la protéine, alors que les positions K213 et K229 sont situées à l'intérieur d'épitopes T mais en dehors des épitopes B. En effet, leur présence au niveau des épitopes B pourrait empêcher la reconnaissance de la mPsaA par des anticorps. Par contre, nous supposons que leur présence à proximité d'épitopes T peut favoriser l'amplification et la pérennisation de la réponse anti-sucre comme cela a pu être démontré par Avci *et al.* (Chapitre 2, section II.2.5.2, (Avci et al. 2011)). De plus ces positions sont toutes des lysines. En effet, la méthode aléatoire qui a été employée afin de générer le conjugué lors du travail réalisé par Maruthi Prasanna a utilisé la réactivité des lysines, et la méthode ciblée utilisant des mutations en cystéines a également été réalisée à des positions correspondant à des lysines dans la protéine sauvage. Effectuer le greffage du tétrasaccharide à la position d'une lysine devrait donc faciliter la comparaison de nos résultats avec les résultats issus de travaux parallèles.

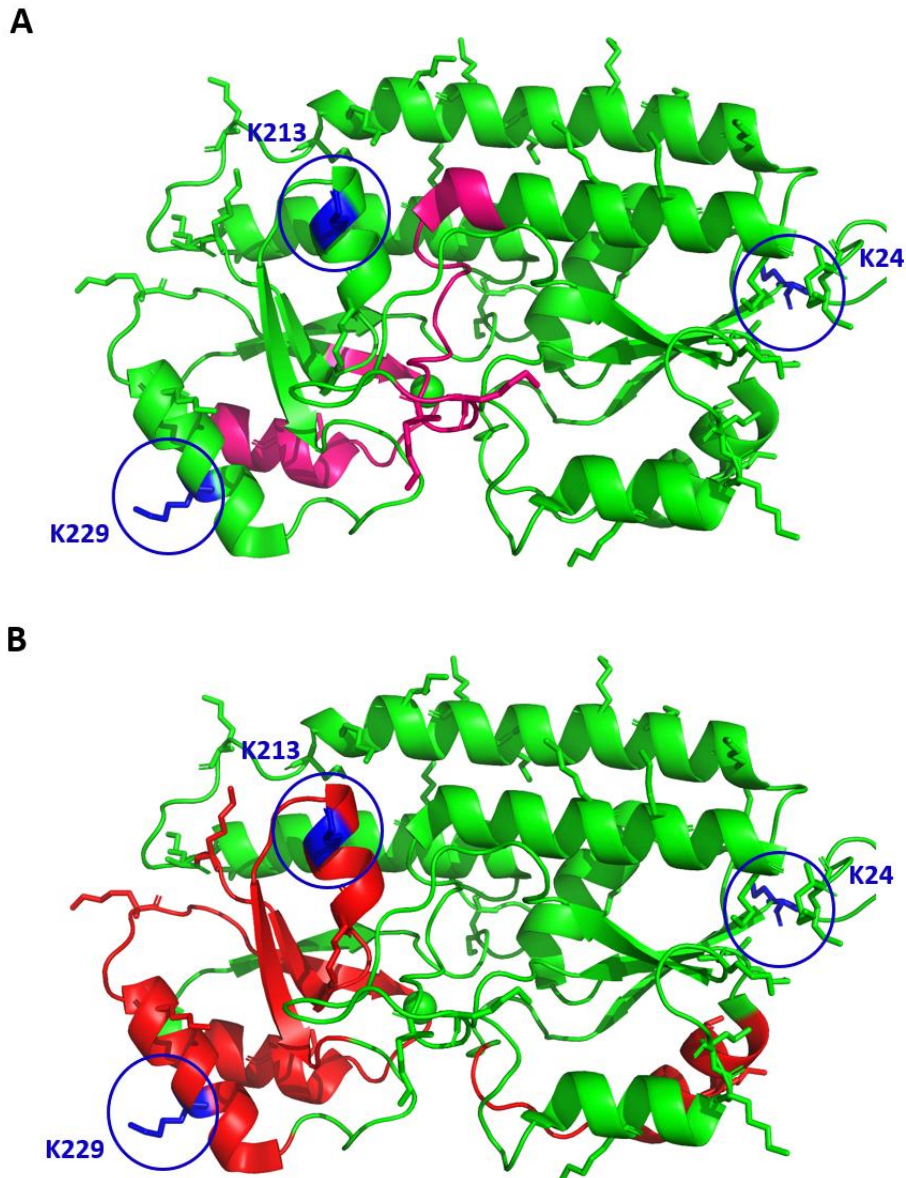


Figure 56: Choix des positions d'incorporation de la PrK au sein de la mPsaA (PDB: 1PSZ)

La protéine est représentée en cartoon, les chaînes latérales des lysines ont été représentées en sticks. (A) Les zones immunogènes ont été colorées en rose pour les épitopes B et (B) en rouge pour les épitopes T helper. Les positions choisies pour incorporer la PrK ont été colorées en bleu et entourées par un cercle bleu.

L'incorporation d'acides aminés non naturels de façon site-spécifique nécessite la mutation du gène de la protéine cible afin de générer un codon stop TAG à la position où il sera incorporé (rappel de la méthode d'incorporation section I.1). Ces mutations ont donc été générées par mutagenèse dirigée (méthode décrite Chapitre 4, section III). L'expression de ces protéines au sein d'*E. coli* aboutit alors à la production de protéines tronquées (Figure 57). En effet, en l'absence de PrK ou de la machinerie orthogonale permettant son incorporation (aaRS et ARNt), la traduction s'interrompt au codon stop TAG. Comme nous pouvons le voir sur la Figure 57, il y a donc production des protéines tronquées K213 et K229 d'environ 21,5kDa et 24,2kDa respectivement (flèche verte et rouge respectivement) au lieu de 34,14kDa pour la protéine sauvage (flèche bleue). La protéine tronquée K24 n'est pas visible, mais cette protéine est très petite (environ 2,5kDa) et est probablement dégradée chez *E. coli*.

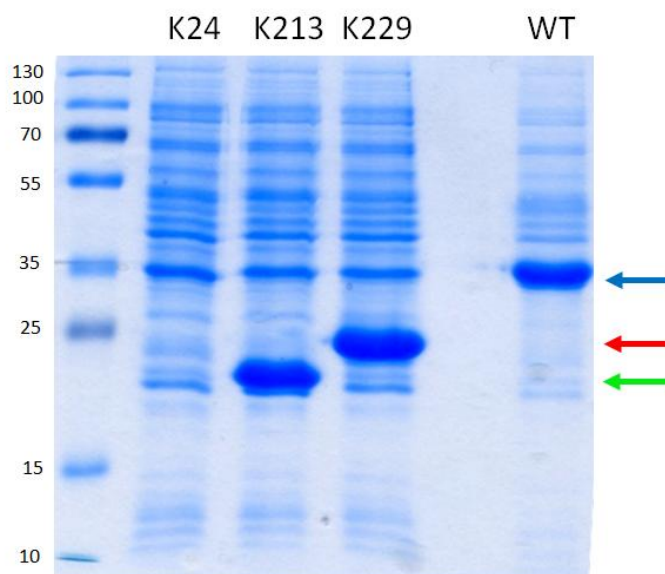


Figure 57: Formes tronquées de la Pneumococcal surface adhesine A mature (mPsaA)
Analyse SDS-PAGE (gel d'acrylamide 12%) d'extraits totaux de cultures bactériennes BL21(DE3) transformés avec les plasmides d'expression pET11a contenant les gènes tronqués de la mPsaA en positions K24, K213 et K229 ou le gène de la protéine sauvage mature (sans le peptide d'adressage) mPsaA-WT.

I.4 Vérification de l'entrée de la Propargyl-Lysine dans la bactérie

La Propargyl-Lysine est un acide aminé non naturel dont le mécanisme d'entrée dans la cellule n'est pas connu. Comme les premiers essais d'incorporation de la Propargyl-Lysine n'étaient pas fructueux (données non montrées), nous avons voulu vérifier qu'elle pouvait bien pénétrer à l'intérieur des bactéries lors des conditions utilisées pour l'expression de protéine. Pour cela, nous avons réalisé une culture bactérienne sur 24h dans les conditions d'expression sans ou à une concentration de 10mM de Propargyl-Lysine. Après lavage et lyse des cellules, nous avons analysé 1µL des lysats bactériens par chromatographie sur couche mince (CCM) en le comparant à de la PrK pure à 10 et 1mM. (méthode décrite Chapitre 4 section II). Comme nous pouvons le voir sur la Figure 58, lorsque la culture est réalisée avec la Propargyl-Lysine, il y a apparition d'une tâche orange (indiquée par une flèche rouge) qui correspond à la PrK. Cette tâche n'est pas présente au niveau de la culture bactérienne sans PrK. Comme les bactéries ont été lavées plusieurs fois afin d'éviter toute contamination avec la PrK du milieu avant la lyse des cellules, cela confirme l'entrée de la Propargyl-Lysine au sein des bactéries. Une analyse de l'image par le logiciel Image Studio Lite a permis d'évaluer approximativement l'intensité de la tâche (indiquée par la flèche rouge Figure 58). Au niveau du lysat de culture, elle correspond à environ 34% de la tâche correspondant à la tâche de la PrK à 1mM, soit environ 2% de la PrK mise dans le milieu de culture. De plus, nous avons également réalisé cette expérience avec un temps d'incubation d'environ 8h, et nous avons pu noter l'absence de cette tâche correspondant à la Propargyl-Lysine. A une concentration de 10mM, la PrK semble donc pénétrer au sein des BL21(DE3). Cependant, il est possible que ce transport soit lent.

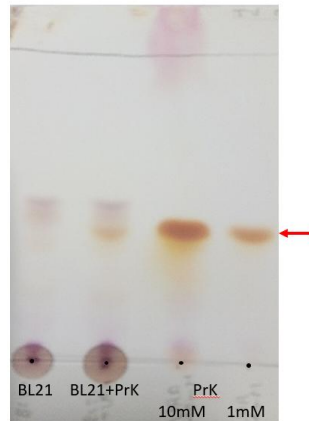


Figure 58: Vérification de l'entrée de la Propargyl-Lysine dans les bactéries par Chromatographie sur Couche mince (CCM) Une culture de 5mL de BL21(DE3) a été réalisée en présence ou non de 10mM de PrK durant 24h. La culture a ensuite été centrifugée à 5000g pendant 10min. Le culot a été lavé 3 fois avec 5mL de tampon phosphate froid, puis resuspendu dans 300µL avant d'être lysé par sonication. Un microlitre de ce lysat a alors été déposé sur CCM, ainsi qu'1µL d'une solution de Propargyl-Lysine à 10mM et 1mM. La plaque CCM a ensuite été migrée et révélée (solvant d'éluion : Butanol/acide acétique/eau en proportions (10/2.5/2.5) et révélateur : ninhydrine).

1.5 Production des mutants mPsaA-PrK au sein du plasmide pET11a

Afin d'incorporer la Propargyl-Lysine au sein de la mPsaA, les plasmides pET11a-mPsaA portant les différentes mutations K24TAG, K213TAG, K229TAG et pEVOL ont été co-transformés au sein de la souche bactérienne d'expression (BL21(DE3)).

Le plasmide pET11a porte donc le gène d'expression de la mPsaA avec les mutations TAG nécessaires pour l'incorporation de la PrK aux positions choisies (K24TAG, ou K213TAG, ou K229TAG), ainsi qu'une étiquette histidine en N-terminal suivi d'un site de coupure de la protéase TEV. Les protéines produites peuvent alors être détectées par WesternBlot à l'aide d'un anticorps dirigé contre l'étiquette histidine. Comme cette étiquette est située en N-terminal de la protéine cible, cela permet alors la détection des formes tronquées n'ayant pas incorporé la PrK et des protéines complètes ayant incorporé l'acide aminé non naturel. Cette construction permet donc l'analyse de la production de la protéine cible ainsi qu'une évaluation qualitative du rendement d'incorporation. Ce plasmide d'expression est présenté en annexe 2 et 3.

Le plasmide pEVOL porte les gènes d'expression de la pyrrolysyl-ARNt synthétase et de l'ARNt_{CUA} issus de l'archée *Methanosarcina mazei*. Le plasmide pEVOL porte deux copies de la synthétase, une sous le contrôle d'un promoteur constitutif (GlnS) et l'autre sous le contrôle d'un promoteur inductible à l'arabinose (Chapitre 1, section IV.4.3). La copie de l'ARNt_{CUA} est sous le contrôle d'un promoteur constitutif (ProK). Ce plasmide a été décrit dans la publication (T. S. Young et al. 2010) et est présenté en annexe 2 et 3.

Afin de vérifier l'incorporation spécifique de la PrK au sein de la protéine, des cultures ont été réalisées sans PrK ou avec 1mM de PrK ainsi que sans arabinose. Des extraits bruts de ces cultures ont ensuite été analysés par WesternBlot. Comme nous pouvons voir sur la Figure 59, l'absence de bande de la taille de la protéine sauvage dans ces conditions confirme que la synthétase n'incorpore pas d'autre acide aminé que la PrK au niveau du codon stop TAG et qu'une autre synthétase que celle de *M. mazei* ne charge pas l'ARNt orthogonal avec un acide aminé naturel. De plus, en présence d'arabinose et en absence de PrK, il y a arrêt de la traduction au niveau du codon stop TAG, nous pouvons donc voir les protéines tronquées K213 et K229 (flèche rouge Figure 59). Les protéines complètes ayant la même taille que la protéine sauvage ne sont exprimées qu'en présence d'arabinose et de PrK, ce qui est un signe de l'incorporation spécifique de cette dernière (flèche bleue Figure 59). Par contre, nous pouvons voir que des protéines tronquées sont toujours exprimées

même en présence de PrK (flèche rouge Figure 59). Ces protéines peuvent s'expliquer par la présence d'un facteur d'arrêt de la traduction RF1 au sein de la souche d'expression. Au sein de la bactérie, RF1 reconnaît les codons stop UAA et UAG. (Chapitre 1, section IV.4.2.2). Lorsqu'il y a incorporation d'un acide aminé non naturel, il y a donc une compétition entre l'ARNt_{CUA} qui porte la PrK et le facteur de terminaison RF1 qui reconnaît le codon UAG en tant que codon stop. De plus, cet arrêt prématuré de la traduction peut également s'expliquer par l'efficacité catalytique plus faible de l'aaRS orthogonale comparativement aux aaRS endogènes.

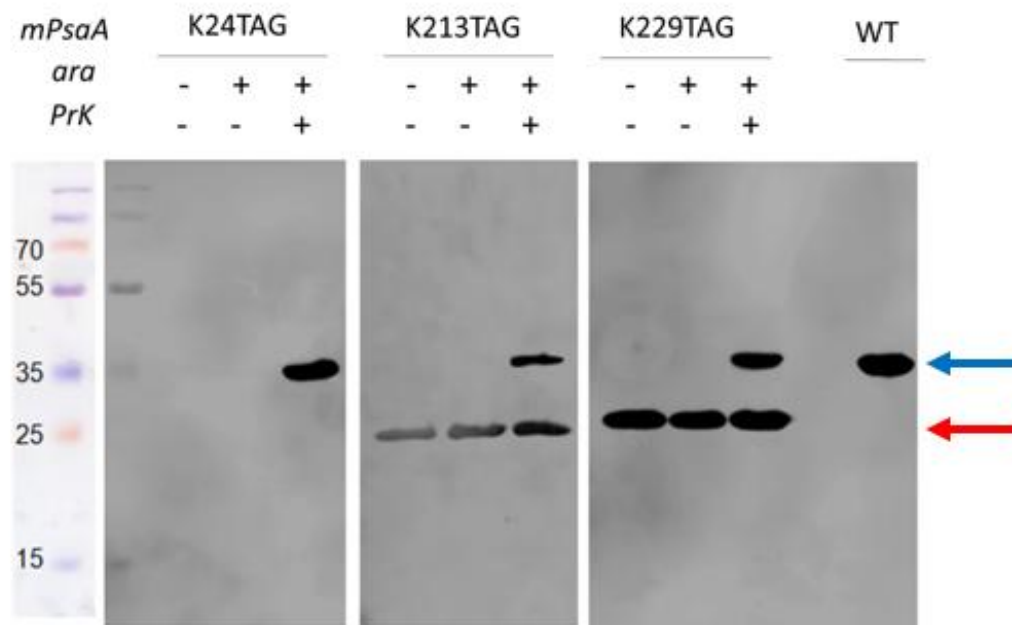


Figure 59: Vérification par WesternBlot de la production des différents mutants *mPsaA* contenant la Propargyl-Lysine. Analyse par Westernblot d'extraits bruts de production des différents mutants de la *mPsaA* en présence ou en absence de PrK et d'arabinose. La protéine sauvage et les protéines complètes ayant incorporé la PrK sont indiquées par une flèche bleu, tandis que les protéines tronquées sont indiquées par une flèche rouge.

Afin de produire les différents vaccins glycoconjugués, il faut purifier les protéines ayant incorporé la PrK. Cependant, lorsque l'étiquette Histidine est située en N-terminal comme au sein du plasmide pET11a, la purification sur colonne de nickel conduira à un mélange de protéines tronquées et de protéines complètes. La différence de taille n'est pas assez importante pour permettre de les séparer sur la colonne Superdex de chromatographie d'exclusion stérique disponible au laboratoire (donnée non montrées). Afin de les séparer efficacement, nous avons reclone le gène de la *mPsaA* sauvage et des formes mutées avec l'étiquette histidine en C-terminal au sein du plasmide pET24d. En effet, lorsque cette étiquette histidine est en C-terminal, seules les protéines complètes ayant incorporé la PrK seront alors purifiées par chromatographie d'affinité sur colonne de nickel. En effet, la traduction des protéines tronquées s'est arrêtée avant la traduction de l'étiquette histidine.

1.6 Clonage des mutants *mPsaA*-TAG au sein du plasmide pET24, production et purification des protéines recombinantes

Le gène de la *mPsaA* sauvage et les gènes présentant les mutations K24TAG, K213TAG et K229TAG ont donc été recloneés au sein du plasmide pET24d avec une étiquette histidine en C-terminal et un site de coupure pour la protéase TEV. Comme présenté dans la Figure 60 ci-dessous, un mélange de protéines tronquées (flèches rouges Figure 60) et de protéines complètes (flèche bleue Figure 60) est observé comme pour la construction précédente. Cependant, seules les

protéines complètes contenant la PrK possèdent l'étiquette histidine. Ces protéines complètes peuvent donc être purifiées par chromatographie d'affinité sur des billes de nickel comme nous pouvons le voir sur la Figure 61 avec la purification du mutant K213 malgré la présence abondante de la protéine tronquée.

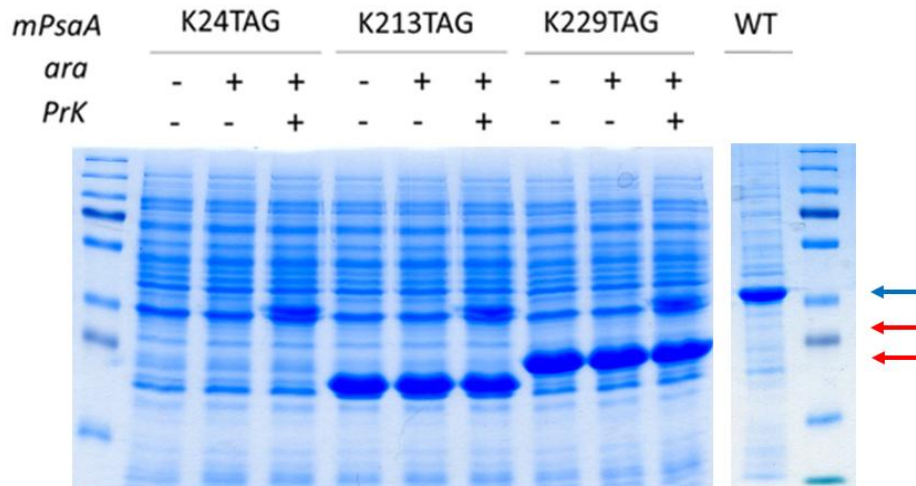


Figure 60 : Vérification par SDS-PAGE de la production des différents mutants *mPsaA* contenant la Propargyl-Lysine. Analyse SDS-PAGE (gel d'acrylamide 12%) d'extraits bruts de production (fraction soluble) des différents mutants de la *mPsaA* en présence ou en absence de *PrK* et d'arabinose. La protéine sauvage et les protéines complètes ayant incorporées la *PrK* sont indiquées par une flèche bleue, tandis que les protéines tronquées sont indiquées par des flèches rouges.

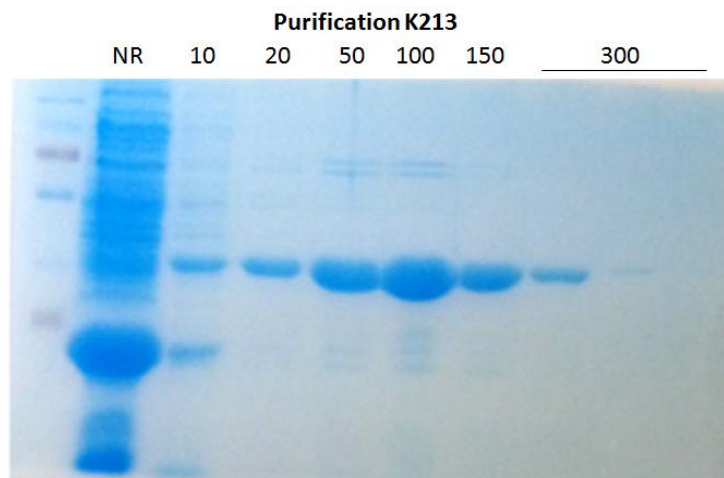


Figure 61 : Purification par chromatographie d'affinité sur colonne de nickel du mutant *mPsaA*-K213. Analyse SDS-PAGE (gel acrylamide 12%) de 15µL des différentes fractions de purification du mutant *mPsaA*-K213PrK ayant incorporé la *PrK*. Ce mutant a été produit au sein du plasmide pET24d portant l'étiquette histidine en C-terminal. La purification a été effectuée grâce à des billes de Nickel et en utilisant du tampon phosphate contenant une quantité croissante d'imidazole. La fraction non retenue est notée NR, puis la concentration en mM d'imidazole des différentes fractions sont indiquées.

1.7 Conjugaisons par chimie click (CuAAC)

La protéine *mPsaA* contenant une *PrK* à différentes positions a ensuite été conjuguée avec le tétrasaccharide par chimie click grâce aux propriétés de la *PrK*. La chimie click utilisée est la CuAAC (*Copper(I)-catalyzed Alkyne Azide Cycloaddition*). Cette réaction a lieu entre une fonction alcyne et une fonction azoture, elle est catalysée par du Cuivre I et peut être réalisée en milieu aqueux complexe. De plus, elle est orthogonale et donc très spécifique. Cependant, cette réaction est

sensible à la température, elle sera donc plus efficace si elle est réalisée à température ambiante plutôt qu'à 4°C.

Nous avons tout d'abord voulu tester la résistance de notre protéine aux différents réactifs de conjugaison et à une incubation à température ambiante. Pour cela, nous avons réalisé des mesures de dichroïsme circulaire de la protéine ayant été incubée selon les différentes durées de conjugaison (1h, 1h30). Le dichroïsme circulaire est la mesure de l'absorption d'une lumière polarisée. En effet, les protéines possèdent des structures asymétriques qui vont l'absorber différemment (hélices α , feuillets β) et si la protéine est dénaturée le signal d'absorption sera plus faible. Comme nous pouvons le voir sur la Figure 62, le signal n'a pas changé d'intensité malgré l'exposition de la protéine aux différents réactifs de chimie click et à son incubation à température ambiante. La protéine ne semble donc pas se dénaturer et ces conditions de conjugaison peuvent être utilisées.

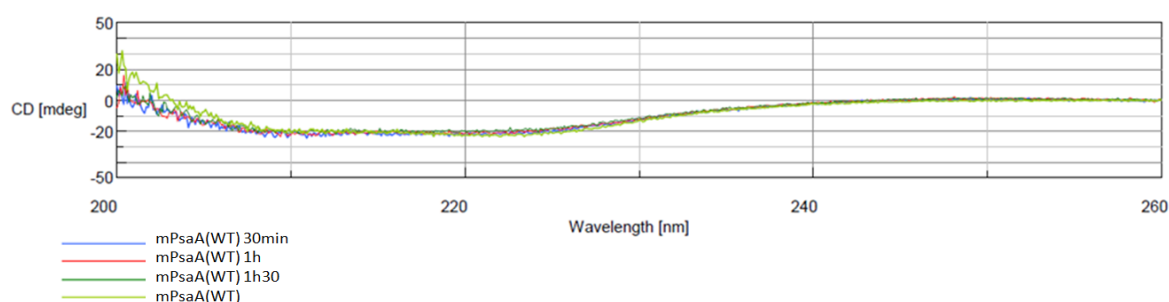


Figure 62 : Analyse par Dichroïsme circulaire de la résistance de la protéine mPsaA aux réactifs de conjugaison. La mPsaA a été mis au contact des différents réactifs de conjugaison pendant 30min, 1h ou 1h30. Un contrôle négatif de la mPsaA sans les réactifs de conjugaison a également été réalisé. Puis l'intégrité de la structure de la mPsaA a été vérifiée par dichroïsme circulaire.

La conjugaison a été analysée par SDS-PAGE. En effet la conjugaison entraîne une augmentation de la taille de la protéine de 776 Da. Cependant, deux bandes ont été observées reflétant une conjugaison incomplète (données non montrées). Le temps d'incubation a donc été augmenté à 10h-14h (sur la nuit). De plus, lors du procédé final, la quantité de cuivre a été augmentée (section I.7). Ces conditions ont également été vérifiées en dichroïsme circulaire (Figure 63). Le signal est relativement stable, une différence 15% du signal entre la protéine dans les conditions de conjugaison par rapport à la protéine contrôle (signal de -17 au lieu de -20) a été observée à 220 nm, ce qui peut correspondre à une dénaturation partielle des protéines. Cet écart semble cependant assez minime et considéré comme acceptable car ces conditions sont nécessaires pour la conjugaison complète de nos protéines, elles ont donc été conservées.

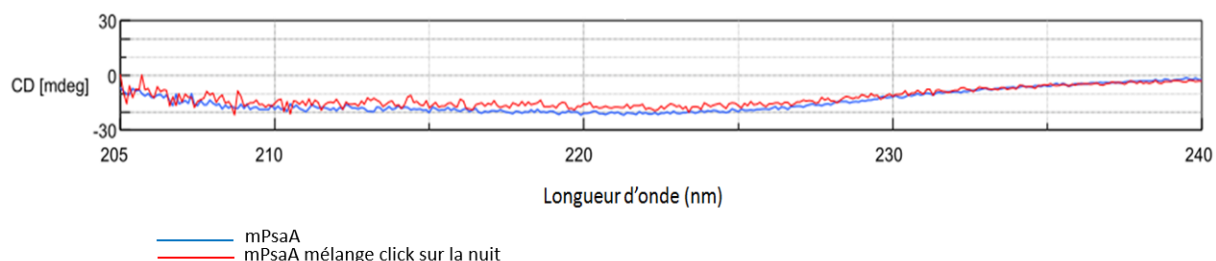


Figure 63 : Analyse par Dichroïsme circulaire de la résistance de la protéine mPsaA aux réactifs de conjugaison sur la nuit. La mPsaA a été mise au contact des différents réactifs de conjugaison pendant 10-14h (sur la nuit) avec une quantité de cuivre augmentée. Un contrôle négatif de la mPsaA sans les réactifs de conjugaison a également été réalisé. Puis l'intégrité de la structure de la mPsaA a été vérifiée par dichroïsme circulaire.

La conjugaison entre la protéine mPsaA et le tétrasaccharide est difficile à analyser. En effet, le tétrasaccharide est petit (776Da) et le léger changement de taille de la protéine après conjugaison de ce composé est donc difficile à observer sur gel d'acrylamide. Dans un premier temps, l'accessibilité de la fonction alcyne de la protéine a donc été vérifiée grâce à la conjugaison de la mPsaA avec une hexachloro-fluorescéine-azide disponible commercialement (Figure 64). Ce composé étant fluorescent, sa conjugaison avec la protéine va donc la rendre visible sur gel d'acrylamide illuminé par des rayons UV.

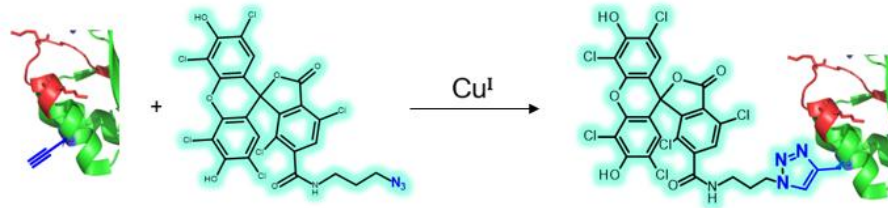


Figure 64 : Schéma représentant la réaction de conjugaison CuAAC entre la protéine et la fluorescéine. La fluorescéine porte une fonction azoture, ce qui permet de la conjuguer à la protéine ayant incorporé la Propargyl-Lysine (qui porte une fonction alcyne). La protéine conjuguée est donc fluorescente et peut donc être visualisée sur gel SDS-PAGE sous une lumière UV (max absorption : 535nm, max émission : 556nm).

L'expérience de conjugaison a été réalisée avec la protéine mPsaA-K24PrK et avec la mPsaA sauvage en tant que contrôle négatif. L'histidine-tag a préalablement été clivé à l'aide de la TEV protéase. En effet, il est plus propice pour la réaction de conjugaison par click d'enlever l'étiquette histidine car celle-ci pourrait complexer les ions cuivres ce qui les rendrait moins disponibles pour la catalyse de la réaction de conjugaison. Comme on peut le voir sur la Figure 65, seule la protéine mPsaA-K24PrK est fluorescente après conjugaison avec la fluorescéine. Ce résultat confirme la présence de la propargyl-lysine au sein de la protéine et sa disponibilité pour une réaction de conjugaison.

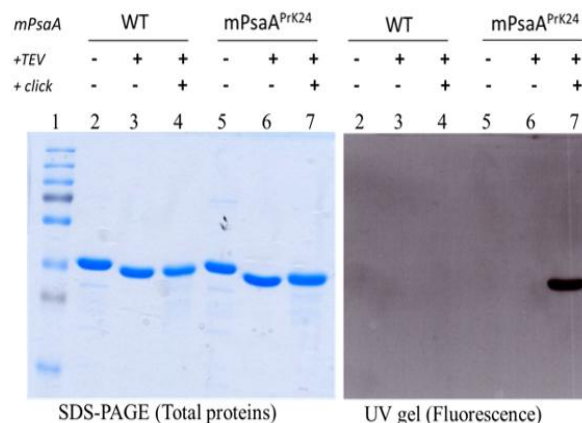


Figure 65 : Conjugaison de la mPsaA WT et K24PrK avec la fluorescéine. La mPsaA sauvage et la protéine mPsaA portant la mutation K24PrK ont été digérées par la TEV protéase avec d'enlever l'étiquette histidine, puis ont été conjuguées avec la fluorescéine. Un échantillon de chacune des étapes a ensuite été déposé sur gel SDS-PAGE (gel d'acrylamide 12%), puis le gel a été visualisé sous une lumière UV (image à droite), puis coloré au bleu de coomassie (image à gauche). La fluorescéine fait 665,09Da.

L'étape de conjugaison ayant été validée par la conjugaison avec la fluorescéine, une expérience de conjugaison avec le tétrasaccharide a ensuite été réalisée Figure 66. Comme le tétrasaccharide est petit (776Da), la différence de taille entre la protéine WT ou K24PrK ayant été conjuguée avec celui-ci, n'est pas très visible (puits 4 et 7 dans la Figure 66). Afin de pouvoir mieux

l'analyser, nous avons fait co-migrer les échantillons avant et après conjugaison de la protéine WT et de la protéine K24PrK (puits (3+4) et (6+7) dans la Figure 66). Nous pouvons alors voir une seule bande pour la protéine WT et une double bande pour la protéine K24PrK, ce qui semble bien confirmer la conjugaison spécifique du tétrasaccharide sur la PrK.

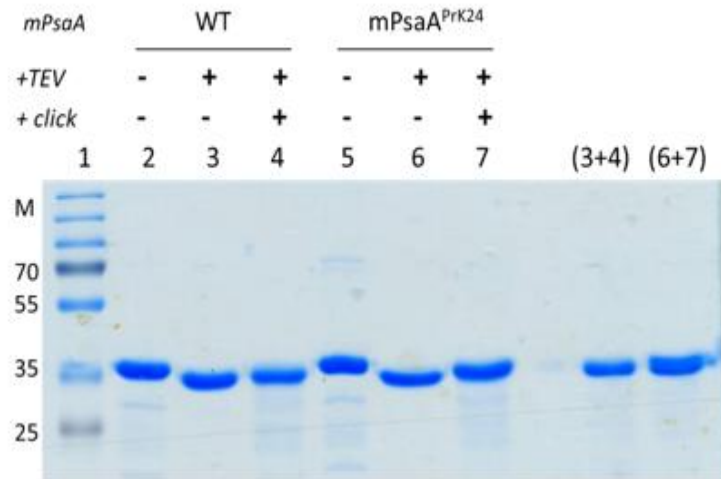


Figure 66: Conjugaison par chimie click de la mPsaA (WT et K24) avec le tétrasaccharide.

La mPsaA sauvage et la protéine mPsaA portant la mutation K24PrK ont été digérées par la TEV protéase avec d'enlever l'étiquette histidine, puis ont été conjuguées avec le tétrasaccharide. Un échantillon de chacune des étapes a ensuite été déposé sur gel SDS-PAGE (gel d'acrylamide 12%), puis le gel a été coloré au bleu de coomassie. Le tétrasaccharide fait 776Da.

La réaction de conjugaison est ensuite stoppée par ajout d'EDTA en excès suivi d'une chromatographie d'exclusion stérique (Chapitre 4, section III.7). Cette étape permet alors de séparer la protéine conjuguée du cuivre et des réactifs de conjugaison qui pourraient dénaturer la protéine. De plus, cela permet aussi de séparer le conjugué du tétrasaccharide libre. En effet, celui-ci est également ajouté en fort excès lors de la conjugaison. Afin de valider cette étape, la protéine mPsaA-K24PrK ayant été conjuguée avec le tétrasaccharide a également été analysée par spectrométrie de masse. La masse trouvée est alors de 34 088Da pour une masse calculée de 34 091Da. Ce résultat confirme alors la conjugaison car la faible différence de la masse (3 Da) est due à la variabilité de la mesure.

Le procédé de production semble alors validé, cependant pour permettre les immunisations des souris après l'ensemble des étapes de purification, il est nécessaire de produire environ 2mg de chaque conjugué. Or la digestion par la TEV protéase de la mPsaA ayant l'étiquette histidine en C-terminal s'est avérée très inefficace. En effet, la quasi-totalité de la protéine n'est pas clivée. Et ce alors que la protéine mPsaA possédant l'étiquette histidine en N-terminal a pu être efficacement digérée dans une étude précédente (Prasanna et al. 2019). Il est possible que la structure 3D de la protéine ne favorise pas l'accès de la séquence TEV par la protéase lorsque celle-ci est positionné en C-terminal. Afin de résoudre ce problème, plusieurs conditions de digestion ont été testées avec différents détergents (Triton, CHAPS, Tween20, Tween80). Cependant, aucune de ces digestions n'ont permis d'augmenter l'efficacité de clivage de la séquence TEV (données non montrées).

Afin de produire nos conjugués, nous avons alors décidé d'utiliser notre construction ayant l'étiquette histidine en N-terminal et de développer un système de purification afin de purifier uniquement les protéines contenant la Propargyl-Lysine et ayant été conjuguées avec le tétrasaccharide.

1.8 Développement d'une méthode de purification par chromatographie d'affinité à l'aide de la Galectine-3

Lorsque la purification de la protéine était réalisée à l'aide d'une étiquette histidine en C-terminal, celle-ci permettait de purifier uniquement les protéines ayant incorporé la Propargyl-Lysine. Or comme nous l'avons vu au-dessus, cette forme de mPsaA pose problème car l'étiquette histidine ne peut pas être clivée efficacement. D'autre part après purification de la protéine avec une étiquette histidine en N-terminal, on obtient un mélange de protéines tronquées et de protéines ayant incorporé la Propargyl-Lysine. Or la conjugaison étant une étape spécifique à la présence de la Propargyl-Lysine, une purification d'affinité sur le tétrasaccharide permettrait d'isoler uniquement nos protéines conjuguées, même lorsqu'elles sont mélangées à des protéines tronquées. Nous avons dans un premier temps étudié cette faisabilité.

Au sein du laboratoire, la galectine-3 est produite dans le cadre d'un projet de recherche sur le développement d'inhibiteurs de galectines (Dion et al. 2017). Notre tétrasaccharide est composé de D-Galp-(1→4)-β-D-Glcp-(1→6)-[β-D-Galp-(1→4)]-β-D-GlcpNAc, et possède donc une unité lactose et lactosamine au sein de celui-ci qui pourraient être reconnues par la galectine 3 (Figure 67).

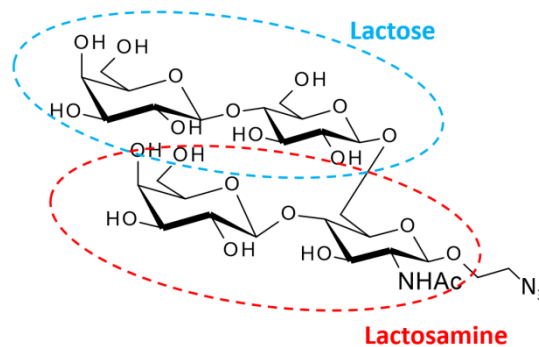


Figure 67 : Structure du tétrasaccharide-N3.
Le tétrasaccharide est composée de D-Galp-(1→4)-β-D-Glcp-(1→6)-[β-D-Galp-(1→4)]-β-D-GlcpNAc.

La technicienne en charge des mesures d'affinités d'inhibiteurs de galectine (Annie Lambert) a réalisé une mesure d'affinité de la galectine-3 avec le tétrasaccharide par polarisation de fluorescence. La constante d'affinité mesurée est d'environ 45 μM. La galectine-3 recombinante est purifiée au laboratoire par chromatographie d'affinité utilisant le lactose et la constante de dissociation pour ce dernier est de 231 μM et de 60μM pour la lactosamine (Diehl et al. 2010). L'affinité de la galectine-3 pour le tétrasaccharide semblait donc suffisante pour réaliser une purification par chromatographie d'affinité.

Lors du procédé initial, l'étiquette histidine est enlevée avant conjugaison. En effet, celui-ci risque de complexer le Cuivre et de le rendre moins disponible pour catalyser la réaction de conjugaison. Cependant, comme nous l'avons vu précédemment lorsque cette étiquette histidine est présente en N-terminal, le procédé de purification sur billes de Nickel entraîne la co-purification des protéines tronquées et des protéines complètes. Leur co-digestion entraînerait la consommation d'une grande quantité de TEV protéase. En effet, afin de tester l'efficacité biologique du vaccin conjugué celui-ci sera à terme injecté au sein de souris. Nous avons donc besoin de grande quantité de protéine ayant subi tout le procédé (environ 2mg par conjugué) et la proportion relative de ces protéines n'est pas en faveur des protéines ayant incorporé la Propargyl-Lysine. La production de la TEV protéase est réalisée au sein du laboratoire mais sa production à grande échelle est fastidieuse. Nous avons donc décidé d'essayer de réaliser la réaction de conjugaison malgré la présence de

l'étiquette histidine en utilisant un gradient de cuivre. En effet, augmenter sa concentration permet d'avoir une concentration de cuivre disponible suffisante pour catalyser la réaction même si une partie de cuivre est complexée par l'étiquette histidine.

Comme nous pouvons le voir sur la Figure 68, la réaction de conjugaison entre la fluorescéine et le mélange protéines tronquées et protéines complètes semble efficace. L'intensité du signal varie cependant entre 0,1 et 0,25 mM de cuivre mais reste stable entre les conditions avec 0,25 mM et 0,5 mM de cuivre. Afin de s'assurer d'une réaction complète, la réaction de conjugaison aura donc lieu en présence de 0,25 mM de cuivre.

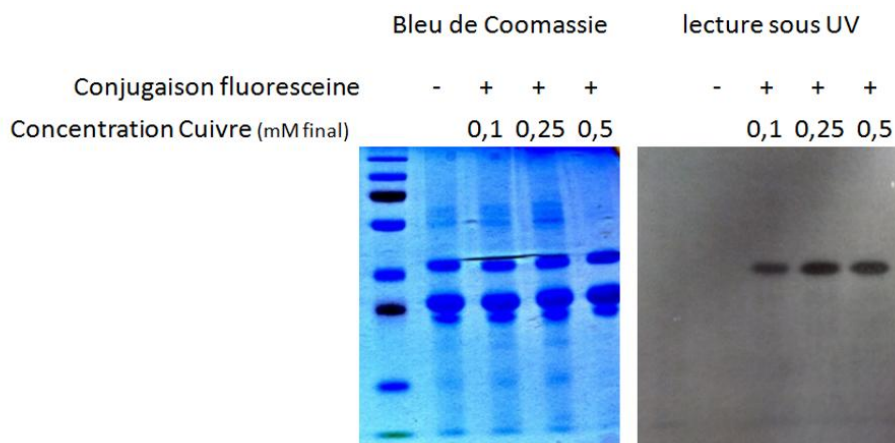


Figure 68 : Conjugaison du mutant mPsaAK24PrK en présence d'un gradient de concentration en cuivre. La protéine mPsaA portant la mutation K24PrK et l'étiquette histidine en N-terminal a été préalablement purifiée sur billes de Nickel. La purification aboutie à un mélange de protéines tronquées et ayant incorporé la PrK ainsi que de protéines contaminantes. La réaction de conjugaison a été réalisée avec la fluorescéine sur ce mélange avec différentes concentrations de Cuivre (0,1, 0,25 et 0,5mM final). Des échantillons de cette conjugaison ont ensuite été déposés sur gel SDS-PAGE (gel d'acrylamide 12%). Celui-ci a été visualisé sous lumière UV (image à droite) afin de visualiser les protéines conjuguées, puis le gel a été coloré au bleu de Coomassie (image à gauche).

Suite à cette vérification, la purification de conjugués mPsaA-tétrasaccharide à l'aide de la galectine-3 est donc apparue comme faisable. La purification se déroule en 4 principales étapes (Figure 69). Tout d'abord, les galectines ont été fixées sur une colonne commerciale portant des groupements NHS qui réagissent avec des amines primaires (présentes à la surface de la protéine, par exemple au niveau des lysines). Puis, la colonne a été incubée avec le mélange réactionnel de conjugaison, les interactions non spécifiques ont été éliminées par des lavages et enfin les conjugués ont été élués à l'aide d'un sucre compétiteur (le lactose).

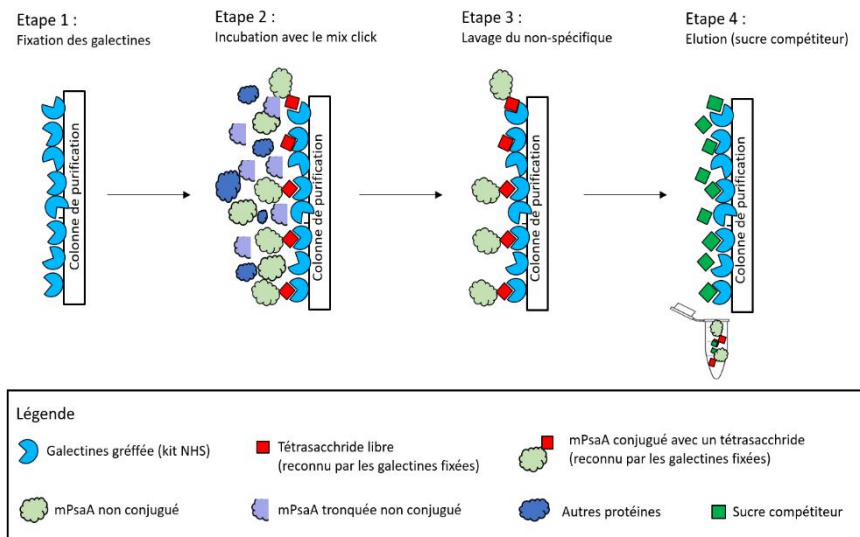


Figure 69: Méthode de purification des conjugués mPsaA-tétrасaccharide à l'aide de galactine-3. La purification se déroule en 4 étapes. Les galectines sont d'abord fixées sur une colonne NHS. Le mélange à purifié est ensuite incubé avec la colonne. La colonne est lavée et les protéines conjuguées ainsi que le tétrasaccharide libre restent fixés sur la colonne. Le mélange est ensuite élué avec un sucre compétiteur.

Cependant, il s'est avéré que la capacité de fixation des colonnes de galactine-3 était trop faible. En effet, la capacité de rétention d'une colonne de galactine-3 fixée à une colonne commerciale NHS par rapport à un sucre libre a été décrite comme étant de 1% (Hirabayashi et al. 2002). Afin de vérifier cette capacité de rétention, nous avons cherché à purifier une autre protéine conjuguée au tétrasaccharide (BSA-tétrасaccharide) qui était disponible au laboratoire et nous n'avons pu retenir que 0,7 % de la protéine chargée sur la colonne soit 6,3 nmoles pour 870nmoles de Galactine-3 fixées. Cependant, nous cherchons à purifier 2 mg de protéine conjuguée, soit environ 59 nmoles. Ceci correspondrait à répéter la purification au moins 10 fois. De plus, étant donné que le conjugué BSA-tétrасaccharide est un conjugué aléatoire avec environ 5 tétrasaccharides fixés sur une protéine, et que celui-ci n'était pas en mélange compétitif avec du tétrasaccharide libre, il est possible que la capacité de rétention de la colonne soit encore réduite pour le conjugué mPsaA-tétrасaccharide. Nous avons donc considéré que cette méthode n'était pas adaptée aux échelles de purification envisagées.

1.9 Production des conjugués monovalents site spécifiques

Nous avons ainsi testé deux approches pour produire les glycoconjugués (procédés 1 et 2 Figure 70). Cependant, l'efficacité de coupure de l'étiquette histidine par la TEV protéase sur les constructions possédant celui-ci en C-terminal s'est avérée inefficace (section 1.7) et la purification d'affinité possède une capacité de rétention insuffisante (section 1.8). Nous avons donc choisi de supprimer l'étape de coupure de l'étiquette histidine (procédé 3 Figure 70). En effet, bien que celui-ci puisse avoir un effet sur la réponse immunitaire, chaque conjugué aura la même étiquette histidine et sera donc potentiellement affecté de manière égale par celle-ci. Deux milligrammes de chaque conjugué ont donc été produits selon le procédé 3 (Figure 70) et ont été utilisés pour des tests d'immunisation sur des souris. Les composés seront également analysés par spectrométrie de masse ce qui pourra confirmer la conjugaison.

Procédés de production des glycoconjugués

Procédé 1 : Etiquette Histidine en C-terminal	Procédé 2 : Etiquette Histidine en N-terminal	Procédé 3 : Etiquette Histidine en C-terminal
1- Production mPsaA-PrK (pET24d)	1- Production mPsaA-PrK (pET11a)	1- Production mPsaA-PrK (pET24d)
2- Clivage de l'étiquette histidine	2- Conjugaison avec PnTS14	2- Conjugaison avec PnTS14
3- Conjugaison avec PnTS14	3- Purification en Chromatographie d'affinité (galectine)	3- Purification en Chromatographie d'exclusion stérique
4- Purification en Chromatographie d'exclusion stérique	4- Purification en Chromatographie d'exclusion stérique	➔ Production de 2mg de : -mPsaA-K24 ^{PrK} -PnTS14-Histag -mPsaA-K213 ^{PrK} -PnTS14-Histag -mPsaA-K229 ^{PrK} -PnTS14-Histag

Figure 70 : Présentation des différents procédés testés pour la production de vaccins glycoconjugués. Les étapes ayant entraîné des difficultés sont indiquées en rouge. Le procédé 1 a été compromis par l'inefficacité du clivage de l'étiquette histidine par la TEV protéase sur les constructions possédant celui-ci en C-terminal, tandis que le procédé 2 a été compris par la capacité de la colonne d'affinité qui s'est avérée trop faible pour notre objectif de production (2mg par conjugué). Le procédé 3 a alors été choisi et a permis de réaliser la production de 2mg de chaque conjugué (en position K24, K213 ou K229) avec une étiquette histidine en C-terminal. Ces conjugués ont alors été utilisés pour des tests d'immunisation sur des souris.

I.10 Conclusion sur la production d'un vaccin monovalent site spécifique

La production de glycoconjugués est généralement réalisée par des méthodes aléatoires basées sur la réactivité des lysines. Cependant, cela génère des glycoconjugués non homogènes et la conjugaison n'est pas ciblée. Il est ainsi difficile d'étudier l'impact de la position du tétrasaccharide sur la réponse immunitaire. Des cystéines peuvent être introduites dans la séquence de la protéine porteuse afin de réaliser des conjugaisons ciblées. Cependant, les cystéines sont parfois impliquées dans des rôles structurels. Ce projet nous a donc permis d'apporter une nouvelle réactivité ciblée indépendante de ces contraintes et nous avons pu ainsi produire 2mg de trois conjugués différents permettant d'étudier la position du greffage du tétrasaccharide sur la PsaA. De plus, cette technologie pourra être utilisée dans le cadre d'autres projets du laboratoire au sein de l'axe de recherche sur les vaccins glycoconjugués. Ainsi, dans la continuité de ce projet, par l'utilisation combinée de la technologie d'incorporation d'acides aminés non naturels et de mutants cystéines, il serait alors possible de greffer spécifiquement plusieurs éléments sur une protéine porteuse. En effet, au sein d'un vaccin glycoconjugué des adjuvants sont utilisés afin de booster la réponse immunitaire. La liaison covalente de ceux-ci à la protéine peut favoriser leur efficacité en évitant leur diffusion. Grâce au couplage de ces technologies, l'impact de la position et du ratio des différents éléments du vaccin pourrait alors être étudié.

II. Optimisation de l'incorporation de la Propargyl-Lysine

L'incorporation d'acide aminé non naturel n'est pas aussi efficace que l'incorporation d'un acide aminé classique. En effet, des arrêts prématurés de la traduction au niveau du codon stop TAG conduisent en général à la production d'un mélange entre des protéines tronquées et des protéines complètes ayant incorporé l'acide aminé non naturel. Plusieurs pistes d'optimisation sont actuellement étudiées afin d'améliorer l'efficacité d'incorporation (Chapitre 1, section IV.4). Nous avons choisi trois pistes afin d'essayer d'améliorer l'efficacité d'incorporation de la PrK dans la protéine mPsaA: i) surexpression d'un transporteur de lysine ii) mutagenèse de l'aaRS selon la publication (Vangmayee Sharma et al. 2018) et iii) réalisation de plans d'expérience.

II.1 Surexpression d'un transporteur de lysine

La Propargyl-Lysine est un acide aminé non naturel dont le transport au sein de la cellule n'est pas connu. Cependant, comme nous l'avons vu section I.4, la Propargyl-Lysine rentre bien dans la bactérie. L'efficacité catalytique d'une enzyme est dépendante de la concentration en substrat. Nous pouvons donc supposer qu'en augmentant la concentration intracellulaire de la Propargyl-Lysine, la catalyse de l'aminocyclisation de l'ARNt orthogonal pourrait être augmentée. Nous supposons que cela pourrait entraîner une augmentation de la production de protéines contenant la Propargyl-Lysine. Au sein de la bactérie, il existe plusieurs transporteurs d'acides aminés plus ou moins spécialisés. Pourtant encore peu d'études ont été réalisées afin d'améliorer le transport d'acides aminés non naturels. De par sa structure, nous avons émis l'hypothèse que la Propargyl-Lysine pourrait être transportée vers la bactérie à l'aide du transporteur de la Lysine (LysP). Un plasmide pEVOL-1copie a été construit pour faciliter les expériences de criblage au cours du deuxième axe d'étude de ma thèse (Chapitre 7 section I.3.3.1). Au sein du plasmide pEVOL-1copie, seule la copie de l'aARS sous le contrôle du promoteur inductible à l'arabinose a été conservée. Cependant, l'efficacité d'incorporation n'en semble pas altérée (données non montrées). Le gène du transporteur LysP a donc été cloné au sein du plasmide pEVOL à la place de la copie de l'aARS sous le contrôle du promoteur constitutif GlnS, tout en conservant ce promoteur. L'efficacité du transporteur LysP a été évaluée par la comparaison du taux d'incorporation de la PrK au sein de la protéine de fusion mPsaA-K193TAG-eGFP dans des bactéries portant le plasmide pEVOL-1copie ou le plasmide portant également le gène du transporteur LysP, pEVOL-1copie-LysP. L'incorporation de la PrK dans cette protéine de fusion conduit à la production d'une protéine complète fluorescente, le taux de fluorescence émise par les bactéries reflétant l'efficacité de production de la protéine et donc l'efficacité d'incorporation de l'acide aminé non naturel. Comme nous pouvons le voir sur la Figure 71, la différence d'incorporation entre les deux souches n'est pas significative. Nous pouvons voir qu'à de fortes concentrations de PrK (2,5 mM ou 3 mM), on pourrait supposer une meilleure incorporation mais après répétitions des mesures, cela ne semble pas significatif.

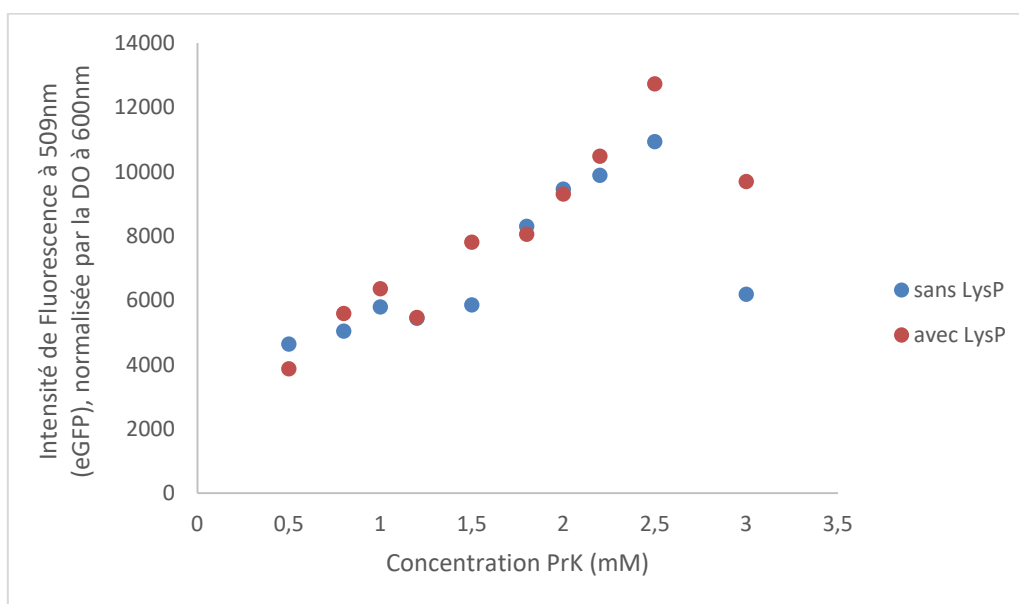


Figure 71 : Comparaison de l'efficacité de l'incorporation de la PrK avec ou sans surexpression de LysP.

Des tests d'incorporations ont été réalisés au sein de BL21(DE3) co-transformée avec le plasmide pEVOL-1copie portant une copie de la synthétase ou le plasmide pEVOL-1copie-LysP portant également une copie du gène du transporteur de la lysine et le plasmide pET24d portant le gène de la protéine de fusion mPsaA-GFP avec un codon stop à la position K193. Lorsque la Propargyl-Lysine est incorporée au sein de la protéine, la protéine fluorescente mPsaA-GFP est donc produite. Cette fluorescence est mesurée grâce au spectrophotomètre TECAN (maximum d'absorption : 488nm, maximum d'émission : 509nm).

Il est possible que le transporteur LysP ne prenne pas en charge la Propargyl-Lysine ou que sa surexpression n'ait pas été fonctionnelle. A défaut de résultat préliminaire positif, nous avons néanmoins choisi de ne pas poursuivre cette étude. De nombreux autres paramètres affectent le taux d'incorporation d'un acide aminé non naturel comme la concentration en ARNt orthogonal, ou les compatibilités de l'aaRS et de l'ARNt aux éléments endogènes à la traduction (ribosome, facteur d'élongation EF-Tu) (Chapitre 1, section IV.4).

II.2 Mutagenèse du domaine N-terminal de l'aminocyl ARNt synthétase

Au cours de ma thèse, une publication sur l'évolution du domaine N-terminal de la pyrrolysyl ARNt synthétase issue de *M. mazei* a été réalisée par Sharma *et al.* afin d'améliorer les rendements d'incorporation de deux dérivés de lysines (Boc-Lysine et la 4-azidobenzoxycarbonyl-dehydrolysine, AcDK) (Vangmayee Sharma *et al.* 2018). Cette évolution dirigée converge vers 3 mutations : R19H, H29R, et T122S. Comme cette mutagenèse concerne la partie N-terminal de la synthétase, elle n'affecte théoriquement pas la zone d'interaction avec le substrat de l'enzyme et serait donc transférable à n'importe quelle aminocyl ARNt synthétase dérivée de la pyrrolysyl ARNt synthétase de *M.mazei*. Nous avons donc réalisé les mutations R19H, H29R, et T122S sur le gène de l'aaRS du plasmide pEVOL-1copie. Puis nous avons réalisé des tests d'efficacité d'incorporation de la PrK au sein de la protéine fluorescente mPsaA-K193-eGFP en utilisant le plasmide pEVOL-1copie ou le plasmide portant les mutations définies par Sharma *et al.*, pEVOL-1copie- R19H-H29R-T122S. Cependant, comme nous pouvons le voir sur la Figure 72, il n'y a pas de différence significative entre ces deux conditions.

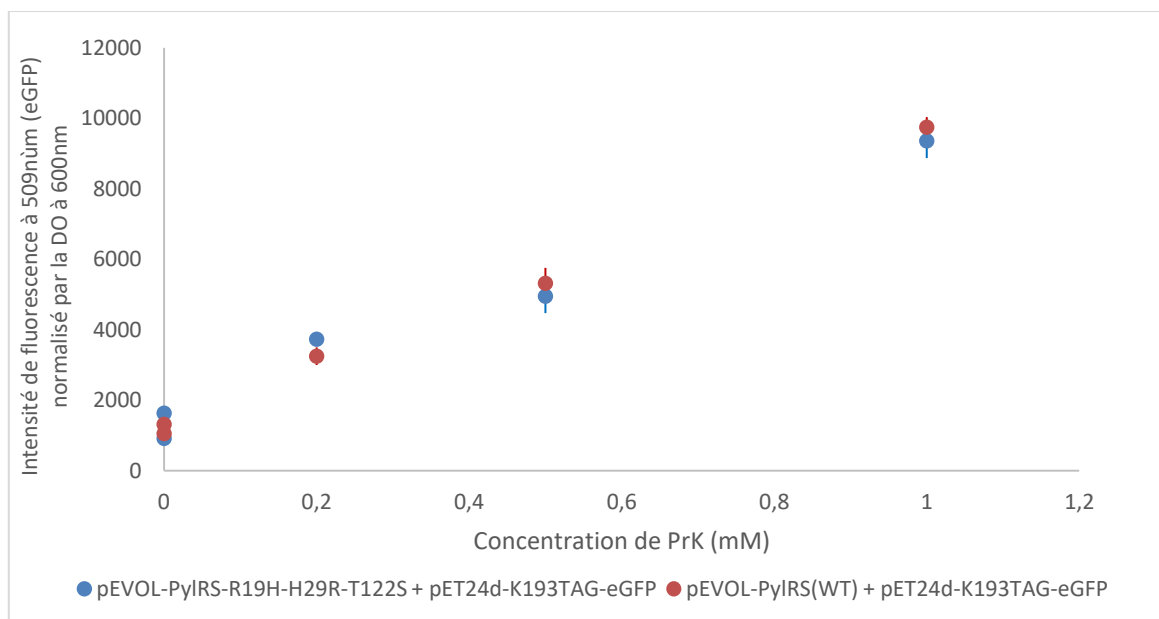


Figure 72 : Evaluation de l'impact des mutations R19H, H29R et T122S sur l'incorporation de la Propargyl-Lysine. Des tests d'incorporations ont été réalisés au sein de BL21(DE3) co-transformées avec le plasmide pEVOL-1copie portant une copie de la synthétase ou le plasmide pEVOL-1copie portant les mutations R19H, H29R et T122S et le plasmide pET24d portant le gène de la protéine de fusion mPsaA-GFP avec un codon stop à la position K193. Lorsque la Propargyl-Lysine est incorporée au sein de la protéine, la protéine fluorescente mPsaA-GFP est donc produite. Cette fluorescence est mesurée grâce au spectrophotomètre TECAN (maximum d'absorption : 488nm, maximum d'émission : 509nm).

Les mutations R19H, H29R, et T122S ne semblent donc pas efficaces dans notre cas pour améliorer l'efficacité d'incorporation de la Propargyl-Lysine. Nous n'avons donc pas continué à utiliser ces mutations lors des autres expérimentations.

II.3 Optimisation par plans d'expérience

Etant donné que de nombreux paramètres influent sur la production de protéine ayant incorporé un acide aminé non naturel, nous avons décidé de réaliser un plan d'expérience plutôt que de réaliser une optimisation par des méthodes d'expérimentations classiques. En effet, lorsque l'on étudie un phénomène biologique par une méthodologie d'expérimentation classique, les expériences sont réalisées de façon séquentielle. Un seul paramètre est modifié à la fois (également appelé facteur) et les expériences suivantes sont planifiées en fonction des résultats obtenus. Cependant, selon le modèle d'étude cela peut se traduire par de nombreuses expériences. De plus, lors d'une optimisation, cette méthodologie n'est pas efficace car l'arrêt des expérimentations se fera de façon subjective lorsqu'une réponse sera jugée suffisante. La solution choisie ne sera alors pas nécessairement la plus optimale et cette méthode ne permet pas de prédire le résultat d'expérimentations non réalisées.

Un plan d'expérience est une méthodologie qui permet d'étudier l'impact de différents facteurs sur une réponse en réalisant un minimum d'expériences afin de pouvoir dégager un modèle qui permette de trouver les conditions optimales. Cette stratégie permet alors d'étudier l'influence de chaque facteur, leurs interactions et de pouvoir définir de façon méthodologique si la réponse choisie est optimale.

Cette stratégie de plan d'expérience se déroule en trois étapes : i) la définition du problème ii) le choix de stratégie et la réalisation des expériences iii) l'analyse des résultats.

II.3.1 Définition du problème

Avant de réaliser un plan d'expérience, une étape de préparation est nécessaire. Cette étape va consister à identifier la réponse à mesurer, les facteurs que nous voulons étudier ainsi que le domaine d'étude et l'objectif fixé.

Réponse mesurée :

Afin de réaliser un plan d'expérience, la réponse étudiée doit être quantitative et mesurable de façon fiable et robuste. La réponse doit également être une variable continue au sein du domaine d'étude. Afin de pouvoir quantifier l'incorporation de la Propargyl-Lysine (PrK), nous avons choisi de mesurer son incorporation au sein d'une protéine de fusion entre la mPsaA et la GFP à une position que nous avons validée précédemment (mPsaA-K193TAG-GFP). Ainsi, l'incorporation de la PrK donne lieu à la production d'une protéine fluorescente facilement détectable. L'intensité de fluorescence mesurée sera alors proportionnelle à l'efficacité de production de la protéine ayant incorporé la PrK.

Facteurs étudiés :

Nous avons alors listé les différents facteurs qui pourraient influencer notre réponse (Tableau 16). Puis nous avons sélectionné les facteurs que nous souhaitons étudier (indiqués par un astérisque au sein du Tableau 16 : la température, la concentration de l'inducteur permettant la production de l'aaRS (ici l'arabinose), le temps d'ajout de celui-ci et le pH du milieu dont nous allons discuter les effets lors de la mise au point du domaine d'étude ci-dessous). Nous avons volontairement évité d'inclure les facteurs correspondant à la Propargyl-Lysine car ceux-ci ont été étudiés lors d'un travail réalisé par Wandrey *et al.*, ce qui a conduit à une condition optimale notamment par l'ajout de la PrK à une concentration de 38 mM (Wandrey et al. 2016). Cependant, la quantité d'acide aminé non naturel étant souvent un facteur limitant, nous avons donc plutôt orienté notre travail sur la maximisation de la production de la protéine ayant incorporé la Propargyl-Lysine sans augmenter la concentration de cette dernière. De même, nous avons écarté les facteurs liés à la souche d'expression et au milieu d'expression (le milieu autoinductif). En effet, les essais précédents réalisés au sein de différents systèmes d'expression et de types de milieux se sont avérés non concluants, et nous avons plutôt cherché à améliorer les conditions existantes. Cependant, il est

important de s'assurer de lister le maximum de facteurs pouvant influencer la réponse car ceux-ci doivent être fixés afin d'éviter des fluctuations qui fausseraient les mesures et donc le modèle déduit.

Tableau 16 : Facteurs pouvant influencer le taux d'incorporation de la Propargyl-Lysine

- Facteurs
- Température d'expression***
 - Milieu d'expression**
 - Type d'inducteur protéine cible**
 - Type inducteur aaRS**
 - Concentration inducteur protéine cible**
 - Concentration inducteur aaRS***
 - Temps d'expression**
 - pH du milieu***
 - concentration de PrK**
 - temps d'ajout de la PrK**
 - Temps ajout inducteur aaRS***
 - Temps ajout inducteur protéine cible**
 - Plasmides utilisés**
 - Souche d'expression**
 - Expérimentateur**

Définition du domaine d'étude :

Un domaine d'étude doit être défini en fixant les valeurs minimales et maximales de chaque facteur étudié. Le modèle d'étude correspondra alors à l'espace mathématique au sein des bornes du domaine d'étude (Figure 73). Le modèle permettant de modéliser la réponse en fonction des différents facteurs, qui sera déduit expérimentalement, sera alors valable à l'intérieur de ce domaine d'étude.

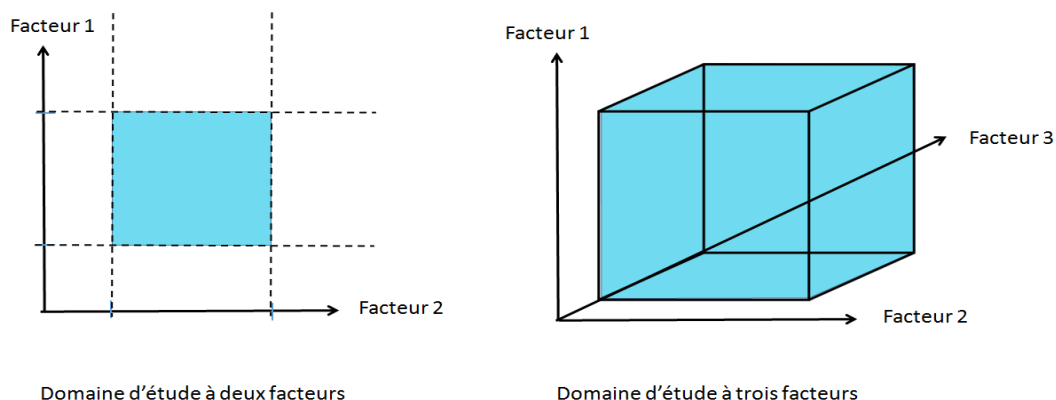


Figure 73: Schéma représentant un domaine d'étude pour un plan d'expérience à deux facteurs (à gauche) ou trois facteurs (à droite)

Nous avons alors fixé les valeurs présentées dans le Tableau 17 selon l'analyse suivante :

-Température : la production de la mPsaA se déroule de façon optimale au laboratoire à 37°C. Cependant, afin de favoriser l'incorporation de la Propargyl-Lysine, il peut être utile de baisser la température d'expression. En effet, nous supposons que ralentir la vitesse d'expression protéique pourrait permettre une meilleure incorporation des acides aminés non naturels qui peuvent nécessiter plus de temps pour être pris en charge par la machinerie de traduction. Afin de ne pas trop ralentir la production de la protéine cible, nous avons cependant choisi de ne pas descendre en dessous de 20°C.

-Concentration de l'arabinose : l'induction de l'aaRS est un paramètre important, la production de protéines induites à l'aide de promoteur inductible à l'arabinose est en générale testée entre une concentration de 0,02% et 2%.

-Temps d'ajout de l'arabinose : afin que l'incorporation de la PrK soit efficace, il faut que l'aaRS soit produite avant la protéine cible pour charger la PrK sur l'ARNt et permettre son incorporation dans la protéine. Nous avons donc réalisé une expérience de production de la protéine cible mPsaA-GFP avec nos conditions initiales (température 37°C, pH7, Concentration d'arabinose 0,2% et ajout de l'arabinose à t2h (ce qui équivaut à une densité optique à 600nm d'environ 0,2 à 0,3) afin de déterminer à partir de combien de temps celle-ci s'exprime au sein du milieu autoinductif et nous avons détecté l'apparition d'un signal fluorescent à partir de 420 minutes. L'arabinose doit donc être ajouté entre le temps (t0) de départ de la culture et 420 minutes.

-pH : L'acidité du milieu est décrite comme permettant d'augmenter l'expression du transporteur LysP (Steffes et al. 1992). Etant donné que nous supposons que la PrK rentre au sein de la bactérie par ce transporteur, nous avons voulu vérifier l'impact de ce facteur sur son incorporation. Le milieu autoinductif utilisé est un milieu à pH stable tamponné cependant un pH trop acide risque d'être délétère pour la croissance. Nous avons donc choisi d'étudier ce facteur entre 5,5 et 7.

Tableau 17 : Détermination du domaine d'étude (niveaux hauts et bas des facteurs)

Facteur étudié	Valeur basse	Valeur haute
Température	20°C	37°C
Concentration d'arabinose	0,02%	2%
Temps d'ajout de l'arabinose	t0	420
pH	5,5	7

Objectif :

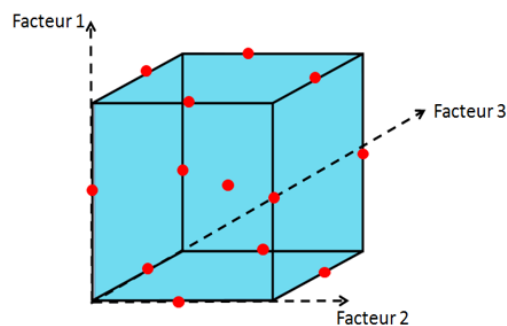
Afin de déterminer la fin de l'optimisation de l'incorporation de la PrK dans la protéine cible, nous devons choisir le but que nous voulons atteindre. Ici nous cherchons à augmenter la production de protéine ayant incorporé la PrK. Nos conditions utilisées initialement (température 37°C, pH7, Concentration d'arabinose 0,2% et ajout de l'arabinose à t2h, soit DO à 600nm 0,2-0,3) permettent une production entraînant un niveau de fluorescence d'environ 2000 Unités de fluorescence. L'augmentation maximale que nous pouvons atteindre consiste en une efficacité d'incorporation similaire à l'incorporation d'un acide aminé naturel et donc à l'expression d'une protéine sauvage mPsaA-GFP. Ce maximum d'intensité de fluorescence en théorie atteignable correspond donc à environ 10 000 unités de fluorescence, soit une augmentation d'un facteur 5.

II.3.2 Choix du modèle et réalisation des expériences

Choix du modèle :

Les plans d'expériences consistent à choisir quelles expériences vont être réalisées afin de pouvoir déterminer un modèle fiable statistiquement sur l'ensemble du domaine d'étude. Tout d'abord, les relations facteurs-réponses peuvent être supposées linéaires ou non linéaires. Lors de l'analyse de facteurs linéaires, deux niveaux de ce facteur seront alors testés. Cependant, si l'on suppose que le facteur aura une influence non linéaire, trois niveaux seront alors testés. De plus, il existe des plans factoriels complets, c'est-à-dire qu'ils permettent d'établir les effets de tous les facteurs et de leurs interactions. Cependant, ce type de plan d'expérience aboutit à de nombreuses expérimentations. Par exemple, lorsqu'on étudie 4 facteurs à 3 niveaux, cela entraîne la réalisation de 3^4 soit 81 expériences. Il existe également des plans factoriels fractionnaires, où les 81 expériences ne seront pas réalisées car certaines interactions seront considérées comme négligeables. Ainsi lorsqu'il y a quatre facteurs (x_1, x_2, x_3, x_4), les interactions du plus haut degré seront négligées, par exemple l'interaction entre ces quatre facteurs $x_1x_2x_3x_4$.

Nous avons supposé que nos 4 facteurs étudiés (la température d'expression, la concentration de l'inducteur permettant la production de l'aaRS (ici l'arabinose), le temps d'ajout de celui-ci et le pH du milieu) avaient une influence non linéaire sur l'incorporation de la PrK. Différents modèles permettent l'étude de facteurs non linéaire de façon optimale. Nous avons donc choisi d'utiliser un modèle de type Box-Behnken (Goupy and Creighton 2006). Au sein de ce modèle, les points expérimentaux sont distribués au milieu des arêtes comme indiqués au sein de la Figure 74. L'avantage de ce modèle est que si, après une première expérimentation, un nouveau facteur doit être pris en compte, les résultats expérimentaux déjà réalisés peuvent être conservés. De plus, cette répartition des points expérimentaux permet de répartir les points à égale distance du point central et donc d'avoir une meilleure précision mathématique. Un point au centre du domaine d'étude est également réalisé en triplicata afin de déterminer l'erreur expérimentale, ce qui nous permettra de déduire quels facteurs sont influents (section II.3.3).



Box-Behnken

Figure 74 : Répartition des points expérimentaux selon un plan d'expérience de type Box-Behnken. Les différents points expérimentaux sont représentés en rouge et sont situés au milieu des arêtes du domaine d'étude, ainsi qu'au centre. De plus, le point central sera répété afin de connaître la variabilité des expérimentations.

Le modèle choisi étant de second degré avec 4 facteurs, il sera décrit par la formule mathématique suivante: « $y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{14}x_1x_4 + a_{23}x_2x_3 + a_{24}x_2x_4 + a_{34}x_3x_4 + a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + a_{44}x_4^2 + e$ »

où y est la réponse (soit l'intensité de fluorescence proportionnelle à l'incorporation de la PrK) en fonction des différents facteurs x associés à leur coefficient a (qui indiqueront l'effet de chaque facteur sur la réponse) et e l'erreur expérimentale.

Les différents facteurs sont:

x_1 : la température

x_2 : le pH

x_3 : la concentration d'arabinose

x_4 : le temps d'ajout de l'arabinose

$x_i x_j$: l'interaction entre le facteur i et le facteur j

Réalisation des expériences :

Le plan d'expérience a été construit à l'aide du logiciel Rstudio et des différents packages listés en annexe 5. Les facteurs comportent trois niveaux qui sont utilisés de manière centrée réduite. Pour cela, la moyenne est soustraite à la valeur et divisée par l'écart type. Par exemple, la valeur de 37°C est issue du domaine d'étude compris entre 20°C et 37°C, dont la moyenne est 28,5 et d'écart type 8.5, cette valeur sera donc redéfinie par 1. Les valeurs extrêmes du domaine d'études seront alors -1 et 1 tandis que la valeur centrale sera égale à 0. Cette manière de coder les données permet d'avoir une homogénéité parmi les différents facteurs étudiés. Une attention particulière est portée sur la précision des manipulations afin de limiter l'erreur expérimentale et les expériences sont réalisées selon le protocole décrit chapitre 4, section IV.1. Le plan d'expérience a abouti à la réalisation de 28 expériences et les valeurs réponses (intensité de fluorescence) ont ensuite été intégrées au logiciel afin de définir un modèle.

II.3.3 Analyse des résultats

Modèle :

Les données brutes obtenues sont présentées en Annexe 6. Leur analyse a permis de déterminer le modèle suivant grâce au logiciel Rstudio :

$$y = 405,36 + 262,73x_1 + 727,74x_2 - 1332,31x_3 + 319,29x_4 + 145,93x_1x_2 - 589,95x_1x_3 - 2,2675x_1x_4 - 730,10x_2x_3 + 970,43x_2x_4 + 516,91x_3x_4 - 367,85x_1^2 + 864,18x_2^2 + 1233,36x_3^2 + 377,25x_4^2 + e$$

Fiabilité du modèle :

Le coefficient R^2 permet de déterminer la qualité du modèle proposé. Il correspond au rapport entre la somme des carrés des réponses calculées corrigées de la moyenne et la somme des carrés mesurés corrigés de la moyenne. Sa valeur varie entre 0 et 1 et est un indicateur de la qualité du modèle proposé. Cependant, comme la valeur du R^2 augmente avec le nombre de facteur, la valeur du R^2 ajustée permet de prendre en considération de nombre de facteurs étudiés. La valeur du R^2 ajustée du modèle obtenu est égale à 0,67. L'équation de la droite de régression est donc capable de déterminer environ 67% de la distribution des points. Cette valeur n'est pas très élevée, il est donc possible qu'un facteur important n'ait pas été correctement fixé et qu'il entraîne une variabilité dans la réponse.

De plus, afin de vérifier la validité du modèle nous avons représenté graphiquement les résidus calculés par le modèle Figure 75. Les résidus étant les distances qui séparent chaque point expérimental de la courbe définie par le modèle, ceux-ci doivent présenter une distribution aléatoire autour de 0 (ligne rouge Figure 75) car ils représentent la variabilité expérimentale. Si la distribution aléatoire n'est pas observée, cela signifie qu'un paramètre qui influe sur la réponse n'a pas été pris en compte. Dans notre cas, la répartition des résidus selon les valeurs expérimentales semble être plutôt aléatoire et donc de part et d'autre de la valeur zéro, bien que l'on puisse remarquer qu'il y a

peu de données pour les valeurs les plus élevées (au-dessus de 3000 unités de fluorescence normalisée) et que les résidus calculés pour ces valeurs sont plus élevés.

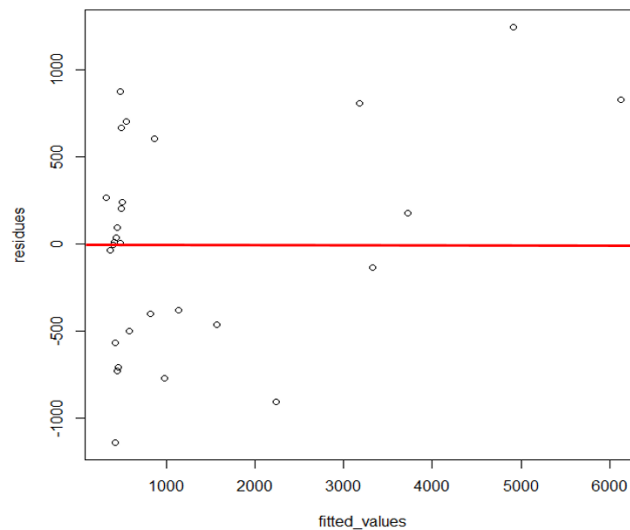


Figure 75: Représentation graphique des résidus en fonction des valeurs ajustées grâce au modèle
Les résidus correspondent à l'écart calculé entre la valeur ajustée grâce au modèle (c'est-à-dire prédite) et la valeur obtenue expérimentalement.

Influence des différents facteurs :

La valeur devant chaque facteur x_i représente l'effet du facteur. Cependant tous les effets ne sont pas nécessairement significatifs. C'est pour cela qu'un point expérimental au centre du domaine d'étude est répété en triplicata afin de connaître la variabilité de l'expérience.

Le logiciel Rstudio permet alors de calculer une *pvalue* pour chaque facteur et pour chaque interaction entre les facteurs afin de déterminer si leur effet est significatif. Cette *pvalue* est calculée à l'aide d'un test de Student consistant à comparer une moyenne observée avec une moyenne théorique. D'après les données, les facteurs significatifs sont x_2 , x_3 , x_2x_4 , soit le pH, la concentration en arabinose et l'interaction pH/Temps d'ajout de l'arabinose. x_2^2 et x_3^2 (pH et concentration en arabinose) ont également des *pvalues* inférieures à 0,05 ce qui signifie que ces facteurs ont une influence non linéaire.

Représentation de la Fluorescence en fonction des différents facteurs :

Nous pouvons alors tracer des graphiques représentant la réponse en fonction des facteurs par couples de deux facteurs. Les facteurs non représentés ont ici été fixés à leur valeur médiane.

La représentation graphique permet de simplifier l'analyse des résultats (Figure 76). Ainsi, nous pouvons voir qu'un coefficient du facteur négatif indique que lorsque le facteur augmente la réponse diminue. C'est le cas notamment pour la concentration en arabinose. C'est-à-dire qu'au sein du domaine d'étude, la meilleure réponse est en fait obtenue lorsque cette concentration est la plus faible, à son niveau -1 (soit 0,02%). Au contraire, la réponse est plus forte lorsque le pH est à sa valeur la plus haute (pH 7). La baisse du pH a donc finalement eu un effet plutôt délétère sur l'incorporation de la Propargyl-Lysine. Bien que les effets des autres facteurs soient moins significatifs, nous pouvons noter que la température à 37°C permet une meilleure efficacité d'incorporation et que l'arabinose peut être ajouté assez tard dans la réalisation de la culture. De façon assez surprenante, l'interaction des facteurs du pH et du temps d'ajout de l'arabinose semble significative.

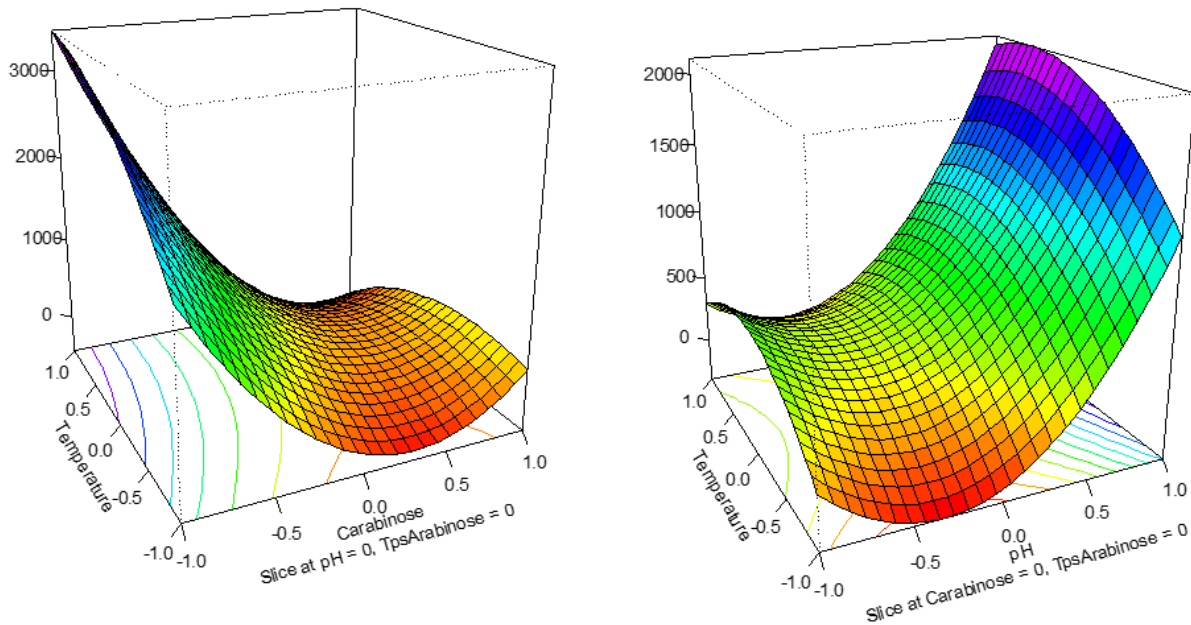


Figure 76: Représentation de la fluorescence en fonction de la température et de la concentration en arabinose (à gauche) et en fonction de la température et du pH (à droite)

Maximum de la réponse :

Le logiciel permet le calcul d'un point stationnaire, c'est à dire un point où la dérivée sera égale à 0. Si les valeurs propres calculées appelées "Eigens values" sont de signes identiques, cela signifie que la courbe change de direction et qu'il s'agit donc d'un minimum ou maximum global. Ici, le logiciel nous a fourni des valeurs de signes mixtes (données en Annexe 6), ce qui signifie que le point donné n'est pas un maximum global. De plus les valeurs données ne sont pas dans le domaine d'étude et ne sont pas cohérentes biologiquement. En effet, le logiciel calcule une concentration négative d'arabinose, et un pH négatif.

Afin de déterminer un maximum de notre fonction sur le domaine d'étude, nous avons généré des données à partir de la fonction modèle, et déterminé une valeur maximale de 7071 unités de fluorescence normalisée pour des valeurs des facteurs suivants :

- Température 37°C
- pH7
- Concentration en arabinose 0,02%
- Temps d'ajout de l'arabinose 420 minutes

Nous pouvons noter que ces valeurs sont situées aux bornes du domaine d'étude. Nous venons de voir que la fonction du modèle ne prend pas de valeur maximale. Ces valeurs sont cohérentes avec les coefficients de chaque facteur et les graphiques présentés Figure 76. En effet, la valeur de la température prédite est à l'extrémité supérieure du domaine d'étude et le coefficient a_1 qui correspond à l'effet de ce facteur est positif. Il est probable que la fenêtre du domaine d'étude doit être élargie afin de préciser cette réponse. De plus, nous avons vu Figure 75 que la précision du modèle était faible pour les valeurs élevées de fluorescence, il est donc possible qu'augmenter notre domaine d'étude pourrait nous permettre d'améliorer nos prédictions. Cependant, biologiquement il est logique que la valeur n'augmente pas pour une température supérieure à 37°C, ou pour un pH plus élevé. Pour la valeur de la concentration d'arabinose, nous savons qu'elle doit cependant être minimale pour induire la production de la synthétase, bien que des concentrations plus faibles peuvent être testées. De plus, la synthétase ne peut être induite après la production de la protéine

cible, sa production tardive pourrait favoriser des erreurs d'incorporation. Il est donc probable que si la fonction n'atteint pas son maximum au sein du domaine d'étude, nous supposons que celle-ci atteint un maximum aux environs proches du domaine d'étude.

Vérification du modèle et Conclusion :

Le point expérimental correspondant au maximum de la réponse calculée a été répété afin de vérifier la fiabilité du modèle (Tableau 18). Le point répété correspond à un pH de 7, une température d'expression de 37°C, une concentration d'ajout de l'arabinose de 0,02% et l'ajout de l'arabinose à t=7h. Le résultat expérimental est de 2077 unités de fluorescence normalisé par la DO, sachant que le résultat prédit par le modèle est de 7071 et que le coefficient de variation des manipulations précédentes de 9,57%, soit 676 unités de fluorescence normalisé par la DO. Le modèle n'a donc pas prédit correctement la valeur du maximum de fluorescence. Le temps d'ajout de l'arabinose semble être un facteur important qui n'a pas été correctement prédit par le modèle. Nous avons donc répété des tests en gardant les autres facteurs fixes (pH 7, température 37°C, concentration de l'arabinose 0,02%) et en faisant varier ce facteur « temps d'ajout de l'arabinose » comme lors d'une expérimentation classique. Nous avons obtenu les résultats présentés dans le tableau suivant :

Tableau 18 : Vérification du modèle prédit

Des expériences d'incorporation de la Propargyl-Lysine au sein de la protéine de fusion mPsaA-GFP contenant un codon stop en position K193 ont été réalisées. L'arabinose a été ajoutés à différents temps (t0, 120min, 210min, 300min et 420min). La fluorescence produite est un indicateur de l'efficacité d'incorporation de la Propargyl-Lysine. Elle a été mesurée à l'aide du TECAN (max absorption : 488nm et max émission : 509nm). Les valeurs sont indiquées en unités de fluorescence normalisées par la densité optique à 600nm.

Temps d'ajout de l'arabinose	t 0	t 120min	t 210min	t 300min	t 420min
Valeur expérimentale	4609	4606	5932	2914	2077
Valeur calculée à l'aide du modèle	5530	5607	5923	6369	7071
Variabilité (issue % variation)	529	536	566	609	676
Correctement prédite	Non	non	oui	non	non

Malgré l'utilisation d'un modèle permettant la prédiction de relations non linéaires, le modèle n'a correctement prédit la valeur de fluorescence qu'au centre du domaine d'étude, et le maximum de la réponse n'est donc pas obtenu à la valeur prédite. Plusieurs raisons pourraient l'expliquer : un facteur n'est pas pris en compte, ou des interactions entre les facteurs de plus haut niveau (entre trois facteurs par exemple) ont été négligées mais ont une influence significative. Les données forment un pic avec une valeur maximale pour 3h30. Ce type de relation n'est cependant pas modélisable par une seule équation de second degré. En effet, le facteur ne peut pas à la fois prendre une valeur positive (entre t0 et t3h30) et négative (entre t3h30 et t7h). De plus, nous pouvons noter que, contrairement à ce qui est prédit, l'influence du facteur « temps d'ajout de l'arabinose » n'est pas négligeable.

Nous avons pu voir précédemment Figure 75 que la précision du modèle est moindre pour les valeurs de fluorescence élevée. De plus, le plan d'expérience que nous avons établi étant un plan fractionnaire, il n'a pas fourni un modèle optimal. Cependant, les conditions d'expression des protéines influent de façon importante sur leur rendement de production. Nous avons testé 4 paramètres qui nous semblaient importants : la température de production, le pH, la quantité d'arabinose ajoutée et le temps de l'ajout de l'arabinose. Le plan d'expérience initialement réalisé est de type Box Behnken fractionnaire et consiste en 28 expériences, alors qu'un plan factoriel complet

avec 4 facteurs et 3 niveaux par facteurs aurait demandé la réalisation de 81 expériences. Le plan d'expérience choisi permet de réduire significativement le nombre d'expériences. Cependant, un plan plus complet aurait peut-être permis d'améliorer les performances de celui-ci. Le modèle obtenu n'est pas très prédictif et le bruit expérimental est assez fort (9,57% de coefficient de variation) mais cette rapide étude a cependant permis d'améliorer les conditions d'incorporation de la Propargyl-Lysine au sein de la mPsaA d'un facteur 3, soit environ 6000 unités de fluorescences normalisées. En effet, les fluorescences obtenues ne dépassaient pas 2000 unités de fluorescences normalisées avec les conditions utilisées avant cette étude (conditions 37°C, pH7, Concentration d'arabinose 0,2% et ajout de l'arabinose à t2h, soit DO à 600nm 0,2-0,3). Ainsi, nous pouvons remarquer que la concentration de l'arabinose et le temps d'ajout de l'arabinose sont des paramètres très influents sur la réponse. L'arabinose est l'inducteur de l'aaRS qui permet le chargement de la Propargyl-Lysine sur l'ARNt. Cet ARNt chargé doit ensuite être pris en charge par le ribosome lors de l'expression de la protéine cible. Il paraît donc cohérent que l'expression de l'aaRS au préalable permette une meilleure expression. L'utilisation d'une faible concentration d'arabinose paraît plus surprenant. Cependant, il est possible qu'une expression trop forte de l'aaRS soit inutile et contre-productive sur l'expression protéique globale.

III. Génération de conjugués multivalents site-spécifiques

Tout comme la position du tétrasaccharide sur la protéine porteuse, le ratio protéine porteuse : tétrasaccharide semble être un paramètre important pour la réponse immunitaire. Afin de l'étudier tout en gardant des sites d'incorporations définis, nous avons incorporé plusieurs Propargyl-Lysines au sein de la mPsaA, aux positions K24, K213 et K229 combinées, grâce aux plasmides pET24d-K24TAG-K213TAG-K229TAG et pEVOL (sections suivantes). Sachant que l'immunisation de souris avec des conjugués monovalents (Pillot et al. 2019) a conduit à une réponse immunitaire plutôt faible vis à vis du tétrasaccharide et que l'immunisation de souris par un conjugué aléatoire possédant en moyenne 5 tétrasaccharides a réduit la réponse anti-mPsaA, l'utilisation de conjugués multivalents avec 3 tétrasaccharides conjugués pourrait permettre une réponse plus mixte anti-mPsaA et anti-tétrasaccharide (section Présentations des objectifs).

Le conjugué trivalent ciblé produit (issu de la conjugaison de mPsaA-K24^{PrK}-K213^{PrK}-K229^{PrK} avec le tétrasaccharide) sera ultérieurement utilisé pour des immunisations en parallèle des conjugués monovalents produits précédemment (section I) et en parallèle d'un conjugué aléatoire de la mPsaA avec en moyenne 3 tétrasaccharides (conjugaison réalisée par Aline Pillot lors de ses travaux de thèse). En effet, cela permettra de comparer à la fois la position de greffage du sucre sur la protéine (par comparaison entre les conjugués monovalents), le ratio protéine sucre (en comparant les conjugués monovalentes avec les conjugués trivalentes), et l'apport d'une conjugaison multiple ciblée (en comparant le conjugué trivalent aléatoire avec le conjugué trivalent ciblé).

III.1 Triple-incorporation de la Propargyl-Lysine au sein de la mPsaA

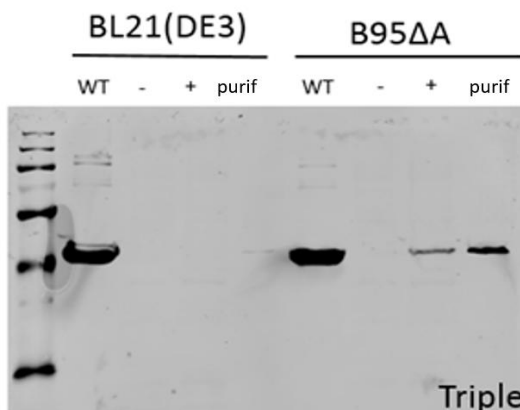
Comme nous l'avons vu lors de l'introduction (section IV.4.4), l'incorporation multiple d'acide aminé non naturel pose des problèmes de rendement. En effet, l'incorporation d'un acide aminé non naturel n'est généralement pas optimale et la production protéique aboutit à un mélange de protéines tronquées et de protéines ayant incorporé l'acide aminé non naturel (sections I.5 et I.6). Lors d'une incorporation multiple, il y a une perte de rendement à chaque position d'incorporation ce qui *in fine* réduit considérablement la quantité de protéine produite. Ainsi, comme nous pouvons le voir dans la Figure 77, la tentative d'incorporation de 3 Propargyl-Lysines dans la protéine mPsaA au

sein de la souche BL21(DE3) n'aboutit pas à une production de protéine en quantités suffisantes pour être visible sur WesternBlot.

Plusieurs hypothèses permettent d'expliquer ce rendement faible d'incorporation. Tout d'abord la compétition entre l'ARNt chargé et le facteur de terminaison RF1. En effet, le facteur RF1 reconnaît le codon stop TAG et favorise ainsi l'arrêt de la traduction. De plus, la capacité catalytique de la pyrrolysyl-ARNt synthetase serait plus faible que celle des synthétases endogènes.

L'hypothèse concernant la compétition entre le facteur de terminaison RF1 et l'ARNt chargé a conduit au développement de stratégies afin de supprimer ou de baisser l'activité du facteur de terminaison RF1 comme vu précédemment au cours du chapitre 1 d'introduction (section IV.4.2.2). Nous avons tout d'abord testé une souche nommée C321ΔAexp (Lajoie et al. 2013) disponible sur Addgene. Cette souche a été modifiée génétiquement afin de supprimer le gène *prfA* codant pour le facteur de terminaison RF1 et de remplacer les 321 codons stop TAG connus de la souche *E. coli* MG1665 par un codon stop TAA. En effet, malgré la rareté de codons stops TAG au sein de gènes essentiels chez *E. coli*, la présence dans la bactérie de la machinerie orthogonale implique la modification de protéines endogènes et la consommation non ciblée de l'acide aminé non naturel. Cependant, cette souche ne possède pas le système d'expression T7. La mPsaA a donc été clonée au sein d'un plasmide permettant son expression sous le contrôle d'un promoteur inductible à l'arabinose (pBAD) et donc indépendant du système T7 mais son expression n'a pas pu être détectée au sein de cette souche (données non montrées). Malgré son développement, cette souche reste peu utilisée pour la production de protéines car les nombreuses modifications génétiques qui ont été réalisées semblent l'avoir trop impactée, ce qui est notamment perçu à travers son défaut de croissance (Lajoie et al. 2013). En effet, au sein du génome d'*E. coli*, de nombreux gènes se chevauchent et la modification du codon stop TAG par un autre codon stop TAA peut donc entraîner des modifications additionnelles au sein d'autres gènes de la souche bactérienne. De plus, l'arrêt de la traduction par un codon stop TAG est présent au sein de seulement 7 gènes essentiels à la survie d'*E. coli* (*coaD*, *hda*, *hema*, *mreC*, *murF*, *lola*, et *lpxK*), et 95 gènes importants pour sa croissance (Smolskaya and Andreev 2019). Une autre souche bactérienne, appelée B95ΔA, a ainsi été développée par Mukai *et al.* à partir de la souche BL21(DE3) avec le remplacement de ces 95 codons stop TAG par un codon TAA (Mukai, Hoshi, et al. 2015). Nous avons pu nous procurer cette souche et effectuer un test d'expression à l'aide des plasmides pET24d-K24TAG-K213TAG-K229TAG et pEVOL et comme nous pouvons le voir sur la Figure 77, nous avons pu observer la production spécifique de la protéine mPsaA contenant trois Propargyl-Lysines aux positions K24, K213 et K229. Nous avons ensuite voulu vérifier que cette protéine avait bien incorporé trois Propargyl-Lysines et que leurs fonctions alcynes étaient fonctionnelles pour la conjugaison.

Triple incorporation : K24-K213-K229



Mono-incorporation : K24

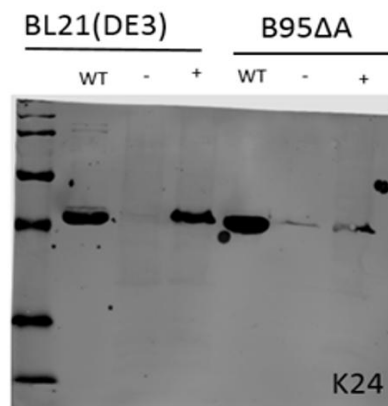


Figure 77 : Analyse par WesternBlot d'une incorporation triple ou une mono-incorporation de la Propargyl-Lysine au sein des souches BL21(DE3) ou B95ΔA

La protéine mPsaA-K24PrK-K213PrK-K229PrK (à gauche) et de la protéine mPsaA-K24PrK (à droite) ont été produites au sein des souches *E. coli* BL21(DE3) et B95ΔA. Puis un extrait total a été déposé sur gel d'acrylamide 12% et analysée par WesternBlot à l'aide d'un anticorps anti-étiquette histidine. Une amplification du signal a été réalisé pour les échantillons annotés « purif » à l'aide d'une purification sur billes de Nickel à petite échelle.

III.2 Conjugaison du Triple mutant

La conjugaison de la protéine mPsaA-K24PrK-K213PrK-K229PrK est ensuite réalisée à l'aide d'une réaction de chimie click selon le même protocole que pour les protéines mPsaA contenant une seule PrK. Cette réaction a tout d'abord été réalisée avec la fluorescéine afin de vérifier la fonctionnalité des Propargyl-Lysines. De plus, la manipulation a été effectuée en parallèle avec la mPsaA sauvage (WT) et la mPsaA-K24PrK. La conjugaison avec la protéine sauvage permet de vérifier la spécificité de la conjugaison. Quant à la mPsaA-K24PrK, elle permet de comparer la conjugaison monovalente à la conjugaison trivalente.

Comme nous pouvons le voir sur la Figure 78, la conjugaison est spécifique de la présence de la Propargyl-Lysine car aucun signal fluorescent n'est visible avec la protéine sauvage ne portant pas de fonction alcyne. De plus, l'intensité de fluorescence est plus importante pour le conjugué trivalent et la bande semble également légèrement plus haute donc de plus haut poids moléculaire ce qui indique la correcte multi-incorporation de la Propargyl-Lysine et la conjugaison de plusieurs molécules de fluorescéine sur cette protéine.



Figure 78: Conjugaison de la mPsaA sauvage, simple mutant (K24PrK) ou triple mutant (K24PrK-K213PrK-K229PrK) avec la fluorescéine.

La mPsaA sauvage et les protéines mPsaA K24TAG et Triple mutant (K24TAG, K213TAG, K229TAG) ont été conjuguées avec le tétrasaccharide (annoté sucre sur la figure) ou la fluorescéine (annoté fluo sur la figure). Un contrôle négatif (appelé neg sur la figure) a été effectué et correspond à la réaction de conjugaison réalisée sans partenaire portant une fonction azoture. Un échantillon de chacune de ces conjugaisons ainsi été déposé sur un gel d'acrylamide 12%, puis visualisé sous une lumière UV.

La conjugaison a ensuite été réalisée avec le tétrasaccharide. Bien que plus difficile à analyser, nous pouvons voir sur la Figure 79 que la bande correspondant à la mPsaA-K24PrK-K213PrK-K229PrK conjuguée au tétrasaccharide a un poids moléculaire apparemment plus élevé ce qui semble indiquer une conjugaison de plusieurs tétrasaccharides sur la protéine.

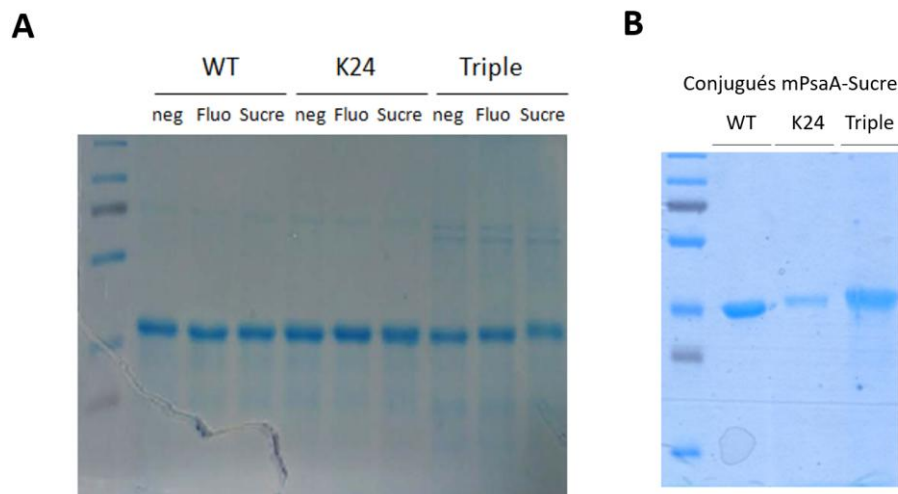


Figure 79: Conjugaison du Triple mutant mPsaA-K24PrK-K213PrK-K229PrK avec le tétrasaccharide
 La mPsaA sauvage et les protéines mPsaA K24TAG et Triple mutant (K24TAG, K213TAG, K229TAG) ont été conjuguées avec le tétrasaccharide (annoté sucre sur la figure) ou la fluorescéine (annoté fluo sur la figure). Un contrôle négatif (appelé neg sur la figure) a été effectué et correspond à la réaction de conjugaison réalisée sans partenaire portant une fonction azoture. Un échantillon de chacune de ces conjugaisons ainsi été déposé sur un gel d'acrylamide 12%, puis analysé par SDS-PAGE. Au niveau de la figure B, les conjugaisons des protéines sauvages, K24 et Triple mutant avec le tétrasaccharide ont été déposés sur gel d'acrylamide 12%.

Le conjugué trivalent a pu être produit en grande quantité (environ 0,66mg final) selon le procédé utilisé pour la production des conjugués monovalents afin de réaliser des tests d'immunisation chez les souris et sera également confirmé par Spectrométrie de Masse.

IV. Analyse de la réponse immunitaire des glycoconjugués

Afin de déterminer l'efficacité de nos conjugués, ceux-ci ont été absorbés sur du chitosane afin d'immuniser des souris. Celles-ci ont reçu trois injections sous-cutanée à J1, J14 et J28 avec une dose de 3µg de tétrasaccharide du sérotype 14 quel que soit le glycoconjugué testé et donc une dose variable de la protéine mPsaA (3 fois moins pour les conjugués portant 3 tétrasaccharides à leur surface). Les sérums ont ensuite été analysés en ELISA afin de déterminer le titre des anticorps produits dirigés contre la mPsaA ou contre le polysaccharide capsulaire ainsi que leur type (IgG ou IgM).

IV.1. Réponse immunitaire dirigée contre la mPsaA

La réponse immunitaire dirigée contre la mPsaA est constituée d'IgG. Les titres d'IgM mesurés se sont avérés nuls (d'un titre inférieur à la dilution minimale testée soit une dilution par 100, données non montrées). Comme nous pouvons le voir sur la Figure 80 et sur le Tableau 19, il existe des différences au niveau de la quantité d'IgG dirigés contre la mPsaA selon le conjugué utilisé pour l'immunisation des souris. Le conjugué K229 présente un titre d'anticorps IgG dirigé contre la mPsaA supérieur, bien que plus variable selon la souris immunisée, aux conjugués K213 et K24. Cela pourrait s'expliquer si l'on considère que cette position est plus favorable à une meilleure reconnaissance des épitopes de la protéine. Cependant, il est surprenant que le conjugué K24 induise une réponse

immunitaire réduite contre la mPsaA par rapport aux autres conjugués. En effet, il s'agit d'une position de couplage éloignée des zones immunogènes connues de la protéine et les résultats préliminaires issus de la thèse d'Aline Pillot semblaient montrer que le couplage du tétrasaccharide adjacent ou éloigné des épitopes ne diminuait pas la réponse contre la mPsaA (section I.3 et Tableau 20). Le conjugué triple qui associe les trois positions K24, K213 et K229 (PrK Triple) présente une réponse immunitaire similaire aux conjugués K24 et K213, ce qui pourrait corroborer cette hypothèse. Par ailleurs, nous pouvons noter que le conjugué aléatoire portant trois tétrasaccharides en moyenne (Random Triple) présente une réponse très faible contre la mPsaA. Il est possible que les méthodes de modifications chimiques de la protéine (dérivatisation des chaînes amines à la surface) aient modifié significativement certains épitopes et donc réduit l'immunogénicité de la protéine. De plus, il est également possible que la conjugaison aléatoire des tétrasaccharides ait entraîné le masquage de ces mêmes zones.

Tableau 19 : Titres des anticorps IgG dirigés contre la mPsaA (à J35)

Les titres d'IgG anti-mPsaA ont été déterminés par ELISA et correspondent à la dilution maximale permettant d'obtenir une valeur d'absorbance 2 fois supérieure au bruit de fond.

Random Triple	K24	K213	K229	PrK Triple
8000	4000	8000	128000	16000
<100	32000	4000	16000	4000
<100	4000	4000	32000	64000
<100	4000	16000	64000	4000
4000	16000	16000	16000	16000

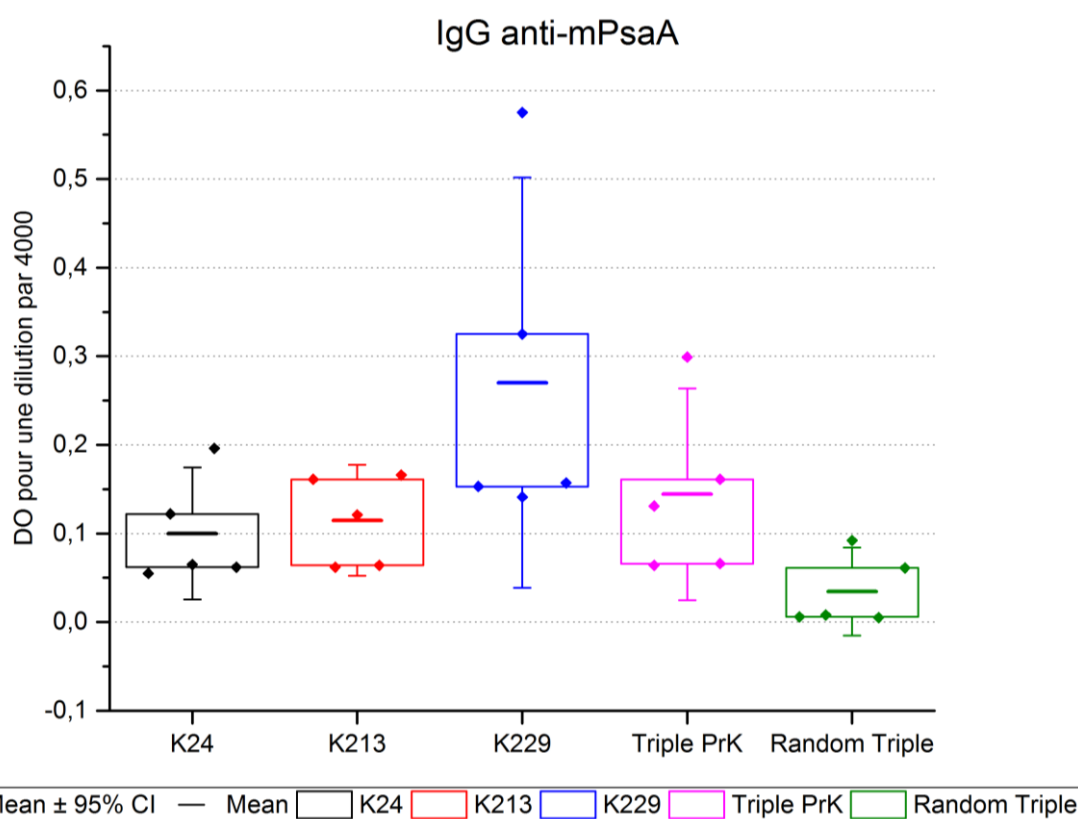


Figure 80 : Analyse de la réponse immunitaire dirigée contre la mPsaA (IgG) à J35

La réponse immunitaire a été analysée en ELISA. Pour cela, la mPsaA est adsorbée à une concentration de 1mg/mL pendant 16h à 4°C sur une plaque maxisorb. Puis les interactions non spécifiques sont limitées grâce à une incubation dans du PBST-Lait. Enfin, le sérum immunitaire est dilué et testé à différentes concentrations et les IgG sont spécifiquement détectés par la suite grâce à un anticorps secondaire dirigé contre la partie constante des IgGs et couplé à l'HRP. La lecture s'effectue ensuite à l'aide d'OPD qui se convertit en substrat chromogène grâce à l'action de l'HRP. La réaction est incubée 20 min à température ambiante puis stoppée avec de l'HCl 3M et lue à 492nm Sur ce graphique, les densités optiques obtenues à 492nm pour les différents sérums ayant été dilué par un facteur 4000. Cette dilution permet d'avoir des valeurs significatives pour chacun des échantillons. Toutes les mesures sont réalisées en duplicatas et des contrôles négatifs (sans sérum) et positifs (avec un sérum positif issu d'un projet précédent du laboratoire) sont également réalisés. Les valeurs de densité optiques obtenues pour chaque sérum testé sont indiquées par des losanges sur la figure, la moyenne pour chaque catégorie est indiquée par un trait horizontal, l'intervalle de confiance correspondant à 95% de l'échantillon est indiqué par des barres horizontales, et les boîtes représentés correspondent aux percentiles 25 et 75.

IV.2. Réponse immunitaire contre le polysaccharide capsulaire du sérotype 14

L'analyse de la réponse immunitaire dirigée contre le polysaccharide capsulaire du sérotype 14 de *S. pneumoniae* a été également réalisée. Pour cela du polysaccharide capsulaire commercial purifié a été adsorbé sur des plaques 96 puits avant une analyse ELISA. Cependant, aucun titre d'IgG dirigé contre cet antigène n'a pu être détecté car les niveaux produits sont probablement trop faibles ou inexistant. De même, les niveaux d'IgM détectés sont négligeables.

Les anticorps potentiellement produits ne sont donc pas capables de reconnaître le polysaccharide capsulaire natif. Ce tétrasaccharide est pourtant connu comme représentatif du sérotype 14 et a pu être identifié comme l'un des polysaccharides les plus immunogènes contre le sérotype 14 parmi 16 structures testées (Safari et al. 2008). De plus il a également été utilisé avec succès lors d'autres études au sein du laboratoire (Prasanna et al. 2019; Pillot et al. 2019). Une réponse contre le polysaccharide capsulaire était donc attendue. Cependant, lors de l'exposition de l'organisme à un antigène, il est possible que la réponse immunitaire soit préférentiellement dirigée

contre des zones de l'antigène appelés épitopes dominants (Wiley 2006). Les conjugués K24, K213, K229 ainsi que le triple conjugué K24-K213-K229 ont une étiquette histidine dues aux difficultés pour l'éliminer (section I.9). Des anticorps peuvent être dirigés contre celle-ci. Il est donc possible que sa présence ait dévié la réponse immunitaire (Khan et al. 2012). Cependant, le mutant aléatoire (Random Triple) ne présente pas non plus de réponse anti-polysaccharide capsulaire alors que celui-ci ne présente pas d'étiquette histidine. Une autre possibilité serait une réponse dirigée contre le cycle triazole généré au niveau de la liaison entre le tétrasaccharide et la protéine. En effet, celui-ci a été décrit comme étant immunogène (Adamo et al. 2014). De nouvelles analyses ELISA sont donc actuellement en cours afin de déterminer la réponse immunitaire dirigée contre ce lien triazole et contre l'étiquette histidine. Si cette réponse s'avère non négligeable, cela pourrait nous donner des informations sur l'intensité de la réponse obtenue selon la position de greffage utilisée et potentiellement également expliquer les différences observées au niveau de la réponse dirigée contre la mPsaA (section IV.1).

IV.3. Bilan de la réponse immunitaire

Globalement, nos résultats sont en accord avec les différentes études bibliographiques sur la position de greffage du polysaccharide sur la protéine et sur le ratio protéine / polysaccharide (Ishioka et al. 1992; Avci et al. 2011; Pozsgay et al. 1999; G Stefanetti et al. 2014; Giuseppe Stefanetti et al. 2015). La réponse immunitaire semble bien impactée par la position de greffage et par le ratio protéine / polysaccharide. Néanmoins, les caractéristiques précises qui permettraient une réponse optimales ne sont pas connues.

La génération de ces vaccins glycoconjugués monovalents ciblés et trivalents ciblés au cours de ce travail de thèse a été réalisée dans le contexte de deux études précédentes réalisées au laboratoire (Prasanna et al. 2019; Pillot et al. 2019). Nous allons donc particulièrement mettre en relief nos résultats par rapport à ces deux précédentes études. Les paramètres des différents vaccins glycoconjugués utilisés ainsi que les résultats principaux sont décrits au sein du Tableau 20. Au cours d'une première étude réalisée par Prasanna *et al.* la mPsaA a été couplée de façon aléatoire à la protéine avec un ratio de 5,4 tétrasaccharides pour 1 mPsaA. La réponse dirigée contre la mPsaA s'est avérée amoindrie avec le conjugué mPsaA-PnTS14 par rapport à une injection avec la mPsaA seule. Quant à la réponse dirigée contre le polysaccharide capsulaire (anti-CP14), les titres d'anticorps ont été évalués à environ 8 000 ce qui est équivalent aux études bibliographiques utilisant cet antigène, même s'il faut noter que les doses utilisés, la chimie de conjugaison et la nature des protéines porteuses sont différentes ce qui a un rôle non négligable (Kurbatova et al. 2019). La baisse de la réponse anti-mPsaA au niveau du conjugué peut s'expliquer, entre autre, par un masquage des épitopes (par dérivatisation ou à cause du couplage), ou par un phénomène de dominance des épitopes saccharidiques sur les épitopes protéiques. Une deuxième étude réalisée par Pillot *et al.* permet alors de réaliser une première étude sur la position de greffage du tétrasaccharide sur la protéine. Pour cela, des conjugués ciblés ont été réalisés grâce à des mutations cystéines. Quatre positions ont pu être testées, dont une éloignée des séquences immunogènes connus (K309C). La réponse immunitaire dirigée contre la mPsaA s'est avérée généralement forte avec une réponse maximale pour le conjugué K309C, ce qui semble indiquer que la présence d'un unique tétrasaccharide en dehors ou à proximité d'un épitope T n'a pas d'effet délétère sur les épitopes protéiques. Quant à la réponse contre le polysaccharide capsulaire, celle-ci s'est avérée plutôt faible avec une réponse plus faible pour le conjugué K309C. Ces résultats ne permettent donc pas de déterminer si la baisse de la réponse immunitaire contre la mPsaA pour les conjugués K237C, K242C et K247C sont dues à un masquage d'épitopes ou à un phénomène de dominance d'épitopes.

Lors de notre étude, des conjugués ciblés entre la mPsaA et le tétrasaccharide ont été réalisés à l'aide d'acides aminés non naturels aux positions K24, K213 et K229 avec la position K24 qui est plutôt éloignée des épitopes connus (B et T) de la mPsaA. Un conjugué trivalent a également été produit en combinant ces trois positions, et un conjugué trivalent aléatoire (appelé "Random Triple") a également été réalisé. La réponse dirigée contre la mPsaA est plutôt moyenne comparée aux réponses obtenues lors des deux études précédentes. Cette réponse est néanmoins d'intensité comparable pour les conjugués monovalents K24, K213 et le conjugué trivalent K24-K213-K229. La réponse anti-mPsaA la plus forte est obtenue pour le conjugué K229 tandis que la réponse est très faible, voire négligeable pour le conjugué Triple aléatoire. Etant donné que la réponse contre le conjugué K24 est plutôt réduite malgré la conjugaison éloignée des séquences immunogènes connus, cela suggère que la baisse de la réponse n'est pas due à un masquage d'épitope comme potentiellement suggéré par les travaux d'Aline Pillot. Cependant, nous pouvons noter que les intensités de réponses anti-mPsaA sont difficilement comparables au niveau de ces deux études de par la différence de quantité de protéine injecté. Nous pouvons néanmoins noter la réponse très faible au niveau du conjugué trivalent aléatoire. Ce résultat peut indiquer que la dérivatisation de la protéine peut engendrer une perte partielle de l'immunogénicité de la protéine par une modification des épitopes. L'utilisation d'acides aminés non naturels ou de mutants cystéines apparaît donc comme une méthode très appréciable.

Cependant, la réponse immunitaire dirigée contre le polysaccharide s'est malheureusement avérée inefficace. Ce résultat ne va donc pas dans le sens d'un effet d'immunodominance. Cependant, il est possible que le lien triazole généré lors de la conjugaison ait dévié la réponse immunitaire. Des analyses sont actuellement en cours afin de vérifier cette hypothèse. D'autres acides aminés non naturels peuvent être utilisés afin de générer des liens moins immunogènes, comme l'Allyl-Lysine qui permet de réaliser des réactions de type thiol/ènes. Des tests d'incorporation de cet acide aminé ont été réalisés avec succès au niveau des trois positions ciblées au sein de la PsaA (données non montrées). Il est donc possible de reproduire ce travail afin d'éviter cette réponse non spécifique. De plus, l'utilisation de cet acide aminé permettrait d'éviter l'utilisation de cuivre nécessaire à la réaction de conjugaison CuAAC, qui peut être délétère pour des applications *in vivo*. Le taux de cuivre résiduel n'a d'ailleurs pas été mesuré mais des expériences afin de l'évaluer seront prochainement réalisées.

Tableau 20 : Comparaison des résultats pour des immunisations avec les précédentes études

	Conjugués aléatoires	Conjugués cystéine	Conjugués NAA
Partenaires	mPsaA et PnTS14	mPsaA et PnTS14	mPsaA et PnTS14
Méthode de couplage	Maleimide/thiol après dérivatisation de la mPsaA	Maleimide/thiol après mutation Cystéine	CuAAC après incorporation de NAA Ou CuAAC après dérivatisation de la mPsaA (Random Triple)
Distribution de la protéine	Aléatoire	Ciblée aux positions K237C, K242C, K247C, K309C	Ciblée aux positions : K24, K213, K229 ou K24-K213-K229 (Triple NAA) ou aléatoire (Random Triple)

Ratio PnTS14/mPsaA	5.4/1	1/1	1/1 ou 3/1
Présence d'une étiquette histidine	Non	Non	Oui sauf pour le trivalent aléatoire (Random Triple)
Adjuvant	Oui α Gal-Cer (cible les iNKT)	Non	Non
Dose de protéine	25 μ g	25 μ g	150 μ g ou 50 μ g
Dose de sucre	3 μ g	0,5 μ g	3 μ g
Schéma d'immunisation	J0, J21, J90, J120	J0, J14, J28	J0, J14, J28
Prélèvements analysés en ELISA	J28, J98, J128	J21, J35	J21, J35
Résultats réponse anti-mPsaA	Réponse plus faible pour le conjugué mPsaA-PnTS14, Réponse présente dès J28	Réponse forte contre la mPsaA, avec la position K309C ayant la meilleure réponse	Réponse moyenne, avec la position K229 ayant la meilleure réponse
Résultats réponse anti-CP14	Réponse forte	Réponse faible, réponse la plus faible pour K309C	Absence de réponse anti-CP14
Référence	(Prasanna et al. 2019)	(Pillot et al. 2019)	/

La réponse immunitaire a été évaluée par une titration des anticorps grâce à une analyse ELISA. Une autre méthode d'étude *in vitro* est également envisagée afin d'évaluer la réponse grâce à des tests d'opsonophagocytose. Pour cela, le sérum est mis en contact avec *S. pneumoniae* tué et des macrophages activés. Si des anticorps reconnaissent la bactérie, ceux-ci vont s'y agglutiner et seront ensuite reconnus par les macrophages afin de stimuler la phagocytose de celles-ci. Il s'agit alors d'un test plus fonctionnel. En effet, la présence d'anticorps est un indice important mais ils peuvent être plus ou moins efficaces pour l'élimination de la bactérie. Cette méthode permettrait alors de vérifier cette efficacité.

Par ailleurs, nous pouvons noter que la réactivité croisée des anticorps produits entre différents sérotypes capsulaires n'a pas été vérifiée. Cette réactivité croisée est néanmoins suggérée grâce à la présence d'anticorps dirigés contre la mPsaA qui ont déjà été montrée dans la littérature comme efficace contre différents sérotypes. Cependant, il serait également possible de vérifier cette assertion par des expériences d'ELISA ou d'opsonophagocytose.

Conclusion

Dans un contexte de résistance croissante aux antibiotiques, les vaccins contre *S. pneumoniae* fournissent un rempart décisif contre ces pathologies. Cependant, la variabilité des sérotypes entraîne des difficultés pour fournir un vaccin ayant une couverture sérotypique optimale. En effet, les sérotypes qui ne sont pas couverts par les vaccins ont tendance à se développer et à entraîner des pathologies pouvant conduire à l'utilisation d'antibiotiques et donc à l'aggravation du phénomène de résistance. Augmenter la couverture sérotypique des vaccins actuels est donc un des objectifs prépondérants des nouvelles stratégies vaccinales. Deux types de vaccins sont présents sur le marché : les vaccins polysaccharidiques et les vaccins glycoconjugués. Néanmoins, seuls les vaccins glycoconjugués apportent une protection chez les jeunes enfants, qui constituent une part importante de la population à risque. Cependant, les méthodes actuellement employées pour produire ces vaccins glycoconjugués ne permettent pas de contrôler la position ou le nombre de tétrasaccharides sur la protéine porteuse et fournissent des antigènes vaccinaux non homogènes. Ces caractéristiques sont pourtant importantes pour la qualité de la réponse immunitaire.

Par cette étude, nous avons alors cherché à atteindre 3 objectifs :

- Objectif 1 : Fournir un vaccin contre *S. pneumoniae* ayant une couverture sérotypique plus large.
- Objectif 2 : Fournir un glycoconjugué site-spécifique et homogène afin de permettre l'étude de l'influence de la position de greffage du saccharide sur la réponse immunitaire.
- Objectif 3 : Fournir un glycoconjugué multivalent homogène afin de permettre l'étude de l'influence du ratio sucre greffé/protéine porteuse sur la réponse immunitaire.

Notre stratégie a consisté à coupler le PnTS14 avec la mPsaA afin d'atteindre l'objectif 1. Une réactivité croisée des anticorps anti-mPsaA pour plusieurs sérotypes est implicite mais n'a pas été étudiée au sein de cette étude. Le cœur de ces travaux est centré sur la génération de glycoconjugués monovalents et trivalents ciblés et homogènes (objectif 2 et 3). Pour cela, nous avons utilisé la technologie d'incorporation d'acides aminés non naturels afin de fournir des vaccins glycoconjugués homogènes et dont le nombre et la position des unités saccharidiques greffées sont contrôlés. Cette stratégie de conjugaison originale a encore été très peu utilisée afin de fournir des vaccins (Fairman, Heinrichs, and Chan 2018; Kapoor et al. 2018). Nous avons alors produit trois conjugués monovalents homogènes (K24, K213 et K229) et un conjugué trivalent homogène (K24-K213-K229). Ces conjugués ont tous été produits en quantité suffisante pour réaliser des tests d'immunogénicité sur des souris. A notre connaissance, il s'agit du premier exemple utilisant la technologie des acides aminés non naturels *in vivo* afin de générer des vaccins glycoconjugués ayant des positions contrôlées. De plus, il s'agit également du premier exemple de vaccin glycoconjugué multivalent homogène généré par cette technologie. L'utilisation de la méthode d'incorporation des acides aminés non naturels est donc un outil innovant permettant de maîtriser des paramètres de conjugaison jusqu'alors difficilement étudiables. Les résultats concernant la réponse immunitaire présentent des différences selon les conjugués utilisés, ce qui semble indiquer que la position de

greffage a impacterait la réponse produite. Néanmoins, cela nécessite des analyses complémentaires afin de pouvoir expliquer ces différences observées au niveau de la réponse immunitaire anti-mPsaA.

La méthode d'incorporation des acides aminés non naturels est une méthode présentant des avantages importants lorsqu'il s'agit de générer des conjugués ciblés sans modifier la protéine porteuse. De plus, cette méthode peut s'adapter à de nombreuses protéines vaccinales issues de différents organismes. La technologie d'incorporation d'acides aminés non naturels s'ajoute donc aux outils maîtrisés au laboratoire afin de fournir des vaccins conçus de manière rationnelle. Ces vaccins permettront alors d'étudier les paramètres de couplage et à terme de déduire leurs effets sur la réponse immunitaire dont beaucoup d'éléments restent encore incompris.

Produire des protéines contenant un acide aminé non naturel en grande quantité est néanmoins difficile et donc limitant pour la génération de protéines vaccinales. Lorsque l'on cherche à conjuguer deux éléments ciblés sur une protéine, l'incorporation de deux acides aminés non naturels de manière site-spécifique s'avère encore plus difficile. Afin de pouvoir conjuguer plusieurs éléments sur une protéine porteuse, il peut alors être intéressant de coupler cette technologie avec l'utilisation de mutants cystéines lorsque c'est possible ou avec l'incorporation résidu-spécifique d'un autre acide aminé non naturel. La méthode d'incorporation de manière résidu-spécifique n'a cependant pas encore été implémentée au sein du laboratoire mais son utilisation conjointe ou complémentaire à l'incorporation site-spécifique a un fort potentiel. Par exemple, afin de conjuguer plusieurs éléments sur la PsaA, comme par exemple un adjuvant, il est également possible de coupler la méthode d'incorporation d'acides aminés non naturels avec la mutagenèse d'acides aminés en cystéines. En effet, la PsaA ne possède naturellement aucune cystéine et des mutants cystéines ont été utilisés avec succès au sein du laboratoire afin de fournir des conjugués spécifiques. Cela permettrait alors de générer un vaccin composé de trois éléments (protéine, saccharides, adjuvant) dont les paramètres de couplage resteraient maîtrisés.

***Chapitre 4 : Matériel et Méthode pour la
génération de vaccins glycoconjugués***

I. Synthèse de la Propargyl-Lysine (PrK) : (échelle 500 mg)

La synthèse ci-dessous est présentée à une échelle de 500mg, cependant cette synthèse a également été montée en échelle (jusqu'à 2g) mais en gardant les mêmes proportions et sans adaptation particulière.

Etape (1) : synthèse du composé intermédiaire

Cinq cent milligrammes de Boc-Lys-OH (soit 2.03mmol) sont dissous dans une solution de 1M NaOH (5mL) et THF (5mL) puis la solution est refroidie à 0°C à l'aide d'un bain de glace. Ensuite, du propargylchloroformate (158µl, soit 192,5 mg, 1,62mmol) est ajouté goutte à goutte. La solution est ensuite laissée à température ambiante sous agitation sur la nuit (environ 10h). Puis, la solution obtenue est refroidie.

La solution refroidie est alors lavée avec de l'éther froid (50mL), afin d'éliminer les réactifs résiduels. La phase étherée est retirée de l'ampoule. La phase aqueuse restante est acidifiée avec de l'acide chlorhydrique froid (1M, 50mL). Le produit est ensuite extrait de la phase aqueuse vers la phase organique par l'ajout d'acétate d'éthyle froid (2 fois 30mL). La phase organique est récupérée dans un erlenmeyer et contient le produit.

Enfin, cette phase organique est séchée par l'ajout de poudre MgSO₄ et filtrée, puis les solvants sont évaporés à l'aide de l'évaporateur rotatif. Le produit intermédiaire est alors obtenu et se présente sous la forme d'une huile marron-orangée. Environ 50mg de ce produit ont été dissouts dans du CDCl₃ (Chloroforme) pour analyse par RMN.

Etape (2) : dé-protection et production de la molécule finale :

Le composé produit précédemment est dissout dans 4mL de Dichlorométhane (milieu anhydre, sous argon). Puis, 4mL de TFA est ajouté goutte à goutte et la réaction est incubée à température ambiante pendant 2h sous agitation. Enfin, les solvants sont évaporés à l'aide de l'évaporateur rotatif et le produit est précipité au froid grâce à l'ajout de diéthyléther, puis filtré et séché afin d'obtenir l'acide aminé non naturel pur sous forme d'un solide blanc. Une analyse RMN a été effectuée et a permis de valider la synthèse de la propargyl-Lysine.

La propargyl-Lysine est ensuite dissoute dans de l'eau à une concentration de 100mM et stockée sous forme d'aliquot à -20°C jusqu'à son utilisation.

II. Test de pénétration de la Propargyl-Lysine dans les bactéries

L'entrée de la propargyl-lysine dans les bactéries a été vérifiée par une analyse sur chromatographie sur couche mince (CCM). Pour cela, une culture de 5mL a été réalisée selon le protocole (section III.3) en absence et en présence de 10mM de Propargyl-Lysine. La culture a ensuite été culotée par centrifugation à 5000g pendant 10min, puis le culot a été lavé 3 fois avec 5mL de tampon froid (Na₂HPO₄/NaH₂PO₄ 50mM, NaCl 150mM, pH8). Le culot a ensuite été resuspendu dans 300µL de ce tampon et a été lysé par sonication (30 secondes, cycles 5sec-5sec, amplification 50%, appareil : Thermo Fischer, FB120-220), puis centrifugé à 10 000rpm pendant 15 minutes à 4°C. Enfin 1µL du surnageant de lyse a été déposé sur CCM (Eluant : Butanol, Acide acétique, Eau (proportions 10/2,5/2,5), et révélateur : Ninhydrine) en parallèle de la PrK pure.

III. Production et analyse des protéines recombinantes et vaccins glycoconjugués

III.1 Transformation au sein des souches d'expression

Les souches d'expression pour la production des protéines recombinantes utilisées lors de ce travail sont BL21(DE3) (Merck Novagen, VWR, 69450) et B95ΔA (souche issue du laboratoire de Sakamoto, Japon, (Mukai, Hoshi, et al. 2015).

Le génotype des souches est le suivant :

-BL21(DE3) : *E. coli* str. B, F⁻ *ompT gal dcm lon hsdS_B(r_B⁻m_B⁻) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB⁺]_{K-12}(λ^S)*

Cette souche est une souche couramment utilisée pour la production de protéines recombinantes.

-B95ΔA : *E. coli* str. BL21(DE3), ΔRF1, Δ95TAG

Concernant la souche B95ΔA, le gène codant pour le Releasing Factor 1 (RF1) a été délété dans le génome et 95 codons stop « ambre » TAG parmi les 273 présents dans le génome ont été remplacés à l'aide de mutations silencieuses. Ces modifications du génome bactérien ont pour objectif de favoriser l'incorporation multiple d'acides aminés non naturels au niveau du codon TAG dans les protéines cibles.

Ces souches sont rendues compétentes à l'aide de la méthode TB+. Pour cela, une préculture est diluée au centième dans du LB, puis incubée à 37°C jusqu'à que la densité optique à 600nm soit comprise entre 0,45-0,5 (début de phase exponentielle). La croissance de la culture est ensuite arrêtée par incubation dans la glace pendant 10 min et centrifugation des cellules pendant 10 min à 2500rpm à 4°C. Le culot bactérien est ensuite resuspendu dans 16% du volume de culture initial dans du TB+ (Pipes 10mM, CaCl₂ 15mM, KCl 250mM, Mn²⁺ 55mM). Les cellules sont ensuite incubées dans la glace pendant 10min avant d'être centrifugées à nouveau à froid à 2500rpm pendant 10min. Le surnageant est éliminé et le culot est de nouveau resuspendu dans 4% du volume de culture initiale. Enfin, 7% de DMSO final est ajouté et les cellules sont aliquotées par 100μL puis congelées pour un usage ultérieur. La compétence des cellules est vérifiée par transformation de 1μL d'un plasmide test pET11a (50ng) avant validation.

III.2 Production des protéines sauvages (mPsaA WT)

La production de la protéine recombinante mPsaA sauvage (WT) s'effectue au sein de la souche BL21(DE3). Pour la partie I.5, la production a été réalisée à l'aide du plasmide pET11a-mPsaA-WT (portant une résistance à l'ampicilline et l'étiquette Histidine en N-terminal de la séquence protéique d'intérêt), alors que pour la partie I.6, la production a été réalisée à l'aide du plasmide pET24d-mPsaA-WT (portant une résistance à la kanamycine et l'étiquette Histidine en C-terminal de la séquence protéique d'intérêt). Les séquences plasmidiques sont présentées en Annexe 2 et 3.

A partir d'une colonie bactérienne, une préculture est réalisée avec 5mL de milieu LB en présence d'un antibiotique de sélection selon la construction plasmidique utilisée (ampicilline 100μg/mL pour la production avec les plasmides pET11a et kanamycine 25μg/mL pour la production avec les plasmides pET24d) et incubée sur la nuit à 37°C sous agitation. Le lendemain, la préculture est diluée au centième dans du milieu auto-inductif (FORMEDIUM, n° AIMLB0210) en présence d'un antibiotique de sélection puis incubée 24h à 37°C sous agitation.

III.3 Production des protéines mutées contenant une Propargyl-Lysine (mPsaA-K24PrK, mPsaA-K213PrK, et mPsaA-K229PrK)

La production s'effectue au sein de la souche BL21(DE3) pour les protéines mPsaA-K24PrK, mPsaA-K213PrK et mPsaA-K229PrK et au sein de la souche B95ΔA pour la protéine mPsaA-K24PrK-K213PrK-K229PrK.

Au sein de la souche d'expression, deux plasmides ont été co-transformés : un plasmide d'expression (pET11a ou pET24d) et un plasmide portant un ARNt et une aaRS orthogonaux, les éléments nécessaires à l'incorporation de l'acide aminé non naturel (pEVOL). Pour la partie I.5, la production a été réalisée à l'aide du plasmide d'expression pET11a-mPsaA-K24, K213 ou K229 en co-transformation avec le plasmide pEVOL alors que pour la partie I.6, la production a été réalisée à l'aide du plasmide pET24d-mPsaA-K24, K213 ou K229 en co-transformation avec le plasmide pEVOL. Les séquences plasmidiques sont présentées en Annexe 2 et 3.

A partir d'une colonie bactérienne, une pré-culture est réalisée avec 5mL de milieu LB en présence d'antibiotiques de sélection (ampicilline 100μg/mL et chloramphenicol 30μg/mL pour la production avec les plasmides pET11a et pEVOL ou kanamycine 25μg/mL et chloramphenicol 30μg/mL pour la production avec les plasmides pET24d et pEVOL) et incubée sur la nuit à 37°C sous agitation. Le lendemain, la pré-culture est diluée au centième dans du milieu auto-inductif (FORMEDIUM, n° AIMLB0210) en présence d'antibiotiques de sélection, de PropargylLysine à 1mM et d'arabinose à 0.02% final puis incubée 24h à 37°C sous agitation.

III.4 Vérification de l'incorporation de l'acide aminé non naturel par SDS-PAGE/WesternBlot

Des cultures sont réalisées selon le protocole précédent, avec ou sans acide aminé non naturel afin de vérifier son incorporation par une analyse SDS-PAGE.

Les échantillons composés de 10μL de culture avec 5μL d'eau et 5μL de tampon de charge (bleu de bromophénol, SDS, et βmercaptoéthanol) sont chauffés 5min à 100°C avant d'être déposés sur gel d'acrylamide 12%.

Pour une analyse des protéines totales par SDS-PAGE, le gel est coloré dans du Bleu de Coomassie pendant 1h puis décoloré pendant 1h dans de l'éthanol 20%.

Pour une analyse de la protéine portant l'étiquette histidine en Westernblot, le protocole suivant est réalisé. Les protéines sont transférées du gel de SDS-PAGE sur une membrane de nitrocellulose (PALL, BioTrace 66485) à l'aide d'un appareil de transfert (Thermo scientifique, Pierce Power Blotter), suivant le protocole de transfert des protéines de tailles mixtes (7min, 1,3A, 25V). La membrane est ensuite incubée 2h dans du TBS-Tween 0,1%-Lait5% pour limiter les interactions non spécifiques, puis lavée dans du TBS-Tween 0,1% avant une incubation sur la nuit dans l'anticorps primaire souris anti-étiquette Histidine (anticorps monoclonal sigma H1029, concentration 1/1000, TBS-Tween 0,1%-BSA-5%). La membrane est lavée 3 fois avec du TBS-Tween 0,1% pendant 15min. La membrane est ensuite incubée dans l'anticorps secondaire (anticorps chèvre anti souris couplé à un fluorophore Alexa 680 (Invitrogen A21058) à une dilution 1/10 000, TBS-Tween 0,1%-BSA-5%) pendant 1h à température ambiante. Enfin, après 3 lavages dans du TBS-Tween 0,1% durant 15min, la membrane est lue à l'aide de l'appareil de lecture (Odyssey CLX, LI-COR).

III.5 Purification des protéines recombinantes

Les protéines recombinantes qui ont été produites lors de ce projet possèdent un tag histidine qui a été utilisé afin de les purifier par chromatographie d'affinité sur colonne de Nickel selon le protocole suivant.

La culture (section III.3) a été centrifugée à 5000g pendant 10min, puis le culot a été resuspendu dans du tampon de lyse ($\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ 50mM, NaCl 150mM, pH8, 5mM imidazole, PMSF 0,2mM), et les cellules ont été lysées par sonication (5min, cycles 5sec-5sec, amplification 50%, appareil : Thermo Fischer, FB120-220). Le lysat a ensuite été centrifugé 30 minutes à 10 000rpm à 4°C, puis le surnageant de lyse cellulaire a été filtré sur une membrane de 0,45 μm avant d'être incubé 1h avec des billes de Nickel (Macherey Nagel, Protino) conditionnées dans du tampon ($\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ 50mM, NaCl 150mM, pH8 imidazole 5mM). La purification a ensuite été effectuée à l'aide de différentes solutions contenant des concentrations d'imidazole croissantes (5mM à 300mM). Chaque fraction d'élution a été récupérée et pour chaque fraction un aliquot de 15 μL a été analysé sur gel SDS-PAGE 12%. Les fractions contenant la protéine d'intérêt pure ont ensuite été dialysées dans du PBS à l'aide de membranes de dialyses (tailles des pores à 6-8000 Da) afin d'éliminer l'imidazole. Les fractions ont ensuite été utilisées pour valider la protéine par spectrométrie de masse ou pour une analyse en dichroïsme circulaire ou enfin pour réaliser une conjugaison par chimie click.

III.6 Dichroïsme Circulaire

Après purification, la protéine mPsaA a été analysée par dichroïsme circulaire, à l'aide de l'appareil Jasco-J810. Pour avoir un signal exploitable et reproductible, les protéines ont été diluées pour obtenir un signal aux alentours de -20/-30 à une longueur d'onde de 220nm, soit une concentration de 0,09mg/mL pour la mPsaA. Les mesures de concentrations ont été réalisées au Nanodrop en utilisant les valeurs caractéristiques de la mPsaA ($\epsilon=37,360\text{cm}^{-1}\text{M}^{-1}$ et $M=34,142\text{kDa}$). Les signaux de différentes conditions ont ensuite été comparés par superposition des courbes obtenues.

III.7 Conjugaison des protéines avec le tétrasaccharide- N_3 ou avec la fluorescéine- N_3

Afin de produire le vaccin glycoconjugué, la protéine mPsaA a été conjuguée avec un tétrasaccharide constitutif de la capsule du sérotype 14 de *S. pneumoniae* (Pn14TS). La protéine mPsaA a également été conjuguée dans certaines expériences avec la fluorescéine afin de tester la fonctionnalité des groupements alcynes portés par l'acide aminé non naturel incorporé.

Une fois purifiées les protéines sont conjuguées à du tétrasaccharide- N_3 ou avec la fluorescéine- N_3 par réaction de conjugaison catalysée par le cuivre I (aussi appelée copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC)) selon le protocole qui a été décrit dans la publication de Presolski et al. (Presolski, Hong, and Finn 2011) En bref, 432,5 μL de protéine purifiée (dont la concentration varie entre 0,2 et 2mg/mL) est mélangée à 10 μL de tétrasaccharide- N_3 ou avec la fluorescéine- N_3 à 5mM, ainsi qu'à un mélange de 2,5 μL de CuSO_4 à 50mM et 5 μL de THPTA à 50mM, à 25 μL de solution d'aminoguanidine à 100mM et à 25 μL de solution d'ascorbate de sodium à 100mM. La réaction a ensuite été incubée 3h à 24h à température ambiante. Cinquante microlitres d'EDTA à 0,5M sont alors ajoutés à la réaction afin de piéger le cuivre (catalyseur de la réaction) et de stopper la réaction. Les échantillons sont ensuite purifiés sur colonne d'exclusion stérique (section suivante III.8).

III.8 Purification sur colonne d'exclusion stérique et lyophilisation

Après la réaction de conjugaison, les tétrasaccharides non conjugués et les réactifs de conjugaison sont rapidement éliminés à l'aide d'une purification sur une colonne d'exclusion stérique (GE, Superdex 200, 16/600). La colonne est d'abord équilibrée à l'aide de tampon PBS avec un débit

de 0,8mL/min, puis l'échantillon est injecté dans la colonne (volume d'échantillon 2-4mL, boucle d'injection de 5mL) et l'échantillon est élué avec 450mL de PBS en appliquant un débit de 0,5-0,8mL/min. L'apparition de pics d'éluion de protéines est suivie à l'aide d'un détecteur à 280nm (Biologics Optics Module II OM-11). Les fractions contenant la protéine d'intérêt sont ensuite récupérées et vérifiées par une analyse sur SDS-PAGE.

III.9 Purification des glycoconjugués sur Galectine 3 fixée sur colonne NHS

L'incorporation d'un acide aminé non naturel par le détournement d'un codon stop TAG n'est pas efficace à 100% et entraîne la production d'un mélange de la protéine cible complète (ayant incorporé le NAA) et de la protéine cible tronquée (dont la traduction s'est arrêtée au codon stop TAG). Lors de l'utilisation de la construction plasmidique pET11a (avec l'étiquette histidine en N-terminal), les deux types de protéines cibles sont produites et purifiées à l'aide de la purification sur colonne de Nickel car elles possèdent toutes les deux l'étiquette histidine (présent en N-terminal). Une construction comportant l'étiquette histidine en C-terminal a donc été développée (section III.1). Cependant, la coupure de cette étiquette par la TEV (section III.10) n'est pas efficace sur la construction possédant l'étiquette histidine en C-terminal. Une stratégie consistant à purifier les protéines ayant été conjuguées avec le tétrasaccharide a alors été envisagée. En effet, la Galectine 3 (une protéine produite au laboratoire) reconnaît le tétrasaccharide. La conjugaison par chimie click étant très spécifique, uniquement les protéines ayant incorporé l'acide aminé non naturel seraient alors purifiées avec cette méthode. De plus, en cas de réaction de conjugaison incomplète cela permet d'isoler les protéines ayant été conjuguées efficacement. Ces manipulations ont été réalisées par une technicienne du laboratoire (Annie Lambert). La Galectine 3 est d'abord fixée aspécifiquement sur une colonne NHS (GE, HiTrap, 1mL, 0716-30). Pour cela, la colonne NHS a été lavée 3 fois avec 2mL d'HCl 1mM. La colonne est ensuite chargée 4 fois avec de la Galectine 3 par une incubation de 15 minutes à 4°C puis 15 minutes à température ambiante avec la solution de Galectine-3 à environ 8mg/mL (préalablement produite au sein du laboratoire, dans un tampon Carbonate de sodium 0,2M, pH 8,3; NaCl 0,2M, lactose 0,1M) pour un chargement final d'environ 30-40mg par colonne. La colonne est ensuite lavée et désactivée à l'aide de 3 fois 2mL de tampon d'Ethanolamine 0,5M, NaCl 0,5M, et pH 8,3 suivi de 3 fois 2mL de tampon Acétate de sodium 0,1M, NaCl pH 4 et de nouveau 3 fois 2mL de tampon d'Ethanolamine 0,5M, NaCl 0,5M, et pH 8,3. Après une incubation d'une demi-heure à température ambiante, ces étapes de lavages sont répétées. Enfin un tampon d'équilibration (PBS 1mM EDTA) est injecté avant de procéder à la purification. Pour réaliser la purification, la solution protéique contenant la protéine conjuguée avec le tétrasaccharide est ensuite injectée sur la colonne. Puis, la colonne est lavée avec 5mL de tampon (PBS, 1mM EDTA) pour éliminer les protéines interagissant de façon aspécifique. Enfin la protéine retenue est éluée à l'aide d'une solution contenant du lactose (PBS, 1mM EDTA, 100mM lactose). Les fractions sont ensuite analysées sur SDS-PAGE (gel d'acrylamide 12%).

III.10 Coupure de l'étiquette histidine par la protéase TEV

Les protéines produites lors de ce projet sont destinées à la génération d'un vaccin glycoconjugué suivi d'une immunisation chez l'animal. Cependant il est possible que l'étiquette histidine soit reconnue par le système immunitaire et que des anticorps soient produits contre cette étiquette au détriment de la protéine mPsaA et du tétrasaccharide. Nous avons donc cherché à enlever cette étiquette à l'aide de la protéase du Tobacco Etch Virus (TEV protéase, produite au sein du laboratoire) spécifique pour le site de coupure ENLYFQ|S que nous avons inséré dans nos constructions plasmidiques entre la séquence de la protéine mPsaA et l'étiquette histidine. Pour cela,

les protéines ont été dialysées dans un tampon dans lequel la protéase TEV est efficace (Tris-HCl 50mM, EDTA 0,5M, DTT 1mM), puis incubées avec la protéase (le rapport protéine à clivée : protéase étant de 20 pour 1) pendant une nuit à température ambiante. La coupure est ensuite vérifiée par une analyse SDS-PAGE.

III.11 Production d'un vaccin glycoconjugué à l'aide de la méthode aléatoire

Afin de comparer les vaccins conjugués ciblés générés grâce à l'utilisation de la Propargyl-Lysine (méthode sections III.3 à III.9) avec des vaccins glycoconjugués générés avec une méthode aléatoire (basée sur la réactivité des lysines), un vaccin glycoconjugué à l'aide de cette dernière méthode a été généré en suivant le protocole ci-après.

Génération de fonction propargyl à partir des lysines présentes à la surface de la mPsaA :

Du propargylsuccinimide (0,05 mmol, dissout dans l'acétonitrile) est progressivement (un tiers du réactif toutes les 45min) ajouté à la protéine mPsaA sauvage (10mg, soit environ 0,3 μ mol de protéine, mais comme la protéine possède 38 lysines, cela correspond à environ 11,4 μ mol de lysines bien que leurs disponibilités par rapport au solvant soient variables) dans du tampon Phosphate à 0,2M. Puis 45 min après le dernier ajout, la protéine est dialysée afin d'éliminer le propargylsuccinimide n'ayant pas réagi ainsi que l'acétonitrile.

Conjugaison de la mPsaA-dérivée avec le tétrasaccharide :

Afin de générer un vaccin glycoconjugué de façon aléatoire comparable au conjugués formés à l'aide de l'incorporation de trois Propargyl-Lysines au sein de la séquence (mPsaA-K24-Pn14TS-K213-Pn14TS-K229-Pn14TS), la réaction de conjugaison est réalisée en suivant le protocole de la section III.7, mais en utilisant uniquement trois équivalents de tétrasaccharide (Pn14TS) par rapport à la protéine (mPsaA). La limitation de la quantité de tétrasaccharide dans le mélange réactionnel devrait permettre de produire un vaccin conjugué de façon aléatoire mais avec seulement trois tétrasaccharides greffés au sein de la séquence (mPsaA-3Pn14TS). Cette protéine conjuguée a ensuite été vérifiée par spectrométrie de masse.

IV. Biologie moléculaire – Préparation des gènes de la protéine cible mPsaA

II.1 Clonage du gène de la mPsaA dans le plasmide pET24d

Le gène de la mPsaA ou Pneumococcal Adhesin A de *S. pneumoniae* sous sa forme mature, c'est-à-dire sans le peptide signal qui la rend insoluble, a été commandé chez Genscript, cloné dans le plasmide pET11a avec une étiquette de 6 histidines en N-terminal. Cependant, l'incorporation d'un acide aminé non naturel par le détournement d'un codon stop TAG n'est pas efficace à 100% et cela entraîne la production d'un mélange de la protéine cible complète (ayant incorporé le NAA) et de la protéine cible tronquée (dont la traduction s'est arrêtée au codon stop TAG). Lors de l'utilisation de la construction plasmidique pET11a (avec l'étiquette histidine en N-terminal), les deux types de protéines cibles sont produites et purifiées par chromatographie d'affinité sur colonne de Nickel car elles possèdent toutes les deux l'étiquette histidine (présent en N-terminal). Afin de pouvoir séparer plus facilement les formes tronquées des formes complètes, le gène de la mPsaA a été recloné au sein du plasmide pET24d (avec l'étiquette histidine en C-terminal), ainsi seulement les protéines cibles ayant incorporé l'acide aminé non naturel sont purifiées sur colonne de nickel. Le gène de la mPsaA a été amplifié par PCR (Pfu polymérase (Promega, M7741) : 1) dénaturation initiale 95°C

pendant 3min, les étapes 2 à 4 sont répétées 18 fois : 2) dénaturation 95°C pendant 30sec, 3) hybridation 58°C pendant 30sec, 4) élongation 68°C pendant 1min, puis 5) élongation finale 68°C pendant 2min) à l'aide d'amorces contenant des sites de restriction *NcoI* et *XhoI* (amorces 1 et 2 présentées en annexe). Puis les amplicons sont ensuite vérifiés sur gel d'agarose 1% et purifiés à l'aide du kit de PCR purification (QIAGEN, 28104).

Le vecteur et l'insert sont ensuite digérés par les enzymes de restriction FastDigest *NcoI* (Thermofisher, FD0573) et *XhoI* (Thermofisher, FD0694) pendant 30min à 37°C avant d'être déposés sur gel d'agarose 1% et purifiés à l'aide du kit d'extraction de gel (QIAGEN, 28115).

Enfin, la ligation est effectuée grâce à l'enzyme T4 ligase (Thermofischer, EL0014) pendant 15min à température ambiante sachant que le rapport molaire entre le vecteur et l'insert utilisé est de 1/3, et que la réaction a été effectuée sur 50ng de vecteur. Ensuite, 2,5µL de la réaction de ligation (1/4 du volume total) ont été utilisés pour la transformation dans des bactéries XL1b ou DH5α compétentes (méthode TSS décrite ci-dessous).

II.2 Compétence des cellules

Les cellules XL1b compétentes à la transformation par choc thermique sont préparées grâce à la méthode TSS.

Pour cela, une préculture bactérienne est diluée au 100^{ème} dans du LB, puis incubée à 37°C jusqu'à ce que la densité optique à 600nm soit comprise entre 0,40-0,5 (début de phase exponentielle). La culture est ensuite aliquotée (1mL par tube) et la croissance bactérienne est arrêtée en mettant la culture 10min dans la glace. Les tubes sont centrifugés à 4500g pendant 10min à 4°C, le surnageant est enlevé et le culot resuspendu dans 100µL de TSS (PEG₆₀₀₀ 100mg/mL ; MgSO₄ 12.5mg/mL ; DMSO 5% v/v ; pH6.5). Les cellules sont ensuite congelées à -80°C et disponibles pour un usage ultérieur. La compétence des cellules est vérifiée par transformation de 1µL d'un plasmide test pET11a (50ng) avant validation.

II.3 Transformation et vérification de séquences

Les cellules compétentes sont décongelées dans la glace pendant 10min, puis l'ADN plasmidique est ajouté aux cellules avant incubation pendant 30min dans la glace. Le choc thermique est ensuite réalisé en plaçant les tubes à 42°C pendant 45sec avant de les replacer dans la glace pendant 2min. Ensuite, l'expression de la résistance à l'antibiotique est permise par ajout de 900µL de LB (sans antibiotique) et incubation à 37°C sous agitation pendant 1h avant étalement sur des boîtes de pétri contenant du LB-Agar avec l'antibiotique de sélection (kanamycine pour pET24d).

Pour vérifier la séquence des vecteurs générés, des précultures sont réalisées avec 5mL de milieu LB avec antibiotique de sélection (les antibiotiques correspondant aux plasmides sont présentés en annexe) et incubées à 37°C sur la nuit sous agitation. Une Miniprep est réalisée suivant le kit de QIAGEN (27104). Cinq cents à mille nanogrammes de cette Miniprep sont ensuite envoyés à séquencer par l'entreprise Eurofins à l'aide de la méthode de Sanger (avec les amorces T7P et T7T pour les plasmides pET (n°3 et 4 en annexe)). Les séquences sont ensuite analysées par alignement avec la séquence cible sur le logiciel Geneious.

II.4 Mutation du gène de la mPsaA par amplification PCR

Afin d'incorporer l'acide aminé non naturel à une position donnée dans la protéine, un codon TAG doit être généré dans le gène codant pour la protéine cible. Nous avons choisi trois positions pour incorporer l'acide aminé non naturel : K24, K213 et K229. Après mutation, nous obtenons les

gènes mPsaA-K24TAG, mPsaA-K213TAG et mPsaA-K229TAG. Ces mutations ont été générées avec une technique de mutagenèse par amplification par PCR décrite ci-dessous.

Le plasmide à modifier est amplifié par PCR (Pfu polymérase (Promega, M7741) : 1) dénaturation initiale 95°C pendant 3min, les étapes 2 à 4 sont répétées 16 à 18 fois : 2) dénaturation 95°C pendant 30sec, 3) hybridation 58°C pendant 30sec, 4) élongation 68°C pendant 5min (1min/kb), puis 5) élongation finale 68°C pendant 15min avec les amorces portant les mutations (cf amorces n°5-10 en annexe). Puis le plasmide parental non modifié est éliminé par traitement avec l'enzyme *DpnI* (ThermoFisher, ER1701, 10U par réaction, incubation pendant 1h30 à 37°C, puis inactivation par 20min à 80°C). Enfin, un dixième de la réaction de mutagenèse (soit 5µL) est transformé dans des bactéries DH5α ou XL1b compétentes. Les modifications sont ensuite vérifiées par séquençage comme décrit ci-dessus (section II.3) avec les amorces T7P et T7T (n°3 et 4 en annexe).

Pour le triple mutant mPsaA-K24PrK-K213PrK-K229PrK, l'étape de mutagenèse a été répétée 3 fois de manière séquentielle en suivant le protocole décrit ci-dessus.

V. Optimisation de l'incorporation de l'acide aminé non naturel

IV.1 Optimisation de l'incorporation de la Propargyl-Lysine par la méthode de plan d'expérience

Création du plan d'expérience sous le logiciel Rstudio :

Afin de déterminer les conditions optimales d'incorporation de la PrK dans la protéine modèle mPsaA, une méthodologie de plan d'expérience basée sur un modèle Box-Behnken a été mise en place en utilisant le logiciel RStudio (Version 1.1.453) et plusieurs packages (Annexe 5). Ce modèle correspond à un modèle du second degré de type « $y = a_0 + a_1x_1 + a_2x_2 + a_{12}x_1x_2 + a_{11}x_1^2 + a_{22}x_2^2 + e$ » où y est la réponse (ici la production de protéine ayant incorporé la PrK) en fonction des différents facteurs x (ici pH, température, temps d'induction à l'arabinose, et concentration d'arabinose) associés à leur coefficient a (qui indiqueront leur effet sur la réponse) et e l'erreur expérimentale. Les facteurs étudiés sont entrés avec leurs bornes (valeurs minimale et maximale) selon la définition du domaine d'étude choisi. Le logiciel établit alors une liste de conditions expérimentales à tester sous forme d'un tableau Excel. Une fois les expériences réalisées, les résultats sont importés dans RStudio toujours via Excel. Le logiciel Rstudio calcule ensuite un modèle de second degré qui s'ajuste au mieux aux points expérimentaux. Plusieurs éléments permettent de vérifier la validité du modèle calculé comme le R^2 , ainsi que la possibilité de représenter graphiquement les résidus (différences entre chaque point expérimental et le modèle) afin de s'assurer que les résidus présentent bien un profil aléatoire et donc que le modèle n'est pas biaisé par un oubli de facteur par exemple. Chaque effet de facteur représenté par les coefficients a est également calculé et évalué selon un test de Student, ainsi une p value permet de déterminer quels sont les facteurs significativement influents. Le logiciel calcule également les points stationnaires du modèle, c'est-à-dire des maxima ou minima du modèle.

Clonage GFP en aval du gène de la mPsaA-K193 :

Afin de réaliser des tests d'incorporation par la méthodologie de plan d'expérience, il est nécessaire d'avoir une réponse quantitative. Le gène de l'eGFP a donc été cloné en aval du gène de la mPsaA, puis un codon TAG a été généré au sein du gène mPsaA en position K193. L'incorporation d'un acide aminé non naturel au niveau du codon TAG conduit à la production d'une protéine de fusion fluorescente. La fluorescence peut ainsi être mesurée avec un spectrophotomètre à plaques (TECAN, Infinite M1000) et quantifiée, et reflète l'efficacité d'incorporation de la PrK dans la protéine d'intérêt.

Le plasmide pET24d-mPsaA a été linéarisé par digestion avec l'enzyme *XhoI* (ThermoFisher ER0692). Les fragments ont été déposés sur gel d'agarose 1% et purifiés à l'aide du kit d'extraction de gel (QIAGEN, 28115).

Le gène eGFP a été amplifié par PCR (Pfu (Thermo Fisher F-530S) : 1) dénaturation initiale 95°C pendant 3min, les étapes 2 à 4 sont répétées 25 fois : 2) dénaturation 95°C pendant 30sec, 3) hybridation 58°C pendant 30sec, 4) élongation 72°C pendant 1min, puis 5) élongation finale 72°C pendant 10min). L'amplicon a ensuite été déposé sur gel d'agarose 1% et purifié à l'aide du kit d'extraction de gel (QIAGEN, 28115).

Enfin, le clonage est réalisé à l'aide du kit de Gibson Assembly (DNA Synthetic Genomics, GA1100). Le rapport molaire entre le vecteur et l'insert utilisé est de 1/3 avec 50ng de vecteur. Le mélange contenant l'ADN (insert et vecteur) ainsi que 5µL de mélange Gibson dans un volume final de 10µL a été incubé à 50°C pendant 30min. Un microlitre de ce mélange a ensuite été transformé au sein de bactéries XL1b compétentes (méthode TSS). Le codon stop TAG a ensuite été généré selon la méthode décrite dans la section II.4, à l'aide des amorces n°25-28 décrites en annexe. Les étapes de transformation et de vérification de séquence ont été réalisées comme décrit précédemment (section II.3).

Réalisation des tests expérimentaux :

Pour s'assurer d'une répétabilité optimale sur les tests, ceux-ci ont été effectués par un seul manipulateur (stagiaire M1 : Mehdi Largouet). Une préculture a été réalisée à heure fixe (16h) dans 5mL de LB en présence d'antibiotiques (kanamycine à 25µg/mL et chloramphenicol à 30µg/mL) à partir d'une colonie. Puis le lendemain à 10h, une culture a été réalisée par dilution de la préculture au centième dans du milieu autoinductif (FORMEDIUM, n° AIMLB0210) en présence d'antibiotiques (kanamycine à 25µg/mL et chloramphenicol à 30µg/mL), puis l'arabinose a été ajouté selon les paramètres du test (cf tableau des conditions des tests). Le lendemain, après 24h de culture (à 10h), 200µL de culture sont utilisés pour une mesure de densité optique à 600nm et une mesure de fluorescence (excitation : 488nm, émission : 509nm) par un appareil de mesure (TECAN, Infinite M1000). Les données générées sont ensuite analysées à l'aide du logiciel Rstudio.

IV.2 Optimisation de l'incorporation de la Propargyl-Lysine par méthode d'ajout d'un transporteur

Clonage d'un transporteur membranaire LysP par la méthode de Gibson :

L'efficacité d'incorporation de la Propargyl-Lysine est dépendante de sa concentration utilisée. Nous avons supposé qu'en augmentant sa concentration intracellulaire nous pourrions augmenter la quantité de protéine ayant incorporé la Propargyl-lysine. La structure de la Propargyl-Lysine étant proche de la Lysine, nous avons supposé que surexprimer le gène du transporteur de la lysine (LysP) pourrait augmenter le rendement de protéine produite. Pour cela, nous l'avons cloné au sein du plasmide pEVOL. Une stratégie de clonage par méthode de Gibson Assembly a été testée. Le plasmide pEVOL-1-copie (issu de la modification du plasmide pEVOL par la suppression d'une copie de la synthétase par PCR, voir protocole Chapitre 7, section I.3.3.1) a été linéarisé par digestion avec l'enzyme *PstI* (ThermoFisher ER0611). Ensuite, le gène LysP a été amplifié par PCR (Phusion (Thermo Fisher F-530S) : 1) dénaturation initiale 98°C pendant 10min, les étapes 2 à 4 sont répétées 25 fois : 2) dénaturation 98°C pendant 30sec, 3) hybridation 58°C pendant 30sec, 4) élongation 72°C pendant 1min, puis 5) élongation finale 72°C pendant 2min) sur colonies bactériennes BL21(DE3) en utilisant les amorces n°15 et 16 décrites en annexe qui apportent une complémentarité de séquence avec le plasmide pEVOL-1copie. Enfin, le clonage est réalisé à l'aide du kit de Gibson Assembly (DNA Synthetic Genomics, GA1100). Le rapport molaire entre le vecteur et l'insert utilisé est de 1/3, et la

réaction est incubée à 50°C pendant 30min, avant d'être transformée au sein d'XL1b compétentes (méthode TSS).

LysP clonage par enzyme de restriction :

Le clonage par la méthode de Gibson n'ayant pas fonctionné, une stratégie de clonage par enzyme de restriction a été réalisée suivant le protocole ci-dessous.

Le site *PstI* étant présent dans la séquence de LysP, il a donc d'abord été éliminé par une première étape de PCR. Deux fragments de PCR, correspondant à la partie en 5' du site *PstI* et la partie en 3' du site *PstI* et se recouvrant au niveau du site *PstI* modifié, ont été générés sans le site *PstI* (PCR Phusion (Thermo Fisher F-530S) : 1) dénaturation initiale 98°C pendant 3min, les étapes 2 à 4 ont été répétées 25 fois : 2) dénaturation 98°C pendant 30secondes, 3) hybridation 58°C pendant 30sec, 4) élongation 72°C pendant 1min, puis 5) élongation finale 72°C pendant 2min, en utilisant les amorces n° 17 et 18 en annexe). Puis les amplicons avec recouvrement ont ensuite été déposés sur gel d'agarose 1% et purifiés à l'aide du kit d'extraction de gel (QIAGEN, 28115). Les deux fragments ainsi générés ont ensuite servi de matrice ADN pour une seconde PCR afin de fusionner les deux parties du gène de LysP sans le site *PstI* et générer l'insert contenant des sites de restriction à ses extrémités 5' et 3' ((PCR (Phusion (Thermo Fisher F-530S) : 1) dénaturation initiale 98°C pendant 3min, les étapes 2 à 4 ont été répétées 5 fois : 2) dénaturation 98°C pendant 30sec, 3) hybridation 58°C pendant 30sec, 4) élongation 72°C pendant 1min, 5) ajout des amorces au bout de 5 cycles, puis les étapes 6 à 8 ont été répétées 25 fois : 6) dénaturation 98°C pendant 30sec, 7) hybridation 58°C pendant 30sec, 8) élongation 72°C pendant 1min, 9) élongation finale 72°C pendant 10min, en utilisant les amorces n°15 et 16). L'amplicon généré correspond au gène LysP sans le site de restriction *PstI*. Il est ensuite cloné au sein du plasmide pEVOL (à la place de la deuxième séquence de l'aaRS) d'après le protocole suivant.

L'amplicon de LysP, ainsi que le vecteur pEVOL sont digérés par les enzymes de restrictions de Thermofisher (*PstI* (ThermoFisher ER0611) et *NdeI* (ThermoFisher, ER0581) pendant 1h à 37°C avant d'être purifiés sur kit (QIAGEN, 28115).

Ensuite, la ligation est effectuée grâce à la T4 ligase (Thermofischer, EL0014) pendant 15min à température ambiante, le rapport molaire entre le vecteur et l'insert utilisé est de 1/3. Enfin 2,5µL de la réaction de ligation ont été utilisés pour la transformation dans des bactéries XL1b ou DH5α compétentes (méthode TSS). Les étapes de transformation et de vérification de séquence ont été réalisées comme décrit précédemment (section II.3).

LysP réalisation des tests expérimentaux :

Afin de vérifier si la surexpression de LysP a permis d'augmenter la quantité de Propargyl-Lysine au sein de la mPsaA, des tests d'incorporation de la Propargyl-Lysine ont été réalisés sur les cultures BL21(DE3) co-transformés avec pET24d-mPsaA(K193TAG)-EGFP et pEVOL-1copie-LysP en suivant le protocole décrit précédemment (section III.3).

IV.3 Optimisation de l'incorporation de la Propargyl-Lysine par mutagenèse ciblée de l'aaRS

Les mutations R19H, H29R et T122S ont été décrites précédemment comme améliorant l'affinité de l'aaRS pour l'ARNt et donc comme pouvant améliorer l'incorporation d'acides aminés non naturels (Vangmayee Sharma et al. 2018). Ces mutations ont donc été réalisées par la méthode de mutation par amplification par PCR comme décrit dans la partie I.4 avec les amorces n°19-24

décrites en annexe. Les tests d'incorporation de la Propargyl-Lysine ont été réalisés comme décrit précédemment partie II.3.

VI. Analyse de la réponse immunitaire

V.1 Immunisation des souris pour évaluer la réponse immunitaire induite par les vaccins produits

Afin d'étudier la réponse immunitaire induite par nos vaccins glycoconjugués, plusieurs types de glycoconjugués ont été testés :

-les conjugués entre la mPsaA et le tétrasaccharide (Pn14TS) à une position ciblée à l'aide de la Propargyl-Lysine : mPsaA-K24-Pn14TS, mPsaA-K213-Pn14TS et mPsaA-K229-Pn14TS.

-le conjugué entre la mPsaA et le tétrasaccharide (Pn14TS) à trois positions ciblées à l'aide de la Propargyl-Lysine : mPsaA-K24-Pn14TS-K213-Pn14TS-K229-Pn14TS.

-un conjugué produit à l'aide d'une méthode de conjugaison aléatoire (section III.11) avec trois tétrasaccharides (Pn14TS) liés à la mPsaA de manière non ciblée : mPsaA-3Pn14TS.

Les vaccins (5µg protéine/dose) ont été injectés en sous cutané à des souris femelles BALB/c au jour 0 et une autre injection a été réalisée afin d'augmenter la réponse (boost) aux jours 14 et 28. Deux jours après l'immunisation finale, les sérums des souris sont prélevés et la réponse immunitaire est analysée par ELISA (section suivante).

V.2 Analyse de la réponse immunitaire par ELISA

Les sérums sont analysés par la technique ELISA pour comparer la présence d'anticorps anti-tétrasaccharide et anti-mPsaA. Pour cela, des plaques 96 puits (VWR, 60941-062) sont coatées avec l'antigène étudié (le polysaccharide capsulaire du sérotype 14 à une concentration comprise entre 0,5 à 10mg/mL, la mPsaA ou de la BSA (à 10mg/mL)) sur la nuit à 4°C. Les plaques sont ensuite lavées 3 fois avec 200µL de tampon PBS-Tween 0,5% à l'aide d'un laveur de plaque (Thermo Scientific, WellWash 5165000). Afin d'éviter des interactions non spécifiques, la plaque est ensuite incubée dans du PBS-Lait-10% pendant 2h à 37°C, puis les plaques sont de nouveau lavées 3 fois avec 200µL PBS-Tween 0,5%. Cent microlitres de différentes dilutions en série du sérum à tester sont ensuite ajoutés dans la plaque et incubés 1h à 37°C, puis les plaques sont lavées avec 200µL PBS-Tween 0,5%. Afin de détecter la présence d'IgG ou d'IgM, 100µL d'anticorps anti-IgG (CliniSciences, 1031-05) ou anti-IgM (CliniSciences, 1020-05) dilués au 6000^{ème} sont ajoutés. Les plaques sont ensuite incubées pendant 1h à 37°C avant lavage (3 fois avec 200µL PBS-Tween 0,5%). Les anticorps anti-IgG ou anti-IgM étant couplés à la HRP (horseradish peroxidase), 100µL d'OPD (o-phénylènediamine dihydrochloride) sont ajoutés pour permettre l'apparition d'une coloration liée à la concentration en IgG dosés et les plaques sont incubées 20min dans l'obscurité à température ambiante. La réaction est ensuite stoppée par ajout de 100µL d'HCL 3M, puis la densité optique est lue à 492 nm à l'aide d'un spectrophotomètre (TECAN, Infinite M1000).

***Chapitre 5 : Interactions Lectines -
Saccharides***

I. Introduction générale

I.1 Lectines

I.1.1 Définition et Classification des lectines

Les lectines sont généralement définies comme des protéines non enzymatiques d'origine non immunogénique qui reconnaissent les sucres de manière non covalente. Elles peuvent reconnaître leur ligand saccharidique sous différentes formes : mono-, oligo- ou polysaccharide, greffé ou non à des lipides ou des protéines (Varki et al. 2009). Elles ont été précédemment appelées agglutinines, hémagglutinines, phytohémagglutinines de par leur capacité à agglutiner des cellules. En effet, la découverte de la première lectine a été attribuée à Peter Hermann Stillmark en 1888 lors de l'analyse d'extraits de graines de ricin (*Ricinus communis*) qui agglutinaient des érythrocytes (Varki et al. 2009). Le nom de lectine (qui vient de *legere* en latin qui signifie choisir, sélectionner) a finalement été utilisé lorsque leur capacité à discriminer les groupes sanguins a été découverte.

Les lectines sont impliquées dans des fonctions biologiques très variées comme la défense des plantes contre les champignons, la régulation du système immunitaire chez les animaux ou encore la reconnaissance pathogène-hôte. Elles sont d'ailleurs présentes de manière ubiquitaire au sein du vivant : les virus, les bactéries, les plantes ou encore les animaux. Le Tableau 21 ci-dessous montre quelques exemples de lectines issues de différents types d'organismes avec leur rôle biologique, ainsi que leur type de reconnaissance.

Tableau 21: Exemples de lectines selon leurs origines, leur principale fonction et leur type de reconnaissance

Lectine	Origine	Exemple fonction	Reconnaissance type	Référence
Concanavaleine A	Plantes légumineuses	Induit la prolifération de cellule T (immunité)	α Man et α Glc	(Palacios 1982)
Calnexine	Vertébrés	Chaperone, repliement protéique	Glc1Man9	(Williams 2006)
Galectines	Animaux	Régulation prolifération cellulaire, homéostasie système immunitaire	β -Gal	(Barondes et al. 1994)
FimH	Bactérie	Adhérence bactérienne	Man	(Hung et al. 2002)
Hémagglutinine du virus de la grippe	Virus	Attachement cellulaire aux cellules hôtes	Acide sialique	(Weis 1997)
Lectine type C, (ex : MBL, DCL)	Animaux	Reconnaissance pathogènes par les macrophages	Man	(Kerrigan and Brown 2009)
<i>Aleuria aurantia</i> lectine (AAL)	Champignon	Antifongique	Fucose	(Amano et al. 2012)

Les lectines ont ainsi été historiquement classées selon leurs origines mais cette classification a ensuite laissé place à une classification plus fonctionnelle selon leurs types de substrats (H. Surangi, N. Jayawardena, Xin Wang 2013), leurs structures primaires (Damme 2014), leurs localisations cellulaires (Lakhtin, Lakhtin, and Alyoshkin 2011) ou encore leurs caractéristiques structurales

(Fujimoto, Tateno, and Hirabayashi 2014; Lis and Sharon 1998). En effet, les lectines sont des protéines très hétérogènes (taille, structure, fonction, localisation au sein de la cellule) mais on retrouve des lectines de différentes origines avec des similarités au niveau de leur squelette et de leur type de reconnaissance (Lis and Sharon 1998), ou de leurs structures tridimensionnelles (Fujimoto, Tateno, and Hirabayashi 2014). Les lectines ont ainsi été classées en de nombreuses familles structurales (Familles appelées Pfam définies à partir d'alignements de séquences) avec pour plus de 50 d'entre elles une structure 3D connue ainsi que leur spécificité de substrats définie (Fujimoto, Tateno, and Hirabayashi 2014). Globalement on peut remarquer une abondante présence de structures de type β dans les lectines avec principalement des β -sandwiches mais aussi des β -prismes, trèfles- β et des β -propeller (Imberty and Prestegard 2017). En effet, ces structures permettent la mise en place d'un réseau de liaisons hydrogènes pouvant interagir avec les hydroxyles des sucres (Fujimoto, Tateno, and Hirabayashi 2014; Hirabayashi and Arai 2019). Cependant, il y a également de nombreuses lectines pauvres en structures β telles que les lectines de type C, les lectines Ym1 ou les lectines de type M (Hirabayashi and Arai 2019). Dans la Figure 81, nous pouvons en effet comparer des structures 3D d'une lectine riche en structure β (ici β -sandwich, Galectin-1, PDB : 3OYW) et une lectine pauvre en structure β (C-type, PDB : 5F2Q).

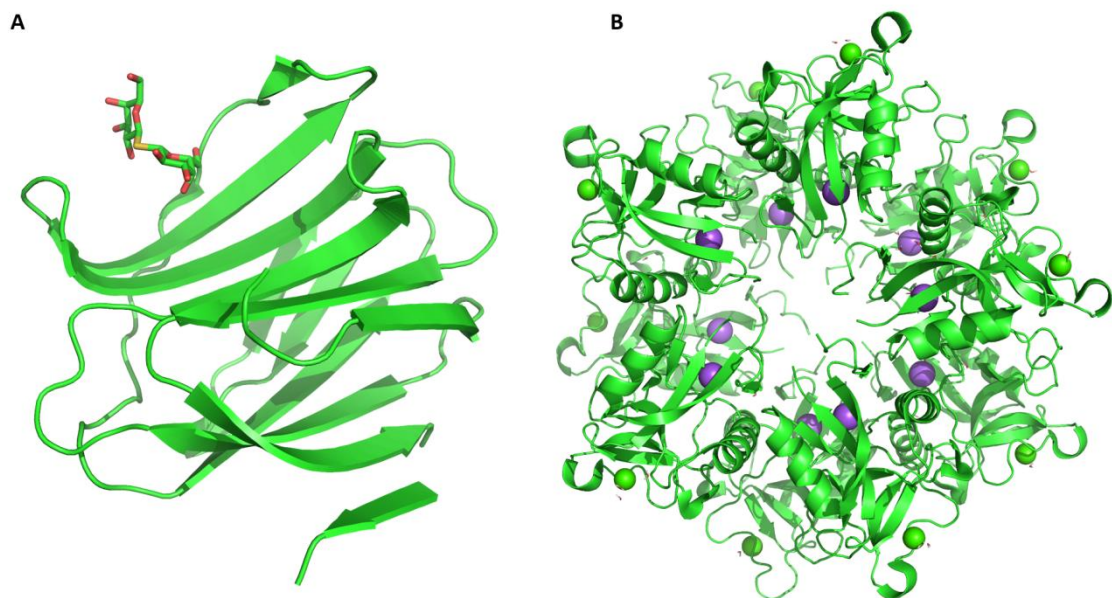


Figure 81: Exemples comparatif de structure 3D de lectines riche en structure β (A) et riche en structure α (B)
 A) Structure 3D de la chaîne A de la Galectin-1 humaine cristallisée avec du thiodigalactoside obtenue par diffraction de rayons X (code PDB : 3OYW); B) Structure 3D de la chaîne A d'une lectine de type C issue de *Bothrops jararacussu* obtenue par diffraction de rayons X (code PDB : 5F2Q)

1.1.2 Valence des lectines

Les lectines ont souvent une très bonne spécificité, elles sont en effet capables de discriminer des épimères de sucre (glucose par rapport au galactose par exemple) et parfois même leurs caractéristiques anomériques (α ou β) (voir Section II.1) (Lis and Sharon 1998). Cependant, l'affinité des lectines est généralement assez faible (mM ou μ M). De nombreuses lectines présentent des structures multivalentes. Cela peut se traduire par des protéines multimériques, une répétition en tandem de domaines lectiniques au sein d'un polypeptide ou de la multi-présentation de lectines à une surface membranaire (Figure 82, Figure 83) (Y. C. Lee and Lee 1995). La liaison à plusieurs substrats permet alors d'augmenter l'affinité apparente par un effet d'avidité en multipliant les sites de reconnaissances (R. T. Lee and Lee 2001).

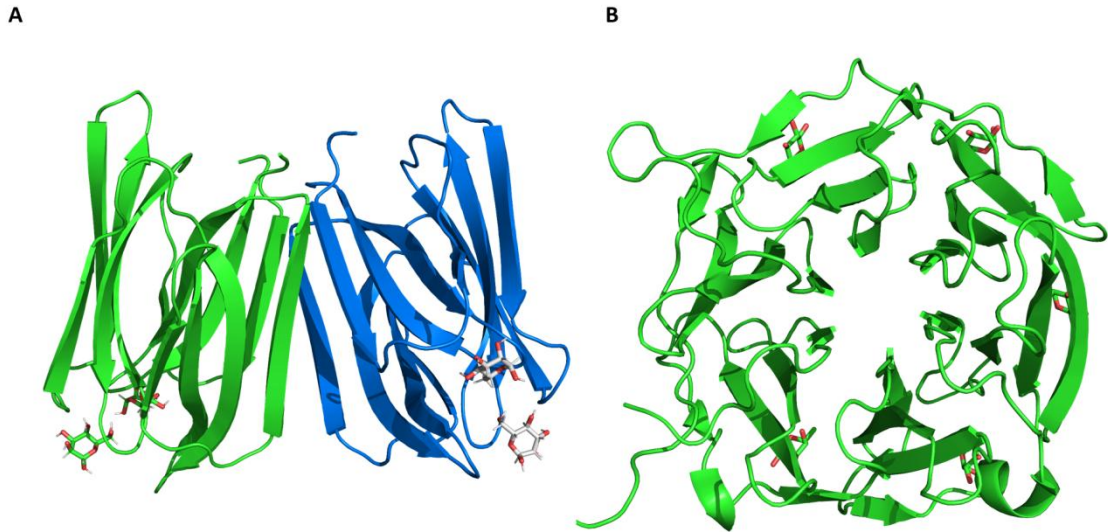


Figure 82: La multivalence des lectines peut se traduire par une multimérisation (exemple A) ou par une répétition en tandem de séquences peptidiques (exemple B).

A) Structure 3D de AcMJRLune lectine issue de *Ananas comosus* obtenue par diffraction des rayons X (Code PDB : 6FLY), B) Structure 3D de la lectine de *Aleuria aurelia* (AAL) par diffraction de rayons X (Code PDB : 1OFZ)

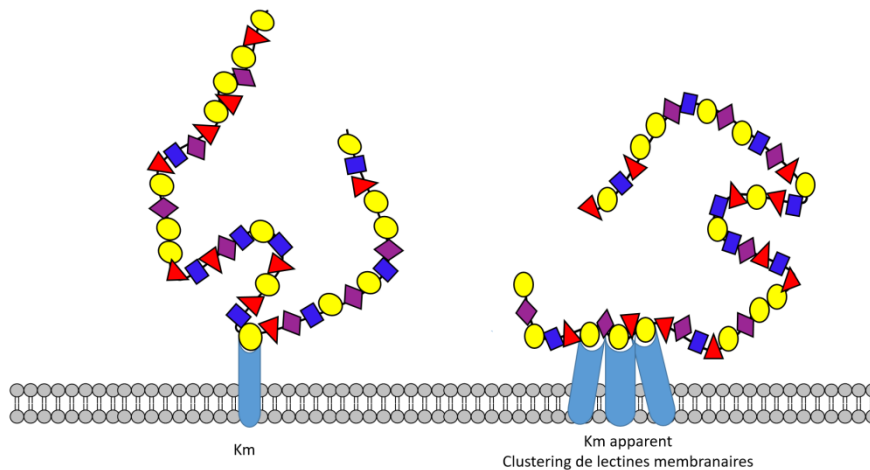


Figure 83: Schéma représentant la clusterisation de lectines à la surface d'une membrane

La multivalence des lectines peut se traduire par une clusterisation des lectines à la surface d'une membrane à la suite de la liaison d'un polysaccharide ou de plusieurs saccharides.

I.2 Particularité des Carbohyrate Binding Modules

Au sein d'enzymes modifiant des saccharides, on peut retrouver des régions caractérisées en tant que domaines lectiniques, également appelées Carbohyrate Binding Modules. Elles ont été originellement appelées Cellulose Binding Domaines (CBD) car les premiers domaines caractérisés étaient spécifiques de la cellulose (Gilkes et al. 1988; Ong et al. 1989; Tomme et al. 1988). Cependant, la diversité de leurs saccharides substrats est mieux représentée sous l'appellation Carbohyrate Binding Module (CBM). Ces CBM sont principalement retrouvés dans des enzymes de dégradation (glycosidases, polysaccharides lyases) ou de synthèse (glycosynthases, transglycosidases, glycosyltransférases) de saccharides. Ce sont des domaines protéiques le plus souvent bien définis qui peuvent donc être exprimés et se replier correctement, indépendamment du domaine catalytique en N- ou C-terminal mais ils peuvent aussi être présents sous la forme de sous-sites appelés "CBM-like" (Shoseyov, Shani, and Levy 2006). Une enzyme peut également contenir plusieurs sous-domaines CBM ce qui augmente l'affinité grâce à un effet de valence (section I.1.2) mais aussi des CBM issus de familles différentes (Guillén, Sánchez, and Rodríguez-Sanoja 2010).

Les différents types de CBM ont été classés en 85 familles au sein de la base de donnée CAZY (Database, n.d.). Cette classification découle de la classification des glycosidases (GH) et est donc basée sur la similarité de séquence. En effet, lorsque la fonction d'un CBM n'est pas connue, sa similarité de séquence avec un autre CBM peut permettre de prédire le type de saccharide ou polysaccharide auquel il pourrait se lier.

Les CBM ont également été classés fonctionnellement en 3 types (A, B, C) selon la façon dont ils se lient au polysaccharide. En effet, ces modes de liaisons ont des répercussions sur leurs caractéristiques structurales (Boraston et al. 2004). Le type A comprend les CBM qui se lient à des surfaces de polysaccharides compactes, voire même cristallines. La présence d'acides aminés hydrophobes permet la création d'une surface plane et l'interaction avec les sucres ligands se fera principalement par des interactions de type CH- π avec le cycle osidique. Le type B est constitué des CBM qui lient des chaînes polysaccharidiques. Ils présentent des structures en tunnel ou cavité avec un site principal et des sous-sites de fixation aux saccharides. Enfin, les CBM qui se lient plutôt à des sucres simples ou à des extrémités de polysaccharides sont du type C. Ceux-ci possèdent des propriétés proches des lectines. Nous pouvons voir dans la Figure 84 ci-dessous, un exemple de structure issu de chaque type.

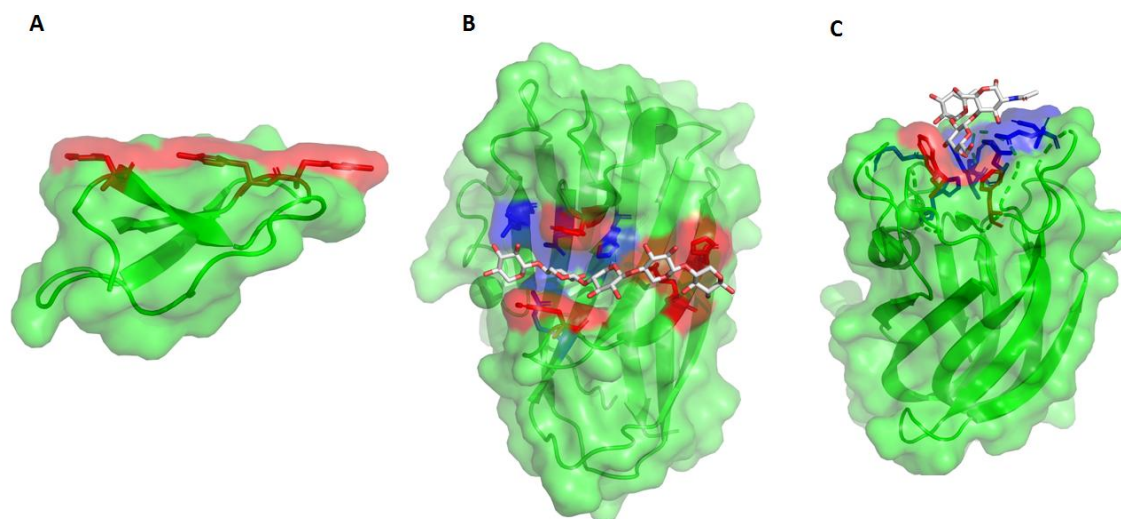


Figure 84: Structures 3D issues des trois types de Carbohydrate Binding Module (CBM)

Les acides aminés intervenant dans des interactions hydrophobes et π -CH sont colorés en rouges, et ceux intervenant dans des interactions polaires sont représentés en bleu. A) CBM de type A: structure RMN du CBM1 de *T. reesei* (code PDB 1CBH). B) CBM de type B: Structure 3D (Diffraction Rayons X) de CBM4 issue d'une Xylanase de *Rhodothermus marinus* (Code PDB 2Y64). C) CBM de type C: Structure 3D (Diffraction Rayons X) CBM32 issue de *Clostridium perfringens* (Code PDB 2J7M)

Trois fonctions principales ont été attribuées aux CBMs : i) effet de proximité ii) ciblage iii) déstabilisation de structure (Shoseyov, Shani, and Levy 2006). En effet, en se liant au polysaccharide substrat de l'enzyme, les CBMs augmentent artificiellement la concentration de l'enzyme à proximité de celui-ci et donc augmentent la vitesse de sa biotransformation. D'ailleurs, les domaines catalytiques de glycoenzymes séparés de leur CBM ont en général une moins bonne activité (Ali et al. 2001; Boraston et al. 2003; Bolam et al. 1998). La présence de CBM peut également permettre de cibler une particularité du polysaccharide, par exemple des CBM9 issus de la xylanase 10A de *Thermotoga maritima*, qui dégradent les parois de plantes, ont une spécificité pour les extrémités réductrices de la cellulose et de polysaccharides solubles, ce qui peut traduire une reconnaissance plus spécialisée envers des portions déjà endommagées de la paroi cellulaire (Boraston et al. 2001). Enfin, la déstabilisation d'interfaces entre des polysaccharides permettant la libération d'extrémités réductrices alors disponibles à la dégradation a également été répertoriée sur des substrats comme l'amidon (Guillén, Sánchez, and Rodríguez-Sanoja 2010). D'autres fonctions ont pu être relevées

comme la stabilisation de l'enzyme (Fontes et al. 1995; Winterhalter et al. 1995), ou la fermeture du site catalytique (Liberato et al. 2016; Venditto et al. 2015).

Les différentes structures tridimensionnelles que l'on peut retrouver au niveau des CBMs sont relativement proches des lectines, avec beaucoup de structures de type β dont la forme β -sandwich est la plus fréquente. Ces structures β issues des CBMs ont été classées en 7 familles (*β -sandwich*, *β -trefoil*, *Cystein knot*, *Unique*, *OB fold*, *Hevein*, *Unique (hevein-like)*) (Guillén, Sánchez, and Rodríguez-Sanoja 2010).

I.3 Génération de lectines artificielles

I.3.1 Les lectines organiques

Les boronolectines :

Les fonctions boronates sont analogues des fonctions carboxylates mais ne sont pas retrouvées dans la nature. Il s'agit d'un atome de bore lié à un carbone et deux fonctions hydroxyles (OH). Les propriétés chimiques de cette fonction et leurs diverses applications sont très bien détaillées dans les revues réalisées par Hall *et al.* (Hall 2019, 2006), Pappin *et al.* (Pappin, J., and A. 2012) et Bromba *et al.* (Bromba et al. 2009). Pour résumer, au sein de cette fonction l'atome de bore possède alors une orbitale p vacante et possède donc les propriétés d'un acide de Lewis faible avec un pKa variant de 5 à 9. Elle peut donc former des liaisons covalentes avec des fonctions donneuses d'électrons (base de Lewis). Les boronates peuvent également former des esters de boronates (analogues des esters carboxyliques) en présence de fonctions hydroxyles vicinales, dont notamment des saccharides. Cette réactivité correspond à un échange de fonctions hydroxyles et est favorisée lorsque le boronate est principalement sous forme tétraédrique, en général à pH basique (Figure 85).

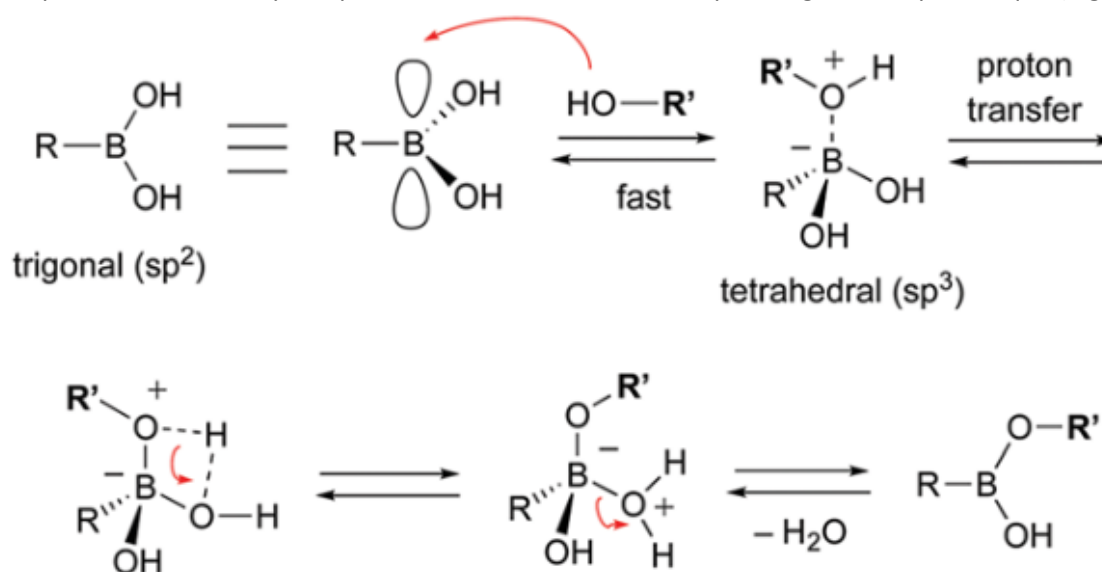


Figure 85: Réactivité d'un boronate permettant l'échange d'une fonction diol avec une forme tétraédrique (sp^3) transitoire. Image extraite de la publication de Hall *et al.*, 2019 (Hall 2019).

La capacité du boronate à réagir avec des saccharides fut remarquée la première fois en 1832 par Biot *et al.* lors d'une analyse du pouvoir optique rotatoire de l'acide tartrique en présence d'acide borique (Pappin, Kiefel, and Houston 2012). Par la suite, l'utilisation de boronate a permis le développement de méthodes pour lier ou fonctionnaliser des sucres de manière régiosélective (Hall 2019). Au niveau des applications de biologie médicale ou biotechnologiques, cela a conduit au développement de récepteurs organiques (appelés aussi boronolectines ou lectines artificielles)

spécifiques pour des sucres. Cela a permis l'amélioration de la délivrance de toxines au sein de cellules mammifères (Ellis, Palte, and Raines 2012), ou la séparation et quantification de protéines glycosylées (Middle et al. 1983; Mallia et al. 1981; Ikeda et al. 1998), mais a également fourni des senseurs fluorescents pour détecter le taux de glucose et ainsi de créer un outil de monitoring chez les diabétiques (Eversense®) (Fang et al. 2018; Worsley et al. 2008; Heo and Kim 2019). Cependant, les défis auxquels doivent faire face les développeurs de lectines organiques d'affinité basée sur la chimie du boronate, sont le manque de sélectivité ainsi que l'affinité faible et dépendante de pH non physiologiques (plutôt très basiques). Pour répondre à ces problématiques, l'utilisation de modifications chimiques, de propriétés de multivalence et le développement d'impression moléculaire sont développés (Z. Liu and He 2017). La technique d'impression moléculaire consiste à générer un polymère fonctionnalisé (à l'aide de liaisons covalentes ou non avec la molécule cible) autour d'une molécule cible ou de structure proche de la molécule cible grâce à une réaction de polymérisation chimique ou thermique d'un mélange entre un agent de réticulation et un initiateur radicalaire dans un solvant. Un fois polymérisée, la molécule cible appelée aussi molécule empreinte est éliminée et le polymère possède alors une mémoire qui devrait permettre la capture spécifique de la molécule cible (Figure 86) (Wulff et al. 1998). La fonctionnalisation de ces polymères avec des acides vinyl-phenyl-boronique (VBA) a notamment permis de créer des lectines organiques contre le D-glucose, le D-fructose, le D-mannose et d'autres dérivés (Parmpi and Kofinas 2004; Tan, Wang, and Yan 2009).

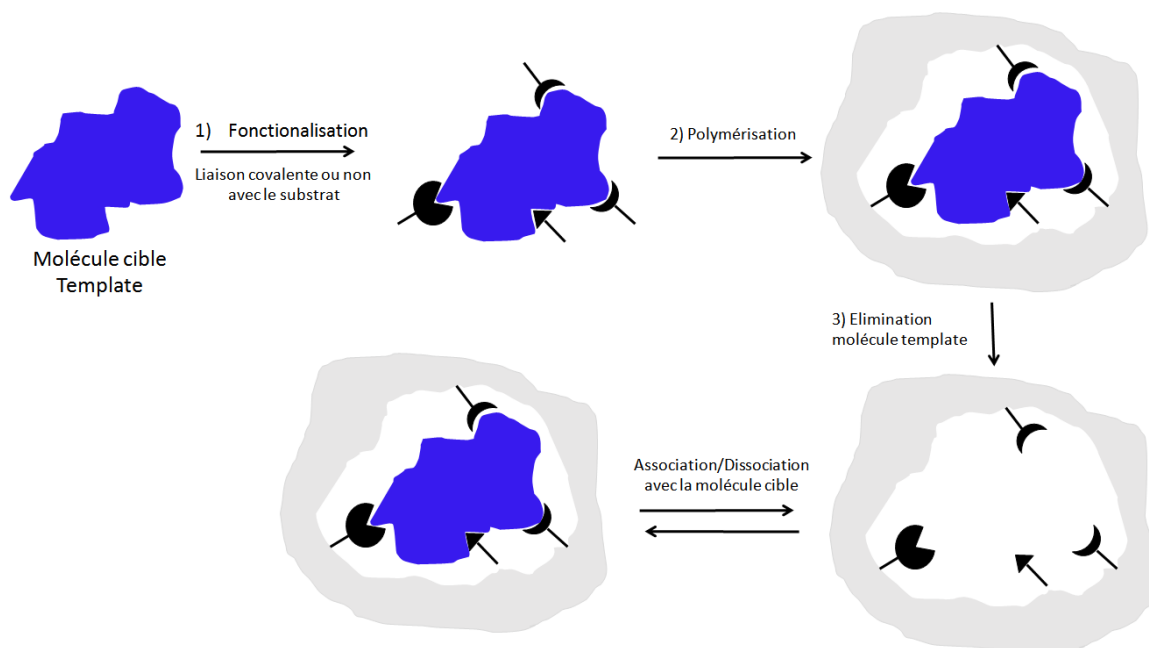


Figure 86: Principe de l'impression moléculaire
Image reproduite à partir de la publication de Lee et al. (S.-H. Lee and Doong 2016)

Autres lectines organiques:

La majorité des lectines organiques utilisent une fonction boronate pour la reconnaissance de saccharides. Cependant, d'autres types de lectines organiques ont pu être développées. Celles-ci permettent l'établissement d'interactions de type CH- π avec le saccharide - cycle osidique (section II.2.2) et liaisons hydrogènes (section II.2.1)- par l'utilisation de cycles aromatiques ou de type aminopyridine. La problématique principale pour créer ce type de lectine est de produire un composé capable de lier les sucres dans un milieu aqueux (Arnaud, Audfray, and Imberty 2013). Cependant, de nombreux travaux ont été réalisés notamment au sein du groupe de Anthony P. Davis, comme par exemple par la création de lectines organiques contenant des surfaces aromatiques

apolaires, espacées par des espaces polaires afin d'interagir avec des sucres, comme représenté dans la Figure 87 ci-dessous (Ferrand, Crump, and Davis 2007; Mooibroek et al. 2016; Carter et al. 2016).

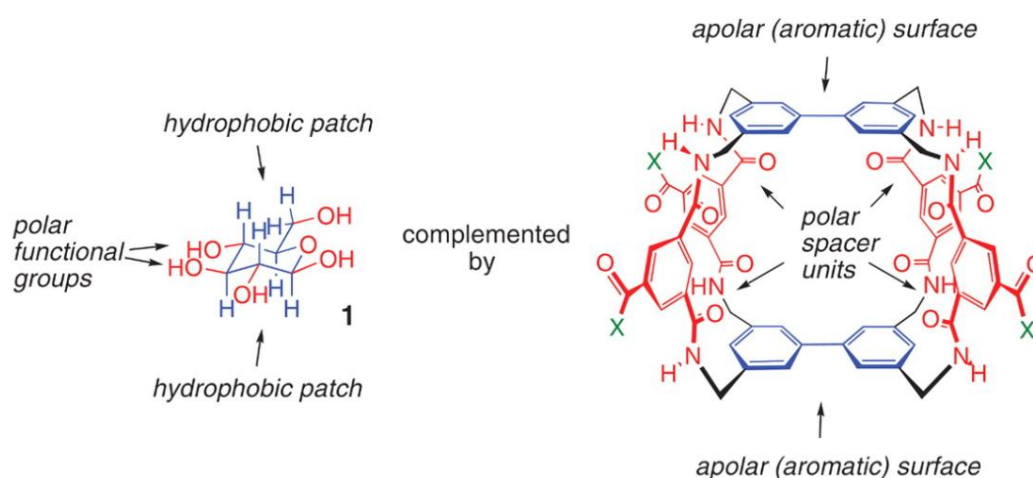


Figure 87: Exemple de lectine organique (Ferrand, Crump, and Davis 2007)

1.3.2 Les aptamères

Les aptamères sont des oligonucléotides (20 à 80 nucléotides) de type ARN ou ADN qui peuvent se fixer à des ligands. Leurs capacités de liaison sont liées aux structures secondaires ou tertiaires formées grâce à l'appariement de séquences, de boucles terminales, *mis-appariement* ou hernies. La découverte des aptamères date des années 1990 et la mise au point concomitante de la méthode SELEX (*Systematic Evolution of Ligands by EXponential enrichment*) ont permis leur développement. C'est en effet par cette méthode que ceux-ci sont généralement sélectionnés à partir de banques de nucléotides synthétiques pouvant contenir jusqu'à 10^{15} variants. La sélection des aptamères est donc basée sur des cycles de sélection au cours desquels la librairie des nucléotides est incubée avec le ligand cible, des lavages sont ensuite réalisés pour éliminer les variants non liants, puis les aptamères qui se sont liés au ligand sont donc récupérés (par élution) et amplifiés par PCR avant d'être réintroduits dans un nouveau cycle de sélection (Ku et al. 2015).

Des aptamères ont été développés avec succès contre des polysaccharides tels que la cellobiose, l'acide sialique, l'antigène de Lewis sLe^x, la cellulose, et du lipopolysaccharide (LPS) (Sun, Du, and Li 2010; Davydova et al. 2016). Cependant, il est difficile de générer des aptamères reconnaissant des motifs saccharidiques sur des glycoprotéines. En effet, la méthode de sélection ne permet pas de cibler une partie d'un ligand complexe et le bruit de fond (sélection d'aptamères qui ne reconnaissent pas la partie ciblée) est considérablement plus fort si le ligand n'est pas bien défini. Comme nous l'avons vu ci-dessus (section 1.3.1), la fonction boronate permet la reconnaissance de sucres. Pour cibler le fibrinogène, l'ajout d'une fonction boronate à un aptamère a permis de sélectionner un aptamère ayant une affinité de 6 nM (M. Li et al. 2008).

Cibler des petites molécules est également plus complexe (Ruscito and DeRosa 2016) et les sucres simples se différencient principalement par la position spatiale des hydroxyles et sont donc des épitopes très similaires. Cela rend difficile la sélection d'aptamères sélectifs. Cependant, des aptamères ont été sélectionnés sur des sucres simples (glucose, fructose, galactose) grâce à une première sélection sur les sucres ligands couplés *in situ* à une boronolectine puis une contre sélection sur la boronolectine seule. Le couplage avec la boronolectine permettrait de mieux présenter les différents épitopes du sucre simple (K. A. Yang et al. 2014).

Le développement des aptamères est cependant limité par certains défis dont leur dégradation rapide (surtout ARN), leur élimination sanguine par filtration rénale, le contrôle sur leur durée d'action, leur délivrance au sein de cellules cibles, leur sélection sur des cibles complexes, leur réactivité croisée ainsi que la méthode de génération des aptamères (Lakhin, Tarantul, and Gening 2013). Différentes pistes sont employées telles que l'utilisation de nucléotides modifiés, leur PEGylation, leur conjugaison à des polymères cationiques, l'utilisation de systèmes de délivrance, ou encore l'optimisation des méthodes de sélections (Lakhin, Tarantul, and Gening 2013).

1.3.3 Inactivation de glycoenzymes

Les enzymes telles que les glycosidases peuvent être modifiées en supprimant leurs activités catalytiques par mutagenèse dirigée au niveau de résidus clés du site catalytique. Ces lectines issues d'enzymes ont été répertoriées sous le nom de Lectenz® (Arnaud, Audfray, and Imberty 2013).

II. Reconnaissance moléculaire protéine - sucre

II.1 La diversité des saccharides ligands

La structure des saccharides a été décrite dans le chapitre 2. Chaque monosaccharide est défini par sa longueur de chaîne, sa fonctionnalisation et la stéréochimie de ses carbones asymétriques. Puis les polysaccharides sont formés par leur enchaînement (en polymère linéaire ou ramifié) selon différents types de liaison (α ou β). Une grande diversité de polysaccharides peut ainsi être générée car en plus de la variété des monosaccharides existants, ceux-ci peuvent former des structures ramifiées très complexes. En effet, le carbone anomérique d'un sucre peut réagir, en α ou en β , avec 5 hydroxyles différents d'un autre sucre, dans le cas d'un hexose.

Au niveau des polysaccharides, un nouveau niveau de complexité apparaît alors avec la rotation potentielle des différents angles de la liaison osidique qui génère de nombreux conformères (Figure 88)(Case et al. 2005) et leurs apporte une flexibilité. Cependant, en pratique uniquement les conformations d'énergie faible sont présentes, et la structure de l'oligosaccharide est parfois rigidifiée par des interactions entre les différents monosaccharides comme des liaisons hydrogènes ou des interactions de type Van der Waals (Taylor and Kurt 2011).

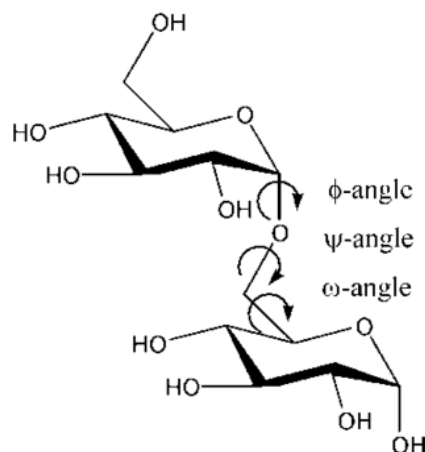


Figure 88: Présentation des angles (ϕ , ψ , et ω) au niveau de la liaison osidique (ici au niveau du α -D-glcp-(1 \rightarrow 6)- α -D-glcp (iso-maltose)).

Image extraite de la publication de Case et al (Case et al. 2005).

Ainsi, les sucres sont très divers de par les différentes stéréochimies des carbones asymétriques, des liaisons osidiques, les différentes ramifications possibles, les différentes conformations adoptées, ainsi que leurs différentes modifications telles que des acétylations, sulfatations ou méthylation. La diversité potentielle générée des polysaccharides est donc bien plus importante que pour les autres biopolymères (composés de nucléotides ou d'acides aminés). Une étude réalisée par Werz *et al.* essaie de donner quelques clés afin de mieux comprendre ces biopolymères complexes à partir de l'étude de 3299 structures de carbohydrates issus de différents mammifères. Ce qu'on peut noter c'est que malgré la diversité des sucres simples, seulement 6 sucres simples représentent la plus grande majorité des unités de sucres biologiques, soit presque 96% des unités monosaccharidiques chez ces polysaccharides mammifères étudiés par Werz *et al.* (Werz et al. 2007). On retrouve d'ailleurs des lectines spécifiques pour chacun de ces 6 monosaccharides principaux (Sharon and Halina 2003) bien que les lectines qui reconnaissent les saccharides simples possèdent en général une affinité faible de l'ordre du millimolaire (Tableau 22).

Tableau 22: Lectines reconnaissant les 6 principaux monosaccharides
(Tableau extrait du livre de Sharon et Halina (Sharon and Halina 2003)).

Saccharide reconnu	Lectines
Man	<i>Allium sativum</i> ; <i>Canavalia ensiformis</i> ; <i>Crocus sativus</i> ; <i>Dioclea grandiflora</i> ; <i>E. coli type 1 fimbriae</i> ; <i>ERGIC-53</i> ; <i>Galanthus nivalis</i> ; <i>MBLs of animals</i> ; <i>Pisum sativum</i> ; <i>Vicia faba</i>
Fuc	<i>Aleuria aurantia</i> ; <i>Anguilla anguilla</i> ; <i>Lotus tetragonolobus</i> ; <i>Pseudomonas aeruginosa lectin II</i> ; <i>Ulex europaeus lectin I</i> ; <i>Ulva lactuca</i> ; <i>Chromobacterium violaceum lectin</i>
Gal/GalNAc	<i>Arachis hypogaea</i> ; <i>Coprinus cinereus</i> ; <i>Entamoeba histolytica</i> ; <i>Erythrina corallodendron</i> ; <i>Dolichos biflorus</i> ; <i>Glycine max</i> ; <i>Griffonia simplicifolia lectin I</i> ; <i>helix pomatia</i> ; <i>Hygrophorus hypothejus</i> ; <i>Phaseolus limensis</i> ; <i>Moluccella laevis</i> ; <i>Polyandrocarpa misakiensis</i> ; <i>Ptilota filicina</i> ; <i>Ricinus communis</i> ; <i>Pseudomonas aeruginosa lectin I</i>
GlcNAc	<i>Conglutinin</i> ; <i>Griffonia simplicifolia lectin II</i> ; <i>Tachylectin-2</i> ; <i>Triticum aestivum</i> ; <i>Ulex europaeus lectin II</i> ; <i>Psathyrella velutina</i>
NeuAc	<i>Achatina fulica</i> ; <i>Cancer antennarius</i> ; <i>Hericium arinaceum</i> ; <i>Homarus americanus lectin I</i> ; <i>Limax flavus</i> ; <i>Scylla serrata</i> ; <i>Triticum aestivum</i> ; <i>Psathyrella velutina</i>

De plus, au sein des oligosaccharides, bien que de nombreuses combinaisons soient possibles, on peut noter que certains types de liaisons sont plus abondants. En effet, Werz *et al.* ont pu montrer que 80% des 3299 structures d'oligosaccharides de mammifères étudiées sont représentés par seulement 13 unités monosaccharidiques en prenant en compte le mode (α ou β) et la position de la liaison (Figure 89).

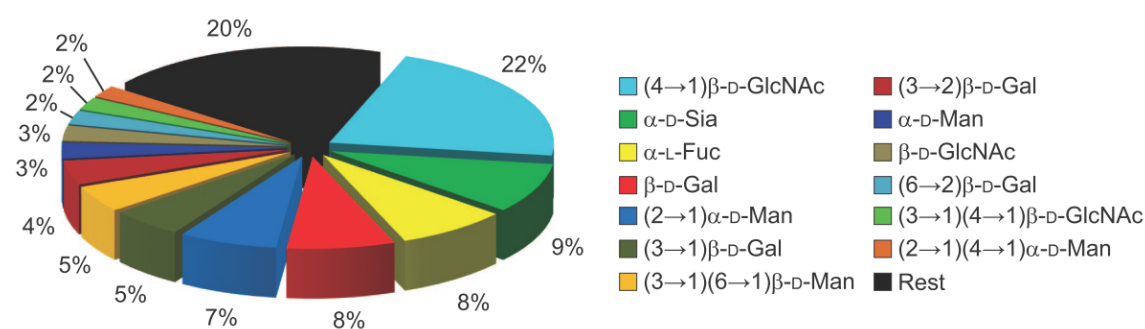


Figure 89: 13 unités monosaccharidiques forment la majeure partie (80%) des oligosaccharides chez les mammifères
Image extraite de la publication de Werz *et al.* (Werz et al. 2007)

Décoder la diversité saccharidique reste aujourd'hui encore un vrai défi (Gabijs 2018). Les lectines sont considérées comme les lecteurs des signaux portés par l'information saccharidique (Gabijs 2018) et malgré une affinité généralement faible, elles montrent une très bonne spécificité avec parfois une reconnaissance très stricte, comme la concanavalin A (ConA) ou l'agglutinine WGA qui sont capables de différencier la conformation du carbone anomérique de leurs ligands (Sharon and Halina 2003).

II.2 Types de liaisons permettant la reconnaissance protéine - sucre

Nous allons voir les principaux types d'interaction qui permettent cette reconnaissance très spécifique protéine - sucre.

II.2.1 Les liaisons Hydrogènes

La liaison d'hydrogène a été décrite pour la première fois par Linus Pauling (Pauling 1963). Elle est formée entre un donneur et un accepteur de liaison d'hydrogène. Le donneur est constitué d'un hydrogène lié à un hétéroatome (fonctions amine, alcool ou thiol, noté X-H), tandis que l'accepteur (noté A) est un hétéroatome portant un doublet d'électrons non liant (azote, oxygène par exemple). La liaison entre l'hydrogène et l'hétéroatome, auquel il est lié de manière covalente, est fortement polarisée à cause de la différence d'électronégativité des deux atomes. Cela crée une charge partielle positive sur l'atome d'hydrogène qui va alors interagir avec un atome électronégatif (accepteur de liaison H). La nature de cette liaison est donc électrostatique et son intensité est considérée comme intermédiaire entre une interaction de Van der Waals et une interaction covalente. La force de la liaison hydrogène ($X^{\delta-}-H^{\delta+}\cdots A$) varie en fonction de la distance, de la polarité de la liaison X-H (donc l'électronégativité de X), et de l'électronégativité de l'accepteur (A). Les saccharides possèdent de nombreuses fonctions hydroxyles qui sont à la fois donneur et accepteur de liaisons hydrogènes (grâce à l'hydrogène porté par l'oxygène et cet atome d'oxygène). On peut donc retrouver plusieurs liaisons hydrogènes sur la même fonction alcool. Les liaisons hydrogènes impliquées dans la reconnaissance des sucres vont ainsi s'établir entre les nombreux groupes hydroxyles ou des fonctions amines portées par le saccharide et directement avec des acides aminés polaires présents dans le site de reconnaissance (fonctions OH, C=O ou NH) ou par l'intermédiaire de molécules d'eau conservées dans le site. L'orientation des fonctions hydroxyles (axiale ou équatoriale) est spécifique de chaque sucre et cette orientation va permettre ou non l'établissement des liaisons hydrogènes au niveau du site de reconnaissance. Les liaisons hydrogènes portent ainsi la spécificité de l'interaction lectine-sucre. Les sucres sont également parfois aminés (glucosamine, galactosamine) et portent donc également des fonctions amines qui peuvent elles aussi générer des liaisons hydrogènes stabilisatrices. Nous pouvons voir sur la Figure 90, le réseau de liaisons hydrogènes qui intervient dans la liaison du 3'sialyl-lactose au sein du CBM40 issu de *S. pneumoniae*.

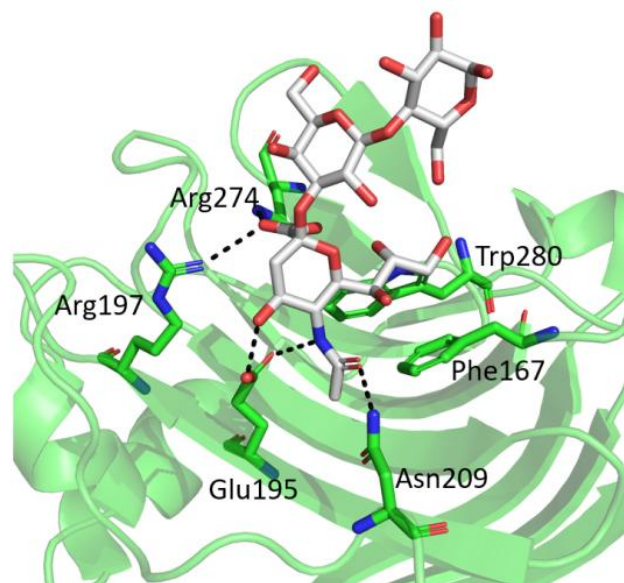


Figure 90: Schéma du site actif de CBM40 (issus de NanA de *Streptococcus pneumoniae*), (PDB 4C1W)
 Les liaisons hydrogènes entre la protéine et les acides aminés sont montrés en pointillés noirs, entre les acides aminés (Glu195, Arg197, Arg274 et Asn209) et le 3'sialyl-lactose.

II.2.2 Interactions cycle aromatique et saccharide

L'interaction entre un cycle aromatique et un saccharide est une interaction complexe car composée de plusieurs types d'interaction. Tout d'abord un contact hydrophobe entre le cycle osidique et le cycle aromatique permet la stabilisation du complexe. En effet, les molécules hydrophobes ont tendance à se regrouper en milieu aqueux afin de limiter leur contact avec l'eau. Le cycle osidique porte des hydroxyles hydrophiles et des hydrogènes hydrophobes. Selon l'orientation de ces fonctions, le saccharide peut présenter une ou plusieurs faces hydrophobes ne présentant que des fonctions hydrogènes comme par exemple le β -D-mannopyranose (Figure 91C) ou les deux faces du glucopyranoside (Figure 91A). Au niveau de la solvatation du saccharide, l'eau sera donc présente au niveau des hydroxyles et pourra par exemple former des liaisons hydrogènes, par contre les hydrogènes ne seront pas en contact avec des molécules d'eau et seront plutôt en contact avec un composé hydrophobe, par exemple un cycle aromatique. A cela vient s'ajouter une interaction de type électrostatique (interaction CH- π) entre les électrons de l'orbitale π d'un cycle aromatique et les charges partielles positives des hydrogènes portés par le cycle osidique (Hudson et al. 2015). En effet, l'oxygène de la fonction hydroxyle est très électronégatif et va générer une charge partielle positive sur le carbone du cycle ainsi que sur l'hydrogène par un effet inductif (effet de propagation). Cette interaction CH- π peut intervenir selon différentes orientations (sur les deux faces du sucre, ou sur différents hydrogènes du cycle) ce qui va favoriser une position spécifique au saccharide ligand comme présentée dans la Figure 91 ci-dessous.

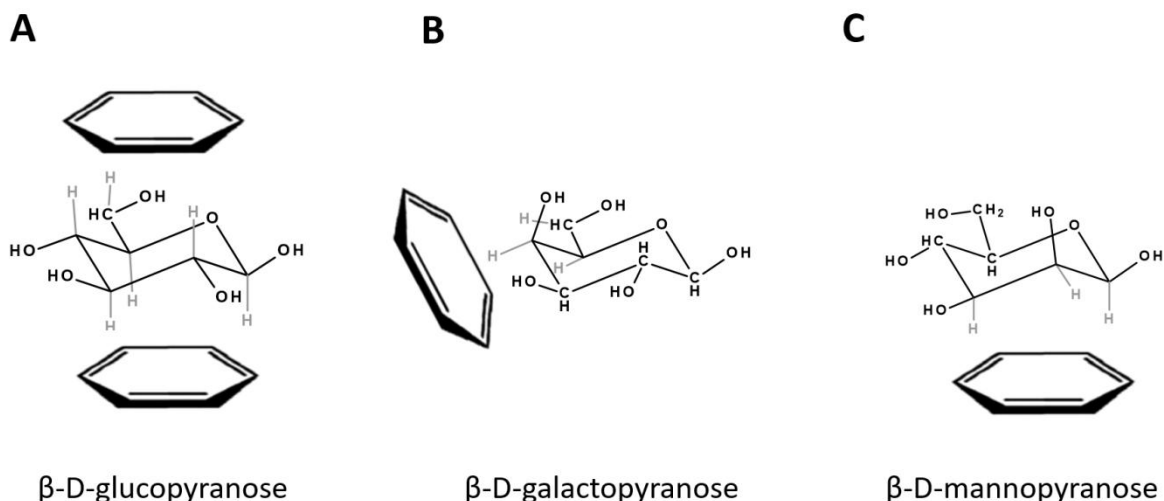


Figure 91: Orientation les plus fréquentes entre le cycle d'un sucre et un cycle aromatique permettant l'établissement d'interactions CH- π .

Image reproduite à partir de la publication de Spiwok et al. (Spiwok 2017) Le β -D-glucopyranose (A) peut interagir avec les cycles aromatiques sur deux faces (A), le sucre β -D-galactopyranose avec les hydrogènes couplés aux carbones C4, C5 et C6 (B) ou le β -D-mannopyranose avec les hydrogènes couplés aux carbones C1, C2 et C3 (C).

La force des interactions CH- π est dépendante du nombre de contact CH- π , ainsi que de l'intensité de polarisation du lien C-H (Spiwok 2017; Nishio 2011). En effet, des études ont récemment montré que celle-ci est modérément affectée par la position et la nature des différents substituants présents sur le sucre (Unione et al. 2017; Hsu et al. 2016). En effet un substituant électronégatif sur le sucre pourra augmenter cette polarité.

La proportion des acides aminés aromatiques au sein des protéines est globalement faible, cependant leur présence très fréquente (Trp>Tyr>Phe) au niveau des sites de reconnaissance de sucres souligne l'importance de cette interaction dans ce mécanisme (Spiwok 2017; Hudson et al. 2015). De plus, leur remplacement par un autre acide aminé par mutagenèse peut entraîner une perte de la reconnaissance (Spiwok 2017).

Malgré sa nature faible, cette interaction a donc une importance cruciale dans la reconnaissance protéine-carbohydate. D'ailleurs, les interactions protéines-sucres sont généralement faibles et cela fait partie des difficultés auxquelles doit faire face la modélisation moléculaire des interactions protéines-sucres. Ainsi, des fonctions de scores spécialisées pour le *docking* (amarrage moléculaire) de sucres ont été créées afin de mieux les prendre en compte (y compris ces interactions CH- π) (Kerzmann et al. 2008).

II.2.3 Autres types d'interactions

Interactions électrostatiques :

Le ligand saccharidique peut être chargé et interagir avec des acides aminés chargés. Ainsi, chez les Siglecs (type de lectines reconnaissant des sucres sialylés) la formation d'un pont salin entre la fonction carboxylate (porté par l'acide sialique) et un acide aminé chargé positivement comme l'arginine est fréquente (Bornhöfft et al. 2018).

Interactions métal-saccharide :

Des ions métalliques sont parfois retrouvés chez des lectines dans un rôle de stabilisation de la structure ou pour favoriser des multimérisations comme c'est souvent le cas chez les lectines légumineuses (R. D. Brown, Brewer, and Koenig 1977). Cependant, au sein de certaines lectines des ions métalliques (le plus souvent le calcium) interagissent aussi directement avec les deux fonctions hydroxyles du saccharide ligand dans une liaison de coordination (Feinberg et al. 2007; Mitchell et al.

2002). Par exemple, comme présenté dans la Figure 92, au niveau du site actif de la lectine LecA de *Pseudomonas aeruginosa*, la reconnaissance du galactose est effectuée par des liaisons de coordinations entre les fonctions hydroxyles (O3 et O4) et le calcium en plus du réseau de liaison hydrogène (Cioci et al. 2003). Ces liaisons de coordinations interviennent également au niveau de sites de reconnaissance du fucose ou du mannose (“Website Scope: [Http://Scop.Mrc-Lmb.Cam.Ac.Uk/Scop/Data/Scop.b.c.Dd.b.A.Html](http://Scop.Mrc-Lmb.Cam.Ac.Uk/Scop/Data/Scop.b.c.Dd.b.A.Html),” n.d.; Loris et al. 2003; Sudakevitz et al. 2004; Mitchell et al. 2002).

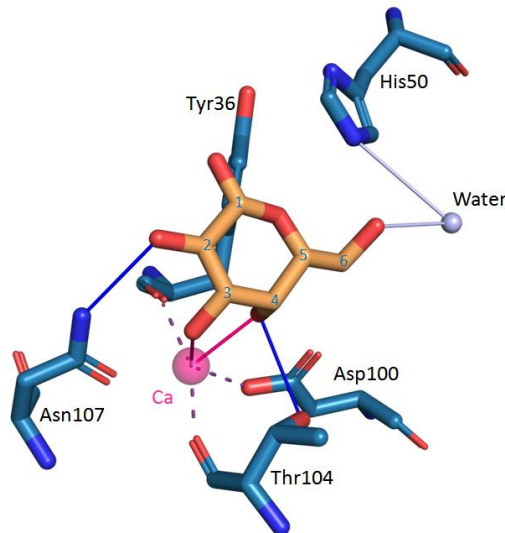


Figure 92: Représentation du site actif de la lectine LecA de *Pseudomonas aeruginosa* complexé avec le galactose (PDB: 4YW7), les carbones du galactose sont numérotés en bleu, les liaisons hydrogènes sont indiquées en bleu ou en blanc si réalisé avec une molécule d'eau intermédiaire, les liaisons de coordinations avec le Calcium sont indiquées en rose. Image réalisée grâce au site web automatisé pour la détection d'interactions protéines ligands : <https://projects.biotec.tu-dresden.de/plip-web/plip/result/588ca750-7e9b-4dbc-981f-5740d497fe90>, ainsi qu'à une analyse complémentaire sur PyMol.

Interactions hydrophobes (Van der Waals) :

Les interactions de type Van der Waals sont issues de la formation de dipôles au niveau de liaisons moléculaires dus à une répartition inégale des nuages électroniques. Ces dipôles peuvent être permanents (dus à des différences d'électronégativité entre les atomes) ou induits (dus à la proximité d'un autre dipôle). Au sein des interactions de Van der Waals, on distingue les forces de Keesom entre deux molécules polaires (dipôles permanents), les forces de Debye entre une molécule polaire et une molécule apolaire (dipôle permanent-dipôle induit), et les forces de London entre deux molécules apolaires (dipôles induits).

Les interactions hydrophobes sont des interactions Van der Waals. Elles sont formées par des forces d'attraction, appelées forces de dispersion de London, entre des dipôles induits formés au sein de molécules non chargées. Au sein d'un solvant aqueux polaire, les molécules hydrophobes vont limiter leur contact avec celui-ci. De plus, les molécules d'eau vont former des liaisons hydrogènes entre elles en excluant la molécule apolaire et former ainsi une interface. L'effet hydrophobe est donc à la fois composé de cette exclusion des composés hydrophobes par l'eau et de la tendance des molécules hydrophobes à s'agréger entre elles grâce aux forces de London. Cependant, les sucres sont généralement considérés comme très hydrophiles de par leurs nombreuses fonctions hydroxyles. Mais comme nous avons pu le voir précédemment (section II.2.2), la disposition des hydroxyles peut entraîner la présence de surfaces hydrophobes au niveau du cycle osidique qui va alors interagir avec des acides aminés aromatiques. De plus, des contacts hydrophobes peuvent également être établis avec les fonctions méthyles de sucres retrouvés dans le fucose, la N-acétylglucosamine (GlcNAc) ou l'acide N-acétylneuraminique (NeuAc).

Par exemple, au sein de la lectine de type C CEL-I, issue de *Cucumaria echinata*, le ligand GalNAc est principalement reconnu par l'intermédiaire de deux liaisons hydrogènes entre l'Arg115 et le groupement carbonyle de l'amide du GalNAc et par des interactions Van der Waals entre le groupe acétyle-amide du GalNAc et les chaînes latérales de la Gln70 et l'Asn123. Le remplacement de ces acides aminés par des alanines à l'aide de mutagenèse dirigée a d'ailleurs pu démontrer l'importance de ces résidus pour les capacités de liaison de la lectine (Sugawara et al. 2004).

III. Méthodes classiques d'ingénierie des lectines

III.1 Applications des lectines

Les saccharides interviennent dans de nombreux phénomènes biologiques. En tant que protéines capables de spécifiquement interagir avec ceux-ci, les lectines ont donc des rôles très variés comme nous avons pu le voir dans la section I.1.1. Des lectines ont donc été utilisées pour des applications en analytique, en diagnostique ou en tant qu'outil thérapeutique.

III.1.1 Utilisation des lectines dans le domaine analytique et diagnostic

La forte spécificité des lectines a été exploitée à des fins d'analyse ou de diagnostic. En effet, l'analyse des sucres peut être effectuée à l'aide de méthodes telle que la spectrométrie de masse ou des méthodes chromatographiques. Cependant, ces méthodes sont limitées par la préparation des échantillons, l'expertise et le matériel nécessaire, et la parallélisation plus complexe des analyses. De nombreux outils basés sur des lectines ont donc été développés et sont répertoriés dans la publication de Hendrickson *et al.* (Hendrickson and Zherdev 2018; Belický, Katrlík, and Tkáč 2016) et présentés dans le Tableau 23. Le développement de biopuces à lectines permet de caractériser rapidement des échantillons biologiques (S. Hu and Wong 2009).

Tableau 23: Présentation des différentes utilisations de lectines en analyse et diagnostic

Application	Méthode	Exemple de lectine utilisé
Identification de sucres	Inhibition de l'hémagglutination	Agglutinine du blé (WGA), Concanavaline A (ConA), Agglutinine <i>Ricinus communis</i> (RCA)
Purification de glycoconjugués	Chromatographie d'affinité à l'aide de lectines	Mélange de la concanavaline A et d'une lectine d' <i>Artocarpus altilis</i>
Marquage histologique et cellulaire	Lectines conjugués à des composés fluorescents, des haptènes ou des particules d'or	Agglutinine du blé (WGA), Concanavaline A (ConA)
FACS	Marquage et tri cellulaire à l'aide de lectines fluorescentes	Lectine recombinante rBC2LCN (domaine N-terminal d'une lectine issue de <i>Burkholderia cenocepacia</i>)
Détection de protéines glycosylées et suivi de changement de glycosylation	Lectin blotting (analogue WesternBlot)	Agglutinine de <i>Helix pomatia</i> (HPA)
Détection de bactéries, champignon, détection de tissus pathologiques (recherche de marqueurs de pathologies)	Microarray de lectines	Différentes lectines immobilisées en spot sur une surface
Détection de différents carbohydate	Biosensor de lectines, label de lectine composé fluorescent, immobilisation sur feuille d'or, ou des nanoparticules d'or	Concanavaline A, PNA (lectine issue de <i>Arachis hypogaea</i>), SNA (lectine issue de <i>Sambucus Nigra</i>), agglutinine de <i>Maackia amurensis</i> (MAA)
Détection de différents carbohydates (par exemple dans le cas de cancers métastatiques)	ELLA (Enzyme-like Lectin Assay) (analogue ELISA)	Agglutinine du blé (WGA), concanavaline A (ConA), agglutinine <i>Ricinus Communis</i> (RCA), SNA (Lectine issue de <i>Sambucus Nigra</i>), agglutinine de <i>Maackia amurensis</i> (MAA)

III.1.2 Utilisation des lectines en thérapeutique

III.1.2.1 Propriétés antitumorales

Des modifications de profils de glycosylations peuvent être impliquées dans de nombreuses maladies, telles que le cancer (Orntoft and Vestergaard 1999; Varki et al. 2009), les maladies liées au foie (cirrhose, cancer du foie, fibrose du foie) (Blomme et al. 2009), la maladie d'Alzheimer (J. Z. Wang, Grundke-Iqbal, and Iqbal 1996) ou des maladies congénitales (Jaeken and Carchon 1993). Les méthodes analytiques précédentes peuvent être ainsi utilisées afin de réaliser des diagnostics (section III.1.1). En plus de cela, des méthodes thérapeutiques contre le cancer impliquant les lectines ont récemment été développées (Bhutia et al. 2019). En effet, des lectines ont démontré des capacités anticancéreuses par plusieurs mécanismes dont des exemples des différents types de mécanismes sont listés dans le Tableau 24 (Yau et al. 2015). Les mécanismes impliqués sont : l'induction de l'apoptose ou de l'autophagie, leurs effets cytotoxiques ou encore par la modulation de la production de cytokines (Yau et al. 2015; S. Coulibaly and C. Youan 2017). Elles peuvent également être utilisées pour cibler des cellules cancéreuses grâce à leur reconnaissance spécifique de profils anormaux de glycosylation (Hashim, Jayapalan, and Lee 2017; S. Coulibaly and C. Youan 2017).

Tableau 24: Différents exemples de lectines ayant des potentiels antitumoraux selon différents mécanismes

Mécanisme	Exemple de lectine
Induction de l'apoptose par différentes voies de signalisations	Galectines (Nakahara and Raz 2008), Plant lectins (Bhutia et al. 2019; Jiang et al. 2015), lectine reconnaissant l'acide sialique (X. Yang et al. 2014; Ning Wang, Whang, and Lee 2008), lectine du gui (Bussing 1996)
Induction de l'autophagie	Concanavalin A (Lei and Chang 2007), Lectine de <i>Agrocybe aegerita</i> (W. Liu et al. 2017), Plant lectins (Bhutia et al. 2019; Jiang et al. 2015)
Immunomodulation (cytokines)	Galectines (Nakahara and Raz 2008), Annexines (Z. Zhang et al. 2010), lectine du gui (Lyu and Park 2007; Yau et al. 2015)
Reconnaissance profils anormaux de glycosylations (conjugué à des composés cytotoxiques)	Lectines de type C (Nonaka et al. 2008; Shimomura et al. 2018)

L'utilisation de lectines en tant qu'agent anti-tumoraux a même été étudiée dans des études précliniques et cliniques. Par exemple, l'injection intratumorale d'une lectine issue de *Viscum album* (Quercus; Iscador® Qu) a montré une grande efficacité pour la suppression d'adénomes du colon sans effets secondaires lourds chez 123 patients. L'utilisation d'une lectine issue du gui a également permis l'amélioration de la qualité de vie de patient lors de l'utilisation de traitements chimiothérapeutiques contre le cancer du sein (Bhutia et al. 2019).

III.1.2.2 Autres propriétés

La reconnaissance des lectines pour les glycoprotéines a également démontré des potentiels antiviraux, antimicrobiens, antifongiques ainsi qu'anti-insectes. En effet, la lectine Griffithsin (GRFT) a

également montré des propriétés anti-virales dirigées contre des virus enveloppés tels que le virus du SIDA, de l'hépatite C, du coronavirus, ainsi que du virus Ebola grâce à ses capacités de liaisons aux glycoprotéines telles que gp120 et gp41 riches en mannoses. (Lusvarghi and Bewley 2016; Barton et al. 2014). De même des lectines peuvent allonger le cycle de développement d'insectes, voire augmenter leur taux de mortalité chez des insectes tels que les pucerons, la mouche du melon, la tique, la bruche du niébé, ou encore certains moustiques (Lam and Ng 2011). De nombreuses lectines issues de plantes et particulièrement des graines de plantes ont également démontré des capacités antifongiques. Ainsi, le développement de leur utilisation pourrait alors être une alternative aux pesticides ou antifongiques chimiques réduisant ainsi les risques sanitaires.

Afin de permettre ces applications, plusieurs stratégies d'ingénierie des lectines ont été développées pour élargir l'éventail des lectines disponibles ou modifier leurs caractéristiques de reconnaissance. Deux types de stratégies sont utilisés pour l'ingénierie des lectines : soit la modification de leur spécificité par des méthodes d'ingénierie protéique classiques, soit la modification de leur valence.

III.2 Ingénierie des lectines - Modification de la spécificité

III.2.1 Production de lectines recombinantes

Traditionnellement, les lectines sont directement extraites de leurs sources d'origine par des méthodes de précipitation couplées à des méthodes chromatographiques. Cependant, les lectines peuvent posséder des isoformes qui ont des activités différentes et ce mélange rend difficile l'obtention de fractions pures. De plus, ces méthodes d'extraction conduisent généralement à des rendements assez faibles (Martínez-Alarcón, Blanco-Labra, and García-Gasca 2018).

Le développement de la production de lectines recombinantes permet de répondre à ces problématiques par la production d'une seule forme de la lectine et conduit généralement à des rendements plus élevés. Les systèmes d'expression les plus utilisés sont les bactéries ou les levures. Les lectines sont naturellement présentes dans tous les compartiments cellulaires (milieu extracellulaire, réticulum endoplasmique, appareil de golgi, cytoplasme, membrane) ce qui influe sur le choix de l'organisme de production. De plus certaines lectines sont glycosylées ou utilisent des ions en tant que co-facteurs. La production de lectines dans les bactéries n'est donc pas toujours le meilleur système d'expression hétérologue, d'autant que les lectines à haut poids moléculaire peuvent être accumulées dans des corps d'inclusions lorsqu'elles sont exprimées à fortes concentrations. Malgré ces limitations, des lectines eucaryotiques ont été produites efficacement chez *E. coli*. Pour les lectines glycosylées, la production chez *Pichia pastoris* est privilégiée. En effet, ce système d'expression permet la production de protéines glycosylées à haute concentration ainsi que leur excrétion dans le milieu ce qui permet une très forte production. De plus, grâce à l'utilisation du promoteur AOX1 les glycosylations de *Pichia pastoris* ne sont pas antigéniques et sont assez proches des glycosylations des organismes eucaryotes supérieurs. La production de lectines a également été réalisée chez les plantes ou au sein de cellules animales. Cependant ces systèmes sont plus complexes d'utilisation et plus coûteux. Le choix du système hétérologue de production sera donc dicté par différents critères tels que la localisation cellulaire, la présence de site de glycosylation et leurs caractéristiques, de ponts disulfures, ou l'utilisation de protéines chaperonnes pour son repliement (Martínez-Alarcón, Blanco-Labra, and García-Gasca 2018).

III.2.2 Méthodes d'ingénierie protéique

La production de lectines par méthodes recombinantes à partir d'un gène d'expression permet l'utilisation de méthodes d'ingénierie protéique afin d'étudier et de modifier les

caractéristiques de reconnaissance des lectines. Ces méthodes d'ingénierie protéique sont essentiellement basées sur des méthodes de mutagenèse couplées à des méthodes de sélection (Figure 93). Différentes tailles de librairie peuvent être générées. En effet, lorsque la structure 3D de la lectine est connue, les stratégies de mutagenèse peuvent être plus ciblées (Rahman et al. 2015). Cependant, les stratégies de mutagenèse très ciblées sont plus risquées car la reconnaissance protéine-sucre est déterminée par des interactions faibles et il est difficile de prédire l'effet d'une mutation (D. Hu, Tateno, and Hirabayashi 2015). Les bibliothèques les plus larges sont générées par des méthodes aléatoires telles que la PCR à erreur, ou le DNA shuffling (Hu et al. 2012; Yabe et al. 2007; Yim, Ono, and Irimura 2002), tandis que les bibliothèques plus ciblées sont générées par mutation saturante ou dirigée (PCR inverse, Quickchange, ou synthèse du gène) (Arango et al. 1993; Jordan and Goldstein 1995; V Sharma, Vijayan, and Surolia 1996; Nishiguchi et al. 1997; Hatakeyama et al. 2013; Salomonsson et al. 2010; Kolatkar and Weis 1996; Topin et al. 2014; Adam et al. 2007; Kříž et al. 2014; Mishra et al. 2012; Keogh et al. 2014; Liang et al. 2016).

Selon la taille de la librairie, différentes méthodes de sélection vont être utilisées. De nombreuses méthodes de type display ont été développées lors de l'ingénierie d'anticorps tels que le phage display, le ribosome display ou encore le yeast display et peuvent également être utilisées pour l'ingénierie des lectines (Ryckaert et al. 2008; Nehmann et al. 2009). Une méthode de type display consiste à lier l'information génétique à la protéine produite. Le mutant est alors sélectionné sur ces caractéristiques de reconnaissance au niveau de la protéine puis la séquence génétique est analysée par séquençage. En effet, le séquençage de nucléotides est moins cher et plus rapide que le séquençage protéique.

Le ribosome display est une des méthodes display les utilisées notamment parce que cette méthode n'est pas limitée par l'efficacité de transformation de la librairie et présente donc de meilleures capacités de criblage (D. Hu, Tateno, and Hirabayashi 2015; Swimmer et al. 1992; K Yamamoto, Maruyama, and Osawa 2000; M. Yamamoto, Kominato, and Yamamoto 1999). Au cours de cette méthode, l'expression protéique est réalisée *in vitro* et les complexes ribosomes, ARN messager et protéine sont purifiés et sélectionnés sur le ligand cible. L'ARNm des lectines liants la cible est ensuite converti en ADN par RT-PCR avant d'être séquencé ou réinjecté dans un nouveau cycle de sélection. Cependant, des ajustements spécifiques à la sélection de lectines par ribosome display ont été réalisés. En effet, l'affinité des lectines est généralement assez faible, les temps d'incubation lors de la sélection sont donc plus longs que pour la sélection d'anticorps et il y a alors un risque pour que le complexe ARNm-ribosome-protéine ne soit pas assez stable, l'utilisation de codons rares en région 3' favorise alors cette interaction (Yabe et al. 2007). D'autre part, l'utilisation de polymères de sucres biotinylés structurellement bien définis à la place de glycoprotéines ainsi que l'ajout d'inhibiteurs compétitifs rendant la sélection plus stringente a permis d'améliorer également les capacités de sélection (D. Hu et al. 2012). Enfin, le développement de puces de glycoconjugués devrait également faciliter la vérification des mutants sélectionnés.

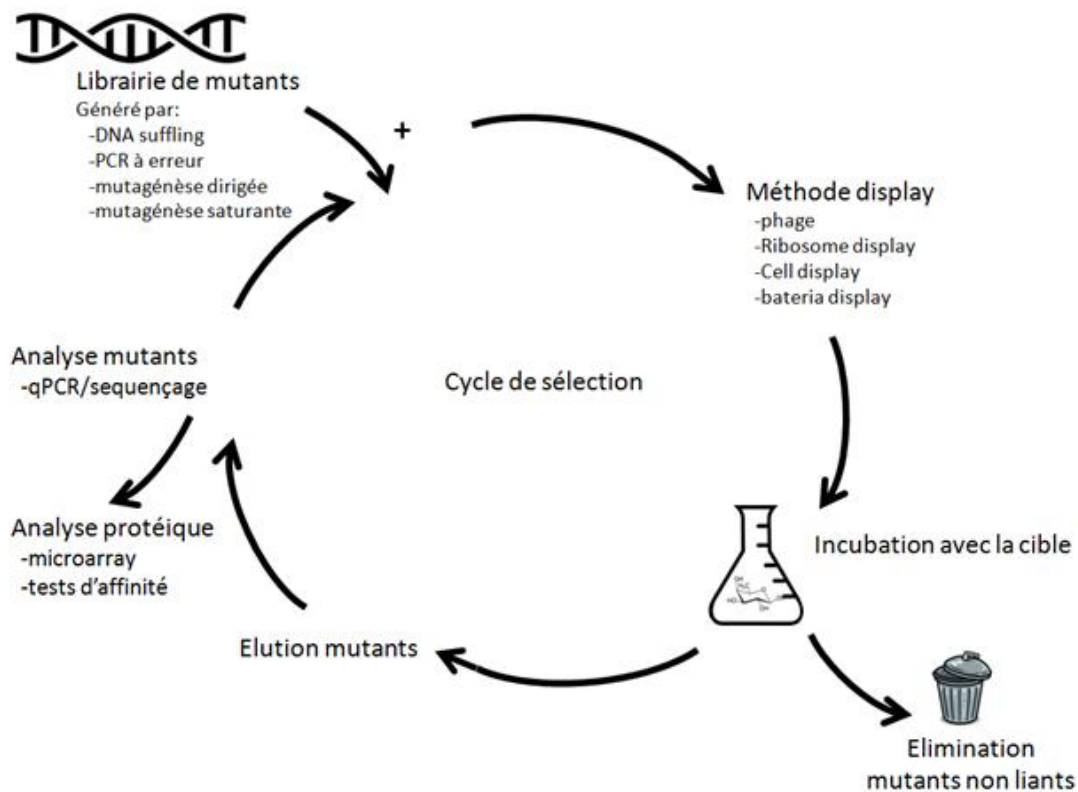


Figure 93: Principe des méthodes de sélections

III.3 Ingénierie des lectines - Modification de la valence des lectines

Les lectines ont la particularité d'être souvent multivalentes, cela a cependant des avantages et des inconvénients. En effet, les lectines multivalentes ont la particularité d'avoir une meilleure affinité apparente mais elles ont aussi la capacité à agglutiner les cellules, ce qui peut être un préjudice selon l'application recherchée comme par exemple pour du tri cellulaire. L'utilisation de modifications chimiques pour prévenir l'oligomérisation (Gunther et al. 1973; Beppu, Terao, and Osawa 1976) ou encore l'élimination des ponts disulfures par réduction (H Kaku et al. 1993; Hanae Kaku and Shibuya 1992) ou par remplacement des cystéines par des sérines (H. Wang et al. 2008) a permis de produire des lectines monovalentes. Plus récemment, un fragment de lectine à tandem de répétition a été produit de manière indépendante chez *E.coli* (Olausson et al. 2011).

Parfois cependant, la multimérisation des lectines a été recherchée afin d'augmenter l'affinité d'une lectine donnée par un effet d'avidité. Pour cela, des lectines ont été formées par des méthodes de protéines chimères à l'aide de *linker* polypeptidique (Yabe et al. 2009), de l'utilisation de biotinylation afin de clustériser des lectines (Kawasaki et al. 2007) ou encore par de la mutagenèse afin d'introduire des cystéines pouvant former des ponts disulfures (McDonald, Hughes, and Davis 2004). Enfin des lectines chimères ont également été produites afin de combiner deux spécificités (Kazuo Yamamoto et al. 1992; Ribeiro et al. 2018).

IV. Ingénierie protéique avec des acides aminés non naturels

IV.1 Utilisation d'acides aminés non naturels au sein de site de reconnaissance ou site catalytique

La reconnaissance spécifique d'un substrat par une protéine est principalement due aux types de liaison que celui-ci peut former avec les chaînes latérales des acides aminés présents au sein

du site de reconnaissance selon leur proximité et selon leur orientation. Les protéines sont essentiellement formées de 20 acides aminés naturels possédant des chaînes latérales variées avec des caractéristiques physicochimiques qui peuvent permettre la formation d'interactions de type polaires, hydrophobes et électrostatiques. Les modifications post-traductionnelles ainsi que l'utilisation de co-facteurs permettent d'assurer les fonctions protéiques qui ne sont pas disponibles par l'utilisation de ces 20 acides aminés naturels. L'utilisation d'acides aminés non naturels permet alors d'apporter de nouvelles chaînes latérales avec des nouvelles caractéristiques ou de mimer les modifications post-traductionnelles. Les méthodes d'incorporation d'acides aminés non naturels au sein des protéines ont déjà permis d'incorporer plus de 200 NAA différents au sein de protéines (Smolskaya and Andreev 2019). Ils peuvent être incorporés de manière site-spécifique ou de manière plus globale en remplaçant un type d'acide aminé naturel (la méthionine par exemple) par un autre (Chapitre 1, section IV).

Des acides aminés non naturels ont déjà été incorporés au sein de sites actifs ou sites de reconnaissance afin de modifier leurs caractéristiques comme répertorié dans les revues récentes de Mayer *et al.* et de Ravikumar *et al.* (Ravikumar *et al.* 2015; Mayer 2019) ainsi que dans la publication méthodologique de Hu *et al.* (C. Hu and Wang 2016). Nous allons voir ci-dessous leurs différentes applications.

IV.1.1 Sondes biophysiques

Des acides aminés peuvent être introduits au sein de sites protéiques afin de faciliter l'étude structurale ou mécanistique par RMN, cristallographie au rayon X, IR ou réticulation comme c'est le cas pour des acides aminés non naturels, portant des atomes lourds, fluorescents ou pouvant servir de sondes radicalaires (Dougherty 2000; Ravikumar *et al.* 2015).

- Utilisation en RMN :

L'utilisation d'acides aminés non naturels en tant que sonde radicalaire est liée à la méthode SDSL (Site directed Spin Labeling). Elle consiste à utiliser à une position donnée un marqueur de spin, c'est à dire un radical, donc une molécule avec un nombre impair d'électrons. Analogue à l'utilisation d'isotopes en RMN, elle permet l'introduction d'un spin non nul actif en RMN, ce qui permet l'étude de l'environnement du radical. Traditionnellement, la méthode SDSL utilisait le greffage de radicaux sur des cystéines, mais les cystéines sont parfois impliquées dans des rôles structuraux chez les protéines. L'utilisation de la réactivité orthogonal d'acides aminés non naturels ou le développement d'acides aminés non naturels radicalaires ont donc été utilisés dans ce sens afin de caractériser la structure de protéines ou la liaison protéine-ligand (Fleissner *et al.* 2009; Jones *et al.* 2010; Cornish *et al.* 1994). L'utilisation d'acides aminés contenant des isotopes, des atomes lourds ou des dérivés de fluor sont également utilisés en RMN pour caractériser la liaison entre la protéine et son ligand (Jones *et al.* 2010).

- Utilisation lors de l'analyse cristallographique par rayon X :

Les atomes lourds sont notamment utilisés en cristallographie au rayon X afin de répondre au problème de phase classique. Ce problème est issu de la perte d'information (concernant la phase de l'onde lumineuse) lors des différentes mesures par des détecteurs qui ont plutôt tendance à uniquement mesurer l'intensité lumineuse. Or, lors de l'étude de protéines par cristallographie aux rayons X, l'information de la phase comporte l'information de la densité électronique obtenue par une transformation de Fourier. Cette information est donc capitale. Différentes solutions ont été appliquées pour répondre à ce problème dont la MAD (Multi-wavelength Anomalous Dispersion), la MIR (Multiple Isomorphous Replacement), la MR (Molecular Replacement) ou la SAD (Single wavelength Anomalous Dispersion). La méthode la plus utilisée est la méthode MAD dans laquelle les méthionines de protéines ont été remplacées par des sélénométhionines ou des tellurométhionines (Hatch *et al.* 2005).

- Analyse par infrarouge IR :

L'introduction d'acide aminé non naturel ayant des propriétés vibrationnelles uniques à une position précise au sein de la protéine permet d'identifier des changements structuraux comme par exemple lors de la liaison d'un ligand. Des acides aminés portant des atomes lourds, des dérivés isotopiques (deutérium) ou des groupements NO, CN, SCN ou N₃ ont pu être utilisés (T. S. Young and Schultz 2010; Ye et al. 2009).

- Acides aminés fluorescents :

Les protéines contiennent naturellement des acides aminés fluorescents : Phénylalanine, Tryptophane, Tyrosine. Ils sont présents en faible proportion au sein des protéines (environ 1% pour le plus abondant). De plus, le Tryptophane est particulièrement sensible à son environnement tel que des changements conformationnels, une dénaturation ou une liaison au substrat. Cela a permis le développement d'une méthode de mesures d'interactions par spectroscopie (Lakowicz and Lakowicz 2006). Cependant, lorsque plusieurs tryptophanes sont présents dans la protéine, il est difficile de différencier leurs émissions spectrales. L'utilisation d'acides aminés fluorescents permet alors l'introduction d'une sonde fluorescente avec des caractéristiques spectrales différentes des acides aminés naturels et permet d'apporter des informations structurales comme par exemple lors de l'étude d'une galactosidase réalisée par Steward *et al.* (Steward et al. 1997).

- Expérience de capture du ligand par réticulation :

Dans le cas d'études récepteur-ligand peptidique ou d'interaction protéine-protéine, l'utilisation d'acides aminés non naturels capables de réaliser de la photo-réticulation, c'est à dire de générer une liaison suite à une irradiation, permet de lier de façon covalente deux partenaires et ainsi pouvoir étudier plus facilement des interactions faibles (Xu and Wu 2014). Cette stratégie permettant de capturer et d'identifier des partenaires protéiques a été utilisée afin d'étudier les interactions entre différentes sous-unités protéiques d'un canal ionique (Murray et al. 2016), ou encore entre un récepteur couplé aux protéines G (GPCR) et son ligand peptidique (Becker and Naider 2015). Cependant ce type de réticulation est très réactif avec tout élément assez proche de la protéine. Un acide aminé non naturel capable de spontanément et spécifiquement réagir avec des résidus thiols à proximité a été récemment développé (B. Yang et al. 2017). De manière générale, l'acide aminé non naturel doit cependant être incorporé à une position permettant la capture du partenaire protéique tout en n'empêchant pas leur interaction.

IV.1.2 Etude des caractéristiques d'un acide aminé clé

Au sein de sites actifs ou de reconnaissance, certains acides aminés peuvent être considérés comme étant clés pour l'efficacité de la protéine. Afin de prouver quelles caractéristiques de l'acide aminé sont favorables, l'utilisation d'acide aminé non naturel permet de remplacer celui-ci par un acide aminé portant une modification (qui augmente l'hydrophobicité ou la polarité par exemple). Contrairement à une mutation choisie parmi des acides aminés naturels, l'acide aminé non naturel sera le plus proche possible de l'acide aminé d'origine ce qui facilitera les conclusions (Dougherty 2000). Cette stratégie a été utilisée pour mieux comprendre l'interaction CH- π au sein d'une lectine (RSL) comme détaillée dans la section IV.2.4, mais également au sein d'enzymes au niveau de la poche hydrophobique d'une nucléase (Ting et al. 1998), ou au sein du récepteur nicotinique de l'acétylcholine (Zhong et al. 1998). Selon une méthode similaire, le rôle d'autres types d'acides aminés a pu également être étudié tels que le rôle de tyrosines au niveau de l'entrée d'un canal ionique (Beene et al. 2004), ou au sein d'oxydases (Y. Yu et al. 2015). En effet, dans le cas de liaisons hydrogènes, il est difficile d'étudier ces interactions par des mutations classiques. Par exemple, dans le cas de la mutation d'un glutamate ou d'un aspartate en glutamine ou asparagine, un fort donneur de liaison hydrogène est remplacé par un fort accepteur de liaison hydrogène, ce qui complique les conclusions. Cependant, des analogues de glutamate ou aspartate contenant des groupes nitro ou

amide peuvent être utilisés afin de seulement neutraliser le donneur de liaison hydrogène de façon plus neutre (Pless and Ahern 2013).

IV.1.3 Liaison covalente de cofacteurs (notamment ions métalliques)

Les enzymes utilisent parfois des cofacteurs, tels que des ions métalliques pour assurer leur activité. Etant donné que des acides aminés non naturels peuvent fixer des co-facteurs à l'aide de réactions orthogonales ou même se lier directement à des ions métalliques avec une forte affinité, ceux-ci ont été utilisés afin de modifier des protéines. Par exemple, l'acide aminé bipyridylalanine a été incorporé au sein de protéines afin de fixer le Cuivre catalyseur (Drienovská et al. 2017, 2015; H. S. Lee and Schultz 2008). Des co-facteurs (Rh₂-tetraacetate, complexes Mn ou Cu-terpyridine) ont également été fixés à la protéine grâce à l'incorporation de l'azidophénylalanine et de la réaction de "chimie click" orthogonale SPAAC (H. Yang et al. 2014). Cette association plus forte entre la protéine et l'ion métallique permet de réduire la formation de sous-produits et facilite la réaction dans un milieu aqueux (Y. Yu et al. 2018; Mirts, Bhagi-Damodaran, and Lu 2019).

IV.1.4 Contrôle optiques des propriétés catalytiques

Des acides aminés non naturels portant des fonctions photoprotégées et dérivés de cystéines, de sérines, de lysines et de tyrosines ont été incorporés. Ils permettent de contrôler l'activité enzymatique grâce à une photodéprotection de l'enzyme pouvant être réalisée *in vitro* ou même *in vivo*. En effet, les chaînes latérales des acides aminés sont directement impliquées dans les mécanismes catalytiques. En ce sens, ces acides aminés photoprotégés ont été introduits au sein de différents sites actifs pour contrôler l'activité d'enzymes telles qu'une protéase (Nguyen et al. 2014), une galactosidase (Deiters et al. 2006), une recombinase (W. Brown and Deiters 2019), une DNA et une ARN polymérase (Chou, Young, and Deiters 2009; Hemphill et al. 2013), une kinase (Gautier, Deiters, and Chin 2011), un canal ionique (Kang et al. 2013) ou encore une nucléase (Chou and Deiters 2011). L'utilisation d'acide 2,3-diaminopropionique photoprotégé a également été exploité afin de capturer à l'aide d'une liaison amide des intermédiaires réactionnels jusqu'alors inobservables au sein de protéases à cystéine (TEV) et du domaine thioestérase d'une synthétase de valinomycine (Huguenin-Dezot et al. 2019). Des séquences comportant ces acides aminés peuvent également être impliquées dans des modifications post-traductionnelles comme la phosphorylation. L'incorporation d'une sérine photoprotégée au sein d'un facteur de transcription a par exemple permis de rendre sa phosphorylation photoactivable (Lemke et al. 2007). L'introduction d'une lysine photoprotégée a également permis de contrôler l'infection de cellules HEK293T par un adénovirus AAV pour la rendre photoinductible (Erickson et al. 2017).

IV.1.4 Modification de sa régiosélectivité ou énantiosélectivité

La capacité des enzymes à modifier régiosélectivement des molécules ainsi que leur capacité à discriminer des énantiomères entre eux est très recherchée.

La régiosélectivité d'une enzyme P450 a été modifiée lors d'une étude réalisée par Kolev par l'incorporation d'analogues de tyrosine. En effet, l'incorporation de NapA (3-(2-naphthyl)alanine) à la place de l'alanine 32 a en effet permis une conversion du (S)-ibuprofène en 5% d'alcool benzylique et 95% d'alcool allylique contre 62% et 38% respectivement pour l'enzyme sauvage (Kolev et al. 2014).

Quant à la sélectivité envers un énantiomère, un tryptophane 222 situé à l'entrée d'un site de reconnaissance d'une réductase qui convertit des cétones en alcools chiraux a été remplacé par des analogues de tyrosine. Le tryptophane 222 permet le positionnement du substrat lors de son entrée dans le site et a donc un rôle critique. La modification de celui-ci a entraîné des changements structuraux et a permis de changer l'énantiosélectivité de l'enzyme. L'enzyme sauvage possède déjà une préférence pour l'énantiomère R de la 2-chloro-1-phényléthanone mais avec un excès de 9,1%

seulement. L'incorporation par exemple de la 2-*ter*-butyle-L-tyrosine a permis d'augmenter cet excès à 33,7% (H. Ma et al. 2014).

Les exemples de modification de la régiosélectivité ou l'énantiosélectivité d'enzymes grâce à des acides aminés non naturels restent encore rares. Mais ces études démontrent le potentiel de leur utilisation pour ces applications.

IV.1.5 Acide aminé non naturel en tant qu'acide aminé catalytique

L'utilisation d'acide aminé non naturel amenant une nouvelle réactivité catalytique est encore très rare. Le groupe de Roelfes a récemment incorporé la *p*-aminophénylalanine au sein de la protéine de la famille des transporteurs ABC (LmrR) issue de *Lactococcus lactis*. Cette protéine forme des homodimères et possède une large poche de liaison hydrophobique à l'interface de ces dimères, ce qui a été également exploité par le groupe de Roelfes au cours de la création de métalloenzymes artificielles (Roelfes 2019). Ici, l'incorporation de la *p*-amino-phénylalanine lui permet alors de catalyser la formation d'hydrazone ou d'oximes par condensation entre des carbonyles et des hydrazines ou hydroxylamines (Drienovská et al. 2018). Puis une évolution dirigée a permis d'augmenter encore cette efficacité catalytique par des mutations qui permettent le positionnement optimal de la *p*-aminophénylalanine. L'enzyme finale possède une capacité catalytique apparente assez importante (k_{cat}/K_m de $132 \text{ M}^{-1}\text{s}^{-1}$) (Mayer et al. 2019).

IV.2 Utilisation des acides aminés non naturels chez les lectines

Bien que moins nombreuses, des stratégies d'ingénierie des lectines ou d'enzymes reconnaissant les sucres et utilisant les acides aminés non naturels ont également été développées afin de modifier leurs caractéristiques de reconnaissance.

IV.2.1 Méthodes de criblage avec des acides aminés non naturels

L'ingénierie classique des lectines est essentiellement basée sur des méthodes de mutagenèse couplées à des méthodes de sélection. Cependant l'utilisation de méthodes systématiques pour tester l'utilisation d'acides aminés non naturels à différentes positions au sein de protéines est plutôt rare. Afin de générer les codons stops TAG au sein des séquences cibles, des méthodes de mutagenèse dirigées classiques (C. C. Liu et al. 2008), de TAG scanning (Daggett, Layer, and Cropp 2009) ou d'échange de trinuécléotides (TriNEx) (Reddington et al. 2015) ont pu être utilisées. A ce jour, l'utilisation des méthodes TriNEx et TAG scanning n'ont cependant pas encore été testées pour l'évolution de lectines. La méthode de mutagenèse dirigée classique utilisée par Liu *et al.* a été couplée avec une méthode de phage display. Cette méthode a cependant nécessité des optimisations de conditions de croissance des phages afin d'avoir une proportion similaire entre les séquences présentant des acides aminés non naturels ou non (C. C. Liu et al. 2008). Cette méthode a ensuite permis de sélectionner un anticorps anti-glucamine (une forme aminée du sorbitol) contenant une boronate-phénylalanine (C C Liu et al. 2009).

Les méthodes précédentes permettent de tester l'incorporation d'un type d'acide aminé non naturel à différentes positions au sein d'une séquence protéique. Cependant, elles ne permettent pas de rapidement tester différentes chaînes latérales. L'incorporation d'acide aminé non naturel de manière site spécifique est très dépendante de la disponibilité de l'aaRS spécifique pour celui-ci, ce qui rend difficile le criblage de plusieurs acides aminés non naturels pour une protéine. Une méthode consiste à générer une déshydroalanine par modification chimique d'une cystéine sur laquelle on pourra ensuite greffer spécifiquement différentes chaînes latérales par addition d'un thiol ou par formation d'un lien $\text{C}(\text{sp}^3)\text{-C}(\text{sp}^3)$. C'est ce qui a pu être testé par Windle *et al.* par la modification de cystéines en déshydroalanine en présence de 2,5-dibromohexan-1,6-diamide. La déshydroalanine a

ensuite réagi avec différents radicaux thiols. Cela a permis de tester 13 radicaux thiols portant des fonctions différentes à 12 positions différentes. La spécificité de l'aldolase a donc pu être modifiée afin de favoriser la reconnaissance de sucres à chaînes plus courtes. En effet, le rapport K_m^{app}/k_{cat}^{app} de la protéine sauvage était de 0,17 et 0,26 pour l'érythrose (C4) et le mannose (C6) respectivement. La mutation de la phénylalanine 190 en 2,3-dihydroxyl-propyl-cystéine (Dpc) a permis de changer ces rapports en 1,7 et 0,09 respectivement (Windle et al. 2017).

IV.2.2 Valence

Comme nous l'avons vu précédemment, les lectines sont souvent multimériques, ce qui permet une augmentation de leur affinité apparente. Lors de la conjugaison de différentes sous-unités de protéines, les acides aminés non naturels ont l'avantage de pouvoir apporter une réactivité orthogonale, ce qui permet d'obtenir des associations spécifiques et contrôlées. Cette stratégie a été utilisée précédemment lors de la création d'anticorps bi-spécifiques (C. H. Kim et al. 2012; Lu, Hua; Zhou 2014; González-miro, Rodríguez-noda, and Fariñas-medina 2017; Mirande 2017; Cao et al. 2015). Lors d'une étude réalisée par Tobola *et al.*, 3 lectines ont été conjuguées à l'aide d'acides aminés non naturels : la *Ralstonia solanacearum* lectine (RSL), la galectine humaine 1 (Gal-1) et la sous-unité B de la Shiga toxine 1 (Stx1B). Ces lectines ont d'abord été associées de manière monospécifique (Gal-1 avec Gal-1, RSL avec RSL et Stx1B avec Stx1B) pour former des lectines multivalentes, puis elles ont été associées entre elles afin de former des lectines chimères bispécifiques. Des liens ont été également ajoutés afin de limiter la gêne stérique entre les différentes sous-unités et améliorer ainsi les réactions de conjugaison. Les affinités des différentes protéines générées n'ont cependant pas été testées (Tobola et al. 2019).

IV.2.3 Liaison boronate

La fonction boronate peut former des liaisons esters avec des cis-diols (section I.3.1) et cette caractéristique a été très exploitée afin de générer des lectines organiques. Cependant, contrairement aux lectines naturelles ces lectines organiques manquent de sélectivité. L'incorporation de la fonction boronate au sein de protéine peut être réalisé par une succession de modifications chimiques basées sur la réactivité des lysines (Lu et al. 2013). Mais cette modification n'est donc pas ciblée. En 2008, une boronate-phénylalanine a été incorporée pour la première fois de manière site-spécifique grâce à une aminoacyl-ARNt synthétase de *Methanocaldococcus jannaschii* (Brustad et al. 2008). La fonction boronate peut donc être incorporée spécifiquement au sein de protéines. En effet, cette fonction est naturellement absente des protéines, même par modifications post-traductionnelles ou au sein de co-facteurs. La synthétase qui permet cette incorporation a été cristallisée (Schiefner et al. 2018). Cette phénylalanine-boronate a pu être utilisée lors de la sélection d'un anticorps contre la glucamine (C C Liu et al. 2009) et pour le design rationnel d'une lectine à partir d'une anticaline humaine (Lcn2) (Edwardraja et al. 2017). Les anticalines sont des petites protéines qui possèdent une région de boucle hypervariable. Des anticalines ont donc pu être générées contre différentes cibles thérapeutiques telles que le CTLA-4, ou le VEGF (Gebauer and Skerra 2012; Schönfeld et al. 2009; Skerra 2007). Dans cette étude, l'équipe de Skerra a montré qu'un environnement électrophile est important pour la fonctionnalité du boronate (Edwardraja et al. 2017). De plus, ils ont aussi cristallisé le boronoester, sous sa forme tétraédrique, formé par l'anticaline modifiée et son ligand (le 4-nitrocatechol) (Sommer, Eichinger, and Skerra 2019). La réactivité du boronate est favorisée dans des conditions basiques et conduit à la production d'ester boronique tétraédrique. Dans cette étude, l'ester tétraédrique est stabilisé par une liaison hydrogène avec un acide aminé polaire au niveau du site de reconnaissance. De plus, ils ont montré que la capacité de liaison de l'anticaline modifiée était dépendante de cette stabilisation. Ce travail montre

l'importance de fournir un environnement favorable à cette fonction. L'incorporation de cette fonction a été réalisée au sein d'un site de reconnaissance large et a permis à l'anticaline humaine Lcn2 d'acquérir des capacités de reconnaissance de sucres. Cependant l'affinité obtenue reste proche des affinités classiques pour des lectines, de l'ordre du μM .

IV.2.4 Interaction tryptophane - carbohydate

Le tryptophane est un acide aminé qui intervient dans trois types de liaison avec des hydrates de carbone : des interactions de type π -CH, des liaisons hydrogènes, et des contacts hydrophobes.

Au sein d'une étude réalisée par Tobola *et al.*, 4 types de tryptophane ayant un atome de fluor en substituant du cycle benzylique ont été incorporés au sein de la *Ralstonia solanacearum* lectine (RSL) à la place des 7 tryptophanes natifs présents dans chaque monomère. Parmi ces 7 tryptophanes, 6 sont à proximité du sucre ligand dont un formant des contacts hydrophobes avec un groupement méthyle du substrat ($\alpha\text{Me-Fuc}$, ou tetrasaccharide Le^x), un formant une liaison hydrogène et un dernier formant des liaisons stacking π -CH. Cependant, les dérivés de tryptophane ont été introduits avec une méthode résidu spécifique, s'accompagnant du remplacement de l'ensemble des résidus tryptophanes, ce qui conduit à une lecture globale sans pouvoir différencier l'effet indépendant de chaque mutation au sein des sites de reconnaissance. Une expérience de dynamique moléculaire basée sur les structures cristallisées des différents mutants a pu montrer que le fluor entraîne une baisse de l'aromaticité, ce qui baisse la force de l'interaction π -CH, mais le fluor a également pu former des liaisons hydrogènes avec les hydroxyles du sucre ligand. Cependant, l'addition de ces effets entraîne un effet global sur l'affinité assez limité (Tobola *et al.* 2018). Le développement d'aaRS permettant l'intégration de dérivés fluorés de Tryptophane de manière site spécifique pourrait permettre de modifier la reconnaissance de lectines.

Les acides aminés non naturels peuvent également être incorporés pour faciliter des études structurales, et des dérivés fluorés de tryptophane ont été particulièrement utilisés pour caractériser des enzymes y compris des galactosidases (Ring, Armitage, and Huber 1985; Luck and Falke 1991). Cependant ces protéines modifiées montraient aussi une activité réduite.

Chapitre 6 : Stratégies de modification des propriétés de reconnaissance de lectines et CBM à l'aide d'acides aminés non naturels

Objectif et Organisation de la modification des lectines

Les lectines ont généralement une forte spécificité malgré une affinité faible envers leur ligand (de l'ordre du mM ou μM). Les types d'interactions mis en jeu lors de cette reconnaissance sont essentiellement des liaisons hydrogènes, des interactions aromatiques - saccharides ainsi que d'autres liaisons faibles. Comme l'organisation spatiale des fonctions hydroxyles définit un sucre, les liaisons hydrogènes ainsi que les liaisons de coordination avec un métal portent donc l'essentiel de la spécificité de la reconnaissance. Au cours de ce travail de thèse, nous avons cherché à exploiter la capacité de ces fonctions hydroxyles à former d'autres types de liaisons plus fortes afin d'augmenter l'affinité de lectines. Pour cela deux stratégies ont été employées : i) l'incorporation d'une fonction aldéhyde protégée et ii) l'incorporation d'une fonction boronate.

I. Incorporation d'une fonction aldéhyde

Les fonctions hydroxyles peuvent réagir avec une fonction aldéhyde pour former des interactions covalentes de type acétal. De plus, les sucres peuvent également porter des fonctions amines qui peuvent aussi réagir avec une fonction aldéhyde pour former une fonction imine (Figure 94). Ces liaisons ainsi formées sont des liaisons covalentes qui, si elles existaient dans un site de reconnaissance d'une lectine, pourraient permettre d'augmenter l'affinité des lectines pour des ligands sucre.

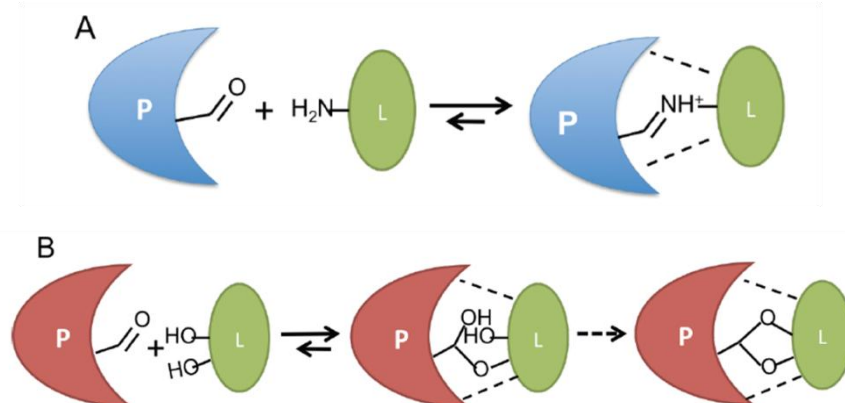


Figure 94: Schéma représentant la réaction entre une fonction aldéhyde incorporée au sein d'une protéine et une fonction amine (A) ou des fonctions alcools (B) présentes sur le substrat

Pendant, la fonction aldéhyde n'est pas présente parmi les acides aminés classiques et est très réactive. Naturellement, il existe des systèmes enzymatiques qui permettent de générer des fonctions aldéhydes mais ces modifications sont dépendantes de la présence d'une séquence signal (Appel and Bertozzi 2015). Or, ces séquences ne sont pas présentes dans le site de reconnaissance des lectines et les incorporer pourrait conduire à une perte de la reconnaissance due à une perturbation trop importante du site. Par contre, l'incorporation d'acides aminés non naturels portant une telle fonction par le détournement d'un codon stop permet son incorporation à la position choisie.

Il existe très peu d'acides aminés non naturels portant une fonction aldéhyde. Au début de la réalisation de cette thèse, dans la littérature, un seul acide aminé non naturel portant une fonction aldéhyde avait été incorporé dans une protéine (Tuley et al. 2014). Cette fonction aldéhyde était portée par un cycle aromatique, ce qui la rendait un peu moins réactive et moins électrophile par un effet de résonance. Elle n'était donc pas protégée pour éviter une réactivité trop précoce de la fonction aldéhyde avec les nombreuses fonctions alcools ou amines présentes dans les cellules. La stratégie initiale de notre projet a plutôt été d'introduire une fonction aldéhyde très réactive mais

protégée que l'on pourrait déprotéger en présence du ligand cible. Il a donc été imaginé un acide aminé non naturel portant une fonction photoactivable. Le travail développé afin de permettre l'incorporation de cet acide aminé non naturel sera développé dans le chapitre 6 suivant, partie I. Compte tenu du caractère ambitieux mais risqué de cette approche, nous avons développé une stratégie alternative.

Au cours de ma thèse, Wang *et al.* ont publié l'incorporation d'un acide aminé non naturel portant une fonction aldéhyde protégée par un groupement 4-azidobenzoyloxycarbonyl afin d'encoder génétiquement la diméthylation de lysines (Figure 95). Nous avons choisi d'exploiter cet acide aminé non naturel afin de générer une fonction aldéhyde au sein des lectines cibles. Pour cela la dernière réaction de transformation ne doit pas être réalisée (Figure 95). Le travail réalisé afin de générer ces lectines modifiées ainsi que leurs caractérisations seront présentées dans le chapitre 6 suivant, partie II.

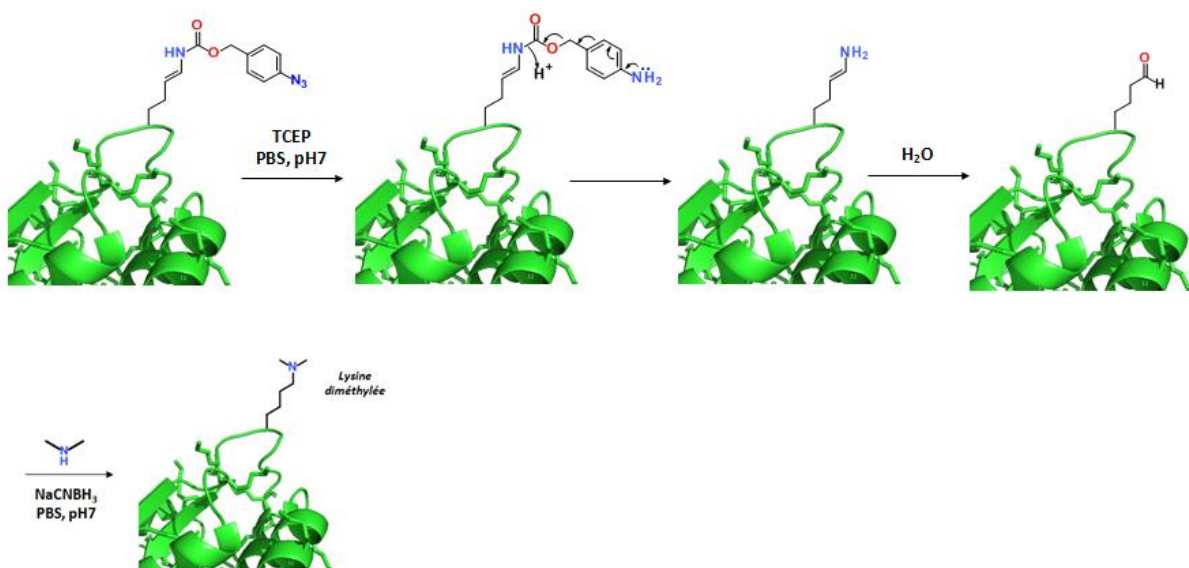


Figure 95: Modification de l'AcidK réalisée par Wang *et al.* afin de générer une diméthyl-lysine (Z. A. Wang *et al.* 2017). La réduction de l'azoture par le TCEP libère une fonction énamine par un effet de mésomérie. Cette fonction est ensuite spontanément hydrolysée dans un milieu aqueux, ce qui génère une fonction aldéhyde. Au cours de l'étude réalisée par Wang *et al.* une amination réductrice conduit à la production d'une lysine diméthylée. Au cours de ce projet de thèse, la dernière étape ne sera pas réalisée afin de conserver une fonction aldéhyde.

II. Incorporation d'une fonction boronate

Comme nous l'avons vu dans les sections I.3.1 et IV.2.3 du chapitre 5, la fonction boronate est capable de former des esters avec les fonctions hydroxyles des sucres. Bien que la formation de ces liaisons soit favorisée à un pH plutôt basique (aux alentours de 9), nous avons évalué la possibilité d'incorporer ces fonctions au sein de lectines cibles. Ce travail est présenté dans le chapitre 6 suivant, partie III.

I. Incorporation d'une fonction aldéhyde photoactivable

I.1 Synthèse d'un acide aminé possédant une fonction aldéhyde photoactivable

La synthèse d'un acide aminé non naturel portant une fonction aldéhyde masquée photoactivable a été imaginée et réalisée par le laboratoire CEISAM (partenaire du projet). La méthodologie de synthèse a été décrite dans le chapitre 7, section II.1. Un bilan de synthèse avec les meilleurs rendements pour chaque étape est présenté ci-dessous : l'acide aminé portant la fonction aldéhyde (composé 6) est d'abord généré selon les étapes indiquées Figure 96. Nous pouvons remarquer que les rendements obtenus sont alors très élevés. La chiralité des composés 4 et 5 a pu être vérifiée par HPLC chirale et est supérieure à 99%.

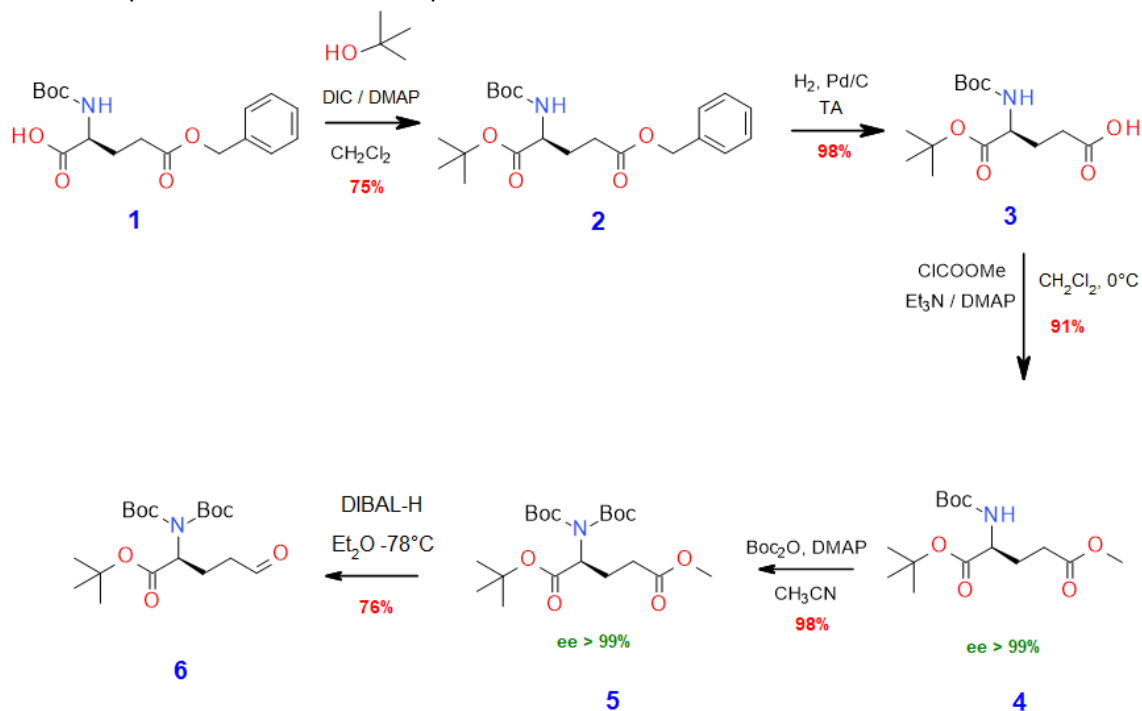


Figure 96: Bilan synthèse de l'aldéhyde intermédiaire

La fonction photolabile a ensuite pu être ajoutée à partir de deux composés diols de chiralité différente (Figure 97) ce qui a abouti à un mélange de composés syn et anti, avec la conformation syn majoritaire (composés 7 et 9, Figure 97). Ces composés syn ont ensuite été purifiés par HPLC et les groupements amine et acide carboxylique de la chaîne principale ont été déprotégés (Figure 98). Cependant, cette dernière étape de synthèse a abouti à la racémisation du produit au niveau du carbone α de l'acide aminé et donc à un mélange d'acides aminés de la série L et D. La mécanistique de cette racémisation n'a pas pu être expliquée. A priori, l'aminoacyl-ARNt synthétase devrait prendre en charge uniquement l'acide aminé issu de la série L (composés 11 et 13 Figure 98). Afin de faciliter la lecture des sections suivantes les mélanges produits ont été appelés 5S-7R-ONpGlu et 5R-7S-ONpGlu (Figure 98). Cinquante milligrammes du mélange 5S-7R-ONpGlu et cent milligrammes du mélange 5R-7S-ONpGlu ont pu être produits.

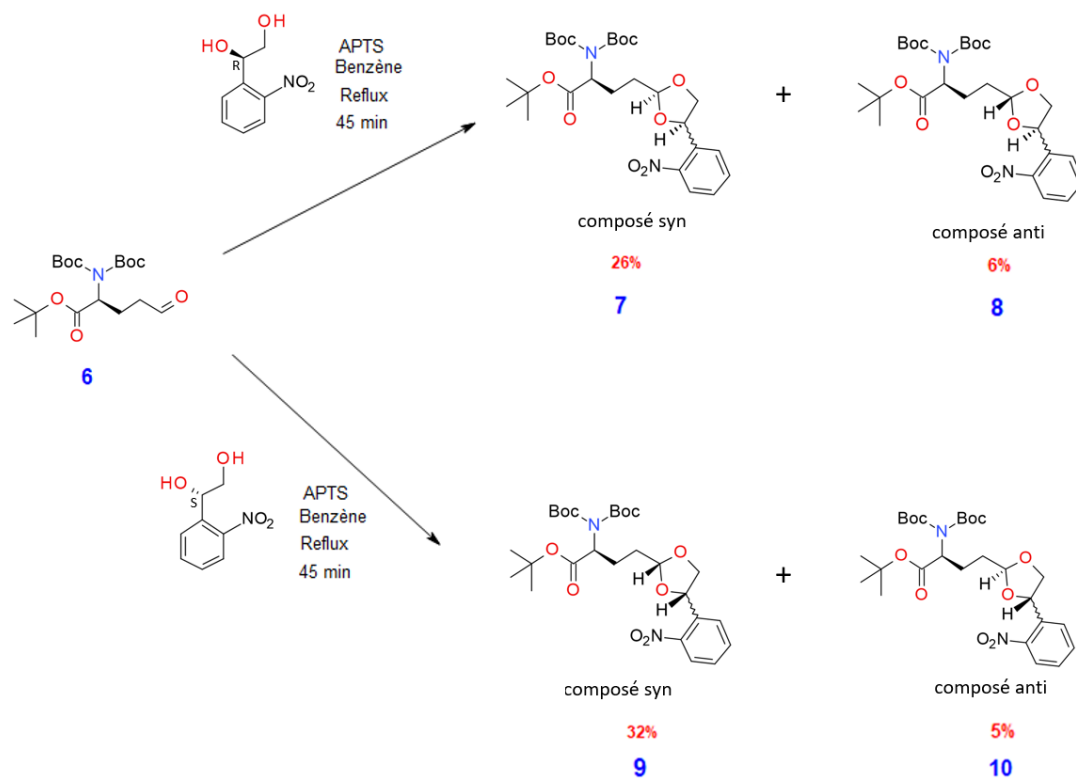


Figure 97 : Ajout du groupement photolabile sur la fonction aldéhyde. Deux configurations ont été testées et aboutissent à un mélange de composé anti et syn, le composé syn étant majoritaire.

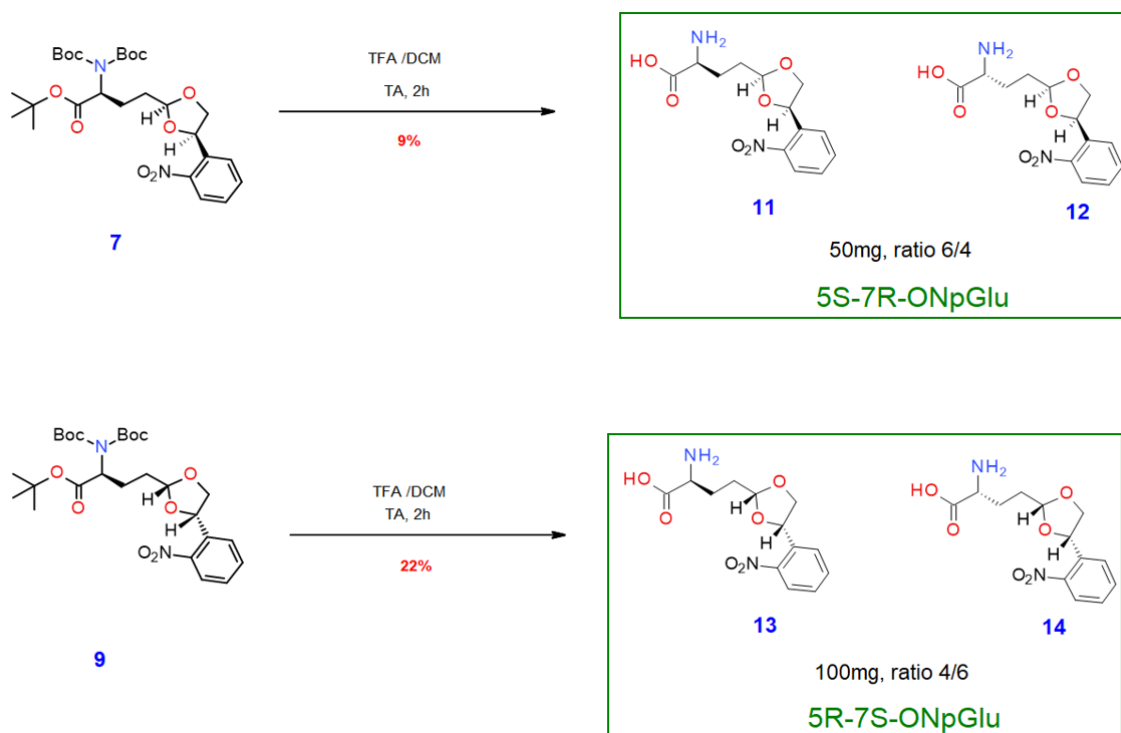


Figure 98 : Dernière étape de synthèse afin de déprotéger les groupements amines et carboxyliques de la chaîne principale de l'acide aminé.

La numérotation commence à partir de la fonction acide carboxylique de la chaîne principale de l'acide aminé, donc 5 s'applique au carbone de l'acétale et 7 le carbone benzylique

Le démasquage de la fonction aldéhyde par irradiation UV des composés 5S-7R-ONpGlu et 5R-7S-ONpGlu ont été vérifiés à partir des composés protégés 7 et 9 (Figure 98). Un suivi cinétique a

donc été réalisé par exposition aux composés à une lumière UV ($\lambda = 330\text{nm}$) et analyse du composé par RMN à différents temps d'exposition. Nous pouvons voir Figure 99 que ce démasquage est assez long et ne dépasse pas 80% après plus de 7h d'irradiation pour le composé 5S-7R-ONpGlu et 70% après plus de 7h de déprotection pour le composé 5R-7R-ONpGlu. Nous pouvons également remarquer que les deux composés ne possèdent pas le même profil de déprotection avec une déprotection plus rapide pour le composé 5R-7S-ONpGlu.

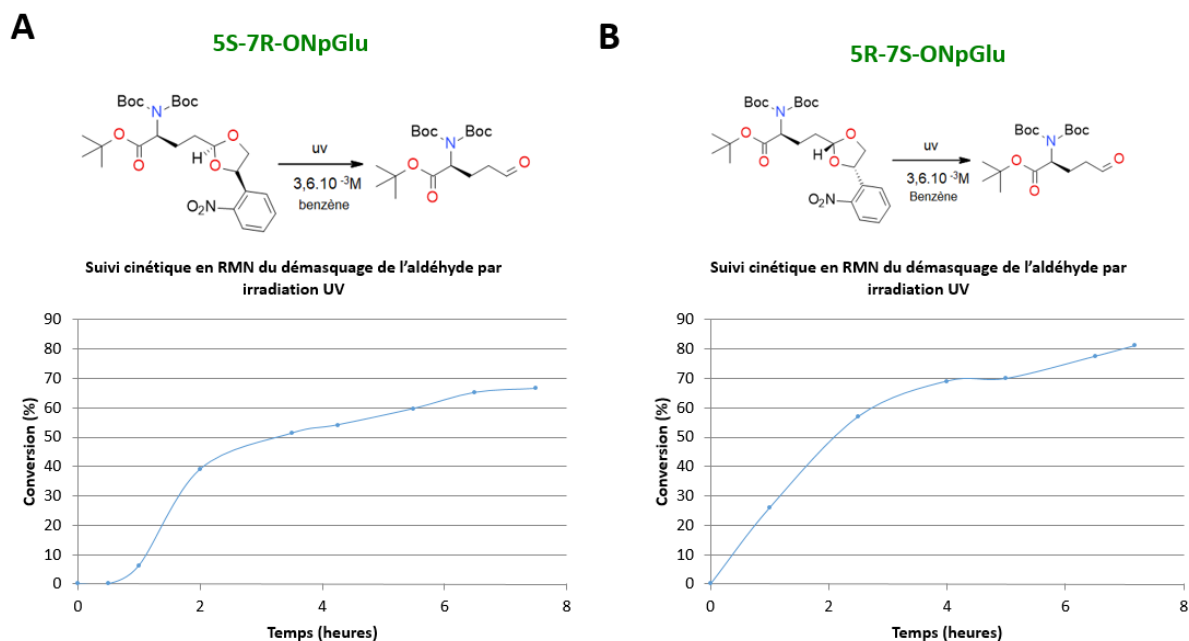


Figure 99: Démasquage de la fonction aldéhyde des composés 7 et 9 par irradiation UV (longueur d'onde 330nm). Les composés sont à une concentration de 3.6mM dans du Benzène et le suivi a été réalisé par une analyse RMN des composés à des points donnés.

1.2 Choix des protéines cibles

Afin de choisir différentes protéines cibles à modifier, plusieurs critères ont été considérés.

- La protéine se produit chez *E.coli*. Afin de limiter les risques liés à des difficultés d'expression des protéines cibles, nous avons choisi de ne considérer que les protéines dont l'expression chez *E.coli* était décrite dans la littérature.
- L'affinité de la protéine envers son substrat est plutôt faible (ordre du mM ou μM). Afin de détecter plus facilement l'effet de notre acide aminé non naturel sur l'affinité de la protéine, les protéines cibles envisagées ont des affinités plutôt faibles.
- La conformation du site de reconnaissance de la protéine permet d'incorporer l'acide aminé non naturel à une position qui, a priori, permet son incorporation et son interaction avec le substrat. La protéine ciblée doit donc avoir une structure 3D connue. Notre étude est rationnelle et l'incorporation de l'acide aminé non naturel sera ciblée. Une étude aléatoire du site d'incorporation entraînerait la multiplication des essais et donc une consommation plus importante de l'acide aminé non naturel.

Pour sélectionner des protéines cibles, la base de données PDB a donc été parcourue en sélectionnant uniquement les protéines ayant été produites chez *E.coli* et en cherchant des lectines ayant une structure compatible avec l'incorporation de l'ONpGlu. Afin de vérifier ce dernier point, le logiciel PyMol a été utilisé. Cependant, la structure de l'ONpGlu n'est pas connue au sein de ce dernier. Afin d'éviter une étude assez lourde sur l'encodage de cet acide aminé non naturel au sein de PyMol, nous avons décidé de considérer *in silico* des mutations Glutamate afin de pouvoir

appréhender l'interaction entre la fonction aldéhyde l'ONpGlu déprotégé et le ligand cible. En effet, une fois déprotégé l'acide aminé OnpGlu est très proche de l'acide aminé Glutamate, ne se différenciant que par la fonction aldéhyde au lieu d'une fonction acide carboxylique (Figure 100). Le site de reconnaissance considéré doit cependant être assez large pour permettre l'incorporation de l'ONpGlu qui est assez volumineux avant déprotection (Figure 100). Pour cela, le site de reconnaissance est inspecté visuellement dans PyMol puis, la fonction "Mutagenesis" est utilisée afin de visualiser une mutation glutamate au site d'incorporation envisagé.

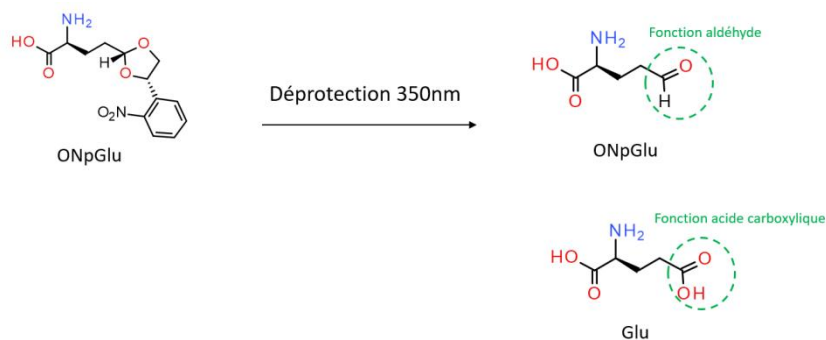


Figure 100: Comparaison des structures de l'ONpGlu et de l'acide glutamique

Trois protéines ont donc été sélectionnées selon ces critères : AAL-S2, CBM32 et CBM40.

1.2.1 Protéine cible n°1 : AAL-S2

La lectine issue de *Aleuria aurantia* est une protéine présentant une multivalence grâce à une répétition de polypeptide en tandem (Chapitre 5, section 1.1.2) et dont un fragment monovalent a été produit chez *E.coli* (AAL-S2) (Olausson et al. 2011). Cette lectine monovalente reconnaît le fucose avec une affinité de 6,6 μ M (Olausson et al. 2011). Si l'incorporation de l'ONpGlu permet d'améliorer l'affinité sur un site de reconnaissance de la lectine, la mutation devrait pouvoir être transférée à chacun des 5 sites de reconnaissance d'AAL car ces sites de reconnaissance sont structurellement très proches. Le site d'incorporation choisi est le Glutamate 89 (en rouge sur la Figure 101). En effet, celui-ci interagit déjà avec les fonctions hydroxyles O3 et O4 du fucose et semble donc correctement positionné pour permettre une réaction de condensation entre l'aldéhyde de l'ONpGlu déprotégé et les mêmes fonctions hydroxyles du fucose (Figure 101). Cependant, nous ne connaissons pas la distance optimale entre ces fonctions et la fonction aldéhyde. Dans un premier temps, nous avons plutôt considéré qu'il faudrait éviter une gêne stérique entre la fonction aldéhyde et les fonctions hydroxyles, bien que leur condensation entraîne une baisse de cette distance et donc un rapprochement du ligand.

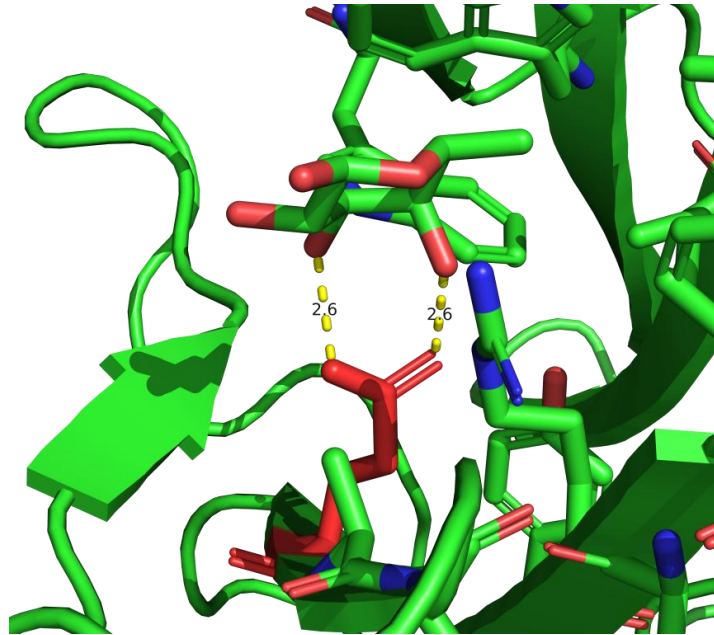


Figure 101: Site de reconnaissance de AAL-S2 (PDB: 1OFZ),

La distance entre les hydroxyles O3 et O4 du fucose et la fonction carboxylique de l'ONpGlu sur PyMol et est indiqué en pointillé jaune en angström.

Une expérience d'*alanine scanning in silico* a été réalisée afin de vérifier le potentiel impact d'une mutation à cette position. La fonction "*alanine scanning*" de l'outil FoldX du logiciel YASARA permet alors de rapidement tester l'impact de la mutation d'un résidu pour la structure de la protéine (Van Durme et al. 2011). Pour cela, le logiciel calcul la différence d'énergie pour chaque mutation. Cette étude ne prend cependant pas en compte la flexibilité ou le mouvement des chaînes latérales des résidus voisins, et comme suggéré dans la publication de Durme *et al.* l'étude peut être plus poussée par l'utilisation de la fonction "*mutate residue*" (Van Durme et al. 2011). Nous nous sommes cependant contentés pour cette analyse de la fonction "*alanine scanning*" qui fournit un score pour chaque mutation en kcal/mol. Plus la valeur de ce score est élevée, plus la mutation a un risque de modifier la structure de la protéine (Figure 102). Ainsi si un score sort de la norme par rapport aux valeurs calculées, cela signifie qu'il y a un risque que la mutation à cette position engendre des changements structuraux trop importants.

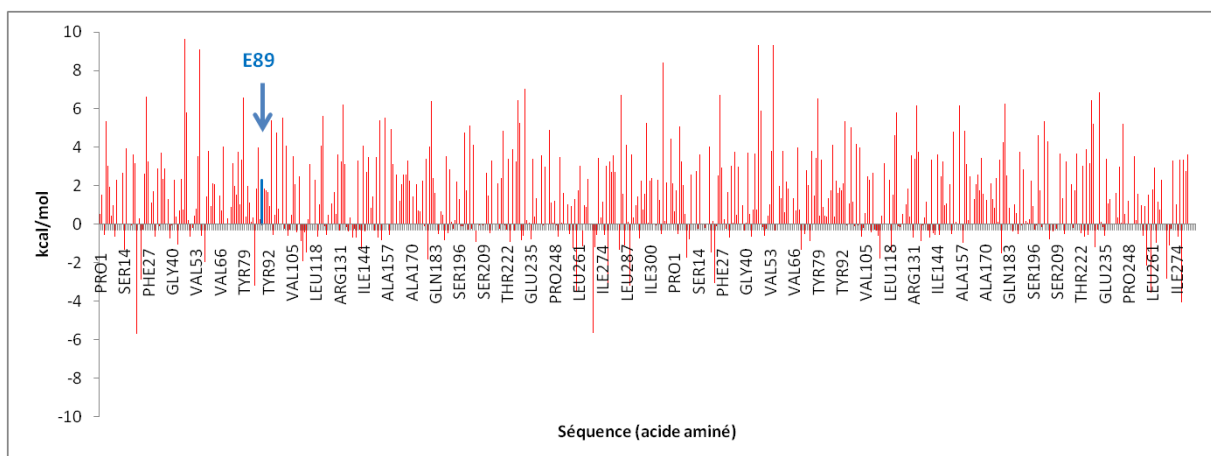


Figure 102: Extrait de résultat d'un alanine scanning sur la protéine AAL-S2 Analyse réalisée à l'aide du logiciel YASARA et du plugins FoldX.

Avec une valeur de 2,29kcal/mol, l'analyse de *l'alanine scanning in silico* n'a pas mis en évidence de problématique particulière à la position envisagée. En effet, cette valeur ne sort pas de la norme des valeurs calculées pour les différentes positions (Figure 102). La mutagenèse de la protéine AAL-S2 a donc été réalisée afin de générer un codon stop TAG à la place du Glutamate 89.

1.2.2 Protéine cible n°2 : CBM32-DD2

Le domaine DD2 du Carbohydre Binding Module 32 (CBM32), issu de la bactérie *Paenibacillus sp. IK-5* a été précédemment exprimé chez *E.coli* (Shinya et al. 2016). Ce domaine reconnaît le chitosane avec une affinité de 84µM (Shinya et al. 2016), et l'incorporation de l'ONpGlu à la position E14 semble permettre son interaction avec la fonction amine du chitosan (Figure 103 à gauche). De plus, il est décrit dans la littérature que la mutation Y36E permettrait une interaction entre le glutamate et la fonction amine (Shinya et al. 2016). En effet, l'homologue CBM32-DD1 possède naturellement cette mutation Y36E. Cette protéine permettrait donc de tester deux positions d'incorporation ce qui pourrait favoriser le succès de la stratégie employée. Une expérience d'*alanine scanning in silico* a donc également été réalisée afin de vérifier l'impact de la mutagenèse envisagée à ces positions ciblées. Cette rapide expérience n'a pas soulevé de problématique particulière obtenant des valeurs de 0,13 kcal/mol pour la position E14 et 1,26 kcal/mol pour la position Y36. La mutagenèse de la protéine a donc été réalisée afin de générer un codon stop TAG à la place du Glutamate 14 et de la Tyrosine 36. Un mutant Y36E a également été réalisé afin de pouvoir vérifier que la modification potentielle de l'affinité serait bien due à la présence de la fonction aldéhyde et non à la mutation de la Tyrosine. En effet, comme nous l'avons vu précédemment l'acide aminé ONpGlu déprotégé est très proche structurellement du glutamate (1.1).

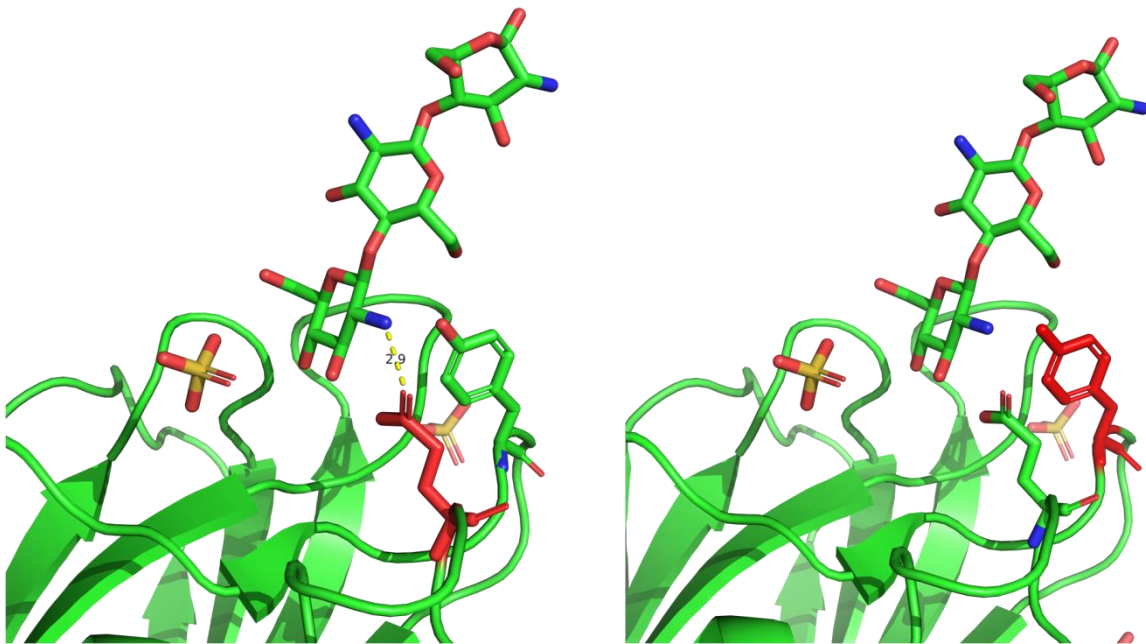


Figure 103: Site reconnaissance du CBM32-DD2 issu de *Paenibacillus sp. IK-5* (PDB 4ZZ8).

A gauche nous pouvons voir en rouge la position E14 en rouge, ainsi que la distance entre la fonction amine du chitosane et la fonction acide carboxylique de l'acide aminé E14 en angström. A droite, nous pouvons voir la position Y36 représentée en rouge. Images réalisées dans PyMol

1.2.3 Protéine cible n°3 : CBM40

Le Carbohydre Binding Module 40 (CBM40) issu de *S. pneumoniae* est une protéine qui est déjà utilisée au laboratoire dans le cadre d'un projet sur le développement d'inhibiteurs. Son expression chez *E.coli* est donc déjà maîtrisée par le laboratoire. De plus, le CBM40 reconnaît le sialyl-lactose

avec une affinité faible d'environ $2\mu\text{M}$, et le positionnement de l'ONpGlu au sein de son site actif à la place de la Phénylalanine 48 pourrait permettre l'interaction entre la fonction aldéhyde et les fonctions cis-diols de l'acide sialique (Figure 104).

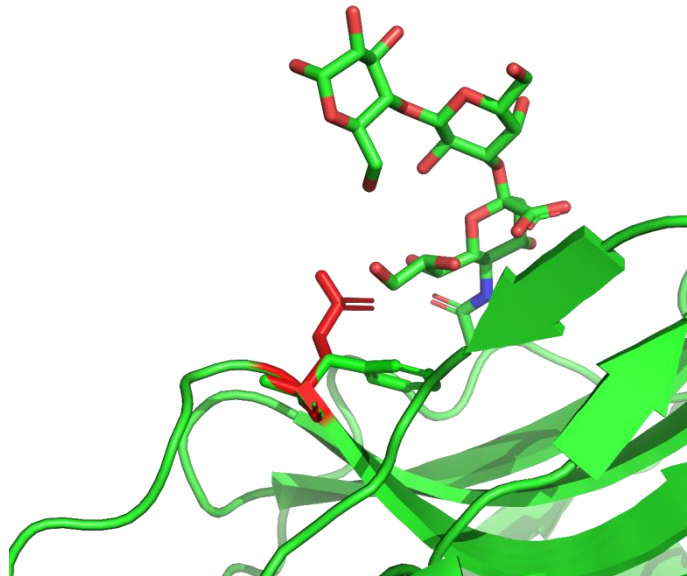


Figure 104: Site actif du CBM40 issu de NanA de *Streptococcus pneumoniae*, (PDB:4C1W)
La position F48 a été mutée en acide glutamique et a été représenté en rouge sur la figure. Image réalisée dans PyMOL

La position choisie nécessite une flexibilité de la protéine et du ligand par rapport à la configuration de la structure 3D cristallographique mais le risque est relativement faible, étant donné que contrairement aux autres protéines cibles, le gène est déjà disponible et l'expression de la protéine est maîtrisée au sein du laboratoire. L'expérience d'alanine scanning n'a d'ailleurs pas été réalisée pour cette protéine cible étant donné que ce risque était amoindri. La mutagenèse de la protéine CBM40 a donc été réalisée afin de générer un codon stop TAG à la place de la phénylalanine 48.

1.3 Mise au point d'une méthode de sélection

1.3.1 Test de pénétration de l'ONpGlu au sein des cellules BL21(DE3)

Le mode de transport d'un acide aminé non naturel original n'est pas connu et il est important de vérifier que celui-ci peut entrer dans la bactérie. Son entrée a donc été vérifiée par CCM grâce à la méthode décrite chapitre 7, section 1.4.1.

L'ONpGlu semble entrer au sein de la bactérie (migration G Figure 105). Cependant, on peut remarquer l'apparition d'un spot orange supplémentaire lorsque l'ONpGlu est mis au contact de bactérie (encadré en noir migration E, F, G Figure 105). Ce spot semble indiquer que l'ONpGlu est en partie modifié par les bactéries. Il pourrait s'agir d'une déprotection partielle. Un contrôle a été réalisé avec de la Propargyl-Lysine car nous savons que celle-ci rentre dans les bactéries. Nous ne pouvons cependant pas clairement voir de spot sur cette CCM (Migration D, Figure 105). Comparativement, l'ONpGlu semble donc rentrer de façon plus importante dans les bactéries.

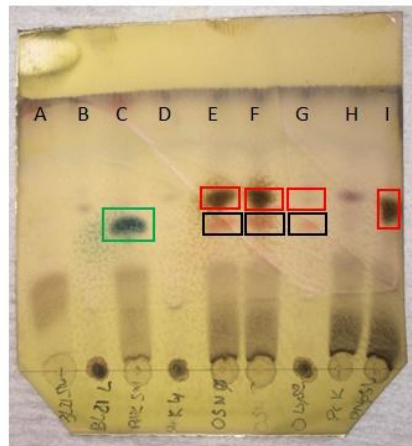


Figure 105: Vérification de l'entrée de l'ONpGlu au sein des bactéries.

A: Surnageant bactérien sans acide aminé, B: Lysat bactérien sans acide aminé, C: Surnageant bactérien avec PrK (cadrant vert), D: Lysat bactérien avec PrK, E: Surnageant bactérien avec ONpGlu (cadrant rouge) sans excitation pour mesure de fluorescence et contaminant visible (cadrant noir), F: Surnageant bactérien avec ONpGlu avec mesure de fluorescence au TECAN, G: Lysat bactérien avec ONpGlu, H: PrK 10mM, I: ONpGlu 10mM

Le composé ONpGlu semble rentrer dans la bactérie mais cette expérience a mis en lumière un problème de modification du substrat. A ce jour cependant, ce composé n'a pas été analysé pour déterminer son identité.

1.3.2 Pré-sélection des mutants de la synthétase par une méthode de bioinformatique

Afin de modifier les capacités de reconnaissance des lectines par l'incorporation de l'ONpGlu, la méthode d'incorporation site spécifique d'acide aminé non naturel (Chapitre 1, section IV.3) doit être utilisée. Pour cela, une aaRS doit permettre le transfert de l'acide aminé sur l'ARNt orthogonal. Cependant, comme il s'agit d'un acide aminé non naturel original, l'aaRS permettant ce transfert n'est pas connue. Une stratégie rationnelle a été envisagée et des mutations ont été prédites par une analyse de bioinformatique réalisée par Johann Hendrickx (ingénieur d'étude au sein du laboratoire) (Chapitre 7, section I.3). Cinq points de mutations avec des mutations préférentielles ont donc été prédits (Y384F, Y306F, N346A/G, C348A/G, W417C/N/A). Les différentes combinaisons des mutations à ces positions ont été classées à l'aide d'un score déterminé grâce au logiciel Rosetta Couples Moves. Plus le score est élevé, plus la probabilité que le mutant puisse prendre en charge l'acide aminé non naturel est grande. Cette expérience a été réalisée pour les deux stéréoisomères de l'ONpGlu disponibles. Les différents mutants sont présentés dans la Figure 106 et les mutants ayant les scores les plus élevés ont été construits par des expériences successives de mutagénèse dirigée afin de pouvoir être testés pour l'incorporation de l'ONpGlu.

séquence :	freq :	score :	min score :
FAGFC	145	1712.8+/-281.8	1255.82
FAGFN	104	1703.3+/-284.4	1246.08
FAGFA	100	1703.1+/-274.7	1250.54
FAAFC	83	1678.7+/-292.8	1250.02
FGGFC	76	1721.3+/-262.2	1268.87
YAGFC	75	1718.7+/-317.9	1254.1
FGGWC	69	1672.7+/-280.7	1275.68
FAGWC	62	1676.9+/-295.6	1242.09
FGGFN	60	1775.3+/-251.2	1301.67
YAGFN	59	1763.3+/-308.5	1635.32
FAGFM	59	1827.9+/-345.2	1266.89
FAGYC	58	1868.0+/-257.9	1844.66
FGGYM	58	1976.8+/-416.9	
FAAFA	57	1719.9+/-236.0	1256.58
FAGFS	57	1698.6+/-282.1	1255.08
YAGFA	55	1759.0+/-338.2	1262.6
YAGFM	54	1952.7+/-380.9	1623.07
FAAFN	51	1673.8+/-276.8	1259.46
AAAFC	50	1671.9+/-272.7	1257.53
FAAGW	49	1656.6+/-283.2	1268.79

5R-7S_ONpGlu

séquence :	freq :	score :	min score :
FGAAW	189	1744.0+/-250.7	1289.08
FAAAW	163	1764.1+/-233.2	1270.14
AAGLA	101	1684.7+/-198.0	1254.27
AAGLC	97	1710.5+/-223.8	1252.38
AAGLH	85	1750.2+/-200.7	1272.31
YAAAW	77	1924.9+/-363.0	1637.21
YGAAW	77	1892.1+/-348.7	1299.92
FGTAW	75	1645.3+/-276.3	1280.98
AAGAA	70	1722.6+/-277.9	1255.17
FGCAW	70	1936.3+/-364.7	1370.68
AAGAH	66	1811.2+/-294.1	1248.21
HGAAH	66	1687.8+/-263.7	1280.72
FATAW	64	1731.3+/-339.0	1276.13
AAGAC	62	1712.1+/-284.3	1253.8
YAGAA	62	1569.4+/-294.4	1258.43
FAGAA	60	1539.1+/-269.5	1245.42
FAGLH	60	1718.6+/-217.8	1254.89
AAAAH	59	1707.5+/-262.4	1256.4
HGAAW	58	1699.3+/-281.4	1292.61
YAGLH	55	1761.4+/-213.8	1260.38
YAGAH	53	1762.8+/-356.6	1259.85

5S-7R_ONpGlu

Figure 106: Présentation des différents mutants de l'aaRS issus de l'analyse bioinformatique. Ces mutants sont classés grâce à une fonction de score permettant de prédire quels sont les mutants les plus favorables à la prise en charge de l'ONpGlu.

1.3.3 Sélection expérimentale

1.3.3.1 Constructions génétiques pour la sélection

Le plasmide pEVOL originellement utilisé dans cette thèse pour l'incorporation d'acide aminé non naturel porte deux copies de l'aaRS. Cependant, lors de la mutagenèse dirigée permettant la génération des mutants la présence de deux copies est une complication supplémentaire. Nous avons alors enlevé la copie de l'aaRS inductible par le promoteur constitutif et n'avons pas observé de différence significative du taux d'incorporation de la Propargyl-Lysine (données non montrées). Ce plasmide pEVOL-1copie a donc été utilisé pour la génération et la sélection des différents mutants prédits par l'analyse bioinformatique (1.3.2). Ce plasmide est présenté en Annexe 2 et 3.

Comme nous l'avons vu précédemment (Chapitre 1, section IV.3.3.1), plusieurs méthodes de sélection existent pour sélectionner les mutants efficaces pour l'incorporation d'un nouvel acide aminé non naturel. Nous avons choisi de développer une méthode basée sur la production d'une protéine fluorescente. En effet, ce crible est moins strict qu'une méthode basée sur la survie bactérienne et il y a moins de phénomènes d'échappement génétique dus à des mutations auxiliaires. De plus, cela peut permettre une sélection quantitative par la comparaison de l'efficacité de différents mutants par analyse de l'intensité de fluorescence produite. Nous avons donc cloné le gène de la GFP en aval de la mPsaA (car nous savons que cette protéine se produit bien et accepte facilement les mutations avec des NAA) au sein du plasmide pET24d et ainsi généré le plasmide pET24d-mPsaA-GFP, puis nous avons généré un codon stop à la position K193 au sein de la mPsaA (plasmide pET24d-mPsaA-K193-GFP). Comme la sélection par une méthode basée sur la fluorescence est moins stricte, nous avons également généré un plasmide où deux codons stops TAG ont été

incorporés au sein de la mPsaA en position K105 et K193 (pET24d-mPsaA-K105-K193-GFP) pour pouvoir tester des conditions plus strictes.

Lors de l'utilisation de ces constructions plusieurs cas de figures peuvent se présenter et sont répertoriés dans le Tableau 25. La sélection sera effectuée en présence et en absence de l'acide aminé non naturel. Afin de sélectionner un mutant efficace et spécifique, une fluorescence doit être visible uniquement lorsque l'acide aminé non naturel est présent. En effet, si une fluorescence apparaît en l'absence de l'acide aminé non naturel, cela signifie qu'une incorporation erronée d'un acide aminé naturel a eu lieu au niveau du codon stop permettant la production de la protéine GFP fluorescente. Ce phénomène devrait cependant être assez rare, et la construction portant deux codons stops TAG permet d'augmenter la rigueur de la sélection.

Tableau 25: Résultats possibles lors de la sélection d'aaRS avec notre méthode de sélection

Mutants aaRS	Croissance sans NAA	Croissance en présence de NAA	Sélection	Caractéristiques
Mutant 1	GFP	GFP	Non	Mutant aspécifique
Mutant 2	Pas de fluorescence GFP	Pas de fluorescence GFP	Non	Mutant inefficace
Mutant 3	Pas de fluorescence GFP	GFP	Oui	Mutant sélectif et efficace
Etc ...	Pas de fluorescence GFP	Faible signal GFP	Oui	Mutant sélectif mais peu efficace

Avant de passer à des sélections des différents mutants pour l'incorporation de l'ONpGlu, les conditions de sélection sont tout d'abord testées. Pour cela, des expériences en milieu solide et liquide sont réalisées sur la synthétase sauvage pour l'incorporation de la Propargyl-Lysine, NAA dont nous maîtrisons déjà l'incorporation dans une protéine.

1.3.3.2 Validation de la méthode de sélection sur milieu solide

Lorsque des bibliothèques larges de mutants sont testées, il est important d'avoir un crible efficace. La sélection en milieu solide peut être une solution efficace. Nous avons donc testé différentes conditions pour l'incorporation de la Propargyl-Lysine. Les constructions pET24d-mPsaA et pET24d-mPsaA-GFP ont été utilisées en tant que contrôle négatif et positif de fluorescence, et les plasmides pET24d-mPsaA-K193-GFP et pET24d-mPsaA-K105-K193-GFP ont été utilisés en tant que test de l'efficacité de la sélection, avec comme modèle l'incorporation de la Propargyl-Lysine. Toutes ces constructions ont été respectivement co-transformées avec le plasmide pEVOL dans une souche BL21(DE3) puis utilisées lors des expériences de sélections modèles.

Nous n'avons pas trouvé de conditions permettant une sélection efficace sur milieu LB agar (données non montrées). Cependant, des boîtes de pétri ont été préparées à partir de milieu autoinductif, et la sélection a pu y être démontrée efficace (Figure 107). En effet, les spots bactériens correspondant à la culture BL21(DE3) cotransformée avec les plasmides pET24-mPsaA-K193-GFP + pEVOL sont fluorescents uniquement en présence de la Propargyl-Lysine (colonne 3 Figure 107B). De plus, la quantité de Propargyl-Lysine utilisée est assez limitée puisque nous l'avons ajoutée par adsorption de 300µL à 100mM, plutôt que directement dans le milieu AIM-agar, diminuant ainsi sa quantité d'environ 83% par rapport à si nous l'avions incorporée dans le gel avant de le couler. En effet, la quantité d'acide aminé non naturel est souvent un paramètre limitant à cause des difficultés

de synthèse qui peuvent être rencontrées. Cependant, la sélection plus stricte avec deux codons stops (pET24-mPsaA-K105-K193-GFP) n'a pas été efficace. Les conditions ne sont pas très propices à l'incorporation d'acide aminé au sein de plusieurs sites comme nous avons pu le tester dans l'autre partie de cette thèse (chapitre 3, section III) mais nous n'avons pas encore ces informations ni les outils nécessaires pour l'incorporation multiple au moment où nous avons réalisé ces tests.

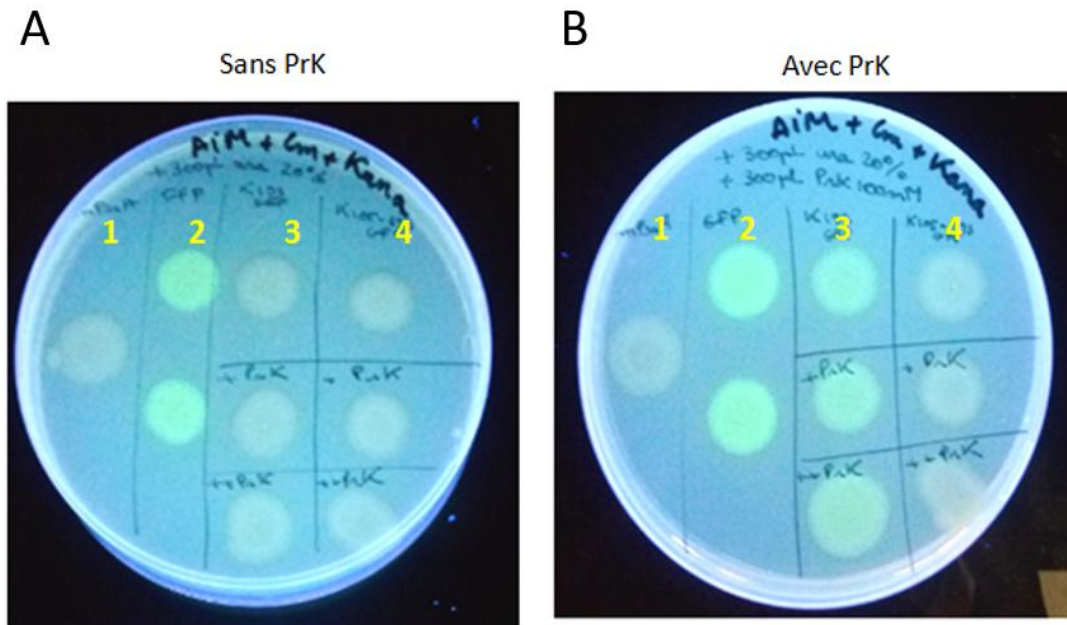


Figure 107: Sélection test sur milieu solide autoinductif.

Les cultures bactériennes des différentes souches: contrôle négatif (mPsaA, colonne 1), contrôles positif (mPsaA-GFP, colonne 2), et de test de sélection plus ou moins strictes (K193-GFP, et K105-K193-GFP, colonne 3 et 4 respectivement) ont été déposées sur boîte AIM+antibiotique (figure A ayant subi une absorption d'arabinose) ou sur une boîte AIM+antibiotiques ayant subi une absorption d'arabinose et de PrK (figure B).

1.3.3.3 Validation de la méthode de sélection sur milieu liquide

Nous avons donc réussi à déterminer des conditions pour réaliser la sélection en milieu solide. Cependant, malgré l'intérêt de cette technique pour cribler de très nombreux clones, la quantité d'acide aminé non naturel utilisé est toujours assez importante. De plus, cette sélection n'est pas quantitative. Nous avons donc cherché à développer une méthode de sélection en milieu liquide, toujours en utilisant la PrK comme NAA modèle, au sein de plaques 96 puits afin de baisser les volumes de culture. Le signal de fluorescence issu de la production de la protéine de fusion complète mPsa-GFP grâce à l'incorporation de la PrK est généralement plus faible car la production de la protéine contenant la PrK est moindre. Nous avons donc essayé d'augmenter les concentrations de NAA afin d'augmenter le signal par rapport au bruit de fond (Figure 108). Comme nous pouvons le voir sur la Figure 108, augmenter la concentration de PrK par 10 permet d'avoir un signal équivalent à la protéine sauvage. De plus, augmenter la concentration de la PrK par 2 suffit à avoir un signal environ 10 fois supérieur au bruit de fond.

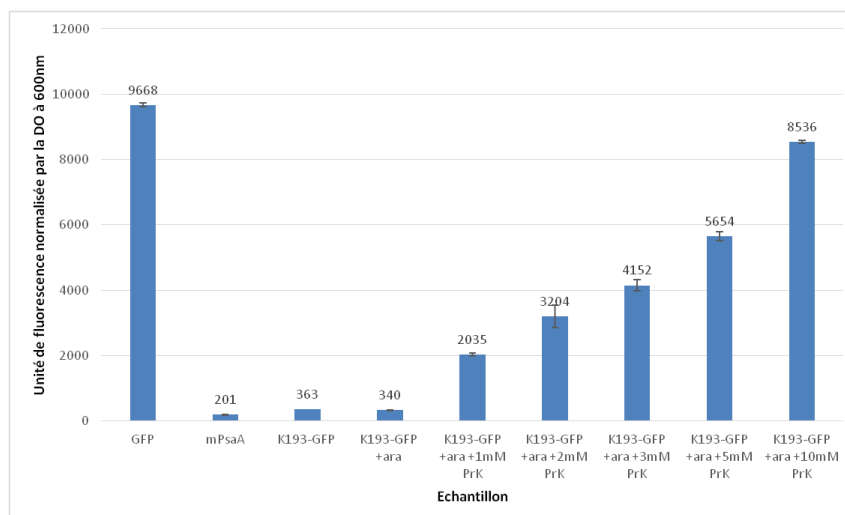


Figure 108: Test de différentes concentrations de la Propargyl-Lysine afin de déterminer des conditions de sélection en plaque 96 puits.

Les cultures GFP et mPsaA sont respectivement des contrôles positifs et négatifs. La culture de BL21(DE3) ayant été transformée avec les plasmides pEVOL-1copie et pET24d-K193-GFP (abrégé ici K193-GFP) ont été mis en cultures en milieu autoinductif avec des concentrations variables de PrK sur 24h afin de tester différentes conditions d'expression pour déterminer des conditions de sélection les plus efficace.

I.3.3.4 Test des différents mutants de l'aaRS afin d'incorporer l'ONpGlu

La mise au point de la méthode de sélection en milieu solide et en milieu liquide a été réalisée en parallèle de l'analyse bioinformatique. Le choix de la méthode employée dépend de la quantité de clones à tester. L'analyse bioinformatique a fourni une liste de mutants classés par une fonction de score qui reflète la stabilisation du ligand au sein du site actif et donc la probabilité que le mutant soit efficace pour l'incorporation de l'ONpGlu (section I.3.2). Cette analyse a été fournie pour les deux diastéréoisomères (5R-7S-ONpGlu et 5S-7R-ONpGlu) fournis par la synthèse chimique (section I.1). La quantité d'acide aminé fourni est assez faible (80 mg du mélange 5R-7S-ONpGlu et 30 mg du mélange 5S-7R-ONpGlu) à cause des difficultés de synthèse rencontrées (données non montrées). Nous avons donc choisi d'avoir une approche très rationnelle et nous avons testé dans un premier temps les 12 mutants ayant les scores les plus élevés pour le 5R-7S-ONpGlu et les 6 premiers mutants pour le 7R-5S-ONpGlu en milieu liquide grâce à des cultures en plaques 96 puits.

Ces sélections n'ont malheureusement pas permis de mettre en évidence un mutant capable d'incorporer l'ONpGlu. En effet, comme nous pouvons le voir Figure 109 le niveau de fluorescence obtenu à la suite de la culture en milieu contenant l'ONpGlu est négligeable et équivalente au bruit de fond malgré une concentration de 5mM de l'ONpGlu dans le milieu (en tenant compte de la pureté du mélange I.1). Le mutant FAGFN a un niveau de fluorescence légèrement supérieur aux niveaux obtenus avec les autres mutants mais une analyse complémentaire d'un extrait brut de production par WesternBlot n'a pas permis de détecter la présence de la protéine mPsaA-GFP ayant incorporé l'acide aminé non naturel en position K193 (données non montrées).

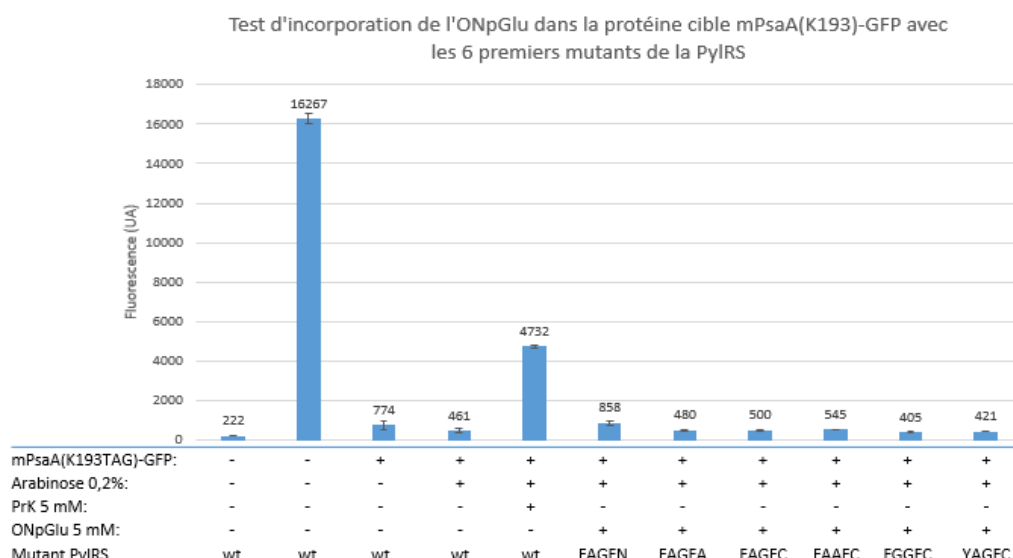


Figure 109: Extrait de données de sélections des différents mutants prédits par l'analyse bioinformatique pour l'incorporation de l'ONpGlu

Il est possible que notre stratégie de sélection ait été trop ciblée et que des mutations clés n'aient donc pas été testées. En effet, l'étude bioinformatique réalisée présente certaines limitations. Elle ne tient notamment pas compte de mutations trop éloignées du site ou allostériques, et ne considère que les mutations sens et exclue donc les mutations d'insertion et de délétion. L'analyse ne permet pas non plus de modéliser des étapes chimiques d'une réaction enzymatique ou la présence d'eau au niveau du site actif. Nous avons considéré que la position de la fonction carboxylate qui intervient dans la réaction enzymatique devait être conservée mais l'encodage de cette condition s'est réalisé grâce à des distances intermoléculaires au niveau du site actif qui sont donc dépendantes de la flexibilité de celui-ci. De plus, la grande flexibilité de l'ONpGlu a compliqué l'étude.

L'incorporation de l'ONpGlu (5S-7R), a pu également être testée au cours d'un Workshop réalisé en Oregon sur l'expansion génétique. Il a été testé également à l'aide d'une méthode de sélection par fluorescence en milieu liquide sur une centaine de synthétase (une plaque 96 puits). Cependant, aucun mutant n'a généré l'apparition de fluorescence spécifique à l'ONpGlu (données non montrées). Le choix des synthétases testées au cours de ce workshop n'a cependant pas pu être clairement identifié, si ce n'est que ce sont des synthétases mutantes généralement assez permissives habituellement utilisées par le laboratoire d'accueil du workshop pour tester l'incorporation de nouveaux acides aminés non naturels.

II. Incorporation d'une fonction aldéhyde protégée

Afin de poursuivre les travaux de modification des capacités de reconnaissance des lectines par l'utilisation d'un acide aminé non naturel, nous avons donc changé de stratégie (section suivante II). En effet, l'incorporation d'un acide aminé non naturel (le $N\epsilon$ -(4-azidobenzoxycarbonyl)- δ,ϵ -dehydrolysine, appelé AcdK) portant une fonction aldéhyde protégée a été décrite dans les travaux de Wang *et al.* (Z. A. Wang et al. 2017) après le début de ma thèse, indiquant notamment la nature des éléments orthogonaux (aaRS et ARNt) permettant cette incorporation.

Cet acide aminé non naturel, une fois déprotégé, est structurellement proche de l'ONpGlu déprotégé que nous souhaitons incorporer (Figure 110). En effet, après déprotection par action du réducteur TCEP suivi d'une hydrolyse spontanée, l'AcdK possède également une fonction aldéhyde avec une chaîne latérale alkyle constituée de 5 atomes de carbones au lieu de 4 pour l'ONpGlu.

Cependant, comme nous l'avons mentionné dans la section portant sur le choix des protéines cibles (section I.2), nous ne connaissons pas la distance idéale pour une réactivité optimale entre la fonction aldéhyde et les fonctions hydroxyles ou la fonction amine portées par le ligand. La possibilité d'étendre la chaîne latérale de l'ONpGlu lors de sa synthèse avait d'ailleurs été envisagée lors de la construction de ce projet de thèse. Néanmoins cette étape n'avait pas pu être réalisée par manque de temps. Nous avons donc considéré l'acide aminé non naturel AcdK comme une bonne alternative pour poursuivre les travaux de modification des lectines cibles. Nous avons d'ailleurs cherché à l'incorporer au sein des cibles présélectionnées pour l'ONpGlu (sections suivantes II.2).

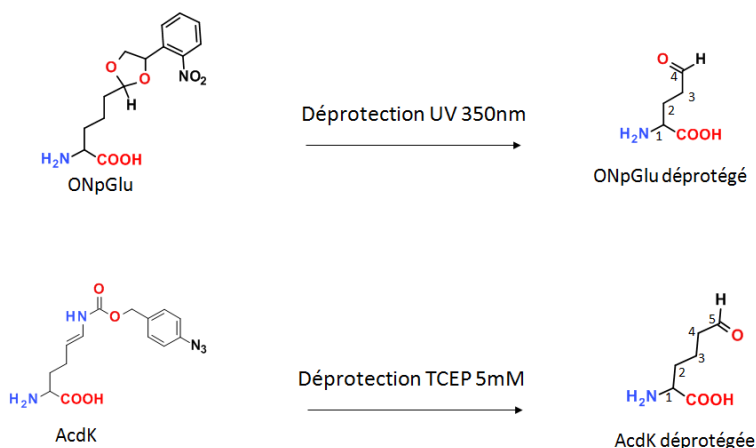


Figure 110: Comparaison des structures de l'ONpGlu et de l'AcdK avant et après déprotection par réduction grâce aux UV pour ONpGlu ou au TCEP pour AcdK.

II.1 Synthèse d'un acide aminé non naturel portant une fonction aldéhyde protégée

La synthèse de l'AcdK a donc été décrite au sein de la publication réalisée par Wang *et al.* (Z. A. Wang et al. 2017) et sa mise au point a été réalisée au sein du laboratoire du CEISAM par un post-doctorant (Antoine Richieu), puis a été montée en échelle au sein de l'UFIP par un stagiaire M1 (Denis Cadet) sous la supervision de Cyrille Grandjean. Environ deux grammes d'AcdK ont donc pu être produits et utilisés pour des tests d'incorporation. Le produit est pur à 80%, et nous avons considéré que l'aaRS ne devrait sélectionner que l'AcdK introduite au sein du milieu.

II.2 Incorporation de l'AcdK au sein des protéines cibles

II.2.1 Production des protéines sauvages et vérification de leur résistance à la lyophilisation et au traitement de déprotection

L'expression des différentes protéines sauvages a été testée. L'expression de AAL-S2 n'a cependant pas été concluante en milieu autoinductif et seulement un niveau très faible d'expression a été observé en milieu LB (données non montrées). Par contre l'expression des protéines sauvages et mutants Glutamate de CBM40 et CBM32 a été concluante en milieu autoinductif (données non montrées). Le stockage des protéines à -80°C après lyophilisation ainsi que leur résistance à une incubation prolongée avec le TCEP ont été vérifiées par dichroïsme circulaire. En effet, le TCEP est utilisé à 5mM durant une incubation de deux heures à température ambiante afin de déprotéger la fonction aldéhyde. Cependant en tant qu'agent réducteur fort celui-ci pourrait perturber la structure protéique. Le dichroïsme circulaire est une méthode d'analyse rapide et non destructrice qui permet de vérifier l'intégrité des structures secondaires (feuillettes β et hélices α). En effet, celles-ci ont des profils asymétriques et donc absorbent la lumière polarisée de manière différentielle.

Le signal de dichroïsme circulaire reste stable avant et après traitement des protéines par le TCEP, les structures des protéines CBM40 et CBM32-DD2 ne semblent pas être altérées par ce traitement de déprotection ou par la lyophilisation (Figure 111 et Figure 112).

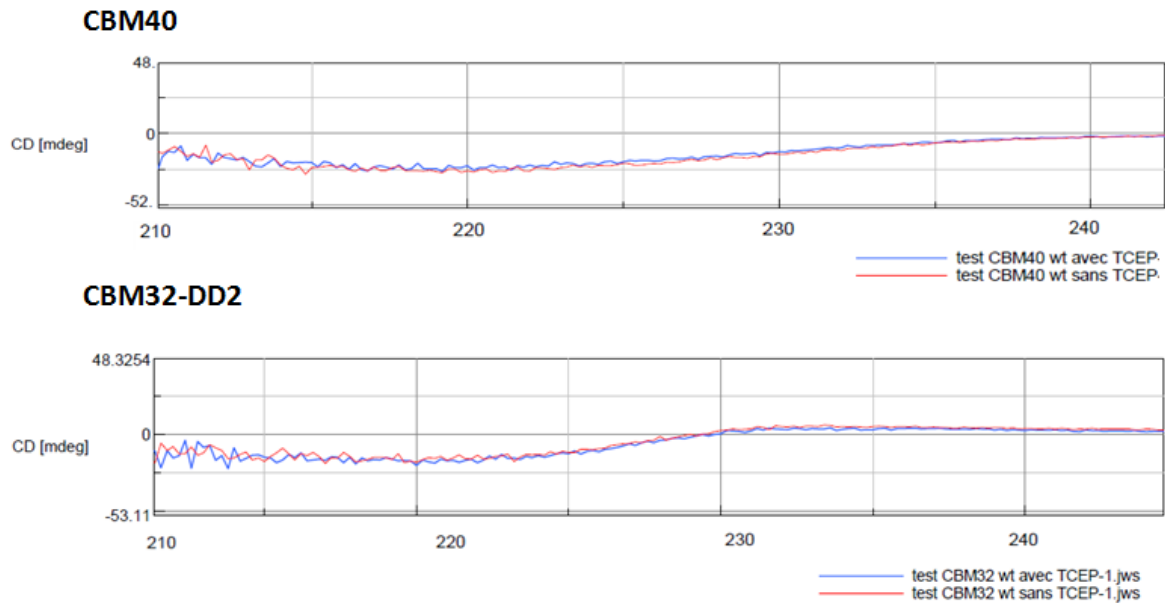


Figure 111: Profils de dichroïsme circulaire montrant la résistance des protéines CBM40 et CBM32(DD2) au traitement par le TCEP

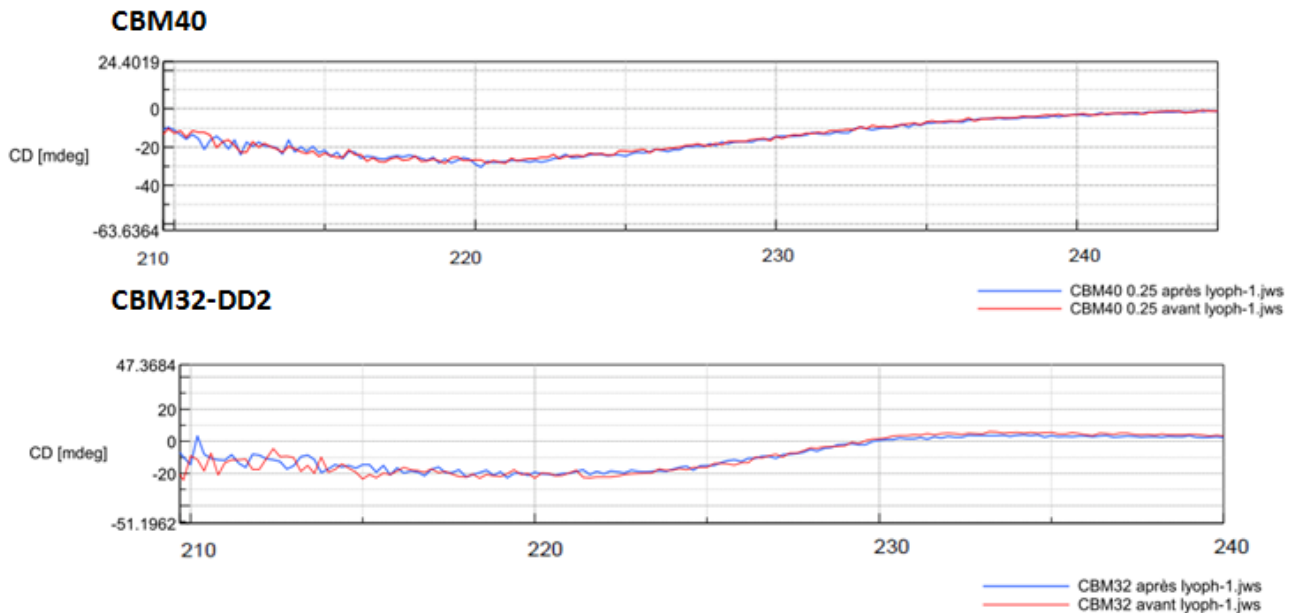


Figure 112: Profils de dichroïsme circulaire montrant la résistance des protéines CBM40 et CBM32(DD2) à la conservation par lyophilisation

II.2.2 Incorporation de l'AcidK au sein des protéines cibles et preuve de la déprotection

Afin d'incorporer l'AcidK, la synthétase de *M. mazei* doit porter les mutations L309T, C348G et Y384F qui permettent sa prise en charge. Ces mutations ont donc été réalisées par mutagenèse

dirigée sur le plasmide pEVOL-1copie. Ce plasmide a ensuite été co-transformé au sein de la souche d'expression BL21(DE3) avec l'un des plasmides portant les gènes des différentes protéines cibles ayant des mutations TAG au niveau du site d'incorporation choisi.

L'expression des différents mutants a donc été réalisée dans les conditions décrites dans la section II.3 (Chapitre 7). Nous avons noté que l'ajout de l'AcidK dans le milieu engendrait un précipité, et après quelques tests rapides de dissolution celui-ci semble précipiter à pH neutre (données non montrées). Cependant, après échange avec les auteurs de la publication de Wang *et al.* (Z. A. Wang et al. 2017), il s'avère qu'un ajout goutte à goutte sous agitation à suffit à réduire ce précipité et l'acide aminé non naturel semble être correctement dissout. Cependant, les concentrations employées ne dépassent pas 0,5 mM. En effet, différentes concentrations d'AcidK ont été testées pour l'incorporation de l'AcidK au sein de la protéine K193-GFP mais l'utilisation de l'AcidK à 1 mM ne permet pas un meilleur taux d'incorporation (Figure 113). La production de la protéine ayant incorporée l'AcidK reste toujours très inférieure à la protéine sauvage (Figure 113). Cependant, comparativement à l'incorporation de la PrK au sein de la protéine à la même concentration (0,5mM), l'efficacité d'incorporation semble similaire (Figure 113).

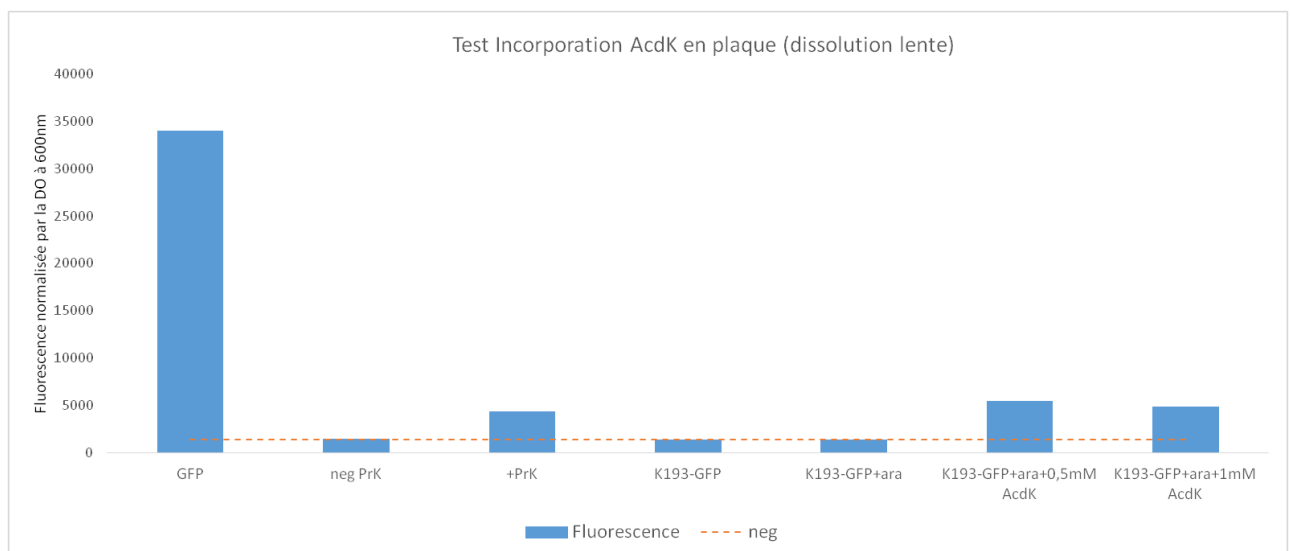


Figure 113: Test de conditions pour l'incorporation de l'AcidK

L'incorporation a été testée dans la protéine mPsaA-K193-GFP portée par le plasmide d'expression cotransformé avec le plasmide pEVOL-AcidK et aux contrôles (protéine mPsaA-GFP sauvage (GFP), le double transformant cultivé en absence (neg PrK) ou en présence (+PrK) de 0,5mM de PrK.

L'incorporation effective de l'AcidK au sein des protéines CBM40-F48TAG, CBM32-DD2-E14TAG et CBM32-DD2-Y36TAG a été testée et vérifiée par une analyse en SDS-PAGE ou en WesternBlot. Comme nous pouvons le voir sur la Figure 114, la production de protéine complète spécifiquement en présence de l'AcidK indique une incorporation correcte de celle-ci au sein de nos protéines cibles.

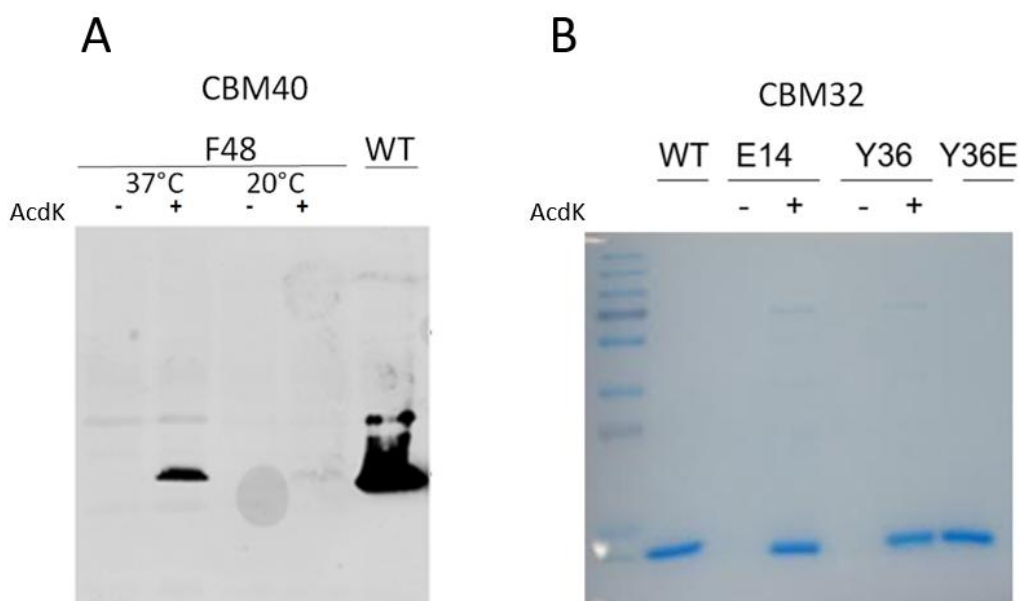


Figure 114: Vérification de l'incorporation de l'AcidK au sein des protéines (A) CBM40 par western blot et (B) CBM32-DD2 par SDS PAGE ;

L'incorporation de l'AcidK a été testée au niveau des positions F48 pour le CBM40, et E14 et Y36 pour le CBM32. Pour cela, les cultures ont été réalisées en milieu autoinductif auquel a été ajouté l'arabinose à une concentration de 0,2% final ainsi que l'AcidK goutte à goutte jusqu'à dissolution complète à une concentration de 0,5mM final. Les cultures ont été incubées 24h à 37°C ou 20°C (pour une des conditions du CBM40). Pour le CBM40 (figure A), 10µL de culture a été déposé sur gel qui a été analysé en Westernblot (anticorps anti étiquette histidine). Pour le CBM32 (figure B), les différentes cultures ont été purifiées sur billes de nickel afin d'être déposé sur gel d'acrylamide 12% qui a été coloré au bleu de Coomassie.

Ces protéines ont par la suite été purifiées sur colonne de nickel grâce à leur étiquette-histidine avant d'être utilisées pour des tests d'affinités (II.3).

II.3 Test d'affinité des lectines cibles

II.3.1 Biotinylation des ligands (sialyl-lactose pour CBM40 et chitosane pour CBM32)

Afin de pouvoir mesurer l'affinité entre les protéines modifiées avec AcidK et leurs ligands respectifs, plusieurs méthodes peuvent être utilisées au sein du laboratoire : du titrage calorimétrique isotherme (ITC), ou du Bio-layer interferometry (BLItz®). La méthode ITC est une méthode basée sur la mesure du dégagement de chaleur libéré lorsque le ligand se lie à la protéine réceptrice. Cette méthode est précise et permet d'accéder à des informations thermodynamiques de la liaison. Cependant, les affinités de nos protéines pour leurs ligands sont assez faibles et les conditions permettant de mesurer celles-ci par ITC n'ont donc pas pu être déterminées malgré plusieurs essais (données non montrées). La méthode BLItz® est une méthode qui permet de trouver rapidement des constantes d'affinités même si les valeurs obtenues sont moins précises que des mesures réalisées par ITC ou par Biacore. Elle est basée sur la mesure du déphasage de l'onde lumineuse ($\Delta\lambda$) issue de la nature de l'échantillon traversé par la lumière (en haut de la Figure 115). Lorsque la protéine sera en interaction avec le ligand fixé, il y aura alors un fort déphasage de la lumière. Au cours de la mesure, le ligand fixé à l'extrémité d'une fibre optique est mis au contact d'une solution contenant la protéine à tester afin de mesurer la phase d'association de celle-ci avec le ligand, puis la fibre optique avec le ligand fixé est ensuite transférée dans une solution ne

contenant plus cette protéine afin d'observer la phase de dissociation (bas de la Figure 115). Une constante d'affinité peut alors être calculée à partir de ces deux phases.

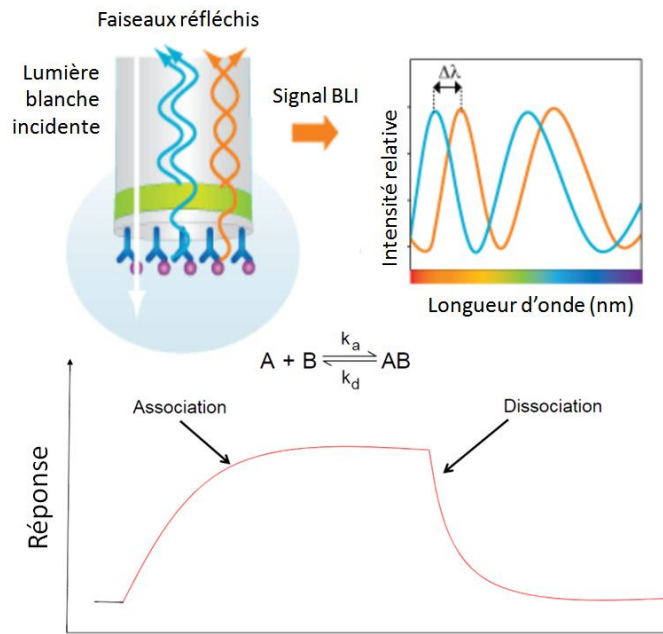


Figure 115: Principe de la mesure d'affinité par méthode Biolayer Light Interferometry (BLItz®)
Image reproduite à partir du site commercial de FortéBio

Afin de réaliser des mesures d'affinité par la méthode BLItz®, nous avons biotinylé les ligands afin de pouvoir les fixer sur des fibres optiques couvertes de streptavidines commerciales (méthode décrite chapitre 7, section II.3.8). Lors de la biotinylation des ligands, la biotine a été ajoutée en excès. En effet, les ligands non biotinylés ne pourront pas se lier à la streptavidine présente sur les fibres optiques et donc ne perturberont pas les mesures. Par contre, un excès de biotine pourrait réduire la fixation du ligand sur cette fibre optique et donc réduire la précision du signal. L'excès de biotine a donc été éliminé par le passage des échantillons sur des membranes en microfibrilles de Quartz lavées alternativement avec de l'acétonitrile ou de l'eau (chapitre 7, section II.3.8). La biotinylation des ligands a été vérifiée par chromatographie sur couche mince et la différence de migration pour les échantillons 3'SL et 6'SL avant et après biotinylation semble indiquer une correcte biotinylation (Figure 116A). De plus, pour ces échantillons biotinylés, il n'y a pas de biotine résiduelle détectée par CCM. Pour le chitosan (DP6, poids moléculaire d'environ 2 kDa), la migration sur CCM est moins nette, mais il semble y avoir une différence du profil de migration avant et après biotinylation. Cependant, deux purifications successives ont été nécessaires afin d'éliminer efficacement l'excès de biotine (Figure 116B).

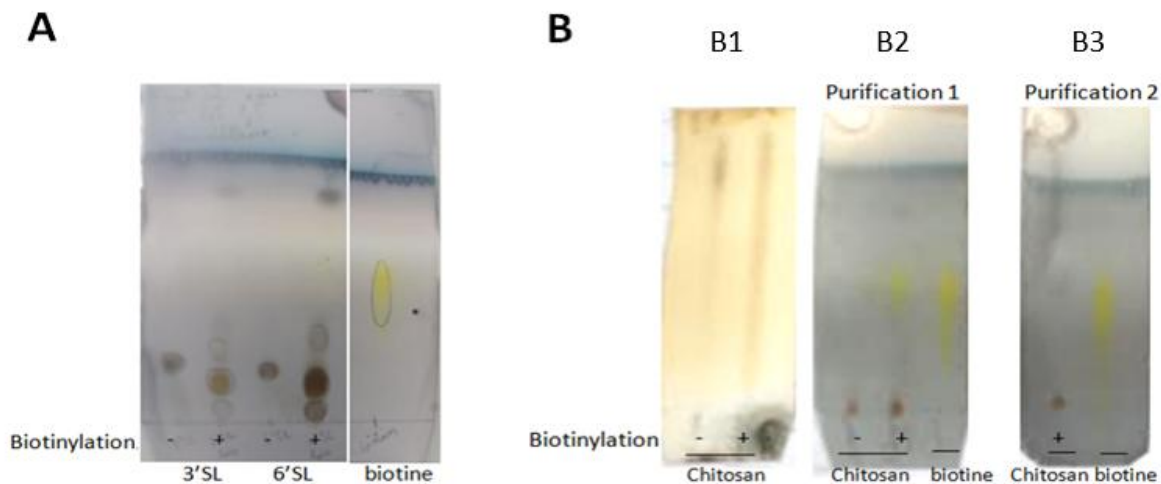


Figure 116: Vérification de la biotinylation des ligands (A) 3'Sialyl-lactose (3'SL), et 6'Sialyl-lactose (6'SL) et (B) du chitosan par chromatographie sur couche mince.

Les échantillons avant (-) et après biotinylation (+) ont été déposés sur CCM. Afin d'éliminer la biotine excédentaire (en jaune sur les CCM) des étapes de purification sur une membrane de microfibre de quartz a été réalisée. Les solvants de migration pour les CCM figure A et figure B2 et B3 sont acide acétique/butanol/eau en proportion (2/5/2) avec une révélation à la vanilline. Pour la figure B1 permettant l'analyse du chitosane biotinylé, les solvants de migration sont ammoniacque/acétate d'éthyle/eau en proportion (5/5/2) suivi d'une révélation à l'orcinol.

II.3.2 Mesures d'affinité

Les affinités des différentes protéines avec leurs ligands respectifs ont donc été mesurées à l'aide du BLItz® selon la méthode décrite chapitre 7, section II.3.8. Dans un premier temps, l'affinité des protéines sauvages CBM40 et CBM32-DD2 ont été mesurées pour plusieurs concentrations protéiques afin d'obtenir des constantes d'affinités fiables (données non montrées). Cependant, lors du test des différents mutants, la quantité de protéine ayant incorporé l'AcdK étant très limitée, nous n'avons pu tester qu'une seule concentration protéique. Les constantes de dissociation ne sont alors pas fiables avec un écart type calculé de l'ordre de la valeur du Kd calculé (données non montrées). Nous discuterons donc seulement des courbes des différents mutants entre eux, sachant qu'elles ont été mesurées à une concentration protéique donnée identique pour chaque mutant.

II.3.2.1 Mesures d'affinité : cible CBM32-DD2

Les protéines CBM32-DD2 ayant incorporé l'AcdK aux positions E14 ou Y36 ont été produites à une concentration de 0,15 mg/mL soit 5,86 μ M. Les courbes d'associations et de dissociation des protéines ont donc été mesurées à 5,86 μ M pour tous les mutants. Les échantillons sont purs (Figure 114) et les concentrations protéiques ont été mesurées par absorption à 280 nm avec un Nanodrop. Afin de limiter les interactions non spécifiques au cours des mesures, de la BSA (5%) a été ajoutée au sein des solutions protéiques et a permis d'obtenir des courbes plus nettes.

Nous pouvons voir sur la Figure 117, les affinités des différents mutants avant et après traitement par le réducteur TCEP qui permet de déprotéger la fonction aldéhyde de l'AcdK. Le mutant Y36E montre la plus forte affinité parmi les différentes formes de la protéine testées et ceci pourrait s'expliquer par une possible formation de liaisons hydrogènes supplémentaires entre le glutamate en position 36 et le ligand en plus des liaisons hydrogènes formées avec le glutamate en position 14. Les mutants AcdK (en position E14 et Y36), montrent une affinité plus faible que la protéine sauvage avant traitement TCEP, ce qui peut s'expliquer par l'encombrement stérique du site de liaison par le groupement protecteur. Cependant, le mutant E14AcdK ne montre pas de modification d'affinité après déprotection, il est possible que cette position ne permette pas

d'interaction avec le ligand. Par contre, après déprotection du mutant Y36AcDK, on observe une augmentation de l'affinité par rapport au mutant Y36 avant déprotection. Celle-ci reste cependant plus faible que pour la protéine sauvage ou pour la protéine Y36E après traitement au TCEP. Le profil de dissociation semble cependant similaire aux profils observés dans les mêmes conditions pour la protéine sauvage ou pour la protéine Y36E, ce qui ne va pas dans le sens d'une interaction entre la fonction aldéhyde et la fonction amine du substrat. En effet, la formation d'une liaison covalente devrait entraîner le ralentissement de la phase de dissociation, voire sa disparition. Nous pouvons supposer que le traitement de la protéine Y36AcDK par le TCEP a éliminé la présence du groupement protecteur en position 36 et permis ainsi de restaurer partiellement l'affinité de la protéine grâce à l'interaction entre le glutamate 14 et le ligand. Ces hypothèses ne sont cependant pas confirmées.

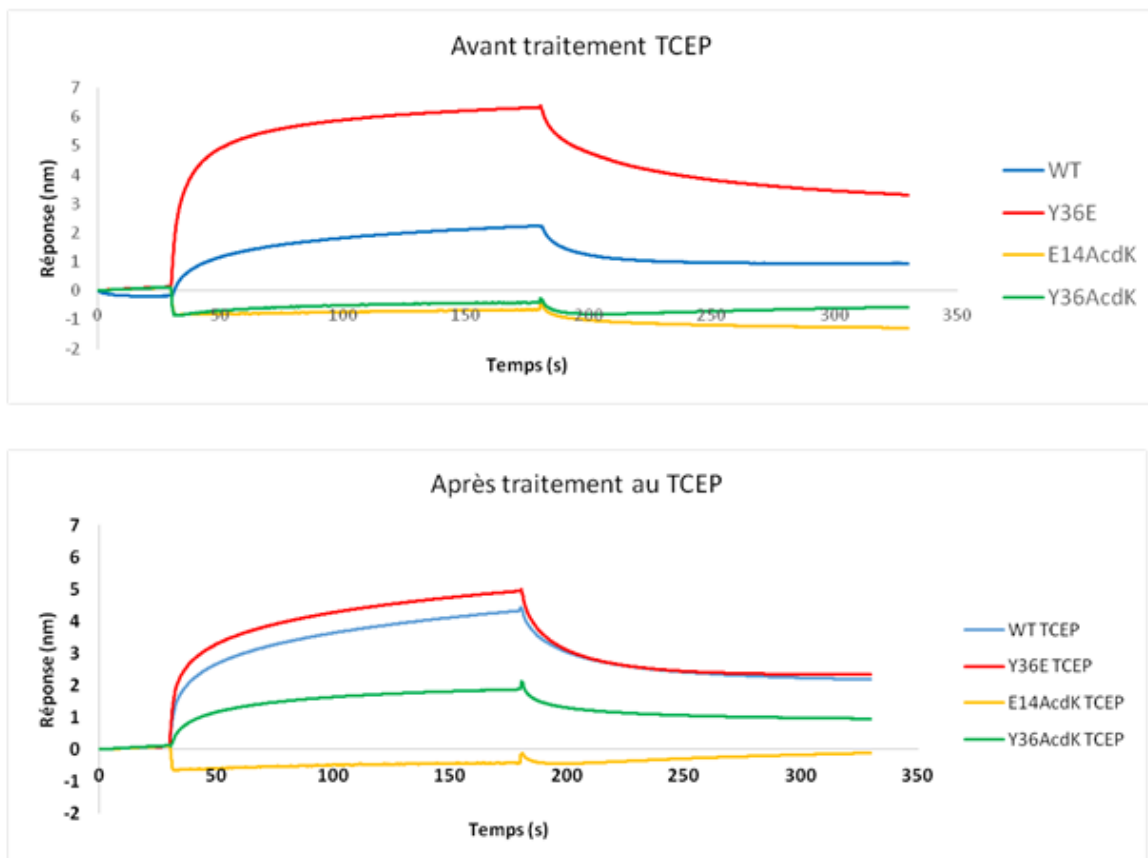


Figure 117: Courbes d'association et de dissociation des différents mutants CBM32-DD2 avant et après traitement TCEP

Le traitement TCEP n'a cependant pas eu un effet neutre sur la mesure de l'affinité des protéines. Celui-ci entraîne une forte baisse du pH mais cette baisse a ensuite été neutralisée par l'ajout de KOH 5M. Nous pouvons voir l'effet du traitement TCEP sur l'affinité des différents mutants testés (Figure 118). Pour la protéine Y36E, le traitement TCEP a abaissé l'intensité du signal, cela peut être dû à une perte de la structure de la protéine, malgré l'ajout de KOH et la vérification de l'intégrité de la protéine par dichroïsme circulaire. Pour la protéine sauvage (WT), il est surprenant que le traitement TCEP ait un effet positif sur la courbe observée. Cependant, la courbe d'association/dissociation de la protéine sauvage (WT) n'ayant pas subi le traitement TCEP a été répétée à cause d'un problème de signal observé au cours de la première expérience. Un standard interne (Y36E) avait été réalisé afin de normaliser le signal pour éviter toute déviation due à l'utilisation d'une nouvelle fibre optique couverte de streptavidine. Cependant, il est possible que les mesures ne soient pas suffisamment comparables malgré cette correction. Pour la protéine E14AcDK,

il n'y a pas de changement de profil d'association/dissociation et les mesures restent très faibles au niveau du bruit de fond de l'expérience. Par contre pour la protéine Y36AcidK, on observe bien une augmentation du profil de l'affinité, ce qui peut indiquer la correcte déprotection de l'AcidK.

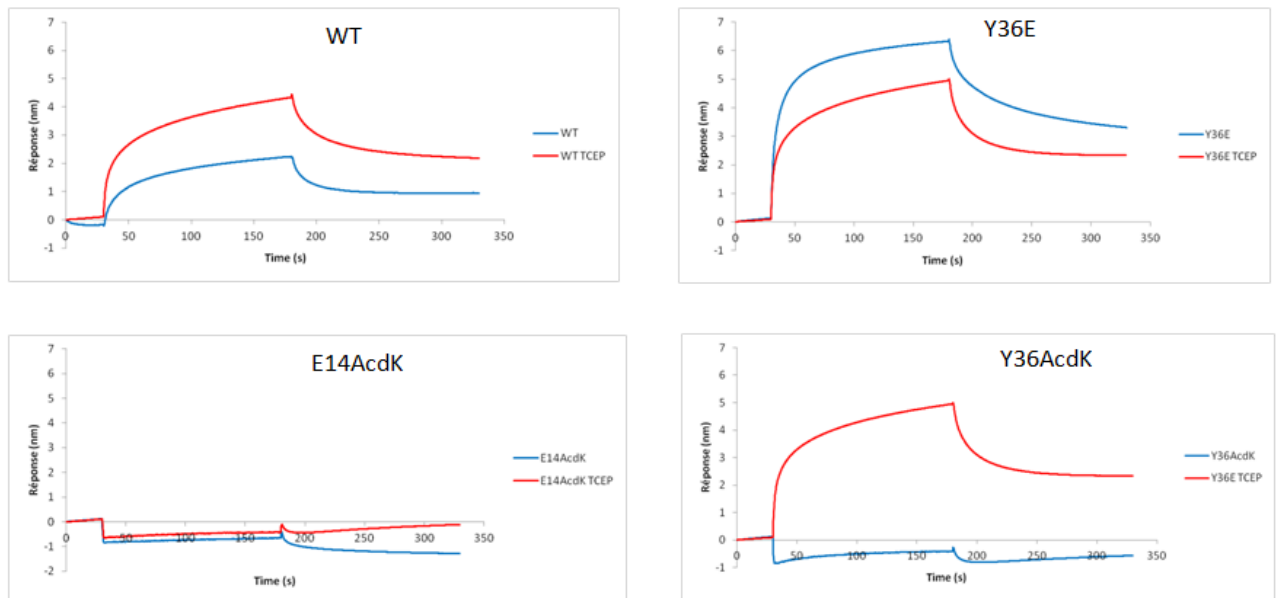


Figure 118: Effet du traitement TCEP sur les courbes de dissociation des différents mutants étudiés

Les mesures d'affinités sont imprécises, à cause de l'utilisation d'une seule concentration protéique. Cependant, les courbes d'association et de dissociation ont été réalisées à une concentration identique de protéine pour tous les mutants testés et cette expérience nous donne donc des informations qualitatives sur leurs affinités. Le mutant Y36E a montré une affinité globalement plus haute que la protéine sauvage, ce qui peut s'expliquer grâce à une interaction supplémentaire, qui est observée chez un variant naturel au niveau de cette position. La mutation E14AcidK n'a pas montré de profil d'interaction avec le ligand, il semble donc que cette mutation ait eu un effet délétère sur l'affinité de la protéine. Par contre, la mutation Y36AcidK a montré un profil différentiel avant et après déprotection, validant la déprotection de NAA. Cependant, son affinité ne semble pas supérieure à la protéine sauvage (WT) ou à la protéine CBM32-DD2-Y36E.

II.3.2.1 Mesures d'affinité : cible CBM40

La protéine CBM40 ayant incorporé l'AcidK en position F48 a été produite à une concentration de 0,14mg/mL. Les échantillons présentent cependant une contamination protéique à faible poids moléculaire qui se co-purifie sur colonne de nickel (Figure 119A). Malgré des étapes de dialyses et de concentrations des protéines avec des membranes ayant des limites d'exclusion de poids moléculaire proche de celui de la protéine à éliminer, la présence du contaminant a été atténuée dans l'échantillon F48AcidK mais reste présente (Figure 119B). Les concentrations des mutants F48E et F48AcidK ont donc été normalisées par rapport à la concentration de la protéine sauvage à l'aide d'un WesternBlot semi-quantitatif (Figure 119C et D). En effet, la protéine sauvage a été produite à une concentration très élevée (3,2mg/mL) et très pure et a été déposée sur gel selon une gamme de concentration puis analysée sur un logiciel permettant la mesure d'intensité (Image Studio Lite Version 5.2). Pour la protéine F48AcidK, l'intensité de la bande observée était dans la gamme de concentration ce qui a donc permis l'estimation directe de sa concentration à 0,0767mg/mL, soit 3 μ M. Pour la protéine F48E, l'intensité de la bande n'était pas dans la gamme. J'ai donc considéré un ratio entre la protéine CBM40-F48E et la protéine contaminante (ratio calculé à

91,9%), et corrigé la concentration de cette protéine à 0,57mg/mL (par rapport à 0,63 mg/mL initialement). Cette correction implique cependant une approximation qui estime que la constante d'absorption de la protéine contaminante est équivalente à celle du CBM40.

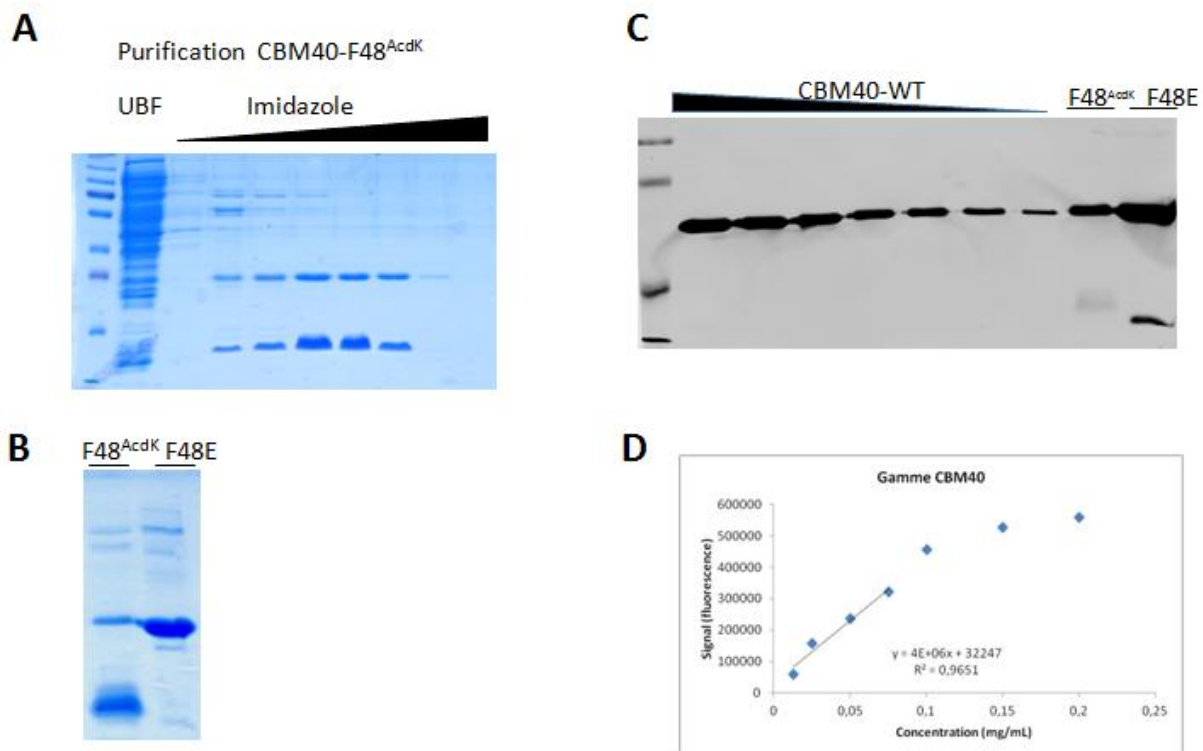


Figure 119: Stratégie de correction de la concentration protéique des mutants CBM40 comparativement à la protéine sauvage

Les différentes fractions de purification de la protéine CBM40-F48AcidK ont été analysées en SDS-PAGE (figure A). Nous pouvons voir que la purification présente une bande contaminante. Après utilisation d'une membrane d'exclusion, les échantillons F48AcidK et F48E ont été déposés sur gel d'acrylamide 12% et analysé en bleu de Coomassie. Finalement une gamme de concentration de la protéine CBM40 a été déposée sur gel acrylamide 12% ainsi que les échantillons à analyser (CBM40-F48AcidK et CBM40-F48E) puis le gel a été analysé en WesternBlot (anticorps anti étiquette histidine) (figure C). L'intensité des bandes de la gamme de CBM40 a été mesurée à l'aide du logiciel Image Studio Lite Version 5.2, ce qui a permis de déterminer une courbe de dosage.

Le mutant F48AcidK ayant la concentration la plus faible, nous avons réalisé les expériences de BLItz® à 3µM pour tous les mutants testés. Les profils de liaison obtenus sont caractéristiques d'une liaison complexe composée probablement d'une part de liaison aspécifique. En effet, les phases d'association n'atteignent pas de plateau et semblent composées de deux temps avec une association rapide suivie d'une association lente Figure 120. Afin de limiter ces interactions non spécifiques au cours des mesures nous avons essayé d'ajouter de la BSA, du lait, ou du Tween au sein du tampon. Cependant, ceux-ci présentaient, lorsqu'ils étaient testés seuls, des profils d'interactions avec la sonde (données non montrées) et n'ont donc pas pu être utilisés pour les mesures d'affinités ci-dessous.

Nous pouvons voir sur la Figure 120, les courbes d'association suivies de dissociation des différents mutants de CBM40 avant et après traitement par le réducteur TCEP. La protéine sauvage présente l'affinité la plus forte. En effet, la mutation F48E semble avoir un effet délétère sur l'affinité observée. La phase de dissociation du mutant F48E ne présente pas de dissociation nette ce qui semble suggérer que la phase d'association pourrait correspondre à de l'association non spécifique. Quant au mutant F48AcidK, il présente une affinité faible comparativement à la protéine sauvage.

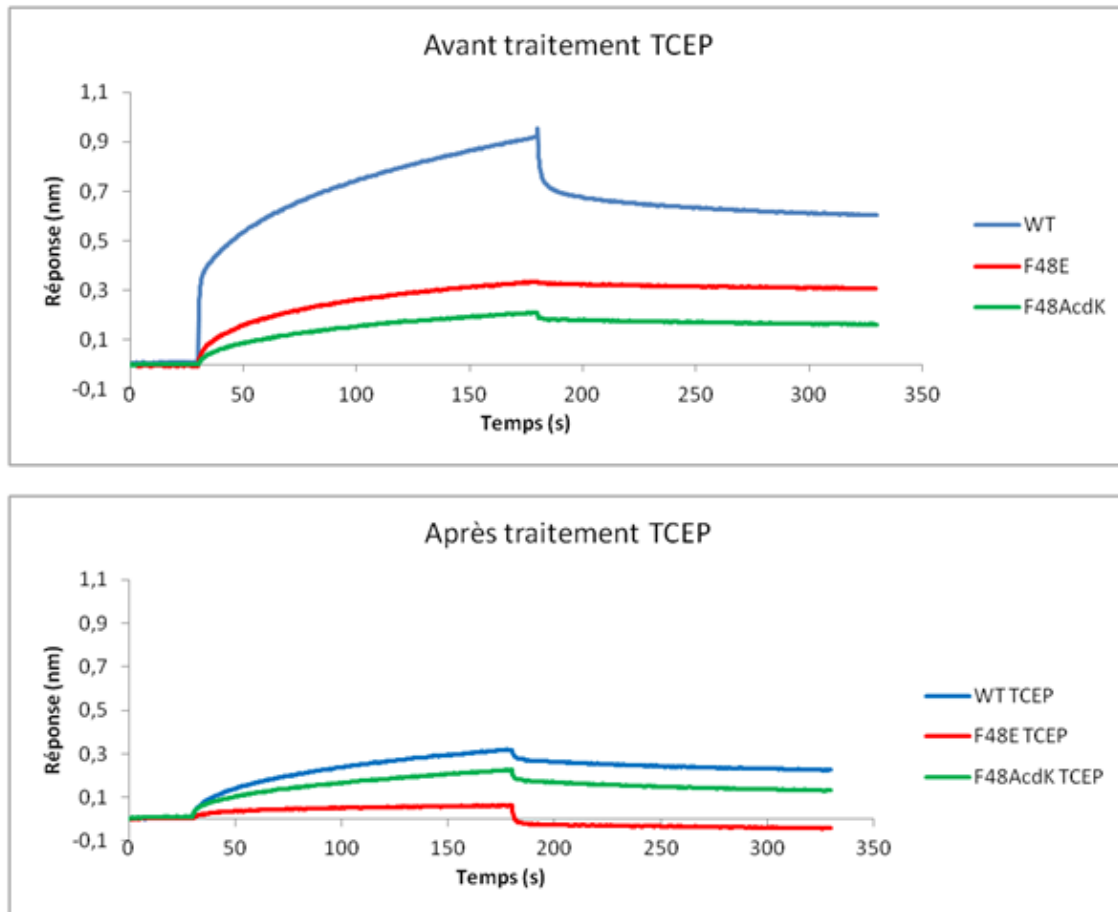


Figure 120: Profils de liaison des différents mutants de CBM40 avant et après traitement au TCEP

Le traitement TCEP n'a pas eu un effet neutre sur la mesure de l'affinité des protéines malgré l'ajout de KOH 5M pour neutraliser le pH de la solution. Nous pouvons voir l'effet du traitement au TCEP sur l'affinité des différents mutants testés (Figure 121). Les protéines sauvages (WT) et F48E présentent une forte baisse de l'intensité du signal. Ce qui pourrait s'expliquer par une partielle dénaturation de la protéine ou par une perturbation des conditions de lecture. Cependant, il n'y a pas de changement observé pour le mutant F48AcidK. Cela pourrait s'expliquer par une potentielle compensation entre un effet positif lié à la libération de la fonction aldéhyde et par une baisse de signal comme observé pour les protéines WT et F48E.

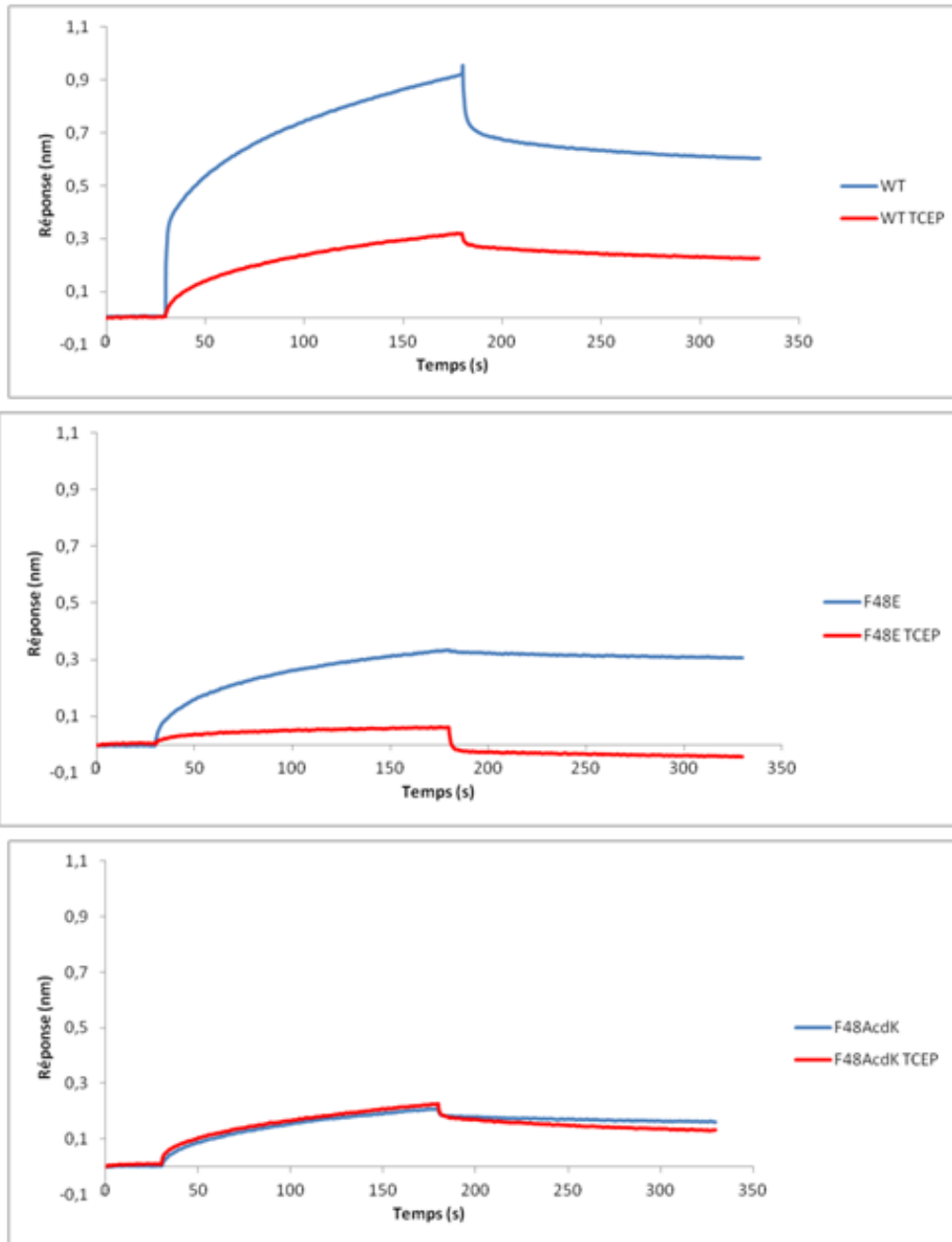


Figure 121: Effet du traitement TCEP sur les courbes d'affinité des différents mutants

Les mesures d'affinités sont imprécises, à cause de l'utilisation d'une seule concentration protéique. Cependant, les courbes d'association et de dissociation ont été réalisées à une concentration identique pour tous les mutants testés. Cette expérience nous donne donc des informations qualitatives sur leurs affinités. Les mutations testées ici n'ont cependant pas permis d'améliorer l'affinité de la protéine CBM40.

III. Incorporation d'une fonction boronate

Au terme de ma thèse, nous avons cherché à développer une méthode alternative supplémentaire afin de modifier l'affinité de lectines. Pour cela, nous avons cherché à incorporer la phénylalanine boronate dans des protéines cibles. En effet, cet acide aminé peut interagir avec des fonctions hydroxyles pour former des liaisons covalentes (Chapitre 5, section I.3.1). De plus, la phénylalanine boronate (Bpa) est un acide aminé non naturel commercial, nous serons donc moins limités par son approvisionnement, et la synthétase qui permet son incorporation a été décrite au sein de la bibliographie. Il s'agit de l'aaRS issue de *Methanococcus jannaschii* possédant les mutations Y32S, L65A, H70M, D158S et L162E (Schiefner et al. 2018).

III.1. Choix des protéines cibles et tests d'expression

Pour choisir les différentes protéines cibles en vue de tester l'incorporation de la phénylalanine boronate afin de modifier l'affinité de lectines, nous avons conservé les critères et la méthodologie utilisée section I.2. Cependant, afin de tester l'effet potentiel de la phénylalanine boronate sur la reconnaissance, nous avons réalisé des mutations de différents acides aminés en phénylalanine sur PyMOL avec la fonction "Mutagenesis". De plus, nous avons considéré un intervalle minimum de 2,5Å entre le cycle aromatique de la phénylalanine et les fonctions hydroxyles du substrat. En effet, c'est la distance que l'on peut mesurer entre le carbone du cycle aromatique portant la fonction boronate et l'extrémité de cette fonction au sein de structures cristallisées (PDB 5MHH et 5N5U). Trois protéines ont donc été sélectionnées selon ces critères : NagBb, GH98Sp et CLC1b.

III.1.1 Protéine cible n°1 : NagBb

La protéine NagBb est une alpha-N-acétylgalactosaminidase, dont la constante d'affinité K_m est de 0,23mM pour le GalNAc. La mutation D435A permet de supprimer son activité enzymatique et a été utilisée pour cristalliser la protéine avec le ligand (Sato et al. 2017). L'incorporation de la phénylalanine boronate (Bpa) en position Y329 devrait permettre son interaction avec les hydroxyles O3 et O4 du GalNAc comme nous pouvons le voir sur la Figure 122.

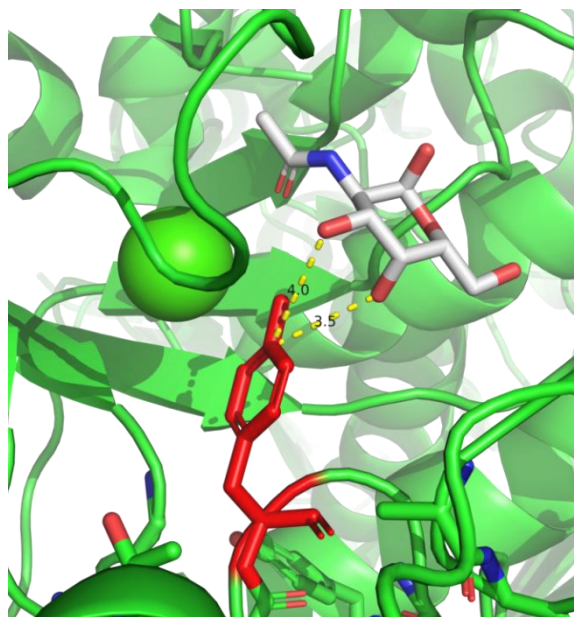


Figure 122: Site de reconnaissance de la protéine NagBb (PDB: 5WZN)

La Phénylalanine 329 (Y329) est colorée en rouge sur la figure, et les distances entre les hydroxyles et le cycle aromatique de la Y329 sont indiqués en angström au niveau des pointillés jaunes. Image réalisée dans PyMOL.

De plus, NagBb est une protéine qui est stable sur une grande gamme de pH (3-11) et ce malgré un maximum d'activité enzymatique à pH 5 (Kiyohara et al. 2012). Cette résistance au pH nous permettra de tester des pH élevés (pH aux alentours de 9) auxquels la fonction boronate est la plus réactive.

Une expérience d'*alanine scanning in silico* a également été réalisée à l'aide du logiciel YASARA et, avec une valeur de -0,47 kcal/mol, la mutation Y329 ne semble pas être une position dont la mutation affecterait la protéine structurellement (données non montrées). La mutagenèse de la protéine NagBb a donc été réalisée afin de générer un codon stop TAG à la place de la tyrosine 329.

III.1.2 Protéine cible n°2 : GH98Sp

La protéine GH98Sp est une glycoside hydrolase issue de *S. pneumoniae* de la famille 98, qui reconnaît qui est présent au niveau de cellules tumorales avec une affinité de l'ordre de 440-570 μ M avec les antigènes de Lewis A et B. La mutation E558A permet de supprimer son activité enzymatique et a été utilisée pour cristalliser la protéine avec le ligand (Higgins et al. 2009). L'incorporation de la Bpa en position H60 devrait permettre son interaction avec les hydroxyles O3 et O4 au niveau du fucose de l'antigène A-Lewis Y (qui correspond à Fuc-Gal-GalNAc) comme nous pouvons le voir Figure 123.

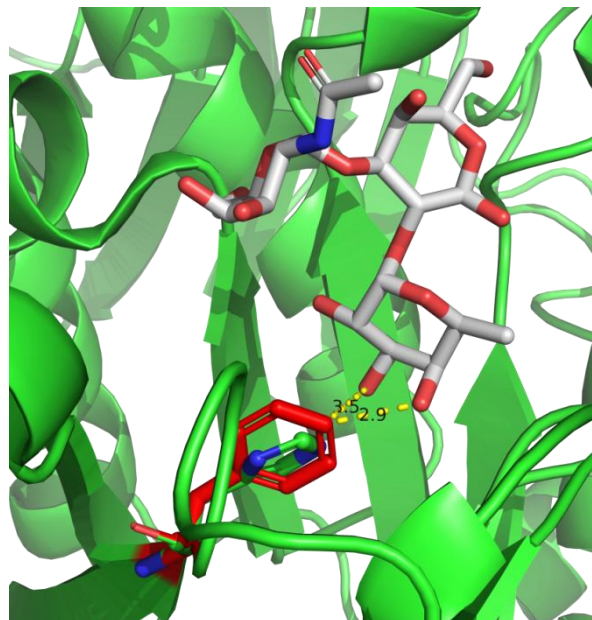


Figure 123: Site de reconnaissance de la protéine GH98Sp (PDB: 2WMI)

La position Histidine 60 (H60) a été mutée en phénylalanine qui a été représentée en rouge sur la figure. De plus, la distance entre le cycle aromatique de cette phénylalanine en position 60 et les hydroxyles O3 et O4 du fucose de l'antigène A-Lewis-Y (qui correspond à Fuc-Gal-GalNAc) a été indiquée en angström au niveau des pointillés jaunes sur la figure. Image réalisée dans PyMOL

Une expérience d'*alanine scanning in silico* a également été réalisée à l'aide du logiciel YASARA et, avec une valeur de 1,62 kcal/mol, la mutation H60 ne semble pas être une position dont la mutation affecterait la protéine structurellement (données non montrées). La mutagenèse de la protéine GH98Sp a donc été réalisée afin de générer un codon stop TAG à la place de l'histidine 60.

III.1.3 Protéine cible n°3 : CLC1b

La protéine CLC1b (aussi appelée CLC2) est une lectine de type C humaine, qui reconnaît une glycoprotéine, la Popadaline, au niveau de l'antigène Tn et également lorsque celui-ci est sialylé avec une affinité de l'ordre du micromolaire (Nagae et al. 2014). Cette protéine glycosylée est présente au

niveau de cellules cancéreuses, et notamment de cancers de la peau (Schacht et al. 2005). De plus l'incorporation de la phénylalanine boronate au sein du site de reconnaissance en position H25 pourrait interagir avec la fonction acide de l'acide sialique et donc favoriser sa reconnaissance (Figure 124).

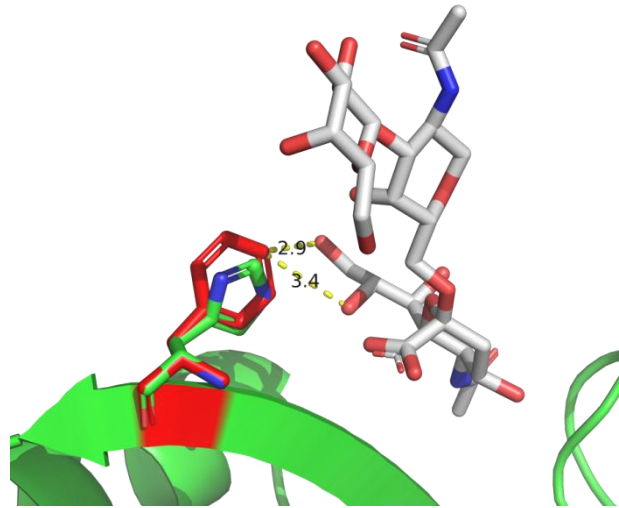


Figure 124 : Site de reconnaissance de la protéine CLC1b, (PDB: 3WSR)

L'histidine en position 25 a été mutée en phénylalanine qui a été représentée en rouge sur la figure. La distance entre les deux fonctions hydroxyles de l'acide sialique et le cycle aromatique de la phénylalanine en position 25 est indiquée en angström au niveau des pointillés jaunes sur la figure. Image réalisée dans PyMOL

Une expérience d'*alanine scanning in silico* a également été réalisée à l'aide du logiciel YASARA et, avec une valeur de 0,11 kcal/mol, la mutation H25 ne semble pas être une position dont la mutation affecterait la protéine structurellement (données non montrées). La mutagenèse de la protéine CLC1b a donc été réalisée afin de générer un codon stop TAG à la place de l'histidine 25.

III.2. Test d'expression des protéines cibles

Les trois protéines cibles n'ayant jamais été exprimées au laboratoire l'expression de leurs formes sauvages a d'abord été testée en milieu autoinductif et LB. Les protéines NagBb et CLC1b s'expriment efficacement au sein du milieu autoinductif ou LB, contrairement à la protéine GH98Sp dont les conditions d'expression en milieu autoinductif n'ont pas pu être déterminées (Figure 125). Etant donné, qu'au sein du laboratoire nous n'avons pas encore déterminé de conditions efficaces pour l'incorporation de NAA en milieu LB, nous avons donc décidé de mettre de côté la protéine GH98Sp. Etant donné que la protéine CLC1b est produite sous forme insoluble, nous avons donc privilégié la protéine NagBb afin de réaliser les premiers tests d'incorporation de la Bpa.

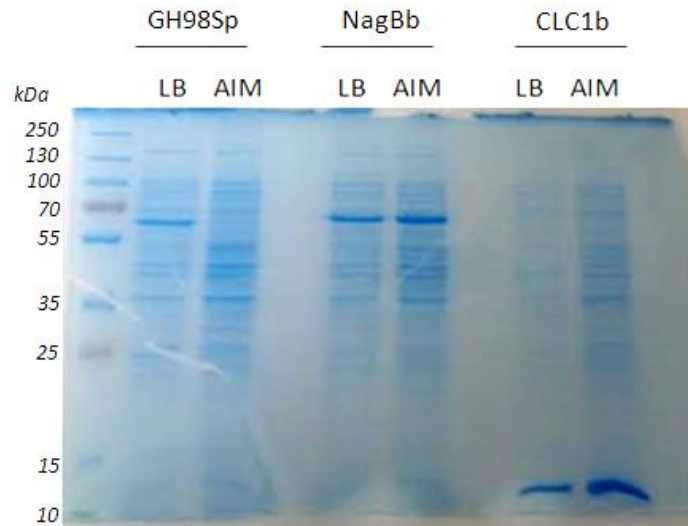


Figure 125: Analyse SDS-PAGE d'extraits bruts de cultures montrant l'expression des différentes protéines cibles en milieu LB ou autoinductif.

Les tailles respectives des protéines GH98Sp, NagBb et CLC1b sont respectivement : 69,8kDa, 72,1kDa et 15,9kDa.

III.3. Incorporation de la phénylalanine boronate au sein de NagBb

Afin d'incorporer la phénylalanine boronate au sein de notre protéine cible NagBb, nous avons généré les mutations Y32S, L65A, H70M, D158S et L162E au sein de la synthétase de *Methanococcus jannaschii* au niveau du plasmide pDULE.2 par méthode de mutagenèse dirigée, puis nous avons co-transformé ce plasmide avec les différents plasmides portant les différentes protéines cibles ayant une mutation TAG à la position d'incorporation choisie. Cependant, aucune incorporation n'a pu être visible sur SDS-PAGE ou WesternBlot. Après contact avec l'équipe du professeur Skerra (Munich Center for Integrated Protein Science), les chercheurs nous ont transmis le plasmide qu'ils ont utilisé afin d'incorporer efficacement la Bpa. A l'aide de ce plasmide, nous avons pu observer l'incorporation de la phénylalanine boronate au sein de la protéine NagBb (Figure 126). Cependant, à ce jour par manque de temps, nous n'avons pas encore testé l'incorporation de la Bpa au sein d'autres cibles protéiques, ni produit NagBb avec Bpa à grande échelle ni réalisé de tests d'affinité sur cette protéine.

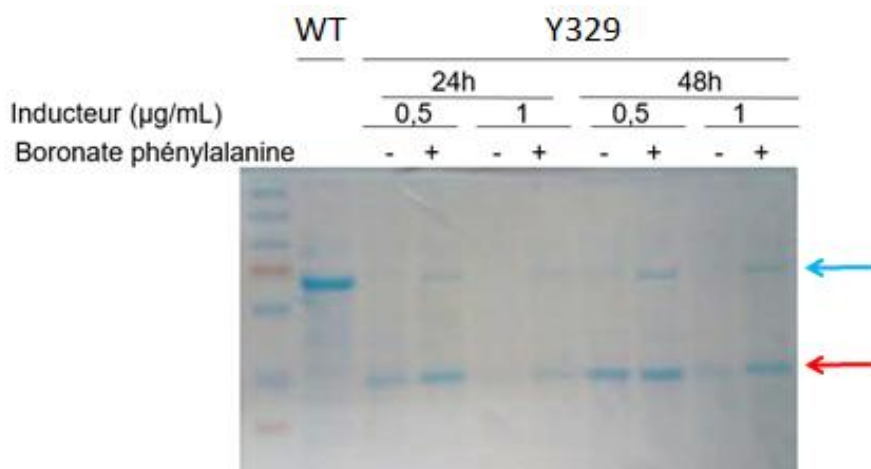


Figure 126: Analyse SDS-PAGE montrant l'incorporation de la Boronate phénylalanine au sein de la protéine NagBb-Y329TAG.

La taille de la protéine tronquée (environ 36,9kDa) est indiquée par une flèche rouge et la taille de la protéine complète ou sauvage (WT) est indiquée par une flèche bleue.

Conclusion

Les lectines sont des protéines avec généralement une haute spécificité mais une affinité faible envers leur ligand saccharidique. Afin d'augmenter cette affinité, il est possible d'exploiter des propriétés de multivalence qui sont d'ailleurs naturellement présentes chez de nombreuses lectines. La modification de valence de lectines a d'ailleurs été explorée par Tobola *et al.*, lors de la mise au point de 3 lectines multivalentes (RSL, Gal-1 et Stx1B) monospécifiques ou bispécifiques grâce à leur couplage par l'intermédiaire d'acides aminés non naturels (Tobola et al. 2019). Les mesures d'affinité de ces différentes lectines multivalentes n'étaient pas présentées dans cette publication.

Au cours de cette thèse, notre stratégie a consisté à augmenter l'affinité de lectines monovalentes par l'incorporation d'un acide aminé non naturel ayant une réactivité forte envers le ligand. Il s'agit d'une méthode très originale. En effet, l'incorporation d'acides aminés non naturels au sein de site actif n'est pas très fréquente, d'autant plus lorsqu'il s'agit de modifier les caractéristiques de reconnaissance de protéines.

Au sein de la littérature, nous pouvons noter deux exemples utilisant des méthodes aléatoires : par phage display afin de générer un anticorps contre la glucamine et la modification de la spécificité d'une aldolase (C C Liu et al. 2009; Windle et al. 2017). Dans le cas de l'anticorps, des mutations aléatoires ont été générées, puis les différents mutants ont été sélectionnés par phage display. Un anticorps contenant une Phénylalanine-boronate avec une affinité de l'ordre du micromolaire a pu être ainsi obtenu (C C Liu et al. 2009). Dans le cas de l'aldolase, une cystéine a été modifiée chimiquement en dehydroalanine afin de tester rapidement plusieurs radicaux thiols non naturels. Cette méthode a alors permis de modifier la spécificité de la protéine pour les chaînes carbonées plus courtes (Windle et al. 2017). Nous pouvons également noter deux exemples ayant des approches plus ciblées : la modification d'une anticaline humaine, et la modification de la lectine RSL par des fluorotryptophanes (Tobola et al. 2018; Sommer, Eichinger, and Skerra 2019). Dans le cas de l'anticaline, une mutation au sein du site actif en phénylalanine-boronate (Bpa) a permis la génération d'un lien covalent avec un ligand (4-nitrocatechol) (Sommer, Eichinger, and Skerra 2019). Cette liaison était favorisée par la présence d'un environnement électrophile et a permis à l'anticaline d'avoir des capacités de reconnaissance de sucre de l'ordre du micromolaire. Dans le cas de la lectine RSL, les fluorotryptophanes ont été incorporés de façon résidu spécifique et, bien qu'une liaison hydrogène supplémentaire ait pu être observée avec le ligand, l'effet sur l'affinité global reste limité (Tobola et al. 2018).

Au niveau de ce travail de thèse, nous avons cherché à changer le mode de liaison de lectines à leur ligand par la génération d'une liaison covalente avec le ligand naturel. Les lectines reconnaissent leur ligand principalement par l'intermédiaire de liaisons hydrogènes avec les fonctions hydroxyles. En effet, l'orientation des fonctions hydroxyles définit un saccharide. Afin de générer ce lien covalent, nous avons alors envisagé d'incorporer une fonction aldéhyde au sein de la protéine. Cette fonction pourrait ainsi former des liaisons covalentes de type acétal avec des fonctions hydroxyles ou imines avec des fonctions amines. Les liaisons imines sont fréquentes au sein d'enzymes ou récepteurs par le biais de chaînes aminées telles que les lysines. Dans notre stratégie, la fonction amine ou alcool est portée par le ligand et la fonction aldéhyde est portée par la protéine. Cependant, étant donné qu'il s'agit d'une fonction très réactive, nous avons ciblé l'incorporation d'une fonction aldéhyde masquée qui peut être libérée en présence du ligand. Ce type de fonction est absent au sein des acides aminés naturels et également très rares au sein des acides aminés non naturels ayant été incorporé de manière site spécifique. Nous avons alors entrepris d'incorporer un acide aminé non naturel original portant une fonction aldéhyde photoactivable (l'ONpGlu). Pour cela, un mutant de l'aminocacyl-ARNt synthétase capable de le prendre en charge doit être déterminé.

Nous avons donc mis en place une méthode de sélection en complément d'une méthode de pré-sélection *in silico*. Malheureusement, aucune incorporation de l'ONpGlu n'a pu être observée dans une protéine modèle. En parallèle, nous avons alors incorporé un acide aminé non naturel portant une fonction aldéhyde activable par réduction (AcdK). Celui-ci a été incorporé efficacement au sein de deux cibles protéiques dont les affinités ont pu être testées. Cependant, bien que la déprotection de l'acide aminé semble être efficace, l'affinité n'apparaît pas modifiée. Une troisième stratégie est en cours et consiste en l'incorporation d'un acide aminé non naturel portant une fonction boronate (Bpa). En effet, cette fonction peut également former des fonctions covalentes avec des fonctions hydroxyles. Cet acide aminé a alors pu être incorporé au sein d'une cible protéique. Cette incorporation doit être répétée en grande quantité afin de pouvoir en tester l'affinité.

Ces résultats démontrent la difficulté de l'utilisation d'acides aminés non naturels appliqués à l'ingénierie de protéine, particulièrement afin de modifier ces caractéristiques de liaison. En effet, nous pouvons remarquer que les études bibliographiques existantes n'ont également eu qu'un effet limité sur l'affinité obtenue lorsque celle-ci avait pu être testée.

Généralement, les méthodologies traditionnelles utilisent des méthodes de mutagenèse semi-rationnelles ou aléatoires afin de tester un maximum de mutations car il est difficile d'en prédire l'effet. Au cours de ce travail, le nombre de mutant a été limité à un ou deux par cible protéique, ce qui paraît hasardeux. Cependant, les limites liées à la quantité d'acide aminé non naturel disponible, ainsi que son efficacité d'incorporation génèrent des difficultés pour utiliser des méthodes traditionnelles de mutagenèses suivies de sélections. Afin de favoriser l'aboutissement de ce type de stratégie, il apparaît bénéfique de s'assurer de la disponibilité commerciale de l'acide aminé non naturel ou la facilité de sa synthèse. Afin de contourner cette problématique, il pourrait être intéressant d'exploiter l'incorporation d'un acide aminé non naturel commun ayant une réactivité particulière qui servira alors d'ancrage pour tester plusieurs radicaux comme au cours de l'étude réalisée par Windle *et al.* (Windle *et al.* 2017). Ainsi, de nombreuses possibilités pourraient être testées à différentes positions avec un plus grand débit. Cette méthode pourrait alors donner une première approche plus large pour générer la mutagenèse de cibles protéiques. Nous pouvons cependant noter qu'il faut que les conditions de couplage soient compatibles avec la protéine cible.

Par ailleurs, l'acide aminé non naturel incorporé doit également être le plus adapté possible au site actif cible. En effet, afin de s'assurer la meilleure réactivité possible, il faut que les fonctions réactives se retrouvent selon un angle et une distance favorable. Une stratégie appliquée à une protéine largement décrite, dont les interactions sont déjà connues a donc probablement plus de chance d'aboutir. Ainsi le contexte dynamique et l'espace disponible pour l'incorporation de l'acide aminé non naturel sera mieux maîtrisé. En effet, la flexibilité de la chaîne principale, des différentes chaînes latérales ainsi que l'impact dynamique de mutations contextuelles peut permettre de mieux prédire les interactions réalisables par l'acide aminé non naturel incorporé. Par exemple, au niveau de l'anticaline ayant été modifiée avec la boronate phénylalanine, les auteurs ont pu montrer que la proximité d'un acide aminé hydrophile pouvait stabiliser la position de la fonction boronate afin de favoriser sa réactivité (Sommer, Eichinger, and Skerra 2019). De plus, il pourrait être envisageable d'utiliser des méthodes de prédiction bioinformatique afin de préciser la stratégie de mutagenèse employée. En effet, il est possible de générer les champs de force encodant un acide aminé non naturel afin d'en déterminer la flexibilité, les degrés de rotation, ainsi que les différents rotamères les plus probables, puis d'intégrer ces paramètres au sein d'une stratégie de sélection de mutants *in silico*. Ainsi, une librairie d'acides aminés non naturels pourrait être testée *in silico* à plusieurs positions afin de réduire la quantité de tests biologiques à réaliser et donc la quantité d'acide aminé non naturel nécessaire.

Les acides aminés non naturels peuvent apporter de nouvelles réactivités très fortes et très spécifiques. Leur utilisation afin d'évoluer des protéines est donc très prometteuse. Cependant, leur utilisation est encore complexe et n'est pas aussi maîtrisée que dans le cadre de mutagénèse avec des acides aminés classiques. Il est possible que les progrès technologiques permettant un meilleur taux d'incorporation, l'incorporation de nouveaux acides aminés non naturels aux réactivités nouvelles, ainsi que la mise au point de nouvelles méthodes de sélection adaptées et à plus haut débit favoriseront l'aboutissement de ces stratégies d'ingénierie originales.

***Chapitre 7 : Matériels et Méthodes pour la
modification de lectines***

I. ONpGlu : Mise au point d'une méthode de sélection

De manière générale, les protéines qui reconnaissent les sucres ont une affinité faible. Le but de ce projet est d'utiliser un acide aminé non naturel afin de modifier l'affinité des protéines cibles. Etant donné que les fonctions disponibles pour former des interactions sur un sucre sont la plupart du temps des fonctions hydroxyles ou des fonctions amines, nous avons essayé d'utiliser un acide aminé qui porte une fonction aldéhyde car cette fonction peut former des liaisons covalentes avec des fonctions hydroxyles ou amines. Cependant, la fonction aldéhyde étant très réactive, elle doit être protégée. Nous avons donc développé une stratégie consistant en l'incorporation d'un acide aminé non naturel original portant une fonction aldéhyde protégée par un groupement photolabile : l'Ortho-Nitro-para-Glutamate (ONpGlu). Cependant cet acide aminé non naturel étant original, il n'y a donc pas de synthétase décrite permettant son incorporation. Des mutations doivent donc être générées au sein du site actif de l'enzyme afin de trouver un mutant capable de le prendre en charge. Une méthode de sélection basée sur une analyse bioinformatique (section I.3 ci-dessous) couplée à une technique de fluorescence (section I.2 et I.4 ci-dessous) a donc été mise en place.

I.1 Synthèse de l'ONpGlu

La synthèse d'un acide aminé non naturel portant une fonction aldéhyde photoprotégée a été imaginée et réalisée par le CEISAM partenaire du projet. Tout d'abord la synthèse du composé aldéhyde intermédiaire (n°6, Figure 127) a été planifiée à partir des synthèses du composé intermédiaire 3 décrite par Y.Itoh *et al.* (Itoh et al. 2016), et de la synthèse de l'intermédiaire 2 au composé 6 décrite par Illies *et al.* (Illies et al. 2010). Elle se réalise en 5 étapes à partir d'un dérivé de l'acide glutamique (Figure 127).

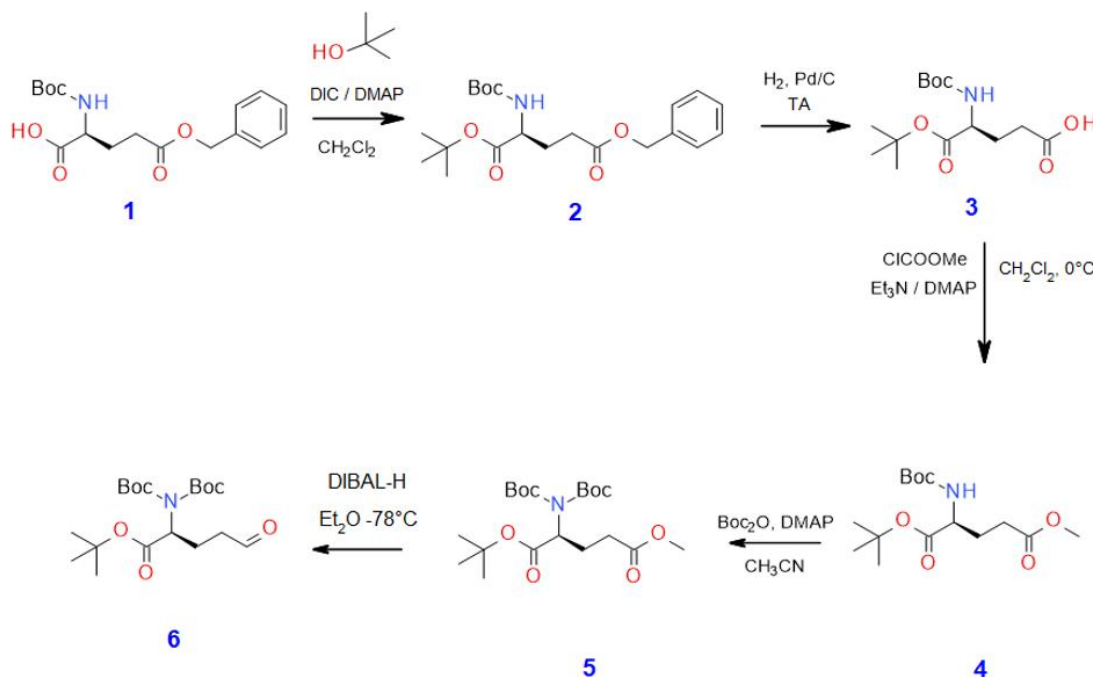


Figure 127 : Méthode de synthèse de l'aldéhyde intermédiaire (composé 6).

Une fois ce composé aldéhyde produit, la fonction aldéhyde sera masquée par un groupement photolabile. Pour cela, deux composés diols aux configurations différentes peuvent être utilisés (Figure 128). La production de ces deux composés a été envisagée, ce qui signifie donc que l'évolution de l'aminoacyl-ARNt synthétase devra être réalisée indépendamment pour ces deux composés. Il est en effet peu probable qu'un même mutant de cette enzyme reconnaisse les deux composés. Néanmoins, nous pouvons remarquer qu'après démasquage de la fonction aldéhyde,

l'acide aminé non naturel généré est le même pour ces deux composés, qui peuvent donc être incorporés indifféremment au sein des mêmes cibles protéiques.

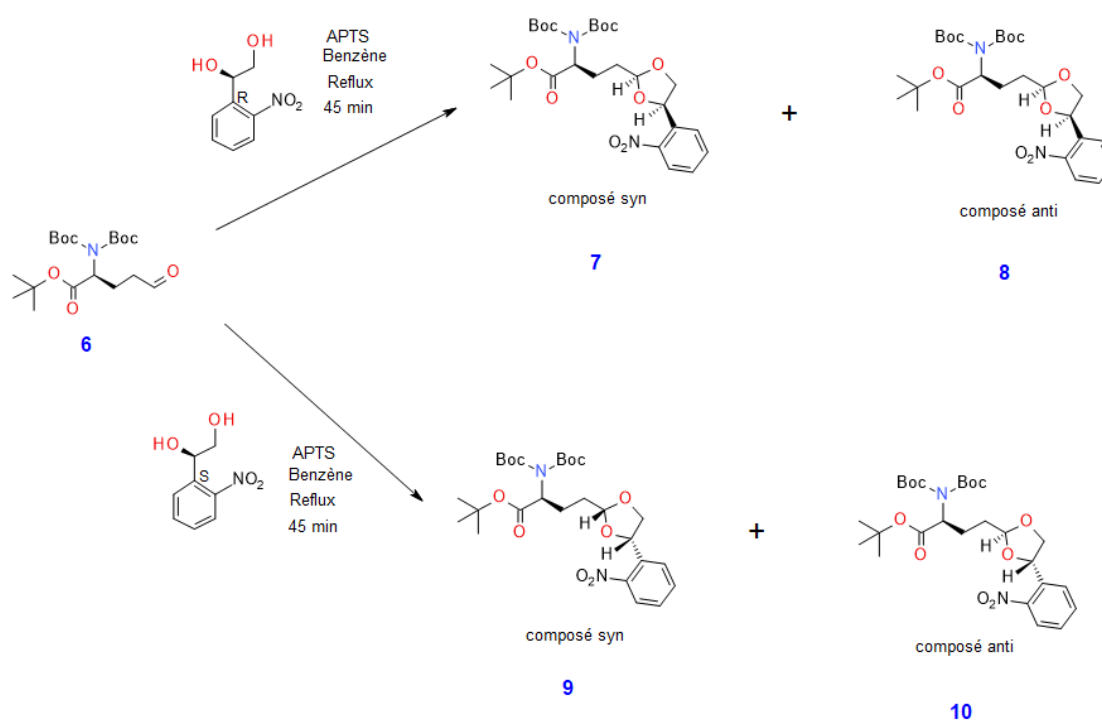


Figure 128 : Ajout du groupement permettant le masquage de la fonction aldéhyde. Deux composés diols avec des configurations différentes ont été testés.

Le composé majoritaire sera ensuite purifié et les fonctions amine et acide carboxylique de la chaîne principale seront déprotégées afin de générer l'acide aminé non naturel final.

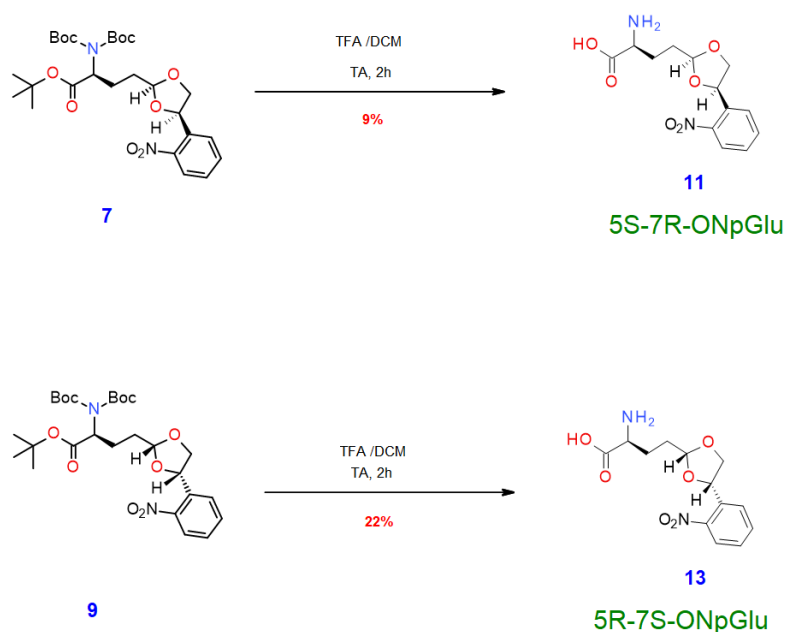


Figure 129 : Etape finale de la synthèse permettant de produire l'acide aminé non naturel déprotégé

I.2 Biologie moléculaire

I.2.1 Clonages EGFP en aval de mPsaA-K193TAG

Afin de sélectionner les différents mutants de la synthétase qui seront efficaces lors de l'incorporation de l'ONpGlu, une stratégie basée sur la production d'une protéine de fluorescence est développée. Pour cela un gène de l'eGFP est cloné au sein du plasmide pET24d en fusion d'un gène de la mPsaA, qui a été muté pour générer un ou deux codons stop TAG (K193 ou K105-K193) afin de réaliser l'incorporation d'acides aminés non naturels.

Grâce à l'utilisation de ces constructions, l'apparition de fluorescence sera alors conditionnée par l'incorporation de l'acide aminé non naturel. En effet, en l'absence de celui-ci, la protéine produite sera tronquée au niveau du codon stop TAG. Cependant, si l'acide aminé a été incorporé, il y aura suppression du codon ambre TAG et donc production de la protéine complète, y compris l'eGFP et donc apparition de fluorescence. Ainsi, ce plasmide sera co-transformé avec le plasmide pEVOL-1copie (défini ci-dessous, section I.2.2) portant les différents mutants des aaRS à sélectionner (voir section I.4.2). Lors de la sélection, plusieurs cas de figures peuvent se présenter (ils sont résumés dans le tableau ci-dessous). Dans un premier cas type, le mutant aaRS est inefficace pour l'incorporation de l'ONpGlu, il n'y aura donc pas de fluorescence avec ou sans acide aminé non naturel dans le milieu. Dans un autre cas, le mutant aaRS n'est pas spécifique de l'acide aminé non naturel et il y aura apparition de fluorescence avec et sans acide aminé non naturel. Enfin troisième possibilité, le mutant aaRS testé est efficace et spécifique de l'ONpGlu, et il y aura apparition de fluorescence uniquement en présence de celui-ci. Ce sont ces derniers mutants que nous cherchons à sélectionner.

Tableau 26 : Exemples récapitulants les possibilités liées à la méthode de sélection employée

Exemple mutant aaRS	Caractéristique	Fluorescence en présence de l'ONpGlu	Fluorescence en l'absence de l'ONpGlu
Mutant 1	Pas efficace	Non	Non
Mutant 2	Pas spécifique	Oui	Oui
Mutant 3	Efficace et spécifique	Oui	Non

Le gène eGFP est donc cloné en aval du gène de la mPsaA (WT, puis muté K193TAG ou K105TAG-K193TAG) au sein du plasmide pET24d. Cette construction a déjà été utilisée lors de l'optimisation de l'incorporation de la Propargyl-Lysine par la méthode de plan d'expérience, le protocole du clonage y est décrit dans le chapitre 4 (section IV.2). Les mutations permettant de générer les codons stops TAG ont été générées selon la méthode décrite chapitre 4 (section II.4), à l'aide des amorces n°25-28 décrites en annexe.

I.2.2 Modification du plasmide pEVOL : suppression d'une des deux copies de l'aaRS

Le plasmide pEVOL, originellement utilisé pour l'incorporation d'acides aminés non naturels, contient deux copies de la pyrrolysyl-ARNt synthétase de *M. mazei* dont une sous le contrôle d'un promoteur constitutif et une sous le contrôle d'un promoteur inductible à l'arabinose. Ce plasmide nous a été transmis par Edward Lemke de l'EMBL grâce à la réalisation d'un *Material Transfert Agreement* (MTA) (T. S. Young et al. 2010). Lors de la sélection d'une nouvelle synthétase, plusieurs mutants de ce gène vont devoir être testés. Afin de faciliter les étapes de mutagenèse et de

séquençage, la copie de la synthétase sous le contrôle du promoteur constitutif a été supprimée par PCR inverse.

Pour cela le plasmide a été amplifié à l'aide de deux amorces (amorces 11 et 12 en annexe 4) excluant ce gène de la synthétase en suivant le protocole de PCR suivant : Q5 polymérase (NEB, M0491S), 1) dénaturation initiale 98°C pendant 3min, les étapes 2 à 4 sont répétées 25 fois : 2) dénaturation 98°C pendant 30sec, 3) hybridation 58°C pendant 30sec, 4) élongation 72°C pendant 5min (1min/kb), puis 5) élongation finale 72°C pendant 15min. Le mélange PCR a ensuite été incubé avec le mélange KLD (NEB, M0554S) contenant une kinase, une ligase, et *DpnI* afin de procéder en une seule étape à la destruction du plasmide matrice non modifié et à la phosphorylation et ligation du nouveau plasmide généré. Pour cela, 1µL de la PCR a été incubé 5min à température ambiante avec 5µL de tampon 2X KLD et 1µL du mélange d'enzymes dans 10µL de volume final. Cinq microlitres de ce mélange ont ensuite été transformés au sein d'XL1b compétentes (méthode TSS). Les étapes de transformation et de vérification de séquence ont été réalisées comme décrit précédemment (cf chapitre 4, section II.3) avec les amorces n° 13 et 14 présentées en annexe.

I.3 Génération des mutants *in silico*

L'ONpGlu n'ayant jamais été incorporé, il faut trouver les mutations nécessaires au sein du site actif de la synthétase afin que celui-ci puisse être incorporé. Pour cela, une stratégie rationnelle a été envisagée, une analyse bioinformatique réalisée par Johann Hendrickx (Ingénieur d'études au sein du laboratoire) qui a permis de cibler des mutations clés. Cette étude a été réalisée pour deux stéréoisomères de l'acide aminé non naturel ONpGlu (le 5R-7S_ONpGlu et le 5S-7R-ONpGlu). L'analyse s'est déroulée en trois parties : une analyse pour préparer la structure de la protéine à muter, ainsi qu'une analyse pour préparer les acides aminés substrats, puis une sélection des mutations.

I.3.1 Préparation de la protéine

La structure de la protéine sauvage a d'abord été générée à partir de structures PDB (2ZIM, 4BW9 et 5K1P) à l'aide du logiciel MODELER (version 9.19) (Šali and Blundell 1993). La protéine Pyrrolysyl-ARnt synthétase issue de *Methanosarcina mazei* possède une boucle ($\beta 7$ - $\beta 8$) qui peut se placer en deux conformations possibles (fermant le site actif en le protégeant d'une potentielle hydrolyse ou ouverte). Les Trois structures PDB ont donc été utilisées afin de générer une structure qui prenne en compte les deux types de conformations tout en positionnant l'ATP (co-substrat nécessaire à l'activité de l'enzyme) et deux molécules d'eau très conservées à proximité de l'ATP au niveau du site actif.

I.3.2 Préparation des acides aminés ONpGlu substrats

3240 différentes conformations de l'ONpGlu (libres rotations, déformation du cycle acétal) ont été générées, puis une minimisation d'énergie a été appliquée pour sélectionner les conformations les plus stables énergétiquement et donc les plus probables, ce qui a permis de cibler entre 1800 et 2600 conformations qui ont été classées en environ 200-300 clusters ayant des conformations très proches (RMSD < 0.1Å). Ces différents clusters ont ensuite été testés sur la topologie du site de reconnaissance. Pour cela, les chaînes latérales des acides aminés présents dans le site de reconnaissance de la protéine ont été enlevées et si les conformations de l'ONpGlu testées entraînent un clash stérique avec le squelette même de la protéine (chaînes principales), cela indique

que la conformation ne pourra pas être prise en charge par la protéine quelques soient les mutations appliquées au sein du site actif et ces conformations sont donc éliminées. *In fine*, cela a permis de sélectionner 5 conformations pour le 5R-7S_ONpGlu et 18 conformations pour le 5S-7R-ONpGlu.

1.3.3 Génération des mutations et sélections des séquences les plus probables

Douze acides aminés au niveau du site actif (A302, L305, Y306, L309, I322, N346, C348, Y384, V401, I413, W417 et G419) ont été sélectionnés en fonction de leur proximité avec l'acide aminé non naturel au niveau du site et évolués *in silico*. Le protocole Couples Moves au sein de Rosetta développé par Ollikainen et al., a ensuite été appliqué afin de sélectionner les mutants les plus probables (Ollikainen, de Jong, and Kortemme 2015). Ce protocole consiste à faire des cycles au cours desquels des échantillonnages de protéines sont générés (par la prise en compte de la flexibilité de la protéine) ou au cours desquels des échantillonnages de substrat sont générés (degré de liberté du ligand au niveau des conformations possibles). Ces différents échantillonnages sont ensuite classés à l'aide d'une fonction d'énergie tout-atome de la suite Rosetta (Alford et al. 2017) Ces étapes cycliques ont été répétées 200 fois pour un modèle de MmPyIRS. Afin de compenser l'insuffisance de la flexibilité des boucles générée par le protocole, dix modèles de départ de MmPyIRS avec des conformations variées de boucle ont été utilisés. Ainsi, le protocole correspondant à un cycle a été appliqué 2000 fois, générant ainsi environ 50000 séquences redondantes. Les séquences ayant une énergie prédite défavorable sont éliminées. Les séquences restantes sont classées par fréquence d'observation, la fréquence la plus importante correspondant à celle du meilleur mutant potentiel.

Etant donné les énergies obtenues lors de la réalisation du protocole et pour plus de finesse de prédiction, certaines positions les moins évoluées ont été exclues et le protocole a été relancé avec uniquement cinq positions à évoluer (Y306, N346, C348, Y384, et W417). L'élimination de certaines positions a été déterminée à partir des résultats de la première simulation (le score associé à ces positions privilégiait de ne pas les muter ou très peu fréquemment), couplés à une analyse visuelle de la structure. Cela permet de réduire la diversité de l'échantillonnage de départ et donc de favoriser la fiabilité de la sélection de mutants par le protocole Couples Moves de Rosetta.

1.4 Méthodes de sélection

1.4.1 Test de pénétration de l'ONpGlu au sein des cellules BL21(DE3)

Avant de sélectionner les différentes synthétases qui pourraient incorporer l'ONpGlu, sa pénétration dans les cellules bactériennes a été vérifiée. En effet, le mode de transport d'un acide aminé non naturel original n'est pas connu et il est important de vérifier que celui-ci peut effectivement entrer dans la bactérie. Son entrée dans les bactéries a été vérifiée par une analyse sur CCM suivant la méthode décrite chapitre 4, partie IV.1. Pour cela, une culture de 5mL a été réalisée selon le protocole (Chapitre 4 III.3) en absence et en présence d'ONpGlu. La culture a ensuite été culotée par centrifugation à 5000g pendant 10min, puis le culot a été lavé 3 fois avec 5mL de tampon froid ($\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ 50mM, NaCl 150mM, pH8). Le culot a ensuite été resuspendu dans 300 μL de ce tampon et a été lysé par sonication (30sec, cycles 5sec-5sec, amplitude 50%, appareil : Thermo Fischer, FB120-220), puis centrifugé à 10 000rpm pendant 15min à 4°C. Enfin, 1 μL du surnageant de lyse a été déposé sur CCM (Eluant : Butanol, Acide acétique, Eau (proportions 10/2,5/2,5), et révélateur : Ninhydrine) en parallèle de l'ONpGlu pur.

1.4.2 Génération des mutations prédites par l'analyse bioinformatique

Il faut une synthétase spécifique pour pouvoir transférer l'ONpGlu sur l'ARNt orthogonal portant le codon complémentaire au codon TAG et ainsi permettre son incorporation dans une protéine d'intérêt. Cependant, comme il s'agit d'un acide aminé non naturel original, les mutations nécessaires pour permettre à l'aaRS de reconnaître et prendre en charge l'ONpGlu ne sont pas connues. Une analyse bioinformatique (section 1.3) a permis de prédire des mutations susceptibles d'être efficaces.

Cinq points de mutations ont été prédits (Y384F, Y306F, N346A/G, C348A/G et W417C/N/A). Ces mutations ont été générées de façon combinatoire à l'aide de mutagénèses successives selon le protocole suivant.

Le plasmide à modifier est amplifié par PCR (PCR (Phusion (Thermo Fisher F-530S), 1) dénaturation initiale 98°C pendant 2 min, puis les étapes 2 à 4 sont répétées 25 fois 2) dénaturation 98°C pendant 30 sec, 3) hybridation 58°C pendant 30 sec, 4) élongation 72°C pendant 1min enfin 5) élongation finale 72°C pendant 10min, en utilisant les amorces n°29-44). Puis le plasmide parental non modifié est éliminé par traitement avec l'enzyme *DpnI* (ThermoFisher, ER1701, 10U par réaction, incubation pendant 1h30 à 37°C, puis inactivation par 20 min à 80°C). Enfin, un dixième de la réaction de mutagénèse (soit 5µL) est transformé dans des bactéries DH5α ou XL1b compétentes. Les modifications sont ensuite vérifiées par séquençage des miniprep obtenues à partir des clones transformés, comme décrit ci-dessus avec des amorces de séquençage (n°13 et 14 en annexe).

1.4.3 Méthodes expérimentales de sélection

La méthode de sélection des synthétases est basée sur l'apparition de fluorescence lorsqu'il y a une incorporation de l'acide aminé non naturel (description section 1.2) dans la protéine de fusion mPsaA-GFP. Il est possible de réaliser cette sélection sur milieu solide ou milieu liquide. Nous avons d'abord mis au point la sélection sur milieu solide puis en milieu liquide. Cette méthode a d'abord été mise au point sur l'incorporation de la Propargyl-Lysine avec l'aaRS adaptée afin de trouver les bonnes conditions avant de pouvoir procéder à des sélections pour trouver des synthétases efficaces pour l'incorporation de l'ONpGlu. Plusieurs souches ont été utilisées :

- une souche BL21(DE3) contenant le plasmide pEVOL co-transformé avec le plasmide pET24-mPsaA(WT) en tant que contrôle négatif de fluorescence
- une souche BL21(DE3) contenant le plasmide pEVOL co-transformé avec le plasmide pET24-mPsaA-eGFP(WT) en tant que contrôle positif de fluorescence
- une souche BL21(DE3) contenant le plasmide pEVOL co-transformé avec le plasmide pET24-mPsaAK193TAG-eGFP(WT) pour tester la sélection
- une souche BL21(DE3) contenant le plasmide pEVOL co-transformé avec le plasmide pET24-mPsaAK105TAG-K193TAG-eGFP(WT) pour tester la sélection dans des conditions plus stringentes. En effet, le rendement de l'incorporation double d'acide aminé non naturel est toujours beaucoup plus faible.

-Méthodes de sélections sur boîte de Pétri

Le protocole de sélection sur milieu solide est réalisé selon le protocole suivant. Des précultures ont été réalisées à partir de colonies des différentes souches listées ci-dessus dans 5mL de LB avec antibiotiques (kanamycine et chloramphenicol) et incubées sur la nuit à 37°C. Le lendemain, les cultures ont été diluées par 4 dans du LB, puis une goutte de 5µL a été déposée sur milieu solide contenant du milieu autoinductif (FORMEDIUM, n° AIMLB0210), de l'agar et sur laquelle nous avons absorbé des solutions d'arabinose (300µL à 20%) et de Propargyl-Lysine (300µL à

100mM). Les boîtes de pétri ont ensuite été incubées à 37°C sur la nuit. La fluorescence est ensuite visualisée à l'aide d'une plaque UV (VilberLourmat, UV/White Converter plate).

-Méthodes de sélection sur plaque 96 puits

La sélection en milieu solide permet d'avoir une indication qualitative de la fluorescence à l'aide du lecteur UV. Cependant, pour avoir une sélection plus quantitative ainsi que pour baisser la quantité d'acide aminé non naturel utilisée lors des tests de sélections, nous avons développé une méthode de sélection en milieu liquide dont le protocole est le suivant. Des précultures à partir de colonies des différentes souches listées ci-dessus ont été réalisées dans 5mL de LB avec antibiotiques (kanamycine et chloramphenicol) et incubées sur la nuit à 37°C. Le lendemain, les précultures sont diluées au 100^{ème} dans 1mL de milieu autoinductif (FORMEDIUM, n° AIMLB0210) avec de l'arabinose à 0,02% final et avec ou sans Propargyl-Lysine (1mM final). Deux cent microlitres de cette culture sont alors répartis en triplicatas dans des puits de plaques 96 (GrenierBioOne, 655101), puis incubés à 37°C avec agitation (à l'aide de l'agitateur Heidolph Titramax incubator 1000) sur environ 8h avec des films respirants (VWR 60941-086), puis sur la nuit à 37°C avec des films non respirant (VWR 60941-062) pour limiter l'évaporation. Des mesures de densité optique et de fluorescence (excitation : 488nm, émission : 509nm) sont ensuite prises à l'aide d'un appareil de mesure (TECAN, Infinite M1000).

Cette méthode a ensuite été utilisée avec l'ONpGlu afin de tester les différents mutants de la synthétase prédit par bioinformatique (section I.3)

I.4.4 Sélection sur librairie (réalisée en Oregon)

Alors que nous n'avions pas réussi à trouver des mutants de l'aaRS efficace pour l'incorporation de l'ONpGlu, nous avons pu tester l'incorporation de l'ONpGlu sur une petite librairie issue du laboratoire de Ryan Mehl en Oregon (USA). En effet, cette expérience a été réalisée au cours d'un workshop sur l'incorporation d'acides aminés non naturels avec la méthode site spécifique, appelé « Genetic Code Expansion Workshop » à l'Université de l'Etat d'Oregon. Nous avons la possibilité de tester une centaine d'aminoacyl-ARNt synthétase issues de leur librairie afin d'essayer de trouver un mutant capable d'incorporer l'ONpGlu. Une souche bactérienne BL21(DE3) avait été préalablement transformée avec un plasmide portant les différents mutants de la synthétase et un plasmide pET28a portant un gène d'expression pour la GFP avec un codon stop TAG en position 150. Le principe de sélection est donc le même que celui que nous avons mis en place (section I.2.1). En pratique, des précultures de leur librairie réalisées en plaques 96 puits (Deep Wells) ont été diluées au 100^{ème} au sein de nouvelles plaques (Deep Wells) dans du milieu autoinductif (milieu décrit dans la publication de Hammill et al.) avec l'ONpGlu à 1mM et les antibiotiques de sélection (ici Kanamycine 25µg/mL et Tétracycline 12µg/mL) (Hammill et al. 2007). La culture est incubée à 37°C pendant 48h et des mesures de fluorescence et de densité optique (600nm) sont réalisées à 24h, 36h et 48h.

II. Utilisation de l'AcidK

L'incorporation de l'ONpGlu n'a pas pu être observée, cependant un acide aminé non naturel qui possède une fonction aldéhyde protégée a été incorporé et ce travail a été publié courant décembre 2016 par Wang *et al.* (Z. A. Wang et al. 2017). Afin de poursuivre le projet de modification de l'affinité d'une protéine qui reconnaît des sucres à l'aide d'un NAA, nous avons poursuivi l'étude avec cet acide aminé non naturel : la 4-azidobenzoxycarbonyl)-δ, ε-dehydrolysine (ici appelée AcidK).

II.1 Synthèse de l'acide aminé non naturel AcdK

La synthèse de l'AcdK a d'abord été reproduite à partir du protocole décrit par Wang *et al.* (Z. A. Wang et al. 2017) par Antoine Richieux (Post-doctorant au CEISAM), puis a été montée en échelle par Denis Cadet (Stagiaire M1 à l'UFIP).

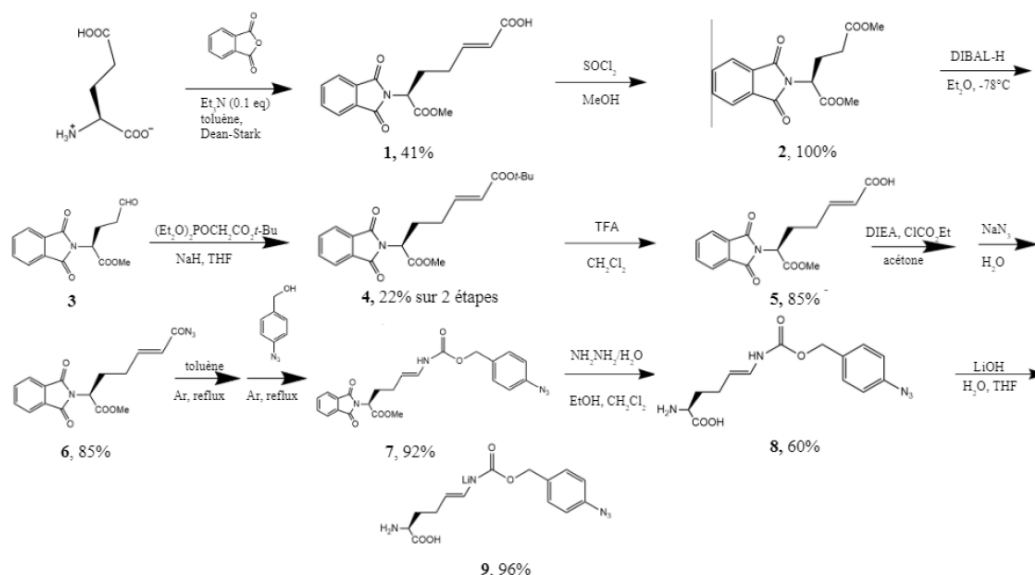


Figure 130 : Schéma générale de la synthèse de l'AcdK

Les neuf étapes nécessaires à la synthèse de l'AcdK sont présentées dans le schéma ci-dessus. Les étapes clés de cette synthèse sont une protection de l'amine α de l'acide L-glutamique, utilisé comme produit commercial de départ, pour ensuite orienter sélectivement une étape de réduction au DIBAL-H du δ COOH par rapport au α COOH qui sera préalablement estérifié. Puis une élongation de la chaîne latérale est réalisée par l'utilisation d'un phosphonate présentant une fonction acide carboxylique masquée par un groupement tert-butyle (composé 4, Figure 130). Après déprotection sélective, cette fonction acide est soumise à un réarrangement de Curtius en utilisant l'alcool 4-azidobenzyle pour piéger l'isocyanate intermédiaire (intermédiaire 6', Figure 130). L'acide aminé non naturel cible, porteur d'une fonction aldéhyde masquée sous la forme d'énamine, est finalement obtenu après élimination des groupements protecteurs. La synthèse des composés a ensuite été validée par RMN.

Comme montré dans la Figure 131 ci-dessous, le groupement 4-azidobenzoyloxycarbonyl au sein de l'AcdK assure le double rôle de groupement protecteur et de gâchette réactionnelle. La réduction de la fonction azoture par réaction de Staudinger conduira à la formation d'une énamine par effet mésomère. Cette énamine n'étant pas stable dans l'eau, elle va pouvoir s'hydrolyser afin de donner un δ aldéhyde qui pourra être disponible pour une réaction avec le sucre ligand.

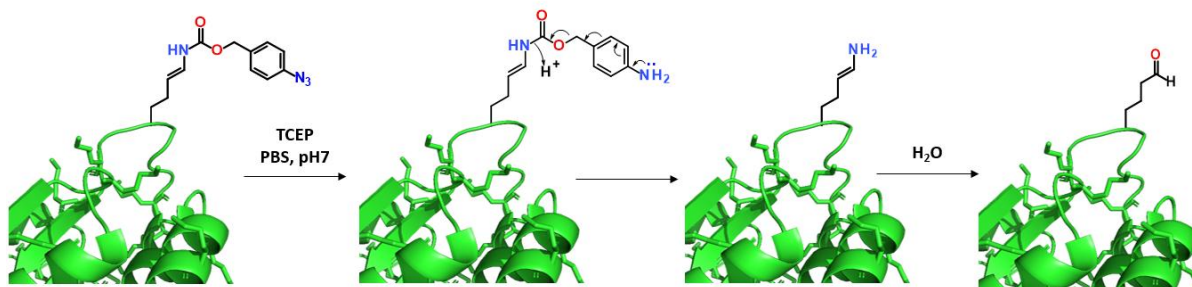


Figure 131 : Schéma représentant la déprotection de la fonction aldéhyde par action du TCEP sur le groupement 4-azidobenzoyloxycarbonyl

II.2 Choix des protéines cibles à modifier avec l'AcidK

Le but de ce projet est d'améliorer l'affinité de protéines reconnaissant les sucres à l'aide d'un acide aminé non naturel (ici AcidK). Plusieurs protéines modèles ont été choisies selon les critères suivants :

-la production de la protéine cible doit être décrite chez *E.coli* afin de limiter les mises au points nécessaires

-la structure 3D doit être disponible afin d'identifier un potentiel site d'incorporation de l'AcidK qui pourrait permettre une interaction avec le ligand en formant un lien covalent. Au niveau d'un site d'incorporation potentiel, une mutation pour un acide aminé naturel présentant une structure proche du NAA une fois déprotégé (ici glutamate) a été générée à l'aide de la fonction «*Mutagenesis* » sur PyMol afin d'analyser visuellement la validité de ce choix.

-la structure 3D de la protéine doit être le moins perturbée possible par une mutation au niveau du site d'incorporation de l'AcidK choisi. Cela a été vérifié à l'aide d'une expérience « d'alanine scanning » réalisée *in silico* à l'aide du logiciel YASARA. Le logiciel YASARA permet de générer *in silico* des mutations alanine à chaque position de la protéine et de calculer leur impact sur la stabilité de la structure globale de la protéine.

Selon ces critères, trois protéines cibles ont été sélectionnées :

-Le CBM40 : c'était une protéine déjà étudiée au laboratoire dans le cadre d'un autre projet. Il s'agit du Carbohydre Binding Module de l'enzyme NanA de *S. pneumoniae*. Cette protéine reconnaît l'acide sialique qui possède deux fonctions hydroxyles en position cis. Celles-ci pourraient former un lien covalent avec l'aldéhyde porté par l'AcidK. La position cible est la phénylalanine 48.

-Le CBM32 : cette protéine correspond au Carbohydre Binding Module (sous-unité DD2) d'une chitosanase issue de *Paenibacillus sp.* IK-5. Elle reconnaît le chitosan qui possède une fonction amine qui pourrait former un lien covalent avec l'aldéhyde porté par l'AcidK. Ce gène a été commandé, grâce à GeneScript, déjà cloné au sein du plasmide pET11a. Deux positions cibles ont été choisies : le glutamate 14 et la tyrosine 36.

-L'AAL-S2 : cette protéine est une sous-unité monomérique de la protéine *Aleuria aurantia* lectin (AAL), elle a été produite pour la première fois par Olausson et al. et reconnaît le fucose, dont deux fonctions hydroxyles (OH en positions 3 et 4 du fucose) pourraient former un lien covalent avec l'aldéhyde porté par l'AcidK (Olausson et al. 2011). Ce gène a été commandé, grâce à GeneScript, déjà cloné au sein du plasmide pET11a. La position cible choisie est le glutamate 89.

II.3 Incorporation de l'AcidK

II.3.1 Protéines cibles : génération des mutants

Afin d'incorporer l'acide aminé non naturel à une position donnée dans la protéine, un codon stop TAG doit être généré dans le gène codant pour la protéine cible. De plus, afin de s'assurer que l'effet potentiel de l'acide aminé AcidK sur l'affinité est bien dû à la fonction aldéhyde et non à un changement de conformation, une mutation Glutamate est également réalisée au niveau des positions choisies. En effet, cet acide aminé naturel est structurellement proche de l'AcidK.

Ces mutations ont été générées avec une technique de mutagenèse par amplification par PCR décrite ci-dessous à partir des plasmides pET11a portant les gènes des protéines d'intérêt avec l'étiquette histidine en C-terminal commandés sur GeneScript. Le plasmide à modifier est amplifié par PCR (PCR (Phusion (Thermo Fisher F-530S), dénaturation initiale 98°C pendant 3min, dénaturation 98°C pendant 30sec, hybridation 58°C pendant 30sec, élongation 72°C pendant 1min, 5cycles, ajout des primers au bout de 5 cycles, puis dénaturation 98°C pendant 30sec, hybridation 58°C pendant 30sec, élongation 72°C pendant 1min, 25 cycles élongation finale 72°C pendant 10min, en utilisant les amorces n°45-56). Puis, le plasmide parental non modifié est éliminé par traitement

avec l'enzyme *DpnI* (ThermoFisher, ER1701, 10U par réaction, incubation pendant 1h30 à 37°C, puis inactivation par 20min à 80°C). Enfin, un dixième de la réaction de mutagenèse (soit 5µL) est transformé dans des bactéries DH5α ou XL1b compétentes. Les modifications sont ensuite vérifiées par séquençage comme décrit ci-dessus avec les amorces T7P et T7T (n°3 et 4 en annexe).

II.3.2 Mutagenèse pour obtenir la synthétase spécifique à l'AcdK

L'incorporation d'acide aminé non naturel nécessite une synthétase spécifique pour l'acide aminé non naturel à insérer. Cela consiste le plus souvent à des mutations au sein du site actif de cette enzyme. La synthétase spécifique pour l'incorporation de l'AcdK a été décrite par Wang *et al.* (Z. A. Wang *et al.* 2017) il s'agit de la pyrrolysyl-ARNt-synthétase de *M. mazei* portant les mutations L309T, C348G et Y384F. Ces mutations ont donc été générées par amplification par PCR à partir du gène de l'aaRS de *M. mazei* sauvage, selon la méthode décrite précédemment (section II.3.1) à l'aide des amorces n°57 à 62. Puis, les mutations ont été vérifiées par séquençage selon le protocole décrit chapitre 4, section II.3 à l'aide des amorces n°13 et 14.

II.3.3 Production des protéines sauvages(WT) et mutants Glutamate (Glu)

La production des protéines recombinantes sauvages (WT) et des mutants Glutamate (Glu) s'effectue au sein de la souche BL21(DE3) transformée avec le plasmide pET11a (conférant une résistance à l'ampicilline) portant les gènes codant pour les différentes formes de ces protéines. Les séquences plasmidiques sont présentées en Annexe 2 et 3.

A partir d'une colonie bactérienne, une préculture est réalisée dans 5mL de milieu LB avec antibiotique de sélection (ampicilline 100µg/mL) et incubée sur la nuit à 37°C sous agitation. Le lendemain, la préculture est diluée au centième dans du milieu auto-inductif (FORMEDIUM, n° AIMLB0210) avec antibiotique puis la culture est incubée 24h à 37°C sous agitation.

II.3.4 Production des protéines mutées avec l'AcdK

La production s'effectue également au sein de la souche BL21(DE3) pour les protéines modifiées avec l'acide aminé non naturel AcdK. Au sein de la souche d'expression, deux plasmides ont été co-transformés : un plasmide d'expression portant la séquence de la protéine cible mutée avec un TAG au niveau du site d'incorporation (pET11a) et un plasmide portant les éléments nécessaires à l'incorporation de l'acide aminé non naturel c'est-à-dire l'ARNt et l'aaRS orthogonaux (pEVOL-AcdK, Annexe 2 et 3).

A partir d'une colonie bactérienne, une préculture est réalisée dans 5mL de milieu LB en présence d'antibiotiques de sélection (ampicilline 100µg/mL et chloramphénicol 30µg/mL) et incubée sur la nuit à 37°C sous agitation. Le lendemain, la préculture est diluée au centième dans du milieu auto-inductif (FORMEDIUM, n°AIMLB0210) avec arabinose à 0.02% final, l'AcdK est ajouté goutte à goutte sous agitation jusqu'à une concentration finale de 0,5mM puis la culture est incubée 24h à 37°C sous agitation.

II.3.5 Vérification de l'incorporation de l'AcdK dans une protéine

Des cultures sont réalisées selon le protocole précédent, avec ou sans acide aminé non naturel et son incorporation est vérifiée par une analyse SDS-PAGE.

Les échantillons composés de 10µL de culture avec 5µL d'eau et 5µL de tampon de charge (bleu de bromophénol, SDS, et βmercaptoéthanol) sont chauffés 5min à 100°C avant d'être déposés sur gel d'acrylamide 12%.

Pour une analyse des protéines totales, le gel de SDS-PAGE est coloré dans du Bleu de Coomassie pendant 1h puis décoloré pendant 1h dans de l'éthanol 20%.

Une analyse des protéines d'intérêt en Westernblot, est réalisée selon le protocole suivant. Le gel est transféré sur une membrane de nitrocellulose (PALL, BioTrace 66485) à l'aide d'un appareil de transfert (Thermo scientifique, Pierce Power Blotter), suivant le protocole de transfert des protéines de taille mixte (7min, 1,3A, 25V). La membrane est ensuite incubée 2h dans du TBS-Tween 0.1%-Lait 5% pour limiter les interactions non spécifiques puis lavée dans du TBS-Tween 0,1% avant une incubation sur la nuit dans l'anticorps primaire souris anti-étiquette Histidine (anticorps monoclonal sigma H1029, concentration 1/1000, TBS-Tween 0,1%-BSA-5%). La membrane est ensuite lavée 3 fois avec du TBS-Tween 0,1% pendant 15min. La membrane est ensuite incubée dans l'anticorps secondaire (anticorps chèvre anti souris couplé à un fluorophore Alexa 680 (Invitrogen A21058) à une dilution 1/10 000, TBS-Tween 0,1%-BSA-5%) pendant 1h à température ambiante. Enfin, après 3 lavages dans du TBS-Tween 0,1% durant 15min, la membrane est lue à l'aide de l'appareil de lecture (Odyssey CLX, LI-COR).

II.3.6 Purification des protéines

Les protéines recombinantes qui ont été produites lors de ce projet possèdent une étiquette histidine qui a été utilisée afin de les purifier sur colonne de Nickel selon le protocole suivant. La culture a été centrifugée à 5000g pendant 10min, puis le culot a été resuspendu dans du tampon de lyse ($\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ 50mM, NaCl 150mM, pH8, 5mM imidazole, PMSF 0.2mM), et les cellules ont été lysées par sonication (5min, cycles 5sec-5sec, amplification 50%, appareil : Thermo Fischer, FB120-220). Le lysat a ensuite été centrifugé 30min à 10 000rpm à 4°C, puis le surnageant de lyse cellulaire a été filtré sur une membrane de 0,45 μm avant d'être incubé 1h avec des billes de Nickel (Macherey Nagel, Protino) conditionnées dans du tampon ($\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ 50mM, NaCl 150mM, pH8 imidazole 5mM). La purification a ensuite été effectuée à l'aide de différentes solutions contenant des concentrations d'imidazole croissantes (5mM à 300mM). Chaque fraction d'élution a été récupérée et pour chaque fraction un aliquot de 15 μL a été analysé sur gel SDS-PAGE 12%. Les fractions contenant la protéine d'intérêt pure ont ensuite été dialysées dans du tampon (PBS pour la protéine cible CBM40 et sodium acétate 50mM pH5 pour CBM32) à l'aide de membranes de dialyses (tailles des pores à 6-8000 Da) afin d'éliminer l'imidazole. Les fractions ont ensuite été utilisées pour une validation des échantillons par spectrométrie de masse suivie d'une analyse en dichroïsme circulaire ainsi que pour réaliser des mesures d'affinités.

II.3.7 Analyse des protéines par dichroïsme circulaire

Après purification, les protéines cibles CBM40 et CBM32 ont été analysées par dichroïsme circulaire, à l'aide de l'appareil Jasco-J810. Pour avoir un signal exploitable et reproductible, les protéines ont été diluées pour obtenir un signal aux alentours de -20/-30 à une longueur d'onde de 220 nm, soit une concentration de 0,25mg/mL pour CBM40 et 0,15mg/mL pour CBM32. Les mesures de concentrations ont été réalisées au Nanodrop en utilisant les valeurs caractéristiques des protéines (CBM40 : $\epsilon=20,340\text{cm}^{-1}\text{M}^{-1}$ et $M=23,677\text{kDa}$; CBM32 : $\epsilon=40,91\text{cm}^{-1}\text{M}^{-1}$ et $M=15,3\text{kDa}$). Les signaux de différentes conditions ont ensuite été comparés par superposition des courbes obtenues.

II.3.8 Mesures d'affinités par méthode BLITZ

Afin de déterminer l'effet de l'incorporation de l'AcidK au sein des protéines cibles, des mesures d'affinité ont été réalisées à l'aide de la technologie BLITZTM (Bio-Layer Interferometry,

ForteBio). Les ligands sont donc fixés à l'aide d'une liaison streptavidine biotine sur un tip, puis celui-ci est mis en contact avec la protéine pour une phase d'association, puis dans du tampon pour une phase de dissociation. Les cinétiques d'association et de dissociation sont suivies en temps réel, ce qui permet d'en déduire une affinité. Ces phases sont déterminées grâce à des mesures basées sur le principe de l'interférométrie. C'est-à-dire une modification de l'onde lumineuse (déphasage, déviation) suite à son passage à travers l'échantillon.

-biotinylation des sucres :

Afin de fixer les sucres ligands (acide sialique pour le CBM40 et chitosan pour le CBM32), sur les *tips*, ceux-ci sont biotinylés selon le protocole suivant. Deux cents milligramme de NaCNBH₃ sont purifiés à l'aide de THF sous agitation pendant 20min afin de précipiter des contaminants. Le mélange est ensuite filtré puis les solvants sont évaporés afin de récupérer le NaCNBH₃, sous forme de poudre blanche pure. Dix-huit milligrammes de biotine sont ensuite dissouts dans 420µL de DMSO, auquel sont ajoutés 180µL d'acide acétique glacial. Ce mélange est ensuite ajouté à l'agent réducteur NaCNBH₃ puis chauffé 1 minute à 65°C. Enfin un milligramme de sucre (acide sialique ou chitosan) est mis au contact de la biotine en proportion molaire équivalente au sein d'un mélange acide acétique glacial/DMSO (2,3/1) pendant 2h à 65°C. Le mélange est ensuite mis au contact d'une membrane (Whatman QM-A-Quartz MicrofiberFilter, préalablement rincée à l'eau et à l'acétonitrile) pendant 15min à température ambiante, avant d'être lavée avec un mélange eau-acétonitrile (4% d'eau dans acétonitrile) afin d'éliminer la biotine en excès. Le sucre biotinylé est ensuite élué à l'aide de 4mL d'eau. Cette solution est ensuite lyophilisée puis le composé est resuspendu dans un volume réduit (200µL). La biotinylation des sucres est ensuite vérifiée par chromatographie sur couche mince (CCM, éluant pour l'acide sialique : acide acétique, butanol, eau (2/5/2), éluant pour le chitosan : ammoniaque, acétate d'éthyle et eau (5/5/2) et révélateur : vanilline). Les sucres biotinylés sont ensuite stockés à -20°C avant leur utilisation pour les mesures d'affinité par BLItz.

-mesures d'affinités :

Les *tips* sont d'abord hydratés pendant 10min au sein du tampon (PBS pour CBM40, et 50mM sodium acétate pH 5 pour CBM32).

-Chargement du sucre ligand : Après une étape de 30sec dans du tampon afin de définir une ligne de base, le tip est mis en contact avec le sucre ligand pendant 300sec pour permettre son association, puis est transféré dans la solution tampon pendant 300sec pour rincer les sucres non liés.

-Mesure d'affinité : Après une étape de 30sec dans du tampon afin de définir une ligne de base, le tip préalablement chargé par le ligand est mis en contact avec la protéine portant l'AcidK sous sa forme protégée pendant 150sec (étape d'association), puis transféré dans la solution tampon pendant 150sec afin de procéder à l'étape de dissociation. Enfin, afin de réutiliser le tip pour une autre expérience la protéine non dissociée est décrochée en transférant le tip dans un tampon Glycine 100mM à pH 2 pendant 20sec, avant d'être rééquilibré dans le tampon de référence (PBS pour CBM40 et 50mM sodium acétate pH 5 pour CBM32)

Les mesures d'affinités sont ensuite réitérées après exposition de la protéine avec 5mM de TCEP pendant 2h à température ambiante afin de déprotéger la fonction aldéhyde portée par l'AcidK (section II.1)

III. Utilisation de la boronate phénylalanine

La synthèse de l'AcidK étant assez complexe, une stratégie basée sur la réactivité du boronate a été développée en parallèle. En effet, la fonction boronate peut réagir en milieu basique avec des

cis-diols (potentiellement présents sur des sucres ligands) afin de former des liaisons covalentes. De plus, l'acide aminé phényl-boronate est un acide aminé commercialement disponible qui a déjà été incorporé précédemment dans des protéines et pour lequel une aaRS adaptée est donc décrite (Chapitre 5, section IV.2.3).

III.1 Choix des cibles

Le but de ce projet est d'améliorer l'affinité de protéines reconnaissant les sucres à l'aide d'un acide aminé non naturel (ici le phényl Boronate). Les cibles ont été choisies selon la même démarche que décrite précédemment (section II.2) selon ces critères :

-La GH98Sp (Glycoside Hydrolase 98 issue de *S. pneumoniae*) : cette protéine reconnaît le tetrasaccharide Lewis-Y (α -Fuc-(1→2)- β -Gal-(1→4)-[α -Fuc-(1→3)]-GlcNAc) qui possède deux fonctions diols (OH en 3 et en 4 du fucose) qui pourraient former un lien covalent avec l'aldéhyde porté par la phénylalanine boronate. Ce gène a été commandé, grâce à GeneScript, déjà cloné au sein du plasmide pET11a. La position cible pour cette protéine est l'histidine 479.

-Le CLC-1 (C-type lectine domain 1 member B) : cette protéine reconnaît l'antigène Tn sialilé (α Neu5Ac-(2-6)- α GalNAc-O-Ser/Thr) dont les deux fonctions diols portées par l'acide sialique pourraient former un lien covalent avec l'aldéhyde porté par la phénylalanine boronate. Ce gène a été commandé, grâce à GeneScript, déjà cloné au sein du plasmide pET11a. La position cible pour cette protéine est l'histidine 119.

-NagBb (Alpha-N-acetylgalactosaminidase) : cette protéine reconnaît le Galactosamine (GalNAc) qui possède deux fonctions diols (OH en 3 et en 4) qui pourraient former un lien covalent avec l'aldéhyde porté par la phénylalanine boronate. Ce gène a été commandé, grâce à GeneScript, déjà cloné au sein du plasmide pET11a. La position cible pour cette protéine est la tyrosine 329.

III.2 Incorporation du boronate au sein de protéines

III.2.1 Souches et plasmides utilisés

La production a été effectuée au sein de la souche BL21(DE3) avec les plasmides pET11a portant les gènes codant pour les protéines cibles et pDULE.2 ou pSB4-Bpa (plasmide fourni par l'équipe de Skerra (Schiefner et al. 2018)) portant les éléments orthogonaux nécessaires à l'incorporation de la phénylalanine boronate (aaRS spécifique et ARNt).

Clonage pour obtenir la synthétase spécifique pour l'incorporation de la phénylalanine boronate au sein du plasmide pDULE.2 :

La synthétase spécifique pour l'incorporation du phényl-boronate a été décrite par Schiefner et al., il s'agit de la Tyrosyl-ARNt-synthetase issue de *Methanocaldococcus jannaschii* portant les mutations Y32S, L65A, H70M, D158S, and L162E (Schiefner et al. 2018).

Le plasmide pDULE.2 contenant la Tyrosyl-ARNt-synthetase issue de *Methanocaldococcus jannaschii* portant des mutations spécifiques pour l'incorporation de l'azido-phénylalanine (Y32L, L65V, F108W, Q109M, D158G et I159P), nous a été transmis par Ryan Mehl de l'Université d'Oregon (Miyake-Stoner et al. 2009). Afin d'obtenir la séquence finale de l'aaRS efficace pour l'incorporation de la Bpa, nous avons commandé la partie de la séquence portant les mutations (E28-V166), puis nous avons linéarisé le vecteur pDULE.2 à l'aide des amorces n°63 et 64 par PCR (Phusion (Thermo Fisher F-530S), 1) dénaturation initiale 98°C pendant 10min, les étapes 2 à 4 sont répétées 18 fois : 2) dénaturation 98°C pendant 30sec, 3) hybridation 58°C pendant 30sec, 4) élongation 72°C pendant 7min, puis 5) élongation finale 72°C pendant 10min). Le plasmide matrice non linéarisé est ensuite dégradé par traitement avec l'enzyme *DpnI* (ThermoFisher, ER1701, 10U par réaction, incubation

pendant 1h30 à 37°C, puis inactivation par 20min à 80°C). Le vecteur est ensuite déposé sur gel d'agarose 1%, puis purifié à l'aide du kit d'extraction de gel (QIAGEN, 28115).

Enfin, le clonage de l'insert et le vecteur, préparés de manière à présenter des parties chevauchantes, est directement réalisé à l'aide du kit de Gibson Assembly (DNA SyntheticGenomics, GA1100). Le rapport molaire entre le vecteur et l'insert utilisé est de 1/2, et la réaction est incubée à 50°C pendant 30min, avant de transformer 2µL (1/10^{ème} de la réaction) au sein d'XL1b compétentes (méthode TSS). Le clonage a été vérifié par séquençage selon le protocole décrit chapitre 4, section II.3 à l'aide des amorces n°66 et 67 (annexe 4).

Modification du promoteur en amont de l'aaRS spécifique pour la phénylalanine boronate :

Etant donné que l'incorporation de la phénylalanine boronate n'avait pu être observée avec l'utilisation du plasmide pDULE.2-Bpa, nous avons essayé de changer le promoteur constitutif alors en place (Ipp) par un promoteur inductible à l'arabinose.

Pour cela, le promoteur inductible à l'arabinose a été amplifié à partir d'un plasmide pBAD (disponible au laboratoire) à l'aide des amorces n°68 et 69 par PCR (Phusion (Thermo Fisher F-530S) : 1) dénaturation initiale 98°C pendant 10min, les étapes 2 à 4 sont répétées 18 fois : 2) dénaturation 98°C pendant 30sec, 3) hybridation 58°C pendant 30sec, 4) élongation 72°C pendant 1min, puis 5) élongation finale 72°C pendant 1min). Le vecteur a également été linéarisé par PCR (Phusion (Thermo Fisher F-530S) : 1) dénaturation initiale 98°C pendant 10min, les étapes 2 à 4 sont répétées 18 fois : 2) dénaturation 98°C pendant 30sec, 3) hybridation 58°C pendant 30sec, 4) élongation 72°C pendant 7min, puis 5) élongation finale 72°C pendant 10min) à l'aide des amorces n°70 et 71. Les deux fragments ont ensuite été déposés sur gel d'agarose 1% et purifiés à l'aide du kit d'extraction de gel (QIAGEN, 28115). Le clonage a ensuite été réalisé à l'aide du kit de Gibson Assembly (DNA SyntheticGenomics, GA1100). Le rapport molaire entre le vecteur et l'insert utilisé est de 1/2, et la réaction est incubée à 50°C pendant 30min, avant de transformer 2µL (1/10^{ème} de la réaction) au sein d'XL1b compétentes (méthode TSS). Le clonage a été vérifié par séquençage selon le protocole décrit chapitre 4, section II.3 à l'aide des amorces n°66 et 63.

Mutagenèse au sein des gènes des protéines cibles :

Afin d'incorporer l'acide aminé non naturel à une position donnée dans la protéine, un codon stop TAG doit être généré dans le gène codant à la position choisie, soit H479TAG pour GH98Sp, H119TAG pour CLC1, et Y329TAG pour NagBb. Ces mutations sont réalisées avec une technique de mutagenèse par amplification par PCR décrite section II.3.2 à l'aide des amorces n°72-77 (annexe 4).

III.2.2 Production des protéines

La production des protéines sauvages s'est réalisée selon le protocole décrit section II.3.3.

Pour les protéines contenant la phénylalanine boronate, la production s'est déroulée comme suit. La production s'effectue au sein de la souche BL21(DE3) pour les protéines modifiées avec l'acide aminé non naturel Phénylalanine-boronate. Au sein de la souche d'expression, deux plasmides ont été co-transformés : un plasmide d'expression portant la séquence de la protéine cible mutée avec un TAG au niveau du site d'incorporation et une étiquette de 6 histidine en C-terminal (pET11a) et un plasmide portant les éléments nécessaires à l'incorporation de l'acide aminé non naturel (pDULE.2-Bpa ou pSB4-Bpa, Annexe 2 et 3).

A partir d'une colonie bactérienne, une préculture est réalisée avec 5mL de milieu LB en présence d'antibiotiques de sélection (ampicilline 100µg/mL et Spectinomycine 50µg/mL pour pDULE.2 ou chloramphénicol 30µg/mL pour pSB4-Bpa) et incubée sur la nuit à 37°C sous agitation. Le

lendemain, la préculture est diluée au centième dans du milieu auto-inductif (FORMEDIUM, n° AIMLB0210) avec arabinose à 0.02% final, ainsi que de la Phénylalanine-boronate à une concentration de 1mM final puis la culture est incubée 24h à 37°C sous agitation.

III-2-3) Vérification de l'incorporation de la phénylalanine boronate

L'incorporation de la phénylalanine boronate est vérifiée en suivant le protocole décrit section II.3.5.

Conclusion Générale

Ce travail de thèse a permis le développement d'un nouvel axe au sein du laboratoire : la biologie de synthèse et plus particulièrement l'incorporation d'acides aminés non naturels dans des protéines impliquées en glycobiochimie. Cette technologie repose sur la caractérisation et le transfert dans un organisme modèle d'aminocyl ARNt synthétases et d'ARNt issus d'organismes orthogonaux afin d'installer une nouvelle voie indépendante permettant l'incorporation ciblée d'un acide aminé non naturel. Le transfert de ces "biobriques" au sein de l'organisme de production permet alors de générer des protéines artificielles n'existant pas dans le vivant. Par cette technologie, il est alors possible de produire des protéines aux caractéristiques uniques.

Au cours de ce travail, nous avons pu dans un premier temps mettre au point la technologie d'incorporation des acides aminés non naturels au sein du laboratoire puis appliquer cette technologie à deux projets de recherche distincts.

- Le premier axe de recherche consiste en la génération de vaccins glycoconjugués homogènes dont la position de greffage est maîtrisée. L'utilisation d'acides aminés non naturels apporte un potentiel très important pour ce type d'applications. Cependant, cette technologie reste encore sous-exploitée au sein de ce domaine. Nous l'avons ici exploitée afin de générer trois conjugués homogènes monovalents ainsi qu'un conjugué homogène trivalent dont les positions sont également ciblées. Ces conjugués ont ensuite été injectés au sein de souris afin d'en tester la réponse immunitaire. Ces travaux ont permis de fournir un outil efficace afin d'étudier l'impact de la position de greffage des oligosaccharides au sein de la protéine porteuse, ainsi que l'impact du ratio protéine / oligosaccharide.
- Le deuxième axe de recherche étudié consiste en l'ingénierie de lectines afin d'en augmenter les affinités. Les acides aminés non naturels peuvent alors apporter des réactivités uniques mais leur utilisation lors d'ingénierie de sites de reconnaissance reste encore complexe. Au cours de ce travail, nous avons mis au point une méthode de sélection afin d'incorporer un acide aminé non naturel original. Nous avons également incorporé un acide aminé non naturel connu portant une fonction aldéhyde activable *in situ* au sein de deux cibles protéiques dont nous avons pu tester les affinités et enfin nous avons également incorporé un autre acide aminé non naturel portant une fonction boronate. Ces acides aminés non naturels ont un fort potentiel pour aller plus loin lors de l'ingénierie de lectines et cette méthodologie bénéficiera des futures avancées techniques.

Au cours de ce travail, nous avons donc abordé l'essentiel des grandes problématiques de cette technologie : l'optimisation par l'étude des différents paramètres permettant l'incorporation, la production de protéine avec un acide aminé non naturel à une ou plusieurs positions, l'incorporation d'acides aminés non naturels à la surface de protéine ou au sein de sites de reconnaissance et enfin l'incorporation de nouveaux acides aminés non naturels avec la mise en place de méthodes de sélections.

Globalement, la technologie d'incorporation des acides aminés non naturels de manière site-spécifique permet de modifier les protéines avec une finesse inégalée afin d'en ajouter des réactivités uniques. Ainsi, contrairement à des méthodes chimiques moins ciblées, cette méthodologie permet de contrôler la position de ces modifications avec moins de limitations quant à la position choisie. De plus, les réactivités disponibles sont de plus en plus variées au fur et à mesure que de nouveaux acides aminés non naturels sont incorporés au sein de protéines. Cette technologie

a donc un potentiel d'applications très important au sein de différents domaines scientifiques et pourrait bien modifier radicalement notre vision de l'ingénierie protéique en étendant radicalement ses possibilités.

Enfin, nous pouvons souligner que cette thèse a été financée dans le cadre des "Paris scientifiques" de la Région qui permettent la mise en place de nouveaux axes de recherche et la réalisation de projets de rupture. Comme tout "Pari", il s'agit d'un risque plus ou moins maîtrisé ce qui peut expliquer les difficultés pour atteindre les objectifs finaux. Mais c'est également l'occasion d'oser travailler sur des projets plus ambitieux qui peuvent apporter de nouvelles expertises. Au travers de cette thèse, nous avons ainsi implémenté cette nouvelle technologie à l'UFIP, qui pourra donc servir aux différents projets scientifiques liés aux problématiques vaccinales ou de modification des protéines.

Références

- Aanensen, David M., Angeliki Mavroidi, Stephen D. Bentley, Peter R. Reeves, and Brian G. Spratt. 2007. "Predicted Functions and Linkage Specificities of the Products of the Streptococcus Pneumoniae Capsular Biosynthetic Loci." *Journal of Bacteriology* 189 (21): 7856–76. <https://doi.org/10.1128/JB.00837-07>.
- Adam, Jan, Martina Pokorná, Charles Sabin, Edward P. Mitchell, Anne Imberty, and Michaela Wimmerová. 2007. "Engineering of PA-IIL Lectin from Pseudomonas Aeruginosa - Unravelling the Role of the Specificity Loop for Sugar Preference." *BMC Structural Biology*. <https://doi.org/10.1186/1472-6807-7-36>.
- Adamo, Roberto, Qi Ying Hu, Antonella Torosantucci, Stefano Crotti, Giulia Brogioni, Martin Allan, Paola Chiani, et al. 2014. "Deciphering the Structure-Immunogenicity Relationship of Anti-Candida Glycoconjugate Vaccines." *Chemical Science* 5 (11): 4302–11. <https://doi.org/10.1039/c4sc01361a>.
- Agard, Nicholas J., Jennifer A. Prescher, and Carolyn R. Bertozzi. 2004. "A Strain-Promoted [3 + 2] Azide-Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems." *Journal of the American Chemical Society* 126 (46): 15046–47. <https://doi.org/10.1021/ja044996f>.
- Agou, Fabrice, Sophie Quevillon, Pierre Kerjan, and Marc Mirande. 1998. "Switching the Amino Acid Specificity of an Aminoacyl-TRNA Synthetase." *Biochemistry* 37 (32): 11309–14. <https://doi.org/10.1021/bi9805590>.
- Agris, Paul F., Emily R. Eruysal, Amithi Narendran, Ville Y.P. Väre, Sweta Vangaveti, and Srivathsan V. Ranganathan. 2018. "Celebrating Wobble Decoding: Half a Century and Still Much Is New." *RNA Biology*. Taylor and Francis Inc. <https://doi.org/10.1080/15476286.2017.1356562>.
- Alexandre Ambrogelly, Dieter Söll, Osamu Nureki, Shigeyuki Yokoyama, and Michael Ibba. 2000. "Class I Lysyl-TRNA Synthetases - Madame Curie Bioscience Database - NCBI Bookshelf." 2000. <https://www.ncbi.nlm.nih.gov/books/NBK6444/>.
- Alfonta, Lital, Zhiwen Zhang, Sean Uryu, Joseph A. Loo, and Peter G. Schultz. 2003. "Site-Specific Incorporation of a Redox-Active Amino Acid into Proteins." *Journal of the American Chemical Society* 125 (48): 14662–63. <https://doi.org/10.1021/ja038242x>.
- Alford, Rebecca F., Andrew Leaver-Fay, Jeliasko R. Jeliaskov, Matthew J. O'Meara, Frank P. DiMaio, Hahnbeom Park, Maxim V. Shapovalov, et al. 2017. "The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design." *Journal of Chemical Theory and Computation* 13 (6): 3031–48. <https://doi.org/10.1021/acs.jctc.7b00125>.
- Ali, M K, H Hayashi, S Karita, M Goto, T Kimura, K Sakka, and K Ohmiya. 2001. "Importance of the Carbohydrate-Binding Module of Clostridium Stercorarium Xyn10B to Xylan Hydrolysis." *Bioscience, Biotechnology, and Biochemistry* 65 (1): 41–47. <http://www.ncbi.nlm.nih.gov/pubmed/11272844>.
- Alpe, Mia, and Stefan Oscarson. 2002. "Synthesis of Oligosaccharides Corresponding to Streptococcus Pneumoniae Type 9 Capsular Polysaccharide Structures." *Carbohydrate Research* 337 (19): 1715–22. [https://doi.org/10.1016/s0008-6215\(02\)00263-x](https://doi.org/10.1016/s0008-6215(02)00263-x).
- Amano, Koh, Hiroe Katayama, Akihiro Saito, Akikazu Ando, and Yoshiho Nagata. 2012. "Aleuria Aurantia Lectin Exhibits Antifungal Activity against Mucor Racemosus." *Bioscience, Biotechnology and Biochemistry* 76 (5): 967–70. <https://doi.org/10.1271/bbb.110982>.
- Anderson, Porter W, Ronald John Eby, and University of Rochester. 1994. Immunogenic conjugates of Streptococcus pneumoniae capsular polymer and toxin or in toxiad, issued 1994.
- Anderson, R, and C Feldman. 2011. "Key Virulence Factors of Streptococcus Pneumoniae and Non-Typeable Haemophilus Influenzae : Roles in Host Defence and Immunisation ." *Southern African Journal of Epidemiology and Infection* 26 (1): 6–12. <https://doi.org/10.1080/10158782.2011.11441412>.
- Aniansson, Gustaf, Bernt Alm, Bengt Andersson, Peter Larsson, Olle Nylén, Hans Peterson, Peter Rignér, Marianne Svanborg, and Catharina Svanborg. 1992. "Nasopharyngeal Colonization during the First Year of Life." *Journal of Infectious Diseases* 165: S38–41. https://doi.org/10.1093/infdis/165-Supplement_1-S38.
- Appel, Mason J., and Carolyn R. Bertozzi. 2015. "Formylglycine, a Post-Translationally Generated Residue with Unique Catalytic Capabilities and Biotechnology Applications." *ACS Chemical Biology* 10 (1): 72–84. <https://doi.org/10.1021/cb500897w>.

- Appelbaum, Peter C. 2002. "Resistance among Streptococcus Pneumoniae: Implications for Drug Selection." *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America* 34 (12): 1613–20. <https://doi.org/10.1086/340400>.
- Appella, Daniel H. 2009. "Non-Natural Nucleic Acids for Synthetic Biology." *Current Opinion in Chemical Biology*. <https://doi.org/10.1016/j.cbpa.2009.09.030>.
- Arango, R, E Rodriguez-Arango, R Adar, D Belenky, F G Loontjens, S Rozenblatt, and N Sharon. 1993. "Modification by Site-Directed Mutagenesis of the Specificity of Erythrina Corallodendron Lectin for Galactose Derivatives with Bulky Substituents at C-2." *FEBS Letters* 330 (2): 133–36. [https://doi.org/10.1016/0014-5793\(93\)80258-v](https://doi.org/10.1016/0014-5793(93)80258-v).
- Archontis, Georgios, Thomas Simonson, Dino Moras, and Martin Karplus. 1998. "Specific Amino Acid Recognition by Aspartyl-TRNA Synthetase Studied by Free Energy Simulations." *Journal of Molecular Biology* 275 (5): 823–46. <https://doi.org/10.1006/jmbi.1997.1470>.
- Arnaud, Julie, Aymeric Audfray, and Anne Imberty. 2013. "Binding Sugars: From Natural Lectins to Synthetic Receptors and Engineered Neolectins." *Chemical Society Reviews*. <https://doi.org/10.1039/c2cs35435g>.
- Auburtin, M., and J. F. Timsit. 2001. "Meningites à Pneumocoque: Actualités, Perspectives." *Reanimation Urgences* 10 (3): 291–301.
- Austin, Stuart, and Kurt Nordström. 1990. "Partition-Mediated Incompatibility of Bacterial Plasmids." *Cell* 60 (3): 351–54. [https://doi.org/10.1016/0092-8674\(90\)90584-2](https://doi.org/10.1016/0092-8674(90)90584-2).
- Avci, Fikri Y., Xiangming Li, Moriya Tsuji, and Dennis L. Kasper. 2011. "A Mechanism for Glycoconjugate Vaccine Activation of the Adaptive Immune System and Its Implications for Vaccine Design." *Nature Medicine* 17 (12): 1602–9. <https://doi.org/10.1038/nm.2535>.
- Avery, O T, and M Heidelberger. 1925. "IMMUNOLOGICAL RELATIONSHIPS OF CELL CONSTITUENTS OF PNEUMOCOCCUS : SECOND PAPER." *The Journal of Experimental Medicine* 42 (3): 367–76. <https://doi.org/10.1084/jem.42.3.367>.
- Avery, Oswald T., and René Dubos. 1931. "The Protective Action of a Specific Enzyme against Type Iii Pneumococcus Infection in Mice." *Journal of Experimental Medicine* 54 (1): 73–88. <https://doi.org/10.1084/jem.54.1.73>.
- Ayyadurai, Niraikulam, Kanagavel Deepankumar, Nadarajan Saravanan Prabhu, Sungu Lee, and Hyungdon Yun. 2011. "A Facile and Efficient Method for the Incorporation of Multiple Unnatural Amino Acids into a Single Protein." *Chemical Communications* 47 (12): 3430–32. <https://doi.org/10.1039/c0cc04672h>.
- Backhaus, Erik, Stefan Berg, Rune Andersson, Gunilla Ockborn, Petter Malmström, Mats Dahl, Salmir Nasic, and Birger Trollfors. 2016. "Epidemiology of Invasive Pneumococcal Infections: Manifestations, Incidence and Case Fatality Rate Correlated to Age, Gender and Risk Factors." *BMC Infectious Diseases* 16 (1): 367. <https://doi.org/10.1186/s12879-016-1648-2>.
- Bakshani, Cassie R., Ana L. Morales-Garcia, Mike Althaus, Matthew D. Wilcox, Jeffrey P. Pearson, John C. Bythell, and J. Grant Burgess. 2018. "Evolutionary Conservation of the Antimicrobial Function of Mucus: A First Defence against Infection." *Npj Biofilms and Microbiomes*. Nature Publishing Group. <https://doi.org/10.1038/s41522-018-0057-2>.
- Barondes, S H, V Castronovo, D N Cooper, R D Cummings, K Drickamer, T Feizi, M A Gitt, J Hirabayashi, C Hughes, and K Kasai. 1994. "Galactins: A Family of Animal Beta-Galactoside-Binding Lectins." *Cell* 76 (4): 597–98. [https://doi.org/10.1016/0092-8674\(94\)90498-7](https://doi.org/10.1016/0092-8674(94)90498-7).
- Barret, Laurie-Anne. 2014. "Influence Des Tensioactifs Dans La Cristallisation Du Complexe Photosynthétique RC-LH1-PufX de Rhodobacter Blasticus."
- Barton, Christopher, J Calvin Kouokam, Amanda B Lasnik, Oded Foreman, Alexander Cambon, Guy Brock, David C Montefiori, et al. 2014. "Activity of and Effect of Subcutaneous Treatment with the Broad-Spectrum Antiviral Lectin Griffithsin in Two Laboratory Rodent Models." *Antimicrobial Agents and Chemotherapy* 58 (1): 120–27. <https://doi.org/10.1128/AAC.01407-13>.
- Becker, Jeffrey M., and Fred Naider. 2015. "Cross-Linking Strategies to Study Peptide Ligand-Receptor Interactions." In *Methods in Enzymology*, 556:527–47. Academic Press Inc. <https://doi.org/10.1016/bs.mie.2014.12.001>.
- Beene, Darren L, Kerry L Price, Henry A Lester, Dennis A Dougherty, and Sarah C R Lummis. 2004. "Tyrosine Residues That Control Binding and Gating in the 5-Hydroxytryptamine3 Receptor Revealed by Unnatural Amino Acid Mutagenesis." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 24 (41): 9097–9104. <https://doi.org/10.1523/JNEUROSCI.2429-04.2004>.
- Belický, Štefan, Jaroslav Katrlík, and Ján Tkáč. 2016. "Glycan and Lectin Biosensors." *Essays in Biochemistry* 60 (1): 37–47.

<https://doi.org/10.1042/EBC20150005>.

- Bentley, Stephen D., David M. Aanensen, Angeliki Mavroidi, David Saunders, Ester Rabinowitsch, Matthew Collins, Kathy Donohoe, et al. 2006. "Genetic Analysis of the Capsular Biosynthetic Locus from All 90 Pneumococcal Serotypes." *PLoS Genetics* 2 (3): 0262–69. <https://doi.org/10.1371/journal.pgen.0020031>.
- Bentley, Stephen, Lesley McGee, and Robert Breiman. 2016. "GPS Global Pneumococcal Sequencing Project." 2016. <https://www.pneumogen.net/gps/publications.html>.
- Beppu, Masatoshi, Tadao Terao, and Toshiaki Osawa. 1976. "Preparation of Monovalent Succinyl-Concanavalin A and Its Mitogenic Activity." *Journal of Biochemistry*. <https://doi.org/10.1093/oxfordjournals.jbchem.a131152>.
- Berg JM, Tymoczko JL, and Stryer L. 2002. *Section 29.2 : Aminoacyl-Transfer RNA Synthetases Read the Genetic Code - Biochemistry - NCBI Bookshelf*. <https://www.ncbi.nlm.nih.gov/books/NBK22356/#A4154>.
- Bergmann, Simone, and Sven Hammerschmidt. 2006. "Versatility of Pneumococcal Surface Proteins." *Microbiology*. <https://doi.org/10.1099/mic.0.28610-0>.
- Berti, Francesco, and Roberto Adamo. 2018. "Antimicrobial Glycoconjugate Vaccines: An Overview of Classic and Modern Approaches for Protein Modification." *Chemical Society Reviews*. Royal Society of Chemistry. <https://doi.org/10.1039/c8cs00495a>.
- Bhutia, Sujit K, Prashanta K Panda, Niharika Sinha, Prakash P Praharaj, Chandra S Bhol, Debasna P Panigrahi, Kewal K Mahapatra, et al. 2019. "Plant Lectins in Cancer Therapeutics: Targeting Apoptosis and Autophagy-Dependent Cell Death." *Pharmacological Research* 144 (June): 8–18. <https://doi.org/10.1016/j.phrs.2019.04.001>.
- Blomme, Bram, Christophe Van Steenkiste, Nico Callewaert, and Hans Van Vlierberghe. 2009. "Alteration of Protein Glycosylation in Liver Diseases." *Journal of Hepatology* 50 (3): 592–603. <https://doi.org/10.1016/j.jhep.2008.12.010>.
- Boelsen, Laura K., Eileen M. Dunne, Karen E. Lamb, Kathryn Bright, Yin Bun Cheung, Lisi Tikoduadua, Fiona M. Russell, E. Kim Mulholland, Paul V. Licciardi, and Catherine Satzke. 2015. "Long-Term Impact of Pneumococcal Polysaccharide Vaccination on Nasopharyngeal Carriage in Children Previously Vaccinated with Various Pneumococcal Conjugate Vaccine Regimes." *Vaccine* 33 (42): 5708–14. <https://doi.org/10.1016/j.vaccine.2015.07.059>.
- Bogaert, D., R. De Groot, and P. W.M. Hermans. 2004. "Streptococcus Pneumoniae Colonisation: The Key to Pneumococcal Disease." *Lancet Infectious Diseases*. [https://doi.org/10.1016/S1473-3099\(04\)00938-7](https://doi.org/10.1016/S1473-3099(04)00938-7).
- Bolam, D N, A Ciruela, S McQueen-Mason, P Simpson, M P Williamson, J E Rixon, A Boraston, G P Hazlewood, and H J Gilbert. 1998. "Pseudomonas Cellulose-Binding Domains Mediate Their Effects by Increasing Enzyme Substrate Proximity." *The Biochemical Journal* 331 (Pt 3) (May): 775–81. <https://doi.org/10.1042/bj3310775>.
- Boraston, Alisdair B., David N. Bolam, Harry J. Gilbert, and Gideon J. Davies. 2004. "Carbohydrate-Binding Modules: Fine-Tuning Polysaccharide Recognition." *Biochemical Journal*. <https://doi.org/10.1042/BJ20040892>.
- Boraston, Alisdair B., A. Louise Creagh, Md Mustafa Alam, Jeffery M. Kormos, Peter Tomme, Charles A. Haynes, R. Antony J. Warren, and Douglas G. Kilburn. 2001. "Binding Specificity and Thermodynamics of a Family 9 Carbohydrate-Binding Module from *Thermotoga Maritima* Xylanase 10A." *Biochemistry* 40 (21): 6240–47. <https://doi.org/10.1021/bi0101695>.
- Boraston, Alisdair B., Emily Kwan, Patrick Chiu, R. A J Warren, and Douglas G. Kilburn. 2003. "Recognition and Hydrolysis of Noncrystalline Cellulose." *Journal of Biological Chemistry* 278 (8): 6120–27. <https://doi.org/10.1074/jbc.M209554200>.
- Bornhöfft, Kim F., Tom Goldammer, Alexander Rebl, and Sebastian P. Galuska. 2018. "Siglecs: A Journey through the Evolution of Sialic Acid-Binding Immunoglobulin-Type Lectins." *Developmental and Comparative Immunology* 86: 219–31. <https://doi.org/10.1016/j.dci.2018.05.008>.
- Bromba, Caleb, Philippa Carrie, Jonathan K.W. Chui, and Thomas M. Fyles. 2009. "Phenyl Boronic Acid Complexes of Diols and Hydroxyacids." *Supramolecular Chemistry* 21 (1–2): 81–88. <https://doi.org/10.1080/10610270802527044>.
- Brook, Itzhak. 2011. "Microbiology of Sinusitis," no. 21. <https://doi.org/10.1513/pats.201006-038RN>.
- Brown, Rodney D., C. Fred Brewer, and Seymour H. Koenig. 1977. "Conformation States of Concanavalin A: Kinetics of Transitions Induced by Interaction with Mn^{2+} and Ca^{2+} ." *Biochemistry* 16 (17): 3883–96. <https://doi.org/10.1021/bi00636a026>.
- Brown, Wes, and Alexander Deiters. 2019. "Light-Activation of Cre Recombinase in Zebrafish Embryos through Genetic Code Expansion." In *Methods in Enzymology*. <https://doi.org/10.1016/bs.mie.2019.04.004>.

- Brueggemann, Angela B, Rekha Pai, Derrick W Crook, and Bernard Beall. 2007. "Vaccine Escape Recombinants Emerge after Pneumococcal Vaccination in the United States." *PLoS Pathogens* 3 (11): e168. <https://doi.org/10.1371/journal.ppat.0030168>.
- Brustad, Eric, Mark L. Bushey, Jae Wook Lee, Dan Groff, Wenshe Liu, and Peter G. Schultz. 2008. "A Genetically Encoded Boronate-Containing Amino Acid." *Angewandte Chemie - International Edition* 47 (43): 8220–23. <https://doi.org/10.1002/anie.200803240>.
- Bryson, David I., Chenguang Fan, Li Tao Guo, Corwin Miller, Dieter Söll, and David R. Liu. 2017. "Continuous Directed Evolution of Aminoacyl-TRNA Synthetases." *Nature Chemical Biology* 13 (12): 1253–60. <https://doi.org/10.1038/nchembio.2474>.
- Buechter, Douglas D., David N. Paoletta, Bridget S. Leslie, Melissa S. Brown, Karen A. Mehos, and Elliott A. Gruskin. 2003. "Co-Translational Incorporation of Trans-4-Hydroxyproline into Recombinant Proteins in Bacteria." *Journal of Biological Chemistry* 278 (1): 645–50. <https://doi.org/10.1074/jbc.M209364200>.
- Bussing, A. 1996. "Induction of Apoptosis by the Mistletoe Lectins: A Review on the Mechanisms of Cytotoxicity Mediated by *Viscum Album* L." *Apoptosis* 1 (1): 25–32. <https://doi.org/10.1007/BF00142075>.
- Butler, Jay C., Robert F. Breiman, John F. Campbell, Harvey B. Lipman, Claire V. Broome, and Richard R. Facklam. 1993. "Pneumococcal Polysaccharide Vaccine Efficacy: An Evaluation of Current Recommendations." *JAMA: The Journal of the American Medical Association* 270 (15): 1826–31. <https://doi.org/10.1001/jama.1993.03510150060030>.
- Butler, Jay C., Eugene D. Shapiro, and George M. Carlone. 1999. "Pneumococcal Vaccines: History, Current Status, and Future Directions." In *American Journal of Medicine*, 107:69–76. Elsevier Inc. [https://doi.org/10.1016/s0002-9343\(99\)00105-9](https://doi.org/10.1016/s0002-9343(99)00105-9).
- Cao, Yu, Jun Y Axup, Jennifer S Y Ma, Rongsheng E Wang, Seihyun Choi, Virginie Tardif, Reyna K V Lim, et al. 2015. "Multiformat T-Cell-Engaging Bispecific Antibodies Targeting Human Breast Cancers." *Angewandte Chemie (International Ed. in English)* 54 (24): 7022–27. <https://doi.org/10.1002/anie.201500799>.
- Carboni, F, F Angiolini, M Fabbri, B Brogioni, A Corrado, F Berti, R Adamo, and I Margarit. 2019. "GBS Type III Oligosaccharides Containing a Minimal Protective Epitope Can Be Turned into Effective Vaccines by Multivalent Presentation." *The Journal of Infectious Diseases*, no. Xx Xxxx: 1–5. <https://doi.org/10.1093/infdis/jiz551>.
- Carlo, Unverzagt, and Kajihara Yasuhiro. 2018. "Recent Advances in the Chemical Synthesis of N-Linked Glycoproteins." *Current Opinion in Chemical Biology* 46: 130–37. <https://doi.org/10.1016/j.cbpa.2018.07.004>.
- Carter, Tom S., Tidlo J. Mooibroek, Patrick F.N. Stewart, Matthew P. Crump, M. Carmen Galan, and Anthony P. Davis. 2016. "Platform Synthetic Lectins for Divalent Carbohydrate Recognition in Water." *Angewandte Chemie - International Edition* 55 (32): 9311–15. <https://doi.org/10.1002/anie.201603082>.
- Case, David A., Thomas E. Cheatham, Tom Darden, Holger Gohlke, Ray Luo, Kenneth M. Merz, Alexey Onufriev, Carlos Simmerling, Bing Wang, and Robert J. Woods. 2005. "The Amber Biomolecular Simulation Programs." *Journal of Computational Chemistry*. <https://doi.org/10.1002/jcc.20290>.
- CDC. 2017. "Pneumococcal Vaccination | What You Should Know | CDC." 2017. <https://www.cdc.gov/vaccines/vpd/pneumo/public/index.html>.
- Cech, Thomas, and Sidney Atlman. 1989. "Nobel Prize of Chemistry 1989." 1989.
- Chaguza, Chrispin, Jennifer E. Cornick, and Dean B. Everett. 2015. "Mechanisms and Impact of Genetic Recombination in the Evolution of *Streptococcus Pneumoniae*." *Computational and Structural Biotechnology Journal* 13: 241–47. <https://doi.org/10.1016/j.csbj.2015.03.007>.
- Chai, Melissa H., Florian Weiland, Richard M. Harvey, Peter Hoffmann, Abiodun D. Ogunniyi, and James C. Paton. 2017. "Proteomic Comparisons of Opaque and Transparent Variants of *Streptococcus Pneumoniae* by Two Dimensional-Differential Gel Electrophoresis." *Scientific Reports* 7 (1). <https://doi.org/10.1038/s41598-017-02465-x>.
- Chalker, Justin M., Smita B. Gunnoo, Omar Boutureira, Stefanie C. Gerstberger, Marta Fernández-González, Gonalo J.L. Bernardes, Laura Griffin, Hanna Hailu, Christopher J. Schofield, and Benjamin G. Davis. 2011. "Methods for Converting Cysteine to Dehydroalanine on Peptides and Proteins." *Chemical Science* 2 (9): 1666–76. <https://doi.org/10.1039/c1sc00185j>.
- Chatterjee, Abhishek, Sophie B. Sun, Jennifer L. Furman, Han Xiao, and Peter G. Schultz. 2013. "A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in *Escherichia Coli*." *Biochemistry* 52 (10): 1828–37. <https://doi.org/10.1021/bi4000244>.

- Chen, Hao, Sumana Venkat, Paige McGuire, Qinglei Gan, and Chenguang Fan. 2018. "Recent Development of Genetic Code Expansion for Posttranslational Modification Studies." *Molecules* 23 (7): 1–19. <https://doi.org/10.3390/molecules23071662>.
- Chen, Ridong, Ann Greer, and Antony M. Dean. 1996. "Redesigning Secondary Structure to Invert Coenzyme Specificity in Isopropylmalate Dehydrogenase." *Proceedings of the National Academy of Sciences of the United States of America* 93 (22): 12171–76. <https://doi.org/10.1073/pnas.93.22.12171>.
- Chen, Zeyu. 2016. "Immunogenicity and Protective Immunity against Otitis Media Caused by Pneumococcus in Mice of Hib Conjugate Vaccine with PsaA Protein Carrier." <https://doi.org/10.1007/s11684-016-0470-y>.
- Cherazard, Regine, Marcia Epstein, Thien Ly Doan, Tanzila Salim, Sheena Bharti, and Miriam A. Smith. 2017. "Antimicrobial Resistant Streptococcus Pneumoniae: Prevalence, Mechanisms, and Clinical Implications." *American Journal of Therapeutics* 24 (3): e361–69. <https://doi.org/10.1097/MJT.0000000000000551>.
- Chin, Jason W., T. Ashton Cropp, Stephanie Chu, Eric Meggers, and Peter G. Schultz. 2003. "Progress toward an Expanded Eukaryotic Genetic Code." *Chemistry and Biology* 10 (6): 511–19. [https://doi.org/10.1016/S1074-5521\(03\)00123-6](https://doi.org/10.1016/S1074-5521(03)00123-6).
- Chou, Chungjung, and Alexander Deiters. 2011. "Light-Activated Gene Editing with a Photocaged Zinc-Finger Nuclease." *Angewandte Chemie - International Edition* 50 (30): 6839–42. <https://doi.org/10.1002/anie.201101157>.
- Chou, Chungjung, Douglas D. Young, and Alexander Deiters. 2009. "A Light-Activated DNA Polymerase." *Angewandte Chemie - International Edition* 48 (32): 5950–53. <https://doi.org/10.1002/anie.200901115>.
- Cioci, Gianluca, Edward P. Mitchell, Catherine Gautier, Michaela Wimmerová, Dvora Sudakevitz, Serge Pérez, Nechama Gilboa-Garber, and Anne Imberty. 2003. "Structural Basis of Calcium and Galactose Recognition by the Lectin PA-IL of *Pseudomonas Aeruginosa*." *FEBS Letters* 555 (2): 297–301. [https://doi.org/10.1016/S0014-5793\(03\)01249-3](https://doi.org/10.1016/S0014-5793(03)01249-3).
- Clarke, Lionel. 2012. "A Synthetic Biology Roadmap for the UK." *UK Synthetic Biology Roadmap Coordination Group*.
- Colombo, Cinzia, Olimpia Pitirollo, and Luigi Lay. 2018. "Recent Advances in the Synthesis of Glycoconjugates for Vaccine Development." *Molecules (Basel, Switzerland)*. <https://doi.org/10.3390/molecules23071712>.
- Coraux, Christelle, Rodolphe Hajj, Pierre Lesimple, and Edith Puchelle. 2005. "Réparation et Régénération de l'épithélium Respiratoire." *Medecine/Sciences*. Elsevier Masson SAS. <https://doi.org/10.1051/medsci/200521121063>.
- Cornish, V. W., D. R. Benson, C. A. Altenbach, K. Hideg, W. L. Hubbell, and P. G. Schultz. 1994. "Site-Specific Incorporation of Biophysical Probes into Proteins." *Proceedings of the National Academy of Sciences of the United States of America* 91 (8): 2910–14. <https://doi.org/10.1073/pnas.91.8.2910>.
- Cowie, Dean B., and Georges N. Cohen. 1957. "Biosynthesis by *Escherichia Coli* of Active Altered Proteins Containing Selenium Instead of Sulfur." *BBA - Biochimica et Biophysica Acta* 26 (2): 252–61. [https://doi.org/10.1016/0006-3002\(57\)90003-3](https://doi.org/10.1016/0006-3002(57)90003-3).
- Csordas, Fátima C.L., Cátia T. Perciani, Michelle Darrieux, Viviane M. Gonçalves, Joaquim Cabrera-Crespo, Mickie Takagi, Maria E. Sbrógio-Almeida, Luciana C.C. Leite, and Martha M. Tanizaki. 2008. "Protection Induced by Pneumococcal Surface Protein A (PspA) Is Enhanced by Conjugation to a *Streptococcus Pneumoniae* Capsular Polysaccharide." *Vaccine* 26 (23): 2925–29. <https://doi.org/10.1016/j.vaccine.2008.03.038>.
- Daggett, Kelly A., Mark Layer, and T. Ashton Cropp. 2009. "A General Method for Scanning Unnatural Amino Acid Mutagenesis." *ACS Chemical Biology* 23 (1): 1–7. <https://doi.org/10.1038/jid.2014.371>.
- Damme, J.M Van. 2014. "History of Plant Lectin Research." *Encyclopedia of Biological Chemistry: Second Edition* 1200: 701–5. <https://doi.org/10.1016/B978-0-12-378630-2.00217-6>.
- Database, CAZY. n.d. "CAZY Database." <Http://Www.Cazy.Org/Carbohydrate-Binding-Modules.Html>.
- Datta, Deepshikha, Nagarajan Vaidehi, Deqiang Zhang, and William A. Goddard. 2009. "Selectivity and Specificity of Substrate Binding in Methionyl-TRNA Synthetase." *Protein Science* 13 (10): 2693–2705. <https://doi.org/10.1110/ps.04792204>.
- Davis, Kimberly M., Henry T. Akinbi, Alistair J. Standish, and Jeffrey N. Weiser. 2008. "Resistance to Mucosal Lysozyme Compensates for the Fitness Deficit of Peptidoglycan Modifications by *Streptococcus Pneumoniae*." *PLoS Pathogens* 4 (12). <https://doi.org/10.1371/journal.ppat.1000241>.
- Davydova, Anna, Maria Vorobjeva, Dmitrii Pyshnyi, Sidney Altman, Valentin Vlassov, and Alya Venyaminova. 2016. "Aptamers against Pathogenic Microorganisms." *Critical Reviews in Microbiology*. Taylor and Francis Ltd. <https://doi.org/10.3109/1040841X.2015.1070115>.

- Dawson, P E, and S B Kent. 2000. "Synthesis of Native Proteins by Chemical Ligation." *Annual Review of Biochemistry* 69: 923–60. <https://doi.org/10.1146/annurev.biochem.69.1.923>.
- Debelouchina, Galia T., and Tom W. Muir. 2017. "A Molecular Engineering Toolbox for the Structural Biologist." *Quarterly Reviews of Biophysics* 50 (May): 1–41. <https://doi.org/10.1017/S0033583517000051>.
- Dedkova, Larisa M., Nour Eddine Fahmi, Serguei Y. Golovine, and Sidney M. Hecht. 2003. "Enhanced D-Amino Acid Incorporation into Protein by Modified Ribosomes." *Journal of the American Chemical Society* 125 (22): 6616–17. <https://doi.org/10.1021/ja035141q>.
- Dedkova, Larisa M., Nour Eddine Fahmi, Rakesh Paul, Melissa Del Rosario, Liqiang Zhang, Shengxi Chen, Glen Feder, and Sidney M. Hecht. 2012. "β-Puromycin Selection of Modified Ribosomes for in Vitro Incorporation of β-Amino Acids." *Biochemistry* 51 (1): 401–15. <https://doi.org/10.1021/bi2016124>.
- Deiters, Alexander, Dan Groff, Youngha Ryu, Jianming Xie, and Peter G. Schultz. 2006. "A Genetically Encoded Photocaged Tyrosine." *Angewandte Chemie - International Edition* 45 (17): 2728–31. <https://doi.org/10.1002/anie.200600264>.
- Deley Cox, Vanessa E., Megan F. Cole, and Eric A. Gaucher. 2019. "Incorporation of Modified Amino Acids by Engineered Elongation Factors with Expanded Substrate Capabilities." *ACS Synthetic Biology* 8 (2): 287–96. <https://doi.org/10.1021/acssynbio.8b00305>.
- Delmont, J, Eric Pichard, S Jaureguiberry, B Marchou, P Parola, F Simon, E Adehossi, Bâ Fall, B Baldin, and A Berrebi. 2012. *Maladies Infectieuses Tropicales*.
- Devaraj, Neal K. 2018. "The Future of Bioorthogonal Chemistry." *ACS Central Science* 4 (8): 952–59. <https://doi.org/10.1021/acscentsci.8b00251>.
- Diavatopoulos, Dimitri A., Kirsty R. Short, John T. Price, Jonathan J. Wilksch, Lorena E. Brown, David E. Briles, Richard A. Strugnell, and Odilia L. Wijburg. 2010. "Influenza A Virus Facilitates Streptococcus Pneumoniae Transmission and Disease." *FASEB Journal* 24 (6): 1789–98. <https://doi.org/10.1096/fj.09-146779>.
- Diehl, Carl, Olof Engström, Tamara Delaine, Maria Håkansson, Samuel Genheden, Kristofer Modig, Hakon Leffler, Ulf Ryde, Ulf J. Nilsson, and Mikael Akke. 2010. "Protein Flexibility and Conformational Entropy in Ligand Design Targeting the Carbohydrate Recognition Domain of Galectin-3." *Journal of the American Chemical Society* 132 (41): 14577–89. <https://doi.org/10.1021/ja105852y>.
- Dion, Johann, Tamara Advedissian, Nataliya Storozhylova, Samir Dahbi, Annie Lambert, Frédérique Deshayes, Mireille Viguier, et al. 2017. "Development of a Sensitive Microarray Platform for the Ranking of Galectin Inhibitors: Identification of a Selective Galectin-3 Inhibitor." *ChemBioChem* 18 (24): 2428–40. <https://doi.org/10.1002/cbic.201700544>.
- Donkor, Eric S. 2013. "Understanding the Pneumococcus: Transmission and Evolution." *Frontiers in Cellular and Infection Microbiology*. <https://doi.org/10.3389/fcimb.2013.00007>.
- Dougherty, Dennis A. 2000. "Unnatural Amino Acids as Probes of Protein Structure and Function." *Current Opinion in Chemical Biology* 4 (6): 645–52. [https://doi.org/10.1016/S1367-5931\(00\)00148-4](https://doi.org/10.1016/S1367-5931(00)00148-4).
- Drienovská, Ivana, Lur Alonso-Cotchico, Pietro Vidossich, Agustí Lledós, Jean-Didier Maréchal, and Gerard Roelfes. 2017. "Design of an Enantioselective Artificial Metallo-Hydratase Enzyme Containing an Unnatural Metal-Binding Amino Acid." *Chemical Science* 8 (10): 7228–35. <https://doi.org/10.1039/c7sc03477f>.
- Drienovská, Ivana, Clemens Mayer, Christopher Dulson, and Gerard Roelfes. 2018. "A Designer Enzyme for Hydrazone and Oxime Formation Featuring an Unnatural Catalytic Aniline Residue." *Nature Chemistry* 10 (9): 946–52. <https://doi.org/10.1038/s41557-018-0082-z>.
- Drienovská, Ivana, Ana Rioz-Martínez, Apparao Draksharapu, and Gerard Roelfes. 2015. "Novel Artificial Metalloenzymes by in Vivo Incorporation of Metal-Binding Unnatural Amino Acids." *Chemical Science* 6 (1): 770–76. <https://doi.org/10.1039/c4sc01525h>.
- Drücker, Patrick, Simon Bachler, Heidi Wolfmeier, Roman Schoenauer, René Köffel, Viktoria S. Babiychuk, Petra S. Dittrich, Annette Draeger, and Eduard B. Babiychuk. 2018. "Pneumolysin-Damaged Cells Benefit from Non-Homogeneous Toxin Binding to Cholesterol-Rich Membrane Domains." *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* 1863 (8): 795–805. <https://doi.org/10.1016/j.bbalip.2018.04.010>.
- Dumas, Anaëlle, Lukas Lercher, Christopher D. Spicer, and Benjamin G. Davis. 2015. "Designing Logical Codon Reassignment-Expanding the Chemistry in Biology." *Chemical Science* 6 (1): 50–69. <https://doi.org/10.1039/c4sc01534g>.

- Durme, Joost Van, Javier Delgado, Francois Stricher, Luis Serrano, Joost Schymkowitz, and Frederic Rousseau. 2011. "A Graphical Interface for the FoldX Forcefield." *Bioinformatics (Oxford, England)* 27 (12): 1711–12. <https://doi.org/10.1093/bioinformatics/btr254>.
- Edwardraja, Selvakumar, Andreas Eichinger, Ina Theobald, Carina Andrea Sommer, Andreas J. Reichert, and Arne Skerra. 2017. "Rational Design of an Anticalin-Type Sugar-Binding Protein Using a Genetically Encoded Boronate Side Chain." *ACS Synthetic Biology* 6 (12): 2241–47. <https://doi.org/10.1021/acssynbio.7b00199>.
- Ehrlich, Michael, Michael J. Gattner, Bastien Viverge, Johanna Bretzler, David Eisen, Michael Stadlmeier, Milan Vrabel, and Thomas Carell. 2015. "Orchestrating the Biosynthesis of an Unnatural Pyrrolysine Amino Acid for Its Direct Incorporation into Proteins Inside Living Cells." *Chemistry - A European Journal* 21 (21): 7701–4. <https://doi.org/10.1002/chem.201500971>.
- Ellis, Gregory A., Michael J. Palte, and Ronald T. Raines. 2012. "Boronate-Mediated Biologic Delivery." *Journal of the American Chemical Society* 134 (8): 3631–34. <https://doi.org/10.1021/ja210719s>.
- Engelen-Lee, Joo Yeon, Matthijs C. Brouwer, Eleonora Aronica, and Diederik van de Beek. 2016. "Pneumococcal Meningitis: Clinical-Pathological Correlations (MeninGene-Path)." *Acta Neuropathologica Communications* 4 (March): 26. <https://doi.org/10.1186/s40478-016-0297-4>.
- Erickson, Sarah B., Raja Mukherjee, Rachel E. Kelemen, Chester J.J. Wrobel, Xiaofu Cao, and Abhishek Chatterjee. 2017. "Precise Photoremovable Perturbation of a Virus–Host Interaction." *Angewandte Chemie - International Edition* 56 (15): 4234–37. <https://doi.org/10.1002/anie.201700683>.
- Esposito, Susanna, and Nicola Principi. 2015. "Impacts of the 13-Valent Pneumococcal Conjugate Vaccine in Children." *Journal of Immunology Research*. Hindawi Publishing Corporation. <https://doi.org/10.1155/2015/591580>.
- Esvelt, Kevin M., Jacob C. Carlson, and David R. Liu. 2011. "A System for the Continuous Directed Evolution of Biomolecules." *Nature* 472 (7344): 499–503. <https://doi.org/10.1038/nature09929>.
- Fairman, Jeffery, Jon Heinrichs, and Wei Chan. 2018. SutroVax_WO2018126229, issued 2018.
- Fan, Chenguang, Kevan Ip, and Dieter Söll. 2016. "Expanding the Genetic Code of Escherichia Coli with Phosphotyrosine." *FEBS Letters*. Wiley Blackwell. <https://doi.org/10.1002/1873-3468.12333>.
- Fan, Chenguang, Hai Xiong, Noah M. Reynolds, and Dieter Söll. 2015. "Rationally Evolving TRNAPyl for Efficient Incorporation of Noncanonical Amino Acids." *Nucleic Acids Research* 43 (22). <https://doi.org/10.1093/nar/gkv800>.
- Fang, Guiqian, Hao Wang, Zhancun Bian, Jie Sun, Aiqin Liu, Hao Fang, Bo Liu, Qingqiang Yao, and Zhongyu Wu. 2018. "Recent Development of Boronic Acid-Based Fluorescent Sensors." *RSC Advances*. Royal Society of Chemistry. <https://doi.org/10.1039/c8ra04503h>.
- Farabaugh, Philip. 2000. "Translational Frameshifting : Implications for the Mechanism of Translational Frame Maintenance I . Introduction." *Progress in Nucleic Acid Research and Molecular Biology* 64: 131–70.
- Feinberg, Hadar, Riccardo Castelli, Kurt Drickamer, Peter H Seeberger, and William I Weis. 2007. "Multiple Modes of Binding Enhance the Affinity of DC-SIGN for High Mannose N-Linked Glycans Found on Viral Glycoproteins." *The Journal of Biological Chemistry* 282 (6): 4202–9. <https://doi.org/10.1074/jbc.M609689200>.
- Ferrand, Yann, Matthew P Crump, and Anthony P Davis. 2007. "A Synthetic Lectin Analog for Biomimetic Disaccharide Recognition." *Science (New York, N.Y.)* 318 (5850): 619–22. <https://doi.org/10.1126/science.1148735>.
- Fleissner, Mark R., Eric M. Brustad, Tamás Kálai, Christian Altenbach, Duilio Cascio, Francis B. Peters, Kálmán Hideg, Peter G. Schultz, and Wayne L. Hubbell. 2009. "Site-Directed Spin Labeling of a Genetically Encoded Unnatural Amino Acid." *Proceedings of the National Academy of Sciences of the United States of America* 106 (51): 21637–42. <https://doi.org/10.1073/pnas.0912009106>.
- Fontes, C. M.G.A., G. P. Hazlewood, E. Morag, J. Hall, B. H. Hirst, and H. J. Gilbert. 1995. "Evidence for a General Role for Non-Catalytic Thermostabilizing Domains in Xylanases from Thermophilic Bacteria." *Biochemical Journal* 307 (1): 151–58. <https://doi.org/10.1042/bj3070151>.
- Freeland, Stephen J., Robin D. Knight, Laura F. Landweber, and Laurence D. Hurst. 2000. "Early Fixation of an Optimal Genetic Code." *Molecular Biology and Evolution* 17 (4): 511–18. <https://doi.org/10.1093/oxfordjournals.molbev.a026331>.
- Frottier, Jacques, and Pierre Ambroise-Thomas. 2005. *Dictionnaire Des Maladies Infectieuses*. http://www.cilf.fr/images/livre_grand_16.jpg.

- Fujimoto, Zui, Hiroaki Tateno, and Jun Hirabayashi. 2014. "Chapter 46 Lectin Structures : Classification Based on the 3-D Structures" 1200. <https://doi.org/10.1007/978-1-4939-1292-6>.
- Gabius, Hans Joachim. 2018. "The Sugar Code: Why Glycans Are so Important." *BioSystems* 164: 102–11. <https://doi.org/10.1016/j.biosystems.2017.07.003>.
- Gaieski, David F., J. Matthew Edwards, Michael J. Kallan, and Brendan G. Carr. 2013. "Benchmarking the Incidence and Mortality of Severe Sepsis in the United States." *Critical Care Medicine* 41 (5): 1167–74. <https://doi.org/10.1097/CCM.0b013e31827c09f8>.
- Gan, Rui, Jessica G. Perez, Erik D. Carlson, Ioanna Ntai, Farren J. Isaacs, Neil L. Kelleher, and Michael C. Jewett. 2017. "Translation System Engineering in Escherichia Coli Enhances Non-Canonical Amino Acid Incorporation into Proteins." *Biotechnology and Bioengineering* 114 (5): 1074–86. <https://doi.org/10.1002/bit.26239>.
- Gautier, Arnaud, Alexander Deiters, and Jason W. Chin. 2011. "Light-Activated Kinases Enable Temporal Dissection of Signaling Networks in Living Cells." *Journal of the American Chemical Society* 133 (7): 2124–27. <https://doi.org/10.1021/ja1109979>.
- Gebauer, Michaela, and Arne Skerra. 2012. "Anticalins: Small Engineered Binding Proteins Based on the Lipocalin Scaffold." In *Methods in Enzymology*, 503:157–88. Academic Press Inc. <https://doi.org/10.1016/B978-0-12-396962-0.00007-0>.
- Geno, K. Aaron, Gwendolyn L. Gilbert, Joon Young Song, Ian C. Skovsted, Keith P. Klugman, Christopher Jones, Helle B. Konradsen, and Moon H. Nahm. 2015. "Pneumococcal Capsules and Their Types: Past, Present, and Future." *Clinical Microbiology Reviews* 28 (3): 871–99. <https://doi.org/10.1128/CMR.00024-15>.
- Gerding, Hanne R., Christiaan Karreman, Andreas Daiber, Johannes Delp, Daniel Hammler, Martin Mex, Stefan Schildknecht, and Marcel Leist. 2019. "Reductive Modification of Genetically Encoded 3-Nitrotyrosine Sites in Alpha Synuclein Expressed in E.Coli." *Redox Biology* 26 (September). <https://doi.org/10.1016/j.redox.2019.101251>.
- Giege, R., M. Sissler, and C. Florentz. 1998. "Universal Rules and Idiosyncratic Features in tRNA Identity." *Nucleic Acids Research* 26 (22): 5017–35. <https://doi.org/10.1093/nar/26.22.5017>.
- Giegé, Richard, and Gilbert Eriani. 2014. "Transfer RNA Recognition and Aminoacylation by Synthetases." *ELS*. <https://doi.org/10.1002/9780470015902.a0000531.pub3>.
- Giel-Pietraszuk, Malgorzata, and Jan Barciszewski. 2006. "Charging of tRNA with Non-Natural Amino Acids at High Pressure." *FEBS Journal* 273 (13): 3014–23. <https://doi.org/10.1111/j.1742-4658.2006.05312.x>.
- Gilkes, N R, R A Warren, R C Miller, and D G Kilburn. 1988. "Precise Excision of the Cellulose Binding Domains from Two Cellulomonas Fimi Cellulases by a Homologous Protease and the Effect on Catalysis." *The Journal of Biological Chemistry* 263 (21): 10401–7. <http://www.ncbi.nlm.nih.gov/pubmed/3134347>.
- González-miro, Majela, Laura Rodríguez-noda, and Mildrey Fariñas-medina. 2017. "Self-Assembled Particulate PsaA as Vaccine against Streptococcus Pneumoniae Infection." *Heliyon* 3 (March): e00291. <https://doi.org/10.1016/j.heliyon.2017.e00291>.
- Goto, Yuki, Takayuki Katoh, and Hiroaki Suga. 2011. "Flexizymes for Genetic Code Reprogramming." *Nature Protocols* 6 (6): 779–90. <https://doi.org/10.1038/nprot.2011.331>.
- Goupy, Jacques, and Lee. Creighton. 2006. *Introduction Aux Plans d'expériences*. Dunod.
- Grabenstein, J. D., and K. P. Klugman. 2012. "A Century of Pneumococcal Vaccination Research in Humans." *Clinical Microbiology and Infection*. Blackwell Publishing Ltd. <https://doi.org/10.1111/j.1469-0691.2012.03943.x>.
- Gragani, Tiziana, Doretta Cuffaro, Silvia Fallarini, Grazia Lombardi, Felicia D'Andrea, and Lorenzo Guazzelli. 2019. "Selectively Charged and Zwitterionic Analogues of the Smallest Immunogenic Structure of Streptococcus Pneumoniae Type 14." *Molecules* 24 (18): 3414. <https://doi.org/10.3390/molecules24183414>.
- Grammel, Markus, and Howard C Hang. 2013. "Chemical Reporters for Biological Discovery." *Nature Chemical Biology* 9 (8): 475–84. <https://doi.org/10.1038/nchembio.1296>.
- Grandjean, Cyrille, Alain Boutonnier, Catherine Guerreiro, Jean Michel Fournier, and Laurence A. Mulard. 2005. "On the Preparation of Carbohydrate-Protein Conjugates Using the Traceless Staudinger Ligation." *Journal of Organic Chemistry* 70 (18): 7123–32. <https://doi.org/10.1021/jo0505472>.
- Gratten, M, H Gratten, A Poli, E Carrad, M Raymer, and G Koki. 1986. "Colonisation of Haemophilus Influenzae and Streptococcus Pneumoniae in the Upper Respiratory Tract of Neonates in Papua New Guinea: Primary Acquisition, Duration of Carriage, and Relationship to Carriage in Mothers." *Biology of the Neonate* 50 (2): 114–20.

<https://doi.org/10.1159/000242576>.

- Gregoriadis, Gregory, Brenda MacCormack, Anthony C Allison, and Georges Poste. 1992. *New Generation Vaccines The Role of Basic Immunology*.
- Guberman, Mónica, and Peter H. Seeberger. 2019. "Automated Glycan Assembly: A Perspective." *Journal of the American Chemical Society* 141 (14): 5581–92. <https://doi.org/10.1021/jacs.9b00638>.
- Guillén, Daniel, Sergio Sánchez, and Romina Rodríguez-Sanoja. 2010. "Carbohydrate-Binding Domains: Multiplicity of Biological Roles." *Applied Microbiology and Biotechnology* 85 (5): 1241–49. <https://doi.org/10.1007/s00253-009-2331-y>.
- Gunther, G R, J L Wang, I Yahara, B A Cunningham, and G M Edelman. 1973. "Concanavalin A Derivatives with Altered Biological Activities." *Proceedings of the National Academy of Sciences of the United States of America* 70 (4): 1012–16. <https://doi.org/10.1073/pnas.70.4.1012>.
- H. Surangi, N. Jayawardena, Xin Wang, Mingdi Yan. 2013. "Classification of Lectins by PATTERN Recognition Using Glyconanoparticles." *Anal Chem.* 85 (21). <https://doi.org/10.1021/ac402069j>.Classification.
- Habets, Michelle G.J.L., Daniel E. Rozen, and Michael A. Brockhurst. 2012. "Variation in Streptococcus Pneumoniae Susceptibility to Human Antimicrobial Peptides May Mediate Intraspecific Competition." *Proceedings of the Royal Society B: Biological Sciences* 279 (1743): 3803–11. <https://doi.org/10.1098/rspb.2012.1118>.
- Habib, Maha, Barbara D. Porter, and Catherine Satzke. 2014. "Capsular Serotyping of Streptococcus Pneumoniae Using the Quellung Reaction." *Journal of Visualized Experiments*, no. 84 (February). <https://doi.org/10.3791/51208>.
- Hadd, Andrew, and John J. Perona. 2014. "Recoding Aminoacyl-TRNA Synthetases for Synthetic Biology by Rational Protein-Rna Engineering." *ACS Chemical Biology* 9 (12): 2761–66. <https://doi.org/10.1021/cb5006596>.
- Hall, Dennis G. 2006. *Structure, Properties, and Preparation of Boronic Acid Derivatives. Overview of Their Reactions and Applications. Boronic Acids: Preparation and Applications in Organic Synthesis and Medicine*. <https://doi.org/10.1002/3527606548.ch1>.
- . 2019. "Boronic Acid Catalysis." *Chemical Society Reviews*. Royal Society of Chemistry. <https://doi.org/10.1039/c9cs00191c>.
- Hammill, Jared T., Shigeki Miyake-Stoner, Jennifer L. Hazen, Jennifer C. Jackson, and Ryan A. Mehl. 2007. "Preparation of Site-Specifically Labeled Fluorinated Proteins for 19F-Nmr Structural Characterization." *Nature Protocols* 2 (10): 2601–7. <https://doi.org/10.1038/nprot.2007.379>.
- Hammitt, Laura L., Lisa R. Bulkow, Rosalyn J. Singleton, J. Pekka Nuorti, Kim Boyd Hummel, Karen M. Miernyk, Carolyn Zanis, et al. 2011. "Repeat Revaccination with 23-Valent Pneumococcal Polysaccharide Vaccine among Adults Aged 55-74 Years Living in Alaska: No Evidence of Hyporesponsiveness." *Vaccine* 29 (12): 2287–95. <https://doi.org/10.1016/j.vaccine.2011.01.029>.
- Harding, Christian M., Mohamed A. Nasr, Nichollas E. Scott, Guillaume Goyette-Desjardins, Harald Nothhaft, Anne E. Mayer, Sthefany M. Chavez, et al. 2019. "A Platform for Glycoengineering a Polyvalent Pneumococcal Bioconjugate Vaccine Using E. Coli as a Host." *Nature Communications* 10 (1). <https://doi.org/10.1038/s41467-019-08869-9>.
- Harrison, L I, H N Christensen, M E Handlogten, D L Oxender, and S C Quay. 1975. "Transport of L-4-Azaleucine in Escherichia Coli." *Journal of Bacteriology* 122 (3): 957–65. <http://www.ncbi.nlm.nih.gov/pubmed/238951>.
- Hartman, Matthew C.T., Kristopher Josephson, Chi Wang Lin, and Jack W. Szostak. 2007. "An Expanded Set of Amino Acid Analogs for the Ribosomal Translation of Unnatural Peptides." *PLoS ONE* 2 (10). <https://doi.org/10.1371/journal.pone.0000972>.
- Hartzell, Joshua D., Charles N. Oster, and Joel C. Gaydos. 2003. "How Contagious Are Common Respiratory Tract Infections? [4]." *New England Journal of Medicine* 349 (1): 95. <https://doi.org/10.1056/NEJMc031032>.
- Haruna, Ken Ichi, Muhammad H. Alkazemi, Yuchen Liu, Dieter Söll, and Markus Englert. 2014. "Engineering the Elongation Factor Tu for Efficient Selenoprotein Synthesis." *Nucleic Acids Research* 42 (15): 9976–83. <https://doi.org/10.1093/nar/gku691>.
- Hashim, Onn Haji, Jaime Jacqueline Jayapalan, and Cheng Siang Lee. 2017. "Lectins: An Effective Tool for Screening of Potential Cancer Biomarkers." *PeerJ* 2017 (9). <https://doi.org/10.7717/peerj.3784>.
- Hatakeyama, Tomomitsu, Tomohiro Ishimine, Tomohiro Baba, Masanari Kimura, Hideaki Unno, and Shuichiro Goda. 2013. "Alteration of the Carbohydrate-Binding Specificity of a C-Type Lectin CEL-I Mutant with an EPN Carbohydrate-

- Binding Motif." *Protein and Peptide Letters* 20 (7): 796–801. <http://www.ncbi.nlm.nih.gov/pubmed/23157284>.
- Hatch, Duane, Jeffrey Boles, Zizhong Li, and L. "Pete Silks. 2005. "Methods for the Synthesis of Heavy-Atom Derivatized Amino Acids: Useful Probes for X-Ray Crystallography, Vibrational, and NMR Spectroscopy of Proteins." *Current Organic Chemistry* 8 (1): 47–64. <https://doi.org/10.2174/1385272043486098>.
- Hauf, Matthias, Florian Richter, Tobias Schneider, Thomas Faidt, Berta M. Martins, Tobias Baumann, Patrick Durkin, et al. 2017. "Photoactivatable Mussel-Based Underwater Adhesive Proteins by an Expanded Genetic Code." *ChemBioChem* 18 (18): 1819–23. <https://doi.org/10.1002/cbic.201700327>.
- Hausdorff, William P., Bernard Hoet, and Lode Schuerman. 2010. "Do Pneumococcal Conjugate Vaccines Provide Any Cross-Protection against Serotype 19A?" *BMC Pediatrics* 10 (February). <https://doi.org/10.1186/1471-2431-10-4>.
- HCSP. 2018. "Calendrier Des Vaccinations et Recommandations Vaccinales 2018." https://solidarites-sante.gouv.fr/IMG/pdf/calendrier_vaccinations_2018.pdf.
- Hecht, S M. 2009. "Protein Engineering : Synthesis of Modified Proteins Using Misacylated TRNAs." *Protein Engineering (P255-270)* ISBN: 978-.
- Hemphill, James, Chungjung Chou, Jason W. Chin, and Alexander Deiters. 2013. "Genetically Encoded Light-Activated Transcription for Spatiotemporal Control of Gene Expression and Gene Silencing in Mammalian Cells." *Journal of the American Chemical Society* 135 (36): 13433–39. <https://doi.org/10.1021/ja4051026>.
- Hendrickson, O. D., and A. V. Zherdev. 2018. "Analytical Application of Lectins." *Critical Reviews in Analytical Chemistry*. Taylor and Francis Ltd. <https://doi.org/10.1080/10408347.2017.1422965>.
- Henrichsen, Jørgen. 1979. "The Pneumococcal Typing System and Pneumococcal Surveillance." *Journal of Infection* 1 (SUPPL. 2). [https://doi.org/10.1016/S0163-4453\(79\)80029-8](https://doi.org/10.1016/S0163-4453(79)80029-8).
- Henriques-Normark, Birgitta, and Elaine I. Tuomanen. 2013. "The Pneumococcus: Epidemiology, Microbiology, and Pathogenesis." *Cold Spring Harbor Perspectives in Medicine* 3 (7). <https://doi.org/10.1101/cshperspect.a010215>.
- Heo, Yun Jung, and Seong Hyok Kim. 2019. "Toward Long-Term Implantable Glucose Biosensors for Clinical Use." *Applied Sciences (Switzerland)*. MDPI AG. <https://doi.org/10.3390/app9102158>.
- Hermodson, M. 1996. "Current Protocols in Protein Science." *Proteins: Structure, Function and Genetics* 24 (3): 409. <https://doi.org/10.1002/prot.340240303>.
- Hest, Jan C.M van, and David A Tirrell. 1998. "Efficient Introduction of Alkene Functionality into Proteins in Vivo." *FEBS Letters* 428 (1–2): 68–70. [https://doi.org/10.1016/S0014-5793\(98\)00489-X](https://doi.org/10.1016/S0014-5793(98)00489-X).
- Higgins, Melanie A., Garrett E. Whitworth, Nihida El Warry, Mialy Randriantsoa, Eric Samain, Robert D. Burke, David J. Voadlo, and Alisdair B. Boraston. 2009. "Differential Recognition and Hydrolysis of Host Carbohydrate Antigens by Streptococcus Pneumoniae Family 98 Glycoside Hydrolases." *Journal of Biological Chemistry* 284 (38): 26161–73. <https://doi.org/10.1074/jbc.M109.024067>.
- Hiller, N. Luisa, Azad Ahmed, Evan Powell, Darren P. Martin, Rory Eutsey, Josh Earl, Benjamin Janto, et al. 2010. "Generation of Genic Diversity among Streptococcus Pneumoniae Strains via Horizontal Gene Transfer during a Chronic Polyclonal Pediatric Infection." *PLoS Pathogens* 6 (9). <https://doi.org/10.1371/journal.ppat.1001108>.
- Hirabayashi, Jun, and Ryoichi Arai. 2019. "Lectin Engineering : The Possible and the Actual." <https://doi.org/10.1007/978-1-4939-1292-6>.
- Hirabayashi, Jun, Tomomi Hashidate, Yoichiro Arata, and Nozomu Nishi. 2002. "Oligosaccharide Specificity of Galectins: A Search by Frontal Affinity Chromatography." *Biochimica et Biophysica Acta (BBA)* 1572: 1–23. [https://doi.org/10.1016/S0167-4834\(02\)00001-9](https://doi.org/10.1016/S0167-4834(02)00001-9).
- Hirst, R. A., A. Kadioglu, C. O'Callaghan, and P. W. Andrew. 2004. "The Role of Pneumolysin in Pneumococcal Pneumonia and Meningitis." *Clinical and Experimental Immunology*. <https://doi.org/10.1111/j.1365-2249.2004.02611.x>.
- Hoppmann, Christian, Allison Wong, Bing Yang, Shuwei Li, Tony Hunter, Kevan M. Shokat, and Lei Wang. 2017. "Site-Specific Incorporation of Phosphotyrosine Using an Expanded Genetic Code." *Nature Chemical Biology* 13 (8): 842–44. <https://doi.org/10.1038/nchembio.2406>.
- Hou, Ya Ming. 2010. "CCA Addition to TRNA: Implications for TRNA Quality Control." *IUBMB Life*. <https://doi.org/10.1002/iub.301>.
- Hsu, Che Hsiung, Sangho Park, David E. Mortenson, B. Lachele Foley, Xiacong Wang, Robert J. Woods, David A. Case, et al. 2016. "The Dependence of Carbohydrate-Aromatic Interaction Strengths on the Structure of the Carbohydrate."

- Journal of the American Chemical Society*. <https://doi.org/10.1021/jacs.6b02879>.
- Hu, C., and J. Wang. 2016. "Method for Enzyme Design with Genetically Encoded Unnatural Amino Acids." *Methods in Enzymology* 580: 109–33. <https://doi.org/10.1016/bs.mie.2016.06.005>.
- Hu, Dan, Hiroaki Tateno, and Jun Hirabayashi. 2015. "Lectin Engineering, a Molecular Evolutionary Approach to Expanding the Lectin Utilities." *Molecules* 20 (5): 7637–56. <https://doi.org/10.3390/molecules20057637>.
- Hu, Dan, Hiroaki Tateno, Atsushi Kuno, Rikio Yabe, and Jun Hirabayashi. 2012. "Directed Evolution of Lectins with Sugar-Binding Specificity for 6-Sulfo-Galactose." *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.M112.351965>.
- Hu, Shen, and David T. Wong. 2009. "Lectin Microarray." *Proteomics - Clinical Applications*. <https://doi.org/10.1002/prca.200800153>.
- Hudson, Kieran L., Gail J. Bartlett, Roger C. Diehl, Jon Agirre, Timothy Gallagher, Laura L. Kiessling, and Derek N. Woolfson. 2015. "Carbohydrate-Aromatic Interactions in Proteins." *Journal of the American Chemical Society*. <https://doi.org/10.1021/jacs.5b08424>.
- Huguenin-Dezot, Nicolas, Diego A. Alonzo, Graham W. Heberlig, Mohan Mahesh, Duy P. Nguyen, Mark H. Dornan, Christopher N. Boddy, T. Martin Schmeing, and Jason W. Chin. 2019. "Trapping Biosynthetic Acyl-Enzyme Intermediates with Encoded 2,3-Diaminopropionic Acid." *Nature*. <https://doi.org/10.1038/s41586-018-0781-z>.
- Hung, Chia Suei, Julie Bouckaert, Danielle Hung, Jerome Pinkner, Charlotte Widberg, Anthony DeFusco, C. Gale Auguste, et al. 2002. "Structural Basis of Tropism of Escherichia Coli to the Bladder during Urinary Tract Infection." *Molecular Microbiology* 44 (4): 903–15. <https://doi.org/10.1046/j.1365-2958.2002.02915.x>.
- Hyams, Catherine, Emilie Camberlein, Jonathan M. Cohen, Katie Bax, and Jeremy S. Brown. 2010. "The Streptococcus Pneumoniae Capsule Inhibits Complement Activity and Neutrophil Phagocytosis by Multiple Mechanisms." *Infection and Immunity* 78 (2): 704–15. <https://doi.org/10.1128/IAI.00881-09>.
- Hyams, Catherine, Jose Yuste, Katie Bax, Emilie Camberlein, Jeffrey N. Weiser, and Jeremy S. Brown. 2010. "Streptococcus Pneumoniae Resistance to Complement-Mediated Immunity Is Dependent on the Capsular Serotype." *Infection and Immunity* 78 (2): 716–25. <https://doi.org/10.1128/IAI.01056-09>.
- Ibba, Michael, and Hauke Hennecke. 1995. "Relaxing the Substrate Specificity of an Aminoacyl-TRNA Synthetase Allows in Vitro and in Vivo Synthesis of Proteins Containing Unnatural Amino Acids." *FEBS Letters* 364 (3): 272–75. [https://doi.org/10.1016/0014-5793\(95\)00408-2](https://doi.org/10.1016/0014-5793(95)00408-2).
- Ikeda, K, Y Sakamoto, Y Kawasaki, T Miyake, K Tanaka, T Urata, Y Katayama, S Ueda, and S Horiuchi. 1998. "Determination of Glycated Albumin by Enzyme-Linked Boronate Immunoassay (ELBIA)." *Clinical Chemistry* 44 (2): 256–63. <http://www.ncbi.nlm.nih.gov/pubmed/9474021>.
- Illardo, Melissa A., and Stephen J. Freeland. 2014. "Testing for Adaptive Signatures of Amino Acid Alphabet Evolution Using Chemistry Space." *Journal of Systems Chemistry* 5 (1). <https://doi.org/10.1186/1759-2208-5-1>.
- Ilies, Monica, Luigi Di Costanzo, Michelle L. North, Jeremy A. Scott, and David W. Christianson. 2010. "2-Aminoimidazole Amino Acids as Inhibitors of the Binuclear Manganese Metalloenzyme Human Arginase I." *Journal of Medicinal Chemistry* 53 (10): 4266–76. <https://doi.org/10.1021/jm100306a>.
- Imberty, Anne, and James H. Prestegard. 2017. "Structural Biology of Glycan Recognition." <https://doi.org/10.1101/glycobiology.3e.030>.
- Ishioka, G Y, A G Lamont, D Thomson, N Bullbow, F C Gaeta, A Sette, and H M Grey. 1992. "MHC Interaction and T Cell Recognition of Carbohydrates and Glycopeptides." *Journal of Immunology (Baltimore, Md. : 1950)* 148 (8): 2446–51. <http://www.ncbi.nlm.nih.gov/pubmed/1560200>.
- Italia, James S., Partha Sarathi Addy, Sarah B. Erickson, Jennifer C. Peeler, Eranthie Weerapana, and Abhishek Chatterjee. 2019. "Mutually Orthogonal Nonsense-Suppression Systems and Conjugation Chemistries for Precise Protein Labeling at up to Three Distinct Sites." Research-article. *Journal of the American Chemical Society* 141 (15): 6204–12. <https://doi.org/10.1021/jacs.8b12954>.
- Italia, James S., Partha Sarathi Addy, Chester J.J. Wrobel, Lisa A. Crawford, Marc J. Lajoie, Yunan Zheng, and Abhishek Chatterjee. 2017. "An Orthogonalized Platform for Genetic Code Expansion in Both Bacteria and Eukaryotes." *Nature Chemical Biology* 13 (4): 446–50. <https://doi.org/10.1038/nchembio.2312>.
- Itoh, Yukihiro, Keisuke Aihara, Paolo Mellini, Toshifumi Tojo, Yosuke Ota, Hiroki Tsumoto, Viswas Raja Solomon, et al. 2016. "Identification of SNAIL1 Peptide-Based Irreversible Lysine-Specific Demethylase 1-Selective Inactivators." *Journal of Medicinal Chemistry* 59 (4): 1531–44. <https://doi.org/10.1021/acs.jmedchem.5b01323>.

- Iwabe, N., K. Kuma, M. Hasegawa, S. Osawa, and T. Miyata. 1989. "Evolutionary Relationship of Archaeobacteria, Eubacteria, and Eukaryotes Inferred from Phylogenetic Trees of Duplicated Genes." *Proceedings of the National Academy of Sciences of the United States of America* 86 (23): 9355–59. <https://doi.org/10.1073/pnas.86.23.9355>.
- Jaeken, J., and H. Carchon. 1993. "The Carbohydrate-Deficient Glycoprotein Syndromes: An Overview." *Journal of Inherited Metabolic Disease* 16 (5): 813–20. <https://doi.org/10.1007/BF00714272>.
- Janoff, E. N., J. B. Rubins, C. Fasching, D. Charboneau, J. T. Rakhola, A. G. Plaut, and J. N. Weiser. 2014. "Pneumococcal IgA1 Protease Subverts Specific Protection by Human IgA1." *Mucosal Immunology* 7 (2): 249–56. <https://doi.org/10.1038/mi.2013.41>.
- Jean-Baptiste, Lamy, H  l  ne Berthelot, and Madeleine Favre. 2016. "Rainbow Boxes: A Technique for Visualizing Overlapping Sets and an Application to the Comparison of Drugs Properties." In *Proceedings of the International Conference on Information Visualisation*, 2016-August:253–60. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/IV.2016.26>.
- Jiang, Q. L., S. Zhang, M. Tian, S. Y. Zhang, T. Xie, D. Y. Chen, Y. J. Chen, et al. 2015. "Plant Lectins, from Ancient Sugar-Binding Proteins to Emerging Anti-Cancer Drugs in Apoptosis and Autophagy." *Cell Proliferation* 48 (1): 17–28. <https://doi.org/10.1111/cpr.12155>.
- Johnson, Scott E, Janet K Dykes, Danny L Jue, Jaquelyn S Sampson, George M Carlone, and Edwin W Ades. 2002. "Inhibition of Pneumococcal Carriage in Mice by Subcutaneous Immunization with Peptides from the Common Surface Protein Pneumococcal Surface Adhesin A." *The Journal of Infectious Diseases*, 19–23.
- Jones, David H., Susan E. Cellitti, Xueshi Hao, Qiong Zhang, Michael Jahnz, Daniel Summerer, Peter G. Schultz, Tetsuo Uno, and Bernhard H. Geierstanger. 2010. "Site-Specific Labeling of Proteins with NMR-Active Unnatural Amino Acids." *Journal of Biomolecular NMR* 46 (1): 89–100. <https://doi.org/10.1007/s10858-009-9365-4>.
- Jordan, E T, and I J Goldstein. 1995. "Site-Directed Mutagenesis Studies on the Lima Bean Lectin. Altered Carbohydrate-Binding Specificities Result from Single Amino Acid Substitutions." *European Journal of Biochemistry* 230 (3): 958–64. <https://doi.org/10.1111/j.1432-1033.1995.tb20642.x>.
- Jung, Jae Eun, Sang Yeul Lee, Hyojin Park, Hyojin Cha, Wooseok Ko, Kalme Sachin, Dong Wook Kim, Dae Yoon Chi, and Hyun Soo Lee. 2014. "Genetic Incorporation of Unnatural Amino Acids Biosynthesized from α -Keto Acids by an Aminotransferase." *Chemical Science* 5 (5): 1881–85. <https://doi.org/10.1039/c3sc51617b>.
- Kaku, H, Y Mori, I J Goldstein, and N Shibuya. 1993. "Monomeric, Monovalent Derivative of Maackia Amurensis Leukoagglutinin. Preparation and Application to the Study of Cell Surface Glycoconjugates by Flow Cytometry." *The Journal of Biological Chemistry* 268 (18): 13237–41. <http://www.ncbi.nlm.nih.gov/pubmed/7685757>.
- Kaku, Hanae, and Naoto Shibuya. 1992. "Preparation of a Stable Subunit of Japanese Elderberry (Sambucus Sieboldiana) Bark Lectin and Its Application for the Study of Cell Surface Carbohydrates by Flow Cytometry." *FEBS Letters*. [https://doi.org/10.1016/0014-5793\(92\)80994-R](https://doi.org/10.1016/0014-5793(92)80994-R).
- Kalstrup, Tanja, and Rikard Blunck. 2017. "Voltage-Clamp Fluorometry in Xenopus Oocytes Using Fluorescent Unnatural Amino Acids." *Journal of Visualized Experiments* 2017 (123). <https://doi.org/10.3791/55598>.
- Kane, James F. 1995. "Effects of Rare Codon Clusters on High-Level Expression of Heterologous Proteins in Escherichia Coli." *Current Opinion in Biotechnology* 6 (5): 494–500. [https://doi.org/10.1016/0958-1669\(95\)80082-4](https://doi.org/10.1016/0958-1669(95)80082-4).
- Kang, Ji Yong, Daichi Kawaguchi, Irene Coin, Zheng Xiang, Dennis D.M. O'Leary, Paul A. Slesinger, and Lei Wang. 2013. "In Vivo Expression of a Light-Activatable Potassium Channel Using Unnatural Amino Acids." *Neuron* 80 (2): 358–70. <https://doi.org/10.1016/j.neuron.2013.08.016>.
- Kapatai, Georgia, Carmen L. Sheppard, Ali Al-Shahib, David J. Litt, Anthony P. Underwood, Timothy G. Harrison, and Norman K. Fry. 2016. "Whole Genome Sequencing of Streptococcus Pneumoniae: Development, Evaluation and Verification of Targets for Serogroup and Serotype Prediction Using an Automated Pipeline." *PeerJ* 2016 (9). <https://doi.org/10.7717/peerj.2477>.
- Kapoor, Neeraj, Ivana Vanjak, James Rozzelle, Aym Berges, Wei Chan, Gang Yin, Cuong Tran, et al. 2018. "Malaria Derived Glycosylphosphatidylinositol Anchor Enhances Anti-Pfs25 Functional Antibodies That Block Malaria Transmission." *Biochemistry* 57 (5): 516–19. <https://doi.org/10.1021/acs.biochem.7b01099>.
- Kast, Peter, and Hauke Hennecke. 1991. "Amino Acid Substrate Specificity of Escherichia Coli Phenylalanyl-TRNA Synthetase Altered by Distinct Mutations." *Journal of Molecular Biology* 222 (1): 99–124. [https://doi.org/10.1016/0022-2836\(91\)90740-W](https://doi.org/10.1016/0022-2836(91)90740-W).
- Kawasaki, Norihito, Ichiro Matsuo, Kiichiro Totani, Daisuke Nawa, Noriko Suzuki, Daisuke Yamaguchi, Naoki Matsumoto,

- Yukishige Ito, and Kazuo Yamamoto. 2007. "Detection of Weak Sugar Binding Activity of VIP36 Using VIP36-Streptavidin Complex and Membrane-Based Sugar Chains." *Journal of Biochemistry*.
<https://doi.org/10.1093/jb/mvm024>.
- Kay, Emily, Jon Cuccui, and Brendan W. Wren. 2019. "Recent Advances in the Production of Recombinant Glycoconjugate Vaccines." *Npj Vaccines* 4 (1): 1–8. <https://doi.org/10.1038/s41541-019-0110-z>.
- Kellner, JD. 2011. "Mise à Jour Sur Le Succès Du Vaccin Conjugué Contre Le Pneumocoque." <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3076179/>.
- Keogh, Damien, Roisin Thompson, Ruth Larragy, Kenneth McMahon, Michael O'Connell, Brendan O'Connor, and Paul Clarke. 2014. "Generating Novel Recombinant Prokaryotic Lectins with Altered Carbohydrate Binding Properties through Mutagenesis of the PA-IL Protein from *Pseudomonas Aeruginosa*." *Biochimica et Biophysica Acta - General Subjects*. <https://doi.org/10.1016/j.bbagen.2014.01.020>.
- Kerrigan, Ann M., and Gordon D. Brown. 2009. "C-Type Lectins and Phagocytosis." *Immunobiology* 214 (7): 562–75. <https://doi.org/10.1016/j.imbio.2008.11.003>.
- Kerzmann, Andreas, Jan Fuhrmann, Oliver Kohlbacher, and Dirk Neumann. 2008. "BALLDock / SLICK : A New Method for Protein-Carbohydrate Docking," 1616–25.
- Khan, Farhat, Patricia M. Legler, Ryan M. Mease, Elizabeth H. Duncan, Elke S. Bergmann-Leitner, and Evelina Angov. 2012. "Histidine Affinity Tags Affect MSP1 42 Structural Stability and Immunodominance in Mice." *Biotechnology Journal* 7 (1): 133–47. <https://doi.org/10.1002/biot.201100331>.
- Khatun, Farjana, Rachel J Stephenson, and Istvan Toth. 2017. "An Overview of Structural Features of Antibacterial Glycoconjugate Vaccines That Influence Their Immunogenicity." *Chemistry (Weinheim an Der Bergstrasse, Germany)* 23 (18): 4233–54. <https://doi.org/10.1002/chem.201603599>.
- Kiick, Kristi L., Jan C. M. van Hest, and David A. Tirrell. 2000. "Expanding the Scope of Protein Biosynthesis by Altering the Methionyl-TRNA Synthetase Activity of a Bacterial Expression Host." *Angewandte Chemie International Edition* 39 (12): 2148–52. [https://doi.org/10.1002/1521-3773\(20000616\)39:12<2148::AID-ANIE2148>3.0.CO;2-7](https://doi.org/10.1002/1521-3773(20000616)39:12<2148::AID-ANIE2148>3.0.CO;2-7).
- Kiick, Kristi L., Eliana Saxon, David A. Tirrell, and Carolyn R. Bertozzi. 2002. "Incorporation of Azides into Recombinant Proteins for Chemoselective Modification by the Staudinger Ligation." *Proceedings of the National Academy of Sciences of the United States of America* 99 (1): 19–24. <https://doi.org/10.1073/pnas.012583299>.
- Kiick, Kristi L., and David A. Tirrell. 2000. "Protein Engineering by in Vivo Incorporation of Non-Natural Amino Acids: Control of Incorporation of Methionine Analogues by Methionyl-TRNA Synthetase." *Tetrahedron* 56 (48): 9487–93. [https://doi.org/10.1016/S0040-4020\(00\)00833-4](https://doi.org/10.1016/S0040-4020(00)00833-4).
- Kim, Chan Hyuk, Jun Y. Axup, Anna Dubrovskaya, Stephanie A. Kazane, Benjamin A. Hutchins, Erik D. Wold, Vaughn V. Smider, and Peter G. Schultz. 2012. "Synthesis of Bispecific Antibodies Using Genetically Encoded Unnatural Amino Acids." *Journal of the American Chemical Society* 134 (24): 9918–21. <https://doi.org/10.1021/ja303904e>.
- Kim, Eunha, and Heebeom Koo. 2019. "Biomedical Applications of Copper-Free Click Chemistry: In Vitro, in Vivo, and Ex Vivo." *Chemical Science* 10 (34): 7835–51. <https://doi.org/10.1039/c9sc03368h>.
- Kimmerlin, T., and Dieter Seebach. 2005. "'100 Years of Peptide Synthesis': Ligation Methods for Peptide and Protein Synthesis with Applications to β -Peptide Assemblies." *Journal of Peptide Research* 65 (2): 229–60. <https://doi.org/10.1111/j.1399-3011.2005.00214.x>.
- Kiyohara, Masashi, Takashi Nakatomi, Shin Kurihara, Shinya Fushinobu, Hideyuki Suzuki, Tomonari Tanaka, Shin Ichiro Shoda, et al. 2012. " α -N-Acetylgalactosaminidase from Infant-Associated Bifidobacteria Belonging to Novel Glycoside Hydrolase Family 129 Is Implicated in Alternative Mucin Degradation Pathway." *Journal of Biological Chemistry* 287 (1): 693–700. <https://doi.org/10.1074/jbc.M111.277384>.
- Klugman, Keith P., and Steven Black. 2018. "Impact of Existing Vaccines in Reducing Antibiotic Resistance: Primary and Secondary Effects." *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences. <https://doi.org/10.1073/pnas.1721095115>.
- Ko, Wooseok, Rahul Kumar, Sanggil Kim, and Hyun Soo Lee. 2019. "Construction of Bacterial Cells with an Active Transport System for Unnatural Amino Acids." Research-article. *ACS Synthetic Biology* 8 (5): 1195–1203. <https://doi.org/10.1021/acssynbio.9b00076>.
- Kobayashi, Takatsugu, Osamu Nurekil, Ryuichiro Ishitani, Anna Yaremchuk, Michael Tukalo, Stephen Cusack, Kensaku Sakamoto, and Shigeyuki Yokoyama. 2003. "Structural Basis for Orthogonal TRNA Specificities of Tyrosyl-TRNA Synthetases for Genetic Code Expansion." *Nature Structural Biology* 10 (6): 425–32. <https://doi.org/10.1038/nsb934>.

- Kolatkár, A R, and W I Weis. 1996. "Structural Basis of Galactose Recognition by C-Type Animal Lectins." *The Journal of Biological Chemistry* 271 (12): 6679–85. <http://www.ncbi.nlm.nih.gov/pubmed/8636086>.
- Kolev, Joshua N., Jacqueline M. Zaengle, Rajesh Ravikumar, and Rudi Fasan. 2014. "Enhancing the Efficiency and Regioselectivity of P450 Oxidation Catalysts by Unnatural Amino Acid Mutagenesis." *ChemBioChem* 15 (7): 1001–10. <https://doi.org/10.1002/cbic.201400060>.
- Kothari, Neha, Kristopher R. Genschmer, Sudeep Kothari, Jeong Ah Kim, David E. Briles, Dong Kwon Rhee, and Rodney Carbis. 2014. "Preparation and Testing of a Vi Conjugate Vaccine Using Pneumococcal Surface Protein A (PspA) from *Streptococcus Pneumoniae* as the Carrier Protein." *Vaccine* 32 (43): 5755–60. <https://doi.org/10.1016/j.vaccine.2014.08.041>.
- Krassnig, Christian. 2007. *Synthetic Biology - Applying Engineering to Biology - Report of a NEST High-Level Expert Group. The Journal of Pharmacy and Pharmacology*. Vol. 27 Suppl?-. <http://www.ncbi.nlm.nih.gov/pubmed/2700>.
- Krishnakumar, Radha, and Jiqiang Ling. 2014. "Experimental Challenges of Sense Codon Reassignment: An Innovative Approach to Genetic Code Expansion." *FEBS Letters* 588 (3): 383–88. <https://doi.org/10.1016/j.febslet.2013.11.039>.
- Kříž, Zdeněk, Jan Adam, Jana Mrázková, Petros Zotos, Thomais Chatzipavlou, Michaela Wimmerová, and Jaroslav Koča. 2014. "Engineering the *Pseudomonas Aeruginosa* II Lectin: Designing Mutants with Changed Affinity and Specificity." *Journal of Computer-Aided Molecular Design*. <https://doi.org/10.1007/s10822-014-9774-7>.
- Ku, Ti Hsuan, Tiantian Zhang, Hua Luo, Tony M. Yen, Ping Wei Chen, Yuanyuan Han, and Yu Hwa Lo. 2015. "Nucleic Acid Aptamers: An Emerging Tool for Biotechnology and Biomedical Sensing." *Sensors (Switzerland)*. MDPI AG. <https://doi.org/10.3390/s150716281>.
- Kurbatova, I. V. Yakovleva, E. Akhmatova, V. Sukhova, D. V. Yashunsky, Yu E. Tsvetkov, and N. E. Nifantiev. 2019. "Carbohydrate Specificity and Isotypes of Monoclonal and Polyclonal Antibodies to Conjugated Tetrasaccharide, a Synthetic Analogue of Repeating Unit of Capsular Polysaccharide of *Streptococcus Pneumoniae* Serotype 14." *Bulletin of Experimental Biology and Medicine* 166 (4): 477–80. <https://doi.org/10.1007/s10517-019-04376-5>.
- Kushwaha, Rekha, Kim R. Schäfermeyer, and A. Bruce Downie. 2014. "A Protocol for Phage Display and Affinity Selection Using Recombinant Protein Baits." *Journal of Visualized Experiments*, no. 84. <https://doi.org/10.3791/50685>.
- Kuttel, Michelle M., Graham E. Jackson, Mpho Mafata, and Neil Ravenscroft. 2015. "Capsular Polysaccharide Conformations in Pneumococcal Serotypes 19F and 19A." *Carbohydrate Research* 406 (April): 27–33. <https://doi.org/10.1016/j.carres.2014.12.013>.
- Lafond, Betty. 2015. "Revue Science et Avenir Numéro 826." In , 37. https://www.sciencesetavenir.fr/sante/infographie-comment-le-vaccin-protège-l-organisme_30454.
- Lagousi, Theano, Paraskevi Basdeki, John Routsias, and Vana Spoulou. 2019. "Novel Protein-Based Pneumococcal Vaccines: Assessing the Use of Distinct Protein Fragments Instead of Full-Length Proteins as Vaccine Antigens." *Vaccines* 7 (1). <https://doi.org/10.3390/vaccines7010009>.
- Lajoie, Marc J., Alexis J. Rovner, Daniel B. Goodman, Hans Rudolf Aerni, Adrian D. Haimovich, Gleb Kuznetsov, Jaron A. Mercer, et al. 2013. "Genomically Recoded Organisms Expand Biological Functions." *Science* 342 (6156): 357–60. <https://doi.org/10.1126/science.1241459>.
- Lakhin, A V, V Z Tarantul, and L V Gening. 2013. "Aptamers: Problems, Solutions and Prospects." *Acta Naturae* 5 (4): 34–43. <http://www.ncbi.nlm.nih.gov/pubmed/24455181>.
- Lakhtin, Vladimir, Mikhail Lakhtin, and Vladimir Alyoshkin. 2011. "Lectins of Living Organisms. The Overview." *Anaerobe* 17 (6): 452–55. <https://doi.org/10.1016/j.anaerobe.2011.06.004>.
- Lakowicz, Joseph R, and Joseph R Lakowicz. 2006. "Principles of Fluorescence Spectroscopy Third Edition."
- Lam, Sze Kwan, and Tzi Bun Ng. 2011. "Lectins: Production and Practical Applications." *Applied Microbiology and Biotechnology*. <https://doi.org/10.1007/s00253-010-2892-9>.
- Lammers, Michael. 2018. "Expression and Purification of Site-Specifically Lysine-Acetylated and Natively-Folded Proteins for Biophysical Investigations." In *Methods in Molecular Biology*, 1728:169–90. Humana Press Inc. https://doi.org/10.1007/978-1-4939-7574-7_11.
- Larson, Thomas R, and Janet Yother. 2017. "*Streptococcus Pneumoniae* Capsular Polysaccharide Is Linked to Peptidoglycan via a Direct Glycosidic Bond to β -D-N-Acetylglucosamine." *Proceedings of the National Academy of Sciences of the United States of America* 114 (22): 5695–5700. <https://doi.org/10.1073/pnas.1620431114>.

- Lassfolk, Robert, Jani Rahkila, Mikael P. Johansson, Filip S. Ekholm, Johan Wärnå, and Reko Leino. 2019. "Acetyl Group Migration across the Saccharide Units in Oligomannoside Model Compound." *Journal of the American Chemical Society* 141 (4): 1646–54. <https://doi.org/10.1021/jacs.8b11563>.
- Lee, Hyun Soo, and Peter G Schultz. 2008. "Biosynthesis of a Site-Specific DNA Cleaving Protein." *Journal of the American Chemical Society* 130 (40): 13194–95. <https://doi.org/10.1021/ja804653f>.
- Lee, Kyung Jin, Deokhee Kang, and Hee Sung Park. 2019. "Site-Specific Labeling of Proteins Using Unnatural Amino Acids." *Molecules and Cells* 42 (5): 386–96. <https://doi.org/10.14348/molcells.2019.0078>.
- Lee, Lucia H., Xin Xing Gu, and Moon H. Nahm. 2014. "Towards New Broader Spectrum Pneumococcal Vaccines: The Future of Pneumococcal Disease Prevention." *Vaccines*. MDPI AG. <https://doi.org/10.3390/vaccines2010112>.
- Lee, Reiko T, and Yuan C Lee. 2001. "Affinity Enhancement by Multivalent Lectin - Carbohydrate Interaction," 543–44.
- Lee, Shih-Hui, and Ruey-An Doong. 2016. "Design of Size-Tunable Molecularly Imprinted Polymer for Selective Adsorption of Pharmaceuticals and Biomolecules." *J Biosens Bioelectron*. <https://www.omicsonline.org/open-access/design-of-size-tunable-molecularly-imprinted-polymer-for-selective-adsorption-of-pharmaceuticals-and-biomolecules-2155-6210-1000228.pdf>.
- Lee, Taehyung C., Minjin Kang, Chan Hyuk Kim, Peter G. Schultz, Eli Chapman, and Ashok A. Deniz. 2016. "Dual Unnatural Amino Acid Incorporation and Click-Chemistry Labeling to Enable Single-Molecule FRET Studies of P97 Folding." *ChemBioChem* 17 (11): 981–84. <https://doi.org/10.1002/cbic.201500695>.
- Lee, Yuan C., and Reiko T. Lee. 1995. "Carbohydrate-Protein Interactions: Basis of Glycobiology." *Accounts of Chemical Research*. <https://doi.org/10.1021/ar00056a001>.
- Lefeber, Dirk J, Eneko Aldaba Arevalo, Johannes P Kamerling, and Johannes FG Vilegenthart. 2002. "Synthesis of a Hexasaccharide Fragment of the Capsular Polysaccharide of Streptococcus Pneumoniae Type 3." https://www.researchgate.net/publication/27697360_Synthesis_of_a_hexasaccharide_fragment_of_the_capsular_polysaccharide_of_Streptococcus_Pneumoniae_type_3.
- Lei, Huan-Yao, and Chih-Peng Chang. 2007. "Induction of Autophagy by Concanavalin A and Its Application in Anti-Tumor Therapy." *Autophagy* 3 (4): 402–4. <https://doi.org/10.4161/auto.4280>.
- Leman, Raphaël. 2014. "Comment Choisir Un Inhalateur Prêt-à-l'emploi Adapté Dans Le Traitement de l'asthme Chez l'adulte ?"
- Lemke, Edward A, Daniel Summerer, Bernhard H Geierstanger, Scott M Brittain, and Peter G Schultz. 2007. "Control of Protein Phosphorylation with a Genetically Encoded Photocaged Amino Acid." *Nature Chemical Biology* 3 (12): 769–72. <https://doi.org/10.1038/nchembio.2007.44>.
- Li, Minyong, Na Lin, Zhen Huang, Lupei Du, Craig Altier, Hao Fang, and Binghe Wang. 2008. "Selecting Aptamers for a Glycoprotein through the Incorporation of the Boronic Acid Moiety." *Journal of the American Chemical Society* 130 (38): 12636–38. <https://doi.org/10.1021/ja801510d>.
- Li, Yuan, Daniel M. Weinberger, Claudette M. Thompson, Krzysztof Trzciński, and Marc Lipsitch. 2013. "Surface Charge of Streptococcus Pneumoniae Predicts Serotype Distribution." *Infection and Immunity* 81 (12): 4519–24. <https://doi.org/10.1128/IAI.00724-13>.
- Liang, Chaoning, Yi Zhang, Yan Jia, Wenzhao Wang, Youhai Li, Shikun Lu, Jian Ming Jin, and Shuang Yan Tang. 2016. "Engineering a Carbohydrate-Processing Transglycosidase into Glycosyltransferase for Natural Product Glycodiversification." *Scientific Reports* 6 (February). <https://doi.org/10.1038/srep21051>.
- Liberato, Marcelo V, Rodrigo L Silveira, Érica T Prates, Evandro A de Araujo, Vanessa O A Pellegrini, Cesar M Camilo, Marco A Kadowaki, et al. 2016. "Molecular Characterization of a Family 5 Glycoside Hydrolase Suggests an Induced-Fit Enzymatic Mechanism." *Scientific Reports* 6 (April): 23473. <https://doi.org/10.1038/srep23473>.
- Licciardi, Paul V., Anne Balloch, Fiona M. Russell, Robert L. Burton, Jisheng Lin, Moon H. Nahm, Edward K. Mulholland, and Mimi L.K. Tang. 2012. "Pneumococcal Polysaccharide Vaccine at 12 Months of Age Produces Functional Immune Responses." *Journal of Allergy and Clinical Immunology* 129 (3). <https://doi.org/10.1016/j.jaci.2011.11.043>.
- Lin, Andrew E., and Qing Lin. 2019. "Rapid Identification of Functional Pyrrolysyl-TRNA Synthetases via Fluorescence-Activated Cell Sorting." *International Journal of Molecular Sciences* 20 (1). <https://doi.org/10.3390/ijms20010029>.
- Lin, Haiying, Zilin Lin, Chun Meng, Jingzhong Huang, and Yanghao Guo. 2010. "Preparation and Immunogenicity of Capsular Polysaccharide-Surface Adhesin A (PsaA) Conjugate of Streptococcus Pneumoniae." *Immunobiology* 215 (7): 545–50. <https://doi.org/10.1016/j.imbio.2009.08.008>.

- Lin, Haiying, Yonghui Peng, Zi Lin Lin, Shuangling Zhang, and Yanghao Guo. 2015. "Development of a Conjugate Vaccine against Invasive Pneumococcal Disease Based on Capsular Polysaccharides Coupled with PspA/Family 1 Protein of *Streptococcus Pneumoniae*." *Microbial Pathogenesis* 83–84 (June): 35–40. <https://doi.org/10.1016/j.micpath.2015.04.006>.
- Ling, Jiqiang, Noah Reynolds, and Michael Ibba. 2009. "Aminoacyl-TRNA Synthesis and Translational Quality Control." *Annual Review of Microbiology* 63 (1): 61–78. <https://doi.org/10.1146/annurev.micro.091208.073210>.
- Link, A. James, and David A. Tirrell. 2005. "Reassignment of Sense Codons in Vivo." *Methods* 36 (3): 291–98. <https://doi.org/10.1016/j.ymeth.2005.04.005>.
- Lis, Halina, and Nathan Sharon. 1998. "Lectins : Carbohydrate-Specific Proteins That Mediate Cellular Recognition †" 2665 (94).
- Liu, C. C., A. V. Mack, M.-L. Tsao, J. H. Mills, H. S. Lee, H. Choe, M. Farzan, P. G. Schultz, and V. V. Smider. 2008. "Protein Evolution with an Expanded Genetic Code." *Proceedings of the National Academy of Sciences* 105 (46): 17688–93. <https://doi.org/10.1073/pnas.0809543105>.
- Liu, C C, A V Mack, E M Brustad, J H Mills, D Groff, V V Smider, and P G Schultz. 2009. "Evolution of Proteins with Genetically Encoded "Chemical Warheads"." *J Am Chem Soc*, 9616–17. <https://doi.org/10.1021/ja902985e>.
- Liu, Chang C., Susan E. Cellitti, Bernhard H. Geierstanger, and Peter G. Schultz. 2009. "Efficient Expression of Tyrosine-Sulfated Proteins in *E. Coli* Using an Expanded Genetic Code." *Nature Protocols* 4 (12): 1784–89. <https://doi.org/10.1038/nprot.2009.188>.
- Liu, Chang C., and Peter G. Schultz. 2010. "Adding New Chemistries to the Genetic Code." *Annual Review of Biochemistry* 79 (1): 413–44. <https://doi.org/10.1146/annurev.biochem.052308.105824>.
- Liu, Chang C, and Peter G Schultz. 2010. "Adding New Chemistries to the Genetic Code." *Annual Review of Biochemistry* 79 (1): 413–44. <https://doi.org/10.1146/annurev.biochem.052308.105824>.
- Liu, David R., and Peter G. Schultz. 1999. "Progress toward the Evolution of an Organism with an Expanded Genetic Code." *Proceedings of the National Academy of Sciences of the United States of America* 96 (9): 4780–85. <https://doi.org/10.1073/pnas.96.9.4780>.
- Liu, Wei, Guojun Yu, Wenhui Yu, Xiangdong Ye, Yanxia Jin, Alok Shrestha, Qing Yang, and Hui Sun. 2017. "Autophagy Inhibits Apoptosis Induced by *Agrocybe Aegerita* Lectin in Hepatocellular Carcinoma." *Anti-Cancer Agents in Medicinal Chemistry* 17 (2): 221–29. <http://www.ncbi.nlm.nih.gov/pubmed/27039921>.
- Liu, Zhen, and Hui He. 2017. "Synthesis and Applications of Boronate Affinity Materials: From Class Selectivity to Biomimetic Specificity." *Accounts of Chemical Research* 50 (9): 2185–93. <https://doi.org/10.1021/acs.accounts.7b00179>.
- Lo, Stephanie W., Rebecca A. Gladstone, Andries J. van Tonder, John A. Lees, Mignon du Plessis, Rachel Benisty, Noga Givon-Lavi, et al. 2019. "Pneumococcal Lineages Associated with Serotype Replacement and Antibiotic Resistance in Childhood Invasive Pneumococcal Disease in the Post-PCV13 Era: An International Whole-Genome Sequencing Study." *The Lancet Infectious Diseases* 19 (7): 759–69. [https://doi.org/10.1016/S1473-3099\(19\)30297-X](https://doi.org/10.1016/S1473-3099(19)30297-X).
- Loris, Remy, Denis Tielker, Karl-Erich Jaeger, and Lode Wyns. 2003. "Structural Basis of Carbohydrate Recognition by the Lectin LecB from *Pseudomonas Aeruginosa*." *Journal of Molecular Biology* 331 (4): 861–70. [https://doi.org/10.1016/s0022-2836\(03\)00754-x](https://doi.org/10.1016/s0022-2836(03)00754-x).
- Lu, Hua; Zhou, Quan; Deshmukh Vishal. 2014. "Targeting Human C-Type Lectin-Like Molecule-1 (CLL1) with a Bispecific Antibody for Acute Myeloid Leukemia Immunotherapy." *Angew Chem Int Ed Engl* 53 (37): 9841–45. <https://doi.org/10.1002/anie.201405353>.Targeting.
- Lu, Ying Wei, Chih Wei Chien, Po Chiao Lin, Li De Huang, Chang Yang Chen, Sz Wei Wu, Chia Li Han, Kay Hooi Khoo, Chun Cheng Lin, and Yu Ju Chen. 2013. "BAD-Lectins: Boronic Acid-Decorated Lectins with Enhanced Binding Affinity for the Selective Enrichment of Glycoproteins." *Analytical Chemistry* 85 (17): 8268–76. <https://doi.org/10.1021/ac401581u>.
- Luck, Linda A, and Joseph J Falke. 1991. "F NMR Studies of the Galactose Chemosensory Receptor 1 Sugar Binding Yields a Global Structural Change †." *Biochemistry* 30 (17): 4248–56.
- Lusvarghi, Sabrina, and Carole A Bewley. 2016. "Griffithsin: An Antiviral Lectin with Outstanding Therapeutic Potential." *Viruses* 8 (10). <https://doi.org/10.3390/v8100296>.
- Lykke-Andersen, Søren, and Torben Heick Jensen. 2015. "Nonsense-Mediated mRNA Decay: An Intricate Machinery That Shapes Transcriptomes." *Nature Reviews Molecular Cell Biology* 16 (11): 665–77. <https://doi.org/10.1038/nrm4063>.

- Lyu, Su-Yun, and Won-Bong Park. 2007. "Effects of Korean Mistletoe Lectin (*Viscum Album Coloratum*) on Proliferation and Cytokine Expression in Human Peripheral Blood Mononuclear Cells and T-Lymphocytes." *Archives of Pharmacal Research* 30 (10): 1252–64. <https://doi.org/10.1007/bf02980266>.
- Ma, Hairong, Xin Yang, Zhuo Lu, Nan Liu, and Yijun Chen. 2014. "The 'Gate Keeper' Role of Trp222 Determines the Enantioselectivity of Diketoreductase toward 2-Chloro-1-Phenylethanone." *PLoS ONE* 9 (7). <https://doi.org/10.1371/journal.pone.0103792>.
- Ma, Zhongrui, Huajie Zhang, Peng George Wang, Xian Wei Liu, and Min Chen. 2018. "Peptide Adjacent to Glycosylation Sites Impacts Immunogenicity of Glycoconjugate Vaccine." *Oncotarget* 9 (1): 75–82. <https://doi.org/10.18632/oncotarget.19944>.
- Mackenzie, Grant A., Amanda J. Leach, Jonathan R. Carapetis, Janelle Fisher, and Peter S. Morris. 2010. "Epidemiology of Nasopharyngeal Carriage of Respiratory Bacterial Pathogens in Children and Adults: Cross-Sectional Surveys in a Population with High Rates of Pneumococcal Disease." *BMC Infectious Diseases* 10 (October). <https://doi.org/10.1186/1471-2334-10-304>.
- Maini, Rumit, Dan T. Nguyen, Shengxi Chen, Larisa M. Dedkova, Sandipan Roy Chowdhury, Rafael Alcala-Torano, and Sidney M. Hecht. 2013. "Incorporation of β -Amino Acids into Dihydrofolate Reductase by Ribosomes Having Modifications in the Peptidyltransferase Center." *Bioorganic and Medicinal Chemistry* 21 (5): 1088–96. <https://doi.org/10.1016/j.bmc.2013.01.002>.
- Mallia, A. Krishna, Greg T. Hermanson, Randall I. Krohn, Edward K. Fujimoto, and Paul K. Smith. 1981. "Preparation and Use of a Boronic Acid Affinity Support for Separation and Qualification of Glycosylated Hemoglobins." *Analytical Letters* 14 (8): 649–61. <https://doi.org/10.1080/00032718108055476>.
- Manso, Ana Sousa, Melissa H. Chai, John M. Atack, Leonardo Furi, Megan De Ste Croix, Richard Haigh, Claudia Trappetti, et al. 2014. "A Random Six-Phase Switch Regulates Pneumococcal Virulence via Global Epigenetic Changes." *Nature Communications* 5. <https://doi.org/10.1038/ncomms6055>.
- Marks, Laura R., Ryan M. Reddinger, and Anders P. Hakansson. 2014. "Biofilm Formation Enhances Fomite Survival of *Streptococcus Pneumoniae* and *Streptococcus Pyogenes*." *Infection and Immunity* 82 (3): 1141–46. <https://doi.org/10.1128/IAI.01310-13>.
- Martínez-Alarcón, Dania, Alejandro Blanco-Labra, and Teresa García-Gasca. 2018. "Expression of Lectins in Heterologous Systems." *International Journal of Molecular Sciences* 19 (2). <https://doi.org/10.3390/ijms19020616>.
- Matos, Maria J., Bruno L. Oliveira, Nuria Martínez-Sáez, Ana Guerreiro, Pedro M.S.D. Cal, Jean Bertoldo, María Maneiro, et al. 2018. "Chemo- and Regioselective Lysine Modification on Native Proteins." *Journal of the American Chemical Society* 140 (11): 4004–17. <https://doi.org/10.1021/jacs.7b12874>.
- Mawas, Fatme, Jutta Niggemann, Christopher Jones, Michael J. Corbel, Johannes P. Kamerling, and Johannes F.G. Vliegthart. 2002. "Immunogenicity in a Mouse Model of a Conjugate Vaccine Made with a Synthetic Single Repeating Unit of Type 14 Pneumococcal Polysaccharide Coupled to CRM197." *Infection and Immunity* 70 (9): 5107–14. <https://doi.org/10.1128/IAI.70.9.5107-5114.2002>.
- Mayer, Clemens. 2019. "Selection, Addition and Catalysis: Emerging Trends for the Incorporation of Noncanonical Amino Acids into Peptides and Proteins in Vivo." *ChemBioChem* 20 (11): 1357–64. <https://doi.org/10.1002/cbic.201800733>.
- Mayer, Clemens, Christopher Dulson, Eswar Reddem, Andy Mark W.H. Thunnissen, and Gerard Roelfes. 2019. "Directed Evolution of a Designer Enzyme Featuring an Unnatural Catalytic Amino Acid." *Angewandte Chemie - International Edition* 58 (7): 2083–87. <https://doi.org/10.1002/anie.201813499>.
- McDonald, Rhona E., David J. Hughes, and Benjamin G. Davis. 2004. "Modular Control of Lectin Function: Redox-Switchable Agglutination." *Angewandte Chemie - International Edition*. <https://doi.org/10.1002/anie.200353320>.
- Mehl, Ryan A., J. Christopher Anderson, Stephen W. Santoro, Lei Wang, Andrew B. Martin, David S. King, David M. Horn, and Peter G. Schultz. 2003. "Generation of a Bacterium with a 21 Amino Acid Genetic Code." *Journal of the American Chemical Society* 125 (4): 935–39. <https://doi.org/10.1021/ja0284153>.
- Melançon, Charles E., and Peter G. Schultz. 2009. "One Plasmid Selection System for the Rapid Evolution of Aminoacyl-TRNA Synthetases." *Bioorganic and Medicinal Chemistry Letters* 19 (14): 3845–47. <https://doi.org/10.1016/j.bmcl.2009.04.007>.
- Melnikov, Sergey V., and Dieter Söll. 2019. "Aminoacyl-TRNA Synthetases and TRNAS for an Expanded Genetic Code: What Makes Them Orthogonal?" *International Journal of Molecular Sciences* 20 (8). <https://doi.org/10.3390/ijms20081929>.
- Merrifield, R. B. 1963. "Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide." *Journal of the American Chemical Society* 85 (1): 2149–54. <https://doi.org/10.1021/ja00291a021>.

- Society* 85 (14): 2149–54. <https://doi.org/10.1021/ja00897a025>.
- Michon, Francis, Peter C. Fusco, Conceição A.S.A. Minetti, Maryline Laude-Sharp, Catherine Uitz, Chun Hsien Huang, Anello J. D'Ambra, et al. 1998. "Multivalent Pneumococcal Capsular Polysaccharide Conjugate Vaccines Employing Genetically Detoxified Pneumolysin as a Carrier Protein." *Vaccine* 16 (18): 1732–41. [https://doi.org/10.1016/S0264-410X\(98\)00225-4](https://doi.org/10.1016/S0264-410X(98)00225-4).
- Micoli, Francesca, Roberto Adamo, and Paolo Costantino. 2018. "Protein Carriers for Glycoconjugate Vaccines: History, Selection Criteria, Characterization and New Trends." *Molecules* 23 (6): 1–18. <https://doi.org/10.3390/molecules23061451>.
- Middle, F. A., A. Bannister, A. J. Bellingham, and P. D.G. Dean. 1983. "Separation of Glycosylated Haemoglobins Using Immobilized Phenylboronic Acid. Effect of Ligand Concentration, Column Operating Conditions, and Comparison with Ion-Exchange and Isoelectric-Focusing." *Biochemical Journal* 209 (3): 771–79. <https://doi.org/10.1042/bj2090771>.
- Mills, Kenneth V, Margaret A Johnson, and Francine B Perler. 2014. "Protein Splicing: How Inteins Escape from Precursor Proteins." *The Journal of Biological Chemistry* 289 (21): 14498–505. <https://doi.org/10.1074/jbc.R113.540310>.
- Minks, Caroline, Stefan Alefelder, Luis Moroder, Robert Huber, and Nediljko Budisa. 2000. "Towards New Protein Engineering: In Vivo Building and Folding of Protein Shuttles for Drug Delivery and Targeting by the Selective Pressure Incorporation (SPI) Method." *Tetrahedron* 56 (48): 9431–42. [https://doi.org/10.1016/S0040-4020\(00\)00827-9](https://doi.org/10.1016/S0040-4020(00)00827-9).
- Mirande, Marc. 2017. "Macromolecular Protein Complexes" 83: 505–22. <https://doi.org/10.1007/978-3-319-46503-6>.
- Mirts, Evan N, Ambika Bhagi-Damodaran, and Yi Lu. 2019. "Understanding and Modulating Metalloenzymes with Unnatural Amino Acids, Non-Native Metal Ions, and Non-Native Metallocofactors." *Accounts of Chemical Research* 52 (4): 935–44. <https://doi.org/10.1021/acs.accounts.9b00011>.
- Mishra, Sushil Kumar, Jan Adam, Michaela Wimmerová, and Jaroslav Koča. 2012. "In Silico Mutagenesis and Docking Study of Ralstonia Solanacearum RSL Lectin: Performance of Docking Software to Predict Saccharide Binding." *Journal of Chemical Information and Modeling*. <https://doi.org/10.1021/ci200529n>.
- Mitchell, Edward, Corinne Houles, Dvora Sudakevitz, Michaela Wimmerova, Catherine Gautier, Serge Pérez, Albert M Wu, Nechama Gilboa-Garber, and Anne Imberty. 2002. "Structural Basis for Oligosaccharide-Mediated Adhesion of Pseudomonas Aeruginosa in the Lungs of Cystic Fibrosis Patients." *Nature Structural Biology* 9 (12): 918–21. <https://doi.org/10.1038/nsb865>.
- Miyake-Stoner, Shigeki J., Andrew M. Miller, Jared T. Hammill, Jennifer C. Peeler, Kenneth R. Hess, Ryan A. Mehl, and Scott H. Brewer. 2009. "Probing Protein Folding Using Site-Specifically Encoded Unnatural Amino Acids as FRET Donors with Tryptophan." *Biochemistry* 48 (25): 5953–62. <https://doi.org/10.1021/bi900426d>.
- Monahan, Sarah L., Henry A. Lester, and Dennis A. Dougherty. 2003. "Site-Specific Incorporation of Unnatural Amino Acids into Receptors Expressed in Mammalian Cells." *Chemistry and Biology* 10 (6): 573–80. [https://doi.org/10.1016/S1074-5521\(03\)00124-8](https://doi.org/10.1016/S1074-5521(03)00124-8).
- Mooibroek, Tiddo J., Juan M. Casas-Solvas, Robert L. Harniman, Charles M. Renney, Tom S. Carter, Matthew P. Crump, and Anthony P. Davis. 2016. "A Threading Receptor for Polysaccharides." *Nature Chemistry* 8 (1): 69–74. <https://doi.org/10.1038/nchem.2395>.
- Moore, Matthew, and Tamara Pilishvilli. 2010. "Pinkbook | Pneumococcal | Epidemiology of Vaccine Preventable Diseases | CDC." 2010. <https://www.cdc.gov/vaccines/pubs/pinkbook/pneumo.html>.
- Morais, Teixeira, and Suarez. 2019. "Next-Generation Whole-Cell Pneumococcal Vaccine." *Vaccines* 7 (4): 151. <https://doi.org/10.3390/vaccines7040151>.
- Morais, Victor, Valerie Dee, and Norma Suárez. 2018. "Purification of Capsular Polysaccharides of Streptococcus Pneumoniae: Traditional and New Methods." *Frontiers in Bioengineering and Biotechnology*. Frontiers Media S.A. <https://doi.org/10.3389/fbioe.2018.00145>.
- Morelli, Laura, Silvia Fallarini, Grazia Lombardi, Cinzia Colombo, Luigi Lay, and Federica Compostella. 2018. "Synthesis and Biological Evaluation of a Trisaccharide Repeating Unit Derivative of Streptococcus Pneumoniae 19A Capsular Polysaccharide." *Bioorganic & Medicinal Chemistry* 26 (21): 5682–90. <https://doi.org/10.1016/j.bmc.2018.10.016>.
- Mukai, Takahito, Hiroko Hoshi, Kazumasa Ohtake, Mihoko Takahashi, Atsushi Yamaguchi, Akiko Hayashi, Shigeyuki Yokoyama, and Kensaku Sakamoto. 2015. "Highly Reproductive Escherichia Coli Cells with No Specific Assignment to the UAG Codon." *Scientific Reports* 5 (May): 9699. <https://doi.org/10.1038/srep09699>.
- Mukai, Takahito, Atsushi Yamaguchi, Kazumasa Ohtake, Mihoko Takahashi, Akiko Hayashi, Fumie Iraha, Satoshi Kira, et al.

2015. "Reassignment of a Rare Sense Codon to a Non-Canonical Amino Acid in *Escherichia Coli*." *Nucleic Acids Research* 43 (16): 8111–22. <https://doi.org/10.1093/nar/gkv787>.
- Murakami, Hiroshi, Hirohide Saito, and Hiroaki Suga. 2003. "A Versatile TRNA Aminoacylation Catalyst Based on RNA." *Chemistry and Biology* 10 (7): 655–62. [https://doi.org/10.1016/S1074-5521\(03\)00145-5](https://doi.org/10.1016/S1074-5521(03)00145-5).
- Murray, Christopher I., Maartje Westhoff, Jodene Eldstrom, Emely Thompson, Robert Emes, and David Fedida. 2016. "Unnatural Amino Acid Photo-Crosslinking of the IKs Channel Complex Demonstrates a KCNE1:KCNQ1 Stoichiometry of up to 4:4." *ELife* 5 (January). <https://doi.org/10.7554/eLife.11815>.
- Musher, Daniel M., Susan B. Manoff, Charlie Liss, Richard D. McFetridge, Rocio D. Marchese, Bonnie Bushnell, Frances Alvarez, Carla Painter, Michael D. Blum, and Jeffrey L. Silber. 2010. "Safety and Antibody Response, Including Antibody Persistence for 5 Years, after Primary Vaccination or Revaccination with Pneumococcal Polysaccharide Vaccine in Middle-Aged and Older Adults." *The Journal of Infectious Diseases* 201 (4): 516–24. <https://doi.org/10.1086/649839>.
- Nagae, Masamichi, Kana Morita-Matsumoto, Masaki Kato, Mika Kato Kaneko, Yukinari Kato, and Yoshiki Yamaguchi. 2014. "A Platform of C-Type Lectin-like Receptor CLEC-2 for Binding O-Glycosylated Podoplanin and Nonglycosylated Rhodocytin." *Structure* 22 (12): 1711–21. <https://doi.org/10.1016/j.str.2014.09.009>.
- Nakahara, Susumu, and Avraham Raz. 2008. "Biological Modulation by Lectins and Their Ligands in Tumor Progression and Metastasis." *Anti-Cancer Agents in Medicinal Chemistry* 8 (1): 22–36. <http://www.ncbi.nlm.nih.gov/pubmed/18220503>.
- NCT03480802. 2018. A Phase 3, Multicenter, Randomized, Double-Blind, Active Comparator-Controlled Study to Evaluate the Safety, Tolerability, and Immunogenicity of V114 Followed by Administration of PNEUMOVAX™23 Eight Weeks Later in Adults Infected With HIV (PNEU-WAY), issued 2018.
- Nehmann, Nina, Ulrika M Schade, Uwe Pfüller, Melitta Schachner, and Udo Schumacher. 2009. "Mistletoe Lectin Binds to Multidrug Resistance-Associated Protein MRP5." *Anticancer Research* 29 (12): 4941–48. <http://www.ncbi.nlm.nih.gov/pubmed/20044600>.
- Nesin, M, M Ramirez, and A Tomasz. 1998. "Capsular Transformation of a Multidrug-Resistant *Streptococcus Pneumoniae* in Vivo." *The Journal of Infectious Diseases* 177 (3): 707–13. <https://doi.org/10.1086/514242>.
- Neumann, Heinz, Sew Y. Peak-Chew, and Jason W. Chin. 2008. "Genetically Encoding Nε-Acetyllysine in Recombinant Proteins." *Nature Chemical Biology* 4 (4): 232–34. <https://doi.org/10.1038/nchembio.73>.
- Nguyen, Duy P., Mohan Mahesh, Simon J. Elsässer, Susan M. Hancock, Chayasith Uttamapinant, and Jason W. Chin. 2014. "Genetic Encoding of Photocaged Cysteine Allows Photoactivation of TEV Protease in Live Mammalian Cells." *Journal of the American Chemical Society* 136 (6): 2240–43. <https://doi.org/10.1021/ja412191m>.
- Nikos, Parisis. 2012. "Xenopus Laevis as a Model System." *Materials and Methods* 2 (December). <https://doi.org/10.13070/mm.en.2.151>.
- Nilsson, Bradley L, Matthew B Soellner, and Ronald T Raines. 2005. "Chemical Synthesis of Proteins." *Annual Review Biophysics and Biomolecular Structure* 34: 91–118. <https://doi.org/10.1146/annurev.biophys.34.040204.144700>.
- Nishat, Sharmeen, and Peter R. Andreana. 2016. "Entirely Carbohydrate-Based Vaccines: An Emerging Field for Specific and Selective Immune Responses." *Vaccines*. MDPI AG. <https://doi.org/10.3390/vaccines4020019>.
- Nishiguchi, Mitsuru, Kazumasa Yoshida, Toshiro Sumizono, and Kiyoshi Tazaki. 1997. "Studies by Site-Directed Mutagenesis of the Carbohydrate-Binding Properties of a Bark Lectin from Robinia Pseudoacacia." *FEBS Letters*. [https://doi.org/10.1016/S0014-5793\(97\)00068-9](https://doi.org/10.1016/S0014-5793(97)00068-9).
- Nishio, Motohiro. 2011. "The CH/π Hydrogen Bond in Chemistry. Conformation, Supramolecules, Optical Resolution and Interactions Involving Carbohydrates." *Physical Chemistry Chemical Physics*. <https://doi.org/10.1039/c1cp20404a>.
- Nokso-Koivisto, Johanna, Tal Marom, and Tasnee Chonmaitree. 2015. "Importance of Viruses in Acute Otitis Media." *Current Opinion in Pediatrics*. Lippincott Williams and Wilkins. <https://doi.org/10.1097/MOP.0000000000000184>.
- Nonaka, Motohiro, Bruce Yong Ma, Ryuuya Murai, Natsuko Nakamura, Makoto Baba, Nobuko Kawasaki, Keiko Hodohara, Shinji Asano, and Toshiyuki Kawasaki. 2008. "Glycosylation-Dependent Interactions of C-Type Lectin DC-SIGN with Colorectal Tumor-Associated Lewis Glycans Impair the Function and Differentiation of Monocyte-Derived Dendritic Cells." *Journal of Immunology (Baltimore, Md. : 1950)* 180 (5): 3347–56. <https://doi.org/10.4049/jimmunol.180.5.3347>.
- Nozawa, Kayo, Patrick O Donoghue, Sarath Gundllapalli, Yuhei Arais, Takuya Umehara, Dieter Soll, and Osamu Nureki.

2009. "Pyrrolysyl-TRNA Synthetase : TRNA Structure Reveals the Molecular Basis of Orthogonality" 457 (7233): 1163–67. <https://doi.org/10.1038/nature07611>. Pyrrolysyl-tRNA.
- Numminen, Elina, Claire Chewapreecha, Claudia Turner, David Goldblatt, Francois Nosten, Stephen D. Bentley, Paul Turner, and Jukka Corander. 2015. "Climate Induces Seasonality in Pneumococcal Transmission." *Scientific Reports* 5 (June). <https://doi.org/10.1038/srep11344>.
- Nyamai, Dorothy Wavinya, and Özlem Tastan Bishop. 2019. "Aminoacyl TRNA Synthetases as Malarial Drug Targets: A Comparative Bioinformatics Study 11 Medical and Health Sciences 1108 Medical Microbiology 06 Biological Sciences 0601 Biochemistry and Cell Biology." *Malaria Journal* 18 (1). <https://doi.org/10.1186/s12936-019-2665-6>.
- Oba, Takanori, Yoshiki Andachi, Akira Muto, and Syozo Osawa. 1991. "CGG: An Unassigned or Nonsense Codon in *Mycoplasma Capricolum*." *Proceedings of the National Academy of Sciences of the United States of America* 88 (3): 921–25. <https://doi.org/10.1073/pnas.88.3.921>.
- Ohuchi, Masaki, Hiroshi Murakami, and Hiroaki Suga. 2007. "The Flexizyme System: A Highly Flexible TRNA Aminoacylation Tool for the Translation Apparatus." *Current Opinion in Chemical Biology* 11 (5): 537–42. <https://doi.org/10.1016/j.cbpa.2007.08.011>.
- Olausson, Johan, Eva Aström, Bengt-Harald Jonsson, Lena A E Tibell, and Peter Pålsson. 2011. "Production and Characterization of a Monomeric Form and a Single-Site Form of Aleuria Aurantia Lectin." *Glycobiology* 21 (1): 34–44. <https://doi.org/10.1093/glycob/cwq129>.
- Ollikainen, Noah, René M. de Jong, and Tanja Kortemme. 2015. "Coupling Protein Side-Chain and Backbone Flexibility Improves the Re-Design of Protein-Ligand Specificity." Edited by Robert L Jernigan. *PLOS Computational Biology* 11 (9): e1004335. <https://doi.org/10.1371/journal.pcbi.1004335>.
- Ong, Edgar, Jeffrey M. Greenwood, Neil R. Gilkes, Douglas G. Kilburn, Robert C. Miller, and R. Anthony J. Warren. 1989. "The Cellulose-Binding Domains of Cellulases: Tools for Biotechnology." *Trends in Biotechnology*. [https://doi.org/10.1016/0167-7799\(89\)90014-0](https://doi.org/10.1016/0167-7799(89)90014-0).
- Orntoft, Torben F., and Else Marie Vestergaard. 1999. "Clinical Aspects of Altered Glycosylation of Glycoproteins in Cancer." *Electrophoresis*. Wiley-VCH Verlag. [https://doi.org/10.1002/\(SICI\)1522-2683\(19990201\)20:2<362::AID-ELPS362>3.0.CO;2-V](https://doi.org/10.1002/(SICI)1522-2683(19990201)20:2<362::AID-ELPS362>3.0.CO;2-V).
- Örtqvist, Åke, Jonas Hedlund, and Mats Kalin. 2005. "Streptococcus Pneumoniae: Epidemiology, Risk Factors, and Clinical Features." *Seminars in Respiratory and Critical Care Medicine*. <https://doi.org/10.1055/s-2005-925523>.
- Osbourn, Anne E., Paul E. O'Maille, Susan J. Rosser, and Keith Lindsey. 2012. "Meetings Synthetic Biology." *New Phytologist* 196 (June): 671–77.
- Pak, Daewoo, Yunsoo Kim, and Zachary F. Burton. 2018. "Aminoacyl-TRNA Synthetase Evolution and Sectoring of the Genetic Code." *Transcription*, May, 1–15. <https://doi.org/10.1080/21541264.2018.1467718>.
- Palacios, R. 1982. "Concanavalin A Triggers T Lymphocytes by Directly Interacting with Their Receptors for Activation." *Journal of Immunology (Baltimore, Md. : 1950)* 128 (1): 337–42. <http://www.ncbi.nlm.nih.gov/pubmed/6459373>.
- Pappin, Brighid, Milton J., and Todd A. 2012. "Boron-Carbohydrate Interactions." In *Carbohydrates - Comprehensive Studies on Glycobiology and Glycotechnology*. InTech. <https://doi.org/10.5772/50630>.
- Pappin, Brighid, Milton J Kiefel, and Todd A. Houston. 2012. "Boron Carbohydrate Interactions." *Carbohydrates – Comprehensive Studies on Glycobiology and Glycotechnology* 59 (2): 577–84.
- Parmpi, Paraskevi, and Peter Kofinas. 2004. "Biomimetic Glucose Recognition Using Molecularly Imprinted Polymer Hydrogels." *Biomaterials* 25 (10): 1969–73. <https://doi.org/10.1016/j.biomaterials.2003.08.025>.
- Pasteur, Chamberland, and Roux. 1881. "Sur Une Maladie Nouvelle Provoquée Par La Salive d'un Enfant Mort de La Rage." *Compt. Rend.*
- Patterson, David M., Lidia A. Nazarova, and Jennifer A. Prescher. 2014. "Finding the Right (Bioorthogonal) Chemistry." *ACS Chemical Biology* 9 (3): 592–605. <https://doi.org/10.1021/cb400828a>.
- Pauksens, Karlis, Anna C. Nilsson, Magalie Caubet, Thierry G. Pascal, Pascale Van Belle, Jan T. Poolman, Pierre G. Vandepapelière, Vincent Verlant, and Peter E. Vink. 2014. "Randomized Controlled Study of the Safety and Immunogenicity of Pneumococcal Vaccine Formulations Containing PhtD and Detoxified Pneumolysin with Alum or Adjuvant System AS02Vin Elderly Adults." *Clinical and Vaccine Immunology* 21 (5): 651–60. <https://doi.org/10.1128/CVI.00807-13>.

- Pauling, Linus. 1963. "The Nature of the Chemical Bond and the Structure of Molecules and Crystals." *Cornell University Press: Ithaca, 3rd Edition* 30 (1): 30. <https://doi.org/10.1002/jps.3030300111>.
- Perciani, Catia T., Giovana C. Barazzone, Cibelly Goulart, Eneas Carvalho, Joaquin Cabrera-Crespo, Viviane M. Gonçalves, Luciana C.C. Leite, and Martha M. Tanizaki. 2013. "Conjugation of Polysaccharide 6B from *Streptococcus Pneumoniae* with Pneumococcal Surface Protein A: PspA Conformation and Its Effect on the Immune Response." *Clinical and Vaccine Immunology* 20 (6): 818–26. <https://doi.org/10.1128/CVI.00754-12>.
- Pereira, Clancy L., Andreas Geissner, Chakkumkal Anish, and Peter H. Seeberger. 2015. "Chemical Synthesis Elucidates the Immunological Importance of a Pyruvate Modification in the Capsular Polysaccharide of *Streptococcus Pneumoniae* Serotype 4." *Angewandte Chemie - International Edition* 54 (34): 10016–19. <https://doi.org/10.1002/anie.201504847>.
- Perona, John J., and Ita Gruic-Sovulj. 2014. "Synthetic and Editing Mechanisms of Aminoacyl-TRNA Synthetases." *Topics in Current Chemistry*. https://doi.org/10.1007/128_2013_456.
- Pichichero, Michael E. 2013. "Protein Carriers of Conjugate Vaccines Characteristics, Development, and Clinical Trials." *Human Vaccines and Immunotherapeutics*. <https://doi.org/10.4161/hv.26109>.
- Pichichero, Michael E., M. Nadeem Khan, and Qingfu Xu. 2016a. "Next Generation Protein Based *Streptococcus Pneumoniae* Vaccines." *Human Vaccines and Immunotherapeutics*. Taylor and Francis Inc. <https://doi.org/10.1080/21645515.2015.1052198>.
- Pichichero, Michael E, M Nadeem Khan, and Qingfu Xu. 2016b. "Next Generation Protein Based *Streptococcus Pneumoniae* Vaccines" 12 (1): 194–205.
- Pillot, Aline, Alain Defontaine, Amina Fateh, Annie Lambert, Maruthi Prasanna, Mathieu Fanuel, Muriel Pipelier, et al. 2019. "Site-Specific Conjugation for Fully Controlled Glycoconjugate Vaccine Preparation." *Frontiers in Chemistry* 7 (November). <https://doi.org/10.3389/fchem.2019.00726>.
- Pine, M. J. 1978. "Comparative Physiological Effects of Incorporated Amino Acid Analogs in *Escherichia Coli*." *Antimicrobial Agents and Chemotherapy* 13 (4): 676–85. <https://doi.org/10.1128/AAC.13.4.676>.
- Pless, Stephan A, and Christopher A Ahern. 2013. "Unnatural Amino Acids as Probes of Ligand-Receptor Interactions and Their Conformational Consequences," 211–32. <https://doi.org/10.1146/annurev-pharmtox-011112-140343>.
- Plumbridge, Jacqueline, and Dieter Söll. 1987. "The Effect of Dam Methylation on the Expression of GlnS in *E. Coli*." *Biochimie* 69 (5): 539–41. [https://doi.org/10.1016/0300-9084\(87\)90091-5](https://doi.org/10.1016/0300-9084(87)90091-5).
- Pozsgay, Vince, Chiayung Chu, Lewis Pannell, Jennifer Wolfe, John B. Robbins, and Rachel Schneerson. 1999. "Protein Conjugates of Synthetic Saccharides Elicit Higher Levels of Serum IgG Lipopolysaccharide Antibodies in Mice than Do Those of the O-Specific Polysaccharide from *Shigella Dysenteriae* Type 1." *Proceedings of the National Academy of Sciences of the United States of America* 96 (9): 5194–97. <https://doi.org/10.1073/pnas.96.9.5194>.
- Prasad Chakraborty, Subhankari, Panchanan Pramanik, and Somenath Roy. 2012. "A Review on Emergence of Antibiotic Resistant *Staphylococcus Aureus* and Role of Chitosan Nanoparticle in Drug Delivery."
- Prasanna, Maruthi, Daphnée Soulard, Emilie Camberlein, Nicolas Ruffier, Annie Lambert, François Trottein, Noemi Csaba, and Cyrille Grandjean. 2019. "Semisynthetic Glycoconjugate Based on Dual Role Protein/PsaA as a Pneumococcal Vaccine." *European Journal of Pharmaceutical Sciences* 129 (March): 31–41. <https://doi.org/10.1016/j.ejps.2018.12.013>.
- Presolski, S. L., V. P. Hong, and M. G. Finn. 2011. "Copper-Catalyzed Azide–Alkyne Click Chemistry for Bioconjugation." *Current Protocols in Chemical Biology* 3 (4): 153–62. <https://doi.org/10.1002/9780470559277.ch110148>. Copper-Catalyzed.
- Prudhomme, Marc, Laetitia Attaiech, Guillaume Sanchez, Bernard Martin, and Jean Pierre Claverys. 2006. "Antibiotic Stress Induces Genetic Transformability in the Human Pathogen *Streptococcus Pneumoniae*." *Science* 313 (5783): 89–92. <https://doi.org/10.1126/science.1127912>.
- Quast, Robert B., Fataneh Fatemi, Michel Kranendonk, Emmanuel Margeat, and Gilles Truan. 2019. "Accurate Determination of Human CPR Conformational Equilibrium by SmFRET Using Dual Orthogonal Noncanonical Amino Acid Labeling." *ChemBioChem* 20 (5): 659–66. <https://doi.org/10.1002/cbic.201800607>.
- Rahman, Mohammad Mizanur, Indrani Sarker, Ahmed H Elmenoufy, Hagar S Bahlol, and Shahriar Zaman. 2015. "Recent Advances in Structure-Based Protein Engineering." *Scholars Journal of Applied Medical Science* 3 (September): 1224–37.

- Rajam, Gowrisankar, Julie M Anderton, George M Carlone, Jacquelyn S Sampson, and Edwin W Ades. 2008. "Pneumococcal Surface Adhesin A (PsaA): A Review." *Critical Reviews in Microbiology* 34 (3–4): 163–73. [https://doi.org/903023267 \[pii\]\r10.1080/10408410802383610](https://doi.org/903023267[pii]\r10.1080/10408410802383610).
- Rajendran, Vijayakumar, Parismita Kalita, Harish Shukla, Awanish Kumar, and Timir Tripathi. 2018. "Aminoacyl-TRNA Synthetases: Structure, Function, and Drug Discovery." *International Journal of Biological Macromolecules*. Elsevier B.V. <https://doi.org/10.1016/j.ijbiomac.2017.12.157>.
- Ramasubramanian, V. 2015. "Pneumococcal Disease in Older Adults-An Overview." *Journal of the Association of Physicians of India*.
- Ramil, Carlo P, and Qing Lin. 2013. "Bioorthogonal Chemistry: Strategies and Recent Developments." *Chemical Communications (Cambridge, England)* 49 (94): 11007–22. <https://doi.org/10.1039/c3cc44272a>.
- Randle, Elise, Nelly Ninis, and David Inwald. 2011. "Invasive Pneumococcal Disease." *Archives of Disease in Childhood: Education and Practice Edition* 96 (5): 183–90. <https://doi.org/10.1136/adc.2010.191718>.
- Ravikumar, Yuvaraj, Saravanan Prabhu Nadarajan, Tae Hyeon Yoo, Chong Soon Lee, and Hyungdon Yun. 2015. "Unnatural Amino Acid Mutagenesis-Based Enzyme Engineering." *Trends in Biotechnology* 33 (8): 462–70. <https://doi.org/10.1016/j.tibtech.2015.05.002>.
- Reddington, Samuel C, Amy J Baldwin, Rebecca Thompson, Andrea Brancale, M Tippmann, and D Dafydd Jones. 2015. "Directed Evolution of GFP with Non-Natural Amino Acids Identifies Residues for Augmenting and Photoswitching Fluorescence." *Chemical Science*. <https://doi.org/10.1039/C4SC02827A>.
- Reglinski, Mark, Giuseppe Ercoli, Charlie Plumtre, Emily Kay, Fernanda C. Petersen, James C. Paton, Brendan W. Wren, and Jeremy S. Brown. 2018. "A Recombinant Conjugated Pneumococcal Vaccine That Protects against Murine Infections with a Similar Efficacy to Prevnar-13." *Npj Vaccines* 3 (1). <https://doi.org/10.1038/s41541-018-0090-4>.
- Régloix, Stanislas Ballivet De, Olga Maurin, Anna Crambert, Louise Genestier, Gratien Bonfort, and Yoann Pons. 2017. "Complications Des Sinusites." *La Presse Medicale*. <https://doi.org/10.1016/j.lpm.2017.05.027>.
- Reinert, Ralf René, Michael R. Jacobs, Peter C. Appelbaum, Saralee Bajaksouzian, Soraia Cordeiro, Mark Van Der Linden, and Adnan Al-Lahham. 2005. "Relationship between the Original Multiply Resistant South African Isolates of *Streptococcus Pneumoniae* from 1977 to 1978 and Contemporary International Resistant Clones." *Journal of Clinical Microbiology* 43 (12): 6035–41. <https://doi.org/10.1128/JCM.43.12.6035-6041.2005>.
- Reinert, Rene, Michael R. Jacobs, and Sheldon L. Kaplan. 2010. "Pneumococcal Disease Caused by Serotype 19A: Review of the Literature and Implications for Future Vaccine Development." *Vaccine*. <https://doi.org/10.1016/j.vaccine.2010.04.020>.
- Rennels, M B, K M Edwards, H L Keyserling, K S Reisinger, D A Hogerman, D V Madore, I Chang, P R Paradiso, F J Malinoski, and A Kimura. 1998. "Safety and Immunogenicity of Heptavalent Pneumococcal Vaccine Conjugated to CRM197 in United States Infants." *Pediatrics* 101 (4 Pt 1): 604–11. <https://doi.org/10.1542/peds.101.4.604>.
- Ribeiro, João P., Sarah Villringer, David Goyard, Liliane Coche-Guerente, Manuela Höferlin, Olivier Renaudet, Winfried Römer, and Anne Imberty. 2018. "Tailor-Made Janus Lectin with Dual Avidity Assembles Glycoconjugate Multilayers and Crosslinks Protocells." *Chemical Science* 9 (39): 7634–41. <https://doi.org/10.1039/c8sc02730g>.
- Rieux, V. 2002. "Les Facteurs de Virulence de *Streptococcus Pneumoniae*." *Médecine et Maladies Infectieuses* 32 (March): 1–12. [https://doi.org/10.1016/S0399-077X\(02\)80002-6](https://doi.org/10.1016/S0399-077X(02)80002-6).
- Ring, M, IM Armitage, and RE Huber. 1985. "M-Fluorotyrosine Substitution in Beta-Galactosidase; Evidence for the Existence of a Catalytically Active Tyrosine." *Biochemical and Biophysical Research Communications* 131 (2): 675–80.
- Robb, Melissa, Joanne K. Hobbs, Shireen A. Woodiga, Sarah Shapiro-Ward, Michael D.L. Suits, Nicholas McGregor, Harry Brumer, Hasan Yesilkaya, Samantha J. King, and Alisdair B. Boraston. 2017. "Molecular Characterization of N-Glycan Degradation and Transport in *Streptococcus Pneumoniae* and Its Contribution to Virulence." *PLoS Pathogens* 13 (1). <https://doi.org/10.1371/journal.ppat.1006090>.
- Robertson, Stephanie A., Jonathan A. Ellman, and Peter G. Schultz. 1991. "A General and Efficient Route for Chemical Aminoacylation of Transfer RNAs." *Journal of the American Chemical Society* 113 (7): 2722–29. <https://doi.org/10.1021/ja00007a055>.
- Robinson, Keven M., Jay K. Kolls, and John F. Alcorn. 2015. "The Immunology of Influenza Virus-Associated Bacterial Pneumonia." *Current Opinion in Immunology*. Elsevier Ltd. <https://doi.org/10.1016/j.coi.2015.02.002>.
- Roelfes, Gerard. 2019. "LmrR: A Privileged Scaffold for Artificial Metalloenzymes." *Accounts of Chemical Research* 52 (3):

- 545–56. <https://doi.org/10.1021/acs.accounts.9b00004>.
- Rogerson, Daniel T., Amit Sachdeva, Kaihang Wang, Tamanna Haq, Agne Kazlauskaitė, Susan M. Hancock, Nicolas Huguenin-Dezot, et al. 2015. "Efficient Genetic Encoding of Phosphoserine and Its Nonhydrolyzable Analog." *Nature Chemical Biology* 11 (7): 496–503. <https://doi.org/10.1038/nchembio.1823>.
- Rolf, Jascha, Katrin Rosenthal, and Stephan Lütz. 2019. "Application of Cell-Free Protein Synthesis for Faster Biocatalyst Development." *Catalysts* 9 (2): 190. <https://doi.org/10.3390/catal9020190>.
- Ronchi, Paolo, Catalina Scarponi, Matteo Salvi, Silvia Fallarini, Laura Polito, Enrico Caneva, Luana Bagnoli, and Luigi Lay. 2013. "Synthesis of a Structural Analogue of the Repeating Unit from *Streptococcus Pneumoniae* 19F Capsular Polysaccharide Based on the Cross-Metathesis-Selenocyclization Reaction Sequence." *Journal of Organic Chemistry* 78 (11): 5172–83. <https://doi.org/10.1021/jo4001146>.
- Rossin, Raffaella, Sandra M. Van Den Bosch, Wolter Ten Hoeve, Marco Carvelli, Ron M. Versteegen, Johan Lub, and Marc S. Robillard. 2013. "Highly Reactive Trans-Cyclooctene Tags with Improved Stability for Diels-Alder Chemistry in Living Systems." *Bioconjugate Chemistry* 24 (7): 1210–17. <https://doi.org/10.1021/bc400153y>.
- Rostovtsev, Vsevolod V., Luke G. Green, Valery V. Fokin, and K. Barry Sharpless. 2002. "A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective 'Ligation' of Azides and Terminal Alkynes." *Angewandte Chemie - International Edition* 41 (14): 2596–99. [https://doi.org/10.1002/1521-3773\(20020715\)41:14<2596::AID-ANIE2596>3.0.CO;2-4](https://doi.org/10.1002/1521-3773(20020715)41:14<2596::AID-ANIE2596>3.0.CO;2-4).
- Ruscito, Annamaria, and Maria C. DeRosa. 2016. "Small-Molecule Binding Aptamers: Selection Strategies, Characterization, and Applications." *Frontiers in Chemistry*. Frontiers Media S. A. <https://doi.org/10.3389/fchem.2016.00014>.
- Ryckaert, Stefan, Nico Callewaert, Pieter P Jacobs, Sylviane Dewaele, Isabelle Dewerte, and Roland Contreras. 2008. "Fishing for Lectins from Diverse Sequence Libraries by Yeast Surface Display - an Exploratory Study." *Glycobiology* 18 (2): 137–44. <https://doi.org/10.1093/glycob/cwm131>.
- Ryu, Youngha. 2015. "Unnatural Protein Engineering: Biochemical and Medicinal Applications." 2015. <https://slideplayer.com/slide/4460530/>.
- S. Coulibaly, Fohona, and Bi-Botti C. Youan. 2017. "Current Status of Lectin-Based Cancer Diagnosis and Therapy." *AIMS Molecular Science* 4 (1): 1–27. <https://doi.org/10.3934/molsci.2017.1.1>.
- Sachdeva, Amit, Kaihang Wang, Thomas Elliott, and Jason W. Chin. 2014. "Concerted, Rapid, Quantitative, and Site-Specific Dual Labeling of Proteins." *Journal of the American Chemical Society* 136 (22): 7785–88. <https://doi.org/10.1021/ja4129789>.
- Safari, Dodi, Huberta A.T. Dekker, John A.F. Joosten, Dirk Michalik, Adriana Carvalho De Souza, Roberto Adamo, Martina Lahmann, et al. 2008. "Identification of the Smallest Structure Capable of Evoking Opsonophagocytic Antibodies against *Streptococcus Pneumoniae* Type 14." *Infection and Immunity* 76 (10): 4615–23. <https://doi.org/10.1128/IAI.00472-08>.
- Saint-Léger, Adélaïde, Carla Bello, Pablo D. Dans, Adrian Gabriel Torres, Eva Maria Novoa, Noelia Camacho, Modesto Orozco, Fyodor A. Kondrashov, and Lluís Ribas De Pouplana. 2016. "Saturation of Recognition Elements Blocks Evolution of New tRNA Identities." *Science Advances* 2 (4). <https://doi.org/10.1126/sciadv.1501860>.
- Sakamoto, Seiji, and Itaru Hamachi. 2019. "Recent Progress in Chemical Modification of Proteins." *Analytical Sciences: The International Journal of the Japan Society for Analytical Chemistry* 35 (1): 5–27. <https://doi.org/10.2116/analsci.18R003>.
- Šali, Andrej, and Tom L. Blundell. 1993. "Comparative Protein Modelling by Satisfaction of Spatial Restraints." *Journal of Molecular Biology* 234 (3): 779–815. <https://doi.org/10.1006/jmbi.1993.1626>.
- Salomonsson, Emma, Michael C. Carlsson, Veronica Osla, Ruth Hendus-Altenburger, Barbro Kahl-Knutson, Christopher T. Öberg, Anders Sundin, et al. 2010. "Mutational Tuning of Galectin-3 Specificity and Biological Function." *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.M109.098160>.
- Sanchez-Trincado, Jose L., Marta Gomez-Perosanz, and Pedro A. Reche. 2017. "Fundamentals and Methods for T- and B-Cell Epitope Prediction." *Journal of Immunology Research*. Hindawi Limited. <https://doi.org/10.1155/2017/2680160>.
- Santamaria, Raquel, Cibelly Goulart, Catia T. Perciani, Giovana C. Barazzone, Rimenys J. Carvalho, Viviane M. Gonçalves, Luciana C.C. Leite, and Martha M. Tanizaki. 2011. "Humoral Immune Response of a Pneumococcal Conjugate Vaccine: Capsular Polysaccharide Serotype 14-Lysine Modified PspA." *Vaccine* 29 (47): 8689–95. <https://doi.org/10.1016/j.vaccine.2011.08.109>.

- Sasaki, Hiroshi M., Shun Ichi Sekine, Toru Sengoku, Ryuya Fukunaga, Motoyuki Hattori, Yukiko Utsunomiya, Chizu Kuroishi, Seiki Kuramitsu, Mikako Shirouzu, and Shigeyuki Yokoyama. 2006. "Structural and Mutational Studies of the Amino Acid-Editing Domain from Archaeal/Eukaryal Phenylalanyl-TRNA Synthetase." *Proceedings of the National Academy of Sciences of the United States of America* 103 (40): 14744–49. <https://doi.org/10.1073/pnas.0603182103>.
- Sato, Mayo, Dorothee Liebschner, Yusuke Yamada, Naohiro Matsugaki, Takatoshi Arakawa, Siobhán S. Wills, Mitchell Hattie, et al. 2017. "The First Crystal Structure of a Family 129 Glycoside Hydrolase from a Probiotic Bacterium Reveals Critical Residues and Metal Cofactors." *Journal of Biological Chemistry* 292 (29): 12126–38. <https://doi.org/10.1074/jbc.M117.777391>.
- Schacht, Vivien, Soheil S. Dadras, Louise A. Johnson, David G. Jackson, Young Kwon Hong, and Michael Detmar. 2005. "Up-Regulation of the Lymphatic Marker Podoplanin, a Mucin-Type Transmembrane Glycoprotein, in Human Squamous Cell Carcinomas and Germ Cell Tumors." *American Journal of Pathology* 166 (3): 913–21. [https://doi.org/10.1016/S0002-9440\(10\)62311-5](https://doi.org/10.1016/S0002-9440(10)62311-5).
- Schiefner, André, Lea Nästle, Marietta Landgraf, Andreas J Reichert, and Arne Skerra. 2018. "Structural Basis for the Specific Cotranslational Incorporation of P-Boronophenylalanine into Biosynthetic Proteins." *Biochemistry* 57 (18): 2597–2600. <https://doi.org/10.1021/acs.biochem.8b00171>.
- Schilling, Christine I, Nicole Jung, Moritz Biskup, Ute Schepers, and Stefan Bräse. 2011. "Bioconjugation via Azide-Staudinger Ligation: An Overview." *Chemical Society Reviews* 40 (9): 4840–71. <https://doi.org/10.1039/c0cs00123f>.
- Schönfeld, D, G Matschiner, L Chatwell, S Trentmann, H Gille, M Hülsmeier, N Brown, et al. 2009. "An Engineered Lipocalin Specific for CTLA-4 Reveals a Combining Site with Structural and Conformational Features Similar to Antibodies." *Proceedings of the National Academy of Sciences of the United States of America* 106 (20): 8198–8203. <https://doi.org/10.1073/pnas.0813399106>.
- Schramm, Laura, and Nouria Hernandez. 2002. "Recruitment of RNA Polymerase III to Its Target Promoters." *Genes and Development*. <https://doi.org/10.1101/gad.1018902>.
- Schulman, LaDonne H., and Heike Pelka. 1990. "An Anticodon Change Switches the Identity of E.Coli TRNA M^{Met} from Methionine to Threonine." *Nucleic Acids Research* 18 (2): 285–89. <https://doi.org/10.1093/nar/18.2.285>.
- Schumann, Benjamin, Heung Sik Hahm, Sharavathi G. Parameswarappa, Katrin Reppe, Annette Wahlbrink, Subramanian Govindan, Paulina Kaplonek, et al. 2017. "A Semisynthetic Streptococcus Pneumoniae Serotype 8 Glycoconjugate Vaccine." *Science Translational Medicine* 9 (380). <https://doi.org/10.1126/scitranslmed.aaf5347>.
- Seeberger, Peter H., Clancy L. Pereira, and Subramanian Govindan. 2017. "Total Synthesis of a Streptococcus Pneumoniae Serotype 12F CPS Repeating Unit Hexasaccharide." *Beilstein Journal of Organic Chemistry* 13 (January): 164–73. <https://doi.org/10.3762/bjoc.13.19>.
- Shaper, Mirza, Susan K. Hollingshead, William H. Benjamin, and David E. Briles. 2004. "PspA Protects Streptococcus Pneumoniae from Killing by Apolactoferrin, and Antibody to PspA Enhances Killing of Pneumococci by Apolactoferrin." *Infection and Immunity* 72 (9): 5031–40. <https://doi.org/10.1128/IAI.72.9.5031-5040.2004>.
- Sharma, V, M Vijayan, and A Suroli. 1996. "Imparting Exquisite Specificity to Peanut Agglutinin for the Tumor-Associated Thomsen-Friedenreich Antigen by Redesign of Its Combining Site." *The Journal of Biological Chemistry* 271 (35): 21209–13. <https://doi.org/10.1074/jbc.271.35.21209>.
- Sharma, Vangmayee, Yu Zeng, W. Wesley Wang, Yuchen Qiao, Yadagiri Kurra, and Wenshe R. Liu. 2018. "Evolving the N-Terminal Domain of Pyrrolysyl-TRNA Synthetase for Improved Incorporation of Noncanonical Amino Acids." *ChemBioChem* 19 (1): 26–30. <https://doi.org/10.1002/cbic.201700268>.
- Sharon, N., and Lis Halina. 2003. *Lectins*. Kluwer Academic Publishers.
- Shibue, Rei, Takahiro Sasamoto, Masami Shimada, Bowen Zhang, Akihiko Yamagishi, and Satoshi Akanuma. 2018. "Comprehensive Reduction of Amino Acid Set in a Protein Suggests the Importance of Prebiotic Amino Acids for Stable Proteins." *Scientific Reports* 8 (1). <https://doi.org/10.1038/s41598-018-19561-1>.
- Shimomura, Osamu, Tatsuya Oda, Hiroaki Tateno, Yusuke Ozawa, Sota Kimura, Shingo Sakashita, Masayuki Noguchi, Jun Hirabayashi, Makoto Asashima, and Nobuhiro Ohkohchi. 2018. "A Novel Therapeutic Strategy for Pancreatic Cancer: Targeting Cell Surface Glycan Using RBC2LC-N Lectin-Drug Conjugate (LDC)." *Molecular Cancer Therapeutics* 17 (1): 183–95. <https://doi.org/10.1158/1535-7163.MCT-17-0232>.
- Shinya, Shoko, Shigenori Nishimura, Yoshihito Kitaoku, Tomoyuki Numata, Hisashi Kimoto, Hideo Kusaoke, Takayuki Ohnuma, and Tamoto Fukamizo. 2016. "Mechanism of Chitosan Recognition by CBM32 Carbohydrate-Binding Modules from a Paenibacillus Sp. IK-5 Chitosanase/Glucanase." *Biochemical Journal* 473 (8): 1085–95. <https://doi.org/10.1042/BCJ20160045>.

- Short, Kirsty R., Patrick C. Reading, Nancy Wang, Dimitri A. Diavatopoulos, and Odilia L. Wijburg. 2012. "Increased Nasopharyngeal Bacterial Titers and Local Inflammation Facilitate Transmission of *Streptococcus Pneumoniae*." *MBio* 3 (5). <https://doi.org/10.1128/mBio.00255-12>.
- Shoseyov, Oded, Ziv Shani, and Ilan Levy. 2006. "Carbohydrate Binding Modules: Biochemical Properties and Novel Applications." *Microbiology and Molecular Biology Reviews* 70 (2): 283–95. <https://doi.org/10.1128/MMBR.00028-05>.
- Silva, Míriam A. Da, Thiago R. Converso, Viviane M. Gonçalves, Luciana C.C. Leite, Martha M. Tanizaki, and Giovana C. Barazzone. 2017. "Conjugation of PspA4Pro with Capsular *Streptococcus Pneumoniae* Polysaccharide Serotype 14 Does Not Reduce the Induction of Cross-Reactive Antibodies." *Clinical and Vaccine Immunology* 24 (8). <https://doi.org/10.1128/CI.00118-17>.
- Singh-Blom, Amrita, Randall A Hughes, and Andrew D. Ellington. 2013. "Chapter 7 : Residue-Specific Incorporation of Unnatural Amino Acids into Proteins In Vitro and In Vivo." *Enzyme Engineering : Methods and Protocols* 978 (1): 21–30. <https://doi.org/10.1007/978-1-62703-293-3>.
- Singh, Rajesh, Pranav Gupta, Praveen K Sharma, W Edwin, Susan K Hollingshead, Shailesh Singh, and W James. 2013. "Prediction and Characterization of Helper T-Cell Epitopes from Pneumococcal Surface Adhesin A," 514–30. <https://doi.org/10.1111/imm.12194>.
- Skerra, Arne. 2007. "Anticalins as Alternative Binding Proteins for Therapeutic Use." *Current Opinion in Molecular Therapeutics* 9 (4): 336–44. <http://www.ncbi.nlm.nih.gov/pubmed/17694446>.
- Smolskaya, and Andreev. 2019. "Site-Specific Incorporation of Unnatural Amino Acids into Escherichia Coli Recombinant Protein: Methodology Development and Recent Achievement." *Biomolecules* 9 (7): 255. <https://doi.org/10.3390/biom9070255>.
- Sommer, Carina A., Andreas Eichinger, and Arne Skerra. 2019. "A Tetrahedral Boronic Acid Diester Formed by a Non-natural Amino Acid in the Ligand Pocket of an Engineered Lipocalin." *ChemBioChem*, cbic.201900405. <https://doi.org/10.1002/cbic.201900405>.
- Song, Jae Hoon, Ron Dagan, Keith P. Klugman, and Bernard Fritzell. 2012. "The Relationship between Pneumococcal Serotypes and Antibiotic Resistance." *Vaccine*. <https://doi.org/10.1016/j.vaccine.2012.01.091>.
- Spicer, Christopher D., E. Thomas Pashuck, and Molly M. Stevens. 2018. "Achieving Controlled Biomolecule-Biomaterial Conjugation." *Chemical Reviews*. American Chemical Society. <https://doi.org/10.1021/acs.chemrev.8b00253>.
- Spiwok, Vojtěch. 2017. "CH/π Interactions in Carbohydrate Recognition." *Molecules* 22 (7): 1–11. <https://doi.org/10.3390/molecules22071038>.
- Srivastava, N., J. L. Zeiler, S. L. Smithson, G. M. Carlone, E. W. Ades, J. S. Sampson, S. E. Johnson, T. Kieber-Emmons, and M. A.J. Westerink. 2000. "Selection of an Immunogenic and Protective Epitope of the PsaA Protein of *Streptococcus Pneumoniae* Using a Phage Display Library." *Hybridoma* 19 (1): 23–31. <https://doi.org/10.1089/027245700315761>.
- Stacevičiene, Indre, Sigita Petraitienė, Daiva Vaičiuniene, Tomas Alasevičius, Jurate Kiršlienė, and Vytautas Usonis. 2016. "Antibiotic Resistance of *Streptococcus Pneumoniae*, Isolated from Nasopharynx of Preschool Children with Acute Respiratory Tract Infection in Lithuania." *BMC Infectious Diseases* 16 (1). <https://doi.org/10.1186/s12879-016-1544-9>.
- Stefanetti, G, S Rondini, L Lanzilao, A Saul, C A MacLennan, and F Micoli. 2014. "Impact of Conjugation Chemistry on the Immunogenicity of *S. Typhimurium* Conjugate Vaccines." *Vaccine* 32 (46): 6122–29. <https://doi.org/10.1016/j.vaccine.2014.08.056>.
- Stefanetti, Giuseppe, Qi-Ying Hu, Aimee Usera, Zack Robinson, Martin Allan, Alok Singh, Hidetomo Imase, et al. 2015. "Sugar–Protein Connectivity Impacts on the Immunogenicity of Site-Selective *Salmonella* O-Antigen Glycoconjugate Vaccines." *Angewandte Chemie* 127 (45): 13396–401. <https://doi.org/10.1002/ange.201506112>.
- Steffes, C., J. Ellis, J. Wu, and B. P. Rosen. 1992. "The LysP Gene Encodes the Lysine-Specific Permease." *Journal of Bacteriology* 174 (10): 3242–49. <https://doi.org/10.1128/jb.174.10.3242-3249.1992>.
- Steitz, Thomas A. 2008. "A Structural Understanding of the Dynamic Ribosome Machine." *Nature Reviews Molecular Cell Biology*. <https://doi.org/10.1038/nrm2352>.
- Sternburg. 1881. "A Fatal Form of Septicemia in the Rabbit Produced by the Subcutaneous Injection of Human Saliva." *Natl. Bd. Health Bull.*
- Steward, Lance E, Cynthia S Collins, Marcella A Gilmore, Justin E Carlson, J B Alexander Ross, and A Richard Chamberlin.

1997. "In Vitro Site-Specific Incorporation of Fluorescent Probes into Beta-Galactosidase" 7863 (7): 761–66.
- Sudakevitz, Dvora, Nikola Kostlánová, Gitit Blatman-Jan, Edward P Mitchell, Batia Lerrer, Michaela Wimmerová, Don J Katcoff, Anne Imberty, and Nechama Gilboa-Garber. 2004. "A New Ralstonia Solanacearum High-Affinity Mannose-Binding Lectin RS-III Structurally Resembling the Pseudomonas Aeruginosa Fucose-Specific Lectin PA-III." *Molecular Microbiology* 52 (3): 691–700. <https://doi.org/10.1111/j.1365-2958.2004.04020.x>.
- Sugawara, Hajime, Masami Kusunoki, Genji Kurisu, Tokiko Fujimoto, Haruhiko Aoyagi, and Tomomitsu Hatakeyama. 2004. "Characteristic Recognition of N-Acetylgalactosamine by an Invertebrate C-Type Lectin, CEL-I, Revealed by x-Ray Crystallographic Analysis." *Journal of Biological Chemistry* 279 (43): 45219–25. <https://doi.org/10.1074/jbc.M408840200>.
- Sun, Wei, Lupei Du, and Minyong Li. 2010. "Aptamer-Based Carbohydrate Recognition." *Current Pharmaceutical Design*, 2269–78.
- Suzuki, Tateki, Corwin Miller, Li Tao Guo, Joanne M.L. Ho, David I. Bryson, Yane Shih Wang, David R. Liu, and Dieter Söll. 2017. "Crystal Structures Reveal an Elusive Functional Domain of Pyrrolysyl-TRNA Synthetase." *Nature Chemical Biology* 13 (12): 1261–66. <https://doi.org/10.1038/nchembio.2497>.
- Swimmer, C, S M Lehar, J McCafferty, D J Chiswell, W A Blättler, and B C Guild. 1992. "Phage Display of Ricin B Chain and Its Single Binding Domains: System for Screening Galactose-Binding Mutants." *Proceedings of the National Academy of Sciences of the United States of America* 89 (9): 3756–60. <https://doi.org/10.1073/pnas.89.9.3756>.
- Taillardet, Morgan, Ghina Haffar, Paul Mondière, Marie-Jeanne Asensio, Hanane Gheit, Nicolas Burdin, Thierry Defrance, and Laurent Genestier. 2009. "The Thymus-Independent Immunity Conferred by a Pneumococcal Polysaccharide Is Mediated by Long-Lived Plasma Cells." *Blood* 114 (20): 4432–40. <https://doi.org/10.1182/blood-2009-01-200014>.
- Takimoto, Jeffrey K., Nikki Dellas, Joseph P. Noel, and Lei Wang. 2011. "Stereochemical Basis for Engineered Pyrrolysyl-TRNA Synthetase and the Efficient in Vivo Incorporation of Structurally Divergent Non-Native Amino Acids." *ACS Chemical Biology* 6 (7): 733–43. <https://doi.org/10.1021/cb200057a>.
- Tan, Jin, He Fang Wang, and Xiu Ping Yan. 2009. "Discrimination of Saccharides with a Fluorescent Molecular Imprinting Sensor Array Based on Phenylboronic Acid Functionalized Mesoporous Silica." *Analytical Chemistry* 81 (13): 5273–80. <https://doi.org/10.1021/ac900484x>.
- Taylor, Maureen E., and Drickaner Kurt. 2011. *Introduction to Glycobiology*. 3rd Edition, Oxford University Press.
- Terasaka, Naohiro, Gosuke Hayashi, Takayuki Katoh, and Hiroaki Suga. 2014. "An Orthogonal Ribosome-TRNA Pair via Engineering of the Peptidyl Transferase Center." *Nature Chemical Biology* 10 (7): 555–57. <https://doi.org/10.1038/nchembio.1549>.
- Tharp, Jeffery M., Yane Shih Wang, Yan Jiun Lee, Yanyan Yang, and Wenshe R. Liu. 2014. "Genetic Incorporation of Seven Ortho-Substituted Phenylalanine Derivatives." *ACS Chemical Biology* 9 (4): 884–90. <https://doi.org/10.1021/cb400917a>.
- Thibodeaux, Gabrielle Nina, Xiang Liang, Kathryn Moncivais, Aiko Umeda, Oded Singer, Lital Alfonta, and Zhiwen Jonathan Zhang. 2010. "Transforming a Pair of Orthogonal TRNA-Aminoacyl-TRNA Synthetase from Archaea to Function in Mammalian Cells." *PLoS ONE* 5 (6). <https://doi.org/10.1371/journal.pone.0011263>.
- Ting, A. Y., I. Shin, C. Lucero, and P. G. Schultz. 1998. "Energetic Analysis of an Engineered Cation- π Interaction in Staphylococcal Nuclease [18]." *Journal of the American Chemical Society*. <https://doi.org/10.1021/ja981520l>.
- Tobola, Felix, Mickael Lelimosin, Annabelle Varrot, Emilie Gillon, Barbara Darnhofer, Ola Blixt, Ruth Birner-Gruenberger, Anne Imberty, and Birgit Wiltschi. 2018. "Effect of Noncanonical Amino Acids on Protein-Carbohydrate Interactions: Structure, Dynamics, and Carbohydrate Affinity of a Lectin Engineered with Fluorinated Tryptophan Analogs." *ACS Chemical Biology* 13 (8): 2211–19. <https://doi.org/10.1021/acschembio.8b00377>.
- Tobola, Felix, Birgit Wiltschi, Elise Sylvander, and Claudia Gafko. 2019. "Clickable Lectins: Bioorthogonal Handles Facilitate the Directed Conjugation of Lectins in a Modular Fashion." <https://doi.org/10.1098/rsfs.2018.0072>.
- Tomasz, A. 1997. "Antibiotic Resistance in Streptococcus Pneumoniae." *Clinical Infectious Diseases* 173 (SUPPL. OCT.). https://doi.org/10.1093/clinids/24.supplement_1.s85.
- Tomme, P, H Van Tilbeurgh, G Pettersson, J Van Damme, J Vandekerckhove, J Knowles, T Teeri, and M Claeysens. 1988. "Studies of the Cellulolytic System of Trichoderma Reesei QM 9414. Analysis of Domain Function in Two Cellobiohydrolases by Limited Proteolysis." *European Journal of Biochemistry* 170 (3): 575–81. <https://doi.org/10.1111/j.1432-1033.1988.tb13736.x>.

- Tomos, Ioannis, Elias Perros, and George Dimopoulos. 2017. "Prevention of Invasive Pneumococcal Disease (IPD)." *Pneumon* 30 (1): 32–40.
- Topin, Jérémie, Julien Diharce, Sébastien Fiorucci, Serge Antonczak, and Jérôme Golebiowski. 2014. "O₂ Migration Rates in [NiFe] Hydrogenases. A Joint Approach Combining Free-Energy Calculations and Kinetic Modeling." *Journal of Physical Chemistry B*. <https://doi.org/10.1021/jp4093964>.
- Troxler, Lukas J., Joel P. Werren, Thierry O. Schaffner, Nadezda Mostacci, Peter Vermathen, Martina Vermathen, Daniel Wüthrich, et al. 2019. "Carbon Source Regulates Polysaccharide Capsule Biosynthesis in *Streptococcus Pneumoniae*." *Journal of Biological Chemistry* 294 (46): 17224–38. <https://doi.org/10.1074/jbc.RA119.010764>.
- Tsunoda, Masaru, Yoshio Kusakabe, Nobutada Tanaka, Satoshi Ohno, Masashi Nakamura, Toshiya Senda, Tomohisa Moriguchi, et al. 2007. "Structural Basis for Recognition of Cognate tRNA by Tyrosyl-tRNA Synthetase from Three Kingdoms." *Nucleic Acids Research* 35 (13): 4289–4300. <https://doi.org/10.1093/nar/gkm417>.
- Tuley, Alfred, Yan Jiun Lee, Bo Wu, Zhiyong U. Wang, and Wenshe R. Liu. 2014. "A Genetically Encoded Aldehyde for Rapid Protein Labelling." *Chemical Communications* 50 (56): 7424–26. <https://doi.org/10.1039/c4cc02000f>.
- Turner, James M., James Graziano, Glen Spraggon, and Peter G. Schultz. 2006. "Structural Plasticity of an Aminoacyl-tRNA Synthetase Active Site." *Proceedings of the National Academy of Sciences of the United States of America* 103 (17): 6483–88. <https://doi.org/10.1073/pnas.0601756103>.
- Tze-Fei Wong, J. 1980. "Role of Minimization of Chemical Distances between Amino Acids in the Evolution of the Genetic Code." *Proceedings of the National Academy of Sciences of the United States of America* 77 (2 II): 1083–86. <https://doi.org/10.1073/pnas.77.2.1083>.
- Unione, Luca, Maria Alcalá, Begoña Echeverría, Sonia Serna, Ana Ardá, Antonio Franconetti, F. Javier Cañada, Tammo Diercks, Niels Reichardt, and Jesús Jiménez-Barbero. 2017. "Fluoroacetamide Moieties as NMR Spectroscopy Probes for the Molecular Recognition of GlcNAc-Containing Sugars: Modulation of the CH– π Stacking Interactions by Different Fluorination Patterns." *Chemistry - A European Journal*. <https://doi.org/10.1002/chem.201605573>.
- Vargas-Rodríguez, Oscar, Anastasia Sevostyanova, Dieter Söll, and Ana Crnković. 2018. "Upgrading Aminoacyl-tRNA Synthetases for Genetic Code Expansion." *Current Opinion in Chemical Biology*. Elsevier Ltd. <https://doi.org/10.1016/j.cbpa.2018.07.014>.
- Varki, Ajit, Richard D Cummings, Jeffrey D Esko, Hudson H Freeze, Pamela Stanley, Carolyn R Bertozzi, Gerald W Hart, and Marilynn E Etzler. 2009. *Essentials of Glycobiology*. *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press. <http://www.ncbi.nlm.nih.gov/pubmed/20301239>.
- Varon, E, and L Gutmann. 2004. "Epidémiologie de La Résistance Aux Antibiotiques Des Pneumocoques - Revue Médicale Suisse." <https://www.revmed.ch/RMS/2004/RMS-2475/23719>.
- Venditto, Immacolata, Shabir Najmudin, Ana S Luís, Luís M A Ferreira, Kazuo Sakka, J Paul Knox, Harry J Gilbert, and Carlos M G A Fontes. 2015. "Family 46 Carbohydrate-Binding Modules Contribute to the Enzymatic Hydrolysis of Xyloglucan and β -1,3-1,4-Glucans through Distinct Mechanisms." *The Journal of Biological Chemistry* 290 (17): 10572–86. <https://doi.org/10.1074/jbc.M115.637827>.
- Verhagen, Lilly M., Marien I. De Jonge, Peter Burghout, Kiki Schraa, Lorenza Spagnuolo, Svenja Mennens, Marc J. Eleveld, et al. 2014. "Genome-Wide Identification of Genes Essential for the Survival of *Streptococcus Pneumoniae* in Human Saliva." *PLoS ONE* 9 (2). <https://doi.org/10.1371/journal.pone.0089541>.
- Vinuesa, C. G., C. De Lucas, and M. C. Cook. 2001. "Clinical Implications of the Specialised B Cell Response to Polysaccharide Encapsulated Pathogens." *Postgraduate Medical Journal*. <https://doi.org/10.1136/pmj.77.911.562>.
- Walsh, Rebecca L, and Andrew Camilli. 2011. "Streptococcus Pneumoniae Is Desiccation Tolerant and Infectious upon Rehydration." *MBio* 2 (3): e00092-11. <https://doi.org/10.1128/mBio.00092-11>.
- Walter, Nils G., and David R. Engelke. 2002. "Ribozymes: Catalytic RNAs That Cut Things, Make Things, and Do Odd and Useful Jobs." *Biologist*.
- Wan, Wei, Jeffery M. Tharp, and Wenshe R. Liu. 2014. "Pyrrolysyl-tRNA Synthetase: An Ordinary Enzyme but an Outstanding Genetic Code Expansion Tool." *Biochimica et Biophysica Acta - Proteins and Proteomics* 1844 (6): 1059–70. <https://doi.org/10.1016/j.bbapap.2014.03.002>.
- Wandrey, Georg, Joel Wurzel, Kyra Hoffmann, Tobias Ladner, Jochen Büchs, Lorenz Meinel, and Tessa Lühmann. 2016. "Probing Unnatural Amino Acid Integration into Enhanced Green Fluorescent Protein by Genetic Code Expansion with a High-Throughput Screening Platform." *Journal of Biological Engineering* 10 (1). <https://doi.org/10.1186/s13036-016-0031-6>.

- Wang, Hui, Lizhong He, Martin Lensch, Hans Joachim Gabius, Conan J. Fee, and Anton P.J. Middelberg. 2008. "Single-Site Cys-Substituting Mutation of Human Lectin Galectin-2: Modulating Solubility in Recombinant Production, Reducing Long-Term Aggregation, and Enabling Site-Specific MonoPEGylation." *Biomacromolecules*.
<https://doi.org/10.1021/bm800801b>.
- Wang, Jian Zhi, Inge Grundke-Iqbal, and Khalid Iqbal. 1996. "Glycosylation of Microtubule-Associated Protein Tau: An Abnormal Posttranslational Modification in Alzheimer's Disease." *Nature Medicine* 2 (8): 871–75.
<https://doi.org/10.1038/nm0896-871>.
- Wang, Kaihang, Heinz Neumann, Sew Y. Peak-Chew, and Jason W. Chin. 2007. "Evolved Orthogonal Ribosomes Enhance the Efficiency of Synthetic Genetic Code Expansion." *Nature Biotechnology* 25 (7): 770–77.
<https://doi.org/10.1038/nbt1314>.
- Wang, Kaihang, Amit Sachdeva, Daniel J. Cox, Nabil W. Wilf, Kathrin Lang, Stephen Wallace, Ryan A. Mehl, and Jason W. Chin. 2014. "Optimized Orthogonal Translation of Unnatural Amino Acids Enables Spontaneous Protein Double-Labeling and FRET." *Nature Chemistry* 6 (5): 393–403. <https://doi.org/10.1038/nchem.1919>.
- Wang, Kaihang, Wolfgang H. Schmieid, and Jason W. Chin. 2012. "Reprogramming the Genetic Code: From Triplet to Quadruplet Codes." *Angewandte Chemie - International Edition* 51 (10): 2288–97.
<https://doi.org/10.1002/anie.201105016>.
- Wang, L., A. Brock, B. Herberich, and P. G. Schultz. 2001. "Expanding the Genetic Code of Escherichia Coli." *Science* 292 (5516): 498–500. <https://doi.org/10.1126/science.1060077>.
- Wang, Lei. 2017. "Engineering the Genetic Code in Cells and Animals: Biological Considerations and Impacts." *Acc Chem Res*.
<https://doi.org/10.1016/j.physbeh.2017.03.040>.
- Wang, Lei, Jianming Xie, and Peter G. Schultz. 2006. "Expanding the Genetic Code." *Annual Review of Biophysics and Biomolecular Structure* 35: 225–49. <https://doi.org/10.1146/annurev.biophys.35.101105.121507>.
- Wang, Nanxi, Tong Ju, Wei Niu, and Jiantao Guo. 2015. "Fine-Tuning Interaction between Aminoacyl-TRNA Synthetase and TRNA for Efficient Synthesis of Proteins Containing Unnatural Amino Acids."
- Wang, Ning, Ilson Whang, and Jehee Lee. 2008. "A Novel C-Type Lectin from Abalone, Haliotis Discus Discus, Agglutinates *Vibrio Alginolyticus*." *Developmental and Comparative Immunology* 32 (9): 1034–40.
<https://doi.org/10.1016/j.dci.2008.02.001>.
- Wang, Pin, Yi Tang, and David A. Tirrell. 2003. "Incorporation of Trifluoroisoleucine into Proteins in Vivo." *Journal of the American Chemical Society* 125 (23): 6900–6906. <https://doi.org/10.1021/ja0298287>.
- Wang, Qian, Angela R. Parrish, and Lei Wang. 2009. "Expanding the Genetic Code for Biological Studies." *Chemistry and Biology*. <https://doi.org/10.1016/j.chembiol.2009.03.001>.
- Wang, Zhipeng A., Yu Zeng, Yadagiri Kurra, Xin Wang, Jeffery M. Tharp, Erol C. Vatansver, Willie W. Hsu, Susie Dai, Xinqiang Fang, and Wenshe R. Liu. 2017. "A Genetically Encoded Allylsine for the Synthesis of Proteins with Site-Specific Lysine Dimethylation." *Angewandte Chemie - International Edition* 56 (1): 212–16. <https://doi.org/10.1002/anie.201609452>.
- "Website Scope: [Http://Scop.Mrc-Lmb.Cam.Ac.Uk/Scop/Data/Scop.b.c.Dd.b.A.Html](http://Scop.Mrc-Lmb.Cam.Ac.Uk/Scop/Data/Scop.b.c.Dd.b.A.Html)." n.d.
- Weinberger, Daniel M., Richard Malley, and Marc Lipsitch. 2011. "Serotype Replacement in Disease after Pneumococcal Vaccination." *The Lancet*. [https://doi.org/10.1016/S0140-6736\(10\)62225-8](https://doi.org/10.1016/S0140-6736(10)62225-8).
- Weis, William I. 1997. "Cell-Surface Carbohydrate Recognition by Animal and Viral Lectins." *Current Opinion in Structural Biology* 7 (5): 624–30. [https://doi.org/10.1016/S0959-440X\(97\)80070-X](https://doi.org/10.1016/S0959-440X(97)80070-X).
- Weiser, J N, D Bae, H Epino, S B Gordon, M Kapoor, L A Zenewicz, and M Shchepetov. 2001. "Changes in Availability of Oxygen Accentuate Differences in Capsular Polysaccharide Expression by Phenotypic Variants and Clinical Isolates of *Streptococcus Pneumoniae*." *Infection and Immunity* 69 (9): 5430–39. <https://doi.org/10.1128/iai.69.9.5430-5439.2001>.
- Weiser, Jeffrey N, Daniela M Ferreira, and James C Paton. 2018. *Streptococcus Pneumoniae : Transmission, Colonization and Invasion*. Vol. 16. <https://doi.org/10.1038/s41579-018-0001-8>.Streptococcus.
- Weishaupt, Markus W., Stefan Matthies, Mattan Hurevich, Claney L. Pereira, Heung Sik Hahm, and Peter H. Seeberger. 2016. "Automated Glycan Assembly of a *S. Pneumoniae* Serotype 3 Cps Antigen." *Beilstein Journal of Organic Chemistry* 12 (July): 1440–46. <https://doi.org/10.3762/bjoc.12.139>.
- Werz, Daniel B., René Ranzinger, Stephan Herget, Alexander Adibekian, Claus Wilhelm Von der Lieth, and Peter H.

- Seeberger. 2007. "Exploring the Structural Diversity of Mammalian Carbohydrates ('glycospace') by Statistical Databank Analysis." *ACS Chemical Biology* 2 (10): 685–91. <https://doi.org/10.1021/cb700178s>.
- Westerink, M. A. Julie, Harry W. Schroeder, and Moon H. Nahm. 2012. "Immune Responses to Pneumococcal Vaccines in Children and Adults: Rationale for Age-Specific Vaccination." *Aging and Disease*. International Society on Aging and Disease.
- White, Benjamin, Elliott Stirling Robinson, and Laverne Almon Barnes. 2008. *The Biology of Pneumococcus; the Bacteriological, Biochemical, and Immunological Characters and Activities of Diplococcus Pneumoniae*.
- Whitney, Cynthia G., Monica M. Farley, James Hadler, Lee H. Harrison, Nancy M. Bennett, Ruth Lynfield, Arthur Reingold, et al. 2003. "Decline in Invasive Pneumococcal Disease after the Introduction of Protein-Polysaccharide Conjugate Vaccine." *New England Journal of Medicine* 348 (18): 1737–46. <https://doi.org/10.1056/NEJMoa022823>.
- Who. 2017. "https://www.who.int/news-room/fact-sheets/detail/pneumonia." 2017.
- WHO. 2017. "WHO Publishes List of Bacteria for Which New Antibiotics Are Urgently Needed." 2017. <https://www.who.int/news-room/detail/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed>.
- . 2018. "WHO Report Meningitis." 2018. <https://www.who.int/emergencies/diseases/meningitis/en/>.
- . 2019. "WHO | Sepsis." 2019. <https://www.who.int/sepsis/en/>.
- Wiley, John and sons. 2006. *Immunodominance : The Choice of the Immune System*. Wiley-VCH.
- Williams, David B. 2006. "Beyond Lectins: The Calnexin/Calreticulin Chaperone System of the Endoplasmic Reticulum." *Journal of Cell Science*. <https://doi.org/10.1242/jcs.02856>.
- Windle, Claire L., Katie J. Simmons, James R. Ault, Chi H. Trinh, Adam Nelson, Arwen R. Pearson, and Alan Berry. 2017. "Extending Enzyme Molecular Recognition with an Expanded Amino Acid Alphabet." *Proceedings of the National Academy of Sciences* 114 (10): 2610–15. <https://doi.org/10.1073/pnas.1616816114>.
- Winterhalter, C, P Heinrich, A Candussio, G Wich, and W Liebl. 1995. "Identification of a Novel Cellulose-Binding Domain within the Multidomain 120 KDa Xylanase XynA of the Hyperthermophilic Bacterium Thermotoga Maritima." *Molecular Microbiology* 15 (3): 431–44. <https://doi.org/10.1111/j.1365-2958.1995.tb02257.x>.
- Worsley, Graham J., Guilhem A. Tourniaire, Kathryn E.S. Medlock, Felicity K. Sartain, Hazel E. Harmer, Michael Thatcher, Adrian M. Horgan, and John Pritchard. 2008. "Measurement of Glucose in Blood with a Phenylboronic Acid Optical Sensor." *Journal of Diabetes Science and Technology* 2 (2): 213–20. <https://doi.org/10.1177/193229680800200207>.
- Wren, John T., Lance K. Blevins, Bing Pang, Lauren B. King, Antonia C. Perez, Kyle A. Murrah, Jennifer L. Reimche, Martha A. Alexander-Miller, and W. Edward Swords. 2014. "Influenza A Virus Alters Pneumococcal Nasal Colonization and Middle Ear Infection Independently of Phase Variation." *Infection and Immunity* 82 (11): 4802–12. <https://doi.org/10.1128/IAI.01856-14>.
- Wulff, Günter, Thomas Gross, Rainer Schönfeld, Thomas Schrader, and Christian Kirsten. 1998. "Molecular Imprinting for the Preparation of Enzyme-Analogous Polymers." *ACS Symposium Series* 703 (2): 10–28.
- Xu, Bin, and Ling Wu. 2014. "Analysis of Receptor-Ligand Binding by Photoaffinity Cross-Linking." *Science China Chemistry* 57 (2): 232–42. <https://doi.org/10.1007/s11426-013-5045-z>.
- Yabe, Rikio, Yoko Itakura, Sachiko Nakamura-Tsuruta, Jun Iwaki, Atsushi Kuno, and Jun Hirabayashi. 2009. "Engineering a Versatile Tandem Repeat-Type A2-6sialic Acid-Binding Lectin." *Biochemical and Biophysical Research Communications*. <https://doi.org/10.1016/j.bbrc.2009.04.090>.
- Yabe, Rikio, Ryuichiro Suzuki, Atsushi Kuno, Zui Fujimoto, Yoshifumi Jigami, and Jun Hirabayashi. 2007. "Tailoring a Novel Sialic Acid-Binding Lectin from a Ricin-B Chain-like Galactose-Binding Protein by Natural Evolution-Mimicry." *Journal of Biochemistry*. <https://doi.org/10.1093/jb/mvm043>.
- Yamamoto, K, I N Maruyama, and T Osawa. 2000. "Cyborg Lectins: Novel Leguminous Lectins with Unique Specificities." *Journal of Biochemistry* 127 (1): 137–42. <https://doi.org/10.1093/oxfordjournals.jbchem.a022575>.
- Yamamoto, Kazuo, Yukiko Konami, Toshiaki Osawa, and Tatsuro Irimura. 1992. "Alteration of the Carbohydrate-Binding Specificity of the Bauhinia Purpurea Lectin through the Preparation of a Chimeric Lectin." *Journal of Biochemistry*. <https://doi.org/10.1093/oxfordjournals.jbchem.a123724>.
- Yamamoto, Miyako, Yoshihiko Kominato, and Fumiichiro Yamamoto. 1999. "Phage Display cDNA Cloning of Protein with Carbohydrate Affinity." *Biochemical and Biophysical Research Communications* 255 (2): 194–99.

<https://doi.org/10.1006/bbrc.1999.0175>.

- Yanagisawa, Tatsuo, Ryohei Ishii, Ryuya Fukunaga, Takatsugu Kobayashi, Kensaku Sakamoto, and Shigeyuki Yokoyama. 2008a. "Crystallographic Studies on Multiple Conformational States of Active-Site Loops in Pyrrolysyl-TRNA Synthetase." *Journal of Molecular Biology* 378 (3): 634–52. <https://doi.org/10.1016/j.jmb.2008.02.045>.
- . 2008b. "Multistep Engineering of Pyrrolysyl-TRNA Synthetase to Genetically Encode N ϵ -(o-Azidobenzoyloxycarbonyl) Lysine for Site-Specific Protein Modification." *Chemistry and Biology* 15 (11): 1187–97. <https://doi.org/10.1016/j.chembiol.2008.10.004>.
- Yang, Bing, Shibing Tang, Cheng Ma, Shang-tong Li, Guang-can Shao, Bobo Dang, William F Degrado, et al. 2017. "Spontaneous and Specific Chemical Cross-Linking in Live Cells to Capture and Identify Protein Interactions." *Nature Communications*, 1–10. <https://doi.org/10.1038/s41467-017-02409-z>.
- Yang, Hao, Poonam Srivastava, Chen Zhang, and Jared C. Lewis. 2014. "A General Method for Artificial Metalloenzyme Formation through Strain-Promoted Azide-Alkyne Cycloaddition." *ChemBioChem* 15 (2): 223–27. <https://doi.org/10.1002/cbic.201300661>.
- Yang, Kyung Ae, Mihaela Barbu, Marlin Halim, Payal Pallavi, Benjamin Kim, Dmitry M. Kolpashchikov, Stevan Pecic, Steven Taylor, Tilla S. Worgall, and Milan N. Stojanovic. 2014. "Recognition and Sensing of Low-Epitope Targets via Ternary Complexes with Oligonucleotides and Synthetic Receptors." *Nature Chemistry* 6 (11): 1003–8. <https://doi.org/10.1038/NCHEM.2058>.
- Yang, Xinyan, Liqin Wu, Xuemei Duan, Lianzhen Cui, Jingjing Luo, and Gongchu Li. 2014. "Adenovirus Carrying Gene Encoding Haliotis Discus Discus Sialic Acid Binding Lectin Induces Cancer Cell Apoptosis." *Marine Drugs* 12 (7): 3994–4004. <https://doi.org/10.3390/md12073994>.
- Yau, Tammy, Xiuli Dan, Charlene Cheuk Wing Ng, and Tzi Bun Ng. 2015. "Lectins with Potential for Anti-Cancer Therapy." *Molecules* 20 (3): 3791–3810. <https://doi.org/10.3390/molecules20033791>.
- Ye, Shixin, Thomas Huber, Reiner Vogel, and Thomas P. Sakmar. 2009. "FTIR Analysis of GPCR Activation Using Azido Probes." *Nature Chemical Biology* 5 (6): 397–99. <https://doi.org/10.1038/nchembio.167>.
- Yim, M., T. Ono, and T. Irimura. 2002. "Mutated Plant Lectin Library Useful to Identify Different Cells." *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.041621998>.
- Young, Douglas D., Travis S. Young, Michael Jahnz, Insha Ahmad, Glen Spraggon, and Peter G. Schultz. 2011. "An Evolved Aminoacyl-TRNA Synthetase with Atypical Polysubstrate Specificity." *Biochemistry* 50 (11): 1894–1900. <https://doi.org/10.1021/bi101929e>.
- Young, Travis S., and Peter G. Schultz. 2010. "Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon." *Journal of Biological Chemistry* 285 (15): 11039–44. <https://doi.org/10.1074/jbc.R109.091306>.
- Young, Travis S, Insha Ahmad, Jun A Yin, and Peter G Schultz. 2010. "An Enhanced System for Unnatural Amino Acid Mutagenesis in E. Coli." *Journal of Molecular Biology* 395 (2): 361–74. <https://doi.org/10.1016/j.jmb.2009.10.030>.
- Yu, Kang, Yin Qiao, Guofeng Gu, Jian Gao, Shuihong Cai, Zhongzhu Long, and Zhongwu Guo. 2016. "Synthesis of the Biological Repeating Unit of Streptococcus Pneumoniae Serotype 23F Capsular Polysaccharide." *Organic and Biomolecular Chemistry* 14 (48): 11462–72. <https://doi.org/10.1039/c6ob02363k>.
- Yu, Yang, Cheng Hu, Lin Xia, and Jianguyun Wang. 2018. "Artificial Metalloenzyme Design with Unnatural Amino Acids and Non-Native Cofactors." *ACS Catalysis*. <https://doi.org/10.1021/acscatal.7b03754>.
- Yu, Yang, Xiaoxuan Lv, Jiasong Li, Qing Zhou, Chang Cui, Parisa Hosseinzadeh, Arnab Mukherjee, Mark J. Nilges, Jianguyun Wang, and Yi Lu. 2015. "Defining the Role of Tyrosine and Rational Tuning of Oxidase Activity by Genetic Incorporation of Unnatural Tyrosine Analogs." *Journal of the American Chemical Society* 137 (14): 4594–97. <https://doi.org/10.1021/ja5109936>.
- Yuste, Jose, Marina Botto, James C. Paton, David W. Holden, and Jeremy S. Brown. 2005. "Additive Inhibition of Complement Deposition by Pneumolysin and PspA Facilitates Streptococcus Pneumoniae Septicemia." *The Journal of Immunology* 175 (3): 1813–19. <https://doi.org/10.4049/jimmunol.175.3.1813>.
- Zafar, M. Ammar, Yang Wang, Shigeto Hamaguchi, and Jeffrey N. Weiser. 2017. "Host-to-Host Transmission of Streptococcus Pneumoniae Is Driven by Its Inflammatory Toxin, Pneumolysin." *Cell Host and Microbe* 21 (1): 73–83. <https://doi.org/10.1016/j.chom.2016.12.005>.
- Zhang, Chun Mei, John J. Perona, Kang Ryu, Christopher Francklyn, and Ya Ming Hou. 2006. "Distinct Kinetic Mechanisms of the Two Classes of Aminoacyl-TRNA Synthetases." *Journal of Molecular Biology* 361 (2): 300–311.

<https://doi.org/10.1016/j.jmb.2006.06.015>.

- Zhang, Qingju, Herman S Overkleeft, Gijsbert A van der Marel, and Jeroen DC Codée. 2017. "Synthetic Zwitterionic Polysaccharides." *Current Opinion in Chemical Biology* 40 (October): 95–101. <https://doi.org/10.1016/j.cbpa.2017.07.010>.
- Zhang, Y A N, L E I Wang, Peter G Schultz, and Ian A Wilson. 2005. "Tyrosyl-TRNA Synthetase (TyrRS) and an Engineered TyrRS Specific for O-Methyl-L-Tyrosine." *Protein Science* 14: 1340–49. <https://doi.org/10.1110/ps.041239305.4>.
- Zhang, Yorke, Brian M. Lamb, Aaron W. Feldman, Anne Xiaozhou Zhou, Thomas Lavergne, Lingjun Li, and Floyd E. Romesberg. 2017. "A Semisynthetic Organism Engineered for the Stable Expansion of the Genetic Alphabet." *Proceedings of the National Academy of Sciences of the United States of America* 114 (6): 1317–22. <https://doi.org/10.1073/pnas.1616443114>.
- Zhang, Zhiquan, Liqun Huang, Wenping Zhao, and Basil Rigas. 2010. "Annexin 1 Induced by Anti-Inflammatory Drugs Binds to NF-KappaB and Inhibits Its Activation: Anticancer Effects in Vitro and in Vivo." *Cancer Research* 70 (6): 2379–88. <https://doi.org/10.1158/0008-5472.CAN-09-4204>.
- Zheng, Lei, Ulrich Baumann, and Jean Louis Reymond. 2004. "An Efficient One-Step Site-Directed and Site-Saturation Mutagenesis Protocol." *Nucleic Acids Research* 32 (14). <https://doi.org/10.1093/nar/gnh110>.
- Zhong, W, J P Gallivan, Y Zhang, L Li, H A Lester, and D A Dougherty. 1998. "From Ab Initio Quantum Mechanics to Molecular Neurobiology: A Cation-Pi Binding Site in the Nicotinic Receptor." *Proceedings of the National Academy of Sciences of the United States of America* 95 (21): 12088–93. <https://doi.org/10.1073/pnas.95.21.12088>.

Annexes

Annexe 1 : Tableau de la composition des principaux sérotypes de *S. pneumoniae*

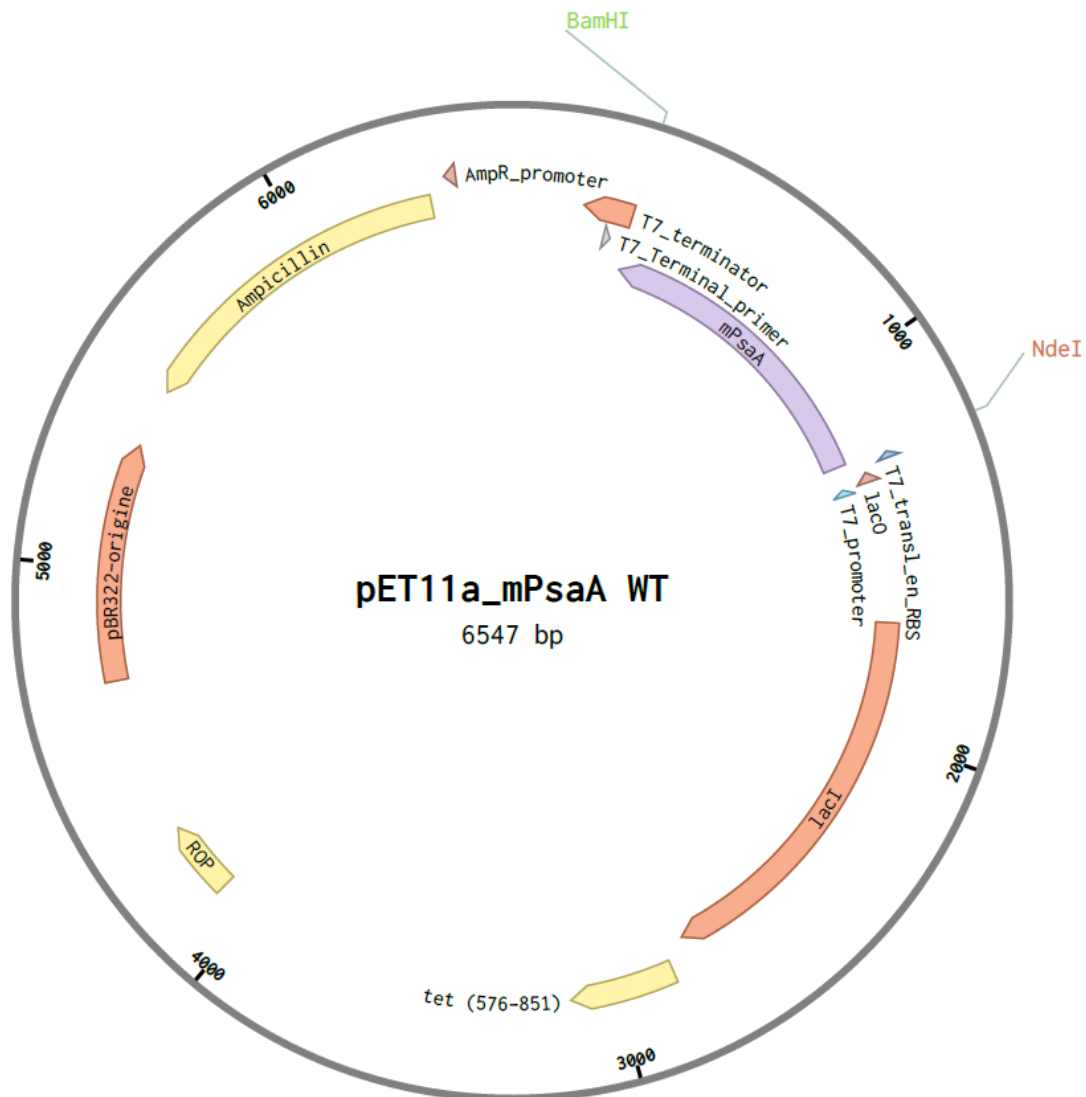
Sérotype	Composition saccharidique
1	→3)-α-AATGalp-(1→4)-α-D-GalpA _{2,0,3,3,0,3} Ac ₂ -(1→3)-α-D-GalpA-(1→
2	→4)-β-D-Glcp-(1→3)-α-L-Rhap-(1→3)-α-L-Rhap-(1→3)-β-L-Rhap-(1→ 2 ↑ 1 α-D-GlcpA-(1→6)-α-D-Glcp
3	→3)-β-D-GlcpA-(1→4)-β-D-Glcp-(1→
4	→3)-β-D-ManpNAc-(1→3)-α-L-FucpNAc-(1→3)-α-D-GalpNAc-(1→4)-α-D-Galp _{2,3(S)} Pyr-(1→
5	→4)-β-D-Glcp-(1→4)-α-L-FucpNAc-(1→3)-β-D-Sugp-(1→ 3 ↑ 1 α-L-PnepNAc-(1→2)-β-D-GlcpA
6B	→2)-α-D-Galp-(1→3)-α-D-Glcp-(1→3)-α-L-Rhap-(1→4)-D-Rib-ol-(5→P→
7F	→6)-α-D-Galp-(1→3)-β-L-Rhap ₂ Ac-(1→4)-β-D-Glcp-(1→3)-β-D-GalpNAc-(1→ 2 ↑ 1 β-D-Galp 4 ↑ 1 α-D-GlcpNAc-(1→2)-α-L-Rhap
8	→4)-β-D-GlcpA-(1→4)-β-D-Glcp-(1→4)-α-D-Glcp-(1→4)-α-D-Galp-(1→
9N	→4)-α-D-GlcpA-(1→3)-α-D-Glcp-(1→3)-β-D-ManpNAc-(1→4)-β-D-Glcp-(1→4)-α-D-GlcpNAc-(1→
9V	→4)-α-D-GlcpA _{2,0,25,3,0,55} Ac ₂ -(1→3)-α-D-Galp-(1→3)-β-D-ManpNAc _{4,0,09,6,1,04} Ac ₂ -(1→4)-β-D-Glcp-(1→4)-α-D-Glcp-(1→
10A	→5)-β-D-Galf-(1→3)-β-D-Galp-(1→4)-β-D-GalpNAc-(1→3)-α-D-Galp-(1→4)-D-Rib-ol-(5→P→ β-D-Galp 1 ↓ 6
11A	→6)-α-D-Glcp _{2,0,6,3,0,5} Ac ₂ -(1→4)-α-D-Galp-(1→3)-β-D-Galp _{4,6,0,5} Ac ₂ -(1→4)-β-D-Glcp-(1→ 4 ↑ Gro-(1→P

12F	$\begin{array}{ccc} \rightarrow 4) - \alpha\text{-L-FucpNAc} - (1 \rightarrow 3) - \beta\text{-D-GalpNAc} - (1 \rightarrow 4) - \beta\text{-D-ManpNAcA} - (1 \rightarrow & & \\ & \begin{array}{c} 3 \\ \uparrow \\ 1 \end{array} & \begin{array}{c} 3 \\ \uparrow \\ 1 \end{array} \\ & \alpha\text{-D-Galp} & \alpha\text{-D-Glcp} - (1 \rightarrow 2) - \alpha\text{-D-Glcp} \end{array}$
14	$\begin{array}{c} \rightarrow 6) - \beta\text{-D-GlcpNAc} - (1 \rightarrow 3) - \beta\text{-D-Galp} - (1 \rightarrow 4) - \beta\text{-D-Glcp} - (1 \rightarrow \\ 4 \\ \uparrow \\ 1 \\ \beta\text{-D-Galp} \end{array}$
15B	$\begin{array}{c} \rightarrow 6) - \beta\text{-D-GlcpNAc} - (1 \rightarrow 3) - \beta\text{-D-Galp} - (1 \rightarrow 4) - \beta\text{-D-Glcp} - (1 \rightarrow \\ 4 \\ \uparrow \\ 1 \\ \alpha\text{-D-Galp}_{2_{0.06}, 3_{0.12}, 4_{0.12}, 6_{0.55}} \text{Ac}_4 (1 \rightarrow 2) - \beta\text{-D-Galp} \\ 3 \\ \uparrow \\ \text{Gro}_{0.7} - (2 \rightarrow P) \end{array}$
17F	$\begin{array}{c} \rightarrow 3) - \beta\text{-L-Rhap} - (1 \rightarrow 4) - \beta\text{-D-Glcp} - (1 \rightarrow 3) - \alpha\text{-D-Galp} - (1 \rightarrow 3) - \beta\text{-L-Rhap} 2\text{Ac} - (1 \rightarrow 4) - \alpha\text{-L-Rhap} - (1 \rightarrow 2) - \text{D-Ara-ol} - (1 \rightarrow P \rightarrow \\ 4 \\ \uparrow \\ 1 \\ \alpha\text{-D-Galp} \end{array}$
18C	$\begin{array}{c} \text{Gro} - (1 \rightarrow P \\ \downarrow \\ 3 \\ \rightarrow 4) - \beta\text{-D-Glcp} - (1 \rightarrow 4) - \beta\text{-D-Galp} - (1 \rightarrow 4) - \alpha\text{-D-Glcp} - (1 \rightarrow 3) - \beta\text{-L-Rhap} - (1 \rightarrow \\ 2 \\ \uparrow \\ 1 \\ \alpha\text{-D-Glcp} 6\text{Ac}_{0.3} \end{array}$
19A	$\rightarrow 4) - \beta\text{-D-ManpNAc} - (1 \rightarrow 4) - \alpha\text{-D-Glcp} - (1 \rightarrow 3) - \alpha\text{-L-Rhap} - (1 \rightarrow P \rightarrow$
20	$\begin{array}{c} \beta\text{-Gal} f 2\text{Ac}_{0.9} \\ 1 \\ \downarrow \\ 4 \\ \rightarrow 3) - \alpha\text{-GlcpNAc} - (1-P-6) - \alpha\text{-Glcp} - (1 \rightarrow 6) - \beta\text{-Glcp} - (1 \rightarrow 3) - \beta\text{-Gal} f 5_{0.9}, 6_{0.9} \text{Ac}_2 - (1 \rightarrow 3) - \beta\text{-Glcp} - (1 \rightarrow \end{array}$

22F	$\begin{array}{c} \rightarrow 4)-\beta\text{-D-GlcpA}-(1\rightarrow 4)-\beta\text{-L-Rhap}2\text{Ac}_{0.8}-(1\rightarrow 4)-\alpha\text{-D-Glcp}-(1\rightarrow 3)-\alpha\text{-D-Galf}-(1\rightarrow 2)-\alpha\text{-L-Rhap}-(1\rightarrow \\ 3 \\ \uparrow \\ 1 \\ \alpha\text{-D-Glcp} \end{array}$
23F	$\begin{array}{c} \text{Gro}-(2\rightarrow P \\ \downarrow \\ 3 \\ \rightarrow 4)-\beta\text{-D-Glcp}-(1\rightarrow 4)-\beta\text{-D-Galp}-(1\rightarrow 4)-\beta\text{-L-Rhap}-(1\rightarrow \\ 2 \\ \uparrow \\ 1 \\ \alpha\text{-L-Rhap} \end{array}$
33F	$\begin{array}{c} \rightarrow 3)-\beta\text{-D-Galp}-(1\rightarrow 3)-\alpha\text{-D-Galp}-(1\rightarrow 3)-\beta\text{-D-Galf}-(1\rightarrow 3)-\beta\text{-D-Glcp}-(1\rightarrow 5)-\beta\text{-D-Galf}2\text{Ac}_{0.5}-(1\rightarrow \\ 2 \\ \uparrow \\ 1 \\ \alpha\text{-D-Galp} \end{array}$

Annexe 2 : Présentation des plasmides

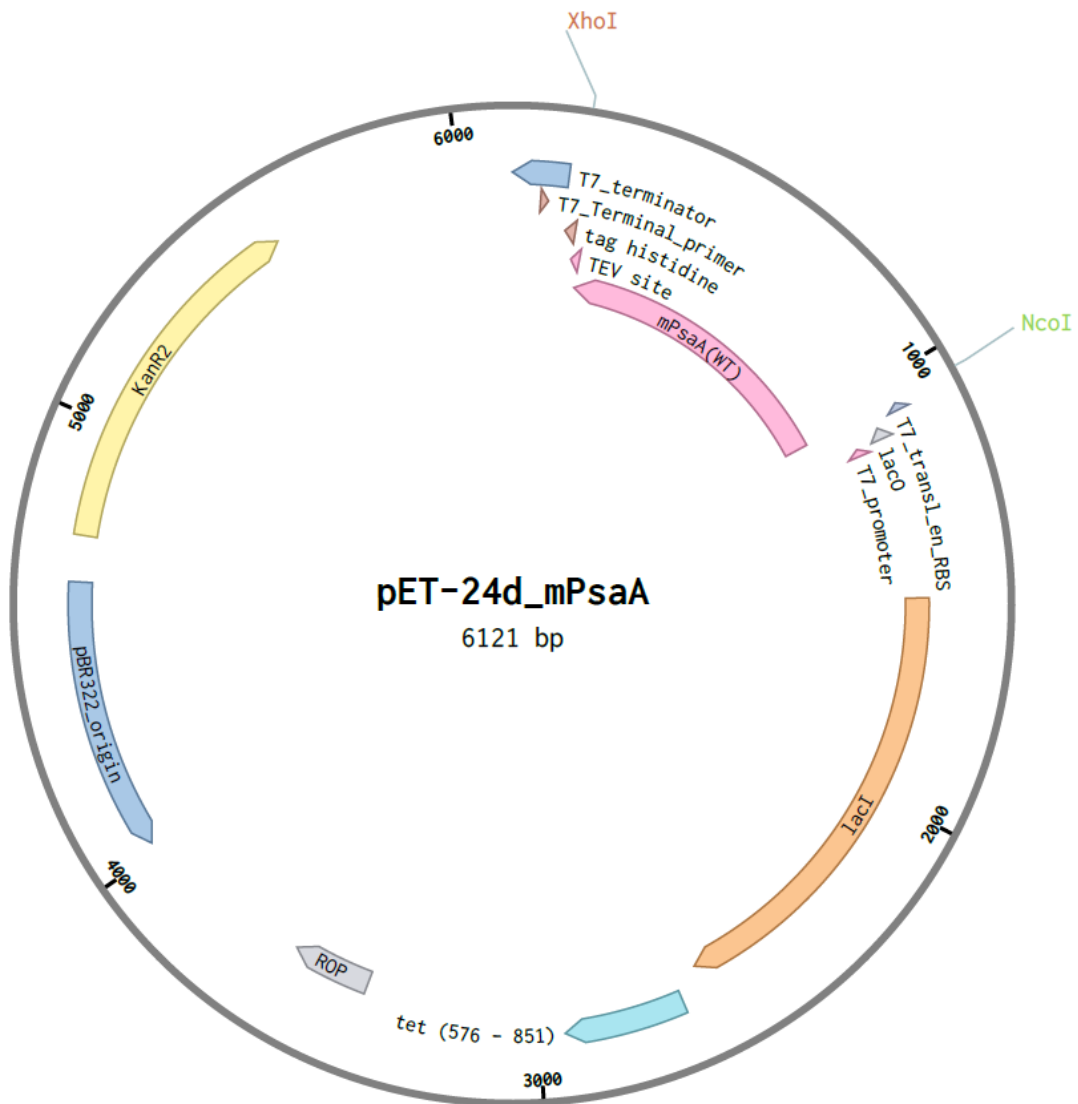
Plasmide pET11a



Ce plasmide a permis la production de la mPsaA avec un tag histidine en N-terminal. Il porte un gène de résistance à l'ampicilline, ainsi que le gène lacI et possède l'origine de répliation pBR322. Les mutants K24, K213 et K229, ont été générés au sein de l'insert mPsaA sur ce plasmide.

Ce plasmide a également été utilisé pour la production des protéines cibles CBM40, CBM32, AAL-S2, NagBb, GH98Sp, ainsi que CLC1. Pour cela, les gènes correspondant à ces protéines avec le tag Histidine présent en C-terminal ont été clonés à la place de l'insert mPsaA. Cela a conduit aux plasmides pEt11a-CBM40, pET11a-CBM32, pET11a-AAL(S2), pET11a-NagBb, pET11a-GH98Sp et pET11a-CLC1.

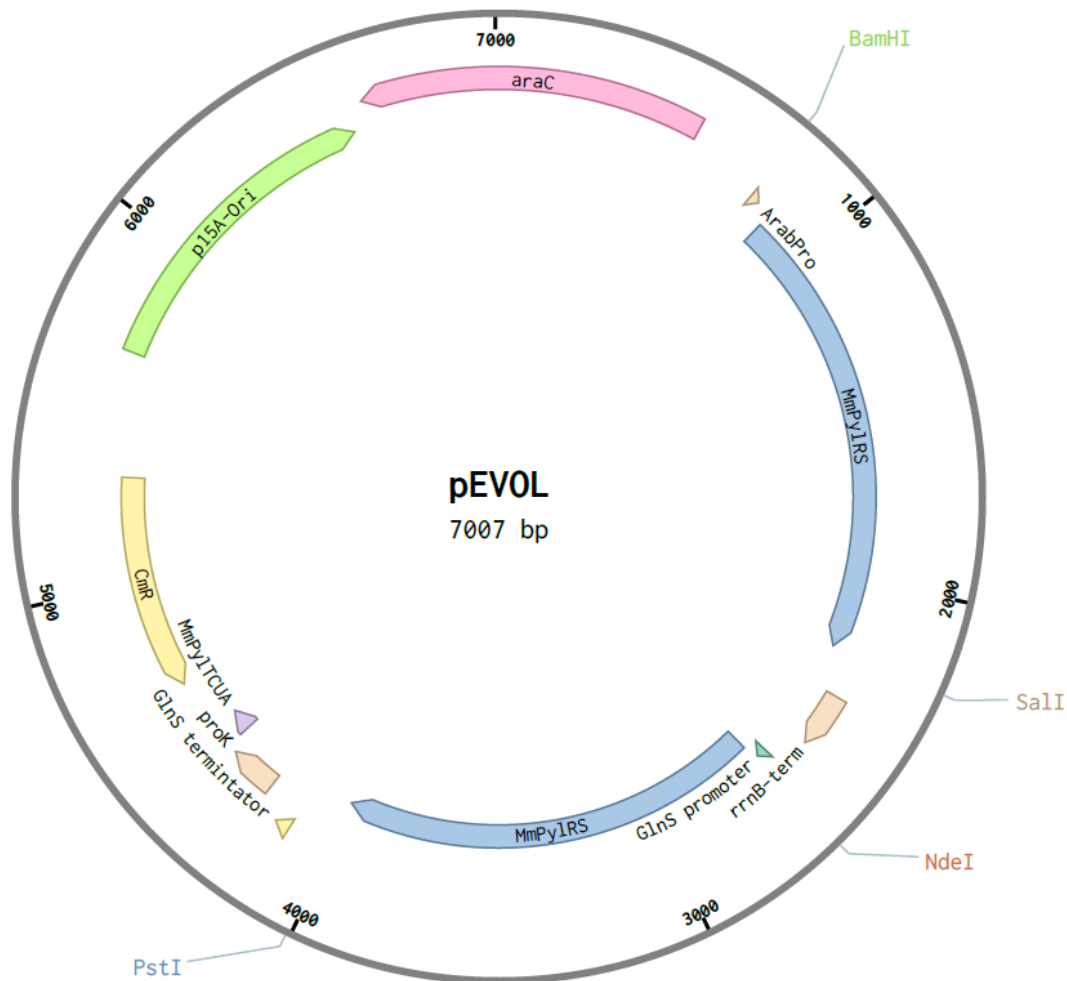
Plasmide pET24d



Ce plasmide a permis la production de la mPsaA avec un tag histidine en C-terminal. Il porte un gène de résistance à la Kanamycine, ainsi que le gène lacI et possède l'origine de réplcation pBR322. Les mutants K24, K213 et K229, ont été générés au sein de l'insert mPsaA sur ce plasmide.

Ce plasmide a également permis l'optimisation de l'incorporation de la PrK et la sélection des différentes synthétases. Pour cela, le gène de l'eGFP a été cloné en fusion du gène de la mPsaA. Ce qui a conduit au plasmide pET24d-mPsaA-GFP. Des mutations ont ensuite été générées au sein de la mPsaA en position K193 et K105 afin de générer les plasmides pET24d-K193GFP et pET24d-K105-K193-GFP.

Plasmide pEVOL



Ce plasmide a permis l'incorporation de la Propargyl-Lysine. Il porte un gène de résistance au chloramphenicol, ainsi que le gène *araC* et possède l'origine de réplication *p15a*. Il porte deux copies de gènes de la Pyrrolysyl-tRNA synthetase sauvage issue de *Methanosarcina mazei* (*MmPyIRS*), une sous le contrôle d'un promoteur inductible à l'arabinose (*AraPro*), et une sous le contrôle d'un promoteur constitutif (*GlnS promoter*). Il porte également une copie de l'ARNt ayant un codon complémentaire au codon TAG sous le contrôle d'un promoteur constitutif (*proK*).

Ce plasmide a également été modifié en enlevant la copie présente sous le contrôle du promoteur constitutif, ce qui a conduit au plasmide pEVOL-1copie.

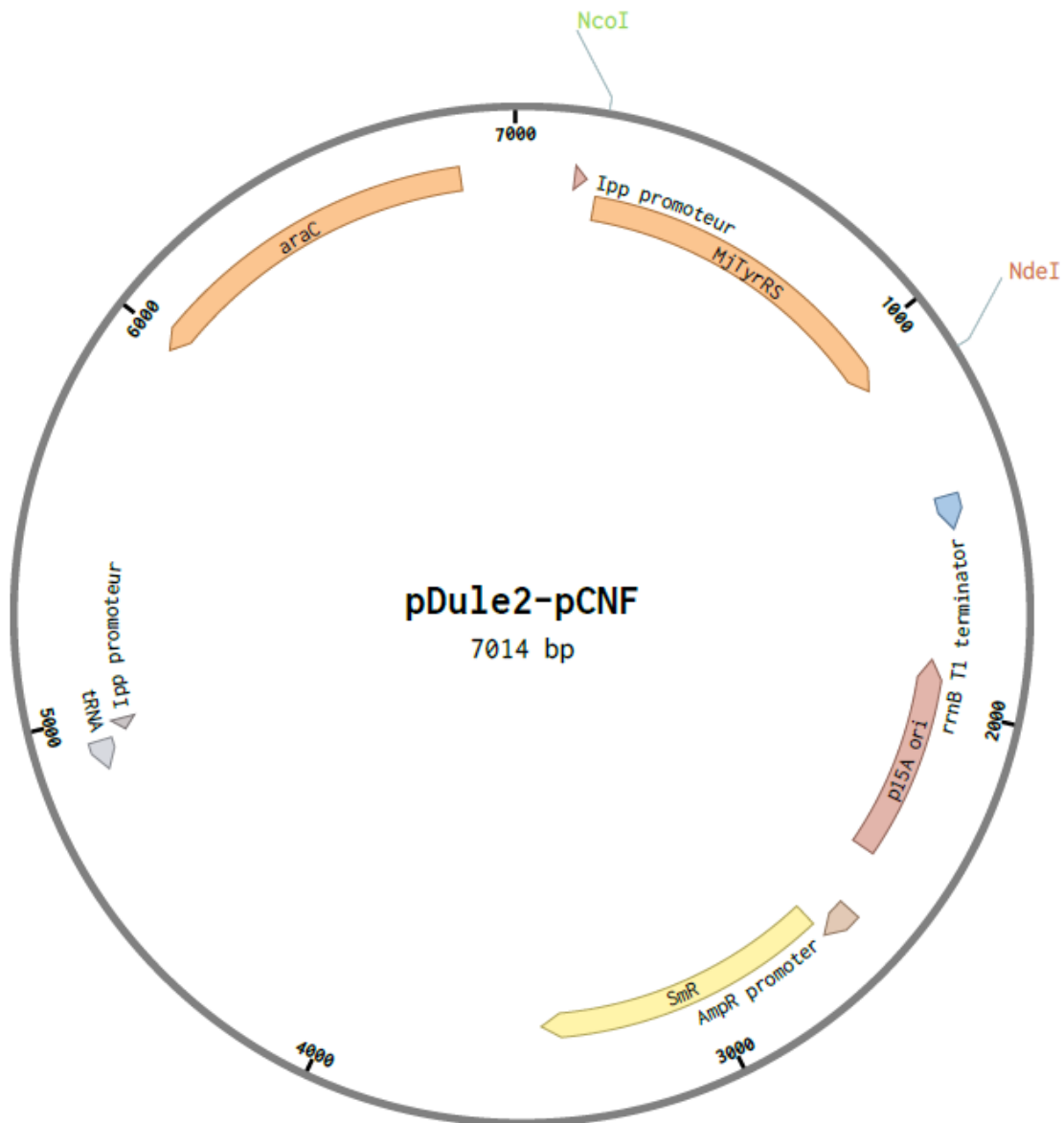
Ce plasmide pEVOL-1copie a ensuite été modifié de différente façon :

- pour l'optimisation de l'incorporation de la PrK : le gène du transporteur *LysP* a été inséré à la place du gène de la Pyrrolysyl-tRNA synthetase (*MmPyIRS*) sous le contrôle du promoteur *GlnS*, ce qui a conduit au plasmide pEVOL-1copie-*LysP*.

- pour tester les différentes mutations de l'aaRS prédites pour permettre d'incorporer l'ONpGlu : des mutations (Y384F, Y306F, N346A/G, C348A/G et W417C/N/A) ont été générées de façon combinatoire au sein du gène de la Pyrrolysyl-tRNA synthetase (*MmPyIRS*), ce qui a conduit aux différents plasmides testés lors de la sélection.

- pour l'incorporation de l'AcidK : les mutations L309T, C348G et Y384F ont été générées au sein du gène de la Pyrrolysyl-tRNA synthetase (*MmPyIRS*), ce qui a conduit au plasmide pEVOL-AcK

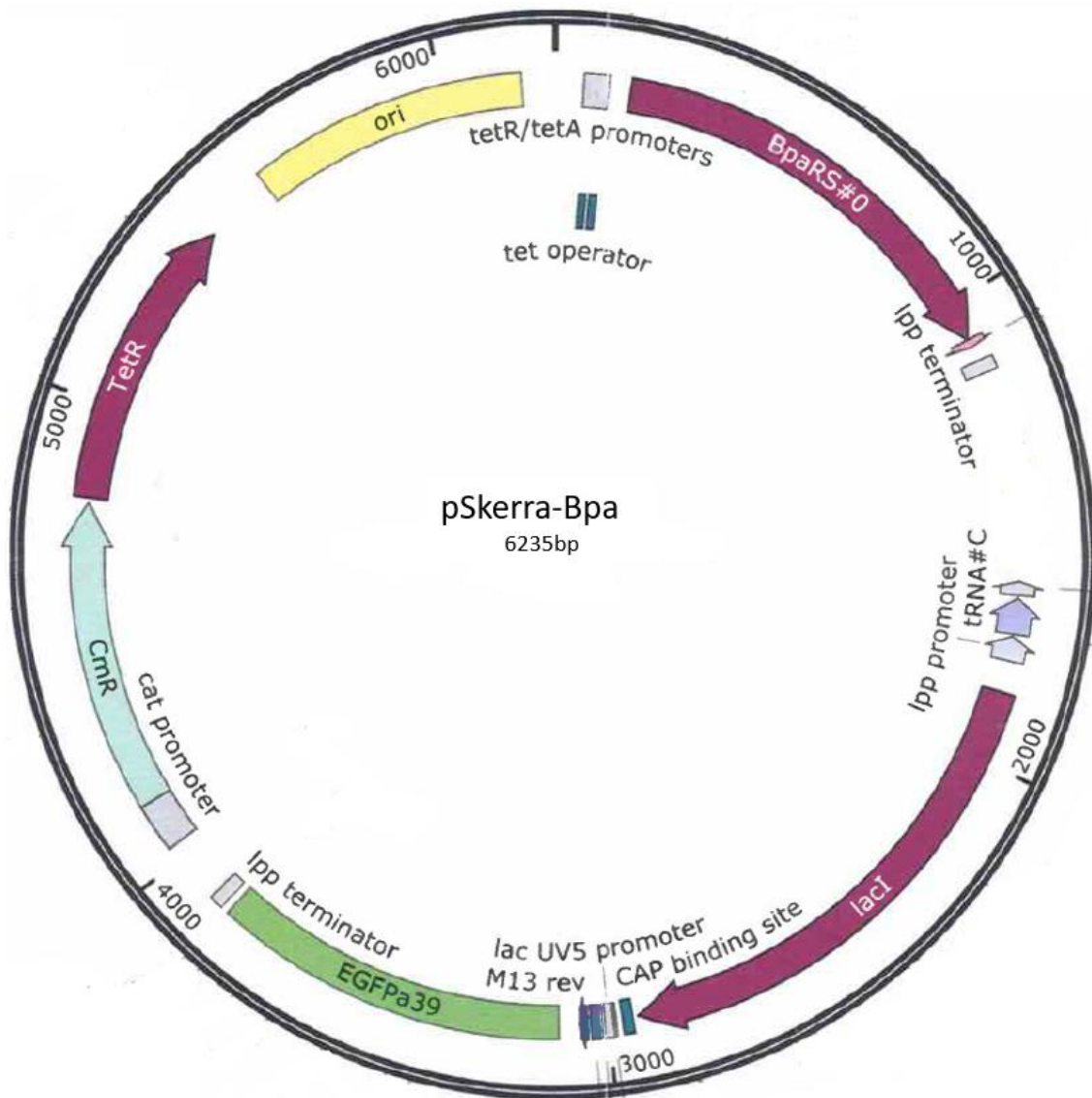
pDULE.2



Ce plasmide porte un gène de résistance à la Spectinomycine (SmR) ; ainsi que le gène *araC* et le gène de la Tyrosyl-tRNA synthétase issue de *Methanocaldococcus jannaschii* (MjTyrRS) portant les mutations nécessaires pour l'incorporation de l'azido-phénylalanine (Y32L, L65V, F108W, Q109M, D158G et I159P). Ce gène est sous le contrôle d'un promoteur constitutif Ipp, de même que le gène codant pour le tRNA ayant un codon complémentaire au codon stop TAG.

Ce plasmide a été modifié afin de générer les mutations au sein du gène de la Tyrosyl-tRNA synthétase nécessaire à l'incorporation de la phénylalanine boronate (Y32S, L65A, H70M, D158S, and L162E), ce qui a conduit au plasmide pDULE.2-Bpa.

pSkerra-Bpa



Ce plasmide porte un gène de résistance au chloramphenicol (CmR) ; ainsi que le gène la Tyrosyl-tRNA synthétase issue de *Methanocaldococcus jannaschii* (MjTyrRS) portant les mutations nécessaires pour l'incorporation de la boronophénylanine (Y32S, L65A, H70M, D158S, L162E et D286R). Ce gène est sous le contrôle d'un promoteur inductif à la doxycycline (tetR/tetA), le gène codant pour le tRNA ayant un codon complémentaire au codon stop TAG est quant à lui sous le contrôle d'un promoteur constitutif (lpp).

Annexe 3 : Liste des souches et des plasmides

Souche d'expression	Utilité	Particularité
BL21(DE3)	Souche d'expression protéique	
B95ΔA	Souche d'expression optimisée pour l'incorporation multiple d'acides aminés non naturels, dérivée de la B21(DE3)	Gène du Releasing Factor 1 (RF1) délété, et remplacement de 95 codons stops TAG par d'autres codons stops (TAA ou TGA)

Numéro	Vecteur	Utilité	Résistance
1	pET11a-mPsaA	Expression mPsaA avec le tag histidine en N terminal	Ampicilline
2	pET24d-mPsaA	Expression mPsaA avec le tag histidine en C terminal	Kanamycine
3	pEVOL	Incorporation d'acides aminés non naturels (PrK) à l'aide de la pyrrolysine tRNA synthétase issue de <i>Methanosarcina mazei</i> , issu du laboratoire de Lemke (EMBL)	Chloramphenicol
4	pEVOL-1copie	Suppression de la copie de la pyrrolysine tRNA synthétase sous le contrôle du promoteur constitutif, simplification des procédures de clonages et séquençages lors de la sélection d'une nouvelle synthétase	Chloramphenicol
5	pEVOL-1copie-LysP	Ajout d'un transporteur pour la Lysine pour augmenter l'efficacité de l'incorporation de la PrK	Chloramphenicol
6	pEVOL-AcdK	Plasmide pEVOL-1copie possédant la version mutée de la tRNA synthétase (mutations : L309T, C3648G et Y384F)	Chloramphenicol
7	pET11a	Plasmide également utilisé pour l'expression des protéines cibles de l'AcdK	Ampicilline

		et du Bpa, mais avec un tag histidine en C terminal	
8	pDULE.2	Plasmide portant la Tyrosyl tRNA synthetase issue de <i>Methanocaldococcus jannaschii</i> , Issu du laboratoire de Ryan Mehl (Oregon)	Spectinomycine
9	pDULE.2-Bpa	Plasmide portant la Tyrosyl tRNA synthetase issue de <i>Methanocaldococcus jannaschii</i> , possédant les mutations : Y32S, L65A, H70M, D158S, and L162E.	Spectinomycine
10	pSkerra-Bpa	Plasmide portant la Tyrosyl tRNA synthetase issue de <i>Methanocaldococcus jannaschii</i> , possédant les mutations : Y32S, L65A, H70M, D158S, and L162E issus du laboratoire de Skerra (Allemagne)	Chloramphenicol
11	pBAD	Plasmide disponible au sein du laboratoire, à partir duquel nous avons amplifié le promoteur inductible à l'arabinose	Chloramphenicol

Annexe 4 : Liste des amorces oligonucléotidiques

Numéro amorce	Amorce	Séquence (5' → 3')	Utilité
1	mPsaA_XhoI_pET24d	GAGACTCGAGACCGCTTTGGAAGTACAGATTTTCTTTTCGC CAGACCCTCCGCAATCT	Amplification mPsaA pour clonage dans pET24d
2	mPsaA_NcoI_pET24d	GAGACCATGGGTAAGAAAGATACCACCAGCG	Amplification mPsaA pour clonage dans pET24d
3	T7P	TAATACGACTCACTATAGG	Vérification clonage dans les plasmides pET
4	T7T	GCTAGTTATTGCTCAGCGG	Vérification clonage dans les plasmides pET
5	F_K24ambre_mPsaA	CCAGCGGCCAATAGCTGAAAGTGGTTG	Mutagenèse pour générer un codon TAG au niveau d'un site d'incorporation choisi
6	R_K24ambre_mPsaA	CAACCACTTTTCAGCTATTGGCCGCTGG	Mutagenèse pour générer un codon TAG au niveau d'un site d'incorporation choisi
7	mPsaA-K213TAG-fw	CGCCATACGCCTAGCTAAAGTATTTGAACGCACCCTCGCT GG	Mutagenèse pour générer un codon TAG au niveau d'un site d'incorporation choisi
8	mPsaA-K213TAG-rev	CTTTAGCTAGGCGTATGGCGTTCGAGCGGTACATCTG GGAGATTAACAC	Mutagenèse pour générer un codon TAG au niveau d'un site d'incorporation choisi
9	mPsaA_K229TAG_fw	CGCAGCTTTTCCACCAGGGTCTAGATTTGCTCCGGGGTG CCC	Mutagenèse pour générer un codon TAG au niveau d'un site d'incorporation choisi
10	mPsaA_K229TAG_rev	GGGCACCCCGGAGCAAATCTAGACCCTGGTGGAAAAGC TGCG	Mutagenèse pour générer un codon TAG au niveau d'un site d'incorporation choisi
11	pEVOL_1copie_fw	CTGCAGTTTCAAACGCTAAAT	Suppression de la copie de la pyrrolysine tRNA synthétase sous le contrôle du promoteur constitutif
12	pEVOL_1copie_rev	ATGGGATTCCCTCAAAGCGTA	Suppression de la copie de la pyrrolysine tRNA synthétase sous le

			contrôle du promoteur constitutif
13	pEVOL1copy_seq_Fw	GGATCCTACCTGACGCTTTTTATCGC	Séquençage de la copie restante sur le pEVOL-1copie
14	pEVOL1copy_seq_Rev	CTGCGTTCTGATTTAATCTGTATC	Séquençage de la copie restante sur le pEVOL-1copie
15	LysP-gibson-F	ACGCTTTGAGGAATCCCATATGGTTTCCGAAAC	Amplification du transporteur LysP pour son clonage dans pEVOL-1copie (en méthode Gibson et méthode par Enzymes de restriction)
16	LysP-gibson-R	GCGTTTGAAACTGCAGTTATTTCTTATCGTTC	Amplification du transporteur LysP pour son clonage dans pEVOL-1copie (en méthode Gibson et méthode par Enzymes de restriction)
17	pLysP-mut-fw	GAGCTGATCGGTATTGCCGCAGGCGAGTCCG	Suppression du site de restriction PstI au sein du gène LysP
18	pLysP-mut-rev	CGGACTCGCCTGCGGCAATACCGATCAGCTC	Suppression du site de restriction PstI au sein du gène LysP
19	pEVOL-aaRS-R19H-fw	GTGGATGAGTCATACCGGAACCATTCAAAAATCAAACA C	Mutagenèse de la partie N-terminale de la pyrrolysine tRNA synthetase pour augmenter son affinité avec le tRNA
20	pEVOL-aaRS-R19H-rev	CCGGTATGACTCATCCACAGACCAGTAGCAGAGATCAGG GTATTC	Mutagenèse de la partie N-terminale de la pyrrolysine tRNA synthetase pour augmenter son affinité avec le tRNA
21	pEVOL-aaRS-H29R-fw	CAAACACCGCGAGGTTAGCCGTTGAAAATCTATATTGA G	Mutagenèse de la partie N-terminale de la pyrrolysine tRNA synthetase pour augmenter son affinité avec le tRNA
22	pEVOL-aaRS-H29R-rev	CCTCGCGGTGTTTGATTTTATGAATGGTTCCGGTACGAC	Mutagenèse de la partie N-terminale de la pyrrolysine tRNA synthetase pour augmenter son affinité avec le tRNA
23	pEVOL-aaRS-T122S-fw	CTGGAAAACCTCTGAAGCAGCACAGGCACAGCCGTCTGGA AGC	Mutagenèse de la partie N-terminale de la pyrrolysine tRNA synthetase pour augmenter son affinité avec le tRNA
24	pEVOL-aaRS-T122S-rev	CTTCAGAGTTTTCCAGTGGTTTAGGGGCACGAGCAACGG	Mutagenèse de la partie N-terminale de la pyrrolysine tRNA synthetase pour augmenter son affinité avec le tRNA
25	F_K193ambre_mPsaA	GGACAAATTTAACTAGATCCCGCGGAGAAG	Mutagenèse pour incorporation NAA en position K193

26	R_K193ambre_mPsaA	CTTCTCCGCCGGGATCTAGTTAAATTTGTCC	Mutagenèse pour incorporation NAA en position K193
27	F_K105ambre_mPsaA	CGAAGAAAACCGAAAAGTAGGATTACTTCGCGG	Mutagenèse pour incorporation NAA en position K105
28	R_K105ambre_mPsaA	CCGCGAAGTAATCCTAGTTTTCGGTTTTCTTCG	Mutagenèse pour incorporation NAA en position K105
29	pLibr-Y306F-fw	CACCAAATCTGTTTAACTATCTGCGCAAAGTGGACCGTGC CCTGCC	Oligo dégénérés pour mutagenèse sur aaRS
30	pLibr-Y306F-rev	GATAGTTAAACAGATTTGGTGCCAGCATAGGGCGCAGAC AGAAGTTTTTATCC	Oligo dégénérés pour mutagenèse sur aaRS
31	pLibr-Y384F-fw	CTGTATGGTGTGGCGACACCCTGGATGTCATGCACGG CGAC	Oligo dégénérés pour mutagenèse sur aaRS
32	pLibr-Y384F-rev	GTCGCCAAACACCATACAGCTGTCGCCACAATTTGAAG TCAATGCCC	Oligo dégénérés pour mutagenèse sur aaRS
33	pLibr-W417C-fw	CGACAAACCGTGCATCGGAGCAGGATTCGGTCTGGAAC GCC	Oligo dégénérés pour mutagenèse sur aaRS
34	pLibr-W417C-rev	CGATGCACGGTTTGTGATACCCCACTCACGGTCCAGCG GAATTGG	Oligo dégénérés pour mutagenèse sur aaRS
35	pLibr-W417N-fw	CGACAAACCGAACATCGGAGCAGGATTCGGTCTGGAAC GCC	Oligo dégénérés pour mutagenèse sur aaRS
36	pLibr-W417N-rev	CGATGTTCCGGTTTGTGATACCCCACTCACGGTCCAGC GGAATTGG	Oligo dégénérés pour mutagenèse sur aaRS
37	pLibr-W417A-fw	CGACAAACCGGCGATCGGAGCAGGATTCGGTCTGGAAC GCC	Oligo dégénérés pour mutagenèse sur aaRS
38	pLibr-W417A-rev	CGATCGCCGGTTTGTGATACCCCACTCACGGTCCAGC GGAATTGG	Oligo dégénérés pour mutagenèse sur aaRS

39	pLibr-N346A-C348G-fw	CCATGCTGGCCTTTGGCCAAATGGGTTTCAGGTTGTACTC GTG	Oligo dégénérés pour mutagénèse sur aaRS
40	pLibr-N346A-C348G-rev	CATTTGGCCAAAGGCCAGCATGGTAACTCCTCCAGATG TTCTTTACCGTCGGAC	Oligo dégénérés pour mutagénèse sur aaRS
41	pLibr-N346A-C348A-fw	CCATGCTGGCCTTTGCCAAATGGGTTTCAGGTTGTACTCG TG	Oligo dégénérés pour mutagénèse sur aaRS
42	pLibr-N346A-C348A-rev	CATTTGGGCAAAGGCCAGCATGGTAACTCCTCCAGATG TTCTTTACCGTCGGAC	Oligo dégénérés pour mutagénèse sur aaRS
43	pLibr-N346G-C348G-fw	CCATGCTGGGCTTTGGCCAAATGGGTTTCAGGTTGTACTC GTG	Oligo dégénérés pour mutagénèse sur aaRS
44	pLibr-N346G-C348G-rev	CATTTGGCCAAAGCCCAGCATGGTAACTCCTCCAGATGT TCTTTACCGTCGGAC	Oligo dégénérés pour mutagénèse sur aaRS
45	CBM40-F48E-fw	CAGGTTGTATTCCGCCGGCGCTTTCGCGTCCGGCTTAA ATTC	Mutagénèse pour changer F48 en E
46	CBM40-F48E-rev	CCGGCGGAATACAACCTGTTTAGCGTGAG CAGCGCGACCAAGAAAG	Mutagénèse pour changer F48 en E
47	CBM40-F48-fw	CAGGTTGTACTACGCCGGCGCTTTCGCGT CCGGCTTAAATTC	Mutagénèse pour incorporation NAA en position F48
48	CBM40-F48-rev	CCGGCGTAGTACAACCTGTTTAGCGTGAGCAGCGCGACC AAGAAAG	Mutagénèse pour incorporation NAA en position F48
49	CBM32-Y36E-fw	GCGCCTTCCGCGCTCGCCCAACGGGTCGCCG	Mutagénèse pour changer Y36 en E
50	CBM32-Y36E-rev	GCGCGGAAGGCGCGAGCCCGCAGTGGA	Mutagénèse pour changer Y36 en E
51	CBM32-E14*-fw-2	CCGCGGTCTAAATGCTGCTGGTCGCGGTGCCTTGTTTC	Mutagénèse pour incorporation NAA en position E14
52	CBM32-E14*-rev-2	GCATTTAGACCGCGGGCCACGAAGGCGACAAAGCGG	Mutagénèse pour incorporation NAA en position E14

53	CBM32-Y36*-fw-2	GCGCCCTACGCGCTCGCCCAACGGGTCGCCG	Mutagenèse pour incorporation NAA en position Y36
54	CBM32-Y36*-rev-2	GCGCGTAGGGCGCGAGCCCGCAGTGA	Mutagenèse pour incorporation NAA en position Y36
55	AAL-E41TAG-for	GAGCTAGTTCGTTTATGATGGCAGCAAATGGATTACC	Mutagenèse pour incorporation NAA en position E41
56	AAL-E41TAG-rev	CATAAACGAACTACGACAGGATGTTGTCCTTGTTACG	Mutagenèse pour incorporation NAA en position E41
57	paaRS-L309T-fw	GTATAACTATACGCGCAAACCTGGACCGTGCCTGCCTGA TCC	Mutagenèse pour obtenir la mutation L309T sur la tRNA synthetase
58	paaRS-L309T-rev	CAGTTTGCGCGTATAGTTATACAGATTTGGTGCCAGCATA GGGCGC	Mutagenèse pour obtenir la mutation L309T sur la tRNA synthetase
59	paaRS-C348G-fw	GAACCTTGCCAAATGGGTTTCAGGTTGACTCGTGAGAA CCTG	Mutagenèse pour obtenir la mutation C348G sur la tRNA synthetase
60	paaRS-C348G-rev	CCATTTGGCCAAAGTTCAGCATGGTAACTCCTCCAGATG TTC	Mutagenèse pour obtenir la mutation C348G sur la tRNA synthetase
61	paaRS-Y384F-fw	CTGTATGGTGTGTTGGCGACACCCTGGATGTCATGCACGG CGAC	Mutagenèse pour obtenir la mutation Y384F sur la tRNA synthetase
62	paaRS-Y384F-rev	GTCGCCAAACACCATACAGCTGTCGCCACAATTTGAAG TCAATGCCC	Mutagenèse pour obtenir la mutation Y384F sur la tRNA synthetase
63	pcloning-Bpa1	CAAAACCTATCGAAGCAGATTTTTCATCTTTTTTAAAACC TC	Linéarisation du vecteur pDULE.2 afin de cloner une partie du gène de l'aaRS pour obtenir l'aaRS de Bpa
64	pcloning-Bpa2	CATTATGAAGGCGTTGATGTTGCAGTTGGAGGGATGGA GCAGAG	Linéarisation du vecteur pDULE.2 afin de cloner une partie du gène de l'aaRS pour obtenir l'aaRS de Bpa
66	pDULE.2 seq-fw	GGCGCCGCTTCTTTGAGC	Séquençage de l'aaRS (TyrRS de <i>M.jannaschii</i>) présente sur le plasmide pDULE.2

67	pDULE.2 seq-rev	GACGGCGCTATTCAGATCC	Séquençage de l'aaRS (TyrRS de <i>M.jannaschii</i>) présente sur le plasmide pDULE.2
68	Para-fw	CTGGCGCCGCTTCTTTGAGCGCTATGCCATAGCAT	Récupération du promoteur arabinose sur pEVOL
69	Para-rev	ATCATTTCAAATTCGTCCATAGATCTAATTCCTCC	Récupération du promoteur arabinose sur pEVOL
70	plinear-bis-pDULE.1	GCGCTCAAAGAAGCGGCGCCAGGGTTGTTTTCTTTTCAC C	Linéarisation du vecteur pDULE
71	plinear-bis-pDULE.2	CTATGGACGAATTTGAAATGATAAAGAGAAACACATCTG AAATTATCAG	Linéarisation du vecteur pDULE
72	pGH98sp-H60*-fw	CCGGCTACAGTTGAATCACGGTGTACGGTTTCAGATCG	Mutagenèse pour générer un codon stop TAG
73	pGH98sp-H60*-rev	GATTCAACTGTAGCCGGATGATCTGCCGAAGCGTGATGG TGCGGCGC	Mutagenèse pour générer un codon stop TAG
74	pCLC1-H25*-fw	GTCAGGTTCTAACGAAAGAAGCCGTAGCAGCTATCACCA TAGTAACGC	Mutagenèse pour générer un codon stop TAG
75	pCLC1-H25*-rev	GCTTCTTTCGTTAGAACCTGACCTGGGAGGAAA GCAAACAGTATTGCACC	Mutagenèse pour générer un codon stop TAG
76	pNagBb-Y329*-fw	CGTTATCCTAACCCGGCTGCGCCCAACCATCCAGGTGC AG	Mutagenèse pour générer un codon stop TAG
77	pNagBb-Y329*-rev	CCGGGTTAGGATAACGGTCATCCGGAATCTGCGGCG TG	Mutagenèse pour générer un codon stop TAG

Annexe 5 : Packages de Rstudio pour réaliser des plans d'expériences

Commandes pour l'installation des différents packages :

- `install.packages(« DoE.base »)`
- `install.packages(« FrF2 »)`
- `install.packages(« DoE.wrapper »)`
- `install.packages(« RcmdrPlugin.DoE »)`
- `install.packages(« RcmdrPlugin.Export »)`
- `install.packages(« RcmdrPlugin.FactoMineR »)`
- `install.packages(« RcmdrPlugin.HH »)`
- `install.packages(« RcmdrPlugin.IPSUR »)`
- `install.packages(« RcmdrPlugin.SurvivalT »)`
- `install.packages(« RcmdrPlugin.TeachingDemos »)`
- `install.packages(« RcmdrPlugin.epack »)`
- `install.packages(« RcmdrPlugin.orloca »)`
- `install.packages(« Rcmdr »)`
- `install.packages(« conf.design »)`
- `install.packages(« lhs »)`
- `install.packages(« AlgDesign »)`
- `install.packages(« DiceDesign »)`
- `install.packages(« rsm »)`

Annexe 6 : Résultats d'une analyse de plan d'expérience Rstudio

```
rsm(formula = fluorescence ~ FO(x1, x2, x3, x4) + TWI(x1, x2,
x3, x4) + PQ(x1, x2, x3, x4), data = correction.coded)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	405.3600	439.2994	0.9227	0.3729508
x1	262.7308	253.6296	1.0359	0.3191419
x2	727.7367	253.6296	2.8693	0.0131611 *
x3	-1332.3133	253.6296	-5.2530	0.0001561 ***
x4	319.2875	253.6296	1.2589	0.2302131
x1:x2	145.9250	439.2994	0.3322	0.7450508
x1:x3	-589.9550	439.2994	-1.3429	0.2022675
x1:x4	-2.2675	439.2994	-0.0052	0.9959600
x2:x3	-730.1025	439.2994	-1.6620	0.1204290
x2:x4	970.4325	439.2994	2.2090	0.0457310 *
x3:x4	516.9125	439.2994	1.1767	0.2604219
x1^2	-367.8554	358.6864	-1.0256	0.3238033
x2^2	864.1833	358.6864	2.4093	0.0315311 *
x3^2	1233.3583	358.6864	3.4385	0.0044039 **
x4^2	377.2471	358.6864	1.0517	0.3120741

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Multiple R-squared:  0.8419,    Adjusted R-squared:  0.6717
F-statistic: 4.945 on 14 and 13 DF,  p-value: 0.003278
```

Analysis of Variance Table

Response: fluorescence

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FO(x1, x2, x3, x4)	4	29707578	7426894	9.6211	0.000763
TWI(x1, x2, x3, x4)	6	8445334	1407556	1.8234	0.171174
PQ(x1, x2, x3, x4)	4	15291614	3822903	4.9524	0.011995
Residuals	13	10035164	771936		
Lack of fit	10	10032595	1003260	1171.6594	0.00003693
Pure error	3	2569	856		

Stationary point of response surface:

x1	x2	x3	x4
0.2808803	-1.7564454	-0.3473650	2.0747919

Stationary point in original units:

Temperature	pH	Carabinose	TpsArabinose
0.2808803	-1.7564454	-0.3473650	2.0747919

Eigenanalysis:

eigen() decomposition

\$values

```
[1] 1503.64021 1076.28513 -49.27809 -423.71391
```

\$vectors

	[,1]	[,2]	[,3]	[,4]
x1	-0.15259447	-0.05358883	0.1334899	0.97776460
x2	-0.52052809	0.66182125	-0.5387262	0.02858666
x3	0.83950978	0.42369131	-0.2804334	0.19252558
x4	-0.03144474	0.61612218	0.7831421	-0.07805825

Annexe 7 : Liste des publications et communications

Publié :

Pillot A, Defontaine A, Fateh A, Lambert A, Prasanna M, Fanuel M, Pipelier M, Csaba N, **Violo T**, Camberlein E, Grandjean C. Site-Specific Conjugation for Fully Controlled Glycoconjugate Vaccine Preparation. *Front Chem.* 2019 Nov 1;7:726. doi: 10.3389/fchem.2019.00726. PMID: 31737603; PMCID: PMC6839274.

En cours de publication (en révision) :

Soumis après corrections en décembre 2019, en attente de réponse

Violo T, Dussouy C, Tellier C, Grandjean C, Camberlein E. Homogenous Glycoconjugate Vaccine Produced by Combined Unnatural Amino Acid Incorporation and Click-Chemistry. *JoVE*.

Communications :

-“Insertion of an unnatural amino acid to create a glycoconjugate vaccine against *Streptococcus pneumoniae*”, BioSynSys, 18-20th October 2017 (3rd Conference of the French Research Group on Synthetic Biology), La Grande Motte (France), oral presentation

-“Insertion of an unnatural amino acid to create a glycoconjugate vaccine against *Streptococcus pneumoniae*”, GlycoOuest, 12th January 2018 (Bretagne and Pays de la Loire research Network in Glycobiology), Rennes (France), oral presentation

-“Insertion of an unnatural amino acid to create a glycoconjugate vaccine against *Streptococcus pneumoniae*”, GCE Conferences, 9-11th august 2018 (Genetic code Expansion Conferences), Corvallis (Oregon, USA), poster

-“A new strategy to improve protein-sugar interaction properties with an Unnatural Amino Acid”, GFPP Conferences, 21-22th October 2019, (Paris), poster

-“Generation of homogenous monovalent and trivalent glycoconjugate vaccines against *Streptococcus pneumoniae* using unnatural amino acid incorporation”, JBS doctoral school, 10-11th December 2019, Anger (France), oral presentation

Annexe 8 : Publication JoVE (soumise)

Homogeneous Glycoconjugate Produced by Combined Unnatural Amino Acid Incorporation and Click-Chemistry for vaccine purpose

Typhaine Violo,¹ Christophe Dussouy,¹ Charles Tellier,¹ Cyrille Grandjean,¹ Emilie Camberlein¹

¹Université de Nantes, CNRS, Unité Fonctionnalité et Ingénierie des Protéines (UFIP), UMR 6286, F-44000 Nantes, France

E-MAIL ADDRESSES:

typhaine.violo@univ-nantes.fr; c.dussouy@gmail.com; charles.tellier@univ-nantes.fr;
cyrille.grandjean@univ-nantes.fr; emilie.camberlein@univ-nantes.fr

CORRESPONDING AUTHOR:

Emilie Camberlein at emilie.camberlein@univ-nantes.fr

KEYWORDS:

Synthetic biology, homogeneous glycoconjugate, site-selective bioconjugation, unnatural amino acid, genetic code expansion, click-chemistry, propargyl-L-Lysine, carbohydrate

SHORT ABSTRACT:

Genetic code expansion is applied for the introduction of an unnatural amino acid bearing a biorthogonal functional group on a carrier protein at a defined site. The biorthogonal function is further used for the site-selective coupling of a carbohydrate antigen to provide a homogeneous glycoconjugate vaccine.

LONG ABSTRACT:

Genetic code expansion is a powerful tool to introduce unnatural amino acids (UAAs) into proteins to modify their characteristics, to study or create new protein functions or to have an access to protein conjugates. Stop codon suppression, in particular amber codon suppression, has emerged as the most popular method to genetically introduce UAAs at defined positions. This methodology is herein applied to the preparation of a carrier protein containing an UAA harboring a biorthogonal functional group. This reactive handle can next be used to specifically and efficiently graft a synthetic oligosaccharide hapten to provide a homogeneous glycoconjugate vaccine. The protocol is limited to the synthesis of glycoconjugates in a one to one carbohydrate hapten/carrier protein ratio but amenable to numerous pairs of biorthogonal functional groups. Glycoconjugate vaccine homogeneity is an important criterion to ensure complete physico-chemical characterization thereby satisfying more and more demanding drug regulatory agencies recommendations, a criterion which is unmet by classical conjugation strategies. Moreover, this protocol makes possible to finely tune the structure of the actual conjugate vaccine giving rise to unprecedented tools to address structure-immunogenicity relationships.

INTRODUCTION:

Glycoconjugate vaccines are essential elements of the vaccine arsenal available for the prophylactic treatment of infectious diseases. They are safe, well-tolerated and efficient in a broad age group including young infants. They provide the optimal defense against infections caused by capsulated bacteria like meningococcus, pneumococcus or Haemophilus influenzae type b¹. Glycoconjugate vaccines are made of purified bacterial polysaccharides which form the capsules of bacteria or synthetic oligosaccharides that mimic these surface-expressed polysaccharides² which are covalently linked to a carrier protein. Presence of a carrier protein is essential to promote protective humoral immune response directed against the antigenic determinant expressed by the carbohydrate antigens³. Apart from a careful selection and production of the carbohydrate antigen, the features known to exert an influence on the efficacy of a glycoconjugate vaccine are: the nature of the carrier protein, the conjugation chemistry including the nature and the length of the linker if used or the saccharide/protein ratio³. Obviously the positions at which the saccharide is conjugated to the protein as well as the number of connectivity are relevant for immunogenicity. To date, these two parameters have hardly been studied because the preparation of the glycoconjugates remains largely empirical. Their synthesis usually relies on the use of amine or carboxylic acid functions of, respectively, lysine or aspartic/glutamic acid side-chain residues present on the carrier protein sequence. This leads not to a single but to a heterogeneous mixture of glycoconjugates. Playing on the reactivity, accessibility or distribution of the amino acid residues in the protein gives rise to more defined glycoconjugates which are more reliable to document the effect of saccharide/protein connectivity⁴. A step forward towards this goal can be achieved by applying the Protein Glycan Coupling Technology, a recombinant process which allows the production of controlled glycoconjugate vaccines in cell factories^{5,6}. However, the glycosylation exclusively takes place at an asparagine residue within D/EXNYS/T sequons (whereby X is any out of the 20 natural amino acids), not naturally present on the carrier proteins. Site selective mutagenesis and in particular incorporation of cysteines to exploit their highly and selective reactivity appears as an alternative^{7,8}. Production of carrier proteins incorporating UAAs in their sequence can offer even more flexibility for homogeneous glycoconjugate vaccine preparation. More than 100 UAAs have been developed and further incorporated into various proteins^{9,10}. Many of them contain bioorthogonal functions usually used to carry out Post Translational Modifications¹¹ or to graft biophysical probes¹² or drugs¹³ but which are ideal handles for further conjugation with carbohydrate antigens. Successful examples have been claimed by a Biotech¹⁴ using cell-free protein synthesis¹⁵ but preparation of glycoconjugate vaccines according to this strategy still waits for becoming popularized. Application of this strategy for the production of mutated carrier protein needs a modified translational machinery that includes a specific codon, a tRNA recognizing the codon and an aminoacyl-tRNA synthetase (aaRS) which specifically catalyzes the transfer of the UAA on the tRNA (Figure 1)¹⁶. Pyrrolysine amber stop codon suppression is one of the most widely used methods to incorporate UAA in particular the propargyl-lysine (PrK)¹⁷. The later can in turn react with azido-functionalized carbohydrate haptens to provide fully defined, homogeneous glycoconjugates. In the present manuscript we describe how to synthesize the propargyl-L-lysine, an UAA carrying an alkyne handle, how to incorporate it into a target protein during its translation in a bacteria and finally how to perform conjugation between the modified protein and a hapten carrying an azide function using click chemistry.

PROTOCOL:

1. Synthesis of the UAA: propargyl-lysine (PrK)

1.1 To synthesize N^α -Boc-propargyl-lysine¹⁸

- 1.1.1 Dissolve 500 mg of Boc-L-Lys-OH (2.03 mmol) in a mixture of aqueous 1 M NaOH (5 mL) and THF (5 mL) in a flask and fit the flask with a silicon septum.
- 1.1.2 Cool the flask in an ice-bath and then add dropwise (over a 2-3 min period) using a microsyringe, 158 μ L of propargyl chloroformate (1.62 mmol) under stirring.
- 1.1.3 Warm the reaction mixture at room temperature and continue the stirring for 10 h.
- 1.1.4 Cool down solutions of 50 mL of diethyl ether, 50 mL of aqueous 1 M hydrochloric acid and 60 mL of ethyl acetate.
- 1.1.5 Cool the crude reaction mixture in an ice-bath and pour the mixture in a separating funnel. Extract the mixture with 50 mL of diethyl ether. Discard the organic layer.
- 1.1.6 Add cautiously aqueous 1 M hydrochloric acid to the aqueous phase in the separating funnel. Then extract the aqueous layer twice using 30 mL of ethyl acetate. Verify the presence of N^α -Boc-propargyl-lysine in the organic phase by TLC using CH_2Cl_2 -Methanol (9:1) as eluent.
- 1.1.7 Dry the combined organic layers over MgSO_4 , filter-off the solid phase and concentrate the filtrate under reduced pressure on a rotary evaporator.
- 1.1.8 Dissolve a sample of the crude oily N^α -Boc-propargyl-lysine in deuterated chloroform (CDCl_3) and control its identity by ^1H NMR.

CAUTION: Extraction may result in a buildup of pressure. Release any pressure buildup frequently.

1.2 To synthesize the Unnatural Amino Acid propargyl-L-lysine (PrK)

- 1.2.1 Introduce N^α -Boc-propargyl-lysine in a round bottom flask equipped with a septum.
- 1.2.2 Add 4 mL of anhydrous dichloromethane (CH_2Cl_2) into the flask under argon to dissolve the N^α -Boc-propargyl-lysine.
- 1.2.3 Add dropwise, using a syringe, 4 mL of trifluoroacetic acid (TFA) under stirring.
- 1.2.4 Stir the reaction mixture for 1h at RT. Monitor the reaction by Thin Layer Chromatography (TLC) using CH_2Cl_2 -Methanol (9:1) as eluent.
- 1.2.5 Concentrate the reaction mixture under reduced pressure.
- 1.2.6 Add diethyl ether to the crude residue and incubate it at 4°C for 1h to precipitate the PrK. When working on higher scale, if the PrK is not completely precipitated, triturate

to precipitate it and extend incubation time if needed.

- 1.2.7 Filter the PrK in the form of a white solid on a fritted-glass.
- 1.2.8 Dissolve an aliquot of the PrK in D₂O then carry out NMR analyses to control its identity and purity.
- 1.2.9 For further use, the Unnatural Amino Acid PrK is dissolved in distilled water at a final concentration of 100 mM and stored at -20 °C as 1 mL aliquots.

2. Production of the recombinant protein modified by PrK

2.1 Plasmid preparation

- 2.1.1 Construct an expression plasmid (pET24d-mPsaAK32TAG-ENLYFQ-HHHHHH) that contains the target mature Pneumococcal surface adhesin A (mPsaA) gene (pET24d-mPsaA-WT) followed by a Tobacco Etch Virus (TEV) Protease sequence by cloning the insert between the *Bam*HI and *Xho*I restriction sites of the pET24d plasmid. This will introduce a His₆ tag at the C-terminus of the protein. Replace the codon of Lysine-32 with the amber codon (TAG), using conventional site-directed mutagenesis technique.
- 2.1.2 Construct a second expression plasmid (pEVOL-MmPylRS) containing two copies of the gene coding for the pyrrolysyl-tRNA synthetase from *Methanosarcina mazei* (MmPylRS) and the gene coding for the corresponding tRNA^{Pyr} as previously described¹⁹ Use this specially designed plasmid vector, pEVOL, for efficient incorporation of UAAs.

NOTE: The detailed plasmids information is described in Supplementary MATERIALS.

2.2 To co-transform plasmids into the expression strain

- 2.2.1 Thaw a 100 µL aliquot of chemically competent *Escherichia coli* BL21(DE3) on ice during 5 min.
- 2.2.2 Add 1 µL of each plasmid (50-100 ng of each) into the cells and incubate 30 min in ice.
- 2.2.3 Transfer the 1.5mL microtube with thawed competent cells in an incubator at 42 °C during 45 seconds, then put it back on ice for 2 min.
- 2.2.4 Add 900 µL of LB medium and incubate under shaking for 1h at 37°C to allow antibiotic expression, then plate the bacteria onto LB agar with 25 µg/mL of kanamycin and 30 µg/mL of chloramphenicol. Allow bacteria growth overnight at 37 °C.

2.3 To express proteins modified with PrK

- 2.3.1 Inoculate a single co-transformed colony in 5 mL of LB medium with antibiotics (25

- $\mu\text{g/mL}$ of kanamycin and $30 \mu\text{g/mL}$ of chloramphenicol). Incubate overnight at $37 \text{ }^\circ\text{C}$ with shaking.
- 2.3.2 Dilute the primary culture (5 mL) into 500 mL of auto-induction medium containing antibiotics, 0.02 % of L-arabinose and 1 mM of the Unnatural Amino Acid PrK and incubate it at $37 \text{ }^\circ\text{C}$ for 24 h with shaking. Include a negative control by performing the culture without PrK in parallel and a positive control by performing the culture of a clone containing the wt protein.
- 2.3.3 Aliquot 5 mL out of the 500 mL culture medium and centrifuge 10 min at 5000g. Discard the supernatant and freeze the pellet at $-20 \text{ }^\circ\text{C}$. Harvest cells from the 495 mL left by centrifugation at 5000g during 10 min. Discard the supernatant and freeze the pellet at $-20 \text{ }^\circ\text{C}$.
- 2.4 Analyze crude cell extracts from the 5mL culture samples by SDS-PAGE and Western Blot analysis
- 2.4.1 Resuspend 5mL cell pellets into 250 μL of lysis buffer (50 mM $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$, 150mM NaCl, pH 8, 5mM imidazole, 0.2mM PMSF) and transfer it into a 1.5 mL microtube.
- 2.4.2 Lyze cells by freezing tubes in liquid nitrogen followed by thawing into a $42 \text{ }^\circ\text{C}$ bath and vortexing at high speed during 30 sec. Repeat this step 3 times.
- 2.4.3 Centrifuge samples at 17 000g during 10 min to eliminate cell debris.
- 2.4.4 Take 10 μL of the supernatant and add 5 μL of water and 5 μL of loading buffer (Bromophenol blue, SDS, β -mercaptoethanol). Heat the samples 5 min at $100 \text{ }^\circ\text{C}$ and carry out SDS-PAGE and Western Blot analyses.
- 2.5 Protein purification by gravity flow-bench affinity chromatography using Nickel-NTA beads
- 2.5.1 Resuspend the cell pellets (from 495mL culture) into 20 mL of lysis buffer (50 mM $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$, 150 mM NaCl, pH 8, 5mM imidazole, 0.2mM PMSF).
- 2.5.2 Add 5 μL of DNase I (1mg/mL) and 500 μL of lysozyme (50 mg/mL) into the suspension and allow lysis by incubating the suspension at $37 \text{ }^\circ\text{C}$ during 30 min.
- 2.5.3 Sonicate the cells during 5 min (cycles of 5 sec-5 sec, amplitude 50%) and then remove the cell debris by centrifugation at 20 000 g during 30 min followed by filtration on 0.45 μm filter.
- 2.5.4 Add Ni-NTA resin to the suspension (500 μL for 500 mL of cell culture) and mix gently at $4 \text{ }^\circ\text{C}$ for 1 h.
- 2.5.5 Pour the suspension into a polypropylene column and collect the unbound fraction.
- 2.5.6 Wash the resin with 10 mL of washing buffer containing 50 mM $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$, 150 mM NaCl, 10 mM imidazole. Wash the resin a second time with 5 mL of washing

buffer (50 mM Na₂HPO₄/NaH₂PO₄, 150 mM NaCl, 20 mM imidazole). Collect the wash fractions.

- 2.5.7 Elute the his-tagged protein with 1 mL of elution buffer (50 mM Na₂HPO₄/NaH₂PO₄, 150 mM NaCl, 300 mM imidazole). Repeat this step 4 times and collect all the elution fractions.
- 2.5.8 Analyze the crude lysate as well as the 7 purification fractions by SDS-PAGE on 12% acrylamide gel.
- 2.5.9 Combine the fractions containing pure His-tagged protein and dialyze it against 1 L of TEV protease buffer (50 mM Tris-HCl, 0.5 mM EDTA, pH 8) overnight by using a dialysis membrane (cut-off MW 6000-8000 Da). Measure the concentration of the protein at 280nm with a molar extinction coefficient of 37 360 cm⁻¹.M⁻¹ and a molecular weight of 34.14 KDa for mPsaA.

3. Removal of the histidine tag by TEV protease digestion

- 3.1 Collect the protein sample into a 50mL tube and add TEV buffer (50 mM Tris HCl, 0.5 mM EDTA, pH 8) up to 1 mL at a concentration of 2 mg/mL.

NOTE: the concentration may vary according to previous results. Protein concentrations that we have tested are in a typical 2-3 mg/mL range.

- 3.2 Add 100 µL of TEV protease (you need to add 1 µL containing 10 units of TEV protease for 20 µg of protein to digest).
- 3.3 Add 50 µL of 0.1 M dithiothreitol (DTT).
- 3.4 Complete with TEV buffer (50 mM Tris HCl, 0.5 mM EDTA, pH 8) up to 5 mL.
- 3.5 Incubate overnight at 4 °C with slow shaking.

NOTE: If digestion is not complete, add more TEV protease, incubate for longer time or at higher temperature up to 30 °C.

- 3.6 Dialyze the digested protein to remove EDTA at 4 °C overnight by using a dialysis membrane (cut-off 6000-8000 Da) against phosphate buffer (50 mM Na₂HPO₄/NaH₂PO₄, 150 mM NaCl, 5 mM imidazole).
- 3.7 To eliminate the TEV protease and the undigested protein, incubate the mix with Ni-NTA beads and mix gently for 1 h at 4 °C.
- 3.8 Pour the suspension into a polypropylene column. Collect the unbound fraction and wash the column with 5 mL of washing buffer (50 mM Na₂HPO₄/NaH₂PO₄, 150 mM NaCl, 10 mM imidazole)

NOTE: the protein of interest should be recovered in the unbound and washing fractions.

- 3.9 Elute the TEV protease and the undigested protein by adding 5 mL of elution buffer (50 mM Na₂HPO₄/NaH₂PO₄, 150 mM NaCl, 300 mM imidazole) on the column. Check the fractions for protein contents at 280 nm and by SDS-PAGE analysis.
- 3.10 Check the efficiency of the digestion by loading digested samples on an SDS PAGE with the undigested protein as a control.
- 3.11 Dialyze the digested protein against 1L of Click buffer (50 mM Na₂HPO₄/NaH₂PO₄, pH8) at 4 °C overnight with a dialysis membrane (cut-off 6000-8000 Da) to remove imidazole as well as to exchange the buffer, and measure the concentration of the protein at 280nm with molar extinction coefficient and molecular weight of mPsaA (37 360 cm⁻¹.M⁻¹ and MW 34.14 KDa).

4. Assessment of the Unnatural Amino Acid propargyl-lysine accessibility and functionality for click chemistry

To conjugate the mPsaA with 6-hexachloro-fluorescein-azide (The protocol is essentially the one described by Presolski *et al.* in 2011²⁰ for click chemistry).

- 4.1 Take 432.5 µL of PrK-mutated protein at a concentration of 57.8 µM into a 2 mL Microtube tube.

NOTE: A minimum concentration of 2 µM of alkyne is acceptable, if the protein concentration is lower, concentrate it with a centrifugal concentrator or favor the balance of the reaction by increasing azide/alkyne molar ratio.

- 4.2 Add 10 µL of 5 mM 6-hexachloro-fluorescein-azide then add a premix of 2.5 µL of CuSO₄ solution at 20 mM and 7.5µL of tris(benzyltriazolylmethyl)amine (THPTA) at 50 mM (stock solutions concentrations).
 - 4.2.1 Add 25 µL of aqueous 100 mM aminoguanidine hydrochloride.
 - 4.2.2 Add 25 µL of 20 mg/mL an extemporaneously prepared aqueous solution of sodium ascorbate.
 - 4.2.3 Close the tube, mix by inverting several times and incubate at room temperature for 2 h.
 - 4.2.4 Stop the reaction by adding 50 µL of 0.5 M EDTA.
 - 4.2.5 Take 15 µL of the reaction mixture and put it a microtube, add 5 µL of loading buffer (Bromophenol blue, SDS, β-mercaptoethanol), heat the mixture at 100 °C for 5 min, and then load it into a 12% acrylamide gel. After migration, visualize the fluorescent conjugate on the gel under UV light at 312nm.

5. Conjugation of mPsaA with an azido-functionalized carbohydrate antigen (Pn14TS-N₃) by click chemistry

5.1 Coupling

- 5.1.1 Take 432.5 μL of PrK-mutated protein at 57.8 μM into a 2 mL microtube tube.
- 5.1.2 Add 10 μL of 5 mM Pn14TS-N₃²¹ in water then add a premix of 2.5 μL of CuSO₄ solution at 20 mM and 7.5 μL of THPTA at 50 mM.

NOTE: synthesis of Pn14TS-N₃, a tetrasaccharide mimicking the *Streptococcus pneumoniae* serotype 14 capsular polysaccharide, has been described in reference 21. Theoretically, any carbohydrate antigen containing an azide function can be used.

- 5.1.3 Add 25 μL of 100 mM aminoguanidine hydrochloride.
- 5.1.4 Add 25 μL of 20 mg/mL extemporaneously prepared aqueous solution of sodium ascorbate.
- 5.1.5 Close the tube, mix by inverting several times and incubate at RT during 2 h.
- 5.1.6 Stop the reaction by adding 50 μL of 0.5 M EDTA.
- 5.1.7 Take 15 μL of samples and analyze by SDS-PAGE.

5.2 Gel filtration purification of the glycoconjugate

- 5.2.1. Purify the glycoconjugate by applying it to a steric exclusion agarose column (15 \times 600 bed dimensions, 3,000-70,000 fractionation range), equilibrated with 100 mM PBS buffer, pH 7.3 at a 0.8 mL/min flow with detection at 280 nm.
- 5.2.2. Collect the fractions containing the glycoconjugate.

NOTE: For prolonged storage, dialyze the glycoconjugate against 1 L of H₂O twice for 2 h and then overnight at 4 °C by using dialysis membrane (cut-off Mw 6000-8000 Da), then freeze-dry and store the glycoconjugate at -80 °C.

REPRESENTATIVE RESULTS:

In this project, a homogeneous glycoconjugate vaccine was prepared using the amber stop codon suppression strategy to introduce an UAA at a defined site (**Figure 1**). Pneumococcal surface adhesion A was selected as the carrier protein moiety. This protein is highly conserved and expressed by all strains of *Streptococcus pneumoniae*²². It is highly immunogenic and previously used as a carrier in pneumococcal vaccine formulations^{23,21} As a proof-of-concept, the UAA propargyl-lysine efficiently charged by the wild type pyrrolysyl-tRNA synthetase (PylRS) / tRNA pair of the archaea *Methanosarcina mazei* was investigated. The propargyl-lysine is commercially available but can be advantageously prepared from Boc-L-lysine in only two synthetic steps (**Figure 2**). An amber codon was generated at a desired position in a pET24d plasmid containing the *mPsaA* gene. This plasmid was co-transformed with a pEVOL plasmid (a kind gift from Edward Lemke (EMBL)¹⁹ containing orthogonal tools necessary to incorporate the propargyl-lysine, into competent *E. coli* BL21(DE3) strain. Positive co-transformed clones were selected using kanamycin 25 $\mu\text{g}/\text{mL}$ and chloramphenicol 30 $\mu\text{g}/\text{mL}$. The plasmid pEVOL contains originally not one but two copies of the gene coding for MmPylRS to incorporate the propargyl-lysine residue: the first copy is under the control of a

constitutive promoter while the expression of the other one is inducible in the presence of arabinose. However, we have noticed no dramatic decrease of propargyl-lysine incorporation if the MmPylRS gene under the control of constitutive promoter is suppressed.

The propargyl-lysine was introduced at position 32 in replacement of a lysine near the N-terminus of the PsaA. Any residue which side-chain is surface-exposed can virtually be exchanged in view of carrying out further conjugation. The mutated protein was produced in its mature form (mPsaA^{K32PrK}) with inclusion of a cleavable 6-histidine tag sequence at its C-terminus. The efficiency of the mPsaA^{K32PrK} production was checked by SDS-PAGE and Western Blot analysis using anti-Histidine tag antibody, when growth was performed in the presence or the absence of the UAA propargyl-lysine and in comparison with the production of the wild type mPsaA (**Figure 3**). Visualization revealed a protein band at an expected molecular weight (Lanes 4 Figure 3A & 3B). Presence of a full-length protein strongly indicates the successful incorporation of the PrK into mPsaA. The intensity is, however, lower than that observed for wild type mPsaA (Lanes 2 Figure 3A & 3B). Leakage – *i.e.* production of the full length protein without incorporation of the UAA – and premature release of the protein by the Release Factor RF1 during translation are two main drawbacks frequently encountered during this process. On one hand, no band at the expected molecular weight is visualized in the absence of propargyl-lysine meaning that no leakage occurred (Lanes 3, Figure 3A and 3B) and indirectly confirmed that the band observed on Lanes 4 corresponds to mPsaA^{K32PrK}. On the other hand, no band can be seen at low molecular weight that could correspond to the truncated form of mPsaA (Lane 4 on Figure 3A).. The mPsaA^{K32PrK} was then purified by affinity chromatography, with a typical yield of 8 mg/L (in comparison of 12-20 mg/L for the wild type protein) and the incorporation of the propargyl-lysine residue was finally confirmed by mass spectrometry (**Figure 4A**). The histidine tag was removed upon proteolytic cleavage using TEV protease (**Figure 4B**). The stability of the mPsaA^{K32PrK} thus obtained was assessed by circular dichroism which showed the structure of the protein was not disturbed by the mutation of the Lysine 32 into a propargyl-lysine (data not shown).

Having the mPsaA^{K32PrK}, the reactivity of the alkyne for click chemistry was assessed using an azido-functionalized fluorescein and further used to conjugate a synthetic oligosaccharide antigen β -2-azidoethyl D-Galp-(1→4)- β -D-Glcp-(1→6)-[β -D-Galp-(1→4)]- β -D-GlcpNAc (Pn14TS) (**Figure 5**). This tetrasaccharide is related to the *S. pneumoniae* type 14 capsular polysaccharide and has previously been conjugated to mPsaA using different conjugation chemistries^{24,21,8}. Experiments here were done in comparison with wild type mPsaA as a control. The histidine tag was first removed upon proteolytic cleavage using the TEV protease. The digested mPsaA^{K32PrK} and mPsaA WT were then conjugated to the fluorophore (**Figure 6A**) or Pn14TS (**Figure 6B**). The reaction was assessed by SDS-PAGE. The small increase in the molecular weight of the sample between lane 6 and 7 (**Figure 6B**) indicates a successful conjugation with the tetrasaccharide Pn14TS. Finally the glycoconjugate was purified by gel filtration and its identity confirmed by mass spectrometry (**Figure 6C**). The conjugation by click chemistry being quantitative the majority of the mPsaA^{K32PrK} was conjugated with the Pn14TS-N₃ as illustrated by the mass spectrometry results (**Figure 6C**).

Figures :

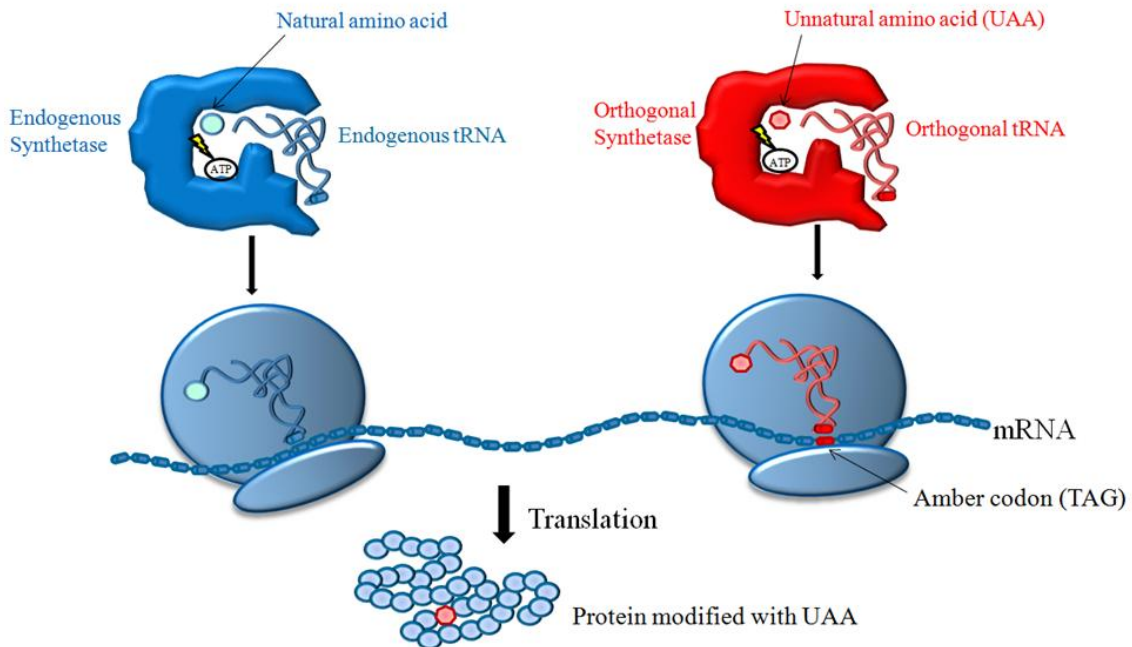


Figure 1 Incorporation of propargyl-lysine (PrK) into mPsaA during translation using an orthogonal pyrrolysyl-tRNA synthetase/tRNA pair and TAG codon reassignment.²⁵ During translation, endogenous synthetases catalyze the link between amino acids and corresponding tRNAs. Then, loaded tRNAs are used by the ribosomal machinery to generate the neo-synthesized polypeptide. According to the amber stop codon suppression strategy, an orthogonal aminoacyl-tRNA synthetase (aaRS) (herein a pyrrolysyl-tRNA synthetase from *M. mazei*), loads an UAA (herein PrK) on its cognate tRNA which designed anticodon can read the amber stop codon (TAG) on the mRNA. This specific recognition directs the incorporation of the UAA into the specific site on the target protein. Figure reproduced from Wang *et al.*²⁵

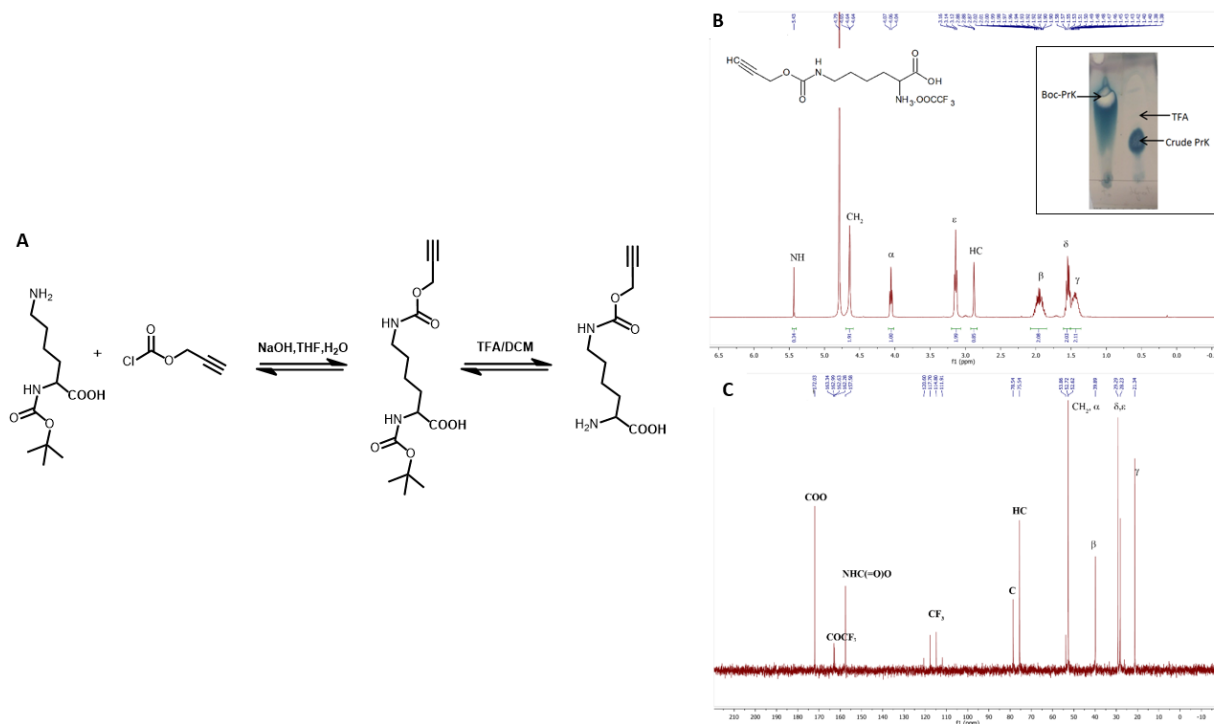


Figure 2 Propargyl-lysine synthesis

(A) Steps of propargyl-lysine synthesis. Insert: Monitoring of the deprotection of Boc-L-Lys(prop-2-ynyloxycarbonyl)-OH intermediate: thin-layer chromatography on 0.25 mm silica gel plates with fluorescent indicator (GF254) and visualised by charring with vanillin in sulfuric acid/ethanol (1.5:95 v/v); eluent: $\text{CH}_2\text{Cl}_2/\text{MeOH}$ (9:1), left lane: Boc-L-Lys(prop-2-ynyloxycarbonyl)-OH (R_f 0.90), right lane: crude propargyl-lysine, (R_f 0.38). 400 MHz ^1H (B) and ^{13}C NMR spectra (C) of propargyl-lysine recorded in D_2O .

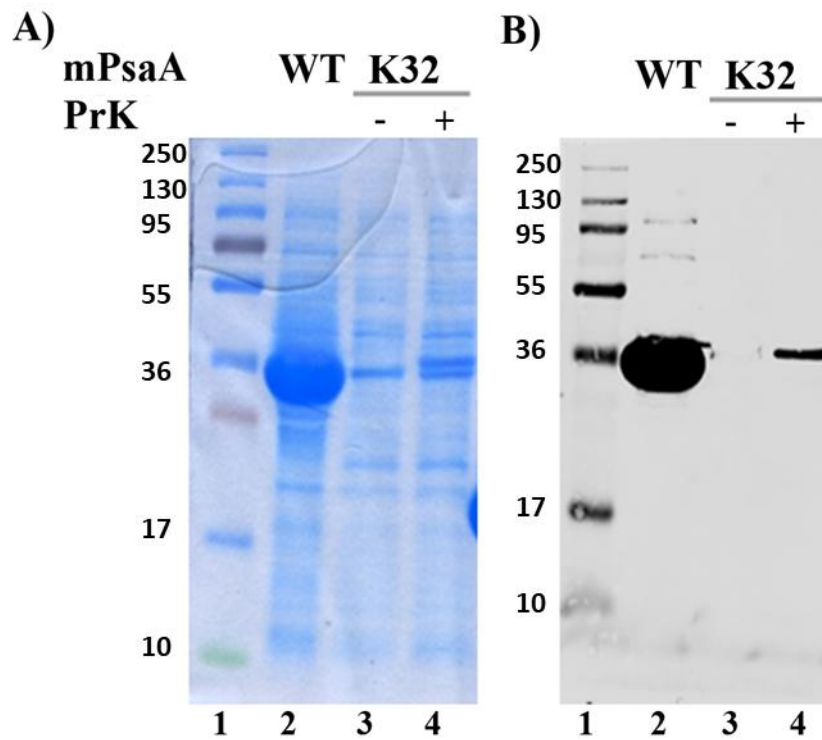


Figure 3 Analysis of crude cell samples

SDS-PAGE analysis (A) and Western blot analysis (B) on crude cell samples. Lane 1: unstained protein marker; Lane 2: crude cell extract of wild type mPsaA; Lane 3: crude cell extract of mPsaA^{K32TAG} grown in the absence of PrK; Lane 4: crude cell extraction of mPsaA^{K32TAG} grown in the presence of PrK. Conditions: 12% acrylamide gel, running at 100 V, 2 h. SDS-PAGE stained by Coomassie blue; WesternBlot revealed using anti-histidine tag antibody and secondary antibody coupled with AlexaFluor680.

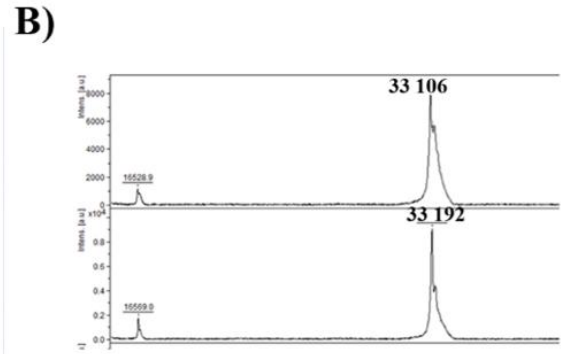
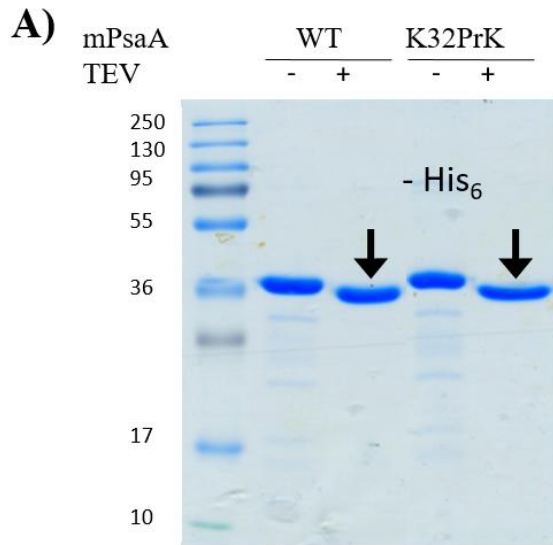


Figure 4 : Histidine tag removal and mass spectrometry analysis

A) SDS-PAGE analysis. Lane 1: unstained protein marker; Lane 2: undigested WT mPsaA; Lane 3: TEV digested WT mPsaA; Lane 4: undigested K32PrK mPsaA; Lane 5: TEV digested K32PrK mPsaA; Conditions: 12% acrylamide gel running at 100 V for 2 h, and stained by Coomassie blue; MALDI-TOF-MS spectra of (top) mPsaA WT, theoretical MW 33 103 Da, found 33 105 Da and (bottom) mPsaA K32PrK, theoretical 33 184 Da, found 33 192 Da. The found masses are within the expected margin error.

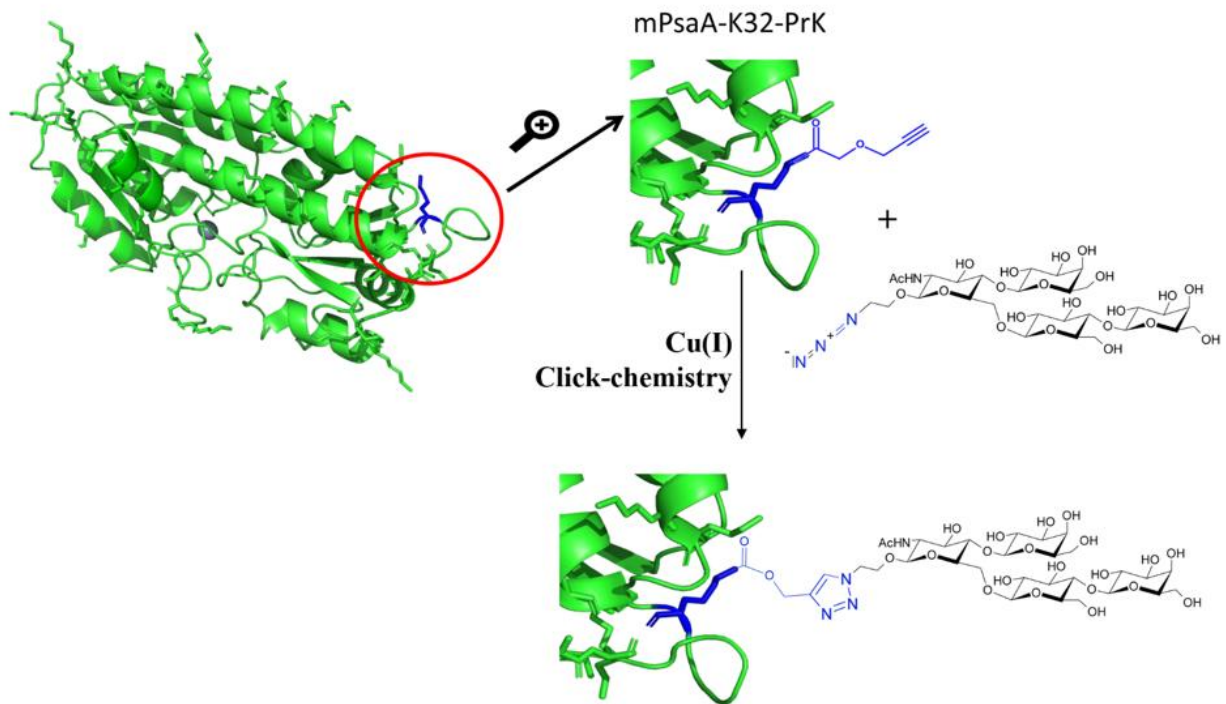


Figure 5 Schematic representation of the conjugation strategy by click chemistry

A single tetrasaccharide bearing an azide is specifically coupled to its complementary biorthogonal alkyne group on mPsaA K32PrK (mPsaA representation based on the 1PSZ PDB file, with a resolution of 2.0 \AA^{26}).

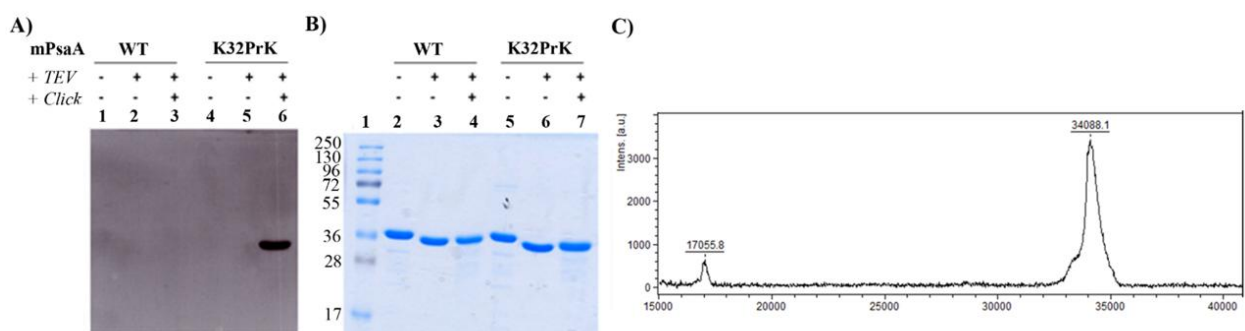


Figure 6 Histidine-tag digestion and conjugation of mPsaA with fluorescein- N_3 (A) and with Pn14TS (B)

SDS-PAGE (A) Lanes 1-3: WT mPsaA; Lane 4-6: mPsaA^{K32PrK}; (B) Lane 1: unstained protein marker; Lane 2-4: WT mPsaA; Lane 5-7: mPsaA^{K32PrK}. 2 μg protein sample/lane, 12% acrylamide, 100V, 2 h; (C) MALDI-TOF-MS spectra of the Pn14TS-mPsaA^{K32PrK} theoretical MW 34 091 Da, found 34 088 Da.

DISCUSSION:

Site-directed mutagenesis is a straightforward strategy to incorporate specific amino acids at a defined position of a protein which remains barely used with the aim of preparing glycoconjugate vaccines^{7,8,14}. “Classical” mutagenesis based on the 20 natural amino acids approach is highly efficient since no modification of the translation machinery is required. Cysteine mutations are usually targeted to further explore the unique thiol reactivity either directly or in two steps *e.g.* after its modification into a dehydroalanine intermediate, a strategy called post-translational mutagenesis^{27,28}. Genetic code expansion is perhaps even more attractive and flexible since it allows the direct incorporation of a wide range of UAAs with diverse functionalities^{9,10}. While several UAAs can be incorporated simultaneously within a protein²⁹, the number of mutations is usually more limited. We herein applied the related amber stop codon strategy to introduce a single propargyl-lysine in a carrier protein. The incorporation can take place at any position provided the side-chain of the initial amino acid was surface-exposed, a criterion easily determined from X-ray crystallographic structures or *in silico* modeling. Moreover it is not limited to propargyl-lysine but can be extended to any UAA functionalized with a biorthogonal function which will later serve as an anchor to graft the incoming carbohydrate antigen and for which an orthogonal aaRS / tRNA pair exists.

One of the drawbacks of the strategy is the possible production of truncated protein, resulting from the release of the peptidyl side-chain when reading the amber stop codon, as a side-product. Even if we did not observe any truncated form here (probably degraded by the bacteria because of its very small size) a histidine tag has been added at C-terminus of the protein to facilitate the purification of the expected full-length mutated protein from impurities and noticeably from the truncated protein (which by essence does not express the histidine tag sequence). This can become essential if the UAA incorporation is carried out near the protein C-terminus since purification cannot be attempted using alternative chromatography techniques like gel filtration.

For most applications removal of the histidine tag is not mandatory. However, it may be useful regarding the design of glycoconjugate vaccine as part of the immune system may be diverted against the tag sequence. For this proof-of-concept we inserted an amino acid length sequence specifically cleaved by the TEV protease which leaves five extra amino acids on the carrier protein after digestion.

The conjugation step between the alkyne of the propargyl-lysine and a representative synthetic oligosaccharide Pn14TS related to a pneumococcal capsule and bearing a complementary azide was carried out according to a click chemistry protocol reported by Presolski *et al.*²⁰ If necessary, completion of the reaction can easily be reached by increasing the reaction time or by modifying the ratio between the alkyne, azide and copper reactants and reagents. Copper salts are eliminated by treatment with excess EDTA followed by a short purification by steric exclusion chromatography.

The glycoconjugate obtained with the technique described in the present work can then be used to immunize mice. Having such fully-defined and easily modulated glycoconjugate in hands provides invaluable tools to evaluate the impact of the hapten/protein carrier connectivity on the immune response⁸. Since increasing the hapten/protein ratio is often correlated with enhanced anti-hapten humoral response when using short haptens,³⁰ one might be interested in testing conjugates with multiple haptens. The incorporation of multiple UAAs however needs some adjustments of the

protocol as the incorporation of an UAA in the protein tends to decrease the yield of protein production due to the RF1 activity.

In definitive, this method is a powerful tool to gain access to homogeneous glycoconjugate vaccines facilitating their physico-chemical characterization and further carbohydrate antigen/carrier connectivity-immunogenicity relationship studies.

ACKNOWLEDGMENTS:

E.C. gratefully acknowledges the financial support from La Région Pays de la Loire (Pari Scientifique Program “BioSynProt”), in particular a doctoral fellowship to T.V. We also acknowledge Dr Robert B. Quast (INRA UMR0792, CNRS UMR5504, LISBP, Toulouse, France) for his precious technical advices.

DISCLOSURES:

The authors have nothing to disclose.

REFERENCES

1. Rappuoli, R. Glycoconjugate vaccines: Principles and mechanisms. *Science Translational Medicine*. 10 (456), doi: 10.1126/scitranslmed.aat4615 (2018).
2. Verez-Bencomo, V. et al. A synthetic conjugate polysaccharide vaccine against *Haemophilus influenzae* type b. *Science (New York, N.Y.)*. 305 (5683), 522–525, doi: 10.1126/science.1095209 (2004).
3. Berti, F., Adamo, R. Antimicrobial glycoconjugate vaccines: an overview of classic and modern approaches for protein modification. *Chemical Society Reviews*. 47 (24), 9015–9025, doi: 10.1039/c8cs00495a (2018).
4. Stefanetti, G. et al. Sugar-Protein Connectivity Impacts on the Immunogenicity of Site-Selective *Salmonella* O-Antigen Glycoconjugate Vaccines. *Angewandte Chemie (International Ed. in English)*. 54 (45), 13198–13203, doi: 10.1002/anie.201506112 (2015).
5. Kay, E., Cuccui, J., Wren, B.W. Recent advances in the production of recombinant glycoconjugate vaccines. *NPJ vaccines*. 4, 16, doi: 10.1038/s41541-019-0110-z (2019).
6. Ma, Z., Zhang, H., Wang, P.G., Liu, X.-W., Chen, M. Peptide adjacent to glycosylation sites impacts immunogenicity of glycoconjugate vaccine. *Oncotarget*. 9 (1), 75–82, doi: 10.18632/oncotarget.19944 (2018).
7. Grayson, E.J., Bernardes, G.J.L., Chalker, J.M., Boutureira, O., Koeppel, J.R., Davis, B.G. A coordinated synthesis and conjugation strategy for the preparation of homogeneous glycoconjugate vaccine candidates. *Angewandte Chemie (International Ed. in English)*. 50 (18), 4127–4132, doi: 10.1002/anie.201006327 (2011).
8. Pillot, A. et al. Site-Specific Conjugation for Fully Controlled Glycoconjugate Vaccine Preparation. *Frontiers in Chemistry*.

9. Neumann-Staubitz, P., Neumann, H. The use of unnatural amino acids to study and engineer protein function. *Current Opinion in Structural Biology*. 38, 119–128, doi: 10.1016/j.sbi.2016.06.006 (2016).
10. Dumas, A., Lercher, L., Spicer, C.D., Davis, B.G. Designing logical codon reassignment - Expanding the chemistry in biology. *Chemical Science*. 6 (1), 50–69, doi: 10.1039/c4sc01534g (2015).
11. Chen, H., Venkat, S., McGuire, P., Gan, Q., Fan, C. Recent Development of Genetic Code Expansion for Posttranslational Modification Studies. *Molecules (Basel, Switzerland)*. 23 (7), doi: 10.3390/molecules23071662 (2018).
12. Adumeau, P., Sharma, S.K., Brent, C., Zeglis, B.M. Site-Specifically Labeled Immunoconjugates for Molecular Imaging--Part 2: Peptide Tags and Unnatural Amino Acids. *Molecular imaging and biology: MIB: the official publication of the Academy of Molecular Imaging*. 18 (2), 153–165, doi: 10.1007/s11307-015-0920-y (2016).
13. Kularatne, S.A. et al. A CXCR4-targeted site-specific antibody-drug conjugate. *Angewandte Chemie (International Ed. in English)*. 53 (44), 11863–11867, doi: 10.1002/anie.201408103 (2014).
14. US20180333484 Polypeptide-Antigen Conjugates with Non-Natural Amino Acids. at <[https://patentscope.wipo.int/search/en/detail.jsf?docId=US233548973&recNum=65&docAn=15859251&queryString=\(GBS\)&maxRec=11858](https://patentscope.wipo.int/search/en/detail.jsf?docId=US233548973&recNum=65&docAn=15859251&queryString=(GBS)&maxRec=11858)>.
15. Quast, R.B., Mrusek, D., Hoffmeister, C., Sonnabend, A., Kubick, S. Cotranslational incorporation of non-standard amino acids using cell-free protein synthesis. *FEBS letters*. 589 (15), 1703–1712, doi: 10.1016/j.febslet.2015.04.041 (2015).
16. Wang, L. Engineering the Genetic Code in Cells and Animals: Biological Considerations and Impacts. *Accounts of Chemical Research*. 50 (11), 2767–2775, doi: 10.1021/acs.accounts.7b00376 (2017).
17. Brabham, R., Fascione, M.A. Pyrrolysine Amber Stop-Codon Suppression: Development and Applications. *Chembiochem: A European Journal of Chemical Biology*. 18 (20), 1973–1983, doi: 10.1002/cbic.201700148 (2017).
18. Genetic Encoding of a Non-Canonical Amino Acid for the Generation of Antibody-Drug Conjugates Through a Fast Bioorthogonal Reaction. - PubMed - NCBI. at <<https://www.ncbi.nlm.nih.gov/insb.bib.cnrs.fr/pubmed/?term=Genetic+Encoding+of+a+Non-Canonical+Amino+Acid+for+the+Generation+of+Antibody-Drug+Conjugates+Through+a+Fast+Bioorthogonal+Reaction>>.
19. Young, T.S., Ahmad, I., Yin, J.A., Schultz, P.G. An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *Journal of Molecular Biology*. 395 (2), 361–374, doi: 10.1016/j.jmb.2009.10.030 (2010).
20. Presolski, S.I., Hong, V.P., Finn, M.G. Copper-Catalyzed Azide-Alkyne Click Chemistry for Bioconjugation. *Current Protocols in Chemical Biology*. 3 (4), 153–162, doi: 10.1002/9780470559277.ch110148 (2011).
21. Prasanna, M. et al. Semisynthetic glycoconjugate based on dual role protein/PsaA as a pneumococcal vaccine. *European Journal of Pharmaceutical Sciences: Official Journal of the European Federation for Pharmaceutical Sciences*. 129, 31–41, doi: 10.1016/j.ejps.2018.12.013 (2019).

22. Morrison, K.E. et al. Confirmation of *psaA* in all 90 serotypes of *Streptococcus pneumoniae* by PCR and potential of this assay for identification and diagnosis. *Journal of Clinical Microbiology*. 38 (1), 434–437 (2000).
23. Lin, H., Lin, Z., Meng, C., Huang, J., Guo, Y. Preparation and immunogenicity of capsular polysaccharide-surface adhesin A (PsaA) conjugate of *Streptococcus pneumoniae*. *Immunobiology*. 215 (7), 545–550, doi: 10.1016/j.imbio.2009.08.008 (2010).
24. Safari, D. et al. Identification of the smallest structure capable of evoking opsonophagocytic antibodies against *Streptococcus pneumoniae* type 14. *Infection and Immunity*. 76 (10), 4615–4623, doi: 10.1128/IAI.00472-08 (2008).
25. Wang, Q., Parrish, A.R., Wang, L. Expanding the genetic code for biological studies. *Chemistry & Biology*. 16 (3), 323–336, doi: 10.1016/j.chembiol.2009.03.001 (2009).
26. Lawrence, M.C., Pilling, P.A., Epa, V.C., Berry, A.M., Ogunniyi, A.D., Paton, J.C. The crystal structure of pneumococcal surface antigen PsaA reveals a metal-binding site and a novel structure for a putative ABC-type binding protein. *Structure (London, England: 1993)*. 6 (12), 1553–1561 (1998).
27. Wright, T.H., Davis, B.G. Post-translational mutagenesis for installation of natural and unnatural amino acid side chains into recombinant proteins. *Nature Protocols*. 12 (10), 2243–2250, doi: 10.1038/nprot.2017.087 (2017).
28. Dadová, J., Galan, S.R., Davis, B.G. Synthesis of modified proteins via functionalization of dehydroalanine. *Current Opinion in Chemical Biology*. 46, 71–81, doi: 10.1016/j.cbpa.2018.05.022 (2018).
29. Worst, E.G. et al. Residue-specific Incorporation of Noncanonical Amino Acids into Model Proteins Using an *Escherichia coli* Cell-free Transcription-translation System. *Journal of Visualized Experiments: JoVE*. (114), doi: 10.3791/54273 (2016).
30. Carboni, F. et al. GBS type III oligosaccharides containing a minimal protective epitope can be turned into effective vaccines by multivalent presentation. *The Journal of Infectious Diseases*. doi: 10.1093/infdis/jiz551 (2019).

Titre : Développement de méthodes de biologie synthétique appliquées à la génération de vaccins glycoconjugués contrôlés et à la modification des capacités de reconnaissance de protéines.

Mots clés : Incorporation d'acides aminés non naturels, Vaccins glycoconjugués, Lectines

Résumé :

Au cours de cette thèse, nous avons mis en place une technologie d'incorporation d'acides aminés non naturels de manière site-spécifique au sein du laboratoire et l'avons appliqué à deux problématiques.

Les vaccins glycoconjugués actuels ne couvrent que 13 des 97 sérotypes de *Streptococcus pneumoniae*. L'utilisation d'une protéine porteuse immunogène permet d'élargir la couverture sérotypique en couplant l'utilisation d'antigènes polysaccharidiques avec des antigènes protéiques. Cependant, les méthodes de conjugaison actuelles ne permettent pas l'étude de l'association optimale entre ces deux antigènes. Dans une première partie, nous avons généré différents vaccins glycoconjugués homogènes en incorporant la Propargyl-Lysine dans la protéine porteuse du pneumocoque (mPsaA) afin de maîtriser la position ainsi que le nombre des haptens.

Les lectines reconnaissent leur ligand saccharidique avec une très bonne spécificité mais avec une affinité généralement faible. Dans une seconde partie de la thèse, nous nous sommes intéressés à l'incorporation d'acides aminés non naturels permettant de former des liaisons covalentes avec les sucres et ainsi augmenter les forces d'interaction entre des lectines et leurs ligands. Nous avons tenté d'incorporer un premier acide aminé non naturel original sans succès. Mais deux autres ont pu être incorporés, dont l'un d'eux au sein de 2 cibles protéiques, en quantité suffisante pour réaliser des tests d'affinité.

Cette thèse a donc permis la mise en place d'un nouvel axe technologique au sein du laboratoire et d'en explorer des applications potentielles.

Title : Development of synthetic biology techniques applied to generate defined glycoconjugate vaccines and to broaden the recognition properties of proteins

Keywords : Unnatural amino acid incorporation, glycoconjugate vaccines, Lectin engineering

Abstract :

In this thesis, we set up the technology of non natural amino acids site-specific incorporation in the laboratory and we applied it to two projects.

Current glycoconjugate vaccines cover 13 out of 97 serotypes from *Streptococcus pneumoniae*. Using an immunogenic carrier protein could allow to boarden this efficiency by coupling carbohydrate antigens with protein antigens. However, currently used conjugation methods do not allow to study the optimal association between these two antigens. In a first part of the thesis, we generated different homogenous glycoconjugate vaccines using Propargyl-Lysine incorporation into a pneumococcal carrier protein (mPsaA) to master hapten positions on the protein as well

as their number.

Lectins recognize their carbohydrate ligand with a high specificity but usually with a low affinity. In a second part of the thesis, we focused on the incorporation of unnatural amino acids which can form covalent link with carbohydrates to increase the interaction forces between the lectins and their ligands. We first tried to incorporate an original amino acid without success. However, two others were incorporated, with one of them in two target proteins, with sufficient yield to perform affinity measurements.

To conclude, during this thesis we have implemented a new technology in the laboratory which allows exploring its potential applications.