



HAL
open science

Nanomachines bactériennes impliquées dans le transfert horizontal de gènes : étude structurale des protéines de l'ICESt3, un élément intégratif et conjugatif de *Streptococcus thermophilus*

Julien Capelle

► **To cite this version:**

Julien Capelle. Nanomachines bactériennes impliquées dans le transfert horizontal de gènes : étude structurale des protéines de l'ICESt3, un élément intégratif et conjugatif de *Streptococcus thermophilus*. Chimie. Université de Lorraine, 2021. Français. NNT : 2021LORR0193 . tel-03632862

HAL Id: tel-03632862

<https://hal.univ-lorraine.fr/tel-03632862v1>

Submitted on 6 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

Nanomachines bactériennes impliquées dans le transfert
horizontal de gènes : étude structurale des protéines de
l'ICESt3, un élément intégratif et conjugatif de
l'organisme *Streptococcus thermophilus*

THÈSE

présentée et soutenue publiquement le 23 Septembre 2021

pour l'obtention du

Doctorat de l'Université de Lorraine
(mention Physique et Chimie des matériaux)

par

Julien Cappèle

Composition du jury

Rapporteurs : Dr. Laurent Terradot
Dr. Vincent Villeret

Examineurs : Pr. Nathalie Leblond-Bourget
Dr. Carole Gardiennet
Dr. Frédérique Favier
Dr. Claude Sauter

Invités : Dr. Pascale Tsan
Dr. Claude Didierjean

Table des matières

Remerciements	
Abréviations	
<u>État de l'art</u>	
I. Introduction	1
II. Les subtilités des systèmes de sécrétion	5
II.1 Sec & Tat	7
II.1.1 La voie Sec	8
II.1.2 La voie Tat	10
II.2 Système de sécrétion de type I	12
II.3 Système de sécrétion de type II	14
II.4 Système de sécrétion de type III	15
II.5 Système de sécrétion de type V	18
II.6 Système de sécrétion de type VI	19
II.7 Système de sécrétion de type VII	22
II.8 Système de sécrétion de type VIII	25
II.9 Système de sécrétion de type IX	27
II.10 Les derniers systèmes de sécrétion	29
III. Le système de sécrétion de type IV	31
III.1 Architecture des systèmes de sécrétion de type IV	33
III.1.1 Système de sécrétion de type IV de classe A	33
III.1.2 Système de sécrétion de type IV de classe B	40
III.2 La conjugaison via un système de sécrétion de type IV	43
III.2.1 Les éléments mobiles conjugatifs	43
III.2.2 Mécanisme chez les bactéries à Gram négatif	46
III.2.3 Mécanisme chez les bactéries à Gram positif	48

IV.	Fonction et structure des principales protéines de la conjugaison	51
IV.1	Relaxases conjugatives	51
IV.2	Protéines de couplage (T4CP)	56
IV.3	L'ATPase VirB4	59
IV.4	Hydrolases de peptidoglycane VirB1	61
IV.5	La protéine structurale VirB8	63
V.	Le système de sécrétion de type IV de l'ICE <i>St3</i> chez <i>Streptococcus thermophilus</i> . . .	67
V.1	Les streptocoques	67
V.1.1	Principales espèces pathogènes	67
V.1.2	Résistances aux antibiotiques	67
V.2	Les éléments intégratifs et conjugatifs chez <i>S. thermophilus</i>	68
V.3	Les protéines conjugatives de l'ICE <i>St3</i>	69
	<u>Objectifs</u>	71
VI.	Objectifs de la thèse	73
	<u>Résultats</u>	75
VII.	Étude structurale de la protéine OrfG	77
VII.1	Article de l'analyse biochimique d'OrfG et de la structure cristallographique de son domaine central OrfG ₆₄₋₂₀₄	77
VII.2	Complément cristallographique à l'article pour le domaine central OrfG ₆₄₋₂₀₄	103
VII.2.1	Collecte des données de diffraction pour la résolution de la structure	103
VII.2.2	Phasage à l'osmium	105
VII.2.3	Blob de densité	110
VII.3	Compléments de résultats obtenus après la publication sur OrfG ₆₄₋₂₀₄	112
VII.4	Étude structurale du domaine soluble OrfG ₆₄₋₃₃₁	117
VII.4.1	La structure d'OrfG ₆₄₋₃₃₁	118
VII.4.2	Oligomérisation	125
VII.4.3	Étude de l'interface	128
VII.5	Discussion	131
VIII.	Étude structurale de la protéine OrfM	133
VIII.1	Production et marquage isotopique	133
VIII.2	Protocoles de production	136

VIII.2.1	Protocole initial de production en milieu riche	136
VIII.2.2	Variation des conditions physico-chimiques	137
VIII.2.3	Protocole hybride	139
VIII.3	Attribution RMN de la protéine OrfM	140
VIII.3.1	Analyse du spectre HSQC- ¹⁵ N	140
VIII.3.2	Attribution des spectres de corrélations hétéronucléaires ¹ H - ¹⁵ N	143
VIII.3.3	Attribution avec l'aide d'expériences triple résonance ¹ H - ¹⁵ N - ¹³ C	145
VIII.3.4	Détermination de la structure secondaire	152
VIII.3.5	Caractérisation dynamique à partir d'expériences de relaxation ¹⁵ N	156
VIII.3.6	Calculs préliminaires de structure	162
VIII.4	Analyse préliminaire du modèle d'OrfM	163
VIII.5	Discussion	167
IX.	MPSA_Viewer	171
IX.1	Limitations de la méthode mTM-align	172
IX.2	Description du fonctionnement de MPSA_Viewer	174
IX.2.1	Mise en place et traitement des données initiales	174
IX.2.2	Traitement des résultats	175
IX.3	Exemples d'utilisation de MPSA_Viewer	176
IX.3.1	Domaines VirB8-like	176
IX.3.2	Glutathion transférases de classe Tau	179
IX.4	Avantages et limitations de MPSA_Viewer	183
IX.4.1	Simplicité	183
IX.4.2	Dépendance et performance	183
IX.5	Discussion	184
X.	Conclusion et Perspectives	187
	<u>Matériel et méthodes</u>	191
XI.	Production et purification des protéines OrfL et OrfM	193
XI.1	Souche et plasmide d'expression	193
XI.2	Composition des milieux	193
XI.3	Protocole initial de production de la protéine 6His-OrfM	194
XI.4	Protocole final de production de la protéine 6His-OrfM	194

XI.5	Purification de la protéine OrfM	195
XII.	Résonance magnétique nucléaire	199
XII.1	Préparation de l'échantillon	199
XII.2	Collecte des spectres RMN et attribution	199
XII.3	Calcul de structure préliminaire	200
XIII.	Résolution de structure cristallographiques par rayons X	200
XIII.1	Cristallisation des protéines OrfG ₆₄₋₂₀₄ et OrfG ₆₄₋₃₃₁	200
XIII.2	Collecte des données de diffraction et résolution de la structure d'OrfG ₆₄₋₃₃₁	201
Annexes	203
Bibliographie	225

Remerciements

Je tiens à commencer en remerciant le Pr. Dominik Schaniel pour m'avoir accepté au sein du laboratoire CRM² (Cristallographie, Résonance magnétique et Modélisation ; UMR CNRS 7036) ainsi que l'école Doctorale C2MP (Chimie-Mécanique-Matériaux-Physique) pour m'avoir choisi et fait confiance dans la réalisation de ces travaux.

Je remercie le Dr. Frédérique Favier et le Dr. Claude Didierjean de m'avoir encadré durant mon stage de Master 2 et durant cette thèse. Votre soutien sans faille, votre expertise et vos encouragements ont été déterminants sur toute la durée de cette superbe collaboration. En particulier les très longues discussions assez souvent hors-sujet qui nous avons eues !

Je remercie chaleureusement le Dr. Pascale Tsan pour son aide durant ces 3 années concernant la RMN. Ta patience, ton enthousiasme constant et ton expertise ont grandement participé à la réussite du projet dans lequel nous nous sommes lancés.

J'exprime toute ma gratitude au Dr. Laurent Terradot et au Dr. Vincent Villeret qui ont accepté de juger mon travail en tant que Rapporteurs de cette thèse ainsi qu'au Pr. Nathalie Leblond-Bourget et le Dr. Claude Sauter pour avoir pris part à mon jury.

Je tiens à remercier tout particulièrement le Dr. Carole Gardiennet pour avoir accepté de participer à ce jury et pour m'avoir accompagné durant ces 3 années avec le Dr. Christophe Charon que je remercie également. Vos conseils avisés m'ont réellement permis de progresser.

Je veux également remercier très chaleureusement le Pr. Nathalie Leblond-Bourget, le Dr. Nicolas Soler et le Dr. Badreddine Douzi pour cette exceptionnelle collaboration entre nos laboratoires ainsi que le Pr. Bertrand Aigle pour m'avoir permis de réaliser les expériences nécessaires dans le laboratoire DynAMic durant cette collaboration.

Je souhaite remercier tous mes camarades de classe, collègues, stagiaires et thésards qui ont partagé ma vie professionnelle (et perso avec les fêtards !) depuis que j'ai commencé à m'intéresser à la Science. Vous êtes bien trop nombreux pour tous vous nommer mais je vous remercie du fond du coeur et je vous souhaite très sincèrement de réussir dans la vie et d'en profiter au maximum.

Je tiens à remercier mes amis (pas dans l'ordre, ne vous prenez pas le chou) Xavier, Dindon,

Ianis, Romane, Pierre, Léo, Laura, Théo, Amandine, Abigail, Gillan, Mathilde, Mélodie, Lysis, Antoine, Pauline et Quentin. Merci de m'avoir supporté, saoulé (dans le bon sens du terme) et accompagné dans ma vie. Elle serait bien fade sans vous mes chers amis. Si je vous ai oublié ici, j'en suis navré et je pense fort à vous quand même.

Mes dernières pensées vont à ma famille, à ma maman et à mon papa, mon frangin, ma belle-famille et à tous mes neveux et nièces. Vous êtes tous extraordinaires et je n'aurais jamais réussi sans vous.

Julie, ma chère et tendre. Je ne pourrai jamais te remercier assez pour tout ce que tu as fait pour moi. Quand je pense que tu as même proposé de relire ma thèse! N'est-ce pas une véritable preuve d'amour?

Abréviations

- ADN** acide désoxyribonucléique
- ADN-T** ADN de transfert
- AT** *Auto Transporter proteins*, protéines auto-transporteuses
- ATP** adénosine tri-phosphate
- ATPase** adénosine tri-phosphate synthase
- BCG** vaccin contre le Bacille de Calmette et Guérin (tuberculose)
- CC** *Core Complex*, complexe de coeur
- CD** *Central Domain*, domaine central
- CHAP** cystéine-histidine amidohydrolase/péptidase dépendent
- Cryo-EM** *Cryo electron microscopy*, cryo-microscopie électronique
- CSI** *Chemical Shift Index*, répertoire des déplacements chimiques
- CTD** *C-terminal Domain*, domaine c-terminal
- Dot** *Defective in organelle trafficking*, defectueux dans le trafic d'organelle
- HP** hydrolase de peptidoglycane
- Icm** *Intracellular multiplication*, multiplication intracellulaire
- IHF** *Integration Host Factor*, facteur d'intégration de l'hôte
- IM** *Internal Membrane*, membrane interne
- IMC** *Internal Membrane Complex*, complexe de la membrane interne
- IVA** système de sécrétion de type IV A (ou T4ASS)
- IVB** système de sécrétion de type IV B (ou T4BSS)
- OM** *Outer Membrane*, membrane externe
- OMC** *Outer Membrane Complex*, complexe de la membrane externe
- RMN** résonance magnétique nucléaire
- SAD** *Single-wavelength Anomalous Diffusion*, diffusion anormale à longueur d'onde unique
- Sec** *General Secretion Pathway*, système de sécrétion général
- SEC-MALS** *Size Exclusion Chromatography MultiAngle Light Scattering*, chromatographie d'exclusion de taille, diffusion de lumière à angle multiple
- SRP** *Signal Recognition Particle*, particule de reconnaissance du signal
- T4CC** *Type IV Coupling Complex*, complexe de couplage de type IV
- T4P** *Type IV Pili*, pili de type IV
- Tat** *Twin-Arginine Translocation*, système de translocation à arginines jumelles
- Ti** *Tumor inducing*, inducteur de tumeur
- TMD** *TransMembrane Domain*, domaine transmembranaire
- TPS** *Two-partners secretion*, sécrétion à deux partenaires
- TxSS** *Type x Secretion System*, système de sécrétion de type x

ÉTAT DE L'ART

I Introduction

Dans un contexte sanitaire qui se dégrade, l'humanité doit faire face à de nombreux nouveaux dangers d'origine invisible. Dès le 17^{ème} siècle, la découverte des premiers micro-organismes a changé la façon dont les maladies étaient perçues. Autrefois de provenance inconnue, de nombreuses maladies sont alors attribuées à des agents pathogènes bactériens qui reçoivent des noms, une classification complète et surtout, une façon de les combattre tout au long du 19^{ème} siècle avec la contribution des médecins (et microbiologistes) du moment, comme Louis Pasteur, Robert Koch et bien d'autres. Il restera toutefois des maladies sans origine identifiée, comme les maladies virales ou encore certains empoisonnements aux métaux, impossibles à diagnostiquer avec les moyens de l'époque. Il a fallu attendre la première moitié du 20^{ème} siècle pour découvrir l'existence de substances qui ralentissent la croissance des bactéries, voire qui l'annulent complètement en fonction des doses. De façon erronée, dans la conscience collective on attribue ce succès à l'Institut Louis Pasteur. Bien que les chercheurs de l'institut aient joué un rôle majeur et même décisif dans l'élaboration et l'administration des traitements antibiotiques, la découverte en revient à Alexander Fleming, qui à la fin des années 30, remarque que l'apparition d'une moisissure a réduit significativement la croissance des bactéries dans une gélose nutritive. Ce champignon, un *Penicillium* du même genre que le champignon du Roquefort, secrète une molécule, la pénicilline, qui inhibe la création de la paroi bactérienne. Les bactéries à Gram positif sont particulièrement sensibles à cette molécule, possédant une très épaisse paroi de peptidoglycane et ne pouvant pas se dupliquer si elle est absente. La pénicilline est une molécule cyclique appartenant à la classe des beta-lactames, et existe sous plusieurs formes, comme la pénicilline G ou encore la benzyl-pénicilline.

Après plus de 80 ans de traitements antibiotiques, le nombre de personnes sauvées des griffes des infections bactériennes est gigantesque. Prenons l'exemple de la peste noire qui a dévasté le continent Européen pendant le Moyen Âge en réduisant la population de 30 à 50%, soit 25 millions de victimes. La bactérie responsable, *Yersinia pestis*, n'a pas été médicalement recensée en France depuis les années 40, et ce parce que les traitements antibiotiques sont nombreux, et sont administrés à temps, ne permettant plus à la maladie de sévir sur notre territoire. Malheureusement, ces deux assertions ne sont pas effectives dans tous les pays. Actuellement, la peste continue de tuer en Asie, en Amérique et en Afrique, dans les zones rurales majoritairement.

Comme tout bon organisme à croissance rapide, les bactéries savent s'adapter à leur environnement, si tant est qu'on leur en laisse la possibilité. Les antibiotiques primordiaux faisant partie intégrante de l'environnement, étant des substances naturellement présentes et sécrétées par d'autres organismes, les bactéries ont eu largement le temps de s'y adapter et de trouver des moyens pour survivre en leur présence. Bien que les modifications apportées sur les antibiotiques soient nombreuses pour améliorer leur potentiel, de plus en plus de résistances apparaissent ou réapparaissent aujourd'hui. Malheureusement, depuis quelques années l'on découvre plus de résistances que de nouveaux antibiotiques à étudier. Si l'on reprend l'exemple des pénicillines, les bactéries résistantes utilisent des enzymes appelées des beta-lactamases, qui sont spécialisées dans la dénaturation des beta-lactames et donc des pénicillines, réduisant significativement leur effet. La réponse à cette résistance est généralement l'augmentation de la dose administrée. Mais comme toute substance chimique peut devenir poison, l'augmentation de la dose qui permettrait de tuer la bactérie résistante peut aussi signifier la mort du patient.

Pour ajouter à ces nouvelles alarmantes, les résistances observées peuvent relever de plusieurs mécanismes. Parfois, cachés dans le génome des bactéries se trouvent des gènes silencieux qui peuvent être réactivés. De manière plus rare, des modifications spontanées (mutations) peuvent être observées sur un gène existant, changeant son rôle et permettant à une nouvelle protéine issue de ce gène de dégrader un nouvel antibiotique du milieu. Bien entendu, comme la bactérie résistante pourra s'adapter à son environnement, c'est elle qui se reproduira le plus, et donc deviendra la souche majoritaire. Enfin, la façon la plus facile et la plus utilisée pour acquérir une résistance pour une bactérie, c'est de l'échanger avec ses voisins. Le transfert horizontal de gène permet ainsi la dissémination des gènes entre deux cellules sans qu'il y ait eu reproduction, au contraire du transfert vertical de gène se déroulant lorsqu'une cellule mère transmet ses gènes à une cellule fille. Naturellement ces échanges ont lieu de façon très limitée, car cela coûte énormément à une cellule de mettre en place ce système d'échange. On parle ici d'une machinerie de conjugaison impliquant plusieurs dizaines de protéines au minimum. Toutefois, si le milieu devient hostile, cet échange s'accélère pour multiplier les chances que les gènes échangés permettent la survie d'un organisme (et sa descendance).

Il paraît plus qu'important de multiplier les efforts concernant les résistances aux antibiotiques. Non seulement les résistances progressent, mais elles s'additionnent, engendrant des "super bactéries" qui peuvent être résistantes à deux, trois, voire une dizaine d'antibiotiques dans les cas les

plus graves. Découvrir de nouveaux antibiotiques reste une voie valable, car si certains antibiotiques sont plus favorables aux résistances que d'autres, leurs modifications chimiques avancées peuvent souvent augmenter leur efficacité. Étudier en parallèle la façon dont ces résistances peuvent être mises en place, échangées et régulées grâce aux systèmes de sécrétion est une seconde voie tout à fait pertinente, car elle pourrait mener à des traitements complémentaires.

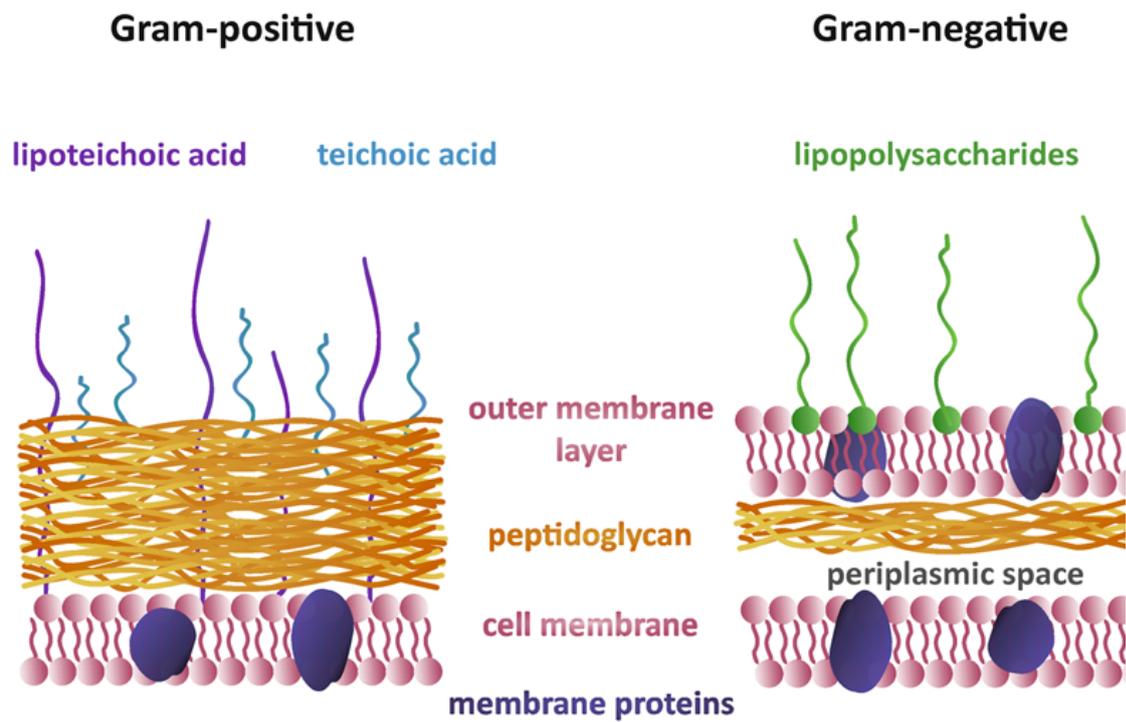
Dans ce manuscrit de thèse, les études structurales de protéines d'un système de sécrétion de type IV sont présentées. Après une revue des principaux systèmes de sécrétion connus à ce jour, le système de sécrétion de type IV qui est au coeur de la conjugaison entre bactéries est détaillé. Ce système est le moyen le plus utilisé pour le transfert horizontal de gènes. Le T4SS étudié dans le cadre de ce travail provient d'un élément intégratif et conjugatif, à savoir l'élément mobile *ICESt3* de *Streptococcus thermophilus*. A la suite de la présentation de l'élément *ICESt3*, mes travaux en cristallographie et en RMN de protéines de cet ICE sont exposés. Enfin leur implication individuelle dans l'environnement d'un système de sécrétion de type IV chez les bactéries à Gram positif est discuté.

II Les subtilités des systèmes de sécrétion

Parmi les grands types de molécules utilisées par le vivant, les protéines sont à la fois ouvrières et outils. Elles peuvent servir de catalyseur chimique en favorisant et accélérant des réactions qui sont très lentes dans les conditions usuelles. Elles ont aussi le pouvoir de structurer les cellules, de leur permettre de se mouvoir, de réguler l'activité d'autres protéines, de transporter des molécules, et bien d'autres fonctions remarquables. Il est donc logique de retrouver abondamment ces molécules tant dans les organismes simples que complexes, au sein des cellules ou dans leur voisinage. Toutefois, les protéines et les autres molécules du vivant ne peuvent pas, sauf exception comme les lipides, traverser aisément la membrane complexe d'une cellule. En effet, cette membrane est beaucoup plus élaborée qu'une simple bicouche lipidique : sa résistance accrue découle de sa nature très structurée, impliquant de nombreuses protéines, des lipides complexes et divers sucres (voir Figure 1, page 6). Chez certaines cellules, comme les bactéries à Gram négatif, se trouvent deux membranes séparées par un interstice appelé le périplasme, où se loge une fine mais résistante couche de peptidoglycane (5 nm). Chez les bactéries à Gram positif, une seule membrane est présente mais avec une très épaisse couche de peptidoglycane (30-100 nm) à l'extérieur, lui permettant de se protéger de son environnement (Clark *et al.*, 2018 ; Pajerski *et al.*, 2019 ; Silhavy *et al.*, 2010).

Comment des molécules peuvent-elles être transportées au travers d'une membrane ? La membrane étant constituée majoritairement de molécules hydrophobes, les petites molécules hydrophobes sont capables de la traverser par diffusion. En revanche, les ions et les molécules hydrophiles ne peuvent la traverser et doivent emprunter des transporteurs ou des pores. Il existe différents types de transporteurs pour prendre en charge à peu près toutes les molécules. Ceux qui nous intéressent ici sont les transporteurs de très grandes molécules comme les acides nucléiques ou protéines. De par leur très grande taille, ces molécules utilisent divers systèmes pour traverser les membranes. Ces systèmes peuvent être à la fois simples et complexes. Ainsi, on retrouve par exemple des petites protéines qui sont directement produites avec leur transporteur au sein de leur séquence primaire, ou bien des complexes multi-protéiques de plusieurs millions de Daltons, appelés systèmes de sécrétion, qui s'ancrent dans la membrane afin de transporter leurs substrats. Il est important de préciser que plus d'un tiers des protéines cellulaires vont être transportées au sein et au-delà des membranes.

FIGURE 1 – Différence entre les membranes des bactéries à Gram positif et négatif



tiré de Pajerski *et al.*, 2019

Représentation schématique des membranes des bactéries à Gram positif (gauche) et négatif (droite). La paroi des bactéries à Gram positif se caractérise par une seule membrane constituée d'une bicouche phospholipidique. Une épaisse couche de peptidoglycane offre une excellente protection contre l'environnement. Chez les bactéries à Gram négatif, la couche de peptidoglycane est bien plus fine et se trouve au niveau du périplasme entre la membrane plasmique et une deuxième membrane appelée membrane externe. Dans les deux cas, les membranes sont riches en protéines et lipides simples et/ou complexes. La structuration différente des membranes induit une différence inévitable dans les systèmes de transport de molécules chez ces organismes.

Ces systèmes sont non seulement importants, ils peuvent parfois être essentiels à la survie de ces organismes.

Dans les paragraphes suivants, nous nous concentrerons en particulier sur les systèmes de sécrétion chez les bactéries, qui permettent aux organismes qui les possèdent de déplacer les grandes molécules depuis leur cytoplasme (environnement interne) vers un autre milieu. Parmi ces milieux, nous pouvons mentionner les compartiments cellulaires, l'environnement extérieur ou même une autre cellule (eucaryote ou bactérienne). Nous distinguerons également les systèmes en fonction de leur appartenance aux bactéries à Gram positif et négatif, car les différences structurales de leurs membranes/parois impliquent des variations dans la structure et le mécanisme de ces systèmes, augmentant une fois de plus la complexité. Cette partie sur les systèmes de sécrétion bactériens ne sera pas exhaustive, mais tentera d'apporter des informations sur les systèmes les plus connus à l'exception du système de sécrétion de type IV qui fera l'objet d'un chapitre particulier.

II.1 Sec & Tat

Commençons avec les systèmes Sec et Tat, respectivement *General Secretion Pathway*, système de sécrétion général (Sec) et *Twin-Arginine Translocation*, système de translocation à arginines jumelles (Tat). Ces deux modes de sécrétion sont largement utilisés par la plupart des organismes vivants et sont retrouvés chez les eucaryotes, procaryotes et archées. Leur rôle est similaire, ils vont transporter une protéine au travers de la membrane cytoplasmique. Ce qui les différencie, ce sont leurs caractéristiques structurales et le fait que l'un est capable de prendre en charge les protéines non repliées, l'autre étant spécifique des protéines repliées (Natale *et al.*, 2008). Ces deux systèmes reconnaissent leurs substrats protéiques lorsqu'ils portent un peptide signal. Ce dernier est une succession d'acides aminés situés à l'extrémité N-terminale et qui est spécifique de la future localisation de la protéine car il sera reconnu par les transporteurs qui lui sont associés.

Ces systèmes sont généraux dans le sens où ils vont accompagner la très grande majorité des protéines membranaires et extra-cellulaires (plus ou moins 90%) vers leur destination. Lorsque les protéines sont membranaires, elles peuvent être à leur tour des transporteurs membranaires ou bien des récepteurs tandis que les protéines complètement excrétées peuvent être des toxines, des enzymes importantes pour le métabolisme ou bien de petits peptides permettant la communication

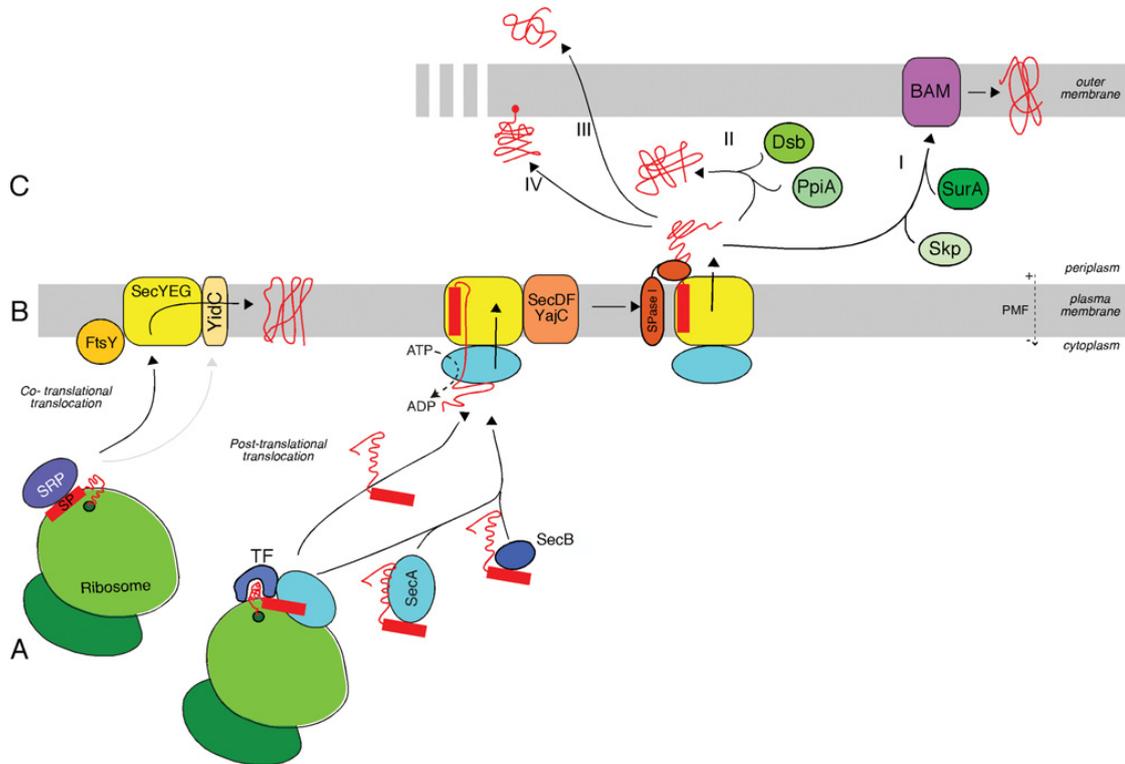
entre les différentes bactéries environnantes. Il est également possible que ces systèmes servent d'intermédiaires pour un autre système de sécrétion, si nécessaire.

II.1.1 La voie Sec

La voie Sec est l'un des systèmes de sécrétion les plus importants et le mieux étudié. Il est généralement essentiel à la viabilité cellulaire chez les procaryotes, c'est pourquoi on le retrouve chez quasiment toutes les bactéries avec quelques améliorations et/ou additions chez certaines d'entre elles (Chatzi *et al.*, 2013). La voie Sec est également utilisée dans le transport de facteurs de virulences chez certaines bactéries pathogènes (Bensing & Sullam, 2002 ; Braunstein *et al.*, 2003 ; Korotkov *et al.*, 2021 ; Lenz *et al.*, 2003). La fonction principale de la voie Sec permet l'export de protéines non repliées depuis le cytoplasme vers la membrane plasmique. Ces fragments protéiques sont ensuite intégrés latéralement dans la membrane, ou bien transportés au-delà de la membrane, que cela soit le périplasma chez les bactéries à Gram négatif ou bien vers le milieu extérieur dans le cas des bactéries à Gram positif. Bien que l'épaisse couche de peptidoglycane chez les bactéries à Gram positif soit complexe et protège efficacement la bactérie des différentes contraintes environnementales, elle reste perméable aux protéines car il n'y a pas de deuxième membrane comme pour les bactéries à Gram négatif.

Compte tenu de la quantité de protéines qu'il transporte, le système Sec est complexe, impliquant de nombreux éléments. Tout d'abord, les protéines non repliées qui sont en route vers la membrane plasmique pour être exportées doivent impérativement rester non repliées et solubles. On appelle ces molécules des pré-protéines, car elles n'ont pas encore atteint leur structure fonctionnelle. Toutes les protéines transportées par la voie Sec sont initialement des pré-protéines et c'est ici que le peptide signal joue un rôle majeur (Gouridis *et al.*, 2009). Essentiellement constitué de résidus hydrophobes, il est situé au niveau de l'extrémité N-terminale et permet un ciblage et un tri spécifique des protéines, selon leur fonction et leur localisation cellulaire (Papanikou *et al.*, 2007). Dans le cas de la voie Sec, c'est la translocase/ATPase SecA qui reconnaît spécifiquement le peptide signal. La translocase est couplée au canal SecYEG, et enclenche le transport en consommant de l'adénosine tri-phosphate (ATP). Chez les eucaryotes, SecYEG est connue sous le nom de SEC61, et il existe des systèmes paralogues chez les procaryotes (SecA2/SecY2) (Papanikou *et al.*, 2007).

FIGURE 2 – La voie de sécrétion Sec



tiré de Chatzi *et al.*, 2013

Représentation schématique des modules principaux du système Sec ainsi que du processus de sécrétion. (A) Pour les protéines membranaires, l'hélice membranaire néoformée qui émerge du ribosome est recrutée par la SRP ("Signal Recognition Particle", particule de reconnaissance du signal). Le complexe protéique est transféré vers le canal SecYEG à l'aide du récepteur à SRP, FtsY. Certaines protéines peuvent directement emprunter le chemin grisé via YidC. Dans le cas des protéines non membranaires, elles sont reconnues par la protéine TF ("Trigger Factor", facteur déclencheur) et seront prises en charge par SecA (soluble ou liée au canal SecYEG). Certaines chaperonnes comme SecB peuvent aider cette prise en charge. Note : SecA peut participer aux translocations des protéines membranaires lorsque celles-ci possèdent de grandes régions hydrophiles. (B) Une fois que le complexe entre SRP et la protéine membranaire atteint SecYEG, la pré-protéine entre dans la membrane au travers du canal pendant qu'elle est synthétisée, avec l'aide de YidC. Dans le cas de la protéine non membranaire, la pré-protéine interagit avec SecA, elle-même liée au canal SecYEG. SecA utilise l'ATP comme source d'énergie pour transférer la protéine au travers du canal. La protéine SecDF-YajC peut aider ce processus. La protéine SPase clive ensuite le peptide signal permettant de relâcher la protéine mature dans le périplasme. La force proto-motrice peut aider dans les deux modes de translocation. (C) Enfin, les protéines solubles libérées dans le périplasme peuvent (I) interagir avec les chaperonnes (Skp et SurA dans cet exemple) afin d'être prises en charge par le système BAM pour être insérées dans la membrane externe, (II) se replier spontanément ou avec l'aide d'autres chaperonnes (Dsb, PpiA qui catalysent des réactions enzymatiques facilitant le repliement de la protéine), (III) être prises en charge par un autre système spécialisé (comme les systèmes de sécrétion, qui ne sont pas représentés ici), ou (IV) être ancrées dans la membrane externe (côté périplasme) via de petits lipides.

Afin d'amener la protéine vers la membrane et le système de translocation, plusieurs acteurs peuvent entrer en jeu (voir Figure 2, page 9). En majorité, les protéines membranaires sont prises en charge pendant leur synthèse par le ribosome, on parle d'une prise en charge co-traductionnelle ; au contraire, la plupart des protéines solubles qui seront sécrétées seront prises en charge après leur synthèse, de façon post-traductionnelle. La protéine SRP ("*ribonucleoprotein signal recognition particle*") est l'un de ces acteurs (Luirink & Sinning, 2004). D'autres protéines appelées chaperonnes peuvent participer pour guider les protéines qui seront excrétées, favoriser leur solubilité (protéines transmembranaires) et réguler le transport de ces dernières.

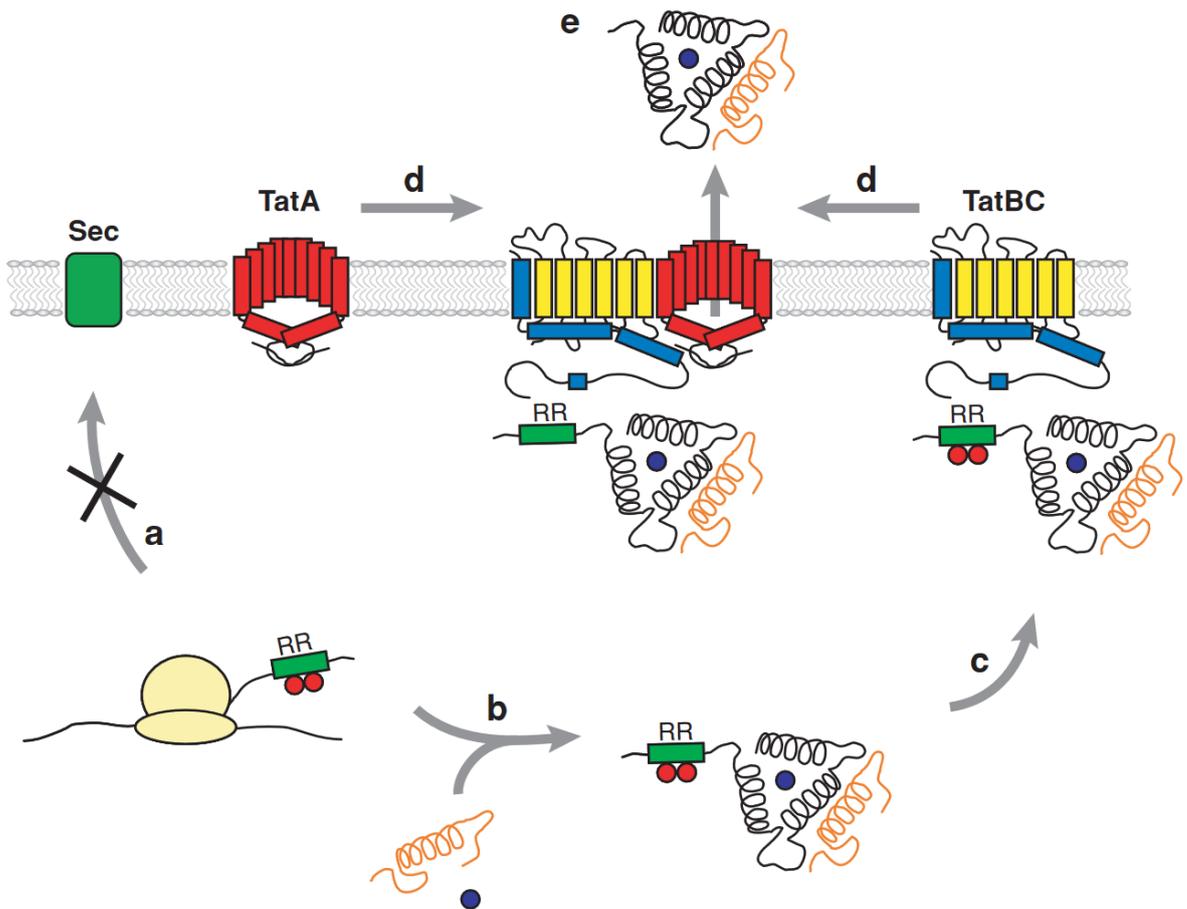
II.1.2 La voie Tat

A l'inverse de la voie Sec, la voie Tat se distingue par le transport de protéines repliées, et plus spécifiquement au travers de la membrane cytoplasmique (C. Robinson & Bolhuis, 2004). Ce système peut transporter des grandes protéines ou complexes sans dépenser la moindre molécule d'ATP : il repose sur la force protomotrice (Cline *et al.*, 1992), cette même force qui permet d'actionner l'ATP synthase (voir Figure 3, page 11). Bien que cette voie soit utilisée par les bactéries, les archées et dans les chloroplastes des plantes, le mécanisme global est encore mal compris à cause de l'absence de données structurales à haute résolution sur le système et ses composants. En revanche, les événements qui précèdent le transport ont été décrits récemment (Frain *et al.*, 2019). Les protéines qui jouent un rôle important dans ce système sont TatA, TatB et TatC chez les bactéries à Gram négatif, tandis que chez les bactéries à Gram positif, TatA et TatB sont regroupées en une seule protéine multidomaine (Frain *et al.*, 2019 ; Pop *et al.*, 2002 ; Sargent *et al.*, 1999).

Chez les bactéries, les substrats de ce système varient beaucoup. En effet, entre 20 et 30 substrats ont été identifiés chez les bactéries entériques comme *E. coli* (Berks *et al.*, 2000). En revanche, chez les bactéries à Gram positif, seulement quelques substrats ont été prédits (Dilks *et al.*, 2003 ; Sibbald *et al.*, 2006 ; Tjalsma *et al.*, 2004). Les raisons précises qui déterminent l'utilisation de ce système restent encore mal comprises, mais plusieurs pistes ont été découvertes et documentées. La première est que la formation d'un complexe protéine/cofacteur grâce à une enzyme est plus simple dans le cytoplasme, plutôt que dans le périplasme entre autres parce que le transport de tous les partenaires vers le périplasme coûte beaucoup d'énergie (Berks, 1996). En effet, il est bien plus compliqué d'employer plusieurs mécanismes de transport à la fois pour la protéine, pour le

cofacteur et pour l'enzyme vers le périplasma afin de réaliser la réaction. Deuxièmement, le contenu en ion métallique dans le cytoplasme est bien plus strict et permet d'éviter une certaine compétition entre les ions pour les sites actifs des protéines (Monteferrante *et al.*, 2012; Tottey *et al.*, 2008). Enfin, certains complexes s'assemblent de façon optimale dans le cytoplasme. Il est donc logique de favoriser leur assemblage avant leur transport.

FIGURE 3 – La voie de sécrétion Tat



tiré de Lee *et al.*, 2006

Représentation schématique du système de sécrétion Tat. (A) Après avoir été synthétisée au sein du ribosome (en beige), la protéine peut éviter les autres modes de sécrétion grâce à son peptide signal et à son repliement. (B) Après repliement, les co-facteurs et autres sous-unités protéiques s'ajoutent à la protéine (C) avant que l'ensemble soit dirigé vers le récepteur TatBC (en bleu et jaune). (D) La force proto-motrice permet la formation d'un complexe de translocation actif TatABC qui permet ensuite au substrat de traverser le canal formé majoritairement par TatA (en rouge). (E) Après que le peptide signal a été clivé, la protéine mature est relâchée dans le périplasma.

Tout comme la voie Sec, la translocation par la voie Tat nécessite la présence d'un peptide signal en N-terminal de la protéine à transporter (Rodrigue *et al.*, 1999). Il semblerait que ce peptide ne

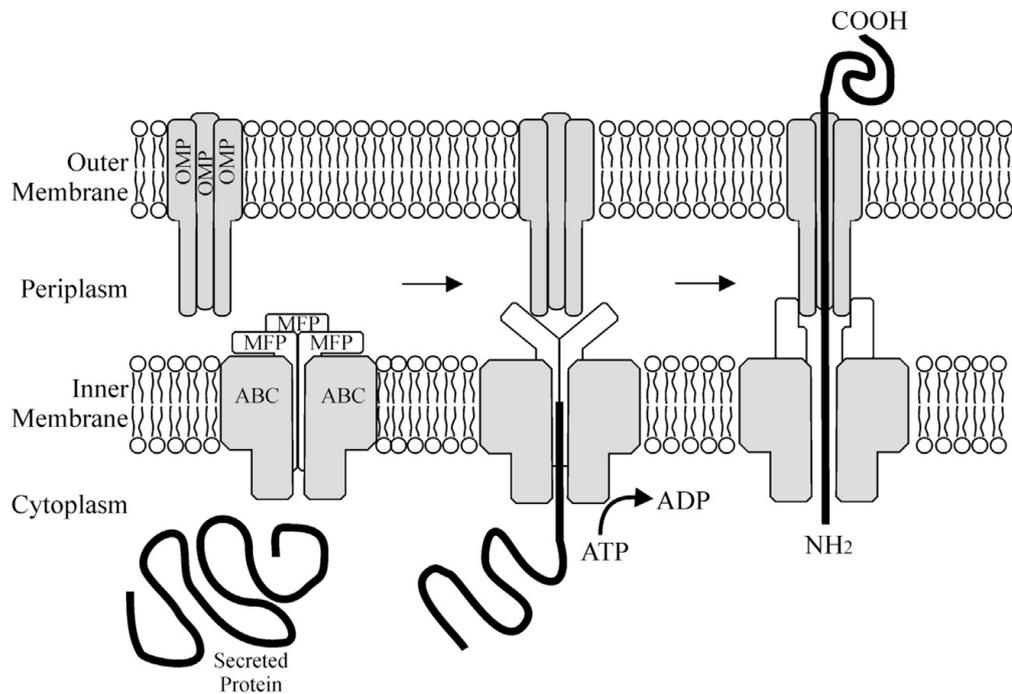
soit pas obligatoire pour les autres partenaires du complexe. Bien que possédant une structure et composition assez similaires au peptide signal Sec, le peptide signal Tat présente un motif particulier contenant deux arginines, d'où le nom de système de translocation à arginines jumelles. La délétion de l'une de ces arginines diminue le taux de transport chez les bactéries, et la double délétion réduit ce taux à zéro (Buchanan *et al.*, 2001). Chez les plantes, les deux arginines sont essentielles (Chaddock *et al.*, 1995).

II.2 Système de sécrétion de type I

Le système de sécrétion de type I (T1SS) est une machinerie assez simple retrouvée chez les bactéries à Gram négatif (Wells & Henderson, 2013). Le plus connu des T1SS chez *E. coli* permet de transporter l'hémolysine, une toxine formant des pores dans des cellules entériques que l'on trouve dans la souche entérohémorragique (ECEH) et qui a par exemple causé une épidémie en Allemagne en 2011 (Frank *et al.*, 2011) avec plus de 3300 malades, dont une centaine de cas graves et plus de trente décès.

Le T1SS permet à la bactérie de sécréter des protéines non repliées directement depuis le cytoplasme vers le milieu extérieur, en agissant comme deux plateformes indépendantes sur les deux membranes (voir Figure 4, page 13). Une fois que le substrat a été reconnu, grâce à une séquence située en C-terminal dans ce cas, la partie inférieure (IMC, "*Inner membrane complex*", complexe de la membrane interne) située dans la membrane interne va interagir avec la partie supérieure (OMC, "*Outer membrane complex*", complexe de la membrane externe) de la membrane externe. Cet appariement permet la formation d'un canal mature pour transporter la protéine du cytoplasme vers le milieu extracellulaire. Les substrats ont des tailles très variables, allant de 80 à plus de 8000 résidus et des fonctions également diverses (Deleplaire, 2004), comme des protéases, des lipases, des toxines, etc... Comme précisé un peu plus haut, c'est majoritairement par l'extrémité C-terminale que sont reconnus les substrats du système, toutefois il existe certaines exceptions pour lesquelles la voie Sec peut intervenir (Ghigo & Wandersman, 1994 ; Koronakis *et al.*, 1989 ; Sheikh *et al.*, 2002).

FIGURE 4 – Le système de sécrétion de type I



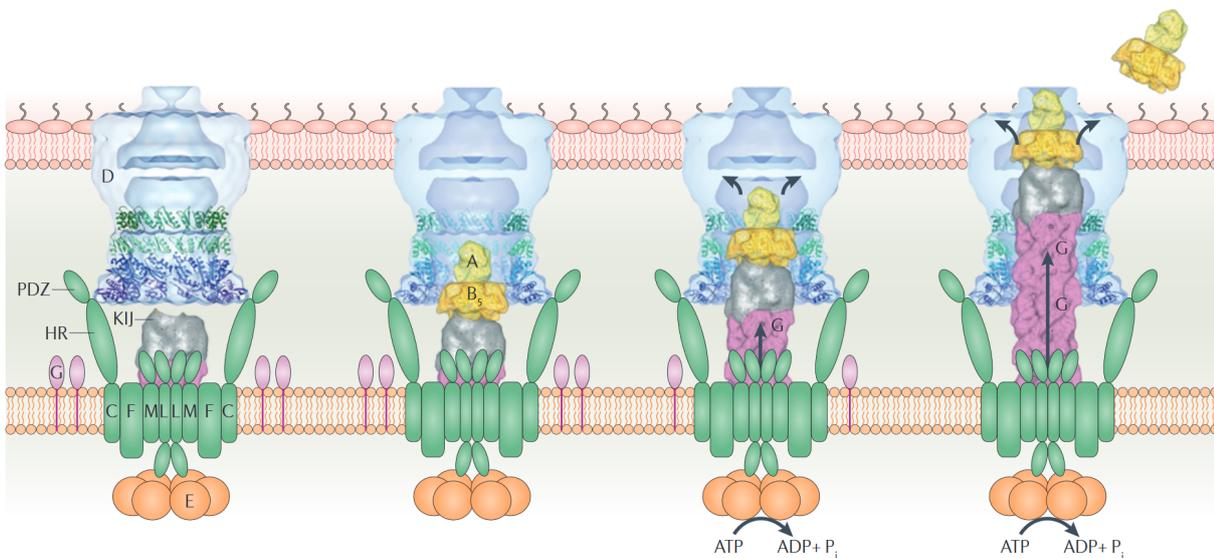
tiré de Wells et Henderson, 2013

Représentation schématique du système de sécrétion de type I. Après reconnaissance d'un peptide signal en C-terminal du substrat (la protéine sécrétée), le complexe de la membrane interne composé de l'ATPase ABC et de protéines MFP interagit avec le trimère OMP situé sur la membrane externe. S'ensuit la formation d'un canal scellé qui traverse les deux membranes permettant au substrat d'être sécrété en une seule étape sans passer par le périplasm.

II.3 Système de sécrétion de type II

Découvert il y a une vingtaine d'années, le système de sécrétion de type II est reconnu comme facteur important pour la virulence de bactéries à Gram négatif. Il favorise l'acquisition de nutriments, la création de biofilms, l'invasion et la colonisation par ces organismes. Les bactéries du groupe "ESKAPE" (De Oliveira *et al.*, 2020), connues pour leur dangerosité dans les milieux hospitaliers, utilisent le T2SS pour sécréter certaines protéines favorisant l'infection d'un patient. Les protéines sécrétées par ce système ont toutefois des fonctions diverses et variées, comme des protéases, lipases, etc. Dans certaines espèces de bactéries, le T2SS peut n'avoir qu'un seul substrat, tandis que chez d'autres espèces il y a en a plusieurs (Cianciotto, 2005). Petit point supplémentaire sur ce système, il serait également le précurseur des systèmes de pili de type IV (T4P) (Melville & Craig, 2013).

FIGURE 5 – Le système de sécrétion de type II



tiré de Korotkov *et al.*, 2012

Représentation schématique du fonctionnement hypothétique d'un système de sécrétion de type II. Les protéines Gsp ("General secretory pathway", voie de sécrétion générale) sont symbolisées ici par leur préfixe en lettres capitales. L'ATPase GspE est en orange, les protéines du complexe de la membrane interne sont en vert, l'extrémité du pseudopilus est grise, le pseudopilus est en rose et la sécrétine située sur la membrane externe est en bleu, cyan et pourpre. De gauche à droite est représenté le mécanisme possible de sécrétion. Tout à gauche se trouve l'état fonctionnel, en attente du substrat. Sur la seconde représentation, le substrat à sécréter (ici en jaune, la toxine cholérique) se place sur l'extrémité du pseudopilus. Le troisième panneau représente la croissance du pseudopilus par assemblage des sous-unités via la base, l'ouverture de la sécrétine jusqu'au quatrième et dernier état où la protéine est libérée à l'extérieur de la cellule.

Le T2SS est un complexe qui permet de transporter des protéines repliées depuis le périplasma des bactéries à Gram négatif jusqu'à l'extérieur de la cellule (voir Figure 5, page 14). Il faut donc que les protéines soient d'abord transportées vers le périplasma par la voie Sec ou Tat (Korotkov *et al.*, 2012; Voulhoux *et al.*, 2001), et qu'elles présentent en N-terminal une séquence peptide spécifique à l'une de ces deux voies (Korotkov *et al.*, 2012). De plus, le T2SS transportant les protéines repliées, il faut donc impérativement que cette étape de repliement se passe dans le périplasma pour les protéines utilisant la voie Sec.

La composition des T2SS est bien plus élaborée que celle du T1SS. C'est un complexe d'environ 15 protéines différentes dont les copies définissent sommairement 4 parties distinctes. Ces dernières s'étendent sur les deux membranes et sont respectivement nommées complexe de la membrane interne (IMC) et complexe de la membrane externe (OMC) (Korotkov *et al.*, 2011). Au sein du complexe de la membrane externe réside une protéine multimérique appelée la sécrétine qui possède notamment une extrémité N-terminale suffisamment longue pour entrer en contact avec le complexe de la membrane interne qui s'étend dans le périplasma (Korotkov *et al.*, 2011). La plateforme du complexe de la membrane interne joue un rôle essentiel car elle interagit avec toutes les autres parties du T2SS (Korotkov *et al.*, 2012). L'ATPase, qui permet d'alimenter l'ensemble du système en énergie, se trouve dans le cytoplasme. A l'inverse, le pseudopilus, qui se trouve au niveau du complexe de la membrane externe, agirait comme un piston : le contact entre le substrat et les protéines de la plateforme (sécrétine et pseudopilus) activerait l'ATPase, permettant l'élongation du pseudopilus jusqu'à l'extérieur de la cellule (Hobbs & Mattick, 1993; Korotkov *et al.*, 2012; Michel-Souzy *et al.*, 2018).

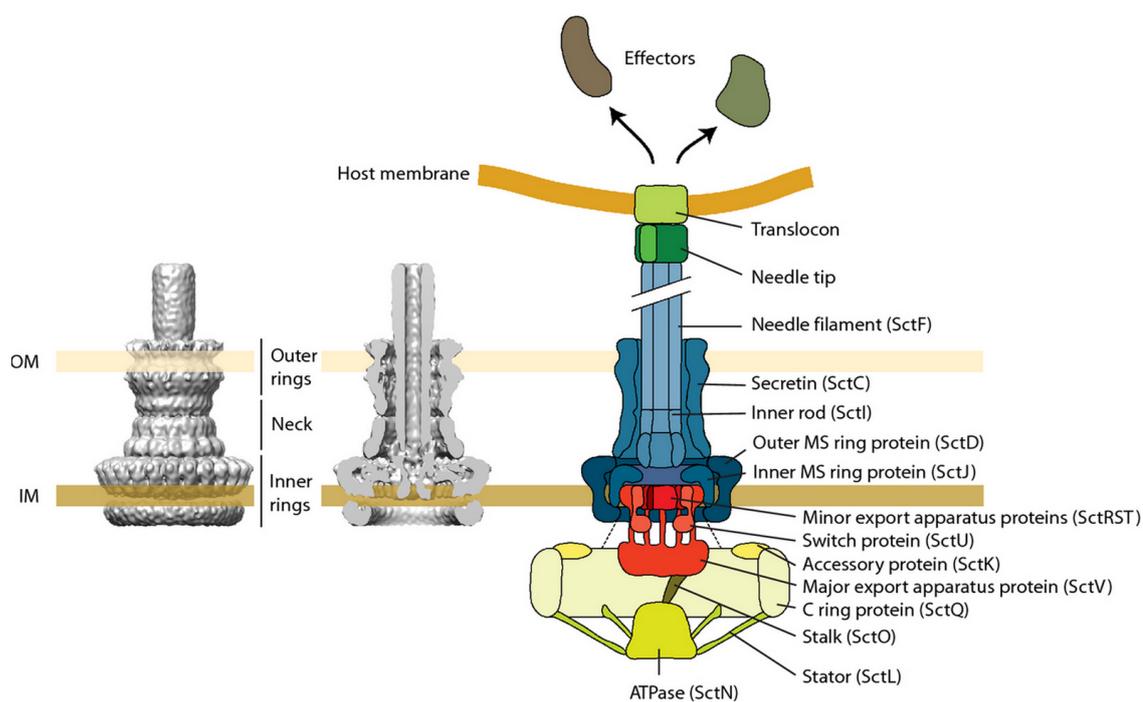
II.4 Système de sécrétion de type III

Retrouvés chez les bactéries à Gram négatif, les T3SS font partie des systèmes les plus complexes que l'on connaisse aujourd'hui chez ces organismes, avec une seule à une centaine de copies pour chacune des 20 à 30 protéines que l'on peut retrouver au sein de ces systèmes, selon les espèces (Büttner, 2012; Cornelis, 2006). Le coeur du complexe, aussi appelé "injectisome", est capable de transporter ses substrats protéiques probablement repliés directement depuis le cytoplasme vers le milieu extérieur et est conservé entre les différentes espèces bactériennes (Cornelis, 2006; Galán & Wolf-Watz, 2006). La plupart des T3SS sont impliqués dans la virulence entre la bactérie hôte et

une cellule eucaryote. C'est le premier des systèmes présentés dans cette introduction à pouvoir faire traverser 3 membranes à son substrat protéique, que l'on nomme effecteur protéique (spécifique au T3SS). La plupart des bactéries pathogènes utilisant ce système perdent partiellement ou totalement leur virulence lorsque l'on inactive leur T3SS (Hueck, 1998). Ces effecteurs protéiques portent une séquence spécifique en N-terminal nécessaire pour être transportés par le T3SS. Bien que certains d'entre eux n'en aient pas besoin, la plupart des substrats protéiques sont accompagnés par des chaperonnes jusqu'à la base du T3SS. Tout comme le T2SS, le T3SS possède également une base commune avec certains flagelles (Van Gijsegem *et al.*, 1995). En effet, les deux systèmes emploient un long filament pour remplir leur fonction. L'un utilise une sorte de seringue pour injecter un effecteur protéique dans une autre cellule, et l'autre une hélice qui, lorsqu'une rotation est appliquée à sa base, permet à la cellule de se mouvoir dans son environnement.

Les T3SS peuvent être décomposés en 3 parties (Green et Mecsas, 2016), le corps du complexe ancré dans la bactérie, "l'aiguille" qui permet à un effecteur d'être transporté, et le translocon qui crée un pore dans la cellule ciblée (voir Figure 6, page 17). La base du complexe comprend l'ensemble de la machinerie depuis la membrane interne côté cytoplasme jusqu'à la membrane externe, et forme une gigantesque structure à base d'anneaux de protéines empilés, contenant 1 à une centaine de copies de 15 protéines différentes (Burkinshaw & Strynadka, 2014; Kubori *et al.*, 1998). Cette base contient, entre autres, une sorte de plateforme d'où émerge un filament cylindrique qui s'étend depuis une sécrétine située au niveau de la membrane vers le milieu extra-cellulaire. Ce filament serait creux et assez large pour pouvoir transporter directement les effecteurs protéiques comme le contenu d'une seringue au travers d'une aiguille (Deane *et al.*, 2006; Demers *et al.*, 2014; Dohlich *et al.*, 2014; Radics *et al.*, 2014). Au bout du filament se trouve la machinerie permettant de reconnaître la cellule cible ainsi que le translocon (Holmström *et al.*, 2001; Picking *et al.*, 2005). Une fois que les conditions sont réunies, ces deux éléments vont créer un pore dans la membrane de la cellule cible afin de permettre la translocation des effecteurs protéiques directement dans le cytoplasme.

FIGURE 6 – Le système de sécrétion de type III



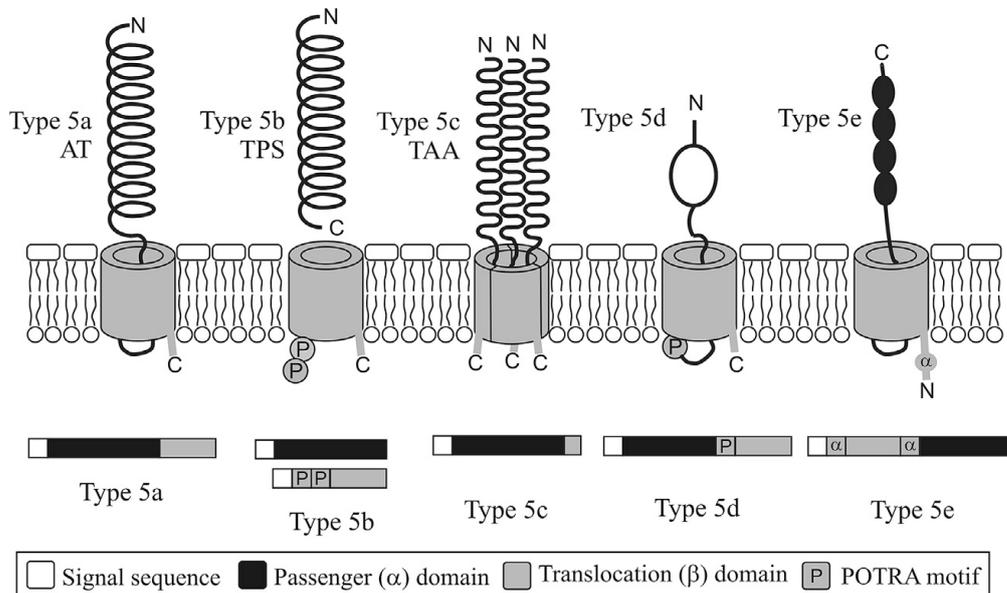
tiré de Diepold et Wagner, 2014

Représentation schématique du système de sécrétion de type III. Les panneaux de gauche et du milieu sont des visualisations 3D de données de cryomicroscopie électronique (pleine et en coupe). Le panneau de droite est un schéma qui indique l'emplacement de la plupart des éléments importants. En jaune sont représentés les composants cytoplasmiques, en rouge l'appareil d'export, en bleu la base et les composants de l'aiguille, puis l'extrémité de l'aiguille et le translocon en vert.

II.5 Système de sécrétion de type V

Les T5SS sont les systèmes de sécrétion les plus simples et probablement les plus fascinants. Parmi leurs différentes classes, on trouve par exemple les protéines "*Autotransporters*", ou AT, qui représentent la classe 5a des T5SS. Ces protéines multi-domaines possèdent leur propre système de transport. En effet, sur les 3 domaines principaux, le premier est une séquence signal N-terminale qui permet à la protéine d'être transportée vers le périplasma par la voie Sec. Ensuite, le dernier domaine s'assemble en un grand tonneau β constitué de 12 brins β et s'intègre dans la membrane externe (Henderson *et al.*, 1998 ; Henderson & Nataro, 2001). Enfin, le domaine de la protéine fonctionnelle, situé au milieu des deux domaines de transport, traverse la membrane externe où la protéine peut soit rester ancrée, soit être clivée (Henderson *et al.*, 2004). D'autres classes existent comme 5b, rassemblant les "*Two-partners secretion*" ou TPS dont le mécanisme ressemble à celui des AT, mais qui sont composées de deux protéines au lieu d'une seule (voir Figure 7, page 18).

FIGURE 7 – Le système de sécrétion de type V



tiré de Wells et Henderson, 2013

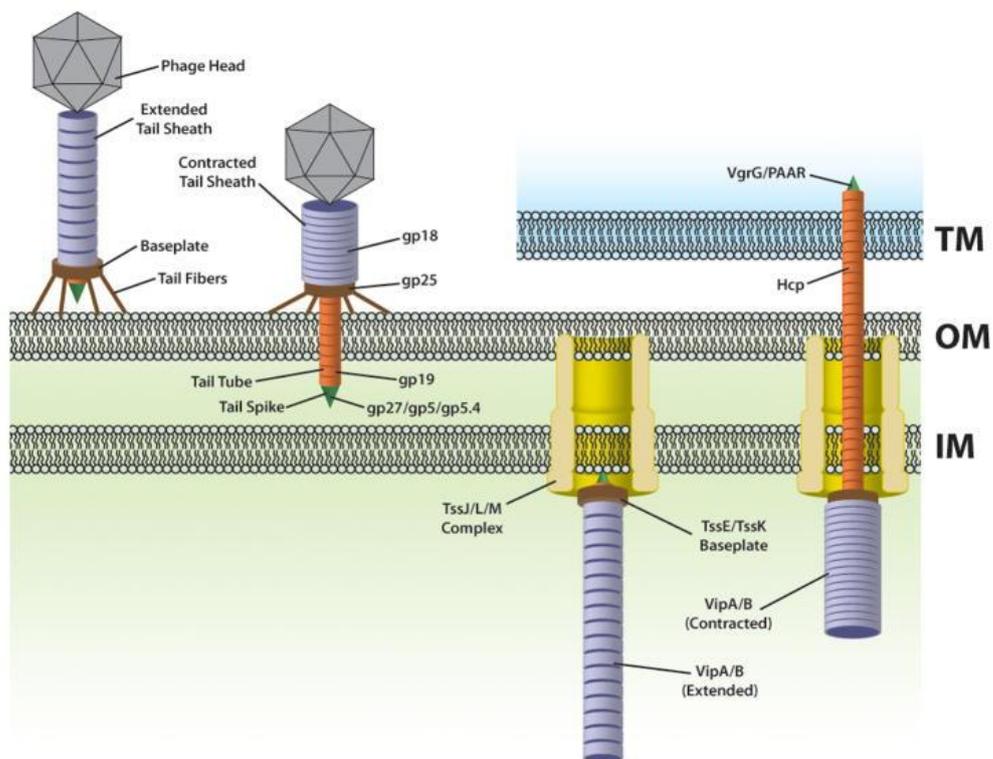
Représentation schématique du système de sécrétion de type V, divisé en 5 classes différentes : les autotransporteurs classiques ou AT (5a), la sécrétion double-partenaire ou TPS (5b), l'adhésine trimérique autotransportrice TAA (5c) ainsi que les types 5d et 5e. Toutes ces protéines ont un peptide signal situé en N-terminal, possèdent leur propre canal de translocation sous la forme d'un tonneau β et un domaine "passager" qui sera sécrété à la surface. Cependant, ils diffèrent en nombre de gènes, en organisation et par la structure du canal.

II.6 Système de sécrétion de type VI

Le T6SS fait partie des machineries moléculaires qui nous font dire que l'Homme n'a fait que réinventer ce que la nature utilise depuis des centaines de millions d'années. Tout comme l'ATP synthase agit comme un barrage hydroélectrique en utilisant la force proto-motrice pour actionner sa turbine afin de créer de l'énergie sous forme d'ATP, le T6SS est une sorte d'arbalète, propulsant une grande pique depuis l'intérieur de la cellule vers une cellule cible dans le but d'y injecter des toxines. Le T6SS a été découvert assez récemment par rapport aux autres systèmes de sécrétion, en 2006 chez le pathogène *Vibro cholerae* (Pukatzki *et al.*, 2006), et depuis a été détecté chez de nombreux autres pathogènes humains (Dudley *et al.*, 2006 ; Folkesson *et al.*, 2002 ; Lertpiriyapong *et al.*, 2012 ; Mougous *et al.*, 2006). Toutefois, le T6SS n'est pas réservé à l'agression de cellules eucaryotes, il joue un rôle tout aussi important contre les bactéries, y compris lors des co-infections entre pathogènes chez l'humain. Les T6SS reprennent en partie le fonctionnement des bactériophages en arborant une sorte de tube rigide et un filament flexible qui l'entoure qui s'appelle également le "fourreau" (Leiman & Shneider, 2012 ; Pukatzki *et al.*, 2007) (voir Figure 8, page 20). Ce dernier est capable de se contracter lors d'un contact avec une cellule "proie". La contraction du filament permet au tube central d'être propulsé, au bout de ce dernier se trouve la protéine qui facilite la perforation de la membrane, et permet ensuite à l'effecteur d'entrer dans la cellule (voir Figure 9, page 21). Ces systèmes sont capables de transporter de larges protéines repliées hydrophiles (Vettiger *et al.*, 2017).

Les T6SS connus aujourd'hui présentent au moins 13 protéines conservées, avec un maximum d'environ 21 protéines. Par exemple, la protéine du tube rigide aussi appelée Hcp est connue pour former des hexamères qui peuvent s'empiler (Ho *et al.*, 2014). Le "fourreau" contractile qui entoure le tube est composé de la protéine VipA/B qui forme également des hexamères empilés, avec une ouverture bien plus large, compatible avec le tube (Wang *et al.*, 2017).

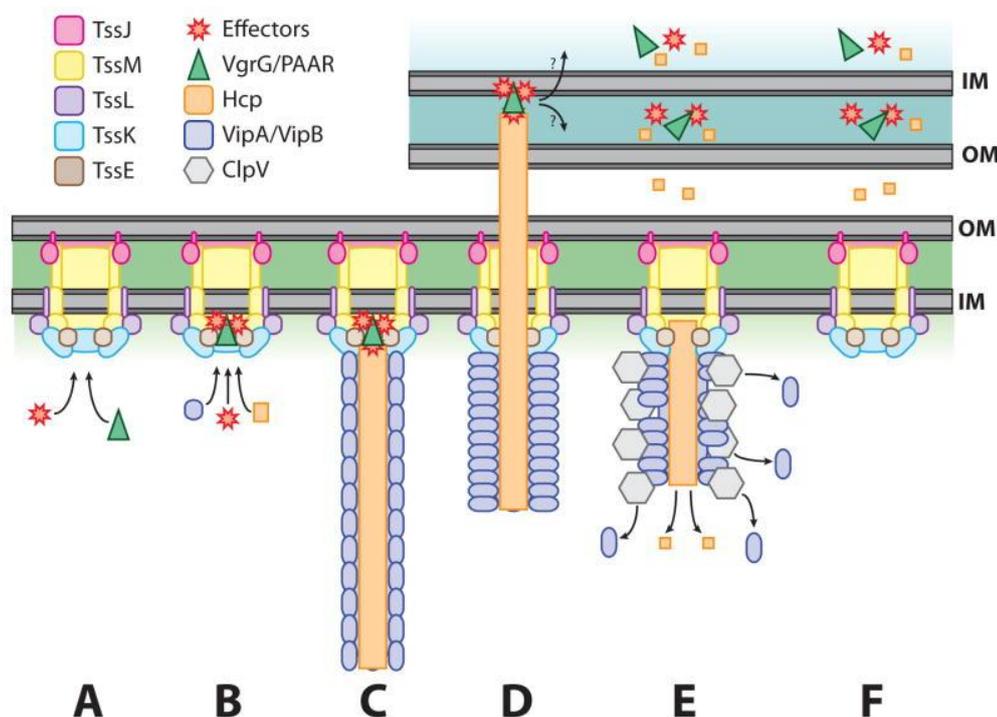
FIGURE 8 – Similarité de fonctionnement des bactériophages T4 et des T6SS



tiré de Ho *et al.*, 2014

Représentation schématique de bactériophages T4 et du système de sécrétion de type VI. Le processus d'injection possède des points communs (représentés par le même couleur) pour ces deux systèmes. La contraction du "fourreau" (gp18 ou VipA/B) permet au tube central (gp19 ou Hcp) d'être propulsé au travers de la membrane afin de délivrer le substrat, que ce soit de l'ADN pour le bactériophage ou bien un effecteur protéique dans le cas du T6SS.

FIGURE 9 – Le système de sécrétion de type VI



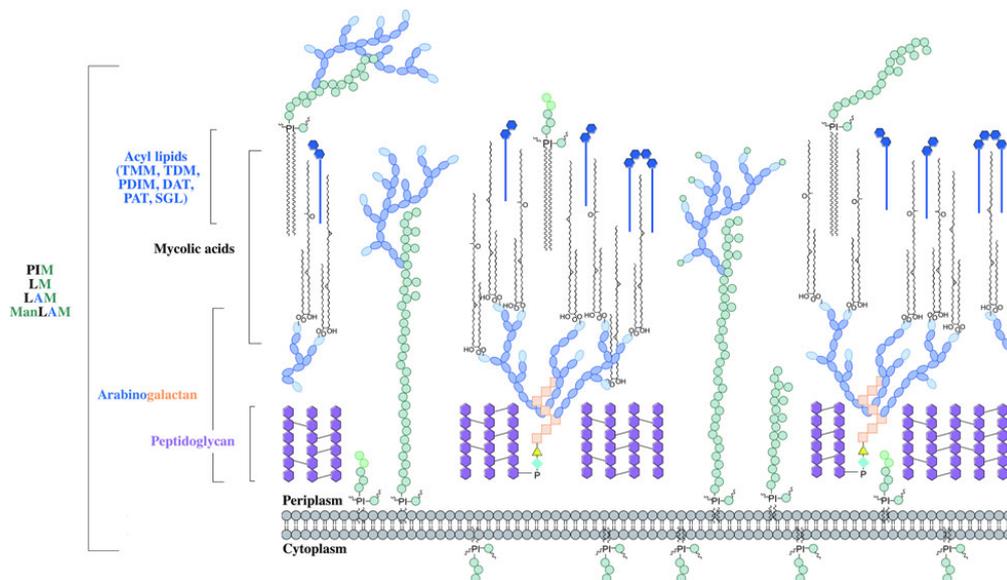
tiré de Ho *et al.*, 2014

(A) La plateforme centrale est formée par les protéines TssE, TssJ, TssK, TssL and TssM. Les autres composants qui ne sont pas indiqués sont TssA, TssF et TssG. Dans certains T6SS, Fha est un composant essentiel de ce complexe. (B) VgrG, PAAR et les effecteurs protéiques sont recrutés vers ce complexe et s'assemblent dans le complexe. L'interaction de VgrG avec PAAR et/ou l'effecteur protéique participe probablement à la stabilité de la structure globale. (C) Le tube de Hcp est polymérisé depuis le complexe VgrG/PAAR qui est lié à l'effecteur protéique, et le fourreau VipA/VipB se construit autour de Hcp. (D) Tout comme le fonctionnement chez le phage, un changement conformationnel du fourreau VipA/VipB engendre une contraction qui propulse le tube Hcp en dehors de l'hôte jusque dans la membrane de la cellule cible. Cette contraction permet de délivrer l'effecteur protéique dans la cellule cible. La protéine ClpV utilise de l'énergie sous forme d'ATP pour remodeler le fourreau contracté, permettant ensuite au tube Hcp d'être désassemblé à son tour. (F) La plateforme centrale est ensuite prête à recommencer le processus ou à être désassemblée.

II.7 Système de sécrétion de type VII

Particulièrement étudié chez le bacille de Koch également connu sous le nom de *Mycobacterium tuberculosis*, le système de sécrétion de type VII possède un rôle direct dans le développement et la virulence de la maladie (Bunduc, Bitter *et al.*, 2020). Il a également été étudié chez certains *Streptomyces* (Roman *et al.*, 2010), ainsi que les Firmicutes comme *Staphylococcus aureus* (Bottai *et al.*, 2017). Pour comprendre pourquoi ce système est à la fois unique et important pour les mycobactéries, il faut s'intéresser à leur physiologie. En effet, les mycobactéries sont des bactéries filamenteuses appartenant au groupe des bactéries à Gram positif mais possèdent une seconde bicouche lipidique, de la même manière que les bactéries à Gram négatif (voir Figure 10, page 22).

FIGURE 10 – La paroi des mycobactéries



tiré de Abrahams et Besra, 2018

Représentation schématique de la paroi des mycobactéries. Celle-ci est composée de glycolipides (PIM, ou phosphatidyl-myo-inositol mannosides ; LM, lipomannane ; LAM, lipoarabinomannane ; ManLAM, lipoarabinomannane mannosylé), de peptidoglycane, d'arabinogalactane et d'acides mycoliques. Les acyl-lipides sont intercalés dans la couche mycolique (avec TMM, trehalose monomycolate ; TDM, trehalosedimycolate ; DAT, diacyltrehalose ; PAT, polyacyltrehalose ; PDIM, phthiocerol dimycocerosate et enfin SGL, sulfoglycolipide). Les éléments de la capsule ne sont pas représentés ici.

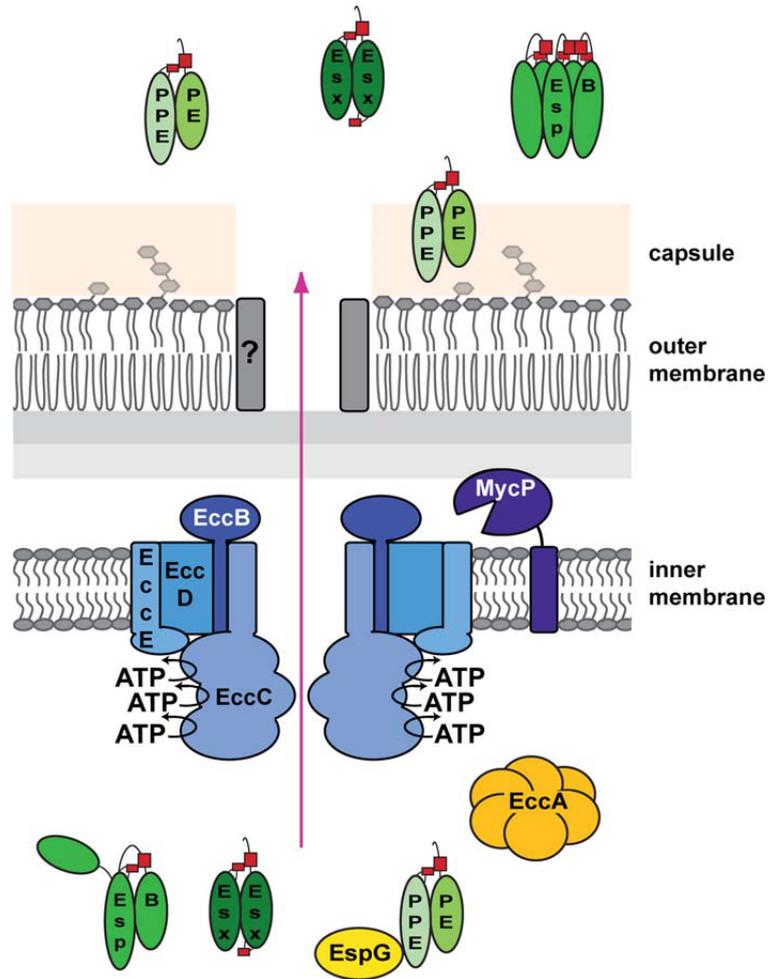
En plus de la couche de peptidoglycane épaisse caractéristique des bactéries à Gram positif, les mycobactéries présentent également une couche d'arabinogalactane liée covalamment au peptidoglycane, la partie "arabinose" étant orientée vers le milieu extérieur. Les acides mycoliques dont

la chaîne carbonée est particulièrement longue et ramifiée se lie aux résidus arabinoses. Afin de compléter la bicouche lipidique, d'autres lipides libres se lient aux acides mycoliques et forment la "mycomembrane" (Dulberger *et al.*, 2020 ; Hett & Rubin, 2008 ; Kieser & Rubin, 2014). Ces derniers sont particulièrement intéressants car nombre d'entre eux sont considérés comme des facteurs de virulence, par exemple le lipoarabinomannane (LAM) (Karakousis *et al.*, 2004). Enfin, les mycobactéries sécrètent des protéines, des lipides et des glucides qui forment une capsule, ajoutant encore à l'imperméabilité de l'ensemble de la paroi cellulaire (Daffé & Etienne, 1999). En tout, près de 40% de la masse cellulaire sèche des mycobactéries est constituée de lipides, ce qui représente presque le double par rapport aux bactéries à double membrane plus conventionnelles que nous manipulons en laboratoire (Brennan & Goren, 1979).

Cette imperméabilité permet à la cellule de se protéger efficacement de son environnement, notamment des agents antibactériens ou des lysozymes utilisés par les phagocytes. Cela dit, elle fonctionne également dans le sens inverse. Chez les mycobactéries, on retrouve les systèmes Sec et Tat afin de transporter les protéines depuis le cytoplasme vers le périplasme, mais le système dédié pour transporter depuis le périplasme vers le milieu extérieur reste encore inconnu (van Winden *et al.*, 2019). Le système de sécrétion de type VII quant à lui devrait permettre de traverser l'ensemble de la paroi (voir Figure 11, page 24), car aucun des substrats du système n'a encore été détecté dans le périplasme (van Winden *et al.*, 2019). Les effecteurs sont repliés et sont souvent transportés sous forme de dimères (Houben *et al.*, 2014).

Dans le cas des T7SS chez les mycobactéries, on parle plutôt de systèmes ESX ou Ecc, les premiers étant nommés de ESX-1 à ESX-5. Pour donner un exemple, le vaccin BCG contient une souche de *Mycobacterium bovis*, et cette souche est atténuée grâce à la délétion de ESX-1 (Pym *et al.*, 2002). Pour une revue plus complète sur ces systèmes et leurs substrats, voir (Bunduc, Bitter *et al.*, 2020). Concernant la structure des ESX, l'une des structures les plus récentes est celle de l'ESX-5 par cryomicroscopie électronique (Bunduc, Fahrenkamp *et al.*, 2020).

FIGURE 11 – La paroi des mycobactéries



tiré de Ates *et al.*, 2016

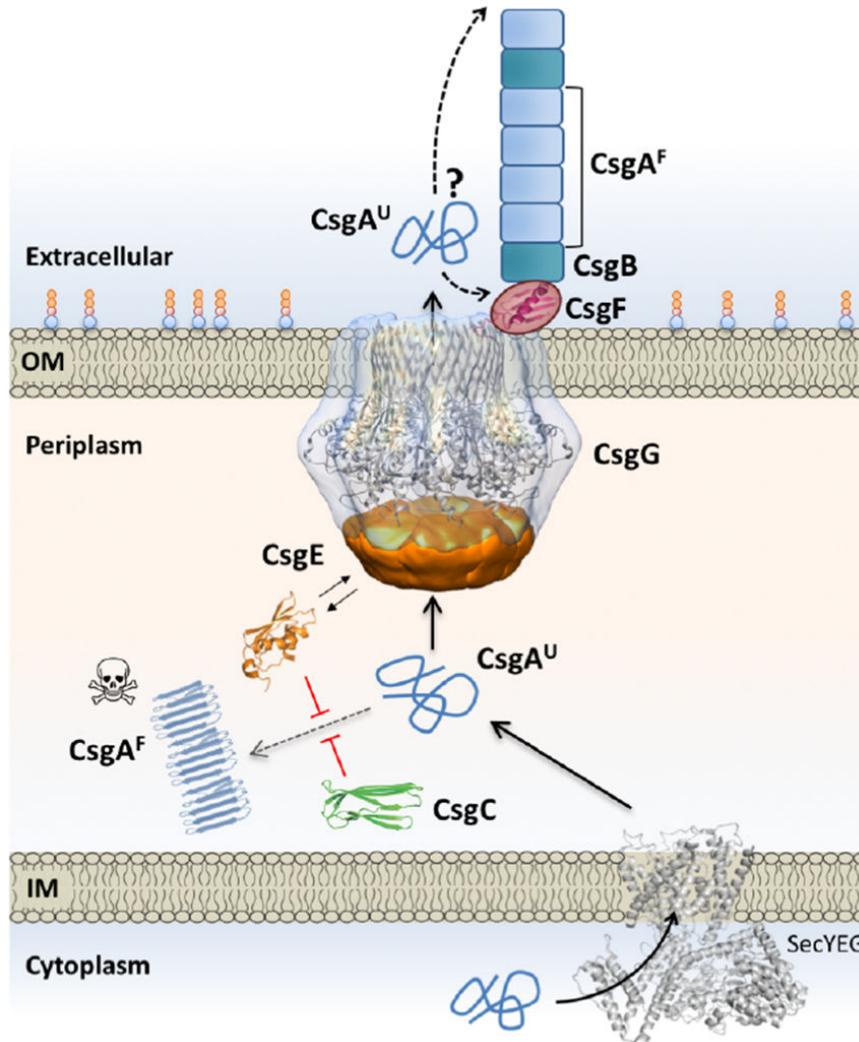
Représentation schématique d'un système de sécrétion de type VII chez une mycobactérie. Les éléments de la membrane en bleu forment un complexe dont l'homologue EccC est probablement l'ATPase qui fournit l'énergie pour la sécrétion. La mycosine (MycP) ne fait pas partie de ce complexe mais est essentielle pour la sécrétion. Les substrats du T7SS (en vert) sont reconnus grâce à leur signaux YxxxD/E et WxG (en rouge). La sécrétion de dimères PE-PPE dépend de chaperonnes cytosoliques EspG et EccA (en jaune et orange). Le processus intégral est encore inconnu, la partie de la membrane externe est symbolisée par un point d'interrogation (en gris).

II.8 Système de sécrétion de type VIII

Comme les autres systèmes, le T8SS permet de transporter une protéine vers le milieu extérieur. Jusque là, c'est assez classique. Ce qui le différencie, c'est la protéine qu'il va transporter. En effet, le T8SS est responsable de la sécrétion de la protéine "Curli". Un bien joli nom, tout à fait approprié puisque cette protéine va s'assembler sous forme de filaments qui vont boucler sur eux-mêmes. On retrouve exclusivement ces filaments chez les bactéries des groupes *Proteobacteria* et *Bacteroidetes* (Barnhart & Chapman, 2006). Ces fibres vont participer activement au biofilm bactérien, en s'enroulant autour des bactéries afin de favoriser leur adhérence sur une surface (Bhoite *et al.*, 2019; Kikuchi *et al.*, 2005). Chez certains pathogènes, ces filaments facilitent la colonisation de l'hôte (Gophna *et al.*, 2001; Van Gerven *et al.*, 2018).

Chez *E.coli*, le T8SS est composé de 7 protéines distinctes nommées CsgA à CsgG (Hammar *et al.*, 1995). CsgA et CsgB sont les composants principaux du filament, tandis que CsgE, CsgF et CsgG permettent l'assemblage de la machinerie de sécrétion. CsgC est une chaperonne et CsgD régule l'expression des 3 premières protéines. CsgA et CsgB sont tout d'abord transportées vers le périplasma par le système Sec et transportées vers le milieu extérieur par la machinerie de sécrétion CsgEFG (Nenninger *et al.*, 2011; Nenninger *et al.*, 2009; L. S. Robinson *et al.*, 2006). Les deux protéines s'assemblent ensuite pour former le filament. Le repliement de la protéine et la formation des filaments à l'intérieur de la cellule seraient catastrophiques pour la bactérie, et c'est le rôle de la chaperonne CsgC (et CsgE) d'éviter cet assemblage prématuré (Evans *et al.*, 2015) (voir Figure 12, page 26).

FIGURE 12 – Le système de sécrétion de type VIII



tiré de Bhoite *et al.*, 2019

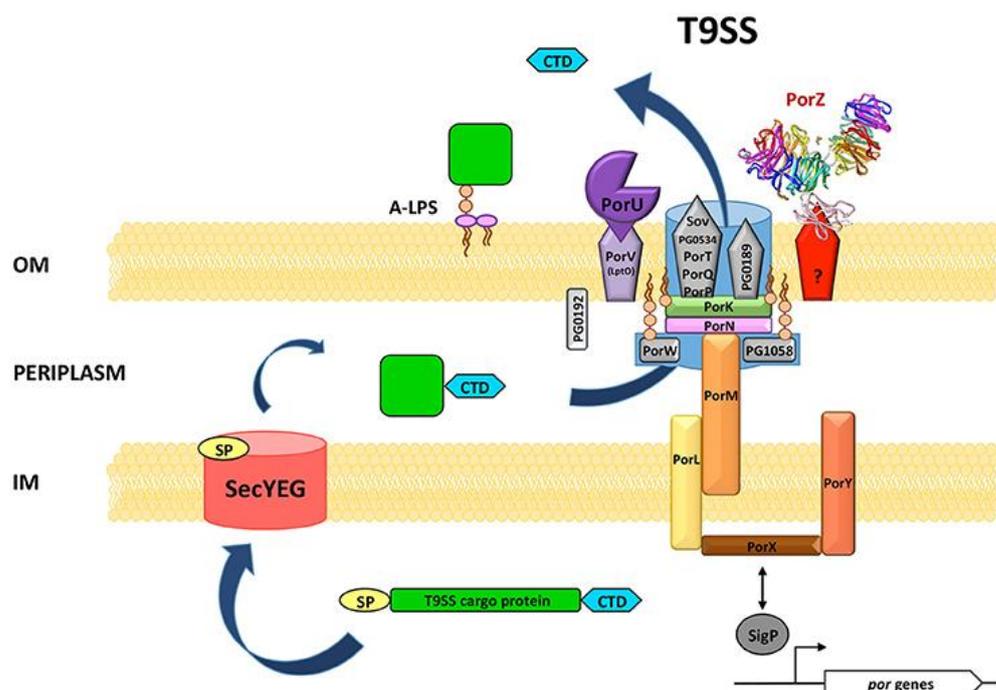
Les sous-unités de Curli CsgA entrent dans le périplasm via la voie générale de sécrétion Sec, d'où elles vont être dirigées vers l'extérieur de la cellule sous forme de protéines non repliées (CsgA-U) via le transporteur CsgG. Le repliement prématuré et la polymérisation de CsgA sont réprimés par CsgE et CsgC, sans quoi les filaments peuvent engendrer la mort cellulaire. CsgG forme un complexe nonamérique qui agit comme un canal de diffusion de peptides en coopérant avec CsgE qui recouvre la partie inférieure du canal en formant une sorte de bouchon. Le recrutement de CsgA au niveau du canal permet de créer un gradient qui favorise sa diffusion au travers du canal vers le milieu extérieur où son repliement et l'assemblage des filaments de Curli sont dirigés par CsgB et CsgF. CsgF est certainement positionnée proche du canal. Le rôle exact de CsgF ainsi que le lieu d'élongation des filaments sont encore inconnus (symbolisé par des flèches en pointillé). IM : membrane interne, OM = membrane externe.

II.9 Système de sécrétion de type IX

Le système de sécrétion de type IX est l'un des derniers systèmes de sécrétion à avoir été découverts et est pour l'instant uniquement décrit dans le groupe *Bacteroidetes* (Lasica *et al.*, 2017; McBride & Zhu, 2013b) (voir Figure 13, page 28). La première fonction étudiée du T9SS est la sécrétion des "gingipainés", des protéases qui dégradent les cytokines qui sont des molécules signal essentielles pour la défense et la réponse immunitaire chez un hôte (Guo *et al.*, 2010). En l'occurrence, ce facteur de virulence est sécrété ou recruté sur la membrane par le T9SS après avoir été transporté vers le périplasme par le système Sec/Tat, tout comme l'ensemble des substrats des T9SS. La deuxième fonction étudiée est tout autre, et on la retrouve chez les bactéries non pathogènes. Il s'agit de la motilité par glissement et l'assemblage du système a beaucoup en commun avec le T9SS cité plus haut (McBride & Nakane, 2015; McBride & Zhu, 2013a). Enfin, ces bactéries également retrouvées dans les sols sont capables de dégrader la cellulose et la chitine via la sécrétion d'enzymes, encore une fois transportées par un T9SS (Kharade & McBride, 2014; Zhu & McBride, 2014). L'ensemble de ces fonctions rendent le T9SS polyvalent pour ces organismes.

Assez curieusement, la partie membranaire interne des T9SS est constituée d'un moteur utilisant la force proto-motrice afin de faire tourner un stator incliné dont l'extrémité décrit des cercles d'environ 18 nm de diamètre. Cette rotation serait ensuite transmise via interaction avec l'assemblage situé sur la membrane qui s'étend dans le périplasme, fournissant l'énergie nécessaire pour la translocation de protéines ou bien la motilité par glissement (Gorasia, Veith *et al.*, 2020; Hennell-James *et al.*, 2020; Leone *et al.*, 2018). Le complexe de la membrane externe, appelé translocon ou SprA/Sov (respectivement retrouvé chez *F. johnsoniae* et *P. gingivalis*), est aussi très intéressant. L'ensemble forme un des plus grands tonneaux β jamais découverts avec 36 brins β au total et une ouverture interne mesurée d'environ 7 nm (Lauber *et al.*, 2018). Deux autres protéines nommées PorV et Plug ont un rôle important dans le mécanisme. L'ouverture du canal SprA/Sov se fait par interaction avec PorV, et la fermeture se fait par fixation de Plug au niveau de l'entrée du canal.

FIGURE 13 – Le système de sécrétion de type IX



tiré de Lasica *et al.*, 2017

Représentation schématique du système de sécrétion de type IX. La structure globale du translocon ainsi que les protéines qui forment le pore dans la membrane externe ont été résolues et présentées dans une autre publication (Gorasia, Chreifi *et al.*, 2020), en cours d'édition. Le substrat du T9SS (aussi appelé protéine cargo) doit posséder deux signaux : un signal peptide en N-terminal qui permet de diriger la protéine vers la voie Sec, et un domaine C-terminal (CTD) reconnu par le T9SS. Après translocation au travers de la membrane interne, la plupart des protéines obtiennent leur repliement final dans le périplasma. Ensuite, le CTD permet de diriger la protéine vers le T9SS pour être transférée vers le milieu extérieur. Enfin, le CTD est clivé de la protéine par la sortase PorU, et la protéine sécrétée est modifiée via l'addition d'un lipopolysaccharide anionique (A-LPS) qui fixe la protéine sur la surface de la membrane externe. Les deux composants PorX/PorY ainsi que le facteur sigma SigP ont un effet de régulation sur les gènes Por.

II.10 Les derniers systèmes de sécrétion

Assez récemment, les systèmes de sécrétion de type X et XI ont été décrits (TXSS, Palmer *et al.*, 2021) (TXISS, Grossman *et al.*, 2021). L'étude du TXSS a notamment permis de répondre à l'un des mystères autour de la toxine typhoïde qui avait besoin d'une hydrolase de peptidoglycane spécifique (muramidase) pour être transportée au travers des membranes chez les salmonelles (Gram négatif). Également, ce système a été décrit chez *Serratia marcescens* (Gram négatif) dans le cas de la sécrétion de la chitine qui a également besoin d'une hydrolase de peptidoglycane (endopeptidase). Dans le cas du TXISS, ce système de sécrétion permet de transporter de nombreuses molécules au travers de la membrane externe chez les bactéries à Gram négatif. Les acteurs principaux de ce système ont d'abord été découverts chez les bactéries pathogènes animales (comme *Pasteurella multocida*, le choléra des poules, Heungens *et al.*, 2002), il a ensuite été associé aux bactéries pathogènes humaines (Hooda *et al.*, 2017 ; Hooda *et al.*, 2016). Ce système permet notamment de sécréter des lipoprotéines et des protéines solubles qui semblent avoir un rôle dans le biofilm bactérien.

Nous pouvons nous attendre à de nombreux nouveaux systèmes de sécrétion dans le futur, notamment grâce à l'énorme diversité d'organismes et de substrats transportés au travers des membranes. Pour le moment, les classifications suivent l'ordre des découvertes, mais il n'est pas exclu qu'une nouvelle classification soit proposée au fur et à mesure des découvertes, permettant de regrouper les systèmes en fonction de la membrane traversée, du type de substrat et du groupe bactérien (Gram positif ou négatif).

III Le système de sécrétion de type IV

Après avoir fait un tour d'horizon des systèmes de sécrétion, on se rend bien compte de la diversité des systèmes employés par le vivant pour transporter les macromolécules biologiques vers le milieu extracellulaire. Pour leur part, les systèmes T4SS sont capables de transporter tout type de molécules et complexes depuis une cellule vers une autre, y compris de l'ADN dans le processus de conjugaison. Ces systèmes sont présents chez un grand nombre de procaryotes et peuvent posséder plusieurs fonctions de transport à la fois (Alvarez-Martinez & Christie, 2009). On classe les T4SS selon leur rôle en 3 types différents : effecteur de protéines, libération/absorption d'ADN, et conjugatif.

Le premier type, les T4SS effecteurs de macromolécules, ressemblent aux autres systèmes de sécrétion dans leur mode de fonctionnement. La principale différence est que les T4SS injectent directement leur substrat dans une cellule cible, qu'elle soit une bactérie ou une cellule eucaryote. Ce type de sécrétion s'associe à une forme de virulence, car les substrats peuvent être protéiques (toxines, etc.) ou ADN (plasmides, etc.). Dans le cas de toxines anti-bactériennes, elles favorisent la survie de l'espèce qui les sécrète en tuant d'autres micro-organismes et/ou ralentissant leur croissance (exemple : les bactéries du genre *Xanthomonas*). Il est également possible que ces effecteurs soient dirigés directement au sein d'une cellule eucaryote comme par exemple l'ADN de transfert (ADN-T) délivré par le T4SS d'*Agrobacterium tumefaciens* dans une cellule végétale, permettant de détourner la/les cellule(s) au profit de la population bactérienne environnante.

Le deuxième système permet à une cellule de transporter un ADN vers le milieu extérieur. Représenté par les organismes *Helicobacter pylori* et *Neisseria gonorrhoeae*, ce système est appelé libération/absorption d'ADN ("*DNA release/uptake*") car il associe à la sécrétion d'ADN la capacité d'en intégrer par un processus de transformation naturelle, en partie via le T4SS.

Quant au troisième et dernier type de T4SS, aussi appelé T4SS conjugatif, il permet le transfert d'un substrat ADN directement dans une autre cellule. A noter que l'ADN est dans la majorité des cas en complexe avec un partenaire protéique (une relaxase). Cette forme de transmission, appelée conjugaison, permet à une population bactérienne soumise à un stress environnemental de maximiser ses chances de survie en partageant une partie de son patrimoine génétique avec d'autres organismes procaryotes. Cela dit, pour que ces gènes soient effectivement transmis, il faut qu'ils soient présents au sein du vecteur qui sera véhiculé par le système en question. Non seulement

ces vecteurs possèdent ces gènes favorisant la survie et la virulence de ces organismes, mais ils contiennent souvent l'ensemble des gènes permettant de construire, de réguler et d'orchestrer le système de sécrétion (certains gènes mobiles ne codent toutefois qu'une partie du système et utilisent un système codé ailleurs dans le génome cellulaire). En ayant conservé la machinerie, la nouvelle bactérie hôte et ses descendants peuvent à leur tour transmettre leurs gènes si nécessaire via ce système. On appelle ce mécanisme le transfert horizontal de gènes, par comparaison au transfert vertical de gènes qui représente la transmission de gènes lors de la reproduction (asexuée, par division binaire dans le cas des bactéries). Le transfert horizontal de gènes est l'un des moyens les plus utilisés pour la transmission de gènes de résistance aux antibiotiques et de gènes de virulence.

Les vecteurs de ces systèmes de sécrétion se retrouvent généralement sous deux formes distinctes. La première est extra-chromosomique, c'est-à-dire qu'elle a un cycle de répllication différent de celui du chromosome bactérien, il s'agit d'un plasmide. La seconde forme est plus singulière et se situe au sein du chromosome bactérien, et on la nomme élément intégratif et conjugatif (ICE). Pour être conservé dans les futures générations bactériennes, un ICE profite de la répllication du chromosome. Par contre dans le cas d'un événement de conjugaison, l'ICE doit être extrait du chromosome bactérien afin de se recirculariser. Enfin, les deux types de vecteurs, les plasmides ou les ICE circulaires, sont répliqués et une copie va être transférée vers la cellule hôte tandis que l'autre va rester dans la cellule donneuse. Dans le cas des ICE, la copie non transférée est réintégrée dans le génome bactérien.

On différencie deux classes de T4SS chez les bactéries à Gram négatif, caractérisées par deux représentants types. Le premier, étudié chez l'organisme *Agrobacterium tumefaciens*, caractérise la classe IVA (T4ASS) et est aussi appelé le système VirB/D4. Le second système (T4BSS), aussi appelé DotI ou Icm, a d'abord été étudié chez *Legionella pneumophila*. Ces deux systèmes sont composés d'une douzaine à une vingtaine de protéines différentes qui peuvent être présentes en plusieurs copies au sein des systèmes assemblés. En comptant les bactéries à Gram positif, on augmente le nombre de classes de T4SS à 8 au total (Guglielmini *et al.*, 2014).

Bien qu'une distinction ait été faite entre les différentes fonctions des T4SS, il semble plutôt que chaque système soit capables de remplir plusieurs fonctions. Ainsi, un T4SS pourrait à la fois conjuguer mais aussi transférer des effecteurs protéiques à un autre moment (Christie, 2004 ; Nagai & Kubori, 2011). Cette plasticité rend les T4SS particulièrement intéressants. Malheureusement,

il en découle une nomenclature très touffue, sans standard établi, qui complique grandement la lecture de la bibliographie sur les T4SS. Par exemple, certaines protéines homologues ne possèdent pas un nom similaire alors que certaines protéines ayant des noms similaires ne sont pas du tout des homologues. Par chance, le système "Vir" d'*Agrobacterium tumefaciens* est bien étudié et possède une nomenclature claire. Lorsque ce sera possible, nous essayerons d'utiliser ce système Vir comme point de repère. Dans cette partie seront discutées les architectures respectives des deux systèmes de sécrétion de type IVA d'*A. tumefaciens* et IVB de *L. pneumophila*. Les mécanismes mis en place pour la conjugaison via les systèmes de sécrétion de type IV chez les bactéries à Gram négatif et positif seront ensuite présentés.

III.1 Architecture des systèmes de sécrétion de type IV

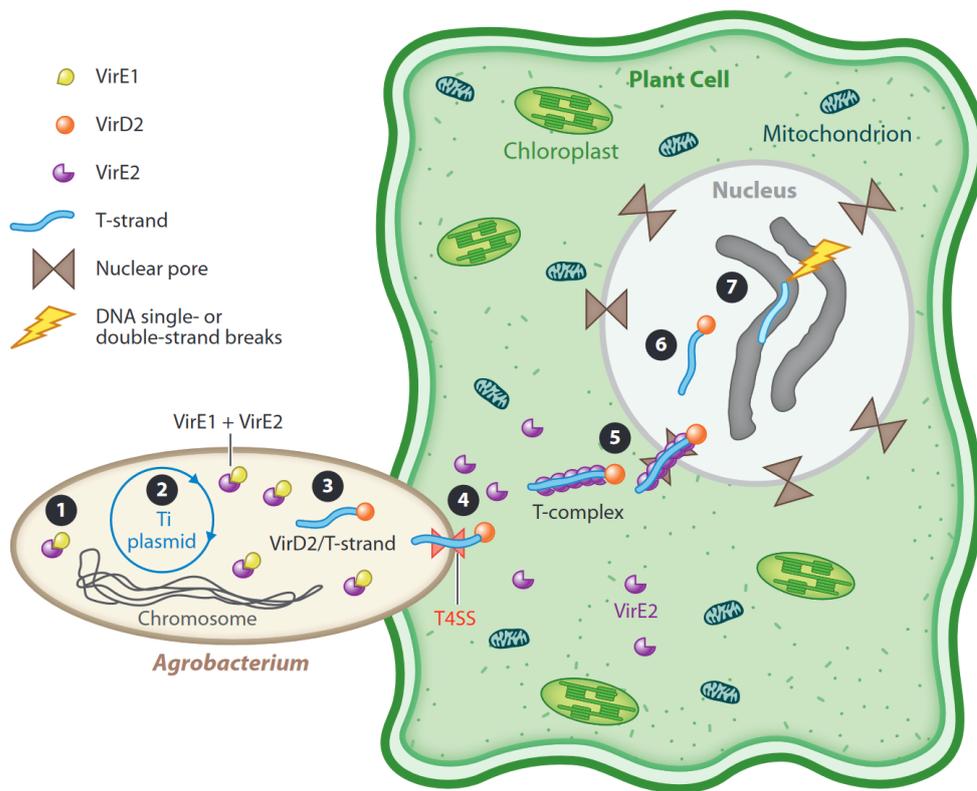
Dans cette première partie sur les systèmes de sécrétion de type IV, nous allons aborder les différentes approches employées par les bactéries afin d'infecter un hôte eucaryote ou de contrôler la population bactérienne environnante. Il ne sera donc pas encore fait mention de conjugaison bien que ces systèmes en soient capables, mais bien d'un transfert de macromolécules ayant un rôle dans la virulence de la bactérie.

III.1.1 Système de sécrétion de type IV de classe A

La bactérie à Gram négatif *Rhizobium radiobacter* plus connue sous le nom d'*Agrobacterium tumefaciens* possède un T4SS qui est l'un des plus connus et mieux étudiés de son genre (Aguilar *et al.*, 2010; Christie *et al.*, 2014; Li & Christie, 2018; Mary & Baron, 2018). Appelé système VirB/D4, ce système de sécrétion permet de délivrer un ADN simple brin oncogénique dans une cellule de plante afin qu'il soit intégré au sein du génome de cette dernière. Résultant de cette infection, la cellule eucaryote va se répliquer de manière non contrôlée et va être forcée à produire toute sorte de nutriments pour la bactérie qui infecte la plante. Sur le lieu de l'infection, se forme une sorte de boursouffure sphérique que l'on appelle une galle donnant le nom à la maladie : la galle du collet ("*Crown-gall disease*"). L'ADN transféré par ce T4SS possède le nom d'ADN-T ("*Transfer DNA*", *T-DNA*) et est conservé via la répllication du plasmide Ti ("*Tumor inducing*"). Cette méthode d'infection et d'intégration génomique a d'ailleurs été longuement étudiée et détournée pour créer

des organismes génétiquement modifiés. L'ADN-T possède plusieurs gènes qui seront exprimés une fois dans le chromosome de la plante. Entre autres, on y retrouve des gènes pour la synthèse des opines nutritives, qui sont des dérivés d'acides aminés que la bactérie est capable de métaboliser grâce à un gène spécifique situé sur le plasmide Ti. On y retrouve également des gènes oncogènes qui enlèvent les limitations du cycle cellulaire de l'hôte engendrant un cycle de division anormal et incontrôlé. Le transfert de cet ADN-T est expliqué plus en détail dans la figure 14, page 34.

FIGURE 14 – Infection par *Agrobacterium tumefaciens*



tiré de Gelvin, 2017

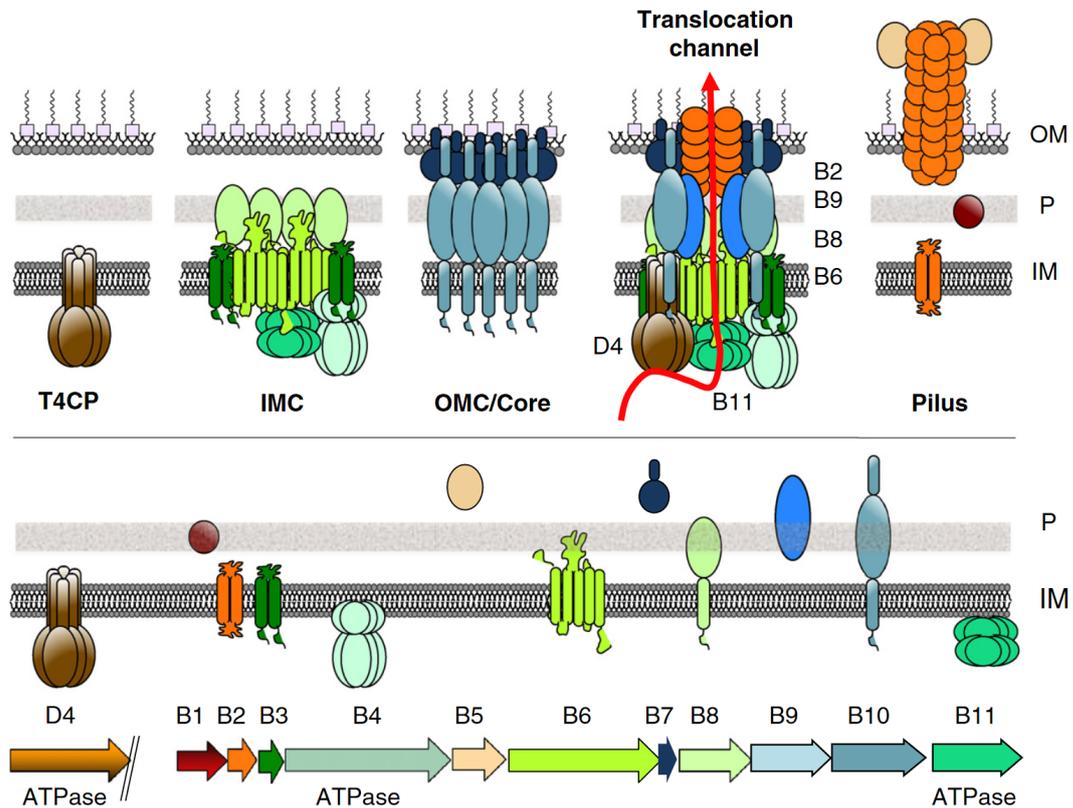
Représentation schématique du processus d'infection d'une cellule de plante par *Agrobacterium tumefaciens*. Lors d'une blessure chez la plante, des composés phénoliques sont libérés et permettent l'expression des gènes *vir*, impliqués dans la virulence. (1) Les phénols et les sucres de la plante sont captés par le duo VirA/VirG permettant l'induction des gènes *vir* et la production des autres protéines Vir. (2) VirD1 et VirD2 se lient aux répétitions de séquence sur le plasmide Ti bordant la région de l'ADN-T. (3) VirD2 s'attache covalamment sur l'extrémité 5' de l'un des brins-T afin de former un complexe nucléoprotéique VirD2/brin-T. (4) Le complexe VirD2/brin-T ainsi que des effecteurs protéiques Vir entrent dans la cellule de plante via un T4SS. Bien que VirE2 soit associé à des chaperonnes au sein de la bactérie, les chaperonnes ne sont pas transportées dans la cellule de plante. (5) Au sein de la plante, la protéine VirE2 s'associe au brin-T et le complexe entier (nommé "*T-complex*") voyage dans le cytoplasme et entre dans le noyau grâce aux pores nucléiques. Les autres protéines bactériennes et de plantes qui s'associent potentiellement avec le complexe ne sont pas montrées. (6) VirE2 se dissocie du brin-T. (7) Le brin-T est intégré dans le chromosome de la plante et les trans-gènes (à savoir les gènes bactériens) peuvent être exprimés.

Afin de mieux comprendre le fonctionnement et l'architecture globale des T4SS, il nous faut passer en revue les différentes protéines qui les composent. Le T4SS de classe IVA d'*A. tumefaciens* (système VirB/D4) est composé de douze protéines différentes numérotées de VirB1 à VirB11, et de VirD4. Ces protéines constituent l'ensemble du T4SS qui s'étend depuis le cytoplasme jusqu'à la membrane externe (voir Figure 15, page 36).

La protéine VirB1 est une hydrolase du peptidoglycane qui permet de remodeler le peptidoglycane pour que le T4SS puisse s'assembler et s'étendre sur les deux membranes. Elle est généralement essentielle à la sécrétion (Baron *et al.*, 1997; Llosa *et al.*, 2000; Zupan *et al.*, 2007). VirB2 et VirB5 sont respectivement des composants majeurs et mineurs du pilus, qui s'assemblent pour former une structure de 8 à 12 nm de diamètre et d'une longueur ne dépassant pas 1 μm (Bradley, 1980). VirB2 s'assemble de façon cyclique pour former le pilus (Eisenbrandt *et al.*, 1999), et VirB5 se trouve au bout du pilus et agirait de paire avec VirB2 pour stabiliser l'interaction à la membrane de la cellule cible (Aly & Baron, 2007; Backert *et al.*, 2008; H.-J. Yeo *et al.*, 2003).

Au sein du T4SS existent 3 ATPases différentes, VirB4, VirB11 et VirD4, qui permettent de fournir de l'énergie pour l'assemblage du système, notamment la formation du pilus, mais aussi pour la sécrétion. VirD4, aussi connue sous le nom de protéine de couplage, est l'une des premières protéines du T4SS à entrer en contact avec le substrat nucléo-protéique (Christie & Vogel, 2000; Christie *et al.*, 2014). Elle s'assemble généralement en hexamère (Gomis-Rüth *et al.*, 2001) et dynamise la première étape de la sécrétion par le T4SS grâce à l'utilisation d'ATP. VirB4 est l'une des protéines les plus conservées au sein des T4SS (Fullner *et al.*, 1994; Watarai *et al.*, 2002). Tout comme VirD4, cette protéine possède une activité ATPase lorsqu'elle est hexamérique (Arechaga *et al.*, 2008; Durand *et al.*, 2010) et est située au niveau de la face cytoplasmique de la membrane interne. Au sein de certains T4SS, il semblerait qu'elle interagisse avec VirD4 (Redzej *et al.*, 2017). VirB11 est la troisième et dernière ATPase retrouvée au sein des T4SS. Également hexamérique (Hare *et al.*, 2006; Savvides *et al.*, 2003; H. J. Yeo *et al.*, 2000), VirB11 est une ATPase dite de "trafic" et est associée à plusieurs autres systèmes de sécrétion chez les bactéries à Gram négatif, notamment les types II, III et VI (Fronzes, Christie *et al.*, 2009).

FIGURE 15 – Le système de sécrétion de type IV chez *A. tumefaciens*

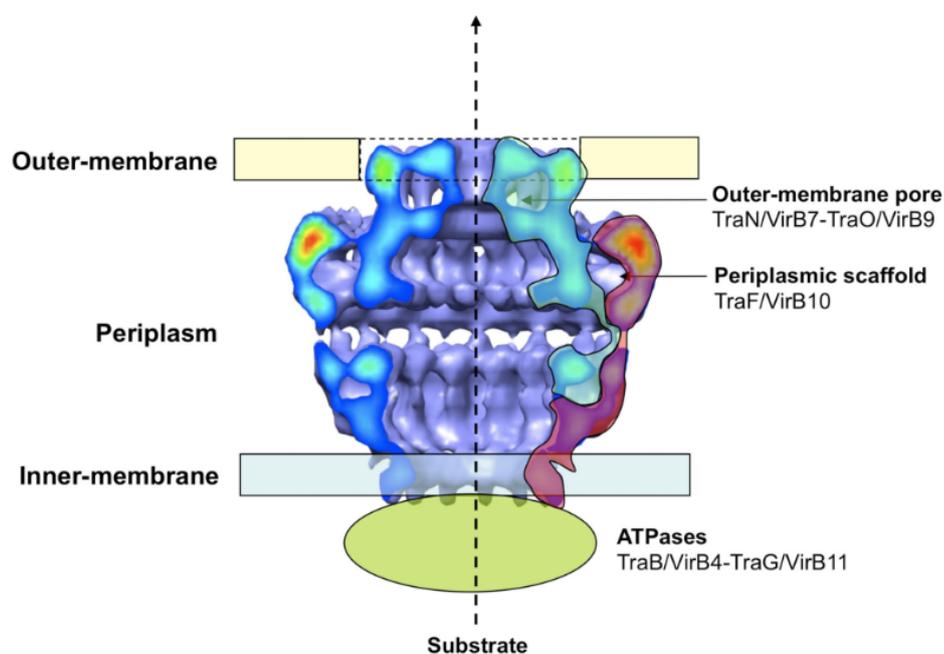


tiré de Christie *et al.*, 2014

Représentation schématique du système de sécrétion de type IV d'*A. tumefaciens* surnommé système VirB/D4. En bas de la figure nous retrouvons les 11 composants du T4SS qui sont synthétisés sous forme de deux opérons séparés avec leur localisation et leur organisation structurale connue dans le complexe global. En haut sont représentés de gauche à droite : l'hexamère de la protéine de couplage T4CP ancré dans la membrane interne avec les domaines fonctionnels situés au niveau de la face cytoplasmique; le complexe de la membrane interne composé des ATPases VirB4 et VirB11 et des protéines structurales VirB3, VirB6 et VirB8; le complexe de la membrane externe aussi appelé "coeur" composé des protéines VirB7, VirB9 et VirB10; le T4SS avec le chemin du substrat symbolisé par une flèche rouge ainsi que l'ensemble des constituants protéiques y compris le pilus qui est décrit dans le tout dernier panneau à droite composé des protéines de pili majoritaires VirB2 ainsi que de la protéine de pili minoritaire VirB5 située à l'extrémité du pilus, en plus de la peptidoglycane hydrolase VirB1 qui remodèle le peptidoglycane (P) située dans le périplasme.

Les protéines VirB6 à VirB10 que l'on peut qualifier de protéines structurales sont probablement les éléments qui permettent à l'ensemble du système de s'étendre sur les deux membranes. VirB7, VirB9 et VirB10 sont présentes en plusieurs copies chacune (14 dans le cas du système codé par le plasmide pKM101, voir Figure 16, page 37). L'ensemble forme une grande partie du complexe de la membrane externe que l'on appelle complexe de coeur ("*Core Complex*") et s'étend jusqu'au complexe de la membrane interne (Chandran *et al.*, 2009; Fronzes, Schäfer *et al.*, 2009). VirB6, VirB8 et (peut-être) VirB3 sont les protéines qui se positionnent pour former le complexe de la membrane interne (Christie *et al.*, 2014). VirB8 est intéressante car elle interagit avec de nombreux partenaires au sein du T4SS d'une part, et d'autre part elle est impliquée dans le transfert du substrat (Cascales & P. J., 2004; Christie *et al.*, 2014; Sivanesan & Baron, 2011; Yuan *et al.*, 2005).

FIGURE 16 – Structure du complexe VirB7/B9/B10 au sein du T4SS du plasmide pKM101



tiré de Fronzes, Schäfer *et al.*, 2009

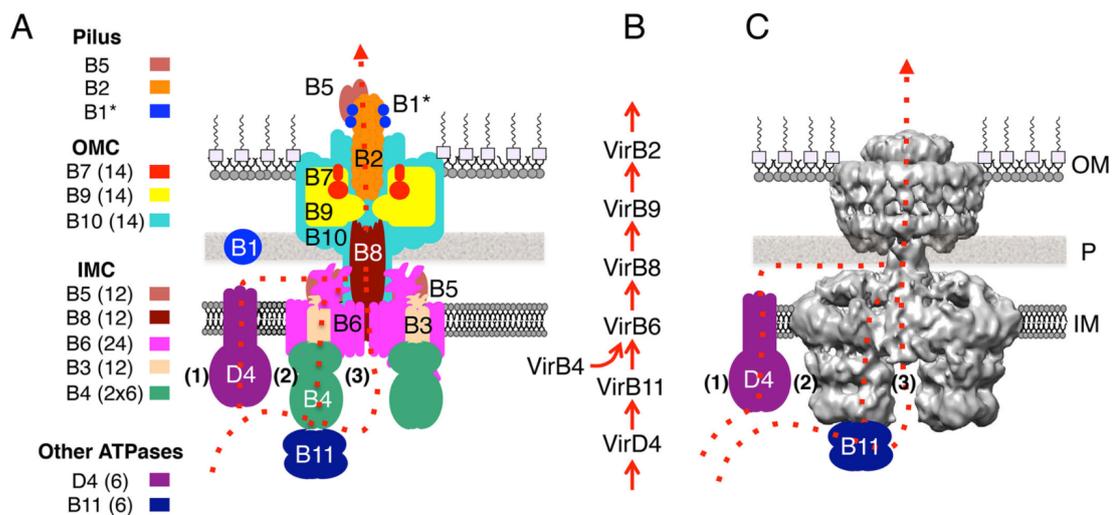
Localisation des protéines VirB7/TraN, VirB9/TraO et VirB10/TraF sur une représentation de la structure de cryo-microscopie électronique d'un assemblage provenant du plasmide pKM101 (Chandran *et al.*, 2009).

De ces différentes protéines nous pouvons dégager 4 groupes importants retrouvés chez les T4ASS de bactéries à Gram négatif (Christie, 2016; Darbari & Waksman, 2015) : 1) Les ATPases cytoplasmiques (VirB4, VirD4 et VirB11) qui permettent d'apporter l'énergie nécessaire au transfert du

substrat ; 2) Les composants de la membrane interne (VirB3, VirB6 et VirB8) qui forment la base et permettent l'ancrage du reste du système ; 3) Les composants du complexe de coeur de la membrane externe (VirB7, VirB9 et VirB10) qui s'étendent depuis la membrane externe jusqu'au complexe de la membrane interne ; 4) Les composants majeurs et mineurs du pilus (VirB2 et VirB5 respectivement) essentiels pour le transfert du substrat, ainsi que l'hydrolase de peptidoglycane (VirB1) qui permet de remodeler le peptidoglycane pour que l'ensemble du système puisse être assemblé.

L'une des structures les plus complètes de T4SS de classe A a été résolue en 2014 par l'équipe de G. Waksman (Harry H *et al.*, 2014) par cryomicroscopie électronique, chez le plasmide R388 d'*E. coli*. Reprise dans la revue de P. Christie en 2016 (Christie, 2016) et présentée dans la figure 17, page 39, elle permet de retrouver les 4 groupes importants relevés plus haut, ainsi que leur architecture globale et les chemins possibles des substrats.

FIGURE 17 – Structure du T4SS de classe A du plasmide R388 chez *Escherichia coli*



tiré de Christie, 2016

Représentations schématiques de la structure du T4SS du plasmide R388 d'*E. coli* en utilisant la nomenclature de référence d'*A. tumefaciens*. (A) Le T4SS est composé d'un pilus extracellulaire, d'un complexe de la membrane externe (OMC), d'un complexe de la membrane interne (IMC), de la protéine de couplage T4CP et d'une ATPase VirB11. Bien que la structure du pilus n'ait pas été résolue en même temps que la structure de microscopie électronique, il est proposé que le pilus fasse partie du système comme montré dans ce schéma. VirB1 est la peptidoglycane hydrolase. (B) Chemin du substrat obtenu grâce à des expériences de cross-linking (Cascales & P. J., 2004). (C) La structure du complexe VirB3-10 du plasmide R388 d'*E. coli* résolu par cryomicroscopie électronique à transmission (Harry H *et al.*, 2014). Le complexe de la membrane externe est supposé contenir le canal qui s'étend depuis le périplasm jusqu'au milieu extérieur. Les complexes de la membrane interne et externe sont reliés par une tige ("stalk"). Deux hexamères de l'ATPase VirB4 s'étendent dans le cytoplasme et sont supposés entrer en interaction avec les deux autres ATPases VirD4 et VirB11. (A et C) Pour les deux schémas, trois routes possibles pour le substrat sont représentées grâce à des lignes en pointillés rouges. Ces routes se rejoignent au niveau du complexe de coeur de la membrane externe où un seul chemin serait possible. OM = membrane externe; P = périplasm; IM = membrane interne.

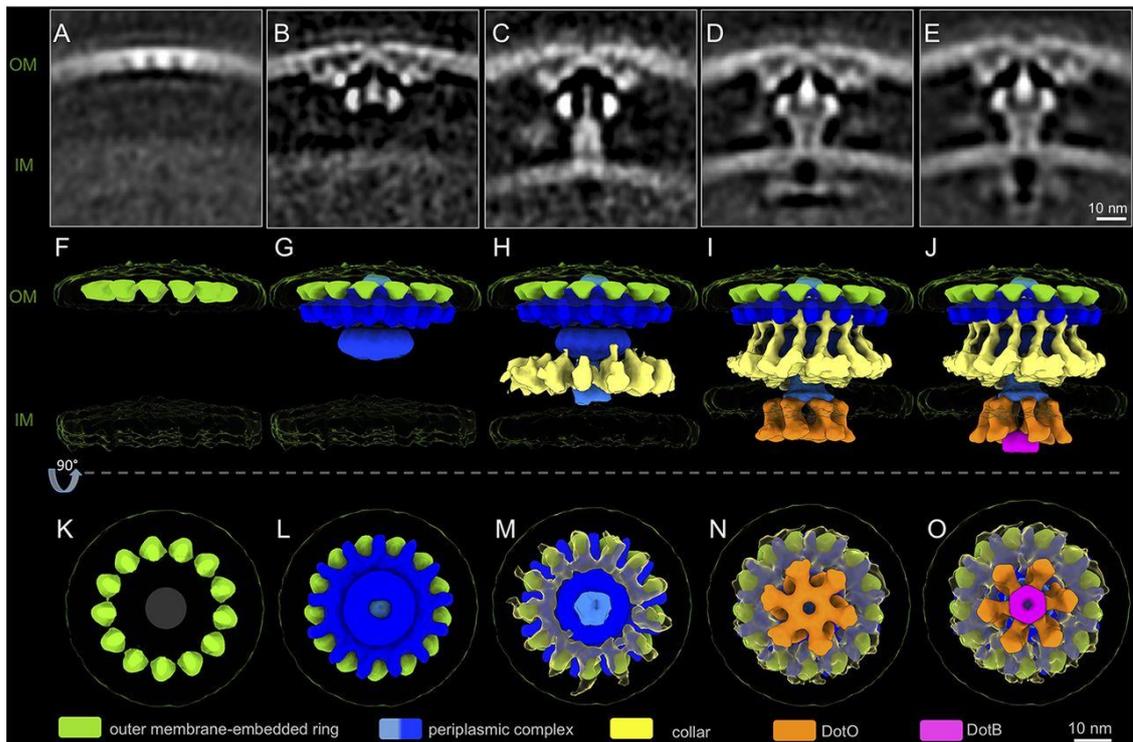
III.1.2 Système de sécrétion de type IV de classe B

La maladie du légionnaire (ou légionellose) est une forme de pneumonie grave causée par l'organisme *Legionella pneumophila*, un bacille à Gram négatif (Brenner *et al.*, 2020). La bactérie utilise un T4SS pour transférer de très nombreux effecteurs protéiques au sein d'une cellule eucaryote (Meir *et al.*, 2020), lui permettant d'en "prendre le contrôle" et d'y créer une organelle appelée LCV (pour "*Legionella-containing vacuole*") spécialement conçue pour lui permettre de s'y répliquer (Qiu & Luo, 2017). La bactérie a pour hôte naturel des protozoaires dans les eaux douces, mais peut être amenée à intégrer des macrophages pulmonaires humains après inspiration de micro-gouttelettes d'eau contaminée. La prolifération des bactéries dans le poumon et la destruction des macrophages engendrent une pneumonie sévère qui peut entraîner la mort dans 15 à 20% des cas.

Le T4SS de *L. pneumophila* appartient à la classe IVB. Nommé système Dot/Icm, il est composé d'un peu plus d'une vingtaine de protéines contre une douzaine pour la classe IVA (ex : *A. tumefaciens*). Deux groupes indépendants ont découvert ce système à peu près en même temps et l'ont nommé Dot ("*Defective in organelle trafficking*") et Icm ("*Intracellular multiplication*") ; c'est de là que vient le nom du système Dot/Icm (Segal *et al.*, 1998 ; Vogel *et al.*, 1998 ; Voth *et al.*, 2012). Les gènes qui le composent font donc partie des clusters *dot* ou *icm* (Kubori & Nagai, 2016). Cependant, pour des raisons de clarté, nous n'utiliserons que l'appellation Dot. Le système Dot ne partage qu'une petite quantité de gènes présentant une homologie de séquence significative avec ceux des T4ASS. Il en découle une architecture vraisemblablement différente de celle des T4ASS. C'est pour cela qu'aujourd'hui le système Dot/Icm fait partie de la classe IVB (T4BSS) (Christie & Vogel, 2000 ; Kubori & Nagai, 2016 ; Nagai & Kubori, 2011 ; Sexton & Vogel, 2002) des systèmes de sécrétion de type IV. Parmi les gènes qui sont partagés entre les deux systèmes, on retrouve les ATPases DotO, DotB et DotL qui sont respectivement des homologues distants de VirB4, VirB11 et VirD4. Enfin, DotG est une protéine homologue distante de VirB10 qui est l'une des protéines structurales du complexe de coeur de la membrane externe au sein des T4ASS. La liste complète des protéines du système Dot/Icm est disponible dans la publication de Segal *et al.*, 2005 mais également en annexe S1, page 205.

Deux publications datant de 2020 s'intéressent à la structure du système Dot/Icm. L'une présente des résultats de cryotomographie électronique (voir Figure 18, page 41) sur plusieurs cellules

FIGURE 18 – Intermédiaires structuraux du système de sécrétion de type IV Dot/Icm

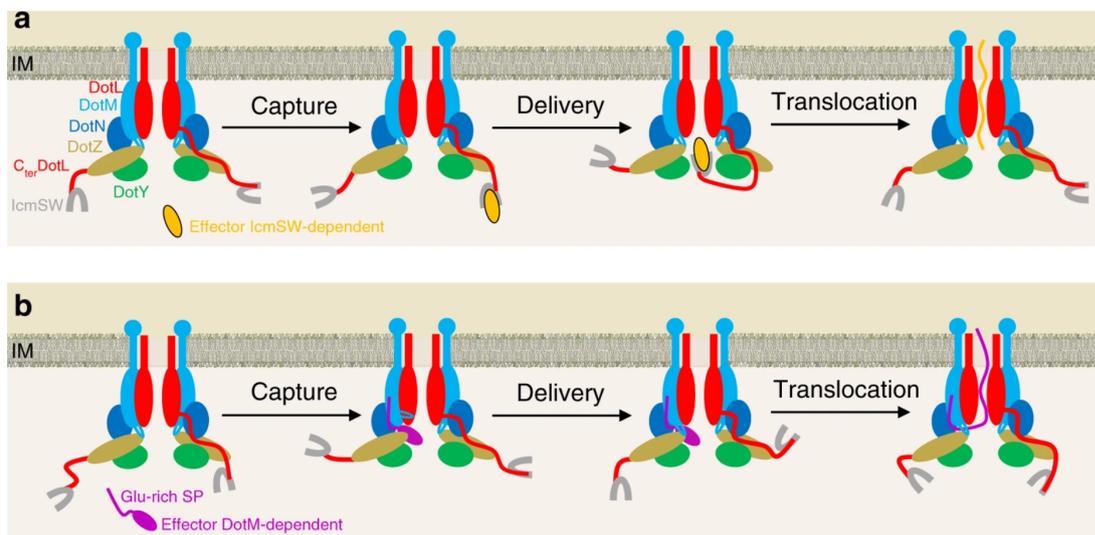


tiré de Park *et al.*, 2020a

(A à E) Représentation des moyennes de structures observées en cryotomographie électronique sur plusieurs intermédiaires d'assemblage du système Dot/Icm. De gauche à droite, les structures sont de plus en plus complexes. (F à O) Surfaces tridimensionnelles des modèles intermédiaires observés de A à E, avec F à J les vues latérales, et K à O les vues par le dessous. DotO est apparentée à VirB4 et DotB est apparentée à VirB11.

entières, montrant le système dans sa globalité, ainsi que certains intermédiaires (Park *et al.*, 2020b) d'assemblage du système. La seconde publication présente des résultats de radiocristallographie sur un complexe entier situé au niveau de la face cytoplasmique de la membrane interne appelé complexe de couplage de type IV (T4CC, "*Type IV Coupling Complex*"). Ce T4CC s'assemble en hexamères et permet de recruter et d'initier le transfert des substrats via le T4SS (Meir *et al.*, 2020). Les auteurs proposent un modèle fonctionnel de ce système T4CC (voir Figure 19, page 42). Il est important de noter qu'un autre modèle du T4CC avait été proposé précédemment par Kwak *et al.*, 2017, moins complet et composé de multiple structures tandis que le nouveau modèle intègre de nouveaux partenaires et présente la structure du complexe complètement assemblé.

FIGURE 19 – Modèle fonctionnel du complexe de couplage de type IV chez *L. pneumophila*



tiré de Meir *et al.*, 2020

La flexibilité du module IcmSW par rapport au reste de la structure permet de faciliter la capture des substrats situés dans l'environnement du complexe par le complexe de couplage de type IV. Une fois capturé, l'effeteur est dirigé vers le canal DotL. DotL est l'équivalent de l'ATPase VirD4. L'activité ATPase de DotL pourrait avoir un rôle double : déstructurer la protéine et fournir l'énergie nécessaire au transfert.

Le T4CC est situé sur la membrane interne, au niveau de la face cytoplasmique. Ce système n'est pas visible dans le modèle de cryomicroscopie tomographique. A la place, la structure globale met en évidence les protéines DotO/VirB4 et DotB/VirB11 qui forment un assemblage de symétrie 6, ressemblant aux structures que l'on connaît pour les systèmes de sécrétion de type IV de classe A. Le T4CC présente d'autres protéines dont DotL/VirD4. Bien que le T4CC soit absent du modèle

présenté par (Park *et al.*, 2020b), aucune contradiction apparente n'existe qui invaliderait l'un des deux modèles. D'autres études doivent être menées pour élucider le positionnement et la fonction de ces deux modèles.

Tout comme les T4SS de classe A, les systèmes de sécrétion de type IVB sont retrouvés chez d'autres organismes que *Legionella*. D'ailleurs, le système Dot/Icm se retrouve lui-même chez un autre genre de bactérie à Gram négatif, *Coxiella*. En l'occurrence, *Coxiella burnetii* utilise ce système dans un but similaire qui a été majestueusement résumé dans le titre de la publication de Newton et Roy, 2011 : "*The Coxiella burnetii Dot/Icm System Creates a Comfortable Home through Lysosomal Renovation*". Pour en savoir plus sur les T4BSS et leurs similarités avec les autres systèmes de sécrétion, voir la revue de Kubori et Nagai, 2016.

III.2 La conjugaison via un système de sécrétion de type IV

Maintenant que deux exemples d'architecture de système de sécrétion de type IV ont été décrits, nous allons exposer les mécanismes de la conjugaison utilisant ces systèmes. Nous allons d'abord parler des éléments mobiles qui sont transportés entre les cellules, leur structure ainsi que leur fonctionnement, et ce au sein des bactéries à Gram négatif et positif. Comme l'architecture des T4SS Gram positif n'est pas encore connue, nous détaillerons également les connaissances actuelles sur ces derniers, avant de développer les différents mécanismes de conjugaison.

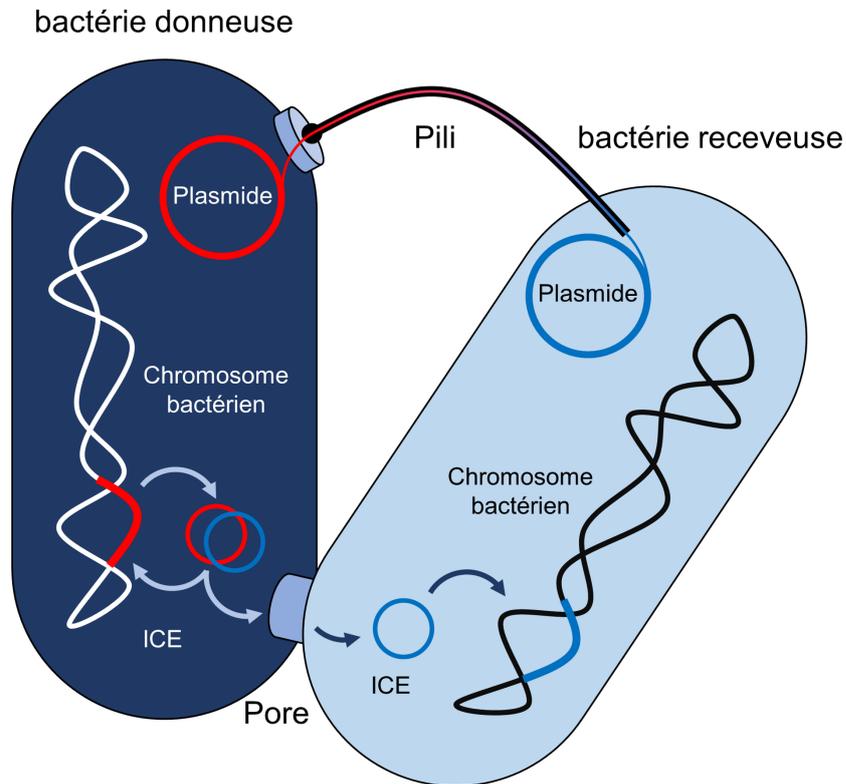
III.2.1 Les éléments mobiles conjuguatifs

On distingue deux types d'éléments mobiles principaux qui sont impliqués dans les mécanismes de conjugaison. Les deux sont des molécules d'ADN regroupant un ensemble de gènes et sites qui sont exprimés/utilisés par la cellule pour maintenir l'élément au cours des générations lors de la division cellulaire (transfert vertical de gènes, Lorenzo-Díaz *et al.*, 2017). Le premier type correspond aux plasmides, des molécules circulaires tout comme le chromosome bactérien, localisés dans le cytoplasme. Les plasmides possèdent une origine de réplication autonome de la réplication du chromosome bactérien et se retrouvent en plusieurs copies (entre une vingtaine et plusieurs centaines). Les copies sont équitablement réparties entre les deux nouvelles cellules lors de la division cellulaire. Le second type d'élément mobile correspond aux éléments intégratifs et conjuguatifs, ou

ICE ("*Integrative and Conjugative Element*"). Les ICE sont également des molécules d'ADN double brin, mais à l'inverse des plasmides, ils ne sont pas dissociables du chromosome bactérien car ils y sont directement intégrés. De ce fait, les ICE sont répliqués en même temps que le chromosome bactérien, et donc transmis aux générations suivantes. Pour éviter toute modification de gènes importants, les ICE ne sont intégrés qu'à des régions spécifiques du génome d'un organisme. Ceci est possible grâce à des sites spécifiques sur le chromosome bactérien qui seront reconnus et utilisés par les recombinases codées sur l'ICE, permettant son insertion et son excision du génome de l'hôte. En plus de leur architecture propre, les éléments mobiles impliqués dans la conjugaison possèdent des modules spécifiques : 1) une région variable codant pour différentes protéines qui apporteront potentiellement un avantage à la bactérie receveuse dans son environnement ; 2) une région conservée qui code pour la majorité des protéines permettant à l'élément mobile d'être transféré, à savoir un système de sécrétion de type IV (Voir Figure 20, page 45)

Les plasmides conjugatifs et les ICE sont retrouvés chez les organismes à Gram négatif et positif : *Bartonella henselae* (Gram négatif, plasmide R388, Fernández-González *et al.*, 2011) ; *Clostridium perfringens* (Gram positif, plasmide pCW3, Revitt-Mills *et al.*, 2021) ; *Acidithiobacillus ferrooxidans* (Gram négatif, ICEAfe1, Flores-Ríos *et al.*, 2019) ; *Bacillus subtilis* (Gram positif, ICEBs1, Auchtung *et al.*, 2016). Toutefois, les ICE ont pour l'instant été majoritairement retrouvés chez les Firmicutes.

FIGURE 20 – Mécanisme simplifié du transfert d'un élément mobile



Représentation schématique des mécanismes de transfert pour les plasmides et les ICE. Au cours de la vie d'une cellule, les plasmides possèdent une répllication autonome permettant à leur population de rester stable au cours des différentes générations. Pour un ICE, le fait qu'il soit intégré lui offre une certaine sécurité car il est répliqué en même temps que le génome de la bactérie. Lorsqu'un événement de conjugaison se prépare, les deux systèmes (plasmide en haut de la bactérie donneuse ; ICE en bas de la bactérie donneuse) expriment les protéines nécessaires à la répllication correcte de l'élément génétique, à l'évènement de conjugaison et à la formation du T4SS. Une fois la relaxase fixée sur le plasmide, la répllication débute et le brin d'ADN lié à la relaxase est transporté via le T4SS. Dans le cas de l'ICE, une étape d'excision depuis le génome est nécessaire, mais le processus reste similaire à celui du plasmide.

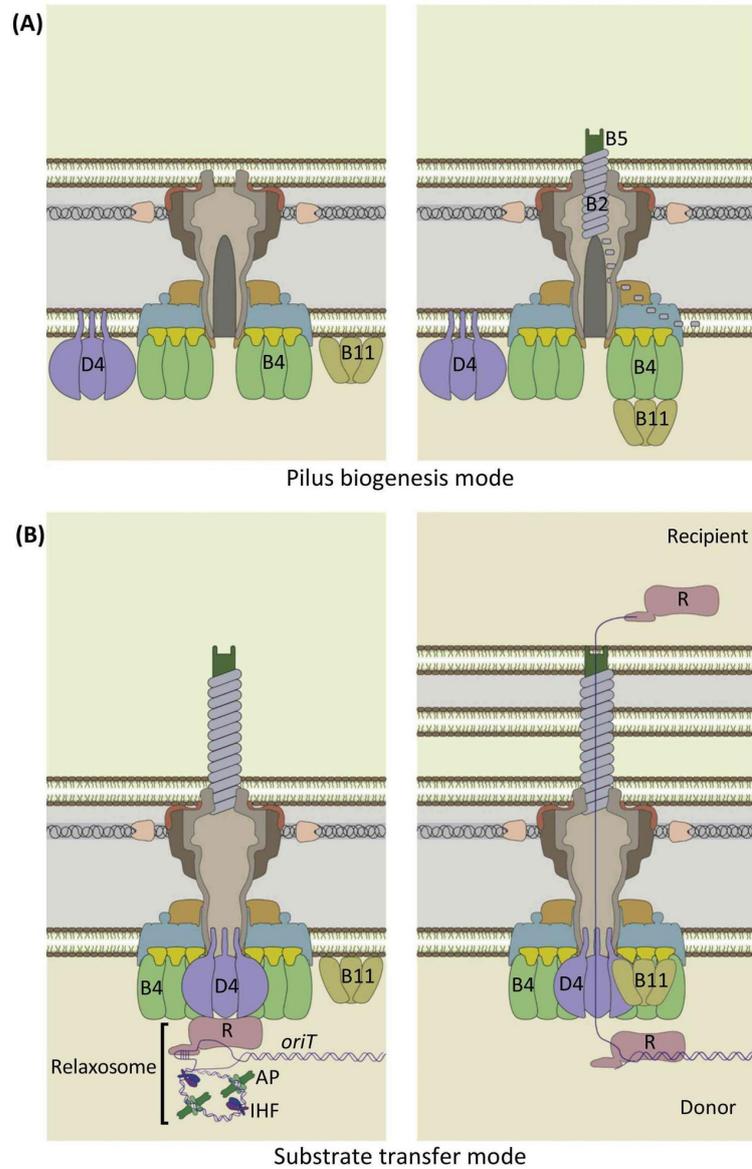
Remarques : contrairement à la représentation, le pili et le pore de conjugaison ne sont pas associés spécifiquement aux plasmides et aux ICE. Par ailleurs, il n'y a pas encore de preuve quant au transfert du complexe ADN-relaxase à l'intérieur du pili comme représenté sur le schéma.

III.2.2 Mécanisme chez les bactéries à Gram négatif

L'ADN qui est transféré pendant la conjugaison via un système de sécrétion de type IV peut être sous la forme simple brin ou double brin. Le transfert d'ADN double brin est notamment retrouvé chez les bactéries mycéliennes (Gram positif) du groupe *Actinobacteria* (Ghinet *et al.*, 2011); nous nous concentrerons sur le transfert d'ADN simple brin. Agissant sur un ADN double brin intégré dans le génome ou dans un plasmide, entre en jeu une protéine appelée relaxase qui joue un rôle essentiel dans la conjugaison. Cette enzyme est parfois accompagnée par un ou plusieurs autres partenaires protéiques pour former le Relaxosome (de la Cruz *et al.*, 2010). Au niveau du site appelé OriT (pour "origine de transfert"), la relaxase réalise un clivage. Elle se lie ensuite à l'extrémité 5' de l'un des brins de l'élément conjugatif (plasmide ou ICE). Ce processus met souvent en jeu une tyrosine catalytique au niveau du site catalytique de la relaxase. Le complexe Relaxosome-ADN est ensuite pris en charge par la protéine de couplage T4CP (VirD4-like) et est transporté au travers du T4SS pour arriver dans la cellule cible (voir Figure 21, page 47). Il faut toutefois noter que les protéines qui s'ajoutent à la relaxase pour former le Relaxosome peuvent être originaires de l'élément mobile, ou du génome de la bactérie hôte. Chez *A. tumefaciens*, la relaxase VirD2 se lie covalamment à une portion du plasmide Ti, l'ADN-T qui sera ensuite pris en charge successivement par différentes protéines du T4SS (Cascales & P. J., 2004) afin d'être transféré dans une cellule de plante. Les relaxases sont de grandes protéines multi-domaines avec plusieurs fonctions (Gomis-Rüth & Coll, 2006), une activité hélicase ou encore catalytique comme la transesterification permettant à la relaxase de cliver l'ADN au niveau de l'OriT, de manière similaire à certaines enzymes de restriction (Sasnauskas *et al.*, 2007).

En prenant l'exemple des T4SS de classe A, le substrat interagit avec différentes protéines en suivant les chemins qui sont représentés sur la figure 17, page 39. La protéine de couplage VirD4 est la première protéine/ATPase à rencontrer le substrat lié à la relaxase depuis le cytoplasme. Les ATPases VirB4 et VirB11 entrent ensuite en jeu et permettent le transfert du substrat grâce à l'hydrolyse de l'ATP. Les protéines structurales VirB6 et VirB8 présentes au niveau du complexe de la membrane interne sont les derniers intermédiaires avant que le substrat atteigne le complexe de coeur de la membrane externe. Le substrat entre ensuite en interaction avec la protéine structurale VirB9 au niveau du complexe de la membrane externe avant de rencontrer les protéines du pilus VirB2 et VirB5.

FIGURE 21 – Les deux modes de fonctionnement d'un T4SS conjugatif Gram négatif



tiré de Ilangovan *et al.*, 2015

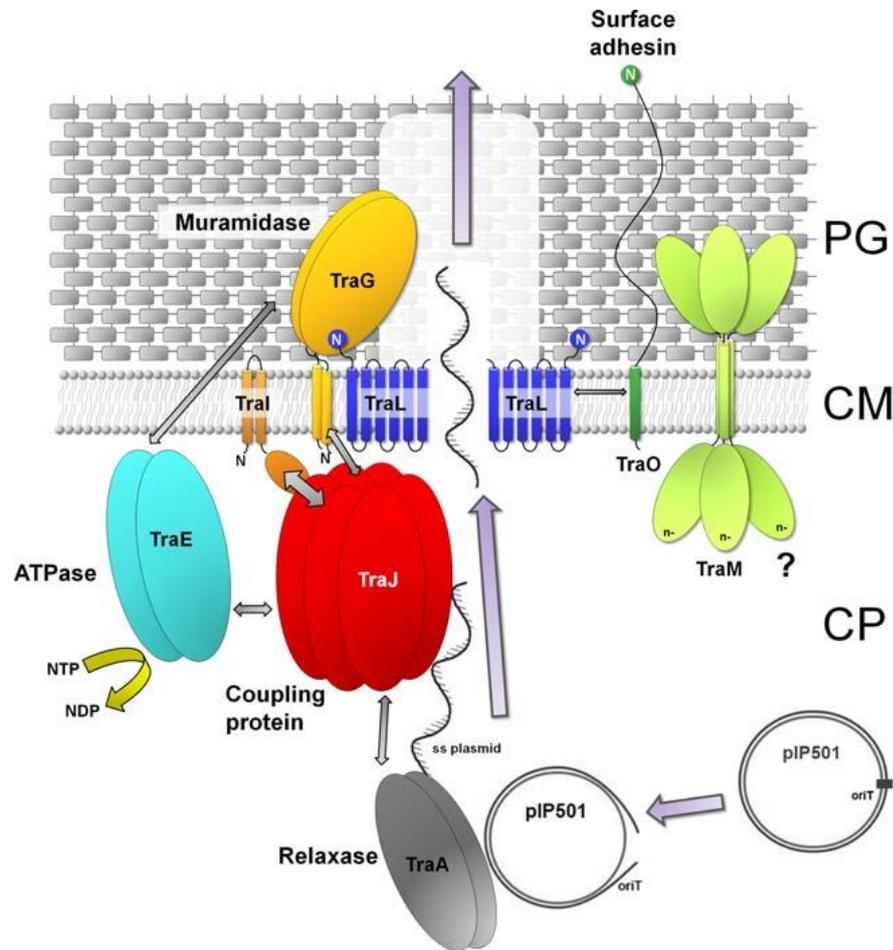
Représentation schématique montrant les deux états de fonctionnement possibles pour un T4SS de bactéries à Gram négatif. (A) Biogenèse du pilus : l'interaction entre VirB4 (vert) et VirB11 (brun clair) permet la croissance du pilus du T4SS. (B) Translocation du substrat : l'interaction de VirB11 (brun clair) et VirD4 (violet) permet au complexe relaxase/ADN de traverser le T4SS jusqu'à la cellule cible. Ici sont montrés les protéines accessoires (AP), la relaxase (R) et les facteurs d'intégration de l'hôte (IHF) qui forment le Relaxosome et prennent en charge l'ADN double brin. En l'occurrence, la relaxase se lie covalamment à l'extrémité 5'-OH de l'un des brins d'ADN, au niveau de la région spécifique *OriT*.

III.2.3 Mécanisme chez les bactéries à Gram positif

A l'inverse des systèmes des bactéries à Gram négatif, les T4SS de bactéries à Gram positif sont moins bien connus (Goessweiner-Mohr, Arends *et al.*, 2013). L'absence d'une seconde membrane induit inévitablement des différences nettes entre les architectures des T4SS, bien qu'une partie de leurs protéines semblent être conservées, notamment celles qui sont retrouvées sur la membrane interne qui s'apparente à la (seule) membrane plasmique des bactéries à Gram positif. En suivant ce principe, on retrouve dans les génomes des bactéries à Gram positif des équivalents aux protéines du complexe de la membrane interne comme VirB8 et VirB6, mais aussi aux ATPases cytoplasmiques comme VirD4 et VirB4. A l'inverse, les protéines du pilus VirB2 et VirB5 ainsi que les protéines du complexe de la membrane externe VirB7, VirB9 et VirB10 n'ont pas encore d'équivalents au sein des systèmes de bactéries à Gram positif. L'hydrolase de peptidoglycane VirB1 présente plusieurs domaines catalytiques au sein des T4SS Gram positif et devient essentielle pour la conjugaison à cause de la couche bien plus épaisse de peptidoglycane chez ces organismes (quelques nm chez les Gram négatif contre 30 à 100 nm chez les Gram positif; Arends *et al.*, 2013; Silhavy *et al.*, 2010; Vollmer *et al.*, 2008). Jouant un rôle important dans le contact entre les deux cellules, les adhésines de surface permettent de renforcer le contact entre deux cellules, nécessaire pour la conjugaison (Navarre & Schneewind, 1999).

D'un point de vue fonctionnel, les premières étapes de conjugaison chez les T4SS Gram positif sont similaires aux T4SS Gram négatif, de la prise en charge de l'ADN par le Relaxosome jusqu'à l'action de la protéine de couplage (Goessweiner-Mohr *et al.*, 2014; Grohmann *et al.*, 1999; Grohmann *et al.*, 2003; Zechner *et al.*, 2012). En revanche, la translocation du complexe relaxase/ADN diffère considérablement. Cela est probablement dû aux différences de membranes (Goessweiner-Mohr *et al.*, 2014). Les protéines de la membrane externe n'étant pas retrouvées au sein de ces systèmes, les protéines de la seule membrane interne doivent permettre le transfert. Parmi les protéines structurales qui sont retrouvées au sein des systèmes Gram positif, VirB8 serait l'un des candidats pour former le pore de conjugaison grâce aux nombreuses interactions qu'elle semble mettre en oeuvre dans les systèmes étudiés (Fercher *et al.*, 2016; Porter *et al.*, 2011)

FIGURE 22 – Modèle actuel de transfert de plasmide via un T4SS Gram positif



tiré de Goessweiner-Mohr *et al.*, 2014

Représentation schématique du transfert d'ADN par le T4SS de pIP501 chez *Enterococcus faecalis*. Tout d'abord, la relaxase TraA/VirD2 se lie à l'origine de transfert OriT. Ensuite, le plasmide simple brin est transporté vers le canal de transfert par la protéine de couplage TraJ/VirD4. La peptidoglycane hydrolase TraG/VirB1 permet de remodeler le peptidoglycane pour permettre le transfert. La protéine structurale TraM/VirB8 est située sur la membrane interne et aurait un rôle essentiel dans la structuration du canal de transfert. L'adhésine TraO est une protéine permettant à la bactérie d'adhérer à la cellule cible durant le transfert. Elle jouerait notamment un rôle d'adhérence équivalent au pilus dans les systèmes à Gram négatif. L'ATPase TraE/VirD4 qui s'assemble en oligomère permettrait de fournir l'énergie nécessaire à l'assemblage et/ou au transfert du complexe nucléoprotéique. Les flèches indiquent l'interaction obtenue par double hybride (levure), et l'épaisseur de la flèche est à la mesure de la force d'interaction.

IV Fonction et structure des principales protéines de la conjugaison

Parmi les différentes protéines qui composent et animent les systèmes de sécrétion de type IV conjugatifs, certaines sont retrouvées dans la quasi totalité des organismes étudiés, qu'ils soient à Gram négatif ou positif. Toutefois, la faible connaissance des systèmes à Gram positif ne nous permet pas de faire une comparaison directe entre les différentes fonctions possibles de ces protéines. Il apparaît logique que, si elles sont présentes de manière ubiquitaire, elles aient au moins un rôle important, voire essentiel. Dans cette partie nous allons faire le point des connaissances actuelles sur ces quelques protéines, et, lorsque ce sera possible, nous détaillerons notamment la relation structure-fonction de ces molécules qui semblent indispensables au bon fonctionnement du processus de conjugaison. Bien que la conjugaison d'ADN double brin soit tout à fait possible (cf. les Actinobactéries), et que des processus apparentés à la conjugaison existent entre bactéries et eucaryotes (Lacroix & Citovsky, 2018), nous nous concentrerons sur l'échange d'ADN simple brin entre bactéries.

Citons par exemple la relaxase, qui est l'endonucléase permettant à la fois de cliver l'élément mobile et de le guider jusqu'au canal créé par la machinerie du T4SS. L'hydrolase de peptidoglycane est également retrouvée dans de très nombreux systèmes, permettant de remodeler le peptidoglycane présent à la fois chez les bactéries à Gram négatif, et d'autant plus chez celles à Gram positif. L'ATPase VirB4 et la protéine de couplage VirD4 sont également conservées quel que soit le Gram, de même que les protéines structurales de la membrane interne. En particulier, VirB8 a probablement un rôle clé dans le transfert du substrat au sein des T4SS de bactéries à Gram positif, puisque c'est l'une des dernières protéines de la membrane interne à entrer en jeu.

IV.1 Relaxases conjugatives

La toute première fois qu'il est fait mention de complexes de relaxation remonte aux années 1970, pour un complexe protéine-ADN qui avait la curieuse tendance, une fois isolé, à passer d'un état superenroulé à un état linéaire (Clewell & Helinski, 1969 ; Helinski & Clewell, 1970). La protéine en question a ensuite été nommée relaxase, pour sa capacité à "relaxer" un ADN circulaire en réalisant une coupure sur un site spécifique par une réaction de transestérification (Traxler & Minkley, 1988). A partir de la découverte des premières relaxases, de nombreuses études ont suivi et ont permis de

mieux comprendre le fonctionnement de cette protéine, par exemple sa capacité à se lier à l'extrémité 5'-phosphate de l'un des brins au site de coupure (Pansegrau *et al.*, 1990 ; Scherzinger *et al.*, 1993). Au sein de toutes les relaxases étudiées à l'époque, même les relaxases Gram positif, un motif particulier "HUH" a été découvert (Guzmán & Espinosa, 1997 ; Ilyina & Koonin, 1992 ; Mendiola & de la Cruz, 1992). Ce motif de résidus histidine/hydrophobe/histidine complété d'une troisième histidine située ailleurs dans la séquence forme une triplette permettant notamment de coordonner un ion métallique à proximité du site catalytique. La présence de ce motif HUH a dicté l'élaboration d'une classification préliminaire dans laquelle les différentes relaxases sont nommées "MOB" en suivant la classification des plasmides en vigueur à cette époque. En effet, ces derniers étaient regroupés par groupe d'incompatibilité. Deux plasmides du même groupe ne peuvent coexister dans une cellule, puisqu'ils utilisent les mêmes outils de régulation et de répllication ; l'un des deux sera donc éliminé au fil des générations. Ces groupes d'incompatibilité, par exemple F, ont été utilisés pour nommer les relaxases conjugatives retrouvées dans leurs séquences, par exemple MOB_F. "MOB" signifie protéine de "mobilisation" et désigne la capacité des relaxases à "mobiliser" un ADN, notamment pour la conjugaison.

Depuis lors, de nombreuses nouvelles relaxases conjugatives ont été découvertes, et cette classification préliminaire qui semblait concorder n'a pas tenu. En effet, sur les familles que l'on connaît aujourd'hui, certaines n'utilisent pas une tyrosine catalytique, mais deux tyrosines ou bien une histidine. D'autres ne présentent pas la même organisation de motifs HUH, et certaines relaxases ne se lient pas covalamment à l'ADN. Aujourd'hui, une classification plus complète a été construite (Guzmán-Herrador & Llosa, 2019 ; Zechner *et al.*, 2018) et classe les relaxases en différentes familles de protéines "MOB" avec des caractéristiques similaires (voir Figure 23, page 53). Cette classification n'est pas forcément concordante avec l'ancienne, si ce n'est que les lettres en indice ont été conservées sans forcément avoir de rapport avec les anciennes classes.

Les relaxases appartenant à de nombreuses classes différentes et ayant des différences significatives entre elles, nous allons nous attarder pour exemple sur une structure en particulier, résolue par cryomicroscopie électronique, celle de la relaxase TraI des plasmides de la famille F (Ilangovan *et al.*, 2017). Cette famille comporte par exemple le plasmide R1 ou encore pED208, isolés depuis des bactéries pathogènes humaines. La relaxase TraI, conservée chez les plasmides de cette famille, possède tout ce que l'on attend d'une relaxase de la famille F : un domaine N-terminal transtesté-

FIGURE 23 – Classification des relaxases connues

MOB Family ^a	F	P	Q	V	C	H	T	TcpM ^b
Prototype relaxase	R388-TrwC	RP4-TraI	RSF1010-MobA	pMV158-MobM	pAD1-TraX	GGI-TraI	Tn916-Orf20	pCW3-TcpM
3D Fold ^c	HUH	HUH,HEN	HUH	HUH	RE	HD	Rep-trans	Y-rec
Catalytic residue	Tyr x2	Tyr	Tyr	His	Tyr	Tyr		Tyr
Covalent complex ^d	Yes	Yes	Yes	Yes	No	Yes?		
2nd Function ^e	Pre Mob	Pre Mob rTn	Pre* Mob	Pre Mob Cop	Pre	Pre Rep	Rep	

tiré de Guzmán-Herrador et Llosa, 2019

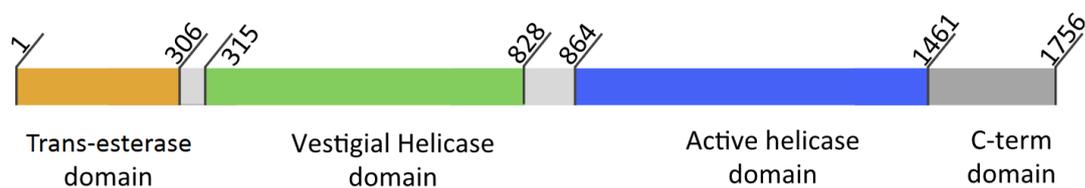
a. Tiré de la classification originale par Garcillán-Barcia *et al.*, 2009 et Guglielmini *et al.*, 2011. b. Cette relaxase a été décrite après que la classification MOB a été créée, mais ne correspond à aucune des familles existantes. c. Familles structurales basées sur la présence de motifs/repliements signatures (ex : HUH) d. "Yes" : complexe relaxase-ADN détecté expérimentalement ; "Yes ?" : preuve indirecte qui suggère que l'extrémité 5' de l'ADN est protégée ; "No" : a été étudié mais pas détecté ; Vide : pas d'information. e. Une fonction autre que le transfert conjugatif a été rapportée. Mob : activité sur séquences hétérologues OriT ; Pre : enzyme de recombinaison de plasmide (Pre* : seulement sur des substrats ADN simple brin) ; Rep : initiateur de la réplication plasmidique ; Cop : régulation du nombre de copies de plasmide ; rTn : activateur de rétrotransposition (ARN vers ADN).

rase, un domaine hélicase inactif qui joue plutôt le rôle de domaine de fixation à l'ADN, un domaine hélicase actif et un domaine C-terminal (voir Figure 24, page 54, explications dans les légendes).

Le positionnement des domaines de la relaxase TraI est illustré dans la Figure 25, page 55. Les deux domaines hélicases permettent à l'ensemble de la relaxase d'interagir et de relaxer l'ADN au niveau du site spécifique OriT sur le plasmide afin de réaliser la coupure ainsi qu'une liaison covalente grâce à une tyrosine catalytique située sur le domaine transestérase. Situé sur ce même domaine, la triplette d'histidines (formée en partie par le motif HUH) coordine l'ion métallique. Malheureusement, de par sa nature supposée flexible, le domaine C-terminal n'apparaît pas sur le modèle structural.

De nombreuses autres données sur le fonctionnement de cette relaxase sont disponibles dans la publication de Ilangovan *et al.*, 2017, notamment l'existence d'un probable couplage entre la relaxase et VirD4 (ATPase, protéine de couplage) au sein du T4SS. La relaxase, étant indispensable à la conjugaison dans les systèmes de sécrétion de type IV, est évidemment une cible potentielle dans la lutte contre le transfert horizontal de gènes (Soler *et al.*, 2019), et plus particulièrement la lutte contre la diffusion des résistances aux antibiotiques. Mieux comprendre comment les différentes classes de relaxases jouent leur rôle pourrait permettre de trouver de nouvelles molécules inhibitrices.

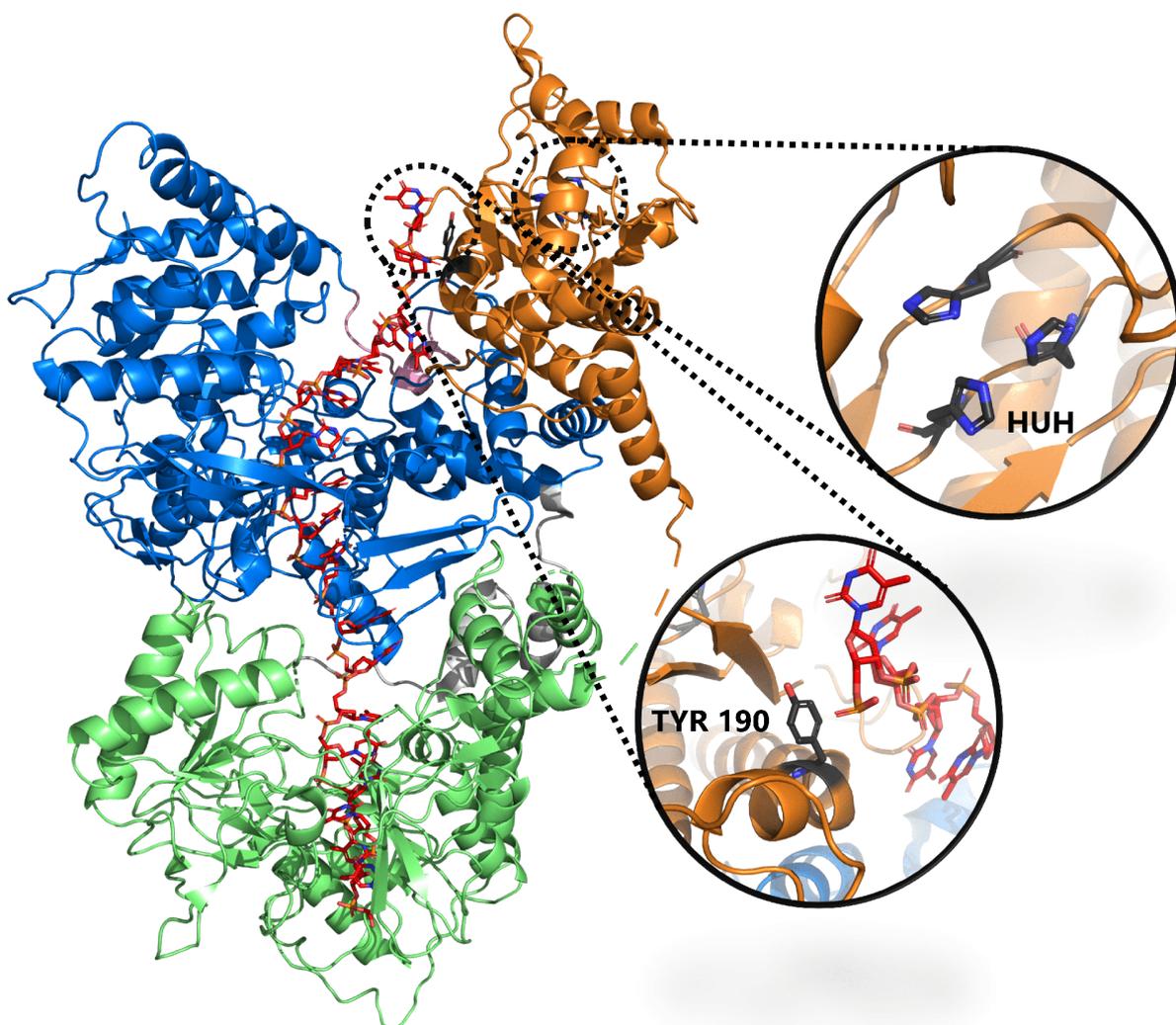
FIGURE 24 – Organisation des domaines de la relaxase TraI



tiré de Ilangovan *et al.*, 2017

Schéma de l'organisation des différents domaines fonctionnels connus au sein de la relaxase TraI. En orange entre l'extrémité N-terminale et l'acide aminé 306 se trouve le domaine transestérase, responsable de la liaison covalente à l'ADN simple brin grâce à une tyrosine catalytique Tyr190 qui est située proche de l'extrémité 5'-phosphate de l'ADN simple brin. En vert, les acides aminés 315 à 828 correspondent à une hélicase non active, mais qui joue le rôle de plate-forme pour l'ADN simple brin. En bleu, entre les acides aminés 864 à 1461 forment une hélicase active qui sépare efficacement l'ADN double brin en deux brins distincts. La relaxase progresse sur le brin avec l'extrémité 5', tandis que le second brin avec l'extrémité 3' est repoussé dans l'environnement. Enfin, de l'acide aminé 1461 jusqu'à l'extrémité C-terminale se trouve le domaine C-terminal, qui jouerait le rôle de plate-forme d'interaction pour les autres partenaires du Relaxosome.

FIGURE 25 – Modèle de la relaxase TraI obtenu par cryomicroscopie électronique



Représentation de la structure de TraI (PDB : 5N8O) via le logiciel de modélisation PyMOL (Schrödinger, LLC, 2021). Les couleurs et délimitations des domaines correspondent à ceux de la figure 24, page 54 à l'exception du domaine C-terminal qui n'apparaît pas sur le modèle, probablement à cause de la flexibilité. A gauche se trouve une vue globale du modèle, et à droite des zooms, en bas sur la tyrosine catalytique Tyr190 proche de l'extrémité 5' de l'ADN simple brin, et en haut sur la tripléte d'histidines permettant de coordonner un ion métallique (absent de la structure). Les orientations de ces zooms diffèrent de celle de la vue principale pour représenter au mieux chaque élément visualisé.

IV.2 Protéines de couplage (T4CP)

Les protéines de couplages des T4SS (ou T4CP) sont des ATPases membranaires situées au niveau de la face cytoplasmique, comme nous avons pu nous en apercevoir dans certains des exemples précédents. Elles sont appelées VirD4 par comparaison à la nomenclature du T4SS d'*A. tumefaciens*. Ces protéines sont essentielles à la grande majorité des T4SS (Gonzalez-Rivera *et al.*, 2016; Grohmann *et al.*, 2018) et jouent un rôle d'intermédiaire entre le substrat accompagné par le Relaxosome, et le canal membranaire formé par les différentes autres protéines du T4SS. Toutefois, cette nature d'intermédiaire n'est pas favorable à la détermination d'une structure montrant le T4SS avec la protéine de couplage. Pour l'instant, les structures disponibles pour les protéines de couplage et leurs partenaires ne permettent pas encore de positionner la T4CP dans le T4SS. A l'heure actuelle, une seule étude rapporte l'isolation d'un T4SS contenant la protéine VirD4 (Redzej *et al.*, 2017), mais la structure de ce complexe n'est pas encore disponible.

Cette nature dynamique reste très intéressante, et puisque la protéine de couplage VirD4 est généralement essentielle à la conjugaison via les T4SS, elle reste une cible principale dans l'étude de ces systèmes. De par sa fonction, VirD4 doit également interagir avec différents acteurs protéiques, notamment le Relaxosome et les autres protéines du complexe de la membrane interne du T4SS chez les bactéries à Gram négatif. Les protéines de couplage sont également retrouvées chez les bactéries à Gram positif et adopteraient la même fonction puisque la membrane interne (et probablement un équivalent du complexe de la membrane interne) est retrouvée chez ces organismes. Aucune structure n'en est disponible. Tout comme les relaxases, les T4CP sont également des cibles pour limiter la dissémination de résistances aux antibiotiques chez les pathogènes humains (Álvarez-Rodríguez *et al.*, 2020).

Sur la base des domaines présents dans la protéine, une classification originale a été établie par Alvarez-Martinez et Christie, 2009. La Figure 26, page 57 présente ces différentes familles (Llosa et Alkorta, 2017). Le domaine fonctionnel des T4CP est le seul domaine universellement conservé contenant les motifs Walker A et Walker B qui sont des séquences d'acides aminés permettant la fixation et l'hydrolyse de l'ATP (Walker *et al.*, 1982), vérifiées *in vitro* (Christie *et al.*, 2014). La plupart des T4CP étudiées possèdent également un domaine transmembranaire et forment des hexamères, mais il existe des T4CP dimériques et des T4CP sans domaine transmembranaire. La

FIGURE 26 – Classification des protéines de couplage connues

Name Prototype	Membrane insertion			CTD	Examples
	TM helixes	Periplasmic loop	Associated membrane protein ^a		
VirD4-like TrwB _{R388}	≥ 2	30–60 aa	No	Yes (some members)	Gram-negatives: VirD4 _{Al} , TrwB _{R388} , TraD _F , TraG _{RP4} , HP0524 _{Hpy} (Cagβ) Gram-positives: PcfC _{pCF10} , YdcQB _{ICEB51}
TraG-J pairs TraG _{R27}	≥ 2	≈4 aa	TraJ	No	Gram-negatives: TraG _{R27} , TraG _{R478} , P081 _{pWWO}
No TMD TraJ _{pIP501}	No	No	One subfamily	No	Gram-negatives: <i>Pseudomonas syringae</i> pSR1 Gram-positives: <i>Staphylococcus aureus</i> pSK41
FtsK-like TcpA _{pCW3}	≥ 2	Yes	No	Yes (TcpA _{pCW3}) ^b	Gram-negatives: TraG _{pKLC102} , Orf16 _{pAb559} Gram-positives: TcpA _{pCW3} , TraB _{pSG5}
Archaeal TraG _{Sis}	No	No	No	No	<i>Sulfolobus islandicus</i> pKEF9

tiré de Llosa et Alkorta, 2017

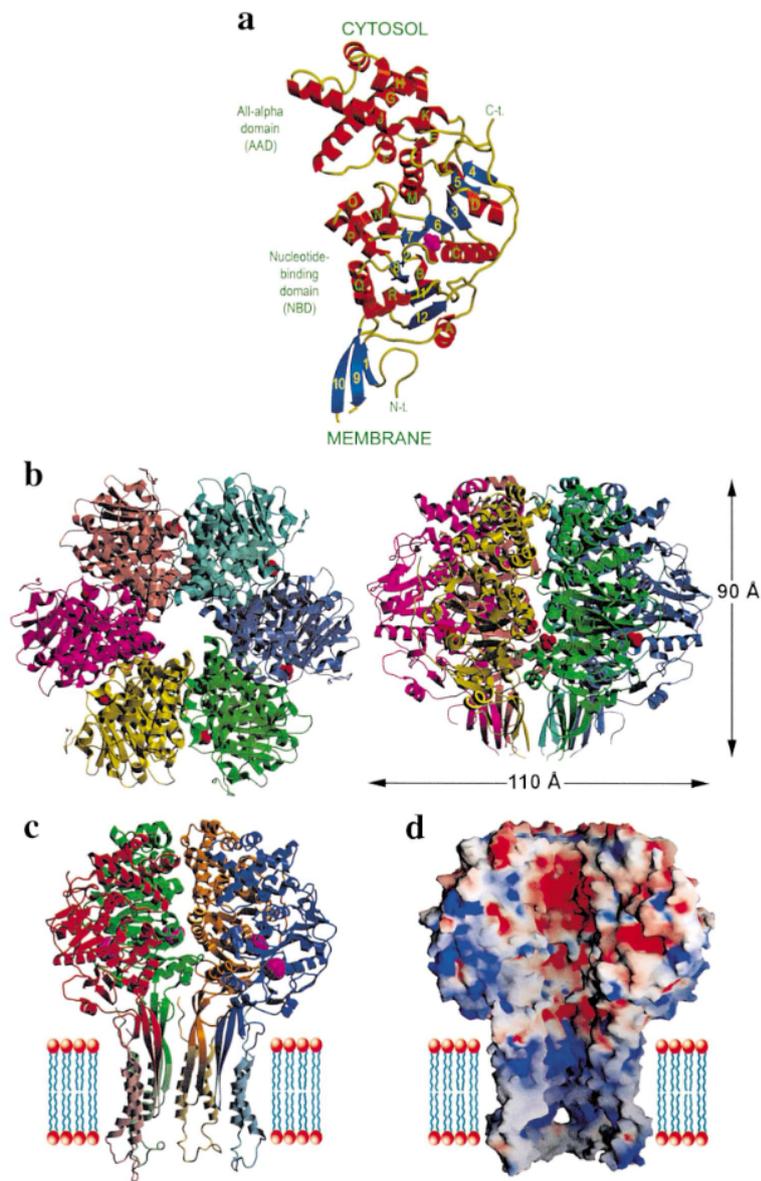
- a. Présence d'un gène adjacent codant pour une petite protéine (+/- 200 aa, 2 à 5 segments transmembranaires)
b. L'extrémité C-terminale de 61 acides aminés n'est pas essentielle pour l'activité de la protéine de couplage TcpA codée par le plasmide pCW3 chez *Clostridium perfringens* (Parsons *et al.*, 2007)

revue de Llosa et Alkorta, 2017 détaille les différentes T4CP et leur rôle au sein des systèmes de sécrétion.

Concernant les structures de T4CP disponibles, la première à avoir été résolue est celle du domaine cytoplasmique de TrwB codée par le plasmide R388 (Gomis-Rüth & Coll, 2001). Depuis, TrwB a été considérée comme la T4CP de référence, étant la seule structure disponible pendant de nombreuses années suivant la publication originale. Dans la Figure 27, page 58 TrwBΔN70 est représentée sous différentes orientations. Cette protéine de couplage de classe VirD4-like possède un segment transmembranaire (non présent dans la structure) de 70 résidus lui permettant de s'intégrer à la membrane interne. Dans la structure, TrwBΔN70 s'assemble en homohexamères d'environ 110 Å de large comportant un canal central. Des clichés de la protéine TrwB obtenus par microscopie électronique à transmission par Hormaeche *et al.*, 2002 confirment la structure résolue par Gomis-Rüth *et al.*, 2001. Plus récemment, en 2017 (Kwak *et al.*, 2017) et en 2020 (Meir *et al.*, 2020) sont parus les modèles du complexe de couplage T4CC que nous avons déjà mentionnés dans la partie sur l'architecture des T4SS de classe B, chez *L. pneumophila*.

Dans le dernier modèle par Meir *et al.*, 2020, la T4CP (DotL) joue un rôle central en formant un hexamère fonctionnel intégré dans la membrane, accompagné de plusieurs partenaires protéiques. En l'occurrence, chaque T4CP intègre un hétéropentamère avec les protéines DotN, DotM, DotZ et DotY

FIGURE 27 – Structure du domaine cytoplasmique et fonctionnel de la protéine de couplage TrwB



tiré de Gomis-Rüth et Coll, 2001

a. Représentation d'un monomère de TrwB Δ N70 . Les hélices annotées par des lettres majuscules sont en rouge, les brins sont numérotés et en bleu. Le domaine transmembranaire est absent puisqu'il n'est pas compris dans la construction cristallisée. Deux domaines sont présents, un domaine tout hélice alpha (AAD) et le domaine fixateur de nucléotide (NBD) contenant les motifs Walkers A et B. b. Représentation de la structure complète contenant un homo-hexamère de TrwB Δ N70 vu depuis le cytosol selon l'axe de symétrie de l'hexamère (vue gauche) et perpendiculairement à cet axe (vue de droite). c. Représentation de 4 des 6 monomères de TrwB Δ N70 , vus perpendiculairement à l'axe, avec les domaines transmembranaires modélisés (résidus 1-70). d. Identique à c mais en montrant la surface de la protéine (Connolly) afin de mettre en évidence le canal central qui permet de connecter le cytoplasme (situé vers le haut) au périplasma (situé vers le bas). Codes PDB : 1E9S, 1E9R, 1GKI, 1GL6, 1GL7

dans ce modèle. Répété 6 fois dans le modèle, on obtient à la fin un hexamère d'hétéropentamère, expression particulièrement difficile à expliquer durant un dîner de famille. On retrouve le même canal chez la T4CP DotL que chez TrwB Δ N70, et le modèle fonctionnel propose que ce canal soit utilisé par le substrat pour traverser la membrane interne vers le périplasme. Toutefois, nous n'avons toujours pas d'information structurale permettant de faire le lien entre le T4CC et les complexes membranaires du T4SS.

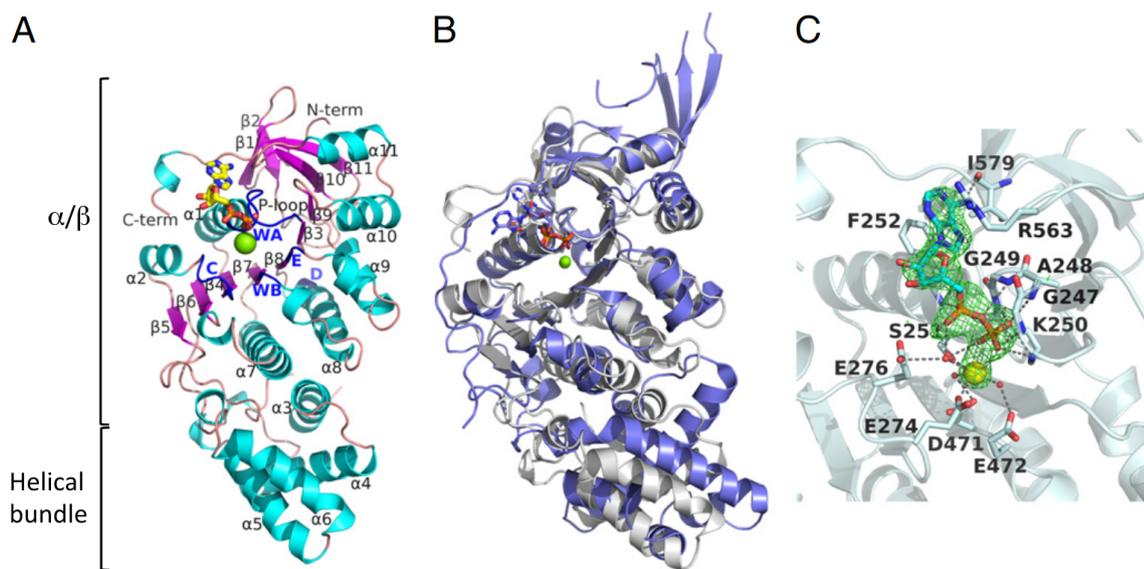
IV.3 L'ATPase VirB4

Tout comme la protéine de couplage, VirB4 est une ATPase essentielle pour le bon fonctionnement d'un système de sécrétion de type IV, et elle est globalement l'une des protéines les mieux conservées au sein de ces systèmes, tout organisme confondu (Wallden *et al.*, 2010). Son activité dépend également de son état d'oligomérisation, et dans les cas étudiés elle est active sous la forme d'hexamères bien qu'elle soit stable sous différentes autres formes en solution (Arechaga *et al.*, 2008; Durand *et al.*, 2010; Wallden *et al.*, 2010). VirB4 est la dernière ATPase à intervenir dans le transfert du substrat entre l'ATPase VirB11 et les protéines structurales VirB6 et VirB8, mais elle n'interagit pas directement avec l'ADN (Cascales & P. J., 2004). A l'heure actuelle, une seule structure de la protéine seule existe : elle a été résolue en radiocristallographie par Walldén *et al.*, 2012 et est présentée dans la Figure 28, page 60. En plus d'y présenter les différentes particularités de la structure, une comparaison y est faite avec la protéine de couplage VirD4 qui est également une ATPase. Malgré une identité de séquence faible de l'ordre de 12% (Walldén *et al.*, 2012), VirB4 et D4 présentent des organisations similaires de leur domaine C-terminal avec un sous-domaine tout alpha et un second sous-domaine fonctionnel contenant, entre autres, les motifs Walker A & B permettant à la protéine de fixer et hydrolyser l'ATP.

Tout comme la protéine de couplage TrwB Δ N70 présentée dans la sous-partie précédente, l'ATPase VirB4 est membranaire et la structure cristallographique résolue ne comprend pas cette partie membranaire. Cette comparaison s'arrête là puisqu'à l'inverse des protéines VirD4, plusieurs modèles résolus en cryomicroscopie électronique comportent des oligomères de VirB4, notamment le double hexamère présent dans le T4SS de classe A du plasmide R388 (revoir Figure 17 page 39) et l'hexamère de la protéine DotO présent dans le T4SS de classe B du système Dot/Icm (revoir Figure 18 page 41). Dans la publication de Walldén *et al.*, 2012 concernant la structure de TpsVirB4_{CTD}, est

également présentée une structure de cryomicroscopie électronique montrant le segment N-terminal en interaction avec la membrane et avec les protéines du complexe de la membrane interne composé des protéines VirB3, VirB6 et VirB8. Ceci concorde avec les observations des autres T4SS de classe A chez les bactéries à Gram négatif ainsi qu'avec les données obtenues quant au chemin emprunté par le substrat. Malheureusement, comme avec la protéine de couplage VirD4, bien que la fonction et l'importance de ces protéines soient certainement très similaires dans les systèmes retrouvés chez les bactéries à Gram positif, nous ne disposons pas encore de structure. Certaines publications font mention d'hybrides entre VirB4 et VirB3, l'une des protéines structurales retrouvées dans le complexe de la membrane interne dont la fonction et le rôle restent encore peu documentés (Batchelor *et al.*, 2004; Christie *et al.*, 2005; Fronzes, Christie *et al.*, 2009).

FIGURE 28 – Structure du domaine cytoplasmique et fonctionnel de la protéine VirB4



tiré de Walldén *et al.*, 2012

(A) Représentation de la structure de TpsVirB4_{CTD}, résidus 205 à 587 (PDB : 4AG5). Les motifs Walker A et Walker B, ainsi que les motifs C, D et E sont indiqués en bleu. (B) Superposition des structures TpsVirB4_{CTD} et TrwBΔN70 (PDB : 1GKI). Les molécules d'ADP et les ions Mg²⁺ pour chaque structure sont présents. (C) Site de fixation de nucléotide (NBD). La carte Fo-Fc (3σ) est représentée pour l'ADP (en vert) et pour l'ion catalytique Mg²⁺. Les molécules d'eau sont représentées sous forme de sphères rouges. Tous les résidus impliqués dans l'interaction avec le nucléotide sont représentés en bâtonnets.

IV.4 Hydrolases de peptidoglycane VirB1

Comme son nom l'indique, l'hydrolase de peptidoglycane HP (aussi connue sous le nom d'hydrolase de paroi cellulaire) permet de remodeler localement le peptidoglycane afin de faciliter l'insertion dans la membrane des différents constituants qui s'y trouvent. A l'échelle cellulaire, de nombreuses HP existent et agissent en permanence en équilibre avec la synthèse et l'assemblage du peptidoglycane ; il est estimé qu'un peu moins de la moitié de la masse en peptidoglycane d'une cellule est renouvelée durant une génération cellulaire (Goodell, 1985). Pour une revue complète sur les différentes hydrolases de la paroi bactérienne, consultez l'article de Vermassen *et al.*, 2019. Dans le cas des T4SS, les HP sont importantes chez les bactéries à Gram négatif, et deviennent presque toujours essentielles chez les bactéries à Gram positif du fait des différences dans leurs parois bactériennes que nous avons déjà discutées précédemment dans la Figure 1, page 6 (Clark *et al.*, 2018 ; Pajerski *et al.*, 2019). En l'occurrence, une HP possède un ou plusieurs domaines catalytiques qui permettent de spécifiquement hydrolyser des liaisons clés dans la structure du peptidoglycane. Pour rappel, le peptidoglycane est un assemblage hybride entre des petits peptides et des sucres, et les HP peuvent le cliver soit dans la partie sucre, soit dans la partie peptide, soit les deux. Toutefois, il faut noter que bien que certaines enzymes de cette famille soient appelées hydrolases, toutes n'utilisent pas l'hydrolyse comme mécanisme d'action. Il existe donc 3 types d'HP : 1) les amidases qui coupent une liaison amide spécifique entre un sucre et un acide aminé ; 2) les glycosylases qui coupent les liaisons entre deux glycanes ; 3) les peptidases qui coupent les liaisons peptidiques entre acides aminés.

La première hydrolase de peptidoglycane à avoir été étudiée dans le cadre d'un système de sécrétion de type IV de classe A est VirB1 chez *A. tumefaciens* (Baron *et al.*, 1997). VirB1 est une transglycosylase lytique qui est retrouvée sous forme membranaire en contact avec les différentes protéines structurales et fonctionnelles du T4SS comme VirB8, VirB9 et VirB11 (Höppner *et al.*, 2005). La délétion de VirB1 réduit le taux de transfert sans l'abolir (Baron *et al.*, 1997 ; Berger & Christie, 1994 ; Fronzes, Christie *et al.*, 2009 ; Llosa *et al.*, 2000), mais VirB1 est essentielle à la formation du pilus (Fronzes, Christie *et al.*, 2009 ; Fullner *et al.*, 1996). VirB1 est divisée en trois parties. En premier, un peptide signal qui permet à la protéine d'être transportée vers le périplasme et qui est clivé par la même occasion. Le deuxième domaine, ou domaine N-terminal une fois le peptide signal clivé, possède une activité transglycosylase (nommé SLT pour transglycosylase lytique soluble) et est retrouvé dans le périplasme tandis que le domaine C-terminal (CTD, aussi appelé

VirB1*) est clivé puis libéré dans le milieu extérieur (Baron *et al.*, 1997 ; Sgro *et al.*, 2019). VirB1* est supposé favoriser la formation du pilus (Baron *et al.*, 1997), d'où le fait que VirB1 soit souvent associée aux protéines du pilus VirB2 et VirB5. L'une des caractéristiques principales des VirB1-like chez *Xanthomonas* est la présence d'une sérine conservée précisément au site de clivage entre le domaine SLT et le domaine VirB1*-like (Sgro *et al.*, 2019). Malheureusement, peu d'informations sont disponibles sur VirB1, et aucune structure n'est disponible dans le contexte des T4SS. Il est possible que cela soit dû à un manque d'intérêt concernant VirB1 puisque ces protéines ne sont pas essentielles pour la conjugaison, à l'inverse des ATPases ou de certaines protéines structurales comme VirB8 qui sont tout à fait indispensables au bon fonctionnement de ces systèmes. En ce sens, les HP chez les bactéries à Gram négatif ne semblent pas des cibles prioritaires pour l'élaboration d'un traitement.

Chez les bactéries à Gram positif, plusieurs hydrolases de peptidoglycane ont été étudiées, notamment chez *Enterococcus faecalis* avec TraG retrouvé au sein du plasmide pIP501 (Arends *et al.*, 2013) et PrgK sur le plasmide pCF10 (Gomez *et al.*, 2014), ou encore TcpG chez *Clostridium perfringens* (Bantwal *et al.*, 2012). Ces HP comportent plusieurs domaines mais sont différents des transglycosylases retrouvées au sein des T4SS Gram négatif. En effet, on retrouve au moins le domaine SLT chez ces deux protéines, ainsi qu'un domaine amidase CHAP (pour Cystéine-Histidine amidohydrolase/peptidase) ou NplC/P60 (Bateman & Rawlings, 2003 ; Duchêne *et al.*, 2019). Considérant leur importance à cause de l'épaisse couche de peptidoglycane, il n'est pas étonnant de trouver plus d'une HP par système conjugatif Gram positif (TcpG et TcpI chez *Clostridium perfringens*). Toutefois, seul TcpG est indispensable au bon fonctionnement du système (Bantwal *et al.*, 2012).

L'une des principales différences avec les VirB1-like est la présence d'une hélice transmembranaire chez certaines HP Gram positif qui permet à la protéine de s'ancrer dans la membrane plasmique. Cela soulève plusieurs questions sans réponse, en particulier : comment une protéine qui fait approximativement 5 nm de long (en considérant ses régions flexibles) peut-elle agir sur une couche si épaisse de peptidoglycane en étant attachée à la membrane ? Ces protéines subissent-elles également un clivage ? Comment se fait-il qu'aucune de ces études n'ait détecté les HP dans le milieu environnant des bactéries ? Malheureusement, tout comme VirB1, nous n'avons que peu d'information sur ces protéines, et aucune structure résolue. Cela étant dit, l'importance des hydrolases de peptidoglycane des T4SS Gram positif devrait favoriser les études sur ces protéines, étant des cibles

potentielles contre la diffusion de résistances aux antibiotiques au sein de ces populations.

IV.5 La protéine structurale VirB8

Les protéines structurales VirB8 sont des petites protéines membranaires d'environ 200 à 300 acides aminés qui font partie du complexe de coeur de la membrane interne. Elles jouent à la fois le rôle de facteur d'assemblage en interagissant avec de nombreux partenaires protéiques (VirB1, VirB4, VirB6, VirB9, VirB10 et VirB11) et permettent également la translocation du substrat en interagissant avec ce dernier (Baron, 2006 ; Casu *et al.*, 2018 ; Höppner *et al.*, 2005 ; Kumar *et al.*, 2000 ; Sharifahmadian *et al.*, 2017 ; Yuan *et al.*, 2005). Par exemple, chez *A. tumefaciens* : VirB4, VirB6 et VirB8 interagissent ensemble et contribuent à l'assemblage du pilus (Yuan *et al.*, 2005). De plus, VirB8 est essentielle au bon fonctionnement de ces systèmes, que ce soit chez les bactéries à Gram négatif ou positif (Baron, 2006 ; Buhrdorf *et al.*, 2003 ; Fercher *et al.*, 2016 ; Kuroda *et al.*, 2015). Par opposition aux protéines que nous avons déjà pu discuter, VirB8 n'a pas été aussi timide que ses camarades quant aux structures disponibles. A l'heure actuelle, on compte plus d'une dizaine d'homologues structuraux à VirB8 d'*A. tumefaciens*, dont quelques unes chez des bactéries à Gram positif (Voir Figure 29, page 64). En revanche, bien que sa localisation soit à peu près connue (au sein du complexe de la membrane interne), son positionnement exact vis à vis de ses autres partenaires n'a pas encore été résolu.

L'une des toutes premières VirB8 à avoir été résolue correspond à son domaine soluble chez *Brucella suis* (Terradot *et al.*, 2005) et est présentée dans la Figure 30, page 65. A première vue, VirB8 adopte un repliement de type "hotdog" avec un feuillet β de 4 brins représentant le bun qui entoure plusieurs hélices α , ou saucisses selon le point de vue. Ce repliement global appartient à la superfamille des domaines NTF2-like qui comprend notamment les facteurs nucléaires eucaryotes NTF2 ou encore des protéines kinases à Calcium/Calmoduline (Eberhardt *et al.*, 2013), mais avec moins de brins dans les feuillets. Il est important de noter que le repliement NTF2 est également impliqué dans des interactions protéines-protéines permettant le transfert de molécules au travers de l'enveloppe nucléaire (Bayliss *et al.*, 2002 ; Bayliss *et al.*, 1999 ; Chaillan-Huntington *et al.*, 2001 ; Chaillan-Huntington *et al.*, 2000).

Au sein de l'unité asymétrique du cristal du domaine soluble de VirB8 (code PDB : 2BHM)

FIGURE 29 – Liste des structures de VirB8-like

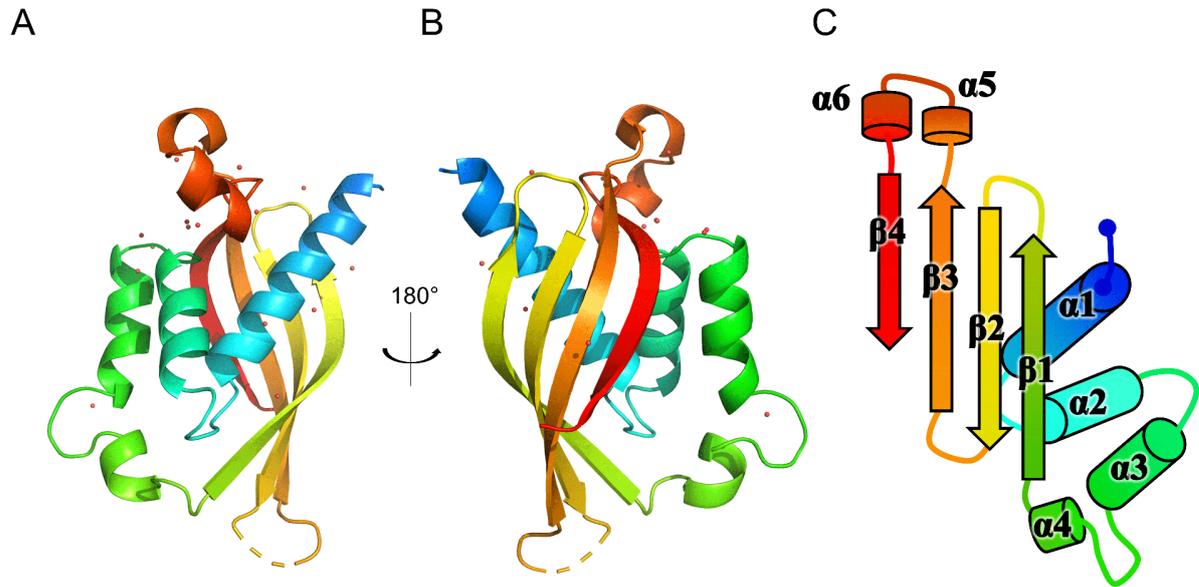
Code PDB	Protéine	Organisme	ID Uniprot	Gram	Oligomérisation en cristal	Methode
4NHF	TrwG	Bartonella grahamii	C6AAT5	-	dimère	RX
4JF8	TrwG	Bartonella birtlesii	D0AAZ5	-	dimère	RX
5I97	TraE	Escherichia coli	Q17U16	-	dimère*	RX
4AKZ	VirB8	Brucella suis	Q7CEG3	-	dimère	RX
2CC3	VirB8	Agrobacterium tumefaciens	P17798	-	dimère	RX
4LSO	VirB8	Bartonella quintana	Q6FYW3	-	monomère	RX
4MEI	VirB8	Bartonella tribocurum	A9IWN6	-	dimère	RX
4KZ1	VirB8	Bartonella grahamii	C6AER9	-	dimère	RX
4O3V	VirB8	Rickettsia typhi	Q68X84	-	dimère*	RX
6IQT	VirB8	Helicobacter pylori	O25265	-	dimère*	RX
3WZ4	DotI	Legionella pneumophila	O54626	-	octamère	RX
3WZ3	TraM	-	Q9R2H2	-	monomère	RX
5CNL	lpg0120	Legionella pneumophila	Q5ZZ91	-	dimère*	RX
3UB1	TcpC	Clostridium tumefaciens	Q1PLH8	+	trimère	RX
4EC6	TraM	Enterococcus faecalis	Q8L1C7	+	trimère	RX
5AIW	TraH	Enterococcus faecalis	AOA140UHJ9	+	-	RMN

Tableau contenant toutes les structures de VirB8 non redondantes, la méthode de résolution ainsi que l'état oligomérique dans le cristal dans le cas de la radiocristallographie. Les dimères avec astérisques présentent un arrangement singulier des monomères, tandis que les dimères n'ayant pas d'astérisque montrent tous le même arrangement. Le code Uniprot permet d'accéder aux informations importantes de la protéine sur le site Uniprot. RX = radiocristallographie, RMN = résonance magnétique nucléaire.

se trouvent 5 monomères, dont 4 sont agencés en homodimères ; le 5ème monomère dimérise avec un équivalent cristallographique (Terradot *et al.*, 2005). Ce même dimère, observé dans d'autres structures de domaine soluble de VirB8, a été supposé avoir une réalité biologique. L'interaction majeure implique l'hélice $\alpha 4$ d'une sous-unité et l'hélice $\alpha 1$ de l'autre. D'autres assemblages ont été découverts pour les VirB8-like de T4SS Gram négatif avec des dimères contenant des monomères orientés différemment, et des structures sous forme d'anneaux hexamériques (Casu *et al.*, 2018) ou encore octamériques. Au sein des T4SS Gram positifs, seulement 3 structures cristallographiques existent et montrent un dimère et deux trimères. Toutefois, il faut prendre en compte que l'architecture des VirB8 de T4SS Gram positif est différente. Dans la Figure 31, page 66 sont représentés quelques modes d'assemblage du domaine soluble de VirB8.

Grâce au travail bioinformatique réalisé par Fercher *et al.*, 2016, les différents domaines structuraux VirB8-like ont été classés selon leur composition en structures secondaires délimitant plusieurs modules. Trois classes permettent de distinguer 3 organisations différentes des domaines VirB8-like. La classe α (alpha) retrouvée à la fois chez les bactéries à Gram négatif et positif présente un domaine transmembranaire (TMD) proche de l'extrémité N-terminale, et un long domaine C-terminal VirB8-like. Uniquement représentées chez les bactéries à Gram positif, il existe également les classes

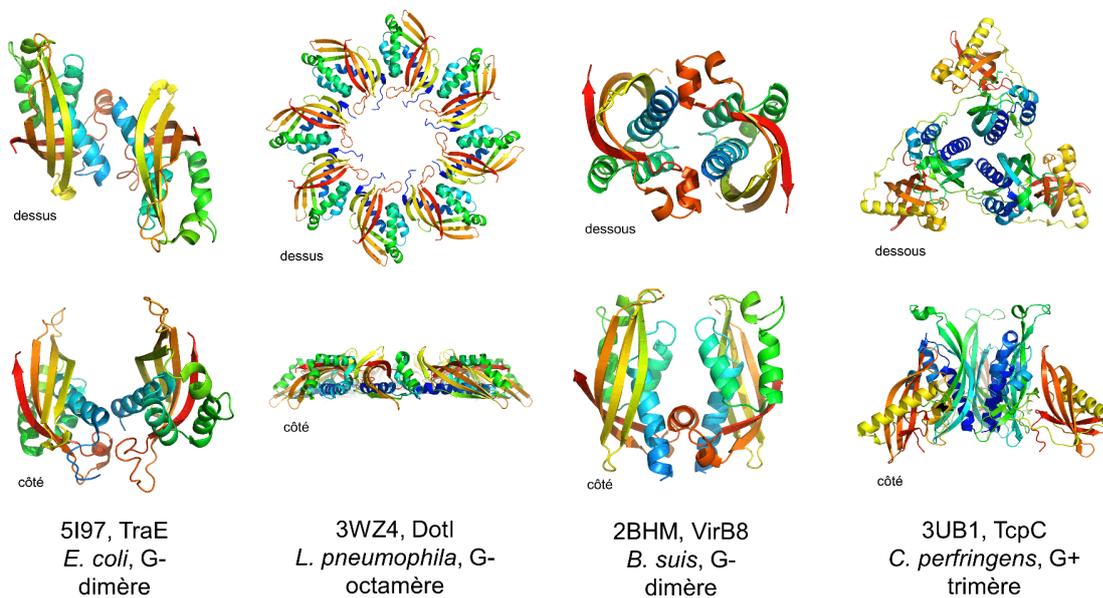
FIGURE 30 – Structure du domaine soluble de la protéine VirB8 chez *Brucella suis*



Différentes représentations de la structure de VirB8 de l'organisme *B. suis* résolue par Terradot *et al.*, 2005. (A) et (B) correspondent à deux vues cartoon de la protéine et se déduisent par rotation de 180° autour de l'axe y. La coloration en arc-en-ciel est appliquée de l'extrémité N-terminale (en bleu) jusqu'à l'extrémité C-terminale (en rouge). Les molécules d'eau y sont représentées par de petites sphères rouges. (C) Représentation schématique des structures secondaires sur un plan 2D en suivant la même coloration que dans (A) et (B). La taille des brins β et des hélices α correspond sensiblement à celles trouvées dans la structure; les boucles sont de taille arbitraire.

β (beta) et γ (gamma). La classe β ressemble à la classe α mais avec deux domaines solubles après le TMD, appelés respectivement domaine central et domaine C-terminal, tandis que dans la classe γ les deux domaines VirB8-like sont séparés par le TMD.

FIGURE 31 – Représentation de quelques oligomères de domaines solubles de VirB8



Quelques exemples d'assemblage des domaines solubles d'homologues de VirB8 observés dans les cristaux. Chaque structure est représentée sous deux orientations à 90° (en haut et en bas). La coloration pour chaque chaîne en arc-en-ciel va de l'extrémité N-terminale (en bleu) vers l'extrémité C-terminale (en rouge). Pour chaque oligomère sont indiqués le code PDB, le nom, l'organisme d'origine ainsi que le type d'oligomère.

V Le système de sécrétion de type IV de l'ICESt3 chez *Streptococcus thermophilus*

V.1 Les streptocoques

Les streptocoques sont des bactéries à Gram positif ubiquitaires, et comprennent de nombreuses espèces commensales et pathogènes. L'une des principales caractéristiques des streptocoques est leur capacité à former des chaînettes dans les milieux liquides, facilement reconnaissables au microscope.

V.1.1 Principales espèces pathogènes

Parmi les espèces pathogènes connues, on retrouve notamment *S. pyogenes*, *S. pneumoniae* et *S. agalactiae*. *S. pyogenes* est surnommée bactérie "mangeuse de chair" car elle peut être à l'origine de graves nécroses cutanées, aussi appelées fasciites nécrosantes. Cette maladie grave progresse rapidement et est due à la libération de toxines par le streptocoque qui tuent les cellules des tissus environnants. De nombreuses autres pathologies peuvent être associées à *S. pyogenes*, et en 2011 il a été estimé que la bactérie était responsable d'environ 600 millions d'infections par an (Lynskey *et al.*, 2011). *S. agalactiae* provoque, entre autres, des méningites et des septicémies chez les nourrissons ainsi que des infections du pis chez la vache qui produit un lait contenant une quantité anormalement élevée de streptocoques. Enfin, *S. pneumoniae* est également surnommé le pneumocoque. Il est responsable de nombreuses co-infections assez graves chez l'Homme, en particulier des infections pulmonaires. A l'inverse de *S. pyogenes*, c'est le pouvoir invasif de *S. pneumoniae* qui le rend dangereux, et non ses toxines.

V.1.2 Résistances aux antibiotiques

La pénicilline a été l'un des premiers antibiotiques utilisés pour soigner les infections de bactéries à Gram positif. En effet, elle appartient aux β -lactames et inhibe les activités transpeptidases, notamment impliquées dans la synthèse du peptidoglycane, élément essentiel de la paroi des bactéries à Gram positif. Malheureusement, comme pour beaucoup d'antibiotiques, des résistances chez les populations bactériennes pathogènes ont commencé à apparaître. C'est dans les années 1940 que les

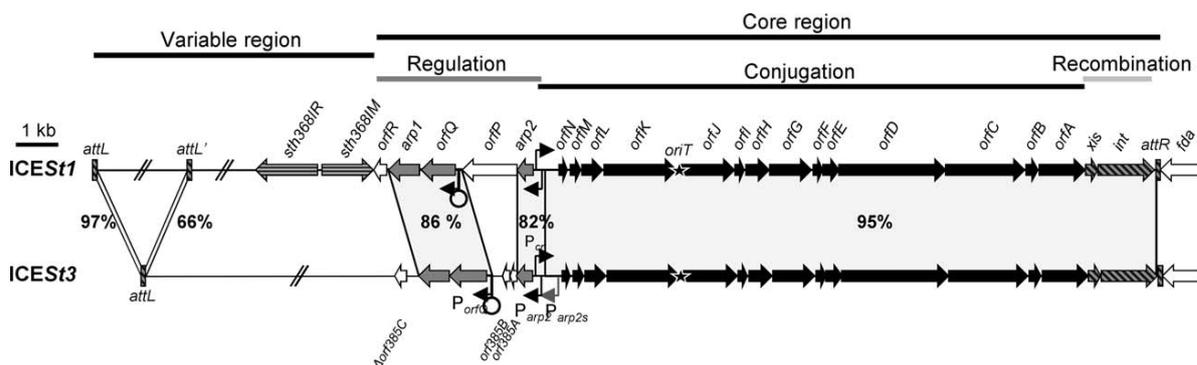
premiers cas de traitements inefficaces ont été décrits (Cattoir, 2016 ; Gillespie, 1998). Dans le cas de *S. pyogenes*, on explique ces échecs en prenant en compte plusieurs facteurs dont 1) la résistance induite par l'internalisation de la bactérie dans certaines cellules (notamment les cellules de la tonsille dans le pharynx) qui n'absorbent pas ou peu les β -lactames ; 2) la sécrétion de β -lactamases (enzymes qui inhibent ou lysent les β -lactames) par les bactéries qui sont présentes naturellement dans les lieux d'infections comme les staphylocoques (Brook, 2013 ; Kaplan *et al.*, 2006 ; Pichichero & Casey, 2007 ; Schaar *et al.*, 2014). L'ensemble des facteurs et mécanismes mis en jeu dans la résistance aux antibiotiques sont détaillés dans la revue écrite par Cattoir, 2016. Bien que les cas de multi-résistances chez *S. pyogenes* et plus généralement chez les streptocoques soient rares, quelques cas ont été documentés et pourraient résulter du transfert de gènes par conjugaison (Lakshmi & Kim, 1989 ; van Asselt *et al.*, 1992)

V.2 Les éléments intégratifs et conjugatifs chez *S. thermophilus*

S. thermophilus est l'une des espèces non pathogènes du genre *Streptococcus*, et est largement utilisée dans l'industrie alimentaire pour la fabrication du yaourt. Le caractère inoffensif de cette bactérie en fait un organisme d'étude idéal, puisqu'elle ne nécessite aucune précaution supplémentaire par rapport aux pathogènes retrouvés chez les streptocoques. L'analyse récente de plus d'une centaine de génomes de streptocoques a permis d'y identifier une super-famille d'éléments intégratifs et conjugatifs qui s'appelle ICE*St3* /ICE*Bs1*/Tn*916* (Ambroset *et al.*, 2016). La famille des Tn*916* est connue pour être impliquée dans le transfert de gènes de résistances à des antibiotiques, en particulier aux tetracyclines et macrolides (Roberts & Mullany, 2011 ; Santoro *et al.*, 2014 ; Schroeder & Stephens, 2016). Chez *S. thermophilus*, on retrouve notamment les ICE ICE*St3* et ICE*St1*, respectivement de 28 et 35 kilobases. Intégrés dans le génome de deux souches distinctes, ils se situent à l'extrémité 3' du gène de la 1,6-biphosphate aldolase (aussi nommé gène *fda*) des souches CNRZ368 (ICE*St1*) et CNRZ 385 (ICE*St3*) (Carraro & Burrus, 2014 ; Dahmane, 2017).

L'ICE*St3* présente une organisation modulaire, avec une région variable aussi appelée module d'adaptation pouvant contenir des gènes avantageux pour la bactérie, comme des gènes de résistances aux antibiotiques, enzymes pour le métabolisme ou encore de protection contre les virus (comme l'ICE*St1*, avec les gènes *sth*). La deuxième région appelée de coeur, est séparée en 3 modules de fonctions différentes. Le premier, appelé module de régulation, permet comme son nom l'indique

FIGURE 32 – Les ICE chez *Streptococcus thermophilus*



tiré de Carraro et Burrus, 2014

Représentation schématique de l'organisation génétique d'ICESt3 et ICESt1. Les différentes régions sont indiquées par des lignes permettant de différencier la région de coeur et la région variable. La région de coeur contient le module de régulation, de recombinaison et de conjugaison. Les différents gènes sont représentés par des flèches colorées selon leur appartenance à un module : prise en charge et traitement du substrat ADN et T4SS (en noir), régulation (en gris), gènes et sites d'excision et d'intégration de l'ICE (en gris avec traits noirs en diagonale), gène de restriction/modification (en gris avec traits noirs horizontaux), inconnus ou n'appartenant pas à l'ICE (en blanc). Les promoteurs sont indiqués par des traits fléchés avec un angle de 90°. L'origine de transfert OriT est représentée par une étoile noire. Les régions présentant des similarités de séquences sont délimitées par des traits noirs reliant les deux ICE, avec le pourcentage au centre.

de limiter ou favoriser le transfert de l'ICE, comprenant également la régulation de l'expression des gènes de l'ICE. Le second module, appelé recombinaison, contient les gènes des enzymes qui réalisent l'excision et l'intégration de l'ICE dans le génome de la bactérie. Ces enzymes spécifiques utilisent notamment les sites *att* de part et d'autre de l'ICE pour l'exciser du génome, mais également pour le dé-circulariser et l'y intégrer. Enfin, le dernier module appelé conjugaison contient l'ensemble des gènes qui sont impliqués dans la prise en charge de l'ADN et son transfert, incluant le système de sécrétion de type IV (voir Figure 32, page 69).

V.3 Les protéines conjugatives de l'ICESt3

Nommés *orfA* à *orfN*, les gènes qui codent pour les protéines conjugatives sont regroupées au sein du même module. Sur ce module se trouve également l'origine de répllication, entre le gène *orfK* et *orfJ*. Plusieurs de ces gènes ont été identifiés et ont des équivalents dans le système Vir.

OrfA, la protéine issue de l'expression du gène *orfA* est une peptidoglycane hydrolase et est assimilée à la protéine VirB1. Par comparaison de séquence avec les autres hydrolases de peptido-

glycane, elle possède un domaine transmembranaire, un domaine Lyt/SLT et un domaine CHAP. A l'heure actuelle, le domaine CHAP semble conservé alors que le domaine Lyt/SLT possède une faible identité de séquence avec d'autres hydrolases de peptidoglycane (Bellanger *et al.*, 2009).

La protéine OrfG issue du gène *orfG* possède une faible identité de séquence avec une protéine VirB8-like appelée TcpC, présente chez l'organisme *Clostridium perfringens*. Résolue par cristallographie (code PDB : 3UB1), la structure de TcpC présente une organisation modulaire avec un segment transmembranaire puis deux domaines VirB8-like, ceux-ci formant un ensemble enregistré dans la base de données Pfam (PF12842) sous le nom de famille "TcpC". Par comparaison de séquence, OrfG semble posséder ce même domaine Pfam, et devrait appartenir à la même famille de protéine. En effet, OrfG et TcpC appartiennent toutes deux à la famille β dans la classification proposée par Fercher *et al.*, 2016.

La protéine OrfJ est une relaxase qui a été caractérisée et qui appartient à la famille des MOB_T (Monteiro, 2016 ; Soler *et al.*, 2019). La famille de cette relaxase est retrouvée chez les Firmicutes qui sont spécialisées dans l'échange des éléments intégratifs et conjugatifs (Soler *et al.*, 2019). OrfJ également appelée RelSt3 possède un homologue structural chez *Geobacillus stearothermophilus* qui est une protéine impliquée dans la réplication en cercle roulant (Carr *et al.*, 2016).

OrfK est l'équivalent de la protéine de couplage VirD4. Cependant, elle ne semble pas apparentée aux VirD4 présente chez les organismes à Gram négatif (Dahmane, 2017).

OrfM et OrfL sont deux petites protéines qui ont une faible identité de séquence avec une protéine appelée HelP (pour "*Helicase Processing*") retrouvée chez l'élément ICEBs1. Cette protéine est un facteur de processivité de l'hélicase PcrA (Thomas *et al.*, 2013), c'est-à-dire qu'elle aide l'hélicase à réaliser la séparation de deux brins d'ADN. Au sein de l'ICE Tn916 existent également deux protéines Orf22 et Orf23 qui sont similaires à HelP (Wright & Grossman, 2016).

OBJECTIFS

VI Objectifs de la thèse

Les systèmes de sécrétion de type IV sont, comme nous avons pu le voir, des assemblages complexes qui permettent aux bactéries d'échanger des éléments mobiles qui facilitent la survie de ces organismes dans des environnements plus ou moins hostiles. Parce qu'ils permettent notamment le transfert de résistances aux antibiotiques, ces systèmes sont donc d'importantes cibles. De nombreuses avancées ont été faites sur les T4SS de bactéries à Gram négatif ce qui nous permet de mieux comprendre comment un élément mobile est transféré vers un autre organisme, que cela soit via la conjugaison ou dans un contexte d'invasion. Les informations sur les systèmes de sécrétion de type IV Gram positif sont très limitées, alors que les bactéries à Gram positif pathogènes sont souvent mentionnées lorsque l'on parle de multirésistances aux antibiotiques, notamment *Staphylococcus* et *Streptococcus*. Peu importe l'organisme, les dégâts causés par les résistances aux antibiotiques progressent plus vite que l'élaboration de nouvelles molécules ou de dérivés antibiotiques.

Dans ce travail de recherche, nous nous sommes donc concentrés sur l'élément mobile ICE*St3* retrouvé chez certaines souches de *Streptococcus thermophilus* qui est un modèle intéressant car totalement inoffensif au contraire de certains autres staphylocoques. Cette étude a été réalisée en proche collaboration avec l'équipe du Pr. N. Leblond-Bourget du laboratoire DynAMic (UMR UL-INRAE 1128). Au sein de l'ICE*St3*, de nombreuses protéines appartenant au T4SS avaient été identifiées en tant que des cibles d'une étude structurale. Cette thèse avait pour but premier d'initier l'étude des relations structure-fonction de ces protéines. Dans ce manuscrit seront décrites les **études structurales des protéines OrfG (par cristallographie RX) et OrfM (par résonance magnétique nucléaire)**, qui sont respectivement la protéine structurale VirB8 et une petite protéine qui permettrait de faciliter l'activité de l'hélicase PcrA à l'aide de son partenaire OrfL; ces résultats seront présentés en deux parties différentes. Des études cristallographiques ont également été entreprises sur la relaxase OrfJ (aussi appelée RelSt3), la protéine OrfL et l'hydrolase de peptidoglycane OrfA pendant la thèse, mais n'ont pas été concluantes. Le domaine soluble de la protéine de couplage OrfK avait été testé en cristallisation avant mon arrivée au laboratoire mais n'avait également rien donné, n'étant pas stable en solution; l'étude structurale en est donc également en attente. Enfin, une troisième partie des résultats sera réservée au **développement d'un outil simple qui permet de représenter efficacement les résultats d'un alignement structural** effectué sur le serveur mTM-align directement dans le logiciel de modélisation PyMOL.

Cet outil a été important dans l'analyse effectuée sur le domaine central de la protéine OrfG qui a donné lieu à un article scientifique. Enfin, les méthodes et informations techniques relatives à ce travail de recherche sont détaillées dans une partie Matériels et Méthodes présentée après le bilan d'ensemble de ces travaux.

RÉSULTATS

VII Étude structurale de la protéine OrfG

Dans ce chapitre sont décrits et discutés les résultats sur la protéine OrfG, l'homologue de VirB8 issue du gène *orfG* de l'*ICESt3*. OrfG est une protéine de 331 acides aminés répartis en 3 domaines incluant un domaine N-terminal transmembranaire et non soluble, ainsi que deux domaines respectivement nommés domaine central (OrfG₆₄₋₂₀₄) et domaine C-terminal (OrfG₂₂₃₋₃₃₁). Reliés par un long linker, ils forment la partie soluble OrfG₆₄₋₃₃₁. Les séquences sont disponibles en annexe S2 et S3 page 206 et 207.

Nos collaborateurs de l'équipe du Pr. Nathalie Leblond-Bourget (laboratoire DynAMic, UMR UL-INRAE 1128) ont produit et purifié plusieurs formes de cette protéine en vue de son étude structurale. En effet, les premiers essais de cristallisation sur la partie OrfG₆₄₋₃₃₁ n'ayant tout d'abord pas porté leurs fruits, les domaines isolés OrfG₆₄₋₂₀₄ et OrfG₂₂₃₋₃₃₁ ont été testés. La structure d'OrfG₆₄₋₂₀₄ a ainsi pu être déterminée, avant que ne le soit finalement celle d'OrfG₆₄₋₃₃₁.

Le présent chapitre résume en premier lieu l'analyse phylogénétique de la protéine OrfG, les données biochimiques sur les différents domaines solubles de la protéine et les résultats structuraux sur le domaine central OrfG₆₄₋₂₀₄, présentés ensuite sous la forme d'un article scientifique (Cappele *et al.*, 2021). Puis nous développons certains points cristallographiques qui n'ont pas pu être détaillés dans la publication, en tenant compte notamment d'une nouvelle structure VirB8-like Gram positif et de son implication sur les résultats présentés dans l'article. En dernier lieu sont proposés les résultats structuraux sur OrfG₆₄₋₃₃₁ et leur apport dans la connaissance des VirB8-like chez les bactéries à Gram positif.

VII.1 Article de l'analyse biochimique d'OrfG et de la structure cristallographique de son domaine central OrfG₆₄₋₂₀₄

Cet article regroupe l'analyse phylogénétique d'OrfG et les analyses biochimiques menées par nos collaborateurs du laboratoire DynAMic sur les différents domaines solubles de la protéine OrfG, c'est-à-dire OrfG₆₄₋₂₀₄ (domaine central) et OrfG₂₂₃₋₃₃₁ (domaine C-terminal) et sur l'ensemble de ces deux domaines formant la partie soluble OrfG₆₄₋₃₃₁. Il présente également la structure du domaine central (OrfG₆₄₋₂₀₄) résolu par cristallographie ainsi que la comparaison de sa structure

aux domaines VirB8-like existants.

Selon la classification réalisée par Fercher *et al.*, 2016, OrfG appartient à la classe β tout comme TcpC dont la structure cristallographique était disponible (code PDB : 3UB1; *Clostridium perfringens*), même si l'identité de séquence entre les domaines centraux de ces deux protéines était très faible. Les autres modèles de VirB8-like de bactéries à Gram positif disponibles appartenaient à d'autres classes et présentaient des identités également très faibles : TraM (code PDB : 4EC6; *Enterococcus faecalis*) de classe γ et TraH (code PDB : 5AIW; *Enterococcus faecalis*) de classe α .

La structure du domaine central OrfG₆₄₋₂₀₄ a été résolue par la méthode SAD. Ce domaine central d'OrfG appartient à la superfamille des domaines NTF2-like et à la famille des domaines VirB8-like. OrfG₆₄₋₂₀₄ partage notamment une grande similarité structurale (1,2 Å de rmsd pour 88 C α superposés) mais une faible identité de séquence (< 10%) avec le domaine central de la protéine TcpC (*Clostridium perfringens*, pCW3). Par la suite, nous avons décidé de comparer les différents domaines VirB8-like et nous proposons une classification qui se base à la fois sur des analyses phylogénétiques, et sur les superpositions structurales fournies par mTM-align (Dong, Pan *et al.*, 2018; Dong, Peng *et al.*, 2018). Afin de représenter ces superpositions, nous avons décidé de développer un script appelé MPSA_Viewer qui permet de faire le pont entre la superposition robuste de mTM-align et le logiciel de modélisation PyMOL (Schrödinger, LLC, 2021). Nous avons constaté une absence globale de conservation de séquence entre les domaines VirB8-like, alors que le repliement est quant à lui bien conservé. On retrouve une similarité de séquence lorsque l'on explore les différents groupes que nous proposons, en particulier au niveau du coeur hydrophobe.

En solution, OrfG₆₄₋₃₃₁, OrfG₆₄₋₂₀₄ et OrfG₂₂₃₋₃₃₁ se comportent majoritairement comme un monomère. Toutefois, une petite proportion de ces protéines en solution ont tendance à former des oligomères lors des chromatographies d'exclusion. Pour confirmer leur présence, des expériences de SEC-MALS et crosslink ont été réalisées démontrant la possibilité qu'ont ces différentes constructions de multimériser en petite proportion, sous la forme de multiples de 6. Il est important de noter que ces différentes constructions protéiques ne possèdent pas le domaine N-terminal transmembranaire normalement présent chez OrfG. La petite proportion d'oligomères pourrait être expliquée par cette absence. Les résultats obtenus semblent coïncider avec les résultats obtenus sur les différentes VirB8-like Gram positive comme TcpC (*Clostridium perfringens*, pCW3), TraM et TraH (*Enterococcus faecalis*, pIP501) qui ont tendance à multimériser. Toutefois, la nature de cette oligomérisation ainsi

que son intérêt dans le cadre biologique d'un système de sécrétion reste une question en suspens.



Structural and Biochemical Analysis of OrfG: The VirB8-like Component of the Conjugative Type IV Secretion System of ICESt3 From *Streptococcus thermophilus*

Julien Cappelé¹, Abbas Mohamad Ali^{2†}, Nathalie Leblond-Bourget², Sandrine Mathiot¹, Tiphaine Dhalleine³, Sophie Payot², Martin Savko⁴, Claude Didierjean¹, Frédérique Favier^{1*} and Badreddine Douzi^{2*}

OPEN ACCESS

Edited by:

Anastassios C. Papageorgiou,
University of Turku, Finland

Reviewed by:

Julien Bergeron,
King's College London,
United Kingdom
Zhao Wang,
Baylor College of Medicine,
United States

*Correspondence:

Frédérique Favier
frederique.favier@univ-lorraine.fr
Badreddine Douzi
badreddine.douzi@inrae.fr

†Present address:

Abbas Mohamad Ali,
CEACadarache/BIAM/St Paul Les
Durance, France

Specialty section:

This article was submitted to
Structural Biology,
a section of the journal
Frontiers in Molecular Biosciences

Received: 16 December 2020

Accepted: 01 February 2021

Published: 18 March 2021

Citation:

Cappelé J, Mohamad Ali A, Leblond-Bourget N, Mathiot S, Dhalleine T, Payot S, Savko M, Didierjean C, Favier Fé and Douzi B (2021) Structural and Biochemical Analysis of OrfG: The VirB8-like Component of the Conjugative Type IV Secretion System of ICESt3 From *Streptococcus thermophilus*. *Front. Mol. Biosci.* 8:642606. doi: 10.3389/fmolb.2021.642606

¹Université de Lorraine, CNRS, CRM2, Nancy, France, ²Université de Lorraine, INRAE, DynAMic, Nancy, France, ³Université de Lorraine, INRAE, IAM, Nancy, France, ⁴Synchrotron SOLEIL, L'Orme des Merisiers, Saint Aubin BP 48, Gif-sur-Yvette, France

Conjugative transfer is a major threat to global health since it contributes to the spread of antibiotic resistance genes and virulence factors among commensal and pathogenic bacteria. To allow their transfer, mobile genetic elements including Integrative and Conjugative Elements (ICEs) use a specialized conjugative apparatus related to Type IV secretion systems (Conj-T4SS). Therefore, Conj-T4SSs are excellent targets for strategies that aim to limit the spread of antibiotic resistance. In this study, we combined structural, biochemical and biophysical approaches to study OrfG, a protein that belongs to Conj-T4SS of ICESt3 from *Streptococcus thermophilus*. Structural analysis of OrfG by X-ray crystallography revealed that OrfG central domain is similar to VirB8-like proteins but displays a different quaternary structure in the crystal. To understand, at a structural level, the common and the diverse features between VirB8-like proteins from both Gram-negative and -positive bacteria, we used an *in silico* structural alignment method that allowed us to identify different structural classes of VirB8-like proteins. Biochemical and biophysical characterizations of purified OrfG soluble domain and its central and C-terminal subdomains indicated that they are mainly monomeric in solution but able to form an unprecedented 6-mer oligomers. Our study provides new insights into the structural analysis of VirB8-like proteins and discusses the interplay between tertiary and quaternary structures of these proteins as an essential component of the conjugative transfer.

Keywords: conjugation, Gram - positive bacteria, type IV secretion system, Integrative and Conjugative Element (ICE), VirB8-like proteins

INTRODUCTION

Conjugation constitutes one of the major mechanisms of horizontal gene transfer (Frost et al., 2005). Besides their crucial role in prokaryotic evolution and adaptation, conjugative elements including plasmids and Integrative and Conjugative Element (ICEs) constitute the most important vectors of antibiotic resistance spreading among bacteria. Conjugative transfer is ensured by dynamic multi-

subunit cell-envelope spanning nanomachines called conjugative type IV secretion systems (Conj-T4SSs) encoded by conjugative elements.

Based on the well-studied archetypal VirB/D4 Conj-T4SS system of the Ti conjugative plasmid from the Gram-negative bacterium *Agrobacterium tumefaciens*, it is commonly assumed that Vir subunits can be grouped into three classes (Cascales and Christie, 2003; Alvarez-Martinez and Christie, 2009; Fronzes et al., 2009; Grohmann et al., 2018). One encloses the energetic components VirB4, VirB11, and VirD4. VirB4 and VirB11 belong to traffic ATPases and are involved in the functioning of T4SS apparatus and in pili biogenesis whereas the coupling protein VirD4 (TC4P) is involved in DNA substrate recognition. The second class corresponds to the inner membrane platform (IMP), which is composed of the membrane proteins VirB3, VirB6, and VirB8. The IMP serves as a docking station for DNA substrate and constitutes the inner membrane channel. The third class forms the outer membrane core complex. It is composed of VirB7, VirB9, and VirB10 proteins. The outer membrane core complex forms a conducting channel that traverses the periplasmic space and the outer membrane barrier, establishing a continuity for the inner membrane channel. Other proteins are involved in the conjugative process. These proteins include the transglycosylase VirB1 that is important for the establishment of the T4SS apparatus within the peptidoglycan layer in the periplasmic space and of the extracellular pili involved in DNA transfer or in host cell attachment. Conjugative pili are predominantly composed by the VirB2 pilin and are tipped by the VirB5 adhesin (Grohmann et al., 2018; Gonzalez-Rivera et al., 2019). DNA transfer process through Conj-T4SS requires DNA recruitment and delivery to the T4SS apparatus. Initial steps of DNA recruitment are conducted by the relaxosome complex constituted by the relaxase and accessory factors that recognise and bind specifically to the origin of transfer *oriT* sequence located in the conjugative element. The nick of the *oriT* sequence is catalysed by the relaxase that binds covalently to the nascent 5' end of the DNA strand to be transferred (T-strand). Straight after, the relaxase-T-strand nucleoprotein docks to T4CP, which consequently unfolds the relaxase and ensures the transfer of the nucleoprotein to the Conj-T4SS channel. Once transferred to the recipient cell, relaxase catalyses the recircularization of the T-strand followed by the second strand synthesis (Ilangoan et al., 2017).

Conj-T4SSs in Gram-positive bacteria are more puzzling compared to those found in Gram-negative bacteria. The lack of the outer membrane components together with the incapacity to detect associated extracellular pili raise questions about the assembly mode of a conducting channel that crosses the cell-wall and ensures a cell-to-cell contact. Several lines of evidence support the idea that Conj-T4SSs from Gram-positive bacteria emerged from diderm Conj-T4SSs and form a “minimized” system spanning the cytoplasmic membrane and the murein cell wall (Bhatty et al., 2013). The best-characterized systems are those encoded by conjugative plasmids pCW3 from *Clostridium perfringens*, and pIP501 and pCF10 from *Enterococcus faecalis* (Dunny, 2007; Goessweiner-Mohr et al.,

2013a; Wisniewski and Rood, 2017). Based on the structural and genetic information gathered from previous studies, it is assumed that the Conj-T4SSs from Gram-positive bacteria are essentially composed of subunits related to components recovered in Gram-negative T4SS. These ones comprise 1) VirB4 and VirD4-like energetic components, 2) VirB3-, VirB6- and VirB8-like membrane subunits that assemble in the cytoplasmic membrane to form the membrane platform, 3) VirB1-like transglycosylase and 4) surface adhesins crucial in cell-to-cell contact or cells aggregation (Bantwal et al., 2012; Arends et al., 2013; Laverde Gomez et al., 2014; Kohler et al., 2017; Dahmane et al., 2018; Kohler et al., 2018). Finally, Conj-T4SSs from Gram-positive bacteria also comprise a set of accessory proteins with undefined functions (Wisniewski et al., 2015). However, the limited structural and mechanistic information on these components restrict our understanding of their mode of assembly and function.

Despite the similarities in the early steps of DNA processing, fundamental differences exist in terms of channel architecture between Gram-negative and -positive Conj-T4SSs. Indeed, two components, a VirB1-like transglycosylase and a VirB8-like protein were predicted to extend their C-terminal domains along the exterior face of the cytoplasmic membrane and form a protruding channel for DNA translocation through the cell wall (Bhatty et al., 2013). Unfortunately, such structural organization was not fully compatible with the crystal structures of the VirB8-like subunits TcpC and TraM from pCW3 and pIP501, respectively (Porter et al., 2012; Goessweiner-Mohr et al., 2013a). Nevertheless, is it not excluded that channel establishment depends on VirB8-like interaction with other cytoplasmic membrane subunits including the VirB1-like transglycosylase.

Two types of mobile genetic elements are autonomous for their conjugative transfer: plasmids and ICEs. With the expansion of available genomes, it emerges that ICEs are the most widespread mobile genetic elements that use conjugation to propagate autonomously intra or interspecies (Guglielmini et al., 2011; Santoro et al., 2014). The 28 kb ICE*St3* from *Streptococcus thermophilus* is the prototype of the ICE*St3*/ICE*Bs1*/Tn*916* superfamily, the most widespread superfamily among Streptococci (Ambroset et al., 2015). A large number of derivatives from ICE*St3*/ICE*Bs1*/Tn*916* contribute to the spread of various antimicrobial resistance genes in human or zoonotic pathogens including tetracycline and macrolide resistance genes (Roberts and Mullany, 2011; Schroeder and Stephens, 2016). ICE*St3* was shown to be self-transferable intra- or inter-specifically including towards the human pathogen *Streptococcus pyogenes* and the opportunistic pathogen *E. faecalis* (Bellanger et al., 2009). The ICE*St3* conjugation module encodes 14 proteins (OrfA to OrfN) (Burrus et al., 2002). Several of these proteins are homologous to the IMP components of the Gram-negative conjugative T4SS, whereas others (OrfB, E, F, H, L, M, N) are not. The presence of these latter Orfs with undefined structural homologues and functional information suggests distinct structural and assembly features of the conjugative apparatus of ICE*St3* and may reflect a specific functioning mechanism.

This study belongs to a multiapproach project that aims to understand the architecture and the assembly mode of Conj-T4SS from ICEs found in Gram-positive bacteria. Here, by using the ICES_{t3} as biological model, we focus on the structure-function analysis of several proteins of the Conj-T4SS. Among them, OrfG shares a low sequence identity (22%) with TcpC, the VirB8-like component of pCW3. A structural analysis revealed that the OrfG central domain adopts a NTF2-like fold and likely presents more structural similarities with VirB8-like proteins TcpC and TraM from Gram-positive bacteria than Gram-negative VirB8 proteins. We used structural alignment tools to identify different structural classes among VirB8-like proteins. Our *in-silico* analysis revealed that Gram-positive VirB8-like proteins form a distinct class from Gram-negative ones, supporting a different evolutionary pathway. As described for soluble domains of Gram-positive VirB8-like proteins, OrfG soluble domain was found to be mainly monomeric in solution. However, a small fraction of OrfG was able to self-assemble into 6-mer oligomers. The significance of this quaternary structure is discussed in view of the role of this protein in Gram-positive Conj-T4SSs.

MATERIALS AND METHODS

Bacterial Strains, Media and Chemicals

Escherichia coli DH5 α was used for cloning procedures, BL21 (DE3) was used for protein expression and production. The genotype of each strain is listed in **Supplementary Table S1**. Lysogenic Broth (LB) was used for bacterial culture. Kanamycin (50 μ g/ml) was added to the culture medium.

Plasmid Construction

Plasmids used in this study are listed in **Supplementary Table S1**. Polymerase Chain Reactions (PCR) were performed using a BioRad T100 Thermocycler thermal Cycler using Phusion High-fidelity DNA Polymerase (Thermo Scientific). *Streptococcus thermophilus* LMG18311 harboring ICES_{t3} genomic DNA was used as template for PCR amplifications. Custom oligonucleotides (Eurogentec) used for cloning procedures were listed in **Supplementary Table S1**. For protein expression and production, the DNA sequence encoding OrfG_{64–331}, OrfG_{64–204}, OrfG_{64–215} and OrfG_{223–331} were cloned into the pET28-TRX leading to pET28-TRX-OrfG_{64–204}, pET28-TRX-OrfG_{64–215}, and pET28-TRX-OrfG_{223–331}, respectively. For *orfG*_{64–331} and *orfG*_{223–331}, restriction-ligation procedure was used for genes cloning. pET28-TRX-OrfG_{64–204} and pET28-TRX-OrfG_{64–215}, plasmid derivatives were obtained by insertion of Stop codon at position 205 and 216 respectively using pET28-TRX-OrfG_{64–331} as template.

Production and Purification of OrfG_{64–331}, OrfG_{64–204}, OrfG_{64–215}, and OrfG_{223–331}

E. coli BL21 (DE3) cells carrying the pET28-TRX derivatives were grown in LB supplemented with Kanamycin (50 μ g/ml), at 37°C to an OD₆₀₀ ~ 0.5. Expression of the constructions was then induced by addition of IPTG (0.5 mM) and cultures were pursued

for 18 h at 25°C. Cells were harvested, resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 1 mM EDTA, 50 μ g/ml lysozyme, 0.1 mM phenylmethylsulfonyl fluoride (PMSF) and submitted to five cycles of sonication. After addition of 20 μ g/ml DNase and 20 mM MgCl₂, the soluble fraction was obtained by centrifugation for 40 min at 16,000 \times g. Recombinant proteins were purified by ion metal affinity chromatography using a 5-ml Nickel HisTrap™ HP Column on an ÄKTA prime apparatus (GE healthcare) pre-equilibrated in 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 mM Imidazole (buffer A). After several washes in buffer A, 6 \times His tagged proteins were eluted in buffer A supplemented with 250 mM Imidazole, cleaved with the 1 mg/ml TEV protease and dialyzed against 50 mM Tris-HCl, 300 mM NaCl pH 8.0 for 18 h at 4°C and loaded onto a HisTrap™ HP pre-equilibrated in buffer A, which selectively retains the TEV protease, the uncleaved 6 \times His-tagged fusion proteins and contaminants. The native proteins were collected in the flow-through, concentrated on Centricon (Millipore; 10 kDa cutoff), and passed through a Sephadex 200 26/60 column pre-equilibrated with 50 mM Tris-HCl pH 8.0, 100 mM NaCl or 20 mM HEPES pH 8.0, 100 mM NaCl.

Crystallization and Structure Determination

Crystallization was conducted using the sitting drop, vapor diffusion method at 293 K. OrfG_{64–204} crystals were obtained by mixing 0.3 μ l of protein (35 mg/ml, 100 mM Tris-HCl pH 8.0) with 0.3 μ l of the reservoir solution containing 2 M ammonium sulfate, 100 mM Tris-HCl, pH 8.5. All drops were prepared by the crystallization robot Oryx8 (Douglas Instruments, Hungerford, United Kingdom) in 96 well 2-lens MRC plates (SWISSCI AG, Neuheim, Switzerland). Crystals reached their maximal size after approximately 1 week. To ensure phasing, osmium (IV) hexachloride di-potassium salt was added to crystal-containing drops (10 mM final concentration) and left for 24 h. Crystals were then mounted on nylon loops and quickly back-soaked into their reservoir solution +20% glycerol and immediately flash frozen in liquid nitrogen to be stored until data collection.

All X-ray diffraction experiments were carried out at Synchrotron SOLEIL (St-Aubin, France) on Proxima-2A beamline (**Table 1**). Measurements were performed at 100 K. A first crystal without osmium was used to collect a native dataset at $\lambda = 0.9801$ Å, up to a resolution of 1.59 Å. At an advanced stage of the refinement the Fourier difference map remained noisy. The associated R_{free} factor was 21.2% (42.8% in the highest resolution shell). Thus, we decided to limit the resolution to 1.75 Å (data processed with XDS (Kabsch, 2010) and scaled with Aimless (Evans and Murshudov, 2013)). From an osmium containing crystal, eight datasets (400°span, 4,000 images each) were collected at osmium peak absorption $\lambda = 1.1397$ Å, at different kappa angles. Merging all of them was mandatory to get a final data set at 2.0 Å with sufficient quality for successful phasing (data processed with XDS and XSCALE). Preliminary phases were determined using the SAD Pipeline implemented in PHASER CCP4i (McCoy et al., 2007) and the partial results were enhanced by ARP/wARP (Langer et al., 2008). The resulting model was further improved with the native dataset at 1.75 Å, using PHENIX (Liebschner et al., 2019) and COOT (Emsley et al.,

TABLE 1 | Statistics of data collection for the native crystal and for the osmium derivative crystal, and refinement of OrfG_{64–204} model.

	Native	Osmium derivative
Data collection		
Space group	<i>P</i> 4 ₁ 2 ₁ 2	<i>P</i> 4 ₁ 2 ₁ 2
Beamline	Proxima-2A, SOLEIL	Proxima-2A, SOLEIL
Unit cell parameters (Å, °)	50.2 50.2 120.9 90. 90. 90.	50.4 50.4 121.3 90. 90. 90.
Wavelength (Å)	0.9801	1.1398
Resolution (Å) ^a	46.33–1.75 (1.78–1.75)	46.62–2.00 (2.05–2.00)
Observations ^a	411,273 (20,684)	1,637,839 (40,984)
Unique reflections	16,372 (1,116)	20,272 (1,494) ^b
R _{merge} ^a	0.027 (0.499)	0.107 (1.028)
R _{meas} ^a	0.028 (0.514)	0.107 (1.048)
CC _{1/2}	1.0 (0.975)	1.0 (0.845)
Mean I/σ ^a	67.5 (7.3)	55.6 (2.8)
Completeness (%) ^a	100. (100.)	100. (99.4)
Anomalous completeness (%) ^a	—	100. (98.6)
Multiplicity ^a	25.1 (23.8)	80.1 (27.4)
Mean anomalous difference ΔF/σ ^a	—	1.21 (0.74)
Figure of merit (I) ^c	—	0.13/0.37
First model ^d	—	4 chains, 63 aa, 31 sequenced
Model refinement		
Resolution (Å) ^a	46.33–1.75 (1.80–1.75)	
No. of protein residues	101 (3 chains)	
No. of non-hydrogen protein atoms	842	
No. of waters	81	
R _{cryst} ^a	0.200 (0.294)	
R _{free} ^{ae}	0.208 (0.427)	
RMSD bonds (Å)	0.014	
RMSD angles (°)	1.635	
Ramachadran plot ^f		
Favored (%)	100	
Outliers (%)	0.0	
Rotamer outliers (%)	1.1	
Clashscore ^f	5.42 (94)	
MolProbity score (PR) ^f	1.32 (98)	
Average B (Å ²)	35.4	
Protein	33.0	
Water	42.1	
Coord. error (Å)	0.10	
PDB code	6zgn ^g	

^aValues for the outer shell of data are given in parentheses.

^bNumber of reflections in the data set where the Friedel mates are not merged.

^cBefore/after solvent flattening; the same values were obtained for both hands.

^dAs built by *Buccaneer* (CCP4i); no model was obtained for the other hand.

^eFive percent test set.

^fFrom MolProbity. The percentile ranks (PR) for Clashscore and MolProbity score are given in parentheses.

^gPDB validation report is available in supplementary materials.

2010) (**Table 1**). The asymmetric unit contained a single OrfG_{64–204} monomer. Several regions were unobserved in the electron density when contoured at 1.2 σ, especially at the N-terminal end. These regions were assumed to be disordered and were not included in the model. The final model contained residues Asp92-Val133 + Val139-Ile168 + Glu176-Ala204, i.e., 70% of the 143 amino acids of the OrfG_{64–204} subdomain. It is worth noting an important blob of density in the pocket delineated by Tyr106-Phe107, Arg140-Ser143, Tyr164, Glu166 and Phe202 (**Supplementary Figure S1**). This blob could be assigned to none of the compounds used for protein crystallization. It probably resulted from a significant affinity for a ligand present in the protein expression cells or in the successive media used to prepare the protein sample. The final OrfG_{64–204} structure was deposited in the Protein Data Bank with

ID 6zgn. All three-dimensional structure representations were obtained using PYMOL (Schrödinger, LLC).

In Silico Analysis, Sequence Alignment, Structure Comparison and Interface Analysis

Protein sequences were retrieved from NCBI and compared using clustalW from MEGA7 (Kumar et al., 2016). Phylogenetic trees were built with MEGA7 by using the Maximum Likelihood method based on the Tamura-Nei model (Tamura and Nei, 1993) without outgroup.

TMD prediction was conducted using CCTOP (<http://cctop.enzim.ttk.mta.hu/>). PSIPRED (Buchan and Jones, 2019) and HHPRED (Zimmermann et al., 2018) were used for OrfG secondary structures and subdomains structures predictions, respectively. COILS (Lupas et al., 1991), MULTICOIL (Wolf et al.,

1997) and DEEPCOIL (Gabler et al., 2020) were used to examine the presence of a coiled-coil domain on OrfG.

The first step of structure comparison was conducted with both DALI (Holm and Sander, 1995) and PDBefold (Krissinel and Henrick, 2004) to fetch a complete list of VirB8 structures (26 PDB entries). From them, a non-redundant set of 16 structures was conserved (2cc3, 3ub1, 3wz3, 3wz4, 4akz, 4ec6, 4jff8, 4kz1, 4lso, 4mei, 4nhf, 4o3v, 5aiw, 5cnl, 5i97, 6iqt) and added to the new OrfG_{64–204}. Then the mTM-align server (Dong et al., 2018a) based on the robust TM-Align method (Dong et al., 2018b) performed the structural alignment and produced the structure-based “phylogenetic tree.” PyMOL and an associated homemade python script were used to emphasize visually structure differences and similarities (available upon request). Briefly, a color gradient was applied to each position depending on the rmsd value calculated on Ca carbons of structurally aligned amino acids at this position. Conserved or similar residues were also automatically highlighted. Furthermore, the superposition of all VirB8 monomers produced by the mTM-align server was used to compare oligomers interfaces and to visually identify interesting features.

Biophysical Characterization of Native and Truncated OrfG_{64–331} Proteins

To determine the molecular weight of the purified proteins, size exclusion chromatography (SEC) coupled to MALS analysis was performed on MiniDAWN TREOS II coupled to Superdex 200 10/300 increase column (GE Healthcare) mounted on FPLC system (AKTA purifier). We used 20 mM Tris-HCl pH 8.0, 100 mM NaCl buffer at 0.5 ml/min flow.

In Vitro Chemical Crosslinking

Crosslinking experiments were performed for OrfG_{64–331}, OrfG_{64–204} and OrfG_{223–331} obtained after SEC purification and present in 20 mM HEPES pH 8.0, 100 mM NaCl buffer. The used protocol is described as follow: 20 μ M of each protein were mixed or not with increasing concentration of paraformaldehyde (PFA) from 0.05 to 5% in incubated for 1 h at 37°C. The crosslinking reaction was stopped by the adding of 1 M Tris-HCl at a final concentration of 80 mM and the mixture was incubated for 10 min at RT. To examine crosslinked products, 15 μ l of SDS-PAGE loading buffer (4 \times) were added and 10 μ l from each reaction was analyzed by 12% SDS-PAGE for OrfG_{64–331} and 15% SDS-PAGE for OrfG_{64–204} and OrfG_{223–331}.

RESULTS

OrfG Belongs to the TcpC Family

Previous studies showed that the integrative and conjugative element ICES_{t3} from *S. thermophilus* is an autonomous element able to self-transfer by conjugation intra and interspecifically (Bellanger et al., 2009). Bioinformatic analyses of ICES_{t3} predicted the presence of a conjugation module that harbors 14 genes. Among them, *orfJ* was recently showed to encode the relaxase that belongs to the MobT superfamily (Soler et al., 2019). In this study, we focused on the structural and

biochemical studies of a putative transfer protein OrfG. Blastp analysis of OrfG indicated that this protein is composed of a TcpC domain (pfam12642) from position 98–316 (e-value $7.4e^{-39}$) (Figure 1A). This domain is homologous to that of several TcpC conjugal transfer proteins encoded by diverse ICEs or conjugative plasmids, including the VirB8-like subunit TcpC from pCW3. The archetypal TcpC from pCW3 is a membrane protein that was shown to be essential for the conjugation of pCW3 from *C. perfringens* (Porter et al., 2012). TcpC consists of an N-terminal cytoplasmic domain followed by a transmembrane domain (TMD) which encompasses residues 57–79 and a soluble domain extended in the cell-wall. Despite a poor sequence identity/similarity between OrfG and TcpC (23% identity/38% similarity) in their TcpC domain, *in silico* analysis using PSIPRED, HHPRED and CCTOP revealed that OrfG shares the same domain organization as observed for TcpC. Thus, OrfG is composed of an N-terminal tail followed by a TMD (residues 37–59) and a soluble domain that we called OrfG_{64–331} (Figure 1A).

Our phylogenetic analysis of OrfG and TcpC orthologs indicated that OrfG is more closely related to the VirB8-like protein (conB) from ICEBs1 (Leonetti et al., 2015) than to the other TcpC proteins (including TcpC from pCW3) (Porter et al., 2012). This is consistent with the existence of at least two sub-families of TcpC proteins (Figure 1B). Given the similarities in domain organization between OrfG and TcpC, their distant evolutionary relationship motivated us to focus on the structural and biochemical characterization of OrfG.

OrfG Central Domain (OrfG_{64–204}) Displays a VirB8-like Structure

Owing to the lack of structural data on Gram-positive Conj-T4SS components especially those involved in ICE conjugation, we envisaged structure determination of OrfG by X-ray diffraction. For this purpose, we purified to homogeneity a large amount of the OrfG soluble domain (OrfG_{64–331}: from Ser64 to Asp331). Despite an easy growing of crystals in multiple and different conditions, their poor X-ray diffraction power prevented the acquisition of data of sufficient quality. To overcome these technical issues, and based on the secondary structure prediction of OrfG as well as the sequence alignment between OrfG and TcpC, we chose to express and purify to homogeneity the separate central (OrfG_{64–204}: from Ser64 to Ala204) and C-terminal (OrfG_{223–331}: from Ala223 to Asp331) subdomains (Figure 1A). Crystallization trials conducted on both forms resulted in suitable tetragonal crystals for OrfG_{64–204} while no positive result was obtained for OrfG_{223–331}. Experimental phases were determined by using a single-wavelength anomalous dispersion method (Porter et al., 2012) on crystals soaked with an osmium derivative. A native crystal allowed final model refinement to 1.75 Å resolution ($R_{\text{free}} = 20.8\%$). Detailed data and refinement statistics are given in Table 1. Note that the electron density of OrfG central domain displayed a blob (Supplementary Figure S1) that remained unassigned despite attempts to fit various compounds in it. This density was located between β_4 and the C-terminal end of α_1 .

The polypeptide chain adopts an alpha + beta fold, composed of two α -helices named α_1 and α_2 , followed by a β -sheet of four

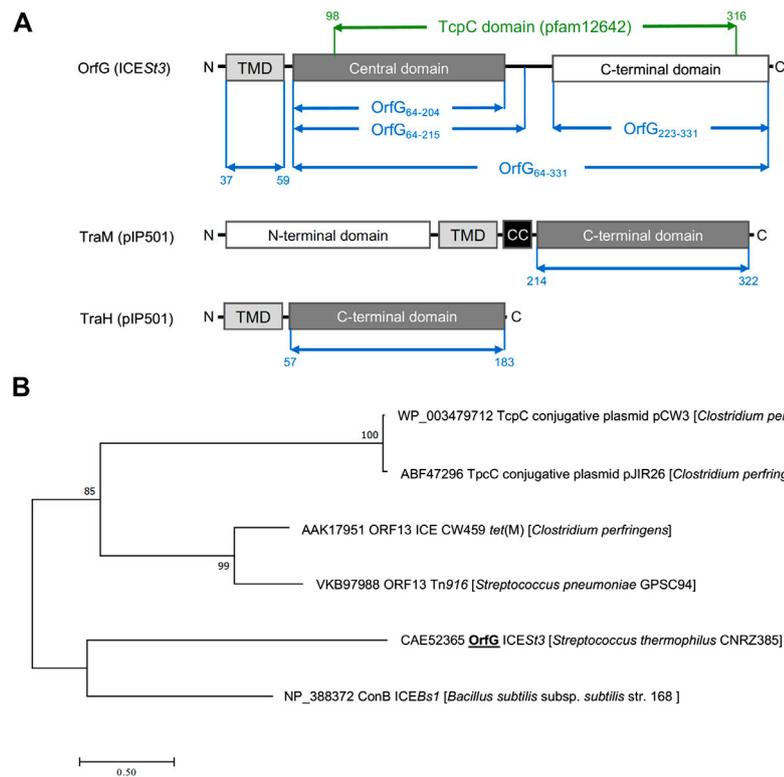


FIGURE 1 | OrfG belongs to the TcpC superfamily **(A)** Schematic representation of OrfG, TraM and TraH subdomains organization. Boundaries of OrfG subdomains and TraM and TraH NTF2-like domains are stated. TMD for transmembrane domain; CC for coiled-coil domain **(B)** Phylogenetic tree of representative members from the TcpC superfamily including OrfG. This tree was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Guindon et al., 2010). The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured as the number of substitutions per site. All positions containing gaps and missing data were eliminated. There was a total of 284 positions in the final dataset. Evolutionary analyses were conducted using MEGA7 (Tamura and Nei, 1993).

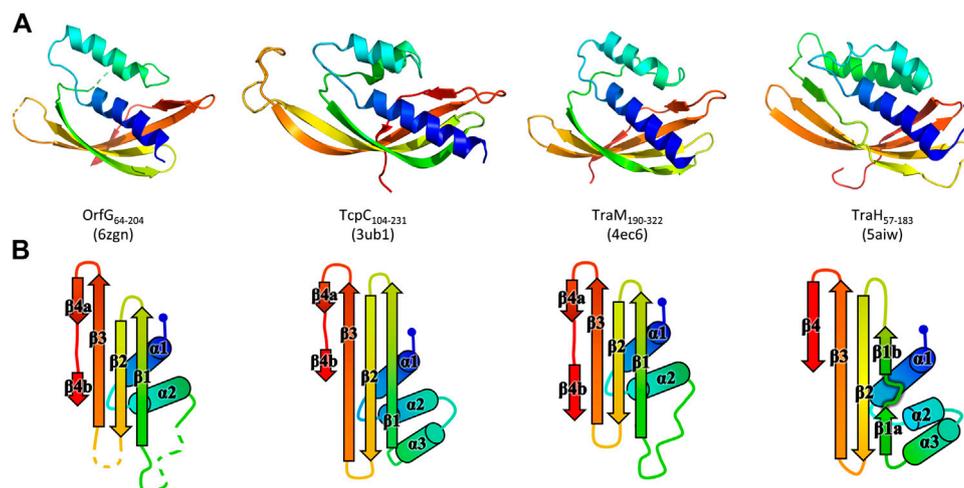


FIGURE 2 | Structures of OrfG₆₄₋₂₀₄, TcpC₁₀₄₋₂₃₁, TraM₁₉₀₋₃₂₂, and TraH₅₇₋₁₈₃. From left to right respectively, the four monomers (PDB id 6zgn, 4ec6, 3ub1, and 5aiw) are displayed in the same orientation and represented as rainbow-colored cartoons, with the N-terminal end of the polypeptide chain in blue and the C-terminal end in red **(A)**. Their topology is schematized with the same rainbow-color code and the name of the secondary structures **(B)**.

strands (β_1 – β_4), the last of which has a marked disruption into two subparts β_{4a} and β_{4b} (Figure 2). Cross-comparisons performed by DALI (Holm and Sander, 1995) and PDBfold (Krissinel and Henrick, 2004) classify the spark OrfG central domain in the NTF2-like superfamily. Its folding pattern mainly consists in a highly curved antiparallel beta sheet that wraps around a central alpha helix surrounded by shorter additional ones. As expected from sequence analysis, the closest VirB8-like structural neighbours of OrfG_{64–204} is the TcpC central domain (TcpC_{104–231}, i.e., the first domain of TcpC_{99–359}, PDB entry 3ub1 (Porter et al., 2012)), with an rmsd of 1.77 Å for 92 aligned amino acids. This search also highlighted the structure of the C-terminal domain of TraM from pIP501 (TraM_{190–322}, PDB entry 4ec6 (Goessweiner-Mohr et al., 2013b)), rmsd of 1.90 Å for 96 aligned amino acids) as a close structural homolog (Figure 2). The folds of TcpC_{104–231}, TraM_{190–322} and OrfG_{64–204} differ by the length of their secondary structures and connecting loops. Furthermore, TcpC_{104–231} has an additional helix α_3 prior to the β -sheet. Both TcpC and TraM were described as Gram-positive equivalents of VirB8 from VirB/D4 T4SS from *A. tumefaciens*. Despite their low sequence identity and their different domain organization, it was proposed that proteins from Gram-positive Conj-T4SSs that adopt a NTF2-like fold form the “VirB8-like” family (Ambroset et al., 2015). This family also includes the soluble domain of the conjugative component TraH from pIP501 (TraH_{57–183}, PDB entry 5aiw (Fercher et al., 2016)) (Figure 2). Since the structure of OrfG_{64–204} belongs to the NTF2-like superfamily as well, we propose that OrfG from ICES τ 3 is a new member of the VirB8-like family found in Gram-positive Conj-T4SS.

Structural Comparison of Available VirB8-like Structures

Unexpectedly, known structures of VirB8-like proteins show subdomains that harbor an overall fold similar to the unique NTF2-like domain of Gram-negative VirB8s. This observation sparked our interest in a structural comparison of these proteins. A previous study using sequence analysis, secondary structure prediction and domain composition of a large set of sequences distinguished three classes of proteins (Goessweiner-Mohr et al., 2013b). Here we conducted a distinct analysis that focused on the available three-dimensional structures of the NTF2-like domain of these proteins, with the aim to point out the shared features and the specific characteristics of both Gram-positive and Gram-negative proteins. It was based on the superposition of their atomic coordinates, independently of their amino acid sequence. The results emphasized a surprising diversity despite the simplicity and small size of the NTF2-like fold.

The set of Gram-negative VirB8 structures encompasses 13 independent entries in the protein database (Protein Data Bank). That of known Gram-positive VirB8-like structures is poorer: the crystallographic analysis of OrfG_{64–204} brings their number to four. The NTF2-like pattern is globally conserved in the 17 structures. On average, their fold consists in three antiparallel alpha helices followed by an antiparallel four-stranded sheet, for which we propose the nomenclature $\alpha_1\alpha_2\alpha_3\beta_1\beta_2\beta_3\beta_4$. However,

the length of the secondary structures (especially beta strands) and their relative orientation, the curvature of the beta sheet, plus possible decorations inserted in the minimal fold bring enough diversity to make the structural superimposition delicate. We found the mTM-align algorithm as the most suited one to deal with these differences. The global superimposition of the 17 structures and the corresponding structural alignment provided by mTM-align were analyzed using an in-house developed script to allow their visualization in Pymol (see *Materials and Methods*) (Figure 3). The structure-based tree obtained using mTM-align indicated that Gram-positive VirB8-like structures clearly separate from the group formed by the Gram-negative ones (Figure 4). The presence of clusters with members that share high sequence similarities and others where modest to low sequence similarities are observed confirms that the structural alignment was performed independently from primary sequences (Supplementary Table S2). At least three classes appear in the Gram-negative VirB8s (named I⁻, II⁻ and III⁻ for the convenience of their description) and one in the Gram-positive VirB8-like structures (I⁺), while the three other proteins did not belong to identified classes (Figure 4). In the Gram-negative group, the three classes differ by the way their helix α_3 is disrupted in two parts α_{3a} and α_{3b} , separated by a protruding loop of four residues with two distinct conformations in classes I⁻ and II⁻, or by a strong kink in class III⁻ (Figure 4). This disruption of α_3 is not a distinctive trait of the Gram-negative proteins since CagV from *Helicobacter pylori* (PDB entry 6igt (Wu et al., 2019)), not assigned to a class, has a straight helix. All members of classes I⁻ and II⁻ have a small additional helix α_4 between β_3 and β_4 , which is involved in protein dimerization, while members of class III⁻ have a loop instead. Furthermore, class III⁻ and CagV display a bulge that separates β_1 in two parts β_{1a} and β_{1b} (Figure 4).

Strikingly, the four Gram-positive VirB8-like proteins have no specific structural feature that unambiguously distinguishes them from the Gram-negative ones. However, the overall structure superposition separated them in the structural tree, which seems to be explained by the small differences disseminated along their polypeptide chains (Figure 4). Indeed, OrfG_{64–204} has no α_3 (replaced by an unobservable region) and a strand β_4 that clearly splits in two short portions β_{4a} and β_{4b} separated by a bulge that ends with Pro200 (Figure 2). By comparison, TraM_{190–322} (PDB entry 4ec6) has a shorter α_2 , also misses α_3 and has a slightly longer loop between β_1 and β_2 . In contrast, TcpC_{104–231} (PDB entry 3ub1) possesses α_3 and has much longer strands β_1 , β_2 , and β_3 , together with a longer loop between them. TraH_{57–183} is the most distant structure, as it is the only one with a continuous strand β_4 with no bulge insertion, while on the contrary β_1 possesses two parts β_{1a} and β_{1b} with a bulge in between (PDB entry 5aiw) (Figure 2), more similar to some Gram-negative proteins (class III⁻ and CagV). Furthermore, TraH has a pronounced helix α_3 with four turns and a long loop towards β_1 .

The most distinguishable difference between Gram-negative and Gram-positive VirB8 proteins lies in the length of the connection between β_3 and β_4 . It has 10–17 amino acids in Gram-negative VirB8s (including α_4 in classes I⁻ and II⁻), and

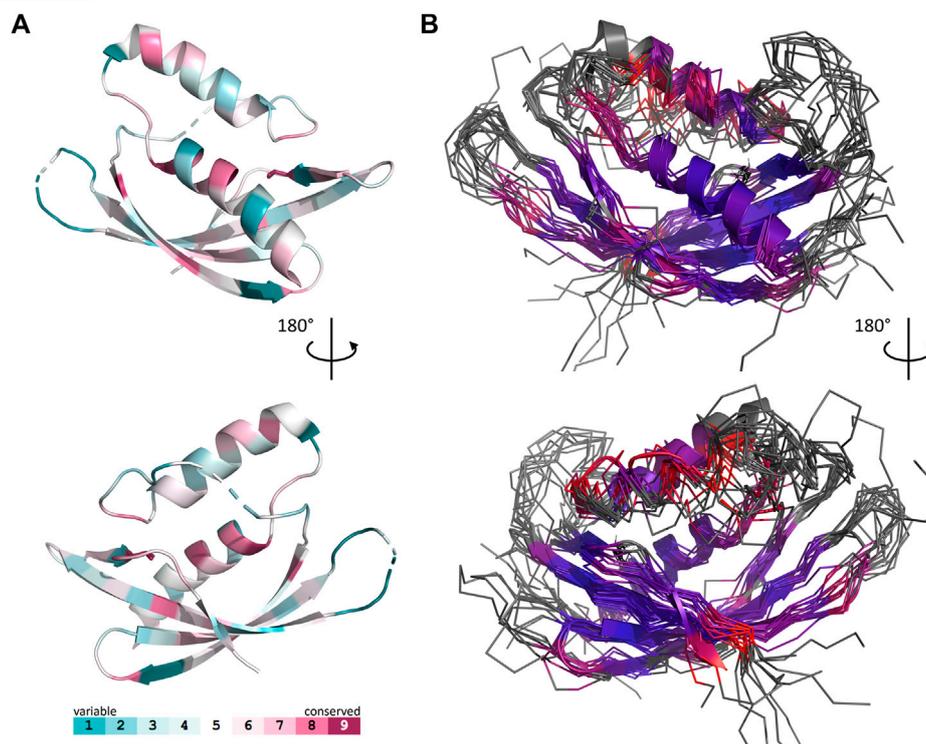


FIGURE 3 | Comparison of 17 representative three-dimensional structures of VirB8 and VirB8-like proteins. The structure of OrfG₆₄₋₂₀₄ (PDB id 6zgn) is represented as a cartoon (A), with its amino acids colored according to the ConSurf server (Ashkenazy et al., 2016) which analyzed the similarity calculated in the sequence alignment deduced from the structure superposition produced by mTM-Align (Dong et al., 2018a) for a set of seventeen selected proteins. The amino acid shown with sticks (Val191 in OrfG) and localised in β_4 facing α_1 is the only one whose side chain shares a similar nature in all structures, while there is no conserved residue. The 17 structures (PDB ids 2cc3, 3ub1 (limited to Tcpc₁₀₄₋₂₃₁), 3wz3, 3wz4, 4akz, 4ec6, 4jf8, 4kz1, 4lso, 4mei, 4nhf, 4o3v, 5aiw, 5cnl, 5i97, 6iqt, 6zgn) are represented (B) as superposed wires (except OrfG as a cartoon) using a color-code based on the quality of the superposition of these structures. For each position, a gradient from violet to red is used depending on the low to high values of the root mean square calculated on the distance between all the possible pairs of the 17 superposed C α carbons at this position (grey means that mTM-Align found at least one structure impossible to align). This representation highlights the resemblances and differences in the NTF2-like folds of these proteins.

only four to seven in Gram-positive ones. However, deleting all these atoms in the seventeen structure files prior to calculation of the structural tree resulted in the same classification, with only minor modifications. Thus, it seems that Gram-positive and -negative VirB8 and VirB8-like proteins separate according to details evenly distributed across the structures instead of localized marked traits that would characterize each group.

The script that we developed to analyse the structural alignment of the seventeen structures easily identified the conserved or similar residues within the different sets of proteins that we considered. Interestingly, we found no conserved residue common to all 17 structures. In the subset of the 13 Gram-negative VirB8 structures, once again, no amino acid is conserved. On the contrary, up to 30 residues are conserved within each Gram-negative class, among which many form hydrophobic contacts where the β -sheet faces α_1 (Supplementary Figure S2). By comparison, in the four Gram-positive VirB8-like structures, only six residues share similar natures and constitute a hydrophobic patch within the core of the protein near the C-terminal side of α_1 . In addition, it is worthy to notice the presence of a conserved glutamate residue (Glu154 in OrfG) at the end of β_1 , with its side chain towards the surface.

Its conservation might be the result of happenstance due to the low number of analyzed structures and no apparent involvement in any observed assemblies. As for Gram-negative VirB8s, only a subset of Gram-positive VirB8-like proteins display more conserved residues (designated as Class I⁺) (Supplementary Figure S2). The absence of any signature was already noticed in the NTF2-like superfamily (Eberhardt et al., 2013). However, it is surprising that, despite probably having the same function, this observation also applies to the restricted subset of NTF2-like superfamily proteins that constitute the VirB8-like proteins.

Crystal Packing of VirB8-like Proteins Raises the Question of their Multimerization

In the tetragonal crystal, the OrfG₆₄₋₂₀₄ monomer interacts with symmetry-related partners, among which one forms a questionable pair with a nice arrangement of the facing alpha helices (Figure 5A). However, the interaction involves a low surface area of 1,600 Å², or only 12% of the solvent accessible surface, as determined by PISA (Krissinel and Henrick, 2007), which rejects the arrangement as an effective quaternary structure. As well, we struggled to find an explanation to the resulting location of both N-termini in this

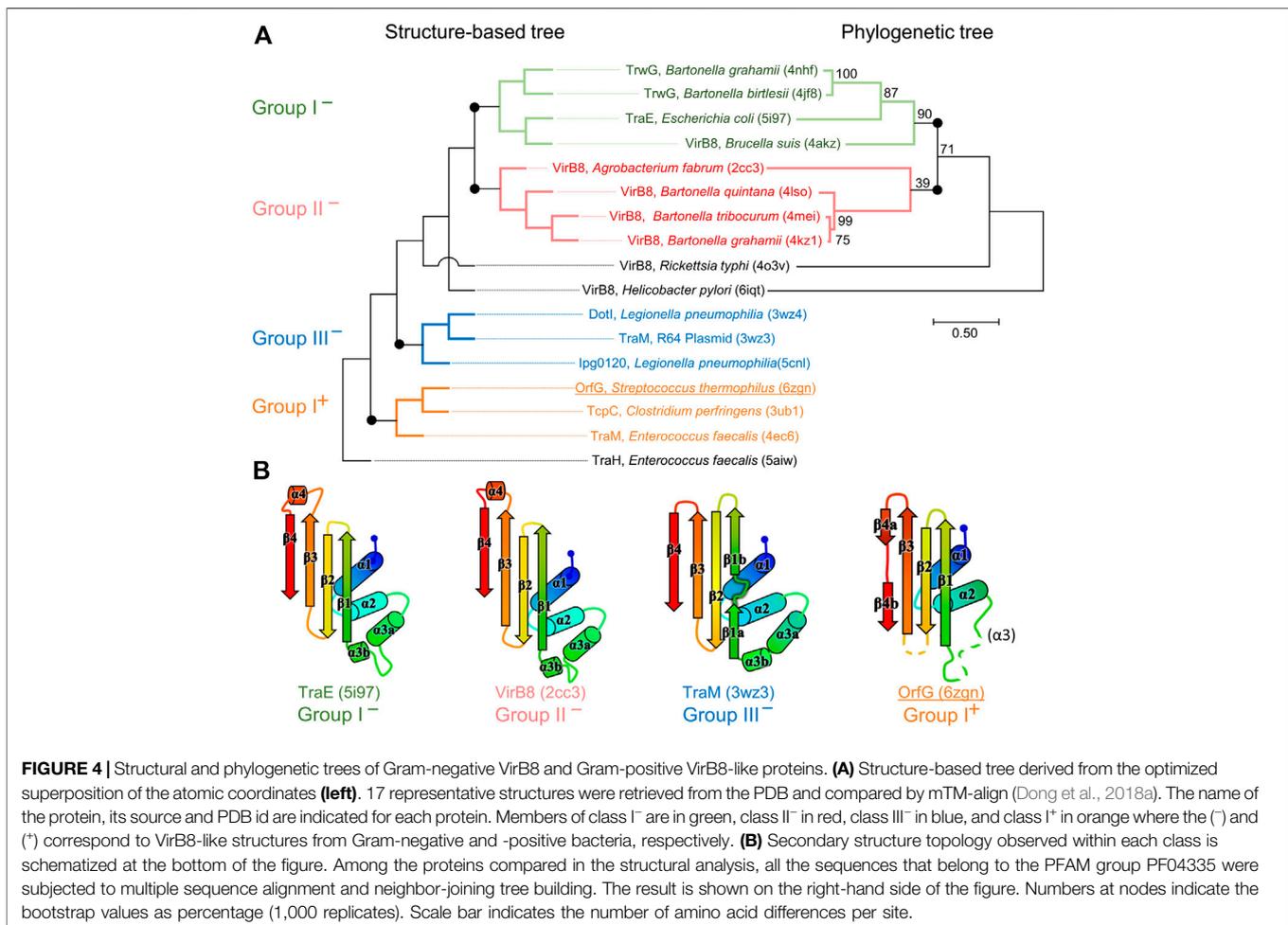
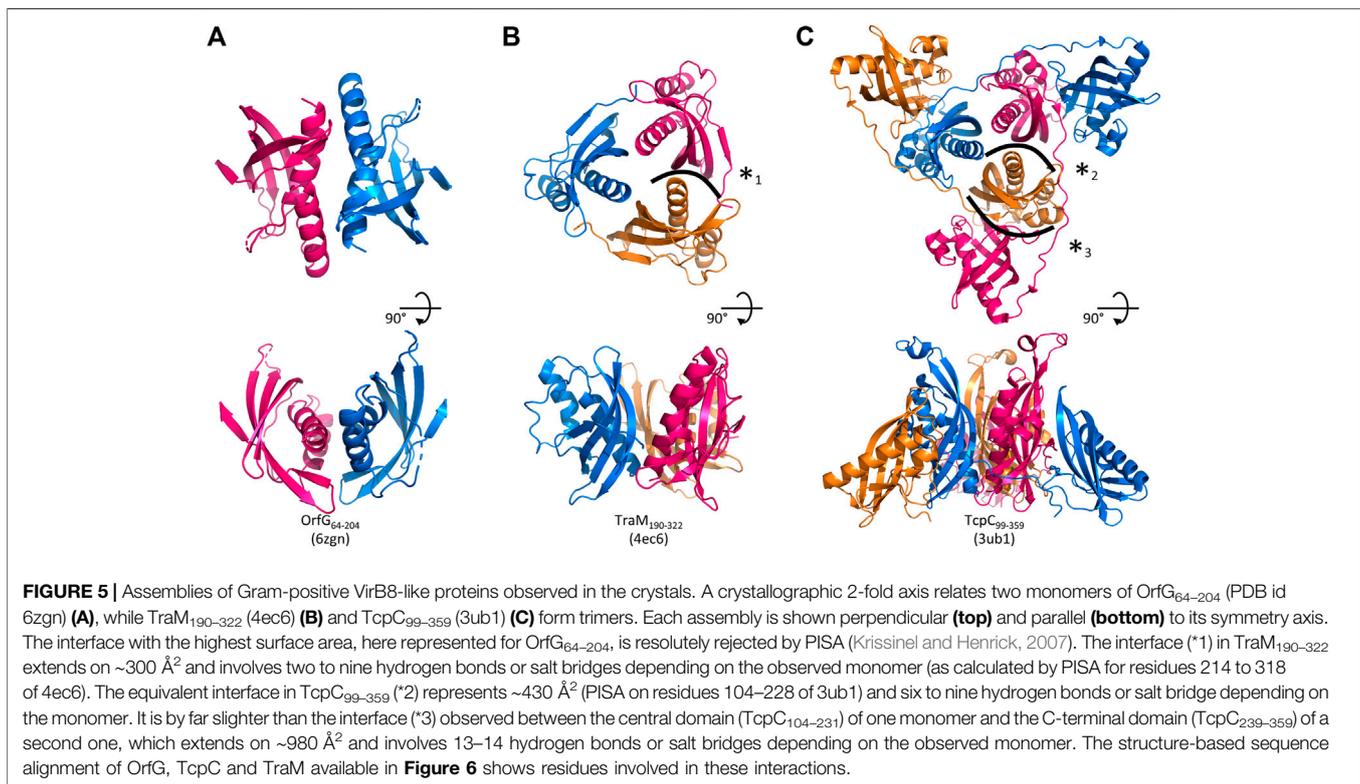


FIGURE 4 | Structural and phylogenetic trees of Gram-negative VirB8 and Gram-positive VirB8-like proteins. **(A)** Structure-based tree derived from the optimized superposition of the atomic coordinates (left). 17 representative structures were retrieved from the PDB and compared by mTM-align (Dong et al., 2018a). The name of the protein, its source and PDB id are indicated for each protein. Members of class I⁻ are in green, class II⁻ in red, class III⁻ in blue, and class I⁺ in orange where the (-) and (+) correspond to VirB8-like structures from Gram-negative and -positive bacteria, respectively. **(B)** Secondary structure topology observed within each class is schematized at the bottom of the figure. Among the proteins compared in the structural analysis, all the sequences that belong to the PFAM group PF04335 were subjected to multiple sequence alignment and neighbor-joining tree building. The result is shown on the right-hand side of the figure. Numbers at nodes indicate the bootstrap values as percentage (1,000 replicates). Scale bar indicates the number of amino acid differences per site.

association. Indeed, it would place the transmembrane domain of each OrfG monomer at opposite sides of the dimer, but this argument should be used with caution since 27 N-terminal residues are missing in the model with respect to the crystallized protein. Altogether, the pair of OrfG₆₄₋₂₀₄ monomers observed in the crystal is not a convincing proof of the existence of a dimer with a biological relevance.

Analysis of structural similarity positioned OrfG₆₄₋₂₀₄ in the same class as TraM₁₉₀₋₃₂₂ and TcpC₁₀₄₋₂₃₁, while TraH₅₇₋₁₈₃ is more distant (Figure 4). The proximity of these three proteins raises once again the question of a possible multimerization of OrfG. Indeed, both TraM₁₉₀₋₃₂₂ and TcpC₉₉₋₃₅₉ form trimers in their crystal assembly (Figures 5B,C). Yet, TraM₁₉₀₋₃₂₂ corresponds to a single domain, like the fragment of OrfG that has been crystallized (OrfG₆₄₋₂₀₄), while TcpC₉₉₋₃₅₉ has two subdomains equivalent to the central and C-terminal domains of OrfG (OrfG₆₄₋₃₃₁, no structure available). Do the protein interfaces share homology that allows prediction of an equivalent assembly of OrfG? Alternatively, does the latter has specificities that would result in a different oligomeric form? Comparison of the trimers formed by TraM₁₉₀₋₃₂₂ and TcpC₉₉₋₃₅₉ emphasizes a surprising similarity in the positioning of their common domain (Figures 5B,C).

TraM₁₉₀₋₃₂₂ and the central domain of TcpC (TcpC₁₀₄₋₂₃₁) bring their N-termini near the 3-fold axis that defines the trimer. Interactions that build this assembly are mainly located at the α_1 N-terminal moiety of one partner and near the β_1 C-terminal end of the facing partner, in a circular pattern of interactions. Detailed analyses are provided in the papers that describe these crystallographic structures (Porter et al., 2012, Goessweiner-Mohr et al., 2013b). Briefly, TraM₁₉₀₋₃₂₂ hides 14% of the calculated solvent-accessible surface area of one isolated monomer in this contact, which involves three hydrogen bonds and several van der Waals contacts (Figure 5B). Consistently with this observation, the weakness of this interaction gives little credibility to the existence of a trimer in solution. However, its existence in the crystal makes sense with respect to the prediction of a coiled-coil motif formed by some thirty residues that precede the NTF2-like domain, which could stabilize a trimeric assembly of TraM (Goessweiner-Mohr et al., 2013b). In the case of TcpC₉₉₋₃₅₉, the equivalent interface between the central domains TcpC₁₀₄₋₂₃₁ is a bit wider and based on more hydrogen bonds (a total of eight) (Porter et al., 2012) (Figure 5C). The striking observation that arises from a precise comparison of these interfaces is their total lack of commonality, except for the region they concern. Two

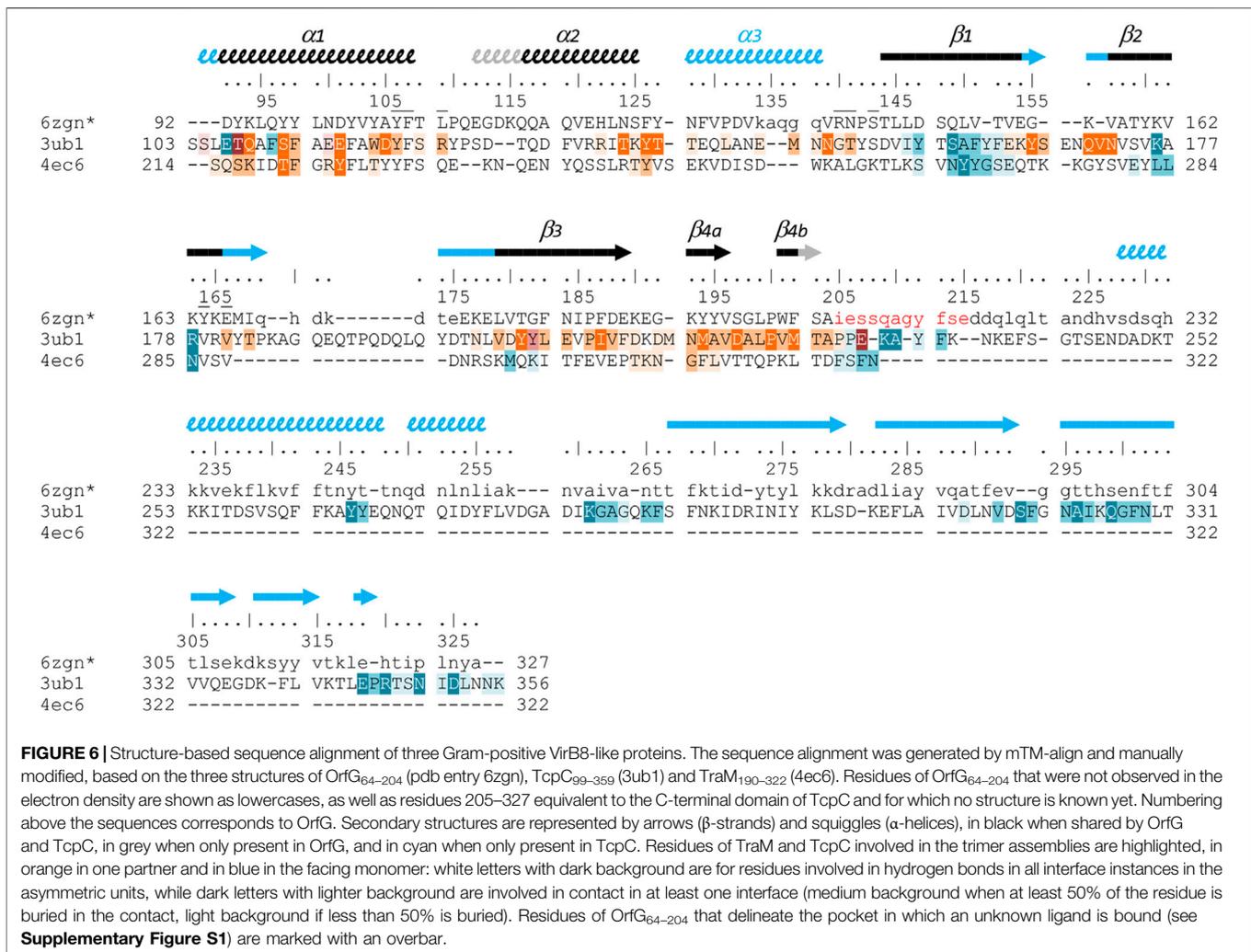


particular examples illustrate this conclusion. First, in TraM_{190–322}, the side chain of Lys217 at the beginning of α_1 interacts with Tyr269 in β_1 of the partner, via van der Waals contact with its aromatic ring and the addition of a hydrogen bond with its main chain on four of the six instances observed in the asymmetric unit. At the equivalent positions of TcpC_{99–359} (**Figure 6**), Gln108 and especially Phe161 have different orientations so that instead Gln108 interacts with Glu106 at the beginning of α_1 in the neighbouring monomer. The second example concerns Tyr224 (α_1) in TraM_{190–322}, the side chain of which forms a hydrogen bond with Asn285 (β_2) of the associated monomer. In TcpC_{99–359}, the residue Glu115 at the equivalent position in α_1 prefers an interaction with Ser159 in β_1 of the facing monomer. As a conclusion, the driving force that governs the trimer formation does not originate from conserved residues at this interface. However, the observation of similar multimeric forms with similar arrangements is suggestive of their probable relevance. Structure and sequence comparisons with OrfG_{64–204} revealed no conservation of interacting pairs of residues (neither with TcpC_{99–359} nor with TraM_{190–322}) but they also revealed no element that would prevent formation of the same interaction (**Figure 6**).

No trimer is observed in the crystal form of OrfG_{64–204}. Deletion of the C-terminal domain is a plausible explanation but it is not appropriate for TraM_{190–322}. A closer analysis of the different structures shows that TraM_{190–322} has a C-terminal extension of five residues after β_{4b} , which follows the same direction as the first residues of the linker between the two domains of TcpC_{99–359}. Its penultimate residue Phe321

positions its side chain in a hydrophobic environment formed by several aromatic residues in α_1 , α_2 and β_{4a} of the facing monomer. In TcpC_{99–359}, the linker residue Lys233 also contributes to stabilization of the trimer, but via a hydrogen bond with the polypeptide main chain following α_2 . In OrfG, the equivalent Gln209 could play the same role (**Figure 6**). In order to evaluate the contribution of this extension, we solved the structure of OrfG_{64–215}. However, no change was induced, neither in the crystallization conditions nor in monomer packing in the crystal (data not shown) demonstrating that the linker was not sufficient to promote OrfG_{64–215} multimerization.

The crystal structure of TcpC soluble domain (TcpC_{99–359}) shows that its C-terminal domain (TcpC_{239–359}) also belongs to the NTF2-like superfamily (Porter et al., 2012). It has a peripheral position in the trimeric assembly, far from the N-terminal domain (TcpC_{104–231}) to which it is connected through an extended coil that allows to the three monomers to intertwine (**Figure 5C**). This organization mode results in a large interface (1000 Å²) between the central domain of one monomer and the C-terminal domain of the partner. Thirteen hydrogen bonds are formed, together with hydrophobic interactions. It alone accounts for much more than the interactions formed between the central domains and described above. However, this convincing arrangement, which buries 24% of the accessible surface area of each monomer, was not observed in solution (Porter et al., 2012). The authors proposed that the deleted N-terminal residues, in particular a transmembrane domain, could explain this discrepancy. The similarity between OrfG and TcpC in their domain organization and in the structure of

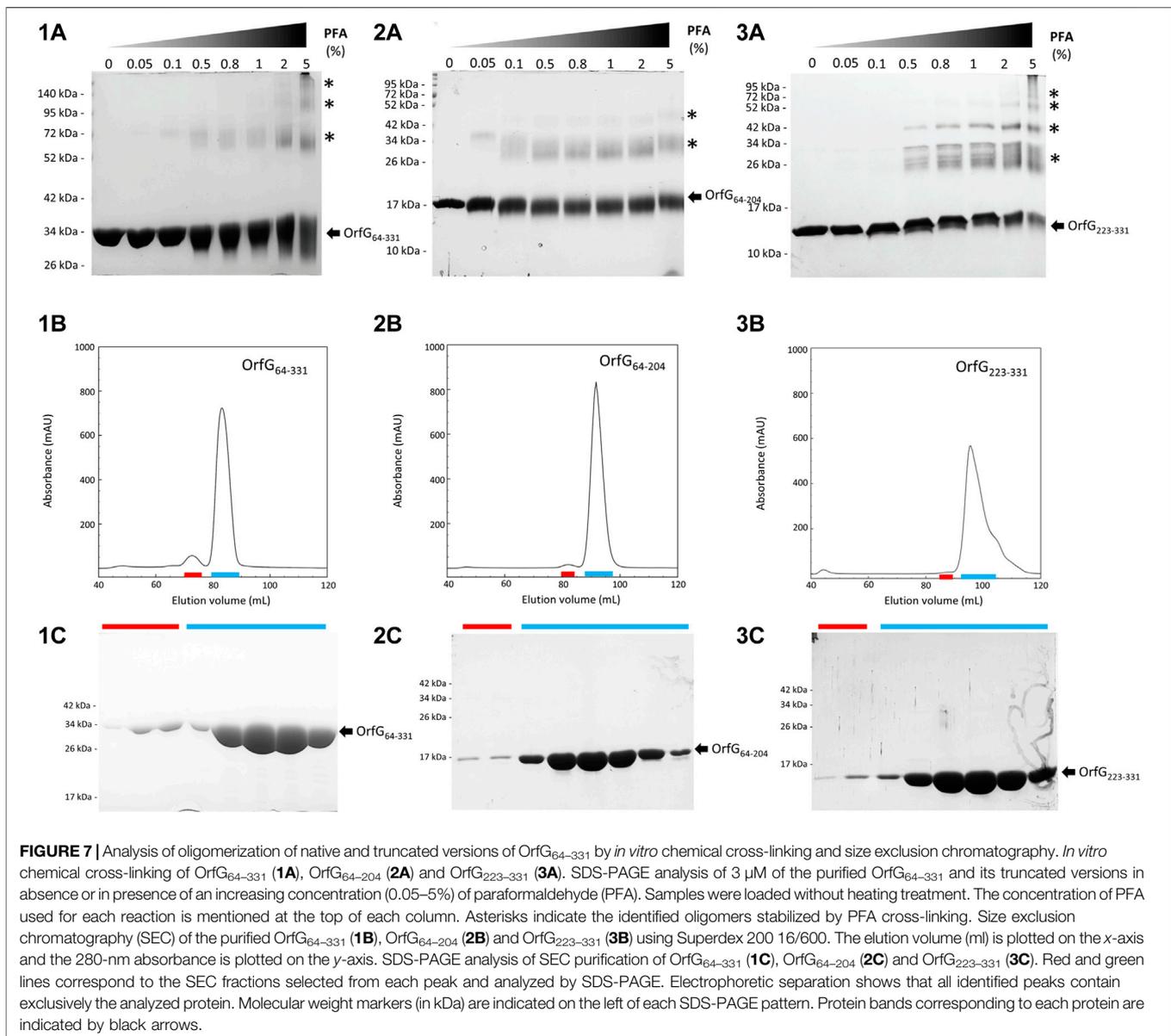


their central domain, despite a low sequence identity, suggests a similar overall structure of the whole protein, and a similar oligomeric state. The missing linker and C-terminal domain probably hinder trimerization in the crystallization conditions used.

In Vitro Self-Oligomerization of OrfG₆₄₋₃₃₁ and Its Subdomains

The closest structural homologs to OrfG₆₄₋₂₀₄ form trimers in their crystal arrangement (**Figures 5B,C**). Thus, it was proposed that trimers constitute the functional biological assembly of both proteins (Porter et al., 2012, Goessweiner-Mohr et al., 2013b). These observations strongly call for the assessment of the oligomeric state of OrfG₆₄₋₂₀₄ in solution. Protein multimerization can be monitored by chemical cross-linking. We performed *in vitro* chemical cross-linking experiments with a constant amount of the protein in presence of an increasing concentration of the cross-linking agent paraformaldehyde (PFA). As observed in **Figure 7A**, in the presence of a low

amount of PFA (0.05–1%), OrfG₆₄₋₃₃₁ predominantly forms dimers. High molecular-mass oligomers of OrfG₆₄₋₃₃₁ are visible at higher PFA concentrations (2 and 5%) demonstrating the predisposition of OrfG₆₄₋₃₃₁ to oligomerize in solution. The ability of OrfG₆₄₋₃₃₁ to self-assemble was confirmed by the analysis of the size exclusion chromatography (SEC) profile where the 31 kDa OrfG₆₄₋₃₃₁ was eluted into two defined peaks. SDS-PAGE analysis showed that both peaks are exclusively composed by OrfG₆₄₋₃₃₁ protein (**Figure 7B**). These results suggest that OrfG₆₄₋₃₃₁ adopts two oligomeric states in solution. To further examine the apparent existence of these two oligomeric states of OrfG₆₄₋₃₃₁, we analyzed purified OrfG₆₄₋₃₃₁ by SEC coupled to multiple angle light scattering (SEC-MALS) (**Figure 8A**). The predominant form (97.3% of the total mass fraction) was eluted at 29 min and corresponds to the monomeric form of OrfG₆₄₋₃₃₁ with a calculated MW of 32 kDa (theoretical MW = 31 kDa). The less present form (2.7% of the total mass fraction) was eluted earlier at 25 min with a calculated MW of 172 ± 11 kDa which can be assigned as 6-mer of OrfG₆₄₋₃₃₁ (186 kDa theoretical MW). SEC-



MALS data confirm the ability of OrfG₆₄₋₃₃₁ to form ordered homo-multimers in solution. Intriguingly, such behavior has not been reported for other characterized members of the Gram-positive VirB8-like proteins. In fact, AUC and SEC analysis performed for TcpC and TraH truncated for their TMD, respectively, revealed that both proteins were strictly monomeric in solution (Porter et al., 2012; Fercher et al., 2016). TraM₁₉₀₋₃₂₂ (TraM lacking its N-terminal domain and the TMD) was able to form oligomers only in the presence of cross-linking agents whereas DLS and SAXS analysis indicated a monomeric form of TraM in solution (Goessweiner-Mohr et al., 2013).

The knowledge that OrfG can assemble into ordered oligomers in solution led us to question whether OrfG central domain or

OrfG C-terminal domain or both would promote/contribute to OrfG₆₄₋₃₃₁ oligomerization. For that purpose, we investigate separately their capacity to self-assemble by using chemical cross-linking. Regarding OrfG₆₄₋₂₀₄, the *in vitro* chemical cross-linking using PFA gave mostly a dimeric form when using at least 0.05% of PFA (Figure 7C). At higher PFA concentrations, an additional band with a higher mass was observed that might correspond to a trimeric form. On the contrary to OrfG₆₄₋₂₀₄, the chemical cross-linking of OrfG₂₂₃₋₃₃₁ easily distinguished when analyzed by SDS-PAGE (Figure 7E). These results suggest that both subdomains were independently able to form oligomers in solution. Interestingly, during SEC analysis of both OrfG₆₄₋₂₀₄ and OrfG₂₂₃₋₃₃₁, we observed the presence of two well-defined peaks as

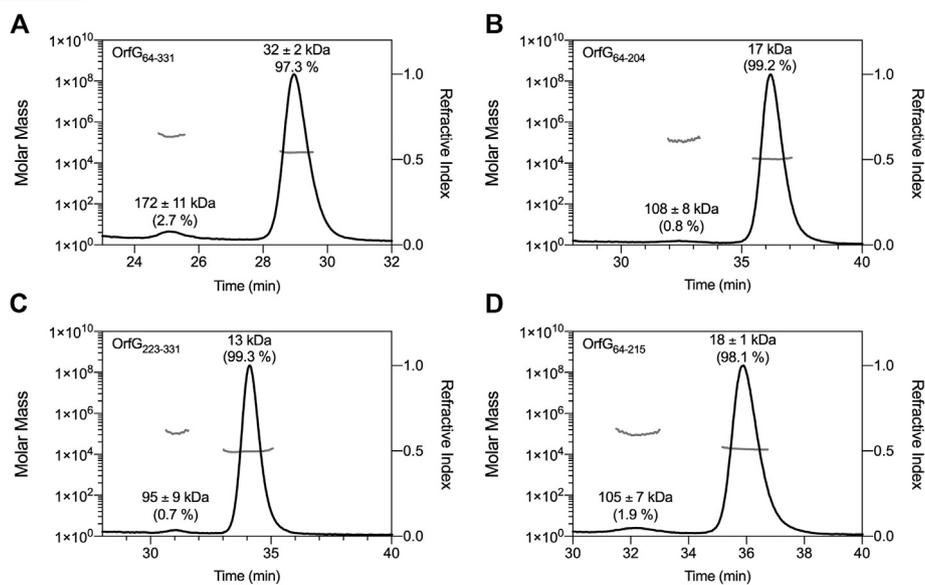


FIGURE 8 | SEC-MALS analysis. Elution profile (black lines) of OrfG₆₄₋₃₃₁ (A), OrfG₆₄₋₂₀₄ (B), OrfG₂₂₃₋₃₃₁ (C) and OrfG₆₄₋₂₁₅ (D) are shown with the molecular weight calculated by MALS (gray lines). The elution time (min) is plotted on the x-axis. The molar mass (in logarithmic scale) is plotted in the first y-axis and the refractive index is plotted in the second y-axis. The molecular weight (in kDa) and the contribution in mass fraction (in %) of each visible peak are shown at the top of the corresponding peak.

seen on the SEC profile of OrfG₆₄₋₃₃₁ (Figures 7B,D,F). These findings prompted us to characterize further the oligomerization of OrfG₆₄₋₂₀₄ and OrfG₂₂₃₋₃₃₁ using SEC-MALS. As observed in Figures 8B,C, both OrfG₆₄₋₂₀₄ and OrfG₂₂₃₋₃₃₁ form two oligomeric states. A predominant monomeric form was eluted at 36 min for OrfG₆₄₋₂₀₄ and at 34 min for OrfG₂₂₃₋₃₃₁ with a calculated molar mass of 17 and 13 kDa, respectively. Additional early peaks were barely detected at 32.5 min and 31 min with calculated molar mass of 108 ± 8 and 95 ± 9 kDa for OrfG₆₄₋₂₀₄ and OrfG₂₂₃₋₃₃₁, respectively. Interestingly, calculated molar masses for the ordered oligomers detected for the OrfG₆₄₋₃₃₁ subdomains were compatible with 6-mer assembly for each subdomain (102 and 78 kDa theoretical MW for OrfG₆₄₋₂₀₄ and OrfG₂₂₃₋₃₃₁, respectively).

DISCUSSION

During the last decades, a great effort was made to understand the mode of action of conjugative type IV secretion systems (Conj-T4SSs) due to their major role in antibiotic resistance spreading among bacteria. More attention was devoted to the study of Conj-T4SS from Gram-negative bacteria. Multiple structural information has been gained for individual components/proteins and isolated complexes improving our understanding of their architecture and assembly mode (reviewed in (Grohmann et al., 2018)). Conj-T4SS from Gram-positive bacteria have been less investigated resulting in little knowledge on their architecture and on the molecular mechanisms underlying DNA transfer in these bacteria (reviewed in (Goessweiner-Mohr et al., 2013a)).

In order to gain further information on the architecture and the assembly mode of Conj-T4SS from Gram-positive bacteria, we focused on the structural, biochemical and biophysical study of OrfG, a putative transfer protein encoded within the conjugative module of ICES_{t3} from *S. thermophilus*. By using X-ray crystallography, we solved the structure of the central domain of OrfG. Structural analysis revealed that OrfG central domain adopts an NTF2-like fold, common to all known VirB8-like proteins in both Gram-negative and -positive Conj-T4SSs. VirB8-like proteins are essential structural and functional component of Conj-T4SSs. Consequently, we propose that OrfG operates as a VirB8-like protein in the conjugative apparatus of ICES_{t3}. The structure of OrfG central domain represents the first structure of a VirB8-like protein from ICEs described so far and rises to four the VirB8-like protein solved structures for Gram-positive Conj-T4SS. Notwithstanding the tricky task to identify VirB8-like proteins in Gram-positive Conj-T4SS clusters due to their low sequence identity, the growing number of VirB8-like structures indicates that it is a conserved protein in conjugation machines and underlines the similarity between protein complexes that govern the transport of mobile genetic elements across the cytoplasmic membrane in both Gram-negative and -positive bacteria.

OrfG central domain shares strong resemblance with the TcpC central domain of pCW3 from *C. perfringens*. Despite their low sequence similarity/identity, *in silico* analysis (see Materials and Methods section) suggests that both OrfG and TcpC adopt the same domain organization with an N-terminal transmembrane domain (TMD) followed by two soluble subdomains. This domain organization appears to be conserved in a large number of putative transfer proteins from both conjugative

plasmids and ICEs alike TcpC from pJIR26, Orf13 from ICECW459, Tn916 and Tn5397 and the VirB8-like protein (conB) from ICEBs1. Our phylogenetic analysis supports this observation showing that these proteins belong to the TcpC family (Figure 1B).

In addition to TcpC and OrfG, two other transfer proteins, TraM and TraH from the conjugative plasmid pIP501 from *E. faecalis*, hold a NTF2-like domain in their C-terminal domains and were proposed to be VirB8-like proteins. Interestingly, both TraM and TraH adopt a different domain organization compared to TcpC and OrfG (Figure 1A). In fact, TraM is composed of two soluble subdomains separated by a TMD while TraH is composed of an N-terminal TMD followed by a unique C-terminal subdomain. The variability in domain organization is restricted to VirB8-like proteins from Gram-positive bacteria since VirB8-like proteins from Gram-negative bacteria share a common domain organization with an N-terminal TMD followed by a single NTF2-like C-terminal periplasmic domain. The modularity/flexibility in domain organization perceived for Gram-positive VirB8-like proteins could be explained in part by the evolutionary scenario of conjugative systems. In fact, it was proposed that Gram-positive conjugative systems emerge from Gram-negative ones by gene deletion and such transfer was followed by diverse adaptive routes (Guglielmini et al., 2013). This significant variation could have originated from the varied cell envelope composition and thickness and/or the functional adaptation of conjugative systems regarding the nature of the donor and recipient cells as well as the variation of the conjugation environment (Alvarez-Martinez and Christie, 2009). Actually, most of the NTF2-like domains of the VirB8-like proteins from Gram-positive Conj-T4SS are localized in the cell-wall, such localization makes these domains more exposed to the environmental stress and this may contribute to their modularity compared to the periplasmic localization of VirB8-like protein in Gram-negative bacteria (Christie, 2016; Waksman, 2019).

Strikingly, although they are subtle, the differences observed when comparing the structures of all the NTF2-like domains of VirB8-like proteins lead again to distinguish Gram-positive and -negative bacteria. The tree based on structural similarities seems significant as it groups structures in a way comparable to the one obtained through phylogenetic analysis (when applicable i.e., for sequences that belong to the PFAM group PF04335, see Figure 4). Moreover, this tree, strictly based on the spatial superposition of atoms without valuing sequence conservation, associates in a same class, proteins that share low sequence similarity. For instance, in the case of Gram-negative bacteria with more known structures, VirB8s from *A. tumefaciens* and from *Bartonella quintana* gather in class II⁻ although they only have 22% of sequence identity; Class III⁻ contains TraM from Plasmid R64 (Kuroda et al., 2015) and the IcmL-like protein from *Legionella pneumophila* despite only 13% of identity (Supplementary Table S2). Thus, the partition of VirB8-like structures from Gram-negative bacteria on one side, and Gram-positive ones on the other side in the tree is highly significant since the low sequence similarity of the latter with any other could have distribute them anywhere in the tree. However, if clear in the tree, this distinction is less apparent when looking at the structures

themselves, so that it is hard to define peculiar features that cause this distinction. Enrichment of the database by new structures of VirB8-like proteins from Gram-positive bacteria will probably bring elements that underpin this partition, as well as it will certainly draw several new classes in this part of the tree. The current distribution separates the monomeric *E. faecalis* TraH/Orf8 from the class I⁺ which contains TcpC from *C. perfringens* and TraM from *E. faecalis*. The presence of OrfG with these two members that share the same trimeric crystal packing questions the physiological assembly of this protein.

One of the most puzzling questions on the study of bacterial secretion systems including the Conj-T4SS concerns the oligomeric state and dynamics of their components inside the cells. In Conj-T4SS, VirB8 proteins are bitopic proteins of the cytoplasmic membrane. In Gram-negative bacteria, many studies suggested/supported that VirB8 function as a multimers. Most of the periplasmic domains of all studied VirB8 proteins adopt a dimeric assembly in solution and it is believed that these dimers are the forming subunits of higher-order multimers which should be structural parts of the functional Conj-T4SS (Casu et al., 2018). Structural and biochemical data proposed that the full-length VirB8 protein self-assembles into a homo-multimers with C3 symmetry. Indeed, the stoichiometry of VirB8 was estimated to 12 in Conj-T4SS from *E. coli* R388 (Low et al., 2014). In line with, the purification of full-length TraE and TraM, a VirB8 homolog from pKM101 and R64, respectively, revealed that TraE and TraM forms hexamers in solution (Kuroda et al., 2015; Casu et al., 2018). Dissimilar to what was described in Gram-negative bacteria, VirB8-like proteins in Gram-positive bacteria are proposed to function as trimers *in vivo* (Porter et al., 2012; Goessweiner-Mohr et al., 2013). Up to date, no higher-order oligomers were obtained for VirB8-like proteins from Gram-positive bacteria supporting a different assembly mode compared to VirB8-like proteins recovered in Gram-negative bacteria. Nevertheless, this variation in the assembly mode do not amend the capacity of VirB8-like proteins to self-assemble following a C3 symmetry independently from the origin of the conjugative system. For Gram-positive VirB8-like proteins, TMD was proposed to trigger trimeric assembly since isolated TMD-less VirB8-like sub-domains were monomeric in solution. In addition, a predicted coiled-coil motif preceding the NTF2-like domain, not conserved in all VirB8-like proteins, seems to play an important role in the stabilization of the trimeric assembly (Goessweiner-Mohr et al., 2013b). In this context, our in-solution characterization revealed intriguing properties concerning OrfG: 1) significant but small fraction of OrfG soluble domain multimerize independently of its TMD and/or a coiled-coil domain since *in silico* analyses failed to find a predicted coiled-coil motif in the purified OrfG soluble domain, 2) OrfG soluble domain has the ability to spontaneously assembles into 6-mer. This ordered oligomeric state observed in solution appears to be a physiological property of OrfG since SEC-MALS analysis performed on OrfG central and C-terminal sub-domains revealed that both form the same ordered-oligomers in solution. The low proportion of OrfG oligomers observed in solution could be explained by the absence of the TMD known to be essential for VirB8

oligomers stabilization (Kuroda et al., 2015; Casu et al., 2018). These data exposed a discrepancy on the number of the forming subunits between OrfG 6-mer oligomers and TcpC and TraM trimers. This could be explained by the fact that OrfG 6-mer oligomers could result from the dimerization of trimeric complexes. Another probable explanation lies on the ability of OrfG to adopt a different assembly mode not yet described for Gram-positive VirB8-like proteins.

Nevertheless, our results combined with published data obtained for TcpC, TraM and TraH consolidate the fact that VirB8-like proteins from Gram-positive bacteria act as multimeric entity on their biological systems but interrogates on their physiological stoichiometry that may depends on their dedicated systems.

VirB8 proteins from Gram-negative bacteria were considered to be interesting targets to develop specific inhibitors of Conj-T4SS by abolishing VirB8 dimerization. These inhibitors interact with a specific site localized in a structurally conserved groove not directly involved in the dimerization of VirB8 proteins (Smith et al., 2012; Casu et al., 2016). Since the inhibitor-binding interface is on the opposite site of dimerization interface, it was proposed that conformational changes on this groove directly affect VirB8 dimerization. Surprisingly, the electron density of OrfG central domain displayed an unassigned blob (**Supplementary Figure S1**) nicely docked in a pocket localized in a hydrophobic cavity that shares intriguing similarity with the groove identified in Gram-negative VirB8 proteins. In fact, residues located on the C-terminal end of α_1 and β_4 delimit both of the identified pockets. Thus, the localization of this groove suggests a probable oligomerization inhibition site for OrfG. The presence of this pocket in VirB8-like proteins could be explained by the nature of the NTF2-like fold. In fact, this fold is widely distributed in bacteria and is associated with various functions including enzymatically active and non-enzymatically active proteins. When extracellular, these proteins often possess non-catalytic ligand-binding activities (Eberhardt et al., 2013).

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

REFERENCES

- Alvarez-Martinez, C. E., and Christie, P. J. (2009). Biological diversity of prokaryotic type IV secretion systems. *Microbiol. Mol. Biol. Rev.* 73, 775–808. doi:10.1128/MMBR.00023-09
- Ambroset, C., Coluzzi, C., Guédon, G., Devignes, M. D., Loux, V., Lacroix, T., et al. (2015). New insights into the classification and integration specificity of *Streptococcus* integrative conjugative elements through extensive genome exploration. *Front. Microbiol.* 6, 1483. doi:10.3389/fmicb.2015.01483
- Arends, K., Celik, E. K., Probst, I., Goessweiner-Mohr, N., Fercher, C., Grumet, L., et al. (2013). TraG encoded by the pIP501 type IV secretion system is a two-domain peptidoglycan-degrading enzyme essential for conjugative transfer. *J. Bacteriol.* 195, 4436–4444. doi:10.1128/JB.02263-12
- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., et al. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 44, W344–W350. doi:10.1093/nar/gkw408
- Bantwal, R., Bannam, T. L., Porter, C. J., Quinsey, N. S., Lyras, D., Adams, V., et al. (2012). The peptidoglycan hydrolase TcpG is required for efficient conjugative transfer of pCW3 in *Clostridium perfringens*. *Plasmid* 67, 139–147. doi:10.1016/j.plasmid.2011.12.016
- Bellanger, X., Roberts, A. P., Morel, C., Choulet, F., Pavlovic, G., Mullany, P., et al. (2009). Conjugative transfer of the integrative conjugative elements ICES1 and ICES3 from *Streptococcus thermophilus*. *J. Bacteriol.* 191, 2764–2775. doi:10.1128/JB.01412-08
- Bhatty, M., Laverde Gomez, J. A., and Christie, P. J. (2013). The expanding bacterial type IV secretion lexicon. *Res. Microbiol.* 164, 620–639. doi:10.1016/j.resmic.2013.03.012

AUTHOR CONTRIBUTION

JC, AM-A, SM, TD, CD, and BD performed the experiments. NL-B phylogeny analysis. JC developed the script used for structural analysis. CD, FF, NL-B, and BD supervision, validation, investigation, methodology. FF and BD Writing original draft. All authors: review and editing of the manuscript.

FUNDING

This work was funded by the National Research Institute for Agriculture, Food and Environment INRAE, the Centre National de la Recherche Scientifique CNRS, the Université de Lorraine and its A2F and CPM scientific poles.

ACKNOWLEDGMENTS

We thank the members of the DynAMic and CRM2 laboratories for insightful discussions; Laurence Hotel, Emilie Piotrowski, Louise Thiriet, Stéphane Bertin, Johan Staub and Anthony Gauthier for technical assistance; Nicolas Soler and Yvonne Roussel for valuable discussions; Emilie Robert and Jean-Michel Girardet from ASIA platform for encouragements. SEC-MALS analysis were conducted using MiniDAWN TREOS II mounted on FPLC system (AKTA purifier) available on the ASIA platform (Université de Lorraine-INRAE, <https://a2f.univ-lorraine.fr/asia/>). The Soleil Synchrotron radiation facility is acknowledged for beamline allocation. The authors appreciated the access to the 'Plateforme de mesures de diffraction X' of the Université de Lorraine.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.642606/full#supplementary-material>.

- Buchan, D. W. A., and Jones, D. T. (2019). The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* 47, W402–W407. doi:10.1093/nar/gkz297
- Burrus, V., Pavlovic, G., Decaris, B., and Guédon, G. (2002). The ICES_{t1} element of *Streptococcus thermophilus* belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration. *Plasmid* 48, 77–97. doi:10.1016/S0147-619X(02)00102-6
- Cascales, E., and Christie, P. J. (2003). The versatile bacterial type IV secretion systems. *Nat. Rev. Microbiol.* 1, 137–149. doi:10.1038/nrmicro753
- Casu, B., Mary, C., Sverzhinsky, A., Fouillen, A., Nanci, A., and Baron, C. (2018). VirB8 homolog TraE from plasmid pKM101 forms a hexameric ring structure and interacts with the VirB6 homolog TraD. *Proc. Natl. Acad. Sci. USA* 115, 5950–5955. doi:10.1073/pnas.1802501115
- Casu, B., Smart, J., Hancock, M. A., Smith, M., Sygusch, J., and Baron, C. (2016). Structural analysis and inhibition of TraE from the pKM101 type IV secretion system. *J. Biol. Chem.* 291, 23817–23829. doi:10.1074/jbc.M116.753327
- Christie, P. J. (2016). The mosaic type IV secretion systems. *EcoSal Plus* 7. doi:10.1128/ecosalplus.ESP-0020-2015
- Dahmane, N., Robert, E., Deschamps, J., Meylheuc, T., Delorme, C., Briandet, R., et al. (2018). Impact of cell surface molecules on conjugative transfer of the integrative and conjugative element ICES_{t3} of *Streptococcus thermophilus*. *Appl. Environ. Microbiol.* 84, e02109. doi:10.1128/AEM.02109-17
- Dong, R., Pan, S., Peng, Z., Zhang, Y., and Yang, J. (2018a). mTM-align: a server for fast protein structure database search and multiple protein structure alignment. *Nucleic Acids Res.* 46, W380–W386. doi:10.1093/nar/gky430
- Dong, R., Peng, Z., Zhang, Y., and Yang, J. (2018b). mTM-align: an algorithm for fast and accurate multiple protein structure alignment. *Bioinformatics* 34, 1719–1725. doi:10.1093/bioinformatics/btx828
- Dunny, G. M. (2007). The peptide pheromone-inducible conjugation system of *Enterococcus faecalis* plasmid pCF10: cell-cell signalling, gene transfer, complexity and evolution. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 362, 1185–1193. doi:10.1098/rstb.2007.2043
- Eberhardt, R. Y., Chang, Y., Bateman, A., Murzin, A. G., Axelrod, H. L., Hwang, W. C., et al. (2013). Filling out the structural map of the NTF2-like superfamily. *BMC Bioinformatics* 14, 327. doi:10.1186/1471-2105-14-327
- Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501. doi:10.1107/S0907444910007493
- Evans, P. R., and Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* 69, 1204–1214. doi:10.1107/S0907444913000061
- Fercher, C., Probst, I., Kohler, V., Goessweiner-Mohr, N., Arends, K., Grohmann, E., et al. (2016). VirB8-like protein TraH is crucial for DNA transfer in *Enterococcus faecalis*. *Sci. Rep.* 6, 24643. doi:10.1038/srep24643
- Fronzes, R., Christie, P. J., and Waksman, G. (2009). The structural biology of type IV secretion systems. *Nat. Rev. Microbiol.* 7, 703–714. doi:10.1038/nrmicro2218
- Frost, L. S., Leplae, R., Summers, A. O., and Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* 3, 722–732. doi:10.1038/nrmicro1235
- Gabler, F., Nam, S. Z., Till, S., Mirdita, M., Steinegger, M., Söding, J., et al. (2020). Protein sequence analysis using the MPI bioinformatics toolkit. *Curr. Protoc. Bioinformatics* 72, e108. doi:10.1002/cpbi.108
- Goessweiner-Mohr, N., Arends, K., Keller, W., and Grohmann, E. (2013a). Conjugative type IV secretion systems in Gram-positive bacteria. *Plasmid* 70, 289–302. doi:10.1016/j.plasmid.2013.09.005
- Goessweiner-Mohr, N., Grumet, L., Arends, K., Pavkov-Keller, T., Gruber, C. C., Gruber, K., et al. (2013b). The 2.5 Å structure of the enterococcus conjugation protein TraM resembles VirB8 type IV secretion proteins. *J. Biol. Chem.* 288, 2018–2028. doi:10.1074/jbc.M112.428847
- González-Rivera, C., Khara, P., Awad, D., Patel, R., Li, Y. G., Bogisch, M., et al. (2019). Two pKM101-encoded proteins, the pilus-tip protein TraC and Pep, assemble on the *Escherichia coli* cell surface as adhesins required for efficient conjugative DNA transfer. *Mol. Microbiol.* 111, 96–117. doi:10.1111/mmi.14141
- Grohmann, E., Christie, P. J., Waksman, G., and Backert, S. (2018). Type IV secretion in gram-negative and gram-positive bacteria. *Mol. Microbiol.* 107, 455–471. doi:10.1111/mmi.13896
- Guglielmini, J., de la Cruz, F., and Rocha, E. P. (2013). Evolution of conjugation and type IV secretion systems. *Mol. Biol. Evol.* 30, 315–331. doi:10.1093/molbev/mss221
- Guglielmini, J., Quintais, L., Garcillán-Barcia, M. P., de la Cruz, F., and Rocha, E. P. (2011). The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *Plos Genet.* 7, e1002222. doi:10.1371/journal.pgen.1002222
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi:10.1093/sysbio/syq010
- Holm, L., and Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* 20, 478–480. doi:10.1016/S0968-0004(00)89105-7
- Ilangovan, A., Kay, C. W. M., Roier, S., El Mkami, H., Salvadori, E., Zechner, E. L., et al. (2017). Cryo-EM structure of a relaxase reveals the molecular basis of DNA unwinding during bacterial conjugation. *Cell* 169, 708–721. doi:10.1016/j.cell.2017.04.010
- Kabsch, W. (2010). Xds. *Acta Crystallogr. D Biol. Crystallogr.* 66, 125–132. doi:10.1107/S0907444909047337
- Kohler, V., Keller, W., and Grohmann, E. (2018). Enterococcus adhesin PrgB facilitates type IV secretion by condensation of extracellular DNA. *Mol. Microbiol.* 109, 263–267. doi:10.1111/mmi.13994
- Kohler, V., Probst, I., Aufschnaiter, A., Büttner, S., Schaden, L., Rechberger, G. N., et al. (2017). Conjugative type IV secretion in Gram-positive pathogens: TraG, a lytic transglycosylase and endopeptidase, interacts with translocation channel protein TraM. *Plasmid* 91, 9–18. doi:10.1016/j.plasmid.2017.02.002
- Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797. doi:10.1016/j.jmb.2007.05.022
- Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2256–2268. doi:10.1107/S0907444904026460
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi:10.1093/molbev/msw054
- Kuroda, T., Kubori, T., Thanh Bui, X., Hyakutake, A., Uchida, Y., Imada, K., et al. (2015). Molecular and structural analysis of Legionella DotI gives insights into an inner membrane complex essential for type IV secretion. *Sci. Rep.* 5, 10912. doi:10.1038/srep10912
- Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* 3, 1171–1179. doi:10.1038/nprot.2008.91
- Laverde Gomez, J. A., Bhaty, M., and Christie, P. J. (2014). PrgK, a multidomain peptidoglycan hydrolase, is essential for conjugative transfer of the pheromone-responsive plasmid pCF10. *J. Bacteriol.* 196, 527–539. doi:10.1128/JB.00950-13
- Leonetti, C. T., Hamada, M. A., Laurer, S. J., Broulidakis, M. P., Swerdlow, K. J., Lee, C. A., et al. (2015). Critical components of the conjugation machinery of the integrative and conjugative element ICEBs1 of *Bacillus subtilis*. *J. Bacteriol.* 197, 2558–2567. doi:10.1128/JB.00142-15
- Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* 75, 861–877. doi:10.1107/S2059798319011471
- Low, H. H., Gubellini, F., Rivera-Calzada, A., Braun, N., Connery, S., Dujancourt, A., et al. (2014). Structure of a type IV secretion system. *Nature* 508, 550–553. doi:10.1038/nature13081
- Lupas, A., Van Dyke, M., and Stock, J. (1991). Predicting coiled coils from protein sequences. *Science* 252, 1162–1164. doi:10.1126/science.252.5009.1162
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674. doi:10.1107/S0021889807021206
- Porter, C. J., Bantwal, R., Bannam, T. L., Rosado, C. J., Pearce, M. C., Adams, V., et al. (2012). The conjugation protein TcpC from *Clostridium perfringens* is structurally related to the type IV secretion system protein VirB8 from Gram-negative bacteria. *Mol. Microbiol.* 83, 275–288. doi:10.1111/j.1365-2958.2011.07930.x
- Roberts, A. P., and Mullany, P. (2011). Tn916-like genetic elements: a diverse group of modular mobile elements conferring antibiotic resistance. *FEMS Microbiol. Rev.* 35, 856–871. doi:10.1111/j.1574-6976.2011.00283.x
- Santorio, F., Vianna, M. E., and Roberts, A. P. (2014). Variation on a theme; an overview of the Tn916/Tn1545 family of mobile genetic elements in the oral and nasopharyngeal streptococci. *Front. Microbiol.* 5, 535. doi:10.3389/fmicb.2014.00535

- Schroeder, M. R., and Stephens, D. S. (2016). Macrolide resistance in *Streptococcus pneumoniae*. *Front Cell Infect Microbiol* 6, 98. doi:10.3389/fcimb.2016.00098
- Smith, M. A., Coinçon, M., Paschos, A., Jolicoeur, B., Lavallée, P., Sygusch, J., et al. (2012). Identification of the binding site of Brucella VirB8 interaction inhibitors. *Chem. Biol.* 19, 1041–1048. doi:10.1016/j.chembiol.2012.07.007
- Soler, N., Robert, E., Chauvot de Beauchêne, I., Monteiro, P., Libante, V., Maigret, B., et al. (2019). Characterization of a relaxase belonging to the MOB T family, a widespread family in Firmicutes mediating the transfer of ICEs. *Mob DNA* 10, 18. doi:10.1186/s13100-019-0160-9
- Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10, 512–526. doi:10.1093/oxfordjournals.molbev.a040023
- Waksman, G. (2019). From conjugation to T4S systems in Gram-negative bacteria: a mechanistic biology perspective. *EMBO Rep.* 20. doi:10.15252/embr.201847012
- Wisniewski, J. A., and Rood, J. I. (2017). The Tcp conjugation system of *Clostridium perfringens*. *Plasmid* 91, 28–36. doi:10.1016/j.plasmid.2017.03.001
- Wisniewski, J. A., Teng, W. L., Bannam, T. L., and Rood, J. I. (2015). Two novel membrane proteins, TcpD and TcpE, are essential for conjugative transfer of pCW3 in *Clostridium perfringens*. *J. Bacteriol.* 197, 774–781. doi:10.1128/JB.02466-14
- Wolf, E., Kim, P. S., and Berger, B. (1997). MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci.* 6, 1179–1189. doi:10.1002/pro.5560060606
- Wu, X., Zhao, Y., Sun, L., Jiang, M., Wang, Q., Wang, Q., et al. (2019). Crystal structure of CagV, the *Helicobacter pylori* homologue of the T4SS protein VirB8. *FEBS J.* 286, 4294–4309. doi:10.1111/febs.14971
- Zimmermann, L., Stephens, A., Nam, S. Z., Rau, D., Kübler, J., Lozajic, M., et al. (2018). A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* 430, 2237–2243. doi:10.1016/j.jmb.2017.12.007

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Cappele, Mohamad Ali, Leblond-Bourget, Mathiot, Dhalleine, Payot, Savko, Didierjean, Favier and Douzi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Table S1. Strains, plasmids and oligonucleotides used in this project

Strains		
Name	Genotype	Source/reference
<i>Streptococcus thermophilus</i>	LMG18311 ICESf3	Laboratory collection
<i>DH5α</i>	<i>F-</i> , Δ (<i>argF-lacZ</i>)U169, <i>phoA</i> , <i>supE44</i> , Δ (<i>lacZ</i>)M15, <i>relA</i> , <i>endA</i> , <i>thi</i> , <i>hsdR</i>	New England Biolabs
<i>BL21 (DE3)</i>	<i>F-</i> <i>ompT gal dcm lon hsdSB(rB- mB-) λ(DE3)</i>	Invitrogen
Plasmids		
Name	Description	Source/reference
pET28a-TRX	6His-tag-TEV protease sequences inserted into pET28a+	Laboratory collection
pET28-TRX-OrfG ₆₄₋₃₃₁	<i>orfG</i> ₆₄₋₃₃₁ cloned in frame of 6His-TEV in pET28-TRX	This project
pET28-TRX-OrfG ₆₄₋₂₀₄	<i>orfG</i> ₆₄₋₂₀₄ cloned in frame of 6His-TEV in pET28-TRX	This project
pET28-TRX-OrfG ₆₄₋₂₁₅	<i>orfG</i> ₆₄₋₂₁₅ cloned in frame of 6His-TEV in pET28-TRX	This project
pET28-TRX-OrfG ₂₂₃₋₃₃₁	<i>orfG</i> ₂₂₃₋₃₃₁ cloned in frame of 6His-TEV in pET28-TRX	This project
Oligonucleotides		
For cloning in pET28-TRX*		
OrfG-BamHI-F	CCCCGGATCCGAGAACCTGTACTTCCAATCATCAAAATAAAGTGGGAACACTAC	
OrfG-HindIII-R	CCCCAAGCTTCTAAATCTTTATCGTTTGGGTAG	
OrfGc-BamHI-F	CCCCGGATCCGAGAACCTGTACTTCCAATCAGCCAACGACCACCGTCTCAG	
*Restrictions sites are underlined and TEV cleavage encoding sequence are highlighted in italics		
Cloning by site directed mutagenesis (Stop codon insertion by PCR)*		
OrfG ₆₄₋₂₀₄ -Stop-F	ACTTCCTTGGTTCTCAGCGTAAAGAGTCTTCACAAGTGGATAC	
OrfG ₆₄₋₂₀₄ -Stop-R	GTATCCAGCTTGTGAAGACTCTTAGCGCTGAGAACCAAGGAAGT	
OrfG ₆₄₋₂₁₅ -Stop-F	AGCTGGATACTTTAGCGGAATAGGATCAACTCCAACCTCACAG	
OrfG ₆₄₋₂₁₅ -Stop-R	CTGTGAGTTGGAGTTGATCCCTATTCGGCTAAAAGTATCCAGCT	
* Mutagenized codons are underlined		

Tables S2. Pairwise alignment lengths, identities and similarities between the 13 Gram-negative VirB8 and 4 Gram-positive VirB8-like proteins considered in this comparison (**Table S3a**). The sequences are derived from the structures, i.e. there might be missing residues (see protein description below) corresponding to unobserved regions in the electron density. Values are calculated from the alignment of the sequences corresponding to the overall superposition of the 17 structures, as performed by mTM-align (Dong *et al.*, 2018a). The TM-scores (Dong *et al.*, 2018b) and the pairwise rmsd (Å) calculated from the same overall superposition are also given (**Tables S3b and S3c**). Clusters that appear in the structural tree are coloured as in Fig.5 and numbered for their convenient description in the main text.

For each following PDB code, the source organism, the Uniprot code, the sequence that was used for the experiment (X-Ray or NMR), where appropriate the unobserved residues (which are therefore missing in the sequence) and the resulting coverage of the whole sequence are given hereafter.

- 2cc3: VirB8 from *Agrobacterium tumefaciens*; Uniprot: P17798; Crystallized sequence: 92-237; C-ter (232-237) unobserved in the electron density; Coverage 59%.
- 3ub1: TcpC from *Clostridium perfringens*; Uniprot: Q1PLH8; Crystallized sequence: 99-309, central domain as defined by authors: 104-231; loop β 2- β 3 (190-192) unobserved in the electron density (coverage 35%).
- 3wz3: TraM from Plasmid R64; Uniprot Q9R2H2; Crystallized sequence: 91-230; C-ter (227-230) unobserved in the electron density (coverage 59%).
- 3wz4: IcmL/DotI from *Legionella pneumophila*; Uniprot: O54626; Crystallized sequence: 73-212; C-ter (209-212) unobserved in the electron density (coverage 64%).
- 4akz: VirB8 from *Brucella suis*; Uniprot: Q7CEG3; Crystallized sequence: 97-234 (coverage 57%).
- 4ec6: TraM from *Enterococcus faecalis*; Uniprot: Q8L1C7; Crystallized sequence: 190-322; N-ter (190-213) unobserved in the electron density (coverage 34%).
- 4jf8: TrwG from *Bartonella birtlesii*; Uniprot: D0AAZ5; Crystallized sequence: 63-232; N-ter (63-88) unobserved in the electron density (coverage 62%).
- 4kz1: VirB8 from *Bartonella grahamii*; Uniprot: C6AER9; Crystallized sequence: 51-222; N-ter (51-84) and C-ter (220-222) unobserved in the electron density; Coverage 61%.
- 4lso: VirB8 from *Bartonella quintana*; Uniprot: Q6FYW3; Crystallized sequence: 65-234; N-ter (65-93) unobserved in the electron density; Coverage 60%.
- 4mei: VirB8 from *Bartonella tribocurum*; Uniprot: A9IWN6; Crystallized sequence: 51-222; N-ter (51-84) and C-ter (221-222) unobserved in the electron density; Coverage 61%.
- 4nhf: VirB8 from *Bartonella grahamii*; Uniprot: C6AAT5; Crystallized sequence: 63-233; N-ter (63-88) and C-ter (233) unobserved in the electron density (coverage 61%).
- 4o3v: VirB8 from *Rickettsia typhi*; Uniprot: Q68X84; Crystallized sequence: 61-232; N-ter (61-89), loop β 3- β 4 (208-209) and C-ter (232) unobserved in the electron density (coverage 60%).
- 5aiw: TraH/Orf8 from *Enterococcus faecalis*; Uniprot: Q9AL12; NMR sequence: 56-183 (coverage 69%).
- 5cnl: IcmL-like from *Legionella pneumophila*; Uniprot: Q5ZZ91; Crystallized sequence: 22-176; C-ter (162-176) unobserved in the electron density (coverage 79%).
- 5i97: TraE from *Escherichia coli*; Uniprot: Q17U16; Crystallized sequence: 70-232; N-ter (70-87) and C-ter (231-232) unobserved in the electron density (coverage 59%).
- 6iqf: CagV from *Helicobacter pylori*; Uniprot: O25265; Crystallized sequence: 74-229; N-ter (74-85), loop β 3- β 4 (202-209) and C-ter (229) unobserved in the electron density (coverage 51%).
- 6zgn: OrfG from *Streptococcus thermophilus*; Uniprot: Q70CA4; Crystallized sequence: 65-204; N-ter (65-91), loop before β 1 (134-138) and loop β 2- β 3 (169-175) unobserved in the electron density (coverage 31%).

Table S2a: Sequence comparison

		Gram-negative VirB8												Gram-positive VirB8-like				
		Class I ⁻				Class II ⁻				Class III ⁻				Class I ⁺				
		5i97	4akz	4nhf	4jf8	4mei	4kz1	4lso	2cc3	6iqt	4o3v	3wz4	3wz3	5cnl	6zgn	3ub1	4ec6	5aiw
5i97	length	143																
	ident.	1																
	simil.	1																
4akz	length	145	138															
	ident.	0.414	1															
	simil.	0.586	1															
4nhf	length	149	148	144														
	ident.	0.383	0.419	1														
	simil.	0.557	0.534	1														
4jf8	length	150	149	145	144													
	ident.	0.367	0.422	0.976	1													
	simil.	0.573	0.564	0.924	1													
4mei	length	150	145	149	150	136												
	ident.	0.200	0.338	0.282	0.287	1												
	simil.	0.413	0.51	0.456	0.473	1												
4kz1	length	149	144	149	150	136	135											
	ident.	0.215	0.340	0.295	0.300	0.956	1											
	simil.	0.436	0.535	0.456	0.473	0.971	1											
4lso	length	152	150	150	151	141	141	141										
	ident.	0.204	0.300	0.253	0.265	0.837	0.837	1										
	simil.	0.454	0.493	0.460	0.477	0.879	0.894	1										
2cc3	length	157	153	158	159	146	146	151	144									
	ident.	0.248	0.242	0.241	0.233	0.240	0.240	0.219	1									
	simil.	0.331	0.279	0.261	0.346	0.425	0.425	0.431	1									
6iqt	length	154	151	157	158	151	150	155	157	135								
	ident.	0.091	0.139	0.134	0.139	0.133	0.133	0.136	0.115	1								
	simil.	0.227	0.258	0.280	0.279	0.265	0.267	0.252	0.274	1								
4o3v	length	163	158	163	164	152	152	156	156	156	138							
	ident.	0.147	0.196	0.172	0.195	0.151	0.145	0.160	0.186	0.103	1							
	simil.	0.282	0.310	0.319	0.335	0.265	0.322	0.308	0.333	0.244	1							
3wz4	length	165	161	165	166	154	153	158	163	154	162	137						
	ident.	0.091	0.093	0.091	0.102	0.104	0.085	0.089	0.074	0.104	0.062	1						
	simil.	0.249	0.248	0.285	0.295	0.292	0.288	0.272	0.282	0.266	0.235	1						
3wz3	length	166	161	165	166	154	153	159	163	155	161	139	137					
	ident.	0.042	0.081	0.097	0.054	0.091	0.084	0.076	0.074	0.058	0.037	0.266	1					
	simil.	0.187	0.205	0.206	0.175	0.260	0.261	0.245	0.227	0.187	0.199	0.453	1					
5cnl	length	169	164	167	168	157	157	161	166	158	164	148	149	139				
	ident.	0.112	0.110	0.132	0.125	0.096	0.096	0.099	0.060	0.101	0.073	0.264	0.134	1				
	simil.	0.231	0.226	0.252	0.262	0.236	0.255	0.248	0.217	0.234	0.220	0.432	0.322	1				
6zgn	length	152	147	153	153	145	144	150	153	145	149	144	144	146	101			
	ident.	0.059	0.068	0.059	0.072	0.062	0.063	0.053	0.059	0.076	0.060	0.056	0.069	0.062	1			
	simil.	0.197	0.204	0.177	0.170	0.172	0.181	0.160	0.163	0.186	0.168	0.188	0.208	0.164	1			
3ub1	length	169	165	171	171	163	162	168	170	158	165	159	159	160	127	122		
	ident.	0.041	0.042	0.035	0.041	0.080	0.086	0.083	0.035	0.101	0.042	0.063	0.063	0.063	0.095	1		
	simil.	0.189	0.188	0.164	0.175	0.178	0.185	0.173	0.171	0.203	0.182	0.182	0.201	0.181	0.323	1		
4ec6	length	158	153	157	157	151	150	155	159	147	155	146	146	147	114	129	109	
	ident.	0.051	0.046	0.045	0.051	0.046	0.047	0.032	0.038	0.088	0.039	0.055	0.089	0.075	0.105	0.109	1	
	simil.	0.190	0.206	0.153	0.159	0.166	0.167	0.155	0.182	0.218	0.194	0.240	0.206	0.197	0.290	0.256	1	
5aiw	length	165	161	166	167	158	157	162	167	154	161	154	155	159	141	155	144	128
	ident.	0.067	0.087	0.108	0.096	0.063	0.064	0.056	0.066	0.065	0.118	0.065	0.071	0.031	0.028	0.077	0.049	1
	simil.	0.200	0.193	0.217	0.198	0.203	0.204	0.191	0.186	0.175	0.236	0.214	0.232	0.195	0.177	0.200	0.208	1

Table S2b : TM-scores

	Gram-negative VirB8														Gram-positive VirB8-like			
	Class I ⁻				Class II ⁻				Class III ⁻						Class I ⁺			
	5i97	4akz	4nhf	4jf8	4mei	4kz1	4lso	3cc3	6iqt	4o3v	3wz4	3wz3	5cnl	6zgn	3ub1	4ec6	5aiw	
5i97	1																	
4akz	0.9167	1																
4nhf	0.8942	0.9071	1															
4jf8	0.893	0.9259	0.9438	1														
4mei	0.8768	0.8754	0.8789	0.8699	1													
4kz1	0.8854	0.8842	0.879	0.8748	0.9948	1												
4lso	0.8665	0.8619	0.878	0.8702	0.9819	0.984	1											
2cc3	0.825	0.8901	0.8365	0.838	0.9123	0.9154	0.883	1										
6iqt	0.8129	0.8043	0.8053	0.7898	0.8311	0.8298	0.8377	0.8139	1									
4o3v	0.7696	0.7971	0.8158	0.7971	0.8208	0.8247	0.8242	0.8326	0.784	1								
3wz4	0.7218	0.7233	0.7432	0.7402	0.7669	0.7691	0.7674	0.7462	0.7733	0.739	1							
3wz3	0.7054	0.7159	0.7391	0.7371	0.7257	0.7291	0.7221	0.7282	0.7408	0.7414	0.8873	1						
5cnl	0.7108	0.7177	0.7397	0.7324	0.7519	0.75	0.7468	0.7258	0.7738	0.7559	0.8823	0.8142	1					
6zgn	0.733	0.7587	0.7454	0.7442	0.743	0.7445	0.7426	0.7586	0.7462	0.7548	0.7725	0.7626	0.7655	1				
3ub1	0.6146	0.6399	0.6202	0.6237	0.6259	0.6266	0.6342	0.6268	0.6432	0.656	0.6907	0.6641	0.6714	0.8122	1			
4ec6	0.6871	0.7121	0.7052	0.715	0.7119	0.7081	0.7212	0.6703	0.7231	0.6991	0.7363	0.716	0.7169	0.8057	0.7764	1		
5aiw	0.6882	0.6891	0.7054	0.6988	0.6775	0.6757	0.6976	0.6962	0.7257	0.7022	0.7033	0.7047	0.678	0.7036	0.6537	0.656	1	

Table S2c: rmsd (Å)

	Gram-negative VirB8														Gram-positive VirB8-like			
	Class I ⁻				Class II ⁻				Class III ⁻						Class I ⁺			
	5i97	4akz	4nhf	4jf8	4mei	4kz1	4lso	3cc3	6iqt	4o3v	3wz4	3wz3	5cnl	6zgn	3ub1	4ec6	5aiw	
5i97	0																	
4akz	1.426	0																
4nhf	1.628	1.631	0															
4jf8	1.56	1.525	1.316	0														
4mei	2.059	1.837	1.738	1.812	0													
4kz1	2.062	1.818	1.754	1.758	0.326	0												
4lso	2.133	1.86	1.969	2.015	0.602	0.562	0											
2cc3	2.128	1.503	1.996	2.114	1.483	1.416	1.704	0										
6iqt	2.286	2.013	2.256	2.505	1.513	1.395	1.738	1.979	0									
4o3v	2.942	2.172	2.495	2.465	2.061	2	2.348	2.488	2.037	0								
3wz4	3.108	2.843	2.687	2.628	2.096	2.103	2.134	2.456	2.408	2.379	0							
3wz3	3.179	2.95	2.694	2.715	2.799	2.82	2.84	2.759	2.77	2.638	1.691	0						
5cnl	3.045	2.823	2.902	3.152	2.301	2.301	2.475	2.814	2.498	2.452	1.631	2.161	0					
6zgn	2.704	2.444	2.605	2.62	2.401	2.402	2.411	2.314	2.359	2.33	2.262	2.393	2.399	0				
3ub1	3.362	3.271	3.675	3.415	3.312	3.248	3.332	3.263	3.219	3.002	2.849	3.18	3.042	1.99	0			
4ec6	2.841	2.718	2.939	2.767	2.691	2.673	2.558	3.114	2.762	2.757	2.599	2.939	2.97	1.878	2.301	0		
5aiw	2.943	2.777	2.931	2.888	2.774	2.757	2.852	2.997	2.774	3.13	2.939	2.863	3.034	2.474	2.637	3.12	0	

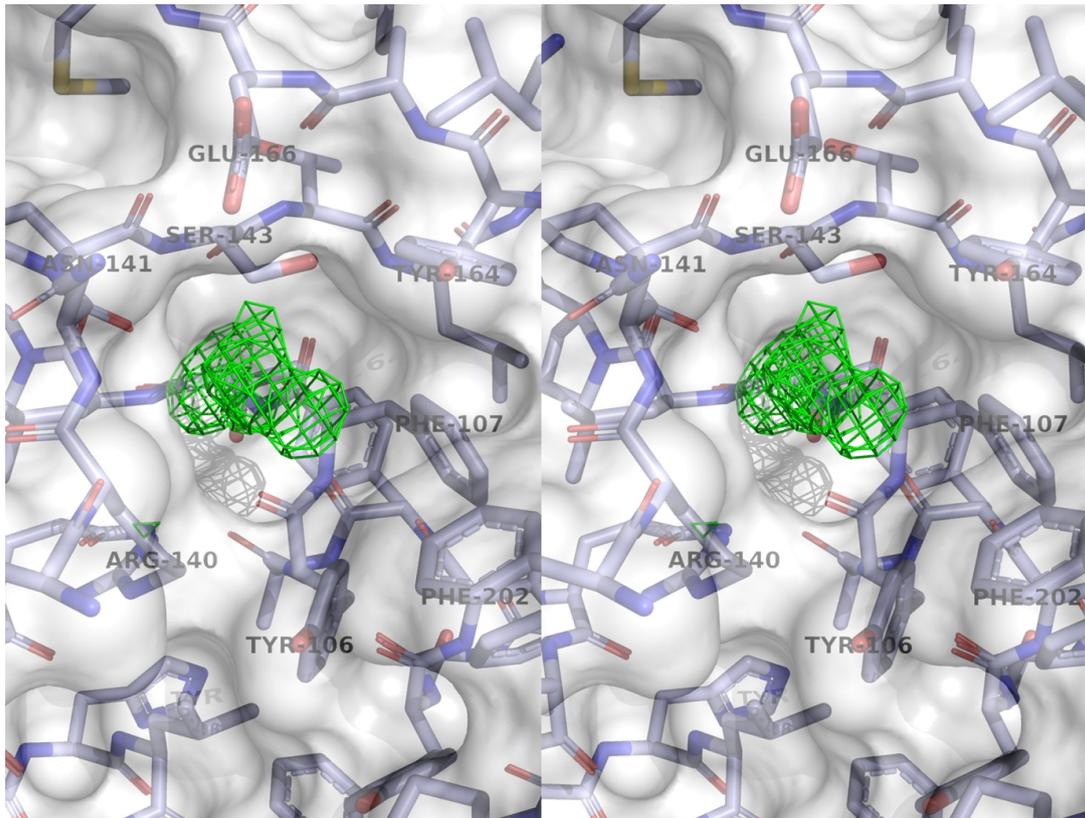


Figure S1: Presence of an unknown ligand bound to OrfG₆₄₋₂₀₄. Stereo view of an unattributed blob of electron density observed in the Fo-Fc map contoured at $+3\sigma$. Surrounding amino acids are shown in sticks and labeled.

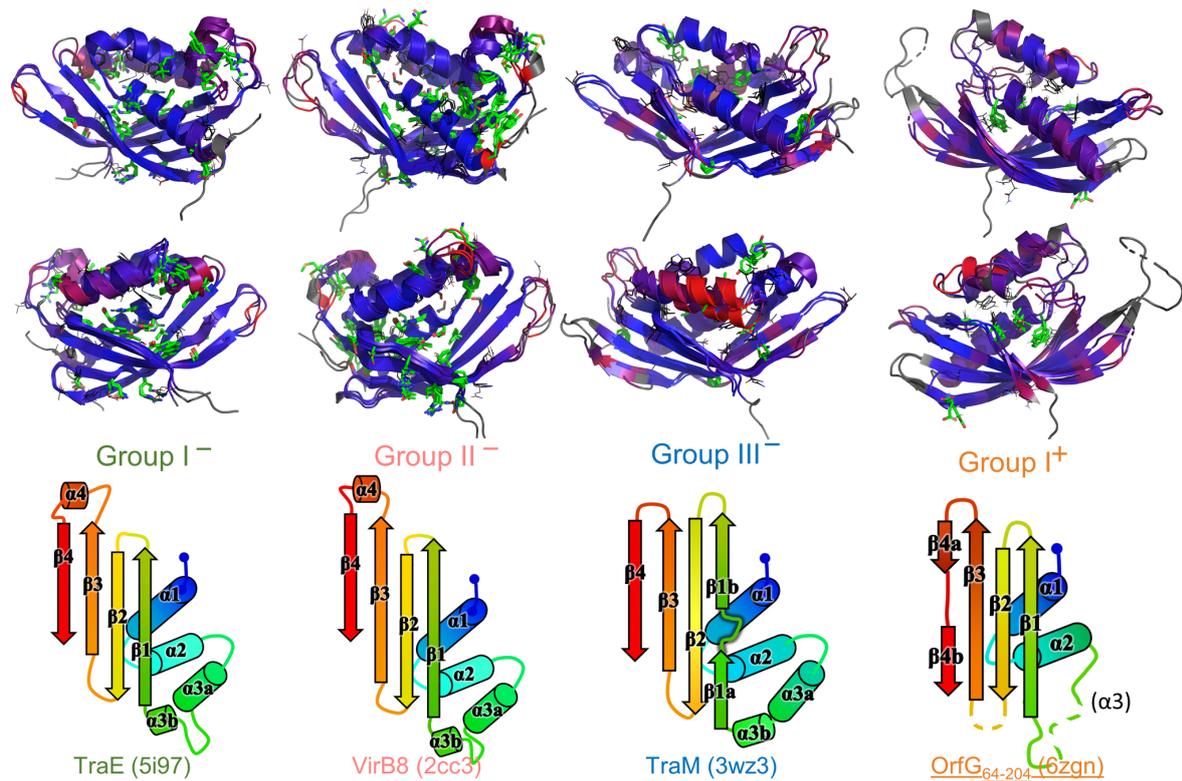
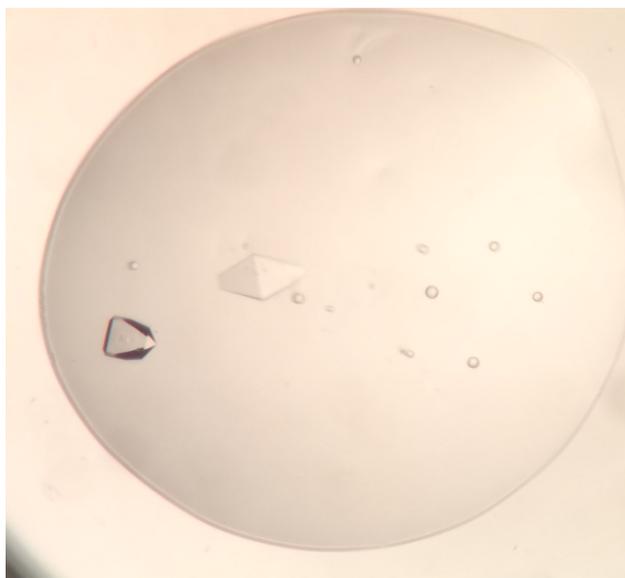


Figure S2: Superposition of VirB8-like structures within each class identified in the structural tree. Structures were superposed by mTM-Align, and represented using the same color code as in Figure 3 (see also Materials and Methods). Amino acid sidechains that are conserved are shown in green sticks, those that share a similar nature within a class are drawn as black lines. For each class, front view, rear view and secondary structure topology of a representative member are shown vertically.

VII.2 Complément cristallographique à l'article pour le domaine central OrfG₆₄₋₂₀₄

VII.2.1 Collecte des données de diffraction pour la résolution de la structure

FIGURE 33 – Cristaux de la construction OrfG₆₄₋₂₀₄

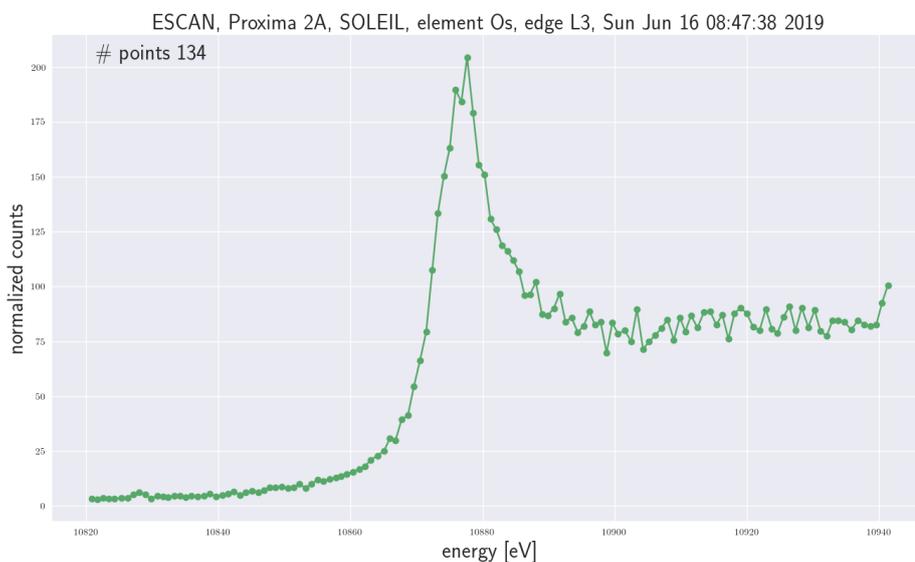


Cristaux de la construction OrfG₆₄₋₂₀₄ dans la condition contenant 2 M de sulfate d'ammonium, 100 mM Tris-HCl, pH 8,5. Détails dans la partie "*Materials & Methods*" de l'article Cappele *et al.*, 2021.

L'ensemble des collectes a été réalisé au synchrotron SOLEIL sur la ligne Proxima-2. Pour cela, 2 cristaux d'OrfG₆₄₋₂₀₄ (voir Figure 33, page 103) avaient été trempés dans une solution contenant de l'hexachloroosmate(IV) de potassium, c'est-à-dire des diffuseurs anomaux (les atomes d'osmium) dont la présence a pour but de permettre l'utilisation de la méthode SAD ("*Single-wavelength anomalous dispersion*") afin de résoudre le problème de la phase. En effet, les homologues structuraux d'OrfG₆₄₋₂₀₄ n'ayant pas une identité de séquence atteignant les 20%, nous ne pensions pas pouvoir déterminer les phases grâce au remplacement moléculaire. Le spectre d'émission de fluorescence autour du seuil d'excitation L3 de l'osmium dans le cristal est disponible sur la figure 34 (page 104) et les facteurs anomaux qui en découlent sont représentés sur la figure 35 (page 105).

L'analyse des données de diffraction de la première collecte au pic de l'osmium ($\lambda = 1,1398 \text{ \AA}$) révélait un signal anomal faible et insuffisant pour résoudre le problème de la phase. Martin Savko, notre local contact, nous a alors conseillé de collecter plusieurs jeux de données en modifiant à

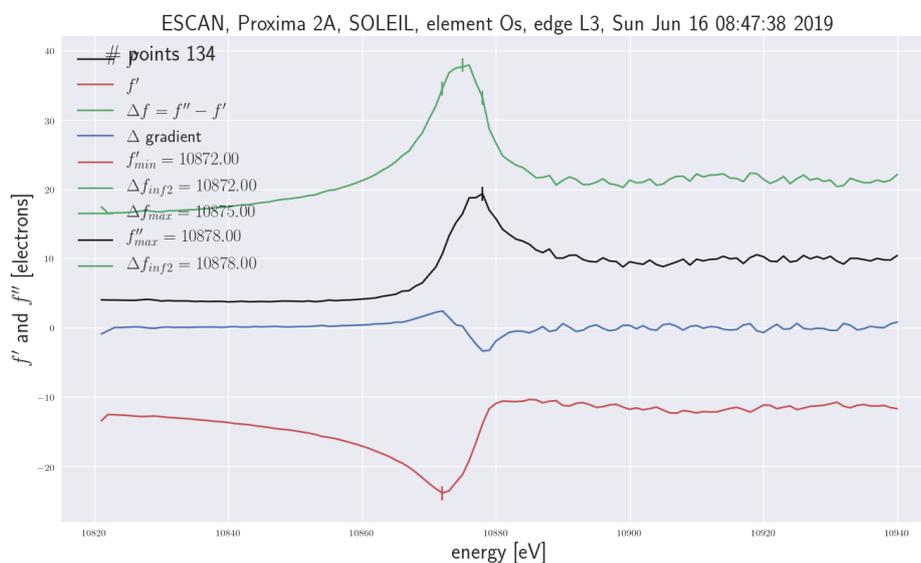
FIGURE 34 – Spectre de fluorescence d'un cristal d'OrfG₆₄₋₂₀₄ enrichi en osmium



Spectre brut de l'émission de fluorescence en fonction de l'énergie excitation (eV) au voisinage du seuil L3 de l'osmium, pour un cristal d'OrfG₆₄₋₂₀₄ trempé dans l'hexachloroosmate de potassium. La mesure, réalisée sur la ligne Proxima 2A du synchrotron SOLEIL, confirme la présence du diffuseur anormal dans les cristaux.

chaque collecte l'angle kappa du goniomètre et en étant centré à différents endroits du cristal afin de limiter les dommages causés par les irradiations répétées. 7 jeux de données ont été collectés en utilisant les deux cristaux, pour une redondance moyenne anormale de 81. Les statistiques décrivant ces jeux de données sont disponibles en annexe S4, S5 et S6 respectivement page 208 à 210. Elles montrent que la qualité de chaque jeu est très satisfaisante au regard des critères usuels.

FIGURE 35 – Contribution anormale de l'atome d'osmium



Les contributions anormales f' (en rouge) et f'' (en noir) au facteur de diffusion de l'osmium contenu dans le cristal d'OrfG₆₄₋₂₀₄, au voisinage du seuil d'absorption L3.

VII.2.2 Phasage à l'osmium

La localisation des diffuseurs anomaux puis le phasage ont été réalisés en utilisant le pipeline SAD Phaser disponible dans la suite CCP4 (Collaborative Computational Project, 1994). Toutefois, il faut bien préciser que la procédure n'a réussi que lorsque les 7 jeux ont été regroupés ensemble. En testant l'ajout progressif des jeux un à un, aucun phasage n'a abouti avant que le dernier jeu ne soit ajouté. Les données statistiques fournies par le logiciel XDS de traitement des données (Kabsch, 2010) puis le pipeline SAD Phaser sont disponibles dans la Figure 36, page 106. Les 5 premiers jeux correspondent à un même cristal et sont ordonnés selon l'angle κ croissant, de 0° à 67° . Les deux derniers jeux correspondent au second cristal aux angles 0° et 27° . Aucun essai n'a réussi avec les jeux du premier cristal ou du second. Ce n'est qu'avec l'addition des jeux des deux cristaux que le phasage réussit.

L'une des plus grosses interrogations dans l'obtention des phases dans ce cas particulier concerne le compte-rendu que SAD Phaser produit après avoir affiné les positions, occupations et facteurs d'agitation des atomes d'osmium localisés par ShelX (Sheldrick, 2008), dans la procédure de phasage.

FIGURE 36 – Analyse des données de diffraction obtenues pour les cristaux d'OrfG₆₄₋₂₀₄ comportant de l'osmium

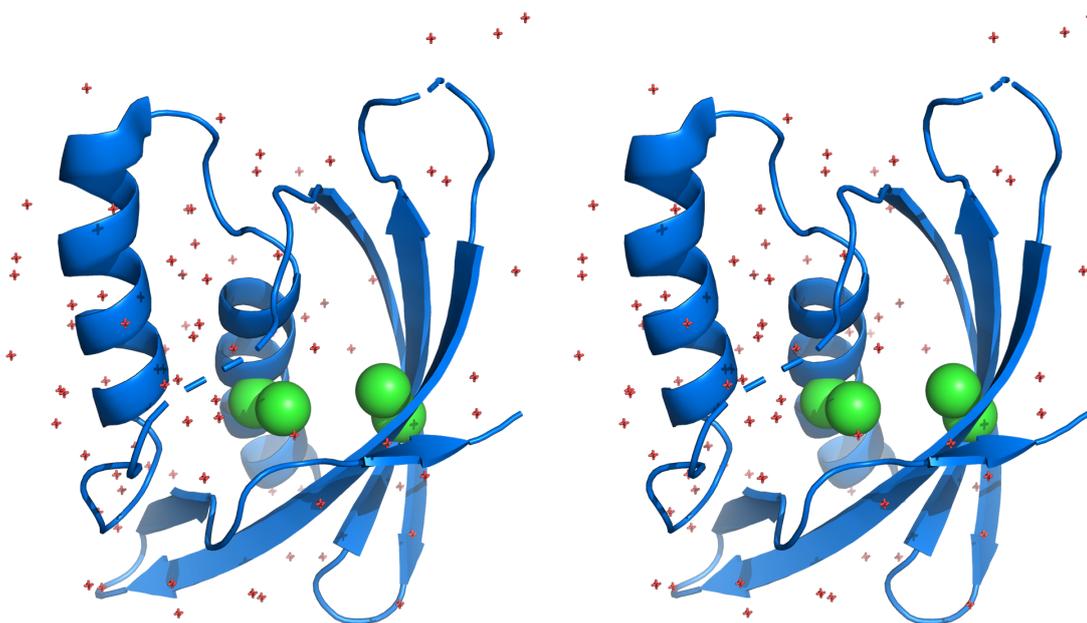
Nombre jeux	2	3	4	5	6	7
Total réflexions	546 708	853 149	1 015 357	1 176 026	1 413 770	1 637 839
Redondance	27,0	42,1	50,1	58,0	69,7	80,7
Complétude	99,9%	100%	100%	100%	100%	100%
R-factor (2 Å)	101,9%	102,7%	102,7%	102,7%	102,8%	102,6%
R-factor total	4,6%	4,8%	6,1%	5,7%	8,5%	10,7%
I/σ (2 Å)	1,97	2,8	2,81	2,8	2,79	2,77
I/σ total	45,24	54,97	55,2	59,6	58,21	55,6
CC 1/2 (2 Å)	74,3%	83,3%	83,2%	83,4%	83,1%	84,5%
CC 1/2 total	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Signal Anomal	1,128	1,208	1,21	1,253	1,238	1,207
Corr. Anomale	31%	29%	29%	31%	30%	31%
Phasage	Non	Non	Non	Non	Non	Oui

Statistiques analysant la qualité des données de diffraction obtenues selon le nombre de jeux rassemblés. Les jeux 1 à 5 sont issus d'un premier cristal, les jeux 6 et 7 d'un second. Les jeux sont ajoutés successivement à l'ensemble précédent : la colonne "2" signifie que les deux premiers jeux ont été rassemblés, la colonne "3" signifie que le jeu 3 a été ajouté à l'ensemble précédent, etc. Les statistiques sont issues d'XDS.

4 atomes d'osmium ont été détectés dans l'unité asymétrique, mais avec des occupations très faibles, c'est-à-dire inférieures à 6%. Les figures de mérite pour les deux mains possibles sont identiques, avec une valeur de 0,13 proche de signifier que les phases sont aléatoires. L'inspection des comptes-rendus de Phaser obtenus en exploitant 2 à 6 jeux de données de diffraction (c'est-à-dire les procédures ayant échoué) montre que le nombre d'osmium détectés oscille entre 3 et 4 atomes dans l'unité asymétrique, mais toujours avec des occupations faibles et des positions variables. Et surtout, les positions de ces atomes, évaluées a posteriori à la lumière de la structure affinée, sont invraisemblables car complètement enfouies dans le coeur de la protéine, et très proches les unes des autres (voir Figure 37, page 107).

Quelques vérifications supplémentaires ont été effectuées concernant la localisation et l'affinement des paramètres des 4 atomes d'osmium. Tout d'abord, la carte de densité électronique liée au signal anomal a été déterminée avec pour module de chaque paire de réflexions la différence entre F_{hkl} et F_{-h-1-l} et pour phase associée celle calculée à partir du modèle. Assez curieusement, la carte ne

FIGURE 37 – Représentation des sites osmium d'OrfG₆₄₋₂₀₄ en vue stéréoscopique



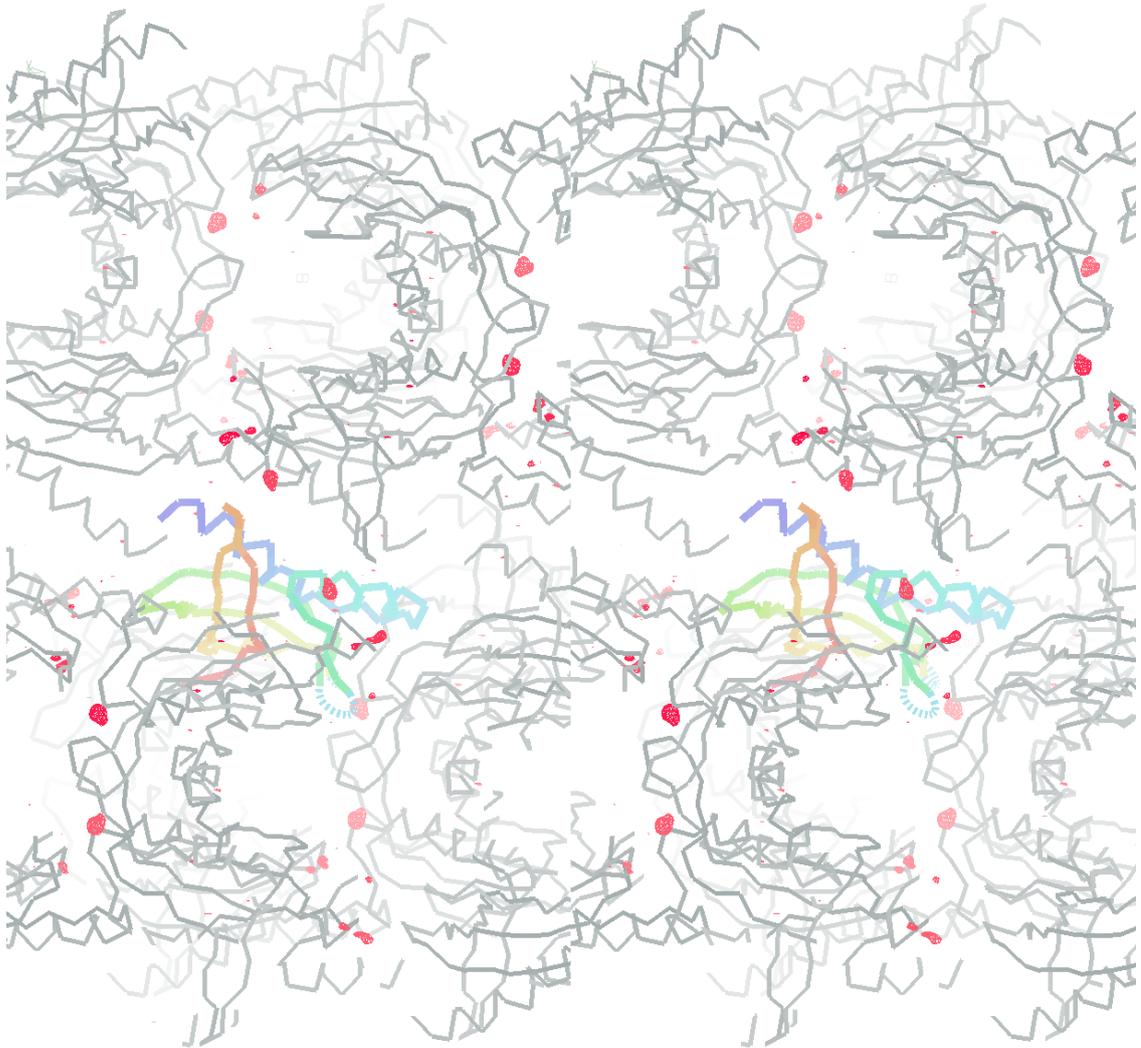
Représentation en mode "cartoon" de la protéine OrfG₆₄₋₂₀₄ en vue stéréoscopique par PyMOL. Les positions des atomes d'osmium détectés par SAD Phaser de la suite CCP4 sont symbolisées par des sphères vertes. Ces positions ne sont tout simplement pas possibles à cause de l'encombrement stérique induit par les ions chlorure qui entourent l'osmium dans le contexte du coeur hydrophobe de la protéine. Malgré ces positions que l'on peut qualifier de fictives, le phasage a fonctionné.

montre pas de signal aux positions des diffuseurs anomaux utilisées dans le phasage. Au contraire, des pics à 4σ sont présents à de nouvelles positions à la surface de la protéine, bien plus logiques et compatibles avec l'encombrement stérique du diffuseur anomal qui est entouré par 6 ions chlorures (voir Figure 38, page 109). Également, les sections de Harker au sein des cartes de Patterson anomaux ont été explorées à la recherche des signaux interprétés par ShelX (Sheldrick, 2008) pour trouver les diffuseurs anomaux. Les coordonnées atomiques des 4 atomes d'osmium sont associées soit à des pics larges couvrant les axes de symétrie de la fonction de Patterson différence anormale, soit à des zones vides de signal. Enfin, une évaluation visuelle de la qualité de la densité électronique à la fin du phasage en utilisant un contour à 3σ permet de constater une nette différence entre le phasage réussi, et tous les échecs précédents. Il est délicat de comprendre pourquoi l'utilisation des

paramètres de ces 4 atomes d'osmium, visiblement faux, permet de déterminer des phases initiales suffisamment correctes.

Quoi qu'il en soit, comme le phasage a été réussi et que la structure qui en a découlé correspondait à celle des autres protéines VirB8-like, nous n'avons pas persévéré dans l'étude de ce cas particulier. En effet, même si la compréhension et l'étude d'un cas difficile de phasage avec la méthode SAD aurait été formateur, il était plus utile à l'avancement du projet de se concentrer sur l'analyse de la structure et de la comparaison avec ses homologues dans le cadre des systèmes de sécrétions. Il existe certainement une explication à ces résultats d'apparence surprenante, mais nous en avons reporté l'analyse à plus tard.

FIGURE 38 – Représentation de la carte anomale OrfG₆₄₋₂₀₄ en vue stéréoscopique



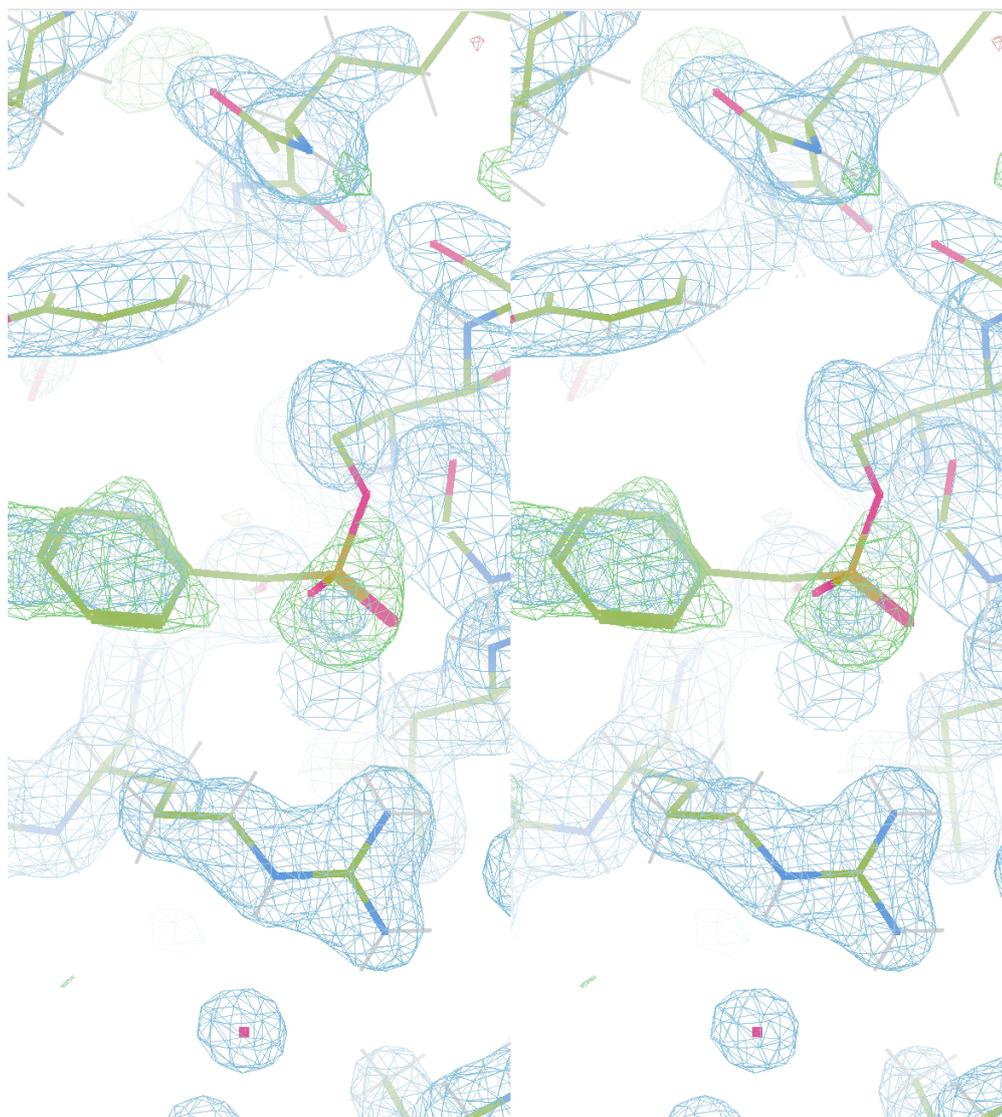
Représentation filament de la protéine OrfG₆₄₋₂₀₄ en vue stéréoscopique par Coot (Emsley *et al.*, 2010). Le monomère central est coloré en arc-en-ciel depuis son extrémité N-terminale (en bleu) jusqu'à son extrémité C-terminale (en rouge), les autres monomères générés par symétrie sont en gris. La carte anomale représentant les pics d'osmium est en rouge et est ajustée à +4 sigma pour réduire le bruit.

VII.2.3 Blob de densité

Ce point a été très rapidement discuté dans la publication. Proches de la sérine 82 et relativement enfouis dans le modèle se trouvent deux pics de densité positive dans la carte différence Fo-Fc contourée à 3σ (voir Annexe S1 de l'article). La modélisation de molécules d'eau n'a pas été probante, n'expliquant pas l'intensité du signal. Nous nous sommes intéressés à ces deux pics car ils sont localisés dans une région similaire à une poche ciblée dans la protéine VirB8 pour inhiber la conjugaison des T4SS dans les bactéries à Gram négatif (Eberhardt *et al.*, 2013). Ont été testés "manuellement" à titre de ligand potentiel les composés présents dans les solutions de préparation des cristaux comme l'ion sulfate, le glycérol ou encore le tampon tris(hydroxyméthyl)aminométhane (ou Tris). Aucun n'a été placé de façon satisfaisante dans la densité et les interactions environnantes n'étaient pas convaincantes. Suite à cela, le protocole automatisé d'identification de ligand du logiciel Phenix (Liebschner *et al.*, 2019) a été tenté, avec plus de 200 ligands testés. Aucun ligand n'a été trouvé par cette procédure automatique.

L'une des molécules utilisées dans le protocole de purification de nos collaborateurs est le PMSF ("*PhenylMethylSulfonyl Fluoride*"), afin d'inhiber les protéases à sérine lors de la purification de la protéine. En modifiant de façon spécifique la sérine catalytique de ces protéases en O-benzylsulfonyl-sérine, l'activité est totalement inhibée. L'application de cette modification sur la sérine 82 dans le modèle d'OrfG₆₄₋₂₀₄ pourrait correspondre à la densité inexpliquée (voir Figure 39, page 111). Cependant, quelques gênes stériques et interactions défavorables sont observées entre les atomes d'oxygène portés par le soufre et leur environnement. De plus, l'absence des autres résidus importants de la triade catalytique des protéases à sérine permettant à cette dernière d'être réactive rend douteux le fait que cette modification ait eu lieu spontanément. Cet essai permet cependant d'évaluer visuellement la taille du ligand recherché, et montre qu'une modification de la sérine est une éventualité plausible. Pour autant, face à l'absence d'explication et de ligand qui correspondrait à ces deux pics de densités positives, nous avons décidé de laisser cette zone vide.

FIGURE 39 – Vue stéréo de la modification potentielle de la sérine 82 en O-benzylsulfonyl-sérine chez OrfG₆₄₋₂₀₄



Vue détaillée de la région de la sérine 82. La densité électronique représentée en bleu correspond à la carte 2Fo-Fc (niveau de contour : $+1,2 \sigma$), celle en vert à la carte Fo-Fc ($+3,0 \sigma$), toutes deux calculées à une étape avancée de l'affinement du modèle comportant la sérine 82 non modifiée. Le résidu O-benzylsulfonyl-sérine (code pdb SEB) a ici été substitué à Ser82 à titre de test de son positionnement (affinement non effectué).

VII.3 Compléments de résultats obtenus après la publication sur OrfG₆₄₋₂₀₄

Pendant que l'article était en cours d'édition avant publication, une nouvelle structure de VirB8-like d'un organisme à Gram positif est parue dans la base de données des protéines. L'article associé à cette nouvelle structure n'ayant pas encore été accepté par un éditeur, une version préliminaire est disponible sur les archives bioRxiv (Jäger *et al.*, 2020). La structure du domaine soluble de la protéine PrgL retrouvée sur le plasmide pCF10 chez *Enterococcus faecalis* a été résolue et est disponible sous le code PDB 7AED. L'unité asymétrique contient 2 monomères de la protéine assemblés en un dimère inédit. Par comparaison avec les autres VirB8-like de bactéries à Gram positif, cette protéine appartient à la classe α introduite par Fercher *et al.*, 2016, en présentant un domaine transmembranaire dans les 30 premiers résidus de son extrémité N-terminale ainsi qu'un long domaine C-terminal NTF2-like qui constitue le domaine VirB8-like que nous avons étudié dans l'article. Cette structure appartient à la même classe que TraH dont la structure (code PDB : 5AIW) a été résolue par RMN. Pour rappel, TraM appartient à la classe γ et les protéines TcpC et OrfG à la classe β .

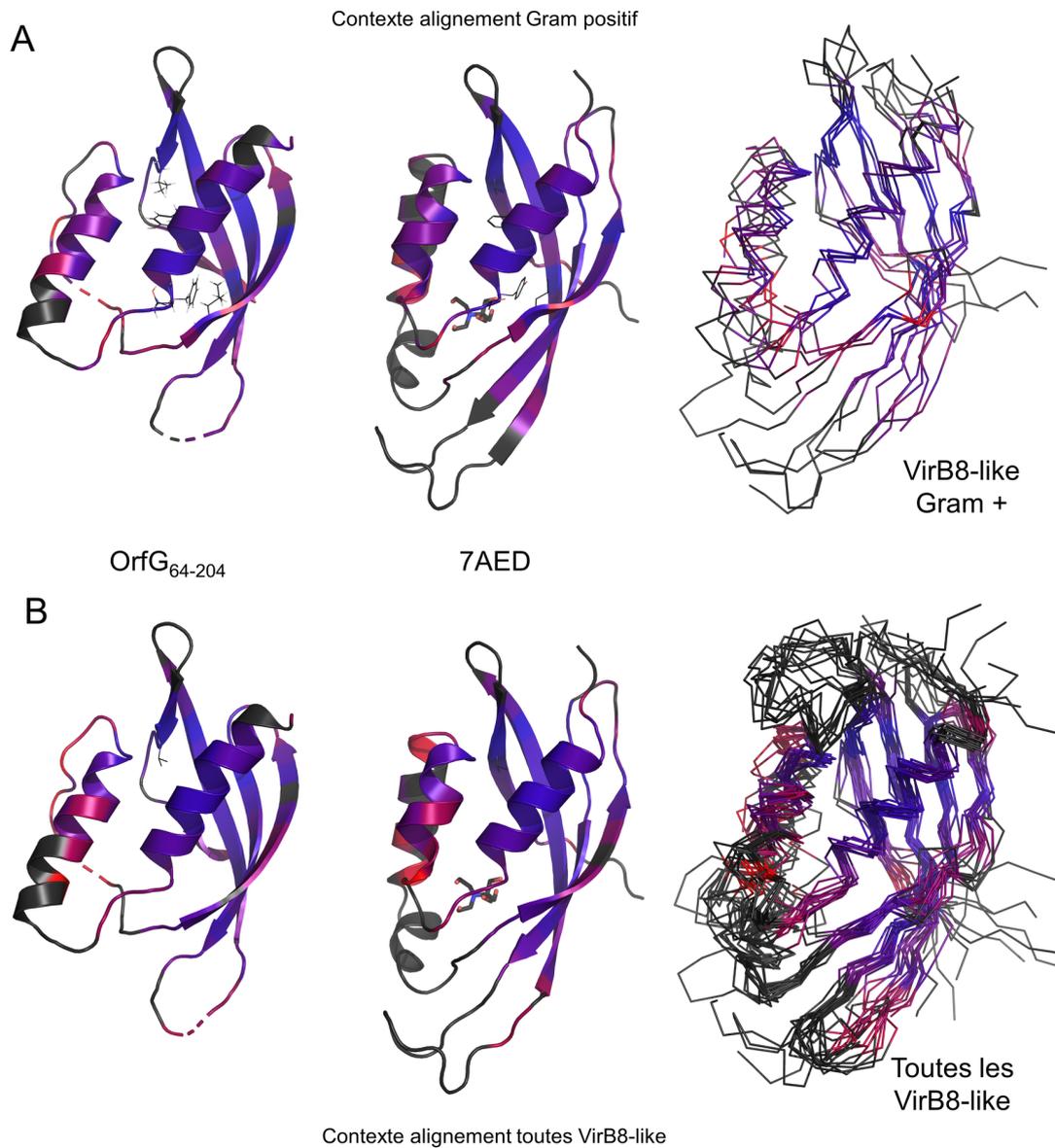
Les nouveaux alignements structuraux obtenus après l'ajout de la nouvelle structure PrgL sont disponibles dans les Figures 40 et 41 respectivement pages 114 et 115. L'arbre structural issu de mTM-align (Dong, Pan *et al.*, 2018; Dong, Peng *et al.*, 2018) a peu évolué par rapport à l'arbre présenté dans la publication. Les groupes formés précédemment restent similaires. En effet, le groupe I⁺ comportant les VirB8-like de bactéries à Gram positif des classes γ et β est toujours séparé des groupes de VirB8-like de bactéries à Gram négatif. TraH (5AIW) et PrgL (7AED) s'intercalent entre ces deux secteurs Gram positif et Gram négatif, sans former de groupe.

Concernant la structure tertiaire, PrgL possède le repliement attendu des VirB8-like, avec un enchaînement $\alpha 1-\alpha 2-\alpha 3-\beta 1-\beta 2-\beta 3-\beta 4$. Les structures de PrgL et TraH sont semblables avec des points communs qui les distinguent du groupe I⁺ associé aux bactéries à Gram positif, à l'instar du brin $\beta 4$ en une seule partie alors qu'il est scindé en deux parties $\beta 4a$ et $\beta 4b$ au sein du groupe I⁺ (voir Figure 42, page 116). Toutefois des différences claires sont à noter. L'hélice $\alpha 3$ est scindée en deux chez PrgL alors qu'elle est à peu près de même longueur mais en une seule partie chez TraH. La différence principale réside dans le brin $\beta 1$ qui est en une seule partie chez PrgL et scindé en deux chez TraH avec une boucle qui semble flexible. Cette dernière différence explique peut-être pourquoi

PrgL et TraH n'ont pas été rassemblées dans l'arbre structural. Toutefois, il est important de préciser que TraH est une structure RMN alors que toutes les autres structures proviennent de données de radiocristallographie. Les biais inhérents aux deux méthodes (et cela s'applique à toutes les structures) n'étant pas parfaitement maîtrisés, seuls davantage de structures permettront de voir si PrgL et TraH peuvent être regroupées.

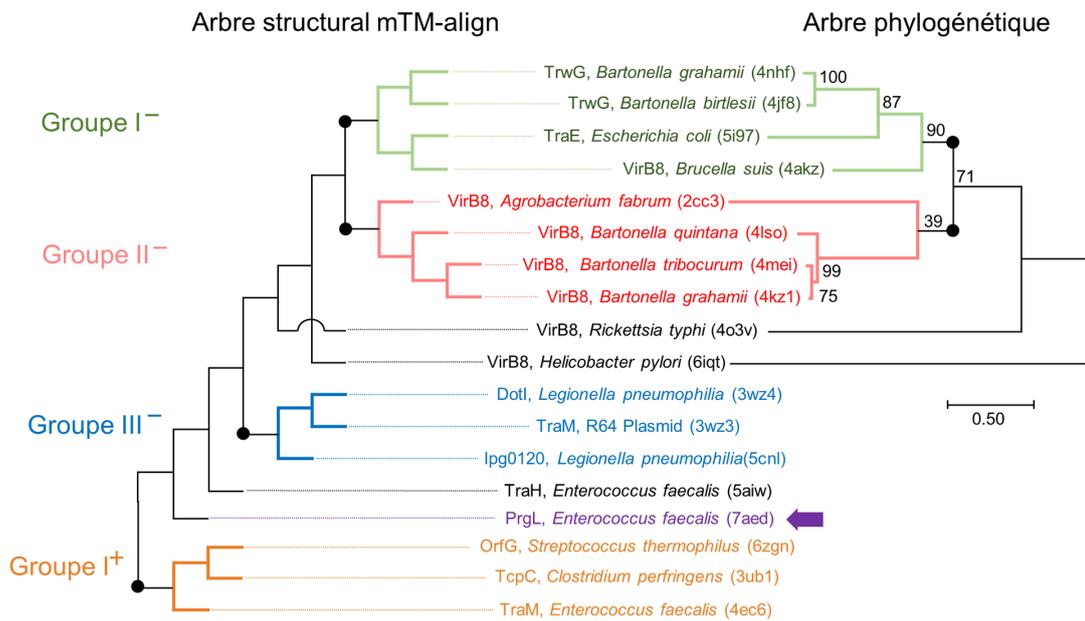
Les VirB8 Gram négatif connues à ce jour appartiennent à la classe α dont le domaine soluble est localisé dans le périplasme. Les protéines TraH et PrgL sont également deux protéines de la classe α où la position du domaine soluble reste sujette à discussion. En effet les données expérimentales obtenues pour ces deux protéines ont permis de proposer dans les deux cas que le domaine soluble serait localisé dans le cytoplasme et non dans l'environnement extra-cellulaire (Fercher *et al.*, 2016). Cette topologie différente interroge. Dans le cas de TraH, des interactions faibles avec les protéines ViB4-like (TraE) et VirD4-like (TraJ) ont été observées (Fercher *et al.*, 2016). Un rôle putatif de TraH serait le recrutement de ces facteurs essentiels au T4SS pour assembler le canal de translocation. Toutefois, TraM de la classe γ appartient au même plasmide que TraH (pIP501) et possède également un domaine soluble cytoplasmique (Goessweiner-Mohr, Grumet *et al.*, 2013). Deux questions restent ouvertes : (1) Pourquoi existe-il deux VirB8-like pour un même T4SS ? (2) Quel est le rôle exact de ces domaines cytoplasmiques ?

FIGURE 40 – Nouveaux alignements structuraux de protéines VirB8-like



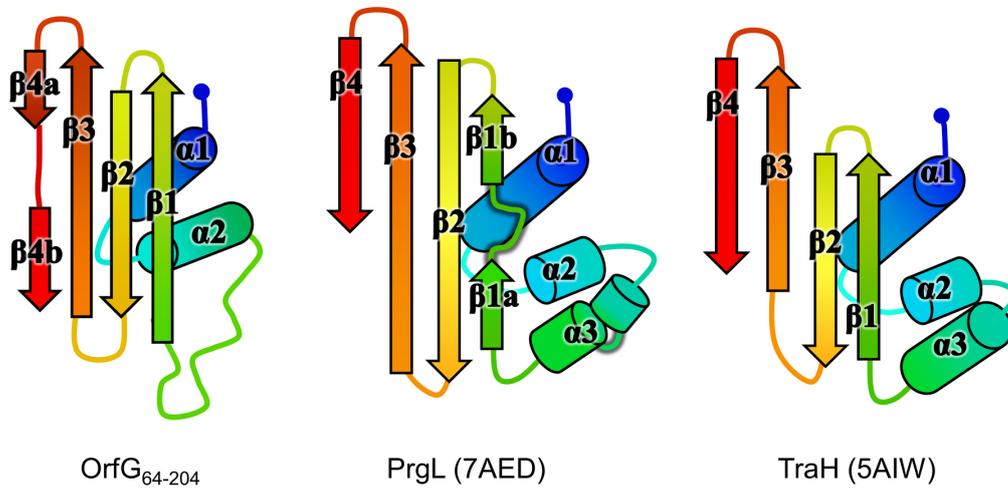
Représentation par échelle de couleur de la superposition structurale des domaines VirB8-like présentés dans la publication, ajoutée de la structure de PrgL (code PDB : 7AED). Les colorations utilisées dans A et B correspondent aux colorations appliquées par MPSA_Viewer. Pour la position dans l'espace d'un acide aminé donné, le RMSD entre tous les carbones C α des structures superposées est calculé et la position est colorée en fonction du résultat. Un RMSD de 1 Å est associé au bleu profond, tandis qu'un RMSD de 10 Å est associé au rouge sang. Un RMSD moyen sera représenté par une couleur violette (moyenne entre rouge et bleu, 5 Å). Une coloration noire/grise signifie que la distance entre les carbones alpha sur la position est supérieure à 10 Å ou que l'alignement n'a pas pu être fait sur cette position avec au moins une des structures. A. De gauche à droite, représentation d'OrfG₆₄₋₂₀₄, de PrgL et de la superposition de toutes les VirB8-like appartenant à des organismes à Gram positif. B. Similaire à A, sauf que l'alignement comprend l'ensemble des domaines VirB8-like, tout organisme confondu.

FIGURE 41 – Arbre structural des protéines VirB8-like



Arbre structural et phylogénétique des protéines VirB8-like, adapté des résultats du serveur mTM-align en prenant en compte la nouvelle structure de PrgL (code PDB : 7AED), par rapport à l'alignement initial de Cappele *et al.*, 2021. Cette nouvelle structure est colorée en violet.

FIGURE 42 – Structures secondaires de OrfG₆₄₋₂₀₄ , PrgL et TraH

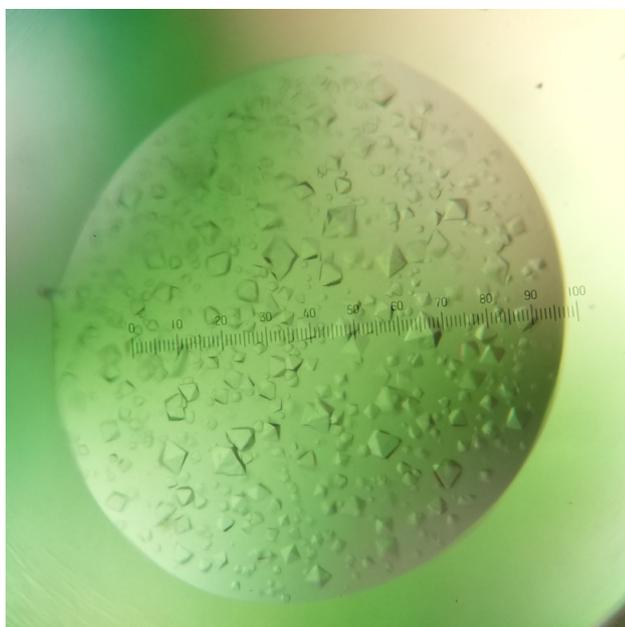


Représentation schématique des structures secondaires des domaines VirB8-like des protéines PrgL (code PDB : 7AED), TraH (code PDB : 5AIW) et OrfG₆₄₋₂₀₄ (code PDB : 6ZGN). La coloration en arc-en-ciel est appliquée depuis l'extrémité N-terminale (en bleu) vers l'extrémité C-terminale (en rouge). Les brins sont symbolisés par des flèches pleines, les hélices par des cylindres pleins. Les structures secondaires sont numérotées selon leur ordre depuis l'extrémité N-terminale jusqu'à l'extrémité C-terminale.

VII.4 Étude structurale du domaine soluble OrfG₆₄₋₃₃₁

Le domaine soluble OrfG₆₄₋₃₃₁, rappelons-le, est le domaine constitué des résidus 64 à 331, correspondant à la protéine OrfG dépourvue de son extrémité N-terminale et donc de son domaine transmembranaire (TMD). OrfG₆₄₋₃₃₁ comporte deux domaines : le domaine OrfG₆₄₋₂₀₄ dont la structure a initialement été résolue et discutée précédemment, et le domaine C-terminal comprenant les résidus 223 à 331. Les deux domaines sont connectés par une chaîne de 19 résidus. OrfG₆₄₋₃₃₁, notre cible initiale, a pendant longtemps posé des difficultés pour obtenir des cristaux avec un bon pouvoir diffractant. Assez rapidement et dans de nombreuses conditions, des cristaux réguliers d'une taille suffisante ont été obtenus. Cependant, les tests de diffraction au laboratoire comme au synchrotron montraient systématiquement un faible pouvoir de diffraction avec une résolution maximale ne dépassant pas les 6 Å. Après 3 mois, de nouveaux cristaux qui semblaient identiques ont été observés dans les mêmes conditions d'obtention, mais cette fois-ci diffractant jusqu'à 1,8 Å (voir Figure 43 page 117). Deux jeux de données ont pu être collectés sur la ligne Proxima2 du synchrotron SOLEIL et ont été exploités par la suite.

FIGURE 43 – Cristaux de la construction OrfG₆₄₋₃₃₁



Cristaux de la construction OrfG₆₄₋₃₃₁ dans la condition contenant 2,4 M de malonate de sodium dibasique monohydraté.

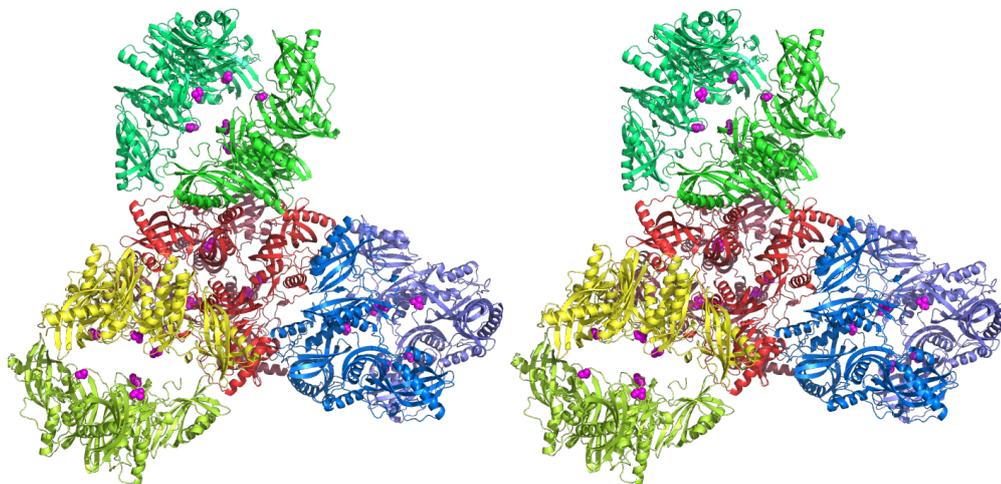
VII.4.1 La structure d'OrfG₆₄₋₃₃₁

En utilisant le domaine central OrfG₆₄₋₂₀₄ (code PDB 6ZGN) ainsi que le domaine C-terminal de TcpC (code PDB 3UB1) comme modèles pour le remplacement moléculaire, la structure d'OrfG₆₄₋₃₃₁ a été résolue facilement, et est présentée dans la Figure 45 page 120. Nous n'avons pas tenté le remplacement moléculaire avec un monomère complet de TcpC : d'une part les deux domaines sont séparés par un linker d'une vingtaine d'acides aminés dont la fixité aurait empêcher les domaines de s'orienter correctement si nécessaire ; d'autre part la structure d'OrfG₆₄₋₂₀₄ précédemment résolue donnait un point de départ a priori beaucoup plus juste pour ce domaine. Le cristal appartient au groupe d'espace cubique P4₁32. Un seul monomère d'OrfG₆₄₋₃₃₁ est présent dans l'unité asymétrique. Lorsque l'on en génère les équivalents cristallographiques, on observe un assemblage plausible en trimères dont les trois TMD seraient localisés du même côté du multimère. En revanche, l'empilement cristallin ne révèle pas de super-assemblage vraisemblable. En effet, les associations de trimères observées ne semblent pas avoir d'intérêt biologique puisqu'elles ne respectent pas la caractéristique membranaire de la protéine : elles correspondent par exemple à des trimères orientés tête-bêche (Voir Figure 44, 119).

La structure du monomère d'OrfG₆₄₋₃₃₁ contient les acides aminés 87 à 328 incluant le domaine central (94-204) et le domaine C-terminal (223-328). Le domaine central (94-204) conserve la topologie observée pour le domaine central seul (i.e. OrfG₆₄₋₂₀₄), soit l'enchaînement de structures secondaires $\alpha 1$ - $\alpha 2$ - $\beta 1$ - $\beta 2$ - $\beta 3$ - $\beta 4$, avec le brin $\beta 4$ scindé en deux brins $\beta 4a$ et $\beta 4b$. Le domaine C-terminal est également un domaine VirB8-like. Il adopte également la topologie $\alpha 1$ - $\alpha 2$ - $\beta 1$ - $\beta 2$ - $\beta 3$ - $\beta 4$. Le brin $\beta 4$ est continu, et l'hélice $\alpha 2$ est plus courte par rapport à celle observée dans le domaine central. Comme dans OrfG₆₄₋₂₀₄, l'extrémité N-terminale de 64 à 86, ainsi que les résidus 168 à 176 n'ont pas été modélisés par manque de densité électronique. En revanche, la boucle entre l'hélice $\alpha 2$ et le brin $\beta 1$, en partie absente de la structure d'OrfG₆₄₋₂₀₄, est présente dans la structure d'OrfG₆₄₋₃₃₁ (voir ci-dessous).

OrfG et TcpC sont deux protéines appartenant à la classe γ (Fercher *et al.*, 2016) et leurs domaines solubles cristallisent avec le même arrangement trimérique. Toutefois les orientations entre les domaines centraux et C-terminaux dans chacune des structures sont suffisamment distinctes pour rendre médiocre la superposition globale des domaines solubles d'OrfG et de TcpC. Une comparaison

FIGURE 44 – Représentation stéréo de l’assemblage dans le cristal d’OrfG₆₄₋₃₃₁

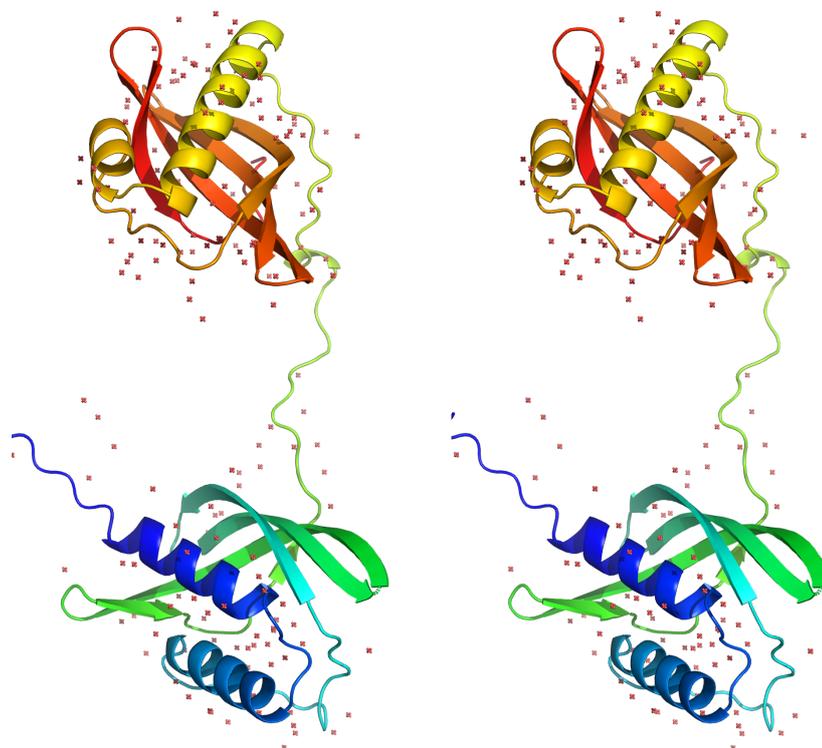


Visualisation stéréoscopique réalisée avec PyMOL. OrfG₆₄₋₃₃₁ s’assemble en trimères dans le cristal, et chaque trimère est représenté avec une couleur différente. L’extrémité N-terminale de chaque monomère, avant laquelle se situerait la partie transmembranaire, est montrée en sphères magenta. Des paires de trimères orientés tête-bêche sont ainsi visibles, ici représentées avec des couleurs voisines associées à chacun des trimères qui les composent.

des domaines séparés est ainsi plus pertinente.

Le domaine central (94-204) observé dans la structure d’OrfG₆₄₋₃₃₁ est assez logiquement très similaire à celui d’OrfG₆₄₋₂₀₄ déjà résolu, mais certaines différences sont à noter. En effet, l’écart quadratique moyen (RMSD) résultant de la superposition du domaine central (94-204) dans OrfG₆₄₋₃₃₁ avec le domaine central seul (OrfG₆₄₋₂₀₄) est de 0,71 Å (87 résidus alignés). Les boucles supposées flexibles et non modélisées dans OrfG₆₄₋₂₀₄ le sont dans la présente structure du domaine central, à l’exception de deux régions. Le nouvel empilement cristallin entraîne une stabilisation et/ou un changement de conformation des régions 165-168 et 175-179, qui allonge de quatre résidus à la fois la fin du brin β 2 et le début du brin β 3 (voir Figure 46 page 121). La proximité du domaine C-terminal d’un autre monomère permet la formation de nouvelles interactions. On citera par exemple la présence de deux nouvelles liaisons hydrogène réalisées par la chaîne latérale du glutamate 166 (extrémité C-terminale du brin β 2, domaine central), avec le groupement amide de la glycine 295 d’une part, et avec le groupement hydroxyle de la thréonine 296 d’autre part. Ces résidus 295 et 296 appartiennent

FIGURE 45 – Monomère d'OrfG₆₄₋₃₃₁

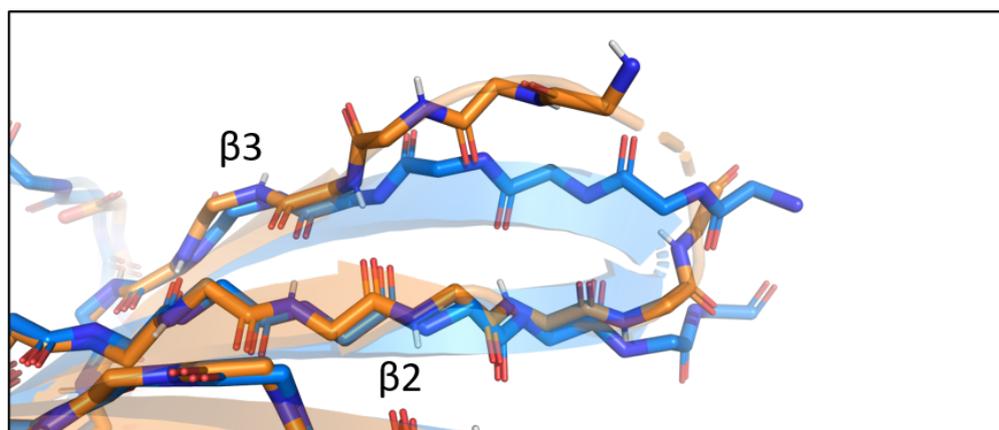


Représentation de la structure d'OrfG₆₄₋₃₃₁ en stéréo. La structure est colorée en arc-en-ciel du bleu pour l'extrémité N-terminale au rouge à l'extrémité C-terminale. Le domaine central apparaît en teinte bleutée/verte et le domaine C-terminal apparaît orangé/rouge.

à la boucle entre $\beta 2$ et $\beta 3$ du domaine C-terminal.

L'une des originalités de cette nouvelle structure est la présence d'un deuxième domaine de type VirB8-like (résidus 228 à 319), de la même manière que chez TcpC (résidus 250 à 344). En effet, les deux domaines C-terminaux possèdent deux hélices $\alpha 1$ et $\alpha 2$, suivies de 4 brins $\beta 1$ à $\beta 4$ formant un feuillet anti-parallèle. Dans la Figure 48 page 123 sont présentés les différents domaines VirB8-like de ces deux protéines dans une orientation conservée en fonction de l'alignement réalisé. La Figure 49 page 123 présente la mesure des écarts entre ces différents domaines. On y constate que les domaines centraux OrfG₆₄₋₂₀₄ et TcpC₁₀₄₋₂₂₉ se superposent très bien, avec un RMSD entre eux de 1,27 Å, comme nous avons déjà pu le constater dans la partie dédiée à OrfG₆₄₋₂₀₄. De la même manière, les domaines C-terminaux des deux protéines (OrfG₂₂₃₋₃₃₁ et TcpC₂₃₉₋₃₅₆) sont

FIGURE 46 – Vue détaillée des brins $\beta 2$ et $\beta 3$ du domaine central d'OrfG



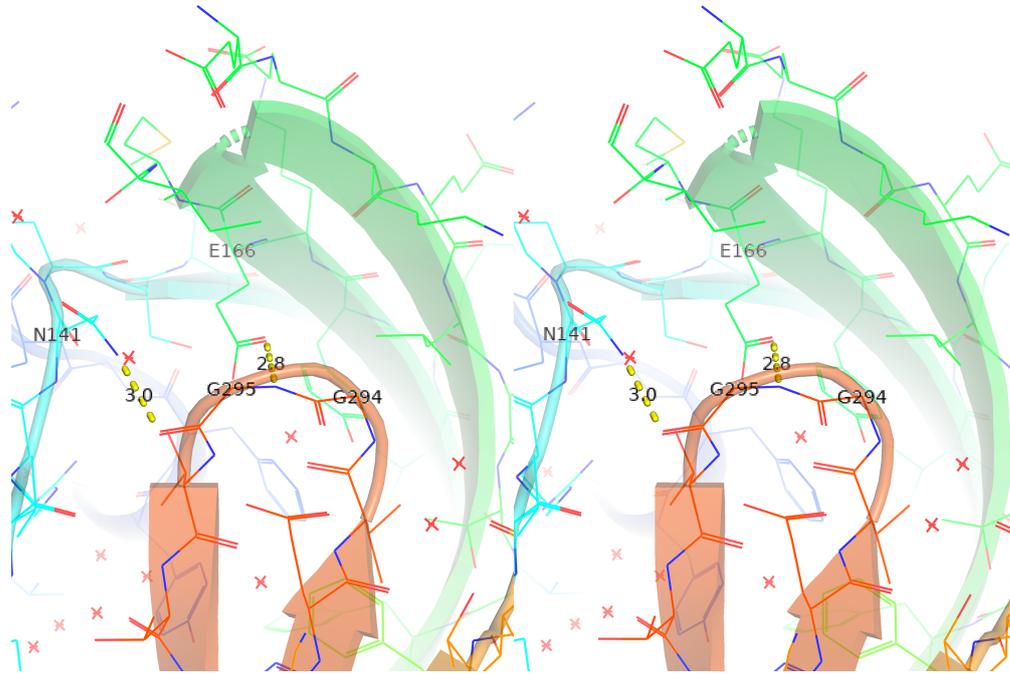
Comparaison de la région de connexion des brins $\beta 2$ et $\beta 3$ du domaine central des structures OrfG₆₄₋₃₃₁ (en bleu) et OrfG₆₄₋₂₀₄ (en orange). Les chaînes principales sont représentées en bâtons, les chaînes latérales sont masquées et les structures secondaires en ruban sont montrées en transparence en suivant la même coloration. Les acides aminés n'ayant pas de densité suffisante ne sont pas modélisés et sont donc absents des structures. Ces régions sont figurées par des traits pointillés (résidus 169-174 en bleu chez OrfG₆₄₋₃₃₁ et 169-175 en orange chez OrfG₆₄₋₂₀₄).

très similaires et présentent un RMSD de seulement 1,13 Å. Dans ces superpositions l'alignement des feuillettes est presque parfait ainsi que celui de l'hélice $\alpha 1$. Des différences plus importantes apparaissent dans les comparaisons entre les domaines centraux et C-terminaux où les RMSD varient de 2,1 à 4,2 Å. Ces superpositions "croisées" sont similaires. Quelle que soit la superposition, aucune structure secondaire n'apparaît mieux alignée qu'une autre. Les variations des valeurs des RMSD d'une superposition croisée à l'autre semblent plutôt corrélées au nombre de résidus superposés (Figure 49 page 123).

Avant de comparer plus en détail les différents domaines VirB8-like en incluant les domaines centraux dans l'arbre structural et en les analysant grâce à mTM-align, notons que dans la structure du monomère d'OrfG₆₄₋₃₃₁ et plus particulièrement dans le domaine central, un blob de densité moins intense est retrouvé à la même position que pour OrfG₆₄₋₂₀₄ . La probabilité que ce blob présent dans deux structures similaires ne soit qu'un artefact diminue, mais nous n'avons pas non plus réussi à y placer de molécule permettant de l'expliquer.

Lorsque l'on inclut dans la superposition structurale des domaines VirB8-like les domaines centraux et C-terminaux séparés des protéines OrfG₆₄₋₃₃₁ et TcpC, on obtient l'arbre présenté dans la figure 50 page 124. Un changement mineur par rapport à l'arbre précédent est constaté. Le do-

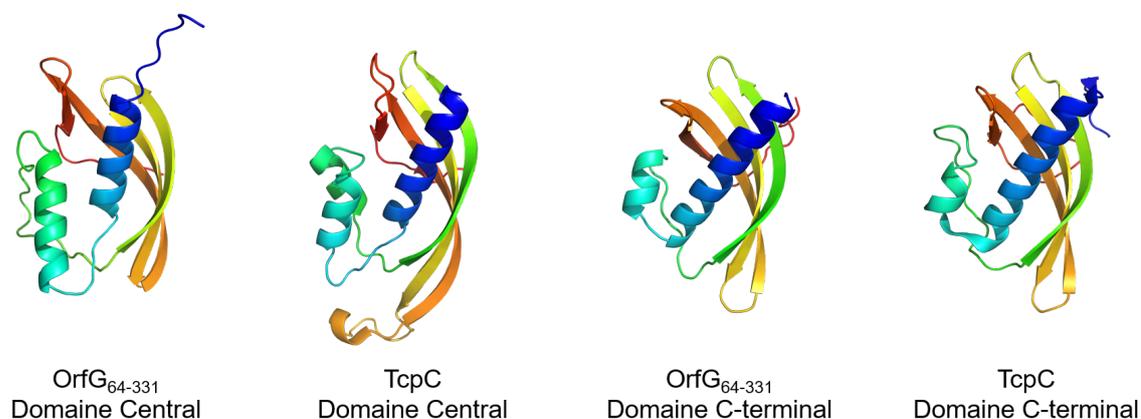
FIGURE 47 – Vue stéréoscopique des brins $\beta 2$ et $\beta 3$ du domaine central d'OrfG₆₄₋₃₃₁



Représentation des brins $\beta 2$ et $\beta 3$ du domaine central en vert et de la boucle entre les brins $\beta 2$ et $\beta 3$ du domaine C-terminal. La boucle entre l'hélice $\alpha 2$ et le brin $\beta 1$ du domaine central est représentée en bleu.

maine central d'OrfG₆₄₋₃₃₁ et le domaine C-terminal de TraM (code PDB 4EC6) sont maintenant inversés, de sorte que le score d'alignement de paire entre TraM et le domaine central de TcpC est meilleur qu'avec le domaine central d'OrfG₆₄₋₃₃₁. Les différences structurales entre le domaine central seul (OrfG₆₄₋₂₀₄) et son équivalent dans OrfG₆₄₋₃₃₁ pourraient en être une des causes, résultant par exemple de l'allongement des brins $\beta 2$ et $\beta 3$. Cette différence observée au sein de l'arbre est accessoire car elle ne remet pas en cause les groupes définis auparavant. Pour finir, les domaines C-terminaux d'OrfG₆₄₋₃₃₁ et TcpC n'intègrent pas l'arbre précédent dans le sens où ils ne sont pas affiliés aux autres groupes. Ainsi les groupes définis auparavant possèdent en commun des caractéristiques structurales non partagés avec les domaines C-terminaux d'OrfG₆₄₋₃₃₁ et TcpC. Ces deux domaines pourraient appartenir à un nouveau groupe de VirB8-like spécifique aux bactéries à Gram positif, et plus particulièrement à la classe β .

FIGURE 48 – Comparaison des différents domaines VirB8-like d'OrfG₆₄₋₃₃₁ et de TcpC



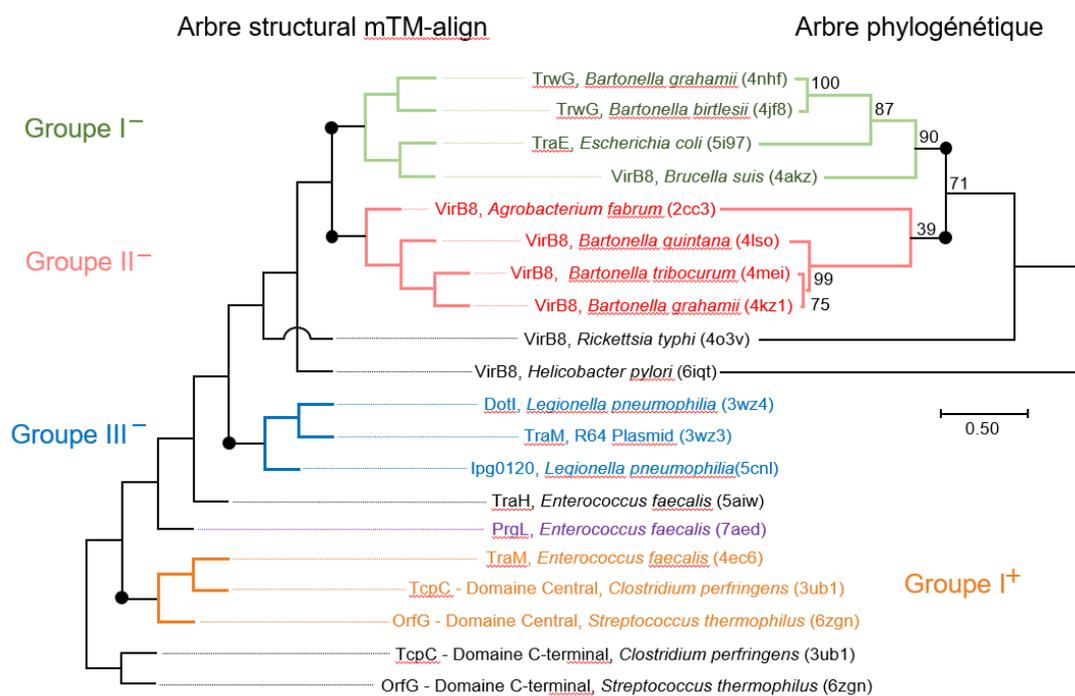
Représentation des domaines VirB8-like isolés des protéines OrfG₆₄₋₃₃₁ et TcpC en arc-en-ciel. La coloration débute à l'extrémité N-terminale en bleu jusqu'à l'extrémité C-terminale en rouge. De gauche à droite sont représentés les domaines centraux d'OrfG₆₄₋₃₃₁ (présent dans OrfG₆₄₋₂₀₄) et TcpC, puis leurs domaines C-terminaux. Toutes les structures sont alignées sur le domaine C-terminal d'OrfG, l'orientation de ces représentations est donc conservée.

FIGURE 49 – Tableau comparatif des RMSD entre les différents domaines d'OrfG₆₄₋₃₃₁ et TcpC

RMSD	OrfG-Central (117aa)	OrfG-Cterminal (101aa)	TcpC-Central (125aa)	TcpC-Cterminal (117aa)
OrfG-Central (117aa)	-	4,103	1,176	3,991
OrfG-Cterminal (101aa)	88	-	2,541	1,061
TcpC-Central (125aa)	69	61	-	2,123
TcpC-Cterminal (117aa)	90	88	58	-

La partie supérieure droite du tableau indique la valeur de RMSD calculé entre les C α des deux structures grâce à l'algorithme SUPER intégré dans PyMOL. La partie inférieure gauche du tableau indique le nombre de C α pris en compte dans le calcul du RMSD correspondant. En bleu est indiqué le RMSD entre les domaines centraux, et en rouge est indiqué le RMSD entre les domaines C-terminaux. La longueur de chaque domaine est précisée à droite de son nom (nombre d'acides aminés).

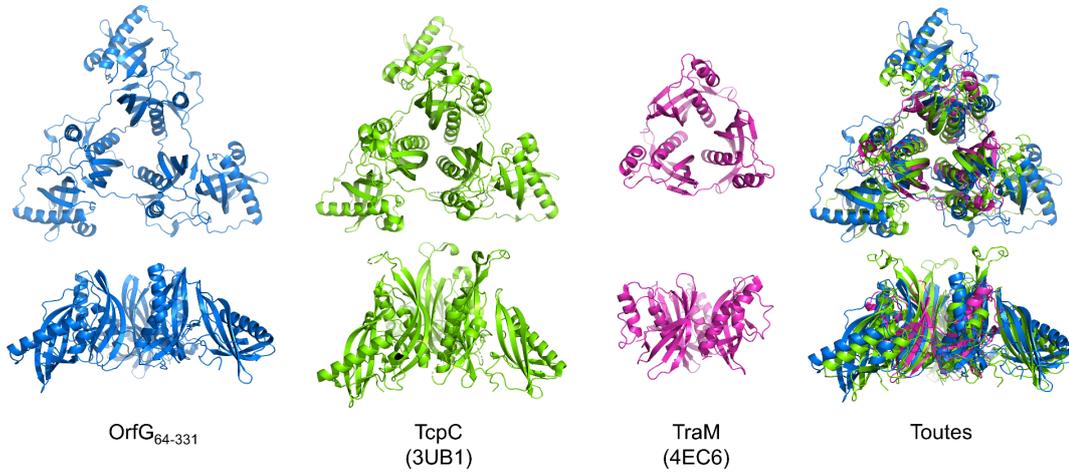
FIGURE 50 – Arbre structural incluant les domaines C-terminaux d'OrfG₆₄₋₃₃₁ et TcpC



Nouvel arbre structural et phylogénétique adapté des résultats du serveur mTM-align, en prenant en compte la nouvelle structure du domaine central observée dans OrfG₆₄₋₃₃₁, son domaine C-terminal, ainsi que celui de TcpC.

VII.4.2 Oligomérisation

FIGURE 51 – Structures trimériques d'OrfG₆₄₋₃₃₁ et ses homologues

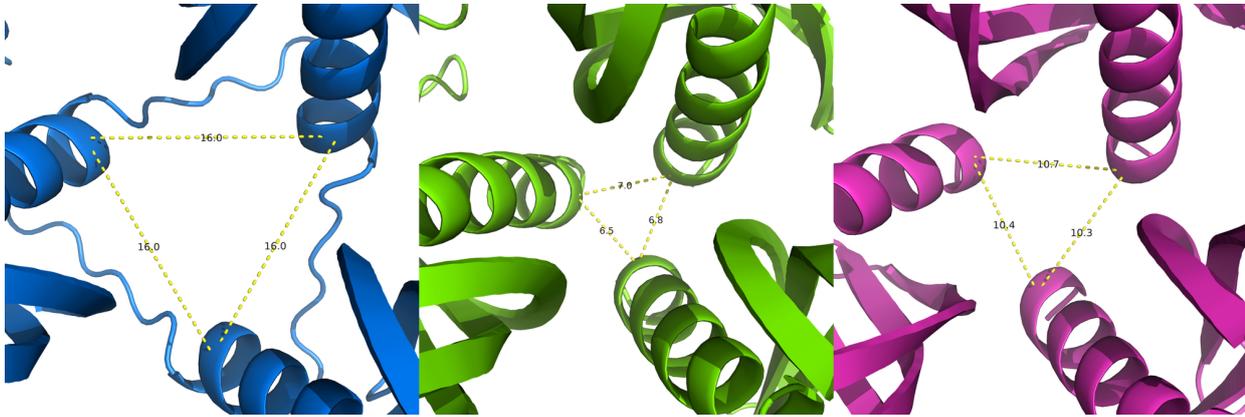


Représentation schématique de l'organisation trimérique des différentes VirB8-like Gram positif. De gauche à droite : OrfG₆₄₋₃₃₁ est en bleu, TcpC est en vert, TraM est en rose et en dernière position se trouve une superposition structurale des 3 molécules. Les trimères sont représentés perpendiculairement (en haut) ou parallèlement (en bas) à l'axe 3 reliant leurs monomères.

Concernant la structure quaternaire de la protéine, les monomères d'OrfG₆₄₋₃₃₁ s'enroulent ensemble pour former le même trimère que TcpC (code PDB : 3UB1). Dans cet assemblage caractéristique, les domaines centraux sont proches de l'axe ternaire et chaque domaine C-terminal s'enroule autour du domaine central d'un autre monomère grâce au grand linker qui relie les domaines. Sans avoir de domaine C-terminal comme TcpC et OrfG₆₄₋₃₃₁, la première structure de VirB8-like Gram positif publiée, TraM, présente également un trimère similaire. A noter que pour TcpC et TraM, l'assemblage trimérique est directement observé dans l'unité asymétrique, sans recours aux symétries cristallographiques. Les trimères sont représentés alignés dans la Figure 51 page 125. Cet assemblage très particulier présent dans la structure cristallographique de TcpC n'a étonnamment pas été discuté dans la publication associée, malgré sa similitude avec l'assemblage formé par TraM. Cela s'explique probablement par l'état monomérique observé en solution pour les domaines solubles de TcpC et TraM. La structure cristallographique d'OrfG₆₄₋₃₃₁ vient renforcer l'idée que cette oligomérisation a un sens biologique. Ce même argument avait été utilisé par Bailey et collaborateurs pour proposer un rôle biologique à la forme dimérique de VirB8 du T4SS d'*Agrobacterium tumefaciens* (Bailey *et al.*, 2006). En effet le même dimère avait été observé dans la structure cristallographique

de VirB8 du T4SS de *Brucella suis* (Terradot *et al.*, 2005). Concernant le trimère d'OrfG, une analyse détaillée et comparative des différentes interfaces entre les domaines sera faite dans la partie VII.4.3, page 128.

FIGURE 52 – Région centrale du trimère d'OrfG₆₄₋₃₃₁ et de ces homologues



Représentation schématique de l'organisation trimérique des différentes VirB8-like Gram positif avec zoom sur le centre du trimère, au niveau du canal potentiel formé par les hélices $\alpha 1$ des différents domaines centraux des trimères. De gauche à droite : OrfG, TcpC et TraM avec respectivement des distances mesurées d'environ 16, 7 et 10 Å entre les hélices.

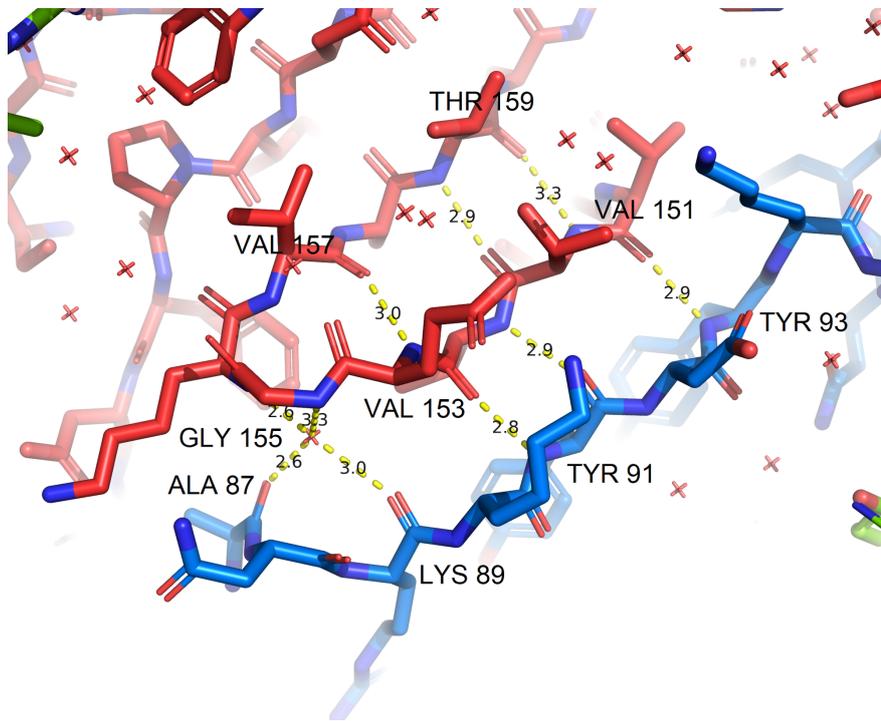
En comparant les trimères, on se rend compte qu'OrfG₆₄₋₃₃₁ adopte une conformation un peu plus étendue que TcpC, ce qui a peut-être pour conséquence la présence d'une ouverture légèrement plus grande au centre de la protéine. En effet, les distances entre les hélices $\alpha 1$ des domaines centraux sont d'environ 16 Å chez OrfG₆₄₋₃₃₁ contre seulement 7 Å chez TcpC et 10 Å chez TraM (voir Figure 52 page 126). Ces différentes conformations au niveau des centres de ces multimères pourraient constituer des canaux, permettant à un substrat protéique de traverser dans une conformation ouverte, ou bien de bloquer le passage dans une conformation fermée. En effet, une publication récente (Miletic *et al.*, 2021) présente une structure dans le contexte du système de sécrétion de type III contenant le substrat protéique déplié en cours de transfert. La mesure de l'ouverture du canal entre les C α des acides aminés en bout d'hélice révèle une ouverture de 15 à 18 Å permettant au substrat protéique sous forme dépliée de traverser le canal. Dans le cas du T4SS Gram négatif, il a été montré récemment que la relaxase, couplée à l'ADN simple brin, est dépliée lors de la translocation (Trokter & Waksman, 2018). En mesurant sur des protéines les sites de fixation à l'ADN simple ou double brin, on se rend compte que l'ouverture respective est d'environ 18 Å ou

30 Å entre les deux points les plus proches du squelette peptidique. Dans le cas d'OrfG, en l'état, l'ouverture observée n'est pas compatible avec l'ADN simple brin (Dickey *et al.*, 2013; McGrath *et al.*, 2018).

Il est également possible d'imaginer que le trimère observé pour OrfG (ou TcpC) ne soit qu'une forme fermée du canal, tandis qu'une oligomérisation supérieure telle qu'un hexamère permettrait d'ouvrir considérablement ce canal pour que traversent facilement les substrats protéiques et ADN, ou encore que la forme trimérique soit spécifique au passage du substrat protéique et qu'une hypothétique oligomérisation supérieure se produise par la suite pour permettre le transfert de l'ADN simple brin. Il est important de noter que cette hypothèse est potentiellement valable pour OrfG et TcpC car pour ces protéines les domaines transmembranaires ne s'associent pas pour former un *coiled-coil* (superhélice formée d'hélices enroulées en spirale) qui condamnerait la possibilité d'un tel canal. Une prédiction et une simulation ont été réalisées dans le cas de TraM (Goessweiner-Mohr, Grumet *et al.*, 2013), démontrant la possibilité que les hélices transmembranaires s'associent ensemble, favorisant le trimère. De plus, TraM possédant une organisation différente avec le domaine transmembranaire au milieu de ses deux domaines solubles, il est possible que les T4SS des organismes qui hébergent ces protéines et éléments mobiles soient complètement différents.

Dans la structure d'OrfG₆₄₋₃₃₁, une interaction quaternaire est intéressante à noter au centre du trimère et entre les domaines centraux. L'extrémité N-terminale d'un domaine central d'OrfG₆₄₋₃₃₁ interagit avec le brin $\beta 1$ du domaine central voisin de sorte que ce segment N-terminal et les brins $\beta 1$ et $\beta 2$ forment un court feuillet à 3 brins (voir Figure 53 page 128). Cette interaction également inédite au sein des trimères VirB8-like induirait, en admettant qu'elle ait une certaine vérité biologique, que le domaine transmembranaire ne soit pas situé en dessous du domaine central du même monomère, mais en dessous du domaine central du monomère voisin.

FIGURE 53 – Interaction du segment N-terminal d'un monomère d'OrfG₆₄₋₃₃₁ avec les brins $\beta 1$ et $\beta 2$ du monomère voisin



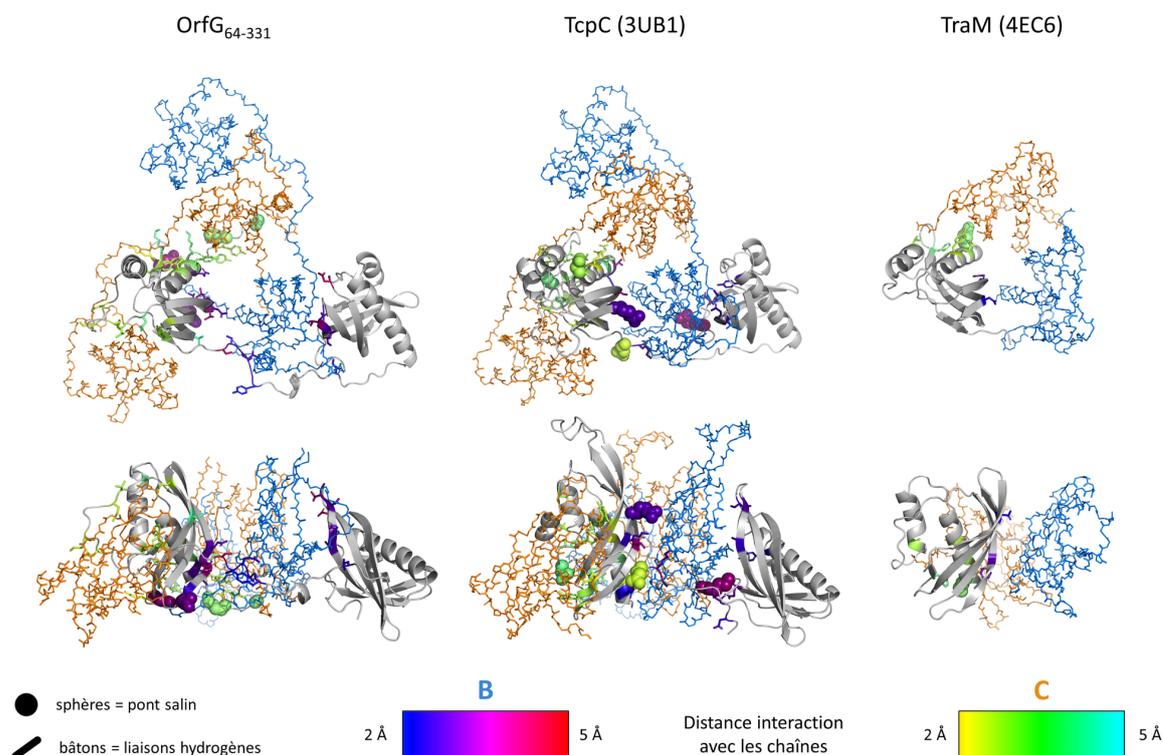
Représentation de l'interaction du segment N-terminal de l'un des monomères d'OrfG₆₄₋₃₃₁ modélisé en bleu avec les brins $\beta 1$ et $\beta 2$ d'un second monomère en rouge. Les distances indiquées représentent des liaisons hydrogènes et sont en Å (les hydrogènes ne sont pas représentés).

VII.4.3 Étude de l'interface

Au niveau de l'interface entre les différents monomères d'OrfG₆₄₋₃₃₁, le logiciel PDBePISA (Krissinel & Henrick, 2007) détecte plus d'une trentaine de résidus impliqués dans des liaisons hydrogène et 4 résidus impliqués dans des ponts salins, tout domaine confondu. Un domaine central donné est en interaction avec les deux autres domaines centraux et l'un des domaines C-terminaux. Pour représenter au mieux ces interactions au sein des trimères des VirB8-like Gram positif, la Figure 54 est présentée en page 129 et est accompagnée des annexes S7 page 211, S8 page 211 et S9 page 212 permettant de visualiser plus précisément les interfaces.

Pour OrfG₆₄₋₃₃₁, la représentation de la figure 54 page 129 montre que le domaine central du monomère A (en "cartoon" gris) interagit fortement avec les deux domaines de la chaîne C, ainsi qu'avec le domaine central de la chaîne B. Il résulte de cet agencement global que l'extrémité N-

FIGURE 54 – Visualisation des interfaces hydrophiles des trimères



Représentation PyMOL des trimères de VirB8-like Gram positif incluant OrfG₆₄₋₃₃₁, TcpC et TraM (de gauche à droite). La chaîne A est représentée en cartoon, avec une coloration grise. Les chaînes latérales des acides aminés représentées en bâtonnets sont impliquées dans des liaisons hydrogènes, tandis que celles représentées en sphère sont impliquées dans des ponts salins. Comme indiqué dans la légende, la coloration de ces chaînes latérales dépend de deux facteurs : 1. la distance de l'interaction ; 2. avec quelle chaîne elles interagissent. Les interactions entre les chaînes A et B sont colorées en suivant un gradient bleu-rouge de 2 à 5 Å de distance, et les interactions entre les chaînes A et C sont colorées en suivant un gradient jaune-cyan de 2 à 5 Å de distance.

Remarque : cette visualisation étant très difficile à représenter en deux dimensions, les figures disponibles en annexe contiennent des vues stéréoscopiques ainsi que des QRcodes vers des vidéos également en stéréo.

terminale de chaque domaine central interagit avec le monomère "à sa gauche" (comme déjà illustré à la Figure S7, 211) ; simultanément, le linker et le domaine C-terminal s'enroulent quant à eux autour du domaine central du monomère "à sa droite". Le domaine C-terminal n'interagit qu'avec un seul domaine central, celui qui s'est inséré entre son propre domaine central et lui-même.

TcpC reprend dans les grandes lignes cette même conformation, plus serrée autour de l'axe 3 non cristallographique reliant les monomères, mais qui s'étend plus en hauteur le long de ce même axe. Entre les différents domaines, PDBePISA (Krissinel & Henrick, 2007) détecte également une trentaine de résidus impliqués dans des liaisons hydrogène et 4 ponts salins. Toutefois, ces interac-

tions ne semblent pas être similaires à celles d'OrfG₆₄₋₃₃₁ tant leurs positions et leurs natures sont différentes. Ce jeu d'interactions a pour origine (ou pour conséquence) une inclinaison différente du domaine central par rapport à l'axe du trimère. En particulier, les axes d'allongement des hélices $\alpha 1$ des domaines centraux de TcpC semblent se diriger vers un même point sur l'axe du trimère, tandis que ceux des hélices $\alpha 1$ des domaines centraux d'OrfG₆₄₋₃₃₁ sont plutôt dirigées vers le domaine central du monomère voisin, créant une interaction avec un de ses brins. On remarque d'ailleurs l'absence du segment N-terminal de TcpC qui n'apparaît pas dans la structure, et n'est donc vraisemblablement pas en interaction avec la partie inférieure du domaine central voisin comme dans OrfG₆₄₋₃₃₁.

Cette notion est assez importante puisque dans le cas d'OrfG₆₄₋₃₃₁, l'interaction du segment N-terminal est associée à une plus grande ouverture du "canal", alors que dans le cas de TcpC, l'absence d'interaction est associée à ce que l'on pourrait qualifier de fermeture du "canal". Cette interaction pourrait également être influencée par la position et les interactions des hélices transmembranaires qui se situeraient en dessous des trimères. Si des partenaires protéiques venaient à interagir avec ces hélices afin de les déplacer (ainsi que l'ensemble du linker N-terminal) et faciliter une interaction avec les brins des domaines centraux voisins, cela pourrait comme dans le cas d'OrfG₆₄₋₃₃₁, permettre d'ouvrir le "canal". Cependant, cette hypothèse est encore mince et nécessiterait la connaissance des structures de nombreuses autres VirB8-like Gram positif pour être confirmée.

Le dernier trimère dont les interfaces restent à analyser est TraM. Selon PDBePISA chaque monomère possède 6 résidus impliqués dans des liaisons hydrogène et 1 résidu impliqué dans un pont salin. Cette interface compense son manque en résidus hydrophiles en étant plus riche en résidus aromatiques (au moins 8) qui semblent former une poche hydrophobe assez importante dans la partie inférieure de chaque interface, là où les monomères sont les plus proches. Encore une fois, il est difficile de comparer ces interactions puisqu'elles sont de nature et de position différentes par rapport aux deux autres trimères. Les extrémités N-terminales n'y sont pas non plus modélisées, mais l'inclinaison des hélices et la direction de leurs axes sont intermédiaires entre celles de TcpC et d'OrfG₆₄₋₃₃₁.

VII.5 Discussion

Les études structurales des différents domaines VirB8-like au sein des organismes Gram positif et négatif permettent un premier constat : le repliement est globalement conservé avec une identité de séquence très faible. Cela confirme les approches d'alignement utilisant des informations structurales comme le repliement et les structures secondaires pour identifier des protéines similaires là où les identités de séquences sont très faibles (Geourjon *et al.*, 2001). Par exemple en se basant sur leur superposition structurale, les monomères solubles d'OrfG et de TcpC partagent une identité de séquence de 9,5% ; OrfG et TraM 10,5%. Sur les superpositions structurales, on remarque que le coeur hydrophobe au sein des différents groupes reste similaire, avec au minimum plus d'une dizaine de résidus identiques. Dès lors que l'on réalise une superposition plus large entre groupes, il n'y a plus aucun acide aminé conservé, même dans le coeur hydrophobe. Ces informations mettent en évidence l'importance du repliement VirB8-like et de ses structures secondaires qui restent conservées alors que les identités de séquence sont très faibles (inférieures ou égales à 10%).

Chez les bactéries à Gram négatif, on trouve à l'heure actuelle des VirB8 appartenant à la classe α avec un domaine transmembranaire et un domaine VirB8-like situé dans le périplasma qui interagit avec d'autres partenaires du T4SS entre les deux membranes. Chez les bactéries à Gram positif, on trouve des membres appartenant aux 3 classes α , β et γ . De plus, les domaines VirB8-like peuvent être à la fois présents en plusieurs exemplaires dans le milieu extérieur (équivalent périplasma), ou bien dans le cytoplasme des bactéries. A l'heure actuelle, les fonctions exactes de ces différents types de VirB8 ne sont pas connues, et ces protéines pourraient posséder des domaines VirB8-like sans être des VirB8, notamment PrgL et TraH. Ces domaines VirB8-like gardent toutefois leur rôle présumé dans les interactions protéine-protéine vu qu'aucune fonction catalytique n'a été décrite pour le moment.

La plupart des VirB8 Gram négatif s'agencent en dimères dans un cristal, tandis que l'on observe plutôt des conformations trimériques pour les principales VirB8 Gram positif. Le fait que les VirB8 n'aient pas été observées au sein des structures de T4SS Gram négatif jusqu'à l'heure actuelle ne permet pas de savoir si ces dimères sont biologiquement pertinents, ou bien s'ils reflètent des oligomères transitoires, voire des artefacts de cristallisation. Cette dernière hypothèse reste peu probable vu le nombre d'exemples disponibles adoptant cette structure quaternaire. Cette interrogation se

transpose également aux VirB8 Gram positif. Les trimères des trois VirB8 principales Gram positif (OrfG, TcpC et TraM) possèdent de nombreuses interactions hydrophiles et hydrophobes permettant aux trimères d'exister dans le cristal. Surtout dans le cas d'OrfG et TcpC, l'agencement commun de leurs domaines enroulés est difficilement interprétable comme un artefact de cristallisation. Au minimum, nous proposons que ce trimère est soit transitoire, soit fonctionnel avec une ouverture flexible permettant la formation du pore. Toutefois, comme pour les bactéries Gram négatif, de nombreuses données suggèrent que les différentes protéines du T4SS sont présentes en au moins 6 à 12 copies. Les données biologiques concernant OrfG₆₄₋₃₃₁ indiquent qu'une très faible proportion de 6-mères existe, caractéristique qui semble être liée à la concentration de la protéine (B. Douzi, communication personnelle) et qui pourrait dépendre du domaine transmembranaire absent des constructions étudiées en solution.

Lors de nos essais de cristallisation, un cristal d'OrfG₆₄₋₃₃₁ diffractant à 6 Å de résolution a présenté un groupe d'espace hexagonal (l'axe de symétrie 6 étant toutefois avec glissement). Les données étant de très basse résolution, le remplacement moléculaire a été tenté mais n'a pas abouti. Les statistiques indiquent que l'unité asymétrique pourrait accueillir 6 monomères d'OrfG. De nouveaux essais de cristallisation ont été réalisés afin d'optimiser ces cristaux. Ils ont conduit à une forme monoclinique contenant plusieurs monomères assemblés en trimères dans l'unité asymétrique, sans nouveauté. L'obtention de nouveaux cristaux de cette forme hexagonale singulière pourrait résoudre la question d'un oligomère supérieur pour les VirB8 Gram positif.

Les VirB8 Gram positif restent encore mal connues et peu étudiées. On en compte aujourd'hui 4 qui ont été résolues par cristallographie des rayons X et 1 par résonance magnétique nucléaire. De plus, nous ne sommes pas certains que les protéines ayant un domaine VirB8-like qui se situe à l'intérieur de la bactérie possèdent une fonction équivalente aux VirB8 Gram négatif, ni pourquoi certaines possèdent deux domaines VirB8-like. Toutefois, comme VirB8 est l'une des dernières protéines du complexe de la membrane interne des T4SS Gram négatif à interagir avec le substrat, il est probable qu'il en soit de même au sein des T4SS Gram positif au niveau de la membrane. Mieux comprendre cette fonction potentielle ainsi que les interactions avec d'autres partenaires protéiques est déterminant dans l'étude des T4SS Gram positif.

VIII Étude structurale de la protéine OrfM

Dans cette avant-dernière partie, nous allons discuter des résultats de résonance magnétique nucléaire (RMN) sur la protéine OrfM. En premier lieu, les résultats concernant la production et la purification de la protéine marquée seront présentés ainsi que les conditions physico-chimiques qui ont été testées dans le but de trouver le protocole final, dont le détail est décrit dans la partie Matériels et Méthodes XI, page 193. Les expériences de RMN et les résultats sur la protéine OrfM seront ensuite présentés et discutés, en finissant par les expériences de dynamique et d'interaction avec le partenaire OrfL.

VIII.1 Production et marquage isotopique

Habituellement, les textes traitant de la méthode de production/purification d'une protéine se trouvent dans la partie Matériels et Méthodes de ce type de document. Cependant, dans le cas de la protéine OrfM, l'ensemble du processus d'optimisation de cette production et du marquage isotopique a pris plusieurs mois, il mérite donc (à mes yeux) sa propre partie.

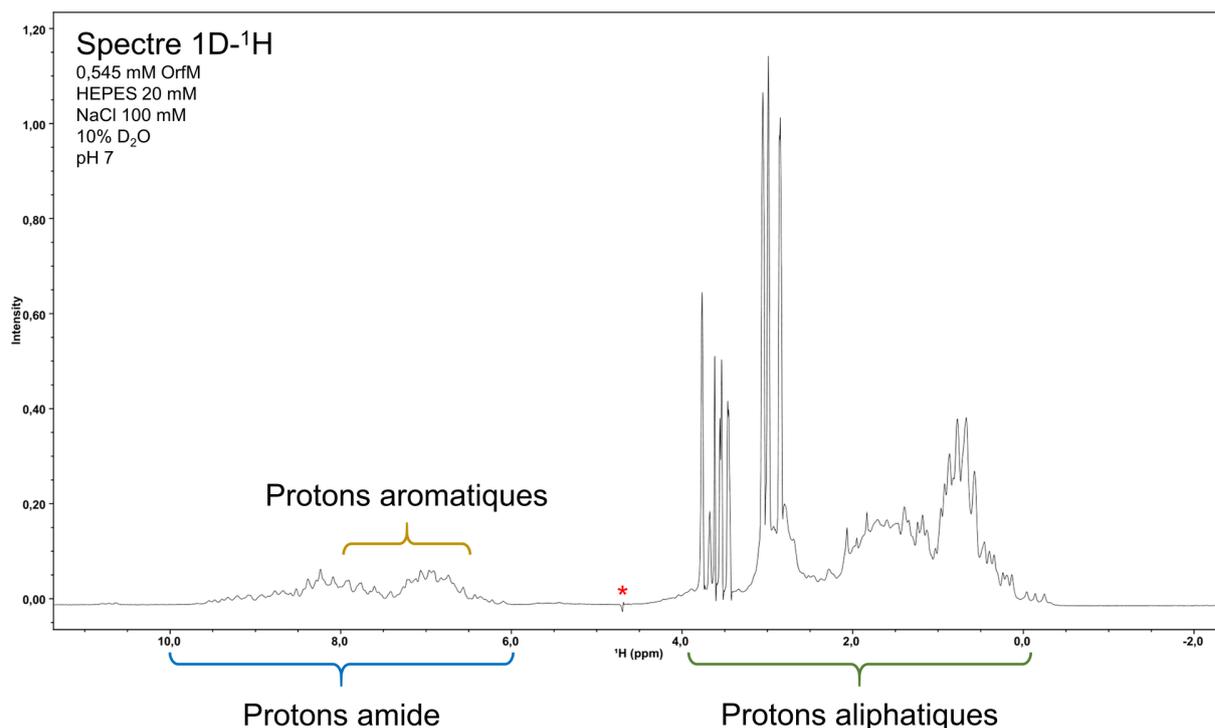
La RMN permet l'observation de noyaux atomiques présentant des propriétés magnétiques caractérisées par un spin nucléaire non nul. Le proton, c'est-à-dire le noyau de l'isotope ^1H le plus commun de l'hydrogène, est très aisément détectable en RMN. Toutefois, du fait de la multitude d'atomes d'hydrogène présents dans une protéine, l'attribution des signaux RMN est impossible à partir d'une expérience de RMN du proton à une dimension. En effet, les recouvrements de signaux sont bien trop nombreux, comme l'illustre le spectre 1D proton d'OrfM sur la Figure 55, page 135.

L'étendue des déplacements chimiques proton observée sur ce spectre, allant de -0,5 à 10 ppm environ, indique néanmoins que la protéine OrfM semble structurée globalement. En effet, le déplacement chimique d'un noyau est directement lié à son environnement chimique. Les protons dans une protéine repliée possèdent des environnements chimiques distincts les uns des autres, ce qui se traduit par une gamme de déplacements chimiques relativement étendue. Au contraire, une protéine désordonnée (en "*random coil*" ou "pelote aléatoire") a une gamme de déplacements chimiques réduite, typiquement entre 0,8 et 8,5 ppm, puisque le déplacement chimique d'un proton résulte de la moyenne de tous les environnements chimiques dans lesquels ce proton est trouvé et il

diffère donc assez peu de celui des autres protons. En solution, l'étude des protéines par RMN est généralement possible pour des poids moléculaires inférieurs à 10 kDa. Elle est réalisée par le biais d'un jeu de spectres homonucléaires proton-proton à deux dimensions montrant soit des couplages proton-proton à travers les liaisons soit des couplages proton-proton à travers l'espace. Pour des protéines de plus grande taille se pose le problème de superposition de signaux à cause du très grand nombre d'atomes d'hydrogène présents. Le recours à des expériences RMN mettant en jeu les atomes d'azote et/ou de carbone devient indispensable pour différencier les atomes d'hydrogène possédant le même déplacement chimique proton. Les isotopes azote ^{15}N et carbone ^{13}C possédant un spin nucléaire adéquat pour pouvoir être observé par RMN sont malheureusement en faible abondance naturelle, n'étant que de 1,1 % pour le carbone ^{13}C et de 0,36 % pour l'azote ^{15}N . Les expériences de RMN hétéronucléaire bi-, voire tri-dimensionnelles nécessaires à l'attribution, requièrent alors un marquage ^{15}N et/ou ^{13}C de l'échantillon, c'est-à-dire que la protéine doit contenir majoritairement les isotopes azote ^{15}N et/ou carbone ^{13}C au lieu des isotopes habituels azote ^{14}N et/ou carbone ^{12}C naturellement présents. Le marquage isotopique d'une protéine est effectué lors de l'expression de celle-ci. Si la cellule ne possède que des acides aminés enrichis en isotopes ^{15}N et/ou ^{13}C , la protéine synthétisée le sera également. Toutefois, pour que cela soit possible, il faut suivre plusieurs règles strictes.

Un marquage isotopique ne peut pas se faire sans précaution. Il faut minutieusement et scrupuleusement contrôler les sources de nutriments pour les isotopes qui nous intéressent. Dans le cas d'un double marquage azote ^{15}N et ^{13}C (nous verrons un peu plus tard pourquoi cela a été nécessaire dans notre cas), il faut contrôler à la fois la source de carbone et la source d'azote. Il n'est donc pas possible d'utiliser les sources usuelles type peptones et sucres variés. Comme les hommes, les cellules n'apprécient guère les régimes. La seule source de carbone passe par le glucose et la seule source d'azote passe par le chlorure d'ammonium, marqués avec l'isotope désiré.

FIGURE 55 – Spectre uni-dimensionnel proton ^1H de la protéine OrfM



Les conditions physico-chimiques sont indiquées en haut à gauche. L'axe des abscisses représente le déplacement chimique proton en ppm. L'axe des ordonnées représente l'intensité du signal. La représentation du spectre a été réalisée avec NMRView (Johnson & Blevins, 1994).

Le déplacement chimique d'un noyau dépend de son environnement chimique donc du groupement chimique auquel il appartient. Un proton de groupement amide se trouve à un déplacement chimique entre 6 et 10 ppm et est observable lorsque sa vitesse d'échange avec le solvant n'est pas trop élevée. La gamme de déplacement chimique des protons aromatiques est entre 6,5 et 8 ppm et se chevauche avec celles des protons amides. Enfin, les protons aliphatiques se situent entre 0 et 5 ppm environ, les CH_3 étant généralement les plus blindés en dessous de 1,5 ppm, et les protons $\text{H}\alpha$ figurant parmi les plus déblindés au-dessus de 3,8 ppm. Le signal de l'eau normalement présent vers 4,7 ppm (*) a été supprimé grâce à une séquence d'impulsions appelée *WATERGATE* qui touche également les signaux de même déplacement chimique. Les signaux fins et intenses entre 2,8 et 3,9 ppm proviennent de l'HEPES.

VIII.2 Protocoles de production

VIII.2.1 Protocole initial de production en milieu riche

La souche *E. coli* BL21 a été transformée avec le plasmide pSKB3-OrfM pour permettre d'exprimer la protéine. Cela a été réalisé dans le laboratoire de nos collaborateurs, je n'y ai pas participé. Tous les détails et les protocoles qui ne sont pas présentés dans cette section sont précisés dans le chapitre "Matériel et Méthodes" dans la partie production/purification d'OrfM XI, page 193. Le protocole initial de production non marqué est le suivant :

(1) Ensemencement de la pré-culture en milieu LB (*lysogeny broth*, "bouillon lysogène") + kanamycine :

Il est très important de commencer une culture viable par une pré-culture de la souche ensemencée la veille de l'induction. Afin de sélectionner uniquement les bactéries comportant le plasmide, un antibiotique est utilisé, la kanamycine. Le plasmide pSK3B utilisé contient un gène de résistance à l'antibiotique. Les bactéries qui se débarrassent du plasmide ne seront pas en mesure de survivre. L'incubation se fait la veille en fin de journée, à température optimale de croissance de la bactérie, c'est à dire 37°C.

(2) Ensemencement d'une culture intermédiaire :

Le lendemain, après 13 à 16h de pré-culture, on ensemence un plus grand volume avec une quantité donnée de bactéries. En mesurant la turbidité de la pré-culture par spectrométrie à une longueur d'onde λ de 600 nm (également appelée densité optique, ou DO), on ensemence à environ 0,1 unité de DO par mL de milieu de culture. Encore une fois, le milieu LB contient l'antibiotique permettant de sélectionner les bactéries contenant le plasmide. On incube à 37°C environ 90 minutes, puis on mesure la DO par intervalle de 15 minutes jusqu'à ce qu'elle atteigne 0,7 unité de DO par mL de milieu de culture. A ce moment là, l'induction doit être faite.

(3) L'induction par l'IPTG (isopropyl β -D-1-thiogalactopyranoside) :

Sur le vecteur plasmidique est présent le gène de la protéine OrfM. Toutefois, pour éviter toute expression précoce qui pourrait impacter sérieusement l'évolution de la culture, le gène est contrôlé par un promoteur lactose qui ne peut être activé qu'en présence de ce dernier, ou d'un équivalent comme l'IPTG. L'induction est réalisée lorsque la souche est dans sa phase de croissance exponen-

tielle et c'est en mesurant la DO que l'on se situe dans son évolution. A ce moment là, les bactéries sont dans un état favorable pour la production de protéines, notamment grâce à une quantité accrue de ribosomes, entités essentielles à la traduction des protéines depuis les ARN messagers.

(4) Arrêt de l'induction et congélation du culot bactérien :

L'induction dure environ 4 à 5 heures, et est arrêtée par centrifugation de l'ensemble du milieu de culture, élimination du milieu, récupération du culot bactérien et congélation à -80°C . Cette dernière permet de retarder l'étape de purification qui peut se faire plus tard, à discrétion. OrfM, qui se trouve en quantité dans le cytoplasme de la bactérie, est dans des conditions favorables pour la conservation à très basse température.

VIII.2.2 Variation des conditions physico-chimiques

La principale différence dans le protocole de production avec marquage est le contrôle strict des sources de nutriments, en particulier des sources de carbone et d'azote comme précisé précédemment. Pour réaliser cela, on utilise des milieux dits "minimum" qui sont entièrement préparés avec les seuls ingrédients choisis par l'expérimentateur. Pour permettre un développement correct des bactéries en milieu minimum, plusieurs éléments sont importants, incluant les conditions physico-chimiques. Il faut noter qu'au cours de l'optimisation du protocole de production en milieu minimum, ces éléments ont été testés au fur et à mesure, sauf le premier point (1) qui est essentiel à la croissance des bactéries. Avant de réaliser un marquage avec des composés comportant les isotopes qui sont très chers, les essais sont réalisés avec les équivalents non marqués.

(1) Les sels, les tampons et l'antibiotique :

Les composés essentiels dans le développement d'un organisme telle qu'une bactérie comprennent : l'eau ; une source de carbone (glucose) ; une source d'azote (chlorure d'ammonium) ; une source de soufre (sulfate) ; une source de phosphore (sels de phosphate) et les oligo-éléments usuels comme le sodium, le potassium, le calcium et le magnésium qui sont retrouvés avec les sources citées plus haut ou sous forme de sels (chlorure de sodium, chlorure de calcium, etc). L'antibiotique est systématiquement présent pour sélectionner les organismes comportant le vecteur. Théoriquement, même si la croissance est ralentie, les bactéries sont dans un environnement adéquat à leur prolifération. Les quantités additionnées des solutés permettent une condition osmotique favorable pour les bactéries,

c'est-à-dire que l'eau située à l'intérieur de la bactérie ne sort pas pour diluer le milieu extérieur (s'il est hypertonique), et inversement si le milieu n'est pas assez riche (hypotonique). L'utilisation d'une quantité double de source de carbone en cas de culture supérieure à 12 heures est une possibilité pour réaliser des cultures sur la nuit.

(2) Température, nombre de pré-cultures, durée d'induction et croissance :

De nombreux protocoles sont possibles en utilisant le milieu minimum et plusieurs ont été testés. La pré-culture est systématiquement réalisée en milieu riche et toujours sur la nuit à température optimale de croissance. En revanche, 2 cultures en milieu minimum à volume croissant sont réalisées le lendemain en ayant pris soin d'éliminer toute trace de milieu riche grâce à des cycles de centrifugation et de rinçage du culot. Ces cultures successives en milieu minimum sont réalisées dans le but de conserver une densité bactérienne ni trop faible ni trop élevée, tout en augmentant la masse bactérienne totale au fur et à mesure. En ayant permis aux bactéries de croître dans le milieu minimum sur de nombreuses générations, on a très largement augmenté la proportion qu'elles contiennent en isotopes issus des sources de carbone et d'azote. Une fois la phase de croissance optimale atteinte dans le volume désiré, on réalise la même induction que dans le protocole classique. Celui-ci stipule que l'induction à 37°C pendant 4 à 5 heures permet d'obtenir une très grande quantité de protéines à purifier par la suite. Cependant, plusieurs autres approches ont dû être testées en milieu minimum. En effet, les essais préliminaires n'ont montré que des quantités très faibles de protéines, ou bien une proportion importante était sous une forme insoluble, irrécupérable durant la purification. Aussi différentes températures (4°C, 30°C et 37°C) durant l'induction ainsi que la durée d'induction (4 h, 6 h et sur la nuit) ont été testées.

(3) Les vitamines et les oligo-éléments :

L'un des derniers points qui a été testé durant ces optimisations est l'ajout de compléments sous forme de vitamines (biotine et thiamine) ainsi que des oligo-éléments plus rares comme le cobalt, le fer, le zinc, le manganèse et d'autres métaux sous forme de sels en petite quantité (de l'ordre du μM ou mM en concentration finale). De nombreux enzymes et processus cellulaires dépendent de ces éléments présents en petite quantité. Concernant les vitamines, leur synthèse peut être ralentie lorsque la population bactérienne évolue dans des milieux pauvres voire dénués d'acides aminés. Dans le but d'accélérer la croissance des bactéries, l'ajout de vitamines est une méthode simple et relativement efficace. Comme le marquage isotopique de ce type de molécules est difficile et très

coûteux, il n'est pas effectué et l'on considère que le bénéfice à la croissance obtenu par l'ajout de ces molécules non marquées est bien supérieur aux désavantages liés à l'apport d'isotopes non désirés.

VIII.2.3 Protocole hybride

Produire et purifier la protéine OrfM en milieu riche est un jeu d'enfant, et il est possible en fin de purification d'obtenir plusieurs dizaines voire centaines de mg de la protéine. En milieu minimum et en utilisant les protocoles usuels de marquage isotopiques, c'est une autre histoire... Sur plusieurs mois, les différents essais mis en place (dont les motivations sont décrites dans la partie précédente) dans le but d'obtenir de plus grandes quantités de protéines ont permis de passer de 1 à 2 mg de protéine en fin de purification, à environ 4 mg. Les premiers essais à l'azote ^{15}N ont permis d'obtenir les premiers spectres RMN que l'on discutera un peu plus tard. Quoiqu'il en soit, la très faible quantité de protéine produite ne nous permettait pas d'envisager un marquage au carbone ^{13}C (beaucoup plus cher) alors que ce marquage était pourtant nécessaire dans le but de résoudre la structure de la protéine.

Parmi les méthodes peu prisées des microbiologistes et des RMNistes, il existe un protocole hybride (ironiquement surnommé "la méthode barbare") qui allie milieu riche et milieu minimum, à l'inverse de ce qu'il faut théoriquement faire pour espérer un marquage uniforme. Les quantités de protéines que nous obtenions en fin de purification étaient si faibles que nous nous devions de l'essayer.

Au risque de diminuer le pourcentage d'isotopes ^{15}N et ^{13}C dans la protéine, la solution a été de combiner le protocole classique en milieu riche avec le protocole en milieu minimum, de sorte que la pré-culture de la veille puis la culture principale visant à placer les bactéries dans un état idéal pour la production des protéines se font en milieu riche (LB). Une fois que la culture primaire atteint entre 0,8 et 1 de DO par mL de milieu, le culot bactérien est nettoyé par des cycles de centrifugations. Le culot est ensuite directement ajouté dans un plus petit volume de milieu minimum que l'on place en incubation à température optimale de croissance pour la bactérie durant environ 1 heure, temps nécessaire à la bactérie pour s'adapter. On induit à l'IPTG, et la culture est de nouveau placée à 37°C sur la nuit. Le culot bactérien est récupéré de la même manière que pour les autres protocoles.

Grâce à ce nouveau protocole, plus de 12 mg de protéine OrfM ont été obtenus en fin de purification, donc trois fois plus qu'en suivant le protocole classique. Cela a permis la préparation d'un échantillon doublement marqué ^{15}N et ^{13}C pour la RMN, de concentration molaire 0,55 mM de protéine, ce qui apporte un gain important par rapport au précédent échantillon à 0,15 mM.

VIII.3 Attribution RMN de la protéine OrfM

Le détail des paramètres utilisés pour les expériences RMN figure dans le chapitre "Matériel et Méthodes" XII, page 199. Nous allons ici expliquer le cheminement depuis les spectres bruts jusqu'aux résultats et leur interprétation, en présentant rapidement l'intérêt de chacun des spectres.

VIII.3.1 Analyse du spectre HSQC- ^{15}N

Un des intérêts du marquage isotopique est de résoudre, grâce à des expériences hétéronucléaires, des problèmes de superposition de signaux de protons. Un marquage simple ^{15}N permet l'enregistrement d'une expérience hétéronucléaire proton- ^{15}N , à deux dimensions, appelée HSQC- ^{15}N (pour "*Heteronuclear Single Quantum Correlation*"). Ce spectre est déterminant dans une étude RMN car il sert de base à de nombreuses autres expériences RMN.

Le spectre HSQC- ^{15}N montre les couplages ^1J entre les noyaux de deux atomes H et N séparés par une liaison covalente. Sont visibles sur ce spectre toutes les liaisons H-N de la protéine, à condition que le proton ne s'échange pas trop rapidement avec l'eau, c'est-à-dire qu'il ne soit pas trop labile. Les liaisons amides de la chaîne principale, celles des chaînes latérales ainsi que la liaison H-N présente dans un cycle de la chaîne latérale des tryptophanes remplissent généralement ces conditions. Le spectre, bi-dimensionnel, comporte alors des pics de corrélation à l'intersection entre le déplacement chimique proton (entre 6 et 10 ppm environ) et le déplacement chimique ^{15}N de ces groupements H-N. Ainsi, des protons qui présenteraient le même déplacement chimique proton ont une chance d'être différenciés sur un spectre HSQC- ^{15}N si les noyaux ^{15}N auxquels ils sont couplés possèdent des déplacements chimiques ^{15}N différents. Par ailleurs, ce spectre renseigne sur le degré de repliement de la protéine de façon plus précise que le spectre 1D proton grâce à l'ajout de la deuxième dimension qui permet d'individualiser la plupart des signaux. Des signaux bien dispersés sont le signe d'une protéine bien repliée, tandis qu'un spectre composé de signaux empacés et

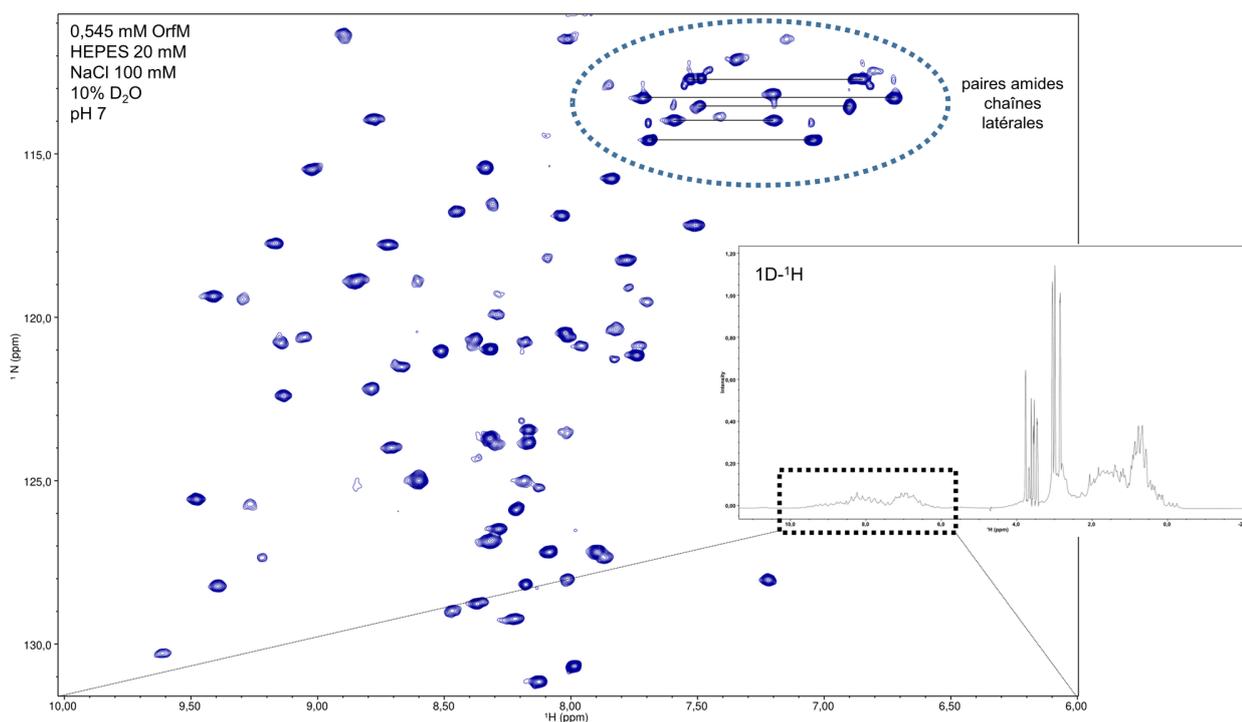
peu résolus signifie généralement que la protéine est peu structurée.

De plus, l'enregistrement d'une expérience HSQC- ^{15}N , qui ne nécessite qu'une dizaine de minutes, permet de surveiller l'état de la protéine au fur et à mesure des expériences. En effet, les spectres multi-dimensionnels en RMN sont généralement enregistrés sur plusieurs heures sur les appareils tels que ceux que nous utilisons. Il faut donc que la protéine reste stable plusieurs jours d'enregistrement à une température d'environ 25°C , ce qui est facilement vérifiable par l'acquisition d'un spectre HSQC- ^{15}N qui est très sensible à la moindre évolution de l'échantillon. Finalement, le spectre HSQC- ^{15}N constitue une "empreinte" spécifique à la protéine et représentative de l'ensemble de son squelette. Il permet d'estimer si la protéine est structurée ou non, et de contrôler la stabilité de l'échantillon au cours du temps.

Le spectre HSQC- ^{15}N de la protéine OrfM est présenté dans la Figure 56, page 142. Les signaux présents sur le spectre sont bien dispersés, entre 6,3 et 9,7 ppm dans la dimension proton et entre 107 et 131 ppm dans la dimension ^{15}N . En tout, un peu plus de cent pics peuvent être comptés, toutefois il est à remarquer qu'une quinzaine d'entre eux sont très peu intenses et il ne peut être exclu qu'ils ne soient artéfactuels. OrfM contient 98 résidus, dont six asparagines et une glutamine possédant un groupement amide NH_2 sur leurs chaînes latérales. Chaque groupement NH_2 donne deux pics supplémentaires à un déplacement chimique azote identique mais des déplacements chimiques proton légèrement différents généralement, formant des paires caractéristiques dans une région blindée du spectre HSQC- ^{15}N , en haut à droite. La protéine OrfM contient aussi neuf prolines, ne possédant pas de proton amide et ne donnant alors pas de signaux sur ce spectre. Par ailleurs, l'azote de la chaîne principale du résidu N-terminal est sous forme d'une amine et non d'un amide, et ne sera pas non plus observé. Par conséquent, le nombre de pics attendus est de $98 + 2 \times 7 - 9 - 1$ soit 102, ce qui est en accord avec le nombre de pics observés. Au final, le spectre HSQC- ^{15}N indique que la protéine semble globalement structurée mais que certains résidus semblent donner des signaux peu intenses.

FIGURE 56 – Spectre bi-dimensionnel HSQC- ^{15}N de la protéine OrfM

Spectre HSQC ^{15}N - ^1H



L'axe des abscisses représente le déplacement chimique proton et l'axe des ordonnées représente le déplacement chimique azote. La représentation du spectre a été réalisée avec NMRView (Johnson & Blevins, 1994).

Le spectre bi-dimensionnel HSQC- ^{15}N est spécifique aux liaisons NH, que l'on peut retrouver entre 6 et 10 ppm sur l'axe des déplacements chimiques protons (représenté sur le spectre 1D- ^1H au centre-gauche de la figure). Les signaux, représentés en bleu sur le spectre HSQC- ^{15}N , correspondent aux pics de corrélation proton-azote dans une liaison NH. Les "paires" de pics en haut à droite du spectre correspondent aux liaisons amides NH_2 des chaînes latérales de glutamine et asparagine dans lesquelles les deux protons sont couplés au même noyau azote, mais ont des déplacements chimiques protons distincts à cause d'un environnement chimique différent.

VIII.3.2 Attribution des spectres de corrélations hétéronucléaires ^1H - ^{15}N

L'étude de protéines par RMN s'apparente à un gigantesque puzzle en 3 dimensions. Avant de pouvoir visualiser des interactions inter-atomiques à travers l'espace et déterminer des angles de torsion phi et psi, afin de réaliser un calcul de structure de la protéine, il faut d'abord l'attribuer c'est-à-dire associer des signaux à des atomes au sein de la molécule.

En 2021, la Protein Data Bank compte un peu plus de 13 000 structures de protéines réalisées par RMN liquide. Le recensement des déplacements chimiques de ces structures permet d'établir des tables de déplacements chimiques pour l'ensemble des protons, azotes et carbones contenus dans les acides aminés d'une protéine. Ce type de tables aide à attribuer un signal non identifié car le déplacement chimique renseigne sur le type de groupement chimique et le type de résidu. Sur le spectre 1D- ^1H , nous avons vu que les protons amides ont généralement un déplacement chimique entre 6 et 10 ppm, les protons aromatiques entre 6 et 8 ppm et les protons aliphatiques entre 0 et 5 ppm, avec les protons $\text{H}\alpha$ entre 3,5 et 5 ppm. La nature du résidu est plus ou moins facile à reconnaître. Les alanines et les glycines sont parmi les plus aisées à identifier car elles présentent une signature caractéristique. Les alanines possèdent en plus du proton $\text{H}\alpha$ un groupement méthyle CH_3 qui donne un signal très intense au voisinage de 1,3 ppm sur la dimension ^1H , et les glycines possèdent deux protons $\text{H}\alpha$ donnant deux signaux proches, entre 3,5 et 5 ppm. Certains acides aminés ont des signatures similaires, à la fois par le nombre et la position de signaux, et sont alors difficiles à différencier. Les prolines sont absentes du spectre HSQC- ^{15}N car elles ne possèdent pas de proton amide. Leurs déplacements chimiques doivent alors être repérés indirectement, grâce aux corrélations avec le proton amide suivant dans la séquence.

De nombreuses expériences peuvent en effet être mises en oeuvre pour corréler certains signaux de manière spécifique, par le biais de couplages à travers les liaisons chimiques ou à travers l'espace. Tout comme un gel d'électrophorèse peut avoir plusieurs dimensions afin d'obtenir une meilleure résolution des taches et faciliter l'analyse, les expériences RMN multidimensionnelles utilisent le même type d'astuce. En ajoutant au spectre HSQC- ^{15}N une troisième dimension, il est possible d'associer à chacun des pics du spectre 2D un ou plusieurs signaux provenant de noyaux corrélés au proton amide considéré et qui permettront d'attribuer ce pic à un résidu précis, comme nous allons l'expliquer.

Les expériences TOCSY-HSQC-¹⁵N (TOCSY pour "*T*otal *C*orrelation *S*pectroscop*Y*") et NOESY-HSQC-¹⁵N (NOESY pour "*N*uclear *O*verhauser *E*ffect *S*pectroscop*Y*") nécessitent seulement un marquage isotopique ¹⁵N car la troisième dimension est une dimension proton qui sera représentée cette fois-ci dans son intégralité afin d'inclure tous les signaux, dont les protons aliphatiques. Le spectre HSQC-TOCSY-¹⁵N met en évidence des couplages relayés avec le proton amide. Ces transferts d'aimantation se font à travers les liaisons covalentes et ne peuvent avoir lieu que vers des protons d'un même résidu. Le spectre NOESY-HSQC-¹⁵N repose sur l'effet Overhauser nucléaire (NOE, prononcer "noé"), correspondant à un couplage à travers l'espace. Il montre les corrélations du proton amide avec tout proton proche dans l'espace, qu'il appartienne ou non au même résidu. L'intensité du pic de corrélation décroît en $1/r^6$ avec la distance r entre protons et devient négligeable au-delà de 5 à 6 Å. L'analyse combinée de ces deux spectres peut suffire pour obtenir l'attribution des signaux d'une protéine de petite taille, le spectre TOCSY-HSQC-¹⁵N renseignant sur le type de résidu et le spectre NOESY-HSQC-¹⁵N permettant l'identification du résidu voisin. L'autre voisin de ce dernier sera ensuite recherché, et en répétant cette procédure de proche en proche jusqu'à rencontrer une difficulté, on pourra obtenir ainsi l'attribution de segments de la séquence. La position des protons H α du résidu i peut être confirmée par l'analyse d'un spectre 3D HNHA montrant les couplages scalaires 3J entre le proton H α (i) et le proton amide (i).

La Figure 57, page 148 présente un exemple d'attribution, pour le segment allant de la sérine 42 jusqu'à la phénylalanine 48, en utilisant les spectres NOESY-HSQC-¹⁵N, TOCSY-HSQC-¹⁵N et HNHA. Une bandelette est associée à chaque signal présent sur le spectre 2D HSQC-¹⁵N, donc à chaque groupement amide de la protéine. Elle représente une portion du plan ¹H -¹H extrait du spectre 3D au déplacement chimique azote de l'amide et centrée dans la dimension horizontale autour du déplacement chimique du proton amide. Ces expériences 3D consistent à disperser le spectre 2D HSQC-¹⁵N dans une troisième dimension, proton ¹H, perpendiculaire aux deux autres dimensions. Cela donne alors lieu à une enfilade de signaux qui apparaîtra le long d'une ligne verticale sur la bandelette.

Au départ, nous disposons d'autant de bandelettes que de pics sur le spectre HSQC-¹⁵N mais ignorions en fait à quel résidu chacune de ces bandelettes correspondait. Choisissons ensemble une bande, celle qui sera plus tard attribuée à la glycine 46. Sur cette bande, deux signaux rouges et plusieurs signaux noirs apparaissent dans la zone entre 3,5 et 5 ppm, où sont trouvés la majorité

des protons $H\alpha$. En rouge est représenté le spectre TOCSY-HSQC- ^{15}N qui est spécifique aux protons du résidu i . La présence de ces deux signaux rouges dans la région des $H\alpha$ suggère que cet acide aminé est une glycine. Des pics noirs aux mêmes déplacements chimiques sont retrouvés sur la bande représentée à droite de celle-ci sur la Figure 57. Ceci indique que cette dernière correspond probablement au résidu suivant dans la séquence car les protons $H\alpha$ d'un résidu i donnent généralement des corrélations NOE bien visibles avec le proton amide du résidu $i+1$. Par ailleurs, il est facile d'identifier ce résidu comme étant une alanine, grâce aux signaux rouges (intra-résidu) montrant la présence d'un proton $H\alpha$ et des protons $H\beta$ de type méthyle caractérisés par un signal unique très intense aux alentours de 1,3 ppm. L'analyse de ces spectres montre finalement que les deux bandes considérées correspondent à une glycine suivie d'une alanine. En regardant la séquence protéique, cet enchaînement glycine-alanine n'a été trouvé que pour les résidus glycine 46 et alanine 47, ce qui a permis d'attribuer précisément ces bandes, avec le numéro de résidu correspondant. En poursuivant cette analyse combinant les informations des différents spectres 3D, il a été possible à partir de la bandelette de l'alanine 47 de trouver celle qui correspondait à la phénylalanine 48 et à partir de la bandelette de la glycine 46, celle de l'histidine 45, puis celle de la lysine 44, etc. La figure 57 représente le résultat de ce travail, les bandes correspondant aux acides aminés 42 à 48 de la séquence d'OrfM ont été extraites et mises dans l'ordre de la séquence.

Ces expériences ont d'abord été menées sur l'échantillon simplement marqué ^{15}N d'OrfM. Malheureusement, leur analyse s'est avérée difficile et n'a abouti qu'à l'attribution de 40 % des résidus. La concentration de l'échantillon de 0,15 mM était faible et un certain nombre de pics visibles sur le spectre 2D HSQC- ^{15}N ne donnaient que très peu voire aucun pic de corrélation sur le spectre 3D NOESY-HSQC- ^{15}N , sensible aux couplages à travers l'espace. Afin de confirmer et compléter l'attribution obtenue, un marquage double ^{15}N et ^{13}C , nettement plus coûteux, a été entrepris afin d'accéder à un ensemble d'expériences triple-résonance 1H , ^{15}N et ^{13}C complémentaires, basées sur d'autres types de transferts, à travers les liaisons.

VIII.3.3 Attribution avec l'aide d'expériences triple résonance 1H - ^{15}N - ^{13}C

D'une part, un spectre HBHA(CO)NH a été enregistré pour faciliter et compléter l'attribution des protons. Cette expérience met en jeu des couplages avec les carbones $C\alpha$, $C\beta$ et CO et permet de visualiser sur la bandelette du résidu i les protons $H\alpha$ et $H\beta$ de l'acide aminé précédent $i-1$. Ce

spectre HBHA(CO)NH a été superposé aux spectres TOCSY-HSQC ^{15}N , NOESY-HSQC ^{15}N et HNHA et apparaît en violet sur la Figure 57. Dans l'exemple présenté, le spectre montre bien, sur la bandelette de la phénylalanine 48, des signaux aux déplacements chimiques du proton $\text{H}\alpha$ et des $\text{H}\beta$ de l'alanine 47 et sur la bandelette de l'alanine 47, des signaux aux déplacements chimiques des deux protons $\text{H}\alpha$ de la glycine 46. Le spectre HBHA(CO)NH apporte dans ce cas une confirmation que les deux bandelettes correspondent bien à des résidus consécutifs dans la séquence.

D'autre part, l'acquisition de plusieurs expériences 3D ^{13}C - ^{15}N - ^1H basées également sur un spectre HSQC- ^{15}N mais montrant cette fois des corrélations du proton amide du résidu i non pas avec des protons mais cette fois avec des ^{13}C . Les bandelettes font donc apparaître verticalement une échelle de déplacement chimique ^{13}C . L'attribution peut être grandement facilitée grâce aux spectres 3D HNCACB et CBCA(CO)NH permettant d'observer les signaux des carbones $\text{C}\alpha$ et $\text{C}\beta$, et les spectres HNCO et HNCACO, ceux des carbones CO.

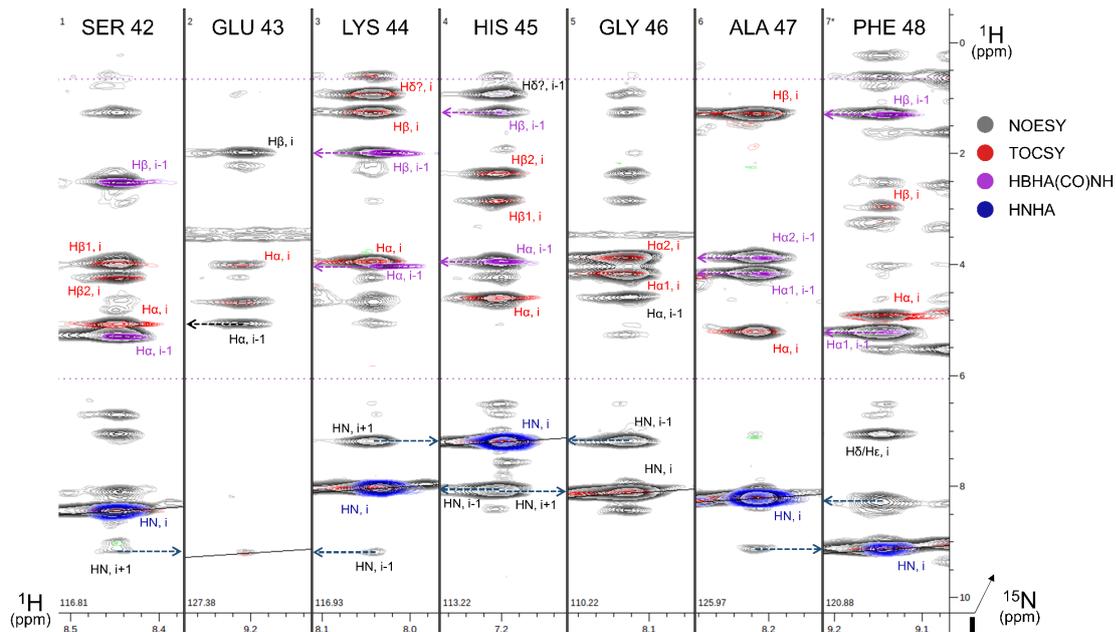
La Figure 58, page 149 représente les bandes extraites des spectres 3D HNCACO et HNCO superposés, pour les résidus 42 à 48. Le premier spectre HNCACO permet l'observation des signaux des carbones CO du résidu i ainsi que du résidu $i-1$, bien que le signal de ce dernier soit généralement plus faible. Le spectre HNCO ne permet d'observer que le signal du carbone CO du résidu $i-1$, mais offre une intensité de signal bien plus élevée.

La Figure 59, page 150 représente les bandes des spectres 3D HNCACB et CBCA(CO)NH superposés pour les résidus 42 à 48. De la même manière que pour les carbones CO, le spectre HNCACB permet d'observer les signaux des carbones $\text{C}\alpha$ et $\text{C}\beta$ du résidu i et de ceux du résidu précédent $i-1$. L'une des particularités de ce spectre est que les signaux des carbones $\text{C}\alpha$ sont de signe positif et apparaissent ici en rouge tandis que les $\text{C}\beta$ sont de signe négatif et sont représentés en vert. Cette différence de signe peut être utile pour les distinguer l'un de l'autre et pour reconnaître le type de résidu, notamment dans le cas des sérines pour lesquels le déplacement chimique du $\text{C}\beta$ est particulièrement bas. Le spectre CBCA(CO)NH, en orange sur la Figure 59, ne montre que des signaux positifs pour les carbones $\text{C}\alpha$ et $\text{C}\beta$ du résidu $i-1$ et permet de trouver l'enchaînement des bandelettes.

Les spectres enregistrés sur l'échantillon doublement marqué ont permis de passer de 40 % à 93 % d'attribution des groupements amides NH de la chaîne principale, montrée sur la Figure

60. Les résidus manquants sont l'asparagine 23, l'alanine 24, le glutamate 32, la thréonine 56, la phénylalanine 72 et la leucine 73. De plus, 91 % des CO et 97 % des C α et C β ont également été attribués.

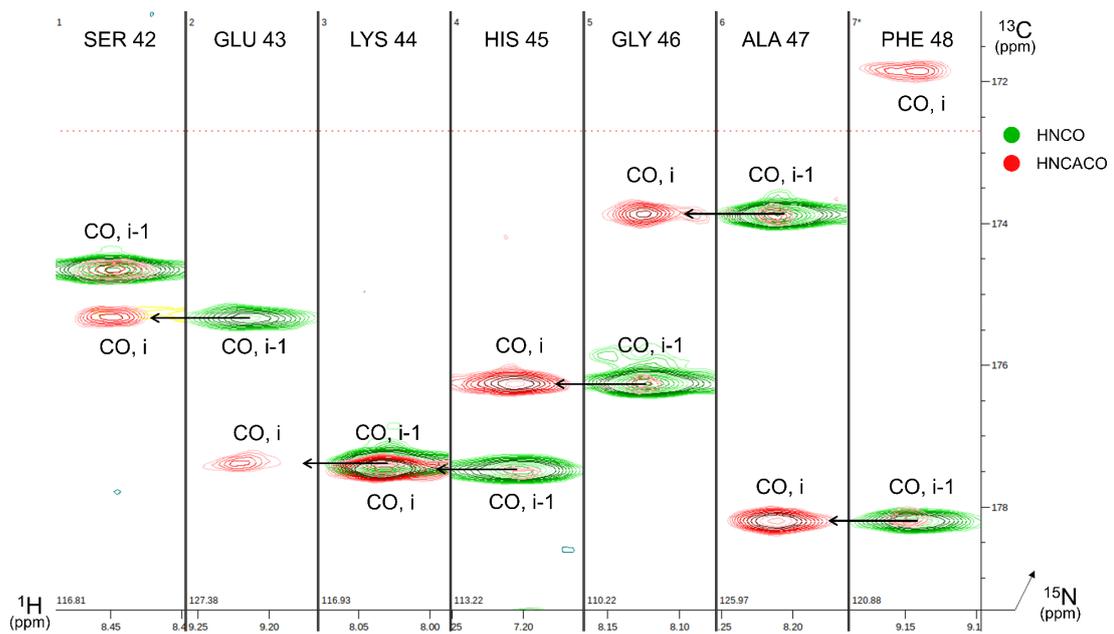
FIGURE 57 – Bandes extraites des spectres 3D ^1H - ^1H - ^{15}N illustrant l'attribution séquentielle des résidus 42 à 48



Attribution séquentielle de la protéine OrfM grâce aux spectres 3D enregistrés dans les dimensions ^1H - ^1H - ^{15}N . Le niveau de contour de chaque spectre est calibré à la limite de l'apparition des artefacts causés par le bruit de fond de chaque expérience. Les signaux sont colorés en fonction des spectres : NOESY-HSQC- ^{15}N (en gris), TOCSY-HSQC- ^{15}N (en rouge), HBHA(CO)NH (en violet), HNHA (en bleu). Les différentes flèches montrent des corrélations entre résidus voisins ayant permis l'attribution séquentielle. La représentation du spectre a été réalisée avec CCPNMR (Vranken *et al.*, 2005).

Chaque bande représente une portion du plan ^1H - ^1H , centrée autour du déplacement chimique proton (dans la dimension horizontale), et à la valeur du déplacement chimique azote (indiquée en bas à droite de chaque bande) correspondant au pic d'un groupement amide i sur le spectre 2D HSQC- ^{15}N . Le spectre NOESY-HSQC- ^{15}N fait apparaître, dans la deuxième dimension proton, l'ensemble des protons dans l'environnement proche (à 5 Å de distance maximum) du proton amide. Le spectre TOCSY-HSQC- ^{15}N fait apparaître uniquement des protons du résidu i et le spectre HBHA(CO)NH, les protons $\text{H}\alpha$ et $\text{H}\beta$ du résidu précédent $i-1$. Le spectre HNHA permet théoriquement de repérer le(s) proton(s) $\text{H}\alpha$ du résidu i , mais aucun n'est visible à un niveau de contour convenable dans cet exemple.

FIGURE 58 – Bandes extraites des spectres 3D ^1H - ^{13}C - ^{15}N pour l'attribution ^{13}C des carbonyles CO

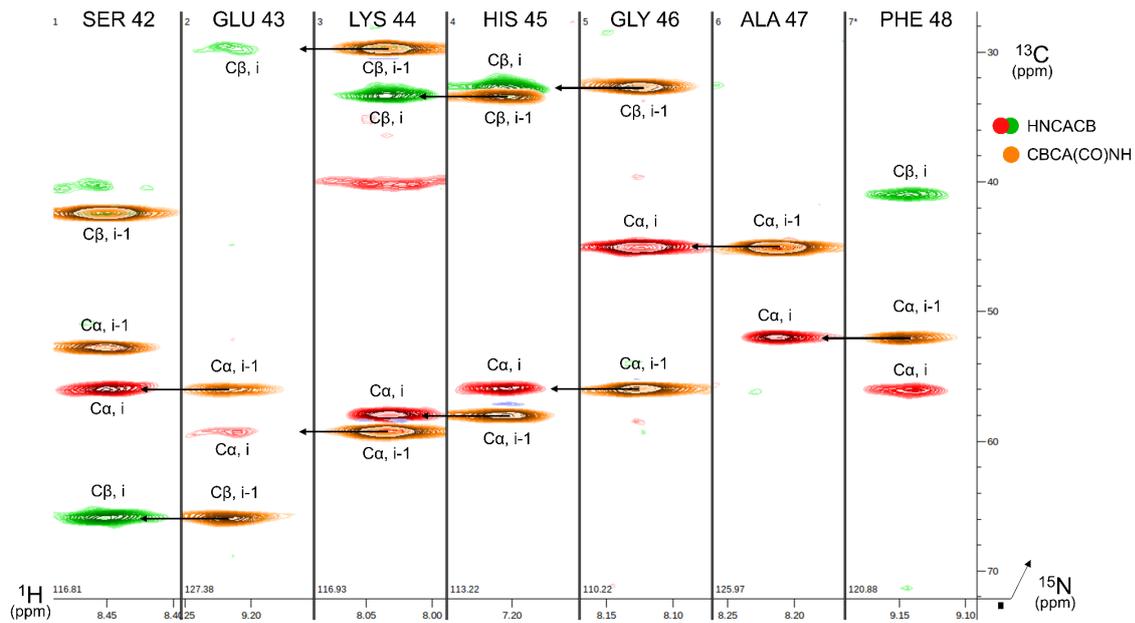


Exemple d'attribution séquentielle de la protéine OrfM.

Le niveau de contour de chaque spectre est calibré à la limite de l'apparition des artefacts causés par le bruit de fond de chaque expérience. Sont superposés les signaux des spectres HN(CA)CO (en rouge) et HNCO (en vert). Les différentes flèches indiquent les signaux provenant du résidu i-1 ayant permis l'attribution séquentielle. La représentation du spectre a été réalisée avec CCPNMR (Vranken *et al.*, 2005).

Le spectre HN(CA)CO permet de visualiser les corrélations d'un proton amide d'un résidu i avec les carbones CO du résidu i et i-1 (en rouge), tandis que le spectre HNCO permet de différencier les deux en ne montrant que le signal du CO du résidu i-1 (en vert).

FIGURE 59 – Bandes issues de spectres 3D ^1H - ^{13}C - ^{15}N pour l'attribution des carbones $\text{C}\alpha$, $\text{C}\beta$

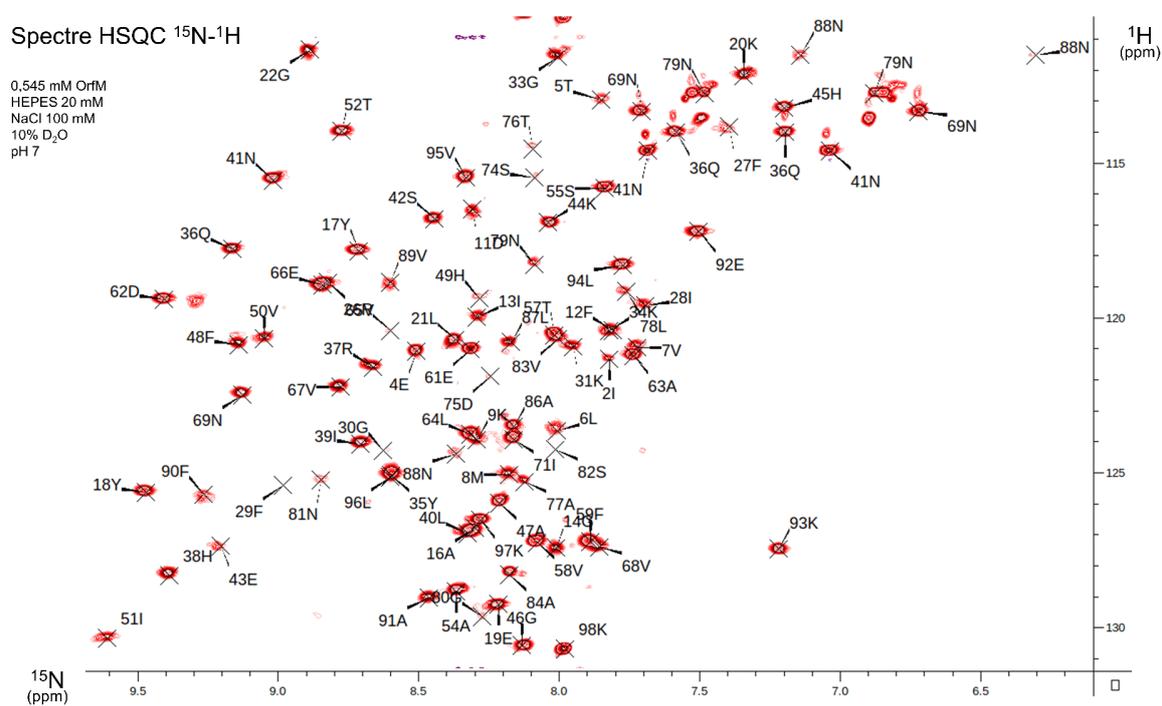


Exemple d'attribution séquentielle de la protéine OrfM à l'aide de spectres 3D enregistrés dans les dimensions ^1H - ^{13}C - ^{15}N .

Le niveau de contour de chaque spectre est calibré à la limite de l'apparition des artefacts causés par le bruit de fond de chaque expérience. Sont superposés les spectres HNCACB (signaux positifs en rouge et négatifs en vert) et CBCA(CO)NH (en orange). Les différentes flèches indiquent les corrélations entre résidus voisins permettant l'attribution séquentielle. La représentation du spectre a été réalisée avec CcpNmr (Vranken *et al.*, 2005).

Le spectre HNCACB montre les corrélations du proton amide d'un résidu i avec les carbones $\text{C}\alpha$ en rouge et $\text{C}\beta$ en vert, à la fois pour le résidu i et pour le résidu $i-1$ (avec une intensité généralement plus faible). La position des carbones $\text{C}\alpha$ et $\text{C}\beta$ du résidu $i-1$ est confirmée par les signaux en orange observés sur le spectre CBCA(CO)NH. Le carbone $\text{C}\beta$ de l'Ala47 est absent du spectre car il est en dehors de la zone de représentation sélectionnée pour la figure.

FIGURE 60 – Spectre 2D HSQC- ^{15}N de la protéine OrfM annoté après attribution



L'axe des abscisses représente le déplacement chimique proton et l'axe des ordonnées représente le déplacement chimique azote. La représentation du spectre a été réalisée avec CCPNMR (Vranken *et al.*, 2005).

VIII.3.4 Détermination de la structure secondaire

L'attribution obtenue donne accès à un grand nombre d'informations qui vont permettre notamment d'identifier les éléments de structure secondaire présents dans la protéine. En effet, les déplacements chimiques observés pour les carbones $^{13}\text{C } \alpha$, $^{13}\text{C } \beta$ et ^{13}C (notés C' sur la Figure 61, page 154) et pour les protons $\text{H}\alpha$ peuvent être utilisés comme indicateurs de structure secondaire. La Figure 61 montre, pour chaque résidu, les écarts entre valeurs observées et valeurs dans une "pelote aléatoire". Dans un feuillet β , les différences $\Delta\delta(^{13}\text{C } \alpha)$ et $\Delta\delta(^{13}\text{C } \beta)$ sont généralement négatives tandis que les différences $\Delta\delta(^{13}\text{C } \beta)$ et $\Delta\delta(^1\text{H } \alpha)$ sont généralement positives. Dans une hélice α , ces différences de déplacements chimiques sont de signe opposé par rapport à celles observées dans les feuillets. Il est courant de déterminer, à partir de ces déplacements chimiques du $\text{C}\alpha$, $\text{C}\beta$, CO et $\text{H}\alpha$, un paramètre appelé indice de déplacement chimique (CSI pour "*Chemical shift index*") qui rendra compte de l'ensemble de ces informations. Une valeur de CSI positive est indicatrice d'une structure en feuillet β , alors qu'une valeur négative indique plutôt une hélice α . L'observation d'un CSI de même signe sur plusieurs acides aminés successifs est nécessaire pour définir une structure secondaire.

En réalité, le déplacement chimique dépend d'un grand nombre de facteurs autres que le type de structure secondaire dans lequel il se trouve, l'interprétation du CSI doit donc être corroborée par l'analyse des corrélations NOE observées sur le spectre NOESY-HSQC- ^{15}N . La Figure 61, page 154 répertorie les corrélations NOE séquentielles observées entre les protons $\alpha(i)$, amide(i) et $\beta(i)$ et le proton amide du résidu suivant (i+1) ainsi que certaines corrélations à moyenne portée entre un résidu i et un résidu i+2, i+3 ou i+4.

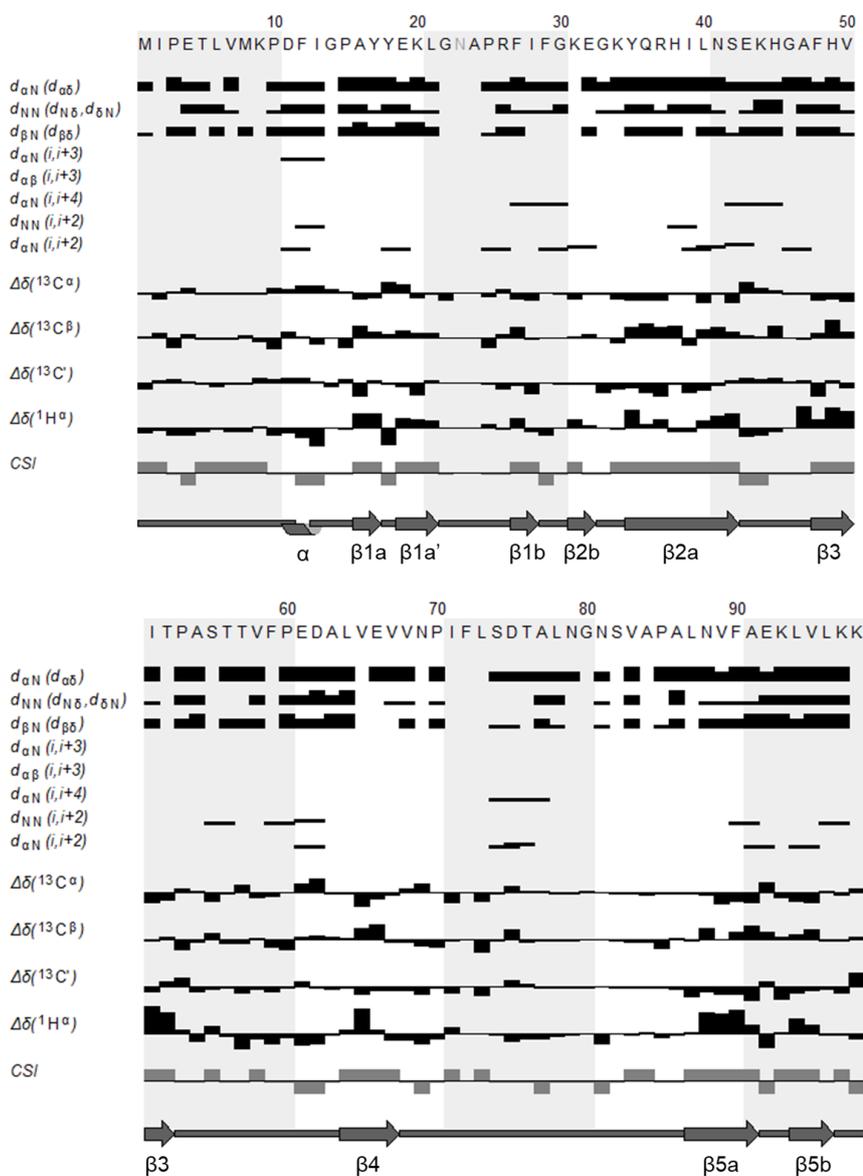
Dans une structure en hélice α , le $\text{H}\alpha$ et le proton amide du résidu i sont respectivement à une distance de 3,5 et 2,8 Å du proton amide i+1, ce qui correspond respectivement à une intensité moyenne et moyenne-forte. Dans une structure en feuillet β , ces distances sont de 2,2 et 4,2 Å, ce qui correspond respectivement à une intensité forte et faible. En plus de ces différences d'intensité, une hélice α présente des corrélations à moyenne portée entre un résidu i et les résidus i+2, i+3 et i+4 alors que les feuillets β font apparaître des corrélations à longue portée entre des résidus appartenant à des brins adjacents mais pouvant parfois être éloignés dans la séquence de la protéine.

Ces corrélations à longue portée qui permettent de définir l'appariement des brins β apparaissent

sur la matrice d'interaction générée par le programme CcpNmr en Figure 62. Cette carte montre les numéros des résidus entre lesquels une corrélation NOE a été observée. Par ailleurs, plus le nombre de corrélations trouvées est grand, plus la case à l'intersection des numéros de résidus est foncée. De nombreuses interactions entre le brin 36-42 et le brin 46-52 sont visibles et sont le long d'une ligne perpendiculaire à la diagonale, ce qui montre que l'appariement est antiparallèle. De même, une structure en feuillet β antiparallèle entre les très longs segments 16-28 et 31-42 est suggérée, mais il est important de noter qu'aucune corrélation au niveau des résidus 23-24 n'est visible car ces résidus n'ont pu être attribués. Par ailleurs, la Figure 61 montre que les caractéristiques RMN de feuillet β ne sont pas trouvées tout le long de ces segments. Par conséquent, ces deux long brins β pourraient être entrecoupés ou du moins ne pas suivre une structure canonique. La Figure 62, page 155 met également en évidence un appariement antiparallèle entre les brins 16-19 et 62-65 d'une part et entre les brins 65-68 et 93-96 d'autre part. Enfin, un petit nombre de corrélations sont visibles entre les segments 49-51 et 86-91 mais sont cette fois le long d'une ligne parallèle à la diagonale et suggère la présence d'un appariement parallèle de ces portions de brins.

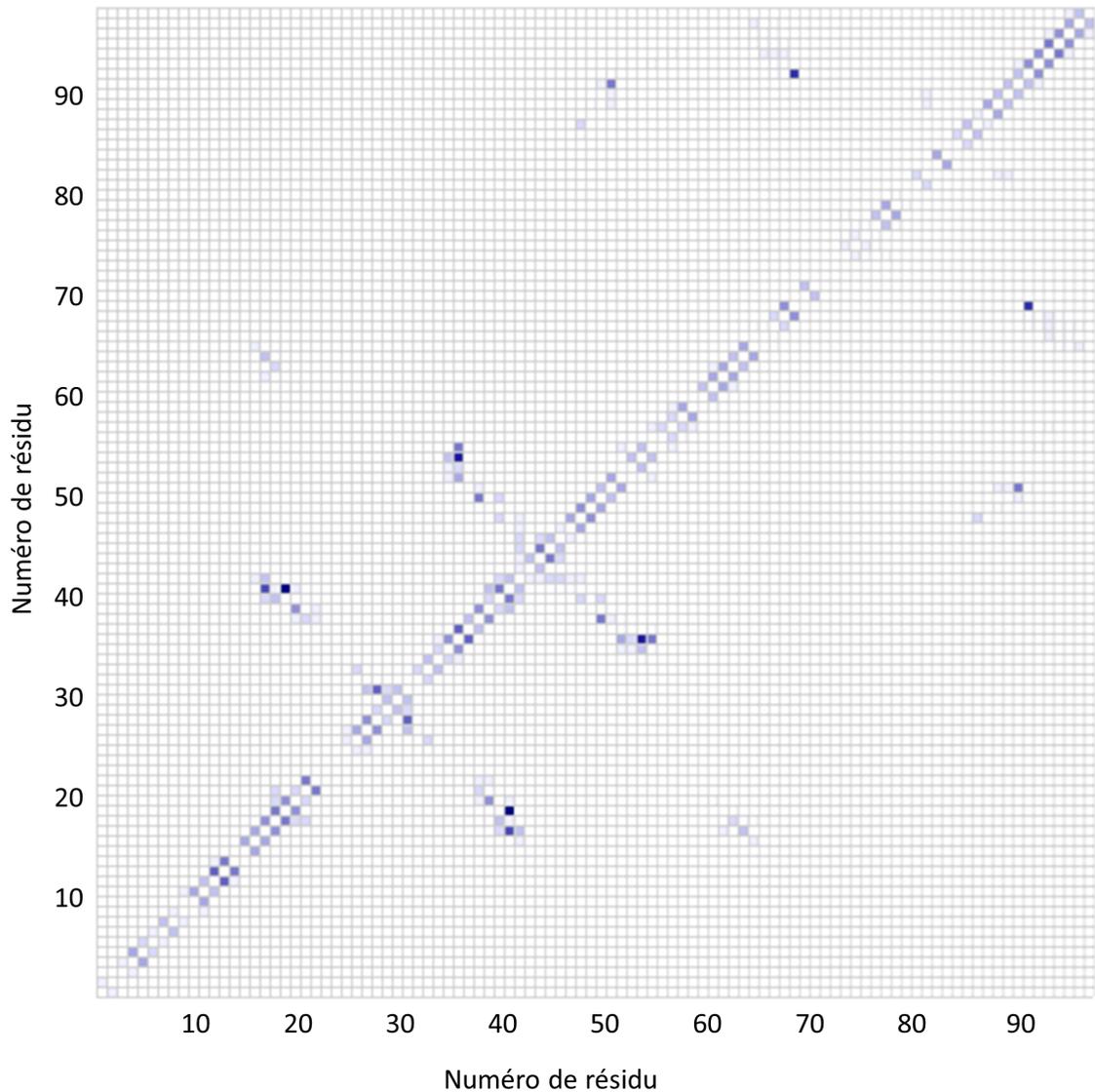
Pour résumer, dans l'état actuel de l'analyse des données RMN, la protéine OrfM peut être décrite comme étant composée de cinq brins, β_1 (16-28), β_2 (31-42), β_3 (46-53), β_4 (62-68) et β_5 (86-96), en précisant que les longs brins β_1 , β_2 et β_5 n'ont pas une structure canonique et pourraient être plutôt décrits comme la succession de plusieurs brins plus courts reliés par des segments d'un à trois résidus (Figure 61). Ces brins sont appariés suivant l'ordre β_1 - β_2 - β_3 - β_5 - β_4 (- β_1), le dernier des brins, β_4 , revenant vers le premier brin β_1 pour former un feuillet antiparallèle. Tous les autres appariements sont également antiparallèles mis à part les brins β_3 - β_5 qui sont parallèles.

FIGURE 61 – Identification des structures secondaires d'OrfM à partir des données RMN



La hauteur des bâtons figurant sur les lignes $d_{\alpha N}$, d_{NN} et $d_{\beta N}$ représentent l'intensité des corrélations NOE séquentielles observées respectivement entre les protons $\alpha(i)$, amide(i) et $\beta(i)$ et le proton amide du résidu suivant (i+1). Les corrélations à moyenne portée $d_{\alpha N}(i, i+3)$, $d_{\alpha\beta}(i, i+3)$, $d_{\alpha N}(i, i+4)$, $d_{NN}(i, i+2)$ et $d_{\alpha N}(i, i+2)$ sont représentées par des segments reliant les résidus en question. L'index de déplacement chimique (Chemical Shift Index ou CSI) a été calculé à partir des déplacements chimiques des $C\alpha$, $C\beta$, CO et $H\alpha$. Les structures secondaires représentées en bas ont été définies à partir de l'analyse de l'ensemble des données RMN. Les corrélations à longue portée (i,j) avec $j > i+4$ permettant de caractériser l'appariement des brins β ne sont pas représentées ici. La figure a été générée par CcpNmr.

FIGURE 62 – Matrice d'interaction des résidus d'OrfM



L'axe des ordonnées et l'axe des abscisses correspondent aux numéros croissants de résidu dans la matrice d'interaction. L'intensité de la couleur de chaque case (i,j) représente le nombre de pics de corrélation NOE trouvés entre le résidu i et le résidu j . Elle va du blanc (aucun pic) au bleu le plus foncé (28 pics). Les cases situées sur la diagonale (d'en bas à gauche jusqu'à en haut à droite) représenteraient les interactions d'un acide aminé avec lui-même, et ne sont pas colorées. Les interactions entre acides aminés voisins $(i,i-1)$ et $(i,i+1)$ apparaissent sur deux lignes autour de la diagonale. L'analyse des interactions à longue portée (i,j) avec $|j-i|>4$ permet de visualiser l'appariement des brins dans les feuillets β : les feuillets β parallèles se caractérisent par des bandes parallèles à la diagonale et les feuillets anti-parallèles par des bandes perpendiculaires à la diagonale. En effet, si le résidu i est proche du résidu j , le résidu $i+1$ le sera avec $j+1$ dans le cas d'un feuillet parallèle ou avec $j-1$ dans celui d'un feuillet anti-parallèle, puis le résidu $i+2$ sera proche de $j+2$ ou du $j-2$, etc.

VIII.3.5 Caractérisation dynamique à partir d'expériences de relaxation ^{15}N

La RMN est assez unique par rapport à d'autres méthodes structurales car elle permet de caractériser des processus dynamiques à l'intérieur d'une molécule se faisant à des échelles de temps très variées, de la picoseconde (10^{-12} seconde) jusqu'à la seconde et au-delà. Les mouvements les plus rapides recouvrent par exemple les vibrations et librations des liaisons et les rotations de chaînes latérales tandis que les plus lents incluent les changements conformationnels suite à la fixation d'un ligand, les processus de repliement, les mouvements allostériques (Chao & Byrd, 2020). Le marquage ^{15}N d'une protéine permet de déterminer des paramètres de relaxation de spin des noyaux ^{15}N qui apporteront des informations dynamiques sur les liaisons NH amides et donc sur le squelette. Les paramètres classiquement mesurés pour l'étude dynamique d'une protéine sont la vitesse de relaxation longitudinale R_1 , la vitesse de relaxation transversale R_2 et le NOE hétéronucléaire ^1H - ^{15}N .

Le principe général des expériences de relaxation est le suivant. Le champ magnétique statique du spectromètre induit un alignement des spins qui crée une aimantation. Lors des expériences de mesure de R_1 et R_2 , on va d'abord exciter les spins ^{15}N d'une certaine manière, puis laisser le système retourner à son état d'équilibre pendant une durée appelée délai de relaxation, et enfin mesurer l'intensité de l'aimantation restante. La vitesse de relaxation longitudinale correspond à la vitesse à laquelle l'intensité de l'aimantation évolue suivant l'axe du champ magnétique z pendant cette période de relaxation et la vitesse de relaxation transversale, à celle à laquelle l'intensité de l'aimantation diminue dans le plan transversal xy . Ces vitesses peuvent être déterminées en acquérant une série d'expériences avec des valeurs différentes de délai de relaxation. L'intensité des signaux est ensuite mesurée sur ces spectres, qui se présentent comme des spectres 2D HSQC- ^{15}N . Enfin, les valeurs de R_1 ou R_2 sont ensuite obtenues par un ajustement, avec une fonction exponentielle décroissante, de l'intensité des pics en fonction de délai de relaxation. Dix délais entre 2 ms à 1,4 s ont été utilisés pour les expériences R_1 . Quant à l'expérience NOE hétéronucléaire ^1H - ^{15}N , elle consiste à évaluer le couplage dipolaire ^1H - ^{15}N dans une liaison NH en mesurant l'effet d'une irradiation du proton ^1H sur le noyau ^{15}N . L'intensité des pics sur le spectre obtenu est comparée à celle obtenue en l'absence d'irradiation. Le NOE hétéronucléaire ^1H - ^{15}N correspond au rapport entre l'intensité de la première et celle de la deuxième, il est sans unité et a une valeur inférieure à 1.

Les expériences de relaxation ^{15}N ont été réalisées sur le premier échantillon de la protéine OrfM marquée uniquement à l'azote ^{15}N . Les valeurs de R_1 , R_2 et de NOE hétéronucléaires sont présentées dans la Figure 63 page 160. Un jeu de données complet avec ces trois paramètres a été obtenu pour 77 acides aminés. Les résidus manquants sont (1) ceux qui n'ont pas été attribués sur l'HSQC- ^{15}N ; (2) ceux dont le pic sur l'HSQC- ^{15}N n'est pas suffisamment résolu pour permettre une mesure fiable de l'intensité; (3) ceux dont le signal est déjà au départ de très faible intensité et s'affaiblit encore avec le délai de relaxation, ce qui donne une incertitude sur la vitesse de relaxation supérieure à sa valeur. Les valeurs et les incertitudes moyennes observées sont $\langle R_1 \rangle = 1,7 \pm 0,1 \text{ s}^{-1}$, $\langle R_2 \rangle = 10,1 \pm 0,6 \text{ s}^{-1}$, $\langle \text{NOE hétéronucléaire} \rangle = 0,67 \pm 0,06$.

La valeur du NOE hétéronucléaire est très sensible aux mouvements très rapides, à une échelle de temps entre la ps et la ns, et reflète le degré de flexibilité du squelette polypeptidique. Plus la valeur est faible, plus la chaîne est flexible. Les résidus 2-11, 29, 71-76, 78-80, 84-88, 98 ont des valeurs de NOE hétéronucléaire inférieures à 0,6, par conséquent la chaîne polypeptidique peut être décrite comme non ou peu structurée au niveau de ces résidus. Le paramètre R_1 montre des valeurs supérieures à la moyenne dans les segments 2-13, 26, 29-33, 77-84 indiquant la présence de mouvements rapides. Le paramètre R_2 n'est pas seulement lié aux mouvements très rapides de l'ordre de la ps et ns, mais également aux mouvements plus lents à l'échelle de la microseconde à la milliseconde qui est la gamme des échanges conformationnels ou chimiques. Ces derniers peuvent correspondre par exemple à des équilibres entre plusieurs conformations, entre une forme protonée et une forme déprotonée, entre une forme liée et une forme libre (en cas d'interaction avec un ligand ou un partenaire protéique...) ou entre une forme monomérique et une forme oligomérique.... La présence de tels échanges se traduit par une augmentation de la valeur du R_2 . Des résidus montrant des valeurs particulièrement élevées sont situés dans les segments 6-14, 19-21, 29-30, 38, 48-54, 69, 74-75, 88-93, ce qui suggère qu'ils pourraient être affectés par un phénomène d'échange.

Il est possible d'aller un peu plus loin dans l'interprétation des paramètres de relaxation, par une analyse de type Lipari-Szabo qui peut être effectuée par le logiciel Tensor 2.0 (Dosset *et al.*, 2000). Dans ce formalisme, on considère que les mouvements des liaisons NH résultent du mouvement global de réorientation de la molécule dans la solution et du mouvement interne de cette liaison dans la molécule. On cherchera donc à déterminer les paramètres dynamiques qui permettront de décrire ces deux mouvements. Il faut dans une première étape caractériser le mouvement global

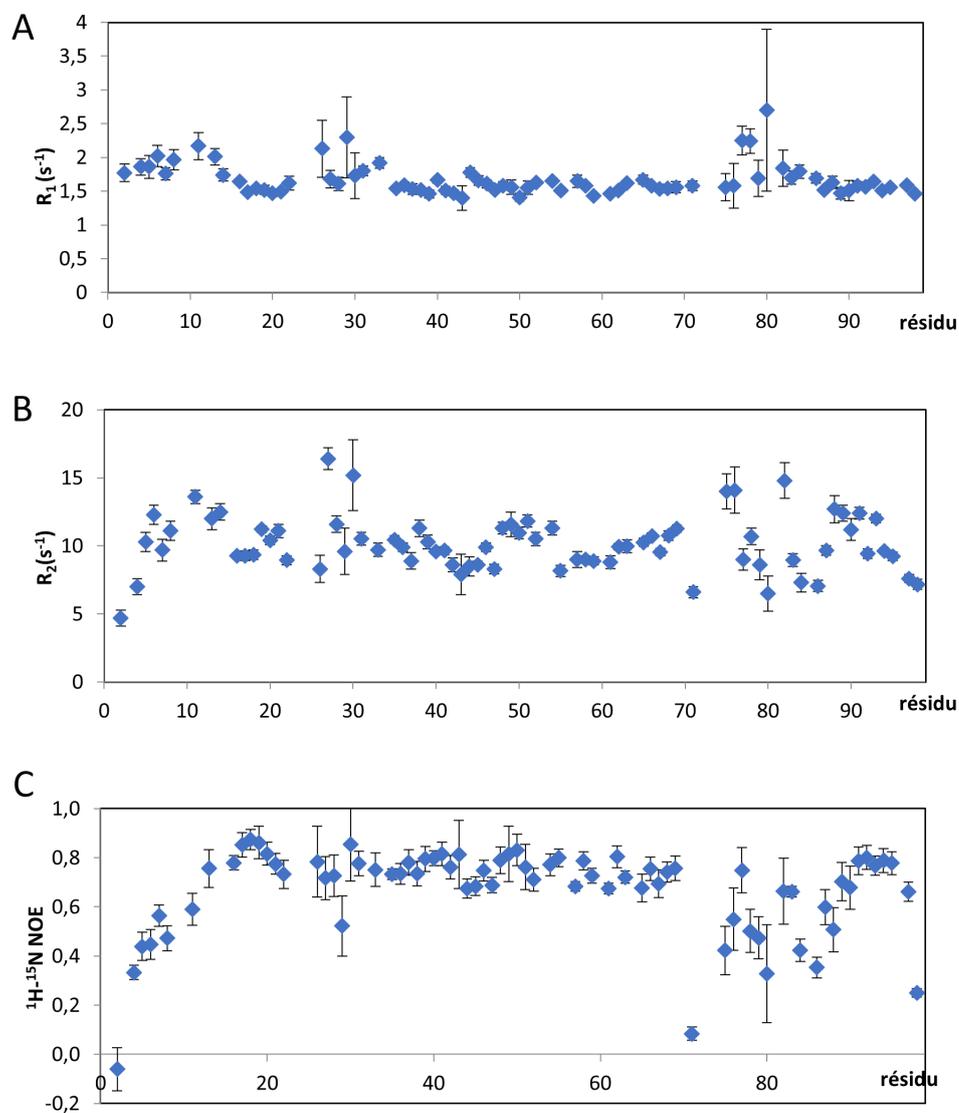
de la protéine. Le modèle le plus simple d'une réorientation isotrope a été choisi, c'est-à-dire que la molécule est supposée tourner sur elle-même à la même vitesse quelle que soit la direction. Ce mouvement global est alors décrit par un seul paramètre appelé temps de corrélation global et noté τ_c , qui correspond au temps moyen nécessaire à la molécule pour effectuer une rotation sur elle-même. Dans ce formalisme de Lipari-Szabo, ce temps de corrélation peut être évalué à partir des rapports R_2/R_1 , à condition d'utiliser uniquement les valeurs des résidus relativement rigides ne présentant pas de mobilité interne complexe. En choisissant des résidus présents dans les brins β clairement identifiés, une valeur de 7,3 ns ($\pm 0,1$ ns) a été obtenue.

La mobilité interne au niveau de chacun des résidus est analysée dans une seconde étape, à partir de ses valeurs de R_1 , R_2 et NOE hétéronucléaire et du temps de corrélation global trouvé pour la protéine. Les mouvements internes des liaisons NH sont modélisés par des mouvements de libration ultra-rapides à une échelle de temps inférieure à 20 ps qui correspondent à des oscillations de ces liaisons NH à l'intérieur d'un cône dont l'angle d'ouverture est plus ou moins grand. Le paramètre d'ordre S^2 est un paramètre associé à cet angle d'ouverture qui prend une valeur entre 0 et 1. Un paramètre d'ordre de 0 correspondant à l'angle d'ouverture le plus grand donc à un mouvement non restreint dans l'espace, alors qu'un paramètre d'ordre proche de 1 correspond à un angle tendant vers 0, donc à un mouvement extrêmement restreint. Le paramètre d'ordre S^2 et le temps de corrélation global interviennent dans l'expression littérale établie par Lipari-Szabo pour ces trois paramètres de relaxation R_1 , R_2 et NOE hétéronucléaire. Le logiciel va donc chercher à déterminer le paramètre S^2 qui permettra de rendre compte des trois paramètres de relaxation expérimentaux. En cas d'échec, il va considérer un modèle de mouvement plus complexe faisant intervenir un, voire plusieurs paramètres supplémentaires en cas de nécessité. Un temps de corrélation interne τ_i peut être introduit, lorsque s'ajoutent aux librations ultra-rapides, des mouvements à une échelle de temps entre 20 ps et le temps de corrélation. Un terme d'échange R_{ex} peut également être ajouté, dans le cas où des mouvements encore plus lents de l'ordre de la μ s à ms auraient lieu. Ainsi, seuls les résidus possédant une mobilité interne complexe présentent des valeurs de τ_i et/ou R_{ex} .

Les 77 résidus de la protéine OrfM dont les valeurs de R_1 , R_2 et NOE hétéronucléaire ont été déterminées ont pu être analysées par le logiciel Tensor2 et les résultats sont présentés en Figure 64, page 161. Les paramètres d'ordre S^2 les plus bas sont observés principalement dans la région

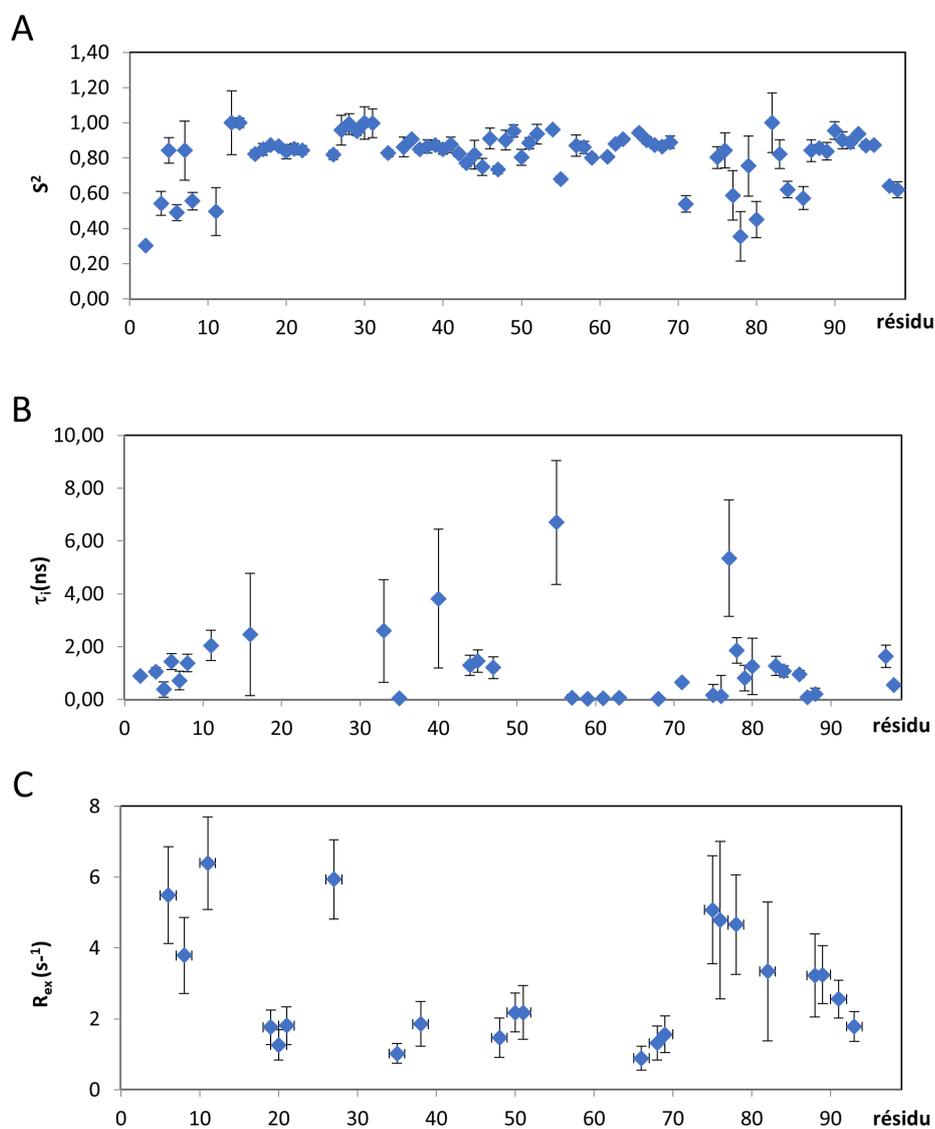
N-terminale jusqu'au résidu 11, pour le résidu 71 et dans la région 77-86 (entre les brins $\beta 4$ - $\beta 5$) ainsi que pour les deux derniers résidus. Leurs valeurs, pour la plupart inférieures à 0,7, témoignent d'une très grande flexibilité de squelette. La majorité des résidus dans ces zones présentent de plus un temps de corrélation interne τ_i autour de 1-2 ns et certains même un terme d'échange R_{ex} , ce qui indique la présence de mouvements internes de l'ordre de la ns et parfois d'échange conformationnel ou chimique. Dans une plus faible mesure, plusieurs résidus du segment 43-47 (dans la boucle $\beta 2$ - $\beta 3$) et le résidu 55 (dans la boucle $\beta 3$ - $\beta 4$) possèdent des paramètres d'ordre inférieurs à la moyenne et présentent une flexibilité notable. L'obtention d'un temps de corrélation interne τ_i de l'ordre de la nanoseconde suggère par ailleurs l'existence de mouvements internes rapides. Enfin, le reste des résidus présentent un paramètre d'ordre S^2 supérieur à 0,8 et forment donc des zones qui peuvent être considérées comme structurées. L'ensemble des résidus constituant les brins β identifiés précédemment figurent bien parmi cette dernière catégorie. Toutefois, on peut remarquer que nombre de ces résidus notamment 19-21, 48-51, 66-69 présentent un terme d'échange, de l'ordre de la micro- à la milliseconde, reflétant un équilibre conformationnel ou chimique.

FIGURE 63 – Paramètres de relaxation ^{15}N R_1 , R_2 et NOE hétéronucléaire de la protéine OrfM



Paramètres de relaxation ^{15}N mesurés sur le premier échantillon de la protéine OrfM marquée à l'azote ^{15}N . Protéine à une concentration de 0,15 mM dans un tampon Tris-Hcl pH 7,5 et 200 mM NaCl, à 298 K. (A) Vitesse de relaxation longitudinale R_1 . (B) Vitesse de relaxation transversale R_2 . (C) NOE hétéronucléaire $^1\text{H}-^{15}\text{N}$.

FIGURE 64 – Paramètres de dynamique interne de la protéine OrfM



Paramètres de dynamique interne obtenus avec Tensor2 pour la protéine OrfM marquée à l'azote ^{15}N à une concentration de 0,15 mM dans un tampon Tris-HCl pH 7,5 et 200 mM NaCl, à 298 K. (A) Paramètre d'ordre S^2 , (B) Temps de corrélation interne τ_c , (C) Terme d'échange R_{ex} .

VIII.3.6 Calculs préliminaires de structure

Le logiciel CcpNmr est capable grâce à plusieurs sous-programmes de générer des contraintes pour les calculs de structure. En l'occurrence, des contraintes d'angles peuvent être définies grâce au sous-programme DANGLE (Cheung *et al.*, 2010) qui utilise les déplacements chimiques $C\alpha$, $C\beta$, CO, ^{15}N et $\text{H}\alpha$ attribués, elles réduisent l'amplitude des angles dièdres (ou angles de torsion) ϕ et ψ de la chaîne principale. Comme les signaux NOE possèdent une intensité variant en $1/r^6$ avec la distance proton-proton, il est possible (après calibration automatique ou manuelle) de déduire des contraintes de distances pour chaque pic de corrélation NOE attribué. Enfin, il est possible de générer des contraintes de liaisons hydrogène, par exemple dans des structures secondaires. Ces contraintes, ainsi que la séquence primaire de la protéine à modéliser, sont ensuite fournies à un logiciel de calcul de structure comme ARIAWeb (Allain *et al.*, 2020) qui va chercher à générer des conformations en accord avec l'ensemble des contraintes introduites. Pour cela, un terme énergétique est associé à chaque contrainte et est ajouté à l'énergie de la molécule, calculée dans le champ de force implémenté dans le logiciel. Ce terme additionnel vaut zéro si la contrainte est respectée dans le modèle structural mais augmente d'autant plus que la contrainte n'est pas respectée. Ainsi, en cherchant à minimiser l'énergie totale de la molécule, le logiciel va proposer un certain nombre de modèles répondant au mieux aux contraintes RMN. Il est important de noter que le calcul de la structure d'OrfM est encore en cours, les résultats présentés sont donc provisoires. Actuellement, 907 contraintes de distance, dont 282 intra-résiduelles, 291 séquentielles, 77 à moyenne portée et 257 à longue portée, ont été dérivées des pics NOE attribués sur le spectre 3D NOESY-HSQC ^{15}N et ont été introduites dans ARIAWeb comme des contraintes inambiguës. Le calcul effectué utilise de plus 152 contraintes d'angle générées par DANGLE et 24 contraintes de liaisons hydrogène définies à partir des feuillets β clairement identifiés.

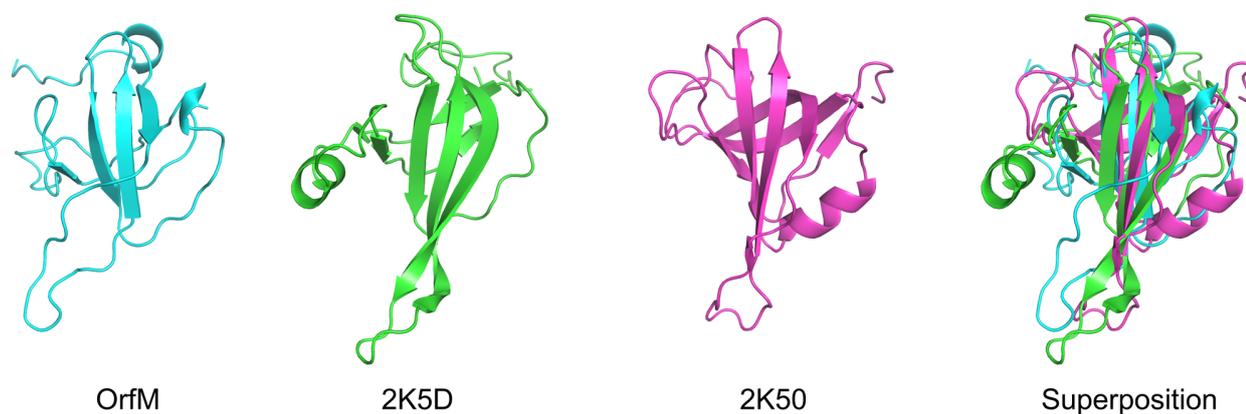
Parmi la centaine de modèles générés par ARIAWeb, 14 modèles de basse énergie ont été sélectionnés et superposés. En excluant les régions très désordonnées 1-13, 23-36, 70-78, 82-85 et 95-98, des RMSD moyens de $2,2 \pm 0,4$ Å sur les atomes du squelette et de $3,4 \pm 0,7$ Å sur les atomes lourds ont été obtenus. Le nombre moyen de violations de distance supérieures à $0,5$ Å est de 58 et celui des violations d'angle supérieures à 5° est de 19. L'énergie moyenne des modèles est de 900 kcal/mol. La structure d'OrfM présente plusieurs brins β formant un feuillet mixte de cinq brins (voir Figure 66, page 166) en accord avec les structures secondaires identifiées sur la Figure 61, page

154. Les résidus 16 à 42 forment une sorte de très longue épingle à cheveux dont la structure est bien définie à ses extrémités, montre un appariement antiparallèle des brins β_{1a} et $\beta_{1a'}$ entre les résidus 16-21 avec les résidus 38-42 du brin β_{2a} . Le rapprochement des brins β_{1b} (résidus 26-27) et β_{2b} (résidus 30-31) est bien visible et définit clairement la tête de l'épingle, cependant la conformation des résidus 22-25 et 32-34 situés respectivement entre $\beta_{1a'}$ et β_{1b} et entre β_{2a} et β_{2b} est mal définie, ce qui se traduit par l'observation d'un ensemble d'épingles plus ou moins distordues au niveau de leurs parties centrales. L'échec dans l'attribution des résidus 23, 24 et 32 est en partie responsable du faible nombre de contraintes RMN trouvées dans cette région et donc de sa mauvaise définition. Toutefois, il est à noter qu'aucune de ces épingles n'apparaît totalement droite, une boursoufflure au niveau des résidus 23-24 et/ou une torsion autour de l'axe de symétrie de l'épingle sont observées dans les 14 modèles. Ces modèles reflètent donc bien les données RMN présentées précédemment sur la Figure 61, page 154, à savoir la présence d'un très long feuillet antiparallèle entrecoupé ou pour le moins déformé. Quant aux autres brins β , l'appariement parallèle des longs brins β_3 et β_{5a} apparaît clairement mais le très court feuillet antiparallèle β_{5b} - β_4 - β_1 est plus difficilement visible. Le modèle montre toutefois que la chaîne polypeptidique entre les résidus 86 et 96 doit se recourber pour que β_{5a} et β_{5b} puissent s'apparier respectivement avec β_4 et β_3 et ainsi contribuer à enrouler ce feuillet sur lui-même. Si on désigne l'ensemble du segment 86-96 par brin β_5 , il faut donc garder à l'esprit qu'il s'agit en fait de deux courts brins β_{5a} et β_{5b} reliés par deux résidus. Finalement, la modélisation confirme le feuillet mixte β_1 - β_2 - β_3 - β_5 - β_4 (- β_1) déduit de l'analyse des données RMN présentée sur la Figure 61. Il peut être noté que ce feuillet présente en son centre un coeur dense, riche en acides aminés hydrophobes non aromatiques. Un tour d'hélice α est parfois présent entre les acides aminés 10 et 13. La très longue boucle 69-85 située entre les brins β_4 - β_5 est assez mal définie globalement, peu de contraintes ont en effet été identifiées (Figure 62, page 155), en particulier pour les résidus 72 et 73 qui manquent à l'attribution.

VIII.4 Analyse préliminaire du modèle d'OrfM

Le modèle préliminaire d'OrfM a été analysé grâce au serveur de comparaison structurale DALI (Holm, 2020). Deux structures de protéine se distinguent, de codes PDB 2K5D et 2K50. Même si les scores d'homologie structurale sont modestes ($Z = 4,6$ et $3,6$, respectivement), il ne fait aucun doute que ces trois protéines partagent le même repliement. La superposition des structures est

FIGURE 65 – Superposition des structures homologues d'OrfM



Superposition réalisée par le serveur DALI (Holm, 2020) des structures 2K5D (en vert), 2K50 (en rose) et de l'un des états d'OrfM sélectionné parmi les 14 modèles de plus basses énergies générés par AriaWEB (en bleu).

présentée dans la Figure 65, page 164. Le RMSD est de 7,1 Å (sur 81 résidus alignés) entre OrfM et 2K5D ; il est de 3,6 Å (sur 71 résidus) entre OrfM et 2K50. Bien que la longueur des brins ne soit pas identique avec les deux autres structures, les 4 brins β qui participent au feuillet ont la même orientation et sont à des positions similaires dans l'alignement.

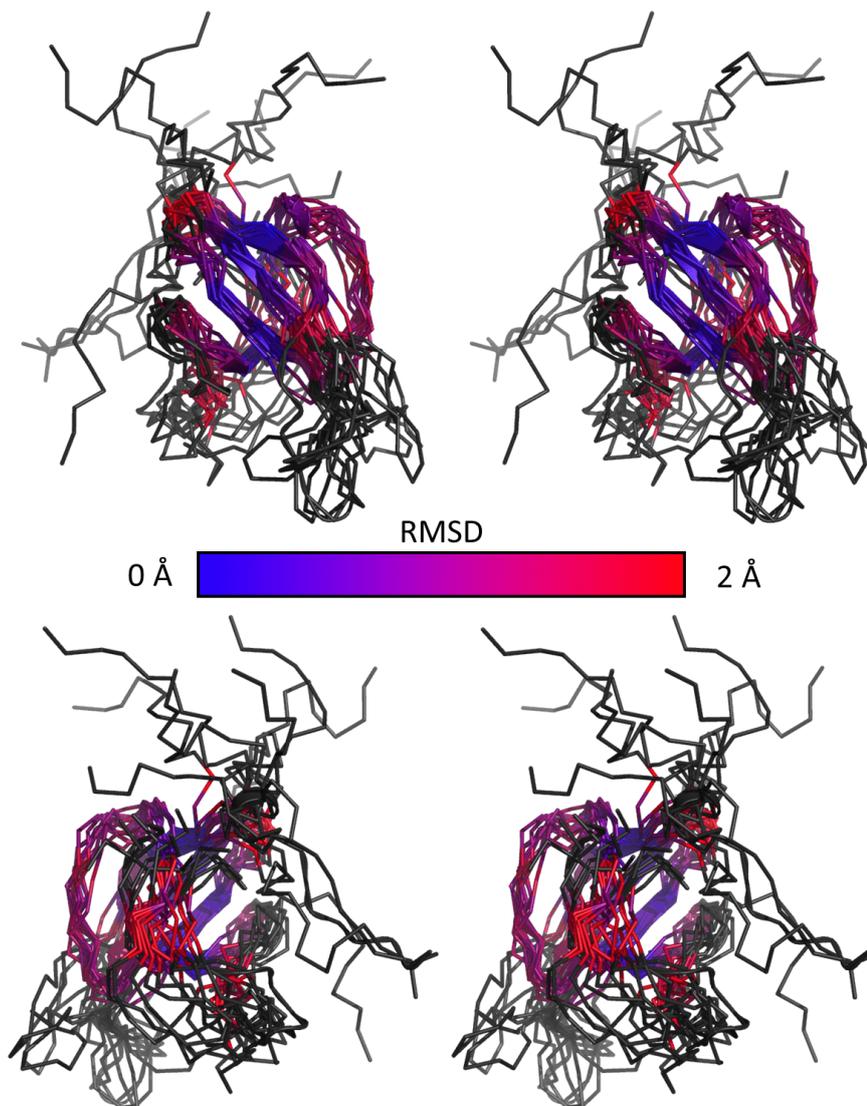
La protéine associée au code 2K5D est SAG0934 (code Uniprot : Q8E006), issue de *Streptococcus agalactiae serotype V*. Cette structure a été déterminée en 2008 dans le cadre d'un programme de génomique structurale, et n'a pas été publiée. Elle a cependant été citée dans la publication de Thomas *et al.*, 2013, comme homologue de la protéine HelP de l'ICEBs1 de la bactérie à Gram positif *Bacillus subtilis*. HelP est démontrée comme nécessaire à l'hélicase hôte (PcrA) pour progresser au-delà de l'origine de transfert oriT de l'ICEBs1. Ces informations fonctionnelles valident le modèle RMN préliminaire d'OrfM et laissent penser qu'OrfM comme SAG0934 pourraient jouer un rôle similaire à celui de HelP. De façon intéressante, c'est cette même protéine qui avait été initialement identifiée comme homologue de OrfM et OrfL sur la base des analyses de séquences (Dahmane, 2017). Cette faible similarité de séquences entre protéines au même rôle potentiel dans les T4SS de bactéries à Gram positif rejoint la faible similarité de séquences déjà constatée entre OrfG et ses homologues Gram positif.

Le code PDB 2K50 est quant à lui associé à la structure d'une protéine de *Methanobacterium thermoautotrophicum*, semblable à un domaine du facteur de réplication A, et déterminée par RMN

dans le cadre d'un programme de génomique structurale (et non publiée). Ces facteurs de réplication A (ou "*Replication Factor A*", RPA) auraient un rôle important dans la réplication, la réparation et la recombinaison de l'ADN (Sakaguchi *et al.*, 2009).

Parmi les résultats plus éloignés de la superposition DALI, d'autres protéines comme PriB (Code PDB : 2CCZ) auraient aussi des fonctions associées aux facteurs de réplication (Huang *et al.*, 2006). Dans cette dernière structure, on peut d'ailleurs observer une molécule d'ADN simple brin modélisée. Dans l'hypothèse où HelP posséderait un motif fixateur de l'ADN appelé motif OB pour "*Oligonucleotide Binding*" (Thomas *et al.*, 2013), il serait possible de retrouver un motif similaire sur OrfM et OrfL. Notons toutefois que la plupart des structures associées à ce terme de facteur de réplication A sont issues de programmes de génomiques structurales et ne sont pas (encore) publiées.

FIGURE 66 – Calculs de structure préliminaires d'OrfM



Superposition mTM-align de 14 modèles de basse énergie de la protéine OrfM générés par ARIAWeb en vue stéréo. MPSA_Viewer a été utilisé pour colorer l'alignement structural en fonction du RMSD pour chaque position des carbones C α dans la séquence d'acides aminés. En bleu, le RMSD est proche de 0 Å tandis que le rouge indique que le RMSD est proche de 2 Å. Au delà de 2 Å, les positions des acides aminés sont colorées en noir.

VIII.5 Discussion

OrfM est une petite protéine de 98 acides aminés qui fonctionnerait de paire avec OrfL comme un facteur de processivité d'hélicase. L'étude RMN a permis l'attribution des noyaux ^1H , ^{15}N et ^{13}C de l'essentiel de la chaîne principale. Sur l'ensemble des spectres collectés, il reste les spectres CC(CO)NH, H(CCO)NH et NOESY-HSQC- ^{13}C à analyser, ce qui devrait permettre de compléter l'attribution ^1H et ^{13}C des chaînes latérales, notamment des cycles aromatiques, et de générer des contraintes supplémentaires afin de préciser la structure qui est en cours de résolution. Les résultats préliminaires de modélisation sous contraintes RMN révèlent une topologie globale s'apparentant au motif "OB-fold" (Youkharibache *et al.*, 2019), caractérisé par un tonneau β avec deux feuillets antiparallèles β_1 - β_2 - β_3 et β_5 - β_4 - β_1 partageant le même brin β_1 et se positionnant orthogonalement. Dans OrfM, le premier feuillet apparaît assez nettement, cependant les brins β_1 et β_2 sont particulièrement longs et forment une épingle à cheveux dont seules les extrémités sont bien définies. Les résonances ^1H , ^{15}N de l'asparagine 23, l'alanine 24 et le glutamate 32, localisés dans la partie centrale de l'épingle, n'ont pu être attribuées. Le faible nombre de contraintes qui en découle semble être la cause principale de la valeur élevée du RMSD observée sur tout le reste de l'épingle, puisque les données de relaxation indiquent une flexibilité relativement limitée et l'existence d'échange conformationnel qui affecte seulement quelques résidus. Dans l'état actuel d'avancement, le deuxième feuillet est moins clairement défini, notamment au niveau du brin β_4 qui est suivi d'une région où très peu de contraintes ont été trouvées. En effet, les signaux correspondant aux résidus 71-82 sont de faible intensité, une proline est présente en position 70 et les résonances ^1H et ^{15}N des résidus 72 et 73 n'ont pu être attribuées. Les expériences de relaxation montrent d'ailleurs dans cette région une dynamique complexe pour la plupart des résidus, avec une flexibilité accrue, des mouvements à l'échelle de la pico- à nanoseconde mais aussi de la micro- à milliseconde. Une analyse méticuleuse des spectres non encore exploités pourrait aider à lever des ambiguïtés d'attribution et à caractériser plus précisément cette région. Certains paramètres RMN présentés sur la Figure 61, page 154 seraient compatibles avec la présence d'un tour d'hélice α juste avant le brin β_4 , comme souvent observé dans les protéines à "OB-fold". De même, il n'est pas exclu qu'il y ait une extension du brin β_4 en 71-73 et un tour d'hélice entre 74 et 79 mais ces hypothèses reposant sur seulement quelques indices nécessiteraient d'être corroborées par des éléments complémentaires. De la même manière, le tour d'hélice à la fin du segment N-terminal 1 à 14 globalement très flexible n'est actuellement pas

fermement défini. Une amélioration de la résolution de la structure en solution d'OrfM peut être espérée en poursuivant l'analyse des spectres, toutefois les résultats de relaxation laissent prévoir que certaines régions resteront mal définies, car peu structurées, par exemple dans la région N-terminale et dans le segment 71-86. L'étude dynamique a en outre permis d'estimer le temps de corrélation τ_c de la protéine à 7,3 ns à 25°C. Cette valeur est dans la gamme attendue pour une protéine de 11 kDa, ce qui indique qu'OrfM est majoritairement monomérique en solution, dans les conditions d'expérience utilisées. Si une forme dimérique est présente, elle doit l'être en faible proportion. Ces résultats vont alors dans le même sens que ceux de nos collaborateurs qui ont trouvé par diffusion de la lumière multi-angle (SEC-MALS) 95 % d'état monomérique et 5 % d'état dimérique (N. Soler, communication personnelle).

Des expériences d'interaction ont été menées par RMN avec OrfL, dans le but de cartographier la zone d'interaction au niveau de la protéine OrfM. Le principe consiste à ajouter OrfL non marqué à un échantillon d'OrfM marqué ^{15}N et de repérer par une expérience 2D HSQC- ^{15}N les résidus d'OrfM qui voient leurs déplacements chimiques perturbés par OrfL. Néanmoins, de telles variations de déplacements chimique n'ont pu être observées pour des ajouts de 0,1 à plus de 2 équivalents d'OrfL. Une diminution brutale de l'intensité des signaux a été à la place constatée sur l'ensemble des pics du spectre 2D HSQC- ^{15}N . Ceci peut signifier une interaction mais il est difficile de savoir si elle est spécifique. Les forts élargissements de signaux observés pourraient être dus à la formation d'un complexe de grande taille et/ou à un équilibre de formation du complexe se faisant à une vitesse d'échange "intermédiaire". Il faudrait en fait se trouver dans un régime d'échange "rapide" pour observer des variations de déplacement chimique sans élargissement important de signaux. En effet, l'évolution d'un spectre RMN lors de la formation d'un complexe dépend de la constante de vitesse de dissociation k_{off} par rapport à la différence de fréquence de résonance du noyau entre la forme complexée et la forme libre noyau. Pour être dans un régime d'échange rapide, il faut que la constante k_{off} soit très supérieure à cette différence de fréquence, qui est typiquement de l'ordre de $3\,000\text{ s}^{-1}$ pour des expériences menées sur un spectromètre 600 MHz. Cela correspond très approximativement à un K_d qui doit être supérieur à $3\ \mu\text{M}$ (Williamson, 2013). Bien que cette valeur ne soit qu'un ordre de grandeur, obtenue en supposant le modèle le plus simple d'une fixation sur un seul site, elle permet d'expliquer les difficultés rencontrées. D'après les résultats obtenus par SwitchSense et Titration Calorimétrique Isotherme (ITC), la valeur du K_d est de 300 nM (N. Soler, communication personnelle), qui est inférieure au seuil de $3\ \mu\text{M}$ et correspond à une

vitesse d'échange intermédiaire. Des changements de tampon (Tris, HEPES), de pH (de 7 à 6,5), de la concentration saline (de 200 mM à 500 mM) et de la température (de 25 à 35°C) ont été testés mais n'ont malheureusement pas apporté d'amélioration significative. D'autres essais n'ont pu être effectués car limités par les quantités disponibles de protéines OrfM marquée et OrfL.

A partir de la structure préliminaire d'OrfM, une similarité a été trouvée avec des modèles de la banque de données des structures de protéines, révélant un motif de fixation aux oligonucléotides, appelé motif OB ("*oligonucleotide binding*"). Cela est tout à fait plausible dans le cadre d'un événement de conjugaison permettant de transporter un ADN. OrfM et OrfL avaient été identifiées comme homologues à la protéine HelP, qui elle-même avait été associée à la protéine SAG0934 (Thomas *et al.*, 2013). La structure de cette dernière est similaire au modèle préliminaire d'OrfM. Le calcul de structure, encore inachevé, semble donc cohérent. Il n'avait pas été possible de prédire cette similarité à partir de la structure primaire d'OrfM car les similarités de séquences sont trop faibles. En rassemblant toutes ces informations, on peut émettre l'hypothèse qu'OrfM (et OrfL) possède(nt) un rôle dans la fixation à l'ADN dont les conditions restent à définir dans le cadre de l'évènement de conjugaison en lien avec l'ICE.

IX MPSA_Viewer

Les comparaisons structurales entre protéines permettent d'identifier des régions conservées dans l'espace et donc probablement dans leur fonction, sans que la séquence qui leur corresponde ne soit la même. En effet, lorsque des protéines ont la même fonction, il est possible qu'elles présentent des similarités de séquence très locales, par exemple au niveau d'un site actif, tandis que la conformation globale conservée ne se traduit que par des identités de séquence bien inférieures aux 30% habituellement attendus pour des homologues structuraux. Il en est de même pour les sites de fixations aux ligands, sites d'interactions protéine-protéine ou protéine-ADN. Notamment, l'évolution convergente conduit à de telles situations, avec des repliements comparables malgré de très faibles similarités de séquences.

L'identité de séquence ne pouvant servir de base de comparaison dans ces situations, il faut utiliser d'autres outils pour ces cas spécifiques. Ils ne doivent prendre en compte que les informations spatiales des atomes de la chaîne principale, traduisant ainsi le repliement conservé malgré la séquence différente. Il en existe plusieurs méthodes, et celle que nous avons retenue comme bien appropriée à la problématique de la protéine OrfG est TM-Align (Zhang & Skolnick, 2005). Cette méthode permet la superposition structurale d'une paire de protéines, évaluée par le score "TM" (mesure de la similarité structurale), et s'est montrée particulièrement efficace dans notre cas pour constituer une matrice en vue d'une comparaison multiple, réalisée par le serveur mTM-align (Dong, Pan *et al.*, 2018; Dong, Peng *et al.*, 2018). En effet, parmi tous les serveurs de comparaison multiple testés sur les différents domaines VirB8, mTM-align a fourni le résultat le plus convaincant. Toutefois, le format et la représentation donnés par le serveur après traitement étaient très limités. L'alignement au format "fasta" qui en découle est pertinent, mais le fichier PDB que l'on peut télécharger a perdu beaucoup d'informations durant le traitement (identifiant de chaîne, position de l'acide aminé dans la séquence originale) en plus de n'avoir pas traité les acides aminés marginaux tels que les sélénométhionines, laissant des trous dans l'alignement. La représentation visuelle sur la page des résultats montre une structure 3D avec une couleur uniforme appliquée pour chaque position de l'alignement lorsque les distances entre tous les carbones alpha de cette même position ne dépasse pas 4 Å (Voir Figure 67, page 172).

Cette représentation très intéressante quoiqu'incomplète couplée à l'absence d'un fichier PDB

options pour débiter l'alignement est de renseigner dans le formulaire l'ensemble des codes PDB à traiter, ou bien téléverser un fichier compressé zip contenant toutes les structures désirées. Un maximum de 30 structures est accepté, pour un nombre supérieur la version installée en local est impérative. L'alignement en lui-même est très rapide. Les résultats présentés comprennent l'arbre phylogénétique et un alignement de séquences, tous les deux basés sur la superposition structurale multiple et le score TM de chaque paire de protéines. Ensuite est dessinée la visualisation 3D de l'alignement avec une coloration des zones "conservées". On pourra toutefois regretter les défauts suivants :

(1) mTM-align n'est pas encore capable de gérer les acides aminés spéciaux tels que les sélénométhionines.

(2) Il n'est pas encore possible, lorsque l'on envoie les structures sous un format compressé, de renseigner les identifiants de chaînes du fichier PDB qui seront utilisées pour l'alignement. Ces deux limitations sont contournables en modifiant manuellement les fichiers PDB ce qui peut s'avérer extrêmement fastidieux lors d'alignements à grande échelle.

(3) La coloration par position dans l'alignement se limite à mesurer les distances entre tous les carbones C α sur cette position, et si la distance maximale est inférieure à 4 Å, alors y sera appliquée une coloration uniforme pour toutes les positions respectant cette condition.

(4) Cette coloration n'est disponible que sur la visualisation très simplifiée du site.

(5) Les chaînes latérales des acides aminés ne sont pas du tout montrées dans la visualisation.

(6) La superposition structurale et l'alignement des séquences (basé sur la superposition des structures) que l'on peut télécharger sont très simples, notamment le fichier PDB qui perd l'ensemble des informations qui peuvent ensuite être utilisées dans PyMOL pour travailler la représentation.

L'ensemble de ces défauts mineurs ne sont pas suffisants pour rejeter le résultat, car de notre point de vue, sur un projet tel que les VirB8 contenant de nombreuses protéines avec de très faibles identités de séquences mais un repliement bien conservé, mTM-align reste le meilleur algorithme de superposition structurale, outil incontournable pour un alignement des séquences associées. De plus, mTM-align propose un arbre "phylogénétique" basé sur cette superposition structurale.

Enfin, PyMOL étant l'un des logiciels de représentation tridimensionnelle des modèles les plus utilisés pour les protéines et étant compatible avec les fichiers de script python, il était évident de le choisir pour améliorer la représentation fournie par le serveur mTM-align.

IX.2 Description du fonctionnement de MPSA_Viewer

Dans cette partie seront détaillées les différentes implémentations réalisées dans le script de MPSA_Viewer afin de contourner ou de régler les différents problèmes de mTM-align, dans le but d'obtenir une représentation fidèle et parlante dans PyMOL.

IX.2.1 Mise en place et traitement des données initiales

Afin de faciliter la tâche de l'utilisateur, le script MPSA_Viewer compresse automatiquement les structures choisies en un fichier zip qui peut ensuite être utilisé pour lancer l'alignement multiple par mTM-align. Avant cela, plusieurs étapes sont nécessaires, notamment le paramétrage du logiciel MPSA_Viewer en fonction de ce que l'utilisateur souhaite.

(1) Choix des paramètres du programme par l'utilisateur

De très nombreux paramètres peuvent être ajustés par l'utilisateur pour personnaliser la visualisation finale de la superposition structurale multiple dans PyMOL. En l'occurrence, il est possible de choisir les couleurs, les types de représentations, les valeurs limites pour les distances calculées, etc.

(2) Le fichier est-il disponible sur la PDB ?

Le script analyse automatiquement le nom d'entrée de chaque protéine que l'utilisateur soumet. Si ce nom présente une nomenclature correspondant à la PDB disponible sur internet, le fichier .pdb est automatiquement téléchargé puis stocké dans un dossier. Sinon, l'utilisateur doit fournir lui-même le fichier de coordonnées du modèle concerné, au format .pdb également.

(3) Tous les modèles sont-ils à disposition ?

Il faut ensuite vérifier que toutes les structures ont été téléchargées correctement, ou que l'utilisateur a bien ajouté toutes ses structures manuellement.

(4) Modification des fichiers PDB

mTM-align ne considère que les 20 acides aminés standards et n'accepte donc pas les acides aminés spéciaux tels que les sélénométhionines, tout comme il ne lit pas les conformations multiples des acides aminés ni les ligands. Le script MPSA_Viewer crée donc une copie simplifiée des fichiers PDB ne contenant que les informations compatibles avec mTM-align.

(5) Compression des données

L'ensemble des structures modifiées est ensuite compressé dans un fichier zip. L'utilisateur n'a plus qu'à le soumettre au serveur mTM-align ("<https://yanglab.nankai.edu.cn/mTM-align/>").

IX.2.2 Traitement des résultats

Maintenant que les structures ont été traitées par mTM-align, un lien vers la page des résultats est fourni. A partir de ce lien, le script MPSA_Viewer peut automatiquement télécharger les résultats fournis par le serveur, à savoir un fichier PDB simplifié contenant toutes les structures superposées, et un alignement de séquences fasta basé sur cette superposition de structures.

(1) Création d'une superposition avec les modèles complets

A partir du fichier .pdb simplifié de la superposition structurale fourni par mTM-align, MPSA_Viewer utilise PyMOL pour rétablir les modèles originaux qui contiennent les informations précédemment retirées avant de lancer mTM-align. Cela permet notamment de retrouver les sélénométhionines, les conformations multiples ou encore les ligands. Grâce à la méthode d'alignement SUPER de PyMOL et la création d'un fichier de scripting compatible avec le logiciel, MPSA_Viewer indique à PyMOL de créer un fichier .pdb de superposition avec les structures originales. Il sauvegarde ensuite les modèles superposés un à un pour une utilisation ultérieure.

(2) Calcul de distance entre les résidus à chaque position de l'alignement de séquence

MPSA_Viewer analyse chaque position de l'alignement de séquences dès lors que chaque structure étudiée y présente un résidu. Si un acide aminé est absent dans l'une des séquences, cette position est passée. Cela permet de négliger les extrémités et les boucles flexibles qui ne sont pas pertinentes lors d'une superposition multiple. Ensuite, pour une position retenue pour l'analyse, le logiciel extrait les coordonnées des carbones C α superposés, et calcule l'écart quadratique moyen (RMSD) correspondant qu'il stocke pour son exploitation ultérieure.

(3) Comparaison de la nature des acides aminés pour chaque position

En même temps que le calcul de RMSD, une comparaison de la nature des résidus est réalisée. Si l'utilisateur a activé l'option en paramétrant `MPSA_Viewer` à son lancement, les résidus identiques ou de natures similaires en une position seront mis en exergue sur la visualisation finale, permettant d'identifier certaines zones conservées tel qu'un site actif ou une interface d'interaction. Ces données sont également stockées pour une utilisation ultérieure.

(4) Création d'un script PyMOL pour la visualisation finale

Une fois toutes les informations de distances et de nature des acides aminés obtenues, `MPSA_Viewer` va créer un long fichier script qui permettra de colorer, représenter et donc mettre en valeur la superposition structurale dans PyMOL. Ceci est réalisé en utilisant les informations des deux étapes précédentes.

IX.3 Exemples d'utilisation de `MPSA_Viewer`

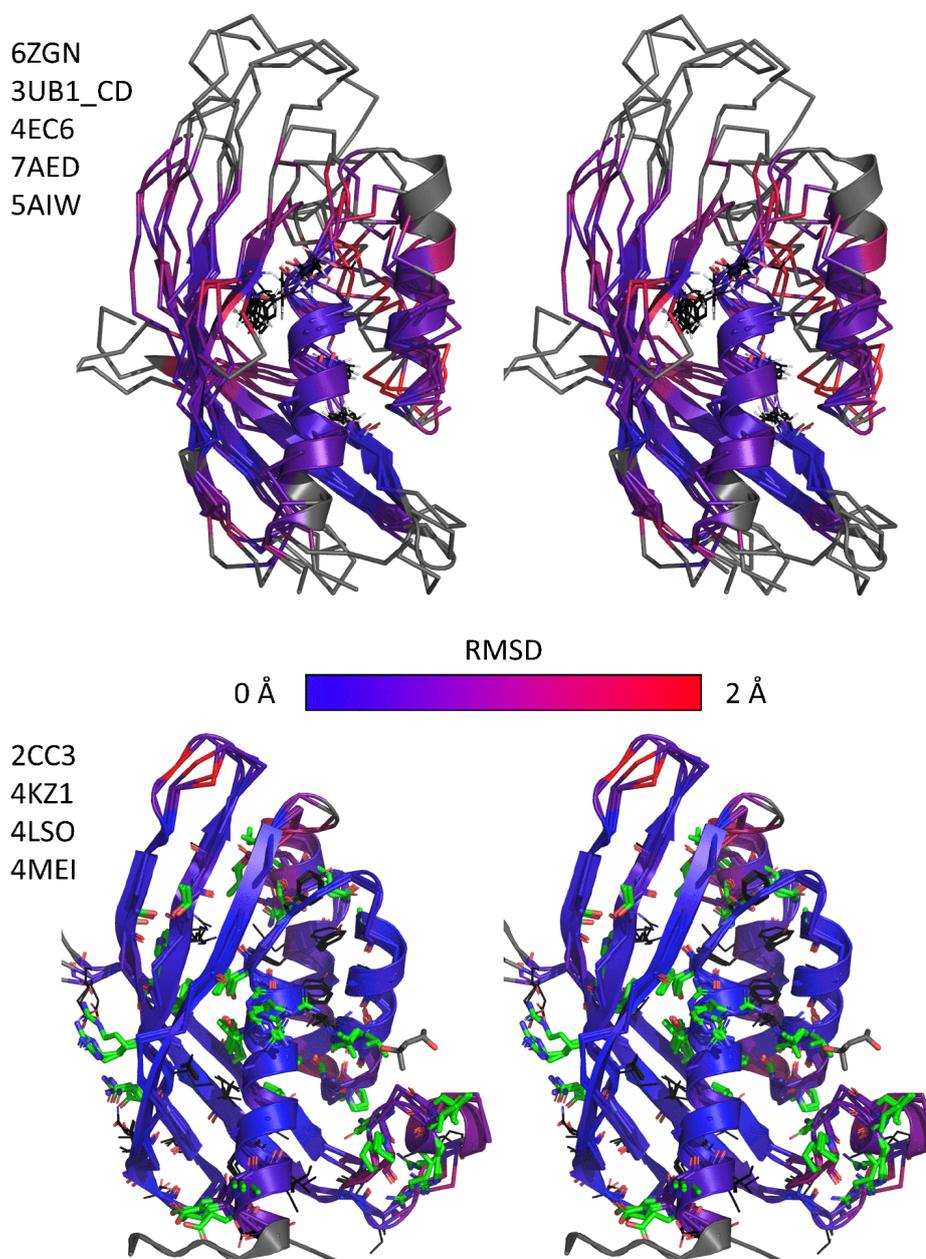
Le premier cas présenté ci-dessous concerne les protéines VirB8-like et n'est donc pas à proprement parler un nouvel exemple, puisqu'une partie en a déjà été présentée dans l'article (Cappele *et al.*, 2021). Il constitue cependant un bon point de départ pour comprendre comment le logiciel fonctionne, avant de montrer un exemple d'application sur une autre famille de protéines, les glutathion transférases. Ces deux exemples ont en commun l'analyse de protéines présentant une très faible identité de séquences mais un repliement globalement conservé.

IX.3.1 Domaines VirB8-like

Dans la Figure 68, page 178 sont montrés deux exemples différents de représentation avec `MPSA_Viewer`. Le premier exemple reprend les domaines VirB8-like issus de bactéries à Gram positif introduits dans l'article, et y ajoute PrgL (code PDB : 7AED). Ces 5 protéines ont deux à deux des identités de séquences faibles (< 10%) ainsi qu'une absence totale de résidus conservés sur l'alignement d'ensemble. Seulement quelques acides aminés de nature similaire (hydrophobes) sont visibles au coeur des domaines sur la représentation tridimensionnelle. La superposition structurale montre un nombre de résidus colorés (et donc alignés) plutôt faible. Les teintes des différentes zones sont plus souvent violet/rouge que bleu, indiquant un RMSD élevé sur ces positions. Les régions

représentées en gris correspondent logiquement aux boucles, très variables en position (ayant donc des RMSD supérieurs à 2 Å) ou en longueur (étant alors rejetées de l'alignement). A l'inverse, le second exemple proposé est un cas où les identités de séquences par paires sont plus élevées (> 25%) ce qui est logiquement associé à une bien meilleure superposition. En effet, les domaines VirB8-like sélectionnés ici font partie de la même classe II (Cappele *et al.*, 2021). Au sein de la superposition, on repère de nombreux résidus conservés (représentés en vert) dans le coeur et à la surface de la protéine.

FIGURE 68 – Superposition structurale de protéines VirB8-like par MPSA_Viewer



Deux exemples de superpositions structurales réalisées par mTM-Align, analysées par MPSA_Viewer et représentées par PyMOL, en stéréo. En haut, 5 domaines VirB8-like de bactéries à Gram positif. En bas, 4 domaines VirB8-like de bactéries à Gram négatif appartenant à la classe II (Cappele *et al.*, 2021). Les acides aminés identiques sont représentés en sticks verts; les acides aminés similaires en lignes fines noires. Le squelette carboné de chaque structure est en représentation "ribbon", sauf 6ZGN (exemple en haut) et 2CC3 (exemple en bas) qui sont représentés en "cartoon". Ce squelette est coloré en fonction du RMSD des carbones C α pour chaque position des acides aminés, avec une coloration allant du bleu lorsque le RMSD est nul, au rouge lorsqu'il atteint 2 Å. La coloration grise correspond aux positions d'acides aminés qui ne sont pas alignées ou ayant un RMSD des carbones C α supérieur à 2 Å.

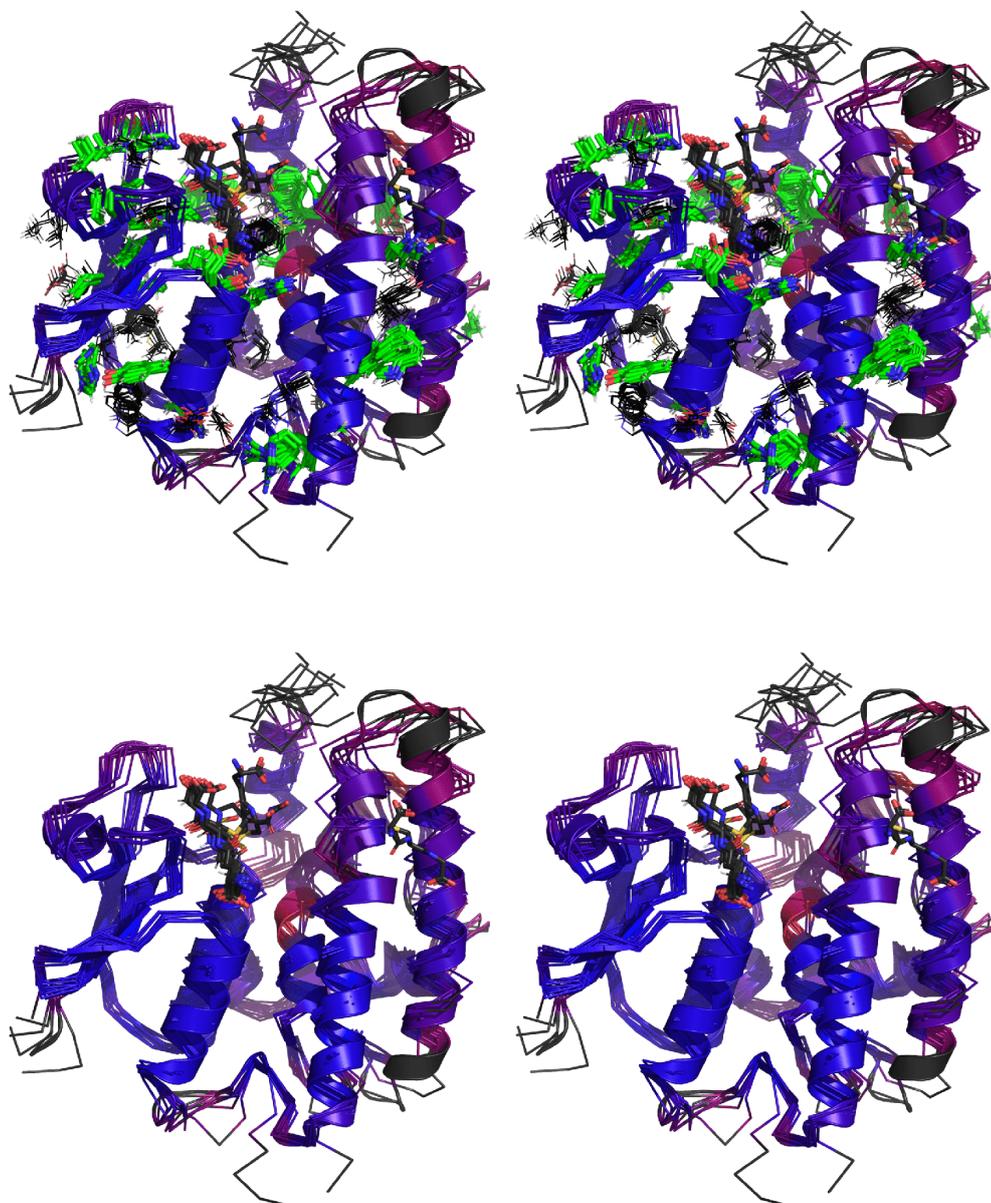
IX.3.2 Glutathion transférases de classe Tau

Le second exemple présenté concerne les glutathion transférases (GST) qui constituent des modèles intéressants pour des alignements structuraux multiples. Les GST cytosoliques sont principalement connues pour leur rôle dans la phase de conjugaison du processus de détoxification des xénobiotiques. De nombreux autres rôles leur sont attribués comme le transport de ligands. Près de 500 structures de ces enzymes sont disponibles dans la PDB. Les GST cytosoliques sont regroupées en classes portant des noms de lettres grecques (Sylvestre-Gonon *et al.*, 2019). Le CRM2 s'intéresse depuis de nombreuses années à cette famille d'enzymes en travaillant plus particulièrement sur les GST des règnes végétal et fongique. Il a été choisi de tester MPSA_Viewer sur une seule classe pour montrer l'utilité de l'outil dans le cas de protéines structurellement proches. La classe Tau des GST a été choisie car c'est une classe spécifique et importante dans les plantes et un nombre significatif de modèles tridimensionnels sont disponibles. Même si les GST de la classe Tau sont dimériques, les superpositions ont été réalisées à partir d'une seule sous-unité des isoformes (voir Figure 69, page 181). 9 structures proviennent de la PDB de divers organismes et 2 d'une étude en cours au laboratoire. Ces deux dernières sont celles de deux proches paralogues du peuplier (PtGSTU19 et PtGSTU20) qui présentent des propriétés biochimiques distinctes en ce qui concerne la fixation de flavonoïdes (C. Didierjean, communication personnelle).

La figure 69 représente la superposition des monomères des GST sélectionnées avec les résidus identiques et similaires et sans ceux-ci. La sous-unité est divisée en deux domaines, le domaine N-terminal composé d'un feuillet mixte à 4 brins entouré de trois hélices (à gauche) et le domaine C-terminal constitué de 6 hélices (à droite). Le site actif se trouve à l'interface entre les deux domaines et contient deux poches : une pour le glutathion (représenté en bâtonnets) et l'autre pour le substrat électrophile. La superposition montre comme attendu que les régions les mieux superposées sont les hélices et les brins. Toutefois il est possible de remarquer la couleur rouge au centre de l'hélice $\alpha 6$ (au coeur de la protéine) qui révèle une variabilité structurale à proximité du site actif. Cette région est intéressante à noter car c'est cette région qui fixe les flavonoïdes dans la PtGSTU20. Pour la PtGSTU19, il n'a pas été possible d'obtenir les structures en complexe avec ces mêmes ligands. Cette région contient un résidu tryptophane (W161 dans la PtGSTU20) conservé qui présente une variabilité conformationnelle modifiant de manière substantielle la taille de la poche (C. Didierjean, communication personnelle). L'alignement des séquences issu de la superposition des structures

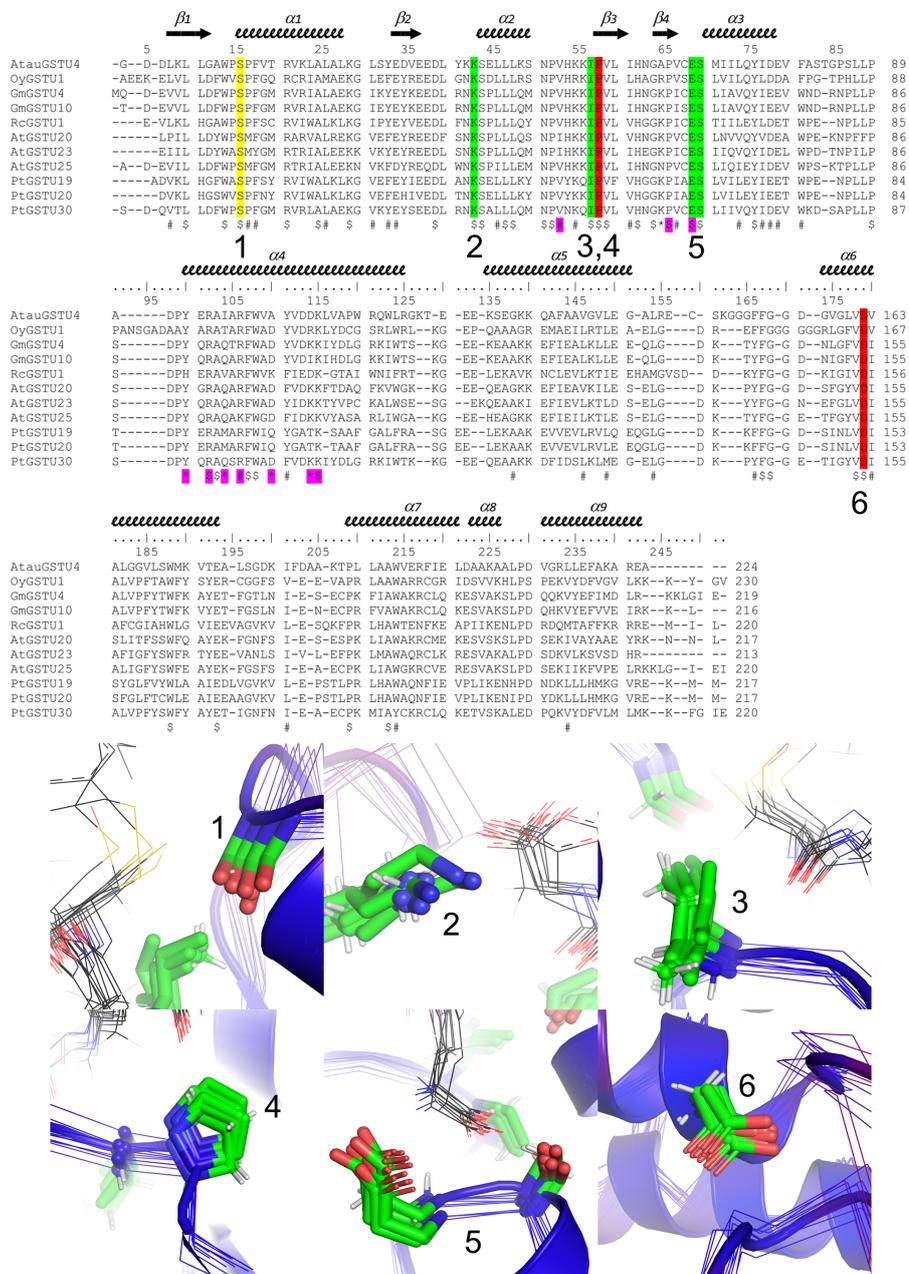
montre, comme déjà observé, que la similarité est la plus importante dans le domaine N-terminal (brin $\beta 1$ à hélice $\alpha 3$) (voir Figure 70, page 182). Toutefois l'information sur la distribution spatiale des résidus conservés et similaires n'est que partielle sur l'alignement. MPSA_Viewer donne clairement cette distribution spatiale. La figure 69 mais plus encore la session PyMOL révèle trois clusters majeurs de similarité. Le plus étendu couvre le site actif et les deux autres constituent les coeurs hydrophobes des domaines N- et C-terminal. De façon intéressante ces deux types de sites avaient été respectivement qualifiés de "super-site" et "noyau de repliement" dans une étude fondamentale sur la recherche des positions universellement conservés dans les repliements des protéines (Mirny & Shakhnovich, 1999). Le premier a été qualifié de "super-site" car il peut correspondre à un site actif de protéines ayant la même structure tertiaire mais n'ayant pas une fonction commune. Le "noyau de repliement" est directement relié à la stabilité de la structure et/ou à sa capacité à se replier. Il apparaît intéressant de poursuivre cette étude en l'étendant à plusieurs classes pour vérifier en particulier la conservation des clusters.

FIGURE 69 – Superposition structurale de GST de classe Tau



Superposition structurale des 11 GST choisies dans la classe Tau, réalisé par mTM-align puis représenté dans PyMOL grâce à MPSA_Viewer, en vue stéréoscopique. Le squelette carboné de chaque structure est en représentation "ribbon", sauf AtauGSTU4 représenté en "cartoon". La coloration du squelette varie du bleu lorsque le RMSD de superposition des C α est nul au rouge lorsqu'il atteint 2 Å. La coloration grise correspond aux positions d'acides aminés qui ne sont pas alignés ou ayant un RMSD des carbones C α supérieur à 2 Å. Les ligands, correspondant aux molécules de la famille du glutathion, sont indiqués en bâtons noirs. La vue du haut présente les chaînes latérales des acides aminés conservés (en bâtons verts) et les acides aminés de nature similaire (en lignes noires) tandis que la vue inférieure ne montre pas les chaînes latérales des acides aminés et permet donc de mieux analyser la superposition des chaînes principales.

FIGURE 70 – Les Glutathion transférases de classe Tau retenues pour tester MPSA_Viewer



En haut, alignement des séquences de 11 GST de la classe Tau. Le code des organismes est le suivant : *Aegilops tauschii* (Atau), *Oryza sativa* (Oy), *Glycine max* (Gm), *Ricinus communis* (Rc), *Arabidopsis thaliana* (At), *Populus trichocarpa* (Pt). L'acide aminé surligné en jaune est l'acide aminé catalytique. Les acides aminés conservés en interaction avec le glutathion sont surlignés en vert, et ceux invariants au sein des GST, en rouge. Les positions surlignées en violet sont des résidus impliqués au moins dans une structure dans la stabilisation du dimère. Les positions conservées dans les 11 structures sont marquées par le signe \$ et les positions similaires par le signe #. En bas, les vues rapprochées des positions conservées 1, 2, 3, 4, 5 et 6 sont proposées à partir de la représentation réalisée grâce à MPSA_Viewer. Elles montrent que tous ces résidus ont bien été détectés par MPSA_Viewer.

IX.4 Avantages et limitations de MPSA_Viewer

IX.4.1 Simplicité

MPSA_Viewer est *open source* et est soumis à une licence publique. Toute personne peut donc modifier et améliorer le logiciel puis le distribuer, tant que la licence est conservée, ainsi qu'une mention des auteurs initiaux. Chacun devrait et doit pouvoir adapter un programme à ses envies.

Le script est également doté d'une interface utilisateur. Pour modérer le temps dévolu à son développement, l'interface est limitée à une simple console et quelques commandes, suffisantes pour utiliser les fonctions du logiciel et les appliquer à un projet. La première version d'une documentation d'installation et d'utilisation a également été créée, et est disponible en annexe page 213.

L'un des arguments pour l'utilisation d'un logiciel de ce type est la façon simple dont il permet l'analyse d'un alignement structural en passant par la visualisation tridimensionnelle de ses différentes caractéristiques. Par exemple, observer directement sur la structure la localisation des acides aminés conservés au sein d'un groupe de protéines est une possibilité intéressante que peu de logiciels offrent. De même, mettre en lumière les zones les plus (ou les moins) variables d'un ensemble de chaînes polypeptidiques permet souvent un lien avec la fonction des protéines étudiées.

Cependant, cette simplicité induit une certaine rigidité pour l'utilisateur. Ainsi, lors de la définition des groupes de résidus que MPSA_Viewer doit considérer comme similaires, le logiciel ne tolère aucune erreur, sous peine d'arrêt brutal. Également lorsqu'une vingtaine de structures sont par exemple superposées, si en une position donnée une seule des structures a un acide aminé non modélisé, la position ne sera pas prise en compte quand bien même le résidu est identique chez les 19 autres structures. Il faudrait implémenter tout au long du logiciel plusieurs moyens statistiques qui ajouteraient un certain degré de flexibilité dans la représentation.

IX.4.2 Dépendance et performance

Malheureusement, MPSA_Viewer est dépendant des logiciels principaux qui lui permettent d'exister, notamment mTM-align. La robustesse et la puissance de l'algorithme TM-align permettent ce type d'analyses puisque la superposition structurale résultante est (à nos yeux) bien plus réussie

qu'avec d'autres outils d'alignement structuraux classiques. Également mais en moindre mesure, le programme est dépendant d'un logiciel de modélisation, PyMOL. L'ensemble des fonctions utilisées ont été développées pour être compatibles avec ce dernier. Transposer la modélisation sur un autre logiciel comme Chimera est possible pour ceux qui le souhaitent. En revanche mTM-align est totalement indispensable. Il existe bien une version portable de mTM-align, mais elle ne fonctionne que sur Linux. L'automatisation sur plusieurs systèmes d'exploitation n'est donc pas possible en l'état sans un travail sur le code source du programme. Enfin, le site ne supporte qu'un alignement structural de moins de 30 structures, il faut donc utiliser la version linux pour analyser un nombre supérieur de structures.

MPSA_Vieweur a été développé en Python, par un étudiant en biologie qui n'avait jamais codé auparavant. L'écriture n'en est sans doute pas optimale et la rapidité du logiciel est très vite impactée par le nombre de structures que l'on souhaite aligner. Un développeur expérimenté y pointerait sans doute plusieurs défauts de débutant. Par ailleurs, PyMOL est assez frileux lorsqu'on lui demande d'ouvrir des scripts de plusieurs dizaines de milliers de lignes. Cette situation peut survenir à partir de la version actuelle de MPSA_Vieweur si l'on travaille avec des ensembles dépassant la trentaine de structures.

IX.5 Discussion

MPSA_Vieweur est un outil intéressant pour faciliter l'analyse d'une superposition multiple de structures, par la visualisation. De nombreuses fonctionnalités y sont implémentées et plusieurs paramètres sont personnalisables pour que l'expérience de l'utilisateur soit la plus libre possible. Toutefois, plusieurs défauts restent présents, notamment un débogage encore incomplet, des fonctionnalités à modifier ou introduire pour gagner en souplesse, et une dépendance à deux logiciels externes. MPSA_Vieweur reste toutefois facile d'utilisation avec une interface console, et permet d'obtenir ses résultats en quelques minutes. Le format de sortie retenu, une session pour PyMOL, permet même la personnalisation à posteriori de la représentation des structures afin d'y mettre en valeur d'autres caractéristiques propres à la famille de protéines étudiées.

A plus long terme pourrait être introduit un autre concept : donner à l'utilisateur la possibilité de s'appuyer sur des matrices de similarité de type BLOSUM ou PFAM au lieu de définir manuellement

les groupes d'acides aminés similaires. Cette approche permettrait d'attribuer des scores de similarité à chaque position, à la place de l'actuel "tout ou rien". Pour cette position pourrait en découler un mode de représentation de l'acide aminé reflétant donc une évaluation plus conventionnelle de l'alignement de séquences, tout en ayant pourtant basé cet alignement sur la superposition des structures tridimensionnelles.

X Conclusion et Perspectives

Les systèmes de sécrétion permettent aux organismes de transporter différentes molécules au travers de leurs membranes et vers d'autres organismes lors de la conjugaison ou des mécanismes d'infection. Dans le cas du transfert horizontal de gènes, c'est-à-dire le transfert de gènes entre deux cellules, la conjugaison est l'un des principaux mécanismes employés par ces organismes. En l'occurrence, les systèmes de sécrétion de type IV (ou T4SS) ont été et sont encore étudiés pour mieux comprendre comment ces échanges ont lieu, en particulier les transferts de gènes de résistances aux antibiotiques qui représentent aujourd'hui un problème majeur dans le domaine médical. Chez les bactéries à Gram négatif, de nombreuses structures de protéines sont résolues, et des modèles structuraux et fonctionnels de T4SS existent, notamment chez *Agrobacterium tumefaciens* et *Legionella pneumophila* qui représentent respectivement les groupes A et B des T4SS de bactéries à Gram négatif. En ce qui concerne le groupe des bactéries à Gram positif, le nombre de structures évolue d'année en année, mais les modèles complets de leur T4SS ne sont pas encore disponibles.

En débutant le projet T4SS chez la bactérie à Gram positif *Streptococcus thermophilus* pour en étudier l'élément intégratif et conjugatif ICESt3, nous avons réussi à produire les premières informations structurales des protéines qui le composent. En particulier, OrfG est l'homologue de VirB8 qui est attendue pour être l'une des protéines importantes et indispensables à l'assemblage et au fonctionnement du T4SS. Nous avons pu résoudre la structure cristallographique du domaine central OrfG₆₄₋₂₀₄ puis de la protéine sans son domaine transmembranaire OrfG₆₄₋₃₃₁. Le repliement de chacun des deux domaines de cette construction est similaire à celui des protéines VirB8 des bactéries à Gram négatif, tout en présentant une identité de séquence très faible avec ces dernières. Une analyse fine de la superposition de tous les domaines VirB8 a permis de mettre en évidence la conservation globale du repliement, malgré l'absence d'acide aminé conservé en surface ou au coeur. Toutefois, plusieurs groupes se sont dégagés de cet alignement de structures, et nous avons proposé une première classification des domaines VirB8-like actuellement connus, publiée dans un article et discutée dans ce manuscrit. Il y apparaît notamment la séparation des groupes des VirB8-like Gram positif et négatif. Les différentes protéines utilisées dans cette comparaison ont été superposées grâce à mTM-align et analysées en utilisant le script MPSA_Viewer. Développé dans le cadre de ce projet, il allie la superposition multiple de structures réalisée par l'algorithme TM-Align et la qualité de modélisation de PyMOL. Ce script permet notamment d'apprécier visuellement les zones

alignées ainsi que les acides aminés conservés afin de faciliter la comparaison des structures et des séquences de protéines. Par ailleurs, la structure oligomérique des VirB8 Gram positif a été discutée bien qu'elle n'ait pas encore été observée dans un contexte biologique. En effet, l'observation d'un même arrangement trimérique original dans les structures cristallines d'OrfG et TcpC, une protéine de même organisation issue de bactérie à Gram positif, donne un sens biologique très probable à cette association. Cependant, il n'a pour l'instant été question que des domaines solubles d'OrfG sans avoir pu étudier la protéine complète, c'est-à-dire avec son domaine transmembranaire en N-terminal. En particulier, nous n'avons pas encore pu démontrer que la protéine interagit avec d'autres partenaires protéiques (ou autres!) dans un contexte plus proche du système de sécrétion. Des expériences de mutation et de crosslink peuvent notamment être mises en place pour confirmer la réalité du trimère dans un contexte biologique, et pourquoi pas en présence de son domaine transmembranaire. La purification et l'étude d'un complexe à l'échelle d'un T4SS pour l'ICE*St3* n'est pas encore à l'ordre du jour vu la complexité du travail, mais petit à petit les expériences à venir permettront d'éclaircir le rôle exact d'OrfG dans le T4SS de l'ICE*St3* chez *S. thermophilus*.

La protéine OrfM, supposée agir de paire avec OrfL comme facteur de processivité d'hélicase, a également été étudiée, cette fois par RMN, bien adaptée à la petite taille de cette protéine d'autant que les essais préliminaires de cristallisation n'avaient pas été encourageants. Après la production d'un échantillon doublement marqué ^{15}N et ^{13}C et l'acquisition de spectres standards, l'attribution des résonances ^1H , ^{15}N et ^{13}C du squelette de la protéine a été réussie à plus de 90%, et les premiers calculs de structure montrent un repliement en feuillet mixte à 5 brins, avec une boucle entre les deux derniers brins ainsi qu'une extrémité N-terminale peu structurées. La topologie globale trouvée s'avère similaire à celle de plusieurs protéines qui pourraient fonctionner comme des facteurs se liant à l'ADN simple brin (HelP, Thomas *et al.*, 2013). A la lumière de ces récents résultats, il semblerait que la recherche d'une interaction d'OrfM avec l'ADN mériterait d'être poursuivie. Un test avec de l'ADN double brin avait montré une absence d'affinité mais il serait intéressant de la réaliser avec l'ADN simple brin notamment. Si cette interaction existe, elle pourrait être cartographiée par RMN. De plus, nos collaborateurs ont observé qu'OrfM et OrfL améliorent l'activité de la relaxase Rel*St3* de l'élément mobile ICE*St3* (N. Soler, communication personnelle). Malheureusement, les expériences d'interaction menées par RMN entre OrfM et OrfL afin de mieux comprendre le rôle de ces protéines n'ont pas été concluantes dans les conditions testées. L'ajout d'OrfL à l'échantillon d'OrfM a entraîné une forte diminution de l'ensemble des signaux qui semble indiquer que la vitesse

de dissociation du complexe est trop faible pour que des variations de déplacement chimique puissent être observées sur le spectre RMN. Une autre explication pourrait être aussi la formation d'un complexe de très haut poids moléculaire. Concernant la structure d'OrfM, l'analyse des spectres CC(CO)NH, H(CCO)NH et NOESY-HSQC-¹³C, qui ont déjà été collectés, doit être effectuée afin de compléter l'attribution des chaînes latérales et d'ajouter de nouvelles contraintes de structure. Celles-ci devraient permettre d'obtenir une structure d'une qualité suffisante pour être déposée, publiée et analysée de façon plus complète dans le contexte des systèmes de sécrétion de type IV.

Il serait également possible de réaliser l'étude structurale d'OrfL par RMN, étape qui ne devrait pas poser de problème. En revanche, l'étude RMN des interactions entre OrfM et OrfL nécessitera une mise au point importante. La stoechiométrie du complexe OrfM/OrfL en solution n'étant pas encore établie, une alternative possible serait de co-exprimer et co-purifier le complexe OrfM/OrfL en vue d'une étude cristallographique : la surexpression des deux protéines non marquées en milieu riche étant déjà maîtrisée. La fonction des protéines OrfM et OrfL reste à préciser, en particulier leur lien avec la Relaxase RelSt3 (OrfJ). La structure de cette dernière n'a pas encore été résolue. En effet, la construction et les conditions de cristallisation actuelles n'ont pas encore permis l'obtention de cristaux exploitables. La modification de la construction, la mutation d'acides aminés prédits désordonnés ou l'étude d'un homologue d'un autre organisme sont des pistes envisagées. L'étude structurale des autres partenaires impliqués dans la conjugaison du T4SS de l'élément ICE*St3* est encore à un stade préliminaire, notamment l'hydrolase de peptidoglycane OrfA et la protéine de couplage OrfK. La caractérisation des interactions entre ces protéines sera nécessaire pour préciser le modèle de T4SS Gram positif. Pour autant, cette étude ouvre la voie vers la caractérisation structurale du T4SS chez *Streptococcus thermophilus*, une bactérie à Gram positif dont le système conjugatif est encore mal compris.

MATÉRIEL ET MÉTHODES

XI Production et purification des protéines OrfL et OrfM

XI.1 Souche et plasmide d'expression

La souche utilisée pour l'expression des protéines est *E. coli* BL21 (DE3) : F- *ompT lon hsdSB* (rB- mB-) *gal dcm*. Cette souche est compatible avec les vecteurs d'expression disposant d'un promoteur à l'ARN polymérase du phage T7. La souche est additionnée d'un plasmide pLysE qui exprime le Lysozyme T7, permettant de diminuer l'expression basale dans le cas où la protéine d'intérêt est toxique pour la cellule. Ce plasmide pLysE est conservé si nécessaire par l'addition de chloramphénicol. Nous n'utiliserons pas cette caractéristique de la souche durant l'étude et nous nous limiterons à ajouter la kanamycine pour conserver le vecteur d'expression pSKB3 (voir Figure 71, page 198).

XI.2 Composition des milieux

Lysogeny Broth (LB) : Pour 1 L : Peptone 10 g; Extrait de levures 5 g, Chlorure de sodium 10 g; qsp 1 L eau distillée. Autoclavage du milieu à 121°C pendant 20 min puis conservation avec bouchon étanche à température ambiante ou à 37 °C pour une utilisation rapide. Les milieux solides LB contiennent de l'Agar à raison de 10 g par litres et sont conservés en surfusion à 50°C après autoclavage pour éviter la solidification du milieu.

Milieu minium (M9) sans additif : Pour 1 L : sels M9 (5X) 200 mL; Glucose 1,6 g; Chlorure d'ammonium 800 mg; Sulfate de magnésium heptahydraté (1 M) 1,6 mL; Chlorure de calcium (1 M) 160 µL; q.s.p. 1 L eau distillée MilliQ.

Milieu minimum (M9+) avec additifs : Milieu M9 + additifs, protocole de préparation détaillé (3 pages) disponible en annexe page 221.

Tous les milieux servant à la croissance de la souche BL21 (DE3) en vue d'une surexpression de protéine doivent contenir un antibiotique pour sélectionner la souche de surexpression et éliminer les organismes indésirables. En l'occurrence la concentration de kanamycine effective dans tous les milieux utilisés est de 50 µg/mL. La solution stock est conservée à -20 °C concentrée 1000x, à 50 mg/mL.

Lors d'un marquage isotopique en milieu minimum, les espèces non marquées (glucose, chlorure d'ammonium) sont simplement remplacées par leurs équivalents marqués en suivant strictement le même protocole.

XI.3 Protocole initial de production de la protéine 6His-OrfM

A partir d'une colonie fraîche (< 24h) sur milieu LB Agar + kanamycine de la souche BL21 contenant le plasmide pSKB3-OrfM, on ensemence 5 mL de milieu LB liquide + kanamycine et on incube à 37°C sur la nuit. Le lendemain matin, la culture est centrifugée à 4200 g à 4 °C pendant 10 min. Après élimination du surnageant, on ajoute 1 mL de milieu minimum (M9) + kanamycine et on centrifuge à 4200 g pendant 5 minute à 4 °C. Cette étape est répétée encore une fois afin d'éliminer les derniers résidus de milieu riche. Le culot est solubilisé dans 500 µL de milieu M9 + kanamycine et ajouté dans 20 mL de milieu M9 + kanamycine puis incubé à 37 °C pendant 8h. 100 mL de milieu M9 + kanamycine dont la concentration de glucose a été doublée sont ensuite ensemencés en y versant l'intégralité des 20 mL de pré-culture, puis incubés à 37 °C sur la nuit. Le lendemain, 1 L de milieu M9 + kanamycine est ensemencé avec les 120 mL de pré-culture. Après environ 4 heures d'incubation à 37 °C, il faut mesurer la densité optique à 600 nm grâce à un spectromètre toutes les 30 minutes jusqu'à ce que la densité optique du milieu atteigne 0,6 unité, généralement au bout de 5 à 6 heures. L'IPTG est ajouté à raison d'une concentration de 1 mM final pour débiter l'induction, puis la culture est incubée à 4°C sur la nuit. Le lendemain, la culture est centrifugée à 4200 g pendant 15 minutes à 4°C. Le surnageant est éliminé et le culot est récolté, pesé puis congelé à -80 °C pour utilisation ultérieure.

XI.4 Protocole final de production de la protéine 6His-OrfM

A partir d'une colonie fraîche (< 24h) sur milieu LB Agar + kanamycine de la souche BL21 contenant le plasmide pSKB3-OrfM, on ensemence 5 mL de milieu LB liquide + kanamycine et on incube à 37°C sur la nuit. Le lendemain matin, on ensemence 20 mL de milieu LB liquide + kanamycine et on incube à 37°C pendant 5 heures. Ensuite, une dernière pré-culture de 1 L de milieu LB + kanamycine est lancée puis incubée à 37°C pendant au moins 1 heure ; la densité optique de la culture est mesurée toutes les 15 min à 600 nm. Dès que la densité optique atteint 0,8 unité,

la culture est centrifugée à 2500 g pendant 15 min à 20 °C. Le culot est soigneusement nettoyé par 3 cycles de lavages/centrifugation successives en éliminant le surnageant et en le solubilisant dans du milieu minimum avec additifs (M9+) + kanamycine. 250 mL de milieu minimum (M9+) + kanamycine sont ensuiteensemencés avec le culot obtenu et incubés une heure à 37°C. L'induction est démarrée en ajoutant l'IPTG à 1 mM dans la culture, puis la culture est incubée toute la nuit à 37°C. Le lendemain, la culture est centrifugée à 4200 g pendant 15 minutes à 4°C. Le surnageant est éliminé et le culot est récolté, pesé, puis congelé à -80 °C pour utilisation ultérieure.

XI.5 Purification de la protéine OrfM

Le culot d'induction est décongelé en le solubilisant dans un tampon de lyse contenant 20 mM Tris-HCl pH 7,5 ; 200 mM NaCl, 10 mM Imidazole, 0,04 mg/mL Lysozyme, 0,1% (v/v) Triton X. Le volume de tampon de lyse à ajouter est proportionnel au poids du culot noté juste avant la congélation, à raison de 15 mL de tampon de lyse par gramme de culot. On laisse le culot se solubiliser dans une chambre froide pendant au moins 30 minutes en appliquant une agitation légère grâce à un agitateur magnétique dans le bêcher. On applique ensuite une sonication de 3 fois 1 minute à 50 % de l'amplitude du sonicateur (Omega) par intervalles de 20 secondes toutes les minutes. La sonication se fait sur glace et les cycles sont courts pour limiter au maximum l'augmentation de la température au dessus de 10°C. Une sonde de température permet d'appliquer une pause si la température à l'intérieur du mélange dépasse 10°C. Le tampon de lyse ne doit pas mousser, et l'amplitude peut être réduite à 30% si le premier cycle produit déjà de la mousse. Les densités optiques mesurées à 600 nm avant et après sonication doivent montrer au minimum une diminution d'un facteur 3 pour assurer une lyse cellulaire correcte, la mesure est donc réalisée entre chaque cycle. Le mélange est ensuite centrifugé au moins 30 minutes à 53700 g, à 4 °C, pour précipiter les débris cellulaires. On récupère le surnageant qui contient alors la fraction totale des protéines cellulaires comprenant la protéine surexprimée. Un gel SDS-PAGE est réalisé avec l'ensemble des fractions collectées tout au long du processus de lyse et permet de vérifier que la majorité de la protéine a été récupérée et est soluble en l'état.

La résine Ni-Sepharose *HisTrap HP* (GE Healthcare) est utilisée pour réaliser une chromatographie d'affinité sur ions métalliques. Le nickel (Ni^{2+}) présent sur la résine possède une forte affinité

avec les résidus histidines, surtout avec l'étiquette poly-histidine introduite à l'extrémité de la séquence de la protéine surexprimée. Un volume de résine homogène de 1 mL est ajouté à une colonne vide et est couplée à une chaîne de chromatographie FPLC (ÄKTA-Prime Plus ; GE Healthcare). La colonne est équilibrée avec 120 mL de tampon HEPES à 20 mM, pH 7 contenant 100 mM de NaCl et 10 mM d'imidazole : il est appelé tampon A. Le débit du liquide doit être inférieur ou égal à 1 mL par minute pour un flux optimal. Après équilibrage, on injecte l'intégralité du surnageant contenant les protéines sur la colonne, et on termine par 15 mL de tampon A pour s'assurer qu'il ne reste plus de protéine non fixée dans l'appareillage. Ensuite, on utilise un gradient de tampon similaire au tampon A mais contenant cette fois entre 50 et 300 mM d'imidazole pour éluer les différentes protéines. Rappelons que l'imidazole va entrer en compétition avec les protéines fixées sur la colonne comportant le nickel. Sur un volume de 60 à 120 mL, on récupère des fractions de 2 mL durant l'intégralité de l'élution. Grâce à un spectromètre UV intégré à la sortie de l'appareillage, les mesures à 280 nm de longueur d'onde sont constamment enregistrées. Les fractions correspondant à la protéine surexprimée sont ensuite choisies et un gel SDS-PAGE est réalisé dans le but de vérifier la nature et le degré de pureté de l'échantillon. Une fois que les fractions contenant la protéine ont été identifiées, on les rassemble et on réalise une dialyse en boudin contre 2 litres de tampon A sur la nuit à 4°C. Cependant, avant d'intégrer les fractions dans le boudin, une mesure à 280 nm permet de calculer empiriquement la quantité totale de protéine, et on y ajoute 1 mg de protéase TEV ("*Tobacco Etch Virus*") pour 8 mg de protéine. Cette enzyme va spécifiquement cliver le site situé entre OrfM et son étiquette poly-histidine. En temps normal, la protéase TEV est sensible à la présence d'imidazole mais comme la lyse est réalisée sur la nuit jusqu'au lendemain, la quasi-totalité de la protéine recombinante sera effectivement libérée de son étiquette.

Une fois la dialyse terminée, l'échantillon de protéine purifiée et clivée de son étiquette poly-histidine va être soumise à la seconde étape de la chromatographie d'affinité aux ions métalliques. En effet, il faut séparer OrfM de l'étiquettes poly-histidine et de la protéase TEV encore présentes dans l'échantillon. La protéase TEV possède également une étiquette poly-histidine. Pour cette seconde étape, on utilise une colonne de paillasse de diamètre 0,5 cm contenant 1,5 cm de Ni-Sepharose *6-Fast-Flow* (GE Healthcare). La colonne est nettoyée avec 2 volumes d'eau puis équilibrée avec 1 volume de tampon A + 10 mM d'imidazole. En prenant soin de récupérer les fractions qui sortent de la colonne par tubes de 5 mL, on fait passer l'échantillon de protéine au travers de la résine. Comme pour la chromatographie automatique, il faut s'assurer que le flux ne soit pas trop rapide,

à raison d'une goutte par seconde (environ 1 mL/min). Comme OrfM ne possède plus d'étiquette poly-histidine, elle ne sera pas fixée par la résine, et les premières fractions appelées "*flow-through*" contiennent théoriquement OrfM. Une fois la totalité du volume de la fraction de protéine passée, on continue avec 1 volume de tampon A + 10 mM d'imidazole, puis on termine avec 1 volume de tampon A + 300 mM d'imidazole pour décrocher les étiquettes poly-histidines de la résine. Un gel SDS-PAGE est réalisé sur l'ensemble des fractions pour identifier celles qui contiennent OrfM purifiée.

Dans le cas d'OrfM, une troisième étape de chromatographie d'exclusion n'a pas été nécessaire car la protéine est suffisamment pure pour les études RMN. On réalise une concentration des fractions contenant OrfM grâce à des cellules Vivaspin Turbo (Sartorius) avec un seuil de coupure à 5 kDa et une limite de volume de 250 μ L. La concentration est réalisée par cycles de 10 min à 2300 g à 4°C en vérifiant systématiquement que le volume ayant traversé la colonne de concentration ne contient aucune protéine en mesurant l'absorbance à 280 nm. OrfM concentrée est ensuite conservée à 4°C.

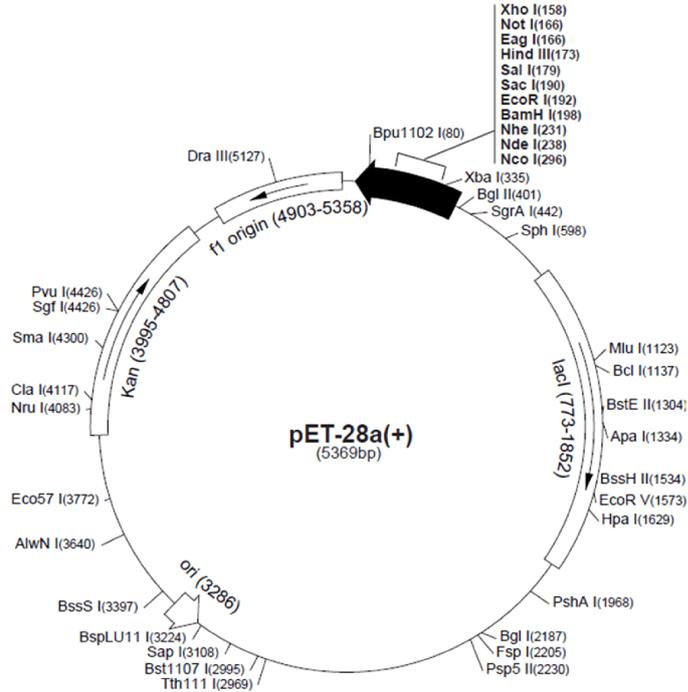
FIGURE 71 – Plasmide pSKB3 dérivé du plasmide pET-28a(+)

A)

pET-28a(+) sequence landmarks

T7 promoter	370-386
T7 transcription start	369
His•Tag coding sequence	270-287
T7•Tag coding sequence	207-239
Multiple cloning sites (<i>Bam</i> H I - <i>Xho</i> I)	158-203
His•Tag coding sequence	140-157
T7 terminator	26-72
<i>lacI</i> coding sequence	773-1852
pBR322 origin	3286
Kan coding sequence	3995-4807
f1 origin	4903-5358

The maps for pET-28b(+) and pET-28c(+) are the same as pET-28a(+) (shown) with the following exceptions: pET-28b(+) is a 5368bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-28c(+) is a 5367bp plasmid; subtract 2bp from each site beyond *Bam*H I at 198.



B)



a. Carte du plasmide pET-28a (NovaGene).

b. Le plasmide modifié contient un site de digestion à la protéase TEV ("Tobacco Etch Virus") ainsi qu'un codon stop à l'extrémité 3' du gène recombinant intégré par l'amorce reverse pour le site *HindIII*.

XII Résonance magnétique nucléaire

XII.1 Préparation de l'échantillon

En suivant le protocole de production en milieu M9+, OrfM a été uniformément marquée en utilisant le ^{13}C -glucose comme seule source de carbone et le chlorure d' ^{15}N -ammonium comme seule source d'azote durant la surexpression. La protéine a été purifiée en tampon HEPES 20 mM pH 7 avec 100 mM de chlorure de sodium afin de garantir sa solubilité ainsi que sa stabilité dans le temps. L'échantillon protéique a ensuite été concentré jusqu'à ce que le volume d'échantillon soit égal à 250 μL puis supplémenté par 10% de D_2O . La concentration molaire finale calculée est de 0,545 mM pour la protéine OrfM.

XII.2 Collecte des spectres RMN et attribution

Tous les spectres ont été collectés à 25°C sur un spectromètre Bruker 600 MHz (Plateforme SCPIM, Brabois) équipé d'une cryosonde. Les données RMN collectées ont ensuite été traitées en utilisant TopSpin 3.0 (Bruker) puis analysées via CcpNmr (Vranken *et al.*, 2005). Les attributions actuelles ^1H , ^{15}N et ^{13}C ont été obtenues en utilisant les spectres suivants : ^1H - ^{15}N -HSQC, ^1H - ^{15}N -TOCSY-HSQC (temps de mélange 60 ms), ^1H - ^{15}N -NOESY-HSQC (temps de mélange 120 ms), HNHA, HBHA(CO)NH, HNCA, HNCACB, CBCA(CO)NH, HNCOC et HNCACO. Les expériences de relaxation ^{15}N R_1 , R_2 et NOE hétéronucléaire ont été collectées en utilisant 10 délais de relaxation pour les mesures de R_1 (2, 10, 50, 100(x2), 200, 400, 600, 1000 et 1400 ms) et de R_2 (0, 17(x2), 234, 51, 68, 85, 102, 119, 136 ms) et ont été analysées en utilisant NMRViewJ et Tensor 2.0 (Dosset *et al.*, 2000). Les expériences d'interaction ont été menées sur un échantillon d'OrfM marqué ^{15}N à une concentration de 0,12 mM dans un tampon phosphate 50mM pH 7 avec 200 mM NaCl auquel ont été ajoutés entre 0,1 et 2,2 équivalents d'OrfL non marqué à 0,6 mM. Les expériences ont été renouvelées sur un échantillon d'OrfM marqué ^{15}N à une concentration de 0,31 mM dans un tampon phosphate 50 mM pH 6,5 avec 200 mM NaCl ainsi qu'à un échantillon à une concentration de 0,38 mM dans un tampon phosphate 50mM pH 6,5 avec 500 mM NaCl.

XII.3 Calcul de structure préliminaire

En utilisant 902 contraintes de distance (à partir de pics NOE attribués manuellement), 152 contraintes d'angles (générées par le sous-programme DANGLE dans Ccpnmr) et 24 contraintes de liaisons hydrogène, ont été calculés 120 modèles grâce à ARIAWeb (Allain *et al.*, 2020). Les 14 modèles de plus basse énergie ont été sélectionnés et superposés grâce à mTM-align et visualisés grâce à PyMOL en utilisant MPSA_Viewer.

XIII Résolution de structure cristallographiques par rayons X

XIII.1 Cristallisation des protéines OrfG₆₄₋₂₀₄ et OrfG₆₄₋₃₃₁

Les conditions de cristallisation de OrfG₆₄₋₂₀₄ sont décrites dans la partie "*Materials & Methods*" de l'article correspondant (Cappele *et al.*, 2021).

La protéine OrfG₆₄₋₃₃₁ a été cristallisée en utilisant la méthode de diffusion de vapeur en goutte assise à 20 °C (293 K). Les cristaux ayant diffracté ont été obtenus dans plusieurs conditions. Le meilleur jeu de données a été collecté à partir d'un cristal obtenu dans une goutte de 0,6 µL contenant 0,3 µL de solution de puits (1,26 M de sulfate d'ammonium et 100 mM de cacodylate de sodium/HCL à un pH de 6,5) et 0,3 µL de la protéine concentrée à 31 mg/mL dans un tampon contenant 100 mM Tris/HCl pH 8. De la même manière que pour OrfG₆₄₋₂₀₄, toutes les gouttes ont été préparées par le robot de cristallisation Orxy8 (Douglas Instruments, Hungerford, Grande-Bretagne) dans des plaques MRC 96 réservoirs avec 2 puits chacun pour les gouttes de cristallisation (SWISSCI AG, Neuheim, Suisse). La plupart des cristaux n'ayant pas diffracté ont été obtenus dans les deux jours suite à la préparation des gouttes tandis que les cristaux ayant permis d'obtenir les jeux de données utilisés pour résoudre la structure ont été observés 3 mois après. Les cristaux ont été récupérés grâce à des boucles en nylon puis ont été rapidement trempés dans une goutte de solution de cristallisation contenant en plus 20 % (v/v) de glycérol, puis instantanément congelés dans de l'azote liquide pour leur conservation et leur transport pour une collecte au synchrotron ultérieure.

XIII.2 Collecte des données de diffraction et résolution de la structure d'OrfG₆₄₋₃₃₁

Les expériences de diffraction de rayons X ont toutes été réalisées au Synchrotron SOLEIL (St-Aubin, France) sur la ligne Proxima-2A (voir Figure 72, page 202). Les mesures du cristal sont réalisées à une température de -173 °C (100 K) maintenue grâce à un flux d'azote gazeux. La longueur d'onde de mesure est de 0,9801 Å et le cristal mesuré diffracte jusqu'à 1,84 Å. 3600 images ont été collectées en réalisant un 360° autour du cristal. Les données ont été traitées en utilisant XDS (Kabsch, 2010) et plus particulièrement XDSME (XDS Made Easier, Pierre Legrand 2016). Le phasage a été réussi en utilisant la suite CCP4i2 (Potterton *et al.*, 2018). Le programme de remplacement moléculaire MOLREP (Vagin & Teplyakov, 1997) a été utilisé en suivant la procédure suivante : (1) le modèle d'OrfG₆₄₋₂₀₄ comme premier modèle en ayant pris soin d'éliminer les boucles flexibles entre les structures secondaires tout en conservant les chaînes latérales des acides aminés ; (2) le modèle du domaine c-terminal de TcpC (3UB1) en ayant retiré les boucles flexibles et en ayant remplacé les résidus par des alanines. Une première étape de construction de modèle grâce au programme Buccaneer (Cowtan, 2006) permet de fournir une base solide qui est ensuite utilisé par le logiciel ARPwARP (Langer *et al.*, 2008), fournissant un modèle quasiment complet. Le modèle est ensuite amélioré avec quelques cycles de modification manuelle en utilisant Coot (Emsley *et al.*, 2010) et Refmac5 (Murshudov *et al.*, 2011).

FIGURE 72 – Statistiques de collecte du cristal d'OrfG₆₄₋₃₃₁ et d'affinement du modèle

	Native
Data collection	
Space group	P 4 ₁ 3 2
Beamline	Proxima-2A. SOLEIL
Unit cell parameters (Å. °)	122.922 122.922 122.922 90° 90° 90°
Wavelength (Å)	0.980107
Resolution (Å) ^a	50.0-1.84 (1.95-1.84)
Observations ^a	2 298 453 (360 000)
Unique reflections	29 698 (4 676)
R _{meas} (/) ^a	0.093 (1.261)
CC _{1/2}	1.0 (0.948)
Mean I/σ ^a	48.17 (4.62)
Completeness (%) ^a	99.8 (99.4)
Multiplicity ^a	77.4 (77.0)
Figure of merit (/) ^b	-
First model ^c	-
Model refinement *	
Resolution (Å) ^a	44.17-1.835 (1.835-1.883)
No. of protein residues	235
No. of non-hydrogen protein atoms	1935
No. of waters	175
R _{work} ^a	0.199 (0.337)
R _{free} ^{a,d}	0.230 (0.358)
RMSD bonds (Å)	0.012
RMSD angles (°)	1.601
Ramachadran plot ^e	
Favored (%)	95.96
Outliers (%)	0.0
Rotamer outliers (%)	0.93
Clashscore ^e	8.58
MolProbity score ^e	1.76
Average B (Å ²)	28.53
Protein	30.4
Water	37.0
Coord. error (Å) ^f	0.118
PDB code	-

* Data based on internal refinement statistics from CCP4

^a Values for the outer shell of data are given in parentheses.

^b Before/after solvent flattening; the same values were obtained for both hands.

^c As built by Buccaneer (CCP4i); no model was obtained for the other hand.

^d Five percent test set.

^e From Molprobity.

^f Value based on R_{free}.

ANNEXES

Annexe S1 – Table comparative des protéines du système Dot/Icm

Table 2
Homologies to the *L. pneumophila* Icm/Dot proteins

Protein	<i>C. burnetii</i> Ident./Sim. ^a	R64 (Incl)	Type-IVA Vir/Incp ^g	Location ^b	Conserved motifs
IcmT	47.1/63.2	TraK		IM	
IcmS	52.6/67.5	–		Cyt	
IcmR	–	–		Cyt	
IcmQ ^b	22.6/34.0	–		Cyt	
IcmP/DotM	36.5/50.3	TrbA		IM	
IcmO/DotL	59.6/70.9	TrbC	VirD4/TraG	IM	Walker-A
IcmN/DotK	24.8/32.7	–		Perp	Lipobox
IcmM/DotJ ^c	–	–		IM	
IcmL/DotI ^c	35.8/48.5	TraM		IM	
	37.6/50.9				
IcmK/DotH	48.2/58.6	TraN		Perp	
IcmE/DotG ^d	44.2/51.4	TraO	VirB10/TrbI	IM	
IcmG/DotF	27.1/42.4	TraP		IM	Coiled-coil
IcmC/DotE ^e	27.7/47.1	TraQ		IM	
IcmD/DotP ^e	20.8/34.0	TraR		IM	
IcmJ/DotN	51.2/61.9	TraT		Cyt	
IcmB/DotO	62.2/73.7	TraU	VirB4/TrbE	Cyt	Walker-A
IcmF ^f	23.1/34.4	–		IM	Walker-A
IcmH	27.7/37.6	–		IM	
IcmW	58.6/68.4	–		Cyt	
IcmX	21.9/30.2	TraW		Perp	
IcmV	27.1/44.5	TraX		IM	
DotA	25.0/34.4	TraY		IM	
DotB	63.4/74.8	TraJ	VirB11/TrbB	Cyt	Walker-A
DotC	40.1/51.1	TraI		Perp	Lipobox
DotD	39.2/48.5	TraH		Perp	Lipobox

–, No homologous protein was found.

^a Identity/similarity, similar amino acids are I, L, V, and M; H, K and R; D and E; T and S; and Q and N.

^b The *C. burnetii* IcmQ is about 50 amino acids larger than the *L. pneumophila* IcmQ.

^c The *L. pneumophila* IcmM/DotJ and IcmL/DotI proteins are homologous to one another, see text.

^d Homology is calculated for the C-terminal end, see text.

^e The *L. pneumophila* IcmC/DotE and IcmD/DotP proteins are homologous to one another, see text.

^f The *C. burnetii* IcmF homologous protein contain a stop codon after amino acid number 373, the second “part” of IcmF (additional 628 amino acids) continues in the same frame.

^g Homology was found using the PSI-BLAST program.

^h IM, inner membrane; Cyt, cytoplasm; Perp, periplasm.

tiré de Segal *et al.*, 2005

Annexe S2 – Séquences protéiques 1/2

OrfG 1-331

```
10      20      30      40      50      60
MIKAKQYLLK AREFLSNFKK VDKKGGPPKL KVTTKKTVNI AVIGGIAFLL FVGLMGSIRA

70      80      90      100     110     120
ITLSNKVGTL QAQVEATQKQ KAQPIDANRK YDYKLQYYLN DYVYAYFTLP QEGDKQQAQV

130     140     150     160     170     180
EHLNSFYNFV PDVKAQGQVR NPSTLLDSQL VTVEGKVATY KVKYKEMIQH DKDTEEKELV

190     200     210     220     230     240
TGFNIPFDEK EGKYYVSGLP WFSAISSQA GYFSEDDQLQ LTANDHVSQS QHKKVEKFLK

250     260     270     280     290     300
VFFNTYTTNQ DNLNLIAKNV AIVANTTFKT IDYTYLKKDR ADLIAYVQAT FEVGGTTHSE

310     320     330
NFTFTLSEKD KSYVVTKLEH TIPLNYANDK D
```

Theoretical pI/Mw: 8.73 / 37735.82

OrfG 64-331

```
10      20      30      40      50      60
SMSNKVGTLQ AQVEATQKQK AQPIDANRKY DYKLQYYLND YVYAYFTLPQ EGDKQQAQVE

70      80      90      100     110     120
HLNSFYNFVP DVKAQGQVRN PSTLLDSQLV TVEGKVATYK VKYKEMIQHD KDTEEKELVT

130     140     150     160     170     180
GFNIPFDEKE GKYYVSGLPW FSAIESSQAG YFSEDDQLQL TANDHVSQSQ HKKVEKFLKV

190     200     210     220     230     240
FFNTYTTNQD NLNLIAKNVA IVANTTFKTI DYTYLKKDRA DLIAYVQATF EVGGTTHSEN

250     260     270
FTFTLSEKDK SYVVTKLEHT IPLNYANDKD
```

Theoretical pI/Mw: 5.44 / 31066.58

OrfG 64-204

```
10      20      30      40      50      60
SMSNKVGTLQ AQVEATQKQK AQPIDANRKY DYKLQYYLND YVYAYFTLPQ EGDKQQAQVE

70      80      90      100     110     120
HLNSFYNFVP DVKAQGQVRN PSTLLDSQLV TVEGKVATYK VKYKEMIQHD KDTEEKELVT

130     140
GFNIPFDEKE GKYYVSGLPW FSA
```

Theoretical pI/Mw: 5.70 / 16536.52

Annexe S3 – Séquences protéiques 2/2

OrfG 223-331

10 20 30 40 50 60
MAKIGLNYGE KLQIADKTYR VITDGLNVPĒ VLGKLSFRGĪ EGPDIIFEED RTQRNPDGSF
70 80 90 100 110 120
VQVNTGEIRG IVVAIHSSVQ HETLFFTIAD MAQSEIEDLG IKYREEVELE DIVITHTHVN
130 140 150
RNDIYKLFAS KIKKKNGGTP LKENKPLDNK KDN

Theoretical pI/Mw: 5.60 / 17282.62

OrfM

10 20 30 40 50 60
MIPETLVMKP DFIGPAYYEK LGNAPRFIFG KEGKYQRHIL NSEKHGAFHV ITPASTTVFP
70 80 90
EDALVEVVNP IFLSDTALNG NSVAPALNVF AEKLVLKK

Theoretical pI/Mw: 6.83 / 10813.56

OrfL

10 20 30 40 50 60
NDHVSDSQHK KVEKFLKVFF TNYTTNQDNĪ NLIAKNVAIV ANTTFKTIDY TYLKKDRADĪ
70 80 90 100
IAYVQATFEV GGTTHSENFĪ FTLSEKDĪSY YVTKLEHTIP LNYANDKD

Theoretical pI/Mw: 6.46 / 12463.90

Cristal 1

Remote $\kappa = 45^\circ$ Limite Rés.	Nombre Reflections		Completeness of Data (%)	R-FACTOR (%)		Compared	I/ σ	R-meas	CC(1/2)	Anomal		
	Observed	Unique		Possible	Observed					Expected	A. Corr.	A. Signal
5.97	8145	317	100	2.8	3.1	8145	101.55	2.8	100	38*	1,304	186
4.22	11453	440	100	2.7	3.1	11453	109.39	2.8	100	-5	0.757	314
3.45	15059	548	100	3.8	3.6	15059	89.73	3.9	100	-9	0.775	419
2.98	15694	612	100	5.8	5.3	15694	59.27	5.9	100	-6	0.793	490
2.67	19586	687	100	10.2	9.7	19586	37.72	10.4	99.9	-5	0.784	555
2.44	22134	752	100	15.9	15.9	22134	25.18	16.2	99.8	-3	0.775	625
2.18	24960	839	100	33.4	33.7	24960	12.77	34	99.7	-8	0.764	707
1.94	27057	930	100	109.5	118.4	27057	3.74	111.4	95.4	-7	0.7	801
1.89	24833	945	98	147.0	162.2	24815	2.49	149.8	86	-9	0.656	785
Total	372 824	13 360	99.8	5.9	6.0	372 806	36.09	6	100	-7	0.754	10 747

$\kappa = 0^\circ$ Limite Rés.	Nombre Reflections		Completeness of Data (%)	R-FACTOR (%)		Compared	I/ σ	R-meas	CC(1/2)	Anomal		
	Observed	Unique		Possible	Observed					Expected	A. Corr.	A. Signal
6.47	9277	591	99.8	2.2	2.2	9277	104.97	2.2	100	95*	4,156	196
4.6	16587	1066	100	2.4	2.6	16587	93.24	2.4	100	60*	1,555	429
3.76	21748	1371	100	2.5	2.7	21748	90.64	2.6	100	38*	1,011	580
3.26	26935	1634	100	3.3	3.6	26935	70.61	3.4	100	19*	0.862	712
2.92	28540	1841	100	5	5.4	28540	45.99	5.1	99.9	10	0.832	813
2.66	31395	2019	100	8.4	8.7	31395	29.52	8.7	99.9	1	0.798	906
2.47	33965	2223	100	14.4	13.7	33965	18.71	14.9	99.7	-8	0.766	1004
2.31	35621	2360	100	23.2	21.2	35621	12.03	24	99.3	2	0.86	1077
2.18	33904	2510	98.9	38.2	34.6	33882	6.73	39.8	98.1	48*	1.174	1134
total	237972	15615	99.8	4.3	4.4	237950	40.49	4.5	100	32*	1,038	6851

$\kappa = 27^\circ$ Limite Rés.	Nombre Reflections		Completeness of Data (%)	R-FACTOR (%)		Compared	I/ σ	R-meas	CC(1/2)	Anomal		
	Observed	Unique		Possible	Observed					Expected	A. Corr.	A. Signal
5.94	12371	766	99.7	2.2	2.3	12371	103.57	2.3	100	92*	3,811	264
4.22	21517	1379	100	2.3	2.5	21517	98.64	2.4	100	52*	1,345	566
3.45	28089	1789	100	2.8	2.9	28089	81.49	2.9	100	26*	0.999	767
2.99	34541	2102	100	4.1	4.3	34541	55.99	4.2	100	8	0.802	930
2.68	37104	2392	100	7.3	7.5	37104	31.6	7.5	99.9	-2	0.751	1068
2.44	41466	2667	100	13.5	13.1	41466	18.5	14	99.7	-12	0.751	1204
2.26	42095	2876	100	25.5	23.1	42095	10.43	26.4	99	6	0.852	1314
2.12	45344	3076	100	48.6	44.5	45344	5.35	50.5	97.4	42*	1.01	1413
2	46461	3239	98.7	87	87.1	46413	2.43	90.3	89.3	24*	0.818	1468
total	308988	20286	99.8	4.8	4.8	308940	32.44	4.9	100	26*	0,971	8994

Cristal 1

κ = 45°	Nombre Reflections			Completeness of Data (%)	R-FACTOR (%)		Compared	I/σ	R-meas	CC(1/2)	Anomal	
	Observed	Unique	Possible		Observed	Expected					A. Corr	A. Signal
Limite Rés.	12144	766	767	99.9	2.3	2.4	21144	97.49	2.4	100	91*	3.569
5.95	21332	1378	1378	100	2.5	2.6	21332	94.23	2.5	100	48*	1.275
4.22	27457	1773	1773	100	3	3	27457	79.32	3.1	100	28*	0.947
3.45	34272	2105	2105	100	4.3	4.4	34272	56.13	4.4	100	10	0.879
2.99	36422	2386	2386	100	7.7	7.6	36422	32.27	8	99.9	1	0.8
2.68	40281	2643	2643	100	13.6	13.2	40281	19.3	14.1	99.7	6	0.821
2.44	41697	2866	2866	100	23.6	22.8	41697	11.23	24.5	99.2	38*	0.968
2.26	45589	3064	3064	100	47.1	45.8	45589	5.5	48.8	97	57*	1.145
2.12	47551	3258	3276	99.5	86.4	88.8	47541	2.49	89.6	86.4	20*	0.792
2	306745	20239	20259	99.9	4.9	5	306735	32.04	5.1	100	39*	1.012
total												8995

κ = 60°	Nombre Reflections			Completeness of Data (%)	R-FACTOR (%)		Compared	I/σ	R-meas	CC(1/2)	Anomal	
	Observed	Unique	Possible		Observed	Expected					A. Corr	A. Signal
Limite Rés.	6366	405	405	100	2.6	2.9	6366	79.13	2.7	100	92*	3.516
7.4	11132	715	715	100	3.8	4.1	11132	57.79	4	100	52*	1.418
5.26	14814	920	920	100	3.7	3.8	14814	66.18	3.8	100	28*	0.954
4.31	17865	1082	1082	100	4.7	4.6	17865	53.92	4.9	99.9	16	0.875
3.73	18776	1215	1215	100	6.8	6.6	18776	37.62	7	99.9	1	0.818
3.34	21108	1369	1369	100	11.2	11	21108	23.83	11.5	99.8	-5	0.739
3.05	22722	1464	1464	100	19.4	19	22722	14.57	20.1	99.4	-2	0.824
2.83	24774	1581	1581	100	32.8	31.3	24774	8.72	33.9	98.2	18*	0.9
2.64	25189	1710	1712	99.9	48.8	48.9	25170	5.19	50.5	93.8	8	0.818
2.49	162746	10461	10463	100	6.6	6.6	162727	30.1	6.8	100	17*	0.952
total												4499

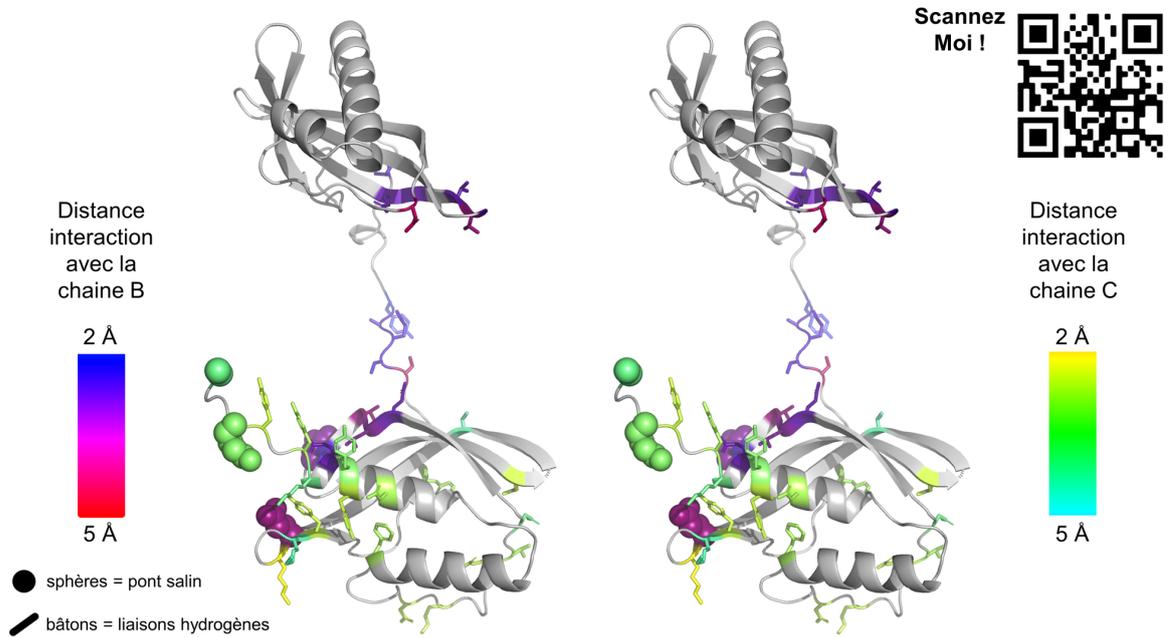
κ = 69°	Nombre Reflections			Completeness of Data (%)	R-FACTOR (%)		Compared	I/σ	R-meas	CC(1/2)	Anomal	
	Observed	Unique	Possible		Observed	Expected					A. Corr	A. Signal
Limite Rés.	6185	401	403	99.5	1.9	2.2	6185	109.1	2	100	96*	4.661
7.4	11037	717	717	100	2.5	2.7	11037	87.23	2.6	100	75*	2.002
5.27	14538	920	920	100	2.3	2.5	14538	99.34	2.4	100	42*	1.238
4.31	17588	1080	1080	100	2.8	2.8	17588	88.49	2.9	100	30*	1.041
3.74	18826	1215	1215	100	3.4	3.4	18826	70.32	3.5	100	16	0.9
3.34	21073	1364	1364	100	4.6	4.5	21073	51.73	4.8	100	8	0.821
3.05	22729	1467	1467	100	6.8	6.5	22729	36.48	7.1	99.9	-3	0.806
2.83	24870	1580	1580	100	10.3	9.9	24870	24.75	10.7	99.8	12	0.84
2.65	24232	1667	1712	97.4	15.3	15	24213	16.13	15.9	99.4	39*	1.033
2.49	161078	10411	10458	99.6	3.5	3.5	161059	54.63	3.6	100	35*	1.106
total												4477

Cristal 2

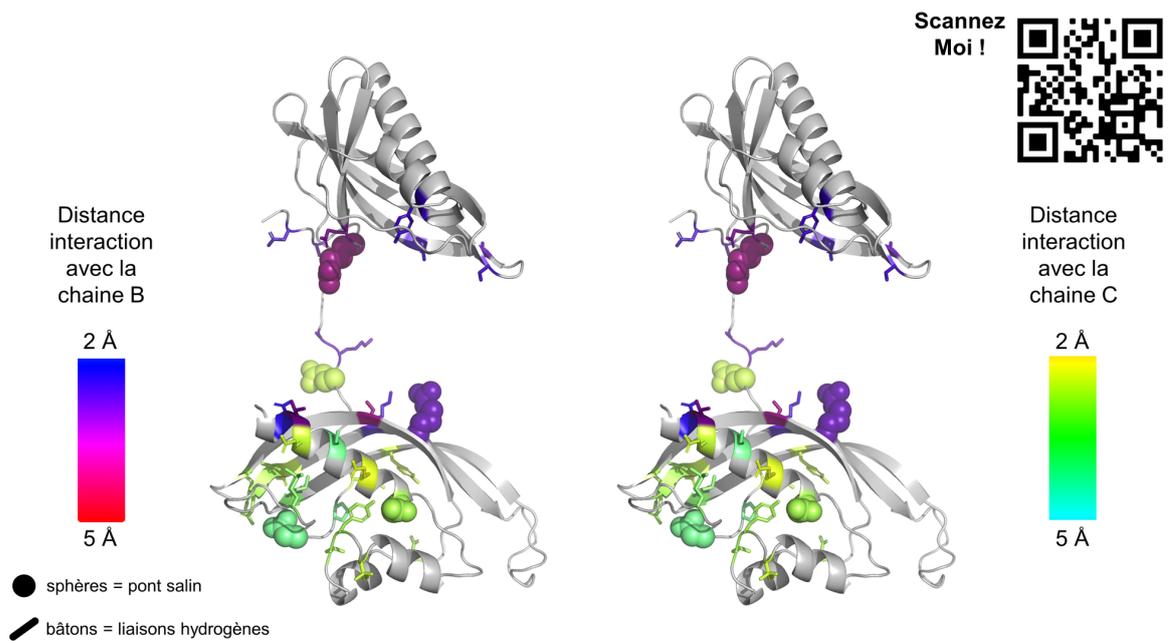
$\kappa = 0^\circ$	Nombre Reflections			Completeness of Data (%)	R-FACTOR (%)		Compared	I/ σ	R-meas	CC(1/2)	Anomal	
	Observed	Unique	Possible		Observed	Expected					A. Corr	A. Signal
Limite Rés.	9706	589	589	100	3.2	3.8	9706	67,21	3,3	100	90*	3,001
4.62	16507	1053	1053	100	4.3	4.2	16507	57,78	4,4	100	33*	1,188
3.78	20318	1344	1344	100	5	4.7	20318	51,5	5,2	99,9	0	0,883
3.27	25979	1618	1618	100	7.7	7.2	25979	36,33	7,9	99,9	2	0,927
2.93	29508	1811	1811	100	14.9	14.8	29508	20,1	15,4	99,7	-1	0,871
2.67	30118	1996	1996	100	28.8	29.1	30118	10,51	29,8	99	-4	0,796
2.48	34116	2189	2189	100	54.9	56.6	34116	5,57	56,8	96,4	10	0,818
2.32	34895	2350	2350	100	85.3	92.9	34895	3,3	88,4	91,7	2	0,746
2.18	37137	2454	2478	99	132.9	152.7	37120	1,98	137,5	84,2	3	0,679
total	238284	15404	15428	99.8	9.3	9.5	238267	20,16	9,6	100	9	0,89

$\kappa = 25^\circ$	Nombre Reflections			Completeness of Data (%)	R-FACTOR (%)		Compared	I/ σ	R-meas	CC(1/2)	Anomal	
	Observed	Unique	Possible		Observed	Expected					A. Corr	A. Signal
Limite Rés.	9107	555	555	100	3.4	3.9	9107	65,06	3,5	99,9	89*	2,862
4.7	15604	1000	1000	100	4.6	4.4	15604	54,79	4,8	100	27*	1,147
3.84	19122	1280	1280	100	5.1	4.7	19122	50,68	5,3	99,9	10	0,95
3.33	24307	1522	1522	100	7.6	6.9	24305	36,25	7,9	99,9	-1	0,842
2.98	27971	1717	1718	99.9	14	13.7	27971	20,89	14,5	99,7	-3	0,79
2.72	29251	1921	1922	99.9	27.4	28.4	29251	10,68	28,3	99,2	-6	0,77
2.52	32184	2043	2048	99.8	50.7	54.1	32183	5,89	52,4	96,4	-5	0,746
2.36	33526	2239	2242	99.9	76	85.1	33526	3,65	78,7	92,3	-5	0,696
2.22	34603	2304	2366	97.4	123.9	147	34581	2,06	128,2	72,4	-4	0,666
total	225675	14581	14653	99.5	9.5	9.7	225650	20,04	9,8	99,9	4	0,846

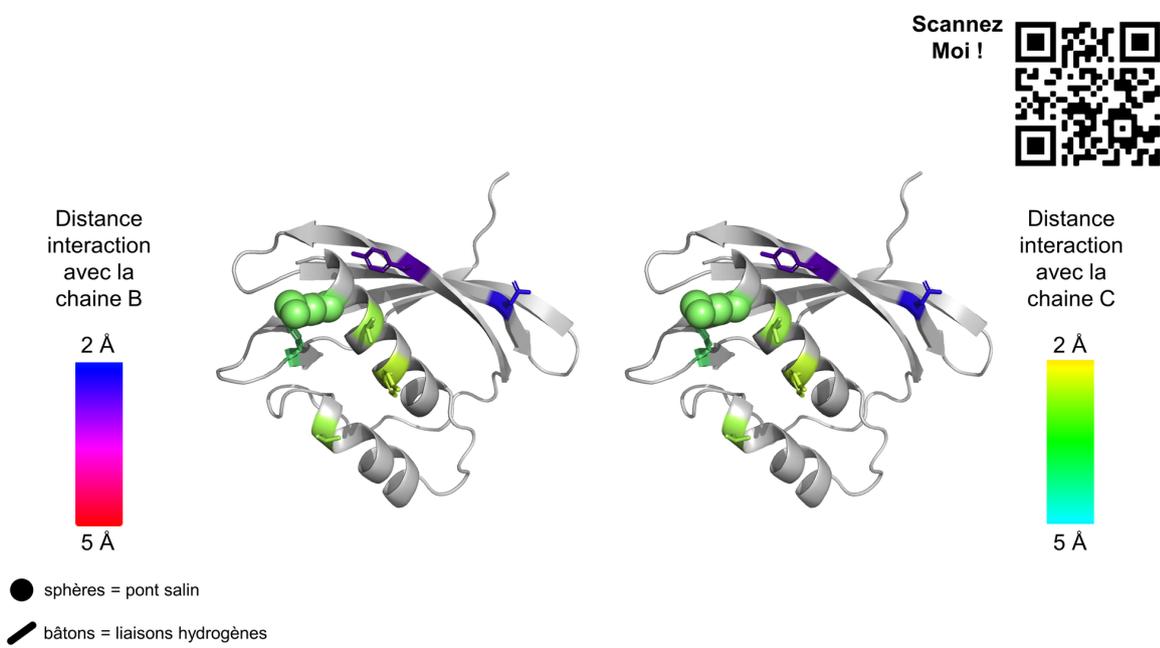
Annexe S7 – Interface hydrophile de OrfG₆₄₋₃₃₁ en vue stéréoscopique



Annexe S8 – Interface hydrophile de TcpC en vue stéréoscopique



Annexe S9 – Interface hydrophile de TraM en vue stéréoscopique



Documentation MPSA Viewer

Julien Cappèle, Claude Didierjean, Frédérique Favier

Documentation version 0.3 ; Mars 2021

Contents

1 Introduction

2 How install MPSA Viewer

2.1	Compatibility	
2.2	Python	
2.3	Installation and folder preparation	

3 Some things to know about MPSA Viewer

3.1	Terminal	
3.2	Commands	
3.2.1	help	
3.2.2	new_project <name> (<template>)	
3.2.3	zip <name>	
3.2.4	fetch <name> <url>	
3.2.5	analysis <name>	
3.3	Example of a project with MPSA Viewer	

1 Introduction

MPSA Viewer has been created to visually emphasize multi-protein structural alignment by using both PyMOL and mTM-align, two robust and powerful analysis tools. Firstly, the mTM-align server will handle your structures using TM-align and produce a protein data bank file containing all structurally aligned atoms of your proteins, and a sequence alignment based on the later. Analyzing those results with MPSA Viewer means that all software detectable similarities and differences within your alignment will be shown visually using PyMOL.

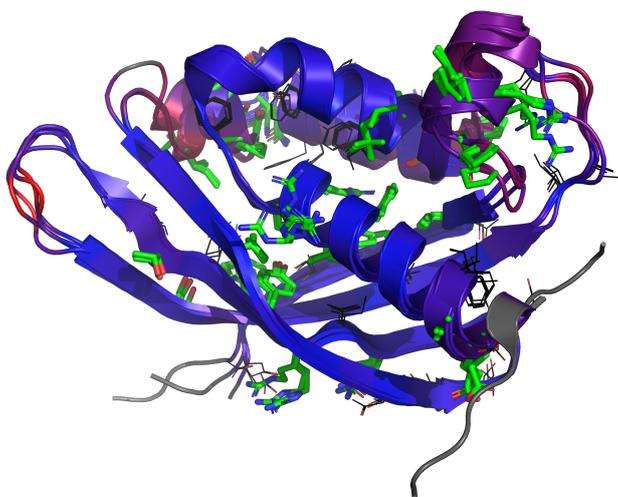


Figure 1: MPSA Viewer representation of a structural alignment in PyMOL

In the Figure 1, you can see an alignment performed by mTM-align with 4 structures (2cc3, 4kz1, 4lso, 4mei) shown in PyMOL. Green and black lateral chains represent respectively identical and same class amino acids if all structures are aligned at this exact position in space. If one of those structure would be too far, none of those amino acids would show up. Additionally, a backbone gradient coloration is applied based on the RMSD of alpha carbons of those amino acids. In this example, blue means close to 0 Å, red means close or above to 2 Å. A grey coloration is applied whenever the highest distance between two alpha carbons at a given position exceeds 4 Å, or if mTM-align placed a gap (-) in the alignment at this position for one or more structure. Those parameters are the default for MPSA Viewer, but those parameters can be changed pretty easily to your preferences.

In this documentation we will go through the installation process and the every-day use of the software. As always with third-party software, make sure you downloaded it for free from a trust worthy source. Authors of the software are usually a good start!

2 How install MPSA Viewer

2.1 Compatibility

As of now, MPSA Viewer *should* be natively compatible with any recent linux distribution, and is compatible with Windows 10 (20H2). MPSA Viewer has not been tested on OSX systems (yet).

2.2 Python

MPSA Viewer is a python script, meaning that you need to install Python. Python is *usually* natively installed on Unix (Linux & OSX) systems, however, you will need a recent version of Python. For Windows users, you will have to install it.

Python 3.9 is one of the latest release of Python as of now and is recommended to use this software. MPSA Viewer might be compatible with Python 3.4, but has not been tested with those versions.

Here is a complete set of URLs to the python documentation on all platforms:

Windows : <https://docs.python.org/3/using/windows.html>

Linux : <https://docs.python.org/3/using/unix.html>

Mac : <https://docs.python.org/3/using/mac.html>

2.3 Installation and folder preparation

MPSA Viewer is portable, meaning it is contained into a single script file that you can put anywhere you want on your system (except in your system files, obviously). On startup, MPSA Viewer will create a folder named MPSA_Projects that will be used to store all of your projects and data. If you move the script file, the project folder won't be used anymore and another one will be created. The best way to move the software is to also move the projects folder with it.

3 Some things to know about MPSA Viewer

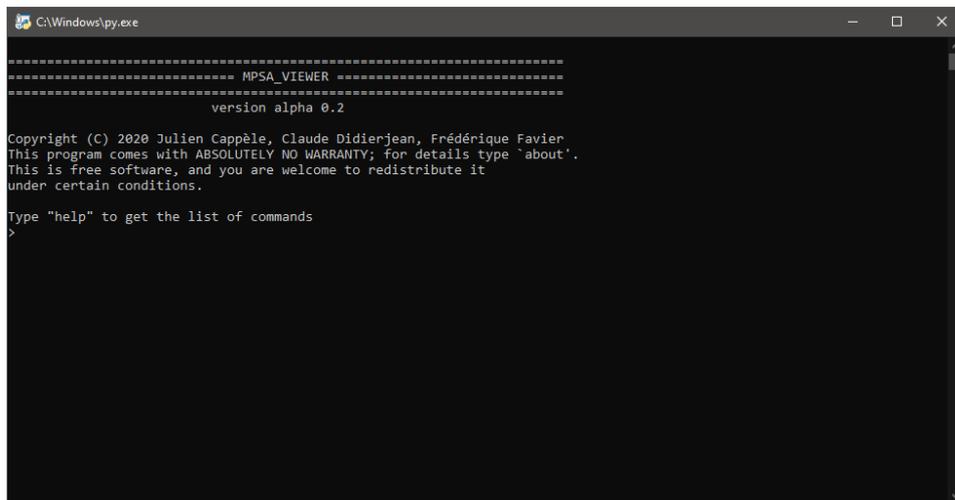
3.1 Terminal

MPSA Viewer will work as a command line terminal. Without any additional library for Python, this means that it will be very basic without auto-completion, previous command history or any advanced command line terminal extensions. But it will get the job done.

In order to start the terminal, you have to use your operating system terminal first, navigate to the executable folder and use the command :

```
$ python mpsa_executable.py
```

This will start the terminal and should look like this:



```
C:\Windows\py.exe
----- MPSA_VIEWER -----
version alpha 0.2
Copyright (C) 2020 Julien Cappèle, Claude Didierjean, Frédérique Favier
This program comes with ABSOLUTELY NO WARRANTY; for details type 'about'.
This is free software, and you are welcome to redistribute it
under certain conditions.
Type "help" to get the list of commands
>
```

Figure 2: Default look of MPSA Viewer in Windows terminal

Starting now, all commands that are typed within the software terminal will start with the superior sign (>), indicating a command in MPSA Viewer. You don't need to type it.

3.2 Commands

3.2.1 help

When the program starts, you will be prompted for a command. If you don't remember the syntax of a command, you can ask the program for help. By using help, the terminal will print a basic description of every implemented commands. Example of use:

```
> help
```

3.2.2 `new_project` <name> (<template>)

The command `new_project` creates a folder within the `MPSA_Project` folder alongside the script executable file. You have to specify the name of your project and the folder will be given the same name. Inside your new project folder, will be created everything needed for MPSA Viewer to work, especially some configuration files that are important.

Make sure to check the configuration file and specify your PyMOL path, and specify your desired structure names in the structure configuration file for every new project.

Example of use:

```
> new_project gfp_18_Oct_2020
```

If you want to create a project based on another one, there is an optional argument <template> that you can use to specify the project that you want to use as a template for the new one. Your preferences will be imported from the template to your new project.

Example of use:

```
> new_project gfp_21_Oct_2020 gfp_18_Oct_2020
```

3.2.3 `zip` <name>

The command `zip` will prepare your desired structure files that you have specified, will download them if possible and ask for you to place your own structures in the correct folder. Once all structures are present, they will be searched for seleno-methionines that are unfortunately not compatible with mTM-align for now. Finally, all corrected structures are compressed into a zip file, ready to be sent on the mTM-align server to be worked on.

Example of use:

```
> zip gfp_21_Oct_2020
```

3.2.4 `fetch` <name> <url>

The command `fetch` will, as its name suggest, fetch results from mTM-align server based on the url of the finished job and place them in the correct folder within your project.

Example of use:

```
> fetch gfp_21_Oct_2020 https://yanglab.nankai.edu.cn/mTM-align/output/...
```

3.2.5 analysis <name>

The command analysis is the main command of MPSA Viewer and will need mTM-align results and the structures. If you used the commands zip and fetch, you should be set already. If not, you have to place all needed structures in the "Processed_Structure" folder, and the mTM-align results in "MSTA_result". This command will both use PyMOL and the mTM-align result to produce a PyMOL session file.

Example of use:

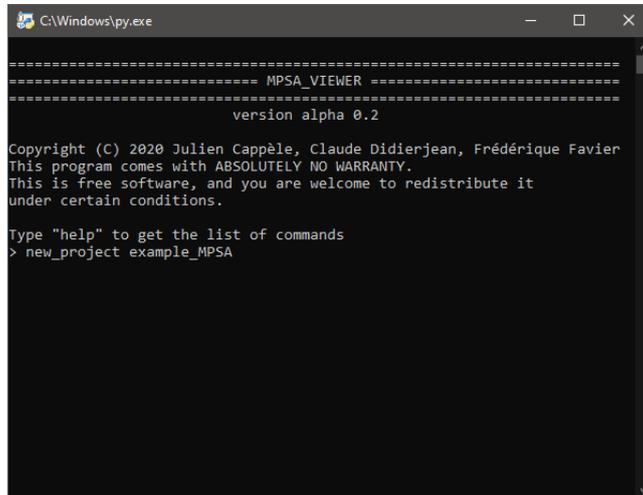
```
> analysis gfp_21_Oct_2020
```

3.3 Example of a project with MPSA Viewer

Let's dive together into the software and try it with some structures. Because a multiple protein alignment can only be used with very similar structures, we choose a set of 4 similar structures for this example.

Those are 2cc3, 4kz1, 4lso and 4mei. These 4 proteins are known as VirB8-like proteins and are involved in gram negative bacteria conjugation.

Starting with a new project, I will give it the name of "example_MPSA"



```
C:\Windows\py.exe
===== MPSA_VIEWER =====
version alpha 0.2

Copyright (C) 2020 Julien Cappèle, Claude Didierjean, Frédérique Favier
This program comes with ABSOLUTELY NO WARRANTY.
This is free software, and you are welcome to redistribute it
under certain conditions.

Type "help" to get the list of commands
> new_project example_MPSA
```

Figure 3: Creating a new project

The last message tells us that we need to modify configurations files. Because the software use PyMOL, the first thing that we need to do is to change that into the configuration_file.txt of the project. In my case, PyMOL is located at this path :

```
C:\Users\Cappèle\PyMOL\PyMOLWin.exe
```

```

C:\Windows\py.exe
=====
===== MPSA_VIEWER =====
=====
version alpha 0.2

Copyright (C) 2020 Julien Cappèle, Claude Didierjean, Frédérique Favier
This program comes with ABSOLUTELY NO WARRANTY.
This is free software, and you are welcome to redistribute it
under certain conditions.

Type "help" to get the list of commands
> new_project example_MPSA
'example_MPSA' folder created successfully
'Original_Structures' folder created successfully
'Alignement' folder created successfully
'PyMOL_Session' folder created successfully
'PyMOL_Script' folder created successfully
'Processed_Structures' folder created successfully
'Aligned_Structures' folder created successfully
'Zip_Files' folder created successfully
'MSTA_result' folder created successfully
Project example_MPSA successfully generated. You may proceed to modify s
tructure, configuration and parameters files
>

```

Figure 4: The new project has been created

The path could be different if you are using Linux or Mac. Make sure that your path is exactly pointing to the executable of PyMOL.

```

configuration_file.txt - Notepad
File Edit Format View Help
# CONFIGURATION FILE
# PyMOL_path ex: C:\Users\User\PyMOL\PyMOLWin.exe # Make sure the separators are compatible with your operating system
pymol_path = C:\Users\Cappèle\PyMOL\PyMOLWin.exe

# GENERATED BY THE PROGRAM, DO NOT TOUCH UNLESS YOU KNOW WHAT YOU ARE DOING
project_name = example_MPSA
project_folder = D:\Google Drive\MPSA_Projects\example_MPSA
msta_results_folder = D:\Google Drive\MPSA_Projects\example_MPSA\MSTA_result
structure_folder = D:\Google Drive\MPSA_Projects\example_MPSA\Original_Structures
zip_folder = D:\Google Drive\MPSA_Projects\example_MPSA\Zip_Files
fasta_folder = D:\Google Drive\MPSA_Projects\example_MPSA\Alignement
pymol_session = D:\Google Drive\MPSA_Projects\example_MPSA\PyMOL_Session
pymol_script_folder = D:\Google Drive\MPSA_Projects\example_MPSA\PyMOL_Script
processed_folder = D:\Google Drive\MPSA_Projects\example_MPSA\Processed_Structures
aligned_structures_folder = D:\Google Drive\MPSA_Projects\example_MPSA\Aligned_Structures

```

Figure 5: The configuration file

Now that PyMOL path is set, we need to specify our structures into the structure list file "structure_list.txt"

We can enter at each line a new pdb name. If you have a non published structure, just enter its name here, and you will be asked to manually place the pdb file in the "Original_Structures" folder. It is only compatible with pdb files. By default, the chain A will be used for the alignment but if you want to test another chain, you can specify it by using an underscore and the letter of the chain. This can only be used for mono-letters chains. (Figure 6)

If you want to change the PyMOL session visuals, you can modify any variable you want in the "parameters.txt" file, but we will use the default for this example, because we really like it ! ;-)

Everything is set and we can create the zip file by using the "zip" command.

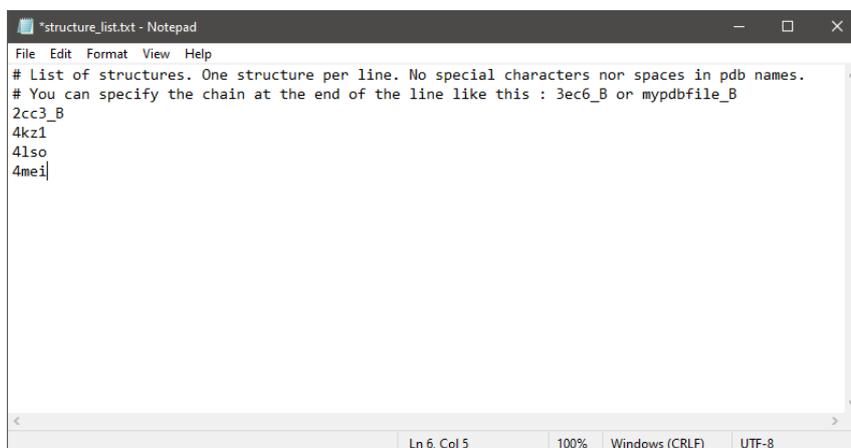


Figure 6: Structures list

```
> zip example_MPSA
```

After a longer and quite cryptic output, all structures have been downloaded from the protein data bank. Because we aren't using any personal structure, the software can handle everything by himself. If you are curious, you can check the processed structure, they should all be corrected from any seleno-methionines and have only one chain, the chain A or the one we specified.

It's time to upload our zip file to the mTM-align server! The zip file will be in the "Zip-File" folder inside your project folder.

On mTM-align, we can go to the multiple protein alignment section, and upload our zip file, then submit it.

Once the work is completed, we can paste the URL to the command fetch, in order to directly download results. Sometimes, the download can take several minutes, and is probably related to the server itself. If you don't want to wait, you can download them directly and place them in the "MSTA_results" folder

Make sure to put your project name as well, so it can place the results in the correct project.

```
> fetch example_MPSA https://yanglab.nankai.edu.cn/mTM-align/output/mTM013906/
```

The download is complete! We can start analysis now !

```
> analysis example_MPSA
```

And voila! The session is open and you can take a look at all similarities or differences within your structural alignment. It should look like the Figure 1.

Enjoy!

M9+ Milieu Minimum avec additif

Compatible avec un marquage isotopique ^{15}N et/ou ^{13}C

Pour 1 L de milieu M9+, ajouter dans l'ordre :

A ajouter	A partir de	Espèce chimique	Concentration finale
100 mL	Sels M9 (10X)	Na_2HPO_4	33,7 mM
		KH_2PO_4	22,0 mM
		NaCl	8,55 mM
10 mL	Stock 2 g/L	NH_4Cl^*	2 g/L
10 mL	Stock 2 g/L	Glucose*	2 g/L
1 mL	Stock 1 M	MgSO_4	1 mM
0,3 mL	Stock 1 M	CaCl_2	0,3 mM
Un précipité peut apparaître, mélanger vigoureusement avant de continuer.			
1 mL	Aliquots -20°C	Biotine	1 μg
1 mL	Aliquots -20°C	Thiamine	1 μg
10 mL	Stock (100X)	Oligoéléments	1X
Ajouter q.s.p. 1 L d'eau distillée stérile, ajouter les antibiotiques, et mélanger.			

* Dans le cas d'un marquage isotopique, remplacer l'espèce non marquée par l'espèce marquée. La présence d'antibiotique permet l'ajout de poudre directement dans le milieu. Pour une induction sur la nuit, il est conseillé de rajouter 2 g/L de glucose supplémentaire pour éviter la limitation de la source de carbone.

I. Solutions à autoclaver 15 min à 121 °C

1. Sels M9 10X (1 L)

Na_2HPO_4	75,2 g
KH_2PO_4	30 g
NaCl	5 g
Dissoudre les sels dans 800 mL d'eau, ajuster le pH à 7,2 avec du NaOH. Compléter jusqu'à 1 L avec de l'eau distillée.	

2. 1 M MgSO_4 (100 mL)

MgSO_4	24,65 g
Dissoudre la poudre dans 80 mL d'eau, puis compléter à 100 mL de volume avec de l'eau distillée.	

3. 1 M CaCl_2 (100 mL)

CaCl_2	14,7 g
Dissoudre la poudre dans 80 mL d'eau, puis compléter à 100 mL de volume avec de l'eau distillée.	

II. Solutions à stériliser sur filtre 0,22 µm

1. Biotine (50 mL)

Biotine	50 mg
Dissoudre la poudre dans 40 mL d'eau distillée, compléter à 50 mL puis stériliser sur filtre 0,22 µm. Aliquoter en tube de 1 mL et conserver à -20°C	

2. Thiamine (50 mL)

Thiamine-HCl	50 mg
Dissoudre la poudre dans 40 mL d'eau distillée, compléter à 50 mL puis stériliser sur filtre 0,22 µm. Aliquoter en tube de 1 mL et conserver à -20°C	

3. Solution d'oligoéléments 100X, nommée « Oligo » (1L)

Élément chimique	A ajouter	Concentration finale
EDTA	5 g	13,4 mM
Dissoudre dans 800 mL d'eau et ajuster à pH 7,2 avec du NaOH		
FeCl₃-6H₂O	830 mg	3,1 mM
ZnCl₂	84 mg	0,62 mM
CuCl₂-2H₂O	13 mg**	76 µM
CoCl₂-2H₂O	10 mg**	42 µM
H₃BO₃	10 mg**	162 µM
MnCl₂-4H₂O	1,6 mg**	8,1 µM
Ajouter tous les composants à la solution d'EDTA, compléter à 1 L avec l'eau distillée puis stériliser par autoclavage. Conserver à 4°C dans le noir.		

** Les éléments suivants sont en très faibles proportions. De façon à être précis, il est conseillé de réaliser des solutions plus concentrées de stock puis d'ajouter un certain volume, voir tableau ci-dessous.

Élément chimique	Solution Stock	Volume à ajouter***
CuCl₂-2H₂O	0,1 M (1,70 g/100 mL)	765 µL
CoCl₂-2H₂O	0,2 M (4,76 g/ 100 mL)	210 µL
H₃BO₃	0,1 M (0,62 g/100 mL)	1,6 mL
MnCl₂-4H₂O	1 M (19,8 g/ 100 mL)	8,1 µL

*** Le « volume à ajouter » remplace la « masse à ajouter » dans le tableau du dessus pour la préparation de la solution stock.

III. Tableau des différents volumes de solutions à rajouter pour différents volumes de M9 :

Volume Final (mL)	5	20	50	250	1000
Eau	4,3285	17,314	43,285	216,425	865,7
M9	0,5	2	5	25	100
NH₄Cl	0,05	0,2	0,5	2,5	10
Glucose	0,05	0,2	0,5	2,5	10
MgSO₄	0,005	0,02	0,05	0,25	1
CaCl₂	0,0015	0,006	0,015	0,075	0,3
Biotine	0,005	0,02	0,05	0,25	1
Thiamine	0,005	0,02	0,05	0,25	1
Oligo	0,05	0,2	0,5	2,5	10
Kanamycine	0,005	0,02	0,05	0,25	1

BIBLIOGRAPHIE

Bibliographie

- Abrahams, K. A. & Besra, G. S. (2018). Mycobacterial cell wall biosynthesis : a multifaceted anti-biotic target. *Parasitology*, *145*(2), 116. <https://doi.org/10.1017/S0031182016002377>
- Aguilar, J., Zupan, J., Cameron, T. A. & Zambryski, P. C. (2010). Agrobacterium type IV secretion system and its substrates form helical arrays around the circumference of virulence-induced cells. *Proc. Natl. Acad. Sci. U.S.A.*, *107*(8), 3758. <https://doi.org/10.1073/pnas.0914940107>
- Allain, F., Mareuil, F., Ménager, H., Nilges, M. & Bardiaux, B. (2020). ARIAweb : a server for automated NMR structure calculation. *Nucleic Acids Res.*, *48*(W1), 41-47. <https://doi.org/10.1093/nar/gkaa362>
- Alvarez-Martinez, C. E. & Christie, P. J. (2009). Biological Diversity of Prokaryotic Type IV Secretion Systems. *Microbiol. Mol. Biol. Rev.*, *73*(4), 775. <https://doi.org/10.1128/MMBR.00023-09>
- Álvarez-Rodríguez, I., Arana, L., Ugarte-Uribe, B., Gómez-Rubio, E., Martín-Santamaría, S., Garbisu, C. & Alkorta, I. (2020). Type IV Coupling Proteins as Potential Targets to Control the Dissemination of Antibiotic Resistance. *Frontiers in Molecular Biosciences*, *7*. <https://doi.org/10.3389/fmolb.2020.00201>
- Aly, K. A. & Baron, C. (2007). The VirB5 protein localizes to the T-pilus tips in *Agrobacterium tumefaciens*. *Microbiology (Reading)*., *153*(Pt), 11. <https://doi.org/10.1099/mic.0.2007/010462-0>
- Ambroset, C., Coluzzi, C., Guédon, G., Devignes, M.-D., Loux, V., Lacroix, T., Payot, S. & Leblond-Bourget, N. (2016). New Insights into the Classification and Integration Specificity of *Streptococcus* Integrative Conjugative Elements through Extensive Genome Exploration. *Frontiers in Microbiology*, *6*, 1483. <https://doi.org/10.3389/fmicb.2015.01483>

- Arechaga, I., Peña, A., Zunzunegui, S., del Carmen Fernández-Alonso, M., Rivas, G. & de la Cruz, F. (2008). ATPase activity and oligomeric state of TrwK, the VirB4 homologue of the plasmid R388 type IV secretion system. *J. Bacteriol.*, *190*(15), 5472-5479. <https://doi.org/10.1128/JB.00321-08>
- Arends, K., Celik, E.-K., Probst, I., Goessweiner-Mohr, N., Fercher, C., Grumet, L., Soellue, C., Abajy, M. Y., Sakinc, T., Broszat, M., Schiwon, K., Koraimann, G., Keller, W. & Grohmann, E. (2013). TraG encoded by the pIP501 type IV secretion system is a two-domain peptidoglycan-degrading enzyme essential for conjugative transfer. *J. Bacteriol.*, *195*(19), 4436-4444. <https://doi.org/10.1128/JB.02263-12>
- Ates, L. S., Houben, E. N. G. & Bitter, W. (2016). Type VII Secretion : A Highly Versatile Secretion System. *Microbiol. Spectr.*, *4*(1), . <https://doi.org/10.1128/microbiolspec.VMBF-0011-2015>
- Auchtung, J. M., Aleksanyan, N., Bulku, A. & Berkmen, M. B. (2016). Biology of ICEBs1, an integrative and conjugative element in *Bacillus subtilis*. *Plasmid*, *86*, 14-25. <https://doi.org/10.1016/j.plasmid.2016.07.001>
- Backert, S., Fronzes, R. & Waksman, G. (2008). VirB2 and VirB5 proteins : specialized adhesins in bacterial type-IV secretion systems? *Trends Microbiol.*, *16*(9), 409-413. <https://doi.org/10.1016/j.tim.2008.07.001>
- Bailey, S., Ward, D., Middleton, R., Grossmann, J. G. & Zambryski, P. C. (2006). *Agrobacterium tumefaciens* VirB8 structure reveals potential protein–protein interaction sites. *Proc. Natl. Acad. Sci. U.S.A.*, *103*(8), 2582-2587. <https://doi.org/10.1073/pnas.0511216103>
- Bantwal, R., Bannam, T. L., Porter, C. J., Quinsey, N. S., Lyras, D., Adams, V. & Rood, J. I. (2012). The peptidoglycan hydrolase TcpG is required for efficient conjugative transfer of pCW3 in *Clostridium perfringens*. *Plasmid*, *67*(2), 139-147. <https://doi.org/10.1016/j.plasmid.2011.12.016>
- Barnhart, M. M. & Chapman, M. R. (2006). Curli biogenesis and function. *Annu. Rev. Microbiol.*, *60*(131-47). <https://doi.org/10.1146/annurev.micro.60.080805.142106>
- Baron, C., Llosa, M., Zhou, S. & Zambryski, P. C. (1997). VirB1, a component of the T-complex transfer machinery of *Agrobacterium tumefaciens*, is processed to a C-terminal secreted product, VirB1. *J. Bacteriol.*, *179*(4), 1203-1210. <https://doi.org/10.1128/jb.179.4.1203-1210.1997>

- Baron, C. (2006). VirB8 : a conserved type IV secretion system assembly factor and drug target. *Biochem. Cell Biol.*, *84*(6), 890-899. <https://doi.org/10.1139/o06-148>
- Batchelor, R. A., Pearson, B. M., Friis, L. M., Guerry, P. & Wells, J. M. (2004). Nucleotide sequences and comparison of two large conjugative plasmids from different *Campylobacter* species. *Microbiology (Reading, England)*, *150*(Pt), 10. <https://doi.org/10.1099/mic.0.27112-0>
- Bateman, A. & Rawlings, N. D. (2003). The CHAP domain : a large family of amidases including GSP amidase and peptidoglycan hydrolases. *Trends in Biochemical Sciences*, *28*(5), 234-237. [https://doi.org/10.1016/S0968-0004\(03\)00061-6](https://doi.org/10.1016/S0968-0004(03)00061-6)
- Bayliss, R., Leung, S. W., Baker, R. P., Quimby, B. B., Corbett, A. H. & Stewart, M. (2002). Structural basis for the interaction between NTF2 and nucleoporin FxFG repeats. *EMBO Journal*, *21*(12), 2843. <https://doi.org/10.1093/emboj/cdf305>
- Bayliss, R., Ribbeck, K., Akin, D., Kent, H. M., Feldherr, C. M., Görlich, D. & Stewart, M. (1999). Interaction between NTF2 and xFxFG-containing nucleoporins is required to mediate nuclear import of RanGDP 11Edited by I. B. Holland. *Journal of Molecular Biology*, *293*(3), 579-593. <https://doi.org/10.1006/jmbi.1999.3166>
- Bellanger, X., Roberts, A. P., Morel, C., Choulet, F., Pavlovic, G., Mullany, P., Decaris, B. & Guédon, G. (2009). Conjugative Transfer of the Integrative Conjugative Elements ICES_{t1} and ICES_{t3} from *Streptococcus thermophilus*. *J. Bacteriol.*, *191*(8), 2764. <https://doi.org/10.1128/JB.01412-08>
- Bensing, B. A. & Sullam, P. M. (2002). An accessory sec locus of *Streptococcus gordonii* is required for export of the surface protein GspB and for normal levels of binding to human platelets. *Mol. Microbiol.*, *44*(4), 1081-1094. <https://doi.org/10.1046/j.1365-2958.2002.02949.x>
- Berger, B. R. & Christie, P. J. (1994). Genetic complementation analysis of the *Agrobacterium tumefaciens* virB operon : virB2 through virB11 are essential virulence genes. *Journal of Bacteriology*, *176*(12), 3646. <https://doi.org/10.1128/jb.176.12.3646-3660.1994>
- Berks, B. C. (1996). A common export pathway for proteins binding complex redox cofactors ? *Mol. Microbiol.*, *22*(3), 393-404. <https://doi.org/10.1046/j.1365-2958.1996.00114.x>
- Berks, B. C., Sargent, F. & Palmer, T. (2000). The Tat protein export pathway. *Mol. Microbiol.*, *35*(2), 260-274. <https://doi.org/10.1046/j.1365-2958.2000.01719.x>

- Bhoite, S., Van Gerven, N., Chapman, M. R. & Remaut, H. (2019). Curli Biogenesis : Bacterial amyloid assembly by the Type VIII secretion pathway. *EcoSal Plus*, 8(2). <https://doi.org/10.1128/ecosalplus.ESP-0037-2018>
- Büttner, D. (2012). Protein export according to schedule : architecture, assembly, and regulation of type III secretion systems from plant- and animal-pathogenic bacteria. *Microbiol. Mol. Biol. Rev.*, 76(2), 262-310. <https://doi.org/10.1128/MMBR.05017-11>
- Bottai, D., Gröschel, M. I. & Brosch, R. (2017). Type VII Secretion Systems in Gram-Positive Bacteria. *Curr. Top. Microbiol. Immunol.*, 404(235-265.). https://doi.org/10.1007/82_2015_5015
- Bradley, D. E. (1980). Morphological and serological relationships of conjugative pili. *Plasmid*, 4(2), 155-169. [https://doi.org/10.1016/0147-619x\(80\)90005-0](https://doi.org/10.1016/0147-619x(80)90005-0)
- Braunstein, M., Espinosa, B. J., Chan, J., Belisle, J. T. & Jacobs, W. R., Jr. (2003). SecA2 functions in the secretion of superoxide dismutase A and in the virulence of Mycobacterium tuberculosis. *Mol. Microbiol.*, 48(2), 453-464. <https://doi.org/10.1046/j.1365-2958.2003.03438.x>
- Brennan, P. J. & Goren, M. B. (1979). Structural studies on the type-specific antigens and lipids of the mycobacterium avium. Mycobacterium intracellulare. Mycobacterium scrofulaceum serocomplex. Mycobacterium intracellulare serotype 9. *J. Biol. Chem.*, 254(10), 4205-4211.
- Brenner, D. J., Steigerwalt, A. G. & McDade, J. E. (2020). Classification of the Legionnaires' Disease Bacterium : Legionella pneumophila, genus novum, species nova, of the Family Legionellaceae, familia nova. *Ann. Intern. Med.*
- Brook, I. (2013). Penicillin failure in the treatment of streptococcal pharyngo-tonsillitis. *Current Infectious Disease Reports*, 15(3), 232-235. <https://doi.org/10.1007/s11908-013-0338-0>
- Buchanan, G., Sargent, F., Berks, B. C. & Palmer, T. (2001). A genetic screen for suppressors of Escherichia coli Tat signal peptide mutations establishes a critical role for the second arginine within the twin-arginine motif. *Arch. Microbiol.*, 177(1), 107-112. <https://doi.org/10.1007/s00203-001-0366-2>
- Buhrdorf, R., Förster, C., Haas, R. & Fischer, W. (2003). Topological analysis of a putative virB8 homologue essential for the cag type IV secretion system in Helicobacter pylori. *Int. J. Med. Microbiol.*, 293(2), 213-217. <https://doi.org/10.1078/1438-4221-00260>

- Bunduc, C. M., Bitter, W. & Houben, E. N. G. (2020). Structure and Function of the Mycobacterial Type VII Secretion Systems. *Annu. Rev. Microbiol.*, 74(1), 315-335. <https://doi.org/10.1146/annurev-micro-012420-081657>
- Bunduc, C. M., Fahrenkamp, D., Wald, J., Ummels, R., Bitter, W., Houben, E. N. G. & Marlovits, T. C. (2020). Structure and dynamics of the ESX-5 type VII secretion system of *Mycobacterium tuberculosis*. *bioRxiv*, 2020.12.02.408906.
- Burkinshaw, B. J. & Strynadka, N. C. J. (2014). Assembly and structure of the T3SS. *Biochim. Biophys. Acta*, 1843(8), 1649-1663. <https://doi.org/10.1016/j.bbamcr.2014.01.035>
- Cappele, J., Ali, A. M., Leblond-Bourget, N., Mathiot, S., Dhalleine, T., Payot, S., Savko, M., Didierjean, C., Favier, F. & Douzi, B. (2021). Structural and Biochemical Analysis of OrfG : The VirB8-like Component of the Conjugative Type IV Secretion System of ICE St3 From *Streptococcus thermophilus*. *Frontiers in Molecular Biosciences*, 8, 642606. <https://doi.org/10.3389/fmolb.2021.642606>
- Carr, S. B., Phillips, S. E. V. & Thomas, C. D. (2016). Structures of replication initiation proteins from staphylococcal antibiotic resistance plasmids reveal protein asymmetry and flexibility are necessary for replication. *Nucleic Acids Res.*, 44(5), 2417-2428. <https://doi.org/10.1093/nar/gkv1539>
- Carraro, N. & Burrus, V. (2014). Biology of Three ICE Families : SXT/R391, ICEBs1, and ICESt1/ICESt3. *Microbiol. Spectr.*, 2(6), . <https://doi.org/10.1128/microbiolspec.MDNA3-0008-2014>
- Cascales, E. & P. J., C. (2004). Definition of a bacterial type IV secretion pathway for a DNA substrate. *Science*, 304(5674), 1170-1173. <https://doi.org/10.1126/science.1095211>
- Casu, B., Mary, C., Sverzhinsky, A., Fouillen, A., Nanci, A. & Baron, C. (2018). VirB8 homolog TraE from plasmid pKM101 forms a hexameric ring structure and interacts with the VirB6 homolog TraD. *Proc. Natl. Acad. Sci. U.S.A.*, 115(23), 5950-5955. <https://doi.org/10.1073/pnas.1802501115>
- Cattoir, V. (2016). Mechanisms of Antibiotic Resistance. *Streptococcus pyogenes : Basic Biology to Clinical Manifestations [Internet]*. University of Oklahoma Health Sciences Center.
- Chaddock, A. M., Mant, A., Karnachov, I., Brink, S., Herrmann, R. G., Klösgen, R. B. & Robinson, C. (1995). A new type of signal peptide : central role of a twin-arginine motif in transfer signals for the delta pH-dependent thylakoidal protein translocase. *EMBO J.*, 14(12), 2715-2722. <https://doi.org/10.1002/j.1460-2075.1995.tb07272.x>

- Chaillan-Huntington, C., Butler, P. J., Huntington, J. A., Akin, D., Feldherr, C. & Stewart, M. (2001). NTF2 monomer-dimer equilibrium. *Journal of Molecular Biology*, *314*(3), 465-477. <https://doi.org/10.1006/jmbi.2001.5136>
- Chaillan-Huntington, C., Braslavsky, C. V., Kuhlmann, J. & Stewart, M. (2000). Dissecting the Interactions between NTF2, RanGDP, and the Nucleoporin XFXFG Repeats *. *Journal of Biological Chemistry*, *275*(8), 5874-5879. <https://doi.org/10.1074/jbc.275.8.5874>
- Chandran, V., Fronzes, R., Duquerroy, S., Cronin, N., Navaza, J. & Waksman, G. (2009). Structure of the outer membrane complex of a type IV secretion system. *Nature*, *462*(7276), 1011. <https://doi.org/10.1038/nature08588>
- Chao, F.-A. & Byrd, R. A. (2020). Protein Dynamics revealed by NMR Relaxation Methods. *Emerging Top. Life Sci.*, *2*(1), 93. <https://doi.org/10.1042/etls20170139>
- Chatzi, K. E., Sardis, M. F., Karamanou, S. & Economou, A. (2013). Breaking on through to the other side : protein export through the bacterial Sec system. *Biochem. J.*, *449*(1), 25-37. <https://doi.org/10.1042/BJ20121227>
- Cheung, M.-S., Maguire, M. L., Stevens, T. J. & Broadhurst, R. W. (2010). DANGLE : A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *J. Magn. Reson.*, *202*(2), 223-233. <https://doi.org/10.1016/j.jmr.2009.11.008>
- Christie, P. J. & Vogel, J. P. (2000). Bacterial type IV secretion : conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol.*, *8*(8), 354-360. [https://doi.org/10.1016/s0966-842x\(00\)01792-3](https://doi.org/10.1016/s0966-842x(00)01792-3)
- Christie, P. J. (2004). Type IV secretion : the Agrobacterium VirB/D4 and related conjugation systems. *Biochim. Biophys. Acta*, *1694*(1-3), 219. <https://doi.org/10.1016/j.bbamcr.2004.02.013>
- Christie, P. J. (2016). The Mosaic Type IV Secretion Systems. *EcoSal Plus*, *7*(1), 101128-0020. <https://doi.org/10.1128/ecosalplus.ESP-0020-2015>
- Christie, P. J., Atmakuri, K., Krishnamoorthy, V., Jakubowski, S. & Cascales, E. (2005). Biogenesis, architecture, and function of bacterial type IV secretion systems. *Annual Review of Microbiology*, *59*(451-85.). <https://doi.org/10.1146/annurev.micro.58.030603.123630>
- Christie, P. J., Whitaker, N. & González-Rivera, C. (2014). Mechanism and structure of the bacterial type IV secretion systems. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1843*(8), 1578-1591. <https://doi.org/10.1016/j.bbamcr.2013.12.019>

- Cianciotto, N. P. (2005). Type II secretion : a protein secretion system for all seasons. *Trends Microbiol.*, 13(12), 581-588. <https://doi.org/10.1016/j.tim.2005.09.005>
- Clark, M. A., Choi, J. & Douglas, M. (2018). Structure of Prokaryotes : Bacteria and Archaea. *OpenStax*.
- Clewell, D. B. & Helinski, D. R. (1969). SUPERCOILED CIRCULAR DNA-PROTEIN COMPLEX IN *Escherichia coli* : PURIFICATION AND INDUCED CONVERSION TO AN OPEN CIRCULAR DNA FORM. *Proceedings of the National Academy of Sciences*, 62(4), 1159-1166. <https://doi.org/10.1073/pnas.62.4.1159>
- Cline, K., Ettinger, W. F. & Theg, S. M. (1992). Protein-specific energy requirements for protein transport across or into thylakoid membranes. Two luminal proteins are transported in the absence of ATP. *J. Biol. Chem.*, 267(4), 2688-2696. [https://doi.org/10.1016/S0021-9258\(18\)45935-2](https://doi.org/10.1016/S0021-9258(18)45935-2)
- Collaborative Computational Project, N. (1994). The CCP4 suite : programs for protein crystallography. *Acta Crystallogr., Sect. D : Biol. Crystallogr.*, 50(Pt), 5. <https://doi.org/10.1107/S0907444994003112>
- Cornelis, G. R. (2006). The type III secretion injectisome. *Nat. Rev. Microbiol.*, 4(11), 811-825. <https://doi.org/10.1038/nrmicro1526>
- Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr., Sect. D : Biol. Crystallogr.*, 62(Pt), 9. <https://doi.org/10.1107/S0907444906022116>
- Daffé, M. & Etienne, G. (1999). The capsule of *Mycobacterium tuberculosis* and its implications for pathogenicity. *Tuber. Lung Dis.*, 79(3), 153-169. <https://doi.org/10.1054/tuld.1998.0200>
- Dahmane, N. (2017). *Caractérisation des éléments intégratifs conjuguatifs de la famille ICES_{t3} et des facteurs influençant leur mobilité*. Université de Lorraine.
- Darbari, V. C. & Waksman, G. (2015). Structural Biology of Bacterial Type IV Secretion Systems. *Annu. Rev. Biochem.*, 84(603-29.). <https://doi.org/10.1146/annurev-biochem-062911-102821>
- de la Cruz, F., Frost, L. S., Meyer, R. J. & Zechner, E. L. (2010). Conjugative DNA metabolism in Gram-negative bacteria. *FEMS Microbiology Reviews*, 34(1), 18-40. <https://doi.org/10.1111/j.1574-6976.2009.00195.x>

- De Oliveira, D. M. P., Forde, B. M., Kidd, T. J., Harris, P. N. A., Schembri, M. A., Beatson, S. A., Paterson, D. L. & Walker, M. J. (2020). Antimicrobial Resistance in ESKAPE Pathogens. *Clin. Microbiol. Rev.*, *33*(3), 00181-00119. <https://doi.org/10.1128/CMR.00181-19>
- Deane, J. E., Cordes, F. S., Roversi, P., Johnson, S., Kenjale, R., Picking, W. D., Picking, W. L., Lea, S. M. & Blocker, A. (2006). Expression, purification, crystallization and preliminary crystallographic analysis of MxiH, a subunit of the *Shigella flexneri* type III secretion system needle. *Acta Crystallogr., Sect. F : Struct. Biol. Cryst. Commun.*, *62*(Pt), 3. <https://doi.org/10.1107/S1744309106006555>
- Delepelaire, P. (2004). Type I secretion in gram-negative bacteria. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1694*(1), 149-161. <https://doi.org/10.1016/j.bbamcr.2004.05.001>
- Demers, J.-P., Habenstein, B., Loquet, A., Vasa, S. K., Giller, K., Becker, S., Baker, D., Lange, A. & Sgourakis, N. G. (2014). High-resolution structure of the *Shigella* type-III secretion needle by solid-state NMR and cryo-electron microscopy. *Nat. Commun.*, *5*, 4976. <https://doi.org/10.1038/ncomms5976>
- Dickey, T. H., Altschuler, S. E. & Wuttke, D. S. (2013). Single-Stranded DNA-Binding Proteins : Multiple Domains for Multiple Functions. *Structure (London, England : 1993)*, *21*(7). <https://doi.org/10.1016/j.str.2013.05.013>
- Diepold, A. & Wagner, S. (2014). Assembly of the bacterial type III secretion machinery. *FEMS Microbiol. Rev.*, *38*(4), 802-822. <https://doi.org/10.1111/1574-6976.12061>
- Dilks, K., Rose, R. W., Hartmann, E. & Pohlschröder, M. (2003). Prokaryotic Utilization of the Twin-Arginine Translocation Pathway : a Genomic Survey. *J. Bacteriol.*, *185*(4), 1478-1483. <https://doi.org/10.1128/JB.185.4.1478-1483.2003>
- Dohlich, K., Zumsteg, A. B., Goosmann, C. & Kolbe, M. (2014). A Substrate-Fusion Protein Is Trapped inside the Type III Secretion System Channel in *Shigella flexneri*. *PLoS Pathog.*, *10*(1), e1003881. <https://doi.org/10.1371/journal.ppat.1003881>
- Dong, R., Pan, S., Peng, Z., Zhang, Y. & Yang, J. (2018). mTM-align : a server for fast protein structure database search and multiple protein structure alignment. *Nucleic Acids Research*, *46*(W1), 380-386. <https://doi.org/10.1093/nar/gky430>

- Dong, R., Peng, Z., Zhang, Y. & Yang, J. (2018). mTM-align : an algorithm for fast and accurate multiple protein structure alignment. *Bioinformatics (Oxford, England)*, *34*(10), 1719-1725. <https://doi.org/10.1093/bioinformatics/btx828>
- Dosset, P., Hus, J.-C., Blackledge, M. & Marion, D. (2000). Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR*, *16*(1), 23-28. <https://doi.org/10.1023/A:1008305808620>
- Duchêne, M.-C., Rolain, T., Knoops, A., Courtin, P., Chapot-Chartier, M.-P., Dufrêne, Y. F., Hallet, B. F. & Hols, P. (2019). Distinct and Specific Role of NlpC/P60 Endopeptidases LytA and LytB in Cell Elongation and Division of *Lactobacillus plantarum*. *Frontiers in Microbiology*, *10*. <https://doi.org/10.3389/fmicb.2019.00713>
- Dudley, E. G., Thomson, N. R., Parkhill, J., Morin, N. P. & Nataro, J. P. (2006). Proteomic and microarray characterization of the AggR regulon identifies a pheU pathogenicity island in enteroaggregative *Escherichia coli*. *Mol. Microbiol.*, *61*(5), 1267-1282. <https://doi.org/10.1111/j.1365-2958.2006.05281.x>
- Dulberger, C. L., Rubin, E. J. & Boutte, C. C. (2020). The mycobacterial cell envelope — a moving target. *Nat. Rev. Microbiol.*, *18*(1), 47-59. <https://doi.org/10.1038/s41579-019-0273-7>
- Durand, E., Oomen, C. & Waksman, G. (2010). Biochemical dissection of the ATPase TraB, the VirB4 homologue of the *Escherichia coli* pKM101 conjugation machinery. *J. Bacteriol.*, *192*(9), 2315-2323. <https://doi.org/10.1128/JB.01384-09>
- Eberhardt, R. Y., Chang, Y., Bateman, A., Murzin, A. G., Axelrod, H. L., Hwang, W. C. & Aravind, L. (2013). Filling out the structural map of the NTF2-like superfamily. *BMC Bioinformatics*, *14*(1), 1-11. <https://doi.org/10.1186/1471-2105-14-327>
- Eisenbrandt, R., Kalkum, M., Lai, E. M., Lurz, R., Kado, C. I. & Lanka, E. (1999). Conjugative pili of IncP plasmids, and the Ti plasmid T pilus are composed of cyclic subunits. *J. Biol. Chem.*, *274*(32), 22548-22555. <https://doi.org/10.1074/jbc.274.32.22548>
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr., Sect. D : Biol. Crystallogr.*, *66*(4), 486-501. <https://doi.org/10.1107/S0907444910007493>
- Evans, M. L., Chorell, E., Taylor, J. D., Åden, J., Götheson, A., Li, F., Koch, M., Sefer, L., Matthews, S. J., Wittung-Stafshede, P., Almqvist, F. & Chapman, M. R. (2015). The bacterial curli

- system possesses a potent and selective inhibitor of amyloid formation. *Mol. Cell*, 57(3), 445-455. <https://doi.org/10.1016/j.molcel.2014.12.025>
- Fercher, C., Probst, I., Kohler, V., Goessweiner-Mohr, N., Arends, K., Grohmann, E., Zangger, K., Meyer, N. H. & Keller, W. (2016). VirB8-like protein TraH is crucial for DNA transfer in *Enterococcus faecalis*. *Scientific Reports*, 6(24643), 1-13. <https://doi.org/10.1038/srep24643>
- Fernández-González, E., de Paz, H. D., Alperi, A., Agúndez, L., Faustmann, M., Sangari, F. J., Dehio, C. & Llosa, M. (2011). Transfer of R388 Derivatives by a Pathogenesis-Associated Type IV Secretion System into both Bacteria and Human Cells. *J. Bacteriol.*, 193(22), 6257. <https://doi.org/10.1128/JB.05905-11>
- Flores-Ríos, R., Moya-Beltrán, A., Pareja-Barrueto, C., Arenas-Salinas, M., Valenzuela, S., Orellana, O. & Quatrini, R. (2019). The Type IV Secretion System of ICEAfe1 : Formation of a Conjugative Pilus in *Acidithiobacillus ferrooxidans*. *Frontiers in Microbiology*, 10. <https://doi.org/10.3389/fmicb.2019.00030>
- Folkesson, A., Löfdahl, S. & Normark, S. (2002). The *Salmonella enterica* subspecies I specific centisome 7 genomic island encodes novel protein families present in bacteria living in close contact with eukaryotic cells. *Res. Microbiol.*, 153(8), 537-545. [https://doi.org/10.1016/s0923-2508\(02\)01348-7](https://doi.org/10.1016/s0923-2508(02)01348-7)
- Frain, K. M., Robinson, C. & van Dijl, J. M. (2019). Transport of Folded Proteins by the Tat System. *Protein Journal*, 38(4), 377-388. <https://doi.org/10.1007/s10930-019-09859-y>
- Frank, C., Werber, D., Cramer, J. P., Askar, M., Faber, M., An der Heiden, M., Bernard, H., Fruth, A., Prager, R., Spode, A., Wadl, M., Zoufaly, A., Jordan, S., Kemper, M. J., Follin, P., Müller, L., King, L. A., Rosner, B., Buchholz, U., . . . Krause, G. (2011). Epidemic Profile of Shiga-Toxin–Producing *Escherichia coli* O104 :H4 Outbreak in Germany. *N. Engl. J. Med.*, 365(19), 1771-1780. <https://doi.org/10.1056/NEJMoa1106483>
- Fronzes, R., Christie, P. J. & Waksman, G. (2009). The structural biology of type IV secretion systems. *Nat. Rev. Microbiol.*, 7(10), 703-714. <https://doi.org/10.1038/nrmicro2218>
- Fronzes, R., Schäfer, E., Wang, L., Saibil, H. R., Orlova, E. V. & Waksman, G. (2009). Structure of a type IV secretion system core complex. *Science (New York, N.Y.)*, 323(5911), 266. <https://doi.org/10.1126/science.1166101>

- Fullner, K. J., Lara, J. C. & Nester, E. W. (1996). Pilus assembly by *Agrobacterium* T-DNA transfer genes. *Science (New York, N.Y.)*, *273*(5278), 1107-1109. <https://doi.org/10.1126/science.273.5278.1107>
- Fullner, K. J., Stephens, K. M. & Nester, E. W. (1994). An essential virulence protein of *Agrobacterium tumefaciens*, VirB4, requires an intact mononucleotide binding domain to function in transfer of T-DNA. *Mol. Gen. Genet.*, *245*(6), 704-715. <https://doi.org/10.1007/BF00297277>
- Galán, J. E. & Wolf-Watz, H. (2006). Protein delivery into eukaryotic cells by type III secretion machines. *Nature*, *444*(7119), 567-573. <https://doi.org/10.1038/nature05272>
- Garcillán-Barcia, M. P., Francia, M. V. & de La Cruz, F. (2009). The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiology Reviews*, *33*(3), 657-687. <https://doi.org/10.1111/j.1574-6976.2009.00168.x>
- Gelvin, S. B. (2017). Integration of *Agrobacterium* T-DNA into the Plant Genome. *Annu. Rev. Genet.*, *51*, 195-217. <https://doi.org/10.1146/annurev-genet-120215-035320>
- Geourjon, C., Combet, C., Blanchet, C. & Deléage, G. (2001). Identification of related proteins with weak sequence identity using secondary structure information. *Protein Sci.*, *10*(4), 788. <https://doi.org/10.1110/ps.30001>
- Ghigo, J. M. & Wandersman, C. (1994). A carboxyl-terminal four-amino acid motif is required for secretion of the metalloprotease PrtG through the *Erwinia chrysanthemi* protease secretion pathway. *J. Biol. Chem.*, *269*(12), 8979-8985. [https://doi.org/10.1016/S0021-9258\(17\)37064-3](https://doi.org/10.1016/S0021-9258(17)37064-3)
- Ghinet, M. G., Bordeleau, E., Beaudin, J., Brzezinski, R., Roy, S. & Burrus, V. (2011). Uncovering the Prevalence and Diversity of Integrating Conjugative Elements in Actinobacteria. *PLoS ONE*, *6*(11). <https://doi.org/10.1371/journal.pone.0027846>
- Gillespie, S. H. (1998). Failure of penicillin in *Streptococcus pyogenes* pharyngeal infection. *Lancet (London, England)*, *-26*;352(9145), 1954-1956. [https://doi.org/10.1016/s0140-6736\(05\)61327-x](https://doi.org/10.1016/s0140-6736(05)61327-x)
- Goessweiner-Mohr, N., Arends, K., Keller, W. & Grohmann, E. (2013). Conjugative type IV secretion systems in Gram-positive bacteria. *Plasmid*, *70*(3), 289. <https://doi.org/10.1016/j.plasmid.2013.09.005>

- Goessweiner-Mohr, N., Arends, K., Keller, W. & Grohmann, E. (2014). Conjugation in Gram-Positive Bacteria. *Microbiol. Spectr.*, *2*(4), 0004-2013. <https://doi.org/10.1128/microbiolspec.PLAS-0004-2013>
- Goessweiner-Mohr, N., Grumet, L., Arends, K., Pavkov-Keller, T., Gruber, C. C., Gruber, K., Birner-Gruenberger, R., Kropec-Huebner, A., Huebner, J., Grohmann, E. & Keller, W. (2013). The 2.5 Å structure of the enterococcus conjugation protein TraM resembles VirB8 type IV secretion proteins. *J. Biol. Chem.*, *288*(3), 2018-2028. <https://doi.org/10.1074/jbc.M112.428847>
- Gomez, J. A. L., Bhatta, M. & Christie, P. J. (2014). PrgK, a Multidomain Peptidoglycan Hydrolase, Is Essential for Conjugative Transfer of the Pheromone-Responsive Plasmid pCF10. *Journal of Bacteriology*, *196*(3), 527. <https://doi.org/10.1128/JB.00950-13>
- Gomis-Rüth, F. X. & Coll, M. (2001). Structure of TrwB, a gatekeeper in bacterial conjugation. *International Journal of Biochemistry & Cell Biology*, *33*(9), 839-843. [https://doi.org/10.1016/s1357-2725\(01\)00060-7](https://doi.org/10.1016/s1357-2725(01)00060-7)
- Gomis-Rüth, F. X., Moncalián, G., Pérez-Luque, R., González, A., Cabezón, E., de la Cruz, F. & Coll, M. (2001). The bacterial conjugation protein TrwB resembles ring helicases and F1-ATPase. *Nature*, *409*(6820), 637-641. <https://doi.org/10.1038/35054586>
- Gomis-Rüth, F. X. & Coll, M. (2006). Cut and move : protein machinery for DNA processing in bacterial conjugation. *Current Opinion in Structural Biology*, *16*(6), 744-752. <https://doi.org/10.1016/j.sbi.2006.10.004>
- Gonzalez-Rivera, C., Bhatta, M. & Christie, P. J. (2016). Mechanism and Function of Type IV Secretion During Infection of the Human Host. *Microbiology Spectrum*, *4*(3). <https://doi.org/10.1128/microbiolspec.VMBF-0024-2015>
- Goodell, E. W. (1985). Recycling of murein by Escherichia coli. *Journal of Bacteriology*, *163*(1), 305.
- Gophna, U., Barlev, M., Seiffers, R., Oelschlager, T. A., Hacker, J. & Ron, E. Z. (2001). Curli Fibers Mediate Internalization of Escherichia coli by Eukaryotic Cells. *Infect. Immun.*, *69*(4), 2659. <https://doi.org/10.1128/IAI.69.4.2659-2665.2001>
- Gorasia, D. G., Chreifi, G., Seers, C. A., Butler, C. A., Heath, J. E., Glew, M. D., McBride, M. J., Subramanian, P., Kjaer, A., Jensen, G. J., Veith, P. D. & Reynolds, E. C. (2020). In situ structure and organisation of the type IX secretion system [[Online ; accessed 26. Feb. 2021]].

- Gorasia, D. G., Veith, P. D. & Reynolds, E. C. (2020). The Type IX Secretion System : Advances in Structure, Function and Organisation. *Microorganisms*, 8(8). <https://doi.org/10.3390/microorganisms8081173>
- Gouridis, G., Karamanou, S., Gelis, I., Kalodimos, C. G. & Economou, A. (2009). Signal peptides are allosteric activators of the protein translocase. *Nature*, 462(7271), 363-367. <https://doi.org/10.1038/nature08559>
- Green, E. R. & Mecsas, J. (2016). Bacterial Secretion Systems – An overview. *Microbiology spectrum*, 4(1). <https://doi.org/10.1128/microbiolspec.VMBF-0012-2015>
- Grohmann, E., Guzmán, L. M. & Espinosa, M. (1999). Mobilisation of the streptococcal plasmid pMV158 : interactions of MobM protein with its cognate oriT DNA region. *Mol. Gen. Genet.*, 261(4-5), 707-715. <https://doi.org/10.1007/s004380050014>
- Grohmann, E., Christie, P. J., Waksman, G. & Backert, S. (2018). Type IV secretion in Gram-negative and Gram-positive bacteria. *Molecular Microbiology*, 107(4), 455-471. <https://doi.org/10.1111/mmi.13896>
- Grohmann, E., Muth, G. & Espinosa, M. (2003). Conjugative plasmid transfer in gram-positive bacteria. *Microbiol. Mol. Biol. Rev.*, 67(2), 277-301. <https://doi.org/10.1128/mnbr.67.2.277-301.2003>
- Grossman, A. S., Mauer, T. J., Forest, K. T. & Goodrich-Blair, H. (2021). Description of a widespread bacterial secretion system with chemically diverse substrates. *bioRxiv*, 2020.01.20.912956.
- Guglielmini, J., Néron, B., Abby, S. S., Garcillán-Barcia, M. P., de la Cruz, F. & Rocha, E. P. C. (2014). Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. *Nucleic Acids Res.*, 42(9), 5715. <https://doi.org/10.1093/nar/gku194>
- Guglielmini, J., Quintais, L., Garcillán-Barcia, M. P., de la Cruz, F. & Rocha, E. P. C. (2011). The Repertoire of ICE in Prokaryotes Underscores the Unity, Diversity, and Ubiquity of Conjugation. *PLOS Genetics*, 7(8), e1002222. <https://doi.org/10.1371/journal.pgen.1002222>
- Guo, Y., Nguyen, K.-A. & Potempa, J. (2010). Dichotomy of gingipains action as virulence factors : from cleaving substrates with the precision of a surgeon's knife to a meat chopper-like brutal degradation of proteins. *Periodontol. 2000*, 54(1), 15. <https://doi.org/10.1111/j.1600-0757.2010.00377.x>

- Guzmán, L. M. & Espinosa, M. (1997). The mobilization protein, MobM, of the streptococcal plasmid pMV158 specifically cleaves supercoiled DNA at the plasmid oriT11 Edited by I. B. Holland. *Journal of Molecular Biology*, 266(4), 688-702. <https://doi.org/10.1006/jmbi.1996.0824>
- Guzmán-Herrador, D. L. & Llosa, M. (2019). The secret life of conjugative relaxases. *Plasmid*, 104, 102415. <https://doi.org/10.1016/j.plasmid.2019.102415>
- Hammar, M., Arnqvist, A., Bian, Z., Olsén, A. & Normark, S. (1995). Expression of two csg operons is required for production of fibronectin- and congo red-binding curli polymers in Escherichia coli K-12. *Mol. Microbiol.*, 18(4), 661-670. https://doi.org/10.1111/j.1365-2958.1995.mmi_18040661.x.
- Hare, S., Bayliss, R., Baron, C. & Waksman, G. (2006). A large domain swap in the VirB11 ATPase of Brucella suis leaves the hexameric assembly intact. *J. Mol. Biol.*, 360(1), 56-66. <https://doi.org/10.1016/j.jmb.2006.04.060>
- Harry H, L., Francesca, G., Rivera-Calzada, A., Braun, N., Connery, S., Dujeancourt, A., Lu, F., Redzej, A., Fronzes, R., Orlova, E. V. & Waksman, G. (2014). Structure of a type IV secretion system. *Nature*, 508(7497), 550-553. <https://doi.org/10.1038/nature13081>
- Helinski, D. R. & Clewell, D. B. (1970). Properties of a supercoiled deoxyribonucleic acid-protein relaxation complex and strand specificity of the relaxation event. *Biochemistry*, 9(22), 4428-4440. <https://doi.org/10.1021/bi00824a026>
- Henderson, I. R., Navarro-Garcia, F. & Nataro, J. P. (1998). The great escape : structure and function of the autotransporter proteins. *Trends Microbiol.*, 6(9), 370-378. [https://doi.org/10.1016/s0966-842x\(98\)01318-3](https://doi.org/10.1016/s0966-842x(98)01318-3)
- Henderson, I. R. & Nataro, J. P. (2001). Virulence Functions of Autotransporter Proteins. *Infect. Immun.*, 69(3), 1231. <https://doi.org/10.1128/IAI.69.3.1231-1243.2001>
- Henderson, I. R., Navarro-Garcia, F., Desvaux, M., Fernandez, R. C. & Ala'Aldeen, D. (2004). Type V Protein Secretion Pathway : the Autotransporter Story. *Microbiol. Mol. Biol. Rev.*, 68(4), 692-744. <https://doi.org/10.1128/MMBR.68.4.692-744.2004>
- Hennell-James, R., Deme, J. C., Kjaer, A., Alcock, F., Silale, A., Lauber, F., Johnson, S., Berks, B. C. & Lea, S. (2020). Structure of a Proton-Powered Motor that Drives Protein Transport and Gliding Motility [[Online ; accessed 4. Feb. 2021]]. <https://doi.org/10.2139/ssrn.3624460>

- Hett, E. C. & Rubin, E. J. (2008). Bacterial Growth and Cell Division : a Mycobacterial Perspective. *Microbiol. Mol. Biol. Rev.*, *72*(1), 126-156. <https://doi.org/10.1128/MMBR.00028-07>
- Heungens, K., Cowles, C. E. & Goodrich-Blair, H. (2002). Identification of *Xenorhabdus nematophila* genes required for mutualistic colonization of *Steinernema carpocapsae* nematodes. *Mol. Microbiol.*, *45*(5), 1337-1353. <https://doi.org/10.1046/j.1365-2958.2002.03100.x>
- Ho, B. T., Dong, T. G. & Mekalanos, J. J. (2014). A view to a kill : the bacterial type 6 secretion system. *Cell Host Microbe*, *15*(1), 9. <https://doi.org/10.1016/j.chom.2013.11.008>
- Hobbs, M. & Mattick, J. S. (1993). Common components in the assembly of type 4 fimbriae, DNA transfer systems, filamentous phage and protein-secretion apparatus : a general system for the formation of surface-associated protein complexes. *Mol. Microbiol.*, *10*(2), 233-243. <https://doi.org/10.1111/j.1365-2958.1993.tb01949.x>
- Holm, L. (2020). DALI and the persistence of protein shape. *Protein Sci.*, *29*(1), 128-140. <https://doi.org/10.1002/pro.3749>
- Holmström, A., Olsson, J., Cherepanov, P., Maier, E., Nordfelth, R., Pettersson, J., Benz, R., Wolf-Watz, H. & Forsberg, A. (2001). LcrV is a channel size-determining component of the Yop effector translocon of *Yersinia*. *Mol. Microbiol.*, *39*(3), 620-632. <https://doi.org/10.1046/j.1365-2958.2001.02259.x>
- Hooda, Y., Lai, C. C. L. & Moraes, T. F. (2017). Identification of a Large Family of Slam-Dependent Surface Lipoproteins in Gram-Negative Bacteria. *Front. Cell. Infect. Microbiol.*, *7*, 207. <https://doi.org/10.3389/fcimb.2017.00207>
- Hooda, Y., Lai, C. C.-L., Judd, A., Buckwalter, C. M., Shin, H. E., Gray-Owen, S. D. & Moraes, T. F. (2016). Slam is an outer membrane protein that is required for the surface display of lipidated virulence factors in *Neisseria*. *Nat. Microbiol.*, *1*, 16009. <https://doi.org/10.1038/nmicrobiol.2016.9>
- Höppner, C., Carle, A., Sivanesan, D., Hoëppner, S. & Baron, C. (2005). The putative lytic transglycosylase VirB1 from *Brucella suis* interacts with the type IV secretion system core components VirB8, VirB9 and VirB11. *Microbiology (Reading, England)*, *151*(Pt), 11. <https://doi.org/10.1099/mic.0.28326-0>
- Hormaeche, I., Alkorta, I., Moro, F., Valpuesta, J. M., Goni, F. M. & De La Cruz, F. (2002). Purification and properties of TrwB, a hexameric, ATP-binding integral membrane protein

- essential for R388 plasmid conjugation. *Journal of Biological Chemistry*, 277(48), 46456-46462. <https://doi.org/10.1074/jbc.M207250200>
- Houben, E. N. G., Korotkov, K. V. & Bitter, W. (2014). Take five — Type VII secretion systems of Mycobacteria. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1843(8), 1707-1716. <https://doi.org/10.1016/j.bbamcr.2013.11.003>
- Huang, C.-Y., Hsu, C.-H., Sun, Y.-J., Wu, H.-N. & Hsiao, C.-D. (2006). Complexed crystal structure of replication restart primosome protein PriB reveals a novel single-stranded DNA-binding mode. *Nucleic Acids Res.*, 34(14), 3878-3886. <https://doi.org/10.1093/nar/gkl536>
- Hueck, C. J. (1998). Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol. Mol. Biol. Rev.*, 62(2), 379-433.
- Ilangovan, A., Connery, S. & Waksman, G. (2015). Structural biology of the Gram-negative bacterial conjugation systems. *Trends Microbiol.*, 23(5), 301-310. <https://doi.org/10.1016/j.tim.2015.02.012>
- Ilangovan, A., Kay, C. W. M., Roier, S., Mkami, H. E., Salvadori, E., Zechner, E. L., Zanetti, G. & Waksman, G. (2017). Cryo-EM Structure of a Relaxase Reveals the Molecular Basis of DNA Unwinding during Bacterial Conjugation. *Cell*, 169(4), 708. <https://doi.org/10.1016/j.cell.2017.04.010>
- Ilyina, T. V. & Koonin, E. V. (1992). Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaeobacteria. *Nucleic Acids Research*, 20(13), 3279-3285. <https://doi.org/10.1093/nar/20.13.3279>
- Jäger, F., Lamy, A., Guerini, N., Sun, W.-S. & Berntsson, R. P.-A. (2020). Structure of the enterococcal T4SS protein PrgL reveals unique dimerization interface in the VirB8 protein family. *bioRxiv*, 2020.10.30.342212.
- Johnson, B. A. & Blevins, R. A. (1994). NMR View : A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, 4(5), 603-614. <https://doi.org/10.1007/BF00404272>
- Kabsch, W. (2010). XDS. *Acta Crystallogr., Sect. D : Biol. Crystallogr.*, 66(Pt 2), 125. <https://doi.org/10.1107/S0907444909047337>
- Kaplan, E. L., Chhatwal, G. S. & Rohde, M. (2006). Reduced ability of penicillin to eradicate ingested group A streptococci from epithelial cells : clinical and pathogenetic implications.

Clinical Infectious Diseases : an Official Publication of the Infectious Diseases Society of America, 43(11), 1398-1406. <https://doi.org/10.1086/508773>

- Karakousis, P. C., Bishai, W. R. & Dorman, S. E. (2004). Mycobacterium tuberculosis cell envelope lipids and the host immune response. *Cell. Microbiol.*, 6(2), 105-116. <https://doi.org/10.1046/j.1462-5822.2003.00351.x>
- Kharade, S. S. & 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(5), 961-970. <https://doi.org/10.1128/JB.01170-13>
- Kieser, K. J. & Rubin, E. J. (2014). How sisters grow apart : mycobacterial growth and division. *Nat. Rev. Microbiol.*, 12(8), 550-562. <https://doi.org/10.1038/nrmicro3299>
- Kikuchi, T., Mizunoe, Y., Takade, A., Naito, S. & Yoshida, S.-i. (2005). Curli fibers are required for development of biofilm architecture in Escherichia coli K-12 and enhance bacterial adherence to human uroepithelial cells. *Microbiol. Immunol.*, 49(9), 875-884. <https://doi.org/10.1111/j.1348-0421.2005.tb03678.x>
- Koronakis, V., Koronakis, E. & Hughes, C. (1989). Isolation and analysis of the C-terminal signal directing export of Escherichia coli hemolysin protein across both bacterial membranes. *EMBO J.*, 8(2), 595-605.
- Korotkov, K. V., Gonen, T. & Hol, W. G. J. (2011). Secretins : dynamic channels for protein transport across membranes. *Trends Biochem. Sci.*, 36(8), 433. <https://doi.org/10.1016/j.tibs.2011.04.002>
- Korotkov, K. V., Sandkvist, M. & Hol, W. G. J. (2012). The type II secretion system : biogenesis, molecular architecture and mechanism. *Nat. Rev. Microbiol.*, 10(5), 336-351. <https://doi.org/10.1038/nrmicro2762>
- Korotkov, K. V., Sandkvist, M. & Hol, W. G. J. (2021). The type II secretion system : biogenesis, molecular architecture and mechanism. *Nat. Rev. Microbiol.*, 10(5), 336. <https://doi.org/10.1038/nrmicro2762>
- Krissinel, E. & Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.*, 372(3), 774-797. <https://doi.org/10.1016/j.jmb.2007.05.022>
- Kubori, T., Matsushima, Y., Nakamura, D., Uralil, J., Lara-Tejero, M., Sukhan, A., Galán, J. E. & Aizawa, S. I. (1998). Supramolecular structure of the Salmonella typhimurium type III

- protein secretion system. *Science*, *280*(5363), 602-605. <https://doi.org/10.1126/science.280.5363.602>
- Kubori, T. & Nagai, H. (2016). The Type IVB secretion system : an enigmatic chimera. *Curr. Opin. Microbiol.*, *29*, 22-29. <https://doi.org/10.1016/j.mib.2015.10.001>
- Kumar, R. B., Xie, Y.-H. & Das, A. (2000). Subcellular localization of the *Agrobacterium tumefaciens* T-DNA transport pore proteins : VirB8 is essential for the assembly of the transport pore. *Mol. Microbiol.*, *36*(3), 608-617. <https://doi.org/10.1046/j.1365-2958.2000.01876.x>
- Kuroda, T., Kubori, T., Bui, X. T., Hyakutake, A., Uchida, Y., Imada, K. & Nagai, H. (2015). Molecular and structural analysis of *Legionella* DotI gives insights into an inner membrane complex essential for type IV secretion. *Sci. Rep.*, *5*, 10912. <https://doi.org/10.1038/srep10912>
- Kwak, M.-J., Kim, J. D., Kim, H., Kim, C., Bowman, J. W., Kim, S., Joo, K., Lee, J., Jin, K. S., Kim, Y.-G., Lee, N. K., Jung, J. U. & Oh, B.-H. (2017). Architecture of the type IV coupling protein complex of *Legionella pneumophila*. *Nature microbiology*, *2*, 17114. <https://doi.org/10.1038/nmicrobiol.2017.114>
- Lacroix, B. & Citovsky, V. (2018). Beyond *Agrobacterium*-Mediated Transformation : Horizontal Gene Transfer from Bacteria to Eukaryotes. *Agrobacterium Biology : From Basic Science to Biotechnology* (p. 443-462). Springer. https://doi.org/10.1007/82_2018_82
- Lakshmi, T. M. & Kim, K. S. (1989). Conjugative co-transfer of penicillin tolerance and high-level resistance to kanamycin in group A streptococci. *FEMS Microbiology Letters*, *57*(3), 329-333. [https://doi.org/10.1016/0378-1097\(89\)90323-6](https://doi.org/10.1016/0378-1097(89)90323-6)
- Langer, G. G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.*, *3*(7), 1171. <https://doi.org/10.1038/nprot.2008.91>
- Lasica, A. M., Ksiazek, M., Madej, M. & Potempa, J. (2017). The Type IX Secretion System (T9SS) : Highlights and Recent Insights into Its Structure and Function. *Front. Cell. Infect. Microbiol.*, *7*. <https://doi.org/10.3389/fcimb.2017.00215>
- Lauber, F., Deme, J. C., Lea, S. M. & Berks, B. C. (2018). Type 9 secretion system structures reveal a new protein transport mechanism. *Nature*, *564*(7734), 77-82. <https://doi.org/10.1038/s41586-018-0693-y>

- Lee, P. A., Tullman-Ercek, D. & Georgiou, G. (2006). The bacterial twin-arginine translocation pathway. *Annu. Rev. Microbiol.*, *60*(373-95.). <https://doi.org/10.1146/annurev.micro.60.080805.142212>
- Leiman, P. G. & Shneider, M. M. (2012). Contractile tail machines of bacteriophages. *Adv. Exp. Med. Biol.*, *726*(93-114.). https://doi.org/10.1007/978-1-4614-0980-9_5
- Lenz, L. L., Mohammadi, S., Geissler, A. & Portnoy, D. A. (2003). SecA2-dependent secretion of autolytic enzymes promotes *Listeria monocytogenes* pathogenesis. *Proc. Natl. Acad. Sci. U.S.A.*, *100*(21), 12432. <https://doi.org/10.1073/pnas.2133653100>
- Leone, P., Roche, J., Vincent, M. S., Tran, Q. H., Desmyter, A., Cascales, E., Kellenberger, C., Cambillau, C. & Roussel, A. (2018). Type IX secretion system PorM and gliding machinery GldM form arches spanning the periplasmic space. *Nat. Commun.*, *9*(429), 1-8. <https://doi.org/10.1038/s41467-017-02784-7>
- Lertpiriyapong, K., Gamazon, E. R., Feng, Y., Park, D. S., Pang, J., Botka, G., Graffam, M. E., Ge, Z. & Fox, J. G. (2012). *Campylobacter jejuni* type VI secretion system : roles in adaptation to deoxycholic acid, host cell adherence, invasion, and in vivo colonization. *PLoS One*, *7*(8), e42842. <https://doi.org/10.1371/journal.pone.0042842>
- Li, Y. G. & Christie, P. J. (2018). The *Agrobacterium* VirB/VirD4 T4SS : Mechanism and Architecture Defined Through In Vivo Mutagenesis and Chimeric Systems. *Curr. Top. Microbiol. Immunol.*, *418*, 233. https://doi.org/10.1007/82_2018_94
- Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., ... Adams, P. D. (2019). Macromolecular structure determination using X-rays, neutrons and electrons : recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.*, *75*(Pt), 10. <https://doi.org/10.1107/S2059798319011471>
- Llosa, M., Zupan, J., Baron, C. & Zambryski, P. (2000). The N- and C-terminal portions of the *Agrobacterium* VirB1 protein independently enhance tumorigenesis. *J. Bacteriol.*, *182*(12), 3437-3445. <https://doi.org/10.1128/jb.182.12.3437-3445.2000>
- Llosa, M. & Alkorta, I. (2017). Coupling Proteins in Type IV Secretion. *Current Topics in Microbiology and Immunology*, *413*(143-168.). https://doi.org/10.1007/978-3-319-75241-9_6

- Lorenzo-Díaz, F., Fernández-López, C., Lurz, R., Bravo, A. & Espinosa, M. (2017). Crosstalk between vertical and horizontal gene transfer : plasmid replication control by a conjugative relaxase. *Nucleic Acids Res.*, *45*(13), 7774. <https://doi.org/10.1093/nar/gkx450>
- Luirink, J. & Sinning, I. (2004). SRP-mediated protein targeting : structure and function revisited. *Biochim. Biophys. Acta*, *1694*(1-3), 17-35. <https://doi.org/10.1016/j.bbamcr.2004.03.013>
- Lynskey, N. N., Lawrenson, R. A. & Sriskandan, S. (2011). New understandings in *Streptococcus pyogenes*. *Current Opinion in Infectious Diseases*, *24*(3), 196-202. <https://doi.org/10.1097/QCO.0b013e3283458f7e>
- Mary, C. & Baron, C. (2018). Bases moléculaires de l'infection de plantes par *Agrobacterium tumefaciens* via un système de sécrétion de type IV1. *Biochem. Cell Biol.*
- McBride, M. J. & Nakane, D. (2015). *Flavobacterium* gliding motility and the type IX secretion system. *Curr. Opin. Microbiol.*, *28*, 72-77. <https://doi.org/10.1016/j.mib.2015.07.016>
- McBride, M. J. & Zhu, Y. (2013a). Gliding Motility and Por Secretion System Genes Are Widespread among Members of the Phylum Bacteroidetes. *J. Bacteriol.*, *195*(2), 270-278. <https://doi.org/10.1128/JB.01962-12>
- McBride, M. J. & Zhu, Y. (2013b). Gliding motility and Por secretion system genes are widespread among members of the phylum bacteroidetes. *J. Bacteriol.*, *195*(2), 270-278. <https://doi.org/10.1128/JB.01962-12>
- McGrath, A. E., Martyn, A. P., Whittell, L. R., Dawes, F. E., Beck, J. L., Dixon, N. E., Kelso, M. J. & Oakley, A. J. (2018). Crystal structures and biochemical characterization of DNA sliding clamps from three Gram-negative bacterial pathogens. *J. Struct. Biol.*, *204*(3), 396-405. <https://doi.org/10.1016/j.jsb.2018.10.008>
- Meir, A., Macé, K., Lukyanova, N., Chetrit, D., Hospenthal, M. K., Redzej, A., Roy, C. & Waksman, G. (2020). Mechanism of effector capture and delivery by the type IV secretion system from *Legionella pneumophila*. *Nat. Commun.*, *11*(2864), 1-11. <https://doi.org/10.1038/s41467-020-16681-z>
- Melville, S. & Craig, L. (2013). Type IV Pili in Gram-Positive Bacteria. *Microbiol. Mol. Biol. Rev.*, *77*(3), 323. <https://doi.org/10.1128/MMBR.00063-12>
- Mendiola, M. V. & de la Cruz, F. (1992). IS91 transposase is related to the rolling-circle-type replication proteins of the pUB110 family of plasmids. *Nucleic Acids Research*, *20*(13), 3521. <https://doi.org/10.1093/nar/20.13.3521>

- Michel-Souzy, S., Douzi, B., Cadoret, F., Raynaud, C., Quinton, L., Ball, G. & Voulhoux, R. (2018). Direct interactions between the secreted effector and the T2SS components GspL and GspM reveal a new effector-sensing step during type 2 secretion. *J. Biol. Chem.*, *293*(50), 19441-19450. <https://doi.org/10.1074/jbc.RA117.001127>
- Miletic, S., Fahrenkamp, D., Goessweiner-Mohr, N., Wald, J., Pantel, M., Vesper, O., Kotov, V. & Marlovits, T. C. (2021). Substrate-engaged type III secretion system structures reveal gating mechanism for unfolded protein translocation. *Nat. Commun.*, *12*(1546), 1-14. <https://doi.org/10.1038/s41467-021-21143-1>
- Mirny, L. A. & Shakhnovich, E. I. (1999). Universally conserved positions in protein folds : reading evolutionary signals about stability, folding kinetics and function11Edited by A. R. Fersht. *J. Mol. Biol.*, *291*(1), 177-196. <https://doi.org/10.1006/jmbi.1999.2911>
- Monteferrante, C. G., Miethke, M., van der Ploeg, R., Glasner, C. & van Dijl, J. M. (2012). Specific Targeting of the Metallophosphoesterase YkuE to the Bacillus Cell Wall Requires the Twin-arginine Translocation System. *J. Biol. Chem.*, *287*(35), 29789-29800. <https://doi.org/10.1074/jbc.M112.378190>
- Monteiro, P. (2016). Caractérisation du site actif de la relaxase d'ICESt3 de *Streptococcus thermophilus*. *Université de Lorraine*, 48.
- Mougous, J. D., Cuff, M. E., Raunser, S., Shen, A., Zhou, M., Gifford, C. A., Goodman, A. L., Joachimiak, G., Ordoñez, C. L., Lory, S., Walz, T., Joachimiak, A. & Mekalanos, J. J. (2006). A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science*, *312*(5779), 1526-1530. <https://doi.org/10.1126/science.1128393>
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr., Sect. D : Biol. Crystallogr.*, *67*(Pt 4), 355. <https://doi.org/10.1107/S0907444911001314>
- Nagai, H. & Kubori, T. (2011). Type IVB Secretion Systems of *Legionella* and Other Gram-Negative Bacteria. *Front. Microbiol.*, *2*, 136. <https://doi.org/10.3389/fmicb.2011.00136>
- Natale, P., Brüser, T. & Driessen, A. J. M. (2008). Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—distinct translocases and mechanisms. *Biochim. Biophys. Acta*, *1778*(9), 1735-1756. <https://doi.org/10.1016/j.bbamem.2007.07.015>

- Navarre, W. W. & Schneewind, O. (1999). Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiology and Molecular Biology Reviews : MMBR*, *63*(1), 174-229.
- Nenninger, A. A., Robinson, L. S., Hammer, N. D., Epstein, E. A., Badtke, M. P., Hultgren, S. J. & Chapman, M. R. (2011). CsgE is a curli secretion specificity factor that prevents amyloid fibre aggregation. *Mol. Microbiol.*, *81*(2), 486-499. <https://doi.org/10.1111/j.1365-2958.2011.07706.x>
- Nenninger, A. A., Robinson, L. S. & Hultgren, S. J. (2009). Localized and efficient curli nucleation requires the chaperone-like amyloid assembly protein CsgF. *Proc. Natl. Acad. Sci. U.S.A.*, *106*(3), 900-905. <https://doi.org/10.1073/pnas.0812143106>
- Newton, H. J. & Roy, C. R. (2011). The *Coxiella burnetii* Dot/Icm System Creates a Comfortable Home through Lysosomal Renovation. *mBio*, *2*(5). <https://doi.org/10.1128/mBio.00226-11>
- Pajerski, W., Ochonska, D., Brzychczy-Wloch, M., Indyka, P., Jarosz, M., Golda-Cepa, M., Sojka, Z. & Kotarba, A. (2019). Attachment efficiency of gold nanoparticles by Gram-positive and Gram-negative bacterial strains governed by surface charges. *J. Nanopart. Res.*, *21*(8), 186-192. <https://doi.org/10.1007/s11051-019-4617-z>
- Palmer, T., Finney, A. J., Saha, C. K., Atkinson, G. C. & Sargent, F. (2021). A holin/peptidoglycan hydrolase-dependent protein secretion system. *Mol. Microbiol.*, *115*(3), 345-355. <https://doi.org/10.1111/mmi.14599>
- Pansegrau, W., Ziegelin, G. & Lanka, E. (1990). Covalent association of the traI gene product of plasmid RP4 with the 5'-terminal nucleotide at the relaxation nick site. *Journal of Biological Chemistry*, *265*(18), 10637-10644. [https://doi.org/10.1016/S0021-9258\(18\)86994-0](https://doi.org/10.1016/S0021-9258(18)86994-0)
- Papanikou, E., Karamanou, S. & Economou, A. (2007). Bacterial protein secretion through the translocase nanomachine. *Nat. Rev. Microbiol.*, *5*(11), 839-851. <https://doi.org/10.1038/nrmicro1771>
- Park, D., Chetrit, D., Hu, B., Roy, C. R. & Liu, J. (2020a). Analysis of Dot/Icm Type IVB Secretion System Subassemblies by Cryoelectron Tomography Reveals Conformational Changes Induced by DotB Binding. *mBio*, *11*(1). <https://doi.org/10.1128/mBio.03328-19>
- Park, D., Chetrit, D., Hu, B., Roy, C. R. & Liu, J. (2020b). Analysis of Dot/Icm Type IVB Secretion System Subassemblies by Cryoelectron Tomography Reveals Conformational Changes Induced by DotB Binding. *mBio*, *11*(1). <https://doi.org/10.1128/mBio.03328-19>

- Parsons, J. A., Bannam, T. L., Devenish, R. J. & Rood, J. I. (2007). TcpA, an FtsK/SpoIIIE homolog, is essential for transfer of the conjugative plasmid pCW3 in *Clostridium perfringens*. *Journal of Bacteriology*, *189*(21), 7782-7790. <https://doi.org/10.1128/JB.00783-07>
- Pichichero, M. E. & Casey, J. R. (2007). Systematic review of factors contributing to penicillin treatment failure in *Streptococcus pyogenes* pharyngitis. *Otolaryngology–Head and Neck Surgery : Official Journal of American Academy of Otolaryngology-Head and Neck Surgery*, *137*(6), 851-857. <https://doi.org/10.1016/j.otohns.2007.07.033>
- Picking, W. L., Nishioka, H., Hearn, P. D., Baxter, M. A., Harrington, A. T., Blocker, A. & Picking, W. D. (2005). IpaD of *Shigella flexneri* is independently required for regulation of Ipa protein secretion and efficient insertion of IpaB and IpaC into host membranes. *Infect. Immun.*, *73*(3), 1432-1440. <https://doi.org/10.1128/IAI.73.3.1432-1440.2005>
- Pop, O., Martin, U., Abel, C. & Müller, J. P. (2002). The twin-arginine signal peptide of PhoD and the TatAd/Cd proteins of *Bacillus subtilis* form an autonomous Tat translocation system. *J. Biol. Chem.*, *277*(5), 3268-3273. <https://doi.org/10.1074/jbc.M110829200>
- Porter, C., Bantwal, R., Bannam, T., Rosado, C., Pearce, M., Adams, V., Lyras, D., Whisstock, J. & Rood, J. (2011). The conjugation protein TcpC from *Clostridium perfringens* is structurally related to the type IV secretion system protein VirB8 from Gram-negative bacteria. *Molecular microbiology*, *83*(2), 275-88. <https://doi.org/10.1111/j.1365-2958.2011.07930.x>
- Potterton, L., Agirre, J., Ballard, C., Cowtan, K., Dodson, E., Evans, P. R., Jenkins, H. T., Keegan, R., Krissinel, E., Stevenson, K., Lebedev, A., McNicholas, S. J., Nicholls, R. A., Noble, M., Pannu, N. S., Roth, C., Sheldrick, G., Skubak, P., Turkenburg, J., ... Wojdyr, M. (2018). CCP4i2 : the new graphical user interface to the CCP4 program suite. *Acta Cryst. D*, *74*(2), 68-84. <https://doi.org/10.1107/S2059798317016035>
- Pukatzki, S., Ma, A. T., Revel, A. T., Sturtevant, D. & Mekalanos, J. J. (2007). Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proc. Natl. Acad. Sci. U.S.A.*, *104*(39), 15508-15513. <https://doi.org/10.1073/pnas.0706532104>
- Pukatzki, S., Ma, A. T., Sturtevant, D., Krastins, B., Sarracino, D., Nelson, W. C., Heidelberg, J. F. & Mekalanos, J. J. (2006). Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. *Proc. Natl. Acad. Sci. U.S.A.*, *103*(5), 1528-1533. <https://doi.org/10.1073/pnas.0510322103>

- Pym, A. S., Brodin, P., Brosch, R., Huerre, M. & Cole, S. T. (2002). Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*. *Mol. Microbiol.*, *46*(3), 709-717. <https://doi.org/10.1046/j.1365-2958.2002.03237.x>
- Qiu, J. & Luo, Z.-Q. (2017). Legionella and Coxiella effectors : strength in diversity and activity. *Nat. Rev. Microbiol.*, *15*(10), 591-605. <https://doi.org/10.1038/nrmicro.2017.67>
- Radics, J., Königsmaier, L. & Marlovits, T. C. (2014). Structure of a pathogenic type 3 secretion system in action. *Nat. Struct. Mol. Biol.*, *21*(1), 82-87. <https://doi.org/10.1038/nsmb.2722>
- Redzej, A., Ukleja, M., Connery, S., Trokter, M., Felisberto-Rodrigues, C., Cryar, A., Thalassinos, K., Hayward, R. D., Orlova, E. V. & Waksman, G. (2017). Structure of a VirD4 coupling protein bound to a VirB type IV secretion machinery. *EMBO J.*, *36*(20), 3080-3095. <https://doi.org/10.15252/embj.201796629>
- Revitt-Mills, S. A., Watts, T. D., Lyras, D., Adams, V. & Rood, J. I. (2021). The ever-expanding tcp conjugation locus of pCW3 from *Clostridium perfringens*. *Plasmid*, *113*, 102516. <https://doi.org/10.1016/j.plasmid.2020.102516>
- Roberts, A. P. & Mullany, P. (2011). Tn916-like genetic elements : a diverse group of modular mobile elements conferring antibiotic resistance. *FEMS Microbiology Reviews*, *35*(5), 856-871. <https://doi.org/10.1111/j.1574-6976.2011.00283.x>
- Robinson, C. & Bolhuis, A. (2004). Tat-dependent protein targeting in prokaryotes and chloroplasts. *Biochim. Biophys. Acta*, *1694*(1-3), 135-147. <https://doi.org/10.1016/j.bbamcr.2004.03.010>
- Robinson, L. S., Ashman, E. M., Hultgren, S. J. & Chapman, M. R. (2006). Secretion of curli fibre subunits is mediated by the outer membrane-localized CsgG protein. *Mol. Microbiol.*, *59*(3), 870-881. <https://doi.org/10.1111/j.1365-2958.2005.04997.x>
- Rodrigue, A., Chanal, A., Beck, K., Müller, M. & Wu, L.-F. (1999). Co-translocation of a Periplasmic Enzyme Complex by a Hitchhiker Mechanism through the Bacterial Tat Pathway. *J. Biol. Chem.*, *274*(19), 13223-13228. <https://doi.org/10.1074/jbc.274.19.13223>
- Roman, S. A. S., Facey, P. D., Fernandez-Martinez, L., Rodriguez, C., Vallin, C., Del Sol, R. & Dyson, P. (2010). A heterodimer of EsxA and EsxB is involved in sporulation and is secreted by a type VII secretion system in *Streptomyces coelicolor*. *Microbiology (Reading)*, *156*(Pt), 6. <https://doi.org/10.1099/mic.0.037069-0>

- Sakaguchi, K., Ishibashi, T., Uchiyama, Y. & Iwabata, K. (2009). The multi-replication protein A (RPA) system – a new perspective. *FEBS J.*, *276*(4), 943-963. <https://doi.org/10.1111/j.1742-4658.2008.06841.x>
- Santoro, F., Vianna, M. E. & Roberts, A. P. (2014). Variation on a theme; an overview of the Tn916/Tn1545 family of mobile genetic elements in the oral and nasopharyngeal streptococci. *Frontiers in Microbiology*, *5*, 535. <https://doi.org/10.3389/fmicb.2014.00535>
- Sargent, F., Stanley, N. R., Berks, B. C. & Palmer, T. (1999). Sec-independent protein translocation in *Escherichia coli*. A distinct and pivotal role for the TatB protein. *J. Biol. Chem.*, *274*(51), 36073-36082. <https://doi.org/10.1074/jbc.274.51.36073>
- Sasnauskas, G., Connolly, B. A., Halford, S. E. & Siksnys, V. (2007). Site-specific DNA transesterification catalyzed by a restriction enzyme. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(7), 2115-2120. <https://doi.org/10.1073/pnas.0608689104>
- Savvides, S. N., Yeo, H.-J., Beck, M. R., Blaesing, F., Lurz, R., Lanka, E., Buhrdorf, R., Fischer, W., Haas, R. & Waksman, G. (2003). VirB11 ATPases are dynamic hexameric assemblies : new insights into bacterial type IV secretion. *EMBO J.*, *22*(9), 1969-1980. <https://doi.org/10.1093/emboj/cdg223>
- Schaar, V., Uddbäck, I., Nordström, T. & Riesbeck, K. (2014). Group A streptococci are protected from amoxicillin-mediated killing by vesicles containing β -lactamase derived from *Haemophilus influenzae*. *Journal of Antimicrobial Chemotherapy*, *69*(1), 117-120. <https://doi.org/10.1093/jac/dkt307>
- Scherzinger, E., Kruft, V. & Otto, S. (1993). Purification of the large mobilization protein of plasmid RSF1010 and characterization of its site-specific DNA-cleaving/DNA-joining activity. *European Journal of Biochemistry*, *217*(3), 929-938. <https://doi.org/10.1111/j.1432-1033.1993.tb18323.x>
- Schrödinger, LLC. (2021). *The PyMOL Molecular Graphics System, Version 2.4*
PyMOL The PyMOL Molecular Graphics System, Version 2.4, Schrödinger, LLC.
- Schroeder, M. R. & Stephens, D. S. (2016). Macrolide Resistance in *Streptococcus pneumoniae*. *Frontiers in Cellular and Infection Microbiology*, *6*, 98. <https://doi.org/10.3389/fcimb.2016.00098>
- Segal, G., Purcell, M. & Shuman, H. A. (1998). Host cell killing and bacterial conjugation require overlapping sets of genes within a 22-kb region of the *Legionella pneumophila* genome.

- Proceedings of the National Academy of Sciences of the United States of America*, 95(4), 1669-1674. <https://doi.org/10.1073/pnas.95.4.1669>
- Segal, G., Feldman, M. & Zusman, T. (2005). The Icm/Dot type-IV secretion systems of *Legionella pneumophila* and *Coxiella burnetii*. *FEMS Microbiology Reviews*, 29(1), 65-81. <https://doi.org/10.1016/j.femsre.2004.07.001>
- Sexton, J. A. & Vogel, J. P. (2002). Type IVB secretion by intracellular pathogens. *Traffic (Copenhagen, Denmark)*, 3(3), 178-185. <https://doi.org/10.1034/j.1600-0854.2002.030303.x>
- Sgro, G. G., Oka, G. U., Souza, D. P., Cenens, W., Bayer-Santos, E., Matsuyama, B. Y., Bueno, N. F., dos Santos, T. R., Alvarez-Martinez, C. E., Salinas, R. K. & Farah, C. S. (2019). Bacteria-Killing Type IV Secretion Systems. *Frontiers in Microbiology*, 10. <https://doi.org/10.3389/fmicb.2019.01078>
- Sharifahmadian, M., Nlend, I. U., Lecoq, L., Omichinski, J. G. & Baron, C. (2017). The type IV secretion system core component VirB8 interacts via the β 1-strand with VirB10. *FEBS Lett.*, 591(16), 2491-2500. <https://doi.org/10.1002/1873-3468.12770>
- Sheikh, J., Czczulin, J. R., Harrington, S., Hicks, S., Henderson, I. R., Le Bougu nec, C., Gounon, P., Phillips, A. & Nataro, J. P. (2002). A novel dispersin protein in enteroaggregative *Escherichia coli*. *J. Clin. Invest.*, 110(9), 1329-1337. <https://doi.org/10.1172/JCI16172>
- Sheldrick, G. M. (2008). A short history of SHELX. *Acta Crystallographica Section A : Foundations of Crystallography*, 64(1), 112-122.
- Sibbald, M. J. J. B., Ziebandt, A. K., Engelmann, S., Hecker, M., de Jong, A., Harmsen, H. J. M., Raangs, G. C., Stokroos, I., Arends, J. P., Dubois, J. Y. F. & van Dijl, J. M. (2006). Mapping the Pathways to Staphylococcal Pathogenesis by Comparative Secretomics. *Microbiol. Mol. Biol. Rev.*, 70(3), 755-788. <https://doi.org/10.1128/MMBR.00008-06>
- Silhavy, T. J., Kahne, D. & Walker, S. (2010). The bacterial cell envelope. *Cold Spring Harbor Perspect. Biol.*, 2(5), a000414. <https://doi.org/10.1101/cshperspect.a000414>
- Sivanesan, D. & Baron, C. (2011). The Dimer Interface of *Agrobacterium tumefaciens* VirB8 Is Important for Type IV Secretion System Function, Stability, and Association of VirB2 with the Core Complex. *J. Bacteriol.*, 193(9), 2097. <https://doi.org/10.1128/JB.00907-10>
- Soler, N., Robert, E., de Beauch ne, I. C., Monteiro, P., Libante, V., Maigret, B., Staub, J., Ritchie, D. W., Gu don, G., Payot, S., Devignes, M.-D. & Leblond-Bourget, N. (2019). Charac-

- terization of a relaxase belonging to the MOB family, a widespread family in Firmicutes mediating the transfer of ICEs. *Mobile DNA*, 10. <https://doi.org/10.1186/s13100-019-0160-9>
- Sylvestre-Gonon, E., Law, S. R., Schwartz, M., Robe, K., Keech, O., Didierjean, C., Dubos, C., Rouhier, N. & Hecker, A. (2019). Functional, Structural and Biochemical Features of Plant Serinyl-Glutathione Transferases. *Front. Plant Sci.*, 10. <https://doi.org/10.3389/fpls.2019.00608>
- Terradot, L., Bayliss, R., Oomen, C., Leonard, G. A., Baron, C. & Waksman, G. (2005). Structures of two core subunits of the bacterial type IV secretion system, VirB8 from *Brucella suis* and ComB10 from *Helicobacter pylori*. *Proceedings of the National Academy of Sciences of the United States of America*, 102(12), 4596. <https://doi.org/10.1073/pnas.0408927102>
- Thomas, J., Lee, C. A. & Grossman, A. D. (2013). A Conserved Helicase Processivity Factor Is Needed for Conjugation and Replication of an Integrative and Conjugative Element. *PLoS Genet.*, 9(1), e1003198. <https://doi.org/10.1371/journal.pgen.1003198>
- Tjalsma, H., Antelmann, H., Jongbloed, J. D. H., Braun, P. G., Darmon, E., Dorenbos, R., Dubois, J.-Y. F., Westers, H., Zanen, G., Quax, W. J., Kuipers, O. P., Bron, S., Hecker, M. & van Dijl, J. M. (2004). Proteomics of Protein Secretion by *Bacillus subtilis* : Separating the “Secrets” of the Secretome. *Microbiol. Mol. Biol. Rev.*, 68(2), 207-233. <https://doi.org/10.1128/MMBR.68.2.207-233.2004>
- Tottey, S., Waldron, K. J., Firbank, S. J., Reale, B., Bessant, C., Sato, K., Cheek, T. R., Gray, J., Banfield, M. J., Dennison, C. & Robinson, N. J. (2008). Protein-folding location can regulate manganese-binding versus copper- or zinc-binding. *Nature*, 455(7216), 1138-1142. <https://doi.org/10.1038/nature07340>
- Traxler, B. A. & Minkley, E. G. (1988). Evidence that DNA helicase I and oriT site-specific nicking are both functions of the F TraI protein. *Journal of Molecular Biology*, 204(1), 205-209. [https://doi.org/10.1016/0022-2836\(88\)90609-2](https://doi.org/10.1016/0022-2836(88)90609-2)
- Troker, M. & Waksman, G. (2018). Translocation through the Conjugative Type IV Secretion System Requires Unfolding of Its Protein Substrate. *J. Bacteriol.*, 200(6). <https://doi.org/10.1128/JB.00615-17>
- Vagin, A. & Teplyakov, A. (1997). MOLREP : an Automated Program for Molecular Replacement. *J. Appl. Crystallogr.*, 30(6), 1022-1025. <https://doi.org/10.1107/S0021889897006766>

- Van Gerven, N., Van der Verren, S. E., Reiter, D. M. & Remaut, H. (2018). The Role of Functional Amyloids in Bacterial Virulence. *J. Mol. Biol.*, *430*(20), 3657-3684. <https://doi.org/10.1016/j.jmb.2018.07.010>
- Van Gijsegem, F., Gough, C., Zischek, C., Niqueux, E., Arlat, M., Genin, S., Barberis, P., German, S., Castello, P. & Boucher, C. (1995). The hrp gene locus of *Pseudomonas solanacearum*, which controls the production of a type III secretion system, encodes eight proteins related to components of the bacterial flagellar biogenesis complex. *Mol. Microbiol.*, *15*(6), 1095-1114. <https://doi.org/10.1111/j.1365-2958.1995.tb02284.x>
- van Asselt, G. J., Vliegthart, J. S., Petit, P. L., van de Klundert, J. A. & Mouton, R. P. (1992). High-level aminoglycoside resistance among enterococci and group A streptococci. *Journal of Antimicrobial Chemotherapy*, *30*(5), 651-659. <https://doi.org/10.1093/jac/30.5.651>
- van Winden, V. J. C., Houben, E. N. G. & Braunstein, M. (2019). Protein Export into and across the Atypical Diderm Cell Envelope of Mycobacteria. *Microbiology Spectrum*, *7*(4). <https://doi.org/10.1128/microbiolspec.GPP3-0043-2018>
- Vermassen, A., Leroy, S., Talon, R., Provot, C., Popowska, M. & Desvaux, M. (2019). Cell Wall Hydrolases in Bacteria : Insight on the Diversity of Cell Wall Amidases, Glycosidases and Peptidases Toward Peptidoglycan. *Frontiers in Microbiology*, *10*. <https://doi.org/10.3389/fmicb.2019.00331>
- Vettiger, A., Winter, J., Lin, L. & Basler, M. (2017). The type VI secretion system sheath assembles at the end distal from the membrane anchor. *Nat. Commun.*, *8*(16088), 1-9. <https://doi.org/10.1038/ncomms16088>
- Vogel, J. P., Andrews, H. L., Wong, S. K. & Isberg, R. R. (1998). Conjugative transfer by the virulence system of *Legionella pneumophila*. *Science (New York, N.Y.)*, *279*(5352), 873-876. <https://doi.org/10.1126/science.279.5352.873>
- Vollmer, W., Blanot, D. & de Pedro, M. A. (2008). Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.*, *32*(2), 149-167. <https://doi.org/10.1111/j.1574-6976.2007.00094.x>
- Voth, D. E., Broederdorf, L. J. & Graham, J. G. (2012). Bacterial Type IV Secretion Systems : Versatile Virulence Machines. *Future microbiology*, *7*(2), 241. <https://doi.org/10.2217/fmb.11.150>

- Voulhoux, R., Ball, G., Ize, B., Vasil, M. L., Lazdunski, A., Wu, L.-F. & Filloux, A. (2001). Involvement of the twin-arginine translocation system in protein secretion via the type II pathway. *EMBO J.*, *20*(23), 6735-6741. <https://doi.org/10.1093/emboj/20.23.6735>
- Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J. & Laue, E. D. (2005). The CCPN data model for NMR spectroscopy : development of a software pipeline. *Proteins*, *59*(4), 687-696. <https://doi.org/10.1002/prot.20449>
- Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.*, *1*(8), 945-951.
- Wallden, K., Rivera-Calzada, A. & Waksman, G. (2010). Type IV secretion systems : versatility and diversity in function. *Cell. Microbiol.*, *12*(9), 1203. <https://doi.org/10.1111/j.1462-5822.2010.01499.x>
- Walldén, K., Williams, R., Yan, J., Lian, P. W., Wang, L., Thalassinou, K., Orlova, E. V. & Waksman, G. (2012). Structure of the VirB4 ATPase, alone and bound to the core complex of a type IV secretion system. *Proceedings of the National Academy of Sciences*, *109*(28), 11348-11353. <https://doi.org/10.1073/pnas.1201428109>
- Wang, J., Brackmann, M., Castaño-Díez, D., Kudryashev, M., Goldie, K. N., Maier, T., Stahlberg, H. & Basler, M. (2017). Cryo-EM structure of the extended type VI secretion system sheath–tube complex. *Nat. Microbiol.*, *2*(11), 1507-1512. <https://doi.org/10.1038/s41564-017-0020-7>
- Watarai, M., Makino, S.-i. & Shirahata, T. (2002). An essential virulence protein of *Brucella abortus*, VirB4, requires an intact nucleoside-triphosphate-binding domain. *Microbiology*, *148*(5), 1439-1446. <https://doi.org/10.1099/00221287-148-5-1439>
- Wells, T. J. & Henderson, I. R. (2013). Type 1 and 5 secretion systems and associated toxins. *Escherichia coli (Second Edition)* (p. 499-532). Academic Press. <https://doi.org/10.1016/B978-0-12-397048-0.00016-4>
- Williamson, M. P. (2013). Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.*, *73*, 1-16. <https://doi.org/10.1016/j.pnmrs.2013.02.001>
- Wright, L. D. & Grossman, A. D. (2016). Autonomous Replication of the Conjugative Transposon Tn916. *J. Bacteriol.*, *198*(24), 3355. <https://doi.org/10.1128/JB.00639-16>

- Yeo, H. J., Savvides, S. N., Herr, A. B., Lanka, E. & Waksman, G. (2000). Crystal structure of the hexameric traffic ATPase of the *Helicobacter pylori* type IV secretion system. *Mol. Cell*, *6*(6), 1461-1472. [https://doi.org/10.1016/s1097-2765\(00\)00142-8](https://doi.org/10.1016/s1097-2765(00)00142-8)
- Yeo, H.-J., Yuan, Q., Beck, M. R., Baron, C. & Waksman, G. (2003). Structural and functional characterization of the VirB5 protein from the type IV secretion system encoded by the conjugative plasmid pKM101. *Proc. Natl. Acad. Sci. U.S.A.*, *100*(26), 15947-15952. <https://doi.org/10.1073/pnas.2535211100>
- Youkharibache, P., Veretnik, S., Li, Q., Stanek, K. A., Mura, C. & Bourne, P. E. (2019). The Small β -Barrel Domain : A Survey-Based Structural Analysis. *Structure*, *27*(1), 6-26. <https://doi.org/10.1016/j.str.2018.09.012>
- Yuan, Q., Carle, A., Gao, C., Sivanesan, D., Aly, K. A., Höppner, C., Krall, L., Domke, N. & Baron, C. (2005). Identification of the VirB4-VirB8-VirB5-VirB2 pilus assembly sequence of type IV secretion systems. *J. Biol. Chem.*, *280*(28), 26349-26359. <https://doi.org/10.1074/jbc.M502347200>
- Zechner, E. L., Lang, S. & Schildbach, J. F. (2012). Assembly and mechanisms of bacterial type IV secretion machines. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *367*(1592), 1073. <https://doi.org/10.1098/rstb.2011.0207>
- Zechner, E. L., Moncalián, G. & de la Cruz, F. (2018). Relaxases and Plasmid Transfer in Gram-Negative Bacteria. *Type IV Secretion in Gram-Negative and Gram-Positive Bacteria* (p. 93-113). Springer. https://doi.org/10.1007/978-3-319-75241-9_4
- Zhang, Y. & Skolnick, J. (2005). TM-align : a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research*, *33*(7), 2302-2309. <https://doi.org/10.1093/nar/gki524>
- Zhu, Y. & McBride, M. J. (2014). Deletion of the *Cytophaga hutchinsonii* type IX secretion system gene sprP results in defects in gliding motility and cellulose utilization. *Appl. Microbiol. Biotechnol.*, *98*(2), 763-775. <https://doi.org/10.1007/s00253-013-5355-2>
- Zupan, J., Hackworth, C. A., Aguilar, J., Ward, D. & Zambryski, P. (2007). VirB1* promotes T-pilus formation in the vir-Type IV secretion system of *Agrobacterium tumefaciens*. *J. Bacteriol.*, *189*(18), 6551-6563. <https://doi.org/10.1128/JB.00480-07>

Résumé

Les résistances aux antibiotiques chez les bactéries font aujourd'hui partie de nos vies quotidiennes. Ces résistances progressent plus vite que l'on arrive à trouver et/ou synthétiser de nouveaux antibiotiques : un problème grandissant qui n'a aucune frontière. Pour acquérir cette résistance, les bactéries s'échangent des éléments génétiques mobiles par transfert horizontal et en utilisant le système de sécrétion de type IV (T4SS). Ces T4SS conjugatifs sont des complexes membranaires multiprotéiques, bien étudiés chez les bactéries à Gram négatif où plusieurs modèles fonctionnels existent. En revanche, chez les bactéries à Gram positif peu d'informations sont disponibles. Cette thèse avait pour but d'initier un travail structural sur les protéines d'un T4SS de *Streptococcus thermophilus*. Plus précisément, il s'agit du T4SS de l'élément intégratif et conjugatif ICE*St3*. D'une part la structure cristallographique de la protéine OrfG a été déterminée et d'autre part la structure et la dynamique d'OrfM ont été étudiés par résonance magnétique nucléaire. OrfM est majoritairement monomérique en solution et le modèle préliminaire obtenu montre un feuillet mixte à 5 brins présentant des similarités avec des protéines fixant l'ADN. La finalisation de cette structure permettra l'établissement des relations structure-fonction d'OrfM qui est supposée agir de paire avec OrfL comme un facteur de processivité d'hélicase pour les T4SS. OrfG est une protéine homologue à la protéine VirB8 qui est un facteur d'assemblage des T4SS Gram négatif. Un script appelé MPSA_Viewer a été développé dans le langage Python pour comparer OrfG à ses homologues structuraux. Cela était nécessaire car les similarités de séquence entre ces protéines étaient très faibles. Toutes ces protéines possèdent un repliement commun appelé VirB8-like mais aussi des disparités séparant les modèles Gram négatif des modèles Gram positif. Ces résultats pourraient être corrélées à leurs structures quaternaires. Un assemblage dimérique fonctionnel a été proposé chez les bactéries à Gram négatif. L'assemblage trimérique d'OrfG observé dans le cristal et conforté par les structures quaternaires similaires de deux autres protéines du même groupe suggère un T4SS original et spécifique aux bactéries à Gram positif.

Mots-clé :

système de sécrétion de type IV, élément intégratif et conjugatif, VirB8-like, cristallographie des rayons X, Résonance Magnétique Nucléaire

Summary

Antibiotic resistance in bacteria has become part of our daily lives. These resistances are progressing faster than new antibiotics can be found and/or synthesized : a growing problem that knows no boundaries. To acquire this resistance, bacteria exchange mobile genetic elements by horizontal transfer and by using the type IV secretion system (T4SS). These conjugative T4SS are multiprotein membrane complexes, well studied in Gram-negative bacteria where several functional models exist. However, in Gram-positive bacteria, little information is available. The aim of this thesis was to initiate a structural work on the proteins of a T4SS of *Streptococcus thermophilus*. More precisely, this T4SS is encoded within the integrative and conjugative element ICE*St3*. On the one hand the crystallographic structure of the OrfG protein was determined and on the other hand the structure and dynamics of OrfM were studied by nuclear magnetic resonance. OrfM is predominantly monomeric in solution and the preliminary model obtained shows a mixed 5-stranded sheet with similarities with DNA-binding proteins. The finalization of this structure will allow the establishment of structure-function relationships of OrfM which is supposed to act in pair with OrfL as a helicase processivity factor for T4SS. OrfG is a protein homologous to the VirB8 protein which is an assembly factor for Gram negative T4SS. A script called MPSA_Viewer was developed in the Python language to compare OrfG to its structural counterparts. This was necessary because the sequence similarities between these proteins were very low. All these proteins have a common fold called VirB8-like but also disparities separating Gram-negative from Gram-positive models. These results could be correlated to their quaternary structures. A functional dimeric assembly has been proposed in Gram negative bacteria. The trimeric assembly of OrfG observed in the crystal and supported by the similar quaternary structures of two other proteins of the same group suggests an original T4SS specific to Gram positive bacteria.

Keywords :

type IV secretion system, integrative and conjugative element, VirB8-like, X-ray crystallography, Nuclear Magnetic Resonance