

# Table des matières

REMERCIEMENTS .....	1
TABLE DES MATIERES .....	4
LISTE DES FIGURES .....	7
LISTE DES TABLEAUX .....	7
INTRODUCTION GENERALE.....	10
CONTEXTE SCIENTIFIQUE.....	14
<b>Chapitre 1 : L'utilisation des anticorps dans les maladies infectieuses.....</b>	<b>14</b>
I. Historique.....	14
II. Les différents mécanismes de protection .....	16
A. Neutralisation.....	16
1. Neutralisation des toxines.....	16
2. Neutralisation de l'entrée des pathogènes et de leur réplication.....	16
3. Neutralisation des facteurs de virulence microbiens.....	17
B. Recrutement des cellules effectrices du système immunitaire .....	17
1. Activation du complément .....	17
2. Antibody dependent cellular toxicity (ADCC) .....	18
3. Antibody-dependent cellular phagocytosis (ADCP) .....	18
III. Les anticorps monoclonaux sur le marché pour le traitement des maladies infectieuses.....	18
<b>Chapitre 2 : Infections par les Shigelles et les Salmonelles : Epidémiologie, mécanismes de pathogénicité et solutions thérapeutiques.....</b>	<b>20</b>
A. Les bactéries entéropathogènes du genre <i>Salmonella</i> et <i>Shigella</i> .....	20
1. Données épidémiologiques (généralités) .....	20
2. Espèces, sous-espèces et nomenclatures.....	21
B. Mécanismes de virulence des Shigelles et des Salmonelles .....	21
1. Les gènes impliqués dans la pathogénèse.....	21
2. Le système de sécrétion de type 3 .....	21
i. Structure générale.....	22
ii. Assemblage de la coiffe de l'aiguille d'injection.....	23
C. La vaccination.....	24
D. L'immunothérapie passive : les anticorps dirigés contre le SST3 .....	24
<b>Chapitre 3 : L'immunogénicité des anticorps thérapeutiques .....</b>	<b>26</b>
I. Les anticorps thérapeutiques : structure et modes de production.....	26
A. Structure et fonction des IgG.....	26
1. Généralités.....	26
2. Structure des IgG .....	26
i. Les domaines variables.....	27
3. Propriétés particulières des IgG .....	28
B. La production des anticorps recombinants.....	28
II. Mécanisme : La réponse cellulaire adaptative et la génération d'anti-drug antibodies (ADA).....	30
A. Mécanisme général aboutissant à la sécrétion d'ADA .....	30
B. Les cellules dendritiques .....	31
1. Internalisation .....	31

2.	L'apprêtement de l'antigène .....	32
3.	Présentation de l'antigène .....	33
i.	Les molécules HLA de classe II : structure et génétique.....	33
ii.	Les molécules HLA de classe II : interaction HLA-II/peptide.....	35
iii.	Apprêtement des peptides sur les molécules HLA-II.....	36
C.	Les lymphocytes T et la sélection thymique .....	37
1.	Le TCR .....	37
2.	La sélection thymique .....	38
D.	Les lymphocytes B et la différenciation en plasmocytes.....	39
1.	Génération des B-cell Receptors (BCR).....	39
2.	Activation des lymphocytes B .....	41
III.	Détection des ADA et données sur les anticorps thérapeutiques .....	42
A.	Les tests de détection des ADA .....	42
B.	Résumé des données sur les anticorps thérapeutiques .....	43
IV.	Conséquences et facteurs .....	46
A.	Conséquences sur le traitement et effets secondaires .....	46
1.	Altération de la pharmacocinétique et pharmacodynamique .....	46
2.	Neutralisation du traitement .....	46
3.	Risques pour le patient.....	46
B.	Les facteurs de l'immunogénicité.....	47
1.	Facteurs liés au produit.....	47
2.	Facteurs liés au traitement .....	48
3.	Facteurs liés au patient.....	48
V.	Prédiction de l'immunogénicité .....	49
A.	<i>La prédiction in silico</i> de la liaison aux molécules HLA-II.....	49
1.	Les algorithmes de prédiction de la liaison aux molécules HLA-II.....	49
i.	Les bases de données sur la liaison aux molécules HLA-II.....	49
ii.	Les algorithmes de prédiction principaux .....	50
iii.	Le choix des sets de molécules HLA pour la prédiction .....	53
2.	Les paramètres de sortie de la prédiction .....	54
B.	<i>In vitro</i> .....	56
1.	Tests biochimiques.....	56
i.	Le test de liaison (ELISA).....	56
ii.	Le MAPPS.....	56
2.	Test de stimulation des lymphocytes T.....	56
VI.	Dé-immunisation de protéines thérapeutiques .....	58
A.	Humanisation .....	58
1.	Historique.....	58
2.	CDR grafting .....	59
3.	Resurfacing.....	61
4.	Human string content (HSC) optimization .....	61
B.	Dé-immunisation .....	61
1.	Suppression des épitopes T.....	62
i.	Les exemples à bas débit (alanine scanning).....	62
ii.	Les exemples à haut débit (criblage in silico ou in vitro).....	62
2.	<i>B-cell epitope removal</i> .....	64

#### Chapitre 4 : Les méthodes de *display* pour l'ingénierie et l'étude des interactions antigène-anticorps

		66
I.	Les technologies de <i>display</i> .....	67
A.	Phage <i>Display</i> .....	67
B.	Autres méthodes de <i>display</i> .....	69
1.	Les systèmes cellulaires : <i>mammalian</i> , <i>yeast</i> et <i>bacterial display</i> .....	69
2.	Les systèmes acellulaires : mRNA, ribosome et DNA <i>display</i> .....	70
II.	Le Yeast Surface Display (YSD) et ses applications .....	70
A.	Historique et principe.....	70
B.	Application à l'affichage d'anticorps .....	72

1.	Les plasmides d'expression .....	72
2.	Les formats d'anticorps .....	73
3.	Génération et criblage des banques .....	73
C.	Applications du Yeast Surface Display à l'ingénierie des anticorps .....	76
1.	La mesure d'affinité à l'équilibre .....	76
2.	Le criblage de banques naïves et immunes .....	77
3.	La maturation d'affinité .....	77
4.	Ingénierie des propriétés autres que l'affinité .....	79
III.	Les apports du séquençage à haut débit pour l'ingénierie des anticorps .....	81
A.	Principe du séquençage à haut débit, le cas d'Illumina .....	81
B.	Apport du séquençage à haut débit pour l'ingénierie des anticorps .....	81
C.	Deep Mutational Scanning .....	82
1.	Principe .....	82
2.	Exemples d'applications .....	84
i.	Epitope mapping .....	84
ii.	Maturation d'affinité .....	84
iii.	Ingénierie de la sélectivité .....	84
<b>RESULTATS .....</b>		<b>87</b>
<b>Partie 1 : Détermination de l'épitope d'un anticorps protecteur anti-Salmonelle/Shigelle .....</b>		<b>88</b>
I.	Introduction de l'article .....	88
II.	Résultats additionnels et conclusion .....	121
<b>Partie 2 : Etude du format d'expression des anticorps à la surface de la levure .....</b>		<b>123</b>
I.	Introduction de l'article .....	123
II.	Résultats additionnels et conclusion .....	138
<b>Partie 3 : Ingénierie de l'anticorps 318 : humanisation et dé-immunisation .....</b>		<b>140</b>
I.	Mise en place de Dezimulab : une plateforme de dé-immunisation <i>in silico</i> .....	140
II.	Evaluation <i>in silico</i> de l'immunogénicité des parties variables de l'anticorps IpaD-318 .....	142
III.	Mise en place du DMS et application à l'anticorps 318 .....	143
IV.	Humanisation de l'anticorps 318 .....	151
V.	Dé-immunisation .....	153
VI.	Evaluation <i>in vitro</i> préliminaire de l'immunogénicité .....	164
Matériel et méthodes .....		167
<b>CONCLUSION GÉNÉRALE .....</b>		<b>175</b>
<b>LISTE DES ABREVIATIONS .....</b>		<b>182</b>
<b>BIBLIOGRAPHIE .....</b>		<b>184</b>

## Liste des Figures

Figure 1 : Vue structurale globale du système de sécrétion de type III de <i>Salmonella</i> et <i>Shigella</i> .....	21
Figure 2 : Représentation structurale d'un tétramère d'IpaD.....	22
Figure 3 : Représentation hypothétique du fonctionnement du SST3 de <i>Salmonella</i> et <i>Shigella</i> .....	24
Figure 4 : Structure générale d'un anticorps au format IgG1.....	26
Figure 5 : La coopération cellulaire dans la réponse humorale dirigée contre les protéines thérapeutiques.....	29
Figure 6 : Représentation structurale d'un CMH-II (HLA-DR1) en complexe avec un peptide.....	33
Figure 7 : Représentation schématique des 5 haplotypes HLA-DR humains.....	34
Figure 8 : Vue structurale en trois dimensions des poches de liaison d'une molécule HLA-DR1 en complexe avec un peptide.....	35
Figure 9 : Génération des complexes peptide/CMH-II dans les cellules dendritiques.....	36
Figure 10 : Modèle de sélection des thymocytes par affinité.....	37
Figure 11 : Organisation génétique et synthèse des anticorps chez l'homme.....	39
Figure 12 : Activation des lymphocytes B par les lymphocytes T CD4.....	41
Figure 13 : Base de données des scores de liaison pour chaque poche utilisée par l'algorithme de prédiction TEPITOPE.....	49
Figure 14 : Arbre de distance regroupant les allèles de CMH-II selon leurs motifs de liaison.....	51
Figure 15 : Les différents degrés d'humanisation des anticorps.....	56
Figure 16 : Les différentes méthodes de dé-immunisation décrites dans la littérature.....	63
Figure 17 : Ingénierie du phage M13 pour réaliser des expériences de <i>phage display</i> .....	65
Figure 18 : Le cycle du <i>phage display</i> .....	66
Figure 19 : Principe de la sélection d'anticorps en <i>Yeast Surface Display</i> en utilisant le système Aga1p/Aga2p.....	69
Figure 20 : Carte simplifiée du plasmide pYD1.....	71
Figure 21 : Principe de la transformation de levures par gap-repair et par mating.....	73
Figure 22 : Courbe de liaison obtenue lors d'une mesure d'affinité à l'équilibre.....	74
Figure 23 : Principe du séquençage Illumina.....	79
Figure 24 : Vue générale des étapes du DMS.....	81
Figure 25 : Exemple de DMS sur les CDR de la chaîne lourde d'un anticorps anti-HEL24.....	83
Figure 26 : Ingénierie de la spécificité d'un anticorps par DMS.....	83

## Liste des tableaux

Tableau 1 : Nombre d'allèles par gène HLA-II.....	32
Tableau 2 : Dénombrement des gènes V, D, J et C par chaîne d'anticorps.....	39
Tableau 3 : Liste des anticorps commercialisés ainsi que des données disponibles sur leur immunogénicité.....	43
Tableau 4 : Comparaison des différents jeux d'allèles décrits dans la littérature.....	53
Tableau 5 : Les différentes conditions utilisées par différents laboratoires pour réaliser des tests de stimulation des lymphocytes T.....	55
Tableau 6 : Recensement des différentes expériences de génération d'anticorps et de maturation d'affinité en <i>Yeast Surface Display</i> .....	76







## Introduction générale

Les anticorps thérapeutiques connaissent un succès très important depuis le début de leur commercialisation à la fin des années 1990. Ces molécules ont été utilisées pour des applications thérapeutiques variées notamment en cancérologie, pour le traitement des maladies inflammatoires et en immunologie contre le rejet de greffes. D'autres domaines d'utilisation des anticorps sont en train d'émerger avec des applications récentes dans le blocage des « immune-checkpoints » (PD-1/PDL-1) en cancérologie, dans le traitement des migraines ou encore dans le domaine des maladies infectieuses comme le VIH ou Ebola. Les nombreux succès cliniques ainsi que les perspectives d'application de ces molécules ont amené les laboratoires académiques et ceux de l'industrie pharmaceutique à renforcer leurs activités de recherche autour de ces molécules.

Les anticorps sont des molécules très performantes, notamment en terme d'affinité et de spécificité mais comme toute molécule biologique et à l'inverse des petites molécules chimiques, ils peuvent être immunogènes. En effet, ils sont susceptibles de déclencher une réponse immunitaire dirigée à leur encontre après avoir été injectés. En réagissant contre la protéine thérapeutique, le système immunitaire cible le médicament en produisant des anticorps dirigés contre ce dernier. Ce phénomène a été observé pour de nombreux anticorps thérapeutiques quel que soit leur degré d'humanisation.

Le laboratoire du Dr. Bernard Maillère travaille depuis plus d'une dizaine d'années sur la question de l'immunogénicité des protéines thérapeutiques et la localisation des zones immunogènes (épitopes T). Plus récemment, l'équipe s'est intéressée à la question de la dé-immunisation. Cette approche consiste à modifier la séquence en acides aminés des protéines thérapeutiques tout en conservant leur fonctionnalité, en particulier l'affinité et la sélectivité dans le cas d'un anticorps thérapeutique. Pour ce faire, des méthodes d'ingénierie des protéines comme le Yeast Surface Display ont été récemment développées au laboratoire.

D'autre part, le laboratoire du Dr. Stéphanie Simon travaille sur le développement d'anticorps dans le domaine des maladies infectieuses. Dans le contexte du problème grandissant de l'antibiorésistance ainsi que de l'émergence de nouveaux virus, l'utilisation des anticorps dans le domaine des maladies infectieuses pourrait permettre d'offrir de nouvelles solutions thérapeutiques. Les Shigelles et les Salmonelles font partie des bactéries pathogènes montrant une résistance de plus en plus fréquente à de nombreux antibiotiques. Dans ce contexte, des anticorps murins ciblant conjointement ces deux pathogènes ont été développés dans ce laboratoire, en ciblant une protéine du système de sécrétion de type III (SST3) chez les deux espèces bactériennes. Un anticorps s'est avéré être protecteur dans un modèle murin d'infection par *Salmonella*

Typhimurium et *Shigella flexneri*. Nous avons donc décidé d'utiliser cet anticorps comme preuve de concept pour le développement d'une technologie de dé-immunisation d'anticorps, généralisable par la suite. A plus long terme, et en fonction de l'efficacité thérapeutique de l'anticorps, il pourra être envisagé d'en faire un outil de l'arsenal thérapeutique utilisé pour combattre et prévenir les infections par les Shigelles et les Salmonelles, potentiellement utilisables comme arme du bioterrorisme, et encore mortelles dans de nombreux pays en voie de développement.

Dans ce contexte, ma thèse a consisté à de-immuniser un anticorps anti-Salmonelles/Shigelles. L'objectif est de transformer un anticorps murin en un anticorps dépourvu de séquence immunogène ayant un risque diminué de déclencher des réactions chez les patients traités. Dès lors, comment modifier la séquence d'un anticorps afin de réduire son immunogénicité tout en conservant ses caractéristiques fonctionnelles ?

Pour répondre à cette question, nous avons développé différents outils, combinant l'analyse *in silico* de l'immunogénicité des séquences, la génération de séquences potentiellement moins immunogènes et l'identification de variants dé-immunisés et actifs grâce à un crible fonctionnel en Yeast Surface Display. Le Yeast Surface Display est une méthode d'ingénierie moléculaire des protéines permettant d'exprimer et d'évaluer la fonctionnalité de millions de variants d'anticorps simultanément à la surface de levures. Une étape préliminaire à l'ingénierie de variants est d'identifier le format de fragment d'anticorps le plus adapté pour leur criblage fonctionnel à la surface des levures.

L'évaluation *in silico* de l'immunogénicité des séquences repose sur la prédiction de la liaison des peptides immunogènes (épitopes T) aux molécules HLA de classe II. Ces algorithmes sont utilisés pour localiser les peptides mais également pour identifier les mutations susceptibles de baisser leur risque d'immunogénicité. Ainsi, l'objectif principal de la thèse a été de combiner la prédiction de l'immunogénicité grâce à des algorithmes avec la sélection fonctionnelle de variants d'anticorps afin d'obtenir un anticorps fonctionnel *in vitro* et dé-immunisé *in silico*.

Au cours de ce travail, nous nous sommes également intéressés à la reconnaissance croisée des antigènes SipD de *Salmonella* et d'IpaD de *Shigella* par les anticorps monoclonaux développés au SPI. Cette double spécificité suggère la reconnaissance d'un épitope ayant une séquence en acides aminés et une conformation tridimensionnelle identique pour les deux antigènes, posant ainsi la question du mode d'action de cet anticorps anti-Salmonelles/Shigelles. Pour répondre à cette problématique, nous avons cherché à définir expérimentalement l'épitope de liaison de l'anticorps sur ses deux antigènes.

Le contexte scientifique du projet de thèse sera présenté en quatre chapitres. Le premier chapitre sera consacré à l'utilisation des anticorps thérapeutiques dans le domaine des maladies infectieuses. Un deuxième chapitre détaillera les mécanismes de pathogénicité des bactéries

entéropathogènes *Salmonella* et *Shigella* ainsi que les moyens mis en œuvre pour les combattre. Un troisième chapitre sera consacré aux mécanismes responsables de l'immunogénicité des anticorps thérapeutiques, aux méthodes permettant de la prédire et de l'évaluer ainsi qu'aux stratégies mises en place dans la littérature pour réduire l'immunogénicité des protéines thérapeutiques. Enfin, les méthodes d'ingénierie des anticorps par display permettant de modifier la séquence des anticorps tout en contrôlant leur fonctionnalité seront présentées dans un dernier chapitre.

Les résultats seront présentés en trois parties :

- 1) La définition de l'épitope de l'anticorps sur ses antigènes IpaD et SipD
- 2) Les résultats qui ont conduit au choix du format d'expression de l'anticorps à la surface des levures
- 3) L'humanisation ainsi que la dé-immunisation de l'anticorps murin

Les résultats des travaux de thèse ont donné lieu à une publication dans le journal Mabs ainsi qu'une publication en cours de soumission :

- Sierocki, R. & Sivelles, C. *et al.* Fab is the most efficient format to express functional antibodies by yeast surface display. *MAbs* **10**, 1–10 (2018).
- Sierocki, R. & Jneid, B. *et al.* An antibody targeting type III secretion system induces broad protection against *Salmonella* and *Shigella* infections.



## Contexte scientifique

### Chapitre 1 : L'utilisation des anticorps dans les maladies infectieuses

#### I. Historique

Les premiers essais d'immunothérapie passive à base d'anticorps pour le traitement des maladies infectieuses remontent à la fin du 19<sup>e</sup> siècle. A l'origine, on utilisait le sérum d'animaux immunisés contre l'agent infectieux pour traiter les patients infectés. Ces thérapies à base de sérum d'animaux ont été mises en œuvre pour la première fois dans les années 1890 et ont été utilisées jusque dans les années 1940. L'apparition des antibiotiques a rapidement conduit à l'abandon de ces thérapies.

De nombreux types d'infections bactériennes ont donné lieu au développement d'anti-sérum animaux ou humains, notamment celles causées par : *Corynebacterium diphtheriae*, *Streptococcus pneumoniae*, *Neisseria meningitidis*, *Haemophilus influenzae* et *Clostridium tetani*<sup>1</sup>. Au début des années 1890, Von Behring et Shibasaburo Kitasato ont généré un sérum animal capable de neutraliser l'effet des toxines produites par *Corynebacterium diphtheriae* et *Clostridium tetani*<sup>2</sup>. Behring a d'ailleurs reçu le Prix Nobel en 1901 pour avoir mis au point ce traitement efficace contre la diphtérie. Pendant la même période, Klemperer a montré l'efficacité des sérums dans les modèles d'infection par *Streptococcus pneumoniae* chez le lapin<sup>3</sup>. Ces résultats ont permis le développement d'un traitement similaire pour l'homme. Ces premières preuves de concept ont ouvert la voie à l'apparition de traitements à base de sérum pour les maladies causées par *Neisseria meningitidis*, *Haemophilus influenzae* et les streptocoques du groupe A<sup>4</sup>. L'efficacité de ces traitements a été évaluée par des études cliniques jusque dans les années 1930<sup>5</sup>. Par exemple, les traitements à base de sérum permettaient de faire chuter la mortalité de 30,7% à 20,6% chez les patients traités contre la pneumonie à pneumocoques. Pour la méningite à méningocoques, le taux de mortalité moyen sans traitement était situé entre 68% et 85% et entre 26% et 64% grâce à la thérapie sérique. Cependant, les effets secondaires associés à ce type de traitement, qui nécessite l'injection de sérum animal contenant de grandes quantités de protéines animales étaient très importants.

Après la découverte de la pénicilline par Fleming en 1928 et l'introduction des antibiotiques à partir des années 1930, la thérapie sérique a été largement abandonnée en l'espace de 10 ans<sup>1</sup>. Cette disparition est due à l'efficacité plus importante de ces nouveaux traitements à base d'antibiotiques ainsi qu'à leur coût et leurs faibles effets secondaires. Aujourd'hui encore, l'utilisation des thérapies sériques est limitée par le coût de fabrication de ces médicaments, leur

faible reproductivité entre lots, la faible spécificité du traitement et les effets secondaires liés au traitement.

Malgré l'amélioration des processus de purification des anticorps et la réduction des effets secondaires constatés, les immunothérapies passives sont principalement restreintes à l'utilisation des anticorps dirigés contre les toxines de venins, les toxines bactériennes et les virus pour lesquels les antibiotiques sont sans effet <sup>1,5</sup>. Aujourd'hui, les thérapies sériques consistent en des anticorps ou des fragments d'anticorps purifiés à partir de sérums de donneurs ou d'animaux ayant été préalablement exposés à des virus (virus de l'hépatite B, virus de la rage, virus respiratoire syncytial humain, virus de la vaccine, les echovirus et les entérovirus) ou à des toxines (venins, toxine tétanique, toxine botulique) <sup>4</sup>.

L'utilisation massive des antibiotiques dans le monde a conduit à l'apparition de souches bactériennes gram(-) ou gram(+) résistantes aux antibiotiques voire multi-résistantes comme c'est le cas par exemple de *Klebsiella pneumoniae* ou *Staphylococcus aureus* (Staphylocoque doré). Ce phénomène est reconnu depuis des années par les autorités comme un problème de santé publique majeur <sup>6</sup>. Au-delà du champ des maladies infectieuses, les thérapies à base d'anticorps connaissent un fort succès depuis la fin des années 1990 sous la forme des anticorps monoclonaux. Les immunothérapies sont en effet de plus en plus utilisées, notamment pour le traitement de nombreuses maladies inflammatoires et différents cancers <sup>7</sup>.

Dans ce contexte, l'utilisation des anticorps dans les maladies infectieuses pourrait permettre d'offrir de nouvelles solutions thérapeutiques contre les pathogènes multi-résistants aux antibiotiques. Des événements récents comme l'apparition de nouveaux virus (syndrome respiratoire aigu sévère et syndrome respiratoire du Moyen-Orient), la réémergence d'anciens virus (Ebola), les difficultés rencontrées dans le traitement des infections de patients immunodéprimés et l'émergence de menaces bioterroristes ont renforcé le besoin en nouvelles thérapies <sup>4</sup>.

L'exemple récent d'épidémie d'Ebola en Afrique de l'ouest à partir de 2014 montre bien l'intérêt et le besoin de développer de nouvelles solutions thérapeutiques, notamment à base d'anticorps thérapeutiques. Un cocktail expérimental d'anticorps a été utilisé (le ZMapp) constitué de 3 anticorps monoclonaux produits dans des plants de tabac. Ce traitement d'abord expérimenté sur des primates non-humains (PNH) avait donné des résultats très prometteurs : 100% de survie avec une injection des anticorps 5 jours après l'infection <sup>8</sup>. Dès 2015, le cocktail ZMapp a été utilisé sur des sujets humains au Libéria, en Sierra Leone, en Guinée et aux Etats-Unis. Sur les 72 patients enrôlés, le taux de mortalité de 22% semblait inférieur à celui observé pour le traitement anti-viral classique (37%) <sup>9</sup>. Depuis, de nouveaux cocktails d'anticorps ont été décrits et confèrent une protection large contre différentes souches Ebola dans différents modèles animaux <sup>10</sup>.



## II. Les différents mécanismes de protection

Dans le cas des maladies infectieuses, les anticorps thérapeutiques peuvent agir de deux façons différentes : soit ils neutralisent le pathogène en empêchant son entrée et sa multiplication dans les cellules humaines soit ils recrutent des effecteurs du système immunitaire qui vont venir détruire le pathogène et/ou les cellules infectées <sup>11</sup>.

### A. Neutralisation

#### 1. Neutralisation des toxines

De nombreuses bactéries pathogènes sécrètent des toxines qui sont la cause principale des symptômes (botulisme, tétanos, anthrax, ...). Les anticorps thérapeutiques monoclonaux ou polyclonaux provenant de patients immunisés contre la toxine peuvent alors être la solution thérapeutique classique pour le traitement de ces maladies <sup>11</sup>. Par exemple, pour le traitement de l'anthrax causé par des toxines sécrétées par *Bacillus anthracis* [PA (Protective Antigen), qui s'associe soit à EF (Edema Factor) ou LF (Lethal Factor)], deux anticorps monoclonaux sont disponibles ainsi qu'une solution d'anticorps polyclonaux provenant de donneurs ayant été immunisés grâce à un vaccin contre l'anthrax (Anthraxil). Des expériences *in vitro* ont montré que des anticorps monoclonaux dirigés contre PA bloquent l'interaction de la toxine avec les récepteurs membranaires <sup>12</sup> et sont donc neutralisants. Ces anticorps sont également protecteurs lorsque la toxine est injectée chez l'animal <sup>13</sup>. Des essais mettant en œuvre seulement la partie Fab d'anticorps protecteurs contre *B. anthracis* ont montré des résultats de protection similaires au traitement par les immunoglobulines complètes. Ceci démontre que le mécanisme de protection est bien de la neutralisation, les fragments Fab étant incapables de recruter les effecteurs du système immunitaire <sup>14</sup>.

#### 2. Neutralisation de l'entrée des pathogènes et de leur réplication

Les anticorps peuvent empêcher l'entrée des pathogènes, leur réplication et leur dissémination. La localisation intracellulaire de certains pathogènes pourrait laisser penser qu'ils seraient protégés des anticorps qui eux resteraient à l'extérieur de la cellule. En réalité, l'entrée du pathogène dans les cellules humaines peut être bloquée en utilisant des anticorps dirigés contre un antigène de surface de ce dernier. C'est le cas par exemple de différents anticorps dirigés contre le HIV <sup>15</sup>, le virus de l'hépatite C <sup>16</sup>, le virus Zika <sup>17</sup>, la dengue <sup>18</sup> et le paludisme <sup>19</sup>.

D'autre part, des protéines présentes à la surface des cellules ciblées par les pathogènes sont utilisées pour médier leur internalisation. De ce fait, ces protéines ayant un rôle clé dans l'infection, elles représentent également une cible intéressante pour bloquer l'entrée des pathogènes dans la cellule.

La cinétique de l'infection peut également être ralentie en ciblant les mécanismes de prolifération intracellulaire. Certains anticorps sont connus pour interférer avec les mécanismes de transport de l'ADN viral du papillomavirus humain (HPV) vers le noyau des cellules cibles <sup>20</sup>. Un autre anticorps bloque la machinerie de transcription du rotavirus <sup>21</sup> et un dernier inhibe la réplication de bactéries comme *Listeria monocytogenes* au sein des endosomes <sup>22</sup>. Dans ces exemples, la pénétration de l'anticorps à l'intérieur des cellules est assurée de différentes façons : il peut rentrer en même temps que le virus s'il est spécifique d'un antigène à sa surface (HPV), utiliser la capacité de transcytose du format IgA (rotavirus) ou entrer dans les endosomes grâce à l'endocytose constitutive ou la phagocytose (*Listeria monocytogenes*).

### 3. Neutralisation des facteurs de virulence microbiens

Les anticorps thérapeutiques peuvent également moduler la prolifération des pathogènes en neutralisant leurs facteurs de virulence. Pour de nombreux microorganismes pathogènes, une étape préalable à l'infection est l'adhérence aux surfaces mucosales de l'hôte. Aussi, les anticorps peuvent jouer un rôle neutralisant en interférant avec les mécanismes permettant aux bactéries d'adhérer aux cellules de l'hôte, aux muqueuses ou en empêchant la formation de biofilms.

Par exemple, un anticorps polyclonal dirigé contre les protéines de surface de *Streptococcus pneumoniae* empêche l'adhérence de la bactérie aux muqueuses des voies aériennes <sup>23</sup>. Les biofilms sont des communautés de microorganismes qui adhèrent entre eux en formant une structure plus ou moins complexe. Les biofilms permettent aux microorganismes qui sont situés à l'intérieur d'être moins exposés aux éventuelles substances antimicrobiennes qui seraient utilisées contre eux. Il a également été montré que certains anticorps dirigés contre une protéine de surface de *Staphylococcus aureus* et *Staphylococcus epidermidis* sont capables d'inhiber la formation de biofilms *in vitro* <sup>24</sup>.

## B. Recrutement des cellules effectrices du système immunitaire

### 1. Activation du complément

Les IgG monoclonaux (sauf les IgG4) ont la capacité d'activer la voie du complément. La voie du complément est une cascade de réactions qui permet la formation d'un complexe d'attaque membranaire aboutissant à la destruction de la cellule <sup>25,26</sup>. La voie classique du complément est activée par les anticorps faisant partie de l'immunité adaptative. Il est établi que l'activation du complément médiée par les anticorps joue un rôle majeur dans la résolution de nombreuses infections bactériennes et virales <sup>11</sup>. Cependant le mécanisme exact par lequel l'activation du complément contribue à la résolution de l'infection (cytotoxicité, mécanismes inflammatoires, phagocytose, ...) n'est pas totalement élucidé <sup>11</sup>.

L'activation du complément par les anticorps joue également un rôle dans le développement de l'immunité adaptative. Les complexes immuns peuvent être capturés par des récepteurs du complément à la surface des cellules B naïves puis transférés aux cellules dendritiques folliculaires qui vont présenter les antigènes du complexe immun dans les centres germinatifs et générer des anticorps spécifiques.

## 2. Antibody dependent cellular toxicity (ADCC)

L'ADCC résulte de l'interaction entre la partie Fc des anticorps et le récepteur FcγRIIIa des cellules NK (Natural Killer) qui provoque la sécrétion de perforine et/ou de granzyme et déclenche la mort de la cellule ciblée par les anticorps. Il a été montré que l'ADCC intervenait dans le contrôle et la clairance de différents pathogènes comme *Chlamydia trachomatis* <sup>27</sup>, *Mycobacterium tuberculosis* <sup>28</sup>, *Schistosoma mansoni* <sup>29</sup> et le HIV <sup>30</sup>.

Le rôle de l'ADCC dans le contexte du HIV est particulièrement marqué pour les thérapies à base d'anticorps monoclonaux. Ceux-ci peuvent non seulement empêcher la reconnaissance des lymphocytes T par le virus mais également se fixer sur les antigènes présents à la surface des lymphocytes infectés et déclencher une lyse des cellules-réservoirs via le mécanisme d'ADCC. Cependant, une fois que l'anticorps n'est plus dans le système, dans une grande majorité des cas le virus finit par réapparaître <sup>30</sup>, ce qui suggère qu'il y a une élimination incomplète des réservoirs de HIV médiée par le mécanisme d'ADCC.

## 3. Antibody-dependent cellular phagocytosis (ADCP)

Le dernier mécanisme permettant de recruter les cellules effectrices du système immunitaire est l'ADCP. Ce mécanisme permet d'éliminer les pathogènes opsonisés par des anticorps. L'opsonisation est un processus par lequel une cellule est recouverte d'un agent permettant de favoriser sa phagocytose par des cellules dotées de récepteurs spécifiques (monocytes, macrophages, cellules dendritiques et granulocytes). Bien que la phagocytose non médiée par les anticorps puisse être une voie d'entrée pour les pathogènes intracellulaires comme les Salmonelles, l'ADCP favorise le transport des vésicules contenant les pathogènes vers les vésicules lysosomales favorisant à la fois leur destruction et le développement d'une réponse immunitaire adaptative <sup>31</sup>.

# III. Les anticorps monoclonaux sur le marché pour le traitement des maladies infectieuses

Jusqu'en 2016 le palivizumab (Synagis, MedImmune) était le seul anticorps monoclonal sur le marché dans le domaine des maladies infectieuses. Approuvé en 1998 par la FDA, il est indiqué dans le traitement du virus respiratoire syncytial qui est une cause fréquente d'infection respiratoire chez l'enfant.

En 2016, l'obiltoxaximab <sup>32</sup> et le bezlotoxumab <sup>33</sup> ont été approuvés pour traiter les infections causées respectivement par *Bacillus anthracis* et *Clostridium difficile*. Ces deux anticorps ont le même mécanisme d'action : ils neutralisent la toxine produite par la bactérie en l'empêchant d'interagir avec les cellules de l'hôte.

En 2018 un nouvel anticorps a été approuvé : l'ibalizumab dans le traitement du HIV. L'ibalizumab est dirigé contre le CD4 des lymphocytes T CD4+ qui est le récepteur primaire du HIV. L'anticorps ne bloque pas l'interaction entre le CD4 du lymphocyte et le virus mais bloque la liaison du HIV avec les co-récepteurs CCR5 et CXCR4, empêchant son entrée dans les cellules cibles <sup>34</sup>.

Les anticorps anti-bactériens n'ont pas été développés car les antibiotiques étaient efficaces et peu chers. Cependant le développement d'anticorps anti-bactériens (autre que les anti-toxines) devrait s'accélérer dans le contexte actuel de l'antibiorésistance. Les plans de lutte nationaux contre la résistance aux antibiotiques passeront par le développement d'alternatives thérapeutiques dans lesquelles les anticorps auront toute leur place.

## Chapitre 2 : Infections par les Shigelles et les Salmonelles : Épidémiologie, mécanismes de pathogénicité et solutions thérapeutiques

### A. Les bactéries entéropathogènes du genre *Salmonella* et *Shigella*

#### 1. Données épidémiologiques (généralités)

Les genres bactériens *Salmonella* et *Shigella* appartiennent à la famille des Enterobacteriaceae. Les espèces pathogènes de ces genres bactériens sont des pathogènes stricts : elles sont absorbées par voie orale et provoquent des troubles intestinaux de gravité variable et peuvent également causer des infections systémiques. Les diarrhées aiguës bactériennes représentent encore aujourd'hui un problème de santé publique important, avec des conséquences graves notamment dans les pays en voie de développement. Plusieurs espèces bactériennes peuvent être responsables de diarrhées aiguës, en particulier *Salmonella enterica* et différentes espèces de *Shigella*. En termes de fréquence, les infections par *S. enterica* et *Shigella* spp. sont respectivement les troisième et quatrième causes de diarrhées bactériennes. Elles sont responsables d'une gamme de maladies allant de l'intoxication alimentaire aux infections systémiques mortelles, particulièrement dans les pays d'Afrique et d'Amérique du Sud<sup>35</sup>. Elles constituent une cause importante de mortalité chez les enfants âgés de moins de 5 ans et sont responsables de 20% des décès des enfants de 0 à 4 ans dans certains pays d'Afrique de l'Est et centrale<sup>36,37</sup>.

Conséquence de l'utilisation massive d'antibiotiques à large spectre pour traiter les infections aux salmonelles et aux shigelles, les souches d'entérobactéries sont de plus en plus multi-résistantes aux antibiotiques. Dans les pays endémiques, 50 à 75% des souches de Salmonelles non typhoïdales sont multi-résistantes aux antibiotiques notamment aux fluoroquinolones et aux céphalosporines de 3<sup>ème</sup> génération<sup>38</sup>. En Ethiopie, une étude menée sur 204 enfants a montré que 100% des *Shigella* étaient résistantes à l'ampicilline, 76,5% à la gentamicine et 64,7% au cotrimoxazole<sup>39</sup>.

Les Shigelles et les Salmonelles peuvent être considérées comme une arme biologique et constituent également une menace bioterroriste. Ces deux genres bactériens sont considérés comme agents du bioterrorisme de classe B (modérément facile à disséminer, taux de morbidité modéré et faible taux de mortalité) par le CDC (Centers for Disease Control and Prevention) aux Etats-unis<sup>40</sup>.

## 2. Espèces, sous-espèces et nomenclatures

Le genre *Salmonella* est composé de 2 espèces : *Salmonella enterica* et *Salmonella bongori*<sup>41</sup>. C'est l'espèce *Salmonella enterica* sous-espèce *enterica* qui est à l'origine de la majorité des cas cliniques d'infections aux salmonelles<sup>42</sup>. Plus de 2500 sérotypes (ou sérovars) de *Salmonella enterica enterica* ont été identifiés<sup>43</sup>. Par souci de clarification, l'espèce *enterica* et la sous-espèce *enterica* sont généralement sous-entendus et on nomme uniquement le genre *Salmonella* directement suivi du nom du sérovar. Les sérovars de cette sous-espèce sont divisés en deux groupes. Les sérovars *Salmonella* Typhi et Paratyphi A, B et C forment un premier groupe responsable de la fièvre typhoïde. Le second groupe comprend les sérovars non typhiques ou NTS (Non Typhoidal *Salmonella* ou NTS) dont *Salmonella* Typhimurium et Enteritidis.

Le genre *Shigella* est composé de 4 espèces : *dysenteriae*, *flexneri*, *boydii* et *sonnei*. Dans les pays en voie de développement, *S. flexneri* et *S. sonnei* sont à l'origine de la majorité des cas de shigellose. *S. dysenteriae* est quant à elle la cause des cas les plus sévères et est responsable de la majorité des cas mortels de shigellose<sup>44</sup>. En Europe, c'est *S. sonnei* qui est responsable de 60% des cas de shigellose<sup>45</sup>.

## B. Mécanismes de virulence des Shigelles et des Salmonelles

### 1. Les gènes impliqués dans la pathogénèse

L'identification des gènes impliqués dans la pathogénèse de *Salmonella* a été rendu possible grâce au séquençage complet de son génome<sup>42</sup>. Ce séquençage a révélé que *Salmonella* possède plusieurs îlots de virulence au sein de son chromosome bactérien. Les protéines codées au sein de ces îlots permettent à la bactérie d'infecter l'hôte et de se multiplier. Parmi les 12 îlots identifiés chez *Salmonella*, les deux principaux sont SPI-1 et SPI-2 (*Salmonella* Pathogenicity Island)<sup>46</sup>. Ces deux îlots contiennent plus de 40 gènes de virulence codant pour les protéines des systèmes de sécrétion de type 3 de *Salmonella* (SST3-1 et SST3-2). Le SST3-1 est requis pour l'invasion de l'hôte et le SST3-2 permet la dissémination systémique de la bactérie<sup>47</sup>.

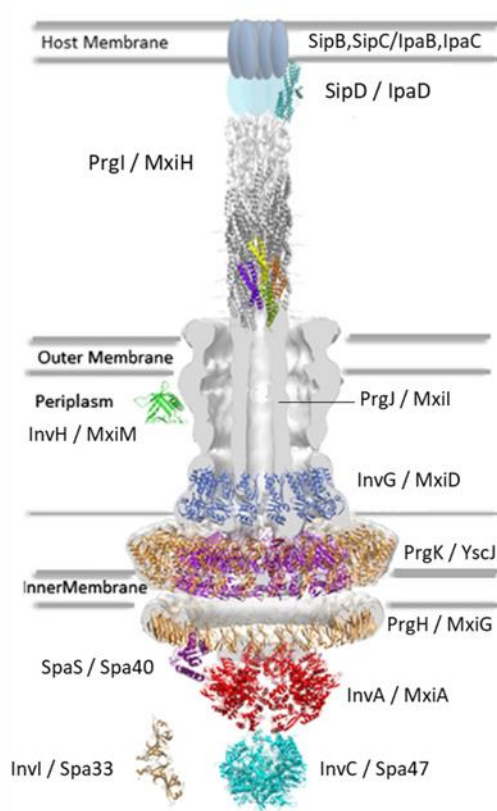
Chez les *Shigella spp.*, les principaux gènes de virulence sont regroupés dans un grand plasmide de virulence d'environ 230 kb<sup>48</sup>. Les gènes requis pour la pénétration de la bactérie au sein des cellules hôtes sont regroupés dans une région du plasmide appelée « entry region » qui code pour les composants du SST3 de *Shigella*<sup>49</sup>. En plus du SST3, certaines souches de *Shigella* sont capables de produire des entéro-toxines comme la *Shigella* enterotoxin 1 (*S. flexneri* 2a) et la Shiga-toxine (*S. dysenteriae*).

### 2. Le système de sécrétion de type 3

*Shigella* et *Salmonella* utilisent un système de sécrétion de type 3 pour transporter des protéines effectrices depuis le cytoplasme de la bactérie jusqu'à l'intérieur de la cellule hôte. Ce

système permet le passage de trois membranes : la membrane interne bactérienne, la membrane externe bactérienne et la membrane plasmique de la cellule hôte. Le SST3 également appelé injectisome est composé d'une vingtaine de protéines principalement insérées dans les membranes de la bactérie sous forme d'oligomères. Les structures principales des différents SST3 des différents genres bactériens qui en possèdent (*Shigella*, *Salmonella*, *Burkholderia*, *Yersinia*, etc...) sont très conservées<sup>50</sup>.

#### i. Structure générale



**Figure 1 : Vue structurale globale du système de sécrétion de type III de *Salmonella* et *Shigella*.** Pour chaque protéine du complexe, la protéine homologue chez *Salmonella* est indiquée à gauche et celle de *Shigella* est précisée à droite. Adapté de Burkinshaw *et al*<sup>52</sup>.

Les SST3 de *Shigella* et *Salmonella* ont une structure très proche. Ainsi il existe une protéine homologue entre les deux espèces pour la quasi-totalité des composants de l'injectisome (**Figure 1**). Le SST3 est divisé en trois parties cylindriques : un compartiment intracellulaire appelé « C-ring », une partie basale insérée dans les membranes bactériennes et une aiguille d'injection extracellulaire avec une coiffe à son extrémité.

La partie basale est composée de deux paires d'anneaux qui assurent l'ancrage du SST3 ainsi qu'une continuité entre le cytoplasme de la bactérie et le milieu extracellulaire<sup>51</sup>. L'aiguille d'injection mesure environ 60 nm de longueur pour un diamètre d'environ 2,5 nm. Le corps de l'aiguille est constitué d'un assemblage polymérique de la protéine PrgI (*Salmonella*) ou MxiH (*Shigella*). Le rôle de l'aiguille est d'exporter des protéines effectrices vers le milieu extracellulaire. Néanmoins, le diamètre étant trop petit pour faire passer des protéines entières repliées, celles-ci doivent transiter au sein de

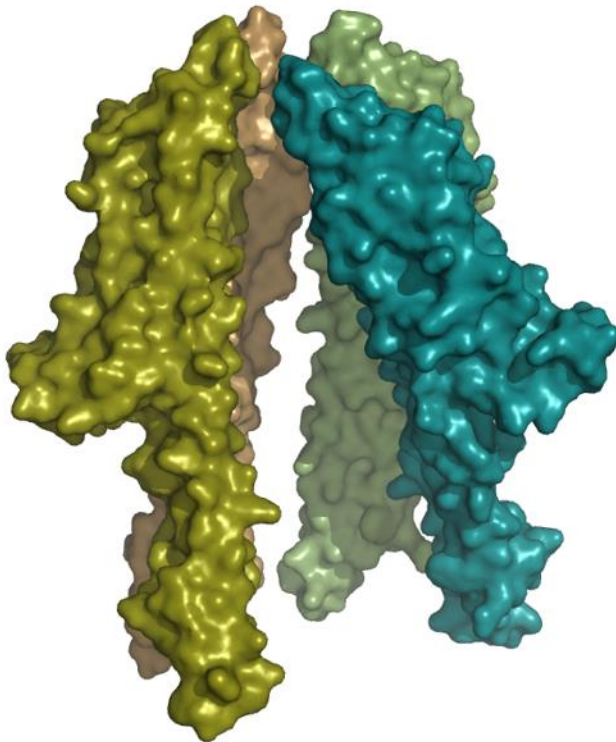
l'aiguille d'injection sous forme dépliée<sup>52</sup>. A l'extrémité de l'aiguille se trouve la coiffe ou TC (Tip Complex) dont la fonction est de détecter la présence de cellules hôtes, permettre au SST3 de s'insérer dans la membrane de la cellule hôte et réguler la sécrétion des protéines effectrices<sup>53</sup>.



## ii. Assemblage de la coiffe de l'aiguille d'injection

La coiffe de l'aiguille d'injection est le dernier complexe du SST3 à s'assembler. Le TC est un complexe multimérique qui contient la protéine IpaD (*Shigella*)<sup>53-59</sup> ou SipD (*Salmonella*)<sup>56,60</sup>. Les protéines IpaD et SipD s'assemblent à l'extrémité distale de l'aiguille d'injection et forment un anneau pentamérique qui empêche la sécrétion des effecteurs jusqu'à ce que les protéines soient en contact avec des petites molécules comme les sels biliaires<sup>61-63</sup> qui provoquent un réarrangement structural du pentamère et le recrutement des translocateurs hydrophobes IpaB (*Shigella*)<sup>62</sup> et SipB (*Salmonella*)<sup>64</sup> au niveau du TC pour préparer l'interaction avec la membrane de la cellule hôte. Lorsque le TC est au contact de la membrane de la cellule hôte, le deuxième translocateur hydrophobe IpaC chez *Shigella* ou son homologue SipC chez *Salmonella* est recruté à l'extrémité du SST3 pour former un pore dans la membrane et induire ensuite la sécrétion des protéines effectrices dans le cytoplasme de la cellule hôte<sup>65</sup>.

Différents modèles ont été proposés pour la structure de la coiffe de l'aiguille d'injection de *Salmonella* et *Shigella*. Tous les modèles s'accordent sur le fait qu'un pentamère se forme à l'extrémité de l'aiguille et qu'il contient au moins quatre copies d'IpaD ou SipD<sup>56,60</sup>. Cependant,



**Figure 2 : Représentation structurale d'un tétramère d'IpaD.** Représentation structurale de l'hypothèse d'assemblage de la coiffe de l'aiguille d'injection de *Shigella* selon le modèle 4 IpaD pour 1 IpaB (non représenté ici). PDB#4D3E<sup>66</sup>

l'identité du cinquième membre du pentamère reste sujet à controverse dans le cas d'IpaD et n'est pas discutée dans le cas de SipD. Pour IpaD, deux théories coexistent : soit le cinquième membre du pentamère est une autre copie d'IpaD ou ce cinquième membre est le premier translocateur IpaB. Etant donné la similitude de fonctionnement entre les SST3 de *Shigella* et *Salmonella*, une question identique sur l'identité du cinquième membre du pentamère serait probablement intéressante à explorer. L'étude de microscopie électronique la plus récente<sup>66</sup> émet l'hypothèse d'un IpaB pour quatre IpaD, hypothèse qui correspond aux cartes de densités électroniques obtenues à partir de TC en solution. Cheung *et al.* ont montré que les TC homopentamériques et hétéropentamériques coexistent, et



proposent que l'hétéropentamère est l'état actif du complexe. Ainsi, une molécule IpaB pourrait remplacer une molécule IpaD dans le pentamère après le signal d'activation. Blocker et ses collaborateurs ont ainsi proposé un modèle avec quatre monomères d'IpaD (**Figure 2**), sans la cinquième sous-unité supposée être IpaB dont la structure n'a été que partiellement résolue récemment <sup>67</sup>.

### C. La vaccination

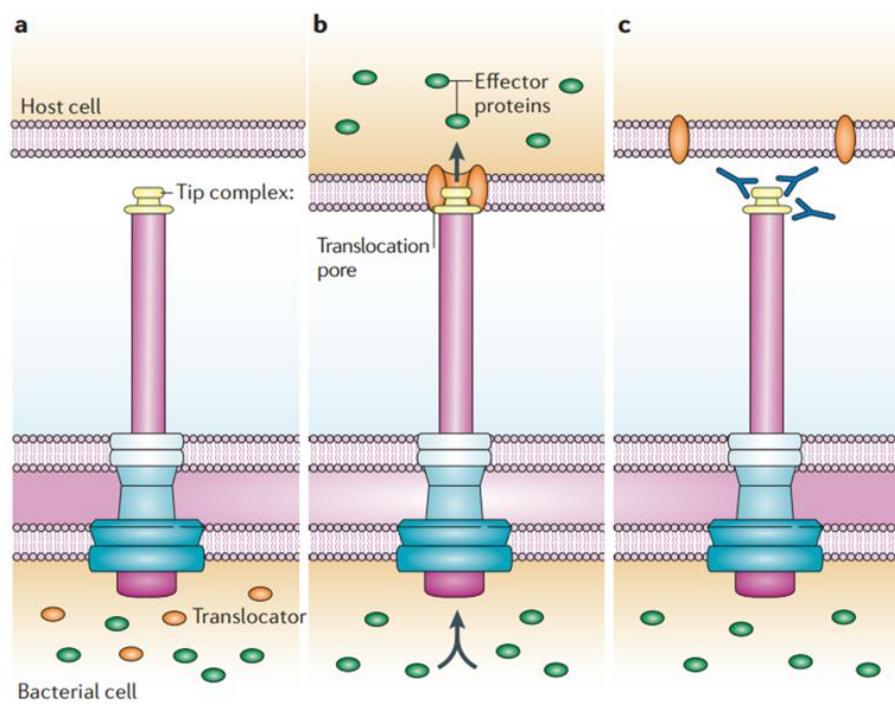
A ce jour, il existe deux vaccins autorisés contre une infection à *Salmonella* Typhi : le Ty21a qui est un vaccin atténué vivant <sup>68</sup> et la Vi CPS constitué de capsules polysaccharidiques de *S. Typhi* <sup>69</sup>. Ces deux vaccins sont dirigés contre *S. Typhi* et il n'existe pas de vaccin contre les trois autres sérovars de *Salmonella enterica enterica* (Paratyphi, Typhimurium et Enteritidis). Le principal défi des vaccins qui seront développés dans le futur est la protection croisée entre les différents sérovars de *Salmonella enterica*. Pour cela, l'approche utilisant des protéines recombinantes purifiées semble être prometteuse <sup>70</sup> en utilisant des protéines conservées dans la plupart des sérovars de *Salmonella* comme la flagelline ainsi que des protéines membranaires comme OmpC, F et D qui sont des porines présentes dans la membrane externe de *Salmonella*. Cependant ces protéines sont difficiles à produire et à conserver dans leur conformation repliée <sup>71</sup>, ce qui limite leur utilisation en tant que vaccin.

En ce qui concerne *Shigella*, il n'existe actuellement aucun vaccin autorisé sur le marché. Cependant, il existe plusieurs types de vaccins en développement : cellulaires, glycoconjugués ou à base de protéines recombinantes. Comme dans le cas de *Salmonella*, le principal défi de ces vaccins est de conférer une protection croisée entre les différentes espèces de *Shigella*. Il existe deux vaccins en cours de développement qui contiennent des protéines recombinantes purifiées du SST3 de *Shigella* : Invaplex (IpaD, IpaB et IpaC) et DB fusion (IpaD et IpaB fusionnées) <sup>72</sup>.

### D. L'immunothérapie passive : les anticorps dirigés contre le SST3

Le SST3 semble être une cible intéressante pour le développement de vaccins ou comme immunogène pour générer des anticorps monoclonaux spécifiques dirigés contre ce dernier (**Figure 3**) <sup>73</sup>. Différents anticorps monoclonaux dirigés contre des protéines du SST3 et plus particulièrement les protéines de la coiffe de l'aiguille d'injection du SST3 de différentes bactéries ont montré une efficacité *in vitro* et *in vivo* en bloquant les mécanismes de virulence de ces bactéries. Il a notamment été montré que des anticorps dirigés contre IpaD de *Shigella* peuvent empêcher la formation des pores dans la membrane des cellules cibles <sup>74</sup>. Des résultats similaires ont été obtenus *in vitro* grâce à des anticorps spécifiques de protéines de la coiffe de *Yersinia pestis* et de *Pseudomonas aeruginosa* <sup>75</sup>. Enfin, il a été montré que des anticorps dirigés contre la coiffe de *Pseudomonas aeruginosa* pouvaient être protecteurs dans des modèles animaux de pneumonie <sup>76</sup>.

Récemment, une thèse a été réalisée entre 2013 et 2016 par Bakhos Jneid sur le sujet suivant : « Evaluation de l'effet protecteur de protéines du Système de Sécrétion de Type III de bactéries entéropathogènes pour la vaccination et l'immunothérapie »<sup>77</sup>. Lors de cette thèse, des anticorps capables de reconnaître à la fois la protéine IpaD *Shigella* et son homologue SipD chez *Salmonella* ont été générés. Ces anticorps se sont montrés protecteurs dans les modèles d'infection murins aux *Salmonelles* et aux *Shigelles*. La première partie de la publication présentée en première partie des résultats du présent manuscrit détaille l'obtention de ces anticorps. Cependant, une étape d'ingénierie de ces anticorps murins pour réduire leur immunogénicité potentielle (humanisation, dé-immunisation) est requise pour que ces anticorps murins parviennent à satisfaire les exigences des autorités de régulation et puissent un jour être utilisés chez l'homme.



**Figure 3 : Représentation hypothétique du fonctionnement du SST3 de *Salmonella* et *Shigella*.** (a) Pas de contact avec la cellule hôte, un polymère d'IpaD/SipD forme un complexe à l'extrémité de l'aiguille d'injection. (b) Au contact de la membrane de la cellule hôte, les translocateurs IpaB/SipB et IpaC/SipC forment un pore dans la membrane de la cellule hôte. Cela permet au complexe IpaD/SipD de se fixer au pore néoformé et d'injecter les protéines effectrices. (c) Des anticorps anti-IpaD/SipD pourraient être protecteurs en empêchant la formation du pore et/ou la fixation de l'aiguille au pore. Adapté de Cornelis *et al.*, 2006

## Chapitre 3 : L'immunogénicité des anticorps thérapeutiques

### I. Les anticorps thérapeutiques : structure et modes de production

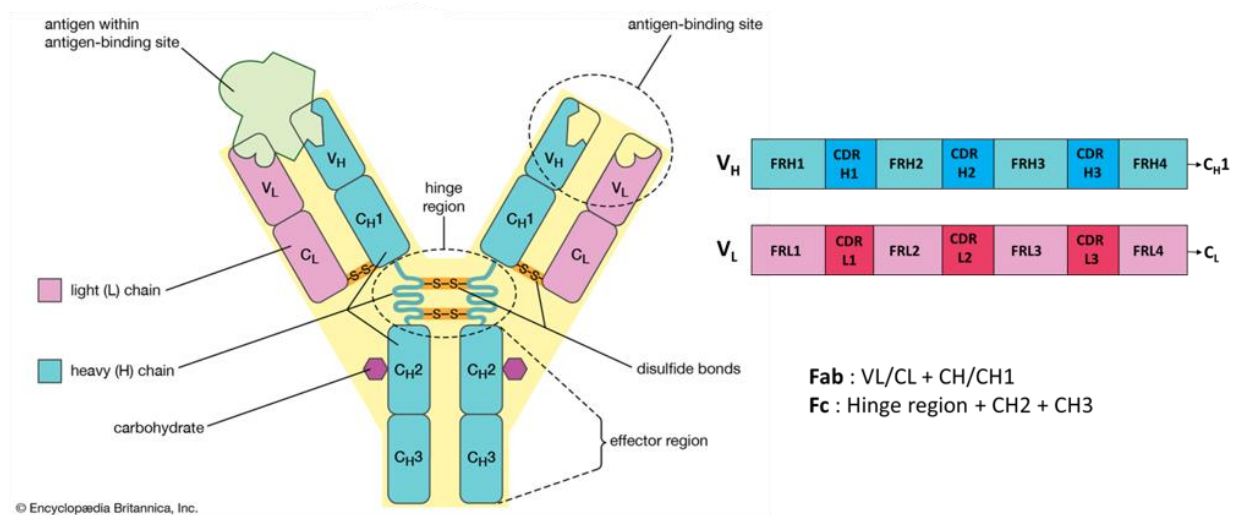
#### A. Structure et fonction des IgG

##### 1. Généralités

Les anticorps (ou immunoglobulines) sont des protéines glycosylées produites par le système immunitaire capables de se lier et de neutraliser des antigènes solubles ou présents à la surface de cellules ou de virus. Les anticorps sont constitués de deux chaînes lourdes et deux chaînes légères formant des hétérodimères reliés entre eux par des ponts disulfures. Ils sont classés en 5 isotypes en fonction de leur chaîne lourde : IgA, IgD, IgE, IgG et IgM. Chaque classe comporte une séquence en acides aminés, un schéma de glycosylation ainsi que des fonctions effectrices qui lui sont propres. En ce qui concerne les anticorps thérapeutiques actuellement sur le marché, seul l'isotype IgG est représenté. Les IgG sont divisées en 4 sous-classes qui diffèrent par leur structure et leur fonction.

##### 2. Structure des IgG

La structure générale des IgG est décrite dans la **Figure 4**, les chaînes lourdes des IgG sont composées de 4 domaines structurés (VH, CH1, CH2 et CH3) et d'une région plus flexible appelée région charnière qui relie CH1 et CH2. Les chaînes lourdes sont liées entre elles par des liaisons non covalentes ainsi que par un ou plusieurs ponts disulfures localisés dans la région charnière. Les chaînes légères sont composées de 2 domaines (VL et CL) et sont liées aux chaînes lourdes par des liaisons non covalentes entre les domaines VH/VL et CH1/CL ainsi qu'un pont disulfure interchaîne entre CH1 et CL. L'ensemble des domaines VH/ CH1 et VL/CL est appelé Fab (Fragment antigen-binding) et le dimère des domaines CH2 et CH3 forme le Fc (Fragment cristallisable). Les IgG ont la particularité de ne porter qu'un site de N-glycosylation dans le domaine CH2.



**Figure 4 : Structure générale d'un anticorps au format IgG1.** V<sub>H</sub>, domaine variable de la chaîne lourde. V<sub>L</sub>, domaine variable de la chaîne légère. C<sub>H</sub>, domaines constants de la chaîne lourde. C<sub>L</sub>, domaine constant de la chaîne légère. CDR<sub>H</sub>, régions hypervariables de la chaîne lourde. CDR<sub>L</sub>, régions hypervariables de la chaîne légère. FRH, framework de la chaîne lourde. FRL, framework de la chaîne légère.

Adapté de : <https://www.britannica.com/science/monoclonalantibody/images-videos#/media/1/389772/17658>

### i. Les domaines variables

Les domaines V<sub>H</sub> et V<sub>L</sub> confèrent à l'anticorps son affinité et sa spécificité grâce aux boucles hypervariables ou CDR (Complementary Determining Region) qu'ils portent. Il existe 3 CDR par chaîne variable, soit 6 par Fab. Dans la séquence primaire des chaînes variables, les zones situées entre les CDR sont des régions appelées *frameworks* qui ont un rôle structurant permettant de former le châssis protéique qui porte les CDR. Les séquences et les structures de ces domaines variables ont fait l'objet de nombreuses études qui ont conduit à la mise en place d'une classification structurale (structures canoniques) pour les CDR<sup>78</sup> ainsi que des systèmes de numérotation des résidus (Kabat, Chothia, Martin, Gelfand et IMGT)<sup>79</sup>. Au sein de la séquence des Fv (V<sub>H</sub> + V<sub>L</sub>), certaines positions clé ont été identifiées comme étant essentielles pour la structuration des chaînes variables : il s'agit des résidus appartenant à la « zone de vernier »<sup>80</sup>. Ces positions clé pour la structuration des CDR ainsi que les résidus dits « canoniques » sont utilisés pour définir l'appartenance de chaque CDR à une structure canonique.

### 3. Propriétés particulières des IgG

Les IgG représentent 75% des immunoglobulines sériques et ont la particularité d'avoir une demi-vie dans le sérum très longue pouvant atteindre 21 jours ainsi qu'une capacité d'opsonisation élevée. Exception faite de la sous-classe IgG4, les trois autres sous-classes d'IgG sont parmi les seules immunoglobulines à pouvoir activer la voie du complément et recruter des cellules effectrices du système immunitaire comme les natural killers (NK) ou les monocytes.<sup>81</sup>

#### B. La production des anticorps recombinants

Les anticorps thérapeutiques nécessitent souvent l'administration de grandes quantités chez les patients, parfois plusieurs centaines de milligrammes par injection. Pour répondre à cette problématique, de nombreux outils sont aujourd'hui disponibles pour la production recombinante des immunoglobulines<sup>82</sup>.

En laboratoire, il existe de nombreuses possibilités dans le choix de l'hôte de production en ce qui concerne les anticorps recombinants. La production d'anticorps recombinants a été décrite dans des systèmes d'expression tels que les plantes, les levures, les cellules d'insectes ainsi que les cellules de mammifères ou humaines. La production d'IgG est également décrite en bactéries mais elle est souvent restreinte à la production de fragments<sup>83</sup>.

Du fait de leur taille et de la présence de nombreuses modifications post-traductionnelles, la production d'anticorps recombinants est souvent réalisée grâce à des cellules d'organismes eucaryotes, notamment des cellules de type CHO (Chinese Hamster Ovary) ou HEK (Humain Embryonic Kidney) en transfection transitoire ou stable<sup>82</sup>.

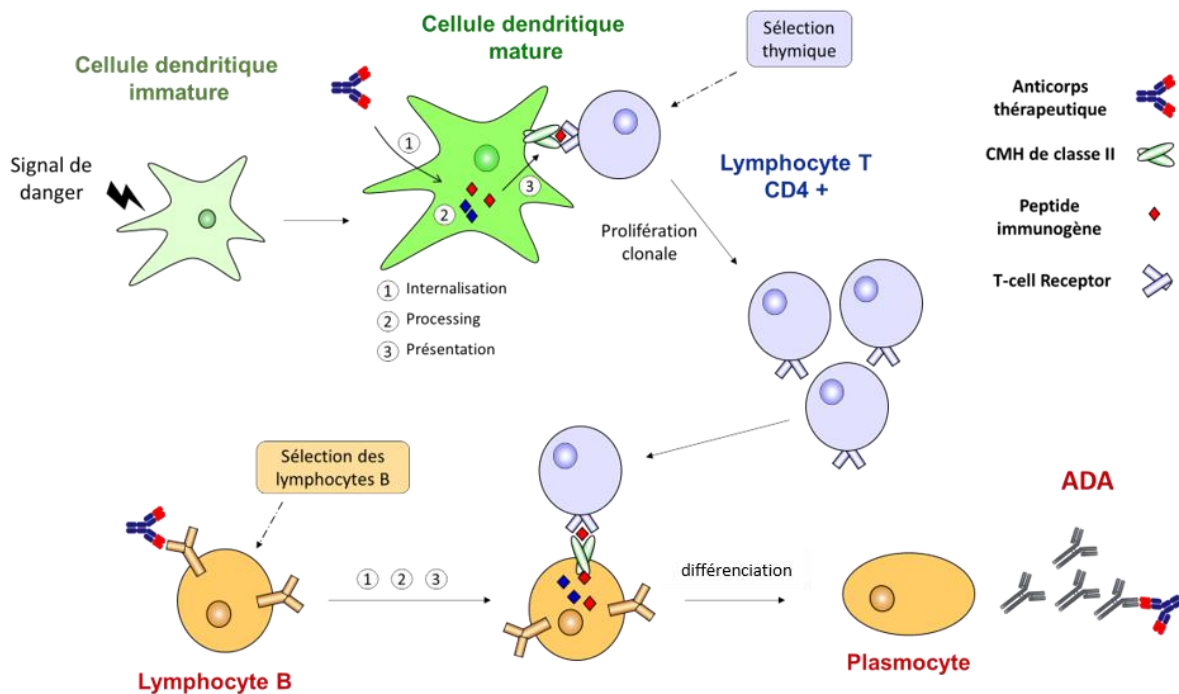
Pour la production industrielle, les anticorps sont quasiment exclusivement produits par des lignées stables de cellules CHO. Du point de vue des acteurs de la production industrielle, les cellules CHO présentent plusieurs avantages. Le système CHO est utilisé depuis de nombreuses années et il peut ainsi être plus aisé d'obtenir l'accord des autorités de régulation. De plus, le système de production a été largement optimisé permettant des rendements de l'ordre du gramme par litre. Enfin, les cellules CHO peuvent être cultivées dans un milieu synthétique sans sérum, apprécié des autorités de régulation.<sup>84</sup>

Les anticorps sont des protéines thérapeutiques et peuvent être reconnues par le système immunitaire du patient comme étranger. Cette reconnaissance est appelée immunogénicité et peut avoir de nombreux effets délétères sur le patient et le traitement.



## II. Mécanisme : La réponse cellulaire adaptative et la génération d'anti-drug antibodies (ADA)

### A. Mécanisme général aboutissant à la sécrétion d'ADA



**Figure 5 : La coopération cellulaire dans la réponse humorale dirigée contre les protéines thérapeutiques.** La réponse humorale vis-à-vis d'une protéine thérapeutique fait intervenir au moins trois types cellulaires. Les cellules dendritiques immatures capturent la protéine et la dégradent en peptides. Parmi les peptides générés, certains possèdent une affinité pour les molécules HLA de classe II et peuvent être présentés aux lymphocytes T CD4. Sous l'effet de facteurs inflammatoires, les cellules dendritiques se différencient en cellules dendritiques matures et deviennent aptes à stimuler les lymphocytes T naïfs. Les cellules dendritiques matures activent les lymphocytes T naïfs spécifiques du peptide présenté. La protéine thérapeutique parvient également aux lymphocytes B possédant un anticorps de surface spécifique de la protéine et selon un processus similaire à celui décrit pour les cellules dendritiques, des peptides sont présentés aux lymphocytes T CD4 spécifiques et les activent. Leur activation permet aux lymphocytes B de se différencier en plasmocytes sécréteurs d'anticorps spécifiques de la protéine appelés ADA (Anti-Drug Antibodies). Adapté de Maillère *et al* <sup>85</sup>.

Le mécanisme aboutissant à une réponse immunitaire dirigée contre un anticorps thérapeutique est présenté en **Figure 5** et repose sur la coopération de trois acteurs du système immunitaire : les cellules présentatrices d'antigènes (CPA), les lymphocytes T CD4+ et les lymphocytes B <sup>85</sup>. Les CPA professionnelles que sont les cellules dendritiques internalisent les protéines présentes dans le milieu extracellulaire puis des protéases digèrent la protéine en peptides de taille variable. Ces peptides sont chargés sur les molécules HLA de classe II et les complexes HLA-II/peptides sont exportés et présentés à la surface des cellules dendritiques <sup>86</sup>. Les lymphocytes T CD4+ expriment à leur surface un récepteur T appelé TCR (T-cell Receptor) capable de reconnaître spécifiquement les complexes HLA-II/peptides. Ces lymphocytes ont



préalablement été contre-sélectionnés dans le thymus pour ne pas reconnaître des peptides issus de protéines du soi.<sup>87</sup> S'il existe des lymphocytes T CD4+ spécifiques du complexe HLA-II/peptide issus de l'anticorps thérapeutique, ceux-ci s'activent et se multiplient. Les derniers acteurs de cette réponse immunitaire sont les lymphocytes B. Ces cellules présentent à leur surface des anticorps membranaires et sont également contre-sélectionnées dans la moelle osseuse pour ne pas reconnaître de protéines du soi. S'il existe des lymphocytes B portant un BCR (B-cell receptor) spécifique de l'anticorps thérapeutique, celui-ci est internalisé, découpé en peptides et présenté sous la forme de complexes HLA-II/peptide à la surface des lymphocytes B. Les lymphocytes T CD4+ qui ont préalablement été activés par les cellules dendritiques grâce aux mêmes complexes HLA-II/peptide vont à leur tour reconnaître ces complexes à la surface des lymphocytes B, ce qui provoque la différenciation des lymphocytes B en plasmocytes sécréteurs d'anticorps.<sup>86</sup>

## B. Les cellules dendritiques

Les cellules dendritiques font partie des cellules présentatrices d'antigènes professionnelles c'est à dire qui participent à l'initiation des réponses immunitaires spécifiques. Les cellules dendritiques sont en effet parmi les seules cellules à pouvoir activer les lymphocytes T naïfs<sup>88</sup>.

### 1. Internalisation

Les anticorps présents dans le milieu extracellulaire peuvent être internalisés par les cellules dendritiques via différentes voies : la macropinocytose, l'endocytose médiée par les récepteurs et la phagocytose<sup>89</sup>.

La macropinocytose est un processus d'endocytose non sélectif permettant de capturer de grandes quantités de matériel extracellulaire<sup>89</sup>. Ce processus repose sur la déformation de la membrane plasmique par les filaments d'actine du cytosquelette<sup>86</sup>. Les cellules dendritiques immatures utilisent ce processus de façon importante et constitutive. On pense que cela constitue le mécanisme principal d'internalisation des antigènes extracellulaires par les cellules dendritiques<sup>90</sup>.

L'endocytose médiée par les récepteurs, est, à l'inverse du processus précédent, spécifique. Il existe deux types de récepteurs aux parties constantes des immunoglobulines présents à la surface des cellules dendritiques : le FcRn et les FcγRs<sup>86</sup>. Le récepteur néonatal FcRn est responsable du recyclage des IgG en les protégeant de la dégradation lysosomale. Le FcRn possède une forte affinité pour les parties constantes des IgG à pH acide (pH 6,0) mais pas à pH physiologique (pH 7,4)<sup>91</sup>. Cette propriété permet de capturer les anticorps présents dans les endosomes acides et de les relâcher à la surface de la cellule. Les récepteurs FcγRs peuvent lier les IgG à pH physiologique et sont divisés en deux familles suivant qu'ils inhibent ou activent la maturation des endosomes en endosomes acides tardifs<sup>92</sup>. Deux récepteurs FcγRs sont



principalement exprimés à la surface des cellules dendritiques : FcγRI, FcγRIIa et FcγRIIb. Le FcγRI et le FcγRIIa activent la maturation des endosomes et le FcγRIIb l'inhibe <sup>92,93</sup>. Les cellules dendritiques matures expriment faiblement le récepteur FcγRIIb, favorisant la maturation des endosomes. Au laboratoire, les cellules dendritiques sont obtenues à partir de monocytes cultivés en présence de GM-CSF et IL-4 <sup>94</sup>. Le récepteur FcγRI est exprimé par la cellules dendritiques humaines mais n'est pas exprimé par les cellules dendritiques dérivées de monocytes *in vitro* (moDC). Son importance peut donc être sous-estimée dans des expériences utilisant des moDC <sup>92</sup>.

La phagocytose est un processus qui peut également conduire à l'internalisation des anticorps du milieu extracellulaire par les cellules dendritiques <sup>95</sup>. La phagocytose intervient dans l'internalisation des corps opsonisés. Cette phagocytose spécifique est médiée par les récepteurs Fc présents à leur surface <sup>86</sup>. Les anticorps sont donc internalisés en même temps que les antigènes qu'ils ciblent.

La macropinocytose, l'endocytose médiée par les récepteurs et la phagocytose sont des processus qui aboutissent à l'internalisation des anticorps circulants au sein d'endosomes précoces dans le cytoplasme des cellules dendritiques.

## 2. L'apprêtement de l'antigène

Les antigènes internalisés par les cellules dendritiques sont dégradés dans les endosomes tardifs ou dans les structures lysosomales <sup>96</sup>. Ces compartiments cellulaires sont riches en disulphides réductases et en enzymes protéolytiques activées à pH acide <sup>86</sup>. Les protéases lysosomales comme l'asparaginyl endopeptidase <sup>97</sup> ainsi que la cathepsine S <sup>98,99</sup> jouent un rôle important pour la dégradation des peptides endogènes mais aussi dans la génération des peptides antigéniques qui vont se lier au complexe HLA-II. Plusieurs changements surviennent lors de l'activation des cellules dendritiques : les différentes cathepsines (S, B, H et L) sont redistribuées des lysosomes vers les endosomes tardifs où se trouvent les CMH-II <sup>100</sup> et l'activité des pompes à protons responsables de l'acidification des endosomes augmente <sup>101</sup>. Ces phénomènes participent à favoriser la génération des peptides à partir des antigènes ainsi que la formation de complexes peptides/HLA-II dans les endosomes tardifs. Des travaux ont été conduits pour élucider les différents sites de clivage des différentes cathepsines <sup>102</sup>. Malgré leur préférence pour des glycines en amont et en aval du site de clivage, ces protéases n'ont pas de site de clivage très spécifique qui permettrait de prédire les peptides antigéniques obtenus après dégradation dans le lysosome à partir d'une protéine donnée.

La protéolyse des antigènes est un processus régulé. En effet, une activité lysosomale trop forte conduirait à la destruction complète des peptides antigéniques. Beaucoup d'épitopes T sont générés dans les endosomes précoces et sont détruits par la protéolyse lysosomale <sup>103,104</sup>.

Cependant, il existe des mécanismes permettant de ménager cette dégradation. Les extraits de lysosomes de cellules dendritiques sont 50 fois moins actifs que ceux issus de macrophages<sup>103</sup> ce qui leur permet de garder des antigènes internalisés intacts pendant plusieurs heures. L'équilibre entre une protéolyse ménagée et une destruction complète de l'antigène dans les phagosomes est modulé par la NADPH oxydase 2 (NOX2) qui participe à augmenter le pH du compartiment et prolonge la persistance des antigènes ainsi que la présentation aux lymphocytes T CD4+ par les CMH-II<sup>105</sup>.

### 3. Présentation de l'antigène

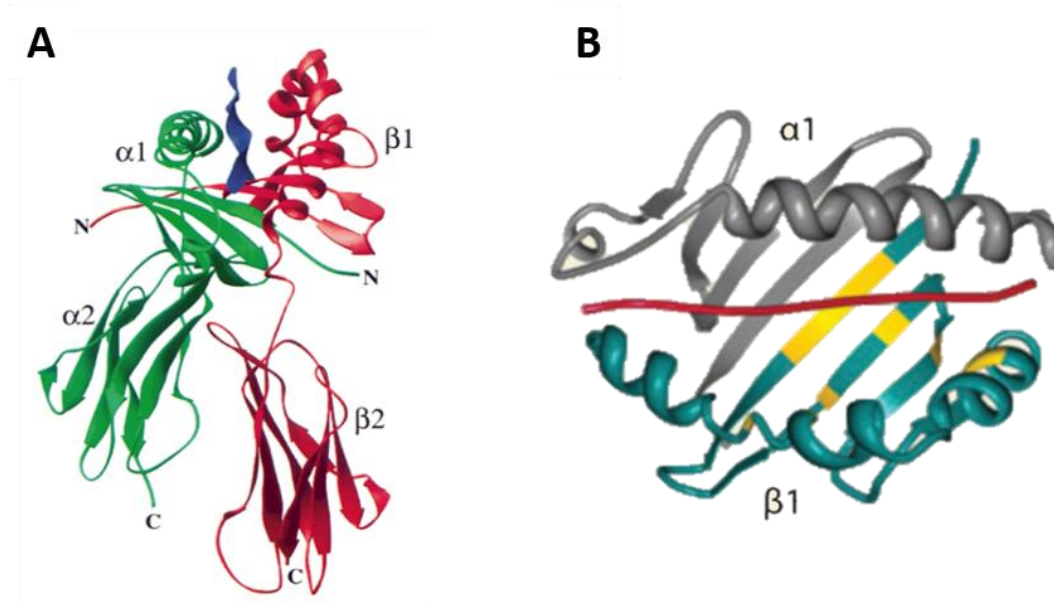
#### *i. Les molécules HLA de classe II : structure et génétique*

Les molécules HLA de classe II sont des hétérodimères constitués de deux chaînes polypeptidiques transmembranaires  $\alpha$  et  $\beta$ <sup>106</sup>. Le domaine proximal (par rapport à la membrane plasmique) est constitué des sous-domaines  $\alpha_2$  et  $\beta_2$  et le domaine distal est formé par les sous-domaines  $\alpha_1$  et  $\beta_1$  qui constituent le site de liaison des peptides. Ce domaine forme un sillon ouvert aux extrémités dans lequel un peptide de taille variable peut venir se loger (**Figure 6**)<sup>89,107,108</sup>. Il existe trois séries de molécules HLA-II : les molécules DR, DQ et DP, chacune étant codée par les gènes correspondants (DRA/DRBx, DQA1/DQB1 et DPA1/DPB1). Pour la molécule DR il existe plusieurs gènes DRB : DRB1, DRB3, DRB4 et DRB5. Les chaînes  $\beta$  des molécules HLA-II sont très polymorphes tandis que les chaînes  $\alpha$  sont considérées comme étant beaucoup moins variables (il n'existe que 2 allèles pour HLA-DRA dont un est majoritaire) (**Tableau 1**).

**Tableau 1** : Nombre d'allèles par gène HLA-II

<i>Gène HLA-II</i>	<i>DRA</i>	<i>DQA1</i>	<i>DPA1</i>	<i><b>DRB1</b></i>	<i>DRB3</i>	<i>DRB4</i>	<i>DRB5</i>	<i><b>DQB1</b></i>	<i><b>DPB1</b></i>
<i>Allèles</i>	7	52	38	<b>1540</b>	58	15	20	<b>664</b>	<b>422</b>
<i>Protéines</i>	2	32	19	<b>1139</b>	46	8	18	<b>435</b>	<b>351</b>

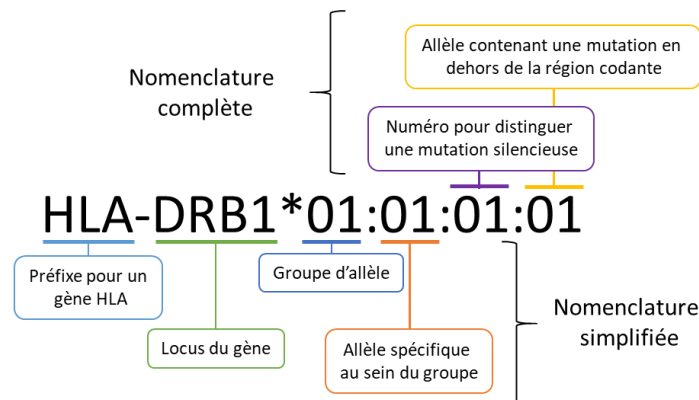
Adapté de : <https://www.ebi.ac.uk/ipd/imgt/hla/stats.html>



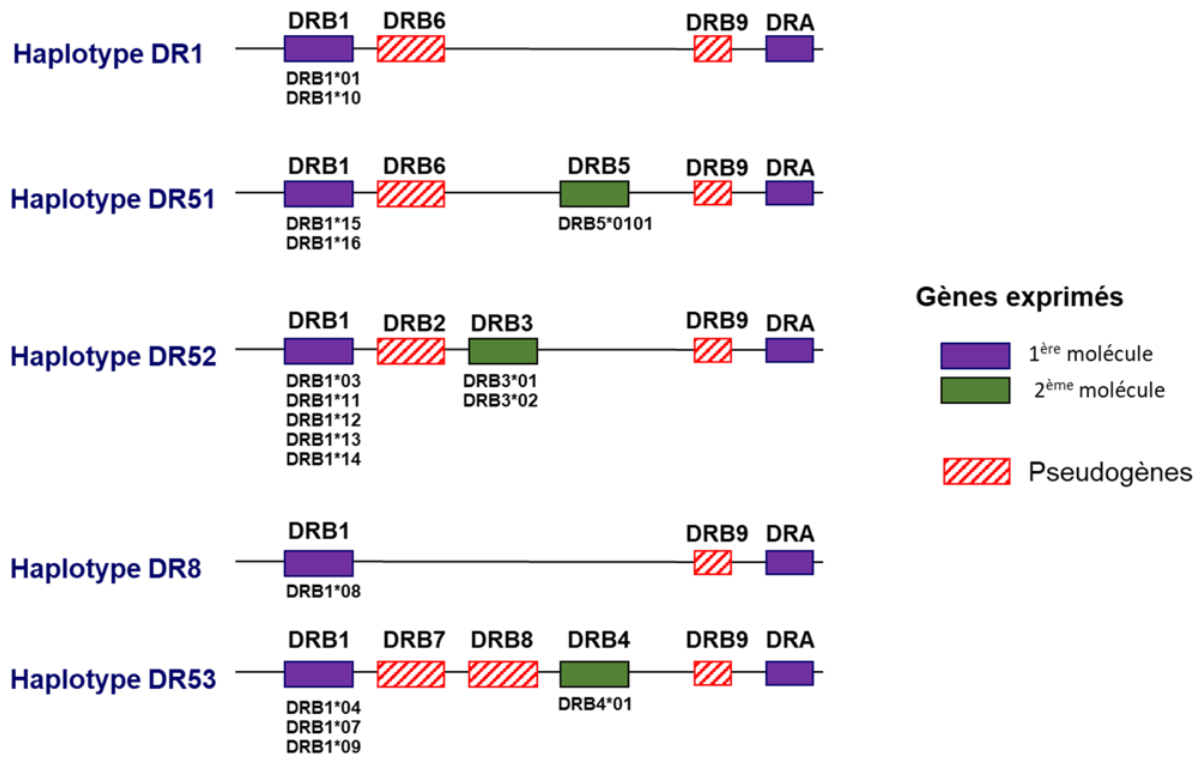
**Figure 6 : Représentation structurale d'un CMH-II (HLA-DR1) en complexe avec un peptide.** A, complexe vu de profil. La chaîne  $\alpha$  et la chaîne  $\beta$  forment une structure qui permet au peptide (en bleu) de se lier au complexe. Les domaines  $\alpha 1$  et  $\beta 1$  forment une gorge ouverte à ses deux extrémités. Les domaines  $\alpha 2$  et  $\beta 2$  forment un domaine de type immunoglobuline qui vient s'attacher à la membrane via un domaine transmembranaire. B, Gorge du CMH-II vue du dessus en complexe avec un peptide. Adapté de Batalia *et al*<sup>108</sup> et Blum *et al*<sup>89</sup>.

Les gènes des molécules HLA-II sont localisés sur le chromosome 6<sup>109</sup>. La proximité entre les loci des gènes des molécules HLA-II conduit à l'existence de 5 haplotypes (**Figure 7**)<sup>110</sup>, c'est-à-dire la transmission d'une combinaison d'allèles en bloc à la descendance. Ainsi, en fonction de l'état homozygote ou hétérozygote de l'individu pour les loci des molécules HLA, il peut avoir entre 1 et 4 molécules DR différentes (en considérant DRA comme monomorphe), 1 et 4 molécules DQ et 1 et 4 molécules DP soit entre 3 et 12 molécules HLA-II différentes.

Afin d'uniformiser la nomenclature des molécules HLA une nomenclature spécifique a été mise en place<sup>109</sup>:



En ce qui concerne les molécules DR, il est d'usage de ne préciser que le nom de la chaîne  $\beta$ , la chaîne  $\alpha$  étant considérée comme monomorphe au niveau protéique.

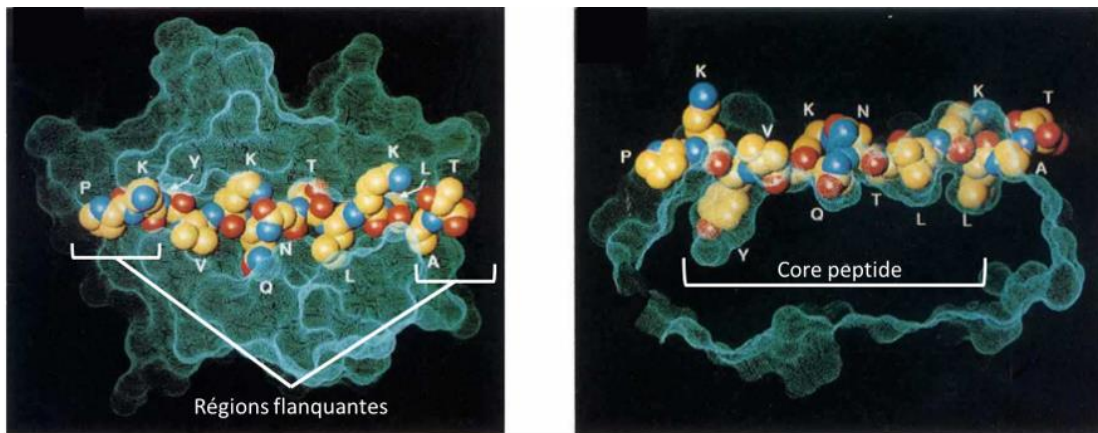


**Figure 7 : Représentation schématique des 5 haplotypes HLA-DR humains.** Les gènes non exprimés (pseudogènes) sont représentés dans des boîtes hachurées. Les groupes d'allèles correspondant aux différents haplotypes sont précisés en dessous des loci DRBX. Par exemple, posséder une molécule DRB du groupe d'allèles DRB1\*15 implique de posséder également une molécule DRB5. Adapté de Anderson *et al* <sup>110</sup>.

## ii. Les molécules HLA de classe II : interaction HLA-II/peptide

La cristallographie de complexes HLA-II/peptide a permis de mieux décrire les modalités d'interaction entre les peptides et les molécules HLA <sup>111</sup>. Le site de liaison s'apparente à un sillon formé par un feuillet  $\beta$  et deux hélices  $\alpha$ . L'agencement et la nature des acides aminés qui composent le sillon forment des cavités que l'on appelle « poches » dont les caractéristiques et les acides aminés qui peuvent s'y loger varient en fonction des molécules HLA et de leurs allèles. Certains acides aminés du peptide qui vient se loger dans le sillon de la molécule HLA-II interagissent avec des poches de liaison présentes dans ce sillon (**Figure 8**). Ces acides aminés sont appelés « positions d'ancrage » et sont notés PX. Dans le cas des molécules HLA de classe II, il existe 5 positions d'ancrage : P1, P4, P6, P7, P9. On distingue également la zone du peptide qui comprend les acides aminés situés entre la première et la dernière position d'ancrage (typiquement 9 acides aminés) que l'on appelle « *core peptide* » ou cœur faisant référence aux 9 acides aminés responsables de la liaison du peptide à la molécule HLA-II. Les acides aminés situés à l'extérieur du core appartiennent aux régions flanquantes qui participent également à

l'interaction via des liaisons hydrogène entre le squelette carboné du peptide et les acides aminés de la molécule HLA.

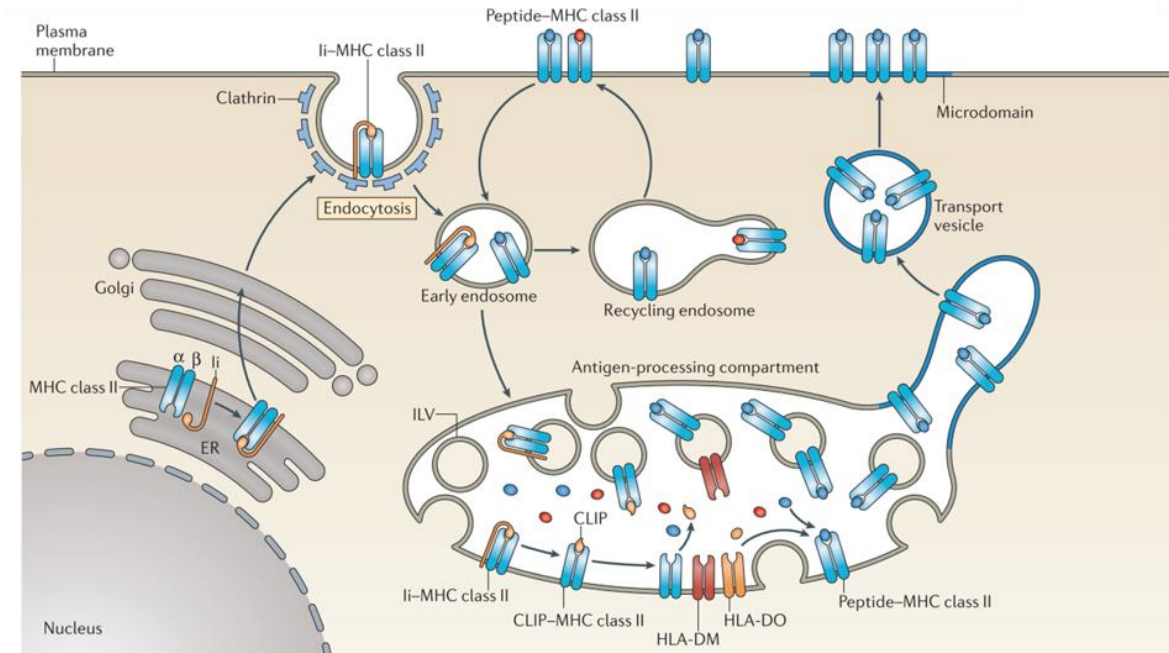


**Figure 8 : Vue structurale en trois dimensions des poches de liaison d'une molécule HLA-DR1 en complexe avec un peptide.** A gauche, le complexe HLA-II/peptide vu du dessus où l'on peut distinguer les régions flanquantes du peptide (PK et AT). A droite, coupe longitudinale du complexe où l'on peut distinguer les poches de liaisons P1 (Y), P4 (Q), P7 (L) et P9 (L) au sein du core peptide de 9 acides aminés (YVKQNTLKL). Adapté de Stern *et al* <sup>111</sup>.

### iii. *Apprêtement des peptides sur les molécules HLA-II*

Les différentes étapes conduisant au chargement des peptides antigéniques sur les molécules HLA de classe II sont résumées sur la **Figure 9**. Les molécules HLA sont synthétisées dans l'appareil de Golgi en complexe avec une protéine appelée « invariant chain » ou « Ii » qui contient un motif d'adressage pour les compartiments où se déroule le traitement des antigènes (endosome/lysosome) <sup>112</sup>. Après avoir été envoyé à la membrane plasmique, le motif d'adressage de la protéine Ii est reconnu et provoque l'endocytose du complexe HLA/Ii <sup>113</sup>. Les endosomes contenant les complexes HLA/Ii trafiquent à travers la voie endocytaire jusqu'aux endosomes tardifs et aux lysosomes qui contiennent les peptides antigéniques <sup>114</sup>. Une fois arrivé dans le compartiment de traitement des antigènes, le peptide Ii est dégradé par les cathepsines et seul un petit fragment du peptide appelé CLIP (Class II-associated Invariant chain Peptide) reste lié à la molécule HLA. Pour que la molécule HLA puisse lier un peptide antigénique, CLIP doit se dissocier du sillon de la molécule HLA. Cette dissociation est facilitée par l'enzyme HLA-DM dont l'activité est elle-même régulée par HLA-DO. Ces deux protéines ont une structure proche de celle des molécules HLA-II et sont également encodées au sein du locus du complexe majeur d'histocompatibilité <sup>115</sup>. Après la dissociation de CLIP, la molécule HLA-II se lie à un peptide antigénique et quitte le compartiment d'apprêtement des antigènes via des vésicules de transport vers la membrane plasmique et ainsi permet la reconnaissance du complexe HLA-II/peptide par un lymphocyte T CD4+.





**Figure 9 : Génération des complexes peptide/CMH-II dans les cellules dendritiques.** Les CMH-II s'associent à la chaîne invariante (Ii) dans le réticulum endoplasmique (ER) et trafiquent à travers l'appareil de Golgi jusqu'à la membrane plasmique. Les complexes Ii/MHC-II sont internalisés par endocytose médiée par les clathrines et trafiquent vers le compartiment d'apprêtement des antigènes. Dans ce compartiment la protéolyse séquentielle de Ii conduit à la génération d'un fragment de Ii, appelé peptide à chaîne invariante (CLIP), qui reste dans la rainure de liaison du CMH-II. CLIP est éliminé des complexes CLIP/MHC-II par l'enzyme HLA-DM permettant ainsi la liaison d'un nouveau peptide sur le MHC-II naissant. L'activité de HLA-DM est régulée par HLA-DO, mais le mécanisme de régulation reste inconnu. Les complexes peptide/MHC-II sont transportés jusqu'à la membrane plasmique via des vésicules de transport qui fusionnent avec cette dernière. Adapté de Roche *et al* <sup>114</sup>.

## C. Les lymphocytes T et la sélection thymique

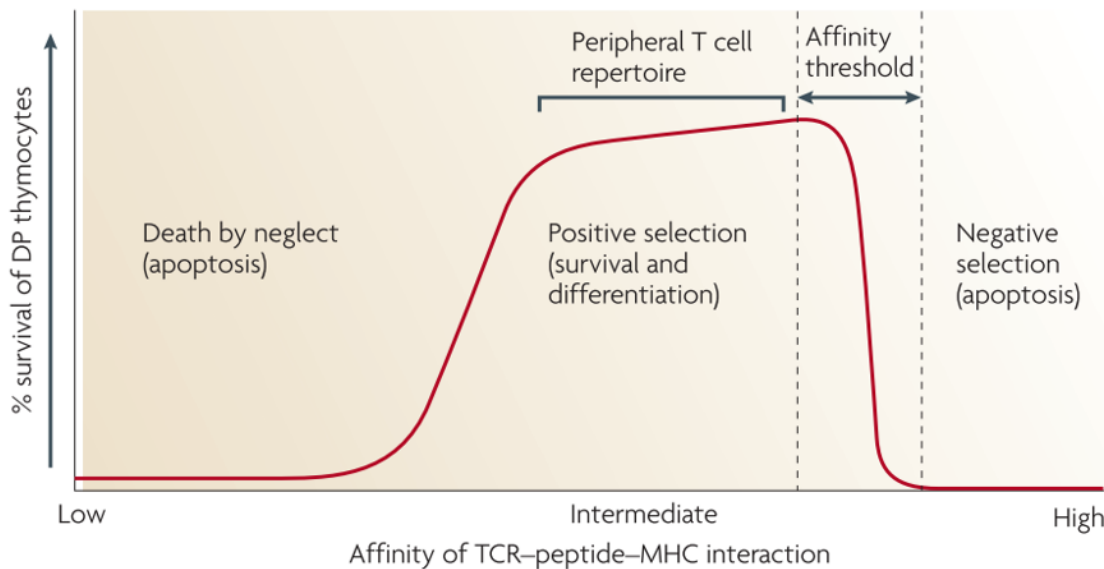
### 1. Le TCR

Le TCR (T-Cell Receptor) est une protéine hétérodimerique membranaire présente à la surface des lymphocytes T et responsable de la reconnaissance des complexes peptide/HLA-II. Le TCR appartient à la famille des immunoglobulines et présente une structure similaire à celle d'un Fab. Le TCR est composé de deux chaînes  $\alpha$  et  $\beta$  comportant chacune un domaine variable (qui porte les 3 CDR) et un domaine constant relié à la membrane <sup>116</sup>. Si l'on fait une analogie avec les anticorps, la chaîne  $\alpha$  correspondrait à la chaîne légère et la chaîne  $\beta$  à la chaîne lourde, comportant un CDR3 hypervariable. La reconnaissance de la molécule HLA est majoritairement portée par les CDR1 et 2 (très peu variables) des deux chaînes et la reconnaissance du peptide est plutôt portée par les CDR3 des deux chaînes <sup>117</sup>. Les domaines constants des TCR possèdent un domaine transmembranaire ainsi qu'un domaine intracellulaire trop court pour la transduction du

signal de liaison du TCR au complexe HLA-II/peptide<sup>118</sup>. La transduction du signal de liaison du TCR au complexe HLA-II/peptide est assurée par le CD3 qui est un complexe composé de 6 chaînes polypeptidiques. Les domaines extracellulaires du CD3 possèdent une affinité pour les parties constantes du TCR et les domaines intracellulaires permettant une transduction du signal via des motifs ITAM (Immunoreceptor tyrosine-based activation motif)<sup>119</sup>.

## 2. La sélection thymique

Les lymphocytes T CD4 sont formés dans la moelle osseuse puis sélectionnés et maturés dans le thymus. La sélection thymique est un processus qui aboutit à la formation du répertoire T qui constitue l'ensemble des lymphocytes T capables de reconnaître un antigène étranger présenté par une molécule HLA. La génération de ce répertoire T passe par deux grands mécanismes décrits en **Figure 10** : la sélection positive (sélection des lymphocytes T ayant une affinité minimale pour une molécule HLA) et la sélection négative (déplétion des lymphocytes T qui reconnaissent un antigène du soi avec une trop forte affinité).



**Figure 10 : Modèle de sélection des thymocytes par affinité.** Selon ce modèle, l'affinité du TCR (T-cell receptor) pour les complexes peptide/CMH-II est un facteur clé de la sélection des lymphocytes T. Les thymocytes qui expriment un TCR avec pas ou peu d'affinité pour les complexes peptide du soi/CHM-II meurent par apoptose, faute de signal de survie. Les thymocytes avec une affinité intermédiaire pour les complexes peptide du soi/CHM-II reçoivent un signal de survie, c'est la sélection positive. Les thymocytes portant des TCR de haute affinité pour les complexes peptide du soi/CHM-II meurent par apoptose, c'est la sélection négative. Adapté de Klein *et al*<sup>120</sup>.

Lors de la sélection positive des lymphocytes T, seuls ceux qui sont capables de lier un complexe HLA/peptide reçoivent un signal de survie (IL-7) de la part des cellules qui présentent ces complexes dans le thymus. Les lymphocytes qui n'expriment pas de TCR fonctionnels

capables de se lier à un complexe peptide/HLA meurent par apoptose. 80 à 90% des lymphocytes ne parviennent pas à passer cette étape <sup>120</sup>.

La sélection négative est le mécanisme principal de la tolérance centrale. Les mTEC (cellules thymiques endothéliales médullaires) sont des cellules spécialisées dans l'expression des protéines du soi <sup>121</sup>. Les mTEC expriment un facteur de transcription particulier, le facteur AutoImmune REgulator (AIRE), qui permet d'activer l'expression de gènes normalement tissu-spécifiques <sup>122</sup>. Ce mécanisme permet d'obtenir une représentation des protéines du soi dans le thymus. Ces antigènes sont ensuite captés par les cellules dendritiques présentes dans le thymus et présentés aux lymphocytes T qui avaient passé la première étape de sélection. Pour compléter le répertoire d'antigènes du soi exprimés par les mTEC, des cellules dendritiques périphériques peuvent capter des antigènes dans la circulation ou dans d'autres tissus et migrer dans le thymus pour les présenter aux LT CD4+ <sup>123</sup>. Les LT CD4+ qui présentent une trop forte affinité pour les complexes HLA-II/peptide provenant des protéines du soi sont éliminés par apoptose, c'est la délétion clonale <sup>124</sup>. Il est également possible qu'un lymphocyte présente une forte affinité pour un peptide du soi mais pas suffisamment pour induire la délétion clonale. Dans ce cas, le lymphocyte peut se différencier en lymphocyte T régulateur (Treg) par un mécanisme appelé sélection agoniste <sup>125</sup>, les Tregs permettent de moduler une éventuelle réponse auto-immune.

## D. Les lymphocytes B et la différenciation en plasmocytes

### 1. Génération des B-cell Receptors (BCR)

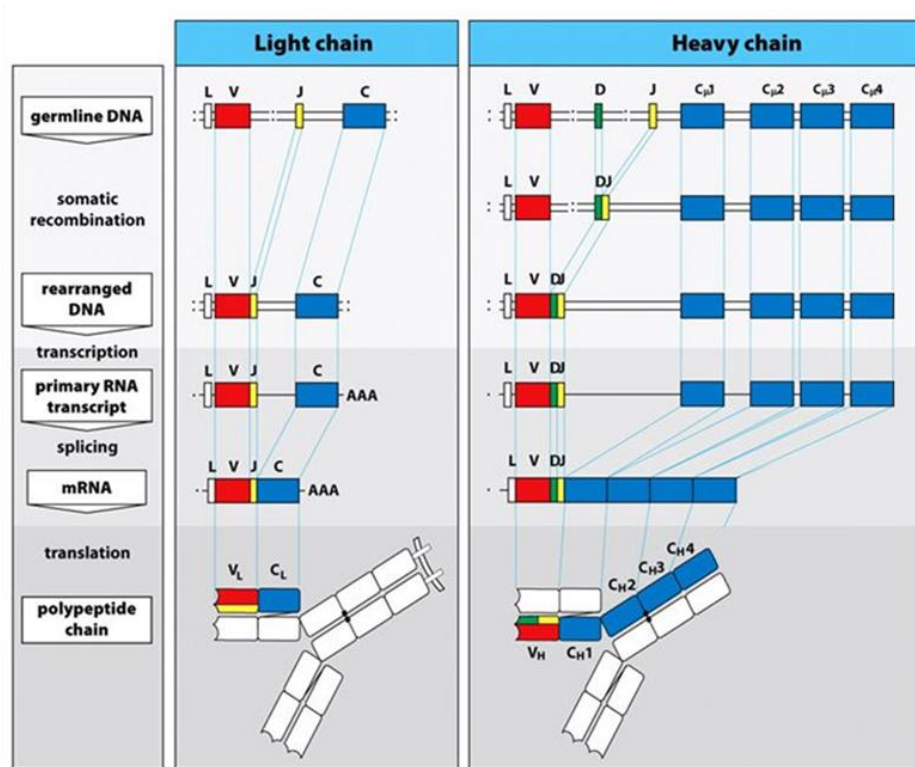
Le BCR est la caractéristique principale des lymphocytes B qui les différencie des lymphocytes T. Le BCR est un anticorps complet avec un domaine membranaire permettant de capter des antigènes spécifiques pour les présenter aux LT via les molécules HLA-II <sup>126</sup>. Les lymphocytes B sont générés dans la moelle osseuse à partir des CLP (common lymphocyte progenitor) et la première étape de différenciation d'un LB aboutit à l'acquisition d'un BCR fonctionnel <sup>127</sup>.

Sur le plan génétique, les régions codant pour les différentes parties d'un anticorps sont réparties en segments pouvant être recombinaés. Il existe 3 familles de gènes chez l'homme : celle codant pour les gènes de la chaîne lourde (chromosome 14), celle de la chaîne légère  $\kappa$  (chromosome 2) et celle de la chaîne légère  $\lambda$  (chromosome 22) <sup>128</sup>. Chaque famille comporte plusieurs gènes correspondant à différents domaines de l'anticorps. Les gènes V codant pour la région FW1-CDR1-FW2-CDR2-FW3-CDR3 (début), les gènes J codant pour le CDR3 (fin)-FW4 et les gènes C pour la partie constante CH1-hinge-CH2-CH3 (chaîne lourde) et C $\kappa$  ou C $\lambda$  (chaîne légère). Dans le cas de la chaîne lourde il existe une 4<sup>ème</sup> famille de gènes, les gènes D qui codent pour la partie centrale du CDR3. Le nombre de gènes par famille est résumé dans le **tableau 2**.



**Tableau 2** : Dénombrement des gènes V, D, J et C par chaîne d'anticorps

Famille	V	D	J	C
Chaîne lourde	123-129	27	9	11
Chaîne légère $\kappa$	76	-	5	1
Chaîne légère $\lambda$	73-74	-	7-11	7-11

Adapté de Nishana *et al.*, 2012

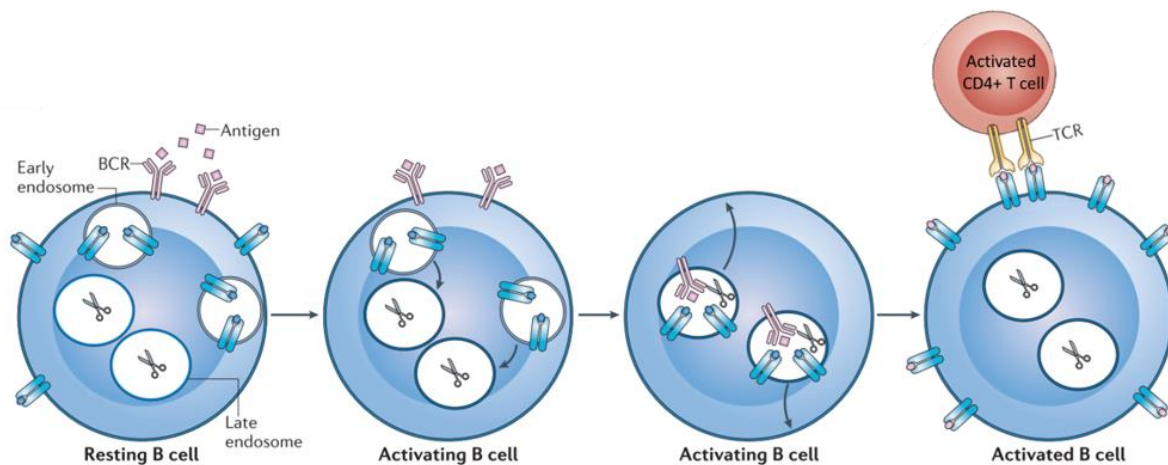
**Figure 11 : Organisation génétique et synthèse des anticorps chez l'homme.** Une recombinaison somatique de l'ADN génomique des lymphocytes B permet de sélectionner un gène V, un gène D (uniquement pour la chaîne lourde) et un gène J pour constituer les domaines variables de l'anticorps. L'épissage alternatif permet de générer un ARN messager codant pour la chaîne lourde (ou légère) complète. L, séquence codant pour le peptide signal. V, séquence codant pour le domaine de variabilité. D, fragment codant pour la séquence de diversité. J, séquence codant pour le domaine de jonction. VL, domaine variable de la chaîne légère. VH, domaine variable de la chaîne lourde. CH1, CH2, CH3, CH4, domaines constants de la chaîne lourde. CL, domaine constant de la chaîne légère. Adapté de Janeway *et al.* <sup>128</sup>.

Pour aboutir à un ARN messenger complet, le génome du lymphocyte B subit plusieurs recombinaisons qui aboutissent à un assemblage aléatoire d'un gène de chaque famille. Par exemple pour la chaîne lourde, un gène V, un gène D, un gène J et un gène C s'assemblent par recombinaison pour ne donner qu'un seul gène, qui après épissage alternatif donnera un ARN messenger complet codant pour la chaîne lourde (**Figure 11**). Le même mécanisme a lieu pour la chaîne légère (sans le gène D). A l'endroit des jonctions V-D-J, des ajouts et des délétions de bases ont lieu grâce à l'enzyme TdT (Terminal desoxyribonucleotidyl Transferase). Ce mécanisme augmente encore la diversité théorique du répertoire B, et en particulier la variabilité du CDRH3.

Les lymphocytes B précoces dont la recombinaison des gènes codant pour les immunoglobulines aboutit à un BCR non fonctionnel (décalage du cadre de lecture, domaine manquant, appariement chaîne lourde/légère défectueux) ne peuvent pas quitter la moelle osseuse et meurent par apoptose. C'est aussi le cas des lymphocytes B auto-réactifs qui reconnaissent des antigènes du soi <sup>127</sup>. En plus de la diversité engendrée par la recombinaison des gènes V, D et J, les anticorps peuvent également subir des mutations somatiques afin d'accroître l'affinité pour leur cible.

## 2. Activation des lymphocytes B

Les cellules B matures naïves possèdent un BCR fonctionnel et normalement non auto-réactif. Ces lymphocytes migrent dans les organes lymphoïdes secondaires comme la rate où de nombreuses cellules présentatrices d'antigène résident. Lorsqu'un lymphocyte B naïf rencontre un antigène pour lequel son BCR est spécifique, celui-ci s'active. Cette activation provoque l'endocytose des BCR complexés avec les antigènes. Ensuite l'antigène est traité et présenté à la surface du lymphocyte grâce aux complexes HLA-II/peptide. S'il existe un lymphocyte T CD4+ préalablement activé par une cellule dendritique ayant auparavant rencontré l'antigène, celui-ci reconnaîtra les mêmes complexes HLA-II/peptide qu'il a déjà rencontrés à la surface des cellules dendritiques (**Figure 12**). Cette reconnaissance à la surface du lymphocyte B par un lymphocyte T effecteur est à l'origine de la cascade de processus qui aboutit à la différenciation du lymphocyte B en plasmocyte sécréteur d'anticorps spécifiques de l'antigène <sup>129</sup>.



**Figure 12 : Activation des lymphocytes B par les lymphocytes T CD4.** Dans les lymphocytes B non activés, la liaison de l'antigène au BCR (B-cell receptor) déclenche l'internalisation des complexes antigène-anticorps. Les complexes sont dégradés dans les compartiments d'apprêtement et les peptides en résultant sont chargés sur des CMH-II et présentés à la surface des lymphocytes B. S'il existe des lymphocytes T CD4 portant des TCR spécifiques des peptides présentés, ceux-ci activent le lymphocyte B correspondant. Adapté de Roche *et al* <sup>86</sup>.

### III. Détection des ADA et données sur les anticorps thérapeutiques

#### A. Les tests de détection des ADA

Il existe 5 méthodes principales pour détecter les ADA suite à l'injection d'un anticorps thérapeutique chez un patient <sup>130,131</sup>.

- L'ELISA/ECLA (Electrochemiluminescence-assay) : Les anticorps du sérum du patient sont capturés sur une membrane et la révélation se fait avec une version marquée de l'anticorps thérapeutique (Bridge). Alternativement la révélation peut se faire en sandwich à l'aide d'un anticorps anti-anticorps humains.
- Radio-ImmunoAssay (RIA) : Les anticorps du sérum sont capturés sur une colonne contenant une résine spécifique des anticorps (Protéine A, Anticorps anti-anticorps). Les ADA capturés sont détectés par l'anticorps thérapeutique radiomarqué.
- Surface plasmon resonance (SPR) : Les ADA présents dans le sérum du patient sont capturés par l'anticorps thérapeutique immobilisé sur un senseur.
- pH-shift anti-Idiotypic Antigen-binding (PIA) : Le sérum du patient est soumis à une baisse de pH, ce qui a pour effet de dissocier les complexes immuns, y compris les complexes ADA-anticorps. Avant la neutralisation, un excès de Fab anti-idiotypic spécifique des parties variables de l'anticorps thérapeutique est ajouté pour bloquer la re-formation des

complexes ADA-anticorps après neutralisation. Après immobilisation sur protéine A et lavage, l'anticorps thérapeutique marqué au format F(ab)<sub>2</sub> (incapable de se lier à la protéine A) est incubé dans la colonne sur la nuit. Après lavage, la mesure du marquage dans la fraction liée à la protéine A reflète le niveau d'ADA présents dans le sérum. <sup>132</sup>

- Homogeneous mobility shift assay (HMSA) : Après une étape de dissociation des complexes immuns à pH acide, l'anticorps thérapeutique fluorescent est ajouté au mélange. Après neutralisation, le mélange est analysé par chromatographie d'exclusion stérique. Les complexes immuns fluorescents sont différenciés de l'anticorps libre fluorescent par leur taille. <sup>133</sup>

Les tests les plus utilisés sont l'ELISA/ECLA et le RIA <sup>130</sup>. Tous les tests de détection des ADA, à l'exception du PIA et de l'HMSA, ont en commun un inconvénient majeur : l'interférence forte de l'anticorps thérapeutique dans le test <sup>131</sup>. En effet, l'anticorps thérapeutique est présent à une forte dose dans le sérum du patient et peut donc interférer de façon importante sur le résultat en créant des faux négatifs. Les deux derniers tests (PIA, HMSA) sont plus récents et ne sont pas encore utilisés en routine pour la détection des ADAs. Ils ont la particularité d'utiliser une étape de dissociation des complexes immuns qui permet de s'affranchir de l'interférence de l'anticorps thérapeutique présent dans le sérum du patient. Ces nouveaux tests sont donc plus sensibles et plus fiables pour la détection des ADAs mais moins utilisés car considérés par les industriels comme moins reproductibles <sup>132,134</sup>.

## B. Résumé des données sur les anticorps thérapeutiques

De nombreux articles documentent l'immunogénicité clinique des anticorps. En particulier, une revue publiée en 2019 <sup>135</sup> recense les données disponibles sur l'immunogénicité des anticorps thérapeutiques disponibles sur le marché (**Tableau 3**). Il est intéressant de noter que le taux de patients qui s'immunisent contre les anticorps thérapeutiques n'est pas forcément corrélé à leur degré d'humanisation. Ainsi, des anticorps humains comme l'adalimumab ou le nivolumab peuvent causer des réponses immunitaires fréquentes chez les patients (respectivement entre 5-26% et 11-38% en fonction des pathologies et des études) alors que certains anticorps chimériques comme le siltuximab (0,2%) ou humanisés comme le bevacizumab (0,6%) présentent un taux d'immunogénicité bien moindre. Ces valeurs sont à prendre avec précaution étant donné que les tests utilisés de détection des ADA peuvent être différents d'une étude à l'autre et sous-estimer la proportion réelle de patients développant des ADA.



**Tableau 3** : Liste des anticorps commercialisés ainsi que des données disponibles sur leur immunogénicité.

Trade Name	INN	Company	Target	Type	FDA Approval	IMG Rate <sup>a</sup>
ReoPro	Abciximab	Eli Lilly	GPIIb/IIIa	Chimeric IgG1 Fab	1994	6%-44%
Humira	Adalimumab	Abbott	TNF- $\alpha$	Human IgG1	2002	5%-26%
Lemtrada	Alemtuzumab	Sanofi	CD-52	Humanized IgG1	2014	1.9%-8.3%
Praluent	Alirocumab	Sanofi	PCSK-9	Human IgG1	2015	1.2%-4.8%
Tecentriq	Atezolizumab	Genentech	PD-L1	Human IgG1	2016	30%-48%
Bavencio	Avelumab	Merck & Co.	PD-L1	Human IgG1	2017	4.1%
Simulect	Basiliximab	Novartis	CD-25	Chimeric IgG1	1998	1.2%
Benlysta	Belimumab	GSK	BLyS	Human IgG1	2011	0%-4.8%
Fasenra	Benralizumab	AstraZeneca	CD-125	Humanized IgG1	2017	13.0%
Avastin	Bevacizumab	Genentech	VEGF	Humanized IgG1	2004	0.6%
Zinplava	Bezlotoxumab	Merck & Co.	C. difficile Toxin B	Human monoclonal antitoxin antibody	2016	0.0%
Adcetris	Brentuximab	Seattle Genetics	CD-30	Chimeric IgG1 ADC	2011	7%-30%
Siliq	Brodalumab	Valeant Pharmaceuticals	IL-17RA	Human IgG2	2017	3.0%
Ilaris	Canakinumab	Novartis	IL-1 $\beta$	Human IgG1	2009	0.0%
Libtayo	Cemiplimab	Regeneron	PD-1	Human IgG	2018	1.3%
Cimzia	Certolizumab pegol	UCB	TNF- $\alpha$	Humanized IgG Fab fragment	2008	6%-23%
Erbix	Cetuximab	Eli Lilly	EGFR	Chimeric IgG1	2004	5.0%
Darzalex	Daratumumab	Janssen	CD-38	Human IgG1	2015	0.0%
Prolia/Xgeva	Denosumab	Amgen	RANKL	Human IgG2	2010	<1%
Dupixent	Dupilumab	Regeneron	IL-4R $\alpha$	Human IgG4	2017	7.0%
Imfinzi	Durvalumab	AstraZeneca	PD-L1	Human IgG1	2017	Unknown
Soliris	Ecilizumab	Alexion Pharmaceuticals	Complement C5	Humanized IgG2/4	2007	1%-3%
Empliciti	Elotuzumab	BMS	SLAMF7	Human IgG1	2015	18.5%
Hemlibra	Emicizumab	Genentech	Factor VIII	Humanized IgG4	2017	Undetected
Aimovig	Erenumab	Amgen	CGRPR	Human IgG2	2018	2.6%-6.2%
Repatha	Evolocumab	Amgen	LDL-C/PCSK9	Human IgG2	2015	0.1%
Ajovy	Fremanezumab	Teva Pharmaceuticals	CGRP	Humanized IgG2	2018	0.4%-1.6%
Emgality	Galcanezumab	Eli Lilly	CGRP	Humanized IgG4	2018	4.8%-12.5%
Mylotarg	Gemtuzumab ozogamicin	Wyeth	CD-33	Humanized IgG4/toxin conjugate	2000	Undetected
Simponi	Golimumab	Janssen	TNF- $\alpha$	Human IgG1	2009	2%-7%
Trogarzo	Ibalizumab	TaiMed Biologics	CD-4	Humanized IgG4	2018	Undetected
Remicade	Infliximab	Janssen	TNF- $\alpha$	Chimeric IgG1	1998	10%-51%
Yervoy	Ipilimumab	BMS	CTLA-4	Human IgG1	2011	1.1%-4.9%
Nucala	Mepolizumab	GSK	IL-5	Human IgG1	2015	6.0%
Poteligeo	Mogamulizumab	Kyowa Kirin	CCR4	Humanized IgG1	2018	3.9%
Lumoxiti	Moxetumomab	AstraZeneca	CD-22	Recombinant immunotoxin	2018	56.0%
Tysabri	Natalizumab	Biogen	VLA-4	Humanized IgG4	2004	9%-10%
Portrazza	Necitumumab	Eli Lilly	EGFR	Human IgG1	2015	4.1%
Opdivo	Nivolumab	BMS	PD-1	Human IgG4	2015	11.2%-37.8%
Gazyvaro	Obinutuzumab	Roche	CD-20	Humanized IgG1	2013	7.0%
Ocrevus	Ocrelizumab	Genentech	CD-20	Humanized IgG1	2017	0.5%-1.5%
Arzerra	Ofatumumab	GSK	CD-20	Human IgG1	2009	Undetected
Lartuvo	Olaratumab	Eli Lilly	PDGFR- $\alpha$	Human IgG1	2016	3.5%
Xolair	Omalizumab	Genentech	IgE	Humanized IgG1	2003	<0.1%
Synagis	Palivizumab	MedImmune	F-protein of RS virus	Humanized IgG1	1998	0.7%
Vectibix	Panitumumab	Amgen	EGFR	Human IgG2	2006	0.8%-4.6%
Keytruda	Pembrolizumab	Merck & Co.	PD-1	Human IgG4	2014	2.0%
Perjeta	Pertuzumab	Roche	HER-2	Humanized IgG1	2012	2.8%
Cyramza	Ramucirumab	Eli Lilly	VEGF	Human IgG1	2014	6.0%
Lucentis	Ranibizumab	Genentech	VEGF-A	Humanized IgG1 Fab fragment	2006	0%-6%
Cinqair	Reslizumab	Teva Pharmaceuticals	IL-5	Human IgG4	2016	4.8%-5.4%
Rituxan	Rituximab	Biogen Idec	CD-20	Chimeric IgG1	1997	1.1%-11%
Cosentyx	Secukinumab	Novartis	IL-17A	Human IgG1	2015	Undetected
Sylvant	Siltuximab	Janssen	cCLB8	Chimeric IgG1	2014	0.2%
Ilumya	Tildrakizumab	Sun Pharma	IL-23	Humanized IgG1	2018	6.5%
RoActemra	Tocilizumab	Chugai	IL-6 receptor	Humanized IgG1	2010	2.0%
Herceptin	Trastuzumab	Genentech	HER-2	Humanized IgG1	1998	0.1%
Stelara	Ustekinumab	Janssen	IL-12/IL-23	Human IgG1	2009	3%-6%
Entyvio	Vedolizumab	Takeda	Integrin- $\alpha$ 4 $\beta$ 7	Humanized IgG1	2014	13.0%

FDA, Food and Drug Administration; INN, international nonproprietary name; IMG, immunogenicity.

<sup>a</sup> Rates of Immunogenicity were derived solely from the FDA licensed package inserts. In cases where IMG rates were not reported, the table was labeled as "unknown." IMG rates listed can be impacted by confounding factors such as comedication, immune status of the patient, and detection assays used.

Adapté de Dingman *et al* <sup>135</sup>.



## IV. Conséquences et facteurs

### A. Conséquences sur le traitement et effets secondaires

Lorsqu'un patient développe une réponse immunitaire contre un anticorps monoclonal, son organisme produit des ADA dirigés contre la molécule thérapeutique. La présence d'ADA peut être sans conséquence mais peut entraîner une altération de la pharmacocinétique et la neutralisation de l'anticorps, avec une perte d'efficacité du traitement pouvant aller jusqu'à des effets secondaires pour le patient et sa mise en danger.

#### 1. Altération de la pharmacocinétique et pharmacodynamique

La demi-vie des anticorps thérapeutiques dans l'organisme peut être perturbée par la présence d'ADA qui accélèrent l'élimination de l'anticorps, on parle alors de CADA (*clearing anti-drug antibodies*). Une étude menée sur 27 patients atteints d'arthrite rhumatoïde traités avec l'infliximab a mis en relation une diminution plus rapide de la concentration d'infliximab circulant avec la présence d'ADA chez les patients en comparaison au groupe qui n'avait pas développé d'ADA <sup>136</sup>.

#### 2. Neutralisation du traitement

Les ADA peuvent reconnaître différentes zones de l'anticorps thérapeutique. Dans la majorité des cas, ce sont les CDR de l'anticorps et donc potentiellement son paratope qui est visé, on parle alors d'anti-idiotypique. Ils peuvent également être neutralisants dans des tests fonctionnels (NADA). Dans ce cas, c'est la capacité de liaison de l'anticorps thérapeutique avec son antigène qui est directement compromise. Des NADA ont été détectés contre l'adalimumab et l'infliximab dans le traitement contre l'arthrite rhumatoïde diminuant l'effet thérapeutique des traitements <sup>137</sup>. Souvent, les patients recevant un traitement anti-TNF $\alpha$  comme l'adalimumab ou l'infliximab et ayant développé des ADA neutralisants se sont vus prescrire une augmentation des doses administrées, augmentant également les risques d'effets secondaires <sup>138</sup>.

Les ADA neutralisants et présents en quantité élevée dans le sang sont à l'origine de résistances aux traitements. Le traitement devient inefficace malgré l'augmentation progressive des doses et la seule solution efficace est le changement de traitement. Par exemple, changer d'anti-TNF $\alpha$  chez les patients atteints d'arthrite rhumatoïde a permis de restaurer l'efficacité du traitement <sup>139</sup>.

#### 3. Risques pour le patient

Il peut exister un risque dû aux anticorps pré-existants. Ce fut le cas du cetuximab pour lequel 22% patients traités présentaient des réactions d'hypersensibilité de type allergique contre l'anticorps. Des études cliniques ont montré qu'il s'agissait en fait d'une réaction allergique

médiée par des IgE pré-existantes dirigées contre une forme de glycosylation particulière présente sur l'anticorps (galactose- $\alpha$ -1,3-galactose)<sup>140</sup>. Des réactions sévères peuvent également avoir lieu lors de l'infusion et peuvent aller jusqu'au choc anaphylactique. C'est le cas de certains patients traités avec l'infliximab et dans la plupart des cas ces réactions sont liées à la présence d'ADAs<sup>141</sup>. Enfin, il existe un certain nombre d'anticorps thérapeutiques qui provoquent une libération de cytokines proinflammatoires pouvant entraîner une dyspnée, une fièvre ou un urticaire. Ce phénomène a été constaté suite à l'injection de différents anticorps comme le rituximab (anti-CD20), l'alemtuzumab (anti-CD52), le muromumab (anti-CD3) et le TGN1412 (anti-CD28)<sup>142-145</sup>.

## B. Les facteurs de l'immunogénicité

L'immunogénicité d'un anticorps thérapeutique, mesurée par l'apparition d'ADA, est influencée par trois catégories de facteurs : les facteurs liés au produit, au traitement ou au patient.

### 1. Facteurs liés au produit

Les facteurs liés au produit sont les plus complexes mais ce sont aussi ceux pour lesquels il est possible d'effectuer des modifications en amont. La séquence en acides aminés de l'anticorps est l'un de ces facteurs. Comme nous l'avons vu précédemment cette séquence peut avoir un impact à plusieurs niveaux :

- La présence de cœurs ou « *core peptides* » présentant une forte affinité pour les molécules HLA-II
- Le degré d'humanisation des peptides présentés par les molécules HLA-II va avoir une importance au niveau de la reconnaissance par les lymphocytes T CD4<sup>+</sup> des peptides présentés. En effet, les lymphocytes T CD4<sup>+</sup> reconnaissant des peptides humains ont été contre-sélectionnés dans le thymus.
- La présence d'épitopes B non contre-sélectionnés dans la moelle osseuse et les organes lymphoïdes, le plus souvent situés au niveau des CDR pour les anticorps thérapeutiques.
- La stabilité de la protéine et sa tendance à agréger va avoir un impact sur sa capacité à activer les cellules dendritiques ainsi que sur l'internalisation de l'anticorps thérapeutique par ces dernières.

Le mode de production, de purification et de stockage de l'anticorps peut également avoir un effet sur la réponse immunitaire. Certains motifs de N-glycosylation sont décrits comme ayant un impact sur l'immunogénicité<sup>140</sup>. Par ailleurs, la présence d'impuretés et de contaminants peut avoir un rôle d'adjuvant<sup>146</sup>. Enfin, les conditions de stockage et de conditionnement peuvent également avoir une influence<sup>147</sup>.



## 2. Facteurs liés au traitement

La fréquence des injections ainsi que la durée du traitement ont un impact important sur le risque d'immunogénicité : un traitement à base de protéines thérapeutique pour une maladie aiguë a moins de probabilité d'engendrer une réponse immunitaire que dans le cas d'un traitement chronique où les injections seront nombreuses et fréquentes <sup>148</sup>. Ce paramètre est d'autant plus important dans le cas des anticorps thérapeutiques puisque les doses injectées au patient sont particulièrement fortes. Les doses importantes augmentent le risque d'internalisation de l'anticorps thérapeutique par les cellules dendritiques et se traduisent souvent par une augmentation du risque d'immunogénicité.

Par ailleurs, tout facteur favorisant un contexte inflammatoire augmente le risque de réponse immunitaire vis à vis de l'anticorps thérapeutique. Ainsi, le mode d'administration de l'anticorps peut avoir un impact significatif. Les facteurs liés au traitement sont ceux où le personnel de santé en charge de la mise en œuvre du traitement a le plus de marge de manœuvre : les doses et les fréquences peuvent être ajustées de façon à réduire le risque d'immunogénicité tout en conservant une bonne efficacité du traitement.

## 3. Facteurs liés au patient

Les facteurs liés au patient regroupent :

- Les facteurs génétiques : en fonction des molécules HLA de classe II exprimées par les cellules du patient, les séquences d'épitopes T reconnues ne seront pas les mêmes et pourront entraîner des réponses très différentes. L'absence de protéine homologue exprimée par le patient ou la présence de mutations dans la séquence de la version endogène de la protéine a un effet sur le répertoire T du patient.
- L'état de maturité du système immunitaire : le système immunitaire présente des stades de développement différents en fonction de l'âge du patient et peut présenter une perte d'efficacité marquée chez les patients âgés (immunosénescence), réduisant ainsi le risque de réponse immunitaire.
- La préexistence de lymphocytes T CD4+ spécifiques de l'anticorps thérapeutique.
- La pathologie traitée : les maladies inflammatoires ou auto-immunes sont un terrain favorable au développement d'une réponse immunitaire dirigée contre l'anticorps thérapeutique. A l'inverse, d'autres pathologies telles que les cancers traités par chimiothérapie et le SIDA pour lesquels le système immunitaire du patient est affaibli, réduisent le risque d'immunogénicité.

- Les antécédents des patients ainsi que la prise de traitements en parallèle peuvent augmenter (ADA développés contre une précédente protéine thérapeutique) ou limiter le risque d'immunogénicité (immunosuppresseurs comme le méthotrexate).

## V. Prédiction de l'immunogénicité

Il existe différentes approches pour l'évaluation préclinique de l'immunogénicité des anticorps à vocation thérapeutique.

Ces méthodes regroupent aussi bien des outils bio-informatiques que des tests biochimiques *in vitro* mais également des tests de prolifération de cellules lymphocytaires humaines. Chacune comporte des avantages et des inconvénients aussi bien en terme de débit, de coût et de fiabilité. Certaines technologies peuvent être utilisées à très haut débit comme la prédiction bio-informatique de la liaison aux molécules HLA-II.

D'autres méthodes font appel à des tests *in vitro* de liaison aux molécules HLA et ne peuvent concerner qu'un nombre plus restreint de molécules HLA. Enfin, les modèles *in vitro* de culture de cellules humaines possèdent de nombreux avantages mais un débit plus réduit. Ces technologies sont complémentaires et permettent de réaliser un arbitrage concernant le choix des candidats anticorps qui vont être développés ou abandonnés.

### A. La prédiction *in silico* de la liaison aux molécules HLA-II

La liaison de très nombreux peptides a été mesurée vis-à-vis de nombreuses molécules HLA-II au cours des 30 dernières années. Il a également été démontré qu'il existe une bonne corrélation entre l'affinité d'un peptide pour une molécule HLA-II et son immunogénicité chez l'homme<sup>149-151</sup>. Les données de ces tests de liaison vis-à-vis de différentes molécules HLA-II ont permis d'alimenter des bases de données consultables sur Internet. Celles-ci ont par la suite servi à générer des algorithmes pour prédire *in silico* la liaison d'un peptide à une molécule HLA-II particulière.

#### 1. Les algorithmes de prédiction de la liaison aux molécules HLA-II

##### i. Les bases de données sur la liaison aux molécules HLA-II

Les principales bases de données de liaison de peptides aux molécules HLA-II sont SYFPEITHI et l'IEDB (The Immune Epitope Database)<sup>152</sup>. La base de données SYFPEITHI n'est plus mise à jour depuis 2012. L'IEDB constitue depuis la ressource principale d'information de liaison entre les peptides et les molécules HLA-II avec plus de 70 000 données de liaison recensées en juillet 2019<sup>153</sup>. D'autres bases de données moins utilisées existent comme MHCBN et AntiJen<sup>152</sup>.

L'IEDB regroupe toutes les données disponibles indépendamment de leur source et de la méthode utilisée pour mesurer l'interaction entre peptide et molécule HLA-II. Ainsi, l'IEDB inclut aussi bien les tests quantitatifs de compétition en ELISA que des mesures (au mieux semi-quantitatives) de liaison mettant en œuvre des cellules entières exprimant des molécules HLA <sup>154</sup>. Pour cette raison, il peut être intéressant de filtrer la base de données en ne conservant que les mesures quantitatives d'interaction peptide/HLA-II utilisant des molécules purifiées pour obtenir des résultats de prédiction plus fiables <sup>152</sup>.

## ii. Les algorithmes de prédiction principaux

Les molécules HLA-II peuvent se lier à des peptides de tailles variables <sup>152,155-157</sup>, le plus souvent entre 13 et 26 résidus <sup>158</sup>. Cette caractéristique constitue une difficulté importante pour la mise en place d'algorithmes fiables et robustes de prédiction de l'affinité des peptides aux molécules HLA-II.

De nombreux algorithmes de prédiction de la liaison aux molécules HLA ont été développés, et 9 sont disponibles sur les plateformes publiques : SVMHC, NetMHCII, NetMHCIIpan, Tepitope/Propred, SYFPEITHI, IEDB\_ARB, IEDB\_Comblib IEDB\_SMM-align, IEDB\_Cons, Rankpep, HLA-DR4pred et EpiToolKit <sup>152</sup>. Ces différents algorithmes ont été comparés par Lin *et al.* <sup>159</sup> et les algorithmes TEPITOPE et NetMHCII/NetMHCIIPan se sont avérés les plus performants.

TEPITOPE est un algorithme développé dans les années 1990 par Sturniolo *et al.* <sup>160</sup>, sur l'idée que chaque molécule HLA-II est constituée de 5 poches de liaison au peptide correspondant aux positions P1, P4, P6, P7 et P9 du cœur ou « *core peptide* » et que ces poches sont indépendantes entre elles.

Ainsi en fonction de la nature de la chaîne latérale des acides aminés à ces positions clé et en tenant compte de la nature biochimique de la poche correspondante, il est possible d'évaluer l'affinité d'un *coeur* donné pour une molécule HLA-II. Sturniolo *et al.* vont encore plus loin dans la définition des poches de liaison : une poche est définie par les acides aminés polymorphes qui la composent et les valeurs d'affinité mesurées pour cette poche sont transposables aux autres poches des autres allèles de molécules HLA-II qui exposent les mêmes acides aminés polymorphes. Ainsi pour chaque poche de liaison, il existe plusieurs profils caractéristiques : 2 pour P1, 11 pour P4, 6 pour P6, 11 pour P7 et 6 pour P9. Par conséquent cette approche permet d'extrapoler les données disponibles pour certains allèles à de nouvelles molécules HLA-II. Ces mesures ont été réalisées grâce à plus de 1000 peptides synthétisés et 10 000 expériences de liaison aux molécules HLA et résumées dans la **Figure 13**. Il est donc possible de calculer un score

			Acide aminé																		
Poche	Allèle	Type	F	W	Y	P	M	I	L	V	A	G	S	T	N	Q	D	E	H	K	R
Poche 1	DRB1*01:01	1	0	0	0	-999	-1	-1	-1	-1	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
	DRB1*13:01	2	-1	-1	-1	-999	0	0	0	0	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
Poche 4	DRB1*01:01	1	0,08	-1,8	-1,1	-1,9	0,8	0,5	0,9	-0,05	0	-0,7	-0,7	-1	0,04	0,1	-2,4	-0,4	-0,7	-2,1	-2,1
	DRB1*15:01	2	2,4	0,4	2,5	-0,3	1	0,6	0,5	0,2	0	0	-0,3	-0,3	-0,2	-0,8	-0,4	-0,6	1,1	-0,7	0,2
	DRB1*03:01	3	-1	-1	0	-1	0	0,5	0	0	0	0,5	0,7	-1	0,2	0	2,3	-1	0	-1	-1
	DRB1*04:01	4	-0,9	-1,2	-1,6	-1,6	0,9	0,8	0,8	-0,9	0	-1,6	0,8	0,7	0,9	0,8	1,4	1,5	1,1	-1,7	-1,9
	DRB1*04:02	5	0,3	1,6	-0,4	-1,3	0,6	0,08	-0,6	-0,7	0	-0,7	-1	-0,5	-0,4	-0,4	-2,3	-2,3	1,2	0,1	1
	DRB1*04:04	6	1	-0,05	-0,4	-1,3	1,8	1,1	1	0,5	0	-2,4	-0,7	-0,9	-0,7	0	1,1	1,1	-1	-1,5	-2,4
	DRB1*11:01	7	0,4	-0,1	-0,7	-1,7	1	0,9	1,1	0,4	0	-1,7	-0,7	-0,6	0	-0,4	-1,7	-1,7	-0,6	-0,5	-0,7
	DRB1*07:01	8	0,2	-1,1	-0,9	-1,2	-0,4	1,1	-0,8	0,9	0	-1,1	1,5	1,4	-1,1	-1,5	-1,6	1,4	0,1	-1,3	-1,1
	DRB1*08:01	9	0,5	0	2,2	-1	1,4	0,3	0,7	0,3	0	-1	-1	-1	0	-1	-1	-1	-1	2,3	2,3
	DRB5*01:01	10	-0,6	-1,4	-0,6	-1,5	1,7	1,3	0,6	1,1	0	-1,6	-0,5	0,3	-1,7	-0,7	-1,9	-1,3	-1,4	-1,7	-1,7
	DRB1*13:02	11	0,8	0,7	0,4	-1,5	0,8	-0,6	0,4	-0,9	0	-1,5	-0,6	-1,1	0,1	0,6	-1,4	-1,1	1,5	0,8	0,2
Poche 6	DRB1*01:01	1	-2,1	-2,4	-2	-0,2	-1,8	-1,9	-2	-1,1	0	-0,3	-0,6	-1,2	-1,1	-1,8	-2,7	2,4	-2,2	-2	-1,8
	DRB1*15:01	2	-0,3	-0,4	0,4	-0,2	0,1	0,05	0,2	-0,3	0	0,5	0,6	-0,04	0,7	-0,8	-0,4	-1	-0,5	-0,3	1
	DRB1*08:01	3	-1,4	-1,4	-1,4	0,5	-0,9	0,7	0,2	1,2	0	-0,7	-0,1	0,8	-0,6	-0,3	-2,4	-1,4	-0,1	1,3	1
	DRB1*04:04	4	-1,1	-1	-1,5	0	-1,1	-0,1	-1,1	0,9	0	-1,5	1	1,9	1,3	-1,5	-1,1	-2,4	-1,4	-2,4	-2,4
	DRB1*07:01	5	-0,8	-0,9	-1	-0,5	-0,8	-0,5	-0,9	0,1	0	-0,6	0,6	-0,08	-0,6	-1,1	-2,5	-2,5	-0,8	-1,1	-1,1
	DRB5*01:01	6	-1,7	-1,7	-1	0,2	-1,5	-1,4	-1	-1,3	0	-0,3	-0,5	-0,8	-1,3	-1,4	-2	-2	-1,2	-1,5	-1,3
Poche 7	DRB1*01:01	1	0,3	-0,08	0,5	0,07	0,09	0,6	0,3	0,7	0	-1,1	-0,2	0,09	0,1	0,2	-2	-0,6	0,1	-0,2	0,09
	DRB1*15:01	2	1,4	0,6	0,7	0,3	1,7	1,5	1,9	0,3	0	0	0,3	0,2	0,7	-0,3	-0,7	-0,7	0,6	-0,3	-0,5
	DRB1*03:01	3	0,5	-0,6	-0,05	0,7	1,1	0,4	0,2	0,2	0	0,1	0,07	-0,1	-0,09	-0,1	-0,6	-0,2	-0,8	-0,9	-0,9
	DRB1*04:01	4	-1	-1,4	-1,2	-0,7	0,8	0,08	0,7	0,08	0	-1,3	-0,2	-0,1	0,6	0	-0,3	0,2	0	-0,3	-1,2
	DRB1*04:02	5	0,5	1,4	0,9	-1	0,8	0,5	1	0,2	0	-2,1	-0,4	0,1	0,6	1,1	2,1	1,2	0,5	0	1,7
	DRB1*04:04	6	-0,05	-0,7	-0,2	-1	0,7	0,08	0,3	-0,1	0	-1,2	0,5	0,4	0,7	-0,2	-1,2	-0,7	-0,4	-1,3	-0,9
	DRB1*11:01	7	-0,4	-0,4	-0,4	-0,05	1,3	0,8	1,5	-0,6	0	-0,4	-1,3	-1,3	-1	-1,1	-2,7	-1,3	-0,2	-0,2	-0,4
	DRB1*07:01	8	2,1	1,4	1,7	-0,2	1,8	2,4	2,2	1,6	0	0	0,4	0,9	1,4	1,1	1,3	0,9	0,9	0,5	0,7
	DRB1*08:01	9	-0,9	-1,2	-1,1	-1,2	-0,3	-0,8	-0,3	-1,1	0	-0,5	-1,3	-2,4	-1,3	-1,2	-2,4	-2,4	-0,9	-0,8	-0,6
	DRB5*01:01	10	1,5	0,4	1,2	-0,6	0,4	1,2	0,6	-0,3	0	0,6	-0,2	0,3	0,5	0,7	-1,5	-0,9	1,2	0,9	1,3
	DRB1*13:02	11	0,6	0,4	-0,2	-0,5	0	-0,5	0,4	-0,1	0	-1,5	-0,9	-0,9	0,1	-0,4	-1,5	-1	0,3	0	1,2
Poche 9	DRB1*01:01	1	-0,4	-1,4	-0,9	-1,1	0,08	0,7	0,5	0,3	0	-0,8	-0,3	-0,2	-1,2	-1,6	-1,9	1,9	-1,1	-1,7	-1
	DRB1*03:01	2	0,9	-1	0,3	-0,3	1,1	0,6	-0,04	0,3	0	0,4	1,1	-0,5	-0,6	-0,2	-0,6	-0,3	-0,5	-0,2	0,5
	DRB1*11:01	3	-1	-1	-1	-1,3	-0,4	-0,3	-1	-0,7	0	-1	0,7	-1,2	-1,4	0,5	-1,7	-1,7	0,08	-0,3	-1
	DRB1*07:01	4	2,1	0,8	1,1	-0,6	2	3,4	3,4	2	0	-0,6	-0,3	-0,4	-0,5	-0,9	-1,2	-0,3	-0,2	-1,1	-0,8
	DRB1*08:01	5	-0,1	-0,1	0,1	-0,9	0,7	-0,1	0	-0,4	0	0,3	0,7	-0,3	0,6	1,3	1	1,3	1,3	-1	-1
	DRB5*01:01	6	1,2	-0,7	1,3	-0,8	0,5	1,2	1,3	-0,2	0	0,4	0,7	-0,2	0	0,7	-1,5	-0,6	1	2,7	2,5

Figure 13 : Base de données des scores de liaison pour chaque poche utilisés par l'algorithme de prédiction TEPITOPE. Adapté de Sturniolo *et al* <sup>160</sup>.

d'affinité pour chaque peptide de 9 résidus et pour chaque molécule HLA présente dans la base en ajoutant les scores de liaison pour chaque poche.

NetMHCII est un algorithme qui fonctionne sur la base du *framework* NN-align <sup>161</sup> qui a été développé en 2009 par Morten Nielsen et qui prédit l'affinité d'un peptide donné pour une molécule HLA-II donnée. NN-align est basé sur un réseau de neurones qui permet d'entraîner un algorithme avec des jeux de données qui contiennent deux paramètres d'entrée : la séquence du peptide et l'allèle de la molécule HLA-II, et une donnée de sortie qui est l'affinité mesurée. La version actuelle de l'algorithme est NetMHCII 2.3 qui a été entraînée sur 134281 couples de données d'entrée réparties sur 36 allèles HLA-DR, 27 HLA-DQ et 9 HLA-DP provenant de la base de données de l'IEDB <sup>157</sup>. Contrairement à TEPITOPE, NetMHCII prédit de façon quantitative la liaison aux molécules HLA-II. Une autre particularité du réseau de neurones NN-align est qu'il prend en compte l'influence des régions flanquantes (résidus situés en dehors du cœur de 9 résidus) du peptide lié à la molécule HLA, mais dont la contribution pour l'affinité a été démontrée <sup>162</sup>, ce qui améliore la prédiction. En plus de prédire l'affinité du peptide, NetMHCII

intègre une prédiction de la position du core au sein du peptide et donc sa modalité de liaison <sup>157</sup>. Avec les données traitées par l'algorithme, il est possible de classer les différents allèles de molécules HLA-II en fonction de leurs motifs de liaison, c'est-à-dire les acides aminés acceptés à chaque position du core (**Figure 14**). Cette représentation permet de mettre en évidence trois choses :

- Les différents loci de molécules HLA (DR/DP/DQ) ont des modalités de liaisons assez différentes. Par exemple, les molécules DR et DP sont plus strictes concernant l'acide aminé en P1 par rapport aux molécules DQ. Les molécules DR et DP n'acceptent pas de résidus acides (D/E) en P9 alors que certaines molécules DQ requièrent ces acides aminés en P9. Les molécules DP ont la particularité de n'accepter pratiquement que des phénylalanines en P1 et en P6 alors que les molécules DR sont plus flexibles sur P6.
- Les motifs de liaison HLA-DP sont moins diversifiés que ceux de DR et DQ.
- HLA-DQ peut être divisé en 3 familles selon la nature des acides aminés acceptés en P7, P8 et P9 : une famille avec une préférence pour les acides aminés chargés positivement, une autre préférant les acides aminés chargés négativement et une dernière avec une préférence pour les petits acides aminés.

Les algorithmes NetMHCII et TEPITOPE existent également en version « Pan », ils sont alors nommés NetMHCIIpan et TEPITOPEpan. Ces versions ont pour particularité d'étendre leurs prédictions aux allèles de molécules HLA qui ne sont pas présentes dans le set d'allèles d'entraînement des algorithmes. Ils proposent donc de réaliser des prédictions de liaison aux molécules HLA pour n'importe quelle molécule HLA dont la séquence protéique est connue en se basant sur la nature des acides aminés polymorphes des différentes poches pour chaque allèle. La différence principale entre NetMHCII et NetMHCIIpan (version 3.2 en Juillet 2019) est qu'il y a autant de réseaux de neurones que d'allèles pour le premier et un seul réseau pour tous les allèles dont la séquence est connue pour le deuxième <sup>157</sup>. TEPITOPEpan étend sa prédiction à 700 allèles de molécules DR mais ne réalise pas de prédiction pour les molécules DP et DQ. Les propriétés des poches de liaison des molécules HLA-DR des allèles non testés expérimentalement sont extrapolées en utilisant les poches déjà connues et la séquence protéique de l'allèle à évaluer <sup>163</sup>.





Figure 14 : Arbre de distance regroupant les allèles de CMH-II selon leurs motifs de liaison. Adapté de Jensen et al.<sup>197</sup>.

### iii. Le choix des sets de molécules HLA pour la prédiction

Il existe des milliers d'allèles de molécules HLA-II différents. Aussi, pour réaliser une prédiction réaliste de la liaison des peptides aux molécules HLA à l'échelle d'une population, des jeux d'allèles contenant un nombre restreint d'allèles représentatifs de la majorité de la population sont souvent utilisés.

Deux critères sont importants dans la constitution d'un panel d'allèles de molécules HLA-II : les fréquences alléliques dans la population et la nécessité de couvrir les grandes familles de répertoires de peptides liés par la molécule HLA-II. Plusieurs jeux d'allèles ont été proposés (**tableau 4**) :

- 27 allèles HLA-DR/DP/DQ <sup>164</sup> : le but de ce panel est de couvrir entre 50 et 75% de la diversité allélique pour chaque locus DR1, DR3/4/5, DP et DQ. Pour cela, les 6 allèles les plus communs des loci DP et DQ ont été choisis ainsi que les 5 plus communs du locus DR3/4/5. Pour le locus DR1, beaucoup plus diversifié, 11 allèles ont été sélectionnés pour une couverture de 50% concernant le locus.
- 46 allèles HLA-DR/DP/DQ <sup>165</sup> : cette collection dérive du jeu de 27 allèles précédents. 19 allèles ont été ajoutés car des cellules les exprimant étaient déjà disponibles pour l'étude.

C'est un jeu qui couvre plus de 90% de la population pour chaque locus et il est composé d'au moins 66% des gènes à chaque locus.

- 15 allèles HLA-DR/DP/DQ <sup>166</sup> : les 13 molécules DR présentes dans ce jeu ont été choisies pour couvrir la majorité des allotypes avec le représentant le plus fréquent par allotype. Seulement 2 allèles DQ sont présents et aucun allèle DP.
- 7 allèles HLA-DR <sup>167</sup> : Dans cette étude les auteurs partent du jeu de 27 allèles et réduisent le nombre d'allèles inclus pour aboutir à un jeu minimal de 7 allèles, qui produit de meilleures prédictions que le jeu de 27 allèles dans la configuration de leurs expériences.

Parmi les jeux d'allèles proposés (voir **tableau 4**), seuls les jeux de 27 et 46 allèles couvrent les 7 « HLA supertypes » mis en évidence par *Greenbaum et al.* En 2018, une étude menée par *Dhanda et al.* n'a pas montré de différence significative de la performance des prédictions obtenues entre les jeux de 27, 15 et 7 allèles <sup>168</sup>.

## 2. Les paramètres de sortie de la prédiction

Lorsqu'on utilise une plateforme web comme l'IEDB pour réaliser des prédictions de liaison aux molécules HLA, les algorithmes de prédiction comme NetMHCII ou TEPITOPE peuvent fournir jusqu'à 3 types de données différentes : IC50, *rank percentile* et core peptide.

L'IC50 est une prédiction de la valeur de l'affinité entre le peptide et la molécule HLA. Cette prédiction n'est disponible qu'avec NetMHCII et NetMHCIIpan.

Le *rank percentile* est une donnée très utilisée dans la prédiction. Ceci permet une évaluation de l'affinité relative par rapport à une large collection de peptides pris au hasard sur les bases de données de protéines comme SWISS-PROT <sup>169</sup>. Ceci permet de s'affranchir des données d'affinité IC50 qui diffèrent grandement entre les molécules HLA, jusqu'à plusieurs ordres de magnitude. Ainsi un *rank percentile* de 10% signifie que le peptide fait partie des 10% meilleurs peptides pour la molécule HLA considérée.

Le *core peptide* : les algorithmes comme NetMHCII et TEPITOPE effectuent également une prédiction du peptide 9-mer qui serait responsable de la liaison à la molécule HLA. La prédiction est un peu différente entre les deux algorithmes : TEPITOPE découpe la séquence de la protéine directement en 9mers <sup>160</sup> et fournit des données directement sur les *core peptides*. NetMHCII découpe la séquence en 15mers (puisqu'il prend en compte l'influence des régions flanquantes) et subdivise ensuite les 15mers en 9mers pour déterminer le core peptide le plus probable <sup>167</sup>.

**Tableau 4 :** Comparaison des différents jeux d'allèles décrits dans la littérature.

	Set 46 allèles	Set 27 allèles	Set 15 allèles	Set 7 allèles	Fréquence génomique	HLA Supertype
	DRB1*0301	DRB1*0301	DRB1*0301	DRB1*0301	7,1	DRB3
	DRB1*0701	DRB1*0701	DRB1*0701	DRB1*0701	7	Main DR
	DRB1*1501	DRB1*1501	DRB1*1501	DRB1*1501	6,3	Main DR
	DRB1*0101	DRB1*0101	DRB1*0101		2,8	Main DR
	DRB1*0401	DRB1*0401	DRB1*0401		2,3	DR4
	DRB1*0405	DRB1*0405	DRB1*0405		3,1	DR4
	DRB1*0802	DRB1*0802	DRB1*0802		2,5	DR4
	DRB1*0901	DRB1*0901	DRB1*0901		3,1	Main DR
	DRB1*1101	DRB1*1101	DRB1*1101		6,1	Main DR
	DRB1*1302	DRB1*1302	DRB1*1302		3,9	DRB3
	DRB1*1201	DRB1*1201			2	Main DR
	DRB1*0404		DRB1*0404		1,9	
	DRB1*0302				1,1	
	DRB1*0402				1,1	
	DRB1*0403				2,3	
	DRB1*0407				2,4	
	DRB1*0411				1,6	
	DRB1*1102				1,1	
	DRB1*1103				0,3	
	DRB1*1104				1,4	
	DRB1*1301				3,2	
	DRB1*1303				1,2	
	DRB1*1304				0,1	
	DRB1*1401				3,4	
	DRB1*1402				2,8	
	DRB1*1601				1	
Total DRB1 génotypique	71,1	46,2	46,1	20,4		
	DRB3*0101	DRB3*0101	DRB3*0101	DRB3*0101	14	DRB3
	DRB4*0101	DRB4*0101	DRB4*0101	DRB4*0101	23,7	DRB3
	DRB5*0101	DRB5*0101	DRB5*0101	DRB5*0101	8,3	Main DR
	DRB3*0202	DRB3*0202		DRB3*0202	18,9	DRB3
	DRB3*0301				6,7	
	DRB5*0102				5,1	
Total DRB3/4/5 génotypique	76,7	64,9	46	64,9		
	DQA1*0501/DQB1*0301	DQA1*0501/DQB1*0301	DQA1*0501/DQB1*0301		19,5	DQ7
	DQA1*0501/DQB1*0201	DQA1*0501/DQB1*0201			5,8	Main DQ
	DQA1*0301/DQB1*0302	DQA1*0301/DQB1*0302			10	Main DQ
	DQA1*0401/DQB1*0402	DQA1*0401/DQB1*0402			6,6	Main DQ
	DQA1*0101/DQB1*0501	DQA1*0101/DQB1*0501			7,6	Main DQ
	DQA1*0102/DQB1*0602	DQA1*0102/DQB1*0602			7,6	DQ7
	DQA1*0201/DQB1*0201				5,7	
	DQA1*0102/DQB1*0502				3,5	
Total DQ génotypique	66,3	57,1	35,1	0		
	DPA1*0201/DPB1*0101	DPA1*0201/DPB1*0101			8,4	Main DP
	DPA1*0103/DPB1*0201	DPA1*0103/DPB1*0201			9,2	DP2
	DPA1*0103/DPB1*0401	DPA1*0103/DPB1*0401			20,1	DP2
	DPA1*0103/DPB1*0402	DPA1*0103/DPB1*0402			23,6	Main DP
	DPA1*0202/DPB1*0501	DPA1*0202/DPB1*0501			11,5	Main DP
	DPA1*0201/DPB1*1401	DPA1*0201/DPB1*1401			3,8	Main DR
Total DP génotypique	76,6	76,6	0	0		
Total HLA Supertype	7/7	7/7	4/7	2/7		

La somme des fréquences génotypiques est indiquée pour chaque groupe de molécules HLA-II.

L'appartenance à un HLA « supertype » tel que défini par Greenbaum *et al.* est défini pour chaque allèle. Le nombre de *supertypes* représentés est indiqué par jeu d'allèle.



## B. *In vitro*

En plus des algorithmes de prédiction *in silico*, il existe différents types de tests biochimiques et cellulaires pour localiser les potentiels épitopes T ou encore évaluer le potentiel immunogène d'une protéine.

### 1. Tests biochimiques

Les tests biochimiques cherchent à évaluer expérimentalement la capacité de liaison des peptides composant une protéine à lier une ou plusieurs molécules HLA-II.

#### i. Le test de liaison (ELISA)

Le premier test est le test de liaison : il s'agit de mesurer par ELISA l'affinité des peptides pour les molécules HLA-II. Ce sont les données de ces tests qui alimentent les algorithmes de prédiction comme NetMHCII.

#### ii. Le MAPPS

Le MAPPS (MHC-Associated Peptides ProteomicS) est une technologie récente de protéomique appliquée à l'étude des peptides liés aux molécules HLA <sup>170-172</sup>. La technique consiste à éluer les peptides présentés par des molécules HLA-II présentes à la surface de cellules présentatrices comme les cellules dendritiques. Une analyse de LC-MS (Liquid chromatography – Mass Spectrometry) est effectuée pour identifier les peptides élués. Contrairement aux tests de liaison et aux algorithmes de prédiction qui ne donnent des informations que sur une étape de la présentation des antigènes, le MAPPS prend en plus en compte les étapes d'internalisation et de traitement de l'antigène. Cependant, cette méthode ne permet pas d'accéder à l'affinité des peptides pour les molécules HLA et sa sensibilité peut être réduite à cause de la présence de peptides autres que ceux issus de la protéine à étudier et provenant d'une protéine présente en grande quantité (CLIP, protéines du milieu, de la cellule...).

### 2. Test de stimulation des lymphocytes T

Le test *in vitro* reconnu comme étant le plus prédictif de l'immunogénicité potentielle d'une protéine thérapeutique est le test de stimulation des lymphocytes T. Contrairement aux techniques précédentes qui peuvent être réalisées à très haut débit (*in silico*) ou à moyen débit (test de liaison, MAPPS), le test de stimulation des lymphocytes T est un test qui est réalisé à bas débit et donc plutôt destiné à comparer une ou quelques protéines thérapeutiques d'intérêt. Il est toutefois le seul à effectivement évaluer la réponse des cellules T, les autres approches n'étant basées que sur l'interaction des peptides avec les molécules HLA et non avec les TCR. En revanche l'activation de lymphocytes B spécifiques de la protéine thérapeutique n'est pas étudiée dans ce test.

Ce test consiste à détecter la présence de lymphocytes T CD4 dans le sang de donneurs sains (avant toute injection du produit chez l'homme) puis à stimuler des lymphocytes T de donneurs sains avec des CPA autologues préalablement mises en culture avec la protéine thérapeutique. Une phase d'amplification des lymphocytes T spécifiques permet d'amplifier le signal. En caractérisant l'activation des lymphocytes T (marqueurs de surface, sécrétion de cytokines, prolifération), il est possible d'évaluer le potentiel immunogénique de la protéine chez ce donneur.

Concrètement, l'évaluation de l'activation de lymphocytes T spécifiques est réalisée en comparant l'expression de cytokines spécifiques de l'activation (INF $\gamma$ ) ou de prolifération (IL-2) des lymphocytes stimulés ou non avec la protéine thérapeutique. Les expériences de « T-cell assay » diffèrent d'un laboratoire à l'autre par la durée d'amplification du signal (entre 1 et 3 semaines) qui influe sur la sensibilité du test, les conditions de culture (types de cytokines utilisées) ainsi que par la méthode de révélation (ELISpot, Fluorospot, cytométrie en flux, ELISA). Il existe quatre principaux protocoles de T-cell assays dont les conditions sont résumées dans le **tableau 5**. La différence principale entre ces tests (sans tenir compte des méthodes de révélation utilisées) est que les tests effectués par Antitope, Lonza et Proimmune sont réalisés sur des durées d'amplification courtes permettant de détecter un signal seulement pour les protéines fortement immunogènes. Seul le protocole du CEA avec 21 jours d'amplification permet d'amplifier suffisamment les lymphocytes T CD4 spécifiques ayant une fréquence très faible dans le sang (cas typique pour les anticorps thérapeutiques). Cela permet d'estimer le nombre de lymphocytes T CD4+ spécifiques de la protéine thérapeutique circulant dans le sang des donneurs sains.

**Tableau 5** : Les conditions utilisées par les différents laboratoires pour réaliser des tests de stimulation des lymphocytes T

Laboratoire	Culture	Révélation
Antitope	5 jours	Prolifération : thymidine tritiée + sécrétion d'IL-2 par ELISpot
CEA	21 jours : stimulation hebdomadaire avec des DC autologues chargés	Activation : sécrétion d'INF $\gamma$ par ELISpot
Lonza	Durée inconnue	Prolifération : cytométrie Activation : sécrétion d'INF $\gamma$ par FluoroSpot
Proimmune	7 jours	Prolifération : cytométrie

## VI. Dé-immunisation de protéines thérapeutiques

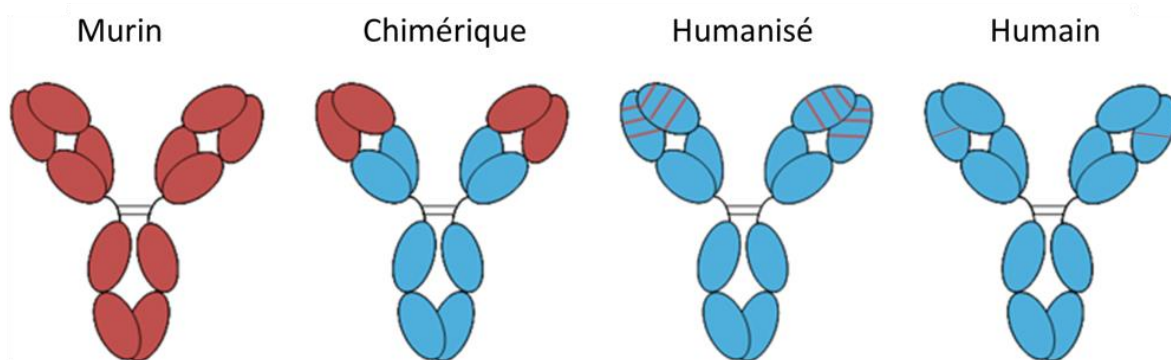
Afin de réduire le risque d'immunogénicité des anticorps à vocation thérapeutique, différentes méthodes de dé-immunisation ont été élaborées. Celles-ci cherchent à réduire l'activation des lymphocytes T ou B en intervenant sur différents paramètres jouant un rôle lors de différentes étapes clés de la réponse immunitaire :

- la maturation des cellules dendritiques
- la présentation de peptides immunogènes via les molécules HLA-II
- la reconnaissance des peptides par les lymphocytes T CD4+
- la reconnaissance de l'anticorps thérapeutique par des BCR de lymphocytes B naïfs

Pour les anticorps thérapeutiques, la chimérisation puis l'humanisation des séquences se sont d'abord imposées pour réduire les risques de réponse vis-à-vis d'anticorps n'étant pas d'origine humaine. Aujourd'hui encore, de nombreux acteurs de l'industrie biopharmaceutique et des laboratoires académiques travaillent sur l'humanisation, voir l'ultra-humanisation (humanisation des CDR) des anticorps afin de réduire les risques d'immunogénicité <sup>173</sup>.

### A. Humanisation

#### 1. Historique



**Figure 15 : Les différents degrés d'humanisation des anticorps.** Les domaines murins sont représentés en rouge, les domaines humains en bleu. Dans la version humanisée le caractère murin des 3 CDR de chaque chaîne est matérialisé en rouge. Dans la version humaine, le CDRH3 est représenté en rouge, sa séquence n'étant pas présente dans les lignées germinales humaines.

Adapté de : <https://absoluteantibody.com/antibody-resources/antibody-engineering/humanisation/>

Préalablement à l'humanisation des anticorps (**Figure 15**), la chimérisation des anticorps est une approche qui a été développée dans les années 1980. Celle-ci consiste à fusionner les parties variables d'un anticorps murin avec les parties constantes d'un anticorps humain pour

aboutir à une molécule dont la séquence est humaine à 70% environ <sup>174</sup>. Les anticorps chimériques conservent bien la fonctionnalité de l'anticorps parental et voient leur immunogénicité bien diminuée mais ils continuent toutefois de provoquer des réponses immunitaires dirigées contre les parties variables murines de l'anticorps <sup>175</sup>.

Dans la seconde moitié des années 1980, l'équipe de Greg Winter a imaginé une méthode appelée « CDR-grafting ». Celle-ci consiste à greffer les CDR de l'anticorps donneur murin sur un anticorps accepteur humain afin de diminuer encore la présence de séquences murines dans l'anticorps <sup>176</sup>. Après avoir réalisé la preuve de concept sur un anticorps murin anti-haptène, la technique s'est généralisée pour devenir le processus principal d'humanisation des anticorps murins. A partir des années 1990, d'autres méthodes d'humanisation ont été développées comme le « resurfacing » <sup>177</sup>, la super-humanisation <sup>178</sup> et l'optimisation du « Human String Content » <sup>179</sup>. Malgré une diminution de l'immunogénicité entre les anticorps chimériques et humanisés, ces derniers provoquent toujours une réponse immunitaire pour une fraction variable de patients en fonction de la maladie et de l'anticorps. Il a été montré que cette réponse est exclusivement dirigée contre les CDR de ces anticorps humanisés <sup>180</sup>.

En parallèle du développement des méthodes d'humanisation, d'autres méthodes de génération d'anticorps complètement humains (dits « *fully human* ») se sont développées. Parmi celles-ci, on peut noter l'utilisation du phage display permettant le criblage de banques naïves d'anticorps issus de lymphocytes B humains <sup>181</sup> et l'utilisation de souris transgéniques porteuses d'un système immunitaire humanisé <sup>182</sup>. Cependant, ces anticorps comportent des mutations dans les CDR par rapport aux lignées germinales des anticorps humains et sont, au même titre que les anticorps humanisés, immunogènes chez une fraction de patients pour lesquels la réponse immunitaire est également dirigée contre ces CDR mutés <sup>180</sup>.

## 2. CDR grafting

Le *CDR grafting* est la méthode la plus commune pour humaniser les anticorps murins. Le processus de *CDR grafting* comporte trois étapes clés :

- 1) la définition des régions qui déterminent la spécificité de l'anticorps, autrement dit les limites des CDR.
- 2) le choix de l'anticorps humain accepteur.
- 3) le choix des *back mutations* du résidu humain vers le résidu non humain dans les régions *frameworks* qui sont parfois nécessaires pour restaurer l'affinité de l'anticorps parental après le *grafting*.

La définition des CDR des anticorps est susceptible de varier en fonction de la méthode et des conventions utilisées. Il existe différentes méthodes de numérotation des résidus d'un anticorps et

autant de méthodes de définition des CDR, les principales étant Kabat<sup>183</sup>, Chothia<sup>78</sup> et IMGT<sup>184</sup>. Ainsi, la définition de la région déterminant la spécificité de l'anticorps est un élément central dans les approches de CDR-grafting.

Afin de réduire le nombre de résidus non humains greffés, différentes approches ont été mises en place comme le SDR-grafting (Specificity Determining Residues) et l'aCDR-grafting (abbreviated CDR)<sup>185</sup> et le SDRU-grafting (SDR usage)<sup>186</sup>. La détermination de la structure du complexe antigène/anticorps permet de définir les SDR qui sont les résidus directement en contact avec l'antigène. Les aCDR sont des CDR écourtés de telle sorte à ce qu'ils incluent l'ensemble des SDR. Les SDRU sont les SDR les plus souvent constatés dans les structures de complexes d'anticorps et leur détermination ne nécessite pas de connaître la structure du complexe *a priori*.

Concernant l'anticorps humain accepteur, plusieurs considérations peuvent participer au choix d'une structure la plus adaptée à la projection des CDR dans une conformation fonctionnelle<sup>187</sup> :

- *Fixed FR* : Les premiers cas d'humanisation ont été réalisés sur des *frameworks* d'anticorps dont les structures étaient connues sans tenir compte du degrés d'homologie entre l'anticorps donneur et l'anticorps accepteur.
- *Best fit* : L'anticorps humain possédant le plus d'homologie avec l'anticorps murin accepteur est sélectionné.
- *Germline* : Seules les lignées germinales d'anticorps (et non des séquences d'anticorps matures) sont utilisées pour constituer les *frameworks*, comme le gène IGHV3-23 pour la chaîne lourde et IGVK3-20 pour la chaîne légère  $\kappa$ <sup>188</sup>.
- *Consensus* : Il existe quelques *frameworks* consensus qui dérivent de germlines d'anticorps humains connus pour leurs bonnes propriétés biophysiques (expression, stabilité, absence d'agrégation).

Une fois le grafting réalisé, il est fréquent que l'anticorps humanisé obtenu soit moins affiné pour sa cible que l'anticorps parental. Dans certains cas, l'anticorps peut même avoir perdu toute capacité de reconnaissance de l'antigène.

Pour cette raison, une dernière étape est souvent nécessaire lors de l'humanisation d'un anticorps. Les pertes d'affinité sont souvent liées au fait qu'un ou plusieurs résidus murins humanisés lors du CDR *grafting* étaient requis pour l'interaction. Il s'agit alors de trouver quelle(s) position(s) sont concernées afin de les faire revenir au résidu murin initial (*back-mutation*). Il existe des positions connues comme étant importantes pour la fonctionnalité d'un anticorps comme les résidus canoniques et les *vernier zone residues*. Une analyse structurale peut être également possible pour identifier ces positions en utilisant une modélisation de l'anticorps ou la structure tridimensionnelle lorsqu'elle est disponible. En l'absence de données structurales, Haidar *et al.* ont répertorié la position de toutes les back-mutations murines qui ont été réalisées

lors des précédentes humanisations décrites dans la littérature. Ceci peut permettre de mieux prévoir les positions qu'il faudra éventuellement muter <sup>188</sup>.

### 3. Resurfacing

Le resurfacing est une méthode qui a émergé au début des années 1990 comme une alternative à la méthode du CDR grafting. L'idée est de supprimer les épitopes B potentiels en humanisant seulement les acides aminés de surface de l'anticorps <sup>189</sup>. Plusieurs anticorps ont été humanisés de cette façon mais en l'absence de données cliniques il n'est pas possible d'évaluer l'impact de cette méthode sur l'immunogénicité des anticorps <sup>187</sup>.

### 4. Human string content (HSC) optimization

Lazar *et al.* ont développé une nouvelle méthode de mesure du caractère humain d'un anticorps en se basant sur le caractère humain de la séquence sur une longueur de 9 résidus consécutifs. En effet, l'immunogénicité des anticorps thérapeutiques est liée à la présence d'épitopes T dont le cœur possède une longueur de 9 résidus, car présentés par les molécules HLA-II. Cette mesure, le HSC (exprimé en %), consiste à compter parmi tous les 9mers possible d'une séquence d'anticorps, combien sont retrouvés dans au moins une lignée germinale humaine <sup>179</sup>. Avec ce raisonnement, les 9mers humains présents dans la séquence de l'anticorps ne sont pas censés être reconnus par un TCR de lymphocyte T CD4+. Ainsi, plus il y a de 9mers humains, moins il y a de chance de constater une réponse immunitaire dirigée contre l'anticorps. Dans cette méthode, lors du CDR grafting, les *frameworks* humains accepteurs ne sont plus choisis par rapport au pourcentage d'identité avec la séquence murine mais par rapport au HSC de toutes les séquences de lignées germinales. Ainsi en choisissant les bonnes combinaisons *framework* humain / CDR murin, il est possible d'obtenir des jonctions *framework*/CDR « plus humaines » (au sens du HSC) qu'en utilisant la méthode du *best fit*.

## B. Dé-immunisation

Certains anticorps, comme l'adalimumab, un anti TNF $\alpha$  utilisé pour le traitement de plusieurs maladies inflammatoires comme la maladie de Crohn ou l'arthrite rhumatoïde, sont décrits comme étant très immunogènes chez les patients alors qu'il s'agit d'anticorps humains (fully human au sens de l'INN, International Nonproprietary Names).

Aussi, d'autres méthodes peuvent être envisagées pour réduire l'immunogénicité des anticorps comme la suppression des épitopes T et B consistant à modifier les séquences potentiellement immunogènes pour qu'elles ne soient plus reconnues par le système immunitaire des patients. Par ailleurs, l'amélioration des propriétés biophysiques de l'anticorps peut également être envisagée afin de ne pas déclencher la maturation des cellules dendritiques à cause de l'inflammation causée par d'éventuels agrégats.

## 1. Suppression des épitopes T

Le T-cell epitope removal (ou suppression des épitopes T) consiste à muter la protéine thérapeutique d'intérêt pour diminuer l'affinité des peptides issus de celle-ci pour les molécules HLA-II. Ceci a pour objectif d'empêcher leur présentation par les cellules présentatrices d'antigènes et éviter l'activation de lymphocytes T spécifiques.

### i. Les exemples à bas débit (*alanine scanning*)

Le premier exemple de dé-immunisation utilisant le *T-cell epitope removal* est celui de la staphylokinase en 2002 par l'équipe de Marc De Maeyer<sup>190</sup>. L'équipe a identifié une région de 17 acides aminés de la protéine bactérienne qui induisait une réponse T chez 90% des donneurs testés. Deux coeurs potentiels ont été identifiés au sein de cette séquence par des méthodes de modélisation *in silico* de la liaison aux principales molécules HLA-II. Des mutations ont été effectuées sur les positions d'ancrage du peptide aux molécules HLA en remplaçant les acides aminés concernés par une alanine. Des tests de stimulation des lymphocytes T montrent que la région immunogène mutée n'induit plus de réponse détectable. Cependant les tests sur protéine entière ne montrent pas de différence significative entre la protéine originale et la protéine mutée. Les auteurs proposent que d'autres épitopes T doivent être éliminés pour pouvoir constater une diminution significative de la réponse. Dans cet exemple, la caractérisation fonctionnelle de la protéine après mutation n'a pas été effectuée.

Un travail a également été mené sur la dé-immunisation des toxines bactériennes depuis 2012 par l'équipe d'Ira Pastan. L'équipe a appliqué la méthode du *T-cell epitope removal* sur une toxine bactérienne à intérêt thérapeutique dans le traitement du cancer : l'exotoxine A de *Pseudomonas aeruginosa*<sup>191,192</sup>. L'identification des épitopes T potentiels a été réalisée à l'aide des outils de prédiction de liaison aux molécules HLA-II disponibles sur l'IEDB. Ensuite les six épitopes T dominants ont été soumis à un *T-cell assay* pour confirmer leur immunogénicité et évaluer les positions qui diminuent la réponse T une fois mutées en alanine. Les six épitopes T ont été supprimés soit par l'élimination d'un domaine de la protéine qui n'était pas requis pour sa fonctionnalité, soit par des mutations en alanine qui n'affectaient pas la fonctionnalité de la toxine. Cependant, aucun test d'évaluation de l'immunogénicité n'a été réalisé sur la protéine dé-immunisée comparée à la protéine native, il s'agit seulement de tests sur des peptides qui ne sont pas représentatifs de l'immunogénicité de la protéine entière.

Ces deux exemples utilisent la méthode illustrée en **Figure 16A**<sup>193</sup>.

### ii. Les exemples à haut débit (*criblage in silico ou in vitro*)

Une des difficultés majeures lors du processus de *T-cell epitope removal* est le contrôle de la fonctionnalité de la protéine mutée. La suppression des épitopes T requiert souvent de muter



des résidus importants pour la fonctionnalité de la protéine. Deux groupes se sont intéressés à ce problème en utilisant des techniques permettant de contrôler la fonctionnalité des protéines mutées au cours du processus de dé-immunisation : le groupe de Bailey-Kellogg et celui de George Georgiou. Le premier utilise la modélisation moléculaire *in silico* pour contrôler la bonne fonctionnalité de la protéine et le second utilise la cytométrie en flux.

Le groupe de Bailey-Kellogg a mis au point un logiciel appelé EpiSweep<sup>194,195</sup> qui prend en compte deux paramètres pour proposer des mutations de dé-immunisation : les données des algorithmes de prédiction de la liaison aux molécules HLA-II et l'impact des mutations sur la structure de la protéine (**Figure 16B**). Cette méthode nécessite de connaître la structure de la protéine à dé-immuniser. Elle a été appliquée pour la dé-immunisation d'une protéine, la  $\beta$ -lactamase d'*Enterobacter cloacae*<sup>196</sup>. Les épitopes T ont été prédits grâce à une version de l'algorithme TEPITOPE appelée ProPred. Ensuite l'algorithme de dé-immunisation DP<sup>2</sup> (maintenant connu sous le nom d'EpiSweep) a proposé des mutations au sein des *cores* de liaison des peptides aux molécules HLA-II en tenant compte de la structure de la protéine. Les auteurs se sont d'abord assurés de la fonctionnalité et de la stabilité de l'enzyme après mutation. Des expériences de liaison avec quatre molécules HLA-II différentes ont été ensuite réalisées pour évaluer la perte d'affinité des peptides mutés pour les molécules HLA-II. Globalement, l'affinité des peptides mutés pour les quatre molécules HLA-II testées est plus faible que celle des peptides natifs. Certains peptides mutés ont vu leur affinité pour les molécules HLA-II augmenter, notamment quand des résidus mutés hydrophiles sont remplacés par des alanines. En revanche, ni l'immunogénicité des peptides mutés ni celle de l'enzyme dé-immunisée n'ont été évaluées par *T-cell assay*.

Enfin, l'équipe de George Georgiou a mis au point une méthode de contrôle de la fonctionnalité par cytométrie en flux qui a l'avantage d'être une mesure directe de la fonctionnalité des mutants et non une prédiction. La protéine dé-immunisée selon cette méthode (**Figure 16C**) est une L-asparaginase d'origine bactérienne indiquée dans le traitement de la leucémie lymphoblastique et contre laquelle jusqu'à 60% des patients s'immunisent<sup>197</sup>. Les épitopes T de l'enzyme ont été prédits grâce à une méthode de prédiction pour 8 allèles DRB1 en incluant DRB1\*04:01, un allèle fortement associé à l'apparition de la maladie chez les jeunes garçons. Trois épitopes ont été sélectionnés et pour chacun les positions P1, P4, P6 et P9 ont été rendues aléatoires via l'utilisation d'un codon dégénéré NNS qui permet l'introduction de chacun des 20 acides aminés aux positions sélectionnées. Ainsi, à l'aide d'un rapporteur fluorescent GFP, les auteurs ont pu évaluer l'activité des enzymes mutées et trier les bactéries exprimant une L-asparaginase fonctionnelle par cytométrie en flux. Les mutants actifs et présentant les scores de prédiction les plus favorables ont été sélectionnés. L'impact des mutations sur l'immunogénicité a été testé grâce à des souris transgéniques porteuses d'un système immunitaire humain et homozygotes pour l'allèle DRB1\*04:01. Bien que l'évaluation de l'immunogénicité sur des

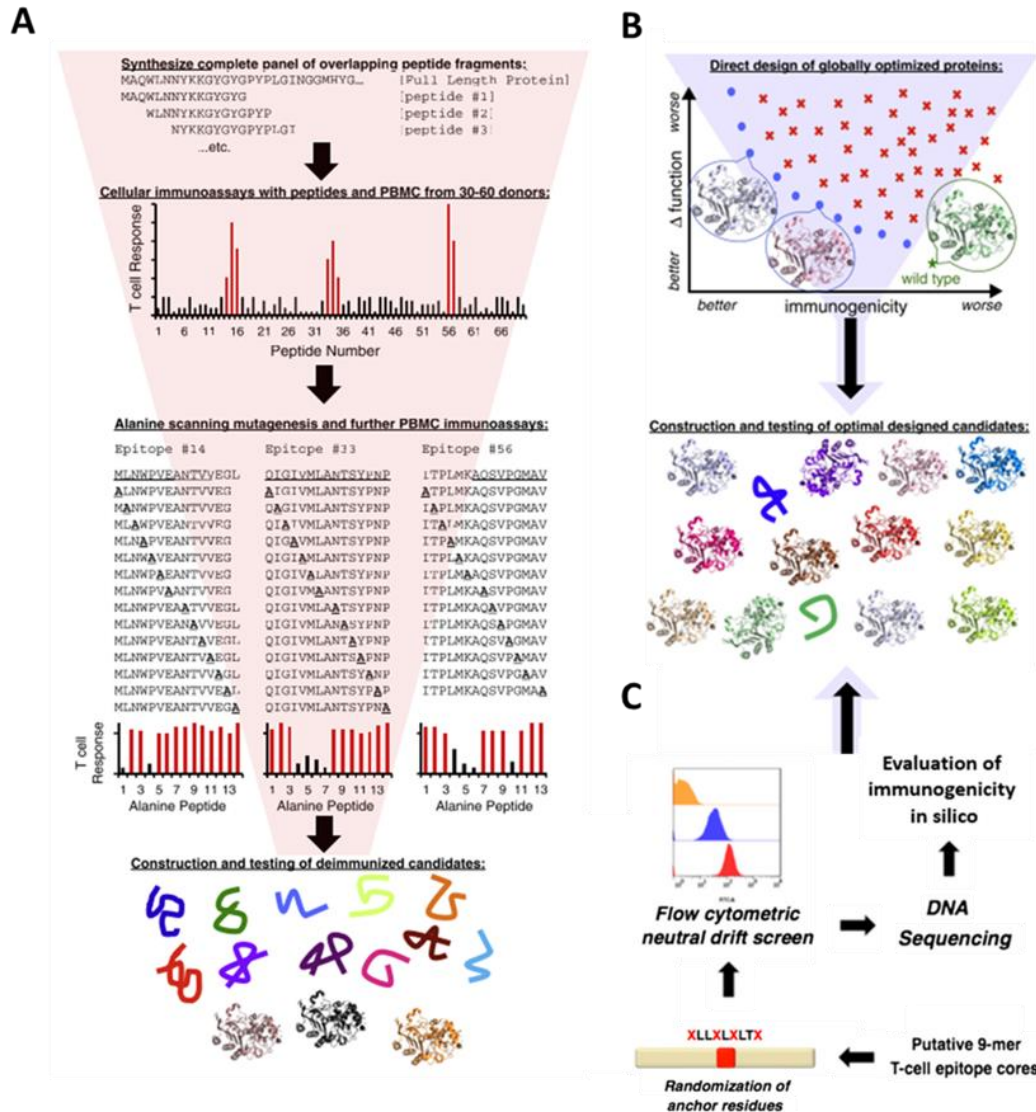


modèles animaux ne soit pas prédictive de l'immunogénicité chez l'homme <sup>198</sup>, les auteurs ont constaté une baisse de l'activation des lymphocytes T en présence des peptides mutés et une forte diminution (90%) du taux d'anticorps dirigés contre l'enzyme dans le sérum des souris.

## 2. *B-cell epitope removal*

Le principe du *B-cell epitope removal* est de modifier des acides aminés de surface de la protéine thérapeutique pour qu'elle ne puisse pas être reconnue par le BCR d'un lymphocyte B. L'identification de tels épitopes et surtout leur prédiction est extrêmement difficile car le répertoire naïf de lymphocytes B n'est pas connu et dépend de chaque individu. Il n'existe pas à ce jour d'exemple de dé-immunisation de protéine thérapeutique utilisant cette méthode où il a été montré une diminution de la réponse immunitaire due à l'absence d'épitope B.

Les méthodes de dé-immunisation des anticorps thérapeutiques impliquent de modifier la séquence de ces derniers. Il peut donc être judicieux de monitorer la fonctionnalité de l'anticorps tout au long du processus de modification de sa séquence primaire. Différentes technologies comme les techniques de display permettent de cribler un nombre important de variants d'anticorps afin de sélectionner les variants intéressants pour la propriété que l'on souhaite améliorer comme par exemple l'immunogénicité tout en conservant sa fonctionnalité.



**Figure 16 : Les différentes méthodes de dé-immunisation décrites dans la littérature. (A)** Une collection de peptides chevauchants sur l'intégralité de la séquence est synthétisée. Chaque peptide fait l'objet d'un test de prolifération des lymphocytes T. Les peptides immunogènes sont soumis à un alanine scanning et les peptides mutés sont également évalués pour leur immunogénicité *in vitro*. Les peptides mutés présentant les réponses les plus faibles au test de stimulation sont sélectionnés et remplacés dans la séquence entière de la protéine. Plusieurs variants dé-immunisés de la protéine sont synthétisés et évalués pour leurs propriétés (activité, stabilité, expression...). **(B)** La prédiction des épitopes T est faite *in silico*. D'autre part l'effet de nombreuses mutations sur la fonctionnalité de la protéine est évalué *in silico*. Les mutations conciliant fonctionnalité et réduction du contenu en épitopes T sont sélectionnés. Les protéines sont ensuite synthétisées et évaluées pour leurs propriétés. **(C)** Les positions d'ancrage des cœurs des épitopes T prédits sont randomisées par mutagenèse dirigée. Les variants fonctionnels sont isolés par cytométrie en flux puis séquencés. Le contenu en épitope T des séquences fonctionnelles mutées est réévalué *in silico*. Les meilleures séquences fonctionnelles et dé-immunisées sont sélectionnées. Adapté de Cantor *et al*<sup>197</sup> et Griswold *et al*<sup>193</sup>.

## Chapitre 4 : Les méthodes de *display* pour l'ingénierie et l'étude des interactions antigène-anticorps

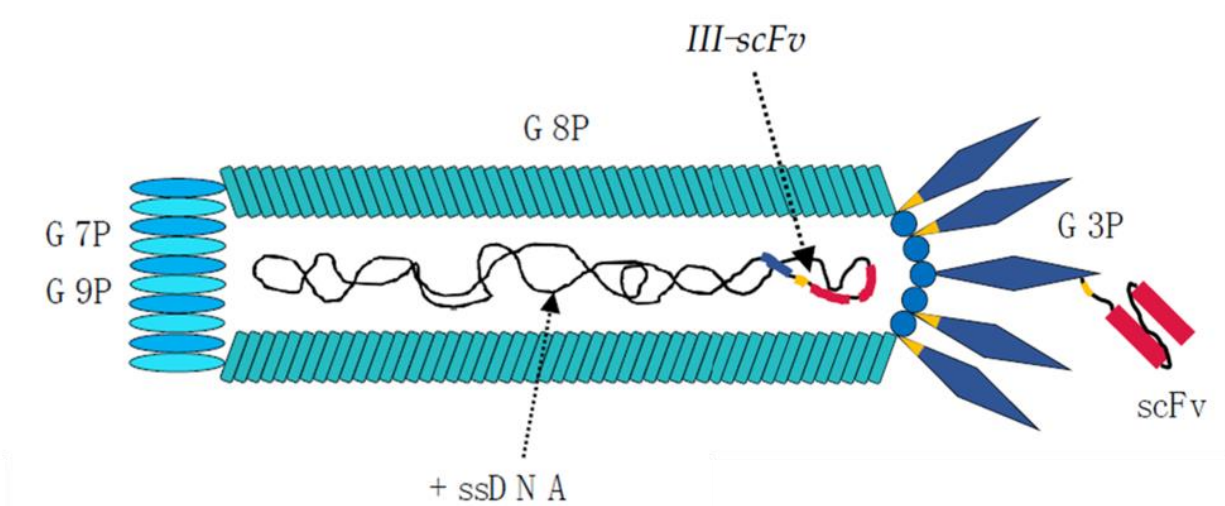
Il existe de nombreux critères d'évaluation d'un potentiel futur anticorps thérapeutique en plus de son efficacité thérapeutique. Les principaux sont l'affinité, la spécificité, la stabilité et le taux d'expression. Il est possible de modifier la séquence primaire des anticorps dans le but d'améliorer ces différentes caractéristiques. Cependant la grande majorité des modifications apportées à un anticorps conduit à la perte partielle ou totale de son activité. Il existe aujourd'hui des méthodes d'affichage ou de « display » qui permettent de sélectionner des variants fonctionnels tout au long des étapes d'ingénierie de l'anticorps. Ces technologies permettent d'exprimer des anticorps le plus souvent sous forme de fragments (scFv, VHH, scFab ou Fab) et elles présentent toutes quatre étapes indispensables :

- 1) Génération de la diversité génotypique : clonage d'un répertoire d'anticorps naïfs ou provenant d'immunisations (chez l'animal ou l'homme). Il peut également s'agir d'une mutagenèse artificielle dirigée ou non.
- 2) Lien entre le phénotype (fonctionnalité de l'anticorps) et le génotype (séquence de l'anticorps). L'information génétique qui code pour l'anticorps (ADN ou ARN) est physiquement liée au fragment d'anticorps.
- 3) Application d'une pression de sélection permettant de conserver seulement les fragments d'anticorps fonctionnels possédant les propriétés souhaitées.
- 4) Amplification de l'information génotypique afin de pouvoir la séquencer.

Dans ce chapitre nous nous restreindrons aux seules méthodes de *display* et plus particulièrement au *display* d'anticorps et de fragments d'anticorps. D'autres méthodes existent comme le double hybride, la complémentation ou la microfluidique mais restent anecdotiques dans la littérature dans le cas des anticorps. Le *display* et l'ingénierie d'autres objets protéiques comme les enzymes, les toxines ou les récepteurs ne seront pas abordés. Le *Yeast Surface Display* sera particulièrement développé puisque c'est la technologie qui a été choisie pour la partie expérimentale de la thèse.

## I. Les technologies de *display*

### A. Phage *Display*



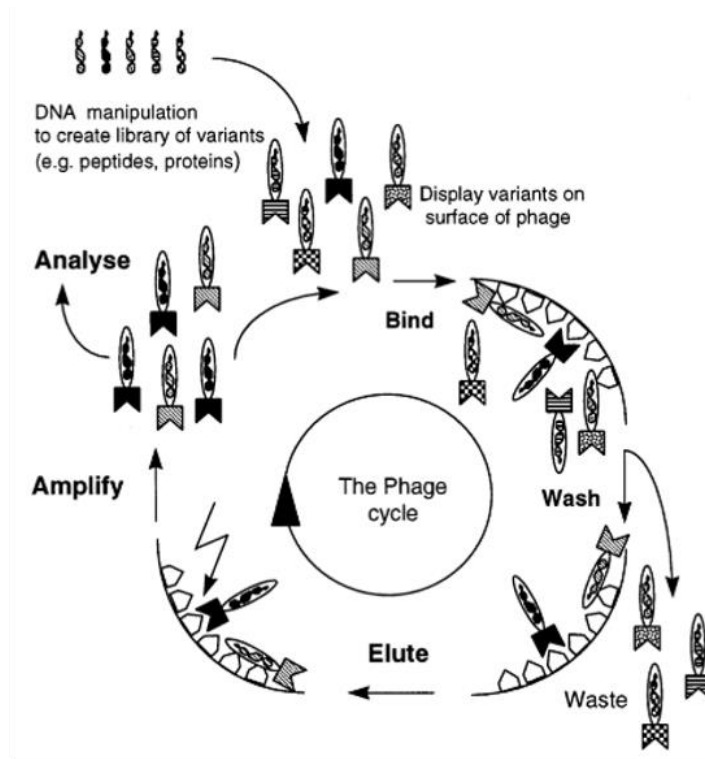
**Figure 17 : Ingénierie du phage M13 pour réaliser des expériences de phage *display*.** Le génome du phage M13 est stocké au sein d'une capside sous forme d'ADN simple brin positif. La capside est constituée de plusieurs protéines (G3P, G9P, G7P et G8P). La protéine G3P est celle qui présente le plus faible nombre de copies (5) et c'est cette protéine qui est fusionnée à la protéine d'intérêt. Dans le génome du phage, la protéine G3P est représentée en bleu et la protéine d'intérêt (ici un scFv) est représentée en rouge. Un site de clivage à la trypsine (en jaune) peut être inséré pour distinguer les copies de G3P natives de celles fusionnées avec la protéine d'intérêt. Adapté de Ledsgaard *et al* <sup>201</sup>.

La première technologie de *display* décrite en 1985 par George Smith <sup>199</sup> repose sur l'affichage de peptides et de protéines à la surface de phages filamenteux. Cette avancée majeure lui a valu l'obtention du prix Nobel de chimie en 2018 pour ses travaux sur le *phage display* et l'évolution dirigée, partagé avec Greg Winter et Frances Arnold. Cinq ans plus tard, McCafferty *et al.* ont montré qu'il était possible de cribler et de sélectionner des anticorps spécifiques (sous forme de scFv) en utilisant cette méthode <sup>200</sup>. Depuis 1990, d'autres formats d'anticorps ont été exprimés avec succès à la surface de phages : anticorps à simple domaine de camélidés (VHH) et de requins (IgNAR), *diabodies* (scFv bivalent) ainsi que le format Fab entier sans linker <sup>201</sup>.

Plusieurs bactériophages peuvent être utilisés pour faire du *phage display*. Parmi eux, le bactériophage filamenteux M13 est le plus utilisé, il sera plus amplement détaillé et est présenté en **Figure 17**. Le bactériophage M13 possède un chromosome d'ADN simple brin d'environ 6400 paires de bases qui code pour 11 protéines virales <sup>202</sup>. Généralement, la protéine d'intérêt (ici un fragment d'anticorps) est fusionnée à la protéine G3P qui est une protéine de la capside présente à 5 copies par phage. Il est préférable que chaque phage ne porte qu'une seule copie de l'anticorps pour éviter les problèmes d'avidité lors de la sélection.

Pour faire co-exister à la surface du phage la protéine G3P native avec la protéine G3P fusionnée à la protéine d'intérêt, plusieurs stratégies sont possibles. Il est ainsi possible d'utiliser

un plasmide auxiliaire appelé phagemide codant pour la protéine G3P-anticorps et qui va entrer en compétition avec la protéine G3P native présente dans le génome du phage présent grâce au « *phage helper* ». Dans ces conditions le gène codant pour la protéine d'intérêt est présent chez 10% des phages générés et majoritairement présent à une seule copie <sup>203</sup>.



**Figure 18 : Le cycle du phage display.** L'ADN codant pour des millions de variants de la protéine d'intérêt est cloné de façon à la fusionner avec une protéine de la capsid du phage. Des ligands spécifiques peuvent être isolés grâce à des cycles de sélections contre l'antigène. Chaque cycle est composé d'une étape de liaison suivie d'un lavage puis d'une élution et enfin une amplification des phages portant une protéine d'intérêt fonctionnelle. Adapté de Hoogenboom *et al* <sup>203</sup>.

Pour les étapes de criblage fonctionnel ou « bio-panning », l'antigène d'intérêt peut être immobilisé sur une surface par adsorption directe ou par un système de capture de type streptavidine-biotine. La banque de phages exprimant une banque d'anticorps est mise au contact de l'antigène puis lavée plusieurs fois pour ne retenir que les phages exprimant un fragment d'anticorps spécifique de l'antigène (**Figure 18**). Ces étapes de sélection peuvent être répétées en criblant itérativement la banque à plusieurs reprises, avec des étapes d'amplification des variants sélectionnés par infection d'*Escherichia coli*. A l'issue des étapes de sélection, la séquence des anticorps d'intérêt peut être obtenue par séquençage.

Cette technologie est devenue une des plateformes principales pour l'obtention d'anticorps thérapeutiques humains en parallèle de l'utilisation de souris transgéniques, des techniques d'humanisation des anticorps murins ainsi que du *single B-cell cloning* <sup>204</sup>. Un exemple impressionnant de l'utilisation du phage display pour la découverte de nouveaux anticorps est



celui de Schofield *et al.* en 2007. Dans cette étude les auteurs ont sélectionné plus de 7200 anticorps recombinants au format scFv contre 292 antigènes différents à l'aide d'une banque de phage de scFv humains contenant plus de  $10^{10}$  clones indépendants <sup>205</sup>. Le *phage display* peut également être utilisé pour la maturation d'affinité des anticorps introduisant de la diversité au sein de la séquence grâce à des méthodes comme le *light chain* et le *heavy chain shuffling*, la mutagenèse aléatoire ou dirigée <sup>201</sup>. Aujourd'hui, six anticorps monoclonaux humains développés par *phage display* ont été approuvés pour un usage thérapeutique et de nombreux autres anticorps développés à l'aide de la même technologie sont en cours d'évaluation clinique <sup>206</sup>.

## B. Autres méthodes de *display*

### 1. Les systèmes cellulaires : *mammalian*, *yeast* et *bacterial display*

Il existe trois technologies de *display* utilisant un système cellulaire basé sur des cellules de mammifères, de levures et des bactéries. Ces trois technologies ont en commun la nature du lien entre le génotype et le phénotype : la protéine d'intérêt est exprimée à la surface de la cellule et l'ADN codant est contenu au sein de la cellule. Pour permettre l'affichage en surface, la protéine d'intérêt est généralement fusionnée à une protéine exprimée à la surface de la cellule appelée protéine d'ancrage.

Le *bacterial display* est apparu peu après la première preuve de concept du *phage display*, en 1986 <sup>207</sup>. L'affichage de la protéine à la surface de la bactérie est réalisé grâce à une fusion génétique avec une protéine présente dans une des deux membranes de la bactérie (pour les Gram négatif) ou dans la paroi cellulaire (pour les Gram positif). Le groupe de George Georgiou a développé le *bacterial display* d'anticorps depuis le début des années 1990. Les anticorps ont été exprimés d'abord au format scFv à la surface de la membrane externe d'*E. coli* <sup>208</sup> et la première banque d'anticorps au format scFv d'une diversité de  $10^6$  a été criblée en utilisant cette plateforme <sup>209</sup>. D'autres systèmes ont ensuite été mis au point pour permettre l'affichage d'anticorps au format Fab <sup>210</sup> ainsi qu'au format IgG complet <sup>211</sup>. Des bibliothèques d'anticorps ont été générées en *bacterial display* puis criblées à l'aide de billes magnétiques ou par cytométrie en flux pour la liaison à un antigène d'intérêt, essentiellement pour des applications de maturation d'affinité <sup>212</sup>. Les techniques développées par le groupe de George Georgiou nécessitent une étape de perméabilisation de la membrane externe d'*E. coli* qui diminue la viabilité des bactéries. Une étape d'amplification par PCR est alors requise pour amplifier les séquences fonctionnelles et effectuer de nouvelles étapes de sélection <sup>212</sup>. Le *bacterial display* reste cependant assez peu utilisé dans le domaine de l'ingénierie des anticorps.

Le *mammalian display* repose sur les mêmes principes : la protéine d'intérêt est fusionnée à une protéine de la membrane de la cellule pour être exprimée à sa surface. Le développement de cette technologie est plus récent et la première preuve de concept d'affichage de scFv à la

surface de cellules HEK a été faite en 2006 <sup>213</sup>. Il existe différentes méthodes de transfection des cellules de mammifères permettant une expression plus ou moins stable de la construction selon qu'il s'agit d'une intégration dans le génome ou d'un plasmide destiné à l'expression transitoire. Les anticorps peuvent être exprimés sous différents formats à l'aide d'une fusion avec le domaine transmembranaire d'un récepteur de surface comme la protéine H-2K<sup>K</sup> ou le PDGF-R (Platelet-derived growth factor receptor) <sup>213,214</sup>. La principale limitation de cette technologie est la faible taille des banques d'anticorps ( $\sim 10^4$ ) obtenues liée aux faibles efficacités de transformation, surtout pour les lignées stables, comparée au phage ( $>10^{10}$ ) et au *yeast display* ( $>10^7$ ). Récemment, une équipe a obtenu une efficacité de transfection stable de l'ordre de  $10^5$  en utilisant le système CRISPR-Cas9 <sup>215</sup>.

De manière similaire, le *Yeast Surface Display* (YSD) est une technologie qui permet d'afficher à la surface de levures des protéines fusionnées à une protéine d'ancrage présente dans la paroi de cette dernière. Une partie dédiée sera consacrée à cette technologie plus loin dans ce chapitre.

## 2. Les systèmes acellulaires : mRNA, ribosome et DNA *display*

Il existe trois technologies de *display* n'ayant pas besoin de cellule pour fonctionner (*cell-free*). Deux d'entre elles sont basées sur une information génétique en ARN, le lien entre le génotype et le phénotype étant réalisé par un groupement puromycine (mRNA *display*) ou un ribosome (ribosome *display*). La dernière est basée sur l'ADN lui-même où le lien se fait grâce à une liaison streptavidine-biotine. Ces technologies présentent l'avantage majeur de ne pas être limitées par l'efficacité de transformation et peuvent atteindre des diversités bien supérieures à celles du *yeast* ou du *phage display* <sup>216</sup>. Différentes banques d'anticorps ont été générées et criblées avec succès en utilisant ces technologies que ce soit au format scFv <sup>217</sup> ou Fab <sup>218</sup>. Ces systèmes acellulaires restent cependant peu utilisés pour l'ingénierie des anticorps, probablement à cause de leur nouveauté et de la complexité du système *in vitro* à mettre en œuvre.

## II. Le Yeast Surface Display (YSD) et ses applications

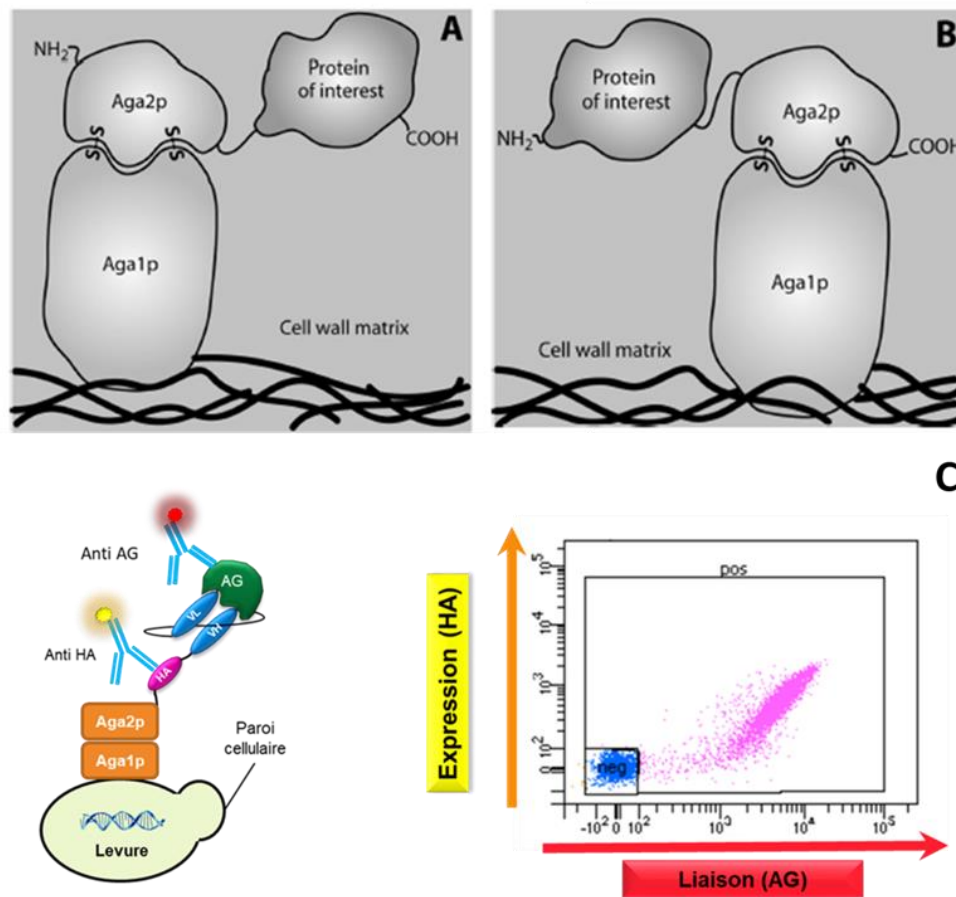
### A. Historique et principe

La première description d'un système permettant d'exprimer une protéine hétérologue à la surface d'une cellule de levure remonte à 1993 par Schreuder *et al* <sup>219</sup>. Les premières protéines exprimées en YSD étaient des enzymes ou des protéines virales et bactériennes pour le développement de vaccins <sup>220</sup>. Ces protéines étaient fusionnées à l' $\alpha$ -agglutinine qui est une protéine de 650 acides aminés elle-même liée aux saccharides de la paroi cellulaire par une ancre GPI située dans sa région C-terminale. D'autres protéines d'ancrage de la famille des ancres GPI



(glycosylphosphatidylinositol) comme Flo1 ont également été utilisées pour l'affichage de protéines à la surface de *Saccharomyces cerevisiae*<sup>221</sup>. D'autres espèces de levures ont également été utilisées comme *Pichia pastoris* ou *Hansenula polymorpha* mais *S. cerevisiae* reste la plus utilisée.

En 1997, Boder et Wittrup s'intéressent à l'affichage de fragments d'anticorps en YSD et publient un nouveau système d'ancrage à la paroi : le système a-agglutinine<sup>222</sup>. Le système est présenté en **Figure 19AB** et est composé de deux partenaires protéiques : Aga1p (725 acides aminés) et Aga2p (87 acides aminés). Aga1p est une protéine à ancre GPI présente à la surface de la paroi de *S. cerevisiae* et Aga2p se lie à Aga1p grâce à deux ponts disulfures.



**Figure 19 : Principe de la sélection d'anticorps en Yeast Surface Display en utilisant le système Aga1p/Aga2p.** (A-B) Construction exprimée en fusion avec la protéine Aga2p en C-terminal (A) ou N-terminal (B). Dans chacune des constructions, la protéine Aga2p s'associe à la protéine Aga1p par deux ponts disulfures et lie de ce fait la protéine d'intérêt à la paroi de la levure. (C) Exemple d'une analyse en cytométrie en flux. Un scFv est exprimé à la surface de la levure avec un tag HA permettant de suivre l'expression de la construction. La détection de l'antigène peut se faire à l'aide d'un antigène marqué ou d'un anticorps anti-antigène marqué. Sur le dot plot de cytométrie, une population de levures exprimant un scFv fonctionnel est visible en rose. En YSD, il existe toujours une population de levures qui n'expriment pas la construction, elle est appelée population double négative (en bleu). (A) et (B) adaptés de Pepper *et al*<sup>220</sup>.

Le système le plus communément utilisé repose sur la surexpression d'Aga1p et d'Aga2p grâce à un promoteur inductible au galactose <sup>223</sup>. Afin d'améliorer les niveaux d'affichage, les souches de *display* ont généralement fait l'objet d'une ingénierie, et une cassette permettant la surexpression d'Aga1p est intégrée dans le génome avec un promoteur Gal1. La surexpression d'Aga2p est rendue possible grâce à un plasmide d'expression contenant le gène de la protéine d'intérêt fusionné à celui d'Aga2p précédé d'un promoteur inductible au galactose ainsi que d'une séquence codant un peptide signal de sécrétion <sup>224</sup>.

Le criblage des banques de YSD peut être réalisé grâce à différentes méthodes comme le tri par billes magnétiques ou la cytométrie en flux. Du fait de sa rapidité, le tri par billes magnétiques est particulièrement utile pour le criblage de banques possédant une grande diversité ( $>10^7$ ).

L'avantage principal des méthodes d'affichage cellulaire est de rendre possible le tri par cytométrie en flux. Deux rapporteurs fluorescents sont généralement utilisés, le premier pour l'expression de la construction et le second pour la fonctionnalité de la protéine exprimée (le plus souvent la présence du ligand de la protéine d'intérêt). Ainsi, les données de cytométrie sont visualisées sous forme de *dot-plots*, où chaque cellule correspond à un point, avec en ordonnée le signal correspondant à l'expression et en abscisses le signal correspondant à la liaison (**Figure 19C**).

Cette méthode de sélection par FACS permet de séparer les mutants d'affinités proches grâce à la définition de fenêtres de tri. Il est alors possible de trier spécifiquement les cellules ayant le niveau d'expression désiré et les propriétés de liaison voulues (haute, moyenne ou basse) à une concentration d'antigène définie. De plus, l'interaction est observée en solution avec une concentration d'antigène donnée ce qui permet de connaître la gamme d'affinité de l'interaction étudiée.

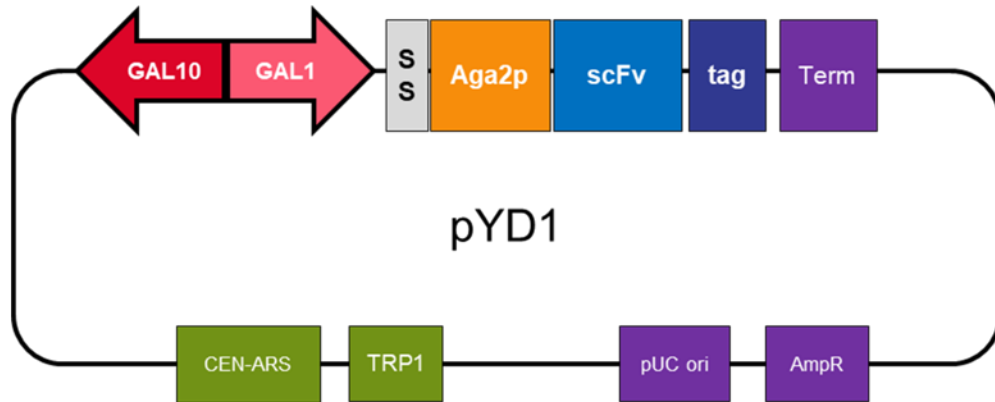
## B. Application à l'affichage d'anticorps

### 1. Les plasmides d'expression

Les plasmides d'expression des fragments d'anticorps en YSD dérivent souvent du plasmide pYD1 créé par l'équipe de Wittrup en 1997 <sup>225</sup> et décrit en **Figure 20**. La cassette d'expression de la protéine d'intérêt à exprimer à la surface de la levure est composée d'un promoteur inductible au galactose (GAL1), du gène codant la protéine Aga2p précédée de son peptide signal, du gène codant pour la protéine d'intérêt fusionné à celui d'Aga2p en 3', suivi d'une séquence codant pour un tag polyhistidine et d'un terminateur. Les plasmides d'expression de YSD comme le pYD1 sont des plasmides navettes entre bactérie et levure. Pour cela, le pYD1 possède une origine de répllication bactérienne ainsi qu'un gène de résistance à l'ampicilline. Le plasmide possède également une origine de répllication centromérique (CEN/ARS) de levure. Cette particularité permet au plasmide de n'être présent qu'à un faible nombre de copies et très

stable car répliqué en même temps que les chromosomes. Cette caractéristique réduit la probabilité d'avoir deux plasmides différents au sein de la même levure après plusieurs passages. Enfin un gène marqueur de sélection TRP1 permet de compléter l'auxotrophie de la souche et la croissance des levures transformées en milieu sélectif dépourvu de tryptophane.

## 2. Les formats d'anticorps



**Figure 20 : Carte simplifiée du plasmide pYD1.** La cassette d'expression est constituée d'un promoteur inducible au galactose (GAL1), d'un peptide signal de sécrétion, suivi de la protéine d'ancrage Aga2p fusionnée à la protéine d'intérêt ainsi qu'un tag permettant de suivre l'expression de la construction. Le maintien du plasmide dans les levures est rendu possible grâce à une origine de répllication levure (CEN-ARS) associée à un marqueur de sélection TRP1 qui permet de compléter l'auxotrophie de la souche. C'est un plasmide navette qui peut être utilisé chez la bactérie grâce à une origine de répllication bactérienne (pUC ori) associée à un gène de résistance à l'ampicilline (AmpR).

Les anticorps ont été exprimés en YSD sous différents formats usuels comme le scFv, le scFab, le Fab et l'IgG entière <sup>224</sup>. L'affichage d'IgG complète est réalisé grâce à un système de sécrétion-recapture <sup>226</sup>. L'affichage d'anticorps complet via une protéine de fusion ancrée dans la paroi de la levure n'a pas encore été décrit dans la littérature. Il n'existe pas de consensus concernant le choix du format d'expression des anticorps en YSD. Cependant, lors du reformatage des fragments d'anticorps en anticorps complet, les formats scFab et Fab se montrent plus performants que le format scFv <sup>227</sup>. Les différents formats ainsi que leurs avantages et inconvénients pour une expression en YSD seront détaillés dans la première partie des résultats.

## 3. Génération et criblage des banques

La première étape de la génération d'une banque d'anticorps consiste à générer de la diversité sous forme d'ADN. Pour cela, il existe généralement deux stratégies. La première consiste à générer une banque de type répertoire en réalisant des PCR sur un répertoire naïf ou immun. La deuxième consiste à générer une banque à partir de la séquence d'un anticorps connu (banque de

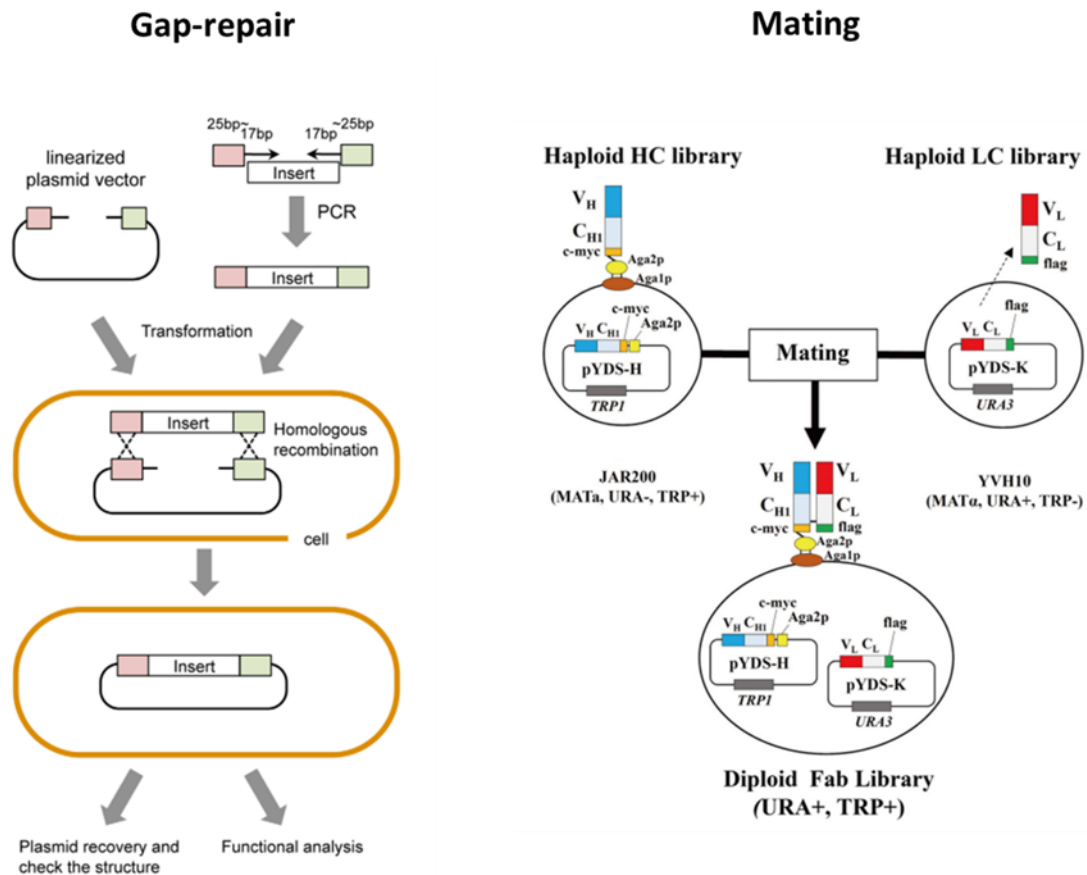
type ingénierie) où l'on introduit de la diversité au sein d'une séquence d'anticorps connue pour améliorer ses propriétés.

Pour les banques de type répertoire, des RT-PCR sont réalisées à partir des ARN messagers codant pour les chaînes lourdes et légères d'anticorps. Un mélange d'amorces est utilisé afin de pouvoir amplifier les différentes familles d'anticorps. L'introduction de la diversité au sein des banques de type ingénierie peut être réalisée par PCR à erreur de manière aléatoire ou par mutagenèse dirigée. Les PCR de mutagenèse dirigée utilisent souvent des codons dégénérés aux positions d'intérêt pour générer la diversité. Il existe une multitude de codons dégénérés permettant d'autoriser certains acides aminés voir certaines familles d'acides aminés. Par exemple, le codon « NNS » qui autorise les 4 bases sur les deux premières positions et seulement un C ou un G pour la dernière position permet de générer 32 codons qui codent pour les 20 acides aminés. Un codon « VVW » autorisera seulement les acides aminés hydrophiles à l'exception de la cystéine. Il existe également un processus de synthèse d'amorces en utilisant des trinucéotides qui permet de sélectionner exactement les acides aminés désirés pour chaque position avec un seul codon par acide aminé <sup>228</sup>. Cette technique permet de réduire de façon importante la diversité génétique des banques générées tout en conservant la même diversité d'acides aminés.

La constitution et le clonage des banques pour le *Yeast Surface Display* peuvent parfaitement être réalisés par des outils conventionnels de biologie moléculaire (restriction-ligature, Gateway, Golden Gate ...). Néanmoins, deux caractéristiques génétiques de la levure permettent d'utiliser des méthodes spécifiques de clonage particulièrement intéressantes : le clonage par recombinaison homologue dite de « *gap-repair* » et le *chain-shuffling* permis par le mécanisme de reproduction sexuée appelé « *mating* » des cellules haploïdes de levures *S. cerevisiae*.

Le *gap-repair* est basé sur un mécanisme de réparation de l'ADN double brin linéaire <sup>229</sup>. Il permet de générer un plasmide circulaire à partir d'un ou plusieurs fragments d'ADN linéaires et d'un plasmide linéarisé par des enzymes de restriction possédant une homologie de séquence à leurs extrémités (de 30 à 50 bases généralement) (**Figure 21**) <sup>230</sup>. Du fait de sa simplicité, le système de *gap-repair* est généralement utilisé pour les clonages des banques en YSD, notamment pour les banques de scFv ou de scFab (à une seule chaîne). Cette méthode peut aussi être utilisée pour générer des plasmides de grande taille en utilisant un nombre de fragments linéaires non limité pour la génération de banques avec plusieurs chaînes <sup>231</sup>.

Pour les protéines hétéro-dimériques comme les Fab, l'utilisation du *mating* permet également le *gene shuffling*. Ainsi, des répertoires indépendants de chaînes légères (LC) et lourdes (HC) peuvent être construits dans des souches de levures possédant des types sexuels opposés. Ces répertoires peuvent alors être combinés grâce à la formation de levures diploïdes par *mating* (**Figure 21**) <sup>232</sup>. Cette approche permet de générer des librairies de Fab d'une diversité pouvant atteindre  $10^9$  à partir de librairies haploïdes d'une diversité de l'ordre de  $10^6$  <sup>233,234</sup>.



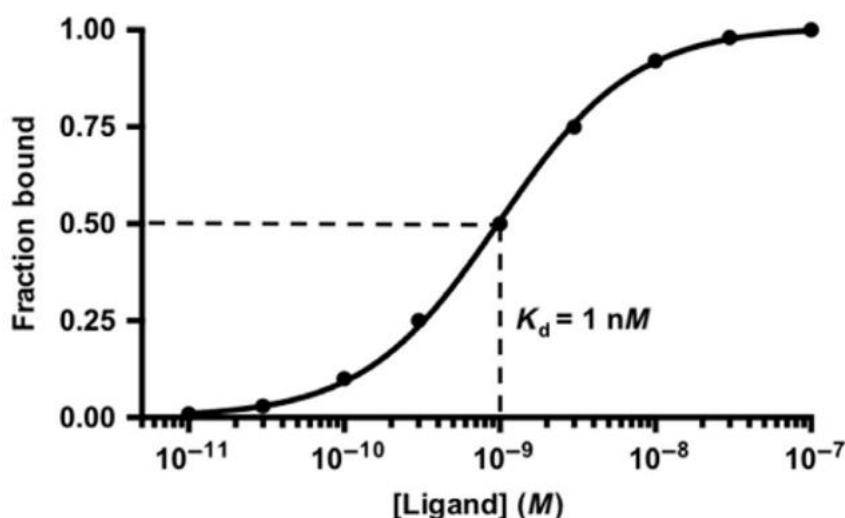
**Figure 21 : Principe de la transformation de levures par gap-repair et par mating. (A gauche)** Procédure générale de gap-repair. Le plasmide d'expression est linéarisé et un insert est généré par PCR avec des régions homologues au plasmide en amont et en aval. Les deux fragments d'ADN linéaires sont transformés dans le levure où a lieu la réaction de recombinaison homologue. Les levures portant un plasmide complet peuvent être sélectionnées grâce à un marqueur de sélection présent dans le plasmide. **(A droite)** Utilisation du mating pour construire une banque de Fab. La librairie combinatoire V<sub>H</sub>/V<sub>L</sub> est générée en réalisant le mating de deux librairies de levures haploïdes possédant des types sexuels opposés (a et α) et deux marqueurs de sélection différents. Adapté de Baek *et al*<sup>232</sup> et Chino *et al*<sup>230</sup>.

Lorsqu'il s'agit de choisir un système de *display*, une donnée importante est l'efficacité de transformation du système et donc la taille maximale des banques que l'on pourra générer dans ce système. Les protocoles de transformation de banques par recombinaison homologue en YSD permettent d'obtenir une efficacité de transformation de l'ordre de  $10^8$  transformants par  $\mu\text{g}$  de plasmide linéarisé. Il est alors possible d'obtenir des banques d'une diversité pouvant atteindre  $10^{10}$  si la réaction de recombinaison homologue est répétée plusieurs fois<sup>235</sup>.

## C. Applications du Yeast Surface Display à l'ingénierie des anticorps

### 1. La mesure d'affinité à l'équilibre

Il existe deux méthodes pour mesurer l'affinité d'une interaction moléculaire : la méthode cinétique et la méthode à l'équilibre. La méthode cinétique consiste à mesurer la vitesse d'association et la vitesse de dissociation de l'anticorps pour obtenir la constante de dissociation, le  $K_D$  ( $\text{mol.L}^{-1}$ ) qui est l'inverse de la constante d'affinité, et est défini par le rapport entre la vitesse de dissociation et la vitesse d'association. A l'équilibre, lorsque 50% des sites de liaison à l'antigène sont occupés, le  $K_D$  est égal à la concentration en antigène (**Figure 22**).



**Figure 22 : Courbe de liaison obtenue lors d'une mesure d'affinité à l'équilibre.** La fluorescence de liaison à la surface des levures est mesurée en présence de différentes concentrations d'antigène. La ligne en pointillé montre la concentration pour laquelle 50% du signal maximal de fluorescence est obtenu, c'est-à-dire le  $K_D$ . Adapté de Hunter *et al* <sup>236</sup>.

Il est possible de mesurer l'affinité d'un anticorps exprimé en YSD grâce à la méthode à l'équilibre. Les levures sont incubées avec une concentration décroissante d'antigène et la fluorescence de liaison est mesurée par cytométrie en flux pour chaque point de la gamme. Le  $K_D$  est obtenu lorsque le signal de fluorescence atteint 50% du signal maximal. Différentes conditions doivent être remplies pour pouvoir mesurer correctement cette interaction <sup>236</sup>. Notamment, il faut éviter le phénomène de déplétion de ligand sachant que le nombre de constructions exprimées à la surface de la levure est estimé à environ  $5 \cdot 10^4$ - $1 \cdot 10^5$  par cellule selon les publications <sup>223,236</sup>.



## 2. Le criblage de banques naïves et immunes

Le YSD est une plateforme performante pour l'identification de nouveaux anticorps à l'aide de bibliothèques naïves ou immunes d'anticorps. Le groupe de Wittrup a fait la démonstration de l'isolement de nouveaux anticorps à partir d'une banque naïve d'anticorps humains au format scFv<sup>237</sup>. La diversité de la banque de scFv obtenue par les auteurs était de  $10^9$ . Cette banque de levures a été utilisée pour isoler plus de 50 scFv dirigés contre 9 antigènes différents (protéines, peptides et haptènes). La sélection des clones a été réalisée par une combinaison des techniques de MACS (Magnetic-activated cell sorting) et FACS. Depuis cette démonstration, de nombreuses études ont été publiées sur l'obtention de nouveaux anticorps en YSD à partir de banques naïves et immunes<sup>224</sup>. Les affinités des anticorps isolés avant une éventuelle étape de maturation d'affinité est généralement comprise entre 1nM et 1μM en fonction de la taille de la banque et de la nature de l'antigène (voir **tableau 6**). Feldhaus et Wittrup montrent que l'affinité des scFv isolés est généralement du même ordre de grandeur que la concentration en antigène utilisée pour réaliser la sélection.

## 3. La maturation d'affinité

Le YSD est un outil d'ingénierie des protéines qui a permis d'améliorer l'affinité de nombreux anticorps. Le premier fragment d'anticorps à avoir été affiché à la surface de la levure en utilisant le système Aga2p est un scFv anti-fluorescéine<sup>222</sup>. Des banques de variants de ce scFv ont été générées par mutagenèse aléatoire et criblées par YSD et FACS. De nombreux mutants avec une affinité améliorée allant jusqu'à 48 fM ont été isolés (l'affinité de départ était de 300 pM)<sup>238</sup>. Depuis ce premier exemple, de nombreux autres cas de maturation d'affinité en YSD ont été décrits dans la littérature. Un exemple important est celui de la maturation d'affinité de l'adalimumab utilisant une méthode de mutagenèse dirigée appelée « look through mutagenesis ». Dans cette méthode les acides aminés sont divisés en 7 groupes selon les propriétés de leurs chaînes latérales et chacun de ces groupes est représenté par un ou deux acides aminés. La permissivité de chaque position au sein des CDR est évaluée en les mutant une à une par un acide aminé appartenant à l'un des 7 groupes. Des banques combinant les meilleures mutations sont ensuite générées pour obtenir des variants multi-mutés avec une affinité améliorée. Dans le cas de l'adalimumab cette méthode a permis d'améliorer son affinité en passant de 950 pM à 1 pM en effectuant 10 mutations.

De nombreux autres anticorps ont été modifiés en YSD et ceux-ci ont été listés par Boder *et al.* (voir **tableau 6**). Au total, 14 anticorps sont passés par une étape de maturation d'affinité en YSD. Parmi ceux-ci, 12 sont exprimés au format scFv et les deux restants l'ont été au format Fab. L'affinité finale obtenue pour l'antigène est généralement sub-nanomolaire. L'introduction de la variabilité au sein de la séquence est généralement réalisée par mutagenèse aléatoire.



**Tableau 6** : Recensement des différentes expériences de génération d'anticorps et de maturation d'affinité en Yeast Surface Display.

Antigen	Ab format	Affin. Mat. or 1° selection	Selection method	Affinity achieved
Activated Mac-1	scFv	1°	FACS	2 nM
Amphiregulin	scFv	1°	MACS + FACS	~100 nM
Amyloid precursor protein	scFv	1°	MACS + FACS	604 nM
Betacellulin	scFv	1°	MACS + FACS	80 nM
BoNT	scFv	AM	FACS	87 pM
BoNT	Fab	AM	FACS	34 pM
BoNT	V <sub>H</sub> H	1°	FACS	147 pM
CEA	scFv	AM	FACS	30 pM
DOTA-yttrium	scFv	AM	FACS	8.2 pM
EGF	scFv	1°	MACS + FACS	3 nM
EGFR	scFv	AM	FACS	15 nM
EGFR	scFv	1°	MACS + FACS	~400 nM
Endosialin/TEM1	scFv	1°	MACS + FACS	4 nM
EPOR	scFv	AM	FACS	330 pM
FITC	scFv	1°	MACS + FACS	5 nM
Fluorescein	scFv	AM	FACS	48 fM
GFP	V <sub>H</sub> H	1°	FACS	3 nM
Glycoprotein NMB	scFv	AM	FACS	300 pM
HIV-1 gp120	scFab	1°	MACS + FACS	500 pM
Lysozyme	scFv	1°	MACS + FACS	6 nM
Malachite green	scFv	1°	MACS + FACS	38 nM
Monkey CD3eY	scFv	AM	FACS	2.9 nM
PCSK9	Fab	1°	FACS	1 nM
Phospho-P53	scFv	1°	MACS + FACS	9 nM
Prostate stem cell antigen	scFv	AM	FACS	11 nM
Streptavidin	Fab	AM	FACS	3–5 nM
T cell receptor	scFv	AM	FACS	~40 nM
Tacrolimus	scFv	AM	FACS	38 pM
Thiazole orange	scFv	1°	MACS + FACS	2 nM
TNF- $\alpha$	scFv	AM	MACS + FACS	1 pM
XPA	scFv	1°	MACS + FACS	82 nM
$\alpha$ -synuclein	scFv	1°	MACS + FACS	n/d <sup>a</sup>

Adapté de Boder *et al* <sup>224</sup>

#### 4. Ingénierie des propriétés autres que l'affinité

L'affinité est l'un des paramètres les plus importants dans l'évaluation d'un anticorps. Cependant, de nombreuses autres propriétés sont également fondamentales comme la stabilité, la sélectivité, la spécificité, le taux d'expression ou les propriétés biophysiques.

Il est possible d'établir une corrélation entre le niveau d'expression d'une protéine et sa stabilité thermique<sup>239</sup>. En effet, la sécrétion des protéines chez les eucaryotes est régulée par un système de contrôle qualité présent dans le réticulum endoplasmique et l'appareil de Golgi<sup>240</sup>. L'ingénierie de la stabilité des protéines peut ainsi reposer sur la sélection des protéines les mieux exprimées. Cette méthode, appliquée avec succès pour stabiliser des TCR<sup>241</sup>, l'est également pour améliorer le taux d'expression des protéines. Elle possède toutefois des limitations puisque des protéines instables peuvent être fortement exprimées à la surface de la levure malgré leur mauvais repliement<sup>242</sup>.

Une deuxième méthode permet d'évaluer directement la stabilité thermique des protéines affichées à la surface des levures. Les levures exprimant la protéine d'intérêt sont soumises à un choc thermique adapté puis triées selon l'activité résiduelle de la fonctionnalité de la protéine d'intérêt<sup>243</sup>. La levure est en effet un organisme très résistant grâce à sa paroi cellulaire et peut supporter une température de 85°C pendant 10 minutes la rendant particulièrement adaptée à ce type d'expériences<sup>241</sup>.

D'autres méthodes permettent d'évaluer et de sélectionner les anticorps exprimés en YSD sur le critère de la spécificité. Il est ainsi possible de réduire la poly-spécificité des anticorps en contre-sélectionnant les levures exprimant des anticorps se liant à un cocktail de protéines biotinylées non pertinentes<sup>244</sup>.

Dans le cas où un épitope spécifique est requis pour l'anticorps, il est possible de sélectionner des anticorps liant la protéine sauvage mais pas un variant spécifiquement muté au sein d'une région clé de la protéine (par exemple pour cibler un épitope réagissant de façon croisée entre plusieurs souches de virus)<sup>245</sup>.

Il est également possible de contre-sélectionner les anticorps se liant à des antigènes dont la structure est proche de la protéine d'intérêt, si ces protéines sont connues et disponibles.

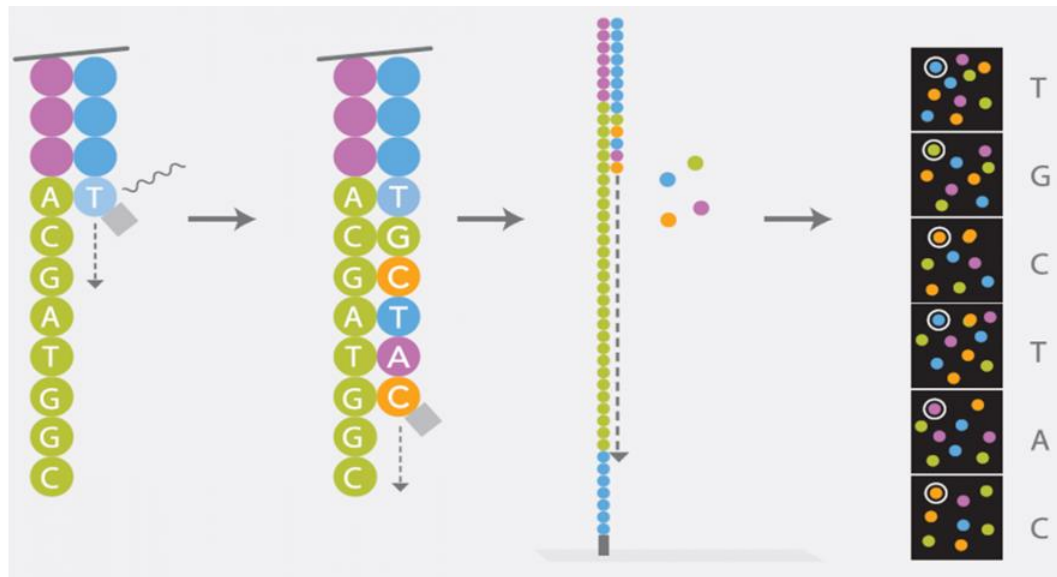


### III. Les apports du séquençage à haut débit pour l'ingénierie des anticorps

#### A. Principe du séquençage à haut débit, le cas d'Illumina

L'arrivée du séquençage à haut débit a révolutionné de nombreux domaines de recherche en biologie à commencer par le séquençage de génomes. Le séquençage à haut débit (NGS, Next-Generation Sequencing) permet de séquencer des dizaines voire des centaines de millions de fragments d'ADN en une seule fois. Il existe différentes plateformes de séquençage à haut débit qui utilisent des principes différents (Pyroséquençage, Ion Torrent, Illumina, Nanopore).

La technologie Illumina est aujourd'hui la plus répandue. Le principe repose sur l'émission d'un signal fluorescent spécifique pour l'incorporation de chaque nouvelle base (**Figure 23**). A la différence du séquençage Sanger, ceci n'est pas réalisé sur un ADN clonal mais au sein d'une micropuce contenant des millions de fragments d'ADN immobilisés grâce à des adaptateurs. Cette technique permet actuellement de lire 350 pb par « run », en *single-end* (lecture dans un seul sens) ou 2x350pb en *paired-end* (lecture combinée dans un sens puis l'autre).



**Figure 23 : Principe du séquençage Illumina.** Le brin complémentaire de l'ADN simple brin fixé à la puce est synthétisé. Chaque base ajoutée émet un signal de fluorescence propre à l'identité de la base ajoutée. Chaque brin d'ADN fixé à la puce est localisé spatialement et les signaux de fluorescence successifs sont enregistrés, ce qui permet de connaître l'enchaînement des bases nouvellement synthétisées et ainsi de séquencer le brins d'ADN fixés à la puce.

Adapté de : <https://www.cephbase.org/en/services/next-generation-sequencing/>

#### B. Apport du séquençage à haut débit pour l'ingénierie des anticorps

Le NGS possède de très vastes domaines d'application <sup>246</sup> (séquençage de génomes, identification de néo-épitopes tumoraux, exploration de la diversité de populations microbiennes) qui ne seront pas abordés ici.

Dans le domaine de l'ingénierie des anticorps, le NGS permet de mieux caractériser les banques, y compris avant les étapes de sélection. Cette technique permet notamment de suivre l'enrichissement des séquences d'intérêt tout au long des processus de sélection <sup>247</sup>. Plusieurs échantillons peuvent être séquencés au sein d'un même *run* grâce à l'emploi de codes-barres qui permettent l'identification des échantillons. Il est également possible de choisir l'ordre de grandeur du nombre de lectures par échantillon.

La grande profondeur de séquençage et la possibilité de travailler sur des populations ont ouvert de nombreuses nouvelles possibilités. Il est ainsi possible de suivre la fréquence de chaque clone au sein de la population tout au long du processus de sélection, chose qui n'était pas possible avec le séquençage Sanger où seulement une petite fraction des clones pouvait être séquencée. Dans les cas classiques d'une sélection d'anticorps en *phage* ou *yeast display*, seule une partie des clones de l'étape de sélection finale était auparavant séquencée, or les meilleurs clones ne sont pas nécessairement les plus abondants à l'issue de la sélection mais ceux qui ont le facteur d'enrichissement le plus fort. Une autre application nouvelle permise par le NGS consiste à identifier la contribution de chaque acide aminé d'une protéine à des propriétés fonctionnelles diverses, comme son expression, sa fonction ou sa spécificité. Cette technique appelée Deep Mutational Scanning (DMS) est particulièrement utile pour identifier les résidus clés dans la séquence d'une protéine et est un outil précieux pour les études fonctionnelles.

## C. Deep Mutational Scanning

### 1. Principe

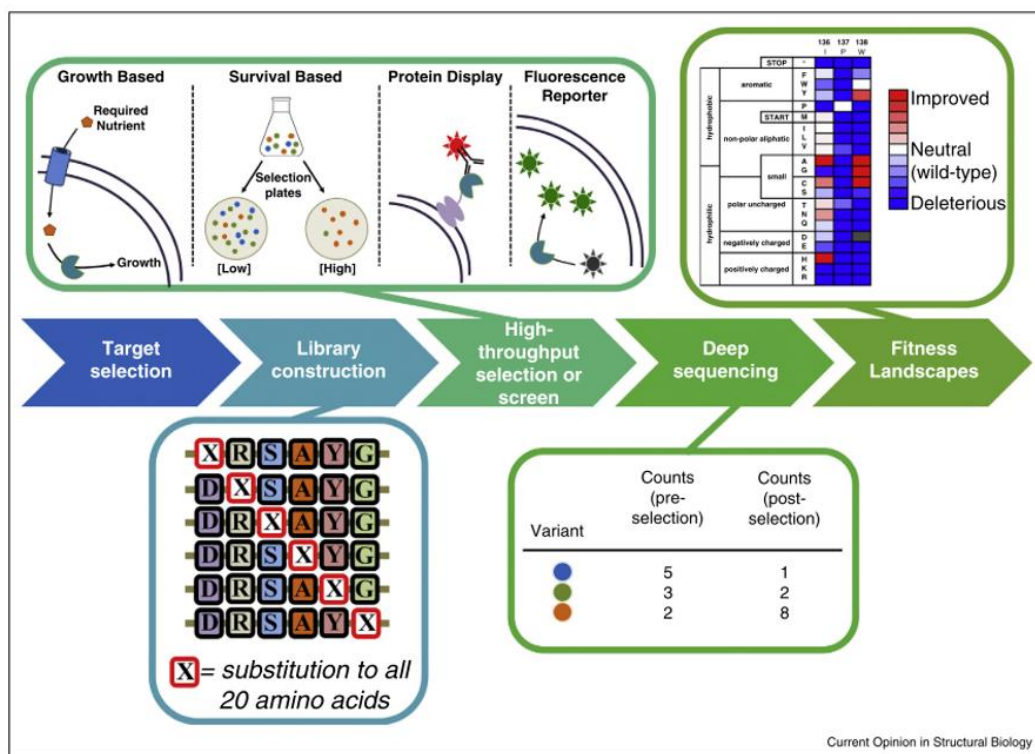
Le DMS est une méthode récente dans le domaine de l'ingénierie des protéines. L'objectif de cette méthode est d'évaluer l'impact de la substitution de chaque résidu d'une protéine d'intérêt par les 19 autres acides aminés naturels sur la fonctionnalité de la protéine. Il s'agit de comparer la fréquence de chaque mutation dans une population triée sur un critère d'intérêt (expression, affinité, sélectivité...) par rapport à sa fréquence dans la population non triée.

Les 5 étapes nécessaires à la mise en œuvre d'un DMS sont résumées en **Figure 24** <sup>248</sup> :

- La première étape consiste à définir les positions qui vont être mutées. Il peut s'agir de la protéine entière ou de zones particulières comme les CDR dans le cas des anticorps.
- La deuxième étape consiste à construire la banque de variants. La particularité des banques de DMS est que chaque individu de la banque est muté sur un seul codon, il s'agit donc de banques de mono-mutants. Il existe 3 techniques décrites dans la littérature pour générer une banque de DMS. La première technique consiste à générer des banques de saturation pour chaque position testée par PCR à l'aide d'amorces dégénérées (codons NNS/NNK). La deuxième technique appelée « pfunkel mutagenesis » utilise un système d'ADN uracilé en

combinaison avec des amorces dégénérées pour générer la banque <sup>249</sup>. La dernière méthode (et la plus récente) est appelée « plamide based one pot saturation mutagenesis » <sup>250</sup>. Cette méthode est basée sur la dégradation sélective d'un des deux brins d'ADN d'un plasmide codant pour la protéine d'intérêt grâce à des enzymes de restriction qui ne coupent qu'un des deux brins du plasmide. Des amorces dégénérées sont ensuite utilisées (une pour chaque position) pour reconstituer le plasmide et générer les mono-mutants.

- La troisième étape est celle du criblage fonctionnel de la banque dans un système choisi par l'expérimentateur.
- Lors de la quatrième étape, la population sélectionnée ainsi que la population de départ sont séquencées par NGS.
- Lors de la dernière étape, les données de séquençage sont traitées afin d'obtenir une valeur d'enrichissement pour chaque mutant. Ces valeurs sont rassemblées dans un tableau résumant l'impact de chacun des 20 acides aminés naturels à chaque position sur l'interaction étudiée.



**Figure 24 : Vue générale des étapes du DMS.** Une librairie d'ADN codant pour des mono-mutants d'une protéine d'intérêt est générée. Cette librairie est criblée par une méthode de criblage à haut débit comme le YSD couplé au FACS. Les fréquences des différents variants sont comparées avant et après sélection grâce au séquençage NGS. Ces fréquences sont utilisées pour générer une carte de score d'enrichissement des différentes mono-mutations tout au long de la séquence de la protéine d'intérêt. Adapté de Wrenbeck *et al* <sup>250</sup>.

## 2. Exemples d'applications

### i. Epitope mapping

De nombreux exemples de déterminations d'épitopes utilisant le DMS ont été décrits dans la littérature<sup>251–253</sup>. Dans ces expériences, c'est l'antigène qui est exprimé et muté. Le principe de l'*epitope mapping* par DMS est de trouver les positions qui une fois mutées réduisent voire abolissent la liaison de l'anticorps avec son antigène. Ainsi, Doolan et Colby ont déterminé les épitopes d'anticorps conformationnels dirigés contre des protéines prions. Van Blarcom *et al.* ont réalisé l'*epitope mapping* d'un panel d'anticorps dirigés contre une toxine de *S. aureus*. Et enfin, Kowalsky *et al.* ont appliqué la méthode à l'identification d'épitopes de trois couples anticorps/antigène différents. Les auteurs ont retrouvé les épitopes conformationnels des trois anticorps et leurs résultats étaient conformes aux données structurales déjà publiées.

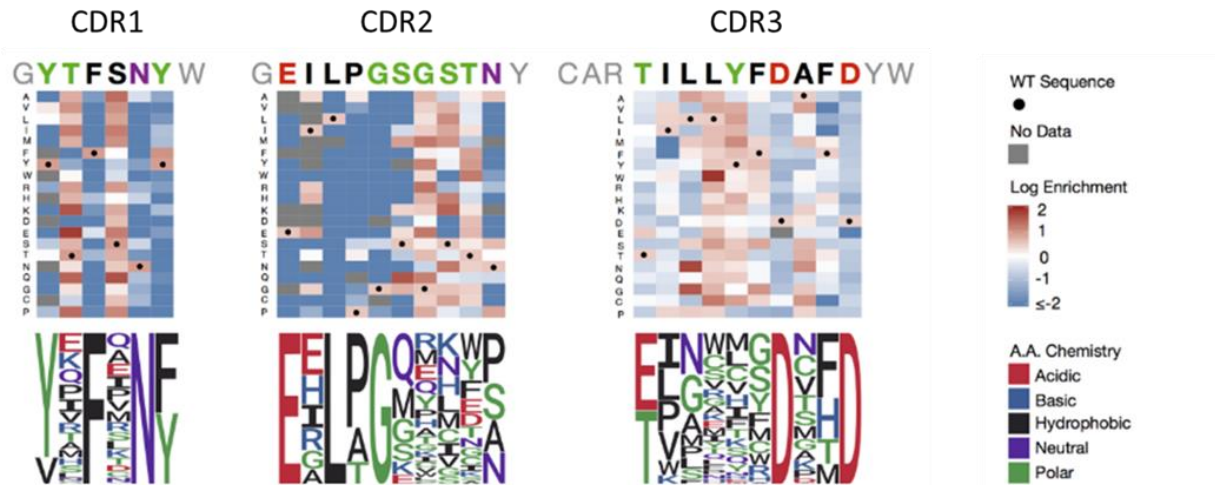
### ii. Maturation d'affinité

Le DMS peut également être utilisé directement sur l'anticorps pour de la maturation d'affinité. Il en existe deux exemples en mammalian display<sup>215,254</sup>. Dans le travail de Mason *et al.*, un DMS a été réalisé sur les trois CDR de la chaîne lourde d'un anticorps anti-HEL (Hen egg lysozyme) afin de déterminer les positions importantes dans l'interaction ainsi que pour identifier les mutations susceptibles d'améliorer l'affinité de l'anticorps (**Figure 25**). Dans cet exemple, trois variants (CDR3) ont été sélectionnés (TILWYFDAFD, TINLYFDAFD, TIGLYFDAFD) et leur affinité était améliorée significativement comparée à la séquence d'origine (les auteurs ne précisent pas les valeurs d'affinité).

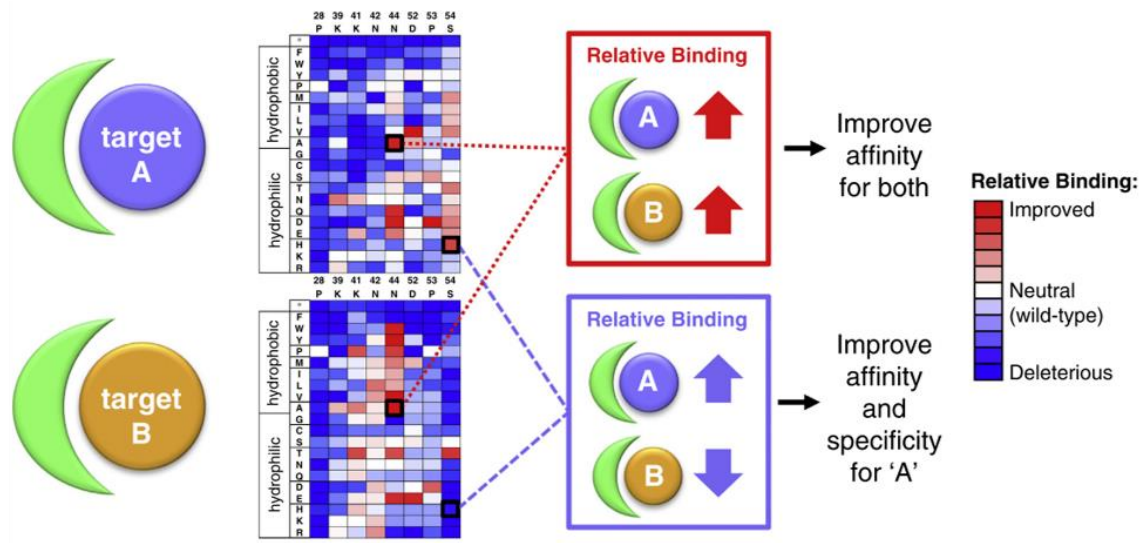
### iii. Ingénierie de la sélectivité

Une autre application du DMS porte sur l'ingénierie de la sélectivité des anticorps, dont le principe est présenté en **Figure 26**. Le DMS est réalisé en utilisant au moins deux antigènes différents. Un exemple important est celui de l'ingénierie de l'affinité d'un Fab reconnaissant deux protéines différentes n'ayant aucun lien de parenté : l'angiotensine 2 et le VEGF<sup>255</sup>. Les mutations améliorant l'affinité pour un antigène sans impacter celle de l'autre antigène ont été combinées. Plusieurs Fabs possédant la double spécificité avec une affinité inférieure au nanomolaire ont ainsi été identifiés.





**Figure 25 : Exemple de DMS sur les CDR de la chaîne lourde d'un anticorps anti-HEL24.** Le DMS est réalisé en mammalian display. On distingue deux types de positions : celles pour lesquelles aucun autre acide aminé que le *wild-type* n'est autorisé (enrichissement des mutants inférieur à 0) et celles qui sont permissives (enrichissement supérieur à 0). Il est également possible de distinguer ponctuellement des mutants avec de forts scores d'enrichissement comme TILWYFDAFD, TILNLYFDAFD et TILGLYFDAFD dans le CDR3. La permissivité de chaque position peut également être représentée à l'aide de WebLogos. Adapté de Mason *et al* <sup>214</sup>.



**Figure 26 : Ingénierie de la spécificité d'un anticorps par DMS.** Un DMS est réalisé sur les CDR de l'anticorps en utilisant chacun des deux antigènes cibles séparément. Une carte d'enrichissement est générée pour chaque antigène. Il existe alors trois types de mutations : celles qui améliorent l'affinité pour les deux antigènes, celles qui améliorent l'affinité pour un seul des deux antigènes et celles qui sont délétères pour la liaison des deux antigènes. Adapté de Wrenbeck <sup>248</sup>.



# Résultats

**Partie 1 : Détermination de l'épitope d'un anticorps protecteur anti-Salmonelle/Shigelle.**

**Partie 2 : Etude du format d'expression des anticorps à la surface de la levure**

**Partie 3 : Ingénierie de l'anticorps 318 : humanisation et dé-immunisation**

## Partie 1 : Détermination de l'épitope d'un anticorps protecteur anti-Salmonelle/Shigelle.

### I. Introduction de l'article

L'article présenté ci-après (en cours de soumission) présente l'obtention et la caractérisation d'un anticorps anti-Salmonelle/Shigelle protecteur dirigé à la fois contre les facteurs de virulence SipD de *Salmonella* et IpaD de *Shigella* qui sont des protéines de la coiffe de l'aiguille d'injection du système de sécrétion de type III de ces bactéries.

La première partie de cet article porte sur l'obtention de l'anticorps IpaD-318, travail qui a été réalisé avant ma thèse dans le cadre de la thèse de Bakhos Jneid. Dans cette partie, des anticorps capables de reconnaître à la fois les protéines SipD et IpaD ont été obtenus, et nous avons pu montrer que l'un d'eux (IpaD-318) était capable d'induire une protection croisée contre une infection à Salmonelle ou Shigelle. Aucun vaccin ou anticorps, capable d'induire une protection large spectre contre deux genres bactériens différents n'a jusqu'alors été décrit.

La seconde partie de ce travail décrit la détermination de l'épitope de l'anticorps IpaD-318 réalisée dans le cadre de ma thèse. Avant d'effectuer le travail présenté dans l'article, nous avons cherché à localiser l'épitope de liaison de l'anticorps IpaD-318 par d'autres moyens. Nous nous sommes d'abord basés sur l'alignement des séquences des deux antigènes et nous avons identifié des régions identiques dans les parties N-terminales d'IpaD et SipD (Figure EV4 de l'article). Des expériences de mutagenèse sur cette région conservée ont été réalisées et aucune perte d'affinité de l'anticorps pour les antigènes mutés n'a été observée. Nous avons ensuite directement travaillé sur la structure des deux antigènes en cherchant un patch d'acides aminés situés en surface et conservés entre les deux protéines. Nous avions également à notre disposition les différentes hypothèses sur la fonction des antigènes IpaD et SipD dans le mécanisme de virulence de Shigelle et Salmonelle. Nous avons émis l'hypothèse que l'anticorps pouvait reconnaître un pentamère d'IpaD ou SipD et que l'épitope devait se trouver sur la face externe du canal formé par le pentamère (pour une question d'accessibilité). Nous avons trouvé un patch d'acides aminés au niveau du début de l'hélice  $\alpha 7$  (voir Figure EV4 de l'article) qui répondait à ces critères. Des mutations ponctuelles sur IpaD et SipD dans cette zone ont aboli la capacité de l'anticorps à reconnaître les deux antigènes. L'épitope déterminé avec précision dans la publication montre en effet que ces acides aminés sont bien présents dans l'épitope de l'anticorps. En revanche, contrairement à l'hypothèse de départ, l'épitope de l'anticorps semble plutôt localisé sur la face interne du canal formé par IpaD ou SipD, posant alors la question de l'accessibilité de l'épitope pour l'anticorps.

# An antibody targeting type III secretion system induces broad protection against *Salmonella* and *Shigella* infections

Raphaël Sierocki<sup>1\*</sup>, Bakhos Jneid<sup>2\*</sup>, Audrey Rouaix<sup>2</sup>, Marc Plaisance<sup>2</sup>, Cécile Féraudet-Tarisse<sup>2</sup>, Bernard Maillère<sup>1</sup>, Hervé Nozach<sup>1</sup> and Stéphanie Simon<sup>2</sup>

<sup>1</sup> Service d'Ingénierie Moléculaire des Protéines, CEA, Université Paris-Saclay, Gif-sur-Yvette, France.

<sup>2</sup> Service de Pharmacologie et Immunoanalyse (SPI), Laboratoire d'Etudes et de Recherches en Immunoanalyse, CEA, INRA, Université Paris-Saclay, Gif-sur-Yvette, France.

\*The authors contributed equally to this work

## Abstract

*Salmonella* and *Shigella* species are food- and waterborne pathogens that are responsible for enteric infections in both humans and animals and are still the major cause of morbidity and mortality in the emerging countries. The existence of multiple *Salmonella* and *Shigella* serotypes as well as the emergence of strains resistant to antibiotics requires the development of broadly protective therapies. Those bacteria utilize a type III secretion system (T3SS) necessary for their pathogenicity. The structural proteins composing the T3SS are common to all virulent *Salmonella* and *Shigella* spp., particularly the needle-tip proteins SipD (*Salmonella*) and IpaD (*Shigella*). We have shown that these proteins are able to induce cross-protection in a murine model of infection by *Salmonella* and *Shigella* and that a monoclonal antibody directed against a conserved common region of their apical part provides good cross-protection efficacy. Our study provides the first *in vivo* experimental evidence of the importance of this common region in the mechanism of virulence of *Salmonella* and *Shigella* and opens the way to the development of cross-protective therapeutic agents.

**Keywords:** T3SS; *Salmonella*; *Shigella*; Broad-spectrum protection; Antibody

## Introduction

*Salmonella* and *Shigella* are Gram-negative enteropathogenic bacteria belonging to the Enterobacteriaceae family <sup>1,2</sup>. Both are responsible for gastrointestinal diseases ranging from moderate to acute, depending on different factors (e.g. pathogen species, ingested dose, and immune status of the host). *Salmonella* and *Shigella* infections are usually responsible for self-limiting gastroenteritis in healthy humans of high-income countries <sup>3,4</sup>. However, they continue to remain a serious health hazard in South and South-East Asia and African countries <sup>5-9</sup>, causing notably severe diarrhea in children under the age of five in sub-Saharan Africa and south Asia <sup>10-12</sup>. Other at-risk populations include military personnel deployed abroad <sup>13-15</sup>, travelers, and victims of bioterrorist attacks <sup>16,17</sup>. While *Salmonella* and *Shigella* consist of few species (two for *Salmonella*: *S. enterica* and *S. bongori* and four for *Shigella*: *S. flexneri*, *S. sonnei*, *S. dysenteriae* and *S. boydii*), there is a multiplicity of subspecies <sup>18-20</sup>, which makes it difficult to develop broad range vaccines.

Currently, three types of *Salmonella* vaccines are licensed: all of them target *S. typhi* and do not offer cross-protection against other *Salmonella* serovars, or against non-typhoidal *Salmonella*. The situation is even worse for *Shigella* for which no licensed vaccine is available despite long-standing efforts. Hopefully, these efforts will pay off in the near future as regards the clinical trials currently being evaluated worldwide <sup>21-23</sup>.

Vaccine strategies can be grouped into two fundamental approaches: live-attenuated vaccines and nonliving vaccines. Live-attenuated vaccines are generally more effective in stimulating the immune response, but generally do not induce broad coverage. Non-living vaccines encompass inactivated whole organisms or purified recombinant subunits. While offering the safest protection, they suffer from lower immunogenicity and generally require supporting strategies to overcome this hurdle <sup>24-27</sup>. Active immune system stimulation induced by vaccination takes days to weeks to be effective and can only be used to prevent infections. As opposed to active immunotherapy, passive immunotherapy provides immediate protection and is more suitable for therapeutic purposes. Although rarely utilized for infectious diseases, and particularly for bacterial

diseases for which there are other therapeutic solutions, the emergence and rapid spread of bacterial strains resistant to multiple antibiotics<sup>28</sup> might change this situation. Because T3SS is essential for virulence and is conserved among all pathogenic *Salmonella* and *Shigella* strains<sup>29</sup>, T3SS proteins appear to be ideal candidates for *Salmonella-Shigella* vaccine and immunotherapy development.

T3SSs or injectisomes are bacterial macromolecular organelles that are involved in the pathogenesis of many important human, animal and plant diseases<sup>30</sup>. Bacteria that have long been in contact with eukaryotic hosts have adapted to survive and reproduce in this environment. Analysis of these interactions leads to in-depth knowledge of the assembly of the injectisome and protein secretion, but also of the mechanisms of pathogenesis (Brown et al, 2005). Components of T3SSs are widely distributed in gram-negative pathogens and are well conserved with regard to their overall structure, architecture, and function. These injectisomes are composed of a basal body that traverses the inner and outer bacterial membrane and a needle-like complex that emerges at its apical end, through which effectors are secreted<sup>31</sup>. The T3SS needle of *Salmonella* and *Shigella* is built by the helical polymerization of several hundred subunits of a single small protein (PrgI and MxiH respectively). The needle-tip is formed by a multimeric hydrophilic protein complex (SipD and IpaD, respectively) connecting the bacterium to the host membrane through the needle<sup>32–34</sup>. During infection, the bacteria receive an external signal from the host environment and begin to assemble coordinately the constituents of the secretion system<sup>35,36</sup> which ultimately leads to the injection of effectors and/or invasion of the targeted host cell by the bacterium<sup>37–42</sup>. Based on the literature and our results, the needle-tip proteins have proved to be immunogenic in mice and in humans, able to elicit good humoral responses protective against salmonellosis and shigellosis<sup>43–47</sup>. Moreover, the sequence identity between IpaD and SipD<sup>48</sup>, led us to the hypothesis that those needle-tip proteins might be suitable targets for the development of cross-protective immunity against *Shigella/Salmonella*.

With this aim, we examined the immunogenicity of the *Salmonella* (SipD) and *Shigella* (IpaD) proteins administered alone by comparing intranasal (IN) and intragastric (IG) immunization routes in a mouse model. Furthermore, a murine monoclonal antibody (mAb) to IpaD was generated and evaluated *in vivo*. Cross-protective efficacy was determined against lethal oral intestinal infection with 100 LD50 of *Salmonella* Typhimurium or lethal intranasal infection with 100 LD50 of *Shigella flexneri* 2a. We provide the first demonstration that SipD-IpaD are both promising target antigens for a cross-protective *Salmonella-Shigella* vaccine and that a monoclonal anti-IpaD/SipD antibody is protective against *Salmonella* and *Shigella* infections in a murine model. Determination of its epitope of recognition sheds new light on the role of the needle tips proposed in the literature on the basis of structural models.

## Results

### Immunizations with SipD or IpaD proteins induce homologous and heterologous antibody responses

The SipD and IpaD proteins used to immunize mice were produced in *E. coli* BL21 (2.3 mg/L and 3 mg/L of culture of SipD and IpaD, respectively). Purity of proteins was assessed by SDS-PAGE electrophoresis and Coomassie blue staining (Fig S1).

Mice immunized by the IN or the intragastric (IG) route with SipD (Fig 1A, left panel, Fig EV1, left panels) or IpaD (Fig 1B, left panel, Fig EV1, left panels) developed antigen-specific humoral responses. Total Ig (G+M), IgG subtypes (G1, G2a and G2b) antibody concentrations as well as IgA titers were measured (principle of the ELISA in Fig S2A) and the results are summarized in Figure 1, Figure EV1 (kinetics) and Table S1. Whatever the routes of immunization, the specific antibody titers against IpaD were superior to those obtained with SipD (except for IgG1 by IG route). IpaD-specific Ig (G+M) concentrations reached the highest values by the IG route. To evaluate the induction of IgA antibodies by the mucosa, the first line of adaptive immune defense against enteric pathogens, SipD and IpaD specific IgA titers in serum from immunized and control mice were measured (Figs 1A and 1B second panels and Table S1). For each protein, the specific IgA titers were equivalent for mice immunized intranasally or intragastrically. It should be noted that for SipD some of the mice did not produce any detectable IgA, contrary to what was noted for IpaD, which supports what we observed for Ig(G+M) responses and the hypothesis of a better immunogenicity of IpaD protein. To investigate further the immune response elicited by both routes of immunization for both proteins, the SipD and IpaD homologous specific IgG1, IgG2a and IgG2b subclasses were measured in serum from immunized SipD, IpaD and control mice after the third immunization for the IN and IG routes (Table S1). Measurement of the IgG isotype concentrations in sera of immunized mice revealed that all main subclasses contributed to the humoral response whatever the route.

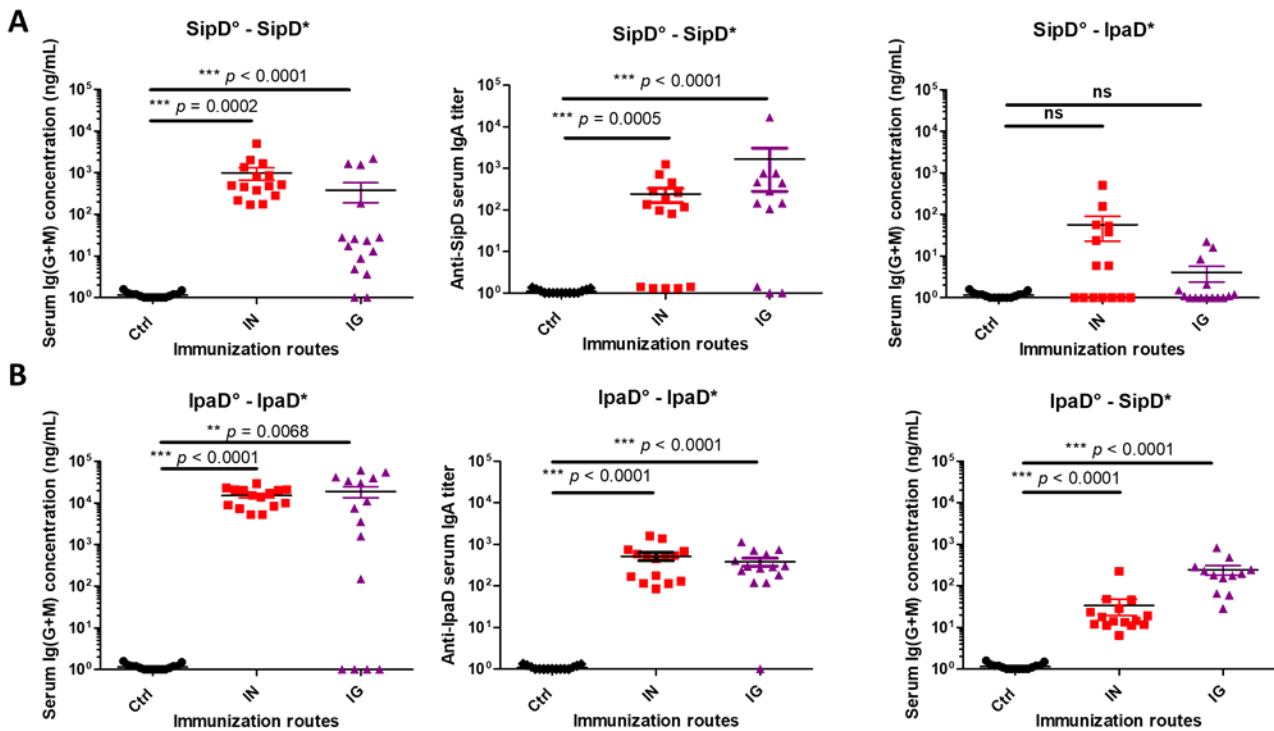
Because the needle-tip proteins SipD and IpaD of the T3SS of *Salmonella* and *Shigella* present sequence identities and with the aim of studying the possibility of cross-protection, the humoral responses against SipD in mice immunized with IpaD and against IpaD in mice immunized with SipD were measured. The crossed (heterologous) antibody responses were significantly lower (approximately 100-fold) than the specific (homologous) responses (Figs 1A and 1B, right panels, Fig EV1, right panels). IpaD immunogen seems to induce a higher heterologous antibody response against SipD than the opposite and particularly by the IG route (Table S1), confirming a better immunogenicity of IpaD protein in agreement with the higher production of homologous anti-IpaD antibodies compared to homologous anti-SipD antibodies.

## Protective efficacy against lethal *S. Typhimurium* or *S. flexneri 2a* challenge

The lethal doses 50% (LD50) of the *S. Typhimurium* (intragastric infection) and *S. flexneri 2a* (intranasal infection) strains used in the experiments (see experimental procedures) were determined at  $10^4$  CFU/mL for *Salmonella* and  $5.10^8$  CFU/mL for *Shigella* (Fig S3) according to the Reed and Muench method<sup>49</sup>. To assess first the homologous protective efficacy induced by SipD against *S. Typhimurium* and IpaD against *S. flexneri 2a*, immunized and control mice were subjected to intragastric or intranasal challenge, six weeks after the last immunization, with a high dose of bacteria:  $\sim 100$  LD50 of *S. Typhimurium* ( $10^6$  CFU/mL) or *S. flexneri 2a* ( $5.10^{10}$  CFU/mL) (Figs 2A and 2B, respectively, and Table S2). In all challenges, the mortality rate of control animals (mice administered phosphate buffer saline [PBS]/adjuvant) was 100% with death occurring at 16-21 days after challenge by *S. Typhimurium* and at 8-13 days after challenge by *S. flexneri 2a*. SipD and IpaD were able to induce efficient homologous protection against challenge by their bacterial counterparts (Figs 2A and 2B). The best homologous protective efficacy was induced by IpaD against a *S. flexneri 2a* challenge (intragastric route, 61% survival

rate). In order to evaluate the cross-protective efficacy of each of the proteins, mice immunized intragastrically or intranasally with IpaD or SipD were challenged by *S. Typhimurium* and *S. flexneri 2a*, respectively (Figs 2C and 2D, and Table S2). Weak cross-protection induced by IpaD was obtained against *S. Typhimurium* infection by the IN and IG routes (27% and 30%, respectively). Cross-protection induced by SipD against *S. flexneri 2a* challenge was significant and even superior to the homologous protection induced by IpaD, whatever the route of immunization (47% by IN route, 67% by IG route). We hypothesized that these cross-protections might be due to the production of specific antibodies directed against common crucial regions shared by the two proteins.

To go further in the hypothesis of a potential cross-protective role of the humoral response induced by needle-tip SipD and IpaD proteins, we decided to produce mAbs recognizing identical conserved regions on the surface of each of the two proteins, and to evaluate their protective capacity against *Salmonella* or *Shigella* infections.



**Figure 1. Serum Ig(G+M) concentrations, IgA titers of mice immunized with SipD (A) or IpaD (B).** Specific (left), cross (right) serum Ig (G+M) antibodies for SipD or IpaD and IgA titers (middle) were quantified by sandwich ELISA 2 weeks after the last immunization as described in experimental procedures. Data represent mean concentrations (ng/mL) and the standard errors (SEM) from 15 individual mice per group. Asterisks indicate P values: \*\*\*  $p < 0.001$  and \*\*  $0.001 < p < 0.01$  when comparing mice immunized by the IN or IG route versus control mice using a nonparametric Mann-Whitney test (exact P value indicated in the figure). °: indicates injected immunogen; \*: indicates biotinylated recombinant protein, ns: non-significant.



## Production and selection of anti-IpaD/SipD mAbs

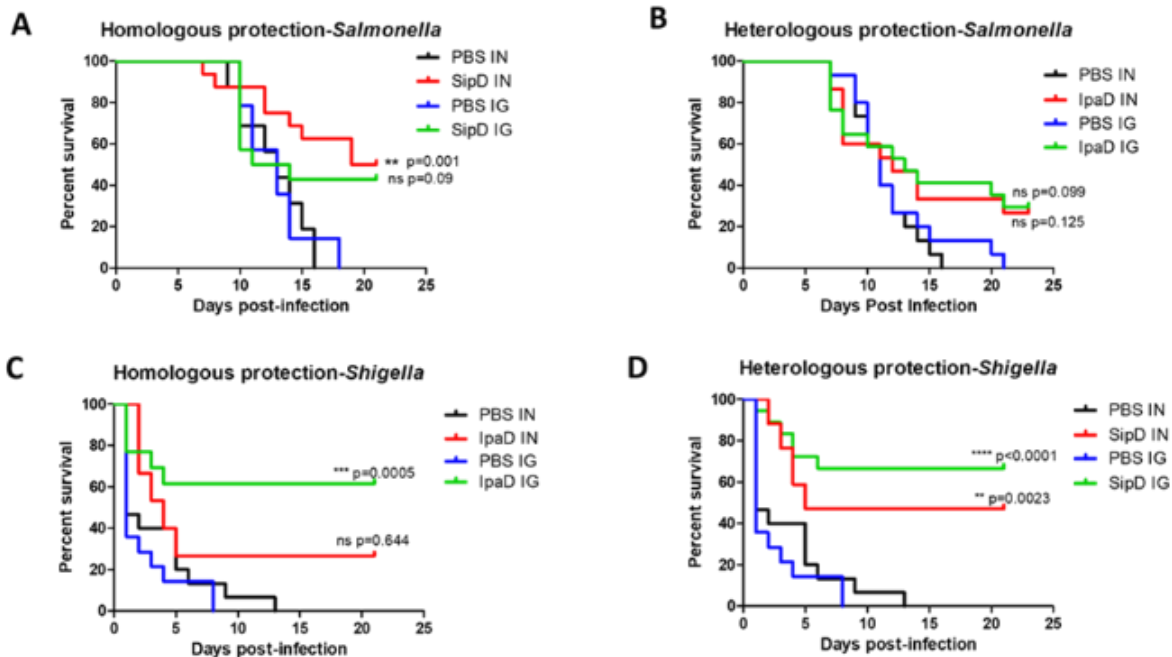
Mice immunized with IpaD and presenting the highest polyclonal antibody titers specific for IpaD (Fig S4) were chosen for fusion of spleen cells with myeloma cells according to the Köhler and Milstein method<sup>50</sup>. As the aim of this study was to obtain antibodies able to recognize homologous sequences of IpaD and SipD of *S. Typhimurium* and *S. flexneri 2a*, selection of antibodies was performed by differential-screening ELISA, using IpaD and SipD biotinylated recombinant proteins (Figure S2A). These screenings led to a final selection of 20 mAbs (Table S3). In order to confirm that the selected mAbs were able to recognize the IpaD and SipD recombinant proteins, an analysis by immunoblot was performed. The immunoblot analysis showed that all tested mAbs recognized a band of approximately 37 kDa (theoretical molecular weight of the native recombinant proteins: 34 kDa for IpaD and 35.1 kDa for SipD) as well as other higher bands (around 75 kDa and higher), that we supposed to be the protein in an oligomerized and not completely denatured form (Fig S5). In fact, these proteins naturally oligomerize and the proportion of multimeric proteins seems to be higher for IpaD than for SipD after denaturation, suggesting a higher quantity of oligomerized IpaD. These mAbs are all able to cross-

recognize both IpaD and SipD proteins, confirming their selection on this criterion during the ELISA screening.

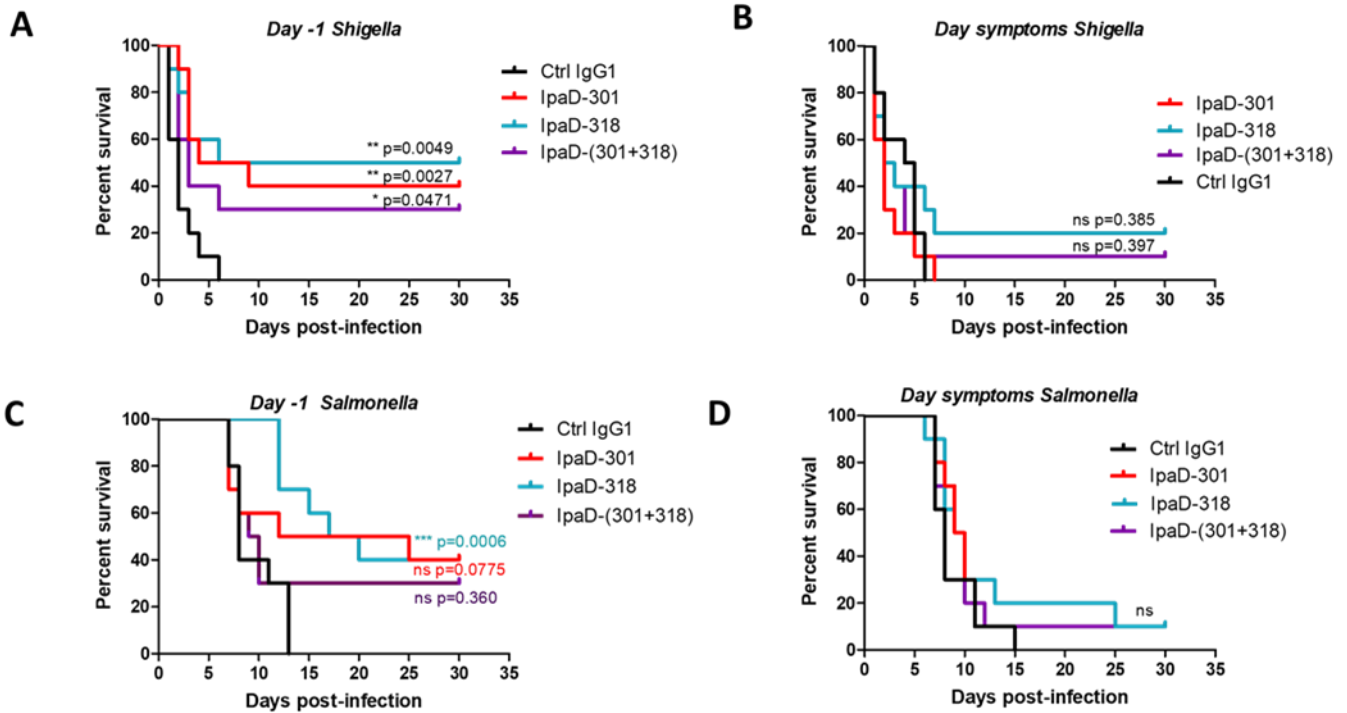
In order to select mAbs able to neutralize bacteria *in vivo*, the 20 mAbs recognizing both IpaD and SipD were injected twice (500 µg each injection) by the intraperitoneal route (IP) into BALB/c mice (3 mice/antibody): first time, 24 hours before intragastric infection with 100 LD<sub>50</sub> of *S. Typhimurium* and, a second time, 4 days after the infection when the signs started to appear. Survival was monitored for 30 days and the percentage protection induced by each antibody is presented in Table S5. This first experiment allowed us to select 7 antibodies able to provide a protection and to retest them (one injection of 300 µg IP 24 h before challenge with 100 LD<sub>50</sub> of *S. Typhimurium*) using ten mice per group (Fig EV2). The two best antibodies IpaD-301 and IpaD-318 mAbs provided 40% of protection and were selected for further analyses.

## IpaD-301 and IpaD-318 mAbs conferred protection against *S. flexneri 2a* and *S. Typhimurium* challenge

The ability of the IpaD-301 and IpaD-318 antibodies to protect against both *S. flexneri 2a* and *S. Typhimurium* was studied using our mouse model (Fig EV2 and Fig 3). When



**Figure 2. Homologous and heterologous protective efficacies induced by SipD and IpaD immunizations against *S. Typhimurium* and *S. flexneri 2a* challenges.** Mice (N = 15) were immunized at days 0, 21 and 42 by the indicated antigens and routes. Six weeks after the last immunization, at day 84, 100 LD<sub>50</sub> of *S. Typhimurium* (A, B) or *S. flexneri 2a* (C, D) was administered (intragastrically and intranasally, respectively) to SipD (A, D) or IpaD (B, C) immunized mice. Survival was monitored for 21 days. Statistical significance was determined using a log-rank (Mantel-Cox) test. Statistically significant differences are indicated by \*\*\* p < 0.001 and \*\* 0.001 < p < 0.01 compared to PBS groups (exact p values indicated in the figure and Table 2).



**Figure 3. *In vivo* protective activity of anti-IpaD/SipD antibodies against *S. flexneri* 2a and *S. Typhimurium* challenges.** Mice (N=10) were injected with the indicated anti-IpaD/SipD antibodies 24 hours before (A, C) the intragastric (*S. Typhimurium*, A, B) or intranasal (*S. flexneri* 2a, C, D) 100 LD 50 challenge and at the onset of the signs (B, D). Survival was monitored for 30 days. Statistical analysis was performed using a log-rank (Mantel-Cox) test comparing mice treated with mAbs versus control mice treated by a control non-relevant IgG1 antibody.

antibodies were administered to mice 24 hours before the 100 LD50 *S. flexneri* 2a intranasal challenge, IpaD-301 and IpaD-318 antibodies conferred 40% and 50% protection, respectively (Fig 3A and Table S4). The antibody combination (IpaD-301 together with IpaD-318) did not increase the protective efficacy (30% protection). When mice were treated 24 h after the *Shigella* challenge, corresponding to the onset of the signs, the protection conferred by IpaD-318 dropped to a low and non-significant protection level of 20% (Fig 3B and Table S4).

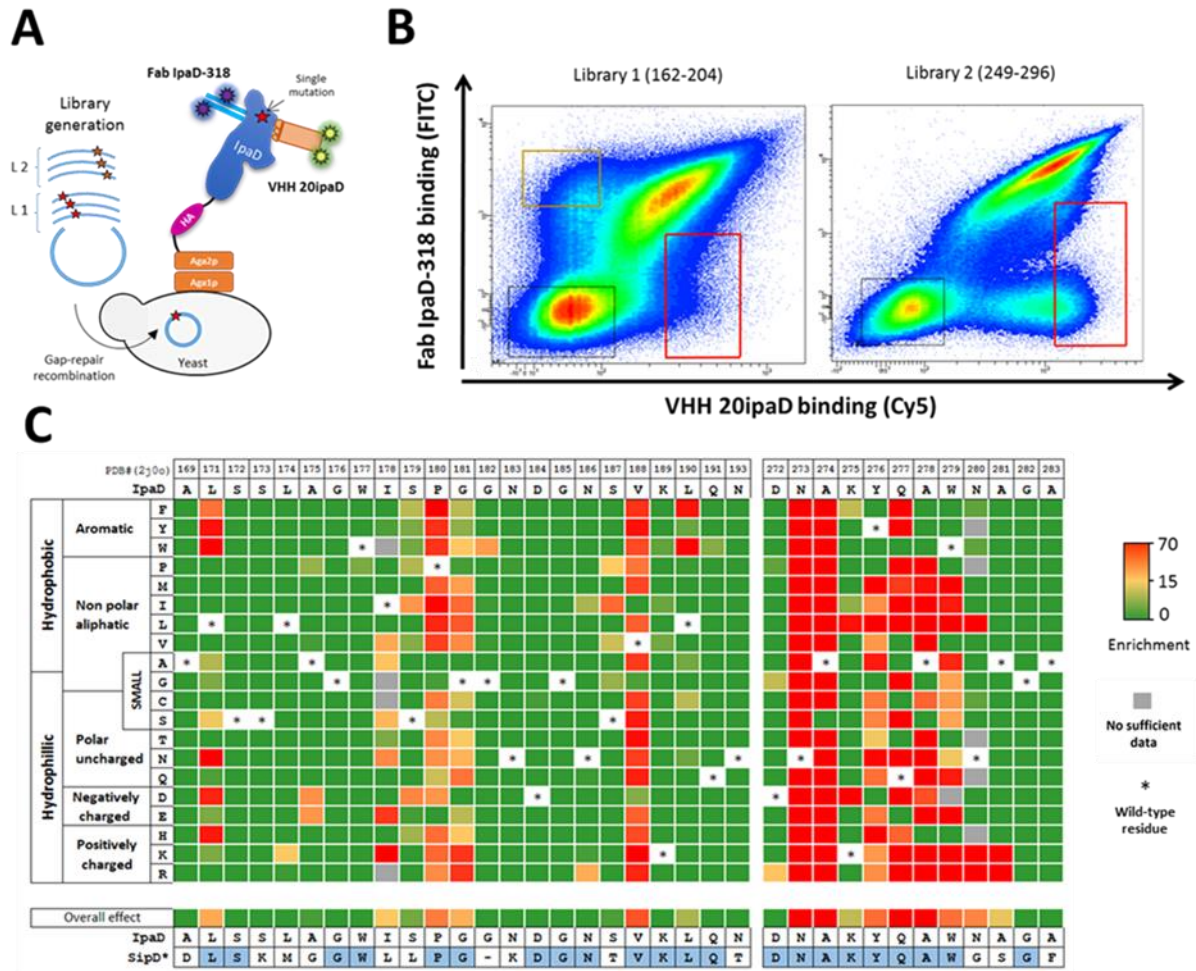
Similarly, IpaD-301 and IpaD-318 provided the same level of protection against an intragastric infection with 100 LD 50 of *S. Typhimurium* (40%, Fig 3C and Table S4) when injected one day before the challenge by *Salmonella*. As for *S. flexneri*, this protective effect was completely abolished when these mAbs were injected at the onset of the signs (5 days after challenge, Fig 3D and Table S4), probably for the same reasons as observed for *S. flexneri*. As for *S. flexneri*, no increase in effectiveness was observed with the combination IpaD-301 and IpaD-318, suggesting that those antibodies might share a common epitope on SipD/IpaD.

For a better understanding of these *in vivo* results, particularly in regard to the potential delay in the availability of antibodies after their administration by the intraperitoneal route, and their time window of action, we measured their pharmacokinetics. The half-lives of the mAbs IpaD-301 and IpaD-318 were measured in plasma after

intraperitoneal injection of antibodies in mice by sandwich ELISA (principle of the test illustrated in Fig S2B) at different times post-injection (see experimental procedures). As shown in Figure S6, an initial rapid plasma increase was observed within the first hours after injection for the two antibodies, corresponding to the antibody transfer from the peritoneum to the blood compartment. The peak was reached approximately 18-24 hours post-injection for the two antibodies, followed by a decrease during the subsequent five weeks. Values were included in a pharmacokinetic model and allowed evaluation of a half-life of 12 and 10 days for IpaD-301 and IpaD-318, respectively. These results shed new light on the results obtained in the *in vivo* experiments. Indeed, the antibodies administered one day before infection reached maximum plasma concentration only the day of the infection. Moreover, a delay of 18 h to 24 h after the onset of the signs should be kept in mind and might partly explain the decrease in efficacy of the antibodies. Because of the greater efficacy of IpaD-318, we decided to go further in the analysis of its mechanism of action.

### Epitope of mAb IpaD-318 is conserved among different *Salmonella* and *Shigella*

To gain insight into the mode of action of mAb IpaD-318, we sought to determine the localization of its epitope on



**Figure 4. Epitope mapping of Fab IpaD-318 on the protein IpaD from *Shigella flexneri*.** (A) General principle of functional screening by yeast surface display. A DNA library of single mutants of IpaD was transformed into yeast using gap repair recombination. Clones harboring IpaD mutations abolishing the binding of Fab IpaD-318 were sorted in FACS. To control the folding of the mutated IpaD antigen, a conformational VHH anti-IpaD compatible with Fab IpaD-318 was used (VHH 20IpaD, PDB #5VXM). (B) Bivariate flow cytometry analysis of libraries L1 and L2 of yeast cells expressing IpaD single amino acid variants on their surface. Each single dot on the plot designates the fluorescence intensity values for a single yeast cell corresponding to Fab IpaD-318 binding (FITC in y-axis) and VHH-20IpaD (Cy5 in x-axis) binding, respectively. Gates (red frames) were designed to sort yeast cells expressing IpaD mutants with altered affinity for Fab IpaD-318 and preserved affinity for VHH-20IpaD. (C) Heatmap representing enrichment values of each IpaD single mutant after functional sorting. For clarity, only the amino acids 169-193 and 272-283 are represented (see complete heatmap in fig. EV4). SipD\*: structural alignment of SipD from *Salmonella enterica* Typhimurium (PDB #3NZZ) with IpaD (PDB #2J00), identical residues are colored in blue. Overall effect corresponds to the mean of the enrichment values for all possible mutations at a selected position.

IpaD/SipD using a yeast surface display (YSD) approach. IpaD was expressed on the surface of *S. cerevisiae* yeast cells in order to screen the ability of antibodies to bind the antigen (Fig 4A).

In a first experiment, we assessed the good folding of IpaD presented on the surface of yeast cells with two conformational VHHs already published and for which epitopes on IpaD have been determined<sup>51</sup>). We found that both VHHs were able to recognize the protein at the surface

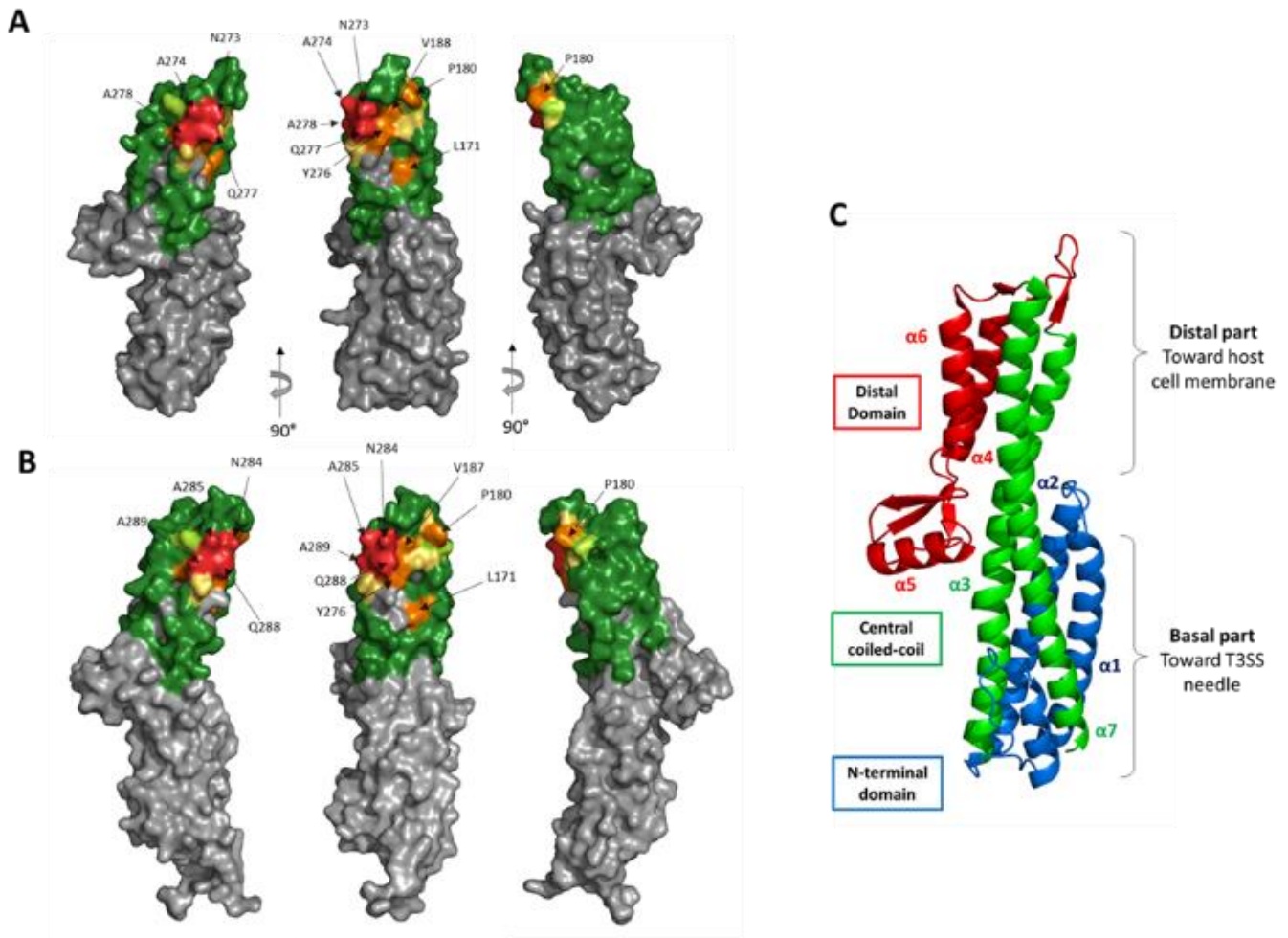
of the yeasts, with high affinity (Fig S7). Moreover, competition and compatibility assays showed that mAb IpaD-318 competed with VHH-JMK-H2 but was compatible with VHH-20IpaD for its binding to wild-type IpaD (Fig S8). Competition of mAb IpaD-318 with VHH-JMK-H2 suggests that these molecules might exhibit overlapping epitopes in the distal part of IpaD (opposite the N- and C-terminals<sup>52</sup>).



To obtain high-resolution conformational epitope mapping, we used deep mutational scanning (DMS) associated with YSD to identify key IpaD amino acid positions necessary for IpaD-318 binding<sup>53–58</sup>. This method consists of the functional assessment of every possible amino acid change at each position in a protein. For DMS analysis, we selected residues located in the distal part of IpaD and accessible to solvent (exposed area > 10 Å<sup>2</sup>). As the distal part of IpaD is encoded by two separate regions in the amino acid sequence, two libraries were designed: library 1 (L1) from positions 162 to 204 (39 mutated residues) and library 2 (L2) from positions 249 to 296 (39 mutated residues). Site-saturation mutagenesis (SSM) libraries were generated following the “plasmid one pot saturation mutagenesis” method<sup>59</sup> to allow any of the 20 natural amino acids at the selected position using an NNS degenerate codon. DNA libraries were transformed into EBY100 yeast cells by gap repair recombination with the pCT-L751 expression plasmid (Fig 4A).

Libraries of yeast cells were induced to express the mutants on their surface before a functional screen in FACS to identify clones harboring mutations abolishing Fab IpaD-318 binding. The conformational VHH-20IpaD, which is compatible with Fab IpaD-318, was used as a control to monitor the proper folding of mutants.

After labeling, analysis by fluorescence-activated cell sorting (FACS) revealed that the majority of cells expressing IpaD mutants had a strong fluorescence signal for both Fab IpaD-318 and VHH-20IpaD binding (Fig 4B). Variants with strong loss of Fab IpaD-318 binding were found in both libraries, although in different proportions, the vast majority of them belonging to library 2 encompassing residues 249 to 296. Conversely, a population exhibiting a loss of binding for VHH-20IpaD while maintaining binding for Fab IpaD-318 was observed in library 1 corresponding to residues 162 to 204, in accordance with the available crystallographic structure<sup>51</sup>. A gating strategy (in red, Fig 4B) was applied to select the fraction corresponding to the 5% of cells with the lowest Fab



**Figure 5. Structural representation of the epitope of mAb IpaD-318 on the antigen IpaD as determined by yeast display. (A)** Residues were colored on the structure of IpaD (PDB #2J00) by the overall effect color code of Fig. 4C, representing the importance of each tested residue in the interaction. Grey residues were not tested. **(B)** Epitope representation of mAb IpaD-318 on the antigen SipD (PDB #3NZZ). **(C)** Structural description of IpaD. Domain segmentation of the protein as described in Barta *et al.*, 2017.

IpaD-318 binding and a preserved VHH-20IpaD binding. Corresponding variants were sorted and cells were recovered in selective medium. Plasmid DNA of each library, before and after functional sorting, was extracted and prepared for deep sequencing on an Illumina MiSeq device (2\*150bp) with at least 750,000 reads per population. For every mutant, enrichment ratios between unsorted and sorted libraries were calculated. Enrichment ratios were coded in color for each mutant and represented as a functionality heatmap (Fig 4C for partial heatmap and Fig EV3 for complete heatmap). These ratios relative to the original binding interaction indicate whether a mutation causes a significant loss of binding (red), a moderate loss of binding (yellow) or no significant change in binding (green). When mutated, nine positions affect particularly the binding of Fab IpaD-318: P180, G181, V188, N273, A274, Y276, Q277, A278 and W279. Positions L171, I178, N280 and A281 also have a significant effect on the interaction when mutated. The alignment of the IpaD sequences of the main *Shigella* serovars (*flexneri*, *sonnei* and *dysenteriae*) as well as of SipD of the main *Salmonella* serovars (Typhi, Paratyphi, Typhimurium (including ST313) and Enteritidis) shows that the amino acids involved in the IpaD-318 epitope are all conserved in these different species (Fig EV4). In particular, eight of the nine main positions determined as the most important in DMS (P180, G181, N273, A274, Y276, Q277, A278 and W279) are strictly conserved. Only valine in position 188 is replaced by isoleucine in *S. Enteritidis*, Paratyphi, Typhi and Typhimurium ST313. According to DMS data, this replacement is the only one that has no impact on antibody binding, suggesting that Fab IpaD-318 should be able to recognize IpaD/SipD from all above-mentioned species/serovars. To confirm this hypothesis, IpaD/SipD from all these species/serovars were expressed at the surface of yeasts and evaluated for Fab IpaD-318 binding (Fig EV5). The measured apparent  $K_D$  showed no difference between all species/serovars with values ranging from 403 pM to 638 pM. The average mutation effect for each position was determined and represented on IpaD structure (Fig 5A). The epitope of IpaD-318 determined by YSD is located in the distal part of IpaD, on the concave side of the protein. Key amino acids represented on the structure belong mainly to the apical end of the two central coiled-coil helices (residues 273-278) and to the apical loop between positions 175 and 191. These results are in good agreement with the epitope determined for the VHH-JMK-H2<sup>51</sup> that competes with Fab IpaD-318 for IpaD binding.

## Discussion

Infections caused by *Shigella* and *Salmonella* (typhoidal as well as nontyphoidal, and particularly invasive nontyphoidal *Salmonella* (iNTS)) are associated with a high burden in terms of mortality and morbidity especially in low income countries and in children under 5 years of age<sup>60</sup>. For this reason, long-standing efforts have been made to understand

the immunological mechanisms underlying these infections and to develop effective therapies against them. Vaccines targeting typhoidal *Salmonella* are already marketed, but none protect against non-typhoidal *Salmonella*. No licensed vaccine exists for *Shigella*, though some developments have been the subject of clinical studies with varying degrees of success. The existence of multiple *Shigella* and *Salmonella* serotypes and the increase of multiresistant iNTS as well as *Shigella* clones highlight the need for development of a broad-spectrum protective vaccine<sup>61,62</sup>.

There are many studies showing the importance of the humoral response in the fight against *Salmonella* and *Shigella* infections<sup>63-66</sup>. For *Shigella*, numerous data provide evidence of the immunogenicity/protective role of T3SS proteins and particularly IpaD<sup>46,67-71</sup>, which has been evaluated as part of a bi-component recombinant vaccine<sup>47</sup>. For *Salmonella*, studies on the importance of a protective humoral response are very scarce and sometimes controversial (for review see<sup>72</sup>). We have shown in a preliminary study that SipD induced a good humoral response and was protective against a *S. Typhimurium* challenge<sup>73</sup>, and more recently Martinez-Becerra and coll. have shown that two fusion proteins mixed together and composed of SipB/SipD and SseB/SseC have the potential to provide a cross-protective effect against two serovars of *Salmonella enterica*<sup>45</sup>. However, the SipB/SipD fusion protein alone was unable to elicit protection.

Based on these data and because of the sequence identity, the strong similarity in the three-dimensional structures and the mechanism of action between SipD and IpaD, as well as the role of the humoral response against these proteins and their importance in protecting against *Shigella* and *Salmonella* infections, we have hypothesized that broad-spectrum cross-protection against *Salmonella* and *Shigella* infections can be induced by using SipD or IpaD as immunogen. The results of this study show that by using indifferently SipD or IpaD, good protection (60%) against *Shigella flexneri* 2a infection is obtained despite very high challenging doses (100 LD50). In a comparative study, immunizations with IpaD yielded 70 to 90% cross-protection against 5 and 11 LD50 of *S. sonnei* and *S. flexneri*, respectively, which decreased dramatically to around 20% with 9 and 24 LD50 of *S. sonnei* and *S. flexneri*, respectively<sup>74</sup>. Interestingly, we found in this study that protection against *S. flexneri* was equivalent using SipD or IpaD, whereas the immune responses induced by SipD were lower than those induced by IpaD. This difference might be due to better immunogenicity of IpaD compared to SipD. This hypothesis is supported by the results obtained with IG immunizations with SipD for which the specific Ig (G+M) responses are more heterogeneous and much lower (two logs) than those obtained for IpaD, which could be partly explained by the heterogeneity in the degradation of the proteins by the gastric acid of the stomach. In addition, while *Shigella* infections are carried out intranasally, protection is better when immunizations are performed intragastrically. This may be related, among other things, to a slightly higher

IgA titer by the IG route than by the IN route, induced by immunizations with SipD.

Although IpaD is also able to induce protection against *Salmonella* infection (100 LD50), it is nevertheless lower than that obtained for *Shigella* with SipD. This might be due to different factors that could be linked altogether: i) *Salmonella* has two type three secretion systems involved in the pathogenicity, ii) pathogenicity mechanisms are different between *Salmonella* and *Shigella* and particularly in regard to the involvement of the innate and adaptative immune response, iii) a more preponderant role of the intracellular cycle of *Salmonella* compared to *Shigella* and iv) a faster systemic dissemination of *Salmonella* in the murine model. Nevertheless, the protective effect obtained using SipD/IpaD as immunogen underscores the importance of the extracellular life cycles of *Salmonella* and *Shigella* for their pathogenicity and dissemination and highlight the role of conserved regions of needle-tip proteins SipD/IpaD in this protection.

To go deeper into the importance of the humoral response against conserved regions of the SipD and IpaD proteins and to provide evidence of its broad protective role, we decided to produce mAbs directed against conserved epitopes of both proteins and to evaluate their cross-protective effect *in vivo*. We selected the mAbs based on their ability to bind both proteins indifferently, on a common epitope at their surface. The most efficient antibodies IpaD-301 and IpaD-318 provided good protection against bacterial infections when administered intraperitoneally to mice one day before challenge by *S. Typhimurium* or *S. flexneri* (40 and 50%, respectively), showing for the first time that mAbs directed to shared epitopes of SipD/IpaD are sufficient to induce good protection against high challenges of *Salmonella* and *Shigella* and confirming the cross-protection observed after immunization by either of the two proteins against infection by either of these pathogens. Administration of a mixture of both antibodies did not improve the protection. Protection fell dramatically when each of the antibodies was administered at the onset of the symptoms (24 h or 5 days after the challenge by *Shigella* and *Salmonella*, respectively), suggesting that multiplication of the bacteria was too great given the very high challenge dose and/or their accessibility to antibodies was impaired. This is reinforced by the pharmacokinetics showing that the circulating antibody concentration only peaks 24 h after administration, thus delaying their neutralizing action.

To our knowledge, this is the first description of mAbs able to provide cross-protection *in vivo* against a high challenge of *Salmonella* and *Shigella* infections and directed against shared epitopes of IpaD and SipD T3SS needle-tip proteins. For this reason, we tried to determine the binding epitope of mAb IpaD-318 on IpaD with the aim of understanding how anti-IpaD/SipD antibodies protect against the virulence of *Salmonella/Shigella* pathogens.

The global structure of the T3SS is conserved between pathogens especially between *Salmonella*, *Shigella* and *Burkholderia*. The T3SS apparatus is made of three main

components: the basal body anchored to the bacterial inner and outer membranes, a needle structure and a multimeric tip complex (TC) including IpaD for *Shigella*<sup>51,75–80</sup> and SipD for *Salmonella*<sup>78,81,82</sup>. The tip proteins IpaD and SipD assemble at the distal end of the needle and form a pentameric ring that prevents premature secretion of effectors until exposure to small molecules such as bile salts<sup>52,83,84</sup>, which triggers a structural rearrangement and recruitment of the hydrophobic translocator protein IpaB<sup>83</sup> for *Shigella* and SipB<sup>85</sup> for *Salmonella* to the tip complex to prepare host membrane interaction. Upon contact with host membrane, the second hydrophobic translocator IpaC for *Shigella* or its counterpart SipC for *Salmonella* is recruited to the tip of the *Shigella* T3SS, thereby promoting translocon pore formation and induction of effector protein secretion into the host cell followed by pathogen invasion<sup>86</sup>.

Multiple models have been proposed for the structure of the tip complex. The structure and the mechanism of action have not been formally demonstrated, but all models agree on the formation of a pentamer at the distal end of the needle containing at least four molecules of IpaD. However, it remains unclear whether the fifth member of the pentamer is another IpaD molecule or the first translocator protein IpaB and no data are available regarding SipD. The most recent electron microscopy study of nascent apparatuses conducted by Blocker and co-workers<sup>87</sup> hypothesizes that active tip complexes carry one IpaB for four IpaD, which fits the electron map densities of tip complexes. They showed that homopentameric and heteropentameric tip complexes coexist in wild-type strains, hypothesizing that the heteropentamer is the active state of the complex and that one IpaB replaces one IpaD in the pentamer after the activation signal. As the complete structure of IpaB is not yet known, although it has recently been partially solved by Barta *et al*<sup>88</sup>, Blocker and co-workers proposed a structure of 4 IpaD lacking the fifth subunit supposed to be IpaB (PDB #4D3E).

The epitope of mAb IpaD-318 is composed of discontinuous amino acids in the peptide sequence, leading to a conformational epitope that is highly conserved among lots of serovars/species of *Salmonella* and *Shigella*, as shown in the sequence alignments for *S. sonnei*, *flexneri*, *dysenteriae* and *S. Typhimurium*, Enteritidis, Paratyphi and Typhi, and as confirmed by the binding experiments and  $K_D$  measurements that are very similar for the different IpaD and SipD of these bacteria. These results allow us to be confident about the ability of this antibody to protect against the serotypes not tested in this study. The epitope is located in the distal part of IpaD/SipD, which is the side of the tip protein directed towards the host membrane, leading to the hypothesis that IpaD-318 could interfere with (i) IpaD-IpaD/SipD-SipD oligomerization, (ii) IpaD-IpaB/SipD-SipB or IpaD-IpaC/SipD-SipC binding or (iii) accessibility of the membrane of the host cell due to steric hindrance. Recently, Barta *et al*<sup>51</sup> generated and characterized four anti-IpaD VHHs. They evaluated the potential of *Shigella*-mediated hemolysis inhibition of their VHHs *in vitro* and also solved their structure in complex with

IpaD. The epitope of one of the VHHs tested (JMK-H2) overlaps but is not identical to that of IpaD-318. In contrast to IpaD-318, which is protective against both *Salmonella* and *Shigella* infections in mice, JMK-H2 was not able to inhibit *Shigella*-mediated hemolysis in *in vitro* experiments. This could be explained by the difference in epitope localization of the VHH compared to mAb IpaD-318. The amino acids of the epitope of JMK-H2, which are not common with those of IpaD-318, are located in the central part (coiled-coil), positioned on the alpha 3 helix. It can then be assumed that in the IpaD pentamer the alpha 3 helix is masked, thus preventing binding of the VHH. The amino acids of the mAb IpaD-318 epitope, located on the loop between the alpha 6 and alpha 7 helices, render them accessible and would explain why IpaD is neutralized by our antibody while JMK-H2 is not. Another explanation might rely on the size difference between VHHs and on a complete antibody resulting in greater steric hindrance.

Given the position of the epitope of mAb IpaD-318 represented in Figure S9 on the complex proposed by Blocker and co-workers, it seems unlikely that IpaD-318 could bind after complete oligomerization of IpaD at the top of the tip platform. IpaD-318 might preferentially prevent interaction between IpaD subunits one by one or IpaB recruitment at the interface between IpaD and IpaB in complex formation. Other studies<sup>51,52,79</sup> suggest that the globular distal domain of IpaD pivots upwards, thus switching the TC from an inactive to active form. As residues of the hinge region are within the epitope, in this hypothesis IpaD-318 could interfere with the movement of the globular distal domain and freeze the TC in its inactive form, blocking its capacity for signal transduction, or could prevent the interaction of IpaD with IpaB after the movement of the globular domain.

In a previous study, Roehrich and coworkers found point mutations in IpaD that led to impairment of effector secretion induction and invasion<sup>89</sup>. Among the 4 most affected mutants, 3 are located at the top of the TC: N186, N273 and Q277. N273 and Q277 belong to and N186 borders the IpaD-318 epitope. These residues could be involved in a structural rearrangement of subunits at the interface with IpaB, leading us to hypothesize that the mAb IpaD-318 might block it and could therefore prevent T3SS activation and secretion by making TCs unable to sense host cells.

Because SipD and IpaD share the same epitope and because of their equivalent protection by the mAb IpaD-318, it should be underlined that this antibody also helps to provide new insights into the role and mechanism of action of SipD in host cell infection by *Salmonella*, which is much less described than that of *Shigella* and which is confirmed to be very similar to that of *Shigella*. In particular, a heteropentamer of SipB/SipD at the tip of T3SS has never been described for *Salmonella* and results obtained in this study could help to gain further insights into the molecular organization of the *Salmonella* tip complex. Interestingly, it has already been demonstrated that IpaB interacts with SipD in certain conditions<sup>90</sup>.

In conclusion, despite relatively weak identity sequence (40%), we have been able to obtain an antibody directed against conserved sequences of needle-tip proteins of T3SS which on its own confers cross-genus protection against *Salmonella* and *Shigella* infections. Considering its epitope, the sequence alignment and binding experiments, this antibody should provide protection against most *Salmonella* serovars (*S. Typhi*, *S. Typhimurium*, *S. Enteritidis*, *S. Paratyphi*, and S.ST313 clone of *S. Typhimurium* known to be responsible for invasive nontyphoidal salmonellosis in sub-Saharan Africa) and *Shigella* species (*S. flexneri*, *S. dysenteriae*, *S. sonnei*) known to be pathogenic in humans. This is the first mAb that is protective against both *Shigella* and *Salmonella* infections *in vivo* in mice. The region targeted by this mAb is essential to the pathogen's virulence and could explain why this region is conserved among *Salmonella* and *Shigella* bacteria. Far from becoming a treatment of shigellosis and salmonellosis, IpaD-318 proved for the first time that antibodies directed against specific and conserved regions of needle-tip proteins alone are essential to provide good cross-protection against high doses of *Salmonella* and *Shigella*. This type of treatment could be very useful in a context of increased multidrug resistance and might help to reduce minimal inhibitory concentrations by a synergistic effect. Moreover, these results have shed new light and provided for the first time *in vivo* evidence of the very similar mechanism of action proposed by structural models of these proteins. The identification of this conserved epitope among lots of serotypes of *Salmonella* and *Shigella* pathogenic for humans and for which no universal vaccine exists might help in the design and development of small molecules targeting this region and helping to prevent infection by those enteropathogenic bacteria and more generally by other bacteria sharing the same mechanism of invasion.

## Materials and Methods

### Ethics statement

Six- to 8-week-old female BALB/c mice from Janvier Labs, France were maintained in accordance with French and European regulations on care and protection of laboratory animals (European Community [EC] Directive 86/609, French Law 2001-486, 6 June 2001) and with the agreement of the ethics committee (CETEA) no. 15-055 delivered to S. Simon and agreement D-91-272-106 from the Veterinary Inspection Department of Essonne (France). Up to eight mice were kept in each cage and housed in a temperature-regulated room and had free access to food and water. All animal experiments minimized suffering in line with the guideline of the CETEA committee.

### Bacterial strains

The *Salmonella enterica* serovar Typhimurium (CIP 104474, Pasteur Institute collection) and *Shigella flexneri* 2a



(generous gift from Dr A. Phalipon, Pasteur Institute) were used in this study. Both strains were first grown overnight in 5 mL of Luria broth (LB) with agitation (200 rpm) at 37°C and then grown with 1:100 of the first culture for 2 h under the same conditions.

## Reagents

Biotin N-hydroxysuccinimide ester and streptavidin were from Sigma-Aldrich. Goat anti-mouse IgG and IgM polyclonal antibodies were from Jackson ImmunoResearch (Interchim, France). Goat anti mouse IgA polyclonal antibodies were from SouthernBiotech (Clinisciences, France). Rat anti-IgG1, -IgG2a and -IgG2b monoclonal antibodies were from Bio-Rad (AbD Serotec). ELISAs were performed with MaxiSorp 96-well microtiter plates (Nunc, Thermoscientific), and all reagents were diluted in enzyme immunoassay (EIA) buffer (0.1 M phosphate buffer [pH 7.4] containing 0.15 M NaCl, 0.1% bovine serum albumin [BSA], and 0.01% sodium azide). Plates coated with proteins were saturated in EIA buffer (18 h at 4°C) and washed with washing buffer (0.01 M potassium phosphate [pH 7.4] containing 0.05% Tween 20). AEBSF (serine protease inhibitor) was from Interchim. Spectra/Por dialysis membranes were from Spectrum Laboratories. Cholera toxin and LB were from Sigma. PBS was from Gibco by Life Technologies.

## Recombinant SipD and IpaD production and immunizations

The SipD and IpaD genes from, respectively, *S. Typhimurium* and *S. flexneri* were synthesized (Genecust) based on the published sequences of *Salmonella* strain CIP 104474 and of *Shigella* strain CIP 82.48T and cloned into *NdeI/XhoI* restriction sites of the IPTG inducible pET22b vector (Novagen), allowing insertion of a poly-histidine tag sequence at the 3' end of the genes. SipD and IpaD were expressed and purified by affinity chromatography (Ni-NTA) as described previously<sup>73</sup>. Protein concentrations were determined by measuring absorbance at 280 nm ( $A_{280}$ ) using the NanoDrop Spectrophotometer and the purity was assessed by SDS PAGE (10-15% gradient Phast Gel, Phast system, GE Healthcare). Purified recombinant proteins were stored at -20°C until use.

Six- to 8-week-old female BALB/c mice were used in groups of 15. For IN immunizations, mice were anesthetized with isoflurane delivered through a vaporizer. Mice were immunized intranasally or intragastrically on days 0, 21 and 42 with 10 µg of SipD or IpaD in 20 µL of PBS (IN) or 300 µg in 200 µL of PBS (IG). The proteins, admixed with 1.5 µg (IN) or 10 µg (IG) cholera toxin adjuvant, were incubated for 1 h in a shaker at room temperature before immunization. Mice that received only adjuvant and PBS were included as controls. Animals were monitored daily after immunizations.

## LD50 determination and challenge procedures

### i) LD50 determination.

5 mL of preculture of *S. Typhimurium* or *S. flexneri* 2a was grown in 200 mL of LB at 37°C with agitation (200 rpm) until OD<sub>600 nm</sub> ~1. Bacteria were centrifuged at 2,000 x g for 15 min at 4°C and pellets were resuspended in PBS. Serial dilutions were performed in sterile PBS and approximately 2.102 to 2.108 CFU of *S. Typhimurium* were administered intragastrically (200 µL) using a curved gavage needle, or 5.105 to 5.1010 CFU of *S. flexneri* 2a were administered intranasally (20 µL) to 20- to 22-week-old female BALB/c mice (5 mice per group). The exact number of CFUs of each challenge dose was recalculated by viable counts (plating serial dilutions on LB agar plates). Mice were monitored twice daily for 25 days. The 50% mouse lethal dose (LD 50) for the challenge strains was calculated by the method of Reed and Muench and determined to be ~104 CFU/mL for *S. Typhimurium* and ~ 5.108 CFU/mL for *S. flexneri* 2a.

### ii) Challenge.

On day 84 after primary immunization, mice (N=15 per group) were challenged with 100 LD 50 of virulent *S. Typhimurium* (~ 106 CFU/mL, 200 µL in sterile PBS) via the oral route or with 100 LD 50 of virulent *S. flexneri* 2a (~ 5.1010 CFU/mL, 20 µL in sterile PBS) via the IN route. Mice were monitored twice daily for 21 days after the challenge and health status, weight and survival were recorded. Any mouse that lost more than 20% of its initial body weight or showed advanced signs of morbidity was euthanized and scored as a death.

## Production of monoclonal antibodies

Six- to 8-week-old female BALB/c mice were immunized 3 times by the IN route with IpaD (20 µg in 100 µL of PBS) admixed with alum hydroxide (1:1). Mice were bled before the first immunization (S0, used as the negative control) and 2 weeks after each injection (S1, S2 and S3). The immune polyclonal response was evaluated by enzyme immunoassay (EIA) using goat anti-mouse Ig(G+M) as coated antibodies (see "Enzyme immunoassays" below). The two mice presenting the highest immune response were selected for preparation of mAbs and given two intravenous booster injections of IpaD recombinant protein 2 months after the last immunization. Two days after the last boost, spleen cells from mice were fused with melanoma NS1 cells as previously described<sup>46,47</sup>. The hybridoma culture supernatants were screened for the presence of anti-IpaD/SipD antibodies (that is to say antibodies recognizing both proteins) by EIA (see below). Selected hybridomas were subsequently cloned by limiting dilution, and mAbs were obtained after inducing ascitic fluid in BALB/c mice. MABs were further purified by affinity chromatography using protein A and dialyzed in 0.05 M phosphate buffer (pH 7.4). Purity was assessed by SDS PAGE and Coomassie blue staining.

## Evaluation of the neutralizing activity of anti-IpaD/SipD mAbs *in vivo*

One day after administration of anti-IpaD/SipD mAbs (300 µg/mouse) by the intraperitoneal route, separately or in combination (for IpaD-301 and IpaD-318), mice were challenged with bacteria at 100 LD 50 via IG (for *S. Typhimurium*) or IN (for *S. flexneri*) infection. The mAbs were also injected into mice after challenge and at onset of signs. Mice administered PBS or control IgG were used as controls. Challenged mice were monitored daily for body weight loss and any signs of sickness. Mice that were considered to have reached the experimental endpoint were euthanized. The number of mice surviving after 30 days of daily observation was used to determine the relative degree of protection.

## Enzyme immunoassays

### i) Labeling with biotin.

One hundred µg of mAb or recombinant protein (SipD or IpaD) in 400 µL borate buffer (0.1 M; pH 8.5) was incubated at a 1:20 molar ratio with biotin-N-hydroxysuccinimide ester dissolved in 6 µL of anhydrous dimethylformamide (DMF). The reaction was stopped after 30 min at RT by adding 100 µL of 1 M Tris-HCl (pH 8) for 30 min. Finally, 500 µL of EIA buffer was added and the preparation was stored frozen at -20°C until use.

### ii) Evaluation of polyclonal response, and screening of mAbs in hybridoma supernatants.

Anti-IpaD/SipD antibodies were measured in sera of immunized mice or hybridoma culture supernatants using sandwich ELISA. Briefly, microtiter plates were coated with 100 µL of goat anti-mouse Ig(G+M) or rat anti IgG1, or IgG2a or IgG2b antibodies at 10 µg/mL (diluted in 50 mM potassium phosphate buffer) overnight at RT. Plates were then saturated overnight at 4°C with 300 µL/well of EIA buffer. After a washing cycle performed with the washing buffer, the plates were incubated with 100 µL/well of each hybridoma culture supernatant or serial dilutions of mouse sera (from 10<sup>-2</sup> to 10<sup>-5</sup>) were added in duplicate and incubated overnight at 4°C (100 µL/well). The plates were then washed 3 times before adding 100 µL/well of biotinylated recombinant SipD or IpaD proteins at 100 ng/mL. After 2 hours of incubation at RT followed by three washing cycles, 100 µL/well of acetylcholinesterase (AChE; EC 3.1.1.7)-labeled streptavidin (1 Ellman unit/mL) was added and incubated for 1 hour at RT. Finally, the plates were washed 3 times and the absorbance was measured at 414 nm after 45 min of reaction with 200 µL/well of Ellman's reagent 92. Concentrations of Ig (G+M) antibodies were calculated by fitting a calibrated control curve with nonlinear regression and interpolation of absorbance values of test samples by two-phase decay analysis. For measurement of IgA antibodies specific to SipD and IpaD, the titers were measured as described above, using goat anti-mouse IgA

antibodies at 2.5 µg/mL for coating. IgA antibody titers were calculated as the reciprocal of the lowest sample dilution giving a signal equal to the average background signal (as measured for 8 wells) plus 10 standard deviations.

### iii) Concentration of IpaD-301 and IpaD-318 in plasma of mice (pharmacokinetic studies).

Measurement was performed using ELISA. 100 µL of recombinant protein IpaD was coated at 1 µg/mL overnight at room temperature. Plates were then saturated overnight at 4°C with 300 µL/well of EIA buffer. After a washing cycle performed with the washing buffer, the plates were incubated with 100 µL/well of serial dilutions of mouse sera (from 10<sup>-2</sup> to 10<sup>-5</sup>) added in duplicate and incubated overnight at 4°C (100 µL/well). The plates were then washed 3 times before adding 100 µL/well of tracer antibody (polyclonal anti Ig(G+M) linked to AChE) at 5 EU (Ellman Units)/mL. After 2 hours of incubation at RT followed by three washing cycles, the plates were washed 3 times and the absorbance was measured at 414 nm after 45 min of reaction with 200 µL/well of Ellman's reagent 92.

## Pharmacokinetic analyses of mAbs in mouse

Female BALB/c mice 6-8 weeks old (body weight 20 to 22 g) were used for the pharmacokinetic study. 50 µg of 0.22 µm filtered antibodies diluted in PBS buffer was administered intraperitoneally to mice (N=4). Blood was collected 1, 7, 24, 48, 96, 360, 576, 768 and 1176 h after mAb injection and centrifuged for 30 min at 3500 x g at 4°C. Plasma was recovered and stored at -20°C until use. Plasma concentration of antibodies was measured by ELISA. Pharmacokinetic parameters were determined using the mean antibody concentrations in mice (N=4) per time point. For calculation of *in vivo* blood clearance, data values were fitted using WinNonlin software (Pharsight).

## Western blot experiment

SipD or IpaD recombinant proteins were suspended in Laemmli buffer (containing 0.25 M Tris-HCl pH 6.8, 4% SDS, 40% glycerol, 0.1% bromophenol blue and 10% β-mercaptoethanol) and denatured for 5 min at 95°C. After migration of 10 ng and 100 ng/well of recombinant SipD or IpaD in SDS-PAGE for 1 h 30 min at 150 V in a 15% resolving gel, proteins were blotted onto a PVDF membrane (Amersham Biosciences) overnight at RT at 30 V. For the saturation step, the membranes were saturated with PBS containing 0.1% Tween 20 (PBST) and 5% skimmed dry milk for 30 min at RT. After two washes in PBST, specific mAbs (4 µg/mL in PBST containing 1% skimmed dry milk) were incubated for 30 min at RT with the membranes. After three washes in PBST, the membranes were reacted for 30 min at RT with HRP-labeled polyclonal goat anti-mouse immunoglobulins (Pierce) diluted to 1:2000 in PBST containing 3% skimmed dry milk. After three washes in PBST and a brief wash in PBS, bands were detected via

chemiluminescence (ECL, Amersham Biosciences) using a VersaDoc imaging system (Bio-Rad).

## Statistical analysis

Graph Pad Prism 5 was used for graphics generation and statistical analyses. Survival rates were analyzed using the log-rank (Mantel-Cox) test. Statistical analyses were performed using the nonparametric Mann-Whitney test to compare antibody concentrations between groups. Data are presented as the mean  $\pm$  standard errors SEM for 10 or 15 samples per group of mice. A P value  $< 0.05$  was considered significant in all determinations.

## Library design & generation

Two libraries of IpaD variants with single amino acid mutations were generated using the “plasmid one pot saturation mutagenesis” method of Wrenbeck et al.<sup>59</sup>. A library was generated for each of the two IpaD regions targeted (library 1: amino acids 162 to 204 and library 2: amino acids 249 to 296, residues numbering referring to PDB#2J00). Only solvent-exposed residues (exposed area  $> 10 \text{ \AA}^2$ ) within those regions were selected, with a theoretical diversity of each library of 1248 variants. After mutagenesis, PCR was performed to recover and amplify the library of IpaD single mutant genes. The final library was obtained by recombining mutant genes in the YSD plasmid pCT-L7.5.1 between NheI and BamHI restriction sites (kindly provided by K. Dane Wittrup, Addgene #42900).

## Yeast surface display

Preparation of competent yeast cells EBY100 (ATCC® MYA-4941) and library transformation were performed according to Benatui et al.<sup>93</sup>, with 2  $\mu\text{g}$  of digested vector and a molar ratio of 25:1 (linear library/digested vector). Gap repair transformations were made in plasmid pCT-L7.5.1 between restriction sites BamHI and NheI. Precultures were performed by inoculating 250 mL of SD-CAA [6.7 g/L yeast nitrogen base without casamino acids, 20 g/L glucose, 5 g/L casamino acids, 100 mM sodium phosphate, pH 6.0] medium with 400  $\mu\text{L}$  of transformed cells and incubated overnight at 30°C, 200 rpm. The saturated pre-culture (typically OD<sub>600</sub> of 8–10) was passaged in order to obtain an initial culture OD<sub>600</sub> of 0.25–0.50 in 50 mL. The culture was grown at 30°C until its OD<sub>600</sub> reached 0.5–1.0. Cells were centrifuged and re-suspended in 50 mL of SG-CAA galactose induction medium [6.7 g/L yeast nitrogen base without casamino acids, 20 g/L galactose, 5 g/L casamino acids, 100 mM sodium phosphate, pH 6.0] and induced for 16–36 h at 20°C, 200 rpm.

## Fab preparation

Fab IpaD-318 was obtained by digestion of a previously described chimeric version of mAb IpaD-318 using the Pierce™ Fab Micro Preparation Kit<sup>94</sup>.

## Flow cytometry

For library sorting, 10<sup>8</sup> induced cells of each library were washed with 1 mL PBSF (PBS, BSA 0.1%) buffer. Cells were resuspended in 50 mL of a solution containing 160 pM of Fab IpaD-318 and 600 pM for VHH-20IpaD (FITC conjugate). These concentrations correspond to the experimental  $\frac{1}{2} K_D$  for both ligands determined on the surface of yeast cells as previously described<sup>94</sup>. Apparent binding affinity  $K_{D \text{ app}}$  values for Fab IpaD-318 and VHH-20IpaD were 314 pM and 1.16 nM, respectively, in good accordance with previously reported values<sup>51,95</sup> (Fig S7).

After incubation at 20°C, with agitation for 4 hours, cells were washed with 1 mL ice-cold PBSF to avoid dissociation. Cells were incubated on ice in the dark for 15 minutes with anti-human Ck antibody (Invitrogen #MA1-10385, APC conjugate), washed with 1 mL ice-cold PBSF and sorted with a BD FACS Aria™ III cytometer using BD FACSDiva™ software. Gates were defined to sort the top 5% of cells with decreased binding for mAb IpaD-318 and maintained binding for VHH-20IpaD. At least 100-fold of the theoretical diversity was sampled and each gate was separately collected and recovered for 2 days in SD-CAA medium at 30°C.

## Deep sequencing and analysis of NGS data

Plasmid DNA of each yeast population was extracted and prepared for sequencing as described in Medina-Cucurella and Whitehead<sup>53</sup>. Two-step PCR was performed to amplify the region of interest and add Illumina adapters and barcodes for multiplexing (see sup data for primer sequences). Deep sequencing was performed with an Illumina MiSeq device (2x150 bp, v2 kit 300 cycles) with at least 250,000 reads per population. Reads were demultiplexed and each sample was processed separately using the Galaxy platform (<https://usegalaxy.org/>) using the functions described in Blankenberg et al.<sup>96</sup>. First, paired reads were joined (Fastq Joiner). A trim was then performed (Fastq Trimmer) on reads to keep just the region of interest in the correct frame. A quality filter (Filter FASTQ) was applied to eliminate reads with a minimum quality score under 30. Next, DNA sequences were translated in protein sequences and identical sequences were grouped. Sequences not repeated at least two times were filtered out. Using the software RStudio, single-mutants were selected to allow calculation of enrichment ratios for each single mutation.

## Acknowledgements

We thank Dr Armelle Phalipon for the generous gift of the *Shigella flexneri* 2a strain. B. Jneid and R. Sierocki were recipients of grants from the PhD program of the Commissariat à l’Energie Atomique et aux Energies Alternatives.

## Author contributions

RS and BJ: design, realization of the experiments and manuscript writing

AR, CFT, and MP: realization of the experiments

BM, HN, SS: design of the experiments, manuscript writing

## Conflict of interest

None.

## References

1. Velge, P. *et al.* Multiplicity of Salmonella entry mechanisms, a new paradigm for Salmonella pathogenesis. *MicrobiologyOpen* **1**, 243–258 (2012).
2. Martinez-Becerra, F. J. *et al.* Broadly protective Shigella vaccine based on type III secretion apparatus proteins. *Infect. Immun.* **80**, 1222–1231 (2012).
3. Feasey, N. A., Dougan, G., Kingsley, R. A., Heyderman, R. S. & Gordon, M. A. Invasive non-typhoidal salmonella disease: an emerging and neglected tropical disease in Africa. *Lancet Lond. Engl.* **379**, 2489–2499 (2012).
4. Niyogi, S. K. Shigellosis. *J. Microbiol. Seoul Korea* **43**, 133–143 (2005).
5. Kotloff, K. L. *et al.* Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet Lond. Engl.* **382**, 209–222 (2013).
6. Ao, T. T. *et al.* Global burden of invasive nontyphoidal Salmonella disease, 2010(1). *Emerg. Infect. Dis.* **21**, (2015).
7. Tennant, S. M., MacLennan, C. A., Simon, R., Martin, L. B. & Khan, M. I. Nontyphoidal salmonella disease: Current status of vaccine research and development. *Vaccine* **34**, 2907–2910 (2016).
8. Levine, M. M., Kotloff, K. L., Barry, E. M., Pasetti, M. F. & Sztein, M. B. Clinical trials of Shigella vaccines: two steps forward and one step back on a long, hard road. *Nat. Rev. Microbiol.* **5**, 540–553 (2007).
9. Levine, M. M., Kotloff, K. L., Nataro, J. P. & Muhsen, K. The Global Enteric Multicenter Study (GEMS): Impetus, Rationale, and Genesis. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **55**, S215–S224 (2012).
10. Farag, T. H. *et al.* Some Epidemiologic, Clinical, Microbiologic, and Organizational Assumptions That Influenced the Design and Performance of the Global Enteric Multicenter Study (GEMS). *Clin. Infect. Dis.* **55**, S225–S231 (2012).
11. Walker, C. L. F., Aryee, M. J., Boschi-Pinto, C. & Black, R. E. Estimating diarrhea mortality among young children in low and middle income countries. *PloS One* **7**, e29151 (2012).
12. Scallan, E., Mahon, B. E., Hoekstra, R. M. & Griffin, P. M. Estimates of illnesses, hospitalizations and deaths caused by major bacterial enteric pathogens in young children in the United States. *Pediatr. Infect. Dis. J.* **32**, 217–221 (2013).
13. Sanders, J. W. *et al.* An observational clinic-based study of diarrheal illness in deployed United States military personnel in Thailand: presentation and outcome of Campylobacter infection. *Am. J. Trop. Med. Hyg.* **67**, 533–538 (2002).
14. Kasper, M. R. *et al.* Infectious etiologies of acute febrile illness among patients seeking health care in south-central Cambodia. *Am. J. Trop. Med. Hyg.* **86**, 246–253 (2012).
15. Kasper, M. R. *et al.* Diarrhea outbreak during U.S. military training in El Salvador. *PloS One* **7**, e40404 (2012).
16. Ma, T., Heywood, A. & MacIntyre, C. R. Chinese travellers visiting friends and relatives--A review of infectious risks. *Travel Med. Infect. Dis.* **13**, 285–294 (2015).
17. Török, T. J. *et al.* A large community outbreak of salmonellosis caused by intentional contamination of restaurant salad bars. *JAMA* **278**, 389–395 (1997).
18. Ansong, C. *et al.* Experimental annotation of post-translational features and translated coding regions in the pathogen Salmonella Typhimurium. *BMC Genomics* **12**, 433 (2011).
19. Bäuml, A. J., Tsois, R. M., Ficht, T. A. & Adams, L. G. Evolution of host adaptation in Salmonella enterica. *Infect. Immun.* **66**, 4579–4587 (1998).
20. Phalipon, A. & Sansonetti, P. J. Shigella's ways of manipulating the host intestinal innate and adaptive immune system: a tool box for survival? *Immunol. Cell Biol.* **85**, 119–129 (2007).
21. Chen, W. H. & Kotloff, K. L. Shigella Vaccine Development: Finding the Path of Least Resistance. *Clin. Vaccine Immunol. CVI* **23**, 904–907 (2016).
22. Launay, O. *et al.* Safety Profile and Immunologic Responses of a Novel Vaccine Against Shigella sonnei Administered Intramuscularly, Intradermally and Intranasally: Results From Two Parallel Randomized Phase 1 Clinical Studies in Healthy Adult Volunteers in Europe. *EBioMedicine* **22**, 164–172 (2017).
23. Kotloff, K. L., Riddle, M. S., Platts-Mills, J. A., Pavlinac, P. & Zaidi, A. K. M. Shigellosis. *Lancet Lond. Engl.* (2017). doi:10.1016/S0140-6736(17)33296-8
24. MacLennan, C. A., Martin, L. B. & Micoli, F. Vaccines against invasive Salmonella disease: current status and future directions. *Hum. Vaccines Immunother.* **10**, 1478–1493 (2014).
25. Kaminski, R. W. & Oaks, E. V. Inactivated and subunit vaccines to prevent shigellosis. *Expert Rev. Vaccines* **8**, 1693–1704 (2009).
26. Tennant, S. M. & Levine, M. M. Live attenuated vaccines for invasive Salmonella infections. *Vaccine* **33** Suppl 3, C36–41 (2015).

27. Venkatesan, M. M. & Ranallo, R. T. Live-attenuated *Shigella* vaccines. *Expert Rev. Vaccines* **5**, 669–686 (2006).
28. Manohar, A., Ahuja, J. & Crane, J. K. Immunotherapy for Infectious Diseases: Past, Present, and Future. *Immunol. Invest.* **44**, 731–737 (2015).
29. Demers, J.-P. *et al.* The common structural architecture of *Shigella flexneri* and *Salmonella typhimurium* type three secretion needles. *PLoS Pathog.* **9**, e1003245 (2013).
30. Galan, J. E. *Salmonella* interactions with host cells: type III secretion at work. *Annu. Rev. Cell Dev. Biol.* **17**, 53–86 (2001).
31. Cornelis, G. R. The type III secretion injectisome. *Nat. Rev. Microbiol.* **4**, 811–825 (2006).
32. Moest, T. P. & Méresse, S. *Salmonella* T3SSs: Successful mission of the secret(ion) agents. *Curr. Opin. Microbiol.* **16**, 38–44 (2013).
33. Blocker, A. *et al.* The tripartite type III secretion of *Shigella flexneri* inserts IpaB and IpaC into host membranes. *J. Cell Biol.* **147**, 683–693 (1999).
34. Cordes, F. S. *et al.* Helical structure of the needle of the type III secretion system of *Shigella flexneri*. *J. Biol. Chem.* **278**, 17103–17107 (2003).
35. Burkinshaw, B. J. & Strynadka, N. C. J. Assembly and structure of the T3SS. *Biochim. Biophys. Acta* **1843**, 1649–1663 (2014).
36. Deng, W. *et al.* Assembly, structure, function and regulation of type III secretion systems. *Nat. Rev. Microbiol.* **15**, 323–337 (2017).
37. Francis, C. L., Starnbach, M. N. & Falkow, S. Morphological and cytoskeletal changes in epithelial cells occur immediately upon interaction with *Salmonella typhimurium* grown under low-oxygen conditions. *Mol. Microbiol.* **6**, 3077–3087 (1992).
38. Goosney, D. L., Knoechel, D. G. & Finlay, B. B. Enteropathogenic *E. coli*, *Salmonella*, and *Shigella*: masters of host cell cytoskeletal exploitation. *Emerg. Infect. Dis.* **5**, 216–223 (1999).
39. Parsot, C. *Shigella* type III secretion effectors: how, where, when, for what purposes? *Curr. Opin. Microbiol.* **12**, 110–116 (2009).
40. Campbell-Valois, F.-X. & Pontier, S. M. Implications of Spatiotemporal Regulation of *Shigella flexneri* Type Three Secretion Activity on Effector Functions: Think Globally, Act Locally. *Front. Cell. Infect. Microbiol.* **6**, 28 (2016).
41. Ashida, H., Mimuro, H. & Sasakawa, C. *Shigella* manipulates host immune responses by delivering effector proteins with specific roles. *Front. Immunol.* **6**, 219 (2015).
42. Ashida, H., Ogawa, M., Mimuro, H. & Sasakawa, C. *Shigella* infection of intestinal epithelium and circumvention of the host innate defense system. *Curr. Top. Microbiol. Immunol.* **337**, 231–255 (2009).
43. Markham, A. P. *et al.* Formulation and immunogenicity of a potential multivalent type III secretion system-based protein vaccine. *J. Pharm. Sci.* **99**, 4497–4509 (2010).
44. Riddle, M. S. *et al.* Safety and immunogenicity of an intranasal *Shigella flexneri* 2a Invaplex 50 vaccine. *Vaccine* **29**, 7009–7019 (2011).
45. Martinez-Becerra, F. J. *et al.* Characterization and Protective Efficacy of Type III Secretion Proteins as a Broadly Protective Subunit Vaccine against *Salmonella enterica* Serotypes. *Infect. Immun.* **86**, (2018).
46. Ndungo, E. *et al.* A Novel *Shigella* Proteome Microarray Discriminates Targets of Human Antibody Reactivity following Oral Vaccination and Experimental Challenge. *mSphere* **3**, (2018).
47. Heine, S. J. *et al.* Evaluation of immunogenicity and protective efficacy of orally delivered *Shigella* type III secretion system proteins IpaB and IpaD. *Vaccine* **31**, 2919–2929 (2013).
48. Sato, H. & Frank, D. W. Multi-Functional Characteristics of the *Pseudomonas aeruginosa* Type III Needle-Tip Protein, PcrV; Comparison to Orthologs in other Gram-negative Bacteria. *Front. Microbiol.* **2**, 142 (2011).
49. Reed, L. J. & Muench, H. A SIMPLE METHOD OF ESTIMATING FIFTY PER CENT ENDPOINTS. *Am. J. Epidemiol.* **27**, 493–497 (1938).
50. Köhler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495–497 (1975).
51. Barta, M. L. *et al.* Single-domain antibodies pinpoint potential targets within *Shigella* invasion plasmid antigen D of the needle tip complex for inhibition of type III secretion. *J. Biol. Chem.* **292**, 16677–16687 (2017).
52. Dickenson, N. E. *et al.* N-terminus of IpaB provides a potential anchor to the *shigella* type III secretion system tip complex protein IpaD. *Biochemistry (Mosc.)* **52**, 8790–8799 (2013).
53. Medina-Cucurella, A. V. & Whitehead, T. A. Characterizing Protein-Protein Interactions Using Deep Sequencing Coupled to Yeast Surface Display. **1764**, 101–121 (2018).
54. Medina-Cucurella, A. V., Zhu, Y., Bowen, S. J., Bergeron, L. M. & Whitehead, T. A. Pro region engineering of nerve growth factor by deep mutational scanning enables a yeast platform for conformational epitope mapping of anti-NGF monoclonal antibodies. *Biotechnol. Bioeng.* 1925–1937 (2018). doi:10.1002/bit.26706
55. Najar, T. A., Khare, S., Pandey, R., Gupta, S. K. & Varadarajan, R. Mapping Protein Binding Sites and Conformational Epitopes Using Cysteine Labeling and Yeast Surface Display. *Structure* **25**, 395–406 (2017).
56. Gaiotto, T. & Hufton, S. E. Cross-neutralising nanobodies bind to a conserved pocket in the hemagglutinin stem region identified using yeast display and deep mutational scanning. *PLoS ONE* **11**, 1–27 (2016).
57. Van Blarcom, T. *et al.* Precise and efficient antibody epitope determination through library design, yeast

- display and next-generation sequencing. *J. Mol. Biol.* **427**, 1513–1534 (2015).
58. Puri, V., Streaker, E., Prabakaran, P., Zhu, Z. & Dimitrov, D. S. Highly efficient selection of epitope specific antibody through competitive yeast display library sorting. *mAbs* **5**, 533–539 (2013).
59. Wrenbeck, E. E. *et al.* Plasmid-based one-pot saturation mutagenesis. *Nat. Methods* **13**, 928–930 (2016).
60. GBD Diarrhoeal Diseases Collaborators. Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Infect. Dis.* **17**, 909–948 (2017).
61. Kotloff, K. L., Riddle, M. S., Platts-Mills, J. A., Pavlinac, P. & Zaidi, A. K. M. Shigellosis. *Lancet Lond. Engl.* **391**, 801–812 (2018).
62. Livio, S. *et al.* Shigella isolates from the global enteric multicenter study inform vaccine development. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **59**, 933–941 (2014).
63. Michetti, P., Mahan, M. J., Schlauch, J. M., Mekalanos, J. J. & Neutra, M. R. Monoclonal secretory immunoglobulin A protects mice against oral challenge with the invasive pathogen *Salmonella typhimurium*. *Infect. Immun.* **60**, 1786–1792 (1992).
64. McSorley, S. J. & Jenkins, M. K. Antibody is required for protection against virulent but not attenuated *Salmonella enterica* serovar typhimurium. *Infect. Immun.* **68**, 3344–3348 (2000).
65. MacLennan, C. A. *et al.* The neglected role of antibody in protection against bacteremia caused by nontyphoidal strains of *Salmonella* in African children. *J. Clin. Invest.* **118**, 1553–1562 (2008).
66. Dougan, G., John, V., Palmer, S. & Mastroeni, P. Immunity to salmonellosis. *Immunol. Rev.* **240**, 196–210 (2011).
67. Oberhelman, R. A. *et al.* Prospective study of systemic and mucosal immune responses in dysenteric patients to specific *Shigella* invasion plasmid antigens and lipopolysaccharides. *Infect. Immun.* **59**, 2341–2350 (1991).
68. Cam, P. D., Pál, T. & Lindberg, A. A. Immune response against lipopolysaccharide and invasion plasmid-coded antigens of shigellae in Vietnamese and Swedish dysenteric patients. *J. Clin. Microbiol.* **31**, 454–457 (1993).
69. Van de Verg, L. L., Herrington, D. A., Boslego, J., Lindberg, A. A. & Levine, M. M. Age-specific prevalence of serum antibodies to the invasion plasmid and lipopolysaccharide antigens of *Shigella* species in Chilean and North American populations. *J. Infect. Dis.* **166**, 158–161 (1992).
70. Kotloff, K. L. *et al.* Safety and immunogenicity of CVD 1208S, a live, oral DeltaguaBA Deltasen Deltaset *Shigella flexneri* 2a vaccine grown on animal-free media. *Hum. Vaccin.* **3**, 268–275 (2007).
71. Simon, J. K. *et al.* Antigen-specific B memory cell responses to lipopolysaccharide (LPS) and invasion plasmid antigen (Ipa) B elicited in volunteers vaccinated with live-attenuated *Shigella flexneri* 2a vaccine candidates. *Vaccine* **27**, 565–572 (2009).
72. Kurtz, J. R., Goggins, J. A. & McLachlan, J. B. *Salmonella* infection: Interplay between the bacteria and host immune system. *Immunol. Lett.* **190**, 42–50 (2017).
73. Jneid, B. *et al.* Role of T3SS-1 SipD Protein in Protecting Mice against Non-typhoidal *Salmonella* Typhimurium. *PLoS Negl. Trop. Dis.* **10**, e0005207 (2016).
74. Martinez-Becerra, F. J. *et al.* Broadly protective *Shigella* vaccine based on type III secretion apparatus proteins. *Infect. Immun.* **80**, 1222–1231 (2012).
75. Johnson, S. *et al.* Self-chaperoning of the type III secretion system needle tip proteins IpaD and BipD. *J. Biol. Chem.* **282**, 4035–4044 (2007).
76. Veenendaal, A. K. J. *et al.* The type III secretion system needle tip complex mediates host cell sensing and translocon insertion. *Mol. Microbiol.* **63**, 1719–1730 (2007).
77. Blocker, A. J. *et al.* What’s the point of the type III secretion system needle? *Proc. Natl. Acad. Sci. U. S. A.* **105**, 6507–6513 (2008).
78. Lunelli, M., Hurwitz, R., Lambers, J. & Kolbe, M. Crystal structure of prgi-SipD: Insight into a secretion competent state of the type three secretion system needle tip and its interaction with host ligands. *PLoS Pathog.* **7**, (2011).
79. Epler, C. R., Dickenson, N. E., Bullitt, E. & Picking, W. L. Ultrastructural Analysis of IpaD at the Tip of the Nascent MxiH Type III Secretion Apparatus of *Shigella flexneri*. *J. Mol. Biol.* **420**, 29–39 (2012).
80. Bernard, A. R., Jessop, T. C., Kumar, P. & Dickenson, N. E. Deoxycholate-Enhanced *Shigella* Virulence Is Regulated by a Rare  $\pi$ -Helix in the Type Three Secretion System Tip Protein IpaD. *Biochemistry (Mosc.)* (2017). doi:10.1021/acs.biochem.7b00836
81. Rathinavelan, T. *et al.* NMR model of PrgI-SipD interaction and its implications in the needle-tip assembly of the salmonella type III secretion system. *J. Mol. Biol.* **426**, 2958–2969 (2014).
82. Rathinavelan, T., Tang, C. & De Guzman, R. N. Characterization of the interaction between the *Salmonella* type III secretion system tip protein SipD and the needle protein PrgI by paramagnetic relaxation enhancement. *J. Biol. Chem.* **286**, 4922–4930 (2011).
83. Barta, M. L. *et al.* Identification of the bile salt binding site on IpaD from *Shigella flexneri* and the influence of ligand binding on IpaD structure. *Proteins* **80**, 935–45 (2012).
84. Chatterjee, S. *et al.* The crystal structures of the *Salmonella* type III secretion system tip protein SipD in complex with deoxycholate and chenodeoxycholate. *Protein Sci.* **20**, 75–86 (2011).
85. McShan, A. C., Anbanandam, A., Patnaik, S. & De Guzman, R. N. Characterization of the Binding of

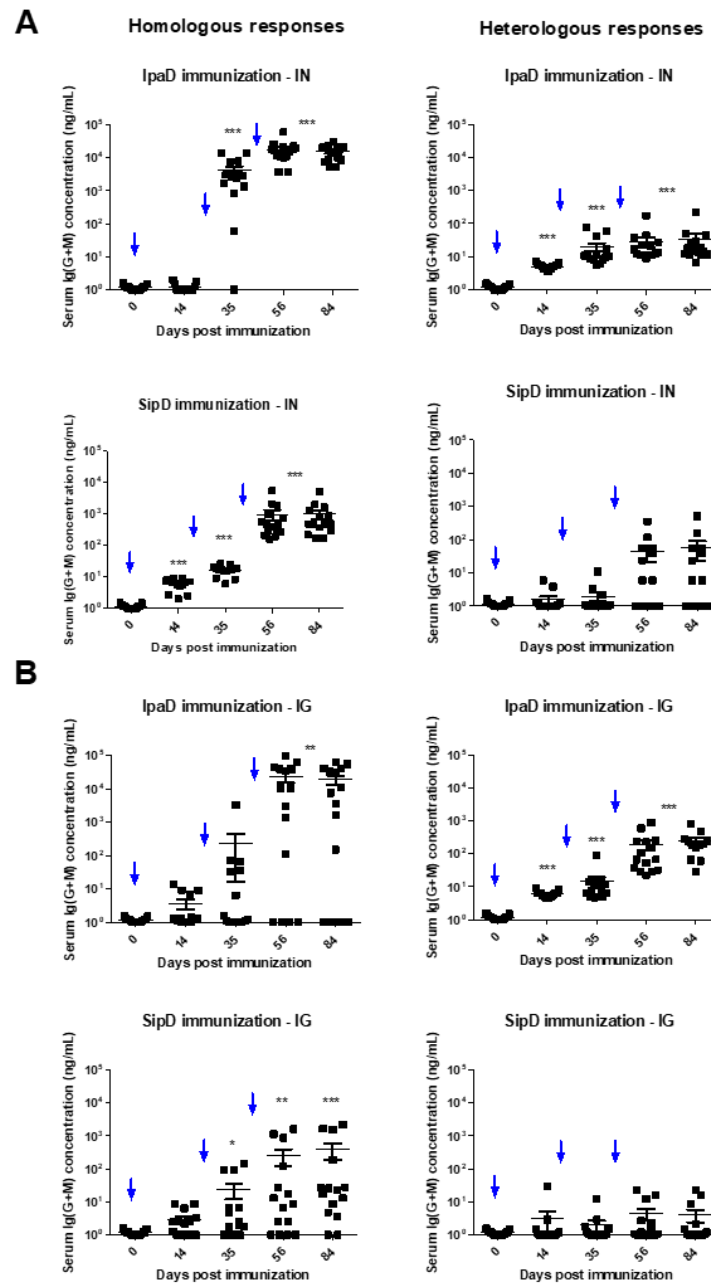
- Hydroxyindole, Indoleacetic acid, and Morpholinoaniline to the Salmonella Type III Secretion System Proteins SipD and SipB. *ChemMedChem* **11**, 963–971 (2016).
86. Epler, C. R., Dickenson, N. E., Olive, A. J., Picking, W. L. & Picking, W. D. Liposomes recruit IpaC to the Shigella flexneri type III secretion apparatus needle as a final step in secretion induction. *Infect. Immun.* **77**, 2754–2761 (2009).
87. Cheung, M. *et al.* Three-dimensional electron microscopy reconstruction and cysteine-mediated crosslinking provide a model of the type III secretion system needle tip complex. *Mol. Microbiol.* **95**, 31–50 (2015).
88. Barta, M. L. *et al.* Using disruptive insertional mutagenesis to identify the in situ structure-function landscape of the Shigella translocator protein IpaB. *Protein Sci.* (2018). doi:10.1002/pro.3428
89. Roehrich, A. D., Guillosoy, E., Blocker, A. J. & Martinez-Argudo, I. Shigella IpaD has a dual role: signal transduction from the type III secretion system needle tip and intracellular secretion regulation. *Mol. Microbiol.* **87**, 690–706 (2013).
90. Dickenson, N. E. *et al.* The N-terminus of IpaB provides a potential anchor to the Shigella type III secretion system tip complex protein IpaD. *Biochemistry (Mosc.)* **52**, 8790–8799 (2013).
91. Laporte, J. *et al.* Fast and sensitive detection of enteropathogenic Yersinia by immunoassays. *J. Clin. Microbiol.* **53**, 146–159 (2015).
92. Ellman, G. L., Courtney, K. D., Andres, V. & Feather-Stone, R. M. A new and rapid colorimetric determination of acetylcholinesterase activity. *Biochem. Pharmacol.* **7**, 88–95 (1961).
93. Benatuil, L., Perez, J. M., Belk, J. & Hsieh, C. M. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng. Des. Sel.* **23**, 155–159 (2010).
94. Sivel, C. *et al.* Fab is the most efficient format to express functional antibodies by yeast surface display. *mAbs* **10**, 1–10 (2018).
95. Sivel, C. *et al.* Fab is the most efficient format to express functional antibodies by yeast surface display. *mAbs* **10**, 720–729 (2018).
96. Blankenberg, D. *et al.* Manipulation of FASTQ data with galaxy. *Bioinformatics* **26**, 1783–1785 (2010).



**Supplementary information  
&  
Expanded views**

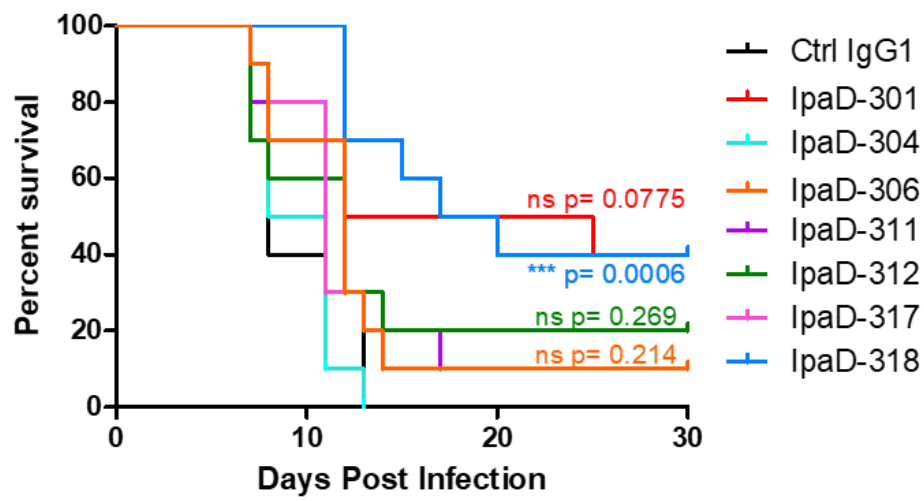
## Expanded Views

Figure EV1



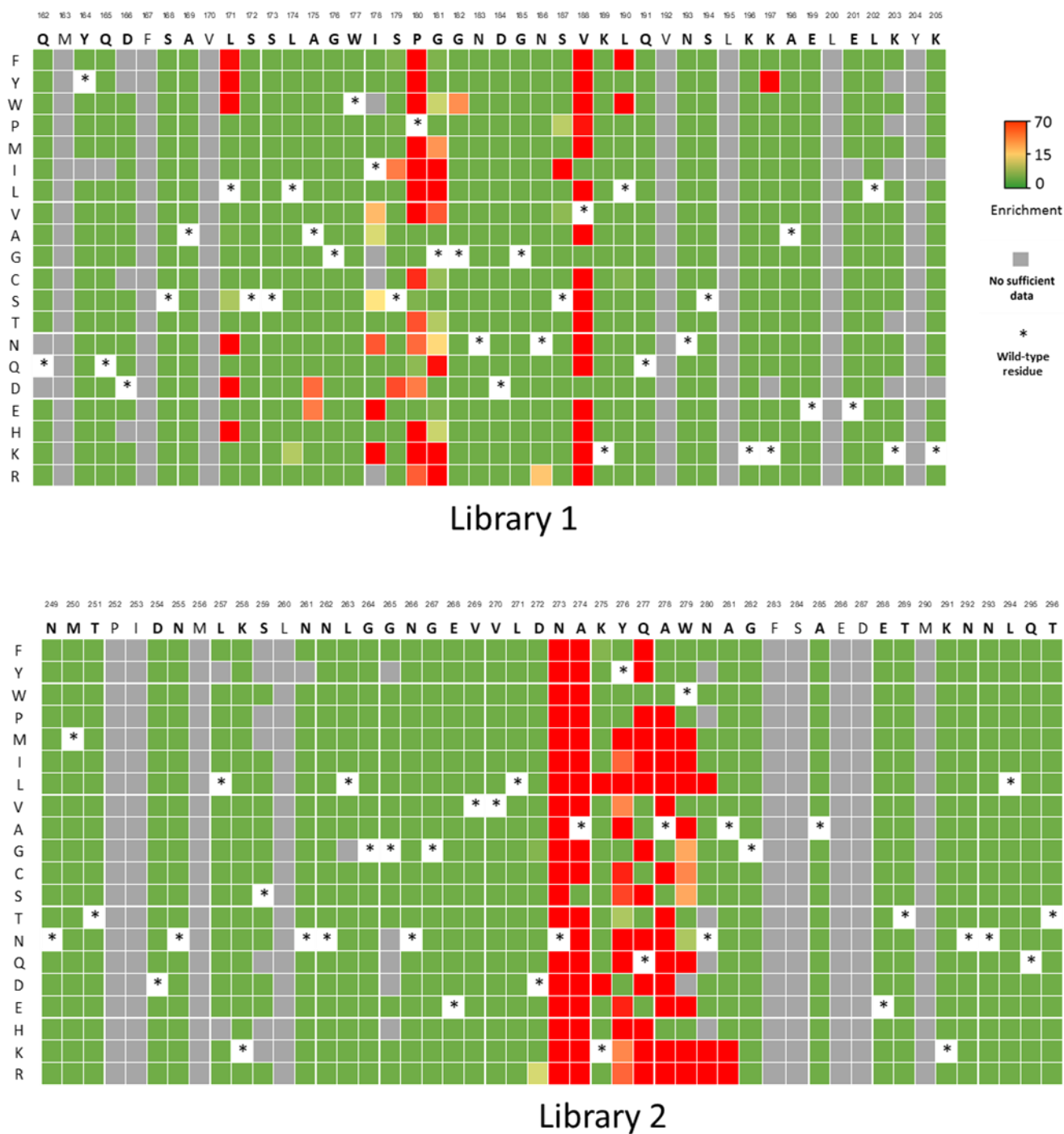
**Figure EV1. Kinetics of polyclonal Ig(G+M) antibody responses to IpaD and SipD antigens.** Mice were immunized three times (time indicated with arrows) with IpaD or SipD by the (A) IN route or (B) IG route as described in Materials and Methods. Homologous responses of Ig(G+M) antibodies specific for IpaD or SipD (A, B left) and heterologous responses of Ig(G+M) antibodies (A, B right) were quantified by ELISA (see figure S2A). Data represent mean concentrations (ng/mL) and the standard errors (SEM) from 14–16 individual mice per group. \*\*\*: P value < 0.001 (except for SipD immunization IG, panel B: P value = 0.002), comparing the antibody responses on days post-immunization versus those on day 0 (nonparametric Mann-Whitney test). \*\*: P value = 0.007 \*: P value = 0.027.

Figure EV2



**Figure EV2. *In vivo* protective mAb screening.** Survival curves of mice treated with monoclonal antibodies (10 mice per antibody) 24 hours before infection with 100 LD<sub>50</sub> of *S. Typhimurium*. P-value was calculated in comparison with the control with the log-rank (Mantel-Cox) test.

Figure EV3



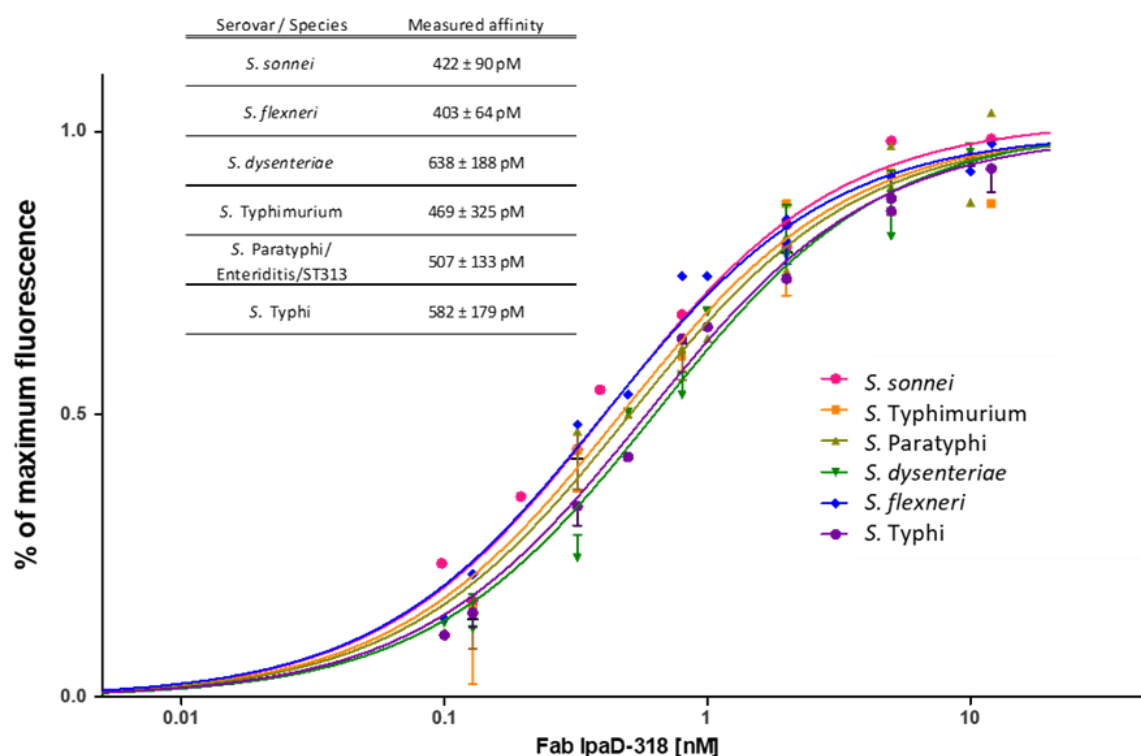
**Figure EV3. Complete heatmap representing enrichment values of each IpaD mutant after sorting of variants with altered Fab IpaD-318 binding.** For a mutation, a strong enrichment value represents a loss of Fab IpaD-318 binding. Only exposed residues (bold letters) were mutated.

Figure EV4



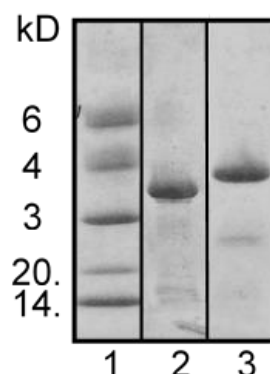
**Figure EV4. Multiple alignment of IpaD and SipD sequences from different serovars / species.** Multiple alignment of IpaD from three *Shigella* species (*S. flexneri*, *S. sonnei* and *S. dysenteriae*) and SipD from five *Salmonella enterica* serovars (Enteritidis, Paratyphi, Typhimurium, Typhi and Typhimurium ST313). *S. Enteritidis*, *S. Paratyphi* and *S. Typhimurium* ST313 all share the same SipD sequence epitope. For vaccination experiments, mice were immunized with IpaD from *Shigella sonnei* and SipD from *Salmonella* Typhimurium. *In vivo* protection experiments were performed using *Shigella flexneri* 2a and *Salmonella* Typhimurium. Clustal omega was used to perform the alignments ([www.ebi.ac.uk/Tools/msa/clustalo/](http://www.ebi.ac.uk/Tools/msa/clustalo/)). Red framing refers to key positions for Fab IpaD-318 binding as determined by deep mutational scanning experiments.

Figure EV5

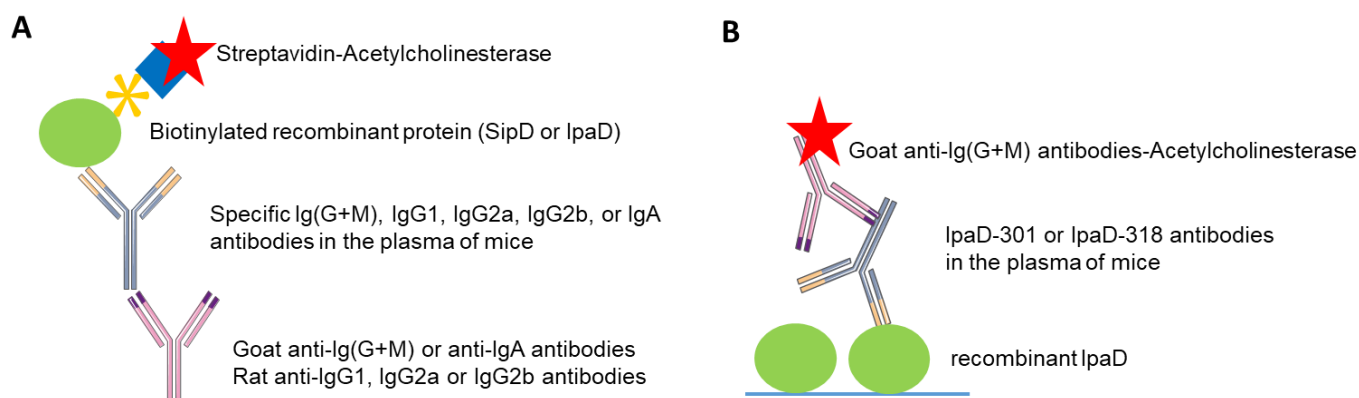


**Figure EV5. Binding affinity determination of Fab IpaD-318 on yeast cells displaying wild-type IpaD from several *Shigella*/*Salmonella enterica* species/serovars.** IpaD or SipD from different species/serovars of *Salmonella* and *Shigella* was expressed at the surface of yeast cells. The fluorescence corresponding to Fab IpaD-318 binding was determined by FACS analysis in the presence of increasing concentrations of Fab. Measurements were done in duplicate with independent cultures and inductions. SipD from *S. Paratyphi*, *S. Enteritidis* and serovar ST313 all share the same sequence epitope. Apparent  $K_D$  values and 95% confidence intervals were determined using PRISM software with a one site-specific binding model.

## Supplementary figures



**Figure S1. Analysis of recombinant SipD and IpaD proteins.** SDS-PAGE / Coomassie blue staining (reducing conditions) of purified recombinant proteins. PolyHis-IpaD (37.1 kDa, lane 2) and polyHis-SipD (38.2 kDa, lane 3) are shown with molecular mass markers in kilodaltons (kDa) (lane 1).

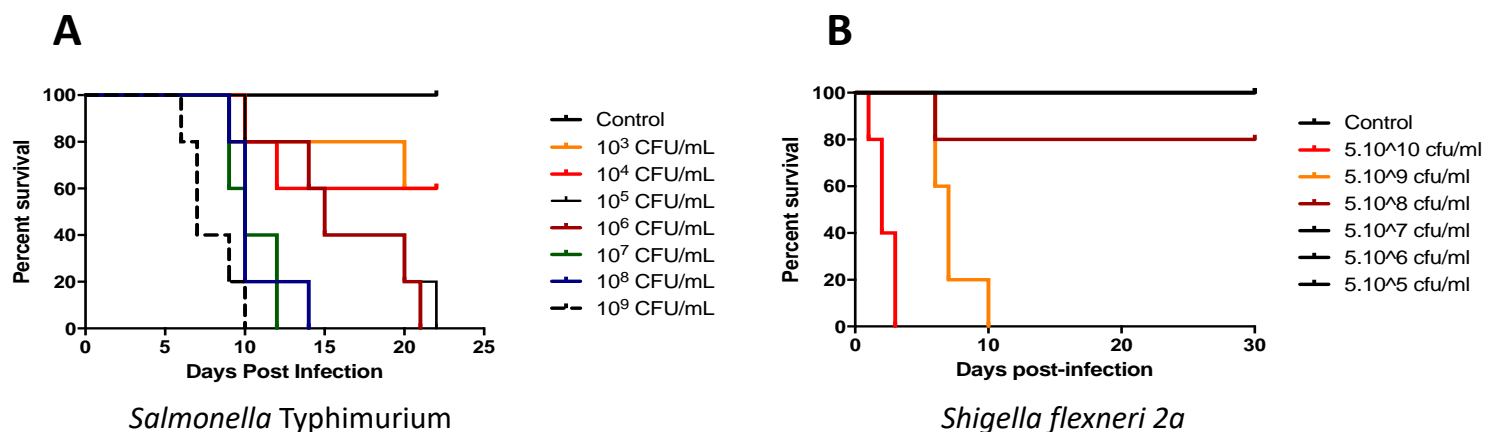


**Figure S2. Principle of the ELISA used for the selection of anti-IpaD/SipD monoclonal antibodies and for measurement of circulating antibodies.**

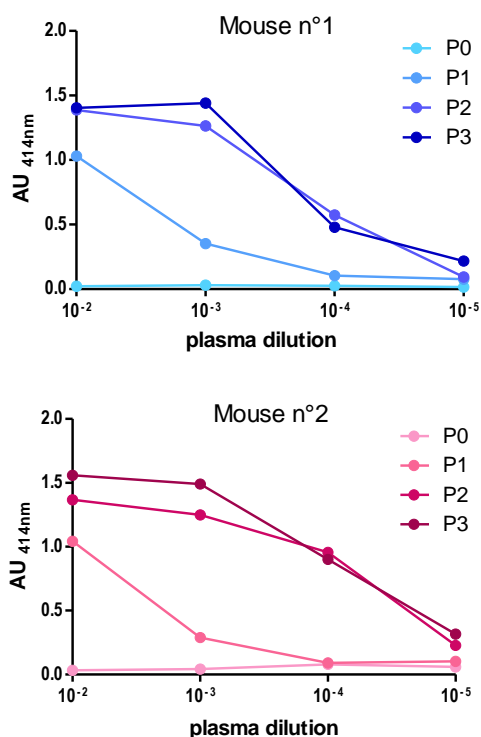
A- Culture supernatants for the final selection of mAbs able to recognize both SipD and IpaD were screened using differential sandwich ELISA. Each culture supernatant containing mAbs of interest was tested using either biotinylated IpaD or SipD. The same ELISA test was performed to measure the concentrations of circulating antibodies (immune response after immunizations (Ig(G+M), IgG1, IgG2a, IgG2b and IgA).

B- Half-lives of IpaD-301 and IpaD-318 were measured using ELISA. Recombinant IpaD was coated on solid phase and dilutions of plasma of mice harvested at different times after injection were tested and revealed using a tracer antibody (polyclonal anti Ig(G+M) antibody coupled to acetylcholinesterase).

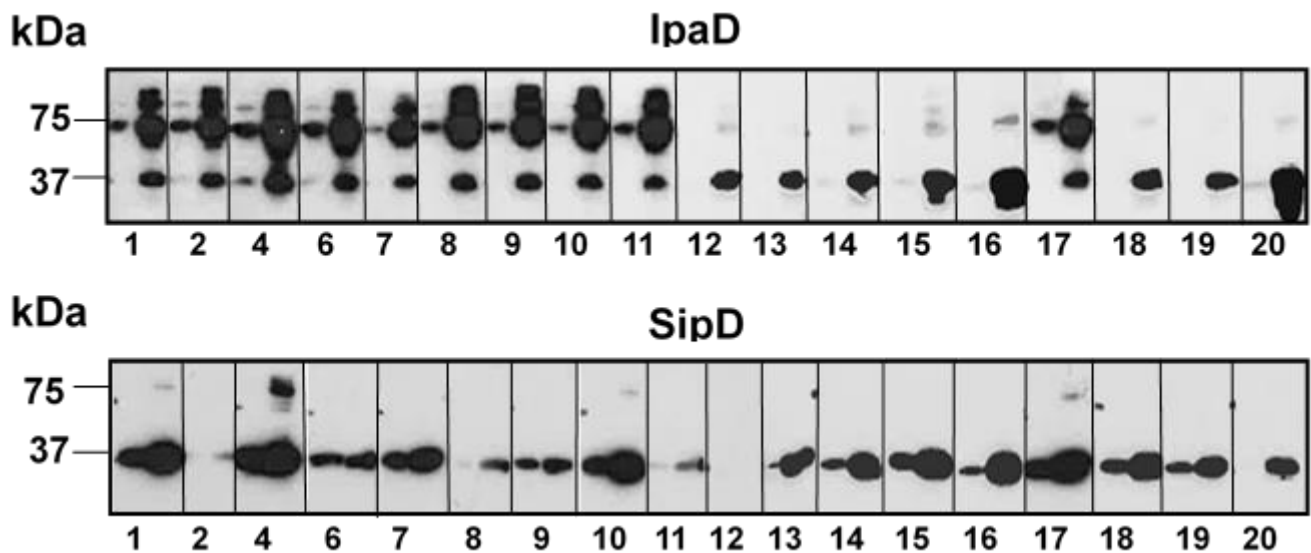




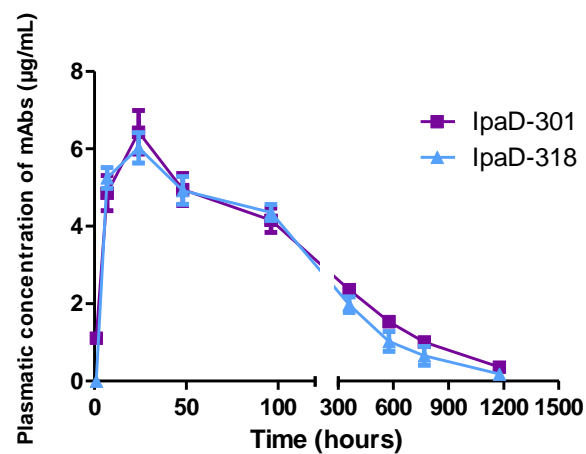
**Figure S3. Determination of LD50 for *S. Typhimurium* and *S. flexneri 2a*.** Serial dilutions of *S. Typhimurium* (from  $2.10^2$  to  $2.10^8$  CFU) and *S. flexneri 2a* ( $5.10^5$  to  $5.10^{10}$  CFU) were administered intragastrically (*S. Typhimurium*) or intranasally (*S. flexneri 2a*) to 20- to 22-week-old female BALB/c mice (5 mice per group). The 50% mouse lethal dose (LD 50) was calculated by the method of Reed and Muench.



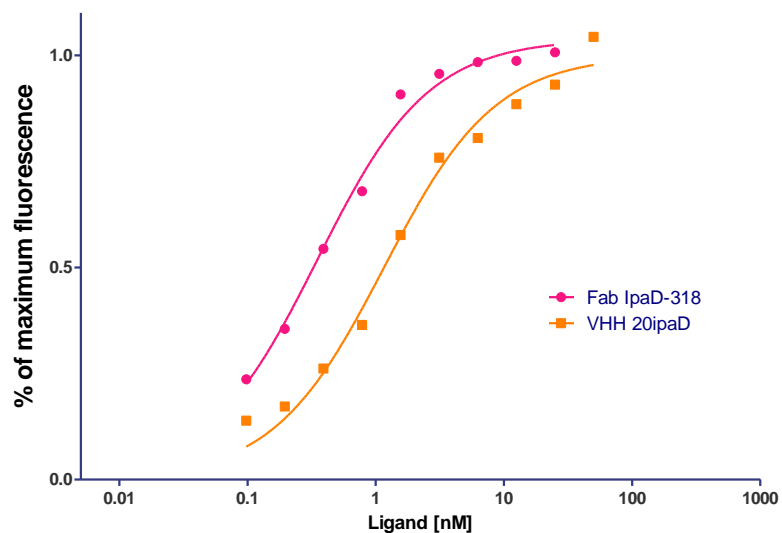
**Figure S4. Polyclonal antibody titers of IpaD-immunized BALB/c mice.** The immune polyclonal response was evaluated by enzyme immunoassay (EIA) with serial dilutions of plasma harvested at different times during intranasal immunization by IpaD recombinant protein. These two mice presenting the highest immune response were selected for selection of mAbs.



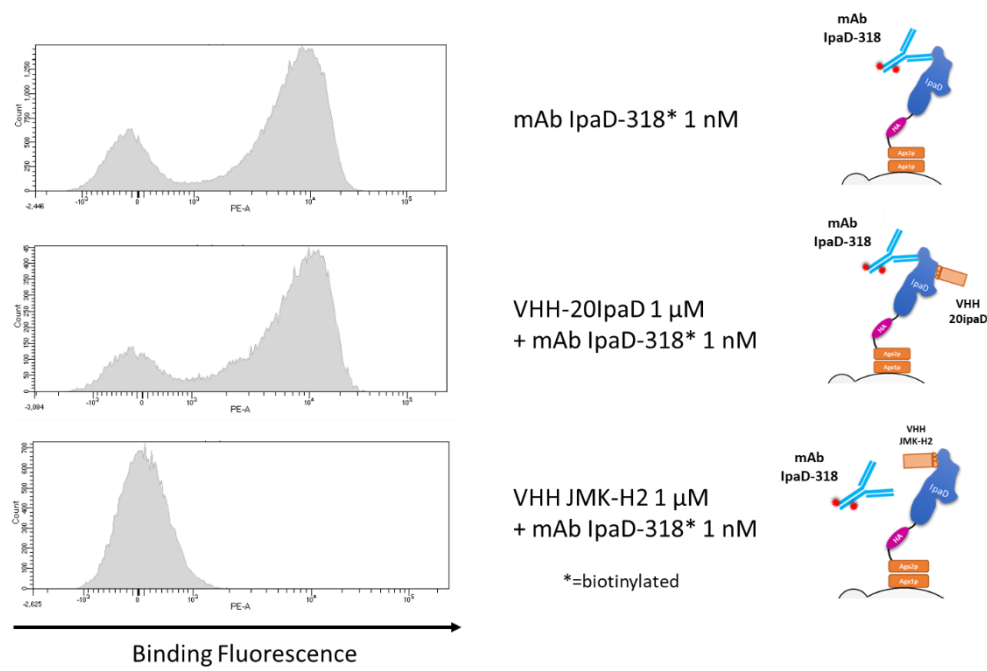
**Figure S5. Recognition of IpaD (top) and SipD (bottom) recombinant T3SS proteins with anti-IpaD/SipD monoclonal antibodies.** Western blotting was performed with 10 ng/well (first well) and 100 ng/well (second well) of IpaD or SipD against 18 purified anti-IpaD/SipD mAbs (4 µg/mL). The numbers of mAbs are indicated below each immunoblot. Numbers on the left indicate the molecular weight markers in kDa.



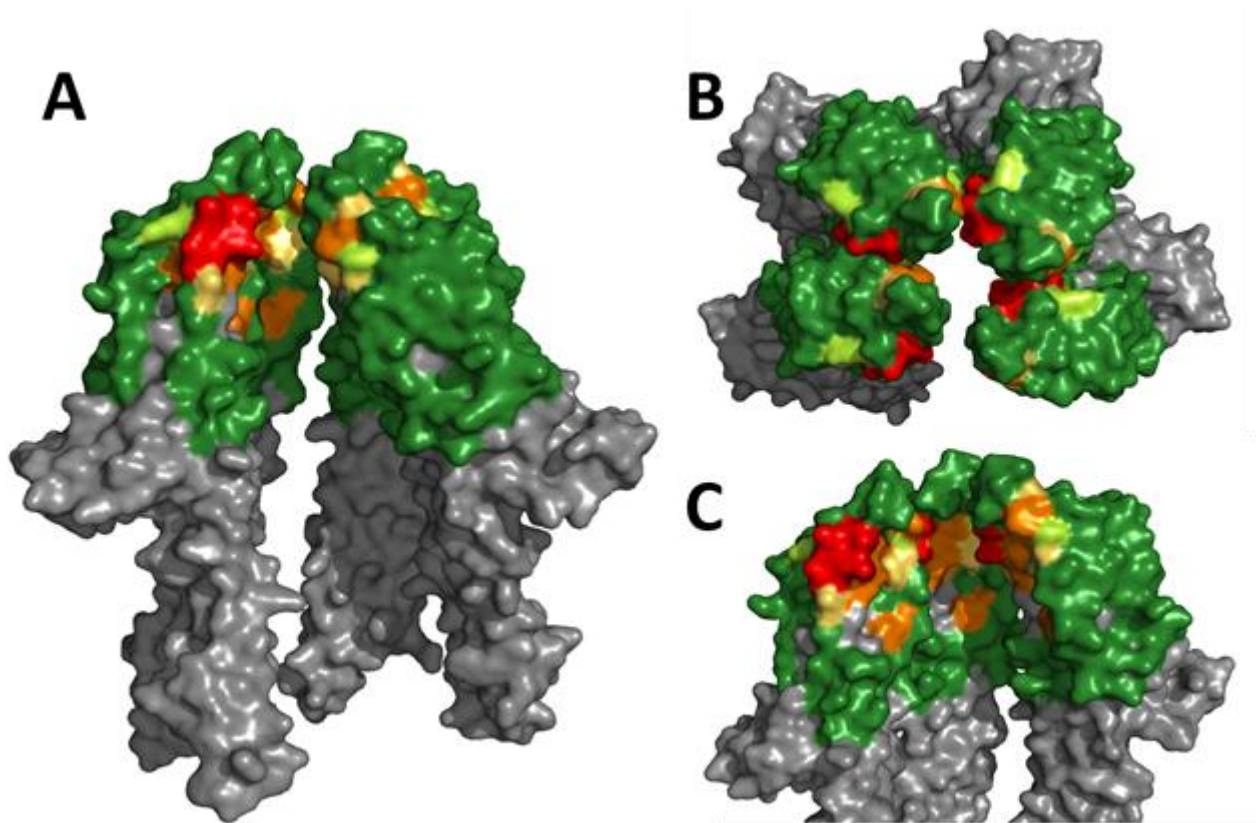
**Figure S6. Pharmacokinetic analysis of anti-IpaD/SipD antibodies in mice.** 50 µg of each purified antibody was injected intraperitoneally into BALB/c mice (N = 4). Plasma was harvested at different times to calculate the circulating concentrations of mAbs (IpaD-301 and IpaD-318) using an ELISA test (see fig S4B).



**Figure S7. Apparent  $K_D$  determination of Fab IpaD-318 and VHH-20ipad on yeast cells expressing wild-type IpaD.** Wild-type IpaD was expressed on the surface of yeast cells. The fluorescence corresponding to Fab or VHH binding was determined by FACS analysis in the presence of increasing concentrations of the corresponding antibody fragment. Apparent  $K_D$  values and 95% confidence intervals were determined using PRISM software with a one site-specific binding model. Apparent  $K_D$  values were measured at  $350 \pm 70$  pM for Fab IpaD-318 and  $1.15 \pm 0.30$  nM for VHH-20Ipad.



**Figure S8. Binding competition and compatibility assay of VHH-JMK-H2 and VHH-20ipad vs mAb IpaD-318.** Wild-type IpaD was expressed on the surface of *S. cerevisiae* cells EBY100 to determine VHH-JMK-H2 and 20ipad competed with mAb IpaD-318 for IpaD binding. The binding fluorescence of biotinylated mAb IpaD-318 on the surface of yeasts is shown as histograms. Prior to mAb IpaD-318 binding, PBS (control), VHH-20ipad (1  $\mu$ M) or VHH-JMK-H2 (1  $\mu$ M) was added to the yeast cells expressing IpaD. Cells were washed and incubated with 1 nM biotinylated mAb IpaD-318. Binding fluorescence (SA-PE) was monitored by flow cytometry.



**Figure S9. Representation of the epitope of mAb IpaD-318 on the proposed model for the tip complex.** Representation of the epitope of mAb IpaD-318 on the proposed model (PDB #4d3e) for the tip complex of Blocker and co-workers. In this model, the fifth member of the pentamer is supposed to be IpaB (not represented). (A) Side view. (B) Upper View. (C) Side view focused on the epitope lacking the 4<sup>th</sup> subunit.

## Supplementary tables

**Table S1. Summary of the homologous (Ig (G+M), IgG1, IgG (2a+2b), IgA) and heterologous (Ig (G+M)) antibody responses after the last immunization with SipD or IpaD by the IN and IG routes.**

Immunization route	Homologous antibody response					Heterologous response
	Immunogen	Ig(G+M)	IgG1	IgG(2a+2b)	IgA titer	Ig (G+M)
IN	SipD	1.7 10 <sup>3</sup>	2.6 10 <sup>3</sup>	4.6 10 <sup>2</sup>	2.4 10 <sup>2</sup>	4.4 10 <sup>1</sup>
	IpaD	2.9 10 <sup>4</sup>	2.0 10 <sup>4</sup>	6.8 10 <sup>3</sup>	5.2 10 <sup>2</sup>	2.9 10 <sup>1</sup>
IG	SipD	1.2 10 <sup>3</sup>	6.8 10 <sup>3</sup>	9.9 10 <sup>2</sup>	1.7 10 <sup>2</sup>	4.4 10 <sup>0</sup>
	IpaD	2.6 10 <sup>4</sup>	4.9 10 <sup>2</sup>	8.3 10 <sup>3</sup>	3.8 10 <sup>2</sup>	1.9 10 <sup>2</sup>

Data represent mean concentrations (ng/mL) for Ig(G+M), IgG1, IgG(2a+2b) responses and IgA titer from each group of mice.

**Table S2. Homologous and cross-protection efficacy induced by SipD and IpaD T3SS protein immunizations by the IN and IG routes from lethal challenge with *S. flexneri 2a* (intranasal) or *S. Typhimurium* (intra-gastric) in mice.**

Immunization route	Immunogen	Challenge	Homologous protection efficacy (%)	Heterologous protection efficacy (%)	<i>P</i> value <sup>a</sup>
IN	IpaD	<i>S. flexneri 2a</i>	27	27	0.644
	IpaD	<i>S. Typhimurium</i>			0.125
	SipD	<i>S. Typhimurium</i>	50	47	0.001
	SipD	<i>S. flexneri 2a</i>			0.002
IG	IpaD	<i>S. flexneri 2a</i>	62	30	0.0005
	IpaD	<i>S. Typhimurium</i>			0.099
	SipD	<i>S. Typhimurium</i>	43	67	0.090
	SipD	<i>S. flexneri 2a</i>			<0.0001

The mice immunized by the intranasal (IN) or intra-gastric (IG) route with SipD and IpaD were challenged with 5.10<sup>10</sup> CFU/mL of *S. flexneri 2a* by the IN route (LD 50 = 5.10<sup>8</sup> CFU/mL) or with 10<sup>6</sup> CFU/mL of *S. Typhimurium* by the IG route (LD 50 = 10<sup>4</sup> CFU/mL). The mortality rate of the immunized group was compared with that of the PBS-immunized control animals using the log-rank (Mantel-Cox) test.

**Table S3. Numbers and isotypes of IpaD/SipD monoclonal antibodies produced against IpaD protein.**

<b>mAbs</b>	<b>Isotype</b>
<b>IpaD-301</b>	IgG1
<b>IpaD-302</b>	IgG1
<b>IpaD-303</b>	IgG1
<b>IpaD-304</b>	IgG1
<b>IpaD-305</b>	IgG1
<b>IpaD-306</b>	IgG1
<b>IpaD-307</b>	IgG1
<b>IpaD-308</b>	IgG2b
<b>IpaD-309</b>	IgG1
<b>IpaD-310</b>	IgG1
<b>IpaD-311</b>	IgG1
<b>IpaD-312</b>	IgG1
<b>IpaD-313</b>	IgG1
<b>IpaD-314</b>	IgG2a
<b>IpaD-315</b>	IgG2a
<b>IpaD-316</b>	IgG2a
<b>IpaD-317</b>	IgG1
<b>IpaD-318</b>	IgG1
<b>IpaD-319</b>	IgG1
<b>IpaD-320</b>	IgG1

The isotypes were identified by a sandwich ELISA test using antibodies specific to each isotype. Although recognizing both proteins and because they were obtained from mice immunized with IpaD, they were named “IpaD”.



Table S4. *In vivo* neutralizing activity of anti-IpaD/SipD monoclonal antibodies against intranasal *S. flexneri 2a* or intragastric *S. Typhimurium* challenge (100 LD50).

mAbs	Survival rate (%) infection <i>S. Typhimurium</i>				Survival rate (%) infection <i>S. flexneri 2a</i>			
	Day-1	p value	Day+5	p value	Day-1	p value	Day+1	p value
<b>Ctrl IgG1</b>	0	na	0	na	0	na	0	na
<b>IpaD-301</b>	40	0.0775	0	na	40	0.0027	0	na
<b>IpaD-318</b>	0	0.0006	10	0.209	50	0.0049	20	0.385
<b>IpaD-(301+318)</b>	0	0.360	10	0.534	30	0.0471	10	0.397

Table S5. Selection of anti-IpaD/SipD monoclonal antibodies with *in vivo* neutralizing activity against intragastric *S. Typhimurium* challenge.

mAb	Survival	mAb	Survival	mAb	Survival	mAb	Survival
<b>IpaD-301</b>	1/3	<b>IpaD-306</b>	2/3	<b>IpaD-311</b>	1/3	<b>IpaD-316</b>	1/3
<b>IpaD-302</b>	1/3	<b>IpaD-307</b>	1/3	<b>IpaD-312</b>	0/3	<b>IpaD-317</b>	3/3
<b>IpaD-303</b>	1/3	<b>IpaD-308</b>	1/3	<b>IpaD-313</b>	0/3	<b>IpaD-318</b>	2/3
<b>IpaD-304</b>	0/3	<b>IpaD-309</b>	0/3	<b>IpaD-314</b>	0/3	<b>IpaD-319</b>	0/3
<b>IpaD-305</b>	0/3	<b>IpaD-310</b>	0/3	<b>IpaD-315</b>	1/3	<b>IpaD-320</b>	0/3

Mice (N = 3) were injected twice with 500 µg of each mAb (24 hours before the challenge and 4 days later) by the intraperitoneal route and were challenged with 100 LD50 of *S. Typhimurium* by the intragastric route. Survival was monitored for 30 days.

## II. Résultats additionnels et conclusion

En complément de ce travail, les séquences des anticorps IpaD-301 et IpaD-318 ont été déterminées par RT-PCR à partir des ARNm extraits des hybridomes. Le séquençage a permis de mettre en évidence que la chaîne légère variable de l'anticorps IpaD-301 partage 93% d'identité avec l'anticorps IpaD-318. D'autre part, les chaînes lourdes variables des deux anticorps partagent 94% d'identité et présentent des CDRH3 identiques. Ces résultats suggèrent que les séquences des deux anticorps dérivent d'une même séquence parentale qui a subi des mutations somatiques différentes. Des tests de compatibilité d'épitopes ont été réalisés en YSD et montrent effectivement que les anticorps IpaD-301 et IpaD-318 ne peuvent pas se lier en même temps à l'antigène IpaD ou SipD (résultats non présentés). Les deux anticorps partagent donc très probablement le même épitope de liaison sur ces antigènes. Ces résultats peuvent expliquer l'absence de synergie entre les deux anticorps dans les tests de protection *in vivo*.



## Partie 2 : Etude du format d'expression des anticorps à la surface de la levure

### I. Introduction de l'article

L'objectif principal de cette thèse a été de mettre en place une méthodologie permettant de modifier des anticorps pour les rendre non immunogènes. Pour cela, il est essentiel de pouvoir cribler à large échelle des banques de variants afin de sélectionner des mutants présentant les propriétés recherchées. La technique du *Yeast Surface Display* disponible au laboratoire a pu être mise en œuvre pour répondre à ce besoin. Dans un premier temps, nous avons cherché à déterminer quel était le format de fragment d'anticorps le plus adapté pour parvenir à l'expression fonctionnelle des anticorps anti-Salmonelles/Shigelles à la surface des levures.

L'article présenté ci-après publié dans le journal *Mabs* en 2018 compare la fonctionnalité de 6 anticorps exprimés en YSD (dont IpaD-301 et IpaD-318) sous différents formats : scFv, scFab et Fab. Le travail que j'ai réalisé dans cet article concerne tous les résultats obtenus pour les anticorps IpaD-301 et IpaD-318.

## REPORT



# Fab is the most efficient format to express functional antibodies by yeast surface display

Coline Sivel  <sup>a,†</sup>, Rapha  l Sierocki<sup>a,†</sup>, Kelly Ferreira-Pinto<sup>a</sup>, St  phanie Simon<sup>b</sup>, Bernard Maill  re<sup>a</sup>, and Herv   Nozach<sup>a</sup>

<sup>a</sup>Service d'Ing  nierie Mol  culaire des Prot  ines (SIMOPRO), CEA, Universit   Paris-Saclay, Gif/Yvette, France; <sup>b</sup>Service de Pharmacologie et Immunoanalyse (SPI), CEA, INRA, Laboratoire d'Etudes et de Recherches en Immunoanalyse, Universit   Paris-Saclay, Gif/Yvette, France

## ARTICLE HISTORY

Received 20 February 2018

Revised 11 April 2018

Accepted 20 April 2018

## KEYWORDS

antibody engineering;  
affinity maturation; fab  
fragment; monoclonal  
antibodies; scFv fragment;  
scFab fragment; yeast surface  
display

## ABSTRACT

Multiple formats are available for engineering of monoclonal antibodies (mAbs) by yeast surface display, but they do not all lead to efficient expression of functional molecules. We therefore expressed four anti-tumor necrosis factor and two anti-IpaD mAbs as single-chain variable fragment (scFv), antigen-binding fragment (Fab) or single-chain Fabs and compared their expression levels and antigen-binding efficiency. Although the scFv and scFab formats are widely used in the literature, 2 of 6 antibodies were either not or weakly expressed. In contrast, all 6 antibodies expressed as Fab revealed strong binding and high affinity, comparable to that of the soluble form. We also demonstrated that the variations in expression did not affect Fab functionality and were due to variations in light chain display and not to misfolded dimers. Our results suggest that Fab is the most versatile format for the engineering of mAbs.

## Introduction

Over the past 20 years, yeast surface display (YSD) has emerged as a very potent tool for protein engineering. It is a very proficient platform for *de novo* identification of new binders in na  ve<sup>1</sup> or immune libraries.<sup>2,3</sup> YSD is also used for the engineering of many antibodies, principally to identify mutants with improved affinity<sup>4,5</sup> selectivity, stability and expression.<sup>4–7</sup> Affinity-matured antibodies have been developed against multiple antigens such as tumor necrosis factor (TNF),<sup>8</sup> fluorescein,<sup>9</sup> botulinum neurotoxin type A,<sup>10</sup> carcinoembryonic antigen,<sup>11</sup> epidermal growth factor receptor<sup>12</sup> and streptavidin.<sup>13</sup> Engineering of other properties of antibodies such as pH sensitivity or antigen specificity has also been described.<sup>14,15</sup>

Engineering of pre-existing antibodies using YSD relies on the anchoring of antibody fragments on the surface of yeast cells via fusion with a cell wall protein. Various formats are described in the literature for the functional display of antibody fragments on yeast cells, such as single-chain variable fragment (scFv), antigen-binding fragment (Fab) or single-chain Fab (scFab), although they are known to lead to variable levels of antibody expression.<sup>2,4,10,16</sup> Yeast display of IgG has also been reported with secretion and capture strategies,<sup>17,18</sup> but, to our knowledge, not as a conventional fusion with a yeast surface anchor protein. In most cases, antibody fragments are presented using the Aga2p-based display system, which permits either N-terminal or C-terminal fusion,<sup>4</sup> but the lack of comparative studies does not facilitate the choice of the appropriate formats to be used.

ScFv molecules are the smallest functional entity (30 kDa) that includes the complete antigen-binding site of an antibody. This entity comprises the two variable domains VH and VL connected by a flexible linker that prevents their dissociation.

Fab are composed of the light chain (VL + CL) and a heavy chain (VH + CH1) connected by an interchain disulfide bond. Several studies have reported the expression of Fab molecules on the surface of yeast cells using either a bigenic plasmid<sup>13</sup> or two distinct plasmids<sup>10,14,19</sup> to allow separate expression of the two chains. Expression levels of Fab molecules are generally considered as lower than those of scFv molecules in *E. coli*,<sup>20,21</sup> at the surface of phages<sup>22,23</sup> and also on the surface of yeast cells.<sup>10</sup>

The scFab derived from Fab is obtained through introduction of a polypeptide linker connecting the light chain and the heavy chain. Thus, it combines some advantages of the scFv (single chain and one expression plasmid, level of expression) with the stability and lower tendency to aggregate of the Fab.<sup>2,23</sup> The linker prevents the display of unpaired heavy chain antibodies (devoid of light chain), which can sometimes interfere with the selection steps.<sup>2</sup> ScFab molecules carry a flexible linker of variable length (32 to 80 amino acids) and an optional inter-chain disulfide bond.<sup>2,24</sup> Deletion of carboxy terminal cysteine residues ( $\Delta$ C) of the CL and CH1 domains results in a scFab $\Delta$ C format, which has increased expression levels in *E. coli*.<sup>23</sup>

The scFv is the most popular format for antibody engineering,<sup>4</sup> probably because of its very high level of expression, its small size and the presence of a single peptide chain, which is an advantage for cloning or sequencing procedures. However,

Fab and scFab formats have undeniable advantages, especially when it comes to stability or for reformatting to IgG molecules after engineering.<sup>22</sup> To date, there is no clear literature consensus regarding the choice of an antibody fragment format suitable for engineering using yeast cell display. Also, no general study has yet explored the influence of antibody fragment formats on a diversity of antibodies in this particular context.

Here, we report the comparison of 4 antibody fragment formats – scFv, scFab with and without interchain disulfide bond and Fab – for the functional display of 6 antibodies. This study includes a first set of well-known antibodies – adalimumab, infliximab, golimumab and certolizumab – already used in clinical settings. This set of antibodies encompasses humanized, chimeric and fully human antibodies, which share TNF as a common antigen.<sup>25</sup> The second set is composed of two chimeric antibodies targeting the protein IpaD from the needle tip of type III secretion system of *Shigella* spp obtained from mouse immunization. The objective of this study was to identify the most suitable format for the functional expression of antibody fragments on the surface of yeast cells.

## Results

### Conception of yeast surface display vectors allowing expression of scFv, scFab and Fab antibody fragments

To determine the influence of formats on the expression and functionality of antibody fragments, a new subset of yeast display plasmids was designed. Four plasmids were constructed for respective expression of scFv, scFab formats and Fab at the surface of *S. cerevisiae* cells. Those plasmids share a common backbone from previously described pCT-L7.5.1,<sup>26</sup> including inducible promoter *Gal 1-10* and Aga2p signal peptide [Fig. 1]. Antibody fragments were displayed with a free N-terminus and tethered to the yeast cell wall by C-terminal fusion with Aga2p,<sup>27</sup> in order to avoid possible steric hindrance of the fusion.

The plasmid pNT scFv encodes a typical scFv composed of VH/VL domains connected by a flexible linker (G4S)<sub>3</sub>. We also compared two previously described scFab formats, scFabΔC<sup>2</sup> and sc60Fab,<sup>24</sup> by constructing the plasmids pNT scFabΔC and pNT sc60Fab. ScFabs are composed of a VL-C<sub>K</sub> light chain connected to the VH-CH1 heavy chain by a flexible linker. The difference between the two formats resides in the nature and length of the linker (34 amino acids for scFabΔC versus 60

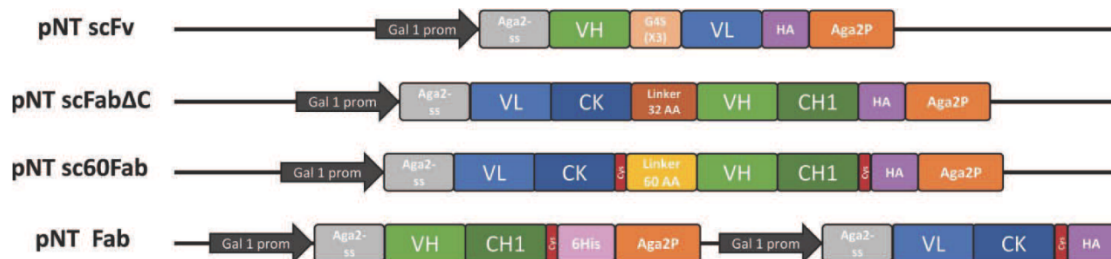
amino acids for sc60Fab) and the presence of an inter-chain disulfide bond in sc60Fab connecting the two C-terminal cysteine residues of C<sub>K</sub> and CH1. Display of Fab molecules on yeast without any linker was achieved using the bigenic plasmid pNT Fab where *Gal 1-10* promoter and terminator are duplicated to obtain separate expression of the two chains.

Each plasmid includes a hemagglutinin (HA) tag to report expression of the complete antibody fragment. For single chain formats (scFv and scFabs), the HA tag is located between the antibody fragment and the Aga2p subunit. Full Fab expression can be controlled for pNT Fab through the expression of the light chain with the HA tag or by direct labeling with anti-C<sub>K</sub> antibody. Expression of heavy chain can be revealed by the HIS tag located between the heavy chain and Aga2p.

The 6 selected antibodies are IgG1 molecules with C<sub>K</sub> light chains. The sequences of TNF-specific antibodies were obtained from the IMGT database.<sup>28</sup> The mAbs IpaD\_301 and IpaD\_318 were obtained by mouse immunization [data not shown] and engineered to generate chimeric antibodies. These two antibodies derive from the same parental genes *IGHV1-14\*01/IGKV1-117\*01* from *Mus musculus* and differ by only 7 and 8 mutations in the heavy chain and the light chain, respectively. Together, 24 plasmids corresponding to the sequences of 6 selected antibodies expressed in the 4 antibody formats were successfully cloned and transformed into yeast. An exhaustive analysis of the 24 constructs was performed in parallel to investigate the influence of the antibody fragment format on antigen binding.

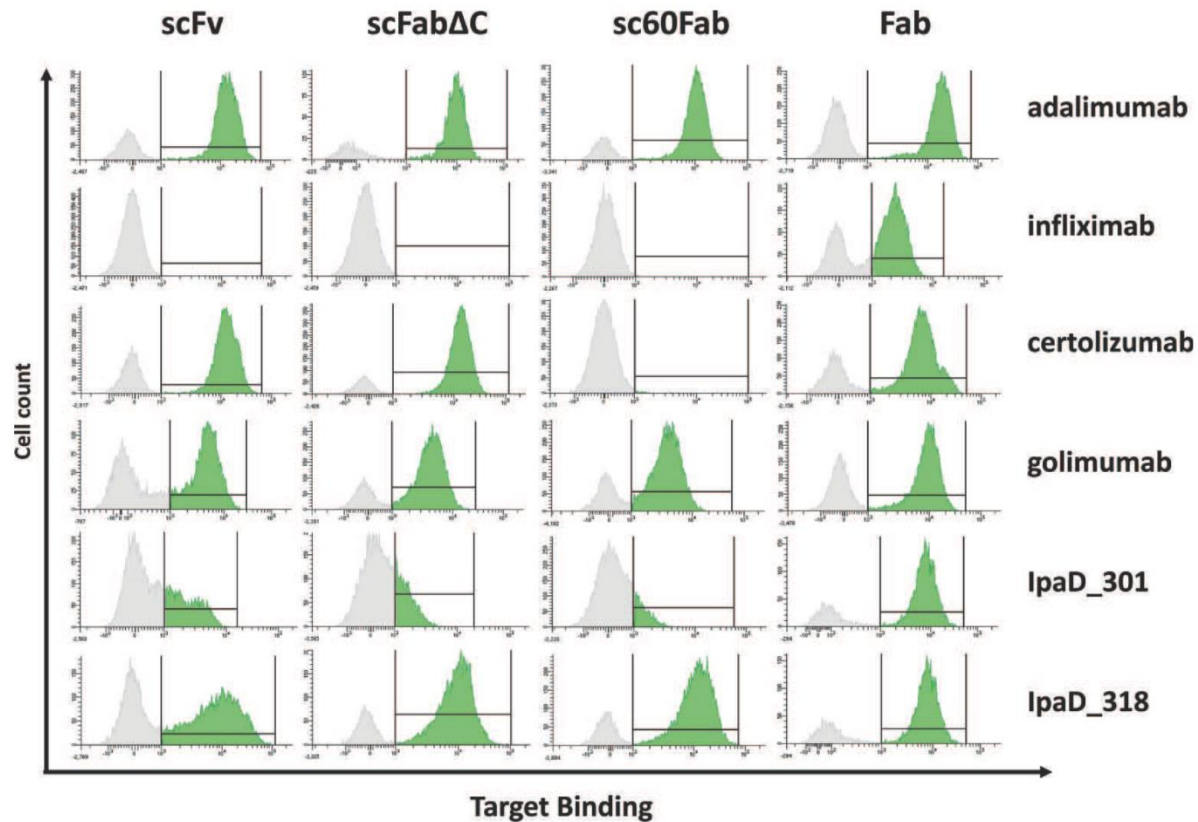
### Fab format grants functional expression of the largest number of antibodies

To assess the expression and functionality of the constructs, yeast cells were incubated with the appropriate biotinylated antigen at a concentration far above the expected K<sub>D</sub> to reach quantitative binding of the antibody fragments to their antigen. When evaluated by fluorescence-activated cell sorting (FACS), cells expressing scFvs of adalimumab, certolizumab, golimumab and IpaD\_318 exhibited a high level of fluorescence, indicating quantitative binding of the antigen [Fig. 2]. Weaker fluorescence levels were observed for IpaD\_301 scFv, whereas no signal was observed for infliximab scFv, indicating antigen binding was weak or absent. Functionality of constructs displaying low or moderate antigen binding was not improved by increased



**Figure 1.** Schematic representation of yeast surface display vectors used in this study. Antibody fragment constructs are tethered to the cell wall of *S. cerevisiae* cells by C-terminal fusion with the Aga2p subunit of the  $\alpha$ -agglutinin yeast cell surface anchor protein. Expression is controlled by the tightly regulated galactose-inducible *Gal1* promoter. Aga2p-ss: native signal sequence of Aga2p for secretion of the constructs. Single-chain antibody fragments (scFv, scFabΔC and sc60Fab) include flexible linkers connecting the light chain and the heavy chain. pNT Fab is a bigenic plasmid including two identical GAL1 promoters for respective expression of light and heavy chains. Cysteine residues responsible for the formation of the inter-chain disulfide bond connecting C<sub>K</sub> and CH1 are indicated in red. Expression of antibody fragments can be monitored using the HA tag for all constructs. A 6His-Tag sequence can also be used for the monitoring of heavy chain expression when using pNT Fab.





**Figure 2.** Antigen binding of the 6 selected antibodies expressed as different antibody fragment formats on the surface of yeast cells. Cells are labeled with biotinylated antigen then with Streptavidin-PE (Target Binding) before FACS analysis. Selected antigen concentrations are far above the expected  $K_D$  to reach a quantitative binding with the antibody fragments. Respective antigen concentration were 20 nM for anti-TNF antibodies and 50 nM for anti-IpaD antibodies. For each construct, a FACS histogram shows the fluorescence intensity corresponding to antigen binding (x-axis) vs the number of detected events (y-axis). Interval gates (bars) are shown to differentiate cells with detectable antigen binding (mean fluorescence intensity (MFI) score above  $10^3$ , in green) from cells with no detectable antigen binding (MFI score under  $10^3$ , in grey). The  $10^3$  cutoff has been obtained by measuring MFI of cells with a non-relevant antigen, cf. Supplementary figure 1.

induction times [Supp Data Fig. 3]. Antigen binding was specific, as demonstrated by the lack of cross-reactivity with the non-relevant antigen: anti-TNF antibodies showed no detectable IpaD binding and anti-IpaD antibodies showed no detectable TNF binding [Supp Data Fig. 1].

Yeast cells expressing scFab corresponding to adalimumab, golimumab and IpaD\_318 displayed quantitative binding to their antigen, both as scFab $\Delta$ C and sc60Fab formats. Infliximab and IpaD\_301 in scFab formats were not or weakly functional, respectively, as attested by the absence of fluorescence for the corresponding cells. Interestingly, certolizumab scFab $\Delta$ C showed significant binding to TNF, while certolizumab sc60Fab with the 60-amino-acid linker and inter-chain disulfide bond failed to bind TNF. All single-chain constructs were expressed at the surface of cells, as attested by labeling with anti-HA antibodies, with the notable exception of certolizumab sc60Fab [Supp Data Fig. 2]. Except for this specific construct, the absence of binding of the antibody fragments to their antigen could not be explained by lack of expression, but rather by a functional deficiency.

In contrast, all of the 6 tested antibodies bound to their antigen when expressed as Fab at the surface of yeast cells. For each antibody, cells expressing functional Fabs form homogeneous populations with a clear shift in fluorescence levels compared to non-expressing cells. The proportion of cells expressing functional Fabs is relatively high (from 46% to 74%).

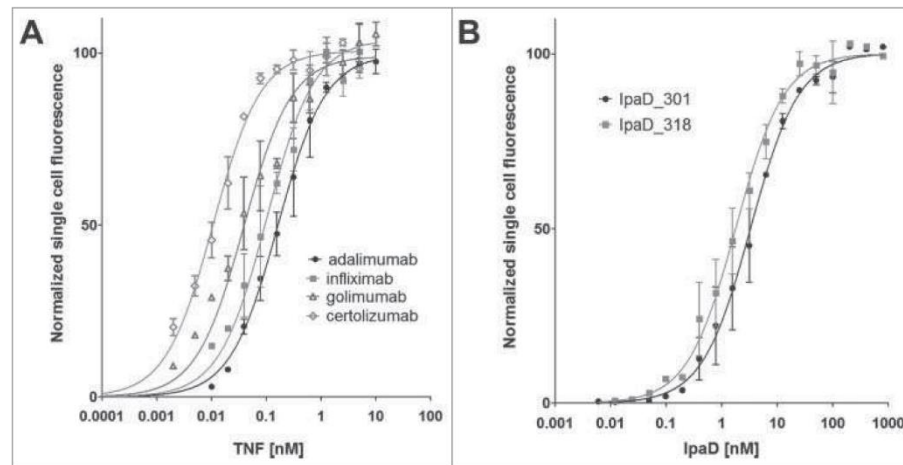
In our set-up, Fab is the only viable format for functional expression of infliximab and IpaD\_301 antibody fragments. Besides, for 5 of 6 antibodies tested, the fluorescence of cells expressing functional Fab was higher than the fluorescence of the equivalent cells expressing either scFvs or scFabs. This suggested that Fab molecules were not only active, but also well expressed. For these reasons, we focused on display of Fab molecules in the rest of the study.

### Variations of expression in Fab format among antibodies do not affect functionality

We observed that maximal fluorescence levels of Fab-expressing cells vary among antibodies targeting a common antigen. More specifically, the mean fluorescence intensity of cells displaying functional infliximab Fabs was much lower than for adalimumab or golimumab, a phenomenon also observed for IpaD\_301 when compared to IpaD\_318.

As variations of signal corresponding to antigen binding could be attributed either to differences of expression levels or to functional differences, dissociation constants ( $K_D$ ) were determined for each displayed Fab [Fig. 3]. The measured  $K_D$  values for IpaD\_301 and IpaD\_318 are 3.2 nM and 1.7 nM, respectively. This result is in accordance with the  $K_D$  evaluated by OctetRed with measured affinities of 1.1 nM and 0.9 nM

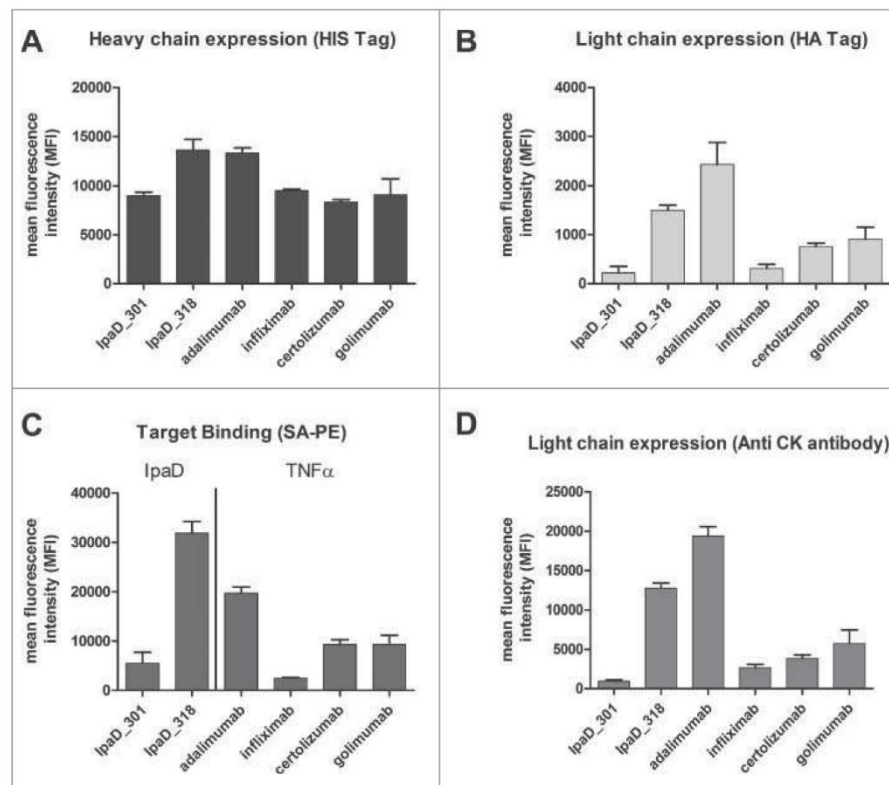




**Figure 3.** Affinity of Fab antibody fragments expressed on the surface of yeast cells. The mean fluorescence intensity (MFI), determined by FACS analysis, is plotted against varying concentrations of antigen for every Fab construct, and fit to a monovalent binding model. Measurements were done in triplicate with independent cultures and inductions. (A) Equilibrium binding constants  $K_D$  of the four anti-TNF Fabs for TNF at 20°C. Curve fitting results in a  $K_D$  of  $114 \pm 6$  nM for adalimumab Fab,  $102 \pm 14$  pM for infliximab Fab,  $39 \pm 4$  pM for golimumab Fab and  $11 \pm 1$  pM for certolizumab Fab. (B) Equilibrium binding constants  $K_D$  of the two anti-IpaD Fabs for IpaD at 20°C.  $K_D$  IpaD\_301 Fab =  $3.2 \pm 0.3$  nM,  $K_D$  IpaD\_318 Fab =  $1.7 \pm 0.2$  nM.

(Supplementary Fig. 4). Yeast expressing anti-TNF Fabs show subnanomolar  $K_D$ , 161 pM for adalimumab, 102 pM for infliximab, 39 pM for golimumab and 11 pM for certolizumab. These  $K_D$  values are consistent with the reported  $K_D$  determined for the corresponding IgG molecules using SPR,<sup>25</sup> indicating that Fabs expressed on the surface of yeast cells have a very strong affinity for their antigen.

Interestingly, higher affinity was not correlated with higher level of binding fluorescence. While having a very good affinity for TNF, infliximab Fab displayed much weaker TNF binding signal compared to adalimumab Fab, suggesting differences in their expression levels. Thus, additional labeling experiments were performed to evaluate different expression reporters. In addition to antigen binding labeling, light chain expression was



**Figure 4.** Antigen binding and expression levels of heavy and light chains of the 6 selected antibodies expressed as Fab on the surface of yeast cells. Cells were labeled with the appropriate expression reporter and the mean fluorescence intensity of positive cells was determined by FACS analysis. For antigen binding, cells were double-labeled with biotinylated antigen and Streptavidin-PE (target binding) before FACS analysis. (A) Heavy chain expression evaluated with anti-HIS antibody reporter. (B) Light chain expression evaluated with anti-HA antibody reporter (C) Biotinylated antigen binding evaluated with Streptavidin-PE (D) Light chain expression evaluated with anti-CK antibody reporter. Measurements were made in triplicate using cells from independent cultures.

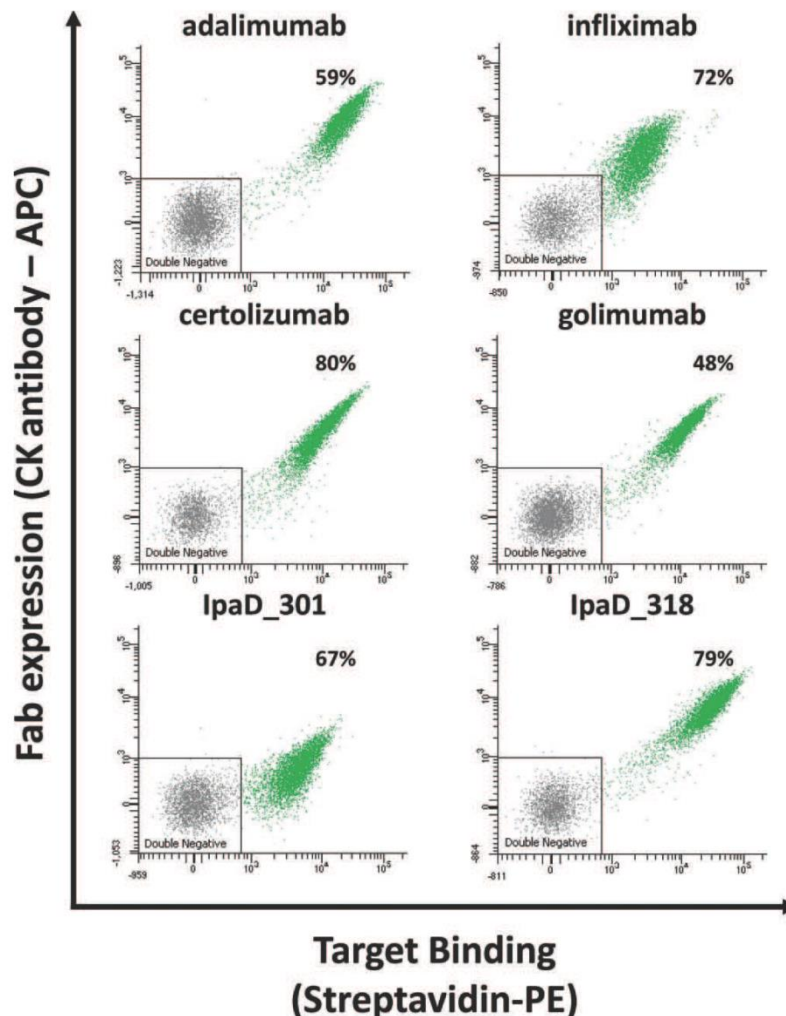
evaluated with anti-HA or anti-C<sub>K</sub> antibodies and heavy chain expression was assessed with anti-HIS antibodies [Fig. 4].

Heavy chain expression showed slight variations between the 6 selected antibodies [Fig. 4a]. Labeling with anti-HIS antibodies showed homogeneous expression of heavy chains, with IpaD\_318 and adalimumab being somewhat above average. This observation contrasted with light chain expression, which showed strong differences across the Fab antibodies [Fig. 4b]. Light chain expression was significantly higher for IpaD\_318 and adalimumab Fabs than for IpaD\_301 or infliximab, while certolizumab and golimumab had intermediate light chain expression levels. Values observed for both reporters anti-HA or anti-C<sub>K</sub> were in good agreement, and the relative intensities of constructs were very similar. Anti-C<sub>K</sub> antibody yielded a very strong fluorescence signal, thus favoring easy discrimination of C<sub>K</sub>-expressing cells.

Antigen binding signals for IpaD\_301 and IpaD\_318 correlated well with their relative light chain expression levels [Fig. 4c and Fig. 5]. Very similarly, the TNF binding signal was proportional to the respective expression levels of adalimumab, infliximab, certolizumab and golimumab light chain of the Fab

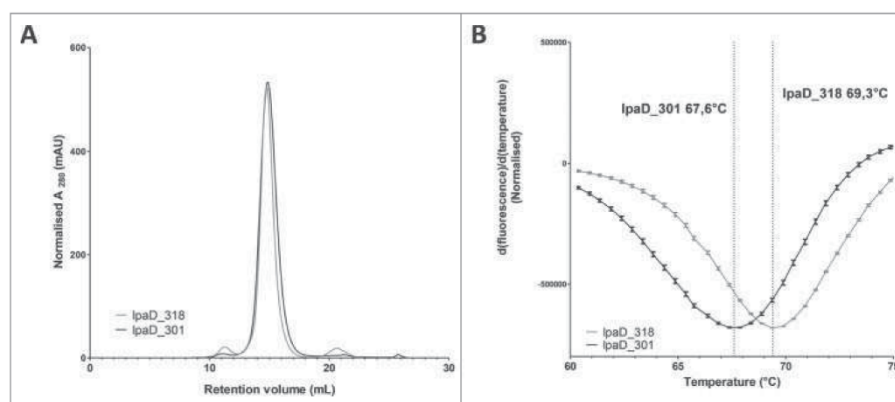
[Fig. 5]. Direct correlation between light chain expression and antigen binding was confirmed by simultaneous labeling with anti-C<sub>K</sub> antibody conjugated to allophycocyanin (APC) and streptavidin R-phycoerythrin (streptavidin-PE) [Fig. 4d].

While having high sequence homology and similar affinities, IpaD\_301 and IpaD\_318 display significant differences in expression and antigen binding in all formats tested. For this reason, we investigated a possible correlation between the biophysical attributes of these two antibodies and their behavior on the surface of yeast cells. Thus, recombinant expression of complete IgG1 IpaD\_301 and IpaD\_318 was performed by transient transfection of HEK293 cells<sup>29</sup> followed by protein A purification. Production yield was slightly higher for IpaD\_318 than for IpaD\_301 (20 and 14 mg per liter, respectively). Size-exclusion chromatography revealed that both antibodies are predominantly produced as full-length IgG with low amounts of aggregates for IpaD\_318 and even lower amounts for IpaD\_301 [Fig. 6]. In contrast, differential scanning fluorimetry (DSF) analysis reveals that IpaD\_301 has a lower melting temperature (T<sub>m</sub>) than IpaD\_318, with T<sub>m</sub> values of 67.6  C and 69.3  C, respectively [Fig. 6].



**Figure 5.** Representative bivariate flow cytometric analysis of yeast cells expressing Fab antibody fragments expressed on their surface. Each single dot on the plot designates the APC (y-axis) and PE (x-axis) fluorescence intensity values for a single yeast cell. Cells were double-labeled with biotinylated antigen/Streptavidin-PE (target binding) and anti-CK APC labels (Fab expression). Antigen concentrations in labeling experiments were 20 nM for anti-TNF antibodies and 50 nM for anti-IpaD antibodies. Functional subpopulations were gated using an inverted gate of the double negative sub-populations. The percentage number correspond to the mean percentage of cells expressing functional Fabs from three separate inductions.





**Figure 6.** Biophysical characterization of IpaD\_301 IgG and IpaD\_318 IgG expressed by HEK cells. (A): Size-exclusion chromatograms of the protein A purified IgGs on a Superdex 200 10/300 GL size-exclusion column for IpaD\_301 IgG (black) and IpaD\_318 IgG (grey). Monomeric IgG species elutes at 15 mL and aggregates or multimeric species elutes between 10 to 12 mL. (B): T<sub>m</sub> determination of IgGs by differential scanning fluorimetry. The fluorescence intensities are normalized to peak values. Error bars correspond to experiment triplicates.

## Discussion

We carried out a comparative study of expression formats using 6 different antibodies specific for two different antigens, TNF and IpaD. In our hands, Fab appears to be the most versatile expression format for display of functional antibodies on the surface of yeast cells.

ScFv is currently the most widespread format for antibody engineering. It has indisputable advantages, such as its small size, which is particularly convenient for the cloning and sequencing of libraries. Also, numerous studies have used scFvs successfully to engineer clones with improved properties,<sup>4</sup> including clones derived from adalimumab with improved affinity towards TNF.<sup>8</sup> However, some antibodies display reduced affinity and impaired biophysical properties when expressed as scFv as compared to the corresponding Fab molecules.<sup>23,30</sup> Indeed, CH1 and CL domains, which are not present in scFv molecules, participate in the greater stability of Fab and scFab.<sup>31,32</sup> The choice of an adapted format has also proved to be an important parameter for other approaches such as phage display<sup>33,34</sup> or recombinant expression of antibody fragments.<sup>35</sup>

Here, we show that some antibodies do not seem to be active either in the scFv format or in the scFab format when they are expressed at the surface of yeast cells. In our set-up, two antibodies, infliximab and IpaD\_301, appeared to lack functionality, as suggested by the total absence of antigen binding in any single-chain format. The presence of C<sub>K</sub> and CH1 domains, which could have a stabilizing role in the scFab formats, had no effect on the functionality of the corresponding infliximab and IpaD\_301 constructs. We cannot exclude that the presence of a linker between the light chain and the heavy chain could also cause steric hindrance that impairs functionality.

We found that adalimumab, golimumab and IpaD\_318 were well expressed and perfectly able to bind their antigen as scFv constructs, but also in scFab or Fab formats. Certolizumab was also a very well expressed and functional antibody in these formats, with the notable exception of the sc60Fab format, which exhibited very poor expression and had no detectable antigen binding (Supp. Fig 2). These particular results emphasize that the presence of the two C-terminal cysteine residues in scFab, as well as the nature and length of the linker sequence, can

dramatically alter expression of scFab molecules for some antibodies. Moreover, the presence of the interchain disulfide bond can have a major influence on the stability of the antibody fragment, as recently shown for the adalimumab Fab.<sup>36</sup> The length of the linker also proved to be an important parameter for the proper packing of light and heavy chains and possibly the accessibility of the paratope, thus ultimately affecting the function of the scFab fragment.<sup>23,24</sup>

In this study set-up, Fab was the only format that allowed functional display for each of the 6 tested antibodies. FACS analysis revealed that a very high proportion of induced yeast cells (up to 70%) expressed Fabs and produced a strong fluorescence signal corresponding to antigen binding. In the context of protein engineering or affinity maturation, a strong signal would favor an easy discrimination of high-affinity clones in libraries containing very high numbers of mutants.

Moreover, for anti-TNF molecules, all Fabs expressed on the surface of yeast cells exhibited a high affinity for their antigen, with K<sub>D</sub> values close to those reported in the literature, strongly suggesting that the K<sub>D</sub> of a protein-protein binding interaction measured on the surface of yeast cells was essentially equal to that measured using soluble proteins.<sup>37,38</sup>

Deeper analysis of expression reporters in the Fab format indicated a lower display level of light chains for infliximab and IpaD\_301 Fabs. Light chain display correlated perfectly with lower target binding of these two constructs compared to the other anti-TNF Fabs or IpaD\_318 Fab, respectively. Coexpression of chaperones such as BiP, use of alternative promoters such as *Gal10* for the light chain expression, or introduction of a new secretory leader sequence optimized for antibody secretion such as  $\alpha$ MFpp8<sup>39</sup> might be useful to promote the assembly of light and heavy chains of antibodies during Fab expression. A recent study demonstrated that co-expression of protein disulfide isomerase (PDI) had a positive effect on the functionality of some antibodies expressed as Fab on the surface of yeast cells.<sup>40</sup> In this study, Georgiou and co-workers also showed that pairing of antibody chains could be promoted by fusing the light and heavy chains with dimerization domains such as leucine-zipper proteins. Forced-dimerization of antibody chains restored functionality for 3 of 13 tested anti-HA

antibodies, although with a lower proportion of cells displaying functional Fab.

Overall, infliximab and IpaD\_301 exhibited relatively low expression in every tested format of antibody fragment. We observed, however, marked differences in the expression profiles of IpaD\_301 and IpaD\_318 on the surface of yeast cells in spite of their high sequence homology, with only 7 and 8 mutations in the heavy chain and the light chain, respectively. Production of these two antibodies in a full-length IgG format reveal that IpaD\_301 has lower production titers in HEK cells and also a lower  $T_m$  (67.6°C vs 69.3°C). Interestingly, infliximab is also described as difficult to express in HEK cells<sup>41</sup> and CHO cells,<sup>42</sup> with relatively low stability over time.<sup>43</sup> A comprehensive study comparing the biophysical properties of clinical-stage antibodies indicated that infliximab has much lower production titers in HEK cells than adalimumab, certolizumab and golimumab (6 mg/L for infliximab vs more than 100 mg/mL for the three other anti-TNF antibodies).<sup>41</sup> Wittrup and co-workers also reported that infliximab had a lower  $T_m$  (64.5°C vs at least 70°C for the three other anti-TNF antibodies). These observations are consistent with previously published studies indicating that protein secretion by yeast cells is correlated with thermodynamic stability.<sup>44,45</sup> Together, this suggests that antibodies with low thermodynamic stability might have low expression on the surface of yeast cells.

Even as many alternative antibody formats are emerging, most of current therapeutic antibodies on the market are antibodies in the IgG format.<sup>21</sup> Thus, it is essential that the conversion between the screening format – the antibody fragment – and the IgG antibody does not impair the desired properties, and in particular affinity for the antigen. From this perspective, use of Fab has strong advantages. Several reports show that the Fab format tends to retain its binding affinity better than scFv when converted into IgG.<sup>19</sup> Moreover, the presence of a peptide linker in the scFv and scFab formats or introduction of additional domains such as leucine-zipper dimerization proteins might introduce bias in the screening of improved clones because of subtle structural changes. Indeed, some affinity-enhanced molecules engineered in scFv and scFab $\Delta$ C formats lost affinity after conversion into IgG.<sup>22</sup>

In conclusion, in spite of their practical convenience, single-chain antibody fragment formats such as scFv and scFab do not always lead to a functional display of antibodies at the surface of yeast cells. In contrast, we showed that all 6 tested antibodies were functional in the Fab format, with high affinities for their antigen. Using our Fab expression plasmid, high proportions of yeast cells expressed Fab molecules and exhibited strong antigen binding signals in FACS, two important properties when getting started with the affinity maturation process by yeast display in good conditions. These advantages and the lower risk of losing engineered properties upon conversion into IgG suggest that the Fab format is the best alternative for the engineering of antibodies using YSD.

## Material and methods

### Yeast display vectors construction

Cloning were performed with the SLiCE recombination cloning method<sup>46</sup> or classic restriction/ligature method using *E.coli*

DH5 $\alpha$  strain (Invitrogen). For each anti-TNF antibody, optimized synthetic genes (VH and VL) were ordered from Eurofins Genomics based on the IMGT sequences. All expression plasmids are derived from previously described pCT-L7.5.1<sup>26</sup> and share the *CEN/ARS* replication origin, *TRP* auxotrophy marker, *colE1* replication origin, Ampicillin resistance gene and galactose inducible promoter *GAL1*. Plasmid pCT-L7.5.1 was a gift from Prof. K. Dane Wittrup (Addgene plasmid # 42900).

Plasmids pNT scFv were modified from parental pCT-L7.5.1 plasmid by modifying the aga2p signal sequence with a dipeptide spacer<sup>27</sup> and a *NheI* restriction site. A stop codon was inserted at the 3' end of aga2p gene to obtain plasmids allowing the display of scFv with a free N-terminus end and a C-terminal anchoring to the yeast cell wall. Synthetic genes corresponding to each scFv were then cloned using *NheI* and *SpeI* restriction sites.

The pNT scFab $\Delta$ C vector was generated by cloning the coding sequence of the scFab $\Delta$ C of the previously described pYDscFab<sup>2</sup> between the restriction sites *NheI* and *XhoI* of the pNT scFv plasmid. Plasmid pYDscFab was a gift from Prof. Dennis Burton. Heavy chains and light chains of antibodies were cloned between *NheI*/*Pfl23II* and *NcoI*/*Sall* restriction sites, respectively.

The pNT sc60Fab vectors were generated from pNT scFab plasmids. The linker and the cysteine residues allowing inter-chain disulfide bond formation were modified by PCR and homologous recombination between *BssHII* and *NcoI* sites. Heavy chains and light chains of antibodies were cloned between *NheI*/*Pfl23II* and *NcoI*/*Sall* restriction sites, respectively.

The bigenic pNT Fab vector was generated by ligation into pNT scFv between *SpeI* and *XhoI* of a synthetic gene (Eurofins Genomics) containing a 6-HIS tag, the Aga2p domain, a *Gall* promoter, an aga2p signal sequence, an HA tag and a transcription terminator. The resulting plasmid has two distinct expression cassettes corresponding to respective expression of light and heavy chains of the Fabs. Heavy chains and light chains of antibodies were then cloned between *NheI*/*Pfl23II* and *NcoI*/*Sall* restriction sites, respectively.

### Yeast transformation

Preparation of competent yeast cells EBY100 (ATCC<sup>®</sup> MYA-4941<sup>TM</sup>; a *GAL1-AGA1::URA3 ura3-52 trp1 leu2  $\Delta$ 1 his3  $\Delta$ 200 pep4::HIS2 prb1 $\Delta$ 1.6R can1 GAL*) was performed according to Suga et al.<sup>47</sup> Then, 100  $\mu$ L of EBY100 electro-competent cells were mixed with 1–2  $\mu$ g plasmid DNA, transferred to a pre-chilled electroporation cuvette (Biorad, 165–2086) and pulsed at 2.5 kV, 25  $\mu$ F (Biorad Gene Pulser Xcell). Yeast cells were then diluted in 1 mL of sorbitol (1 M), and 200  $\mu$ L were streaked on SD-CAA agar plates [6.7 g/L yeast nitrogen base without casamino acids, 20 g/L dextrose, 5 g/L casamino acids, 100 mM sodium phosphate pH 6.0].

### Growth and expression conditions

Yeast cultures were performed in 24 deep-well round-bottom plates. Pre-cultures were performed by inoculating 3 mL of SD-CAA medium with one colony from selective agar plate and incubated overnight at 30°C, 200 rpm. The saturated pre-



culture (typically OD<sub>600</sub> of 8–10) was passaged in order to obtain an initial culture OD<sub>600</sub> of 0.25–0.50. The culture was grown at 30°C until its OD<sub>600</sub> reached 0.5–1.0. Cells were centrifuged and re-suspended in 3 mL of SG-CAA galactose induction medium [6.7 g/L yeast nitrogen base without casamino acids, 20 g/L galactose, 5 g/L casamino acids, 100 mM sodium phosphate, pH 6.0] and induced for 16–36 h at 20°C, 200 rpm.

### Yeast labelling and flow cytometry analysis

10<sup>6</sup> cells were taken from the induced culture and washed with 1 mL PBSF (phosphate-buffered saline (PBS), bovine serum albumin (BSA) 0.1%) buffer. Cells were resuspended in 50 µL of the target protein at the desired concentration and incubated at 20°C, 1000 rpm until equilibrium is reached (depending on target protein concentration). Cells were washed with 1 mL of ice-cold PBSF to avoid dissociation and resuspended in ice-cold PBSF containing the appropriate fluorescent reporters: Streptavidin-PE (Thermo Fisher scientific; catalog number S866; 1:100 dilution), Mouse anti-Human Kappa Light Chain Antibody APC conjugate (Thermo Fisher scientific; catalog number MH10515; 1:100 dilution), 6x-His Epitope Tag antibody Dylight-650 conjugate (Thermo Scientific, catalog number MA1-21315-D650; 1:100 dilution) or HA Tag Monoclonal Antibody APC conjugate (1:50 dilution). Cells were incubated on ice in the dark for 15 minutes and analyzed with BD FACS Aria<sup>TM</sup> III cytometer.

### Cell-Binding assays for determining the affinity of antibodies for their antigen

Cell binding assays were performed according to Hunter *et al.*<sup>38</sup> in order to avoid ligand depletion. 10<sup>4</sup> to 10<sup>5</sup> cells were incubated in 50–500 µL of the target protein at the desired concentrations and incubated at 20°C, 1000 rpm long enough to reach equilibrium. Cells were analyzed by FACS for their binding fluorescence as described above. Auto-fluorescence was subtracted for each measure and resulting values were normalized with maximum fluorescence. Data are then fitted with a non-linear regression analysis to obtain the K<sub>D</sub>.

### Production and purification of IgG IpaD\_301 and IpaD\_318

IpaD\_301 and IpaD\_318 antibodies were produced via transient transfection of HEK293 Freestyle<sup>TM</sup> cells (HEK 293FS). HEK 293FS were transfected as previously described by Subedi *et al.*<sup>48</sup> without the use of Valproic acid using the HEK transient transfection plasmids described by Smith *et al.*<sup>29</sup>

The day before transfection, cells were sub-cultured in order to be at a density of 2–3.10<sup>6</sup> cells/mL the next day. The day of the transfection, cells were centrifuged 5 min at 100 g and resuspended in fresh FreeStyle 293 Expression medium (Thermo-Fisher) to obtain a density of 2.5.10<sup>6</sup> cells/mL. For a transfection volume of 100 mL, 150 µg of each plasmid (coding for heavy chain and light chain) was added to the cells at a concentration of 0.5 µg/µL. Cells were incubated 5 min at 37°C, 180 rpm, 8% CO<sub>2</sub>. 1.8 mL of polyethylenimine (PEI) was added

at a concentration of 0.5 mg/mL to the cells (PEI:DNA ratio = 3:1). Cells were subsequently incubated 24h at 37°C, 180 rpm, 8% CO<sub>2</sub>.

A day after transfection, 100 mL of fresh EX-CELL<sup>®</sup> Serum-Free Medium for HEK 293 (Sigma-Aldrich), supplemented with 6 mM L-Glutamine, was added to the transfected culture. Cells were finally incubated 4 more days at 37°C, 180 rpm, 8% CO<sub>2</sub>. Five days after transfection, the supernatant was harvested by centrifuging the cells at 1000g for 15 min at 4°C and kept on ice until purification. The supernatant was purified using a HiTrap<sup>TM</sup> Protein A HP column of 1 mL (GE Healthcare) on an ÄKTA purifier (GE Healthcare) according to the manufacturer's description.

### Size-exclusion chromatography

A Superdex-200 10/300 GL (GE Healthcare) column was used on an ÄKTA purifier (GE Healthcare). Runs were monitored with Unicorn<sup>TM</sup> 5.31 software (GE Healthcare). 500 µL of each sample was applied on the column using a flow rate of 0.7 mL/min and PBS as running buffer.

### Fab IpaD\_301 and IpaD\_318 affinity determination

Fab were prepared from full-length IgG using a Pierce<sup>®</sup> Fab Micro Preparation Kit. The binding affinity of Fab IpaD\_301 and IpaD\_318 was evaluated by biolayer interferometry using an Octet RED96 instrument (Pall ForteBio). Biotinylated IpaD was loaded onto streptavidin biosensors (ForteBio) at a concentration of 5 µg/mL for 8 min in PBSF (1x PBS, pH 7.4, 0.1% BSA). Fab IpaD\_301 and IpaD\_318 were then associated with IpaD using 4 different concentrations (20, 10, 5 and 2.5 nM). Duration of association and dissociation steps were 20 min and 120 min. All steps were performed at 20°C. Data were collected and analyzed using Octet Software version 10.0.3 (Pall ForteBio). Binding kinetics were fitted using a 1:1 Langmuir-binding model.

### Thermostability evaluation

DSF was performed as previously described by Niesen *et al.*<sup>49</sup> using a StepOne Real-Time PCR System from Applied Biosystems. IgG monomers of IpaD\_301 and IpaD\_318 purified via SEC were mixed with the SYPRO Orange reagent. Final dye concentration was 5x and final IgG concentration was 150 µg/mL. The temperature gradient was run from 20 to 90 °C with 1 min equilibration at each degree centigrade.

### Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

### Abbreviations

BSA	bovine serum albumin
DSF	differential scanning fluorimetry
Fab	antigen-binding fragment
MFI	mean fluorescence intensity
PDI	protein disulfide isomerase

PEI polyethylenimine  
scFab single-chain antigen-binding fragment  
scFv single-chain variable fragment  
SEC size exclusion chromatography  
TNF tumor necrosis factor  
YSD yeast surface display

## Acknowledgments

The authors would like to thank Dr Magali Aumont-Nicaise for her assistance with differential scanning fluorimetry experiments and Dr Herv   Volland for anti-HA antibodies.

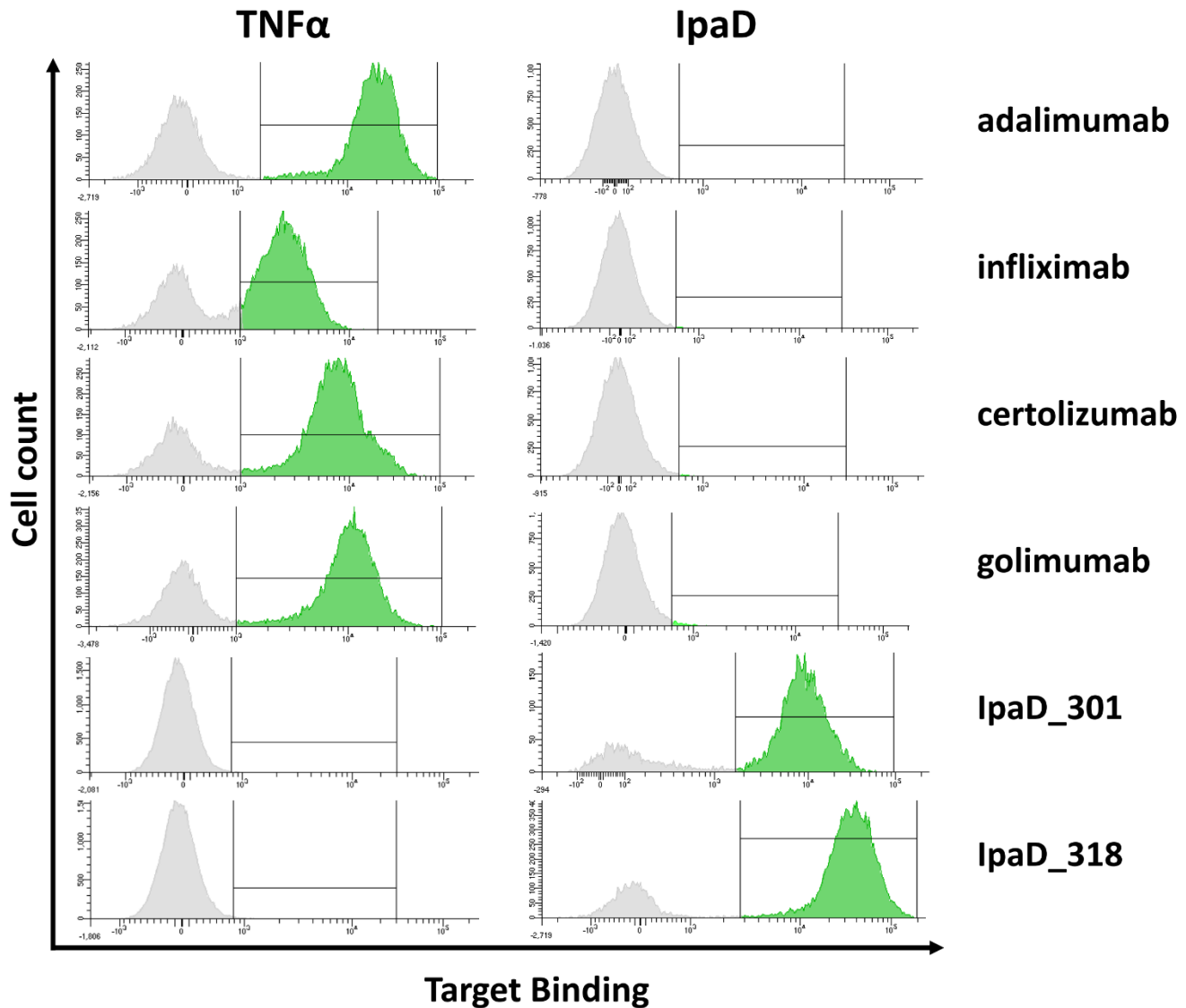
## References

- Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, Wittrup KD. Isolating and engineering human antibodies using yeast surface display. *Nat Protoc.* 2006;1:755–68. doi:10.1038/nprot.2006.94. PMID:17406305.
- Walker LM, Bowley DR, Burton DR. Efficient recovery of high-affinity antibodies from a single-chain Fab yeast display library. *J Mol Biol.* 2009;389:365–75. doi:10.1016/j.jmb.2009.04.019. PMID:19376130.
- Wang B, Lee CH, Johnson EL, Kluwe CA, Cunningham JC, Tanno H, Crooks RM, Georgiou G. Discovery of high affinity anti-ricin antibodies by B cell receptor sequencing and by yeast display of combinatorial VH:VL libraries from immunized animals. *mAbs.* 2016;8:1035–44. doi:10.1080/19420862.2016.1190059. PMID:27224530.
- Boder ET, Raeeszadeh-Sarmazdeh M, Price JV. Engineering antibodies by yeast display. *Arch Biochem Biophys.* 2012;526:99–106. doi:10.1016/j.abb.2012.03.009. PMID:22450168.
- Pepper LR, Cho YK, Boder ET, Shusta EV. A decade of yeast surface display technology: where are we now? *Comb Chem High Throughput Screen.* 2008;11:127–34. doi:10.2174/138620708783744516. PMID:18336206.
- Angelini A, Chen TF, de Picciotto S, Yang NJ, Tzeng A, Santos MS, Van Deventer JA, Traxlmayr MW, Wittrup KD. Protein engineering and selection using yeast surface display. *Methods Mol Biol.* 2015;1319:3–36. doi:10.1007/978-1-4939-2748-7\_1. PMID:26060067.
- Gera N, Hussain M, Rao BM. Protein selection using yeast surface display. *Methods.* 2013;60:15–26. doi:10.1016/j.ymeth.2012.03.014. PMID:22465794.
- Rajpal A, Beyaz N, Haber L, Cappuccilli G, Yee H, Bhatt RR, Takeuchi T, Lerner RA, Crea R. A general method for greatly improving the affinity of antibodies by using combinatorial libraries. *Proc Natl Acad Sci U S A.* 2005;102:8466–71. doi:10.1073/pnas.0503543102. PMID:15939870.
- Boder ET, Midelfort KS, Wittrup KD. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proc Natl Acad Sci U S A.* 2000;97:10701–5. doi:10.1073/pnas.170297297. PMID:10984501.
- Lou J, Geren I, Garcia-Rodriguez C, Forsyth CM, Wen W, Knopp K, Brown J, Smith T, Smith LA, Marks JD. Affinity maturation of human botulinum neurotoxin antibodies by light chain shuffling via yeast mating. *Protein Eng Des Sel.* 2010;23:311–9. doi:10.1093/protein/gzq001. PMID:20156888.
- Graff CP, Chester K, Begent R, Wittrup KD. Directed evolution of an anti-carcinoembryonic antigen scFv with a 4-day monovalent dissociation half-time at 37 degrees C. *Protein Eng Des Sel.* 2004;17:293–304. doi:10.1093/protein/gzh038. PMID:15115853.
- Zhou Y, Drummond DC, Zou H, Hayes ME, Adams GP, Kirpotin DB, Marks JD. Impact of single-chain Fv antibody fragment affinity on nanoparticle targeting of epidermal growth factor receptor-expressing tumor cells. *J Mol Biol.* 2007;371:934–47. doi:10.1016/j.jmb.2007.05.011. PMID:17602702.
- van den Beucken T, Pieters H, Steukers M, van der Vaart M, Ladner RC, Hoogenboom HR, Hufton SE. Affinity maturation of Fab antibody fragments by fluorescent-activated cell sorting of yeast-displayed libraries. *FEBS Lett.* 2003;546:288–94. doi:10.1016/S0014-5793(03)00602-1. PMID:12832056.
- Schroter C, Gunther R, Rhiel L, Becker S, Toleikis L, Doerner A, Becker J, Sch  nemann A, Nasu D, Neuteboom B, et al. A generic approach to engineer antibody pH-switches using combinatorial histidine scanning libraries and yeast display. *mAbs.* 2015;7:138–51. doi:10.4161/19420862.2014.985993. PMID:25523975.
- Garcia-Rodriguez C, Levy R, Arndt JW, Forsyth CM, Razai A, Lou J, Geren I, Stevens RC, Marks JD. Molecular evolution of antibody cross-reactivity for two subtypes of type A botulinum neurotoxin. *Nat Biotechnol.* 2007;25:107–16. doi:10.1038/nbt1269. PMID:17173035.
- Rosowski S, Becker S, Toleikis L, Valldorf B, Grzeschik J, Demir D, Willenb  cher I, Gaa R, Kolmar H, Zielonka S, et al. A novel one-step approach for the construction of yeast surface display Fab antibody libraries. *Microb Cell Fact.* 2018;17:3. doi:10.1186/s12934-017-0853-z. PMID:29316915.
- Rhiel L, Krah S, Gunther R, Becker S, Kolmar H, Hock B. REAL-Select: full-length antibody display and library screening by surface capture on yeast cells. *PLoS One.* 2014;9:e114887. doi:10.1371/journal.pone.0114887. PMID:25501029.
- Rakestraw JA, Aird D, Aha PM, Baynes BM, Lipovsek D. Secretion-and-capture cell-surface display for selection of target-binding proteins. *Protein Eng Des Sel.* 2011;24:525–30. doi:10.1093/protein/gzr008. PMID:21402751.
- Weaver-Feldhaus JM, Lou JL, Coleman JR, Siegel RW, Marks JD, Feldhaus MJ. Yeast mating for combinatorial Fab library generation and surface display. *FEBS Lett.* 2004;564:24–34. doi:10.1016/S0014-5793(04)00309-6. PMID:15094038.
- Gaciarz A, Veijola J, Uchida Y, Saaranen MJ, Wang C, Horkko S, Rudnick LW. Systematic screening of soluble expression of antibody fragments in the cytoplasm of *E. coli*. *Microb Cell Fact.* 2016;15:22. doi:10.1186/s12934-016-0419-5. PMID:26809624.
- Holliger P, Hudson PJ. Engineered antibody fragments and the rise of single domains. *Nat Biotechnol.* 2005;23:1126–36. doi:10.1038/nbt1142. PMID:16151406.
- Steinwand M, Droste P, Frenzel A, Hust M, Dubel S, Schirrmann T. The influence of antibody fragment format on phage display based affinity maturation of IgG. *mAbs.* 2014;6:204–18. doi:10.4161/mabs.27227. PMID:24262918.
- Hust M, Jostock T, Menzel C, Voedisch B, Mohr A, Brenneis M, Kirsch MI, Meier D, D  bel S. Single chain Fab (scFab) fragment. *BMC Biotechnol.* 2007;7:14. doi:10.1186/1472-6750-7-14. PMID:17346344.
- Koerber JT, Hornsby MJ, Wells JA. An improved single-chain Fab platform for efficient display and recombinant expression. *J Mol Biol.* 2015;427:576–86. doi:10.1016/j.jmb.2014.11.017. PMID:25481745.
- van Schie KA, Ooijevaar-de Heer P, Dijk L, Kruithof S, Wolbink G, Rispens T. Therapeutic TNF inhibitors can differentially stabilize Trimeric TNF by inhibiting monomer exchange. *Scientific Reports.* 2016;6:32747. doi:10.1038/srep32747. PMID:27605058.
- Hackel BJ, Kapila A, Wittrup KD. Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling. *J Mol Biol.* 2008;381:1238–52. doi:10.1016/j.jmb.2008.06.051. PMID:18602401.
- Wang Z, Mathias A, Stavrou S, Neville DM, Jr. A new yeast display vector permitting free scFv amino termini can augment ligand binding affinities. *Protein Eng Des Sel.* 2005;18:337–43. doi:10.1093/protein/gzi036. PMID:15976011.
- Lefranc MP, Giudicelli V, Duroux P, Jabado-Michaloud J, Folch G, Aouinti S, Carillon E, Duvergey H, Houles A, Paysan-Lafosse T, et al. IMGT(R), the international ImMunoGeneTics information system(R) 25 years on. *Nucleic Acids Res.* 2015;43:D413–22. doi:10.1093/nar/gku1056. PMID:25378316.
- Smith K, Garman L, Wrammert J, Zheng NY, Capra JD, Ahmed R, Wilson PC. Rapid generation of fully human monoclonal antibodies specific to a vaccinating antigen. *Nat Protoc.* 2009;4:372–84. doi:10.1038/nprot.2009.3. PMID:19247287.
- Worn A, Pluckthun A. Stability engineering of antibody single-chain Fv fragments. *J Mol Biol.* 2001;305:989–1010. doi:10.1006/jmbi.2000.4265. PMID:11162109.

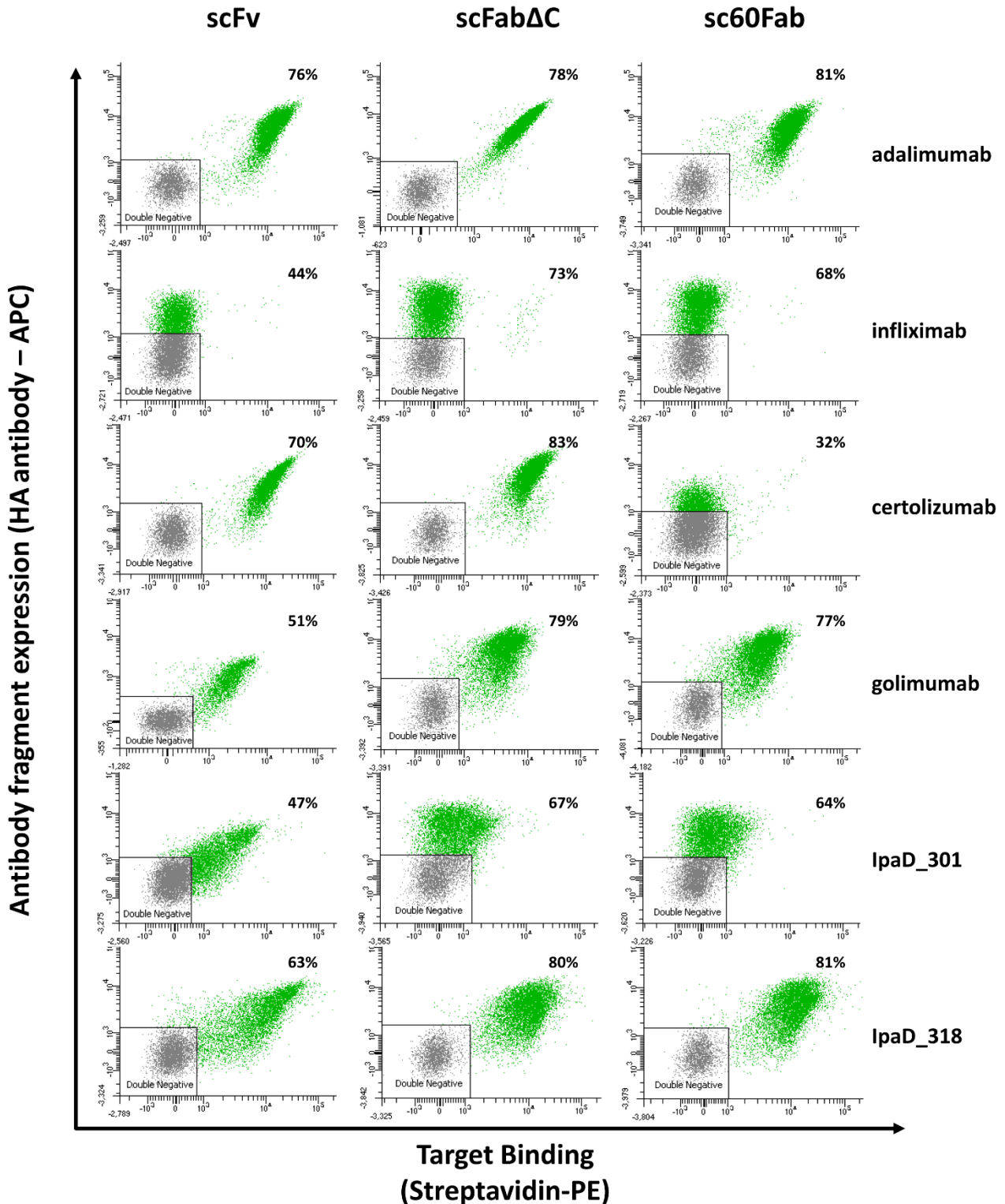


31. Rothlisberger D, Honegger A, Pluckthun A. Domain interactions in the Fab fragment: a comparative evaluation of the single-chain Fv and Fab format engineered with variable domains of different stability. *J Mol Biol.* 2005;347:773–89. doi:10.1016/j.jmb.2005.01.053. PMID:15769469.
32. Quintero-Hernandez V, Juarez-Gonzalez VR, Ortiz-Leon M, Sanchez R, Possani LD, Becerril B. The change of the scFv into the Fab format improves the stability and in vivo toxin neutralization capacity of recombinant antibodies. *Mol Immunol.* 2007;44:1307–15. doi:10.1016/j.molimm.2006.05.009. PMID:16814388.
33. Chan CE, Chan AH, Lim AP, Hanson BJ. Comparison of the efficiency of antibody selection from semi-synthetic scFv and non-immune Fab phage display libraries against protein targets for rapid development of diagnostic immunoassays. *J Immunol Methods.* 2011;373:79–88. doi:10.1016/j.jim.2011.08.005. PMID:21856306.
34. Hust M, Jostock T, Menzel C, Voedisch B, Mohr A, Brenneis M, Kirsch MI, Meier D, Dübel S. Single chain Fab (scFab) fragment. *BMC Biotechnol.* 2007;7:14. doi:10.1186/1472-6750-7-14. PMID:17346344.
35. Rao MF, Li YJ, Dong JX, Wu WJ, Xu ZL, Sun YM, et al. Production and characterization of a single-chain Fab fragment for the detection of O,O-diethyl organophosphorus pesticides. *Anal Methods-Uk.* 2016;8:3140–7. doi:10.1039/C6AY00224B.
36. Nakamura H, Oda-Ueda N, Ueda T, Ohkuri T. A novel engineered interchain disulfide bond in the constant region enhances the thermostability of adalimumab Fab. *Biochem Biophys Res Commun.* 2018;495:7–11. doi:10.1016/j.bbrc.2017.10.140. PMID:29097200.
37. Cherf GM, Cochran JR. Applications of yeast surface display for protein engineering. *Methods Mol Biol.* 2015;1319:155–75. doi:10.1007/978-1-4939-2748-7\_8. PMID:26060074.
38. Hunter SA, Cochran JR. Cell-binding assays for determining the affinity of protein-protein interactions: Technologies and considerations. *Methods Enzymol.* 2016;580:21–44. doi:10.1016/bs.mie.2016.05.002. PMID:27586327.
39. Rakestraw JA, Szazinsky SL, Piatasi A, Antipov E, Wittrup KD. Directed evolution of a secretory leader for the improved expression of heterologous proteins and full-length antibodies in *Saccharomyces cerevisiae*. *Biotechnol Bioeng.* 2009;103:1192–201. doi:10.1002/bit.22338. PMID:19459139.
40. Wang B, DeKosky BJ, Timm MR, Lee J, Normandin E, Misasi J, Kong R, McDaniel JR, Delidakis G, Leigh KE, et al. Functional interrogation and mining of natively paired human VH:VL antibody repertoires. *Nat Biotechnol.* 2018;36(2):152–15. doi:10.1038/nbt.4052. PMID:29309060.
41. Jain T, Sun T, Durand S, Hall A, Houston NR, Nett JH, Sharkey B, Bobrowicz B, Caffry I, Yu Y, et al. Biophysical properties of the clinical-stage antibody landscape. *Proc Natl Acad Sci U S A.* 2017;114:944–9. doi:10.1073/pnas.1616408114. PMID:28096333.
42. Le Fourn V, Girod PA, Buceta M, Regamey A, Mermod N. CHO cell engineering to prevent polypeptide aggregation and improve therapeutic protein secretion. *Metab Eng.* 2014;21:91–102. doi:10.1016/j.ymben.2012.12.003. PMID:23380542.
43. Lerch TF, Sharpe P, Mayclin SJ, Edwards TE, Lee E, Conlon HD, Polleck S, Rouse JC, Luo Y, Zou Q. Infliximab crystal structures reveal insights into self-association. *mAbs.* 2017;9:874–83. doi:10.1080/19420862.2017.1320463. PMID:28421849.
44. Kowalski JM, Parekh RN, Mao J, Wittrup KD. Protein folding stability can determine the efficiency of escape from endoplasmic reticulum quality control. *J Biol Chem.* 1998;273:19453–8. doi:10.1074/jbc.273.31.19453. PMID:9677365.
45. Kowalski JM, Parekh RN, Wittrup KD. Secretion efficiency in *Saccharomyces cerevisiae* of bovine pancreatic trypsin inhibitor mutants lacking disulfide bonds is correlated with thermodynamic stability. *Biochemistry.* 1998;37:1264–73. doi:10.1021/bi9722397. PMID:9477952.
46. Zhang Y, Werling U, Edelmann W. SLICE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res.* 2012;40:e55. doi:10.1093/nar/gkr1288. PMID:22241772.
47. Suga M, Isobe M, Hatakeyama T. Cryopreservation of competent intact yeast cells for efficient electroporation. *Yeast.* 2000;16:889–96. doi:10.1002/1097-0061(200007)16:10%3c889::AID-YEA582%3e3.0.CO;2-R. PMID:10870100.
48. Subedi GP, Johnson RW, Moniz HA, Moremen KW, Barb A. High yield expression of recombinant human proteins with the transient transfection of HEK293 cells in suspension. *J Vis Exp.* 2015;106:e53568. PMID:26779721.
49. Niesen FH, Berglund H, Vedadi M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc.* 2007;2:2212–21. doi:10.1038/nprot.2007.321. PMID:17853878.

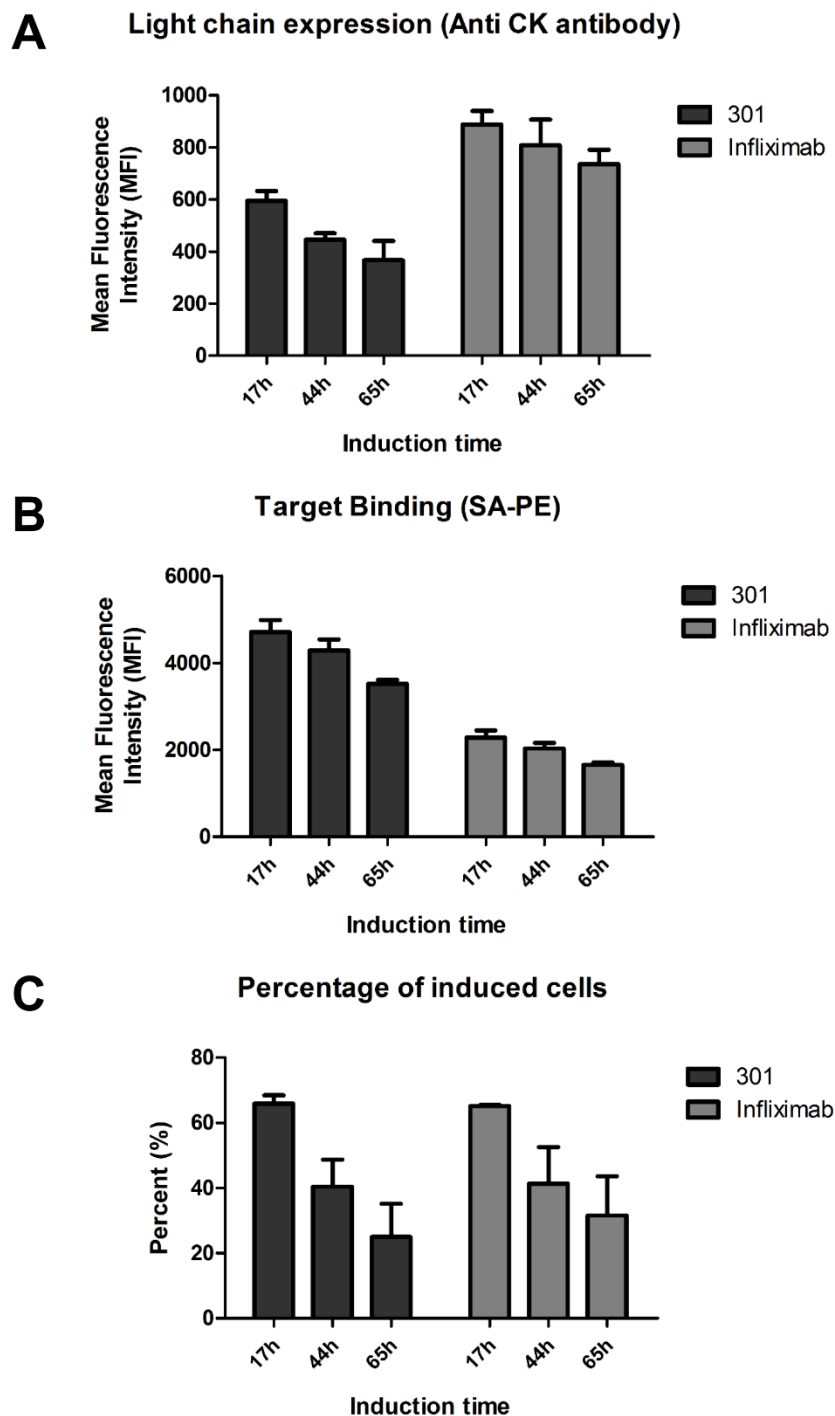




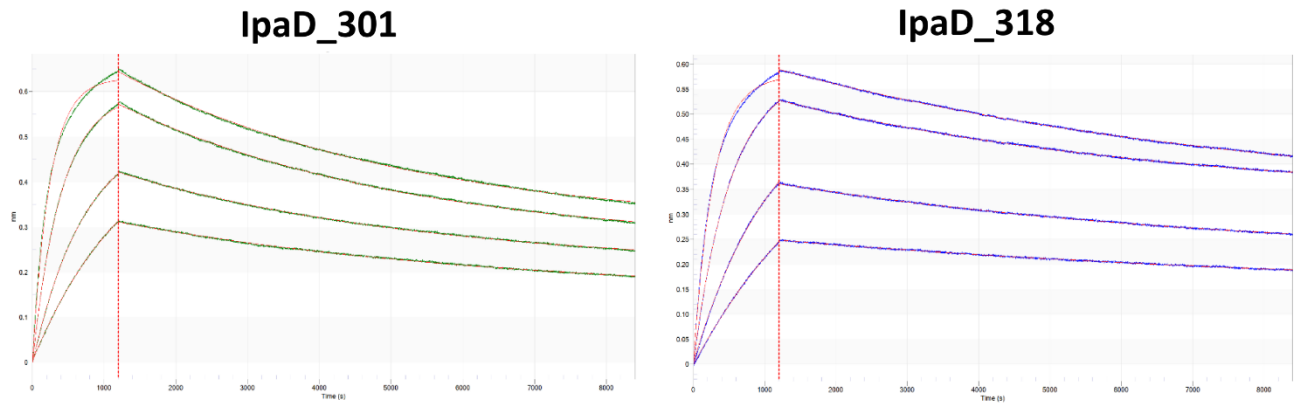
**Supplementary figure 1. TNF $\alpha$  and IpaD binding of the 6 selected antibodies expressed as Fab on the surface of yeast cells.** Cells are labeled with biotinylated antigen then with Streptavidin-PE (Target Binding) before FACS analysis. Selected antigen concentrations are far above the expected  $K_D$  to reach a quantitative binding with the antibody fragments. Respective antigen concentration were 20 nM for TNF $\alpha$  and 50 nM for IpaD. For each construct, a FACS histogram shows the fluorescence intensity corresponding to antigen binding (x-axis) vs the number of detected events (y-axis). Interval gates (bars) are shown to differentiate cells with detectable antigen binding (in green) from cells with no detectable antigen binding (in grey).



**Supplementary figure 2. Bivariate flow-cytometric analysis of yeast cells expressing various antibody fragments on their surface.** Each single dot on the plot designate the APC (y-axis) and PE (x-axis) fluorescence intensity values for a single yeast cell. Cells are double-labeled with biotinylated antigen/Streptavidin-PE (Target Binding) and anti-HA APC labels (Fab expression). Antigen concentration in labelling experiment were respectively 20 nM for anti-TNF antibodies and 50 nM for anti-IpaD antibodies. Functional subpopulations in green were gated using an inverted gate of the double negative subpopulations. The percentage is the subpopulation of cell expressing Fab.



**Supplementary Figure 3. Antigen binding and expression levels of antibodies IpaD\_301 and Infliximab expressed as Fab antibody fragments on the surface of yeast cells.** Cells were labeled with the appropriate expression reporter and the mean fluorescence intensity of positive cells was determined by FACS analysis. For antigen binding, cells were double-labeled with biotinylated antigen and Streptavidin-PE (target binding) before FACS analysis. (A) Light chain expression evaluated with anti-CK antibody reporter (B) Biotinylated antigen binding evaluated with Streptavidin-PE (C) Percentage of cells with detectable antigen binding. Measurements were made in triplicate using cells from independent cultures.



	KD (M)	kon(1/Ms)	koff(1/s)
Fab IpaD_301	1.05E-09	2.16E+05	2.25E-04
Fab IpaD_318	9.25E-10	1.59E+05	1.47E-04

**Supplementary figure 4. Bio-layer interferometry (Octet Red)  $K_D$  determination of IpaD\_301 Fab and IpaD\_318 Fab.** Fab binding to recombinant IpaD was assayed by BLI. Biotinylated IpaD was loaded on Streptavidin biosensors and equilibrated in PBSF buffer (1X PBS pH 7.4, 0.01% (w/v) BSA) before being incubated with varying concentrations of each Fab for 20 min. Dissociation was performed in PBSF buffer to for 2 hours. For each Fab, real time binding profiles using multiple protein concentrations and fitted curves (1:1 Langmuir binding model) are displayed.

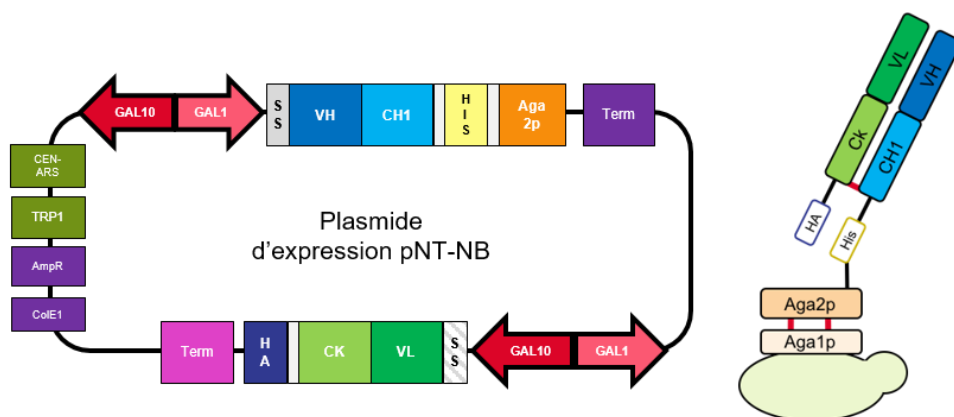
## II. Résultats additionnels et conclusion

A l'issue de ce travail, c'est le format Fab qui a été choisi. Cependant, si le plasmide pNT-Fab décrit dans l'article fonctionne très bien avec des séquences clonales, il présente l'inconvénient de contenir de nombreuses répétitions de séquence : promoteurs, peptide signaux et terminateurs. Nous avons observé que le clonage n'était pas efficace pour la génération de banques dans ce plasmide en recombinaison homologue chez la levure. Il est probable que la présence de ces cassettes répétées interfère avec les mécanismes de recombinaison homologue. De ce fait, un nouveau plasmide d'expression a été conçu pour supprimer les séquences répétées. Ce plasmide est présenté en **Figure 1**. Trois modifications ont été effectuées par rapport au plasmide pNT-Fab original pour supprimer les séquences répétées :

- Inversion du promoteur bidirectionnel de la chaîne légère
- Modification de la séquence nucléotidique du peptide signal de la chaîne légère
- Changement du terminateur de la chaîne lourde

Le nouveau plasmide « pNT-NB » s'est révélé très efficace pour le clonage par recombinaison homologue et permet par ailleurs une expression et une fonctionnalité équivalentes des Fab affichés à la surface de la levure par rapport au plasmide précédent.

Parmi les deux anticorps testés, IpaD-318 paraît être le meilleur candidat à la fois dans les tests de protection *in vivo* et c'est également celui qui a les meilleures propriétés d'expression en YSD au format Fab. C'est donc l'anticorps IpaD-318 qui a été choisi pour être humanisé et dé-immunisé dans la suite de ce travail de thèse.



**Figure 1 : Carte du plasmide pNT-NB.** Ce plasmide est une modification du plasmide pNT-Bicistronique. Pour éviter toute recombinaison du plasmide sur lui-même, le promoteur bidirectionnel de la chaîne légère a été inversé et la séquence nucléotidique du peptide signal de la chaîne légère ainsi que le terminateur ont été modifiés.



## Partie 3 : Ingénierie de l'anticorps 318 : humanisation et dé-immunisation

L'objectif principal de la thèse est de mettre en place des outils ainsi qu'une méthode de dé-immunisation des anticorps à haut débit en prenant comme exemple l'anticorps murin IpaD-318. L'évaluation de l'immunogénicité potentielle de l'anticorps est réalisée grâce aux algorithmes de prédiction *in silico* de la liaison aux molécules HLA-II. L'évaluation à haut débit de l'immunogénicité de nombreuses séquences n'est en effet possible qu'avec des outils *in silico*, les autres méthodes basées sur des tests *in vitro* n'ayant pas le débit requis.

La dé-immunisation de l'anticorps nécessite de modifier sa séquence en acides aminés au sein de régions clés, notamment les CDR, et peut donc avoir un lourd impact sur sa fonctionnalité. Afin de contrôler la fonctionnalité et de sélectionner les mutants actifs de l'anticorps tout au long du processus de dé-immunisation, le *Yeast Surface Display* a été utilisé.

Préalablement à la dé-immunisation de l'anticorps murin IpaD-318 nous avons procédé à sa chimérisation. L'anticorps IpaD-318 chimérique a été exprimé de façon fonctionnelle en YSD au format Fab ainsi qu'au format Fab soluble dans la partie précédente. Ainsi, le travail de dé-immunisation a été focalisé exclusivement sur les parties variables murines de l'anticorps IpaD-318.

### I. Mise en place de Dezimulab : une plateforme de dé-immunisation *in silico*

Il existe différents algorithmes de prédiction de la liaison aux molécules HLA-II et ceux-ci ont été décrits en introduction. Parmi ces algorithmes, NetMHCIIpan a été choisi car nous avons vu en introduction qu'il semblait être le plus performant. Cet algorithme peut être utilisé par le biais de serveurs publics comme l'IEDB. NetMHCIIpan fournit une évaluation de l'affinité des séquences peptidiques pour les différents allèles de molécules HLA-II sélectionnés, ainsi qu'une proposition de leur mode d'interaction. En revanche, il n'a pas été conçu pour comparer l'immunogénicité des séquences entre elles mais plutôt pour évaluer une séquence en particulier. Ainsi, pour la comparaison de nombreuses séquences, les résultats peuvent être difficiles à interpréter et ne sont pas mis en forme de manière exploitable. De plus, la mise en place d'une méthode de dé-immunisation à haut débit implique de devoir prédire la liaison aux molécules HLA-II pour des milliers de séquences à la fois, ce que ne permettent pas les serveurs publics.

Une plateforme permettant d'utiliser de façon pratique et rapide l'algorithme a été mise en place au CEA avec la collaboration de Youen Lespare, Aurore Lomet et Lorène Allano. Cet outil appelé « Dezimulab », permet de réaliser trois opérations principales :



- La mono-prédiction : ce module permet de prédire les cœurs de liaison aux allèles de molécules HLA-II sélectionnés ainsi que les scores de liaison associés à ces cœurs pour une séquence d'acides aminés.
- La multi-prédiction : ce module effectue la prédiction sur une liste de séquences et les classe en fonction du paramètre choisi.
- La dé-immunisation : ce module effectue des mutations sur la séquence d'intérêt et sélectionne les meilleures séquences dé-immunisées selon un critère que l'on choisit. Il fonctionne grâce à un algorithme génétique qui permet de trouver les meilleures combinaisons de mutations sans devoir les énumérer une à une. Il est possible de contraindre l'algorithme en lui fournissant une liste des mutations qu'il peut ou ne peut pas réaliser.

Pour comparer plusieurs séquences entre elles, différents critères de scores peuvent être retenus comme le minimum, la moyenne ou la médiane des *rank percentiles*. De la même façon, différents jeux d'allèles de molécules HLA-II peuvent être utilisés.

Le jeu utilisé dans ce travail contient 25 allèles et est dérivé du jeu de 27 allèles décrits en introduction contenant une majorité d'allèles HLA-DR ainsi que les principaux allèles HLA-DP/DQ. La comparaison de l'immunogénicité de deux séquences proches s'est révélée assez complexe. Nous avons d'abord défini la notion de seuil de prédiction. NetMHCIIpan donne une valeur de *rank percentile* permettant de définir si le peptide fait partie des peptides les plus affins (*rank percentile* faible) ou les moins affins (*rank percentile* élevé) pour un allèle donné. Ceci permet notamment de comparer les allèles entre eux. Nous avons défini un score d'immunogénicité permettant de tenir compte de l'affinité des peptides pour les molécules HLA-II mais aussi de la diversité des allèles pour lesquels une liaison est prédite. Ce score d'immunogénicité correspond au nombre d'épitopes prédits pour une séquence d'acides aminés avec un jeu d'allèles et un seuil de prédiction (*rank percentile*) donné.

Par exemple si une séquence possède un seul cœur de liaison prédit et que le *rank percentile* de ce peptide est inférieur à 10% pour 15 des 25 allèles testés, son score d'immunogénicité en utilisant un seuil de 10% sera de 15. De la même façon, si une séquence présente 5 cœurs de liaison avec un *rank percentile* inférieur à 10% pour un seul des 25 allèles, son score d'immunogénicité à 10% sera de 5. A l'inverse, si les 5 cœurs sont prédits avec un *rank percentile* inférieur à 10% pour chacun des 25 allèles testés, le score d'immunogénicité de la séquence à 10% sera de 125. L'évaluation du score d'immunogénicité à différents seuils permet d'évaluer la force de l'interaction prédite. Dans ce travail, le score d'immunogénicité sera évalué à 5%, 10%, 15%, 20%, 25% et 30%. L'objectif étant de réduire au maximum ce score pour les seuils les plus faibles.

L'outil intègre également une reconnaissance des cœurs de liaison humains. Pour cela Dezimulab dispose d'une liste de séquences germinales d'anticorps humains. Cette liste se compose de l'ensemble des allèles 01 (allèle le plus fréquent) de chaque gène codant pour les fragments V et J des chaînes légères et lourdes. Si un cœur de liaison est prédit comme se liant à un allèle de molécule HLA-II mais que celui-ci est retrouvé dans les séquences germinales d'anticorps humain, il ne sera pas considéré comme potentiellement immunogène.

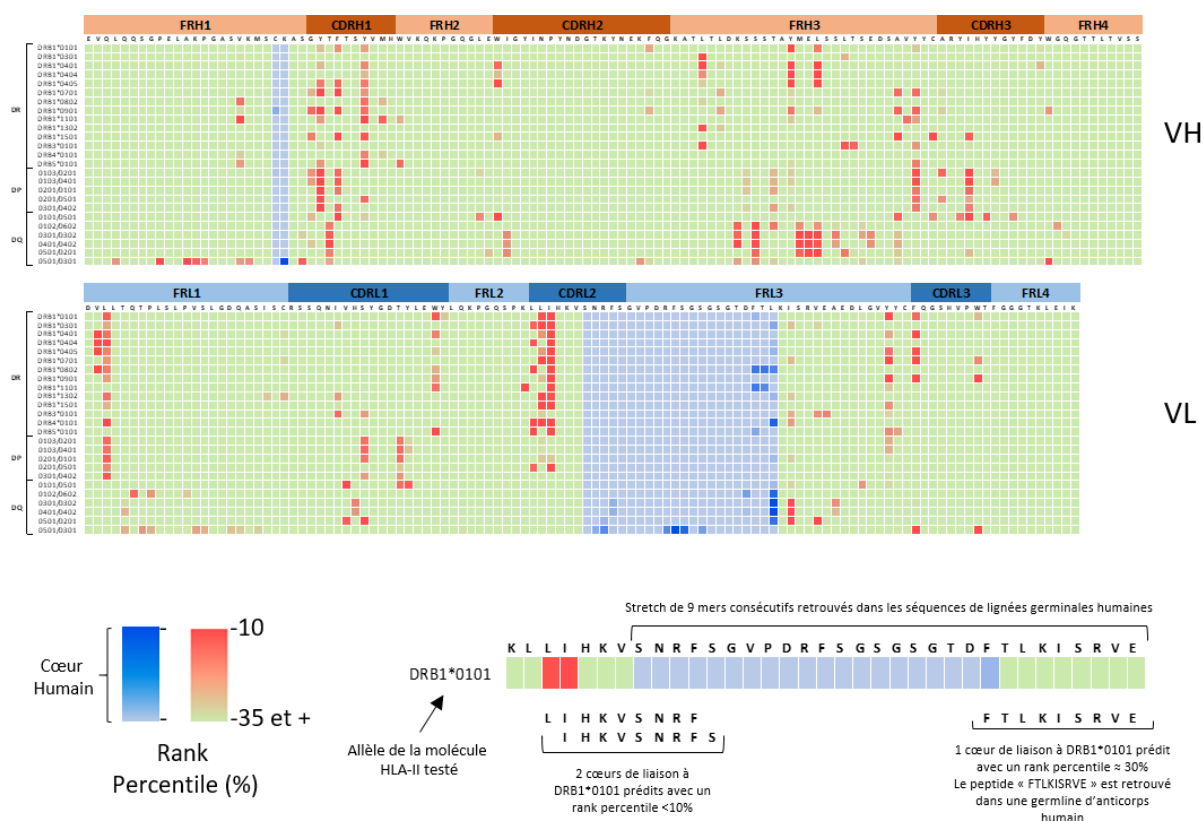
## II. Evaluation *in silico* de l'immunogénicité des parties variables de l'anticorps IpaD-318

Les séquences en acides aminés des chaînes variables lourdes et légères de l'anticorps IpaD-318 a été analysée par l'algorithme NetMHCIIpan au sein de l'outil Dezimulab. La cartographie des cœurs de liaisons prédits pour les 25 molécules HLA-II sélectionnées est présentée en **Figure 1**. Les peptides prédits comme étant de possibles épitopes T sont indiqués en dégradé de couleur rouge (rouge foncé pour les peptides fortement prédits et clair pour les peptides prédits moins fortement). Les peptides prédits pour se lier aux molécules HLA mais ayant une séquence « humaine », c'est-à-dire présente dans au moins une séquence germinale d'anticorps humain, sont indiqués en bleu foncé (les peptides humains non prédits pour se lier aux molécules HLA sont indiqués en bleu clair).

On peut observer que de nombreux cœurs avec un *rank percentile* inférieur à 35% sont prédits sur la chaîne lourde et la chaîne légère et ce pour de nombreux allèles de molécules HLA-II. Ceux-ci sont répartis au sein de la séquence et ne forment pas une zone clairement définie. Ainsi, on observe des zones potentiellement immunogènes au sein des CDR (CDRH1, CDRH3, CDRL1 et CDRL2 principalement) mais également au sein des régions *Frameworks* (en particulier FRH3, FRL1 et FRL3).

Les zones ainsi identifiées sont susceptibles d'être mutées pour réduire leur immunogénicité potentielle. Du fait de leur localisation dans les parties variables de l'anticorps IpaD-318, il est probable que certaines mutations ne soient pas tolérées d'un point de vue fonctionnel.

Nous avons alors cherché à identifier les positions permissives au sein de la séquence ainsi que les acides aminés directement impliqués dans la liaison à l'antigène. Pour répondre à cette problématique, une approche de *Deep Mutational Scanning* (DMS) a été entreprise. La connaissance de la permissivité des positions au sein de la séquence de l'anticorps permet de guider la conception des banques de dé-immunisation pour obtenir des variants fonctionnels.



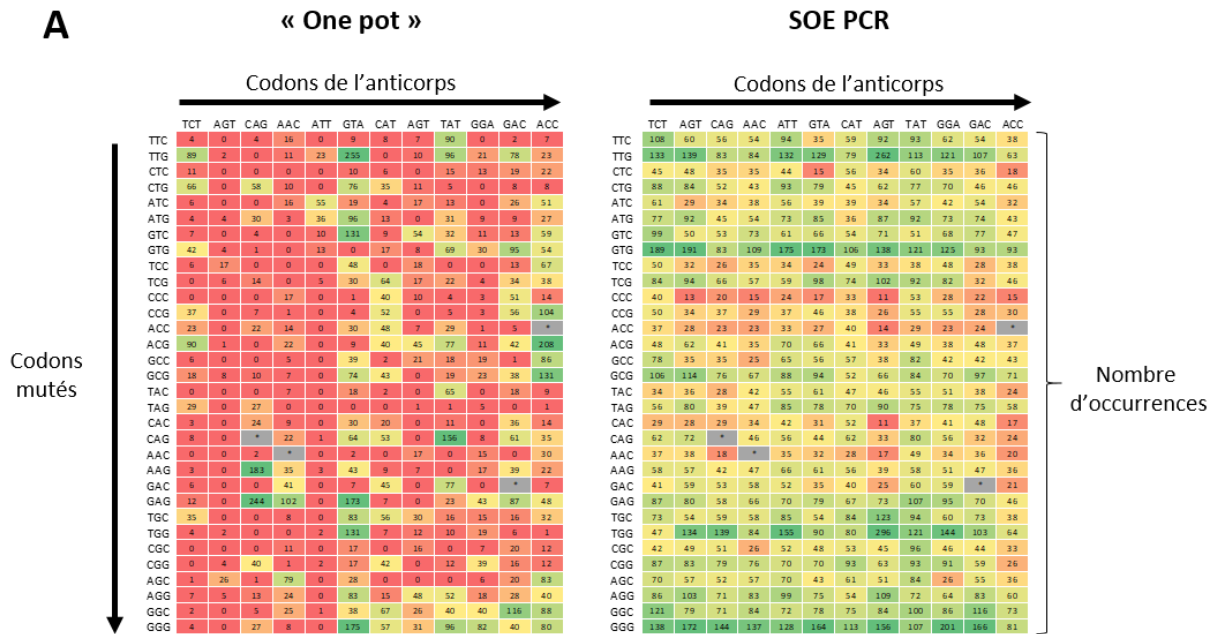
**Figure 1. Prédiction des peptides de liaison aux molécules HLA-II sur les parties variables de l'anticorps murin IpaD-318.** La prédiction a été réalisée à l'aide de l'algorithme NetMHCIIpan 3.1 et mise en forme par le logiciel Dezimulab. Sur la *heatmap*, le score de chaque cœur de 9 acides aminés consécutifs est reporté au niveau de la P1 (premier acide aminé du cœur). L'échelle de couleur considérée est entre 10 et 35% de *rank percentile*. Au-delà de 35%, la liaison entre le cœur et la molécule HLA-II est considérée comme trop faible pour pouvoir déclencher une réponse immunogène. Les séquences humaines sont également considérées comme non immunogènes (résultats de prédiction colorés en bleu). Les prédictions sont réalisées sur le jeu de 25 allèles de molécules HLA-II qui dérive du set de 27 allèles de Greenbaum *et al* présenté en introduction. Les prédictions sur la chaîne lourde sont présentées sur la première *heatmap*, les prédictions sur la chaîne légère sont présentées sur la seconde *heatmap*. Un exemple d'interprétation de cette représentation est présenté en bas de la figure.

### III. Mise en place du DMS et application à l'anticorps 318

La réalisation du DMS sur les parties variables de l'anticorps IpaD-318 implique de générer chacune des 20 substitutions possibles pour chacun des résidus de la séquence étudiée.

Plusieurs méthodes de biologie moléculaire existent pour générer une banque comprenant l'intégralité des mutants. Nous avons dans un premier temps choisi d'utiliser la méthode utilisée dans la première partie des résultats, le « Plasmid-based one pot saturation mutagenesis » décrit par Wrenbeck *et al* <sup>250</sup>. Cette méthode consiste à générer un ADN simple brin à partir d'un plasmide cible et de synthétiser le brin manquant à l'aide d'amorces dégénérées (une amorce par position étudiée). Les plasmides *wild-type* sont spécifiquement digérés en fin de mutagenèse pour ne conserver théoriquement que les plasmides néoformés (voir la section matériel et méthodes).

Nous avons appliqué ce protocole pour générer deux banques comprenant respectivement toutes les mutations possibles pour la chaîne variable lourde ou pour la chaîne légère de l'anticorps IpaD-318 dans le plasmide pNT-NB. Nous avons alors séquencé la banque de plasmides (**Figure 2**). On peut observer qu'une faible proportion de clones contenait les mutations prévues, et que près de 72 à 79% des séquences n'étaient pas mutées. Par ailleurs, toutes les substitutions n'étaient pas présentes, les banques couvrant seulement 53% et 63% de la diversité prévue. Ces résultats sont meilleurs que ceux obtenus dans le cadre de la mutagenèse d'IpaD dans la première partie des résultats. Le DMS sur l'antigène IpaD a été réalisé sur des zones plus courtes et en utilisant un plasmide différent. La proportion de plasmides *wild-type* à l'issue de la mutagenèse s'élevait seulement à 20%. De plus, dans le cas de l'*epitope mapping* de l'anticorps IpaD-318 la représentation exhaustive de toutes les mutations attendues n'était pas nécessaire.



**B**

	"One Pot"		SOE PCR			
	VH318	VL318	VH318 - 1	VH318 - 2	VL318 - 1	VL318 - 2
Diversité <sup>a</sup>	3808	3584	1952	1856	1856	1728
Nombre reads <sup>b</sup>	275 194	321 895	979 579	1 013 843	318 258	342 141
% wild-type <sup>c</sup>	79%	72%	19%	17%	13%	14%
% mono-mutants <sup>d</sup>	19%	24%	59%	63%	69%	69%
% de couverture <sup>e</sup>	53%	63%	100%	100%	100%	99%

- a : nombre de mutants différents attendus en ADN. Nombre de résidus x 32 codons  
b : nombre de reads ayant passé les filtres de qualité (Q>30)  
c : proportion de reads non mutés  
d : proportion de reads mutés sur un seul codon  
e : nombre de mutations attendues constatées / diversité

**Figure 2. Comparaison des méthodes de génération du DMS. (A)** Nombre de mutations par position pour chacune des deux méthodes de DMS utilisées. \* : codon *wild-type*. **(B)** Résumé du traitement des séquençages NGS.

Pour améliorer la qualité des banques, nous avons mis au point une nouvelle méthode de génération des banques de DMS basée sur le principe de la SOE PCR (*Splicing by Overlap Extension PCR*). Cette méthode permet d'introduire des mutations à un emplacement spécifique de la séquence par PCR et de reconstituer le gène complet muté lors d'une PCR d'assemblage (**Figure 2B** et section matériel et méthodes). Par ailleurs pour optimiser la qualité de nos données de séquençage Illumina MiSeq et permettre deux lectures par brin (*paired end sequencing*), nous avons choisi de scinder en deux les banques. Ainsi quatre banques (2 par chaîne variable) ont été générées en utilisant à nouveau le plasmide pNT-NB, et la qualité de la banque a été évaluée par séquençage NGS.

Les banques générées par cette méthode ne comportent qu'une faible proportion de séquences non mutées: entre 13% et 19% (**Figure 2B**). Par ailleurs, la totalité des mutations attendues ont été observées pour trois des quatre banques et près de 99% pour la dernière.

L'analyse de la fréquence des mutations au sein des banques générées par ces deux méthodes révèle que la seconde méthode produit une banque bien plus homogène que la première (**Figure 2A**). Pour cette raison, la méthode SOE-PCR a été retenue pour la génération du DMS sur les parties variables de l'anticorps.

Les quatre banques de DMS ont été transformées dans *S. cerevisiae* puis induites pour exprimer les mutants à la surface des levures. La stratégie de tri est présentée en **Figure 3** avec la banque comprenant la première partie de VH. Nous avons d'abord choisi de sélectionner les mutations sur l'anticorps qui n'affectent ni la liaison avec l'antigène ni l'appariement des chaînes lourde et légère. Pour cela, nous avons utilisé la fenêtre de tri de couleur verte, également appelée « tri positif ». A l'inverse, le « tri négatif » correspond à la fenêtre de tri de couleur rouge et permet de sélectionner les mutations n'affectant pas l'appariement des deux chaînes mais qui abolissent la liaison avec l'antigène.

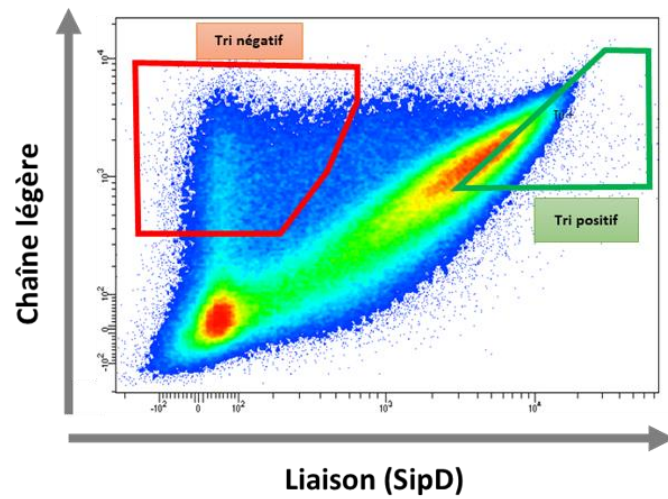
Pour chaque banque, les levures non triées, triées positivement et négativement ont été séquencées par NGS. Pour chaque position, l'enrichissement de chaque mutation dans chacun des tris par rapport à la banque non triée a été calculé. Les valeurs obtenues ont permis d'établir un tableau aussi appelé « *heatmap* » permettant respectivement d'identifier les résidus très importants pour l'interaction avec l'antigène (avec les données issues des tris négatifs et un code couleur en dégradé de rouges) et les résidus pouvant être substitués sans perte d'activité (avec les données issues des tris positifs et un code couleur en dégradé de bleus). Les *heatmaps* du DMS de la chaîne lourde sont présentées en **Figure 4** et celles de chaîne légère en **Figure 5**.

Les *heatmaps* des tris négatifs montrent très clairement les positions importantes pour la fonctionnalité de l'anticorps : ces positions sont matérialisées par une colonne rouge qui indique que la position n'est pas permissive pour la mutation. Dans la plupart des cas ces résidus sont situés au niveau des CDR, ce qui corrèle bien avec le fait que les CDR sont censés porter le

paratope. Le CDRL2 semble être le CDR le moins important pour l'interaction. La position L72 du FRH3 de la chaîne lourde et la position G69 du FRL3 de la chaîne légère semblent également jouer un rôle important pour la reconnaissance de l'antigène. Les *heatmaps* des tris positifs montrent la permissivité de tous les résidus de la chaîne lourde et de la chaîne légère. On peut d'abord constater que des résidus permissifs sont répartis tout au long des séquences des deux chaînes. Les zones correspondant aux CDR semblent moins permissives car on constate des « zones blanches » à leur niveau, indiquant que ces résidus ne peuvent pas être mutés. Cependant, même au sein des CDR à l'exception du CDRH3, il existe systématiquement des positions permissives sur lesquelles il sera possible d'agir dans le cadre de banques de dé-immunisation. Les résidus du CDRH3 semblent très peu permissifs à la mutation (zone blanche) et il est à priori difficile de modifier des acides aminés dans cette zone.

Si l'on compare les *heatmaps* des tris positifs et négatifs, on peut constater qu'elles sont complémentaires et ne se contredisent pas. On peut noter que certaines mutations ne sont enrichies dans aucun des tris. Celles-ci correspondent vraisemblablement à des mutations qui affectent la structure de l'anticorps et qui empêchent l'expression ou l'appariement des deux chaînes.

Ces cartes de DMS sont utiles pour comprendre la contribution des acides aminés de l'anticorps à la reconnaissance de l'antigène. Elles permettent également d'identifier quelles substitutions sont tolérées et permettent de maintenir la fonctionnalité de l'anticorps. Dans la suite du projet, elles ont été utilisées pour guider la conception de banques de variants moins immunogènes de l'anticorps IpaD-318.



**Figure 3. Définition des fenêtres de tri du DMS.** Deux DMS ont été générés par chaîne. Le tri a été effectué de la même façon pour les 4 DMS. Un exemple de définition des fenêtres de tri sur la première partie du DMS de la chaîne lourde est présenté. Deux tris ont été effectués par DMS : le tri négatif qui représente les constructions complètes (présentant une chaîne lourde et une chaîne légère) à la surface de la levure mais qui ont perdu la liaison pour l'antigène. Le tri positif cible les levures exposant une construction complète avec une affinité conservée ou améliorée pour l'antigène. Les tris ont été effectués avec l'antigène SipD à une concentration de 2 nM.



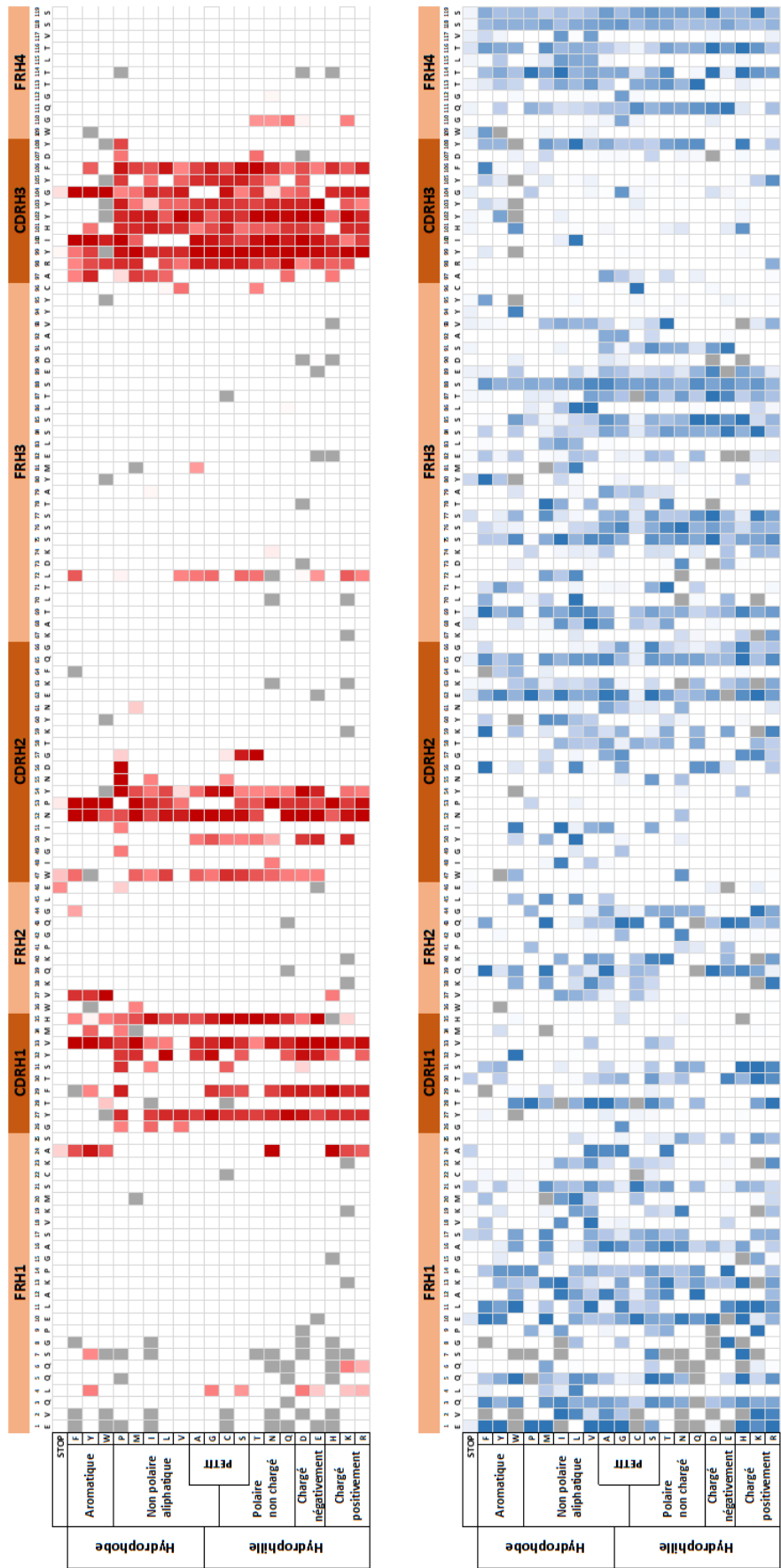
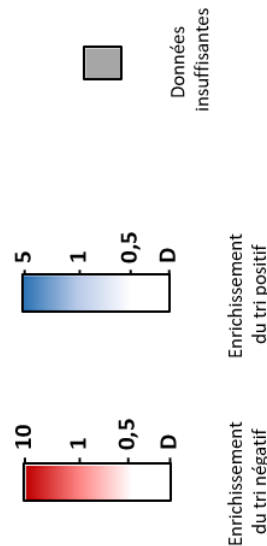
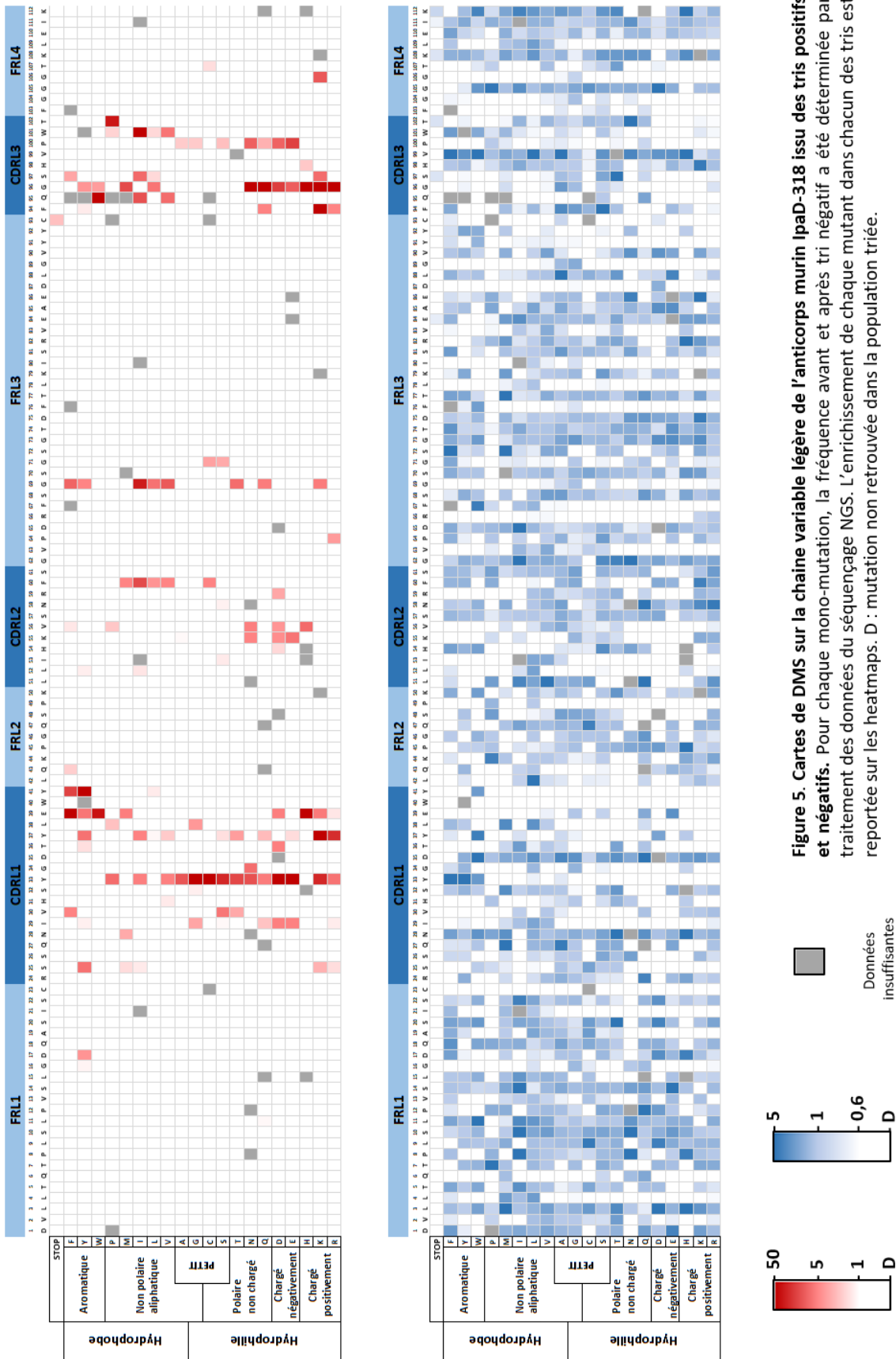
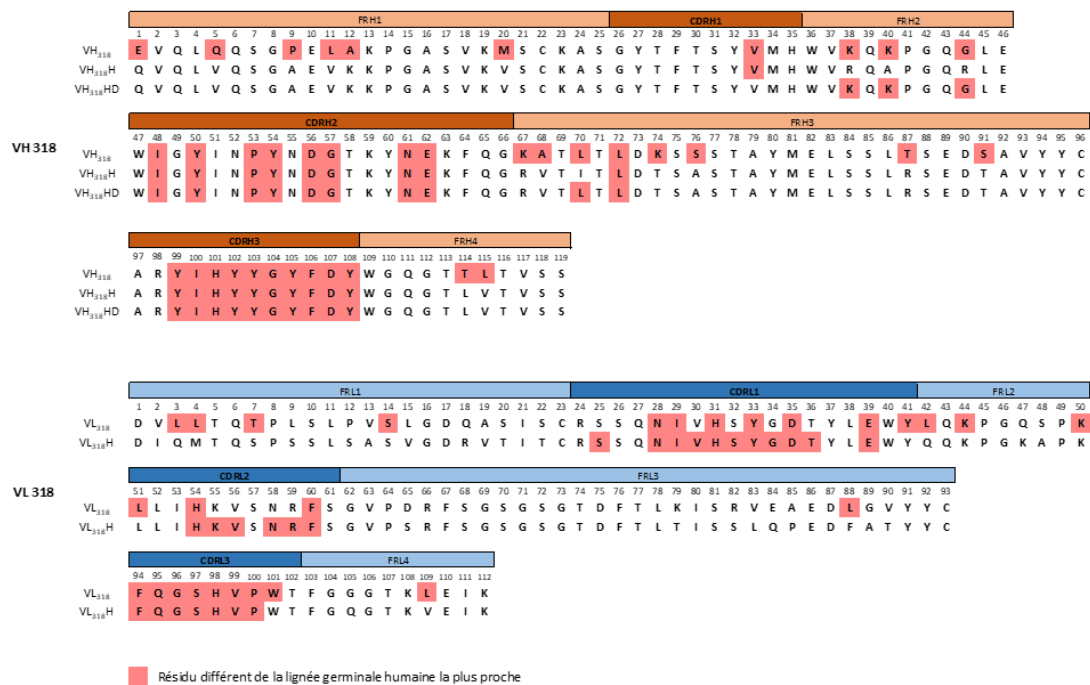


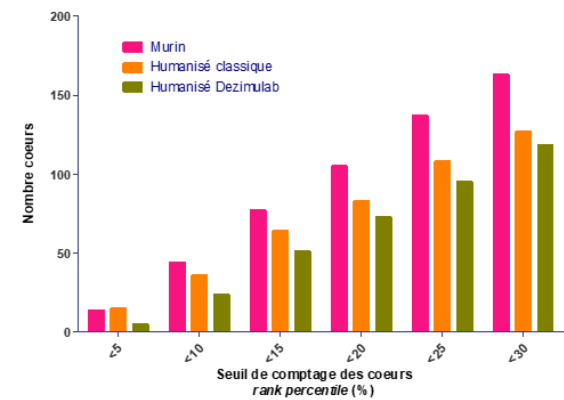
Figure 4. Cartes de DMS sur la chaîne variable lourde de l'anticorps murin IpaD-318 issu des tris positifs et négatifs. Pour chaque mono-mutation, la fréquence avant et après tri négatif a été déterminée par traitement des données du séquençage NGS. L'enrichissement de chaque mutant dans chacun des tris est reportée sur les heatmaps. D : mutation non retrouvée dans la population triée.







**Figure 6. Séquences des parties variables de l'anticorps IpAD-318 murin et humanisé.** VH<sub>318</sub> : chaîne variable lourde de l'anticorps murin IpAD-318. VH<sub>318</sub>H : chaîne variable lourde humanisée de l'anticorps IpAD-318. VH<sub>318</sub>HD : chaîne variable lourde « humanisée Dezimulab » de l'anticorps IpAD-318. VL<sub>318</sub> : chaîne variable légère de l'anticorps murin IpAD-318. VL<sub>318</sub>H : chaîne variable légère humanisée de l'anticorps IpAD-318.



**Figure 7. Bilan de la prédiction du nombre de cœurs de liaison aux molécules HLA-II.** Prédiction du nombre de cœurs de liaison aux molécules HLA-II.

	SipD			IpaD		
	K <sub>D</sub>	k <sub>on</sub>	k <sub>off</sub>	K <sub>D</sub>	k <sub>on</sub>	k <sub>off</sub>
Chimérique	450 pM	4,8.10 <sup>5</sup>	2,2.10 <sup>-4</sup>	150 pM	4,9.10 <sup>5</sup>	7,3.10 <sup>-5</sup>
Humanisé	ND	ND	1,7.10 <sup>-4</sup>	ND	ND	4,7.10 <sup>-5</sup>
Humanisé Dezimulab	ND	ND	1,3.10 <sup>-4</sup>	ND	ND	6,1.10 <sup>-6</sup> *

**Tableau 1. Bilan de l'humanisation de l'anticorps IpAD-318 sur le plan fonctionnel.** Evaluation de la fonctionnalité des versions humanisées au format Fab par rapport à l'anticorps chimérique au format Fab. Les mesures ont été réalisées à l'OCTET Red96. ND : pour ces constructions, seuls les k<sub>off</sub> ont été déterminés. \* : valeur indiquée par l'Ocetet, non réellement mesurable car la dissociation n'est pas visible dans l'échelle de temps considérée.

## IV. Humanisation de l'anticorps 318

L'humanisation a été identifiée comme une solution pour réduire le nombre de zones problématiques à dé-immuniser. Ainsi, l'humanisation permettrait d'exclure les zones prédites comme étant immunogènes situées dans les *frameworks*.

La technique qui a été choisie pour humaniser l'anticorps est la stratégie classique du *CDR grafting*. Pour cela trois paramètres doivent être déterminés : les bornes des six CDR, la séquence de l'anticorps humain accepteur, ainsi que les éventuelles positions importantes dans les *frameworks* qui nécessiteraient de rester murines.

Les bornes des six CDR ont été déterminées grâce aux différents schémas de numérotation conventionnels : Chothia, Kabat, IMGT et Contact. Les bornes des CDR ont été choisies de façon à avoir un consensus entre les quatre schémas de numérotation de sorte que les CDR de chaque schéma y soient inclus.

Pour le choix des anticorps humains accepteurs dans lesquels sont greffés les CDR, nous avons choisi d'évaluer deux stratégies différentes. La première stratégie consiste à choisir dans la base de données IMGT la séquence germinale d'anticorps humains la plus proche pour chacune des chaînes, c'est la stratégie « *best fit* ». La seconde stratégie consiste à utiliser des séquences germinales « consensus » connues pour leurs bonnes propriétés biophysiques et leur bonne expression, décrites par Haidar *et al* <sup>188</sup>. Ainsi, nous avons choisi le gène IGHV1-3\*01 par l'approche *best-fit* et le gène IGHV3-23\*01 pour l'approche consensus pour la chaîne lourde. Pour la chaîne légère, les gènes IGKV2-30\*01 et IGKV1-39\*01 ont été choisis respectivement par les approches *best-fit* et consensus.

Lors du processus d'humanisation, il est parfois nécessaire de conserver certains résidus murins au sein des régions *frameworks*. Ainsi, à la lumière des résultats du DMS, en particulier sur les *heatmaps* issues des tris négatifs, nous avons conservé le résidu L72 du FRH3 murin. En revanche, la mutation du résidu G69 du FRL3 n'a pas été nécessaire car il s'avère être conservé entre les *frameworks* humains sélectionnés et l'anticorps chimérique IpaD-318.

Les chaînes variables humanisées ont été construites en insérant les CDR murins dont les bornes sont présentées en **Figure 6** à la place des CDR présents dans les séquences humaines choisies, tout en conservant le résidu L72 murin de la chaîne lourde.

Ces chaînes variables humanisées ont été évaluées en YSD en deux étapes. D'une part, les chaînes lourdes humanisées ont été testées en combinaison avec la chaîne légère variable murine et d'autre part les chaînes légères humanisées ont été évaluées en combinaison avec la chaîne lourde variable murine. L'expression de ces constructions à la surface des levures a montré que seule la chaîne lourde humanisée sur la base du *best-fit* était fonctionnelle avec un  $K_D$  apparent équivalent à celui de l'anticorps chimérique pour l'antigène SipD. Pour la chaîne légère c'est la

version basée sur la séquence consensus qui s'est avérée fonctionnelle au format Fab. Les deux chaînes humanisées (séquences présentées en **Figure 6**) ont ensuite été assemblées et le Fab correspondant a été produit en cellules HEK293 puis l'affinité du Fab pour IpaD et SipD a été évaluée par BLI (Bio Layer Interferometry) sur un OctetRed96 (**Figure 7B**). La version humanisée de l'anticorps semble avoir conservé son affinité pour ses deux antigènes avec des constantes de dissociation équivalentes à celles de l'anticorps murin (mesures effectuées sur Fab non purifiés, de nouvelles mesures à l'Octet vont être réalisées dans les prochaines semaines pour compléter ces résultats dans le manuscrit final).

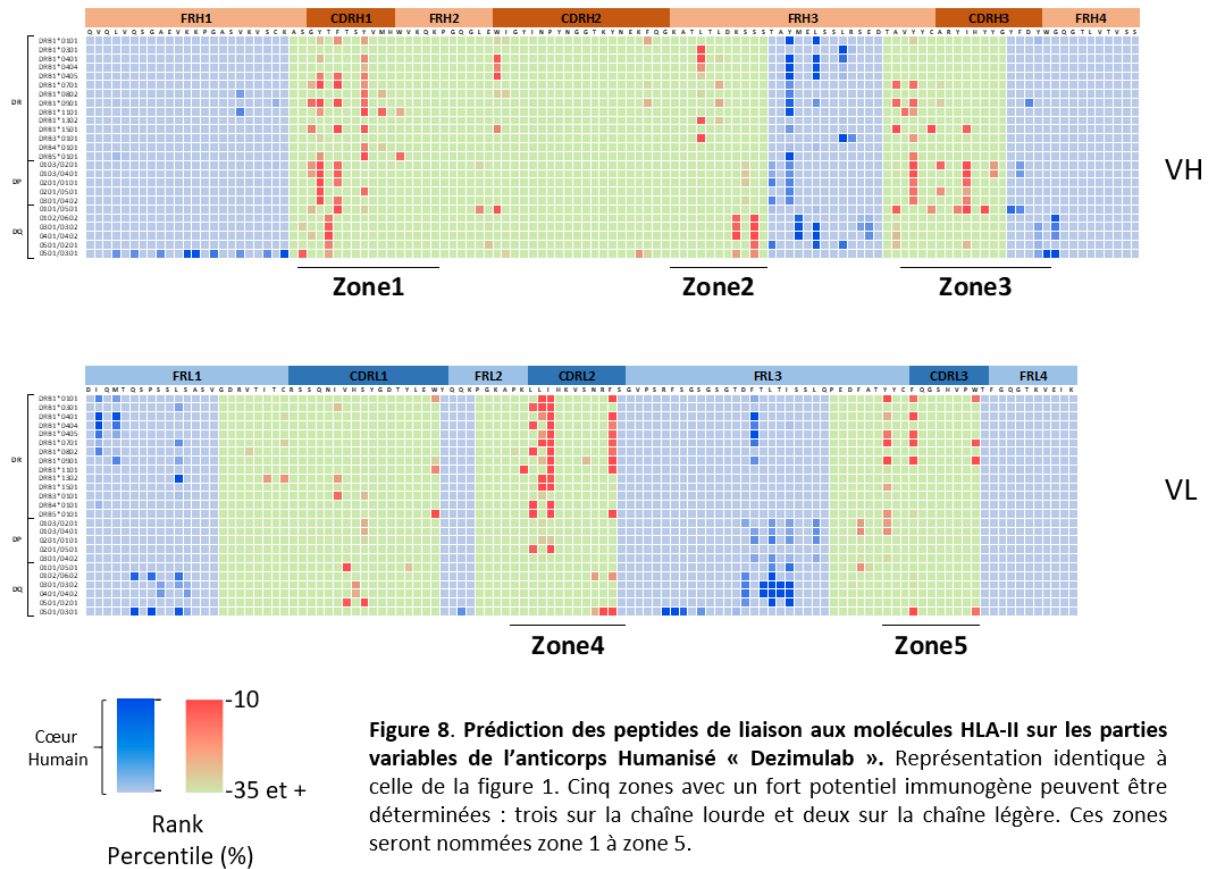
En analysant les résultats de prédiction de la liaison aux molécules HLA-II, nous nous sommes rendu compte que la prédiction pour la zone située en aval du CDRH1 était moins favorable dans la version humanisée que dans la version murine. Nous avons donc muté les positions K38, K40 et G44 en leur équivalent murin pour améliorer la prédiction. Cette version est appelée « humanisée Dezimulab » et sa séquence est présentée en **Figure 6**.

Les scores d'immunogénicité à différents seuils pour les trois versions de l'anticorps IpaD-318 générées sont présentés en **Figure 7**. Sur cette Figure, la baisse du score d'immunogénicité est visible entre l'anticorps murin et l'anticorps humanisé et cette baisse est encore plus importante avec la version « humanisée Dezimulab ». Cette version possède également une fonctionnalité équivalente à celle de l'anticorps chimérique (**tableau 1**) (mesures effectuées sur Fab non purifiés, de nouvelles mesures à l'Octet vont être réalisées dans les prochaines semaines pour compléter ces résultats dans le manuscrit final).

L'humanisation de l'anticorps IpaD-318 a permis de réduire significativement son score d'immunogénicité prédit tout en conservant sa fonctionnalité. Cependant il reste de nombreux cœurs de liaison aux molécules HLA-II encore présents au sein de la séquence de l'anticorps dans les parties de l'anticorps qui ne sont pas humaines comme les six CDR et une partie du FRH3 du fait de la *backmutation* L72. Il est donc nécessaire de réaliser des mutations au sein de ces séquences murines afin de réduire encore davantage le score d'immunogénicité de l'anticorps : c'est l'étape de dé-immunisation.

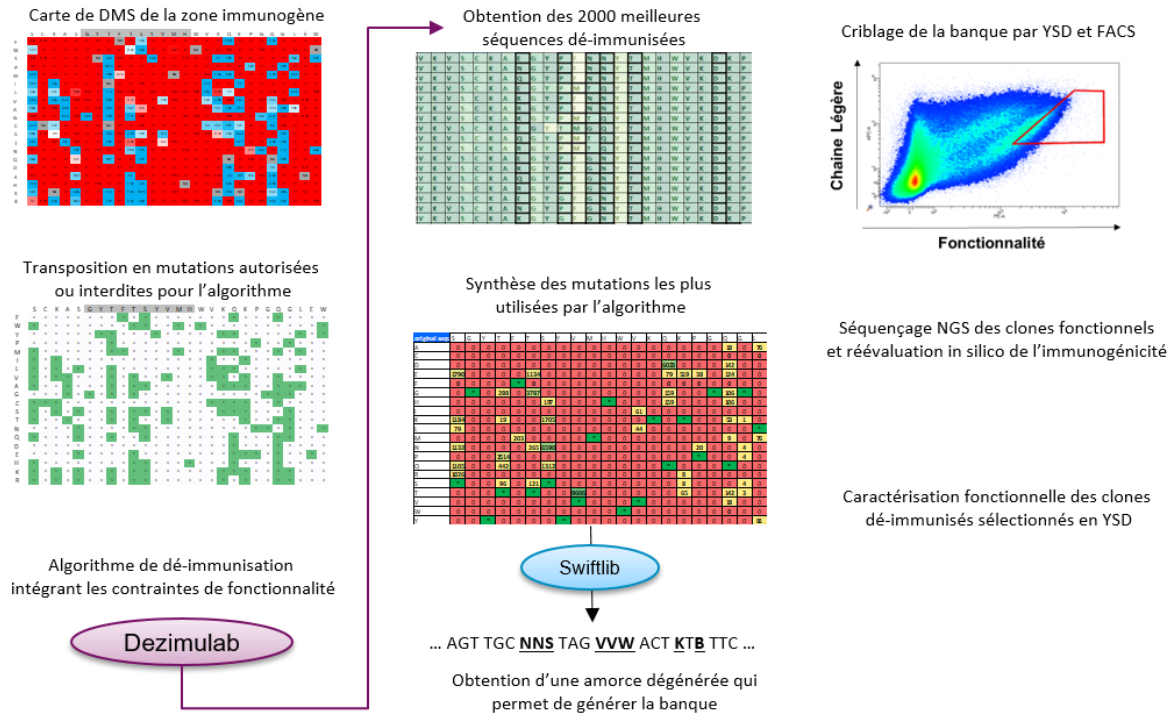
## V. Dé-immunisation

La mono-prédiction des cœurs de liaison aux molécules HLA-II sur les parties variables de l'anticorps humanisé Dezimulab est présentée en **Figure 8**. En comparant avec la prédiction de la **Figure 1**, on constate que l'humanisation a permis de supprimer des zones prédites comme étant immunogènes dans le FRH3, le FRL1 et le CDRL1. On peut également constater que les cœurs de liaison aux molécules HLA-II prédits sont regroupés en 5 zones (indiquées dans la **Figure 8**) situées au niveau du CDRH1 (Zone 1), FRH3 (Zone 2), CDRH3 (Zone 3), CDRL2 (Zone 4) et du CDRL3 (Zone 5). Afin de réduire le score d'immunogénicité de ces 5 zones, un processus de dé-immunisation combinant prédiction, mutagenèse dirigée, *Yeast Surface Display* et NGS a été mis en place et est présenté en **Figure 9**.



L'objectif de ce processus est de combiner des informations fonctionnelles apportées par les cartes fonctionnelles du DMS aux contraintes liées à la liaison des séquences aux molécules HLA-II apportées par Dezimulab. Ceci aboutit à la génération d'une banque contenant des variants potentiellement dé-immunisés et actifs.

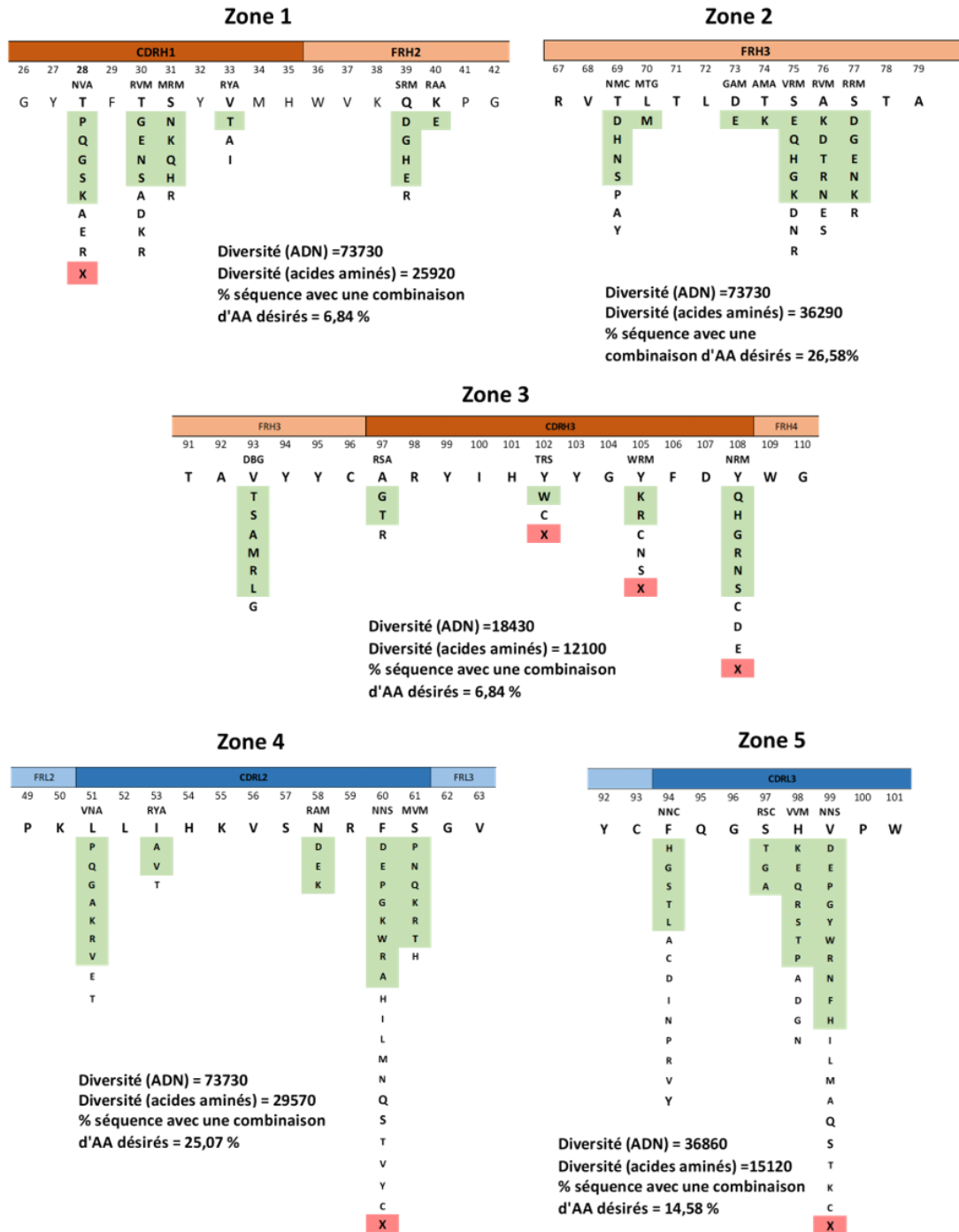




**Figure 9. Stratégie de dé-immunisation.** La carte de DMS positif est utilisée pour connaître les substitutions tolérées. Ensuite, l'algorithme de dé-immunisation cherche les 2000 meilleures combinaisons de substitutions tolérées produisant une séquence avec le moins de cœurs de liaison aux molécules HLA-II prédits. Une synthèse des mutations utilisées par l'algorithme de dé-immunisation est réalisée. Ces données sont ensuite fournies à un algorithme qui permet de générer une amorce dégénérée en fonctions des acides aminés demandés à chaque position. Cette amorce dégénérée permet de générer la banque de dé-immunisation qui est criblée par la suite en YSD pour la fonctionnalité. Les clones fonctionnels sont séquencés par séquençage à haut débit. Les séquences fonctionnelles sont réévaluées pour la prédiction du nombre de cœurs de liaison aux molécules HLA-II. Le ou le meilleurs clones sont produits et évalués pour leur fonctionnalité et/ou leur immunogénicité.

La stratégie de dé-immunisation est présentée en **Figure 9**. Les données du DMS positif sont codées sous forme de matrice de fonctionnalité et servent de point de départ pour l'algorithme pour savoir quelles sont les mutations autorisées. L'algorithme génétique de Dezimulab cherche alors à identifier les meilleures combinaisons de mutations permettant de réduire le score d'immunogénicité. La fréquence des différentes substitutions permettant de réduire ce score est alors extraite et intégrée dans une matrice qui répertorie les mutations les plus à même de réduire l'immunogénicité sans affecter la fonctionnalité.

Pour générer une banque combinant le maximum de substitutions désirées, nous avons utilisé l'outil *Swiftlib* pour identifier pour chaque position le codon dégénéré le plus approprié, tout en conservant une diversité modérée (ici environs  $10^5$ ). La banque ainsi conçue est synthétisée par SOE PCR à l'aide d'une amorce dégénérée puis transformée et exprimée en YSD pour réaliser un tri fonctionnel.

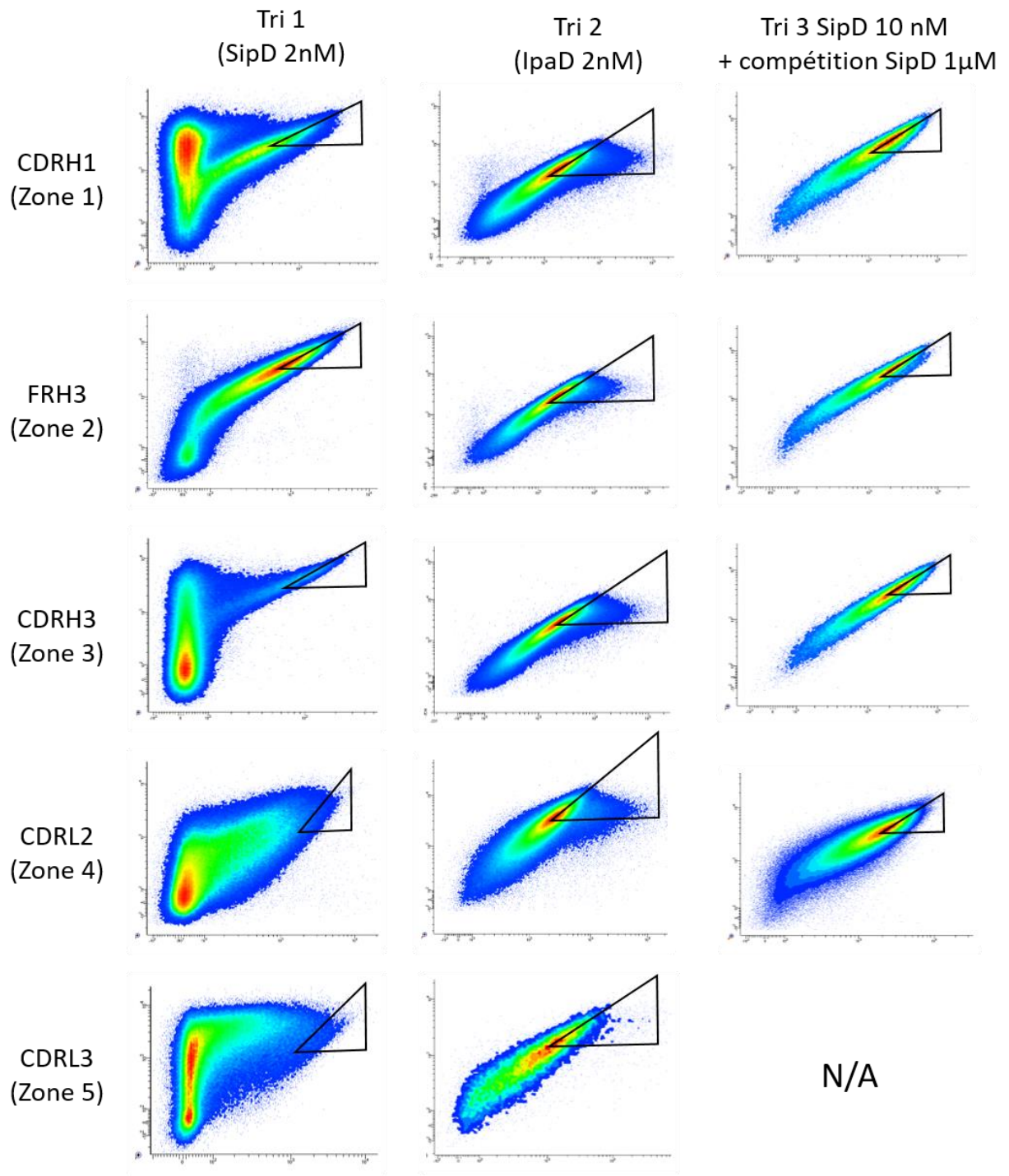


**Figure 10. Détail des banques de dé-immunisation par zone.** Les banques produites par l'algorithme *Swiftlib* sont codées par des amorces dégénérées précisées pour les cinq zones. Les acides aminés présents dans ces codons sont indiqués. Les acides aminés verts sont ceux qui sont désirés. Les « X » représentent un codon STOP. Pour chaque banque, la diversité en ADN et en acides aminés est précisée ainsi que la proportion de séquences ne contenant que des acides aminés désirés.

Le processus de conception des banques a été effectué pour les cinq zones immunogènes. Le détail de chaque banque est donné en **Figure 10**. Pour chaque zone on peut voir quels sont les acides aminés présents à chaque position dans la banque. Les acides aminés en vert correspondent aux résidus les plus utilisés par Dezimulab pour dé-immuniser la séquence. On peut constater que dans la majorité des cas ces acides aminés sont plutôt hydrophiles ou forment un coude (proline). Cela concorde avec le fait que les poches de liaison aux molécules HLA-II accueillent des acides aminés plutôt hydrophobes et n'acceptent généralement pas de prolines. Certains résidus (souvent hydrophobes) ne sont pas mutés car il s'agit de positions non permissives appartenant au paratope de l'anticorps.

La diversité en ADN des banques varie entre  $1,8.10^4$  et  $7,4.10^4$  et la proportion de séquences ne contenant que des mutations demandées par l'algorithme varie entre 6,8% et 26,5%. Cette faible proportion s'explique par les contraintes inhérentes à l'utilisation des codons dégénérés. En effet, ces derniers peuvent coder en grande partie pour des acides aminés non désirés. L'introduction de substitutions non désirées augmente la diversité des banques. Néanmoins, il peut également exister des phénomènes intéressants de compensation ou des synergies entre deux mutations. Certaines mutations se révèlent tolérées en combinaison avec d'autres mutations alors qu'elles sont délétères individuellement.

Les cinq banques générées ont alors été transformées dans *S. cerevisiae* puis induites pour exprimer les variants à la surface des levures. Pour chaque banque, trois tris en FACS ont été effectués. Les Figures de cytométrie correspondant aux différents tris pour chacune des banques sont présentées dans la **Figure 11**. Le premier tri est réalisé avec l'antigène SipD à une concentration de 2 nM, concentration proche du  $K_D$ . Le second tri est effectué en utilisant l'antigène IpaD à 2 nM pour garantir la conservation de la réactivité croisée de l'anticorps. Enfin le dernier tri est un tri cinétique qui permet de conserver les variants possédant une constante de dissociation faible. Pour ce tri, les cellules induites sont saturées avec l'antigène SipD biotinylé puis lavées. Une seconde incubation est réalisée avec l'antigène non biotinylé à très forte concentration (1  $\mu$ M) pendant 2h pour favoriser la dissociation de l'antigène biotinylé. Les fenêtres de tris sont conçues de façon à conserver les mutants à la fois les plus affins et bien exprimés. La forme diagonale de la fenêtre permet de prendre en compte l'hétérogénéité de l'expression au sein de la population de levures. Toutes les levures se trouvant sur la diagonale portent un Fab avec une affinité équivalente, le seul paramètre qui varie étant le nombre de copies de ce Fab se trouvant à la surface. Les populations affichées correspondent aux levures exprimant une chaîne lourde (marquée par un anti-polyhistidine-FITC). Ce marquage permet de ne considérer que les levures qui sont induites et qui ne présentent pas de codon STOP dans la chaîne lourde.



Expression  
 Chaîne légère  
 ↗  
 Liaison à  
 l'antigène

**Figure 11. Figures de cytométrie des différents tris de chaque banque.** Chaque banque a été triée trois fois. Le premier tri est réalisé avec 2nM de l'antigène SipD. Le deuxième tri est réalisé avec 2 nM de l'antigène IpaD. Le dernier tri est un tri cinétique. Les levures sont incubées avec l'antigène SipD biotinylé à forte concentration (10 nM) puis les cellules sont lavées et incubées avec 1  $\mu$ M de SipD froid pendant 2h à 20°C. Ce tri permet de sélectionner les clones avec les meilleures constantes de dissociation. Pour chaque graphique, l'expression de la chaîne légère est reportée en ordonnée et la fluorescence de liaison est en abscisse.

Les Figures de cytométrie du premier tri de chaque banque permettent d'évaluer l'impact moyen des mutations choisies à la fois sur la fonctionnalité mais également sur l'expression des Fab à la surface des levures. On peut observer que chacune des banques montre un comportement différent.

La banque CDRH1 présente deux populations distinctes avec une chaîne légère bien exprimée (et donc un bon appariement chaîne lourde / chaîne légère) dans les deux cas. La plupart des mutations de cette banque altèrent la fonctionnalité du Fab tout en conservant un bon appariement des deux chaînes.

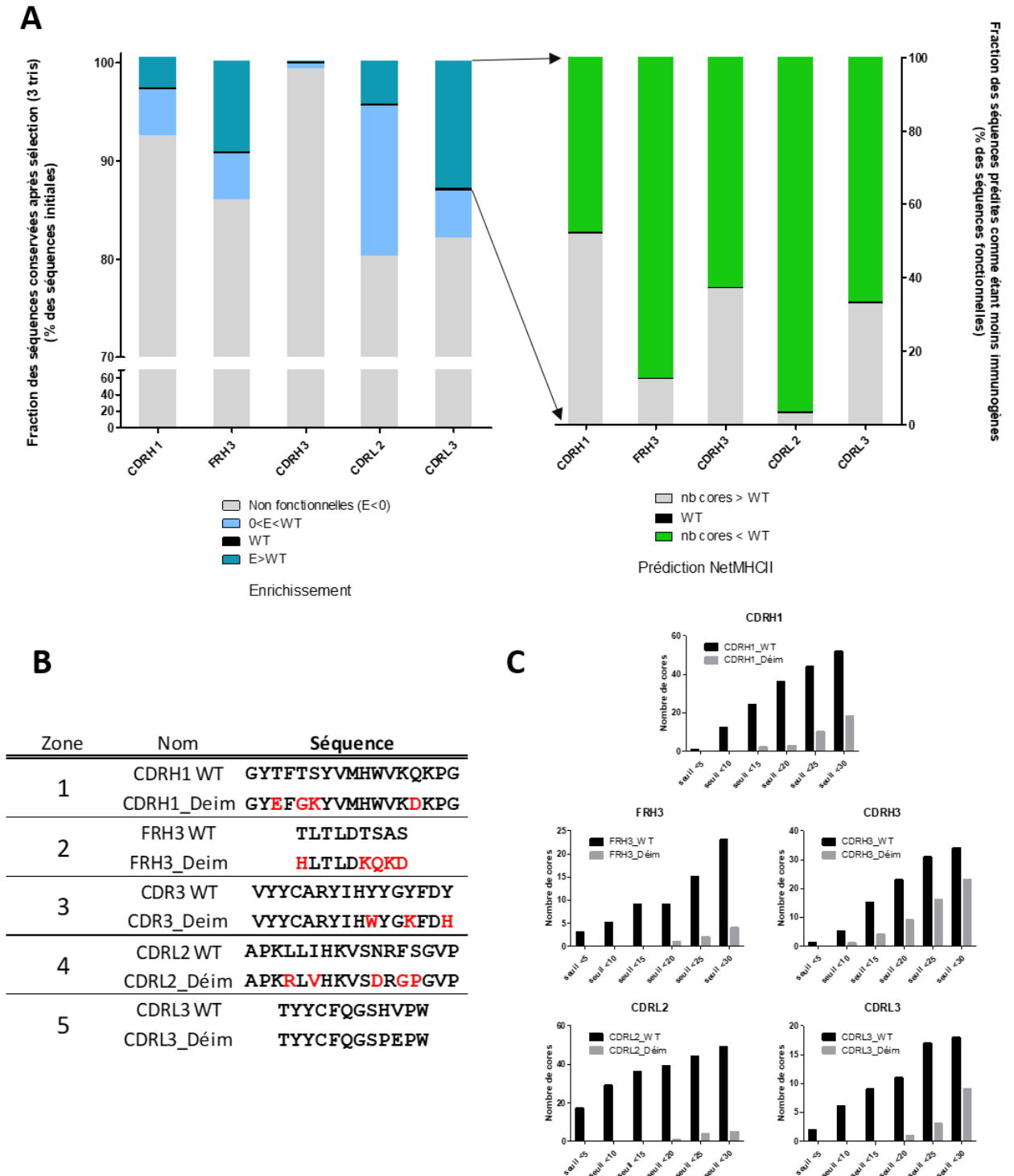
Le profil de la banque FRH3 est très différent. Une portion des mutations semble affecter l'appariement des deux chaînes : ni la liaison à l'antigène ni l'expression de la chaîne légère ne sont détectées (cellules correspondant aux points en bas à gauche). Une autre population est visible montrant à la fois une bonne expression et une bonne reconnaissance de l'antigène.

Les mutations réalisées dans la banque CDRH3 provoquent trois types d'effets : l'altération de l'appariement des deux chaînes (population majoritaire), l'altération de la liaison avec l'antigène sans modifier l'appariement et enfin la conservation de l'affinité pour l'antigène.

Pour les deux banques de la chaîne légère, on peut voir que les mutations effectuées affectent grandement l'appariement des deux chaînes (surtout pour la banque CDRL2) et que la majorité des variants de la banque CDRL3 perdent leur fonctionnalité alors qu'ils conservent un bon appariement, confirmant l'importance du CDRL3 dans la reconnaissance de l'antigène.

Malgré les différences notables de comportement des différentes banques, les résultats montrent qu'il existe au sein de chaque banque des mutants permettant de préserver à la fois l'expression des Fab et la reconnaissance des antigènes SipD et IpaD. Les données du DMS ont permis de concevoir des banques de dé-immunisation contenant une part importante de variants actifs. En effet, un essai préliminaire de conception de banque de variants dé-immunisés sans inclure de contrainte fonctionnelle du DMS a abouti à une banque contenant quasiment exclusivement des anticorps inactifs (résultats non montrés dans ce travail).

Pour chaque banque, la population non triée et la population triée à l'issue du dernier tri ont été séquencées (voir section matériel et méthodes). Les résultats du séquençage des cinq banques sont présentés en **Figure 12A**. La proportion de clones fonctionnels (c'est-à-dire retrouvés à l'issue du tri 3) est très différente en fonction des banques. Celle-ci va de 20% pour la banque CDRL2 à moins de 1% pour la banque CDRH3. Les données de NGS ont également permis de calculer pour chaque zone l'enrichissement de chaque clone après les étapes de tri. Cet enrichissement peut alors être comparé au facteur d'enrichissement de la séquence initiale (l'anticorps humanisé « Dezimulab »).



**Figure 12. Bilan du criblage des banques.** (A, gauche) Synthèse des scores d'enrichissement des variants après les trois tris fonctionnels pour chaque banque. 100% représentent le nombre total de séquences indépendantes retrouvées dans les banques non triées. Trois catégories sont représentées : les séquences perdues après tri (gris), les séquences conservées mais avec un enrichissement inférieur à celui du wild-type (bleu clair) et celles avec un enrichissement supérieur au wild-type (bleu foncé). (A, droite) synthèse des nombres de cœurs de liaison aux molécules HLA-II prédits seulement pour les séquences ayant un score d'enrichissement supérieur au wild-type dans chaque zone. (B) séquences des versions dé-immunisées choisies dans chaque zone. Les mutations par rapport à la séquence originale sont indiquées en rouge. (C) Comparaison de la prédiction du nombre de cœurs de liaison aux molécules HLA-II à différents seuils entre la version wild-type et dé-immunisée pour chaque zone.

Selon les banques, la proportion de clones avec un enrichissement supérieur à celui de l'anticorps parent est comprise entre 0,1% pour la banque CDRH3 à 13% pour la banque CDRL3. Les séquences des variants avec un score d'enrichissement supérieur à celui de l'anticorps de référence ont été réévaluées pour leur score d'immunogénicité (**Figure 12A**). Entre 48% et 97% de ces séquences ont un score d'immunogénicité inférieur à celui de l'anticorps humanisé Dezimulab. On constate que la conception des banques en utilisant les données de DMS couplées à la fonctionnalité de dé-immunisation du logiciel Dezimulab permet d'obtenir une fraction importante de clones fonctionnels pour chacune des banques (en dehors du CDRH3) et que ces clones sont également moins immunogènes sur le plan de la prédiction *in silico*.

Pour chacune des zones, la séquence fonctionnelle la plus dé-immunisée a été sélectionnée. C'est-à-dire que parmi les séquences qui présentent un enrichissement supérieur à celui du *wild-type*, nous avons choisi celle avec le score d'immunogénicité le plus faible. Ces séquences sont présentées en **Figure 12B**. Les séquences sélectionnées contiennent respectivement 4, 5, 3, 5 et 2 mutations par rapport à la séquence initiale.

Pour évaluer la différence d'immunogénicité prédite entre les séquences initiales et les séquences des variants dé-immunisés, nous avons à nouveau utilisé NetMHCIIpan pour comparer le nombre de cœurs prédits pour des seuils compris entre 5 et 30%. Pour chaque zone, les scores d'immunogénicité des variants sélectionnés (en gris) et de la séquence parentale (en noir) sont présentés en **Figure 12C**. Pour chaque zone, on peut observer une baisse significative des scores d'immunogénicité.

Les variants sélectionnés pour les zones 2 (FRH3) et 4 (CDRL2) semblent les mieux dé-immunisés *in silico* avec un très faible nombre de cœurs prédits, y compris pour les seuils les plus élevés. Les variants correspondant aux zones 1 (CDRH1) et 5 (CDRL3) montrent également une absence de cœurs prédits pour les seuils les plus bas (5% et 10%) et une baisse importante pour les seuils plus élevés. Enfin pour la zone 3 (CDRH3), le variant sélectionné montre une baisse du nombre de cœurs prédits supérieure à 50% pour les seuils compris entre 5% et 25%. Néanmoins tous les cœurs ne sont pas abolis au sein de cette zone, ce qui est probablement à mettre en relation avec la faible permissivité de la zone d'un point de vue fonctionnel.

Il est intéressant de noter qu'il existe une probable corrélation entre le degré de dé-immunisation et la proportion de clones fonctionnels à l'issue des étapes de tri. En effet, les banques des zones 2 et 4 ont produit respectivement 9% et 13% de séquences avec un enrichissement au moins égal à celui du *wild-type*, 3% et 4% pour les zones 1 et 5 et 0,1% pour la zone 3. Ainsi, il est probable que plus une zone est permissive d'un point de vue de la fonctionnalité, plus il y aura de chances d'identifier un clone fonctionnel avec un score d'immunogénicité faible.



Les cinq zones immunogènes ont été dé-immunisées séparément. Les variants sélectionnés possèdent des scores de prédiction d'immunogénicité *in silico* plus favorables que l'anticorps parent. La liaison à l'antigène semble préservée, comme l'attestent les scores d'enrichissement élevés. La poursuite du processus de dé-immunisation nécessite l'assemblage des mutations sélectionnées au sein d'un même anticorps. Cet assemblage a été réalisé en plusieurs étapes, d'abord pour les trois zones de la chaîne lourde, puis en ajoutant les zones de la chaîne légère (**Figure 13A**). Les différents assemblages ont alors été produits au format Fab dans des cellules HEK et leur fonctionnalité a été évaluée en BLI.

Dans la construction D, c'est d'abord les trois zones dé-immunisées de la chaîne lourde qui ont été assemblées avec une chaîne légère chimérique. Des mesures d'affinité préliminaires en BLI indiquent que la vitesse de dissociation ( $k_{off}$ ) de cette construction est équivalente à celle du Fab chimérique. Comme ces mesures ont été effectuées directement à partir des surnageants de culture et non à partir de Fab purifiés, la concentration n'est pas connue, donc la vitesse d'association ( $k_{on}$ ) n'a pas pu être calculée. Cette mesure semble indiquer une affinité préservée de la construction et de nouvelles mesures à l'Octet avec des Fab purifiés seront réalisées dans les prochaines semaines pour compléter ces résultats dans le manuscrit final.

La construction E incorpore non seulement les mutations dans la chaîne lourde (comme la construction D), mais également les mutations des deux zones de la chaîne légère. Les mesures d'affinité en Octet ont révélé une moindre fonctionnalité de cette construction par rapport au Fab chimérique, avec des pertes d'affinité atteignant un facteur 20 pour SipD et plus de 20000 pour IpaD. Ces pertes significatives d'affinité suggèrent que les zones dé-immunisées ne peuvent pas systématiquement être traitées de façon indépendante. Il est ainsi possible que certaines mutations ne fonctionnent que dans un contexte donné et qu'elles ne soient pas compatibles avec d'autres mutations introduites par ailleurs au sein d'autres zones. Nous avons alors cherché à identifier quelle zone pouvait être responsable de la perte d'affinité, le CDRL2 ou le CDRL3.

Lors du processus d'humanisation, l'analyse des données du DMS suggérait une moindre participation du CDRL2 dans l'interaction avec les antigènes. En effet, peu de substitutions y sont susceptibles de causer une perte d'affinité (cf. données du DMS négatif **Figure 5**). Nous avons alors généré une version de l'anticorps avec un CDRL2 non pas humanisé (avec seulement les *frameworks* en amont et en aval humains), mais complètement humain (construction C). Les mesures d'affinité réalisées avaient alors démontré que le CDRL2 était tout de même impliqué dans l'interaction mais que la perte d'affinité était modérée. En effet, cette construction conservait une affinité nanomolaire pour les deux antigènes SipD et IpaD (**Figure 13A**).

Nous avons donc choisi de réintroduire le CDRL2 humain dans l'anticorps dé-immunisé pour essayer d'en restaurer la fonctionnalité. La construction F ainsi obtenue s'est révélée beaucoup plus affine pour les deux antigènes, avec des valeurs subnanomolaires pour SipD et nanomolaires

pour IpaD (**Figure 13A**). Cette observation semble suggérer la présence de substitutions au sein de la zone 4 modifiée incompatibles avec d'autres substitutions incluses dans les autres zones de dé-immunisation. Il est possible que les mutations introduites dans le CDRL2 interfèrent avec le repliement du CDRH3 dé-immunisé. En effet, ces deux CDR sont en contact dans la plupart des structures d'anticorps. La modification du repliement du CDRH3 pourrait donc être la cause de cette perte d'activité. Avec davantage de temps, il aurait été préférable de recommencer le processus de dé-immunisation du CDRL2 dans le contexte dé-immunisé (c'est-à-dire avec les quatre autres zones dé-immunisées déjà assemblées).

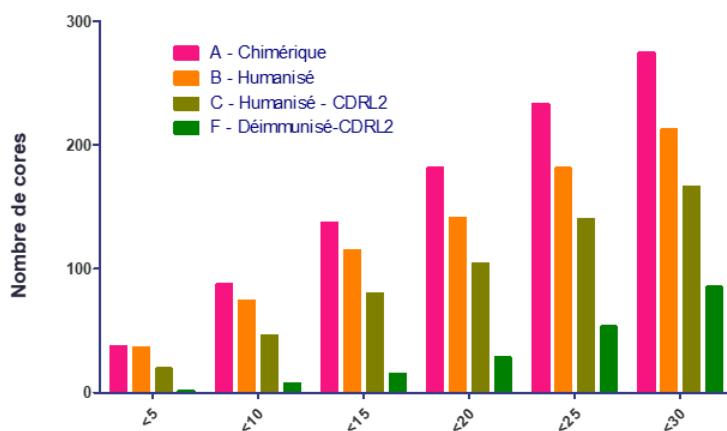
**A**

Construction	VH			VL		KD	
	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	SipD	IpaD
A*	Murin	Murin	Murin	Murin	Murin	338 pM	41 pM
B*	Humanisé	Humanisé	Humanisé	Humanisé	Humanisé	~WT	~WT
C*	Humanisé	Humanisé	Humanisé	Humain	Humanisé	2,8 nM	1,5 nM
D	Déim	Déim	Déim	Murin	Murin	~WT	~WT
E	Déim	Déim	Déim	Déim	Déim	6 nM	> 1 $\mu$ M
F*	Déim	Déim	Déim	Humain	Déim	403 pM	2,7 nM

~WT : Pour cette construction, seul le  $k_{off}$  a été mesuré et est de l'ordre de celui du wild-type

\* : constructions sélectionnées pour être évaluées dans les tests *in vitro* d'immunogénicité

**B**



**Figure 13. Effet de l'assemblage des zones dé-immunisées au sein d'un même anticorps sur la fonctionnalité et prédiction du nombre de cœurs de liaison aux molécules HLA-II.** (A) Sur la base de l'anticorps humanisé Dezimulab, différents assemblages des 5 zones immunogènes ont été testés. Lorsque les zones sont indiquées « murin », l'intégralité de la chaîne variable concernée est murine. Dans les autres cas, les zones 1 à 5 sont soit humanisées (humanisé), dé-immunisées (Déim) ou humaines (humain). Pour chacune de ces constructions, le Fab a été produit et sa fonctionnalité évaluée par BLI (*Bio Layer Interferometry*). (B) Prédiction du nombre de cœurs de liaison aux molécules HLA-II à différents seuils pour les quatre versions sélectionnées.

A l'issue du processus de dé-immunisation, nous avons pu obtenir un anticorps spécifique des deux antigènes (construction F). Le profil de sélectivité se retrouve légèrement modifié par l'ingénierie. En effet, si l'affinité pour SipD est maintenue (403 pM vs 338 pM), une diminution de l'affinité pour IpaD est observée (2,7 nM vs 41 pM). Cette observation est peut-être liée à la pression de sélection appliquée lors des étapes de tri. Une sélection plus stricte vis-à-vis de l'affinité pour IpaD pourrait éventuellement permettre d'identifier des clones avec des profils différents, l'enjeu principal étant ici de chercher des clones dé-immunisés.

Notre approche a permis de sélectionner quatre nouvelles constructions qui devront être évaluées *in vitro* pour leur immunogénicité. Chaque anticorps correspond à une des étapes du processus de dé-immunisation. Ainsi, la **construction A** qui correspond à l'anticorps IpaD-318 chimérique est le point de départ de la dé-immunisation. La **construction B** permettra de voir l'impact d'une humanisation classique sur l'anticorps. La **construction F** sera la construction fonctionnelle dé-immunisée. Nous avons également choisi d'inclure la construction C comme contrôle pour évaluer l'impact de la zone 4 avec une séquence totalement humanisée.

Le bilan des scores d'immunogénicité à différents seuils pour ces quatre constructions est présenté en **Figure 13B**. On constate une nette baisse du nombre de cœurs de liaison prédits tout au long du processus de dé-immunisation. Les cœurs prédits aux seuils les plus faibles (5% et 10%) sont ceux qui ont été le mieux dé-immunisés avec une réduction du score d'immunogénicité d'environ 97% à un seuil de 5% et 92% à un seuil de 10% entre la version dé-immunisée et la version chimérique de l'anticorps IpaD-318. La diminution du score est également importante concernant les seuils intermédiaires : 89% à un seuil de 15% et 85% au seuil de 20%. Aux seuils les plus élevés de 25% et 30%, l'efficacité de la dé-immunisation est moindre avec respectivement 77% et 69% de diminution. Enfin, si l'on se place à un seuil intermédiaire de 20%, la réduction de 85% du score d'immunogénicité peut être attribuée pour 23% à l'humanisation, 20% à l'humanisation du CDRL2 et 42% à la dé-immunisation des quatre autres zones. On constate donc que selon notre critère d'évaluation de l'immunogénicité *in silico*, l'humanisation de l'anticorps représente seulement une petite partie de l'effort de dé-immunisation.

Finalement, notre processus de dé-immunisation a permis de générer des anticorps aux séquences plus « humaines » d'une part et des variants ayant un plus faible nombre de séquences prédites comme potentiel épitopes T par les algorithmes. Une perspective de ce travail consiste naturellement à comparer l'immunogénicité de ces séquences, c'est-à-dire la capacité de ces anticorps à stimuler la prolifération de lymphocytes T CD4+ spécifiques lors de tests cellulaires.

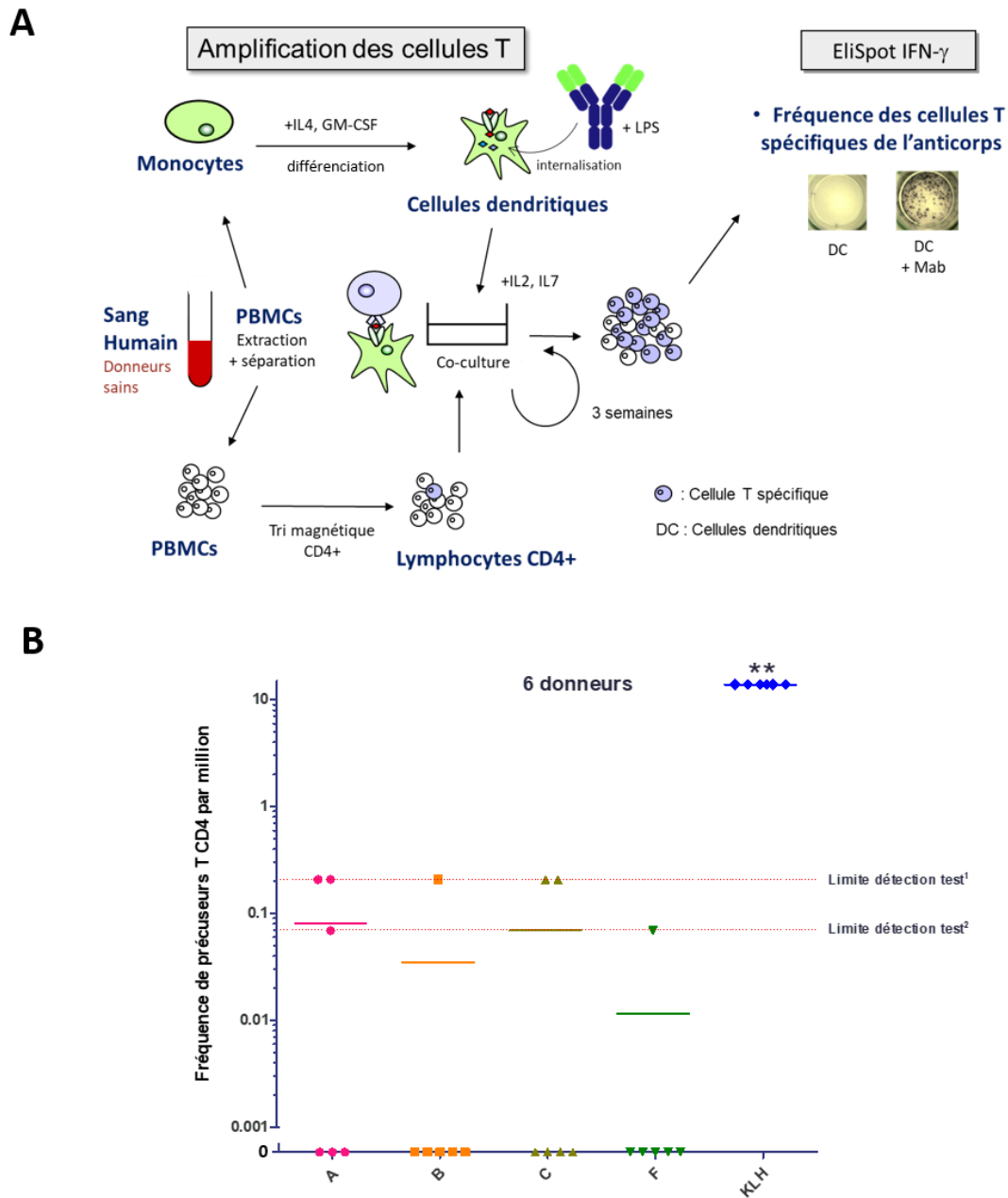
Pour permettre la mise en place de ces tests cellulaires *in vitro*, nous avons procédé au reformatage de ces anticorps au format IgG avec un Fc humain. Ceux-ci ont alors été produits par transfection transitoire de cellules HEK293. Les données de production et de purification ne semblent pas indiquer de différence notable en termes d'agrégation et de production. Cela suggère

que le processus de dé-immunisation n'a pas eu d'influence majeure en termes de productivité et de propriétés biophysiques. Ces quatre anticorps, produits au format IgG seront évalués *in vitro* pour leur immunogénicité et *in vivo* pour leur capacité protectrice.

## VI. Evaluation *in vitro* préliminaire de l'immunogénicité

Le test de prolifération des lymphocytes T CD4, mis au point au sein de l'équipe, fournit une estimation du nombre de lymphocytes T spécifiques de la protéine testée et circulant dans le sang de donneurs sains. Le principe de ce test est décrit en **Figure 14A**. A partir de sang de donneurs sains humains, les lymphocytes T CD4+ et les monocytes sont isolés. Les monocytes sont différenciés en cellules dendritiques puis incubées avec la protéine d'intérêt afin de réaliser les étapes d'internalisation, d'apprêtement de la protéine et de présentation des peptides par les molécules HLA-II. Les MoDC chargés sont co-cultivés avec les lymphocytes T CD4+ pendant une semaine. Les fréquences des cellules T du répertoire naïf spécifique d'un antigène étant très faibles, cette étape de stimulation est répétée 3 fois afin de pouvoir les détecter. On dispose de 200 000 ou 300 000 LT CD4+ par puits et les puits positifs sont détectés par EliSpot anti INF- $\gamma$ . En raison des faibles fréquences de précurseurs en début de culture, la distribution des lymphocytes T suit une loi de poisson. Dans ces conditions, avoir un puits positif signifie qu'au moins un lymphocyte spécifique de la protéine d'intérêt était présent dans le puits. On peut remonter à la fréquence de lymphocytes T spécifiques de la protéine d'intérêt dans le sang du donneur en considérant les puits négatifs et qui ne contiennent donc pas de cellules T spécifiques au début de la culture.

Les fréquences de précurseurs spécifiques ont été déterminées pour six donneurs et pour les quatre anticorps testés. Ces résultats préliminaires sont présentés en **Figure 14B**.



**Figure 14. Test de prolifération des lymphocytes T sur les quatre versions sélectionnées. (A)** Principe du test de prolifération des lymphocytes T CD4. Les PBMCs (*Peripheral Blood Mononuclear Cells*) sont d'abord extraites du sang total de donneurs sains. Ensuite les monocytes sont isolés des autres cellules et différenciés en cellules dendritiques (MoDC). Les MoDC sont ensuite maturés et incubés avec la protéine d'intérêt. Les Lymphocytes T CD4 sont isolés des PBMCs restantes par tri magnétique et mis au contact des MoDC matures présentant les peptides issus de la protéine d'intérêt via les molécules HLA-II. Les Lymphocytes T CD4 sont cultivés au contact des MoDC pendant une semaine. Les lymphocytes sont restimulés de la même façon pendant trois semaines. La culture est réalisée en plaque 96 puits avec entre 24 et 48 puits par condition et entre 200 000 et 300 000 lymphocytes T par puits. A l'issue des trois semaines de culture un Elispot anti interféron  $\gamma$  est réalisé. Les lymphocytes sont mis au contact de MoDC chargés avec la protéine d'intérêt ou non chargés (contrôle négatif) et le nombre de lymphocytes qui sécrètent de l'interféron  $\gamma$  est dénombré par comptage des spots. Un puits est considéré positif si il y a au moins deux fois plus de spot dans la condition « MoDC chargés » que la condition « non chargés » avec un minimum de 25 spots. Il est ainsi possible de remonter à la fréquence de lymphocytes T CD4 spécifiques dans le sang du donneur en considérant qu'un puits positif correspond à un lymphocyte T spécifique. **(B)** Résultats préliminaires des moyennes de fréquences de lymphocytes T CD4 spécifiques obtenues pour six donneurs sur les quatre versions de l'anticorps. La KLH est une protéine immunogène qui joue le rôle de contrôle positif. \*\*: Ces points se trouvent à la limite haute de détection du test pour la KLH avec huit puits, ici  $f > 13,86$ . <sup>1</sup> : limite pour 24 puits avec 200 000 CD4 par puits. <sup>2</sup> : limite pour 48 puits avec 300 000 CD4 par puits

Les six donneurs testés ont tous bien répondu à la KLH qui est le contrôle positif de l'expérience, et pour laquelle il existe une fréquence élevée de précurseurs naïfs chez les donneurs sains. Les réponses mesurées contre les quatre anticorps sont faibles et sont systématiquement à la limite de détection du test. C'est-à-dire que sur les 24 ou 48 puits testés, 1 ou 0 puits est positif. Malgré ces faibles réponses, nous pouvons constater que l'anticorps chimérique est celui qui répond le plus dans le test avec 50% de donneurs répondeurs. Vient ensuite l'anticorps humanisé portant un CDRL2 humain avec deux répondeurs sur six et enfin l'anticorps humanisé et l'anticorps dé-immunisé avec un donneur répondeur sur six. La fréquence de précurseurs naïfs semble plus basse pour la version dé-immunisée ( $f=0,0115$ ) que pour les versions chimérique ( $f=0,0808$ ), humanisée ( $f=0,0346$ ) et humanisée avec le CDRL2 humain ( $f=0,0693$ ). Malgré l'ajout de nombreuses mutations non humaines sur l'anticorps (14 au total), la version dé-immunisée semble moins immunogène que les versions humanisées, ce qui laisse penser que la méthode utilisée pour sélectionner ces mutations n'ajoute pas d'immunogénicité supplémentaire. Ces résultats préliminaires sont encourageants et demandent à être complétés par des essais avec davantage de donneurs sains.

## Matériel et méthodes

### Prédiction de l'interaction peptide/molécule HLA II

Les prédictions d'interaction avec les molécules HLA II ont été réalisées à l'aide de l'algorithme netMHCIIpan 3.2.<sup>157</sup> Brièvement cet algorithme prédit la probabilité d'interaction d'une séquence avec des molécules HLA II choisies et fournit un résultat pour chaque allèle relatif à un ensemble de peptide. Une valeur de 1% pour un peptide signifie qu'il se situe parmi les 1% des peptides ayant la plus forte probabilité d'interaction avec les molécules HLA II. En collaboration avec le LIST au CEA Saclay, une plateforme permettant d'exécuter l'algorithme netMHCIIpan3.2 de manière parallélisée et d'étudier de manière systématique l'effet de chaque substitution a été développée. Pour chacune des prédictions réalisées, c'est le panel de 27 allèles publié par Greenbaum *et al*<sup>164</sup> qui a été utilisé (celui-ci est décrit en introduction). De plus, un ensemble de séquences recouvrant les gènes d'immunoglobulines humaines les plus fréquents a été utilisé afin d'évaluer le caractère humain ou non de chacun des cœurs.<sup>179</sup>

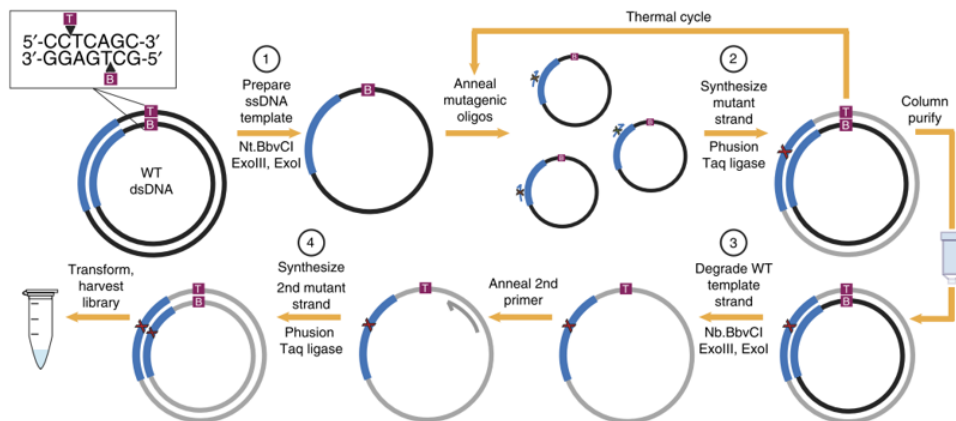
### Construction des banques de DMS

Les banques de type « one pot » ont été construites selon la méthode décrite par Wrenbeck *et al*<sup>250</sup> (**Figure 15A**). Brièvement, un des deux brins du plasmide qui porte la cassette d'intérêt à muter est digéré. Ce brin est resynthétisé à l'aide d'amorces portant un codon dégénéré (une amorce par position à randomiser). Ensuite, le second brin (qui ne porte pas de mutation) est digéré puis resynthétisé à l'aide d'une amorce externe à la zone de mutation. Ainsi, les deux brins du plasmide portent la mutation. Dans une dernière étape, les plasmides qui n'auraient pas été néoformés sont spécifiquement digérés à l'aide de l'enzyme de restriction DpnI. Enfin, le pool de plasmides mono-mutés est transformé dans des bactéries électrocompétentes.

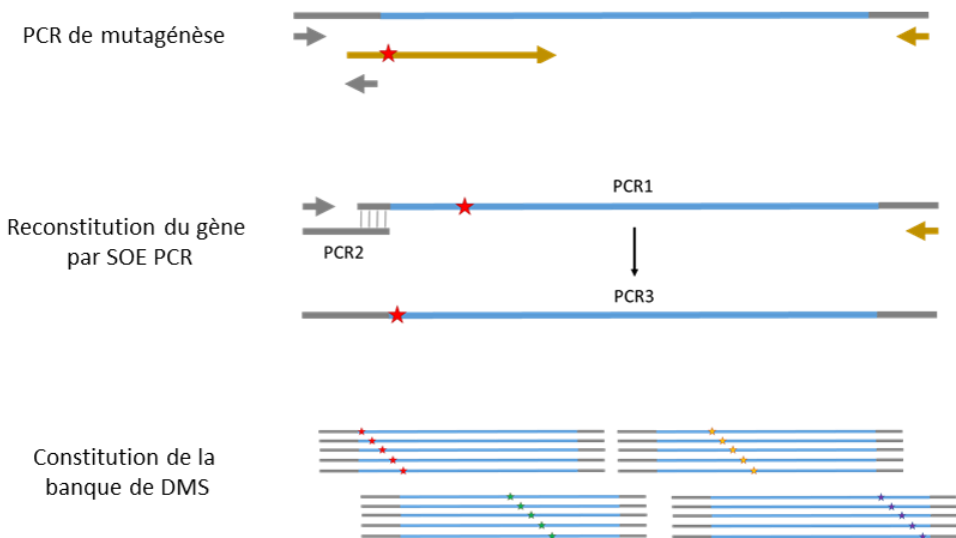
Les banques de DMS de type « SOE PCR » sont générées selon la méthode décrite en **Figure 15B**. Dans le cas du DMS de l'anticorps IpaD-318, pour optimiser la qualité des données de séquençage (Illumina MiSeq) et permettre deux lectures par brin (*paired end sequencing*), nous avons choisi de scinder en deux les banques sur chaque chaîne variable.



**A**



**B**



**Figure 15. Principe des deux méthodes utilisées pour générer les banques de DMS. (A)** La méthode « Plasmid-based one pot saturation mutagenesis ». **(B)** La méthode développée dans ce travail : SOE PCR.

## Construction des banques de dé-immunisation

Pour les banques de dé-immunisation, l'assemblage par SOE PCR a été utilisé. La diversité a été introduite à l'aide de codons dégénérés choisis à l'aide de l'outil Swiftlib (<http://rosettadesign.med.unc.edu/SwiftLib/>).

## Transformation et sélection par YSD

Les banques ont été clonées dans le plasmide pNT-NB dérivé du plasmide pCT-L7.5.126 (Addgene plasmide #429000). Comme le pCT-L7.5.126 il possède une origine de réplication levure CEN/ARS, un gène d'auxotrophie TRP, une origine de réplication bactérienne colE1, un gène de résistance à l'ampicilline et un promoteur bidirectionnel inductible au galactose GLA1/GAL10. Le clonage de la banque dans le plasmide d'expression en YSD a été réalisé par recombinaison homologue lors de la transformation de cellules EBY100 (ATCC® MYA-4941™; a GAL1-AGA1::URA3 ura3-52 trp1 leu2 D1 his3 D200 pep4::HIS2 prb1D1.6R can1 GAL). Pour chacune des transformations 1 µg de plasmide linéarisé (NheI et SalI pour VH, NcoI et Pfl23II pour VL) et 2 µg d'insert ont été utilisés pour chacune des transformations. Toutes les banques ont été générées par électroporation selon la méthode décrite par Benatuil *et al.*<sup>235</sup> Chaque banque a été transformée en une seule réaction (environ  $5 \cdot 10^7$  clones indépendants obtenus par réaction). Le nombre de clones transformés a été déterminé à partir d'un étalement d'une dilution au 1 : 1000, un nombre de clones transformés au moins dix fois supérieur à la taille de la banque a été respecté. Les banques ainsi clonées dans le plasmide ont été cultivées en milieu SD-CAA [6,7 g/L yeast nitrogen base without casamino acids, 20 g/L dextrose, 5 g/L casamino acids, 100 mM sodium phosphate pH 6.0] et ont été passées deux fois avant d'induire l'expression en milieu SG-CAA [6.7 g/L yeast nitrogen base without casamino acids, 20 g/L galactose, 5 g/L casamino acids, 100 mM sodium phosphate, pH 6.0] afin de minimiser les doubles transformants. Les étapes de culture et d'expression ont été réalisées selon la description faite dans les parties précédentes.

La sélection des banques de DMS a été faite en une seule étape par FACS sur un appareil ARIA III (Becton Dickinson, Franklin Lakes, États-Unis). Après induction de l'expression, les banques ont été incubées 3h à 20°C avec l'antigène SipD biotinylé à une concentration de 2 nM. Après lavage des cellules avec du PBS 0,1% BSA celles-ci ont été marquées à l'aide d'un anticorps dirigé contre le domaine C $\kappa$  couplé APC (Thermo Fisher Scientific, Waltham, États-Unis; dilution 1:100) et d'une streptavidine couplée PE (Thermo Fisher Scientific, Waltham, États-Unis; dilution 1:100). La sélection des banques a été faite à l'aide des fenêtres de tri décrites en **Figure 3**.

La sélection des banques de dé-immunisation a été réalisée par FACS en plusieurs étapes. Pour chaque banque, trois tris ont été effectués. Le premier tri est réalisé avec l'antigène SipD à

une concentration de 2 nM. Le deuxième tri est effectué en utilisant l'antigène IpaD à 2 nM. Enfin le dernier tri est un tri cinétique. Pour ce tri, les cellules induites sont saturées avec l'antigène SipD biotinylé puis lavées. Une seconde incubation est réalisée avec l'antigène non biotinylé à 1  $\mu$ M pendant 2h pour favoriser la dissociation de l'antigène biotinylé. La forme diagonale des fenêtres (**Figure 11**) permet de prendre en compte l'hétérogénéité de l'expression au sein de la population de levures.

## Séquençage NGS des banques et analyse des données

Les plasmides ont été extraits des cellules par lyse enzymatique à l'aide du kit Zymoprep Yeast Plasmid Miniprep II (Zymo Research, Irvine, Etats-unis). Les fragments correspondants ont ensuite été amplifiés et les adaptateurs Illumina et étiquettes de multiplexage ajoutés par deux étapes de PCR comme décrit par Kowalsky *et al* <sup>256</sup>. Les banques ont été séquencées en *paired-end* sur un MiSeq à l'aide d'un kit V2 2x150 cycles ou sur un iSeq100 en 2x150 cycles (Illumina, San Diego, Etats-unis). Pour les banques de DMS une profondeur de séquençage de 50X minimum a été respectée.

Les séquences ont été démultiplexées et traitées de manière indépendante sur la plateforme Galaxy (<https://usegalaxy.org/>) à l'aide des fonctions décrites par Blankenberg *et al.* <sup>257</sup> Les séquences sont appariées (fastq-join) et seules les séquences ayant un score de qualité supérieur ou égal à 30 sont conservées (FASTQ Quality Trimmer). Enfin, les séquences identiques sont dénombrées et agrégées (Group). Ces données sont ensuite traitées avec le logiciel R afin de calculer les fréquences des différents mutants et ainsi déterminer leur enrichissement (rapport entre la fréquence du mutant dans la banque triée et sa fréquence dans la banque initiale non triée).

## Production et Purification de Fab et d'IgG

Pour les variants de l'anticorps IpaD-318 produits au format IgG, les chaînes variables lourdes et légères ont été clonées dans les plasmides AbVec2.0-IGHG1 et AbVec1.1-IGKC respectivement. <sup>258</sup> Pour la production de Fab, la chaîne lourde a été clonée dans un plasmide dérivé du AbVec2.0-IGHG1 pour lequel les domaines CH2 et CH3 ont été retirés et remplacés par un tag 6His. La production de Fab et d'IgG a été réalisée par transfection transitoire de cellules HEK293 Freestyle (Thermo Fisher Scientific, Waltham, États-Unis) cultivées dans le milieu associé. La transfection a été réalisée à une densité de  $2,5 \cdot 10^6$  cellules par mL de culture, une solution de PEI (Sigma-Aldrich, Saint-Louis, États-Unis) a été utilisée comme agent de transfection. Les plasmides ont été ajoutés à la culture à un ratio 1 : 1 et à une concentration finale dans la culture de 1,5  $\mu$ g/mL pour chaque plasmide. Après 5 minutes d'agitation à 37°C et 8% de CO<sub>2</sub>, le PEI a été ajouté goutte à goutte à une concentration finale de 9  $\mu$ g/mL de culture. Après

24h sous agitation à 37°C et 8% CO<sub>2</sub>, la culture a été diluée au demi. La production a été stoppée 7 jours après transfection et le surnageant a été récupéré par centrifugation à 4°C, 10 min à 3000G puis 20 min à 20000G. Les protéines ont ensuite été purifiées à l'aide d'un système AKTA Purifier (GE Healthcare, Pittsburgh, États-Unis). Pour les Fab, la purification a été faite à l'aide d'une colonne HisTrap Excel (GE Healthcare, Pittsburgh, États-Unis) avec élution en tampon imidazole 500 mM. Suite à la purification, les Fab ont été dialysés contre du PBS. Les IgG ont été purifiés à l'aide d'une colonne HiTrap Protein A HP (GE Healthcare, Pittsburgh, États-Unis), suivit par une étape de *gel filtration* sur une colonne Sephacryl S-200 HR (GE Healthcare, Pittsburgh, États-Unis) afin de ne garder que la forme monomérique des IgG.

## Mesure d'affinité

Les mesures d'affinité ont été réalisées de manière cinétique avec un Octet Red96 (Molecular Devices, San Jose, États-Unis) selon le protocole décrit par Schroter *et al.*<sup>259</sup> Brièvement, l'antigène biotinylé est immobilisé sur des sensors streptavidine (Streptavidin (SA) Biosensor) à une concentration de 50 nM. Après saturation des sensors dans une solution de blocage contenant de la biotine (Sigma-Aldrich, Saint-Louis, États-Unis) à 10µg/mL, l'association et la dissociation des Fab sont mesurées pendant 20 et 30 minutes respectivement. Lors de l'analyse, la référence est soustraite à chacune courbe et un modèle Langmuir global 1 : 1 est appliqué pour obtenir les paramètres d'affinité.

Dans le cas où seul le paramètre de K<sub>off</sub> est disponible, les Fab n'ont pas été purifiés. Le surnageant de culture de la transfection transitoire des cellules HEK 293 Freestyle est dilué au 10<sup>ème</sup> dans le tampon utilisé pour réaliser les mesures d'affinité à l'Octet (PBS, Tween 0,05%, BSA 0,1%). Ainsi, la concentration du Fab n'est pas connue et seul le K<sub>off</sub> peut être déterminé.

## Induction de lignées T spécifiques

Les cellules mononuclées du sang périphérique (PBMC) ont été purifiées par Ficoll à partir de sang total de donneurs sains anonymes (Etablissement Français du Sang, Rungis). Les cellules dendritiques (DC) dérivées de monocytes ont été obtenues à partir de monocytes adhérents sur le plastique mis en culture cinq jours dans du milieu AIM-V (Invitrogen) supplémenté par 1000 u/mL d'IL-4 et 1000 u/mL de GM-CSF. Les lymphocytes T CD4 (LT CD4) ont été obtenus à partir de la fraction non adhérente au plastique des PBMC par tri magnétique positif à l'aide de billes anti-CD4 (Miltenyi Biotec). Pour les protéines, 20 000 DCs par puits (plaque 96 puits fond rond) ont été chargées sur la nuit à 37°C avec 3µM de protéine ou 0,25µM de KLH et leur maturation a été induite par ajout de lipopolysaccharide (LPS) à 1µg/mL. 200 000 ou 300 000 LT CD4 autologues ont été ajoutés par puits, 24 ou 36 puits par protéine et par donneur ont été mis

en culture selon le nombre de cellules obtenues lors des étapes de sélection. La culture a été réalisée dans 200  $\mu$ L de milieu IMDM (Invitrogen) supplémenté de 10% de sérum humain, 0,24 mM de glutamine, 0,55 mM d'asparagine, 1,5 mM d'arginine, 50 U/mL de pénicilline et 50  $\mu$ g/mL de streptomycine (IMDM complet). Lors de la première semaine de culture le milieu a également été supplémenté par 1000 U/mL d'IL-6 et 10 ng/mL d'IL-12. Les LT CD4 ont été restimulés après 7 jours et 14 jours de culture avec 20 000 DC autologues maturées et chargées sur la nuit précédente. A partir du jour 7 la culture a été réalisée en IMDM complet supplémenté par 10 U/mL d'IL-2 et 5ng/mL d'IL-7. A 21 jours de culture, chaque lignée (LT CD4 ayant proliféré dans un puit) a été collectée et sa spécificité a été évaluée par un test EliSpot.

### Révélation par EliSpot IFN- $\gamma$

Un anticorps anti-IFN $\gamma$  (MabTech) a été adsorbé (4°C sur la nuit) sur des plaques 96 puits multiscreen hemagglutinine (Merck Millipore) à 2,5 ng/mL. Les membranes des plaques ont ensuite été saturées 2h à 37°C avec du milieu IMDM complet puis lavées au PBS. La protéine d'intérêt (à 1  $\mu$ M) ou les peptides (à 5  $\mu$ M ou 0,2  $\mu$ M) ont été chargés sur des DC pendant 4h à 37°C en milieu AIM-V. Les cellules chargées ont ensuite été irradiées 6 minutes sous une lampe UV puis mises en culture avec des LT CD4 autologues dans du milieu AIM-V contenant 0,5 ng/mL d'IL-7 (5.103 DC chargées et 2.104 LT CD4 par puit). Après incubation sur la nuit et lavages, les puits ont été incubés 1h avec 0,25 ng/mL d'un anticorps anti-IFN $\gamma$  biotinylé (Mabtech) dans du PBS contenant 1% de BSA. Après lavage, les puits ont été incubés avec de l'extravidine couplée à une phosphatase (Sigma) diluée à 1 : 3000 dans du PBS/Tween 20 0,05%/BSA 1%. Après lavage, la révélation a ensuite été réalisée par incubation de 5 à 10 min avec du NBT/BCIP. Les *spots* ont finalement été dénombrés à l'aide d'un compteur AID ELISPOT Reader System. Les lignées LT CD4 sont considérées positives et spécifiques lorsqu'on compte au moins deux fois plus de *spots* en présence de la protéine ou du peptide qu'en son absence, avec une différence minimum de 25 *spots*.







## Conclusion générale

### Mise en place des outils d'affichage à la surface de levure pour la sélection de variants d'intérêt

Au cours de ma thèse, le *Yeast Surface Display* s'est révélé utile aussi bien pour la détermination de l'épitope que pour contrôler la fonctionnalité de l'anticorps tout au long du processus de dé-immunisation. Pour répondre aux deux problématiques, les outils permettant l'expression des antigènes et des anticorps à la surface des levures ont été mis en place.

Les premiers essais pour exprimer les anticorps IpaD-301 et IpaD-318 à la surface des levures se sont révélés décevants, en particulier en scFv. La littérature n'indiquait alors pas de consensus clair quant au choix du format d'anticorps à utiliser en YSD. Différents formats sont utilisés dans la littérature mais aucune comparaison entre ces différents formats n'a été effectuée au sein de la même étude<sup>224,233,260</sup>. Ainsi, un travail préalable sur les différents formats d'anticorps utilisables en YSD a été réalisé pour déterminer l'influence du choix du format des fragments sur la fonctionnalité et l'expression des anticorps. Pour les six anticorps testés (IpaD-301, IpaD-318 et 4 anti-TNF $\alpha$ ), le format le plus adapté s'est révélé être le format Fab. Ces résultats vont dans le même sens qu'une étude questionnant l'influence du choix du format d'anticorps en *phage display* sur le reformatage de l'anticorps en IgG<sup>227</sup>. Ce format a l'avantage d'être proche du format IgG car les chaînes lourde et légère ne sont pas liées entre elles par un *linker*. De ce fait, les mutations introduites dans la séquence de l'anticorps susceptibles d'altérer l'appariement des deux chaînes ne seront pas sélectionnées en YSD. Cette caractéristique permet d'espérer un meilleur transfert du format de criblage (Fab lié à la surface de la levure) à l'IgG ou au Fab soluble. Dans ce travail de thèse, tous les anticorps sélectionnés en YSD au format Fab puis produits aux formats Fab soluble ou IgG se sont révélés être fonctionnels. L'article que nous avons publié sur l'étude comparée des différents formats d'anticorps utilisables en YSD a déjà été cité pour justifier le choix du format Fab en YSD<sup>261</sup>.

### Implémentation du DMS et du séquençage à haut débit

De nombreuses approches d'ingénierie des protéines reposent sur la compréhension des interactions protéine/protéine. Ainsi, l'analyse de la permissivité des séquences à la mutation s'avère très utile pour comprendre la fonctionnalité et les modes d'actions de ces protéines. Par

ailleurs, le NGS est de plus en plus utilisé dans les projets d'ingénierie des protéines car il permet d'analyser l'évolution des banques au fur et à mesure des étapes de sélection.

Dans cette optique, nous avons choisi de coupler l'affichage à la surface des levures avec une approche de *Deep Mutational Scanning* et de séquençage haut débit. Cette technique permet l'analyse de la permissivité des résidus d'un anticorps <sup>215</sup>, la détermination d'épitopes <sup>251–253</sup> ou encore la détermination des surfaces d'interaction protéine-protéine <sup>262</sup>.

Au cours de ce travail, les outils de biologie moléculaire nécessaires à la génération des banques de DMS ont été mis en place. Par ailleurs le traitement des données de NGS a nécessité la mise en place d'outils bio-informatiques, avec différents scripts correspondant aux différentes étapes de traitement (qualité, appariement, énumération, calcul de fréquences, etc...) avec les logiciels R et Galaxy.

Aujourd'hui, les solutions de séquençage NGS se sont démocratisées et de nouveaux appareils accessibles comme le iSeq100 permettent aux laboratoires de s'équiper à un coût modéré. Dans notre cas, ceci nous a permis d'augmenter considérablement les débits et de rendre facilement accessibles les approches nécessitant du séquençage à haut débit pour les autres projets du laboratoire. Cependant, le traitement des données de NGS reste compliqué et nécessite des compétences en bio-informatique. Le développement du NGS devra à l'avenir être accompagné de solutions simples de traitement des données pour se démocratiser encore davantage.

## Détermination de l'épitope de l'anticorps IpaD-318

L'anticorps IpaD-318 murin possède une double spécificité vis-à-vis des antigènes IpaD (*Shigella*) et SipD (*Salmonella*), avec de très bonnes affinités, inférieures au nanomolaire. Cet anticorps s'est révélé être protecteur dans les modèles murins d'infection aux Salmonelles et aux Shigelles lorsqu'il est administré avant le début de l'infection.

Ces propriétés intéressantes nous ont conduit à rechercher l'épitope d'IpaD-318 afin de mieux comprendre son mode d'action. Nous avons choisi d'utiliser le *Yeast Surface Display* pour identifier les substitutions de l'antigène IpaD abolissant la reconnaissance par l'anticorps. Cette approche couplant DMS et séquençage haut débit a permis d'identifier un épitope conformationnel au sein de la partie apicale de l'antigène. Cette zone est située à l'intérieur du canal formé par oligomérisation d'IpaD ou de SipD à l'extrémité de l'aiguille d'injection du SST3. Il est intéressant de noter que cet épitope est a priori inaccessible dans le cas où la coiffe de l'aiguille est totalement formée. Néanmoins, lorsqu'il est injecté préalablement à l'infection, IpaD-318 se révèle protecteur *in vivo*. Plusieurs hypothèses ont été formulées dans ce travail de thèse pour expliquer son mode d'action. Celle qui nous semble la plus probable serait le blocage

de l'une des étapes de l'assemblage de la coiffe du SST3 comme l'oligomérisation d'IpaD (ou de SipD) avec lui-même ou la liaison IpaD-IpaB (ou SipD-SipB).

Pour valider cette hypothèse, nous avons étudié la capacité de neutralisation de l'anticorps dans un test *in vitro* de lyse de globules rouges par les Shigelles. Ce test n'a pas permis de mettre en évidence un effet neutralisant de l'anticorps sur la lyse des globules rouges. Il est possible que ce test ne permette pas de mettre en évidence le blocage des étapes d'assemblage de la coiffe. Dans la littérature, des anticorps<sup>74</sup> ou des VHH<sup>59</sup> dirigés contre IpaD (*Shigella*) sont capables de neutraliser le SST3 et empêchent la lyse des globules rouges *in vitro* dans ce même test. L'un de ces VHH nous a servi de témoin positif dans les tests de neutralisation que nous avons réalisés, validant ainsi nos conditions expérimentales. L'anticorps IpaD-318 a donc un mécanisme d'action différent des anticorps décrits dans la littérature : il ne semble pas être neutralisant dans le test *in vitro* considéré mais est tout de même protecteur *in vivo*. Les anticorps et VHH issus de la littérature n'ont pas été évalués *in vivo*. Lors des essais de protection *in vivo*, l'anticorps était efficace seulement s'il était injecté avant le début de l'infection. Il serait pertinent d'évaluer les autres anticorps décrits dans la littérature pour savoir s'ils présentent le même profil de protection *in vivo*.

La détermination de cet épitope de liaison conservé entre les deux antigènes pourra servir à la validation de modèles 3D d'assemblage de la coiffe du SST3 de *Salmonella* et *Shigella*. En effet, différents modèles sont proposés pour l'assemblage et le fonctionnement de la coiffe de l'aiguille d'injection de ces bactéries mais les hypothèses les concernant manquent d'outils de validation. Il est intéressant de noter que l'anticorps IpaD-318 apporte un éclairage nouveau sur la grande similarité du mécanisme d'action entre l'injectisome de *Salmonella* et celui de *Shigella*. En effet, la littérature, prolifique concernant *Shigella*, l'est bien moins concernant *Salmonella*.

Sur un plan plus technique, la méthodologie utilisée pour déterminer l'épitope de l'anticorps s'est révélée efficace et pourra être utilisée dans d'autres projets et à plus large échelle. En effet, lors du développement d'un anticorps thérapeutique, de nombreux candidats sont générés. Si les critères d'affinité et de sélectivité sont souvent des critères importants, le mode d'action des différents candidats peut également s'avérer décisif pour obtenir l'effet thérapeutique escompté. Dans cette optique, une caractérisation rapide des épitopes des différents candidats pourrait permettre une sélection judicieuse des clones.

## Dé-immunisation de l'anticorps IpaD-318

L'objectif principal de ma thèse était de répondre à la question suivante : **comment modifier la séquence d'un anticorps afin de réduire son immunogénicité tout en conservant ses caractéristiques fonctionnelles ?**

Pour répondre à cette question, une méthode de dé-immunisation basée sur le principe de la suppression des épitopes T CD4+, à l'origine des réponses immunitaires chez les patients a été développée durant ce travail de thèse. La suppression des épitopes T a été utilisée dans plusieurs articles traitants de la dé-immunisation de protéines thérapeutiques comme des toxines <sup>191,192</sup> ou des enzymes <sup>197</sup>. A notre connaissance, cette méthode n'a pour l'instant jamais été appliquée à des anticorps. Dans ces projets, les épitopes T sont d'abord identifiés à l'aide de tests de prolifération des lymphocytes T CD4+. Cette étape est longue et nécessite à elle seule souvent plusieurs mois d'expériences. Aussi, nous avons fait le choix de ne travailler que sur la base de la prédiction des épitopes T potentiels par des algorithmes, tels que NetMHCIIpan pour évaluer leur liaison aux molécules HLA-II <sup>157</sup>. C'était un choix délibéré puisque nous voulions évaluer une méthode de dé-immunisation rapide qui ne nécessiterait pas une étape d'identification des épitopes *in vitro* au préalable (étape qui allongerait considérablement le processus de dé-immunisation).

L'anticorps parent IpaD-318 est un anticorps murin que nous avons d'abord chimérisé. Cet anticorps a ensuite été humanisé pour réduire le nombre de zones potentiellement immunogènes au sein de la séquence. Cinq zones potentiellement immunogènes ont été identifiées par l'algorithme de prédiction. A l'aide des cartes de DMS et de l'algorithme de dé-immunisation développé par le LIST (CEA), nous avons défini une stratégie de dé-immunisation consistant à introduire des combinaisons de substitutions potentiellement tolérées d'un point de vue fonctionnel dans les zones immunogènes.

Pour chacune des zones identifiées, des banques de mutants ont été générées et criblées à la surface des levures avec une succession de trois tris fonctionnels. A l'issue des étapes de sélection, des variants avec des scores d'immunogénicité *in silico* réduits et des propriétés fonctionnelles préservées ont été choisis.

Le DMS s'est révélé incontournable pour la conception de banques de mutants. En effet, un essai préliminaire de génération de variants dé-immunisés sans inclure les contraintes fonctionnelles du DMS a abouti à une banque contenant quasiment exclusivement des anticorps inactifs (résultats non montrés dans ce travail). La mise en place du DMS ainsi que de tous les outils bio-informatiques associés a donc été déterminante dans ce projet.

En assemblant les séquences des variants identifiés pour chaque zone, un anticorps dé-immunisé sur l'ensemble de sa séquence a été obtenu. Cette étape a nécessité certains ajustements car l'anticorps obtenu avait perdu son activité d'origine. Nous avons identifié la zone qui posait problème et l'avons remplacé par son équivalent humain, ce qui a permis de retrouver une fonctionnalité satisfaisante (au moins de l'ordre du nanomolaire pour les deux antigènes). Cette étape du processus pourra être améliorée à l'avenir, en procédant par exemple à une dé-immunisation séquentielle de l'anticorps. Il est en effet possible de dé-immuniser la chaîne lourde en premier puis de cribler les banques de la chaîne légère en ayant choisi la chaîne lourde dé-immunisée. Une autre stratégie plus fastidieuse consisterait à dé-immuniser les zones les unes après les autres (ce qui oblige à attendre les résultats sur une zone pour pouvoir passer à la suivante). Une dernière stratégie pourrait être de garder une certaine diversité de solutions fonctionnelles pour chaque zone à dé-immuniser. La dernière étape consisterait alors à recombinaison les différentes zones entre elles pour former une nouvelle banque à cribler.

La version dé-immunisée de l'anticorps IpaD-318 possède un score d'immunogénicité *in silico* fortement réduit tout en conservant sa spécificité pour les deux antigènes. La stratégie de dé-immunisation que nous avons développée semble donc répondre aux objectifs fixés au début de la thèse. Il est cependant important de noter que les seuils utilisés pour calculer les scores d'immunogénicité sont relativement arbitraires. L'anticorps dé-immunisé possède un score d'immunogénicité fortement réduit lorsqu'on le calcul pour un *rank percentile* inférieur à 30%. Il existe peu d'études traitant du problème du choix du *rank percentile* pour évaluer l'immunogénicité potentielle d'une protéine. Deux études portant sur l'évaluation *in silico* de l'immunogénicité se basent sur un seuil de 20% <sup>168,263</sup>, nous avons choisi un seuil de 30% afin d'augmenter la probabilité de réduire significativement l'immunogénicité de l'anticorps.

Il s'agit pour la suite de vérifier que la réduction du score d'immunogénicité mesurée par les algorithmes de prédiction à un seuil de 30% sera suffisante pour avoir un effet sur l'immunogénicité réelle déterminée *in vitro* et que l'anticorps ainsi produit aura conservé son efficacité *in vivo*.

L'immunogénicité de l'anticorps dé-immunisé a été évaluée dans une première expérience préliminaire de prolifération des lymphocytes T CD4+. Les premiers résultats obtenus sur six donneurs sont encourageants, une fréquence de précurseurs plus faible pour l'anticorps dé-immunisé ayant été observée. Ces expériences seront poursuivies pour augmenter le nombre de donneurs et donner une valeur statistique aux résultats.

Pour les mêmes raisons que le *mapping* des épitopes T, la décision a été prise de ne pas évaluer l'immunogénicité *in vitro* initiale de l'anticorps chimérique avant sa dé-immunisation. Son immunogénicité a donc été évaluée en même temps que les trois autres versions de l'anticorps humanisées et dé-immunisée. Les fréquences de précurseurs obtenues par Delluc *et al* <sup>264</sup> dans le



même test avec le rituximab ( $f=0,35$ ) et l'infliximab ( $f=0,2$ ) qui sont des anticorps chimériques sont bien plus élevées que la fréquence préliminaire obtenue dans ce travail pour l'anticorps IpaD-318 chimérique ( $f=0,08$ ). L'anticorps IpaD-318 chimérique est donc beaucoup moins immunogène qu'attendu. Cette faible immunogénicité de départ risque de rendre difficile la mesure d'une éventuelle diminution d'immunogénicité apportée par la dé-immunisation.

Le test de prolifération utilisé dans ce travail de thèse est reconnu comme étant plus sensible que les autres tests disponibles dans les autres laboratoires ou sur le marché <sup>265</sup>. La sensibilité du test vient principalement du fait que les lymphocytes T CD4+ spécifiques sont amplifiés pendant trois semaines, ce qui permet de détecter de faibles réponses. Malgré cela, au-delà de l'exemple de l'anticorps traité dans ce travail de thèse, les réponses sont globalement faibles pour la plupart des autres anticorps évalués dans ce test *in vitro* d'immunogénicité. Les fréquences de précurseurs vont de 0,02 pour le trastuzumab à 0,35 pour le rituximab, à l'inverse de nombreuses autres protéines testées (16 pour la KLH et entre 0,3 et 6,3 pour un anticorps murin) <sup>264</sup>.

La présence du fragment Fc, reconnu par les récepteurs FcRn présents dans les endosomes des cellules dendritiques <sup>86</sup> pourrait freiner le mécanisme d'internalisation des anticorps par les cellules dendritiques. Les cellules dendritiques présentent également les récepteurs FcγRIIb et FcγRI <sup>86</sup> activateurs de la maturation des endosomes, favorisant ainsi l'internalisation des anticorps. Il existerait donc un équilibre entre l'internalisation favorisée par les FcγRs et la protection de la dégradation permise par le FcRn. Cependant, les MoDC (Cellules dendritiques dérivées des monocytes) utilisées dans ce test n'expriment pas l'un d'entre eux, le FcγRI (récepteur de haute affinité) <sup>92</sup>, ce qui pourrait conduire à une moindre internalisation des anticorps par rapport à ce qui se passerait dans des cellules dendritiques conventionnelles.

La faible réponse obtenue pour les anticorps thérapeutiques est également fortement susceptible d'être liée au faible nombre d'épitopes T contenus dans leur séquence. En effet, pour les anticorps avec une région constante humaine, les épitopes T sont localisés dans la partie variable, voir même seulement autour des CDR pour les anticorps humains <sup>180,266,267</sup>, ce qui réduit considérablement le nombre d'épitopes T comparé à des anticorps murins ou à une protéine non humaine comme la KLH. Un faible nombre d'épitopes T pourrait être reconnu par un faible nombre de lymphocytes T spécifiques circulants dans le sang des donneurs sains, ce qui aboutit à une faible fréquence de précurseurs mesurée dans le test. Il est important de noter que même si les fréquences de lymphocytes T CD4+ spécifiques d'un anticorps thérapeutique immunogène sont très faibles (au sens du seuil de détection du test), il existe bien une réponse immunitaire liée à ce faible nombre de lymphocytes spécifiques.

Ainsi, pour augmenter le signal obtenu dans ce test, il pourrait être envisagé d'augmenter le nombre de lymphocytes T CD4+ testés par donneur et par anticorps afin d'abaisser le seuil de

détection. Une autre stratégie complémentaire pourrait être d'essayer de favoriser la dégradation des anticorps par les cellules dendritiques en ne testant que la partie Fab, bien que dans ce cas il n'y aurait plus de liaison au récepteur FcγRIIb.

En plus des questions de sensibilité, le principal inconvénient des tests d'évaluation *in vitro* de l'immunogénicité est leur faible débit. Une piste d'amélioration possible pourrait être d'utiliser une approche NGS pour évaluer la fréquence de précurseurs T CD4<sup>+</sup> en suivant l'enrichissement de chaque lymphocyte au cours du test en séquençant leurs TCR. Le séquençage des TCR est un domaine en plein développement grâce aux progrès effectués ces dernières années sur le séquençage haut débit <sup>268</sup>. Les expériences de séquençage des TCR sont également facilitées grâce à l'apparition de plateformes d'analyse des données NGS spécialisées comme MiTCR <sup>269</sup>. L'utilisation des technologies de séquençage à haut débit permettrait d'augmenter à la fois le nombre de lymphocytes T CD4<sup>+</sup> testés ainsi que le nombre de donneurs testés tout en conservant la même durée d'expérience, ce qui permettrait d'améliorer le débit et la robustesse du test. Cette piste d'amélioration est actuellement en cours d'évaluation au sein du laboratoire.

## Perspectives des applications de ce travail dans le monde industriel

Le développement d'un anticorps thérapeutique passe par différentes étapes comme la génération et le choix de l'anticorps ainsi que son optimisation. Des paramètres comme l'affinité, la stabilité ou la spécificité sont déjà largement pris en compte dans le choix de l'anticorps et/ou son optimisation. En revanche, l'immunogénicité reste encore à l'heure actuelle une problématique importante. L'humanisation ou le « caractère » humain d'un anticorps (dans le cas d'un anticorps « *fully human* ») sont utilisés de façon relativement empirique pour justifier la prise en compte du problème de l'immunogénicité au cours du développement d'un anticorps. Or, comme nous l'avons vu, le degré d'humanisation d'un anticorps n'est pas complètement corrélé à son degré d'immunogénicité.

La méthode de dé-immunisation mise en place dans ce travail de thèse pourrait permettre d'améliorer le processus d'optimisation des anticorps thérapeutiques en intégrant une étape de suppression des épitopes T – réels ou prédits - de l'anticorps. Dans cette optique, ce projet de thèse se poursuivra dans le cadre d'un projet d'essaimage en 2020. Ce projet aura pour objectif d'une part de consolider les résultats sur le processus de dé-immunisation et d'autre part d'évaluer la pertinence de la proposition d'un service de dé-immunisation et/ou de caractérisation des anticorps par DMS au sein de l'écosystème du développement des anticorps thérapeutiques.

## Liste des abréviations

**ADA** : anticorps dirigé contre la protéine thérapeutique (*Anti-Drug Antibody*)  
**ADCC** : réponse cytotoxique dépendante des anticorps (*Antibody Dependant Cellular Cytotoxicity*)  
**ADN** : Acide DésoxyriboNucléique  
**AIRE** : AutoImmune REgulator  
**ARN** : Acide RiboNucléique  
**BCR** : récepteur des cellules B (*B-cell Receptor*)  
**CCR5** : Protéine de surface des leucocytes impliquée dans l'immunité (C-C chimiokine de type 5)  
**CD20** : Protéine jouant un rôle dans la différenciation des lymphocytes B (Cluster de différenciation 20)  
**CD28** : Protéine exprimée par les lymphocytes T permettant le signal de co-stimulation (Cluster de différenciation 28)  
**CD3** : complexe protéique membranaire exprimé par les lymphocytes T (*Cluster de Différenciation 3*)  
**CD3** : Complexe protéique impliqué dans l'activation des lymphocytes T par l'activation du TCR (Cluster de différenciation 3)  
**CD4** : Glycoprotéine exprimée à la surface des Lymphocytes T CD4+ (Cluster de Différenciation 4)  
**CD52** : CAMPATH-1 antigène. Glycoprotéine présente à la surface des lymphocytes matures. (Cluster de différenciation 52)  
**CDC** : Center for Disease Control and Prevention  
**CDR** : région hypervariable d'un anticorps responsable de la liaison spécifique à un antigène (*Complementary Determining Region*)  
**CEA** : Commissariat à l'énergie atomique et aux énergies alternatives  
**CHO** : Chinese Hamster Ovary  
**CLIP** : peptide invariant présent dans le sillon de la molécule CMH II lors de son assemblage (*Class II-associated Invariant Chain Peptide*)  
**CLP** : Common Lymphocyte Progenitor  
**CMH-II** : Complexe Majeur d'Histocompatibilité de type II  
**CPA** : Cellule Présentatrice d'Antigène  
**CXCR4** ou CD184 : Récepteur membranaire de la protéine CXCL12  
**DC** : Cellule dendritique (*Dendritic Cell*)  
**DMS** : Analyse par mutagenèse exhaustive (*Deep Mutational Scanning*)  
**ECLA** : méthode de détection de protéine par électrochimiluminescence (*ElectroChemiLuminescence-Assay*)  
**EF** : Edma Factor  
**ELISA** : méthode immuno-enzymatique de détection des protéines (*Enzyme-Linked Immunosorbent Assay*)  
**ELISPOT** : méthode immuno-enzymatique permettant de dénombrer des cellules à partir de leur sécrétion (*Enzyme-Linked immunospot*)  
**FACS** : tri par cytométrie en flux (*Fluorescence-Assisted Cell Sorting*)  
**Fc** : fragment constant d'un anticorps (*Fragment cristallisable*)  
**FcR** : récepteur au fragment constant des anticorps  
**FcRn** : neonatal Fc Receptor  
**FR** : régions des domaines variables d'anticorps situées entre les CDR (*Framework*)  
**GFP** : Green Fluorescent Protein  
**GM-CSF** : Granulocyte Macrophage Colony Stimulating Factor  
**GPI** : Glycosylphosphatidylinositol  
**HEK** : Human Embryonic Kidney  
**HLA-II** : molécule HLA de classe II  
**HMSA** : Homogeneous Mobility Shift Assay  
**HPV** : Human Papillomavirus  
**HSC** : Human String Content  
**IC50** : concentration inhibant 50% des effets  
**IEDB** : Immune Epitope DataBase  
**IFN-γ** : Interféron gamma, sécrétée par les LT et les NK  
**Ig** : Immunoglobuline  
**Ii** : chaîne invariante  
**IL-2** : interleukine 2, induit la prolifération des LT

**IL-6** : interleukine 6, induit la prolifération et la différenciation des LT et des LB  
**IL-7** : Interleukine 7, impliquée dans la survie des lymphocytes  
**INN** : International Nonproprietary Names  
**ITAM** : Immunoreceptor Tyrosine-Based Activation Motif  
**KLH** : protéine issue de l'hémolymphe du mollusque *Megathura crenulata* (*Keyhole limpet hemocyanin*)  
**LB** : Lymphocyte B  
**LC-MS** : Liquid Chromatography Masse Spectrometry  
**LF** : Lethal Factor  
**MAAPS** : analyse protéomique des peptides naturellement par les molécules HLA II (*Major histocompatibility complex-Associated Peptide Proteomics*)  
**MACS** : tri par bille magnétique (*Magnetic-Activated Cell Sorting*)  
**mTEC** : medullary Thymic Epithelial Cells  
**NADPH** : Nicotinamide adénine dinucléotide  
**NGS** : séquençage haut-débit (*Next Generation Sequencing*)  
**NTS** : Non Thyphoïdal *Salmonella*  
**PA** : Protective Antigen  
**PD-1** : *Programmed cell Death 1*  
**PD-L1** : *Programmed cell Death Ligand 1*  
**PIA** : pH-shift anti-Idiotypic Antigen-binding  
**PNH** : Primate Non Humain  
**RT-PCR** : Reverse Transcriptase PCR  
**SOE-PCR** : Slide Over Extension - PCR  
**SST3** : Système de sécrétion de type 3  
**TCR** : récepteur des cellules T (*T-cell Receptor*)  
**TdT** : enzyme permettant l'ajout de nucléotides lors de la recombinaison des gènes V, D e J des Ig (*Terminal deoxynucleotidyl transferase*)  
**TNF $\alpha$**  : facteur de nécrose tumorale  $\alpha$  (*Tumor Necrosis Factor  $\alpha$* )  
**TNF $\alpha_m$**  : facteur de nécrose tumorale  $\alpha$  membranaire  
**Treg** : Lymphocyte T régulateur  
**TRP1** : Gène impliqué dans la synthèse du tryptophane chez *S.cerevisiae*  
**VH** : Domaine variable de la chaîne lourde  
**VHH** : Anticorps de camélidés à simple domaine  
**Vi CPS** : Vaccin recommandé par la WHO pour la prévention de la typhoïde (Vi Capsular PolySaccharide)  
**VIH/HIV** : Virus de l'Immunodéficience Humaine  
**VL** : Domaine variable de la chaîne légère  
**YSD** : expression à la surface de levure (*Yeast Surface Display*)

# Bibliographie

1. Casadevall, A. Antibody-Based Therapies for Emerging Infectious Diseases. *Emerg. Infect. Dis.* **2**, 200–208 (1996).
2. Behring & Kitasato. Ueber das Zustandekommen der Diphtherie-Immunität und der Tetanus-Immunität bei Thieren. *Dtsch. Medizinische Wochenschrift* (1890) doi:10.1055/s-0029-1207589.
3. Klemperer, G. & Klemperer, F. Versuche über immunisierung und heilung bei der pneumokokkeninfektion. *Berlin Klin Wochenschr* **28**, 833–835 (1891).
4. Hey, A. History and Practice: Antibodies in Infectious Diseases. *Microbiol. Spectr.* **3**, 1–15 (2015).
5. Casadevall, A. & Scharff, M. D. Serum therapy revisited: Animal models of infection and development of passive antibody therapy. *Antimicrob. Agents Chemother.* **38**, 1695–1702 (1994).
6. WHO. Global action plan on antimicrobial resistance. *Who* 1–28 (2017).
7. Maggon, K. Monoclonal antibody ‘gold rush’. *Curr. Med. Chem.* (2007).
8. Qiu, X. *et al.* Reversion of advanced Ebola virus disease in nonhuman primates with ZMapp. *Nature* **514**, 47–53 (2014).
9. PREVAIL II Writing group. A Randomized, Controlled Trial of ZMapp for Ebola Virus Infection. *N. Engl. J. Med.* **375**, 1448–1456 (2016).
10. Wec, A. Z. *et al.* Development of a Human Antibody Cocktail that Deploys Multiple Functions to Confer Pan-Ebolavirus Protection. *Cell Host Microbe* (2019) doi:10.1016/j.chom.2018.12.004.
11. Lu, L. L., Suscovich, T. J., Fortune, S. M. & Alter, G. Beyond binding: Antibody effector functions in infectious diseases. *Nat. Rev. Immunol.* **18**, 46–61 (2018).
12. Little, S. F., Leppla, S. H. & Cora, E. Production and characterization of monoclonal antibodies to the protective antigen component of Bacillus anthracis toxin. *Infect. Immun.* (1988).
13. Maynard, J. A. *et al.* Protection against anthrax toxin by recombinant antibody fragments correlates with antigen affinity. *Nat. Biotechnol.* (2002) doi:10.1038/nbt0602-597.
14. Harvill, E. T. *et al.* Anamnestic protective immunity to Bacillus anthracis is antibody mediated but independent of complement and Fc receptors. *Infect. Immun.* (2008) doi:10.1128/IAI.00647-07.
15. Gautam, R. *et al.* A single injection of anti-HIV-1 antibodies protects against repeated SHIV challenges. *Nature* (2016) doi:10.1038/nature17677.
16. Law, M. *et al.* Broadly neutralizing antibodies protect against hepatitis C virus quasispecies challenge. *Nat. Med.* (2008) doi:10.1038/nm1698.
17. Zhao, H. *et al.* Structural Basis of Zika Virus-Specific Antibody Protection. *Cell* (2016) doi:10.1016/j.cell.2016.07.020.
18. Nybakken, G. E. *et al.* Structural basis of West Nile virus neutralization by a therapeutic antibody. *Nature* (2005) doi:10.1038/nature03956.
19. Irani, V. *et al.* Acquisition of Functional Antibodies That Block the Binding of Erythrocyte-Binding Antigen 175 and Protection Against Plasmodium falciparum Malaria in Children. *Clin. Infect. Dis.* (2015) doi:10.1093/cid/civ525.
20. Ishii, Y. *et al.* Inhibition of nuclear entry of HPV16 pseudovirus-packaged DNA by an anti-HPV16 L2 neutralizing antibody. *Virology* (2010) doi:10.1016/j.virol.2010.07.019.
21. Aiyegbo, M. S. *et al.* Human Rotavirus VP6-Specific Antibodies Mediate Intracellular Neutralization by Binding to a Quaternary Structure in the Transcriptional Pore. *PLoS One* (2013) doi:10.1371/journal.pone.0061101.
22. Edelson, B. T. & Unanue, E. R. Intracellular antibody neutralizes Listeria growth. *Immunity* (2001) doi:10.1016/S1074-7613(01)00139-X.
23. Kaur, R., Surendran, N., Ochs, M. & Pichichero, M. E. Human antibodies to PhtD, PcpA, and ply reduce adherence to human lung epithelial cells and murine nasopharyngeal colonization by Streptococcus pneumoniae. *Infect. Immun.* (2014) doi:10.1128/IAI.02124-14.
24. Lam, H., Kesselly, A., Stegalkina, S., Kleanthous, H. & Yethon, J. A. Antibodies to PhnD inhibit staphylococcal biofilms. *Infect. Immun.* (2014) doi:10.1128/IAI.02168-14.
25. Merle, N. S., Church, S. E., Fremeaux-Bacchi, V. & Roumenina, L. T. Complement system part I - molecular mechanisms of activation and regulation. *Front. Immunol.* **6**, 1–30 (2015).
26. Merle, N. S., Noe, R., Halbwachs-Mecarelli, L., Fremeaux-Bacchi, V. & Roumenina, L. T. Complement system part II: Role in immunity. *Front. Immunol.* **6**, 1–26 (2015).
27. Moore, T. *et al.* Fc receptor regulation of protective immunity against Chlamydia trachomatis. *Immunology* (2002) doi:10.1046/j.0019-2805.2001.01354.x.

28. Lu, L. L. *et al.* A Functional Role for Antibodies in Tuberculosis. *Cell* (2016) doi:10.1016/j.cell.2016.08.072.
29. Gounni, A. S. *et al.* High-affinity IgE receptor on eosinophils is involved in defence against parasites. *Nature* (1994) doi:10.1038/367183a0.
30. Ackerman, M. E. *et al.* Polyfunctional HIV-Specific Antibody Responses Are Associated with Spontaneous HIV Control. *PLoS Pathog.* (2016) doi:10.1371/journal.ppat.1005315.
31. Herrada, A. A., Contreras, F. J., Tobar, J. A., Pacheco, R. & Kalergis, A. M. Immune complex-induced enhancement of bacterial antigen presentation requires Fc Receptor III expression on dendritic cells. *Proc. Natl. Acad. Sci.* (2007) doi:10.1073/pnas.0700999104.
32. Yamamoto, B. J. *et al.* Obiltoximab prevents disseminated Bacillus anthracis infection and improves survival during pre- and postexposure prophylaxis in animal models of inhalational anthrax. *Antimicrob. Agents Chemother.* **60**, 5796–5805 (2016).
33. Bartlett, J. G. Bezlotoxumab — A New Agent for Clostridium difficile Infection. *N. Engl. J. Med.* **376**, 381–382 (2017).
34. Bettiker, R. L., Koren, D. E. & Jacobson, J. M. Ibalizumab. *Curr. Opin. HIV AIDS* **13**, 354–358 (2018).
35. Kirk, M. D. *et al.* World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral Diseases, 2010: A Data Synthesis. *PLoS Med.* (2015) doi:10.1371/journal.pmed.1001921.
36. Dosso, M., Coulibaly, M. & Kadio, A. Place des diarrhées bactériennes dans les pays en développement. *Bull. la Société Pathol. Exot.* **91**, 402–405 (1998).
37. Kotloff, K. L. *et al.* Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet (London, England)* **382**, 209–222 (2013).
38. Kariuki, S., Gordon, M. A., Feasey, N. & Parry, C. M. Antimicrobial resistance and management of invasive Salmonella disease. *Vaccine* (2015) doi:10.1016/j.vaccine.2015.03.102.
39. Abebe, W. *et al.* Prevalence and antibiotic susceptibility patterns of Shigella and Salmonella among children aged below five years with Diarrhoea attending Nigist Eleni Mohammed memorial hospital, South Ethiopia. *BMC Pediatr.* (2018) doi:10.1186/s12887-018-1221-9.
40. CDC. Bioterrorism Agents/Diseases. <https://emergency.cdc.gov/agent/agentlist-category.asp> (2018).
41. Issenhuth-Jeanjean, S. *et al.* Supplement 2008-2010 (no. 48) to the White-Kauffmann-Le Minor scheme. *Res. Microbiol.* (2014) doi:10.1016/j.resmic.2014.07.004.
42. McClelland, M. *et al.* Complete genome sequence of Salmonella enterica serovar Typhimurium LT2. *Nature* (2001) doi:10.1038/35101614.
43. Bula-Rudas, F. J., Rathore, M. H. & Maraqa, N. F. Salmonella Infections in Childhood. *Advances in Pediatrics* (2015) doi:10.1016/j.yapd.2015.04.005.
44. Kotloff, K. L. *et al.* Global burden of Shigella infections: Implications for vaccine development and implementation of control strategies. *Bull. World Health Organ.* (1999).
45. Kahsay, A. G. & Muthupandian, S. A review on Sero diversity and antimicrobial resistance patterns of Shigella species in Africa, Asia and South America, 2001-2014. *BMC Research Notes* (2016) doi:10.1186/s13104-016-2236-7.
46. Marcus, S. L., Brumell, J. H., Pfeifer, C. G. & Finlay, B. B. Salmonella pathogenicity islands: Big virulence in small packages. *Microbes and Infection* (2000) doi:10.1016/S1286-4579(00)00273-2.
47. Ibarra, J. A. & Steele-Mortimer, O. Salmonella - the ultimate insider. Salmonella virulence factors that modulate intracellular survival. *Cellular Microbiology* (2009) doi:10.1111/j.1462-5822.2009.01368.x.
48. Kane, K. A. & Dorman, C. J. VirB-mediated positive feedback control of the virulence gene regulatory cascade of Shigella flexneri. *J. Bacteriol.* (2012) doi:10.1128/JB.00800-12.
49. Schroeder, G. N. & Hilbi, H. Molecular pathogenesis of Shigella spp.: Controlling host cell signaling, invasion, and death by type III secretion. *Clinical Microbiology Reviews* (2008) doi:10.1128/CMR.00032-07.
50. Blocker, A., Komoriya, K. & Aizawa, S.-I. Type III secretion systems and bacterial flagella: Insights into their function from structural similarities. *Proc. Natl. Acad. Sci.* (2003) doi:10.1073/pnas.0535335100.
51. Chatterjee, S., Chaudhury, S., McShan, A. C., Kaur, K. & De Guzman, R. N. Structure and Biophysics of Type III Secretion in Bacteria. *Biochemistry* **52**, 2508–2517 (2013).
52. Burkinshaw, B. J. & Strynadka, N. C. J. Assembly and structure of the T3SS. *Biochim. Biophys. Acta - Mol. Cell Res.* **1843**, 1649–1663 (2014).
53. Veenendaal, A. K. J. *et al.* The type III secretion system needle tip complex mediates host cell sensing and translocon insertion. *Mol. Microbiol.* **63**, 1719–1730 (2007).
54. Johnson, S. *et al.* Self-chaperoning of the type III secretion system needle tip proteins IpaD and BipD. *J. Biol. Chem.* **282**, 4035–4044 (2007).
55. Blocker, A. J. *et al.* What's the point of the type III secretion system needle? *Proc. Natl. Acad. Sci. U. S. A.* **105**, 6507–6513 (2008).
56. Lunelli, M., Hurwitz, R., Lambers, J. & Kolbe, M. Crystal structure of prgi-sipd: Insight into a secretion competent state of the

- type three secretion system needle tip and its interaction with host ligands. *PLoS Pathog.* **7**, (2011).
57. Epler, C. R., Dickenson, N. E., Bullitt, E. & Picking, W. L. Ultrastructural Analysis of IpaD at the Tip of the Nascent MxiH Type III Secretion Apparatus of *Shigella flexneri*. *J. Mol. Biol.* **420**, 29–39 (2012).
  58. Bernard, A. R., Jessop, T. C., Kumar, P. & Dickenson, N. E. Deoxycholate-Enhanced *Shigella* Virulence Is Regulated by a Rare  $\pi$ -Helix in the Type Three Secretion System Tip Protein IpaD. *Biochemistry* (2017) doi:10.1021/acs.biochem.7b00836.
  59. Barta, M. L. *et al.* Single-domain antibodies pinpoint potential targets within *Shigella* invasion plasmid antigen D of the needle tip complex for inhibition of type III secretion. *J. Biol. Chem.* **292**, 16677–16687 (2017).
  60. Rathinavelan, T. *et al.* NMR model of PrgI-SipD interaction and its implications in the needle-tip assembly of the salmonella type III secretion system. *J. Mol. Biol.* **426**, 2958–2969 (2014).
  61. Dickenson, N. E. *et al.* N-terminus of IpaB provides a potential anchor to the shigella type III secretion system tip complex protein IpaD. *Biochemistry* **52**, 8790–8799 (2013).
  62. Barta, M. L. *et al.* Identification of the bile salt binding site on IpaD from *Shigella flexneri* and the influence of ligand binding on IpaD structure. *Proteins* **80**, 935–45 (2012).
  63. Chatterjee, S. *et al.* The crystal structures of the Salmonella type III secretion system tip protein SipD in complex with deoxycholate and chenodeoxycholate. *Protein Sci.* **20**, 75–86 (2011).
  64. McShan, A. C., Anbanandam, A., Patnaik, S. & De Guzman, R. N. Characterization of the Binding of Hydroxyindole, Indoleacetic acid, and Morpholinoaniline to the Salmonella Type III Secretion System Proteins SipD and SipB. *ChemMedChem* **11**, 963–971 (2016).
  65. Epler, C. R., Dickenson, N. E., Olive, A. J., Picking, W. L. & Picking, W. D. Liposomes recruit IpaC to the *Shigella flexneri* type III secretion apparatus needle as a final step in secretion induction. *Infect. Immun.* **77**, 2754–2761 (2009).
  66. Cheung, M. *et al.* Three-dimensional electron microscopy reconstruction and cysteine-mediated crosslinking provide a model of the type III secretion system needle tip complex. *Mol. Microbiol.* **95**, 31–50 (2015).
  67. Barta, M. L. *et al.* Using disruptive insertional mutagenesis to identify the in situ structure-function landscape of the *Shigella* translocator protein IpaB. *Protein Sci.* **27**, 1392–1406 (2018).
  68. Fraser, A., Paul, M., Goldberg, E., Acosta, C. J. & Leibovici, L. Typhoid fever vaccines: Systematic review and meta-analysis of randomised controlled trials. *Vaccine* (2007) doi:10.1016/j.vaccine.2007.08.027.
  69. Khan, M. I., Ochiai, R. L. & Clemens, J. D. Population impact of Vi capsular polysaccharide vaccine. *Expert Review of Vaccines* (2010) doi:10.1586/erv.10.43.
  70. Salazar-González, R. M. *et al.* Induction of cellular immune response and anti-Salmonella enterica serovar typhi bactericidal antibodies in healthy volunteers by immunization with a vaccine candidate against typhoid fever. *Immunol. Lett.* (2004) doi:10.1016/j.imlet.2004.01.010.
  71. Toobak, H. *et al.* Immune response variations to Salmonella enterica serovar Typhi recombinant porin proteins in mice. *Biologicals* (2013) doi:10.1016/j.biologicals.2013.05.005.
  72. Mani, S., Wierzbza, T. & Walker, R. I. Status of vaccine research and development for *Shigella* prepared for WHO PD-VAC. *Vaccine* (2016) doi:10.1016/j.vaccine.2016.02.075.
  73. Cornelis, G. R. The type III secretion injectisome. *Nat. Rev. Microbiol.* **4**, 811–825 (2006).
  74. Espina, M. *et al.* IpaD localizes to the tip of the type III secretion system needle of *Shigella flexneri*. *Infect. Immun.* **74**, 4391–4400 (2006).
  75. Goure, J., Broz, P., Attree, O., Cornelis, G. R. & Attree, I. Protective Anti-V Antibodies Inhibit Pseudomonas and Yersinia Translocon Assembly within Host Membranes. *J. Infect. Dis.* (2005) doi:10.1086/430932.
  76. Warrenner, P. *et al.* A novel anti-PcrV antibody providing enhanced protection against pseudomonas aeruginosa in multiple animal infection models. *Antimicrob. Agents Chemother.* (2014) doi:10.1128/AAC.02643-14.
  77. Jneid, B. Evaluation de l'effet protecteur de protéines du Système de Sécrétion de Type III de bactéries entéropathogènes pour la vaccination et l'immunothérapie. (Paris-Saclay, 2016).
  78. Al-Lazikani, B., Lesk, A. M. & Chothia, C. Standard conformations for the canonical structures of immunoglobulins. *J. Mol. Biol.* (1997) doi:10.1006/jmbi.1997.1354.
  79. Dondelinger, M. *et al.* Understanding the significance and implications of antibody numbering and antigen-binding surface/residue definition. *Front. Immunol.* **9**, 1–15 (2018).
  80. Foote, J. & Winter, G. Antibody framework residues affecting the conformation of the hypervariable loops. *J. Mol. Biol.* **224**, 487–499 (1992).
  81. Schroeder, H. W. & Cavacini, L. Structure and function of immunoglobulins. *J. Allergy Clin. Immunol.* **125**, S41–S52 (2010).
  82. de Marco, A. Recombinant antibody production evolves into multiple options aimed at yielding reagents suitable for application-specific needs. *Microb. Cell Fact.* **14**, 1–17 (2015).
  83. Lee, Y. J. & Jeong, K. J. Challenges to production of antibodies in bacteria and yeast. *J. Biosci. Bioeng.* **120**, 483–490 (2015).



84. Kim, J. Y., Kim, Y. G. & Lee, G. M. CHO cells in biotechnology for production of recombinant proteins: Current state and further potential. *Appl. Microbiol. Biotechnol.* **93**, 917–930 (2012).
85. Maillère, B., Delluc, S. & Ravot, G. La prédiction de l'immunogénicité des protéines thérapeutiques. *médecine/sciences* **28**, 82–88 (2012).
86. Roche, P. A. & Furuta, K. The ins and outs of MHC class II-mediated antigen processing and presentation. *Nat. Rev. Immunol.* **15**, 203–216 (2015).
87. Jenkins, M. K., Chu, H. H., McLachlan, J. B. & Moon, J. J. On the Composition of the Preimmune Repertoire of T Cells Specific for Peptide–Major Histocompatibility Complex Ligands. *Annu. Rev. Immunol.* (2010) doi:10.1146/annurev-immunol-030409-101253.
88. Langenkamp, A., Messi, M., Lanzavecchia, A. & Sallusto, F. Kinetics of dendritic cell activation: impact on priming of TH1, TH2 and nonpolarized T cells. *Nat. Immunol.* **1**, 311–316 (2000).
89. Blum, J. S., Wearsch, P. A. & Cresswell, P. Pathways of Antigen Processing. *Annu. Rev. Immunol.* **31**, 443–473 (2013).
90. Mercer, J. & Helenius, A. Virus entry by macropinocytosis. *Nature Cell Biology* (2009) doi:10.1038/ncb0509-510.
91. Roopenian, D. C. & Akilesh, S. FcRn: the neonatal Fc receptor comes of age. *Nat. Rev. Immunol.* **7**, 715–725 (2007).
92. Guillemins, M., Bruhns, P., Saeys, Y., Hammad, H. & Lambrecht, B. N. The function of Fcγ receptors in dendritic cells and macrophages. *Nat. Rev. Immunol.* **14**, 94–108 (2014).
93. Liu, Y. *et al.* Regulated Expression of Fc R in Human Dendritic Cells Controls Cross-Presentation of Antigen–Antibody Complexes. *J. Immunol.* (2014) doi:10.4049/jimmunol.177.12.8440.
94. Sallusto, F. & Lanzavecchia, A. Efficient presentation of soluble antigen by cultured human dendritic cells is maintained by granulocyte/macrophage colony-stimulating factor plus interleukin 4 and downregulated by tumor necrosis factor α. *J. Exp. Med.* (1994) doi:10.1084/jem.179.4.1109.
95. Bonaccorsi, I. *et al.* Membrane Transfer from Tumor Cells Overcomes Deficient Phagocytic Ability of Plasmacytoid Dendritic Cells for the Acquisition and Presentation of Tumor Antigens. *J. Immunol.* (2014) doi:10.4049/jimmunol.1301039.
96. Humeniuk, P., Dubiela, P. & Hoffmann-Sommergruber, K. Dendritic cells and their role in allergy: Uptake, proteolytic processing and presentation of allergens. *Int. J. Mol. Sci.* **18**, (2017).
97. Manoury, B. *et al.* Asparagine endopeptidase can initiate the removal of the MHC class II invariant chain chaperone. *Immunity* (2003) doi:10.1016/S1074-7613(03)00085-2.
98. Shi, G. P. *et al.* Cathepsin S required for normal MHC class II peptide loading and germinal center development. *Immunity* (1999).
99. Lee, H. K. *et al.* In Vivo Requirement for Atg5 in Antigen Presentation by Dendritic Cells. *Immunity* (2010) doi:10.1016/j.immuni.2009.12.006.
100. Lautwein, A. *et al.* Inflammatory stimuli recruit cathepsin activity to late endosomal compartments in human dendritic cells. *Eur. J. Immunol.* (2002) doi:10.1002/1521-4141(200212)32:12<3348::AID-IMMU3348>3.0.CO;2-S.
101. Trombetta, E. S., Ebersold, M., Garrett, W., Pypaert, M. & Mellman, I. Activation of lysosomal function during dendritic cell maturation. *Science (80- )*. (2003) doi:10.1126/science.1080106.
102. Biniossek, M. L., Nägler, D. K., Becker-Pauly, C. & Schilling, O. Proteomic identification of protease cleavage sites characterizes prime and non-prime specificity of cysteine cathepsins B, L, and S. *J. Proteome Res.* **10**, 5363–5373 (2011).
103. Delamarre, L., Pack, M., Chang, H., Mellman, I. & Trombetta, E. S. Differential lysosomal proteolysis in antigen-presenting cells determines antigen fate. *Science (80- )*. (2005) doi:10.1126/science.1108003.
104. Manoury, B. *et al.* Destructive processing by asparagine endopeptidase limits presentation of a dominant T cell epitope in MBP. *Nat. Immunol.* (2002) doi:10.1038/ni754.
105. Romao, S. *et al.* Autophagy proteins stabilize pathogen-containing phagosomes for prolonged MHC II antigen processing. *J. Cell Biol.* (2013) doi:10.1083/jcb.201308173.
106. Brown, J. H. *et al.* Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* (1993) doi:10.1038/364033a0.
107. O'Brien, C., Flower, D. R. & Feighery, C. Peptide length significantly influences in vitro affinity for MHC class II molecules. *Immunome Res.* **4**, 1–7 (2008).
108. Batalia, M. A. & Collins, E. J. Peptide binding by class I and class II MHC molecules. *Biopolymers* **43**, 281–302 (1997).
109. McCluskey, J., Kanaan, C. & Diviney, M. Nomenclature and serology of HLA class I and class II alleles. *Curr. Protoc. Immunol.* **2017**, A.1S.1–A.1S.6 (2017).
110. Andersson, G. Evolution of the human HLA-DR region. *Front. Biosci.* **3**, d739–745 (2016).
111. Stern, L. J. *et al.* Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* **368**, 215–221 (1994).
112. Cresswell, P. Invariant Chain Structure and MHC Class II Function. *Cell* **84**, 505–507 (1996).

113. McCormick, P. J., Martina, J. A. & Bonifacio, J. S. Involvement of clathrin and AP-2 in the trafficking of MHC class II molecules to antigen-processing compartments. *Proc. Natl. Acad. Sci.* (2005) doi:10.1073/pnas.0502206102.
114. Roche, P. A., Teletski, C. L., Stang, E., Bakke, O. & Long, E. O. Cell surface HLA-DR-invariant chain complexes are targeted to endosomes by rapid internalization. *Proc. Natl. Acad. Sci.* (2006) doi:10.1073/pnas.90.18.8581.
115. Busch, R. *et al.* Achieving stability through editing and chaperoning: Regulation of MHC class II peptide binding and expression. *Immunological Reviews* (2005) doi:10.1111/j.0105-2896.2005.00306.x.
116. Davis, M. M. & Bjorkman, P. J. T-cell antigen receptor genes and T-cell recognition. *Nature* (1988) doi:10.1038/334395a0.
117. Wucherpfennig, K. W., Gagnon, E., Call, M. J., Huseby, E. S. & Call, M. E. Structural biology of the T-cell receptor: insights into receptor assembly, ligand recognition, and initiation of signaling. *Cold Spring Harb. Perspect. Biol.* **2**, 1–14 (2010).
118. Glatzová, D. & Cebecauer, M. Dual Role of CD4 in Peripheral T Lymphocytes. *Front. Immunol.* **10**, 618 (2019).
119. Germain, R. N. & Stefanová, I. THE DYNAMICS OF T CELL RECEPTOR SIGNALING: Complex Orchestration and the Key Roles of Tempo and Cooperation. *Annu. Rev. Immunol.* (2002) doi:10.1146/annurev.immunol.17.1.467.
120. Klein, L., Hinterberger, M., Wirnsberger, G. & Kyewski, B. Antigen presentation in the thymus for positive selection and central tolerance induction. *Nat. Rev. Immunol.* **9**, 833–844 (2009).
121. Anderson, M. S. *et al.* Projection of an immunological self shadow within the thymus by the aire protein. *Science* (80-. ). (2002) doi:10.1126/science.1075958.
122. Peterson, P., Org, T. & Rebane, A. Transcriptional regulation by AIRE: Molecular mechanisms of central tolerance. *Nature Reviews Immunology* (2008) doi:10.1038/nri2450.
123. Bonasio, R. *et al.* Clonal deletion of thymocytes by circulating dendritic cells homing to the thymus. *Nat. Immunol.* (2006) doi:10.1038/ni1385.
124. Kappler, J. W., Roehm, N. & Marrack, P. T cell tolerance by clonal elimination in the thymus. *Cell* (1987) doi:10.1016/0092-8674(87)90568-X.
125. Jordan, M. S. *et al.* Thymic selection of CD4+CD25+ regulatory T cells induced by an agonist self-peptide. *Nat. Immunol.* (2001) doi:10.1038/86302.
126. Leung, C. S. *et al.* Robust T-cell stimulation by Epstein-Barr virus-transformed B cells after antigen targeting to DEC-205. *Blood* (2013) doi:10.1182/blood-2012-08-450775.
127. Melchers, F. Checkpoints that control B cell development. *J. Clin. Invest.* **125**, 2203–2210 (2015).
128. Charles A Janeway, Paul Travers, Mark Walport, and M. J. S. Principles of innate and adaptive immunity - Immunobiology - NCBI Bookshelf. in *Immunobiology: The Immune System in Health and Disease*. (2001).
129. Fearon, D. T., Carroll, M. C. & Carroll, M. C. Regulation of B Lymphocyte Responses to Foreign and Self-Antigens by the CD19/CD21 Complex. *Annu. Rev. Immunol.* **18**, 393–422 (2000).
130. Nencini, F. *et al.* Assays and Strategies for Immunogenicity Assessment of Biological Agents. *Drug Development Research* (2014) doi:10.1002/ddr.21184.
131. Yin, L., Chen, X., Vicini, P., Rup, B. & Hickling, T. P. Therapeutic outcomes, assessments, risk factors and mitigation efforts of immunogenicity of therapeutic protein products. *Cell. Immunol.* **295**, 118–126 (2015).
132. van Schouwenburg, P. A. *et al.* A novel method for the detection of antibodies to adalimumab in the presence of drug reveals 'hidden' immunogenicity in rheumatoid arthritis patients. *J. Immunol. Methods* **362**, 82–88 (2010).
133. Wang, S.-L. *et al.* Development and validation of a homogeneous mobility shift assay for the measurement of infliximab and antibodies-to-infliximab levels in patient serum. *J. Immunol. Methods* **382**, 177–188 (2012).
134. Wang, S. L. *et al.* Development and validation of a homogeneous mobility shift assay for the measurement of infliximab and antibodies-to-infliximab levels in patient serum. *J. Immunol. Methods* (2012) doi:10.1016/j.jim.2012.06.002.
135. Dingman, R. & Balu-Iyer, S. V. Immunogenicity of Protein Pharmaceuticals. *J. Pharm. Sci.* **108**, 1637–1654 (2019).
136. van den Bemt, B. J. *et al.* Anti-infliximab antibodies are already detectable in most patients with rheumatoid arthritis halfway through an infusioncycle: an open-label pharmacokinetic cohort study. *BMC Musculoskelet. Disord.* **12**, 12 (2011).
137. Van Schouwenburg, P. A. *et al.* Adalimumab elicits a restricted anti-idiotypic antibody response in autoimmune patients resulting in functional neutralisation. *Ann. Rheum. Dis.* (2013) doi:10.1136/annrheumdis-2012-201445.
138. Murdaca, G. *et al.* Immunogenicity of infliximab and adalimumab: what is its role in hypersensitivity and modulation of therapeutic efficacy and safety? *Expert Opin. Drug Saf.* **15**, 43–52 (2016).
139. Atzeni, F. *et al.* Autoimmunity and anti-TNF- $\alpha$  agents. in *Annals of the New York Academy of Sciences* (2005). doi:10.1196/annals.1361.100.
140. Chung, C. H. *et al.* Cetuximab-induced anaphylaxis and IgE specific for galactose- $\alpha$ -1,3- galactose. *N. Engl. J. Med.* (2008) doi:10.1056/NEJMoa074943.
141. Vultaggio, A. *et al.* Anti-infliximab IgE and non-IgE antibodies and induction of infusion-related severe anaphylactic reactions. *Allergy Eur. J. Allergy Clin. Immunol.* **65**, 657–661 (2010).

142. Suntharalingam, G. *et al.* Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN1412. *N. Engl. J. Med.* (2006) doi:10.1056/NEJMoa063842.
143. Wing, M. G. *et al.* Mechanism of first-dose cytokine-release syndrome by CAMPATH 1-H: Involvement of CD16 (FcγRIII) and CD11a/CD18 (LFA-1) on NK cells. *J. Clin. Invest.* (1996) doi:10.1172/JCI119110.
144. Chatenoud, L. *et al.* In vivo cell activation following OKT3 administration: Systemic cytokine release and modulation by corticosteroids. *Transplantation* (1990) doi:10.1097/00007890-199004000-00009.
145. Winkler, U. *et al.* Cytokine-release syndrome in patients with B-cell chronic lymphocytic leukemia and high lymphocyte counts after treatment with an anti-CD20 monoclonal antibody (rituximab, IDEC-C2B8). *Blood* (1999).
146. Verthelyi, D. & Wang, V. Trace levels of innate immune response modulating impurities (IIRMI)s synergize to break tolerance to therapeutic proteins. *PLoS One* (2010) doi:10.1371/journal.pone.0015252.
147. Fleischman, M. L., Chung, J., Paul, E. P. & Lewus, R. A. Shipping-Induced Aggregation in Therapeutic Antibodies: Utilization of a Scale-Down Model to Assess Degradation in Monoclonal Antibodies. *J. Pharm. Sci.* (2017) doi:10.1016/j.xphs.2016.11.021.
148. Schellekens, H. Bioequivalence and the immunogenicity of biopharmaceuticals. *Nature Reviews Drug Discovery* (2002).
149. Iwai, L. K. *et al.* In Silico Prediction of Peptides Binding to Multiple HLA-DR Molecules Accurately Identifies Immunodominant Epitopes from gp43 of Paracoccidioides brasiliensis Frequently Recognized in Primary Peripheral Blood Mononuclear Cell Responses from Sensitized Ind. *Mol. Med.* (2003) doi:10.1007/bf03402131.
150. Mustafa, A. S. & Shaban, F. A. ProPred analysis and experimental evaluation of promiscuous T-cell epitopes of three major secreted antigens of Mycobacterium tuberculosis. *Tuberculosis* (2006) doi:10.1016/j.tube.2005.05.001.
151. Al-Attayah, R. & Mustafa, A. S. Computer-Assisted Prediction of HLA-DR Binding and Experimental Analysis for Human Promiscuous Th1-Cell Peptides in the 24 kDa Secreted Lipoprotein (LppX) of Mycobacterium tuberculosis. *Scand. J. Immunol.* (2004) doi:10.1111/j.0300-9475.2004.01349.x.
152. Nielsen, M., Lund, O., Buus, S. & Lundegaard, C. MHC Class II epitope predictive algorithms. *Immunology* **130**, 319–328 (2010).
153. IEDB. <https://www.iedb.org/> (2019).
154. Rothbard, J. B. & Busch, R. Binding of Biotinylated Peptides to MHC Class II Proteins on Cell Surfaces. in *Current Protocols in Immunology* (2004). doi:10.1002/0471142735.im1801s25.
155. Dhanda, S. K. *et al.* Predicting HLA CD4 immunogenicity in human populations. *Front. Immunol.* **9**, 1–14 (2018).
156. Paul, S. *et al.* Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes. *J. Immunol. Methods* **422**, 28–34 (2015).
157. Jensen, K. K. *et al.* Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* **154**, 394–406 (2018).
158. Chicz, R. M. *et al.* Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size. *Nature* (1992) doi:10.1038/358764a0.
159. Lin, H. H., Zhang, G. L., Tongchusak, S., Reinherz, E. L. & Brusci, V. Evaluation of MHC-II peptide binding prediction servers: Applications for vaccine research. in *BMC Bioinformatics* (2008). doi:10.1186/1471-2105-9-S1-S22.
160. Sturniolo, T. *et al.* Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat. Biotechnol.* **17**, 555–561 (1999).
161. Nielsen, M. & Lund, O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics* (2009) doi:10.1186/1471-2105-10-296.
162. Holland, C. J., Cole, D. K. & Godkin, A. Re-directing CD4 + T cell responses with the flanking residues of MHC class II-bound peptides: The core is not enough. *Frontiers in Immunology* (2013) doi:10.3389/fimmu.2013.00172.
163. Zhang, L. *et al.* Tepitopepan: Extending tepitope for peptide binding prediction covering over 700 hla-dr molecules. *PLoS One* (2012) doi:10.1371/journal.pone.0030483.
164. Greenbaum, J. *et al.* Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. *Immunogenetics* **63**, 325–335 (2011).
165. McKinney, D. M. *et al.* A strategy to determine HLA class II restriction broadly covering the DR, DP, and DQ allelic variants most commonly expressed in the general population. *Immunogenetics* **65**, 357–370 (2013).
166. Tangri, S. *et al.* Rationally Engineered Therapeutic Proteins with Reduced Immunogenicity. *J. Immunol.* **174**, 3187–3196 (2005).
167. Paul, S. *et al.* Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes. *J. Immunol. Methods* **422**, 28–34 (2015).
168. Dhanda, S. K. *et al.* Development of a strategy and computational application to select candidate protein analogues with reduced HLA binding and immunogenicity. *Immunology* **153**, 118–132 (2018).
169. Paul, S. *et al.* Evaluating the Immunogenicity of Protein Drugs by Applying In Vitro MHC Binding Data and the Immune Epitope Database and Analysis Resource. *Journal of Immunology Research* (2013).

170. Sekiguchi, N. *et al.* MHC-associated peptide proteomics enabling highly sensitive detection of immunogenic sequences for the development of therapeutic antibodies with low immunogenicity. *MAbs* **10**, 1168–1181 (2018).
171. Röhn, T. A. *et al.* A novel strategy for the discovery of MHC class II-restricted tumor antigens: Identification of a melanotransferrin helper T-cell epitope. *Cancer Res.* (2005) doi:10.1158/0008-5472.CAN-05-1973.
172. Mutschlechner, S. *et al.* Naturally processed T cell-activating peptides of the major birch pollen allergen. *J. Allergy Clin. Immunol.* (2010) doi:10.1016/j.jaci.2009.10.052.
173. Safdari, Y., Farajnia, S., Asgharzadeh, M. & Khalili, M. Antibody humanization methods - A review and update. *Biotechnol. Genet. Eng. Rev.* **29**, 175–186 (2013).
174. Morrison, S. L., Johnstont, M. J., Herzenberg, L. A. & Oi, V. T. Chimeric human antibody molecules: Mouse antigen-binding domains with human constant region domains (transfection/protoplast fusion/calcium phosphate transfection/intronic controlling elements/transfectoma). *Proc. Natl. Acad. Sci. USA* vol. 81 (1984).
175. Hwang, W. Y. K. & Foote, J. Immunogenicity of engineered antibodies. *Methods* **36**, 3–10 (2005).
176. Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. & Winter, G. Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature* **321**, 522–525 (1986).
177. Padlan, E. A. A possible procedure for reducing the immunogenicity of antibody variable domains while preserving their ligand-binding properties. *Mol. Immunol.* (1991) doi:10.1016/0161-5890(91)90163-E.
178. Tan, P. *et al.* “Superhumanized” Antibodies: Reduction of Immunogenic Potential by Complementarity-Determining Region Grafting with Human Germline Sequences: Application to an Anti-CD28. *J. Immunol.* (2002) doi:10.4049/jimmunol.169.2.1119.
179. Lazar, G. A., Desjarlais, J. R., Jacinto, J., Karki, S. & Hammond, P. W. A molecular immunology approach to antibody humanization and functional optimization. *Mol. Immunol.* **44**, 1986–1998 (2007).
180. Harding, F. A., Stickler, M. M., Razo, J. & DuBridge, R. B. The immunogenicity of humanized and fully human antibodies: Residual immunogenicity resides in the CDR regions. *MAbs* **2**, 256–265 (2010).
181. Marks, J. D. *et al.* By-passing immunization. Human antibodies from V-gene libraries displayed on phage. *J. Mol. Biol.* (1991) doi:10.1016/0022-2836(91)90498-U.
182. Fishwild, D. M. *et al.* High-avidity human IgGk monoclonal antibodies from a novel strain of minilocus transgenic mice. *Nat. Biotechnol.* (1996) doi:10.1038/nbt0796-845.
183. E.A. Kabat, T.T. Wu, H. Bilofsky, M. Reid-Miller, H. P. *Sequence of Proteins of Immunological Interest.* (National Institutes of Health, 1983).
184. Lefranc, M. P. *et al.* IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev. Comp. Immunol.* (2003) doi:10.1016/S0145-305X(02)00039-3.
185. Kashmiri, S. V. S., De Pascalis, R., Gonzales, N. R. & Schlom, J. SDR grafting--a new approach to antibody humanization. *Methods* **36**, 25–34 (2005).
186. Almagro, J. C. Identification of differences in the specificity-determining residues of antibodies that recognize antigens of different size: Implications for the rational design of antibody repertoires. *J. Mol. Recognit.* (2004) doi:10.1002/jmr.659.
187. Almagro, J. C. & Fransson, J. Humanization of antibodies. *Front. Biosci.* **13**, 1619–1633 (2008).
188. Haidar, J. N. *et al.* A universal combinatorial design of antibody framework to graft distinct CDR sequences: A bioinformatics approach. *Proteins Struct. Funct. Bioinforma.* **80**, 896–912 (2012).
189. Pedersen, J. T. *et al.* Comparison of surface accessible residues in human and murine immunoglobulin Fv domains. Implication for humanization of murine antibodies. *J. Mol. Biol.* (1994) doi:10.1006/jmbi.1994.1050.
190. Warmerdam, P. A. M. *et al.* Elimination of a human T-cell region in staphylokinase by T-cell screening and computer modeling. *Thromb. Haemost.* **87**, 666–673 (2002).
191. Mazor, R. *et al.* Recombinant immunotoxin for cancer treatment with low immunogenicity by identification and silencing of human T-cell epitopes. *Proc. Natl. Acad. Sci.* **111**, 8571–8576 (2014).
192. Mazor, R. *et al.* Identification and elimination of an immunodominant T-cell epitope in recombinant immunotoxins based on Pseudomonas exotoxin A. *Proc. Natl. Acad. Sci.* **109**, E3597–E3603 (2012).
193. Griswold, K. E. & Bailey-Kellogg, C. Design and engineering of deimmunized biotherapeutics. (2016) doi:10.1016/j.sbi.2016.06.003.
194. Parker, A. S., Zheng, W., Griswold, K. E. & Bailey-Kellogg, C. Optimization algorithms for functional deimmunization of therapeutic proteins. *BMC Bioinformatics* **11**, 1–15 (2010).
195. Choi, Y., Verma, D., Griswold, K. E. & Bailey-Kellogg, C. EpiSweep: Computationally driven reengineering of therapeutic proteins to reduce immunogenicity while maintaining function. in *Methods in Molecular Biology* (2017). doi:10.1007/978-1-4939-6637-0\_20.
196. Osipovitch, D. C. *et al.* Design and analysis of immune-evading enzymes for ADEPT therapy. *Protein Eng. Des. Sel.* **25**, 613–623 (2012).
197. Cantor, J. R. *et al.* Therapeutic enzyme deimmunization by combinatorial T-cell epitope removal using neutral drift. *Proc. Natl.*

- Acad. Sci. U. S. A.* **108**, 1272–1277 (2011).
198. EMA. Guideline on Immunogenicity assessment of therapeutic proteins. (2017) doi:10.1111/j.1559-3584.1956.tb03992.x.
  199. Smith, G. P. Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface. *Science* (80-). (1985) doi:10.1126/science.4001944.
  200. McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* (1990) doi:10.1038/348552a0.
  201. Ledsgaard, L., Kilstrup, M., Karatt-Vellatt, A., McCafferty, J. & Laustsen, A. H. Basics of antibody phage display technology. *Toxins (Basel)*. **10**, (2018).
  202. van Wezenbeek, P. M. G. F., Hulsebos, T. J. M. & Schoenmakers, J. G. G. Nucleotide sequence of the filamentous bacteriophage M13 DNA genome: comparison with phage fd. *Gene* (1980) doi:10.1016/0378-1119(80)90093-1.
  203. Hoogenboom, H. R. *et al.* Antibody phage display technology and its applications. *Immunotechnology* **4**, 1–20 (1998).
  204. Wilson, P. C. & Andrews, S. F. Tools to therapeutically harness the human antibody response. *Nature Reviews Immunology* (2012) doi:10.1038/nri3285.
  205. Schofield, D. J. *et al.* Application of phage display to high throughput antibody generation and characterization. *Genome Biol.* **8**, (2007).
  206. Frenzel, A., Schirrmann, T. & Hust, M. Phage display-derived human antibodies in clinical development and therapy. *MAbs* **8**, 1177–1194 (2016).
  207. Freudl, R., MacIntyre, S., Degen, M. & Henning, U. Cell surface exposure of the outer membrane protein OmpA of Escherichia coli K-12. *J. Mol. Biol.* (1986) doi:10.1016/0022-2836(86)90171-3.
  208. Francisco, J. A., Campbell, R., Iverson, B. L. & Georgiou, G. Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface. *Proc. Natl. Acad. Sci. U. S. A.* (1993) doi:10.1073/pnas.90.22.10444.
  209. Daugherty, P. S., Chen, G., Olsen, M. J., Iverson, B. L. & Georgiou, G. Antibody affinity maturation using bacterial surface display. *Protein Eng.* (1998).
  210. Ki, J. J., Min, J. S., Iverson, B. L. & Georgiou, G. APEX 2-hybrid, a quantitative protein-protein interaction assay for antibody discovery and engineering. *Proc. Natl. Acad. Sci. U. S. A.* (2007) doi:10.1073/pnas.0702650104.
  211. Mazor, Y., Blarcom, T. Van, Mabry, R., Iverson, B. L. & Georgiou, G. Isolation of engineered, full-length antibodies from libraries expressed in Escherichia coli. *Nat. Biotechnol.* **25**, 563–565 (2007).
  212. Löfblom, J. Bacterial display in combinatorial protein engineering. *Biotechnol. J.* (2011) doi:10.1002/biot.201100129.
  213. Ho, M., Nagata, S. & Pastan, I. Isolation of anti-CD22 Fv with high affinity by Fv display on human cells. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 9637–9642 (2006).
  214. Bowers, P. M. *et al.* Mammalian cell display for the discovery and optimization of antibody therapeutics. *Methods* **65**, 44–56 (2014).
  215. Mason, D. M. *et al.* High-throughput antibody engineering in mammalian cells by CRISPR / Cas9-mediated homology-directed mutagenesis. **46**, 7436–7449 (2018).
  216. Tsuruta, L. R., Dos, M. L. & Moro, A. M. Display Technologies for the Selection of Monoclonal Antibodies for Clinical Use. in *Antibody Engineering* vol. i 13 (InTech, 2018).
  217. Fukuda, I. *et al.* In vitro evolution of single-chain antibodies using mRNA display. *Nucleic Acids Res.* (2006) doi:10.1093/nar/gkl618.
  218. Sumida, T., Yanagawa, H. & Doi, N. In vitro selection of fab fragments by mRNA display and gene-linking emulsion PCR. *J. Nucleic Acids* (2012) doi:10.1155/2012/371379.
  219. Schreuder, M. P., Brekelmans, S., Van Den Ende, H. & Klis, F. M. Targeting of a heterologous protein to the cell wall of *Saccharomyces cerevisiae*. *Yeast* (1993) doi:10.1002/yea.320090410.
  220. Pepper, L. R., Cho, Y. K., Boder, E. T. & Shusta, E. V. A decade of yeast surface display technology: where are we now? *Comb. Chem. High Throughput Screen.* **11**, 127–34 (2008).
  221. Nakamura, Y., Ishii, J. & Kondo, A. Bright fluorescence monitoring system utilizing zoanthus sp. green fluorescent protein (ZsGreen) for human g-protein-coupled receptor signaling in microbial yeast cells. *PLoS One* (2013) doi:10.1371/journal.pone.0082237.
  222. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* (1997).
  223. Gera, N., Hussain, M. & Rao, B. M. Protein selection using yeast surface display. *Methods* **60**, 15–26 (2013).
  224. Boder, E. T., Raeeszadeh-Sarmazdeh, M. & Price, J. V. Engineering antibodies by yeast display. *Arch. Biochem. Biophys.* **526**, 99–106 (2012).
  225. Kieke, M. C., Cho, B. K., Boder, E. T., Kranz, D. M. & Wittrup, K. D. Isolation of anti-T cell receptor scFv mutants by yeast surface display. *Protein Eng.* (1997).

226. Rakestraw, J. A., Aird, D., Aha, P. M., Baynes, B. M. & Lipovšek, D. Secretion-and-capture cell-surface display for selection of target-binding proteins. *Protein Eng. Des. Sel.* (2011) doi:10.1093/protein/gzr008.
227. Steinwand, M. *et al.* The influence of antibody fragment format on phage display based affinity maturation of IgG. *MAbs* **6**, 204–218 (2014).
228. Kayushin, A. *et al.* A new approach to the synthesis of trinucleotide phosphoramidites - Synthons for the generation of oligonucleotide/peptide libraries. in *Nucleosides and Nucleotides* (1999). doi:10.1080/07328319908044778.
229. Pâques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* (1999).
230. Chino, A., Watanabe, K. & Moriya, H. Plasmid construction using recombination activity in the fission yeast *Schizosaccharomyces pombe*. *PLoS One* (2010) doi:10.1371/journal.pone.0009652.
231. van Leeuwen, J., Andrews, B., Boone, C. & Tan, G. Rapid and Efficient Plasmid Construction by Homologous Recombination in Yeast. *Cold Spring Harb. Protoc.* **2015**, 853–862 (2015).
232. Baek, D. S. & Kim, Y. S. Construction of a large synthetic human Fab antibody library on yeast cell surface by optimized yeast mating. *J. Microbiol. Biotechnol.* (2014) doi:10.4014/jmb.1401.01002.
233. Weaver-Feldhaus, J. M. *et al.* Yeast mating for combinatorial Fab library generation and surface display. *FEBS Lett.* (2004) doi:10.1016/S0014-5793(04)00309-6.
234. Blaise, L. *et al.* Construction and diversification of yeast cell surface displayed libraries by yeast mating: Application to the affinity maturation of Fab antibody fragments. *Gene* (2004) doi:10.1016/j.gene.2004.08.014.
235. Benatuil, L., Perez, J. M., Belk, J. & Hsieh, C. M. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng. Des. Sel.* **23**, 155–159 (2010).
236. Hunter, S. A. & Cochran, J. R. Cell-Binding Assays for Determining the Affinity of Protein–Protein Interactions. in *Methods in Enzymology* (ed. Pecoraro, V. L.) vol. 580 21–44 (Academic Press, 2016).
237. Feldhaus, M. J. *et al.* Flow-cytometric isolation of human antibodies from a nonimmune *Saccharomyces cerevisiae* surface display library. *Nat. Biotechnol.* **21**, 163–170 (2003).
238. Boder, E. T., Midelfort, K. S. & Wittrup, K. D. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 10701–10705 (2000).
239. Leuenberger, P. *et al.* Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science* (80- ). (2017) doi:10.1126/science.aai7825.
240. Ellgaard, L. & Helenius, A. Quality control in the endoplasmic reticulum. *Nature Reviews Molecular Cell Biology* (2003) doi:10.1038/nrm1052.
241. Shusta, E. V., Kieke, M. C., Parke, E., Kranz, D. M. & Wittrup, K. D. Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency. *J. Mol. Biol.* (1999) doi:10.1006/jmbi.1999.3130.
242. Park, S. *et al.* Limitations of yeast surface display in engineering proteins of high thermostability. *Protein Eng. Des. Sel.* (2006) doi:10.1093/protein/gzl003.
243. Traxlmayr, M. W. & Obinger, C. Directed evolution of proteins for increased stability and expression using yeast display. *Arch. Biochem. Biophys.* **526**, 174–180 (2012).
244. Xu, Y. *et al.* Addressing polyspecificity of antibodies selected from an in vitro yeast presentation system: a FACS-based, high-throughput selection and analytical tool. *Protein Eng. Des. Sel.* **26**, 663–670 (2013).
245. Puri, V., Streaker, E., Prabakaran, P., Zhu, Z. & Dimitrov, D. S. Highly efficient selection of epitope specific antibody through competitive yeast display library sorting. *MAbs* **5**, 533–539 (2013).
246. van Dijk, E. L., Auger, H., Jaszczyszyn, Y. & Thermes, C. Ten years of next-generation sequencing technology. *Trends Genet.* **30**, 418–426 (2014).
247. Ravn, U. *et al.* Deep sequencing of phage display libraries to support antibody discovery. *Methods* (2013) doi:10.1016/j.jymeth.2013.03.001.
248. Wrenbeck, E. E., Faber, M. S. & Whitehead, T. A. Deep sequencing methods for protein engineering and design. *Curr. Opin. Struct. Biol.* **45**, 36–44 (2017).
249. Firnberg, E. & Ostermeier, M. PFunkel: Efficient, Expansive, User-Defined Mutagenesis. *PLoS One* (2012) doi:10.1371/journal.pone.0052031.
250. Wrenbeck, E. E. *et al.* Plasmid-based one-pot saturation mutagenesis. *Nat. Methods* **13**, 928–930 (2016).
251. Van Blarcom, T. *et al.* Precise and efficient antibody epitope determination through library design, yeast display and next-generation sequencing. *J. Mol. Biol.* **427**, 1513–1534 (2015).
252. Kowalsky, C. A. *et al.* Rapid Fine Conformational Epitope Mapping Using Comprehensive Mutagenesis and Deep Sequencing \* □ S. (2015) doi:10.1074/jbc.M115.676635.
253. Doolan, K. M. & Colby, D. W. Conformation-dependent epitopes recognized by prion protein antibodies probed using mutational scanning and deep sequencing. *J. Mol. Biol.* (2015) doi:10.1016/j.jmb.2014.10.024.

254. Forsyth, C. M. *et al.* Deep mutational scanning of an antibody against epidermal growth factor receptor using mammalian cell display and massively parallel pyrosequencing. *MAbs* **5**, 523–532 (2013).
255. Koenig, P. *et al.* Deep sequencing-guided design of a high affinity dual specificity antibody to target two angiogenic factors in neovascular age-related macular degeneration. *J. Biol. Chem.* **290**, 21773–21786 (2015).
256. Kowalsky, C. A. *et al.* High-resolution sequence-function mapping of full-length proteins. *PLoS One* **10**, 1–23 (2015).
257. Blankenberg, D. *et al.* Manipulation of FASTQ data with galaxy. *Bioinformatics* **26**, 1783–1785 (2010).
258. Tiller, T. *et al.* Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning. *J. Immunol. Methods* (2008) doi:10.1016/j.jim.2007.09.017.
259. Schröter, C. *et al.* A generic approach to engineer antibody pH-switches using combinatorial histidine scanning libraries and yeast display. *MAbs* **7**, 138–151 (2015).
260. Walker, L. M., Bowley, D. R. & Burton, D. R. Efficient Recovery of High-Affinity Antibodies from a Single-Chain Fab Yeast Display Library. *J. Mol. Biol.* **389**, 365–375 (2009).
261. Roth, L. *et al.* Facile generation of antibody heavy and light chain diversities for yeast surface display by Golden Gate Cloning. *Biological Chemistry* (2018) doi:10.1515/hsz-2018-0347.
262. Kowalsky, C. A. & Whitehead, T. A. Determination of binding affinity upon mutation for type I dockerin–cohesin complexes from *Clostridium thermocellum* and *Clostridium cellulolyticum* using deep sequencing. *Proteins Struct. Funct. Bioinforma.* (2016) doi:10.1002/prot.25175.
263. Oseroff, C. *et al.* Molecular Determinants of T Cell Epitope Recognition to the Common Timothy Grass Allergen. *J. Immunol.* **185**, 943–955 (2010).
264. Delluc, S., Ravot, G. & Maillere, B. Quantitative analysis of the CD4 T-cell repertoire specific to therapeutic antibodies in healthy donors. *FASEB J.* **25**, 2040–2048 (2011).
265. Jawa, V. *et al.* T-cell dependent immunogenicity of protein therapeutics: Preclinical assessment and mitigation. *Clinical Immunology* (2013) doi:10.1016/j.clim.2013.09.006.
266. Hamze, M. *et al.* Characterization of CD4 T Cell Epitopes of Infliximab and Rituximab Identified from Healthy Donors. *Front. Immunol.* **8**, (2017).
267. Meunier, S. *et al.* Impact of human sequences in variable domains of therapeutic antibodies on the location of CD4 T-cell epitopes. *Cell. Mol. Immunol.* 4–6 (2019) doi:10.1038/s41423-019-0304-3.
268. Dziubianau, M. *et al.* TCR repertoire analysis by next generation sequencing allows complex differential diagnosis of T cell-related pathology. *Am. J. Transplant.* (2013) doi:10.1111/ajt.12431.
269. Bolotin, D. A. *et al.* MiTCR: Software for T-cell receptor sequencing data analysis. *Nature Methods* (2013) doi:10.1038/nmeth.2555.



**Titre :** Conception et caractérisation d'anticorps non immunogènes : application à un anticorps anti-Salmonelles/Shigelles

**Mots clés :** Immunogénicité, Dé-immunisation, Epitopes T, *Salmonella*, *Shigella*, Ingénierie des anticorps, Expression à la surface de levures

**Résumé :** Suite à leur injection, les anticorps thérapeutiques sont susceptibles de déclencher une réponse immunitaire dirigée à leur rencontre. L'objectif principal de ce travail est de mettre au point une méthode de dé-immunisation et de l'appliquer à l'anticorps IpaD-318 anti-Salmonelles/Shigelles (développé au CEA) protecteur dans les modèles d'infection murins. Le second objectif de ce travail est de déterminer l'épitope de liaison de l'anticorps IpaD-318 sur ses antigènes IpaD et SipD. La méthode de dé-immunisation consiste à supprimer les épitopes T responsables de l'immunogénicité de l'anticorps tout en conservant sa fonctionnalité. Afin de maîtriser ce problème d'immunogénicité, ce travail de thèse propose d'utiliser le *Yeast Surface Display* (YSD) afin de contrôler la fonctionnalité de l'anticorps tout au long du processus de dé-immunisation. Pour cela, une première étape de comparaison des différents formats d'expression d'anticorps en YSD a permis de définir

le format d'affichage à utiliser pour la suite du projet. La stratégie de dé-immunisation adoptée ensuite est basée sur la suppression des épitopes T. La première étape de la méthode consiste à réaliser une mutagenèse exhaustive (DMS) pour identifier les mutations qui aboutissent à un anticorps fonctionnel. Après humanisation de l'anticorps, des algorithmes de prédiction de l'interaction HLA II/épitope T sont utilisés afin d'identifier les substitutions à effectuer pour réduire l'immunogénicité prédite. Des banques de dé-immunisation sont alors générées, criblées en YSD et analysées par NGS. Les différentes zones dé-immunisées et fonctionnelles sont ensuite assemblées au sein d'un même anticorps. Un variant de l'anticorps IpaD-318 à l'immunogénicité réduite selon les algorithmes de prédiction et fonctionnel *in vitro* a été identifié et caractérisé. D'autre part, l'épitope de l'anticorps a été déterminé par DMS, permettant de proposer des hypothèses sur son mode d'action.

**Title:** Conception and characterisation of non-immunogenic antibodies: application to an anti-*Salmonella/Shigella* antibody.

**Keywords:** Immunogenicity, Deimmunization, T-cell epitopes, *Shigella*, *Salmonella*, Protein Engineering, Yeast Surface Display

**Abstract:** Following their injection, therapeutic antibodies are likely to trigger an immune response directed against them. The main objective of this work is to develop a deimmunization method and apply it to the protective IpaD-318 anti-*Salmonella/Shigella* antibody (developed at CEA) in mouse infection models. The second objective of this work is to determine the binding epitope of the IpaD-318 antibody on its IpaD and SipD antigens. The de-immunization method consists in removing the T epitopes responsible for the immunogenicity of the antibody while maintaining its functionality. In order to address this immunogenicity problem, this thesis proposes to use the Yeast Surface Display (YSD) to control the functionality of the antibody throughout the deimmunization process. To do this, a first step of comparison between the different antibody expression formats in YSD was performed and we were able to define the display format to be used for the rest of the

project. The de-immunization strategy adopted subsequently is based on the removal of T-cell epitopes. The first step of the method is to perform an exhaustive mutagenesis (DMS) to identify mutations that result in a functional antibody. After humanization of the antibody, prediction algorithms were used to identify substitutions that can be performed to reduce the predicted immunogenicity. Deimmunization banks are then generated, screened in YSD and analyzed by NGS. The different de-immunized and functional areas are then assembled into a single antibody. A variant of the antibody IpaD-318 with reduced immunogenicity according to the *in vitro* prediction and functional algorithms has been identified and characterized. On the other hand, the epitope of the antibody was determined by DMS, allowing hypotheses to be proposed on its mode of action.