



HAL
open science

Structure de la queue du phage T5 et mécanisme de perforation de l'enveloppe bactérienne par les Siphoviridae

Charles-Adrien Arnaud

► **To cite this version:**

Charles-Adrien Arnaud. Structure de la queue du phage T5 et mécanisme de perforation de l'enveloppe bactérienne par les Siphoviridae. Biologie structurale [q-bio.BM]. Université Grenoble Alpes, 2017. Français. NNT : 2017GREAV086 . tel-01771870

HAL Id: tel-01771870

<https://theses.hal.science/tel-01771870>

Submitted on 20 Apr 2018

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.

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : Biologie Structurale et Nanobiologie

Arrêté ministériel : 25 mai 2016

Présentée par

Charles-Adrien ARNAUD

Thèse dirigée par **Cécile Breyton**

préparée au sein de **L'Institut de Biologie Structurale**
dans **l'École Doctoral Chimie et Sciences du Vivant**

Structure de la queue du phage T5 et mécanisme de perforation de l'enveloppe bactérienne par les Siphoviridae

Structure of phage T5 tail and mechanism of bacterial envelope perforation by Siphoviridae

Thèse soutenue publiquement le **19 Octobre 2017**,
devant le jury composé de :

Présidente : Dr. Irina GUTSCHE

DIRECTRICE DE RECHERCHE, CNRS DELEGATION ALPES

Rapporteur : Dr. Stéphanie RAVAUD

MAITRE DE CONFERENCES, UNIVERSITE LYON 1

Rapporteur : Dr. Sophie ZINN-JUSTIN

DIRECTRICE DE RECHERCHE, CEA SACLAY

Examineur : Dr. Yves GAUDIN

DIRECTEUR DE RECHERCHE, CNRS DELEGATION ILE-DE-FRANCE
SUD

Directrice de thèse : Dr. Cécile BREYTON

DIRECTRICE DE RECHERCHE, CNRS DELEGATION ALPES



THÈSE

présentée pour obtenir le titre de DOCTEUR en
BIOLOGIE

STRUCTURE DE LA QUEUE DU PHAGE T5 ET MÉCANISME DE PERFORATION DE L'ENVELOPPE BACTÉRIENNE PAR LES SIPHOVIRIDAE

Charles-Adrien ARNAUD
charles.arnaud.adl@gmail.com

Soutenue publiquement le Thèse à soutenir le 19/10/2017 devant un jury composé de :

<i>Présidente</i>	Irina Gutsche	IBS - CNRS, Grenoble
<i>Rapporteuses</i>	Stéphanie Ravaud Sophie Zinn-Justin	IBCP - Université Claude Bernard, Lyon 1 I2BC - CEA, Gif-Sur-Yvette
<i>Examineur</i>	Yves Gaudin	I2BC - CNRS, Gif-sur-Yvette
<i>Directrice de thèse</i>	Cécile Breyton	IBS - CNRS, Grenoble

Résumé

La grande majorité des bactériophages connus ont un virion équipé d'une queue permettant la reconnaissance de l'hôte, la perforation de l'enveloppe bactérienne et l'éjection du matériel génétique viral directement dans le cytoplasme de la bactérie. La famille des *Siphoviridae* représente 60% des phages caudés et est caractérisée par une queue longue et non-contractile. Le tube de la queue est formé par la protéine majeur de tube (Tail Tube Protein - TTP) polymérisée autour de la protéine vernier (Tape Measure Protein - TMP). L'extrémité distale de la queue est équipée d'un complexe de protéines dans lequel se trouve les protéines de liaison au récepteur (Receptor Binding Protein - RBP). La séquence d'évènements permettant l'éjection de l'ADN et l'infection est encore mal décrite.

Au cours de cette thèse, la structure de pb6, la TTP du phage T5 qui s'assemble de façon non-canonique en trimères, a été résolue par cristallographie à une résolution de 2,2 Å. L'analyse de cette structure confirme cependant une homologie structurale de pb6 avec les autres TTPs et avec des protéines bactériennes du système de sécrétion de type VI et des pyocines R. Une étude RMN comparant pb6 dans ses états de monomère et de tube polymérisé est en cours et permettra à terme une description très fine de cet assemblage.

De plus, les structures du complexe distal de queue et du tube de la queue (tube de pb6) ont été résolues des résolutions intermédiaires avant et après interaction avec le récepteur bactérien. Ces structures obtenues par cryo-microscopie électronique révèlent une absence de changements structuraux au niveau du tube, en contradiction avec le modèle jusque-là proposé.

L'ensemble de ces données ainsi que des expériences biochimiques et la comparaison avec d'autres systèmes permettent de proposer un nouveau modèle pour les premières étapes de l'infection des *Siphoviridae*. Ce modèle a également un intérêt pour l'étude du mécanisme d'autres familles de virus (*Myoviridae*). Les différences, similarités et parentés d'éléments de la queue du phage T5 avec d'autres systèmes de perforation de membrane sont discutés.

Mots clés : virologie bactérienne, cryo-microscopie électronique, cristallographie, RMN, biochimie.

**Structure of phage T5 tail and
mechanism of bacterial envelope
perforation by Siphoviridae**

Abstract

The vast majority (96%) of bacteriophages possess a tail that allows host cell recognition, cell wall perforation and safe viral DNA channelling from the capsid to the cytoplasm of the bacterium. *Siphoviridae* is a family representing 60% of all tailed phages characterized by a long flexible tail. The tail tube is formed by stacks of hexamers of the tail tube protein (TTP) polymerised around the tape measure protein (TMP). At the distal end of the tail, the tail tip complex harbours the receptor binding proteins (RBP). For these phages, little is known on the mechanism that triggers DNA ejection after binding to the host.

We report the crystal structure at 2.2 Å resolution of pb6, an unusual trimeric TTP, of siphophage T5. Structure analysis however confirms the homology of pb6 with all TTPs and related tube proteins of bacterial puncturing devices (type VI secretion system and R-pyocin). We fit this structure into the cryo-electron microscopy map of the tail tube determined at 6 Å resolution. Comparing the structure of the tail tube before and after interaction with the host receptor, we show that unlike previously proposed, the host binding information is not propagated to the capsid by the tail tube, as the two structures, at that resolution, are identical. An ambitious NMR comparative study of the TTP in its monomeric and tube form is underway to further describe this assembly.

Moreover, the structures of the tail tip complex prior and after interaction with the bacterial receptor were solved at intermediate resolution. These structures reveal interesting conformational changes.

Together with biochemical data and comparison with other known system in the literature we are able to propose a model for *Siphoviridae* very first steps of infection. These findings might be of interest for the mechanism of other viral family (notably *Myoviridae*) and similarity with other membrane perforating systems is discussed.

Keywords: bacteriophage, cryo-electron microscopy, X-ray crystallography, NMR, biochemistry.

Remerciements

Tout d'abord, merci à Cécile Breyton de m'avoir accueilli et formé pendant mon master puis de m'avoir donné l'opportunité de réaliser cette thèse sous sa direction. Ces quatre dernières années d'abord rythmées par les soirées au synchrotron puis par les journées de purifs et de scrutations de micrographes ont été improbablement riches et intéressantes ! Merci énormément de m'avoir laissé travailler avec autant d'autonomie, d'avoir systématiquement pris le temps de m'écouter et de débattre, de m'avoir donné autant d'occasions de rencontrer différentes communautés scientifiques et aussi de m'avoir tapé sur les doigts quand j'en avais besoin. J'espère que tu as apprécié le travail de notre duo de phagistes autant que moi et que le sujet T5 perdurera !

Un grand merci à Stéphanie Ravaud et Sophie Zinn-Justin qui ont accepté de faire la révision de ce manuscrit et de faire le voyage jusqu'à Grenoble pour ma soutenance.

Merci également à Yves Gaudin et Irina Gutsche de prendre le temps de venir écouter ma soutenance et évaluer ce travail.

Ce travail ne saurait être aussi consistant sans le concours de remarquables collaborations :

- je remercie chaleureusement Eric Girard et Sylvain Engilberge pour leur aide sur tous les aspects cristallographiques du projet. Le sacrifice de nombreuses heures de sommeil sur le canapé de BM30 aura été long à porter ses fruits, mais les résultats sont enfin là !
- L'enthousiasme et la persévérance de Paul Schanda sur l'étude de pb6 en RMN ne sauraient mériter moins que toute ma gratitude et mon admiration ! Merci aussi à Vilius Kurauskas de s'être attelé à la laborieuse tâche de l'attribution de pb6 en soluble ainsi qu'à Isabel Ayala pour l'accès au produits marqués et ses conseils sur les protocoles de cultures.
- Un grand merci à Guy Schoehn et son équipe pour leur support sur tous les aspects de la microscopie électronique. Merci à Guy pour son intérêt et son soutien du projet, ainsi que sa patience pendant les longues journées derrière le microscope au cours de l'été 2016. Merci à Maria Bacia Verloop pour la préparation de grilles et les collectes d'images, à Daphna pour sa

disponibilité et sa bonne humeur à la coloration négative et aussi un grand merci à Greg Effantin et Léandro Estrozi pour m'avoir accordé du temps pour discuter et m'aider sur le traitement des données.

- Enfin, merci beaucoup à Pascale Boulanger et son équipe de virologie bactérienne pour m'avoir accueilli pour m'initier à la manipulation des phages. Merci beaucoup pour tes nombreux conseils sur les protocoles de microbiologie et biochimie ainsi que ta participation à mon CST.

Merci énormément à Wim Burmeister, sans qui le financement de cette thèse n'aurait pas pu se faire et qui a pris le temps de participer à mon CST. J'espère que d'autres occasions se présenteront pour explorer les homologues structurales de Herpès et phages !

J'aimerais remercier tous les membres des équipes SSIMPA et MIT qui participent grandement à créer un environnement de travail aussi sain et attachant : Christine Ebel et Franck Fieschi qui ont consécutivement dirigé le groupe M&P ; Aline Le Roy, Michel Thepaut, Corrine Deniaud et Isabelle Petit-Hartlein qui sont les quatre piliers d'efficacité et d'agréabilité du laboratoire.

Mes plus humbles excuses au service administratif de l'IBS qui a toujours su gérer mes oublis et retards ainsi que les différents problèmes avec bonne humeur et efficacité.

Je souhaiterais enfin remercier Béatrice Schaack et l'administration de l'UGA pour leur dévouement à faire de moi une personne plus rigoureuse, patiente et tolérante.

Charline Sabin, Aurélien Heinis et Laurent Bergeret ont évidemment toute ma reconnaissance pour leur contribution modeste mais très appréciée à l'élaboration de ce manuscrit.

Sur un plan plus personnel, un certain nombre de personnes ont été des soutiens essentiels et appréciés au cours des dernières années. Cependant, ayant épuisé le nombre de formulations contenant le mot « merci » (le lecteur attentif n'aura pas eu de mal à repérer les nombreuses répétitions, je vous prie néanmoins d'excuser ces maladresses qui ne sauraient entâcher la sincérité de ma gratitude) je pense qu'ils sauront me pardonner un style assez sobre.

Les doctorants et post-doctorants du bureau 369 (présent et passé) sont de loin ceux qui ont eu à me supporter sur les plus grandes plages horaires : Vaness', Pasquale, Annelise, Gaëtan et Janek. Une pensée particulière à ma povera malata di stonzetta gattina, une inoubliable camarade de laboratoire et une amie très chère.

Ceux qui me supportent depuis encore plus longtemps et qui continuent malgré tout à me soutenir sont mes parents et ma famille. Vous avez tant fait pour m'aider et me pousser à continuer, je vous en suis infiniment reconnaissant.

De nombreuses autres personnes que j'ai à citer ici peuvent se subdiviser en 4 groupes :

— IBS

Le nombre de gens que j'ai rencontrés ici, avec qui j'ai eu l'opportunité de discuter, qui m'ont aidé pour du matériel ou des protocoles est tout à fait incroyable... Citons rapidement Hugues Nury, Nicolas Coquelle et Philippe Frachet pour leur discussion et leur aide sur mon projet de thèse.

Les membres de la légendaire « team Porquerolles », dont je me plaît à croire être le cadet, Kermadec, Louise, Alex, Fichoux, Didier... J'espère avoir l'occasion de vous recroiser bientôt.

Les membres de l'expédition islandaise : Sil', Laura, Marko et Quentin. Et Jimmy bien sûr.

Et bien mes compagnons du soir : Joyce, Simon, Raleb, Muge, Sylvain (et Sylvain aussi dans une moindre mesure), Emilie, Gaëlle, Daniel, et aussi Orso, Audrey, Ambroise et Caroline.

— Les « chimistes » Grenoblois et Glob'Alps

Je me dois de commencer par Gaff' et Am', l'adorable couple qui m'a aidé à faire mon nid à Grenoble, qui comptent parmi mes amis les plus proches et m'ont permis de rencontrer un grand nombre de personnages hauts en couleur : Laz, Pau, Charly, Kiki, Geoff', Bobby, Nono, Toc, Binou, Cathy et bien d'autres.

Les membres du bureau de l'association Glob'Alps dans laquelle je me suis modestement investi sont des individus inspirants et agréables, en particulier Présidente Marlène, Félix, Carine, Marie et Racine. (D'ailleurs, cher lecteur, si tu te trouves être un stagiaire ou un jeune doctorant qui a ouvert ce manuscrit pour voir-à-quoi-ressemble-une-thèse-ou-pour-trouver-un-protocole-particulier-mais-qui-finit-par-juste-lire-les-remerciements-parce-que-c'est-le-plus-rigolo, je t'invite à te renseigner sur ce qu'organise cette asso à ton époque, que ce soit les afterworks ou les visites d'entreprises, on s'est bien marré).

Merci également à Philou, pour ses subventions assassines de fin de soirée et à Stouf, un colloc de qualité.

— Les Jurassiens

Les nobles terres des Coteaux de Seille auront été gâtées d'une riche génération de zouaves qui persistent à garder le contact malgré le temps et la distance. Vincent, Ary, Hreinis, cela fait près de 20 ans qu'on traîne ensemble, il ne nous reste qu'à repartir pour 20 de plus ! Une pensée affectueuse à ceux que je ne vois plus si souvent ainsi qu'à Zouzou et clèm' (qui malgré leurs origines doubistes ont moultement démontré leur aptitude à surmonter ce lourd handicap;p).

N'oublions pas les

— Le Pays de Montbéliard

Une compagnie caractérisée par sa belle âme. Aurèl', Laurent, Antoine, Hocine,

Remerciements

Justine et bien d'autres, les soirées à Valentigney, nos périodes de colocation provisoires et nos weekends de réunions ont été autant de bouffées d'air pur au cours des 4 dernières années (plus ou moins pur mais oxygénant néanmoins. A cet égard, il n'est peut être pas nécessaire de trop regretter la fin précoce de la machine à fumée...). Merci également à Pauline, qui m'a initié aux charmes du pays cathare.

Merci enfin à Alexandra dont l'énergie, l'enthousiasme et l'intégrité ne cesseront probablement jamais de m'ébahir.

La taille de ces remerciements devenant démesurée, je vais y mettre un terme en souhaitant au cher lecteur de trouver les informations qu'il cherche dans ce manuscrit et d'y apprendre quelque chose sur le monde merveilleux de la microbiologie et des queues de phages.

Table des matières

Résumé	i
Abstract	v
Remerciements	vii
Table des matières	xi
Introduction	1
1 Introduction biologique	3
1.1 Les bactériophages	3
1.1.1 Historique	3
1.1.1.1 Découverte des bactériophages et phagothérapie	3
1.1.1.2 Phages et biologie moléculaire	4
1.1.1.3 Enjeux de la recherche sur les phages aujourd’hui	5
1.1.2 Classification	5
1.1.3 Cycle de reproduction viral	8
1.2 Place des bactériophages dans la biosphère	9
1.2.1 <i>Terra incognita</i> microbiologique et virale	9
1.2.2 Rôle de la virosphère dans l’environnement	10
1.2.3 Interaction avec d’autres systèmes biologiques	11
1.3 Applications	11
1.3.1 Biologie moléculaire & biotechnologies	11
1.3.2 Médecine	12
1.3.2.1 Phagothérapie	12
1.3.2.2 Evaluation de l’efficacité et de l’innocuité de la phagothérapie	13
1.3.2.3 Traitements dérivés des phages	14

TABLE DES MATIÈRES

1.3.3	Agroalimentaire	14
1.3.3.1	Enjeux de l'étude des phages pour l'industrie agroalim- entaire	14
1.3.3.2	Phagothérapie animale et végétale	15
1.3.4	Impact environnemental de l'utilisation des phages	16
1.3.5	Détection	16
1.3.5.1	Détection de la lyse	17
1.3.5.2	Détection de l'amplification des phages	17
1.3.5.3	Détection par phage rapporteur génétiquement mod- ifié	17
1.3.5.4	Détection par adsorption de phages ou protéines de phage	17
1.4	Perforation de la paroi bactérienne et éjection de l'ADN phagique .	18
1.4.1	Les <i>Podoviridae</i>	18
1.4.1.1	T7	18
1.4.1.2	P22	22
1.4.2	Les <i>Myoviridae</i>	25
1.4.2.1	T4	26
1.4.3	Les <i>Siphophages</i>	31
1.4.3.1	SPP1	32
1.4.3.2	Lactophages	35
1.5	Le bactériophage T5	38
1.5.1	Spécificités de T5	38
1.5.2	Génome	38
1.5.3	Aspects structuraux	39
1.5.4	Assemblage des virions	43
1.5.5	Mécanisme d'infection	43
	Points clés	45
2	Introduction méthodologique	47
2.1	Introduction	47
2.2	Microscopie électronique	48
2.2.1	Généralités	48
2.2.1.1	Historique	48
2.2.1.2	Coloration négative et cryo-microscopie électronique	49
2.2.1.3	Interaction électron-matière	50
2.2.2	Anatomie d'une microscopie électronique à transmission . . .	50
2.2.2.1	Canon à électrons	50
2.2.2.2	Lentille électronique	51
2.2.2.3	colonne du microscope	51

2.2.3	Formation des images en microscopie électronique à transmission	53
2.2.3.1	Contraste d'amplitude et contraste de phase	53
2.2.3.2	Fonction de transfert de contraste	53
2.2.4	Révolution apporté par les détecteurs à détection directe d'électrons	56
2.2.5	Traitement de données	56
2.3	Cristallographie	57
2.3.1	Généralités	57
2.3.2	Cristallogénèse	58
2.3.3	Diffraction des rayons X	60
2.4	Modélisation	60
	Points clés	62

Tube de la queue de T5 63

3 Tube de la queue du phage T5 : contexte & méthodes 65

3.1	Contexte	65
3.2	Matériel et méthodes	66
3.2.1	Production de pb6	66
3.2.1.1	Cultures en milieu riche	66
3.2.1.2	Cultures en milieu minimum	66
3.2.1.3	Production de protéines marquées	66
3.2.2	Purification et caractérisation de pb6 monomère	67
3.2.2.1	Lyse des cellules	67
3.2.2.2	Chromatographie d'affinité au nickel	67
3.2.2.3	Chromatographie échangeuse d'ions	67
3.2.2.4	Dessalage	68
3.2.2.5	Electrophorèse sur gel de polyacrylamide laurylsulfate de sodium (SDS-PAGE)	68
3.2.2.6	Spectroscopie ultraviolet/visible et concentration	68
3.2.2.7	Filtration sur gel analytique	69
3.2.3	Purification et caractérisation des fibres de pb6	69
3.2.3.1	Resolubilisation	69
3.2.3.2	Gradient de sucrose	69
3.2.3.3	Microscopie électronique à transmission (MET) – Coloration négative	69
3.2.4	Tests de polymérisation de pb6	69
3.2.4.1	Filtration sur gel analytique	69
3.2.4.2	Fluorimétrie	70

3.2.5	Mutants de pb6	70
3.2.5.1	Obtention des plasmides des mutants	70
3.2.6	Cristallogénèse de pb6	70
3.2.7	Cristallographie de pb6	71
3.2.7.1	Collecte des données de diffraction de rayons X	71
3.2.7.2	Obtention d'une première carte de densité	71
3.2.7.3	Résolution et affinement de la structure de pb6	71
3.2.8	Cryo-microscopie électronique	72
3.2.8.1	Collecte d'images de fibres	72
3.2.8.2	Obtention du modèle initial	72
3.2.8.3	Obtention des structures de tube	72
3.2.9	Modélisation	73
4	Tube de la queue du phage T5 : données biochimiques sur pb6	75
4.1	Purification de pb6	75
4.2	Polymérisation de pb6	79
4.2.1	Polymérisation initiale à la lyse des cellules	79
4.2.2	Polymérisation de monomère purifié	82
4.2.2.1	Filtration sur gel	82
4.2.2.2	Expériences de fluorescence	82
4.3	Introduction de l'article	85
	Points clés	86
5	Tubes de la queue du phage T5 : données structurales et déclenchement de l'éjection de l'ADN	87
5.1	Article	87
5.2	Suppléments	97
	Points clés	111
6	Tube de la queue du phage T5 : préparation d'échantillons pour la RMN & perspectives	113
6.1	Préparation des échantillons	113
6.1.1	pb6 entière	113
6.1.2	Mutants de pb6	118
6.2	Apport des expériences et perspectives	118
6.2.1	Domaine Ig-like	118
6.2.2	Dynamique de pb6	121
6.2.3	Combinaison d'EM et de RMN	121
	Points clés	124

Complexe distal de la queue de T5 et interaction avec la membrane	125
7 Complexe distal de la queue du phage T5 : contexte & méthodes	127
7.1 Contexte	127
7.2 Matériel et méthodes	127
7.2.1 Amplification de T5sto et T5amD20	127
7.2.2 Purification et caractérisation de T5sto et T5amD20	128
7.2.3 Production et purification de queues de phage T5	128
7.2.4 Production de FhuA	128
7.2.5 Purification de FhuA	129
7.2.6 Purification de MSP1E3D1 (similaire pour MSP1D1)	129
7.2.7 Purification et caractérisation de FhuA en nanodisques	130
7.2.8 Microscopie électronique	131
7.2.8.1 Préparation de grilles pour la coloration négative	131
7.2.8.2 Préparation de grilles pour la cryo-microscopie	131
7.2.8.3 Traitement des données	131
8 Complexe distal de queue du phage T5 : résultats & perspectives	133
8.1 Préparation des échantillons	133
8.1.1 Queues de T5	133
8.1.1.1 T5amD20-Amplification et contrôle de la souche	133
8.1.1.2 Production de queues de phages	134
8.1.1.3 Purification de queues de phages	134
8.1.2 FhuA	138
8.1.2.1 Nanodisques de FhuA	138
8.2 Structure du complexe distal de queue	140
8.2.1 Structure du complexe distal de queue avant interaction avec FhuA	140
8.2.1.1 Coloration négative	140
8.2.1.2 Cryo-microscopie électronique	143
8.2.2 Structure du complexe distal de queue en interaction avec FhuA	149
8.2.2.1 Coloration négative	149
8.2.2.2 Cryo-microscopie électronique	149
8.3 Perspectives	154
8.3.1 Expériences complémentaires	154
8.3.1.1 Protéolyse ménagée	154
8.3.1.2 Expériences à réaliser	156
Points clés	157

Conclusion	159
Annexes	165
A Article : Solid-state NMR H-N-(C)-H and H-N-C-C 3D/4D correlation experiments for resonance assignment of large proteins	167
B Article : Crystallophore: a versatile lanthanide complex for protein crystallography combining nucleating effects, phasing properties, and luminescence	175
C Figures complémentaires	187
C.1 Purification de queues de phages sur colonne échangeuse d’ion : exploration des quantités relatives des protéines du complexe. . . .	189
C.2 Prediction de structure et alignement de la séquence de p140	190
C.3 Protéolyse ménagée de queues de T5 : compléments	190
Bibliographie	191

Liste des Figures

- Figure 1.1** : Répartition des virus par Ordre.
- Figure 1.2** : Tableau des 11 familles de bactériophages connus.
- Figure 1.3** : Structure du phage T7.
- Figure 1.4** : Structure du phage P22.
- Figure 1.4** : Contraction de la queue du phage T4 pendant la perforation de la paroi de la cellule.
- Figure 1.5** : Structures du connecteur et du tube du phage SPP1 avant et après interaction avec le récepteur YueB.
- Figure 1.6** : Structures des lactophages p2 et TP901-1.
- Figure 1.7** : Queue du bactériophage T5.
- Figure 2.1** : Interaction électron-matière.
- Figure 2.2** : Colonne d'un microscope électronique.
- Figure 2.3** : Concept de la correction de CTF des images.
- Figure 2.4** : Concept de la reconstruction 3D.
- Figure 2.5** : Diagramme de phase.
- Figure 4.1** : Purification de pb6 monomères, comparaison de l'expression en milieu riche et en milieu minimum.
- Figure 4.2** : Purification de pb6.
- Figure 4.3** : Effet des conditions de lyse et de culture sur la polymérisation de pb6.
- Figure 4.4** : Comparaison du gradient de sucrose de la fraction insoluble dans des conditions favorisant ou non la polymérisation de pb6.
- Figure 4.5** : Induction de la polymérisation du monomère.
- Figure 6.1** : Purification des différents échantillons destinés à la RMN.
- Figure 6.2** : Microscopie électronique à coloration négative de différentes formes de pb6 précipités pendant la préparation d'échantillons marqués.
- Figure 6.3** : Comparaisons de spectres RMN 2D.
- Figure 6.4** : Données RMN du domaine C-terminale de pb6.
- Figure 6.5** : Stratégie d'étude des différents échantillons de pb6 par RMN.
- Figure 8.1** : Purification des queues de phage T5.
- Figure 8.2** : SDS-PAGE de queues de phages purifiées.

Figure 8.3 : Purification de la protéine MSP et assemblage de nanodisques avec FhuA.

Figure 8.4 : Micrographes de queues de phages.

Figure 8.5 : Micrographes de queues de phages un mois après purification.

Figure 8.6 : Reconstruction 3D du complexe distal de queue.

Figure 8.7 : Reconstruction asymétrique du complexe distal de la queue du phage T5.

Figure 8.8 : Reconstruction 3D du complexe distal de queue - décomposition des éléments du complexe.

Figure 8.9 : Micrographes de queues de phages incubées avec FhuA.

Figure 8.10 : Reconstruction 3D du complexe distal de queue en interaction avec FhuA-Nanodisque.

Figure 8.11 : Comparaison des différents échantillons.

Figure 8.12 : Protéolyse ménagée des queues de T5.

Figure 9.1 : Schéma du mécanisme d'entrée de T5.

Figure 9.2 : Comparaison du tube interne de T4 avec T5.

Introduction

"If we all reacted the same way, we'd be predictable, and there's always more than one way to view a situation. What's true for the group is also true for the individual. It's simple: overspecialize and you breed in weakness. It's slow death."

(Kasunori Ito)

Chapter 1

Introduction biologique

1.1 Les bactériophages

1.1.1 Historique

1.1.1.1 Découverte des bactériophages et phagothérapie

En 1915, Frederick Twort, qui travaillait sur le développement d'un milieu de culture pour le virus de la vaccine, décrit un phénomène de « lyse transmissible ». Cet événement marque la découverte, par sérendipité, du monde des phages. Deux ans plus tard, Félix D'Hérelle publie une découverte similaire d'un élément qui persiste dans les solutions filtrées avec un filtre de Chamberland, capable de provoquer la lyse des bactéries et de s'amplifier. Contrairement à Twort, il avance l'hypothèse qu'il s'agit d'organismes biologiques capables de se reproduire en se « nourrissant » des bactéries et les baptise « bactériophages ». D'Hérelle voit déjà dans ces bactériophages un modèle d'étude des mécanismes fondamentaux de la vie mais aussi un outil pour traiter les infections bactériennes. Il travailla longtemps au développement d'une phagothérapie efficace, traitant dès 1919 des patients atteints de dysenterie à l'hôpital des enfants-malades à Paris. Le « phénomène Twort-D'Hérelle » et son applicabilité thérapeutique furent initialement accueillis avec beaucoup d'enthousiasme malgré les questions en suspens sur la nature des bactériophages. Des travaux semblèrent confirmer leur potentiel thérapeutique et leur innocuité, et la production et la vente de solutions de phages commerciales se développa en France, en Inde (où D'Hérelle avait entrepris d'essayer d'appliquer la phagothérapie contre des épidémies de choléra et de peste), au Brésil et aux Etats Unis [Chanishvili 12, Sulakvelidze 01, Dublanchet 07, Summers 01]. Cependant, la phagothérapie sombra dans l'oubli pour deux raisons :

- Les solutions commerciales de phages qui commencèrent à se développer étaient peu/pas efficaces. En effet, il n'existait pas de méthodologie pour

titrer ou évaluer la spécificité d'une solution de phages et les phénomènes de lysogénie étaient inconnus.

- Après la seconde guerre mondiale, la pénicilline et les antibiotiques s'imposèrent comme solution antibactérienne dans le monde occidental.

La phagothérapie continua cependant à être utilisée dans le bloc de l'Est, notamment à partir de la Géorgie dans un institut fondé par Félix d'Hérelle et Giorgi Eliava (institut toujours opérationnel aujourd'hui). Dans le monde occidental, les phages devinrent un outil d'étude en science fondamentale.

1.1.1.2 Phages et biologie moléculaire

L'étude des phages est intrinsèquement liée au développement de la biologie moléculaire, apportant notamment des informations capitales sur le rôle et le fonctionnement de l'ADN pendant les balbutiements de la biologie moléculaire.

En 1943, une expérience de Luria et Delbrück démontre que l'apparition de la résistance à un phage n'est pas induite par le contact de la bactérie avec le virus mais qu'elle apparaît aléatoirement. Ce travail montre expérimentalement l'apparition de mutations aléatoires dans le matériel génétique, un des prérequis de la théorie de l'évolution [Luria 43].

L'expérience Hershey-Chase s'intéressait à la nature des phages durant leur état végétatif (infection virale non productive : lysogénie ; cf. 1.1.3). Le suivi de marqueurs radioactifs a montré que l'ADN était injecté dans les bactéries infectées alors que les protéines restaient en dehors de la cellule [HERSHEY 52]. Cette expérience a fait la démonstration que l'ADN est le support de l'information génétique.

Alfred Hershey, Salvador Luria et Max Delbrück ont été récompensés par un prix Nobel en 1969 « pour leurs découvertes concernant le mécanisme de la réplication et la structure génétique des virus ».

Au début des années 50, plusieurs observations indépendantes furent faites de phages incapables d'infecter une souche différente de la même espèce que leur hôte. Bertani & Weigle mirent en évidence des capacités d'adaptations dont l'efficacité ne pouvait pas être expliquée par un phénomène de mutation/sélection. L'observation de cette adaptation non héréditaire à l'hôte mènera à la découverte des systèmes de restrictions [BERTANI 53, Meselson 68, Murray 00]. En 1961, l'infectivité de mutants du phage T4 pour différentes souches d'*E. coli* permit de démontrer que l'information génétique est non-chevauchante, que le départ des gènes est déterminé et que le code génétique est dégénéré et utilise des codons de 3 bases [CRICK 61, van Ooij 11].

1.1.1.3 Enjeux de la recherche sur les phages aujourd'hui

En science fondamentale, les phages et leurs interactions avec leurs hôtes sont toujours des outils importants dans la compréhension de mécanismes fondamentaux de la biologie moléculaire, les systèmes CRISPR-CAS sont des systèmes de défense bactérienne contre les phages, qui ont récemment été utilisés et adaptés comme puissant outil de biologie moléculaire [Marraffini 15]. Les études réalisées au XXème siècle s'étant volontairement concentrées sur un nombre limité de phages pour faciliter la comparaison des résultats, il est certain que l'étude de nouveaux systèmes hôte-virus complètera notre compréhension du monde viral.

La phagothérapie redevient aujourd'hui un sujet d'intérêt en tant qu'alternative/complément à l'utilisation des antibiotiques (contre lesquelles on observe un nombre inquiétant et croissant de souches résistantes). D'autres applications peuvent être développées en utilisant les phages, un certain nombre d'exemples seront proposés dans ce manuscrit (cf. 1.3).

1.1.2 Classification

Les virus sont classés sur la base de la nature du matériel génétique porté par le virion et des caractéristiques du virion. Le comité international de la taxonomie des virus (ICTV) répertorie 4404 espèces de virus classées dans 8 ordres et 130 familles. 85 familles regroupant 2526 espèces ne sont classées dans aucun ordre [ICT] (Fig 1.1).

Les bactériophages connus sont classés dans 11 familles (Fig. 1.2). Il existe des phages avec des matériels génétiques double ou simple brins, ADN ou ARN et des virions aux morphologies variées : filamenteuse (*Inoviridae*), enveloppée (*Plasmaviridae*), double capsid (*Tectiviridae*), sphérique (*Leviviridae*)... Cependant, la grande majorité des phages connus appartient à l'ordre des *Caudovirales* : une étude de microscopie électronique portant sur la morphologie de plus de 5500 phages montre que 96% d'entre eux sont des virus caudés [Ackermann 09]. Ces virus se caractérisent par une capsid protéique équipée d'une queue, un ADN double-brin empaqueté de façon très dense dans la capsid et des cycles de production lytiques (cf. 1.1.3). Les 3 familles de *Caudovirales* se distinguent par l'architecture de leur queue : les *Myoviridae* ont de longues queues contractiles, les *Siphoviridae* ont de longues queues non-contractiles et les *Podoviridae* ont des petites queues. Des exemples de ces différentes architectures de queues et leur rôle dans l'infection seront discutés dans le chapitre 1.4.

La classification des virus est un exercice particulièrement complexe et qui impose des choix arbitraires du fait du faible échantillonnage de virus connus (cf. 1.2), de la caractéristique des virus à s'amplifier avec une forte burst size et à répliquer leur ADN avec un haut taux de mutations. Ainsi, des virus ayant 95%

de similarité ADN sont considérés comme 2 souches d'une même espèce et des espèces ayant 40-60% de similarité sont classées dans le même genre. Les données protéomiques et phylogénétiques sont utilisées pour construire les sous-familles [Kropinsky 16].

Figure 1.1 : répartition des espèces de virus par Ordre.

Graphique réalisé à partir des données de l'ICTV.

Non assignés
85 familles
2526 espèces

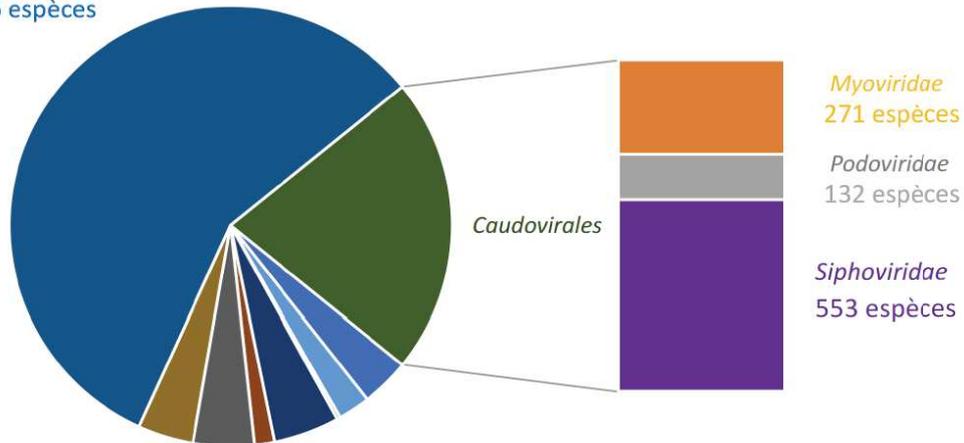


Figure 1.2 : Tableau des 11 familles de bactériophages connues.

Tableau réalisé à partir des données de l'ICTV.

	<u>Genome</u>	<u>Enveloppé</u>	<u>Structure</u>	<u>Nb d'espèces</u>
<u>Podoviridae</u>	ADNdb linéaire	non	Virion avec petite queue non-contractile	132
<u>Myoviridae</u>	ADNdb linéaire	non	Virion avec queue contractile	271
<u>Siphoviridae</u>	ADNdb linéaire	non	Virion avec queue longue non-contractile	553
<u>Sphaerolipoviridae</u>	ADNdb	non	Capside icosaédrique	7
<u>Leviviridae</u>	ARNsb linéaire	non	Capside sphérique	4
<u>Tectiviridae</u>	ADNdb linéaire	non	Capside icosaédrique décorée de fibres autour d'une vésicule membranaire	5
<u>Plasmaviridae</u>	ADNdb circulaire	oui	Enveloppe membranaire sphérique ou pléomorphique	1
<u>Corticoviridae</u>	ADNdb circulaire	oui	Capside icosaédrique décorée de fibres autour d'une vésicule membranaire	1
<u>Inoviridae</u>	ssDNA circulaire	non	Filamenteux	31
<u>Microviridae</u>	ssDNA circulaire	non	Capside protéique sphérique	21
<u>Cystoviridae</u>	ARNdb linéaire	oui	Double capsides de symétries différentes et enveloppées	1

1.1.3 Cycle de reproduction viral

Le cycle de reproduction phagique est initié par l'entrée virale, c'est-à-dire l'adsorption du virion à la surface de l'hôte et l'introduction de matériel viral dans la cellule. Dans le cas des *Caudovirales*, la queue du virion assure cette fonction en permettant l'éjection de l'ADN dans le cytoplasme de l'hôte.

Traditionnellement, on oppose deux cas de figures :

- Le cycle lytique au cours duquel les gènes/opérons de l'ADN viral s'expriment séquentiellement (early, middle et late phase) pour neutraliser l'hôte, détourner ses machineries de synthèse pour la production de nouveaux virions et finalement provoquer la lyse de la cellule infectée.
- Le cycle lysogénique où le matériel viral est intégré dans le génome de l'hôte ou prend une forme de plasmide, on parle dans les deux cas de prophage. Pendant le cycle lysogénique, les gènes de production et de lyse ne sont pas exprimés avant l'induction, c'est-à-dire l'entrée dans un cycle lytique en réponse à un signal (facteur de l'environnement, état métabolique de la cellule...).

On peut ainsi distinguer les phages lytiques qui provoquent rapidement la lyse de leur hôte et les phages lysogènes qui peuvent persister sur plusieurs générations d'une souche infectée. Cette dichotomie est justifiée par le fait que la grande majorité des phages appartiennent à l'ordre des *Caudovirales* qui sont strictement lytiques mais est inadaptée à représenter la diversité des mécanismes connus et laisse planer beaucoup d'ambiguïtés sur certains termes (phages virulents, phages tempérés...). Hobbs & Abedon proposent en 2016 une autre classification ainsi que des définitions précises[Hobbs 16].

Les cycles de reproduction viraux sont classés en 4 catégories en fonction de la nature du cycle de production (lytique ou chronique) et de la capacité à entrer dans un cycle lysogénique.

- Type I : lytique sans lysogénie. Exemples : T4, T5.
- Type II : chronique sans lysogénie. Exemple : M13.
- Type III : lytique avec lysogénie. Exemple : Lambda.
- Type IV : chronique avec lysogénie. Exemple : CTX.

Pour les phages de type I, on parle alors de phages strictement/obligatoirement lytiques. Les phages de type III ou IV sont dits tempérés.

Le terme virulent a été utilisé comme synonyme de strictement lytique mais Hobbs & Abedon proposent de réserver ce terme pour référer aux mutants virulents, c'est-à-dire des phages tempérés ayant perdu l'aptitude à pratiquer la lysogénie. *A contrario*, un phage strictement lytique ne portant aucune parenté avec une souche lysogénique et dépourvu des modules génétiques portant les gènes nécessaires à la lysogénie est décrit comme professionnellement lytique. Ce point de sémantique peut paraître anodin mais prend tout son sens si l'on considère une

application en phagothérapie où l'isolation de souches professionnellement lytiques est primordiale pour être assuré du comportement du phage.

1.2 Place des bactériophages dans la biosphère

1.2.1 *Terra incognita* microbiologique et virale

D'après le catalog of life [CoL], 9982 espèces de bactéries et 377 espèces d'archées sont connues, des nombres surprenamment bas surtout comparés aux nombres d'animaux et de plantes répertoriés (respectivement 1444228 et 357135 espèces) qui sont des organismes plus complexes, plus récents dans l'histoire du monde vivant et ayant des temps de génération beaucoup plus importants. Cette intuition sur un vaste inconnu microbiologique a été confirmée en 2016 quand le nombre d'espèces microbiennes a été estimé à 10^{11} - 10^{12} espèces de microbes, en majorité bactériennes [Lennon 16]. Ce chiffre théorique est étayé par des études métagénomiques des protéines ribosomales. L'utilisation des protéines ribosomales (qui tendent à être synténiques et co-localisées) plutôt qu'un autre marqueur génétique ou de la seule sous-unité 16S permet d'éviter l'élimination de séquences divergentes. Cette approche révèle un nombre important de groupes phylogénétiques majeurs sans représentants isolés et montre que le domaine bactérien est le plus important et le plus diversifié des règnes du vivant [Hug 16].

L'ICTV répertorie seulement 4404 espèces de virus, tous type d'hôtes confondus, et d'autres sources classent plus de 5500 bactériophages isolés de 150 hôtes différents [Ackermann 09]. Seulement 2200 génomes de virus à ADN double brin étaient déposés sur NCBI contre 45000 génomes bactériens en 2016 [Paez-Espino 16]. Des virus ont été isolés dans tous les biomes où des micro-organismes vivants se trouvent : aquatique (marin et eau douce), terrestre, sources géothermiques et dans différents organismes (humains, plantes, animaux). Le ratio de virus par procaryotes généralement évalué à 10, mais peut aller jusqu'à 1000 dans les sols secs [Suttle 07, Srinivasiah 08, Wommack 15]. Des virus ont récemment été isolés de sources géothermiques dans le plancher océanique (500 m sous les sédiments, sous 300 m d'eau) avec une forte abondance (2000 à 20000 virus par mL) et une présence élevée de virus d'archées (environ 50% de phages et 50% de virus d'archées) [Nigro 17]. L'abondance des virus dans les milieux aquatiques fluctue en fonction de cycles saisonniers, de la latitude, de la profondeur, du taux d'oxygène et de la concentration microbienne et est estimée à 10^7 virus par mL dans les couches supérieures de l'océan [Parsons 12, Brum 16].

La biodiversité du monde viral est donc difficile à évaluer et la morphologie caudée pourrait ne pas être la plus répandue chez les phages comme suggéré par des études de métaprotéomique et de microscopie électronique quantitative

[Brum 13, Brum 16]. Des études de métagénomique semblent indiquer une grande diversité des génomes mais une prédominance de certains (1-3% des génomes) ayant une abondance plus importante [Wommack 15]. Ceci est illustré par le fait que certains virus ont été isolés plusieurs fois en des lieux géographiquement éloignés comme PRD1 ou CrAssphage [Dutilh 14], et que de nombreux génomes de virus ont été retrouvés dans de nombreux échantillons marins et, dans une moindre mesure, dans des échantillons d'autres biomes [Paez-Espino 16]. Le développement de nouveaux outils de traitement informatique a permis de découvrir plus de 125000 génomes de virus dans des viromes dans des données métagénomiques de 3000 prélèvements différents et l'analyse des bibliothèques de spacers CRISPR a révélé près de 10000 associations hôtes-virus dont certaines pour des phylums procaryotes pour lesquelles aucuns virus n'est connu [Paez-Espino 16].

Les études génomiques restent complexes car 60-70% des gènes viraux n'ont pas d'homologues connus, il existe peu de gènes marqueurs génétiques (polA et RNR) et il est difficile d'évaluer leur fiabilité ; les lieux et la méthode de prélèvement ainsi que le traitement des échantillons peuvent biaiser les résultats. Le développement de nouvelles méthodes, comme la génomique de virus isolés (single virus genomics) qui combine cytométrie en flux et qPCR [Lance 16], permettront à terme d'évaluer et d'étendre les données obtenues en travaillant sur les bases de données métagénomiques.

1.2.2 Rôle de la virosphère dans l'environnement

Il est à la fois facile d'imaginer que la co-évolution phage-bactérie a un impact sur l'environnement et extrêmement complexe de l'évaluer. Des études sur la lyse bactérienne induite par les phages dans les océans donnent quelques éléments de réponse. Il a été estimé qu'entre 20 et 50% de la population de bactérioplankton disparaît du fait de la lyse phagique chaque jour, assurant un turn-over de ces populations mais aussi une mise à disposition de leurs composants dans l'environnement sous une forme particulièrement accessible, la lyse phagique des cellules étant plus « brutale » que l'apoptose. La remise à disposition du fer en particulier, pourrait assurer 90% des besoins du phytoplankton [Poorvin 04] et 35% du carbone benthique provient de ces lysats phagiques, illustrant le rôle des phages dans le cycle biogéochimique du carbone [Danovaro 08].

Les cyanobactéries, qui assurent 25% de la production d'oxygène par photosynthèse, sont également la cible de phages et 5 à 15% de leur population est lysée chaque jour [Suttle 94]. Les cyanophages portent dans leur génome les gènes du photosystème de leur hôte [Lindell 05] et 5% de la production globale d'oxygène pourrait être assurée par des cellules infectées [Wommack 15].

Ces travaux tendent à prouver que les virus ont un rôle important de prédation et de sélection des micro-organismes mais aussi dans le renouvellement écologique

des matières organiques.

1.2.3 Interaction avec d'autres systèmes biologiques

Comme cité précédemment, les phages ont colonisé tous les milieux, y compris les organismes supérieurs.

Le corps humain accueille des populations de virus spécifiques aux organes (bouche, intestin, nez, oreille, tractus génital) et certaines séquences semblent spécifiques de l'hôte [Paez-Espino 16]. Le CrAssphage, un phage du tractus digestif, semble au contraire, extrêmement commun. La colonisation par les phages peut avoir un impact sur la santé humaine car ils peuvent transporter des îlots de pathogénicité et ainsi rendre pathogène des bactéries commensales du microbiome humain, c'est par exemple le cas du choléra, où le gène de la toxine cholérique est propagé par le phage filamenteux CTX [Karaolis 99, Boyd 12].

Enfin, les interactions entre les phages et eucaryotes sont potentiellement plus proches et complexes que ce que la spécificité d'infection des phages laisse imaginer : le phage WO infecte une bactérie strictement intracellulaire *Wolbachia* et qui a intégré dans son génome un module d'adaptation à l'hôte (cellules d'insectes) de la bactérie.

1.3 Applications

Au-delà d'être des modèles de recherche fondamentale et des acteurs sous-étudiés de notre environnement, les phages peuvent être envisagés comme des outils biotechnologiques. L'utilisation des phages peut servir pour des développements méthodologiques en recherche mais aussi dans le domaine médical pour le diagnostic et/ou la thérapie de maladies infectieuses. Enfin, l'étude et l'usage des phages présentent également un intérêt dans les domaines industriels, en particulier l'agroalimentaire. Ce chapitre propose une présentation non-exhaustive d'exemples d'applications utilisant des bactériophages.

1.3.1 Biologie moléculaire & biotechnologies

Les phages sont depuis longtemps utilisés en biologie moléculaire, promoteurs et polymérase phagiques étant très commun dans des plasmides recombinants (promoteurs et ARN polymérase du phage T7 par exemple). Les phages peuvent aussi être directement utilisés pour réaliser des clonages d'ADN ou d'ADN complémentaire. Le phage Lambda permet par exemple des transformations de plus grand fragment d'ADN qu'avec un plasmide recombinant (jusqu'à 52 kpb contre 20 kpb dans un plasmide), des vecteurs cosmides sont construits avec les sites de

fin cohésives (sites cos) du génome de lambda et permettent l'encapsulation et la transduction d'un fragment d'ADN d'intérêt [Cronan 03].

Les phages peuvent également être utilisés pour explorer des interactions entre une protéine et une molécule cible. Le phage display a été inventé en 1985 [Smith 85] et consiste à préparer une librairie de phages dont les protéines de décorations sont fusionnées avec différentes protéines/peptides. Les phages sont ensuite produits et sélectionnés sur la base de leur adhésion à la molécule cible fixée sur une surface. Les phages sont élués puis amplifiés. Après plusieurs cycles, les phages sont séquencés, permettant de caractériser les molécules ayant la meilleure affinité. Cette technique a notamment permis l'identification d'anticorps monoclonaux d'intérêt thérapeutique [Crisuolo 17].

Les virions de phages peuvent également servir de base à des structures de présentation ou à des cargos de molécules actives. Des virus ou des particules virales peuvent être conçus pour permettre de délivrer des molécules actives, des molécules de marquage pour l'imagerie médicale ou des enzymes qui sont présentés à haute valence. Ces molécules sont protégées du système immunitaire et peuvent être attachées à la particule virale conjointement avec d'autres facteurs permettant par exemple de cibler certains tissus ou de faciliter la libération endosomale [Henry 12]. De plus, les propriétés géométriques précises des capsides et queues de phage en font des objets d'intérêt pour la biologie structurale et les sciences des matériaux [Molek 15, Patel 17].

1.3.2 Médecine

1.3.2.1 Phagothérapie

Depuis le début des années 2000 on observe un renouveau de cette pratique qui ne sera pas reconnue par les autorités sanitaires avant 2010 (la phagothérapie était considérée comme une médecine « alternative », au même titre que l'application de sangsues ou la phytothérapie). Cette approche apporte une alternative antibactérienne pour de multiples affections : infections pulmonaires, gastro-intestinales et osseuses, plaies infectées, acné...

Aujourd'hui, la phagothérapie est pratiquée de manière « traditionnelle » à l'institut Eliava (T'bilisi, Géorgie), à titre compassionnel par quelques médecins (aux Etats-Unis, en France, Allemagne, Belgique et Pologne) et plusieurs entreprises développent des phages ou des cocktails de phages ciblant certaines pathologies. Ces deux approches se distinguent du fait que la manière « traditionnelle » consiste à isoler et amplifier un phage (souvent issue d'une « phagothèque ») spécifique du pathogène du patient alors que l'approche industrielle utilise des phages ou des cocktails de phages à spectre large, ciblant plusieurs souches pathogènes d'une espèce bactérienne donnée.

Faisant partie de ce second type de stratégie, le projet PhagoBurn, initié en 2013, est le premier du genre à atteindre le stade d'essais cliniques. Des cocktails de phages ciblant des souches pathogènes d'*Escherichia coli* et de *Pseudomonas aeruginosa* sont utilisés pour traiter les infections chez les grands brûlés. L'administration se fait par l'imprégnation avec la solution de phages des pansements hydrocolloïdes couvrant la brûlure infectée. La société française Pherecydes Pharma produit ces phages et développe également des cocktails pour les infections respiratoires à *P. aeruginosa* et les infections articulaires et osseuses à *Staphylococcus aureus*. Plusieurs autres entreprises de biotechnologie développent des solutions thérapeutiques à base de phages qui sont en phase pré-cliniques ou dans les premières phases d'essai. On peut citer AmpliPhi (phages de *S. aureus* et *P. aeruginosa*), Enbiotix (phages prévenant la formation de biofilms sur les prothèses), Sarum Biosciences (traitement des infections à *Clostridium difficile*) et TechnoPhage (traitement d'infections respiratoire et gastro-intestinales).

Des mesures prophylactiques peuvent également être envisagées en utilisant des phages codant un système CRISPR-Cas9 pour créer une pression de sélection favorisant les bactéries non résistantes aux antibiotiques (en milieu hospitalier) [Bikard 14, Goren 17].

1.3.2.2 Evaluation de l'efficacité et de l'innocuité de la phagothérapie

L'une des principales inquiétudes quant à l'utilisation des phages est leur interaction avec le système immunitaire. Bien qu'incapables d'infecter les cellules eucaryotes, ils provoquent probablement une réponse des systèmes immunitaires inné et adaptatif. Les effets d'inactivation des phages et de réaction immunitaire potentiellement nocive doivent être plus explorés *in vivo*, mais les observations *in vitro* et sur modèle murin sont encourageantes. La réponse immunitaire innée semble largement impliquée dans l'inactivation des phages, impliquant la nécessité de mieux caractériser ces effets *in vivo* et d'optimiser les doses et conditions d'administrations pour les thérapies à base de phages [Hodyra-Stefaniak 15]. L'utilisation de lysat plutôt que de phage purifié a été suggérée pour cumuler l'effet lytique des phages avec une stimulation du système immunitaire contre la bactérie ciblée [Górski 12].

Un autre point d'inquiétude est la libération de facteurs pro-inflammatoires lors de la lyse bactérienne, des travaux récents montrent que la lyse induite par les phages est moins nocive à cet égard que celle provoquée par des antibiotiques de la classe des bêta-lactames [Dufour 17].

La disparition des phages administrés dans l'organisme s'observe de façon concomitante à la disparition de leur bactérie cible [Hodyra-Stefaniak 15].

Bien qu'il soit nécessaire d'acquérir plus de données *in vivo* pour comprendre et évaluer rigoureusement les effets de l'utilisation de phages chez l'Homme, les études

à disposition semblent montrer que la phagothérapie est sûre et efficace, y compris chez des patients immunodéprimés [Borysowski 08]. Les effets d'inactivations doivent également être mieux décrits et pourraient être problématique pour des applications de thérapies systémiques.

Enfin, le développement de résistances au phage par la bactérie ne représente pas nécessairement une importante limitation, le fait que le stress apporté par le phage soit très spécifique diminue la compétitivité de la bactérie cible face au reste de la flore et augmente sa susceptibilité au système immunitaire, pouvant permettre une réussite de la thérapie malgré le développement de résistance. De plus, ces résistances peuvent être contournées par l'utilisation de cocktails de phages ou par une phagothérapie administrée en complément de l'antibiothérapie. Des phages lytiques se fixant à des protéines de la membrane externe impliquées dans la résistance aux antibiotiques peuvent être utilisés, permettant :

- de cibler spécifiquement les souches résistantes aux antibiotiques
- de repotentialiser les antibiotiques (l'acquisition de la résistance au phage favorisant la perte de la résistance à l'antibiotique).

Ce dernier type d'approche a été utilisé avec succès chez l'Homme [Chan 16].

1.3.2.3 Traitements dérivés des phages

En fin du cycle d'infection, les bactériophages lytiques produisent des endolysines des enzymes dégradant l'enveloppe bactérienne, permettant la libération des virions. Ces enzymes peuvent être produites et utilisées comme agent bactérien contre les bactéries gram-positifs [Borysowski 06]. Une endolysine a été modifiée pour être efficace contre les bactéries gram négatif [Briers 14].

1.3.3 Agroalimentaire

1.3.3.1 Enjeux de l'étude des phages pour l'industrie agroalimentaire

En agroalimentaire, la microbiologie est une préoccupation majeure pour différentes raisons :

- La flore technologique est nécessaire à certains processus de production (produits laitiers, vin, bière. . .).
- Certains microbes peuvent avoir un effet négatif sur l'aspect ou les qualités organoleptiques du produit (flore d'altération).
- Des organismes pathogènes peuvent contaminer les aliments et les consommateurs.

L'étude des phages a beaucoup à apporter à ce milieu que ce soit pour (i) le contrôle ou la détection des flores pathogènes et d'altération (ii) le design de mécanismes de défense pour la flore technologique en milieu industriel (iii) la mise au point

de protocoles pour utiliser les phages naturellement présents dans l'environnement dans le cadre d'une production « Bio ».

Les toxi-infections alimentaires sont responsables de centaines de milliers d'hospitalisations chaque année et évoluent de façon préoccupante du fait de l'émergence des résistances aux antibiotiques ainsi que de l'évolution de la consommation vers des produits prêt-à-manger. Pour ces produits, le contrôle de l'environnement de production est plus complexe et le consommateur ne cuira pas le produit, augmentant le risque de contamination [Velusamy 10].

De plus, les antibiotiques sont très utilisés dans le domaine agroalimentaire. En Europe, 7982 tonnes d'antibiotiques ont été consommées dans l'agroalimentaire contre 3399 tonnes pour la santé humaine [ECDC/EFSA/EMA 15]. Cette tendance est moins marquée en France (761 tonnes d'antibiotiques utilisés dans l'agroalimentaire pour 719 tonnes dans la santé humaine). Les données pour le reste du monde restent vagues et très variables en fonction du pays et du type d'élevage. Cette utilisation importante des antibiotiques dans l'agroalimentaire en fait une source de préoccupation pour la dispersion des molécules actives et l'apparition de souches bactériennes résistantes [ANSES 14, Organisation 15] mais aussi un important marché pour des technologies antimicrobiennes alternatives.

1.3.3.2 Phagothérapie animale et végétale

Dans une démarche similaire à celle de la phagothérapie appliquée à l'homme, les phages peuvent être utilisés pour combattre les infections microbiennes survenant chez les animaux et les plantes. L'utilisation de phages peut être envisagée au cours de la production et/ou pendant le processus de préparation/empaquetage d'un produit.

Des essais ont été réalisés sur des pathogènes des genres *Salmonella*, *Escherichia*, *Campylobacter* et *Listeria* dans des élevages bovins, ovins, porcins et de volailles [Goodridge 11] ainsi qu'en pisciculture, contre *Vibrio anguillarum* [Romero 14]. Les résultats observés démontrent un potentiel antibactérien chez les animaux d'élevage mais la variabilité des résultats suggère peut-être une spécificité des comportements selon les souches et surtout une nécessité d'optimiser ce genre de stratégies. La mise au point et l'harmonisation des protocoles de caractérisation des phages et de leur efficacité semblent nécessaires ainsi que l'étude des modes d'administration. Les entreprises VetoPhages (France) et ACD Pharma (Norvège) développent des solutions de phagothérapie à usages vétérinaires. Les bactériophages peuvent également être envisagés pour traiter les infections bactériennes des plantes, cette application imposant la contrainte supplémentaire d'utiliser des virus résistants à l'exposition aux UVs [Jones 12].

L'utilisation de phages pour les étapes de traitement de la nourriture est juridiquement moins contraignante et plusieurs produits à base de phages sont déjà

commercialisés. Le LISTEXTM, un cocktail de phages anti-Listeria, a été approuvé pour cet usage par la FDA (Food and Drug Administration) en 2006 et en Union Européenne en 2011. La société néerlandaise Microcos, qui a développé le LISTEXTM, commercialise également un cocktail anti-Salmonella et la société Intralytix propose des produits similaires contre *Listeria monocytogenes*, *Shigella spp.*, *Escherichia coli* 0157 :H7 et certains sérotypes de *Salmonella*. La société Fixed Phage développe des emballages sur lesquels des phages ont été immobilisés et la société APS Biocontrol Ltd développe un cocktail de phages dirigé contre la flore d'altération de légumes.

1.3.4 Impact environnemental de l'utilisation des phages

L'utilisation des phages pour les différentes applications évoquées jusqu'à présent pose la légitime question de leurs impacts sur les communautés microbiennes environnementales, particulièrement dans le cadre d'une utilisation facilement dispersible (traitement en milieu agricole, pisciculture...). En effet, les phages ont un potentiel d'amplification et une capacité rapide de mutation qui leur donne le potentiel d'affecter l'équilibre d'une population procaryote (et donc de leur écosystème). L'utilisation de souches qui ne sont pas professionnellement lytiques (phages lysogéniques et mutants virulents) devrait donc être appliquée avec d'autant plus de prudence.

Cette problématique mériterait d'être étendue à d'autres domaines d'activités humaines : la diversité et l'abondance des populations microbiennes ne représentent-elles pas des indicateurs de choix d'impact sur l'environnement ? Ce type d'indicateur nécessite l'acquisition de plus de connaissances fondamentales sur les microbiomes et le développement d'outils permettant d'étudier ces populations microbiennes.

1.3.5 Détection

Le développement de nouvelles méthodes de détection des bactéries pathogènes pour le domaine médical ou industriel représente à la fois un enjeu de santé publique et économique. Les méthodes classiques de détection (cultures bactériennes) sont fiables mais représentent une lourde contrainte car elles requièrent un coût de main-d'œuvre important et sont lentes (48 à 72h en moyenne) impliquant le décalage d'un diagnostic, la prolongation du stockage de denrées périssables et/ou l'utilisation de traitements à très large spectre qui favorisent l'apparition de résistances.

Les bactériophages représentent un axe de développement intéressant du fait de leur spécificité pour leur hôte et du phénomène d'amplification inhérent au cycle de production. Différents types de méthodes développées ou en développement ont

donc le potentiel de répondre aux normes de limites de détection et de spécificité en vigueur.

1.3.5.1 Détection de la lyse

La lyse de la cellule bactérienne provoque la libération de molécules intracellulaires dont la détection peut être plus simple et plus rapide que celle de la cellule. L'ATP peut par exemple être identifiée par bioluminescence après ajout de luciférase. L'activité de la -D-glucosidase a également été explorée comme marqueur suite à une lyse phagique, le produit de la réaction induisant un changement de courant électrique mesurable avec un potentiostat [Schmelcher 14].

1.3.5.2 Détection de l'amplification des phages

Ce type de méthode consiste à substituer la détection du procaryote d'intérêt par celle d'un de ces virus lytiques par le biais d'une molécule ou protéine produite lors de l'amplification du phage. Cela permet d'accélérer la détection d'échantillons contaminés par de faibles taux de bactéries nécessitant une phase d'enrichissement pour être mesurable par d'autres méthodes. Dans le cas de pathogènes ayant une croissance très lente, comme les mycobactériums à croissance lente (8 à 18 semaines de croissance), qui sont notamment des pathogènes animaux préoccupant pour l'élevage bovin (principale agent responsable de la tuberculose bovine), l'utilisation de phages permet d'accélérer considérablement le protocole de détection [Swift 16].

1.3.5.3 Détection par phage rapporteur génétiquement modifié

L'utilisation de phages génétiquement modifiés ouvre la possibilité d'utiliser d'autres méthodes de détection. Il est par exemple possible d'intégrer un gène de protéine pour avoir un marqueur produit dans les bactéries infectées (protéine fluorescente par exemple). Ce principe a été testé pour la détection de *Salmonella enterica* [Vinay 15] avec une détection de la fluorescence par cytométrie en flux. Ce type de technologie ouvre des perspectives intéressantes avec l'utilisation de phages spécifiques de différents hôtes et produisant des protéines fluorescentes distinctes. L'utilisation de ce type de phage rapporteur est étudiée pour le développement de pansements et de revêtements de matériels médicaux « intelligents » [Milo 16].

1.3.5.4 Détection par adsorption de phages ou protéines de phage

Des phages ou des fibres de phages (structures portant les récepteurs de phages, cf. 1.4) peuvent être adsorbés sur des surfaces ou des billes para-magnétiques, ouvrant d'autres perspectives de détections. L'utilisation de billes magnétiques permet de séparer spécifiquement un pathogène d'un milieu pour en accélérer la

phase d'enrichissement [Denyes 17]. L'adsorption de phages ou de récepteurs de phages sur une surface permet l'utilisation et le développement d'autres méthodes de détection comme la microscopie, des dosages d'absorption par enzymes liées ou la conception de puces pour la SPR [Denyes 17, Schmidt 16, Wang 16].

1.4 Perforation de la paroi bactérienne et éjection de l'ADN phagique

L'entrée virale est un phénomène complexe qui peut être décomposé en différents événements :

- Fixation réversible du virion à l'hôte
- Fixation irréversible du virion à l'hôte
- Perforation de la/des membrane(s)
- Ouverture de la capsid
- Ejection de l'ADN viral de la capsid au cytoplasme de l'hôte

Ces étapes ne se déroulent pas de la même façon d'un phage à l'autre et peuvent faire intervenir des dépendances à certaines activités enzymatiques ou conditions physico-chimiques (présence d'ions particuliers, gradient de protons de la membrane plasmique de l'hôte...). L'éjection de l'ADN est un événement dont la nature de la force motrice n'a toujours pas été élucidée et qui a fait l'objet de beaucoup d'études (voir [Molineux 13] pour revue).

Cette thèse se concentre sur les événements de perforation de la membrane et de la transmission d'un signal d'ouverture de la capsid qui précède l'éjection de l'ADN. Ci-après seront décrits les mécanismes d'entrée virale de différents phages appartenant aux trois familles de l'ordre des *Caudovirales*. Des informations seront données sur l'ensemble du mécanisme (dans les limites de ce qui est connu) mais un intérêt particulier sera porté sur les données structurales des éléments impliqués dans la perforation et la transmission de cette information à la capsid.

1.4.1 Les *Podoviridae*

Les *Podoviridae* sont caractérisés par une petite queue non-tractile. Ils ont souvent des protéines à l'intérieur de la capsid qui sont éjectées en même temps que l'ADN lors de l'infection.

1.4.1.1 T7

T7 est un phage lytique de *E. coli* de la famille des *Podoviridae*, il a une architecture proche du phage T3 et est connue pour infecter beaucoup de souches différentes (y compris des souches pathogènes (*E. coli* O157 ; H7) et encapsulées).

Architecture de la queue

La queue du bactériophage T7 est insérée à l'un des vertex de sa capsidie icosaédrique de 600 Å de diamètre ($T=7$). Ce complexe protéique émerge de 185 Å à l'extérieur de la capsidie et est connecté à un noyau interne de 265 Å x 175 Å à l'intérieur de la capsidie (voir ci-dessous).

La queue s'assemble à partir du dodécamère de gp8 formant le portail à un des vertex de la capsidie. Un anneau dodécamérique de gp11 (adaptateur - gatekeeper) se forme au contact de gp8 et sert de base à un hexamère de gp12 formant la pointe de la queue (nozzle). L'interface entre gp11 et gp12 sert de point d'attache pour les 6 fibres (trimères de gp17) [Cuervo 13]. Les fibres sont mobiles mais lorsque le virion est libre elles montrent une tendance à se replier contre la capsidie de façon ordonnée [Hu 13]. La cavité au centre de la queue est a priori remplie par la protéine gp7.3, qui a un rôle dans l'adhésion du virion [Kemp 05].

Le noyau interne forme une structure cylindrique de 3 couches concentriques. Un dodécamère de gp14 est attaché à la face interne de gp8, il est suivi par un octamère de gp15 et, le plus au centre de la capsidie, d'un tétramère de gp16. On note une rupture de symétrie entre qui a fait l'objet d'un développement dans le traitement des données de microscopie électronique (Focused Asymmetric Reconstruction – reconstruction indépendante des différentes parties du complexe n'ayant pas la même symétrie) et qui pourrait avoir une importance fonctionnelle [Guo 13]. La cavité du noyau interne est remplie dans sa partie la plus interne avec de l'ADN et par la protéine nucleocapsidique gp6.7 (head completion protein) proche du vertex [Kemp 05, Cheng 14].

Changement de conformation de la queue pendant l'infection

Lors de l'interaction avec l'hôte, toutes les fibres sont détachées de la capsidie et interagissent avec la membrane. La fixation des fibres à la paroi de la bactérie répercute un changement de conformation au niveau de la jonction pb11-pb12 et l'hexamère de pb12 change également de conformation, il est moins « vrillé » et son extrémité s'ouvre [González-García 15b]. Les protéines du noyau interne (gp14, gp15 et gp16), ainsi que les protéines gp6.7 et 7.3 [Kemp 05], sont éjectées du virion et se réassemblent pour former une structure tubulaire de 8 nm de diamètre et de 45 nm de long à travers les membranes et le périplasme, ainsi qu'une structure discoïdale de 300 Å de diamètre dans le cytoplasme de l'hôte [Hu 13] (Fig 1.3-a-c).

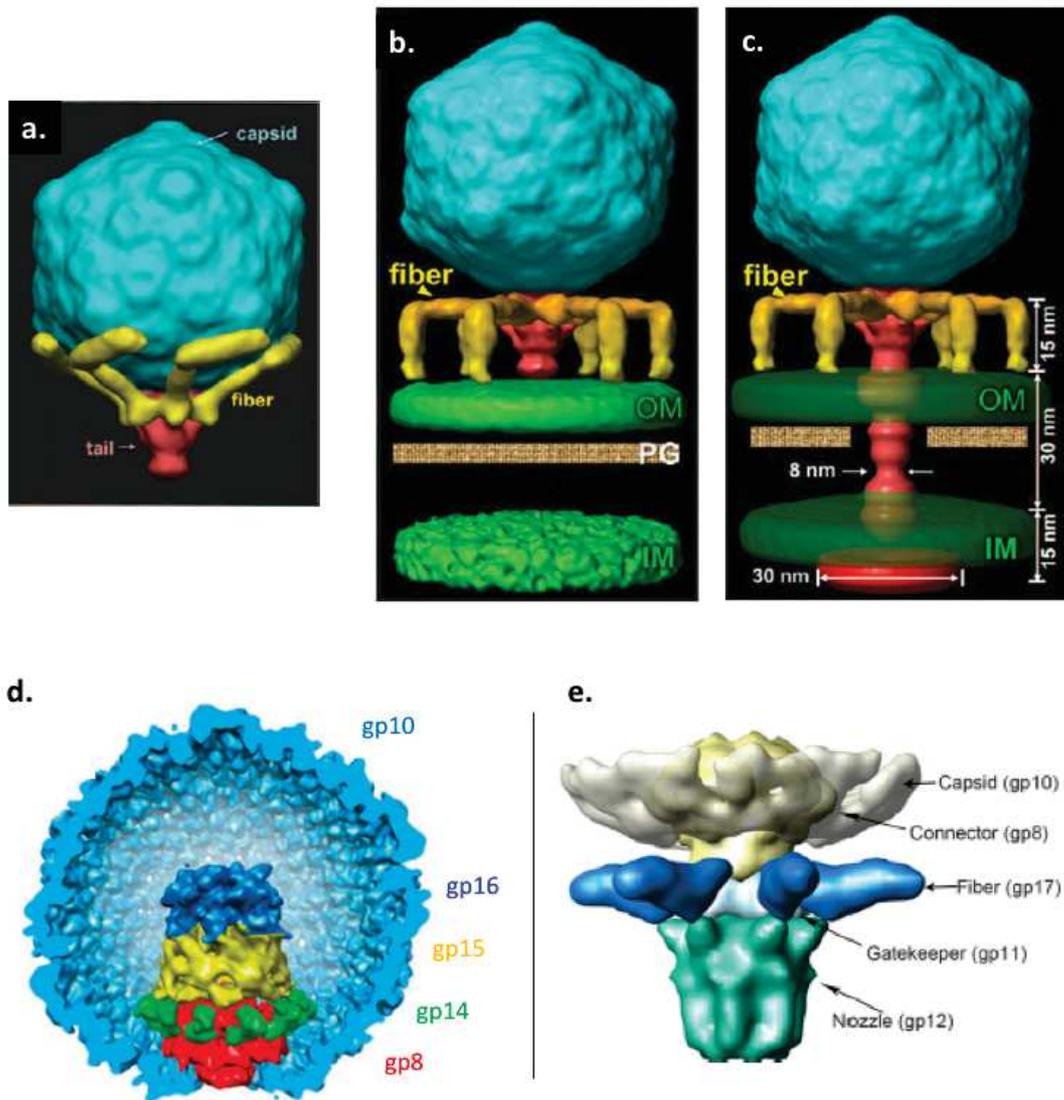


Figure 1.3 : Structures des phages T7.

a.-c. Changements de conformations du virion de T7 lors de l'infection. Lorsque le virion est libre, les fibres sont le plus souvent repliées contre la capsid (a.). Les protéines à l'intérieur de la capsid (noyau interne) sont éjectées pour former un canal transmembranaire pour l'ADN.

Figure adaptée de [Hu, 13].

d. Organisation du noyau interne de T7. Les protéines gp 16, gp15 et gp14 sont éjectées lors de l'interaction de l'hôte.

Figure adaptée de [Guo, 13].

e. Organisation de la queue de T7 après interaction avec du LPS. La partie proximale des fibres indique qu'elles ne sont plus repliées contre la capsid. La protéine gp12 subit également de notables changements de conformation.

Figure issue de [González-García, 13].

Aspects fonctionnels

Le phage T7 interagit de façon réversible avec le LPS-R (rough) puis de façon irréversible avec ce récepteur ou un autre, inconnu. *In vitro*, l'interaction avec du LPS-R est suffisante pour provoquer l'éjection d'ADN [González-García 15b]. Il a été proposé que T7 utilise les porines OmpA et OmpF comme récepteurs secondaires [González-García 15a], mais les expériences soutenant cette hypothèse sont peu convaincantes.

Lors de l'interaction avec la membrane de son hôte, les changements conformationnels menant à la formation d'une structure tubulaire à travers la paroi bactérienne se font sans que le système ne requière un apport d'énergie extérieure. Cependant, la translocation de l'ADN dans le cytoplasme est dépendante du gradient de proton, il a été proposé que la partie transpériplasmique du complexe se comporte comme un transporteur dépendant du gradient de proton et transporte activement l'ADN viral dans l'hôte [Kemp 04].

Le complexe gp8-11-12 peut être séparé du virion ou reconstitué en expression des gènes clonés [Cuervo 13]. Ce complexe peut être complété avec des fibres de gp17 purifiées indépendamment et ce complexe change de conformation en présence de LPS-R [Cuervo 13, González-García 15b].

Le complexe gp15-16 peut être reconstitué *in vitro* : gp15, dimère soluble, mélangé avec gp16, monomère soluble, forme un complexe à la stoechiométrie semblable au complexe dans le virion (4 gp16 pour 8 gp15). Ce complexe a des propriétés d'interaction avec les membranes et de faible interaction avec l'ADN. De plus la protéine gp16 porte une activité de dégradation hydrolytique du peptidoglycane [Lupo 15, Leptihn 16].

Il est important de noter que la localisation, la structure et le rôle de plusieurs protéines est encore mal compris, notamment les protéines gp6.7 et gp7.3 dont la localisation dans le virion et l'éjection dans l'hôte a été montrée [Kemp 05] sans qu'elles apparaissent sur les structures [Cuervo 13, Hu 13]. La protéine gp13, bien qu'absente du virion, semble également avoir un rôle important [Kemp 05].

Modèles proposés

Interaction avec la membrane et déclenchement de l'infection

Le modèle « classique » propose que les fibres positionnent le virion à la surface de l'hôte en interagissant réversiblement avec le LPS et permettent à un autre élément du virion (inconnu, *a priori* une autre zone des fibres ou l'extrémité de gp12) de faire une interaction irréversible (*a priori*, avec les porines de la membrane externe ; voir paragraphe précédent) qui déclenche les changements de conformations menant à la perforation de la paroi bactérienne [González-García 15b, González-García 15a].

Un autre modèle s'appuie sur le fait que les fibres de T7 sont en équilibre entre une conformation « repliée » contre la capsidie et un état « étendu ». Lorsque le virion est libre en solution, les fibres sont la plupart du temps en conformation repliée même si elles sont retrouvées en faible proportion en conformation étendue. Lors de l'interaction avec la membrane, l'interaction faible avec le LPS déplace l'équilibre conformationnel vers la forme étendue [Hu 13]. Lorsque la majorité ou toutes les fibres sont étendues elles imposent un changement de conformation au niveau de la queue [González-García 15b] qui déclenche l'éjection.

Changements de conformations et structures impliquées dans la perforation de la paroi bactérienne

La protéine gp14 est retrouvée dans les membranes externes d'hôtes infectés et semble donc former la partie transmembranaire de la membrane externe du complexe. Les protéines gp15 et gp16 forment un complexe transpériplasmique qui dégrade le peptidoglycane et interagit avec la membrane interne. De plus, il semble que ce complexe « escorte » l'extrémité de l'ADN de l'intérieur de la capsidie jusqu'à l'intérieur de l'hôte. Ces 3 protéines, qui forment le noyau interne du virion libre, ne peuvent pas passer dans une conformation entièrement repliée par le tube de la queue, elles doivent nécessairement passer par un état partiellement déplié. De plus, la conformation que ces protéines adoptent une fois éjectées est différente de celle qu'elles ont au sein du noyau interne. Ces trois protéines sont donc maintenues dans une structure « métastable » au sein du virion, se dénaturent partiellement lors de l'éjection et se restructurent en éléments tubulaires, traversant la paroi bactérienne [Lupo 15, Leptihn 16]. Intuitivement, il semble logique que gp14 soit éjectée et se stabilise dans la membrane externe avant le passage de gp15 et gp16, mais la séquence précise des événements et le rôle des autres protéines éjectées reste à déterminer.

1.4.1.2 P22

P22 est un phage tempéré infectant *Salmonella enterica*. Bien qu'ayant une morphologie typique des *Podoviridae*, P22 a souvent été associé au siphovirus Lambda, cette parenté semble cependant avoir été surestimée [Ackermann 15].

Architecture de la queue

La queue du bactériophage P22 protrude de 320 Å au niveau d'un des vertex de la capsidie icosaédrique de 650 Å de diamètre (T=7). Ce complexe d'environ 2,8 MDa est, à l'instar de T7, connecté à un coeur protéique dans la capsidie (250 Å).

Un premier modèle, réalisé à partir de la structure à résolution intermédiaire du complexe de la queue de P22 et de l'ajustement de structures cristallographiques,

fut proposé en 2009 [Lander 09] mais il a par la suite été beaucoup discuté et modifié (cf. autres références de ce paragraphe). Un dodécamère de gp1 forme le portail au niveau d'un vertex de la capsid. Il est lié, vers l'extérieur, à un dodécamère de gp4 (protéine hub) qui sert de base à un hexamère de gp10, 6 trimères de gp9 en périphérie (formant 6 tailspikes) et un trimère central de gp26 (tail needle) [Bhardwaj 14, Pintilie 16, Wu 16] (Fig. 1.3-d). La protéine gp26 est une protéine en 3 domaines, le domaine N-terminal est exceptionnellement conservé chez les phages proches de P22 (194 phages) et verrouille la capsid (plug domain), le domaine central forme un coiled-coil (11 heptades répétés), et le C-terminal forme un domaine dont la fonction précise, non déterminée, est très variable. Remarque : le domaine N-terminal de gp26 a été utilisé pour classer les p22likevirus et apporter des informations sur l'histoire évolutive de ce genre [Bhardwaj 16, Leavitt 13].

L'intérieur de la capsid est occupé, en plus de l'ADN viral, par la protéine gp1 qui, après son domaine « portail » forme un long tonneau en hélice (15 nm). Il a été proposé, sur la base de travaux de microscopie électronique et d'analyse de séquence, que ce tonneau pourrait être partiellement désordonnée [Wu 16] utilisant en plus les structures de [Lander 09, Chang 06], mais la structure de [Pintilie 16] ainsi que la structure cristallographique de gp1 sont en contradiction avec cette hypothèse. Le rôle de ce tonneau dans l'empaquetage et l'éjection de l'ADN n'est pas connu. Les protéines gp16, gp20 et gp7 (protéines pilotes ou protéines E ou encore protéines éjectées) se localisent également dans la capsid et semblent associées au tonneau en hélices de gp1.

La protéine portail, gp1, sert d'abord d'attache à la machinerie d'empaquetage de l'ADN et une fois la capsid pleine (signal headfull), de subtils changements conformationnels permettent le détachement de la terminase et la fixation de gp4, initiant l'assemblage de la queue [Andres 10].

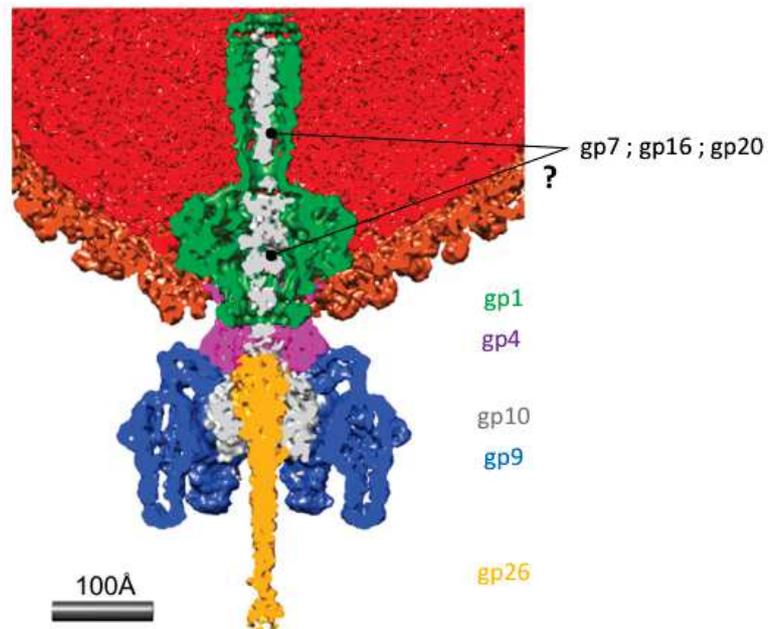


Figure 1.4 : Structures des phages P22.

Structure de la queue du phage P22. L'activité enzymatique portée par les trimères de gp9 permet de rapprocher gp26 de la membrane de l'hôte. La position des protéines gp7, gp16 et gp20 n'a pas encore été établie avec certitude et aucune structure après interaction avec l'hôte n'a encore été publiée.

Figure adaptée de [Pintilie, 16].

Interaction avec l'hôte

Le phage P22 reconnaît spécifiquement les polysaccharides du LPS des *Salmonella* par le biais de ses trimères de gp9 (tailspikes). Les trimères de gp9 portent une activité endorhamnosidase qui clive l'antigène O jusqu'à ce qu'il ne reste que 5 unités répétées. Cette activité est nécessaire à l'infection [Andres 10]. L'existence d'un récepteur secondaire a longtemps été suspectée et des phages de la même famille (sf6) ont besoin d'une interaction supplémentaire entre le domaine N-terminal de gp26 et les protéines bactériennes OmpA/C [Parent 14], mais il semble que le LPS soit le seul récepteur de P22 [Leavitt 13].

Il existe peu d'information structurale sur les changements conformationnels de la queue du virion pendant l'infection. La protéine gp26 est détachée et le segment N-terminal, qui a une conformation étendue et peu structurée dans la particule virale libre, s'organise en épingle à cheveux trimérique après éjection [Bhardwaj 16].

Aspects fonctionnels

Il est intéressant de noter que sans la protéine gp7, l'incorporation dans le virion de gp16 est réduite et celle de gp20 ne se fait plus. Par ailleurs, l'incorporation dans la capsid des protéines E n'est pas dépendante de gp1.

Remarque : des chimères de P22 où le domaine C-terminale de gp26 a été échangé avec celui de la needle d'un autre phage restent infectieuses mais la vitesse d'éjection de l'ADN change [Leavitt 13].

Modèle proposé

Les trimères de gp9 fixent et clivent les polysaccharides en surfaces jusqu'à une hauteur déterminée (5 unités répétées du core). Dans cette configuration, le virion est positionné par rapport à la membrane de l'hôte et il a été proposé que le trimère de gp26 (tail needle), protrudant nettement plus loin de la capsid que les trimères de gp9, déclenche (mécaniquement ?) les changements de conformation menant à sa propre éjection, celle des protéines pilotes et l'acheminement de l'ADN jusqu'au cytoplasme [Andres 10]. Les détails structuraux post-éjection et les rôles précis de gp26 et des protéines pilotes dans l'enveloppe de l'hôte n'ont pas été décrits.

1.4.2 Les *Myoviridae*

Les *Myoviridae* sont caractérisés par une longue queue contractile qui permet la perforation de la membrane externe de l'hôte par une action mécanique comparable à celle d'une seringue.

1.4.2.1 T4

T4 est l'un des virus les plus étudiés, et le principal représentant de la famille des *Myoviridae*. Il est structurellement identique aux phages T2 et T6 et ils sont souvent rassemblés sous l'appellation « T-pairs » (T-even). De plus, la queue du phage T4 a une structure et un mécanisme qui est très proches de systèmes de sécrétion bactériens. T4 infecte *E. coli*.

Architecture de la queue

La queue du phage T4 est composée d'un tube rigide de 925 Å avec une structure en forme de dôme de 270 Å de haut pour 520 Å de diamètre à son extrémité distale (la baseplate) et des structures en anneaux à l'extrémité distale qui fait le lien avec la capsid. La baseplate sert d'attache à 6 fibres longues (1450 Å) et comprend 6 fibres courtes, qui sont repliées et forment la base distale du dôme. Après fixation de la queue avec la capsid, le « cou » sert d'attache à un collier et 6 « moustaches » (500 Å) de fibratine (coiled-coil trimerique).

La partie centrale de la queue se compose du tube et de son fourreau contractile (contractile sheath) polymérisés autour de la protéine vernier gp29 (tape measure protein). Le tube de la queue est une structure de 925 Å de long et de 90 Å de diamètre avec un lumen de 40 Å formé par un empilement d'anneaux hexamériques de gp19. L'empilement de gp19 a une géométrie hélicoïdale avec un pas à droite, un allongement par anneau de 40,2 Å et une rotation par anneau de 17,9°. Le tube est entouré d'un fourreau contractile formé de l'empilement de 23 anneaux d'hexamères de gp18. Ce tube externe, de 925 Å de long et de 240 Å de large, a également une structure en hélice avec un pas à droite, un allongement par anneau de 40,6 Å et une rotation de 17,2 Å. Dans sa partie distale, le tube de gp19 est terminé par 2 anneaux hexamériques, un de gp54 puis un de gp48. Ces deux anneaux ont des structures et des géométries similaires à ceux de gp19. Remarque : alors que gp48 trouve de nombreux équivalents dans les autres assemblages de phages (cf. 1.4.3.2, 1.4.3.1 & 1.5.3), gp53 représente un module assez spécifique de T4.

Le terminateur se compose des protéines gp3 et gp15 qui forment des anneaux hexamériques se fixant respectivement aux l'extrémités proximales du tube et du sheath. Le terminateur de tube, la protéine gp3, forme un anneau hexamérique interagissant avec l'anneau le plus proximal de gp19 [Fokine 13]. Cette protéine qui termine et stabilise le tube est présumée avoir une géométrie similaire aux anneaux du tube et interagir avec gp29. Le terminateur de queue est formé par un anneau hexamérique de gp15 qui interagit avec gp3, gp14 (protéine du « cou ») et l'anneau le plus proximal de gp18.

La baseplate est un assemblage protéique complexe formé de 12 protéines dif-

férentes et se décomposant en 4 parties : une pointe centrale, une partie intérieure, une partie intermédiaire et un réseau de fibres périphérique. La pointe se trouve dans la continuité du tube. Elle est formée de 3 copies de gp27, 3 de gp5 et d'une copie de gp5.4. Gp27 forme un anneau pseudo-hexamérique qui fait la transition entre la symétrie 6 du tube et la symétrie 3 de la pointe. Gp5 est subdivisé en 3 domaines : le domaine N-terminal fait le lien avec l'anneau de gp27, il a un repliement de domaine de liaison aux oligonucléotides. Le domaine central porte une activité lysozyme et entoure la partie supérieure des 3 domaines C-terminaux qui forment une triple hélice- β . Ce domaine C-terminal, dont la topologie est similaire aux fibres amyloïdes, est stabilisé par la coordination de magnésium par des résidus Histidines et la fixation des chaînes d'acides gras [Buth 15]. Cette structure en hélice- β est également retrouvée dans des structures bactériennes (système de sécrétion de type 6, pyocine-R) et d'autres phages, pour les phages P2 et phi92 la coordination d'un atome de fer permet de stabiliser cette structure [Browning 12].

La baseplate intérieure est principalement composée d'un iris dodécamérique de gp6. 6 molécules de gp6 forment un anneau intérieur (gp6A) alors que les 6 autres sont plus éloignés du centre et ont une orientation inverse (gp6B). Les domaines C-terminaux forment des dimères « têtes bêtes » gp6A-gp6B alors que les domaines N-terminaux forment des modules hétéro-trimériques en hélice avec des segments de la protéine gp7. La protéine gp6 a également un troisième domaine impliqué dans des interactions avec d'autres éléments de la baseplate (voir ci-après). Le module trimérique (gp6)₂-(gp7) sert d'attache à 2 autres protéines : gp25 et gp53. gp25 qui se fixe à l'extrémité des hélices et interagit avec gp53 et gp18 vers l'extérieur et l'anneau de gp48 (dernier anneau du tube) vers l'intérieur. Le domaine supplémentaire de gp6A interagit avec l'interface gp27-gp48. Gp53 enserre le cœur en hélices de (gp6)₂-(gp7) et complète l'interaction du domaine supplémentaire de gp6B avec le dernier anneau de gp18 (fourreau contractile) [Taylor 16].

La baseplate intermédiaire est formée du reste de gp7 et de dimères de gp8 qui permettent des interactions avec les attaches des fibres longues et courtes. Un trimère de gp9 sert d'attache aux fibres longues. Un trimère en forme de lettre X de gp10 sert, avec gp11, de fixation aux fibres courtes (trimères de gp12). Gp10 permet la liaison avec la protéine gp7, cette interaction est particulièrement forte puisqu'il existe même un pont disulfure intermoléculaire. Lorsque le virion est libre en solution, les fibres courtes sont repliées contre la baseplate et forment la partie périphérique.

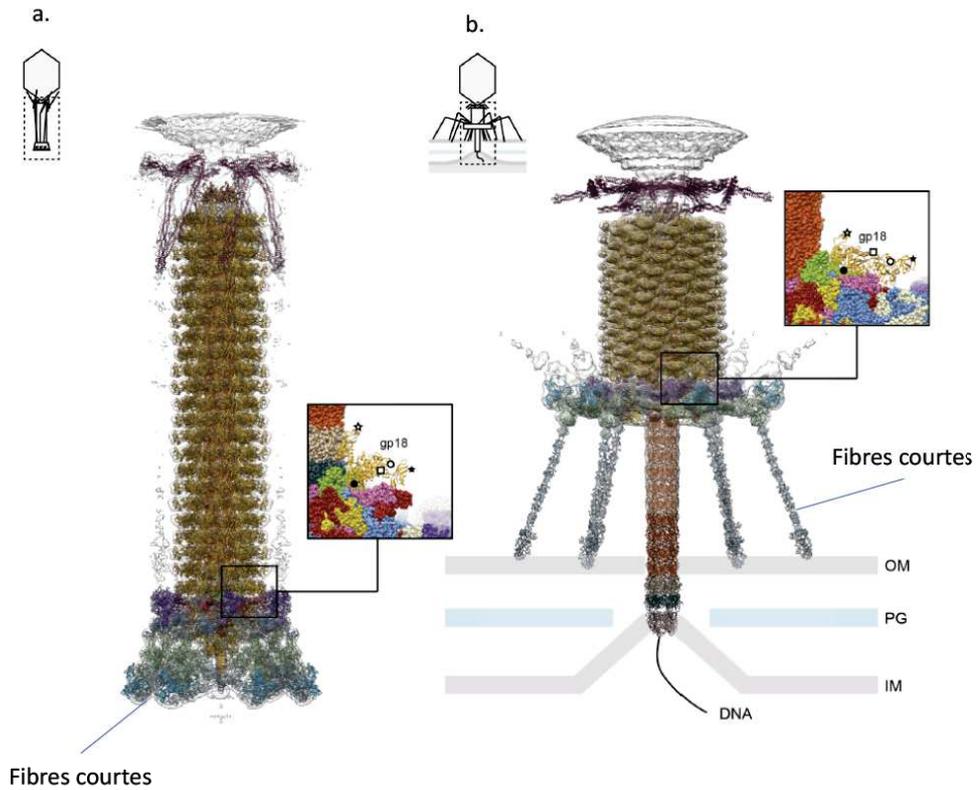


Figure 1.5 : Contraction de la queue du phage T4 permettant la perforation de la paroi de la cellule. a. avant adhésion à la paroi de la cellule, les fibres courtes sont repliées sous la *baseplate*. La fixation des fibres longues (non visibles sur cette reconstruction) libère les fibres courtes. **b.** une fois les fibres courtes irrévocablement fixées à la paroi de l'hôte, la contraction du tube externe se déclenche. Le tube interne, rigide, perce la membrane externe (OM) et dégrade localement le peptidoglycane (PG). Le tube interne ne perce pas la membrane interne (IM) qui s'invagine pour rejoindre l'extrémité du canal (réarrangement peut-être assuré par la TMP).
Figure adaptée de [Taylor, 16].

Changements de conformations pendant l'infection

La contraction de la queue est provoquée par la liaison des longues fibres au LPS à la surface de la bactérie cible. L'information de l'adsorption est véhiculée à la baseplate par l'intermédiaire des trimères de gp9.

La baseplate perd sa conformation en dôme et prend une forme de basse énergie à l'aspect d'étoile à 6 branches de 610 Å pour 120 Å de haut. Il est intéressant de noter que si toutes les protéines participent au changement de conformation, les domaines protéiques pris individuellement sont semblables à leur structure d'origine, à l'exception notable de la boucle 841-862 de gp7, située au niveau du module hétéro-trimérique (gp6)₂-(gp7), qui passe d'un côté à un autre de la trifurcation. Les fibres courtes sont les structures dont les réarrangements sont les plus importants, elles se déplient de la base de la baseplate grâce à une rotation du module gp10-gp11-gp12. Les domaines C-terminaux de gp12, maintenus à la base du dôme par des interactions avec gp11 et les autres fibres sont libérées et peuvent interagir avec un récepteur à la surface de la bactérie. Gp11 interagit alors avec les fibres longues. Ces mouvements sont transférés vers la baseplate intérieure par le module (gp6)₂-(gp7) qui s'incline de 30°, ils provoquent des changements de conformations dans l'anneau de gp6, dont le diamètre augmente, et les protéines gp25 et gp53 qui se dissocient de l'extrémité du tube et de la pointe et provoquent la contraction du fourreau de gp18.

Le tube externe de gp18 se contracte pour que sa longueur finale soit un peu moins de la moitié de ses 925 Å d'origine. L'allongement par anneau passe de 40,6 Å à 16,4 Å et la rotation par anneau de 17,2° à 32,9°. La structure d'un monomère reste inchangée entre les 2 états, la molécule s'incline d'environ 45° dans le sens des aiguilles d'une montre en s'éloignant de l'axe de la queue de 50 Å [Aksyuk 09] suggérant que la contraction se réalise par glissement des molécules de gp18 les unes par rapport aux autres [Yap 14]. La contraction se fait depuis la baseplate vers la capsid. L'anneau le plus proximal de gp19 interagit avec la protéine gp15 et l'interaction entre ses 2 anneaux change de façon importante après la contraction.

La pointe a pour rôle de pénétrer la membrane externe et de dégrader le peptidoglycane, à l'instar du tube interne de gp19, il ne subit pas de changements de conformations. Après dégradation locale du peptidoglycane, le module (gp5)₃-(gp5.4) se dissocie et gp27 devient l'extrémité distale du tube. Des phénomènes complexes et régulés de fusion avec la membrane interne permettent de terminer l'assemblage d'une structure trans-périplasmique assurant le passage de l'ADN phagique. Cette dernière étape pourrait impliquer la protéine vernier gp29 (Fig. 1.4).

Aspects fonctionnels

La nature précise des récepteurs reconnus par les fibres longues et courtes n'a pas été déterminée, les fibres longues sont principalement repliées contre le tube externe, maintenue par les « moustaches », lorsque le virion est libre. Les fibres longues sont capables d'interagir réversiblement avec la membrane, permettant au phage de se déplacer à la surface de son hôte pour trouver son/ses récepteurs. Le phage pénètre plus fréquemment son hôte au niveau du pôle de la cellule. Il était pensé que la fixation des 6 fibres longues était nécessaire pour déclencher la pénétration mais il a été montré que la fixation de 3 seulement permettait de libérer les fibres courtes. La fixation des 6 fibres courtes déclenche la contraction du tube.

In vitro, il existe différentes façons de déclencher la contraction du fourreau, l'incubation avec du LPS ou de l'urée provoque des changements de conformation au niveau de la baseplate et la contraction [Coombs 94] mais l'exposition à la chaleur [Arisaka 81] ou à des détergents cationiques [To 69] provoque la contraction du sheath sans changements conformationnels de la baseplate.

La contraction du tube externe de gp18 permet la pénétration de la membrane externe seulement [Simon 67a, Simon 67b], la liaison avec la membrane interne est un mécanisme indépendant et régulé : l'incubation de phages dont la contraction a été provoquée peuvent infecter des sphéroplastés (bactéries dont la membrane cytoplasmique est accessible), cette infection peut être inhibé par le phosphatidyl-glycerol [Baumann 70]. Ce phospholipide peut donc être considéré comme un récepteur de la membrane interne.

Modèle

L'accumulation de données structurales et fonctionnelles permet de proposer avec un haut degré de certitude des toutes premières étapes de l'entrée virale de T4 : les fibres longues se fixent à la surface de la bactérie et déclenchent un changement de conformation au niveau de gp9 quand au moins 3 fibres se sont fixées. Les fibres courtes sont libérées et peuvent à leur tour fixer leur récepteur, stabilisant définitivement le virion dans une orientation particulière à la surface de l'hôte. La rotation du module gp10-gp11-gp12 se répercute, en particulier via gp7 et la liaison covalente qui l'unit à gp10, vers la baseplate interne qui se détache de la pointe et provoque la contraction du tube extérieur de gp18. La contraction de gp18 provoque la percée de la membrane externe du tube interne et de la pointe, cette dernière dégradant le peptidoglycane une fois dans le périplasme via les domaines lysosymes de gp5.

Ces étapes expliquent le phénomène de perforation, mais de nombreuses zones d'ombre restent à explorer sur la suite de l'entrée virale. Plusieurs hypothèses

peuvent être cependant avancées :

- Les changements de conformations au niveau de l'interaction gp18-gp15 se répercutent le long de gp13-gp14 (« cou ») et provoquent l'ouverture de la capsid et l'éjection de l'ADN et de gp29 [Yap 14].
- La dissociation de la pointe permet la sortie de gp29 qui interagit avec la membrane interne, la sortie de gp29 provoque des changements de conformation au niveau de gp3 (terminateur de tube) qui sert de signal à l'éjection de l'ADN [Hu 15].
- L'interaction de la protéine majeure de tube avec la membrane provoque un changement de conformation répercuté dans tout le tube jusqu'au terminateur [Goldberg 94, Nováček 16].

Parenté avec d'autres systèmes de perforation de membrane

Le mécanisme du phage T4 a été décrit comme représentatif des myophages mais aussi de structures bactériennes utilisant des tubes contractiles. Les protéines du fourreau contractile, du tube et de la pointe ont été décrites comme des orthologues structuraux de leurs équivalents au sein du système de sécrétion de type 6 (SST6) [Leiman 09], les éléments de la structure interne de la baseplate, à l'exception de gp53, et le mécanisme menant à la contraction du tube sont également communs avec le SST6, en particulier gp25 et son interface avec le sheath [Taylor 16]. Alors qu'il n'existe que peu ou pas d'identité/homologie au niveau des séquences nucléotidiques et peptidiques, les études structurales montrent de très fortes ressemblances et suggèrent une origine évolutive commune aux protéines de tubes phagiques, les protéines de tube du SST6, les protéases de capsides phagiques [Fokine 16] et les pyocines bactériennes [Nakayama 00].

Les éléments de la périphérie de la baseplate semblent moins universellement utilisés. Chez T4, la proximité structurale de différentes protéines laisse penser qu'il existe un repliement ancestral commun [Taylor 16]. Une autre différence avec la plupart des myophages est l'absence d'activité enzymatique liée à l'adsorption généralement portée par les fibres courtes [Yap 14].

1.4.3 Les *Siphophages*

Les *Siphoviridae* sont caractérisés par une longue queue non-contractile souvent flexible. Le représentant le plus étudié de cette famille est le phage Lambda. Cependant, ce phage tempéré infectant *Escherichia coli* a surtout été étudié pour comprendre les interactions complexes qu'il entretient avec son hôte pendant le phénomène de lysogénie et le mécanisme d'assemblage du virion [Casjens 15, Davidson 12]. Le travail sur Lambda a notamment permis de comprendre le rôle de la TMP dans l'assemblage du virion et la détermination de la longueur de la queue [Katsura 84].

Les structures de la protéine majeure de tube, gpV [Pell 09a, Pell 10] et de la protéine terminatrice de queue gpU [Pell 09b] ont été résolues. Cependant, et malgré l'abondance de données fonctionnelles (voir [Davidson 12] pour revue), peu de données structurales existent sur le mécanisme d'entrée virale. Ce paragraphe présentera d'une part la structure de la queue et le mécanisme d'infection proposé pour le phage SPP1 puis les données structurales obtenues sur des phages infectant des bactéries du genre *Lactococcus*.

1.4.3.1 SPP1

SPP1 est un phage lytique de la famille des *Siphoviridae* infectant la bactérie gram positive *Bacillus subtilis*.

Architecture de la queue

Le virion de SPP1 possède une queue flexible de 190 nm reliée, par l'un de ses vertex, à une capsidie icosaédrique de 60 nm de diamètre.

Le portail de la capsidie est formé par un dodécamère de gp6 à l'intérieur de la capsidie sur lequel se fixe un dodécamère de gp15, adaptateur capsidie-queue. Dans le virion mature, une extrémité de l'ADN de SPP1 est engagée dans le canal formé par ces deux protéines, interagissant avec des boucles de l'intérieur du canal. A la suite de gp15, se trouve un dodécamère de gp16 dont une boucle de chaque monomère est orientée vers le centre du canal, ces boucles interagissent entre elles pour former un feuillet beta en anneau qui ferme le canal [Chaban 15]. Gp17 et une autre protéine non-identifiée (protéine terminatrice et de complétion de queue, que l'on peut assimiler à gp3 et gp15 de T4 ou gp142 et gp143 de T5) arrivent la queue de SPP1 à gp16.

Le tube de 160 nm de long pour 112 Å de large est formé par l'empilement d'environ 40 anneaux hexamériques de gp17.1 (protéine majeure de queue, TTP) et l'intérieur du tube est rempli par la protéine vernier, gp18 [Plisson 07]. Pour SPP1 et d'autres phages de la même famille, un mécanisme de décalage programmé du cadre de lecture de gp17.1 a été mis en évidence, menant à la production de 2 variantes de la TTP, gp17.1 et gp17.1*, à un ratio de 3 : 1. La variante gp17.1* possède une extension de 9,1 kDa qui correspond à un domaine de type FN3 (domaine en sandwich bêta, similaire à ceux trouvés dans les fibronectines de type III). Le domaine N-terminal commun aux 2 variantes de la TTP forme la paroi du tube et les 2 variantes de la protéine sont capables, indépendamment, de polymériser en des tubes fonctionnels [Auzat 08]. La structure prédite et modélisée du domaine N-terminal de gp17.1 est similaire à celle des TTP de Lambda, de T4 et de protéines du système de sécrétion de type 6 [Langlois 15].

L'extrémité distale du tube est équipée d'un capuchon qui se rétrécit jusqu'à

un diamètre de 40 Å, et passe d'une symétrie 6 à une symétrie 3 avant la liaison avec la fibre droite [Plisson 07]. La partie haute du capuchon est formée par la protéine gp19.1 (protéine distale de queue, Dit) dont le domaine N-terminal forme un anneau hexamérique de structure homologue au repliement commun des tubes de phages et de SST6. Le domaine N-terminal adopte un repliement galectin-like [Veesler 10b]. Le reste du capuchon est formé par un domaine de gp18 qui a été clivé du reste de la protéine [Plisson 07].

L'extrémité distale de la queue forme une fibre rigide de symétrie 3 de 310 Å de long pour 95 Å de large au maximum. Cette fibre est composée de plusieurs domaines et liée au capuchon par un coude flexible permettant environ 50° de liberté par rapport à l'axe du tube. La jonction au capuchon est assurée par un domaine globulaire qui n'a pas été attribué, il est suivi par la protéine gp21 qui forme plusieurs domaines et présente des similarités structurales avec gp9 de P22 (tailspikes) [Plisson 07] et avec la RBP (Receptor Binding Protein) de p2 [Veesler 10a].

Changements de conformations lors de l'interaction avec l'hôte

La perforation de la paroi de l'hôte et l'éjection de l'ADN de SPP1 sont provoquées par la liaison de l'extrémité du phage à l'ectodomaine de la protéine YueB de la paroi cellulaire de *B. subtilis*. Après perforation et éjection de l'ADN, la fibre droite est désolidarisée du complexe et le capuchon à l'extrémité du tube est ouvert pour faire suite au canal permettant le passage de l'ADN. La protéine gp18 est éjectée de l'intérieur du tube, et des changements de conformation de la TTP ont également été décrits [Plisson 07] (Fig. 1.5-b). A l'interface capsid-queue, la protéine gp16 change également de conformation au niveau de son domaine extérieur et les boucles servant de bouchon à la capsid se dénaturent transitoirement pendant le passage de l'ADN et se réassemblent à la fin de l'éjection [Chaban 15] (Fig. 1.5-a).

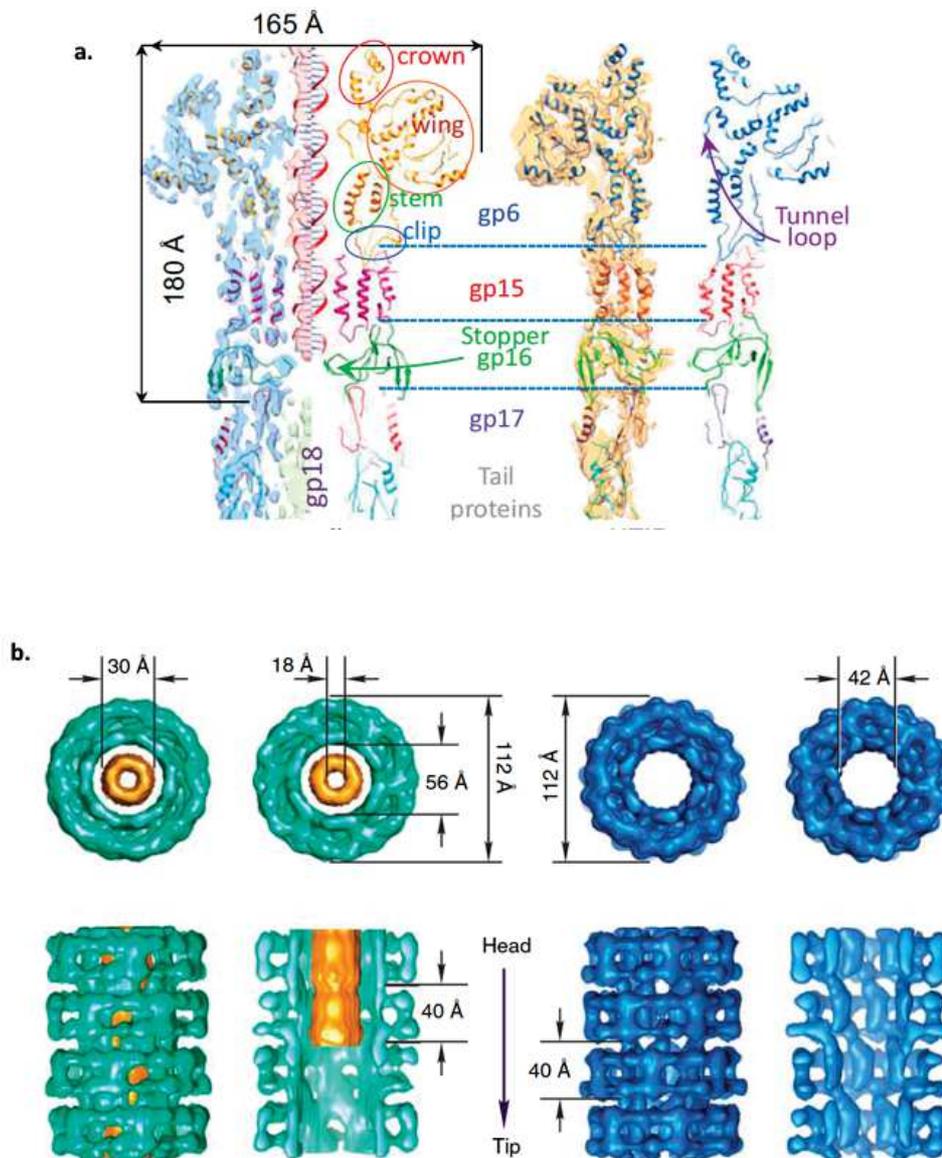


Figure 1.6 : Structures du connecteur et du tube du phage SPP1 avant et après interaction avec le récepteur YueB. a. Modèle du connecteur du phage SPP1 avant et après éjection de l'ADN. La boucle de la protéine gp16 qui bouche le canal dans les 2 conformations (*Stopper loop*) doit transitoirement se dénaturer pour le passage de l'ADN b. Structure du tube du phage SPP1 avant et après éjection. La TMP est déstabilisée et est éjectée après que la TTP ait changé de conformation (la surface d'interaction de l'intérieur du tube et son diamètre ont changé).

Figure adaptée de [Chaban, 15] & [Plisson, 08].

Aspects fonctionnels

Comme dit précédemment, SPP1 utilise YueB comme récepteur, pour lequel la fixation est irréversible et est assurée par la protéine gp21 [Vinga 12]. SPP1 se fixe également réversiblement aux acides teichoïques de la paroi cellulaire, cette interaction est Calcium-dépendante et importante pour l'infectivité du phage. Cette dépendance au calcium peut être expliquée à la fois par un changement conformationnel de la paroi cellulaire [Pal 90] et par une stabilisation du capuchon [Jakutyte 12, Goulet 11].

L'infection de SPP1 provoque une dépolarisation rapide de la membrane de l'hôte (< 1 minutes après incubation) et une fuite lente et faible de potassium [Jakutyte 12].

Un autre élément d'intérêt sur le virion de SPP1 est que la séparation de la queue et de la capsid peut être provoquée par incubation avec un chélateur (EDTA) sans que l'ADN soit éjecté de la capsid [Tavares 96]. Cela implique que la capsid reste "fermée" même après fixation du complexe de la queue et qu'un changement de conformation doit s'opérer au niveau du connecteur tête-queue lors de l'interaction avec l'hôte.

Modèle proposé

Plusieurs hypothèses sont proposées dans [Jakutyte 12] sur le comportement du phage, avant le déclenchement des changements de conformations de perforation de la membrane et d'éjection de l'ADN, pour traverser la paroi cellulaire et accéder à la membrane de l'hôte. Le capuchon s'ouvrirait au niveau de la membrane (ou d'un troisième récepteur non identifié). Le réarrangement du capuchon provoquerait un changement de conformation du premier anneau de TTP qui serait répercuté en cascade le long du tube. Lors de ces changements de conformation, la surface interne du tube changerait, provoquant la déstabilisation de la TMP (gp18) qui serait alors éjectée. Au niveau du connecteur, les changements de conformations de la TTP se répercuteraient au niveau de gp16 qui ouvrirait transitoirement le canal, permettant l'ouverture de la capsid et l'éjection de l'ADN.

1.4.3.2 Lactophages

Les phages lytiques infectant les bactéries à gram positif *Lactococcus lactis*, sont ubiquitaires et très majoritairement des *Siphoviridae* [Moineau 02]. Ils ont dans la grande majorité des grandes plateformes d'adsorption à l'hôte (baseplate) et ont une spécificité d'hôte très étroite.

Les phages TP901-1 et p2, respectivement caractéristiques des groupes P335 et 936, ont été particulièrement étudiés. Ils ont tous les deux un connecteur très semblable à celui de SPP1 avec 2 protéines adaptatrices entre le portail dodécamérique

et le terminateur de queue. Ils ont des tubes d'environ 1170 Å de long, correspondant à environ une trentaine d'anneaux hexamériques de la TTP. Le repliement de la TTP est prédit être le même que les autres protéines de tube (SPP1, Lambda, T4, SSTVI. . .). Les tubes de TP901-1 et p2 se distinguent par la proéminence des domaines de décoration de p2 [Spinelli 14].

L'extrémité distale de leur queue est équipée d'une plateforme d'adsorption qui porte plusieurs RBP. Le fait que ces phages reconnaissent des sucres à la surface de la bactérie plutôt qu'un récepteur protéique explique la nécessité d'augmenter la valence des RBP, l'affinité des interactions protéine-sucre étant plus faibles. La partie centrale des plateformes d'adsorption de ces deux phages est similaire et a également des similarités avec d'autres systèmes (T4 notamment). Cette partie centrale est composée de la protéine distale de queue (Distal tail protein, Dit), homologue structural de la Dit de SPP1 et gp48 de T4, et d'une lysine (Tail associated lysine, Tal) qui forme l'extrémité distale dans l'axe du tube. Les RBPs sont des protéines qui s'organisent en trimères, formées de trois domaines distincts (« épaule » / « cou » / « tête ») qui rappellent la structure de gp12 de T4 ou de certaines protéines de liaison au récepteur de certains virus de mammifères de la famille des adénovirus ou réovirus. Le domaine de tête est très similaire entre les 2 phages et porte le site de liaison aux sucres, le domaine cou forme une hélice Beta dans les deux cas alors que le domaine épaule, qui assure la liaison avec le reste de la plateforme d'adsorption, est variable. En effet, les structures des plateformes périphériques, portant les RBPs, sont très différentes entre TP901-1 et p2 [Veesler 12].

Dans le cas de p2, le domaine C-terminal de la Dit forme une extension qui lie un trimère de RBP. En absence de calcium, les RBPs sont repliées en direction de la capsid, la fixation de calcium à une boucle de la Dit provoque une rotation d'environ 200° des RBPs qui sont alors dans une conformation permettant l'adsorption à l'hôte (Fig. 1.6-b-c). Ce mécanisme d'activation explique la dépendance de p2 au calcium et est absent chez TP901-1.

La plateforme de TP901-1 est équipée d'une protéine supplémentaire : la BppU (Upper Baseplate protein). Cette protéine forme un coude à partir de la partie extérieure de la Dit et porte 3 trimères de RBPs, une BppU étant fixée à chacun des hexamères de la Dit, il y a alors un total de 54 RBP, contre 18 pour p2 [Bebeacua 10] (Fig. 1.6).

Les changements de conformations provoqués par la liaison des sucres et la suite d'évènements qui en découle n'ont pas été décrits à ce jour. Mais ces travaux apportent des informations précieuses sur la modularité et les similarités entre protéines structurales de différents phages.

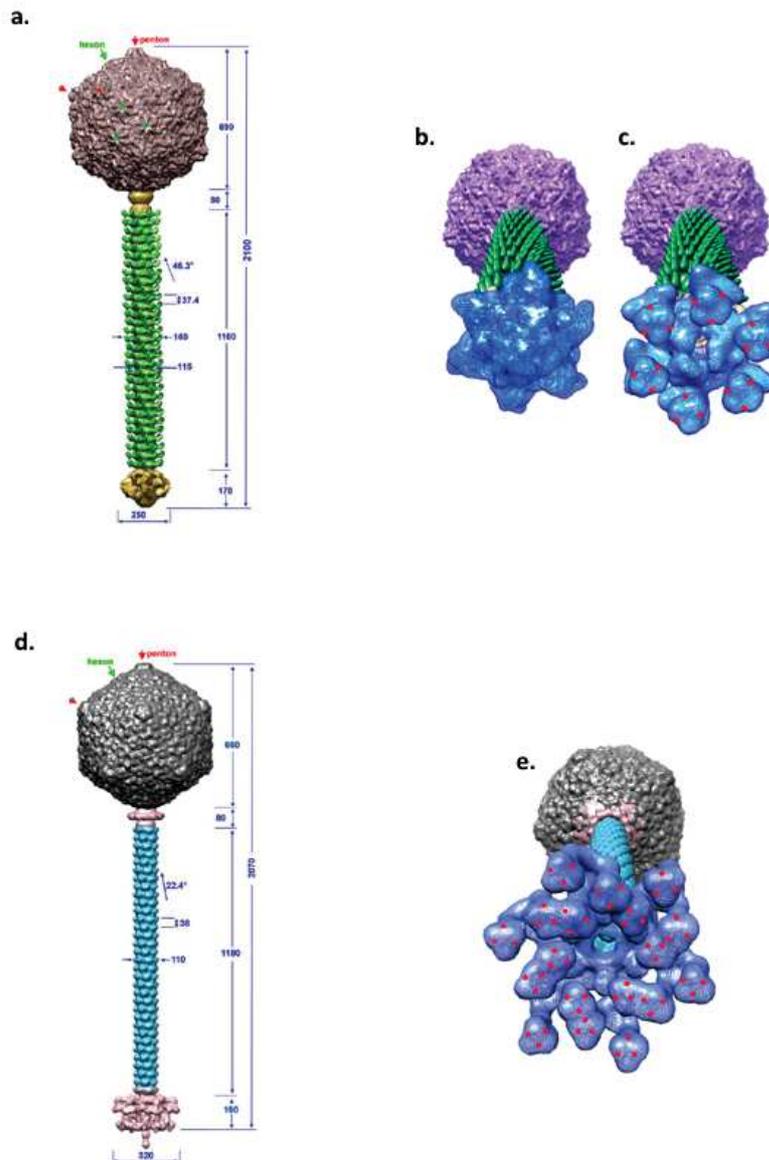


Figure 1.7 : Structures des lactophages p2 et TP901-1. a.-c. Structure du phage p2. b.-c. Vue de la *baseplate* avant et après activation par le calcium. Après activation, les RBPs sont exposées et peuvent interagir avec la membrane (points rouges). De plus, les changements de conformations pendant l'activation mènent également à l'ouverture d'un canal. d.-e. Structure du phage TP901-1 et vue de sa *baseplate*.
Figure adaptée de [Spinelli, 14].

1.5 Le bactériophage T5

1.5.1 Spécificités de T5

Le phage T5 est un *Siphoviridae* (phage à longue queue non contractile) qui infecte *E. coli*, il s'agit d'un phage strictement lytique. C'est un virus à ADN double-brin qui utilise comme récepteur la protéine membranaire FhuA, un transporteur de ferrichrome de la membrane externe d'*E. coli* qui sert également de récepteur aux phages T1, 80 et UC-1 et de transporteur à plusieurs toxines (allomycine, rifamycine, colicine M, microcine J25) [Ferguson 98]. T5 présente plusieurs particularités qui seront pour la plupart développées dans les paragraphes suivants :

- Son ADN a la particularité de présenter des nicks (résidus clivés) sur l'un de ses brins. Les zones de clivages ne semblent pas aléatoires mais leur rôle n'a pas été encore compris. Ces nicks ne sont pas présents lors de la réplication, mais le sont pour l'encapsulation de l'ADN.
- T5 ne synthétise pas d'ARN polymérase, il utilise celle d'*E. coli* grâce à des promoteurs très efficaces.
- Contrairement à d'autres phages (80 par exemple), l'infection n'est pas énergie-dépendante et ne nécessite pas que le récepteur soit lié avec le complexe TonB (permet médiation de l'énergie du gradient électro-chimique de la membrane interne). L'éjection de l'ADN peut être déclenchée *in vitro* par incubation avec FhuA (en détergents ou protéoliposomes)[Böhm 01] (Fig. 1.7-b).
- *In vivo*, l'infection du bactériophage T5 se passe en 2 temps, une première partie du génome viral est injectée et exprimée avant que l'injection de l'ADN ne reprenne jusqu'à éjection totale. *In vitro*, ce mécanisme n'est pas conservé avec la même précision que *in vivo* mais se distingue néanmoins des autres : sensibilité aux pressions non-canonique (à des pressions osmotiques <6atm, il n'y a plus de corrélation entre la quantité d'ADN éjectée et la pression appliquée, contrairement à ce qui est observé chez SPP1 et Lambda) et observation de « pauses » en des régions non-aléatoires. La possibilité que les clivages d'un des brins d'ADN aient un rôle dans ce processus a été évoquée sans qu'aucune expérience déterminante n'ait encore pu être mise en place [Ponchon 05, Leforestier 08].

1.5.2 Génome

T5 a un génome de 121 750 pb composé d'une séquence unique de 111 531 pb suivi d'une zone de 10 219 pb qui est la répétition directe des 10 219 premières paires de base en 5' (zone non codante). L'un des brins de la double hélice est clivé, en moyenne en 8 endroits, 4 de ces sites de clivages (dits majeurs) se font presque systématiquement (80-90% des molécules) en des points précis : à 7,9

(9500e pb), 18,5 (22500e), 32,6 (40000e) et 64,8% de la taille du génome (78 900e). Les autres sites de clivage n'ont pas une occurrence aussi précise et reproductible mais n'est pas strictement aléatoire, ces sites sont plus fréquents sur les extrémités de la chaîne nucléotidique. Remarque : la souche st(0) a une délétion dans la zone de clivage à 32,6%, elle n'a donc que 3 spots de clivages majeur. (Genbank accession number AY543070 & 692264 ; [Wang 05, Scheible 77] ; Rapport HDR Pascale Boulanger).

D'un point de vue fonctionnel, le génome du virus T5 se divise en 3 régions :

- Une région portant les gènes « anté-précoce » qui est la première à être transférée et exprimée dans le cytoplasme de la bactérie. Ces gènes sont localisés à l'extrémité de l'ADN, entre 0% et 8,3% (et donc sur la zone répétée à l'autre extrémité génome). L'expression des protéines de cet « ADN de premier transfert » provoque l'arrêt de la réplication, la transcription et la traduction bactérienne ainsi que la destruction de l'ADN de la bactérie hôte. Cet ADN de la première partie de transfert code également une déoxyribonucléotidase, probablement pour éviter l'accumulation toxique de dNMPs. Les protéines codées par cette partie de l'ADN viral permettent également la reprise du transfert d'ADN (après environ 5 minutes). Remarque : ces gènes ne sont plus exprimés une fois le transfert d'ADN terminé.
- La région entre 8,3 et 67,3% du génome porte la plupart gènes précoces qui sont exprimés une fois le transfert de l'ADN viral terminé (pendant environ 8 minutes). Ces gènes codent pour des protéines impliquées dans la réplication de l'ADN, le métabolisme des nucléotides et des facteurs de lyse de la cellule.
- Enfin, les gènes tardifs, qui sont principalement des gènes structuraux (protéines formant le virion, ainsi que celles permettant son assemblage), sont localisés entre 67,3 et 90,8% du génome. SciA et SciB, prédites comme des endonucléases responsables des clivages de l'ADN viral, sont également des gènes tardifs (p149 a été identifié comme l'un d'entre eux). Les gènes tardifs ne sont exprimés que 12-13min après le début de l'infection.

1.5.3 Aspects structuraux

La capside de T5 a un diamètre moyen d'environ 900 Å, sa queue mesure 2500 Å pour un diamètre de 150 Å et se termine par une structure conique et une fibre droite de 500 Å.

La capside a une forme quasi-icosahédrique de 900 Å de diamètre (830-940), avec des parois de 40 Å d'épaisseur. Cette capside correspond à un icosaèdre ayant un nombre de triangulation $T=13$ laevo : 12 sommets au niveau de pentamères, séparés par 120 hexamères. Remarque : un des sommets est équipé du portail, dodécamère de pb7 et du connecteur capsid-queue, p144. Elle est composée de

730 copies de pb8, protéine majeure de capsid. 120 copies de pb10, protéine de « décoration », viennent compléter l'enveloppe de la capsid en se plaçant au centre de chaque hexamère de pb8. L'intérieur de la capsid est marqué par la présence de cavités au niveau de chaque capsomère. Les structures du portail et du connecteur restent à élucider [Vernhes 17, Huet 10, Zivanovic 14].

De manière contre-intuitive, le diamètre de la capsid augmente de volume lorsque l'ADN est éjecté (-10 Å au niveau des arrêtes mais pas de changement au niveau des sommets) ce changement de diamètre est accompagné par l'apparition d'une légère courbure. En microscopie électronique, l'ADN est visible sous forme de couches concentriques de densités séparées de 24,4 Å (rappel : diamètre ADN-A = 25,5 Å). La capsid de T5 est très similaire en taille et en géométrie à celle de T4 ; mais son organisation, son mécanisme d'assemblage et les cavités intérieures de la capsid sont proches de HK97 [Effantin 06]. La structure du portail (pb7) et du connecteur (p144) n'a pas été déterminée, mais une reconstruction préliminaire montre une organisation en trois domaines, à l'instar des phages T7 et SPP1 (pour lesquels cette structure a été décrite) (suppléments de [Effantin 06]).

La queue du bactériophage T5 est liée à la capsid par la protéine terminatrice (p142) et la protéine de complétion de queue (p143). La structure de cette extrémité proximale de la queue n'est pas connue. La protéine majeure de queue (pb6) forme un tube de 190 nm avec une organisation hélicoïdale suivant une symétrie d'ordre 3. Ce tube est formé par la juxtaposition, tout les 43,8 Å, de 40 anneaux trimériques de pb6, chacun relativement décalé de 39,4° (suppléments de [Effantin 06]). Remarque : la répétition axiale des anneaux est plus petite en microscopie électronique à coloration négative qu'en cryo-microscopie électronique (déshydratation due à la coloration négative). Cette symétrie d'ordre 3 et atypique, une symétrie d'ordre 6 est généralement retrouvée dans les queues de phages. La lumière du tube formé par pb6 est d'environ 50 Å, et est remplie par le domaine N-Ter de pb2 (stoechiométrie inconnue) tant qu'il n'y a pas d'infection.

L'extrémité distale du tube se termine sur un hexamère de pb9 (protéine distale de queue = « Dit »), une protéine homologue des Dits de lactophages, de gp19.1 de SPP1 et de gp48 de T4. La structure de pb9 (monomère) a été déterminée

par cristallographie, la protéine est composée de deux domaines : un domaine d'hexamérisation qui est similaire au repliement commun de protéines de tubes (Dits, TTPs, protéines de SST6...) et un domaine non conservé avec une structure de liaison aux oligonucléotides/oligosaccharides. La fonction de ce second domaine n'a pas été déterminée. L'anneau hexamérique de pb9 (modélisé à partir d'hexamères de Dit connue) présente deux faces de charges opposées, suggérant de fortes liaisons électrostatiques avec ses partenaires au sein de l'assemblage complet de la queue. L'intérieur de l'anneau est chargé négativement, facilitant le passage de l'ADN lors de l'infection [Flayhan 14] (Fig. 1.7).

La partie C-terminale de pb1 a été cristallisée et correspond à l'extrémité des fibres flexibles qui fixent réversiblement le LPS [Garcia-Doval 13, Garcia-Doval 15]. Les fibres flexibles sont rattachées au reste de l'assemblage par un collier, environ deux fois plus large que le tube formé par pb6 (145 Å), formé du N-terminale de pb1 et de p132 et fixé à l'interface entre le dernier anneau de pb6 et pb9. L'axe du tube est poursuivi par une structure conique (formé par un oligomère de pb3) et la fibre droite de 500 Å de long (formé par un oligomère de pb4 et terminé par la protéine de liaison au récepteur : pb5) [Zivanovic 14]. Peu d'informations structurales sont disponibles pour pb3 et pb4. pb3 a été identifiée comme la Tal de T5, homologue de gp27 chez T4 [Flayhan 14]. La structure basse résolution de pb5 en complexe avec FhuA a été déterminée par des expériences de SANS et d'EM [Flayhan 12].

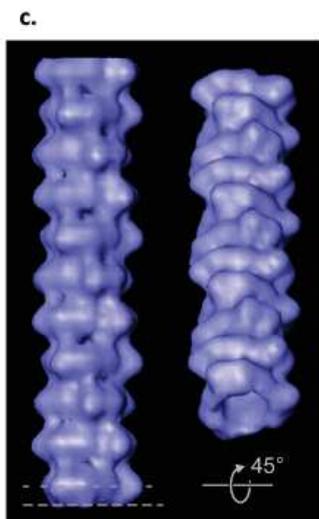
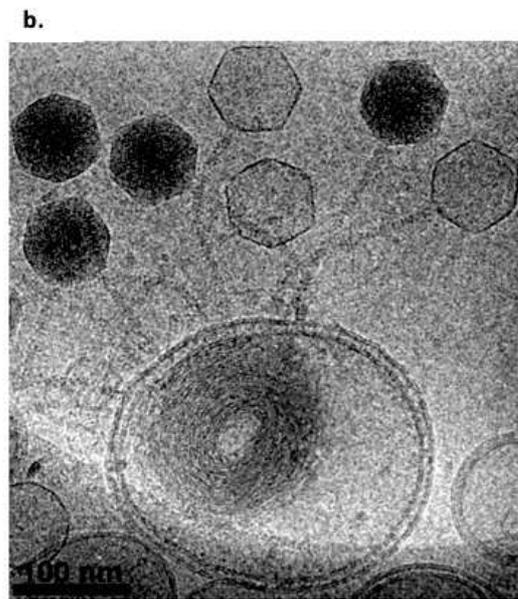
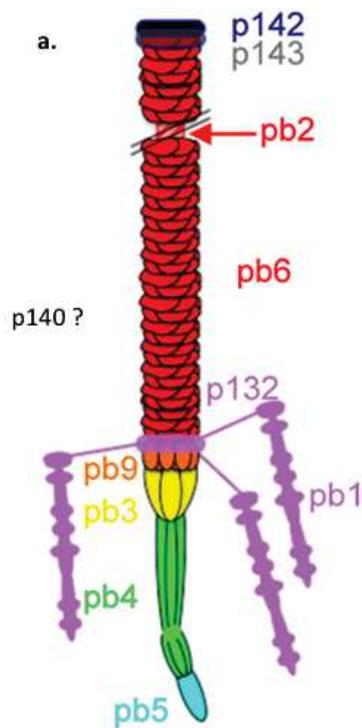


Figure 1.8 : Queue du bacteriophage T5. a. Schéma de la queue du phage T5. 10 des 11 protéines du complexe ont été identifiées. Le code couleur de ce schéma sera réutilisé dans la partie III. b. Micrographe en coloration négative de bactériophages T5 incubés avec des protéoliposomes de FhuA. Certains phages ont réussi à éjecter leur matériel génétique à travers 2 bicouches lipidiques, démontrant que le mécanisme d'entrée virale de T5 est indépendant de tout autre élément bactérien autre que son récepteur. c. Reconstruction 3D du tube de la queue du phage T5 à partir de clichés de microscopie électronique à coloration négative. Le tube est formé par la superposition d'anneaux trimériques. Un domaine de la protéine pointe vers l'extérieur.

1.5.4 Assemblage des virions

Queues et capsides sont assemblées selon 2 séquences d'évènements distinctes :

Pb8 et pb7 forment une pro-capside qui, après clivage, devient une pré-capside prête à accueillir l'ADN virale. Contrairement à beaucoup d'autres phages possédant des capsides de taille $>500 \text{ \AA}$, la capside de T5 n'utilise pas de protéines chaperonnes. Un domaine de la protéine majeur de capside, pb8 (32kDa mature, 51kDa avant clivage), possède un domaine qui joue ce rôle (-domain, 128 résidus N-ter) et est clivé par pb11. Pb11 (16kDa post-clivage) participe également à sa propre maturation (clivages N et C-ter) et à celle de pb7 (44kDa mature). Pb7 est la protéine s'assemblant en un dodécamère pour former le portail. L'encapsidation de l'ADN viral est assurée par l'association transitoire du portail avec un hétérodimère : la terminase. La petite sous-unité de la terminase (TerS = p152 a priori), sert à la fixation de l'ADN phagique et à la régulation de l'encapsidation (pourrait également être impliquée dans la régulation de la formation des « nicks » de l'ADN phagique). La grosse sous-unité (TerL = p151), assure la translocation de l'ADN dans la capside (ATP-dépendant) ainsi que le clivage de l'ADN qui est concatémérique. Un léger changement de morphologie de la capside pendant l'empaquetage permet la fixation de pb10. Après l'encapsidation de l'ADN, la protéine p144 (19,3kDa) vient « fermer » la capside et forme le connecteur avec la queue [Ponchon 05, Ponchon 06, Leforestier 08, Huet 10, Zivanovic 14].

L'assemblage des queues de phages se fait à partir d'un complexe initial qui comprend surement pb3, pb4 et pb2 qui subit une maturation protéolytique dans son domaine C-terminale. La partie N-terminale de pb2 est maintenue dans une conformation de coiled-coil par des chaperonnes qui sont remplacées par pb6 qui polymérise en tube autour de ce domaine. p142 et p143 achèvent l'extrémité proximale de la queue et permettent la liaison à une capside dans laquelle l'ADN a été empaqueté [Boulangier 08, Zivanovic 14]. pb1 et p132 s'assemblent à une étape non déterminée, il est important de noter que si ces éléments augmentent considérablement l'adhésion du virion à l'hôte, l'infection peut être réalisée en absence des fibres flexibles et du collier [Flayhan 14].

1.5.5 Mécanisme d'infection

La caractérisation des premières étapes de l'infection de T5 fait l'objet de cette thèse. Quelques données supplémentaires doivent être discutées avant d'établir précisément les objectifs de ce travail de thèse :

- La protéine p140 a été identifiée comme une protéine structurale mineure de la queue du phage T5 sans que son rôle et sa position dans l'assemblage aient été établis [Zivanovic 14].
- Le domaine C-terminal de pb2 a une activité muralytique, la capacité de

fusionner des membranes et se comporte comme une protéine membranaire [Boulangier 08]. Ces propriétés font qu'il est raisonnable de penser qu'elle est un acteur majeur dans l'assemblage d'une structure transpériplasmique pouvant assurer le passage de l'ADN.

Ce travail de thèse se concentre sur 2 évènements précis de l'entrée virale de T5:

- La transmission de l'information de fixation irréversible du virion du complexe distale à la capsid. Ce mécanisme a été décrit pour SPP1 être assuré par un changement de conformation en cascade de la TTP [Plisson 07] (cf. 1.4.3.1).
- Les changements de conformations dans le complexe distal de queue qui permettent la perforation de la membrane. Ces changements sont suspectés impliquer de façon préminente la protéine pb2. Etablir la position de la protéine p140 dans l'assemblage est un des objectifs de ce travail.

Points clés

Résumé chapitre 1

- ❑ Les phages sont des acteurs sous-étudiés de notre environnement, ils ont un potentiel dans de nombreuses applications.
- ❑ La majorité des phages connus sont des phages caudés, le virion est équipé d'une queue permettant la reconnaissance de l'hôte et l'éjection de l'ADN.
- ❑ Les mécanismes d'entrée virale des phages semblent très variables du point de vue mécanique et fonctionnel mais font intervenir des assemblages, des protéines ou domaines protéiques similaires.
- ❑ Le repliement commun des protéines de tubes semble concerner toutes les TTPs de Myoviridae et de Siphoviridae, d'autres protéines de phages (Dits) ainsi que les protéines de tubes de structures bactériennes comme le SST6 et les pyocines R. Cette homologie structurale n'est pas visible dans les séquences acides-aminés ou nucléotidiques, mais suggère néanmoins une parenté entre ces structures.
- ❑ La TTP de T5 forme des anneaux trimériques contrairement aux autres TTPs (hexamères). Cette thèse vise à élucider la structure de cette protéine et son rôle dans la transmission de l'information de fixation à la protéine FhuA.
- ❑ Le complexe distal de queue de T5 se distingue de ceux décrits pour les lactophages et T4 (grande plateforme d'attache pour de multiples RBPs, récepteurs saccharidiques) et est plus proche de SPP1 (petite plateforme d'attache pour RBP unique, récepteur protéique). Cette thèse s'intéresse à la structure de ce complexe et de ses changements de conformations lors de la fixation de FhuA.

Chapter 2

Introduction méthodologique

2.1 Introduction

Afin d'élucider le mécanisme d'entrée du bactériophage T5, nous avons adopté une stratégie inspirée des travaux réalisée sur d'autres phages (cf. 1.4), en s'attendant à la résolution des structures de la queue du phage avant et après interaction avec la membrane. Le travail réalisé s'appuie principalement sur des travaux de microscopie électronique, mais également sur la cristallographie de la protéine majeure de la queue, pb6. Ce chapitre propose une introduction rapide de ces deux techniques ainsi qu'un paragraphe sur les moyens d'en intégrer les résultats pour modéliser la structure d'assemblages complexes. D'autres méthodes seront évoquées dans les résultats, en particulier la RMN, mais le travail de l'auteur s'étant limité à la préparation des échantillons, elles ne seront pas introduites ici.

Les résultats avancés dans cette thèse sont issus d'études sur un nombre limité d'échantillons :

- La queue du phage T5 obtenue grâce à un mutant ambre du phage T5 contenant un codon Ambre (codon STOP) dans le gène codant pour la protéine majeure de capsid. Le choix de cet assemblage, plutôt que le phage entier, a été motivé par l'espoir d'obtenir de meilleures résolutions en microscopie électronique (meilleure répartition des particules dans la glace et moins d'épaisseur du fait de l'absence de capsid, absence d'ADN dans le bruit de fond lorsque l'éjection est déclenchée). Le complexe de queue reste tout à fait fonctionnel et est capable de changer de conformation pour perforer une bicouche lipidique si tant est que le récepteur du phage y soit inséré.
- La protéine FhuA, un transporteur bactérien de fer-ferichrome, est le récepteur du phage T5. La production, la purification ainsi que la structure cristallographique de cette protéine sont connues depuis longtemps [Ferguson 98].

Solubilisée dans différents environnements (détergents, amphipols, nanodisques), cette protéine est utilisée pour déclencher les événements permettant la perforation de la paroi cellulaire et l'expulsion de l'ADN.

— La protéine pb6, protéine du tube de la queue de T5 (cf 1.5).

La caractérisation structurale de ces différents échantillons a été réalisée essentiellement à l'aide de la microscopie électronique, mais aussi de la cristallographie à rayons X. Ces deux méthodes sont introduites ci-après.

2.2 Microscopie électronique

2.2.1 Généralités

2.2.1.1 Historique

Le concept de l'optique des électrons naît dans les années 1920, avec l'idée que les électrons en mouvement ont un comportement ondulatoire défendue par Louis de Broglie. L'hypothèse que le champ électrique généré par des bobines de fils électrique chargées affecte les électrons de la même manière que des lentilles en verre affecte les photons est avancée par Hans Busch en 1927. Ernst Ruska et Max Knoll travaillent sur cette idée, font breveter la lentille électronique en 1932 et fabriquent un premier prototype de microscope électronique en 1933. Ernst Ruska recevra le prix Nobel de physique en 1986 pour cette invention.

La microscopie électronique à transmission permet de produire des projections 2-D d'un objet biologique, en combinant des projections de différentes orientations d'un même objet, il est possible d'en obtenir une reconstruction 3-D. Cette méthode a été utilisée la première fois en 1968 pour reconstruire la structure de la queue du phage T4 [De Rosier 68]. Huit ans plus tard, Taylor et Glaeser [Taylor 76] introduisent la cryo-microscopie, ils proposent de remplacer la coloration aux métaux lourds par la congélation des échantillons dans un fin film de glace vitreuse, permettant de les observer dans un état proche de l'état natif.

Dans les années 1970 et 1980, la cristallographie 2D permet d'obtenir plusieurs structures à résolution sub-nanométrique et la structure atomique de la bactériorhodopsine en 1990 [Henderson 90]. Cette méthode, bien qu'elle soit l'une des rares à permettre d'accéder à la structure de protéines membranaires dans leur environnement lipidique, est limitée par le prérequis de pouvoir obtenir des cristaux 2D diffractant à haute résolution de la protéine.

Les techniques dites de « particules uniques » (Single Particle analysis), sont mises en avant à partir du début des années 1990. Elles consistent à compenser l'absence de l'organisation rigoureuse des molécules au sein d'un cristal ou d'un arrangement particulier (hélice par exemple), en moyennant de nombreuses molécules

isolées. Cela implique de résoudre les problèmes de bruit de fond et d'alignement des particules (cf 2.2.5).

La reconstruction 3D de particules isolées par cryo-microscopie électronique permet depuis la fin des années 2000, grâce au développement de nouveaux détecteurs et de programmes d'alignement des particules sophistiqués, d'atteindre des résolutions atomiques [Zhang 08] et prend rapidement de l'ampleur, étant considérée aujourd'hui comme une des méthodes majeures de la biologie structurale [Nogales 15].

Il existe trois principaux modes de microscopie électronique. La microscopie électronique à balayage et la microscopie à balayage par transmission fonctionnent par focalisation du faisceau d'électrons en un point précis qui balaye l'échantillon. La microscopie électronique en transmission focalise le faisceau d'électrons en amont et illumine l'échantillon. Ce troisième mode sera le seul discuté dans ce document.

2.2.1.2 Coloration négative et cryo-microscopie électronique

En microscopie électronique à transmission, il existe 2 principales techniques de préparation d'échantillon qui seront discutées dans cette thèse :

- En coloration négative, l'échantillon est placé entre une feuille de carbone et une plaque de mica (quelques μL d'échantillon se répartissent dans cette espace par capillarité). Le mica/échantillon/carbone est déposé sur une solution de métaux lourds (dans le cas des échantillons présentés dans les résultats, une solution de sodium silico-tungstène) côté carbone vers le dessus. Le mica coule et le carbone flotte avec l'échantillon à sa surface inférieur, exposé au métal lourd qui s'accumule autour de l'échantillon. Le carbone est enfin récupéré et déposé sur une grille de microscopie (grille en cuivre de 3 mm avec un grillage ayant une maille de 200 à 400 μm), l'échantillon vers le haut.
- En cryo-microscopie électronique, les grilles sont déjà recouvertes d'un fin film de carbone (15-35 nm) qui est troué. La grille est exposée à un plasma pour atténuer l'hydrophobicité de la surface (glow discharge). Quelques μL d'échantillon sont alors déposés sur la grille, elle est tamponnée pour évacuer un maximum de liquide et éviter que l'épaisseur de glace soit trop importante à la congélation. L'échantillon se retrouve suspendu dans l'eau maintenue par tension de surface dans les trous du carbone. Enfin, la grille est congelée dans de l'éthane liquide qui permet une congélation plus rapide que l'azote liquide et limite donc la formation de glace cristalline plus dense aux électrons que la glace amorphe. Les grilles sont conservées dans l'azote liquide jusqu'à observation. La préparation de ces grilles fait entrer en compte de nombreux facteurs parfois difficiles à maîtriser/optimiser : taille et forme des trous du carbone, temps de glow discharge, temps de tamponnage, température et humidité...

2.2.1.3 Interaction électron-matière

Les électrons n'ont pas une très bonne pénétration de la matière, il est donc nécessaire que les échantillons étudiés soient fins (généralement autour de 50 nm) et que la colonne du microscope soit maintenue sous vide.

Les électrons peuvent traverser un échantillon sans être diffusés, être diffusés de façon inélastique ou élastique (Fig. 2.1).

Lors de la diffusion inélastique, l'électron interagit avec le nuage électronique d'un atome et perd de l'énergie. L'atome disperse l'énergie reçue sous forme de rayons X ou par l'émission d'un électron secondaire, issu de son nuage électronique.

L'électron incident a perdu de l'énergie, il se déplace donc avec une longueur d'onde différente et sera focalisé différemment par les lentilles électroniques. Lors de la diffusion élastique, l'électron incident interagit avec le noyau de charge opposée d'un atome. Cette interaction se fait sans perte d'énergie mais change la phase et la direction de la particule. La diffusion élastique provoque des angles de diffusion plus grands que la diffusion inélastique, lorsque l'angle de diffusion est supérieur à 90° , on parle de rétrodiffusion.

2.2.2 Anatomie d'une microscope électronique à transmission

2.2.2.1 Canon à électrons

La partie supérieure d'un microscope électronique contient une source d'électrons dans laquelle passe un courant haute tension, cette source est entourée d'un cylindre de Wehnelt dans lequel passe un courant à encore plus haute tension. Le cylindre apporte une cohérence spatiale au faisceau d'électrons, parallèle à l'axe de la colonne, qui débouche alors dans l'accélérateur : une pile de plaques sous tension liées les unes aux autres par des résistances, la première plaque étant sous haute tension et la dernière liée à la terre. Ce système génère des champs électromagnétiques perpendiculaires à l'axe de la colonne avec un gradient de tension, apportant leur énergie/vitesse aux électrons. Pour un microscope à 300 kV, les électrons sont accélérés à 76% de la vitesse de la lumière avec une longueur d'onde d'environ 2 pm.

La nature de la source d'électrons va également affecter la cohérence spatiale (direction des électrons parallèles entre eux et à l'axe de la colonne) et temporelle (énergie/vitesse des électrons) du faisceau d'électrons. Dans un ordre croissant de coût et de qualité de cohérence de faisceau, il existe les filaments de tungstène, les cristaux de borure de lanthane et les pistolets à émission électronique.

2.2.2.2 Lentille électronique

Les lentilles des microscopes électroniques sont des bobines de fils conducteurs dans lesquelles un courant électrique passe, générant un champ électrique. Le comportement des faisceaux d'électrons à travers ce champ est analogue à celui des photons à travers une lentille en verre et le voltage appliqué à la bobine détermine la distance focale de la lentille.

La lentille d'un microscope électronique est composée de plusieurs éléments : deux défecteurs en amont de la bobine qui forment des champs perpendiculaires à l'axe de la colonne pour aligner le faisceau incident sur l'axe optique de la lentille ; un système de refroidissement en périphérie de la bobine ; un stigmatiseur en aval qui corrige les défauts du champ électrique formé par la bobine et un diaphragme qui stoppe les faisceaux non alignés autour de l'axe optique.

2.2.2.3 colonne du microscope

La colonne du microscope contient un canon à électrons et 3 systèmes de lentilles : le condenseur qui contrôle l'illumination de l'échantillon, l'objectif qui collecte le faisceau après interaction avec l'échantillon et le système de projection qui augmente le grandissement et projette le faisceau sur l'écran ou le détecteur. Un support pour l'échantillon qui permet d'insérer et déplacer l'échantillon dans la colonne sans briser le vide. La colonne est également équipée d'un filtre à énergie qui permet de séparer les électrons qui ont interagi de façon inélastique de ceux qui ont interagi de façon élastique ou pas interagi avec l'échantillon.

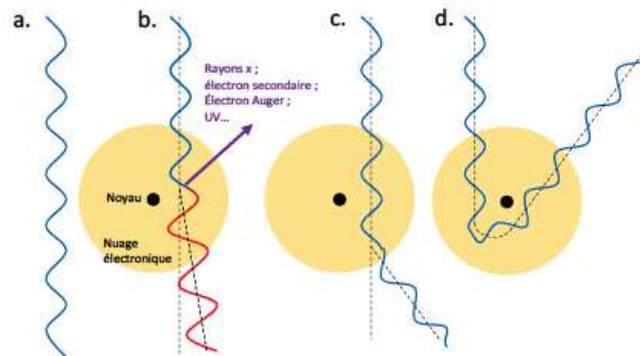


Fig. 2.1 : Interaction électron– matière. a. Electron non diffusé. b. Electron diffusé de manière non-élastique. La longueur d’onde change et l’angle de diffusion est faible. L’énergie perdue par l’électron est dispersée par l’atome sous forme de rayonnement ou de l’émission d’un électron. c. & d. Electrons diffusés de manière élastique, la longueur d’onde n’est pas modifiée mais la phase est décalée et l’angle de diffusion est important. Si l’angle de diffusion dépasse 90°, on parle d’électron rétrodiffusé (d.).

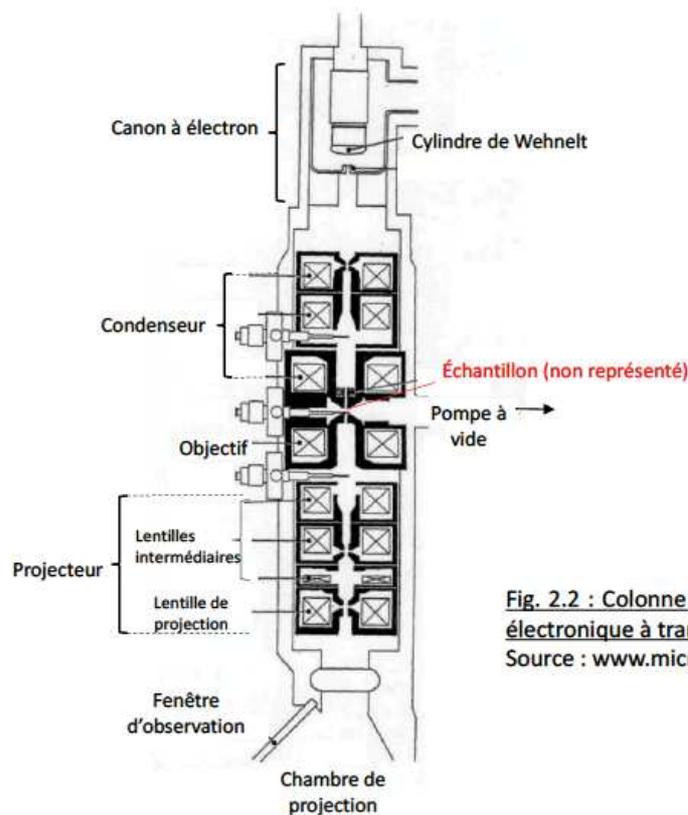


Fig. 2.2 : Colonne d’un microscope électronique à transmission.
Source : www.microscopy.ethz.ch

2.2.3 Formation des images en microscopie électronique à transmission

2.2.3.1 Contraste d'amplitude et contraste de phase

Les images de microscopie électronique à transmission se forment grâce à deux types de contraste :

- Le contraste d'amplitude peut se comprendre comme la soustraction de la contribution des électrons rétrodiffusés et des électrons ayant interagi de façon inélastique avec l'échantillon. En coloration négative, la majorité du contraste se forme de cette façon, contrairement aux images de cryo-microscopie.
- Le contraste de phase provient de l'interférence entre le faisceau n'ayant pas interagi avec l'échantillon et celui diffusé de manière élastique. La phase de ce dernier est décalée lors de l'interaction avec l'échantillon et en fonction de l'angle de diffusion, plus l'angle de diffusion est élevé et plus le trajet entre l'échantillon et le détecteur est long (et donc le décalage de phase important). Ainsi, le contraste de phase lie la contribution au contraste de l'image à la fréquence spatiale (angle de diffusion) et pour certaines fréquences spatiales, le décalage de phase sera proche ou égale à $\pm 90^\circ$, correspondant à une contribution négligeable au contraste de l'image. Ce phénomène est observé dans la fonction de transfert de contraste.

2.2.3.2 Fonction de transfert de contraste

La fonction de transfert de contraste (Contrast Transfer Function – CTF) est une représentation du contraste en fonction de la fréquence spatiale, l'enveloppe de cette fonction diminue vers les hautes fréquences spatiales du fait des défauts de cohérence spatiale et temporelle du faisceau d'électrons et des aberrations du système optique.

En faisant varier le plan de focalisation de l'objectif, le trajet des électrons diffusant élastiquement varie également, ce qui se traduit sur la CTF par un déplacement des points où la courbe passe par zéro. En conservant le plan focal de l'objectif proche du plan conjugué à celui du détecteur (faible défocus) il y a moins de transitions entre les contrastes positifs et négatifs, ce qui se traduit par moins de perte d'informations aux hautes résolutions. Cependant, un défocus plus important apporte plus de contraste aux basses résolutions ce qui rend les particules plus facilement visibles sur les micrographes.

Une autre façon de définir la CTF est comme la transformée de Fourier de la fonction de dispersion de point (Point Spread Function – PSF). La PSF est la représentation de l'image d'un point à travers un système optique donné. Cette

fonction est représentative des déformations de l'image par rapport à l'objet (Fig. 2.3). A partir d'un voltage et de paramètres d'aberrations connus, la CTF peut être déduite de la transformée de Fourier d'une image (à l'aide des anneaux de Thon). Une fois connue, la CTF peut être utilisée pour restaurer une partie des informations perdues en divisant la transformée de l'image par la CTF (Fig. 2.3). Cette opération ne peut pas être réalisée pour les points où la CTF passe par zéro, la correction de CTF d'une superposition d'images d'un même objet prises avec différentes valeurs de focus permet de combler l'information manquante à ces fréquences spatiales. La superposition d'images pour lesquelles la correction de CTF a été faite individuellement a le même effet.

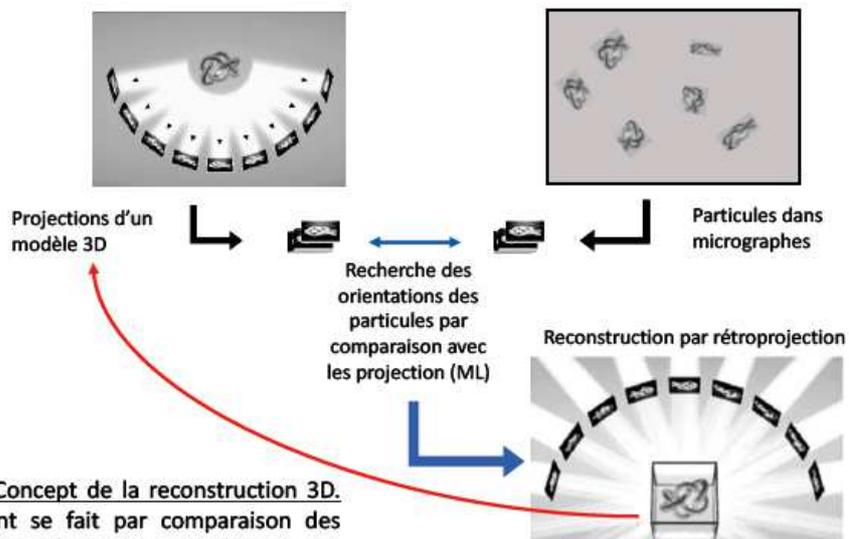
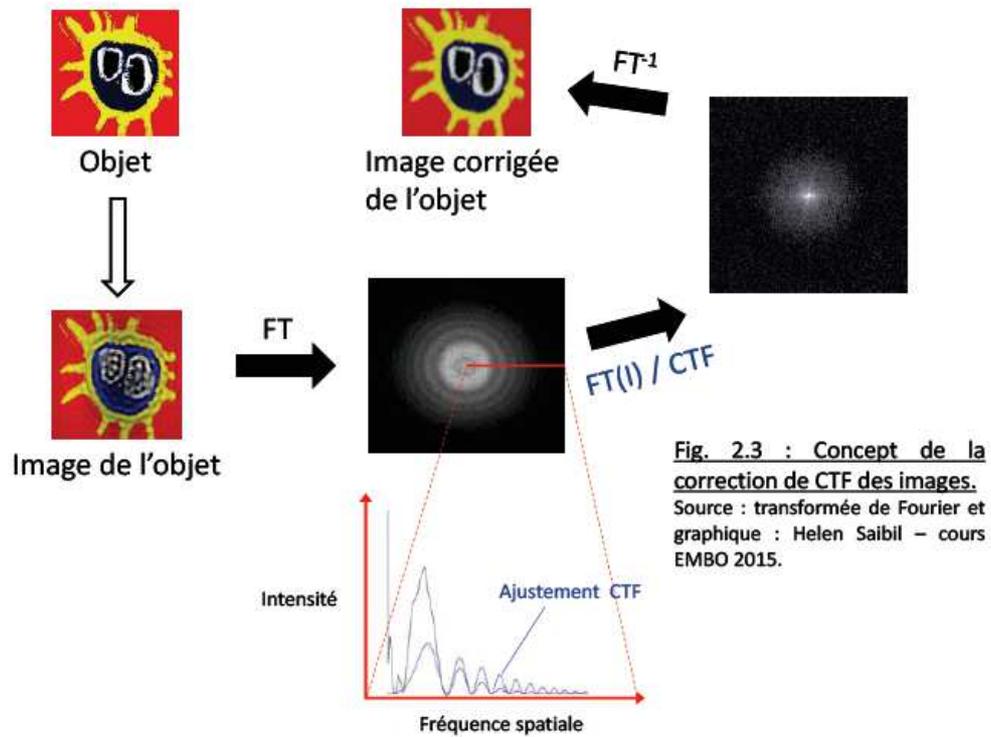


Fig. 2.4 : Concept de la reconstruction 3D.
 L'affinement se fait par comparaison des particules avec les projections du nouveau modèle (processus itératif).

2.2.4 Révolution apporté par les détecteurs à détection directe d'électrons

La détection des électrons se faisait avec des films et avait été partiellement remplacé par des détecteurs CCD à partir des années 1990. Les capteurs CCD permettent d'observer la qualité de l'échantillon sans attendre le développement des films, mais ont plusieurs inconvénients : ils génèrent notamment beaucoup de bruit quand les électrons ont des énergies supérieures à 100 keV, empêchant l'utilisation de voltages plus importants. De plus, le capteur CCD est obligatoirement couplé à une couche de phosphore et un système de fibres optiques, les électrons incidents génèrent des photons en frappant la couche de phosphore qui sont collectés par les fibres optiques et amenés jusqu'au détecteur CCD. Ce système réduit la résolution du détecteur, car il apporte une incertitude sur le trajet de l'électron dans le phosphore, sur le nombre de photons générés et sur le trajet des photons.

Depuis 2012, les détecteurs à détection directe sont commercialisés. Il s'agit de détecteurs CMOS (Complementary Metal Oxide Semi-conducteur) qui sont beaucoup moins sensibles à l'illumination électronique, ont des tailles de pixels plus petites et génèrent beaucoup moins de bruit que les capteurs CCD. De plus, le détecteur exporte ses données en continu, cela permet d'enregistrer des films qui sont utilisés pour corriger les mouvements des particules dans la glace pour reformer un micrographe. Depuis 2012, le nombre de structures obtenues par microscopie électronique à des résolutions sous-nanométrique et atomique augmente exponentiellement [McMullan 16].

2.2.5 Traitement de données

Une fois les micrographes collectés, le mouvement des particules et les CTFs corrigés, le processus de reconstruction 3D peut être amorcé.

La première étape est le picking ou boxing, qui consiste à repérer la position des particules d'intérêt sur chaque micrographe. Plusieurs logiciels sont en développement pour automatiser cette étape, mais ils ne sont pas adaptés à tous les échantillons, dont ceux étudiés dans ce travail de thèse, et ne seront pas décrits dans ce document. Le picking a ici été fait à la main.

Les particules sont ensuite extraites et classées en classes 2D. Ces classes correspondent à une « moyenne » d'images similaires. Un nombre de classes est fixé et les particules sont associées au sein d'une classe sur la base d'un score de corrélation croisée (CC) des intensités de pixels (différentes orientations sont corrélés pour chaque particule). Alternativement, et notamment implémenté dans le programme Relion [Scheres 12], la CC peut être remplacée par un calcul de Maximum de Vraisemblance (Maximum Likelihood – ML) où l'assignement à toute les classes

est associée à par une probabilité. Le calcul ML apporte une description statistique du bruit et empêche les fausses corrélations qui peuvent apparaître, en particulier avec un faible signal sur bruit [Sigworth 10]. Dans le processus proposé par Re-
lion, les classes 2D sont utilisées pour éliminer les particules trop hétérogènes ou de mauvaise qualité.

Les aspects mathématiques de la reconstruction 3D ne seront pas abordés dans ce document. Le principe de rétroprojection qui permet de revenir à un volume 3D à partir des projections obtenues au microscope s'appuie sur le théorème de la section centrale [Penczek 10]. Ce théorème dit que la transformée de Fourier d'une projection d'un objet 3D correspond à une section plane de la transformée de Fourier de l'objet, passant par le centre et perpendiculaire à l'orientation de la projection. Ainsi, l'obtention ou l'affinement d'un volume 3D peut se résumer à l'attribution d'une orientation à chaque image de particule. Lorsqu'un modèle existe déjà, ces angles sont déterminés par comparaison entre image et projection du modèle (processus itératif et mettant en jeu un calcul par ML) (Fig 2.4). En absence de modèle pré-existant, des orientations peuvent être déterminées expérimentalement pour aider le traitement de données (tomographie et inclinaison conique aléatoire) ou un modèle peut être généré *ab initio* (méthodes de reconstruction angulaire et de la ligne commune) [Voss 10].

Remarque : une hétérogénéité conformationnelle de l'échantillon peut amener à une dégradation locale de la résolution ou à un alignement difficile des particules, ces difficultés peuvent être en partie surmontées en utilisant des classifications 3D qui peuvent discriminer un nombre discret et défini d'états conformationnels ou par l'utilisation de masques [Bai 15].

2.3 Cristallographie

2.3.1 Généralités

La cristallographie est la méthode la plus utilisée pour obtenir des structures atomiques de molécules. Au 11 juillet 2017, 118154 structures avaient été résolues avec cette technique et déposées sur dans la PDB (Protein Data Base). Le nombre de structures déposées chaque année progresse et a été multiplié par deux au cours des 10 dernières années (5513 en 2006 contre 10002 en 2016). La résolution de structures de protéines par cristallographie se fait en surmontant 2 étapes limitantes : la cristallisation de la protéine d'intérêt et la résolution du problème de la phase.

2.3.2 Cristallogenèse

Au niveau moléculaire, un cristal correspond à un arrangement spatial précis de la matière. Il est possible d'obtenir des cristaux de protéines dans des conditions particulières que l'on peut schématiser sur un diagramme de phase. Pour certaines concentrations de protéines et de précipitants à une température donnée, la solution se trouve dans un état intermédiaire entre l'état soluble et la précipitation : l'état métastable (ou zone de supersaturation) dans lequel on distingue 2 zones. La zone de nucléation spontanée (ou homogène) correspond à la gamme de concentrations bordant la phase de précipitation où des cristaux peuvent apparaître. La zone de nucléation hétérogène (ou zone de croissance) se trouve entre la zone de nucléation spontanée et la limite de solubilité. Lorsqu'une solution de protéine atteint la zone de nucléation spontanée, un ou des cristaux se forment induisant la diminution de la concentration de protéine soluble. La croissance du cristal se poursuit dans la zone de nucléation hétérogène jusqu'à ce que la solution atteigne la courbe de solubilité (Fig. 2.5).

Le phénomène de cristallisation peut être affecté par la présence de ligands de la protéine, d'additifs et par la cinétique de déplacement à travers le diagramme de phase. Il existe plusieurs méthodes pour obtenir des cristaux de protéines :

- Cristallisation en batch : une solution de protéine est directement mélangée avec une quantité appropriée de précipitant. Cette méthode peut être utilisée pour obtenir des cristaux de grandes dimensions [Rayment 02].
- Cristallisation par diffusion de vapeur : méthode la plus répandue, qui consiste à faire une goutte de solution de protéine et de précipitant dans un puits fermé dont le fond est rempli de la solution de précipitant (la goutte est suspendue sur le couvercle du puits ou déposée sur une « corniche » aménagée spécialement). L'équilibration des concentrations par diffusion de vapeur entraîne une augmentation des concentrations de protéine et de précipitant dans la goutte.
- Cristallisation par dialyse : la solution de protéine est séparée de la solution de précipitant par une membrane semi-perméable. Il est possible d'ajuster la cinétique de cristallisation en contrôlant la concentration de la solution de précipitant [Bunick 00].

Les cristaux sont le plus souvent récupérés dans des boucles en nylon et congelés dans de l'azote liquide. La congélation permet de stocker les cristaux sur des périodes prolongées de temps et de limiter les dommages d'irradiations pendant la collecte des données. Si la solution de précipitant n'a pas les propriétés nécessaires pour permettre une congélation adéquate (vitrification de l'eau plutôt que formation de glace cristalline), le cristal est préalablement trempé dans une solution de cryo-protection.

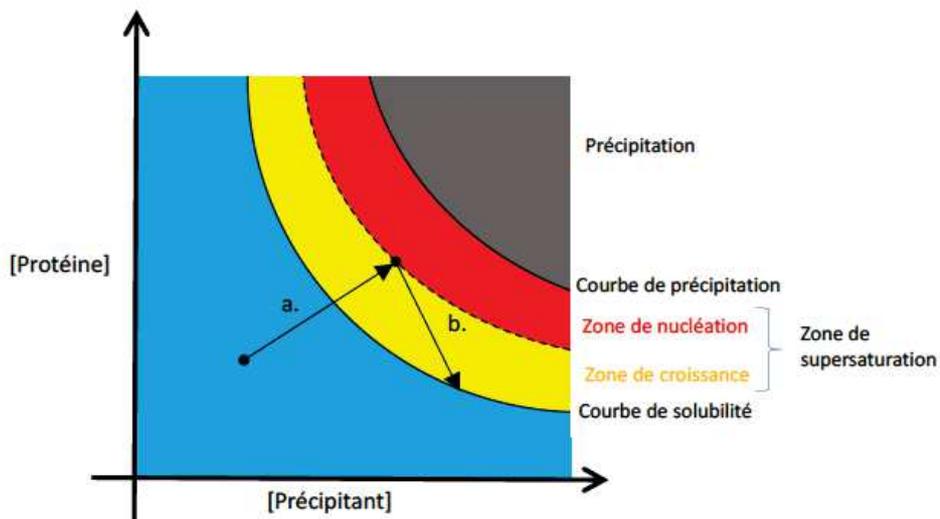


Figure 2.5 : Diagramme de phase de cristallisation. La nucléation d'un cristal ne peut se faire que dans la zone de nucléation, mais sa croissance peut se poursuivre jusqu'à la courbe de solubilité. a.b. Suivi des concentrations de protéine et de précipitant dans une goutte de cristallisation par diffusion de vapeur. Dans un premier temps les concentrations de protéine et de précipitant augmentent (a.) car l'eau diffuse vers la solution du puit qui est plus concentrée en précipitant. Lorsque la zone de nucléation est atteinte, un ou des cristaux de protéine se forme et croissent, diminuant la concentration de protéine soluble (b.).

2.3.3 Diffraction des rayons X

Du fait de l'organisation spatiale précise et répétée des molécules au sein d'un cristal, lors de l'exposition à des rayons X de longueur d'onde inférieure aux distances inter-atomiques, les rayonnements ré-émis par les atomes équivalents interfèrent négativement, sauf à des angles particuliers (définies par la loi de Bragg) où ils interfèrent positivement. Le diffractogramme du cristal est donc composé d'un nombre discret de points de diffraction.

Il est possible d'obtenir l'enveloppe électronique d'une protéine à partir de clichés de diffraction obtenus en exposant différentes orientations d'un même cristal à un faisceau de rayons X de longueur d'onde définie et une distance cristal-détecteur donnée. Le cristal est monté sur un goniomètre qui permet de le faire tourner. Des clichés de diffraction sont collectés pour différentes orientations. L'angle minimum sur lequel il est nécessaire de collecter des données pour résoudre la structure dépend de la symétrie du cristal.

Cependant, les clichés de diffraction correspondent à des images qui n'ont pas été refocalisées (il n'existe pas de lentilles pour rayons X) et ne contiennent plus l'information de phases. Le problème des phases peut être résolu de différentes façons :

- Par méthode directe, il est possible d'isoler des phases initiales grâce à un lien statistique entre les phases et les amplitudes. Ce type de résolution a cependant besoin de faire l'assomption que chaque atome forme un pic dans la densité électronique et nécessite donc des données de très haute résolution (1 \AA).
- Par remplacement moléculaire en utilisant les phases d'une structure homologue (généralement $< 35 \%$ d'identité [Abergel 13]).
- Par remplacement isomorphe, en comparant des jeux de données de cristaux de protéine natifs et marqués avec un atome lourd. Le petit nombre d'atomes ayant une forte contribution au signal de diffraction permet de résoudre les phases.
- Par diffusion anormale, en collectant un ou plusieurs jeux de données en faisant varier la longueur d'onde des rayons X autour et au niveau du seuil d'absorption de certains atomes aux propriétés de diffusion complexe (comme les lanthanides), il est possible d'exploiter les variations de comportements de ces atomes pour résoudre les phases d'un cristal.

2.4 Modélisation

En biologie structurale, les méthodes de simulation de dynamique moléculaire sont connues pour permettre d'accéder à des intermédiaires conformationnels et des

comportements dynamiques de molécules dans leur environnement. Ces approches apportent un autre angle de réflexion pour comprendre le mécanisme de réactions enzymatiques [Bernardi 14] ou de transports trans-membranaires [Musgaard 17] par exemple.

Des méthodes de modélisation peuvent également servir pour combiner des données de cristallographie et de microscopie électronique, en déterminant la structure à résolution atomique de protéines ou sous-unités individuelles, et la structure d'assemblages oligomériques ou multi-moléculaires par microscopie électronique, puis en combinant les différentes structures afin de déterminer la structure atomique de l'assemblage multi-moléculaire. Des méthodes automatisées d'ajustement corps rigide (rigid fitting) ont été développées au début des années 2000, remplaçant les approches manuelles d'ajustement d'une structure atomique dans une enveloppe à plus faible résolution par une comparaison quantifiable entre la densité observée et une densité calculée à partir d'un modèle atomique [Fabiola 05]. Cependant, les structures atomiques disponibles peuvent souvent arborer des différences notables avec la densité correspondante observée en microscopie : états conformationnels/oligomériques différents, artefacts de cristallisation, changement de conformation. . . Différents programmes d'ajustement flexible (flexible fitting) ont été développés pour surmonter ces difficultés, elles se basent soit sur une analyse des modes normaux, comme iMODfit (Lopez-Blanco, 2013) et NMFF (Tama, 2004), soit sur une approche de dynamique moléculaire, comme MDFF [Trabuco 08], MDfit [Whitford 11] ou FlexEM [Topf 08]. Ce dernier a été utilisé dans le cadre de l'étude de la protéine pb6 (cf. 2.4), il fonctionne en réalisant une itération de cycles au cours desquelles un ajustement corps rigide est réalisé sur la protéine entière, puis un recuit simulé est appliqué à des zones pré-déterminées par l'utilisateur. Cette approche a été choisie car la définition des zones rigides laissait peu de place au doute du fait de la qualité de l'ajustement corps rigide d'origine et permettait un gain sur le temps de calcul.

Les méthodes *in silico* sont également utilisées pour résoudre des structures à partir de données de RMN. La combinaison de RMN et d'enveloppes de microscopie électronique est également possible et sera discutée plus loin dans ce document (cf. 6.2.3).

Points clés

Résumé chapitre 2

- ❑ La microscopie électronique permet d'obtenir des structures de résolution intermédiaire, voire haute, sur de grand assemblages macromoléculaires.
- ❑ La cristallographie permet d'obtenir des structures hautes résolution d'éléments cristallisables.
- ❑ Les méthodes de modélisation permettent d'intégrer les résultats provenant de différentes techniques.

Tube de la queue de T5

"Boy, there's nothing worse
than an inscrutable omen."

(Bill Watterson)

Chapter 3

Tube de la queue du phage T5 : contexte & méthodes

3.1 Contexte

Cette première partie s'intéresse à la structure du tube de la queue du bactériophage T5.

La protéine pb6 forme le tube par un empilement hélicoïdal d'anneaux trimériques et orme un canal qui contient la protéine pb2. Cette dernière est éjectée lors de l'interaction avec l'hôte pour que le canal serve au passage de l'ADN.

Un des objectifs de ce travail est également d'élucider la transmission de l'information de fixation à la capsid. Il a été proposé que la protéine de tube assure cette fonction chez les *Siphoviridae* (cf. 1.4.3.1).

Ces résultats sont centrés sur l'étude de la protéine pb6 mais font aussi intervenir des échantillons plus longuement discutés dans la partie 6.2.3. Le chapitre 4 évoquera les résultats obtenus sur la protéine pb6 et sa polymérisation. Le chapitre 5 consiste en l'article Bacteriophage T5 tail tube structure suggests a trigger mechanism for Siphoviridae DNA ejection et présente les données structurales obtenues sur pb6 et le tube de T5. Enfin, le chapitre 6 présentera les résultats préliminaires d'une étude RMN de pb6.

Le travail de cristallographie de la protéine pb6, notamment pour l'obtention des phases, à fait l'objet de nombreux essais qui seront peu évoqués dans les résultats. Après l'obtention des phases avec une protéine marquée à la sélénométhionine, Sylvain Engilberge a continué à travailler sur des cristaux de protéine complexés avec des composés lanthanides. Ce travail a été intégré dans une publication est présentée en annexe : Crystallophore: a versatile lanthanide complex for protein crystallography combining nucleating effects, phasing properties, and luminescence.

Ci-après, le matériel & méthodes.

3.2 Matériel et méthodes

3.2.1 Production de pb6

3.2.1.1 Cultures en milieu riche

La protéine pb6 est surexprimée dans *Escherichia coli* BL21(DE3) transformé avec un plasmide pLIM13 conférant une résistance à la kanamycine et ayant intégré le gène codant pour la protéine pb6 avec un tag hexahistidine précédé d'un site de clivage pour la protéase TEV (*Tobacco etch Virus*) en C terminal.

Toutes les cultures de production de pb6 (milieux riche et minimum) sont réalisées en présence de 50 μM de kanamycine.

Les précultures en LB sontensemencées à partir d'un stock glycérol 25% (conservé à -80°C) ou de bactéries fraîchement transformées. Elles sont incubées sous agitation la nuit à 37°C . Les cultures en LB sontensemencées à une DO_{600} (densité optique à 600 nm) de 0,1 et incubées sous agitation une journée à 37°C .

3.2.1.2 Cultures en milieu minimum

Pour la préparation d'échantillons en milieu minimum, une préculture en LB est utilisée pour ensemencer à une DO_{600} de 0,1 une préculture en milieu M9 (50 mM Na_2HPO_4 ; 20 mM K_2HPO_4 ; 10 mM NaCl ; 20 mM NH_4Cl ; 2 mM MgSO_4 ; 100 μM CaCl_2 ; 2 g/l glucose). La culture M9 estensemencée à une DO_{600} de 0,1 et incubée sous agitation à 37°C . La production de protéines est induite avec 50 μM d'IPTG (isopropyl -D-1-thiogalactopyranoside) lorsque la DO_{600} est d'environ 0,8 et la culture est incubée sur la nuit à 37°C , sous agitation.

3.2.1.3 Production de protéines marquées

La production de pb6 marqué ^{13}C et ^{15}N est réalisée comme une culture en milieu de culture M9 mais en utilisant du glucose marqué ^{13}C (2 g/l) et du $^{15}\text{NH}_4\text{Cl}$ (1 g/l).

La production de pb6 sélénométhionilée est réalisée en milieu M9. Lorsque la DO_{600} atteint 0,6, 50 mg de thréonine, 50 mg de lysine, 50 mg de phénylalanine, 50 mg de leucine 25 mg d'isoleucine, 25 mg de valine et 30 mg de sélénométhionine sont ajoutés au milieu. 15 min plus tard, 50 μM d'IPTG sont ajouté et la culture est incubée sur la nuit à 37°C , sous agitation.

La production de pb6 marquée ^{13}C , ^{15}N et deutérée nécessite des étapes d'adaptation au deutérium : précultures en LB, en M9 puis en M9 50% deutérium et enfin en M9 100% deutérium, glucose ^{13}C et chlorure d'ammonium ^{15}N . L'ensemencement d'une préculture à une autre se fait à une DO_{600} de 0,2, par resolubilisation d'un

culot bactérien obtenu par centrifugation (10', 4000g, 4°C) du volume nécessaire de la préculture précédente. La culture en M9 100% deutéré et marqué $^{13}\text{C}/^{15}\text{N}$ est ensemencée/incubée/induite comme une culture en milieu M9.

La production de pb6 marquée ^{13}C , ^{15}N , deutérée et un marquage particulier des méthyles mais lors de la culture, on ajoute un précurseur de la valine et de la leucine 1h avant induction et un précurseur de l'isoleucine 20' avant induction. Pour un marquage ^{13}C , ^{15}N complet mais avec conservation de H des méthyles, le 2 (D3)methyl 1, 2, 3, 4 ($^{13}\text{C}4$) acetolactate fut utilisé comme précurseur des valine et leucine et le 2 hydroxy 2[1,2 $^{13}\text{C}2$ 1 D2]ethyl 3 oxo [1, 2, 3 $^{13}\text{C}3$]4 [D3]butanoate comme précurseur de l'isoleucine. Pour un marquage ^{13}C , ^{15}N complet mais avec les chaines latérales des valine, leucines et isoleucines marquées ^{13}C uniquement au niveau des méthyles et conservation de leurs H, le 2 (D3)methyl 2, 4 ($^{13}\text{C}2$) acetolactate fut utilisé comme précurseur des valine et leucine et l'acide 2 oxobutyrique 4 (^{13}C), 3, 3(D2) comme précurseur de l'isoleucine.

3.2.2 Purification et caractérisation de pb6 monomère

3.2.2.1 Lyse des cellules

La culture est centrifugé 20', 5000 rpm (JLA 8.1), 4 °C. Le culot bactérien est resuspendu dans un tampon de lyse (50 ml de 50 mM Tris pH8 , 100 mM NaCl si non précisé) et la lyse réalisée par 6 passages au microfluidizer (M-110 P Microfluidics) à 13 Kpsi. Les fractions insolubles et solubles sont séparées par ultracentrifugation 30', 30000 rpm (Ti45), 4 °C.

3.2.2.2 Chromatographie d'affinité au nickel

Les chromatographies sont réalisées à l'aide d'un système de chromatographie en phase liquide en pressions intermédiaires, un BioLogic DuoFlowTM (BIO-RAD). La purification par chromatographie d'affinité au nickel est réalisée sur une colonne HiTrapTM Chelating HP 5ml (GE Healthcare). La fraction soluble du lysat bactérien est supplémentée en imidazole (15 mM final) et chargée sur la colonne d'affinité préalablement équilibrée puis lavée avec un tampon 20 mM Tris pH8, 100 mM NaCl, 15 mM imidazole. La protéine est éluée par chargement d'un tampon 20 mM Tris pH8, 300 mM imidazole.

3.2.2.3 Chromatographie échangeuse d'ions

La purification par chromatographie échangeuse d'anions est réalisée sur une co- lonne HiTrapTM Q HP 5 ml ou 1ml (GE Healthcare). L'éluion de la chromatographie d'affinité est diluée dans du 20 mM Tris pH8 de façon à baisser la

conductivité de la solution à 5 mS/cm (dilution d'un facteur 5 habituellement). L'échantillon est chargé sur la colonne préalablement équilibrée puis lavée avec une solution 20 mM Tris pH8. La protéine est éluée par un gradient linéaire de conductivité obtenu par ajout de proportions croissantes de 20 mM Tris pH8, 1 M NaCl.

3.2.2.4 Dessalage

La protéine éluée de la colonne échangeuse d'anions est concentrée de façon à être contenue dans un volume inférieur à 0,5 ml (cf. 3.2.2.6). L'échantillon est alors injecté sur une colonne HiTrapTM Desalting (GE Healthcare) préalablement équilibré puis rincé avec un tampon (20 mM Tris pH8 si non précisé).

3.2.2.5 Electrophorèse sur gel de polyacrylamide laurylsulfate de sodium (SDS-PAGE)

Les échantillons à analyser sont mélangés avec du tampon de charge (0,1 M Tris pH 6,8, 4% SDS, 20% glycerol, 0,05% bleu de bromophenol final) et chauffés 90'' à 90°. Les échantillons sont déposés dans les puits d'un Gel de PolyAcrylamide pour Electrophorèse en présence de DodecylSulfate de Sodium (SDS-PAGE) 12% acrylamide installé dans un système d'électrophorèse vertical 1-D Mini-PROTEAN Tetra Cell (BIO-RAD) selon les recommandations constructeur. La migration électrophorétique est réalisée par application d'un courant électrique de 220 mV pendant 40' au système rempli d'un tampon de migration (0,19 M glycine, 25 mM Tris base, 0,1% SDS). Le gel est coloré au bleu de coomassie par incubation 15-20' dans un bain 43, 2% ethanol, 3% acide acétique, 1,82 mM bleu brillant R puis décoloré soit par passage au four à micro-ondes dans de l'eau milliQTM, soit par incubation 40-50' dans un bain 30% ethanol, 10% acide acétique.

3.2.2.6 Spectroscopie ultraviolet/visible et concentration

Les spectres d'absorptions 240-500 nm des différents échantillons sont obtenus en utilisant un spectrophotomètre Cary® 50 UV-Vis (Varian) ou Nanodrop 2000c (Thermo Scientific) et leurs logiciels associés et en respectant les recommandations du constructeur. La concentration d'un échantillon est réalisée par centrifugation sur Amicon® Ultra 15 ml, 4 ml ou 0,5 ml avec une limite nominale de séparation de poids moléculaire de 30 kDa. Les centrifugations sont faites aux vitesses recommandées par le constructeur pendant des temps allant de 2 à 6' à 4°. Entre 2 centrifugations, le rétentat est mélangé par pipetage et est analysé par spectroscopie UV-Visible pour confirmer la concentration en protéine et l'absence de précipitation. Le nombre de centrifugations et la taille du concentrateur sont adaptés au volume d'échantillon et la concentration visée.

3.2.2.7 Filtration sur gel analytique

L'analyse par chromatographie d'exclusion stérique est réalisée sur une colonne SuperdexTM 200 5/150 GL (GE Healthcare). 10 µl d'échantillon sont chargés sur la colonne préalablement équilibrée puis rincée en 20 mM Tris pH8, 150 mM NaCl.

3.2.3 Purification et caractérisation des fibres de pb6

3.2.3.1 Resolubilisation

La fraction insoluble du lysat bactérien (cf. 3.2.2.1) est resolubilisée par pipetage et homogénéisée à l'aide d'un potter dans un tampon 50 mM Tris pH 8, 100 mM NaCl, 0,5% Triton X-100, 1 mM EDTA, 100 µg/ml lysosyme. Cette solution est incubée 1h à 37° sous agitation.

3.2.3.2 Gradient de sucrose

Les fibres resolubilisées sont déposées sur un gradient de sucrose de 4 coussins : 50, 40, 35 et 35% , 20 mM Tris pH 8, 100 mM NaCl, 0,05% Triton X-100. Le gradient est centrifugé 30', 10000 rpm (SW41), 4°. Le culot est repris dans un tampon (20 mM Tris pH 8, 100 mM NaCl si non précisé) et centrifugé 5', 2000g, 4° puis resolubilisé dans le même tampon afin d'éliminer les traces de sucres et de détergent.

3.2.3.3 Microscopie électronique à transmission (MET) – Coloration négative

Des grilles de microscopie électronique sont préparées en utilisant le Sodium Silico Tungstate par et selon le protocole de la plateforme contrôle qualité par microscopie électronique de l'IBS. Les grilles sont observées sur un FEI Tecnai12 120 kV LaB6 équipé d'une caméra GATAN Orius 1000.

3.2.4 Tests de polymérisation de pb6

3.2.4.1 Filtration sur gel analytique

Des échantillons de pb6 monomère purifiés sont placés dans différentes conditions de pH, de concentrations salines et en présence de différents ions bivalents par dessalage (cf. 3.2.2.4) ou dilution et concentration (cf. 3.2.2.6) dans un nouveau tampon. Après un temps variable d'incubation, les échantillons sont analysés par chromatographie d'exclusion stérique (cf. 3.2.2.7)

3.2.4.2 Fluorimétrie

Les échantillons de pb6, monomères ou fibres, sont incubés dans des conditions variables. 25 μ M final de Thioflavine T (Sigma-Aldrich) sont ajoutés avant de mesurer la fluorescence ($\text{em} = 482 \text{ nm}$; $\text{ex} = 440 \text{ nm}$). Les mesures de fluorimétrie sont réalisées à température ambiante dans des plaques 96 puits par un CLARIOstar® (BMG LABTECH).

3.2.5 Mutants de pb6

3.2.5.1 Obtention des plasmides des mutants

Les vecteurs d'expression des mutants de pb6 sont obtenus par mutation ponctuelle du plasmide pLIM13 dans lequel le gène de pb6 est inséré. Le *QuickChange Lightning Multi Site-directed Mutagenesis Kit* (Agilent Technologies) est utilisé selon les recommandations du fournisseur, les amorces utilisées sont décrites dans le chapitre 5 (cf. 5.2). La qualité des plasmides obtenus est contrôlée par séquençage (réalisé par un prestataire : GENEWIZ).

3.2.6 Cristallogénèse de pb6

Les cristaux de pb6 ont été obtenus en gouttes suspendues sur des plaques 48 puits siliconés à 4°C. La solution de puits se compose de PEG (PolyEthylène Glycol) 3350, d'HEPES pH 7,5, de lithium sulfate et de chlorure de manganèse dans des quantités variables. Les gouttes sont formées sur une lamelle en verre par un mélange de 1 μ l de solution de puits, de 1 μ l de solution 5 à 15 mg/ml de pb6 et, dans des essais de co-cristallisation, de 1 μ l de solution d'élément susceptible de permettre le phasage expérimental. Les lamelles sont placées sur les puits, gouttes coté intérieur, le tout scellé par le silicone. Les cristaux de pb6 ont été manipulés à l'aide de boucles en nylon CrystalCap™ HT (Hampton Research) de 0,1 à 0,5 mm. Différentes solutions de trempages avec des agents de phasage ont été testées : composés lanthanides 0,01 à 0,3 M sur des temps allant de quelques secondes à une nuit; NaCsI et NiSO₄ testés de 0,1 à 1 M, pendant quelques secondes à quelques minutes. La cryoprotection a été réalisée par trempage dans des solutions identiques à la solution du puit d'origine du cristal plus 20% glycérol ou 25% PEG 400. Une cryoprotection supplémentaire/alternative par trempage dans de l'huile de paraffine ou de paratone a également été utilisée. Une fois le cristal trempé dans une/des solutions de cryoprotection, il est congelé puis conservé dans de l'azote liquide.

3.2.7 Cristallographie de pb6

3.2.7.1 Collecte des données de diffraction de rayons X

Les clichés de diffraction ont été obtenus sur FIP-BM30A (French beamline for Investigation of Proteins – Bending Magnet 30 A, ligne de lumière à l'ESRF - Grenoble) par exposition de cristaux de pb6 congelés à des faisceaux de rayons X d'énergies variable, les clichés de diffraction sont enregistrés par un détecteur CCD ADSC Q315r. Les mesures ont été faites en utilisant des énergies de faisceau différentes sur 180 ou 360° de rotation avec un cliché tous les demi-degrés obtenu avec un temps d'exposition de 15 à 60 s.

3.2.7.2 Obtention d'une première carte de densité

La structure de pb6 a été résolue *de novo* par combinaison de données MAD (*Multi wavelength Anomalous Diffraction*, obtenues sur un cristal de protéine marquée avec de la sélénométhionine) et de données obtenues sur un cristal natif. Les statistiques des jeux de données sont exposées dans le chapitre 5 (cf. 5.2). La position des atomes de sélénium dans l'unité asymétrique a été réalisée avec le programme SHELXD [Sheldrick 10]. L'affinement de la position des atomes lourds et le phasage initial a été réalisé avec le programme Sharp [Bricogne 03]. Les phases de Sharp ont été améliorées par modification de densité en utilisant le programme SOLOMON [Abrahams 98] (FoM = 0,82).

3.2.7.3 Résolution et affinement de la structure de pb6

La construction du modèle atomique fut commencée de façon automatique grâce au logiciel BUCCANEER [Cowtan 06] et complétée manuellement avec COOT [Emsley 10]. L'affinement du modèle à été amorcé avec Phoenix [Adams 10] et terminé avec BUSTER (2016 BUSTER version 2.10.3 Cambridge, United Kingdom: Global Phasing Ltd). L'optimisation du modèle a été réalisée par itération de cycle d'affinement/construction de modèle. Une NCS (Non crystallographic Symmetry) et des groupes TLS (Translation/Libration/Screw) furent appliqués et les hydrogènes (hors molécules du solvant) furent ajoutés en fin d'affinement. Les statistiques de l'affinement sont regroupées dans un tableau dans la section 5.2. Les coordonnées atomiques du modèle de pb6 et les facteurs de structures mesurés ont été déposés sur la PDB (Protein Data Bank) sous la référence 5NGJ.

3.2.8 Cryo-microscopie électronique

3.2.8.1 Collecte d'images de fibres

Les échantillons de fibres et de queues de T5 (en interaction ou non avec FhuA) ont été préparés en suivant le même protocole. 3 μL d'échantillon sont déposés sur des grilles QUANTIFOIL® R 2/1 (Quantioil Micro Tools GmbH, Allemagne) préalablement exposées à du plasma. Les grilles ont été congelées par plongeon dans de l'éthane liquide grâce à un Vitrobot Mark II (FEI, Pays-Bas). Les échantillons ont été observés avec un microscope FEI Polara 300kV. Les micrographes ont été enregistrés avec une caméra à détection direct K2 Summit (Gatan Inc., Etats-Unis) en utilisant le mode super resolution counting. Les films ont été enregistrés avec une exposition totale de 4 secondes et 100 ms par images, résultant en des films de 40 images avec une exposition totale d'environ 40 électrons/Å². Le grossissement utilisé pour les fibres de pb6 et les queues de phage incubées avec FhuA était de 23000 (0,82 Å/pixel au niveau de la caméra) alors qu'il était de 15500 pour les queues de phage (1,24 Å/pixel).

3.2.8.2 Obtention du modèle initial

Un modèle initial à basse résolution a été obtenu par rétro-projection du tube d'une unique queue de phage (6 répétitions) et l'application d'une symétrie 3 avec le programme SPIDER [Frank 96]. Ce modèle initial a été affiné avec une méthode classique de corrélation des projections [Franzetti 02] : après correction de la CTF (Contrast Transfer Function) des images, des segments de tubes ont été pickés dans des boîtes de 128x128 pixels et superposés avec un chevauchement de $\frac{3}{4}$ de la taille de boîte. Les superpositions obtenues (1000 images) ont été alignées sur des projections du modèle et moyennées en différentes classes. Les classes ont été rétro-projectées pour former un nouveau modèle sur lequel a été également appliqué une symétrie 3. Après 10 cycles d'affinement le modèle initial était suffisamment bon pour être utilisé pour la suite du traitement de données.

3.2.8.3 Obtention des structures de tube

La dérive des images a été corrigée en utilisant le programme Digital Micrograph (Gatan Inc., USA) pour les queues de T5 ou le programme MotionCorr [Li 13] pour les queues de T5 incubées avec FhuA et les fibres de pb6. Pour ces 2 derniers échantillons, la taille de pixel fut ramenée à 1,64 Å/pixel. La même procédure fut ensuite appliquée aux trois échantillons. Les segments de tubes ont d'abord été pickés en utilisant le programme boxer [Ludtke 99]. La distance entre chaque pick a été fixée à la distance estimée séparant chaque anneau (40 Å). Les coordonnées ont ensuite été importées sur Relion [Scheres 12]. Les paramètres de CTF furent

estimés avec le programme CTFFIND4 [Rohou 15]. Le programme Relion a été utilisé pour réaliser une classification 2D et retirer du jeux de données les particules ne s’alignant pas avec les autres. Un premier affinement 3D fut réalisé en utilisant le modèle initial (cf. 3.2.8.2) comme référence, une symétrie C3 fut imposée. Comme chaque segment a été extrait avec une distance correspondant à la distance inter-anneau, aucune méthode ne fut appliquée pour imposer la forme hélicoïdale de l’assemblage. Les tubes « pleins » et « vides » furent séparés par des classifications 3D successives. Les classes 3D des queues de phages et des queues de phages incubées avec FhuA furent réparties en 2 groupes : d’une part celles avec une densité visible dans le lumen, d’autre part celles sans. Enfin, un affinement 3D et un post-processing (avec application d’un masque) fut réalisé sur 24243 segments de tube « plein » (queue de T5) résultant en une structure de 6,2 Å ; sur 22704 segments de tube « vide » (queue de T5 incubées avec FhuA) résultant en un structure à 5,8 Å de résolution ; et sur 12244 segments de tube de fibre de pb6 résultant en une structure à 8,8 Å de résolution. Tableau détaillé des statistique regroupées dans un tableau dans la section 5.2. La visualisation de la carte, la segmentation et le fitting ont été réalisé à l’aide du programme CHIMERA [Pettersen 04].

3.2.9 Modélisation

La structure cristallographique de la protéine pb6 a d’abord été manuellement placée dans la densité électronique de la carte « combinée » en utilisant le logiciel CHIMERA [Pettersen 04]. Le rigid-fit a ensuite été affiné en utilisant l’option *fit in density*. En reproduisant ce protocole sur différentes positions du monomère, il est possible de construire un premier modèle du tube. Cependant, bien que le cœur de la protéine soit tout à fait en accord avec les densités de microscopie électronique, la plupart des zones périphériques, impliquées dans les interactions formant le tube, ne s’ajustent pas correctement dans les densités. De plus, une large boucle est absente de la structure cristallographique.

Une version tronquée du fichier pdb de pb6 (où les zones ne rentrant pas dans les densités électroniques ont été retirées) a été utilisée pour segmenter la densité du tube de façon à reconstituer la densité correspondant à un seul monomère de pb6 au sein de l’assemblage. Cette opération a été réalisée en utilisant le logiciel Segger [Pintilie 10]. La densité d’un monomère a ensuite été utilisée pour reconstruire la boucle manquante manuellement sous COOT [Emsley 10]. Afin de mieux modéliser ces zones d’interactions, elles ont été soumises à un recuit-simulé en utilisant le logiciel FlexEM [Topf 08]. Ce recuit-simulé a été fait en apportant des contraintes supplémentaires au programme : la densité électronique d’un monomère et la définition de parties rigides dans le modèle atomique (non soumises au recuit-simulé). La définition des parties rigides a été faite sur le serveur

RIBFIND [Pandurangan 12], le fichier a ensuite été édité manuellement pour permettre plus de flexibilité (détails en annexe). Afin de renforcer la confiance dans le modèle, plusieurs conformations de départ ont été construites manuellement sous COOT et soumises au recuit-simulé, la validité des solutions obtenues a été évaluée sur la base de la convergence des différentes itérations et de la comparaison avec les structures du tube de T4 et du tube de la pyocine R.

Chapter 4

Tube de la queue du phage T5 : données biochimiques sur pb6

Les données structurales obtenus sur la protéine pb6 seront présentés par le biais de l'article « Bacteriophage T5 tail tube structure suggests a trigger mechanism for Siphoviridae DNA ejection » en révision dans le journal Nature Communications.

L'article sera précédé de la description des purifications de pb6 sous ses formes monomériques et les investigations sur les conditions de polymérisation de cette protéine. L'article sera également précédé d'un résumé de ses principaux résultats et suivi de conclusions.

Enfin, des travaux sur l'analyse RMN de la protéine pb6 sous ses formes de monomère et de fibre seront présentés.

4.1 Purification de pb6

Après lyse des cellules surexprimant pb6, la protéine peut être isolée et purifiée sous deux formes, une forme soluble et monomérique et une forme polymérisée en longues fibres. Ces deux types d'échantillons ont été produits et purifiés dans différentes conditions pour servir à trois objectifs :

- La cristallographie de la protéine pb6 monomérique.
- L'étude par cryo-microscopie électronique des fibres polymérisées de pb6.
- L'étude des fibres de pb6 en RMN du solide et de pb6 monomère par RMN en solution.

La protéine pb6 monomérique produite en LB est purifiée sur une colonne d'affinité puis une colonne échangeuse d'ions (élution à 34 mS/cm), avec un rendement de 20 mg de protéine par litre de culture et une pureté > 80% (Fig. 4.1-a). Ce type d'échantillon a principalement servi à obtenir des cristaux natifs.

Lorsque la protéine pb6 est produite en milieu minimum, le rendement diminue à 4 mg par litre de culture et la pureté à <50%. Ce rendement beaucoup moins satisfaisant s'explique par des contaminants supplémentaires (à 30 et 15 kDa, voir Fig. 4.1), mais aussi en partie par des quantités relatives de protéines solubles plus faibles que lorsque la protéine est produite en LB : le ratio de protéine dans la partie soluble du lysat passe de ~ 40% à ~ 30% (Fig. 4.1-b). La production de protéine en milieu minimum étant nécessaire pour réaliser des marquages (sélénométhionines pour le phasage en cristallographie ; ^{13}C , ^{15}N et deutérium pour la RMN), les conditions de polymérisation de pb6 ont été étudiées pour améliorer la production de ce type d'échantillon (cf. 4.2.1).

Les fibres de pb6 sont purifiées à partir de la fraction insoluble du lysat grâce à un gradient de sucrose après une incubation en présence de détergent/EDTA/lysosyme (Fig. 4.2). Ces fibres sont préparées pour être observées en microscopie électronique et analysées en RMN du solide. L'objectif est de déterminer la structure de ces fibres par cryo-microscopie électronique pour comparer leur structure à celles des tubes de queue de phage, ainsi que de faire une étude comparative et intégrative, notamment en dynamique, entre monomère et fibre basée sur la RMN (cf. 6).

La stabilité de la protéine pb6 a été évaluée par mesure de sa température de fusion (T_m). On observe un gain important de la stabilité de la protéine quand elle est sous forme de fibres : la T_m passe de 50°C sous forme monomère à 81°C sous forme de fibres. Il est intéressant de noter que la T_m de pb6 au sein du complexe de la queue de phage (cf. 6.2.3) est encore supérieure, aux alentours de 90°C.

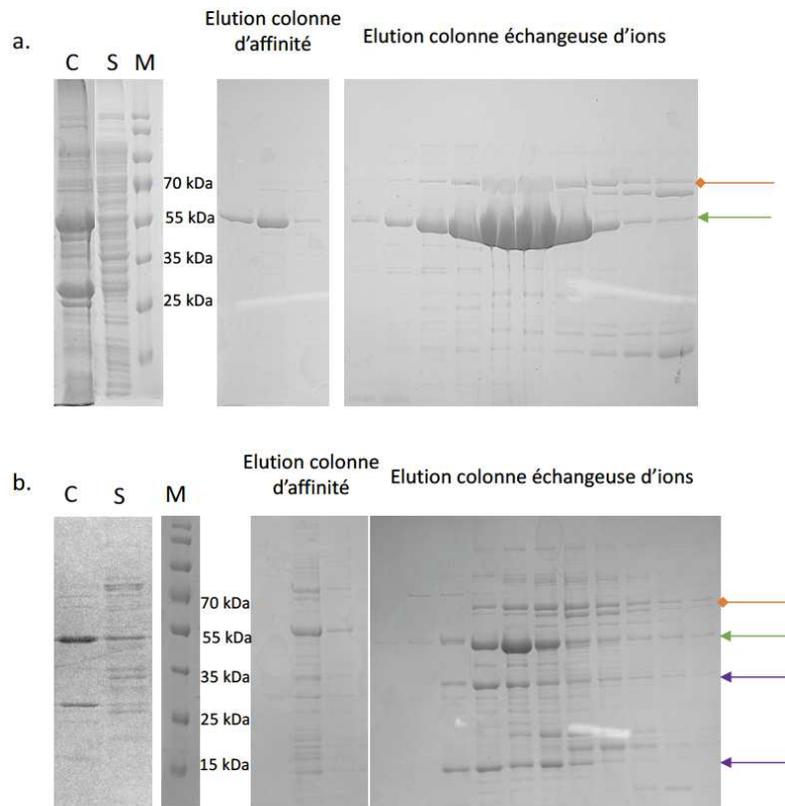


Fig 4.1 : Purifications de pb6 monomères, comparaison de l'expression en milieu minimum et en milieu riche.

Gels de la purification de pb6 à partir d'une culture LB (a.) ou M9 (b.). C = fraction insoluble du lysat ; S = fraction soluble du lysat ; M = marqueur de poids moléculaire. La protéine pb6 migre au même niveau que la bande à 55 kDa du marqueur (flèche verte). Dans les 2 purifications on observe un contaminant d'environ 70 kDa (flèche orange). A partir de la culture en milieu minimum, le rendement de la purification est moins bon et on observe de nouveaux contaminants de poids moléculaire plus bas (flèches violettes).

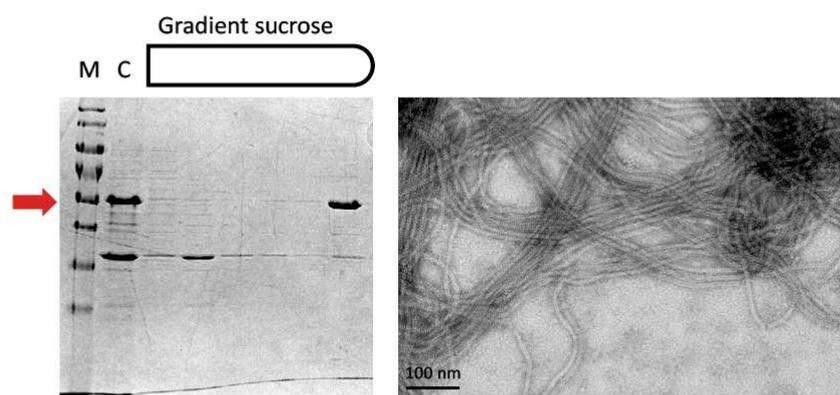


Fig 4.2 : Purifications de fibres de pb6.

C = lysat insoluble, M = marqueur de poids moléculaire. Gel de purification des fibres de pb6 sur gradient de sucrose par centrifugation différentielle et cliché de microscopie électronique à coloration négative.

4.2 Polymérisation de pb6

Valentin Sencio, stagiaire M1 pendant 3 mois au cours de l'été 2015, s'est beaucoup investi, sous ma direction, dans la production d'une grande partie des résultats présentés dans ce paragraphe.

4.2.1 Polymérisation initiale à la lyse des cellules

Comme indiqué en Matériel et Méthodes, lors de la lyse des cellules surexprimant pb6, le jus de lysat se gélifie. L'observation du culot de lyse montre de longues fibres de plusieurs micromètres de long. L'homogénéisation du lysat gélifié conduit à une solution à nouveau fluide, qui ne regélifie pas. Cela suggère que la formation des fibres et surtout leur réticulation se fait après la lyse. Cette hypothèse est confirmée par le criblage de différents tampons de lyse, faisant varier le pH, la concentration en sels et la présence d'ions divalents. De grandes variations du ratio de protéine soluble : insoluble ont été observé (Fig. 4.3-a), l'absence totale de sels dans le tampon de lyse permettant notamment d'obtenir la quasi-totalité de la protéine sous forme de monomère soluble. La même observation a été faite pour la lyse de cellules cultivées en milieu minimum (Fig. 4.3-b).

Cette variation du ratio protéine soluble : fibre selon le tampon de lyse suggère fortement que la polymérisation de la protéine ne se produit pas dans les bactéries mais au moment de la lyse. Cela est cohérent avec ce que nous savons sur l'assemblage des phages, qui est un processus minutieusement contrôlé, où la polymérisation et l'assemblage d'une protéine ne se fait qu'en présence de ses partenaires [Veesler 11]. D'un point de vue évolutif, il est important pour le virus de conserver des protéines structurales capable de s'auto-assembler pour une faible énergie d'activation mais ne formant pas d'oligomères indépendants du virion. Remarque : pour un virus eucaryote, la formation de particules virales non infectieuses et/ou incomplètes peuvent constituer un camouflage au système immunitaire de l'hôte.

En l'absence de sel, il a été confirmé grâce à des gradients de sucrose que la protéine restante dans la fraction insoluble était précipitée et non polymérisée sous forme de fibres (Fig. 4.4).

Le déclencheur de la polymérisation lors de la lyse a également été recherché, en utilisant différentes méthodes pour lyser les cellules : en plus de la lyse classiquement utilisée pour pb6, par le Microfluidizer (passage sous pression dans un capillaire augmentant brusquement de diamètre), la lyse chimique (combinaison de lysozyme et de détergents) et par sonication ont été testées. Aucune conclusion claire n'a pu être tirée de ces expériences.

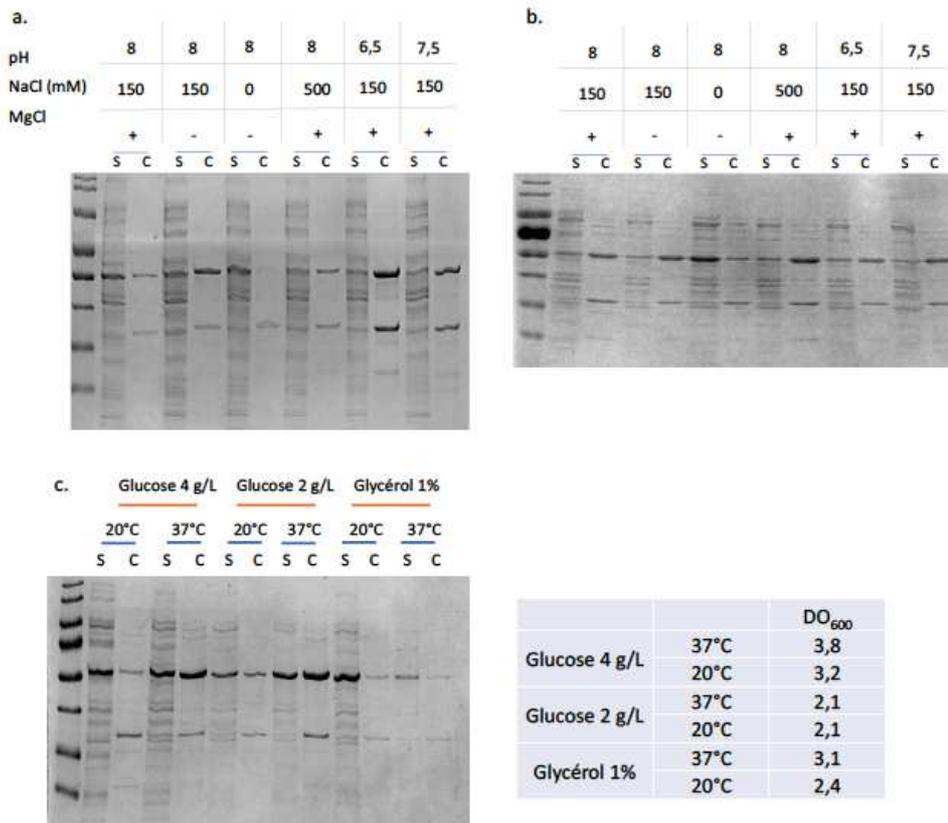


Figure 4.3 : Effet des conditions de lyse et de culture sur la polymérisation de pb6. Effet de la composition du tampon de lyse sur l'état oligomérique de pb6 produit en milieu riche (a.) ou en M9 (b.). L'absence de sels dans le tampon de lyse inhibe la formation de fibres. c. Lyse avec un tampon favorisant la présence de pb6 sous forme monomérique de bactéries cultivées dans différentes conditions (milieu M9 avec variation de la source de carbone et de la température) et turbidité avant lyse des cultures testées (tableau). Le ratio de protéine soluble sur protéine insoluble est moins bon à 37°C mais la quantité totale de pb6 est plus importante.

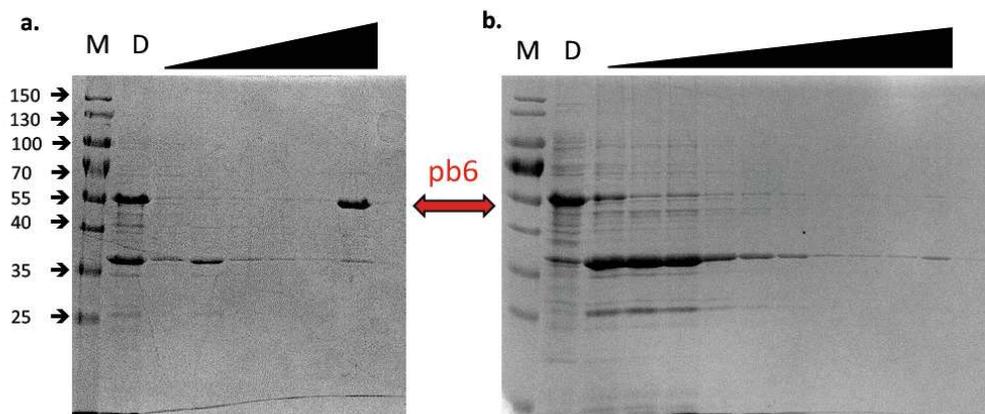


Figure 4.4 : Comparaison du gradient de sucrose de pb6 insoluble dans des conditions favorisant ou non la polymérisation. D = fraction déposée sur le gradient. **a.** Purification sur gradient de sucrose à partir d'e la fraction insoluble d'un lysat ou la lyse a été réalisée dans des conditions standard (150 mM NaCl). **b.** Purification sur gradient de sucrose à partir d'e la fraction insoluble d'un lysat ou la lyse a été réalisée sans sel dans le tampon de lyse. La protéine pb6 reste dans la partie haute du gradient dans les mêmes conditions, indiquant que la protéine dans le culot n'est probablement polymérisée sous forme de fibres mais agrégée.

4.2.2 Polymérisation de monomère purifié

Différentes conditions ont été criblées (pH, sels, ions divalent, température) pour essayer d'induire la formation de fibres à partir du monomère soluble purifié. Deux approches ont été utilisées pour suivre la formation des fibres :

- Analyse d'échantillons de pb6 en chromatographie d'exclusion de taille.
- Analyse en lecteur de plaques d'échantillons de pb6 en présence de la sonde fluorescente thioflavine T.

4.2.2.1 Filtration sur gel

Les profils d'élution de pb6 dans différentes conditions sont présentés dans la Fig. 4.5-a. Trois courbes se démarquent : celle de pb6 incubé dans un milieu très basique (courbe bleu), celle de pb6 incubé à pH 2,5 (courbe orange) et celle de pb6 dans un milieu faiblement acide (courbe grise, dont le pic au niveau du volume mort est partiellement masqué par les superpositions de courbes). L'analyse sur gel (non présenté) et en microscopie électronique à coloration négative (Fig. 4.5-b) révèle que l'échantillon en milieu basique est dégradé et forme des précipités de protéine probablement dénaturée, mais que les deux échantillons en milieu acide contiennent des fibres. Il est cependant étrange que la formation de fibres ne se fasse pas dans des proportions plus importantes, étant donné le comportement de la protéine au cours de la lyse. Nous nous serions attendus à ce qu'une fois un noyau de polymère formé, la totalité/majorité de la protéine s'assemble sous forme de tube.

4.2.2.2 Expériences de fluorescence

La thioflavine T est une sonde fluorescente couramment utilisée pour suivre la polymérisation de fibres amyloïdes. Elle émet un signal à 482 nm lorsqu'elle est excitée à 440 nm et fixée à un feuillet. Du fait de l'augmentation de la stabilité de pb6 sous sa forme polymérisée et de la structure du monomère de la TTP du phage lambda [Pell 09a] où tous les segments en feuillets ne semblent pas formés en solution, l'hypothèse que des feuillets se stabiliseraient lors de la formation du tube nous a amené à tester cette molécule comme témoin de la formation de fibres.

Les données obtenues suggèrent que l'effet du pH et du sel se cumulent et provoquent une polymérisation importante de la protéine (Fig. 4.5-c). D'autres conditions (température, ions divalents) ont été testées sans effets notables, le KCl semble avoir un effet similaire au NaCl mais moins marqué (données non présentées).

Il est d'intérêt de noter que lors des expériences de tests de tampon de lyse (Fig. 4.3-a), l'utilisation d'un tampon acide avait significativement augmenté la quantité de protéine dans la fraction insoluble.

Contrairement à la TTP de SPP1 qui s'assemble sous forme de tube au cours du temps [Langlois 15], pb6 ne polymérise en dehors de la voie d'assemblage de la queue que dans des conditions particulières, des conditions extrêmes de sel et de pH semble induire la polymérisation massive de la protéine mais de tels conditions ne sont pas atteintes pendant la lyse. La concentration de la protéine a été testée jusqu'à 100 mg/mL sans observation de polymérisation [Tissot 13]. Il n'est pas impossible que le phénomène de polymérisation soit influer par d'autres éléments comme la pression, la viscosité de la solution ou la présence de protéines bactériennes.

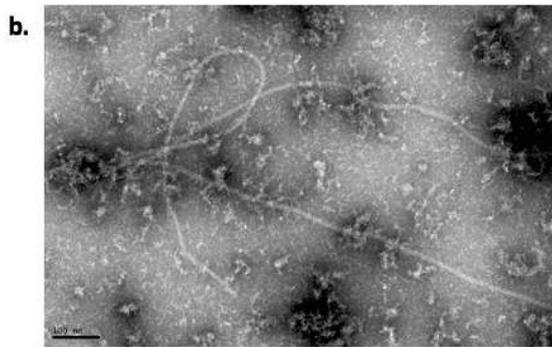
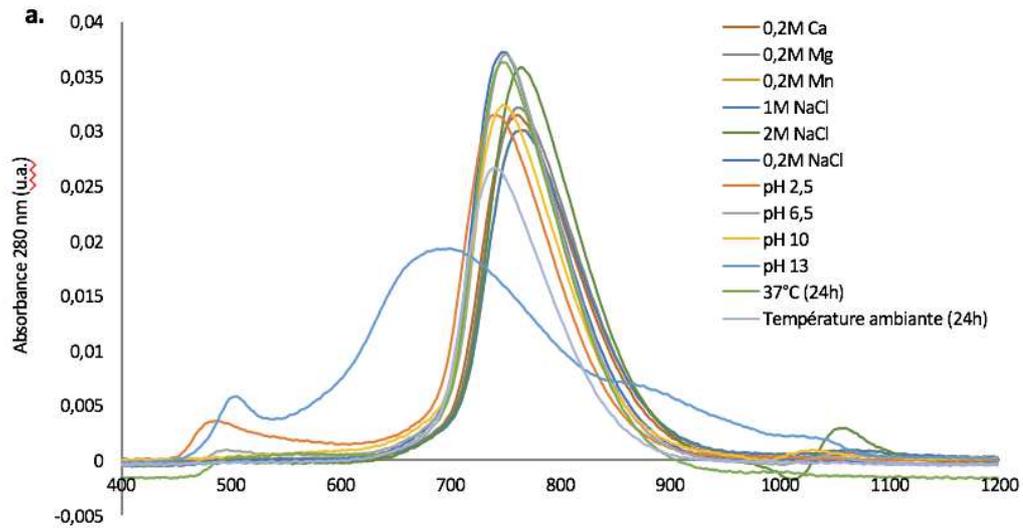
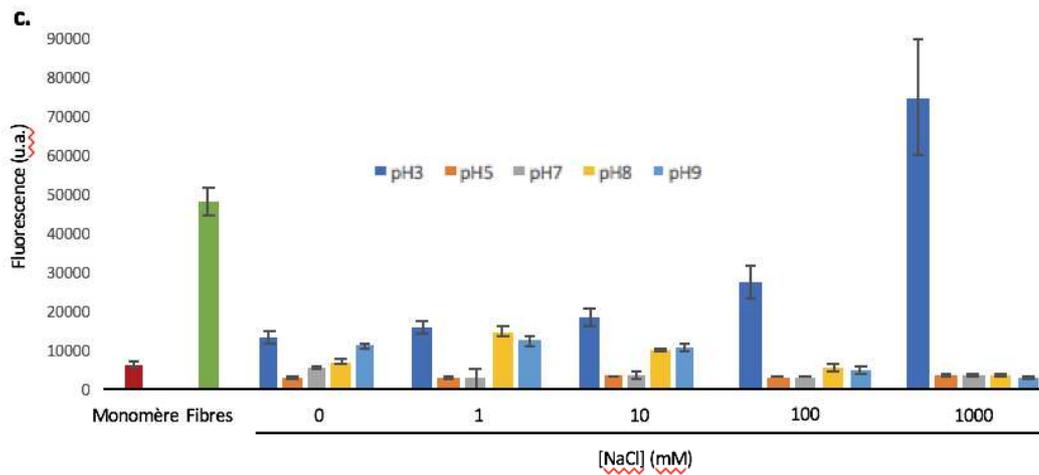


Figure 4.5 : Induction de la polymérisation de pb6 monomère. a. Profils d'élution de chromatographie d'exclusion de taille de pb6 dans différentes conditions. En milieu basique la protéine se dégrade, en milieu acide on observe l'apparition d'un petit pic dans le volume mort mais le pic du monomère est inchangé. b. Microscopie en coloration négative de fibres obtenues par incubation de pb6 monomère dans un milieu acide. c. Histogramme de fluorescence de pb6 incubé avec une sonde (Thioflavine T) dans différentes conditions de pH et de sels.



4.3 Introduction de l'article

Ci-après est présenté l'article « Bacteriophage T5 tail tube structure suggests a trigger mechanism for Siphoviridae DNA ejection ». Dans cet article, la structure cristallographique de pb6, obtenue à partir de cristaux de la protéine monomérique, est analysée et comparée à d'autres TTPs ainsi qu'aux protéines des tubes du système de sécrétion de type 6. Trois reconstructions par cryo-microscopie électronique de l'assemblage de pb6 sous forme de tube sont également discutés : les fibres décrites plus tôt dans ce paragraphe et le tube de pb6 comme il apparaît dans les queues de phages avant et après interaction avec la protéine FhuA. Une description plus détaillée des queues de phages est faite dans la partie 6.2.3. Les données de microscopie électronique et de cristallographie sont combinées pour produire un modèle pseudo-atomique de la protéine au sein de l'assemblage et de mutants sont testés sur leur capacité à inhiber la formation de tubes.

Points clés

Résumé chapitre 4

- La protéine pb6 peut être purifiée sous deux formes : une forme de monomère soluble et une de fibres polymérisés insolubles
- La protéine pb6 semble plus stable au sein de fibres
- Un tampon de lyse sans NaCl limite grandement la polymérisation de la protéine à la lyse des cellules
- Le sel et un pH acide semble induire la polymérisation de la protéine, mais ce phénomène est encore mal compris

Chapter 5

Tubes de la queue du phage T5 : données structurales et déclenchement de l'éjection de l'ADN

5.1 Article

ARTICLE

DOI: 10.1038/s41467-017-02049-3

OPEN

Bacteriophage T5 tail tube structure suggests a trigger mechanism for *Siphoviridae* DNA ejection

Charles-Adrien Arnaud¹, Grégory Effantin¹, Corinne Vivès¹, Sylvain Engilberge¹, Maria Bacia¹, Pascale Boulanger², Eric Girard¹, Guy Schoehn¹ & Cécile Breyton¹

The vast majority of phages, bacterial viruses, possess a tail ensuring host recognition, cell wall perforation and safe viral DNA transfer from the capsid to the host cytoplasm. Long flexible tails are formed from the tail tube protein (TTP) polymerised as hexameric rings around and stacked along the tape measure protein (TMP). Here, we report the crystal structure of T5 TTP pb6 at 2.2 Å resolution. Pb6 is unusual in forming a trimeric ring, although structure analysis reveals homology with all classical TTPs and related tube proteins of bacterial puncturing devices (type VI secretion system and R-pyocin). Structures of T5 tail tubes before and after interaction with the host receptor were determined by cryo-electron microscopy at 6 Å resolution. Comparison of these two structures reveals that host-binding information is not propagated to the capsid through conformational changes in the tail tube, suggesting a role of the TMP in this information transduction process.

¹Université Grenoble Alpes, CNRS, CEA, Institut for Structural Biology (IBS), 38000 Grenoble, France. ²Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, Université Paris-Saclay, 91405 Gif-sur-Yvette, France. Correspondence and requests for materials should be addressed to C.B. (email: Cecile.Breyton@ibs.fr)

Phage tail architectures and strategies of cell wall recognition and perforation are different for each family of tailed phages: *Myoviridae* use a 'syringe-like' mechanism, whereby the long and straight contractile tail mechanically and chemically 'drills' the cell wall with a metal-loaded needle^{1–4}. For the short tailed *Podoviridae*, the sequence of events has also been investigated and proteins involved in DNA delivery have been identified: following attachment of the receptor fibres to the host, conformational changes are transmitted to the core proteins located in the capsid, leading to their expulsion, resulting in the formation of a channel that spans the whole cell wall^{5–8}. Concerning the large family of the *Siphoviridae*, structural information (refs. 9, 10 and references therein) has yet to detail the mechanisms by which receptor binding promotes DNA ejection. A study based on negative-stain electron microscopy (EM) data proposed the tail tube protein (TTP) to be involved in signal transduction from the distal end of the tail to the capsid in *Bacillus subtilis* phage SPP1¹¹. Beyond differences between the phage families, the wealth of structural data on phage tail proteins points to strong structural homologies, highlighting a common protein building block that has been duplicated and decorated with different domains to serve alternative functions within the long phage tails (e.g. refs. 9, 12). Furthermore, structural homologies between tail proteins, proteins of the type VI secretion system (T6SS) of pathogenic bacteria^{13, 14} and of R-pyocins¹⁵ suggest a common evolutionary origin that evaded sequence

analysis because of very low sequence conservation (reviewed in refs. 9, 14). In particular, the inner tube of all these puncturing devices is formed by the stack of doughnut-shaped, structurally very conserved hexamers.

Phage T5 is a *Siphoviridae* infecting the Gram-negative host *Escherichia coli*. Its overall structure has been determined¹⁶, showing an unusual threefold symmetry of the tail tube¹⁷. To our knowledge, such a threefold symmetric tail tube has been observed in only one other case, for siphophage Φ CbK^{18, 19}. In T5, the tail tube is formed by the stack of 40 trimers of the TTP pb6¹⁶ around the tape measure protein (TMP). The latter spans the whole tail tube as a long coiled-coil, and its C terminus is located in the tail tip complex²⁰, at the tip of which is found T5 receptor-binding protein²¹. The mere interaction of T5 with its host receptor, the outer membrane iron-ferrichrome transporter FhuA, triggers in vitro the release of DNA from the capsid²². Moreover, functional tails can be isolated from an amber mutant in the major capsid protein²¹. This makes T5 an attractive and well-suited system to investigate *Siphoviridae* tail structure and reorganisation induced upon DNA release. Here, we determined the structure of pb6, and show that it results in the duplication/fusion of the hexamerisation domain common to all other tubes. We also determined the structure of T5 tail tube by cryo-EM to 6 Å resolution. The fit of pb6 crystal structure in the EM density map allowed proposing a pseudo-atomic model of T5 tail tube. Comparison of the structures of T5 tail tube before and after

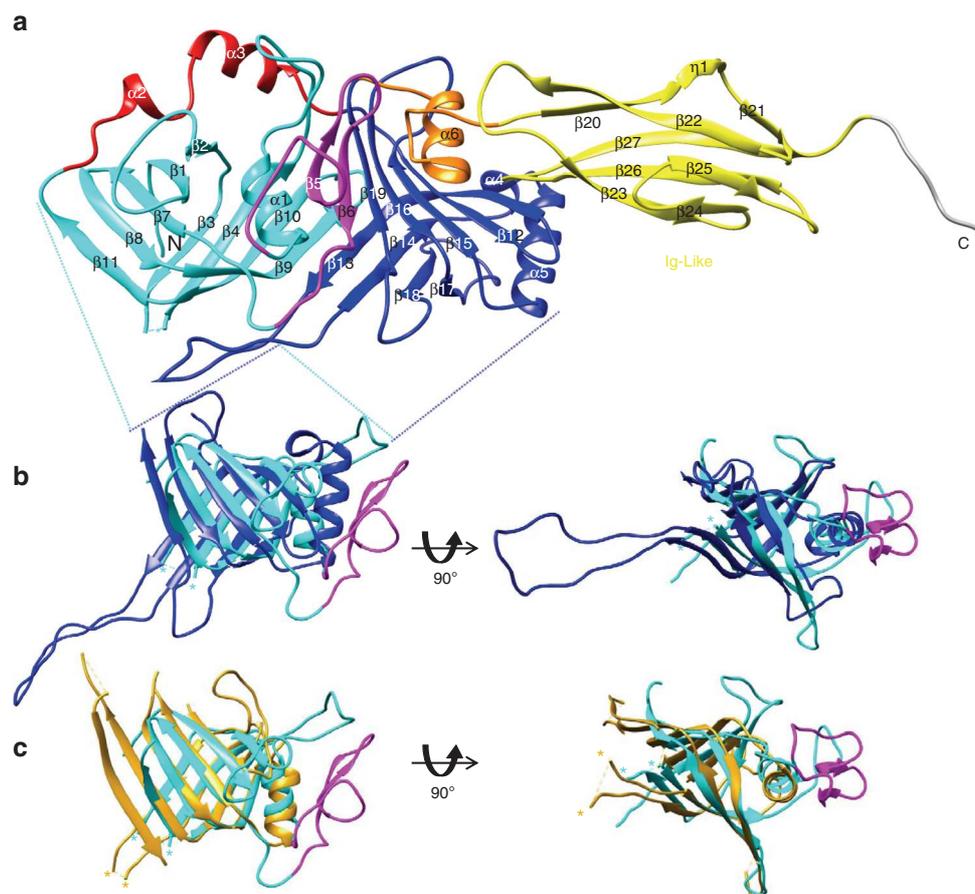


Fig. 1 Crystal structure of pb6. **a** Ribbon representation. Cyan: subdomain 1, blue: subdomain 2, magenta: insertions with respect to all other homologous proteins in subdomain 1 (see Supplementary Fig. 1b), red: linker between the two subdomains, yellow: Ig-like domain, orange: linker between subdomain 2 and the Ig-like domain, grey: TEV cleavage site. N and C: N and C terminus, respectively. **b** Superimposition of the two pb6 subdomains. **c** Superimposition of subdomain 1 (cyan) with the inner tube protein of the T6SS of *Yersinia pestis* (gold, PDB 3V4H). Asterisks point to the departure of the long disordered 28 amino acid loop β 3– β 4 of pb6 and of the tube protein of *Y. pestis* T6SS

Table 1 Crystallography data collection and refinement statistics

	Native PDB 5NGJ	MAD data set		
Data collection	ESRF/BM30A			
Space group	$P12_11$			
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	76.79, 115.08, 83.36			
α , β , γ (°)	90.00, 111.88, 90.00			
Wavelength (Å)	0.97965	Peak	Inflection	Remote
Resolution (Å) (last resolution shell)	57.54–2.20 (2.32–2.20)	0.97932	0.97952	0.97836
		77.80–3.00 (3.16–3.00)	77.83–3.00 (3.16–3.00)	57.93–3.00 (3.16–3.00)
R_{merge}	0.070 (1.724)	0.104 (0.507)	0.111 (0.697)	0.115 (0.683)
$I/\sigma(I)$	14.4 (1.6)	14.6 (3.5)	13 (2.6)	5.4 (1.5)
Completeness (%)	99.7 (97.9)	99.7 (98.3)	99.8 (98.4)	99.8 (98.5)
Multiplicity	7.5 (7.5)	7.5 (7.5)	7.6 (7.5)	3.8 (3.7)
Refinement	Buster-TNT			
No. reflections	67,830 (6631)			
$R_{\text{work}}/R_{\text{free}}$	0.2050 (0.2451)/0.2350 (0.2756)			
No. atoms	6917			
Protein	6684			
Ligand/ion	5			
Water	228			
<i>B</i> factors	85.64			
Protein	86.01			
Ligand/ion	79.61			
Water	75.19			
r.m.s. deviations				
Bond lengths (Å)	0.014			
Bond angles (°)	1.74			
Ramachandran favoured (%)	95.93			
Ramachandran allowed (%)	3.72			
Ramachandran outliers (%)	0.35			

interaction with its receptor shows no differences, suggesting that pb6 plays no role in the transduction of receptor binding from the tip of the tail to the capsid.

Results

Crystal structure of pb6. We have determined the pb6 monomer structure at 2.2 Å resolution (Fig. 1a and Table 1). At first sight, the structure can be divided into two domains encompassing 374 and 85 residues (Fig. 1a). The C-terminal domain (residues 375–462) possesses an immunoglobulin-like (Ig-like) fold of the Big-2 family²³, confirming a previous sequence analysis²¹. Ig-like domains are very common in phage proteins and have been proposed to play accessory roles in the infection process, probably by binding to carbohydrates²⁴. They are particularly found in TTPs of siphophages, as in phage λ ^{23,25} and SPP1²⁶, where it was shown that they are dispensable for phage assembly and infectivity. In λ , however, its absence has an influence on burst size and temperature sensitivity of the phage particle²³. For pb6, formation of tubes occurs even when this domain is absent (see below), and some T5-like TTP lack it (Supplementary Fig. 1a). A closer examination of the N-terminal domain structure reveals subdomain duplication, which is confirmed by DALI pairwise comparison²⁷ (Fig. 1a, b and Supplementary Table 1). The common core shared by both subdomains consists of a β -sandwich flanked by an α -helix, and a long loop. This loop is not resolved in subdomain 1 (loop β 3– β 4), whereas it is stabilised by crystal contacts in subdomain 2 (loop β 13– β 14; Fig. 1b and Supplementary Fig. 2). On a sequence level, T5-like TTPs have no homologues in the databases according to the PSIBlast and HHPRED software tools. However, both subdomains display high structural homology with TTPs of other siphophages, distal tail proteins of siphophages, T6SS tube proteins, the tube

protein of R-pyocin, baseplate hub proteins from myophages and spike proteins from T6SS (Fig. 1c, Supplementary Fig. 1b and Supplementary Table 1). Thus, T5 TTP is not an outlier in the family of TTPs, as it contains the same hexamerisation domain as other phages, T6SS and pyocin tube proteins, but results from a duplication/fusion: the unusual threefold symmetry of the tube is in fact pseudo-hexameric. Such a duplication has been observed before in myophage hub and T6SS spike proteins^{14,28}. However, these should be considered as two independent events, since the two subdomains are linked differently in pb6 and in the hub and spike proteins (Supplementary Fig. 3).

Pseudo-atomic structure of T5 tail tube. In parallel, we have determined the structure of T5 tail tube by cryo-EM at 6.2 Å resolution (Fig. 2a, d, g, j, Supplementary Fig. 4a and Supplementary Table 2). This structure shows a 90-Å large tube (pitch = 40.6 Å, twist = 39.1°) that is decorated by a bulb of density pointing outwards, as is often observed for *Siphoviridae* (ref. ²⁹ and references therein). Electron density is present in the 40-Å wide lumen of the tube, which most likely corresponds to the TMP pb2. However, no structural information can be derived from this density as threefold symmetry and averaging along the tube will have blurred it. The X-ray structure of the pb6 monomer fits remarkably well in the tube density (Chimera³⁰ correlation 0.89 using a pb6 map simulated from atoms at a resolution of 6 Å, 879/6657 atoms being outside the contour), except for the N terminus, the long β 13– β 14 loop, loop β 17– β 18 and helix α 5, all positioned at interfaces between monomers (Fig. 3a–c). Interestingly, these elements have a high temperature factor—i.e., are more flexible—in the crystal structure, and we could model them and loop β 3– β 4 in the unassigned EM densities observed in their vicinity, leading to a complete pseudo-atomic model of the T5 tail

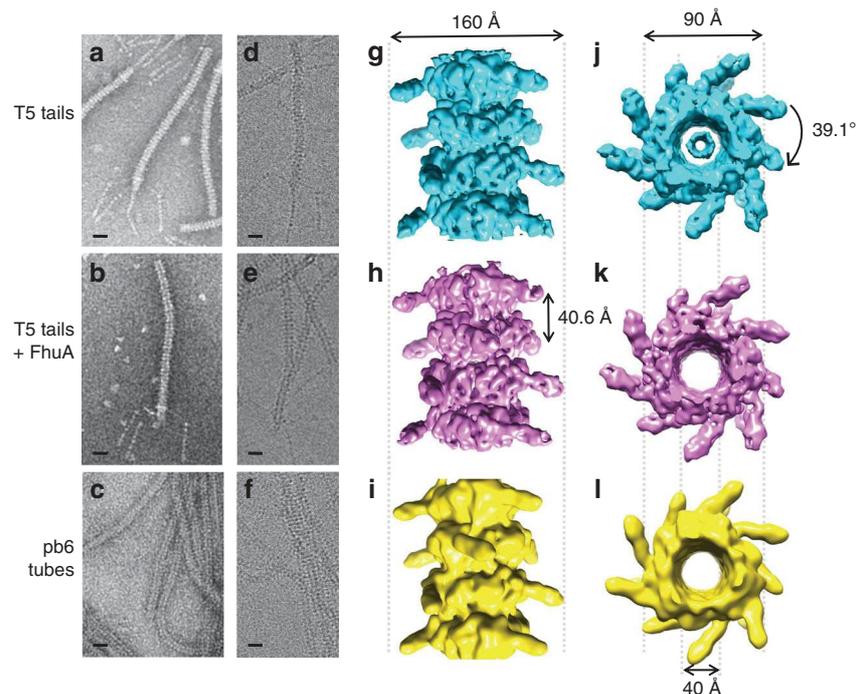


Fig. 2 EM reconstructions of T5 tail before and after interaction with FhuA. Negative stain (**a, b, c**) and cryo-EM (**d, e, f**) images, and cryo-EM 3D reconstruction (side views (**g, h, i**) and top view (**j, k, l**)) of T5 tails before (6.2 Å resolution, **a, d, g, j**) and after (5.8 Å resolution, **b, e, h, k**) interaction with FhuA, and of pb6 tubes (8.8 Å resolution, **c, f, i, l**). Scale bar, 20 nm

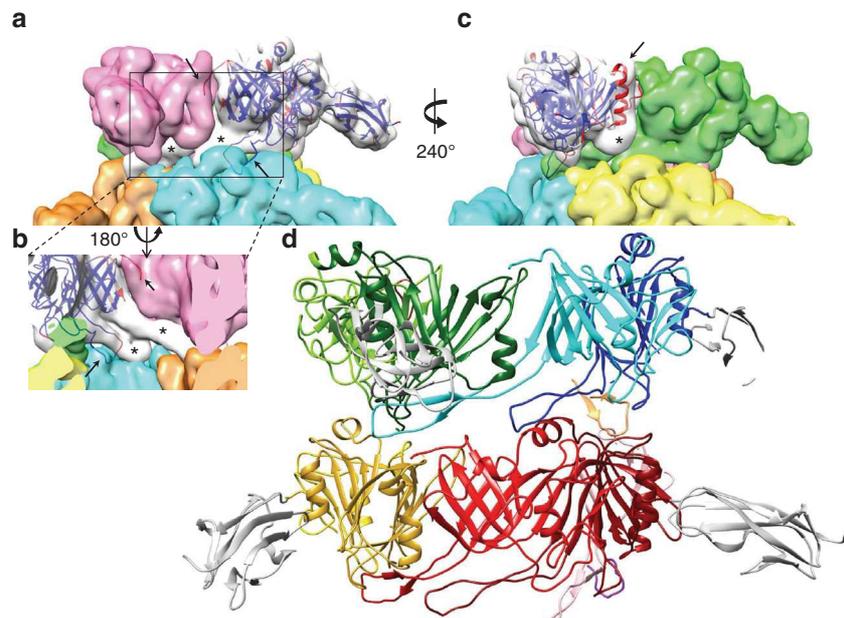


Fig. 3 Rigid fit of pb6 crystal structure in the EM reconstruction. **a** Side view of the segmented EM tail tube density. In the white segment, the crystal structure of the pb6 monomer was rigidly fitted. It is represented as a ribbon coloured according to the averaged residue temperature factor (red: more flexible, blue: more rigid). **b** Enlarged-up view, seen from the inside of the tube and **c** view rotated 240° with respect to **a**. Unattributed densities are indicated by an asterisk, whereas arrows point to structure elements that do not fit the segmented density. **d** Side view of the pseudo-atomic model of the tail tube after flexible fitting of pb6 into each segment (see also Supplementary Fig. 5). Within each pb6 protein, the two subdomains are coloured with two tones of the same colour (light and dark green, cyan and blue, and sandy brown for the upper ring and yellow and gold, light and dark red, and salmon and purple for the lower ring), whereas the Ig-like domain is coloured white in all cases

tube (Figs. 3d and 4a and Supplementary Fig. 5). Mutation of conserved charged amino acids within these loops (Supplementary Fig. 1a), as well as the deletion of the two long loops ($\beta 3$ – $\beta 4$ and $\beta 13$ – $\beta 14$) or the N terminus, impaired tube formation (Fig. 5, Supplementary Table 3 and see below), confirming their

importance in tube formation/stabilisation. This is a common phenomenon among phage proteins, where oligomerisation of a protein is regulated by interaction with its partner, and unstructured regions play a central role in regulating assembly¹⁰. In the tube assembly, a monomer is connected to six

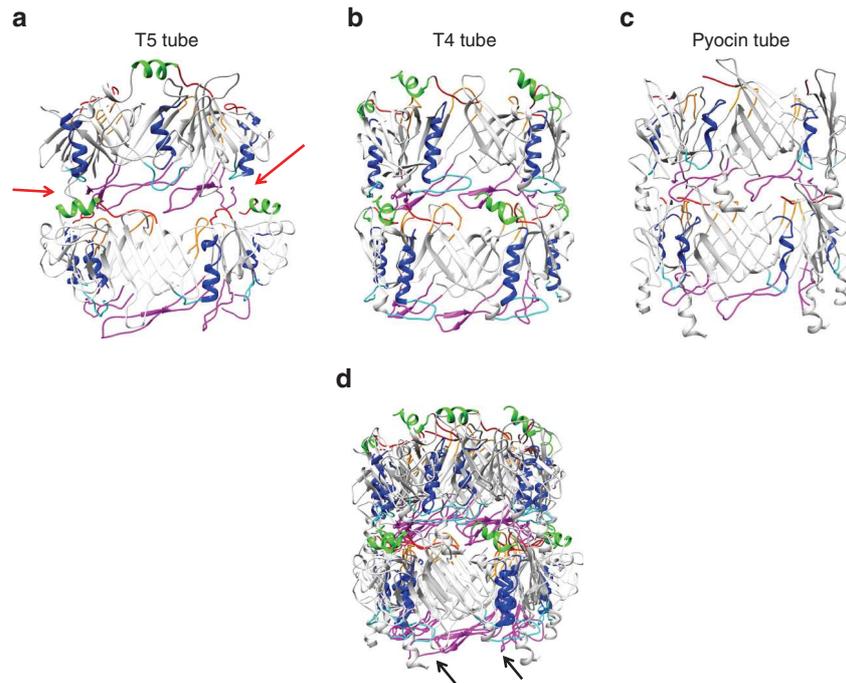


Fig. 4 Comparison of known tube structures. Side views of a stack of two TTP rings are shown, as a ribbon representation. **a** T5 tail tube pseudo-atomic model. pb6 Ig-like domain and its linker were removed for clarity. **b** Structure of the T4 tail tube (PDB 5IV5). **c** Structure of R-pyocin tube (PDB 3J9Q). Red arrows point to notches in the T5 tube that could explain tube flexibility; in T4 and pyocin tubes, additional inter-ring contacts are observed at these positions. **d** Superimposition of the three tubes, aligned on the lower ring. The longer loops of the two pb6 subdomains (black arrows) and asymmetry of the interface of two subdomains seem to impose a larger twist: T5 ring is not in register with the two other tubes in the upper ring. In **a–d**, similar structural features are coloured accordingly: the long loop (magenta) and helix (blue) of the common core fold, the N terminus (red) and two smaller loops (cyan and orange) and an additional N-terminal helix (green)

neighbouring subunits, making this structure very stable (pb6 tube melting temperature = 83 °C vs. 50 °C for pb6 monomer; note that it is 90 °C for purified tails). A comparison with the other tube structures available (phage T4 (PDB 5iv5⁴) and R-pyocin (PDB 3j9q¹⁵)) indicates that the inner and outer tube diameters are identical, as are the distances between the rings. As expected from the common fold of the ring building block in these three tubes, common features in monomer interfaces can be identified (Fig. 4). However, the twists of the tubes differ (39.1° vs. $18.3^\circ \times 2 = 36.6^\circ$ for pyocin and $17.9^\circ \times 2 = 36.8^\circ$ for T4, Fig. 4d), which could stem from the long loops being longer in pb6. Flexibility of the T5 tail tube could stem from fewer inter-ring interactions (red arrows, Fig. 4a). The interior of the tube is strongly negatively charged, as expected for a tube that channels DNA³¹.

No structural change upon interaction with the receptor. We then determined the structure of the tail tube after receptor binding, which results in an empty tail tube²¹, by cryo-EM to 5.8 Å resolution (Fig. 2b, e, h, k, Supplementary Fig. 4a and Supplementary Table 2). We did not observe any structural differences at this resolution with the tail before receptor binding (Chimera correlation 0.98) besides the absence of the internal electron density assigned to the TMP (Supplementary Fig. 6a and Supplementary Movie 1). The resolution of both structures reached a limit around 6 Å; addition of more particles in the refinement was ineffective in improving it (Supplementary Fig. 4a). This is probably linked to the flexibility of both the Ig domain and the tail, which would limit the overall resolution. Local resolution is estimated between 5 and 6 Å for the core of the structure (subdomains 1 and 2), while the more peripheral densities (Ig-like

domains) are less well defined (resolution >6.5 Å) (Supplementary Fig. 4b).

When overexpressed, pb6 is recovered in part in the soluble fraction of broken cells, but also, in majority, in the insoluble fraction. Whereas the soluble protein is monomeric, the protein in the insoluble fraction forms very long tubes of several μm (Fig. 5a). The cryo-EM structure of these tubes at 8.8 Å resolution (Fig. 2c, f, i, l, Supplementary Fig. 4a and Supplementary Table 2) does not show any difference with both the full or empty tail tubes (chimera correlation 0.99, Supplementary Fig. 6b and Supplementary Movie 2). Thus, even though pb6 monomers can be concentrated at least up to 100 mg mL^{-1} without forming tubes or aggregating, they can spontaneously assemble during cell lysis to form a tube that is structurally identical to that of the tail, although much longer in the absence of TMP. However, the trigger for this polymerisation is unclear, although we noticed that it is enhanced by low pH and high salt concentration. Several pb6 deletions and point mutants were overexpressed to evaluate their ability to form tubes (Fig. 5b, c and Supplementary Table 3). The design of these mutants was based on sequence conservation within T5-like TTPs (Supplementary Fig. 1a) or on monomer interfaces identified in the pseudo-atomic tail structure. As expected, and as observed in tube proteins of phage λ ¹³, phage SSP1³² and *Burkholderia pseudomallei* T6SS³³, deletions of/substitutions in the long loops abolish tube formation, confirming their importance in tube stabilisation. Tube formation is also totally or partially impaired in the case of alanine substitutions at monomer–monomer interfaces. In the latter case, tubes are still observed (Supplementary Table 3), indicating that the large interaction surfaces stabilising the tube allow some tolerance for localised perturbation.

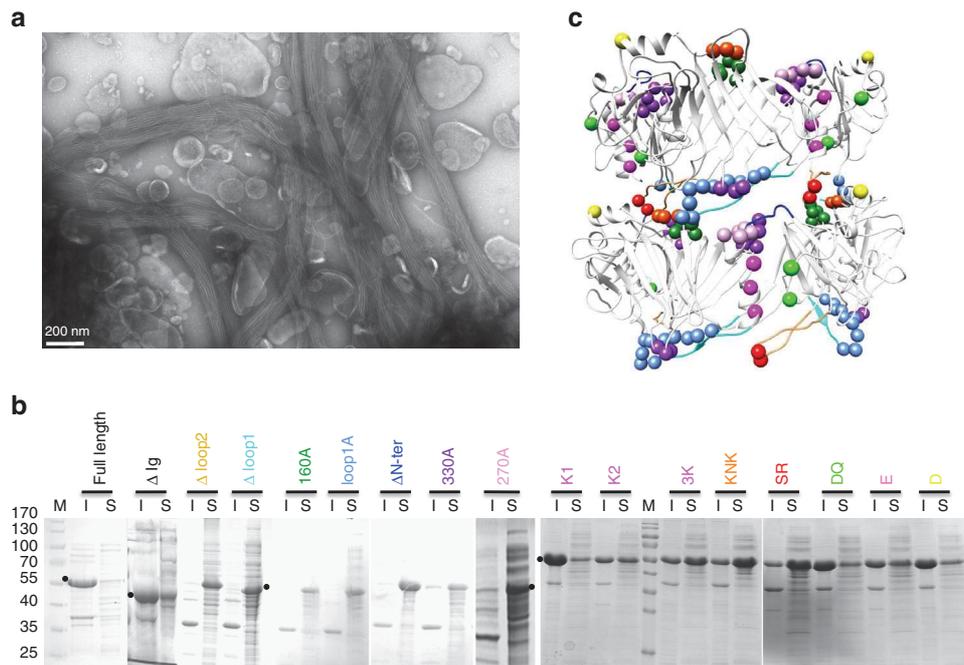


Fig. 5 Mutations affecting tube formation. **a** Negative stain EM of the pellet of broken cells showing micrometer-long pb6 tubes. Bar, 200 nm. **b** Coomassie-stained SDS-PAGE of the soluble (S) and insoluble (I) material after centrifugation of broken cells overexpressing the full-length protein or different mutants (see also Supplementary Table 3). The black dots point to the position of pb6 in the gel. All the generated mutants are presented. M molecular markers (170, 130, 100, 70, 55, 40, 35 and 25 kDa). **c** Two rings of the T5 tail tube ribbon model, with spheres indicating the position of amino acids that were mutated. The Ig-like domain and its linker were removed for clarity. Mutations in the same colour were performed in the same mutant protein. Same colour code in **b** and **c**, see also Supplementary Table 3

Discussion

At the resolution of our EM structures (6 Å), T5 tail tube does not undergo conformational changes upon binding of the tail tip to its receptor. Furthermore, the structure of the building block that forms the different tubes is remarkably conserved in those of siphophage tails, T6SS and R-pyocin, suggesting little or no flexibility of the structured core of the protein. We suggest that in all these structures, the TTP would have only a scaffolding function and would not be involved in signal transduction or in expulsion of the TMP. It is possible that previous contradictory observations on siphophage SPP1¹¹ resulted from negative-stain artefacts that may occur during stain interaction with the protein, interaction of the protein with the carbon layer and/or drying of the sample.

Hence, in *Siphoviridae*, we propose both signal transduction and membrane perforation functions are fulfilled by the TMP with the TMP constrained in a metastable state within the tail assembly, as for the sheath of contractile devices^{3, 15} and for the *Podoviridae* core proteins³⁴. Interaction of the receptor binding proteins with the host receptor would initiate conformational changes, culminating in the release of TMP, leading to its expulsion from the tail tube and its insertion into the cell wall²⁰. Moreover, the membrane insertion property and muralytic activity described for the C-terminal domains of T5 TMP would serve as complementary forces driving TMP ejection/refolding²⁰. The exit of the TMP at the level of the head-to-tail connector would in turn induce conformational changes in the latter, resulting in its opening for DNA release³⁵.

In *Myoviridae* and the other bacterial contractile devices, contraction of the sheath provides the energy to perforate the cell wall^{3, 15, 36}. It does not, however, seem to be the signal that opens the capsid in *Myoviridae*, as contracted tails with full capsids have been isolated³. The signal for capsid opening was proposed to be generated by TMP ejection² or conformational changes in the

TTP³⁷. Given the remarkable conservation of structures and tail organisation (tube, inner baseplate)⁹ within *Siphoviridae* and *Myoviridae* phages, we propose that the TMP of *Siphoviridae* and *Myoviridae* plays a similar role in signal transduction.

Methods

Protein and tail production and purification. Production and purification of the monomeric soluble pb6 protein bearing a hexahistidine tag at its C terminus were performed as described³⁸: *E. coli* BL21 (DE3) strain (BL21(DE3) Singles™ competent cells, Merck), transformed with the pb6-pLIM13 plasmid was grown in LB medium with kanamycin (50 µg mL⁻¹) at 37 °C until stationary phase. We relied on the leak of the promoter for protein expression as induction yielded 100% insoluble protein. Cells were broken in 50 mM Tris-HCl, pH 8—which optimises pb6 monomer fraction—with a microfluidizer, and the clarified cell lysate, supplemented with 15 mM imidazole, was loaded onto a Ni-NTA column (HiTrap™ Chelating HP 5 mL GE Healthcare) equilibrated with 20 mM Tris pH 8, 100 mM NaCl, 15 mM Imidazole. The column was washed with the equilibration buffer, and proteins were eluted with 20 mM Tris pH 8, 300 mM imidazole. Protein containing fractions were pooled and diluted five times before loading onto an ion exchange column (HiTrap™ Q HP 1 mL, GE Healthcare) equilibrated with 20 mM Tris pH 8. The protein was eluted by a 0–1 M NaCl linear gradient. Selenomethionine protein was produced in the same strain but grown in an in-house M9 minimal medium (50 mM Na₂HPO₄, 20 mM K₂HPO₄, 10 mM NaCl, 20 mM NH₄Cl, 2 mM MgSO₄, 100 µM CaCl₂, 2 g L⁻¹ glucose). Fifteen minutes prior to 100 mM IPTG induction, 30 mg L⁻¹ selenomethionine, 50 mg L⁻¹ threonine, lysine, phenylalanine and leucine, 25 mg L⁻¹ isoleucine and valine were added to the culture. Protein was purified as the native protein.

FhuA was produced and purified as described³⁹: *E. coli* AW740 [FhuA31 Δ ompF *zcb::Tn10* Δ ompC] transformed with the pHX405 plasmid, in which the *fhuA* gene is under control of its natural promoter, was grown at 37 °C in LB medium supplemented with ampicillin (125 µg mL⁻¹), tetracycline (10 µg mL⁻¹) and 2,2'-bipyridyl (100 µM), an iron chelator used to induce the production of the protein. After clarification of the cell lysate, total membranes were recovered by ultracentrifugation and solubilised by 50 mM Tris pH 8, 2% (w/w) OPOE (*N*-octylpolyoxyethylene, Bachem) at 37 °C for half an hour. The insoluble material was recovered by ultracentrifugation and solubilised 1 h at 37 °C by 50 mM Tris pH 8, 1 mM EDTA, 1% (w/w) LDAO (*N,N*-dimethyl dodecylamine-*N*-oxide, Anatrace). The solubilised fraction, recovered after ultracentrifugation, was supplemented with 4 mM MgCl₂ and 5 mM imidazole and loaded on a nickel affinity column (HiTrap™ Chelating HP 5 mL GE Healthcare) previously

equilibrated with 0.1% LDAO, 20 mM Tris pH 8, 200 mM NaCl and washed with the same buffer. Prior to elution, a delipidation step was performed with 10 mL of 1% LDAO in the same buffer. The protein was eluted from the column with 0.1% LDAO, 20 mM Tris pH 8, 200 mM imidazole, and loaded onto an ion exchange column (HiTrap™ Q HP 1 mL, GE Healthcare) equilibrated with 0.05% LDAO, 20 mM Tris pH 8. The protein was eluted by a 0–1 M NaCl linear gradient in the same buffer. Purified FhuA was transferred into amphipol (A8–35) at a protein:amphipol ratio of 1:5 (w:w). After 1 h 40 min incubation of the protein:amphipol mixture at room temperature, detergent was removed by the addition of 200 mg mL⁻¹ BioBeads (BioRad) and incubation on a stirring wheel at room temperature for 45 min.

Tubes of pb6 tubes were purified from the insoluble fraction of the cell lysate: the broken cell pellet was resuspended and incubated for 1 h at 37 °C in 50 mM Tris-HCl pH 8, 100 mM NaCl, 0.5% Triton X-100, 1 mM EDTA, 100 µg mL⁻¹ lysosyme before loading onto a 30–50% sucrose gradient (20 mM Tris-HCl pH 8, 100 mM NaCl, 0.05% Triton X-100) centrifuged at 10,000 rpm for 30 min (SW41 rotor) at 4 °C. Pure tubes were recovered in the lower part of the gradient and extensively washed by dilution/centrifugation in 20 mM Tris-HCl pH 8, 100 mM NaCl to remove sucrose and detergent.

Production and purification of T5 tails were developed elsewhere²¹: *E. coli* F cultures at 37 °C were infected during the exponential growth phase with the amber mutant phage T5D20am30d, at a multiplicity of infection of 6–7. After complete cell lysis (OD_{600nm} < 0.15), T5 tails were precipitated from the culture medium by incubation with 0.5 M NaCl and 10% (w/w) PEG 6000 overnight at 4 °C. The pellet was resolubilised in 10 mM Tris pH 8, 100 mM NaCl 1 mM CaCl₂, 1 mM MgCl₂ and purified on a glycerol step gradient (10–40%) in the same buffer, and centrifuged 2 h at 20,000 rpm (SW41 rotor). The gradient fractions containing the tails (usually ~10% glycerol), diluted ten times in 10 mM Tris pH 8, were loaded onto an ion exchange column (HiTrap™ Q HP 1 mL, GE Healthcare) equilibrated with the gradient buffer without NaCl and washed in that same buffer. The tails were eluted by a 0–0.5 M NaCl linear gradient in the same buffer.

Mutations. Expression vectors of pb6 mutants were obtained using the Quick-Change Lightning Site-directed Mutagenesis Kit (Agilent Technologies) on the pb6-pLM13 plasmid³⁸ according to the recommendations of the manufacturer, with primers described in Supplementary Table 4. All pb6 mutants were sequenced prior to transformation. Mutant proteins were produced as wild-type protein, except for pb6-ΔNter, which production had to be induced with 1 mM IPTG. Cells were resuspended and broken in 50 mM Tris-HCl, pH 8, 150 mM NaCl, which enhances the relative amount of pb6 tubes. After ultracentrifugation of the cell lysate, soluble and insoluble fractions were deposited on SDS-PAGE to compare relative quantities of pb6.

Measure of melting temperature. Melting temperatures of purified pb6 monomer (3 mg mL⁻¹), purified pb6 tube and purified T5 tail (~6 mg mL⁻¹) were obtained using a Prometheus NT.48 (NanoTemper Technologies) according to the manufacturer's recommendations.

Crystallisation. Initial hits were obtained on the high throughput crystallisation platform on site. Crystals were reproducibly obtained at 4 °C by the hanging drop vapour diffusion setup by mixing 1:1 (v/v) protein solution (8 mg mL⁻¹) with reservoir buffer containing 10% 3350 PEG, 0.2 M HEPES pH 7.5, 0.3 M LiSO₄, 0.1 mM MnCl₂. Selenomethionine-pb6 crystals were obtained with a protein solution at 7 mg mL⁻¹ and a reservoir buffer containing 11% 3350 PEG, 0.2 M HEPES pH 7.5, 0.1 M LiSO₄, 40 mM MnCl₂. Crystals typically grew within 2 days. All crystals were cryoprotected with 25% PEG 400 and flash-frozen in liquid nitrogen using nylon loops (CryoLoop™, Hampton Research).

Data collection and crystallography. X-ray data sets were collected at the European Synchrotron Radiation Facility (French beamline for Investigation of Proteins – BM30A). A native data set was collected on a native crystal up to 2.2 Å and a MAD data set was collected on a single selenomethionine-pb6 crystal up to 3 Å. Diffraction frames were integrated using the programme XDS⁴⁰ and the integrated intensities were scaled and merged using the CCP4 programmes⁴¹ SCALA and TRUNCATE, respectively. A summary of the processing statistics is given in Table 1. pb6 was solved de novo by combining the MAD and the native data sets. Selenium positions were determined within the asymmetric unit using the programme SHELXD⁴². Heavy-atom refinement and initial phasing were performed using the programme SHARP⁴³. Phases from SHARP were improved by density modification using the programme SOLOMON⁴⁴. The figure of merit after density modification was 0.82. The first steps of model building were performed automatically using BUC-CANEER⁴⁵ and completed manually with COOT⁴⁶. Model refinement was performed with PHENIX⁴⁷ and BUSTER (2016 BUSTER version 2.10.3, Cambridge, UK: Global Phasing Ltd). The model was optimised through iterative rounds of refinement and model building. Automated non-crystallographic symmetry and translation/libration/screw were applied, and at the end stages of the refinement hydrogen were added (except for solvent molecules).

Refinement statistics are summarised in Table 1, and a stereo image of a portion of the electron density map is presented in Supplementary Fig. 7. The atomic

coordinates and measured structure factor amplitudes for pb6 were deposited in the Protein Data Bank with accession code 5NGJ.

Negative stain electron microscopy. Negative-staining EM allowed purity and concentration assessment, as well as adapting the T5 tail:FhuA ratio prior to cryo-grid preparation. Negative-stain grids of all samples were prepared using the mica-carbon flotation technique over a 2% sodium silicotungstate solution and were observed with an FEI/Thermo Fisher Scientific Tecnai 12 microscope (LaB₆ electron source, 120 kV) equipped with a Gatan Orius 1000 CCD camera.

Cryo-electron microscopy. pb6 tubes and purified T5 tails incubated with or without FhuA were prepared following the same procedure. An aliquot of 3 µL of sample was applied to glow discharged 2:1 Quantifoil holey carbon grid (Quantifoil Micro Tools GmbH, Germany) and the grids were plunge-frozen in liquid ethane with a Vitrobot Mark II (FEI/Thermo Fisher Scientific). Samples were observed with an FEI/Thermo Fisher Scientific Polara at 300 kV. Images were recorded on a K2 Summit direct detector (Gatan Inc., USA) in super resolution counting mode. Movies were recorded for a total exposure of 4 s and 100 ms per frame, resulting in 40-frame movies with a total dose of ~40 e⁻ Å⁻². For tails incubated with FhuA and pb6 fibres, the magnification was ×23,000 (0.82 Å per pixel at the sample level) while for T5 tails, the nominal magnification was ×15,500 (1.24 Å per pixel at the sample level).

3D reconstruction. An initial model was created by back-projecting a portion of a single straight T5 tail (total of ~6 consecutive trimeric stacks) and by applying threefold symmetry using SPIDER⁴⁸. This initial model was refined using a classical projection matching method as described⁴⁹. Briefly, the contrast transfer function (CTF) corrected tail images were boxed into 128 × 128 pixel windows with an overlap of ¾ of the box. The generated stacked images (1000 images) were aligned against the projections of the model and averaged into classes. The classes were then back projected to obtain a new threefold symmetrised model. After 10 cycles of refinement, we obtained a map at ~10 Å resolution in which the crystal structure of pb6 could be fitted without ambiguity. In order to further improve the resolution, a larger data set was processed as described below.

Images were first motion corrected either with Digital Micrograph (Gatan Inc., USA) for T5 tails or with MotionCorr⁵⁰ for tails incubated with FhuA and pb6 tubes. The latter were also binned two times by Fourier cropping before further processing (final pixel size of 1.64 Å per pixel). The same image analysis procedure was followed for all samples. Tubes were first boxed as overlapping segments with boxer⁵¹. The interbox distance was set to one axial rise (z = 40 Å) for all samples. The coordinates were then imported into RELION 1.4⁵². CTFIND4⁵³ was used to determine the CTF parameters. 2D classification was then performed in RELION to remove outliers. The low-resolution 3D initial model (see above) or a featureless cylinder were used as the initial reference for a first 3D refinement imposing C3 symmetry. As nearly all possible helical segments were picked by choosing an interbox distance equal to the axial rise, no further helical averaging was needed. This was confirmed later by two different refinement strategies, both enforcing helical symmetry, using IHRSR⁵⁴ and RELION 2⁵⁵. Both approaches gave 3D density maps nearly identical to the one obtained without helical symmetry empty and full tails were then separated by successive 3D classifications. 3D classes displaying central density were grouped together while those having an empty channel formed a second group. A last 3D refinement was then done on the 22704 empty (T5 tail incubated with FhuA) and the 24243 full (T5 tails) particles, which yielded 3D reconstructions at 5.8 and 6.2 Å, respectively (Fig. 2g, j, h, k and Supplementary Table 2) after post processing in RELION with a threshold mask. For the pb6 tubes, no 3D classification was needed as the tubes were all empty, and the final 3D reconstruction had 8.8 Å resolution and included 12,244 particles (Fig. 2i, l and Supplementary Table 2). Map visualisation, segmentation and fitting were done with CHIMERA³⁰.

As the 3D reconstructions obtained for full or empty T5 tails are virtually identical at the pb6 level, the two data sets were combined by scaling the T5 tails acquired at the nominal magnification of 15,500 from a pixel size of 1.24–1.64 Å per pixel. The combined data set was processed as described above but with RELION 2, which implements helical symmetry during image processing. For consistency reasons, the data sets for full and empty T5 tails were re-processed with RELION 2. As expected, the 3D reconstructions for both samples were nearly identical to the one obtained with RELION 1.4. The final 3D reconstruction for the combined data set included 55,447 particles and reached 6 Å resolution at FSC = 0.143, but with enhanced interpretability, notably at the subunit interfaces (Supplementary Fig. 4). This map was used to develop a pseudo-atomic model of the pb6 tube.

Model building. The crystal structure of pb6 was rigidly fitted in the tube density using the 'fit in map' option in CHIMERA³⁰. This original fit was used to guide segmentation of the map using SEGGER⁵⁶. Isolated densities corresponding to a monomer was then used to build the missing loop (loop β3–β4, residues 38–67) using COOT⁴⁶ and model the structural features rearranged in the tube assembly using FLEXEM⁵⁷. Definition of rigid domains was achieved using the RIBBID server⁵⁸, and the output file was slightly edited to allow mobility in protein regions not

correctly fitting density in the original rigid fit. For the most mobile element (N terminus 1–8, the two long loops $\beta 3$ – $\beta 4$ and $\beta 13$ – $\beta 14$ and loop $\beta 17$ – $\beta 18$) different conformations were used as starting points (manually modelled in COOT) and subjected to 5–10 iterations of simulated annealing rigid body (FLEXEM Molecular Dynamics (MD) mode). A new set of starting points was built by combining the best-looking models of all mobile elements (typically twisted or straight for the loops). These new starting points were subjected to five iterations of MD refinement. The convergence of the models generated from different starting points lowered the uncertainty of the orientation of the polypeptide chain.

Data availability. Coordinates and structure factors for pb6 have been deposited in the Protein Data Bank under accession number 5NGJ. Electron microscopy maps have been deposited in the Electron Microscopy Data Bank under accession codes EMD-3689 (full tail tube), EMD-3689 (empty tail tube), EMD-3691 (pb6 tube) and EMD-3692 (combined tail tube). Other data supporting the findings of this study are available from the corresponding author on reasonable request.

Received: 9 June 2017 Accepted: 2 November 2017

Published online: 05 December 2017

References

- Browning, C., Shneider, M. M., Bowman, V. D., Schwarzer, D. & Leiman, P. G. Phage pierces the host cell membrane with the iron-loaded spike. *Structure* **20**, 326–339 (2012).
- Hu, B., Margolin, W., Molineux, I. J. & Liu, J. Structural remodeling of bacteriophage T4 and host membranes during infection initiation. *Proc. Natl Acad. Sci. USA* **112**, E4919–E4928 (2015).
- Leiman, P. G., Chipman, P. R., Kostyuchenko, V. A., Mesyanzhinov, V. V. & Rossmann, M. G. Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. *Cell* **118**, 419–429 (2004).
- Taylor, N. M. I. et al. Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature* **533**, 346–352 (2016).
- Hu, B., Margolin, W., Molineux, I. J. & Liu, J. The bacteriophage $\tau 7$ virion undergoes extensive structural remodeling during infection. *Science* **339**, 576–579 (2013).
- González-García, V. A. et al. Conformational changes leading to T7 DNA delivery upon interaction with the bacterial receptor. *J. Biol. Chem.* **290**, 10038–10044 (2015).
- Lupo, D. et al. The T7 ejection nanomachine components gp15–gp16 form a spiral ring complex that binds DNA and a lipid membrane. *Virology* **486**, 263–271 (2015).
- Xu, J., Gui, M., Wang, D. & Xiang, Y. The bacteriophage $\phi 29$ tail possesses a pore-forming loop for cell membrane penetration. *Nature* **534**, 544–547 (2016).
- Veessler, D. & Cambillau, C. A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. *Microbiol. Mol. Biol. Rev.* **75**, 423–433 (2011).
- Davidson, A. R., Cardarelli, L., Pell, L. G., Radford, D. R. & Maxwell, K. L. Long noncontractile tail machines of bacteriophages. *Adv. Exp. Med. Biol.* **726**, 115–142 (2012).
- Plisson, C. et al. Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection. *EMBO J.* **26**, 3720–3728 (2007).
- Flayhan, A. et al. Crystal structure of pb9, the distal tail protein of bacteriophage T5: a conserved structural motif among all siphophages. *J. Virol.* **88**, 820–828 (2014).
- Pell, L. G., Kanelis, V., Donaldson, L. W., Howell, P. L. & Davidson, A. R. The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. *Proc. Natl Acad. Sci. USA* **106**, 4160–4165 (2009).
- Leiman, P. G. et al. Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc. Natl Acad. Sci. USA* **106**, 4154–4159 (2009).
- Ge, P. et al. Atomic structures of a bactericidal contractile nanotube in its pre- and postcontraction states. *Nat. Struct. Mol. Biol.* **22**, 377–382 (2015).
- Effantin, G., Boulanger, P., Neumann, E., Letellier, L. & Conway, J. F. Bacteriophage T5 structure reveals similarities with HK97 and T4 suggesting evolutionary relationships. *J. Mol. Biol.* **361**, 993–1002 (2006).
- Guénebaud, V. et al. TEM moiré patterns explain STM images of bacteriophage T5 tails. *Ultramicroscopy* **69**, 129–137 (1997).
- Leonard, K. R., Kleinschmidt, A. K. & Lake, J. A. Caulobacter crescentus bacteriophage phiCbK: structure and in vitro self-assembly of the tail. *J. Mol. Biol.* **81**, 349–365 (1973).
- Papadopoulos, S. & Smith, P. R. The structure of the tail of the bacteriophage phi CbK. *J. Ultrastruct. Res.* **80**, 62–70 (1982).
- Boulanger, P. et al. Phage T5 straight tail fiber is a multifunctional protein acting as a tape measure and carrying fusogenic and muralytic activities. *J. Biol. Chem.* **286**, 13556–13564 (2008).
- Zivanovic, Y. et al. Insights into bacteriophage T5 structure from analysis of its morphogenesis genes and protein components. *J. Virol.* **88**, 1162–1174 (2014).
- Boulanger, P. et al. Purification and structural and functional characterization of FhuA, a transporter of the Escherichia coli outer membrane. *Biochemistry* **35**, 14216–14224 (1996).
- Pell, L. G. et al. The solution structure of the C-terminal Ig-like domain of the bacteriophage λ tail tube protein. *J. Mol. Biol.* **403**, 468–479 (2010).
- Fraser, J. S., Maxwell, K. L. & Davidson, A. R. Immunoglobulin-like domains on bacteriophage: weapons of modest damage? *Curr. Opin. Microbiol.* **10**, 382–387 (2007).
- Katsura, I. Structure and function of the major tail protein of bacteriophage lambda. Mutants having small major tail protein molecules in their virion. *J. Mol. Biol.* **146**, 493–512 (1981).
- Auzat, I., Dröge, A., Weise, F., Lurz, R. & Tavares, P. Origin and function of the two major tail proteins of bacteriophage SPP1. *Mol. Microbiol.* **70**, 557–569 (2008).
- Holm, L. & Laakso, L. M. Dali server update. *Nucleic Acids Res.* **44**, W351–355 (2016).
- Kanamaru, S. et al. Structure of the cell-puncturing device of bacteriophage T4. *Nature* **415**, 553–557 (2002).
- Spinelli, S. et al. Cryo-electron microscopy structure of lactococcal siphophage 1358 virion. *J. Virol.* **88**, 8900–8910 (2014).
- Pettersen, E. F. et al. UCSF chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- Lhuillier, S. et al. Structure of bacteriophage SPP1 head-to-tail connection reveals mechanism for viral DNA gating. *Proc. Natl Acad. Sci. USA* **106**, 8507–8512 (2009).
- Langlois, C. et al. Bacteriophage SPP1 tail tube protein self-assembles into β -structure-rich tubes. *J. Biol. Chem.* **290**, 3836–3849 (2015).
- Lim, Y. T. et al. Extended loop region of Hcp1 is critical for the assembly and function of type VI secretion system in Burkholderia pseudomallei. *Sci. Rep.* **5**, 8235 (2015).
- Kemp, P., Garcia, L. R. & Molineux, I. J. Changes in bacteriophage T7 virion structure at the initiation of infection. *Virology* **340**, 307–317 (2005).
- Chaban, Y. et al. Structural rearrangements in the phage head-to-tail interface during assembly and infection. *Proc. Natl Acad. Sci. USA* **112**, 7009–7014 (2015).
- Kostyuchenko, V. A. et al. The tail structure of bacteriophage T4 and its mechanism of contraction. *Nat. Struct. Mol. Biol.* **12**, 810–813 (2005).
- Nováček, J. et al. Structure and genome release of Twort-like Myoviridae phage with a double-layered baseplate. *Proc. Natl Acad. Sci. USA* **113**, 9351–9356 (2016).
- Noirclerc-Savoie, M. et al. Tail proteins of phage T5: investigation of the effect of the His6-tag position, from expression to crystallisation. *Protein Expr. Purif.* **109**, 70–78 (2015).
- Flayhan, A., Wien, F., Paternostre, M., Boulanger, P. & Breyton, C. New insights into pb5, the receptor binding protein of bacteriophage T5, and its interaction with its Escherichia coli receptor FhuA. *Biochimie* **94**, 1982–1989 (2012).
- Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
- Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
- Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 479–485 (2010).
- Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 2023–2030 (2003).
- Abrahams, J. P. & De Graaff, R. A. New developments in phase refinement. *Curr. Opin. Struct. Biol.* **8**, 601–605 (1998).
- Cowan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
- Frank, J. et al. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
- Franzetti, B. et al. Tetrahedral aminopeptidase: a novel large protease complex from archaea. *EMBO J.* **21**, 2132–2138 (2002).
- Li, X. et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).

51. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
52. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
53. Rohou, A. & Grigorieff, N. CTFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
54. Egelman, E. H. The iterative helical real space reconstruction method: surmounting the problems posed by real polymers. *J. Struct. Biol.* **157**, 83–94 (2007).
55. Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5**, e18722 (2016).
56. Pintilie, G. D., Zhang, J., Goddard, T. D., Chiu, W. & Gossard, D. C. Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J. Struct. Biol.* **170**, 427–438 (2010).
57. Topf, M. et al. Protein structure fitting and refinement guided by cryo-EM density. *Structure* **16**, 295–307 (2008).
58. Pandurangan, A. P. & Topf, M. RIBFIND: a web server for identifying rigid bodies in protein structures and to aid flexible fitting into cryo EM maps. *Bioinforma. Oxf. Engl.* **28**, 2391–2393 (2012).

Acknowledgements

We thank Flavien Grégoire, Alice Labaronne and Valentin Sencio for help at different stages of the project, Daphna Fenel for help with negative-stain EM, Christine Ebel for support and James Conway for proof reading of the manuscript. We also acknowledge access to the European Synchrotron Radiation Facility (ESRF) and SOLEIL beamlines for crystal screening, and thank scientists of the FIP-BM30A at ESRF, in particular Michel Pirrochi for help with data collection. This work used the electron microscopy and the high-throughput crystallography platforms from the Grenoble Instruct Center (ISBG: UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). The IBS Electron Microscope facility is supported by the Rhône-Alpes Region, the Fonds Feder, the Fondation pour la Recherche Médicale and GIS-IBISA. We acknowledge funding by the ANR PerfoBac (ANR-16-CE11-0027-01). CAA was funded by GRAL.

Author contributions

C.-A.A., E.G., G.S., and C.B. designed and conceived the experiments. C.-A.A., G.E., S.E., C.V., M.B., P.B., E.G., G.S., and C.B. performed experiments (C.-A.A.: biochemistry, molecular biology, crystallography and crystallography, E.M. image processing, C.V.: molecular biology, S.E. and E.G.: crystallography, M.B. and G.S.: E.M. data acquisition, G.E., G.S. and C.B.: E.M. image processing, P.B.: designed the purification of T5 tails). C.-A.A., G.E., S.E., E.G., G.S. and C.B. analysed the data. C.B. wrote the paper with contributions from C.-A.A., G.E., E.G. All authors revised and approved the manuscript.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-017-02049-3>.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

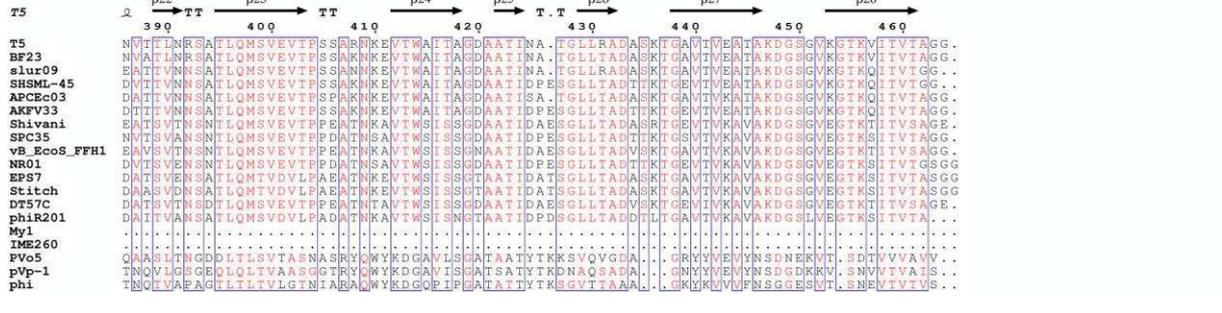
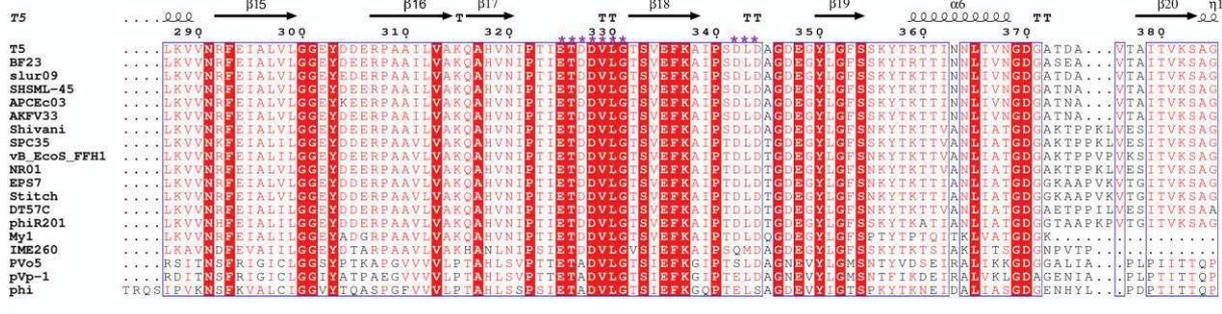
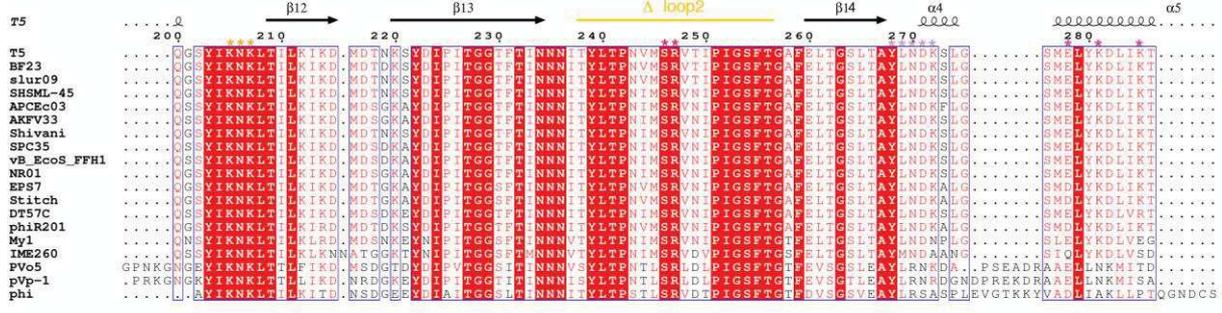
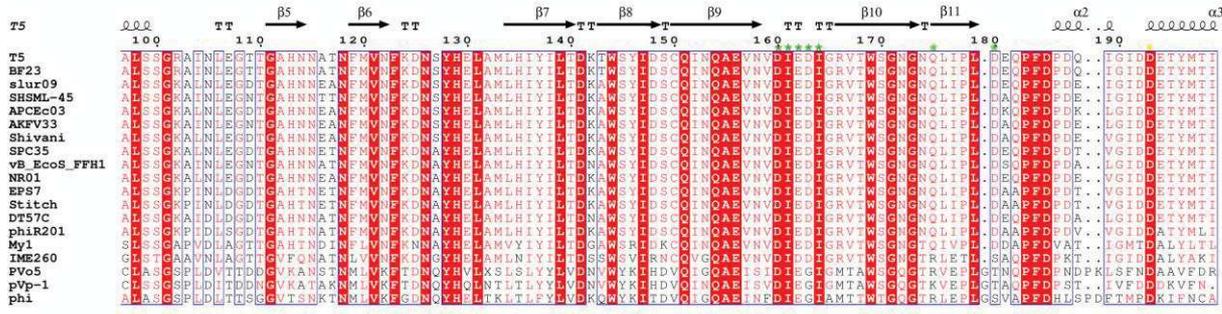
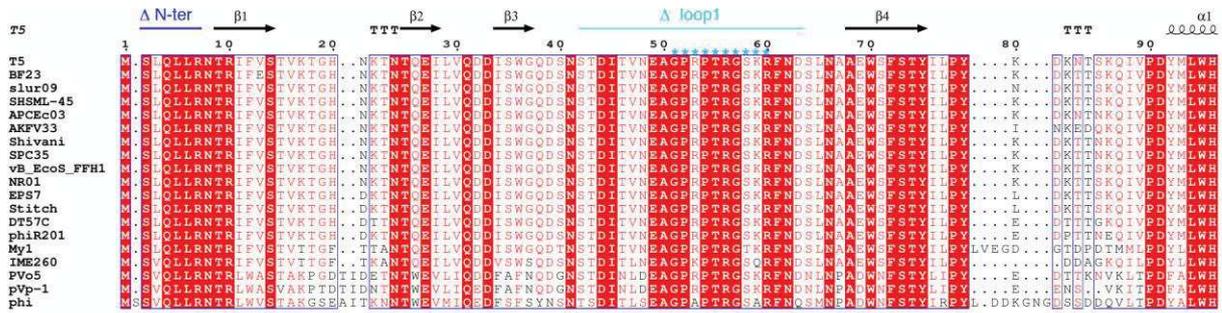


Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

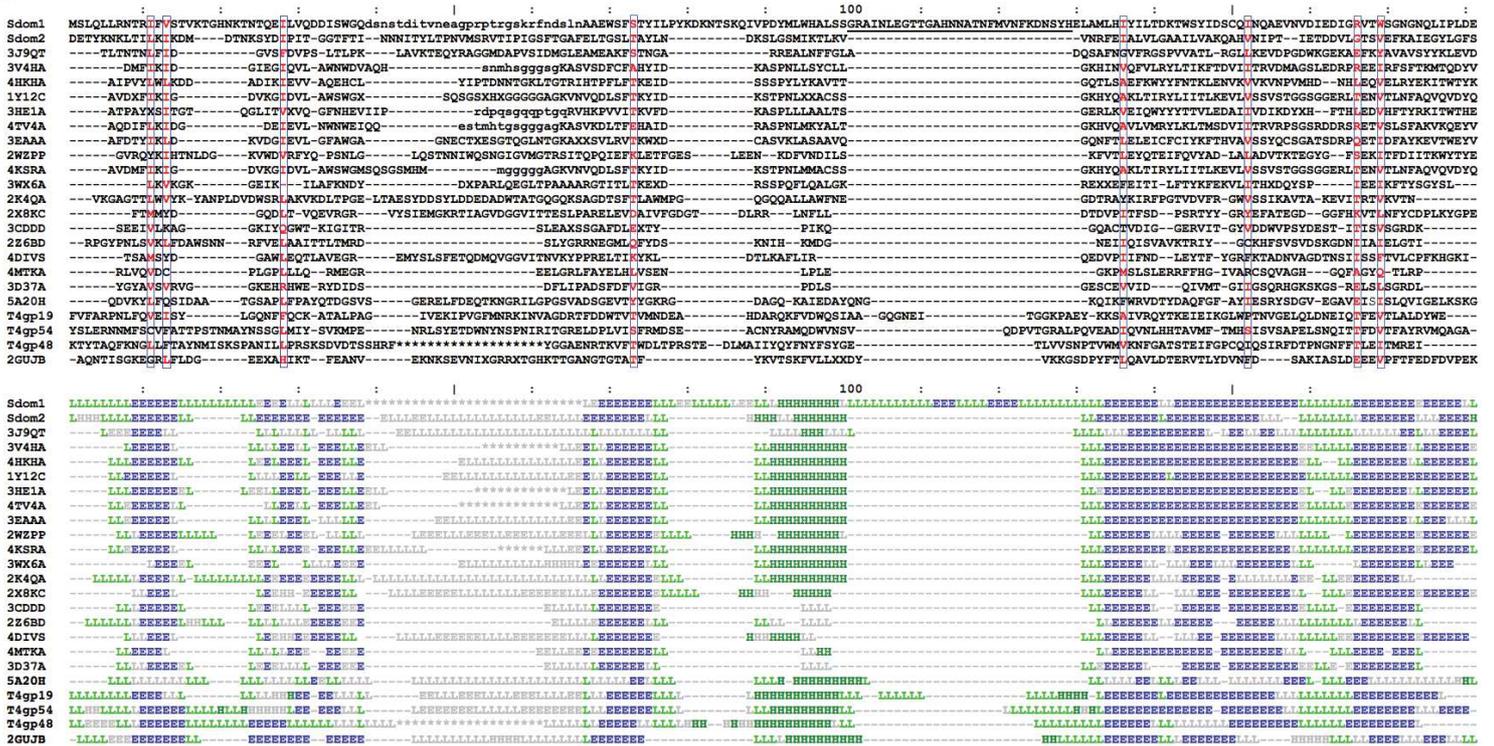
© The Author(s) 2017

5.2 Suppléments

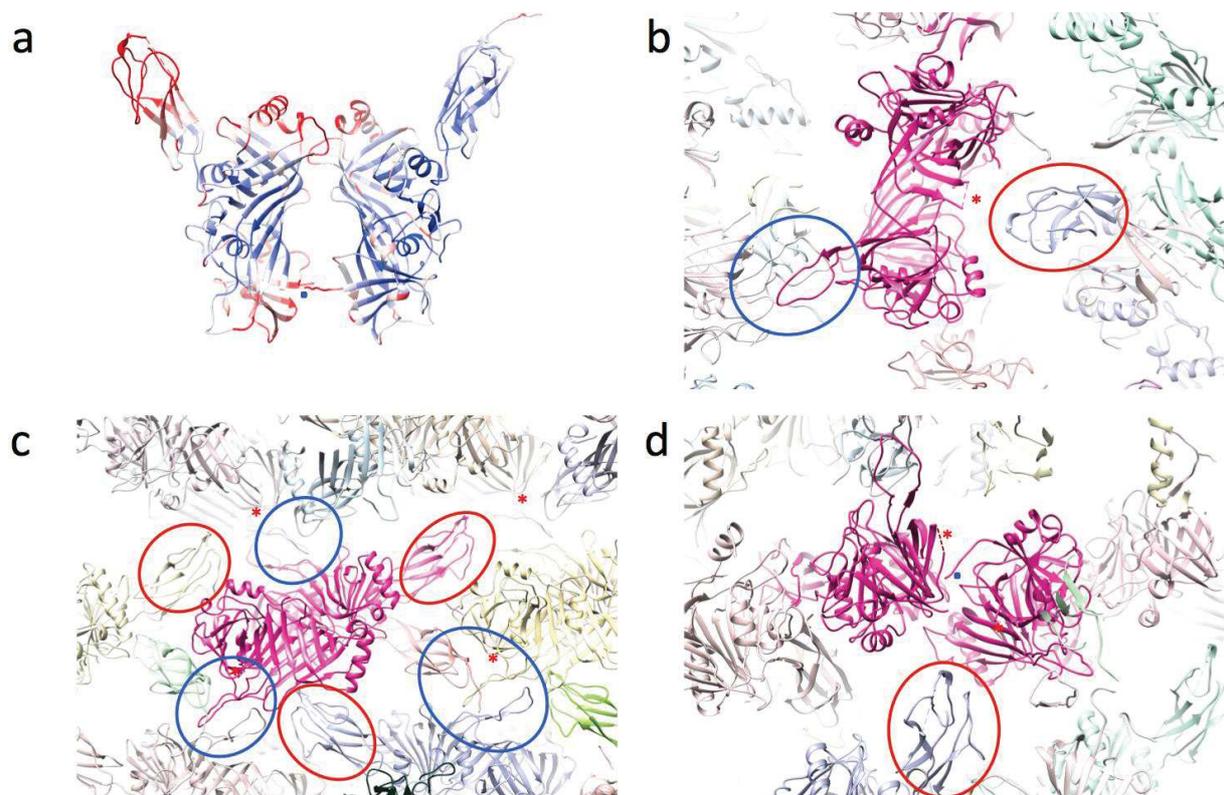
a



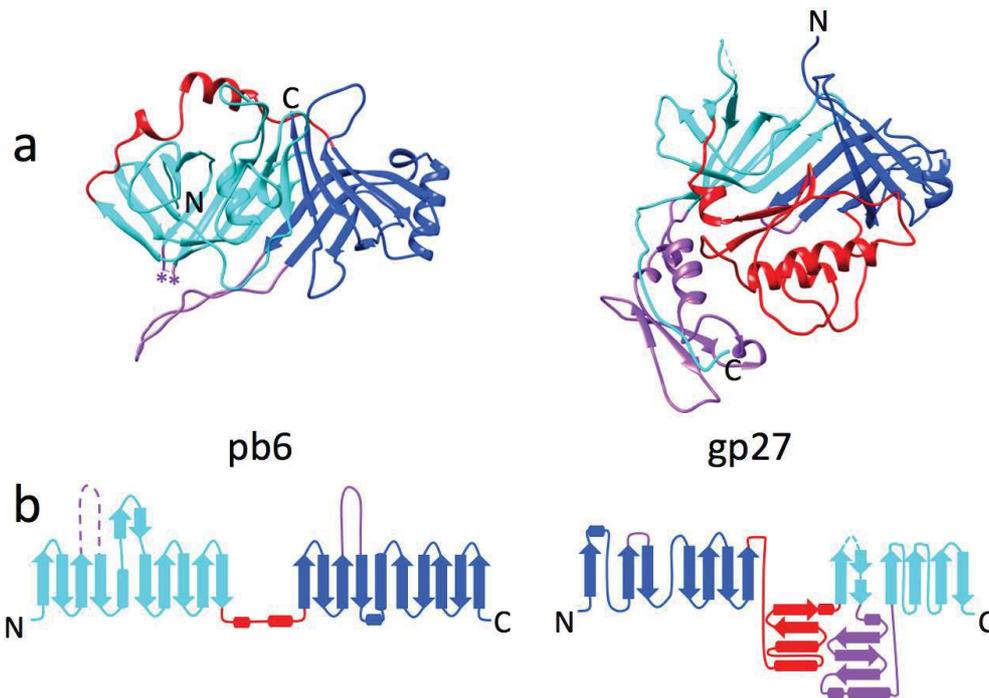
b



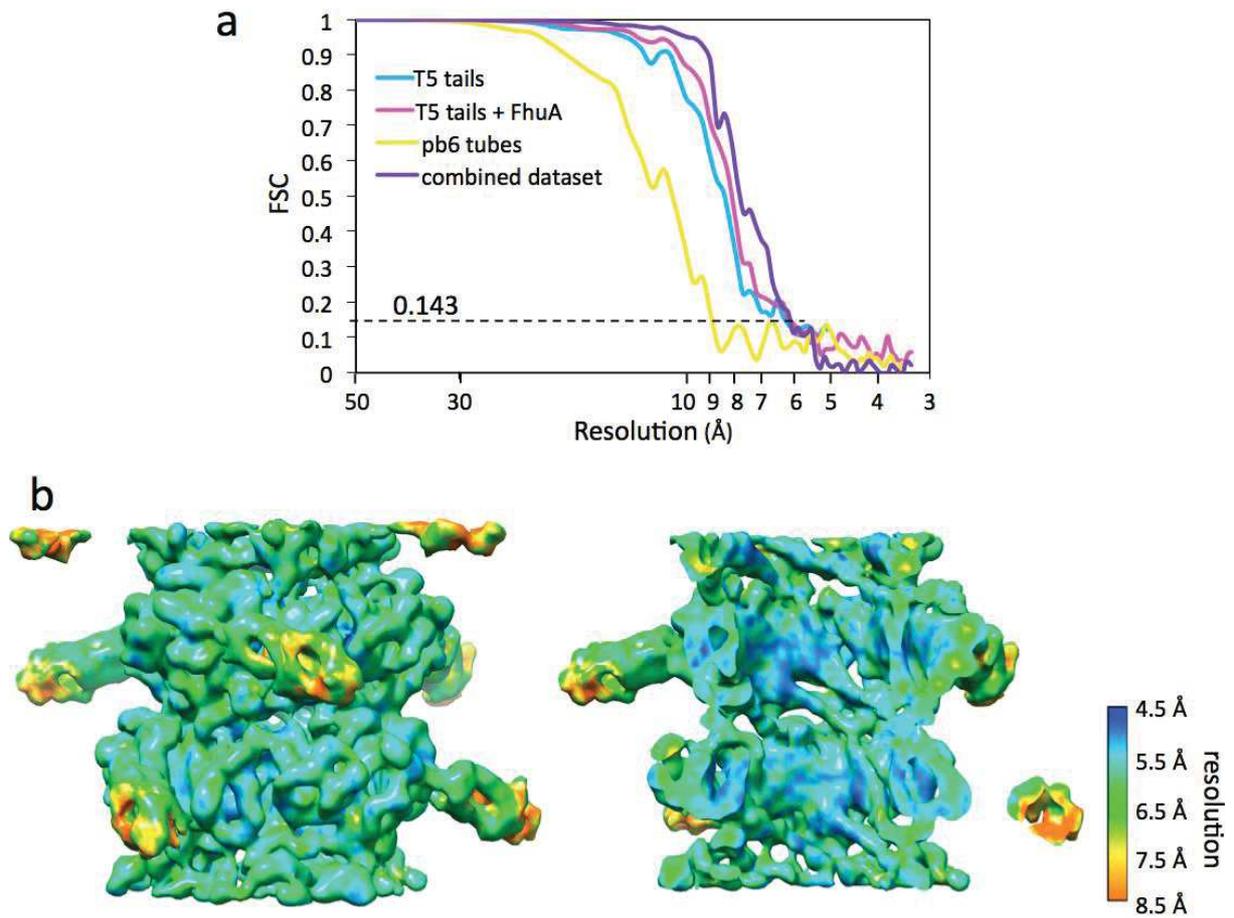
Supplementary Figure 1: a. Amino acid sequence alignment of TTPs from T5-like bacteriophages, displayed using EsPript² (%Equivalent option). The amino acids mutated to alanines in the different pb6 mutants are marked by asterisks with the same colour code as in Figure 4 and Supplementary Tables S4 and S5. b. Pairwise DALI³ sequence alignments of phage, T6SS and pyocin proteins on subdomain1 (Sdom1). The upper panel shows the amino acid sequence alignments, presented using EsPript² and combining the Risler, % Multalin and % Equivalent options (threshold = 0.7). All other options did not highlight any residue in the alignment. In lowercase are indicated the sequences of the long loops that are not ordered. In T4gp48, asterisks stand for the 55 aminoacid unstructured loop. The underlined insertion in subdomain 1 is the coloured magenta in Figure 1a. The lower panel presents the secondary structure assignments by DSSP (H: helix (green), E: strand (blue), L: coil, *: undefined, coloured when common). For clarity, insertions in the structural homologues of pb6 were not expanded. Sdom2: pb6 subdomain2; 3J9Q: inner tail tube protein of *Pseudomonas aeruginosa* pyocin; 3V4H: inner tube protein (Hcp) of the T6SS of *Yersinia pestis*; 4HKH: inner tube protein of the T6SS of *Escherichia coli*; 1Y12: inner tube protein of the T6SS of *Pseudomonas aeruginosa*; 4TV4: inner tube protein of the T6SS of *Burkholderia pseudomallei*; 3EAA: inner tube protein of the T6SS of *Edwardsiella tarda*; 2WZP: Distal tail protein of siphophage p2; 4KSR: inner tube protein of the T6SS of *Vibrio cholerae*; 3WX6: inner tube protein of the T6SS cluster 1 of *Burkholderia pseudomallei*; 2K4Q: tail tube protein of siphophage λ ; 2X8K: Distal tail protein of siphophage SPP1; 3CDD: baseplate hub protein (subdomain 1) of promyophage MuSO2 from *Shewanella oneidensis*; 2Z6B: baseplate hub protein (gp27, subdomain 1) of myophage T4; 4DIV: Distal tail protein of siphophage TP901.1; 4MTK: central spike (VgrG1, subdomain 1) of the T6SS of *Pseudomonas aeruginosa*; 3D37: tail tube protein of a pro-siphophage of *Neisseria meningitidis*; 5A20: tail tube protein of siphophage SPP1; T4gp19: inner tail tube protein of myophage T4; T4gp54 and 48: inner tail baseplate proteins of myophage T4; 2GUJ: tail tube protein of prophage PBSX of *Bacillus subtilis*.



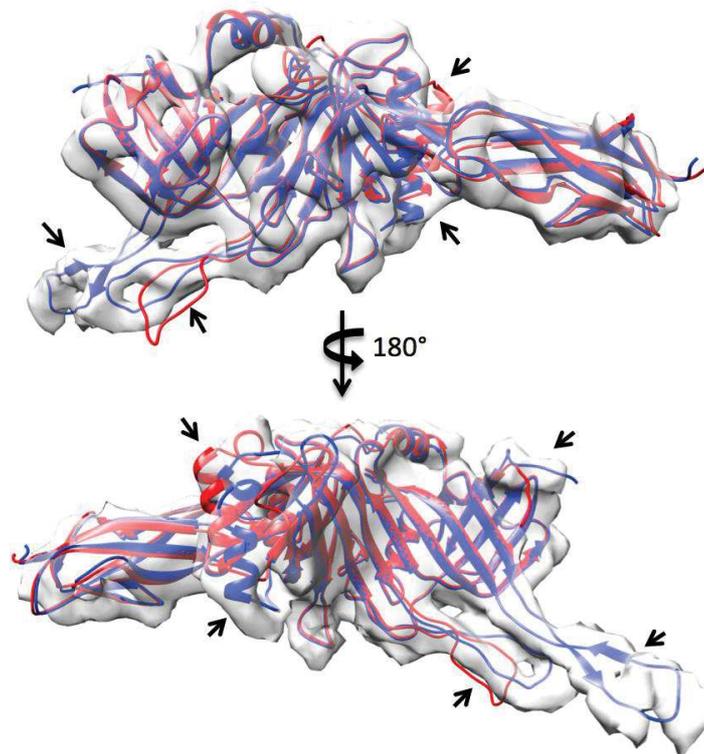
Supplementary Figure 2: pb6 crystal packing. **a.** Ribbon representation, coloured according to the average residue temperature factor (red: more flexible, blue: more rigid) of the asymmetric unit of the pb6 crystal. Both monomers are superimposable with an rmsd of 0.506 Å over 376 residues. This value drops down to 0.355 Å over 332 residues when the Ig-like domain and its linker are omitted. In one monomer, the Ig-like domain appears very flexible and less resolved (high temperature factor, left). This is due to it pointing towards a solvent channel in the crystal (circled red in b-d). In the other monomer, the Ig-like domain is much better resolved (lower temperature factor, right), as it takes part, also through the C-terminal His tag, in crystal contacts with the long loop β 13- β 14 (circled blue in b-c). **b-d.** Crystal packing of pb6. The long loop of subdomain 2 (β 13- β 14) of both monomers form crystal contacts, together with the Ig-domain (circled blue), thus constraining it. The long β 3- β 4 loop of subdomain 1 of both monomers points towards the solvent and is not ordered (indicated by a red asterisk in b-d), as is also observed in solution for the TTP of phage λ^4 . In gpV, the N-terminus is also disordered; in the pb6 crystal however, the N-terminus of one of the monomers also takes part in crystal contacts (blue dot in a and d) and is thus resolved. It still appears flexible, however, as shown by its high temperature factor (blue dot in a).



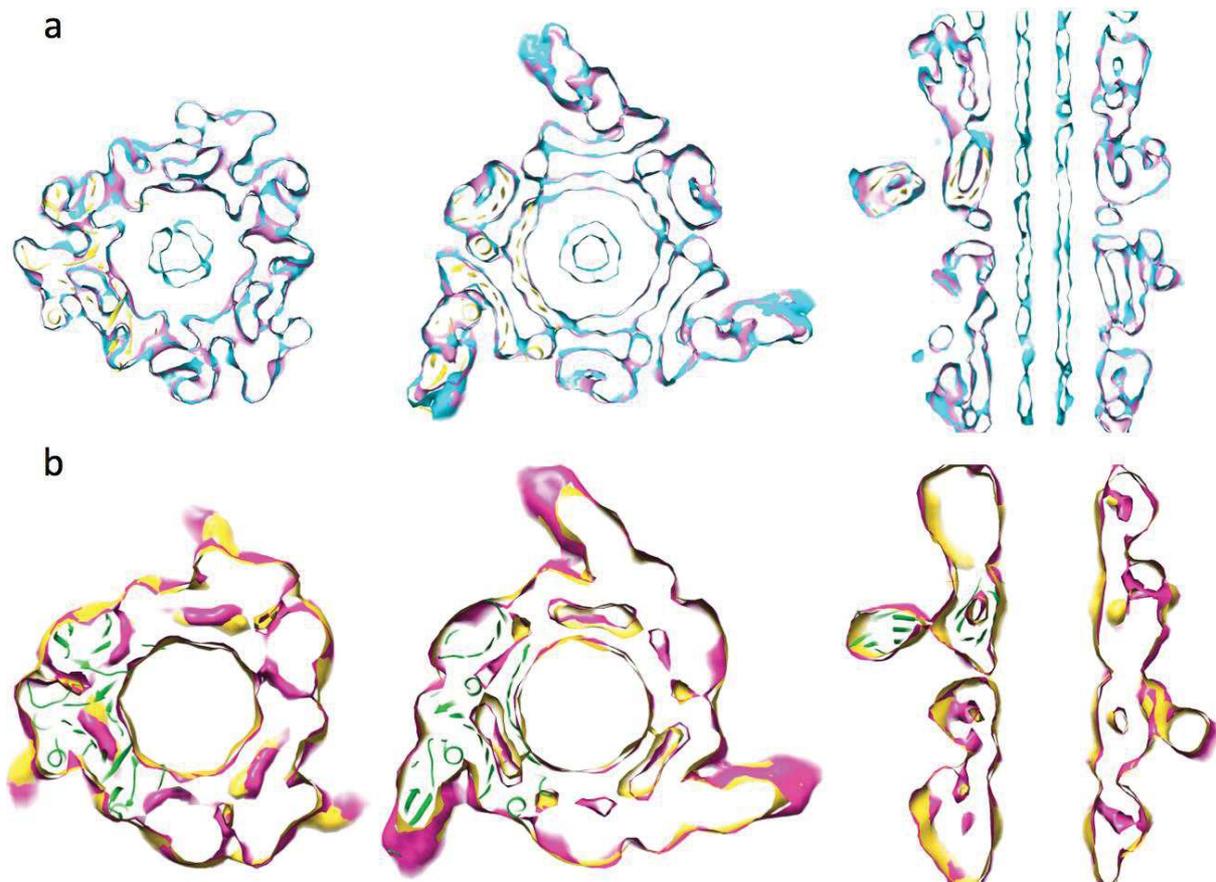
Supplementary Figure 3: Independent domain duplication in pb6 and gp27. **a.** Ribbon representation of pb6 (left) and T4 gp27 (right, PBD 2Z6B), in the same side view orientation. T4 gp27 was chosen as a representative of Myophage hub and T6SS spike proteins. In the latter, the C-terminus is fused to a spike domain. In pb6, the Ig-like domain and its linker have been removed for clarity. pb6 subdomains 1 and 2 and their linker are coloured as in Fig 1 (cyan, blue and red, respectively), and the long loops of the two subdomains are coloured purple; the start of the unstructured subdomain 1 long loop is indicated by asterisks. The same colour code was used for structurally homologous features in gp27. **b.** Topology diagram of pb6 (left) and gp27 (right). The equivalent of the pb6 long loop in gp27 subdomain 2 is short; in the other subdomain it is much longer and forms, together with the linker between the two subdomains, another domain that serves the function of the protein⁵. Furthermore, the link between the two subdomains is topologically different in the two proteins, as it connects the C-terminus of subdomain 1 to the N-terminus of subdomain 2 in pb6, and the N-terminus of subdomain1 to the C-terminus of subdomain2 in gp27. N and C: N- and C-terminus respectively.



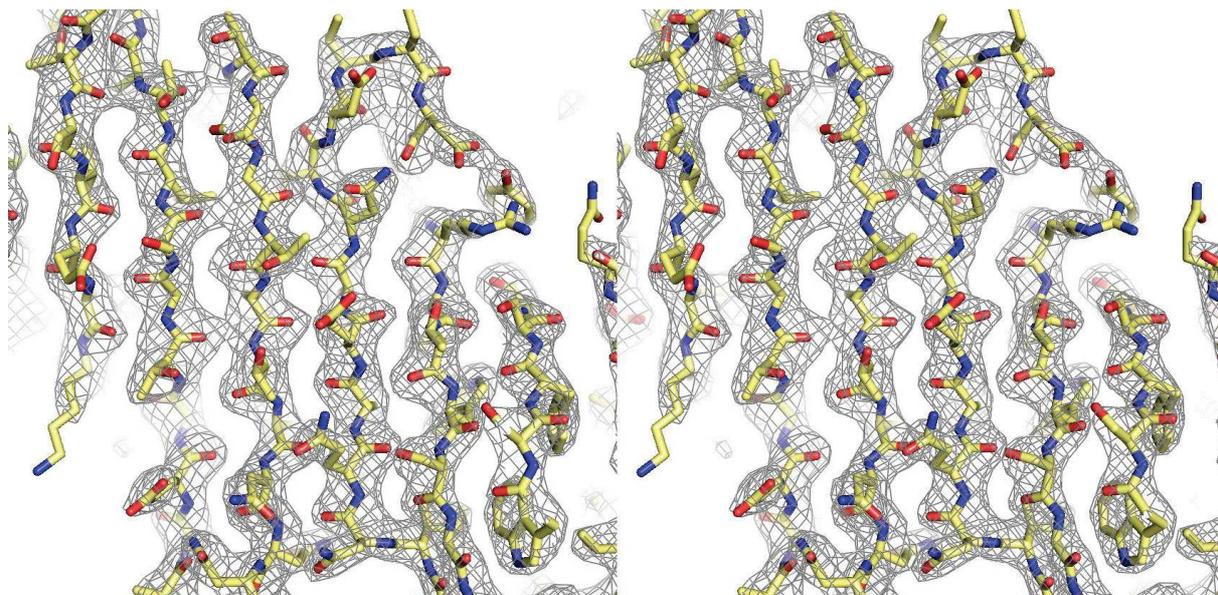
Supplementary Figure 4: EM data. **a.** Fourier shell correlation curves for reconstruction of the full (cyan) and empty (magenta) tails, of the combined dataset (purple) and of pb6 tubes (yellow). **b.** Local resolution of the combined density map shown on the exterior side view of the tube (left) or on its interior (right).



Supplementary Figure 5: Flexible fitting of pb6 crystals in a segment of the EM map of the tail. Ribbon representation of the crystal structure of the pb6 monomer rigidly (red) and flexibly (blue) fitted into a segment of the combined EM tail tube density. Main differences occur in the N-terminus, the 2 large loops and the $\alpha 5$ helix (black arrows).



Supplementary Figure 6: Comparison of full tails, empty tails and pb6 fibres. Three sections through the overlaid EM densities to allow comparison of the full (cyan) and empty (magenta) tail tube at 6 Å resolution (a), and the empty tail tube (magenta) and the tube fibres (yellow) at 9 Å resolution (b). Depicted as a yellow (a) and green (b) ribbon is the crystal structure of pb6 rigidly fitted into the density of the tubes. See also Supplementary Movies S1 and S2.



Supplementary Figure 7: Stereo image of a portion of the 2Fo - Fc electron density map displayed as a map with a contour of 1.5 σ .

Supplementary Table 1: Result of the DALI query on pb6 subdomains 1 (residue 1-192) and 2 (residue 193-374).

PBD	Subdomain 1				Subdomain 2				Protein
	Z-score	rmsd	lali	% id	Z-score	rmsd	lali	% id	
	6.1	3.3	98	7					pb6 subdomain 2
3J9Q	7.5	2.6	97	10	5.6	3.8	111	8	<i>Pseudomonas aeruginosa</i> Pyocin inner TTP
3V4H	7.3	2.8	95	13	7.2	3.0	102	9	<i>Yersinia pestis</i> T6SS Hcp1
4HKH	7.2	2.8	96	7	7.8	3.6	115	9	<i>Escherichia coli</i> T6SS Hcp1
3HE1	7.1	3.1	98	10	8.4	3.4	108	10	<i>Pseudomonas aeruginosa</i> T6SS Hcp3
2WZP	7.1	3.3	99	10	6.1	4.4	120	11	Siphophage p2 distal tail protein
1Y12	6.8	2.9	97	18	7.0	3.1	113	13	<i>Pseudomonas aeruginosa</i> T6SS Hcp1
4TV4	6.6	3.0	94	13	7.3	3.4	103	11	<i>Burkholderia pseudomallei</i> T6SS Hcp1
3EAA	6.5	2.9	97	10	7.1	3.4	117	14	<i>Edwardsiella tarda</i> T6SS Hcp1
4KSR	6.5	2.9	97	16	6.9	2.8	108	13	<i>Vibrio cholerae</i> T6SS Hcp1
3WX6	6.3	2.3	83	6	6.0	2.8	89	15	<i>Burkholderia pseudomallei</i> T6SS Hcp1
2X8K	5.8	3.6	88	13	5.3	3.9	100	14	Siphophage SPP1 distal tail protein
3CDD	5.7	2.8	76	7	5.0	4.2	89	13	<i>Shewanella oneidensis</i> promyophage MuSO2 baseplate hub protein
2Z6B	5.4	3.5	90	7	4.2	3.7	98	12	Myophage T4 baseplate hub protein (gp27)
4MTK	5.3	2.6	76	4	5.4	3.4	95	9	<i>Pseudomonas aeruginosa</i> T6SS central spike (VgrG1)
2K4Q	5.1	4.0	100	12	5.1	4.9	101	10	Siphophage λ TTP
1WRU	5.0	3.5	83	12	3.7	4.1	91	12	Myophage Mu baseplate hub protein
2GUJ	5.0	3.1	87	1	5.9	3.9	100	14	<i>Bacillus subtilis</i> prophage PBSX TTP
4DIV	4.5	3.0	86	5	3.5	4.2	94	9	Siphophage TP901.1 distal tail protein
3D37	4.1	3.8	84	8					<i>Neisseria meningitides</i> prosiphophage TTP
5A20	3.6	3.0	92	5	5.0	4.0	107	11	Siphophage SPP1 TTP
5IV5 T4gp19	7.8	4.3	112	5	4.9	4.1	114	5	Myophage T4 inner TTP
5IV5 T4gp48	5.7	3.5	101	5	6.2	3.5	111	10	Myophage T4 baseplate TTP
5IV5 T4gp54	7.4	3.2	109	6	5.9	3.9	119	13	Myophage T4 baseplate TTP

rmsd: root-mean-square deviation of C α atoms in the least-squares superimposition of the structurally equivalent C α atoms, lali: number of structurally equivalent residues, % id: percentage of identical amino acids over all structurally equivalent residues.

Supplementary Table 2: Cryo-EM data collection and image analysis statistics

Sample	Full tails (EMD-3689)	Empty tails pb6 (EMD-3689)	pb6 tubes (EMD-3691)
Magnification	15,500	23,000	23,000
Voltage (kV)	300	300	300
Defocus range (μm)	-1.0 / -7.0	-1.5 / -4.0	-1.5 / -5.0
Number of micrographs	61	84	53
Pixel size (unbinned) (\AA)	1.24	0.82	0.82
Number of frames per movie	40	40	40
Total dose per movie (e^- $/\text{\AA}^2$)	40	40	40
Frame alignment	Digital Micrograph	MotionCorr	Digital Micrograph
Symmetry imposed	3-fold	3-fold	3-fold
Number of helical segments picked	32891	50014	19392
Number of segments in final reconstruction	24243	22704	12244
Resolution (\AA) FSC = 0.143	6.2	5.8	8.8

Supplementary Table 3: Summary of the performed mutants.

Name	Mutation (protein)	Tube formation	Fraction of soluble protein (%)
pb6-full length		+	21 ± 4 (n = 5)
pb6-Ig	p.S2_D374del	nr	nr
pb6-ΔIg	p.A375_464del	+	18
pb6-ΔNter	p.S2_N8del	-	96
pb6-Δloop1	p.S42_D63del	-	92
pb6-Loop1A	p.G51A_P52A_R53A_P54A_T55A_R56A_G57A_S58A_K59A_R60A	ni	97
pb6-160A	p.D160A_I161A_E162A_D163A_I164A	ni	91
pb6-QD	p.Q175A_D180A	+	24
pb6-D	p.D193A	+	26
pb6-Δloop2	p.T238_A256del	-	91
pb6-KNK	p.K205A_N206A_K207A	+	69
pb6-SR	p.S246A_R247A	+	74
pb6-270A	p.Y268A_L269A_N270A_D271A_K272A	ni	82
pb6-E	p.E278A	+	47
pb6-K1	p.K281A	+	15
pb6-K2	p.K285A	+	48
pb6-3K	p.E278A_K281A_K285A_K288A	+	60
pb6-330A	p.E325A_T326A_D327A_D328A_V329A_L330A_G331A_D342A_L343A_D344A	ni	74

p. indicates point protein mutations. X_Ydel indicates an X-Y deletion, nr: not relevant, ni: not investigated. +: tube formation and -: no tube formation, as determined by negative stain EM of the insoluble fraction after cell lysis, in buffer conditions that favour tube formation. The fraction of soluble protein was determined by comparing the amount of soluble and insoluble pb6 in bacterial lysate on SDS-PAGE (Fig. 4b) using the gel analysis tool in ImageJ¹. Mutants are colour coded as in Figure 4 and Supplementary Figure 1a.

Supplementary Table 4: Primers used for generating the pb6 mutants.

Mutant	Primer1	Primer2
pb6-Ig	CTTTAAGAAGGAGATATACATATGGC TGTAACCGCAATTACTG	CAGTAATTGCGGTTACAGCCATATGTATA TCTCCTTCTTAAAG
pb6-ΔIg	GGTGACGGTGCTACTGATGCTGGCGC TGGCCATCATCATCATCATTGAG CTGTAACCGCAATTACTG	CAGTAATTGCGGTTACAGCTCAATGATGA TGATGATGATGGCCAGCGCCAGCATCAG TAGCACCGTCACC
pb6-ΔNter	CTTTAAGAAGGAGATATACATATGAC TCGAATCTTCGTGTCTACG	CGTAGACACGAAGATTCGAGTCATATGT ATATCTCCTTCTTAAAG
pb6-Δloop1	CTTGGGGTCAGGACAGTAACTCTTTA AACGCAGCTGAGTGG	CCACTCAGCTGCGTTTAAAGAGTTACTGT CCTGACCCCAAG
pb6-Loop1A	CAGATATTACTGTTAATGAAGCTGCG GCGGCGGCGGCGGCGGCGGCGGCGG CGTTTAACTACTCTTTAAACGCAG	CTGCGTTTAAAGAGTCGTTAAACGCCGCC GCCGCCGCCGCCGCCGCCGCCGAGCTTC ATTAACAGTAATATCTG
pb6-160A	GATTAATCAGGCAGAAGTTAACGTTG CGGCGGCGGCGGCGGGTTCGTGTAAC TTGGTCTG	CCAGACCAAGTTACACGACCCGCCGCCG CCGCCGCAACGTTAACTTCTGCCTGATTA ATC
pb6-QD	CTTGGTCTGGTAATGGTAACGCGCTA ATCCCACTGGCGGAGCAACCATTGTA TCCAG	CTGGATCAAATGGTTGCTCCGCCAGTGGG ATTAGCGCGTTACCATTACCAGACCAAG
pb6-D	CCAGATCAAATAGGTATTGATGCGGA AACTTATATGACTATCCAG	CTGGATAGTCATATAAGTTTCCGCATCAA TACCTATTTGATCTGG
pb6-Δloop2	CTTTTACTATTAACAACAACATTGGT GCTTTTGAATTAAGTGG	CCAGTTAATTCAAAAGCACCAATGTTGTT GTTAATAGTAA
pb6-KNK	CTATCCAGGGTTCTTACATTGCGGCG GCGCTAACGATCTTGAAGATCAAG	CTTGATCTTCAAGATCGTTAGCGCCGCCG CAATGTAAGAACCCTGGATAG
pb6-SR	CCTAACACCTAATGTAATGGCGGCGG TTACTATTCCAATTGGTTC	GAACCAATTGGAATAGTAACCGCCGCCA TTACATTAGGTGTTAGG
pb6-270A	GAATTAAGTGGTTCATTAAGTCTGCTGC GGCGGCGGCGGCGGTCTCTTGGTTCTA TGGAAGTCTG	CAGTTCCATAGAACCAAGAGACGCCGCC GCCGCCGAGCAGTTAATGAACAGTTA ATTC
pb6-E	GATAAATCTCTTGGTTCTATGGCGCT GTATAAAGATCTTATCAAAAC	GTTTTGATAAGATCTTTATACAGCGCCAT AGAACCAAGAGATTTATC
pb6-K1	CTTGGTTCTATGGAAGTGTATGCGGA TCTTATCAAACTCTGAAG	CTTCAGAGTTTTGATAAGATCCGCATACA GTTCCATAGAACCAAG
pb6-K2	GGAAGTGTATAAAGATCTTATCGCGA CTCTGAAGGTAGTTAACCG	CGGTAACTACCTTCAGAGTCGCGATAAG ATCTTTATACAGTTCC
pb6-3K	GATAAATCTCTTGGTTCTATGGCGCT GTATGCGGATCTTATCGCGACTCTGG CGGTAGTTAACCGCTTTGAAATC	GATTTCAAAGCGGTTAACTACCGCCAGA GTCGCGATAAGATCCGCATACAGCGCCA TAGAACCAAGAGATTTATC
pb6-330A	GCGCACGTTAATATTCCTACTATTGC GGCGGCGGCGGCGGCGGCGACTTCA GTAGAATTTAAGGCTATTCCGTCAGC GGCGGCGGCTGGTACGAAGGTTAC TTAG	CTAAGTAACCTTCGTCACCAGCCGCCGCC GCTGACGGAATAGCCTTAAATTCTACTGA AGTCGCCGCCGCCGCCGCCGCCGCAATA GTAGGAATATTAACGTGCGC

Mutants are colour coded as in Figure 4 and Supplementary Figure 1a.

Supplementary references

1. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
2. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320-324 (2014).
3. Holm, L. & Laakso, L. M. Dali server update. *Nucleic Acids Res.* **44**, W351-355 (2016).
4. Pell, L. G., Kanelis, V., Donaldson, L. W., Howell, P. L. & Davidson, A. R. The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 4160–4165 (2009).
5. Kanamaru, S. *et al.* Structure of the cell-puncturing device of bacteriophage T4. *Nature* **415**, 553–557 (2002).

Points clés

Résumé chapitre 5

- ❑ La structure cristallographique de pb6 a été élucidé confirmant que le tube de T5 est formé par un élément homologue aux tubes des autre *Siphoviridae*, *Myoviridae* ainsi que le tube de SST6 et de la pyocine-R.
- ❑ Le tube T5 ne change pas de conformation après interaction avec la protéine FhuA, le modèle de transmission de l'information de fixation de à la capsidie proposé pour SPP1 n'est donc pas valide.
- ❑ Un nouveau modèle est proposé : lors de l'interaction avec FhuA, la protéine pb2 est exposée à la membrane et provoque sa propre éjection. Ce modèle suppose que la protéine pb2 est dans un état métastable et interagit assez peu avec la TTP pour que son réarrangement à travers le périplasma soit suffisamment énergétiquement avantageux pour se faire de manière autonome. L'éjection de pb2 provoquera un changement de conformation au niveau de l'interface capsidie-queue qui permettent l'éjection d'ADN.

Chapitre 6

Tube de la queue du phage T5 : préparation d'échantillons pour la RMN & perspectives

L'étude de pb6 par RMN s'est faite en collaboration avec Paul Schanda. Lui et son ancien étudiant Vilius Kurauskas, ont assurés la collecte et l'analyse des spectres RMN.

6.1 Préparation des échantillons

Les échantillons de pb6 préparés pour la RMN, mutants solubles compris, ont été purifiés sur la base du même protocole que pour nos autres études (Fig. 6.1). Les conditions particulières de cultures sont détaillées dans la partie 3.2.1.3.

6.1.1 pb6 entière

Les échantillons de pb6 sous forme de fibre et de monomère avec un marquage ^{13}C , ^{15}N et deutéré ont été les premiers à être produits.

La production / purification de fibres de pb6 pour être étudiées en RMN du solide (solid state NMR, ssRMN) posa peu de problèmes. Des étapes de lavage des fibres purifiées ont été ajoutées au protocole initial car des signaux spécifiques des sucres étaient visibles dans nos premières expériences (Fig. 6.3-b). A une seule occasion, malgré leur grande stabilité, une dégradation des fibres a été observée en quelques jours (Fig. 6.2-b). Après cet incident, que nous avons attribué à une contamination, de petites quantités d'azide (\sim nM) furent systématiquement ajoutées à l'échantillon final.

La préparation de la protéine soluble fût beaucoup moins évidente, et les premières tentatives ne permirent pas de produire un échantillon suffisamment concentré. La difficulté à produire cet échantillon fût l'une des motivations des expériences réalisées pour comprendre la polymérisation de pb6 (cf. 4.2.1). Malheureusement, les cultures deutérées affectent la polymérisation de pb6 et l'adaptation des conditions de culture et de lyse ne permit pas d'améliorer la proportion de protéine soluble obtenue. Un des précipités, isolé du non-retenu de la chromatographie d'affinité, présentait des particularités intéressantes : il était constitué uniquement de fibres de petites tailles, parfois circularisées, et d'anneaux trimériques (Fig. 6.2-a). Remarque : cet échantillon a été utilisé en RMN.

Par ailleurs, la protéine a précipité à plusieurs reprises pendant la purification, la concentration sur centricon ou lors de la conservation de la protéine. La précipitation de la protéine a été observée avec ou sans formation de fibres (Fig. 6.2-c&d).

Un échantillon avec un marquage ^{13}C , ^{15}N et deutéré à 150 μM a néanmoins été produit avec une pureté de 77%. Une comparaison d'un spectre 2D de cet échantillon avec la même protéine non deutérée (marquage ^{13}C et ^{15}N) (Fig. 6.3-a) illustre l'immense gain de sensibilité lié à la deutération sur un échantillon de cette taille.

Une tentative d'étude de pb6 au sein de son assemblage cristallin par ssRMN a été tentée car ce type d'approche aurait pût nous apporter des informations intéressantes sur les différences entre protéine en solution et au sein du cristal (exemples : [Kurauskas 17, Ma 15]). Mais les cristaux de pb6, trop fragiles, ne résistèrent pas à la centrifugation nécessaire pour les placer dans la sonde RMN et donnèrent un signal similaire à de la protéine précipitée.

Des échantillons de fibres et de monomère soluble ont également été produits avec un marquage particulier des méthyles (échantillon soluble à une concentration de 350 μM). Ce marquage consiste en un marquage ^{13}C ^{15}N ^2H de tous les résidus sauf des Isoleucines, Leucines et Valines dont les chaînes latérales qui sont ^{12}C ^{14}N ^2H à l'exception du carbone 1 des isoleucines et des méthyles proR des leucines et valines. Ces derniers sont marqués $^{13}\text{C}^1\text{H}$ ($\text{I}^1/\text{LV}^{\text{proR}}$). Ce marquage des méthyles permet d'alléger les spectres de grosses protéines et d'augmenter la sensibilité du reste des signaux. Cependant, ce marquage ne permet pas de réaliser des expériences d'attribution (Ile, Leu et Val non marqués).

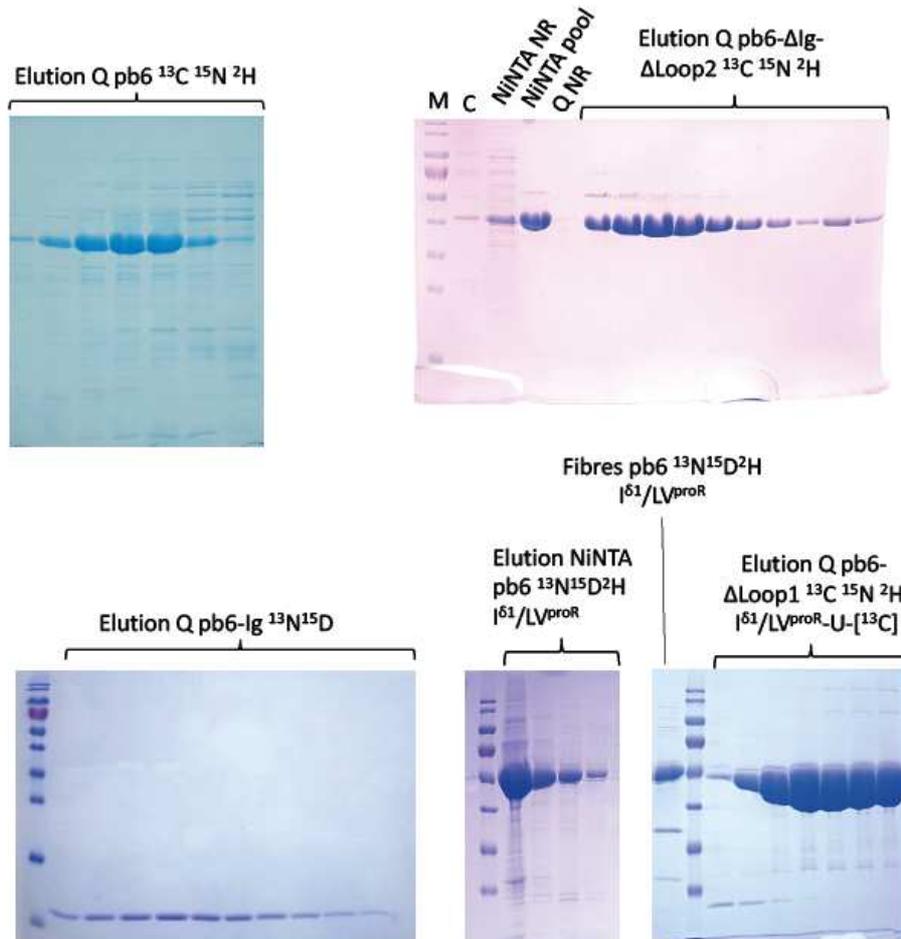


Figure 6.1 : Purifications des différents échantillons destinés à la RMN.

NiNTA NR : non-retenue de la chromatographie d'affinité au Nickel ; Q NR : non-retenue de la colonne échangeuse d'anions ; NiNTA pool : élution de la chromatographie d'affinité ; élution Q : élution de la colonne échangeuse d'anions. Remarque : l'échantillon pb6 $^{13}\text{N}^{15}\text{D}^2\text{H}$ $^{161}/\text{LV}^{\text{proR}}$ n'a pas été purifié sur colonne échangeuse d'ions pour limiter les risques de précipitation.

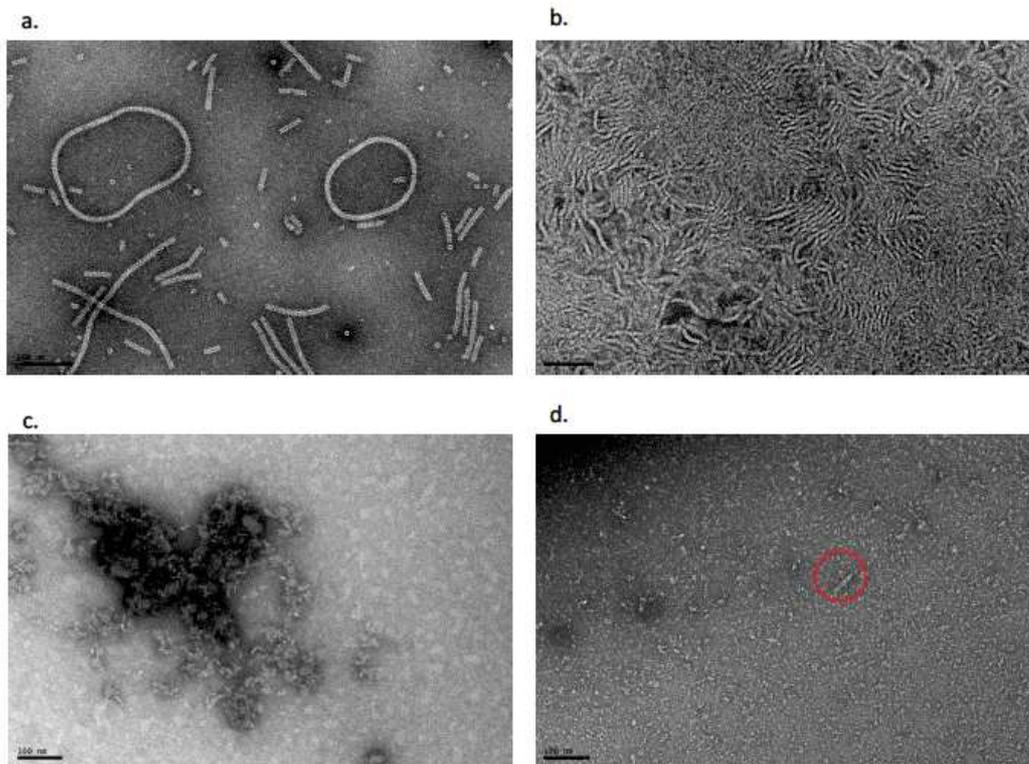


Figure 6.2 : microscopie électronique à coloration négatives de différentes formes de pb6 précipités pendant la préparation d'échantillons marqués. a. fibres de pb6 s'étant formées lors de la chromatographie d'affinité au nickel et extraites du non-retenue. b. Fibres de pb6 purifiés dégradées. c. Aggrégats de pb6 sans fibres. d. Aggrégats de pb6 avec de petites fibres (rond rouge).

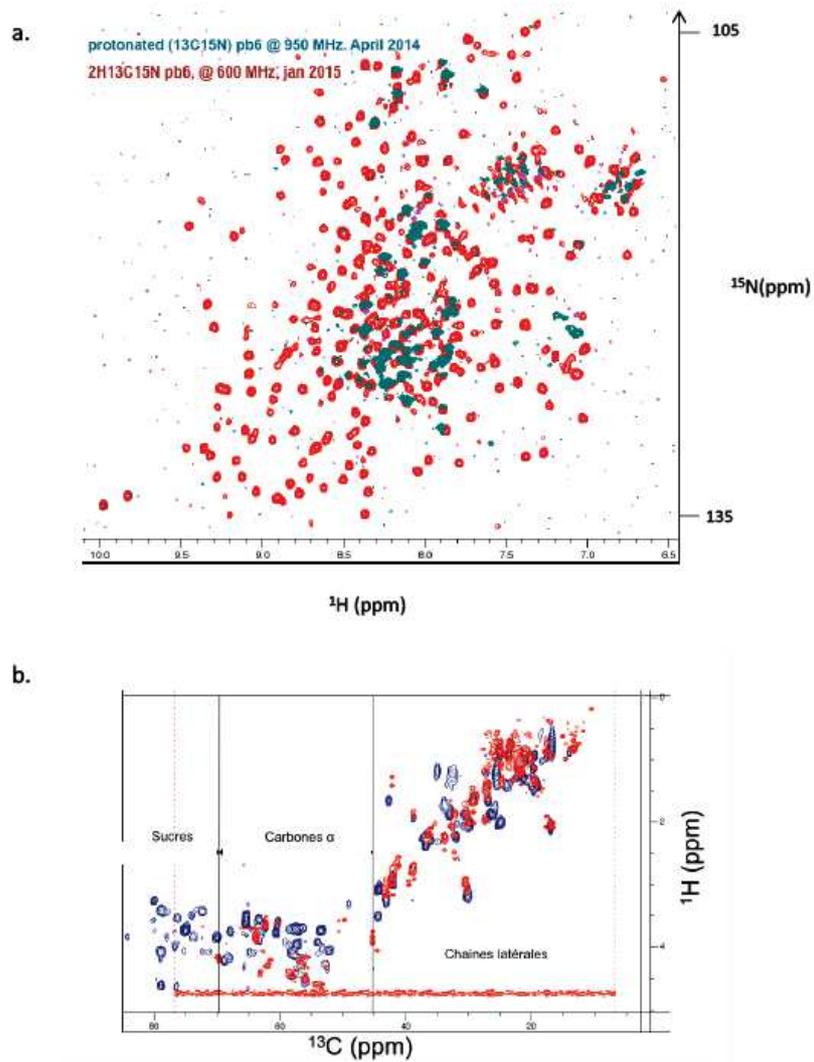


Fig 6.3 : comparaisons de spectre RMN 2D. a. comparaison de pb6 monomère $^{13}\text{C} \ ^{15}\text{N}$ (vert) et de pb6 monomère $^{13}\text{C} \ ^{15}\text{N} \ ^2\text{H}$. En absence de deutération, seules les boucles flexibles de la protéine sont visibles. b. Comparaison de spectres 2D de fibres (bleu) et de monomères (rouge) de pb6. On observe la superposition de plusieurs pics entre confirmant que les résidues correspondant ont le même environnement chimique.

6.1.2 Mutants de pb6

Des mutants de pb6 incapables de former des fibres (présentés dans l'article) ont été utilisés afin de produire des échantillons solubles plus concentrés et pour « fractionner » l'attribution des signaux RMN. Trois mutants ont été produits en utilisant des stratégies de marquages différentes, ils ont tous été purifiés à des puretés >80% et à des concentrations de l'ordre du millimolaire.

Le mutant pb6-Ig-Loop2 cumule 2 délétions : il ne porte plus le domaine de décoration Ig-like et l'une des boucles assurant des contacts majeurs pour la formation du tube (résidus 238 à 256). Cet échantillon a été produit de façon similaire au monomère entier (marquage ^{13}C , ^{15}N , ^2H).

Le mutant pb6-loop1 n'a plus la seconde des grandes boucles impliquée dans l'assemblage (résidus 42 à 63). Un marquage particulier a été utilisé pour cet échantillon (cf. 3.2.1.3) : la totalité de la protéine est marquée ^{13}C ^{15}N ^2H à l'exception des méthyles 1 des isoleucines et des proR des leucines et valines qui sont marqués $^{13}\text{C}^1\text{H}$ ($\text{I}^1/\text{LV}^{\text{proR}}\text{-U-}[^{13}\text{C}]$). Ce marquage permet de faciliter l'attribution de ces acides aminés et donc de l'ensemble de la chaîne (chaîne peptidique marquées mais contribution des chaînes latérales des Ile, Leu et Val retirées).

Enfin, le mutant pb6-Ig consiste en le seul domaine Ig-like. Cet échantillon a été marqué ^{13}C ^{15}N , sans deutérium car sa petite taille (10 kDa) n'en justifiait pas l'usage.

6.2 Apport des expériences et perspectives

Plusieurs spectres 2D, 3D et 4D ont été collectés sur ces différents échantillons pour permettre l'attribution des spectres de pb6 monomère et fibre ainsi qu'une comparaison des déplacements chimiques. Des expériences de dynamique de la protéine sous ses 2 formes ont également été réalisées et sont en cours d'analyse. L'attribution du spectre d'une protéine aussi grande (50,3 kDa), même avec l'aide des différents mutants, représente un travail colossal qui a été commencé par Vilius Kurauskas à la fin de son doctorat et est poursuivi par Paul Schanda. Ci-après sont présentés un certain nombre de résultats préliminaires, dont des données sur le domaine Ig-like de pb6, qui sont intégrés dans une publication de méthodologie d'attribution de spectres de ssRMN (cf. Annexe), ainsi qu'une discussion des résultats qui seront produits.

6.2.1 Domaine Ig-like

Les spectres du mutant pb6-Ig ont été entièrement attribués en soluble. De plus, l'attribution de ce domaine sur des échantillons de fibres a été réalisée à 92% à

partir des spectres de fibres en ssRMN ([Fraga 17] ; cf. Annexe) grâce à des spectres 3D (HcocaNH & HcacoNH) et 4D (hCOCANH & hCOCAcoNH). Une stratégie proche avait été appliquée à la TTP de SPP1 [Zinke 17].

L'attribution du domaine Ig-like de pb6 a pu être transférée sur les spectres de la protéine entière soluble (Fig. 6.4-b), seules quelques différences sont observées pour les résidus du lien entre le corps de la protéine et le domaine Ig-like. La comparaison des spectres du domaine Ig-like seul en solution et dans les fibres en ssRMN révèle des variations de déplacements chimiques qui se localisent vers la zone de ce domaine en contact avec le tube (Fig. 6.4-a). La polymérisation de pb6 provoque des changements dans l'environnement chimique de la partie du domaine Ig-like de pb6 la plus proche de l'axe du tube.

Cette observation est à rapprocher du fait que ces domaines de « décoration » sont particulièrement bien résolus dans les reconstructions de microscopie électronique, indiquant qu'ils sont contraints dans une conformation particulière par rapport au tube. Il est difficile d'imaginer une pression de sélection à conserver ce domaine fixé rigidement dans une direction si son seul rôle est de donner au virion une surface d'adhésion non-spécifique. Il est possible que ce domaine joue un rôle plus complexe dans le comportement de la queue en solution, affectant peut être la rigidité de l'assemblage, comme observé dans [Auzat 08] pour SPP1. Les différences de déplacements chimiques entre monomère et fibres sont cependant assez faibles et l'analyse de la protéine entière devra être réalisée pour discuter plus longuement ce type de comparaisons.

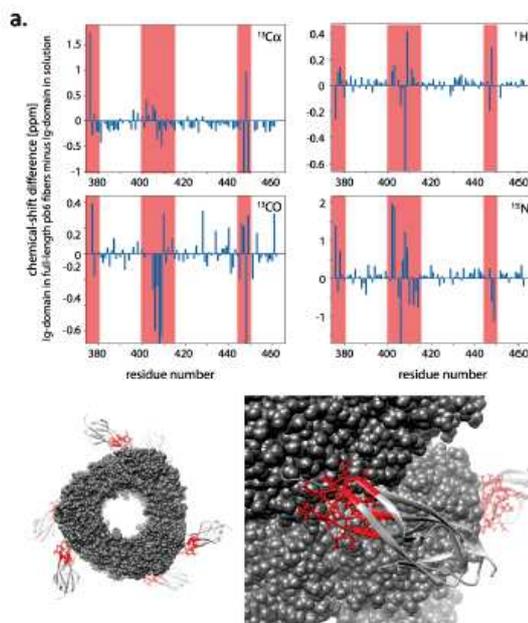
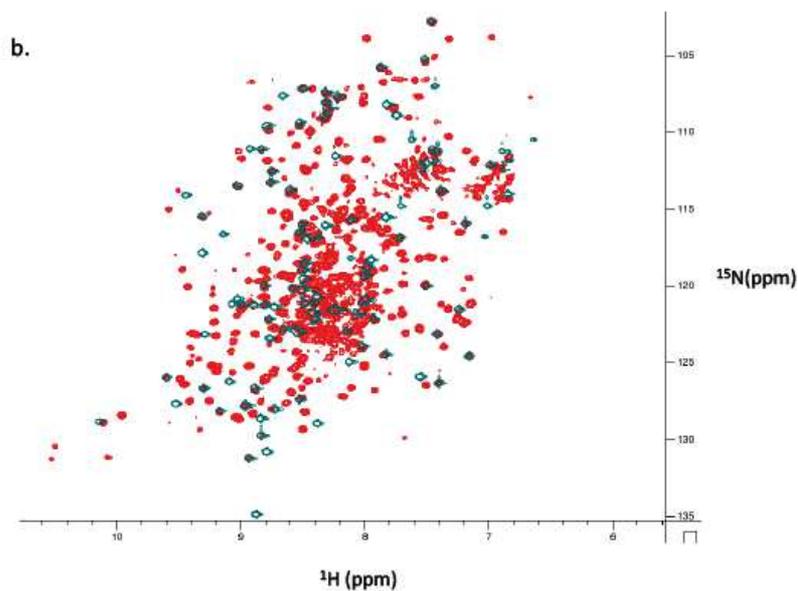


Figure 6.4 : Données RMN du domaine C-terminale de pb6. a. Figure issue de Fraga, 2017 (cf. annexes). Représentation de la variation du déplacement chimique par résidus entre les spectres de pb6-Ig soluble et de fibres. Les structures avec une variation significative sont représentés en rouge sur la structure. **b.** Comparaison de spectres 2D du domaine *Ig-like* en solution (vert) et de pb6 monomères (rouge).



6.2.2 Dynamique de pb6

Des expériences de dynamiques ont été réalisées sur le monomère et les fibres pour comparaison mais nécessitent que l'attribution soit terminée pour être interprétées.

La comparaison des spectres 2D permet, à l'œil, de dire que moins de résidus se trouvent dans une conformation de boucle flexible quand la protéine est polymérisée. Des informations plus précises pourront être obtenues une fois l'attribution terminée.

Les spectres du mutant pb6-Ig-Loop1 ont permis une attribution de 85% des résidus (Fig. 6.5). De façon surprenante, un certain nombre de résidus faisant partis du feuillet du sous-domaine 1 de la protéine n'apparaissent pas sur les spectres. Des expériences de dynamique de cet échantillon (mesure CPMG de la dispersion R2) confirme que ces résidus alternent entre différentes conformations. Ces résultats suggèrent que le feuillet ne se forme pas en solution mais est stabilisé lors de la formation des tubes, ce qui avait été proposé pour la TTP de SPP1 [Langlois 15].

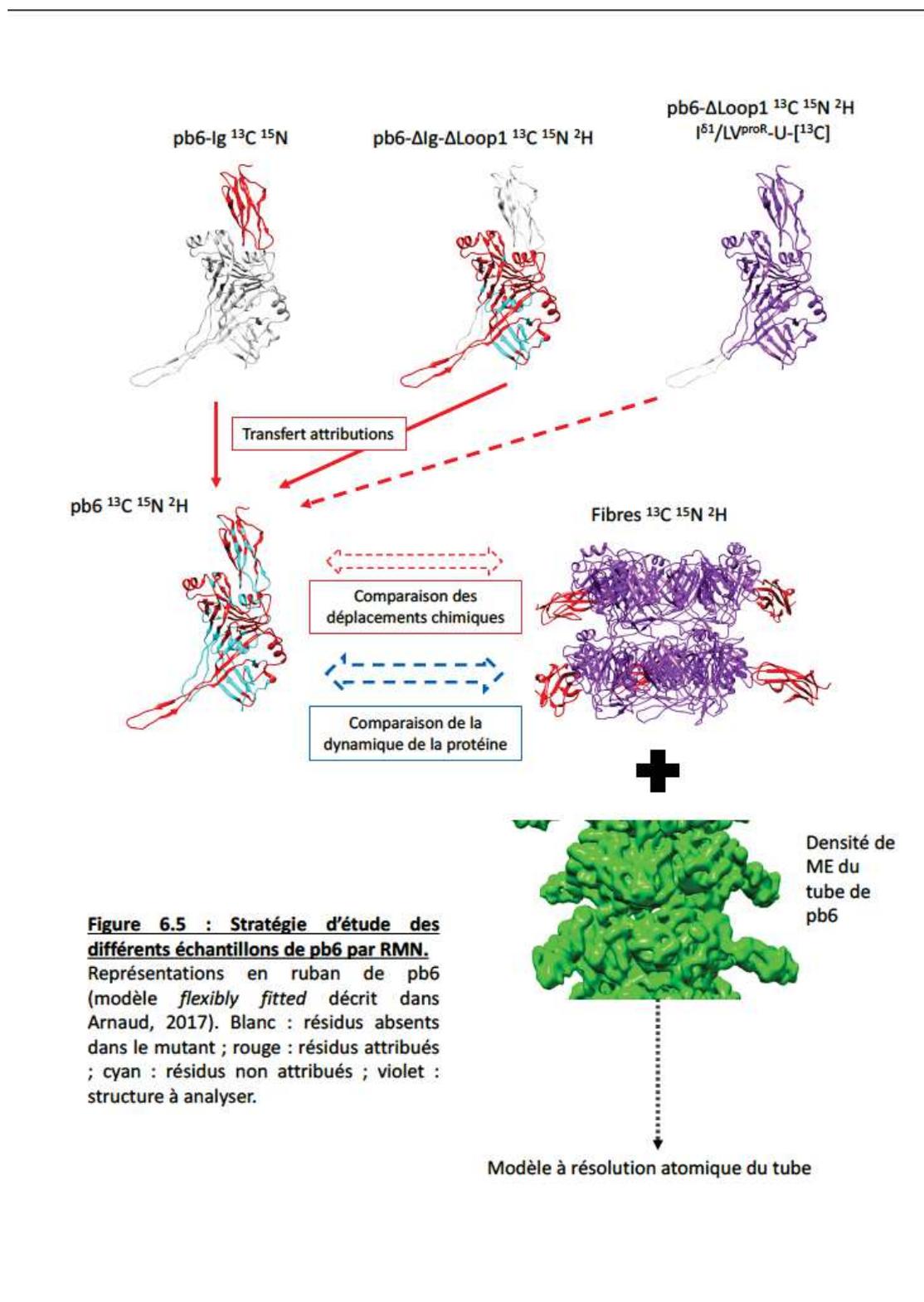
Une partie des attributions de pb6-Ig-Loop1 a pu être directement transférée sur les spectres de pb6. En y ajoutant le transfert de pb6-Ig, 65% de la protéine entière est attribuée. Ce pourcentage d'attribution sera largement augmenté à l'aide des données de pb6-loop2 et d'une analyse plus poussée.

Une fois les attributions de pb6 monomère et de la fibre complétées, une comparaison des variations de déplacement chimiques et de la dynamique (repliement et flexibilité des différents domaines) sera réalisée.

6.2.3 Combinaison d'EM et de RMN

La combinaison des contraintes de RMN et d'une enveloppe de microscopie électronique peuvent être cumulés dans une simulation de dynamique moléculaire pour obtenir des structures à résolution atomique de grands assemblages moléculaire [Loquet 12, Shi 15, Habenstein 15, Cuniassse 17]. La RMN amène des contraintes sur les orientations de chaque acide aminé par le biais des déplacements chimiques ainsi que sur les distances inter-atomiques obtenues par spectroscopie NOE (Nuclear Overhauser Effect) qui peuvent être utilisées pour construire un modèle initial. Une densité électronique obtenue par cryoEM de 4 à 15 Å, peut être utilisée pour apporter une contrainte spatiale. Une limitation peut être le placement du modèle dans la densité électronique qui peut forcer à tester plusieurs possibilités. La solution devrait émerger comme celle convergent vers la structure la plus stable (respect des géométries de polypeptides et faible énergie libre) et respectant le plus la densité électronique. Une fois ce modèle identifié, le cumule des contraintes RMN et EM peut être réalisé pour générer une structure à une résolution atomique.

Cette méthode pourra être appliquée au tube de pb6 en utilisant les données de ssRMN des fibres de pb6 et la reconstruction du tube présentée dans chapitre 5 nous permettant de résoudre la structure de l'assemblage. Dans le cas de pb6, l'existence d'un très bon ajustement rigide de la structure cristallographique dans la densité électronique et d'un modèle réalisé par ajustement flexible devrait faciliter la mise en place de cette stratégie. Cette structure permettra une description plus précise des interactions en jeu dans le tube et d'approfondir notre comparaison avec les tubes de T4 et de la pyocine R (5.1).



Points clés

Résumé chapitre 6

- ❑ Des échantillons de pb6 sous sa forme monomérique soluble et polymérisé insoluble ont été produits avec différents marquage. Plusieurs mutants de pb6 ont également été produits avec succès.
- ❑ L'attribution de la protéine pb6 (50 kDa) est en cours. L'attribution du domaine Ig-like de pb6 a été réalisé en RMN du solide en utilisant une stratégie innovante ayant fait l'objet d'une publication (cf. Annexe)
- ❑ La comparaison des spectres de la protéine monomérique donnera des information sur la dynamique de la protéine, les résidues impliqués dans la formation du tube et peut être son mécanisme d'assemblage.
- ❑ La combinaison des données de Microscopie électronique et de RMN pourra permettre la production d'une structure haute résolution du tube de pb6.

Complexe distal de la queue de T5 et interaction avec la membrane

« (...) souvenez-vous que dans
les champs de l'observation, le
hasard ne favorise que les
esprits préparés. »"

(Louis Pasteur)

Chapitre 7

Complexe distal de la queue du phage T5 : contexte & méthodes

7.1 Contexte

Cette partie s'intéresse au complexe distal de la queue et les changements de structure qui s'y opèrent lors de l'interaction avec FhuA. Du fait de contraintes de temps, une grande partie des résultats présentés sont préliminaires, et l'objectif d'observer des structures transmembranaires n'a, pour l'instant, pas été atteint. Des éléments de réflexions inattendus et intéressants, notamment sur le rôle de la fibre droite (pb4), sont néanmoins ressortis de l'analyse de ces données préliminaires. De plus certains résultats ont un intérêt pour les conclusions apportés par l'article présenté en chapitre 5.

7.2 Matériel et méthodes

7.2.1 Amplification de T5sto et T5amD20

Une culture de *E. coli* F (souche LPS rough issue de la souche FCb [LANNI 58]) en LB, 1 mM MgCl₂, 1 mM CaCl₂, est infectée avec le phage T5sto lorsque la DO₆₀₀ de la culture atteint 0,3 (environ 3.10⁸ bactéries par mL) avec une MOI (Multiplicity of Infection) de 10⁻². La DO₆₀₀ est suivie jusqu'à ce qu'elle chute et se stabilise (en général aux alentours de DO₆₀₀ = 0,1-0,2). Pour le mutant T5amD20 (mutant ambre dans le gène D20, codant pour la protéine majeure de capsid), le même protocole est appliqué, mais en utilisant la souche *E. coli* CR63 (souche codant pour l'ARNt SupD qui corrige la mutation ambre).

7.2.2 Purification et caractérisation de T5sto et T5amD20

La lyse des cellules est complétée par l'ajout de 1 mL de chloroforme, de 10 µg/mL de DNase et de 1 µg/mL de RNase (Roche) et incubation 1h à 37°C. 0.5 M de NaCl sont ajoutés et la solution incubée 30' à température ambiante. Le milieu est ensuite centrifugé 20' à 6000 rpm (JLA8.1), 8-10% de PEG 6000 sont ajoutés au surnageant qui est incubé à 4°C sur une nuit.

La solution est centrifugée 1h30' à 6000 rpm (JLA8.1), le culot est resuspendu dans du tampon (10 mM Tris-HCl pH 7,5, 150 mM NaCl, 1 mM MgCl₂, 1 mM CaCl₂) dans environ 1/100e du volume de culture initial. La solution est centrifugée 30' à 4000 rpm (SX4250) 0,5 M NaCl et 8% PEG 6000 sont ajoutés au surnageant et la solution est à nouveau incubée sur la nuit à 4°C. La solution est centrifugée 30' à 4000 rpm (SX4250) et le culot replit dans un volume équivalent de tampon. La solution est déposée sur un gradient de CsCl et centrifugée 2h à 20000 rpm (SW41), les fractions contenant les bactériophages sont poolées et dialysées.

Le titre de la solution de phage purifié est évalué sur boîte de pétri par décompte de plages de lyse. Pour le phage T5amD20, la titration comparative sur les souches *E. coli* F et *E. coli* CR63 permet de confirmer qu'il s'agit du virus mutant et d'estimer le nombre de révertants.

7.2.3 Production et purification de queues de phage T5

Les queues de phages sont obtenues par infection de *E. coli* F par T5amD20. Le protocole est similaire à celui de production/purification des phages complets jusqu'à la première précipitation PEG/NaCl à l'exception de la MOI utilisée pour l'infection (6-7). Le culot replit dans du tampon est centrifugé 20' à 4000 rpm (rotor SX4250) et le surnageant déposé sur un gradient glycérol qui est centrifugé 2h à 20000 rpm (SW41). Les queues de phages sont récupérées dans les fractions hautes du gradient, poolées et diluées dans du 10 mM Tris-HCl, pH 7,5 de façon à ce que la conductivité de la solution soit <5 mS/cm. Les queues de phages sont alors purifiées sur une colonne d'échange d'anions.

7.2.4 Production de FhuA

FhuA est surexprimée dans une culture d'*E. coli* AW740 {ompF zcb : Tn10 ompC FhuA31} transformée avec le plasmide pHX405. La souche AW740 a pour particularité de ne pas exprimer les 2 porines majeures de la membrane externe d'*E. coli* (ompF et ompC), de présenter une mutation sur FhuA endogène et elle résiste à la tétracycline (Tet) [Ingham 90].

Le plasmide pHX405 permet l'expression de FhuA portant un tag 6 histidines sur la boucle 5 (après le résidu 405), sous le contrôle de son promoteur naturel,

et porte un gène de résistance à l'ampicilline (Amp). L'expression de FhuA est induite par une carence en fer.

Une préculture de LB, 125 µg/ml ampicilline, 10 µg/ml tetracycline est commencé à partir d'un stock glycérol, cette préculture est incubée la nuit à 37°C sous agitation. Des cultures de LB, 125 µg/ml ampicilline, 10 µg/ml tetracycline, 100 µM dipyridine sont ensemencées à 0,05 de DO₆₀₀ initial. Les cultures sont incubées à 37°C sous agitation pendant 8 à 12h. Les cultures sont alors centrifugées 15' à 6000 rpm rotor JA10 à 4°C.

7.2.5 Purification de FhuA

Le culot est repris dans un tampon de lyse (50 mM Tris pH 8, 150 mM NaCl, 2 mM MgCl₂, 1 pointe de spatule DNase, CLAPA 1X) et les cellules sont cassées par 5 passages au microfluidizer à 10000 PSI. Le lysat de bactérie est alors centrifugé 15' à 6000 rpm (JA25-50). Le surnageant est centrifugé 20' à 35000 rpm (Ti45). Le culot est repris dans une solution 50 mM Tris pH 8, 2% (w/w) OPOE (octyl polyoxyéthylène), puis incubé 30' sous agitation à 37°C avant d'être à nouveau centrifugé 20' à 35000 rpm (Ti45). Le culot est repris dans une solution 50 mM Tris pH 8, 1 mM EDTA, 1% (w/w) LDAO (Lauryldiméthylamine oxide) et est incubé 1h sous agitation à 37°C avant d'être centrifugé 20' à 35000 rpm (Ti45). 4 mM de MgCl₂ sont ajoutés au surnageant avant purification sur colonne HisTrap HP 5ml (GE Healthcare). La colonne est préalablement équilibrée avec un tampon d'équilibration (0,1 % LDAO, 20 mM Tris pH 8, 5 mM Imidazole). Le surnageant est déposé sur la colonne qui est ensuite lavée avec du tampon d'équilibration, puis avec un tampon de délipidation (1 % LDAO, 20 mM Tris pH 8, 5 mM Imidazole), la protéine est éluée avec un tampon d'éluion (0,1 % LDAO, 20 mM Tris pH 8, 200 mM imidazole). Tout le protocole se réalise à un débit de 2 ml/min, à l'exception de l'étape de délipidation (0,25 ou 0,1 mL/min). Les fractions d'éluions de la chromatographie précédente sont déposées sur une colonne HiTrap Q HP 1ml (GE Healthcare) préalablement équilibrée avec un tampon (0,05 % Tris LDAO, 20 mM Tris pH 8). La protéine est éluée dans le même tampon par un gradient linéaire de concentration de NaCl allant de 0 à 1 M.

7.2.6 Purification de MSP1E3D1 (similaire pour MSP1D1)

Cultures et protocoles fournies par Yann Huon De Kermadec.

Les culots sont repris dans un tampon de lyse (20 mM NaPi, pH 7.4, CLAPA 1X, 1 % (w/v) Triton X-100, lysozyme 0,2 mg/ml, DNaseI 0,2 mg/ml), puis les cellules sont lysées mécaniquement grâce à un microfluidizer (3 passage à 13000 psi, microfluidic M-110P) ou par sonication (2 secondes à 90% de puissance, 10 seconde de repos, 12 minutes au total, branson digital sanifier). La solution est

ensuite centrifugée pour se débarrasser des débris cellulaires à 50000 g pendant une heure à 4°C. Le surnageant est passé sur une résine de nickel (His-trap, 5ml Quiagen), préalablement équilibrée avec le tampon (40 mM Tris-HCl, pH 8.0, 300 mM NaCl, 1% (w/v) Triton X-100). La résine est lavée successivement par le tampon précédent, puis par le tampon (40 mM Tris-HCl, pH 8.0, 300 mM NaCl, 50 mM NaCholate), et (40 mM Tris-HCl, pH 8.0, 300 mM NaCl) et enfin (40 mM Tris-HCl, pH 8, 300 mM NaCl, 10 mM Imidazole) à raison de 25ml de chaque tampon et la protéine est éluée avec le tampon (40 mM Tris-HCl, pH 8.0, 300 mM NaCl, 300 mM Imidazole)

Lorsqu'il est nécessaire de cliver l'étiquette Histidine de la protéine MSP, la protéase TEV est ajoutée à la fraction éluée de la colonne d'affinité, dans un ratio massique de 1MSP :10 TEV. Le mélange est incubé toute la nuit dans le tampon (50 mM Tris-HCl, pH 8, 0.5 mM EDTA, 1 mM DTT), à température ambiante puis dialysée pendant 2h dans 4 L de tampon (20 mM Tris-HCl, pH 8, 150 mM NaCl). La protéine clivée est de nouveau chargée sur la résine NiNTA, et la résine est lavée avec 20ml de tampon (20 mM Tris-HCl, pH 8, 150 mM NaCl). Les fractions non retenues et celles issues des lavages, qui contiennent la protéine clivée, sont dialysées dans du tampon (20 mM Tris-HCl, pH 7.4, 100 mM NaCl, 0.5 mM EDTA), pendant toute la nuit à température ambiante. La protéine est concentrée jusqu'à environ 1,5 à 2mg/ml, le dosage de la protéine étant fait en mesurant l'absorbance à 280nm. (Mw 23329,3 ; 23950 ; NanoDrop 2000c Thermo scientific)

7.2.7 Purification et caractérisation de FhuA en nanodisques

FhuA (solubilisé en LDAO), MSP1E3D1 et les lipides (Avanti Polar lipids) sont mélangés avec un ratio 1/6/360 et incubés 1h sur glace. Des BioBeads™ (Sigma) sont ajoutés à la solution (0,5 g par mL de solution environ) et elle est incubée 2 h à température ambiante sous agitation. La qualité des échantillon peut être vérifiée sur chromatographie d'exclusion de taille (cf. 3.2.2.7).

Les nanodisques sont purifiés par chromatographie d'affinité au nickel pour diminuer la proportion de nanodisques vides (tampon d'équilibration : 20 mM Tris-HCl pH 7,5, 150 mM NaCl, 5 mM imidazole ; tampon d'éluion : 20 mM Tris-HCl pH 7,5, 150 mM NaCl, 200 mM imidazole). Enfin, l'imidazole est éliminé par dessalage (cf. 3.2.2.4) en utilisant le tampon (10 mM Tris-HCl pH 7,5, 150 mM NaCl, 1 mM MgCl₂, 1 mM CaCl₂).

7.2.8 Microscopie électronique

7.2.8.1 Préparation de grilles pour la coloration négative

cf . 3.2.3.3

7.2.8.2 Préparation de grilles pour la cryo-microscopie

cf. 3.2.8.1

7.2.8.3 Traitement des données

Un modèle initial a été obtenue par Guy Schoehn en utilisant un protocole similaire à ce qui a été fait pour le tube de pb6 (cf. 3.2.8.2).

Sur l'échantillon de queues de T5 avant interaction avec FhuA, 292 micrographes ont été collectés à un grossissement de 15500 (1,24 Å/pixel). 4857 particules ont été pickées en utilisant le programme boxer [Ludtke 99] et l'extraction a été réalisée avec Relion [Scheres 12] dans des boites de tailles différentes. 4533 particules ont été utilisées dans les reconstructions 3D (quelques particules éliminées par classification 2D). Des stratégies de classification 2D et 3D ont été testées mais le nombre restreint de particules ne permettait pas de mesurer d'effet sur la reconstruction. Les reconstructions 3D ont été obtenues par simple Refine3D, avec ou sans imposition de symétrie C3. Une reconstruction sans imposition de symétries à partir de particules extraites dans des boites de 200x200 pixels. Une reconstruction à partir de boite de 500x500 en imposant la symétrie C3 abouti à une reconstruction à 12,5 Å et celle à partir de boites de 1000x1000 une reconstruction à environ 20 Å.

Le même modèle initial a été utilisé pour l'échantillon de queues de phages en interaction avec FhuA. 513 micrographes avec un grossissement de 23000 (0,82 Å/pixel) ont été collectés. Un binning de 2 a été appliqué aux images (regroupement de pixels) pour accélérer le traitement des données. 6672 particules ont été extraites dans des boites de 250x250 pixels. 6640 particules ont été utilisées dans la reconstruction et aucune symétrie n'a été imposée résultant en une structure de 17 Å de résolution environ.

Chapitre 8

Complexe distal de queue du phage T5 : résultats & perspectives

8.1 Préparation des échantillons

8.1.1 Queues de T5

Les queues de phages sont obtenues grâce à un phage mutant, T5 $amD20$ qui possède une mutation ambre (codon stop) dans le gène de la protéine majeure de capsid (pb 11). Les voies d'assemblages de la queue et de la capsid étant indépendantes, l'infection avec ce virus conduit à l'accumulation de queues de phages.

8.1.1.1 T5 $amD20$ -Amplification et contrôle de la souche

L'amplification de T5 $amD20$ est réalisée par infection de *E. coli* CR63, une souche suppressive qui permet l'incorporation d'un acide aminé pour le codon stop ambre. L'infection par T5 $amD20$ produit alors des virions entiers.

Il est important de noter l'importance de réaliser l'amplification à partir d'un clone isolé (plage de lyse) sous peine de voir le nombre de révertants (mutation du codon stop de pb11) augmenter rapidement. Le taux de révertant est évalué par comparaison du nombre de plages de lyse sur tapis de *E. coli* CR63 et *E. coli* F. Le taux normal de révertant se situe aux alentours de 0,0001 %, cependant, si l'infection est réalisée à partir d'un stock plutôt que d'une plage de lyse, ce taux monte à 1 % et à plus de 10 % pour une seconde amplification.

8.1.1.2 Production de queues de phages

Les queues de phages sont obtenues par infection de culture de *E. coli* F par T5amD20 à haute MOI (Multiplicité d'infection, le ratio entre le nombre de bactéries et de virus ; une MOI haute signifie 8-10 fois plus de virus que de bactéries). Environ 1h après infection, la DO₆₀₀ atteint un pic puis chute traduisant la lyse des bactéries. Avant le pic de DO₆₀₀ on observe un petit plateau dans la phase exponentielle, ce comportement a été observé de façon reproductible sur différentes cultures (données non présentées).

8.1.1.3 Purification de queues de phages

Après une étape de précipitation PEG/NaCl du surnageant de culture, les queues de phages sont purifiées sur gradient de glycérol. La plupart des contaminants sont éliminés dans la partie haute du gradient ou dans le culot. Deux contaminants, à environ 55 kDa et 35 kDa sont retrouvés dans les fractions 30 et 40% glycérol. Les queues se trouvent principalement dans la fraction 10% de glycérol, ainsi que dans la fraction 20% (Fig. 8.1-a)

Lors de la colonne échangeuse d'anions, les queues éluent sur une gamme de conductivité assez large (21 à 45 mS/cm). On peut distinguer 5 pics d'élution (Fig. 8.1-b), dont le premier correspond à des queues pures (23,5 mS/cm), les autres correspondent à des queues co-éluant avec différents contaminants. Comme la densité des bandes des protéines de queues de phages (en particulier celle de pb6) ne diminue pas linéairement, il ne s'agit donc pas d'une « trainée » de l'élution de l'échantillon mais cela suggère plutôt des interactions entre les queues de phages et les contaminants. Des efforts ont été fait pour investiguer la possibilité d'une corrélation entre la conductivité d'élution et la désolidarisation des protéines pb1 (fibres longues) et/ou pb2 (protéine vernier), qui a été observée en microscopie, mais sans succès (cf. Annexe).

Des variations notables ont été observées d'une purification à une autre, la fenêtre de conductivité sur laquelle les queues éluent a été observée entre 5 et 55 mS/cm avec une amplitude allant de 10 à 30 mS/cm. Les fractions contenant les queues le plus pures se trouvent systématiquement en début d'élution, souvent entre 5 et 10 mS/cm ou entre 20 et 24 mS/cm. Lorsque la conductivité de la solution chargée sur la colonne est proche ou supérieure à 5 mS/cm, des queues sont trouvées dans la fraction non retenue. Ces comportements semblent affectés par le volume de la colonne et la pente du gradient salin et illustre la capacité des particules virales à entretenir des interactions faibles et peu spécifiques avec les éléments de leur environnement.

Le rendement moyen de la purification de 0,5 L de culture est de 1 mL de queues pures à 100-150 nM soit 10 à 15 nmol d'échantillon. Le complexe de queue

est composé de 11 protéines, toutes visibles sur gel (voir Fig. 8.2).

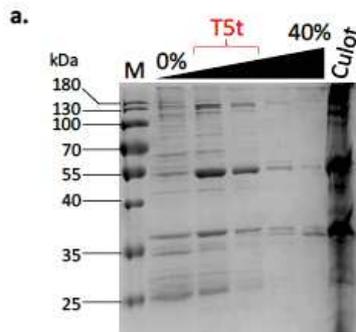
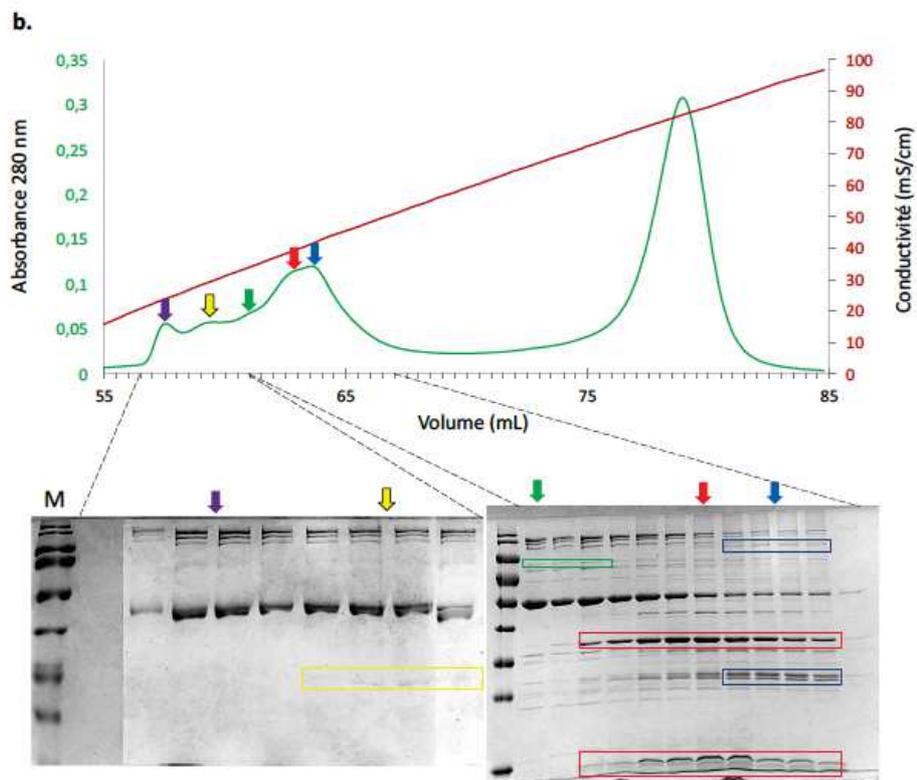


Figure 8.1 : Purification de queues de phage T5. **a.** Gel de la première étape de la purification des queues, un gradient de glycérol. **b.** Chromatogramme de l'éluion de queues de phages. La courbe rouge indique le gradient salin, la verte l'absorbance à 280 nm. Les flèches de différentes couleurs indiquent des pics d'éluion de queues de phages avec différents contaminants.



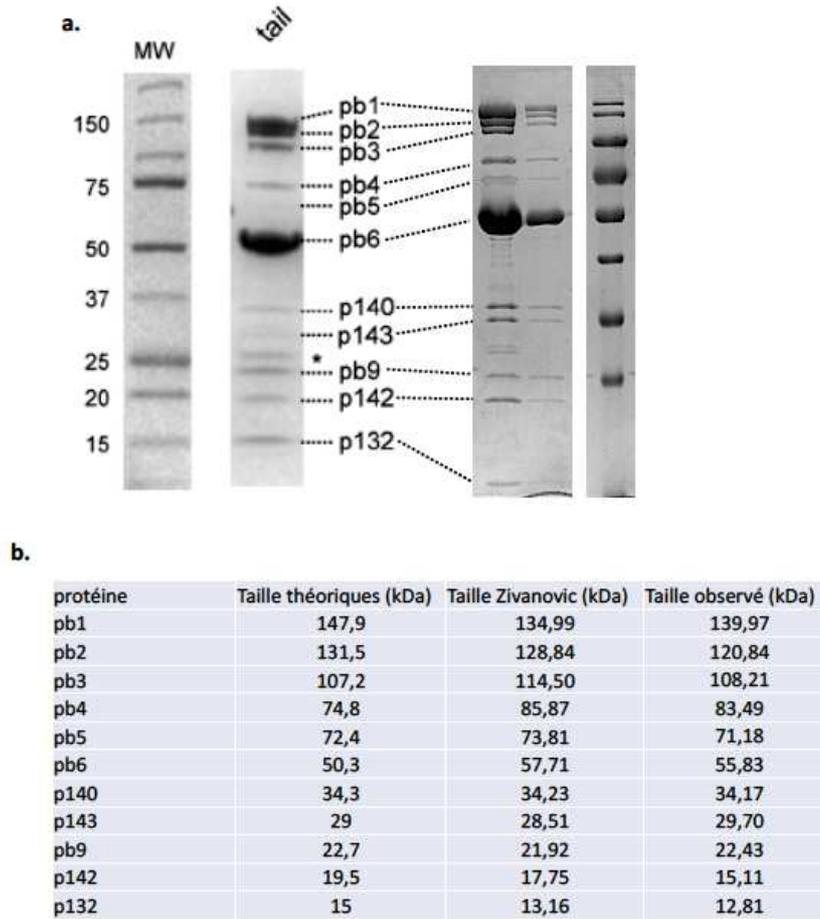


Figure 8.2 : SDS-PAGE de queues de phages purifiées. **a.** Comparaison avec des queues de phages purifiées dans la littérature [Zivanovic, 14]. **b.** Tableau avec les tailles théoriques de toutes les protéines du complexe de queue (calculé à partir de la séquence) et les tailles calculées à partir de la distance de migration sur gel.

8.1.2 FhuA

La surexpression et la purification de la protéine FhuA ne seront pas détaillées dans les résultats car elles ont été décrites dans la littérature il y a près de 20 ans [Ferguson 98].

8.1.2.1 Nanodisques de FhuA

La production et la purification de MSP ainsi que les premiers essais de solubilisation de FhuA en nanodisques ont été réalisés avec Yann Huon-De-Kermadec.

Les nanodisques sont des disques de bicouches lipidiques solubilisés grâce à la protéine MSP (Membrane Scaffold Protein). La protéine MSP est purifiée sur colonne d'affinité au nickel puis son étiquette histidine clivée (Fig. 8.3-a). Les nanodisques contenant FhuA sont formés par le mélange de lipides/FhuA/MSP à un ratio 360/1/6 et où le détergent est éliminé avec des BioBeadsTM après 1h d'incubation. La solution de FhuA solubilisée en nanodisques est à nouveau chargée sur une colonne d'affinité au nickel pour permettre de purifier les nanodisques chargés en FhuA, et se débarrasser des nanodisques vides (Fig. 8.3-b).

Des essais ont été réalisés avec différents lipides : le POPC et un extrait de lipides polaire de *E. coli*. Les nanodisques réalisés avec le POPC ont un profil sur filtration sur gel avec moins de matériel niveau du volume mort et une proportion de nanodisques pleins qui semble meilleure (Fig. 8.3-c).

Il y a deux types de MSP utilisées en routine dans le laboratoire : MSP1D1 et MSP1E3D1 qui forment des disques de 100 Å et 140 Å de diamètre respectivement. D'après ces valeurs, on peut estimer la surface de MSP1S1 à 7854 Å² et celle de MSP1E3D1 à 15394 Å². Mais des données expérimentales [Bayburt 10, Denisov 04] estiment ces valeurs plus basses : 4400 Å² pour MSP1D1 et 8900 Å² pour MSP1E3D1. La surface de FhuA est estimée à 1570 Å². Il faut prendre en compte le fait que le phage forme probablement une structure transmembranaire supplémentaire ne passant pas par le tonneau de FhuA (l'intérieur du tonneau de FhuA est occupé par un domaine qui devrait se désolidariser pour permettre le passage de protéine et d'ADN ; un tel phénomène n'est pas observé sur gel). Cette structure supplémentaire occuperait une surface minimum de 1256 Å², ne laissant que 1574 Å² de surface lipidique, soit 22 molécules de POPC (71 Å² par molécule) si l'on utilise MSP1D1. MSP1E3D1 est donc utilisé car il permet d'avoir une marge de 6053 Å² (soit 85 molécules de POPC).

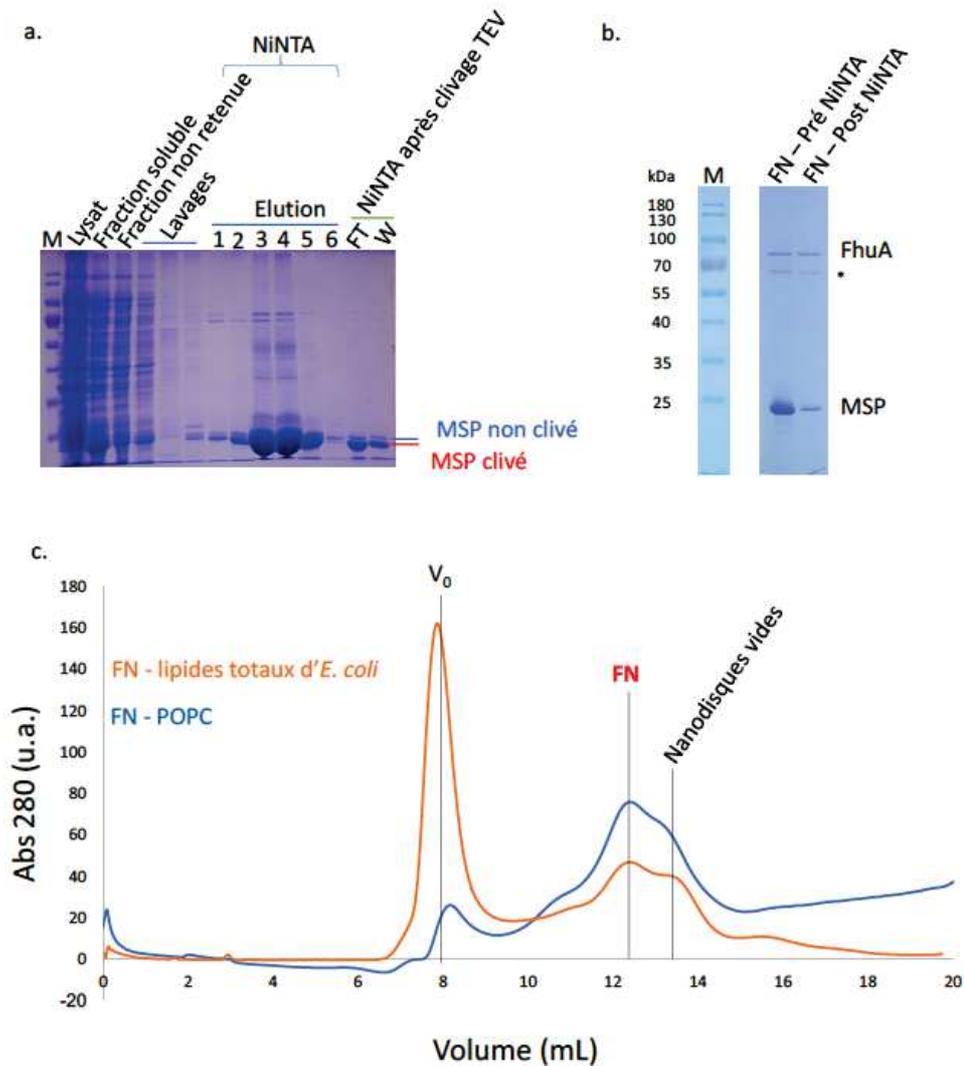


Figure 8.3 : Purification de la protéine MSP et assemblage de nanodisques avec FhuA.
a. SDS PAGE de la purification de MSP. **b.** SDS PAGE comparant FhuA solubilisé en nanodisque avant et après passage sur colonne d'affinité au nickel. La quantité de nanodisques vide est significativement diminuée. **c.** Profils de chromatographie d'exclusion de taille de FhuA solubilisé en nanodisques en utilisant 2 extraits lipidiques différents. FN : FhuA-Nanodisque.

8.2 Structure du complexe distal de queue

8.2.1 Structure du complexe distal de queue avant interaction avec FhuA

8.2.1.1 Coloration négative

Avant que des grilles de microscopie électronique soit congelées, les échantillons sont systématiquement observés en coloration négative pour confirmer leur qualité (pureté, stabilité du complexe et concentration).

Sur ces micrographes (Fig. 8.4), il est possible de distinguer les différents éléments du complexe distal. Le collier, à la base du tube de pb6, sert d'attache aux fibres flexibles. Les trois fibres flexibles ont 2 coudes (zones donnant des degrés de libertés à la structure autrement rigide), un au niveau de la jonction avec le collier et un autre environ 20 nm plus loin. Ces fibres flexibles seront peu discutées dans le présent document mais leur structure a été déterminée dans [Garcia-Doval 15]. Sous le collier se trouve une structure rectangulaire, poursuivi par un rétrécissement en cône qui se termine par la fibre droite au bout de laquelle se trouve pb5 (Receptor Binding Protein, RBP). La fibre droite a un coude à environ 450 Å du collier qui permet à l'extrémité portant la RBP d'être plus mobile, facilitant ses mouvements pour interagir avec la protéine FhuA. Ces observations sont en parfait accord avec ce qui avait été décrit précédemment ([Zivanovic 14, Effantin 06, Böhm 01]; cf. 1.5).

L'échantillon de queues est relativement stable à 4°C, mais des altérations ont été observées après un certain temps. Des micrographes en coloration négative de l'échantillon 1 mois après purification (Fig. 8.5) montre en effet qu'une importante proportion des queues perdent la protéine pb2 (TMP) qui se trouve à l'intérieur du complexe. La sortie de pb2 se fait sans changements de conformations du complexe distal. Elle se fait probablement par l'extrémité proximale, normalement connecté à la capsid, cette assomption est confirmée par l'observation de particules où la TMP est partiellement sortie, laissant la partie distale du tube vide alors que la partie proximale est encore remplie de densités. Cela confirme l'hypothèse présumée dans Arnaud, 2017 que la TMP est contrainte dans un état métastable dans le tube.

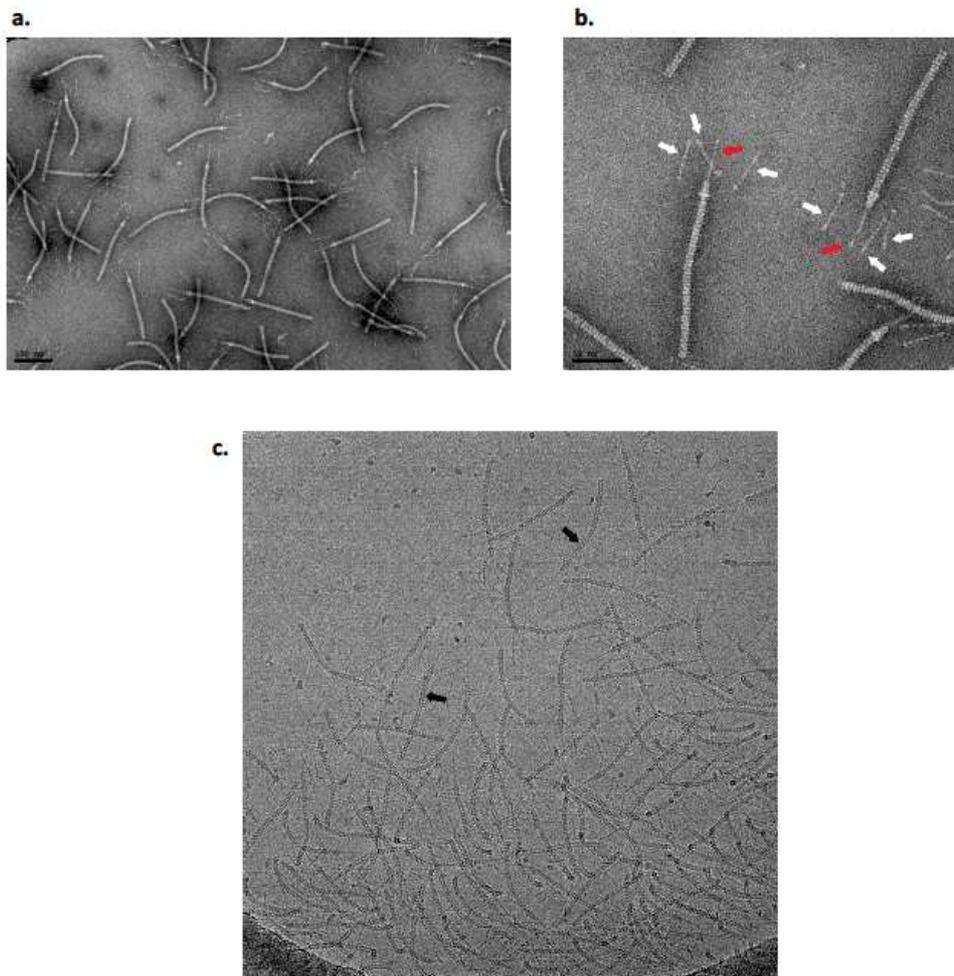


Figure 8.4 : Micrographes de queues de phages. a. et b. Microscopie en coloration négative de queues de T5 à deux grossissements différents . Les trois fibres flexibles liées au collier (flèches blanches) et la fibre droite (flèches rouges) sont clairement visibles. **c.** Micrographe de cryo-microscopie électronique typique de l'échantillon queues de T5. Une proportion importante de l'échantillon est concentrée sur le bord du trou ainsi que sur le carbone. Seules les particules clairement isolées peuvent être utilisée pour la reconstruction 3D. Des queues de phages ayant perdues pb2 (flèches noir) sont visibles.

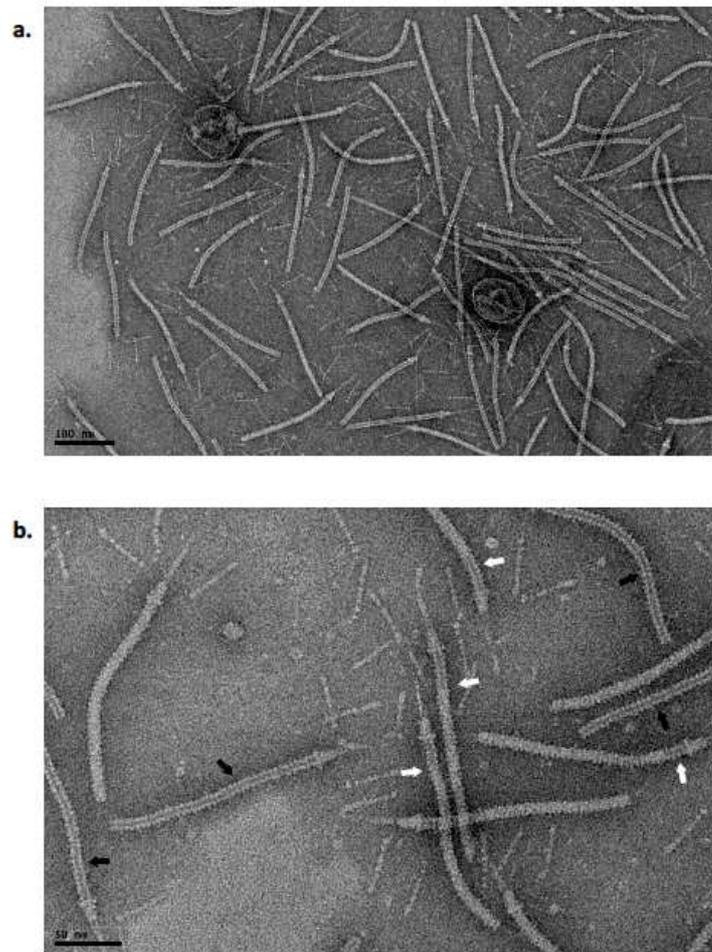


Figure 8.5 : Micrographes de queues de phages un mois après purification. a. et b. Microscopie en coloration négative de queues à deux grossissements différents. **a.** 60% des queues de phages ont perdu la protéine pb2 (TMP), l'intérieur du tube se trouve alors accessible au colorant et devient plus sombre. **b.** Les queues de phages ont complètement (flèches noires) ou partiellement (flèches blanches) perdu pb2. La sortie de la protéine semble se faire par l'extrémité proximale (flèche blanches), et le complexe distale de queue reste intact.

8.2.1.2 Cryo-microscopie électronique

Deux reconstructions 3D du complexe distale ont été obtenues à partir de 4533 particules, dans les 2 cas une symétrie C3 a été imposée (Fig. 8.6). Les 2 reconstructions diffèrent par la taille de boîte utilisée : 1000 et 500 pixels (la boîte est centrée sur la base collier). Une reconstruction sans imposition de symétrie est également présentée (Fig. 8.7) à partir d'une taille de boîte de 200 pixels. Les reconstructions sans imposition de symétrie avec des boîtes de 500 et 1000 pixels ont également été réalisées mais ne sont pas présentées car un alignement correct de ce petit nombre de particules n'a pas été possible et les densités résultantes valident la forme générale de l'objet mais sont assez peu informatives.

Pour des raisons de convenance pour la description de l'assemblage, les termes « supérieur » ou « dessus » font référence à la partie proximale de la queue (côté capsid), « inférieur » ou « dessous » font référence à la partie distale (côté membrane) et « extérieur » et « intérieur » sont utilisés pour positionnés par rapport à l'axe du tube.

Sur la reconstruction obtenue avec la taille de boîte la plus grande (1000 pixels soit ~ 1250 Å ; environ 20 Å de résolution), la densité de la fibre droite est visible jusqu'à 470 Å en dessous du collier. Neufs anneaux de pb6 sont résolus du côté proximal, cependant, comme révélé par l'orientation des domaines Ig-like la reconstruction a été faite avec la mauvaise main (Fig. 8.8-b), c'est aussi le cas des autres reconstructions. Les domaines Ig-like du dernier anneau de pb6 semblent interagir avec la partie supérieure du collier. Les moignons des trois fibres flexibles sont visibles au départ de la partie inférieure de l'anneau. L'observation de la face inférieure du collier (Fig. 8.8-c.) révèle une symétrie 12, indiquant qu'il pourrait être formé d'un anneau dodécamérique. L'observation de la reconstruction asymétrique ne permet cependant pas de valider avec certitude cette assertion, il pourrait s'agir d'un décamère ou d'un dodécamère.

A l'intérieur du collier, une densité a pu être isolée, correspondant probablement à la protéine p140, seule protéine de queue à ne pas avoir été localisée [Zivanovic 14]. La hauteur de cette densité (~ 30 Å) et sa position dans l'assemblage permettent de supposer que la protéine p140 a un repliement analogue aux autres protéines de tubes, formant un anneau hexamérique, et occupe une position dans l'assemblage similaire à gp54 de T4.

Sous le collier, 6 protubérances correspondant au domaine de décoration de pb9 (domaine OB-fold) sont visibles confirmant la position et l'organisation en anneau hexamérique proposée dans [Flayhan 14]. Cette structure est également analogue au repliement commun des protéines de tubes, à l'instar de pb6, des TTPs des *Siphoviridae* et des *Myoviridae* et des protéines de tubes de SST6 et de pycine R.

La densité faisant suite à pb9, qui avait été affiliée à la protéine pb3 [Zivanovic 14], a une hauteur de 125 Å et une largeur d'environ 100 Å. Cette structure semble

avoir une symétrie 3, avec 3 domaines partant vers l'extérieur dans sa partie supérieure et 3 protubérances pointant vers la partie distale sur sa face inférieure (Fig. 8.8-e). La symétrie 3 de cet élément est confirmée par une reconstruction où la symétrie n'est pas imposée (Fig. 8.7).

La protéine pb4 est supposée former la fibre droite jusqu'à la RBP. Dans sa partie supérieure (correspondant au cône), 3 domaines allongés vers la partie distale forment une légère hélice, la partie distale de la fibre droite (visible en partie sur la Fig. 8.6-a), est constitué d'une suite de domaines globulaires et discoïdaux. La symétrie 3 de pb4 n'est pas confirmée par la reconstruction asymétrique sur laquelle cet élément n'est pas visible, mais est validé par les travaux présentés dans le paragraphe 8.2.2.2.

En segmentant la densité du complexe sur la reconstruction avec un boxing de 500 pixels (13 Å de résolution environ), il a été possible d'identifier un domaine continu à travers le tube de pb6, les anneaux de p140 et de pb9 ainsi que pb3 (Fig. 8.6-b et Fig. 8.8-f). Cette densité correspond très probablement à la protéine pb2 (TMP). A cette résolution, la structure et la stoechiométrie de la protéine ne peuvent pas être discutés. De plus, il est difficile d'évaluer la fiabilité de cette segmentation. De manière intéressante, un autre domaine a pu être isolé à l'intérieur de pb3, semblant faire le lien entre pb2 et pb4 (Fig. 8.8-f). Il est difficile d'attribuer cette densité à pb2 ou pb4, il pourrait également s'agir d'un domaine de pb3. La reconstruction asymétrique ne permet pas non plus de discriminer les différents éléments présents à l'intérieur de pb3 : il semble raisonnable d'attribuer la densité cylindrique visible dans la partie proximale à pb2, mais les densités dans la partie distales sont plus floues et des liaisons semblent se faire avec la paroi formé par pb3 (Fig. 8.7-b). Il est également possible que pb2 et/ou pb4 aient des domaines qui participent à la formation du domaine qui a été attribué à pb3, cette hypothèse étant néanmoins peu probable pour pb2 qui se désolidarise facilement du reste de l'assemblage (Fig. 8.5).

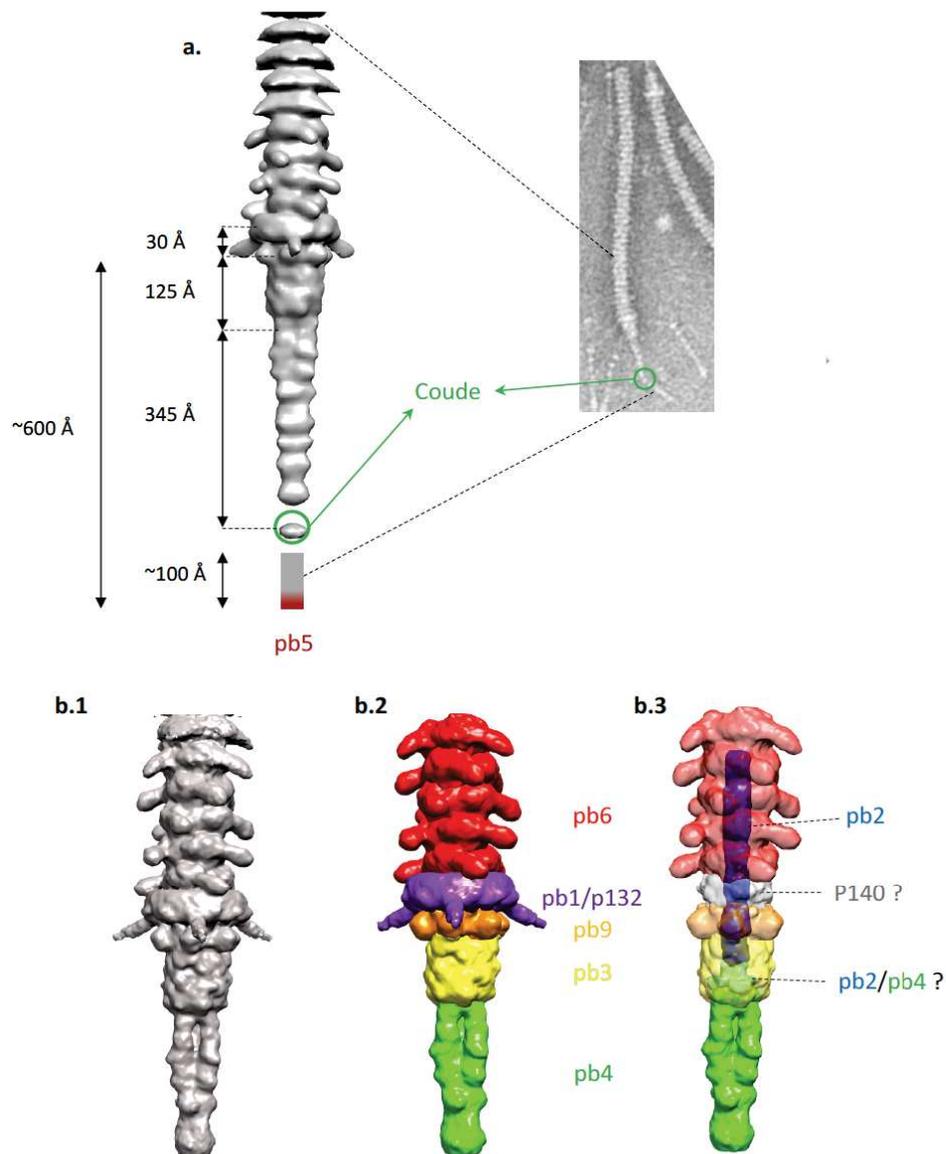


Figure 8.6 : Reconstruction 3D du complexe distal de queue. **a.** Reconstruction 3D du complexe de pointe avec une taille de *boxing* de 1000 pixels (~1250 Å). La résolution est d'environ 20 Å. **b.** Reconstruction 3D du complexe de pointe avec une taille de *boxing* de 500 Å, la résolution est d'environ 13 Å. **b.2. & b.3** La densité du tube a été segmentée et colorée par protéine et domaines dissociables. La densité du collier a été retirée et les autres segments mis en transparence sur le panneau **b.3.** La densité d'un anneau assurant la continuité entre **pb6** et **pb9** a été identifié (densité grise **b.3.**), il s'agit probablement de p140. Une densité a également été identifié à l'intérieur de **pb3** semblant assurée une continuité entre **pb2** et **pb4**.

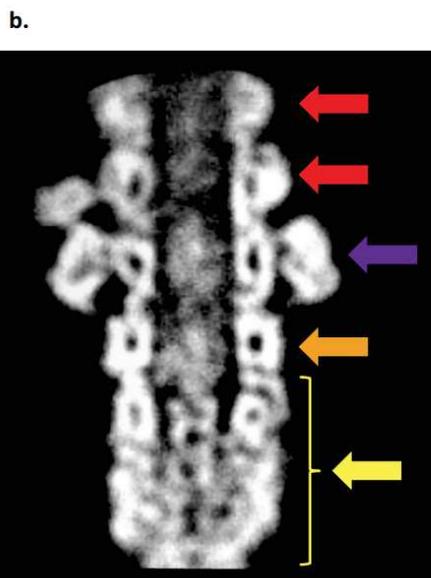
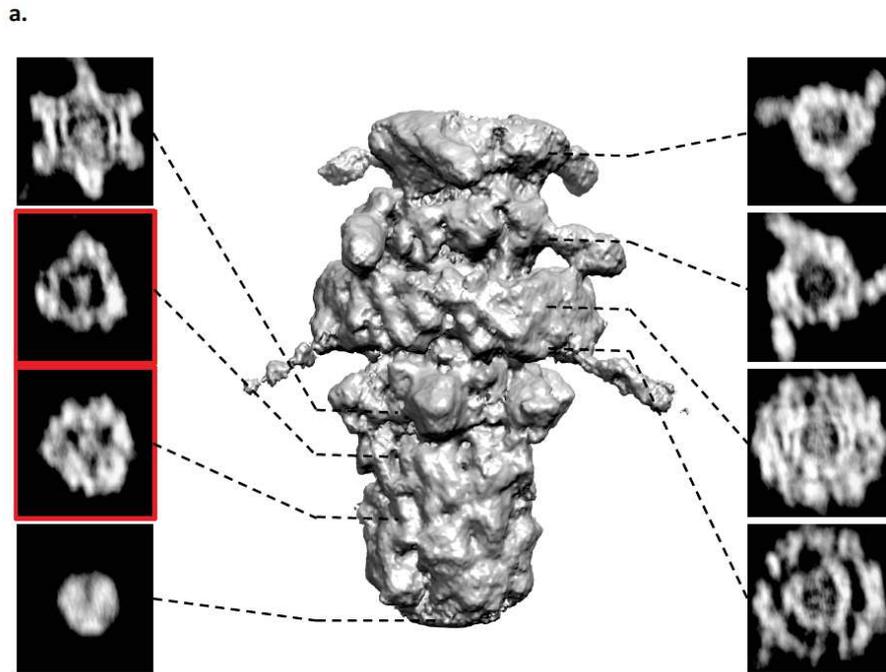


Figure 8.7 : reconstruction asymétrique du complexe distal de la queue du phage T5. reconstruction 3D réalisée à partir d'une taille de boîte de 200 pixels, sans imposer une symétrie C3. **a.** Vue de côté de la structure du complexe distal avec des coupes transverses à différents niveaux de la structure. La symétrie 3 de pb3 est manifeste sur les coupes 3 (encadrés rouge). **b.** Coupe médiane de la structure. Les flèches respectent le code couleur précédemment utilisé pour pb6, le collier, pb9 et pb3. Les sandwichs β des anneaux de pb6 et pb9 sont clairement visibles. Des structures similaires sont visibles au niveau de la partie supérieure de pb3 et de l'intérieur du collier (probablement p140).

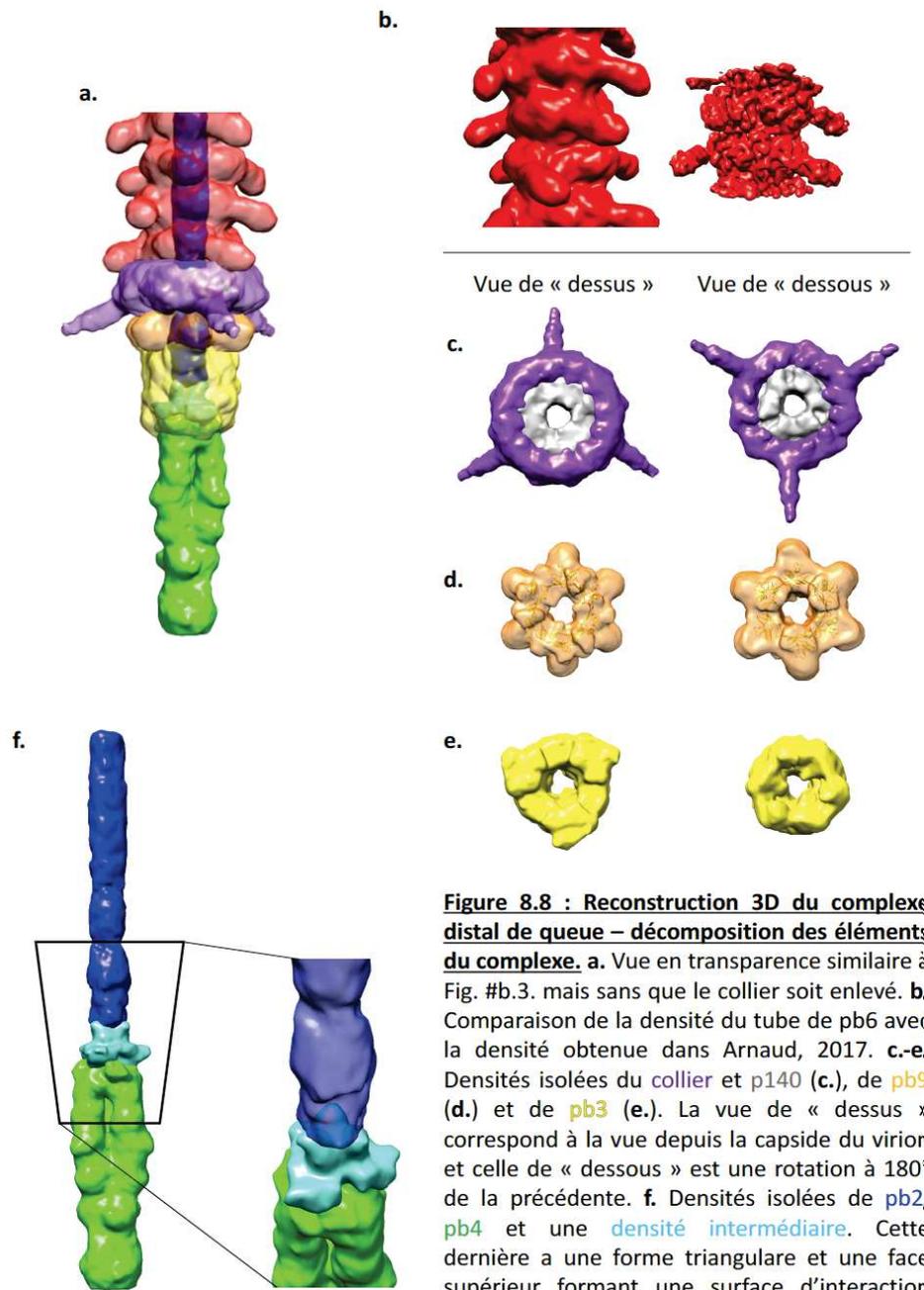


Figure 8.8 : Reconstruction 3D du complexe distal de queue – décomposition des éléments du complexe. **a.** Vue en transparence similaire à Fig. #b.3. mais sans que le collier soit enlevé. **b.** Comparaison de la densité du tube de pb6 avec la densité obtenue dans Arnaud, 2017. **c.-e.** Densités isolées du collier et p140 (**c.**), de pb9 (**d.**) et de pb3 (**e.**). La vue de « dessus » correspond à la vue depuis la capsid du virion et celle de « dessous » est une rotation à 180° de la précédente. **f.** Densités isolées de pb2, pb4 et une densité intermédiaire. Cette dernière a une forme triangulaire et une face supérieur formant une surface d'interaction évasé accueillant pb2.

8.2.2 Structure du complexe distal de queue en interaction avec FhuA

8.2.2.1 Coloration négative

L'interaction entre les queues de T5 et FhuA a été observée en coloration négative avec la protéine FhuA solubilisée en amphipols et en nanodisques (Fig. 8.9). Dans les deux cas, la protéine pb2 est éjectée de l'intérieur du complexe et la fibre droite n'est plus visible, la base de la protéine pb3 formant la nouvelle extrémité distale du complexe. D'autres changements de conformations sont particulièrement visibles sur l'échantillon queues-FhuA-nanodisque : le nanodisque peut être indentifié à la base et perpendiculaire à la queue, des densités peuvent être vues traverser le nanodisque et une structure supplémentaire est visible, faisant un coude aux côtés de pb3, pointant également vers le nanodisque. Il s'agit vraisemblablement de la protéine pb4 qui serait « rétractée » contre le reste du complexe. De plus, toujours dans cet échantillon, la protéine pb2 est souvent observée uniquement partiellement éjectée, alors qu'elle est complètement éjectée lorsque les queues sont incubées avec le récepteur solubilisé en amphipols. Cela suggère que la présence d'une bicouche est nécessaire pour la stabilisation du canal. Nous nous sommes focalisés sur l'échantillon où FhuA est solubilisé en nanodisque dans l'espoir d'observer des structures transmembranaires et le repliement de la protéine pb4 (Fig. 8.11-a-c).

8.2.2.2 Cryo-microscopie électronique

La reconstruction 3D a été réalisée sans imposer de symétrie afin de reconstruire également la structure asymétriquement positionnée sur le côté de pb3. 6640 particules ont permis une reconstruction à environ 15 Å de résolution (Fig. 8.10). A cette résolution, on observe peu de changements de conformations de p140, pb9 et pb3 par rapport à la structure avant éjection (Fig. 8.10-b). La densité au niveau du nanodisque est très fragmentée et difficilement exploitable. Ce manque d'informations au niveau du nanodisque s'explique par une homogénéité moins grande qu'en coloration négative de l'échantillon : de manière similaire à ce qui est observé dans l'échantillon queues-FhuA-Amphipols, le nanodisque avec FhuA et pb2 ne sont plus rattachés au complexe (collecte réalisée sur un échantillon moins homogène que celui observé en coloration négative). Par ailleurs, les différents degrés d'éjection de la protéine pb2 doivent également participer à l'hétérogénéité de l'échantillon.

Pb4 s'est repliée sur un côté de la structure en formant un angle très aigu sur le côté de pb3, la densité au niveau du coude est beaucoup moins bien définie suggérant que la protéine y est plus flexible. La partie proximale de la protéine

est rattachée au corps de pb3 par 3 domaines (peut-être les 3 domaines distincts formant le cône dans l'assemblage avant éjection), deux de ces domaines sont liés du même côté et s'étendent vers le coude en formant une double hélice. Le troisième domaine est rattaché à la face opposée de pb3 (Fig. 8.10-a). La partie distale de la protéine rejoint la zone du nanodisque de façon décalée par rapport à l'axe de la queue. Une densité mal définie est visible à cette position, correspondant probablement au complexe entre pb5 et FhuA.

Dans l'axe de la queue, 3 densités s'enfoncent dans la membrane à partir des 3 protubérances de pb3 et semblent former un canal (très morcelé dans la présente reconstruction) à côté de la densité de FhuA. Il n'y a plus de densité à l'intérieur de la queue (préalablement occupé par pb2), et un canal continu peut être observé de l'extrémité proximale du tube de pb6 jusqu'à pb3. De façon inattendu le canal est obstrué par une très fine densité à l'extrémité de pb3, il pourrait s'agir d'un artefact de la reconstruction ou d'un problème d'homogénéité des particules. D'après ces structures, peu ou pas de changement de conformations sont observé au niveau de pb3 (Fig. 8.11-d&e). Cette observation va à l'encontre de travaux déjà publiés sur la Tail de SPP1 [Goulet 11], il s'agit peut être d'un artefact de notre reconstruction, nous pourrions être plus affirmatif avec plus de données. Le nombre de particules actuellement collectées ne permet pas d'appliquer de stratégies de sélection des particules qui permettrait de discriminer les particules dont le nanodisque s'est désolidarisé.

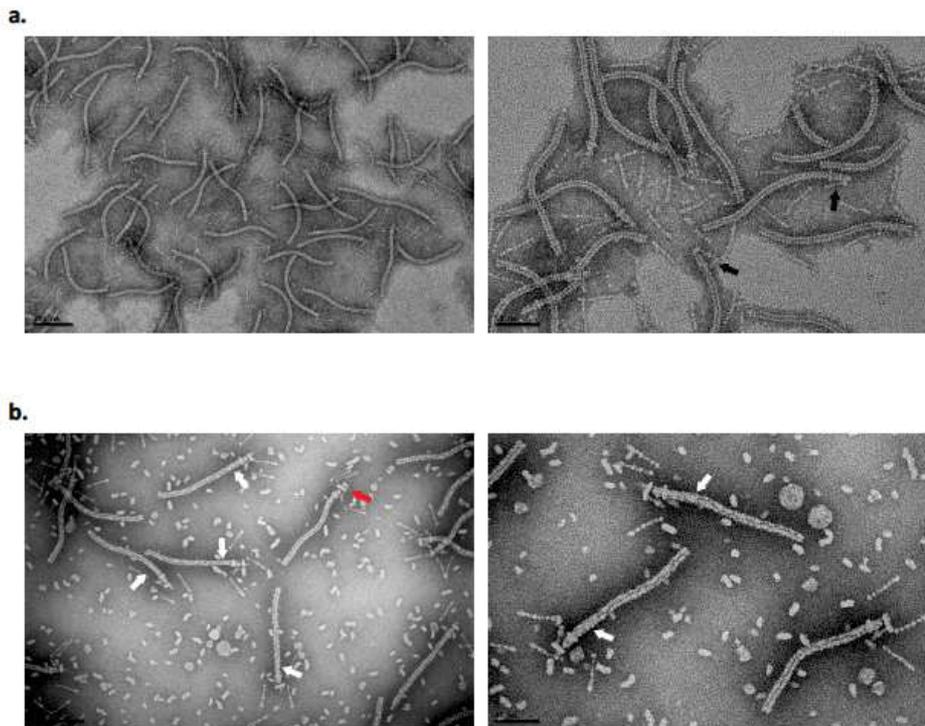
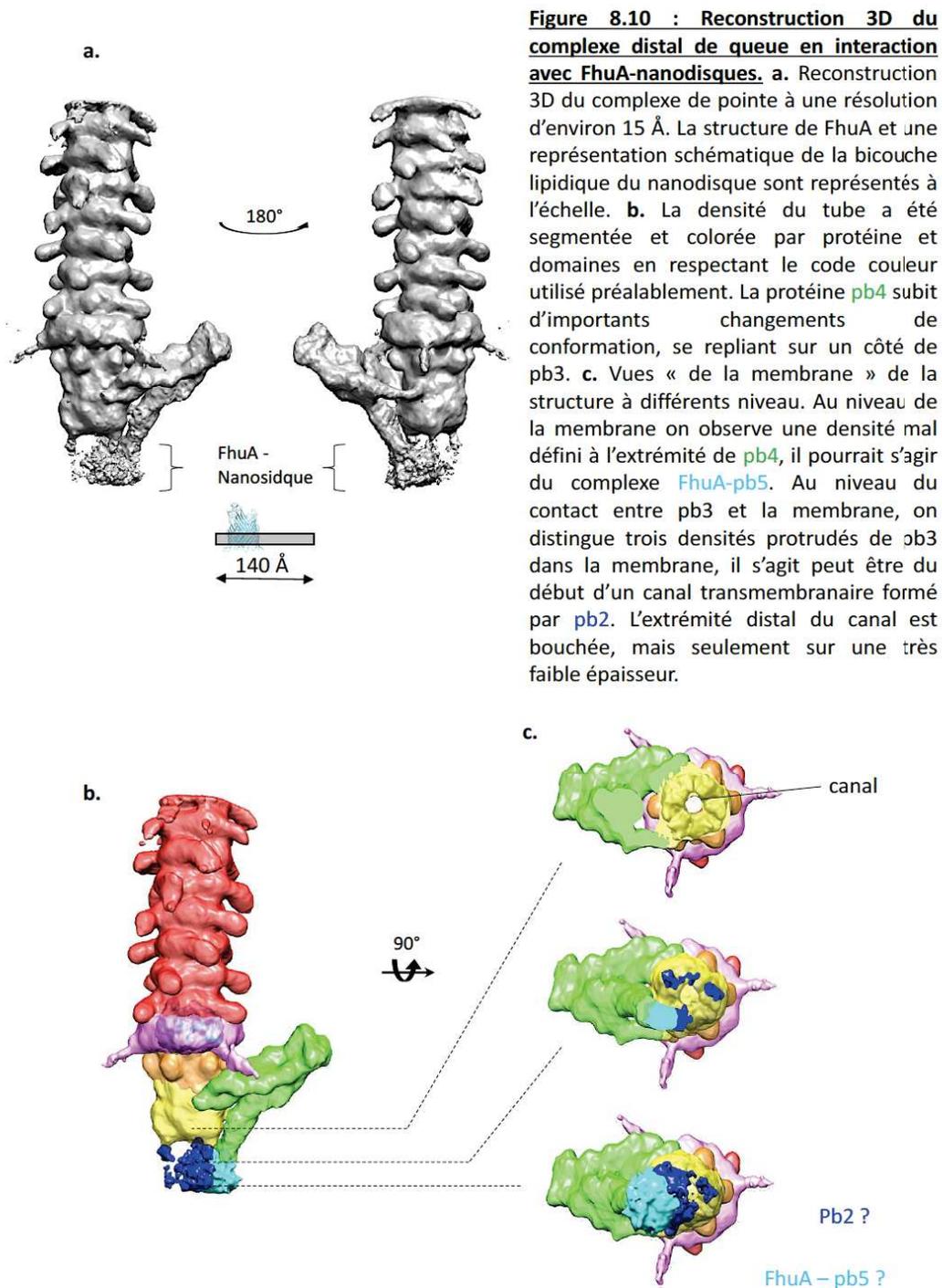


Figure 8.9 : Micrographes de queues de phages incubées avec FhuA. a. et b. Microscopie en coloration négative de queues de T5 à deux grossissements différents (barres). **a.** Queues incubées avec FhuA solubilisé en amphipoles. La fibre droite et la structure en cône à sa base ne sont plus visibles. pb2 a été éjectée des queues, de la coloration est visible à l'intérieur du tube et dans pb3. **b.** Queues incubées avec FhuA solubilisé en nanodisques. Les queues de phages sont fixés au nanodisques (perpendiculaires à l'axe du tube) au niveau de l'extrémité de pb3 (à montrer). Une structure parallèle à pb3 est visible sur presque toutes les particules, elle peut aussi être observée sur de rares particules de queues incubées avec FhuA-amphipole (flèches noires **a.**). pb2 est souvent partiellement éjectée (flèches blanches). Sur de rares particules, pb2 éjectée du tube traverse deux nanodisques (flèche rouge).



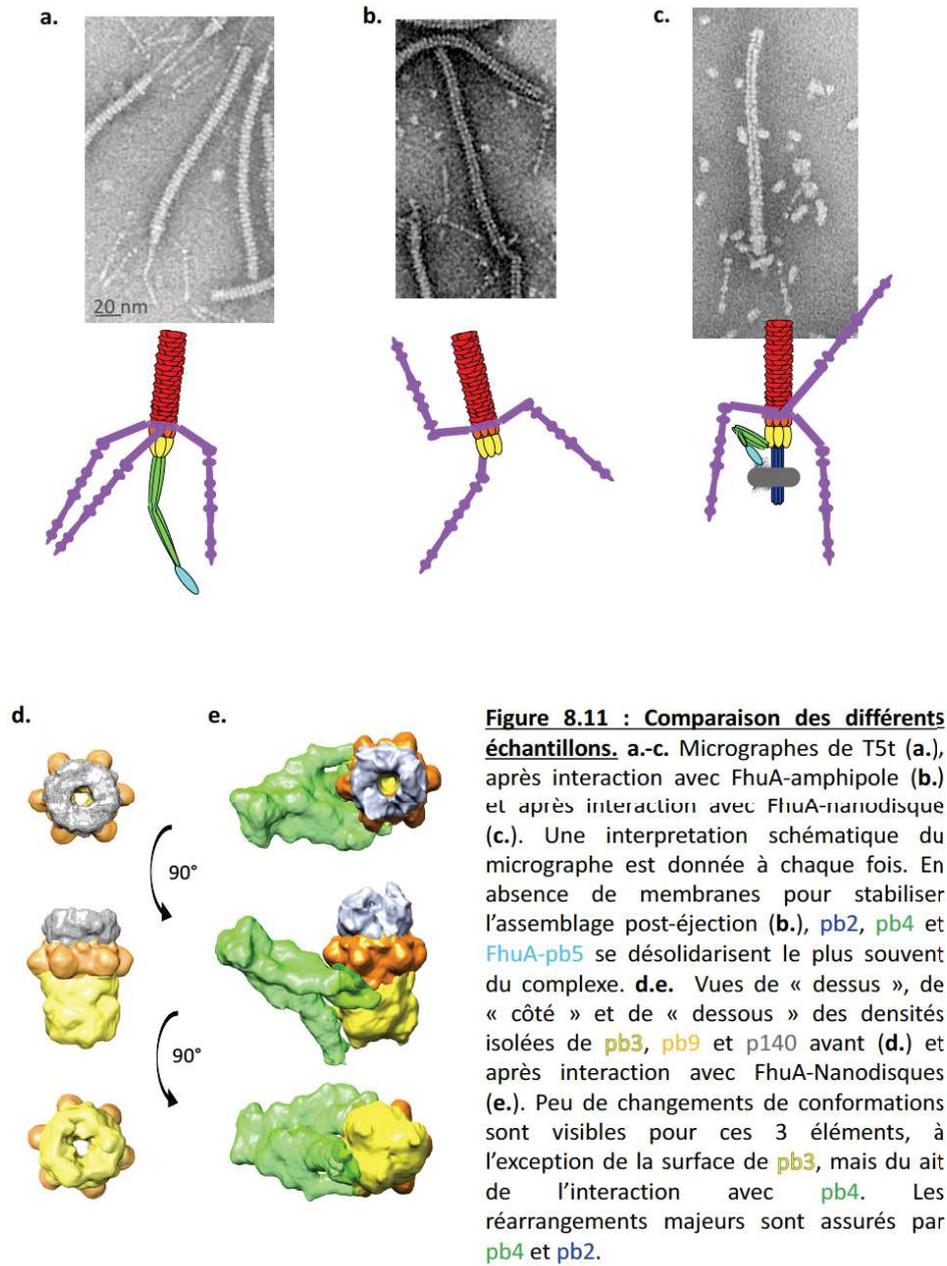


Figure 8.11 : Comparaison des différents échantillons. a.-c. Micrographes de T5t (a.), après interaction avec FhuA-amphipole (b.) et après interaction avec FhuA-nanodisque (c.). Une interprétation schématique du micrographe est donnée à chaque fois. En absence de membranes pour stabiliser l'assemblage post-éjection (b.), pb2, pb4 et FhuA-pb5 se désolidarisent le plus souvent du complexe. d.e. Vues de « dessus », de « côté » et de « dessous » des densités isolées de pb3, pb9 et p140 avant (d.) et après interaction avec FhuA-Nanodisques (e.). Peu de changements de conformations sont visibles pour ces 3 éléments, à l'exception de la surface de pb3, mais du ait de l'interaction avec pb4. Les réarrangements majeurs sont assurés par pb4 et pb2.

8.3 Perspectives

8.3.1 Expériences complémentaires

8.3.1.1 Protéolyse ménagée

Des expériences de protéolyse ménagée sur les queues incubées ou non avec FhuA solubilisé en détergent ont été réalisées (Fig. 8.12). Les changements de conformations induits par la liaison à FhuA ont pour conséquence de rendre les protéines pb2 et pb3 nettement plus sensible à la dégradation, ce qui est cohérent avec l'éjection de pb2 et le réarrangement de pb4 qui augmente la surface exposée de pb3. Aucun commentaire ne peut être fait sur la protéine pb4 dont la bande est masquée par celle de FhuA (volontairement ajouté en excès). Trois observations sont moins triviales à expliquer :

1. La dégradation rapide de pb2 dans l'échantillon de queues. Il est en effet contre-intuitif que la protéine pb2, qui se trouve au cœur du complexe, semble être la première dégradée (diminution notable de la densité de pb2 à partir de 0,4 µg/mL de subtilisine). Une explication possible est qu'une partie du terminateur de queue (p142) ou du connecteur (p143) impliqué dans la fermeture et le maintien de pb2 à l'intérieur du tube soit accessible à la subtilisine, permettant la sortie de pb2. Cette sortie de pb2 par l'extrémité proximale de la queue a déjà été observée en microscopie électronique (cf. 8.2.2.1 et Fig. 8.5) et serait cohérente avec l'hypothèse que pb2 est contrainte dans un état métastable à l'intérieur du tube.
2. La sensibilité accrue de pb1 en présence de FhuA (Fig. 8.12-b&c). Pb1 jouant un rôle accessoire dans le mécanisme de pénétration, il n'est pas évident d'expliquer cette sensibilité accrue. Il est possible que la dégradation de pb3 permette une désolidarisation des fibres flexibles du reste du complexe ou qu'elles soient affectées par le réarrangement de pb3 et pb4.
3. La digestion partielle de pb6 à de faibles concentrations de subtilisine en présence de FhuA (Fig. 8.12-c). Le fait que la digestion observée de pb6 ne se poursuive pas au-delà de la perte d'un segment de 10 kDa rend très probable qu'il s'agisse du domaine Ig-like. Aucun changement de la flexibilité de ce domaine n'a été observé sur les reconstructions 3D du tube et ce segment n'est pas impliqué dans l'assemblage. Il n'est donc pas raisonnable de penser que la protéine pb6 subisse un changement de conformation qui affecterait l'exposition de ce domaine à la subtilisine. Une autre explication pourrait être la présence de détergent dans l'échantillon.

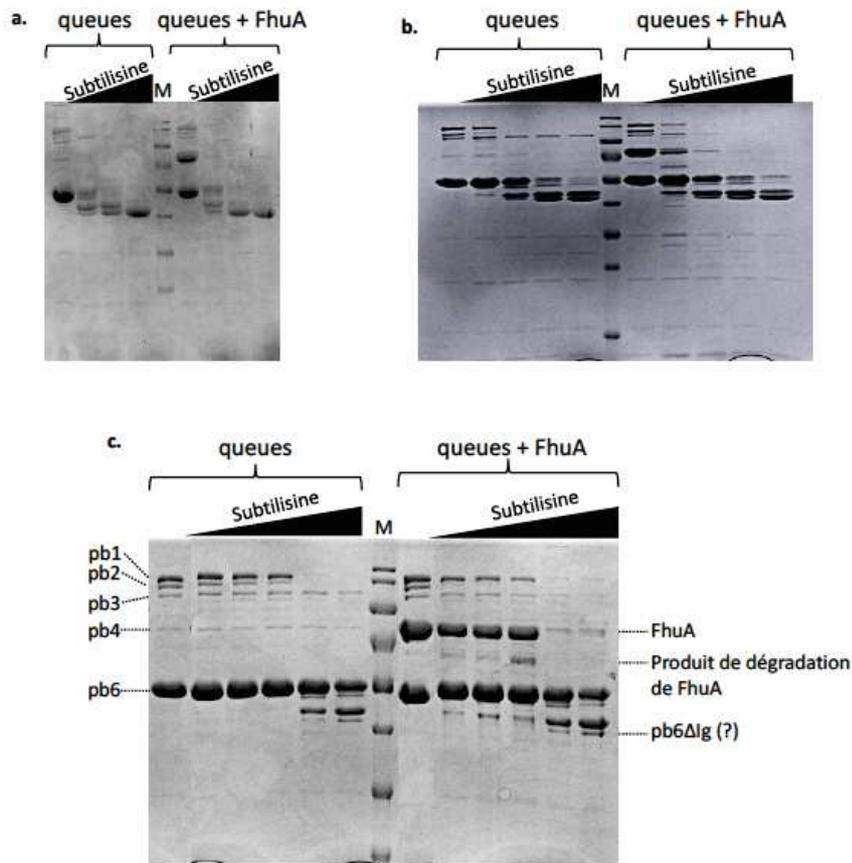


Figure 8.12 : protéolyse ménagée des queues de T5. a-c. Queues de T5, préalablement incubées ou non avec FhuA, incubées 1 h à 37°C avec de la subtilisine. a. 0 ; 0,02 ; 0,1 ; 0,5 mg/mL subtilisine. À 0,02 mg/mL de subtilisine on observe une persistance des bandes de pb3 et pb4 en absence de FhuA. La digestion de la protéine pb6 se fait par 2 paliers majeurs et la bande la plus basse est environ 10 kDa plus petite que la protéine intact, cette taille est cohérente avec le domaine *lg-like* qui n'est pas impliqué dans la formation du tube. Et en présence de FhuA ? b. 0 ; 1 ; 5 ; 25 ; 50 µg/mL subtilisine. En absence de FhuA, la protéine pb3 n'est pas dégradée dans ces conditions et les protéines pb1 et pb2 sont nettement moins dégradées qu'en présence de FhuA pour 1 µg/mL de subtilisine. c. 0 ; 0,1 ; 0,2 ; 0,4 ; 2 ; 4 µg/mL subtilisine. En présence de FhuA, pb2 est entièrement digéré à toutes les concentrations observées de subtilisine. Les protéines pb1 et pb3 commencent également à être dégradées. Le complexe natif est très peu affecté par de faibles concentrations de subtilisine ($\leq 0,4$ µg/mL).

8.3.1.2 Expériences à réaliser

L'acquisition et le traitement de plus de données de cryo-microscopie électronique est nécessaire à des reconstructions de plus hautes résolutions, permettant une segmentation et une description plus fine de l'assemblage, avant et après interaction avec FhuA. Sur l'échantillon queues-FhuA-Nanodisques, plus de données nous permettraient de contourner l'hétérogénéité de l'échantillon grâce à des stratégies de classification 3D.

Par ailleurs, plusieurs types d'expériences sont envisagés pour compléter la caractérisation de la queue du phage T5 :

- Les expériences de protéolyse ménagée réalisée jusqu'à présent ont donné des résultats prometteurs, elles nécessitent néanmoins d'être complétés par des études similaires menées sur le complexe queues-FhuA-Nanodisques qui pourrait permettre d'observer une stabilisation au moins partielle de pb2. De plus, une expérience contrôle de queues incubées avec du détergent est à prévoir. Remarque : ces expériences ont pu être réalisées après le rendu de cette thèse, en confirme les assumptions et sont brièvement présentées dans les annexes de cette édition (C.3).
- Des expériences d'immunomarquage similaire à celle proposées dans [Zivanovic 14] sont prévues. L'objectif est la localisation de pb2 et de pb5 après interaction avec FhuA solubilisé en nanodisques, ainsi que la localisation de p140. Cette dernière étant probablement masquée par le collier d'attache des fibres flexibles, le mutant T5hd1 (dépourvu de fibres flexibles) sera utilisé et l'assemblage natif servira de contrôle.
- L'effet de queue de phages purifiées sur des cultures de *E. coli* doit être exploré. Cette expérience permettra de déterminer si le canal formé par la queue de phage dans la paroi de la cellule reste « ouvert » en absence de capsid et d'ADN à transférer.
- Une caractérisation de l'interaction FhuA – queues par SPR est également envisagée. Une étude sur l'interaction FhuA-pb5 avait conclue sur une affinité très forte de ce complexe mais n'avait pas permit de déterminer de données chiffrée. Nous souhaiterions voir si l'interaction observé est aussi forte avec le complexe entier et étendre la gamme de concentration testée pour déterminer un Kd.

Points clés

Résumé chapitre 8

- ❑ pb3 et pb4 forment des arrangements de symétries 3 et le colier a une symétrie 12
- ❑ La position probable de p140 a été identifiée
- ❑ Les reconstructions préliminaires du complexe distale recèlent des informations qui permettent de faire une ébauche du mécanisme de perforation de la paroi par T5 (cf. Conclusion)
- ❑ Des informations biochimiques et de microscopie électronique à coloration négative confirment l'hypothèse que la TMP se trouve dans un état instable au sein du complexe, renforçant l'hypothèse que la TMP assure la transmission de l'information à la capsidémie émise dans le chapitre 5
- ❑ Un travail d'acquisition et de traitement de données, ainsi que la mise en œuvre d'expériences complémentaires doivent être réalisés pour confirmer et affiner ces résultats.

Conclusion

Bilan des travaux effectués

La protéine du tube de la queue de T5 (pb6) a été purifiée sous forme de monomère et de fibres insolubles. La polymérisation de la protéine semble pouvoir être provoquée en absence des autres partenaires formant la queue en milieu acide et en présence de sels (<100 mM).

La structure cristallographique de pb6 monomère a été déterminée révélant que la protéine est formée de trois domaines, un domaine Ig-like de décoration et deux domaines dupliqués ayant un repliement homologue à celui des autres TTPs de phage connues et des protéines de tube du SST 6 et de la pyocine R. La structure de pb6 sous forme de tube a été résolue à résolution subnanométrique pour les fibres et le tube de la queue de T5, avant et après interaction avec FhuA. Ces structures sont similaires et aucun changement conformationnel n'est observé entre fibres, tube de queue et tube de queue après interaction.

Plusieurs constructions de pb6 ont été produites avec différents marquages pour permettre son analyse par RMN. Les fibres de pb6 peuvent être analysées en ssRMN ouvrant des perspectives de comparaisons entre monomères et fibres ainsi que la possibilité d'utiliser ces données pour résoudre la structure du tube de pb6 à haute résolution.

Des structures du complexe distal de queue ont été obtenues, ses travaux doivent être affinés et conclus d'ici la fin 2017.

Modèle de l'infection de T5

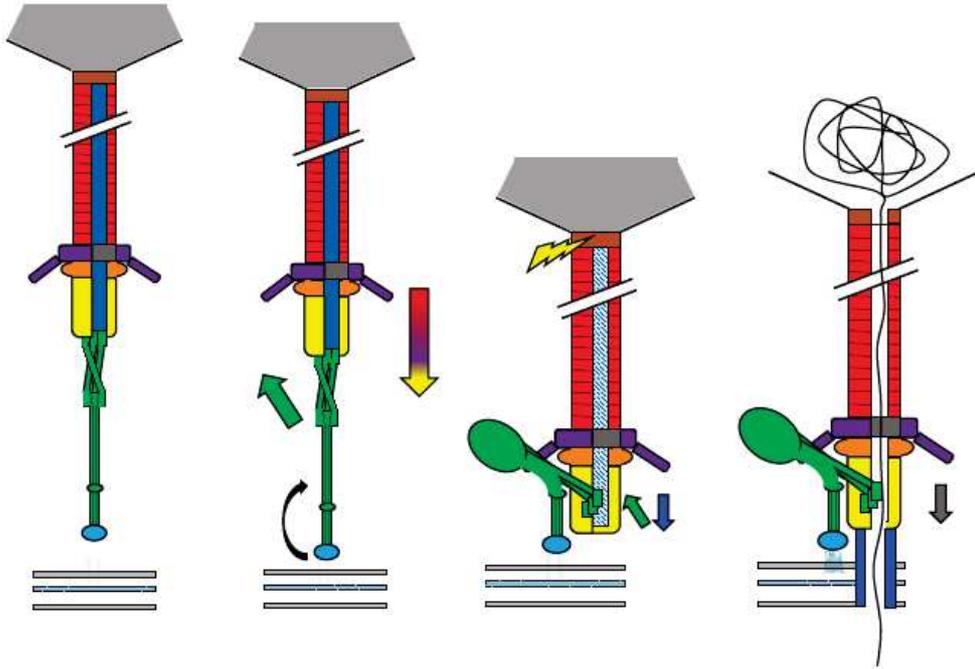
Les travaux produits au cours de cette thèse permettent de compléter le modèle d'infection pour le phage T5 (Fig. 9) :

- Les fibres flexibles permettent de fixer réversiblement le LPS à la surface de l'hôte, leur action augmente l'adhésion et l'infectivité du phage mais ne sont pas nécessaires.
- La protéine pb5 (RBP) à l'extrémité de la fibre droite fixe la protéine FhuA, cette interaction est nécessaire et suffisante pour déclencher la perforation de la paroi cellulaire et l'éjection de l'ADN. Il est probable que cette interaction

provoque un changement de conformation jusqu'au coude de la fibre droite qui induise la désolidarisation et le réarrangement de la fibre entre le coude et la densité attribuée à pb3.

- Le réarrangement de la fibre droite en une structure repliée sur le côté du complexe distal permet de rapprocher l'extrémité de pb3 de la membrane. Ce mécanisme peut être comparé à l'étape où les phages ayant de grandes plateformes d'adhésion avec de multiples RBP (récepteur saccharidiques) sont fixés à la membrane par de multiples interactions, orientant de façon contrôlée la queue par rapport à la paroi de l'hôte. Il sera intéressant de découvrir si ce mécanisme d'arrimage à la membrane décrit pour T5 est commun ou non à tous les phages ayant de petites plateformes d'adhésion (récepteur protéique).
- Le réarrangement de la fibre droite permet également d'exposer (et peut-être de « tirer ») l'extrémité C-terminale de pb2 (TMP). Cette dernière a des propriétés d'insertions dans les membranes et doit ainsi provoquer le reste de son éjection. Nos résultats montrent que la protéine pb2 est maintenue dans un état métastable au sein du complexe, rendant cet important changement de conformation thermodynamiquement viable. Les observations d'éjections partielles de pb2 suggèrent que l'éjection totale de la protéine pourrait être favorisée par la présence d'une seconde membrane pour se stabiliser ou par l'activité de dégradation du peptidoglycane.
- L'absence de changements de conformations de la TTP va à l'encontre de ce qui avait été proposé par SPP1 et suggère que c'est l'éjection de la TMP qui provoque un changement de conformation dans l'interface capsid-queue. Ce mécanisme de transmission de l'ADN pourrait être commun à tous les *Siphoviridae* mais aussi aux *Myoviridae* pour lesquelles le comportement de la TMP au cours de l'infection a été peu étudié.
- L'ouverture de la capsid, l'éjection de l'ADN et la régulation de cet éjection (chez T5 l'éjection de l'ADN a une « pause » ; cf. 1.5) n'ont pas été décrits au niveau moléculaire (y compris la force motrice de l'éjection de l'ADN).

Figure 9 : Schéma du mécanisme d'entrée de T5



Architectures de queues de phage et mécanismes d'entrées

Ces nouvelles informations sur le mécanisme d'entrée de T5 ne permettent pas de tirer un modèle simple généralisable à une famille, un groupe ou sous-groupe identifié de virus.

Cependant, plusieurs hypothèses sont suggérées par ces résultats obtenus au cours de ce travail de thèse et des informations disponibles pour d'autres systèmes :

- Les 3 familles de virus caudés (*Myoviridae*, *Podoviridae* et *Siphoviridae*), sont basées sur les principales architectures de queues et il semblait originellement raisonnable de penser que chaque type d'architecture était associé à un mécanisme. Cependant, le faible échantillonnage de structures de phages à une meilleure résolution révèle que chaque famille (queues longues et contractiles, longues et non-contractiles et petites queues) est constituée de différentes architectures et de mécanismes variés 1.4.
- Par ailleurs, des structures communes entre différents phages ont été mis en avant. Cette homologie de structure a été particulièrement bien décrite pour les protéines de tubes mais a également été suggéré pour d'autres protéines du complexe distal. La comparaison des complexes distaux de T4 et de T5 est particulièrement intéressante (Fig.9.2).

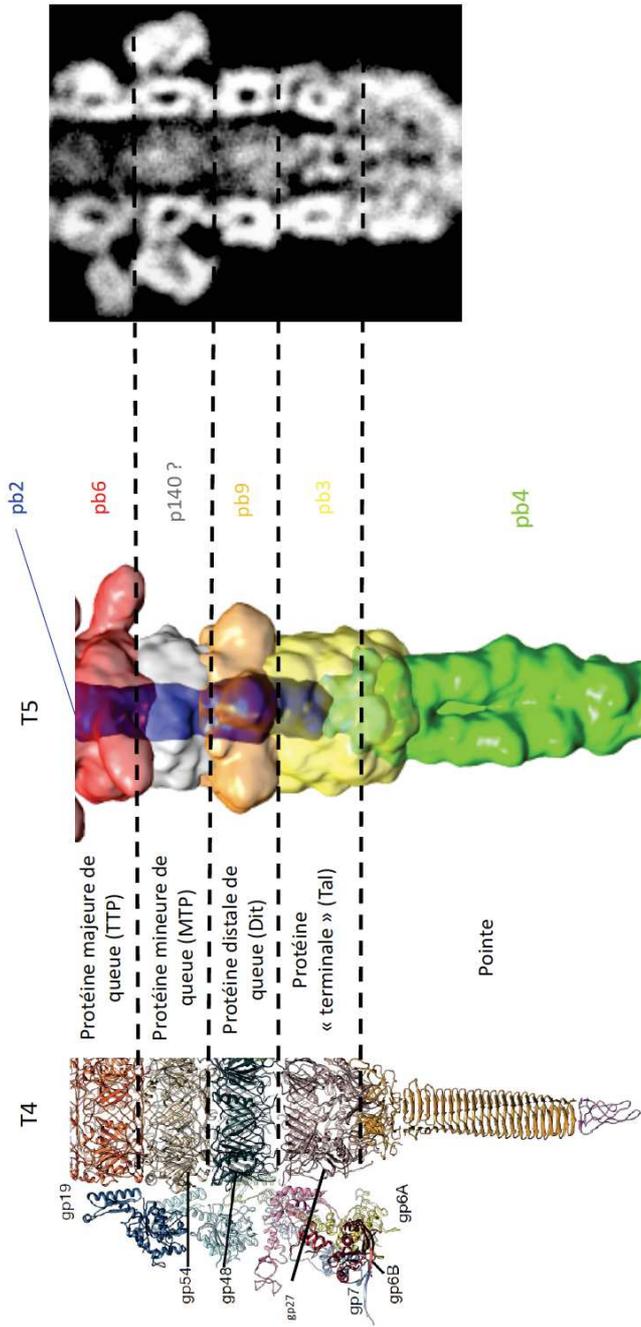


Figure 9.2 : Comparaison du complexe distal de T5 avec la partie interne du complexe distal de T4.

De gauche à droite : structure de la partie interne de l'extrémité de la queue de T4 (figure issue de [Taylor et al. 2016]) ; représentation segmentée de la structure du complexe distal T5 ; représentation en coupe du complexe distal de T5. On observe un enchaînement de structures en anneaux hexamériques : dernier anneau de la protéine majeur de queue, protéine mineure de queue, protéine distale et protéine terminale. Cette homologie pourra être confirmée avec une structure à meilleure résolution du complexe distale de T5. Les pointes des 2 virus sont manifestement différentes et, comme discuté précédemment, la jonction pb4/pb3/pb2 est difficile à décrire chez T5. Néanmoins, sur la base de la comparaison avec T4, il est possible que la partie distale de la densité attribuée à la protéine pb3 soit en réalité un domaine de la protéine pb4.

- Le fait que l'assemblage du tube et du complexe distal interne aient une structures si proche suggère que le rôle de stabilisation d'une structure à travers la paroi cellulaire et de déclenchement de l'ouverture de la capsidie par la TMP que nous avons proposé chez T5 pourrait être commun à plusieurs espèces de *Siphoviridae* et *Myoviridae*. Des études supplémentaire sur le fonctionnement des TMP de différents phages sont nécessaires.
- Plutôt que de classer les *Caudovirales* en 3 groupes figés, il est peut-être plus pertinent de considérer les queues de phages comme des architectures modulaires, basées sur un nombre limité de structures impliquées dans des mécanismes différents. Une queue de phage pourrait se concevoir comme un assemblage de différentes fonctions : adhésion et attachement à la membrane, perforation de la paroi bactérienne, éjection ADN. Cette approche pourrait être généralisable à d'autre structures virales et pourrait permettre une meilleure compréhension de la phylogénie de ces éléments, utilisant ente autre leurs homologies structurales. L'acquisition de plus de structures à haute résolution permise par le développement de la microscopie électronique ainsi que le développement d'outils de bioinformatique spécifiques devrait permettre de répondre à ces interrogations dans les années à venir.

Annexes

Annexe A

**Article : Solid-state NMR
H-N-(C)-H and H-N-C-C 3D/4D
correlation experiments for
resonance assignement of large
proteins**

Special
Issue

Solid-State NMR H–N–(C)–H and H–N–C–C 3D/4D Correlation Experiments for Resonance Assignment of Large Proteins

Hugo Fraga,^[a, b] Charles-Adrien Arnaud,^[a] Diego F. Gauto,^[a] Maxime Audin,^[c] Vilius Kurauskas,^[a] Pavel Macek,^[a] Carsten Krichel,^[a] Jia-Ying Guan,^[a] Jerome Boisbouvier,^[a] Remco Sprangers,^[c, d] Cécile Breyton,^[a] and Paul Schanda^{*[a]}

Solid-state NMR spectroscopy can provide insight into protein structure and dynamics at the atomic level without inherent protein size limitations. However, a major hurdle to studying large proteins by solid-state NMR spectroscopy is related to spectral complexity and resonance overlap, which increase with molecular weight and severely hamper the assignment process. Here the use of two sets of experiments is shown to expand the tool kit of ¹H-detected assignment approaches, which correlate a given amide pair either to the two adjacent CO–CA pairs (4D hCOCANH/hCOCAcoNH), or to the amide ¹H of the neighboring residue (3D HcocaNH/HcacoNH, which can be extended to 5D). The experiments are based on efficient coherence transfers between backbone atoms using INEPT trans-

fers between carbons and cross-polarization for heteronuclear transfers. The utility of these experiments is exemplified with application to assemblies of deuterated, fully amide-protonated proteins from approximately 20 to 60 kDa monomer, at magic-angle spinning (MAS) frequencies from approximately 40 to 55 kHz. These experiments will also be applicable to protonated proteins at higher MAS frequencies. The resonance assignment of a domain within the 50.4 kDa bacteriophage T5 tube protein pb6 is reported, and this is compared to NMR assignments of the isolated domain in solution. This comparison reveals contacts of this domain to the core of the polymeric tail tube assembly.

1. Introduction

Rapid progress has been made in magic-angle spinning (MAS) solid-state NMR (ssNMR) over the last decade, which has enabled the study of the structure, interactions and dynamics of proteins of increasing size and complexity.^[1–5] It has also afforded new opportunities to study proteins in their native environ-

ment, such as in cell walls and membranes.^[6–8] Resonance assignment remains a crucial and often time-consuming step to be accomplished before atomic-resolution information can be obtained from ssNMR data. Resonance overlap is a major challenge for the assignment process, particularly for large proteins due to the large number of resonances. In addition to the overlap problem, the intensity of individual cross-peaks decreases with the size of the protein, because of the lower number of molecules in the fixed volume of the sample rotor. A powerful strategy to increase resolution and sensitivity is based on proton-detected experiments applied at high MAS frequencies, with deuterated samples that are re-protonated at amide sites.^[9–14] Fast MAS and deuteration extend the coherence lifetimes T_2' of ¹H, ¹⁵N and ¹³C nuclei, resulting in two beneficial effects: 1) the associated narrow line widths increase resolution and intensity, and 2) the coherence transfer between nuclei becomes more efficient, which further increases sensitivity in correlation experiments. Different 3D experiments have been proposed to correlate the amide site (¹H–¹⁵N) to the ¹³C α , ¹³C β and ¹³CO nuclei of the same amino acid residue (intra-residue H_i–N_i–C_i) and preceding residue (H_i–N_i–C_{i–1}).^[15–17] Sequential assignment of amides is achieved by matching the ¹³C_i and ¹³C_{i–1} frequencies of different amide sites. Furthermore, it has been proposed to correlate the ¹H–¹⁵N frequency pair to the ¹⁵N frequency of the neighboring residue.^[18,19]

[a] Dr. H. Fraga, C.-A. Arnaud, Dr. D. F. Gauto, Dr. V. Kurauskas, Dr. P. Macek, Dr. C. Krichel, Dr. J.-Y. Guan, Dr. J. Boisbouvier, Dr. C. Breyton, Dr. P. Schanda
Univ. Grenoble Alpes, CEA
CNRS, Institute for Structural Biology (IBS)
71 avenue des martyrs, 38044 Grenoble (France)
E-mail: paul.schanda@ibs.fr

[b] Dr. H. Fraga
Departamento de Bioquímica
Faculdade de Medicina da Universidade do Porto
Porto (Portugal)

[c] Dr. M. Audin, Prof. Dr. R. Sprangers
Max Planck Institute for Developmental Biology
Spemannstrasse 35
72076 Tübingen (Germany)

[d] Prof. Dr. R. Sprangers
Department of Biophysics I
University of Regensburg
93053 Regensburg (Germany)

Supporting Information and the ORCID identification numbers for the authors of this article can be found under:
<https://doi.org/10.1002/cphc.201700572>

An invited contribution to a Special Issue on Physical Chemistry in France

Overlap of the H–N frequencies, that is, peak overlap in the two-dimensional “root” spectrum, increases the ambiguity of connecting a given H–N pair to the corresponding intra- and inter-residue ^{13}C frequencies, and, therefore, severely complicates the assignment process, particularly for large (> 20 kDa) proteins. We address this problem here by introducing two complementary approaches for correlating the amide sites with either the $^{13}\text{C}\alpha$ and ^{13}CO carbon atoms flanking the given amide site, resulting in a pair of 4D ($\text{H}_i\text{-N}_i\text{-C}\alpha_i\text{-CO}_i$, $\text{H}_i\text{-N}_i\text{-C}\alpha_{i-1}\text{-CO}_{i-1}$) spectra, or to the amide protons of either of the neighboring residues, resulting in another set of correlation spectra ($\text{H}_i\text{-N}_i\text{-H}_{i-1}$, $\text{H}_i\text{-N}_i\text{-H}_{i+1}$).

The choice of the transfer elements in sequential correlation experiments must ensure the highest signal-to-noise ratio. For heteronuclear transfers (H–C, C–N, N–H), ramped cross-polarization transfer is generally found to be most favorable. For C–C transfers, different schemes have been proposed, using either scalar-coupling-based INEPT steps^[16,17] or dipolar-based transfers.^[20,21] Under conditions of fast MAS (> 50 kHz) and deuteration, the ^{13}C coherence lifetimes are long (typically of the order of > 10 ms for $^{13}\text{C}\alpha$ and > 25 ms for ^{13}CO ^[22]). Under such conditions, INEPT-based transfers generally outcompete those that are dipolar-based. Of central importance to the present study is the realization that the INEPT-based transfer scheme allows the simultaneous labeling of the $^{13}\text{C}\alpha$ or ^{13}CO resonance frequency during the transfer delay, a property that dipolar-based transfer schemes do not have. Therefore, the addition of ^{13}C frequency dimensions is without further loss of sensitivity (except for the $\sqrt{2}$ loss inherent to quadrature detection in any indirect dimension) because frequency editing and transfer are achieved during the same delay, which represents a significant advantage over dipolar-transfer based schemes.^[23] Constant-time evolution has the additional benefits that the corresponding time-domain signal does not decay, thus improving resolution enhancement through linear prediction, and that the effect of homonuclear J -couplings is removed;^[24] the latter is particularly interesting for $^{13}\text{C}\alpha$, which is difficult to decouple from $^{13}\text{C}\beta$. Here, we present a suite of sequential resonance assignment experiments, based on these general design principles. We investigated the usefulness of this approach through measurements on different assemblies of proteins with monomer sizes ranging from 20 to 60 kDa.

2. Results and Discussion

2.1. Design of 4D and 3D Assignment Experiments

Figure 1 shows the pulse sequences proposed in this study for generating 3D and 4D correlation spectra. One pair of experiments, HcocaNH and HcacoNH, correlates a given H–N pair to the $^{13}\text{C}\alpha$ and ^{13}CO nuclei of the same residue (4D hCOCANH) or the preceding residue (4D hCOCAcNH). In a second pair of experiments, the amide H–N pair is connected to the amide ^1H frequency of the preceding (HcacoNH) or consecutive (HcocaNH) residue. The main aim of this pair of experiments is to provide an additional nucleus—the amide proton—to complement and confirm sequential assignments. Inherently, the

drawback of using two proton frequencies is that proton line widths are generally larger than heteronuclear line widths, and resolution is, thus, limited. For this reason, this pair of spectra is intended primarily to confirm sequential connectivities. Note that it is straightforward to extend these latter pulse sequences to 4D or 5D experiments (two ^1H dimensions, one ^{15}N dimension, and one or two ^{13}C dimensions). In these experiments, the ^{13}C dimensions may be edited in a constant-time manner during the INEPT delays,^[24] in the manner that is used in the hCOCANH and hCOCAcNH experiments shown in Figure 1, without additional relaxation losses.

In this study we used a range of different ^2H , ^{13}C , ^{15}N -labeled proteins that represent a significant challenge for contemporary ssNMR in terms of their size and thus spectral complexity: 1) the caseinolytic protease assembly ClpP from *Thermus thermophilus*, which forms a double-ring structure of a total molecular weight of 300 kDa, which is composed of 14 identical 21 kDa subunits; 2) the 173 kDa archaeal exosome core particle, which consists of three heterodimers (each formed by the proteins Rrp41 and Rrp42); in the sample used here, the 275-residue Rrp42 subunit was isotope-labeled, and in complex with unlabeled Rrp41; 3) pb6, which is the major protein that forms the tail of bacteriophage T5; the sample used in this study consists of long tubular tail-like particles assembled by the approximately 50 kDa pb6 monomers; 4) the thermosome protein from *Pyrococcus horikoshii*, which forms hollow particles of 16 60 kDa subunits; we used a sedimented preparation of these particles. These four proteins differ considerably in their size, three-dimensional structure, oligomeric state and method of preparation, and their spectra and relaxation properties provide, to some extent, a general view of the properties of a range of samples typically amenable to solid-state NMR.

The proton-detected 2D HN and 3D hCANH and 3D hCONH spectra of these proteins are shown in Figure 2, along with several traces across correlation peaks, indicating typical line widths (further discussed below using relaxation measurements).

The coherence lifetimes of the ^1H , ^{13}C and ^{15}N nuclei are key parameters dictating the resolution of correlation spectra, and influence the efficiency of coherence transfer steps. We measured the apparent coherence lifetimes for these four proteins using spin–echo relaxation delays (see the Supporting Information for details on the methods). These experiments were performed as series of 1D experiments, and thus report on the bulk of amide sites in these proteins. The apparent coherence lifetimes, reported in Figure 3A, provide an indication of average lifetimes. Coherence lifetimes of the order of up to 50 ms were found for ^{15}N and ^{13}CO , whereas amide- ^1H spins relax within approximately 5–7 ms, and the apparent decay of $^{13}\text{C}\alpha$ was of the order of 9–15 ms (note that the scalar coupling to $^{13}\text{C}\beta$ was not decoupled in these measurements but deconvoluted, see the Supporting Information and Figure S2). The coherence lifetimes observed in these various samples of oligomeric complexes are similar to previously reported values on smaller and most often crystalline protein samples.^[11,22,27,28] The calculated line widths L , based on these coherence lifetimes T_2' [$L = 1/(T_2'\pi)$], are of the order of 50–70 Hz ($^1\text{H}^N$), and

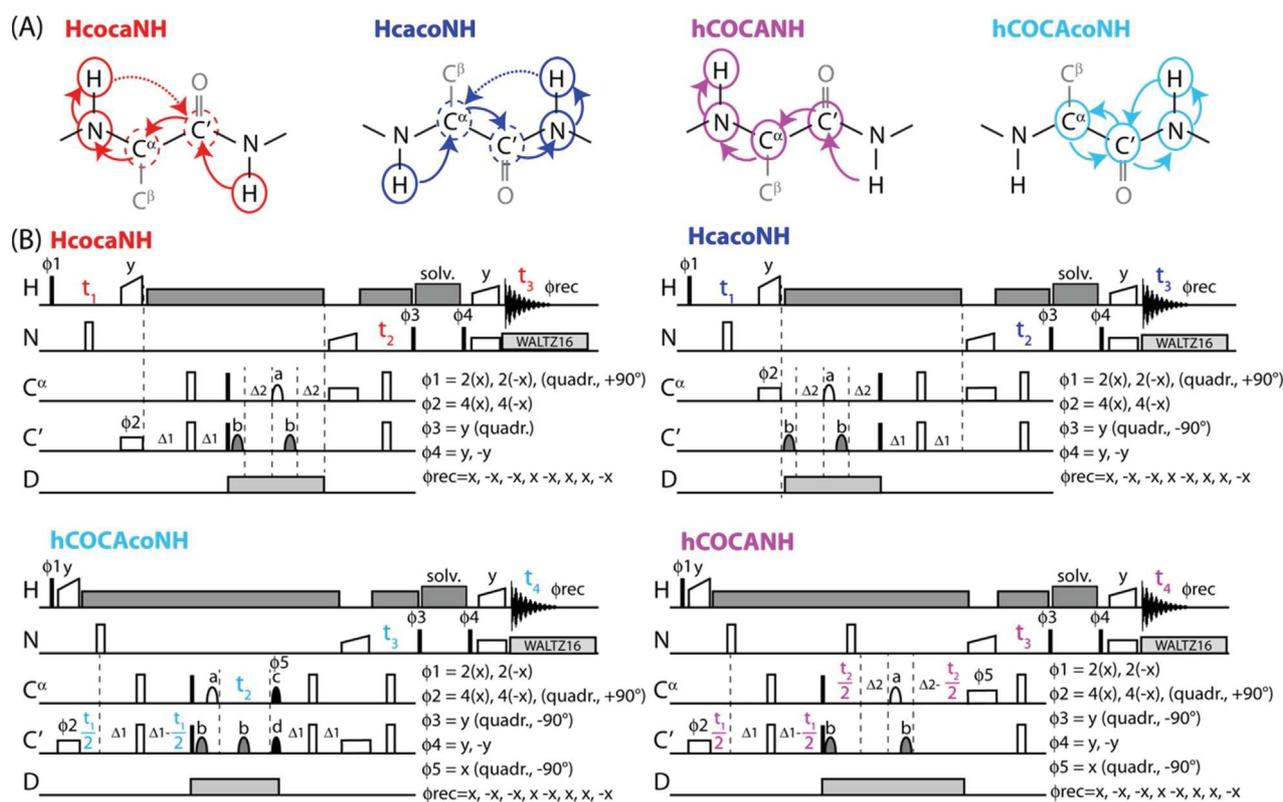


Figure 1. The proton-detected resonance assignment approaches used in this study. A) Coherence transfer pathways in the four proposed experiments. Encircled nuclei are frequency-edited; dashed circles denote nuclei, which could be edited in a constant-time manner during transfer, extending these experiments to 4D or 5D. Capital letters denote frequency dimensions that were edited in the experiments presented in this study. B) Corresponding pulse sequences. Narrow filled (open) rectangles denote hard 90° (180°) pulses; the shaped symbols on the ^{13}C channel represent REBURP^[25] (denoted "a"), ISNOB2^[26] ("b"), EBURP2^[25] ("c") and time-reversed EBURP2 ("d") shapes, with band widths of 70 ppm for pulses applied to CO (centered at 175 ppm), and 32 ppm for pulses applied to $\text{C}\alpha$ (centered at 54 ppm). Cross-polarization steps are denoted by open symbols with a linear ramp on one of the channels. "solv." indicates suppression of solvent signals; it is composed of a composite sequence of RF pulses with constant amplitude, a duration of approximately 850 μs per pulse element, applied with variable phases (either randomized or following a WALTZ-16 scheme), and a total duration of typically 60–130 ms. Decoupling of the ^2H spins (≈ 3 kHz WALTZ16) is optional and requires suitable hardware. ^1H decoupling (10 kHz WALTZ16) was used during all indirect chemical-shift evolution periods and INEPT transfers. During acquisition, 3 kHz WALTZ-16 decoupling was applied on ^{15}N . The INEPT transfer delays were typically 4.2–4.7 ms, and were optimized individually for the periods with transverse magnetization on CO (Δ_1) and $\text{C}\alpha$ (Δ_2). Note that in the present implementation the maximum evolution time, and thus the digital resolution, is limited by the total INEPT transfer delay (≈ 9 ms); although this evolution time is typically sufficient, semi-constant-time evolution schemes^[24] can be implemented in a straightforward manner, albeit at a cost in sensitivity.

25–35 Hz ($^{13}\text{C}\alpha$), which are in reasonable agreement with experimentally observed line widths (cf. Figure 2). For ^{15}N and ^{13}C the predicted line widths (6–10 Hz) are about two- to threefold lower than experimentally observed ones, suggesting the presence of additional line broadening due to sample heterogeneity and/or magnetic field inhomogeneity. Although these calculated and experimentally observed line widths are similar to well-behaving crystalline protein samples, the sheer number of cross-peaks in these samples of large proteins (ca. 200–500 expected cross-peaks) results in heavy resonance overlap, as reflected in Figure 2. The 4D strategy largely resolves these ambiguities and facilitates sequential walks, because two frequencies ($^{13}\text{C}\alpha$, ^{13}C) are simultaneously matched.

We next investigated the efficiency of the different correlation experiments shown in Figure 1. One-dimensional traces of the various experiments obtained from the sample of the phage tail tubes (pb6 protein) are shown in Figure 3B. We found that the most efficient experiments were hCONH and hCANH, with approximately 25% of the signal amplitude rela-

tive to the most basic HN correlation experiment. For the least efficient experiment reported in this study, hCACONH, the sensitivity is approximately 4% of the HN experiment. These values are comparable to previously reported ones.^[16,22,29]

2.2. Application of the Approach to the 14-mer ClpP Assembly

We applied the two sets of experiments first to the 21 kDa protein ClpP. Figure 4A shows an example of a sequential walk through part of the ClpP sequence. As expected, the 4D spectra are particularly useful in cases in which the 3D "sub-spectra" leave several options for connecting the HN frequency pair to the carbon frequency. For example, two or three different intra- $\text{C}\alpha$ frequencies are possible for HN of G109 and inter- $\text{C}\alpha$ frequencies of E107 (Figure 4A); these ambiguities are resolved in the 4D spectra.

Figure 4B shows example sequential walks using the $^1\text{H}^{\text{N}}$ – $^1\text{H}^{\text{N}}$ correlation experiments. In many instances, we found both

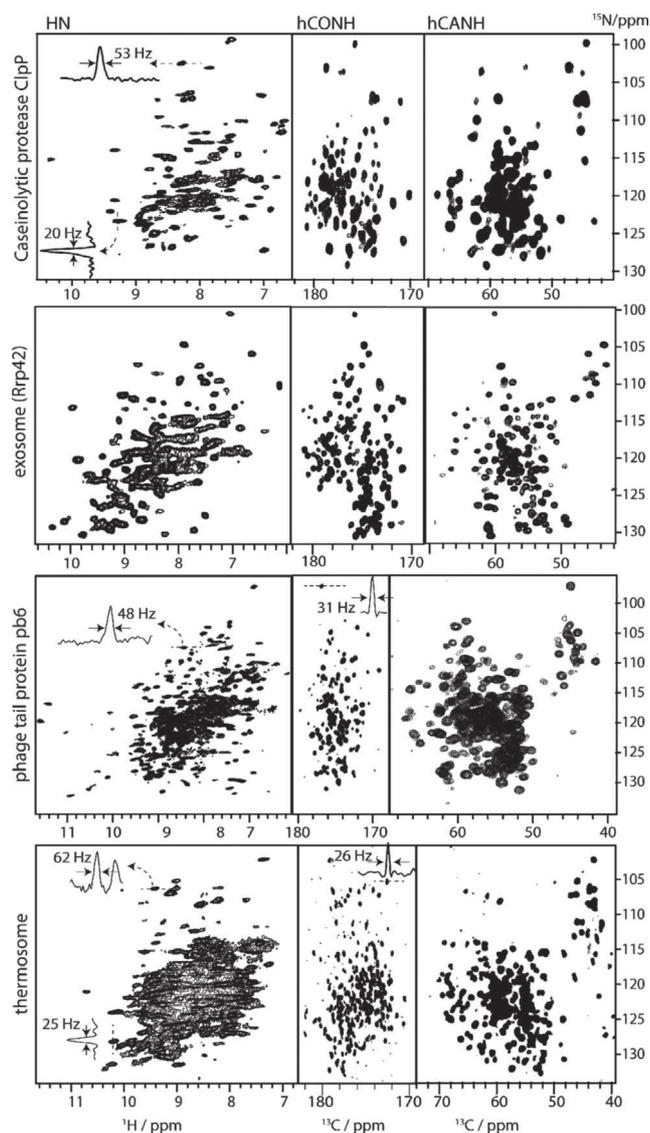


Figure 2. Proton-detected correlation spectra of large protein assemblies at 55 kHz MAS (14.1 T magnetic field strength). The identities of the proteins are indicated on the left of each panel. For the hCONH and hCANH spectra, a projection along the ^1H dimension was generated by superimposing all ^{15}N - ^{13}C planes over the range of ^1H frequencies in the amide region. Representative 1D traces are reported at positions indicated by dashed lines. Note that in the ^{13}C dimensions, the line width is primarily limited by sampling of the indirect dimension, not by inherent line widths.

the sequential cross-peaks ($\text{H}_i\text{-N}_i\text{-H}_{i-1}$, $\text{H}_i\text{-N}_i\text{-H}_{i+1}$, respectively) as well as the diagonal peak ($\text{H}_i\text{-N}_i\text{-H}$). The former cross-peaks arise from initial cross-polarization from the nearest amide proton to the spatially closest proton ($\text{H}_i\text{-CO}_{i-1}$ and $\text{H}_i\text{-}\alpha_i$, for hCONH and hCANH, respectively), whereas the diagonal peak stems from $\text{H}_i\text{-CO}_i$ and $\text{H}_i\text{-}\alpha_{i-1}$ cross-polarization, respectively. The corresponding distances are 2.0–2.1 Å for the nearest H–C pair, and 2.5–3.1 Å for the next H–C. Accordingly, we find that the diagonal peaks have mostly lower signal intensity than the sequential connectivity peaks.

In order to assess the utility of the approach, we performed initial assignments, and achieved approximately 50% of the

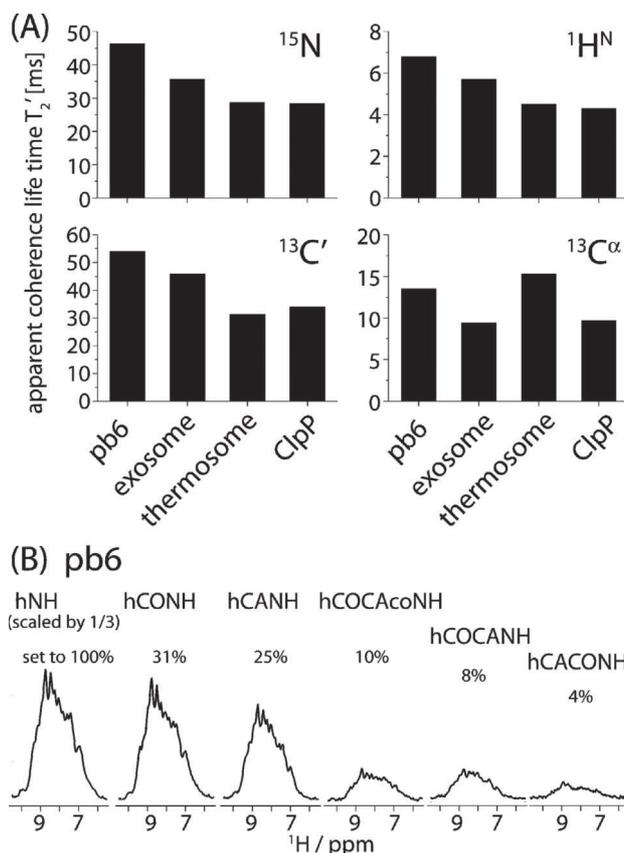


Figure 3. Coherence lifetimes and relative sensitivities of correlation experiments. A) Apparent coherence lifetimes, T_2' , of ^{15}N , ^1H , $^{13}\text{C}'$ and $^{13}\text{C}\alpha$ measured for the four proteins used in this study. For $^{13}\text{C}\alpha$, the apparent decay is heavily influenced by the large ($\approx 35\text{ Hz}^{[24]}$) scalar coupling to $^{13}\text{C}\beta$, which is difficult to decouple, and the reported value corresponds to the decay of the envelope from a fit which takes into account the by the scalar coupling (see Figure S2). B) Relative signal intensities of the different experiments applied to pb6, as obtained from 1D traces. Note that the first spectrum, hNH, has been scaled to one-third. The percentage signal heights are reported above each spectrum.

backbone assignment (Table S1 in the Supporting Information). These assignments are corroborated by ^{13}C -detected 3D experiments. Although a more complete assignment is deferred to a future publication, these results convinced us that the approaches proposed here are useful and applicable to even larger proteins.

2.3. Insight into the 50.4 kDa Bacteriophage T5 Tube Protein

We next investigated whether this approach could be successful for much larger proteins, and collected the set of four experiments on the phage tail tube protein composed of the 50.4 kDa pb6 protein. pb6 is, to our knowledge, larger than any protein that has been assigned so far by ssNMR. De novo assignment of such a large protein is challenging and time-consuming. The full assignment is not within the scope of the current report, which has a methodological focus. To assess the usefulness of the approach we focused on assigning the approximately 10 kDa-large Ig-like domain of pb6 within the

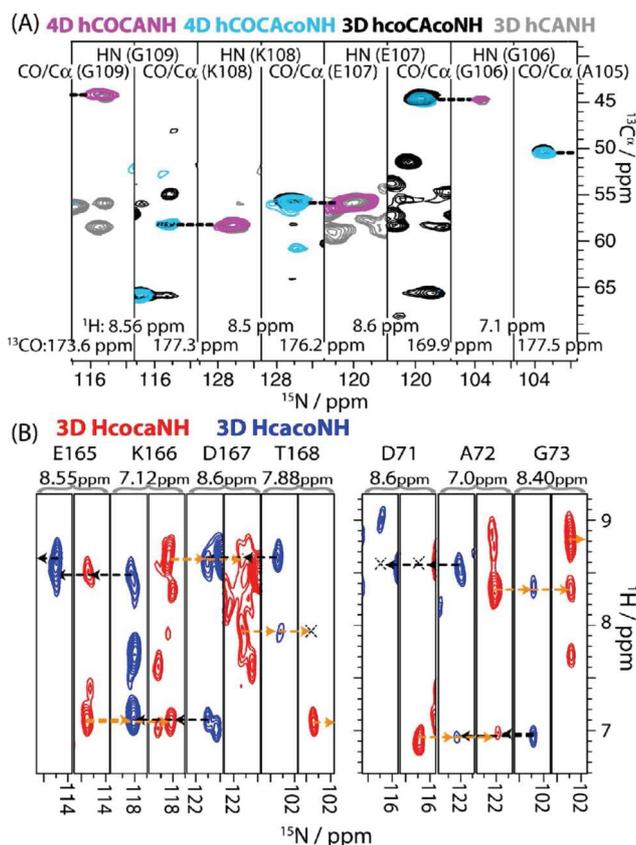


Figure 4. Assignment spectra of ClpP (14×21 kDa) for selected stretches along the protein sequence. A) In the 4D assignment strategy, sequential connectivities between H–N sites are established through two matching frequencies, C α (vertical axis) and CO (frequencies are given in the corresponding panels). B) Assignment using pairs of sequential H–H connectivities. Crosses denote the position of diagonal peaks in cases in which only the sequential cross-peak is observed. Orange and black arrows indicate residues $i+1$ and $i-1$, respectively.

full-length pb6 assembly (i.e., the spectra were collected on a sample of the full-length phage tail tubes, but we only focused on the assignment of this domain). This approach is also interesting because the Ig-like domain can be produced in isolation and studied in solution. The assignment of this domain in solu-

tion is reported in Figure S1. We used these solution-state assignments to facilitate assignment of the Ig domain in the context of the full-length 50.4 kDa protein assembled to tubes.

Figure 5A shows an example sequential walk through a 15-residue stretch using the pair of 4D H–N–C–C experiments, and Figure 5B shows sequential amide-to-amide connectivities for selected residues of this stretch. Remarkably, despite the size of this protein, the detection sensitivity and resolution in these experiments allows straightforward sequential linking. During the assignment process, we also used 3D hcaCBaNH and NCACB experiments, which proved useful for identifying the amino acid type. Using this approach, we could assign 92% of the backbone (H^N, N, C α and CO sites) of the Ig-like domain (i.e., residues 375–462 of the full-length pb6 protein). Further assignment of the rest of the protein is currently ongoing. The assignment of the 10 kDa Ig-like domain establishes that the proposed approach has the capability to provide assignments for proteins that significantly exceed those routinely studied by ssNMR.

With these chemical shift assignments at hand, we investigated if and how the conformation of this domain in the tubes formed by the full-length protein differs from the truncated domain in solution. Figure 6 shows the differences in H^N, N, C α and CO chemical shifts in solution (Ig-like domain alone) and in the solid state (full-length pb6 protein). A first observation is that the chemical shifts are, overall, very similar, establishing that this domain, as part of the assembled tube, retains its conformation. The regions with significant differences in chemical shifts are plotted onto a structural model of the pb6 tail tubes^[30] in Figure 6. These parts are located in close proximity to the remainder of the protein, that is, where the Ig-like domain contacts the core of the tail tube, as one expects from the structural model. We are currently extending the assignments of pb6 to the remaining 374 residues, in order to characterize the structure, interactions and dynamics of these assemblies in detail. The approaches presented herein will be a key element in pursuing these studies.

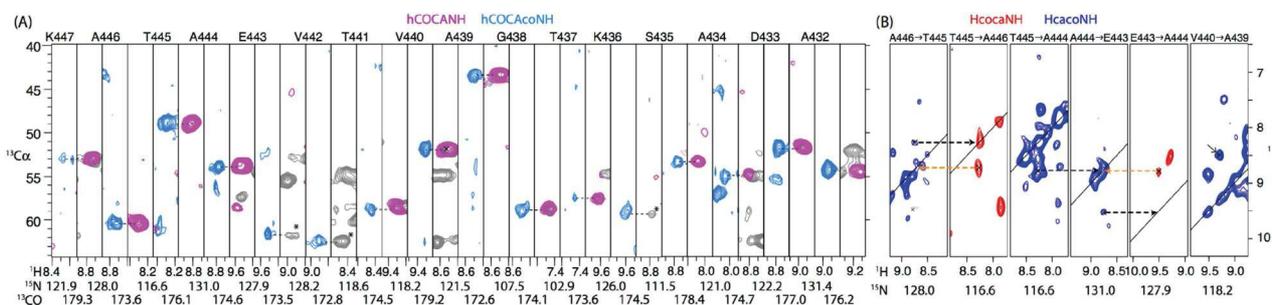


Figure 5. Assignment of the 50.4 kDa phage tail tube protein pb6, assembled into polymeric tubular structures, exemplified with a stretch comprising residues K447 to A432 of this 464-residue protein. A) Assignment strategy based on 4D hCOCANH (pink) and 4D hCOCaNH (blue) experiments. Also shown is an hCANH spectrum (gray), which is useful in cases in which the cross-peak in the hCOCANH experiment is not detected (three instances are shown here, marked with asterisks). Residue linking is obtained by matching the C α and CO frequencies. B) Complementary assignment information, exemplified with five residues from the same stretch of residues shown in panel A.

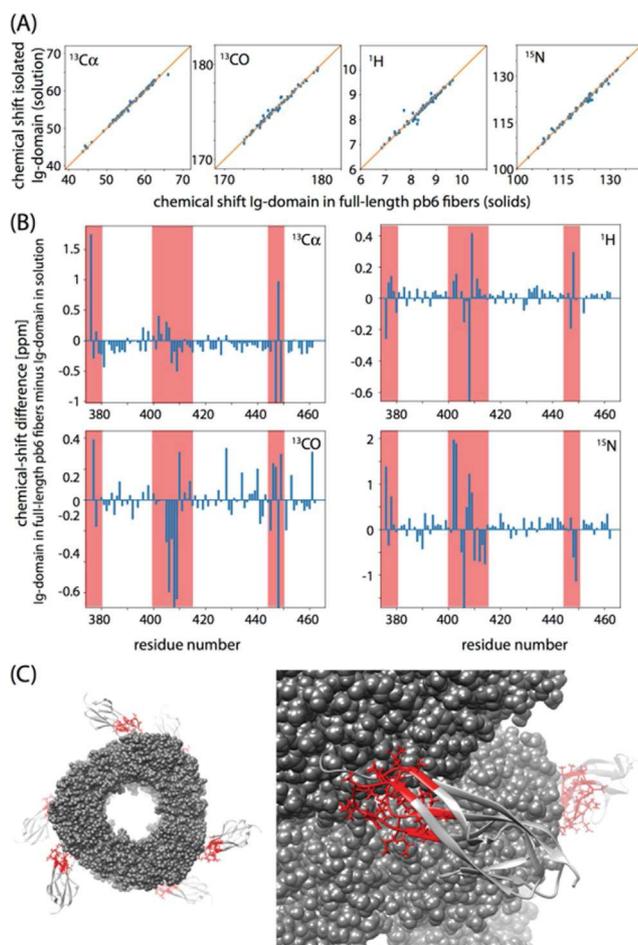


Figure 6. Insight into bacteriophage T5 tail tube protein pb6 by comparison of solution-state and ssNMR chemical shifts. Shown are the chemical shifts of the Ig-like domain (residues 375–364), studied either as a truncated construct in solution, or in the context of the tubes formed by the full-length pb6 protein (solid state). A) Correlation of chemical shifts in these two states. B) Residue-wise chemical-shift differences $\Delta\delta$. Regions with significant $\Delta\delta$ values are highlighted in red, and plotted onto the structural model of the full-length pb6 tubes^[30] (panel C).

3. Conclusions

In summary, we have demonstrated high-dimensional approaches for backbone assignments of proteins by MAS ssNMR. The approaches exploit the fact that ^1H , ^{13}C and ^{15}N line widths are favorable under the experimental conditions chosen here (deuteration and ≈ 40 – 60 kHz MAS). It is expected that this strategy can be extended in a straightforward manner also to other experimental conditions that result in narrow line widths. In particular, if even faster MAS is used (> 100 kHz), deuteration may be abandoned whereas line widths and transfer efficiencies remain comparable. The inherent drawbacks of deuteration, namely that one needs to re-protonate the amide sites, which might not be possible for some proteins, and that for many proteins the expression levels are lower in D_2O -based media, would be circumvented with such an approach. Experiments such as those presented herein will contribute to extending the applicability of MAS ssNMR to proteins of increasing size, complexity and biological interest.

Acknowledgements

This work was financially supported by the European Research Council (ERC-StG-311318-ProtDyn2Function and ERC-StG-2010-260887), INSTRUCT (“Innovative EM/NMR approach for the characterization of the drug target ClpP APPID: 301”), and used the platforms of the Grenoble Instruct Center (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). We thank Guillaume Mas and Elodie Crublet for help with the preparation of thermosome samples.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: NMR methods • NMR pulse sequences • phage proteins • protein assemblies • protein quality control machinery

- [1] L. B. Andreas, K. Jaudzems, J. Stanek, D. Lalli, A. Bertarello, T. Le Marchand, D. Cala-De Paepe, S. Kotelovica, I. Akopjana, B. Knott, S. Wegner, F. Engelke, A. Lesage, L. Emsley, K. Tars, T. Herrmann, G. Pintacuda, *Proc. Natl. Acad. Sci.* **2016**, *113*, 9187–9192.
- [2] A. Loquet, N. G. Sgourakis, R. Gupta, K. Giller, D. Riedel, C. Goosmann, C. Griesinger, M. Kolbe, D. Baker, S. Becker, A. Lange, *Nature* **2012**, *486*, 276–279.
- [3] F. Hu, W. Luo, M. Hong, *Science* **2010**, *330*, 505–508.
- [4] M. Tang, A. E. Nesbitt, L. J. Sperl, D. A. Berthold, C. D. Schwieters, R. B. Gennis, C. M. Rienstra, *J. Mol. Biol.* **2013**, *425*, 1670–1682.
- [5] H. Van Melckebeke, C. Wasmer, A. Lange, E. Ab, A. Loquet, A. Böckmann, B. H. Meier, *J. Am. Chem. Soc.* **2010**, *132*, 13765–75.
- [6] M. Renault, A. Cukkemane, M. Baldus, *Angew. Chem. Int. Ed.* **2010**, *49*, 8346–8357; *Angew. Chem.* **2010**, *122*, 8524–8535.
- [7] P. Schanda, S. Triboulet, C. Laguri, C. Bougault, I. Ayala, M. Callon, M. Arthur, J.-P. Simorre, *J. Am. Chem. Soc.* **2014**, *136*, 17852–17860.
- [8] T. Jacso, W. T. Franks, H. Rose, U. Fink, J. Broecker, S. Keller, H. Oschkinat, B. Reif, *Angew. Chem. Int. Ed.* **2012**, *51*, 432–435; *Angew. Chem.* **2012**, *124*, 447–450.
- [9] B. Reif, *J. Magn. Reson.* **2012**, *216*, 1–12.
- [10] L. B. Andreas, T. Le Marchand, K. Jaudzems, G. Pintacuda, *J. Magn. Reson.* **2015**, *253*, 36–49.
- [11] D. H. Zhou, J. J. Shea, A. J. Nieuwkoop, W. T. Franks, B. J. Wylie, C. Mullen, D. Sandoz, C. M. Rienstra, *Angew. Chem. Int. Ed.* **2007**, *119*, 8532–8535; *Angew. Chem.* **2007**, *119*, 1022–1022.
- [12] V. Agarwal, S. Penzel, K. Szekely, R. Cadalbert, E. Testori, A. Oss, J. Past, A. Samoson, M. Ernst, A. Böckmann, B. H. Meier, *Angew. Chem. Int. Ed.* **2014**, *53*, 12253–12256; *Angew. Chem.* **2014**, *126*, 12450–12453.
- [13] J. Medeiros-Silva, D. Mance, M. Daniëls, S. Jekhmene, K. Houben, M. Baldus, M. Weingarth, *Angew. Chem. Int. Ed.* **2016**, *55*, 13606–13610.
- [14] P. Schanda, M. Huber, R. Verel, M. Ernst, B. H. Meier, *Angew. Chem. Int. Ed.* **2009**, *48*, 9322–9325; *Angew. Chem.* **2009**, *121*, 9486–9489.
- [15] P. Fricke, V. Chevelkov, M. Zinke, K. Giller, S. Becker, A. Lange, *Nat. Protoc.* **2017**, *12*, 764–782.
- [16] E. Barbet-Massin, A. J. Pell, J. S. Retel, L. B. Andreas, K. Jaudzems, W. T. Franks, A. J. Nieuwkoop, M. Hiller, V. Higman, P. Guerry, A. Bertarello, M. J. Knight, M. Felletti, T. Le Marchand, S. Kotelovica, I. Akopjana, K. Tars, M. Stoppini, V. Bellotti, M. Bolognesi, S. Ricagno, J. J. Chou, R. G. Griffin, H. Oschkinat, A. Lesage, L. Emsley, T. Herrmann, G. Pintacuda, *J. Am. Chem. Soc.* **2014**, *136*, 12489–12497.
- [17] R. Linser, U. Fink, B. Reif, *J. Magn. Reson.* **2008**, *193*, 89–93.
- [18] L. B. Andreas, J. Stanek, T. Le Marchand, A. Bertarello, D. Cala-De Paepe, D. Lalli, M. Krejčíková, C. Doyen, C. Öster, B. Knott, S. Wegner, F. Engelke, I. C. Felli, R. Pierattelli, N. E. Dixon, L. Emsley, T. Herrmann, G. Pintacuda, *J. Biomol. NMR* **2015**, *62*, 253–261.

- [19] S. Xiang, K. Grohe, P. Rovó, S. K. Vasa, K. Giller, S. Becker, R. Linser, *J. Biomol. NMR* **2015**, *62*, 303–311.
- [20] C. Shi, H. K. Fasshuber, V. Chevelkov, S. Xiang, B. Habenstein, S. K. Vasa, S. Becker, A. Lange, *J. Biomol. NMR* **2014**, *59*, 15–22.
- [21] S. Xiang, J. Biernat, E. Mandelkow, S. Becker, R. Linser, *Chem. Commun.* **2016**, *52*, 4002–4005.
- [22] E. Barbet-Massin, A. J. Pell, K. Jaudzems, W. T. Franks, J. S. Retel, S. Kotelovica, I. Akopjana, K. Tars, L. Emsley, H. Oschkinat, A. Lesage, G. Pintacuda, *J. Biomol. NMR* **2013**, *56*, 379–386.
- [23] S. Xiang, V. Chevelkov, S. Becker, A. Lange, *J. Biomol. NMR* **2014**, *60*, 85–90.
- [24] M. Sattler, J. Schleucher, C. Griesinger, *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 93–158.
- [25] H. Geen, R. Freeman, *J. Magn. Reson.* **1991**, *93*, 93–141.
- [26] E. Kupce, J. Boyd, I. D. Campbell, *J. Magn. Reson. Ser. B* **1995**, *106*, 300–303.
- [27] M. J. Knight, A. L. Webber, A. J. Pell, P. Guerry, E. Barbet-Massin, I. Bertini, I. C. Felli, L. Gonnelli, R. Pierattelli, L. Emsley, A. Lesage, T. Herrmann, G. Pintacuda, *Angew. Chem. Int. Ed.* **2011**, *50*, 11697–11701; *Angew. Chem.* **2011**, *123*, 11901–11905.
- [28] R. Linser, U. Fink, B. Reif, *J. Biomol. NMR* **2010**, *47*, 1–6.
- [29] S. Penzel, A. A. Smith, V. Agarwal, A. Hunkeler, M.-L. Org, A. Samoson, A. Böckmann, M. Ernst, B. H. Meier, *J. Biomol. NMR* **2015**, *63*, 165–186.
- [30] C.-A. Arnaud, G. Effantin, C. Vivès, S. Engilberge, M. Bacia, P. Boulanger, E. Girard, G. Schoehn, C. Breyton, **2017**, *submitted*.

Manuscript received: May 24, 2017

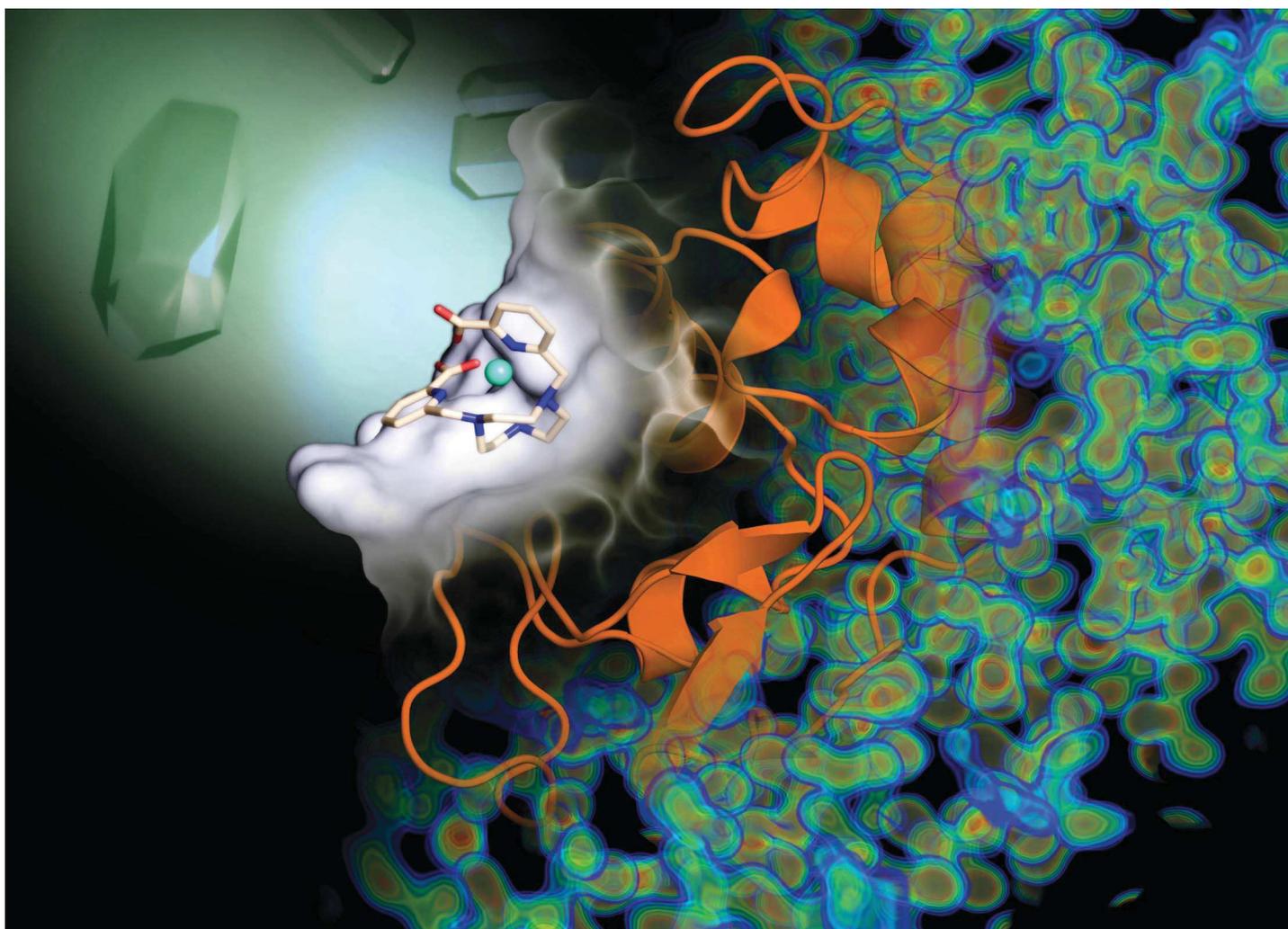
Revised manuscript received: August 8, 2017

Accepted manuscript online: August 9, 2017

Version of record online: September 5, 2017

Annexe B

Article : Crystallophore: a versatile lanthanide complex for protein crystallography combining nucleating effects, phasing properties, and luminescence



Showcasing collaborative research from Dr Eric Girard, IBS, Univ. Grenoble Alpes-CEA-CNRS, France and Dr Olivier Maury, Chemistry Laboratory of ENS Lyon, Univ. Lyon-CNRS-ENS-Univ. Lyon 1, France.

Crystallophore: a versatile lanthanide complex for protein crystallography combining nucleating effects, phasing properties, and luminescence

Introducing the first example of "crystallophore" (Xo4), a lanthanide complex family that tackle both bottlenecks of biocrystallography: to grow diffracting crystals and to solve the phase problem. This terbium complex combines unprecedented nucleating properties with the exceptional phasing power of the Tb(III) heavy atom, facilitating crystal structure determination. Furthermore, easy detection of the protein/Tb-Xo4 co-crystals is enabled thanks to the green luminescence of the complex. We demonstrate the potential of this additive for the crystallisation and structure determination of eight proteins, two of whose structures were unknown.

As featured in:



See Olivier Maury, Eric Girard et al., *Chem. Sci.*, 2017, 8, 5909.



rsc.li/chemical-science

Registered charity number: 207890

Cite this: *Chem. Sci.*, 2017, 8, 5909

Crystallophore: a versatile lanthanide complex for protein crystallography combining nucleating effects, phasing properties, and luminescence†

Sylvain Engilberge,^{‡a} François Riobé,^{‡b} Sebastiano Di Pietro,^{‡b} Louise Lassalle,^a Nicolas Coquelle,^a Charles-Adrien Arnaud,^a Delphine Pitrat,^b Jean-Christophe Mulatier,^b Dominique Madern,^a Cécile Breyton,^a Olivier Maury^{‡*b} and Eric Girard^{‡*a}

Macromolecular crystallography suffers from two major issues: getting well-diffracting crystals and solving the phase problem inherent to large macromolecules. Here, we describe the first example of a lanthanide complex family named "crystallophore" (Xo4), which contributes to tackling both bottlenecks. This terbium complex, Tb–Xo4, is an appealing agent for biocrystallography, combining the exceptional phasing power of the Tb(III) heavy atom with powerful nucleating properties, providing ready-to-use crystals for structure determination. Furthermore, protein/Tb–Xo4 co-crystals can be easily detected and discriminated from other crystalline by-products using luminescence. We demonstrate the potential of this additive for the crystallisation and structure determination of eight proteins, two of whose structures were unknown.

Received 17th February 2017
Accepted 2nd June 2017

DOI: 10.1039/c7sc00758b

rsc.li/chemical-science

Introduction

X-ray crystallography remains the method of choice to obtain high-precision structural information about biological macromolecules. This is illustrated by the fact that almost 90% of protein structures deposited in the Protein Data Bank were elucidated using this method. However, production of high-quality single crystals and phase determination are the two main issues inherent to X-ray crystallography that severely slow down the resolution of new protein structures.¹ Phase determination has been partially solved with the increasing number of known structures and the development of sophisticated algorithms and programs for molecular replacement. When the lack of homologous structures hinders structure determination, *de novo* phasing using anomalous-based methods – Single-wavelength Anomalous Dispersion (SAD) and Multi-wavelength Anomalous Dispersion (MAD)^{2,3} – is used thanks to the development of tuneable high-energy synchrotron radiation sources. These methods rely on the presence of well-ordered heavy atoms within the crystal, and this can be achieved by the conventional seleno-methionine labelling of the protein⁴ or by the recent exploitation of the intrinsic proteins'

sulphur content.^{5,6} Thus, crystallization represents the major issue in the whole structure determination process. Commercially available crystallization screens associated with crystallization robotics have optimised the trial and error process. However, crystallization of biological macromolecules remains very time-consuming, without any guarantee of success, and has triggered the research for various additives to facilitate the nucleation process. In this context, heterogeneous nucleants have been developed and these include minerals,^{7,8} natural materials such as horse and human hair,^{9,10} homologous protein thin films,¹¹ or carbon-nanotubes.^{12,13} The best results have been obtained with porous nucleating agents such as porous silicon,¹⁴ nanoporous gold,¹⁵ microporous zeolite,¹⁶ bioglass,¹⁷ and, more recently, molecularly imprinted polymers.^{18–20}

Over the last 15 years, lanthanide complexes have been widely used to solve the phase problem thanks to the large anomalous contribution of lanthanide ions.²¹ Insertion of these heavy ions into protein crystals has been achieved by (i) substitution of Ca²⁺ in calcium-binding proteins,^{22,23} (ii) covalent grafting of a lanthanide binding tag,^{24,25} and (iii) supra-molecular co-crystallization of lanthanide complexes with protein. In this context, several complexes, including macrocyclic (DOTA, DO₃A, and HPDO₃A), polydentate (DTPA-BMA) or tris-dipicolinate lanthanide complexes (Fig. 1), have been evaluated on commercial proteins^{26–28} but have also allowed the structural determination of new proteins.^{29–35} Herein, we report a new terbium(III) complex that combines nucleating, phasing, and luminescence properties, allowing us to overcome the two major issues of X-ray macromolecular crystallography with

^aUniv. Grenoble Alpes, CEA, CNRS, IBS, F-38000 Grenoble, France. E-mail: eric.girard@ibs.fr

^bUniv Lyon, Ens de Lyon, CNRS UMR 5182, Université Claude Bernard Lyon 1, Laboratoire de Chimie, F-69342 Lyon, France. E-mail: olivier.maury@ens-lyon.fr

† Electronic supplementary information (ESI) available: General procedures, Tb–Xo4 synthesis, Fig. S1–S11 and Tables S1–S4. See DOI: 10.1039/c7sc00758b

‡ These authors contributed equally to the work.



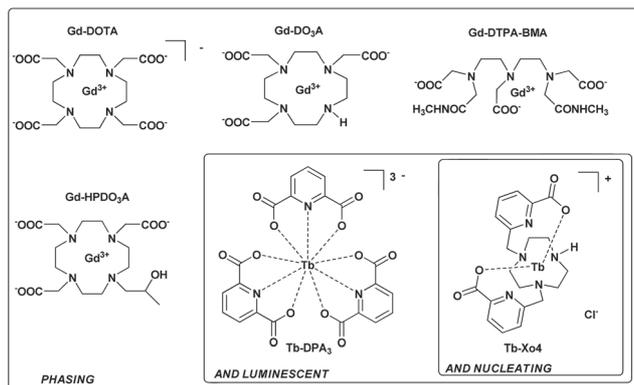


Fig. 1 Three generations of complexes designed for protein structure determination.

a single tool (patent pending PCT/FR2016/053539). We called this type of complex “crystallophore” (Xo4) and illustrated its properties for the crystallisation and structure determination of a series of eight proteins, including two proteins of unknown structure.

Results and discussion

Synthesis of Tb-Xo4

The complex was designed to be stable and highly soluble in aqueous media, and to be cationic with a non-saturated coordination sphere while remaining luminescent. It consists of a triazacyclononane macrocyclic scaffold bearing two picolinate arms, which ensure a good stability in the crystallization media and act as efficient antennae to sensitize terbium(III) luminescence (Fig. 1).³⁶ This macrocyclic ligand strongly chelates the Tb(III) ion with a coordination number of seven; consequently two coordination sites remain accessible for further direct interaction with proteins. In particular, this cationic complex was designed to preferentially target protein carboxylate amino-acids like aspartate or glutamate. The compound was prepared and used as a chloride salt of the complex. Details of the synthesis and characterisation are provided in the ESI (Fig. S1–S8†).

Nucleating properties of Tb-Xo4

The potentialities of this new complex were first evaluated with the benchmark hen egg white lysozyme (HEWL) protein. The standard conditions leading to the tetragonal crystal form, *i.e.* in the presence of NaCl as a precipitant, were used. First, the Tb-Xo4 optimal concentration was determined by evaluating different complex concentrations and by keeping the protein concentration constant at 20 mg mL⁻¹ (Fig. 2). Interestingly, the former generation of complexes are generally used at high concentrations (50–100 mM), whereas crystals of HEWL have been obtained with lower concentrations of Tb-Xo4, with an optimal concentration of about 10 mM.

In order to get a deeper insight into this markedly different behaviour, the crystallization diagram of HEWL was determined

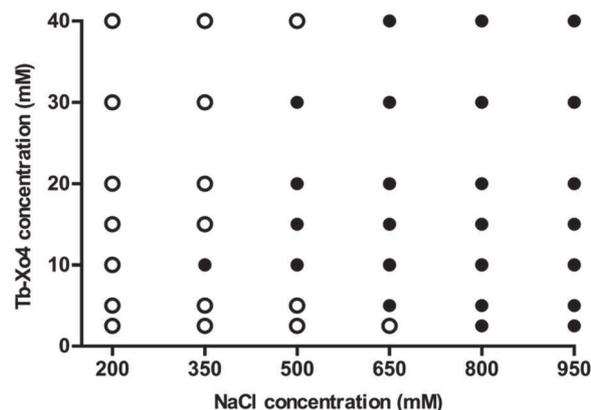


Fig. 2 Crystallization diagrams of HEWL after two days with different Tb-Xo4 concentrations (0 to 100 mM). Black dots represent conditions where crystals are observed. HEWL crystallizes up to a Tb-Xo4 concentration of 100 mM. However, for clarity, results in the presence of 50 mM of the complex (similar to 40 mM) and 100 mM (similar to 2.5 mM) are not displayed.

(Fig. 3). The diagrams in the absence and in the presence of Tb-Xo4 are dramatically different, clearly indicating that the complex has a strong influence on the crystallisation process. The presence of Tb-Xo4 speeds up the appearance of crystals and promotes crystallization in a wider concentration range of the precipitant, here NaCl. Importantly, exploitable protein crystals can be obtained at very low precipitant concentrations as well as low protein concentrations. For example, HEWL tetragonal crystals could be grown at a concentration of NaCl as low as 200 mM, which has never been observed in either rational crystallography studies³⁷ or when using other nucleants.¹⁸ Furthermore, the protein concentration giving crystals after 2 days was reduced from 15 mg mL⁻¹ under native conditions to 5 mg mL⁻¹ in the presence of the terbium complex, an unprecedented result in the case of HEWL.^{18,37} This result clearly indicates that Tb-Xo4 presents a strong nucleating effect in the case of HEWL crystallisation.

The general nucleating properties of Tb-Xo4 were then evaluated with several soluble proteins and under wide-ranging crystallization conditions. For this purpose, automated crystallization experiments were performed at the HTXlab (EMBL, Grenoble) with a set of 6 standard crystallization screens, corresponding to a total of 576 conditions. Six proteins of known structure, including three commercial ones (HEWL, thaumatin and proteinase K), as well as three in-house ones (protease1 from *P. horikoshii*, glyoxylate/hydroxypyruvate reductase (GRHPR) from *P. furiosus* and pb9 from the T5 phage tail) were tested (see ESI†). Two proteins of unknown structure were also tested: a halophilic tetrameric malate dehydrogenase (MDH) and a second protein of the T5 phage tail, pb6. Descriptions of the proteins and of the crystallization kits are provided in Tables S1 and S2† respectively. The nucleating properties of the complex were assessed using a comparative crystallization experiment. In the same crystallization plate, each protein was tested at a given concentration in the presence or in the absence



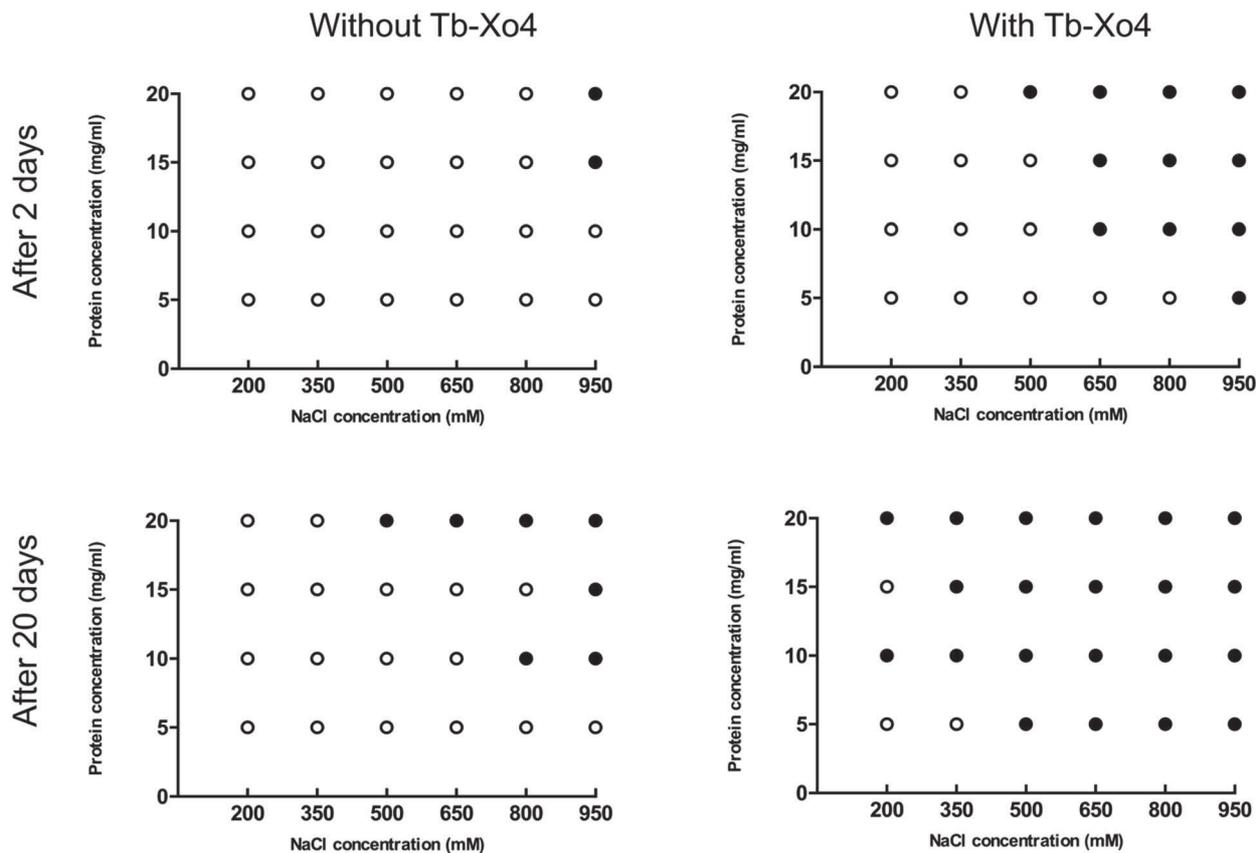


Fig. 3 Crystallization diagrams determined for HEWL. White dots represent clear drops and black dots represent conditions giving crystals (hits). The presented diagrams correspond to the situation after 2 days (top) or 20 days (bottom) where no more evolution is observed. The left diagrams refer to the native proteins and the right ones to crystallization in the presence of Tb-Xo4 (10 mM).

of Tb-Xo4. The samples for automated crystallization were prepared by direct mixing of the complex in each protein solution to reach an optimal Tb-Xo4 concentration of 10 mM (Fig. 2). The use of this molecular additive is thus perfectly compatible with automated crystallization robots, in contrast with heterogeneous nucleants. In the following discussion, we have called a crystallization condition leading to the appearance of a crystal a “hit”. The results after a 34 days observation period are displayed in Fig. 4. We chose this period thanks to the recent analysis of the Oxford Structural Genomics Consortium showing that 98% of all crystals that led to deposited structures appeared within 30 days.³⁸ However, it should also be noted that most of the hits are already present after 7 days (Fig. S9†). For all of the proteins tested, the presence of Tb-Xo4 in the crystallization drops provided additional hits compared to the native ones. This is also the case with the two proteins of unknown structure, MDH and pb6, with 2 and 6 unique hits out of 6 and 12 total hits, respectively. One should also note that, in the presence of the complex, some of the native hits are no longer preserved. Therefore, with the pragmatic goal of obtaining as many exploitable hits as possible, an optimal crystallization experiment would combine screening of the conditions of the protein both under native conditions and in the presence of Tb-Xo4.

To further confirm the nucleating effect of Tb-Xo4, the crystallization diagrams of the unknown protein pb6 were determined similarly to those of HEWL (Fig. 3). The best

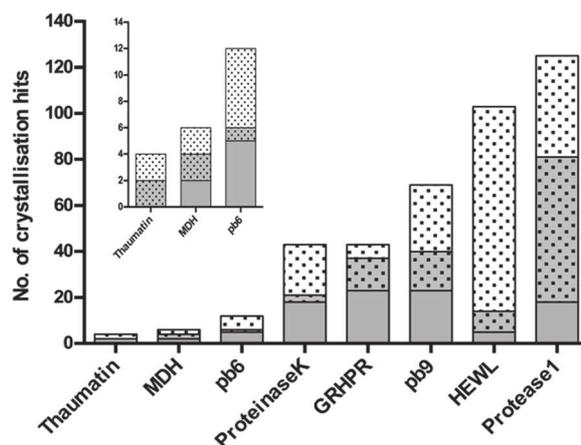


Fig. 4 Details of the results of the automated crystallization screening (576 conditions) performed on the 8 proteins tested. The number of crystallization hits is depicted in grey for the native protein and with dots for protein supplemented with 10 mM of Tb-Xo4. As a result, the shared conditions are represented in grey with dots. Observation after 34 days.

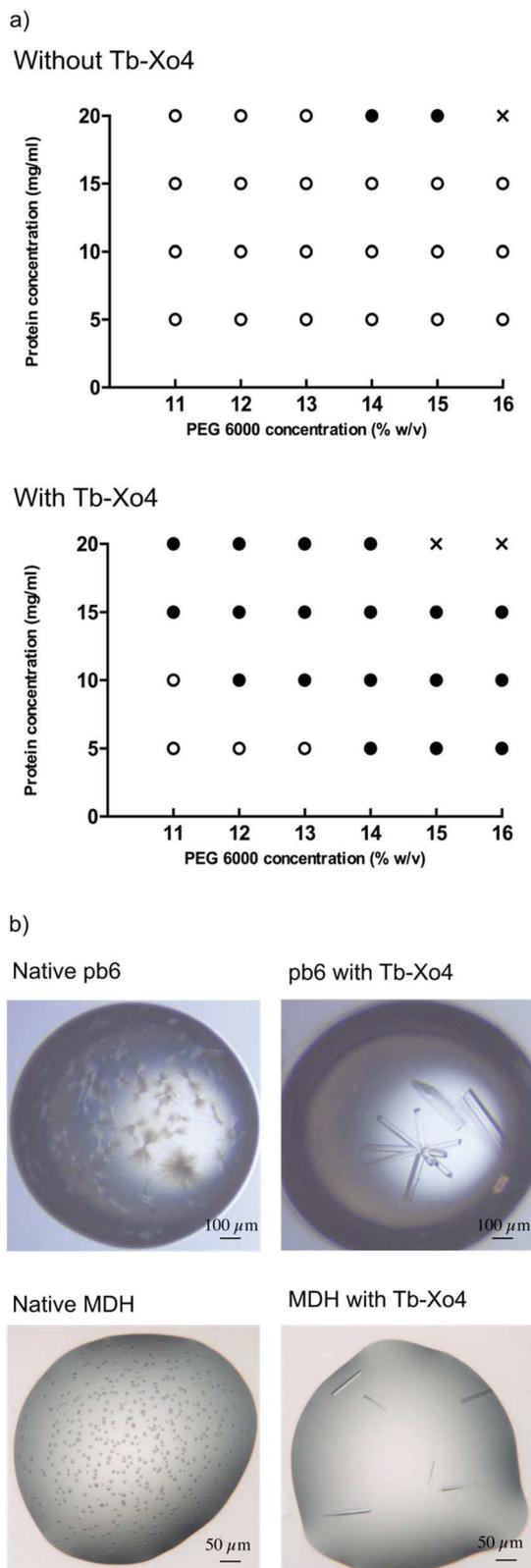


Fig. 5 (a) Crystallization diagrams determined for pb6 in the absence (top) and in the presence (bottom) of Tb-Xo4 (10 mM). Black dots represent conditions leading to crystals (hits). White dots represent clear drops. Crosses represent drops with protein precipitates. The presented diagrams correspond to the stable situation when no more evolution is observed. (b) Comparison of the crystals' quality obtained under strictly similar crystallization conditions for the two proteins of unknown structures pb6 and MDH, in the absence and presence of 10 mM of Tb-Xo4 as indicated.

amongst the 6 additional crystallisation conditions obtained after automated screening, *i.e.* in the presence of PEG 6000 as precipitant, was used (Fig. 5). As was already observed in the case of HEWL, crystals of pb6 appeared faster and in a broader range of precipitant concentrations in the presence of Tb-Xo4. Crystals were again obtained even at low protein concentrations. Consequently, such an additive opens the way to the crystallization of low concentration samples, which is a major benefit when considering precious protein samples. Beyond its nucleation properties, Tb-Xo4 also improves the quality of the crystals. This was illustrated for the crystals of both unknown proteins (pb6 and MDH) grown in the presence of the complex, which were clearly bigger and showed better morphologies than those of the native ones, making them ready-to-use for diffraction experiments (Fig. 5).

The diffraction of all the crystals produced in the presence of Tb-Xo4 was tested, an unprecedented assessment for a new nucleating agent. Two different approaches were used to evaluate the diffraction quality of the crystals. Firstly, complete anomalous data sets were collected for all 8 proteins tested, and the diffraction resolution ranged from 2.85 to 1.80 Å using the best crystallization conditions (Tables 1 and S3†). In particular, for commercial proteins, the resolution is in the highest range of what is classically deposited in the Protein Data Bank. Secondly, we directly evaluated the diffraction quality of all of the crystals obtained on a high-throughput screening assay using a statistical approach. This *in situ* experiment was performed in the case of HEWL and the diffraction quality of the 113 hits obtained in the presence of Tb-Xo4 was compared to that of 20 native hits (Fig. 6). The diffraction quality was assessed using the number of Bragg peaks per frame as a function of the resolution, using the recently developed NanoPeakCell software.³⁹ Compared to native crystals, the overall diffraction quality of the protein/Tb-Xo4 co-crystals, obtained under the same crystallization conditions, is slightly better in terms of the resolution and of number of detected Bragg peaks (Fig. 6). A similar analysis was performed on the crystals obtained only in the presence of the complex (*i.e.* unique hits, Fig. S11†). This led to a diffraction distribution equivalent to the one shown on Fig. 6. Consequently, the crystals grown only in the presence of the complex show identical diffraction quality compared to both native and Tb-Xo4 crystals obtained under shared crystallization conditions.

Overall, Tb-Xo4 exhibits a pronounced nucleating effect, increasing the number of crystallization conditions for all proteins tested, especially at lower protein and precipitant concentrations. Furthermore, it leads to the formation of larger crystals, ready-to-use for structure determination directly from the screening conditions.

Tb-Xo4 facilitates crystal detection

Under 280–370 nm excitation, Tb-Xo4 also presents the characteristic luminescence properties of terbium(III), *i.e.* a green emission arising from the $^5D_4 \rightarrow ^4F_j$ ($J = 0-6$) metal centered transitions (Fig. 7). This luminescence is very useful



Table 1 Experimental phasing statistics obtained for the eight proteins used. Different *de novo* phasing methods, using the strong anomalous signal of Tb–Xo4, were evaluated. All of the crystals were obtained under nucleating conditions (*i.e.* 10 mM Tb–Xo4). The quality of the phasing is assessed by the figure of merit (FOM) obtained after phasing with SHARP and phase improvement with Solomon (ESI†). The quality of the resulting experimental electron density maps is provided by the percentage of the protein model built automatically by Buccaneer

Protein	Soaking in 100 mM Tb–Xo4 (soaking time)	Final Tb–Xo4 concentration for phasing (mM)	Phasing method	Resolution (Å)	Space group	FOM after SHARP	% solvent content ^a	FOM after Solomon	Buccaneer ^b	U.A. content (molecules)
HEWL	No	10	SAD	1.8	$P4_32_12$	0.42	41.3	0.88	97.2	1
Thaumatin	No	10	SAD	1.8	$P4_12_12$	0.40	45.6	0.92	75.4	1
Proteinase K	No	10	SAD	1.8	$P4_32_12$	0.31	45.8	0.86	0.0	1
Proteinase K	Yes (120 s)	100	SAD	1.8	$P4_32_12$	0.26	44.5	0.85	86.4	1
Protease1	No	10	SAD	2.0	$P4_12_12$	0.21	61.9	0.95	62.7	3
Protease1	Yes (45 s)	100	SAD	2.0	$P4_12_12$	0.35	60.1	0.96	87.8	3
GRHPR	Yes (120 s)	100	SAD	2.5	$I4_1$	0.18	68.7	0.92	90.8	1
pb9	Yes (45 s)	100	SAD	2.0	$P2_1$	0.58	50.8	0.87	84.6	4
pb6	Yes (300 s)	100	MAD $2\lambda^c$	2.85	$P2_1$	0.29	65.5	0.79	64.7	4
MDH	No	10	SAD	2.35	$R3$	0.32	50.8	0.94	99.5	4

^a As evaluated by AutoSHARP. ^b The values indicated correspond to the completeness of chains (%) as provided by Buccaneer. ^c Two-wavelength MAD experiment.

for the detection of the derived crystal within the crystallization drops, in particular allowing an easier discrimination between the protein and salt crystals. Besides, this luminescence detection can be an attractive feature for robotized crystallization drop imaging systems. Furthermore, with the appropriate adaptation of the instrumentation, the luminescence properties of Tb–Xo4 could be used for crystal centring in synchrotron beamlines.^{40,41}

Tb–Xo4 as a powerful phasing agent

Throughout the whole periodic table, lanthanide ions possess the largest anomalous contribution making them particularly suitable candidates for anomalous-based phasing of diffraction data.^{3,42} In addition, lanthanides may also be highly efficient for phasing with isomorphous techniques due to their

high Z number. We evaluated the phasing properties of Tb–Xo4 by performing anomalous-based experiments for each aforementioned protein. The data were collected at the terbium L_{III} absorption edge by setting the incident beam energy to optimally exploit the large f'' contribution. Either SAD or MAD phasing protocols were used to solve the phase problem (Table 1). All of the structures of the evaluated proteins were determined, leading to a percentage of the automatically built model that ranged from 46 to 100% (Table 1). All of the crystals were initially obtained under nucleating conditions (*i.e.* co-crystallized with 10 mM of Tb–Xo4). The structure determination can be achieved either directly from crystals obtained under nucleating conditions (HEWL, protease1, and MDH) or after a rapid soaking of the 10 mM co-crystallized crystals in a concentrated solution of the complex (*i.e.* at 100 mM of Tb–Xo4 for proteinase K, protease1, GRHPR, pb9, and pb6). A soaking time ranging from 30 s to 5 min was sufficient to obtain a strong anomalous signal (Table 1) without any degradation of the crystals. In particular, in the case of proteinase K and protease1, the soaking highly facilitated the phase determination, leading to 86% and 88% complete models (Table 1), respectively. These results unambiguously showed that Tb–Xo4 is an excellent phasing agent, and is fully compatible with conventional phasing methods. The combination of its nucleating and phasing properties brings major advantages compared to the tedious and time-consuming seleno-methionine labelling method or the method of heavy-atom incorporation.

Application of the crystallophore for new protein structures determination

As a proof-of-concept, Tb–Xo4 was used to determine the structure of pb6 and MDH proteins as well as to resolve a difficult-to-phase protein, pb9.³⁴ These structures range from 800 to 1800 ordered residues and have resolution limits from

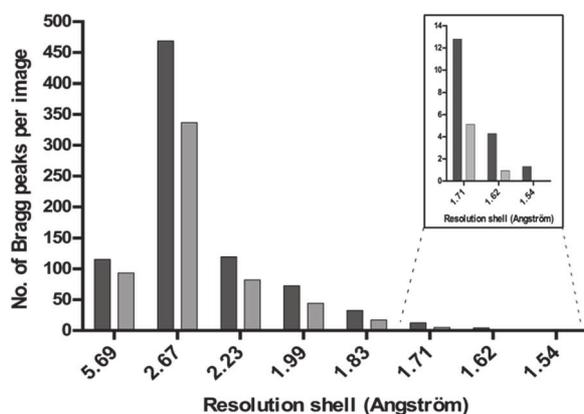


Fig. 6 The diffraction quality (measured as the number of Bragg peaks per image) as a function of the resolution of native (light grey) and Tb–Xo4 (dark grey) HEWL crystals obtained under shared crystallization conditions.



2.85 to 2.0 Å (Table 1). A full description of the pb6 and MDH structures and associated biology will be published elsewhere.

pb6 structure determination. The tail tube protein pb6 forms a stack of trimeric rings composing the bacteriophage T5 tail tube.⁴³ Crystallization screening in the presence of Tb-Xo4 (10 mM) provided 6 unique conditions out of a total of 12 (Fig. 4) and the manual optimization of one of them produced well diffracting crystals (Fig. 5). A two-wavelength MAD experiment was performed on a crystal soaked in a solution containing 100 mM of the complex (Table 1). The resulting experimental electron density map allows us to unambiguously elucidate secondary structure elements (Fig. 8a).

MDH structure determination. The structure determination of MDH, a tetrameric malate dehydrogenase, is a part of a larger project that aims at deciphering the origin of halophilic adaptation. The sequence of MDH was bioinformatically generated *via* ancestral sequence reconstruction based on a consolidated phylogenetic study. As expected, MDH behaves as a halophilic protein since it is only stable in the presence of 1.5 M KCl and such a high salt content is often detrimental to crystallization. As for pb6, the presence of Tb-Xo4 doubled the number of hits (Fig. 4). Crystals suitable for diffraction were obtained after manual optimisation of the best unique conditions (Fig. 5).

These crystals showed excellent diffraction up to 2.35 Å resolution (Table 1). SAD phasing was successful on the sole basis of crystals obtained under nucleating conditions (10 mM of the complex) and provided an almost complete model using an automatic procedure (Table 1) thanks to an excellent electron density map (Fig. 8b).

pb9 “resolving”. pb9 is the distal tail (Dit) protein of bacteriophage T5. Based on its structure, it is suggested to be a paradigm for a large family of Dit proteins of siphophages infecting mostly Gram-negative bacteria.³⁴ Despite a native data set at 1.89 Å resolution, the structure determination was a long and tedious process owing to the absence of a homologous model for molecular replacement, the sensitivity of crystals to radiation damage, the insufficient seleno-methionine labelling, and the difficulty of getting suitable derivatives for *de novo* phasing. Thanks to a lanthanide complex,²⁷ a derivative crystal was obtained by soaking, which provided initial phases at 3 Å resolution. Those were sufficient to build two poly-alanine helices, providing the key to determining the non-crystallographic operators. Then, cycles of partial model rebuilding and phase combination enabled us to obtain a poly-alanine model sufficient to initiate a molecular replacement using the 1.89 Å native data.

Even though the structure of pb9 is already known, the resolving of the structure using the crystallophore fully illustrates the benefits of this agent. As for pb6 and MDH, we optimised one of the most promising additional hits (Table S3†). Based on these conditions, we confirmed that the optimal Tb-Xo4 concentration to obtain an optimal nucleant effect is between 5 and 10 mM (Fig. S10†). A rapid soaking of a crystal in a concentrated solution of the complex provided a successfully derived crystal for SAD phasing, as illustrated by the high figure of merit and by the 85% residues built by the automatic model building (Table 1). For the four pb9 molecules present in the asymmetric unit, 15 Tb-Xo4 binding sites were determined with refined occupancies ranging from 0.1 to 0.7 (Table S4†). A superimposition of our monomer (PDB entry 5MF2) on the PDB entry 4JMQ monomer³⁴ led to a root mean square deviation of 0.55 Å, showing that the binding of Tb-Xo4 does not modify the overall structure of the protein.

As expected from the nucleating properties of Tb-Xo4, the analysis of the pb9 crystal packing showed the involvement of the complex at the interfaces between proteins (Fig. 8c). There is direct interaction between negatively charged residues, in particular glutamate, and the lanthanide atom at the main binding sites (Fig. 8c). Simultaneously, the crystallophore interacts through its hydrophobic macrocycle with the main-chain atoms of residues located on loops of symmetry-related proteins. In light of this binding mode, the nucleating properties of Tb-Xo4 could be explained by a reduction of surface entropy^{44,45} and/or by a slight modification of the surface topology,⁴⁶ favouring contact between molecules. These results open fascinating perspectives for the study of the supramolecular interactions between the crystallophore and the proteins at the beginning of the crystallisation process.

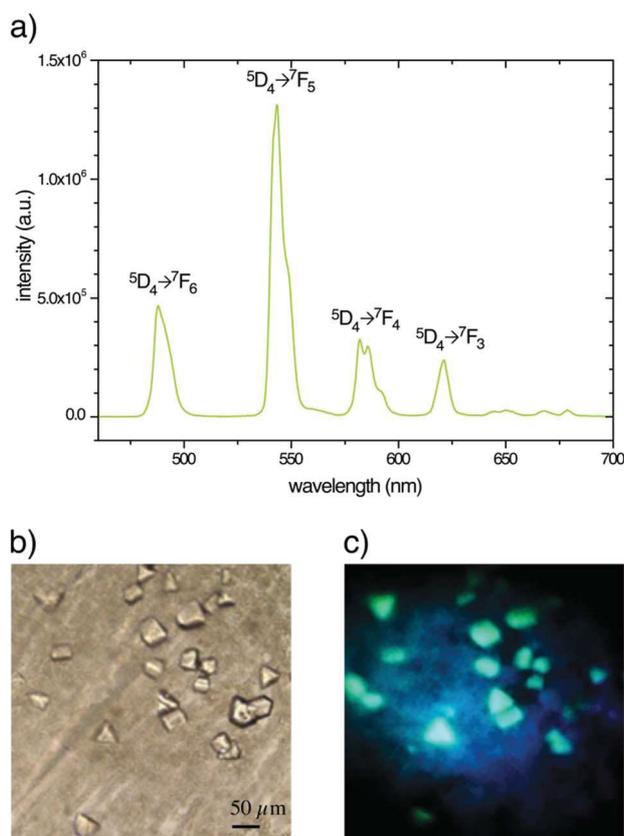


Fig. 7 (a) Luminescence spectrum of Tb-Xo4 in aqueous solution ($\lambda_{\text{ex}} = 290$ nm). MDH crystals obtained with 10 mM of the complex observed (b) in transmission mode and (c) under UV light irradiation (365 nm).



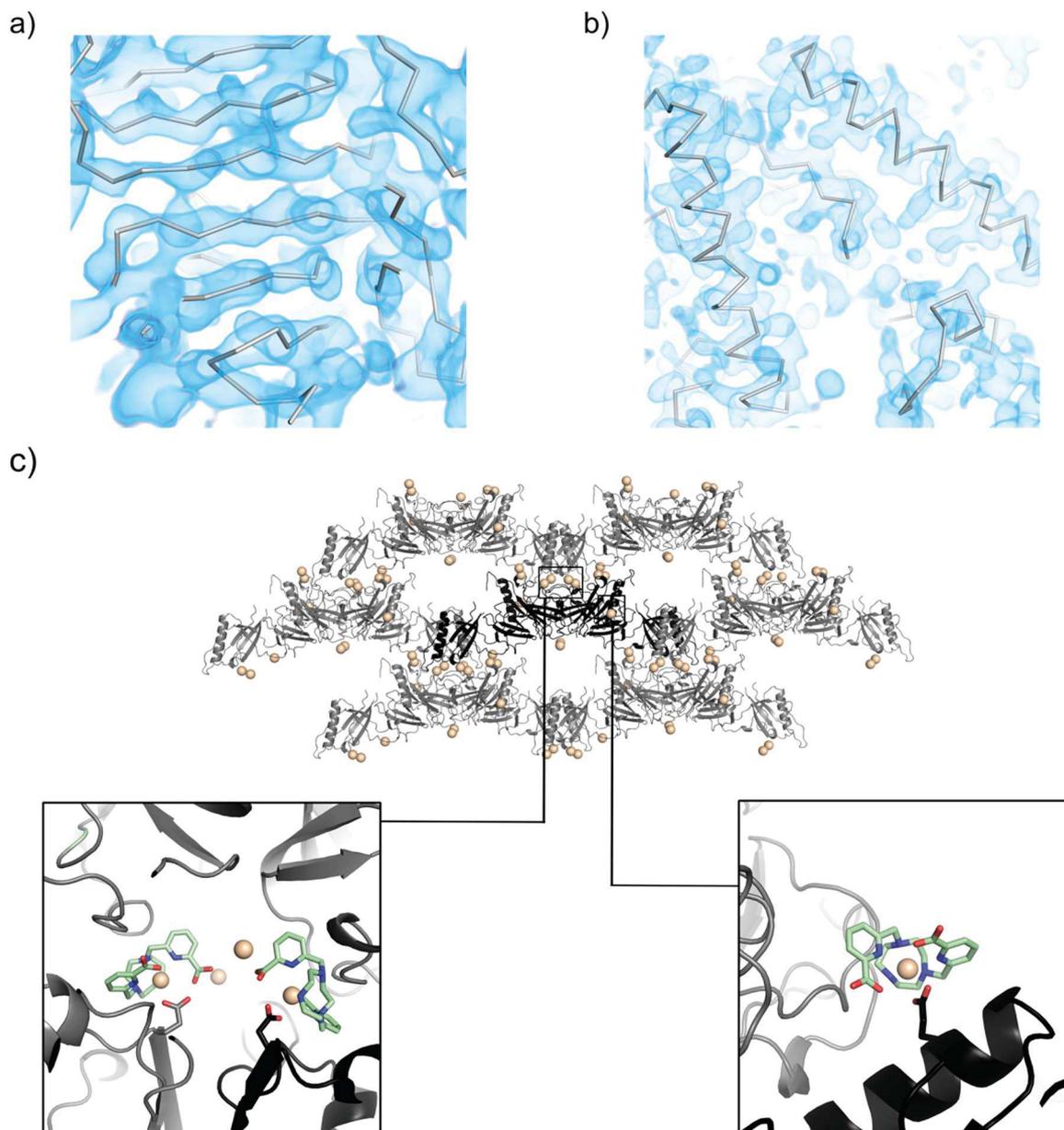


Fig. 8 Experimentally-phased electron density maps for (a) pb6 at 2.85 Å resolution and (b) MDH at 2.35 Å resolution. The maps are contoured at 1.0 sigma. The C_{α} trace corresponds to the automatically built model. (c) The packing of pb9 crystals (asymmetric unit content in black, symmetry-related molecules in grey). All of the Tb positions are shown with spheres. Details of the main Tb-Xo4 binding sites are shown in close-up views. pb9 PDB ID code 5MF2.

Conclusion

In this paper, we describe the first example of a new class of lanthanide complexes, named crystallophore, that can act as powerful auxiliaries for protein crystallography. Tb-Xo4 presents multiple advantages as follows. (i) It is soluble and stable under all crystallisation conditions and consequently fully compatible with high throughput screening. (ii) It acts as a strong nucleating agent, providing new crystallisation conditions compared to the native ones and generally leads to the formation of bigger derived crystals, ready-to-use for X-ray diffraction. (iii) Under UV irradiation, the terbium(III) complex presents a green emission,

enabling very easy detection of the derived crystal, and finally (iv) the crystallophore is a straightforward phasing agent thanks to the exceptional anomalous properties of the f-block elements facilitating crystal structure determination using SAD or MAD methods. Overall, the crystallophore is an all-in-one, easy-to-use tool that aims to overcome the two main bottlenecks of macromolecular crystallography: producing high quality single crystals and solving the phase problem. Future studies will be devoted to understanding the origin of such peculiar nucleating properties, with the aim of optimizing the design of this first compound and bringing new insight into the intricate issue of protein crystallization.



Acknowledgements

Authors acknowledge the ANR agency for funding (Ln²3 – ANR-13-BS07-0007-01) and for a grant to SE and SDP. LL and SDP are also grateful to the SATT Pulsalys (Lyon) for the maturation program (L-1382). We acknowledge the platforms of the Grenoble Instruct Center (ISBG; UMS 3518 CNRS-CEA-UGA-EMBL), supported by the French Infrastructure for Integrated Structural Biology Initiative FRISBI (ANR-10-INSB-05-02) and by the Grenoble Alliance for Integrated Structural Cell Biology GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB).

References

- 1 S. Khurshid, E. Saridakis, L. Govada and N. E. Chayen, *Nat. Protoc.*, 2014, **9**, 1621–1633.
- 2 G. L. Taylor, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 325–338.
- 3 W. A. Hendrickson, *Q. Rev. Biophys.*, 2014, **47**, 49–93.
- 4 S. Doublet, *Methods Enzymol.*, 1997, **276**, 523.
- 5 Q. Liu, T. Dahmane, Z. Zhang, Z. Assur, J. Brasch, L. Shapiro, F. Mancina and W. A. Hendrickson, *Science*, 2012, **336**, 1033–1037.
- 6 T. Weinert, V. Olieric, S. Waltersperger, E. Panepucci, L. Chen, H. Zhang, D. Zhou, J. Rose, A. Ebihara, S. Kuramitsu, D. Li, N. Howe, G. Schnapp, A. Pautsch, K. Bargsten, A. E. Prota, P. Surana, J. Kottur, D. T. Nair, F. Basilico, V. Cecatiello, S. Pasqualato, A. Boland, O. Weichenrieder, B.-C. Wang, M. O. Steinmetz, M. Caffrey and M. Wang, *Nat. Methods*, 2014, **12**, 131–133.
- 7 A. McPherson and P. Shlichta, *Science*, 1988, **239**, 385–387.
- 8 G. Falini, S. Fermani, G. Conforti and A. Ripamonti, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2002, **58**, 1649–1652.
- 9 A. D'Arcy, A. Mac Sweeney and A. Haber, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2003, **59**, 1343–1346.
- 10 D. G. Georgieva, M. E. Kuil, T. H. Oosterkamp, H. W. Zandbergen and J. P. Abrahams, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2007, **63**, 564–570.
- 11 E. Pechkova and C. Nicolini, *J. Cell. Biochem.*, 2002, **85**, 243–251.
- 12 L. Govada, H. S. Leese, E. Saridakis, S. Kassen, B. Chain, S. Khurshid, R. Menzel, S. Hu, M. S. P. Shaffer and N. E. Chayen, *Sci. Rep.*, 2016, **6**, 20053.
- 13 H. S. Leese, L. Govada, E. Saridakis, S. Khurshid, R. Menzel, T. Morishita, A. J. Clancy, E. R. White, N. E. Chayen and M. S. P. Shaffer, *Chem. Sci.*, 2016, **7**, 2916–2923.
- 14 N. E. Chayen, E. Saridakis, R. El-Bahar and Y. Nemirowsky, *J. Mol. Biol.*, 2001, **312**, 591–595.
- 15 F. Kertis, S. Khurshid, O. Okman, J. W. Kysar, L. Govada, N. Chayen and J. Erlebacher, *J. Mater. Chem.*, 2012, **22**, 21928–21934.
- 16 M. Sugahara, Y. Asada, Y. Morikawa, Y. Kageyama and N. Kunishima, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2008, **64**, 686–695.
- 17 N. E. Chayen, E. Saridakis and R. P. Sear, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 597–601.
- 18 E. Saridakis, S. Khurshid, L. Govada, Q. Phan, D. Hawkins, G. V. Crichlow, E. Lolis, S. M. Reddy and N. E. Chayen, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 11081–11086.
- 19 Y. Xing, Y. Hu, L. Jiang, Z. Gao, Z. Chen, Z. Chen and X. Ren, *Cryst. Growth Des.*, 2015, **15**, 4932–4937.
- 20 S. Khurshid, L. Govada, H. F. El-Sharif, S. M. Reddy and N. E. Chayen, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2015, **71**, 534–540.
- 21 K. Djinovic-Carugo and O. Carugo, *J. Inorg. Biochem.*, 2014, **143**, 1–8.
- 22 R. Kahn, R. Fourme, R. Bosshard, M. Chiadmi, J. Risler, O. Dideberg and J. Wery, *FEBS Lett.*, 1985, **179**, 133–137.
- 23 W. I. Weis, R. Kahn, R. Fourme, K. Drickamer and W. A. Hendrickson, *Science*, 1991, **254**, 1608–1615.
- 24 N. R. Silvaggi, L. J. Martin, H. Schwalbe, B. Imperiali and K. N. Allen, *J. Am. Chem. Soc.*, 2007, **129**, 7114–7120.
- 25 K. N. Allen and B. Imperiali, *Curr. Opin. Chem. Biol.*, 2010, **14**, 247–254.
- 26 E. Girard, M. Stelter, P. L. Anelli, J. Vicat and R. Kahn, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2003, **59**, 118–126.
- 27 G. Pompidor, A. D'Aléo, J. Vicat, L. Toupet, N. Giraud, R. Kahn and O. Maury, *Angew. Chem., Int. Ed.*, 2008, **47**, 3388–3391.
- 28 M. Stelter, R. Molina, S. Jeudy, R. Kahn, C. Abergel and J. A. Hermoso, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2014, **70**, 1506–1516.
- 29 B. N. Chaudhuri, M. R. Sawaya, C. Y. Kim, G. S. Waldo, M. S. Park, T. C. Terwilliger and T. O. Yeates, *Structure*, 2003, **11**, 753–764.
- 30 S. de Bono, L. Riechmann, E. Girard, R. L. Williams and G. Winter, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 1396–1401.
- 31 P. Arnoux, T. Morosinotto, G. Saga, R. Bassi and D. Pignol, *Plant Cell*, 2009, **21**, 2036–2044.
- 32 R. Molina, M. Stelter, R. Kahn and J. Hermoso, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2009, **65**, 823–831.
- 33 L. Larivière, C. Plaschka, M. Seizl, L. Wenzek, F. Kurth and P. Cramer, *Nature*, 2012, **492**, 448–451.
- 34 A. Flayhan, F. M. D. Vellieux, R. Lurz, O. Maury, C. Contreras-Martel, E. Girard, P. Boulanger and C. Breyton, *J. Virol.*, 2014, **88**, 820–828.
- 35 C. Eichmann, C. Tzitzilonis, T. Nakamura, W. Kwiatkowski, I. Maslennikov, S. Choe, S. A. Lipton and R. Riek, *J. Mol. Biol.*, 2016, **428**, 3737–3751.
- 36 G. Nocton, A. Nonat and C. Gateau, *Helv. Chim. Acta*, 2009, **92**, 2257–2273.
- 37 P. Retailleau, A. Ducruix and M. Riès-Kautt, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2002, **58**, 1576–1581.
- 38 J. T. Ng, C. Dekker, P. Reardon and F. von Delft, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2016, **72**, 224–235.
- 39 N. Coquelle, A. S. Brewster, U. Kapp, A. Shilova, B. Weinhausen, M. Burghammer and J.-P. Colletier, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2015, **71**, 1184–1196.
- 40 X. Vernede, B. Lavault, J. Ohana, D. Nurizzo, J. Joly, L. Jacquamet, F. Felisaz, F. Cipriani and D. Bourgeois, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2006, **62**, 253–261.



- 41 L. M. G. Chavas, Y. Yamada, M. Hiraki, N. Igarashi, N. Matsugaki and S. Wakatsuki, *J. Synchrotron Radiat.*, 2011, **18**, 11–15.
- 42 E. Girard, M. Stelter, J. Vicat and R. Kahn, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2003, **59**, 1914–1922.
- 43 Y. Zivanovic, F. Confalonieri, L. Ponchon, R. Lurz, M. Chami, A. Flayhan, M. Renouard, A. Huet, P. Decottignies, A. R. Davidson, C. Breyton and P. Boulanger, *J. Virol.*, 2014, **88**, 1162–1174.
- 44 Z. S. Derewenda, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 604–615.
- 45 L. Goldschmidt, D. Eisenberg and Z. S. Derewenda, *Methods Mol. Biol.*, 2014, **1140**, 201–209.
- 46 Y. D. Devedjiev, *Acta Crystallogr., Sect. F: Struct. Biol. Commun.*, 2015, **71**, 157–162.



Annexe C

Figures complémentaires

C.1 Purification de queues de phages sur colonne échangeuse d'ion : exploration des quantités relatives des protéines du complexe.

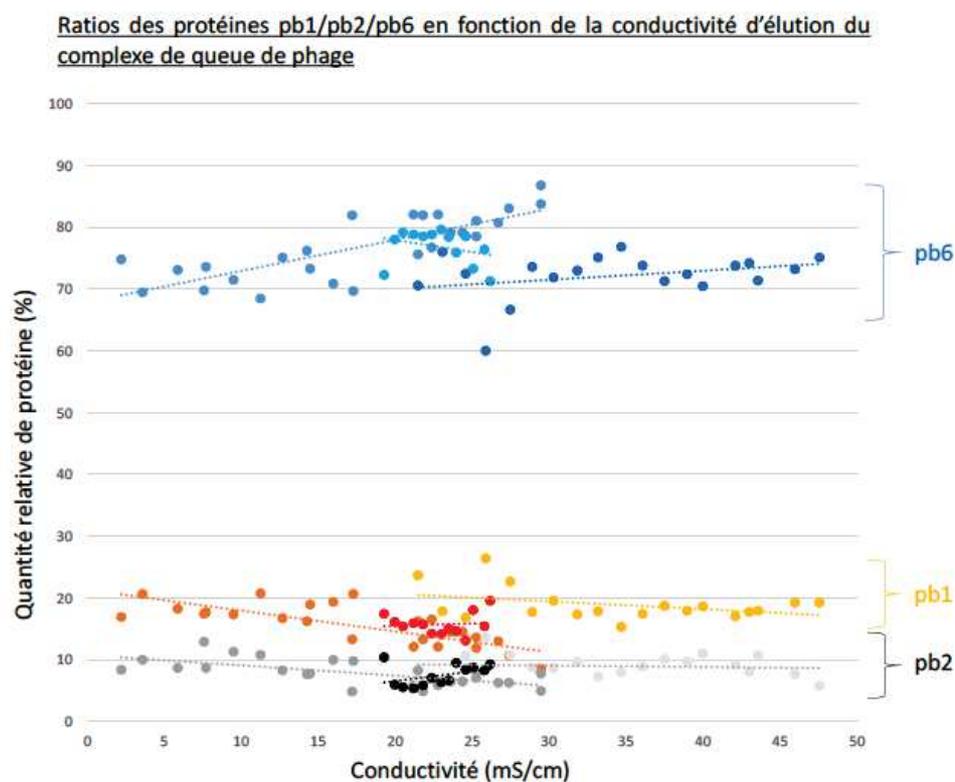
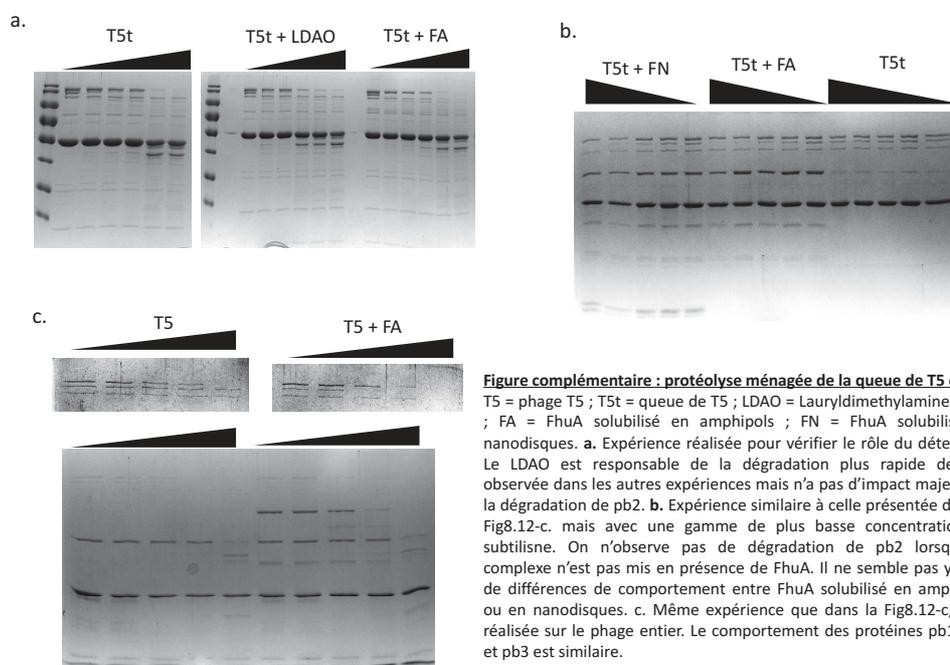


Figure annexe 1 : les quantités relatives de pb1, pb2 et pb6 ont été évaluées par densitométrie de bandes sur gel pour de nombreuses fractions d'éluions de colonne échangeuse d'ions contenant des queues de phages. Ce travail a été réalisé sur les fractions de 3 purifications différentes (3 nuances de couleurs pour chaque protéine).

C.2 Prediction de structure et alignement de la séquence de p140

Le serveur Jpred effectue une prédiction de structure secondaire et effectue un PSI-BLAST de la séquence. Les séquences identifiées par le PSI-BLAST sont celles de protéines mineures de queues des phages différents Siphoviridae (pvp-1, EPS7, my1) et une protéine de fibre du Siphophage PhiR201.

C.3 Protéolyse ménagée de queues de T5 : compléments



Bibliographie

- [Abergel 13] Chantal Abergel. *Molecular replacement : tricks and treats*. Acta Crystallogr D Biol Crystallogr, vol. 69, no. Pt 11, pages 2167–73, Nov 2013.
- [Abrahams 98] J P Abrahams & R A De Graaff. *New developments in phase refinement*. Curr Opin Struct Biol, vol. 8, no. 5, pages 601–5, Oct 1998.
- [Ackermann 09] Hans-W Ackermann. *Phage classification and characterization*. Methods Mol Biol, vol. 501, pages 127–40, 2009.
- [Ackermann 15] Hans-W Ackermann. *The lambda - P22 problem*. Bacteriophage, vol. 5, no. 1, page e1017084, 2015.
- [Adams 10] Paul D Adams, Pavel V Afonine, Gábor Bunkóczi, Vincent B Chen, Ian W Davis, Nathaniel Echols, Jeffrey J Headd, Li-Wei Hung, Gary J Kapral, Ralf W Grosse-Kunstleve, Airlie J McCoy, Nigel W Moriarty, Robert Oeffner, Randy J Read, David C Richardson, Jane S Richardson, Thomas C Terwilliger & Peter H Zwart. *PHENIX : a comprehensive Python-based system for macromolecular structure solution*. Acta Crystallogr D Biol Crystallogr, vol. 66, no. Pt 2, pages 213–21, Feb 2010.
- [Aksyuk 09] Anastasia A Aksyuk, Petr G Leiman, Lidia P Kurochkina, Mikhail M Shneider, Victor A Kostyuchenko, Vadim V Mesyanzhinov & Michael G Rossmann. *The tail sheath structure of bacteriophage T4 : a molecular machine for infecting bacteria*. EMBO J, vol. 28, no. 7, pages 821–9, Apr 2009.
- [Andres 10] Dorothee Andres, Christin Hanke, Ulrich Baxa, Anait Seul, Stefanie Barbirz & Robert Seckler. *Tailspike interactions with lipopolysaccharide effect DNA ejection*

- from phage P22 particles *in vitro*. *J Biol Chem*, vol. 285, no. 47, pages 36768–75, Nov 2010.
- [ANSES 14] ANSES. *Evaluation des risques d'émergence d'antibiorésistances liées aux modes d'utilisation des antibiotiques dans le domaine de la santé animale*. Rapport ANSES, Apr 2014.
- [Arisaka 81] F Arisaka, J Engel & H Klump. *Contraction and dissociation of the bacteriophage T₄ tail sheath induced by heat and urea*. *Prog Clin Biol Res*, vol. 64, pages 365–79, 1981.
- [Auzat 08] Isabelle Auzat, Anja Dröge, Frank Weise, Rudi Lurz & Paulo Tavares. *Origin and function of the two major tail proteins of bacteriophage SPP1*. *Mol Microbiol*, vol. 70, no. 3, pages 557–69, Nov 2008.
- [Bai 15] Xiao-chen Bai, Eeson Rajendra, Guanghui Yang, Yigong Shi & Sjors H W Scheres. *Sampling the conformational space of the catalytic subunit of human -secretase*. *Elife*, vol. 4, Dec 2015.
- [Baumann 70] L Baumann, W C Benz, A Wright & E B Goldberg. *Inactivation of urea-treated phage T₄ by phosphatidylglycerol*. *Virology*, vol. 41, no. 2, pages 356–64, Jun 1970.
- [Bayburt 10] Timothy H Bayburt & Stephen G Sligar. *Membrane protein assembly into Nanodiscs*. *FEBS Lett*, vol. 584, no. 9, pages 1721–7, May 2010.
- [Bebeacua 10] Cecilia Bebeacua, Patrick Bron, Livia Lai, Christina Skovgaard Vegge, Lone Brøndsted, Silvia Spinelli, Valérie Campanacci, David Veessler, Marin van Heel & Christian Cambillau. *Structure and molecular assignment of lactococcal phage TP901-1 baseplate*. *J Biol Chem*, vol. 285, no. 50, pages 39079–86, Dec 2010.
- [Bernardi 14] Rafael C Bernardi, Isaac Cann & Klaus Schulten. *Molecular dynamics study of enhanced Man5B enzymatic activity*. *Biotechnol Biofuels*, vol. 7, page 83, 2014.
- [BERTANI 53] G BERTANI & J J WEIGLE. *Host controlled variation in bacterial viruses*. *J Bacteriol*, vol. 65, no. 2, pages 113–21, Feb 1953.
- [Bhardwaj 14] Anshul Bhardwaj, Adam S Olia & Gino Cingolani. *Architecture of viral genome-delivery molecular machines*. *Curr Opin Struct Biol*, vol. 25, pages 1–8, Apr 2014.

- [Bhardwaj 16] Anshul Bhardwaj, Rajeshwer S Sankhala, Adam S Olia, Dewey Brooke, Sherwood R Casjens, Derek J Taylor, Peter E Prevelige Jr & Gino Cingolani. *Structural Plasticity of the Protein Plug That Traps Newly Packaged Genomes in Podoviridae Virions*. J Biol Chem, vol. 291, no. 1, pages 215–26, Jan 2016.
- [Bikard 14] David Bikard, Chad W Euler, Wenyan Jiang, Philip M Nussenzweig, Gregory W Goldberg, Xavier Duportet, Vincent A Fischetti & Luciano A Marraffini. *Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials*. Nat Biotechnol, vol. 32, no. 11, pages 1146–50, Nov 2014.
- [Böhm 01] J Böhm, O Lambert, A S Frangakis, L Letellier, W Baumeister & J L Rigaud. *FhuA-mediated phage genome transfer into liposomes : a cryo-electron tomography study*. Curr Biol, vol. 11, no. 15, pages 1168–75, Aug 2001.
- [Borysowski 06] Jan Borysowski, Beata Weber-Dabrowska & Andrzej Górski. *Bacteriophage endolysins as a novel class of antibacterial agents*. Exp Biol Med (Maywood), vol. 231, no. 4, pages 366–77, Apr 2006.
- [Borysowski 08] Jan Borysowski & Andrzej Górski. *Is phage therapy acceptable in the immunocompromised host ?* Int J Infect Dis, vol. 12, no. 5, pages 466–71, Sep 2008.
- [Boulangier 08] Pascale Boulangier, Pierre Jacquot, Laure Plançon, Mohamed Chami, Andreas Engel, Claudine Parquet, Chantal Herbeuval & Lucienne Letellier. *Phage T5 straight tail fiber is a multifunctional protein acting as a tape measure and carrying fusogenic and muralytic activities*. J Biol Chem, vol. 283, no. 20, pages 13556–64, May 2008.
- [Boyd 12] E Fidelma Boyd. *Bacteriophage-encoded bacterial virulence factors and phage-pathogenicity island interactions*. Adv Virus Res, vol. 82, pages 91–118, 2012.
- [Bricogne 03] G Bricogne, C Vonrhein, C Flensburg, M Schiltz & W Paciorek. *Generation, representation and flow of phase information in structure determination : recent developments in and around SHARP 2.0*. Acta Crystallogr D Biol Crystallogr, vol. 59, no. Pt 11, pages 2023–30, Nov 2003.

- [Briers 14] Yves Briers, Maarten Walmagh, Victor Van Puyenbroeck, Anneleen Cornelissen, William Cenens, Abram Aertsen, Hugo Oliveira, Joana Azeredo, Gunther Verween, Jean-Paul Pirnay, Stefan Miller, Guido Volckaert & Rob Lavigne. *Engineered endolysin-based "Artilyns" to combat multidrug-resistant gram-negative pathogens*. MBio, vol. 5, no. 4, pages e01379–14, Jul 2014.
- [Browning 12] Christopher Browning, Mikhail M Shneider, Valorie D Bowman, David Schwarzer & Petr G Leiman. *Phage pierces the host cell membrane with the iron-loaded spike*. Structure, vol. 20, no. 2, pages 326–39, Feb 2012.
- [Brum 13] Jennifer R Brum, Ryan O Schenck & Matthew B Sullivan. *Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses*. ISME J, vol. 7, no. 9, pages 1738–51, Sep 2013.
- [Brum 16] Jennifer R Brum, J Cesar Ignacio-Espinoza, Eun-Hae Kim, Gareth Trubl, Robert M Jones, Simon Roux, Nathan C VerBerkmoes, Virginia I Rich & Matthew B Sullivan. *Illuminating structural proteins in viral "dark matter" with metaproteomics*. Proc Natl Acad Sci U S A, vol. 113, no. 9, pages 2436–41, Mar 2016.
- [Bunick 00] C Bunick, A C North & G Stubbs. *Evaporative microdialysis : an effective improvement in an established method of protein crystallization*. Acta Crystallogr D Biol Crystallogr, vol. 56, no. Pt 11, pages 1430–1, Nov 2000.
- [Buth 15] Sergey A Buth, Laure Menin, Mikhail M Shneider, Jürgen Engel, Sergei P Boudko & Petr G Leiman. *Structure and Biophysical Properties of a Triple-Stranded Beta-Helix Comprising the Central Spike of Bacteriophage T4*. Viruses, vol. 7, no. 8, pages 4676–706, Aug 2015.
- [Casjens 15] Sherwood R Casjens & Roger W Hendrix. *Bacteriophage lambda : Early pioneer and still relevant*. Virology, vol. 479-480, pages 310–30, May 2015.
- [Chaban 15] Yuriy Chaban, Rudi Lurz, Sandrine Brasilès, Charène Cornilleau, Matthia Karreman, Sophie Zinn-Justin, Paulo Tavares & Elena V Orlova. *Structural rearrangements in the phage head-to-tail interface during assembly*

- and infection*. Proc Natl Acad Sci U S A, vol. 112, no. 22, pages 7009–14, Jun 2015.
- [Chan 16] Benjamin K Chan, Mark Sistrom, John E Wertz, Kaitlyn E Kortright, Deepak Narayan & Paul E Turner. *Phage selection restores antibiotic sensitivity in MDR Pseudomonas aeruginosa*. Sci Rep, vol. 6, page 26717, May 2016.
- [Chang 06] Juan Chang, Peter Weigle, Jonathan King, Wah Chiu & Wen Jiang. *Cryo-EM asymmetric reconstruction of bacteriophage P22 reveals organization of its DNA packaging and infecting machinery*. Structure, vol. 14, no. 6, pages 1073–82, Jun 2006.
- [Chanishvili 12] Nina Chanishvili. *Phage therapy—history from Twort and d’Herelle through Soviet experience to current approaches*. Adv Virus Res, vol. 83, pages 3–40, 2012.
- [Cheng 14] Naiqian Cheng, Weimin Wu, Norman R Watts & Alasdair C Steven. *Exploiting radiation damage to map proteins in nucleoprotein complexes : the internal structure of bacteriophage T7*. J Struct Biol, vol. 185, no. 3, pages 250–6, Mar 2014.
- [CoL] catalog of life [online]. [www](http://www.ncbi.nlm.nih.gov/taxonomy) [cited 23/06/2017]
- [Coombs 94] D H Coombs & Fumio Arisaka. Molecular biology of bacteriophage t4, chapitre T4 tail structure and function, pages 259–81. American Society for Microbiology : Washington D.C., 1994.
- [Cowtan 06] Kevin Cowtan. *The Buccaneer software for automated model building. 1. Tracing protein chains*. Acta Crystallogr D Biol Crystallogr, vol. 62, no. Pt 9, pages 1002–11, Sep 2006.
- [CRICK 61] F H CRICK, L BARNETT, S BRENNER & R J WATTS-TOBIN. *General nature of the genetic code for proteins*. Nature, vol. 192, pages 1227–32, Dec 1961.
- [Criscuolo 17] Elena Criscuolo, Sara Spadini, Jacopo Lamanna, Mattia Ferro & Roberto Burioni. *Bacteriophages and Their Immunological Applications against Infectious Threats*. J Immunol Res, vol. 2017, page 3780697, 2017.
- [Cronan 03] John E Cronan. *Cosmid-based system for transient expression and absolute off-to-on transcriptional control of*

- Escherichia coli* genes. J Bacteriol, vol. 185, no. 22, pages 6522–9, Nov 2003.
- [Cuervo 13] Ana Cuervo, Mar Pulido-Cid, Mónica Chagoyen, Rocío Arranz, Verónica A González-García, Carmela Garcia-Doval, José R Castón, José M Valpuesta, Mark J van Raaij, Jaime Martín-Benito & José L Carrascosa. *Structural characterization of the bacteriophage T7 tail machinery*. J Biol Chem, vol. 288, no. 36, pages 26290–9, Sep 2013.
- [Cuniasse 17] Philippe Cuniasse, Paulo Tavares, Elena V Orlova & Sophie Zinn-Justin. *Structures of biomolecular complexes by combination of NMR and cryoEM methods*. Curr Opin Struct Biol, vol. 43, pages 104–113, Apr 2017.
- [Danovaro 08] Roberto Danovaro, Antonio Dell’Anno, Cinzia Corinaldesi, Mirko Magagnini, Rachel Noble, Christian Tamburini & Markus Weinbauer. *Major viral impact on the functioning of benthic deep-sea ecosystems*. Nature, vol. 454, no. 7208, pages 1084–7, Aug 2008.
- [Davidson 12] Alan R Davidson, Lia Cardarelli, Lisa G Pell, Devon R Radford & Karen L Maxwell. *Long noncontractile tail machines of bacteriophages*. Adv Exp Med Biol, vol. 726, pages 115–42, 2012.
- [De Rosier 68] D J De Rosier & A Klug. *Reconstruction of three dimensional structures from electron micrographs*. Nature, vol. 217, no. 5124, pages 130–4, Jan 1968.
- [Denisov 04] I G Denisov, Y V Grinkova, A A Lazarides & S G Sligar. *Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size*. J Am Chem Soc, vol. 126, no. 11, pages 3477–87, Mar 2004.
- [Denyes 17] Jenna M Denyes, Matthew Dunne, Stanislava Steiner, Maximilian Mittelviehhaus, Agnes Weiss, Herbert Schmidt, Jochen Klumpp & Martin J Loessner. *Modified Bacteriophage S16 Long Tail Fiber Proteins for Rapid and Specific Immobilization and Detection of Salmonella Cells*. Appl Environ Microbiol, vol. 83, no. 12, Jun 2017.
- [Dublanchet 07] Alain Dublanchet & Shawna Bourne. *The epic of phage therapy*. Can J Infect Dis Med Microbiol, vol. 18, no. 1, pages 15–8, Jan 2007.

- [Dufour 17] Nicolas Dufour, Raphaëlle Delattre, Jean-Damien Ricard & Laurent Debarbieux. *The Lysis of Pathogenic Escherichia coli by Bacteriophages Releases Less Endotoxin Than by β -Lactams*. Clin Infect Dis, vol. 64, no. 11, pages 1582–1588, Jun 2017.
- [Dutilh 14] Bas E Dutilh, Noriko Cassman, Katelyn McNair, Savannah E Sanchez, Genivaldo G Z Silva, Lance Boling, Jeremy J Barr, Daan R Speth, Victor Seguritan, Ramy K Aziz, Ben Felts, Elizabeth A Dinsdale, John L Mokili & Robert A Edwards. *A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes*. Nat Commun, vol. 5, page 4498, Jul 2014.
- [ECDC/EFSA/EMA 15] ECDC/EFSA/EMA. *ECDC/EFSA/EMA first joint report on the integrated analysis of the consumption of antimicrobial agents and occurrence of antimicrobial resistance in bacteria from humans and food-producing animals*. Rapport ECDC, Jan 2015.
- [Effantin 06] G Effantin, P Boulanger, E Neumann, L Letellier & J F Conway. *Bacteriophage T5 structure reveals similarities with HK97 and T4 suggesting evolutionary relationships*. J Mol Biol, vol. 361, no. 5, pages 993–1002, Sep 2006.
- [Emsley 10] P Emsley, B Lohkamp, W G Scott & K Cowtan. *Features and development of Coot*. Acta Crystallogr D Biol Crystallogr, vol. 66, no. Pt 4, pages 486–501, Apr 2010.
- [Fabiola 05] Felcy Fabiola & Michael S Chapman. *Fitting of high-resolution structures into electron microscopy reconstruction images*. Structure, vol. 13, no. 3, pages 389–400, Mar 2005.
- [Ferguson 98] A D Ferguson, E Hofmann, J W Coulton, K Diederichs & W Welte. *Siderophore-mediated iron transport : crystal structure of FhuA with bound lipopolysaccharide*. Science, vol. 282, no. 5397, pages 2215–20, Dec 1998.
- [Flayhan 12] Ali Flayhan, Frank Wien, Maïté Paternostre, Pascale Boulanger & Cécile Breyton. *New insights into pb5, the receptor binding protein of bacteriophage T5, and its interaction with its Escherichia coli receptor FhuA*. Biochimie, vol. 94, no. 9, pages 1982–9, Sep 2012.
- [Flayhan 14] Ali Flayhan, Frédéric M D Vellieux, Rudi Lurz, Olivier Maury, Carlos Contreras-Martel, Eric Girard, Pascale

- Boulanger & Cécile Breyton. *Crystal structure of pb9, the distal tail protein of bacteriophage T5 : a conserved structural motif among all siphophages*. J Virol, vol. 88, no. 2, pages 820–8, Jan 2014.
- [Fokine 13] Andrei Fokine, Zhihong Zhang, Shuji Kanamaru, Valorie D Bowman, Anastasia A Aksyuk, Fumio Arisaka, Venigalla B Rao & Michael G Rossmann. *The molecular architecture of the bacteriophage T4 neck*. J Mol Biol, vol. 425, no. 10, pages 1731–44, May 2013.
- [Fokine 16] Andrei Fokine & Michael G Rossmann. *Common Evolutionary Origin of Procapsid Proteases, Phage Tail Tubes, and Tubes of Bacterial Type VI Secretion Systems*. Structure, vol. 24, no. 11, pages 1928–1935, Nov 2016.
- [Fraga 17] Hugo Fraga, Charles-Adrien Arnaud, Diego F Gauto, Maxime J C Audin, Vilius Kurauskas, Pavel Macek, Carsten Krichel, Jia-Ying Guan, Jérôme Boisbouvier, Remco Sprangers, Cécile Breyton & Paul Schanda. *Solid-state NMR H-N-(C)-H and H-N-C-C 3D/4D correlation experiments for resonance assignment of large proteins*. Chemphyschem, Aug 2017.
- [Frank 96] J Frank, M Radermacher, P Penczek, J Zhu, Y Li, M Ladjadj & A Leith. *SPIDER and WEB : processing and visualization of images in 3D electron microscopy and related fields*. J Struct Biol, vol. 116, no. 1, pages 190–9, 1996.
- [Franzetti 02] B Franzetti, G Schoehn, J-F Hernandez, M Jaquinod, R W H Ruigrok & G Zaccai. *Tetrahedral aminopeptidase : a novel large protease complex from archaea*. EMBO J, vol. 21, no. 9, pages 2132–8, May 2002.
- [Garcia-Doval 13] Carmela Garcia-Doval, Daniel Luque, José R Castón, Pascale Boulanger & Mark J van Raaij. *Crystallization of the C-terminal domain of the bacteriophage T5 L-shaped fibre*. Acta Crystallogr Sect F Struct Biol Cryst Commun, vol. 69, no. Pt 12, pages 1363–7, Dec 2013.
- [Garcia-Doval 15] Carmela Garcia-Doval, José R Castón, Daniel Luque, Meritxell Granell, José M Otero, Antonio L Llamas-Saiz, Madalena Renouard, Pascale Boulanger & Mark J

- van Raaij. *Structure of the Receptor-Binding Carboxy-Terminal Domain of the Bacteriophage T5 L-Shaped Tail Fibre with and without Its Intra-Molecular Chaperone*. *Viruses*, vol. 7, no. 12, pages 6424–40, Dec 2015.
- [Goldberg 94] E Goldberg, L Grinius & L Letellier. *Molecular biology of bacteriophage t4*, chapitre Recognition, attachment and injection, pages 347–56. American Society for Microbiology : Washington D.C., 1994.
- [González-García 15a] Verónica A González-García, Rebeca Bocanegra, Mar Pulido-Cid, Jaime Martín-Benito, Ana Cuervo & José L Carrascosa. *Characterization of the initial steps in the T7 DNA ejection process*. *Bacteriophage*, vol. 5, no. 3, page e1056904, 2015.
- [González-García 15b] Verónica A González-García, Mar Pulido-Cid, Carmela Garcia-Doval, Rebeca Bocanegra, Mark J van Raaij, Jaime Martín-Benito, Ana Cuervo & José L Carrascosa. *Conformational changes leading to T7 DNA delivery upon interaction with the bacterial receptor*. *J Biol Chem*, vol. 290, no. 16, pages 10038–44, Apr 2015.
- [Goodridge 11] Lawrence D Goodridge & Bledar Bisha. *Phage-based bio-control strategies to reduce foodborne pathogens in foods*. *Bacteriophage*, vol. 1, no. 3, pages 130–137, May 2011.
- [Goren 17] Moran Goren, Ido Yosef & Udi Qimron. *Sensitizing pathogens to antibiotics using the CRISPR-Cas system*. *Drug Resist Updat*, vol. 30, pages 1–6, Jan 2017.
- [Górski 12] Andrzej Górski, Ryszard Międzybrodzki, Jan Borysowski, Krystyna Dąbrowska, Piotr Wierzbicki, Monika Ohams, Grażyna Korczak-Kowalska, Natasza Olszowska-Zaremba, Marzena Łusiak-Szelachowska, Marlena Kłak, Ewa Jończyk, Ewelina Kaniuga, Aneta Gołaś, Sylwia Purchla, Beata Weber-Dąbrowska, Sławomir Letkiewicz, Wojciech Fortuna, Krzysztof Szufnarowski, Zdzisław Pawełczyk, Paweł Rogóż & Danuta Kłosowska. *Phage as a modulator of immune responses : practical implications for phage therapy*. *Adv Virus Res*, vol. 83, pages 41–71, 2012.
- [Goulet 11] Adeline Goulet, Joséphine Lai-Kee-Him, David Veessler, Isabelle Auzat, Gautier Robin, Dale A Shepherd, Alison E Ashcroft, Eric Richard, Julie Lichière, Paulo Ta-

- vares, Christian Cambillau & Patrick Bron. *The opening of the SPP1 bacteriophage tail, a prevalent mechanism in Gram-positive-infecting siphophages*. J Biol Chem, vol. 286, no. 28, pages 25397–405, Jul 2011.
- [Guo 13] Fei Guo, Zheng Liu, Frank Vago, Yue Ren, Weimin Wu, Elena T Wright, Philip Serwer & Wen Jiang. *Visualization of uncorrelated, tandem symmetry mismatches in the internal genome packaging apparatus of bacteriophage T7*. Proc Natl Acad Sci U S A, vol. 110, no. 17, pages 6811–6, Apr 2013.
- [Habenstein 15] Birgit Habenstein, Antoine Loquet, Songhwan Hwang, Karin Giller, Suresh Kumar Vasa, Stefan Becker, Michael Habeck & Adam Lange. *Hybrid Structure of the Type 1 Pilus of Uropathogenic Escherichia coli*. Angew Chem Int Ed Engl, vol. 54, no. 40, pages 11691–5, Sep 2015.
- [Henderson 90] R Henderson, J M Baldwin, T A Ceska, F Zemlin, E Beckmann & K H Downing. *Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy*. J Mol Biol, vol. 213, no. 4, pages 899–929, Jun 1990.
- [Henry 12] Marine Henry & Laurent Debarbieux. *Tools from viruses : bacteriophage successes and beyond*. Virology, vol. 434, no. 2, pages 151–61, Dec 2012.
- [HERSHEY 52] A D HERSHEY & M CHASE. *Independent functions of viral protein and nucleic acid in growth of bacteriophage*. J Gen Physiol, vol. 36, no. 1, pages 39–56, May 1952.
- [Hobbs 16] Zack Hobbs & Stephen T Abedon. *Diversity of phage infection types and associated terminology : the problem with 'Lytic or lysogenic'*. FEMS Microbiol Lett, vol. 363, no. 7, Apr 2016.
- [Hodyra-Stefaniak 15] Katarzyna Hodyra-Stefaniak, Paulina Miernikiewicz, Jarosław Drapała, Marek Drab, Ewa Jończyk-Matysiak, Dorota Lecion, Zuzanna Kaźmierczak, Weronika Beta, Joanna Majewska, Marek Harhala, Barbara Bubak, Anna Kłopot, Andrzej Górski & Krystyna Dąbrowska. *Mammalian Host-Versus-Phage immune response determines phage fate in vivo*. Sci Rep, vol. 5, page 14802, Oct 2015.

- [Hu 13] Bo Hu, William Margolin, Ian J Molineux & Jun Liu. *The bacteriophage t7 virion undergoes extensive structural remodeling during infection*. Science, vol. 339, no. 6119, pages 576–9, Feb 2013.
- [Hu 15] Bo Hu, William Margolin, Ian J Molineux & Jun Liu. *Structural remodeling of bacteriophage T4 and host membranes during infection initiation*. Proc Natl Acad Sci U S A, vol. 112, no. 35, pages E4919–28, Sep 2015.
- [Huet 10] Alexis Huet, James F Conway, Lucienne Letellier & Pascale Boulanger. *In vitro assembly of the T=13 procapsid of bacteriophage T5 with its scaffolding domain*. J Virol, vol. 84, no. 18, pages 9350–8, Sep 2010.
- [Hug 16] Laura A Hug, Brett J Baker, Karthik Anantharaman, Christopher T Brown, Alexander J Probst, Cindy J Castelle, Cristina N Butterfield, Alex W HERNSDORF, Yuki Amano, Kotaro Ise, Yohey Suzuki, Natasha Dudek, David A Relman, Kari M Finstad, Ronald Amundson, Brian C Thomas & Jillian F Banfield. *A new view of the tree of life*. Nat Microbiol, vol. 1, page 16048, Apr 2016.
- [ICT] [online][[www](#)] [cited 14/08/2017]
- [Ingham 90] C Ingham, M Buechner & J Adler. *Effect of outer membrane permeability on chemotaxis in Escherichia coli*. J Bacteriol, vol. 172, no. 7, pages 3577–83, Jul 1990.
- [Jakutyte 12] Lina Jakutyte, Rudi Lurz, Catarina Baptista, Rut Carballido-Lopez, Carlos São-José, Paulo Tavares & Rimantas Daugelavicius. *First steps of bacteriophage SPP1 entry into Bacillus subtilis*. Virology, vol. 422, no. 2, pages 425–34, Jan 2012.
- [Jones 12] Jeffrey B Jones, Gary E Vallad, Fanny B Iriarte, Obradovic Aleksa, Mine H Wernsing, Lee E Jackson, Botond Balogh, Jason C Hong & Timur M Momol. *Considerations for using bacteriophages for plant disease control*. Bacteriophage, vol. 2, no. 4, pages 208–14, Dec 2012.
- [Karaolis 99] D K Karaolis, S Somara, D R Maneval Jr, J A Johnson & J B Kaper. *A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria*. Nature, vol. 399, no. 6734, pages 375–9, May 1999.

- [Katsura 84] I Katsura & R W Hendrix. *Length determination in bacteriophage lambda tails*. Cell, vol. 39, no. 3 Pt 2, pages 691–8, Dec 1984.
- [Kemp 04] Priscilla Kemp, Manisha Gupta & Ian J Molineux. *Bacteriophage T7 DNA ejection into cells is initiated by an enzyme-like mechanism*. Mol Microbiol, vol. 53, no. 4, pages 1251–65, Aug 2004.
- [Kemp 05] Priscilla Kemp, L René Garcia & Ian J Molineux. *Changes in bacteriophage T7 virion structure at the initiation of infection*. Virology, vol. 340, no. 2, pages 307–17, Sep 2005.
- [Kropinsky 16] Andrew M Kropinsky. *2016 update from the ICTV bacterial and archaeal viruses subcommittee*. In Viruses of Microbe EMBO conference - 2016 Liverpool, 2016.
- [Kurauskas 17] Vilius Kurauskas, Sergei A Izmailov, Olga N Rogacheva, Audrey Hessel, Isabel Ayala, Joyce Woodhouse, Anastasya Shilova, Yi Xue, Tairan Yuwen, Nicolas Coquelle, Jacques-Philippe Colletier, Nikolai R Skrynnikov & Paul Schanda. *Slow conformational exchange and overall rocking motion in ubiquitin protein crystals*. Nat Commun, vol. 8, no. 1, page 145, Jul 2017.
- [Lance 16] Shea T Lance, David J Sukovich, Kenneth M Stedman & Adam R Abate. *Peering below the diffraction limit : robust and specific sorting of viruses with flow cytometry*. Virol J, vol. 13, no. 1, page 201, Dec 2016.
- [Lander 09] Gabriel C Lander, Reza Khayat, Rui Li, Peter E Prevelige, Clinton S Potter, Bridget Carragher & John E Johnson. *The P22 tail machine at subnanometer resolution reveals the architecture of an infection conduit*. Structure, vol. 17, no. 6, pages 789–99, Jun 2009.
- [Langlois 15] Chantal Langlois, Stéphanie Ramboarina, Abhishek Cukkemane, Isabelle Auzat, Benjamin Chagot, Bernard Gilquin, Athanasios Ignatiou, Isabelle Petitpas, Emmanouil Kasotakis, Maïté Paternostre, Helen E White, Elena V Orlova, Marc Baldus, Paulo Tavares & Sophie Zinn-Justin. *Bacteriophage SPP1 tail tube protein self-assembles into -structure-rich tubes*. J Biol Chem, vol. 290, no. 6, pages 3836–49, Feb 2015.

- [LANNI 58] Y T LANNI. *Lysis inhibition with a mutant of bacteriophage T5*. Virology, vol. 5, no. 3, pages 481–501, Jun 1958.
- [Leavitt 13] Justin C Leavitt, Lasha Gogokhia, Eddie B Gilcrease, Anshul Bhardwaj, Gino Cingolani & Sherwood R Casjens. *The tip of the tail needle affects the rate of DNA delivery by bacteriophage P22*. PLoS One, vol. 8, no. 8, page e70936, 2013.
- [Leforestier 08] A Leforestier, S Brasilès, M de Frutos, E Raspaud, L Letailier, P Tavares & F Livolant. *Bacteriophage T5 DNA ejection under pressure*. J Mol Biol, vol. 384, no. 3, pages 730–9, Dec 2008.
- [Leiman 09] Petr G Leiman, Marek Basler, Udipi A Ramagopal, Jeffrey B Bonanno, J Michael Sauder, Stefan Pukatzki, Stephen K Burley, Steven C Almo & John J Mekalanos. *Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin*. Proc Natl Acad Sci U S A, vol. 106, no. 11, pages 4154–9, Mar 2009.
- [Lennon 16] Jay T Lennon & Kenneth J Locey. *The Underestimation of Global Microbial Diversity*. MBio, vol. 7, no. 5, Sep 2016.
- [Leptihn 16] Sebastian Leptihn, Julia Gottschalk & Andreas Kuhn. *T7 ejectosome assembly : A story unfolds*. Bacteriophage, vol. 6, no. 1, page e1128513, 2016.
- [Li 13] Xueming Li, Paul Mooney, Shawn Zheng, Christopher R Booth, Michael B Braunfeld, Sander Gubbens, David A Agard & Yifan Cheng. *Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM*. Nat Methods, vol. 10, no. 6, pages 584–90, Jun 2013.
- [Lindell 05] Debbie Lindell, Jacob D Jaffe, Zackary I Johnson, George M Church & Sallie W Chisholm. *Photosynthesis genes in marine viruses yield proteins during host infection*. Nature, vol. 438, no. 7064, pages 86–9, Nov 2005.
- [Loquet 12] Antoine Loquet, Nikolaos G Sgourakis, Rashmi Gupta, Karin Giller, Dietmar Riedel, Christian Goosmann, Christian Griesinger, Michael Kolbe, David Baker, Stefan Becker & Adam Lange. *Atomic model of the type*

- III secretion system needle*. Nature, vol. 486, no. 7402, pages 276–9, May 2012.
- [Ludtke 99] S J Ludtke, P R Baldwin & W Chiu. *EMAN : semiautomated software for high-resolution single-particle reconstructions*. J Struct Biol, vol. 128, no. 1, pages 82–97, Dec 1999.
- [Lupo 15] Domenico Lupo, Sebastian Leptihn, Gisela Nagler & Maximilian Haase. *The T7 ejection nanomachine components gp15-gp16 form a spiral ring complex that binds DNA and a lipid membrane*. Virology, vol. 486, pages 263–71, 2015.
- [Luria 43] S E Luria & M Delbrück. *Mutations of Bacteria from Virus Sensitivity to Virus Resistance*. Genetics, vol. 28, no. 6, pages 491–511, Nov 1943.
- [Ma 15] Peixiang Ma, Yi Xue, Nicolas Coquelle, Jens D Haller, Tairan Yuwen, Isabel Ayala, Oleg Mikhailovskii, Dieter Willbold, Jacques-Philippe Colletier, Nikolai R Skrynnikov & Paul Schanda. *Observing the overall rocking motion of a protein in a crystal*. Nat Commun, vol. 6, page 8361, Oct 2015.
- [Marraffini 15] Luciano A Marraffini. *CRISPR-Cas immunity in prokaryotes*. Nature, vol. 526, no. 7571, pages 55–61, Oct 2015.
- [McMullan 16] G McMullan, A R Faruqi & R Henderson. *Direct Electron Detectors*. Methods Enzymol, vol. 579, pages 1–17, 2016.
- [Meselson 68] M Meselson & R Yuan. *DNA restriction enzyme from E. coli*. Nature, vol. 217, no. 5134, pages 1110–4, Mar 1968.
- [Milo 16] Scarlet Milo, Naing Tun Thet, Dan Liu, Jonathan Nza-kizwanayo, Brian V Jones & A Toby A Jenkins. *An in-situ infection detection sensor coating for urinary catheters*. Biosens Bioelectron, vol. 81, pages 166–172, Jul 2016.
- [Moineau 02] Sylvain Moineau, D Tremblay & S Labrie. *Phages of lactic acid bacteria : from genomics to industrial applications*. ASM News, vol. 68, pages 388–93, 2002.
- [Molek 15] Peter Molek & Tomaž Bratkovič. *Bacteriophages as scaffolds for bipartite display : designing swiss army knives*

- on a nanoscale*. Bioconjug Chem, vol. 26, no. 3, pages 367–78, Mar 2015.
- [Molineux 13] Ian J Molineux & Debabrata Panja. *Popping the cork : mechanisms of phage genome ejection*. Nat Rev Microbiol, vol. 11, no. 3, pages 194–204, Mar 2013.
- [Murray 00] N E Murray. *Type I restriction systems : sophisticated molecular machines (a legacy of Bertani and Weigle)*. Microbiol Mol Biol Rev, vol. 64, no. 2, pages 412–34, Jun 2000.
- [Musgaard 17] Maria Musgaard, Teresa Paramo, Laura Domicевичa, Ole Juul Andersen & Philip C Biggin. *Insights into channel dysfunction from modelling and molecular dynamics simulations*. Neuropharmacology, Jun 2017.
- [Nakayama 00] K Nakayama, K Takashima, H Ishihara, T Shinomiya, M Kageyama, S Kanaya, M Ohnishi, T Murata, H Mori & T Hayashi. *The R-type pyocin of Pseudomonas aeruginosa is related to P2 phage, and the F-type is related to lambda phage*. Mol Microbiol, vol. 38, no. 2, pages 213–31, Oct 2000.
- [Nigro 17] Olivia D Nigro, Sean P Jungbluth, Huei-Ting Lin, Chih-Chiang Hsieh, Jaclyn A Miranda, Christopher R Schvarcz, Michael S Rappé & Grieg F Steward. *Viruses in the Oceanic Basement*. MBio, vol. 8, no. 2, Mar 2017.
- [Nogales 15] Eva Nogales. *An electron microscopy journey in the study of microtubule structure and dynamics*. Protein Sci, vol. 24, no. 12, pages 1912–9, Dec 2015.
- [Nováček 16] Jiří Nováček, Marta Šiborová, Martin Benešík, Roman Pantůček, Jiří Doškař & Pavel Plevka. *Structure and genome release of Twort-like Myoviridae phage with a double-layered baseplate*. Proc Natl Acad Sci U S A, vol. 113, no. 33, pages 9351–6, Aug 2016.
- [Organisation 15] World Health Organisation. *WHO estimates of the global burden of foodborne diseases*. Rapport OMS, 2015.
- [Paez-Espino 16] David Paez-Espino, Emiley A Eloé-Fadrosh, Georgios A Pavlopoulos, Alex D Thomas, Marcel Huntemann, Natalia Mikhailova, Edward Rubin, Natalia N Ivanova & Nikos C Kyrpides. *Uncovering Earth's virome*. Nature, vol. 536, no. 7617, pages 425–30, 08 2016.

- [Pal 90] M K Pal, T C Ghosh & J K Ghosh. *Studies on the conformation of and metal ion binding by teichoic acid of Staphylococcus aureus*. Biopolymers, vol. 30, no. 3-4, pages 273–7, 1990.
- [Pandurangan 12] Arun Prasad Pandurangan & Maya Topf. *RIBFIND : a web server for identifying rigid bodies in protein structures and to aid flexible fitting into cryo EM maps*. Bioinformatics, vol. 28, no. 18, pages 2391–3, Sep 2012.
- [Parent 14] Kristin N Parent, Marcella L Erb, Giovanni Cardone, Katrina Nguyen, Eddie B Gilcrease, Natalia B Porcek, Joe Pogliano, Timothy S Baker & Sherwood R Casjens. *OmpA and OmpC are critical host factors for bacteriophage Sf6 entry in Shigella*. Mol Microbiol, vol. 92, no. 1, pages 47–60, Apr 2014.
- [Parsons 12] Rachel J Parsons, Mya Breitbart, Michael W Lomas & Craig A Carlson. *Ocean time-series reveals recurring seasonal patterns of viroplankton dynamics in the northwestern Sargasso Sea*. ISME J, vol. 6, no. 2, pages 273–84, Feb 2012.
- [Patel 17] Anisha N Patel, Agnès Anne, Arnaud Chovin, Christophe Demaille, Eric Grelet, Thierry Michon & Cécilia Taofifenua. *Scaffolding of Enzymes on Virus Nanoarrays : Effects of Confinement and Virus Organization on Biocatalysis*. Small, vol. 13, no. 13, Apr 2017.
- [Pell 09a] Lisa G Pell, Voula Kanelis, Logan W Donaldson, P Lynne Howell & Alan R Davidson. *The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system*. Proc Natl Acad Sci U S A, vol. 106, no. 11, pages 4160–5, Mar 2009.
- [Pell 09b] Lisa G Pell, Amanda Liu, Lizbeth Edmonds, Logan W Donaldson, P Lynne Howell & Alan R Davidson. *The X-ray crystal structure of the phage lambda tail terminator protein reveals the biologically relevant hexameric ring structure and demonstrates a conserved mechanism of tail termination among diverse long-tailed phages*. J Mol Biol, vol. 389, no. 5, pages 938–51, Jun 2009.
- [Pell 10] Lisa G Pell, Genevieve M C Gasmi-Seabrook, Marc Morais, Philipp Neudecker, Voula Kanelis, Diane Bona, Lo-

- gan W Donaldson, Aled M Edwards, P Lynne Howell, Alan R Davidson & Karen L Maxwell. *The solution structure of the C-terminal Ig-like domain of the bacteriophage tail tube protein*. J Mol Biol, vol. 403, no. 3, pages 468–79, Oct 2010.
- [Penczek 10] Pawel A Penczek. *Fundamentals of three-dimensional reconstruction from projections*. Methods Enzymol, vol. 482, pages 1–33, 2010.
- [Pettersen 04] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng & Thomas E Ferrin. *UCSF Chimera—a visualization system for exploratory research and analysis*. J Comput Chem, vol. 25, no. 13, pages 1605–12, Oct 2004.
- [Pintilie 10] Grigore D Pintilie, Junjie Zhang, Thomas D Goddard, Wah Chiu & David C Gossard. *Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions*. J Struct Biol, vol. 170, no. 3, pages 427–38, Jun 2010.
- [Pintilie 16] Grigore Pintilie, Dong-Hua Chen, Cameron A Haase-Pettingell, Jonathan A King & Wah Chiu. *Resolution and Probabilistic Models of Components in CryoEM Maps of Mature P22 Bacteriophage*. Biophys J, vol. 110, no. 4, pages 827–39, Feb 2016.
- [Plisson 07] Celia Plisson, Helen E White, Isabelle Auzat, Amineh Zafarani, Carlos São-José, Sophie Lhuillier, Paulo Tavares & Elena V Orlova. *Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection*. EMBO J, vol. 26, no. 15, pages 3720–8, Aug 2007.
- [Ponchon 05] Luc Ponchon, Stéphanie Mangenot, Pascale Boulanger & Lucienne Letellier. *Encapsidation and transfer of phage DNA into host cells : from in vivo to single particles studies*. Biochim Biophys Acta, vol. 1724, no. 3, pages 255–61, Aug 2005.
- [Ponchon 06] Luc Ponchon, Pascale Boulanger, Gilles Labesse & Lucienne Letellier. *The endonuclease domain of bacteriophage terminases belongs to the resolvase/integrase/ribonuclease H superfamily : a bioinformatics analysis validated by a functional study on bacte-*

- riophage T5*. J Biol Chem, vol. 281, no. 9, pages 5829–36, Mar 2006.
- [Poorvin 04] Leo Poorvin, Johanna M Rinta-Kanto, David A Hutchins & Steven W Wilhelm. *Viral release of iron and its bioavailability to marine plankton*. Limnol oceanogr, vol. 49, no. 5, pages 1734–41, Sept 2004.
- [Rayment 02] Ivan Rayment. *Small-scale batch crystallization of proteins revisited : an underutilized way to grow large protein crystals*. Structure, vol. 10, no. 2, pages 147–51, Feb 2002.
- [Rohou 15] Alexis Rohou & Nikolaus Grigorieff. *CTFFIND4 : Fast and accurate defocus estimation from electron micrographs*. J Struct Biol, vol. 192, no. 2, pages 216–21, Nov 2015.
- [Romero 14] Jaime Romero, Gastón Higuera, Felipe Gajardo, Daniel Castillo, Mathias Middleboe, Katherine García, Carolina Ramírez & Romilio T Espejo. *Complete Genome Sequence of Vibrio anguillarum Phage CHOED Successfully Used for Phage Therapy in Aquaculture*. Genome Announc, vol. 2, no. 4, Jul 2014.
- [Scheible 77] P P Scheible, E A Rhoades & M Rhoades. *Localization of single-chain interruptions in bacteriophage T5 DNA I. Electron microscopic studies*. J Virol, vol. 23, no. 3, pages 725–36, Sep 1977.
- [Scheres 12] Sjors H W Scheres. *RELION : implementation of a Bayesian approach to cryo-EM structure determination*. J Struct Biol, vol. 180, no. 3, pages 519–30, Dec 2012.
- [Schmelcher 14] Mathias Schmelcher & Martin J Loessner. *Use of bacteriophage cell wall-binding proteins for rapid diagnostics of Listeria*. Methods Mol Biol, vol. 1157, pages 141–56, 2014.
- [Schmidt 16] Andreas Schmidt, Wolfgang Rabsch, Nina K Broeker & Stefanie Barbirz. *Bacteriophage tailspike protein based assay to monitor phase variable glucosylations in Salmonella O-antigens*. BMC Microbiol, vol. 16, page 207, Sep 2016.
- [Sheldrick 10] George M Sheldrick. *Experimental phasing with SHELXC/D/E : combining chain tracing with density*

- modification*. Acta Crystallogr D Biol Crystallogr, vol. 66, no. Pt 4, pages 479–85, Apr 2010.
- [Shi 15] Chaowei Shi, Pascal Fricke, Lin Lin, Veniamin Chevelkov, Melanie Wegstroth, Karin Giller, Stefan Becker, Martin Thanbichler & Adam Lange. *Atomic-resolution structure of cytoskeletal bactofilin by solid-state NMR*. Sci Adv, vol. 1, no. 11, page e1501087, Dec 2015.
- [Sigworth 10] Fred J Sigworth, Peter C Doerschuk, Jose-Maria Carazo & Sjors H W Scheres. *An introduction to maximum-likelihood methods in cryo-EM*. Methods Enzymol, vol. 482, pages 263–94, 2010.
- [Simon 67a] L D Simon & T F Anderson. *The infection of Escherichia coli by T2 and T4 bacteriophages as seen in the electron microscope. I. Attachment and penetration*. Virology, vol. 32, no. 2, pages 279–97, Jun 1967.
- [Simon 67b] L D Simon & T F Anderson. *The infection of Escherichia coli by T2 and T4 bacteriophages as seen in the electron microscope. II. Structure and function of the baseplate*. Virology, vol. 32, no. 2, pages 298–305, Jun 1967.
- [Smith 85] G P Smith. *Filamentous fusion phage : novel expression vectors that display cloned antigens on the virion surface*. Science, vol. 228, no. 4705, pages 1315–7, Jun 1985.
- [Spinelli 14] Silvia Spinelli, David Veessler, Cecilia Bebeacua & Christian Cambillau. *Structures and host-adhesion mechanisms of lactococcal siphophages*. Front Microbiol, vol. 5, page 3, 2014.
- [Srinivasiah 08] Sharath Srinivasiah, Jaysheel Bhavsar, Kanika Thapar, Mark Liles, Tom Schoenfeld & K Eric Wommack. *Phages across the biosphere : contrasts of viruses in soil and aquatic environments*. Res Microbiol, vol. 159, no. 5, pages 349–57, Jun 2008.
- [Sulakvelidze 01] A Sulakvelidze, Z Alavidze & J G Morris Jr. *Bacteriophage therapy*. Antimicrob Agents Chemother, vol. 45, no. 3, pages 649–59, Mar 2001.
- [Summers 01] W C Summers. *Bacteriophage therapy*. Annu Rev Microbiol, vol. 55, pages 437–51, 2001.
- [Suttle 94] C A Suttle & A M Chan. *Dynamics and Distribution of Cyanophages and Their Effect on Marine Synechococcus*

- spp.* Appl Environ Microbiol, vol. 60, no. 9, pages 3167–74, Sep 1994.
- [Suttle 07] Curtis A Suttle. *Marine viruses—major players in the global ecosystem*. Nat Rev Microbiol, vol. 5, no. 10, pages 801–12, Oct 2007.
- [Swift 16] Benjamin M C Swift, Thomas W Convery & Catherine E D Rees. *Evidence of Mycobacterium tuberculosis complex bacteraemia in intradermal skin test positive cattle detected using phage-RPA*. Virulence, vol. 7, no. 7, pages 779–88, Oct 2016.
- [Tavares 96] P Tavares, R Lurz, A Stiege, B Rückert & T A Trautner. *Sequential headful packaging and fate of the cleaved DNA ends in bacteriophage SPP1*. J Mol Biol, vol. 264, no. 5, pages 954–67, Dec 1996.
- [Taylor 76] K A Taylor & R M Glaeser. *Electron microscopy of frozen hydrated biological specimens*. J Ultrastruct Res, vol. 55, no. 3, pages 448–56, Jun 1976.
- [Taylor 16] Nicholas M I Taylor, Nikolai S Prokhorov, Ricardo C Guerrero-Ferreira, Mikhail M Shneider, Christopher Browning, Kenneth N Goldie, Henning Stahlberg & Petr G Leiman. *Structure of the T4 baseplate and its function in triggering sheath contraction*. Nature, vol. 533, no. 7603, pages 346–52, 05 2016.
- [Tissot 13] Alice Tissot. Surexpression, purification, caractérisation biochimique et biophysique et étude structurale de pb6, protéine majeure de queue du bactériophage t5. Master’s thesis, Université Claude Bernard - Lyon 1, 2013.
- [To 69] C M To, E Kellenberger & A Eisenstark. *Disassembly of T-even bacteriophage into structural parts and subunits*. J Mol Biol, vol. 46, no. 3, pages 493–511, Dec 1969.
- [Topf 08] Maya Topf, Keren Lasker, Ben Webb, Haim Wolfson, Wah Chiu & Andrej Sali. *Protein structure fitting and refinement guided by cryo-EM density*. Structure, vol. 16, no. 2, pages 295–307, Feb 2008.
- [Trabuco 08] Leonardo G Trabuco, Elizabeth Villa, Kakoli Mitra, Joachim Frank & Klaus Schulten. *Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics*. Structure, vol. 16, no. 5, pages 673–83, May 2008.

- [van Ooij 11] Christiaan van Ooij. *Molecular biology. Nature of the genetic code finally revealed!* Nat Rev Microbiol, vol. 9, no. 12, page 835, Nov 2011.
- [Veesler 10a] David Veesler, Stéphanie Blangy, Silvia Spinelli, Paulo Tavares, Valérie Campanacci & Christian Cambillau. *Crystal structure of Bacillus subtilis SPP1 phage gp22 shares fold similarity with a domain of lactococcal phage p2 RBP.* Protein Sci, vol. 19, no. 7, pages 1439–43, Jul 2010.
- [Veesler 10b] David Veesler, Gautier Robin, Julie Lichère, Isabelle Auzat, Paulo Tavares, Patrick Bron, Valérie Campanacci & Christian Cambillau. *Crystal structure of bacteriophage SPP1 distal tail protein (gp19.1) : a baseplate hub paradigm in gram-positive infecting phages.* J Biol Chem, vol. 285, no. 47, pages 36666–73, Nov 2010.
- [Veesler 11] David Veesler & Christian Cambillau. *A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries.* Microbiol Mol Biol Rev, vol. 75, no. 3, pages 423–33, first page of table of contents, Sep 2011.
- [Veesler 12] David Veesler, Silvia Spinelli, Jennifer Mahony, Julie Lichère, Stéphanie Blangy, Gérard Bricogne, Pierre Legend, Miguel Ortiz-Lombardia, Valérie Campanacci, Douwe van Sinderen & Christian Cambillau. *Structure of the phage TP901-1 1.8 MDa baseplate suggests an alternative host adhesion mechanism.* Proc Natl Acad Sci U S A, vol. 109, no. 23, pages 8954–8, Jun 2012.
- [Velusamy 10] Vijayalakshmi Velusamy, Khalil Arshak, Olga Korostynska, Kamila Oliwa & Catherine Adley. *An overview of foodborne pathogen detection : in the perspective of biosensors.* Biotechnol Adv, vol. 28, no. 2, pages 232–54, 2010.
- [Vernhes 17] Emeline Vernhes, Madalena Renouard, Bernard Gilquin, Philippe Cuniasse, Dominique Durand, Patrick England, Sylviane Hoos, Alexis Huet, James F Conway, Anatoly Glukhov, Vladimir Ksenzenko, Eric Jacquet, Naïma Nhiri, Sophie Zinn-Justin & Pascale Boulanger. *High affinity anchoring of the decoration protein pb10 onto the bacteriophage T5 capsid.* Sci Rep, vol. 7, page 41662, Feb 2017.

- [Vinay 15] Manon Vinay, Nathalie Franche, Gérald Grégori, Jean-Raphaël Fantino, Flavie Pouillot & Mireille Ansaldi. *Phage-Based Fluorescent Biosensor Prototypes to Specifically Detect Enteric Bacteria Such as E. coli and Salmonella enterica Typhimurium*. PLoS One, vol. 10, no. 7, page e0131466, 2015.
- [Vinga 12] Inês Vinga, Catarina Baptista, Isabelle Auzat, Isabelle Petipas, Rudi Lurz, Paulo Tavares, Mário A Santos & Carlos São-José. *Role of bacteriophage SPP1 tail spike protein gp21 on host cell receptor binding and trigger of phage DNA ejection*. Mol Microbiol, vol. 83, no. 2, pages 289–303, Jan 2012.
- [Voss 10] Neil R Voss, Dmitry Lyumkis, Anchi Cheng, Pick-Wei Lau, Anke Mulder, Gabriel C Lander, Edward J Brignole, Denis Fellmann, Christopher Irving, Erica L Jacovetty, Albert Leung, James Pulokas, Joel D Quispe, Hanspeter Winkler, Craig Yoshioka, Bridget Carragher & Clinton S Potter. *A toolbox for ab initio 3-D reconstructions in single-particle electron microscopy*. J Struct Biol, vol. 169, no. 3, pages 389–98, Mar 2010.
- [Wang 05] Jianbin Wang, Yan Jiang, Myriam Vincent, Yongqiao Sun, Hong Yu, Jing Wang, Qiyu Bao, Huimin Kong & Songnian Hu. *Complete genome sequence of bacteriophage T5*. Virology, vol. 332, no. 1, pages 45–65, Feb 2005.
- [Wang 16] Ziyuan Wang, Danhui Wang, Juhong Chen, David A Sela & Sam R Nugen. *Development of a novel bacteriophage based biomagnetic separation method as an aid for sensitive detection of viable Escherichia coli*. Analyst, vol. 141, no. 3, pages 1009–16, Feb 2016.
- [Whitford 11] Paul C Whitford, Aqeel Ahmed, Yanan Yu, Scott P Hennelly, Florence Tama, Christian M T Spahn, José N Onuchic & Karissa Y Sanbonmatsu. *Excited states of ribosome translocation revealed through integrative molecular modeling*. Proc Natl Acad Sci U S A, vol. 108, no. 47, pages 18943–8, Nov 2011.
- [Wommack 15] K Eri Wommack, Daniel J Nasko, Jessica Chopyk & Eric G Sakowski. *Counts and sequences, observations that continue to change our understanding of viruses in*

- nature*. J Microbiol, vol. 53, no. 3, pages 181–92, Mar 2015.
- [Wu 16] Weimin Wu, Justin C Leavitt, Naiqian Cheng, Eddie B Gilcrease, Tina Motwani, Carolyn M Teschke, Sherwood R Casjens & Alasdair C Steven. *Localization of the Houdinisome (Ejection Proteins) inside the Bacteriophage P22 Virion by Bubblegram Imaging*. MBio, vol. 7, no. 4, Aug 2016.
- [Yap 14] Moh Lan Yap & Michael G Rossmann. *Structure and function of bacteriophage T4*. Future Microbiol, vol. 9, no. 12, pages 1319–27, 2014.
- [Zhang 08] Xing Zhang, Ethan Settembre, Chen Xu, Philip R Dornitzter, Richard Bellamy, Stephen C Harrison & Nikolaus Grigorieff. *Near-atomic resolution using electron cryo-microscopy and single-particle reconstruction*. Proc Natl Acad Sci U S A, vol. 105, no. 6, pages 1867–72, Feb 2008.
- [Zinke 17] Maximilian Zinke, Pascal Fricke, Camille Samson, Songhwan Hwang, Joseph S Wall, Sascha Lange, Sophie Zinn-Justin & Adam Lange. *Bacteriophage Tail-Tube Assembly Studied by Proton-Detected 4D Solid-State NMR*. Angew Chem Int Ed Engl, vol. 56, no. 32, pages 9497–9501, Aug 2017.
- [Zivanovic 14] Yvan Zivanovic, Fabrice Confalonieri, Luc Ponchon, Rudi Lurz, Mohamed Chami, Ali Flayhan, Madalena Renouard, Alexis Huet, Paulette Decottignies, Alan R Davidson, Cécile Breyton & Pascale Boulanger. *Insights into bacteriophage T5 structure from analysis of its morphogenesis genes and protein components*. J Virol, vol. 88, no. 2, pages 1162–74, Jan 2014.