



UNIVERSITE D'AIX-MARSEILLE

ECOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE (EDSVS) ED 62

UMR 7257 CNRS

AFMB ARCHITECTURE ET FONCTION DES MACROMOLECULES BIOLOGIQUES

INNATE PHARMA

Thèse présentée pour obtenir le grade universitaire de docteur

Discipline : Biologie

Spécialité : Biochimie structurale

Jennifer ROCHE

Etudes structurales de fragments d'anticorps d'intérêt thérapeutique et biotechnologique

Soutenue le 17/10/2017 devant le jury :

Marie-Hélène LE DU - (I2BC, Gif-sur-Yvette)

RAPPORTEUR

Jean-Luc FERRER – (IBS/CEA, Grenoble)

RAPPORTEUR

Julie MENETREY – (I2BC, Gif-sur-Yvette)

EXAMINATEUR

James STURGIS – (LISM/IBSM, Marseille)

EXAMINATEUR

Laurent GAUTHIER – (INNATE PHARMA, Marseille)

EXAMINATEUR

Alain ROUSSEL – (AFMB, Marseille)

DIRECTEUR DE THESE

Remerciements

Je voudrais commencer par remercier la région PACA, l'université AMU et Innate Pharma de m'avoir accordé le financement nécessaire à la réalisation de cette thèse.

Je voudrais aussi remercier le laboratoire AFMB de m'avoir accueillie, depuis mes stages de master jusqu'à la fin de ce projet de thèse. Merci donc à Yves Bourne pour tout ça. Merci à Alain Roussel de m'avoir fait assez confiance pour me proposer ce sujet et à Sonia Longhi de m'avoir poussée à me lancer dans cette aventure.

Ensuite, je voudrais dire merci à tous les membres de l'équipe « Interaction hôte-pathogène », passés et présents. Vous avez été super au quotidien, chacun à votre manière.

Un grand merci à Silvia Spinelli, Gerling Sulzenbacher, Philippe Canteau, Philippe Leone et Véronique Zamboni pour avoir eu la patience de me transmettre votre savoir en cristallogénèse, pêche des cristaux, diffraction par rayons X, collecte de données etc... J'espère vous rendre fiers !

Des remerciements spéciaux à Stéphanie Blangy et Julie Lichière, les wonderwoman du labo que je garderai toujours en mémoire !

Merci à tous ceux qui ont fait partie de mon quotidien, pour vos sourires et votre bonne humeur, nos discussions scientifiques, sérieuses ou farfelues.

Maintenant les choses sérieuses...

Merci mes petites « tata/maman » : Jenny, Amandine et Béatrice. J'ai tellement appris grâce à vous...

Merci à Clémentos, Marionos, Fathios, Reubenos, Matilde, Amal, Gabichou, Yoan, Laurie et Anaïs. Vous êtes tarés (mais vraiment pas équilibrés !) mais je vous aime quand même. Merci aussi à toute la team Danse (désolée, vous êtes beaucoup trop nombreux ^^).

Merci à tous mes correcteurs de dernière minute (Marjorie, Raphaël et Yoan), vous avez été géniaux !!!

Un grand merci à Marjorie, ma petite sœur de cœur, à ma sœur de sang Nathalie et mon frère Anthony.

Merci à toi John, pour cette grande leçon que tu m'as donné et que je garderai toujours en mémoire : ne jamais baisser la tête.

Antoine, j'ai connu avec toi les joies, espérances, déceptions, angoisses de la thèse, du milieu à sa fin. Merci de m'avoir supportée pendant tous ces moments, d'avoir su être un

réel partenaire et encore merci d'avoir accepté de me lire et corriger, malgré les circonstances.

Sonia... Tu sais tout le bien que je pense de toi. Je suis fière d'être une de tes « petites ».

Alain, certes on aurait pu partager plus durant cette thèse en apprenant plus rapidement à se parler. Mais je pense qu'avec cette rédaction on a finalement réussi à développer une complicité qui, je l'espère, perdurera. Je suis très heureuse d'avoir réalisé cette thèse sous ta direction.

Et le plus important ... Merci Polo !!!

À Pierrot, mon deuxième papa, discret mais toujours présent,

À mon père, qui m'a toujours poussée à aller au plus loin dans la connaissance,

À ma mère, j'espère que tu aurais été fière.

Liste des abréviations

ADN : Acide DésoxyriboNucléique
BLI : Bio-Layer Interferometry
CD : Dichroïsme Circulaire
CDR : Région Déterminante de Complémentarité
CH : « Constant Heavy chain domain »
CL : « Constant Light chain domain »
CMH : Complexe Majeur d'Histocompatibilité
 D_{\max} : Distance maximale
DO : Densité Optique
DTT : DiThioTréitol
EG : Ethylène Glycol
Fab : « Fragment Antigen-Binding »
Fc : Fragment constant
FPM : Force Proton Motrice
GM-CSF : « Granulocyte-Macrophage Colony-Stimulating Factor »
HcAbs : « Heavy chain Antibodies »
His-Tag : Etiquette 6xHistidine
 I_0 : Intensité diffusée extrapolée à une valeur de vecteur diffusé de 0
Ig : Immunoglobuline
IPTG : IsoPropyl-beta-D-1-ThioGalactopyranoside
k_{ass} : constante de vitesse d'association
K_D : constante de dissociation
k_{diss} : constante de vitesse de dissociation
LGL : Grands lymphocytes à granules
Lymphocytes Tc : Lymphocyte T cytotoxique
Lymphocytes Th : Lymphocyte T « helper »
mAbs : « monoclonal Antibodies »
MMS : « Matrix Micro-Seeding »
Nb : Nanobody
NK : Natural Killer
NKG2D : « Natural Killer Group 2-member D »

PCR : « Polymerase Chain Reaction »
PEG : Polyéthylène glycol
PG : Propylène Glycol
PorM : Protéine M de *Porphyromonas gingivalis*
PPG : Polypropylène Glycol
pPorM : Partie périplasmique de la protéine PorM
pPorM-Ct : Partie C-terminale de la protéine pPorM
pPorM-Nt : Partie N-terminale de la protéine pPorM
pPorM-T : Partie résultante après protéolyse limitée à la Trypsine de la protéine pPorM
RCSB PDB : « Research Collaboratory for Structural Bioinformatics Protein Data Bank »
Rg : Rayon de giration
r.m.s.d : « Root Mean Square Deviation »
SAXS : « Small Angle X-ray light Scattering »
SDS-PAGE : « Sodium Dodecyl Sulfate-PolyAcrylamide Gel Electrophoresis »
SEC : « General SECRetion pathway »
SEC-MALS : « Size Exclusion Chromatography with Multi-Angle Light Scattering »
SeMet : Sélénométhionine
S.O.C : « Super Optimal broth with Catabolite repression »
Tat : « Twin
TCR : « T-Cell Receptor »
TNF : Facteur de nécrose tumorale
TxSS : Système de Sécrétion de Type x
VH : « Variable Heavy chain domain »
VL : « Variable Light chain domain »

Table des matières

REMERCIEMENTS	3
LISTE DES ABREVIATIONS	5
1 INTRODUCTION	9
1.1 INTRODUCTION GENERALE :.....	9
1.1.1 <i>Le système immunitaire</i>	9
1.1.2 <i>Les anticorps</i>	11
1.2 PRESENTATION DU SUJET DE THESE.....	15
1.2.1 <i>Présentation du premier volet : Anticorps thérapeutique contre des protéines MIC.</i> ..	15
1.2.2 <i>Présentation du deuxième volet : Anticorps d'intérêt biotechnologique contre la protéine pPorM.</i>	15
2 ETUDES STRUCTURALES D'ANTICORPS THERAPEUTIQUES CONTRE PROTEINES MIC 16	
2.1 INTRODUCTION	16
2.2 STRUCTURE DU FAB MIC12	23
2.3 CRISTALLISATION DES COMPLEXES DE MIC12 AVEC LES PROTEINES MIC	28
2.3.1 <i>Protéines MIC exprimées en système bactérien</i>	28
2.3.2 <i>Protéines MIC exprimées en système baculovirus/cellules d'insectes</i>	33
2.3.3 <i>Protéines MIC exprimées en cellules de mammifères</i>	36
2.4 ETUDES DES COMPLEXES MIC/MIC12 PAR SAXS.....	37
2.5 DOCKING MOLECULAIRE DU COMPLEXE MICA*001/MIC12 ET PROPOSITION D'UN MODELE.....	41
2.6 DISCUSSIONS ET CONCLUSIONS.....	44
3 UTILISATION DE NANOBODY POUR LES ETUDES STRUCTURALES DE LA PROTEINE PPORM	48
3.1 INTRODUCTION	48
3.1.1 <i>Les systèmes de sécrétion bactériens</i>	48
3.1.2 <i>Le système de sécrétion de type 9 (T9SS)</i>	58
3.2 GENERATION ET STRUCTURES DES NANOBODY ANTI-PORM.....	72
3.2.1 <i>Introduction</i>	72
3.2.2 <i>Article 1</i>	74
3.2.3 <i>Cristallisation du nanobody nb02 en ChipX et résolution de la structure</i>	75
3.3 ETUDES STRUCTURALES DE PPORM	79
3.3.1 <i>Structure de pPorM-Nt</i>	79
3.3.2 <i>Etudes SAXS de pPorM</i>	91
3.3.3 <i>Article 2 (manuscrit en préparation)</i>	97
3.4 DISCUSSIONS ET CONCLUSIONS.....	98
4 MATERIELS ET METHODES	102
4.1 DESCRIPTION DES METHODES UTILISEES	102
4.1.1 <i>Clonages</i>	102
4.1.2 <i>Dichroïsme Circulaire (CD)</i>	103
4.1.3 <i>Mesures d'interactions protéine/protéine par Bio-Layer Interferometry (BLI)</i>	104
4.1.4 <i>Cristallographie aux rayons X des protéines</i>	105
4.1.5 <i>« Small Angle Light Scattering » (SAXS)</i>	111
4.2 ETUDES STRUCTURALES D'ANTICORPS THERAPEUTIQUES CONTRE PROTEINES MIC	114
4.2.1 <i>Fab MIC12 : expression et purification</i>	114
4.2.2 <i>Protéines MIC exprimées en système bactérien</i>	115
4.2.3 <i>Protéines MIC : expression en système baculovirus</i>	120

4.2.4	<i>Analyses biochimiques et biophysiques</i>	123
4.2.5	<i>Cristallographie</i>	124
4.2.6	<i>SAXS des protéines MIC et complexes MIC/MIC12</i>	127
4.2.7	<i>Prédictions de structures du complexe MICA*001/MIC12 par docking moléculaire.</i> 129	
4.3	UTILISATION DE NANOBODY POUR LES ETUDES STRUCTURALES DE LA PROTEINE pPORM	130
4.3.1	<i>Nanobody nb01 et nb02 : expression et purification</i>	130
4.3.2	<i>Cristallographie à température ambiante de nb02</i>	131
4.3.3	<i>Clonage des constructions pPorM-Nt et pPorM-Ct</i>	132
4.3.4	<i>Expression des protéines pPorM-Nt native et SeMet</i>	133
4.3.5	<i>Purification des constructions de pPorM</i>	133
4.3.6	<i>Analyses biochimiques et biophysiques de pPorM-Nt</i>	134
4.3.7	<i>Purification des complexes pPorM-Nt/nanobody</i>	134
4.3.8	<i>Cristallographie aux rayons X de pPorM-Nt native et SeMet et des complexes pPorM-Nt/nanobody</i>	134
4.3.9	<i>SAXS des constructions de pPorM</i>	137
	REFERENCES :	139
	RESUME	144
	ABSTRACT	144

1 Introduction

1.1 Introduction générale :

L'environnement d'un organisme est constitué d'une multitude de micro-organismes. Ceux-ci peuvent être de différents types (bactéries, virus, levures, protozoaires ou parasites multicellulaires) et leur interaction avec l'organisme peut être bénéfique à celui-ci comme nocive. Dans ce second cas, ces micro-organismes peuvent entraîner des maladies par leur propagation et prolifération, pouvant aller jusqu'à la mort de l'organisme hôte. Celui-ci doit alors développer des stratégies de défense afin d'anticiper ou au mieux réagir aux attaques des pathogènes.

La première ligne de défense est constituée des défenses naturelles externes. La barrière cutané-muqueuse empêche la plupart de ces micro-organismes infectieux d'entrer en interaction avec l'organisme. Aussi, le lysozyme contenu dans la plupart des sécrétions, le mucus, les sécrétions des glandes sébacées, l'acidité gastrique, des germes ou simplement des caractéristiques physiques de l'organisme participent à l'élimination et non invasion d'agents infectieux (1). Cependant, ceux-ci peuvent passer au travers des épithéliums des tractus digestifs et uro-génitaux ou même au niveau des voies aériennes supérieures ou poumons. Lorsqu'il y a infection, c'est le système immunitaire qui prend en charge l'action de lutte et élimination de ces menaces.

1.1.1 Le système immunitaire

Le système immunitaire regroupe tous les agents et stratégies de défense de l'organisme. Il a pour fonction essentielle l'élimination des agents infectieux et la prévention des dommages qu'ils peuvent induire. Pour cela, il doit pouvoir les distinguer des éléments appartenant à l'organisme. Chaque molécule issue du génome d'un organisme est reconnue par celui-ci comme faisant parti du « soi ». Au contraire, toute molécule étrangère est reconnue comme du « non-soi ». Le système immunitaire doit pouvoir identifier les molécules du « non-soi » et adapter sa réponse en fonction.

Ce système est constitué de deux catégories : l'immunité innée et l'immunité adaptative. Selon la localisation et la nature de la menace, l'une, l'autre ou les deux

stratégies vont être mises en application. L'immunité innée (aussi appelée immunité naturelle) représente la réponse non spécifique tandis que la réponse adaptative se développe et se renforce au contact de l'agent infectieux de manière spécifique. La réponse immunitaire adaptative est toujours initiée par la présence de molécules reconnues spécifiquement, les antigènes.

Les principales cellules de défense du système immunitaire sont les leucocytes. Ils regroupent deux sous-groupes de cellules : le premier est constitué des macrophages, monocytes et polynucléaires neutrophiles et représente la toute première ligne de défense contre l'infection tandis que le second est constitué de lymphocytes. Parmi ces lymphocytes se trouvent les lymphocytes T, les grands lymphocytes à granules (LGL) et les lymphocytes B.

Les lymphocytes T sont des lymphocytes développés dans le thymus. Ils peuvent être classés en deux types selon leurs activités: les lymphocytes T accessoires « helper » Th, regroupant les cellules aidant à l'activité des lymphocytes B ou des phagocytes mononucléés, et les lymphocytes T cytotoxiques Tc. Ils participent ainsi à la fois à la réponse immunitaire innée et adaptative.

Par leur sécrétion de cytokines, molécules de signalisation intercellulaire, les lymphocytes Th aident au contrôle des lymphocytes B, participent au développement d'anticorps et interagissent avec les cellules phagocytaires. Plus généralement, ces lymphocytes contrôlent les réponses immunitaires.

Les lymphocytes Tc ont, quant à eux, une action cytotoxique sur les cellules tumorales ou infectées par un virus. Lors de l'infection virale par exemple, des molécules propres au virus vont être présentés à la surface de la cellule par des molécules du Complexe Majeur d'Histocompatibilité (CMH). Les lymphocytes Tc vont alors les reconnaître à partir de récepteurs « T-cell receptor » TCR qu'ils portent à leur surface.

Les grands lymphocytes à granules (LGL) sont des cellules à action cytotoxique. À la différence des lymphocytes Tc, leur reconnaissance est non spécifique. Ces cellules sont également appelées cellules Natural Killer (NK). Tout comme les lymphocytes T cytotoxiques, elles portent à leur surface des récepteurs reconnaissant les molécules du CMH. Selon la molécule reconnue, une réponse activatrice ou inhibitrice de l'activité de la cellule NK sera déclenchée. Lorsque cette réponse est activatrice, les substances contenues

dans les granules, les perforines et granzymes, sont libérées dans la synapse et lysent les membranes des cellules cibles, qui sont ensuite phagocytées.

Les lymphocytes B sont à l'origine du développement des anticorps. Développés dans la moelle osseuse, chaque cellule est conçue pour ne porter à sa surface qu'un seul type de récepteur à reconnaissance spécifique d'un antigène.

Lorsque l'un de ces récepteurs se lie à un antigène, le clone est sélectionné (on parle de sélection clonale). Les cellules correspondantes sont alors stimulées pour proliférer, arriver à maturation et se différencier en plasmocytes. Ce type cellulaire sécrète ensuite les récepteurs, solubles, dans le sérum. Ils sont alors appelés anticorps. Des cellules maturées mais non différenciées sont conservées comme « mémoire » de l'antigène, les cellules à mémoire. L'efficacité de ce mécanisme est renforcée à chaque contact avec l'antigène.

Lorsqu'un type d'anticorps est développé, ceux-ci vont reconnaître et se lier aux antigènes portés à la surface d'une cellule ou d'un agent infectieux. Cette simple interaction peut suffire à bloquer l'action du pathogène. Dans le cas contraire, la partie accessible de l'anticorps va pouvoir être reconnue par d'autres cellules de défense portant les récepteurs spécifiques et permettre ainsi l'attachement de celles-ci. Elles pourront ensuite prendre en charge l'élimination de la menace avec une efficacité augmentée, puisque mieux ancrées dans l'environnement de l'agent infectieux.

1.1.2 Les anticorps

Les anticorps, ou immunoglobulines (Ig), sont des molécules de reconnaissance d'un antigène spécifique. Développés par les lymphocytes B, ces glycoprotéines se trouvent libres, solubles, dans le sérum après sécrétion par des cellules différenciées, les plasmocytes. La reconnaissance de leur antigène passe par l'interaction du paratope, porté par l'anticorps, avec l'épitope, zone spécifique portée par l'antigène.

Le paratope est porté par la partie variable de l'anticorps. Le reste de sa structure est constante et relativement conservée. Il existe cinq types d'Ig classés selon la séquence de cette partie constante : IgM, IgD, IgG, IgE et IgA.

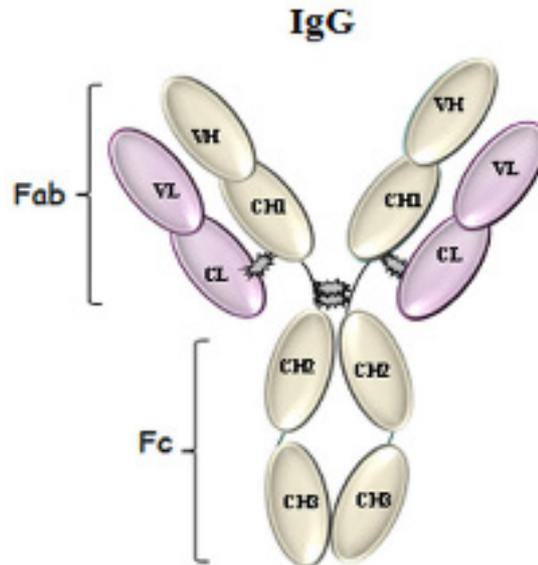


Figure 1.1.1 : Structure d'un anticorps IgG (2). L'anticorps est constitué de deux chaînes lourdes (beige), reliées par des ponts disulfures (en gris), et deux chaînes légères (rose). Les chaînes légères et lourdes sont également maintenues assemblées par des ponts disulfures. Chaque chaîne lourde comprend 3 domaines constants CH et un domaine variable VH tandis que chaque chaîne légère compte un domaine constant CL et un domaine variable VL. Les sites de liaison à l'antigène sont portés par les deux complexes de domaines variables chaîne légère/chaîne lourde. Chez les IgE et IgM, un domaine constant supplémentaire est présent sur la chaîne lourde.

La structure basique d'un anticorps est illustrée en Figure 1.1.1. De taille approximative de 150 kDa, elle peut être divisée en deux parties : le fragment de liaison à l'antigène (Fab) et le fragment constant (Fc). Quatre chaînes polypeptidiques sont assemblées : deux chaînes lourdes et deux légères. Les chaînes lourdes sont constituées pour la plupart de 3 domaines constants (C_{H1} , C_{H2} et C_{H3}) et d'un domaine variable (V_H). Les chaînes légères, elles, sont formées d'un domaine constant C_L et d'un domaine variable V_L . Les chaînes lourdes sont maintenues assemblées par des ponts disulfures et chaque chaîne légère est associée à l'une des chaînes lourdes, également par des ponts disulfures. Le Fc comprend les domaines C_{H2} et C_{H3} des deux chaînes lourdes. Chacun des deux Fab est composé des domaines C_{H1} , V_H , C_L et V_L . La structure des domaines variables et constants, fortement similaires, appartient à la famille des repliements de type « immunoglobuline ». Ce type de repliement consiste en deux feuillets β antiparallèles. Les deux structures sont présentées en Figure 1.1.2. Dans ce Fab, le fragment variable Fv porte les trois régions déterminantes de la complémentarité CDR formant le site de liaison à l'antigène. Le CDR1 est la boucle située entre le brin B et C, le CDR2 entre le C' et C'' et le CDR3 entre les brins F et G. Le CDR3 est généralement le plus long et, dans la majorité des cas, est impliqué dans l'interaction avec l'épitope de l'antigène. Le Fc quant à lui est reconnu par des récepteurs spécifiques portés par certaines cellules de défense.

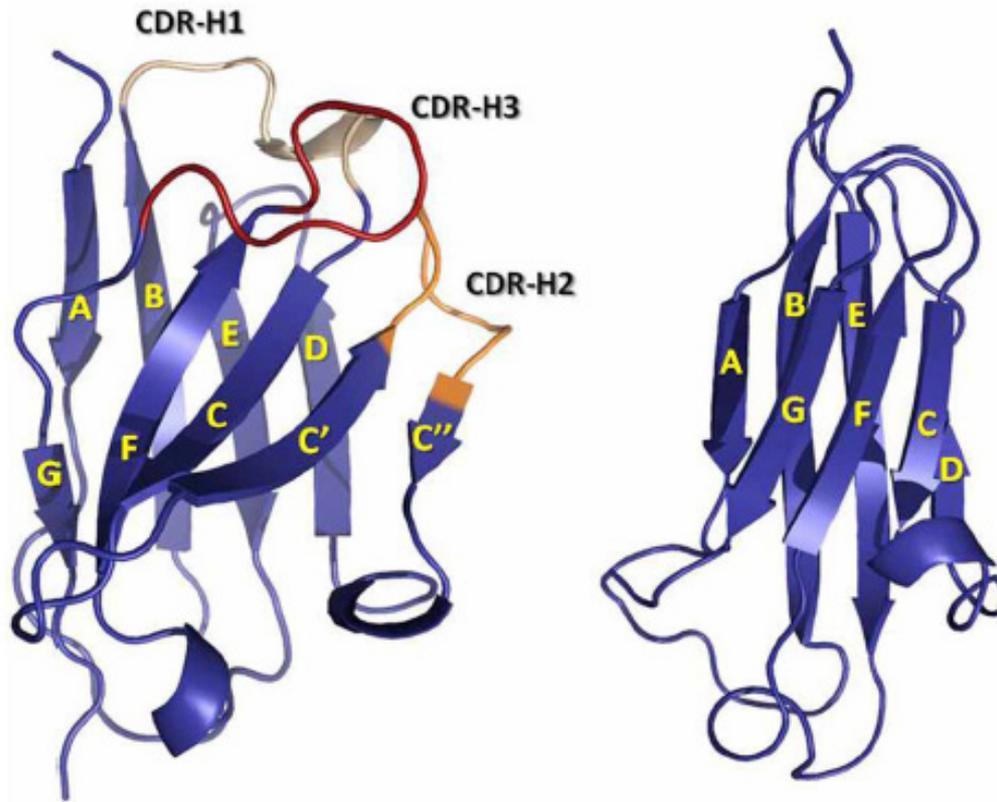


Figure 1.1.2: Structure en rubans des domaines variable et constant d'un Fab (3). Les deux types de domaines appartiennent au repliement de type « immunoglobuline ». À gauche, le domaine variable 9 feuillets β et trois boucles CDR-H1, CDR-H2 et CDR-H3. Le domaine constant à droite est lui constitué de 7 feuillets β .

De nos jours, de plus en plus de stratégies de traitement de maladies rares ou cancer sont développées à partir de l'utilisation d'IgG. La production d'anticorps monoclonaux mAbs dirigés contre des cibles d'intérêt sert à la fois à la prévention, à la détection et au traitement en lui-même. Les anticorps, sous leur forme de Fab, ont beaucoup été utilisés comme outils de détection et diagnostique depuis leur découverte, du fait de leur reconnaissance spécifique d'antigènes.

Pourtant, la taille imposante de cette protéine (150 kDa) présente des inconvénients : elle limite sa pénétration et biodistribution. De plus, son activité est rapidement neutralisée par le système immunitaire et sa durée de vie est de quelques jours.

Un nouveau type d'anticorps a été découvert au début des années 1990. Ces anticorps ont été trouvés dans le sérum de camélidés, incluant les lamas et chameaux, en plus d'anticorps IgG classiques. Ils ont été nommés anticorps de chaîne lourde HcAbs, dû à l'absence de chaîne légère (2) (Figure 1.1.3).

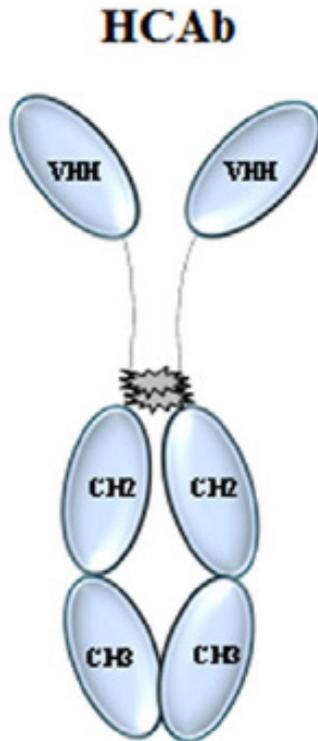


Figure 1.1.3: Anticorps de Camélidé HcAbs (2). Ce type d'anticorps est constitué de deux chaînes lourdes, comprenant les domaines constants CH3 et CH2 et le domaine variable VHH. Les deux chaînes sont reliées par des ponts disulfures (ici en gris). Les VHH contiennent à eux seuls la totale capacité de liaison à l'antigène.

L'anticorps est formé de l'assemblage de deux chaînes lourdes. Chacune compte deux domaines constants, C_{H3} et C_{H2} , et un domaine variable VHH, également appelé nanobody Nb (4). Les deux nanobody portent les CDR interagissant avec l'épitope de l'antigène. Ces CDR sont situés pour le CDR1 entre les brins 3-4, le CDR2 entre les brins 5-6 et le CDR3 entre les brins 9 et 10 (5). Là encore, la reconnaissance est majoritairement réalisée par l'interaction du CDR3, généralement très long. Le repliement des domaines est également de type « immunoglobuline ».

Les nanobody ont une taille approximative de 15 kDa. Ils présentent alors de nombreux avantages vis à vis des mAbs, comme leur capacité biodiffusive plus importante. Aussi, ils sont plus stables dans le temps et face à des conditions très variées (4). Ils ont donc des applications médicales comme biotechnologiques, notamment dans les études structurales. Leur grande stabilité et spécificité pour un antigène leur permettent de former des complexes anticorps/antigènes stables. Les nanobody servent alors de chaperonnes de cristallisation et de marquage moléculaire.

1.2 Présentation du sujet de thèse

Durant ces trois années de thèse, j'ai été chargée de l'étude structurale de protéines en complexe avec ces deux types d'anticorps: le premier volet implique l'utilisation d'anticorps à visée thérapeutique tandis que le second volet utilise les nanobody en tant qu'outils biotechnologiques.

1.2.1 Présentation du premier volet : Anticorps thérapeutique contre des protéines MIC

Le premier volet de cette thèse porte sur l'étude structurale d'un anticorps thérapeutique dirigé contre des protéines humaines homologues à celles du Complexe Majeur Histocompatibilité (CMH) de classe I. J'ai travaillé sur un anticorps, nommé ici MIC12, reconnaissant les protéines MICA*001, MICA*008 et MICB*002. L'objectif était d'étudier MIC12 mais également d'étudier son interaction avec chacune des protéines MIC. Pour cela, je devais utiliser la cristallographie aux rayons X afin de résoudre la structure atomique de MIC12 et des complexes MIC12/MICA*001, MIC12/MICA*008 et MIC12/MICB*002.

Ce sujet a été financé par une bourse d'étude de la région PACA avec comme partenaire économique l'entreprise Innate Pharma.

1.2.2 Présentation du deuxième volet : Anticorps d'intérêt biotechnologique contre la protéine pPorM

Le second volet de cette thèse porte sur l'étude structurale de la protéine pPorM du système de sécrétion de type 9 (T9SS) de *Porphyromonas gingivalis* utilisant des nanobody.

Ce sujet fait partie du projet T9SS, qui est l'une des thématiques phares de l'équipe de recherche « Interactions hôte/pathogène » dans laquelle j'ai effectué ma thèse. Ce projet est financé par l'ANR « Toothpaste », en collaboration avec l'équipe du docteur Eric Cascales au LISM (Marseille).

2 Etudes structurales d'anticorps thérapeutiques contre protéines MIC

2.1 Introduction

Dès les premiers stades du développement, l'organisme élabore son système immunitaire afin de lutter contre tout agent pathogène (6). Ce système englobe deux types de réponses immunitaires : innée et adaptative. Le système immunitaire inné représente les premières lignes de détection et identification de l'infection ou défaillance ainsi que les premières défenses de l'organisme. Le système immunitaire adaptatif, quant à lui, intervient plus tard, avec l'élaboration d'anticorps adaptés au pathogène. À partir des premières semaines de gestation, ces deux systèmes se développent de manière exponentielle et ce jusqu'à l'âge adulte (7). Les macrophages, par exemple, apparaissent à partir de la quatrième semaine de gestation chez l'Homme tandis que les granulocytes et cellules Natural Killer NK se développent à partir de la huitième semaine. Ces acteurs principaux du système immunitaire inné interviennent dans la reconnaissance de cellules infectées, ou plus généralement en stress. Le même mécanisme entre en jeu lors de phénomènes de dégénérescence cellulaire. Des cellules saines adoptent une multiplication anarchique sans subir d'apoptose et entrent en état de stress. Elles forment alors des tumeurs et sont appelées cellules cancéreuses. Afin d'enrayer ce processus et empêcher leur prolifération, le système immunitaire déclenche au plus vite la réponse immunitaire innée par l'action des macrophages et cellules NK. Ce processus est appelé immunosurveillance.

Les cellules NK sont des grands lymphocytes granuleux issus de la moelle osseuse (8). Elles interviennent dans la réponse immunitaire innée par leur action cytotoxique, mais également par leur sécrétion de cytokines (interférons $IFN-\gamma$, facteurs de nécrose tumorale (TNF) et « granulocyte-macrophage colony-stimulating factor » (GM-CSF)) et chemokines. Lorsqu'une de ces cellules reconnaît une cellule tumorale ou infectée, elle libère dans la synapse d'interaction des perforines contenues dans ses granules cytoplasmiques. Celles-ci vont alors former des pores dans la membrane de la cellule cible, entraînant sa lyse. D'autres actions cytotoxiques peuvent avoir lieu, comme par exemple la libération de Granzymes B (9).

La distinction entre cellules saines et en stress est faite par les cellules NK au travers d'une reconnaissance de molécules du soi. Cette reconnaissance est réalisée par l'intermédiaire de récepteurs portés à la surface de la cellule NK (Figure 2.1.1). Ils peuvent être activateurs ou inhibiteurs d'un mécanisme. Dans un cas non pathologique, les cellules NK reconnaissent les protéines du CMH de classe I présentes à la surface de la cellule à l'aide de récepteurs inhibiteurs. Cela a pour effet d'inhiber l'action de ces cellules immunitaires. Lors d'une infection ou développement d'une tumeur, il y a surexpression de récepteurs liés au stress (8). Ceux-ci servent de ligands aux récepteurs activateurs des cellules de défense et leur interaction va activer la réponse immunitaire.

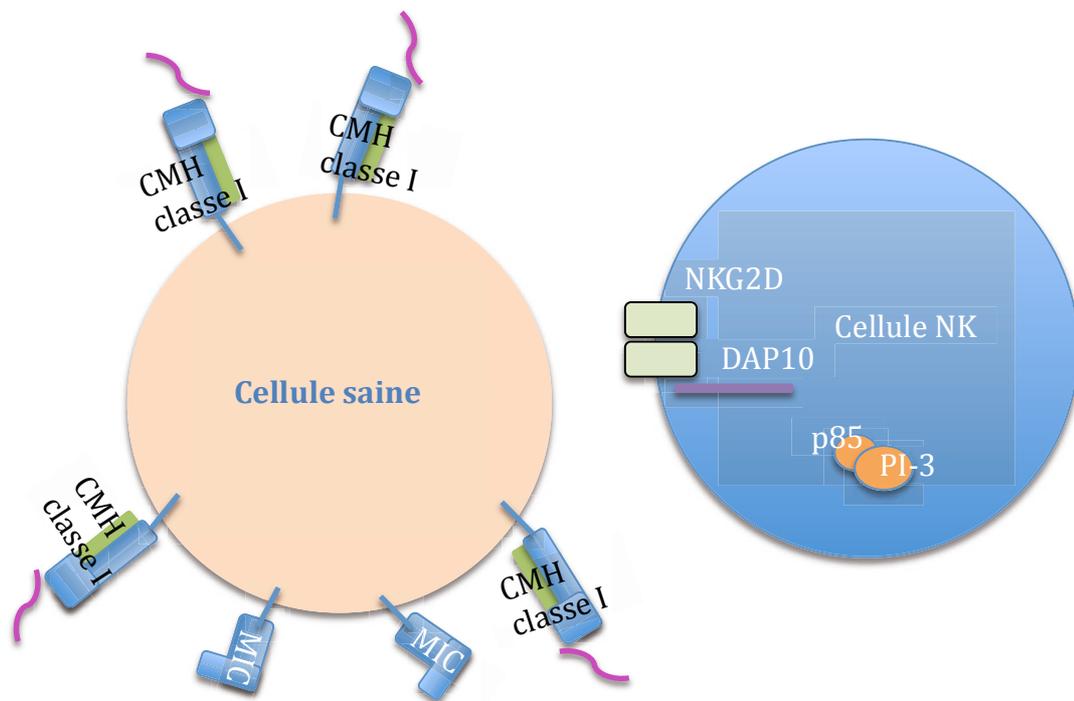


Figure 2.1.1 : Mécanisme d'action des cellules NK, cas non pathologique. La cellule hors état de stress est représentée en beige, la cellule NK inactive en bleu. La cellule saine possède à sa surface une majorité de protéines du CMH de classe I par rapport aux protéines MIC, pas ou très peu exprimées. Elles présentent à leur surface un peptide du soi (en mauve). Cette sous-représentation des protéines MIC ne permet pas leur interaction avec le récepteur NKG2D (en vert pale) et n'active donc pas le recrutement de la kinase Phosphatidylinositol-3 (PI-3) par l'intermédiaire de son domaine p85. La réponse immunitaire n'a pas lieu.

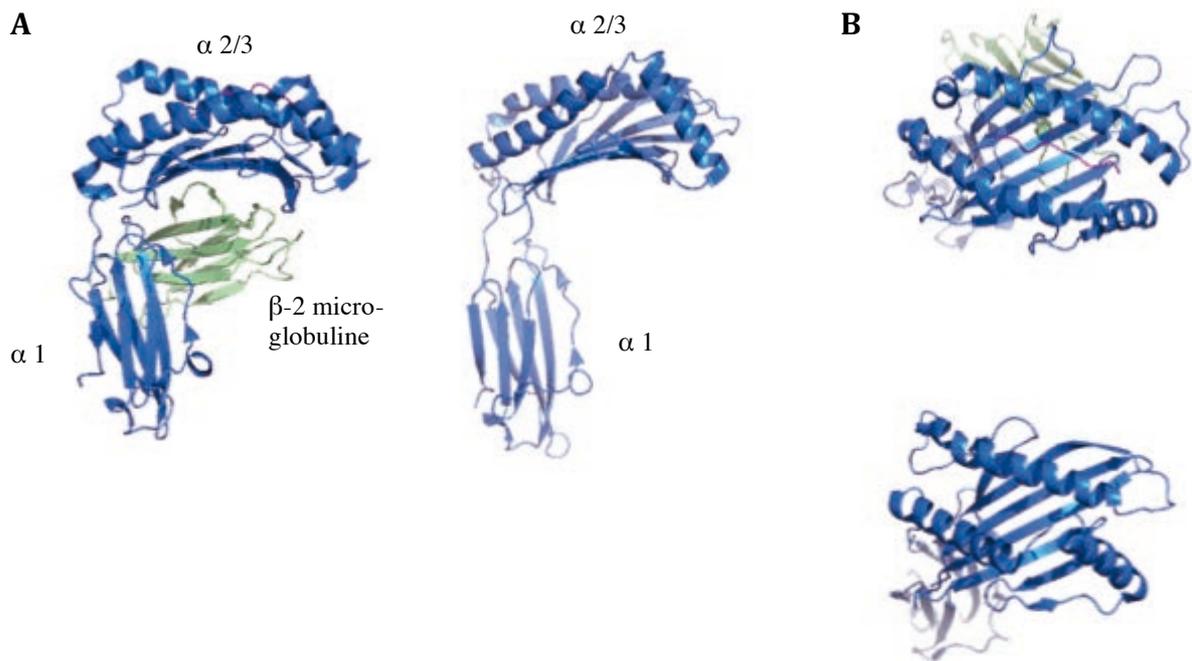


Figure 2.1.2: Exemple de protéine du CMH de classe I et de protéine MIC. Les modèles sélectionnés sont les protéines HLA-B*27:05 (5IB2) et MICA 001 (1HYR). La chaîne alpha est représentée en bleu (avec les domaines $\alpha 1$, 2 et 3), la sous-unité β -2-microglobuline en vert et le peptide en magenta. A) Vue de côté avec à gauche la protéine du CMH et à droite la protéine MIC. B) Vue de dessus avec la protéine du CMH en haut et la protéine MIC en bas.

Le récepteur activateur « Natural Killer Group 2-member D » (NKG2D) est porté à la surface des lymphocytes T $\text{TCR}\gamma\delta^+$ et $\text{CD8}^+ \text{TCR}\alpha\beta^+$ ainsi que des cellules NK. Il s'agit d'un récepteur de type C « lectin-like ». Ancré à la membrane par sa partie transmembranaire, il forme un homodimère interagissant par sa partie cytoplasmique avec la protéine adaptatrice DAP10 (10). Chez l'Homme, les ligands de NKG2D sont les protéines ULBP-1 à -6, MICA et MICB.

MICA et MICB sont des protéines multi-alléliques homologues des protéines du CMH de classe I. Les structures cristallographiques des parties extracellulaires des allèles MICA^*001 et MICB^*002 sont connues et peuvent être retrouvées dans la RCSB PDB sous les codes 1B3J (11) et 1JE6 (12). Celles-ci sont constituées d'une chaîne contenant trois domaines $\alpha 1$, $\alpha 2$ et $\alpha 3$ (Figure 2.1.2). Cette organisation diffère de celle des protéines du CMH de classe I, constituées de la même chaîne α reliée à une sous-unité β -2-microglobuline.

Elles ne présentent pas non plus d'antigènes à la surface cellulaire, malgré la présence dans leur structure d'un sillon de présentation d'antigène. Ces protéines ne sont pas exprimées à la surface de la cellule saine mais le sont lorsque la cellule entre en état de stress. Des études ont prouvé par exemple que cette surexpression intervient chez des

cellules HeLa subissant un choc thermique (8). Il existe plus de 70 allèles de MICA mais l'allèle le plus présent chez l'Homme est MICA*008 (entre 21% et 47%) (13).

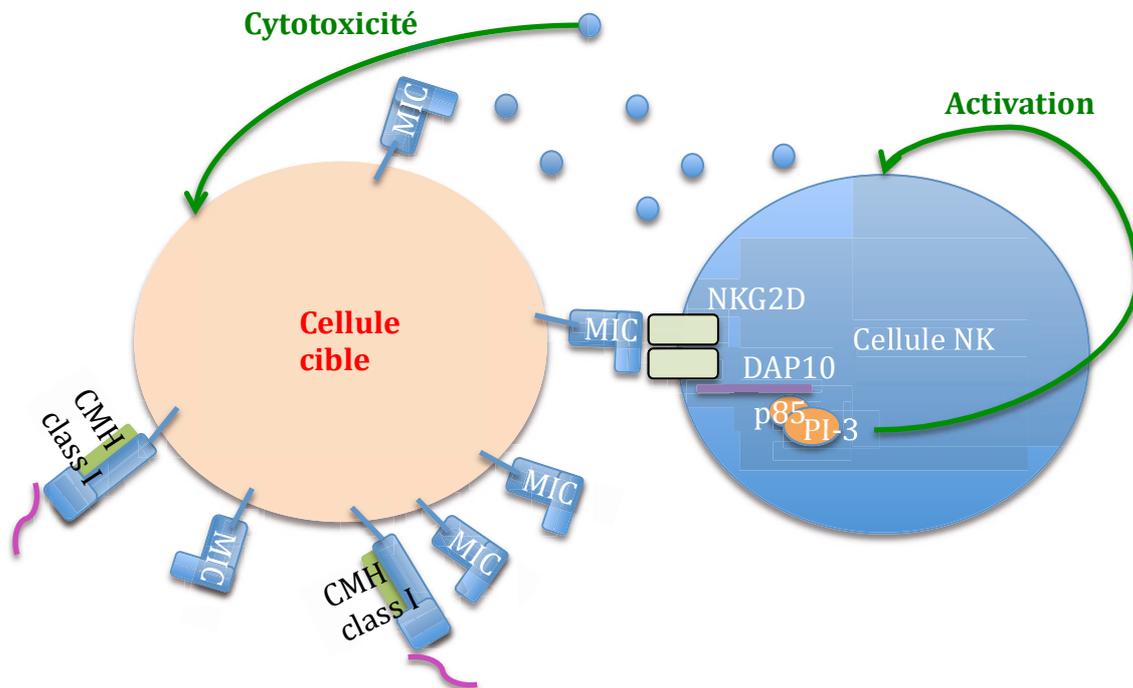


Figure 2.1.3: Mécanisme d'action des cellules NK, cas pathologique. Dans un cas de stress cellulaire, la balance protéines du CMH/protéines MIC va pencher en faveur de ces dernières qui seront sur-représentées à la surface cellulaire. Cela entraîne leur interaction avec le récepteur NKG2D, déclenchant le recrutement de la kinase Phosphatidylinositol-3 (PI-3) par son interaction avec DAP10 via son domaine p85. La cellule NK est alors activée et lyse la cellule cible par sécrétion de perforines (ronds bleus).

Lorsque le récepteur NKG2D reconnaît une protéine MIC, leur interaction entraîne la phosphorylation du motif YxxM de DAP10. Celui-ci recrute alors une kinase Phosphatidylinositol-3 qui s'y fixe par l'intermédiaire de son domaine p85. Celle-ci déclenche alors l'activation de la cellule NK et son action cytotoxique (8) (Figure 2.1.3).

Cependant, une diminution de cette cytotoxicité peut être observée au cours du temps. Ceci est dû à la régulation négative que subissent les cellules NK, due à une stratégie d'« évasion immunitaire » de la part de la cellule cible. Cette régulation intervient au niveau de l'expression de NKG2D et du mécanisme d'action en lui-même (Figure 2.1.4). Un premier mécanisme fait intervenir des métalloprotéases ADAM17 qui viennent cliver les protéines MIC pour les détacher de la surface de la cellule cible. Des protéines solubles MIC se retrouvent alors dans la synapse et interagissent avec NKG2D. Ceci a pour effet de saturer ces récepteurs qui ne peuvent donc plus reconnaître de protéines MIC encore ancrées à la surface des cellules cibles. Un second mécanisme a été observé, pour l'allèle MICA*008 uniquement jusqu'à présent, consistant à son adressage à l'endosome par la voie de synthèse. La protéine est endocytée, amenée à l'endosome qui, une fois à maturité,

relargue dans le milieu extra-cellulaire des exosomes contenant MICA*008. Ceux-ci se fixent également sur NKG2D, entraînant les mêmes conséquences que les protéines MIC solubles (13). Les cellules NK ne sont alors plus capables de reconnaître les cellules cibles et ne surexpriment plus NKG2D. Une autre hypothèse attribue la saturation de NKG2D aux protéines MIC exprimées à la surface des cellules (11). L'action des cellules NK faisant partie de la réponse immunitaire innée, elle n'est pas faite pour perdurer dans le temps et diminuera peu à peu.

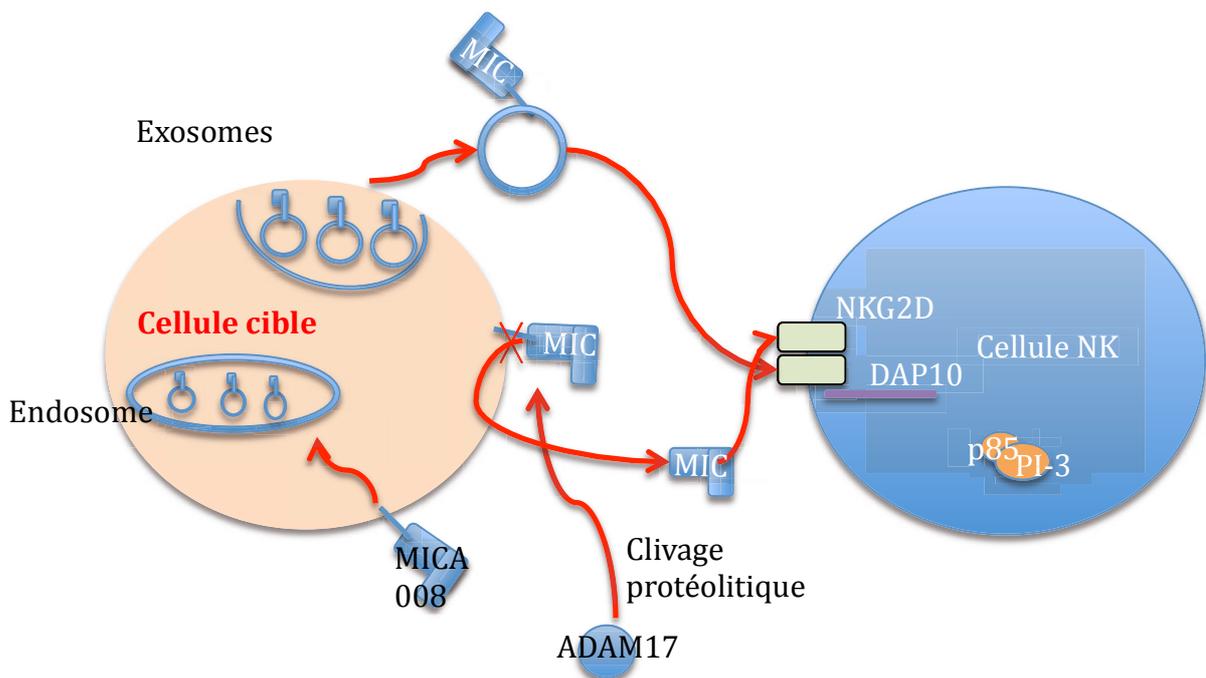


Figure 2.1.4: Mécanisme d'action des cellules NK : régulation négative. La régulation négative de l'action des cellules NK se met en place selon différents mécanismes. Les métalloprotéases ADAM 17, sécrétées par les cellules cibles, réalisent un clivage protéolytique des protéines MIC, qui se retrouvent alors dans le milieu extracellulaire. La cellule cible n'est alors plus reconnue par la cellule NK et les protéines MIC clivées elles-mêmes régulent négativement par saturation des récepteurs NKG2D. L'allèle MICA 008, quant à lui, est importé vers l'endosome pour être ensuite sécrété à la surface d'exosomes. Ceux-ci participent également à la saturation de NKG2D.

La stratégie développée par la société de biotechnologie Innate Pharma est de bloquer l'interaction MIC/NKG2D. Pour cela, des anticorps compétitifs à NKG2D, se fixant aux protéines MIC, ont été générés.

Les anticorps reconnaissent les protéines MIC, qu'elles soient fixées à la membrane, clivées ou portées à la surface d'exosomes (Figure 2.1.5). L'idée est de bloquer l'interaction tout en laissant l'accès à des effecteurs afin de permettre la régénération de l'action des cellules NK. Ces anticorps, élaborés sous une forme d'IgG cytotoxiques, devraient aussi servir de système de marquage des cellules en stress et permettre le recrutement de nouvelles cellules NK par reconnaissance du Fc par d'autres récepteurs.

Le Fab MIC12 est généré à partir d'un de ces anticorps. Testé sur les différents allèles représentatifs de chaque sous groupes structuraux, MIC12 interagit avec tous. Des expériences de mutagenèse dirigées d'allèles de MICA ont été menées sur deux à quatre résidus à la fois et l'interaction avec MIC12 a ensuite été testée. Cette méthode a permis d'identifier des résidus impliqués dans la reconnaissance MIC12/MIC (14). Pourtant, la zone définie est trop petite et ne comble pas la totalité de la zone d'interaction.

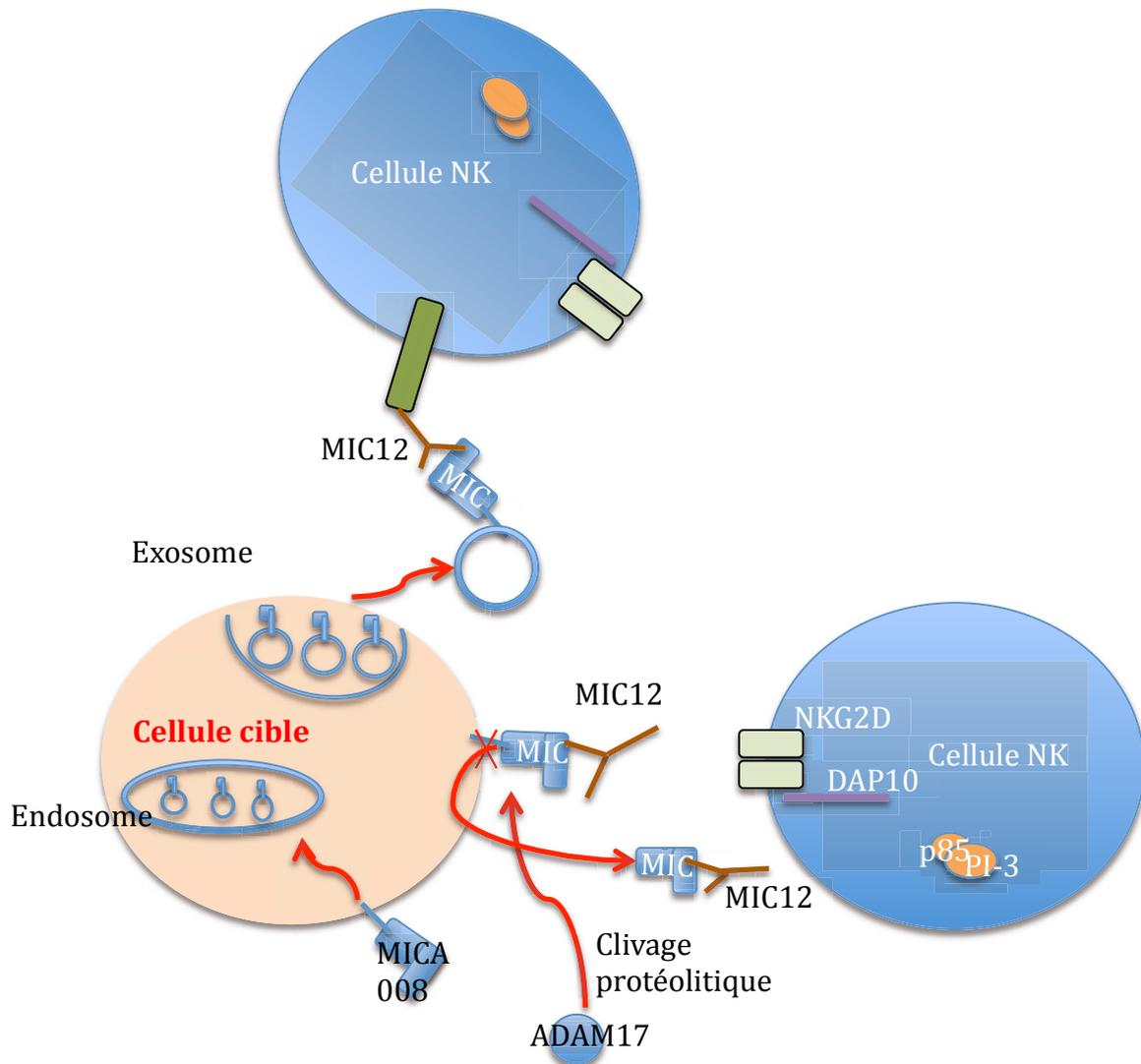


Figure 2.1.5: Inhibition de la régulation négative de l'action des cellules NK par l'anticorps MIC12. La présence de l'anticorps MIC12 d'Innate Pharma dans la synapse cellulaire entraînerait l'inhibition de la régulation négative des cellules NK. MIC12 reconnaît les protéines MIC sous toutes leurs formes (attachées à la surface ou clivées) et empêche donc l'interaction NKG2D/MIC. Le Fc de MIC12 est reconnu par d'autres récepteurs, permettant le recrutement de nouvelles cellules NK.

Le sujet MIC consiste en l'étude structurale de MIC12 ainsi que des complexes qu'il forme avec respectivement les protéines MICA*001, MICA*008 et MICB*002. La caractérisation des interactions anticorps/protéine permettra une description détaillée de la

zone de fixation de l'anticorps et ainsi de son inhibition de l'interaction MIC/NKG2D. Le but de cette étude est de cartographier le site d'interaction afin de le protéger par un brevet et permettre la poursuite des études en phases cliniques.

Dans la littérature, l'expression des protéines MICA*001 et MICB*002 est faite soit en système bactérien, passant par un protocole d'extraction depuis les corps d'inclusion puis repliement de la protéine, soit en système baculovirus. Des premières expressions de l'allèle MICA*001 en système bactérien ont été réalisées par Innate Pharma et les premières analyses ont montré que la protéine s'exprimait là aussi dans les corps d'inclusions. Une étape d'extraction à l'aide de détergents suivie d'une étape de repliement est donc nécessaire à la purification de cet allèle. Un protocole, non optimisé, a été établi et quelques purifications ont pu être menées avec, cependant, un faible rendement.

Quatre anticorps étaient prévus à l'origine, cependant, seul l'anticorps MIC12 fût retenu. Des analyses de l'interaction de cet anticorps ont permis de déterminer des constantes de dissociations de l'ordre du nanomolaire.

2.2 Structure du Fab MIC12

L'anticorps thérapeutique MIC12 a été exprimé en lignée stable de cellules CHO, fournies par Innate Pharma. La construction contient une étiquette histidine en C-terminal. D'après une analyse *in silico* de sa séquence à l'aide de l'outil ProtParam d'ExPaSy (<http://web.expasy.org/protparam/>), il s'agit d'une construction de poids moléculaire 47,755 kDa et ayant un point isoélectrique de 8,65.

La protéine MIC12 est exprimée dans les cellules puis sécrétée dans le milieu de culture. Après récupération de ce surnageant, la protéine est extraite par fixation sur colonne d'affinité Ni²⁺ (GE Healthcare) puis éluée avec un tampon contenant 250 mM d'Imidazole. Les éventuels contaminants sont ensuite éliminés et le tampon changé pour un tampon plus adapté à la cristallogénèse à l'aide d'une chromatographie d'exclusion de taille. Le Fab est élué à un volume de 93 mL sur une colonne Superdex 200 16/60 (GE Healthcare) (Figure 2.2.1). Une fraction du pic a été déposée sur SDS-PAGE pour vérifier que tous les contaminants ont été éliminés. La protéine dénaturée est bien seule et migre sur gel en trois bandes, la première correspondant à la migration d'une protéine de 47 kDa, la deuxième de 25 kDa et la dernière de 23 kDa. Ce résultat était attendu puisque la présence d'agent réducteur (DTT) a brisé une partie des ponts disulfure reliant les chaînes lourde et légère du Fab. Le rendement moyen d'anticorps obtenu pour 1L de culture est de 40 mg.

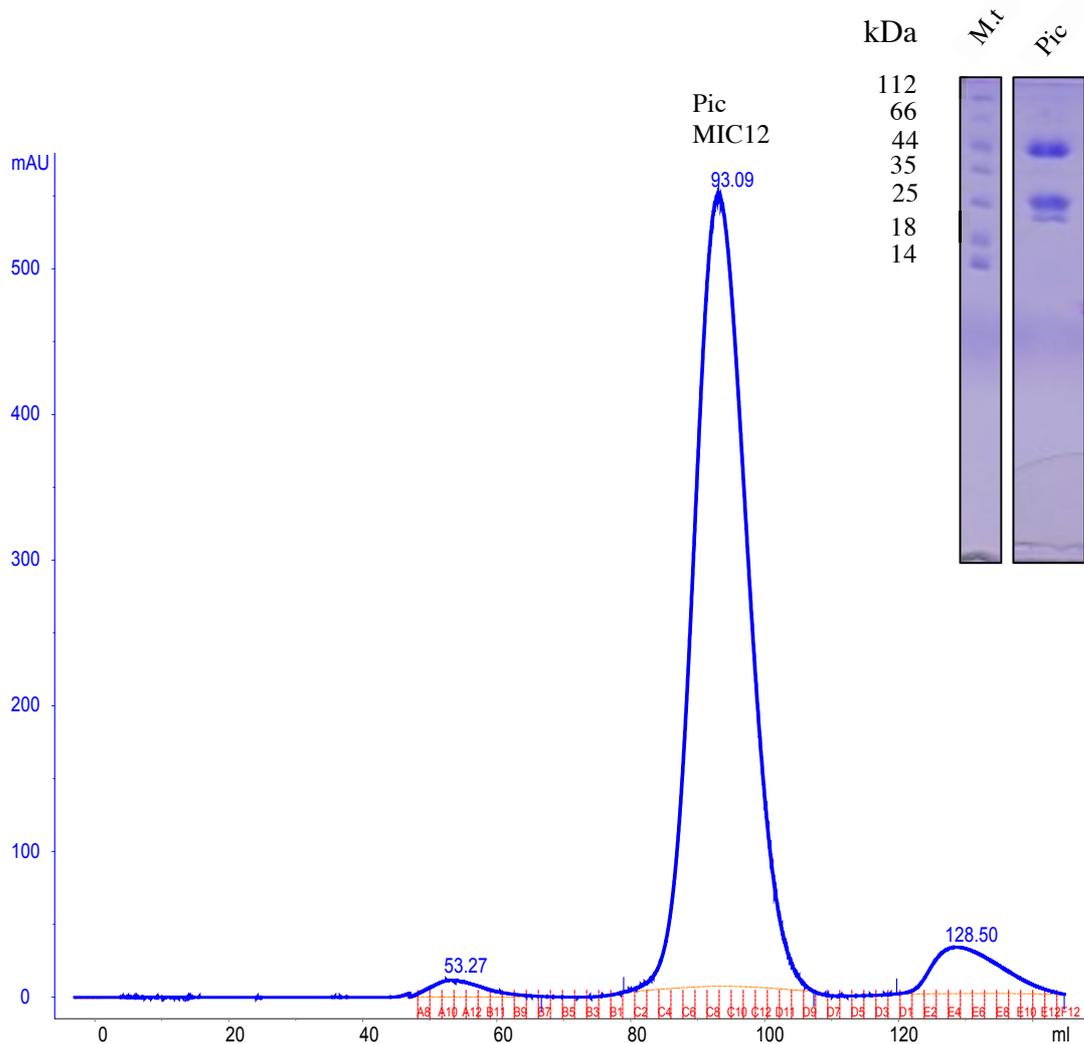


Figure 2.2.1: Purification du Fab MIC12. La protéine est éluée de la colonne Superdex 200 16/60 à un volume d'élution de 93,09 mL. La fraction du pic déposée sur SDS-PAGE 15% migre en 3 bandes à 47, 25 et 23 kDa. M.t: Marqueur de taille.

Des essais de cristallisation en diffusion de vapeur avec des gouttes assises ont été réalisés. La protéine a d'abord été concentrée à 18 mg/mL mais l'évaluation de la concentration par un premier screen général Stura a révélé qu'il fallait encore l'augmenter. Les screens généraux MDL, Index et JCSG+ ont été testés avec une concentration en MIC12 de 23 mg/mL. Après 36h, un cristal est apparu dans la condition C5 du screen MDL (Figure 2.2.2A). La condition de cristallisation est 0,1 M de Sodium Hepes pH 8,2, 1,4 M de Tri-sodium citrate dihydrate. L'optimisation de cette condition a donné des cristaux au bout de

24h (Figure 2.2.2B et 10C) puis 44 jours (Figure 2.2.2D). Après ajout de 10% de glycérol comme agent cryoprotectant puis congélation du cristal de la Figure 2.2.2B, sa diffraction a été testée au laboratoire. La caractérisation du cristal à partir de 2 images a permis de définir une stratégie de collecte et un jeu de données de 120 images a été enregistré. Les caractéristiques de ce jeu ont été reportées dans le Tableau 2.2.1.

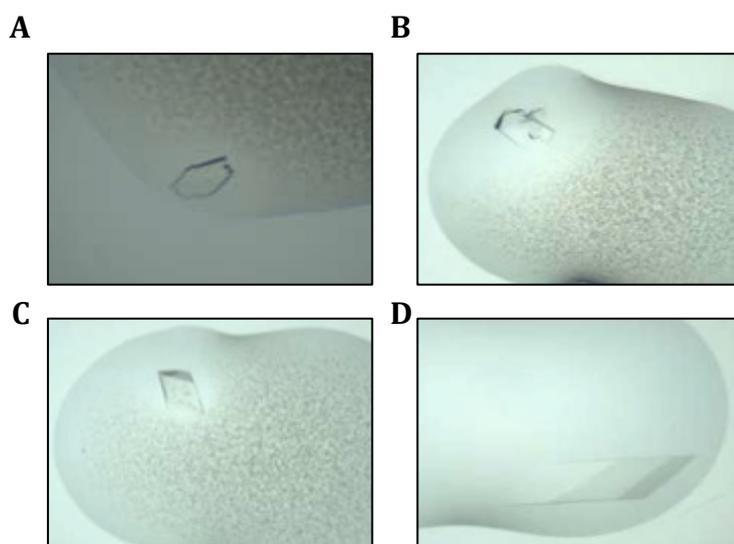


Figure 2.2.2: Cristallogénèse de MIC12. A: Cristal obtenu en condition C5 du screen MDL. B,C et D: Cristaux de la condition optimisée, obtenus respectivement en Na3Citrate 1,36 M pH 7,8 (B) Na3Citrate 1,08 M pH 8 (C) et Na3Citrate 0,78 M pH 8,4 (D).

Bien que le jeu d'images enregistrées suffise à la résolution de la structure de MIC12, le cristal a été stocké en azote liquide pour tester sa diffraction et enregistrer un nouveau jeu de données sur une source synchrotron. Ceci a été réalisé sur la ligne ID23-2 de l'ESRF à Grenoble avec cette fois-ci une diffraction allant jusqu'à 1,4 Å. Un nouveau jeu de données a été enregistré.

La structure a été résolue à 1,4 Å. Les statistiques de collecte, de remplacement moléculaire et d'affinement ont été reportées dans le Tableau 2.2.1. Cette structure est constituée de la chaîne légère et de la chaîne lourde, toutes deux constituées d'un domaine constant et d'un domaine variable. Deux ponts disulfure intra-chaîne sont visibles pour chacune de chaînes : entre les résidus C23-C89 et C135-C195 pour la chaîne légère et entre les résidus C22-C96 et C142-C198 pour la chaîne lourde. Un troisième pont disulfure, inter-chaîne, est présent entre les résidus C215(L)-C218(H). Deux boucles n'ont cependant pas pu être construites dû à leur haut degré de flexibilité : S129(H)-G135(H) et S189(H)-Q194(H). D'après la classification des repliements de SCOP, les quatre domaines VH, CH1, VL et VH appartiennent à la famille des repliements Immunoglobulin-like β -sandwich. Les domaines

variables portent les CDRs, avec pour particularité un CDR3 long d'une dizaine de résidus (Figure 2.2.3).



Figure 2.2.3: Structure du Fab MIC12. La chaîne légère est représentée en bleu gris et la chaîne lourde en vert. Sur chacun des domaines variables (domaines supérieurs) les CDRs sont colorés en orange (CDR1), mauve (CDR2) et rose (CDR3). Ils ont été définis selon la convention entre les feuillets B et C (CDR1), C' et C'' (CDR2) et F et G (CDR3). Les trois ponts disulfures, deux intra-chaînes et un inter-chaîne, sont colorés en jaune. Au cœur de la structure se trouve une partie de l'étiquette 6His de l'un des symétriques.

	MIC12	
	Laboratoire	Synchrotron
<i>Condition de cristallisation</i>	0,1 M Sodium Hepes pH8,2, 1,4 M Tri-Sodium Citrate dihydrate	
<i>Collecte des données</i>	1,5 M Tri-Sodium Citrate dihydrate pH8,2 + 10% glycérol	
Cryo-protectant	-	ESRF
Synchrotron	-	ID23-2
Ligne de lumière	-	ID23-2
Groupe d'espace	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
a, b, c (Å)	65,92, 73,24 119,07	66,264, 73,603 119,583
α, β, γ (°)	90, 90, 90	90, 90, 90
Résolution (Å) ^a	21,05-1,8 (1,9-1,8)	46,41-1.4 (1.4-1.4)
Longueur d'onde (Å)	1,54	0.87260
Réflexions uniques ^a	53761 (7570)	115633 (16740)
Multiplicité ^a	4,7 (4,5)	6,6 (6,5)
Complétude (%) ^a	99,3 (97,3)	100,0 (100,0)
I/σ ^a	17,3 (3,4)	11,7 (0,7)
R _{meas} (%) ^{a,b}	0,057 (0,463)	0,079 (2,658)
R _{p.i.m.} (%) ^{a,c}	0,025 (0,209)	0,031 (1,031)
CC1/2 ^d	0,999 (0,895)	0,999 (0,257)
Mosaicity	0,44	0,05
Contenu en solvant (%)	59,15	59,74
<i>Molecular replacement</i>		
Modèle de départ	3RVT	Structure résolue au laboratoire
<i>Qualité de l'affinement et du modèle</i>		
Résolution (Å)	21,05-1,8 (1,9-1,8)	45,54-1.4
Réflexions	53697 (3582)	115494 (8044)
R _{cryst} /R _{free} (%) ^e	18,28/22,4	18.94/21.10
Nombre d'atomes		
Protéine (chaîne(s) dans l'UA)	3228 (2)	3352 (2)
Eau	643	503
B-factors moyens (Å ²)		
Protéine	28,3	26.93
Eau/ion	51,745	43,605
Rmsd ^f		
Liaison (Å)	0,01	0.005
Angle (°)	1,14	0,822
Ramachandran plot (%)		
Régions les plus favorables	97,09	98.31
Régions additionnelles autorisées	2,18	1.69

Tableau 2.2.1: Statistiques de collecte des données et affinement de la structure de MIC12. Les valeurs entre parenthèses correspondent aux données de la couche de plus haute résolution.

L'affinement de la structure est actuellement terminé, bien que les valeurs de R_{work} et R_{free} soient encore élevés par rapport à la résolution de la structure. En effet, il est généralement admis que ces valeurs finales doivent être égales à environ dix fois la valeur de la résolution maximale mais le rapport final de MolProbity est satisfaisant. Les données ont été transmises à Innate Pharma et sont en attente du dépôt de brevet avant de pouvoir déposer la structure dans la banque de données protéiques RCSB.

2.3 Cristallisation des Complexes de MIC12 avec les protéines MIC

2.3.1 Protéines MIC exprimées en système bactérien

2.3.1.1 Optimisation du protocole de repliement des protéines dénaturées

Les protéines MIC ont tout d'abord été produites en système bactérien, comme préconisé par Innate Pharma. Pour cela, après obtention des constructions MICA*008 et MICB*002 portant une étiquette 6His à leur extrémité C-terminale, j'ai exprimé les trois protéines MIC en cellules *E. coli* Rosetta (DE3) pLysS (Novagen), en corps d'inclusion. Des premiers essais d'extraction, de repliement et de purification ont été réalisés suivant le protocole fourni par le partenaire économique. Après de nombreuses pertes à chaque étape, j'ai obtenu un rendement final de repliement de 1% par rapport à la quantité engagée au moment de la dilution. J'ai donc cherché à optimiser le protocole de repliement afin d'augmenter ce rendement.

La première recherche d'optimisation portait sur la méthode même de repliement. J'ai voulu comparer au total 4 méthodes : celle proposée par le protocole initial, deux méthodes de grande dilution, par criblage haut-débit d'une condition de repliement (15) et par l'utilisation d'un tampon Tris pH8 contenant un couple redox Gluthation (16), et une méthode de dialyses successives (protocole interne, Bruno Coutard). La méthode haut-débit n'ayant pas donné de condition satisfaisante, je me suis focalisée sur les trois autres. J'ai mené les trois techniques en parallèle sur une même culture de protéines MICA*001, dénaturées au préalable selon le protocole d'origine. Après purification sur chromatographie d'affinité Ni²⁺ (GE Healthcare), le meilleur rendement a été obtenu avec la méthode préconisée par Innate Pharma.

Tout en revenant au protocole initial, j'ai cherché à optimiser les étapes en amont du repliement des protéines dénaturées. En m'inspirant du protocole de Bruno Coutard, j'ai gardé l'étape de lavage des culots de corps d'inclusion incluant trois cycles de deux lavages avec et sans Triton 0,01%. Afin de jouer sur un maximum de paramètres à optimiser, j'ai réalisé un criblage basé sur trois paramètres : la méthode de lyse des cellules (sonication ou presse de french), le ratio protéine dénaturée/volume de tampon de dilution (10 mg, 20 mg et 50 mg/100 mL) et le temps de dilution (24h ou 48h). Pour cela, sur un même lot, j'ai réalisé les tests sur des fractions de 10 ml que j'ai menés jusqu'à la purification sur colonne de chromatographie d'affinité Ni²⁺ (GE Healthcare). Les résultats montrent que la sonication

est préférable à la lyse par presse de french. La meilleure condition a été définie comme étant celle avec une lyse par sonication, un ratio de 20 mg/100 mL lors d'une dilution de 24h. Cette condition a ensuite été validée par augmentation du volume à une purification à partir d'un culot bactérien de 1L. Le rendement final a été augmenté à 2,75%. Un tableau récapitulatif des étapes de repliement des protocoles d'origine et optimisés est présenté en Tableau 2.3.1

	Protocole initial	Protocole optimisé
Traitement des corps d'inclusion	1 lavage avec un tampon dénaturant + Potter	3 cycles de deux lavages
Extraction des protéines	Tampon de dénaturation urée 8M	Tampon de dénaturation Guanidinium 6M
Quantification	Nanodrop	Nanodrop + SDS-PAGE
Dilution	20mg/100mL pendant 48h	20mg/100mL pendant 24h
Dialyses	Ratio v/v 1:3 pendant 48h chacune	Ratio v/v 1:5 pendant 36 à 48h chacune

Tableau 2.3.1: Comparatif des protocoles de repliement des protéines MIC "coli".

La quantification des protéines dénaturées a très vite posé problème. Trois méthodes étaient possibles : la quantification par mesure d'absorbance au NanoDrop 2000 (Thermo Scientific), sur gel et l'utilisation d'un kit Pierce BCA protein assay (Thermo Fisher Scientific). Les deux premières n'étant pas idéales avec le tampon de dénaturation en urée 8M et la troisième étant assez longue, j'ai opté pour un tampon dénaturant en Guanidinium, permettant un dépôt sur SDS-PAGE. La quantification a donc été faite par dépôt sur gel avec comparaison à une gamme de BSA, couplée à une estimation par mesure au NanoDrop 2000 (Thermo Scientific). Avec l'expérience et l'optimisation de la verrerie (bêchers de plus gros volumes et agitateurs magnétiques dédiés par exemple), le rendement de repliement a augmenté de 1% au départ à 8-10% pour les trois protéines MIC.

2.3.1.2 Vérification du bon repliement des protéines MIC

Après repliement et purification des protéines MICA*001, MICA*008 et MICB*002, leur bon repliement est vérifié par Dichroïsme circulaire (CD). Les spectres dichroïques des mesures d'ellipticité moyenne par résidu en fonction de la longueur d'onde entre 207 et 260 nm sont représentés en Figure 2.3.1.

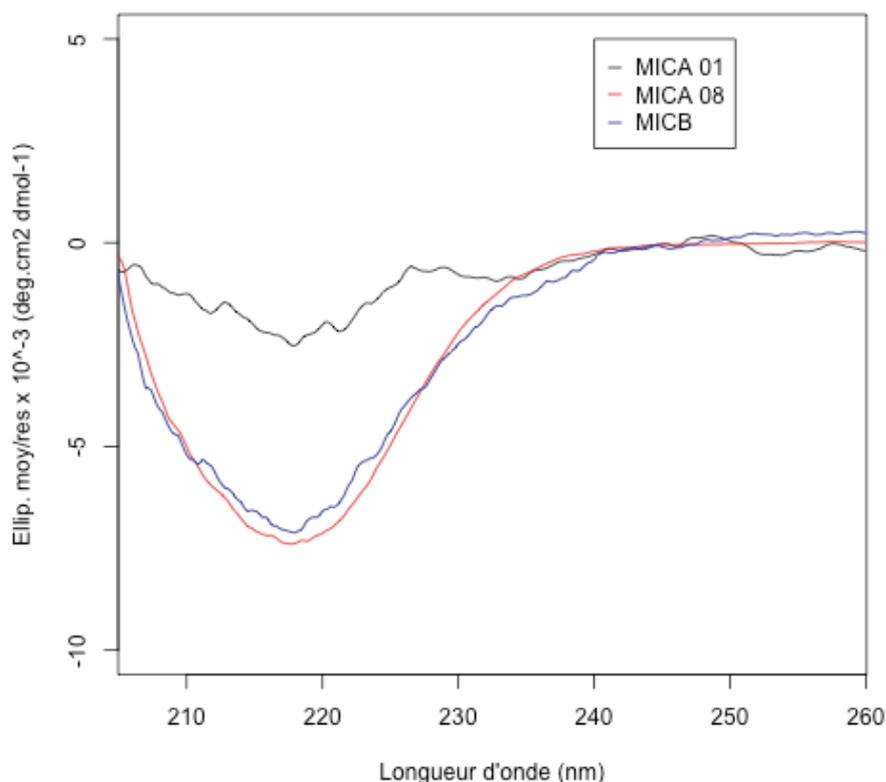


Figure 2.3.1: Spectres de Dichroïsme circulaire des protéines MIC repliées. Les courbes des protéines MICA*001, MICA*008 et MICB*002 sont représentées respectivement en noir, rouge et bleu. Toutes ont un minimum vers 217 nm.

Les profils des trois protéines sont caractéristiques de protéines contenant majoritairement des structures secondaires de type feuillet β , avec un seul minimum spécifique vers 217 nm. L'analyse par le serveur Dichroweb confirme cette tendance. Ces résultats coïncident avec le contenu en structures connu pour ces protéines et ne correspondent pas à des profils de protéines dénaturées, ce qui confirme leur repliement. Aussi, la différence d'ellipticité moyenne de MICA*001 avec les deux autres protéines révèle la mauvaise estimation de la concentration au moment de la mesure.

Une seconde vérification a été faite par analyse sur Size Exclusion Chromatography-Multi Angles Light Scattering (SEC-MALS) pour les protéines MICA*008 et MICB (Figure 2.3.2). Les poids moléculaires calculés sont proches des poids moléculaires théoriques attendus pour chacune des protéines. Cette estimation étant basée sur la détermination du rayon de giration, ces poids moléculaires ne seraient pas si proches si les protéines n'étaient même que partiellement repliées. Là encore, les problèmes d'estimation

de la concentration de MICA*001 n'ont pas permis l'obtention de résultats correctes au moment de l'analyse.

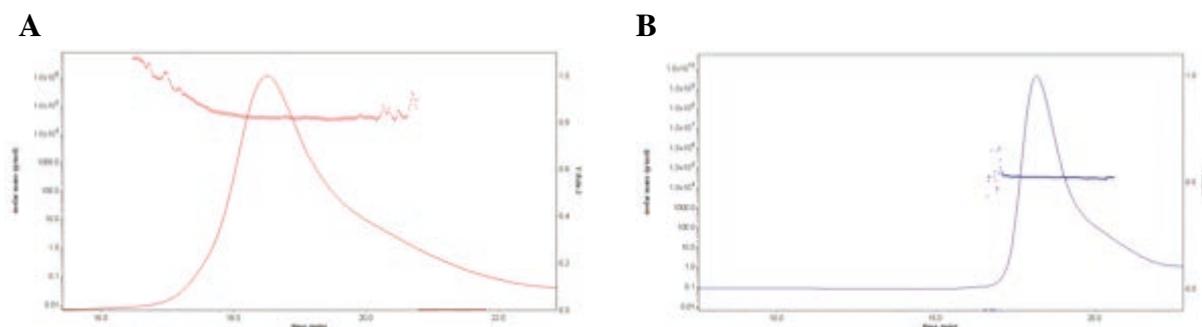


Figure 2.3.2: Analyse par SEC-MALS des protéines MICA*008 et MICB "coli". Colonne Shodex KW-802.5. A : Analyse de MICA*008. Le pic de protéine a été élué à un volume de 18,5 mL et la masse moléculaire a été estimée à 32 kDa. B : Analyse de MICB. La protéine a été élué à un volume de 18,8 mL et la masse moléculaire a été estimée à 33 kDa.

2.3.1.3 Mesures de l'interaction de MIC12/MIC

Bien que Innate Pharma ait mesuré l'interaction du Fab MIC12 avec différents allèles de MICA et MICB, nous voulions vérifier que celui-ci reconnaisse bien les protéines MICA*001, MICA*008 et MICB après repliement. Pour cela, j'ai fait appel à la technique de BioLayer Interferometry (BLI), en service à la plateforme biophysique du laboratoire. Christine Kellenberger, co-responsable de cette plateforme, a réalisé l'expérience.

Les résultats sont représentés dans le Tableau 2.3.2. Les Kd résultants des mesures des constantes d'association et de dissociation sont toutes de l'ordre ou meilleures que du nanomolaire. Cela confirme des interactions très fortes de ce Fab avec les protéines repliées et concorde avec les résultats obtenus par Innate Pharma.

	kass (M)	kdiss (M)	Kd (M)	χ^2	R ²
MICA*001	9×10^4	8.8×10^{-5}	$5 - 9.6 \times 10^{-10}$	2.7	0.97
MICA*008	9×10^5	$< 1 \times 10^{-7}$	$< 10^{-12}$	3.3	0.88
MICB*002	1×10^5	8×10^{-4}	$4 - 7 \times 10^{-9}$	1.8	0.90

Tableau 2.3.2 : Mesures par BLI de l'interaction MIC/MIC12. kass : constante de vitesse d'association ; kdiss : constante de vitesse de dissociation ; Kd : constante de dissociation.

2.3.1.4 Purification des complexes MIC12/MIC « E. coli »

Une fois les protéines MIC purifiées, le Fab MIC12 est ajouté pour la mise en complexe. Pour les trois complexes, les pics correspondant au complexe et au Fab seul se chevauchent en sortie de chromatographie d'exclusion de taille, ce qui peut gêner la cristallisation. Les chromatogrammes correspondant à la dernière étape de purification sont montrés en Figure 2.3.3 pour chaque complexe. Pour le complexe MICA*001/MIC12

(Figure 2.3.3A) il s'agit d'une sortie d'échangeuse d'ions cations, pour les deux autres complexes (Figure 2.3.3B et C) d'une sortie de chromatographie d'exclusion de taille. Les pics de complexes résultants sont homogènes et ne semblent pas être contaminés. Le dépôt sur SDS-PAGE (Figure 2.3.3D) le confirme avec des bandes uniques de MIC12 (vers 45 KDa) et de protéines MIC (vers 35 KDa).

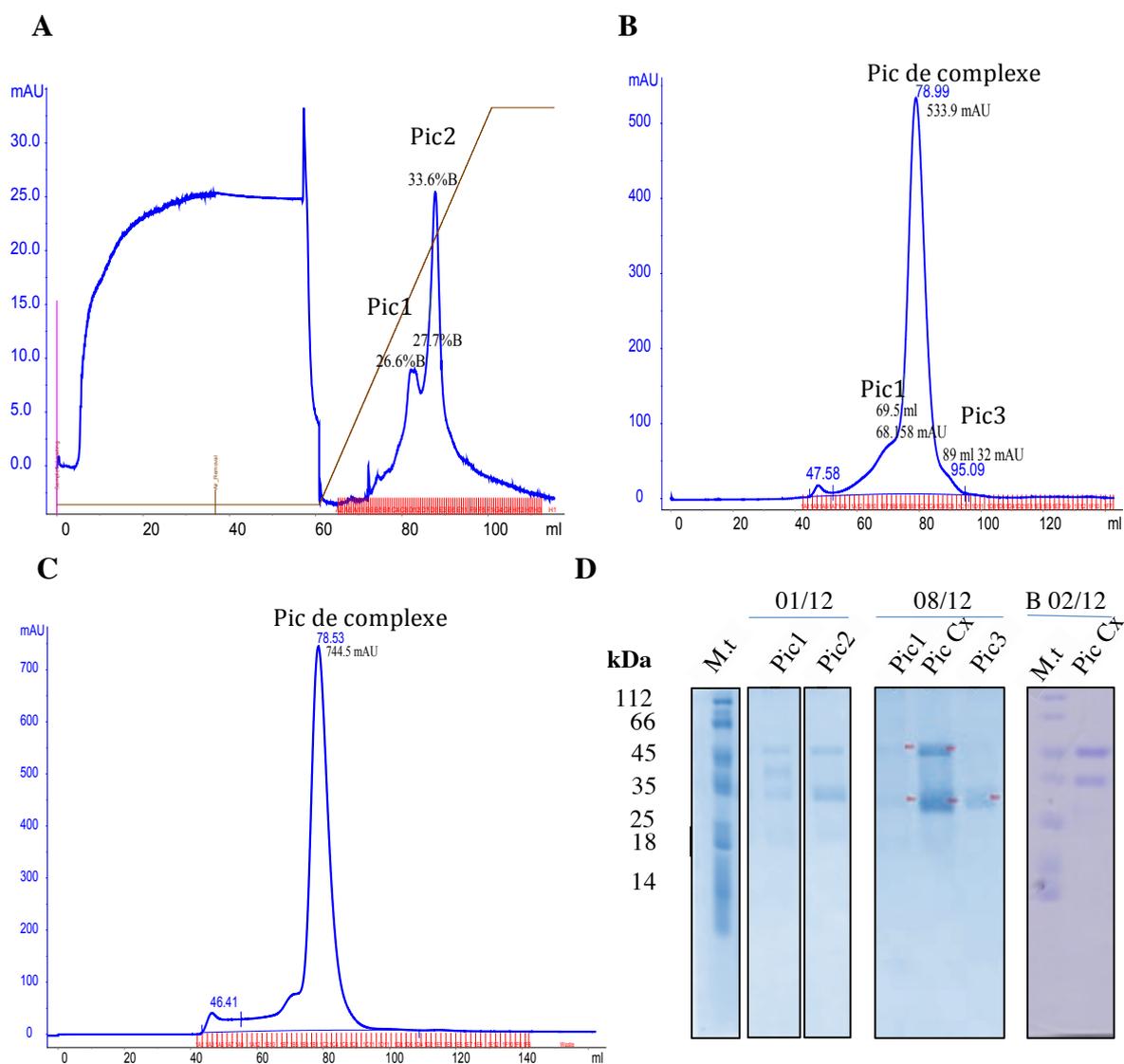


Figure 2.3.3: Purification des complexes "E. coli". A: Echangeuse d'ion cations du complexe MICA*001/MIC12. Un premier pic est élué à un pourcentage du tampon B d'élution de 26,6-27,6%, le second est élué à un pourcentage de 33,6%. B: Chromatographie d'exclusion de taille du complexe MICA*008/MIC12. Le pic de complexe est élué après un volume de tampon de 79 ml. C: Chromatographie d'exclusion de taille du complexe MICB*002/MIC12. Le pic de complexe est élué après un volume de tampon de 78,5 ml. D: SDS-PAGE des étapes finales de purification des trois complexes MIC. M.t: Marqueur de taille; 01/12: complexe MICA*001/MIC12; 08/12: complexe MICA*008/MIC12; B 02/12: complexe MICB*002/MIC12; Pic Cx: pic de complexe.

2.3.1.5 Cristallographie des complexes MIC12/MIC « E. coli »

La cristallogénèse des complexes a d'abord été réalisée à une concentration de 5 mg/mL puis a été ajustée au cas par cas, principalement suite aux résultats d'un screen Stura.

Aucune condition de cristallisation satisfaisante n'a été trouvée pour le complexe MICB*002/MIC12 mais des cristaux ont été obtenus pour MICA*001/MIC12 et MICA*008/MIC12 (Figure 2.3.4).

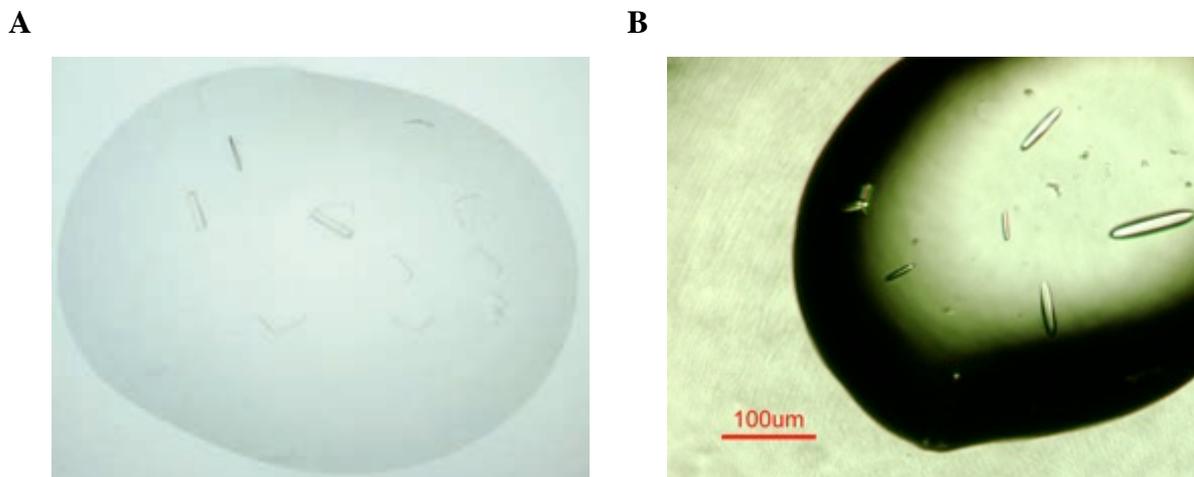


Figure 2.3.4: Cristaux des complexes MIC12/MIC « *E. coli* ». **A:** Cristaux du complexe MICA*001/MIC12. La condition de cristallisation est constituée de 1,32 M de sodium citrate et 10 mM de sodium tetra-borate pH 9 (screen Stura). **B:** Cristaux du complexe MICA*008/MIC12, apparus dans la condition 0,1 M d'acétate pH 4,5, 0,8 M de NaH_2PO_4 et 1,2 M de KH_2PO_4 (screen Wizard).

Les cristaux du complexe MICA*008/MIC12 (Figure 2.3.4B) sont apparus dans le screen général Wizard après 44 jours. La condition de cristallisation est composée de 0,1 M d'acétate pH 4,5 et d'un mélange de 0,8 M de NaH_2PO_4 et 1,2 M de KH_2PO_4 . La diffraction de ces cristaux a été testée sur une ligne de lumière synchrotron (ID30B, ESRF) mais aucune diffraction n'a pu être observée. Les cristaux restant de la goutte ont alors été utilisés pour faire du Matrix Micro-seeding (MMS, (17)) mais au bout de deux cycles d'optimisation les cristaux résultats se sont avérés être du sel.

Concernant le complexe MICA*001/MIC12, les cristaux sont apparus après 1 jour dans le screen général Stura en position B6. La condition, 1,32 M de Sodium citrate et 10 mM de Sodium Tetra-borate pH 9, a été optimisée et les cristaux obtenus (Figure 2.3.4A) sont apparus en 3 jours. Après avoir été cryoprotégés avec 10% de glycérol ajoutés à la liqueur mère, les cristaux ont été testés sur la ligne synchrotron ID23-2 (ESRF). Le meilleur jeu de données collecté a diffracté à une résolution de 7,5 Å. Malgré des essais d'optimisation, de MMS et même de criblage d'additifs, je n'ai jamais réussi à avoir des cristaux de meilleure diffraction.

2.3.2 Protéines MIC exprimées en système baculovirus/cellules d'insectes

Dans la littérature, les protéines recombinantes MIC sont exprimées à partir de deux systèmes d'expression : bactérien ou en cellules d'insectes grâce au baculovirus. Afin de mettre toutes les chances de notre côté pour la résolution des structures des complexes MIC12/MIC, j'ai également tenté ce deuxième type d'expression.

2.3.2.1 Purification des complexes MIC/MIC12 « baculo » :

Les protéines MICA*001, MICA*008 et MICB*002 avec une étiquette Strep ont été purifiées à partir des surnageant de culture, puisque sécrétées. Le fort taux de biotine contenu dans le milieu de culture utilisé (147 mg/L) ne permet pas de directement charger le surnageant sur une colonne Step. En effet, la biotine se fixe de manière irréversible sur le ligand Streptactin et sature définitivement la colonne. Bien que la biotine puisse être chélatée par ajout d'Avidine, j'ai opté pour une stratégie de mise en complexe directement en sortie d'infection par ajout du Fab MIC12 dans le surnageant. MIC12 étant porteur d'une étiquette His, la purification a été faite par chromatographie d'affinité Ni²⁺ (GE Healthcare) puis par chromatographie d'exclusion de taille.

Les profils UV₂₈₀ des premières productions en sortie de cette dernière étape de purification montrent de gros écart de rendement de production (Figure 2.3.5). Alors que pour les complexes impliquant MICA*001 et MICB*002 un pic vers 77 mL et un autre vers 83 mL sont visibles, le complexe MICA*008/MIC12 ne comprend que ce dernier pic, correspondant à MIC12 libre. Ce résultat laisse à penser que MICA*008 n'a pas été exprimée et qu'il y aurait un problème avec son virus, ou bien que la fixation avec MIC12 ne se soit pas faite. La présence des deux partenaires dans le pic de complexe a été vérifiée sur SDS-PAGE avec une bande vers 45 kDa pour MICA*001, une vers 40 kDa pour MICB*002 et toujours les deux bandes à 25 et 23 kDa pour MIC12. Cette différence de poids moléculaire apparent vis à vis des protéines produites en bactéries est due à la glycosylation ayant lieu en production par des cellules d'insectes. Les rendements moyens sont de 132 µg/400 mL pour le complexe MICA*001/MIC12 et de 810 µg/1L pour le complexe MICB*002/MIC12. Malheureusement, l'efficacité des virus a vite chuté et seules quelques infections ont pu être menées avant une perte totale d'expression.

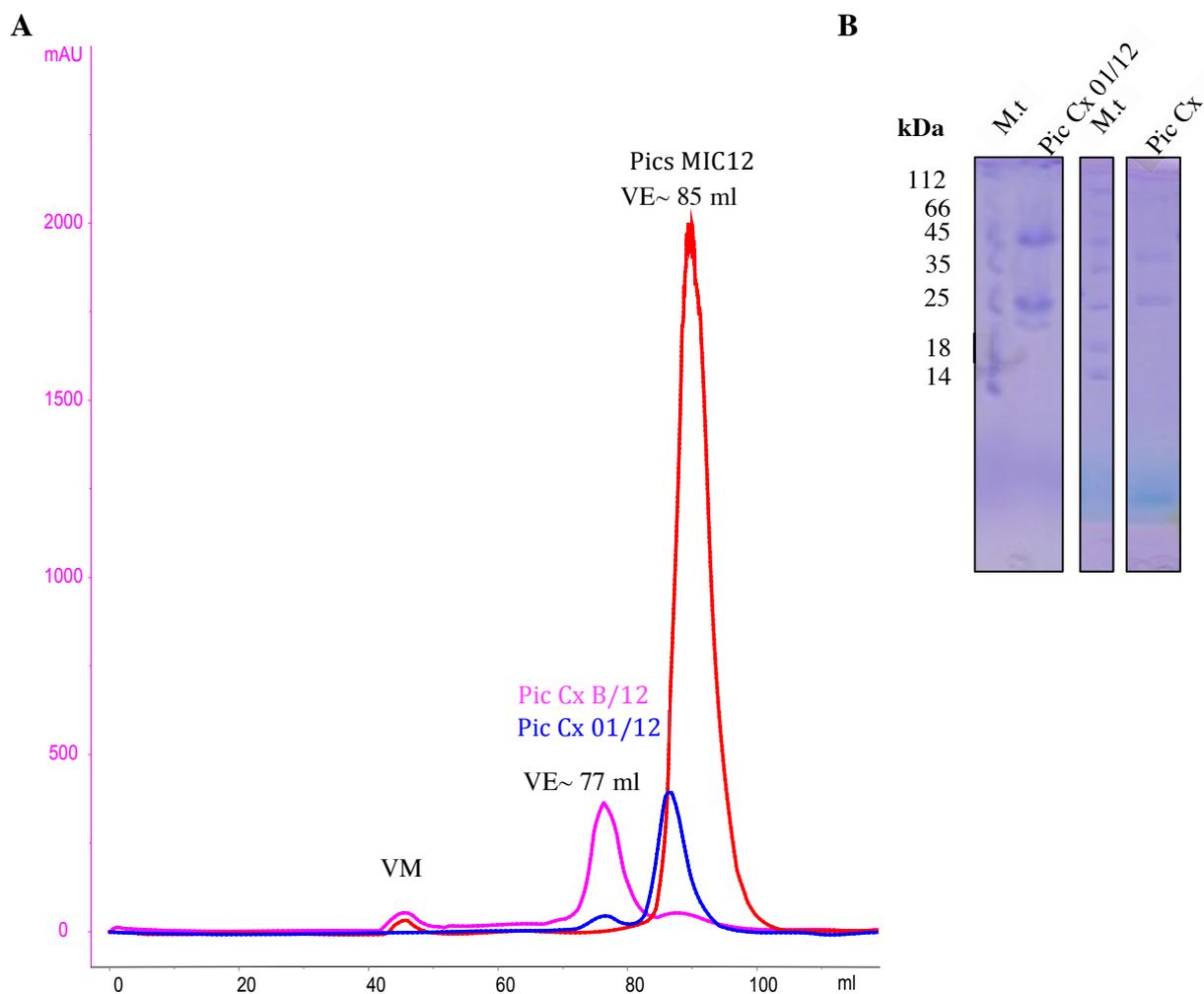


Figure 2.3.5 : Purification des complexes "baculo". A : Chromatogrammes des sorties de colonne Superdex 200 16/60 des trois complexes. Les pics correspondant aux complexes sont élués après environ 77 ml de tampon. Les pics d'excès de MIC12 sont élués après un volume autour de 85 ml. Les courbes de purification des complexes incluant MICA*001, MICA*008 et MICB*002 sont colorées respectivement en bleu, rouge et rose. VM: volume mort de la colonne.; VE: volume d'élution; Pic Cx 01/12 : pic du complexe MICA*001/MIC12; Pic Cx B/12 : pic du complexe MICB*/MIC12. B : Analyse des fractions sur SDS-PAGE. M.t : Marqueur de taille.

2.3.2.2 Cristallographie des complexes « baculo » :

Des essais de cristallogenèse ont tout de même pu être réalisés. Pour le complexe MICA*001/MIC12, les rendements de purification n'ont permis de ne faire qu'une boîte du screen général Stura à 3,78 mg/mL. Aucune condition de cristallisation ne s'en est dégagée.

Pour le complexe MICB*002/MIC12, les screens généraux Stura et MDL ont été testés à une concentration en complexe de 5,4 mg/mL. Des cristaux sont apparus après un jour dans les conditions C3 (0,2 M d'imidazole malate pH 7,7, 17,5% de PEG 10K) et D3 (0,2 M d'imidazole malate pH 7,7, 22,5% de PEG 10K) du screen MDL (Figure 2.3.6). Ces

cristaux ont été optimisés selon la condition 0,2M d'imidazole malate pH 7,2-8,2, 12,5%-27,5% de PEG 10K.

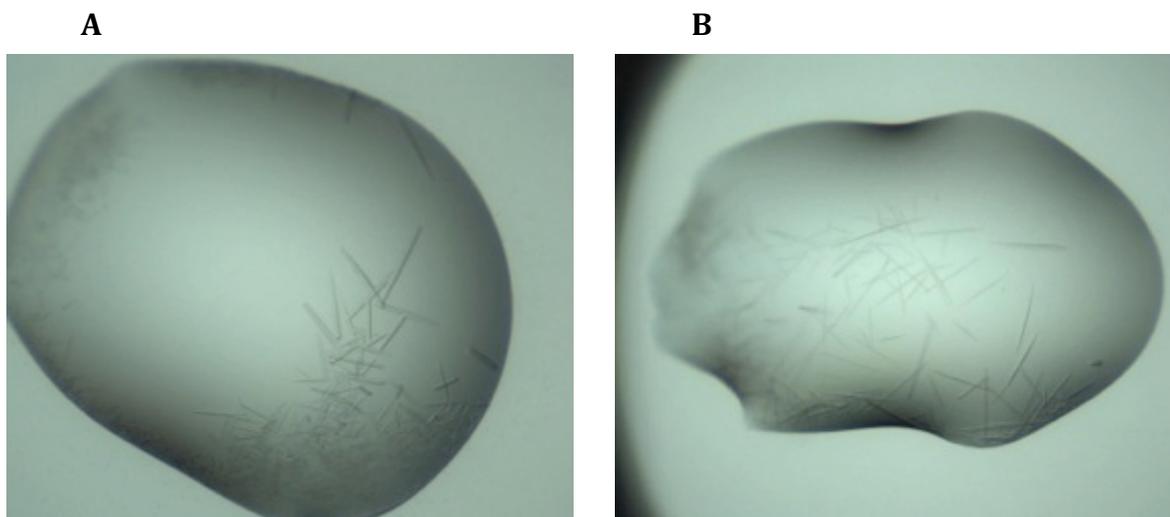


Figure 2.3.6 : Cristaux du complexe "baculo" MICB*002/MIC12. A: Condition C3 du screen MDL, goutte 200/100 nl. B: Condition D3 du screen MDL, goutte 100/100 nl.

Des cristaux sont apparus dans la boîte après seulement quelques heures. Je les ai ensuite cryoprotégés en ajoutant 10% de glycérol à la solution du réservoir et en y transférant les cristaux avant la congélation sous flux d'azote. Ces cristaux n'ont pas diffracté lors de leur test au laboratoire mais ont diffracté jusqu'à 10 Å sur la ligne Proxima 2 (SOLEIL, Gif-sur-Yvette). Bien que ces résultats soient très encourageants, ils n'ont pu être poursuivis dû à la perte d'efficacité du virus.

2.3.3 Protéines MIC exprimées en cellules de mammifères

Les protéines MIC sont, à la base, des allèles des protéines MICA et MICB que l'on retrouve chez l'Homme, donc en cellules de mammifères. Suite à la perte des baculovirus et en parallèle des expressions en système bactérien, ce nouveau système d'expression a été envisagé. Les modifications post-traductionnelles engendrées par l'expression en cellules de mammifères, comme par exemple les N- et O- glycosylations ou phosphorylations, peuvent agir sur la stabilité des protéines et ainsi aider la cristallogénèse des complexes MIC/MIC12. Les glycosylations, par exemple, sont différentes de celles que l'on peut retrouver chez les cellules d'insectes et peuvent jouer un rôle dans l'interaction Fab/protéine.

À la fin de ma thèse, Innate Pharma a réussi à développer des lignées stables de cellules CHO exprimant les protéines MICA*001 et MICA*008. Les cultures des cellules se sont poursuivies chez eux pour des raisons de confidentialité et les surnageants nous ont été

remis. Des collègues de l'équipe travaillant également sur des projets en collaboration avec Innate Pharma (Béatrice Amigues, Vincent Delauzun et Anaïs Gaubert) ont pris en charge la purification de ces protéines, celles des complexes et leur cristallo-genèse. Une stratégie de déglycosylation a été développée après la formation des complexes. Ceci permet à la fois de conserver les glycosylations éventuelles entrant en jeu dans l'interaction mais aussi d'éliminer le plus de parties flexibles. Celles-ci peuvent en effet empêcher la formation de contacts cristallins. Ces essais sont actuellement en cours et des cristaux sont régulièrement testés au synchrotron sur Proxima 2 (SOLEIL).

2.4 Etudes des complexes MIC/MIC12 par SAXS

Pour compléter les données biochimiques et biophysiques de l'interaction MIC/MIC12 ainsi que les futures données structurales cristallographiques, j'ai mené des études de diffusion de rayons X aux petits angles SAXS.

Les protéines et complexes étudiés sont MICA*001, MICA*008, MIC12, MICA*001/MIC12 et MICA*008/MIC12. Les données, collectées sur la ligne BioSAXS (BM29, ESRF), ont toutes pu être traitées à l'exception de celles de la protéine MICA*008. Je n'ai pu superposer aucune des courbes des différentes concentrations, allant de 0,6 à 7,21 mg/ml, probablement dû à des dommages aux radiations (Figure 2.4.1A). Les autres données quant à elles sont stables indépendamment des concentrations, permettant la création de courbes chimères (Figure 2.4.1B).

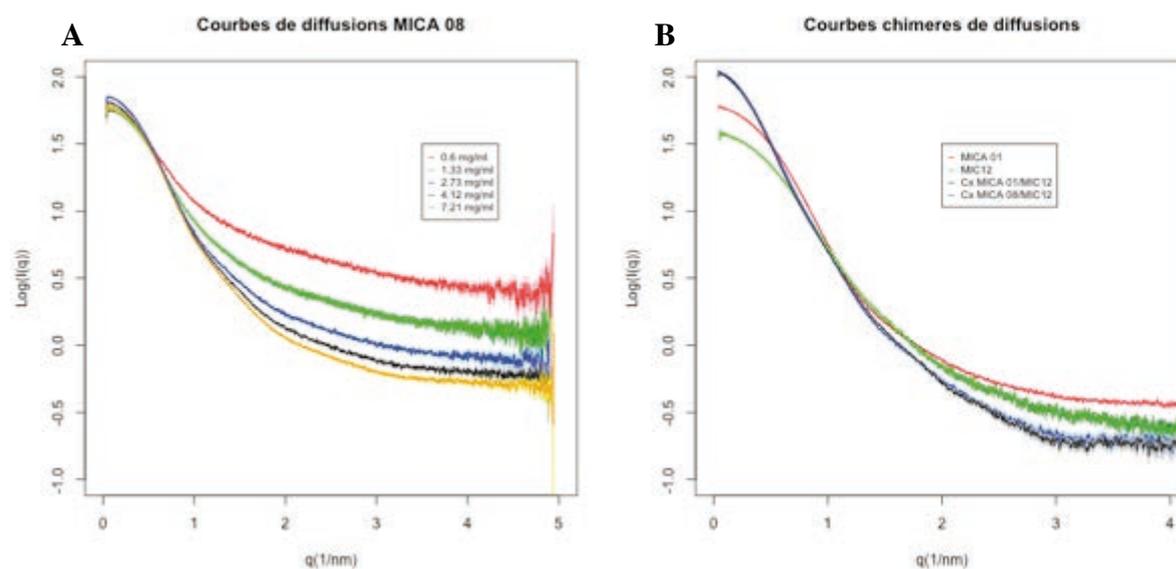


Figure 2.4.1: Courbes de diffusion SAXS des protéines MIC. A : Courbes de diffusions de MICA*008, aux différentes concentrations. **B :** Courbes chimères de diffusions des protéines MICA*001 (rouge), MIC12 (vert), MICA*001/MIC12 (bleu) et MICA*008/MIC12 (noir).

À ce stade, il peut déjà être noté que les courbes chimériques correspondantes aux deux complexes MIC sont presque confondues. L'analyse de Guinier de ces courbes (Figure 2.4.2A) a été possible sur chacune, montrant ainsi que les échantillons ne sont pas agrégés. Les paramètres d'intensité de diffusion extrapolée à une valeur de vecteur diffusé $q=0$ (I_0) et de rayon de giration R_g , estimés à partir de cette analyse, ont été reportés dans le Tableau 2.4.1.

Une analyse de la flexibilité intrinsèque des protéines a également été possible par l'observation des plots de Kratky sans dimensions (Figure 2.4.2B). L'allure des courbes seules et des complexes sont similaires deux à deux. Les valeurs en y des courbes de MICA*001 et MIC12 décrivent un pic suivi d'une remontée, ce qui est typique de protéines contenant des éléments de flexibilité. Ces résultats sont cohérents avec les structures connues des protéines MICA et de MIC12 : toutes deux contiennent des domaines flexibles car reliés par de longues séquences non structurées. Les complexes, quant à eux, sont plus stables avec un pic suivi d'un plateau ne descendant pas à zéro. Ce profil est typique de protéines structurées contenant des parties flexibles.

La distribution des distances est ensuite analysée à l'aide de GNOM (18) (Figure 2.4.2 C et D). Les paramètres I_0 et R_g estimés à partir de l'analyse de distribution des distances par paires (ou Porod) sont reportés dans le Tableau 2.4.1 tout comme la distance maximale D_{max} . Là encore, les complexes MICA*001/MIC12 et MICA*008/MIC12 ont une distribution des distances fortement similaire et une D_{max} quasi-identique. Le profil suivi par la courbe correspondant à MIC12 est typique d'une protéine de forme globulaire, ce qui est cohérent avec sa structure résolue par cristallographie. C'est également le cas pour MICA*001, bien qu'ayant un profil plus allongé. Les complexes, quant à eux, ont une courbe de Porod faisant penser à une forme plus allongée, moins globulaire.

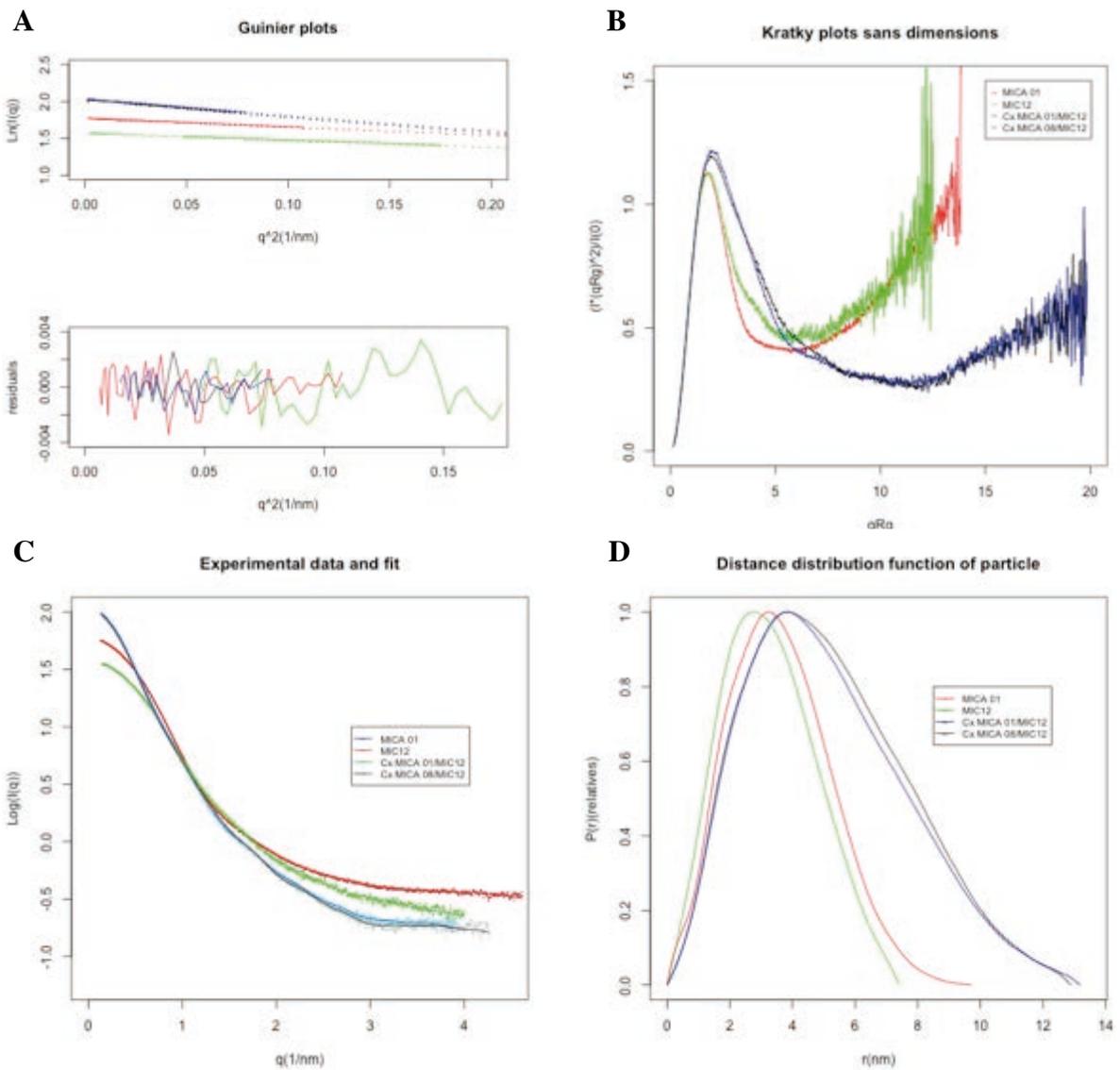


Figure 2.4.2: Traitement des données SAXS des protéines MIC à l'aide de la suite PRIMUS. A: Analyse de Guinier. **B:** Plots de Kratky sans dimensions. **C et D:** Analyse des distributions des distances par paires. Sur chacune de ces représentations, les échantillons de MICA*001, de MIC12, du complexe MICA*001/MIC12 et du complexe MICA*008/MIC12 sont colorés respectivement en bleu, rouge, vert et noir.

	MICA 01	MICA 08	MIC12	MICA 01/MIC12	MICA 08/MIC12
Paramètres de collecte des données					
Instrument	BM29 (ESRF)	BM29 (ESRF)	BM29 (ESRF)	BM29 (ESRF)	BM29 (ESRF)
Détecteur	Pilatus (1M)	Pilatus (1M)	Pilatus (1M)	Pilatus (1M)	Pilatus (1M)
Géométrie du faisceau	4*4	4*4	4*4	4*4	4*4
Longueur d'onde (nm)	0.09919	0.09919	0.09919	0.09919	0.09919
q range (nm ⁻¹)	0.041-4.942	0.041-4.942	0.041-4.942	0.041-4.942	0.041-4.942
Temps d'exposition (s)	1	1	1	1	1
Gamme de concentration (g/L)	0.33-8.9	0.6-7.21	0.68-4.64	0.44-11.17	0.49-14.11
Température (K)	283	283	283	283	283
Paramètres structuraux					
I(0) (cm ⁻¹) (P(r))	59.24	-	37.20 +/- 0.32	108.5 +/- 0.07	105.3 +/- 0.06
I(0) (cm ⁻¹) (Guinier)	59.47 +/- 0.04	-	37.44 +/- 0.05	108.83 +/- 0.08	106.36 +/- 0.12
Rg (Å) (P(r))	28.0	-	25.36 +/- 1.66	40.33 +/- 0.03	40.54 +/- 0.03
Rg (Å) (Guinier)	28.4 +/- 0.6	-	25.4 +/- 0.7	40.0 +/- 0.7	40.9 +/- 0.9
D _{max} (Å) (P(r))	97	-	75	130.6	128.9
D _{max} (Å) (Structure cristallographique)	100 (1B3J)	-	79	-	-
Masse moléculaire (M) (kDa)					
M théorique	32.695	32.749	47.755	80.450	80.504
M SAXS MoW2 ³	35.358	-	27.921	124.853	119.862
Programmes utilisés					
Réduction des données	PRIMUS	PRIMUS	PRIMUS	PRIMUS	PRIMUS
Analyse des données	GNOM		GNOM	GNOM	GNOM
Validation des modèles	CRYSOL		CRYSOL	CRYSOL DAMMIF	DAMMIF

Tableau 2.4.1 : Paramètres des données SAXS des protéines MIC. Les valeurs ont été estimées pour chaque échantillon à partir des courbes chimériques.

Les D_{max} évaluées par cette dernière analyse sont en accord avec celles des structures résolues par cristallographie aux rayons X (Tableau 2.4.1), ce qui est encourageant quant à l'utilisation de ces données pour la création et validation de modèles structuraux.

Puisque MICA*001 et MIC12 sont des protéines plutôt flexibles dû à d'assez longs linkers inter-domaines, et que MICA*001 est issue d'une production en système eucaryote, donc glycosylée, une comparaison de ces données SAXS avec les structures cristallographiques n'aurait pas de sens. En revanche, même s'il s'agit de modèles imprécis

et hypothétiques, les enveloppes que peut générer le logiciel DAMMIF (19) peuvent donner une idée de la forme globale des deux complexes.

Les meilleures enveloppes obtenues, évaluées sur la base de la valeur du χ^2 des fits des courbes théoriques sur les points expérimentaux, sont représentées en Figure 2.4.3.

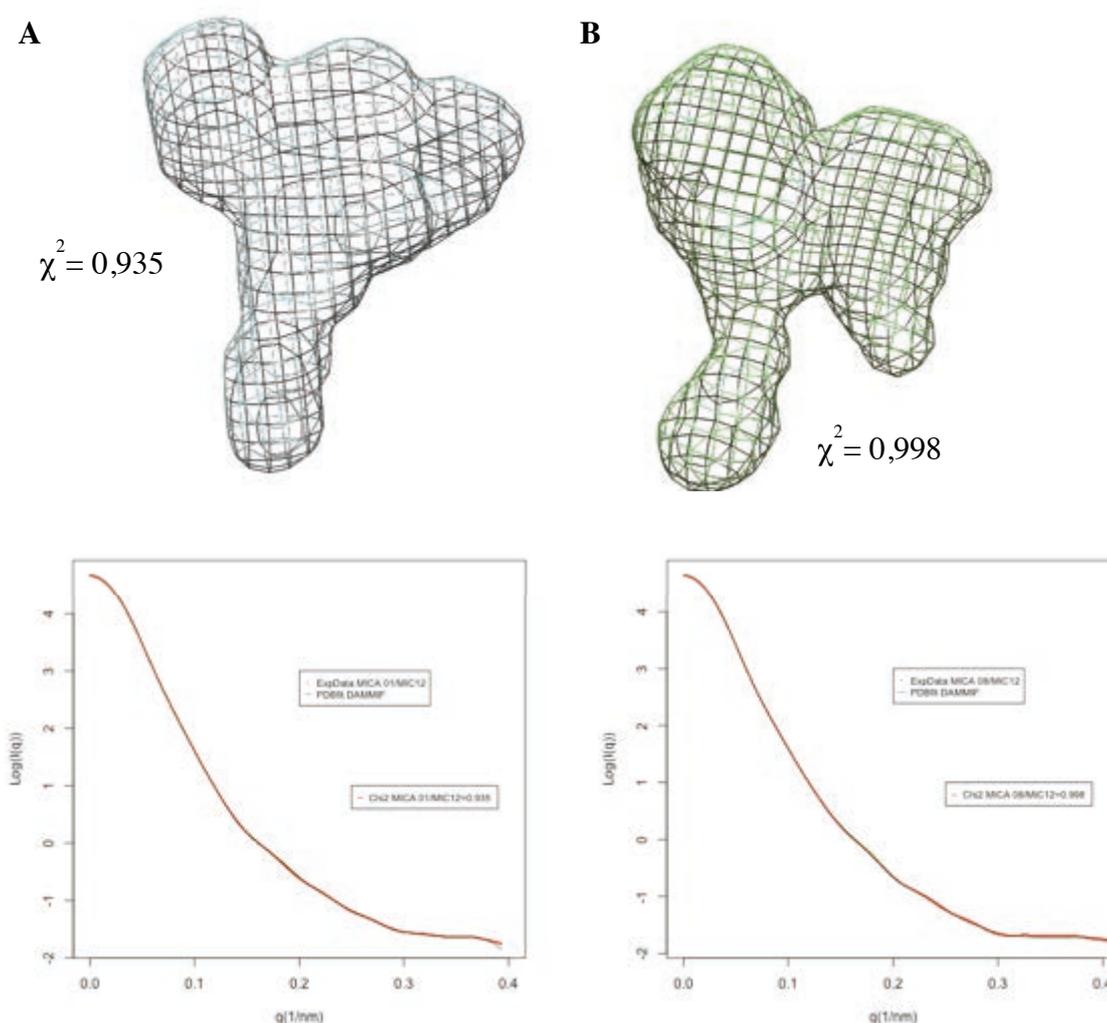


Figure 2.4.3: Enveloppes des complexes MICA*001/MIC12 et MICA*008/MIC12 générées à l'aide de DAMMIF. A: Enveloppe et fit du complexe MICA*001/MIC12. B: Enveloppe et fit du complexe MICA*008/MIC12.

2.5 Docking moléculaire du complexe MICA*001/MIC12 et proposition d'un modèle

Des précédentes études menées par Innate Pharma ont permis l'identification de résidus dans l'interaction MIC/MIC12 (14). En l'attente de données structurales atomiques pouvant les confirmer et étendre la zone d'interaction, des prédictions d'interaction peuvent

être faites à l'aide de logiciels de docking. Les modèles prédits pourront ensuite être validés, ou du moins triés, par combinaison du docking avec les données SAXS.

Le docking du complexe MICA*001/MIC12 a été réalisé à partir de la structure de MICA*001 contenue dans la structure du complexe avec NKG2D (1HYR) ainsi qu'à partir de la structure résolue de MIC12. Le serveur ClusPro 2.0 (<https://peptidock.cluspro.org/login.php>) a généré 29 modèles pdb, chacun représentatif d'un cluster de modèles. A l'issue d'un tri selon l'implication des résidus définis dans le brevet, les modèles 4 et 8 ont été retenus (Figure 2.5.1).

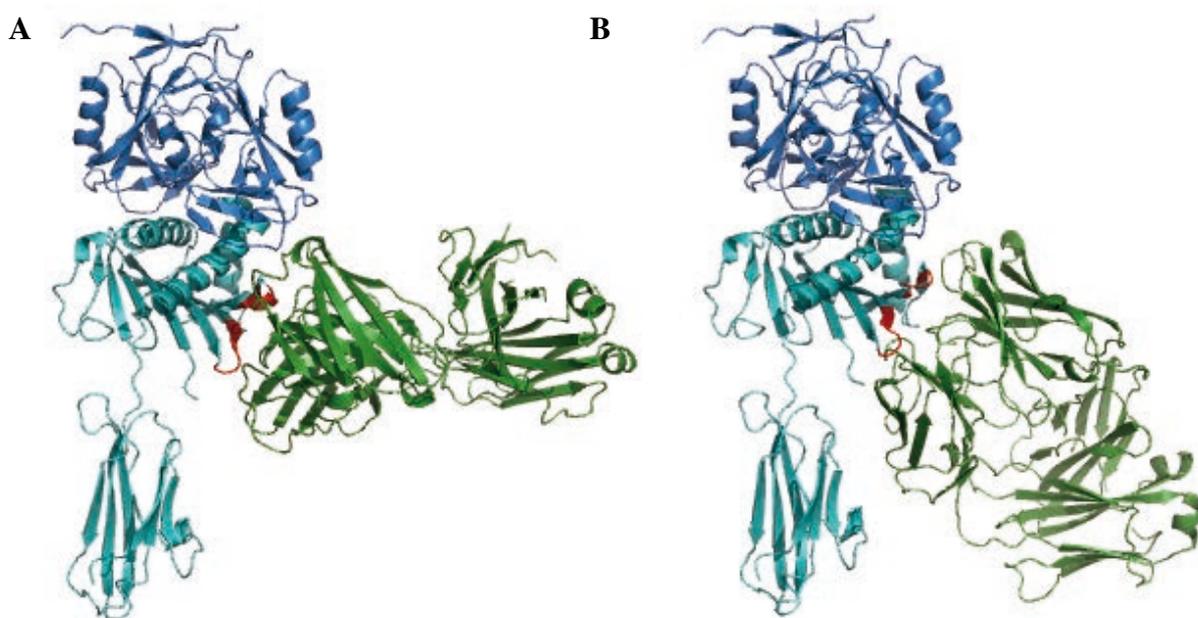


Figure 2.5.1 : Modèles du complexe MICA*001/MIC12, générés par ClusPro. A: Modèle 4. B: Modèle 8. Les protéines MICA*001, MIC12 et NKG2D sont colorées respectivement en cyan, vert et bleu foncé. Les résidus de MICA*001 identifiés dans l'interaction par Innate Pharma sont en rouge.

Le Fab MIC12 étant supposé empêcher l'interaction NKG2D/MIC par compétition, les modèles sont discriminés selon la présence ou non de clashes stériques avec le récepteur. Le modèle 4 a alors été sélectionné comme modèle le plus probable.

Afin de comparer ce modèle avec les données obtenues par SAXS, celui-ci a été introduit dans l'enveloppe générée par DAMMIF (Figure 2.5.2A). Malgré le fait que le modèle du complexe rentre dans cette enveloppe, des régions restent non comblées. Le logiciel CRY SOL, qui génère une couche d'hydratation autour du modèle et le compare aux données expérimentales SAXS, donne en sortie une superposition des courbes de diffusion théorique et expérimentale avec la valeur du test χ^2 . Celle-ci est ici de 151,40, ce qui est bien

loin des valeurs que devrait prendre ce test pour une superposition parfaite (χ^2 autour de 1). Puisque la protéine MICA*001 ayant servie à l'analyse par SAXS est glycosylée par son expression en cellules de mammifères, des glycosylations ont été modélisées à la surface de celle-ci à l'aide du logiciel GLYCOSYLATION de la suite ATSAS. Les sites potentiels de glycosylation proposés sur MICA*001 sont de 33, comprenant les N-glycosylation et O-glycosylation (uniquement sur les sérines). Le modèle 4 glycosylé, introduit dans l'enveloppe SAXS, est représenté en Figure 2.5.2B. Bien que la valeur du χ^2 issu de l'analyse par CRY SOL soit encore bien loin d'une valeur idéale de 1, les quatre glycosylations ajoutées ont largement contribué à l'amélioration de ce paramètre.

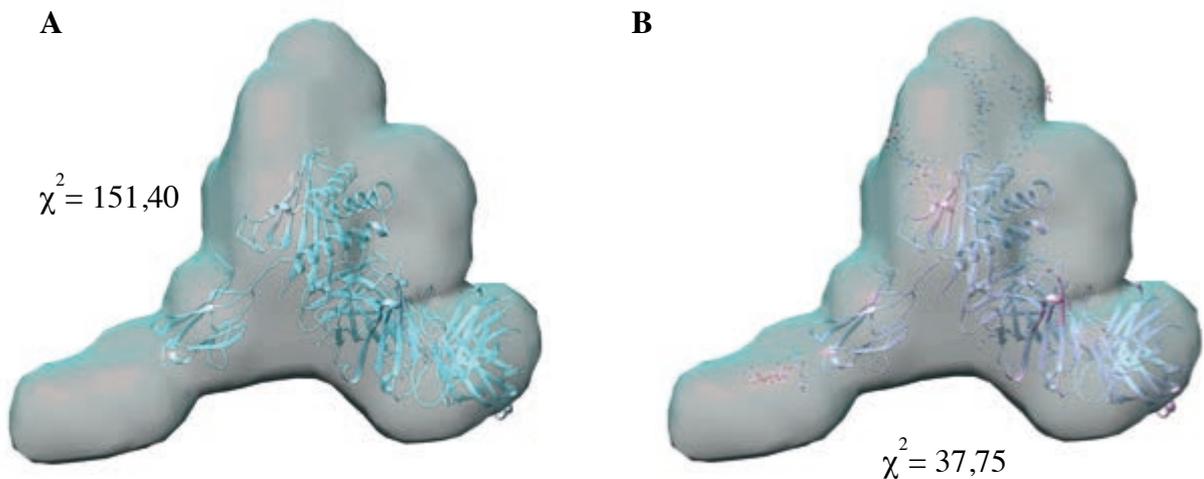


Figure 2.5.2: Validation par SAXS du modèle du complexe MICA*001/MIC12 proposé par docking. A: Modèle 4 intégré à l'enveloppe SAXS. La valeur du χ^2 du CRY SOL est de 151,40. **B:** Modèle 4 glycosylé intégré à la même enveloppe. La valeur du χ^2 du CRY SOL est de 37,75.

2.6 Discussions et conclusions

L'objectif de ce sujet était de caractériser au niveau atomique l'interaction du Fab MIC12 avec ses antigènes MICA*001, MICA*008 et MICB. Pour cela, des études structurales ont été menées par cristallographie aux rayons X, par diffusion de la lumière aux petits angles et par docking moléculaire protéine/protéine, et cela sur des protéines exprimées aussi bien en système procaryote qu'eucaryote.

La structure cristallographique de MIC12 a été déterminée à une résolution de 1,4 Å. Les conformations des CDRs constituant le paratope ont ainsi pu être définies et transmises à Innate Pharma, qui peut maintenant s'appuyer sur des données réelles pour leurs nouveaux tests et mutagenèses dirigées de l'anticorps.

Les essais de cristallographie ont permis l'obtention de cristaux des trois complexes MICA*001/MIC12, MICA*008/MIC12 et MICB/MIC12. Leur diffraction atteint jusqu'à présent une résolution maximale de 7 Å, ce qui reste insuffisant pour permettre la résolution d'une structure par cristallographie. Différentes stratégies ont été mises en place afin d'améliorer la qualité des cristaux.

Au niveau biochimique, des tampons optimums de stabilité des protéines et complexes ont été définis. Les glycosylations apportées par la production en système eucaryote ont été digérées afin d'éliminer toutes parties flexibles pouvant empêcher la formation de contacts cristallins. Aussi, des étapes de purification supplémentaires ont été ajoutées, afin de s'assurer de la pureté des échantillons utilisés pour la cristallogénèse.

Au niveau de la cristallogénèse, différentes méthodes ont été employées pour l'amélioration des conditions de cristallisation définies. Tout d'abord, chacune de ces conditions a été optimisée selon une gamme de pH et de concentration en agent précipitant. Différentes concentrations en protéines ont été étudiées et différentes températures de cristallogénèse ont été testées. Ensuite, des conditions de cristallisation ont été testées comme additifs à la condition de cristallisation définie. Un découplage des étapes de nucléation et de croissance cristalline a également été tenté par l'utilisation de la technique d'ensemencement à partir de nucléis ainsi que du MMS. Malgré l'amélioration visuelle des cristaux, la résolution de la diffraction reste aux alentours de 7 Å.

Différentes options restent encore à tester, comme par exemple le changement de la méthode de cristallisation. La diffusion de vapeur est la méthode la plus utilisée mais il en

existe d'autres telles que la diffusion contraire en micro-capillaires ou bien la cristallisation sous gouttes d'huile. Ces méthodes présentent l'avantage de limiter ou éviter le phénomène de convection et permettrait ainsi une cristallisation plus homogène.

L'expression des protéines MIC en système baculovirus avait donné des résultats intéressants, malheureusement avortés par la perte d'efficacité des virus. Cette expression a été reprise par des collègues (Béatrice Amigues, Vincent Delauzun et Anaïs Gaubert) dans les derniers mois de ma thèse. De nouveaux virus ont été obtenus, plus stables, permettant un réel criblage de conditions de cristallisation.

La principale limitation à l'amélioration des cristaux, dans leur qualité et diffraction, est, selon moi, la flexibilité des protéines et complexes. Les études menées par SAXS ont clairement montré que des parties flexibles sont présentes chez MICA*001 et MIC12 mais également dans les complexes, bien que plus stables. Il existe d'ailleurs des études structurales de la flexibilité de MICA*001 entre la « plateforme » (formée par les sous-unités $\alpha1$ et $\alpha2$) et la sous-unité $\alpha3$ (11). Sous sa forme libre MICA*001 adopte une conformation « ouverte » (1B3J) et se « referme » lorsqu'elle est en interaction avec son récepteur NKG2D (1HYR). Pour prendre un exemple concret, en superposant les deux structures, l'angle formé par les résidus Q83, T181 et Q251 prend pour chacune des structures les valeurs de $142,1^\circ$ pour la forme libre et $106,4^\circ$ pour celle complexée (Figure 2.6.1).

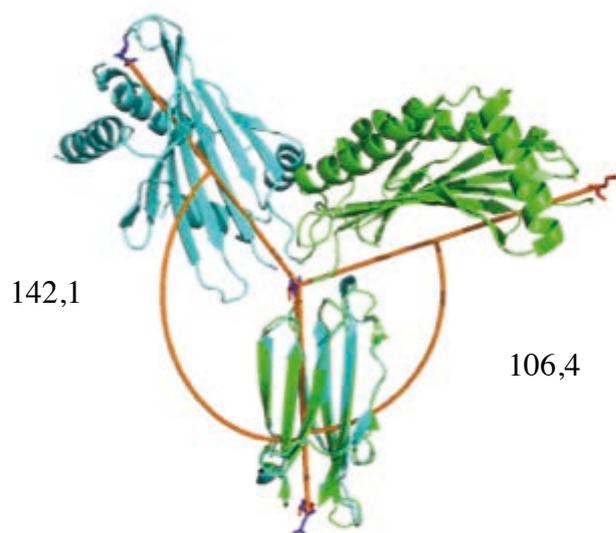


Figure 2.6.1: Flexibilité de la protéine MICA*001. La conformation "ouverte" (1B3J) est colorée en bleu et la conformation "refermée" (1HYR) est colorée en vert. Les deux structures ont été alignées par superposition des domaines $\alpha3$. L'angle formé par les résidus Q83, T181 et Q251 a été mesuré pour chacune et sont représentés en orange.

La conformation « refermée » induite par NKG2D est certes plus contrainte, elle n'en est cependant pas pour autant unique et reste certainement flexible. La fixation du Fab MIC12 ne semble pas non plus contraindre totalement cette flexibilité et ne va d'ailleurs peut-être pas induire cette même conformation refermée. Il est donc nécessaire de trouver un moyen de limiter cette flexibilité.

Une idée serait d'utiliser un deuxième anticorps non compétitif avec MIC12, comme il a déjà été fait par exemple pour la protéine TLR3(20). Innate Pharma dispose d'autres anticorps fixant les protéines MIC permettant de tester cette stratégie et doit prochainement nous en faire parvenir.

Une autre option serait de revoir les constructions MIC en elles même en éliminant la sous-unité $\alpha 3$. En effet, seules les sous-unités $\alpha 1$ et $\alpha 2$ semblent participer à l'interaction avec MIC12. Cette version tronquée permettrait ainsi d'obtenir un complexe bien plus compacte et plus apte à former un réseau cristallin solide.

En parallèle de ces études structurales par cristallographie aux rayons X, des modèles structuraux du complexe MICA*001/MIC12 ont été prédits par docking moléculaire. Un mode optimisé pour des interactions anticorps/antigène a été utilisé avec également un mode SAXS. Le premier mode permet de ne considérer que les CDRs comme épitopes de l'interaction en masquant le reste de l'anticorps tandis que le second mode filtre les clusters de modèles en les comparant aux données SAXS implémentées. Parmi les modèles représentatifs des clusters générés, un modèle semble cohérent avec les résultats préliminaires obtenus par Innate Pharma (14). Il respecte à la fois les conditions de résidus de MICA*001 impliqués dans l'interaction avec MIC12 et de compétition à la fixation du récepteur NKG2D. De plus, ce modèle rentre dans l'enveloppe proposée par le logiciel de DAMMIF, bien que des régions ne soient pas comblées. Ceci peut s'expliquer de différentes manières : 1) le traitement des données, 2) les glycosylations de MICA*001 et 3) la flexibilité du complexe. Concernant la flexibilité du complexe, je suis évidemment convaincue de sa contribution à cette enveloppe, bien que d'un impact bien moins important que celui des glycosylations. Les positions des sites potentiels de glycosylations semblent coïncider avec ces régions vides et les quelques glycosylations que j'ai pu générer avec le programme GLYCOSYLATION semblent nettement améliorer la concordance du modèle avec l'enveloppe. Des études SAXS sur des protéines exprimées en système eucaryotes puis déglycosylées pourraient aider à confirmer cette implication.

Cependant, il faut garder à l'esprit qu'il s'agit bien d'une prédiction et non d'une structure issue de données expérimentales. Il faut également garder un œil critique sur les enveloppes obtenues avec des modélisations ab-initio à partir de données SAXS, qui sont dépendantes de la personne traitant les données. L'expérience et la formation entrent beaucoup en jeu, principalement lors de l'analyse des distributions de distances (Porod) et peu de paramètres permettent d'estimer la qualité du traitement. Le contrôle des résultats par des analyses utilisant le programme CRY SOL est donc important, puisque ne dépendant pas de l'analyste.

Bien que je n'ai pu résoudre dans le temps qu'il m'était imparti les structures des complexes MIC/MIC12, des cristaux qui diffractent ont été obtenus dans plusieurs conditions pour chaque complexe. Les recherches se poursuivent dans l'équipe : Béatrice Amigues, Vincent Delauzun et Anaïs Gaubert ont pris la suite du projet et les idées évoquées plus haut vont être initiées. Le modèle prédit par le docking du complexe MICA*001/MIC12 a permis de cartographier la zone potentielle d'interaction et des études s'appuyant sur ces résultats pourraient être menées.

3 Utilisation de nanobody pour les études structurales de la protéine pPorM

3.1 Introduction

Les bactéries évoluent dans des milieux très divers. Que ce soit pour se mouvoir, trouver les nutriments nécessaires à leur métabolisme, se défendre ou attaquer, elles doivent sans cesse interagir avec leur environnement. L'un des moyens dont elles disposent pour cela est l'export de protéine de leur cytoplasme vers un autre compartiment, qu'il soit intracellulaire, extracellulaire ou dans une cellule cible. Ce processus de sécrétion de protéines est exercé par divers mécanismes, impliquant très souvent les systèmes de sécrétion (21).

3.1.1 Les systèmes de sécrétion bactériens

Les systèmes de sécrétions sont des machines bactériennes dédiées à la sécrétion de protéines. Ils peuvent, selon les types, servir à la sécrétion d'une ou plusieurs protéines spécifiques et n'être, selon leurs rôles et importance, présents uniquement chez quelques espèces bactériennes ou au contraire communs à un ensemble de procaryotes. Ces systèmes existent aussi bien chez les bactéries à Gram positif que négatif. Différentes classes de systèmes de sécrétion sont répertoriées selon le nombre de membranes qu'ils permettent aux protéines de traverser, allant de une à trois, à savoir les membranes interne et externe et la membrane d'un autre organisme.

J'ai choisi de ne traiter ici que des machines de sécrétion présentes chez les bactéries Gram-négatif.

3.1.1.1 Les machineries SEC et Tat

Les machineries de sécrétion SEC « General SECRetion pathway » et Tat « Twin arginine translocation » sont les deux voies majeures de sécrétion de protéines positionnées dans la membrane cytoplasmique (22). Ce sont également les deux machines les plus utilisées pour réaliser le passage de protéines à travers cette membrane, et cela aussi bien chez les bactéries, chez les archae que chez les eucaryotes.

La machinerie SEC transfère des protéines dans leur état déplié. Ces protéines sont d'abord prises en charge par une protéine cargo pour être amenées à une protéine moteur qui

va permettre leur passage à travers la membrane cytoplasmique par la translocase SecYEG (23). Selon la séquence de leur peptide signal, les protéines peuvent être adressées au processus de transfert SecB ou SRP (Figure 3.1.1). Lorsqu'elles sont destinées à aller dans le périplasma ou à être sécrétées dans le milieu extracellulaire, les protéines sont reconnues par la protéine SecB qui, tout en conservant l'état déplié de la protéine, va la transporter jusqu'à la membrane. La protéine SecA sert alors d'ATPase et permet le passage de la protéine à travers SecYEG. Avant d'y entrer, la protéine voit son peptide signal clivé et se replie une fois sortie (Figure 3.1.1A). Pour les protéines à destination de la membrane cytoplasmique, le ribosome se déplace jusqu'à la translocase où la protéine est prise en charge dès la sortie du ribosome par la protéine SRP. Ceci évite le contact des domaines hydrophobes avec le milieu cytoplasmique (24). La SRP recrute la protéine FtsY qui amène le complexe protéine/ribosome à SecYEG pour y transférer la protéine (23). Au cours de la traduction, les domaines transmembranaires sortent par le côté du canal pour s'insérer dans la bicouche lipidique (Figure 3.1.1B).

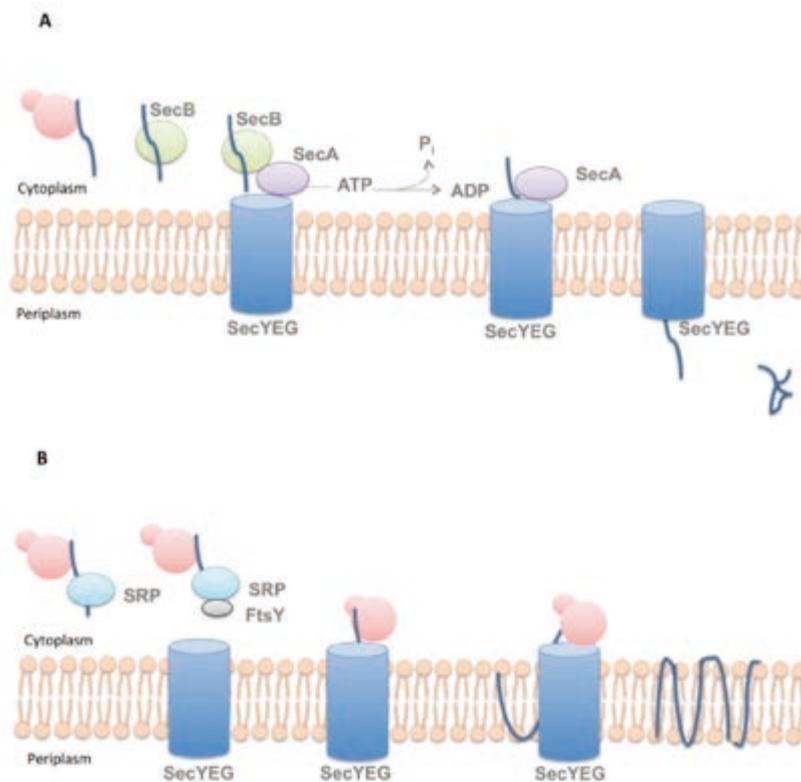


Figure 3.1.1: Export de protéines à travers la machinerie SEC (21). A : Export de protéines vers le périplasma ou milieu extérieur par le processus de transfert SecB. La protéine (bleu) encore dépliée à la sortie du ribosome (rose) est reconnue grâce à la séquence peptide signal par la protéine SecB (vert). Celle-ci sert de chaperonne en la maintenant dans cet état. Elle la transporte jusqu'à la protéine moteur SecA qui, après hydrolyse d'un ATP, la fait passer à travers la translocase SecYEG. La protéine adopte alors un état replié. B : Export de protéines dans la membrane cytoplasmique par le processus de transfert SRP. Dès sa sortie du ribosome, la protéine est reconnue par la protéine SRP qui recrute la protéine FtsY. Cette dernière amène le complexe à la translocase et la protéine naissante sort par le côté du canal pour se placer dans la membrane.

La machinerie Tat, quant à elle, transfère des protéines repliées, ne pouvant l'être hors milieu cytoplasmique (Figure 3.1.2). Les protéines TatB et TatC (ne formant qu'une seule et même protéine, multidomaine, chez les bactéries Gram-positif), reconnaissent un motif « twin » arginine à l'extrémité N-terminale de la protéine sortie du ribosome et s'y fixent. Elles amènent alors cette protéine à la membrane cytoplasmique où sont recrutées des protéines TatA qui viennent former le canal transmembranaire, au travers duquel passe la protéine.

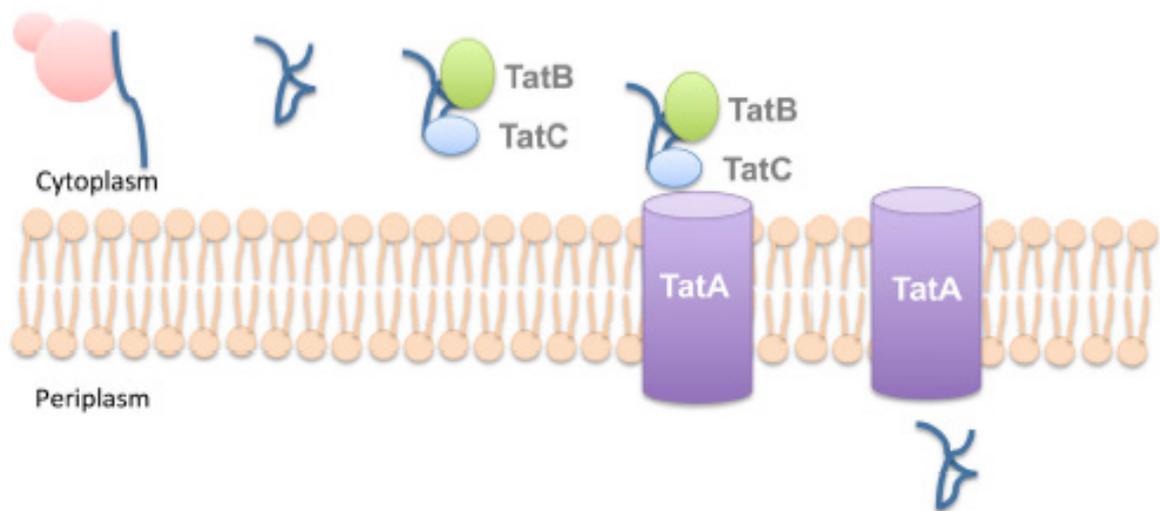


Figure 3.1.2: Export de protéines par la machinerie Tat (21). Une fois la protéine traduite et séparée du ribosome (ici en rose), elle est repliée dans le cytoplasme puis prise en charge par les protéines TatB et TatC (chez les bactéries gram-négatif). Elles transitent jusqu'à la membrane cytoplasmique où des protéines TatA sont recrutées et assemblées pour former le translocon. La protéine traverse alors la membrane en passant par ce canal.

3.1.1.2 Les systèmes de sécrétions chez les bactéries Gram-négatif

Les bactéries Gram-négatif possèdent différents types de systèmes de sécrétion permettant l'export de protéines de leur cytoplasme aux milieux périplasmique, extracellulaire ou directement dans le cytoplasme d'autres cellules. Certains systèmes de sécrétion permettent le transport de protéines en une seule étape à travers les membranes interne et externe, tels que le système de sécrétion de type 1 (T1SS), le système de sécrétion de type 3 (T3SS), le système de sécrétion de type 4 (T4SS) et le système de sécrétion de type 6 (T6SS), et sont désignés comme systèmes de sécrétion de protéines Sec ou Tat indépendants. D'autres, tels que le système de sécrétion de type 2 (T2SS), sont Sec ou Tat dépendants et nécessitent un premier export des protéines via ces machineries jusqu'au périplasme avant de pouvoir les exporter vers le milieu externe (Figure 3.1.3).

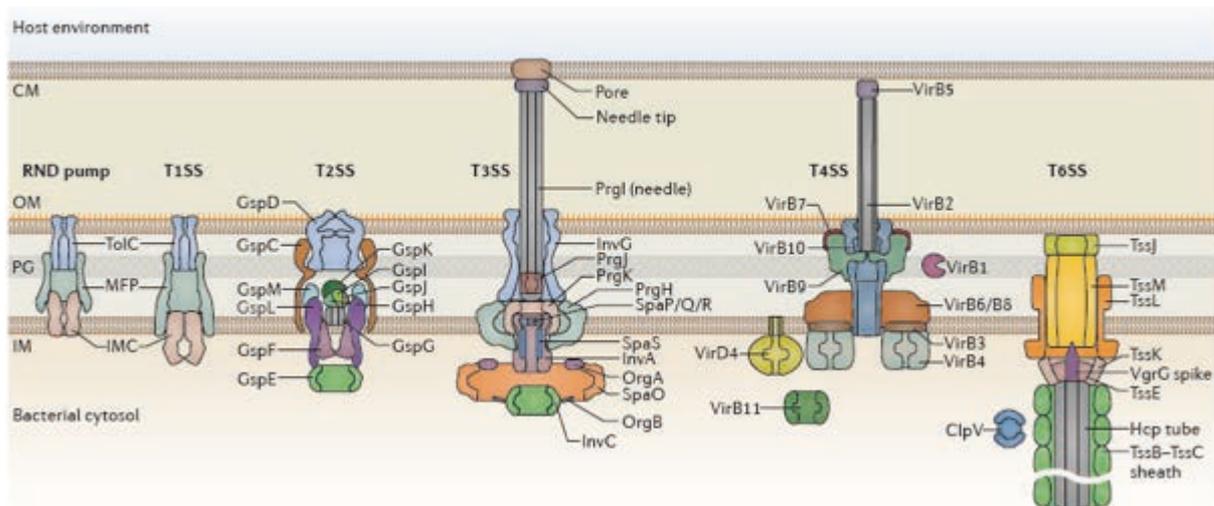


Figure 3.1.3 : Représentation schématique des systèmes de sécrétion chez les bactéries Gram-négatif (25). Les systèmes T1SS, T2SS, T3SS, T4SS et T6SS sont décrits ainsi que la pompe RND, en comparaison avec le T1SS. IM : membrane interne (ou cytoplasmique) ; OM : membrane externe ; PG : couche de polyglycans ; CM : membrane de cellule hôte.

3.1.1.2.1 Le système de sécrétion de type I (T1SS)

Ce premier système de sécrétion est présent chez de nombreuses bactéries, notamment chez des bactéries pathogènes de plantes ou d'animaux. Il s'agit d'un système permettant la sécrétion de protéines dépliées en une étape du cytoplasme au milieu externe. La plupart de ces substrats sont impliqués dans la nutrition ou la virulence de la bactérie (25). Parmi eux se trouvent des protéases, des lipases, des adhésines, des protéines se liant à l'hème (telle que la protéine HasA), des protéines contenant des motifs répétés RTX (par exemple les protéines HlyA ou CyaA) ou bien des protéines de la couche de surface cristalline S comme la protéine RsaA (26). Ces substrats peuvent contenir en C-terminal une séquence signalée reconnue par le T1SS.

Côté architectural, le T1SS est similaire aux pompes « résistance-nodulation-division » (RND) (Figure 3.1.4). Il est composé de trois parties : une protéine transporteur « ATP-Binding Cassette » ABC insérée dans la membrane cytoplasmique, des protéines de fusion à la membrane (MFP) reliant les éléments de la membrane cytoplasmique et externe, et une partie externe constituée de trois facteurs de membrane externe (OMP), formant le canal traversant la membrane externe.

Le transporteur ABC présente différentes fonctions : il catalyse la phosphorylation de l'ATP pour donner l'énergie nécessaire au fonctionnement du système, reconnaît le substrat et interagit avec les MFP. Il existe trois classes de transporteurs, selon leur séquence N-terminale : ceux possédant un domaine peptidase C39 permettant le clivage de la partie

N-terminale des substrats, ceux possédant un domaine « C39 peptidase like » (CLD) qui ne peuvent réaliser ce clivage, et les transporteurs ne possédant pas cette région N-terminale.

Selon la taille et la séquence des protéines à sécréter, celles-ci peuvent être prises en charge par l'un de ces trois types de transporteurs ABC. Le canal de la membrane interne vient alors se sceller sur le canal de la membrane externe et après catalyse de l'ATP, les protéines traversent la membrane interne puis externe en une seule étape (Figure 3.1.4).

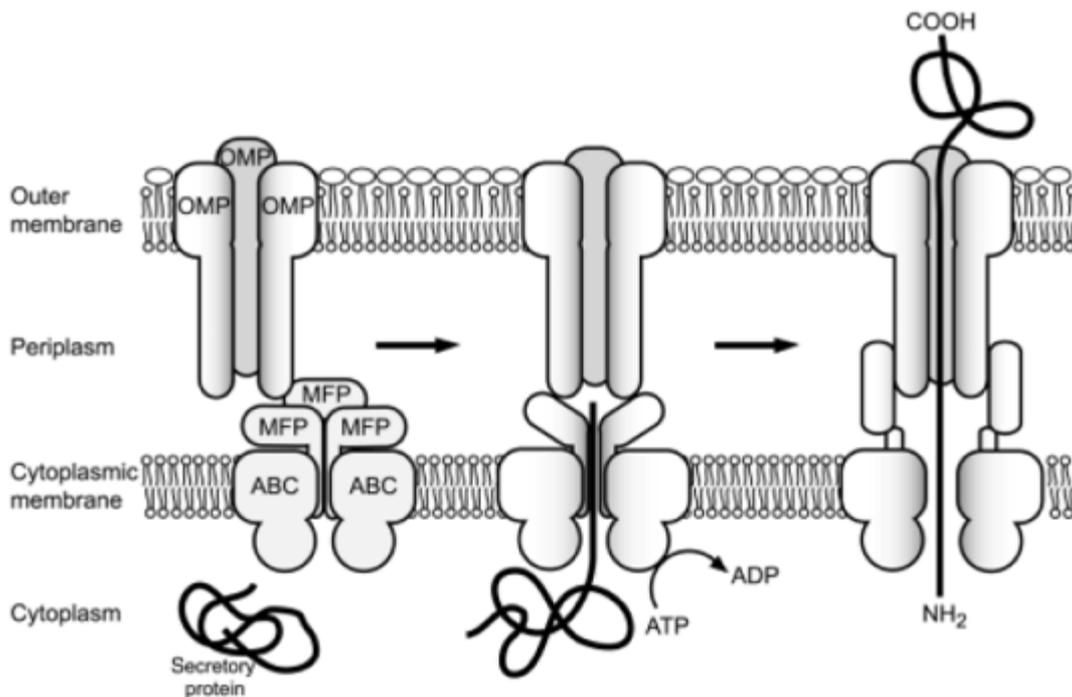


Figure 3.1.4: Description schématique de l'assemblage et fonctionnement du T1SS (27). La protéine à sécréter est reconnue par le transporteur ABC qui va alors recruter le canal formé par les protéines OMP. Après catalyse d'un ATP en ADP, le canal formé par le transporteur s'ouvre, laissant le passage à la protéine qui traverse les membranes interne et externe dans un état déplié.

3.1.1.2.2 Le système de sécrétion de type II (T2SS)

Le T2SS est un système de sécrétion conservé chez la plupart des bactéries Gram-négatif, quelles soient pathogènes ou non (25). Il est notamment utilisé par la bactérie dans des processus d'adaptation à l'environnement ou de transport de facteurs de virulence. Il permet la sécrétion de protéines repliées en deux étapes : le passage des protéines du cytoplasme au périplasme est réalisé à l'aide d'une machinerie SEC ou Tat puis le T2SS effectue le passage du périplasme au milieu extracellulaire.

Ce système comprend 12 éléments essentiels à la biogénèse et sécrétion des protéines (28) : la sécrétine de la membrane externe (GspD), la plateforme de la membrane interne (comprenant les protéines GspC, GspF, GspL et GspM), l'ATPase cytosolique (GspE), le

pseudo pilus (GspG, GspH, GspI, GspJ et GspK) et la peptidase prépiline GspO (Figure 3.1.5).

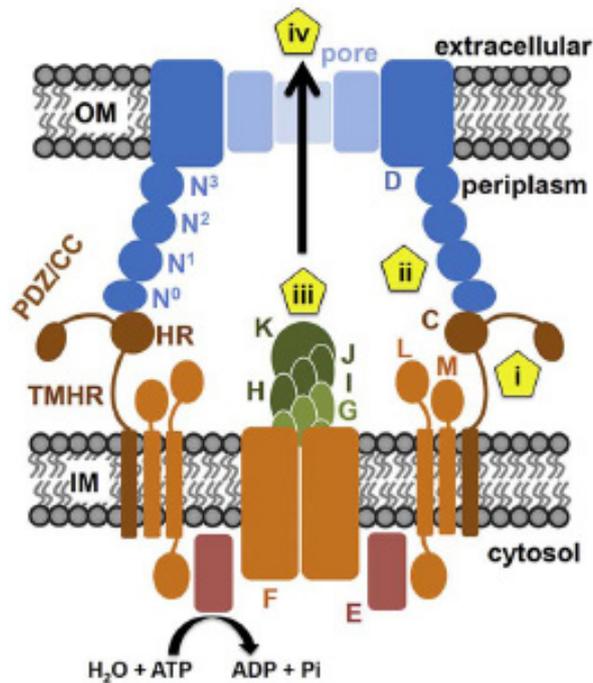


Figure 3.1.5: Représentation schématique du modèle actuel du T2SS (28). Les protéines sont indiquées par la dernière lettre de leur nom (ex : GspJ=J) et le code couleur des lettres correspond à celui des protéines. TMHR désigne un lien flexible et HR une région structurée d'homologie entre plusieurs espèces bactériennes. PDZ/CC : domaines PDZ ou coiled-coil présents à l'extrémité C-terminale de GspC chez certaines espèces. Les éléments notés en jaune de i à iv représentent les étapes de sécrétion des protéines. Les protéines, une fois dans le périplasm, vont être recrutées par GspC et placées dans le vestibule par action combinée de GspC et GspD. Elles sont ensuite placées à l'extrémité du pilus et l'ATPase GspE est recrutée pour l'hydrolyse d'un ATP. L'énergie produite fait grandir le pilus et éjecte la protéine hors de la bactérie.

Les protéines sont d'abord exportées du cytoplasme vers le périplasm par les machineries Sec ou Tat comme décrit plus haut. Le repliement de la protéine avant la prise en charge par le T2SS est nécessaire et doit être réalisé dans le périplasm pour les protéines issues du mécanisme SEC. Les protéines sont ensuite reconnues par l'extrémité de la protéine GspC via leurs régions hautement flexibles ou dépliées lors de leur transition conformationnelle. Un réarrangement est parfois nécessaire pour les substrats les plus volumineux afin de permettre leur entrée dans le « vestibule » du T2SS. Les protéines GspC et GspD (partie périplasmique) amènent le substrat jusqu'à l'intérieur en l'« asseyant » sur le pseudo pilus. L'ATPase GspE, recrutée au moment de l'interaction de GspD avec le substrat (21), procure l'énergie au système par hydrolyse d'un ATP et fait grandir le pseudo pilus tout en le conservant dans le périplasm. Ceci a pour effet d'amener le substrat au milieu externe.

3.1.1.2.3 Le système de sécrétion de type III (T3SS)

Le système de sécrétion de type III, aussi retrouvé sous le terme d'injectisome, est présent chez de nombreuses espèces bactériennes pathogènes telles que *Salmonella*, *Shigella*, *Yersinia*, *Pseudomonas* et *Escherichia coli* entéropathogènes (EPEC). Le terme générique de protéines effecteur est employé pour désigner les protéines sécrétées par ce système, la plupart du temps directement dans le milieu cytoplasmique de cellule eucaryotes. Le T3SS traverse donc jusqu'à trois membranes grâce à son organisation sous forme de « seringue et aiguille ».

Le corps du système est constitué de 9 protéines, dont 8 sont des flagellines, présentent également dans l'appareil flagellaire de nombreuses bactéries. Une vingtaine d'autres protéines jouent également un rôle important dans le fonctionnement du T3SS. Il existe actuellement 7 familles de T3SSs, définies d'après les caractéristiques de leurs translocons, pointes ou aiguilles, mais l'organisation générale est identique. Elle peut être décrite selon trois éléments principaux : le corps basal, l'aiguille et le translocon. La structure du corps basal forme une « chaussette » autour de l'aiguille (21) (Figure 3.1.6A). Elle est constituée d'éléments cytoplasmiques et de deux anneaux, insérés dans les membranes interne et externe. L'anneau externe, formé de 12 à 15 copies de la protéine SctC, s'étend largement dans la partie périplasmique en enveloppant un tube interne formé des protéines SctI. L'aiguille, constituée de protéines de filaments SctF, est comprise dans le tube interne. Sa longueur est modulable et elle s'étend de la partie périplasmique au milieu extracellulaire. Elle contient un fin canal permettant tel quel la sécrétion de protéines dépliées.

Lors de sa formation, l'aiguille est construite peu à peu par la sécrétion des protéines SctF et sa longueur est régulée par la protéine SctP. La sécrétion des protéines de la pointe permet ensuite la formation de cette dernière. Lorsque le T3SS contacte la membrane d'une cellule hôte par l'intermédiaire de cette pointe, le signal est transmis à la base via les filaments et la protéine SctW est libérée. Un pore est alors formé dans la membrane de la cellule hôte et les effecteurs peuvent être sécrétés et injectés dans le cytoplasme (Figure 3.1.6B).

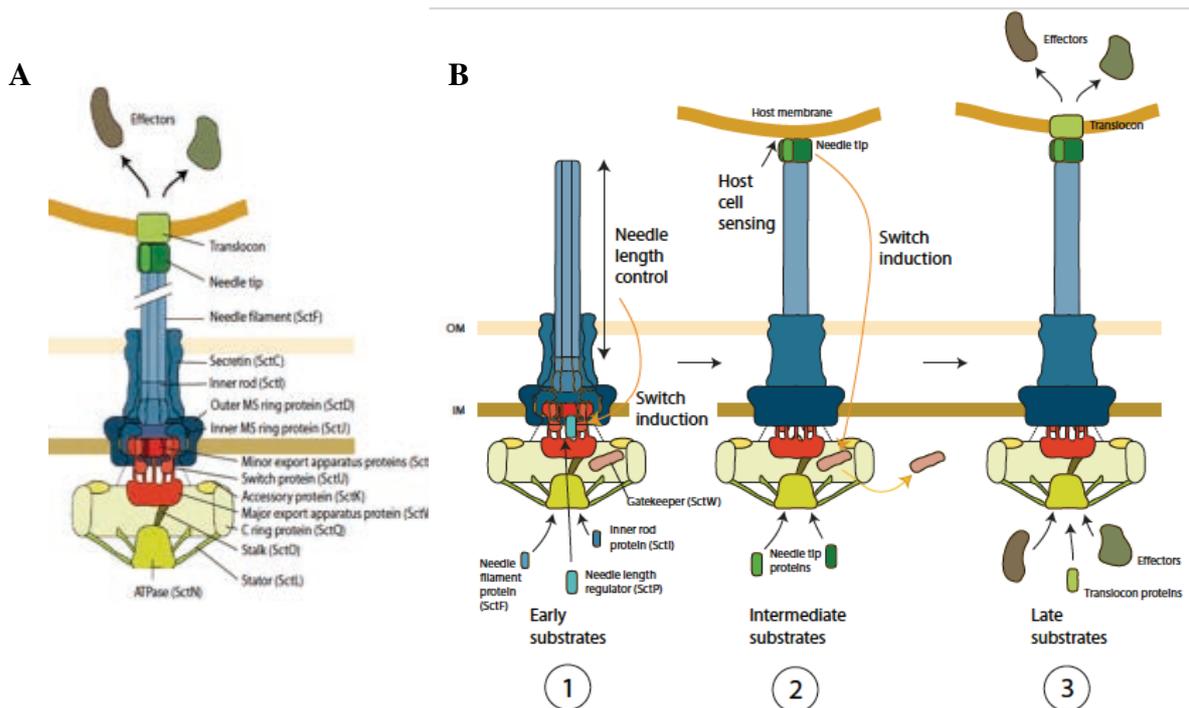


Figure 3.1.6: Description et fonctionnement du T3SS (29). **A :** Description des principaux éléments du système. **B :** Modèle de fonctionnement du système. 1 : La sécrétion précoce de protéines des filaments de l'aiguille (SctF) permet la constitution de l'aiguille jusqu'à une longueur contrôlée par la protéine régulatrice SctP. Des substrats intermédiaires sont alors sécrétés pour constituer la pointe de l'aiguille. 2 : La détection par contact de la pointe avec la membrane de la cellule hôte transmet un signal via les filaments jusqu'à la base du système, entraînant la libération de la protéine de maintien de la porte SctW ; 3 : Le passage est libre pour la sécrétion des substrats tardifs (protéines du translocon et effecteurs) puis leur injection dans le cytoplasme de la cellule hôte.

3.1.1.2.4 Le système de sécrétion de type IV (T4SS)

Ce système de sécrétion est particulier : il s'agit du seul système capable de sécréter des ADNs, seuls ou en complexe avec des protéines (21). Présent chez de nombreuses espèces bactériennes, aussi bien Gram positif que négatif, le T4SS est relié de manière ancestrale aux systèmes de conjugaison d'ADN bactérien. Il traverse, tout comme le T3SS, les membranes interne et externe bactériennes ainsi que celles de cellules hôtes aussi bien de procaryotes que d'eucaryotes. Il existe une grande diversité de T4SSs mais tous partagent une organisation similaire. Le système décrit ici sera basé sur celui de *Agrobacterium tumeficans*, qui possède un T4SS de type VirB/D.

Ce système possède 12 protéines : les protéines VirB (de 1 à 11) et la protéine VirD4 (Figure 3.1.7) (25). Il est constitué d'un core-complexe de membrane externe (VirB7, VirB9 et VirB10) relié à un core-complexe de membrane interne (VirB3, VirB4, VirB6, VirB8 et la partie N-terminale de VirB10) par une tige centrale. VirB4, VirB11 et VirD4 sont les trois ATPases fournissant le système en énergie, situées dans la membrane interne.

Les protéines VirB2 et VirB5 forment le pilus qui s'étend jusqu'au milieu extracellulaire. La protéine VirB1 est une transglycosylase lytique située dans le périplasma, nécessaire à la biogenèse du pilus par sa dégradation de la couche de peptidoglycans (30).

Le mécanisme de sécrétion n'est pas encore tout à fait découvert mais le substrat, qu'il soit ADN ou protéique, semble être pris en charge par la protéine VirD4 qui réalise la sélection des substrats. Cette protéine, qui fonctionne comme une « porte » moléculaire à la base de l'appareil de sécrétion, transfère le substrat à la protéine VirB11 qui l'introduit dans le complexe de membrane interne. Le substrat passe ensuite le périplasma pour arriver au complexe de membrane externe. Le rôle du pilus dans ce processus de sécrétion n'est pas encore défini mais il existe deux hypothèses. Selon la première, le pilus servirait de point d'ancrage de la bactérie à la cellule hôte. La seconde voit plutôt le canal central de ce pilus comme un conduit permettant le passage du substrat. Les difficultés à trancher sur ces hypothèses viennent principalement de l'interconnexion des protéines dans leur fonctionnement (30).

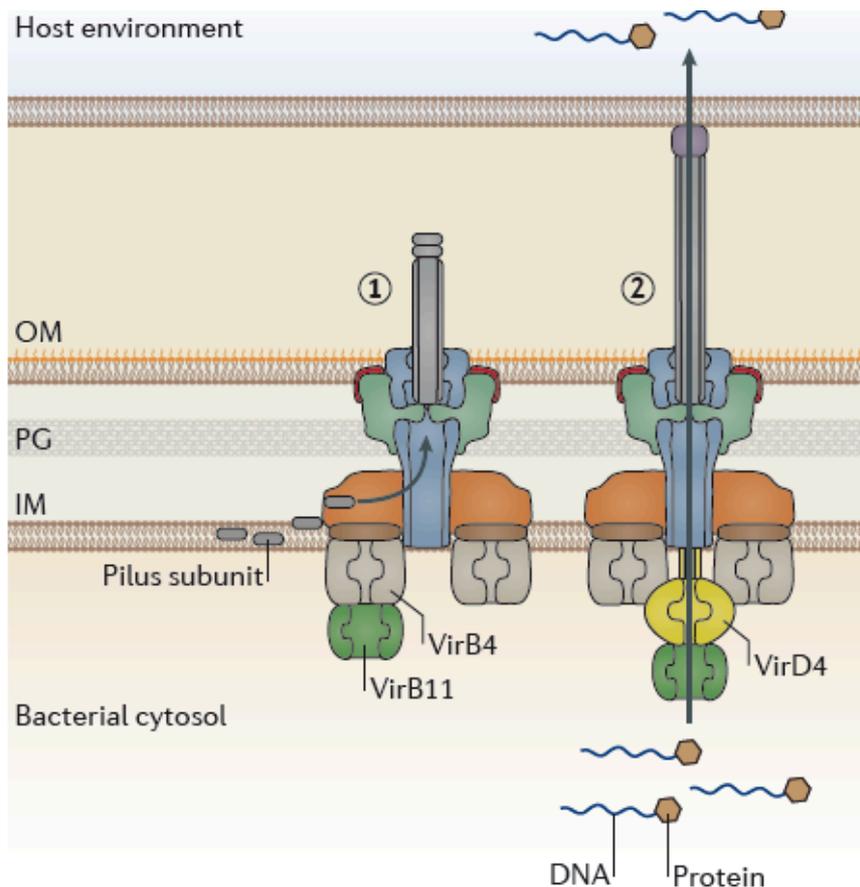


Figure 3.1.7: Représentation schématique du T4SS (25). 1 : La biogenèse du pilus est provoquée par la fixation de la protéine VirB11 sur VirB4. 2 : Mécanisme de sécrétion d'un complexe ADN/protéine. Le complexe est reconnu par VirD4 qui le conduit à VirB11 qui le transfère lui-même dans le système.

3.1.1.2.5 Le système de sécrétion de type VI (T6SS)

Le système de sécrétion de type 6 est un système largement répandu chez les *Proteobacteria* avec une représentation chez environ un quart des espèces. Les différentes classes de T6SSs ont tout de même des spécificités : le T6SSii n'est présent que chez les *Francisella* alors que le T6SSiii l'est uniquement chez les *Bacterioidetes* (31).

Ce système est utilisé dans les processus de compétition bactérienne et de pathogénicité contre des cellules procaryotes et eucaryotes. Des effecteurs toxiques sont sécrétés par son intermédiaire directement dans le milieu cytoplasmique d'autres cellules. Le T6SS a la particularité de se présenter sous une forme et organisation similaire à la queue d'un phage T4. Les protéines la constituant sont d'ailleurs reliés au niveau évolution aux protéines constituant la queue contractile d'un bactériophage (25).

Le T6SS comporte deux complexes : le complexe membranaire, comprenant des protéines de membrane interne similaires à celles du T4SS, et la queue. La queue comporte elle-même trois parties : le complexe de plateforme, la gaine et le tube interne.

Le complexe membranaire comporte les protéines TssJ, TssL et TssM et est communément appelé complexe TssJLM. Celui-ci est inséré dans la membrane interne et ancré à la membrane externe par une ancre lipidique portée par TssJ du côté N-terminal. Le complexe de plateforme est ensuite recruté. Il est formé d'une plateforme, composée des protéines TssE, TssF, TssG et TssK, et d'une pointe, constituée des protéines VgrG et PAAR. Ce complexe de plateforme sert de base à la polymérisation du tube interne et de la gaine constituant la queue. Au moment de l'élongation, des effecteurs peuvent être chargés dans le tube interne ou bien attachés à la pointe VgrG. Lorsque le contact est établi avec une cellule cible, le T6SS est activé et l'éjection du tube transperce la membrane de la cellule cible, permettant la libération des effecteurs dans le milieu cytoplasmique. Il existe deux hypothèses quant au mécanisme d'ouverture du complexe membranaire. Selon la première, celle-ci intervient avant la contraction de la gaine, libérant ainsi le passage en créant un canal de 79 Å de diamètre (Figure 3.1.8). La deuxième propose un changement conformationnel du complexe TssJLM induit par la force du complexe pointe/tube interne, poussé par la contraction de la gaine. Après pénétration de la cellule cible, la pointe et les sous-unités Hcp se dissocient, libérant ainsi les effecteurs. Côté cytoplasmique, l'ATPase ClpV AAA+ est recrutée sur le système pour recycler les sous-unités constituant la gaine (32).

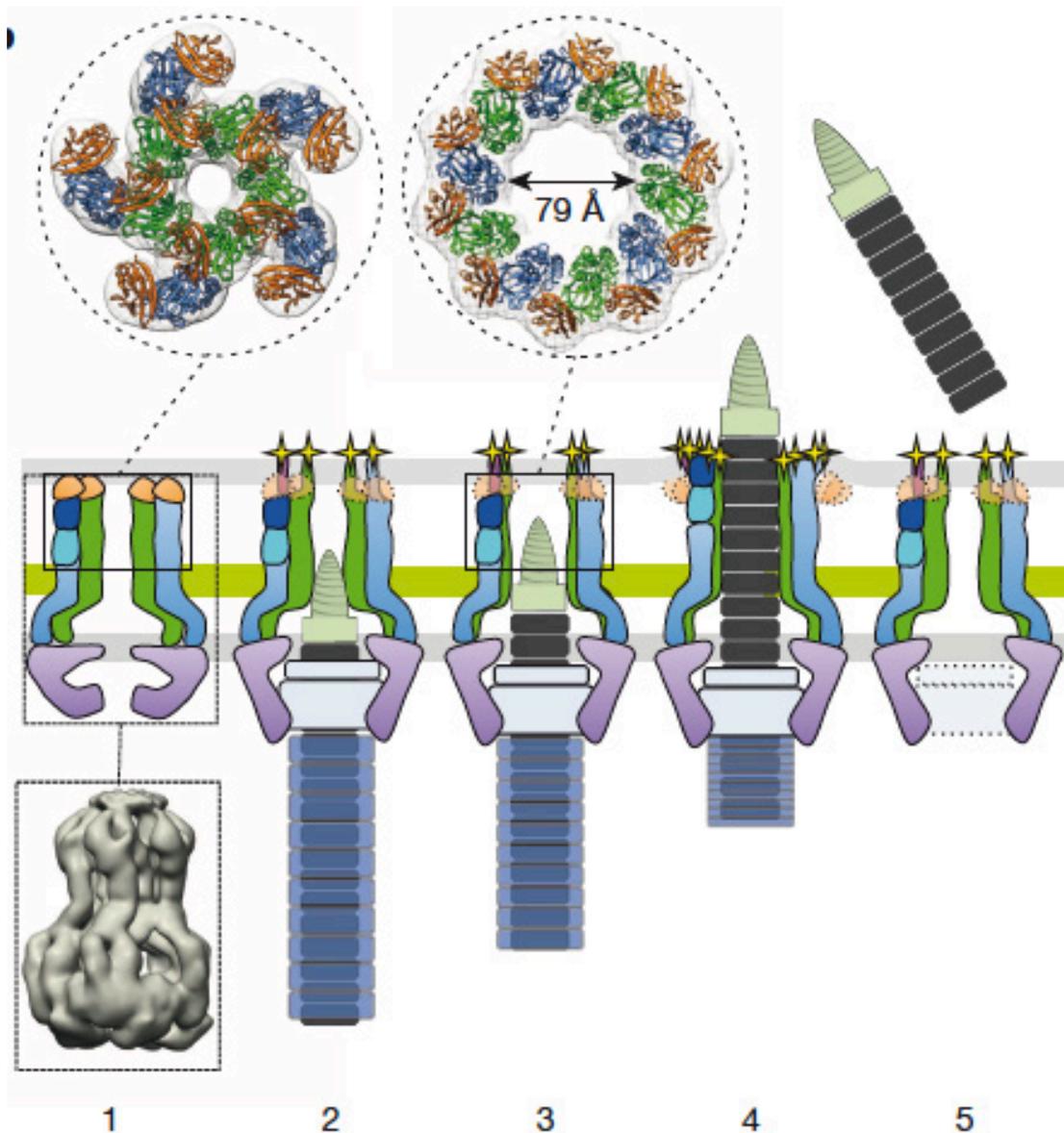


Figure 3.1.8: Mécanisme d'action du T6SS (33). 1 : Le complexe membranaire TssJLM. Il est à ce stade dans une conformation fermée, résolue par microscopie électronique. Il est totalement intégré à la membrane externe puisqu'il y est ancré uniquement par l'ancrage lipidique N-terminale de TssJ. 2 : Suite à l'assemblage de la queue et de la plateforme, la partie C-terminale de TssM s'insère dans la membrane externe (représenté ici par des étoiles jaunes) et est accessible au milieu extracellulaire. Le tube interne est compris dans la partie cytoplasmique, entouré de la gaine. 3 : Le canal du complexe membranaire s'ouvre pour laisser le passage à la queue (4), éjectée par contraction de la gaine. 5 : La plateforme est retirée du complexe membranaire qui est prêt pour un nouveau cycle de sécrétion.

3.1.2 Le système de sécrétion de type 9 (T9SS)

En 2005, une étude chez la bactérie gram-négatif *Porphyromonas gingivalis* révèle que la délétion du gène *porT* entraîne une accumulation de gingipainés et adhésines dans le périplasme (34). Il en a ainsi été conclu que cette protéine était non seulement essentielle à la sécrétion de ces protéines mais également qu'elle devait faire partie d'un système de sécrétion. Devant l'absence, chez *P. gingivalis* et d'autres bactéries, de protéines homologues aux protéines des systèmes de sécrétion déjà connus, l'existence d'un nouveau

système de sécrétion a été proposée. D'abord dénommé système de sécrétion Por (35), ce système a depuis été renommé système de sécrétion de type 9 (T9SS) (36). Bien qu'il ait été initialement étudié chez les bactéries *Flavobacterium johnsoniae* et *Porphyromonas gingivalis* (34) (35) (37) (38), une récente étude a depuis détecté le T9SS dans 62% des espèces de *Bacteroidetes* (31).

Le T9SS est un système de sécrétion Sec-dépendant et se place donc en étape finale du mécanisme de sécrétion général (GSP). Il a été mis en évidence que les substrats qu'il sécrète contiennent un peptide signal porté en N-terminal, les adressant à la machinerie SEC, ainsi qu'un domaine C-terminal (CTD) les dirigeant vers la membrane externe. Après sécrétion, ces protéines sont soit libérées dans le milieu extracellulaire, soit attachées à la surface de la bactérie par un lipopolysaccharide A (A-LPS) (39). Ce dernier cas de figure est une particularité propre à ce système : seulement un à deux substrats fixés à la membrane externe ont été répertoriés jusqu'à présent, issus des systèmes T1SS et T2SS.

Les principales études menées sur le T9SS et les substrats qu'il sécrète ont été faites sur la bactérie *Flavobacterium johnsoniae*, se déplaçant par « gliding » et la bactérie pathogène orale *Porphyromonas gingivalis*.

3.1.2.1 Le système de sécrétion de type 9 chez *Flavobacterium johnsoniae*

Flavobacterium johnsoniae est une bactérie Gram négatif appartenant au phylum des *Bacteroidetes*. Cette longue bactérie de 5 à 10 µm sur 0,4 µm de large (Figure 3.1.9) vit en milieu aérobie et est non-pathogène. Elle peut être trouvée dans le sol où elle est capable de digérer la chitine. Cette bactérie se déplace, tout comme certains membres de ce phylum, sans l'aide de flagelles ou pilus en « glissant » selon le processus de « gliding motility » à une vitesse approximative de 2 µm/s. Elle peut se déplacer, en sécrétant des adhésines principalement, le long d'un axe, inverser son mouvement, pivoter ou même se retourner sur des surfaces diverses telles que le verre, le polystyrène, le Teflon ou l'agar (40). L'analyse de mutants incapables de se mouvoir a permis d'identifier des gènes responsables de cette motilité, dont des gènes *gld*, *spr* et *rem*. De récentes études ont suggéré l'implication de ces gènes dans la sécrétion d'adhésines au travers d'un nouveau système de sécrétion bactérien. Ces mouvements ou glissements seraient donc induits par ceux des adhésines présentes à la surface des cellules. Il en existe deux principales : SprB et RemA.

SprB est une protéine de haut degré de répétition, d'un poids moléculaire de 669 kDa. Cette adhésine forme de longs filaments à la surface de *F. johnsoniae* (Figure 3.1.9) (41). Des bactéries mutantes délétées du gène *sprB* se sont révélées incapables de se déplacer sur un support agar. Leurs mouvements, bien que toujours présents, sur les autres surfaces se sont également trouvés affectés. SprB serait donc nécessaire au gliding sur gelose d'agar et probablement aussi impliquée dans celui d'autres surfaces. Des études par marquage fluorescent de SprB ont permis de suivre les déplacements de ces protéines à la surface de la bactérie et ont montré que SprB suit une trajectoire sous forme de boucle hélicoïdale (41). L'énergie nécessaire à ce mouvement est fournie par une force proton motrice (FPM). La vitesse de ce déplacement est similaire à celle du glissement de la bactérie, alimentant l'hypothèse d'un modèle selon lequel des moteurs FPM-dépendants propulsent SprB le long de sa trajectoire hélicoïdale (40)

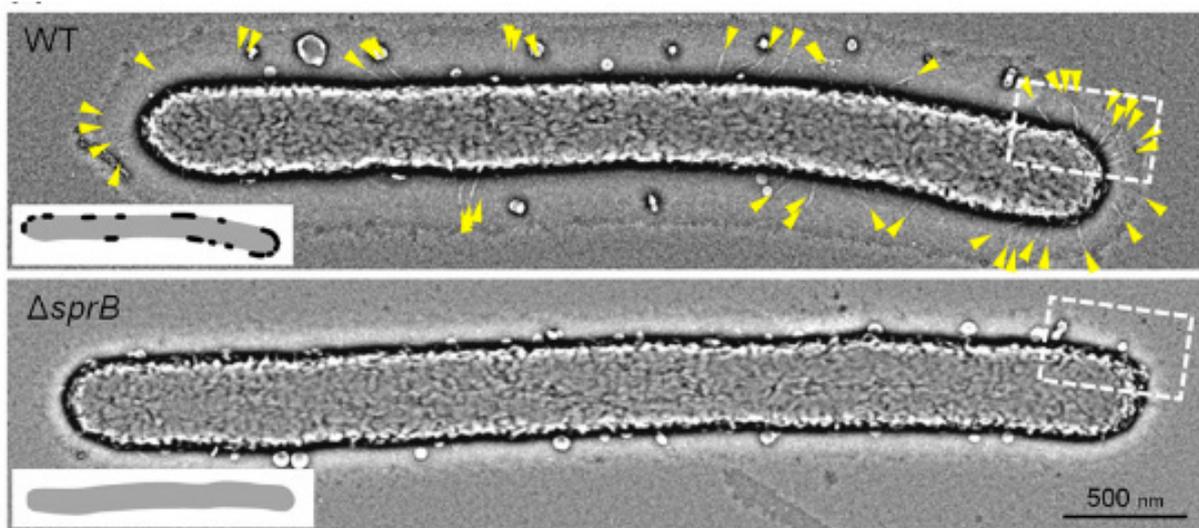


Figure 3.1.9 : Observation par coloration négative des filaments de SprB à la surface de *F. johnsoniae* (41). Les filaments présents à la surface de la souche wild-type (wt) UW101 sont indiqués par des flèches jaunes. (Image du haut). La même expérience réalisée sur une souche mutée n'exprimant pas le gène *sprB* (CJ1922 Δ *sprB*) ne montre aucun filament.

Les mouvements de glissement résiduels des bactéries n'exprimant pas SprB suggèrent l'existence d'autres adhésines de surface, responsables de cette motilité. La protéine RemA a été identifiée comme l'une d'entre elles. Des bactéries doubles mutantes ne produisant ni RemA ni SprB ont une perte totale du glissement tandis que des bactéries simples mutantes pour l'un de ces gènes conservent une motilité partielle. RemA est une lectine se liant aux polysaccharides de surface produits et sécrétés par *F. johnsoniae*. Un modèle du glissement de *F. johnsoniae* a été proposé en 2015 (40) (Figure 3.1.10). Les

moteurs, formés de protéines Gld, propulsent les adhésines le long de leur trajectoire hélicoïdale. Les protéines RemA étant attachées aux polysaccharides, leurs mouvements entraînent ceux de la bactérie qui se met à glisser sur la surface.

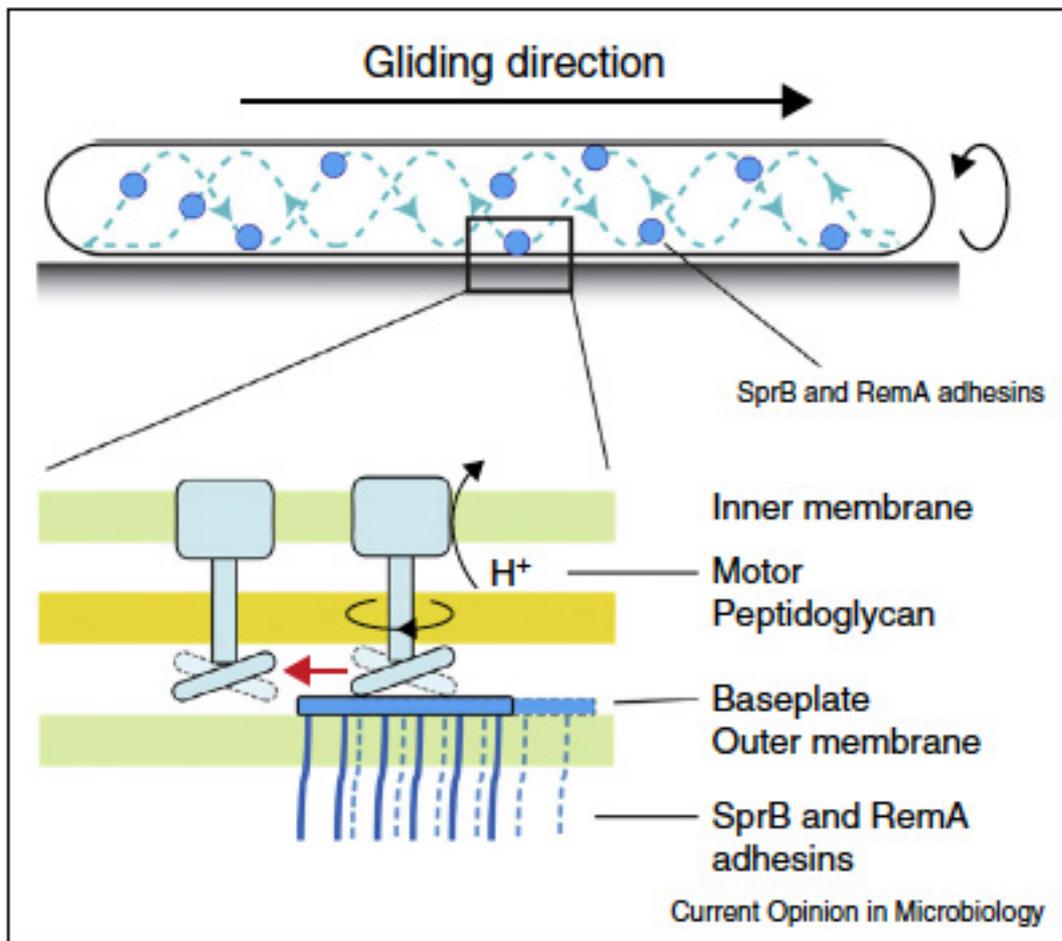


Figure 3.1.10: Modèle du gliding de *F.johnsoniae* (40). Les moteurs FPM-dépendants, formés par les protéines Gld et ancrés dans la membrane interne, procurent l'énergie aux adhésines SprB et RemA pour poursuivre leur trajectoire hélicoïdale le long de la surface de la bactérie. Les adhésines fixées sur les polysaccharides portés à la surface entraînent alors la bactérie dans leurs mouvements, résultant au mouvement de gliding.

Pour que ce système soit mis en place, les adhésines ainsi que les polysaccharides doivent être produits puis sécrétés à la membrane externe ou à la surface de *F. johnsoniae*, par leur passage au travers d'un nouveau système de sécrétion.

Les protéines GldK, GldL, GldM, GldN, SprA et SprE avaient été initialement considérées comme des protéines permettant la motilité. L'existence d'orthologues à ces protéines chez d'autres espèces du phylum des *Bacteroidetes*, incapables d'un tel mécanisme, a permis de reconsidérer leur fonction. Ces protéines, ainsi que la protéine SprT, font partie du système de sécrétion de type 9, le T9SS. Il est maintenant connu que ce système est nécessaire à la sécrétion de SprB, RemA, de la chitinase soluble ChiA (38) ainsi que d'une douzaine d'autres protéines chez *F. johnsoniae*. Il semblerait donc que le T9SS et

la motilité par glissement soient deux mécanismes étroitement liés. Outre les protéines citées ci-dessus, la protéine PorV, intégrée à la membrane externe, est nécessaire à la sécrétion de RemA, de ChiA et d'autres protéines mais pas à celle de SprB. En revanche, la protéine SprF, également présente dans la membrane externe, n'est elle nécessaire qu'à la sécrétion de SprB.

Du fait de la présence en N-terminal d'un peptide signal reconnu par la machinerie SEC et en C-terminal d'un CTD, les protéines à sécréter doivent dans un premier temps être prises en charge par la machinerie SEC sous une forme dépliée, transférées dans le périplasma puis reconnues par le T9SS qui les sécrète au travers la membrane externe. Un modèle du couplage T9SS/motilité par glissement a été proposé en 2015 par McBride & Nakane (Figure 3.1.11).

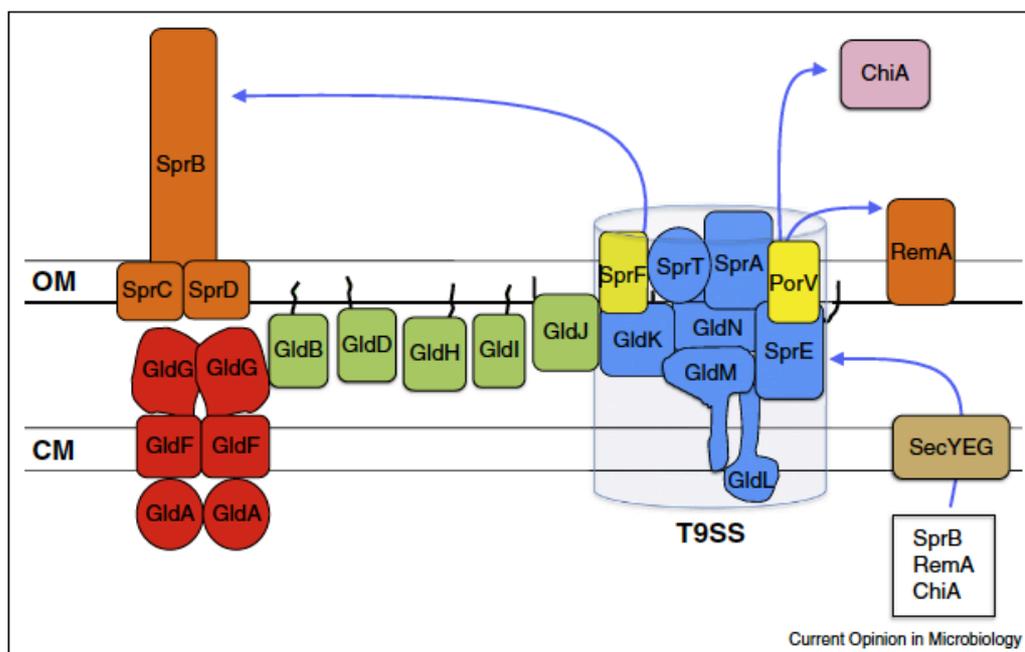


Figure 3.1.11: Model schématique du T9SS chez *Flavobacterium johnsoniae* (40).

Le mécanisme du moteur dirigeant le mécanisme de glissement n'est à ce jour pas encore résolu mais l'implication des protéines Gld dans celui-ci est certain. Seules les protéines GldF, GldG, GldL et GldM sont candidates, dû à leur insertion dans la membrane cytoplasmique. Les protéines GldF et GldG semblent plutôt impliquées dans un processus utilisant de l'ATP, les protéines GldL et GldM restent l'hypothèse la plus probable sur les protéines impliquées dans un processus de génération de force proton motrice.

3.1.2.2 Le système de sécrétion de type 9 chez *Porphyromonas gingivalis*

Le premier modèle d'études historique du T9SS est la bactérie *Porphyromonas gingivalis*. Vivant en milieu anaérobie, cette bactérie a la particularité de présenter une pigmentation noire, provenant de la dégradation de l'hémine, nécessaire à sa croissance et à sa prolifération (Figure 3.1.12). Ce phénotype est plus particulièrement dû à l'activité protéolytique et de liaison de l'hémagglutinine/hème de gingipaïnes (ou porphyrines). Les protéines constituant ce système ont été initialement nommées Por, non pas en référence à *Porphyromonas gingivalis* mais plutôt en référence à ces porphyrines (exemple : la protéine PorT). Les protéines Por sont essentielles à la production et/ou sécrétion de ces porphyrines (39).

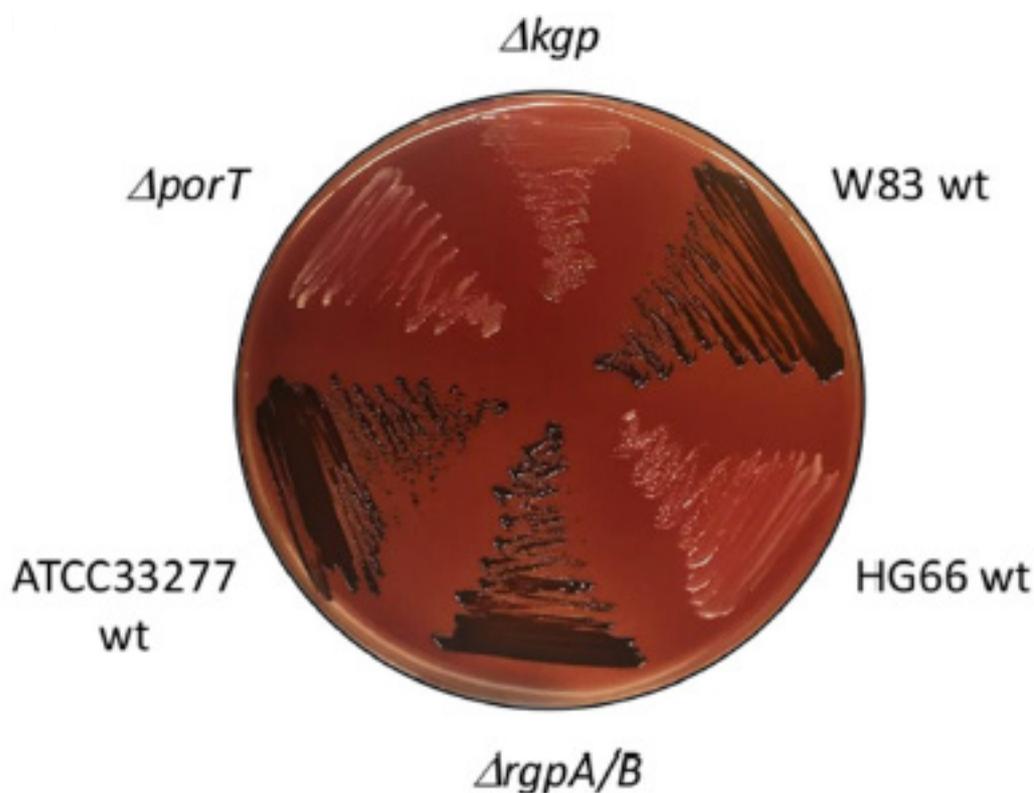


Figure 3.1.12: Pigmentation de différentes souches de *P.gingivalis* W83 (42). Croissance bactérienne anaérobie sur boîte d'agar contenant du sang. Les souches non-mutées (wt) présentent une pigmentation noire due à l'hème accumulée à la surface des cellules. La souche HG66 n'ayant pas de A-LPS, les gingipaïnes sécrétés par le T9SS ne sont pas ancrées à la membrane externe et ne peuvent donc pas retenir l'hème. Lorsque les bactéries n'expriment pas les protéines Kgp et PorT, le phénotype non-pigmenté est présent. En revanche, l'absence d'expression des gingipaïnes RgpA et RgpB n'influence pas le phénotype sauvage.

P.gingivalis est une bactérie pathogène humaine que l'on retrouve, avec d'autres pathogènes, dans de nombreuses maladies telles que l'athérosclérose, la pneumonie d'aspiration, la polyarthrite rhumatoïde ou encore certains cancers (42). Elle joue également un rôle clé dans la parodontite en formant, avec *Tannerella forsythia* et *Treponema denticola*, une triade de bactéries pathogènes appelée le complexe rouge (43).

Ce complexe fait partie du microbiote buccal, constitué de plusieurs centaines d'espèces de bactéries, virus et champignons. Chez un individu sain, l'existence et le développement de ces micro-organismes sont maintenus par des conditions physiologiques particulières de température, pH, potentiel redox, conditions atmosphériques, salinité et par la salive. Cet équilibre dépend principalement de l'hygiène bucco-dentaire de l'individu et, lorsqu'il est rompu ou perturbé suite à une lésion par exemple, certaines espèces telles que *P.gingivalis* peuvent proliférer (44).

Les bactéries présentes dans la cavité buccale peuvent s'organiser en biofilms. Ceux-ci peuvent se former au niveau de la crevasse gingivale (interface de rencontre de la gencive avec la dent), de la crevasse sub-gingivale, de la langue, des muqueuses et des prothèses dentaires.

Le parodonte est l'ensemble des tissus de soutien des dents. Il comprend les gencives, l'os alvéolaire, le ligament alvéolaire-dentaire et le ciment. L'os alvéolaire est le tissu de soutien des dents. Il entoure la dent et est en continuité avec l'os maxillaire et mandibulaire. Il évolue avec la dent : il peut tout aussi bien augmenter (par exemple pendant la poussée dentaire) comme se résorber lorsqu'il y a perte de celle-ci (45).

Parmi les maladies touchant le parodonte se trouve la parodontite. Il s'agit d'une inflammation chronique due à l'accumulation d'une plaque bactérienne entre la gencive et la dent (46) (Figure 3.1.13). Le pourcentage en bactéries Gram-négative de ce biofilm, faible pour un tissu sain, est ici considérablement augmenté (47). Les bactéries constituant le complexe rouge ont été identifiées comme étant impliquées dans cette maladie. Des enzymes protéolytiques sécrétées par ce complexe ont été découvertes comme étant impliquées dans l'échappement à la réponse immunitaire.

Comme dit plus haut, les enzymes protéolytiques sécrétées par *P.gingivalis* sont appelées gingipaïnes. Celles-ci sont les facteurs de virulence responsables de la colonisation, nutrition, neutralisation des défenses de l'hôte et altération de la réponse inflammatoire de cet hôte, ce qui en font les principales responsables des symptômes des parodontites. Parmi ces gingipaïnes se trouvent les protéines Kgp, RgpA et RgpB. Toutes trois sont des protéases clivant les liaisons peptidiques portées en C-terminal de lysines (Kgp) ou arginine (RgpA/RgpB). Elles sont portées à la surface bactérienne par une ancre A-LPS ajoutée après sécrétion et doivent donc auparavant avoir été sécrétées par le T9SS (42). Ce système de sécrétion a également été retrouvé chez *Tannerella forsythia*.

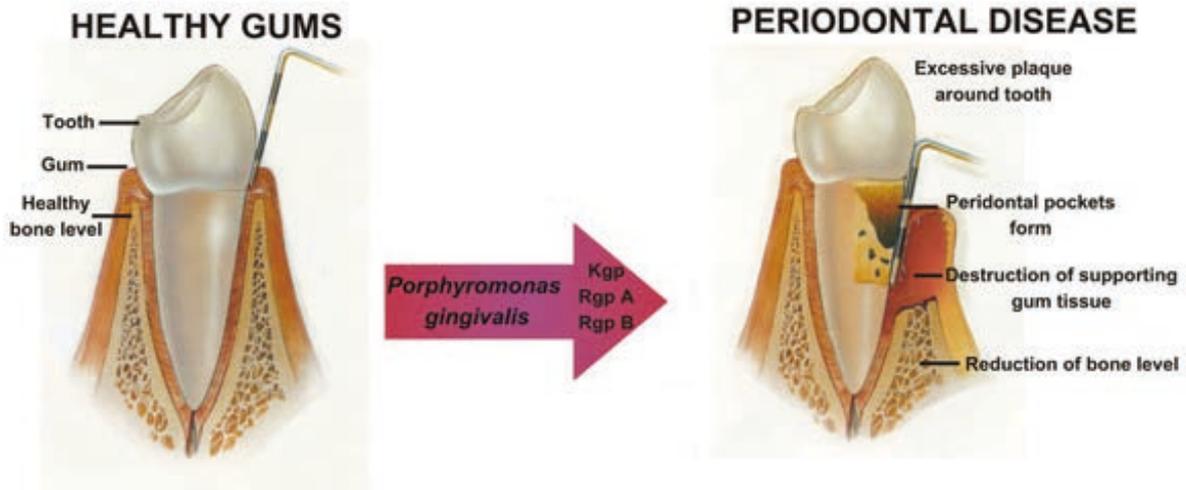


Figure 3.1.13: Effets de la parodontite sur le parodonte, provoquée notamment par *P.gingivalis* (48) (<http://www.stcatherinesdentalpractice.co.uk/dental-treatments-in-grantham/periodontal-treatment-in-grantham.html>). L'activité des gingipaïnes Kgp, RgpA et RgpB permet à la bactérie d'échapper aux défenses immunitaires de l'hôte, ayant pour effet l'accumulation de plaque dentaire dans la crevasse gingivale. Ceci conduit à une destruction progressive de l'os alvéolaire soutenant les dents, provoquant à terme un déchaussement des dents.

Le T9SS de *P.gingivalis* comprend 18 protéines impliquées dans le processus de sécrétion des gingipaïnes mais seulement 11 constitueraient la machinerie en elle-même. Tout comme chez la bactérie *F.johnsoniae*, ce système fonctionne en coordination avec le translocon SecYEG. Bien que les deux types de T9SS présentent des différences, leur « cœurs » sont très similaires, constitués de gènes orthologues : *gldK/porK*, *gldL/porL*, *gldM/porM*, *gldN/porN*, *sprA/sov*, *sprE/porW* et *sprT/porT*.

Dans l'interface cytoplasme-membrane interne se trouvent les protéines membranaires PorL et PorM, possédant toutes deux une partie cytoplasmique (environ 236 résidus pour PorL, résidus 1 à 8 pour PorM), une partie membranaire (résidus 9 à 41 pour PorM) et une partie périplasmique (résidus 42 à 516 pour PorM). L'interaction de ces deux protéines a été prouvée *in vitro* chez *E. coli* (49) et *in vivo* (35). Tout comme pour leurs protéines orthologues chez *F.johnsoniae*, une implication de PorL et PorM dans les processus énergétique du système a été suggérée (49). Ce complexe produirait un changement conformationnel à l'assemblage du T9SS et participerait au processus de sécrétion des substrats grâce à la force proton motrice et aux protéines associées au T9SS PorY et PorX. La partie cytoplasmique C-terminale de PorL interagit *in vitro* avec la protéine régulatrice de réponse PorX (50). PorX est un régulateur de réponse qui, avec son partenaire senseur de kinase PorY, intervient dans ce processus de régulation des gènes du T9SS.

La protéine PorM se trouve également dans le périplasme. Son étude *in vitro* en cellules *E. coli* montre une dimérisation ainsi qu'une interaction avec les protéines PorK et PorN (49). Ces deux dernières protéines interagissent également entre elles (49). PorN a un rôle crucial dans la stabilisation des interactions PorL/PorM et PorN/PorK (51). Le périplasme contient également les protéines PorW et PG1058, toutes deux prédites comme étant des lipoprotéines attachées à la membrane externe.

Les protéines Sov, PorQ, PorP, PorT, PorV, PG0189 et PG0534 sont prédites comme étant des protéines de la membrane externe en tonneau bêta. Très peu de choses sont connues sur ces protéines, excepté pour PorP et PG0189 qui interagissent avec le complexe PorKLMN. PG0189 interagit avec PorK et PorN avec sa boucle périplasmique tandis que des expériences *in vitro* montrent une interaction de PorP avec PorK et PorM. Une étude a prouvé l'interaction de PorV avec PorU et propose qu'elle lui serve d'ancrage à la membrane externe. Les protéines PorU et PorZ sont quant à elles associées à cette membrane par sa surface externe. Alors que PorU est une cystéine peptidase impliquée dans le clivage du CTD des substrats, la structure cristallographique de PorZ suggère également une implication dans le processus de modification post-translocation (42).

Le mécanisme de sécrétion des substrats impliquant le T9SS chez *P.gingivalis* est un processus en deux étapes. Le passage du cytoplasme vers le périplasme est réalisé au travers du translocon SecYEG de la machinerie SEC. Le substrat, déplié, y est adressé par reconnaissance de son peptide signal porté en N-terminal, clivé après sa translocation. Son mécanisme de repliement ainsi que son transfert au système restent inconnus. La protéine est reconnue du T9SS par son domaine CTD et traverse la membrane externe en passant par le translocon de membrane externe (Figure 3.1.14).

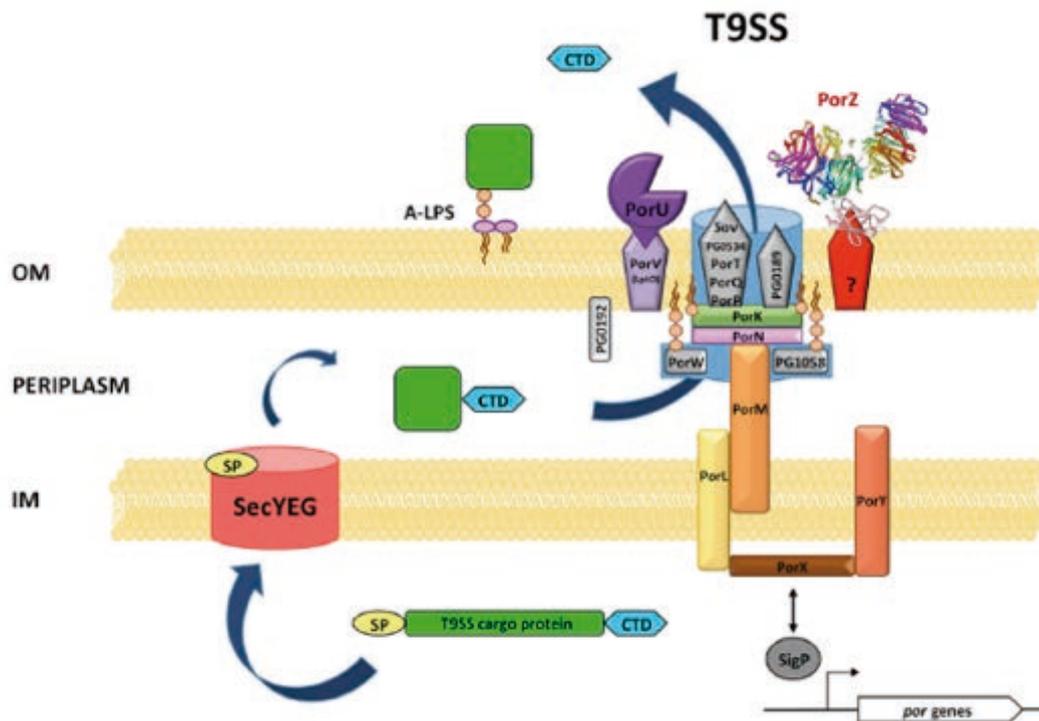


Figure 3.1.14: Modèle schématique hypothétique de la structure et du fonctionnement du T9SS chez *P.gingivalis* (42). Les substrats devant être sécrétés par le système portent en N-terminal un peptide signal (SP) et en C-terminal un domaine C-terminal (CTD). Le substrat est pris en charge par la machinerie de sécrétion Sec qui le fait passer du cytoplasme au périplasm par le translocon SecYEG. Une fois le SP clivé, le substrat est reconnu par le T9SS qui le fait passer dans le milieu externe. Le CTD est clivé par PorU et le substrat est, dans la majorité des cas, attaché à la membrane externe par la lipopolysaccharide A-LPS qui lui a été ajoutée.

Le cargo CTD est alors clivé par PorU puis le substrat est soit libéré dans le milieu extracellulaire, soit attaché à la membrane externe par une A-LPS, après modification post-translocation. Les gènes du T9SS sont régulés par le système à deux composants PorXY(50). L'architecture de la protéine PorY est similaire à une histidine kinase. Après autophosphorylation sur son Histidine 193, le groupement phosphate est transféré à la protéine PorX sur son résidu Asparagine 58. PorX phosphorylé va alors stabiliser le facteur sigma de fonction extracytoplasmique SigP qui va directement se fixer aux régions promotrices des gènes du T9SS. PorX phosphorylé peut également aller se fixer à PorL (50). La partie N-terminale de PorX étant similaire aux régulateurs de réponse de la famille CheY, une hypothèse serait l'implication de PorX dans la modulation de l'architecture du T9SS. Les protéines CheY, après phosphorylation, se fixent à l'anneau C de flagelles, induisant un changement de direction du mouvement de ce flagelle (52). PorX pourrait, par un processus similaire, agir sur la rotation de l'appareil. La Figure 3.1.14 regroupe toutes ces données sous la forme d'un modèle schématique.

3.1.2.3 Le cœur du T9SS chez *P.gingivalis* : le complexe PorKLMN

Bien que la plupart des gènes du T9SS soient répartis dans le génome de *P.gingivalis*, un opéron de 5 de ces gènes est conservé au sein des *Bacteroidetes* : *porP-porK-porL-porM-porN*. L'analyse sur natif PAGE de lysats cellulaires de *P.gingivalis* a révélé la présence d'un complexe d'environ 1,4 MDa, non dissociable. La présence dans ce complexe des protéines PorK, PorL, PorM et PorN a été prouvée par immunoblotting (35). Bien que la protéine PorP interagisse avec au moins deux autres protéines issues de l'opéron (49), elle ne semble pas faire partie du complexe ni stabiliser ce complexe (35). Des expériences de mutagénèse portant sur les protéines du complexe PorKLMN ont montré que ce complexe, ainsi que chacune des protéines le composant, est essentiel à la présence de la couche dense aux électrons observable à la surface de *P.gingivalis* (51). Cette couche étant constituée des gingipaines de surface de la bactérie, son absence est donc significative d'un dysfonctionnement du T9SS. Le complexe PorKLMN est, par conséquent, essentiel au fonctionnement du T9SS.

Il a récemment été montré que ce complexe est constitué de deux sous-complexes : PorKN et PorLM. La structure du complexe PorKN a été étudiée *in-vivo* et *in-vitro* par cryo-microscopie électronique (51) : il s'agit, *in-vivo*, d'un assemblage de deux anneaux dans un ratio PorK:PorN de 1:1, de diamètre interne de 35 nm et externe de 50 nm, attaché à la membrane externe côté périplasma. *In-vitro*, un second assemblage de deux anneaux vient interagir avec le premier, formant un sandwich (Figure 3.1.15). Même s'il n'existe actuellement aucune évidence le prouvant, il est vraisemblable que l'anneau de plus grand diamètre soit constitué de protéines PorK et le second anneau de PorN. En effet, le plus grand anneau est au contact de la membrane *in-vivo*, ce qui serait rendu possible par l'attachement lipidique de PorK à la surface périplasmique de la membrane externe. Un autre argument pour étayer cette hypothèse est la taille de PorK, plus importante que celle de PorN, qui permettrait la formation d'un anneau de diamètre plus important pour un même nombre de sous-unités. Cette structure porte une symétrie soit de 32 unités au total, soit de 36. La structure du complexe *in-vitro* aurait donc un poids moléculaire de 5,9 MDa ou 6,6 MDa, concordant avec la taille observée lors de la migration sur SDS-PAGE, et celle *in-vivo* un poids moléculaire autour de 3 MDa.

Peu de données structurales sont disponibles actuellement pour comprendre le fonctionnement et l'assemblage du complexe PorLM. Comme dit précédemment, les deux protéines sont ancrées dans la membrane interne par leurs hélices transmembranaires. Des

études d'accessibilité des cystéines sur des mutants de PorL et PorM ont permis de déterminer la topologie de ces protéines dans la membrane(49). PorL comporte deux hélices transmembranaires et ses extrémités N et C-terminales se trouvent toutes les deux dans le cytoplasme. La longue région cytoplasmique C-terminale de PorL interagit avec la protéine PorX phosphorylée. La même étude a révélé l'état trimérique de la protéine.

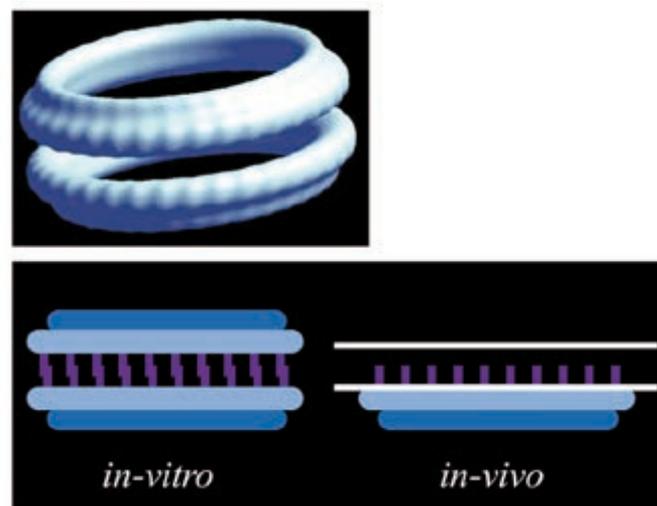


Figure 3.1.15: Reconstruction 3D et modèle schématique du complexe PorKN (51). Le complexe PorKN a été observé par cyro-microscopie électronique *in-vitro* et *in-vivo*. Alors qu'*in-vivo* un seul assemblage d'anneaux PorK/PorN est visible (1x anneau PorK/1x anneau PorN), un second vient interagir tête bêche avec le premier dans la préparation *in-vitro* (2x anneau PorK/2x anneau PorN). Le contact est produit par l'interaction des lipides de PorK (en violet). La reconstruction 3D *in-vitro* est représentée sur l'image du haut tandis que des modèles schématiques le sont sur l'image de bas. Dans l'assemblage, les deux anneaux n'ont pas le même diamètre. Celui de plus haut diamètre figure au contact de la surface périplasmique de la membrane externe sur le modèle *in-vivo*, avec les lipides ancrés dans la membrane.

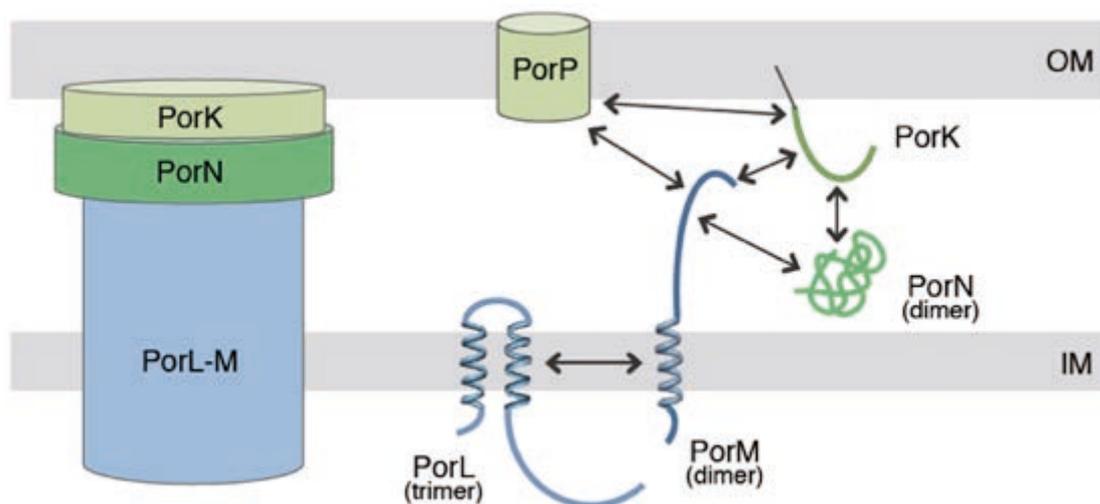


Figure 3.1.16: Modèle schématique du complexe PorKLMN et des interactions entre ses protéines (49). Le complexe PorL-M est ancré dans la membrane interne et s'étend dans le périplasme jusqu'au complexe PorKN, formé de deux anneaux portés à la surface de la membrane externe. La protéine PorL, trimérique, interagit avec le dimère PorM via leurs hélices transmembranaires. La partie C-terminale de PorM interagit avec les protéines PorN (dimérique), PorK et PorP. Le dimère de PorN interagit également avec PorK et cette dernière lie la protéine PorP, ancrée dans la membrane externe.

La partie périplasmique de la protéine PorM est, elle, dimérique en sortie de colonne d'exclusion de taille, laissant supposer que PorM soit dimérique (49). Elle traverse la membrane interne, avec son extrémité N-terminale dans le cytoplasme, une hélice transmembranaire (résidus 9 à 35) et s'étend jusque dans le périplasme. Des études structurales de sa partie périplasmique (pPorM) ont été menées par l'équipe (53). Malgré l'obtention d'un jeu de données structurales par cristallographie aux rayons X, les difficultés de résolution du problème de phases puis de reconstruction n'ont pas permis la résolution complète de la structure. Des expériences menées chez *F. johnsoniae* ont démontré l'existence d'un moteur rotatif du T9SS (54). Ce moteur serait un transducteur d'énergie, générant une force proton motrice utilisée pour l'assemblage et la sécrétion des protéines. Il a été présumé associé aux protéines GldL et GldM, orthologues du complexe PorLM chez *P.gingivalis*. Cette hypothèse a été posée sur la base de leur localisation transmembranaire et de la présence de résidus glutamate chargés négativement dans les hélices transmembranaires. Vincent *et al.* ont proposé l'extension de cette hypothèse au complexe PorLM.

Compte tenu des différents états oligomériques de ses quatre protéines, le complexe PorKLMN minimal serait un complexe $\text{PorK}_2\text{L}_3\text{M}_2\text{N}_2$ ayant un poids moléculaire d'environ 400 KDa. Pour expliquer cette différence avec le poids moléculaire apparent du complexe de 1,4 MDa, deux hypothèses ont été émises : la première propose que d'autres protéines interviennent dans ce complexe tandis que la seconde propose une multimérisation du complexe minimal (49). Plusieurs éléments viennent alimenter ces deux hypothèses puisque de nouveaux partenaires ont été découverts par spectrométrie de masse après étude du complexe PorKN par cryo-microscopie électronique(51), le T6SS comporte lui même une multimérisation du complexe minimal TssJLM. Par ailleurs, il faut garder à l'esprit que l'estimation du poids moléculaire de PorKLMN à 1,4 MDa a été faite sur la base d'un immunotransfert d'un gel dont le marqueur de taille avait pour protéine de plus haut poids moléculaire 1 MDa. Cette estimation est donc imprécise et potentiellement erronée.

La protéine PorM interagit avec les protéines PorL, PorP, PorN et PorK (Figure 3.1.16) (35) (49). Elle tient donc une place centrale dans le complexe PorKLMN. La poursuite de son étude structurale, couplée à une étude fonctionnelle par nos collaborateurs (équipe E. Cascales, LISM), est donc essentielle à l'élucidation de l'assemblage du complexe PorKLMN et du fonctionnement du T9SS.

3.1.2.4 Résultats préliminaires sur la protéine PorM

À mon arrivée dans l'équipe, les études structurales sur la partie périplasmique pPorM (résidus 44 à 516) avaient donné lieu à une publication (53). La rapide dégradation de pPorM y a été expliquée ainsi que la nécessité d'étudier la protéine sous une forme digérée à la trypsine (pPorM-T). Cette stratégie a permis de réduire au mieux les parties flexibles et d'obtenir des cristaux après quelques jours au lieu de 6 mois suivant une dégradation naturelle. La structure cristallographique a alors pu être résolue à une résolution de 2,85 Å en calculant des phases à l'aide de Selenométhionines. Cependant, la flexibilité de certaines boucles et la résolution maximale n'ont pas permis une reconstruction totale.

Afin de compléter cette structure et de poursuivre la détermination de la structure totale de pPorM, une immunisation de Lama contre pPorM a été effectuée. En parallèle, le fragment majoritaire pPorM-T a été analysé par Spectrométrie de Masse et séquençage d'Edman. Les bornes déterminées correspondent à la partie C-terminale de pPorM, comprenant les résidus 225 à 516. Deux sous-constructions de pPorM ont été imaginées : pPorM-Nt pour la partie N-terminale (résidus 44 à 217 de PorM) et pPorM-Ct pour la partie C-terminale (résidus 225 à 516 de PorM).

3.2 Génération et structures des nanobody anti-PorM

3.2.1 Introduction

Comme dit précédemment (voir 1.1.2), les nanobody sont des fragments d'anticorps issus d'anticorps de camélidés, découverts au début des années 1990. Autre nom des domaines VHH, portant à eux seuls la partie reconnaissant et fixant l'antigène, ils présentent l'avantage d'être stables dans des conditions diverses et d'être de petite taille (environ 15 kDa), ce qui leur confère la propriété de chaperonnes de cristallisation de protéines.

Des nanobody peuvent être générés dans ce but (Figure 3.2.1). Pour cela, la protéine servant d'antigène doit être purifiée dans un tampon adapté à l'animal utilisé pour l'immunisation. En effet, la préparation doit être injectée à un camélide, généralement chameau ou lama, et un tampon non adapté peut provoquer un choc osmotique pouvant entraîner la mort de l'animal. Après 45 jours, une prise de sang est faite et les lymphocytes sont isolés. Les ARNm sont alors extraits et les ADN complémentaires (ADNc) sont synthétisés. Les séquences de ces ADNc correspondant aux domaines VHH sont amplifiées par PCR et une librairie de nanobody est ainsi constituée.

Afin de sélectionner les nanobody reconnaissant l'antigène et s'y fixant, un cycle de 3 à 4 tests ELISA est réalisé en fixant l'antigène sur un support. Après plusieurs lavages, seuls les anticorps se fixant à cette protéine sont gardés, puis élués pour être de nouveau amplifiés et recommencer le test. Les nanobody positifs sont alors séquencés puis exprimés dans la partie périplasmique de cellules *E. coli*.

De la même manière, une librairie de nanobody « naïve » peut être constituée. Celle-ci est issue d'un animal auquel l'antigène n'a pas été injecté et qui est donc dit « naïf ». La librairie est constituée des nanobody déjà présents dans le sérum de l'animal et peut être testée par ELISA contre l'antigène afin de vérifier s'il existe déjà un anticorps le reconnaissant.

La totalité du protocole prend entre 2 à 3 mois, de l'immunisation à l'obtention de nanobody positifs à l'antigène.

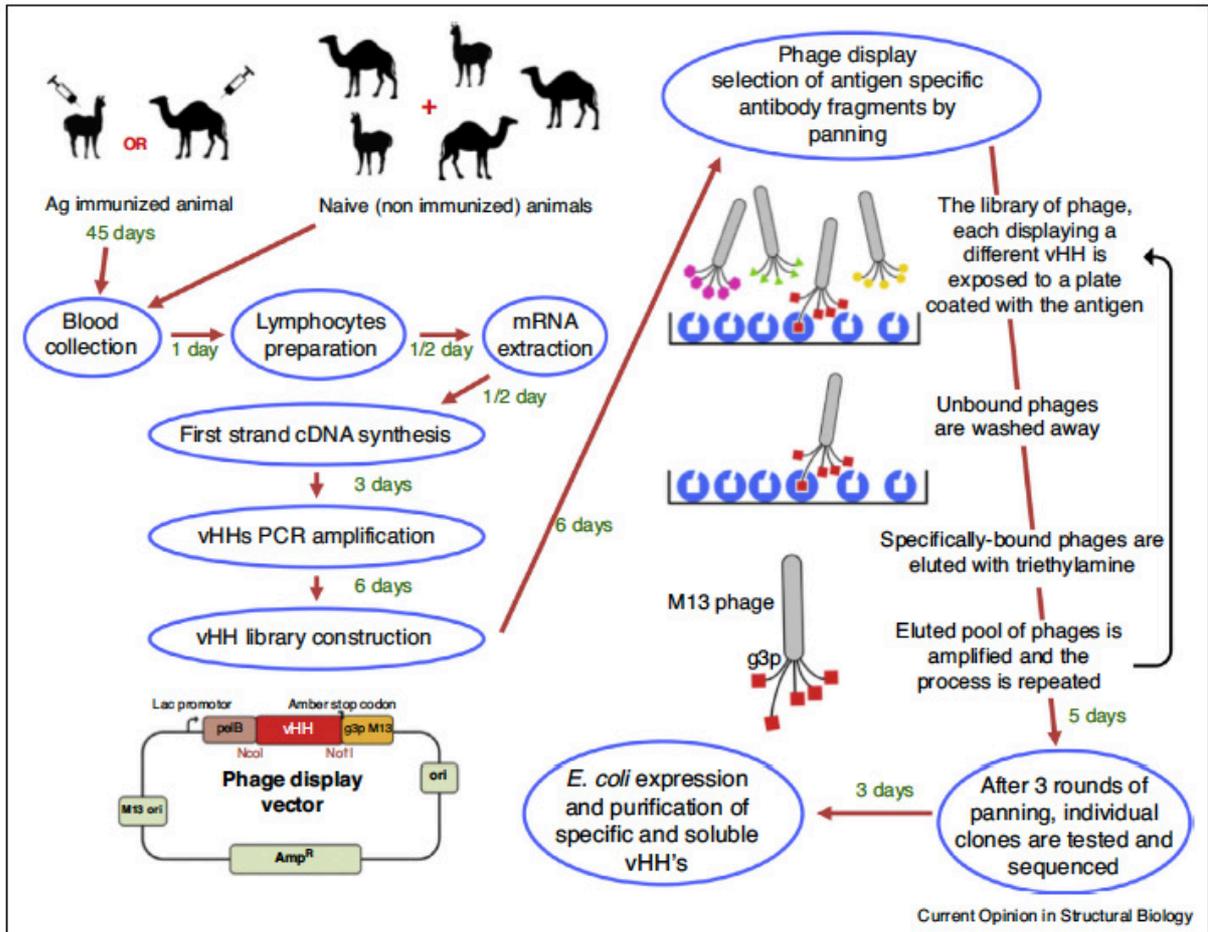


Figure 3.2.1 : Protocole d'immunisation de camélidés et sélection des nanobody contre un antigène (55). Du sang de camélidés est prélevé, préalablement immunisés ou non avec l'antigène. Les lymphocytes sont isolés, les ARNm extraits, puis les ADNc synthétisés. Là, la séquence des domaines VHH est amplifiée pour pouvoir constituer une librairie de nanobody issue d'immunisation ou bien naïve. Après introduction dans des phages, chaque nanobody est testé par test ELISA contre l'antigène et les nanobody positifs sont sélectionnés pour l'expression périplasmique en *E. coli*. Le processus prend 2 à 3 mois.

3.2.2 Article 1

Afin de poursuivre les études structurales de la protéine PorM, des nanobody dirigés contre la partie périplasmique pPorM ont été générés.

La protéine pPorM a été exprimée et purifiée par Philippe Léone. Après 4 injections de 0,5 mg à une semaine d'intervalle, des échantillons de sang ont été prélevés au lama immunisé. Les lymphocytes ont été isolés puis les nanobody ont été sélectionnés par Aline Desmyter selon la méthode précédemment décrite (55) (Figure 3.2.1). Deux librairies de nanobody ont été constituées, l'une issue de cette immunisation, l'autre naïve.

Après vérification par test ELISA, 5 nanobody, portant des séquences de CDR différentes, ont été retenus : nb01, nb02, nb06, nb19 et nb130. Alors que les 4 premiers nanobody sont issus de la librairie de lama immunisé, le nanobody nb130 est lui issu de la librairie naïve.

Les sous-constructions de la protéine pPorM ont été imaginées par Philippe Léone puis clonées par moi-même. Les protéines pPorM-Nt (résidus 44 à 217) et pPorM-Ct (résidus 225 à 516) ont ainsi pu être exprimées en système bactérien puis purifiées.

Des tests d'interaction, menés par Christine Kellenberger, ont été réalisés sur chaque nanobody contre les différentes constructions de pPorM. Les résultats ont permis de déterminer pour chacun la région d'interaction sur la protéine pPorM : les nanobody nb01 et nb02 se fixent sur la partie N-terminale (pPorM-Nt) alors que les nanobody nb130 et nb19 se fixent sur la partie C-terminale (pPorM-T et pPorM-Ct).

Les quatre nanobody ont été exprimés, purifiés et cristallisés. J'ai participé à la collecte des jeux de données de nb01, nb02 et nb130 et j'ai résolu la structure du nanobody nb02.

J'ai également réalisé des essais de cristallisation du complexe pPorM-Nt/nb01 qui se sont avérés concluants. Un jeu de données a été collecté dont les paramètres sont présentés dans l'article suivant, publié, pour lequel je suis co-premier auteur(5).



Camelid nanobodies used as crystallization chaperones for different constructs of PorM, a component of the type IX secretion system from *Porphyromonas gingivalis*

Yoan Duhoo,^{a,b} ‡ Jennifer Roche,^{a,b} ‡ Thi Trang Nhung Trinh,^{a,b} ‡ Aline Desmyter,^{a,b} Anaïs Gaubert,^{a,b} Christine Kellenberger,^{a,b} Christian Cambillau,^{a,b} Alain Roussel^{a,b,*} and Philippe Leone^{a,b,*}

Received 25 January 2017

Accepted 19 April 2017

Edited by R. L. Stanfield, The Scripps Research Institute, USA

‡ These authors contributed equally to this work.

Keywords: camelid nanobodies; type IX secretion system; crystallization chaperones; PorM; *Porphyromonas gingivalis*.

PDB references: nb01, 5lz0; nb02, 5lmw; nb19, 5lmj; nb130, 5fwo

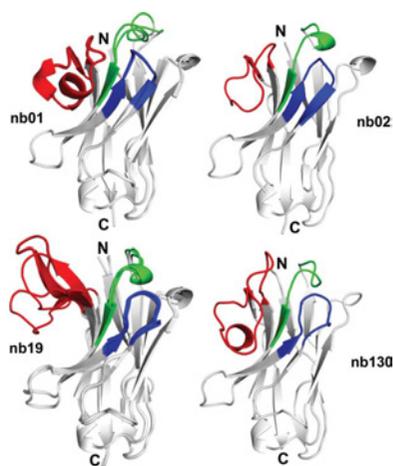
Supporting information: this article has supporting information at journals.iucr.org/f

^aCentre National de la Recherche Scientifique, Architecture et Fonction des Macromolécules Biologiques, UMR 7257, Marseille, France, and ^bAix-Marseille Université, Architecture et Fonction des Macromolécules Biologiques, UMR 7257, Marseille, France. *Correspondence e-mail: alain.roussel@afmb.univ-mrs.fr, philippe.leone@afmb.univ-mrs.fr

PorM is a membrane protein that is involved in the assembly of the type IX secretion system (T9SS) in *Porphyromonas gingivalis*, a major bacterial pathogen that is responsible for periodontal disease in humans. In the context of structural studies of PorM to better understand T9SS assembly, four camelid nanobodies were selected, produced and purified, and their specific interaction with the N-terminal or C-terminal part of the periplasmic domain of PorM was investigated. Diffracting crystals were also obtained, and the structures of the four nanobodies were solved by molecular replacement. Furthermore, two nanobodies were used as crystallization chaperones and turned out to be valuable tools in the structure-determination process of the periplasmic domain of PorM.

1. Introduction

Periodontal disease is one of the most frequently occurring infectious diseases in humans (Armitage, 1996). The main microorganism responsible for periodontitis and gingivitis is the oral pathogenic bacterium *Porphyromonas gingivalis*. Infection by this Gram-negative bacterium causes severe lesions in periodontal tissues, leading to destruction of the alveolar bone and the tooth-supporting structure in the most severe cases. This damage is induced by a cocktail of specialized toxin proteins secreted by *Porphyromonas*, called gingipains. Gingipains act as adhesins and/or proteases that help the bacterium to adhere to periodontal tissues and to promote the invasion of gingival tissues by the degradation of matrix proteins, fibrinogen and collagen. The active release of gingipains at the bacterial cell surface is catalyzed by a recently identified protein complex called the type IX secretion system (T9SS; Sato *et al.*, 2010). The T9SS is composed of 10–14 subunits, encoded by the *por* genes, that are thought to assemble a transenvelope channel that specifically recruits the gingipains and transports them to the cell surface, where they are processed to become active (by cleavage of a C-terminal sequence and glycosylation). Interestingly, four of these proteins, PorK, PorL, PorM and PorN, are sufficient to assemble a stable complex of ~1.2 MDa that resists native PAGE analyses (Sato *et al.*, 2010). In a recent study (Vincent *et al.*, 2017), we have shown that PorK is a lipoprotein anchored



© 2017 International Union of Crystallography

in the outer membrane that interacts strongly with the periplasmic protein PorN, and that PorL and PorM are inner membrane proteins which interact with each other *via* their transmembrane segments. In addition to its transmembrane segment, PorM possesses a long periplasmic domain of nearly 500 residues which is thought to link the inner and outer membrane subcomplexes.

A crystallization strategy for PorM is to use camelid nanobodies as crystallization chaperones. In addition to conventional antibodies, members of the Camelidae family possess a set of unusual antibodies that consist only of heavy chains; the constant domain is missing, and they are therefore named heavy-chain-only antibodies or nanobodies (Hamers-Casterman *et al.*, 1993). These antibodies offer a very interesting new tool in scientific research, as they contain variable domains (VHHs) that are fully responsible for and are fully capable of recognizing and binding to antigens. The variable domains are also more hydrophilic than their IgG counterparts, as they are not required to bind to a complementary domain of a light chain (Spinelli *et al.*, 1996). Moreover, they are very stable, and owing to their rather small size (14 kDa) they are able to bind to epitopes within clefts (Desmyter *et al.*, 1996) that are more difficult to reach for larger antibodies. Both their single-domain nature and their lack of glycosylation allow them to be produced at high levels using bacterial expression systems.

Using the nanobody platform of our laboratory, we were able to raise four nanobodies against the periplasmic domain of PorM (residues 36–516). Here, we present an investigation of their binding specificities and their crystal structures. Moreover, diffracting crystals of the N- and C-terminal parts of PorM in complex with their respective nanobodies were obtained. These crystals were essential in the structure-determination process of the targets, validating the strategy of using nanobodies as crystallization chaperones.

2. Materials and methods

2.1. Generation of nanobodies against pPorM

Four injections of 0.5 mg purified periplasmic domain of PorM (pPorM) in 10 mM HEPES pH 7.4, 150 mM NaCl were performed subcutaneously at one-week intervals followed by a fifth injection two weeks later in one llama (*Llama glama* from Ardèche Lamas, France). Lymphocytes were isolated from blood samples obtained 5 d after the last immunization. The cDNA was synthesized from purified total RNA by reverse transcription and was used as a template for PCR amplification to amplify the sequences corresponding to the variable domains of the heavy-chain antibodies. PCR fragments were then cloned into the phagemid vector pHEN4 (Arbabi Ghahroudi *et al.*, 1997) to create a nanobody phage-display library. The selection and screening of nanobodies were performed as described previously (Desmyter *et al.*, 2013). Three rounds of panning resulted in the isolation of pPorM-specific binders. After sequence analysis, four different positive clones, nb01, nb02, nb06 and nb19, were chosen to be

subcloned into pHEN6 expression vector downstream of the pelB signal peptide and fused to a C-terminal 6×His tag (Conrath *et al.*, 2009).

A nanobody phage-display naïve library was generated by the same procedure as described above, starting from the blood of several non-immunized animals: three camels (*Camelus dromedarius* from C. Gutierrez at Universidad de Las Palmas de Gran Canaria, Spain) and four llamas (*L. glama* from Capralogics Inc., USA). Three rounds of panning resulted in the isolation of pPorM-specific binders. After sequence analysis, three different positive clones, nb103, nb130 and nb105, were found corresponding to nb02, nb06 and nb19, respectively, from the immunized library. Clone nb130 was subcloned into pHEN6 expression vector.

2.2. Production and purification of nanobodies

The expression and purification of the nanobodies was performed as described previously (Conrath *et al.*, 2009). The periplasmic fraction was extracted by osmotic shock and the recombinant nanobodies were purified on a 5 ml Ni-NTA column (GE Healthcare) in 50 mM sodium/potassium phosphate pH 8.0, 300 mM NaCl, 10% glycerol. The fractions eluted in 250 mM imidazole were concentrated by centrifugation using an Amicon Ultra 10 kDa cutoff concentrator prior to being loaded onto a HiLoad 10/30 Superdex 75 pg gel-filtration column (GE Healthcare) equilibrated in phosphate-buffered saline (PBS). The purified nanobodies were concentrated by centrifugation; their concentration was determined by measuring the absorbance at 280 nm with a NanoDrop 2000 (Thermo Scientific).

2.3. Cloning, production and purification of PorM constructs

pPorM and its trypsinized fragment (pPorM-T) were produced as described previously (Stathopoulos *et al.*, 2015). The sequence of pPorM-T was deduced by N-terminal Edmann sequencing and peptide mass fingerprinting (results not shown), and corresponds to the C-terminal part of pPorM (residues 225–516). This sequence (hereafter denoted pPorM-Cter) and the sequence corresponding to the remaining N-terminal part of pPorM (residues 44–217; hereafter denoted pPorM-Nter) were cloned into the pET-28a+ derivative vector pLIC03. The pPorM-Cter and pPorM-Nter proteins were produced and purified using the same protocol as for pPorM.

2.4. Purification of nanobodies in complex with PorM constructs

The nanobodies and the PorM constructs pPorM-Nter and pPorM-Cter were first purified independently by nickel-affinity chromatography as described previously. The proteins were then mixed in a molar ratio of 1:1.2 (PorM constructs: nanobodies) and left for 15 min at room temperature to form the following complexes: pPorM-Nter in complex with nb01 or nb02, and pPorM-Cter in complex with nb19 or nb130. The complexes were purified on a HiLoad 16/60 Superdex 75 pg gel-filtration column (GE Healthcare) equilibrated in 10 mM HEPES pH 7.5, 150 mM NaCl. The purified complexes were

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	nb01	nb02	nb19	nb130
Crystallization condition	0.1 M trisodium citrate pH 5.5, 20% (w/v) PEG 3000	0.1 M trisodium citrate pH 5.5, 20% PEG 3000	0.1 M HEPES pH 4.5–5.5, 0.4–1.2 M NaH ₂ PO ₄ , 0.4–1.2 M KH ₂ PO ₄	0.2 M diammonium citrate pH 5.0, 20% (w/v) PEG 3350
Protein concentration (mg ml ⁻¹)	14.0	13.0	10.1	16.0
Data collection				
Cryoprotectant	20% (v/v) polypropylene glycol	10% glycerol	30% (v/v) glycerol	20% (v/v) glycerol
Diffraction source	ID-23, ESRF	ID-23, ESRF	ID-29, ESRF	PROXIMA-2, SOLEIL
Space group	C2	P4 ₃ 2 ₁ 2	P4 ₁ 2 ₁ 2	P6 ₁ 22
<i>a</i> , <i>b</i> , <i>c</i> (Å)	78.33, 65.83, 48.31	65.68, 65.68, 88.80	113.26, 113.26, 153.45	54.54, 54.54, 126.72
α , β , γ (°)	90, 95.89, 90	90, 90, 90	90, 90, 90	90, 90, 120
Resolution (Å)	28.8–1.60 (1.69–1.60)	46.44–1.50 (1.58–1.50)	48.1–2.10 (2.21–2.10)	47.23–1.70 (1.79–1.70)
Wavelength (Å)	0.87260	0.8729	1.07224	0.9801
Unique reflections	31426 (4389)	31805 (4532)	58799 (8431)	13038 (1821)
Multiplicity	5.3 (5.5)	6.8 (6.5)	8.7 (8.2)	21.0 (21.1)
Completeness (%)	97.7 (94.1)	99.9 (99.4)	99.9 (99.9)	99.9 (99.6)
$\langle I/\sigma(I) \rangle$	10.5 (1.7)	22.0 (1.7)	15.0 (1.7)	24.6 (2.4)
$R_{\text{meas}}^{\dagger}$ (%)	11.8 (111.1)	5.1 (124.5)	7.9 (144.0)	8.5 (139.6)
$R_{\text{p.i.m.}}^{\ddagger}$ (%)	5.1 (46.6)	1.9 (48.2)	2.6 (47.9)	1.9 (30.2)
CC _{1/2} [§]	0.997 (0.693)	1.000 (0.627)	0.998 (0.652)	0.999 (0.850)
Mosaicity (°)	0.21	0.23	0.05	0.15
Solvent content (%)	40.6	63.3	70.8	37.2
Molecular replacement				
PDB code of starting model	4tvs	4qgy	4hem	4fhh
Refinement and model quality				
Resolution (Å)	27.2–1.6	19.58–1.5	46.6–2.1	44.2–1.7
No. of reflections	31426	31733	58723	12963
$R_{\text{cryst}}/R_{\text{free}}^{\parallel}$ (%)	20.5/23.2	18.8/19.6	18.8/21.0	18.6/22.2
No. of atoms				
Protein [chain(s) in asymmetric unit]	1902 [2]	957 [1]	3833 [4]	1000 [1]
Water/ion/ligand	251/—/—	213/—/6	404/55/—	108/—/—
Average <i>B</i> factors (Å ²)				
Protein	21.4	27.5	68.5	33.4
Water/ion/ligand	32.9/—/—	46.4/—/29.9	73.0/81.4/—	46.1/—/—
R.m.s.d. ^{††}				
Bonds (Å)	0.01	0.01	0.01	0.01
Angles (°)	0.97	0.98	1.11	1.05
Ramachandran plot (%)				
Most favoured regions	96.3	94.0	97.9	97.5
Additionally allowed regions	3.7	5.2	1.7	2.5
PDB code	5lz0	5lmw	5lmj	5fwo

	nb01–pPorM-Nter	nb130–pPorM-Cter		
Crystallization condition	0.1 M bis-tris pH 7.0, 25% (w/v) PEG 3350	0.1 M HEPES pH 7.5, 0.2 M NaCl, 25% (w/v) PEG 3350	0.2 M ammonium citrate tribasic pH 7.0, 20% (w/v) PEG 3350	
Protein concentration (mg ml ⁻¹)	10.0	10.1	10.1	10.1
Data collection				
Cryoprotectant	10% (v/v) glycerol	20% (v/v) glycerol	10% (v/v) PEG 400	10% (v/v) PEG 400, 30% (w/v) PEG 3350
Diffraction source	PROXIMA-1, SOLEIL	PROXIMA-1, SOLEIL	PROXIMA-1, SOLEIL	PROXIMA-1, SOLEIL
Space group	P2 ₁	P2	P2 ₁	P1
<i>a</i> , <i>b</i> , <i>c</i> (Å)	80.35, 100.12, 80.41	64.05, 78.11, 76.09	64.87, 157.40, 77.37	55.24, 77.18, 156.30
α , β , γ (°)	90, 93.82, 90	90, 104.69, 90	90, 105.51, 90	90.24, 91.75, 97.16
Resolution (Å)	42.47–2.40 (2.53–2.40)	40–2.50 (2.63–2.50)	40–2.15 (2.27–2.15)	40–2.10 (2.21–2.10)
Wavelength (Å)	0.97857	0.97857	0.97857	0.97857
Unique reflections	49059 (7020)	25205 (3620)	80438 (11626)	145898 (21179)
Multiplicity	6.9 (6.8)	3.8 (3.7)	4.2 (4.3)	2.9 (2.9)
Completeness (%)	98.3 (96.7)	99.4 (97.7)	99.3 (98.7)	97.6 (96.8)
$\langle I/\sigma(I) \rangle$	11.6 (1.0)	8.6 (1.0)	12.5 (1.8)	12.5 (2.0)
$R_{\text{meas}}^{\dagger}$ (%)	7.8 (203.3)	12.1 (70.6)	8.1 (92.1)	6.5 (76.4)
$R_{\text{p.i.m.}}^{\ddagger}$ (%)	3.0 (77.4)	6.2 (36.2)	3.9 (44.0)	3.7 (43.3)
CC _{1/2} [§]	0.999 (0.755)	0.995 (0.766)	0.999 (0.660)	0.998 (0.647)
Mosaicity (°)	0.05	0.15	0.15	0.19
Solvent content (%)	43.4	48.9	50.6	43.1

[†] $R_{\text{meas}} = \sum_{hkl} \{N(hkl)/[N(hkl) - 1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity and $\langle I(hkl) \rangle$ is the average intensity from $N(hkl)$ observations (symmetry-related and duplicate measurements of a unique reflection). [‡] $R_{\text{p.i.m.}} = \sum_{hkl} \{1/[N(hkl) - 1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$. [§] CC_{1/2} values are the half-set correlation coefficients (Karplus & Diederichs, 2012). [¶] $R_{\text{cryst}} = \sum_{hkl} |F_{\text{obs}}| - |F_{\text{calc}}| / \sum_{hkl} |F_{\text{obs}}|$; R_{free} is calculated for a randomly selected 5% of reflections that were excluded from refinement. ^{††} Root-mean-square deviation from ideal values.

concentrated by centrifugation; their concentration was determined by measuring the absorbance at 280 nm with a NanoDrop 2000 (Thermo Scientific).

2.5. Biolayer interferometry (BLI)

The proteins were biotinylated using the EZ-Link NHS-PEG4-Biotin kit (Perbio Science, France). The reaction was quenched by removing the excess biotin using a Zeba Spin Desalting column (Perbio Science, France) which was equilibrated with PBS buffer. BLI studies were performed in black 96-well plates (Greiner) at 25°C using an Octet RED96 (ForteBio, USA; PBS, 0.01% BSA, 0.002% Tween 20, 0.005% sodium azide). Streptavidin biosensor tips (ForteBio, USA) were first hydrated with 0.2 ml kinetic buffer (KB; ForteBio, USA; PBS, 0.01% BSA, 0.002% Tween20 and 0.005% sodium azide) for 20 min and then loaded with biotinylated pPorM-Nter, pPorM-Cter or pPorM (10 µg ml⁻¹ in KB). The surface was then blocked using biocytin. The association of PorM with various concentrations of nanobodies (a zero concentration and concentrations varying from 0.5 to 250 nM, depending on the experiment) was monitored for 300 s, and the dissociation was followed for 600 or 900 s. in KB. These experiments were run two times independently. Control experiments were run to check that there was no nonspecific interaction between the analytes (nanobodies) and control biosensors (loaded with no protein and blocked with biocytin). Fitting of the binding data and the measurement of constants were performed with the Octet RED system software (v.7.1) using a 1:1 model. The error values are calculated from two independent experiments; averaged statistical parameters (χ^2 and R^2) generated during the fittings are reported.

2.6. Crystallization, data collection and processing

Initial crystallization trials of the nanobodies and of their complexes with PorM constructs were performed by the sitting-drop vapour-diffusion method at 293 K in 96-well Greiner plates using a Mosquito Crystal robot (TTP Labtech)

with the following screens: Wizard I and II (Rigaku Reagents), JCSG+ (Qiagen), Index (Hampton Research), Crystal Screen and Crystal Screen 2 (Hampton Research). Drops were prepared by mixing different volumes (100, 200 and 300 nl) of protein solution and 100 nl precipitant solution and were equilibrated against a 150 µl reservoir volume. Initial crystallization conditions were obtained in several conditions. For nb19 only, optimization was then carried out by varying the pH and the concentration of the precipitant (Lartigue *et al.*, 2003). The final crystallization conditions are given in Table 1. Crystals were mounted in cryoloops (Hampton CrystalCap Magnetic) and were briefly soaked in crystallization solution supplemented with the appropriate cryoprotectant before being flash-cooled in a nitrogen-gas stream at 100 K using a home cryocooling device (Oxford Cryosystems). For the nb130-pPorM-Cter complex, a 5 min air-dehydration step was performed in the crystallization solution with a 10% increase in the precipitant concentration and supplemented with the cryoprotectant.

Diffraction data were collected on beamlines ID23 and ID29 at the European Synchrotron Research Facility (ESRF), Grenoble, France and on PROXIMA-1 and PROXIMA-2 at SOLEIL, Paris, France. The data sets were integrated with XDS (Kabsch, 2010) and were scaled with SCALA (Evans, 2006) from the CCP4 suite (Winn *et al.*, 2011). Data-collection statistics are reported in Table 1. The Matthews coefficient and solvent content were calculated with MATTHEWS_COEF (Kantardjieff & Rupp, 2003) from the CCP4 suite.

2.7. Molecular replacement and refinement

Molecular replacement was performed with MOLREP (Vagin & Teplyakov, 2010) from the CCP4 suite. A different starting model was used for each nanobody, according to the result of sequence alignment against the wwPDB. For nb01, nb02 and nb130 molecular replacement was straightforward. For nb19, MOLREP returned a well contrasted solution for three VHHs. After refinement with autoBUSTER (Blanc *et*

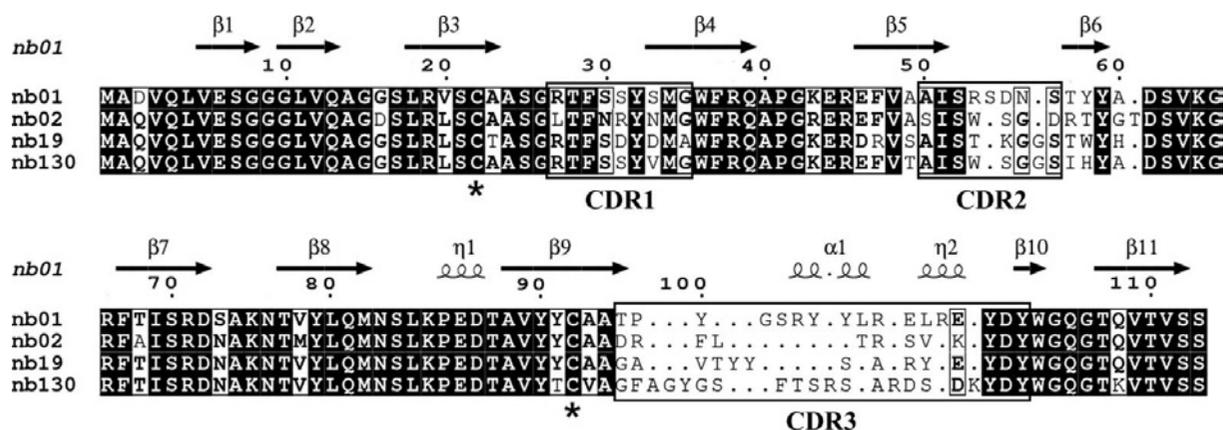


Figure 1 Sequence alignment of the four nanobodies nb01, nb02, nb19 and nb130 raised against the periplasmic domain of PorM. Secondary-structure elements from the nb01 structure (molecule *A*) are displayed above the alignment. The CDR1, CDR2 and CDR3 sequences are boxed; the cysteine residues are denoted by asterisks. Kabat numbering is used (Kabat *et al.*, 1991). The sequence alignment was generated by DeepAlign (Wang *et al.*, 2013) on the basis of the superposition of the four nanobody structures (molecule *A* for nb01 and nb19) and processed by ESPript (Robert & Gouet, 2014).

Table 2

Kinetic and thermodynamic parameters of the interactions between the nanobodies nb01, nb02, nb19 and nb130 with different constructs of the PorM periplasmic domain.

	K_d (nM)	K_{ass} ($M^{-1} \text{ s}^{-1}$)	K_{diss} (s^{-1})	Full χ^2_{full} – R^2
pPorM–nb01	1.8 ± 0.7	$8.1 \pm 2.4 \times 10^5$	$1.4 \pm 0.2 \times 10^{-3}$	1.38–0.98
pPorM–nb02	2.5 ± 1.5	$3.1 \pm 1.0 \times 10^5$	$7.7 \pm 1.5 \times 10^{-4}$	1.88–0.98
pPorM–nb19	3.8 ± 1.4	$1.7 \pm 0.4 \times 10^5$	$6.7 \pm 0.6 \times 10^{-4}$	0.35–0.98
pPorM–nb130	4.5 ± 1.0	$1.4 \pm 0.3 \times 10^5$	$6.3 \pm 0.3 \times 10^{-4}$	1.29–0.99
pPorM–Nter–nb01	4.0 ± 1.1	$2.0 \pm 0.7 \times 10^5$	$8.0 \pm 0.8 \times 10^{-4}$	3.49–0.99
pPorM–Nter–nb02	5.5 ± 1.2	$1.6 \pm 0.5 \times 10^5$	$9.0 \pm 0.6 \times 10^{-4}$	3.98–0.99
pPorM–Cter–nb19	8.6 ± 1.4	$6.7 \pm 1.5 \times 10^4$	$5.8 \pm 0.2 \times 10^{-4}$	5.44–0.99
pPorM–Cter–nb130	8.0 ± 1.5	$6.8 \pm 1.2 \times 10^4$	$5.6 \pm 0.2 \times 10^{-4}$	4.97–0.99

al., 2004), electron density for a fourth VHH could be identified. However, even using the refined model of the three VHHS, *MOLREP* could not localize this fourth VHH. The amino-acid chain was therefore traced manually with *TURBO* (Roussel & Cambillau, 1991) until there was sufficient structural information to allow the positioning of a complete VHH by rotation/translation.

Refinement of the four-nanobody model was performed with *autoBUSTER* and *PHENIX* (Adams *et al.*, 2010), and the structures were corrected with *Coot* (Emsley *et al.*, 2010). Model validations were performed with *MolProbity* (Chen *et al.*, 2010). Refinement statistics are reported in Table 1.

3. Results

3.1. Nanobody generation and binding characterization

Nanobodies were raised by immunization of llamas with the purified periplasmic domain of the PorM protein (residues

36–516; pPorM). Several strong pPorM binders were identified from the immune library by three rounds of panning using phage display coupled to ELISA. Four nanobodies, called nb01, nb02, nb09 and nb130, were selected for further study based on their high affinity for pPorM and on their amino-acid differences in the variable regions [also called the complementarity-determining regions (CDRs); Fig. 1]. Interestingly, nb02, nb19 and nb130 were also present in a naïve library generated in the laboratory (article in preparation). The four nanobodies possess the two cysteines Cys22 and Cys92 that form the conserved disulfide bond that is present in all VHH domains. The CDR1 and CDR2 of the four nanobodies have the same length (nine and eight residues, respectively), but there is a great variation in the length of the CDR3: 17, 12, 14 and 21 residues for nb01, nb02, nb19 and nb130, respectively.

The nanobodies were produced in the periplasm of *Escherichia coli* and purified to homogeneity. They all behaved as monomers in size-exclusion chromatography. In a previous study (Stathopoulos *et al.*, 2015), we have shown that a fragment of the purified pPorM could be obtained by limited proteolysis (hereafter denoted pPorM-T). Using N-terminal Edmann sequencing and peptide mass fingerprinting (results not shown), we deduced that this fragment corresponds to the C-terminal part of pPorM (residues 225–516; hereafter denoted pPorM-Cter). We then produced this domain as well as the remaining N-terminal part of pPorM (residues 44–217; hereafter denoted pPorM-Nter), and their interaction with the four nanobodies was assessed by biolayer interferometry (BLI). The four nanobodies bind to pPorM in the nanomolar range. Interestingly, they present a specificity in their interaction with pPorM: nb01 and nb02 bind specifically to

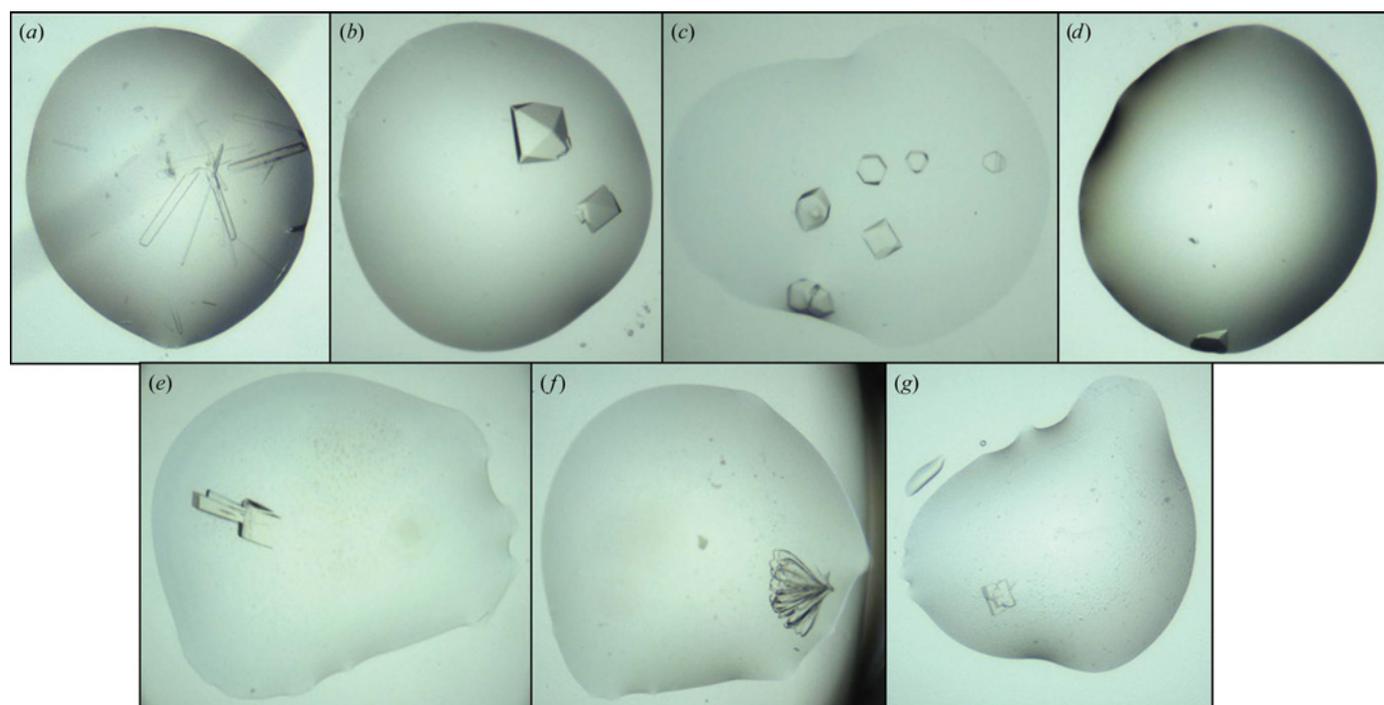


Figure 2

Crystals of nb01 (a), nb02 (b), nb19 (c) and nb130 (d), of pPorM-Cter in complex with nb130 in space group $P2_1$ (e) and space group $P2$ (f), and of pPorM-Nter in complex with nb01 (g).

pPorM-Nter, whereas nb19 and nb130 bind specifically to pPorM-Cter. The results are summarized in Table 2.

3.2. Nanobody structures

We decided to solve the structures of the four nanobodies alone for the following two reasons. On one hand, high-resolution structures of the nanobodies could be used either to solve the structure of the complex with the protein of interest by molecular replacement or to help in building in the case of *ab initio* phasing. On the other hand, as there are only a limited number of unbound nanobody structures in the PDB, new structures will provide more data for potential use in theoretical studies of single-chain antibody binding. Thus, the four nanobodies were crystallized (Fig. 2) and their structures were solved by molecular replacement. As expected, the four nanobodies adopt the classical immunoglobulin fold (Fig. 3) with two β -sheets composed of four antiparallel β -strands:

$\beta 1-\beta 3-\beta 7-\beta 8$ and $\beta 4-\beta 5-\beta 6-\beta 9$. The former β -sheet is flanked by CDR3 and is further prolonged by two parallel β -strands ($\beta 2$ and $\beta 11$). One π -helix is present in the $\beta 8-\beta 9$ loop. The structures of the four nanobodies display a high level of homology, with an r.m.s.d. ranging from 0.69 to 0.99 Å on C $^{\alpha}$ atoms without the CDRs. The main differences arise from the CDRs, which are located at the N-terminal edge of the β -sandwich. Indeed, CDR3, which generally mediates interaction with the epitope, is the most divergent. Its folding is different in the four structures: one α -helix and one π -helix in nb01, random coil only in nb02, random coil only or prolonged β -strands in nb19 and one π -helix in nb130 (Fig. 3). Paradoxically, the long CDR3s of nb01 and nb130 (17 and 21 residues, respectively) adopt a unique conformation in the crystal, suggesting a certain conformational stability. On the contrary, despite its shorter sequence (14 residues), the CDR3 of nb19 is more flexible, as it can adopt multiple distinct conformations. The CDR3 conformational heterogeneity

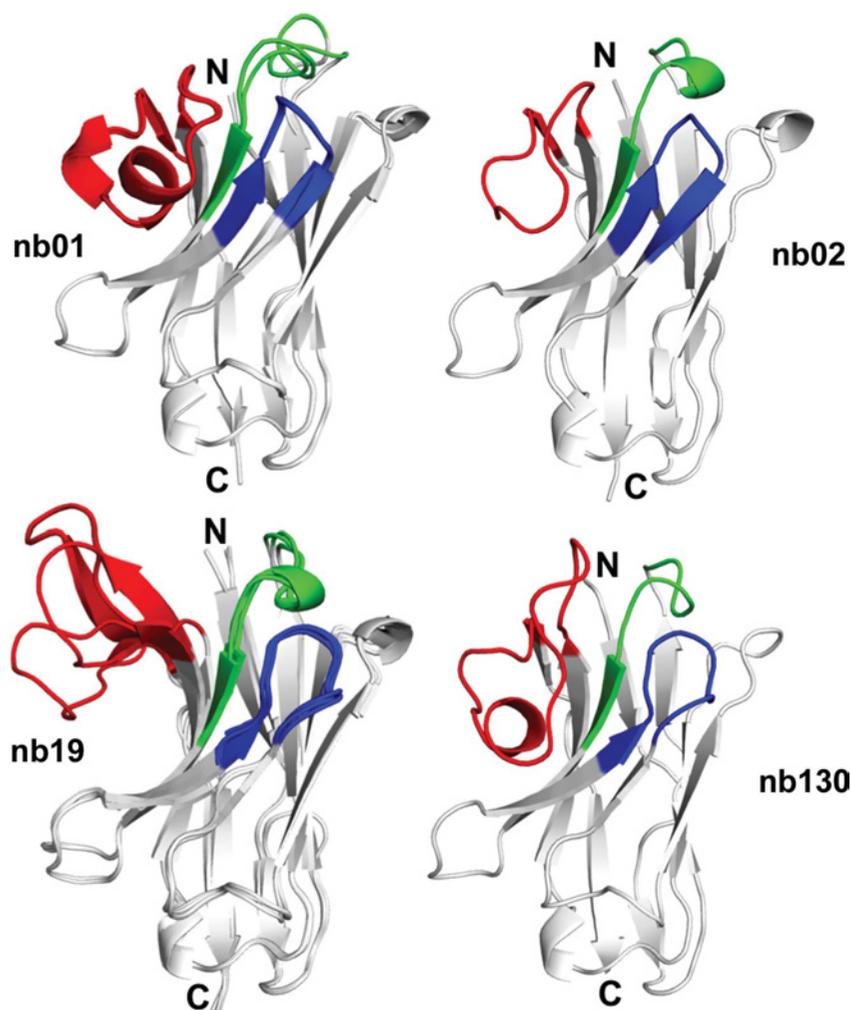


Figure 3

Crystal structures of the four nanobodies nb01, nb02, nb19 and nb130 raised against the periplasmic domain of PorM. CDR1, CDR2 and CDR3 of each nanobody according to Kabat numbering (Kabat *et al.*, 1991) are represented in green, blue and red, respectively. The N- and C-terminal extremities are labelled N and C, respectively. The different molecules present in the asymmetric unit are superposed (molecules A and B of nb01 and molecules A, B and D of nb19; molecule C of nb19, which is identical to molecule B, is not represented for the purposes of clarity). This figure was generated by PyMOL (Schrödinger).

observed only in nb19 could be related to the substitution of the Phe47 present in nb01, nb02 and nb130 by an arginine residue in nb19 (Fig. 1). Indeed, the CDR3 of nb01, nb02 and nb130 folds over and packs against the hydrophobic side chain of the framework Phe47, while the CDR3 of nb19 is somewhat repulsed by the side chain of Arg47.

3.3. Use of nanobodies as crystallization chaperones

We have previously obtained diffracting crystals of pPorM-T and of its selenomethionine derivative; native and MAD data sets were collected and an experimental map was calculated (Stathopoulos *et al.*, 2015). However, building of the C-terminal domain of pPorM-T was challenging, probably owing to the higher thermal motion of this domain. We decided to use nanobodies as crystallization chaperones, with the assumption that complex formation may minimize the conformational flexibility of the target. Therefore, crystallization trials of pPorM-Cter in complex with its specific nanobodies nb19 and nb130 were performed. Crystals of pPorM-Cter in complex with nb130 were obtained in space groups $P2$ and $P2_1$ that diffracted to 2.5 and 2.15 Å resolution, respectively (Fig. 2). In an attempt to improve the resolution, the $P2_1$ crystals were dehydrated, resulting in crystals in a new space group, $P1$, that diffracted to 2.1 Å resolution. Structure determination of the complex is in progress using the nb130 structure and the partial model of pPorM-T as a model for molecular replacement.

The same strategy of co-crystallization with nanobodies was applied in order to overcome our inability to obtain exploitable pPorM-Nter crystals. Indeed, all of the native or selenomethionine-derivative crystals that we obtained for this construct were twinned and therefore were not usable to solve the structure *ab initio*. Crystallization trials of complexes with nb01 or nb02 were then performed and crystals of pPorM-Nter in complex with nb01 were obtained that diffracted to 2.4 Å resolution (Fig. 2).

4. Concluding remarks

In recent years, nanobodies have been intensively used as crystallization chaperones in structural studies. Indeed, they preferentially bind concave and hinge regions of proteins and thus can help in the crystallization process of recalcitrant targets such as multi-domain proteins, large complexes and membrane proteins (Desmyter *et al.*, 2015). In our case, nanobodies also turned out to be valuable tools in the structure-determination process. For pPorM-Cter, co-crystallization with the nanobody nb130 permitted the resolution of the structure to be increased, presumably by stabilizing a flexible domain of the protein. For pPorM-Nter, co-crystallization with the nanobody nb01 permitted us to obtain untwinned crystals, in contrast to the protein alone. Molecular replacement was performed with the structure of the nanobody as a starting model, providing the initial phases used to calculate electron-density maps. Recently, we solved the structure of a truncated construct of TssM using the same

strategy. TssM is a central component of the membrane complex of the type VI secretion system. The construct TssM_{32Ct} (residues 836–1129 of the TssM periplasmic domain) was co-crystallized with a specific nanobody (Nguyen *et al.*, 2015) and its structure was built into the electron-density maps obtained after molecular replacement using the structure of the nanobody as a starting model (Durand *et al.*, 2015). As pPorM-Nter is smaller than TssM_{32Ct} (211 and 293 residues, respectively) and as the crystals of the two proteins in complex with their respective nanobodies diffracted to similar resolutions (2.4 Å for pPorM-Nter and 1.92 Å for TssM_{32Ct}), we are confident that we will succeed in building the structure of pPorM-Nter.

Acknowledgements

We would like to thank the European Synchrotron Research Facility (ESRF) and the SOLEIL synchrotron for beamline allocation.

Funding information

Funding for this research was provided by: Centre National de la Recherche Scientifique; French Infrastructure for Integrated Structural Biology (FRISBI) (award No. ANR-10-INSB-05-01); Agence Nationale de la Recherche (award No. ANR-15-CE11-0019-01); Ministère de l'Éducation Nationale, de l'Enseignement Supérieur et de la Recherche (award No. doctoral fellowship for YD); Région PACA (award No. doctoral fellowship for JR); Eiffel program from the French Ministère des Affaires Étrangères (award No. doctoral fellowship for TNTT).

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Arbabi Ghahroudi, M., Desmyter, A., Wyns, L., Hamers, R. & Muyldermans, S. (1997). *FEBS Lett.* **414**, 521–526.
- Armitage, G. C. (1996). *Ann. Periodontol.* **1**, 37–215.
- Blanc, E., Roversi, P., Vornrhein, C., Flensburg, C., Lea, S. M. & Bricogne, G. (2004). *Acta Cryst.* **D60**, 2210–2221.
- Conrath, K., Pereira, A. S., Martins, C. E., Timóteo, C. G., Tavares, P., Spinelli, S., Kinne, J., Flaudrops, C., Cambillau, C., Muyldermans, S., Moura, I., Moura, J. J. G., Tegoni, M. & Desmyter, A. (2009). *Protein Sci.* **18**, 619–628.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Desmyter, A., Farenc, C., Mahony, J., Spinelli, S., Bebeacua, C., Blangy, S., Veesler, D., van Sinderen, D. & Cambillau, C. (2013). *Proc. Natl Acad. Sci. USA*, **110**, E1371–E1379.
- Desmyter, A., Spinelli, S., Roussel, A. & Cambillau, C. (2015). *Curr. Opin. Struct. Biol.* **32**, 1–8.
- Desmyter, A., Transue, T. R., Ghahroudi, M. A., Dao Thi, M.-H., Poortmans, F., Hamers, R., Muyldermans, S. & Wyns, L. (1996). *Nature Struct. Mol. Biol.* **3**, 803–811.
- Durand, E., Nguyen, V. S., Zoued, A., Logger, L., Péhau-Arnaudet, G., Aschtgen, M.-S., Spinelli, S., Desmyter, A., Bardiaux, B., Dujancourt, A., Roussel, A., Cambillau, C., Cascales, E. & Fronzes, R. (2015). *Nature (London)*, **523**, 555–560.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.

- Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hammers, C., Songa, E. B., Bendahman, N. & Hammers, R. (1993). *Nature (London)*, **363**, 446–448.
- Kabat, E. A., Wu, T. T., Perry, H. M., Gottesmann, K. S. & Foeller, C. (1991). *Sequences of Proteins of Immunological Interest*, 5th ed. Bethesda: National Institutes of Health.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 133–144.
- Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
- Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
- Lartigue, A., Gruez, A., Briand, L., Pernollet, J.-C., Spinelli, S., Tegoni, M. & Cambillau, C. (2003). *Acta Cryst.* **D59**, 919–921.
- Nguyen, V. S., Spinelli, S., Desmyter, A., Le, T. T. H., Kellenberger, C., Cascales, E., Cambillau, C. & Roussel, A. (2015). *Acta Cryst.* **F71**, 266–271.
- Robert, X. & Gouet, P. (2014). *Nucleic Acids Res.* **42**, W320–W324.
- Roussel, A. & Cambillau, C. (1991). *Silicon Graphics Geometry Partners Directory*, p. 81. Mountain View: Silicon Graphics.
- Sato, K., Naito, M., Yukitake, H., Hirakawa, H., Shoji, M., McBride, M. J., Rhodes, R. G. & Nakayama, K. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 276–281.
- Spinelli, S., Frenken, L., Bourgeois, D., de Ron, L., Bos, W., Verrips, T., Anguille, C., Cambillau, C. & Tegoni, M. (1996). *Nature Struct. Biol.* **3**, 752–757.
- Stathopoulos, J., Cambillau, C., Cascales, E., Roussel, A. & Leone, P. (2015). *Acta Cryst.* **F71**, 71–74.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* **D66**, 22–25.
- Vincent, M. S., Canestrari, M. J., Leone, P., Stathopoulos, J., Ize, B., Zoued, A., Cambillau, C., Kellenberger, C., Roussel, A. & Cascales, E. (2017). *J. Biol. Chem.* **292**, 3252–3261.
- Wang, S., Ma, J., Peng, J. & Xu, J. (2013). *Sci. Rep.* **3**, 1448.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.

3.2.3 Cristallisation du nanobody nb02 en ChipX et résolution de la structure

Dans le cadre d'une collaboration avec le laboratoire IBMC (Claude Sauter, équipe Biologie des ARNt et pathogénicité), j'ai eu l'occasion d'étudier la structure du nanobody nb02 à température ambiante. En tant que bêta-testeur de puces micro-fluidiques de cristallisation, les ChipX-3, il me fallait tester la cristallisation et la collecte de jeux de données sur des cristaux non congelés. Pour cela, j'ai choisi de cristalliser à nouveau ce nanobody.

3.2.3.1 Purification et cristallisation en ChipX-3 du nanobody nb02

Un lot de nb02 purifié quelques mois auparavant et conservé jusque là à 4°C a été utilisé. Il a été purifié une nouvelle fois, afin d'éliminer les agrégats et vérifier la qualité de la protéine, à l'aide d'une colonne d'exclusion de taille Superdex 75 16/60 (GE Healthcare) (Figure 3.2.2). Le pic élué en sortie de colonne était symétrique et son analyse sur gel ne présentait pas de contaminant. Une puce ChipX a été utilisée pour réaliser l'optimisation de la condition de cristallisation, initialement trouvée par la technique de cristallisation en diffusion vapeur, dans les mêmes conditions de concentration protéique (13,78 mg/ml). Des cristaux sont apparus après 4h puis ont continué de croître et apparaître pendant près d'un mois (Figure 3.2.3).

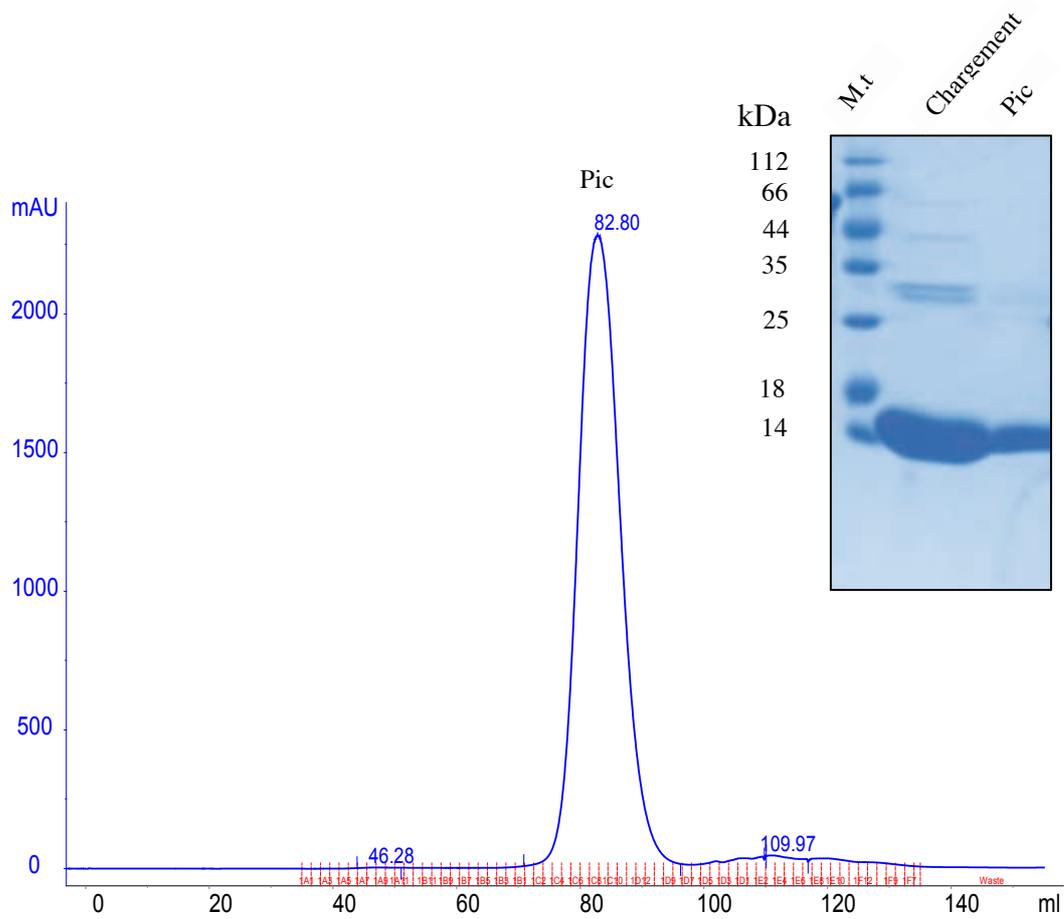


Figure 3.2.2: Purification du nanobody nb02 pour cristallisation en ChipX. Chromatogramme de la sortie de colonne Superdex 75 16/60. Le pic de protéine est élué après 82,8 ml de tampon. Insert: Analyse sur SDS-PAGE des fractions de purification. Une fraction du chargement et une du pic élué ont été déposés. M.t : Marqueur de taille.

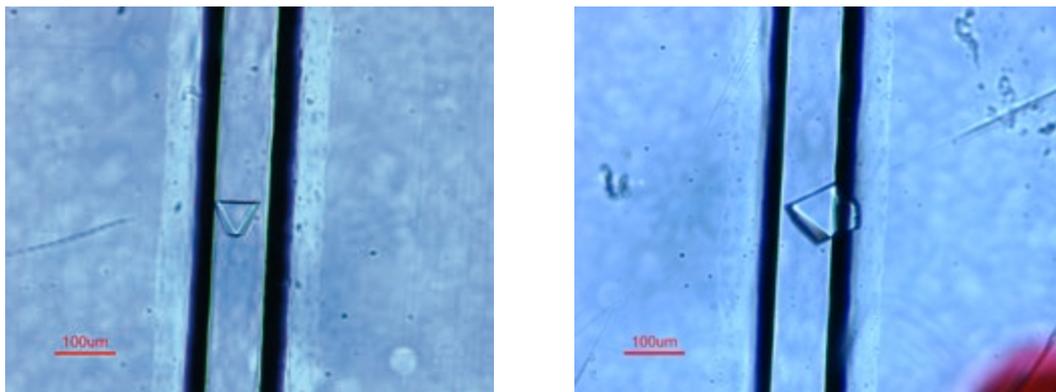


Figure 3.2.3: Exemples de cristaux de nb02 obtenus en ChipX-3. Les précipités cristallins sont apparus dans le micro-capillaire en quelques heures et les cristaux tels que sur les photos ont poussé en une semaine. La barre rouge en bas à gauche des photos représente une échelle de 100 μm .

3.2.3.2 Collecte in-situ des cristaux de nb02

Les cristaux du nanobody nb02 obtenus en ChipX-3 ont été collectés lors d'une mission sur la ligne de lumière PROXIMA 2A à SOLEIL (Gif-sur-Yvette). Deux jeux de données ont été enregistrés sur un cristal situé proche de l'entrée de la protéine, au niveau des capillaires C et D (EmbCD, Figure 3.2.4). Ils ont servi, après combinaison des données, à l'obtention d'un jeu de données complet à une résolution allant jusqu'à 1,9 Å. La structure a été déterminée par remplacement moléculaire à l'aide de la structure de nb02 obtenue à 100°K. Les statistiques de la collecte des données et de l'affinement sont présentées dans le Tableau 3.3.1. Après affinement, les deux structures cristallographiques ont été comparées par alignement à l'aide du logiciel PyMOL (Figure 3.2.5).

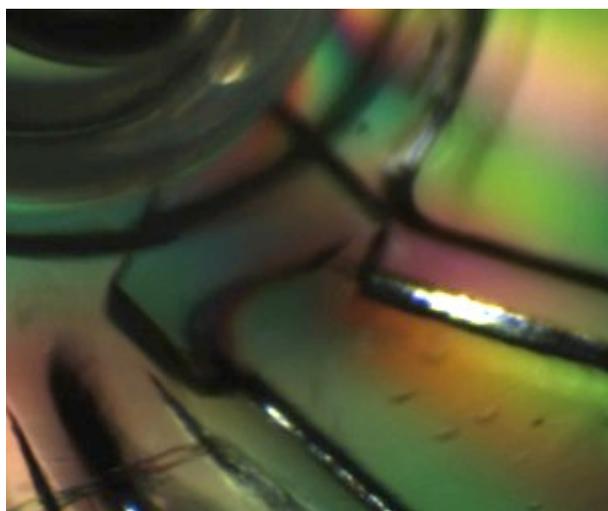


Figure 3.2.4: Cristal de nb02 en ChipX-3, collecté in-situ. Le cristal se trouve au niveau de l'entrée de la protéine, au milieu des capillaires C et D.

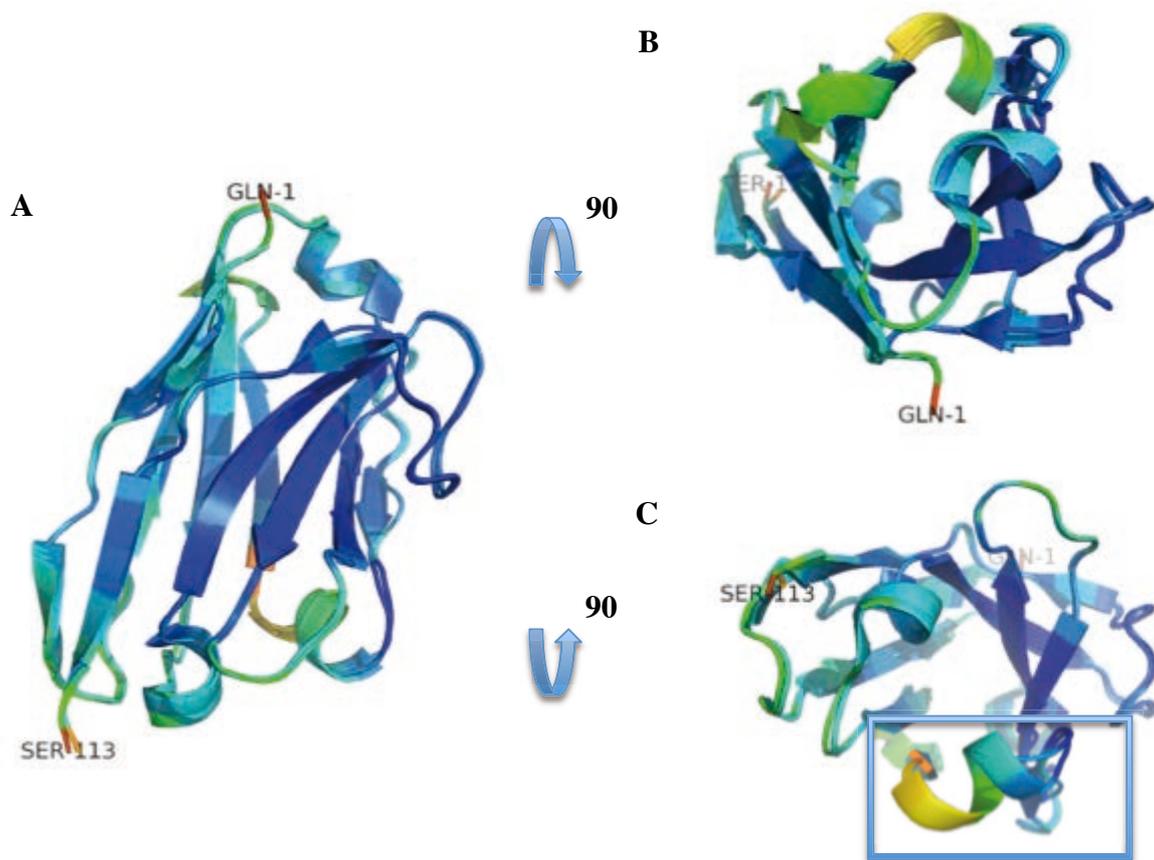


Figure 3.2.5: Comparaison des structures de nb02 collectées à température cryogénique (5LMW) et à température ambiante, par alignement à l'aide du logiciel PyMOL. Les structures ont été colorées selon leurs b-facteurs. Le résidu GLN-1 indique l'extrémité N-terminale et le résidu SER-113 l'extrémité C-terminale. A: Vue de côté. B : Vue de dessus par une rotation de 90°. C : Vue de dessous par une rotation de 90°. L'encadré bleu désigne la boucle G59-F67, principale différence entre les deux structures.

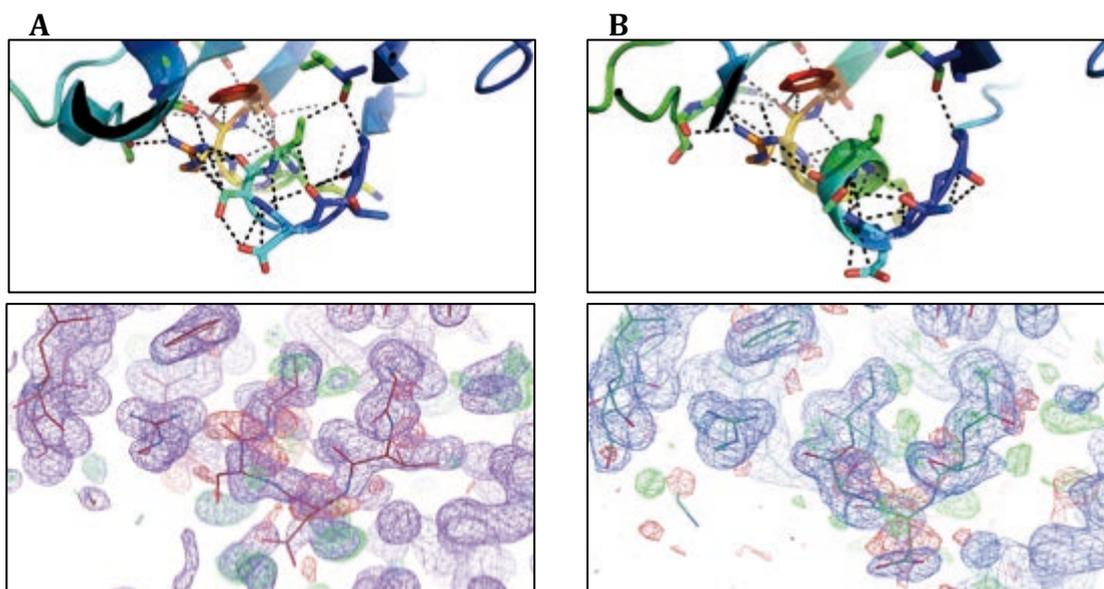


Figure 3.2.6: Zoom avant sur la région 59-67 des structures de nb02 à température cryogénique (A) et ambiante (B). Le réseau d'interaction des résidus a été représenté pour les deux structures à l'aide du logiciel PyMOL. Chaque distance inférieure à 3 Å a été affichée (pointillés noirs). Les cartes de densités Fo-Fc et 2Fo-Fc sont illustrées sous les structures correspondantes, avec des rmsd respectivement de 2,5 et 1,0.

D'un point de vue global, les deux structures n'ont que de légères différences (Figure 3.2.5A). La structure de nb02 collectée à température ambiante comporte de nombreuses conformations alternatives. Malgré tout, il ne manque aucune densité électronique dans la carte et le nanobody a pu être construit entièrement, boucles comprises. La structure issue du cristal congelé, quant à elle, avait donné plus de difficultés pendant la construction et l'affinement, principalement au niveau de la boucle comprenant les résidus 59 à 67, où la densité était faible (Figure 3.2.6). C'est précisément au niveau de cette boucle que se trouve la seule différence notable entre ces deux structures (Figure 3.2.5C). Le réseau des interactions des résidus de cette boucle a été cartographié (Figure 3.2.6A et B). Il présente des différences dues à des rotamères pour les résidus Thr60 et Ser62. Pour la structure à température ambiante, la boucle semble moins compacte, plus distante du feuillet bêta qui précède, résultant à une meilleure définition en hélice alpha.

3.3 Etudes structurales de pPorM

3.3.1 Structure de pPorM-Nt

La protéine pPorM-Nt correspond à la partie périplasmique N-terminale de PorM (résidus 44 à 217). Nous avons choisi dans l'équipe d'étudier cette protéine chez *E. coli*, en expression recombinante. Pour cela, j'ai réalisé le clonage dans le vecteur d'expression pLIC03 (fourni par BioXtal, optimisé pour la technologie « Ligation-Independent Cloning » LIC(56)) en utilisant la technique de RF cloning. Ce vecteur est dérivé du vecteur pET28 a+ (Novagen), il contient en amont de la séquence d'intérêt une étiquette de six histidines et un site de coupure à la protéase TEV, laissant la possibilité d'ôter l'étiquette. Après purification du produit de PCR et transformation en cellules DH5a, la séquence de l'ADN purifié a été vérifiée et validée. D'après une analyse in silico à l'aide de l'outil ProtParam d'Expasy (<http://web.expasy.org/protparam/>), il s'agit d'une construction de poids moléculaire 21,947 kDa.

3.3.1.1 Protéine pPorM-Nt native : expression, purification et analyses

La protéine pPorM-Nt a été exprimée en bactéries Rosetta (DE3) pLysS (Novagen). Après lyse des bactéries, la protéine est extraite du lysat clarifié par purification sur colonne de chromatographie d'affinité au Ni²⁺ (GE Healthcare). Elle est éluée à l'aide d'un tampon

contenant 250 mM d'Imidazole puis injectée sur une colonne d'exclusion de taille Superdex 75 16/60 (GE Healthcare) pour une dernière étape de purification. Le pic correspondant à pPorM-Nt est élué à un volume d'élution de 70 mL (Figure 3.3.1). D'après la calibration de cette colonne, ce volume d'élution correspondrait à une protéine globulaire de 19 kDa, semblant indiquer que pPorM-Nt est monomérique. La fraction du pic déposée sur SDS-PAGE montre une migration allant jusqu'à environ 23 kDa, ce qui correspond à notre poids moléculaire théorique (Figure 3.3.1).

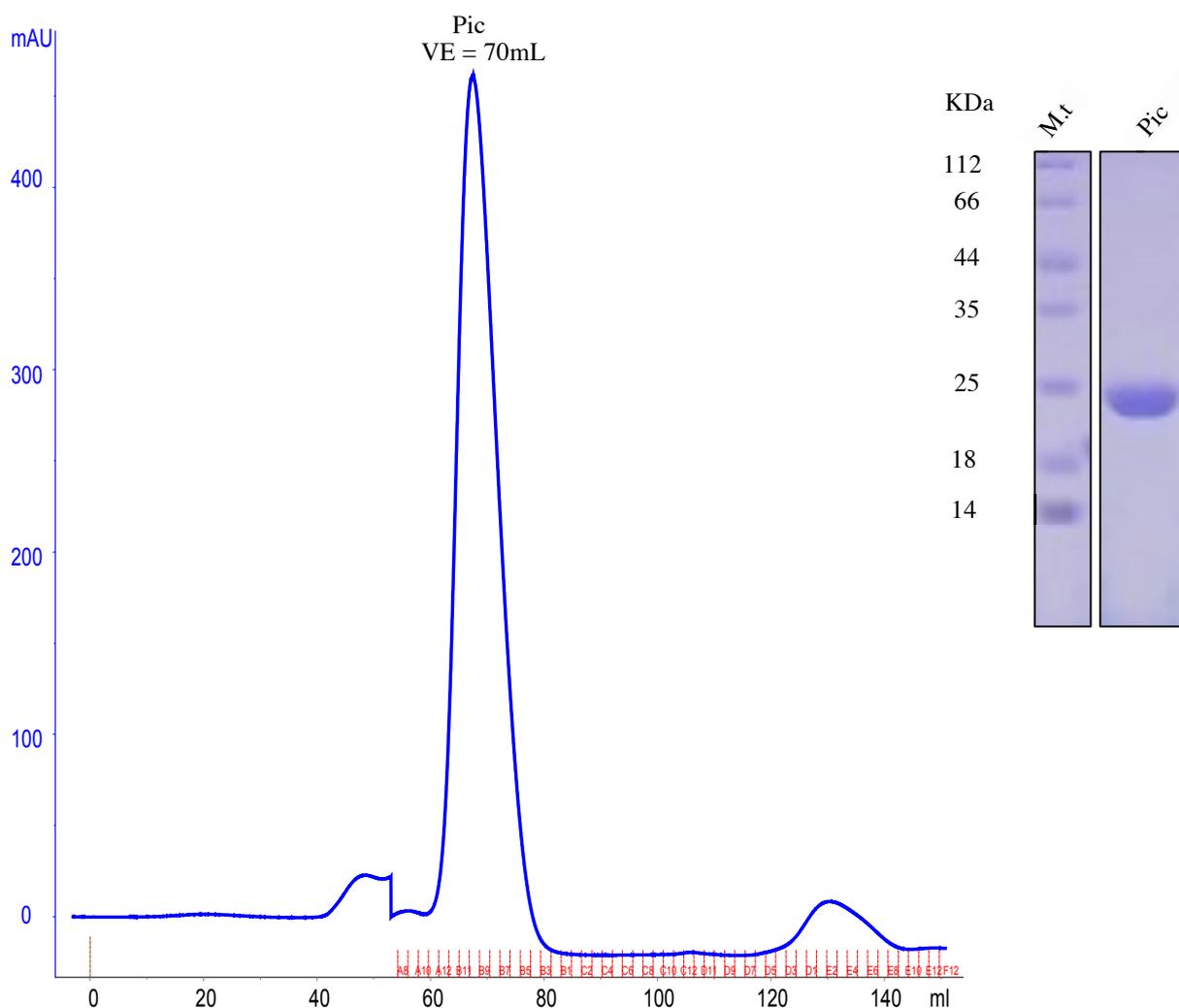


Figure 3.3.1: Purification de la protéine pPorM-Nt sur colonne Superdex 75 16/60. La protéine est éluée avec un volume d'élution de 70 mL de tampon Hepes 10 mM pH 7,3, NaCl 150 mM. La migration sur SDS-PAGE 15% d'une fraction du pic révèle une bande unique migrant en dessous de la bande 25 kDa du marqueur de taille (M.t), ce qui correspond à la taille attendue pour un monomère de pPorM-Nt (environ 22 kDa).

Des tests de stabilité à 4°C ont montrés que la protéine ne subit pas de dégradation jusqu'à 10 jours et est donc stable. Elle supporte également très bien la concentration puisqu'elle peut être concentrée jusqu'à 12 mg/mL sans aucune précipitation.

L'analyse du contenu en structures secondaires de pPorM-Nt par dichroïsme circulaire semble indiquer un repliement « tout alpha » (Figure 3.3.2). En effet, la courbe d'ellipticité moyenne par résidu en fonction de la longueur d'onde est proche de la courbe théorique correspondant à une protéine entièrement constituée d'hélices alpha : elle possède deux minima, autour de 207 nm et 222 nm. L'analyse par le serveur en ligne Dichroweb confirme cette tendance avec un pourcentage de 80% en hélices alpha.

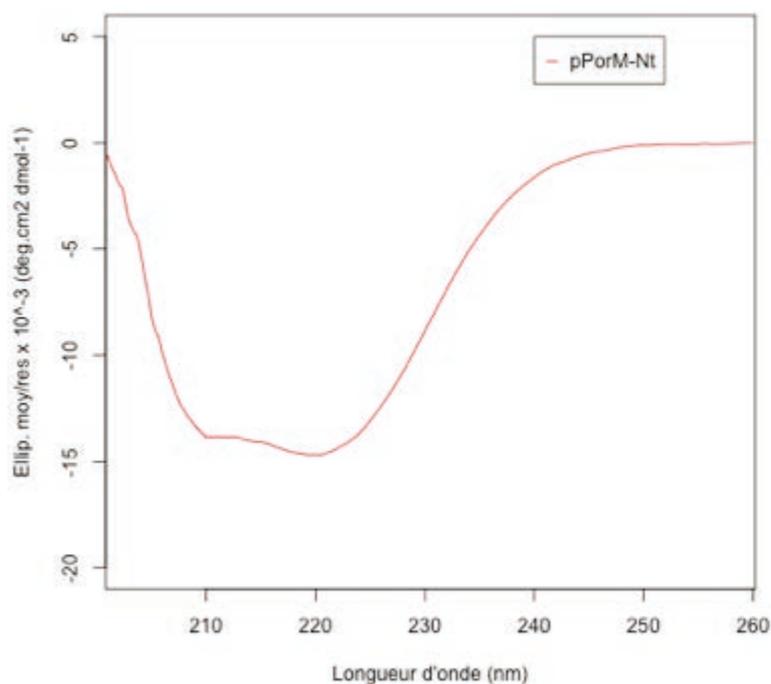


Figure 3.3.2: Courbe de dichroïsme circulaire de pPorM-Nt aux UV lointains. La mesure a été réalisée entre 190 et 260 nm mais seules les données comprises entre 203 et 260 nm sont exploitables. Deux minima sont visibles à 209 nm et 222 nm.

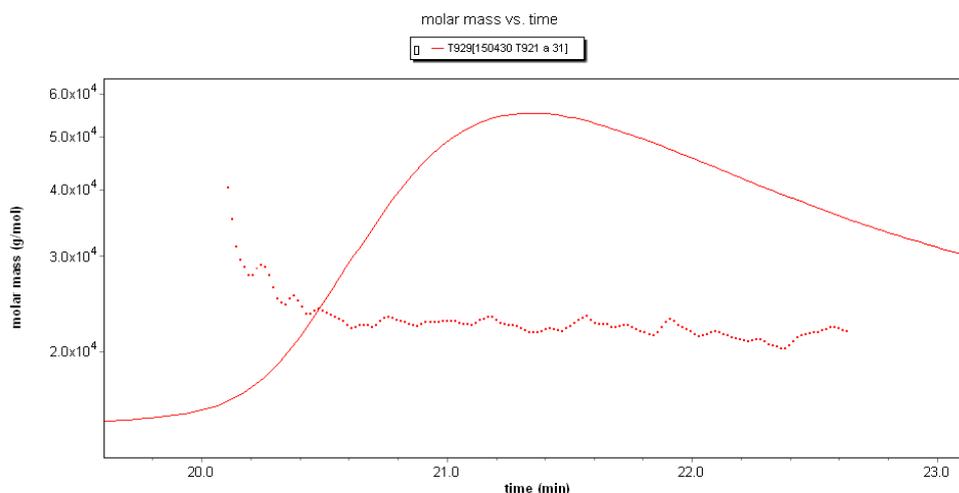


Figure 3.3.3: Analyse par SEC-MALS de la protéine pPorM-Nt. Colonne Shodex KW-802.5. Le pic de protéine a été élué à un volume de 21,25 mL et la masse moléculaire a été estimée à 22 kDa (pointillots rouges).

La protéine a également été analysée par SEC-MALS (Figure 3.3.3). Le poids moléculaire calculé, de 22 kDa, est très proche du poids moléculaire théorique, confirmant ainsi un état monomérique de la protéine.

Afin de déterminer la structure de la protéine pPorM-Nt, des études par cristallographie aux rayons X ont été initiées. Pour obtenir l'information de phase nécessaire à la résolution de la structure, deux méthodes peuvent être appliquées : l'utilisation des phases d'un modèle d'homologie de séquence supérieure à 30% (remplacement moléculaire) ou le calcul des phases de la structure à partir de métaux lourds (détermination de structure *ab-initio*). Aucune des structures actuellement disponibles dans la RCSB PDB ne peut servir de modèle à la protéine pPorM-Nt pour un remplacement moléculaire protéine. Deux stratégies ont alors été appliquées en parallèle : l'utilisation d'atomes de Sélénium incorporés aux Méthionines (SeMet) et la mise en complexe de pPorM-Nt avec un nanobody.

3.3.1.2 Protéine pPorM-Nt produite avec des Selenométhionines : expression et purification

L'incorporation d'atomes de Sélénium à la protéine pPorM-Nt a été réalisée au moment de l'expression bactérienne. La protéine surexprimée, nommée pPorM-Nt SeMet, a été purifiée du lysat bactérien selon le même protocole qu'utilisé pour la protéine native. Après injection du lysat clarifié bactérien sur une colonne de chromatographie d'affinité Ni²⁺ (GE Healthcare), pPorM-Nt SeMet est éluée avec un tampon contenant 250 mM d'Imidazole. Le pic récupéré est alors injecté sur une colonne de chromatographie d'exclusion de taille (Figure 3.3.4). Sur une colonne Superdex 75 26/60 (GE Healthcare), la protéine est éluée à un volume de tampon de 168 mL, ce qui correspond, d'après la calibration, au volume d'éluion attendu pour une protéine globulaire de 33 kDa. Le dépôt sur SDS-PAGE d'une fraction du pic élué montre une bande unique migrant à environ 23 kDa. Le rendement moyen de purification pour cette protéine est de 66,3 mg pour 2L de culture.

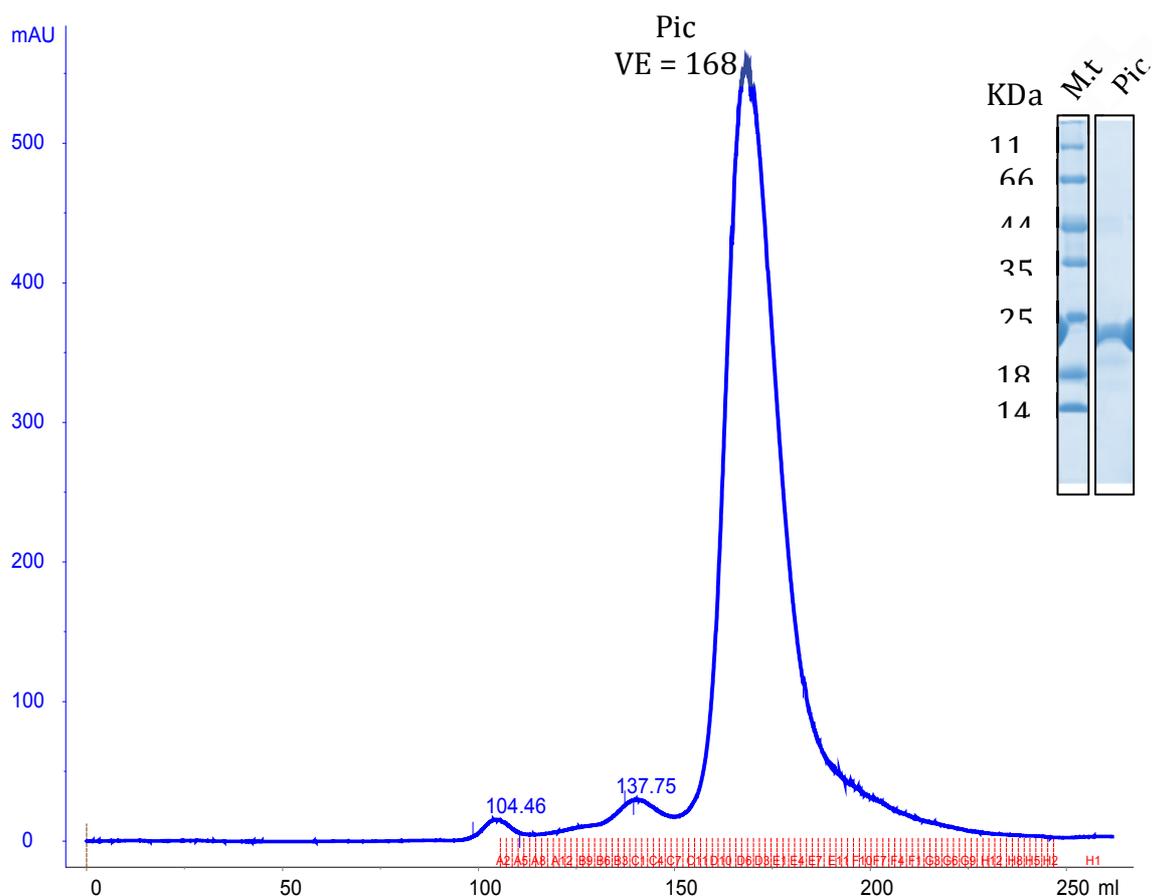


Figure 3.3.4: Purification de la protéine pPorM-Nt SeMet. Colonne Superdex 75 26/60. La protéine est éluée à un volume d'éluion VE de 168 mL. La migration sur SDS-PAGE d'une fraction du pic révèle une bande unique migrant sous la bande de 25 kDa du marqueur de taille (M.t).

3.3.1.3 Complexe pPorM-Nt/nanobody nb01 ou nb02 : mise en complexe et purification

Les nanobody nb01 et nb02 se fixant tout deux sur la protéine pPorM-Nt, les deux anticorps sont utilisés comme chaperonne de cristallisation. Pour les deux complexes, la mise en complexe est faite après l'étape de purification du nanobody sur colonne de chromatographie d'affinité Ni²⁺ (GE Healthcare). L'excès de nanobody est séparé du pic de complexe par purification sur chromatographie d'exclusion de taille (Figure 3.3.5).

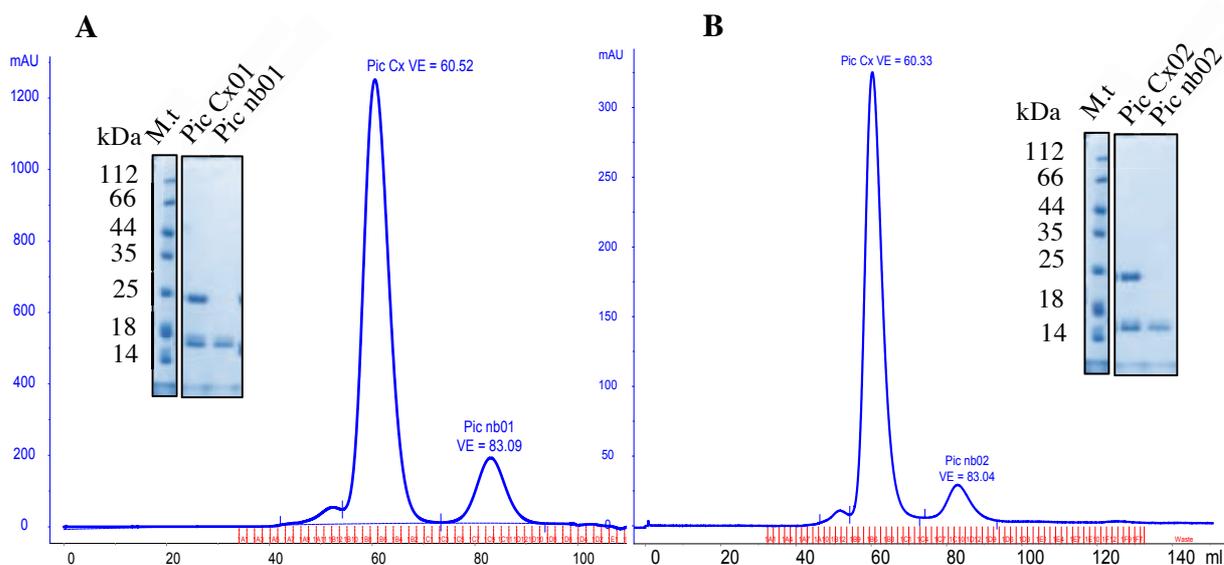


Figure 3.3.5: Purification des complexes pPorM-Nt/nanobody. Colonne Superdex 75 16/60. A : Complexe pPorM-Nt/nb01. Le complexe (Cx) est élué à un volume d'éluéon VE de 60,52 mL. L'excès de nb01 est élué à un volume de 83,09 mL. Deux bandes sont visibles lors de la migration sur SDS-PAGE pour le pic de complexe, l'une sous la bande à 25 kDa du marqueur de taille (M.t) correspondant à pPorM-Nt, l'autre entre les bandes à 18 et 14 kDa, correspondant à nb01. **B :** Complexe pPorM-Nt/nb02. Le complexe est élué à un volume d'éluéon de 60,33 mL et l'excès de nb02 à un volume d'éluéon de 83,04 mL. La même migration que celle du complexe avec nb01 est visible sur SDS-PAGE.

Pour chaque purification, les pics de complexe et d'excès de nanobody sont bien séparés. Les volumes d'éluéon sont similaires : ils sont, pour une purification sur colonne Superdex 75 16/60 (GE Healthcare), autour de 60 mL pour le pic de complexe et 83 mL pour le nanobody. Selon la calibration de cette colonne, un volume d'éluéon de 60 mL correspond à une protéine globulaire de 34 kDa ce qui est en accord avec le poids moléculaire théorique des complexes de 35 kDa environ. Le dépôt sur SDS-PAGE confirme la présence majoritaire du complexe dans le pic élué.

3.3.1.4 Etudes cristallographiques de la protéine pPorM-Nt

Les cristallogénèses de la protéine pPorM-Nt native et SeMet ont été menées en utilisant la diffusion de vapeur en gouttes assises.

Pour la protéine native, les screens généraux STURA (Molecular Dimension), MDL (Molecular Dimension), PEG I et PEG II (Qiagen) ont d'abord été testés. Des premiers cristaux de pPorM-Nt sont apparus après 8 jours dans différentes conditions : 200 mM calcium chloride avec 20% de PEG 3350, 200 mM calcium acétate avec 20% de PEG 3350 et 100 mM de Tris pH 8,5 200 mM calcium chloride avec 20% de PEG 4000. Ces conditions étant similaires, des optimisations ont été lancées en combinant les composés

redondants. Des cristaux sont apparus après une semaine en 200 mM calcium chloride avec 20% de PEG 4000, à pH 7,5 (Figure 3.3.6A).

La protéine pPorM-Nt SeMet a été cristallisée suite à l'obtention des cristaux de la protéine native. Les screens PEG I et PEG II ont donc été privilégiés, étant ceux ayant donné les cristaux. De petits cristaux se sont développés en 100 mM de Tris pH 8,5 200 mM calcium chloride avec 25% de PEG 4000 en une semaine (Figure 3.3.6B). Bien qu'une optimisation ait été faite en gouttes assises ainsi qu'en gouttes suspendues en boîtes LIMBRO, la taille des cristaux n'a pas augmentée.

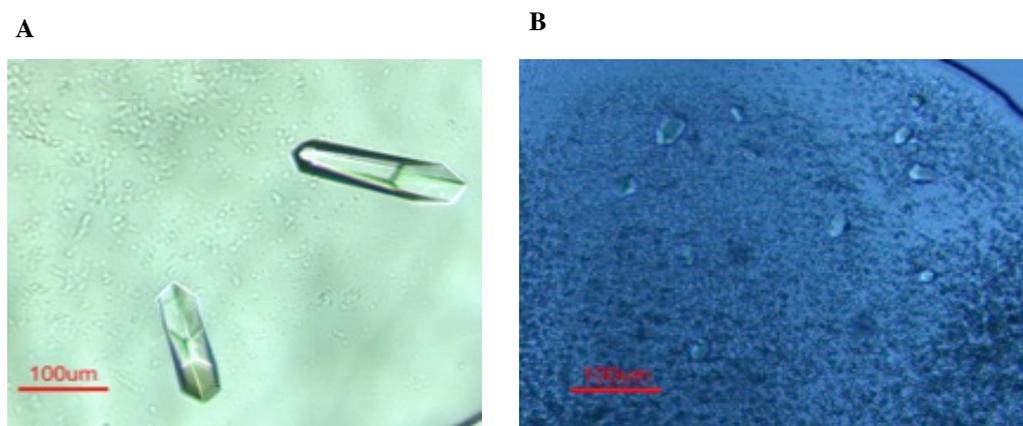


Figure 3.3.6: Cristaux de la protéine pPorM-Nt. A: pPorM-Nt native. B: pPorM-Nt SeMet.

Après avoir été pêchés et cryo-protégés, ces cristaux ont été congelés sous flux d'azote afin de tester leur diffraction à une source de rayonnement synchrotron.

Les cristaux de pPorM-Nt ont été envoyés à une mission synchrotron sur la ligne de lumière ID23-2 à l'ESRF, Grenoble. L'un des cristaux a diffracté jusqu'à une résolution de 1,9 Å et un jeu d'images a été collecté. Cependant, à l'issue du traitement des données, l'analyse menée par le programme SCALA (ccp4) a révélé une fraction de twin de 0,47, valeur se rapprochant de la fraction 0,5 d'un twinning parfait mérohédral. Le groupe d'espace trouvé, $P6_422$, n'est normalement pas sujet au twinning des données, sauf dans le cas de twin parfait du groupe de Laue où il peut s'agir en réalité d'un groupe d'espace $P312$, $P321$ ou $P6$. Le retraitement des données dans ces différents groupes d'espaces n'a pas permis de débloquent la situation. Les programmes spécialisés dans le contournement du twin, utilisant les lois de twin, tels que Detwin de la suite CCP4, phenix.refine avec l'option twin et CNS n'ont pas non plus aidé à finir l'affinement de la structure.

Les cristaux de pPorM-Nt SeMet ont été envoyés sur Proxima 1 à SOLEIL, Gif-sur-Yvette. La présence d'atomes de Sélénium a été confirmée à l'aide d'un scan d'énergie. Un des cristaux a diffracté jusqu'à une résolution maximale de 3,5 Å et trois jeux d'images ont

été collectés : un jeu à la longueur d'onde correspondant au pic d'absorption du Sélénium, un second à celle du point d'inflexion du pic, calculé à partir du spectre d'énergie, et un troisième à une longueur d'onde après le pic. Nous n'avons malheureusement pas réussi à trouver une solution distincte, au moment de la recherche des atomes de Sélénium, permettant le calcul de cartes de Patterson et donc à calculer des phases. La structure n'a donc pu être résolue.

Afin de contourner ces problèmes de twin, plusieurs stratégies ont été envisagées comme le changement de température au moment de la cristallogénèse, le micro-seeding ou la cristallisation en micro-capillaires (ChipX, Claude Sauter IBMC Strasbourg). Mais l'une des idées principales a été de se concentrer sur l'utilisation des nanobody comme chaperonnes de cristallisation.

3.3.1.5 Etudes cristallographiques des complexes pPorM-Nt/nb01 et pPorM-Nt/nb02

Les complexes pPorM-Nt/nanobody ont été cristallisés en utilisant la méthode de diffusion de vapeur en gouttes assises.

Deux conditions de cristallisation ont été définies pour le complexe pPorM-Nt/nb02. La première est constituée de 100 mM de MES pH 6,5 25% PEG 4000 et des aiguilles y sont apparues après 65 jours (Figure 3.3.7A, image de gauche). La seconde est composée de 100 mM d'imidazole malate à pH 6 avec 20% de PEG 4000 et des objets cristallins s'y sont formés après 45 jours (Figure 3.3.7A, image de droite). Les cristaux de ces deux conditions n'ont pu être reproduits.

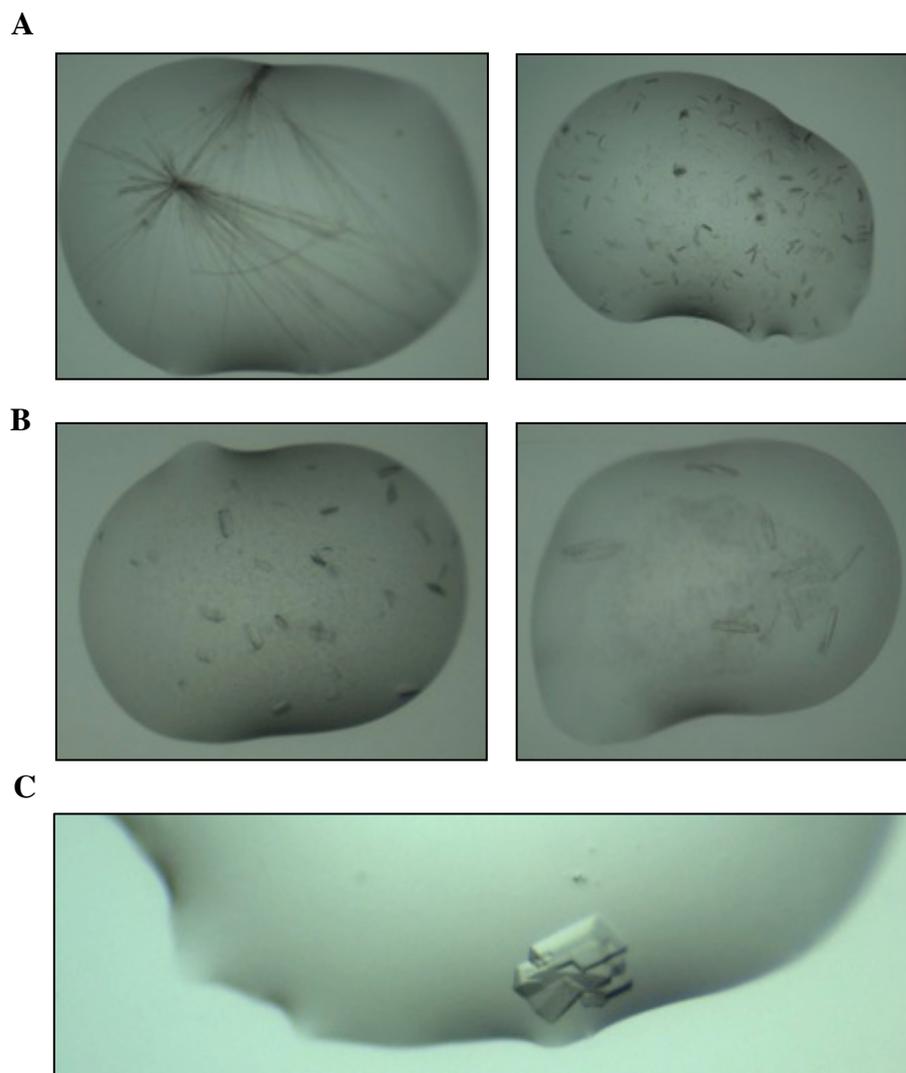


Figure 3.3.7: Cristallisation des complexes pPorM-Nt/nanobody. A: Complexe pPorM-Nt/nb02. B: Complexe pPorM-Nt/nb01 à 12 mg/mL. C: Complexe pPorM-Nt/nb01 à 16 mg/mL.

De nombreuses conditions de cristallisation ont été trouvées pour le complexe pPorM-Nt/nb01. Parmi elles se trouvent la condition 100 mM CHES à pH 9,5, 30% PEG 3000. Les cristaux, apparus après un jour, ont été testés sur la ligne ID23-2 de l'ESRF (Figure 3.3.7B, image de gauche). Leur diffraction aux rayons X n'est cependant pas allée au delà de 6 Å. Il en a été de même pour les cristaux obtenus dans la condition 200 mM de NaCl, 100 mM de Tris pH 8,5 et 25% PEG 3350, qui ont diffracté à une résolution maximale de 5 Å (Figure 3.3.7B, image de droite). Toutes ces conditions ont été trouvées à une concentration en complexe de 12 mg/mL mais une autre condition de cristallisation a pu être définie à 16 mg/mL. Après deux jours, des cristaux superposés sont apparus en 100 mM Bis-Tris pH 7 25% PEG 3350 (Figure 3.3.7C). Au moment de les pêcher, cet amas s'est séparé en trois parties qui ont toutes les trois été congelées et envoyées sur la ligne de

lumière Proxima 2. La partie centrale de cet amas a diffracté à une résolution maximale de 2,3 Å. Un jeu de données a été enregistré puis traité au laboratoire.

Le nanobody nb01 dont la structure avait été résolue au préalable (numéro d'accèsion PDB : 5LZ0) a été positionné par remplacement moléculaire. Ce phasage initial a permis d'obtenir les premières cartes de densité électronique. Plusieurs cycles de construction et affinement du modèle dans les cartes résiduelles ont permis d'obtenir la structure d'une grande partie de pPorM-Nt, allant du résidu 70 à 216 (Figure 3.3.8C). La structure finale du complexe est présentée dans la Figure 3.3.8A. Les statistiques de la collecte des données et de l'affinement sont présentées dans le Tableau 3.3.1. Les structures de nb01 libre ou complexé à pPorM-Nt sont quasiment identiques (Figure 3.3.8B). Seuls quelques changements minimes sont visibles au niveau des boucles des CDR. Le repliement général de pPorM-Nt peut être décrit comme un ensemble de quatre hélices antiparallèles avec une topologie haut-bas haut-bas avec un pas à droite. La comparaison de cette structure avec la base de données PDB en utilisant le serveur DALI a révélé que le domaine pPorM-Nt montre une forte similarité structurale avec une kinase d'adhésion focale (PDB 3B71, (57)) avec un Z-score de 10,3 et un rmsd de 2,7 Å (Figure 3.3.9A).

Les 25 premiers résidus n'ont pas été construits, dû au manque de densité électronique dans la carte. Le domaine pPorM-Nt doit pouvoir relier la partie transmembranaire de PorM au domaine pPorM-Ct. Puisque les résidus 70 et 217 se trouvent du même côté, il est vraisemblable que les résidus manquants pointent du côté opposé. Ce changement de direction peut être induit par la proline PRO71, introduisant ainsi un coude.

Cet assemblage d'hélices présente une « cassure » de la troisième hélice, des résidus 125 à 136. Il s'agit là d'une particularité qui n'est pas présente sur la kinase d'adhésion focale (Figure 3.3.9A). La coloration selon les B-facteurs montre d'ailleurs qu'il s'agit d'une boucle plus mobile.

Le nanobody nb01 interagit avec pPorM-Nt par l'intermédiaire de son CDR3 (résidus 97 à 100B, Figure 3.3.9B). Ces résidus forment des liaisons avec les résidus THR93, PHE94, ASP97, ALA101, ARG104, LYS136 et GLU143 de PorM.

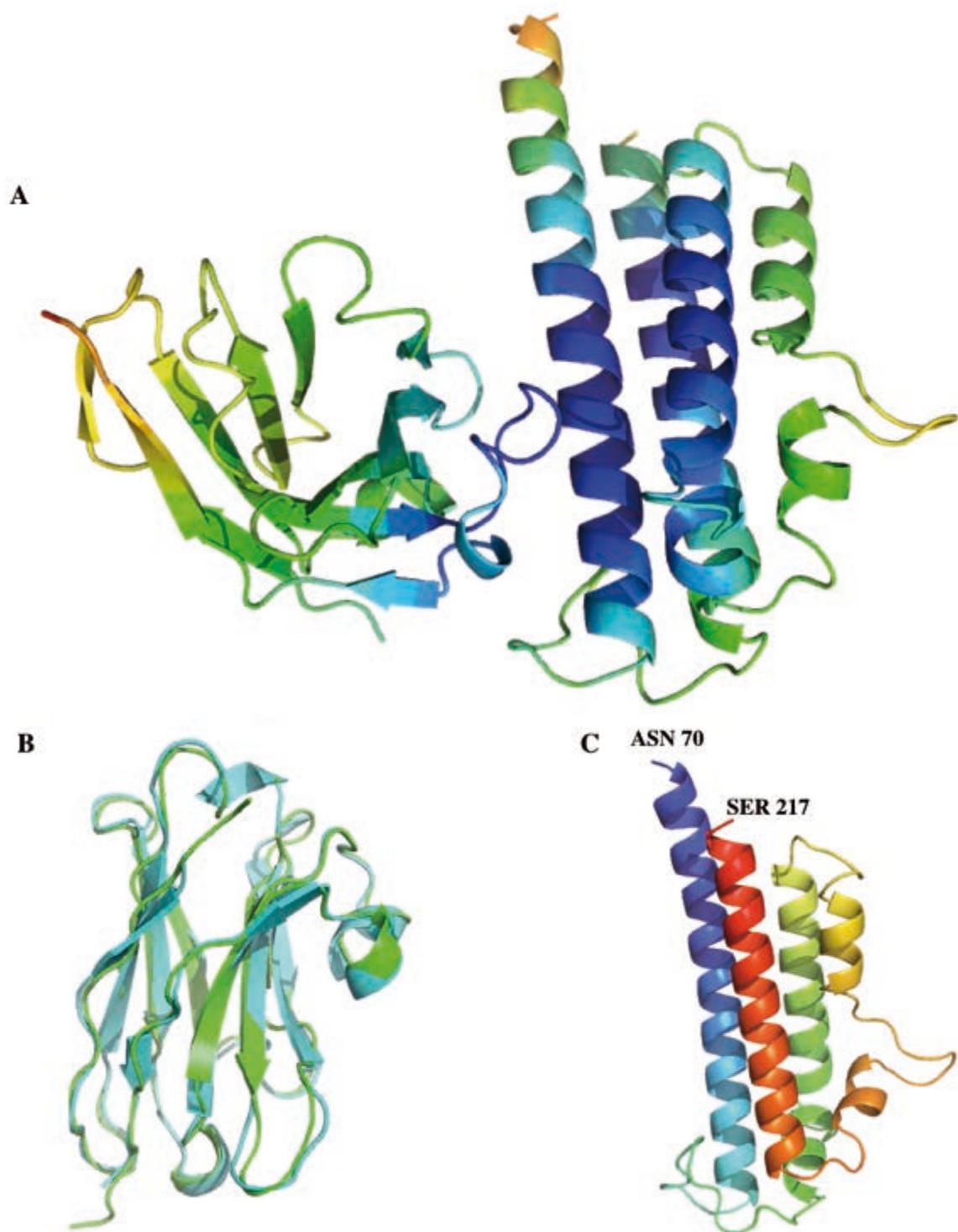
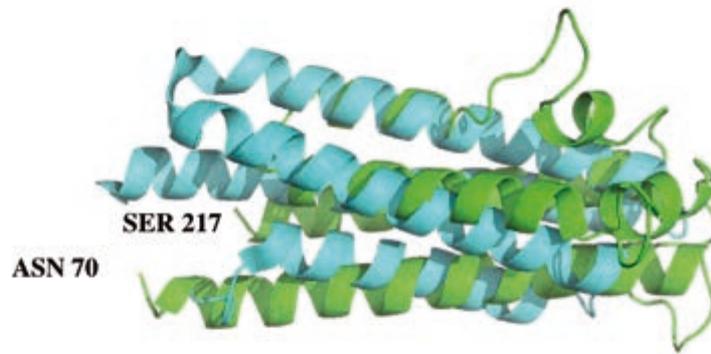


Figure 3.3.8: Structure cristallographique de pPorM-Nt en complexe avec nb01. A : Structure du complexe. Les deux partenaires sont colorés selon leurs b-facteurs. B : Superposition de la structure du nanobody nb01 sous sa forme complexée avec la structure de la forme libre (5LZ0). La première est colorée en vert et la seconde en bleu. C : Présentation de la structure résolue de pPorM-Nt (résidus 70 à 217). Une coloration « Rainbow » est utilisée, dans un dégradé allant du bleu foncé (N-ter) au rouge (C-ter). Elle est constituée d'un ensemble de 4 hélices en formation antiparallèle haut-bas haut-bas. Les résidus 44 à 69 n'ont pu être construits.

A

3B71
Z=10,3
rmsd=2,7



B

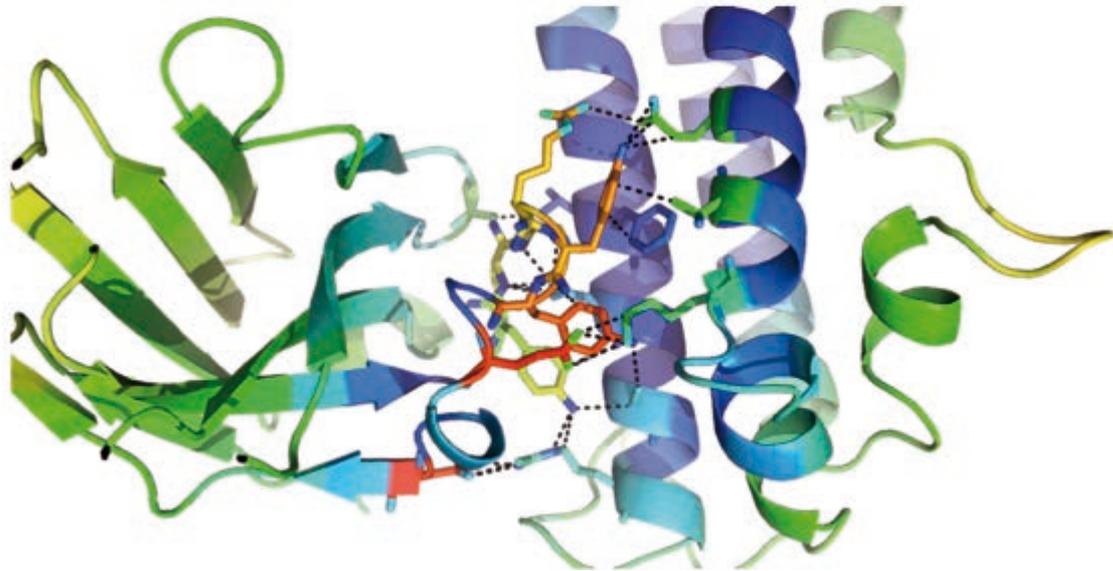


Figure 3.3.9: Analyse structurale du complexe pPorM-Nt/nb01. A : Alignement structural (DALI). Superposition de la structure de pPorM-Nt (vert) avec son plus proche voisin structurale (3B71, bleu). B : Région d'interaction de nb01 avec pPorM-Nt. Les résidus impliqués sont représentés en bâtons et leurs interactions sont affichées en pointillés noirs. Les deux partenaires sont colorés selon leurs b-facteurs.

	Nb02 température ambiante	pPorM-Nt/nb01
<i>Condition de cristallisation</i>	0.1 M trisodium citrate pH 5.5, 20% PEG 3000	0.1 M bis-tris pH 7.0, 25%(w/v) PEG 3350
<i>Collecte des données</i>		10% glycérol
Cryo-protectant	-	SOLEIL
Synchrotron	SOLEIL	SOLEIL
Ligne de lumière	PROXIMA-2A	PROXIMA-1
Groupe d'espace	P4 ₃ 2 ₁ 2	P2 ₁
<i>a, b, c</i> (Å)	66,68 66,68 91,82	80,35 100,12 80,41
α, β, γ (°)	90 90 90	90 93,82 90
Résolution (Å) ^a	31,34 - 1,9 (1,968 - 1,9)	42,47 - 2,4 (2,53 - 2,4)
Longueur d'onde (Å)	0,826561	0,9786
Réflexions uniques ^a	16502 (1622)	49059 (7020)
Multiplicité ^a	3,7 (3,9)	6,9 (6,8)
Complétude (%) ^a	92,41 (80,87)	98,30 (96,7)
<i>I</i> / <i>s</i> ^a	8,61 (0,42)	11,60 (1,0)
<i>R</i> _{meas} (%) ^{a,b}	8,57 (247,9)	7,8 (203,3)
<i>R</i> _{p.i.m.} (%) ^{a,c}	4,255 (123,4)	3,0 (77,4)
CC1/2 ^d	0,997 (0,27)	0,999 (0,755)
Mosaïcité	0,106/0,029	0,05
Contenu en solvant (%)	67,2	43,4
<i>Molecular replacement</i>		
Modèle de départ	5LMW	5LZ0
<i>Qualité de l'affinement et du modèle</i>		
Résolution (Å)	31,338-1,9	19,91-2,5
Réflexions	15648 (1340)	42814 (4179)
<i>R</i> _{cryst} / <i>R</i> _{free} (%) ^e	19,83/23,07	21,89/26,19
Nombre d'atomes		
Protéine (chaîne(s) dans l'UA)	1002 (1)	8373 (8)
Eau	62	88
B-facteurs moyens (Å ²)		
Protéine	30,37	121,26
Eau/ion	41,32	97,47
Rmsd ^f		
Liaison (Å)	0,008	0,013
Angle (°)	0,90	1,71
Ramachandran plot (%)		
Régions les plus favorables	95,80	94,91
Régions additionnelles autorisées	4,20	4,51

Tableau 3.3.1 : Statistiques de collecte et d'affinement des structures du nanobody nb02 à température ambiante et du complexe pPorM-Nt/nb01. Les valeurs entre parenthèses sont celles de la coquille de plus haute résolution. Pour nb02 à température ambiante, les deux valeurs de mosaïcité correspondent aux deux jeux de données initiaux.

3.3.2 Etudes SAXS de pPorM

La structure cristallographique de pPorM-Ct a été résolue dans l'équipe par Philippe Léone, également à l'aide de nanobody. Seulement, cette structure est incomplète : pPorM-Ct est composée de trois domaines D1, D2 et D3, respectivement du côté N- au C-terminal, et la structure résolue ne comporte que les domaines D2 et D3. La structure de la protéine pPorM-T, équivalente à pPorM-Ct, a également été résolue et comporte les trois domaines.

Ces données, associées à la structure de pPorM-Nt, permettent d'avoir une idée de l'organisation structurale de pPorM mais restent incomplètes.

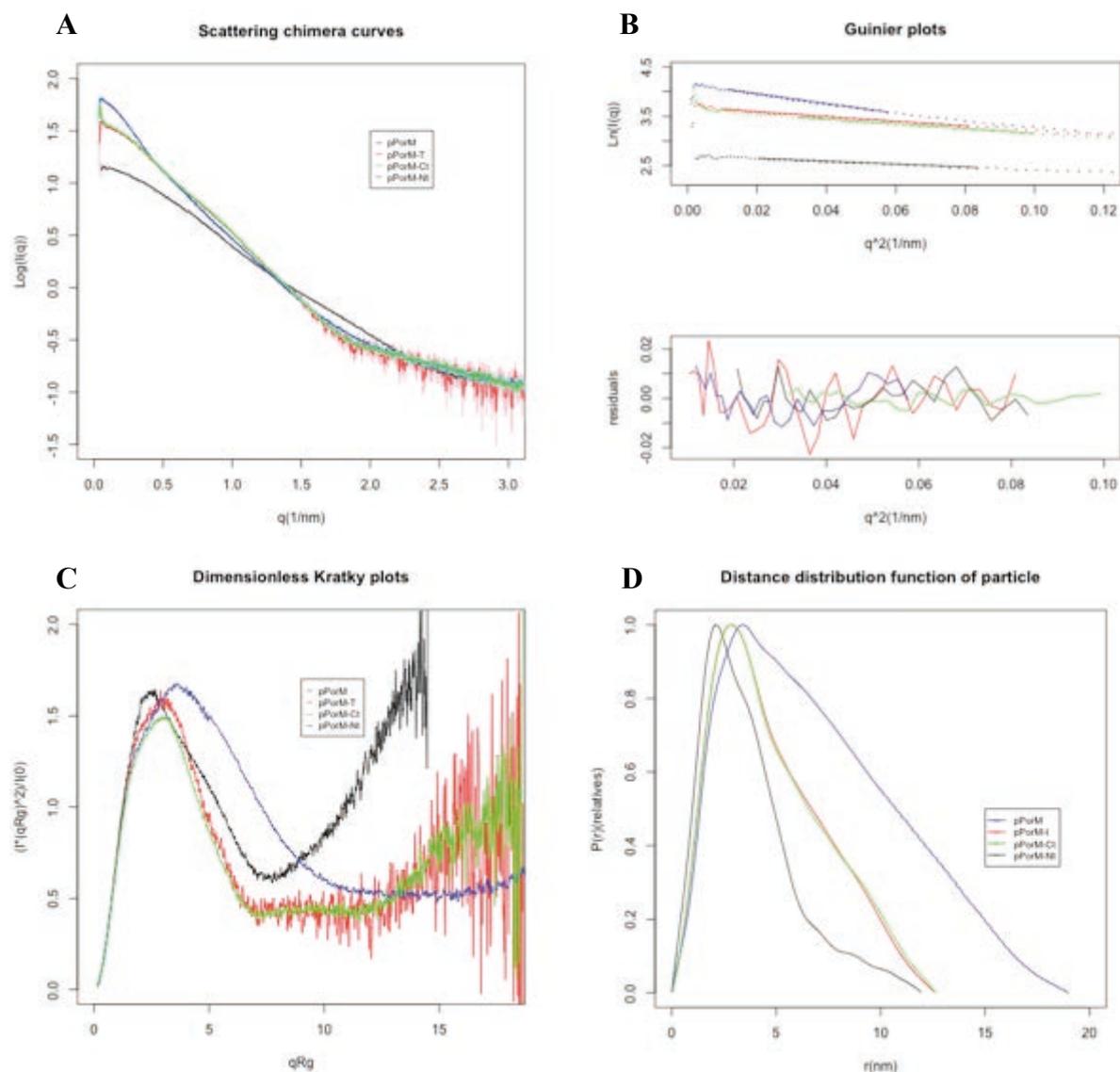


Figure 3.3.10: Analyses SAXS des constructions de pPorM en utilisant la suite ATSAS. A : Courbes de diffusion chimériques. B : Analyses de Guinier. C : Plots de Kratky sans dimensions. D : Fonctions de distribution des distances des particules. Les données de pPorM, pPorM-T, pPorM-Ct et pPorM-Nt sont respectivement colorées en bleu, rouge, vert et noir.

Afin de combler ces lacunes sur l'étude structurale cristallographique de pPorM, des analyses en solution ont été réalisées par SAXS. Des données ont été collectées sur la ligne BioSAXS BM29 du synchrotron ESRF sur les protéines pPorM, pPorM-T, pPorM-Nt et pPorM-Ct.

Pour toutes ces protéines, des données de bonne qualité ont pu être collectées lors de ces expériences SAXS. Les protéines n'ont pas subi de dommages aux radiations, comme le montrent les courbes de diffusion (Figure 3.3.10A).

Une droite a pu être superposée sur la région de Guinier (les 100 premiers points enregistrés environ) de ces courbes de diffusion (Figure 3.3.10B), ce qui indique l'absence d'agrégation des échantillons. Les R_g et $I(0)$ estimés sont reportés dans le Tableau 3.3.2. Les poids moléculaires ont été estimés en utilisant le serveur en ligne SAXS MoW2 (Tableau 3.3.2).

	pPorM	pPorM-T	pPorM-Ct	pPorM-Nt
<i>Data collection parameters</i>				
Instrument	BM29 (ESRF)	BM29 (ESRF)	BM29 (ESRF)	BM29 (ESRF)
Detector	Pilatus (1M)	Pilatus (1M)	Pilatus (1M)	Pilatus (1M)
Beam geometry	4*4	4*4	4*4	4*4
Wavelength (nm)	0.09919	0.09919	0.09919	0.09919
q range (nm ⁻¹)	0.036-4.912	0.036-4.912	0.036-4.912	0.036-4.912
Exposure time (s)	1	1	1	1
Concentration range (g/L)	0.73-10	0.43-4.40		0.52-16.11
Temperature (K)	283	283	283	283
<i>Structural parameters</i>				
$I(0)$ (cm ⁻¹) (from P(r))	63.97	39.68	37.74	15.03
$I(0)$ (cm ⁻¹) (from Guinier)	64.04 +/- 0.33	38.10 +/- 0.12	37.30 +/- 0.05	14.69 +/- 0.03
R_g (Å) (from P(r))	56.6	38.4	38.8	30.2
R_g (Å) (from Guinier)	55.3 +/- 0.05	37.9 +/- 2.4	37.7 +/- 2.8	29.5 +/- 2.3
D_{max} (Å) (from P(r))	190	126	126	119
D_{max} (Å) (from crystal structure)		119	88	59
<i>Molecular mass (M) (kDa)</i>				
M theoretical	55.357	34.328	34.328	21.946
M from SAXS MoW2 ³	147.201	72.259	72.760	20.779
<i>Software employed</i>				
Primary data reduction	PRIMUS	PRIMUS	PRIMUS	PRIMUS
Data processing	GNOM	GNOM	GNOM	GNOM
Validation of structural models		CRYSQL SREFLEX	CRYSQL	CRYSQL DAMMIF

Tableau 3.3.2: Collecte des données SAXS et paramètres structuraux dérivés.

Selon ces estimations, la protéine pPorM-Nt semble être monomérique alors que les protéines pPorM-T et pPorM-Ct semblent plutôt être dimériques. Ces résultats sont en accord avec les structures obtenues par cristallographie. Concernant la protéine pPorM, un état oligomérique de dimère étendu est possible.

Les plots de Kratky sans dimensions (Figure 3.3.10C) renseignent sur la flexibilité et la forme globulaire d'une molécule. Les courbes des protéines pPorM, pPorM-T et pPorM-Ct sont similaires à la courbe de Kratky correspondant à une protéine partiellement flexible, avec un pic d'une valeur maximale à $qR_g=3,5$ suivit d'un plateau à une valeur de $I^*(qR_g)^2/I(0)$ autour de 0,5 . Pour la protéine pPorM-Nt, la diminution lente du pic sans décrire de plateau par la suite peut être le signe d'un degré de flexibilité plus important que pour les autres protéines.

Les fonctions de distribution des distances (Figure 3.3.10D) ont été obtenues en utilisant le programme GNOM de la suite ATSAS. Les estimations des R_g et $I(0)$ sont similaires à celles obtenues par analyses de Guinier (Tableau 3.3.2). La distance D_{\max} calculée à partir de la $P(r)$ pour la protéine pPorM-T est en accord avec celle mesurable sur la structure cristallographique (valeur approximative ne tenant pas compte de la couche d'hydratation, Tableau 3.3.2). Ce n'est pas le cas en revanche pour les protéines pPorM-Ct et pPorM-Nt. Dans le cas de pPorM-Ct, la différence entre la valeur obtenue en solution avec un échantillon fraîchement préparé et celle mesurée avec la structure cristallographique peut s'expliquer par l'absence du domaine D1, probablement clivé avant la formation du cristal. Dans le cas de pPorM-Nt, les 25 premiers résidus non construits sont probablement très flexibles, ce qui pourrait expliquer la valeur importante de la D_{\max} observée en solution. L'allure générale de ces courbes donne l'information de forme de ces protéines. Alors que celles de pPorM-T et pPorM-Ct sont similaires et semblent correspondre à la courbe d'une protéine modulaire, la courbe de pPorM-Nt a une longue extension sur la fin responsable de la D_{\max} élevée. Cette allure peut être due à une partie étendue, probablement désordonnée. La courbe correspondant à la protéine pPorM ressemble à celle d'une protéine allongée.

La comparaison de ces données avec la courbe théorique de diffusion de la structure cristallographique de pPorM-T, à partir de CRY SOL, révèle un accord important avec une valeur du test de χ^2 de 0,97 (Figure 3.3.11A). Ce résultat signifie que la structure obtenue par cristallographie aux rayons X est en accord avec les conformères trouvés en solution et analysés par SAXS. J'ai ensuite cherché à savoir si la structure cristallographique de pPorM-T est le conformère correspondant le mieux à la protéine pPorM-T en solution. Une simulation utilisant une analyse par modes normaux (SREFLEX) a généré 10 structures affinées mais aucune n'ayant une meilleure valeur du test de χ^2 , même si elles restent très bonnes (Figure 3.3.11B et C). Les principales différences consistent en de faibles degrés de flexibilité au niveau de chaque domaine et il n'y a pas de grande variation de l'angle entre les domaines D1 et D2.

Comme dit précédemment, le domaine D1 est manquant dans la structure cristallographique de pPorM-Ct, ce qui peut facilement expliquer le désaccord entre ces données et les données SAXS (Figure 3.3.12A). Cependant, la structure cristallographique de pPorM-T a un meilleur accord avec ces données SAXS, bien que la valeur du test de χ^2 reste mauvaise (Figure 3.3.12A).

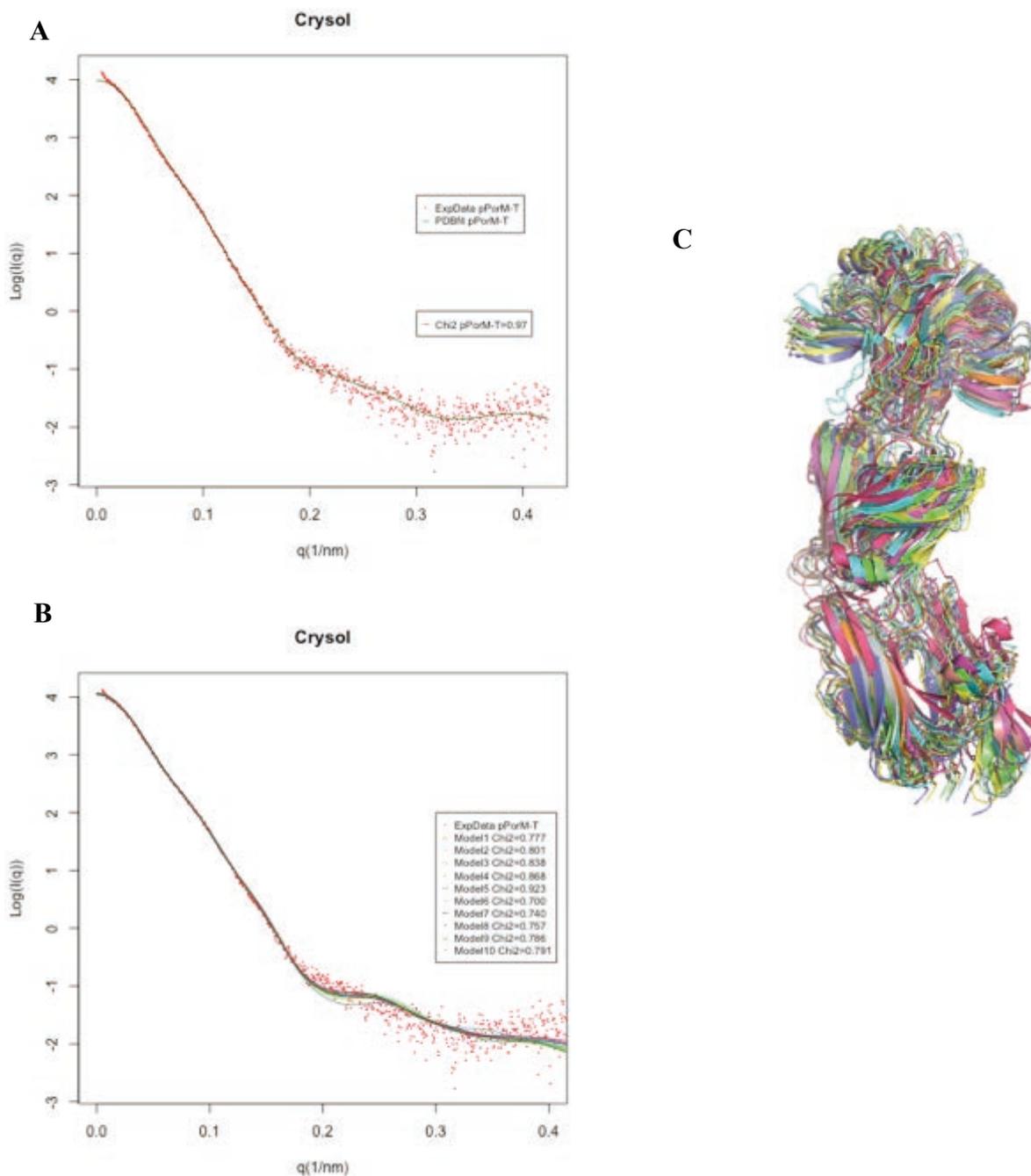


Figure 3.3.11: Analyses SAXS de pPorM-T. A : Analyse CRY SOL. La valeur du test de χ^2 est de 0,97. **B et C :** Modélisation avec SREFLEX de 10 modèles structuraux selon les degrés de flexibilité intrinsèques. Les courbes de diffusions (B) ainsi que les modèles correspondants (C) sont illustrés.

Pendant l’affinement de pPorM-Nt, de faibles densités étaient visibles pour la partie N-terminale manquante de la protéine, ne permettant cependant pas de compléter la structure. La structure cristallographique est alors incomplète et une analyse selon CRY SOL n’aurait aucun sens. Pour avoir tout de même une idée de la forme globale de pPorM-Nt, le logiciel DAMMIF a été utilisé depuis le Webservice ATSAS et le meilleur modèle a été

représenté en Figure 3.3.12B. La très bonne valeur du test de χ^2 permet de considérer cette forme comme forme potentielle du conformère en solution.

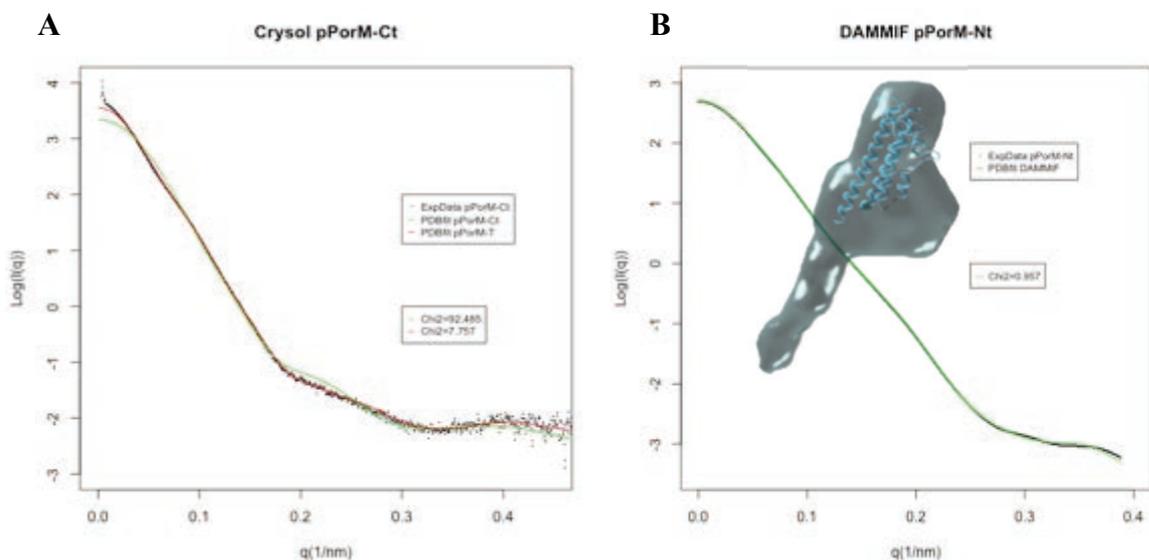


Figure 3.3.12: Analyse SAXS de pPorM-Ct et pPorM-Nt. A : Analyses CRYSOLOG de pPorM-Ct. Les données SAXS ont été comparées aux courbes théoriques de diffusion des structures cristallographiques de pPorM-Ct et pPorM-T. **B :** Modélisation *ab-initio* d'une enveloppe pour pPorM-Nt. La courbe du modèle généré est illustrée ainsi que l'enveloppe, dans laquelle a été insérée la structure actuellement connue de pPorM-Nt.

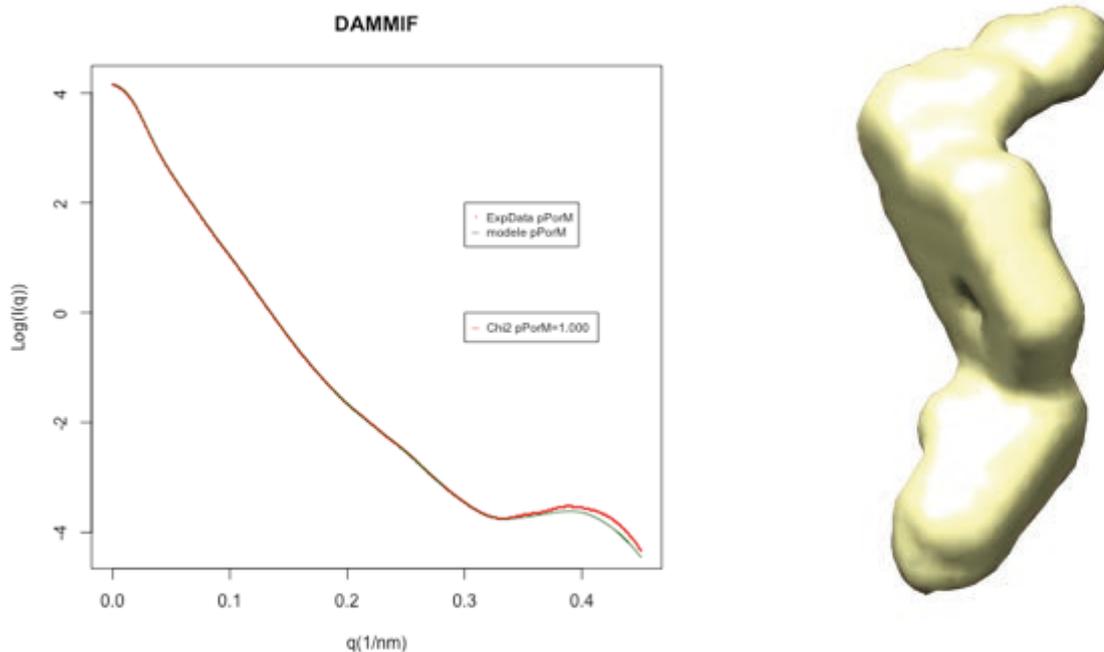


Figure 3.3.13: Modélisation *ab-initio* d'une enveloppe de pPorM à l'aide de DAMMIF. 20 modèles ont été générés mais est représenté ici uniquement celui ayant la meilleure valeur de test de χ^2 . La courbe de diffusion théorique du modèle (vert) est superposée aux données expérimentales (rouge). L'enveloppe du modèle est représentée à côté en mode surfaces.

Il n'est pas réellement possible de comparer les données expérimentales SAXS de pPorM avec les structures cristallographiques obtenues pour pPorM-T, pPorM-Ct ou pPorM-Nt. L'organisation entre ces différents domaines n'est pas connue et bien que l'état dimérique de pPorM soit très probable, les domaines entremêlés de la partie C-terminale (pPorM-T) ne permettent pas d'appliquer une symétrie P2 lors des analyses. La seule analyse qui puisse aider à la compréhension de cette organisation est la génération d'une enveloppe SAXS par modélisation *ab-initio* à l'aide de DAMMIF. Les valeurs du test de χ^2 des 20 modèles générés ont été comparées et le modèle ayant une valeur $\chi^2=1,000$ est illustré en Figure 3.3.13. L'enveloppe de ce modèle a une forme allongée présentant deux angles aux extrémités. Bien que les détails diffèrent selon les modèles, cette allure générale est conservée.

3.3.3 Article 2 (manuscrit en préparation)

X-ray Structure of the periplasmic domain of PorM, a component of the T9SS in *Porphyromonas gingivalis*

Jennifer Roche^{1,2,&}, Philippe Leone^{1,2,&}, Maxence S. Vincent³, Yoan Duhoo^{1,2}, Quang Hieu Tran^{1,2}, Christine Kellenberger^{1,2}, Christian Cambillau^{1,2}, Eric Cascales³ and Alain Roussel^{1,2,*}.

¹ Centre National de la Recherche Scientifique, Architecture et Fonction des Macromolécules Biologiques, UMR 7257, Marseille, France

² Aix-Marseille Université, Architecture et Fonction des Macromolécules Biologiques, UMR 7257, Marseille, France.

³ Laboratoire d'Ingénierie des Systèmes Macromoléculaires (LISM), UMR 7255 CNRS et Aix-Marseille Université, Marseille, France

& These authors have contributed equally to the work

* To whom correspondence should be addressed. Email: alain.roussel@afmb.univ-mrs.fr

Running title: Crystal structure of pPorM

Introduction

Bacteria have evolved specialized macromolecular machines named secretion systems that secrete a large range of substrates, including small molecules, DNA and proteins. In diderm bacteria, secretion may be either a single-step process in which substrates are translocated directly from the cytoplasm to the external milieu or a two-step process in which the substrates first cross the inner membrane (IM) into the periplasm using the Sec, Tat or holins pathway, then cross the outer membrane (OM) through a specialized translocon (1). After secretion the substrates may stay attached to the surface of the OM, be released into the extracellular milieu or be injected into a target cell (1). The recently discovered type 9 secretion system (T9SS) belongs to the two-step process systems group. The T9SS, which is exclusively present in the Bacteroidetes phylum (2-4), has been studied mainly in *Flavobacterium johnsoniae* and *Porphyromonas gingivalis*. In *F. johnsoniae* it contributes to the gliding motility by secreting SprB, a cell surface adhesin that is required for movement on agar (5). SprB localizes to the cell surface, where it forms long slender filaments (6). Unlike *F. johnsoniae*, *P. gingivalis* is a non-motile bacterium. It is a human oral pathogen that is a major causative agent of periodontitis (7). In *P. gingivalis* the T9SS contributes to the production of potent proteolytic enzymes called gingipains (8-10) that are essential virulence factors responsible for corrupting host innate defense mechanisms (7,11). Gingipains are secreted in large amounts and are mainly attached to the surface of the OM, but are also partially released in a soluble form into the extracellular milieu (8,12).

Presently, 18 genes have been proven essential for proper T9SS function in *P. gingivalis* by deletion mutagenesis studies (13). Among them a group of five genes, PorP-PorK-PorL-PorM-PorN, form an operon structure and are co-transcribed (14). In addition it has been shown that PorK, PorL, PorM and PorN, assemble a > 1.4 MDa complex (2). PorK is a lipoprotein anchored in the outer membrane and interacts strongly with the periplasmic protein PorN. PorK and PorN form large disks of around 50 nm in diameter in the OM (15). PorL and PorM are inner membrane proteins which interact with each other via their transmembrane segments. In addition to its transmembrane segment, PorM possesses a

long periplasmic domain of nearly 500 residues which interacts with all other operon partners, and which is thought to link the inner and outer membrane subcomplexes (14). Therefore, we can speculate that PorM is a central scaffolding component of the T9SS core complex.

Now, to the best of our knowledge, there is no atomic structure of any component of the PorKLMN complex. In the present study, we report the three-dimensional structure of PorM based on the X-ray crystallographic structures of two fragments of PorM representing the periplasmic part of the protein. The structure analysis revealed that PorM folds into four domains: one monomeric four-helix bundle domain in the N-terminal part, while the C-terminal part is composed of three domains, whose two are domain-swapped immunoglobulin-like domains. The crystal structures of the PorM constructs allowed modelling of the full-length periplasmic domain, highlighting an extended and bended conformation that is confirmed by SAXS analysis.

Results

Structure determination of the periplasmic domain of PorM

The polypeptide chain of PorM contains a short intracellular segment (residues 1–12), a predicted transmembrane domain (residues 13–35) and a larger periplasmic part (residues 36–516; pPorM). In a previous study, limited proteolysis made it possible to identify and produce a C-terminal domain of PorM (residues 225–516; pPorM-T (16)). Native crystals of pPorM-T diffracted up to 2.85Å, and phases were obtained with a MAD data set from Seleno-Methionine derivative crystals (16). The pPorM-T domain is composed of three subdomains encompassing residues 225-314 (D1), 318-398 (D2) and 400-516 (D3), respectively. However, complete tracing of the domain backbone was not possible as the electron density was very poor for the D3 subdomain. In order to improve the resolution, we cloned and produced a construct corresponding to the region of D1+D2+D3 (residues 225-516; pPorM-Cter). Since this new construct did not yield crystals, we generated llama antibodies against pPorM aiming at co-crystallization with pPorM-Cter. Four nanobodies (nb01, nb02, nb19 and nb130) were selected and produced and their structures were solved by molecular replacement (17). Two of them, namely nb19 and nb130, were shown to bind to pPorM-Cter and were used as co-crystallization chaperones. Diffracting crystals were obtained for nb130 in complex with pPorM-Cter and the structure of this complex was solved by molecular replacement using the nanobody structure and the partial models of D2 and D3. The complete chain of D2 and D3 could be traced and the structure was refined up to 2.1 Å resolution. The structure of the complex revealed that nb130 binds to the D3 subdomain and stabilizes it, in particular through the protruding CDR3 and a deep interaction of Tyr99. Contemporaneously, we produced a construct corresponding to the N-terminal part of the PorM periplasmic domain (residue 44-217, pPorM-Nter). Diffracting crystals of pPorM-Nter were obtained and several data sets (native and Seleno-Methionine derivative) were collected up to 1.9 Å resolution. Unfortunately, these datasets were twinned and therefore not useable for further crystallographic work. This led us to use again a nanobody as a way to obtain different crystals, hopefully non-twinned. This strategy revealed to be a success as diffracting crystals of pPorM-Nter with nanobody nb01 were obtained. The structure of the complex was solved by molecular replacement using only the structure of nb01 as starting model and the structure of pPorM-Nter could be built in the residual electron density map. The structures of pPorM-T, pPorM-Cter/nb130 and pPorM-Nter/nb01 are described in more details in the following sections.

Crystal structure of pPorM-T

pPorM-T crystallized in space group $P4_32_12$ with two molecules in the asymmetric unit forming a tight dimer with 4875\AA^2 buried surface area (corresponding to ~25% of the total surface of each monomer). This dimerization is indeed consistent with the molecular weight of the protein in solution, as measured by SEC-MALLS (16). The structure of pPorM-T is composed of three domains named D1 (residues 225-314), D2 (residues 318-398) and Dom3 (residues 400-516). A very prominent feature of the structure is the presence of domain swapped stretches within both D1 and D2 that lock the dimer. To confirm that this two-domain-swapping is not an artefact due to over-expression of a truncated construct, rational design of disulfide bonds was carried out using the Disulfide Design algorithm (18). The following residues were selected by visual analysis of the structure for cysteine mutation in full-length PorM: A230 with R305, A318 with A391, and M325. Then the two double PorM mutants A320C-R305C and A318C-A91C, and the single PorM mutant M325C were produced in *E. coli*, fused to a 6xHis tag. Crude extract samples in presence or not of DTT were analysed by SDS-PAGE and western-blot. Contrary to wild-type PorM, the three mutants showed one band at molecular weight corresponding to a dimer in absence of DTT, confirming that the introduced cysteines formed disulphide bridges and that domain-swapping is present in wild-type PorM.

Domain D1 displays the first domain swapping in which the N-terminal region from residue 225 to 253 is exchanged between the two chains of the dimer. An artificially deswapped domain (dsD1) encompassing residues 225 to 253 from chain B and residues 255 to 314 from chain A is presented in Figure 1. It assembles a β -sandwich of two antiparallel β -sheets with three β -strands each (strands ABE and GFC, respectively) packed face to face. This topology is similar to the C2-set of the immunoglobulin superfamily (19,20) in which the short C' strand is replaced by a short helical turn. A Dali search (21) with dsD1 revealed the highest similarity with the polycystic kidney disease (PKD) domain of the modular *Clostridium thermocellum* enzyme CtCel9D-Cel44A (PDB code 2C26, (22)) with a Z-value of 9.0 and a rmsd of 2.4 \AA . The PKD domain is an 80–90-amino acids module originally found in the human PKD1 gene encoding the cell surface glycoprotein polycystin-1 (23). The extracellular segments of human polycystin-1 are made up of 15 PKD domains. The PKD domains are usually found in the extracellular parts of proteins involved in protein–protein or protein–carbohydrate interactions (24).

D2 displays the second domain swapping. In this case only the first strand (residues 318 to 322) is exchanged between the two chains of the dimer. A deswapped artificial domain (dsD2) assembles residues 318 to 322 from chain B and residues 328 to 398 from chain A (Figure 1). dsD2 exhibits an immunoglobulin-like fold made up of seven β -strands in two β -sheets containing the ABED strands and the GFC strands, a hallmark of the C1-set topology. The closest structural homologue found using the DALI server (21) is a human filamin C domain (PDB 3V8O) with a Z-score of 8.0 and a rmsd of 1.7 \AA . Human filamins are flexible molecules composed of 24 similar Ig domains. The filamin Ig domain can be regarded as protein interaction modules. They often interact with other proteins by a β -sheet augmentation mechanism (25).

As indicated below, we were not able to build the structure of D3 in this crystal structure as the electron density for this region was too weak, especially in the loop regions. This problem was solved by using a stabilizing nanobody as described in the following part.

Crystal structure of pPorM-Cter in complex with nanobody nb130

Crystals of pPorM-Cter in complex with nanobody nb130 were obtained in several conditions. Data sets were collected for three crystal forms in three different space groups: P2₁ at 2.15 Å resolution with 4 complexes in the asymmetric unit forming 2 hetero tetramers; P2 at 2.5 Å resolution with 2 complexes in the asymmetric unit that are not forming a tetramer, although the hetero tetramer could be reconstituted by applying the crystallographic two-fold axis; P1 at 2.1 Å resolution with 8 complexes in the asymmetric unit forming 4 hetero tetramers. The structures were solved by molecular replacement using the structure of D2 of pPorM-T, and the structure of nb130 that was previously solved (17).

Taking into account the three crystal forms and the numerous complexes present in the different asymmetric units, a total of 7 independent hetero tetramers are structurally available. Two important features are present in all these crystal forms. The first feature is a clear electron density map for D3 that allowed us to trace its amino-acid chain accurately. Indeed, nanobody nb130 binds to D3 and makes several contacts in the crystal packing resulting in a stabilization of the C-terminal extremity of pPorM-Cter. In contrast, the second feature is the absence of electron density for D1 (residues 225-314) although pPorM-Cter includes residues 225 to 516, hence the three domains are present in the structure of pPorM-T. Notably, the produced pPorM-Cter protein shows a single band in SDS-PAGE analysis corresponding to a molecular weight of around 35 kDa. The elution time for pPorM-Cter on gel filtration column is comparable to those obtained for pPorM-T and corresponds to a dimer of around 70 KkDa. However, SDS-PAGE analysis of the crystals shows a shorter form for pPorM-Cter around 20 kDa (Figure S2). This size corresponds to the 202 residues that are present in the electron density map.

The hetero tetramer composed by two PorM-Cter and two nb130 displays a T shape with the PorM-Cter dimer forming the vertical bar and the two nanobodies the lateral ones (Figure 2). The D2 dimer of this complex is fully superimposable to that of PorM-T and displays the same domain swapping. The crystal structure of nb130 revealed a typical immunoglobulin domain with a scaffold consisting of two β-sheets connected by the highly conserved Cys23–Cys97 disulfide bond and presenting three antigen-binding CDR loops (Figure 2). Nb130 inserts the CDR3 in a deep crevice located at the interface between the two D3 domains (Figure 2, top view). The D3 monomer structure is composed of a large four-strand antiparallel β-sheet covered in one end by a short three-strand antiparallel β-sheet and in the other end by two α-helices. Two β-hairpin loops extend outward on each side and contribute to give an overall flattened shape to the domain. The D3 dimer resembles somehow to a flower bouquet with the four β-hairpin loops and the four α-helices pointing outward (Figure 1). A Dali search (21) with D3 revealed the highest similarity with the domain A of the L,D-transpeptidase 2 of *M. tuberculosis* (PDB code 5DZJ, (26)) with a Z-value of 6.1 and a rmsd of 2.9Å. This Ig-like domain belongs to the C-type Ig fold (27) but its function is not yet identified.

Crystal structure of pPorM-Nter in complex with nanobody nb01

The structure of pPorM-Nter in complex with nanobody nb01 was solved to 1.7 Å resolution. The structures of the nanobody nb01 free and in complex with pPorM-Nter are almost identical. Only very slight changes occur in the CDR loops. Several cycles of model building in the residual map and refinement permit to obtain the structure of a large part of pPorM-Nter, from residue 44 to 217. The overall fold of the pPorM-Nter can be described as an antiparallel four-helix bundle with an up–down, up–down, right-handed topology (Figure 3).

Comparison of the pPorM-Nter structure to the PDB database using the DALI search engine (21) revealed that the pPorM-Nter domain shows strong structural similarity to the focal adhesion kinase (PDB 3B71, (28)) with a Z-score of 10.3 and a rmsd of 2.7 Å. The 25 first residues of pPorM-Nt have not been built

SAXS analysis of pPorM

Synchrotron SAXS data were collected for the proteins pPorM, pPorM-T, pPorM-Ct and pPorM-Nt, in order to complete X-ray crystallography structural data with analyses in solution. For all the proteins, SAXS experiments gave good quality data. There are no radiation damages of the protein, as we can observe on the scattering curves (Figure 4A). On these scattering curves, the Guinier region (almost the first 100 points) of each sample can be fitted by a straight line (Figure 4B), that indicates the absence of aggregates for all the studied samples. The estimated R_g and $I(0)$ are reported table 2. The molecular weights have been estimated using SAXS MoW2 online package (Table 2). pPorM-Nt appears to be a monomer while pPorM-T and pPorM-Ct seem more to be in a dimer state, as expected regarding the crystal structures. Concerning pPorM, an extended dimer is possible.

The dimensionless Kratky plots (Figure 4C) give informations on globularity and flexibility of the molecule. The proteins pPorM, pPorM-T and pPorM-Ct curves are similar to the theoretical Kratky curve from partially flexible protein, showing a peak with a maximum around $q=0.8 \text{ nm}^{-1}$, decreasing and followed by plateau around $I \cdot q^2=1$. For the protein pPorM-Nt, the late decreasing of the peak without following plateau may be due to a higher level of flexibility than in the other proteins.

Distance distribution functions ($P(r)$) (Figure 4D) were obtained using the program Gnom. The estimations of R_g and $I(0)$ from $P(r)$, reported in Table 2, are quite similar to those obtained by Guinier analyses. The D_{\max} computed from $P(r)$ for pPorM-T is in agreement with the crystal structure (approximative value, without hydration layer) (Table 2). This is not the case for pPorM-Ct and pPorM-Nt. In the case of pPorM-Ct the discrepancy between the value obtained in solution with a freshly prepared sample and the value measured in the crystal structure could be explain by the absence of the D1 domain in the crystal which is probably cleaved before crystal formation. In the case of pPorM-Nt the 25 first residues that cannot be built due to the lack of electron density map are probably highly flexible and could explain the larger D_{\max} observed in solution. The general look of these curves gives shape information about proteins. While pPorM-T and pPorM-Ct are almost the same and looks like a modular protein, pPorM-Nt has a long tail responsible of this big D_{\max} . This shape can be due to an extended part, maybe disordered. The curve corresponding to the protein pPorM looks like an extended protein.

The comparison of these data with the theoretical scattering of pPorM-T crystal structure using CRY SOL revealed a really good agreement of these data with a value of χ^2 of 0.95 (Figure 5A). This result means that the structure obtained by X-ray crystallography is in an agreement with conformers find in solution and analyzed by SAXS. We then wanted to see if the crystallographic structure of pPorM-T is the conformer with the best fit with pPorM-T in solution. The simulation using normal mode analysis (SREFLEX) gave ten refined structures but none had a best χ^2 value even if there still are very good (Figure 5C). The differences are mainly of week degree of flexibility for the three domains and no big difference of angle is visible between D1 and D2.

As it was previously said, D1 is missing in the crystal structure of pPorM-Ct that can easily explain the bad agreement with SAXS data (Figure S3A). However, the crystal structure of

pPorM-T fits better with SAXS data from pPorM-Ct, even if the χ^2 value still not good (Figure S3A).

During the refinement step of pPorM-Nt, few densities were visible for the N-terminal part of the protein. It was not enough to let us complete the structure. The crystal structure is so uncompleted and a CRY SOL analysis has no sense. To get an idea of the total shape of pPorM-Nt, DAMMIF was performed using ATSAS online and the best model, with its fit on SAXS data, has been reported on Figure S3B. The very good value of χ^2 allows us to consider this shape as the shape of a possible conformer in solution.

PorM homologs

Until now, T9SS genes have been identified only in the genome of bacteria from the *Bacteroidetes* phylum (Abby et al., 2016). To further compare the protein sequences from PorM homologs, we used the BLAST program. As expected, closest homologs among the 100 first aligned sequences are from the *Porphyromonas* genus (with sequence identity from 99% to 47%, with more than 99% of sequence covering), and other *Bacteroidetes* (with sequence identity from 61% to 46%, with more than 96% of sequence covering). Interestingly, when Blast is run with exclusion of *Bacteroidetes* sequences, the closest homolog is a sequence obtained by a shotgun sequencing of the human pathogen *Chlamydia trachomatis*. However, no PorM homologs are present in other *Chlamydia trachomatis* serova.

Concluding remarks

In this study we have solved the crystal structure of the N-terminal and C-terminal domains of the periplasmic part of the protein PorM. The crystal structures combined with the results of the SAXS analysis suggest a rigid extended conformation for pPorM. One of the main feature of the crystal structure of pPorM-T is the presence of domain swapping. Indeed two swapped domains were identified in pPorM-T structure and were experimentally confirmed. These swapped domains are not located at the N- or C-terminus but rather involve internal sequences. Indeed, the swap domain present at the N-terminus of D1 can be considered as internal, as the protein was crystallized after trypsinization. Inspection of 3DSwap, an integrated knowledgebase of proteins involved in domain swapping, revealed that less than 15% of the 293 present in the database possess internal swapped-domains. Furthermore, only very few structures possess two swapped domains. Actually, only one structure in the database possess two internal swapped domains, the hexameric NC1 domains of type IV collagen. We propose that this almost unique feature of PorM stabilises the dimer in a highly rigid conformation. Further studies will be necessary to assess the role of this specular feature of PorM.

Materials and Methods

Protein expression and purification

Expression and purification of the different PorM constructs (pPorM, pPorM-T, pPorM-Nter and pPorM-Cter) and of the nanobodies (nb01 and nb130) were performed as previously described (16,17). The PorM constructs were produced from Rosetta (DE3) pLysS *E. coli* cells cultured over night (ON) in ZYP-5052 auto-induction medium (29). The recombinant proteins were purified by nickel-affinity chromatography on a 5mL Ni-NTA column (GE Healthcare) in 50mM Tris pH8.0, 300mM NaCl, 250mM imidazole, followed by size-exclusion

chromatography on a HiLoad 16/60 Superdex200 PG column (GE Healthcare) equilibrated in 10mM Hepes pH7.5, 150mM NaCl. pPorM-T was obtained by ON trypsinolysis (1:1000 w:w) of pPorM at RT, followed by size-exclusion chromatography. Selenomethionine-substituted pPorM-T (SeMet pPorM-T) was produced from Rosetta (DE3) pLysS *E. coli* cells cultured ON in PSAM-5052 auto-induction medium (29), and purified identically to the native pPorM-T.

The nanobodies nb01 and nb02 were produced from WK6 *E. coli* cells cultured ON in Terrific Broth medium. The periplasmic fraction was extracted by osmotic shock and the recombinant proteins were purified by nickel-affinity chromatography on a 5mL Ni-NTA column (GE Healthcare) in 50mM Na/K phosphate, 300mM NaCl, 10% glycerol, 250mM imidazole, followed by size-exclusion chromatography on a HiLoad 10/30 Superdex75 PG gel-filtration column (GE Healthcare) equilibrated in PBS.

The complexes pPorM-Nter/nb01 and pPorM-Cter/nb130 were obtained by mixing the two partners purified independently by nickel-affinity chromatography, with a molar ratio 1:1.2 (PorM constructs:nanobodies); the complexes were then purified by size-exclusion chromatography on a HiLoad 16/60 Superdex200 PG column (GE Healthcare) equilibrated in 10mM Hepes pH7.5, 150mM NaCl.

The purified proteins and complexes were concentrated by centrifugation and their concentration was determined by measuring the absorbance at 280_{nm} with a NanoDrop 2000 (Thermo Scientific).

SEC/MALS/UV/RI analysis

For the size exclusion chromatography (SEC) with on-line multiangle laser light scattering (MALS), Absorbance (UV), and refractive index (RI) analysis, pure T9-19-tryp was concentrated to XX mg.mL⁻¹ and stored at 4°C. Protein was centrifuged (10000g for 5 min) just before the experiment. Size exclusion chromatography was carried out on an Alliance 2695 HPLC system (Waters) using a Silica Gel KW804 column (Shodex) equilibrated in 10 mM Tris pH 7.5 and 150 mM NaCl at a flow of 0.5 ml/min. Detection was performed using a triple-angle light scattering detector (Mini-DAWN TREOS, Wyatt Technology), a quasi-elastic light scattering instrument (Dynapro, Wyatt Technology), and a differential refractometer (OptilabrEX, Wyatt Technology).

Crystallization, data collection and processing

Crystallization and data collection of pPorM-T, SeMet pPorM-T, and of the complexes pPorM-Nter/nb02 and pPorM-Cter/nb01 were performed as described previously (16,17). Initial crystallization trials were performed by the sitting-drop vapour-diffusion method at 293K in 96-well Swissci plates (Hampton Research), with the Wizard I and II (Emerald BioSystems), JCSG+ (Quiagen), Index (Hampton research) and Crystal Screens 1 and 2 (Hampton Research) screens, using a Mosquito Crystal robot (TTP Labtech). For pPorM-T and SeMet pPorM-T, optimization was then carried out by varying the pH and the concentration of precipitant (30). Crystals were mounted in cryo-loops (Hampton CrystalCap Magnetic) and were briefly soaked in crystallization solution supplemented with the appropriate cryo-protectant before being flash-frozen in a nitrogen gas stream at 100K using a home cryo-cooling device (Oxford Cryosystems). For the complex pPorM-Cter/nb130, a 5min air dehydration step was performed in the crystallization solution with a 10% increase in precipitant concentration, and supplemented with the cryo-protectant, leading to new space group crystals (17).

Diffraction data were collected on beamlines ID23 at the European Synchrotron Research

Facilities (ESRF, Grenoble, France), and Proxima1 at the French synchrotron SOLEIL (Paris, France). For SeMet pPorM-T, a fluorescence scan was performed to determine the peak and inflection wavelengths (0.979109Å and 0.979336Å, respectively) and MAD data were collected with a remote wavelength at 0.976256Å.

The different data sets were integrated with XDS (31) and were scaled with SCALA (32) from CCP4 Suite (33). Data collection statistics are reported in Table 1. The Matthews coefficient and solvent content were calculated with the programs MATTHEWS_COEF (34) from CCP4 Suite.

Structure solution and refinement

The structure of pPorM-T was solved by the multiple-wavelength anomalous diffraction (MAD) method using the SeMet pPorM-T data set at 3.1 Å resolution. Heavy-atom substructure determination, positional refinement, phase calculations and solvent flattening were performed using autoSHARP (35), SHARP (36) and SOLOMON (37), as previously described (16). The partial model of SeMet pPorM-T was built using Turbo-Frodo (38), and was subsequently used as model for molecular replacement with MOLREP (39) from CCP4 Suite to solve the structure of native pPorM-T at 2.85Å. The structures of pPorM-Cter in complex with nanobody nb130 was solved by molecular replacement with MOLREP using the partial model of domains D2 and D3 of pPorM-T, and the structure of nb130 (17) as models. The building of domain D3 of pPorM-T and pPorM-Cter was then completed manually with COOT. The structure of pPorM-Nter in complex with nb02 was solved by molecular replacement with MOLREP using the structure of nb02 (17) as starting model.

Refinement was performed with autoBUSTER using auto-ncs parameter, and the structures were corrected with COOT. Model validations were done with Molprobity. Refinement statistics are reported in Table 1.

SAXS studies

Data acquisition and data reduction. pPorM, pPorM-T, pPorM-Ct and pPorM-Nt samples analysed using SAXS came from size exclusion column. Proteins were eluted in 10 mM Hepes pH 7.4 150 mM NaCl buffer and the different concentrations were taken as the concentration using centricon increase. All small-angle X-ray scattering (SAXS) data for pPorM, pPorM-T, pPorM-Ct and pPorM-Nt proteins in 10 mM Hepes pH 7.4 150 mM NaCl buffer were collected at the BioSAXS beamline (BM29) of the ESRF (Grenoble, France). Data were collected using a PILATUS 1M 2D detector (Dectris) at a sample-detector distance of 2.869 m. The wavelength (λ) used was 0.09919 nm, leading to scattering vectors q ranging from 0.036 to 4.912 nm⁻¹. The scattering vector is defined as $q = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle. SAXS data were collected at 10°C using purified protein samples (30 μ l each) and concentration flowthrough as buffer samples. To avoid radiation damages, 5 mM DTT was added to some samples. Protein concentrations were as follow: 0.73 to 10 mg/mL for pPorM, 0.43 to 4.40 mg/mL for pPorM-T, 0.94 to 8.86 mg/mL for pPorM-Ct and 0.52 to 16.11 mg/mL for pPorM-Nt.

Samples were loaded in a fully automated sample changer. To control radiation damages, 10 successive exposures of 1 s each were made for each protein concentration. Concerning the buffer, 10 first curves were taken before and after protein scattering acquisition.

The forward scattering $I(0)$, the gyration radius R_g , pair distribution of particules distribution $p(r)$ and maximum dimension D_{max} were computed first following the automated SAXS data analysis pipeline and then compared with manually computed values obtained using the

program PRIMUS from the ATSAS package. Data were normalized with transmitted beam intensity to give the average scattering curve for each measurement. Each curve affected by radiation damages was manually excluded at this step. Signal corresponding to buffer scattering was subtracted and all scattering curves were scaled. By use of the Guinier analysis, each curve showing aggregated signal was also excluded, according to their estimated R_g and $I(0)$. After a first analysis of each concentration, a low and a high concentration were merged on at least 100 common points and a chimera curve was also obtained. The level of flexibility was analysed by use of dimensionless Kratky plot and the $I(0)$, R_g and D_{max} were then computed following the program GNOM. The shape of the pair distribution of distances curve also informs about the global shape of the protein.

MW_{SAXS} is computed using the server SAXS MoW 2 (<http://saxs.ifsc.usp.br/>). This estimation gives also an estimation of the oligomerization state. We used CRY SOL to compare the chimera scattering curves obtained by SAXS with theoretical scattering curves computed from X-ray crystallography structure pdb file. The discrepancy χ^2 is measured with a value of 1 for a perfect fit. The program SREFLEX was used to generate 10 models following the flexibility of the high-resolution crystal model of pPorM-T. We determined an ab-initio shape of the pPorM-Nt protein using the program DAMMIF, based on data analysed by GNOM.

Acknowledgments

We would like to thank the European Synchrotron Research Facilities (ESRF, Grenoble, France) and the Soleil synchrotron (Saint-Aubin, France), and in particular the beamlines BM29, ID23 and Proxima1 staffs, respectively, for their assistance.
FRISBI, ANR ToothPaste

References

1. Costa, T. R., Felisberto-Rodrigues, C., Meir, A., Prevost, M. S., Redzej, A., Trokter, M., and Waksman, G. (2015) Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat Rev Microbiol* **13**, 343-359
2. Sato, K., Naito, M., Yukitake, H., Hirakawa, H., Shoji, M., McBride, M. J., Rhodes, R. G., and Nakayama, K. (2010) A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proc Natl Acad Sci U S A* **107**, 276-281
3. McBride, M. J., and Zhu, Y. (2013) Gliding motility and Por secretion system genes are widespread among members of the phylum bacteroidetes. *J Bacteriol* **195**, 270-278
4. Abby, S. S., Cury, J., Guglielmini, J., Neron, B., Touchon, M., and Rocha, E. P. (2016) Identification of protein secretion systems in bacterial genomes. *Sci Rep* **6**, 23080
5. Nelson, S. S., Bollampalli, S., and McBride, M. J. (2008) SprB is a cell surface component of the *Flavobacterium johnsoniae* gliding motility machinery. *J Bacteriol* **190**, 2851-2857
6. Nakane, D., Sato, K., Wada, H., McBride, M. J., and Nakayama, K. (2013) Helical flow of surface protein required for bacterial gliding motility. *Proc Natl Acad Sci U S A* **110**, 11145-11150
7. Hajishengallis, G. (2015) Periodontitis: from microbial immune subversion to systemic inflammation. *Nat Rev Immunol* **15**, 30-44
8. Pike, R., McGraw, W., Potempa, J., and Travis, J. (1994) Lysine- and arginine-specific proteinases from *Porphyromonas gingivalis*. Isolation, characterization, and evidence for the existence of complexes with hemagglutinins. *J Biol Chem* **269**, 406-411

9. Pavloff, N., Potempa, J., Pike, R. N., Prochazka, V., Kiefer, M. C., Travis, J., and Barr, P. J. (1995) Molecular cloning and structural characterization of the Arg-gingipain proteinase of *Porphyromonas gingivalis*. Biosynthesis as a proteinase-adhesin polyprotein. *J Biol Chem* **270**, 1007-1010
10. Curtis, M. A., Kuramitsu, H. K., Lantz, M., Macrina, F. L., Nakayama, K., Potempa, J., Reynolds, E. C., and Aduse-Opoku, J. (1999) Molecular genetics and nomenclature of proteases of *Porphyromonas gingivalis*. *J Periodontal Res* **34**, 464-472
11. Potempa, J., Sroka, A., Imamura, T., and Travis, J. (2003) Gingipains, the major cysteine proteinases and virulence factors of *Porphyromonas gingivalis*: structure, function and assembly of multidomain protein complexes. *Curr Protein Pept Sci* **4**, 397-407
12. Rangarajan, M., Aduse-Opoku, J., Slaney, J. M., Young, K. A., and Curtis, M. A. (1997) The prpR1 and prR2 arginine-specific protease genes of *Porphyromonas gingivalis* W50 produce five biochemically distinct enzymes. *Mol Microbiol* **23**, 955-965
13. Lasica, A. M., Ksiazek, M., Madej, M., and Potempa, J. (2017) The Type IX Secretion System (T9SS): Highlights and Recent Insights into Its Structure and Function. *Front Cell Infect Microbiol* **7**, 215
14. Vincent, M. S., Canestrari, M. J., Leone, P., Stathopoulos, J., Ize, B., Zoued, A., Cambillau, C., Kellenberger, C., Roussel, A., and Cascales, E. (2017) Characterization of the *Porphyromonas gingivalis* Type IX Secretion Trans-envelope PorKLMNP Core Complex. *J Biol Chem* **292**, 3252-3261
15. Gorasia, D. G., Veith, P. D., Hanssen, E. G., Glew, M. D., Sato, K., Yukitake, H., Nakayama, K., and Reynolds, E. C. (2016) Structural Insights into the PorK and PorN Components of the *Porphyromonas gingivalis* Type IX Secretion System. *PLoS Pathog* **12**, e1005820
16. Stathopoulos, J., Cambillau, C., Cascales, E., Roussel, A., and Leone, P. (2015) Crystallization and preliminary X-ray analysis of the C-terminal fragment of PorM, a subunit of the *Porphyromonas gingivalis* type IX secretion system. *Acta Crystallogr F Struct Biol Commun* **71**, 71-74
17. Duhoo, Y., Roche, J., Trinh, T. T. N., Desmyter, A., Gaubert, A., Kellenberger, C., Cambillau, C., Roussel, A., and Leone, P. (2017) Camelid nanobodies used as crystallization chaperones for different constructs of PorM, a component of the type IX secretion system from *Porphyromonas gingivalis*. *Acta Crystallogr F Struct Biol Commun* **73**, 286-293
18. Craig, D. B., and Dombkowski, A. A. (2013) Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinformatics* **14**, 346
19. Mornon, J. P., Halaby, D., Malfois, M., Durand, P., Callebaut, I., and Tardieu, A. (1998) alpha-Crystallin C-terminal domain: on the track of an Ig fold. *Int J Biol Macromol* **22**, 219-227
20. Bodelon, G., Palomino, C., and Fernandez, L. A. (2013) Immunoglobulin domains in *Escherichia coli* and other enterobacteria: from pathogenesis to applications in antibody technologies. *FEMS Microbiol Rev* **37**, 204-250
21. Holm, L., and Laakso, L. M. (2016) Dali server update. *Nucleic Acids Res* **44**, W351-355
22. Najmudin, S., Guerreiro, C. I., Carvalho, A. L., Prates, J. A., Correia, M. A., Alves, V. D., Ferreira, L. M., Romao, M. J., Gilbert, H. J., Bolam, D. N., and Fontes, C. M. (2006) Xyloglucan is recognized by carbohydrate-binding modules that interact with beta-glucan chains. *J Biol Chem* **281**, 8815-8828
23. Hughes, J., Ward, C. J., Peral, B., Aspinwall, R., Clark, K., San Millan, J. L., Gamble, V., and Harris, P. C. (1995) The polycystic kidney disease 1 (PKD1) gene encodes a novel protein with multiple cell recognition domains. *Nat Genet* **10**, 151-160

24. Leettola, C. N., Knight, M. J., Cascio, D., Hoffman, S., and Bowie, J. U. (2014) Characterization of the SAM domain of the PKD-related protein ANKS6 and its interaction with ANKS3. *BMC Struct Biol* **14**, 17
25. Nakamura, F., Pudas, R., Heikkinen, O., Permi, P., Kilpelainen, I., Munday, A. D., Hartwig, J. H., Stossel, T. P., and Ylanne, J. (2006) The structure of the GPIb-filamin A complex. *Blood* **107**, 1925-1932
26. Kumar, P., Kaushik, A., Lloyd, E. P., Li, S. G., Mattoo, R., Ammerman, N. C., Bell, D. T., Perryman, A. L., Zandi, T. A., Ekins, S., Ginell, S. L., Townsend, C. A., Freundlich, J. S., and Lamichhane, G. (2017) Non-classical transpeptidases yield insight into new antibacterials. *Nat Chem Biol* **13**, 54-61
27. Both, D., Steiner, E. M., Stadler, D., Lindqvist, Y., Schnell, R., and Schneider, G. (2013) Structure of LdtMt2, an L,D-transpeptidase from *Mycobacterium tuberculosis*. *Acta Crystallogr D Biol Crystallogr* **69**, 432-441
28. Garron, M. L., Arthos, J., Guichou, J. F., McNally, J., Cicala, C., and Arold, S. T. (2008) Structural basis for the interaction between focal adhesion kinase and CD4. *J Mol Biol* **375**, 1320-1328
29. Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**, 207-234
30. Lartigue, A., Gruez, A., Briand, L., Pernollet, J. C., Spinelli, S., Tegoni, M., and Cambillau, C. (2003) Optimization of crystals from nanodrops: crystallization and preliminary crystallographic study of a pheromone-binding protein from the honeybee *Apis mellifera* L. *Acta Crystallogr D Biol Crystallogr* **59**, 919-921
31. Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* **66**, 133-144
32. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**, 72-82
33. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242
34. Kantardjieff, K. A., and Rupp, B. (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12**, 1865-1871
35. Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) Automated structure solution with autoSHARP. *Methods Mol Biol* **364**, 215-230
36. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003) Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr D Biol Crystallogr* **59**, 2023-2030
37. Abrahams, J. P., and Leslie, A. G. (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D Biol Crystallogr* **52**, 30-42
38. Roussel, A., and Cambillau, C. (1991) Turbo-Frodo. in *Silicon Graphics Geometry Partners Directory* (Graphics, S. ed.), Mountain View, CA. pp 86
39. Vagin, A., and Teplyakov, A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **66**, 22-25

Figure legends

Figure 1: X-ray structure of pPorM-T.

Ribbon diagram of the pPorM-T dimer viewed in two orientations rotated by 90° along the vertical axis. One monomer is colored in grey and the other is colored in a rainbow spectrum with blue for the N-terminus and red for the C-terminus. Close-up view of the D1, D2 and D3 domains are shown.

Figure 2. X-ray structure of pPorM-Ct in complex with nanobody nb130

Ribbon diagram of the pPorM-Ct dimer in complex with two nanobodies nb130 viewed in two orientations rotated by 90° along the horizontal axis (side view and top view). One monomer and one nanobody are colored in grey and the other monomer and nanobody are colored in a rainbow spectrum with blue for the N-terminus and red for the C-terminus.

Figure 3. X-ray structure of pPorM-Nt in complex with nanobody nb01

Ribbon diagram of pPorM-Nt in complex with nanobody nb01. The two molecules are colored in a rainbow spectrum with blue for the N-terminus and red for the C-terminus.

Figure 4. SAXS analyses of pPorM, pPorM-T, pPorM-Ct and pPorM-Nt using PRIMUS software. (A) Scaled scattering curves (B) Guinier fits on scattering curves (C) Dimensionless Kratky plots (D) Normalized distance distribution function of particle curves. Data for pPorM protein are in blue, for pPorM-T in red, for pPorM-Ct in green and for pPorM-Nt in black.

Figure 5 SAXS analyses of pPorM-T using CRY SOL and SREFLEX software. (A) CRY SOL of pPorM-T (B) CRY SOL of the 10 models generated by SREFLEX (C) PDB files of the 10 models.

Figure 6. Model of PorM full length

The structures of pPorM-Nt and pPorM-Ct have been combined according to the SAXS data in an extended confirmation. The transmembrane helix as well as the first helical part of pPorM-Nt have been modelized. PorM is anchored to the inner membrane and binds to the outer membrane thanks to the interaction with PorK/PorN complex.

Supplementary figure legends

Figure S1: Structural alignment of the three domains of pPorM-T with homolog structures given by Dali

Figure S2: SDS-PAGE analysis of pPorM-Ct-nanobody nb130 crystals

Figure S3: (A) SAXS analyses of pPorM-Ct using CRY SOL. (B) SAXS *ab-initio* shape modeling of pPorM-Nt using DAMMIF

Table 1 : data collection and refinement statistics

Table 2 : SAXS data collection and scattering-derived structural parameters.

	pPorM	pPorM-T	pPorM-Ct	pPorM-Nt
<i>Data collection parameters</i>				
Instrument	BM29 (ESRF)	BM29 (ESRF)	BM29 (ESRF)	BM29 (ESRF)
Detector	Pilatus (1M)	Pilatus (1M)	Pilatus (1M)	Pilatus (1M)
Beam geometry	4*4	4*4	4*4	4*4
Wavelength (nm)	0.09919	0.09919	0.09919	0.09919
q range (nm ⁻¹)	0.036-4.912	0.036-4.912	0.036-4.912	0.036-4.912
Exposure time (s)	1	1	1	1
Concentration range (g/L)	0.73-10	0.43-4.40		0.52-16.11
Temperature (K)	283	283	283	283
<i>Structural parameters</i>				
I(0) (cm ⁻¹) (from P(r))	63.97	39.68	37.74	15.03
I(0) (cm ⁻¹) (from Guinier)	64.04 +/- 0.33	38.10 +/- 0.12	37.30 +/- 0.05	14.69 +/- 0.03
Rg (Å) (from P(r))	56.6	38.4	38.8	30.2
Rg (Å) (from Guinier)	55.3 +/- 0.05	37.9 +/- 2.4	37.7 +/- 2.8	29.5 +/- 2.3
D _{max} (Å) (from P(r))	190	126	126	119
D _{max} (Å) (from crystal structure)		119	88	59
<i>Molecular mass (M) (kDa)</i>				
M theoretical	55.357	34.328	34.328	21.946
M from SAXS MoW2 ³	147.201	72.259	72.760	20.779
<i>Software employed</i>				
Primary data reduction	PRIMUS	PRIMUS	PRIMUS	PRIMUS
Data processing	GNOM	GNOM	GNOM	GNOM
Validation of structural models		CRY SOL SREFLEX	CRY SOL	CRY SOL DAMMIF

Table 2. SAXS data collection and scattering-derived structural parameters.

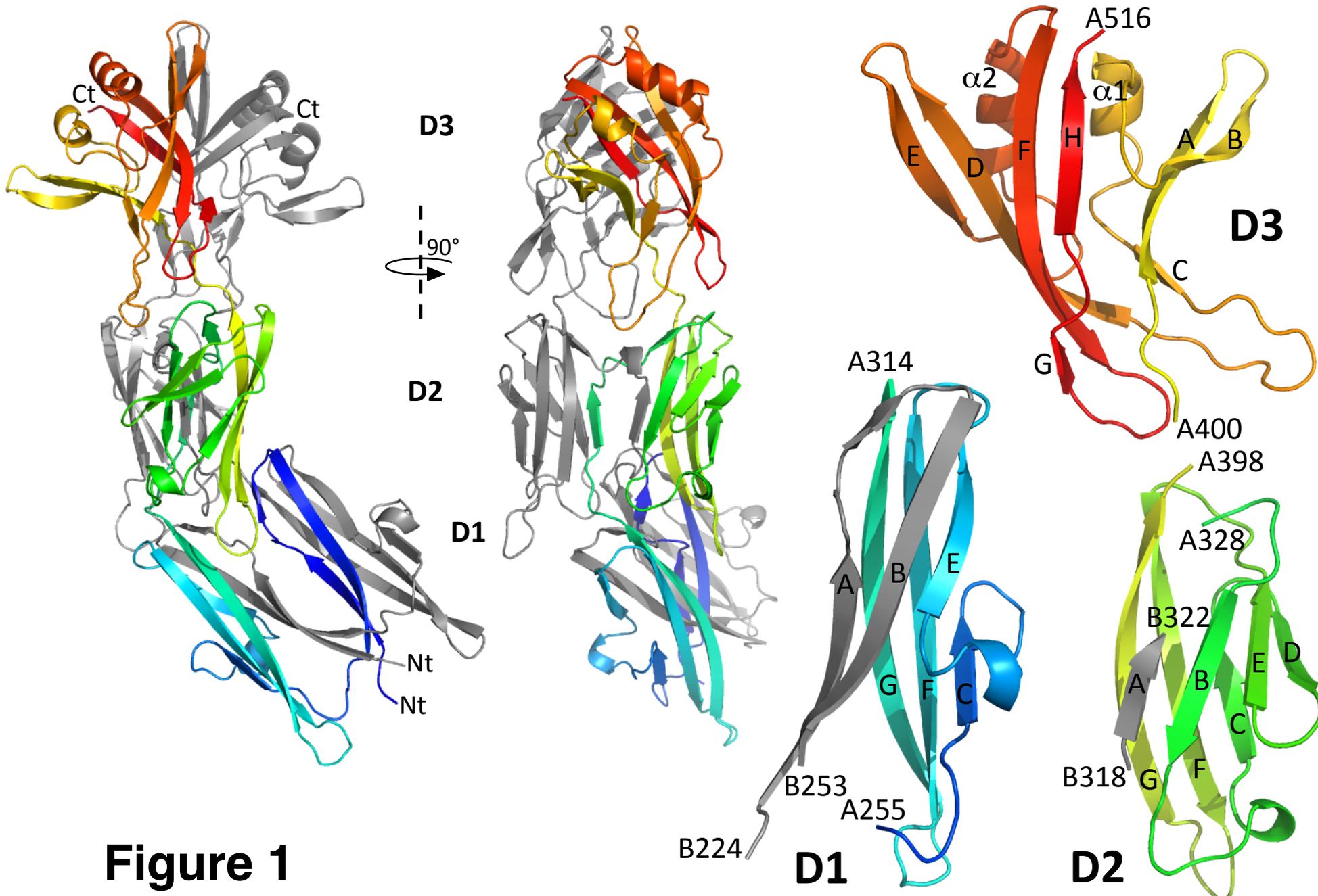


Figure 1

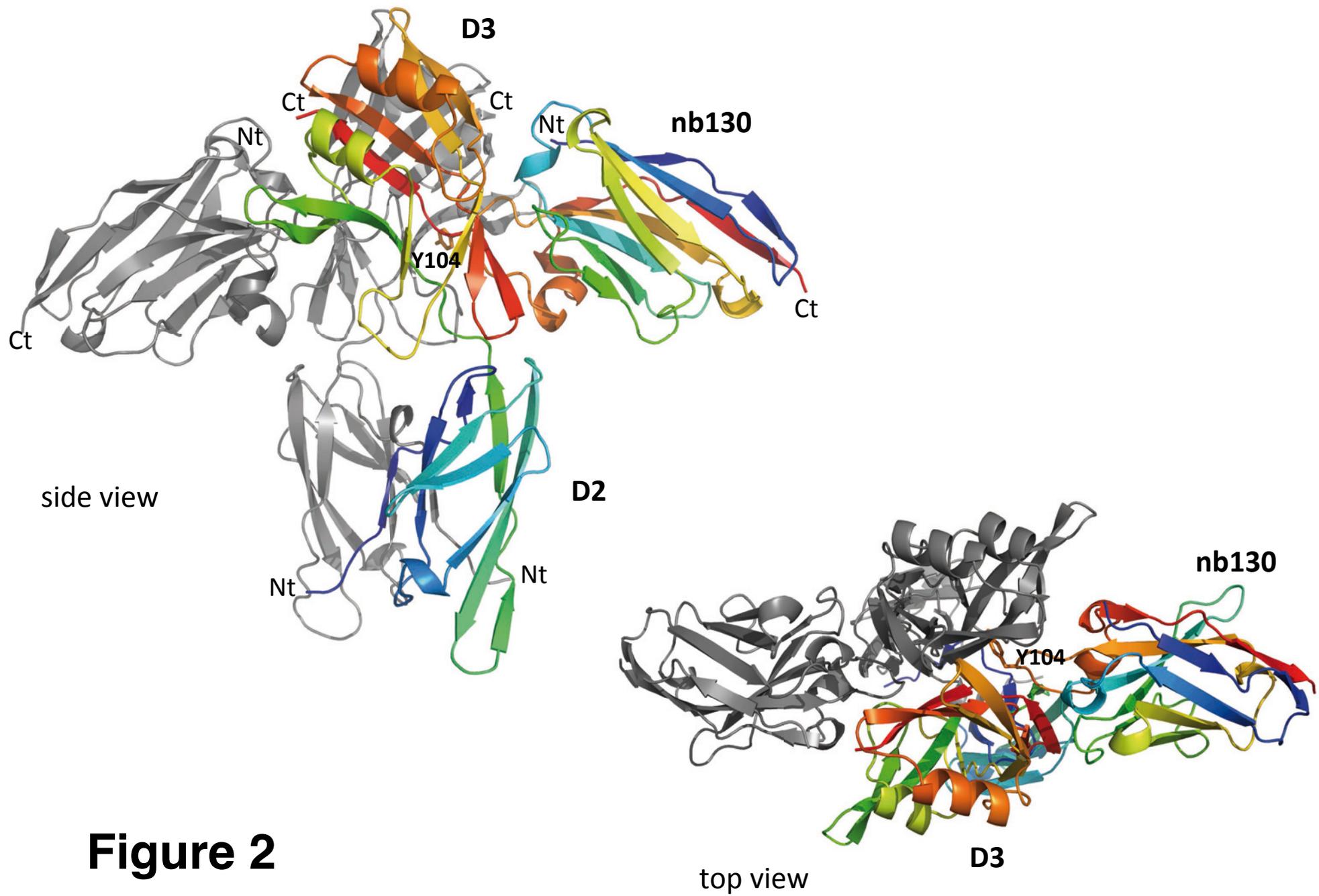


Figure 2

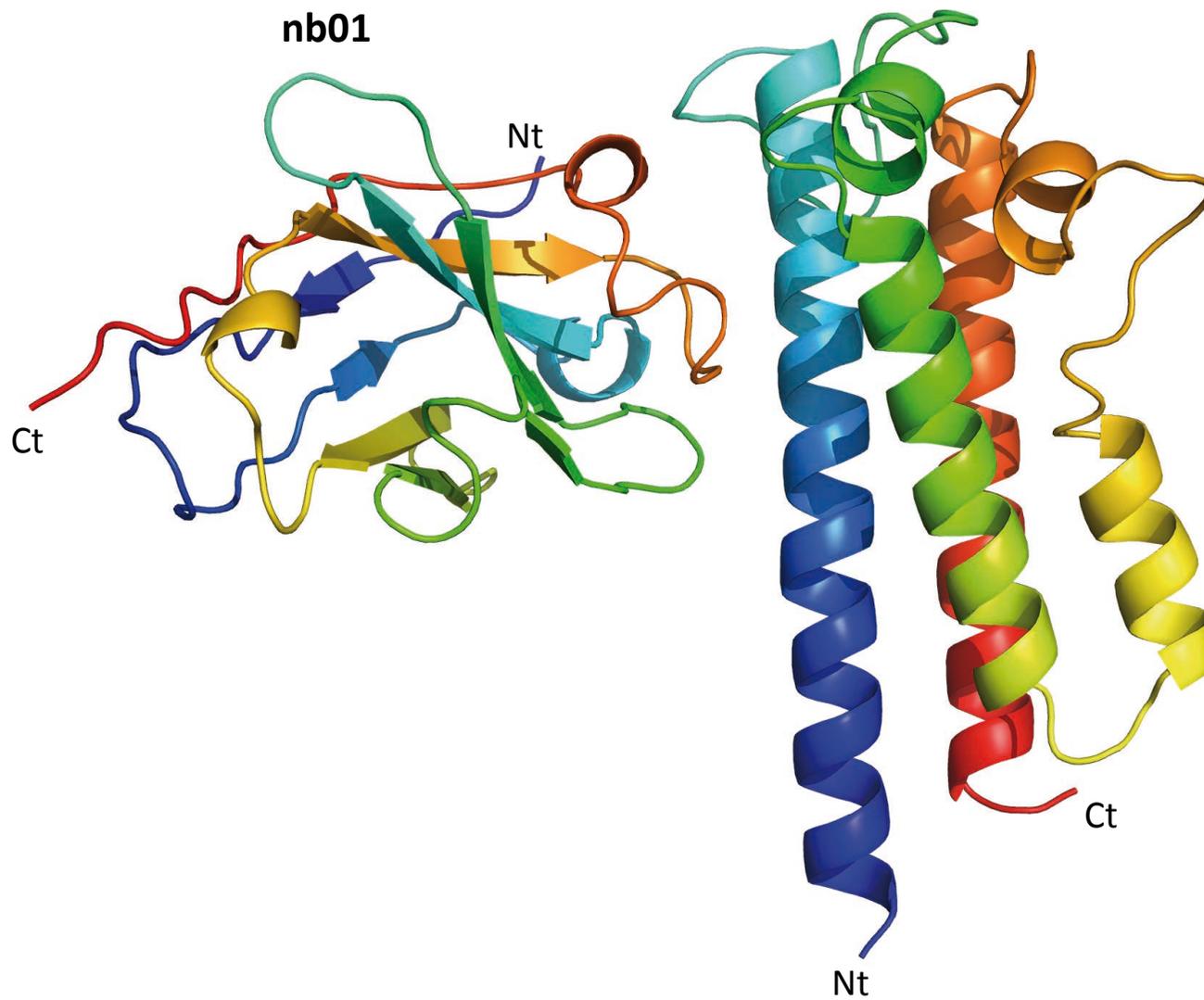


Figure 3

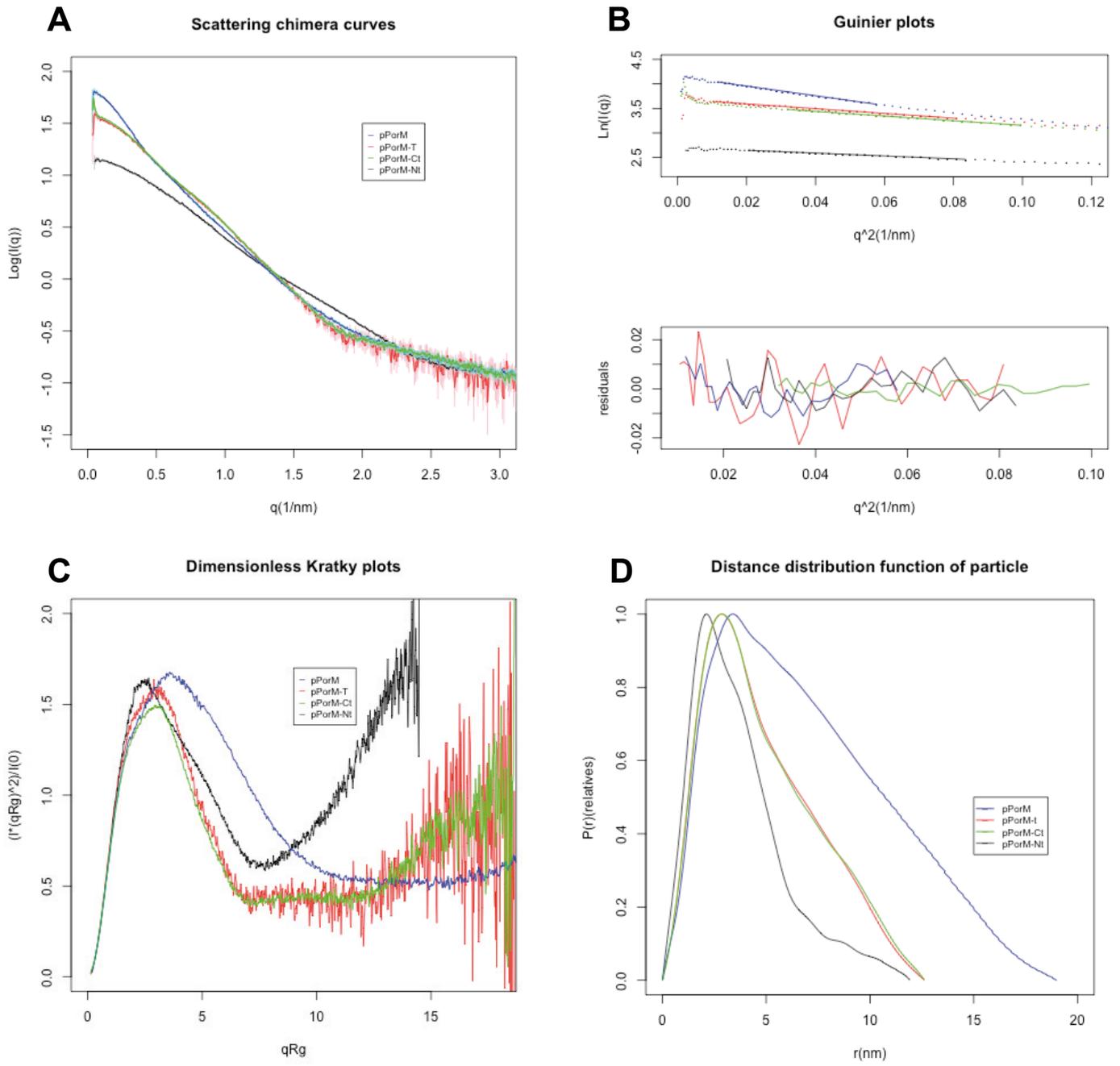
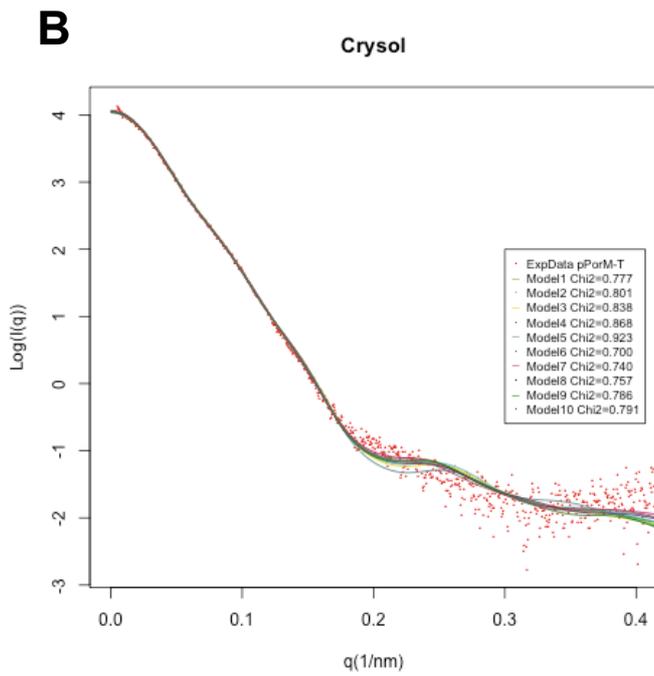
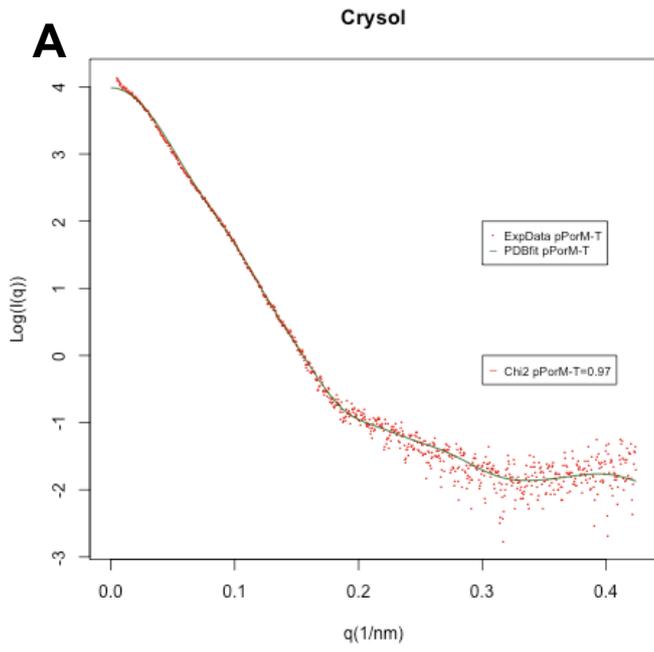


Figure 4



C



Figure 5

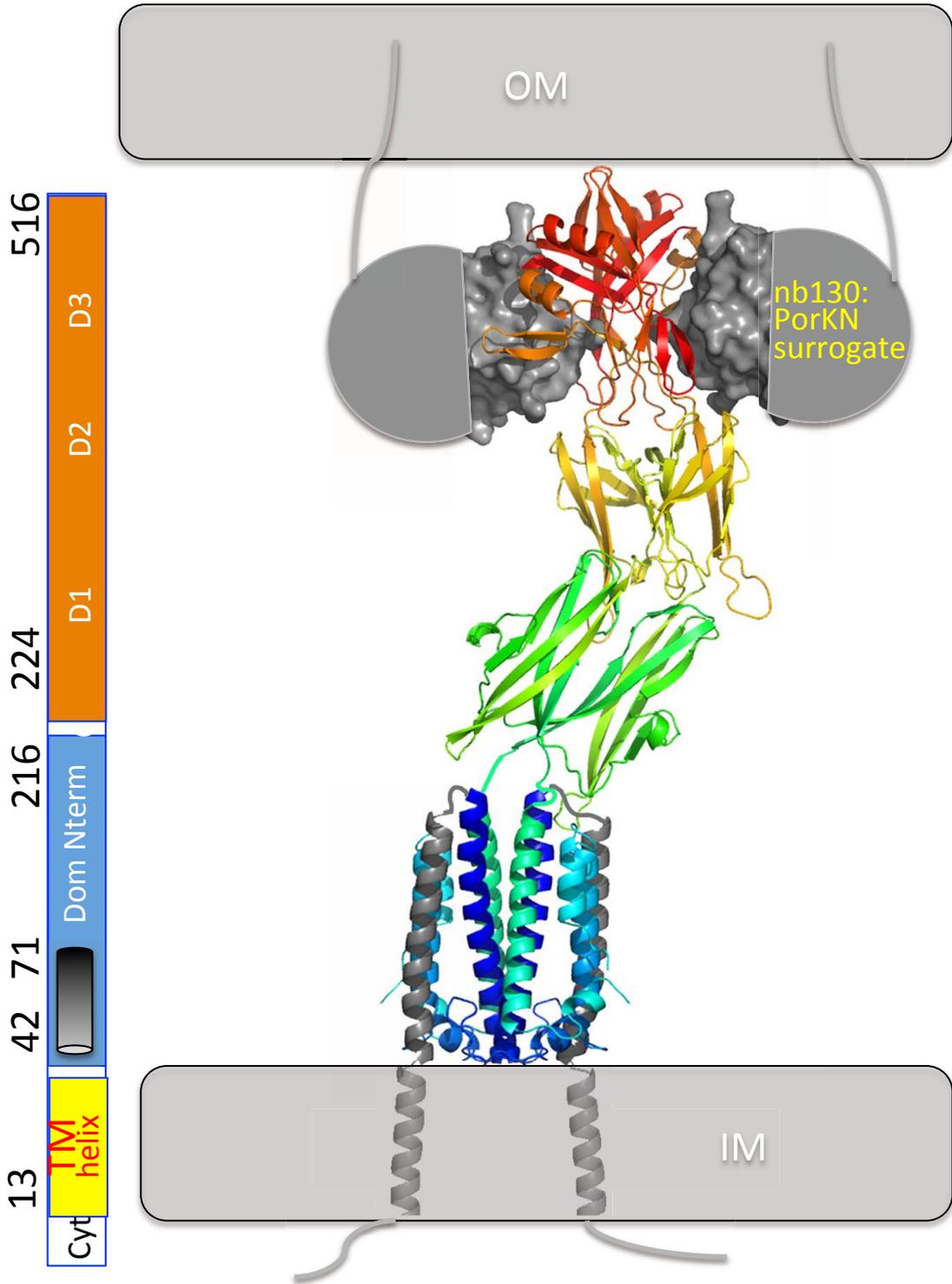


Figure 6

Figure S1

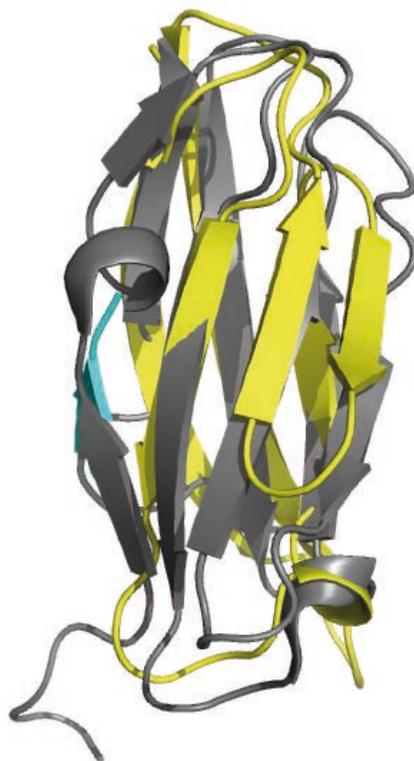
Structural alignment (DALI)

2C26, Z=9.0, rmsd=2.4,
lali=78, id=15



D1

3V80, Z=8.0, rmsd=1.7,
lali=70, id=10



D2

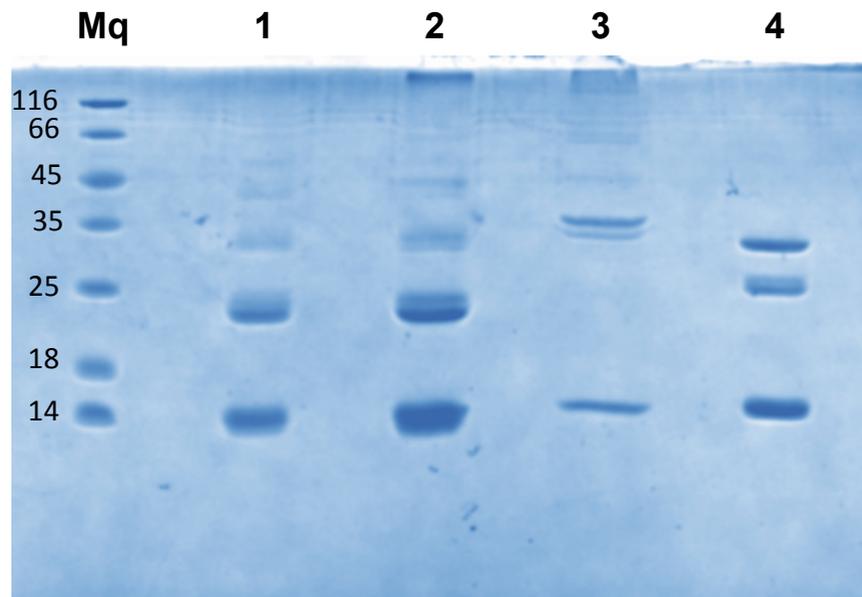
4QTF, Z=6.1, rmsd=2.8,
lali=79, id=14



D3

Figure S2

18% (home-made): J110 after purification/crystallization



1: First Xtals (<21 days) from index G1-F11-H3

2: Late Xtals (>65 days) from index F1-F7-F8-F9-F10
G4-G10-G11

3: drops from JCSG+ A5-A6

4: protein used for Xtallo stored at 4°C
=> 30 lot5a / VHH130 lot2 GF conc (J110)

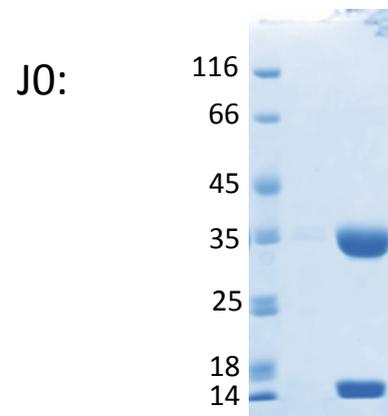
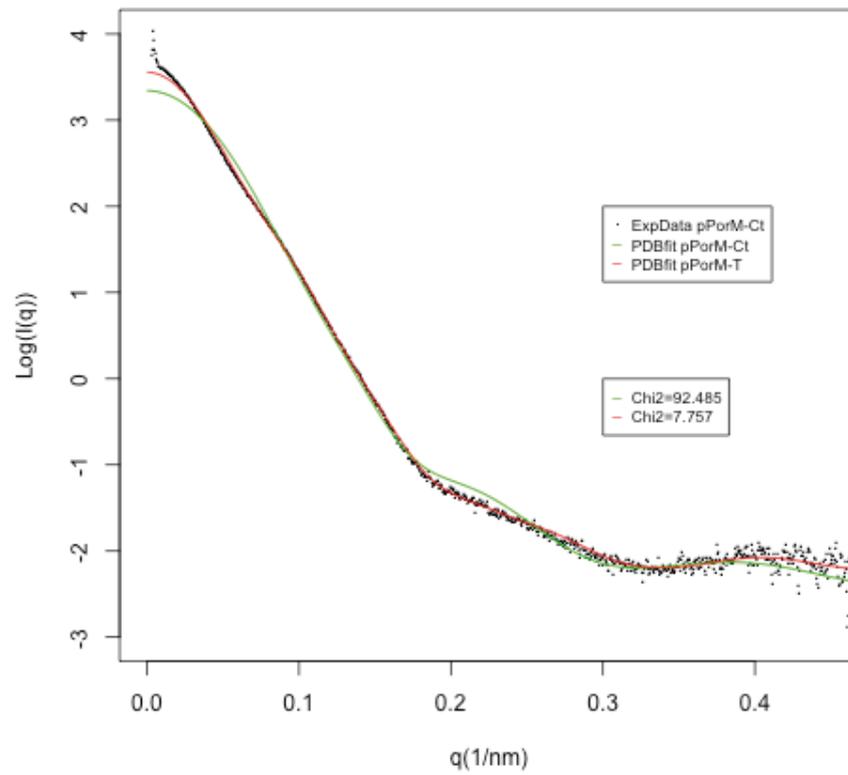


Figure S3

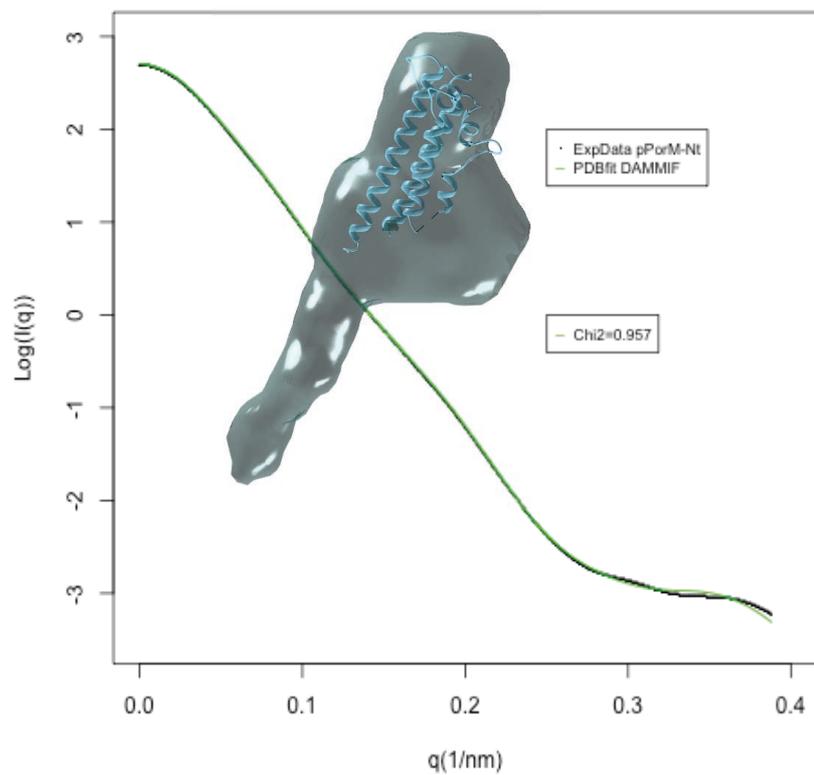
A

Crysol pPorM-Ct



B

DAMMIF pPorM-Nt



3.4 Discussions et conclusions

L'objectif de ce volet de ma thèse était l'étude structurale de la protéine pPorM. Suite aux premiers résultats obtenus sur ce projet avant mon arrivée, l'étude de cette protéine a été divisée en l'étude de deux sous-constructs : le domaine périplasmique N-terminal pPorM-Nt et les domaines suivants C-terminaux pPorM-Ct. Cette dernière construction correspond à la partie de pPorM résultante de la digestion à la Trypsine pPorM-T. Bien qu'ayant contribué à son étude, j'étais en charge de l'étude de la construction pPorM-Nt.

Des nanobody reconnaissant pPorM ont été obtenus après immunisation d'un Lama. Parmi eux, quatre ont été retenus : nb01 et nb02, reconnaissant spécifiquement la région N-terminale de pPorM (pPorM-Nt), et nb130 et nb19, reconnaissant la partie C-terminale (pPorM-Ct et pPorM-T). Les structures de ces quatre nanobody ont été résolues (5), j'ai moi-même résolu celle de nb02 à une résolution de 1,5 Å.

Dans le cadre d'une collaboration avec une équipe de l'IBMC à Strasbourg, j'ai de nouveau cristallisé ce nanobody en puce micro-fluidique ChipX-3. Ce type de puces de cristallisation utilisant la micro-fluidique permet d'éviter tout phénomène de convection et obtenir un cristal plus homogène. L'enregistrement *in-situ* à température ambiante des données a permis de comparer les deux structures de nb02. Curieusement, la seule différence notable entre ces deux structures réside dans une boucle éloignée des CDRs. Alors que cette même boucle avait été compliquée à construire pour la structure du cristal congelé, aucune difficulté n'a été rencontrée dans le cas de la structure à température ambiante, alors que l'on s'attendrait justement à un degré de flexibilité plus important.

L'étude structurale de la protéine pPorM-Nt a d'abord été réalisée par cristallographie aux rayons X, après une caractérisation biophysique et biochimique. Des cristaux ont été rapidement obtenus. Leur diffraction à une résolution de 1,9 Å a permis l'enregistrement d'un jeu de données. Cependant, un twin parfait a été mis en évidence lors du traitement et le jeu n'a alors pu être exploité. Ce domaine de pPorM n'ayant pas de structure connue avec une homologie de séquence suffisante, deux stratégies se présentaient : le calcul de phases à partir de métaux lourds et l'utilisation des phases d'un partenaire en complexe, de structure connue. Les deux méthodes ont été menées en parallèle.

L'obtention de l'information de phases à partir de d'atomes de Sélénium incorporés à des Méthionines n'a pu aboutir, et ce malgré l'obtention d'un jeu de données MAD à 3,5 Å. Ceci est dû à un problème au moment de la recherche du nombre d'atomes de Sélénium pendant le traitement des données, bien que le signal anomal ait été confirmé. Des tests d'optimisation des cristaux ont été réalisés mais ceux-ci restent petits et aucune autre condition de cristallisation n'a été trouvée.

La structure de pPorM-Nt a été résolue grâce à l'utilisation des nanobody comme chaperonne de cristallisation et modèle pour l'information de phases. L'un des cristaux obtenus du complexe pPorM-Nt/nb01 a diffracté au synchrotron à une résolution de 2,3 Å et la structure a pu être résolue à une résolution maximale de 2,5 Å. Les 25 premiers résidus de la protéine n'ont cependant pas pu être construits, dû au manque de densité dans la carte. Les données SAXS obtenues pour cette protéine en solution laissent envisager que cette région manquante est flexible et peut prendre une forme étendue. Cette flexibilité expliquerait en effet bien l'absence de densité électronique dans la carte. Un modèle *ab-initio* a pu être déterminé à l'aide du programme DAMMIF. La valeur du test de χ^2 est autour de 0,96, montrant une cohérence entre la courbe théorique de diffusion du modèle généré avec les données expérimentales en solution. Ce modèle présente une longue extension mais n'est probablement que l'une des représentations de la protéine en solution.

J'ai également mené des expériences SAXS sur les protéines pPorM, pPorM-T et pPorM-Ct. Les résultats obtenus viennent compléter les études structurales cristallographiques déjà réalisées ou, dans le cas de pPorM, apportent des hypothèses sur l'organisation structurale de la protéine.

Concernant la protéine pPorM-T, la question de la validité de l'angle entre les domaines D1 et D2 s'était posée suite à la résolution de la structure par cristallographie. En effet, nous nous demandions s'il ne s'agissait pas là d'un artefact induit par le réseau cristallin. Les données SAXS enregistrées en solution valident la structure cristallographique, avec une valeur du test de χ^2 de 0,97, proche de la valeur correspondant à un accord parfait des données théoriques et expérimentales de diffusion, d'une valeur de 1. Le programme SREFLEX a permis de générer à partir de la structure de pPorM-T des modèles structuraux selon la flexibilité de la protéine et les a comparés aux données SAXS expérimentales. Aucun de ces modèles ne correspond mieux à ces données que la structure d'origine, ce qui a fini de valider la structure cristallographique comme conformère majoritaire en solution de pPorM-T.

La protéine pPorM-Ct correspond aux bornes de pPorM-T déterminées par Spectrométrie de masse et séquençage d'Edman. Elle serait donc potentiellement structurellement proches, ce qui est effectivement le cas pour les domaines D2 et D3. En revanche, la structure cristallographique de pPorM-Ct ne comporte pas le domaine D1, probablement clivé avant la formation du cristal. Les expériences SAXS sur la protéine en solution fraîchement purifiée contiennent, elle, toujours l'information structurale du domaine D1. Ces résultats sont en tout points similaires à ceux de la protéine pPorM-T en solution, autant au niveau du degré de flexibilité que de la forme ou des paramètres dérivés de ces expériences.

Enfin, les expériences SAXS sur la protéine pPorM en solution permettent d'obtenir des informations structurales malgré l'absence de structure cristallographique. Il s'agit d'une protéine allongée, très certainement dimérique et comportant quelques parties flexibles. Le consensus ressortant de la modélisation *ab-initio* confirme une forme allongée de pPorM en forme de Z avec un angle à chaque extrémité. Il est impossible en l'état de discerner une polarité à cette enveloppe et le placement « à la main » des structures de pPorM-T et pPorM-Nt n'est pas concluante.

Les études structurales de la protéine pPorM sont maintenant presque abouties. La résolution des structures cristallographiques des deux sous-constructions pPorM-Ct et pPorM-Nt ainsi que les études complémentaires SAXS donnent une idée de l'agencement globale de cette protéine centrale du T9SS. Cependant, l'assemblage des domaines entre eux reste irrésolu. La modélisation *ab-initio* d'une enveloppe de pPorM à partir des données SAXS était une bonne idée mais la forme obtenue ne permet pas d'attribuer les extrémités N- et C-terminales de la protéine. L'une des solutions pourrait être d'utiliser encore une fois les nanobody afin de « casser » cette symétrie. L'analyse SAXS de complexes pPorM/nb130 ou pPorM/nb01 par exemple devrait permettre d'orienter les enveloppes modélisées et ainsi donner une idée de la forme globale de pPorM.

L'utilisation des nanobody a eu une contribution essentielle à l'étude structurale de la partie périplasmique de la protéine PorM du T9SS. Outre leur fonction de chaperonne de cristallisation, ces anticorps ont servi de modèle pour l'obtention de l'information de phase lors du remplacement moléculaire pendant la résolution des complexes pPorM-Ct/nb130 et pPorM-Nt/nb01. Ils ont ainsi permis soit de stabiliser et aider à la cristallisation de protéines, soit de changer les paramètres du réseau cristallin en établissant de nouvelles conditions de

crystallisation. Par ailleurs, leur fixation sur les régions des protéines qu'ils reconnaissent induit un changement de forme du complexe, ce qui peut aider à la localisation de celles-ci. Dans le cas du SAXS par exemple, la fixation spécifique d'un nanobody sur une région d'une protéine longiligne permet d'apporter une orientation à une enveloppe symétrique.

Les nanobody restent des anticorps pouvant être utilisés pour une application médicale. Cependant, leur utilisation biotechnologique en fait des outils de choix dans l'étude structurale de protéines.

4 Matériels et méthodes

4.1 Description des méthodes utilisées

4.1.1 Clonages

Toutes les constructions réalisées au cours de ma thèse ont été clonées suivant la méthode Restriction Free cloning (RF cloning, Figure 4.1.1). Comme son nom l'indique, cette méthode n'utilise pas d'enzymes de restriction. Cette méthode nécessite deux PCR successives : la première PCR permet d'amplifier la séquence souhaitée, avec l'ajout aux extrémités 3' et 5' d'amorces s'hybridant par la suite sur le plasmide lors de la seconde PCR. Alors que pour la première réaction des amorces doivent être synthétisées, la seconde utilise le produit de la première PCR en ce sens, comme une « méga-amorce ». Lors de cette dernière réaction, la « méga-amorce » s'hybride sur le plasmide d'intérêt. La seconde PCR consiste alors à synthétiser le plasmide avec cette « méga-amorce ». La séquence des amorces correspond donc à une partie s'hybridant à l'ADN à amplifier et une autre partie s'hybridant au plasmide de destination. L'élaboration de chacune doit tenir compte des températures de fusions optimales à une bonne hybridation, comprises entre 60 et 65°C (58).

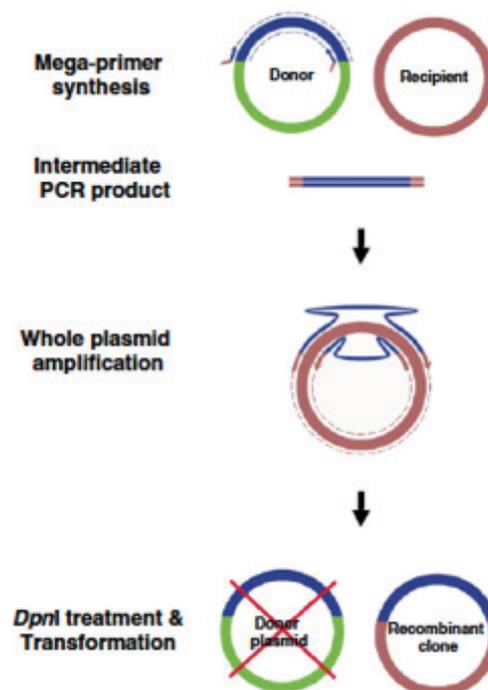


Figure 4.1.1 : Principe du RF cloning (58). L'ADN à cloner (bleu) est amplifié lors d'une première PCR au cours de laquelle sont ajoutées les extrémités s'hybridant au plasmide de destination (rouge). Il sert ensuite d'amorce lors de la seconde PCR. Les plasmides parentaux sont ensuite digérés par traitement à la DpnI et les clones recombinants sont introduits dans des bactéries par transformation.

Le mélange nécessaire à la première réaction comprend l'ADN à amplifier, du tampon de réaction, les dNTP, les amorces sens et anti-sens et l'ADN polymérase. L'amplification est obtenue par plusieurs cycles de dénaturation de l'ADN double brin, d'hybridation des amorces et d'élongation. L'ADN produit est alors vérifié sur gel agarose puis purifié.

Le second mix contient le plasmide de destination, du tampon de réaction, les dNTPs, le produit de la première PCR et l'enzyme.

4.1.2 Dichroïsme Circulaire (CD)

Le CD est une technique permettant l'analyse des structures secondaires (études aux UV lointains) ou tertiaires (UV proches) de biomolécules. Cette spectrométrie d'absorption mesure la différence d'absorbance de la lumière polarisée circulaire gauche et droite par les structures placées en milieu asymétrique ou par des chromophores chiraux. Les mesures prises à des longueurs d'onde comprises entre 180 nm et 260 nm permettent l'étude des structures secondaires. L'analyse des spectres renseigne sur le contenu en hélices α , feuillets β , tours et structures désordonnées de l'échantillon (Figure 4.1.2).

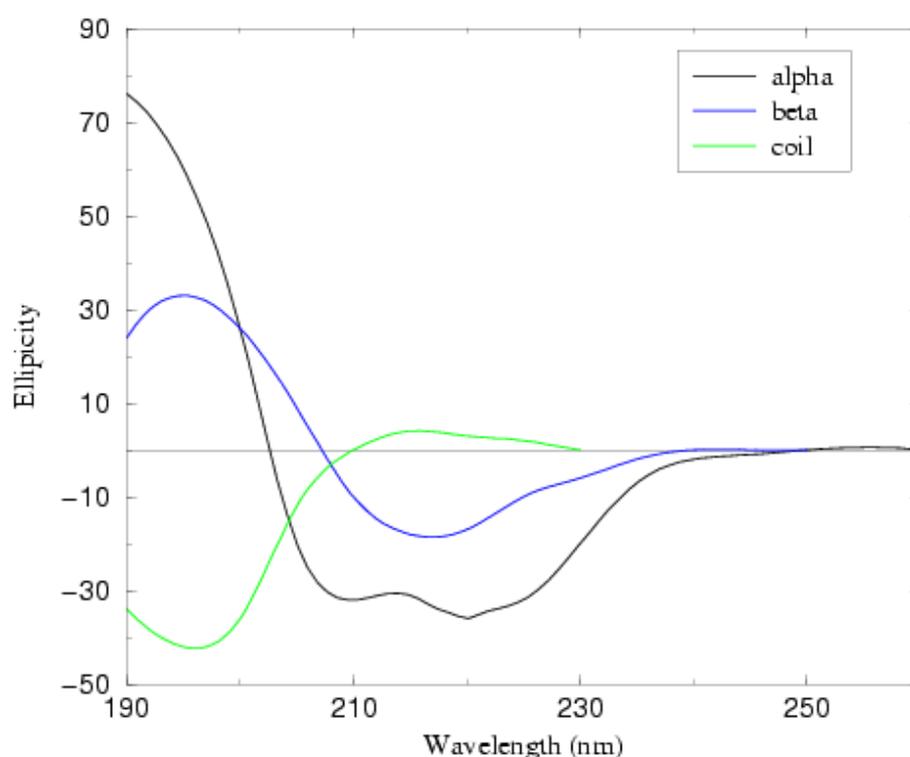


Figure 4.1.2: Courbes de Dichroïsme Circulaire théoriques aux UV lointains(59). La courbe noire représente la courbe théorique d'une molécule constituée d'hélices α , possédant un minimum d'ellipticité à 208 nm et un second à 222 nm. La courbe bleue est représentative d'une molécule composée de feuillets β , avec un minimum d'ellipticité à 214

nm. La troisième courbe (en vert) représente une molécule dépliée. Elle est caractérisée par un maximum d'ellipticité autour de 214 nm.

Le CD peut donc être utilisé pour mesurer des valeurs d'ellipticité moyenne par résidu selon une gamme de longueur d'onde ou selon une gamme de température, à une longueur d'onde spécifique. Dans ce cas, la dénaturation thermique de la protéine permet le calcul d'un T_m en se plaçant à la longueur d'onde spécifique de la structure considérée : pour des hélices α à 222 nm, pour des feuillets β à 214 nm.

4.1.3 Mesures d'interactions protéine/protéine par Bio-Layer Interferometry (BLI)

Les interactions entre protéines peuvent être testées et étudiées à l'aide de différentes méthodes comme par exemple la thermophorèse à micro-échelle MST, la titration calorimétrique isotherme ITC ou bien la Bio-Layer Interferometry (BLI). Chacune de ces techniques permet de calculer et mesurer des paramètres cinétiques différents tels que les constantes d'affinité et d'association/dissociation, l'entropie, l'enthalpie ou la stœchiométrie.

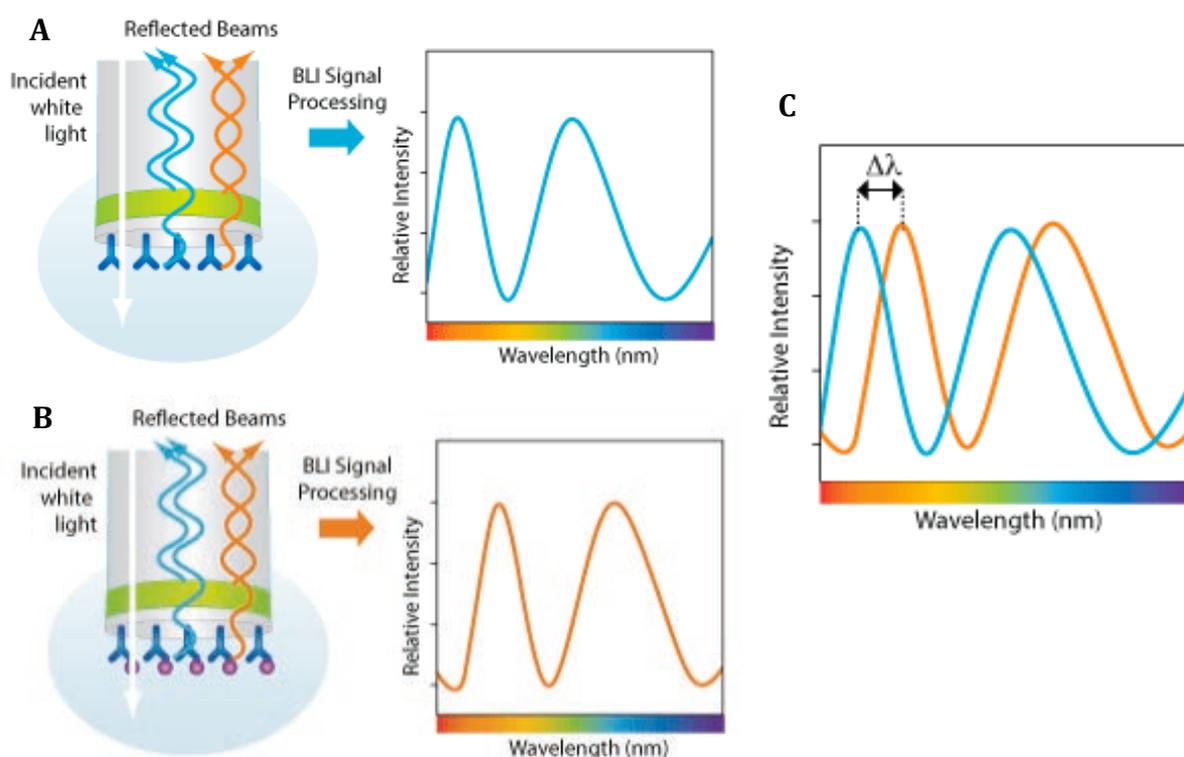


Figure 4.1.3: Principe de la Bio-Layer Interferometry (60). L'un des partenaires de l'interaction à tester est fixé sur la pointe d'un bio-senseur et un rayon de lumière blanche est émis. Les intensités des rayons réfléchis, traversant une couche de protéines fixées et une interne, sont mesurées (A). Lorsque le partenaire est lié aux protéines fixées (B) un décalage des ondes émises est créé (C). Ce signal est ensuite traité pour extraire les paramètres cinétiques.

La BLI est une technologie optique analytique permettant de mesurer l'interférence de la lumière blanche avec deux surfaces : une couche de protéines immobilisées sur la

pointe d'un bio-senseur et une couche de référence interne(60). L'interaction de molécules avec les protéines fixées sur le bio-senseur crée un décalage de longueur d'onde par l'augmentation de l'épaisseur de cette couche, mesuré en temps réel (Figure 4.1.3). Ainsi, des interactions biomoléculaires peuvent être testées et mesurées par calcul des constantes d'affinités, d'association et dissociation. L'un des deux partenaires est fixé sur des bio-senseurs tandis que le second est en solution.

4.1.4 Cristallographie aux rayons X des protéines

La cristallographie aux rayons X est une technique de biochimie structurale consistant à la détermination d'une structure à partir des données de diffraction aux rayons X d'un cristal. Cette technique peut être utilisée dans de nombreux domaines comme la Chimie ou même la Géologie mais seule son application en Biologie Structurale des macromolécules est traitée ici.

Le cristal sert d'amplificateur du signal des macromolécules dont il est constitué. Celles-ci sont organisées en réseau d'un motif identique répété appelé maille. Cette maille est elle-même composée de plusieurs unités asymétriques qui correspondent au plus petit motif permettant la reconstitution complète d'une macromolécule.

Une fois les cristaux obtenus, ceux-ci sont exposés à une source de rayons X pour en enregistrer les données structurales. Ces données peuvent être collectées en utilisant une source de rayons X « maison » de type anode tournante ou de type synchrotron. Bien que le premier type permette tout à fait la détermination d'une structure cristallographique, le développement des sources de lumière synchrotron a eu un impact majeur dans le domaine. L'accélération des électrons dans le vide à une vitesse proche de la vitesse de la lumière permet l'exploitation d'une source de lumière d'intensité et de brillance bien plus importantes que les sources présentent en laboratoire. Les électrons provenant d'un canon à électrons sont accélérés dans un premier anneau dit « booster » jusqu'à atteindre l'énergie de fonctionnement. Le faisceau est ensuite transmis à un deuxième anneau, l'anneau de stockage, permettant par l'utilisation d'aimants de courbure et d'éléments d'insertion de dévier ce faisceau. L'énergie perdue, le rayonnement synchrotron, est alors redirigée et utilisée sur les lignes de lumière expérimentales. Deux sources synchrotron de troisième génération sont disponibles en France : l'European Synchrotron Radiation Facility (ESRF) à Grenoble et la Source Optimisée de Lumière d'Energie Intermédiaire de Lure (SOLEIL) à

Gif-sur-Yvette. Leur développement et renouvellement constant de leur équipement permettent de repousser les limites de résolution à échelle atomique.

Bien que la cristallographie à rayons X des macromolécules comprenne l'étude de composés chimiques, de molécules d'ADN ou même d'ARN, je ne traiterai pour la suite que du cas de l'étude structurale de protéines.

4.1.4.1 Cristallogenèses

La première étape de la cristallographie est l'obtention de cristaux, ou cristallogenèse. La protéine soluble va être mise en contact avec une condition de cristallisation. Par des échanges avec celle-ci et à l'aide d'agents de précipitation, elle va entrer dans un état métastable et former des contacts avec les molécules voisines, créant ainsi plusieurs phases. La formation d'un réseau stable et ordonné donne naissance à un réseau cristallin. Des objets cristallins sont alors formés, appelés nucléis. Cette première étape, l'étape de nucléation, est suivie d'une étape d'association de ces nucléis et de croissance cristalline, jusqu'à l'obtention d'un cristal unique (Figure 4.1.4).

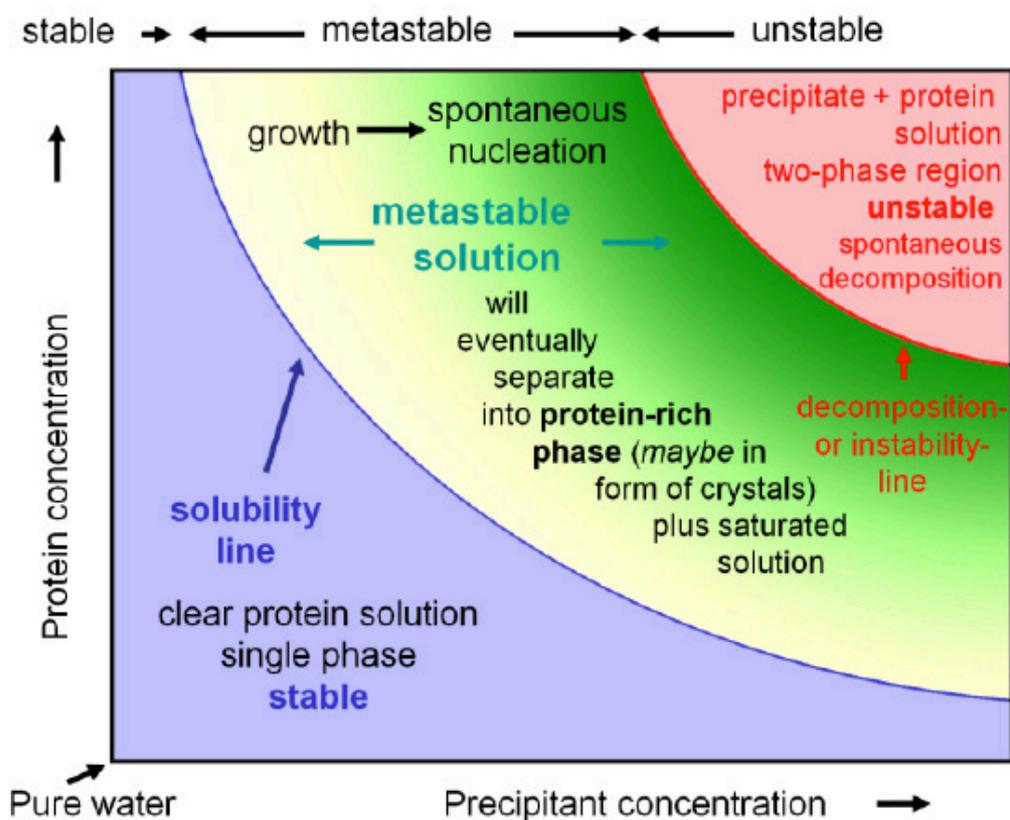


Figure 4.1.4: Les régions fondamentales d'un diagramme de phases de cristallisation de protéines(61). Il existe principalement trois zones : la zone soluble (bleu), la zone métastable (jaune et vert) puis la zone de précipitation (rose). Pour cristalliser, un échantillon protéique doit passer de la zone soluble à la zone métastable (région verte) sans atteindre la zone de précipitation. Les nucléis formés s'assemblent et croissent ensuite jusqu'à retomber dans la région jaune de la zone métastable, jusqu'à l'obtention du cristal unique.

Les méthodes employées pour la cristallisation des protéines sont multiples : diffusion de vapeur, diffusion contraire, microbatch sous gouttes d'huile, dialyse ou bien cristallisation en phase cubique. Au cours de cette thèse, j'ai utilisé les deux premières méthodes.

La diffusion vapeur est une méthode basée sur les échanges vapeurs en milieu hermétique entre une goutte contenant la protéine et un réservoir contenant la condition de cristallisation testée. Les différences de concentration des différents composés de l'échantillon protéique et du réservoir vont induire des échanges vapeur afin de tendre vers un équilibre de ces concentrations. Cela conduit à la favorisation des contacts macromoléculaires au sein de l'échantillon. Il existe deux types de diffusion vapeur : en gouttes assises et en gouttes suspendues. Bien que la première technique soit la plus utilisée, la seconde permet, lors de l'étape d'optimisation des cristaux, l'obtention de cristaux généralement plus gros. Ceci est dû aux volumes plus importants mis en jeu.

La contre-diffusion est une méthode basée sur la diffusion contraire en micro-capillaires de la protéine et de la condition de cristallisation. L'absence de convection ainsi que la mise en place de ces gradients sont les grands avantages de cette méthode : le phénomène de nucléation est plus homogène et une plus large étendue de ratios protéine/condition sont testés. J'ai eu l'occasion de tester des puces microfluidiques de contre-diffusion, fournies par Claude Sauter (laboratoire IBMC, Strasbourg). Ces puces permettent de tester 8 conditions à la fois contre une protéine donnée. Les cristaux obtenus dans les capillaires peuvent être directement exposés aux rayons X sur une ligne synchrotron à température ambiante.

4.1.4.2 Cryoprotection et congélation des cristaux

Afin d'éviter les dommages liés aux radiations et ainsi augmenter la durée de vie des cristaux durant l'exposition aux rayons X, ceux-ci doivent être protégés par congélation à 100°K. Cette protection est devenue d'autant plus nécessaire avec le développement des sources synchrotron, bien plus puissantes et causant donc bien plus de dommages aux cristaux. Mais un autre problème survint alors : la formation de cristaux de glace, due à la congélation de l'eau contenue dans la condition de cristallisation, dans laquelle se trouve le cristal. Outre la contamination du signal de diffraction du cristal de protéine par celle de la glace, la formation de ces cristaux peut endommager le cristal de protéines. Il doit donc être protégé par l'ajout d'un agent cryoprotectant qui va entraîner la formation d'une glace amorphe lors de la congélation. Il vaut tout de même mieux ne pas directement congeler les

cristaux sans tests préalables de la réaction des cristaux à la solution déterminée. En effet, il est possible que cette nouvelle condition endommage ou même dissolvent le cristal. Celui-ci est ensuite pêché, protégé puis congelé par mise en contact avec de l'azote, soit sous forme de flux à 100°K, soit liquide.

4.1.4.3 Collecte des données

Le cristal est exposé à une source de rayons X. Les rayons diffractés sont enregistrés sur un détecteur sous la forme de taches de diffractions, ou réflexions, qui constituent un cliché de diffraction. Chaque tâche de diffraction de cette image représente la somme des contributions de tous les atomes de l'unité asymétrique dans une orientation donnée.

Après avoir monté le cristal sur le goniomètre, permettant de l'orienter entre la source de rayons X et le détecteur sur un axe ϕ , il doit être positionné pour rester centré sur la trajectoire du faisceau au moment de l'exposition. Le centrage peut être manuel ou bien automatisé, comme c'est le cas sur certaines lignes de lumière synchrotron. Il est possible aussi d'opérer un scan du cristal pour déterminer une zone de meilleure diffraction à cibler, sur la ligne ID23-2 de l'ESRF par exemple.

Même s'il est possible de collecter des images en opérant une rotation de 360° du cristal, l'exposition aux rayons X l'endommage. Malgré la cryoprotection, les rayons détériorent à la fois la qualité et la quantité de réflexions détectées, bien que ces dommages soient beaucoup moins importants avec une source « maison ». Seulement deux images, enregistrées à 90° l'une de l'autre, sont nécessaires à la caractérisation des paramètres de maille. Ceux-ci donnent des caractéristiques géométriques renseignant sur les opérateurs de symétrie potentiels des molécules dans le cristal. Ces symétries sont classées sous la forme de groupes d'espaces. Après détermination du groupe d'espace, une stratégie de collecte de données peut être établie.

Le pourcentage d'exposition, le temps d'exposition par image, le degré de rotation par image et entre chaque image, la distance échantillon/détecteur et la longueur d'onde sont les paramètres à définir pour la collecte. Cette collecte peut être centrée sur un même point ou bien réalisée selon une trajectoire hélicoïdale. Dans ce second cas, deux points doivent être définis et caractérisés avant de lancer la collecte.

Pour des projets en remplacement moléculaire, une collecte d'un jeu de données à une longueur d'onde unique est suffisante. Pour des projets de phasage *ab-initio*, des diffuseurs anomaux et/ou des métaux lourds sont introduits dans les cristaux. La procédure de collecte diffère. Un scan d'énergie doit être réalisé afin de déterminer la présence des

métaux dans le cristal. Une fois les métaux identifiés, un spectre d'absorption est enregistré afin de déterminer l'énergie optimale à laquelle se placer pour collecter les données anormales. La courbe obtenue est dérivée deux fois et trois énergies en sont extraites : celle du point d'inflexion du pic, celle du haut du pic et celle après le pic. Le cristal peut alors être collecté à une seule de ces énergies « Single-wavelength Anomalous Diffraction » (SAD) ou à plusieurs « Multi-wavelength anomalous diffraction » (MAD).

4.1.4.4 Traitement des données

Les images collectées contiennent un certain nombre de réflexions. Dans un premier temps les réflexions d'un échantillon d'images sont détectées et analysées. Cette étape est appelée l'indexation. Selon les possibilités de paramètres de maille, des réflexions non visibles sont extrapolées et des groupes d'espaces sont proposés et triés selon un score croissant de pénalité. Les groupes d'espaces de plus faible pénalité sont proposés en premiers par ordre croissant de degré de symétrie puis survient un gap dans la gamme de pénalité. Au delà de ce gap sont proposés des groupes d'espaces peu probables. Le meilleur groupe d'espace est généralement celui offrant le score de plus haute pénalité avant ce seuil tout en ayant la plus haute symétrie. Lorsqu'un groupe d'espace est sélectionné, les paramètres de diffraction sont affinés. Vient ensuite l'intégration des réflexions, c'est à dire la redéfinition de chaque tache dans les trois dimensions et l'évaluation de leur intensité. Chacune de ces tâches est définie selon les indices de Miller h , k et l déterminant une position dans l'espace réciproque selon les axes a , b et c respectivement. Les intensités de ces réflexions sont alors corrigées et les statistiques des données collectées sont calculées. C'est à cette étape que peut être estimée la qualité du jeu de données et la résolution maximale d'exploitation de celles-ci. Une fois les données traitées dans le groupe d'espace le plus probable et à la résolution maximale la plus raisonnable aux vues des statistiques, des programmes tels que POINTLESS permettent de vérifier la validité du groupe d'espace. Les paquets d'images sont alors vérifiés selon certains paramètres incluant la fraction de twin des données. Les observations des réflexions sont ensuite mises à l'échelle et leur intensité moyennée en utilisant des programmes de mise à l'échelle tels que SCALA. C'est également à cette étape qu'une partie des réflexions est marquée afin, plus tard, de ne pas être prises en compte dans l'affinement de la structure. Ces réflexions permettent de calculer le R_{free} et d'estimer le biais du modèle dans la valeur du facteur R . Le fichier de sortie de ce type de

programme, de format .mtz, contient les indices hkl intégrées au bon groupe d'espace, avec toutes les statistiques de collecte et traitement des données.

4.1.4.5 Détermination de la structure et affinement du modèle.

Le cristal de protéine exposé aux rayons X diffracte en émettant des ondes enregistrées par le détecteur. Ces ondes, contenant l'information structurale, sont caractérisées par une amplitude et une phase. Afin de reconstituer cette information, ces deux paramètres sont nécessaires. Seulement, bien que l'information d'amplitude ait été conservée au travers des intensités des réflexions, l'information de phase a été perdue. Il existe deux moyens de résoudre ce problème de phases : l'utilisation de l'information de phase d'un homologue structural, appelée remplacement moléculaire, et la détermination de cette information de phase au sein même de la structure à partir d'un signal isomorphe ou anomal, appelée phasage expérimental. Je ne traiterai ici que du remplacement moléculaire.

Le remplacement moléculaire utilise les phases d'un modèle d'homologie structurale significative avec la structure à déterminer. Il est admis qu'une homologie de séquence inférieure à 30% est insuffisante. Il existe différents programmes permettant le remplacement moléculaire tels que Molrep, AMoRe ou Phaser. Chacun fonctionne à partir d'un modèle structural et du fichier de réflexions mises à l'échelle. Le nombre de molécules par unité asymétrique doit également être indiqué. Celui-ci doit être estimé au préalable à partir du calcul du coefficient de Matthews. Ce coefficient permet aussi le calcul des pourcentages de protéines et de solvant dans l'unité asymétrique. Ces programmes de remplacement cherchent et proposent des solutions de la structure. L'introduction de clashes structuraux, la cohérence des contacts cristallins dans la maille et les facteurs R d'accord des données observées et calculées (R_{work} et R_{free}) doivent être vérifiés avant de valider une solution.

La solution constitue un modèle qu'il faut modifier en s'appuyant sur des cartes de densités électroniques. Ces cartes, $F_o - F_c$ et $2F_o - 2F_c$, sont générées à la suite d'un affinement. Plusieurs programmes d'affinement sont disponibles, dont Refmac5 (suite CCP4), Phenix.refine (suite Phenix) et AutoBUSTER (Global Phasing). Tous ces programmes proposent de nombreuses options d'affinement, classiques, comme le nombre de cycles, la gamme de résolution ou l'ajout de molécules d'eau, ou bien plus spécifiques telles que l'utilisation de groupes TLS, des NCS ou la prise en compte d'une loi de twin. Les cartes de densité électronique obtenues après affinement permettent de modifier la structure

pour le cycle suivant à l'aide d'un programme tel que COOT. Plusieurs cycles d'affinement/construction sont nécessaires pour l'obtention d'un modèle répondant à des statistiques cohérentes.

4.1.4.6 Validation de la structure.

Une fois que le modèle a atteint des facteurs R_{work} et R_{free} satisfaisants (environ 10 fois inférieurs à la résolution maximale de la structure), la structure doit être validée selon différents paramètres : la présence de clashes, la probabilité des rotamères, la géométrie du modèle (angles des chaînes latérales, liaisons, angles selon le plot de Ramachandran), les molécules d'eau etc... Des outils introduits dans les programmes de construction peuvent aider à cette validation, tout comme certains serveurs en ligne comme MolProbity (<http://molprobity.biochem.duke.edu/>). Ces mêmes étapes de validations sont utilisées dans le processus de dépôt des structures dans la banque de données RCSB PDB.

4.1.5 « Small Angle Light Scattering » (SAXS)

La diffusion des rayons X aux petits angles, ou SAXS, est une méthode d'étude structurale des protéines en solution. Elle permet donc, contrairement à la cristallographie, d'étudier des systèmes flexibles, dynamiques en solution et non contraints dans une conformation donnée et cela sans besoin de marquage, contrairement à la Résonance Magnétique Nucléaire (RMN). Il s'agit d'une technique de basse résolution en comparaison avec la cristallographie aux rayons X (autour de 20 Å) : les informations portent principalement sur la forme globale du système, son état d'oligomérisation, son rayon de giration et la présence éventuelle de différentes espèces en solution. Cette technique apporte donc des informations structurales complémentaires aux autres techniques citées plus haut.

4.1.5.1 Collecte des données

Rares sont les laboratoires à posséder l'équipement nécessaire à des expériences SAXS « maison ». Le plus couramment, la source utilisée est de type synchrotron avec des lignes de lumière telles que BM29 (BioSAXS, ESRF) ou Swing (SOLEIL).

L'expérience nécessite un échantillon pure, homogène, dans une gamme de concentration allant de 0,5 mg/ml à 10-15 mg/ml. Cet échantillon peut être directement chargé dans le capillaire et exposé aux rayons X ou bien chargé sur une colonne d'exclusion

de taille branchée en amont du capillaire (SEC-SAXS). Dans le premier cas, différentes concentrations de la protéine doivent être préparées avant l'expérience. Une fois sur place, les échantillons doivent de nouveau être centrifugés pour s'assurer de l'absence d'agrégats et les concentrations sont prises à nouveau. Chaque échantillon est déposé dans des tubes ou dans une plaque et le tout est placé dans le chargeur d'échantillon. Le tampon exact de la protéine doit également y être chargé pour servir de blanc lors de l'acquisition. Dans le second cas, l'acquisition des données est faite dès la sortie de la colonne et une seule injection par protéine, de haute concentration, est nécessaire. La gamme de concentration est réalisée naturellement tout le long du pic. L'autre avantage est la séparation des différentes espèces et agrégats en solution : seul le pic d'intérêt peut être étudié, sans contamination, et le tampon d'élution est directement utilisé comme blanc d'acquisition. Les intensités diffusées sont mesurées en fonction des vecteurs de diffusion (q ou s selon les lignes).

4.1.5.2 *Traitement des données et analyses.*

Les données enregistrées se présentent sous la forme de courbes de diffusion (Figure 4.1.5A). Différentes informations peuvent être obtenues selon la région de la courbe étudiée.

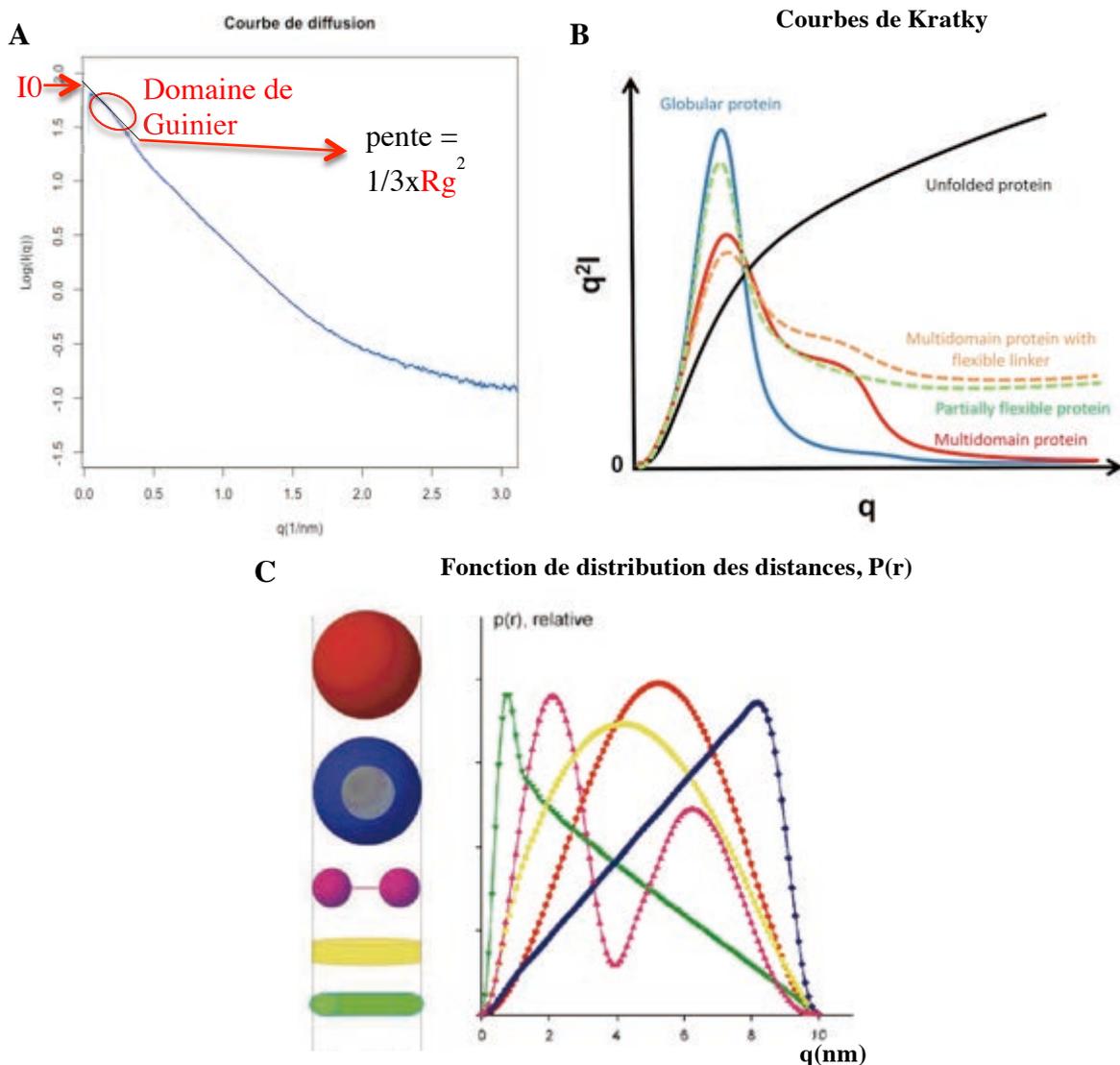


Figure 4.1.5: Courbes théoriques de diffusion et d'analyses SAXS. **A :** Courbe de diffusion. Le signal de diffusion du tampon a été soustrait. Les intensités diffusées I sont tracées en fonction des vecteurs diffusés q . La zone encadrée en rouge représente le domaine de Guinier, utilisé pour l'estimation du rayon de giration (R_g) et pour l'extrapolation de I à $q=0$ (I_0) à l'aide de la droite (en noir). **B :** Courbes de Kratky(62). L'allure de la courbe renseigne sur le niveau de flexibilité de la protéine en solution. **C :** Fonction de distribution des distances par paires (courbes de Porod $P(r)$)(63). L'allure de la courbe renseigne cette fois sur la forme de la protéine en solution avec une représentation de la répartition des distances. La distance maximale D_{max} est la valeur de la distance r lorsque la courbe redescend sur cet axe. Les couleurs des courbes correspondent à celle des formes à gauche.

L'information du rayon de giration peut être extraite des premiers points de la courbe (les 100 premier environ) en calculant la pente d'une droite les alignant. Une valeur d'intensité diffusée I peut également être extrapolée pour une valeur du vecteur de diffusion $q=0$. Ces premières analyses sont définies comme analyse de Guinier. La possibilité d'aligner une droite sur ces points renseigne sur l'état de l'échantillon : agrégé ou non. Le contenu en flexibilité de l'échantillon peut être étudié par le tracé du plot de Kratky (Figure 4.1.5B). Cette courbe, représentant les valeurs « $I \cdot q^2$ » en fonction de q^2 , voit son allure modifiée selon que le système soit stable, contiennent quelques régions flexibles ou alors qu'il soit

désordonné ou déplié. L'analyse de la fonction de distribution des distances par paires donne, elle, une information de forme (Figure 4.1.5C). Pour une protéine globulaire, cette courbe prend une allure en cloche, tandis qu'elle comprendra plusieurs pics pour une protéine multi-domaine. La distance maximale D_{\max} de la protéine peut également être estimée à partir de cette analyse des distances, représentant la distance extrapolée à $P(r)=0$.

Il est courant d'analyser les données obtenues par SAXS non pas en étudiant les résultats d'une seule concentration mais en combinant une faible et une forte concentration. Une mesure SAXS sur une faible concentration apporte de l'information précise sur les enregistrements à très faible résolution, autrement dite sur la région de Guinier. La haute résolution n'est pas accessible car trop bruitée. En revanche, les mesures à forte concentration sont moins précises dans la région de Guinier mais très résolutive et donc précises sur le facteur de forme. Une courbe chimère peut être réalisée en moyennant deux courbes sur une région couvrant quelques dizaines à une centaine de points, juste après la région de Guinier.

Différentes méthodes d'analyses existent selon l'information souhaitée et le système étudié. Les programmes tels que DAMMIN et DAMMIF modélisent « ab-initio » des enveloppes, en se basant sur la valeur de D_{\max} , tandis que d'autres programmes comme CRY SOL se basent uniquement sur les courbes de diffusion expérimentales brutes. Il existe aussi des programmes adaptés à l'étude de complexes (ex : CORAL ou SASREF) ainsi que des programmes de modélisation et déconvolution d'ensemble conformationnels (ex : EOM).

4.2 Etudes structurales d'anticorps thérapeutiques contre protéines MIC

4.2.1 Fab MIC12 : expression et purification

Le Fab MIC12 a été exprimé en cellules de mammifères Chinese Hamster Ovary (CHO) en lignée stable. Les cellules ont été fournies stabilisées en culture par Innate Pharma et j'ai maintenu ces cellules en culture en milieu CD CHO (Gibco) supplémenté de 2 mM de L-Glutamax, 10% de CD-CHO efficient feed B (Gibco), 10 μ M de Na Pyruvate (Gibco) et d'Hygromycine à 250 μ g/mL. Les cellules ont été cultivées à 32°C en roller et maintenues en cultures avec une concentration contrôlée autour de 2 Millions par mL. Lorsque le volume approchait les 1,5 L, la concentration a été laissée augmenter jusqu'à 6 à 7 Million de cellules par mL. Celles-ci ont alors été centrifugées à 1000 rpm pour récolter le surnageant contenant la protéine sécrétée. Si la mortalité n'était pas encore trop importante,

les cellules étaient diluées de nouveau afin de poursuivre les productions. De l'azide à 0,02% final et 1mM de PMSF ont été ajoutés au surnageant qui peut être stocké quelques jours à 4°C ou directement purifié.

Des lots de surnageant contenant le Fab ont régulièrement été récoltés puis purifiés. Une première étape de purification a été réalisée sur colonne de chromatographie d'affinité HisTrap Ni²⁺ (GE Healthcare), pré-équilibrée en tampon de lavage 50 mM de Tris pH 7,5, 300 mM de NaCl et 10 mM d'Imidazole. La construction du Fab MIC12 contenant une étiquette HisTag à son extrémité C-terminale, il est retenu par l'interaction de cette étiquette avec le nickel fixé sur les billes de la colonne, tandis que les autres protéines passent sans être retenues. Le Fab a été chargé sur la colonne puis élué dans le même tampon contenant 250 mM d'Imidazole. Après concentration du pic élué avec un concentrateur Amicon® Ultra (Millipore) de cut off 30 kDa jusqu'à un volume inférieur à 10 mL, la protéine a été centrifugée puis chargée sur une colonne de chromatographie d'exclusion de taille Superdex 200 26/60 (GE Healthcare). Le tampon utilisé pour cette étape de purification contient 10 mM de Hepes à pH 7,3 et 150 mM de NaCl. Les protéines sont éluées selon leur poids moléculaire et leur taille. Une fois le Fab élué et les fractions du pic collectées, il peut être conservé plusieurs mois à 4°C.

4.2.2 Protéines MIC exprimées en système bactérien

4.2.2.1 Clonage

Les protéines MICA*008 et MICB*002 ont été clonées dans le plasmide pLIC06 (fourni par BioXtal) à partir des ADN complémentaires fournis par Innate Pharma. Ce plasmide, dérivé du plasmide d'expression pET28 a+ (Novagen), a été fourni par la société BioXtal. Il permet l'expression d'une protéine recombinante avec une étiquette HisTag en C-terminal. Il comprend une résistance à l'antibiotique kanamycine avec un promoteur T7, permettant une induction à l'IPTG. La taille de chaque insert est d'environ 830 pb pour une taille totale de ces constructions (plasmide plus insert) d'environ 6150 pb.

	Amorces	Séquences
	MIC- pLIC06F	TTTAAGAAGGAGATATACCATGGAGCCCCACAGTCTTCGT
MIC	MIC- pLIC06R	GGTGATGATGGTGGTGTTCCTCCAGAGGGCACAG
	MIC-pDCE- HisF	CGGCGGATCCCGGGGAGCCCCACAGTCTTCGTTATAACC
	MIC-pDCE- HisR	CCTCAATGGTGGTGGTGGTGGTGTTCCTCCAGAGGGCACAGGG
	MIC-pDCE- StrepR	ACCGCCGGCCTGGCCTCATTTTTTCGAACTGCGGGTGGCTCCA TTTCCAGAGGGCACAGGG
T9SS	pPorM-NtF	CCGAGAACCTGTACTTCCAATCAAGCTTAACCTCTTCCATCGACG
	pPorM-NtR	CGGAGCTCGAATTCGGATCCTTATTAGCTCTTTACCAGATCGGCAAG
	pPorM-CtF	CCGAGAACCTGTACTTCCAATCAGTGAACAGTATCACGGCACAAG
	pPorM-CtR	CGGAGCTCGAATTCGGATCCTTATTAGTTCACAATTACTTCAATGGC

Tableau 4.2.1: Amorces utilisées pour le clonage des constructions MIC et T9SS.

Les clonages ont été réalisés en appliquant la technique du RF cloning, décrite en 4.1.1. Les amorces utilisées sont reportées dans le Tableau 4.2.1. Le mélange de la première réaction comprenait 20 ng d'ADN, du tampon de réaction 5X, 200 µM de chaque dNTP, 0,5 µM de chaque amorce et 1 unité d'ADN polymérase PHUSION. Le programme utilisé comprend 25 à 30 cycles de 10s de dénaturation à 98°C, 10s d'hybridation à 60°C et 30s/Kb d'élongation à 72°C, suivis d'une étape finale de 10min à 72°C pour permettre l'élongation des brins incomplets. L'ADN amplifié a été vérifié sur gel agarose 1% par comparaison de sa migration au marqueur de taille 1Kb. L'ADN a été purifié à l'aide d'un kit NucleoSpin® gel and PCR clean up (Macherey-Nagel). Le second mélange réactionnel contenait 20 ng de plasmide pLIC06, du tampon de réaction 5X, 200 µM de chaque dNTP, 100 ng de produit de la première PCR et 1,5 unité de l'enzyme PHUSION. La réaction a été faite suivant 20 cycles de 30s de dénaturation à 95°C, 1min d'hybridation à 60°C et 30s/Kb d'élongation à 72°C, le tout suivi d'une dernière étape à 72°C pendant 10min. Une fois la construction obtenue et vérifiée sur gel d'agarose 1%, l'ADN méthylé a été digéré par ajout de 10 unités DpnI à 26°C pendant 2h. Le produit PCR non digéré a ensuite été purifié à l'aide du kit NucleoSpin® gel and PCR clean up.

Des cellules DH5 α ont été transformées par cet ADN dans un ratio volumique 10:1 par choc thermique (30min à 4°C, 45s à 42°C puis 2min à 4°C). Après ajout de 500 μ L de SOC, les cellules ont été incubées à 37°C sous agitation pendant 45min puis étalées sur boîte LB agar contenant 50 μ g/mL de kanamycine. Les boîtes ont été incubées à 37°C toute la nuit. Une fois les colonies bactériennes développées sur boîte, l'insertion correcte de l'ADN a été vérifiée par screen PCR. Pour cela, une colonie bactérienne a été piquée et déposée dans 10 μ L de LB. Le mélange utilisé est composé du tampon de réaction, 200 μ M de chaque dNTP, 0,5 μ M de chaque amorce, de 1,5 mM de MgCl₂, de 2 μ L de la colonie et 0,5 unité de polymérase Taq. L'amplification comprend 29 cycles de 30s de dénaturation à 94°C, 45s d'hybridation à 55°C et 1min/Kb d'élongation à 72°C avec une étape finale à 72°C de 10min. Les produits de cette PCR ont été analysés sur gel agarose 1%.

Des pré-cultures de quelques colonies positives en volume de 4mL de LB plus kanamycine ont été incubées à 37°C toute la nuit. Les ADN plasmidiques ont alors été extraits et purifiés à l'aide d'un kit NucleoSpin® plasmid (Macherey-Nagel) et les séquences ont été vérifiées par la société GATC Biotech.

4.2.2.2 Expression

Les protéines MICA*001, MICA*008 et MICB*002 sont exprimées en cellules Rosetta (DE3) pLysS (Novagen). Le plasmide d'expression de MICA*001 diffère dans sa résistance aux antibiotiques : il comporte un gène de résistance à l'ampicilline et non pas à la kanamycine.

Les cellules ont été transformées par choc thermique avec l'ADN séquencé. Après pousse des colonies sur boîte, des pré-cultures en milieu LB plus antibiotiques ont étéensemencées avec puis mises à incuber à 37°C toute la nuit. Les cultures ont été réalisées à 37°C en milieu TB plus antibiotiques,ensemencé avec 25mL de pré-culture pour 1L de culture. Une fois que la DO_{600nm} a atteint une valeur située entre 0,6 et 0,8, la production de la protéine a été induite avec 0,5 mM d'IPTG à 17°C pendant 20h. Les cellules ont ensuite été centrifugées à 8000g pendant 8min et les culots bactériens congelés à l'azote liquide puis stockés à -20°C.

4.2.2.3 Extraction et repliement de protéines contenues dans les corps d'inclusion (protocole optimisé)

À la différence des protéines exprimées dans la fraction soluble des bactéries, les trois protéines MIC s'expriment en corps d'inclusion. Les cellules bactériennes, culotées et congelées, ont été dégelées et resuspendues en tampon de lyse (10mL/Unité DO/Litre de culture) 50mM de Tris pH8, 300mM de NaCl, 10mM d'Imidazole, 0,1% de Triton, 0,25mg/mL de lysozyme, 1mM de PMSF, 10µg/mL de Dnase et 20mg/mL de MgSO₄. Après 40 min d'incubation à 4°C sur roue, les cellules ont été lysées par sonication sur glace (pour 30 mL : 3 cycles de 30s de pulses puis 30s de repos). Les débris cellulaires, membranes et corps d'inclusion ont ensuite été culotés par centrifugation 30min à 12000g. Le culot a été lavé suivant 3 cycles de 2 lavages en tampon 50 mM de Tris pH8, 300 mM de NaCl, 1M d'Urée, 25 mM d'Imidazole avec ou sans 0,1% de Triton, chaque lavage étant suivi d'une centrifugation à 12000g pendant 30min à 4°C. À la fin de la dernière centrifugation, le culot a été repris dans 2 mL de tampon de dénaturation constitué de 25 mM de Tris pH8, 150 mM de NaCl, 6 M de Guanidine, 15 mM de β-mercaptoethanol et de 1% de glycérol. Il a ensuite été incubé sur roue 48h à 4°C. La fraction dénaturée a été centrifugée à 11000g pendant 30 min à 4°C et la quantité de protéine dénaturée contenue dans le surnageant a été estimée d'abord par mesure de l'absorbance au NanoDrop 2000 (Thermo Scientific) puis par dépôt sur SDS-PAGE. La mesure de l'absorbance permet d'avoir une première approximation de la concentration protéique, qu'une quantification sur gel SDS avec un dépôt de la même quantité estimée de BSA permet de confirmer. La protéine dénaturée a été ensuite diluée dans un tampon de repliement composé de 50 mM de NaH₂PO₄, 500 mM d'Urée, 400 mM de L-arginine, 150 mM de NaCl et d'un couple de glutathion oxydé/réduit de 0,5/5 mM, suivant un ratio 20 mg de protéine dénaturée pour 100 mL de tampon. Pour cela, la protéine a été injectée à la seringue directement dans un vortex avec ensuite diminution drastique de la vitesse d'agitation. Après 24h à 4°C, les agrégats ont été éliminés par centrifugation à 11000g pendant 30min à 4°C. Le surnageant a été dialysé pendant 48h à 4°C contre un premier tampon de dialyse (tampon de repliement sans Urée) dans un ratio volumique de 1:5 en protéine/tampon. Le tampon de dialyse a ensuite été changé pour le même tampon sans L-arginine et est laissé 48h à 4°C. Après une dernière dialyse de 48h à 4°C en tampon 50 mM de NaH₂PO₄ et 150 mM de NaCl, le contenu du boudin a été centrifugé à 11000 g 30min à 4°C pour éliminer les agrégats.

4.2.2.4 Purification des protéines MIC repliées

Bien que les étapes de lavages de culot aient éliminé beaucoup de contaminants, la fraction protéique en sortie de dialyse n'est pas encore pure et des étapes de purification sur chromatographie d'affinité et colonne d'exclusion de taille sont nécessaires. Ces étapes permettent surtout un contrôle de la qualité du repliement ainsi qu'une séparation des espèces de différents niveaux de repliement. Les protéines MIC portent toutes des étiquettes HisTag, ayant une grande affinité pour les ions Ni^{2+} .

Les fractions protéiques ont été chargées sur une colonne de chromatographie d'affinité Ni^{2+} HiTrap 5mL (GE Healthcare), pré-équilibrée en tampon 50 mM de NaH_2PO_4 et 150 mM de NaCl contenant de l'Imidazole à 10mM. Après une étape de lavage de la colonne avec ce même tampon, la protéine surexprimée a été éluée avec 30 mL de tampon contenant 250 mM d'Imidazole. La purification étant réalisée sur un système AKTA, le signal $\text{UV}_{280\text{nm}}$ permet le suivi de l'éluion de la protéine. Le pic élué de la protéine a ensuite été quantifié par mesure de son absorbance à l'aide d'un NanoDrop 2000 (Thermo Scientific) et concentré si nécessaire afin de réduire le volume.

Chaque fraction récoltée au cours de la purification a été déposée sur SDS-PAGE pour analyse sur gel.

4.2.2.5 Purification des complexes Fab/protéines MIC repliées

L'interaction formée par le Fab MIC12 avec les protéines MIC a été réalisée après la chromatographie d'affinité. Un ratio molaire 1:1,2 est appliqué (excès de Fab) et l'incubation a été faite à température ambiante pendant 15 min. Après centrifugation à 11000g pendant 10min à 4°C, la fraction protéique a été injectée sur une colonne Superdex 75 16/60 (GE Healthcare) équilibrée en tampon Hepes 10 mM pH 6,8, NaCl 150 mM. Les protéines étant éluées selon leur poids moléculaire et leur taille, le pic de complexe a été élué en premier, celui du Fab en excès en second. Le premier pic a été récolté. En cas de chevauchement de ces pics, un second passage sur colonne Superdex 200 16/60 (GE Healthcare) a été réalisé.

Concernant le complexe MICA*001/MIC12, une dernière étape de purification a été réalisée afin de séparer différents isomorphes. Le complexe a été chargé sur une colonne

d'échangeuse d'ion cations Resource S. Un gradient de NaCl allant d'environ 50 mM à 1M est utilisé, avec un tampon contenant 10 mM d'Hepes à pH 7.

4.2.3 Protéines MIC : expression en système baculovirus

4.2.3.1 Clonages

Les protéines MICA*001, MICA*008 et MICB*002 ont également été clonées dans le plasmide pDCE 213 afin de permettre une expression en système baculovirus en cellules d'insectes. En effet, ce plasmide est un pFastBac like, nécessaire pour l'obtention de bacmides après transformation de cellules DH10Bac (Invitrogen). Pour chaque protéine, deux constructions ont été réalisées : une comprenant une étiquette HisTag et l'autre une étiquette Strep, toutes deux situées en C-terminal. Ces constructions contiennent un gène de résistance à l'antibiotique gentamycine avec un promoteur T7. La taille de chaque insert est d'environ 830 pb pour une taille totale de ces constructions (plasmide plus insert) d'environ 3800 pb.

Les amorces utilisées sont reportées dans le Tableau 4.2.1 et le protocole de clonage utilisé est identique à celui décrit plus haut.

4.2.3.2 Obtention des bacmides

Après vérification des séquences de chaque construction, des cellules DH10Bac ont été transformées avec les ADN purifiés des deux constructions de MICA*001 et MICB*002 et la construction avec HisTag de MICA*008. Un choc thermique a été réalisé puis 900 µl de milieu SOC ont été ajoutés. Les mélanges ont été incubés à 37°C sous agitation pendant 4h. Plusieurs dilutions de chaque culture ont été faites en utilisant le milieu SOC puis étalées sur boîtes LB agar contenant 50 µg/mL de kanamycine, 7 µg/mL de gentamicine, 10 µg/mL de tétracycline, 100 µg/mL X-gal et 40 µg/mL d'IPTG. Les boîtes ont été incubées à 37°C pendant 48h. Un criblage « blanc-bleu » a ainsi été réalisé, avec l'apparition de colonies blanches positives (contenant la construction) et d'autres bleues, négatives. Plusieurs colonies des deux couleurs ont été analysées par screen PCR et les ADN ayant une migration sur gel agarose 1% cohérente avec la taille attendue ont été envoyés au séquençage.

4.2.3.3 Obtention des baculovirus

Des cellules d'insectes Sf9 et Hi5 semi-adhérentes sont cultivées en routine pour permettre l'obtention des baculovirus et l'expression des protéines. Pour cela, une ampoule de cellules, généralement concentrée à environ 20 Million de cellules par mL et stockée à -80°C, a été décongelée dans du milieu Insect Xpress et diluée à 0,5 Million de cellules par mL dans une flasque à 27°C. Lorsque les cellules recouvrent la totalité de la surface de la flasque en une monocouche, elles sont transférées en fiole pour permettre une culture en suspension, sous agitation à 27°C. Afin de permettre une bonne aération, un volume de 1:3 maximum (volume de cellules : capacité volumique de la fiole) est maintenu.

Une fois les séquences des bacmides vérifiées, celles-ci ont été introduites par transfection dans des cellules d'insectes *Spodoptera frugiperda* semi-adhérentes Sf9.

Le plaquage des cellules a été réalisé dans une plaque format 6 puits avec, par puit, 2 mL de milieu GRACE sans sérum contenant 0,8 Million de cellules. Les cellules ont été maintenues à 24°C le temps de l'adhésion au fond de la boîte tandis que le mélange de transfection a été préparé à part dans un tube eppendorf. Un volume de 100 µL de milieu GRACE a été ajouté sur 1 à 1,5 µg de bacmide. Dans un second tube, le même volume de milieu a été ajouté à 8 µL de cellfectine puis le mélange a été vortexé avant d'être introduit dans le tube contenant le bacmide. Après l'avoir tapoté, le mélange a été incubé 30 min à température ambiante sous une hôte à flux laminaire puis déposé sur les cellules plaquées. La plaque a été incubée à 24°C pendant 3 à 5h, puis le milieu de chaque puit a été remplacé par 2 mL de milieu SF 900. La mortalité des cellules a été suivie par comptage des cellules au microscope à l'aide d'une cellule de Malassez. Lorsque celle ci a atteint les 100% de mortalité après environ une semaine, le contenu de chaque puit a été centrifugé et le surnageant a été analysé par Western Blot.

Après dépôt sur gel dénaturant des échantillons et migration, le transfert a lieu sur une membrane PVDF préalablement activée dans de l'éthanol 100% puis rincée à l'eau pure. Cette membrane a ensuite été équilibrée en tampon de transfert « one step semi-dry » (Pierce) puis insérée dans le montage suivant : 2 papiers filtres, membrane, SDS-PAGE, 2 papiers filtres. Le tout a été imbibé de tampon de transfert puis déposé sur un système de transfert semi-sec Fast Semi-Dry Blotter (Pierce). Le transfert a été réalisé à 15 V constant pendant 50 min à température ambiante. La membrane a alors été bloquée pendant 1h sous agitation avec 20 mL d'une solution de PBS contenant 5% de lait. Après 3 lavages avec 20 mL de tampon PBS-Tween20, l'anticorps couplé à la « horse radish peroxidase » (HRP),

dilué dans du PBS-Tween20, a été ajouté sur la membrane et l'incubation a été faite à température ambiante sous agitation pendant 1h. Après 2 lavages de 20 mL en PBS-Tween puis 2 autres en PBS, la membrane a été révélée par incubation avec 1 mL de Luminata™ Crescendo Western HRP Substrate dilué par 2 pendant 1 min 30s puis exposition à l'aide d'une station image (Kodak) équipée d'un filtre 490-580nm. Le temps d'exposition est ajusté selon le contraste signal/bruit.

Les virus P1 ont ainsi été obtenus.

Pour augmenter la charge virale, deux cycles d'infections ont été réalisés. Généralement, le nombre de cycle n'est pas supérieur à deux afin d'éviter l'introduction de mutations dans le virus. Pour la première infection, 20 Millions de cellules plaquées dans une flasque ont été infectées avec 1 mL de virus P1. Lorsqu'une mortalité de 100% a été atteinte, une fraction du surnageant a été analysée par Western Blot. Le virus obtenu est appelé virus P2. Afin de déterminer le taux optimum d'infection, le virus a été titré par infection de cellules *Trichoplusia ni* Hi5. Pour cela, des cellules d'insecte Hi5 ont été infectées à différents taux de virus et les aliquots ont été prélevés tous les jours jusqu'à une mortalité de 100% pour chaque test. Une analyse par Western Blot a permis de déterminer à la fois le taux et le temps d'infection optimums pour l'obtention du meilleur rendement protéique.

4.2.3.4 Expression des protéines MIC en cellules d'insectes Hi5 et purification des complexes MIC/MIC12

L'expression des protéines MICA*001, MICA*008 et MICB*002 en système baculovirus a été faite en cellules Hi5. Les taux volumiques respectifs en virus/culture ont été définis respectivement à 1/300, 1/500 et 1/500. Lorsque la concentration a atteint environ 1,5 Million de cellules par mL avec au moins 1L de cultures, les cellules ont été infectées selon les taux déterminés. L'infection a été arrêtée lorsque la mortalité était au moins de 90%. Les cultures ont alors été centrifugées à 8000 g pendant 8 min et le surnageant a été récupéré. Après ajout de 0,02% d'azide et 1mM de PMSF, il peut être gardé deux jours maximum à 4°C ou purifié tout de suite.

Les trois protéines, fusionnées à une étiquette Strep, ont été extraites du surnageant par formation du complexe avec MIC12 directement dans le surnageant de culture puis fixation de ce dernier sur colonne d'affinité Ni²⁺ (GE Healthcare). Pour 400 mL de

surnageant, 8 mg de Fab ont été ajoutés et laissé incubé sous agitation 1h à 4°C. Le surnageant a ensuite été injecté sur colonne HisTrap Excel 5 mL pré-équilibrée en tampon 50 mM de Tris pH 8, 300 mM de NaCl, 10 mM d'Imidazole. La purification s'est poursuivie comme décrit précédemment, avec comme tampon de chromatographie d'exclusion de taille un tampon 10 mM d'Hepes (pH adapté au PI théorique des complexes), 150 mM de NaCl. Le pic de complexe a été collecté, puis le pic d'excès de MIC12.

4.2.4 Analyses biochimiques et biophysiques

4.2.4.1 Dichroïsme circulaire (CD)

Pour vérifier le repliement correcte des protéines MIC, en particulier celles issues d'une expression en système bactérien, une étude du contenu en structures secondaires a été menée par CD (paragraphe 4.1.2). Le laboratoire dispose d'un spectropolarimètre J-810 (Jasco). Les échantillons, dilués dans un tampon non absorbant dans cette gamme de longueur d'onde, sont à une concentration finale de 0,3 mg/mL. Les mesures sont réalisées dans une cuve en Quartz de 1 mm avec tout d'abord une mesure du blanc (tampon seul + volume en tampon de protéine) puis la mesure de l'échantillon. Les protéines ont été diluées dans un tampon 50 mM de NaH₂PO₄ pH 7. Les mesures ont été réalisées entre 190 et 260 nm, avec accumulation de 3 spectres. Seules les données obtenues avec une valeur de HT inférieure à 600 V sont prises en compte.

Les données ont été lissées puis exportées dans un tableau. Les valeurs d'ellipticité moyenne par résidu ont été calculées d'après la formule :

$$\theta = \frac{(Ech - Blanc) \times MM}{[Ech] \times nb.aa}$$

Avec MM : poids moléculaire théorique (Da), [Ech] : concentration de l'échantillon (mg/mL) et nb.aa : nombre d'acides aminés. Θ est exprimé en deg.cm².d.mol⁻¹. Les spectres obtenus sont ensuite déconvolués à l'aide du programme Dichroweb (<http://dichroweb.cryst.bbk.ac.uk/html/home.shtml>) en utilisant le programme K2D.

4.2.4.2 Analyses par Size Exclusion Chromatography-Multi Angles Light Scattering SEC-MALS des protéines MIC

Bien que les analyses précédentes donnent une idée globale du poids moléculaire et de l'état d'oligomérisation des protéines purifiées, la méthode Size Exclusion Chromatography-Multi Angles Light Scattering (SEC-MALS) renseigne de manière précise

sur ces paramètres par mesure de la diffusion de la lumière laser (LS), de l'index de réfraction (RI) et de l'absorbance UV. Le rayon de giration peut être déterminé à partir de la diffusion statique et quasi-élastique de la lumière.

La technologie est disponible au laboratoire au travers d'un service plateforme grâce à un système Wyatt-HPLC (Alliance 2695 HPLC, Waters). Les protéines ont été chargées sur une colonne de chromatographie d'exclusion Shodex KW-802.5 (Phenomenex), pré-équilibrée avec le tampon des protéines, couplée à un détecteur de diffusion de lumière 3-angles MiniDAWN TREOS (Wyatt Technology), un instrument de diffusion de la lumière quasi élastique Dynapro (Wyatt Technology) et un réfractomètre différentiel Optilab rEX (Wyatt Technology). Des analyses ont été faites sur les protéines MICA*001, MICA*008 et MICB*002 issues d'une production en système bactérien. Pour cela, les échantillons ont été concentrés à 2 mg/mL pour un volume minimum de 30 µL.

4.2.4.3 Mesure des interactions Fab/protéine

Les interactions Fab/protéine ont été étudiées en utilisant la technologie Bio-Layer Interferometry (BLI), disponible au laboratoire sous la forme d'un service plateforme. Les mesures ont été réalisées sur un Octect RED96 .

4.2.5 Cristallographie

4.2.5.1 Cristallogenèses

Le laboratoire dispose d'une plateforme de cristallisation comprenant un robot de remplissage de boîtes (TECAN) ainsi qu'un robot de précision permettant de prélever et de déposer des volumes de l'ordre du nanolitre (Mosquito). Les plaques utilisées sont principalement des plaques 96 MRC 3 puits SWISSCI (Jena Bioscience) et, plus occasionnellement des plaques 96 3 puits (Greiner Bio One). Les conditions de cristallisations sont dispensées dans les 96 réservoirs (40 µL pour les SWISSCI, 100 µL pour les Greiner) et un mélange protéine/réservoir est distribué dans les puits. Pour tester trois concentrations en protéine, des mélanges $v(\text{protéine})/v(\text{réservoir})$ 300/100, 200/100 et 100/100 nl sont réalisés pour chaque condition. La protéine est concentrée jusqu'à sa limite de solubilité puis des premiers essais de cristallogenèses sont réalisés en screens commerciaux. Les plaques sont stockées à 20°C dans un système de stockage et prise d'image Formulatrix contrôlé par le programme Rock imager (Formulatrix). L'interface

CRIMS disponible en intranet au laboratoire permet de programmer une prise d'image en lumière visible ou UV sur 200 jours. Lorsqu'une condition de cristallisation est identifiée, une visualisation par UV permet de déterminer si l'objet cristallin est bien de nature protéique et non saline. Le pH de la condition est alors mesuré puis une optimisation est réalisée en faisant varier le pH et la concentration en agent précipitant autour de la condition déterminée.

Les cristallogènes des protéines de ce volet ont toutes été réalisées suivant la méthode de diffusion vapeur, remplies en utilisant les robots pour les plaques 96 puits avec gouttes assises ou à la main pour les plaques 24 puits avec gouttes suspendues.

Pour le Fab MIC12, la protéine a d'abord été concentrée à 18 mg/mL mais l'évaluation de la concentration par un premier screen général Stura (Molecular Dimension) a révélé qu'il fallait encore l'augmenter. Les screens généraux MDL (Molecular Dimension), Index (Hampton Research) et JCSG+ (Molecular Dimension) ont été testés avec une concentration en MIC12 de 23 mg/mL. L'optimisation des cristaux a été réalisée en pratiquant autour de la condition de cristallisation définie une gamme de pH et de concentration en agent précipitant.

Pour les trois complexes MIC12/MIC issus de l'expression bactérienne, les screens généraux Stura (Molecular Dimension), MDL (Molecular Dimension), Wizard I et II (Rigaku), PEGs I et PEGs II (Qiagen), JCSG+ (Molecular Dimension) et Morpheus (Molecular Dimension) ont été testés. Des optimisations ont été faites à partir des conditions déterminées, en faisant varier le pH, la concentration en agent précipitant ou en testant comme additifs les screens Anions et Cations (Qiagen).

4.2.5.2 Cryoprotection et congélation

Chaque cristal obtenu a été testé aux UV_{280nm} ou, le cas échéant, sa diffraction a pu être testée au laboratoire.

La diffraction des cristaux a ensuite été testée en utilisant la source de rayons X de type anode tournante combinée au détecteur MAR345 dont dispose le laboratoire. Tout d'abord, une solution cryoprotectante doit être définie. La diffraction de la condition de cristallisation, congelée sous flux d'azote liquide à 100°K, est d'abord testée seule pour voir si elle n'a pas déjà des propriétés cryoprotectantes. Dans le cas contraire, des additifs tels que le PEG 400, PEG 4000, EG, PPG ou le glycérol peuvent être ajoutés à des pourcentages variant de 10 à 25%. Une fois la solution cryoprotectante définie, un cristal est sacrifié pour

tester le comportement des cristaux face à cette nouvelle condition. Il est pêché puis déposé dans une nano-goutte et observé au microscope binoculaire quelques dizaines de secondes. S'il a pu supporter la solution, celle-ci est validée. Les cristaux sont alors pêchés à l'aide d'une boucle cryogénique en nylon (Hampton Research), de diamètre adapté à la taille du cristal, et déposés dans une nano-goutte de cette solution quelques secondes puis pêchés de nouveaux et montés sur la tête du goniomètre, sous flux d'azote à 100 °K. Le cristal congelé, la diffraction peut être testée puis le cristal collecté ou bien stocké en azote liquide en attendant une prochaine mission synchrotron (ESRF ou SOLEIL).

4.2.5.3 Collecte des données et détermination de la structure de MIC12

Les clichés de diffraction du Fab MIC12 ont été collectés au laboratoire puis à l'ESRF sur la ligne ID23-2.

Au laboratoire, deux images ont été prises avec une rotation de 1° sur l'axe ϕ par image avec un incrément de 90° entre les deux. Celles-ci ont été traitées avec iMOSFLM de la suite CCP4 pour définir des paramètres de maille, un groupe d'espace et donc une stratégie de collecte d'images. Une collecte de 120 images, avec un temps d'exposition de 5 min et une rotation de 1° par image, a été lancée sur la nuit. Les données ont d'abord été traitées avec iMOSFLM pour les étapes d'indexation, d'affinement de la maille, d'intégration et pour l'analyse de la symétrie avec POINTLESS. Les données ont ensuite été mises à l'échelle avec le programme Scala de CCP4 avec définition d'un pourcentage de 5% des réflexions en Rfree. Le nombre de molécules par unités asymétrique a été estimé par calcul du coefficient de Matthews (programme Matthews coef). La structure a alors été résolue par remplacement moléculaire à l'aide du programme MolRep en utilisant la structure d'un Fab de même paramètres de maille : 3RVT. Un premier affinement a été réalisé avec le programme d'affinement REFMAC5 puis avec autoBUSTER (Global Phasing Limited). Le modèle a été modifié manuellement entre chaque affinement à l'aide du logiciel COOT, implémenté dans la suite CCP4.

Le même cristal a également été collecté sur la ligne de lumière ID23-2 de l'ESRF à Grenoble. Il s'agit d'une ligne microfocus équipée à l'époque de la mission d'un détecteur PILATUS 3M. Le cristal a été caractérisé à partir de 4 images à l'aide de la plateforme MXCuBE. Les paramètres de mailles et groupe d'espace définis, une collecte de 1800 images avec une rotation de 0,1° et un temps d'exposition de 0,1 s par image a été lancée. Les données ont été traitées au laboratoire d'abord à l'aide de la suite de programmes XDS

pour les mêmes étapes que citées précédemment pour iMOSFLM. La suite du traitement a été réalisée avec la suite CCP4 comme décrit plus haut. Le remplacement moléculaire a été réalisé à partir de la structure de MIC12 résolue au laboratoire.

La structure affinée a ensuite été évaluée par le serveur MolProbity (Duke Biochemistry, <http://molprobity.biochem.duke.edu/>). Après ajout des hydrogènes, les rotamères, les longueurs et angles des liaisons, les angles selon le plot de Ramachandran et les clashes stériques sont répertoriés et permettent d'améliorer le modèle puis de le valider.

4.2.6 SAXS des protéines MIC et complexes MIC/MIC12

4.2.6.1 Préparation des échantillons et collecte des données

Les études structurales par SAXS ont été menées sur les protéines et complexes MICA*001, MICA*008, MIC12, MICA*001/MIC12 et MICA*008/MIC12. Béatrice Amigues a purifié les protéines par chromatographie d'exclusion de taille, en étape finale, en utilisant un tampon final 50 mM de PIPES à pH 7, 150 mM de NaCl et 1 mM d'EDTA. J'ai concentré chacun de ces échantillons à l'aide d'un concentrateur Amicon® Ultra (Millipore) de cut off adapté. Des échantillons de 100 µl ont été prélevés au fur et à mesure de la concentration lorsque celle-ci atteignait 0,5 mg/mL, 1 mg/mL, 3 mg/mL, 5 mg/mL puis 10 mg/mL. Le filtrat de concentration a été conservé pour être utilisé comme blanc de chaque mesure. Les échantillons ont été collectés sur la ligne SAXS BM29 de l'ESRF à Grenoble (64). Les expériences ont été réalisées en utilisant un détecteur PILATUS 1M 2D (Dectris) à une distance avec l'échantillon de 2,867 m. La longueur d'onde utilisée était de 0,09919 nm ce qui permet des mesures dans une gamme de vecteurs de diffusion allant de 0,0364 à 4,94 nm⁻¹. Les échantillons ont été centrifugés 10 min à 11000g à 4°C pour éliminer d'éventuels agrégats puis transférés dans de nouveaux tubes. Après avoir repris toutes les concentrations, les protéines ont été chargées sur le changeur d'échantillons sur la ligne. Les expériences ont été menées à 10°C en utilisant pour chaque échantillon 30 µl de protéine. Chaque expérience s'est déroulée selon le schéma suivant : 10 acquisitions de 1s du tampon, 10 autres de la protéine puis de nouveau 10 du tampon. Une calibration à la BSA a été faite avant par le scientifique de la ligne et toutes les données ont été transférées et traitées en temps réel via la plateforme ISPyB.

4.2.6.2 Traitement des données et analyses

Bien que les données soient traitées par ISPyB, je préfère généralement les traiter moi-même. Ainsi, j'ai trié manuellement les courbes de diffusion des expériences menées sur les protéines MIC.

Dans un premier temps, les 10 courbes de chaque lot d'acquisition ont été comparées, que ce soit des acquisitions de tampon ou de protéine, afin de vérifier qu'il n'y ait pas eu de dommages aux radiations. Après avoir éliminé les courbes endommagées, les courbes restantes ont été moyennées. Les courbes de tampon avant et après acquisition de la protéine ont été comparées. Encore une fois, l'impossibilité de superposer ces courbes est le signal de probables dommages aux radiations. Ces courbes ont été soustraites à la courbe moyenne de l'échantillon après avoir été moyennées. La courbe de diffusion résultante est la courbe finale de diffusion de la protéine à la concentration donnée.

La première analyse consiste en une comparaison des différentes courbes selon les concentrations de la protéine. L'analyse de Guinier a été réalisée pour chaque concentration et celles présentant des signes d'agrégation ont été écartées. Les valeurs de R_g et I_0 ont été estimées et comparées afin là encore d'écartier les concentrations en désaccord avec l'ensemble des données. Pour chaque protéine, une faible et une forte concentration valables ont été sélectionnées afin de créer une courbe chimérique de diffusion. Le poids moléculaire est estimé d'après cette courbe en utilisant le serveur SAXS MoW2 (<http://saxs.ifsc.usp.br/>). Le degré de flexibilité des protéines a été analysé en observant le plot de Kratky sans dimensions de chacune. Contrairement au plot de Kratky, représentation de $I \cdot q^2$ en fonction de q , ce graphique prend en compte le R_g et le I_0 calculés lors de l'analyse de Guinier. Cela permet de comparer entre elles des protéines de poids moléculaires et forme diverses(65). Cette nouvelle courbe a de nouveau été analysée selon Guinier et le premier point pris en compte pour l'alignement de la droite est sélectionné comme premier point pour l'analyse de distribution des distances. En prenant le maximum de points (256), un compromis a été recherché entre obtenir la meilleure superposition théorique sur la courbe de diffusion expérimentale et éviter la région de plus haute résolution, trop bruitée. La distance D_{max} a été modifiée jusqu'à faire doucement retomber la courbe de Porod à une valeur de $R_{max}=0$.

Pour les protéines MICA*001 et MIC12, les structures cristallographiques sont disponibles. Une étude par modélisation *ab-initio* d'une enveloppe pourrait être informative uniquement sur la dynamique en solution, mais d'autres analyses plus précises peut être réalisées en ce sens. En revanche, il est possible de réaliser une analyse en utilisant

CRY SOL. Une courbe théorique de diffusion SAXS a été générée en prenant en compte une couche d'hydratation superficielle calculée. Cette courbe est superposée à la courbe de diffusion chimérique expérimentale. La valeur de la fonction χ^2 , permettant l'optimisation de la procédure de minimisation, est un indicateur de la qualité du recouvrement des deux courbes. Cette fonction χ^2 est définie telle que :

$$\chi^2 = \sum_{i=1}^N \frac{(I_{exp}(q_i) - I_{teo}(q_i))^2}{\sigma(q_i)^2}$$

avec $I_{exp}(q_i)$ les intensités des N points expérimentaux, $I_{teo}(q_i)$ celles des points théoriques et $\sigma(q_i)$ les déviations standard (66). La valeur du χ^2 pour une superposition de bonne qualité doit être proche ou égale à 1.

Pour les complexes MIC/MIC12, des enveloppes ont été modélisées en utilisant le programme DAMMIF (19). A partir de la distance D_{max} , une sphère a été modélisée sous forme de billes et des modèles ont été générés par recuits simulés. Chacun a été comparé aux données expérimentales et un test de χ^2 a permis l'évaluation de la qualité de la superposition des courbes de diffusions expérimentales et théoriques. Les dix meilleurs modèles ont été sélectionnés et le modèle avec la meilleure valeur de χ^2 a été retenu.

Le complexe MICA*001/MIC12 a également été étudié par docking moléculaire. Des glycosylations ont été ajoutées au modèle prédit le plus probable à l'aide du programme GLYCOSYLATION. J'ai utilisé pour cela le mode expert afin de choisir les sites et le nombre de glycosylations en fonction de l'enveloppe générée par DAMMIF. Les glycosylations apportées sont au nombre de 4, portant sur les résidus N8, S17, N141 et S274.

4.2.7 Prédiction de structures du complexe MICA*001/MIC12 par docking moléculaire.

Le docking moléculaire du complexe MICA*001/MIC12 a été réalisé à partir du serveur ClusPro 2.0 (<https://cluspro.bu.edu/>) (67) (68) (69). Les prédictions ont été lancées en utilisant le mode Anticorps avec MIC12 comme récepteur et MICA*001 comme ligand. Les modèles utilisés pour MICA*001 ont pour numéro d'accèsion pdb 1HYR et 1B3J (données non présentées pour la seconde structure). Après 70000 rotations, 29 clusters de structures ont été générés et un modèle moyen pour chaque est mis à disposition.

Les modèles ont été triés selon 1) la cohérence de positionnement de l'anticorps vis à vis de contraintes spatiales (la présence de la membrane cellulaire par exemple) et 2) les résidus impliqués dans l'interaction. En effet, des résidus ont déjà été identifiés comme étant impliqués dans cette interaction par de précédentes études (14). Les interactions des résidus pour les modèles retenus ont été analysées à l'aide du logiciel PyMOL (résultats non présentés).

Le modèle prédit du complexe ainsi que le modèle glycosylé ont été comparés à l'enveloppe SAXS générée en utilisant le logiciel Chimera (UCSF Chimera). La carte a été générée puis les modèles y ont été introduits à l'aide de l'outil « Fit to map ».

4.3 Utilisation de nanobody pour les études structurales de la protéine pPorM

4.3.1 Nanobody nb01 et nb02 : expression et purification.

L'expression des nanobody nb01 et nb02 a été réalisée dans la partie périplasmique de cellules *E. coli* de souche WK6. La pré-culture a été réalisée dans 10 mL de milieu LB avec 100 µg/mL d'ampicilline à partir de stock glycérolé et incubée 15h à 37°C. Pour chaque litre de culture, 3mL de pré-culture ont été utilisés pourensemencer du milieu TB contenant 100 µg/mL d'ampicilline et 0,1% de glucose. Les cultures ont été incubées à 37°C jusqu'à atteindre une DO_{600nm} entre 0,5 et 0,8 puis induites avec 1mM d'IPTG à 28°C pendant 20h. Les cellules ont ensuite été culotées par centrifugation à 8000g pendant 8min et le surnageant a été éliminé.

Le culot obtenu a été repris dans 9mL de tampon réfrigéré TES (0,2M de Tris-HCl pH8, 0,5mM d'EDTA et 0,5M de sucrose) puis laissé 1h à 4°C. Les protéines périplasmiques ont été extraites par ajout de 54 mL de tampon TES dilué par 4 avec de l'eau et incubation 2h à 4°C. Après deux centrifugations à 24000g pendant 30min, les protéines périplasmiques se trouvent dans le surnageant qui peut être conservé à -20°C ou être purifié tout de suite, selon les mêmes protocoles que décrits plus haut.

4.3.2 Cristallographie à température ambiante de nb02

4.3.2.1 Cristallisation du nanobody nb02 en puce microfluidique.

Dans le cadre d'une collaboration avec Claude Sauter (IBMC, Strasbourg), la cristallisation de protéines dans des puces microfluidiques a été testée : les ChipX-3 (Figure 4.3.1). Ces puces permettent une cristallisation libre de tout phénomène de convection et donc une cristallisation plus homogène. Une fois les cristaux obtenus, leur diffraction est réalisée directement dans la puce (*in-situ*) et permet alors une étude structurale à température ambiante(70).

La puce a d'abord été rincée à l'eau déminéralisée par injection du côté « protéine » puis l'eau a été chassée par injection d'air comprimé. Un volume de 8 μ l de la protéine nb02, concentrée à 13,78 mg/ml, a été injecté à l'entrée des 8 canaux (à gauche de la chip) jusqu'à les remplir totalement. Dans les 8 réservoirs situés à droite, 5 μ l de condition de cristallisation ont été déposés en évitant toute injection d'air dans les canaux. Pour cette puce, la condition de cristallisation définie par diffusion vapeur en gouttes assises(5) a été reprise et optimisée, du réservoir A à H, selon une gamme de pH allant de 5 à 6, en se plaçant à une concentration de PEG 3000 de 20%. L'entrée de la protéine et les réservoirs ont ensuite été fermés hermétiquement avec un scotch cristal clear puis la chip a été stockée à 20°C.



Figure 4.3.1: Puces micro-fluidiques ChipX-3. La protéine est injectée dans les 8 canaux à partir de l'entrée à gauche. A droite, les 8 réservoirs (A à H) permettent de tester 8 conditions de cristallisation par diffusion contraire.

4.3.2.2 Collecte *in-situ* à température ambiante des cristaux de nb02.

Les cristaux obtenus dans la puce ChipX-3 ont été collectés sur la ligne Proxima 2A à SOLEIL (Gif-sur-Yvette). La puce a été montée sur le goniomètre à l'aide d'un portoir créé par les scientifiques de la ligne. Les différents matériaux séparant le faisceau du cristal induisent de nombreux changements d'indices de réfraction, créant une différence non

négligeable entre la position optiquement visible du cristal et sa position réelle. L'alignement de chaque cristal avec le faisceau de rayons X a donc été optimisé à l'aide d'un scan de rayons X, basé sur 3 angles sur l'axe ϕ . Une fois la position enregistrée, les cristaux ont été collectés sur 20° à 180° à une énergie de 15 keV, avec une transmission de 2, 5 ou 10%, un incrément d'angle entre deux images de $0,1^\circ$ et un temps d'exposition de 0,025 s par image. Le cristal EmbCD a, lui, été collecté deux fois. Le premier jeu de données contient 800 images collectées sur 80° avec une transmission de 10% du faisceau. Le second jeu de données contient 1800 images collectées sur 180° avec une transmission de 5%.

4.3.2.3 Traitement des jeux de données et détermination de la structure de nb02 température ambiante.

Les images issues des différents jeux de données ont été triées une à une à l'aide du logiciel de visualisation d'image Albula (DECTRIS). Chaque jeu a été traité à l'aide de la suite XDS puis les réflexions ont été mises à l'échelle à l'aide du programme XSCALE (<https://github.com/legrandp/xdsme/blob/master/bin/noarch/xscale2.py>, Pierre Legrand). Les réflexions des différents jeux de données ont alors été combinées, toujours avec XSCALE, afin d'obtenir un jeu de données complet.

Le jeu de données a été analysé avec le programme Xtriage de PHENIX. Le remplacement moléculaire a été réalisé à l'aide du programme phenix.phaser, en utilisant la structure de nb02 (numéro d'accèsion RCSB PDB : 5LMW).

Les affinements de la structure ont d'abord été réalisés une fois avec Autobuild puis avec phenix.refine. Des « Translation Libration Screw-rotation » TLS ont été définies à l'aide du serveur TLSMD (<http://skuld.bmsc.washington.edu/~tlsmd/>).

4.3.3 Clonage des constructions pPorM-Nt et pPorM-Ct.

Les protéines pPorM-Nt et pPorM-Ct ont été clonées dans le plasmide pLIC03 en utilisant la technique du RF cloning (paragraphe 4.1.1). Tout comme le plasmide pLIC06 précédemment décrit, ce plasmide est dérivé du plasmide pET28 a+ (Novagen) et possède une résistance à l'antibiotique kanamycine. Il permet l'expression de protéines recombinantes à l'aide d'un promoteur T7 avec une étiquette HisTag en N-terminal suivie

d'un site de coupure à la protéase TEV. Les amorces utilisées sont notées dans le Tableau 4.2.1.

4.3.4 Expression des protéines pPorM-Nt native et SeMet.

Les protéines pPorM-Nt native et avec SelenoMethionine (SeMet) ont été exprimées en cellules Rosetta (DE3) pLysS (Novagen) avec les antibiotiques kanamycine et chloramphénicol.

La protéine pPorM-Nt a été exprimée de la même manière que les protéines MIC.

Les cultures permettant l'expression de pPorM-Nt SeMet ont été réalisées en milieu auto-inducteur. Une pré-culture de pPorM-Nt a été réalisée en milieu SB avec kanamycine et chloramphénicol et incubée à 37°C pendant 9h sous agitation. Le milieu auto-inducteur est constitué de 50 mM de Na₂HPO₄, 50 mM de KH₂PO₄, 25 mM de (NH₄)₂SO₄, 2 mM de MgSO₄, 0,2x de mix de métaux (stock à 1000x), 0,5% de glycérol, 0,05% de glucose, 0,2% d' α -lactose, 200 μ g/mL de chaque acide aminé (sauf cystéine, tyrosine et méthionine), 10 μ g/mL de Met, 125 μ g/mL de SeMet et 100 nM de vitamine B12. Ce milieu a étéensemencé avec la pré-culture (25mL pour 1L) puis incubé à 30°C pendant 15h. Une fois que la DO_{600nm} a atteint une valeur de 2, la température a ensuite été augmentée à 37°C pendant 10h puis diminuée à 20°C pendant 15h. Les cultures ont ensuite été arrêtées selon le protocole décrit précédemment.

4.3.5 Purification des constructions de pPorM

Après décongélation des culots bactériens, les cellules ont été lysées suivant la procédure suivie pour les protéines MIC exprimées en cellules *E. coli*, à l'exception que les constructions de pPorM sont toutes exprimées dans la fraction soluble. Après la sonication et centrifugation du lysat, le surnageant a été récupéré cette fois-ci, constituant le lysat clarifié. Toutes les constructions de pPorM, à savoir pPorM, pPorM-Nt et pPorM-Ct, sont exprimées sous forme soluble dans la bactérie *E. coli* Rosetta (DE3) pLysS (Novagen) et contiennent une étiquette HisTag. Leur purification a donc été réalisée selon les mêmes protocoles que décrits précédemment. Le tampon utilisé pour la purification par chromatographie d'affinité Ni²⁺ (GE Healthcare) est constitué de 50 mM de Tris pH8, 300 mM de NaCl, supplémenté de 10 mM ou 250 mM d'Imidazole. Toutes ces protéines ont été éluées avec la seconde élution à 250 mM d'Imidazole. Le tampon d'équilibration et

d'élution de la colonne d'exclusion de taille est un tampon 10 mM d'Hepes à un pH adapté au point isoélectrique de la protéine, 150 mM de NaCl. La colonne utilisée est une Superdex 200 16/60 (GE Healthcare), excepté pour les protéines pPorM-Nt et pPorM-Nt SeMet (Superdex 75 16/60, GE Healthcare). Pour obtenir la protéine pPorM-T, une digestion à la protéase Trypsine a été réalisée après la chromatographie d'exclusion de taille. Un ratio massique 1 :1000 a été appliqué à température ambiante pendant 15 heures. Une fois l'enzyme neutralisée par l'ajout d'un cocktail inhibiteur de protéases, une nouvelle chromatographie d'exclusion de taille a été réalisée.

4.3.6 Analyses biochimiques et biophysiques de pPorM-Nt

4.3.6.1 Dichroïsme circulaire de la protéine pPorM-Nt

Après purification, le contenu en structures secondaires de la protéine pPorM-Nt a été étudié par dichroïsme circulaire. La méthodologie appliquée est strictement la même que pour les protéines MIC (paragraphe 4.2.4.1). Suite à la courbe obtenue, typique d'un contenu majoritairement en hélices α , une dénaturation thermique a été réalisée sur une gamme de température allant de 20°C à 80°C à une longueur d'onde de 222 nm, spécifique de ces structures secondaires.

4.3.7 Purification des complexes pPorM-Nt/nanobody

Les complexes formés par la protéine pPorM-Nt avec soit le nanobody nb01, soit nb02, ont été réalisés avec les protéines purifiées en sortie de chromatographie d'exclusion de taille. Les protéines ont été complexées suivant un ratio molaire protéine/nanobody 1 :1.2 à température ambiante pendant 15 min. après centrifugation à 11000 g pendant 10 min à 4°C, le mélange a été chargé sur une colonne Superdex 75 16/60 (GE Healthcare) équilibrée en tampon 10 mM d'Hepes pH7, 150 mM de NaCl. Le pic correspondant au complexe a été élué avant le pic d'excès de nanobody dans le même tampon.

4.3.8 Cristallographie aux rayons X de pPorM-Nt native et SeMet et des complexes pPorM-Nt/nanobody

4.3.8.1 Cristallogenèses

La cristallogenèse des protéines pPorM-Nt native et SeMet, les complexes pPorM-Nt/nb01 et pPorM-Nt/nb02 (natifs et SeMet) a été réalisée par diffusion vapeur, principalement en gouttes assises.

Pour la protéine pPorM-Nt native, les conditions des screens généraux Stura (Molecular Dimension), Index (Hampton Research), MDL (Molecular Dimension), Wizard I et II (Rigaku), JCSG+ (Molecular Dimension) ont d'abord été criblées à une concentration en protéine de 10 mg/ml. Ensuite, les screens PEGs I et PEGs II (Qiagen) ont été testés à 8 et 12 mg/ml. Les conditions donnant des cristaux ont été combinées et optimisées en réalisant un criblage selon une gamme de pH et de concentration en agent précipitant, à une concentration en protéine de 12 mg/ml. La même optimisation a également été réalisée en gouttes suspendues.

Du trempage d'atome lourd a été réalisé sur deux cristaux optimisés de pPorM-Nt avec du Tantale.

Les essais de cristallisation de la protéine pPorM-Nt SeMet ont été menés à partir des screens généraux Stura (Molecular Dimension), Wizard I et II (Rigaku) et PEGs II (Qiagen) à une concentration en protéine de 8, 10 et 12 mg/ml. Une optimisation suivant la condition identifiée de cristallisation de la protéine native a été faite, ainsi qu'une optimisation classique autour de la condition identifiée pour la protéine SeMet.

De nombreux essais ont été menés pour la cristallisation des complexes pPorM-Nt/nb02, pPorM-Nt SeMet/nb01 et pPorM-Nt SeMet/nb02. Les screens généraux cités plus haut ont tous été testés, ainsi que le screen Morpheus, à une concentration de 5 et 10 mg/ml. Des ensemencements par des précipités cristallins ont également été tentés.

La cristallogénèse du complexe pPorM-Nt/nb01 a été réalisée par criblage des screens généraux Stura et Index à 10 mg/ml.

4.3.8.2 Cryo-protection et congélation des cristaux

Les cristaux obtenus pour les protéines pPorM-Nt, pPorM-Nt SeMet et le complexe pPorM-Nt/nb01 ont été cryo-protégés et congelés comme décrit plus haut pour les cristaux des protéines et complexes MIC.

4.3.8.3 Collecte des données et traitement des données

Les données ont toutes été collectées aux synchrotrons après avoir testé la diffraction des cristaux au laboratoire.

Le jeu de données de la protéine pPorM-Nt a été collecté à l'ESRF, sur la ligne de lumière ID23-2. Une stratégie de collecte a été définie par MxCube à partir de 4 images prises à 45° entre chaque et 2400 images ont été collectées. L'incrément d'angle a été de 0,05° sur une rotation totale de 48°, avec un temps d'exposition de 0,04 s par image. Les

données ont été traitées au laboratoire avec la suite de programmes XDS. Le twin parfait des données a été révélé par le test de détection de twin du programme POINTLESS (CCP4). Ce twin a été traité à l'aide des programmes detwin (CCP4), phenix.refine avec l'introduction d'une loi de twin en option (PHENIX) et du module data_detwin du programme CNSsolve (CNS).

Le jeu de données correspondant à la protéine pPorM-Nt SeMet a été collecté à SOLEIL sur Proxima 1. Un spectre d'absorption a permis de confirmer la présence de signal anomal dû à du Sélénium et un scan d'énergie a aidé à la détermination des énergies du point d'inflexion, du haut du pic et après ce pic. Un set de 1000 images a été collecté à chacune de ces énergies (stratégie de collecte MAD) avec un incrément d'angle de 0,1° et un temps d'exposition par image de 0,1s.

Deux jeux de données du complexe pPorM-Nt/nb01 ont été collectés : l'un sur Proxima 1 et l'autre sur Proxima 2A. Le premier jeu contient 3600 images obtenues avec un incrément d'angle par image de 0,1s et un temps d'exposition de 0,1°. Ces données ont été traitées avec XDSME puis CCP4 à partir de la recherche des groupes de Laue à l'aide de POINTLESS. Les données ont été mises à l'échelle avec SCALA.

4.3.8.4 Détermination de la structure du complexe pPorM-Nt/nb01 et affinement.

Les phases de nb01 ont été utilisées pour le remplacement moléculaire avec PHASER. L'utilisation du programme BUCCANEER de la suite CCP4 a permis de construire des chaînes polyAlanine dans les blobs de densité, de forme hélices alpha, situées en face des 4 chaînes de nb02. Après un premier affinement avec RefMac5, la structure a été construite à l'aide de COOT puis affinée à l'aide du programme autoBUSTER. Des TLS ont été définis à l'aide du serveur TLSMD puis introduits lors de l'affinement. La structure a ensuite été affinée avec autoBUSTER en utilisant les options autoncs et TLS.

4.3.8.5 Evaluation et analyse de la structure cristallographique de pPorM-Nt/nb01.

La qualité de la structure construite a été évaluée avec les outils de validation disponibles dans COOT mais également avec le serveur MolProbity (<http://molprobity.biochem.duke.edu/>).

La structure de pPorM-Nt a été soumise au serveur DALI pour être comparée à d'autres structures contenues dans la RCSB PDB. Les dix structures de plus haut score ont été prises en compte.

L'interaction des deux protéines a été analysée à l'aide du logiciel PyMOL. Les résidus entrant en jeu dans l'interaction ont été relevés. Toujours avec COOT, les structures de nb01 seul et en complexe ont été comparées.

4.3.9 SAXS des constructions de pPorM

4.3.9.1 Préparation des échantillons et collecte des données

Les protéines pPorM, pPorM-T, pPorM-Nt et pPorM-Ct ont été purifiées au laboratoire avec comme dernière étape de purification une chromatographie d'exclusion de taille en utilisant une colonne Superdex 16/60 (75 ou 200 selon la taille de la protéine, GE Healthcare). Toutes ont été éluées dans un tampon Hepes 10 mM pH 7,3, NaCl 150 mM. Les échantillons pour chaque protéine ont été prélevés au fur et à mesure de la concentration de manière à avoir 5 concentrations différentes dans une gamme allant de 0,5 à 15 mg/ml. Les protéines ont été concentrées à l'aide d'un concentrateur Amicon® Ultra (Millipore) adapté au poids moléculaire de la protéine et le filtrat a été conservé pour servir de blanc d'acquisition. Les données ont été collectées sur la ligne BM29 de l'ESRF, en utilisant un détecteur PILATUS 1M 2D à une distance de 2,87 m. La longueur d'onde était de 0,09919 nm, permettant l'enregistrement sur une gamme de vecteurs diffusés q allant de $0,036 \text{ nm}^{-1}$ à $4,912 \text{ nm}^{-1}$. Les échantillons ont été préparés et les données enregistrées de la même manière que pour les études SAXS des protéines MIC. Pour palier aux dommages aux radiations de la protéine pPorM, 5mM de DTT ont été ajoutés lors d'une seconde mesure de chaque concentration.

4.3.9.2 Traitement des données et analyses

Les données ont été traitées suivant le même protocole que celui appliqué pour les protéines MIC. Les courbes de diffusion ont été triées et comparées et des courbes chimères ont été créées. Chacune a été analysée par Guinier, Kratky puis selon la distribution des distances et Porod. Afin de pouvoir comparer les courbes de Kratky des 4 protéines, une représentation sans dimensions, prenant en compte les R_g et I_0 calculés par l'analyse de Guinier, a été adoptée. De même, les courbes de Porod ont été normalisées de manière à ce que le haut de chacune corresponde à une valeur de $P(r)=1$. Les poids moléculaires et état oligomériques ont été estimés à l'aide du serveur SAXS MoW2.

Ces données SAXS ont été comparées aux données cristallographiques pour les protéines pPorM-T, pPorM-Nt et pPorM-Ct à l'aide du programme CRY SOL. Des

modélisations ab-initio des protéines pPorM et pPorM-Nt ont été réalisées à l'aide du programme DAMMIF en générant les dix meilleures enveloppes. Un affinement de la flexibilité du modèle cristallographique de pPorM-T selon les données expérimentales SAXS a été réalisé avec le programme SREFLEX et dix modèles ont été générés. La qualité de tous ces résultats a été évaluée par un test de χ^2 basé sur les données expérimentales.

Références :

1. Roitt, I., Brostoff, J., and Male, D. (1994) *Immunologie*, DeBoeck
2. Fernandes, C. F. C., Pereira, S. dos S., Luiz, M. B., Zuliani, J. P., Furtado, G. P., and Stabeli, R. G. (2017) Camelid Single-Domain Antibodies As an Alternative to Overcome Challenges Related to the Prevention, Detection, and Control of Neglected Tropical Diseases. *Front. Immunol.* 10.3389/fimmu.2017.00653
3. Finlay, W. J. J., and Almagro, J. C. (2012) Natural and man-made V-gene repertoires for antibody discovery. *Front. Immunol.* 10.3389/fimmu.2012.00342
4. Steeland, S., Vandembroucke, R. E., and Libert, C. (2016) Nanobodies as therapeutics: big opportunities for small antibodies. *Drug Discov. Today.* **21**, 1076–1113
5. Duhoo, Y., Roche, J., Trinh, T. T. N., Desmyter, A., Gaubert, A., Kellenberger, C., Cambillau, C., Roussel, A., and Leone, P. (2017) Camelid nanobodies used as crystallization chaperones for different constructs of PorM, a component of the type IX secretion system from *Porphyromonas gingivalis*. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **73**, 286–293
6. Kuper, C. F., van Bilsen, J., Cnossen, H., Houben, G., Garthoff, J., and Wolterbeek, A. (2016) Development of immune organs and functioning in humans and test animals: Implications for immune intervention studies. *Reprod. Toxicol.* **64**, 180–190
7. Gollwitzer, E. S., and Marsland, B. J. (2015) Impact of Early-Life Exposures on Immune Maturation and Susceptibility to Disease. *Trends Immunol.* **36**, 684–696
8. Diefenbach, A., and Raulet, D. H. (1999) Natural killer cells: stress out, turn on, tune in. *Curr. Biol.* **9**, R851–R853
9. Xiong, Y. (2016) *Impact du G-CSF sur le phénotype et les fonctions des cellules NK dans le cadre d'une immunothérapie post-allogreffe de cellules souches hématopoïétiques*. Ph.D. thesis, Université de Lorraine, [online] <https://tel.archives-ouvertes.fr/tel-01441455/> (Accessed June 14, 2017)
10. Li, P., Morris, D. L., Willcox, B. E., Steinle, A., Spies, T., and Strong, R. K. (2001) Complex structure of the activating immunoreceptor NKG2D and its MHC class I-like ligand MICA. *Immunol. Nature.* [online] http://www.nature.com/insb.bib.cnrs.fr/ni/journal/v2/n5/pdf/ni0501_443.pdf (Accessed June 13, 2017)
11. Li, P., Willie, S. T., Bauer, S., Morris, D. L., Spies, T., and Strong, R. K. (1999) Crystal structure of the MHC class I homolog MIC-A, a gamma delta T cell ligand. *Immunity.* **10**, 577–584
12. Holmes, M. A., Li, P., Petersdorf, E. W., and Strong, R. K. (2002) Structural Studies of Allelic Diversity of the MHC Class I Homolog MIC-B, a Stress-Inducible Ligand for the Activating Immunoreceptor NKG2D. *J. Immunol.* **169**, 1395–1400
13. Ashiru, O., Boutet, P., Fernandez-Messina, L., Aguera-Gonzalez, S., Skepper, J. N., Vales-Gomez, M., and Reyburn, H. T. (2010) Natural Killer Cell Cytotoxicity Is Suppressed by Exposure to the Human NKG2D Ligand MICA*008 That Is Shed by Tumor Cells in Exosomes. *Cancer Res.* **70**, 481–489
14. Blery, M., Gauthier, L., PERROT, I., and BONNAFOUS, C. (2013) Mica binding agents. [online] <http://www.google.dj/patents/WO2013117647A1> (Accessed June 27, 2017)
15. Vincentelli, R., Canaan, S., Campanacci, V., Valencia, C., Maurin, D., Frassinetti, F., Scappucini-Calvo, L., Bourne, Y., Cambillau, C., and Bignon, C. (2009) High-throughput automated refolding screening of inclusion bodies. *Protein Sci.* **13**, 2782–2792

16. Kellenberger, C., Porciero, S., and Roussel, A. (2004) Expression, refolding, crystallization and preliminary crystallographic study of MHC H-2K^k complexed with octapeptides and nonapeptides. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1278–1280
17. D'Arcy, A., Bergfors, T., Cowan-Jacob, S. W., and Marsh, M. (2014) Microseed matrix screening for optimization in protein crystallization: what have we learned? *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **70**, 1117–1126
18. Svergun, D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Cryst.* **25**, 495–503
19. Franke, D., and Svergun, D. I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346
20. Malia, T. J., Obmolova, G., Luo, J., Teplyakov, A., Sweet, R., and Gilliland, G. L. (2011) Crystallization of a challenging antigen-antibody complex: TLR3 ECD with three noncompeting Fabs. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* **67**, 1290–1295
21. Green, E. R., and Meccas, J. (2016) Bacterial Secretion Systems: An Overview. in *Virulence Mechanisms of Bacterial Pathogens, Fifth Edition* (Kudva, I. T., Cornick, N. A., Plummer, P. J., Zhang, Q., Nicholson, T. L., Bannantine, J. P., and Bellaire, B. H. eds), pp. 215–239, American Society of Microbiology, 10.1128/microbiolspec.VMBF-0012-2015
22. Natale, P., Brüser, T., and Driessen, A. J. M. (2008) Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—Distinct translocases and mechanisms. *Biochim. Biophys. Acta BBA - Biomembr.* **1778**, 1735–1756
23. Papanikou, E., Karamanou, S., and Economou, A. (2007) Bacterial protein secretion through the translocase nanomachine. *Nat. Rev. Microbiol.* **5**, 839–851
24. Luirink, J., and Sinning, I. (2004) SRP-mediated protein targeting: structure and function revisited. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1694**, 17–35
25. Costa, T. R. D., Felisberto-Rodrigues, C., Meir, A., Prevost, M. S., Redzej, A., Trokter, M., and Waksman, G. (2015) Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat. Rev. Microbiol.* **13**, 343–359
26. Thomas, S., Holland, I. B., and Schmitt, L. (2014) The Type 1 secretion pathway — The hemolysin system and beyond. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1843**, 1629–1641
27. Linhartová, I., Bumba, L., Mašín, J., Basler, M., Osička, R., Kamanová, J., Procházková, K., Adkins, I., Hejnová-Holubová, J., Sadílková, L., Morová, J., and Šebo, P. (2010) RTX proteins: a highly diverse family secreted by a common mechanism. *Fems Microbiol. Rev.* **34**, 1076–1112
28. Gu, S., Shevchik, V. E., Shaw, R., Pickersgill, R. W., and Garnett, J. A. (2017) The role of intrinsic disorder and dynamics in the assembly and function of the type II secretion system. *Biochim. Biophys. Acta BBA - Proteins Proteomics.* **1865**, 1255–1266
29. Diepold, A., and Wagner, S. (2014) Assembly of the bacterial type III secretion machinery. *FEMS Microbiol. Rev.* **38**, 802–822
30. Merino, E., Flores-Encarnación, M., and Aguilar-Gutiérrez, G. R. (2017) Functional interaction and structural characteristics of unique components of *Helicobacter pylori* T4SS. *FEBS J.* 10.1111/febs.14092
31. Abby, S. S., Cury, J., Guglielmini, J., Néron, B., Touchon, M., and Rocha, E. P. C. (2016) Identification of protein secretion systems in bacterial genomes. *Sci. Rep.* 10.1038/srep23080
32. Cascales, E., and Journet, L. (2016) The Type VI Secretion System in *Escherichia coli* and Related Species. *EcoSal Plus.* 10.1128/ecosalplus.ESP-0009-2015

33. Durand, E., Nguyen, V. S., Zoued, A., Logger, L., Péhau-Arnaudet, G., Aschtgen, M.-S., Spinelli, S., Desmyter, A., Bardiaux, B., Dujeancourt, A., Roussel, A., Cambillau, C., Cascales, E., and Fronzes, R. (2015) Biogenesis and structure of a type VI secretion membrane core complex. *Nature*. **523**, 555–560
34. Sato, K., Sakai, E., Veith, P. D., Shoji, M., Kikuchi, Y., Yukitake, H., Ohara, N., Naito, M., Okamoto, K., Reynolds, E. C., and Nakayama, K. (2005) Identification of a New Membrane-associated Protein That Influences Transport/Maturation of Gingipains and Adhesins of *Porphyromonas gingivalis*. *J. Biol. Chem.* **280**, 8668–8677
35. Sato, K., Naito, M., Yukitake, H., Hirakawa, H., Shoji, M., McBride, M. J., Rhodes, R. G., and Nakayama, K. (2010) A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proc. Natl. Acad. Sci.* **107**, 276–281
36. McBride, M. J., and Zhu, Y. (2013) Gliding Motility and Por Secretion System Genes Are Widespread among Members of the Phylum Bacteroidetes. *J. Bacteriol.* **195**, 270–278
37. Sato, K. (2011) Por secretion system of *Porphyromonas gingivalis*. *J. Oral Biosci.* **53**, 187–196
38. Kharade, S. S., and McBride, M. J. (2014) *Flavobacterium johnsoniae* Chitinase ChiA Is Required for Chitin Utilization and Is Secreted by the Type IX Secretion System. *J. Bacteriol.* **196**, 961–970
39. Veith, P. D., Glew, M. D., Gorasia, D. G., and Reynolds, E. C. (2017) Type IX Secretion: the generation of bacterial cell surface coatings involved in virulence, gliding motility and the degradation of complex biopolymers: The Type IX Secretion System. *Mol. Microbiol.* 10.1111/mmi.13752
40. McBride, M. J., and Nakane, D. (2015) *Flavobacterium* gliding motility and the type IX secretion system. *Curr. Opin. Microbiol.* **28**, 72–77
41. Nakane, D., Sato, K., Wada, H., McBride, M. J., and Nakayama, K. (2013) Helical flow of surface protein required for bacterial gliding motility. *Proc. Natl. Acad. Sci.* **110**, 11145–11150
42. Lasica, A. M., Ksiazek, M., Madej, M., and Potempa, J. (2017) The Type IX Secretion System (T9SS): Highlights and Recent Insights into Its Structure and Function. *Front. Cell. Infect. Microbiol.* 10.3389/fcimb.2017.00215
43. Socransky, S. s., Haffajee, A. d., Cugini, M. a., Smith, C., and Kent, R. L. (1998) Microbial complexes in subgingival plaque. *J. Clin. Periodontol.* **25**, 134–144
44. Avila, M., Ojcius, D. M., and Yilmaz, Ö. (2009) The Oral Microbiota: Living with a Permanent Guest. *DNA Cell Biol.* **28**, 405–411
45. Bodic, F., Hamel, L., Lerouxel, E., Baslé, M. F., and Chappard, D. (2005) Bone loss and teeth. *Joint Bone Spine.* **72**, 215–221
46. Sochalska, M., and Potempa, J. (2017) Manipulation of Neutrophils by *Porphyromonas gingivalis* in the Development of Periodontitis. *Front. Cell. Infect. Microbiol.* 10.3389/fcimb.2017.00197
47. Ruby, J., and Barbeau, J. (2002) The buccale puzzle: the symbiotic nature of endogenous infections of the oral cavity. *Can. J. Infect. Dis. Med. Microbiol.* **13**, 34–41
48. Borojevic, T. (2012) Smoking and Periodontal Disease. *Mater. Socio Medica.* **24**, 274
49. Vincent, M. S., Canestrari, M. J., Leone, P., Stathopoulos, J., Ize, B., Zoued, A., Cambillau, C., Kellenberger, C., Roussel, A., and Cascales, E. (2017) Characterization of the *Porphyromonas gingivalis* Type IX Secretion Trans-envelope PorKLMNP Core Complex. *J. Biol. Chem.* **292**, 3252–3261
50. Vincent, M. S., Durand, E., and Cascales, E. (2016) The PorX Response Regulator of the *Porphyromonas gingivalis* PorXY Two-Component System Does Not Directly

- Regulate the Type IX Secretion Genes but Binds the PorL Subunit. *Front. Cell. Infect. Microbiol.* 10.3389/fcimb.2016.00096
51. Gorasia, D. G., Veith, P. D., Hanssen, E. G., Glew, M. D., Sato, K., Yukitake, H., Nakayama, K., and Reynolds, E. C. (2016) Structural Insights into the PorK and PorN Components of the Porphyromonas gingivalis Type IX Secretion System. *PLOS Pathog.* **12**, e1005820
 52. Roman, S. J., Meyers, M., Volz, K., and Matsumura, P. (1992) A chemotactic signaling surface on CheY defined by suppressors of flagellar switch mutations. *J. Bacteriol.* **174**, 6247–6255
 53. Stathopoulos, J., Cambillau, C., Cascales, E., Roussel, A., and Leone, P. (2015) Crystallization and preliminary X-ray analysis of the C-terminal fragment of PorM, a subunit of the Porphyromonas gingivalis type IX secretion system. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **71**, 71–74
 54. Shrivastava, A., Lele, P. P., and Berg, H. C. (2015) A Rotary Motor Drives Flavobacterium Gliding. *Curr. Biol.* **25**, 338–341
 55. Desmyter, A., Spinelli, S., Roussel, A., and Cambillau, C. (2015) Camelid nanobodies: killing two birds with one stone. *Curr. Opin. Struct. Biol.* **32**, 1–8
 56. Aslanidis, C., and de Jong, P. J. (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* **18**, 6069–6074
 57. Garron, M.-L., Arthos, J., Guichou, J.-F., McNally, J., Cicala, C., and Arold, S. T. (2008) Structural Basis for the Interaction between Focal Adhesion Kinase and CD4. *J. Mol. Biol.* **375**, 1320–1328
 58. Erijman, A., Dantes, A., Bernheim, R., Shifman, J. M., and Peleg, Y. (2011) Transfer-PCR (TPCR): A highway for DNA cloning and protein engineering. *J. Struct. Biol.* **175**, 171–177
 59. The Besley Group – University of Nottingham [online] <http://besley.chem.nottingham.ac.uk/research/research-prospec.html> (Accessed August 31, 2017)
 60. Pall ForteBio :: BLI Technology [online] <https://www.fortebio.com/bli-technology.html> (Accessed August 30, 2017)
 61. Rupp, B. (2017) Rupp_crystallization_Granada_2017.pdf
 62. Solution Scattering Data Assessment [online] <http://www-ssrl.slac.stanford.edu/~saxs/analysis/assessment.htm> (Accessed August 31, 2017)
 63. Sévajol, M. (2013) Cours_SAXS_M1_MAS1_2013.pdf
 64. Pernot, P., Round, A., Barrett, R., Antolinos, A. D. M., Gobbo, A., Gordon, E., Huet, J., Kieffer, J., Lentini, M., Mattenet, M., Morawe, C., Mueller-Dieckmann, C., Ohlsson, S., Schmid, W., Surr, J., Theveneau, P., Zerrad, L., and McSweeney, S. (2013) Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J. Synchrotron Radiat.* **20**, 660
 65. Durand, D., Vivès, C., Cannella, D., Pérez, J., Pebay-Peyroula, E., Vachette, P., and Fieschi, F. (2010) NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. *J. Struct. Biol.* **169**, 45–53
 66. Oliveira, C. L. P. (2011) Investigating macromolecular complexes in solution by small angle X-ray scattering. in *Current trends in X-ray Crystallography*, InTech, [online] <https://www.intechopen.com/download/pdf/25147> (Accessed July 10, 2017)
 67. Comeau, S. R., Gatchell, D. W., Vajda, S., and Camacho, C. J. (2004) ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics.* **20**, 45–50
 68. Comeau, S. R., Gatchell, D. W., Vajda, S., and Camacho, C. J. (2004) ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Res.* **32**, W96–W99

69. Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., and Vajda, S. (2017) The ClusPro web server for protein–protein docking. *Nat. Protoc.* **12**, 255–278
70. Pinker, F., Brun, M., Morin, P., Deman, A.-L., Chateaux, J.-F., Oliéric, V., Stirnimann, C., Lorber, B., Terrier, N., Ferrigno, R., and Sauter, C. (2013) ChipX: A Novel Microfluidic Chip for Counter-Diffusion Crystallization of Biomolecules and in Situ Crystal Analysis at Room Temperature. *Cryst. Growth Des.* **13**, 3333–3340

Résumé

Pour lutter contre des micro-organismes pathogènes ou certaines maladies, le système immunitaire acquis des vertébrés développe des anticorps. Ces molécules de reconnaissance du non-soi permettent de distinguer spécifiquement des marqueurs antigéniques appelés épitopes. Deux types d'anticorps ont été découverts jusqu'à présents : les anticorps « classiques » et les anticorps de camélidés, également appelés nanobody. Alors que le paratope des premiers est formé par l'association de deux domaines variables, l'un de chaîne lourde et l'autre de chaîne légère, le paratope des seconds n'est constitué que d'un domaine variable de chaîne lourde. Cette thèse porte sur des études structurales de fragments d'anticorps d'intérêt thérapeutique et biotechnologique. Pour cela, deux volets ont été traités : l'étude structurale d'un anticorps thérapeutique ciblant des protéines homologues du CMH de classe I (protéines MIC) et l'utilisation d'un anticorps de camélidés pour l'étude structurale de la protéine périplasmique PorM (pPorM) du système de sécrétion de type 9 de *Porphyromonas gingivalis*.

Au cours du premier volet, j'ai résolu la structure par cristallographie aux rayons X du fragment d'anticorps MIC12 à une résolution de 1,5 Å. Dans le but de résoudre la structure des complexes de MIC12 avec ses antigènes MIC, j'ai obtenu des cristaux diffractant à une résolution allant jusqu'à 7 Å. En parallèle, des analyses par diffusion de rayons X aux petits angles (SAXS) combinées à des prédictions d'interaction par docking ont été conduites afin d'obtenir une première description de la région globale d'interaction de MIC12 sur l'une de ses cibles MIC.

Concernant le second volet, 4 nanobody ont été obtenus à partir de l'immunisation d'un Lama contre la protéine pPorM. J'ai résolu la structure par cristallographie de l'un d'entre eux, le nanobody nb02, à une résolution de 1,5 Å. L'utilisation de ces nanobody comme chaperonne de cristallisation m'a permis de résoudre la structure de la partie N-terminale de pPorM. J'ai également mené une étude par SAXS de la protéine pPorM entière. L'ensemble des résultats que j'ai obtenu par cristallographie et par SAXS, combinés aux résolutions des structures des autres domaines de pPorM, ont permis de proposer un modèle structural de la protéine pPorM entière.

Mots clés : Anticorps, Fab, VHH, MIC, T9SS, Cristallographie aux rayons X, SAXS

Abstract

To fight against pathogenic microorganisms or diseases, the adaptive immune system in vertebrates uses antibodies. These non-self recognition molecules help to specifically distinguish antigenic markers called epitopes. Two types of antibody were discovered so far: "classic" antibodies and camelid antibodies, also called nanobodies. While the paratope of the first ones is formed by the association of two variable domains, one from the heavy chain and the other from the light chain, the paratope of the second is only composed of the heavy chain variable domain. This PhD deals with structural studies of antibodies fragments of therapeutic and biotechnological interest. For that purpose, two projects were carried out: the structural study of therapeutic antibody targeting homologous proteins of the MHC class I (MIC proteins) and the use of a camelid antibody for the structural study of the periplasmic protein PorM (pPorM) of the secretion system type 9 from *Porphyromonas gingivalis*.

During the first part, I solved the structure of the MIC12 antibody fragment using X-ray crystallography at a resolution of 1.4 Å. In order to solve the structure of the complexes of MIC12 with its MIC antigens, I obtained crystals of the complexes diffracting at a resolution of up to 7 Å. In parallel, analyses by Small Angles X-ray Scattering (SAXS) combined with in silico docking predictions were led to obtain a first description of the global binding region of MIC12 on one of its MIC targets.

Concerning the second part, 4 nanobodies were obtained from the immunization a Llama against the pPorM protein. I solve the structure of one of them by crystallography, nanobody nb 02, at a resolution of 1.5 Å. Use of these nanobodies as chaperones of crystallization help me to solve the structure of the N-terminal part of pPorM. I also conducted a study by SAXS of the whole pPorM protein. All these results, obtained by crystallography and SAXS studies, combined with the solving of the structures of the other domains of pPorM, made it possible to propose a structural model of the entire pPorM protein.

Keywords : Antibodies, Fab, VHH, MIC, T9SS, X-ray Crystallography, SAXS