



HAL
open science

Etudes structurales et dynamiques du système de transport bactérien FhaB/FhaC et du complexe de réplication des Rhabdoviridae

Nicolas Martinez

► **To cite this version:**

Nicolas Martinez. Etudes structurales et dynamiques du système de transport bactérien FhaB/FhaC et du complexe de réplication des Rhabdoviridae. Autre [cond-mat.other]. Université de Grenoble, 2012. Français. NNT : 2012GRENM017 . tel-00721739

HAL Id: tel-00721739

<https://theses.hal.science/tel-00721739>

Submitted on 30 Jul 2012

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 L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Physique pour les sciences du vivant**

Arrêté ministériel : 7 août 2006

Présentée par

Nicolas MARTINEZ

Thèse dirigée par **Giuseppe ZACCAI** et
codirigée par **Marc JAMIN**

préparée au sein de l'**Institut Laue-Langevin** et l'**Unité
Interactions Virus-Cellule hôte**
dans l'**École Doctorale de Physique**

Etudes structurales et dynamiques du système de transport bactérien FhaB/FhaC et du complexe de réplication des *Rhabdoviridae*

Thèse soutenue publiquement le **15 Mars 2012**,
devant le jury composé de :

Pr. Judith PETERS

Université Joseph Fourier, Grenoble, France (Présidente)

Pr. Marie-Christine MAUREL

Laboratoire Acides Nucléiques et Biophotonique, Paris, France
(Rapporteur)

Dr. René WINTJENS

Université Libre de Bruxelles, Bruxelles, Belgique (Rapporteur)

Dr. Isabelle LANDRIEU

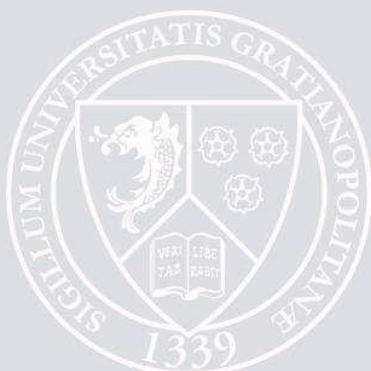
Laboratoire de Glycobiologie Structurale et Fonctionnelle, Lille, France
(Examinatrice)

Dr. Giuseppe ZACCAI

CNRS, Institut Laue-Langevin, Grenoble, France (Examinateur)

Pr. Marc JAMIN

Unit of Viral-Host Cell Interaction, Grenoble, France (Examinateur)



Remerciements

Je voudrais commencer par remercier Jo ZACCAI sans qui je n'aurais jamais pu faire cette thèse, mais surtout pour m'avoir initié aux techniques de diffusion neutronique. Je n'aurais jamais imaginé, à la fin de mon master, que j'allais découvrir les propriétés des neutrons et leur application dans le domaine des sciences du vivant. Je voudrais aussi remercier Marc JAMIN, qui m'a encadré à l'UVHCI durant ces trois années et des brouettes. Je ne pourrais jamais assez le remercier pour son ouverture d'esprit et pour m'avoir impliqué dans de nombreux projets, ainsi que pour sa disponibilité.

Merci à Marie-Christine MAUREL et René WINTJENS d'avoir accepté d'être rapporteurs de ce travail, et aussi à Judith PETERS et Isabelle LANDRIEU pour avoir accepté de faire partie du jury de cette thèse.

Une grande partie de ce travail s'est fait en collaboration avec l'Institut de Biologie de Lille. Mes remerciements vont Françoise JACOB-DUBUISSON pour m'avoir accueilli à Lille, mais surtout pour sa bonne humeur et sa disponibilité, ainsi que son aide précieuse dans le domaine de la biologie moléculaire. Je voudrais aussi remercier Catherine BAUD, mais aussi tous les membres de l'équipe de l'équipe Camille LOCHT pour leur accueil chaleureux. Cette collaboration m'a amené à travailler avec David ALSTEENS et Yves DUFRENE de l'UCL, que je remercie pour leur enthousiasme et le travail formidable qu'ils ont accompli sur le projet dépliement de Fha30.

Je remercie aussi Martin BLACKLEDGE ainsi que tous les membres de l'équipe FDP de l'IBS : Guigui, Malene, Val... pour toute l'aide qu'ils ont pu m'apporter (scripts shell, RMN, latex, etc...). J'en profite pour remercier aussi les lyonnais Emilie et Guido leur collaboration sur le projet Rougeole.

Les expériences de diffusion de neutrons aux petits angles n'auraient jamais pu être réalisées sans l'aide de Phil CALLOW sur D22. Je voudrais aussi remercier Judith PETERS et Francesca NATALI pour leur aide plus que précieuse sur IN13, mais aussi pour leur disponibilité.

Ces trois années et des poussières se sont faites dans un cadre de travail exceptionnel. Je tenais à remercier Euripedes RIBEIRO, mon maître, ainsi que Cédric LEYRAT, dit "le rat" et Ambroise DESFOSES, Ivan IVANOV et Filip YABUKARSKI pour l'ambiance qu'ils ont su donner au laboratoire 008. Merci aussi à tous mes collègues de l'UVHCI : Julien P, Audrey, Charlotte, Eric, Alex M, Nico T, Catherine, Martinellours, David, Aurélien, Emilie, Pauline, Jens, Leandro, Celeste, Stefan et tous ceux dont j'ai pu un jour croiser la route. Je voulais aussi remercier quelques "anciens" : Alex D, Lucy, Julien L, PAC et Habib.

Je tenais aussi à remercier mes amis : Adrien, Cindouche, Louménane, Carlito, Cristian et Olga, Isa, Alex ze Great, La Marie, Anne, Yoz, Ptiyo, Céline, Delph, Marlaine, Toto, Titou et bien d'autres que je ne pourrai pas nommer faute de place.

Finalement, je voulais remercier les membres de ma famille pour m'avoir soutenu pendant ces trois années, particulièrement mon frère Xav, mes parents et mes soeurs.

Table des matières

1	Introduction	1
1.1	Les systèmes de sécrétion bactériens de type V	1
1.1.1	Le système de transport FhaB/FhaC de <i>Bordetella pertussis</i>	3
1.2	Le complexe de réplication des <i>Rhabdoviridae</i>	6
1.2.1	La phosphoprotéine du virus de la stomatite vésiculaire (VSV)	8
1.3	Techniques biophysiques	9
1.3.1	Diffusion de neutrons	9
1.3.2	Diffusion aux petits angles	12
1.3.3	Diffusion incohérente de neutrons	18
1.3.4	Simulations de dynamique moléculaire	21
1.3.5	Microscopie à force atomique	22
2	Analyses structurales du complexe chaperonne-protéine client Fha30-Par27	25
2.1	Introduction	25
2.2	Matériels et méthodes	28
2.2.1	Préparation des échantillons	28
2.2.2	Diffusion de rayons X aux petits angles	29
2.2.3	Diffusion de neutrons aux petits angles (SANS)	29
2.2.4	Modélisation et analyse d'ensembles	29
2.3	Résultats et Discussion	31
2.3.1	Modélisation de Par27	31
2.3.2	Modélisation de Fha30	45
2.3.3	Modélisation de Fha30 dénaturé	46
2.3.4	Modélisation du complexe Fha30/Par27	48
2.4	Conclusion	53
3	Etude du repliement de Fha30	55
3.1	Introduction	55
3.2	Matériels et méthodes	58
3.2.1	Conception d'une séquence stabilisant la partie C-terminale de Fha30	58
3.2.2	Biologie moléculaire	59
3.2.3	Expression et Purification des protéines	59
3.2.4	Mesures de diffusion élastique incohérente	60
3.3	Résultats	60
3.3.1	Conception d'une coiffe sur la partie C-terminale de Fha30	60
3.3.2	Etude du dépliement de FHA par microscopie à force atomique et simulations de dynamique moléculaire	60

3.3.3	Etude des propriétés dynamiques de Fha30 par diffusion de neutrons incohérente	88
3.4	Conclusion	91
4	Analyses structurales des éléments du complexe de réplication des <i>Rhabdoviridae</i>	93
4.1	Introduction	93
4.2	Matériels et méthodes	95
4.2.1	Préparation des échantillons :	95
4.2.2	Collecte des données	96
4.2.3	Modélisation et analyse d'ensembles	96
4.3	Résultats et discussion	96
4.3.1	Modélisation du complexe N ⁰ -P	96
4.3.2	Détermination de la structure haute résolution du domaine C-terminal de la phosphoprotéine du virus de la jaunisse nécrotique de la laitue (LNYV)	102
4.4	Conclusion	138
5	Dynamique des éléments du complexe de réplication du virus de la stomatite vésiculaire	139
5.1	Introduction	139
5.2	Matériels et méthodes	141
5.2.1	Préparation des échantillons	141
5.2.2	Mesure de l'intensité élastique incohérente	141
5.3	Résultats	141
5.3.1	Dynamique de P-CTD	141
5.3.2	Dynamique de P-NTD	144
5.3.3	Dynamique de la P entière	146
5.4	Conclusion	148
6	Conclusion générale et perspectives	149
6.1	L'hémagglutinine filamenteuse de <i>Bordetella pertussis</i>	149
6.2	Le complexe de réplication des <i>Rhabdoviridae</i>	150
A	Etude du désordre dans les nucleocapsides du virus de la rougeole	151
	Bibliographie	159
	Liste des Abréviations	173

Introduction

"Les neutrons voient tout."

J. Zaccai, a propos de la diffusion incohérente de neutrons

Sommaire

1.1	Les systèmes de sécrétion bactériens de type V	1
1.1.1	Le système de transport FhaB/FhaC de <i>Bordetella pertussis</i>	3
1.2	Le complexe de réplication des <i>Rhabdoviridae</i>	6
1.2.1	La phosphoprotéine du virus de la stomatite vésiculaire (VSV)	8
1.3	Techniques biophysiques	9
1.3.1	Diffusion de neutrons	9
1.3.2	Diffusion aux petits angles	12
1.3.3	Diffusion incohérente de neutrons	18
1.3.4	Simulations de dynamique moléculaire	21
1.3.5	Microscopie à force atomique	22

1.1 Les systèmes de sécrétion bactériens de type V

Les bactéries à Gram négatif, possèdent différents types de systèmes leur permettant de sécréter des protéines dans le milieu extracellulaire. Parmi ces systèmes, le système de type V est étonnant par son apparente simplicité.

D'un point de vue fonctionnel, on peut diviser ce système en deux types : le système autotransporteur (AT), qui n'implique qu'une seule protéine, et le système à deux partenaires (TPS), qui implique une protéine transportée (appelée TpsA) et son transporteur spécifique (appelé TpsB) [Yen 2002].

La première étape de translocation membranaire (translocation du cytoplasme vers le périplasme) est commune aux deux systèmes et se fait par le système Sec. L'export dans le périplasme est médié par un peptide signal présent sur les parties N-terminales des protéines exportées, peptide qui est ensuite clivé après translocation [Pugsley 1993].

Une fois transportées dans le périplasme, elles vont pouvoir être sécrétées dans le milieu extracellulaire grâce à leur transporteur spécifique, qui est soit lié de manière covalente sur la partie C-terminale de la protéine sécrétée (système autotransporteur), soit séparé (système TPS) (voir figure 1.1). Le module de transport du système

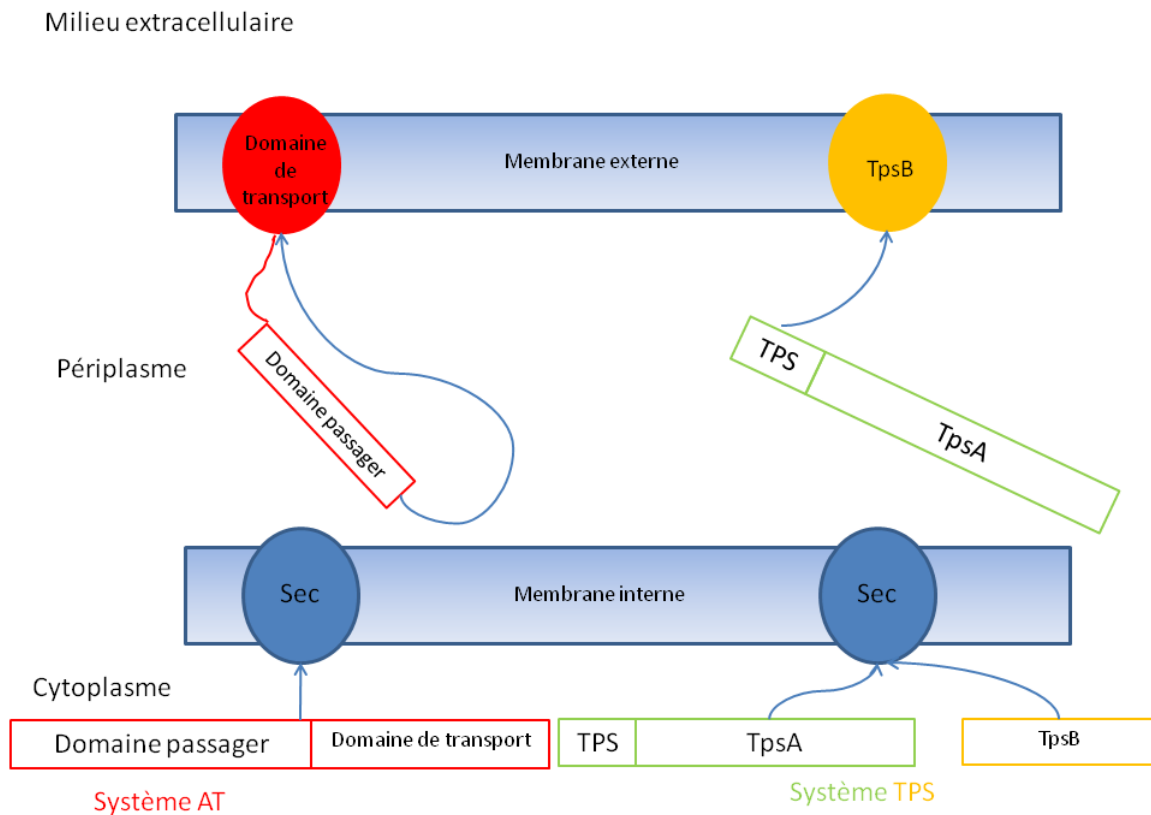


FIGURE 1.1 – Schéma du fonctionnement des systèmes TPS et AT. D'après [Jacob-Dubuisson 2004]

AT n'a pas besoin d'avoir un domaine de reconnaissance de son partenaire puisque celui-ci est attaché de manière covalente sur sa partie C-terminale, alors que dans le cas du système TPS, le partenaire sécrété TpsA doit avoir une séquence spécifique (appelée TPS) située dans sa partie N-terminale afin d'être reconnue par son transporteur TpsB [Renauld-Mongenie 1996]. Les modules de transport AT et TpsB forment généralement des tonneaux β qui viennent s'insérer dans la membrane [Clantin 2007] [Oomen 2004], et qui vont transporter leur partenaire en commençant par la partie N-terminale [Maurer 1999a] [Renauld-Mongenie 1996] pour le système TPS alors que le système AT fait une translocation de la partie C vers N-terminale [Ohnishi 1994]. Ces deux systèmes ont la particularité de requérir une conformation dépliée de la protéine sécrétée lors de la translocation [Clantin 2007] [Klauser 1990]. Les protéines sécrétées sont généralement des protéines de virulence, telles que des hémolysines [Hirono 1997], des adhésines [Benz 1992] ou encore des protéases [Pohlner 1987]. Les protéines TpsB ont les particularités d'être de très grande taille (supérieures à 2000 résidus) [Rojas 2002] [Locht 2001], et d'avoir des motifs de séquence se répétant [Kajava 2001].

Durant ma thèse, j'ai été amené à étudier un système TPS modèle : Le système FhaB/FhaC de *Bordetella pertussis*, que je vais décrire dans les prochains paragraphes.

1.1.1 Le système de transport FhaB/FhaC de *Bordetella pertussis*

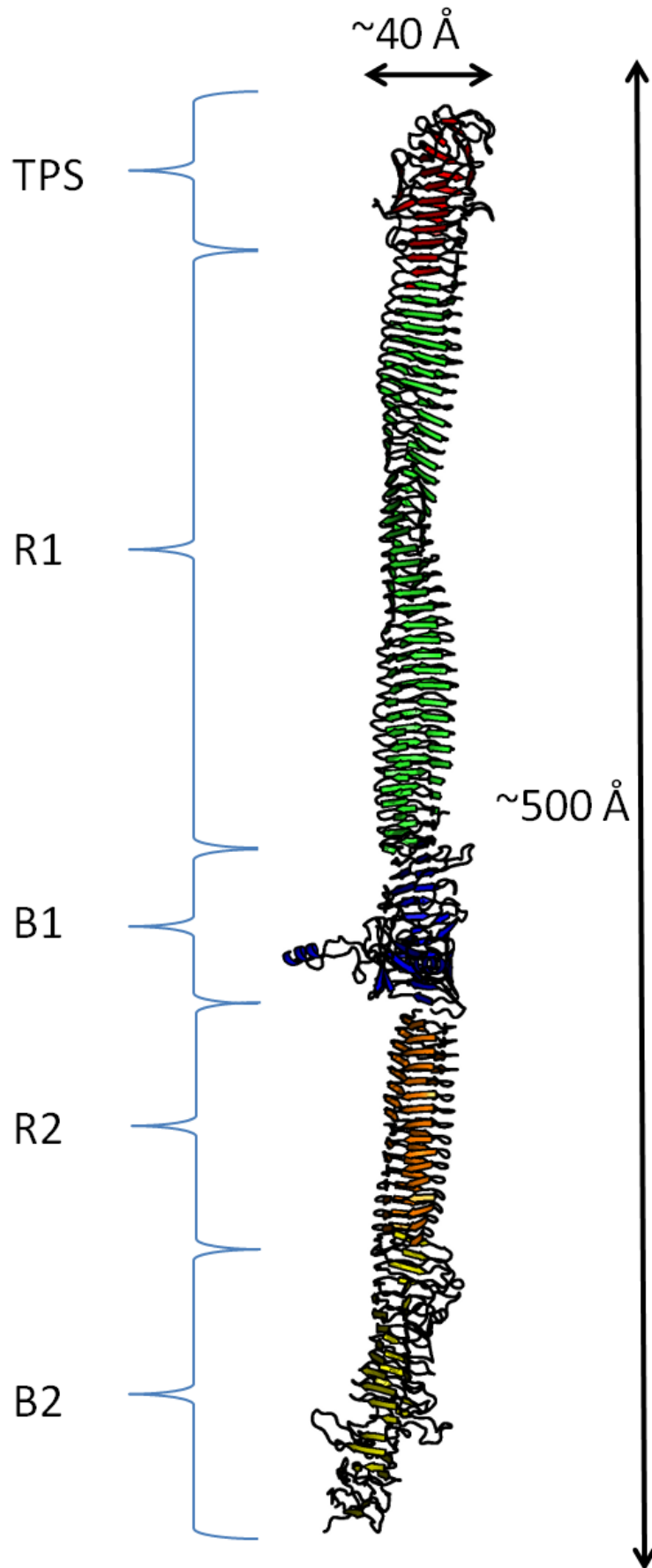
Bordetella pertussis est une bactérie à Gram-négatif responsable de la coqueluche. Cette bactérie sécrète une protéine appelée l'hémagglutinine filamenteuse (FHA), une adhésine de 230 kDa qui est essentielle pour la pathogénie de la bactérie car elle permet son adhésion sur les tissus de l'organisme hôte [Tuomanen 1985]. Elle comporte des sites de liaison à des motifs présents à la surface des cellules eucaryotes mais aussi procaryotes. Son rôle est de maintenir un lien mécanique entre le pathogène et le tissu infecté [Relman 1989] grâce à des sites de reconnaissance et de fixation à des motifs RGD [Relman 1989] et a des carbohydrates [Liu 1997]. Cette protéine est aussi très immunogène et donc largement utilisée dans les vaccins anti-coqueluche [Sato 1984].

Elle est sécrétée à la surface de *Bordetella pertussis* grâce au système TPS. Deux protéines sont donc produites : la protéine TpsA, appelée FhaB et la protéine TpsB, appelée FhaC. FhaB est d'abord produite sous la forme d'un très long précurseur de 367 kDa [Domenighini 1990], qui est d'abord transporté dans le périplasme par la machinerie Sec grâce a son peptide signal [Lambert-Buisine 1998]. Ensuite, FhaB traverse le périplasme dans une conformation dépliée [Meli 2006] pour être enfin sécrété dans le milieu extracellulaire par son partenaire FhaC. FhaB subit diverses étapes de maturation par protéolyse des parties N [Lambert-Buisine 1998] et C-terminales [Domenighini 1990].

Une fois sécrétée, FhaB adopte une forme de cylindre rigide [Makhov 1994] [Kajava 2001] qui peut être divisé en 4 parties (figure 1.2) :

- la partie N-terminale, appelée domaine TPS car ce domaine contient le motif de reconnaissance par le transporteur associé. Elle possède une structure en hélice β avec une coiffe dans la partie N-terminale [Clantin 2004].
- La partie R1, que les données structurales [Clantin 2004] et de modélisation [Kajava 2001] prédisent comme formant une hélice β . Elle contient des répétitions de 19 résidus en moyenne, qui sont compatibles avec la longueur d'un tour d'hélice.
- la partie B1 contient des éléments de reconnaissance aux carbohydrates [Liu 1997] et un motif de reconnaissance RGD [Relman 1989]. La structure de ce domaine n'a pas été résolue, mais il contient lui aussi des motifs répétitifs, supposant aussi une forme en hélice β avec une boucle contenant le motif de reconnaissance RGD [Kajava 2001]
- la partie R2 contient comme la partie R1 des motifs répétitifs prédits en hélice β [Kajava 2001].
- la partie B2 qui contient aussi des motifs répétitifs supposant une continuation de l'hélice dans la partie C-terminale [Kajava 2001].

FIGURE 1.2 – Modèle structural de FHA. La structure à été reconstruite à partir de la structure connue [Clantin 2004], de modèles générés par [Kajava 2001] et sur base de modélisation par méthode d'enfilage de séquence [Wu 2007]. Cette méthode permet de modéliser la structure tri-dimensionnelle d'une protéine à partir de la séquence seule. Le programme force la protéine à adopter tous les repliements virtuellement connus ("partie enfilage"), et calcule un score qui est une mesure quantitative de la compatibilité de la séquence avec ce repliement particulier.



Son partenaire associé FhaC, possède une structure en tonneau β inséré dans la membrane externe [Clantin 2007]. Il forme un pore permettant le passage d'un polypeptide sous forme d'une chaîne désordonnée [Jacob-Dubuisson 1999]. La reconnaissance de FhaB se fait grâce à des domaines de transport de polypeptides (PO-TRA) situées dans la partie périplasmique [Delattre 2011]. L'interaction entre ces domaines de FhaC et FhaB se fait pas l'intermédiaire d'interactions de type hydrophobe [Delattre 2011]. Le canal qui permet le transport à travers la membrane externe a un diamètre maximal de 8.2 Å, ce qui ne permet pas le transport d'une forme repliée de FhaB [Clantin 2007](figure 1.3).

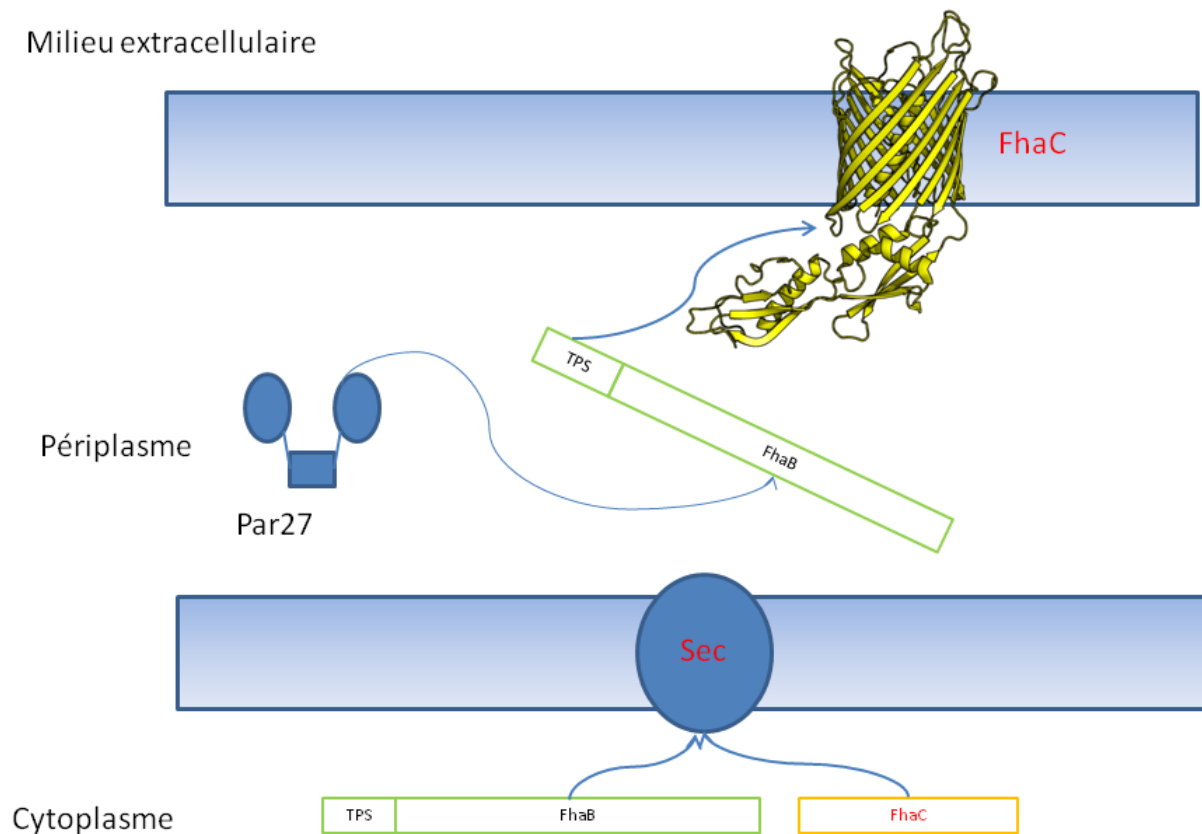


FIGURE 1.3 – Schéma illustrant le transport de FhaB dans le milieu extracellulaire.

FhaB doit donc être maintenu sous une forme dépliée dans le périplasme, ce qui nécessite des mécanismes pour le prémunir contre la dégradation par les protéases périplasmiques et l'agrégation. Plusieurs chaperonnes périplasmiques semblent as-

surer ce rôle, dont DegP [Baud 2009] et Par27 [Hodak 2008]. Le périplasme étant dépourvu d'ATP, le mécanisme de dissociation (permettant la prise en charge de FhaB par FhaC) entre ces chaperonnes et FhaB n'est pas connu.

Une partie de ma thèse s'articule autour de deux aspects de ce système :

- Le devenir de FHA, dans le périplasme. J'ai pour cela entrepris une étude structurale d'un fragment N-terminal de FHA de 30kDa, appelé Fha30, en complexe avec une chaperonne périplasmique, Par27. Le complexe ayant une flexibilité intrinsèque, j'ai opté pour une approche par diffusion aux petits angles, combinée à de la modélisation. Cette partie est présentée dans le chapitre 2.
- Les propriétés mécaniques de FHA. J'ai pour cela effectué une étude des propriétés dynamiques de Fha30 par diffusion incohérente de neutrons. J'ai aussi, avec des collaborateurs de l'institut de biologie de Lille (**Dr. Françoise Jacob-Dubuisson**) et de l'université catholique de Louvain-la-Neuve (**Dr. David Alsteens et Pr. Yves Dufrière**), réalisé une étude de son dépliement mécanique par microscopie à force atomique (AFM) et par dynamique moléculaire. Cette partie est présentée dans le chapitre 3

1.2 Le complexe de réplication des *Rhabdoviridae*

Les virus de la famille des *Rhabdoviridae* ont un spectre d'hôtes très large, dont les poissons, les crustacés, les insectes, les plantes, et les mammifères. Parmi ces virus, on compte notamment le virus de la rage qui est un important pathogène humain, mais aussi le virus de la stomatite vésiculaire. Le virus de la stomatite vésiculaire est un prototype des virus de la famille des *Rhabdoviridae*. Il infecte principalement le bétail, provoquant une maladie bénigne caractérisée par de la fièvre, une salivation excessive et des lésions vésiculaires sur les gencives, la langue, les pis et les sabots. Il est activement étudié, car il a un fort potentiel thérapeutique, notamment comme agent de vaccination [Lichty 2004] [Roediger 2008] ou encore comme agent oncolytique [Ebert 2003] [Shi 2009].

Leur génome est composé d'une molécule d'ARN de polarité négative servant de matrice pour la fabrication d'ARN messagers qui sont utilisés par la machinerie ribosomale pour produire les protéines virales. Ce génome code pour cinq protéines : la phosphoprotéine (P), la nucléoprotéine (N), l'ARN polymérase ARN dépendante (L), la protéine matrice (M) et la glycoprotéine (G).

Le complexe de réplication, composé de trois de ces protéines (P, L et N), est une composante essentielle du cycle viral (figure 1.4), car il est responsable de la multiplication du virus dans la cellule infectée. N encapside l'ARN viral [Albertini 2006], le maintenant dans une forme simple brin et le protégeant des nucléases et des récepteurs de l'immunité innée [Simpson 1966], alors que L et P constituent le complexe polymérasique portant l'activité transcription/réplication à proprement parler. Ce complexe utilise exclusivement le complexe N-ARN comme matrice et non l'ARN nu [Emerson 1975]. La P joue d'autres rôles essentiels dans la réplication virale en

chaperonnant la N [Mavrakis 2003] mais aussi en inhibant la réponse immunitaire [Blondel 2002]. Le cycle viral commence par l'adhésion du virus sur la surface de la

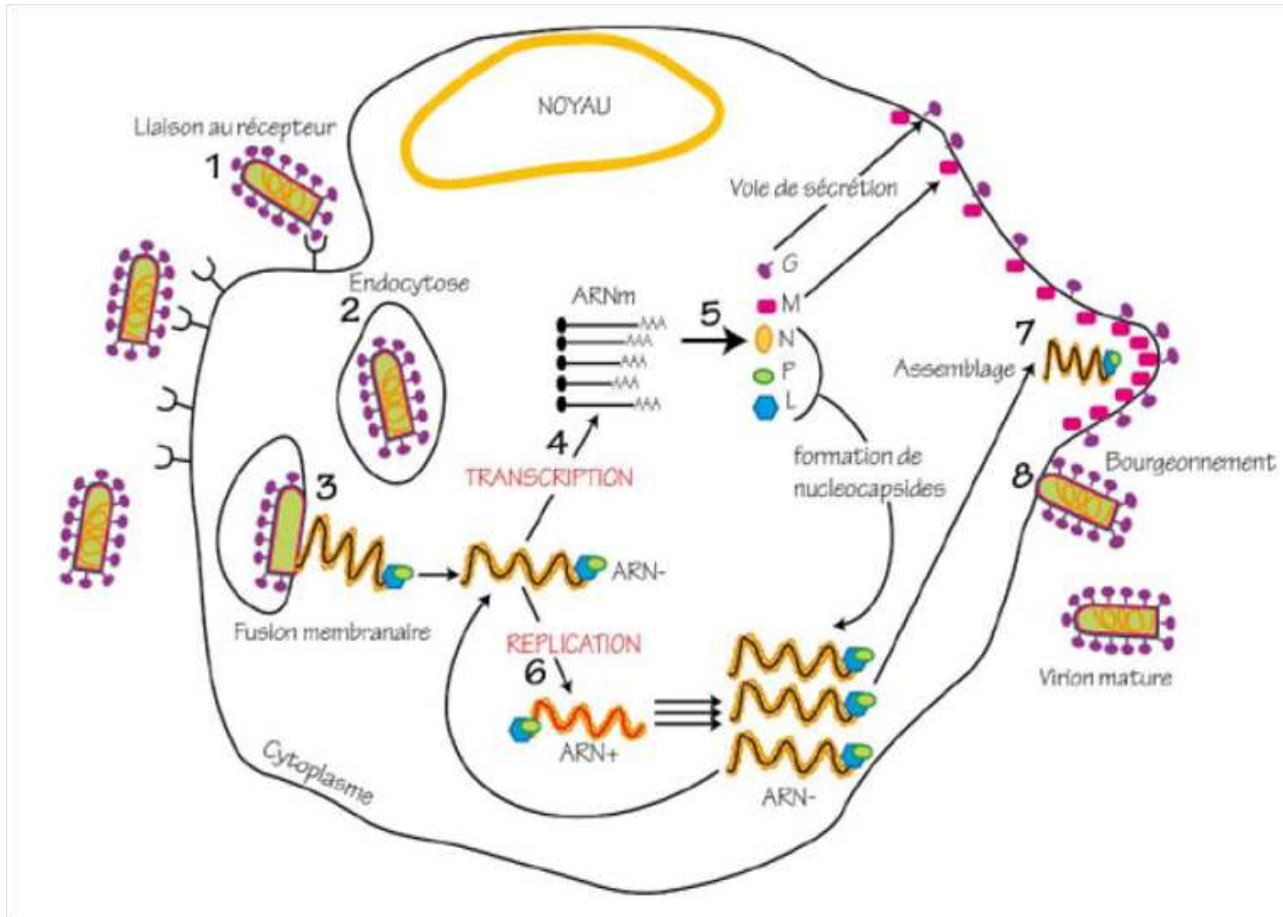


FIGURE 1.4 – Schéma illustrant le cycle viral des *Rhabdoviridae*. Le cycle commence par la liaison du virus aux récepteurs cellulaires (1). Ensuite, la particule virale est endocytée (2) et la fusion des membranes virales et cellulaires libère la machinerie répliquative dans le cytoplasme cellulaire (3). Le complexe de réplication va ensuite transcrire des ARNm (4), qui vont être traduits en protéines par la machinerie cellulaire (5). Le complexe de réplication va aussi répliquer le génome viral (6). Enfin, de nouveaux virions vont s'assembler près de la membrane cellulaire (7) et bourgeonner à la surface de la cellule (8).

cellule hôte, par interaction de G avec des composants de la surface de la cellule hôte [Matlin 1982]. La particule virale est ensuite internalisée par endocytose. L'acidification de l'endosome provoque ensuite des changements de conformation de la protéine G qui induisent la fusion des membranes endosomales et virales [White 1981], et donc la libération du complexe de réplication et de l'ARN viral encapsidé dans le cytoplasme de la cellule hôte.

Une fois que le complexe de réplication se retrouve dans la cellule hôte, le virus va entrer en phase de transcription : l'ARN viral va être transcrit en ARN messagers par le complexe L/P [Emerson 1975]. Cette étape est suivie par la traduction de ces ARN messagers par la cellule hôte, provoquant une production de protéines virales. La phase de réplication va ensuite débiter lorsque la quantité de ces protéines est suffisante. Le complexe de réplication va alors produire une molécule d'antigénome qui servira de matrice pour la production de nouveaux génomes viraux [Banerjee 1987]. Ces ARN néo synthétisés vont être immédiatement encapsidés par N [Patton 1984], formant le complexe N-ARN.

L'assemblage de nouveaux virions se fait par export du complexe N-ARN, de L et de P vers la membrane plasmique [Hunt 1976], où elles vont pouvoir prendre la forme hélicoïdale du virion grâce à M [Newcomb 1981] et bourgeonner à la surface de la cellule [Whitt 1989].

1.2.1 La phosphoprotéine du virus de la stomatite vésiculaire (VSV)

La protéine P est d'une importance capitale pour la réplication et la transcription du virus. Elle est un cofacteur de L indispensable pour la transcription et la réplication [Emerson 1975] [Green 2000]. Elle joue aussi un rôle de chaperonne de N, permettant à celle-ci de ne pas lier les ARN cellulaires [Paul 1988] [Ribeiro 2008] et la maintenant sous forme soluble en absence d'ARN en formant un complexe appelé N⁰-P [Mavrakis 2003] [Leyrat 2011b].

La protéine P forme des dimères en solution [Gerard 2009]. Et possède trois régions :

- La partie C-terminale, dont la structure à été résolue seule [Ribeiro 2008] ou en complexe avec un complexe N-ARN [Green 2006].
- La partie centrale qui est responsable de l'oligomérisation [Ding 2004] dont la structure est connue [Ding 2006].
- La partie N-terminale, qui est désordonnée, mais qui néanmoins forme des structures transitoires en hélice α [Leyrat 2011a]. Une partie de ce domaine se structure en hélice α lors de la formation du complexe N⁰-P [Leyrat 2011b].

Les parties centrales et C-terminales sont reliées entre elles par une courte région flexible [Gerard 2009]. La flexibilité est une caractéristique commune de certaines protéines impliquées dans l'assemblage de complexes macromoléculaires [Tompa 2002], ce qui est le cas de la protéine P qui contient des sites de fixation à L [Emerson 1987], à N-ARN [Chenik 1994] et à N⁰-P [Mavrakis 2006].

Dans ma thèse, je me suis intéressé à deux aspects de cette phosphoprotéine :

- Les propriétés structurales de cette protéine et du complexe que celle-ci forme avec la protéine N en absence d'ARN. Ce complexe, par la nature de P, est hautement flexible. Nous avons étudié ses propriétés structurales par des techniques de diffusion aux petits angles combinées avec une modélisation et une analyse d'ensembles. J'ai aussi réussi à cristalliser et résoudre la structure du domaine C-terminal de la phosphoprotéine du virus de la jaunisse nécrotyque par diffraction de rayons X. Cette partie est présentée dans le chapitre 4

- Les propriétés dynamiques de cette protéine. Cette protéine est un excellent modèle pour l'étude des propriétés dynamiques des protéines flexibles, car elle possède des parties désordonnées et des parties structurées. J'ai pu sonder ces aspects par la technique de diffusion de neutrons incohérente. Cette partie est présentée dans le chapitre 5

1.3 Techniques biophysiques

Dans les prochaines parties, j'exposerai les différentes techniques biophysiques que j'ai utilisées durant ma thèse. Je commencerai d'abord par la technique de diffusion aux petits angles : cette technique permet d'obtenir des informations structurales à basse résolution. Elle est idéale pour l'étude de grosses molécules qui ne peuvent être cristallisées ou étudiées par RMN. Aussi, la technique de variation de contraste est idéale pour l'étude de complexes multimoléculaires (voir partie 1.3.2.4). Ensuite, je continuerai par la technique de diffusion incohérente de neutrons, qui nous permet de sonder les mouvements des atomes d'hydrogène. Elle est idéale pour étudier les propriétés dynamiques des protéines.

Enfin, je finirai par les techniques de simulation de dynamique moléculaire, qui nous permet de simuler les mouvements des atomes dans des systèmes aussi complexes que les protéines.

1.3.1 Diffusion de neutrons

1.3.1.1 Section efficace et longueur de diffusion

Le neutron est une particule élémentaire de charge neutre, que l'on peut caractériser par son énergie :

$$E = \frac{1}{2}mv^2 \quad (1.1)$$

avec v sa vitesse et m sa masse (1.66×10^{-24} kg). On lui associe aussi une longueur d'onde en accord avec la formule de De Broglie :

$$\lambda = \frac{h}{mv} \quad (1.2)$$

avec h la constante de Planck. On peut donc définir un vecteur d'onde \mathbf{k} :

$$\mathbf{k} = \frac{m}{\hbar}\mathbf{v} \quad (1.3)$$

ou \hbar est la constante de Planck réduite ($\frac{h}{2\pi}$). Lorsqu'un neutron rencontre un échantillon trois processus vont pouvoir se produire :

- Le neutron traverse la matière sans que rien ne se passe (cas de la majorité des neutrons, leur probabilité d'interaction avec la matière est très faible).
- Il est absorbé par un noyau.
- Il est diffusé. Sa direction est alors changée, alors que son énergie peut être changée (diffusion inélastique) ou non (diffusion élastique) .

En analysant les changements de direction et d'énergie des neutrons diffusés, on accède aux propriétés structurales et dynamiques de notre échantillon.

Un paramètre important est la section efficace σ . Cette grandeur exprime l'intensité totale diffusée I_d (en neutrons.cm⁻²) en fonction de l'intensité du faisceau incident I_0 (en neutrons.s⁻¹) :

$$I_d = \sigma I_0 \quad (1.4)$$

La section efficace est donc une grandeur équivalente à une surface. Elle représente la fraction de la surface du faisceau incident qui va être diffusée. On l'exprime généralement en barns (1 barn = 10⁻²⁴ cm²), car la probabilité d'interaction avec la matière est très faible. Le tableau 1.1 présente les différentes sections efficaces d'éléments composants les protéines. Les neutrons interagissent avec les noyaux par des forces nucléaires ou par interaction magnétique. Dans les systèmes biologiques, l'interaction magnétique est négligeable devant l'interaction nucléaire. L'interaction neutron-noyau se fait par l'intermédiaire d'un pseudopotentiel de Fermi $V(\mathbf{r})$. Parce que la taille des noyaux est très inférieure à la longueur d'onde des neutrons, la diffusion est isotrope et le potentiel dépend donc d'un seul paramètre b , la longueur de diffusion :

$$V(\mathbf{r}) = \frac{2\pi\hbar^2}{m} b_i \delta(\mathbf{r} - \mathbf{r}_i) \quad (1.5)$$

avec \mathbf{r} la position du neutron, \mathbf{r}_i la position du noyau i de longueur de diffusion b_i . Cette longueur de diffusion est reliée à la section efficace de diffusion $\sigma = 4\pi b^2$.

1.3.1.2 Diffusion cohérente et incohérente

Un échantillon est composé d'une multitude de noyaux différents qui sont eux même hétérogènes de par la présence d'isotopes. La longueur de diffusion b_i ne dépend pas uniquement de l'espèce considérée, mais change également en fonction de l'isotope parce que l'interaction neutron/noyau dépend de la nature du noyau (incohérence isotopique) et du spin total du système neutron-noyau (incohérence de spin). Pour cette raison, on peut scinder les longueurs de diffusion en deux parties : une partie *cohérente* et une partie *incohérente*. La partie cohérente est la moyenne des b_i alors que la partie incohérente est la déviation par rapport à cette moyenne :

$$b_{coh} = \langle b_i \rangle \quad (1.6)$$

$$b_{inc} = \sqrt{\langle b_i^2 \rangle - \langle b_i \rangle^2} \quad (1.7)$$

Le tableau suivant nous donne les valeurs des longueurs de diffusion et des éléments les plus abondants constituant les protéines.

On peut noter deux choses :

- La section efficace incohérente de l'hydrogène est un ordre de grandeur plus élevé que celle des autres éléments. Les spectres de diffusion incohérente seront donc dominés par cette contribution.

- Diffusion cohérente: informations sur la **structure** de l'échantillon
- Diffusion incohérente: informations sur la **dynamique** de l'échantillon

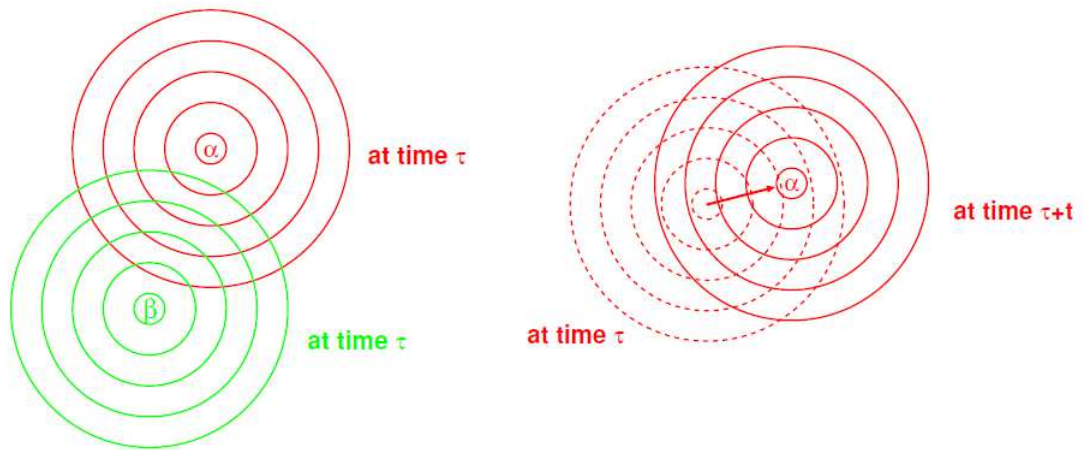


FIGURE 1.5 – Schéma illustrant la différence entre la diffusion cohérente et incohérente.

Element	b_{coh} (10^{-15} m)	b_{inc} (10^{-15} m)	σ_{coh} (barns)	σ_{inc} (barns)	σ_{tot} (barns)
H	-3.74	25.21	1.76	79.91	81.67
D	6.67	4.03	5.60	2.04	7.64
C	6.65	0	5.55	0	5.55
N	9.37	1.98	11.01	00.49	11.52
O	5.81	0	4.24	0	4.24

TABLE 1.1 – Sections efficaces et longueurs de diffusion des principaux éléments composant les protéines [Bée 1988].

- La longueur de diffusion cohérente de l'hydrogène est négative alors que celle du deutérium est positive. Cette différence est à la base de la technique de variation de contraste en diffusion de neutrons aux petits angles.

La diffusion cohérente donne lieu à des interférences entre ondes diffusées par des atomes différents : elle nous donne une information sur la **structure** de notre échantillon. La diffusion incohérente donne lieu à des interférences entre ondes diffusées

par un même atome à des temps différents : elle nous informe sur la **dynamique** de notre système (figure 1.5).

1.3.2 Diffusion aux petits angles

La diffusion aux petits angles est une technique exploitant la diffusion *cohérente* d'un échantillon, elle nous donne donc une information structurale. Elle est basée sur des phénomènes d'interférence. Le principe est de mesurer l'intensité en fonction du transfert de vecteur d'onde \mathbf{Q} défini par :

$$\mathbf{Q} = \mathbf{k}' - \mathbf{k} \quad (1.8)$$

où \mathbf{k} et \mathbf{k}' sont les vecteurs d'onde des faisceaux incidents et diffusés respectivement. On relie la norme de \mathbf{Q} à l'angle de diffusion par :

$$|\mathbf{Q}| = \frac{4\pi \sin\theta}{\lambda} \quad (1.9)$$

ou θ est l'angle de diffusion et λ la longueur d'onde utilisée (figure 1.6). La mesure

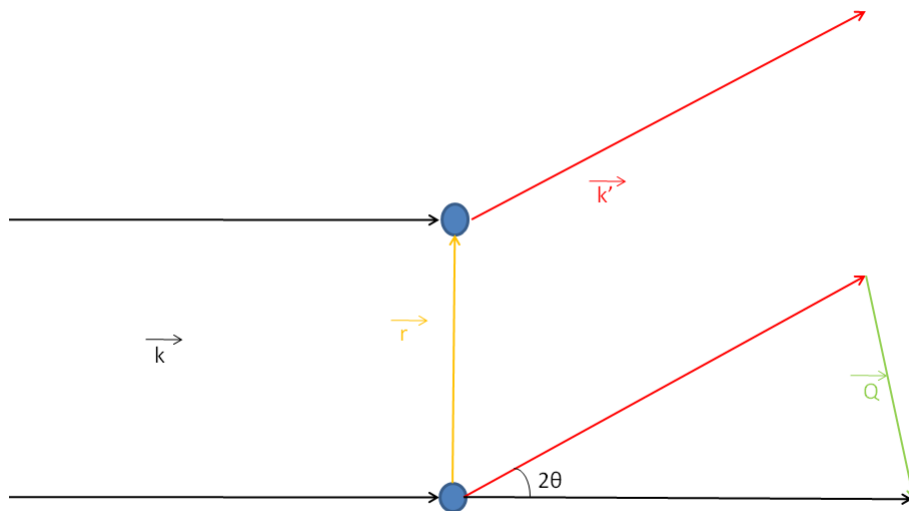


FIGURE 1.6 – Schéma illustrant la géométrie de la diffusion.

se fait sur des angles très petits, qui correspondent à des valeurs de $|\mathbf{Q}|$ entre 0.001 \AA^{-1} et 0.6 \AA^{-1} , ce qui veut dire que, contrairement à la cristallographie, elle nous donne des informations à **basse résolution**.

Pour un système à plusieurs noyaux, on peut d'abord définir une densité de diffusion $\rho(\mathbf{r})$:

$$\rho(\mathbf{r}) = \sum_i b_i \delta(\mathbf{r} - \mathbf{r}_i) \quad (1.10)$$

avec b_i et \mathbf{r}_i la longueur de diffusion et la position de l'atome i respectivement. Cette fonction nous permet de définir le facteur de structure :

$$F(\mathbf{Q}) = \int \rho(\mathbf{r}) e^{i\mathbf{Q}\cdot\mathbf{r}} d\mathbf{r}^3 \quad (1.11)$$

On peut passer au cas discret où b_j est la longueur de diffusion correspondant à la position définie par le vecteur \mathbf{r}_j . L'équation 1.11 devient alors :

$$F(\mathbf{Q}) = \sum b_j e^{i\mathbf{Q}\cdot\mathbf{r}_j} \quad (1.12)$$

La quantité que l'on mesure est l'intensité, que l'on peut exprimer à l'aide du facteur de structure :

$$I \propto F(\mathbf{Q}) \overline{F(\mathbf{Q})} \quad (1.13)$$

avec $\overline{F(\mathbf{Q})}$ le conjugué complexe de $F(\mathbf{Q})$.

Les mesures se font en solution, l'orientation des particules est donc arbitraire. On peut donc faire une moyenne angulaire dénotée par $\langle \dots \rangle$:

$$I(Q) \propto \langle F(\mathbf{Q}) \overline{F(\mathbf{Q})} \rangle \quad (1.14)$$

Le terme exponentiel devient alors :

$$\langle e^{i\mathbf{Q}\cdot\mathbf{r}} \rangle = \frac{\sin Qr}{Qr} \quad (1.15)$$

On obtient donc la formule de Debye [Debye 1915] :

$$I(Q) = \sum_j \sum_k b_j b_k \frac{\sin Qr_{jk}}{Qr_{jk}} \quad (1.16)$$

où r_{jk} est la distance entre les points définis par les vecteurs de position \mathbf{r}_j et \mathbf{r}_k . Une fonction intéressante que l'on peut obtenir à partir des données expérimentales est :

$$p(r) = \frac{2r^2}{\pi} \int Q^2 I(Q) \frac{\sin Qr}{Qr} dQ \quad (1.17)$$

qui n'est autre que la fonction de distribution de paires. Elle représente la probabilité de trouver deux centres diffuseurs à une certaine distance r . Ces équations nous permettent de calculer la courbe de diffusion théorique d'une structure à partir des longueurs de diffusion atomiques [Svergun D.I. 1995], ou encore une enveloppe à basse résolution de notre molécule [Franke 2009].

1.3.2.1 Instrumentation

Les mesures de diffusion de neutrons aux petits angles présentées dans ce manuscrit ont été effectuées sur l'instrument D22 de l'ILL. IL est schématisé dans la figure 1.7.

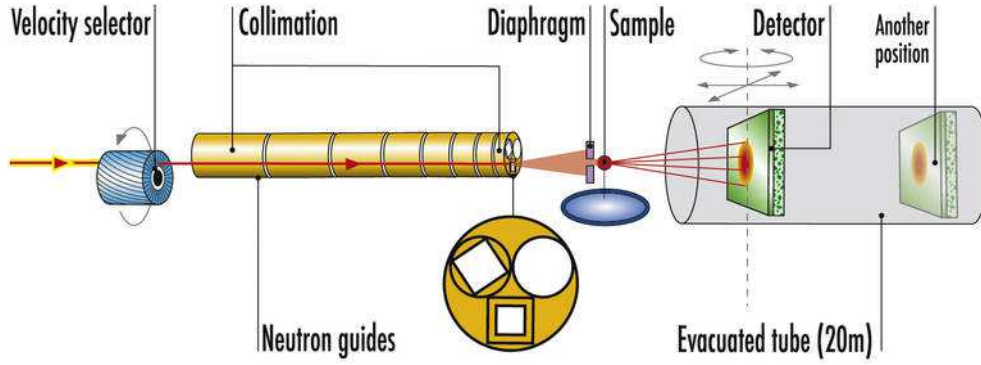


FIGURE 1.7 – Schéma présentant les différentes parties de l'instrument D22 de l'ILL.

Les neutrons sont d'abords sélectionnés en fonction de leur vitesse, ce qui permet de sélectionner leur longueur d'onde grâce aux relations 1.1 et 1.2. Ils passent à travers un guide qui sert de collimateur et permet donc de réduire la divergence du faisceau. L'échantillon (contenu dans une cuve en quartz) est placé dans un porte échantillon permettant un contrôle de la température. Les neutrons sont comptés par un détecteur de surface situé dans un tube sous vide. Une fois les données collectées en deux dimensions, on effectue une intégration radiale pour avoir l'intensité en fonction de l'angle de diffusion.

1.3.2.2 Rayon de giration et intensité à angle zéro

Pour étudier le comportement aux très petits angles, on peut développer le terme $\frac{\sin Qr}{Qr}$ de la loi de Debye en série de McLaurin :

$$\frac{\sin Qr}{Qr} = 1 - \frac{Q^2 r^2}{6} + \frac{Q^4 r^4}{120} - \dots \quad (1.18)$$

En remplaçant dans 1.16 et en négligeant les termes d'ordre supérieur à 2, on obtient :

$$\sum_j \sum_k b_j b_k \left(1 - \frac{Q^2 r_{jk}^2}{6}\right) = \sum_j \sum_k b_j b_k - \frac{1}{6} Q^2 \sum_j \sum_k b_j b_k r_{jk}^2 \quad (1.19)$$

On peut alors définir deux paramètres :

$$I(0) = \sum_j \sum_k b_j b_k = \left(\sum_j b_j\right)^2 \quad (1.20)$$

qui est la somme des longueurs de diffusion au carré, et

$$R_g^2 = \frac{1}{2} \frac{\sum_j \sum_k b_j b_k r_{jk}^2}{\sum_j \sum_k b_j b_k} \quad (1.21)$$

ou si l'on choisit le centre de masse de la molécule comme référence :

$$R_g^2 = \frac{1}{2} \frac{\sum_j b_j r_j^2}{\sum_j b_j} \quad (1.22)$$

On peut voir intuitivement qu'à angle nul, toutes les ondes sont en phase. On somme donc le carré des amplitudes de toutes les ondes diffusées. R_g est appelé le rayon de giration qui est l'équivalent du rayon de giration utilisé en mécanique. On a donc 1.19 qui devient

$$I(Q)(1 - \frac{1}{3} R_g^2 Q^2) \quad (1.23)$$

où l'on reconnaît aisément le développement en série de McLaurin de la fonction exponentielle, on a donc :

$$I(Q) \approx I(0) e^{-\frac{1}{3} R_g^2 Q^2} \quad (1.24)$$

En portant le logarithme de l'intensité en fonction de Q^2 , on peut obtenir R_g à partir de la pente, et $I(0)$ avec l'ordonnée à l'origine. Cette formulation nous vient de Guinier [Guinier 1955]. Elle est valable pour $R_g^2 Q^2 < 2$, pour les objets ayant une forme globulaire [Kratky 1982]. Expérimentalement, $\sqrt{I(0)}/C$ est proportionnel à la somme des amplitudes au carré (C étant la concentration de notre échantillon, il nous permet donc de calculer la masse moléculaire absolue et de définir l'état d'oligomérisation de notre protéine, en faisant d'abord une calibration avec une protéine de masse connue [Yeager 1976] ou avec la diffusion incohérente du solvant [Jacrot 1981].

R_g est un paramètre qui dépend de la structure de la protéine. Pour deux protéines de masse égale, un R_g plus grand signifiera que la molécule est répartie dans un espace plus vaste autour de son centre de masse.

1.3.2.3 Représentation de Kratky

Un polymère désordonné est caractérisé par une fonction de distribution d'équation [Flory 1953] :

$$P(r) = \left(\frac{3}{12\pi R_g^2}\right)^{3/2} e^{-\frac{3r^2}{12R_g^2}} \quad (1.25)$$

En injectant cette expression dans 1.16 [Kratky 1982] :

$$I(Q) = \frac{2((QR_g)^2 - 1 + e^{-(QR_g)^2})}{(QR_g)^2} \quad (1.26)$$

On obtient donc une expression caractéristique, avec une courbe ayant un comportement asymptotique lorsqu'on l'exprime sous la forme $I(Q)Q^2$ en fonction de Q^2 . Une protéine globulaire une forme de courbe en cloche [Koch 2003]. Cette représentation nous permet donc de savoir si notre protéine est désordonnée ou structurée.

1.3.2.4 Variation de contraste

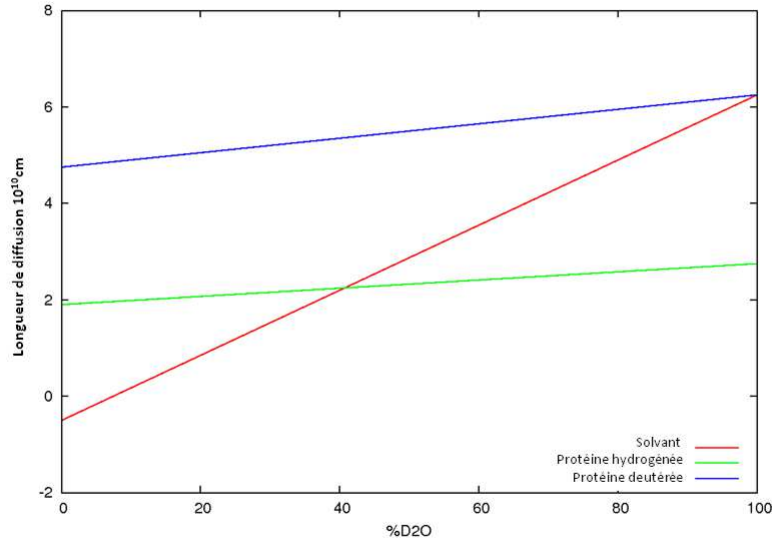


FIGURE 1.8 – Longueurs de diffusion de différents composants en fonction du pourcentage de D₂O. On voit qu'à 40% D₂O, la longueur de diffusion de la protéine-H devient égale à celle du solvant : celle-ci devient "invisible". Le même phénomène a lieu à 100% D₂O pour la protéine-D. Adapté de [Serdyuk 2007]

Pour une solution de N particules identiques n'interagissant pas entre-elles, l'intensité mesurée est due au contraste des particules par rapport au solvant. Le solvant peut être considéré comme un milieu ayant une amplitude de diffusion constante $\bar{\rho}$. On a donc la formule de Debye 1.16 qui devient :

$$I(Q) = N \sum_j \sum_k m_j m_k \frac{\sin Q r_{jk}}{Q r_{jk}} \quad (1.27)$$

avec $m_j m_k$:

$$m_j m_k = (b_j - \bar{\rho})(b_k - \bar{\rho}) \quad (1.28)$$

L'intensité à angle zéro devient alors :

$$I(0) = ((\rho - \bar{\rho})V)^2 \quad (1.29)$$

V étant le volume de la particule et ρ le contraste moyen de la particule. La technique de variation de contraste s'avère très intéressante lorsque l'on utilise un rayonnement neutronique : en effet, les amplitudes de diffusion des isotopes de l'hydrogène sont très différentes (voir tableau 1.1). On peut donc dans certaines proportions de $\text{H}_2\text{O}/\text{D}_2\text{O}$ avoir le terme $\rho(\mathbf{r}) - \bar{\rho}$ qui devient nul : l'objet ne contribue plus à la diffusion. Cette technique est très intéressante pour étudier des complexes protéines-protéines, car on peut produire une version deutérée d'une protéine et former un complexe avec une protéine non deutérée. A certains rapports $\text{H}_2\text{O}/\text{D}_2\text{O}$, on va donc faire "disparaître" la protéine-D ou la protéine-H (figure 1.8).

Expérimentalement, on enregistre des courbes de diffusion d'un complexe protéine-D/protéine-H à différentes proportions $\text{H}_2\text{O}/\text{D}_2\text{O}$. On écarte ensuite tout effet d'aggrégation en portant la racine carrée de l'intensité à angle nul en fonction du contraste. Ce graphe, appelé graphe de Sturhmann, nous permet de vérifier qu'il y a bien une dépendance linéaire entre le contraste et la racine carrée de l'intensité à angle nul (équation 1.29).

1.3.2.5 Analyse d'ensembles

Si l'on possède une solution de molécules ayant plusieurs conformations i , comme par exemple une protéine désordonnée, on a :

$$I(Q) \propto \sum_i N_i |F_i(Q)|^2 \quad (1.30)$$

avec $F_i(Q)$ le facteur de structure de la molécule dans la conformation i en N_i copies en solution. On peut donc exprimer l'intensité diffusée d'une molécule ayant plusieurs conformations comme la somme des facteurs de structures au carré de ses différents conformères pondérés par la probabilité de les trouver en solution.

Pour la courbe de diffusion d'une protéine désordonnée ou flexible, on peut donc générer plusieurs conformères (de manière aléatoire ou par un calcul de dynamique moléculaire), et faire une sélection afin d'avoir un nombre restreint de modèles permettant d'expliquer la courbe de diffusion. C'est un problème d'optimisation : trouver la somme des $N_i |F_i(Q)|^2$ qui minimise :

$$\chi^2 = \sum \left(\frac{I_{exp}(Q) - I_{the}(Q)}{\sigma(Q)} \right)^2 \quad (1.31)$$

avec $I_{exp}(Q)$ l'intensité expérimentale, $I_{the}(Q)$ l'intensité théorique calculée et $\sigma(Q)$ la racine carrée de la variance de la mesure expérimentale. Il existe différentes méthodes, et j'ai utilisé dans ma thèse l'approche de [Bernado 2007] : le problème d'optimisation est résolu en utilisant un algorithme génétique. Le principe consiste à générer un grand nombre de modèles de manière aléatoire ou à l'aide d'un calcul de dynamique moléculaire, de calculer $|F_i(Q)|^2$ pour chaque modèle et de résoudre le problème d'optimisation. On obtient ainsi un nombre restreint de modèles qui représentent les différents états conformationnels de la molécule.

1.3.3 Diffusion incohérente de neutrons

1.3.3.1 Facteur de structure dynamique

Une expérience de spectroscopie neutronique mesure le nombre de neutrons diffusés dans un angle solide $d\Omega$ défini par le vecteur d'onde diffusé et dans un intervalle d'énergie transmis d'énergie $d\omega$ entre les neutrons diffusés et incidents. La quantité mesurée s'appelle la double section efficace différentielle qui s'exprime :

$$\frac{\partial^2 \sigma}{\partial \omega \partial \Omega} = \frac{k}{k'} \frac{\sigma_{coh} + \sigma_{inc}}{4\pi} S(\mathbf{Q}, \omega) \quad (1.32)$$

avec σ_{coh} et σ_{inc} les sections efficaces cohérentes et incohérentes, et $S(\mathbf{Q}, \omega)$ le facteur de structure dynamique. Cette fonction est en fait la double transformée de Fourier d'une fonction de corrélation de Van Hove :

$$S(\mathbf{Q}, \omega) = \int F(\mathbf{Q}, t) e^{-i\omega t} dt \quad (1.33)$$

avec $F(\mathbf{Q}, t)$ la fonction de diffusion intermédiaire :

$$F(\mathbf{Q}, t) = \int G(\mathbf{r}, t) e^{-i\mathbf{Q}\mathbf{r}} d\mathbf{r} \quad (1.34)$$

et $G(\mathbf{r}, t)$ est la fonction de corrélation de Van Hove qui peut être scindée en deux parties : une partie de corrélation de paires G_d et une partie d'autocorrélation G_s . Le facteur de structure dynamique peut être divisée en deux parties : une partie cohérente qui est la double transformée de Fourier de G_d et une partie incohérente qui est la double transformée de Fourier de G_s :

$$G_s(\mathbf{r}, t) = \frac{1}{N} \sum_i \int \delta(\mathbf{r} - \mathbf{r}' + \mathbf{r}_i(0)) \delta(\mathbf{r}' - \mathbf{r}_i(t)) d\mathbf{r}' \quad (1.35)$$

avec $\mathbf{r}_i(t)$ la position de l'atome i à l'instant t et N le nombre d'atomes. cette fonction exprime la probabilité d'avoir une particule se trouvant à la position \mathbf{r} à l'instant t , alors qu'elle était à l'origine à l'instant $t = 0$. Le facteur de structure dynamique devient alors :

$$S(\mathbf{Q}, \omega) = \frac{\sigma_{inc}}{\sigma_{inc} + \sigma_{coh}} S_{inc}(\mathbf{Q}, \omega) + \frac{\sigma_{coh}}{\sigma_{inc} + \sigma_{coh}} S_{coh}(\mathbf{Q}, \omega) \quad (1.36)$$

et donc la double section efficace différentielle peut s'exprimer sous la forme :

$$\frac{d^2 \sigma}{d\omega d\Omega} = \frac{k}{k'} \left[\frac{\sigma_{coh}}{4\pi} S_{coh}(\mathbf{Q}, \omega) + \frac{\sigma_{inc}}{4\pi} S_{inc}(\mathbf{Q}, \omega) \right] \quad (1.37)$$

L'hydrogène est un élément extrêmement abondant dans les protéines (environ 50%), et sa section efficace incohérente dépasse largement celle des autres atomes (table 1.1). De plus, les hydrogènes sont distribués de manière homogène dans les protéines et constituent donc un outil idéal pour mesurer les propriétés dynamiques des protéines.

1.3.3.2 Diffusion élastique incohérente

L'intensité élastique totale nous informe sur les mouvements localisés des atomes. Lorsqu'un atome reste confiné dans un certain volume, qui est défini par la gamme de Q , et dans une certaine gamme de temps définie par la résolution instrumentale, il va contribuer à la diffusion élastique.

A un temps infiniment grand, les positions atomiques ne sont plus corrélées à leur positions de départ, et 1.34 devient donc :

$$F_{inc}(\mathbf{Q}, \infty) = \frac{1}{N} \sum |\langle e^{i\mathbf{Q}\mathbf{r}_i} \rangle|^2 \quad (1.38)$$

avec \mathbf{r}_i les vecteurs de position des particules. On peut donc définir une approximation similaire à celle faite par Guinier pour le cas de la diffusion aux petits angles :

$$S(Q, 0) \approx e^{-\frac{1}{3}\langle u^2 \rangle Q^2} \quad (1.39)$$

qui est appelée l'approximation Gaussienne ??, où $\langle u^2 \rangle$ représente les déplacements carrés moyens des atomes (msd). Comme pour l'approximation de Guinier, celle-ci possède une contrainte spatiale : $\langle u^2 \rangle Q^2 < 2$.

En pratique, on ne mesure pas $S(Q, 0)$, car les instruments ont une résolution énergétique ΔE , c'est à dire qu'ils peuvent mesurer $E \pm \Delta E$, et que les déplacements carrés moyens mesurés dépendent du temps. On définit donc une fenêtre d'espace (défini par la gamme de Q) temps (défini par ΔE) spécifique de chaque instrument, qui nous donnera un signal incohérent élastique. Les deux instruments que j'ai utilisé durant ma thèse, IN13 et IN16, ont une résolution de $8 \mu\text{eV}$ et $0.9 \mu\text{eV}$ respectivement, qui correspondent à des temps caractéristiques de $\approx 1\text{ns}$ et $\approx 100\text{ps}$.

1.3.3.3 Instrumentation

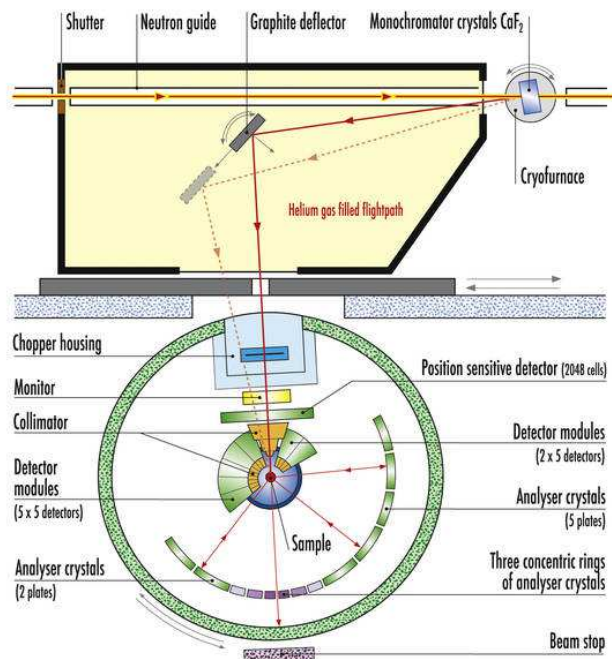
Les mesures de diffusion incohérente élastique présentées dans cette thèse ont été effectuées sur deux instruments : IN13 et IN16 de l'ILL. Ces deux instruments sont des spectromètres à rétrodiffusion, leur schémas sont représentés dans la figure refIN13+IN16. Le principe de la rétrodiffusion consiste à mesurer les neutrons diffractés par un cristal à des angles proches de 180° . On peut alors différencier la loi de Bragg :

$$\lambda = 2d \sin \theta \quad (1.40)$$

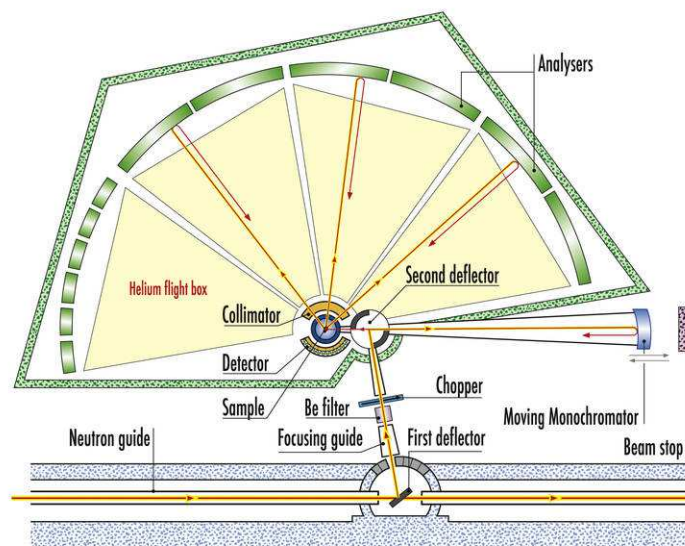
avec λ la longueur d'onde, d l'espacement entre les plans cristallins et θ l'angle de diffusion. Qui devient, une fois différenciée :

$$\frac{\delta \lambda}{\lambda} = \frac{\delta d}{d} + \frac{\Delta \theta}{\tan \theta} \quad (1.41)$$

En conditions de rétrodiffusion ($2\theta = 180^\circ$) le deuxième terme à droite de 1.41 devient nul, ce qui nous permet d'avoir une résolution optimale. Le principe consiste à



(a) Schéma du spectromètre IN13 de l'ILL



(b) Schéma du spectromètre IN16 de l'ILL

FIGURE 1.9 – Schémas de fonctionnement des spectromètres IN13 et IN16.

amener un faisceau monochromatique sur l'échantillon. Les neutrons vont alors être diffusés dans un certain angle et à une certaine énergie qui dépend de la dynamique de notre échantillon. Les neutrons vont être rétrodiffusés par des analyseurs à des angles proches de 180° (condition de rétrodiffusion), repasser dans l'échantillon et être détectés derrière celui-ci. La condition de rétrodiffusion nous permet de mesurer que les neutrons ayant une certaine énergie, qui dans le cas de la résolution élastique est l'énergie des neutrons incidents

1.3.3.4 Evolution des déplacements carrés moyens en fonction de la température et résilience moyenne

Pour un système fluctuant autour d'une position d'équilibre, on peut utiliser un modèle harmonique. Il a été démontré que pour les protéines en dessous de 200K, le déplacement carré moyen des hydrogènes pouvait être modélisé de cette manière [Doster 1989].

Dans le cadre d'un mouvement purement harmonique, la force de rappel $F(x)$ exercée sur le point déplacé d'une longueur x est donnée par la loi de Hooke :

$$F(x) = -kx \quad (1.42)$$

où k est la constante de force du ressort. Par analogie, on peut définir aussi une constante de force effective $\langle k' \rangle$ [Zaccai 2000] pour décrire la dépendance des

déplacements carrés moyens à la température :

$$\langle k' \rangle = 2k_B \left(\frac{d \langle u^2 \rangle}{dT} \right)^{-1} \quad (1.43)$$

où k_B est la constante de Boltzmann. Cette constante de force harmonique (en dessous de 200K) ou quasi-harmonique (au dessus de 200K) nous renseigne sur la rigidité de notre protéine et a déjà été utilisée pour caractériser de manière qualitative différentes protéines [Gabel 2002]. Il est intéressant de noter que les valeurs de $\langle k' \rangle$ mesurées sont du même ordre de grandeur que les forces que l'on mesure en microscopie à force atomique lors du dépliement mécanique.

1.3.4 Simulations de dynamique moléculaire

La dynamique moléculaire consiste à calculer les mouvements des atomes en résolvant les équations du mouvement de Newton :

$$\mathbf{F}(x) = -\nabla U(x) = m \frac{d\mathbf{v}}{dt} \quad (1.44)$$

avec \mathbf{F} la force, $U(x)$ l'énergie potentielle (définie par le champ de force que nous verrons par la suite), m la masse, \mathbf{v} la vitesse et t le temps. Le but est donc de connaître les position et vitesses des atomes au cours du temps pour déterminer la trajectoire des atomes d'une molécule. Pour cela, il est nécessaire de définir une structure de départ (disponible dans la PDB si la structure est résolue), et un champ de force, qui définit l'énergie potentielle qui définit les interactions entre atomes. Les équations du mouvement sont ensuite intégrées par un algorithme en utilisant un pas de temps (typiquement, 2 fs), et les positions et vitesses sont calculées à tous les pas de temps.

Ces calculs sont complexes et longs à cause du grand nombre d'interactions qui sont mises en jeu. Le premier calcul de dynamique moléculaire d'une macromolécule à été effectué en 1977 [McCammon 1977] sur l'inhibiteur de la trypsine pancréatique bovine. Le temps de calcul augmentant avec la taille du système et la durée de la simulation, l'augmentation de la puissance des micro-processeurs ainsi que des nouveaux algorithmes optimisés [Phillips 2005] [Hess 2008] permettent maintenant de simuler des systèmes plus complexes et sur des durées plus longues. Deux exemples traduisent ces avancées : une modélisation du virus entier de la mosaïque du tabac durant 50 ns [Freddolino 2006], qui a permis de montrer que l'ARN était essentiel pour la stabilité du virion. Un autre exemple représentatif est la simulation d'un domaine de la villine durant 500 μ s [Zagrovic 2002] où l'on a simulé le repliement de ce domaine. La simulation a permis de confirmer le fait que la formation du coeur hydrophobe intervient très tôt dans le mécanisme de repliement. Plus récemment, Shaw et ses collaborateurs ont simulé BPTI (inhibiteur de la trypsine pancréatique bovine) sur une milliseconde [Shaw 2010], et ont pu ainsi caractériser complètement la dynamique de cette protéine.

Les temps de calcul étant considérables, on peut utiliser des modèles ayant une représentation simplifiée afin de les réduire.

1.3.4.1 Modèles simplifiés

Afin de réduire les temps de calcul, des efforts ont été mis en oeuvre pour permettre l'élaboration de nouveaux modèles avec des champs de force simplifiés.

La majeure partie du temps de calcul est dédiée au calcul des interactions longue distance, il est donc très avantageux de réduire le nombre d'atomes (réduction du nombre de degrés de liberté du système) mais aussi de simplifier les calculs en approximant les interactions.

Les premiers modèles de ce type ont été introduits par Go [Taketomi 1975], [Go 1983], qui a proposé de représenter chaque résidu par une sphère. Les interactions du système ne prennent en compte que les contacts dans la forme native. Ces modèles ont été utilisés avec succès pour de nombreuses études de repliement, car ils permettent d'avoir un calcul de durée raisonnable pour observer un processus relativement lent. On peut citer entre autres les études de Onuchic qui a réussi avec ce type de modèle à caractériser la structure des intermédiaires de repliement de la dihydrofolate reductase et de l'interleukine-1 β [Clementi 2000a]. Ce champ de force a été utilisé dans mon travail [Clementi 2000b]. Chaque résidu est représenté par son Carbone α . Le champ de force favorise les interactions natives (qui sont présentes dans la forme repliée et fonctionnelle de notre molécule), et défavorise les interactions non-natives. Ces modèles ont pour effet de lisser la surface d'énergie potentielle et sont utilisés pour comprendre les phénomènes de repliement/dépliement qui mettent en jeu des changements sur des échelles de temps plus longues.

1.3.5 Microscopie à force atomique

1.3.5.1 Dépliement mécanique

La microscopie à force atomique nous permet de sonder les propriétés mécaniques des macromolécules biologiques sur une molécule unique. Le schéma du fonctionnement du microscope est représenté figure 1.10.

Le principe consiste à attacher une extrémité d'une protéine sur une surface, et l'autre sur la pointe du microscope. Un système permet d'éloigner le pointe de la surface, et un système de laser couplé à une photodiode permet de mesurer la force que la protéine exerce sur la pointe du microscope au fur et à mesure que l'on éloigne celle-ci. Les premiers travaux se concentraient sur les propriétés élastiques des molécules d'acides nucléiques [Wang 1997] ou encore des protéines [Rief 1997] [Tskhovrebova 1997].

Des expériences de microscopie à force atomique, menées sur sur plusieurs protéines composées de domaines répétitifs comme la titine, on montré un profil en dents de scie (figure 1.11), où chaque pic correspond au dépliement d'un domaine.

La dynamique moléculaire constitue un outil idéal nous permettant de comprendre les phénomènes mis en jeu lorsque l'on effectue ce type d'expérience. Grubmuller [Grubmuller 1996] a été un des premiers à utiliser une simulation de dynamique moléculaire pour reproduire une expérience d'extension sur molécule unique. Par la suite, Schulten a utilisé cette même technique pour comprendre les méca-

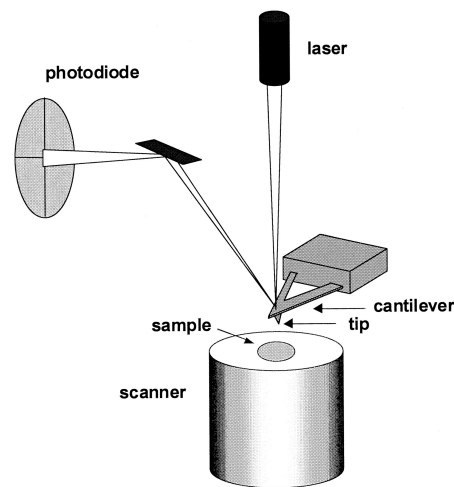


FIGURE 1.10 – Schéma du dispositif expérimental du microscope à force atomique. Adapté de [Dufrene 2002]

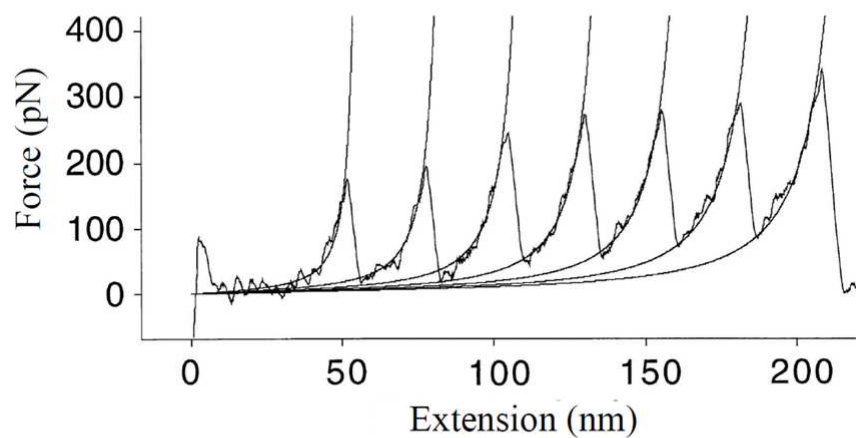


FIGURE 1.11 – Profil de force extension des domaines immunoglobuline de la titine à vitesse constante. Adapté de [Rief 1997]

nismes survenant lors du dépliement mécanique des protéines [Lu 1999]. Ces modèles consistent en représentation explicite de tous les atomes dans notre système, contrairement aux modèles simplifiés où chaque résidu est représenté par son $C\alpha$

uniquement.

De nombreuses études ont montré que la résistance mécanique était dépendante du type de structure secondaire présent dans l'état natif [Li 2007], montrant ainsi que les hélices- α formaient des structures moins résistantes que les feuillets- β . Ces travaux ont été réalisés avec des modèles simplifiés, démontrant ainsi l'utilité de ces modèles pour ce genre d'études.

Enfin, on peut noter que ce type de simulations, combinés a des résultats expérimentaux sont actuellement utilisés pour expliquer la stabilité mécanique des protéines [Lee 2010]. Les simulations de dynamique moléculaire sont donc un outil indispensable nous permettant de comprendre ces phénomènes.

Analyses structurales du complexe chaperonne-protéine client Fha30-Par27

"Le Piprite chaperonné (Piprites pileata) est une espèce de passereau placé dans la famille des Tyrannidae."

Encyclopédie libre Wikipédia

Sommaire

2.1	Introduction	25
2.2	Matériels et méthodes	28
2.2.1	Préparation des échantillons	28
2.2.2	Diffusion de rayons X aux petits angles	29
2.2.3	Diffusion de neutrons aux petits angles (SANS)	29
2.2.4	Modélisation et analyse d'ensembles	29
2.3	Résultats et Discussion	31
2.3.1	Modélisation de Par27	31
2.3.2	Modélisation de Fha30	45
2.3.3	Modélisation de Fha30 dénaturé	46
2.3.4	Modélisation du complexe Fha30/Par27	48
2.4	Conclusion	53

2.1 Introduction

L'excrétion de certaines grandes protéines de virulence par les bactéries à Gram négatif est faite grâce à deux systèmes : le système d'autotransporteur (AT) et le système de sécrétion à deux partenaires (Two Partner Secretion, TPS), la différence entre les deux venant du mécanisme de passage à travers la membrane externe [Yen 2002].

Le système AT implique une seule protéine capable de traverser la membrane de manière autonome. La protéine contient deux domaines, le domaine passager, situé dans la partie N-terminale, et le domaine transporteur situé dans la partie C-terminale [Halter 1984]. Le domaine transporteur s'insère dans la membrane externe, formant

un tonneau β [Maurer 1999b], et transporte le domaine passager dans le milieu extracellulaire. Une fois le transport effectué, le domaine passager est dans la grande majorité des cas clivé par une protéase [Halter 1984].

Le système TPS quant à lui repose sur deux protéines distinctes : la protéine sécrétée (TpsA) et son transporteur associé (TpsB). TpsB s'insère dans la membrane externe, en formant généralement un tonneau β [Clantin 2007], et reconnaît spécifiquement son partenaire TpsA grâce à un domaine présent sur la partie N-terminale de la protéine sécrétée, appelé domaine TPS.

Avant de passer au travers de la membrane externe, les protéines de ces systèmes doivent traverser la membrane plasmique afin d'entrer dans le périplasma. Cette étape est réalisée grâce à la machinerie Sec qui est responsable de la translocation de nombreuses protéines dans l'espace périplasmique [Pugsley 1993]. La reconnaissance de la protéine client se fait au moyen d'une séquence signal présente sur sa partie N-terminale. L'étape de translocation est alors effectuée directement lors de la synthèse de la protéine client, le ribosome formant un complexe avec le transporteur Sec. Le transport est ensuite effectué de manière active, dans le sens où la translocation nécessite l'hydrolyse d'une molécule de GTP [Osborne 2005].

Peu de choses sont connues sur le devenir de ces protéines dans le périplasma, mais on suppose qu'elles s'y trouvent sous une forme désordonnée, donc sensibles à la protéolyse et à l'aggrégation [Klauser 1992].

L'hémagglutinine filamenteuse (FHA) de *Bordetella pertussis* est une protéine de 220kDa, responsable de l'adhérence des bactéries sur les tissus infectés et un des principaux agents de virulence de cette bactérie responsable de la coqueluche. Cette protéine, excrétée dans le milieu extracellulaire, est d'abord transportée à travers la membrane plasmique par la machinerie Sec [Grass 2000]. La reconnaissance de la protéine par la machinerie Sec se fait par la présence d'un peptide signal dans la partie N-terminale de la protéine, qui est clivé une fois le transport effectué [Lambert-Buisine 1998]. Une fois dans le périplasma, FHA est ensuite exportée à travers la membrane externe par un transporteur spécifique, FhaC. Il a été démontré que FHA ne pouvait pas traverser la membrane lorsque la protéine était repliée [Guedin 1998], et que le transporteur FhaC formait des pores dans la membrane [Jacob-Dubuisson 1999]. Cette étape ne requiert donc pas l'hydrolyse d'un quelconque substrat et se fait donc sans apport d'énergie, par un transport passif.

Les structures de Fha30, un fragment de 30kDa correspondant à la partie N-terminale de FHA [Clantin 2004] (qui est la partie exportée en premier, mais aussi plus petit fragment de FHA pouvant être exporté), ainsi que celle de son transporteur FhaC [Clantin 2007], résolues par diffraction de rayons X montrent que FHA doit traverser la membrane dans une conformation dépliée, car le canal de FhaC n'est pas suffisamment large pour laisser passer une protéine repliée (figure 2.1).

Il est donc très probable que FHA se trouve dans une conformation dépliée dans le périplasma, et donc prise en charge par des protéines chaperonnes afin d'éviter toute dégradation protéolytique. C'est en voulant identifier des partenaires qui se fixaient sur une forme dénaturée de Fha30 que la protéine Par27 a été identifiée comme une chaperonne périplasmique. Cette protéine forme des dimères de 54 kDa

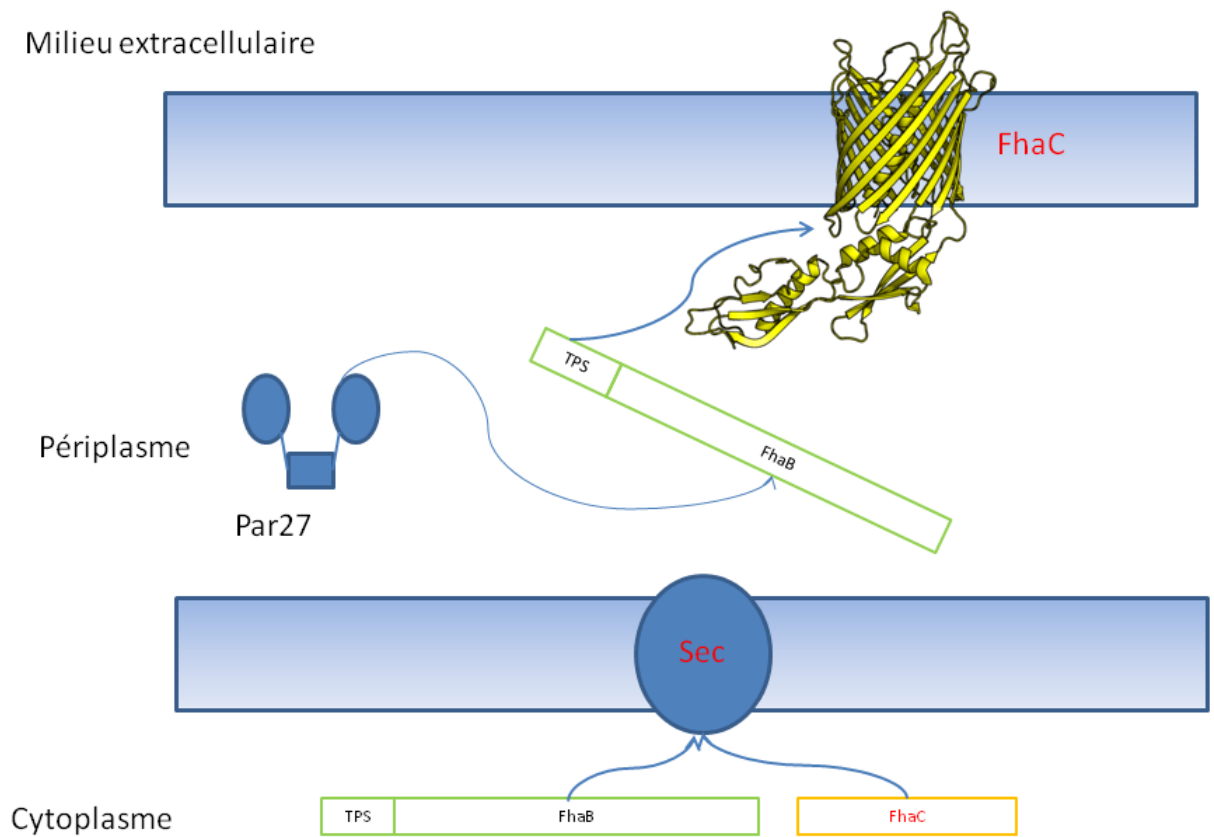


FIGURE 2.1 – Schéma représentant le mécanisme d'exportation de FhaA dans le milieu extracellulaire

et possède une activité peptidyl-prolyl isomérase (PPIase) [Hodak 2008].

L'inactivation du gène de Par27 n'a aucun effet notable sur la croissance bactérienne, ni sur l'excrétion de Fha30 [Hodak 2008]. Il est donc vraisemblable que l'activité de Par27 soit redondante et qu'en son absence, FhaA soit prise en charge par d'autres chaperonnes périplasmiques. Nous avons entrepris l'étude de cette chaperonne et de son interaction avec sa protéine client Fha30, car ce système représente un modèle de chaperonne périplasmique fonctionnant sans hydrolyse d'ATP.

La technique de diffusion aux petits angles combinée aux méthodes de modélisation sous forme d'ensembles conformationnels [Bernado 2007] est une approche parfaitement adaptée à l'étude de ce système. En effet, nous avons une chaperonne qui, malgré le fait d'avoir des régions bien structurées, possède une grande mobilité [Clantin 2010], et elle fixe une protéine qui elle est sous forme dépliée. Il est donc

peu vraisemblable, voire impossible que ces éléments puissent être étudiés par diffraction de rayons X.

Ce chapitre comporte une étude structurale de Par27, Fha30 natif, Fha30 sous forme dénaturée et le complexe entre Par27 et Fha30 par des techniques de diffusion aux petits angles. J'ai utilisé Fha30 car ce fragment est de taille modérée, et il est relativement facile à produire dans un système recombinant. Ce fragment contient en outre la partie TPS mais aussi le début de la partie R1 qui se répète sur toute une partie de FHA. On peut donc supposer que ce fragment est représentatif de la molécule entière.

2.2 Matériels et méthodes

2.2.1 Préparation des échantillons

Préparation de Par27, Fha30 et du complexe Fha30/Par27 : Les différentes protéines ont été produites dans un système recombinant *E. coli*, les gènes étant clonés dans un plasmide contenant une résistance à un antibiotique. Afin de dénaturer Fha30, la protéine a été dialysée dans un tampon contenant 6M de chlorure de guanidinium. La formation du complexe s'est faite ensuite en mélangeant cette solution de Fha30 dénaturée avec une solution de Par27 dans un tampon 20 mM Tris pH 7.5, car Fha30 doit être initialement dénaturée pour pouvoir former le complexe [Hodak 2008]. Pour éviter de se retrouver avec des concentrations trop importantes de chlorure de guanidinium (Gu – HCl), la solution de Fha30 a dû être diluée au moins dix fois, réduisant la concentration finale de Gu – HCl à moins de 0.2 M.

Le mélange a été réalisé de manière stœchiométrique (1 molécule de Fha30 pour 2 molécules de Par27), la masse moléculaire a été vérifiée par SEC-MALLS-RI afin de déterminer l'état d'oligomérisation des différentes protéines et de vérifier la monodispersité des échantillons.

Préparation de Par27 deutéré : Pour les expériences de SANS, il a été choisi de deutérer Par27, car les procédures d'expression et de purification étaient plus simples. La deutération a été faite en cultivant les bactéries dans un milieu contenant du D₂O et dans des conditions de haute densité cellulaire. Cette étape a été réalisée par le service de deutération du D-lab de l'ILL.

Les cultures à haute densité cellulaire nécessitent d'utiliser un marqueur de résistance dont l'effecteur a une localisation cytoplasmique. Pour cette raison, j'ai sous-cloné le gène de Par27 dans un plasmide pET28 contenant un marqueur de résistance à la kanamycine. Ce plasmide ayant aussi une étiquette poly-histidine, la protéine a pu être purifiée sur une colonne Ni²⁺-NTA.

Après vérification la qualité et la pureté des échantillons par SEC-MALLS-RI et SDS-PAGE, la formation du complexe a été réalisée dans les mêmes conditions que

celles décrites dans le paragraphe ci-dessus.

Préparation du complexe Fha30/Par27 deutéré pour l'expérience de SANS :

Le complexe et Par27 seul ont ensuite été dialysés dans des solutions contenant des concentrations de D₂O de 30, 40 et 60%. Malheureusement, le complexe a précipité à des concentrations de D₂O supérieures à 30%. Malgré de nombreuses tentatives (variations du pH, fortes concentrations de sels, différents tampons, différents sels, etc...) aucune condition n'a permis de garder le complexe soluble en présence de D₂O. Seuls les échantillons suivants ont pu être utilisés pour des mesures de SANS :

- le complexe dans un tampon 20mM tris pH 7.5, 150 mM NaCl
- Par27 deutéré dans le même tampon, avec des concentrations en D₂O de 0, 30, 40 et 60%

2.2.2 Diffusion de rayons X aux petits angles

Les mesures de diffusion aux petits angles sur la ligne ID-14-3 de l'ESRF. Chaque échantillon à été mesuré à trois concentrations différentes afin d'écartier des effets d'agrégation ou d'interactions dépendant de la concentration. L'intégration radiale a été faite par un logiciel disponible sur place, et la soustraction du tampon avec le programme primus [P.V.Konarev 2003].

2.2.3 Diffusion de neutrons aux petits angles (SANS)

Les mesures de diffusion de neutrons aux petits angles ont été effectuées sur l'instrument D22 de l'ILL. Les mesures ont été faites à deux distances détecteur-échantillon différentes : 5.6 et 2m. L'intégration radiale et la réduction des données ont été faites avec le programme GRASP, et l'analyse de Guinier [Guinier 1939] et l'extrapolation de l'intensité à angle zéro nous ont révélé des quantités d'agrégation non négligeables dans tous les échantillons.

2.2.4 Modélisation et analyse d'ensembles

Construction du modèle de Fha30 native : La structure résolue par cristallographie est un dimère, ou il manque les 5 dernier résidus de la partie C-terminale de l'un des monomères. Les résidus 292 à 297 de la partie C-terminale d'un des dimères ont du donc être placés manuellement en dehors de l'hélice β , dans une conformation étendue grâce au logiciel UCSF Chimera [Pettersen 2004].

Afin de générer un modèle me permettant d'expliquer les données de diffusion aux petits angles, des calculs de dynamique moléculaire ont été effectués. D'abord, un premier calcul de 100 ps a été fait, afin d'équilibrer le système. Le dimère a été solvaté avec des molécules d'eau dans une boîte périodique, laissant ainsi un espace de 4Å entre les bords de la boîte et la protéine. Des ions Na et Cl ont aussi été rajoutés

afin d'avoir une concentration finale de NaCl de 150mM, qui est celle utilisée dans les expériences de SAXS. Tous les atomes sont représentés de manière explicite. La simulation a été effectuée avec un pas de temps de 2fs, grâce à l'algorithme de contrainte SHAKE [Ryckaert 1977]. Les interactions électrostatiques sont évaluées avec la méthode du maillage particulier d'Ewald [Toukmaji 1996], et les interactions de Van der Waals avec une méthode de "cut-off". La pression et la température sont maintenues constantes grâce à un piston de Langevin, qui simule des forces de frottement aléatoires sur le système, avec une constante de 5ps^{-1} . Cette valeur élevée permet au système de s'équilibrer correctement sans produire des erreurs lors de l'intégration.

Après cette étape, le calcul de dynamique sans contraintes a été lancé pour une durée de 1ns, avec cette fois une constante de friction de 1ps^{-1} . La structure et les vitesses initiales du système de départ sont celles de la fin de la précédente étape. Tous les calculs ont été réalisés avec le programme NAMD [Phillips 2005].

Construction de l'ensemble des modèles de Fha30 dénaturée : Afin de modéliser Fha30 dénaturé, une procédure de génération d'ensembles de conformations suivie d'une étape de sélection de ces modèles ont été réalisées.

La première étape de génération d'ensembles a été faite par le programme RANCH [Bernado 2007]. Ce programme génère des modèles aléatoires tout en respectant certaines contraintes de torsion qui dépendent du type de résidu.

Ces modèles (10000 au total) ont ensuite subi une étape de sélection faite par le programme GAJOE [Bernado 2007] pour obtenir un ensemble de 20 modèles permettant de reproduire les données de SAXS. Ces modèles étant des modèles où il n'y a que la chaîne principale, les chaînes latérales ont dû être reconstruites grâce au programme Sabbac [Maupetit 2006] pour les 20 modèles sélectionnés.

Construction des modèles du complexe Par27/Fha30 : Avant d'utiliser des techniques d'analyse d'ensembles, le programme Dammif [Franke 2009] a été utilisé afin de construire une enveloppe à basse résolution. 20 modèles ont été générés, filtrés, alignés avant d'en faire la moyenne. Cette étape a été faite avec les programmes Supcomb [Kozin 2001] et Damaver [Volkov 2003].

Pour construire les modèles à haute résolution du complexe, les 6 modèles différents de Par27 générés pour la publication présentée dans la partie 2.3.1, et les modèles de Fha30 dénaturé générés dans la partie 2.2.4 ont été utilisés. Ces modèles ont été utilisés dans une procédure d'arrimage moléculaire (docking) : chaque modèle de Fha30 a été accroché sur chacun des modèles de Par27. 120 procédures d'arrimage avec le programme BIGGER [Krippahl 2003] ont ainsi été effectuées. À l'issue de cette procédure, les deux meilleures solutions selon le score du programme ont été sélectionnées, et ce pour chaque conformation de Par27. Un calcul de dynamique moléculaire a ensuite été effectué, comme décrit dans le paragraphe 2.2.4. Un modèle a ensuite été extrait à la fin de chaque simulation, qui a été confronté aux données

de diffusion aux petits angles.

2.3 Résultats et Discussion

2.3.1 Modélisation de Par27

La formation d'un complexe entre Par27 et Fha30 avait été précédemment démontrée par des expériences de type "pull-down", et de résonance plasmonique de surface. Ces deux expériences avaient montré que Par27 possède une affinité forte pour la forme dénaturée de Fha30. La fonction de chaperonne de Par27 avait été quant à elle mise en évidence par une expérience de diffusion de lumière statique, qui montrait que Par27 empêche l'agrégation du lysozyme. La stœchiométrie du complexe avec Fha30 ainsi que l'état d'oligomérisation de Par27 avait ensuite été déterminés par SEC-MALLS-RI [Hodak 2008]. Des cristaux ont été obtenus [Wohlkonig 2008] et un jeu de données complet a été collecté, et l'utilisation d'une forme sélénométhionine de la protéine a permis la résolution d'une partie de la structure (parties N et C-terminales) par nos collaborateurs.

Les cartes de densité électronique obtenues ont permis de construire les parties N et C-terminales de la protéine, mais le domaine PPIase était absent. Le réexamen des données de diffraction indiquait la présence de diffusion diffuse qui pourrait provenir de la présence de régions dynamiques [Wall 1997]. Dans ce travail, nous avons construit un modèle du domaine PPIase par homologie et nous l'avons validé en comparant la courbe SAXS du modèle théorique avec la courbe expérimentale.

La structure du domaine PPIase et des domaines N et C-terminaux ont ensuite été utilisées pour effectuer une procédure d'ajustement de corps rigides, pour obtenir des modèles qui reproduiraient la courbe de diffusion de la protéine entière. Cette procédure a produit des modèles hétérogènes, qui différaient de par la positions des domaines PPIase. Enfin, une simulation de dynamique moléculaire a montré que les parties les plus mobiles étaient celles localisées dans les régions de connections des domaines, confirmant ainsi les résultats de la modélisation par corps rigides. Le fait d'avoir des domaines séparés par des connections flexibles est une caractéristique que partagent d'autres chaperonnes périplasmiques, comme SurA [Xu 2007] ou FkpA [Saul 2004]. Cette flexibilité pourrait être due au fait que ce type de protéine doit avoir une grande adaptabilité afin de pouvoir fixer des protéines désordonnées, qui ont donc une multitude de conformations possibles.



Structure and plasticity of the peptidyl-prolyl isomerase Par27 of *Bordetella pertussis* revealed by X-ray diffraction and small-angle X-ray scattering

Bernard Clantin^c, Cédric Leyrat^a, Alex Wohlkönig^c, Hélène Hodak^b, Euripedes de Almeida Ribeiro Jr.^a, Nicolas Martinez^a, Catherine Baud^b, Caroline Smet-Nocca^d, Vincent Villeret^c, Françoise Jacob-Dubuisson^{b,*}, Marc Jamin^{a,*}

^a Unit of Virus Host Cell Interactions, UMI 3265 UJF-EMBL-CNRS, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France

^b INSERM U629, IFR 142, Institut Pasteur de Lille, 1 rue Professeur Calmette, F-59019 Lille Cedex, France

^c UMR8161 CNRS, Institut de Biologie de Lille – Institut Pasteur de Lille – Université de Lille 1 – Université de Lille 2, 1 rue Calmette, 59021 Lille Cedex, France

^d CNRS UMR 8576, UGSF-IFR 147, Université des Sciences et Technologies de Lille 1, 59655 Villeneuve d'Ascq cedex, France

ARTICLE INFO

Article history:

Received 2 October 2008

Received in revised form 8 September 2009

Accepted 16 November 2009

Available online 20 November 2009

Keywords:

Chaperone

Peptidyl-prolyl isomerase

Bacterial periplasm

Whooping cough

Bordetella pertussis

ABSTRACT

Par27 from *Bordetella pertussis* belongs to a newly discovered class of dimeric peptidyl-prolyl isomerase (PPIase)/chaperones from the parvulin family. It is a tripartite protein with a central PPIase domain surrounded by N- and C-terminal sub-domains (NTD and CTD). Here, the Par27 structure was characterized by X-ray crystallography, small-angle X-ray scattering and template-based modeling. In the crystal lattice, Par27 consists of alternating well ordered and poorly ordered domains. The PPIase domains gave rise to diffuse scattering and could not be solved, whereas a 2.2 Å resolution crystal structure was obtained for the NTD and CTD, revealing a cradle-shaped dimeric platform. Despite a lack of sequence similarity with corresponding sub-domains, the topology of the peptide chain in the NTD/CTD core is similar to that of other monomeric PPIase/chaperones such as SurA and trigger factor from *Escherichia coli*. In Par27, dimerization occurs by sub-domain swapping. Because of the strong amino acid sequence similarity to other parvulin domains, a model for the Par27 PPIase domain was built by template-based modeling and validated against small-angle X-ray scattering (SAXS) data. A model of the full-length dimeric Par27 structure was built by rigid-body modeling and filtering against SAXS data using the partial crystal structure of the NTD/CTD core and the template-based PPIase model. The flexibility of protein was accounted for by representing the structure as an ensemble of different conformations that collectively reproduce the scattering data. The refined models exhibit a cradle-like shape reminiscent of other PPIase/chaperones, and the variability in the orientation of the PPIase domains relative to the NTD/CTD core platform observed in the different models suggests inter-domain flexibility that could be important for the biological activity of this protein.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

The periplasm of Gram-negative bacteria is a site of intense trafficking of proteins destined to the outer membrane, the cell surface or the extracellular milieu (Wülfing and Plückthun, 1994). Following their export across the cytoplasmic membrane, many such proteins cross that compartment in extended, unfolded conformations. Depending on their final destination, some of them acquire their native form in the periplasm, while others remain fully or partially extended prior to their secretion across the outer membrane. Gram-negative bacteria have thus developed periplasmic

systems that provide assistance for protein folding and protection from aggregation or proteolytic degradation (Wülfing and Plückthun, 1994; Missiakas et al., 1996; Miot and Betton, 2004). These systems include broad-spectrum as well as specific chaperones, and folding catalysts such as disulfide oxidases/isomerases and peptidyl-prolyl isomerases (PPIases).

A large group of bacterial folding catalysts exhibit both chaperone and PPIase activities. The peptidyl-prolyl isomerase activity is carried out by small globular domains of approximately 100 amino acids that form three distinct families based on their amino acid sequence: the parvulins, the FK506-binding proteins (FKBP) and the cyclophilins (Rahfeld et al., 1994; Göthel and Marahiel, 1999). Each family comprises single-domain PPIases as well as multi-domain proteins. In the latter category, the additional domains often act as chaperones, increasing the affinity and providing the PPIase domain with selectivity for protein substrates (Scholz et al., 1997a). While the importance of the PPIase activity of such multi-domain proteins

* Corresponding authors. Address: Unit of Virus Host Cell Interactions (UVHCI), UMI 3265 UJF-EMBL-CNRS, 6 rue Jules Horowitz, B.P. 181, 38042 Grenoble Cedex 9, France. Fax: +33 4 76 20 94 00 (M. Jamin).

E-mail addresses: francoise.jacob@ibl.fr (F. Jacob-Dubuisson), jamin@embl.fr (M. Jamin).

in vivo remains unclear, the chaperone activity of some of them plays an essential role *in vivo* (Behrens et al., 2001). Thus, the chaperone activity of SurA in the periplasm of *Escherichia coli* is essential for outer membrane protein biogenesis (Lazar and Kolter, 1996; Rouviere and Gross, 1996; Lazar et al., 1998; Sklar et al., 2007).

We recently reported the identification of the prototype of a new group of parvulins, Par27, from the whooping cough agent *Bordetella pertussis* (Hodak et al., 2008). Par27 was identified in a search for accessory factors involved in the secretion of filamentous hemagglutinin (FHA), the major attachment factor of that bacterium. FHA is a large protein secreted by the two-partner secretion (TPS) pathway (Jacob-Dubuisson et al., 2004). It is exported through the cytoplasmic membrane by the Sec machinery and thus transits through the periplasm in an extended conformation before its transport across the outer membrane by FhaC, a specific transporter of the Omp85/TpsB superfamily (Clantin et al., 2007). At the cell surface, FHA progressively folds into a long amphipathic β -helix (Kajava et al., 2001; Clantin et al., 2004). Because FHA exposes amphipathic, extended structures in the periplasm prior to secretion and subsequent folding, it is a good model to study the role of periplasmic chaperones and catalysts. Par27 was shown to bind to a non-native FHA fragment by affinity chromatography (Hodak et al., 2008). Par27 also displays affinity for other proteins rich in amphipathic β structure such as outer membrane porins, and therefore it has been proposed to serve as a general periplasmic chaperone in *B. pertussis* (Hodak et al., 2008).

Par27, which exhibits both chaperone and PPIase activities *in vitro*, is the first identified parvulin protein that forms dimers in solution (Hodak et al., 2008). The central PPIase domain of Par27 is flanked by N- and C-terminal extensions that are found in a number of putative PPIases present mostly in β proteobacteria (Hodak et al., 2008). In this work, we have characterized the structure of Par27 by X-ray crystallography, template-based modeling and small-angle scattering (SAS). We show that the PPIase domains are attached to a homodimeric cradle-shaped ‘core’ domain made by the assembly of the N- (NTD) and C-terminal domains (CTD) of each monomer. By analogy with other PPIases/chaperones, the structure suggests that the NTD/CTD core constitutes the chaperone domain of the protein.

2. Materials and methods

2.1. Expression and purification of recombinant Par27 and its PPIase domain

Production of the full-length protein was described earlier (Hodak et al., 2008). For the expression of the peptidyl-prolyl isomerase (PPIase) domain of Par27 (aa 107–202), the central portion of the gene encoding Par27 was amplified by PCR with the primers 5'-g gatccggcaagatggaatacaagg-3' and 5'-aagcttttaacgggtatcgtcgactgga t-3', and the resulting amplicon was inserted into the BamHI and HindIII sites of pQE30. Par27 and the isolated PPIase domain were produced in *E. coli* as recombinant proteins with a C-terminal His-tag. The proteins were purified by nickel affinity chromatography followed by gel filtration as previously described (Hodak et al., 2008). The protein concentrations were measured by absorbance at 280 nm using $\epsilon_{280\text{ nm}} = 18,500\text{ M}^{-1}\text{ cm}^{-1}$ for full-length Par27 and $14,000\text{ M}^{-1}\text{ cm}^{-1}$ for the PPIase domain.

The homogeneity of each sample was routinely checked by size exclusion chromatography (SEC) coupled with detection by multi-angle laser light scattering (MALLS) and refractometry (RI) (Wyatt, 1998). SEC was performed with a Superdex S200 column equilibrated in 20 mM Tris-HCl, 150 mM NaCl at pH 7.5 and 20 °C. MALLS detection was performed with a DAWN-EOS detector (Wyatt Technology Corp., Santa Barbara, CA) and the differential

refractive index measurement was performed with a RI2000 detector (Schambeck SFD). A dn/dc value of 0.185 ml g^{-1} was used. Combined detection by MALLS and RI allows the determination of the weight-averaged molecular mass (M_w) and its distribution across the elution peak that informs about the dispersity of the sample (Hodak et al., 2008).

Far-UV circular dichroism spectra were recorded at 20 °C with a JASCO J-810 CD spectropolarimeter equipped with a temperature controller (Peltier system).

2.2. X-ray crystallography

Full-length Par27 was crystallized as reported previously (Wohlkönig et al., 2008). A single SeMet-labeled protein crystal was used to collect diffraction data at beamline ID23-1 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Data collection and refinement statistics are summarized in Tables 1 and 2. Three wavelengths (λ_{peak} , λ_{remote} , $\lambda_{\text{inflexion}}$) were chosen around the selenium K-absorption edge (0.9795 Å). The diffraction data for each set were indexed, integrated and scaled using XDS (Kabsch, 1993).

Data were indexed in space group C222, with cell parameters $a = 54.6\text{ Å}$, $b = 214.1\text{ Å}$, $c = 57.8\text{ Å}$, assuming two molecules in the asymmetric unit, or in space group P2 with unit-cell parameters $a = 42.2\text{ Å}$, $b = 142.8\text{ Å}$, $c = 56.0\text{ Å}$ and $\beta = 95.1^\circ$ (Wohlkönig et al., 2008). The positions of 4 of the 5 expected Se atoms were determined with ShelxD (Uson and Sheldrick, 1999). The Se atoms parameters were refined and initial experimental phases were calculated at 2.2 Å by the MAD or MIR method using Sharp (Bricogne et al., 2003). The density modification procedure and the automatic model building were also carried out using Sharp (Bricogne et al., 2003). Data were also deconvoluted using the program DETWIN (Collaborative Computational Project, Number 4, 1994. “The CCP4 Suite: Programs for Protein Crystallography” 1994). Molecular replacement was carried out using AMoRe (Trapani and Navaza, 2008).

2.3. Homology modeling

Homology modeling was performed by submitting the amino acid sequence to the meta-threading-server LOMETS (Wu and

Table 1
Data collection statistics.

Data collection	Se pk1
Cell parameters (Å)	54.90 215.56 57.81
(°)	90.0 90.0 90.0
Space group	C222
Wavelength (Å)	0.97910
Resolution (Å) ^a	2.2 (2.3–2.2)
Completeness (%) ^a	99.1 (100.0)
Redundancy ^a	3.78 (3.81)
I/si ^a	15.36 (4.90)
R _{mrgd} – F (%) ^a	8.7 (33.5)
Beamline	ESRF ID23-1
<i>Refinement</i>	
R _{work} (%) ^b	30.72
R _{free} (%)	32.87
<i>R.m.s.d.</i>	
Bond lengths (Å)	0.029
Bond angles (°)	2.125
<i>Ramachandran statistics</i>	
Most favored (%)	90.2
Additionally allowed (%)	8.9
Generously allowed (%)	0.8
Disallowed (%)	0

^a Number in parentheses is the statistic for the highest resolution shell.

^b R-factor = $\sum ||F_o| - |F_c|| / \sum |F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively.

Table 2

Experimental and theoretical molecular mass (MM) and molecular dimensions of full-length Par27 and of its isolated PPIase domain.

Proteins	MM (kDa) (from MALLS)	MM (kDa) (from SAXS)	MM (Da) (from sequence)	R_g (nm) (from Guinier)	R_g (nm) (from $P(r)$)	D_{max} (nm)
Full-length Par27	60 ± 5	55 ± 3	(2 × 28,433) 56,866	3.9 ± 0.1	3.8 ± 0.1	11.5
PPIase domain	12.0 ± 0.5	12 ± 1	12,222	1.7 ± 0.1	1.7 ± 0.1	5.9

Zhang, 2007), an automated server that takes predictions from nine threading servers, selects ten consensus models based on structure similarity and generates models with MODELLER. The Par27 PPIase domain was ranked as an 'easy' target because of the good quality of template alignments and the accuracy of constraints. Of the 10 consensus models, the first six were constructed using the PPIase domain of Pin1 (Bayer et al., 2003) as a template, and the next four using the PPIase domain of PrsA (Tossavainen et al., 2006). The r.m.s.d. for all heavy backbone atoms in pair-wise comparisons of the different models ranged from 0.7 to 1.2 Å and those for all atoms ranged from 1.8 to 2.2 Å, showing a well-defined global fold. The quality of the models was checked using PROCHECK (Laskowski et al., 1993). For the best-scoring model, 88% of the residues were in the most favored regions of the Ramachandran plot, and the remaining 12% were in additional allowed regions (Laskowski et al., 1993).

2.4. Small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS) experiments

SANS data were collected at the Laue Langevin Institute (Grenoble, France) on beamline D22 (2005). The sample-to-detector distance was 4 m and the neutron wavelength was 0.6 nm so as to cover a scattering vector (Q) range from 0.18 to 1.20 nm⁻¹, where $Q = 4\pi(\sin \theta)/\lambda$. Samples were loaded in a 1 mm pathlength quartz cuvette. The protein concentration was 6.3 mg ml⁻¹. Reduction of SANS data was performed with a program suite available at D22.

SAXS data were collected at the European Synchrotron Radiation Facility (Grenoble, France) on beamlines ID2 (Narayanan et al., 2001) and ID14-3. The sample-to-detector distance was 1 m and the wavelength of the X-rays was 0.0995 nm. Samples were contained in a 1.9 mm wide quartz capillary. The time of exposure was optimized for reducing radiation damage. Data acquisition was performed at 20 °C. Protein concentrations ranged from 1.1 to 6.7 mg ml⁻¹ for Par27 and from 2.1 to 8.4 mg ml⁻¹ for the PPIase domain. Reduction of the 2D-SAXS patterns to 1D scattering profiles was performed using the established procedure available at ID2, and buffer background runs were subtracted from sample runs.

The radius of gyration and forward intensity at zero angle ($I(0)$) were determined with the programs PRIMUS (Konarev et al., 2003) by using the Guinier approximation at low Q values, in a $Q \cdot R_g$ range up to 1.3:

$$\ln I(Q) = \ln I(0) - \frac{R_g^2 Q^2}{3} \quad (1)$$

For SANS experiments, the molecular mass of Par27 was directly calculated from the forward intensity, $I(0)$, knowing the protein concentration and the scattering length density (Jacrot and Zaccai, 1981). For SAXS experiments, the forward scattering intensity was calibrated using bovine serum albumin and lysozyme as references. The radius of gyration and pairwise distance distribution function, $P(r)$, were calculated with the program GNOM (Semenyuk and Svergun, 1991). The maximum dimension (D_{max}) value was adjusted so that the R_g value obtained from GNOM agreed with that obtained from the Guinier analysis.

Models of full-length Par27 and its PPIase domain were obtained by using the program BUNCH (Petoukhov and Svergun, 2005). For the PPIase domain, the all-atom model obtained by template-based modeling was completed by adding dummy residues in place of the 17 N-terminal residues. For full-length Par27, rigid-body modeling was carried out using the X-ray structure of the core domain encompassing the NTD and CTD domains of Par27 and the structure of the PPIase domain predicted by template-based modeling. The NTD and CTD of the core domain from the crystal structure were used as independent sub-domains, thus allowing small fluctuations in their relative orientation. Amino acids 90–99 and 189–195 were replaced by dummy atoms. Modeling of full-length Par27 was performed assuming twofold symmetry around the NTD dimerization interface. Eighty independent models were built, of which 45 were pseudo-enantiomorphs of the crystal structure and were discarded. The normalized spatial discrepancy (NSD) values were calculated with the program DAMAVER between each pair of models without realignment, to evaluate their resemblance and select the most typical model; i.e., the model with the lowest deviation from the others (Table 3). Models were discarded if their NSD value was larger than two standard deviations.

2.5. Molecular dynamics simulation

An all-atom model was constructed from the best BUNCH model of Par27. Using the C_α coordinates generated by BUNCH, the backbone and side chains for the N-terminal His-tag (aa 1–11) and the flexible linkers joining the PPIase domain to the NTD (aa 114–123) and CTD domains (aa 212–219) were built using the SABBACv1.2 server (Maupetit et al., 2006). The resulting all-atom model was energy minimized and used as input for molecular dynamics simulations. Molecular dynamics refinement of the all-atom model of Par27 was carried out using the GROMACS 4.0.2 software package (Hess et al., 2008) and the GROMOS96 53A6 force field (Oostenbrink et al., 2005). The protein was solvated with 27006 SPC/E water molecules in a triclinic box. The minimum distance between any atom of the protein and the box wall was 0.9 nm. Na⁺ and Cl⁻ ions were added to a concentration of 150 mM in order to approximate physiological conditions. Covalent bonds in the protein were constrained using the LINCS algorithm. The geometry of the water molecules was constrained using the SETTLE algorithm. Electrostatic interactions were calculated using the Particle Mesh Ewald (PME) algorithm with a cut-off of 0.9 nm. Lennard-Jones interactions were calculated with a 0.9–1.4 nm twin-range cut-off. Hydrogen atoms were treated as virtual sites to increase computational efficiency. A dielectric constant of 1 and a time step of 5 fs were used. The temperature and the pressure were maintained at 300 K and 1 bar by using the v-rescale thermostat and the Parrinello–Rahman barostat, respectively. The relaxation times were 0.1 and 2.0 ps, respectively. The energy of the system was initially minimized by 2500 steps of steepest descent, followed by 125 ps of position restraints molecular dynamics to let the solvent equilibrate around the protein. Production simulation was run for 20 ns and snapshots were collected every 5 ps for analysis. Root-mean square deviation (r.m.s.d.) and Root-mean square fluctuation (r.m.s.f.) were calculated on the C_α

Table 3
Normalized spatial discrepancy (NSD) and χ values (fit to the experimental SAXS curve) for Par27 models built with BUNCH – 10 best-fitting models. The average NSD value for all pair-wise comparisons is equal to 1.1 and the standard deviation is equal to 0.4.

Models	Consensus models									Outlier models
	8	7	2	6	1	5	9	4	3	
(NSD) before realignment	0.951	0.966	0.976	0.977	0.979	0.988	1.034	1.036	1.001	2.374
NSD after realignment	Ref.	0.760	0.762	0.751	0.860	0.678	0.909	0.570	0.868	–
χ Value	0.51	0.51	0.50	0.51	0.48	0.60	0.53	0.54	0.45	0.56

atoms after least-square fitting of the backbone using the GRO-MACS routines.

2.6. PPIase activity assays

Two PPIase activity assays were used to characterize the enzymatic activity of the isolated PPIase domain, using either reduced carboxymethylated RNase T1 or a 16-mer peptide as substrates as previously described for the full-length protein (Hodak et al., 2008). For the peptide assay, NMR experiments were performed on a Bruker 600 MHz DMX spectrometer (Bruker, Karlsruhe, Germany) equipped with a cryogenic triple resonance probe head. All ^1H spectra were calibrated with 1 mM sodium 3-trimethylsilyl-d(3,3',2,2')-propionate as a reference. Standard TOCSY and NOESY spectra with 60 and 400 ms mixing times, respectively, were acquired on a 2 mM peptide solution in 25 mM phosphate buffer (pH 6.6), 50 mM NaCl at 293 K to assign peptide ^1H resonances. The content of *cis* conformers was calculated with the integration of the $\text{H}\alpha/\text{H}\gamma$ correlation peaks for both *cis* and *trans* signals in a homonuclear TOCSY experiment. To measure the PPIase activity, a $^1\text{H}-^1\text{H}$ EXSY spectrum was recorded at a mixing time of 300 ms on a 2 mM peptide sample containing 0.050 mM (2.5%) Par27 in 25 mM phosphate buffer (pH 6.6), 50 mM NaCl in 100% D_2O . A series of $^1\text{H}-^1\text{H}$ EXSY spectra were then acquired at various mixing times (100, 200, 300, 400 and 500 ms) to quantify the exchange rate of the Par27 isomerase activity. Integration of the $\text{H}\alpha$ *cis*/ $\text{H}\alpha$ *trans* cross-peaks detected for both proline residues was performed. The exchange cross-peaks were normalized with the integration value of the corresponding diagonal peak to provide the ratio of molecules that underwent exchanges from the *cis* to the *trans* conformation during the mixing time. The exchange rate constant (k_{exch}) was calculated by fitting the theoretical curve obtained with the following equation to the experimental points:

$$\%(cis \rightarrow trans) = \alpha \frac{1 - \exp(-(1 + 1/\alpha)k_{\text{exch}}M_T)}{1 + \exp(-(1 + 1/\alpha)k_{\text{exch}}M_T)} \quad (2)$$

where $\%(cis \rightarrow trans)$ corresponds to the fraction of molecules that underwent changes from the *cis* to the *trans* conformation during the mixing time (M_T), and α is the excess of *trans* over *cis* conformer.

RNase T1 was purchased from Sigma. The reduction and carboxymethylation of RNase T1 to generate RCM-T1 were performed as previously described (Mücke and Schmid, 1994). RCM-T1 was unfolded in 0.1 M Tris-HCl (pH 8) for at least 1 h at room temperature at a final concentration of 20 μM . The refolding was initiated by a 40-fold dilution in 2 M NaCl. The time course of refolding of the RCM-T1 was followed at 320 nm with a QM-4 spectrofluorometer (PTI) using $\lambda_{\text{exc}} = 268$ nm. For the calculation of k_{cat}/k_m , Par27 or the PPIase domain were used at a fixed concentration and RCM-T1 concentration was varied from 100 nM to 1 μM . The reaction was measured in a total volume of 2 ml, in a quartz cuvette under constant agitation at 15 $^\circ\text{C}$. For each RCM-T1 concentration, the initial velocity value measured in the absence of enzyme was subtracted from the initial velocity value measured in its presence.

3. Results

3.1. X-ray diffraction analysis of full-length Par27

Our initial goal was to solve the structure of Par27 using X-ray diffraction techniques. Full-length Par27 was crystallized as reported previously (Wohlkönig et al., 2008). Data were collected for a SeMet derivative, indexed in space group C222 and subjected to either the MAD or MIR method for phasing. Surprisingly, only a partial model corresponding to residues 1–101 and 208–239 could be built using automated procedures. A high quality electron-density map was obtained at 2.2 Å resolution for residues 1–89 and 218–239 and the sequence assignment in this part of the structure was straightforward (Fig. 1a). In addition, the backbone structure could be built for residues 90–101 and 208–217, but no electron density could be observed for residues 102–207 that encompass the previously identified PPIase domain (aa 112–199) (Hodak et al., 2008).

To confirm that residues 102–207 were present in the crystal, we dissolved the crystal and verified protein integrity by mass spectrometry, which yielded the expected mass for the full-length protein. The part of the protein missing in the structure was not expected to be intrinsically unstructured, since it includes a clearly identified PPIase domain of the parvulin family, and it was not expected to be flexible, since it appears to be involved in the crystal packing. However, examination of the diffraction patterns revealed the presence of a pronounced diffuse scattering and of many diffraction spots characteristic of an incommensurable lattice. Inspection of experimentally phased electron-density maps confirmed the absence of electron density for the PPIase domain. These results suggested that, within the crystal, the PPIase domain adopts periodically shifted conformations differing from the crystal periodicity, thus smearing out the electron density corresponding to this domain. The void left in the lattice by the absence of this part of the protein was large enough to accommodate the missing PPIase domains in a conformation similar to that obtained in the model reconstructed from SAXS data (see below). No further attempt was made to carry on further this analysis because of the inherent difficulties encountered in collecting, integrating and interpreting diffuse scattering data (Wall et al., 1997; Wilson and Brunger, 2003) and because of our inability to obtain crystals of sufficient quality for this kind of study.

3.2. The NTD/CTD core structure of Par27

The partial crystal structure and sequence alignments (Hodak et al., 2008) indicated that Par27 can be divided into an N-terminal sub-domain (NTD), a central PPIase domain and a C-terminal sub-domain (CTD) (Fig. 1b). In the crystal structure, Par27 forms symmetrical homodimers in which the NTD and CTD sub-domains of both monomers interact to form a cradle-shaped NTD/CTD core structure (Fig. 1c). The NTD of each monomer consists of five α -helices ($\alpha 1$ – $\alpha 5$) and a small anti-parallel β -sheet formed by two short β -strands ($\beta 1$ and $\beta 2$). The CTD forms a long α -helix

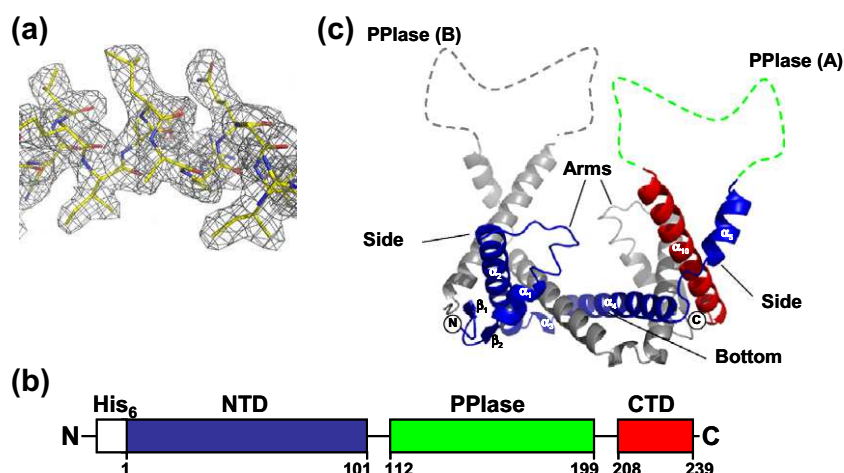


Fig. 1. Crystal structure of the Par27 NTD/CTD core domain. (a) Excerpt of the experimental electron density map of Par27 contoured at 1.4σ level. The illustrated region is part of helix α_3 . (b) Schematic diagram of the modular organization of Par27. The His-tag sequence is shown in white, the N-terminal sub-domain (NTD) in blue, the PPIase domain in green and the C-terminal sub-domain (CTD) in red. Amino acid numbers showing the boundaries of the different domains refer to the amino acid sequence of the mature protein, after cleavage of the signal peptide. (c) Ribbon drawing of the Par27 NTD/CTD core dimer showing one protomer with the same color-code as in (a) and the other in gray. The secondary structure elements are labeled in the colored protomer. The dotted lines show the expected locations of the PPIase domains that could not be solved. (For interpretation of color mentioned in this figure the reader is referred to the web version of the article.)

(α_{10} ; secondary structure elements have been numbered from the N-terminus to the C-terminus in the final model of the entire protein) that packs against helices α_4 and α_5 of the NTD in the same monomer and against helix α_2 of the NTD in the other monomer (Fig. 1c). In each monomer, the anti-parallel β -strands and helices α_1 and α_2 form the 'arms', helices α_5 and α_{10} (CTD) form the 'sides' and helices α_3 and α_4 form a flat dimeric 'bottom' (Fig. 1c).

A sequence homology search using BlastP identified homologues of Par27 mainly in the *Burkholderiales* order (Hodak et al., 2008). However, a search of the Protein Data Bank with the dimeric core domain of Par27 using DALI (Holm and Sander, 1996) revealed structural similarity with the core domain of the periplasmic PPIase/chaperone SurA from *E. coli* (z -score = 6.1) (Bitto and McKay, 2002; Xu et al., 2007). The core domain of SurA also shares structural similarity (Stirling et al., 2006) with the C-terminal clamp-like binding domain of the cytosolic PPIase/chaperone trigger factor (TF) (Ferbitz et al., 2004; Maier et al., 2005) and with the hypothetical *Mycoplasma* protein MPN555, identified as a putative homolog of TF from *Pseudomonas putida* although it contains no PPIase domain (Schulze-Gahmen et al., 2005). These three proteins share with Par27 a modular organization in structural or functional domains although the number of domains and their organization along the amino acid sequence are different (Fig. 2a). The monomeric SurA comprises two PPIase domains of the parvulin family, PPIase 1 and 2, flanked by the N-terminal (NTD) and C-terminal sub-domains (CTD). The NTD, PPIase 1 and CTD domains form a platform to which the enzymatically active PPIase 2 domain is tethered by long extended chains (Bitto and McKay, 2002). TF comprises an N-terminal ribosome binding domain, a central PPIase domain of the FKBP family and two C-terminal ARM domains that binds to nascent proteins leaving the ribosome exit tunnel (Ferbitz et al., 2004; Merz et al., 2006). The MPN555 protein comprises only NTD and CTD sub-domains, but no PPIase domain.

Sequence alignment of the N- and C-terminal sub-domains of Par27 revealed overall 19% identity and 40% similarity with the corresponding sub-domains of SurA (Fig. S1 in Supplementary material), but only 10% similarity with the corresponding domains of TF and MPN555 (data not shown). Although Par27 forms dimers, the topological organization of its core domain is reminiscent of those of the core domains of SurA, TF and MNP555 (Fig. 2b). Com-

parison of the topological organization of Par27 with that of SurA (Fig. 2b) indicated conservation of strands β_1 and β_2 and of helices α_1 and α_2 . In SurA, helices α_3 and α_4 form a close to 90° angle that allows helices α_5 and α_6 to dock against helices α_1 , α_2 and α_{18} , while in dimeric Par27, the corresponding residues form a long helix, α_4 , that interacts with helix α_{10} of the same chain and with helix $\alpha_{2'}$ of the other chain. This geometry of the polypeptide chain introduces a sub-domain swapping in Par27 involved in the dimerization of the protein mainly through interactions between helices α_4 and $\alpha_{4'}$ that lie anti-parallel to each other in the dimer, but also through contacts between residues of helices α_2 and $\alpha_{10'}$, $\alpha_{2'}$ and α_{10} , α_3 and $\alpha_{4'}$ and $\alpha_{3'}$ and α_4 . The long α -helix of the C-terminal domain in both Par27 (α_{10}) and SurA (α_{18}) and the following short β -strand of SurA (β_{11}) dock onto strand β_1 and helix α_5 . Alignment of the three-dimensional structures using the SSM server (Krissinel and Henrick, 2004) indicated that the anti-parallel β -sheet and helices α_1 , α_2 , α_3 , α_5 and α_{18} of SurA can be partially superimposed on the corresponding secondary structure elements of Par27 (Fig. 2c). An r.m.s.d. of 3.6 \AA was obtained for the backbone atoms, suggesting a good similarity between the two proteins, although SurA is monomeric and Par27 dimeric. A SurA variant lacking the second PPIase domain forms a dimer in the presence of a dodecapeptide substrate (Xu et al., 2007), but its topological organization is different from that of Par27 (Fig. 2b and Fig. S2 in Supplementary material). In the dimer of the SurA variant, the main inter-chain interactions are between helices α_{18} and $\alpha_{18'}$ that lie parallel to each other and between helix α_6 and helices $\alpha_{3'}$ and $\alpha_{5'}$ (as well as $\alpha_{6'}$ and α_3 and α_5). In TF, the topological arrangement of helices α_7 – α_{11} forming the two arms is similar to that found in SurA (Fig. 2b). However, the long helix α_6 is located upstream of the ARM domains in the TF amino acid sequence, although it occupies a position in space similar to that of the long CTD helix of SurA. In the MPN555 protein, the topology is similar to that of SurA except for the presence of an additional β -sheet, formed by strands β_1 and β_3 , and an additional helix, α_3 , inserted in the sequence at a position equivalent to the PPIase domains of SurA and Par27 (Fig. 2b and Fig. S2 in Supplementary material). As for SurA, the three-dimensional structure of the Par27 NTD/CTD core domain can be aligned with corresponding regions of TF and MPN555 (Fig. S2 in Supplementary material).

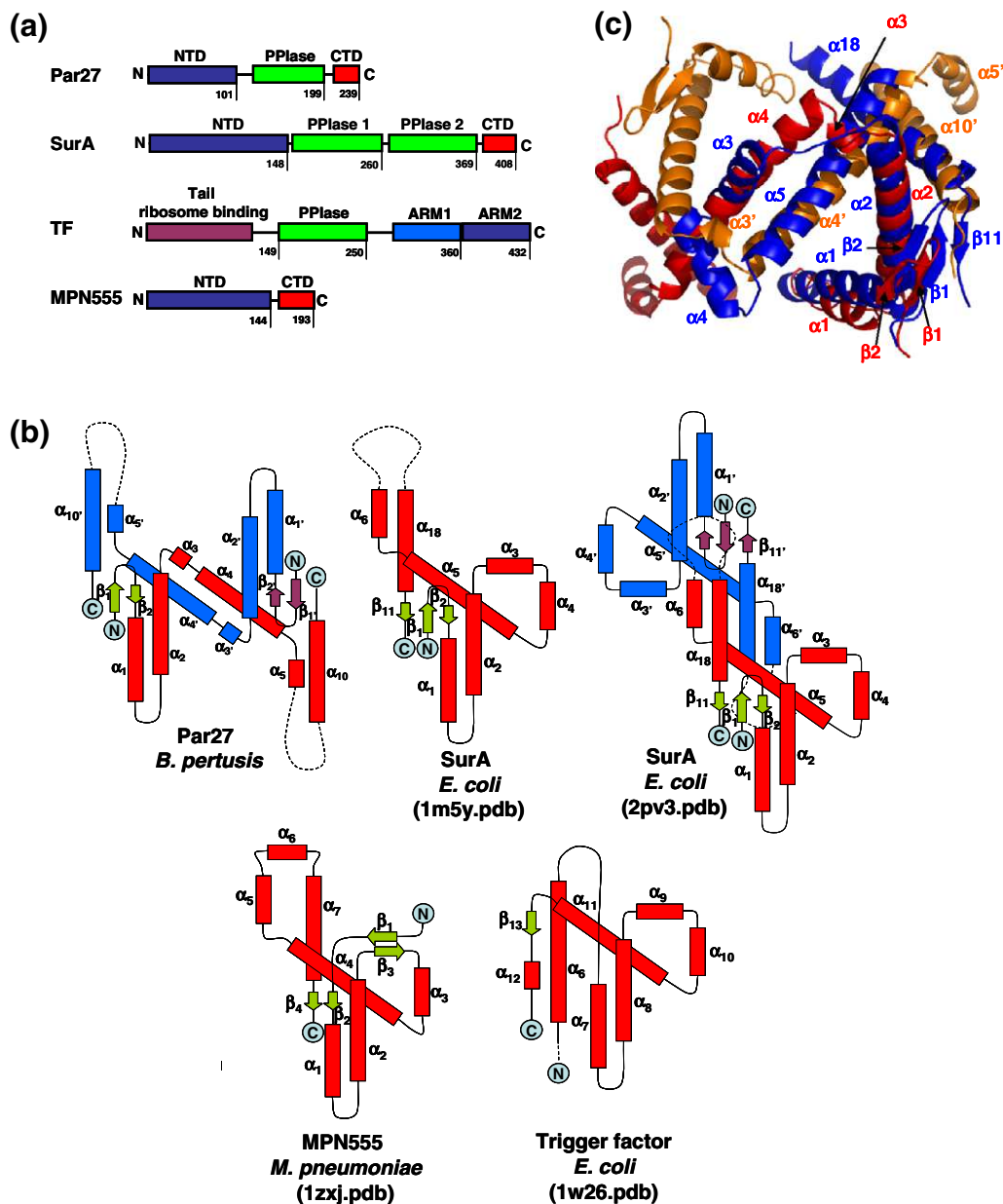


Fig. 2. Structural comparison of the Par27 core with other clamp-like PPIase/chaperones: SurA and trigger factor from *E. coli* and MPN555 protein from *M. pneumoniae*. The numbers indicate the C-terminal boundaries of the different domains in the mature proteins. (b) Topology of the core domains of the different proteins. The green arrows and red boxes show β -strands and α -helices, respectively. In Par27 and SurA dimers, violet arrows and blue boxes show β -strands and α -helices in the second protomer. The secondary structure elements in the second protomer are labeled with a prime. (c) Superposition of the Par27 NTD/CTD core dimer with the SurA core domain (NTD and CTD sub-domains). Superposition of the structure was performed with SSM (Krissinel and Henrick, 2004). The α carbon r.m.s.d. was 3.6 Å. SurA is shown in blue, one Par27 protomer in red and the other in orange. The corresponding secondary structure elements are labeled. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

3.3. Small-angle scattering experiments

In order to build a structural model of the entire protein, small-angle X-ray (SAXS) and neutron (SANS) scattering data were combined with molecular modeling and with the crystal structure of the NTD/CTD core (Putnam et al., 2007; Tsutakawa et al., 2007). The PPIase domain and full-length Par27 were produced as recombinant proteins with an N-terminal His-tag and purified as previously described (Hodak et al., 2008). Circular dichroism spectroscopy revealed that both proteins were folded. The monodispersity of all samples used in SAXS and SANS experiments was checked by size exclusion chromatography (SEC) combined with detection by multi-angle laser light scattering (MALLS) and refractometry (RI) (Gérard et al., 2007; Hodak et al., 2008). The

analysis of the isolated PPIase domain by SEC-MALLS-RI yielded a molecular mass of 12.0 ± 0.5 kDa, very close to the calculated mass, indicating that it is monomeric (Table 2).

SAXS profiles for full-length Par27 and the isolated PPIase domain were collected for scattering vector values ($Q = 4\pi(\sin \theta)/\lambda$) ranging from 0.1 to 3.0 nm^{-1} and from 0.2 to 2.3 nm^{-1} , respectively (Figs. 3a and 4a). A SANS profile was also collected for full-length Par27 (Fig. 4a). Guinier plots were linear for $Q \cdot R_g$ values ranging from 0.4 to 1.3 and yielded radius of gyration (R_g) values independent of protein concentrations (Figs. 3b and 4b and Table 2) (Glatter and Kratky, 1982). Similar R_g values were obtained from the distance distribution function, $P(r)$, calculated as the inverse Fourier transform of $I(Q)$ using the program GNOM (Semenyuk and Svergun, 1991) (Figs. 3c and 4c and Table 2). The

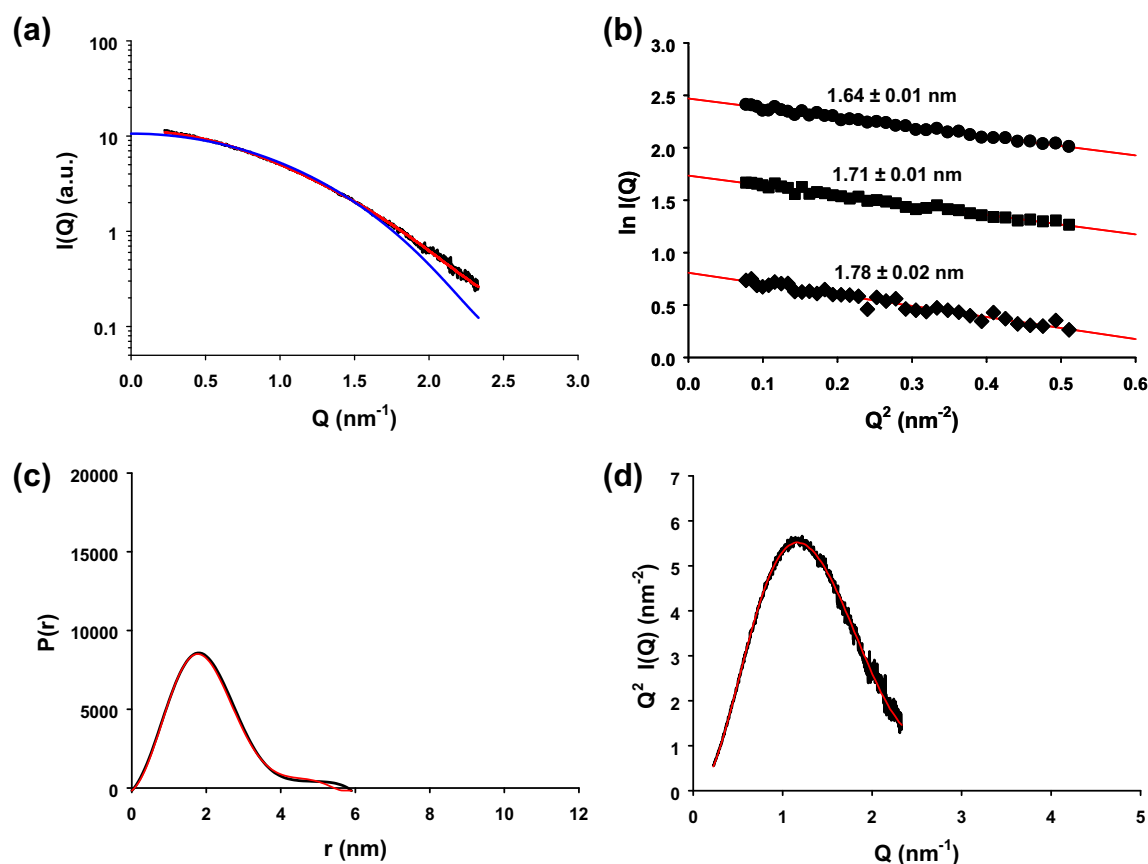


Fig. 3. SAXS profiles of the Par27 PPIase domain. (a) Experimental SAXS data as a function of Q . The black line shows the experimental profile, the blue line the SAXS profile calculated for the best-scoring template-based model (χ value = 4.40) and the red line the curve calculated for the reference model including the N-terminal extension generated with BUNCH (χ value = 0.78). (b) Guinier plots for Q values ranging from 0.22 to 0.73 nm^{-1} ($0.4 < Q \cdot R_g < 1.2$). The symbols are for data measured at protein concentration of 8.4 mg ml^{-1} (circles), 4.2 mg ml^{-1} (squares) and 2.1 mg ml^{-1} (diamonds). The red lines show linear regressions, and the numbers above the curves show the R_g values calculated from the slope of the fitted curves. (c) Distance distribution function and (d) Kratky plot. The $P(r)$ function was calculated using the GNOM software and a D_{max} value of 5.9 nm. The experimental curve is shown in black and the theoretical curve calculated for the reference model generated with BUNCH in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

forward scattering intensity, $I(0)$, provided estimates of the respective molecular masses, in agreement with the calculated molecular masses for the dimeric Par27 and the monomeric PPIase domain and with the molecular masses measured by multi-angle laser light scattering (MALLS) ((Hodak et al., 2008) and this work). These results confirmed that Par27 is dimeric, that the PPIase domain is monomeric and that both proteins are well-behaved and monodisperse in solution.

The $P(r)$ function shows the probability distribution of interatomic distances within the scattering particle. For the PPIase domain, the pair distribution function, $P(r)$, (Fig. 3c) has a symmetrical bell shape typical of a globular protein with a peak centered near 2 nm indicating a maximal dimension of about 4 nm, and the Kratky plot shows a characteristic symmetrical peak, typical of a folded protein (Doniach, 2001; Svergun and Koch, 2003) (Fig. 3d). In the $P(r)$ function, the bell shape curve is followed by a tail extending to a maximum protein dimension (D_{max}) of 5.9 nm. A spherical protein of this size (12.2 kDa, $\bar{v} = 0.73 \text{ cm}^3 \text{ gr}^{-1}$) is expected to have a diameter of about 3.0 nm, suggesting that the PPIase domain is slightly elongated, whereas the N-terminal His-tag could explain the tailing. For the full-length Par27, both the $P(r)$ function (Fig. 4c) and the Kratky plot (Fig. 4d) show complex profiles with several maxima and shoulders. In the $P(r)$ function, the first peak near 2 nm might correspond to the distribution of intra-domain distances, while the other peak centered roughly around $D_{\text{max}}/2$ (5.5 nm), might correspond to the distribution of inter-domain distances. The linear rise at high Q values observed in the Kratky plot

($Q > 3 \text{ nm}^{-1}$) suggests the presence of flexible regions (Doniach et al., 1995; Doniach, 2001).

3.4. Modeling the structure of the PPIase domain of Par27

The PPIase domain of Par27 (aa 112–200) belongs to the parvulin family (Hodak et al., 2008). Pairwise sequence alignments revealed between 35% and 40% identity with six PPIase domains of known three-dimensional structure and multiple sequence alignment revealed several conserved elements (Fig. S3a in Supplementary material) (Ranganathan et al., 1997; Sekerina et al., 2000; Terada et al., 2001; Bitto and McKay, 2002; Bayer et al., 2003; Kühlewein et al., 2004; Li et al., 2005; Tossavainen et al., 2006). A high-resolution structural model was obtained by template-based modeling by using the LOMETS metaserver (Wu and Zhang, 2007) (Fig. S3b in Supplementary material).

The scattering curve calculated from the best-scoring model using the program CRY SOL (Svergun et al., 1995) show, however, discrepancies at low and high Q values, and the quality of the fit to the experimental curve is poor as judged by a χ value of 4.4 (Fig. 3a). The purified protein used in the SAXS experiments comprises a 17-residues N-terminal extension, including a His-tag and five residues of Par27 that were not included in the homology modeling procedure (MRGSHHHHHHGS GKMEY). Twenty independent models of the entire construct were thus generated with the program BUNCH (Petoukhov and Svergun, 2005) by adding dummy residues at the N-terminus of the high-resolution structure obtained

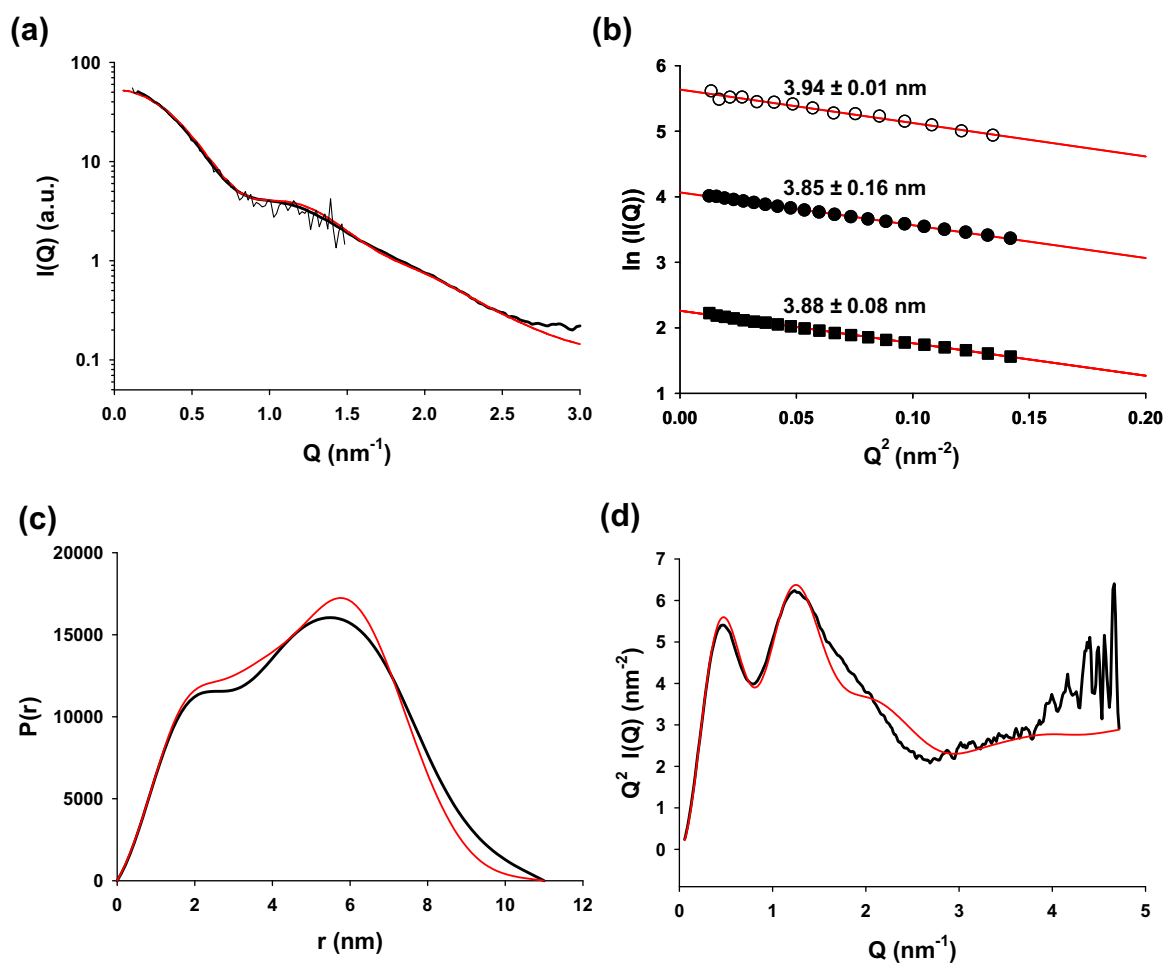


Fig. 4. SAXS and SANS profiles of full-length Par27. (a) Experimental SAXS and SANS data as a function of Q . The thick black line shows the experimental SAXS profile, the thin black line the SANS profile and the red line the theoretical curve calculated for the averaged BUNCH models (χ value = 0.50). (b) Guinier plots for Q values ranging from 0.11 to 0.37 nm^{-1} ($0.4 < Q \cdot R_g < 1.4$). The open circles are for SANS data measured at a protein concentration of 6.1 mg ml^{-1} . The closed symbols are for SAXS data measured at protein concentration of 6.7 mg ml^{-1} (circles) and 1.1 mg ml^{-1} (squares). The red lines show linear regressions, and the numbers above the curves are the R_g values calculated from the slope of the fitted curves. For clarity, each plot has been shifted along the $\ln(I(Q))$ axis. (c) Distance distribution function and (d) Kratky plot. The $P(r)$ function was calculated using the GNOM software and a D_{max} value of 11.5 nm. The experimental curve is shown in black and the theoretical curve calculated for the averaged nine BUNCH models in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

by homology modeling. After alignment of the structured domain using the program DAMAVER (Volkov and Svergun, 2003) (average normalized spatial discrepancy value, (NSD) = 0.48), the added tail exhibits a different conformation in the various models suggesting that it is flexible (Fig. 5). The globular domain is a prolate spheroid with a maximum dimension of 3.9 nm, in good agreement with the SAXS curve analysis, and the presence of the His-tag explains the large dimension of the molecule ($D_{\text{max}} = 5.9 \text{ nm}$) (Fig. 3). The inclusion of dummy residues corresponding to the N-terminal extension significantly improved the quality of the fit to the experimental curve, yielding χ values for the 20 different models that range from 1.41 to 1.43 (Fig. 3a).

3.5. Rigid-body modeling of full-length Par27 structure

Models of the full-length protein were generated by rigid-body modeling with the program BUNCH (Petoukhov and Svergun, 2005) using the crystal structure of the dimeric NTD/CTD core and the template-based model of the PPIase domain. Because building a three-dimensional structure from small-angle scattering data is a problem with more than one solution, 35 independent models were compared. Fig. 6a shows a typical model of Par27 where the PPIase domains are tethered to the dimeric NTD/CTD core through

two shafts, each one formed by helix $\alpha 5$ from the NTD of one monomer and helix $\alpha 10$ from the CTD of the other monomer. The theoretical curves computed for the 35 models yielded χ values that ranged from 0.45 to 1.14. Pair-wise comparison using the program DAMAVER (Volkov and Svergun, 2003) yielded an average normalized spatial discrepancy value ($\langle \text{NSD} \rangle$) of 1.12 (standard deviation = 0.31). Two models were discarded because their NSD values were larger than two standard deviations. If only the best-fitting models with χ values lower or equal to 0.6 were selected (10 models), pair-wise comparison using the program DAMAVER yielded a similar ($\langle \text{NSD} \rangle$) value of 1.14, and the individual NSD values calculated after realignment of each model to the reference model ranged from 0.57 to 0.91 (Table 3). For globular proteins, NSD values larger than 0.7 are indicative of some variability among the models (Volkov and Svergun, 2003). Superposition of the models revealed that they differ mainly in the orientation of the PPIase domains by rotation around the shafts (Fig. 6b). To account for that flexibility, the 10 best-fitting curves were averaged. The averaged curve correctly reproduces the experimental curve (χ value = 0.50) (Fig. 4a) as well as the calculated distance distribution function (Fig. 4c) and Kratky plot (Fig. 4d). Although this procedure provided no improvement to the quality of the fit, an ensemble of conformations appears more relevant for representing a flexible protein.

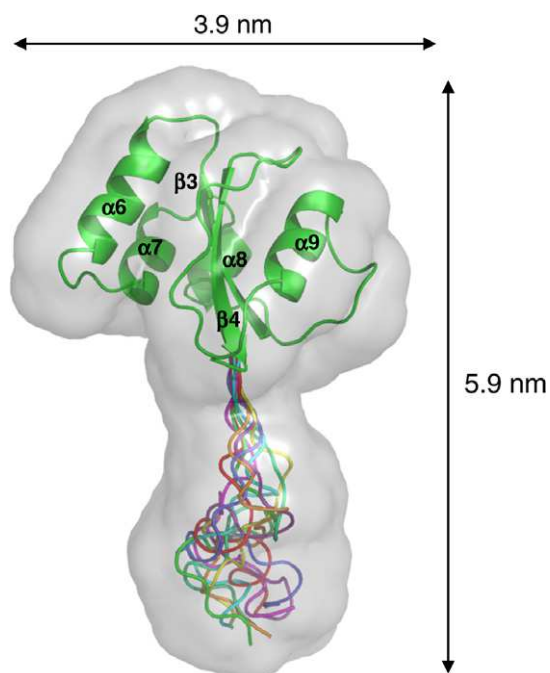


Fig. 5. Solution structure of the Par27 PPIase domain. BUNCH models of the PPIase domain with an N-terminal 6-His-tag. The template-based model of the PPIase domain is shown in ribbon representation and the secondary structure elements are labeled. The extended 17-residues N-terminal extensions from 10 BUNCH models are shown as ribbons of different colors. The gray surface shows the filtered model generated by the program DAMAVER (Damfilt model). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

To test how SAXS data constrain the positioning of the PPIase domains, models were built by manually moving the PPIase domains relative to the dimeric NTD/CTD core, and the calculated SAXS profiles were compared with the experimental data. Models in which the PPIase domains were moved apart (Fig. 7b and f) or toward (Fig. 7d) the NTD/CTD core clearly show poorer fit of the calculated and experimental SAXS profiles (Fig. 7a, c and e), stating that SAXS data are adequate for positioning the PPIase domains within the protein.

3.6. Dynamics in the structure of full-length Par27

To analyze the conformational fluctuations occurring in Par27, a 20 ns-long molecular dynamics simulation was run in explicit solvent and the conformational drift of the structure was measured in terms of root-mean-square deviations (r.m.s.d.) (Fig. 8a). After an initial rise during the first 10 ns in which the structure adjusted to the simulation conditions, the r.m.s.d. value reached a plateau, characteristic of a stable structure. The location of the flexible parts of the protein was determined by calculating the root-mean-square fluctuations (r.m.s.f.) of C_{α} atoms with respect to their time-averaged position over the last 10 ns of the simulation (Fig. 8b). The plot of the r.m.s.f. along the amino acid sequence revealed largest movements of helix $\alpha 1$, the C-terminal part of helix $\alpha 5$ and the PPIase domains (Fig. 8c and d) in agreement with the BUNCH models.

3.7. Functional characterization of the Par27 domains

The PPIase activity of full-length Par27 was previously characterized using two different assays, either by monitoring the catalyzed refolding of reduced and carboxymethylated RNase T1

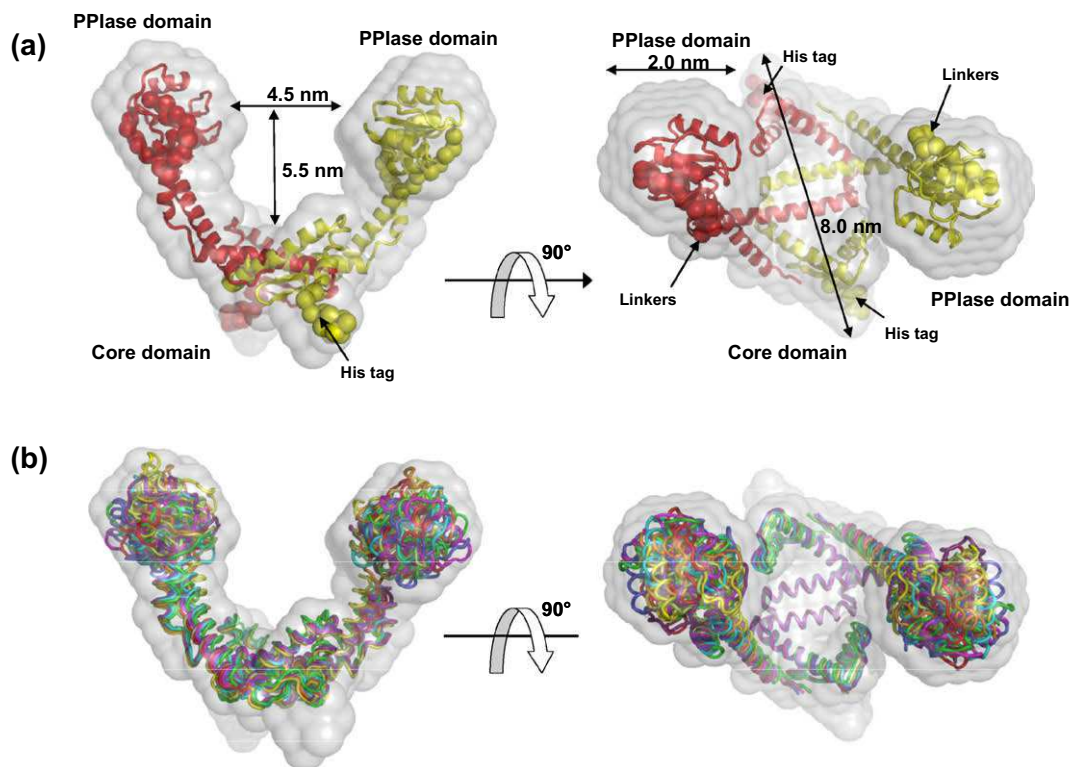


Fig. 6. Solution structure of Par27 and its domain organization. (a and b) Model generated by rigid-body modeling of the SAXS data using BUNCH. The ribbon representations depict the 'core' domain determined by X-ray crystallography and the PPIase domains predicted by homology modeling. The beads show the dummy residues added by BUNCH corresponding to the N-terminal His-tag and the two linkers tethering the PPIase domains to the NTD/CTD core.

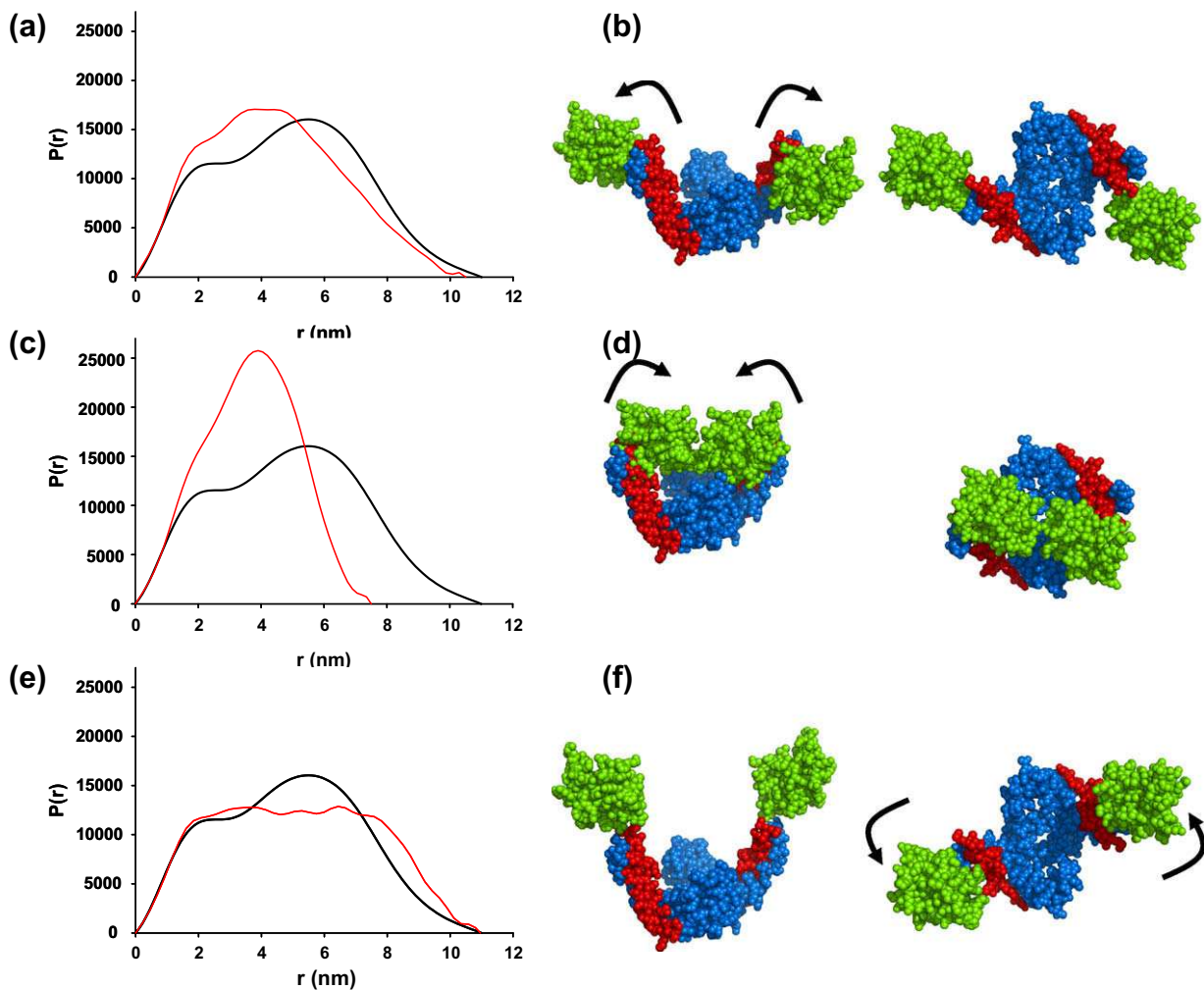


Fig. 7. Models of Par27 and $P(r)$ functions. (b, d, and f) The models were generated by moving the PPIase domains relative to the NTD/CTD core. (a, c, and e) The $P(r)$ function was calculated for the different models and compared with the experimental curve. The experimental curve is shown in black and the calculated curve in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

(RCM-T1), or by measuring the isomerisation rate of two peptidyl-prolyl bonds in a 16-mer peptide (Hodak et al., 2008). The activity of the isolated PPIase domain was significantly lower than that of full-length Par27 in both tests, and in particular the activity with the protein substrate was strongly affected. The k_{cat}/k_m determined for refolding of RCM-T1 catalyzed by the PPIase domain was reduced more than 140-fold as compared with full-length Par27 (Fig. 9a), whereas the exchange rate measured for the 16-mer peptide was reduced only fivefold (Fig. 9b). These differential behaviors suggested that the NTD/CTD core plays a role in binding large protein substrates or in stimulating the catalytic activity of the PPIase domains. It is noteworthy that no stable complex could be formed between denatured Fha30 and the isolated PPIase domain (data not shown), in contrast with what was observed with full-length Par27 (Hodak et al., 2008), supporting the involvement of the NTD/CTD core in binding protein substrates.

Previously, we showed that Par27 prevents the aggregation of lysozyme and forms a tight complex with Fha30, a 30 kDa fragment of *B. pertussis* haemagglutinin, only when Par27 was mixed with the unfolded form of Fha30 (Hodak et al., 2008). Here, we found that the isolated PPIase domain has no effect on lysozyme aggregation and that no complex between the PPIase domain and unfolded Fha30 could be detected by SEC, indicating that the NTD/CTD platform plays a major role in binding to the client protein.

4. Discussion

Par27 of *B. pertussis* belongs to a large group of bacterial folding catalysts that exhibit both chaperone and PPIase activities. All three families of PPIases – cyclophilins, FKBP and parvulins – comprise single domain as well as multi-domain proteins. In the parvulin family, Par10 and Par14 are single domain proteins, whereas Pin1, SurA or Par27 contain additional domains. Most often the additional domains act as chaperone and/or mediate the oligomerization of the protein (Saul et al., 2004). Par27 is the first example of a multi-domain protein containing a parvulin PPIase domain that is dimeric (Hodak et al., 2008).

In Par27, the NTD and CTD sub-domains mediate the dimerization, whereas the central domain carries the PPIase activity. CD spectroscopy and SAXS indicate that the isolated PPIase domain is properly folded, apart from the His-tag tail, but this isolated domain exhibits a reduced peptidyl-prolyl isomerase activity both toward a peptide substrate and a protein substrate when compared to the entire Par27. In both SurA and TF, the isolated active PPIase domain exhibits a reduced PPIase activity toward protein substrates but conserves the same activity than the full-length protein when assayed with a peptide substrate (Scholz et al., 1997b; Behrens et al., 2001). This suggests that the dimeric NTD/CTD core domain of Par27 plays a role in maintaining the proper arrangement of the PPIase active site residues and/or participates in substrate

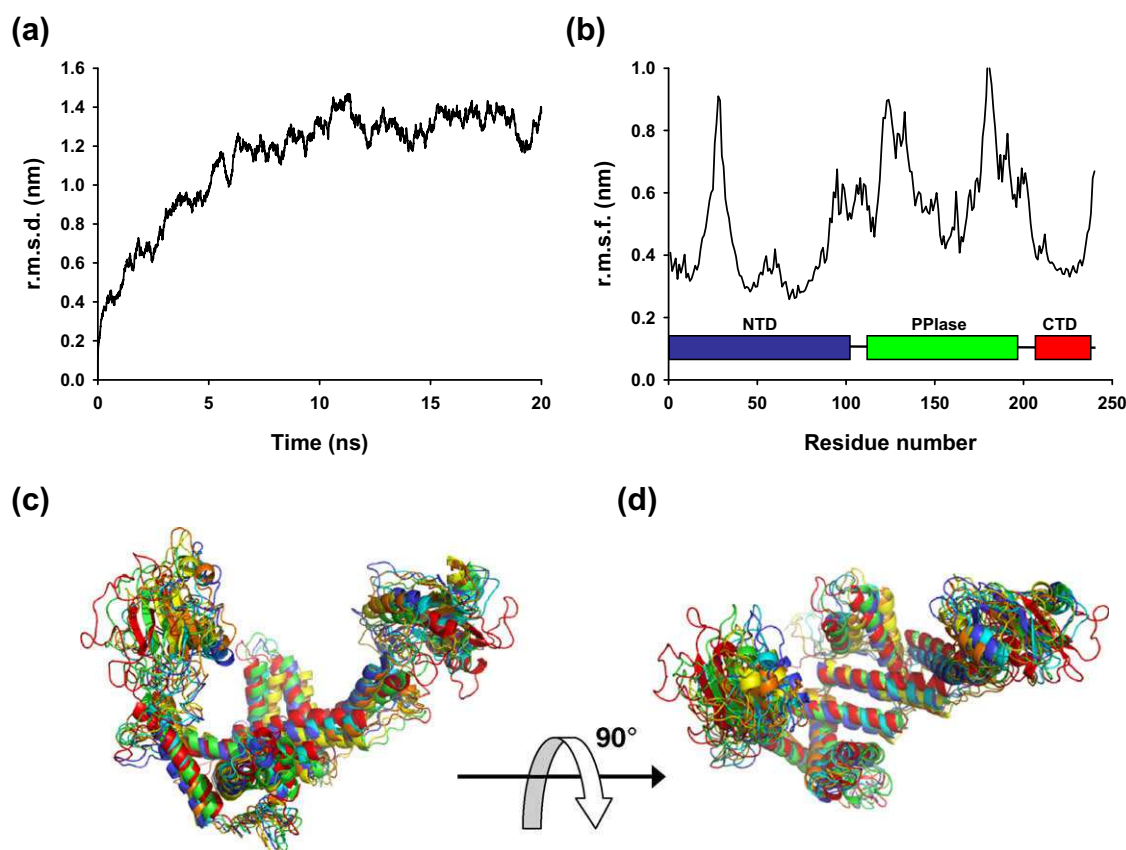


Fig. 8. Molecular dynamics simulation. (a) Root-mean square deviations (r.m.s.d.) of all C α atoms from the starting structure. (b) Root-mean square fluctuations (r.m.s.f.) of all C α atoms from their time-averaged positions. The inset shows the boundaries of the different domains and sub-domains of Par27. (c and d) Superposition of snapshots taken at 12, 14, 16, 18 and 20 ns of the MD simulation.

binding and/or optimizes the presentation of the substrate to the catalytic site.

In other multi-domain protein folding catalysts that exhibit both a protein folding catalytic activity and a chaperone activity, the two functions were assigned to distinct regions of the molecule. Structural comparison with these proteins suggests that the cradle-like NTD/CTD core of Par27 resembles the clamp-like structures found in other chaperone proteins (Stirling et al., 2006) and is a good candidate for the substrate-binding site. On one hand, the overall V-shape of the Par27 homodimer (Fig. 6) resembles that of other dimeric chaperones/folding catalysts. FkpA from the periplasm of *E. coli* (Saul et al., 2004), and Mip, a cell surface virulence factor of *Legionella pneumophila* (Riboldi-Tunncliffe et al., 2001) are chaperone/PPlase protein folding helpers that form V-shaped dimers with the C-terminal PPlase domain tethered to a central dimeric core by a long α -helix. In FkpA deletion studies showed that the chaperone activity resides in the N-terminal dimeric domain (Saul et al., 2004). Another folding catalyst, the disulfide isomerase DsbC from the periplasm of *E. coli* also forms a V-shaped homodimer consisting of separate thioredoxin domains linked to a dimeric core domain by a long α -helix (McCarthy et al., 2000). In DsbC, the core domain formed by the dimerization of NTDs carries the chaperone activity. There is no sequence or structural similarity between the NTD/CTD core of Par27 and the corresponding regions of other dimeric protein folding helpers, apart from the overall V-shape structure, suggesting that these clamp-like structures result from a convergent evolution because they provide an efficient mechanism for seizing large protein substrates, preventing their aggregation, and giving them access to the catalytic domains (Stirling et al., 2006).

On another hand, the dimeric NTD/CTD core domain of Par27 has a topological organization similar to the chaperone substrate-binding domain of monomeric proteins such as SurA and trigger factor. In SurA, the chaperone activity was found to be independent of the presence of the active PPlase 2 domain (Behrens et al., 2001; Bitto and McKay, 2002). The putative substrate-binding site, suggested from intermolecular packing interactions observed in the crystal, is located in the core domain and is made of three helices (α 1, α 2 and α 5) of the NTD, forming a crevice where a short peptide can bind (Bitto, 2002 #4631; Bitto, 2003 #4630). In TF, the chaperone activity is also independent of the presence of the catalytic PPlase activity (Scholz et al., 1998; Kramer et al., 2004a), and the N-terminal and the two 'arm' domains form a cradle-like structure that was suggested to be important for substrate binding (Ferbitz et al., 2004; Stirling et al., 2006), although the location of the chaperone activity could not be clearly assigned to one of the domains (Kramer et al., 2004b). The topological similarity suggests a common ancestry with these monomeric PPlase/chaperone proteins. The dimerization of Par27 could represent an adaptation to its substrates, potentially providing a larger binding surface for its client proteins. In *B. pertussis*, Par27 shows affinity for different amphipathic β -rich proteins or protein domains of outer membrane proteins (Hodak et al., 2008). *In vitro*, it accepts Fha30, a soluble fragment of FHA, only in its unfolded form but not in its native form (Hodak et al., 2008). The dimensions of the cradle-like groove are 8.0×4.5 nm, with a depth of 5.5 nm, sufficient to bind a protein domain or to hide from the solvent a large part of a substrate the size of a partially folded Fha30. The motif recognized in the client protein by Par27 is unknown, but is certainly related to the character of the binding surface. The surface

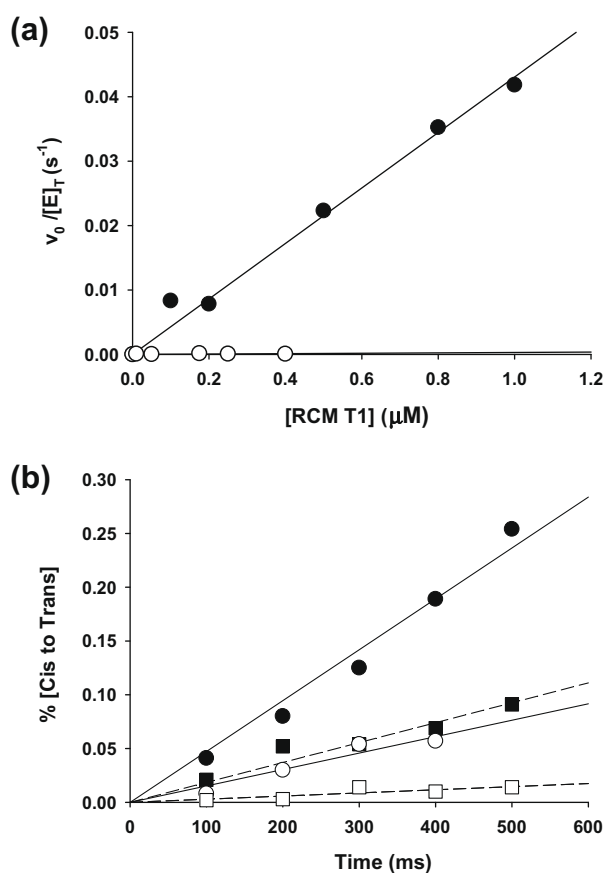


Fig. 9. PPIase activity assays. (a) PPIase activity using an RCM-T1 refolding assay. Refolding of denatured RCM-T1 was initiated by dilution in 2.0 M NaCl. Refolding kinetics was followed by the change of fluorescence at 320 nm at 15 °C as a function of RCM-T1 concentration. Initial velocities were measured in the presence and in the absence of Par27 (filled circles) or of the PPIase domain (open circles). The contribution from the uncatalyzed folding reaction increases linearly with RCM-T1 concentration, and the initial velocities in the presence of Par27 were corrected for the uncatalyzed contribution. Value of $300 \pm 100 \text{ M}^{-1} \text{ s}^{-1}$ was obtained for $k_{\text{cat}}/k_{\text{m}}$ from the slope of the line. The data for full-length Par27 are taken from (Hodak et al., 2008). (b) PPIase activity using a 16-mer peptide containing two Ser-Pro motifs. The peptide was solubilized at a concentration of 2 mM in 50 mM sodium phosphate (pH 6.6) and 50 mM NaCl. The ratios of molecules that undergo from cis to trans conformation [%[cis → trans]] for Pro5 (circles) and Pro13 (squares) residues in the presence of 2.5% Par27 are given as a function of time for Par27 (filled symbols) and for the PPIase domain (open symbols). The data for full-length Par27 are taken from Hodak et al. (2008).

of the cradle-like cavity is mainly positive, with a patch of negative potential on each side (Fig. S4a in Supplementary material). Also, this cavity in Par27 contains hydrophobic residues, forming putative hydrophobic pockets in the ‘armpits’ on each side of the core platform that could accommodate non-polar side chains (Fig. S4b in Supplementary material).

Although Par27 was able to crystallize, the structure of the PPIase domain was not seen in the electron density map, and the structures of the C-terminal part of NTD and the N-terminal part of CTD were not well-defined. These results presented here support a model of Par27 where the PPIase domains are tethered to the NTD/CTD core by flexible linkers and undergo limited motions around the shaft (Fig. 6b). Such flexibility could explain the missing electron density in the crystal structure and could be relevant to the biological activity of the protein. Flexibility appears to be a common feature among the multi-domain chaperone/folding catalysts. In monomeric SurA as well as in dimeric Par27, FkpA, Mip or DsbC, the catalytic domain is tethered to the core by a long α -helix or an extended chain that allows motions between the core and

catalytic domains. SurA is stabilized in the crystal by packing interactions with neighboring molecules, however, in the presence of peptide substrates, the protein adapts its structure and binds peptides in different conformations (Xu et al., 2007). Different crystal forms of FkpA and DsbC have also revealed flexibility in the orientation of the C-terminal domains located at the extremities of the V-shape structures (Saul et al., 2004), and the plasticity of FkpA has been confirmed by NMR (Hu et al., 2006). The results reported here for Par27 suggest a similar plasticity that could allow the binding to substrates of different size or the adaptation to substrate size during folding.

Par27 is an abundant periplasmic protein of *B. pertussis* that might serve as a general periplasmic chaperone partially overlapping SurA functions or serving as a back-up to SurA in the periplasm of *B. pertussis* (Hodak et al., 2008). While SurA appears to be essential in *B. pertussis*, the inactivation of the *par27* gene has a small but detectable effect on bacterial growth, indicating that Par27 contributes to bacterial fitness. As for other PPIase/chaperones, an *in vivo* role for the PPIase activity has not yet been demonstrated. In *E. coli*, a SurA variant lacking the active parvulin domain (PPIase 2) complements a null mutation, arguing that the chaperone rather than the PPIase activity of SurA is important *in vivo* (Behrens et al., 2001). It can nevertheless not be ruled out that Par27 also contributes to protein folding in the periplasm by its enzymatic activity.

Acknowledgments

This work was supported in part by a Grant “ANR MIME 07 TPS Path” to F.J.D. The authors thank Dr. H. Drobecq for his help with mass spectrometry analyses, Ambroise Desfosses for his help in writing scripts and Dr. G. Zaccai for his help with SANS experiments. We thank Jean-Paul Declercq, Miguel Ortiz-Lombardia, Christian Cambillau, Javier Perez, Markus Rudolph and Coralie Bompard for helpful discussions. C.L. was supported by a MENRT predoctoral fellowship from the French government. F.J.-D. is a researcher of the CNRS.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jsb.2009.11.007.

References

- Bayer, E., Goetsch, S., Mueller, J.W., Griewel, B., Guiberman, E., Mayr, L.M., Bayer, P., 2003. Structural analysis of the mitotic regulator hPin1 in solution: insights into domain architecture and substrate binding. *J. Biol. Chem.* 278, 26183–26193.
- Behrens, S., Maier, R., de Cock, H., Schmid, F.X., Gross, C.A., 2001. The SurA periplasmic PPIase lacking its parvulin domains functions *in vivo* and has chaperone activity. *EMBO J.* 20, 285–294.
- Bitto, E., McKay, D.B., 2002. Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins. *Structure* 10, 1489–1498.
- Bricogne, G., Vornrhein, C., Flensburg, C., Schiltz, M., Paciorek, W., 2003. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D Biol. Crystallogr.* 59, 2023–2030.
- Clantin, B., Hodak, H., Willery, E., Loch, C., Jacob-Dubuisson, F., Villeret, V., 2004. The crystal structure of filamentous hemagglutinin secretion domain and its implications for the two-partner secretion pathway. *Proc. Natl. Acad. Sci. USA* 101, 6194–6199.
- Clantin, B., Delattre, A.S., Rucktooa, P., Saint, N., Meli, A.C., Loch, C., Jacob-Dubuisson, F., Villeret, V., 2007. Structure of the membrane protein FhaC: a member of the Omp85-TpsB transporter superfamily. *Science* 317, 957–961.
- Collaborative Computational Project, Number 4, 1994. The CCP4 Suite: Programs for Protein Crystallography, 1994. *Acta Crystallogr. D Biol. Crystallogr.* 50, 760–763.
- Doniach, S., 2001. Changes in biomolecular conformation seen by small angle X-ray scattering. *Chem. Rev.* 101, 1763–1778.
- Doniach, S., Basile, J., Garel, T., Orland, H., 1995. Partially folded states of proteins: characterization by X-ray scattering. *J. Mol. Biol.* 254, 960–967.

- Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E., Ban, N., 2004. Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* 431, 590–596.
- Gérard, F., Ribeiro, E., Albertini, A., Zaccai, G., Ebel, C., Ruigrok, R., Jamin, M., 2007. Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution. *Biochemistry* 46, 10328–10338.
- Glatter, O., Kratky, O., 1982. *Small Angle X-ray Scattering*. Academic Press, London, UK.
- Göthel, S.F., Marahiel, M.A., 1999. Peptidyl-prolyl *cis*–*trans* isomerases, a superfamily of ubiquitous folding catalysts. *Cell. Mol. Life Sci.* 55, 423–436.
- Hess, B., Kutzner, C., Van der Spoel, D., Lindahl, E., 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4, 435–447.
- Hodak, H., Wohlkönig, A., Smet-Nocca, C., Drobecq, H., Wieruszski, J.M., Senechal, M., Landrieu, I., Locht, C., Jamin, M., Jacob-Dubuisson, F., 2008. The peptidyl-prolyl isomerase and chaperone Par27 of *Bordetella pertussis* as the prototype for a new group of parvulins. *J. Mol. Biol.* 376, 414–426.
- Holm, L., Sander, C., 1996. Mapping the protein universe. *Science* 273, 595–603.
- Hu, K., Galius, V., Pervushin, K., 2006. Structural plasticity of peptidyl-prolyl isomerase sFkpA is a key to its chaperone function as revealed by solution NMR. *Biochemistry* 45, 11983–11991.
- ILL, T.y.b.o., 2005. The yellow book of ILL: a guide to neutron research facilities.
- Jacob-Dubuisson, F., Fernandez, R., Coutte, L., 2004. Protein secretion through autotransporter and two-partner pathways. *Biochim. Biophys. Acta* 1694, 235–257.
- Jacrot, B., Zaccai, G., 1981. Determination of molecular weight by neutron scattering. *Biopolymers* 20, 2413–2426.
- Kabsch, W., 1993. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* 26, 795–800.
- Kajava, A.V., Cheng, N., Cleaver, R., Kessel, M., Simon, M.N., Willery, E., Jacob-Dubuisson, F., Locht, C., Steven, A.C., 2001. Beta-helix model for the filamentous haemagglutinin adhesin of *Bordetella pertussis* and related bacterial secretory proteins. *Mol. Microbiol.* 42, 279–292.
- Konarev, P.V., Volkov, V.V., Sokolova, A., M.H.J., K., Svergun, D.I., 2003. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* 36, 1277–1282.
- Kramer, G., Patzelt, H., Rauch, T., Kurz, T.A., Vorderwulbecke, S., Bukau, B., Deuerling, E., 2004a. Trigger factor peptidyl-prolyl *cis/trans* isomerase activity is not essential for the folding of cytosolic proteins in *Escherichia coli*. *J. Biol. Chem.* 279, 14165–14170.
- Kramer, G., Rutkowska, A., Wegrzyn, R.D., Patzelt, H., Kurz, T.A., Merz, F., Rauch, T., Vorderwulbecke, S., Deuerling, E., Bukau, B., 2004b. Functional dissection of *Escherichia coli* trigger factor: unraveling the function of individual domains. *J. Bacteriol.* 186, 3777–3784.
- Krisninel, E., Henrick, K., 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2256–2268.
- Kühlewein, A., Voll, G., Hernandez Alvarez, B., Kessler, H., Fischer, G., Rahfeld, J.U., Gemmecker, G., 2004. Solution structure of *Escherichia coli* Par10: the prototypic member of the Parvulin family of peptidyl-prolyl *cis/trans* isomerases. *Protein Sci.* 13, 2378–2387.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283–291.
- Lazar, S.W., Kolter, R., 1996. SurA assists the folding of *Escherichia coli* outer membrane proteins. *J. Bacteriol.* 178, 1770–1773.
- Lazar, S.W., Almiron, M., Tormo, A., Kolter, R., 1998. Role of the *Escherichia coli* SurA protein in stationary-phase survival. *J. Bacteriol.* 180, 5704–5711.
- Li, Z., Li, H., Devasahayam, G., Gemmill, T., Chaturvedi, V., Hanes, S.D., Van Roey, P., 2005. The structure of the *Candida albicans* Ess1 prolyl isomerase reveals a well-ordered linker that restricts domain mobility. *Biochemistry* 44, 6180–6189.
- Maier, T., Ferbitz, L., Deuerling, E., Ban, N., 2005. A cradle for new proteins: trigger factor at the ribosome. *Curr. Opin. Struct. Biol.* 15, 204–212.
- Maupetit, J., Gautier, R., Tuffery, P., 2006. SABBAC: online Structural Alphabet-based Protein Backbone reconstruction from Alpha-Carbon trace. *Nucleic Acids Res.* 34, W147–W151.
- McCarthy, A.A., Haebel, P.W., Torronen, A., Rybin, V., Baker, E.N., Metcalf, P., 2000. Crystal structure of the protein disulfide bond isomerase, DsbC, from *Escherichia coli*. *Nat. Struct. Biol.* 7, 196–199.
- Merz, F., Hoffmann, A., Rutkowska, A., Zachmann-Brand, B., Bukau, B., Deuerling, E., 2006. The C-terminal domain of *Escherichia coli* trigger factor represents the central module of its chaperone activity. *J. Biol. Chem.* 281, 31963–31971.
- Miot, M., Betton, J.M., 2004. Protein quality control in the bacterial periplasm. *Microb. Cell Fact.* 3, 4.
- Missiakas, D., Betton, J.M., Raina, S., 1996. New components of protein folding in extracytoplasmic compartments of *Escherichia coli* SurA, FkpA and Skp/OmpH. *Mol. Microbiol.* 21, 871–884.
- Mücke, M., Schmid, F.X., 1994. Intact disulfide bonds decelerate the folding of ribonuclease T1. *J. Mol. Biol.* 239, 713–725.
- Narayanan, T., Diat, O., Bösecke, P., 2001. SAXS and USAXS on the high brilliance beamline at the ESRF. *Nucl. Instrum. Methods Phys. Res. Sect. A* 467–468, 1005–1009.
- Oostenbrink, C., Soares, T.A., van der Vegt, N.F., van Gunsteren, W.F., 2005. Validation of the 53A6 GROMOS force field. *Eur. Biophys. J.* 34, 273–284.
- Petoukhov, M.V., Svergun, D.I., 2005. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* 89, 1237–1250.
- Putnam, C.D., Hammel, M., Hura, G.L., Tainer, J.A., 2007. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* 40, 191–285.
- Rahfeld, J.U., Schierhorn, A., Mann, K., Fischer, G., 1994. A novel peptidyl-prolyl *cis/trans* isomerase from *Escherichia coli*. *FEBS Lett.* 343, 65–69.
- Ranganathan, R., Lu, K.P., Hunter, T., Noel, J.P., 1997. Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. *Cell* 89, 875–886.
- Riboldi-Tunncliffe, A., König, B., Jessen, S., Weiss, M.S., Rahfeld, J., Hacker, J., Fischer, G., Hilgenfeld, R., 2001. Crystal structure of Mip, a prolylisomerase from *Legionella pneumophila*. *Nat. Struct. Biol.* 8, 779–783.
- Rouviere, P.E., Gross, C.A., 1996. SurA, a periplasmic protein with peptidyl-prolyl isomerase activity, participates in the assembly of outer membrane porins. *Genes Dev.* 10, 3170–3182.
- Saul, F.A., Arié, J.P., Vulliez-le Normand, B., Kahn, R., Betton, J.M., Bentley, G.A., 2004. Structural and functional studies of FkpA from *Escherichia coli*, a *cis/trans* peptidyl-prolyl isomerase with chaperone activity. *J. Mol. Biol.* 335, 595–608.
- Scholz, C., Rahfeld, J., Fischer, G., Schmid, F.X., 1997a. Catalysis of protein folding by parvulin. *J. Mol. Biol.* 273, 752–762.
- Scholz, C., Stoller, G., Zarnt, T., Fischer, G., Schmid, F.X., 1997b. Cooperation of enzymatic and chaperone functions of trigger factor in the catalysis of protein folding. *EMBO J.* 16, 54–58.
- Scholz, C., Mücke, M., Rape, M., Pecht, A., Pahl, A., Bang, H., Schmid, F.X., 1998. Recognition of protein substrates by the prolyl isomerase trigger factor is independent of proline residues. *J. Mol. Biol.* 277, 723–732.
- Schulze-Gahmen, U., Aono, S., Chen, S., Yokota, H., Kim, R., Kim, S.H., 2005. Structure of the hypothetical *Mycoplasma* protein MPN555 suggests a chaperone function. *Acta Crystallogr. D Biol. Crystallogr.* 61, 1343–1347.
- Sekerina, E., Rahfeld, J.U., Müller, J., Fanghanel, J., Rascher, C., Fischer, G., Bayer, P., 2000. NMR solution structure of hPar14 reveals similarity to the peptidyl prolyl *cis/trans* isomerase domain of the mitotic regulator hPin1 but indicates a different functionality of the protein. *J. Mol. Biol.* 301, 1003–1017.
- Semenyuk, A.V., Svergun, D., 1991. GNOM – a program package for small-angle scattering data processing. *J. Appl. Crystallogr.* 24, 537–540.
- Sklar, J.G., Wu, T., Kahne, D., Silhavy, T.J., 2007. Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in *Escherichia coli*. *Genes Dev.* 21, 2473–2484.
- Stirling, P.C., Bakhoum, S.F., Feigl, A.B., Leroux, M.R., 2006. Convergent evolution of clamp-like binding sites in diverse chaperones. *Nat. Struct. Mol. Biol.* 13, 865–870.
- Svergun, D.I., Koch, M.H.J., 2003. Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.* 66, 1735–1782.
- Svergun, D., Barberato, C., Koch, M.H., 1995. CRYSOLO – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* 28, 768–773.
- Terada, T., Shirouzu, M., Fukumori, Y., Fujimori, F., Ito, Y., Kigawa, T., Yokoyama, S., Uchida, T., 2001. Solution structure of the human parvulin-like peptidyl prolyl *cis/trans* isomerase, hPar14. *J. Mol. Biol.* 305, 917–926.
- Tossavainen, H., Permi, P., Purhonen, S.L., Sarvas, M., Kilpeläinen, I., Seppälä, R., 2006. NMR solution structure and characterization of substrate binding site of the PPIase domain of PrsA protein from *Bacillus subtilis*. *FEBS Lett.* 580, 1822–1826.
- Trapani, S., Navaza, J., 2008. AMOR: classical and modern. *Acta Crystallogr. D Biol. Crystallogr.* 64, 11–16.
- Tsutakawa, S.E., Hura, G.L., Frankel, K.A., Cooper, P.K., Tainer, J.A., 2007. Structural analysis of flexible proteins in solution by small angle X-ray scattering combined with crystallography. *J. Struct. Biol.* 158, 214–223.
- Uson, I., Sheldrick, G.M., 1999. Advances in direct methods for protein crystallography. *Curr. Opin. Struct. Biol.* 9, 643–648.
- Volkov, V.V., Svergun, D.I., 2003. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* 36, 860–864.
- Wall, M.E., Ealick, S.E., Gruner, S.M., 1997. Three-dimensional diffuse x-ray scattering from crystals of Staphylococcal nuclease. *Proc. Natl. Acad. Sci. USA* 94, 6180–6184.
- Wilson, M.A., Brunger, A.T., 2003. Domain flexibility in the 1.75 Å resolution structure of Pb2+-calmodulin. *Acta Crystallogr. D Biol. Crystallogr.* 59, 1782–1792.
- Wohlkönig, A., Hodak, H., Clantin, B., Senechal, M., Bompard, C., Jacob-Dubuisson, F., Villeret, V., 2008. Crystallization and preliminary X-ray diffraction analysis of the peptidylprolyl isomerase Par27 of *Bordetella pertussis*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 64, 809–812.
- Wu, S., Zhang, Y., 2007. LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res.* 35, 3375–3382.
- Wülfing, C., Plückthun, A., 1994. Protein folding in the periplasm of *Escherichia coli*. *Mol. Microbiol.* 12, 685–692.
- Wyatt, P.J., 1998. Submicrometer particle sizing by multiangle light scattering following fractionation. *J. Colloid Interface Sci.* 197, 9–20.
- Xu, X., Wang, S., Hu, Y.X., McKay, D.B., 2007. The periplasmic bacterial molecular chaperone SurA adapts its structure to bind peptides in different conformations to assert a sequence preference for aromatic residues. *J. Mol. Biol.* 373, 367–381.

2.3.2 Modélisation de Fha30

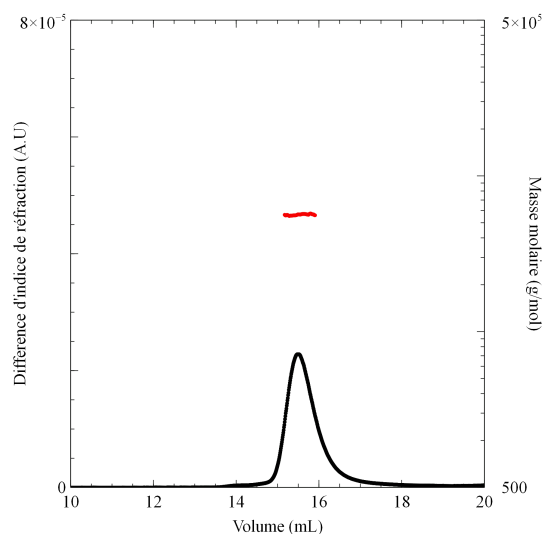


FIGURE 2.2 – Analyse de SEC-MALLS-RI et masse calculée de Fha30 native, le profil de chromatographie est présenté en noir, la masse calculée à partir de la diffusion de lumière en rouge.

La structure de Fha30 a été résolue par cristallographie [Clantin 2004] dans une forme dimérique. En solution comme dans le cristal, la protéine forme des dimères dans lesquels l'hélice β d'un monomère coiffe et prolonge l'hélice β de l'autre monomère. Toutefois, le dernier brin β d'un des monomères est exclu de cette structure et existe sous forme flexible car il n'est pas visible sur les cartes de densité électronique. Afin d'avoir la conformation de cette protéine en solution, j'ai entrepris une étude basée sur la technique de diffusion de rayons X aux petits angles.

J'ai tout d'abord effectué une expérience de SEC-MALLS-RI afin de confirmer l'état d'oligomérisation (figure 2.2). Cette analyse nous montre que cette protéine forme des dimères en solution, comme dans le cristal [Clantin 2004]. J'ai donc utilisé ce dimère cristallographique, auquel j'ai rajouté le dernier brin β , sous une forme étendue (figure 2.4).

J'ai ensuite lancé un calcul de dynamique moléculaire, afin de générer un ensemble de conformations qui représente l'espace conformationnel accessible à la protéine. La distribution des valeurs de R_g dans l'ensemble initial est centrée sur une valeur de 34.4 Å. L'ensemble sélectionné qui représente la courbe expérimentale est centré sur une valeur de R_g de 33.5 Å en accord avec la valeur expérimentale déterminée à partir du graphe de Guinier. Les courbes théorique de l'ensemble et expérimentale sont montrées figure 2.4a. Le fait d'utiliser cette procédure permet d'améliorer le χ^2 ($\chi^2 = 2$ avec l'ensemble alors que le χ^2 de départ est de 2.2) que très légèrement, car la taille du brin manquant est trop faible pour voir une grande différence. Les rayons de giration de l'ensemble et celui du dimère cristallographique

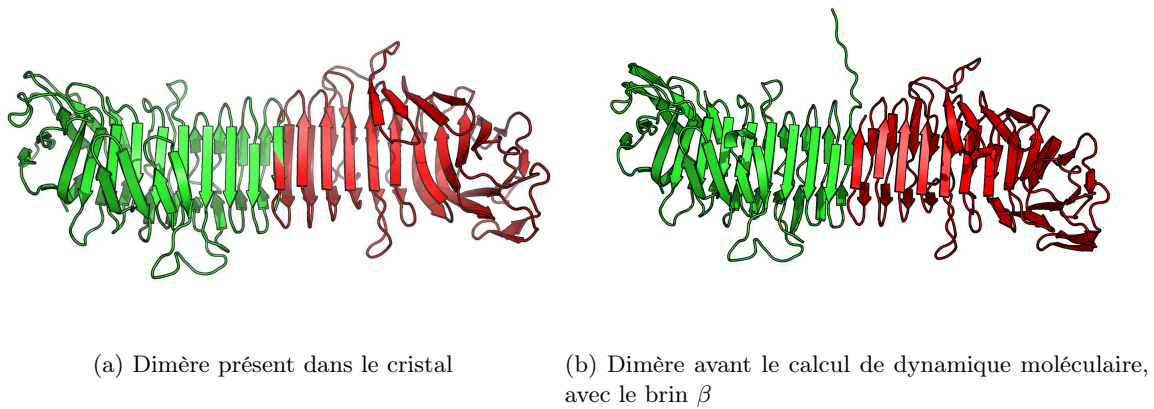


FIGURE 2.3 – Modèles de Fha30 native

ont la même valeur. L'ensemble sélectionné est composé exclusivement de modèles avec les brin en position étendue (figure 2.4b), et ce malgré la présence de modèles avec le feuillet rabattu sur l'hélice dans l'ensemble de départ.

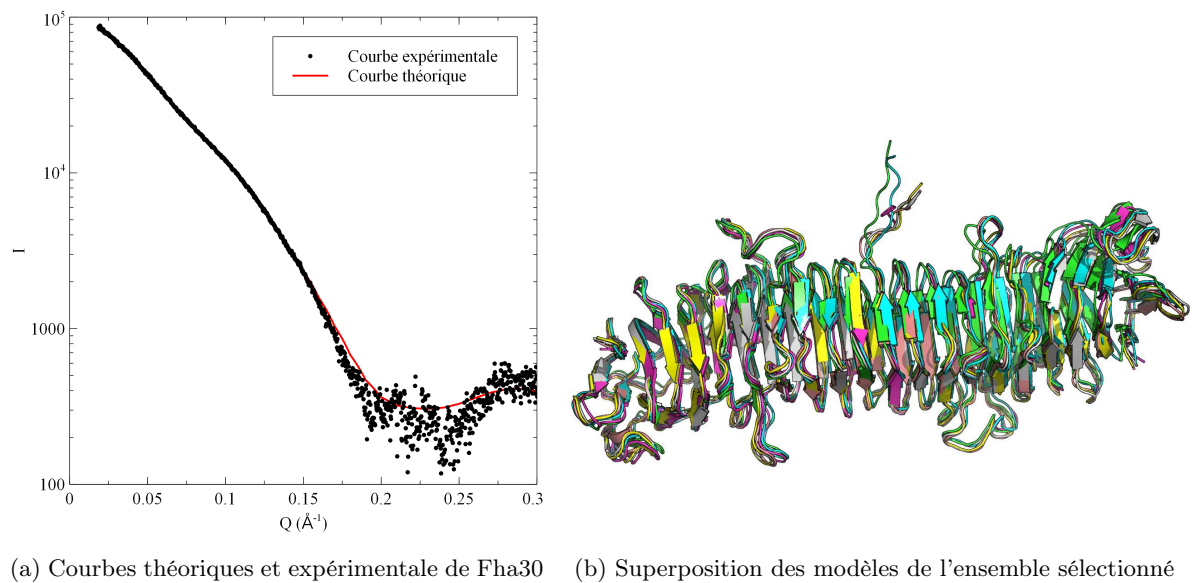


FIGURE 2.4 – Résultats de la procédure de sélection d'ensemble

2.3.3 Modélisation de Fha30 dénaturé

La forme de Fha30 dénaturée par Gu-HCl est monomérique, comme indiqué par les expérience de SEC-MALLS-RI, qui nous permet de calculer une masse de 30 kDa (figure 2.5). J'ai donc entrepris de modéliser Fha30 sous sa forme dénaturée dans le chlorure de Guanidinium, afin de comprendre son comportement, mais aussi

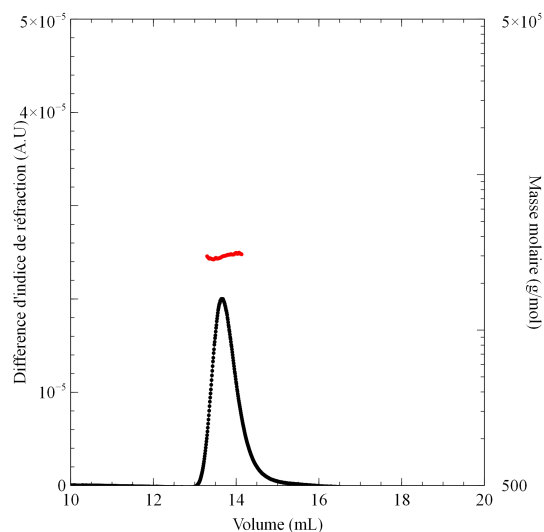


FIGURE 2.5 – Profils de SEC-RI-MALLS et masse calculée de Fha30 dénaturée, le profil de chromatographie est présenté en noir, la masse calculée en rouge.

pour obtenir des modèles que je pourrais utiliser pour une procédure d'arrimage moléculaire sur Par27 et ainsi obtenir un modèle du complexe.

La superposition des courbes de diffusion expérimentale et théorique calculée à

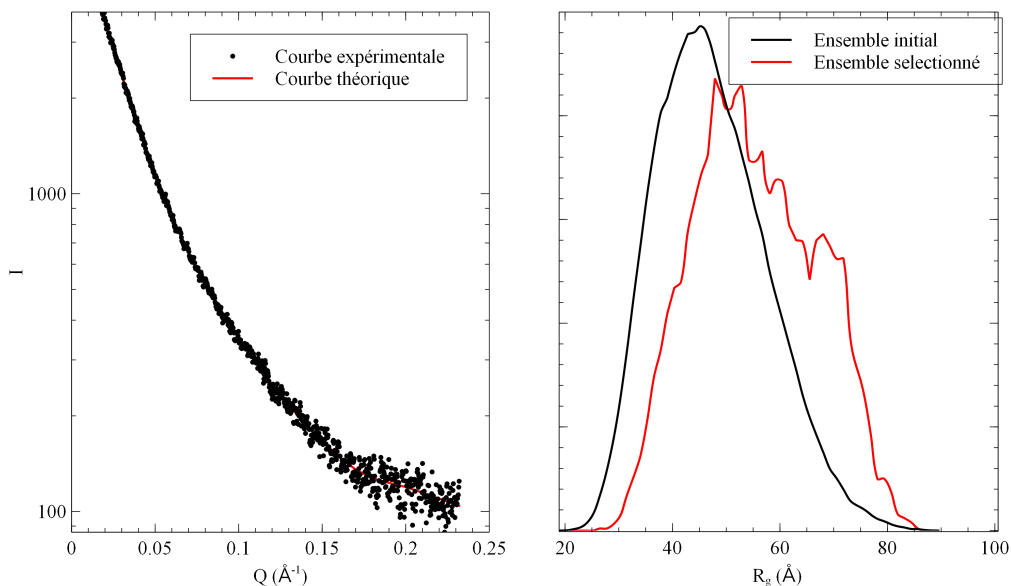


FIGURE 2.6 – courbes de diffusion expérimentales et de l'ensemble final et distribution des rayons de gyration de l'ensemble de départ et de l'ensemble sélectionné.

partir de l'ensemble , ainsi que la distribution des rayons de gyration des ensembles

sélectionnées et de départ sont montrés figure 2.6. L'ensemble initial de 10000 modèles générés par Ranch a une distribution quasi gaussienne centrée sur 45 Å. L'ensemble sélectionné par Gajoe présente une distribution d'une largeur similaire à celle de l'ensemble initial mais décalée vers des valeurs de R_g plus grandes.

Ce phénomène est typique des protéines dénaturées chimiquement, qui ont tendance à prendre des formes plus étendues que les chaînes intrinsèquement désordonnées [Bernado 2009], et ce à cause du mécanisme d'action des dénaturants : ils peuvent former des interactions avec la chaîne principale [Schellman 1958] et donc ainsi favoriser des conformations plus étendues.

2.3.4 Modélisation du complexe Fha30/Par27

Par 27 forme un complexe avec Fha30 dans sa forme dénaturée, et possède un pouvoir protecteur contre la protéolyse et l'agrégation, mais aussi d'une activité PPIase [Hodak 2008]. Nous avons décidé d'entreprendre des études structurales afin de comprendre où la protéine client se positionne sur la chaperonne, et quel est l'état de compaction de la protéine client dans le complexe.

2.3.4.1 Enveloppe basse résolution

L'analyse de la courbe de diffusion du complexe nous permet de calculer un rayon de giration de 46 Å, qui est plus grand que le R_g de la chaperonne seule qui est de 39 Å [Clantin 2010], mais plus faible que celui de Fha30 dénaturé, qui est de 51 Å. On peut donc supposer que dans le complexe il y a compaction partielle de Fha30. Pour confirmer cette hypothèse, nous avons entrepris des expériences de SANS en utilisant la technique de variation de contraste qui pourrait nous fournir des indications sur la conformation de chacun des composants dans le complexe. Les expériences de SANS n'ont pas pu être réalisées parce que le complexe Par27 deutéré (D-Par27)/Fha30 et D-Par27 s'agrègent en présence de D_2O (figure 2.7).

Avant de générer des modèles tout-atomes, j'ai tout d'abord commencé par essayer de produire des modèles à basse résolution composés de billes [Franke 2009]. J'ai d'abord produit 20 modèles, dont j'ai fait la moyenne [Volkov 2003] qui est montrée figure 2.8. Les différents modèles sont homogènes, avec une valeur de NSD (qui est un indicateur de la similarité entre les modèles [Kozin 2001]) de 1.1, indiquant une bonne similitude entre les enveloppes. On voit que sur cette enveloppe que l'on retrouve la forme de pince de Par27 et on peut supposer que Fha30 apparaît sous la forme d'un cylindre.

Néanmoins, on peut se poser des questions sur ce type de démarche : l'analyse de la courbes en représentation de Kratky [Kratky 1982] nous montre une courbe qui est caractéristique des protéines contenant des zones structurées mais aussi des parties désordonnées : la courbe présentée figure 2.9 avec deux maxima à 0.05 et 0.12 Å⁻¹ typique de ce genre de protéines [Putnam 2007]. Il est donc apparu plus judicieux de représenter ce complexe par un ensemble de structures que par un modèle unique. C'est pour cette raison que j'ai décidé de combiner des techniques

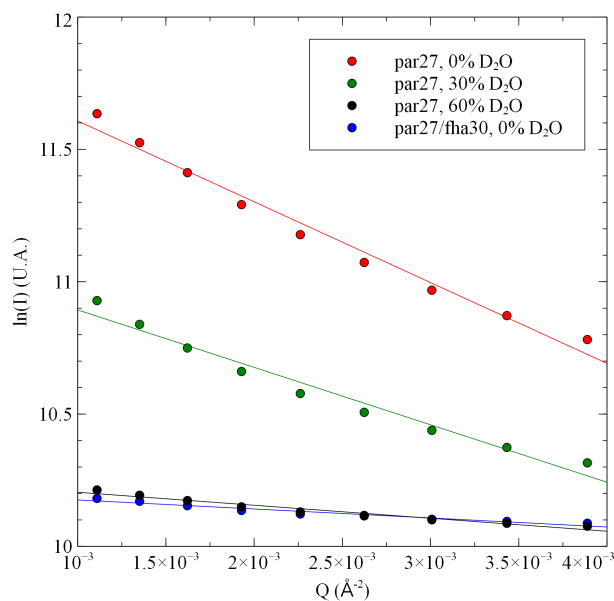


FIGURE 2.7 – Courbes de diffusion de neutrons de Par27 et du complexe Par27/Fha30 à différentes concentrations de D₂O en représentation de Guinier.

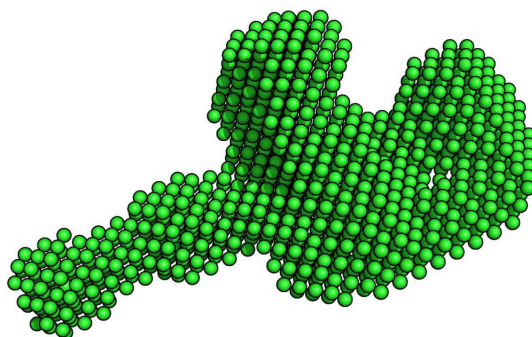


FIGURE 2.8 – enveloppe basse résolution du complexe Fha30/Par27

de modélisation par un ensemble de conformations.

2.3.4.2 Arrimage moléculaire

Les modèles tout atome de Fha30 dénaturé ont été utilisés dans une procédure d'arrimage sur différentes conformations de Par27 et les modèles générés par cette méthodes ont été confrontés aux résultats expérimentaux obtenus en diffusion aux

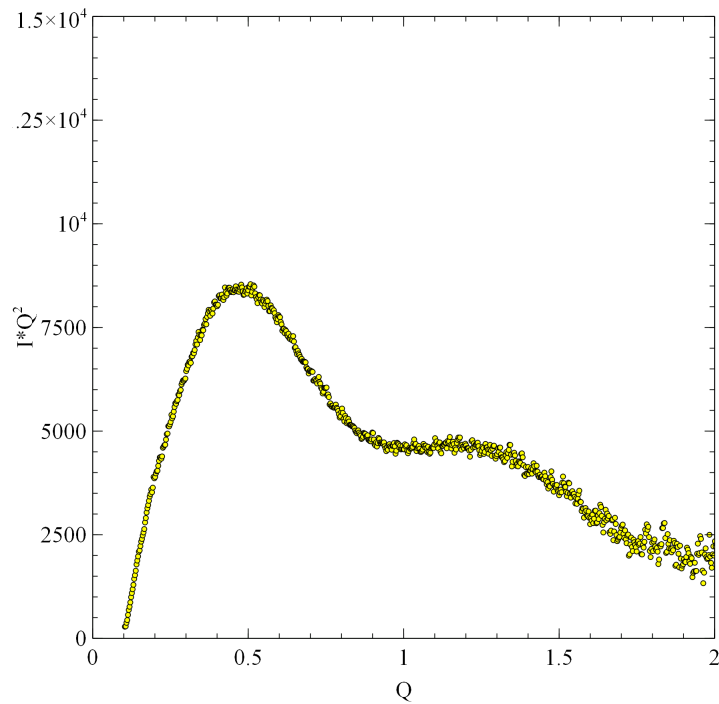


FIGURE 2.9 – Courbe de diffusion du complexe Fha30/Par27 en représentation de Kratky.

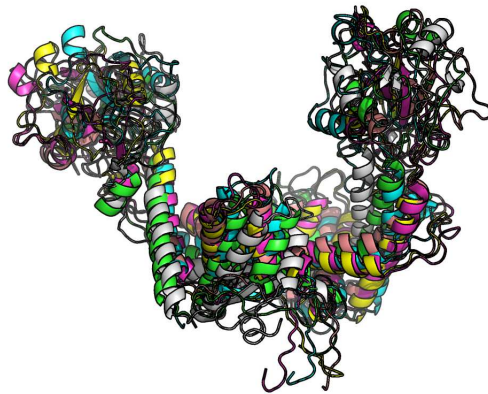


FIGURE 2.10 – Superposition des 6 modèles de Par27 utilisés pour l'arrimage moléculaire

petits angles. Cette procédure est illustrée sur la figure 2.11.

Les six modèles de Par27 utilisés dans cette procédure correspondent aux modèles qui avaient été utilisés pour modéliser les données de diffusion aux petits angles

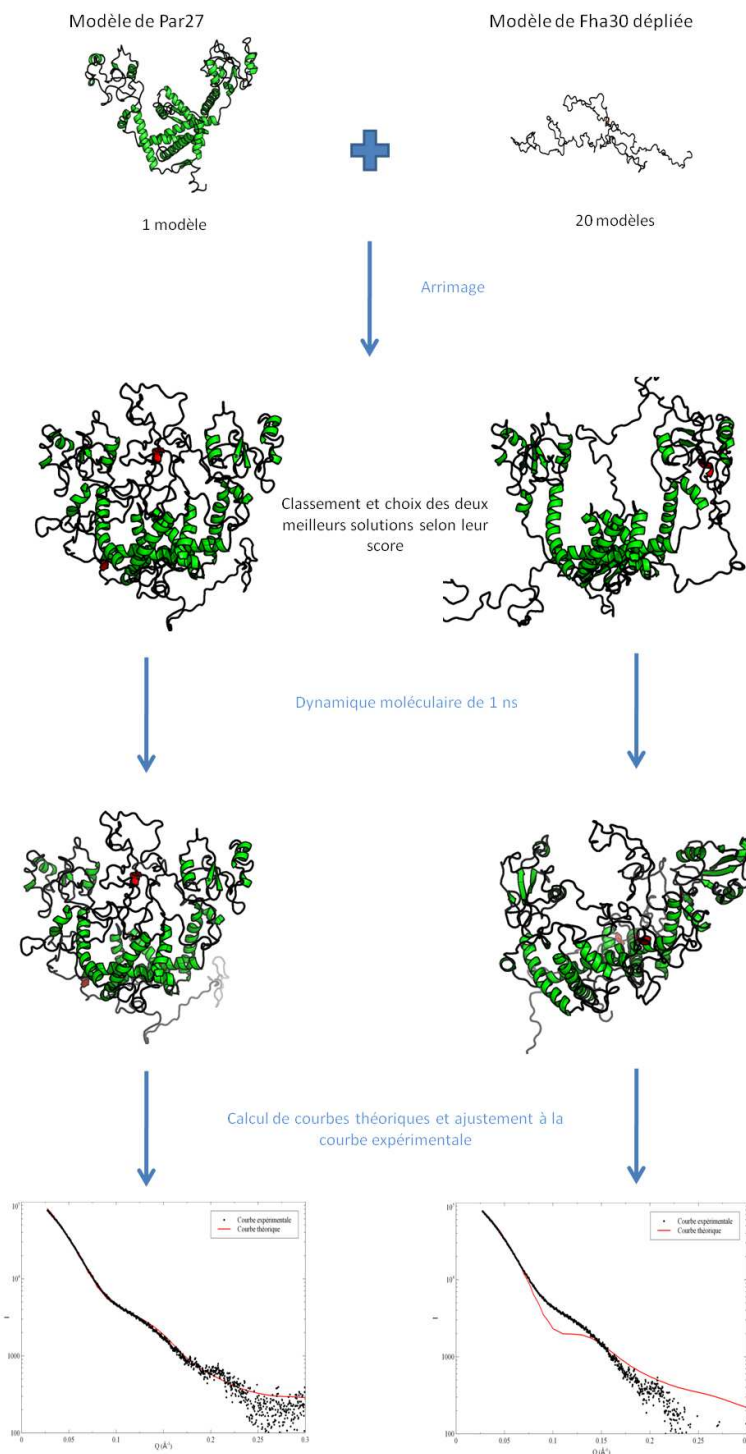


FIGURE 2.11 – Schéma représentant la procédure d'arrimage moléculaire et de sélection d'ensembles : tout d'abord les 20 modèles de Fha30 dépliés sont arrimés sur Par27 avec le programme BIGGER [Krippahl 2003]. Ensuite, les différentes solutions sont triées en fonction du score attribué par le programme, et seul les deux meilleures solutions sont gardées. Ces deux solutions sont ensuite utilisées chacune comme structures de départ pour une simulation de dynamique moléculaire avec le programme NAMD [Phillips 2005]. Cette simulation permet de générer un modèle qui est comparé aux données expérimentales.

de la chaperonne seule. Les modèles de Fha30 dénaturé sont les 20 modèles générés par une analyse d'ensembles de Fha30 dénaturé. Je suis donc parti de l'hypothèse que Par27 reconnaissait une forme dénaturée de Fha30, ce qui est en accord avec les données biochimiques [Hodak 2008].

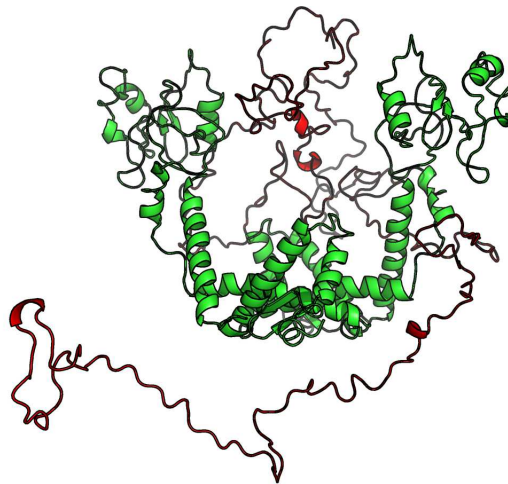


FIGURE 2.12 – Solution de arrimage moléculaire ayant le meilleur score, Par27 est représenté en vert, Fha30 en rouge

Pour chaque paire de modèles, le programme BIGGER [Krippahl 2003] réalise un arrimage moléculaire de corps rigides et classe les solutions selon un score calculé sur la base de paramètres de complémentarité de formes, mais aussi sur des critères d'interactions entre résidus, de satisfaction de liaisons électrostatiques et de surface exposée au solvant [Palma 2000]. La solution présentant le meilleur score est montrée figure 2.12. Les modèles ayant les meilleurs scores sont exclusivement des modèles où Fha30 est situé entre les deux domaines PPIase de Par27 (figure 2.12), ce qui peut être expliqué par le fait que le score comporte un terme qui favorise une zone de contact large. On peut noter que le programme a tendance à situer la zone d'interaction sur la partie centrale de Fha30, laissant les parties N et C-terminales libres. Cela est en accord avec le fait que la partie N-terminale doit être disponible pour pouvoir interagir avec le transporteur qui l'excrète dans le milieu extra-cellulaire [Clantin 2004]. Ce résultat est d'autant plus surprenant qu'aucune contrainte n'est imposée au programme quant au placement des molécules.

A la fin de cette procédure, j'ai sélectionné les 2 meilleures solutions de l'arrimage moléculaire pour chaque conformère de Par27 (faisant 12 modèles au total), et généré des courbes théoriques de ces modèles que j'ai comparé aux données expérimentales. La meilleure solution nous donnait un χ^2 de 4.4. J'ai donc effectué une simulation de dynamique moléculaire pour permettre à la structure de s'équilibrer et ainsi d'améliorer les valeurs de χ^2 .

2.3.4.3 Comparaison des courbes expérimentales et théoriques

La simulation de dynamique moléculaire a permis d'équilibrer la structure, et elle a permis d'obtenir un modèle qui expliquait mieux les résultats expérimentaux. Les valeurs de χ^2 ont ainsi pu être divisées par deux en moyenne, passant de 4.3 à 2.2 pour le modèle ayant les meilleurs statistiques, et qui est présenté figure 2.14. Les autres modèles ont donné des valeurs de χ^2 nettement moins bonnes toutes supérieures à 4. Les courbes expérimentales et théoriques du modèle expliquant le mieux les données, ainsi que celle d'un autre modèle (avec une valeur de χ^2 de 4.7) sont présentées figure 2.13.

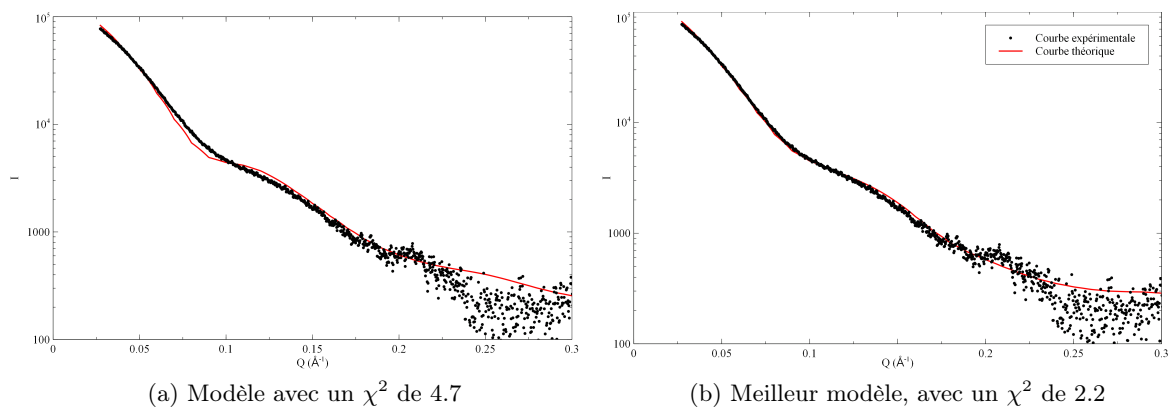


FIGURE 2.13 – Superposition des courbes de SAXS théoriques et expérimentales pour les deux modèles ayant les meilleurs χ^2 .

Sur le modèle expliquant mieux les données de diffusion, la partie N-terminale de Fha30 vient se loger à l'intérieur de la pince de Par27. Un autre détail intéressant est que l'on trouve une proline qui est proche du domaine PPIase de Par27. Aussi, on note que la partie N-terminale, qui contient le domaine de reconnaissance par le transporteur membranaire ne fait pas d'interactions avec la chaperonne, et est donc directement disponible pour interagir avec le domaine POTRA du transporteur membranaire.

Pour avoir une description plus complète de l'état de Fha30 au sein du complexe chaperonne-protéine client, il faudrait utiliser une technique d'analyses d'ensembles pour prendre en compte le fait que Fha30 possède de larges zones flexibles. Aussi, des informations biophysiques complémentaires nous permettraient d'avoir d'autres preuves nous permettant de valider un modèle.

2.4 Conclusion

Les résultats obtenus pour Par27 nous montrent que cette chaperonne possède une flexibilité intrinsèque qui, comme d'autres chaperonnes, lui permet de reconnaître un grand nombre de conformations d'une protéine client totalement dénaturée.

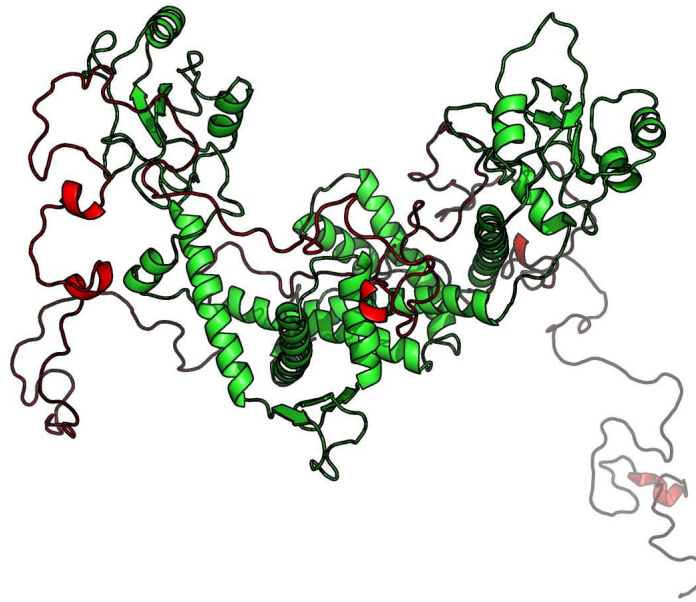


FIGURE 2.14 – Modèle final ayant le meilleur χ^2 .

L'analyse des courbes SAXS de Fha30 dénaturé nous montre que lorsque cette protéine est dénaturée par un agent chimique, elle va adopter des conformations plus étendues qu'une chaîne intrinsèquement désordonnée. Ceci confirme un mode d'action où les dénaturants chimiques viennent interagir avec la chaîne principale et la "déroulent", favorisant ainsi des formes ayant une très grande étendue spatiale.

Enfin, en combinant les différents résultats obtenus, nous avons pu proposer un premier modèle pour le complexe de Fha30/Par27, où la protéine client vient se lier sur la plateforme centrale de Par27 entre les deux domaines PPIase, laissant la partie N-terminale de Fha30 disponible pour être reconnue par le transporteur. La combinaison des techniques de modélisation et de diffusion aux petits angles s'avère donc indispensable pour étudier l'aspect structural de ce système, qui est trop flexible pour être étudié par cristallographie et dont la taille trop importante (77 kDa) rend impossible l'enregistrement des spectres de RMN interprétables. Deux choses cependant restent à explorer afin de caractériser complètement ce système : une analyse d'ensembles du complexe afin de prendre en compte le caractère flexible dans la description, et une meilleure caractérisation biophysique, permettant d'apporter plus de preuves afin de valider le modèle.

Etude du repliement de Fha30

"Stand, and unfold yourself!"

W. Shakespeare, Hamlet

Sommaire

3.1	Introduction	55
3.2	Matériels et méthodes	58
3.2.1	Conception d'une séquence stabilisant la partie C-terminale de Fha30	58
3.2.2	Biologie moléculaire	59
3.2.3	Expression et Purification des protéines	59
3.2.4	Mesures de diffusion élastique incohérente	60
3.3	Résultats	60
3.3.1	Conception d'une coiffe sur la partie C-terminale de Fha30	60
3.3.2	Etude du dépliement de FHA par microscopie à force atomique et simulations de dynamique moléculaire	60
3.3.3	Etude des propriétés dynamiques de Fha30 par diffusion de neutrons incohérente	88
3.4	Conclusion	91

3.1 Introduction

Les protéines ayant une structure en ce que l'on appelle 'hélice β ' forment une classe particulière ayant des caractéristiques communes. Les premières structures haute résolution de protéines en hélice β ont été celles d'une pectate lyase C [Yoder 1993] et d'une protéase alcaline [Baumann 1993]. Leur structure est composée de brins β tournant en hélice. Ces brins peuvent être au nombre de 2, 3 ou 4 par tour d'hélice (figure 3.3).

La structure est maintenue par des liaisons hydrogène entre les brins β parallèles, mais aussi par un confinement des chaînes latérales des résidus hydrophobes dans l'intérieur de l'hélice (figure 3.2).

Ces structures peuvent être classifiées en type droite ou gauche selon le sens dans lequel l'hélice tourne, la grande majorité des hélices β (70%) formant une hélice droite qui est en général plus allongée [Kobe 2000]. Une caractéristique intéressante est que

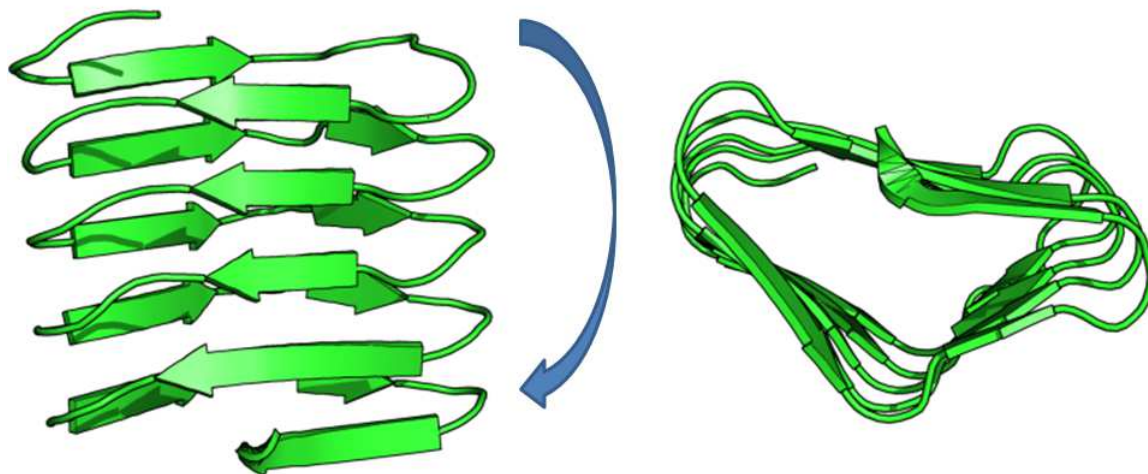


FIGURE 3.1 – Exemple d'une hélice β droite avec trois brins par tour d'hélice (cas de FHA).

les hélices droites sont monomériques, alors que les hélices gauches forment généralement des oligomères, ce qui suggère que les hélices gauches sont moins stables que les hélices droites, cette lacune étant "compensée" par la formation de multimères [Kajava 2006].

D'un point de vue géométrique, ces hélices peuvent former divers motifs qui conditionnent la forme de leur section. On en distingue 4 types : la forme en L (type L), la forme en O (type O), la forme de rectangle (type R) la forme de triangle (type T) [Kajava 2006], les formes les plus courantes étant les types L et T. Les parties terminales de ces hélices comportent un fort potentiel d'agrégation car elles sont constituées de surfaces hydrophobes qui auraient tendance à s'associer pour ne pas être exposées au solvant. C'est pour cela que ces hélices possèdent généralement une coiffe.

Ces coiffes peuvent être constituées par un dernier tour d'hélice qui comporte des résidus chargés, afin de favoriser l'interaction avec le solvant, comme c'est le cas pour la pertactine P69 [Emsley 1996]. Ce type de coiffe est ce que l'on peut trouver de plus simple, mais d'autres structures plus complexes existent. On peut trouver par

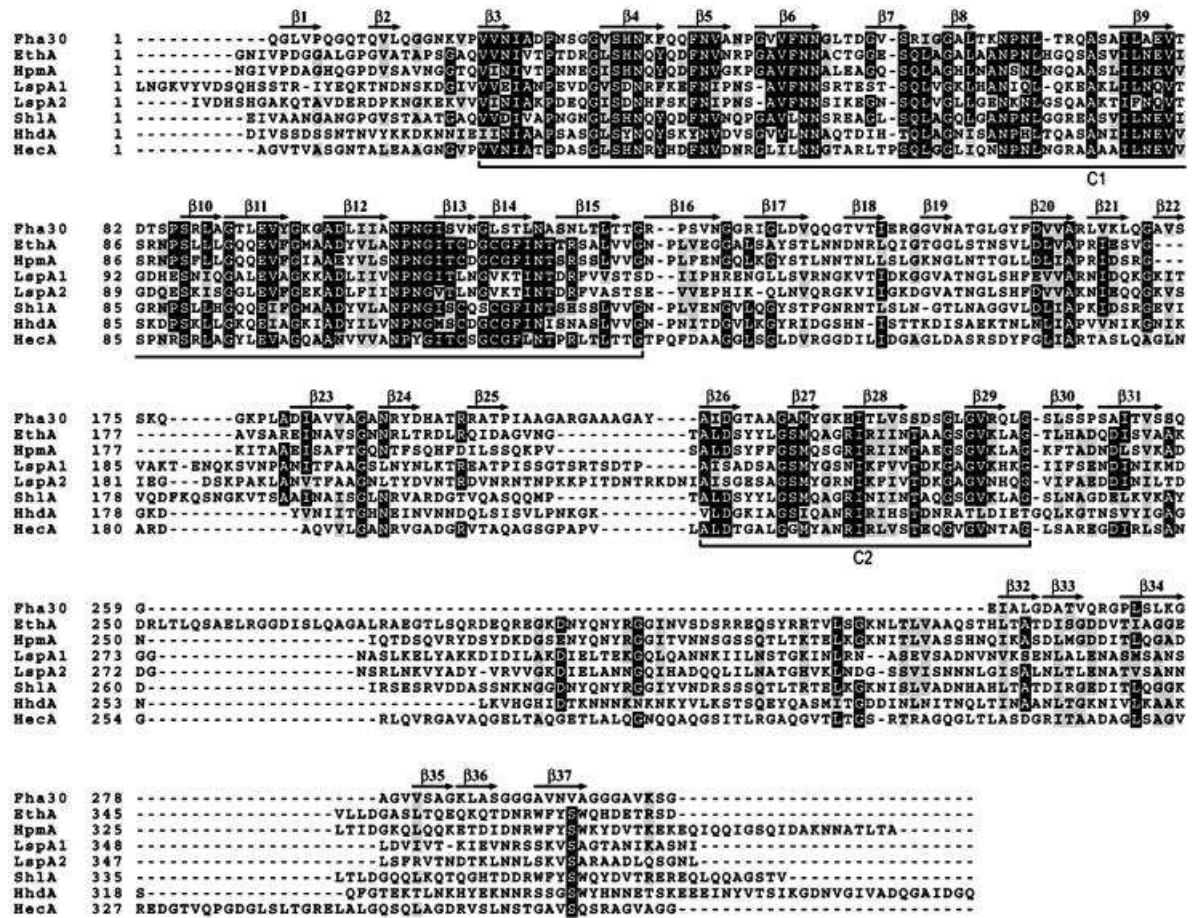


FIGURE 3.2 – Alignement de séquence des parties N-terminales de plusieurs protéines TpsA. Les éléments de structure secondaire de FHA sont représentés au dessus. D'après [Clantin 2004].

exemple des brins β qui viennent recouvrir le dernier tour (comme par exemple la partie N-terminale de FHA [Clantin 2004]) ou encore une hélice α [Pickersgill 1998]. Le mécanisme de repliement et les propriétés mécaniques de ces protéines sont encore mal connus.

La grande majorité de ces protéines sont des facteurs de virulence ou des adhésines qui sont exprimées à la surface d'agents pathogènes et qui doivent reconnaître divers substrats à la surface des cellules de l'organisme hôte. Leur forme allongée serait donc un atout leur fournissant une large surface d'interaction.

Fha30 a une structure en hélice β droite de type T. La partie N-terminale forme une coiffe composée de brins β qui viennent recouvrir la partie C-terminale. Elle forme des dimères en solution, et ne se replie pas correctement une fois dépliée. La protéine ne peut être produite de manière native que si elle est effectivement excrétée par le transporteur.

Ceci est problématique d'un point de vue pratique, car la réaction de repliement n'est plus un phénomène réversible pour lequel on peut atteindre un équilibre thermodynamique. La première approche que nous avons essayé était donc de modifier la surface de dimerisation afin d'obtenir une surface moins hydrophobe et ainsi stabiliser la protéine. Les protéines que l'on trouve à l'état naturel non tronquées ont en effet une coiffe sur leur parties N et C terminales, afin de masquer les parties hydrophobes, nous avons donc décidé de concevoir une coiffe afin de stabiliser la protéine.

La deuxième partie porte sur l'utilisation de techniques de simulation combinées avec des données de microscopie à force atomique afin de comprendre comment s'effectue le dépliement mécanique de la protéine pour obtenir des informations sur les propriétés mécaniques, qui dans le cas de cette molécule présente un intérêt majeur puisque c'est elle qui permet à la bactérie de rester accrochée sur les tissus infectés (rôle d'adhésine).

Enfin, afin de comprendre les propriétés dynamiques d'une structure en hélice- β et le comparer à d'autres type de structures, nous avons entrepris une étude de diffusion incohérente de neutrons.

3.2 Matériels et méthodes

3.2.1 Conception d'une séquence stabilisant la partie C-terminale de Fha30

La construction de la coiffe a été réalisée in-silico avec le programme Rosetta Design. La philosophie de ce programme est de se poser les problème suivant : peut-on à partir d'une structure donnée, retrouver une séquence en acides aminés qui nous donnera ce type de structure ? Ce programme a déjà permis de concevoir une séquence en acides aminés qui se replie dans une topologie préalablement sélectionnée [Kuhlman 2003], ou d'identifier des mutations qui rendent une protéine monomérique dimérique [Kuhlman 2001].

Le choix de la structure qui viendrait coiffer la fin de l'hélice s'est porté sur une hélice α , car c'est un motif relativement simple (ne nécessitant que peu de résidus) que l'on retrouve dans d'autres protéines du même type [Pickersgill 1998]

J'ai du tout d'abord prendre une hélice α de cinq tours (d'une longueur suffisante pour masquer l'extrémité de l'hélice) en polyalanine, et la placer sur la surface de dimerisation de Fha30. Cette étape a été faite avec le logiciel Chimera (figure 3.4). J'ai ensuite utilisé le programme Rosetta design [Liu 2006] pour choisir les résidus qui conviendraient le mieux pour ce genre de structure (séquence optimisée). Le programme a été utilisé pour faire de mutants ou l'on a juste une séquence optimisée de l'hélice- α , l'hélice- α et le dernier tour de l'hélice- β , mais aussi un modèle sans coiffe mais avec une séquence optimisée du dernier tour de l'hélice- β . Ces différentes étapes ont permis la génération de 6 variants : 1 variant avec des résidus optimisés sur le dernier tour de l'hélice β (donc sans hélice α venant masquer la surface hydrophobe), 1 variant avec une coiffe composée d'une hélice α sur la partie

hydrophobe terminale, et enfin 4 variants qui avaient une optimisation des résidus du dernier tour de l'hélice β et une coiffe en hélice α . Afin de vérifier la stabilité des modèles, j'ai entrepris une simulation de dynamique moléculaire afin de vérifier si les constructions resteraient stables. J'ai pour cela utilisé le programme NAMD [Phillips 2005] avec les paramètres décrits dans le chapitre 2.

3.2.2 Biologie moléculaire

Le gène naturel de FHA possède un site de clivage NotI, et il a été tronqué grâce à un site de clivage BamHI introduit pour avoir la construction Fha30 [Clantin 2004]. Un fragment synthétique (Geneart) codant pour la partie C-terminale de Fha30 a été utilisé pour effectuer les mutations.

J'ai d'abord échangé les fragments synthétisés dans un plasmide navette, pEC139, entre les sites de restriction NotI-BamHI. Ce plasmide contient le gène codant pour Fha30 plus une séquence d'adressage qui est clivée lors de la translocation dans le périplasma. J'ai ensuite cloné ce fragment dans le plasmide pBB-I-tac qui est le plasmide final d'expression entre les sites EcoRI-HindIII. Ce plasmide permet d'exprimer Fha30, mais pour que la protéine soit effectivement sécrétée, la machinerie de sécrétion est nécessaire, i.e. le transporteur FhaC. La souche finale est obtenue par transformant une bactérie *E. coli* UT5600 (souche déficiente en protéases périplasmiques) avec le plasmide pT7Fw, contenant le gène codant pour FhaC et le plasmide pBB-I-tac avec la mutation souhaitée. Les constructions étaient donc des souches qui exprimaient la protéine dans le milieu extracellulaire.

3.2.3 Expression et Purification des protéines

La purification de Fha30 et de ses différents mutants sous forme native se fait à partir du surnageant. Des bactéries *E. coli* exprimant Fha30 dans le surnageant sont d'abord cultivées à 37°C dans un milieu LB + Ampicilline jusqu'à D.O \approx 0.6. Ensuite, la production est induite en rajoutant de l'IPTG à concentration finale de 1mM. La culture est ensuite laissée 4h à 37°C. Les cultures sont ensuite centrifugées, le culot bactérien est jeté et le surnageant est mis à pH 5.5, afin que la protéine puisse interagir avec la résine échangeuse de cations utilisée plus tard. Enfin, le surnageant avec le pH ajusté est filtré sur un filtre 0.3 μ m.

La protéine étant quasiment pure dans le surnageant, celui-ci est chargé sur une colonne Poros-HS20 échangeuse de cations préalablement équilibrée dans un tampon MES pH 5.5. Après avoir lavé la colonne avec un tampon MES pH 5.5, la protéine est éluée avec un gradient linéaire allant de 0 à 1M NaCl. Les fractions contenant la protéine sont vérifiées par SDS-PAGE et concentrées avec un concentrateur avec un cut-off de 10kDa avant d'être dialysées contre d'abord contre un tampon Tris 20mM pH 7.5, 150mM NaCl. Pour les expériences de diffusion incohérente, les échantillons ont ensuite été dialysés contre un tampon Tris 0.2mM pH 7.5, 1.5mM NaCl. L'échantillon de 30 mL a ensuite été lyophilisé et resuspendu dans un volume de 200 μ L d'H₂O (2g d'H₂O pour 1g de protéine), puis chargé dans un porte-échantillon

plat en aluminium de 0.3 mm. Afin de faire une soustraction de tampon correcte, j'ai aussi préparé une cellule contenant 200 μ L de 20mM tris pH7.5, 150mM NaCl dans le même type de cellule.

3.2.4 Mesures de diffusion élastique incohérente

Les mesures sur IN13 ont été faites dans la gamme de température allant de 280 à 310K avec un pas de 5K.

La réduction des données a été réalisée en utilisant la procédure décrite dans [Bée 1988]. La mesure du tampon est soustraite à la mesure de la cellule avec l'échantillon, pondérée par les facteurs d'absorption. Ces valeurs sont ensuite normalisées à une mesure d'une plaque de vanadium (diffuseur purement incohérent) à laquelle on a soustrait la cellule vide. Les mesures sur IN16 ont été faites dans les mêmes conditions que sur IN13, à l'exception que le solvant utilisé était du D₂O et la gamme de température allait de 280-315K avec un pas de température de 5K.

3.3 Résultats

3.3.1 Conception d'une coiffe sur la partie C-terminale de Fha30

Les protéines en hélice- β de structure connue possèdent une coiffe qui leur permet de masquer leur parties hydrophobes présentes à l'intérieur de l'hélice. Une structure que l'on retrouve couramment est le recouvrement par une hélice α , comme observé chez *Erwinia carotovora ssp. carotovora* [Pickersgill 1998](figure 3.3).

J'ai donc décidé d'ajouter une séquence permettant de former une hélice dans la partie C-terminale, qui est représentée figure 3.4. J'ai aussi conçu un mutant sans coiffe, mais avec des résidus sur le dernier tour de l'hélice- β optimisés. Pour les mutants ayant une hélice α en guise de coiffe, le programme Rosetta Design [Liu 2006] donnant plusieurs solutions (10 au total), j'ai sélectionné les modèles où les résidus exposés au solvant de la dernière hélice étaient des résidus polaires, ce qui a permis de réduire leur nombre à 5. Sur les 6 mutants, 2 seulement ont pu être exprimés, les autres ne donnant pas d'expression. La détermination du degré d'oligomérisation par SEC-MALLS-RI nous indique que ces mutants sont dimériques (figure 3.5), avec une masse autour des 60kDa. Ils ne pourront donc pas être utilisés pour des études de dépliement/repliement nous permettant de mesurer un ΔG de dépliement.

3.3.2 Etude du dépliement de FHA par microscopie à force atomique et simulations de dynamique moléculaire

Les protéines impliquées dans les phénomènes d'adhésion doivent, de par leur fonction, avoir des propriétés mécaniques remarquables afin de résister aux stress tout en restant fonctionnelles, mais aussi avoir un chemin de repliement compatible avec l'excrétion dans le milieu extracellulaire.

L'hémagglutinine filamenteuse est une protéine de 230 kDa [Locht 1993] impliquée

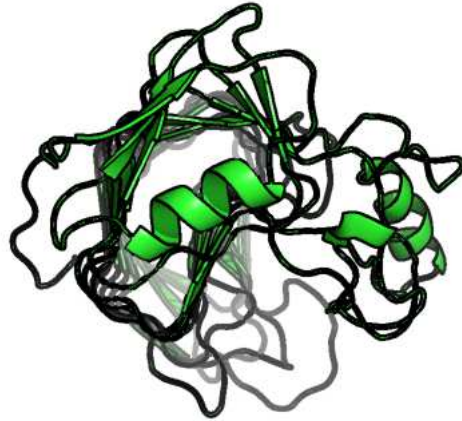


FIGURE 3.3 – Coiffe de la polygalacturonase de *Erwinia carotovora ssp. carotovora* (code PDB 1BHE)

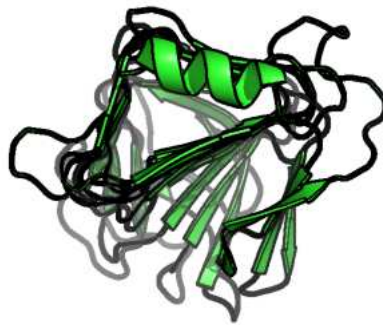


FIGURE 3.4 – Conception in-silico de la coiffe de Fha30

dans l'adhésion de *Bordetella pertussis*, une bactérie à Gram négatif responsable

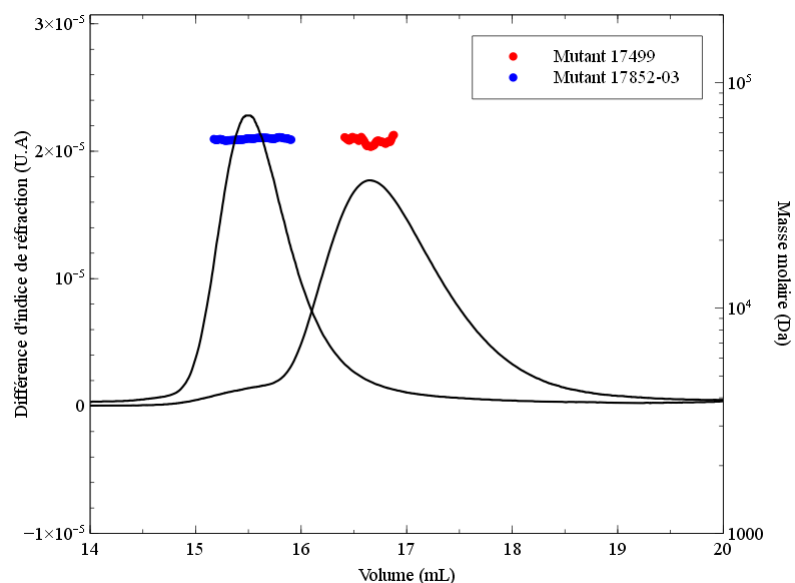


FIGURE 3.5 – Profil de chromatographie (noir) et masse (rouge) déterminées par SEC-MALLS-RI

de la coqueluche. Elle possède un repliement en hélice β [Clantin 2004], formant de longs cylindres observables en microscopie électronique [Kajava 2001]. Sa sécrétion se fait par le système de transport à deux partenaires (TPS). FhaB (la protéine excrétée) est donc transportée au travers de la membrane externe grâce à son transporteur associé, FhaC. Ce transporteur fonctionne sans apport d'ATP, et a la particularité de transporter FhaB que sous sa forme dépliée [Clantin 2007] : La protéine va donc se replier à la surface de la bactérie, sans l'assistance d'aucune protéine chaperonne.

Nous avons étudié le dépliement de FhaB par microscopie à force atomique (AFM) et dynamique moléculaire afin de sonder ses propriétés mécaniques, mais aussi pour comprendre le mécanisme de repliement, qui est intimement relié à son transport au travers la membrane. Les profils de force-extension obtenus par AFM pour différentes constructions (Fha30, Fha60 et FhaB complet) nous montrent que la protéine se déplie de manière séquentielle, ce qui est surprenant pour une protéine monodomaine [Rief 1997]. Afin de comprendre les différents événements observés lors du dépliement mécanique, l'expérience d'AFM a été simulée en effectuant un calcul de dynamique moléculaire avec un champ de force simplifié [Clementi 2000b]. Ce calcul nous montre que le dépliement se fait d'abord sur les parties N et C terminales, et que la partie la plus stable se trouve entre les parties LC1 et C2 [Clantin 2004], qui se trouve au milieu de la protéine, dans une zone où un feuillet β conservé [Hodak 2006] vient se replier sur l'hélice β . Ce mécanisme est compatible avec un repliement progressif couplé au transport au travers de la membrane externe. Il nous permet de suggérer un modèle de repliement où FhaB va diffuser librement de part

et d'autre du canal. Lorsque un nombre des résidus suffisant se retrouve a la surface de la cellule, ceux-ci vont se replier, et vont donc se retrouver 'bloqués' à la surface de la cellule. La protéine va ensuite répéter ce processus jusqu'au repliement complet.

Manuscrit soumis à *Proceedings of the National Academy of Science of the United States of America*

Sequential unfolding of a beta helical protein

by single-molecule atomic force microscopy

David Alsteens^{1#}, Nicolas Martinez^{2,3#},

Yves F. Dufrêne^{1*}, Marc Jamin^{2*}, Françoise Jacob-Dubuisson^{4*}

¹ Université catholique de Louvain, Institute of Condensed Matter and Nanosciences, Croix du Sud 2/18, 1348 Louvain-la-Neuve, Belgium

² UMI3265 UJF-EMBL-CNRS Unit for Virus Host Cell Interactions, 6 rue Jules Horowitz 38042 Grenoble, France

³ Institut Laue Langevin, 6 rue Jules Horowitz 38042 Grenoble, France

⁴ Center for Infection and Immunity of Lille, 1 rue Calmette, F-59019 Lille, France; Univ. Lille-Nord de France, F-59000 Lille, France; CNRS UMR8204, F-59021 Lille, France; INSERM U1019, F-59019 Lille, France.

contributed equally

* for correspondence: yves.dufrene@uclouvain.be; tel +32 10 47 36 00; jamin@embl.fr; tel +33 476 20 94 62; fax + 33 476 20 94 00; francoise.jacob@ibl.fr; tel +33 320 87 11 55; fax +33 320 87 11 58.

key words: β helix, protein secretion, protein folding, bacterial pathogen, single-molecule atomic force microscopy, steered molecular dynamics

Abstract

The parallel β helix is a common fold among extracellular proteins, particularly microbial adhesins and enzymes. In Gram-negative bacteria, a widespread pathway called Two-Partner Secretion (TPS) is dedicated to the secretion of long β -helical proteins. Those proteins are too long for intracellular folding and they have been hypothesized to cross the outer membrane linearly and fold only after secretion. However, no conventional energy source is available to assist secretion and folding at the outer membrane. Here, single-molecule atomic force microscopy (AFM) and steered molecular dynamics (SMD) simulations were combined to explore the mechanical properties of a prototypic β -helical TPS substrate, filamentous hemagglutinin of the whooping cough agent *Bordetella pertussis*. Both AFM and SMD demonstrated that strong extension forces are required for full protein unfolding. Strikingly, a conserved N-terminal region critical for secretion was shown to unfold hierarchically, a remarkable property for a mono-domain protein, while the full-length protein unraveled one to a few coils at a time. The sequential uncoiling of the β helix supports an *in vivo* vectorial folding process initiated from one extremity of the protein and which progressively drives translocation. The mechanical properties uncovered here may apply to other proteins that emerge from a channel in a linear conformation and have to fold in the absence of energy and chaperone assistance.

Introduction

Extracellular proteins are overwhelmingly involved in interactions with the environment, and therefore they have evolved to withstand various forms of stress in order to maintain function in non-controllable conditions. Importantly, exported proteins must also adopt folds that are compatible with their transport across membranes. Those constraints

probably account for the large proportion of repetitive extracellular proteins. β helices are such repetitive proteins, composed of coils formed of three short, amphipathic β strands connected by variable-length turn regions (1-3). The strands of successive coils stack on top of one another, forming parallel β sheets with conserved hydrogen-bonding patterns (2). The interior of β helices is mostly hydrophobic and shielded from the milieu by capping structures at its extremities (4). β helices display extensive surface area well suited for binding, and their scaffold is also used for the presentation of specific functional determinants by the insertion of loops or domains. The β helix and related solenoid folds have been adopted throughout the phylogenetic tree, from microbial and phage proteins to insulin-like growth factor receptors in higher eukaryotes and antifreeze proteins in insects (1,2).

The first β helix structure was that of pectate lyase C (PelC) (3). This relatively small enzyme contains seven complete β coils and is thought to fold in the periplasm before export to the extracellular milieu (5). In contrast, other families of secreted bacterial proteins that are overwhelmingly β -helical, TpsA proteins (for Two-partner secretion, component A; see below) and autotransporters, appear to be too long for intracellular folding (6-8). They have thus been proposed to cross the membrane in an extended conformation and to fold at the cell surface after they exit from their proteinaceous transport channels. The mechanisms of secretion of TpsA proteins and autotransporters across the outer membrane are the focus of much interest, as those long proteins translocate and fold in the absence of cellular energy such as ATP or the proton motive force.

TpsA proteins can form extremely long β helices of several thousand residues. A model TpsA protein is filamentous hemagglutinin (FHA), the major adhesin of the whooping cough agent *Bordetella pertussis* (9). FHA is a 230-kDa protein that appears by electron microscopy as a rigid, 40 nm-long rod (10). It is mostly composed of approximately 80

imperfect sequence repeats of 19 residues or more in tandem that form the coils of an extended β helix, followed by a 50-kDa globular domain at the C terminus (11) (Fig. 1A). FHA is secreted by the widespread bacterial Two-Partner secretion (TPS) pathway, which appears to be dedicated to β -helical proteins (8). It is translocated across the outer membrane by its specific TpsB transporter, called FhaC (12). FHA harbours a 250-residue-long N terminal ‘TPS domain’, which is essential for secretion and conserved among proteins secreted by TPS systems (13,14). The TPS domain itself forms a β helix capped by anti-parallel β structure at the top and flanked by a small anti-parallel β sheet on one side (13,15,16). The recombinant protein Fha30 (13), which comprises the TPS domain and the first three coils of the repeated region, is the smallest FHA fragment competent for secretion and is therefore used as a model (Fig. 1B).

Because FHA is hypothesized to be translocated in a linear conformation through a channel and to fold at the cell surface, its mechanical properties might play an important role for its secretion mechanism. Single-molecule atomic force microscopy (AFM) (17-19) is a technique of choice to investigate the mechanical properties of this long, repetitive protein by exerting a force along the helix axis. In this work, we used a combination of single-molecule AFM and steered molecular dynamics (SMD) simulations to characterize the molecular elasticity of FHA. This represents the first such study of a β -helical protein. We show that FHA unfolds sequentially when subjected to force. These remarkable properties have implications for the secretion and folding pathways of FHA and related proteins.

Results

Mechanical unfolding of TPS domain by single-molecule atomic force microscopy. A modified Fha30 was bound to a gold surface through a C-terminal Cys residue and to a Ni-

NTA functionalized AFM tip through an N-terminal 6-His tag. Single-molecule AFM unfolding was performed by upward extension from the surface (Fig. 2A). A substantial fraction of the force-distance (F-D) curves (~5 % from a total of more than 5000 curves) exhibited a saw-tooth pattern with successive force peaks, in which the mechanical tension increased and then sharply decreased, suggesting that the polypeptide chain experienced a multi-step unfolding process as a result of force application (Fig. 2B). In those curves, the unfolding process is characterized by an initial series of small force peaks followed by several major peaks of increasing forces. Superposition of several F-D curves showed that the last three, strongest force peaks were systematically present, supporting a sequential process along which progressive forces hierarchically unfolded increasingly stable regions of the protein (Fig. 2C). This behavior is quite different from that observed for individual domains in most multi-domain proteins that generally unfold in a highly concerted manner, in one or two force peaks (20,21). The strongest peak occurred at applied forces of ≈ 250 pN, consistent with the unfolding of domains with β structures, and much higher than those for α -helical proteins (20,21). The worm-like chain (WLC) model (22) that characterizes the elastic response of a polymer to extension could be used to fit the force curve preceding each unfolding event (Fig. 2B). The rupture length of 114 nm was compatible with full extension of the protein assuming that each amino acid contributes 0.36 nm to the contour length of a fully extended polypeptide chain.

To probe the order in which the different regions of Fha30 unfold upon extension, we produced variant proteins with a stretch of 20 alternating Gly and Ser residues inserted either into the loop following the first extra-helical β hairpin (Fha30_GS1) or into the long loop close to the end of the TPS domain (Fha30_GS2) (Fig. 1B). Fha30_GS1 was hardly secreted and heavily degraded, most likely because the inserted linker interfered with the folding of the protein. Fha30_GS2 was secreted efficiently and was therefore purified and analyzed by

AFM. The F-D curves were similar to those obtained for the wild type (wt) protein, except for the presence of an additional force peak in the second half of the extension process (Fig. 3A,B). The contour length of this peak appeared shifted by 6.6 nm from the preceding force peak, consistent with the length increment expected for a 20-residues segment in the WLC model. This result shows that the region of Fha30 encompassing the loop unfolds rather late upon forced elongation, and the inserted linker has no major effect on the early stages of the unfolding reaction. This supports a mechanism of progressive unraveling of the structural elements. The facts that the inserted peptide unfolded as a separate element *in vitro* and did not affect secretion *in vivo* argue that neither Fha30 structure nor its *in vivo* folding pathway was significantly altered and that folding and secretion are intimately linked.

Simulated unfolding. To provide an interpretation for these experiments, we used SMD simulations with a structured-based coarse-grained model of Fha30. We reproduced the extension process by keeping the C terminus fixed and moving the N terminus at different constant pulling speeds (23). For the wt protein, the simulated force-distance curves reproduced the main events of the experimental AFM profiles, exhibiting several small force peaks in the early part of the extension process, followed by several major force peaks of increasing forces (Fig. 4A). The F-D profiles obtained at different pulling speeds were similar, indicating a good convergence of the simulations (Fig. S1A).

The right-handed β helix of Fha30 is capped by three β -strands at the N terminus and comprises six irregular coils, called here coils_A to _F, followed by three regular R1 coils, named R1A to R1C (13). In addition, the structure includes three extra-helical elements – a β hairpin between coils_A and _B (extra-helix element I), another within coil_D (extra-helical element II) and a β -hairpin followed by a loop in coil_F (extra-helical element III). The β hairpins of the extra-helical elements II and III form a β sheet that packs against the β helix. Based on sequence alignments (13), the coils of the β helix and the extra-helical elements can

be assigned to successive less conserved (LC) and conserved (C) regions of the TPS domain, called LC1, C1, LC2 and C2, respectively (Fig. 1B).

To analyze the unfolding trajectories, a native contact map of Fha30 was built, and its evolution was monitored over time. In the early stages of the extension process, the small force peaks represented the sequential unraveling of the R1C and R1B coils (in black and red in Fig. 4 A-C), and of the LC1 region, which disappeared simultaneously with R1C (in black) and a strand of R1A (in orange). This was followed in the next three peaks by the unfolding of the last strands of R1A together with the C-terminal strands of coil_F (in yellow), then unfolding of the last strand of coil_F and of the extra-helical element III (in green), and finally that of the last strand of coil_E (in light blue). In the next step, the N-terminal part of coil_A and the C-terminal part of coil_E, from opposite sides of the molecule, unraveled simultaneously (in violet). This was followed by the unfolding in two successive peaks of the second part of coil_A and of extra-helical element I (in magenta), and then of coil_B (in dark blue). The portion of Fha30 to unfold last (in gray) comprised coil_C, coil_D and the extra-helical element II. In conclusion, unfolding of Fha30 proceeded alternatively from both ends and progressed towards the center of the molecule, with a stable 'nucleus' straddling parts of C1 and LC2.

A model of the Fha30_GS2 variant was built by homology, with the insertion as a long β hairpin. The SMD simulation showed similar force-extension profiles with small force peaks in the first half of the extension followed by several larger force peaks in the second half (Fig. S1B). In accordance with the experimental results, the contour lengths calculated for the large force peaks were shifted by about 6 nm relative to their positions in the wt profile, showing that the region that carries the insertion unfolded in the second part of the extension process.

Unfolding of longer variant. We next investigated the unfolding pathway of a longer form called Fha60, with 15 additional R1 coils (Fig. 1A). Fha30 and Fha60 were attached to the AFM surface by their C termini and stretched with a non-functionalized AFM tip (Fig. S2). The F-D curves displayed a number of peaks for both proteins (Fig. S2A,B), and their superposition documented several common unfolding peaks, which were shifted to longer elongation distances for Fha60 (Fig. S2C). This indicated a pathway in which the additional R1 coils unfolded early during the extension of Fha60, whereas the TPS domain, which is common to both Fha30 and Fha60, unfolded late. This pathway is in agreement with our interpretation of the sequential unfolding of Fha30, in which the R1 coils appeared less stable than the TPS domain and unfolded at the beginning of the extension. This also showed that the TPS domain is a mechanically resistant portion of FHA compared with the more regular repeats of the central part of the protein.

Mechanical unfolding of full-length FHA. The unfolding of FHA was also analyzed by single-molecule AFM. FHA, which is composed of more than 80 β -helical rungs (Fig. 1A), was immobilized randomly and pulled with a non-functionalized AFM tip at a loading rate of 10,000 pN/sec (Fig. 5A). A saw-tooth pattern with a number of unfolding peaks was observed in a substantial fraction of the F-D curves ($\sim 10\%$ in a total of > 1000 curves) (Fig. 5B), and the peaks could also be fitted with a WLC model (Fig. 5B). For a subset of curves, the total separation distance between the chip and the AFM tip after rupture was consistent with the expected length of fully extended FHA, showing that the protein eventually becomes completely unfolded. Although no regular patterns of unfolding were detected, the most frequent peak-to-peak distances were multiples of 8-9 nm (Fig. 5C). As the majority of the FHA coils comprise 19 residues and each coil is 2-3 nm wide, 8-9 nm roughly correspond to the rise and unfolding of one β -helical coil ($0.36 \text{ nm} \times 19 + 2\text{-}3 \text{ nm}$). Thus, FHA unfolded one to several coils at a time. Similar results were obtained with Fha44, which contains the TPS

domain followed by 29 R1 repeats (Fig. 1A) (10). The F-D curves of Fha44 showed a number of peaks distributed in a non-random manner, with the most frequent occurrences of separation distances equal to 9.0, 15.6, 22.6 and 31.3 nm (Fig. S3).

Discussion

The mechanisms by which a linear polypeptide folds at the exit of a channel is a major issue in biology, pertinent to proteins exiting from the ribosome or from the Sec machinery and to many proteins exported to the milieu (24,25). Specific chaperones are present in the various intracellular compartments to assist protein folding; however this is not necessarily the case for proteins destined to the extracellular milieu or the cell surface. In bacteria, several protein secretion pathways involve the translocation of secretory polypeptides in extended conformations through outer membrane-embedded protein channels. Remarkably, many such proteins that are kept largely unfolded prior to secretion appear to adopt solenoid folds (7,8). In spite of the large representation of β -helical structures and of their potential importance for secretion and/or function, their mechanical properties remain essentially unknown. Combined single-molecule AFM and SMD simulations to stretch the TPS substrate FHA revealed that this model β -helical protein unfolds in a sequential manner, which has major implications for its secretion and folding *in vivo*.

Most AFM studies thus far have been performed on mono-domain proteins or on proteins naturally or artificially composed of a string of small independent domains (20,21,23). Upon forced extension, such protein domains generally unfold in a concerted manner, giving rise to one or two force peaks. Accordingly, many mono-domain proteins subjected to *in vitro* thermal or chemical denaturation undergo reversible two-state transitions (26,27). β -helical PelC was also reported to undergo highly cooperative unfolding and

refolding transitions, with the reformation of most of its structure in one step (28). In contrast, Fha30 refolded slowly and incompletely *in vitro* after chemical denaturation (14) and similarly, denatured β -helical autotransporter proteins did not refold readily (e.g. 29). Those observations and the progressive unfolding patterns uncovered by this work strongly support an *in vivo* vectorial folding process for those long secreted proteins, initiated at one extremity of the polypeptide and extending towards the other end over the course of secretion. Vectorial folding is facilitated by the short-range character of most interactions in β helices. Of note, long-range disulfide bonds are absent from TpsA proteins.

The high forces necessary to unravel the TPS domain and full-length FHA are strong indications that once a stable nucleus has formed at the cell surface, the free energy released by the addition of tertiary structure to the nascent protein might power translocation of subsequent segments in the absence of another source of energy. We thus propose that after exiting from the FhaC channel, early folding of a TPS nucleus occurs at the cell surface. At this stage, interactions between FhaC and the C2 part of FHA's TPS domain in the periplasm (14) may transiently promote nucleation at the surface, by hindering backsliding of the translocated portion to the periplasm. Formation of a β helical template then favors the formation of additional β -helical coils, until the acquisition of a stably folded portion at the surface promotes translocation and folding of the following segments. Initially, the formation of the folded TPS nucleus at the cell surface may be kinetically favored, because the low contact order topology of β helices should result in rather high folding rate constants (30). The possibility that some of the extracellular loops of the FhaC transporter facilitate initial folding steps of FHA remains to be investigated.

The TPS domain was previously shown to mediate molecular recognition between FHA and its transporter prior to secretion (14). This work has identified another, most likely

crucial role for the TPS domain, which is to initiate FHA folding at the bacterial cell surface and template β helix formation in the secretion process. This essential function most likely explains why some single substitutions were found to abolish secretion without affecting molecular recognition with the transporter (14,31). According to SMD simulations, the most stable portion of the TPS domain comprises a part of C1, whose sequence is highly conserved among TPS domains, as well as the anti-parallel β sheet of LC2 that packs against the β helix and most likely stabilizes it, which is a structurally conserved element in TPS domains as well (13,15,16). Similarly, a semi-conserved C-terminal domain of autotransporters was shown to initiate vectorial secretion and folding, albeit in a C to N direction (32-34). The capacity to build up extremely long, stable structures in a simple manner, by coupling the progressive folding of a polypeptide at the cell surface with its translocation, most likely explains why proteins secreted by distinct pathways and that carry out various functions have converged toward the β -helical fold.

Among the many advantages of repetitive folds is the ability to evolve rapidly by the addition, deletion, duplication or rearrangement of repeated units, with beneficial consequences such as the potentiality to generate new binding capabilities or to escape the immune system (35). In addition to their role for the vectorial folding and secretion of long proteins, the mechanical properties of the solenoid fold may confer elasticity, which is likely to be important for adhesin function. The mechanical properties uncovered here will likely apply to many other prokaryotic and eukaryotic proteins that adopt β -helical or related folds and might play key roles in their secretion, folding and functions.

Materials and Methods

Plasmids. To produce Fha30 with a C-terminal 6-His tag in *E. coli*, the EcoRI-BamHI fragment coding for Fha30 (13) was introduced into pUC19, and the internal HindIII site in the signal peptide coding sequence was removed by site-directed mutagenesis, yielding pEC138 Δ . pEC138 Δ was modified by the introduction of a 6-His coding sequence and a nonsense codon at the end of *fha30*, yielding pEC138NM1-His. The EcoRI-HindIII fragment of the latter plasmid was introduced into pBBItac, yielding pTac138NM1-H. To construct pBBItac, *lacI* and the *tac* promoter of pMMB91 (36) were amplified by PCR as a NcoI-XbaI fragment, and the amplicon was introduced into pBBR1-MCS5 (37). To produce Fha60, pEC138 Δ was modified by introducing successively at the end of *fha30* a linker coding for 10 Gly-Ser residues, then an amplicon of the *fhaB* sequence that follows *fha30* (861 additional nucleotides) and finally a 6-His tag coding sequence followed by a nonsense codon. The resulting fragment EcoRI-HindIII fragment was inserted into pBBItac, yielding pTac138-NM2-lk-His. For production of Fha30 with an N-terminal 6-His tag and a C-terminal Cys residue, pEC138 Δ was modified by the successive introductions of a 6-His coding sequence after the first three codons of mature FHA by overlapping PCR and of a linker sequence coding for a short segment with a helical propensity to cap the C-terminus of the truncated β helix and harbouring 2 Cys codons before the nonsense codon into the BamHI site, yielding pEC138H-LX2C. The EcoRI-HindIII fragment of the latter plasmid was introduced into pBBItac, yielding pTac138H-LX2C. For Fha30_GS1 and Fha30_GS2 (20-residues linker insertions between residues 73 and 74 and between residues 210 and 211 of mature FHA, respectively), synthetic gene portions were obtained from GeneArt (Würzburg, Germany) and introduced by ligation to replace the respective wild type fragments in pEC138H-LX2C. Finally, the EcoRI-HindIII fragments of the latter two plasmids were introduced into pBBItac as above. For production of Fha44 with N- and C-terminal 6-His tags in *B. pertussis*, *fha44*-encoding pCB4 (38) was modified as follows. A 6-His coding sequence followed by a

nonsense codon was inserted into the BamHI site of pCB4, and the promoter and 5' coding region of *fhaB* were exchanged for a fragment obtained by overlapping PCR to insert a 6-His coding sequence between the first and second codons of mature FHA, yielding pJe3. The 2.8 kb EcoRI-XbaI fragment of pJe3 was introduced into pBBR1-MC5, yielding pB5Je3. *B. pertussis* BPGR4 (*fhaB* KO; ref. 39) was transformed with the latter plasmid for production of the recombinant protein.

Protein production and purification. FHA and Fha44 were produced in *B. pertussis* culture supernatants and purified by affinity chromatography on Heparin Sepharose (40). The shorter forms were produced in *E. coli* supernatants and purified as described (13).

Atomic force microscopy. AFM measurements were performed at room temperature (20°C) in buffered solutions (PBS; pH 7.2), using a Nanoscope V Multimode AFM (Bruker Corporation, Santa Barbara, CA) and oxide-sharpened microfabricated Si₃N₄ cantilevers (Microlevers, Bruker Corporation) or gold-coated cantilevers (Olympus, Atomic Force, Germany). The spring constants of the cantilevers were measured using the thermal noise method (Bruker Corporation), yielding values ranging from 0.011 to 0.031 N/m. All force measurements were recorded with a loading rate of 10,000 pN/s, calculated by multiplying the tip pulling velocity (nm/s) by the spring constant of the cantilever (pN/nm). Gold-coated surfaces were prepared by coating silicon wafers (Siltronix) by electron beam thermal evaporation with a 5-nm thick chromium layer followed by a 30-nm thick gold layer.

To attach FHA fragments using His-tags, gold-coated surfaces and gold cantilevers were cleaned for 15 min by UV and ozone treatment, rinsed with ethanol, dried with a gentle nitrogen flow and immersed overnight in ethanol containing 0.05 mM of nitrilotriacetate-terminated (5%) and triethylene glycol-terminated (95%) alkanethiols (Prochimia, Poland). After rinsing with ethanol, the tips and supports were immersed in a 40 mM aqueous solution of NiSO₄ (pH 7.2) for 1 h, rinsed with water, incubated in PBS buffer containing 200 µg/mL

FHA fragments for 2 h, and finally rinsed with buffer.

To graft the fragments using Cys residues, the gold-coated surfaces and gold cantilevers were cleaned for 15 min by UV and ozone treatment, rinsed with ethanol, dried with a gentle nitrogen flow and incubated in PBS buffer containing 200 $\mu\text{g}/\text{mL}$ FHA fragments for 2 h, and finally rinsed with buffer.

The full protein was kept in PBS buffer at a concentration of 200 $\mu\text{g}/\text{mL}$ and allowed to adsorb onto freshly evaporated gold-coated surface for 1 h.

The force peaks of the F-D curves were fitted with the worm-like-chain model, using a persistence length of 0.4 nm: $F(x) = k_B T / l_p [0.25(1 - x/L_c)^{-2} + x/L_c - 0.25]$, where L_c and l_p are the contour length and persistence length of the molecule, respectively, k_B is the Boltzmann constant, and T is the absolute temperature.

Steered molecular dynamics simulations using a coarse-grained model. SMD simulations were performed using a structure-based model of Fha30 as described in (23). The crystal structure of Fha30 (1RWR) was reduced to a coarse-grained $C\alpha$ model with the CG force field (41) using the SMOG@ctbp webserver (42). The constant velocity SMD simulations were performed with the GROMACS package (43). The C terminus of the protein was fixed, while the N terminus was moved at constant velocity. The force constant of the spring was set to 6 $\text{pN}\cdot\text{nm}^{-1}$ and the velocity was varied from 1 to 20 $\text{nm}\cdot\text{ns}^{-1}$. The protocol tested with the I27 domain of titin reproduced the simulated unfolding process described in the literature (20).

Acknowledgements

We thank Jérémy Guérin and Emmunuelle Petit for help with some of the constructs. A. Kajava is acknowledged for providing calculated coordinates of the repeated regions of FHA. This work was supported in part by the ANR “DYN FHAC” (2010-2014) grant to F. J.-D., and NM was supported by a fellowship from the Region Rhone-Alpes. Work in the Y.F.D. team was supported by the National Foundation for Scientific Research (FNRS), the Université catholique de Louvain (Fonds Spéciaux de Recherche), the Federal Office for Scientific, Technical and Cultural Affairs (Interuniversity Poles of Attraction Programme), and the Research Department of the Communauté française de Belgique (Concerted Research Action). Y.F.D. and D.A. are Senior Research Associate and Postdoctoral Researcher of the FRS-FNRS.

References

1. Kajava AV, Steven AC (2006) β -rolls, β -helices, and other β -solenoid proteins. *Adv Protein Chem* 73:55-96.
2. Jenkins J, Pickersgill R (2001) The architecture of parallel β -helices and related folds. *Prog Biophys Mol Biol* 77:111-175.
3. Yoder MD, Keen NT, Jurnak F (1993) New domain motif: the structure of pectate lyase C, a secreted plant virulence factor. *Science* 260:1503-1507.
4. Bryan AW Jr., Starner-Kreinbrink JL, Hosur R, Clark PL, Berger B (2011) Structure-based prediction reveals capping motifs that inhibit β -helix aggregation. *Proc Natl Acad Sci USA* 108:11099-11104.
5. Lindeberg M, Boyd CM, Keen NT, Collmer A (1998) External loops at the C terminus of *Erwinia chrysanthemi* pectate lyase C are required for species-specific secretion through the out type II pathway. *J Bacteriol* 180:1431-1437.
6. Bradley P, Cowen L, Menke M, King J, Berger B (2001) BETAwrap: Successful prediction of parallel β -helices from primary sequence reveals an association with many microbial pathogens. *Proc Natl Acad Sci USA* 98:14819-14824.
7. Kajava A, Stevens A (2006) The turn of the screw: variations of the abundant beta-solenoid motif in passenger domains of Type V secretory proteins. *J Struct Biol* 155: 306-315.
8. Jacob-Dubuisson F, Fernandez R, Coutte L (2004) Protein secretion through autotransporter and two-partner pathways. *Biochim Biophys Acta* 1694:235-257.
9. Locht C, Bertin P, Menozzi FD, Renauld G (1993) The filamentous haemagglutinin, a multifaceted adhesin produced by virulent *Bordetella* spp. *Mol Microbiol* 9:653-660.

10. Makhov AM, *et al.* (1994) Filamentous hemagglutinin of *Bordetella pertussis*. A bacterial adhesin formed as a 50-nm monomeric rigid rod based on a 19-residue repeat motif rich in β strands and turns. *J Mol Biol* 241:110-124.
11. Kajava AV, *et al.* (2001) β -helix model for the filamentous haemagglutinin adhesin of *Bordetella pertussis* and related bacterial secretory proteins. *Mol Microbiol* 42:279-292.
12. Clantin, B, *et al.* (2007) Structure of the membrane protein FhaC: a member of the Omp85-TpsB transporter superfamily. *Science* 317:957-61.
13. Clantin B, *et al.* (2004) The crystal structure of filamentous hemagglutinin secretion domain and its implications for the two-partner secretion pathway. *Proc Natl Acad Sci USA* 101:6194-6199.
14. Hodak H, *et al.* (2006) Secretion signal of the filamentous haemagglutinin, a model two-partner secretion substrate. *Mol Microbiol* 61:368-382.
15. Yeo HJ, *et al.* (2007) The structure of the *Haemophilus influenzae* HMW1 pro-piece reveals a structural domain essential for bacterial two-partner secretion. *J Biol Chem* 282:31076-31084.
16. Weaver TM, *et al.* (2009) Structural and functional studies of truncated hemolysin a from *Proteus mirabilis*. *J Biol Chem* 284:22297-22309.
17. Rief M, Oesterhelt F, Heymann B, Gaub HE (1997) Single Molecule Force Spectroscopy on Polysaccharides by Atomic Force Microscopy. *Science* 275:1295-1297.
18. Marszalek PE, Oberhauser AF, Pang YP, Fernandez JM (1998) Polysaccharide elasticity governed by chair-boat transitions of the glucopyranose ring. *Nature* 396:661-664.
19. Oesterhelt F, *et al.* (2000) Unfolding pathways of individual bacteriorhodopsins. *Science* 288:143-146
20. Rief M, Gautel M, Oesterhelt F, Fernandez JM, Gaub HE (1997) Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* 276:1109-1112.
21. Rief M, Pascual J, Saraste M, Gaub HE (1999) Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. *J Mol Biol* 286:553-561.
22. Bustamante C, Marko JF, Siggia ED, Smith S (1994) Entropic elasticity of lambda-phage DNA. *Science* 265:1599-1600.
23. Lee W, *et al.* (2010) Full reconstruction of a vectorial protein folding pathway by atomic force microscopy and molecular dynamics simulations. *J Biol Chem* 285:38167-38172.
24. Cabrita LD, Hsu ST, Launay H, Dobson CM, Christodoulou J (2009) Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy. *Proc Natl Acad Sci USA* 106 :22239-44.
25. Clark PL (2004) Protein folding in the cell: reshaping the folding funnel. *Trends Biochem Sci* 29 :527-534.
26. Tanford C (1968) Protein denaturation. *Adv Prot Chem* 23:121-282.
27. Privalov PL (1979) Stability of proteins: small globular proteins. *Adv Prot Chem* 33:167-241.
28. Kamen DE, Griko Y, Woody RW (2000) The stability, structural organization and denaturation of pectate lyase C, a parallel beta-helix protein. *Biochemistry* 39:15932-15943.
29. Junker M, Clark PL (2009) Slow formation of aggregation-resistant beta-sheet folding intermediates. *Proteins* 78:812-824.
30. Makarov DE, Keller CA, Plaxco KW, Metiu H. (2002) How the folding rate constant of simple, single-domain proteins depends on the number of native contacts. *Proc Natl Acad Sci USA* 99:3535-3539.
31. Schönherr R, Tsolis R, Focareta, T, Braun V (1993) Amino acid replacements in the *Serratia marcescens* haemolysin Sh1A define sites involved in activation and secretion. *Mol Microbiol* 6:1229-1237.

32. Junker M, Besingi RN, Clark PL (2009) Vectorial transport and folding of an autotransporter virulence protein during outer membrane secretion. *Mol Microbiol* 71:1323-1332.
33. Oliver DC, Huang G, Nodel E, Pleasance S, Fernandez RC (2003) A conserved region within the *Bordetella pertussis* autotransporter BrkA is necessary for folding of its passenger domain *Mol Microbiol* 47: 1367-1383
34. Peterson JH, Tian P, Ieva R, Dautin N, Bernstein HD (2010) Secretion of a bacterial virulence factor is driven by the folding of a C-terminal segment. *Proc Natl Acad Sci USA* 107:17739-17744.
35. Andrade MA, Perez-Iratxeta C, Ponting CP (2001) Protein repeats: structures, functions, and evolution. *J Struct Biol* 134:117-131.
36. Fürste JP, *et al.* (1986) Molecular cloning of the plasmid RP4 primase region in a multi-host-range *tacP* expression vector. *Gene* 48:119-131.
37. Kovach ME, *et al.* (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* 166:175-176.
38. Lambert-Buisine C, Willery E, Locht C, Jacob-Dubuisson F (1998) N-terminal characterization of the *Bordetella pertussis* filamentous haemagglutinin. *Mol Microbiol* 28:1283-1293.
39. Locht C, Geoffroy MC, Renauld G (1992) Common accessory genes for the *Bordetella pertussis* filamentous hemagglutinin and fimbriae share sequence similarities with the *papC* and *papD* gene families. *EMBO J* 11:3175-3183.
40. Baud C, *et al.* (2009) Role of DegP for two-partner secretion in *Bordetella*. *Mol Microbiol* 74:315-329.
41. Clementi C, Nymeyer H, Onuchic JN (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol* 298:937-953.
42. Noel JK, Whitford PC, Sanbonmatsu KY, Onuchic, JN (2010) SMOG@ctbp: simplified deployment of structure-based models in GROMACS *Nucl Acid Res* 38:W657–W661.
43. Hess B, Lindhal, E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput* 4:435-447.

Figure legends.

Figure 1. Schematic representation of the proteins used in this work. A. Model of FHA based on ref 11. The TPS domain is red, and the R1, B1, R2 and B2 repeats are shown in green, blue, orange and yellow, respectively. An additional, non-repetitive domain of unknown structure is found at the C terminus. The four dotted vertical lines at the right side of the model indicate the approximate lengths of the proteins used in this work, i.e. FHA, Fha44, Fha60 and Fha30 (from left to right). B. Structure of Fha30. The less conserved subdomains LC1 and LC2 are shown in blue, the conserved subdomains C1 and C2 in grey, and the first

R1 coils in magenta. Arrowheads indicate the positions of linker insertions.

Figure 2. Unfolding of Fha30 by single-molecule AFM. A. Experimental set-up. B. Representative force-distance curve obtained by stretching a single fragment, showing periodic features reflecting sequential unfolding peaks. Force peaks were well described by the worm-like-chain model (red line), using a persistence length of 0.4 nm. The curve shown is representative of a total of more than 1,000 curves obtained using 3 independent tips and 2 sample preparations. C. Superimpositions of 10 typical force curves showing that the last three force peaks in particular were reproducibly observed.

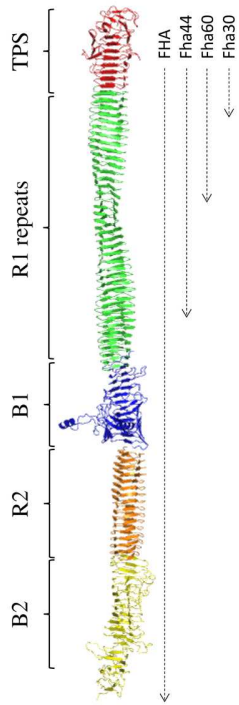
Figure 3. Superimpositions of 10 typical force curves of Fha30 (A) and Fha30_GS2 (B) showing the apparition of an additional force peak 6.6 nm after the penultimate one.

Figure 4. Unfolding of Fha30 simulated by SMD simulations. A. A complete unfolding F-D curve aligned with the experimental curve. The experimental curve is 12 nm longer, which corresponds to the rise of the molecule (5 nm) + the extended length of the 20 C-terminal residues present in the experimental construct. The colored arrows indicate the positions of the simulation where the contact map was calculated. B. Contact map of Fha30 using the same color code. Each point represents a distance < 0.6 nm in the structure. Each colored region of the map indicates the structure that unraveled in the force peak preceding the corresponding arrow in figure A. C. Representation of the Fha30 structure using the same color code as in A and B. The substructures with the same color unraveled simultaneously (see text).

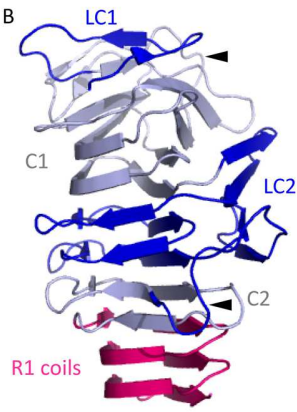
Figure 5. Unfolding of full-length FHA by single-molecule AFM. A. Representative F-D

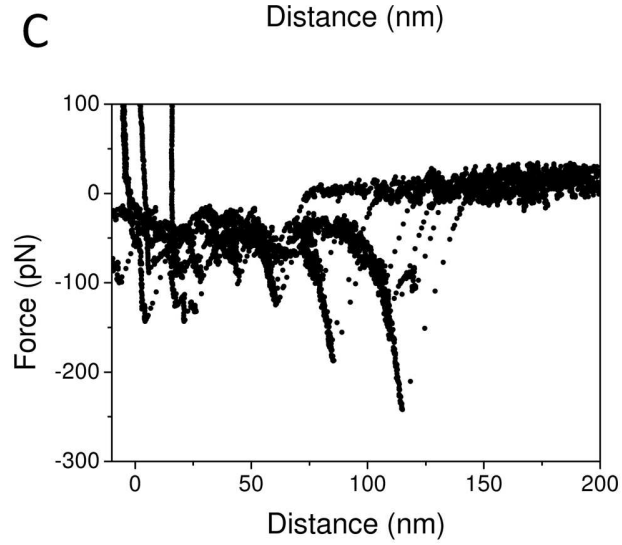
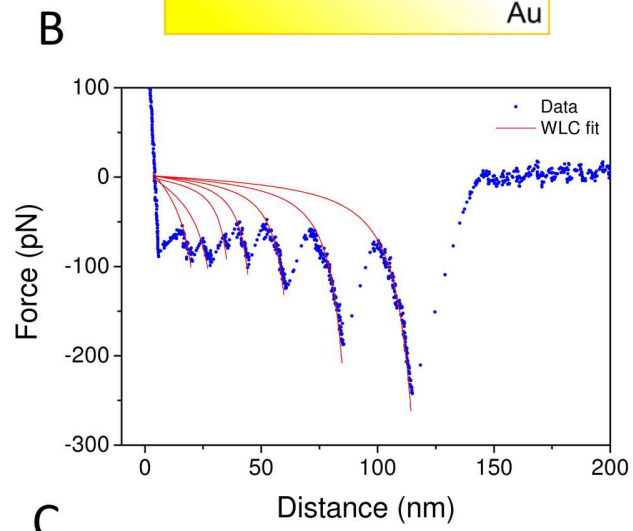
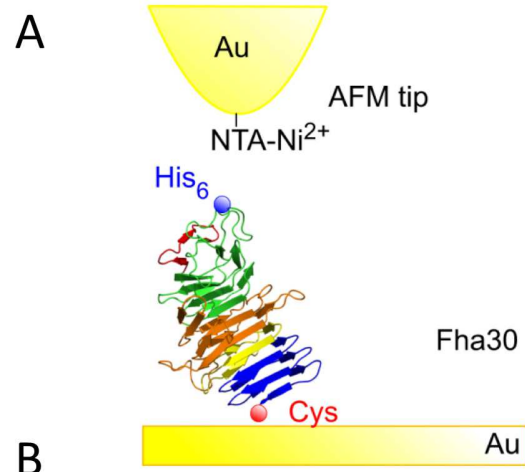
curves of FHA adsorbed on an Au surface and stretched by a Si_3N_4 AFM tip. Variations of the total extension lengths are due to the random immobilization of FHA and the use of a non-functionalized tip. B. WLC analysis of a representative curve. C. Histogram showing the frequency of peak-to-peak distances (measured on 100 force curves). Gaussian analyses were performed to extract most frequent values.

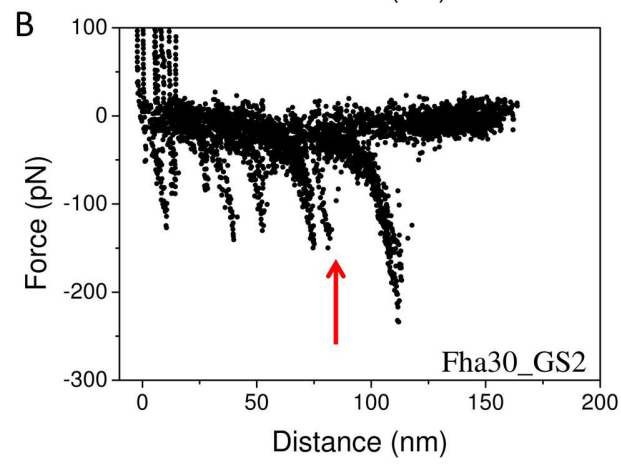
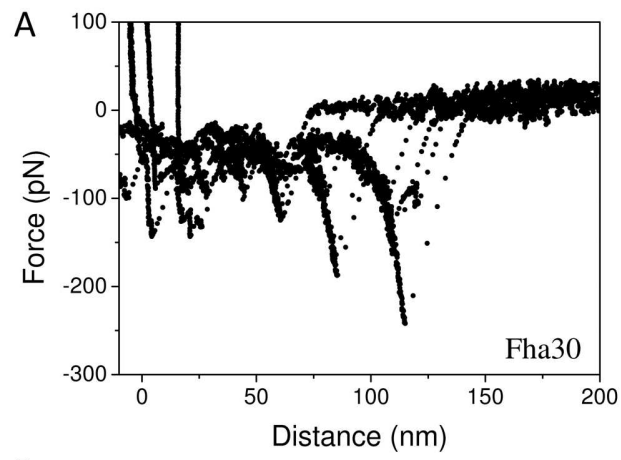
A

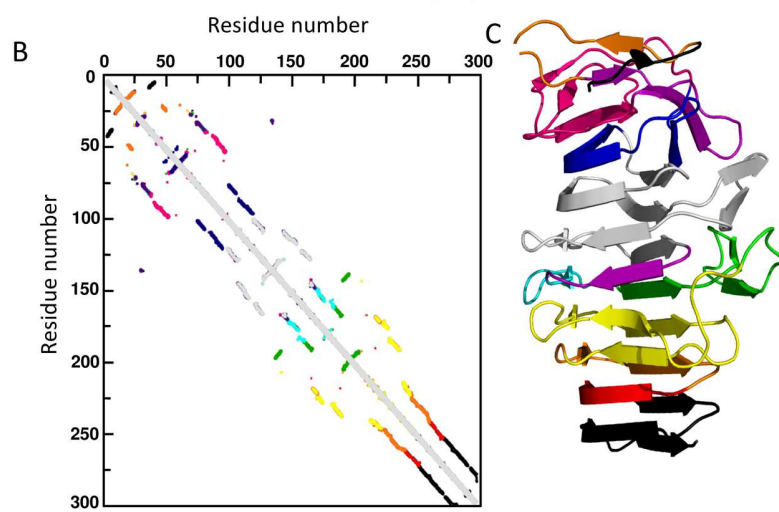
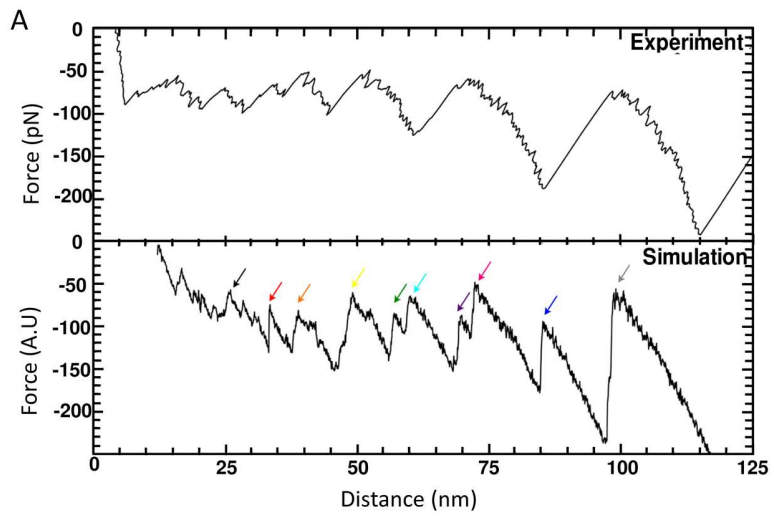


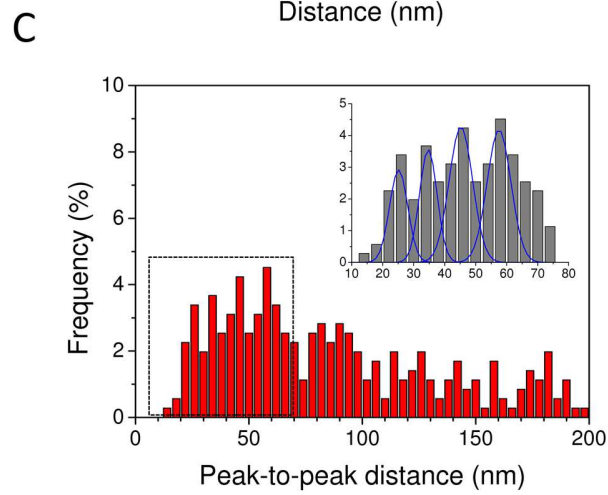
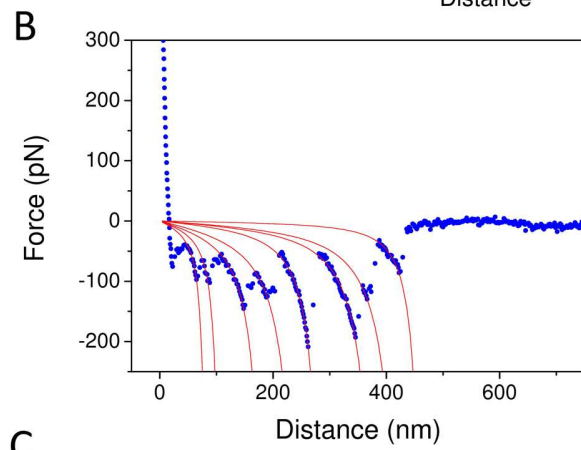
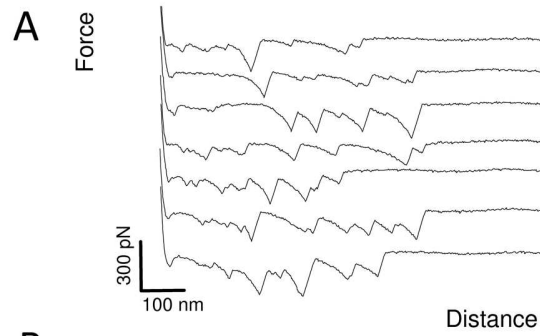
B











3.3.3 Etude des propriétés dynamiques de Fha30 par diffusion de neutrons incohérente

Afin de caractériser cette structure en hélice- β d'un point de vue dynamique, nous avons utilisé la diffusion de neutrons incohérente élastique sur deux instruments : IN13 et IN16. IN13 possède une résolution instrumentales de $8 \mu\text{eV}$, ce qui correspond à des mouvements ayant un temps caractéristique de l'ordre de 100 ps ou plus. IN16 à une résolution instrumentale plus fine $1 \mu\text{eV}$, nous permettant de mesurer de mouvements ayant temps caractéristique de 1 ns ou plus. Ces mesures permettent de calculer la résilience de la protéine [Zaccai 2000], ce qui caractérise la rigidité de la protéine. Le taux d'hydratation des échantillons est 300mg/mL . Les mesures sur IN13 ont été faites dans un solvant H_2O , alors que les mesures sur IN16, qui possède une résolution plus fine, on été réalisées dans un solvant D_2O afin de n'avoir que la contribution des hydrogènes de la protéines.

3.3.3.1 Mesures de diffusion élastique sur IN13

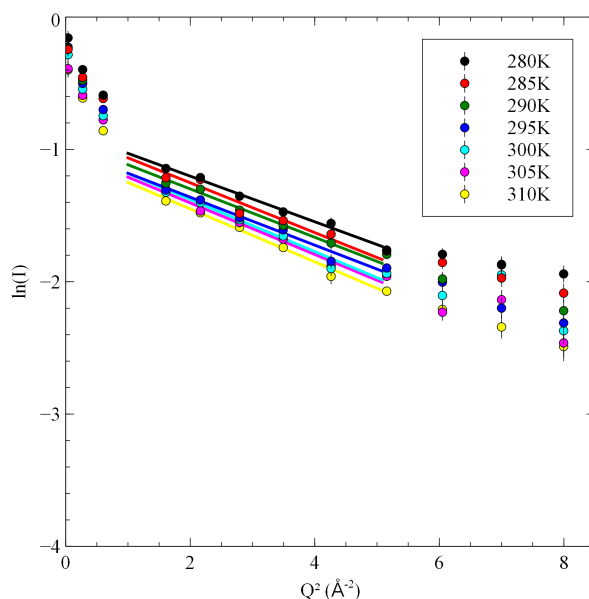


FIGURE 3.6 – Courbes de diffusion élastique incohérente sur IN13, $\ln I$ en fonction de Q^2 . Les lignes représentent la courbe obtenue dans le cadre de l'approximation gaussienne.

Les données ont été analysées en utilisant l'approximation gaussienne dans une gamme de Q^2 allant de 1.6 à 4.2\AA^{-2} (figure 3.6), ce qui nous a permis d'extraire des déplacements carrés moyens. L'analyse de l'évolution des déplacements carrés moyens en fonction de la température nous a permis d'extraire une constante de force quasi-harmonique ([Zaccai 2000], [Bicout 2001]) de 0.858 N/m . Les résultats

de cette analyse sont montrés figure 3.7.

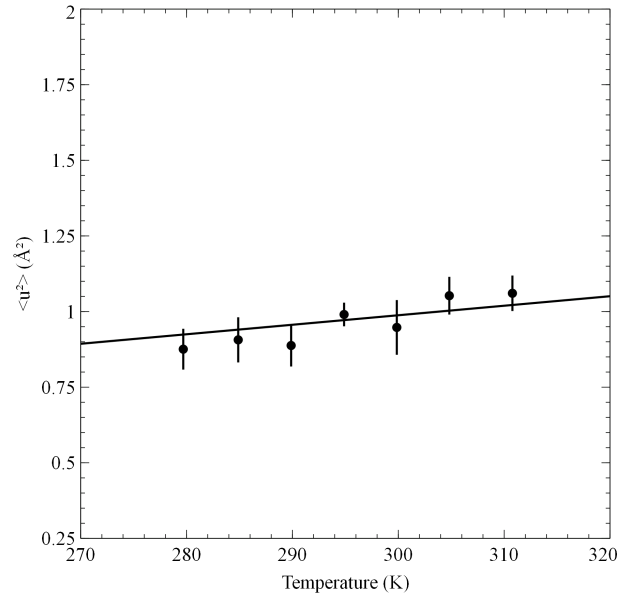


FIGURE 3.7 – Déplacement carrés moyens mesurés à partir des courbes de diffusion sur IN13, en fonction de la température

La mesure de la constante de force nous donne une valeur de 0.858 N/m. La résilience est donc plus élevée pour cette protéine que celle mesurée pour l'hémoglobine (0.28 N/m) [Doster 1989] [Bicout 2001], structurée exclusivement en hélice- α . Cette rigidité augmentée des protéines en feuillets β par rapport aux protéines ayant une structure en hélices α avait déjà été décrite dans [Andreani 1995]. Les auteurs avaient mesuré des spectres quasiélastiques de la superoxide dismutase (composée exclusivement de feuillets β), et avaient comparé leurs spectres à ceux de la myoglobine (composée exclusivement d'hélices α) [Doster 1989]. Leurs mesures indiquaient que la composante quasiélastique était plus faible pour la superoxide dismutase, indiquant une plus grande rigidité.

Il faut aussi noter que les expériences sur la myoglobine(exclusivement hélices- α)[Doster 1989] et la Superoxide dismutase (exclusivement feuillets- β) [Andreani 1995] ont été faites sur des protéines sous forme de poudres hydratées à 0.4g de solvant pour 1g de protéine. Néanmoins, il a été démontré pour la myoglobine que la résilience reste la même dans ces conditions [Zaccai, G. 2000].

3.3.3.2 Mesures de diffusion élastique sur IN16

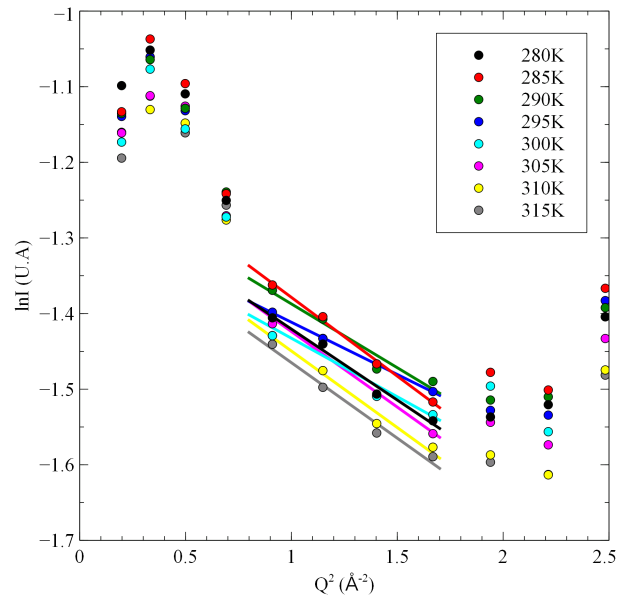


FIGURE 3.8 – Courbes de diffusion élastique incohérente sur IN16, $\ln(I)$ en fonction de Q^2 .

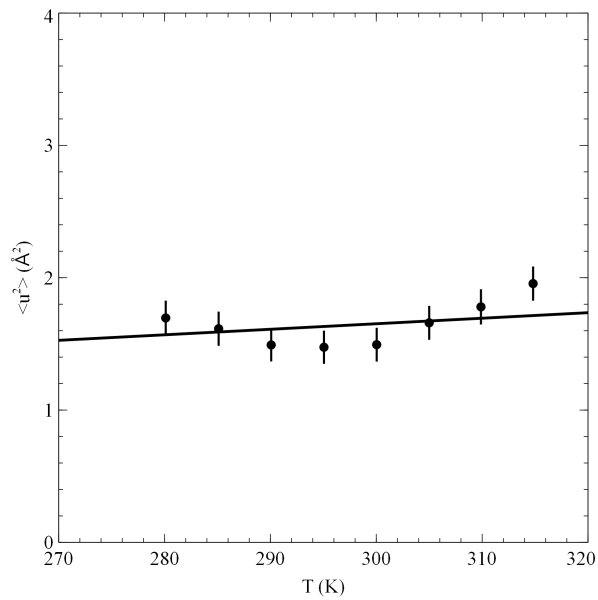


FIGURE 3.9 – Déplacement carrés moyens mesurés à partir des courbes de diffusion obtenues sur IN16, en fonction de la température

L'analyse des données dans le cadre de l'approximation gaussienne est montrée figure 3.8, en utilisant une gamme de Q^2 allant de 0.9 à 1.6 \AA^{-2} . L'évolution des déplacements carrés moyens en fonction de la température sont montrés figure 3.9. On calcule une résilience de 0.673 N/m. Cette mesure est plus basse que celle mesurée sur IN13, et ce à cause de la résolution instrumentale. Cet effet (augmentation de $\langle k' \rangle$ lorsque l'on baisse la résolution instrumentale) avait déjà été observé pour des sucres [Jasnin 2010]. On ne peut donc pas conclure sur un quelconque effet du D_2O sur la résilience de cette protéine, comme cela l'avait déjà été fait pour la BSA ou encore la malate déshydrogénase [Tehei 2001].

3.4 Conclusion

Les essais de conception d'un mutant de Fha30 pouvant se replier correctement n'ont pas abouti soit parce que je n'ai pas réussi à exprimer la protéine recombinante dans un système d'excrétion, soit parce que la protéine possédait encore une interface de dimérisation. Ceci peut s'expliquer par le fait le système d'excrétion ne fonctionne que si la protéine est incapable de se replier correctement. En effet, cette protéine doit traverser le canal formé par FhaC sous une forme dépliée, et le fait d'avoir un système se repliant spontanément peut donc abolir cette étape. Des expériences où l'on avait fusionné un peptide globulaire sur la partie C-terminale de Fha30 avaient aussi aboli l'excrétion, et il avait été suggéré que ce domaine se repliait dans le périplasme, empêchant le transport au travers la membrane [Guedin 1998]. Les expériences de dépliement mécanique vont dans le sens d'un repliement de FHA vectoriel à la surface de la cellule. Le repliement pourrait donc se faire dès que la chaîne polypeptidique sort du canal, sans qu'elle soit entièrement excrétée. Le transporteur pourrait agir comme une contrainte, en "obligeant" la protéine à rester dépliée dans sa partie C-terminale pour que celle-ci puisse traverser le canal. Cette hypothèse va aussi dans le sens d'un couplage entre le transport au travers la membrane et le repliement.

Enfin, les expériences de diffusion incohérente nous montrent que cette structure en feuillets β est plus résiliente qu'une structure en hélice α . Des mesures pour comparer l'effet du D_2O sur la résilience de cette protéine pourraient permettre de mieux la caractériser. On pourrait alors voir si le D_2O a un effet "entropique" en favorisant l'accès à plus d'états sous-conformationnels ou "enthalpique" en favorisant la force de liaison D entre la protéine et le solvant [Tehei 2001]. Les auteurs avaient montré que le D_2O avait un effet stabilisant sur ces deux protéines, alors les mesures montraient une résilience diminuée dans le cas de la BSA dans un solvant D_2O alors que pour la malate déshydrogénase celle-ci était augmentée lorsque l'on passait dans un solvant D_2O . Cette résilience très élevée pourrait être aussi le reflet du rôle biologique de cette protéine, dont le rôle est de maintenir la cellule accrochée sur le tissu infecté.

Analyses structurales des éléments du complexe de réplication des *Rhabdoviridae*

"Somewhere between the sacred silence and sleep, disorder, disorder, disorder."

System of a Down, Toxicity

Sommaire

4.1 Introduction	93
4.2 Matériels et méthodes	95
4.2.1 Préparation des échantillons :	95
4.2.2 Collecte des données	96
4.2.3 Modélisation et analyse d'ensembles	96
4.3 Résultats et discussion	96
4.3.1 Modélisation du complexe N ⁰ -P	96
4.3.2 Détermination de la structure haute résolution du domaine C-terminal de la phosphoprotéine du virus de la jaunisse né- crotique de la laitue (LNYV)	102
4.4 Conclusion	138

4.1 Introduction

Les *Rhabdoviridae* sont des virus enveloppés à ARN négatif. Ils possèdent une morphologie caractéristique en forme de balle de fusil [Howatson 1962]. Leur génome est composé d'une molécule d'ARN négatif et code pour 5 protéines structurales :

- La glycoprotéine (G), présente à la surface du virion [David 1973], qui permet l'entrée du virus dans la cellule hôte [Matlin 1982].
- La protéine matrice (M) est une protéine multifonctionnelle, qui intervient dans l'assemblage du virion [Chong 1993], mais aussi dans les processus de réplication/transcription [Finke 2003].
- La polymérase (L) qui porte l'activité enzymatique de transcription et de réplication [Baltimore 1970] de l'ARN viral, et est aussi responsable de la maturation des ARNm viraux en formant leur coiffe [Rhodes 1974] et leur queue

poly-A [Banerjee 1974]. Son substrat est exclusivement la matrice ARN-nucléoprotéine (N-ARN) [Emerson 1972]

- La nucléoprotéine (N), qui encapside l'ARN génomique, formant le complexe N-ARN qui permet à l'ARN d'avoir une grande résistance aux ribonucléases [Soria 1974]. Ce complexe forme des structures oligomériques hélicoïdales [Sokol 1969], et sa structure haute résolution est connue chez différents virus [Albertini 2006], [Green 2006]. Elle forme aussi un complexe avec la phosphoprotéine [Mavrakis 2003] appelé N⁰-P qui empêche N d'encapsider des ARN cellulaires.
- La phosphoprotéine joue le rôle de chaperonne de la N, mais elle a aussi des fonctions de cofacteur de L [Emerson 1975], et elle interagit avec plusieurs protéines de l'immunité innée [Brzozka 2005] ou du cytosquelette [Jacob 2000] de la cellule hôte.

Les phosphoprotéines des *Rhabdoviridae* partagent une organisation structurale similaire. Elles se composent de trois parties reliées par des domaines flexibles [Gerard 2009](figure 4.1) :

- La partie N-terminale est responsable de la fonction chaperonne de la nucléoprotéine et elle permet à la phosphoprotéine de former le complexe N⁰P [Mavrakis 2003] [Leyrat 2011b].
- Le domaine central est responsable de l'oligomérisation et dont la structure a été résolue par cristallographie de rayons X pour le virus de la rage [Ivanov 2010] et le VSV [Ding 2006]
- Le domaine C-terminal fixe le complexe N-ARN permettant ainsi de faire le lien entre L et sa matrice N-ARN [Paul 1988]. Les structures de VSV [Ribeiro 2008] et du virus de la rage [Mavrakis 2004] sont connues.

Ces différents domaines sont reliés entre eux par des parties flexibles non structurées [Ribeiro 2008] [Gerard 2009].

Le complexe N⁰-P est un élément indispensable lors de la réplication et la transcription, car il empêche l'encapsidation d'ARN cellulaires et permet d'encapsider génomes et antigénomes synthétisés lors de la réplication. Ce complexe est aussi impliqué dans la régulation du basculement entre les modes réplication et transcription de la polymérase [Gupta 1997]. Récemment la structure d'un complexe mimant le complexe N⁰-P a été résolue. Ce complexe est formé entre une version délétée de la N et la partie N-terminale de la P (résidus de 1 à 60). La première partie de ce chapitre est consacrée à l'étude structurale du complexe N⁰-P de VSV avec la P entière. Comme nous connaissons les structures des différents domaines structurés, et que la P possède un caractère hautement flexible, nous avons utilisé des techniques de diffusion aux petits angles afin d'obtenir des informations structurales.

La deuxième partie du chapitre est centrée sur la structure du domaine C-terminal de la P du virus de la nécrose jaune de la laitue (LNYV). Les domaines C-terminaux des *Rhabdoviridae* de type vesiculovirus et lyssavirus, bien que possédant des séquences très différentes, possèdent des structures similaires [Delmas 2010]. LNYV fait partie

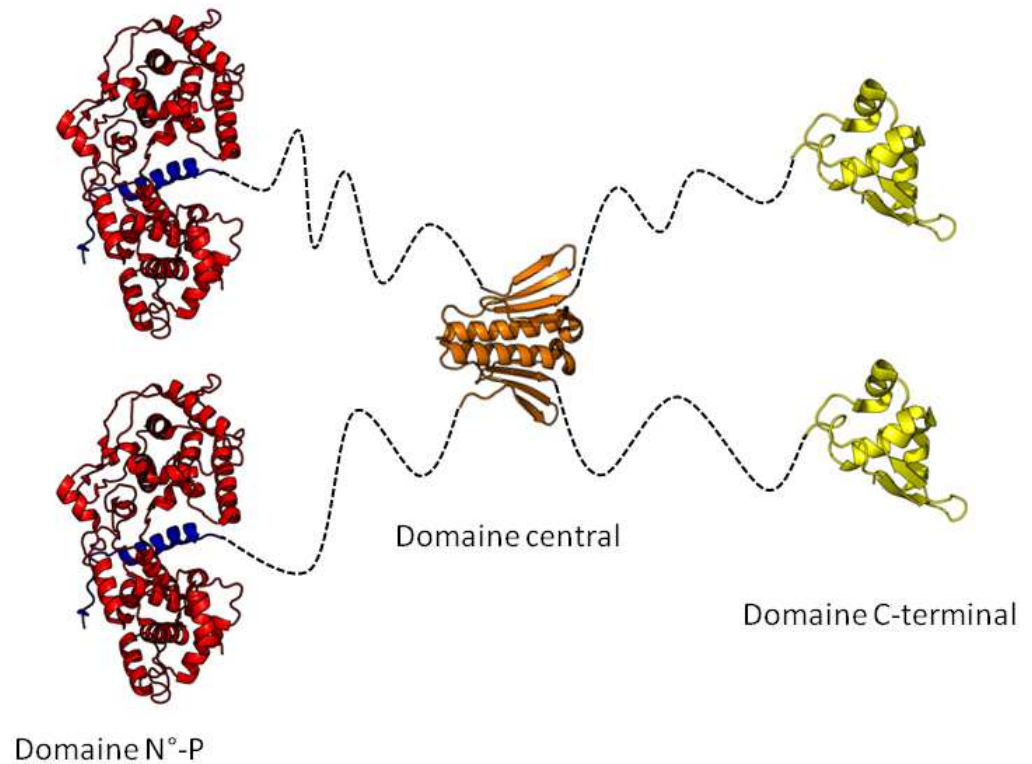


FIGURE 4.1 – Schéma illustrant l'organisation du complexe N°-P.

du genre cytorhabdovirus, un type de *Rhabdoviridae* infectant les plantes. Il était donc intéressant de voir si la structure du domaine C-terminal était aussi conservée chez ce type de virus, malgré une identité de séquence extrêmement faible.

4.2 Matériels et méthodes

4.2.1 Préparation des échantillons :

Les échantillons ont été produits par **Filip Yabukarski**. Toutes les protéines sont exprimées dans des cellules *E. coli* BL21. Le domaine N-terminal (P-NTD) de la phosphoprotéine a été purifié sur colonne Ni-NTA puis par chromatographie d'exclusion de taille. Un mutant de la nucléoprotéine (N Δ 21) [Leyrat 2011b] fusionné avec un tag MBP clivable à la TEV a été purifié sur colonne d'amylose. Cette colonne

permet d'immobiliser les protéines possédant une affinité pour le maltose (qui sont donc fusionnées à la MBP). Ces deux échantillons sont mélangés, puis purifiés sur colonne Ni-NTA et colonne d'exclusion de taille. La phosphoprotéine entière, produite en milieu deutéré (au D-lab de l'ILL) est aussi purifiée sur colonne Ni-NTA suivie d'une chromatographie d'exclusion de taille. On mélange ensuite le complexe N/P-NTD avec la phosphoprotéine entière, en mettant un excès molaire de P de 5 pour 1. Ce mélange est ensuite passé sur une colonne d'amylose et incubé pendant 40h avec la protéase TEV et purifié sur colonne Ni-NTA, suivie d'une chromatographie d'exclusion de taille. On obtient ainsi le complexe N⁰-P. Le complexe à ensuite été dialysé dans des concentrations de 20, 40 et 100% de D₂O.

4.2.2 Collecte des données

Les données de diffusion ont été collectées sur l'instrument D22 de l'ILL à deux distances échantillon-détecteur différentes. L'intégration radiale et la soustraction du tampon ont été faites avec le programme SPOLLY.

4.2.3 Modélisation et analyse d'ensembles

Les modèles de départ ont été générés par le **Dr. Cédric Leyrat** à partir des structures haute résolution du domaine central de P (P-CED) [Ding 2006], du domaine C-terminal de P (P-CTD) [Ribeiro 2008] du complexe N-ARN/P-CTD [Green 2006], et du complexe N⁰-P. Les linkers flexibles ont été générés avec le programme Gajoe [Bernado 2007], et les différentes parties ont été reliées entre elles manuellement.

Les simulations de dynamique moléculaire ont été réalisées comme décrit dans [Jamros 2010] : le complexe est représenté par un champ de force simplifié décrit dans [Clementi 2000b]. Ce champ de force ne prend pas en compte les interactions électrostatiques et a une représentation des forces de type Van der Waals qui est simplifiée [Noel 2010]. Cette représentation permet d'accélérer les calculs pour pouvoir générer un ensemble maximal de conformères.

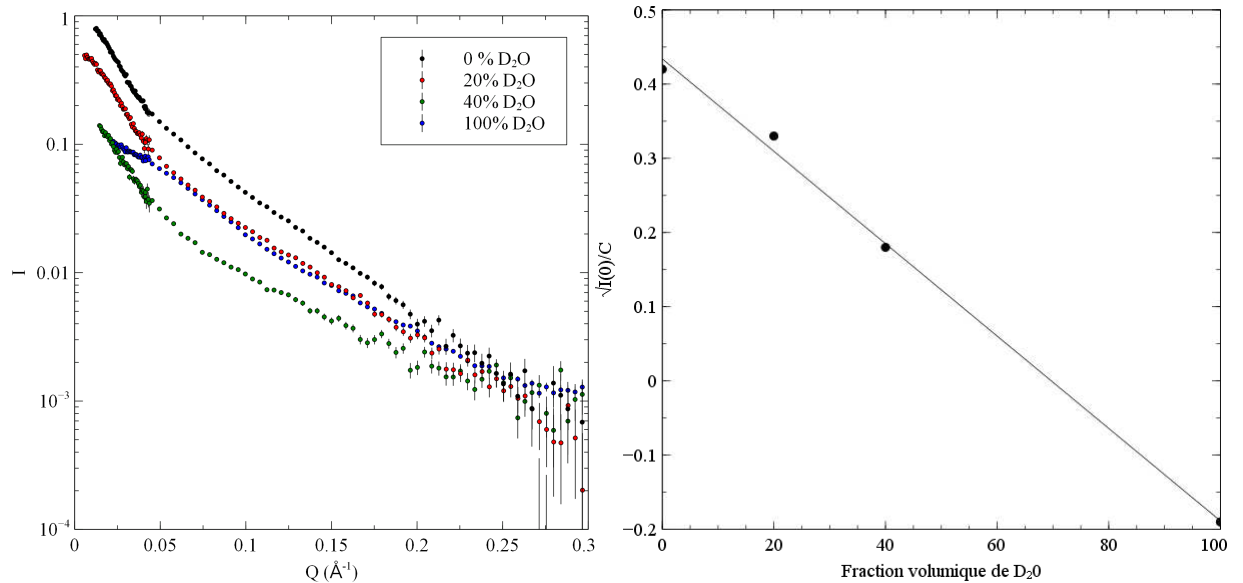
Ensuite, des courbes de diffusion théoriques ont été calculées pour chaque modèle avec le programme Cryson [Svergun 1998] dans les 4 conditions de contraste (0, 20, 40 et 100% de D₂O). Le programme Gajoe [Bernado 2007] a été utilisé pour faire une sélection et trouver un ensemble permettant de reproduire les données expérimentales.

4.3 Résultats et discussion

4.3.1 Modélisation du complexe N⁰-P

Les courbes de diffusion sont montrées figure 4.2a.

Afin de valider l'expérience de variation de contraste, j'ai fait une analyse de Stuhmann, qui est montrée figure 4.2b, sur laquelle on voit que $\sqrt{I(0)/C}$ est

(a) Courbes de diffusion de neutrons aux petits angles du complexe N⁰-P

(b) Analyse de Sturhman

FIGURE 4.2 – Données expérimentales de diffusion de neutrons aux petits angles

proportionnel à la fraction volumique de D₂O. Ceci nous indique que l'expérience de variation de contraste a été concluante, et que nos échantillons sont dépourvus d'agrégats.

Les expériences de SEC-MALLS-RI sur le complexe indiquent que la stœchiométrie du complexe dépend de la manière dont l'échantillon est préparé, principalement sur la proportion N/P utilisée lors du mélange. Selon le rapport N :P utilisé dans la procédure de reconstitution du complexe, l'expérience de SEC-MALLS-RI indique que l'échantillon contient soit 2N par dimère de P, soit un mélange de 2N par dimère de P et de 1 N par dimère de P.

Nous avons donc fait deux types de modèles : des modèles avec une stœchiométrie 1N :2P (figure 4.3a), et des modèles avec une stœchiométrie 2N :2P (figure 4.3b). Nous avons fait la procédure de sélection pour chaque condition de contraste (figure 4.4). Les ensembles sélectionnés dans les 4 conditions de contraste contiennent une forte proportion de modèles avec une stœchiométrie 1 N par dimère de P (70% des modèles en moyenne). La figure 4.6 nous montre une superposition de 5 modèles représentatifs avec 1 N alors que la figure 4.7 nous montre 3 modèles représentatifs avec 2N. Afin de vérifier si notre modélisation était correcte, j'ai effectué une procédure de sélection sur le même ensemble, mais soit en enlevant les modèles 1N :2P, soit les modèles 2N :2P (figure 4.5). On observe que l'on ne peut pas trouver une solution correcte pour toutes les conditions de contraste, et plus particulièrement pour la condition à 100% de D₂O. Cette condition nous donne un contraste nul pour P, la stœchiométrie du complexe prend donc une importance capitale, car seule N

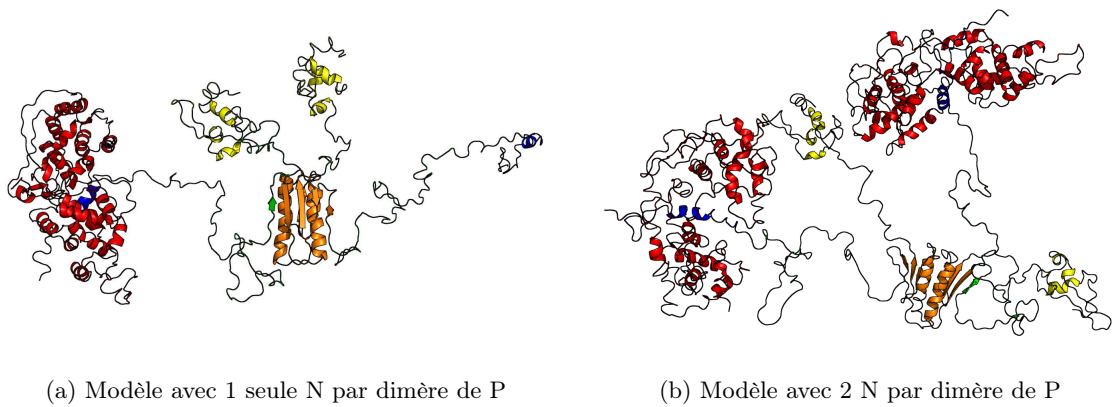


FIGURE 4.3 – Différents types de modèles utilisés lors de la modélisation du complexe N^0 -P. Les domaines centraux de P, PCTD, P-NTD et N sont représentés en orange, jaune, bleu et rouge respectivement.

contribue à la diffusion.

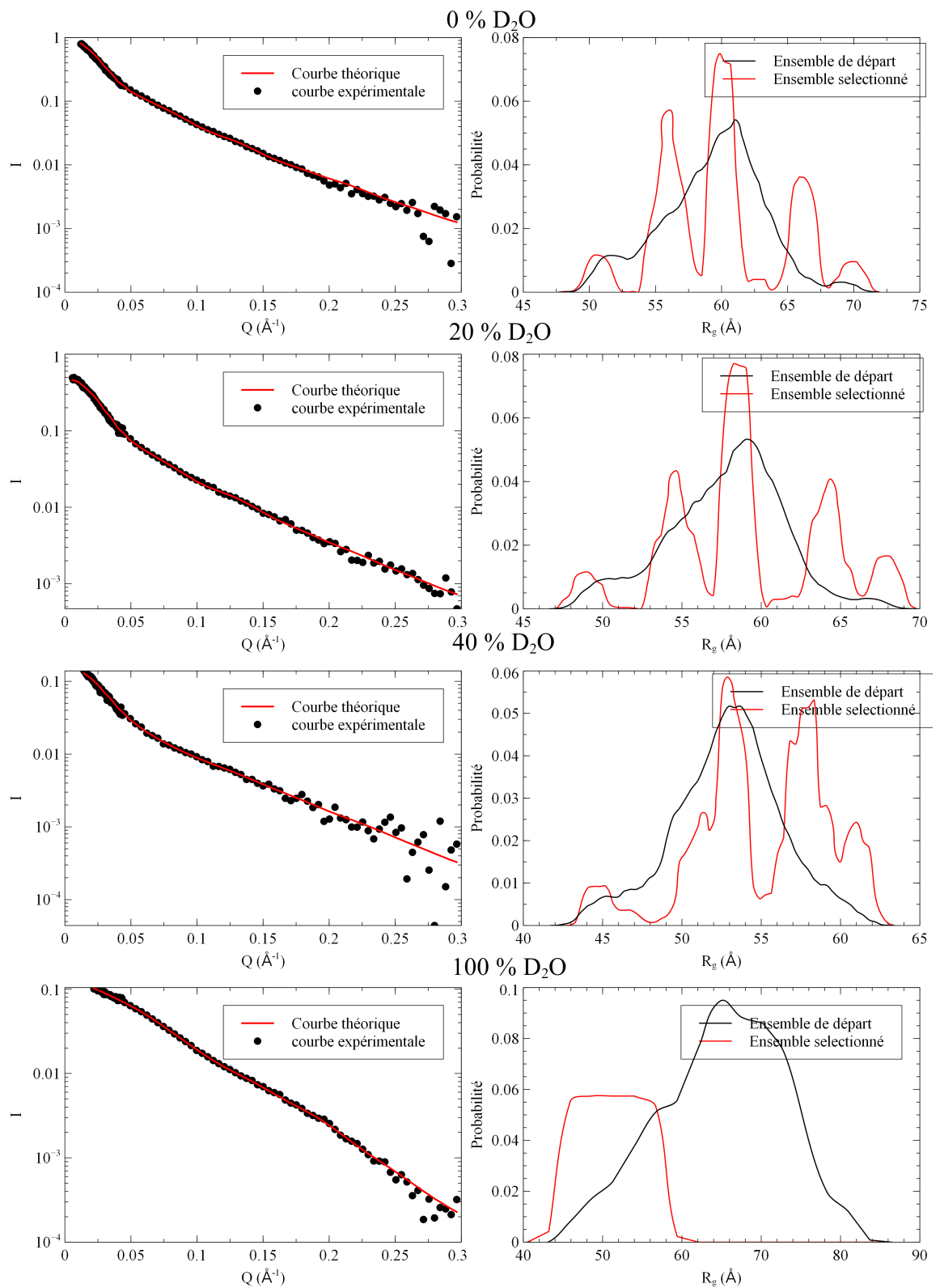


FIGURE 4.4 – Courbes théoriques et expérimentales, et distribution des rayons de giration de l'ensemble de départ et de l'ensemble sélectionné pour l'ensemble final, dans les différentes conditions de contraste.

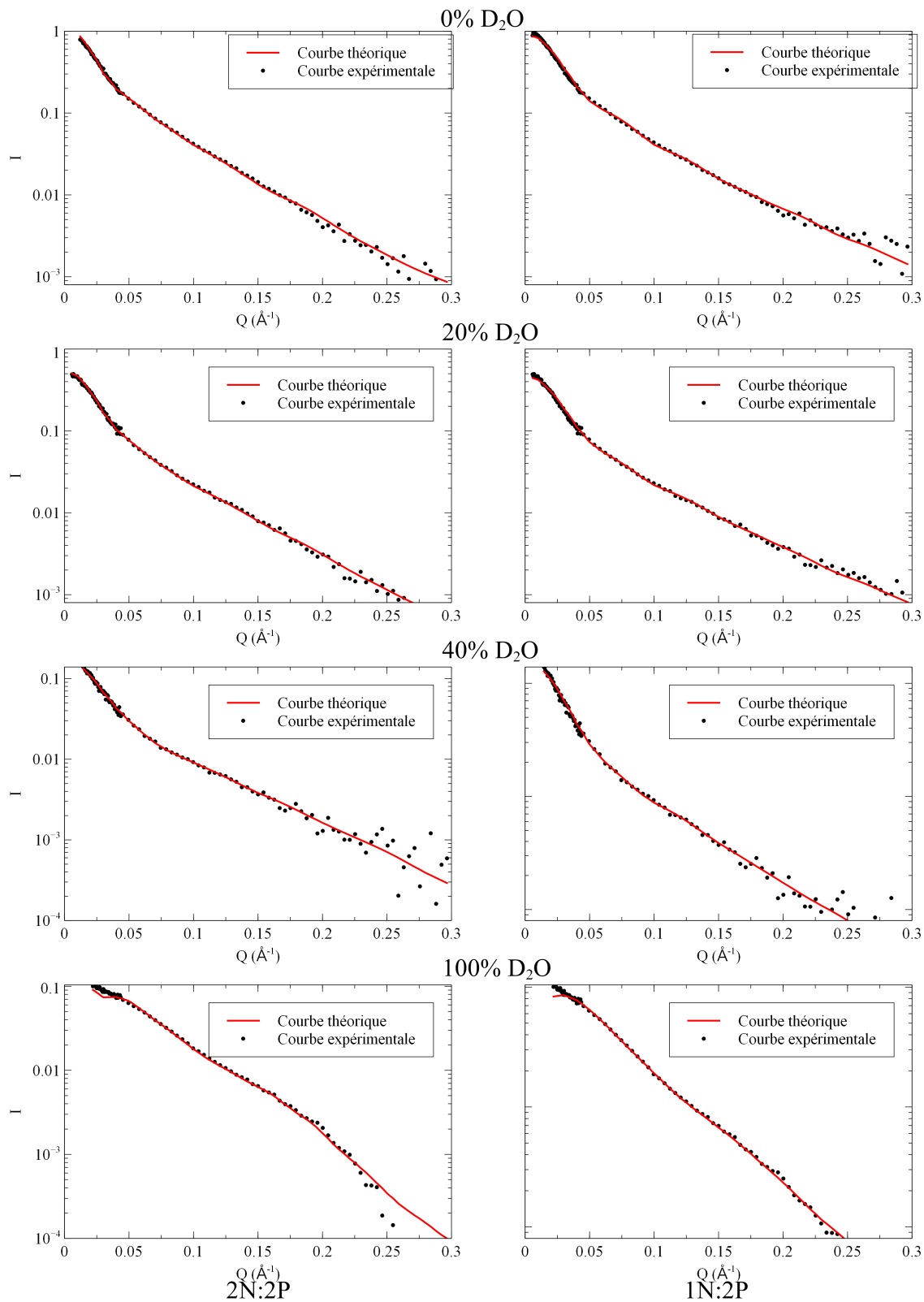
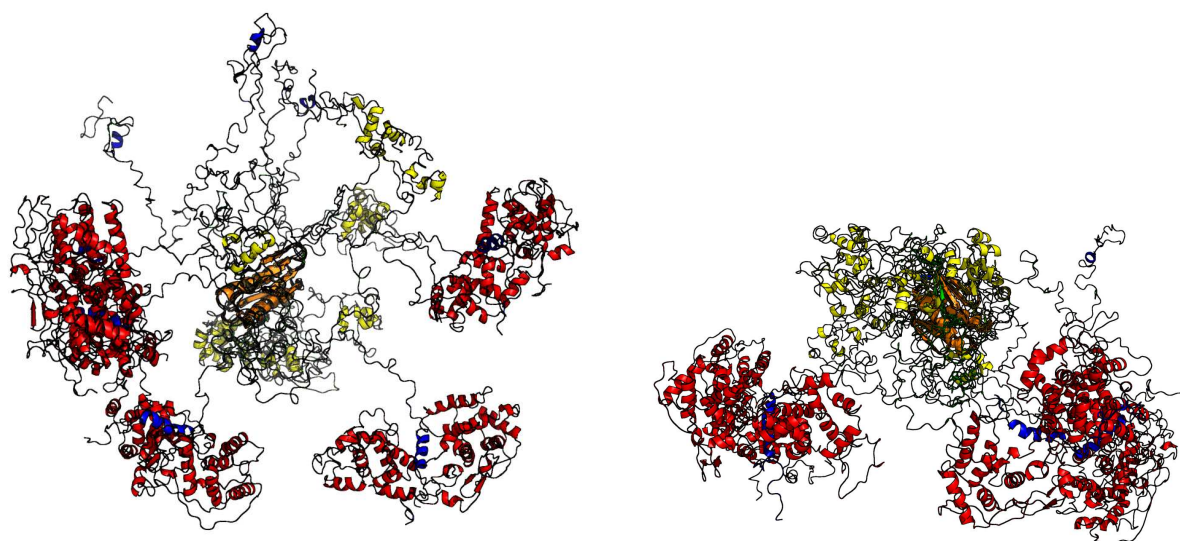


FIGURE 4.5 – Courbes expérimentales et théoriques obtenues avec un ensemble contenant deux N par dimère de P (gauche) et 1N par dimère de P (droite).

Si l'on se repère par rapport au domaine central de la phosphoprotéine, on observe que la zone explorée par la nucléoprotéine est très large, tout comme la partie N-terminale de P libre. Aussi, on remarque que N est plutôt éloignée de la partie N-terminale de P libre (figure 4.6a). Le domaine C-terminal de P a tendance à se retrouver proche du domaine central, mais plutôt dans une région de l'espace opposée à la partie N-terminale (figure 4.6b), avec une distance moyenne de 9 nm entre leurs centres de masse.

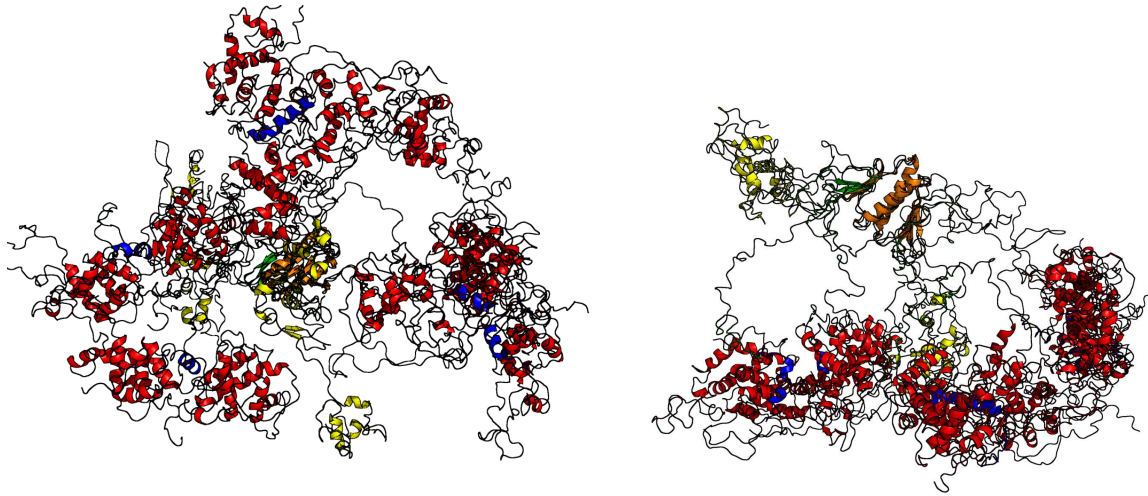


(a) Ensemble représentatif des modèles avec 1 N par dimère de P

(b) Même que 4.6a, tourné à 90°

FIGURE 4.6 – Superposition des modèles avec 1 N par dimère de P du complexe N^0 -P représentatifs des ensembles sélectionnés dans les différentes conditions de contraste. Les domaines centraux qui ont été superposés sont représentés en orange, N, PCTD et P-NTD en rouge, jaune et bleu respectivement.

Pour les modèles ayant deux N par dimère de P, qui représentent donc environ 30% de l'ensemble, on obtient des modèles où les 2 N sont plus proches (figure 4.7a), la distance moyenne entre les centres de masses des parties N-terminales de P est de 11 nm, contre 15 nm lorsque l'on a une seule N.



(a) Ensemble représentatif des modèles avec 2 N par dimère de P

(b) Même que 4.7a, tourné à 90°

FIGURE 4.7 – Superposition des modèles avec 2 N par dimère de P du complexe N^0 -P représentatifs des ensembles sélectionnés dans les différentes conditions de contraste. Les domaines centraux qui ont été superposés sont représentés en orange, N, PCTD et P-NTD en rouge, jaune et bleu respectivement.

Ces résultats nous montrent que N explore un espace conformationnel très large. Ce volume très large exploré par N et la partie C-terminale de N pourrait permettre au complexe, si celui-ci est fixé à une nucléocapside de faciliter l'encapsidation d'un ARN néo-synthétisé, mais aussi pour libérer l'ARN (avec sa partie N-terminale libre) afin qu'il soit pris en charge par L.

Afin d'avoir un seul ensemble représentatif du complexe, il serait plus judicieux d'utiliser cette même méthode en minimisant une χ^2 global, et non pas de faire des sélections séparées pour chaque condition de contraste. Cela nous permettrait d'avoir un seul ensemble et d'avoir une approche plus rigoureuse du problème.

4.3.2 Détermination de la structure haute résolution du domaine C-terminal de la phosphoprotéine du virus de la jaunisse nécrotique de la laitue (LNYV)

Le virus de la jaunisse nécrotique de la laitue (LNYV) a été classé dans le genre des *rhabdoviridae* principalement sur des critères morphologiques (virion en forme de balle typique des *rhabdoviridae*) [Jackson 2005]. Le séquençage de son génome a permis d'identifier plusieurs de ses protéines, dont la phosphoprotéine et la nucléoprotéine, mais aucune de ces protéines n'avait été étudiée structuralement. Sur la base d'une analyse de prédiction de désordre [Gerard 2009], j'ai choisi d'exprimer une partie C-terminale de 9 kDa. Cette construction s'est révélée, comme la plupart des domaines P-CTD de *Rhabdoviridae* connus, une unité de repliement autonome [Ribeiro 2008] [Mavrakis 2004]. J'ai donc entrepris une étude structurale sur ce domaine, que j'ai réussi à cristalliser et j'ai pu obtenir une structure haute résolution

par cristallographie aux rayons X.

La structure est composée de 4 hélices α qui sont organisées de manière similaire à ce que l'on peut trouver chez les autres virus de la famille des *Rhabdoviridae*, comme par exemple le virus de la rage [Mavrakis 2004] ou encore le virus de Mokola, et ce malgré une identité au niveau de la séquence en acides aminés très faible. Cette structure nous permet de confirmer l'appartenance de LNYV à la famille des *Rhabdoviridae*, et aussi de supposer un mode de fonctionnement similaire de la machinerie répliquative des virus de cette même famille.

Manuscrit en préparation

**Structure of the C-terminal domain of
Lettuce Necrotic Yellows Virus phosphoprotein**

**Nicolas Martinez^{1,2}, Euripedes A. Ribeiro Jr¹, Nicolas Tarbouriech¹, Rob W.H. Ruigrok¹
and Marc Jamin^{1*}**

¹UMI 3265 UJF-EMBL-CNRS, Unit of Virus Host Cell Interactions

6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France

²Institut Laue Langevin, 6 Rue Jules Horowitz 38042 Grenoble Cedex 9, France.

Running title:

Keywords: phosphoprotein, Rhabdoviridae, plant virus

* **Corresponding author:** Unit of Virus Host Cell Interactions (UVHCI), UMI 3265 UJF-EMBL-CNRS, 6, rue Jules Horowitz, B.P. 181, 38042 Grenoble Cedex 9, France

E-mail : jamin@embl.fr, Phone: + 33 4 76 20 94 62, Fax : + 33 4 76 20 94 00

ABSTRACT

The Lettuce Necrotic Yellows Virus (LNYV) is a plant virus considered as a prototype of the cytorhabdoviruses. By a meta-prediction of disorder, the folded C-terminal domain of its phosphoprotein was localized in the amino acid sequence, and the protein fragment was expressed as a recombinant and purified. This domain is an autonomous folding unit that is monomeric in solution. The crystal structure reveals a lollipop-shaped structure comprising five helices. The topology of this domain shows similarities with that of corresponding phosphoprotein domains from other Rhabdoviridae as well as from Paramyxoviridae and Filoviridae, revealing an evolutionary relationship between the phosphoproteins of the non-segmented negative strand RNA viruses. In addition, the structure of the nucleoprotein of LNYV was built by homology modeling. The predicted structure is similar to that of RAV and VSV N proteins providing a molecular confirmation that LNYV belongs to the Rhabdoviridae.

INTRODUCTION

Numerous viruses infecting plants are classified as rhabdoviruses most generally on the basis of electron microscopic observations of distinctive enveloped bacilliform or bullet-shaped particles in the infected cells (Jackson et al., 2005). Only a few of them were confirmed to be true rhabdoviruses by sequencing of their genome and molecular studies (Kormelink et al., 2011). These non-segmented negative strand RNA viruses (NNV) form the family Rhabdoviridae that is classified in the order Mononegavirales with three other families of NNV, the Paramyxoviridae, the Filoviridae and the Bornaviridae. The Rhabdoviridae is currently subdivided into six genera and the putative plant rhabdoviruses are classified in two genera on the basis of their main site of morphogenesis; the Cytorhabdoviruses replicate and assemble in the cytoplasm whereas the Nucleorhabdoviruses replicate in the nucleus (Tordo et al., 2004).

The lettuce necrotic yellows virus (LNYV) was identified in 1963 in Australia (John, 2004). It is spread in Australia and New Zealand, but isolation of this virus or of a similar virus was also reported in Italy, Spain and Great Britain. Infection by LNYV causes fading of the leaves and necrosis of the heads making the lettuce unsuitable for consumption. LNYV is transmitted from plant to plant by an aphid of the genus *Hyperomyzus* (John, 2004) and the main natural reservoir of both virus and vector is the common sowthistle (*Sonchus oleraceus*) that is symptomlessly infected. Disease outbreaks occur where the natural reservoir is densely populated and coincide with flights of the vector. The virus is not transmitted by seeds.

LNYV is a prototypic member of the Cytorhabdovirus genus that infects different tissues of the plants (Chambers et al., 1965; Chambers & Francki, 1966), but little is known about its mechanisms of replication. Primarily, RNA synthesis occurs in the nucleus (Wolanski & Chambers, 1971), but in a later stage of viral replication, RNA synthesis relocated in cytoplasmic inclusions containing virus particles (Wolanski & Chambers, 1971). The genome of LNYV was completely sequenced revealing an organization conserved among other rhabdoviruses. The genome consists of six genes separated by intergenic regions that contain highly conserved consensus sequences and flanked by untranslated 3' leader and 5' trailer (Dietzgen et al., 2006). On the basis of sequence analysis and immunolabeling, the genes were proposed to encode successively from the 3' extremity to the 5' extremity, the nucleoprotein (N), the phosphoprotein (P), a plant specific 4b protein assisting in cell-to-cell movement of viral complexes through plasmodesmata and the vascular system, the matrix protein (M), the

glycoprotein (G) and the large subunit of the RNA-dependent RNA polymerase (L) (Callaghan & Dietzgen, 2005; Dietzgen et al., 2006; Dietzgen & Francki, 1988; Wetzel et al., 1994a; Wetzel et al., 1994b). However, no experimental data are available about the functions and structure of these proteins or about the relationships with the corresponding proteins from other NNV.

The structure and function of the P protein from two other genera of rhabdoviruses have been largely studied, namely the P protein from vesicular stomatitis virus (VSV), a Vesiculovirus, and that of rabies virus (RAV), a Lyssavirus (Ivanov et al., 2011). In VSV and RAV, the phosphoprotein plays multiple roles during the replication process. P forms a two-subunit RNA-dependent RNA polymerase with the L protein, in which P acts as a non-catalytic cofactor, connecting and correctly positioning the polymerase on its template and participating in the motion of the polymerase along its template (Emerson & Yu, 1975). The genome of the rhabdoviruses, like that of all non-segmented negative-sense RNA viruses, is entirely coated by the viral nucleoprotein (N), and it is the N-RNA complex rather than the naked RNA that serves as template for the polymerase (Arnheiter et al., 1985). The replication of the viral genome thus requires a continuous supply of N, and P acts as a chaperone of the nascent nucleoprotein, preventing the auto-assembly of N in the absence of RNA and the encapsidation of cellular RNA (Masters & Banerjee, 1988; Peluso & Moyer, 1988).

VSV and RAV P form dimers that behave as non-globular proteins (Gérard et al., 2007). Limited proteolysis, disorder predictions and biochemical studies of protein fragments revealed that these proteins are multi-domain proteins that consist of two structured domains and two disordered regions (Gérard et al., 2009). A central domain is responsible for the dimerization of the protein, whereas a C-terminal domain binds to the N-RNA template. The structures of both domains of VSV and RAV P, as well as the structure of the C-terminal domain of Mokola virus (MOKV) were solved recently (Assenberg et al., 2009; Ding et al., 2006; Ivanov et al., 2010; Mavrakis et al., 2004; Ribeiro et al., 2008), and a model for the entire protein has been proposed (Leyrat et al., 2010; Wertz, 1978). Despite a lack of sequence conservation, the structure of the C-terminal domain of VSV is homologous to those of RAV and MOKV (Delmas et al., 2010; Ribeiro et al., 2008). Other recent studies revealed the structure of the complex formed between the N-RNA template and a C-terminal domain of P and showed that the binding site for P_{CTD} is made of two adjacent N subunits and thus exists only in polymerized N complexes (Green & Luo, 2009; Ribeiro et al., 2009). The long N-terminal disordered region contains the binding site

for the polymerase and for the RNA-free N molecule (N⁰). The N⁰ binding site of VSV P comprises transiently populated helical segments and constitutes a molecular recognition element (MoRE) that adopts a stable structure upon assembly with its partner (Leyrat et al., 2011a; Leyrat et al., 2011b). A recent structural characterization of a reconstituted complex between the N-terminal region of P and a N variant deleted of its N-terminal sub-domain revealed the molecular mechanisms of the chaperone activities of VSV P (Leyrat et al., 2011b). Structural studies and disorder predictions showed that this modular organization of the P protein is conserved between members of the Rhabdoviridae (VSV and RAV), but resembles also that found among members of the Paramyxoviridae (Karlin et al., 2003; Llorente et al., 2006), the Bornaviridae (Hock et al., 2009) and the Filoviridae (Leung et al., 2009)

Because no functional or structural information was available for any Cytorhabdovirus, we used a meta-prediction of disorder to localize the disordered and structured regions in the sequence of LNYV P. The D-score provides a consensus from sixteen different web-servers that predict the localization of disorder regions from the amino acid sequence of a protein (Gérard et al., 2009; Ribeiro et al., 2008). Consequently, for a multi-domain protein, this meta-prediction of disorder allows to localize the boundaries between ordered and disordered regions. This approach was previously used successfully for localizing structured domains in P proteins from Sendai virus, VSV and RAV (Gérard et al., 2009; Ivanov et al., 2010; Ribeiro et al., 2008). On the basis of this analysis, we cloned, expressed and purified the C-terminal domain of the protein. Its crystal structure revealed a fold that is slightly different from those of VSV, RAV and from those of paramyxoviruses and filoviruses, but that clearly indicates common ancestry. In addition, the predicted structure of the nucleoprotein also demonstrates that the fold of this protein is well conserved among the Rhabdoviridae.

RESULTS

Meta-prediction of disorder

Multiple sequence alignments revealed significant conservation within each genus but not between genii (Supplementary material: Figure S1), and even the known localization of P_{CED} and P_{CTD} in the Vesiculovirus and Lyssavirus was not sufficient for localizing corresponding domains in the LNYV P sequence. To localize the C-terminal domain, we therefore used the method previously developed for calculating a consensus from multiple disorder predictions (Gérard et al., 2009; Ribeiro et al., 2008). This D-score calculated from the amino acid sequence of LNVY P predicted the presence of two intrinsically disordered regions (IDR) with contiguous D-score values lower than 0.5 (IDR_{NT}: aa 1-63 and IDR_{CT}: aa 184-230) and two folded regions with contiguous D-score values higher than 0.5, a central bipartite domain (P_{CED}: aa 64-183) and a C-terminal domain (P_{CTD}: aa 231-300) (Figure 1A). This modular organization was similar to that previously observed for RAV and VSV P (Figures 1B and 1C) and for other viruses of the order MNV (Karlin et al., 2003). This suggested that the predicted C-terminal domain of LNYV P corresponded to the N-RNA binding domain of the other rhabdovirus P proteins (Gérard et al., 2009; Mavrakis et al., 2004; Ribeiro et al., 2008).

The D-score profile of RAV and VSV also predicted the presence of a N-terminal and a central domains (Gérard et al., 2009) (Figures 1B and 1C). The central region predicted to be folded corresponded to the homo-dimerization domains (Ding et al., 2006; Ivanov et al., 2010), whereas the N-terminal region was shown to contain two transient helical elements in the isolated peptide of VSV (Leyrat et al., 2011a) and to constitute a molecular recognition element (MoRE) that is stabilized when it binds to its physiological partner, the RNA-free nucleoprotein (Leyrat et al., 2011b). For LNYV, a folded region is predicted in the central part of the molecule, which could correspond to a homo-oligomerization domain, but no folded region is predicted in the N-terminal region of P. However, a bipartite region with D-score values above 0.25 is found (aa 5-22 and aa 33-37) that may also correspond to a MoRE (Figure 1 A).

Expression and purification of LNYV P_{CTD}

On the basis of the D-score analysis, a synthetic cDNA corresponding to residues 230 to 300 of LNYV P was cloned and expressed in the Escherichia coli BL21 (DE3-RIL) strain as a

His-tag fusion protein. The protein was purified by nickel-affinity chromatography and size-exclusion chromatography.

In solution, the C-terminal domain of the protein (P_{CTD}) behaves as a monomer with a molecular mass of 9.2 ± 0.5 kDa in good agreement with the mass of 9,198 Da calculated from the amino acid sequence (Figure 2A). The Stokes' radius of 1.7 ± 0.1 nm is close to that expected for a globular protein of this molecular mass (1.6 nm) (Uversky, 1993). The far-UV circular dichroism spectrum exhibited two minima at 208 and 222 nm typical of a protein rich in α -helices (Figure 2B). The reversible thermal unfolding exhibited a sigmoid transition typical of a folded protein with a hydrophobic core (Figure 2C). These results clearly showed that the C-terminal region of LNYV P_{CTD} consisted of an autonomous structural unit that adopts a globular fold with large content of α -helical structure in the absence of the remaining part of the protein.

Crystal structure

A selenomethionine derivative of the LNYV P_{CTD} was crystallized at 20°C in 20 mM sodium cacodylate buffer, at pH 6.5, containing 200 mM magnesium acetate and 15% MPD. Diffraction data on frozen crystals were collected at Se K-edge on beam line ID14-4 (ESRF, Grenoble, France). The crystal diffracted to a 2.0 Å resolution and belonged to space-group $P4_122$ with one monomer in the asymmetric unit. The phases were solved by the SAD method using the ShelX-C-D-E pipeline (Sheldrick, 2010). A preliminary model was automatically generated in the electron density with the software Resolve (Terwilliger, 2003) and was refined by hand and finally with REFMAC5 (Vagin et al., 2004). The quality of the model was checked with PROCHECK (Laskowski et al., 1993) (Table 1).

The C-terminal domain of LNYV P comprises five α -helices, named α_1 to α_5 (Figure 3). The structure of P_{CTD} appeared as a lollipop with a globular region formed by helices α_1 to α_4 and the N-terminus of helix α_5 and a stalk formed by the C-terminal part of helix α_5 . The globular region had one flat face made by helices α_4 and α_5 and one rounded face made by the three remaining helices. Side-chains from residues in helices α_1 to α_4 and residues in the N-terminal moiety of helix α_5 constitute the hydrophobic core. A hydrophobic groove including a hole is located on the protein surface between the hairpin formed by helices α_1 and α_2 and helix α_5 (Figure 4A, hydrophobic groove 1) It comprises Trp17, Tyr25, Ile29, Phe32, Leu33 and Leu65. A second deep groove at the top of the domain is lined with Leu11, Leu15, Leu39 and

Leu46 (Figure 4A, hydrophobic groove 2). Analysis of the distribution of electrostatic potentials (Figure 4B) showed the presence of two patches of negative charges on opposite faces of the molecule. Asp36, Asp37, Glu57 and Glu60 (Negative patch 1) form a ridge of negatively charged residues on the flat face whereas Asp16, Glu18, Glu21, Glu49 and Glu52 (Negative patch 2) line the ridge along the hydrophobic cavity on the round face of P_{CTD}.

Comparison with other P proteins

The three-dimensional structure of P C-terminal domain has been solved for three rhabdoviruses (Assenberg et al., 2009; Mavrakis et al., 2004; Ribeiro et al., 2008), three paramyxoviruses (Blanchard et al., 2004; Johansson et al., 2003; Kingston et al., 2008) and for the corresponding VP35 protein of a filovirus (Leung et al., 2009; Leung et al., 2010). Despite the lack of sequence conservation between LNYV P and these other P proteins, we performed pairwise alignment of protein structures using the PDBeFold webserver (Krissinel & Henrick, 2004). This program uses the empirical Q function to align backbone C_α in a two-step process, first matching the secondary structure elements by comparing their positions and orientations in 3D and then refining the alignment of the C_α. The Q-score accounts for the alignment length and the root-mean square deviation (RMSD), and is thus less sensitive to local structural variations than the RMSD. The Q value increases from 0 to 1 with structural similarity.

Although, the Q and RMSD values obtained in these alignments suggested a low structural similarity, the alignment of the secondary structure elements is remarkable. PDBeFold aligned helices $\alpha 1$, $\alpha 2$, $\alpha 3$, $\alpha 4$ and $\alpha 5$ of LNYV P_{CTD} with helices $\alpha 1$, $\eta 1$ (3¹⁰ helix), $\alpha 2$, $\alpha 3$ and $\alpha 4$ of RAV and MOKV P_{CTD} (Figures 5A, 5E and S1), respectively, suggesting that the topological organization of the secondary structure elements is conserved. LNYV P_{CTD} lacks the β -hairpin localized between helix $\alpha 1$ and helix $\eta 1$ of both RAV and MOKV as well as helices $\alpha 5$ and $\alpha 6$. Previously, we have shown that the topology of RAV P_{CTD} is conserved in VSV P_{CTD} but the latter lacks helices $\alpha 4$ and $\alpha 6$ of the RAV protein (Ribeiro et al., 2008). In accordance, PDBeFold matched helices $\alpha 1$, $\alpha 2$, $\alpha 3$ and $\alpha 5$ of LNYV P_{CTD} with $\alpha 1$, $\eta 1$, $\alpha 3$ and $\alpha 4$ of VSV (Figures 5B and 5E). These alignments clearly suggested that P proteins from animal rhabdoviruses and from LNYV share a common topology and have evolved from a common ancestor by acquiring or losing secondary structure elements.

The C-terminal domain of paramyxovirus P proteins consists of an anti-parallel triple-helix bundle that is well-conserved within this family and is different, at first sight, from the structure of rhabdovirus P_{CTD}. PDBeFold aligned helices α_2 , α_3/α_4 and α_5 of LNYV P_{CTD} with helices α_1 , α_2 and α_3 of measles virus, mumps virus and Sendai virus P_{CTD} (Figures 5C, 5E and S1).

Finally, the LNYV P_{CTD} structure was aligned with the C-terminal domain of Ebola virus (EBOV) VP35 protein. The structure of the EBOV domain is composed of an α -helical subdomain and a β -sheet sub-domain. The four α -helices of the α -helical subdomain of EBOV P_{CTD} were aligned with helices α_2 , α_3/α_4 and α_5 of LNYV P_{CTD}, and superimposed on the three helices of the paramyxovirus proteins, revealing a similarity between the proteins of paramyxoviruses and filoviruses (Figures 5D and 5E).

A search of the Protein Data Bank (PDB) using the DALI server (Holm & Rosenstrom, 2010) yielded numerous hits with Z values lower than 5.0 that appeared to be unrelated to our viral protein domain and failed to recognize the C-terminal domains of (-)RNA phosphoproteins. A search with the PDBeFold server (Krissinel & Henrick, 2004) also failed to recognize the phosphoproteins of (-)RNA viruses, but identified the auto-inhibited GTPase-binding domain of the Wiskott-Aldrich syndrome protein (WASP) as a putative structural homolog with a Z-score of 2.0. The overall topology of this protein is very similar to that of LNYV P_{CTD} and the structures can be aligned on the basis of their α -helices (Figure S3). The functional relationship between the virus P and this protein is unclear. It is possible that the structural similarity results from a process of convergent evolution for packing α -helices.

Modeling the structure of LNYV nucleoprotein

The nucleoprotein of LNYV showed no significant sequence similarity with the nucleoproteins of other NNV viruses. The structure of the nucleoprotein of LNYV was predicted by submitting its amino acid sequence to the I-TASSER server (Zhang, 2008). I-TASSER uses a multi-step protocol that includes meta-threading prediction, extraction and assembly of fragments and structure selection and refinement (Roy et al., 2010). An estimation of the quality of the predicted model is provided by the server in the form of a C-score. Typical values of this parameter ranges from -5 to +2, and higher values indicates a higher confidence. The best model predicted by the server had a C-score of -3.09, which is in the lower range of the confidence

interval, and the 10 most used protein structures for the construction of this model were unrelated to viral nucleoproteins. Nevertheless, a search of the PDB showed that the model exhibited a close structural relationship with the nucleoprotein of VSV and RAV (Figure 6A). Structural alignments of the LNYV N model with VSV and RAV N yielded RMSD values of 2.39 Å and 3.22 Å, respectively (Figure 6B).

The LNYV N model consists of two domains, an N-terminal (N_{NTD}) and a C-terminal domain (N_{CTD}). As in the other rhabdovirus N proteins, a deep groove extending at the interface between these two domains is lined with basic residues and likely corresponds to the RNA binding site (Figure 6C). The model also contains two subdomains extending on opposite side of the molecules that in the polymeric N-RNA complex of VSV and RAV stabilize the assembly by binding to the neighboring N molecules (Albertini et al., 2006; Green et al., 2006). In both RAV and VSV polymeric N-RNA complex, the N_{CT} -loop of two adjacent N subunits participates in binding the C-terminal domain of P (Green & Luo, 2009; Ribeiro et al., 2009). LNYV N is longer than VSV and RAV N and contains several insertions in the N_{NTD} domain (Figure 6 D).

DISCUSSION

Structure of LNYV P

The prediction of disorder from the amino acid sequence suggested that LNYV P has a structural organization similar to that of other rhabdoviruses (Gérard et al., 2009), with a long N-terminal disordered region, a central folded domain, a short disordered region and a C-terminal folded domain. Despite low sequence conservation, the modular organization with alternating disordered and folded regions appears conserved between animal and plant rhabdoviruses.

The isolated C-terminal domain localized by this prediction behaved as an autonomous folding unit like those of RAV, MOKV and VSV. The structural comparison clearly revealed the relationship between these structures. The largest domains are those of RAV and MOKV, which both belongs to the same genus Lyssavirus, and share the same topology. P_{CTD} of LNYV and VSV are shorter and lack different secondary structure elements of the RAV and MOKV proteins. VSV lacks helices $\alpha 4$ and $\alpha 6$, whereas LNYV lacks the β -hairpin and helix $\alpha 6$. The

conserved secondary structure elements adopt similar 3D positions and orientations than in RAV and MOKV and appear superimposed when the domains are aligned pairwise. This suggests that animal lyssaviruses occupy an intermediate position between the animal vesiculoviruses and the plant cytorhabdoviruses in the phylogenetic tree, in accordance with phylogenetic analysis based on the amino-acid sequence of the L protein.

The function of the C-terminal domain is conserved between the rhabdoviruses and the paramyxoviruses, but the P_{CTD} of the paramyxoviruses is shorter and comprises only three α -helices. Here, we showed that these three α -helices can be aligned in 3D onto α -helices of the rhabdoviruses P_{CTD} and also onto α -helices of the corresponding domain of a filovirus. Despite the low structural similarity between these proteins, the location of these structural elements in the C-terminal part of P protein and the similar position and orientation in 3D unlikely result from a convergent evolution and argue for a common ancestry of these proteins. Indeed, given the large structural variations already observed between rhabdovirus P_{CTD}, it is not surprising to see even larger variations between members of different virus families. The comparison with EBOV VP35 showed that the P protein of filoviruses is also structurally related to the P protein of paramyxoviruses, in accordance with other phylogenetic analyses based on the amino acid sequence of L.

Finally, the structural resemblance of the LNYV N model with the structures of RAV and VSV N provides an additional argument for classifying this plant virus within the Rhabdoviridae.

Structural relationship between NNV phosphoproteins

Among the proteins commonly encoded by all NNV, the L protein that carries all enzymatic activities required for mRNA synthesis and genome replication is the most conserved protein. (Poch)

This of course indicates an evolutionary relationship between all these viruses but provides no evidence that all these proteins have followed a similar evolution. In particular, the sequence of the phosphoprotein is highly variable and its relationship with that of other viruses rarely extends beyond the limit of the viral genus. Thus, between VSV and RAV P, sequence identity is only about 10%.

A structural relationship between the C-terminal domain of Paramyxoviridae and Rhabdoviridae was already proposed. However, the helices of RAV and MOK that were matched to those of MEV were different than in this study.

Materials and Methods

Amino acid sequence analysis, cloning, production, and purification of the C-terminal domain of LNYV P

The location of disordered regions within LNYV P was predicted by submitting its amino acid sequence to 16 different algorithms accessible through web servers and by calculating a consensus prediction as described previously (Gérard et al., 2009; Ribeiro et al., 2008). On the basis of this prediction, a synthetic gene (GeneArt) optimized for expression in *E. coli* and encoding residues 230-300 of LNYV P was cloned between the NcoI and BamHI restriction sites of a pET28 vector, such that the construct includes a C-terminal His₆-tag and a Leu-Glu linker. The construction was verified by DNA sequencing. The protein was expressed in *E. coli* BL21-RIL. Cells were grown in LB medium at 37°C until O.D._{600nm} reached a value 0.6. Then, protein expression was induced by adding 1mM IPTG and cells were grown for 3h. The cells were harvested and suspended in 20 mM Tris-HCl buffer at pH 7.5 containing 150 mM NaCl and EDTA-free protease inhibitor cocktail (Roche) (Buffer A). The cells were broken by sonication and the soluble fraction was loaded onto a Ni²⁺ column equilibrated in buffer A. The column was washed with buffer A containing 20mM imidazole and the protein was then eluted with buffer A containing 300 mM imidazole. The protein was further purified by gel filtration with a Superdex S75 (GE Healthcare) equilibrated in buffer A. The protein preparations were checked by SDS-PAGE.

Size exclusion chromatography (SEC) combined with detection by multi-angle laser light scattering (MALLS) and refractometry

SEC was performed with a Superdex S75 column (GE Healthcare) equilibrated in 20 mM Tris-HCl, pH 7.5, 150 mM NaCl. Separations were performed at 20 °C with a flow rate of 0.5 ml.min⁻¹. 50 µL of a protein solution at a concentration of 7 mg.mL⁻¹ were injected. On-line multi-angle laser light scattering (MALLS) detection was performed with a DAWN-EOS detector (Wyatt Technology Corp., Santa Barbara, CA) using a laser emitting at 690 nm. Protein concentration was measured on-line by refractive index measurements using a RI2000 detector (Schambeck SFD) and a refractive index increment $dn/dc = 0.185 \text{ mL.g}^{-1}$. Data were analyzed

and weight-averaged molecular masses (M_w) were calculated using the software ASTRA V (Wyatt Technology Corp., Santa Barbara, CA) as described previously (Gérard et al., 2007). For size determination, the column was calibrated with proteins of known Stokes' radius (R_S) (Uversky, 1993).

Crystallization, data collection and structure determination and refinement.

Crystallization conditions were screened by the hanging drop vapor diffusion method using a PixSys4200 Cartesian robot (high-throughput crystallization laboratory at EMBL Grenoble, France). The screen was performed by combining 0.1 μ l of protein solution at 8 mg.mL⁻¹ in buffer A with 0.1 μ l of Hampton Crystal Screen solutions.

The selenomethionine derivative of LNYV P_{CTD} was crystallized in a solution 20 mM sodium cacodylate buffer at pH 6.5 containing 200 mM magnesium acetate and 15% MPD. Crystals were frozen in the same solution at a final concentration of 30% MPD as cryoprotectant.

Diffraction data were collected at Se K-edge on the ID14-4 beam line (ESRF Grenoble, France). The crystals diffracted to 2.0 Å resolution and belonged to space-group P4122. Data was integrated and scaled using XDS [7]. Two selenium sites and initial phases were found by SAD method using the ShelX-C-D-E pipeline [11] and the software Resolve [13] was used to automatically trace a model in the resulting electron density. The model was improved by hand and finally refined with REFMAC5 [9].

Hits were reproduced manually. The N_{Δ21}⁰-P₆₀ complex crystallized at 20 °C in 0.1 M sodium acetate buffer, pH 4.6, containing 4 % (w/v) of PEG4000. Single crystals were harvested from the drop, briefly soaked in the reservoir solution supplemented with 25 % glycerol and flash frozen in liquid nitrogen at 100 K before data collection. X-ray diffraction data were collected at a wavelength of 0.933 Å on the ID14-2 beamline at the ESRF (Grenoble, France).

The data were processed using the program iMosflm (Powell, 1999) and scaled with the program Scala from the ccp4 suite (Potterton et al., 2003). The structure was solved by molecular replacement with the program Phaser (McCoy et al., 2007) using residues 22 to 422 of a protomer of N (2GIC, chain E) extracted from the N-RNA crystal structure (Green et al., 2006) as a search model. The visible part of P₆₀ was assigned and constructed with the program Buccaneer (Cowtan, 2006), and the overall structure was refined to 3.0 Å resolution using Coot (Emsley et al., 2010) and Refmac5 (Murshudov et al., 1997). The quality of the model was

checked with PROCHECK (Laskowski et al., 1993). Data collection and refinement statistics are summarized in Table S2.

LNYV P_{CTD} was crystallized in space group P2₁2₁2. The crystallographic asymmetric unit contained five protomers of N_{Δ21}⁰-P₆₀ complex. Each protomer includes residues 22 to 422 of N and residues 6 to 33 of P₆₀. In some protomers, residues 34 and 35 could also be constructed.

Structure analysis and comparisons

Sequence alignments were made using the CLUSTALW2 server,⁶⁷ and structure alignments were made with the SSM program.²⁷ Electrostatic maps were calculated using the Delphi program.⁶⁸ Conservation score was calculated with the AL2CO program.⁶⁹ Figures of the structure were prepared with PyMOL⁷⁰ and Chimera.⁷¹

PDB accession number

Coordinates have been deposited in the Research Collaboratory for Structural Bioinformatics PDB with the accession number [2k47](#).

Acknowledgments

This work was supported by a grant from the French ANR (ANR-07-001-01 (ANRAGE)). E.A.R. was supported by postdoctoral fellowships from the ANR. N.M. was supported by fellowship from the Région Rhone-Alpes. We thank the Partnership for Structural Biology for the excellent structural biology environment.

References

- Albertini, A. A., Wernimont, A. K., Muziol, T., Ravelli, R. B., Clapier, C. R., Schoehn, G., Weissenhorn, W. & Ruigrok, R. W. (2006). Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* **313**, 360-363.
- Arnheiter, H., Davis, N. L., Wertz, G., Schubert, M. & Lazzarini, R. A. (1985). Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. *Cell* **41**, 259-267.
- Assenberg, R., Delmas, O., Ren, J., Vidalain, P. O., Verma, A., Larrous, F., Graham, S. C., Tangy, F., Grimes, J. M. & Bourhy, H. (2009). The Structure of the N-RNA Binding Domain of the Mokola virus Phosphoprotein. *J Virol* **84**, 1089-1096.
- Blanchard, L., Tarbouriech, N., Blackledge, M., Timmins, P., Burmeister, W. P., Ruigrok, R. W. & Marion, D. (2004). Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. *Virology* **319**, 201-211.
- Callaghan, B. & Dietzgen, R. G. (2005). Nucleocapsid gene variability reveals two subgroups of Lettuce necrotic yellows virus. *Arch Virol* **150**, 1661-1667.
- Chambers, T. C., Crowley, N. C. & Francki, R. I. (1965). Localization of lettuce necrotic yellows virus in host leaf tissue. *Virology* **27**, 320-328.
- Chambers, T. C. & Francki, R. I. (1966). Localization and recovery of lettuce necrotic yellows virus from xylem tissues of *Nicotiana glutinosa*. *Virology* **29**, 673-676.
- Cowtan, K. D. (2006). The Buccaneer software for automated model building. *Acta Crystallographica Section D-Biological Crystallography* **62**, 1002-1011.
- Delmas, O., Assenberg, R., Grimes, J. M. & Bourhy, H. (2010). The structure of the nucleoprotein binding domain of lyssavirus phosphoprotein reveals a structural relationship between the N-RNA binding domains of Rhabdoviridae and Paramyxoviridae. *RNA Biol* **7**, 322-327.
- Dietzgen, R. G., Callaghan, B., Wetzell, T. & Dale, J. L. (2006). Completion of the genome sequence of Lettuce necrotic yellows virus, type species of the genus *Cytorhabdovirus*. *Virus Res* **118**, 16-22.
- Dietzgen, R. G. & Francki, R. I. (1988). Analysis of lettuce necrotic yellows virus structural proteins with monoclonal antibodies and concanavalin A. *Virology* **166**, 486-494.
- Ding, H., Green, T. J., Lu, S. & Luo, M. (2006). Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. *J Virol* **80**, 2808-2814.
- Emerson, S. U. & Yu, Y. (1975). Both NS and L proteins are required for in vitro RNA synthesis by vesicular stomatitis virus. *J Virol* **15**, 1348-1356.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501.
- Gérard, F. C. A., Ribeiro, E., Albertini, A., Zaccari, G., Ebel, C., Ruigrok, R. & Jamin, M. (2007). Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution. *Biochemistry* **46**, 10328-10338.
- Gérard, F. C. A., Ribeiro, E. A., Leyrat, C., Ivanov, I., Blondel, D., Longhi, S., Ruigrok, R. W. H. & Jamin, M. (2009). Modular organization of rabies virus phosphoprotein. *J Mol Biol* **388**, 978-996.
- Green, T. J. & Luo, M. (2009). Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P. *Proc Natl Acad Sci U S A* **106**, 11721-11726.

- Green, T. J., Zhang, X., Wertz, G. W. & Luo, M. (2006).** Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* **313**, 357-360.
- Hock, M., Kraus, I., Schoehn, G., Jamin, M., Andrei-Selmer, C., Garten, W. & Weissenhorn, W. (2009).** RNA induced polymerization of the Borna disease virus nucleoprotein. *Virology* **397**, 64-72.
- Holm, L. & Rosenstrom, P. (2010).** Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545-W549.
- Ivanov, I., Crepin, T., Jamin, M. & Ruigrok, R. (2010).** Structure of the dimerisation domain of the rabies virus phosphoprotein. *J Virol* **84**, 3707-3710.
- Ivanov, I., Yabukarski, F., Ruigrok, R. W. H. & Jamin, M. (2011).** Structural insights into the rhabdovirus transcription/replication complex. *Virus Res*.
- Jackson, A. O., Dietzgen, R. G., Goodin, M. M., Bragg, J. N. & Deng, M. (2005).** Biology of plant rhabdoviruses. *Annu Rev Phytopathol* **43**, 623-660.
- Johansson, K., Bourhis, J. M., Campanacci, V., Cambillau, C., Canard, B. & Longhi, S. (2003).** Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein. *J Biol Chem* **278**, 44567-44573.
- John, T. J. (2004).** Chandipura virus, encephalitis, and epidemic brain attack in India. *Lancet* **364**, 2175; author reply 2175-2176.
- Karlin, D., Ferron, F., Canard, B. & Longhi, S. (2003).** Structural disorder and modular organization in Paramyxovirinae N and P. *J Gen Virol* **84**, 3239-3252.
- Kingston, R. L., Gay, L. S., Baase, W. S. & Matthews, B. W. (2008).** Structure of the nucleocapsid-binding domain from the mumps virus polymerase; an example of protein folding induced by crystallization. *J Mol Biol* **379**, 719-731.
- Kormelink, R., Garcia, M. L., Goodin, M., Sasaya, T. & Haenni, A. L. (2011).** Negative-strand RNA viruses: The plant-infecting counterparts. *Virus Res*.
- Krissinel, E. & Henrick, K. (2004).** Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**, 2256-2268.
- Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993).** PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* **26**, 283-291.
- Leung, D. W., Ginder, N. D., Fulton, D. B., Nix, J., Basler, C. F., Honzatko, R. B. & Amarasinghe, G. K. (2009).** Structure of the Ebola VP35 interferon inhibitory domain. *Proc Natl Acad Sci U S A* **106**, 411-416.
- Leung, D. W., Shabman, R. S., Farahbakhsh, M., Prins, K. C., Borek, D. M., Wang, T., Muhlberger, E., Basler, C. F. & Amarasinghe, G. K. (2010).** Structural and functional characterization of Reston Ebola virus VP35 interferon inhibitory domain. *J Mol Biol* **399**, 347-357.
- Leyrat, C., Gerard, F. C., de Almeida Ribeiro, E., Jr., Ivanov, I., Ruigrok, R. W. & Jamin, M. (2010).** Structural disorder in proteins of the rhabdoviridae replication complex. *Protein Pept Lett* **17**, 979-987.
- Leyrat, C., Jensen, M. R., Ribeiro, E. A., Gérard, F., Ruigrok, R., Blackledge, M. & Jamin, M. (2011a).** The N⁰-binding region of the vesicular stomatitis virus phosphoprotein is globally disordered but contains transient α -helices. *Prot Sci* **20**, 542-556.

- Leyrat, C., Yabukarski, F., Tarbouriech, N., Ribeiro, E. A., Jensen, M. R., Blackledge, M., Ruigrok, R. & Jamin, M. (2011b).** The structure of the vesicular stomatitis virus N⁰-P complex PLoS Path In press.
- Llorente, M. T., Garcia-Barreno, B., Calero, M., Camafeita, E., Lopez, J. A., Longhi, S., Ferron, F., Varela, P. F. & Melero, J. A. (2006).** Structural analysis of the human respiratory syncytial virus phosphoprotein: characterization of an alpha-helical domain involved in oligomerization. *J Gen Virol* **87**, 159-169.
- Masters, P. S. & Banerjee, A. K. (1988).** Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA. *J Virol* **62**, 2658-2664.
- Mavrakis, M., McCarthy, A. A., Roche, S., Blondel, D. & Ruigrok, R. W. (2004).** Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. *J Mol Biol* **343**, 819-831.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007).** Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997).** Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography* **53**, 240-255.
- Peluso, R. W. & Moyer, S. A. (1988).** Viral proteins required for the in vitro replication of vesicular stomatitis virus defective interfering particle genome RNA. *Virology* **162**, 369-376.
- Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003).** A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr* **59**, 1131-1137.
- Powell, H. R. (1999).** The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr D Biol Crystallogr* **55**, 1690-1695.
- Ribeiro, E. A., Jr., Favier, A., Gerard, F. C., Leyrat, C., Brutscher, B., Blondel, D., Ruigrok, R. W., Blackledge, M. & Jamin, M. (2008).** Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein. *J Mol Biol* **382**, 525-538.
- Ribeiro, E. A., Leyrat, C., Gérard, F. C., Albertini, A. A., Falk, C., Ruigrok, R. W. & Jamin, M. (2009).** Binding of rabies virus polymerase cofactor to recombinant circular nucleoprotein-RNA complexes. *J Mol Biol* **394**, 558-575.
- Roy, A., Kucukural, A. & Zhang, Y. (2010).** I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **5**, 725-738.
- Sheldrick, G. M. (2010).** Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **66**, 479-485.
- Terwilliger, T. C. (2003).** SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol* **374**, 22-37.
- Tordo, N., Benmansour, A., Calisher, C., Dietzgen, R. G., Fang, R.-X., Jackson, A. O., Kurath, G., Nadin-Davis, S., Tesh, R. B. & Walker, P. J. (2004).** Rhabdoviridae. In *Virus taxonomy, VIIIth Report of the ICTV*, pp. 623-644. Edited by M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger & L. A. Ball. London: Elsevier/Academic Press.
- Uversky, V. N. (1993).** Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry* **32**, 13288-13298.

- Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. & Murshudov, G. N. (2004).** REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* **60**, 2184-2195.
- Wertz, G. W. (1978).** Isolation of possible replicative intermediate structures from vesicular stomatitis virus-infected cells. *Virology* **85**, 271-285.
- Wetzel, T., Dietzgen, R. G. & Dale, J. L. (1994a).** Genomic organization of lettuce necrotic yellows rhabdovirus. *Virology* **200**, 401-412.
- Wetzel, T., Dietzgen, R. G., Geering, A. D. & Dale, J. L. (1994b).** Analysis of the nucleocapsid gene of lettuce necrotic yellows rhabdovirus. *Virology* **202**, 1054-1057.
- Wolanski, B. S. & Chambers, T. C. (1971).** The multiplication of lettuce necrotic yellows virus. *Virology* **44**, 582-591.
- Zhang, Y. (2008).** I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40.

Table 1. Data collection and refinement statistics (molecular replacement)

Crystal 1	
Data collection	
Space group	P4 ₁ 22
Cell dimensions	
a, b, c (Å)	43.45, 43.45, 89.71
α , β , γ (°)	90.0, 90.0, 90.0
Resolution (Å) ^a	43.28-2.00 (2.16-2.00)
R _{sym} ^a	4.3 % (26.6%)
I / σ I ^a	16.05(4.03)
Completeness (%) ^a	98.9 (100)
Redundancy	3.76 (3.80)
Refinement	
Resolution (Å)	43.28-2.00 (2.16-2.00)
No. reflections ^a	47380 (7366)
R _{work} / R _{free}	22.6% / 29.0%
No. atoms	
Protein	610
Water	22
R.m.s. deviations	
Bond lengths (Å)	0.031
Bond angles (°)	2.095
Ramachandran favored	90.8 %
Ramachandran allowed	98.2 %

^aValues in parentheses are for highest-resolution shell.

Legends for the figures

Figure 1. Predicting boundaries between structured and disordered regions. The D-score as a function of residue number is calculated as a consensus of different predictions as described in material and methods. Regions of the protein with contiguous D-score value above 0.5 are predicted to be structured, whereas regions with D-score value below 0.5 are predicted to be disordered. The data for RAV and VSV are taken from (Gérard et al., 2009) and are shown here for comparison. **A. Lettuce Necrotic Yellowing virus (LNYV) P.** The red box above the plot shows the region of the protein that has been selected, cloned and characterized in this work. **B. Rabies virus (RAV) P.** The orange and red boxes show the localization of the central (P_{CED}) and C-terminal (P_{CTD}) domains for which the three-dimensional structure has been experimentally determined. The PDB code is given above each box. The dotted box indicates the position of the N-terminal region involved in binding to the RNA-free nucleoprotein (Mavrakis et al., 2006). **C. Vesicular stomatitis virus (VSV) P.** The orange and red boxes show the localization of the P_{CED} and P_{CTD} for which the three-dimensional structure has been experimentally determined. The dotted box indicates the position of the N-terminal region MoRE that has been shown to fold upon binding to the RNA-free nucleoprotein (Leyrat et al., 2011).

Figure 2. Solution properties of LNYV C-terminal domain. The experiments were performed in a 20 mM Tris/HCl buffer at pH 7.5 containing 150 mM NaCl at 20°C. **A. Molecular mass and size determined by size exclusion chromatography combined with detection by multi-angle laser light scattering and refractometry (SEC-MALLS-RI).** 50 μ l of a protein sample at 5 mg.mL^{-1} . The line shows the SEC elution profile as monitored by refractometry. The crosses show the molecular mass calculated from light scattering and refractometry data. **B. Secondary structure by far-UV circular dichroism.** The spectrum was measured with a protein sample at a concentration of mg.mL^{-1} in a 1 mm cuvette. **C. Thermal unfolding monitored by far-UV circular dichroism at 222 nm.** The experiment was performed with a protein sample at a concentration of mg.mL^{-1} in a 10 mm cuvette under continuous stirring. The temperature was raised at a speed of 4 $^{\circ}\text{C.min}^{-1}$.

Figure 3. Crystal structure and topology of LNYV P_{CTD} . **A. Schematic representation of LNYV P_{CTD} .** Helices are numbered from $\alpha 1$ to $\alpha 4$. The numbers indicate the positions of the N- and C-terminal residues. The figure was drawn with PYMOL. **B. Topology of LNYV P_{CTD} .**

Figure 4. Mapping of hydrophobic residues and electrostatic surface potential on the surface of LNYV P_{CTD} . **A. Hydrophobic surface map.** The figure shows the distribution of hydrophobic residues on the surface; green indicates hydrophobic residues and yellow hydrophilic residues. **B. Electrostatic surface potential.** The surface potential was calculated with the Delphi program and is colour-coded on the surface from red (negatively charged residues; $-12 \text{ k}_B\text{T}$) to blue (positively charged residues; $+12 \text{ k}_B\text{T}$).

Figure 5. Structural comparison with the C-terminal domain of other MNV viruses. The structure of LNYV was aligned pairwise with the structure of the C-terminal of different rhabdoviruses (RAV, VSV, MOKV), paramyxoviruses (MEV, MUV, SeV) and filoviruses (EBOV). The superposition of each of these protein domains (in blue) with the structure of LNYV P_{CTD} (in red) and the topology of the protein domain are shown. **A. RAV. B. VSV. C. MeV. D. EBOV. E. Amino acid sequence alignment based on the structural alignments.**

Figure 6. Homology model of the LNYV N protein and comparison with the RAV and VSV N proteins. **A. Schematic representation of the LNYV N model generated by the I-TASSER**

server. **B. Superposition of the LNYV N model with the structure of RAV and VSV N. C. Electrostatic surface potential.** The surface potential was calculated with the Delphi program and is colour-coded on the surface from red (negatively charged residues; $-12 \text{ k}_B\text{T}$) to blue (positively charged residues; $+12 \text{ k}_B\text{T}$). **D. Sequence alignment.**

Figure S1. Amino-acid sequence alignment. The alignment was performed with T-coffee and colored according to the ClustalX scheme.

Figure S2. Structural comparison with the C-terminal domain of other MNV viruses.

Figure S3. Structural comparison with the auto-inhibited GTPase-binding domain of the Wiskott-Aldrich syndrome protein (WASP).

Gérard, F. C. A., Ribeiro, E. A., Leyrat, C., Ivanov, I., Blondel, D., Longhi, S., Ruigrok, R. W. H. & Jamin, M. (2009). Modular organization of rabies virus phosphoprotein. *J Mol Biol* **388**, 978-996.

Leyrat, C., Yabukarski, F., Tarbouriech, N., Ribeiro, E. A., Jensen, M. R., Blackledge, M., Ruigrok, R. & Jamin, M. (2011). Structure of the vesicular stomatitis virus N⁰-P complex. *PLoS Path* **7**, e1002248.

Mavrakis, M., Mehoulas, S., Real, E., Iseni, F., Blondel, D., Tordo, N. & Ruigrok, R. W. (2006). Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA. *Virology* **349**, 422-429.

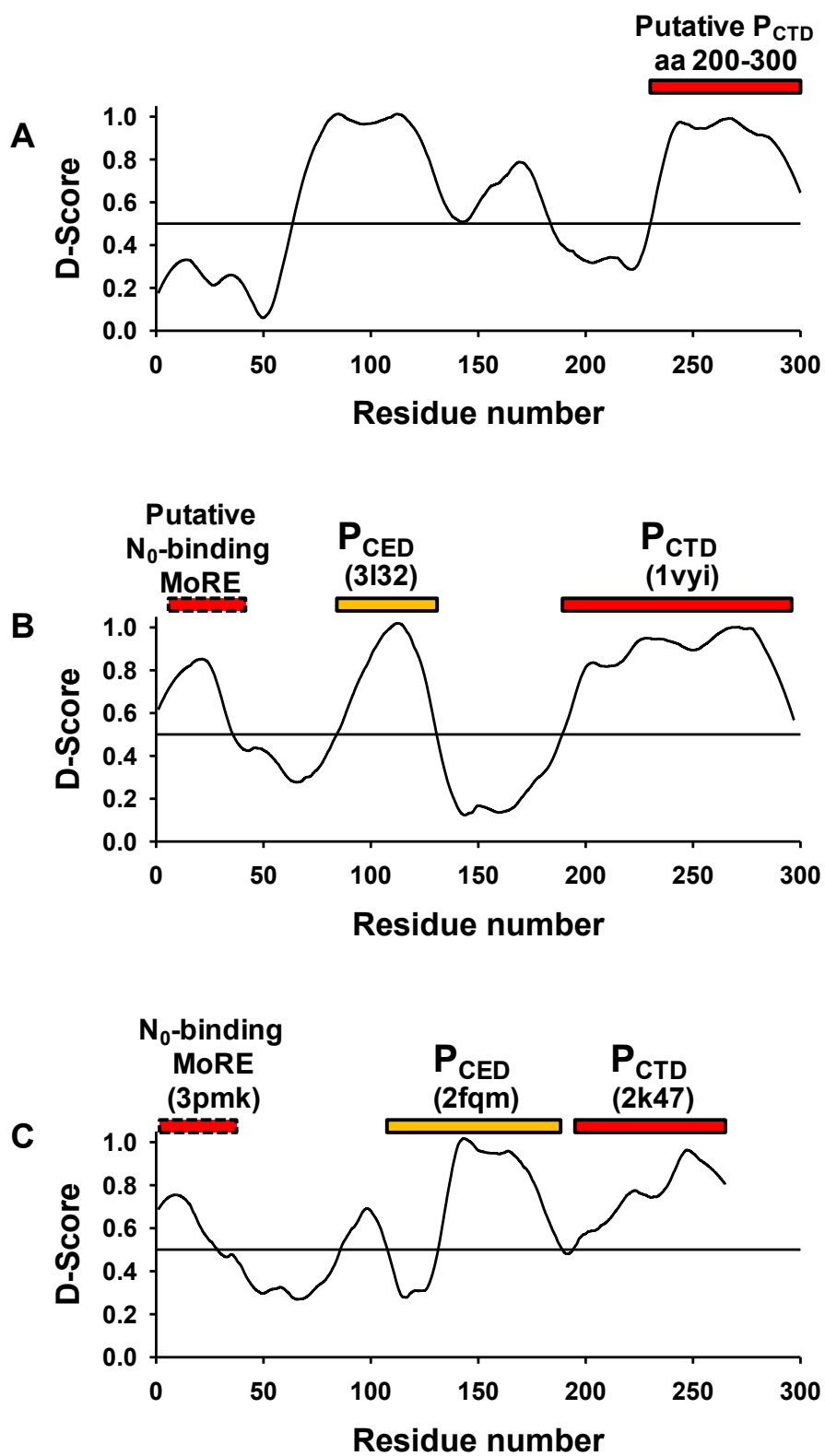


Figure 1.

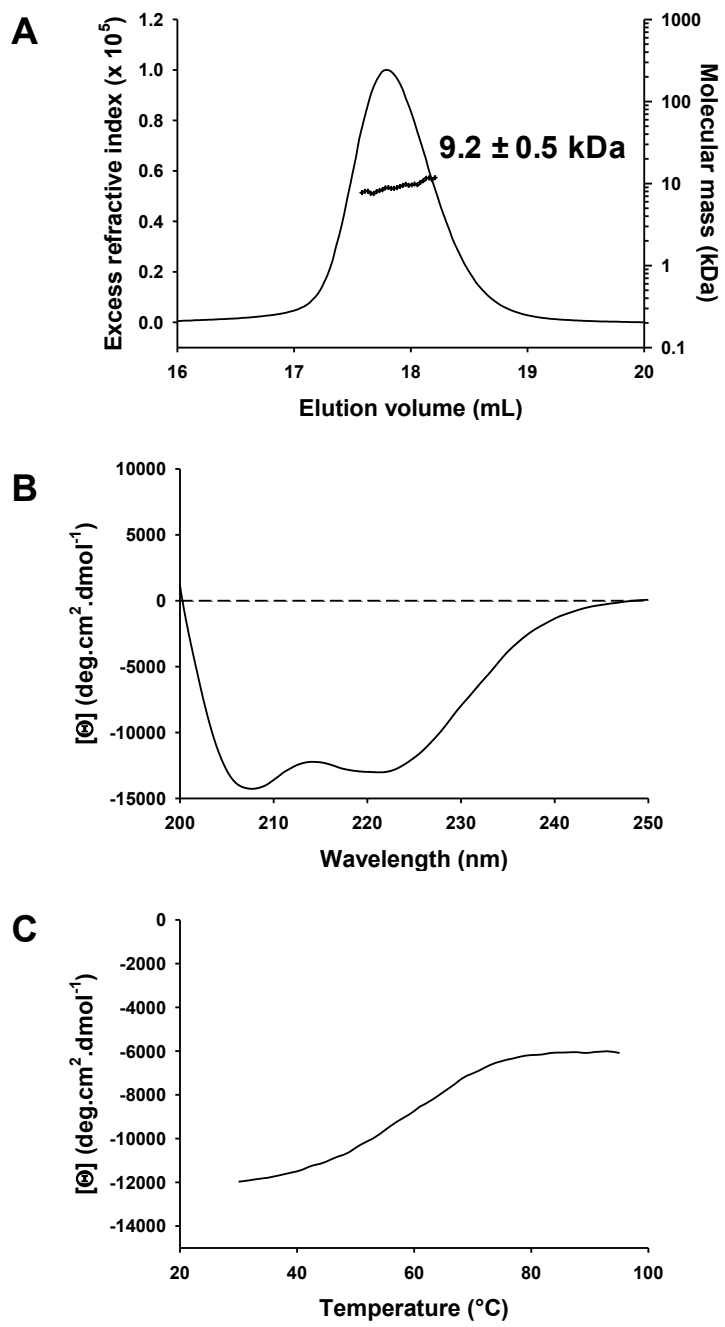


Figure 2.

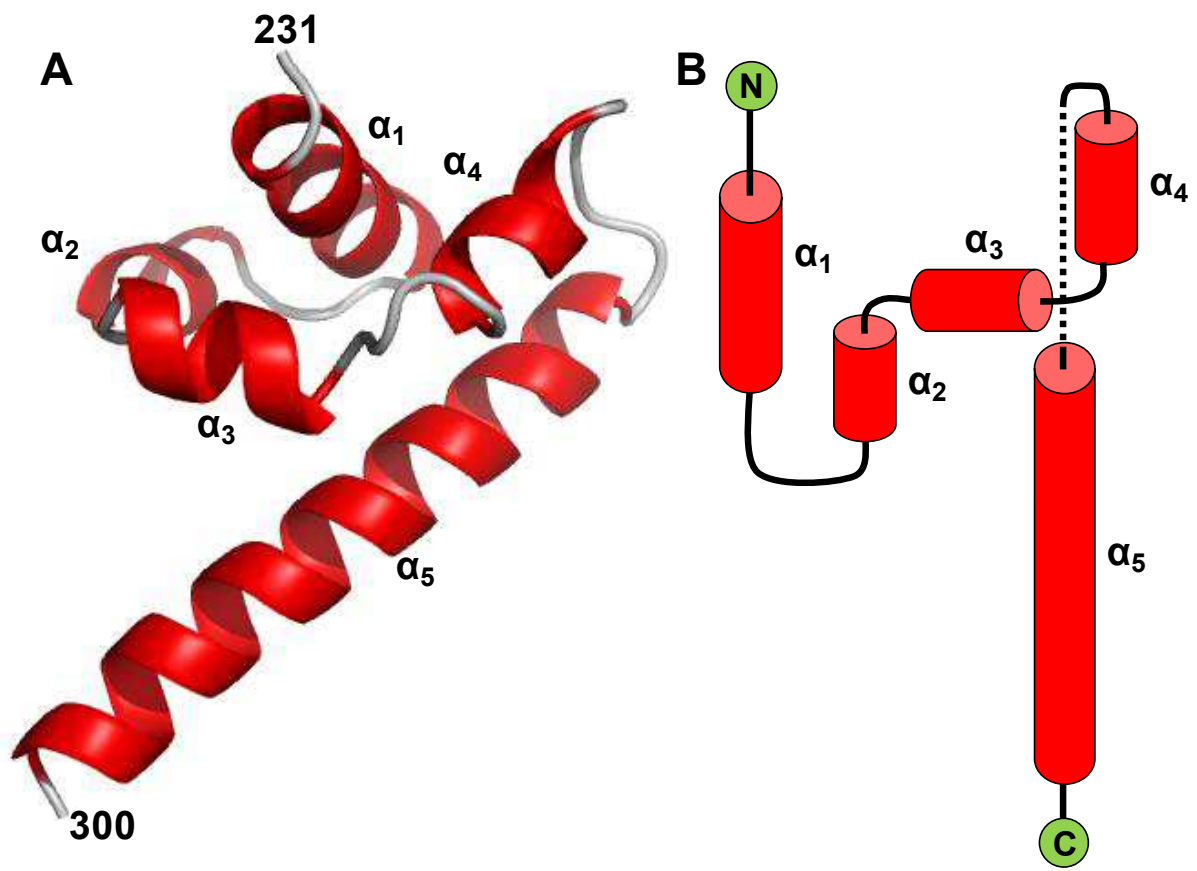


Figure 3.

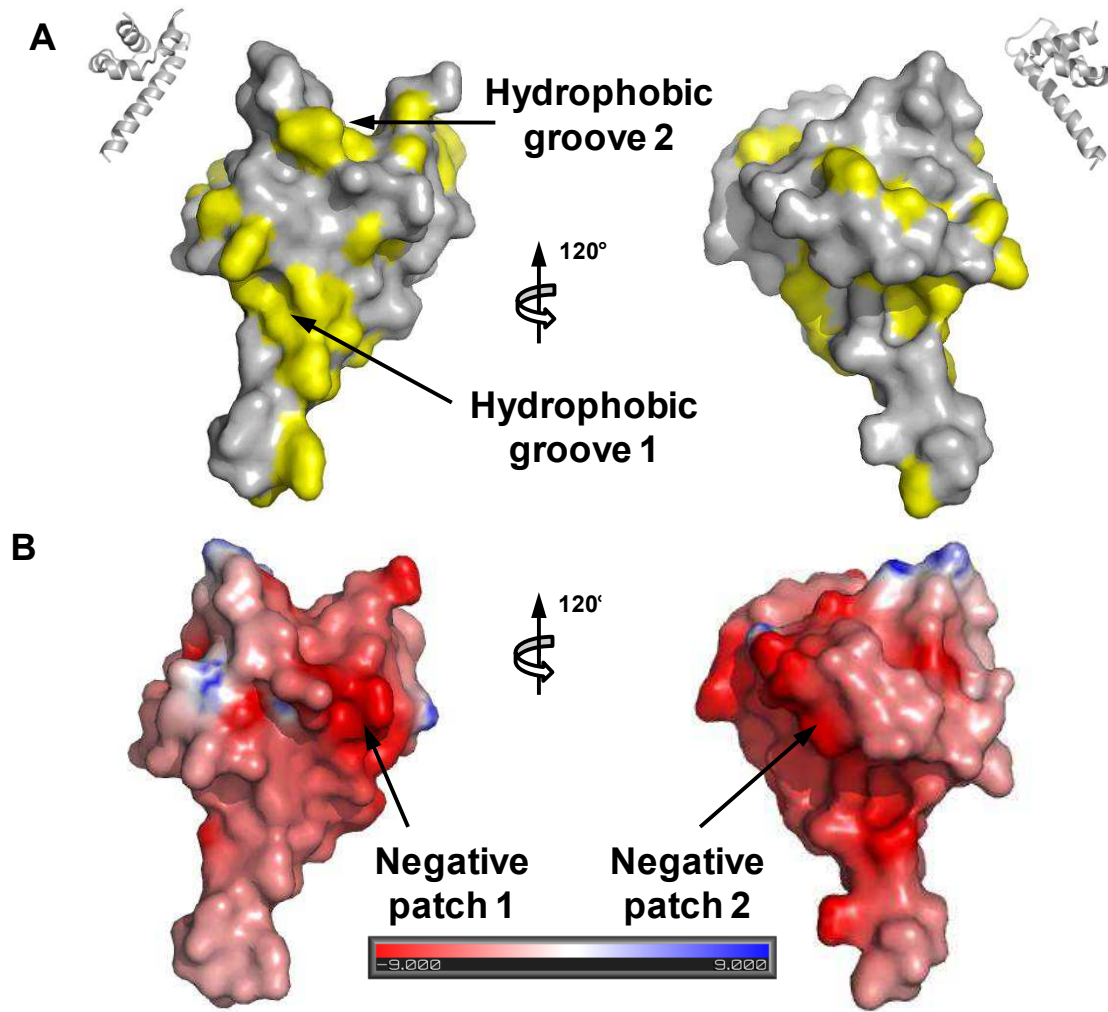


Figure 4.

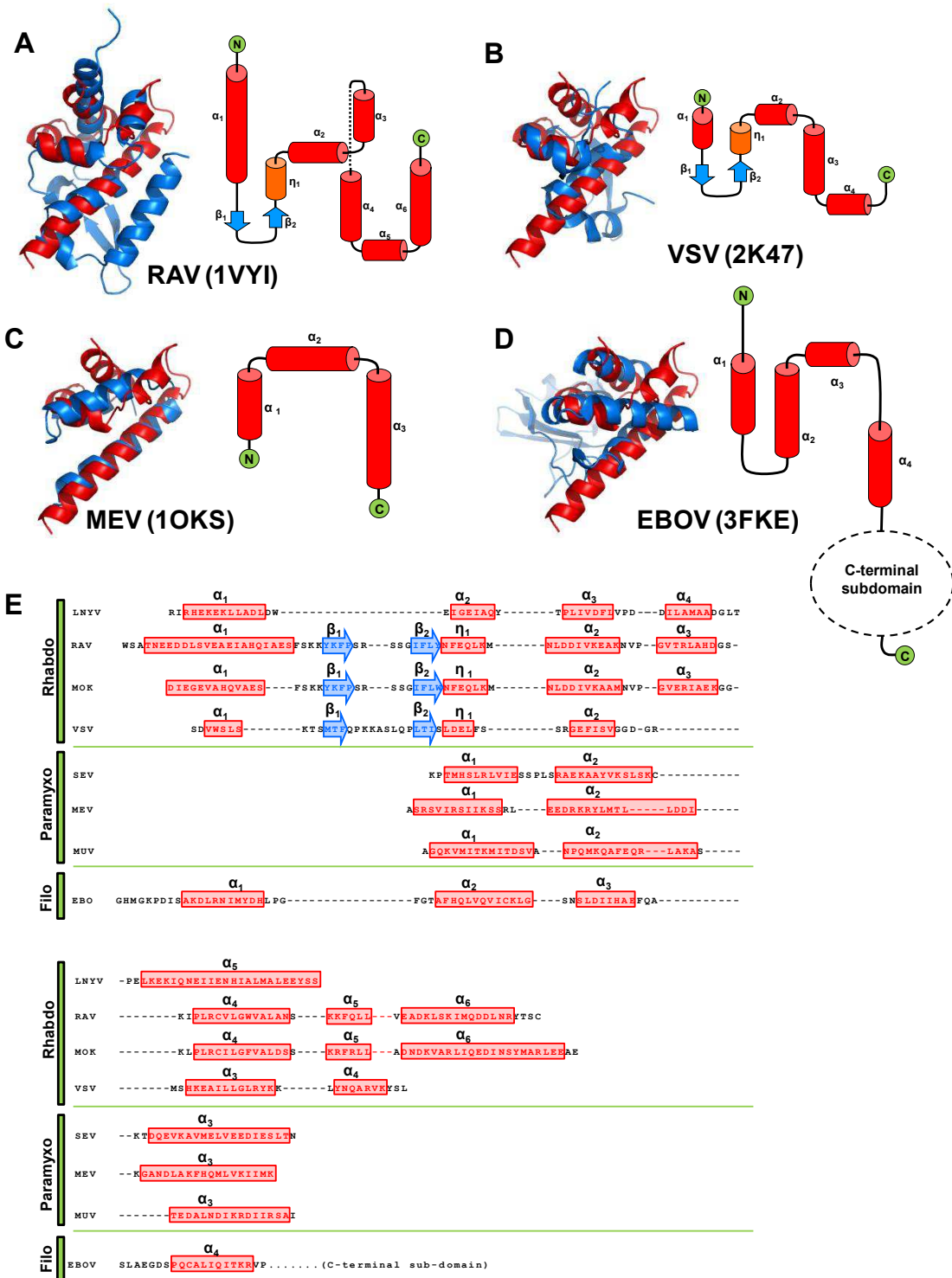


Figure 5.

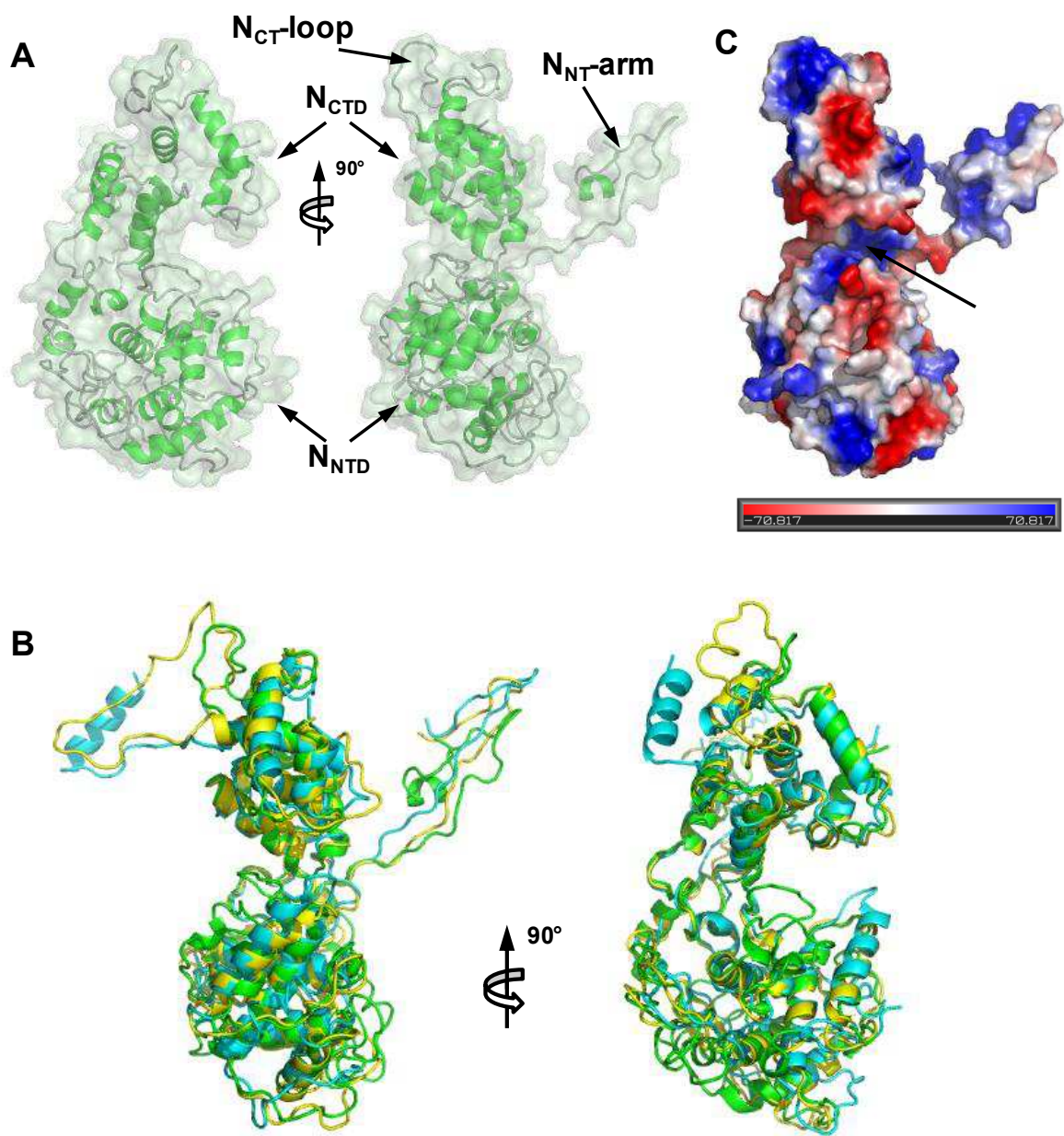
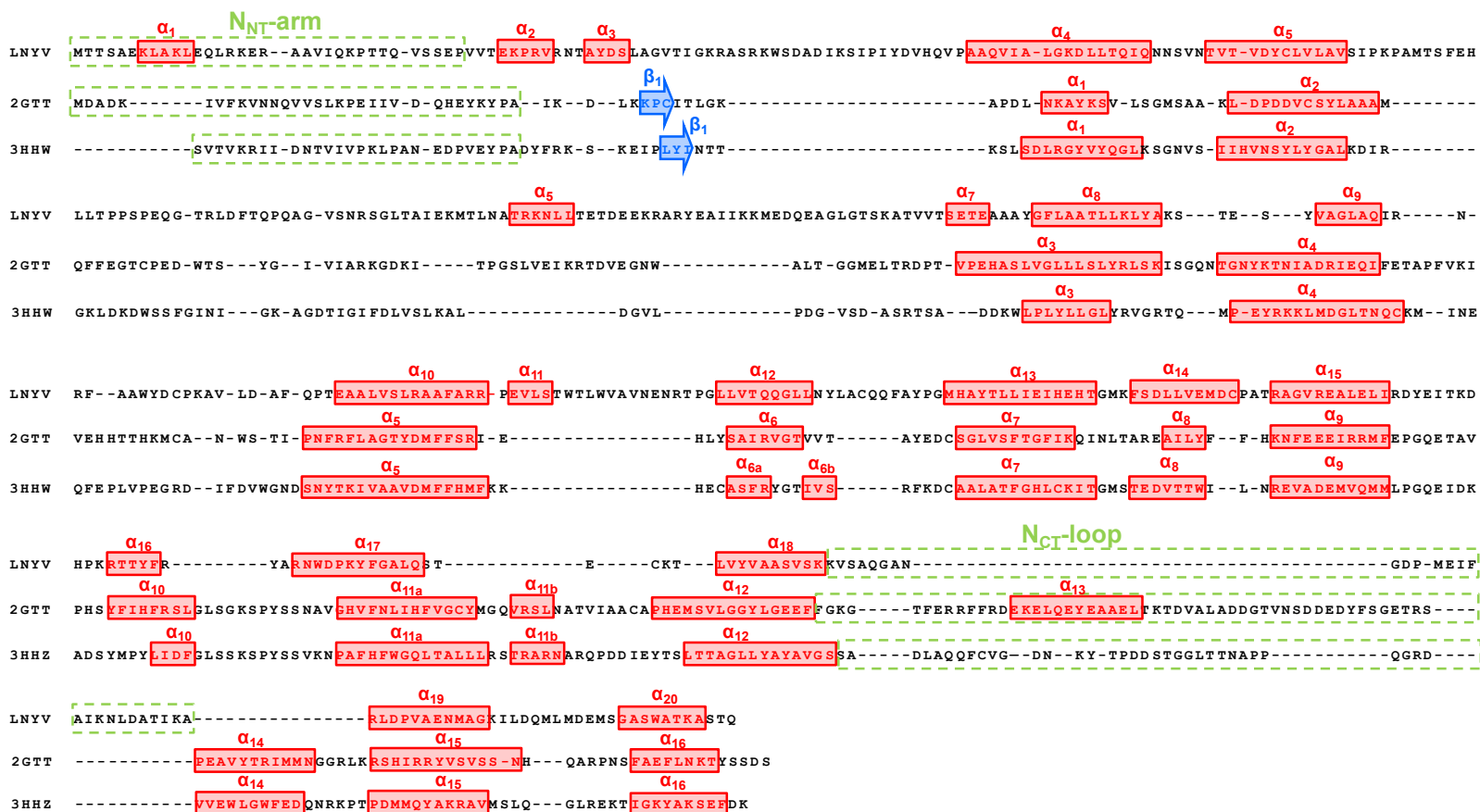


Figure 6.

Figure 6D



SUPPLEMENTARY FIGURES

Sequence alignments

Sequence alignment showing amino acid sequences for various proteins across different conditions and species. The sequences are color-coded to highlight specific residues.

Section 1 (Top): Aligns sequences for proteins like LNYV/1-300, LYMN/1-306, NCMV/1-286, CVS-11/1-297, Pasteur/1-297, ABLV/1-297, ABLV2/1-297, EBLV/1-298, EBLV2/1-297, Duvenhage/1-298, Lagoa/1-305, Mokoia/1-303, Aravan/1-297, Hut1/1-298, Khujandi/1-297, WCBV/1-297, Ocray/1-265, San/1-265, Central/1-265, Glasgow/1-265, Mudd-Summers/1-265, North/1-265, South/1-265, Misouini/1-274, Ogden/1-274, Piry/1-327, Chandipura/1-293, Ifahan/1-289, COCAL/1-264, and MARABA/1-265. The alignment spans from position 1 to 56.

Section 2 (Middle): Aligns sequences for LNYV/1-300, LYMN/1-306, NCMV/1-286, CVS-11/1-297, Pasteur/1-297, ABLV/1-297, EBLV/1-298, EBLV2/1-297, Duvenhage/1-298, Lagoa/1-305, Mokoia/1-303, Aravan/1-297, Hut1/1-298, Khujandi/1-297, WCBV/1-297, Ocray/1-265, San/1-265, Central/1-265, Glasgow/1-265, Mudd-Summers/1-265, North/1-265, South/1-265, Misouini/1-274, Ogden/1-274, Piry/1-327, Chandipura/1-293, Ifahan/1-289, COCAL/1-264, and MARABA/1-265. The alignment spans from position 116 to 187.

Section 3 (Bottom): Aligns sequences for LNYV/1-300, LYMN/1-306, NCMV/1-286, CVS-11/1-297, Pasteur/1-297, ABLV/1-297, EBLV/1-298, EBLV2/1-297, Duvenhage/1-298, Lagoa/1-305, Mokoia/1-303, Aravan/1-297, Hut1/1-298, Khujandi/1-297, WCBV/1-297, Ocray/1-265, San/1-265, Central/1-265, Glasgow/1-265, Mudd-Summers/1-265, North/1-265, South/1-265, Misouini/1-274, Ogden/1-274, Piry/1-327, Chandipura/1-293, Ifahan/1-289, COCAL/1-264, and MARABA/1-265. The alignment spans from position 57 to 187.

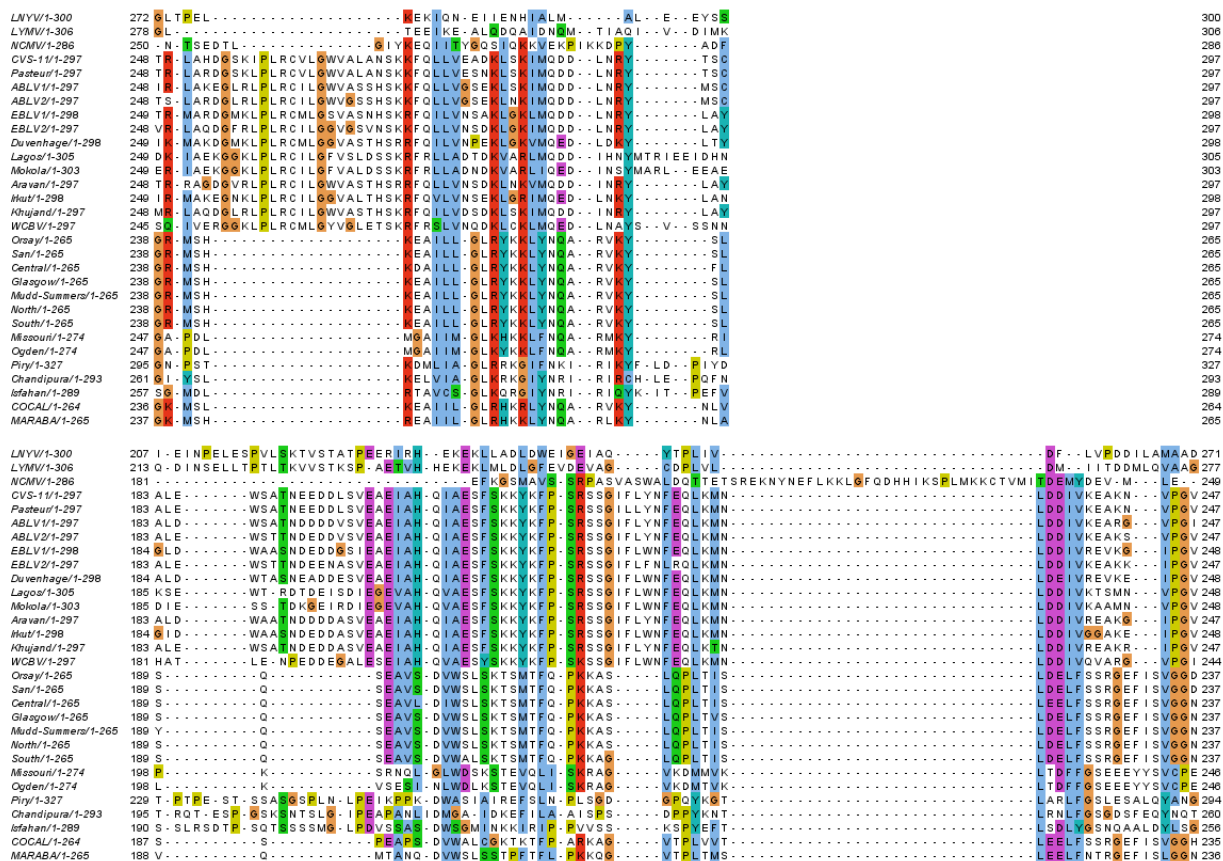
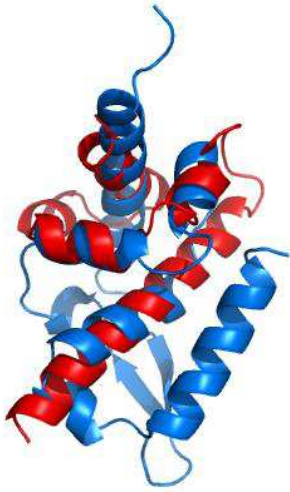
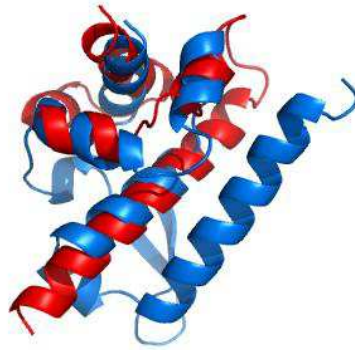


Figure S1

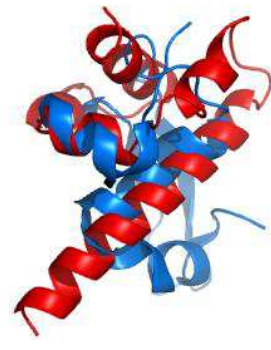
Rhabdoviridae



RAV (1VYI)

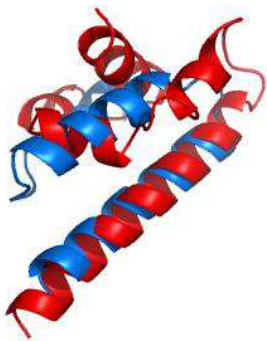


MOKV (2WZL)

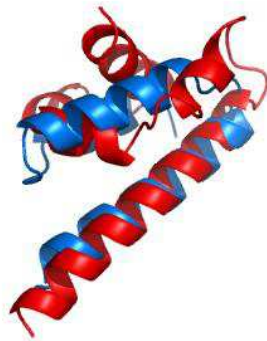


VSV (2K47)

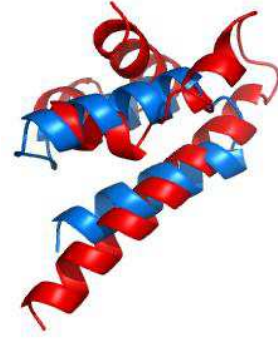
Paramyxoviridae



SEV (1R4G)



MEV (1OKS)



MUV (3BBZ)

Filoviridae

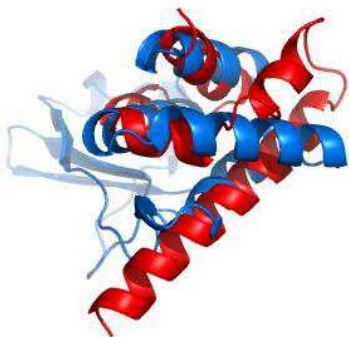


Figure S2

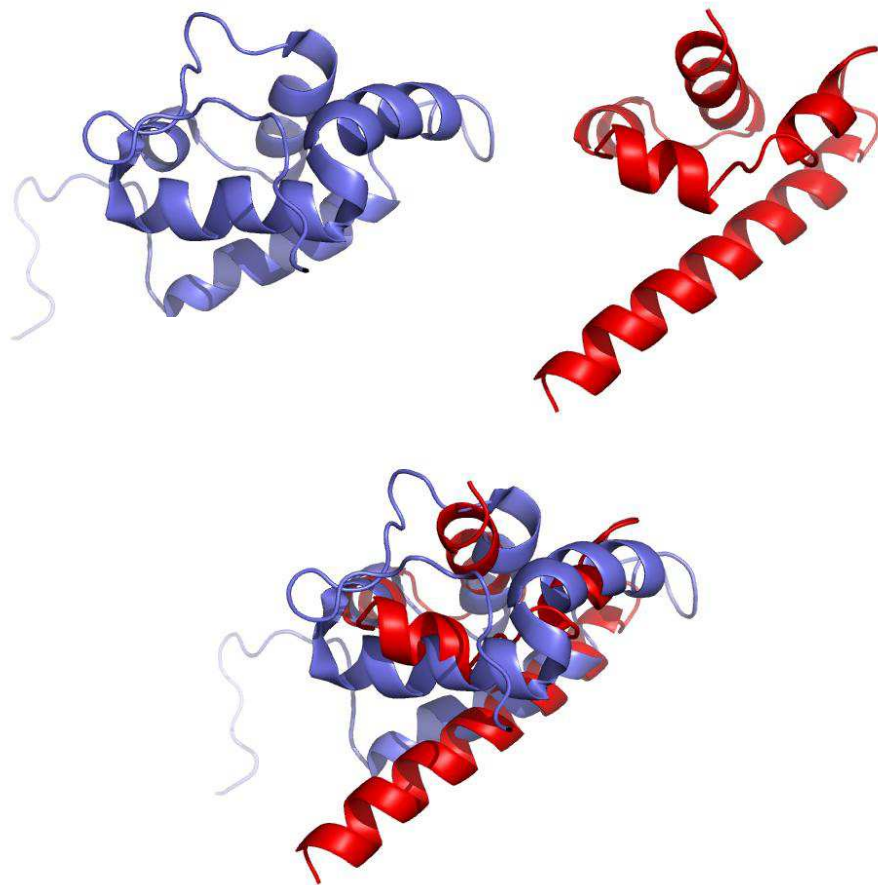


Figure S3

4.4 Conclusion

Les résultats de modélisation et de diffusion aux petits angles nous permettent de proposer un modèle où P occupe un espace très large, et plus particulièrement ses extrémités N et C-terminales qui doivent interagir avec d'autres partenaires. Cette organisation pourrait permettre au système d'être plus souple et de fonctionner efficacement en posant des contraintes minimales vis-à-vis du positionnement des différents éléments avec lesquels il doit interagir. La prochaine étape est d'utiliser l'algorithme d'optimisation qui minimiserait un χ^2 global obtenu en sommant les χ^2 des courbes dans les différentes conditions de contraste.

La structure à haute résolution du domaine C-terminal de la phosphoprotéine de LNYV nous montre que la structure des domaines C-terminaux des virus de la famille des *Rhabdoviridae* est conservée malgré des séquences en acides aminés différentes, et nous suggère un mode de fonctionnement du complexe de réplication similaire à ce que l'on observe chez d'autres membres des *Rhabdoviridae*. Cette structure confirme aussi l'appartenance de ce virus à la famille des *Rhabdoviridae*.

Dynamique des éléments du complexe de réplication du virus de la stomatite vésiculaire

"That's not my problem!"

I. Ivanov

Sommaire

5.1	Introduction	139
5.2	Matériels et méthodes	141
5.2.1	Préparation des échantillons	141
5.2.2	Mesure de l'intensité élastique incohérente	141
5.3	Résultats	141
5.3.1	Dynamique de P-CTD	141
5.3.2	Dynamique de P-NTD	144
5.3.3	Dynamique de la P entière	146
5.4	Conclusion	148

5.1 Introduction

Les phosphoprotéines des virus de la famille des *Rhabdoviridae* possèdent une organisation similaire [Gerard 2009] : elles contiennent de longues régions désordonnées séparées par des domaines structurés [Ribeiro 2008] [Ding 2006].

On sait que les régions désordonnées sont présentes de manière abondante (environ 60% des séquences génomiques eucaryotes sont prédites comme codant pour une région désordonnée de plus de 40 résidus [Ward 2004]). Ces régions sont connues pour avoir des fonctions associées, notamment dans les phénomènes de reconnaissance moléculaire [Tompa 2002]. Leur état désordonné leur permettrait d'explorer un volume 2 à 3 fois plus important qu'un peptide replié de même taille [Gunasekaran 2003].

La première étude portant sur la dynamique des protéines intrinsèquement désordonnées explorée par diffusion de neutrons incohérente [Gaspar 2008] date de 2008. Les auteurs ont montré que les amplitudes des mouvements dans les protéines désordonnées étaient plus importantes dans des protéines désordonnées, mais aussi que les

protéine désordonnées présentaient un élargissement quasiélastique plus important, indiquant une plus grande mobilité des protons de la protéine. La phosphoprotéine a des propriétés d'une protéine désordonnée, mais aussi d'une protéine ordonnée, et ce grâce à ses parties structurées. Ce chapitre présente une étude de cette protéine et de ses fragments par diffusion de neutrons incohérente. Elle vise à caractériser la dynamique de cette protéine, mais aussi à comprendre les différences de comportement entre les protéines intrinsèquement désordonnées et les protéines repliées d'un point de vue dynamique.

L'étude présentée dans ce chapitre porte sur la dynamique dans une fenêtre d'espace et de temps de 2\AA^2 et 100 ps environ, sur différentes constructions aux propriétés remarquables de la phosphoprotéine : la partie C-terminale, la partie N-terminale, et la protéine entière. Les prochains paragraphes présentent une description de ces différentes constructions.

Le domaine C-terminal de la phosphoprotéine : Le domaine C-terminal porte la fonction de liaison au complexe N-ARN [Green 2009]. Il forme une unité de repliement autonome de 9 kDa dont la structure est connue [Ribeiro 2008]. La construction que j'ai utilisée dans ma thèse, contenant les résidus 195 à 265, est une partie contenant exclusivement des résidus ordonnés [Ribeiro 2008].

Le domaine N-terminal de la phosphoprotéine : Le domaine N-terminal est le domaine permettant de former le complexe N⁰-P [Leyrat 2011b]. C'est un peptide de 7 kDa désordonné en solution qui a tendance à former des hélices- α de manière transiente [Leyrat 2011a]. Les auteurs ont démontré par une approche combinant RMN, SAXS et modélisation que ces hélices transientes étaient présentes en moyenne sur 9% des conformères. Celles-ci, allant des résidus 8 à 12 et 25 à 38, sont un élément de reconnaissance moléculaire [Mohan 2006] permettant la formation du complexe N⁰-P. Les résultats présentés dans cette thèse concernent la partie allant des résidus 1 à 60.

La phosphoprotéine entière : La phosphoprotéine est une composante essentielle du complexe de réplication, possédant des sites de fixation aux différents éléments de ce complexe [Green 2009] [Mavrikakis 2003] [Emerson 1975]. Elle forme des dimères de 69 kDa [Gerard 2007]. Sa structure désordonnée entrecoupée par des domaines structurés suggère que ses parties impliquées dans des phénomènes d'interaction avec les différents éléments du complexe de réplication occupent un volume très large, facilitant leur interaction [Gerard 2009].

On a donc 3 échantillons aux propriétés remarquables : une protéine désordonnée, une protéine repliée, et une protéine qui possède des régions désordonnées séparées par des domaines repliés.

5.2 Matériels et méthodes

5.2.1 Préparation des échantillons

Les différentes protéines ont été purifiées par le **Dr. Ivan Ivanov** : la production s'est fait en surexprimant la protéine voulue dans *E. coli*, dans un système inductible à l'IPTG. Les cellules ont été ensuite lysées par sonication, et la fraction soluble passée sur une colonne Ni-NTA. Les fractions contenant la protéine ont été vérifiées par SDS-PAGE, rassemblées, concentrées et passées sur une colonne de chromatographie d'exclusion de taille. Ensuite, les fractions contenant la protéine ont été concentrées puis dialysées dans un tampon préparé de manière à ce que la concentration finale (dans la cellule de mesure) soit de 150mM NaCl, 20mM tris pH 7.5. Les cellules en aluminium de 0.3 mm contenaient environ 100 mg de protéine au total pour chaque échantillon, dilué dans 200 μ L d'H₂O ou D₂O. Une cellule contenant 200 μ L de tampon 150mM NaCl, 20mM tris pH 7.5 a aussi été préparée pour la soustraction du tampon.

5.2.2 Mesure de l'intensité élastique incohérente

Les mesures de diffusion élastique ont été faites sur IN13. La réduction des données a été faite de la même manière que dans le chapitre 3.

5.3 Résultats

5.3.1 Dynamique de P-CTD

5.3.1.1 Mesures de la diffusion élastique sur IN13

Les données réduites sont montrées figure 5.1. Les mesures ont été faites en solvant deutéré et non deutéré, sur une gamme de températures de 280-310 K pour l'échantillon en solvant hydrogéné et 280-300 K pour l'échantillon en solvant deutéré. J'ai poursuivi l'analyse en analysant les courbes dans le cadre de l'approximation gaussienne, afin d'extraire les déplacements carrés moyens pour chaque condition de température et de solvant. L'analyse est montrée figure 5.2

Les différences entre les deux courbes sont minimales, et elles présentent une résilience très forte, les $\langle u^2 \rangle$ restent constants sur la gamme de température explorée (figure 5.3). Le D₂O n'a pas d'effet sur la dynamique de cette protéine dans la fenêtre de temps définie par la résolution énergétique instrumentale.

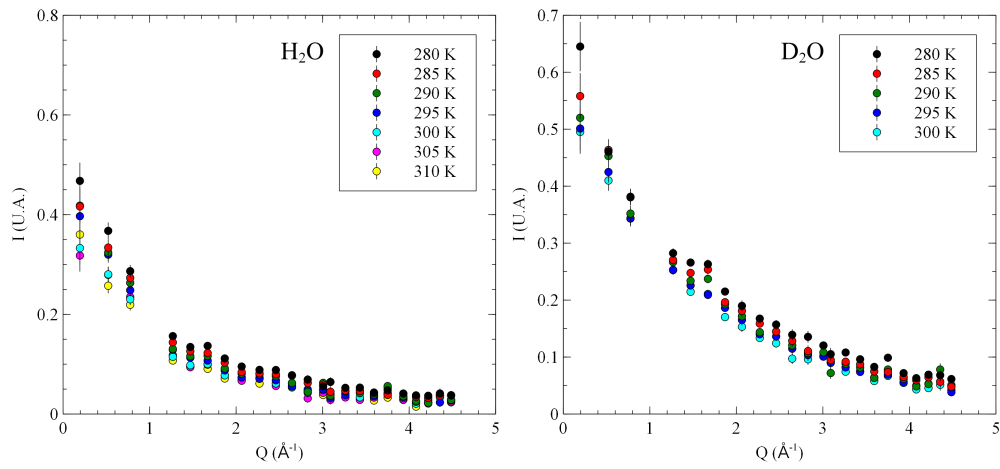


FIGURE 5.1 – Courbe de diffusion élastique incohérente sur IN13 de P-CTD en solvant H_2O et D_2O , après réduction des données

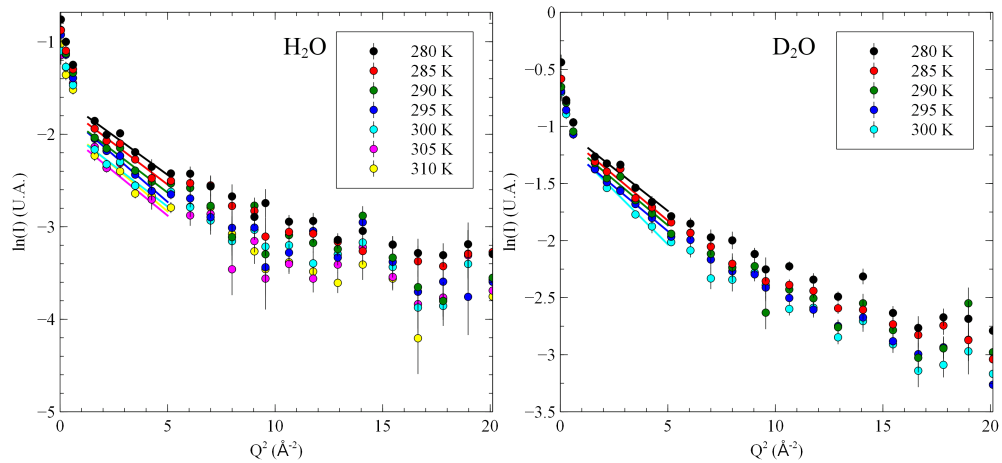


FIGURE 5.2 – Courbe de diffusion élastique incohérente sur IN13 de P-CTD en solvant H_2O et D_2O en représentation $\ln(I)$ en fonction de Q^2 . Les lignes représentent la courbe obtenue dans le cadre de l'approximation gaussienne.

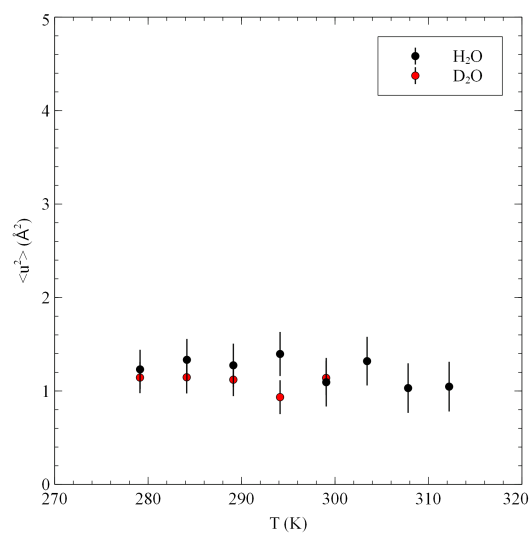


FIGURE 5.3 – Evolution des déplacements carrés moyens de P-CTD en solvant H₂O et D₂O en fonction de la température.

5.3.2 Dynamique de P-NTD

5.3.2.1 Mesures de la diffusion élastique sur IN13

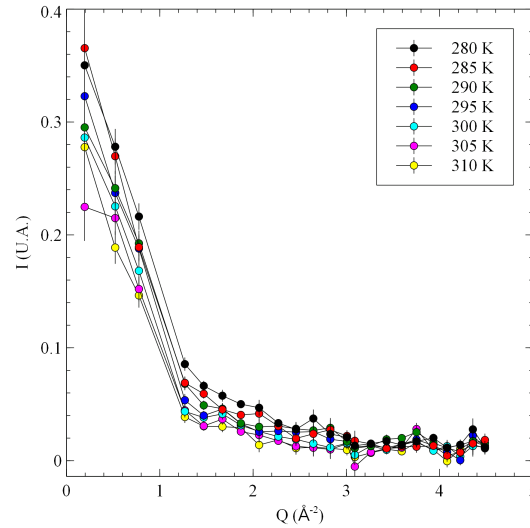


FIGURE 5.4 – Courbe de diffusion élastique incohérente sur IN13 de P-NTD, après réduction des données.

Les données réduites sont montrées figure 5.4. Les mesures n'ont été faites que pour un solvant hydrogéné. On note une baisse du rapport signal sur bruit par rapport aux échantillons P-CTD et P entière, et ce malgré la même quantité de matériel dans la cuve. Ceci peut s'expliquer par une dynamique trop forte, les mouvements trop rapides vont alors donner lieu à de la diffusion quasiélastique (sortie des mouvements de la résolution instrumentale). Le calcul des déplacements carrés moyens ainsi que l'évolution de ceux ci en fonction de la température sont montrés figure 5.5. Ce qui est surprenant est que les déplacements carrés moyens sont identiques à ceux de P-CTD, alors que l'on devrait s'attendre à des valeurs beaucoup plus élevées, comme cela avait été décrit précédemment [Gaspar 2008]. On mesure donc une résilience de 0.169 N/m, ce qui est deux fois plus faible que la valeur mesurée pour la myoglobine [Bicout 2001].

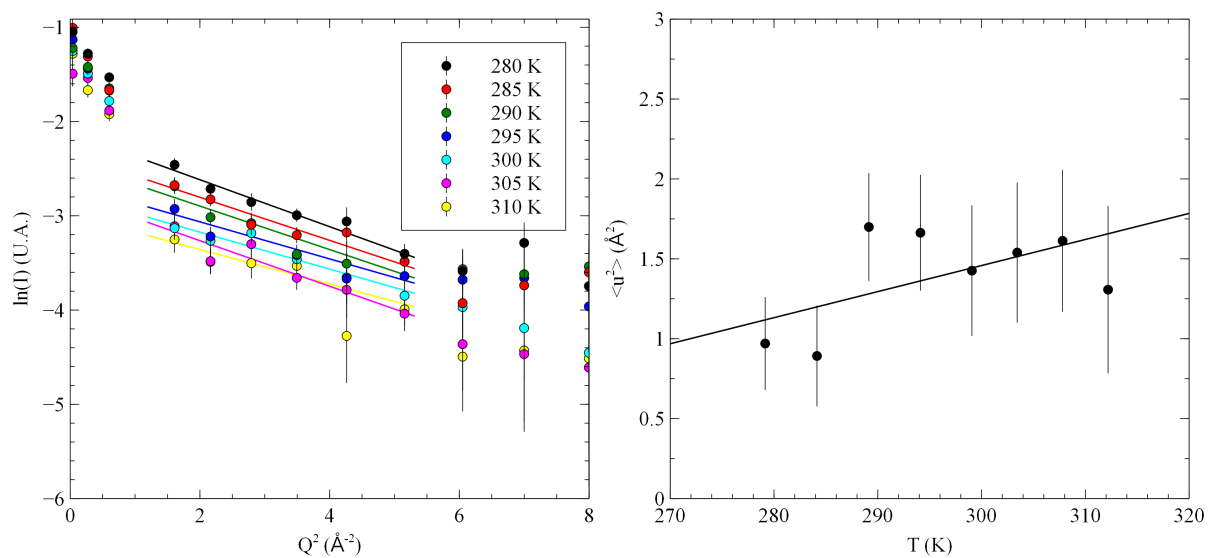


FIGURE 5.5 – Courbe de diffusion élastique incohérente sur IN13 de P-NTD en représentation $\ln(I)$ en fonction de Q^2 , les courbes représentant l'analyse des déplacements carrés moyens en utilisant l'approximation gaussienne (gauche), et déplacements carrés moyens en fonction de la température (droite).

5.3.3 Dynamique de la P entière

5.3.3.1 Mesures de la diffusion élastique sur IN13

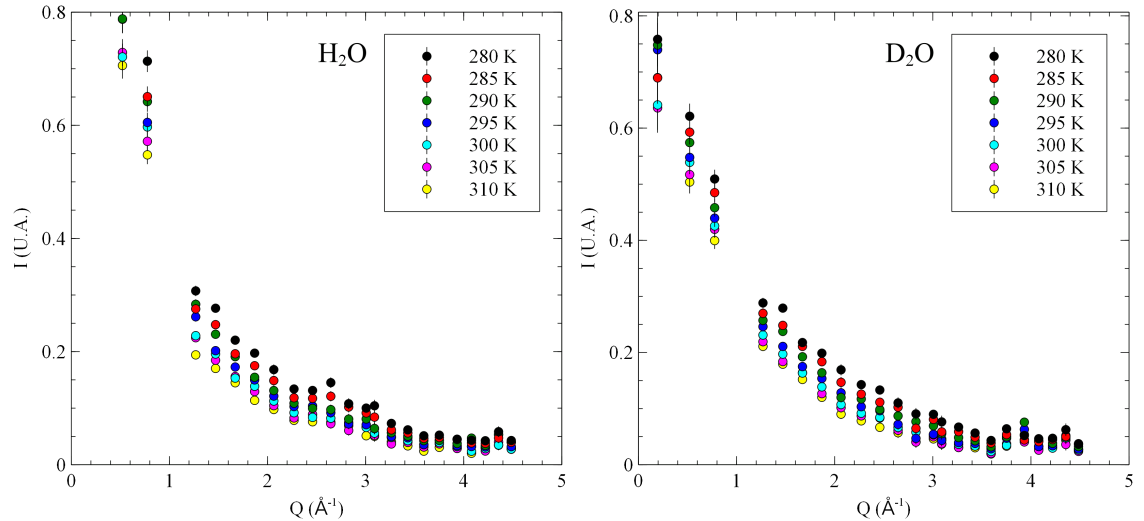


FIGURE 5.6 – Courbe de diffusion élastique incohérente sur IN13 de P entière en solvant H_2O et D_2O , après réduction des données

Les données brutes sont montrées figure 5.6 Comme précédemment, j’ai fait analysé les données en utilisant l’approximation gaussienne afin d’en extraire des déplacements carrés moyens (figure 5.7).

Les déplacements carrés moyens sont plus élevés que dans le cas de P-NTD et P-CTD (figure 5.8), ce qui peut être expliqué par le fait que les parties désordonnées nous donnent une contribution dans notre signal de diffusion. J’ai ensuite mesuré une résilience de 0.154 N/m dans le solvant D_2O et de 0.224 N/m dans le solvant H_2O (figure 5.8). Cette valeur est comparable à celle trouvée pour la partie P-NTD, nous montrant donc que cette protéine possède des propriétés d’une protéine désordonnée sur une échelle de temps de l’ordre d’une centaine de picosecondes. La différence de résilience entre les mesures faites en solvant deutéré et hydrogéné n’est pas significative pour que l’on puisse conclure sur une différence de comportement.

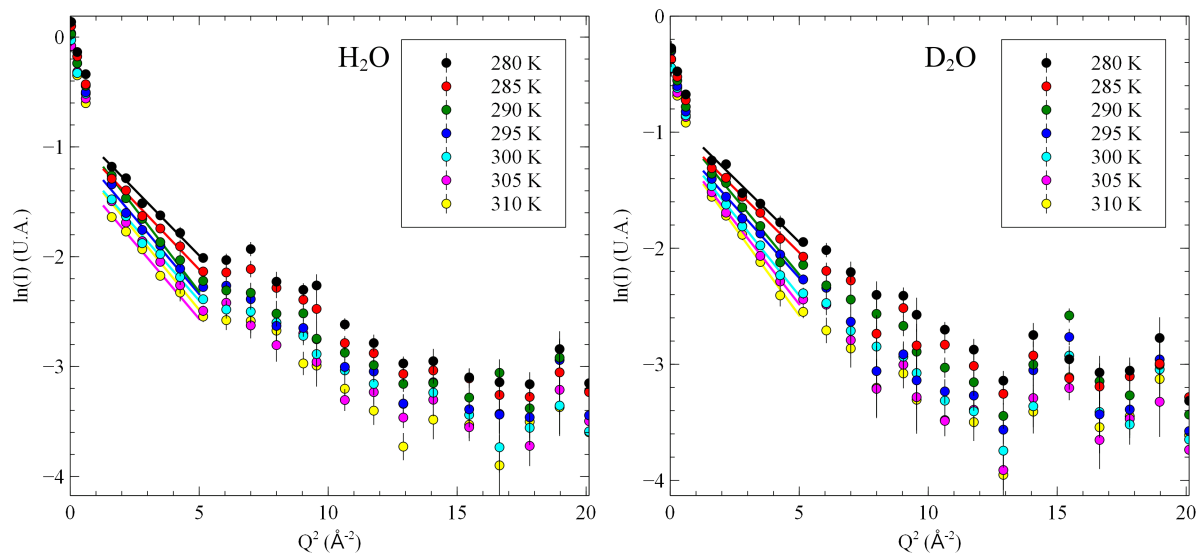


FIGURE 5.7 – Courbe de diffusion élastique incohérente sur IN13 de P entière en solvant H_2O et D_2O en représentation $\ln(I)$ en fonction de Q^2 . Les lignes représentent la courbe obtenue dans le cadre de l'approximation gaussienne.

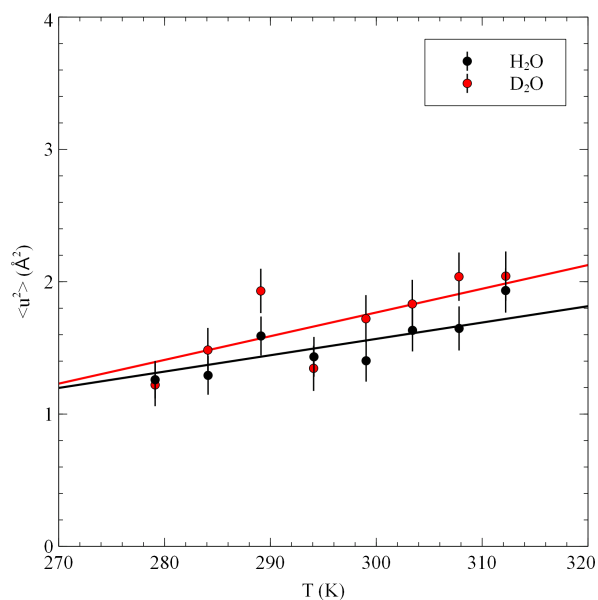


FIGURE 5.8 – Evolution des déplacements carrés moyens calculés pour PFL en fonction de la température, en solvant D_2O et H_2O . Les lignes correspondent aux courbes utilisées pour le calcul de la résilience.

5.4 Conclusion

Le résumé des valeurs $\langle k' \rangle$ et $\langle u^2 \rangle$ des différents échantillons dans différentes conditions de solvant sont montrés tableau 5.1.

Echantillon	$\langle k' \rangle$ H ₂ O(N/m)	$\langle u^2 \rangle$ H ₂ O(Å ²)	$\langle k' \rangle$ H ₂ O(N/m)	$\langle u^2 \rangle$ H ₂ O(Å ²)
P-FL	0.224	1.51	0.154	1.69
P-NTD	0.169	1.36	-	-
P-CTD	N/A	1.22	N/A	1.10

TABLE 5.1 – Résiliences et déplacements carrés moyens des échantillons de P-FL, P-NTD, et P-CTD dans des solvants deutérés ou hydrogénés.

Les déplacements carrés moyens mesurés pour tous les échantillons sont plutôt bas si on les compare à des mesures faites dans les mêmes conditions sur la myoglobine [Zaccai, G. 2000]. Les auteurs avaient alors mesuré des valeurs de $\langle u^2 \rangle$ aux alentours de 2 Å² pour une protéine concentrée à 200 mg/mL. Ils avaient alors suggéré que l'augmentation des déplacements carrés moyens par rapport à un échantillon sous forme de poudre était due aux mouvements diffusifs translationnels de la molécule entière, mais ils avaient souligné que leurs mesures ne permettaient pas d'exclure d'autres effets. La différence que l'on observe avec nos échantillons peut donc venir d'une différence dans le comportement diffusif des différentes constructions. Des mesures de spectres résolus en énergie auraient permis de séparer ces différentes contributions [Bée 1988], et mesurer les différences entre la dynamique interne des échantillons.

Les résiliences mesurées sont consistantes avec l'état désordonné ou ordonné des différents fragments [Gerard 2009]. La partie C-terminale structurée est très résiliente, la partie N-terminale désordonnée est beaucoup plus molle, alors que la protéine entière possède une résilience se situant entre celle d'une protéine désordonnée et celle d'une protéine structurée.

La comparaison des mesures de résilience de la phosphoprotéine dans des solvants D₂O et H₂O nous indique que cette protéine voit sa résilience baisser, ce qui indique un effet "entropique" du D₂O qui favoriserait l'accès à un nombre plus élevé de sous-états conformationnels. Ce comportement avait déjà été décrit pour la BSA [Tehei 2001]. Les auteurs avaient effectué des expériences comparant les résiliences d'une protéine mésophile et une protéine halophile dans des solvants non-deutérés et deutérés. Ils avaient observé une augmentation de la résilience lors du passage à un solvant deutéré pour la protéine en solvant deutéré, alors que le contraire se passait pour la protéine mésophile [Tehei 2001], et ce malgré le fait que les solvants deutérés avaient des propriétés stabilisantes. Leur interprétation était que, pour la protéine mésophile, l'effet stabilisant du D₂O était dû à une augmentation de l'entropie, ce qui se traduisait par une résilience plus faible. Pour la protéine halophile par contre, l'augmentation de la résilience était due à une stabilisation des liaisons deutérium, plus forte que la liaison hydrogène.

Conclusion générale et perspectives

"C'est comme un singe qui essaye d'attraper une branche pour ne pas tomber."

L. Estrozi

Sommaire

6.1	L'hémagglutinine filamenteuse de <i>Bordetella pertussis</i>	. . . 149
6.2	Le complexe de réplication des <i>Rhabdoviridae</i> 150

6.1 L'hémagglutinine filamenteuse de *Bordetella pertussis*

L'hémagglutinine filamenteuse de *Bordetella pertussis* constitue un système TPS modèle dont les différentes composantes ont été caractérisées structuralement [Clantin 2004] [Clantin 2007] et ont permis de proposer un modèle où FhaB traverse la membrane externe sous une forme dépliée. Ce modèle a conduit à la découverte de plusieurs chaperonnes périplasmiques [Baud 2009] protégeant FHA dans le périplasma, dont Par27 [Hodak 2008].

Le travail présenté dans cette thèse nous a permis de proposer un modèle structural haute résolution de Par27. Les mesures de diffusion aux petits angles ont pu établir que cette protéine, bien que possédant des domaines structurés, possédait une flexibilité intrinsèque que l'on retrouve dans de nombreuses chaperonnes [Saul 2004] [Xu 2007]. Nous avons pu aussi, en combinant une approche de modélisation avec la technique de diffusion aux petits angles, fournir un premier modèle haute résolution du complexe entre cette chaperonne et Fha30. Ce modèle suggère que partie N-terminale de Fha30 reste désordonnée et donc accessible à FhaC qui peut donc ainsi la transporter à travers la membrane externe. Des études biophysiques complémentaires pourraient permettre de valider ce modèle, et de représenter ce système sous la forme d'un ensemble de modèles qui refléterait mieux le caractère flexible de ce complexe.

L'autre partie de cet travail concerne le repliement et les propriétés dynamiques de Fha30. Cette protéine possède une structure composée exclusivement de feuillets β ,

formant une hélice β [Clantin 2004]. Les mesures de diffusion de neutrons incohérente nous montrent que cette structure est plus rigide qu'une protéine globulaire composée exclusivement d'hélices α , suggérant que cette structure en hélice β lui confère des propriétés mécaniques particulières qui doivent être importantes pour la fonction biologique d'adhésion cellulaire. Nous avons pu montrer, en combinant des techniques de dynamique moléculaire et de microscopie à force atomique que Fha30 se déplie par 'morceaux', comme si il s'agissait d'une protéine ayant plusieurs domaines. Ces résultats suggèrent que le repliement à la surface cellulaire se fait de manière vectorielle dès la sortie du transporteur membranaire. Le transporteur pourrait ainsi agir comme une contrainte qui "oblige" la protéine à se replier de la partie N vers la partie C-terminales, et permettrait d'expliquer pourquoi ce transporteur est essentiel au bon repliement de cette protéine. Une étude du repliement de cette protéine nous permettrait de répondre à ces questions.

6.2 Le complexe de réplication des *Rhabdoviridae*

La phosphoprotéine joue un rôle central dans le complexe réplication du virus par des interactions avec les différentes composantes du système. Elle possède une flexibilité intrinsèque [Gerard 2009] mais aussi des domaines structurés [Ribeiro 2008] [Ding 2006]. Nous avons étudié les propriétés structurales de la phosphoprotéine de VSV dans un contexte de complexe avec la nucléoprotéine. Les structures à haute résolution du domaine C-terminal de la phosphoprotéine [Ribeiro 2008], son domaine central [Ding 2006] et du complexe que sa partie N-terminale forme avec la nucléoprotéine dépourvue d'ARN [Leyrat 2011b] étant connues, la techniques de diffusion aux petits angles est idéale pour étudier la structure de ce complexe. Nos résultats montrent que la flexibilité facilite l'interaction avec les autres partenaires (la partie C-terminale avec le complexe nucléoprotéine-ARN, la partie N-terminale avec la polymérase et la nucléoprotéine avec l'ARN viral) car la zone explorée par les différentes parties est très large. Afin de compléter ce travail et de préciser le rôle de ce complexe *in-vivo*, il serait intéressant de mesurer l'influence de ce complexe sur l'activité de la polymérase, et de mesurer l'affinité que ce complexe a pour l'ARN. La flexibilité de cette protéine en fait aussi une candidate idéale pour des études de dynamiques. Nous avons mesuré la dynamique de la partie N-terminale dans un temps de 100ps. Nos résultats montrent que la partie N-terminale est plus "molle" que la partie C-terminale dans cette gamme de temps, et que la résilience de la protéine entière se situe entre les deux. La comparaison des résiliences entre la protéine entière dans des solvants hydrogénés et deutérés nous a montré que le D₂O a un effet entropique, favorisant l'accès à un plus grand nombre de sous-états conformationnels. Des mesures de diffusion incohérente résolues en temps nous permettraient de séparer les différents types de mouvements et avoir un description plus complète de la dynamique de ce système.

ANNEXE A

Etude du désordre dans les nucleocapsides du virus de la rougeole

"This arrived very recently, just showing that measles is a problem in Grenoble."

R. Ruigrok, a propos de l'épidémie de rougeole de 2011

Intrinsic disorder in measles virus nucleocapsids

Malene Ringkjøbing Jensen^a, Guillaume Communie^{a,b}, Euripedes Almeida Ribeiro, Jr.^{a,b}, Nicolas Martinez^b, Ambroise Desfosses^b, Loïc Salmon^a, Luca Mollica^a, Frank Gabel^a, Marc Jamin^b, Sonia Longhi^c, Rob W. H. Ruigrok^b, and Martin Blackledge^{a,1}

^aInstitut de Biologie Structurale Jean-Pierre Ebel, Commissariat à l'Énergie Atomique, Centre National de la Recherche Scientifique, Université Joseph Fourier, 41, Rue Jules Horowitz, 38027 Grenoble, France; ^bUnit for Virus Host Cell Interactions, Centre National de la Recherche Scientifique, Université Joseph Fourier, European Molecular Biology Laboratory, 6, Rue Jules Horowitz, 38042 Grenoble, France; and ^cArchitecture et Fonction des Macromolécules Biologiques, Centre National de la Recherche Scientifique, Universités d'Aix-Marseille I and II, 163 Avenue de Luminy, 13288 Marseille Cedex 09, France

Edited by Angela M. Gronenborn, University of Pittsburgh School of Medicine, Pittsburgh, PA, and approved April 25, 2011 (received for review February 28, 2011)

The genome of measles virus is encapsidated by multiple copies of the nucleoprotein (N), forming helical nucleocapsids of molecular mass approaching 150 Megadalton. The intrinsically disordered C-terminal domain of N (N_{TAIL}) is essential for transcription and replication of the virus via interaction with the phosphoprotein P of the viral polymerase complex. The molecular recognition element (MoRE) of N_{TAIL} that binds P is situated 90 amino acids from the folded RNA-binding domain (N_{CORE}) of N, raising questions about the functional role of this disordered chain. Here we report the first in situ structural characterization of N_{TAIL} in the context of the entire N-RNA capsid. Using nuclear magnetic resonance spectroscopy, small angle scattering, and electron microscopy, we demonstrate that N_{TAIL} is highly flexible in intact nucleocapsids and that the MoRE is in transient interaction with N_{CORE} . We present a model in which the first 50 disordered amino acids of N_{TAIL} are conformationally restricted as the chain escapes to the outside of the nucleocapsid via the interstitial space between successive N_{CORE} helical turns. The model provides a structural framework for understanding the role of N_{TAIL} in the initiation of viral transcription and replication, placing the flexible MoRE close to the viral RNA and, thus, positioning the polymerase complex in its functional environment.

NMR | SAXS | ensemble description | dynamics | unfolded protein

Measles virus (MeV) is a member of the *Paramyxoviridae* family of the *Mononegavirales* order of negative sense, single stranded RNA viruses. The viral genome is encapsidated by multiple copies of the nucleoprotein (N) forming a helical nucleocapsid. Transcription and replication of the viral RNA are initiated by an interaction between N and the polymerase complex, composed of the phosphoprotein (P) and the RNA-dependent RNA polymerase (1). N consists of two domains: N_{CORE} (residues 1–400), responsible for the interaction with the viral RNA and for maintaining the nucleocapsid structure, and a long intrinsically disordered domain, N_{TAIL} (residues 401–525) serving as the anchor point for the polymerase complex (2, 3). The molecular recognition element (MoRE) (residues 485–502) of the disordered N_{TAIL} interacts with the C-terminal three-helix bundle domain, XD, of P (residues 459–507) (4) and thereby recruits the polymerase complex onto the nucleocapsid template (5, 6).

The realization that intrinsically disordered proteins (IDPs) are functional despite a lack of structure (7–9) has revealed entirely new paradigms that appear to redefine our understanding of the role of conformational flexibility in molecular interactions (10–12). Until now most IDPs have been studied in isolation, or in the presence of a single interaction partner, although it is evident that a real physiological environment could influence the nature and relevance of apparent intrinsic disorder. In this context resolving the question of whether the protein is actually disordered in situ is of paramount importance. In this case the mechanistic role of the extensive disorder present in N_{TAIL}

is particularly intriguing, because the MoRE is located at a distance of 90 apparently unfolded amino acids away from the folded N_{CORE} domain that binds the RNA (13). In order to resolve the mechanism by which the remote interaction between N_{TAIL} and the polymerase complex initiates transcription and replication, it is necessary to develop an atomic resolution understanding of molecular disorder in the context of the intact nucleocapsid. Here we use Nuclear Magnetic Resonance (NMR) spectroscopy, small angle scattering (SAS), and electron microscopy (EM) to describe the conformational behavior and mechanistic role of N_{TAIL} in situ.

Results

N_{TAIL} Populates a Dynamic Equilibrium Comprising Preencoded Helical Conformers at the Phosphoprotein Recognition Site. In this study we have developed an atomic resolution ensemble description of isolated N_{TAIL} from MeV using recently developed tools designed to provide quantitative descriptions of conformational equilibria in IDPs on the basis of experimental NMR data (14–16). Chemical shifts (17, 18) and residual dipolar couplings (RDCs) (19, 20), measured in a weakly ordering alignment medium were combined to directly probe the level and nature of residual structure in N_{TAIL} , revealing that while the majority of N_{TAIL} behaves like an intrinsically disordered chain, the MoRE exists in a rapidly interconverting conformational equilibrium between an unfolded form and conformers containing one of four discrete α -helical elements situated around the interaction site (Fig. 1, Fig. S1, Tables S1 and S2). All of these α -helices are found to be stabilized by N-capping interactions mediated by side chains of four different aspartic acids or serines that precede the observed helices (21, 22). N-capping stabilization of helices or turns represents an important mechanism by which the primary sequence encodes prerecognition states in disordered proteins, and has been observed in the proteins Tau (23), Sendai virus N_{TAIL} (19), the N-terminal transactivation domain of p53 (24), and the ribosomal protein L9 (25).

A crystal structure of the chimeric complex between a short construct of N_{TAIL} and XD shows that N_{TAIL} docks as a helix between residues Q486 and A502 (26). This helix is similar to the longest of the four helical elements present in isolated N_{TAIL} . Changes in chemical shifts and RDCs (Fig. 2) confirm that upon binding to XD, the MoRE of N_{TAIL} folds into a helix. However the decreasing values of secondary structure propensity (SSP) (17) and the RDCs towards the ends of the helix indicate

Author contributions: M.R.J., F.G., R.W.H.R., and M.B. designed research; M.R.J., G.C., E.A.R.J., N.M., A.D., L.S., L.M., F.G., M.J., S.L., and M.B. performed research; M.R.J., G.C., F.G., and M.B. analyzed data; and M.R.J., F.G., R.W.H.R., and M.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. E-mail: martin.blackledge@ibs.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1103270108/-DCSupplemental.

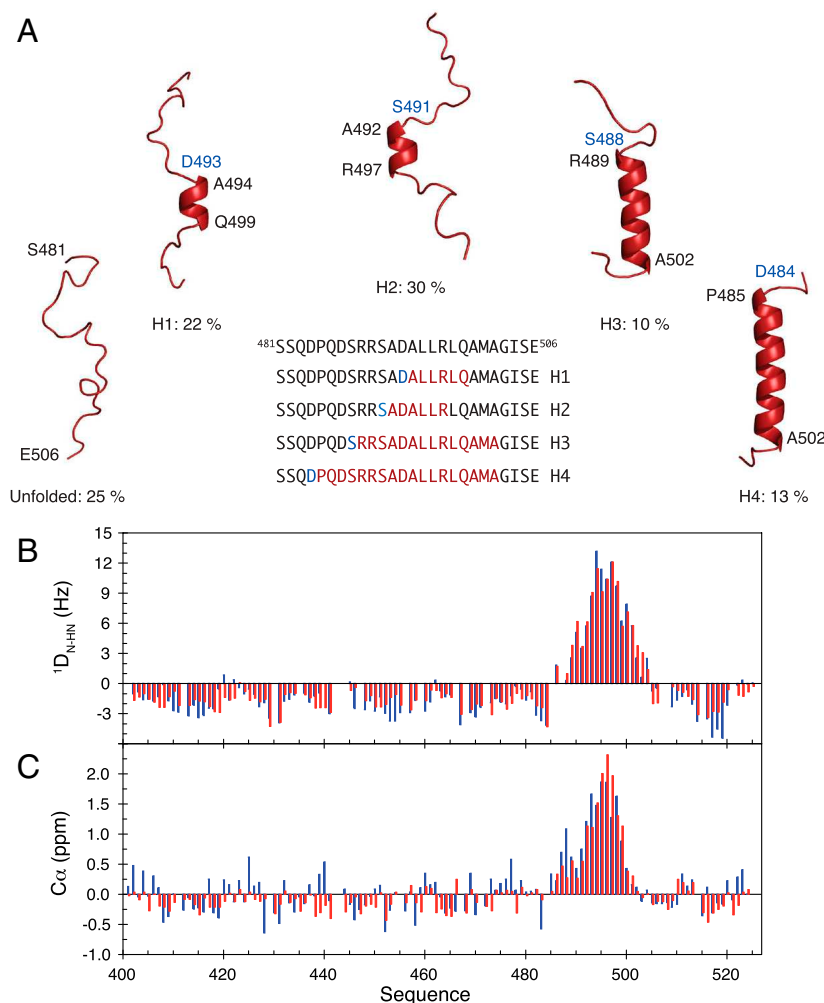


Fig. 1. Ensemble description of the MoRE of N_{TAIL} . (A) N_{TAIL} preferentially adopts a dynamic equilibrium between a completely unfolded state and different partially helical conformations each represented by a single cartoon structure for clarity. All helices are stabilized by N-capping interactions through aspartic acids or serines (blue residues). The location of the helices within the MoRE is shown in the primary sequence. (B) Comparison of experimental (blue) and back-calculated (red) D_{N-HN} RDCs from the model of N_{TAIL} shown in (A). (C) Comparison of experimental (blue) and back-calculated (red) C_{α} secondary chemical shifts from the model of N_{TAIL} shown in (A).

some residual degree of dynamics in the complex. In addition, exchange line broadening persists for residues surrounding the two smallest helices (H1 and H2) present in the conformational

equilibrium, even for a large excess of XD compared to N_{TAIL} . There is therefore evidence that both conformational selection from the equilibrium free-form ensemble, and coupled folding and binding, drive the interaction between N_{TAIL} and XD, testifying to the complexity of this highly dynamic interaction.

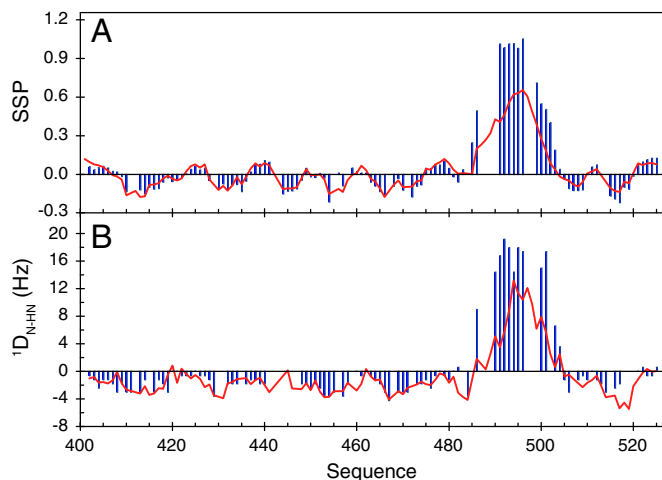


Fig. 2. The MoRE of N_{TAIL} folds upon binding to the XD domain of P protein. (A) SSP (17) of N_{TAIL} obtained from experimental C_{α} and C_{β} chemical shifts in free (red) and P (XD) bound (blue) form. (B) N-HN RDCs in free (red) and bound (blue) form of N_{TAIL} .

N_{TAIL} Remains Flexible in Intact Nucleocapsids and Binds Transiently to the Capsid Surface. Although the MoRE folds upon binding, the remainder of the 90 amino acid long N-terminal chain between the interaction site and N_{CORE} remains flexible (Fig. 2), again raising the intriguing question of the functional role of this long strand. To extend the investigation of N_{TAIL} to a physiologically relevant environment, we have therefore used solution state NMR to characterize the conformational behavior and flexibility of ^{15}N , ^{13}C labeled nucleocapsids. From EM (Fig. 3) we estimate the molecular mass distribution of the objects in the NMR sample to fall in a range between 2 to 50 Megadalton that would normally preclude detection of solution state NMR signals of a folded globular protein (27). The heteronuclear single quantum coherence (HSQC) spectrum of the intact capsids however reveals that N_{TAIL} remains flexible when attached to the nucleocapsid. Comparisons of 1H - ^{15}N (Fig. 3), and ^{13}C - ^{13}C (Fig. S2) correlation spectra of the isolated N_{TAIL} domain and intact nucleocapsids show that the NMR resonances superimpose, demonstrating that the local conformational behavior of residues 450–525 of N_{TAIL} is

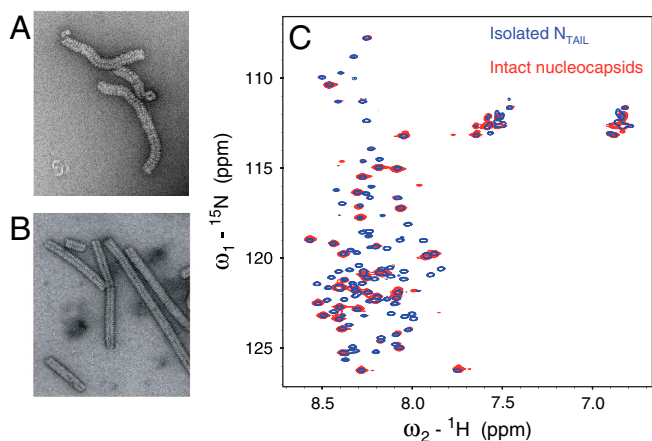


Fig. 3. Electron microscopy and NMR studies of Measles virus nucleocapsids. (A) Electron micrograph (negative staining) of the ^{13}C , ^{15}N labeled nucleocapsid sample used for solution NMR studies. (B) Electron micrograph of trypsin-digested ^{13}C , ^{15}N labeled nucleocapsids. The solution NMR spectrum of this sample was empty. (C) Superposition of the ^1H - ^{15}N HSQC spectrum of isolated N_{TAIL} (blue) and intact nucleocapsids (red).

retained in situ. However, signals for the first 50 amino acids (residues 401–450) are absent, while large variations of peak intensities indicate differential flexibility along the remainder of the chain, with the MoRE having particularly low intensities (Fig. 4A).

To further probe the conformational dynamics of N_{TAIL} , we have measured ^{15}N R_2 spin relaxation rates in isolated N_{TAIL} and intact nucleocapsids (Fig. 4B). Isolated N_{TAIL} shows uniform R_2 relaxation rates throughout the sequence, except in the MoRE where the presence of residual helical structure results in elevated rates. ^{15}N R_2 values of N_{TAIL} in the capsid exhibit a very different profile. In the center of the MoRE (around residue 495) R_2 values are similar to the rates in the isolated N_{TAIL} domain,

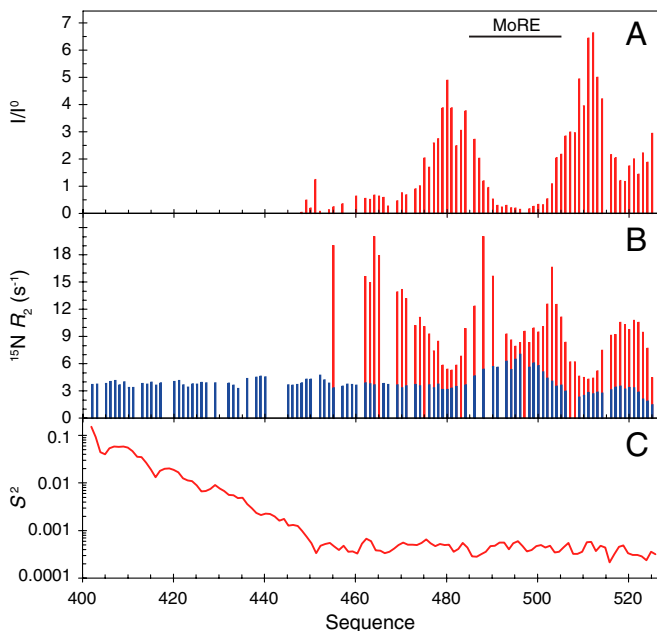


Fig. 4. Dynamics of N_{TAIL} in intact capsids. (A) Intensity profile of the ^1H - ^{15}N HSQC spectrum of intact nucleocapsids. The intensity profile was calculated as the ratio of the intensities (I) in the capsid spectrum and the intensities in the spectrum of the free N_{TAIL} domain (I^0). (B) Comparison of ^{15}N R_2 relaxation rates measured on a 1 GHz spectrometer in the free form of N_{TAIL} (blue) and in intact nucleocapsids (red). (C) N-H angular order parameter S^2 averaged over an ensemble of 5,000 conformers of N_{TAIL} that were calculated as shown in Fig. 5 and described in the *Methods* section.

indicating that the MoRE is in slow exchange, while the larger relaxation rates observed at the edges of the MoRE indicate that the same exchange rate appears faster (smaller chemical shift differences) for these sites. These results suggest that the MoRE of N_{TAIL} slowly exchanges on and off the surface of the nucleocapsids. Analysis of the intensity of the peaks shows that more than 95% of the MoRE population is bound. The R_2 values increase dramatically around residue 460, which, combined with the absence of signals of the first 50 residues of N_{TAIL} , indicates that the first stretch of 50 amino acids of the unfolded domain is conformationally restricted. We note that the C terminus of the protein also interacts, either directly with the capsid, or folds back onto the MoRE as it interacts with the capsid.

N_{TAIL} Exfiltrates from Inside to Outside of the Capsid Helix Through the Interstitial Space Between Successive N_{CORE} Helical Turns. MeV nucleocapsids have previously been visualized by EM, exhibiting a characteristic herring-bone appearance (5, 28–31). Nothing is known about the location and conformational state of N_{TAIL} in intact nucleocapsids because N_{TAIL} does not appear to contribute coherently to the reconstructed density from EM, however it is apparent that both the structure and dynamics of the nucleocapsids are significantly modulated by N_{TAIL} . Whereas full-length capsids adopt flexible structures, the capsids become significantly more compact and rigid upon cleavage of the disordered tail (Fig. 3A and B) (5, 32, 33). EM also reveals that the diameter of the capsid decreases from 200 to 190 Å and that the pitch decreases from 57.2 to 48.7 Å upon removal of N_{TAIL} (34).

The atomic resolution structure of N_{CORE} of MeV is unknown, however, the structure of the N-RNA complex of Respiratory Syncytial Virus (RSV), another member of the *Paramyxoviridae* family, was recently solved using X-ray crystallography (35), and docked into the EM density map of MeV N-RNA on the basis of secondary structural homology (34). Notably, this coarse docking places the C terminus of N_{CORE} , and therefore the N terminus of N_{TAIL} , at the interior of the helix capsid, raising intriguing questions about the position of N_{TAIL} within the capsid. Due to steric hindrance, the 13 copies of N_{TAIL} per turn of the capsid helix cannot reside in the interior of the capsid and remain flexible enough to give rise to NMR signals. We have therefore investigated whether the disordered N_{TAIL} can escape from the interior of the MeV nucleocapsid helix, as reconstructed using EM, by building explicit models that obey random coil statistics for the conformational sampling of the primary sequence, while avoiding the N_{CORE} domains in the capsid. This model (Fig. 5) demonstrates that N_{TAIL} can indeed exfiltrate from the interior of the capsid via the interstitial space between the N_{CORE} moieties. Importantly, reorientational sampling of the chain calculated over the entire ensemble (Fig. 4C), demonstrates that maximal angular freedom is only achieved after approximately 50 amino acids, providing a reasonable explanation for the lack of solution NMR signals up to residue 450. In this case the first 50 amino acids of N_{TAIL} retain conformational disorder, which would also explain why they could not be resolved in the EM reconstruction of the capsids (34).

Small Angle Scattering Confirms Transient Binding of N_{TAIL} MoRE to Capsid Surface. Small angle X-ray and neutron scattering (SAS) provides important information concerning the dimensions of N_{TAIL} in intact nucleocapsids. Despite significant polydispersity in terms of length, the cross-sectional radii of gyration, R_C , of the capsids can be accurately determined from these data. SAS analysis of the scattering length density distribution around the nucleocapsid symmetry axis gives R_C values of (78.0 ± 0.6) Å and (69.5 ± 2.4) Å for the intact and cleaved forms respectively (Fig. 6, Fig. S3). The expected value of R_C calculated from the atomic coordinates of RSV N-RNA docked into the reconstructed electron density of the cleaved MeV capsid gives very

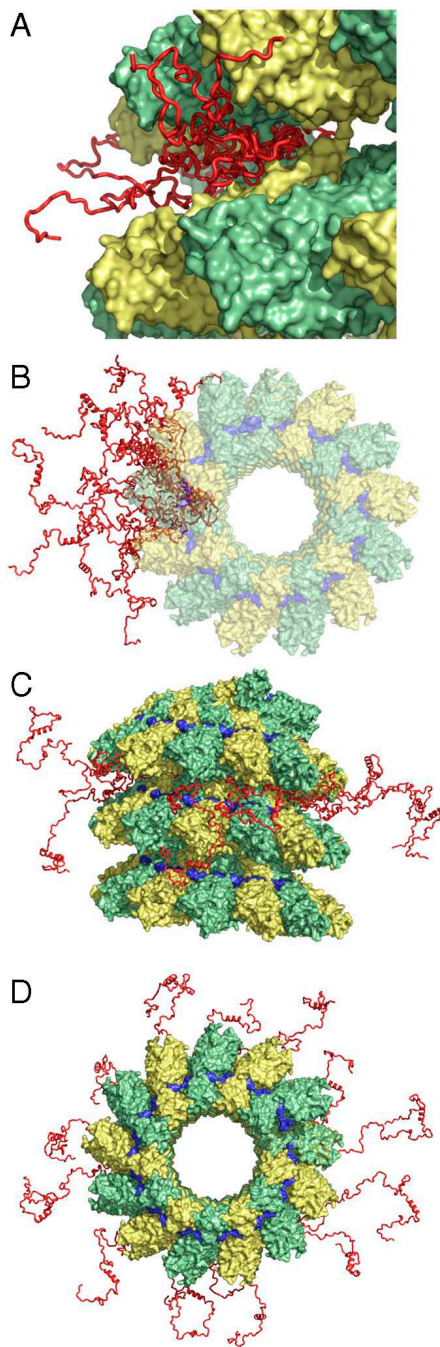


Fig. 5. Proposed model of the location of N_{TAIL} in intact nucleocapsids. The three-dimensional coordinates of the RSV N-RNA subunit docked into the EM density map of MeV N-RNA were used (34). The conformational sampling algorithm *flexible-meccano* was used to build chains from the C terminus of the folded domain of N_{CORE} (successive N_{CORE} monomers are coloured green and yellow). Amino-acid specific conformational sampling allows the chain to escape from the interstitial space of the capsid helix. (A) Representation of the conformational sampling of N_{TAIL} from a single N protein in the capsid. Different copies of N_{TAIL} (red) are shown to indicate the available volume sampling of the chain. The first 50 amino acids of N_{TAIL} are shown. (B) Representation of the conformational sampling of N_{TAIL} from a single N protein in the capsid, shown along the axis of the nucleocapsid. (C, D) Representation of the 13 N_{TAIL} conformers from a single turn of the nucleocapsid. In the interests of clarity, (B–D) deliberately show more conformers outside the capsid, and fewer bound to the surface, than are probable at any one time (see text). The position of the RNA is shown in blue.

good agreement with experiment (68.0 Å), while the calculated model of the capsid with the full-length chain gives a value of 83.8 Å when the MoRE is entirely free, and 78.4 Å when the

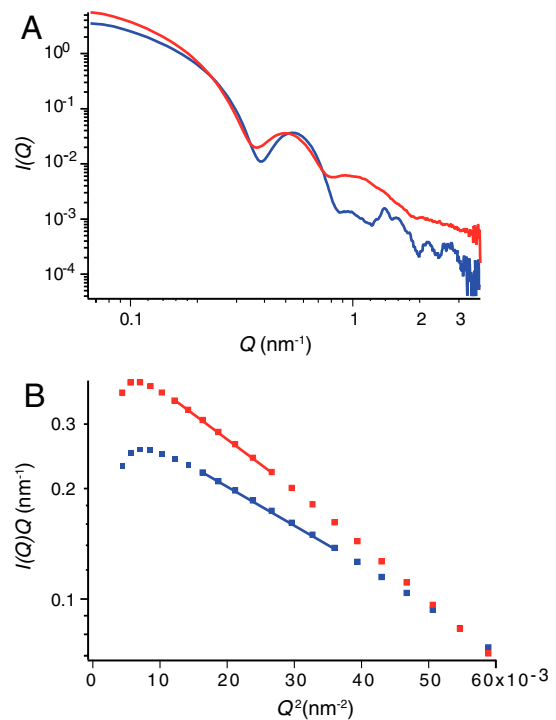


Fig. 6. Small angle X-ray scattering of nucleocapsids. (A) Data from intact full-length nucleocapsids (red) and N_{TAIL} -cleaved nucleocapsids (blue). (B) Linear fits of $\ln[I(Q)Q] = \ln[I(0)Q] - \frac{1}{2}R_C^2 Q^2$ used to extract values of R_C from SAXS and SANS data (Fig. S3) from the cleaved (blue) and noncleaved (red) helical capsids.

center of the MoRE is positioned less than 8 Å from any of the folded domains of the capsid. The NMR-based model of a transient interaction between the MoRE and the capsid is therefore strongly supported by the SAS data. These results also provide a steric explanation for the observed decrease in pitch between intact and cleaved capsids (34), as parts of the disordered N_{TAIL} reside in the interstitial space between the N_{CORE} lobes.

Discussion

Measurements of NMR, SAS, and EM on nucleocapsids therefore provide the basis for the development of an in situ ensemble model describing the conformational behavior of N_{TAIL} in intact nucleocapsids. On the basis of this model we are able to provide a structural framework for understanding the dual role of the 125 amino acid intrinsically disordered N_{TAIL} domain. The first 50 disordered amino acids form an articulated spacer that allows the MoRE to escape from the interior of the capsid via the confined interstitial space between successive turns of the helix. The remainder of the chain, on the other hand, is more mobile, and retains the conformational sampling that exists in the isolated form of the protein. This sampling includes the conformational equilibrium of rapidly interconverting helical elements in the MoRE that is predefined by the primary sequence. At the same time the MoRE exchanges on and off the surface of the nucleocapsids, with the majority of conformers in contact with the capsid. The NMR and SAS data indicate that at any given time approximately one of the 13 copies of the nucleoprotein per helical turn is completely free in solution, while the remainder are bound to the capsid surface. While we currently have no information about the position of the binding site, or whether this binding is specific, such a mode of action would provide an efficient mechanism by which N_{TAIL} could “catch” the viral polymerase complex when in free solution, and colocalize the complex on the nucleocapsid surface, thereby initiating transcription and replication of the viral RNA. Interestingly the RNA is sequestered on the outer surface of the RSV and MeV capsids (34, 35), which

would place the RNA in the immediate vicinity of N_{TAIL} as it emerges from the interstitial space, providing a mechanistic rationalization of the entire disordered domain of the nucleocapsid. Further structural and dynamic information will be necessary in order to determine the subsequent sequence of events that follow this initial recognition step, and ultimately lead to transcription and replication.

Methods

Cloning, expression, and purification of the isotopically labeled isolated MeV N_{TAIL} domain and the C-terminal domain of P (XD) were described previously (13). Cloning, expression, and purification procedures for MeV nucleoproteins are described in *SI Text* (34). Random cellular RNA forms the basis of the reconstituted nucleocapsids which are therefore of variable length.

NMR Experiments. All NMR experiments were carried out at 25 °C. For the measurement of RDCs ^{13}C , ^{15}N labeled N_{TAIL} was aligned in a liquid crystal composed of poly-ethylene glycol (PEG) and 1-hexanol (36) giving rise to a residual deuterium splitting of 21 Hz. $^1D_{N-HN}$, $^1D_{C-C}$ and $^1D_{C-His}$ RDCs were obtained using 3D BEST-type HNCO and HN(CO)CA experiments modified to allow for coupling evolution in the ^{13}C dimension (37). Spectra were acquired with a sweep width of 7.5 kHz and 512 complex points in the 1H dimension and a sweep width of 1.32 kHz and 36 complex points in the ^{15}N dimension. For the ^{13}C dimension, the spectra were acquired with a sweep width of 1.2 kHz and 60 complex points (HNCO-type spectra) and 3 kHz and 60 complex points [HN(CO)CA-type spectrum]. Estimates of experimental errors on the RDCs were obtained through repeated measurements: 1.0 Hz ($^1D_{N-HN}$), 2.0 Hz ($^1D_{C-His}$) and 0.5 Hz ($^1D_{C-C}$). Spectra were processed in NMRPipe (38) and analyzed using Sparky (39) and CCPN (40).

The complex between N_{TAIL} and XD was obtained by preparing a sample containing 0.14 mM ^{15}N , ^{13}C N_{TAIL} and 1.4 mM unlabeled XD. The complex was aligned in a liquid crystal composed of PEG and 1-hexanol giving rise to a residual deuterium splitting of 26 Hz. $^1D_{N-HN}$ were obtained for N_{TAIL} in the complex using a 2D IPAP SOFAST-HMQC (41) experiment containing 1,024 complex points in the 1H dimension and 150 complex points in the ^{15}N dimension. All RDCs (free and bound form of N_{TAIL}) were measured at a 1H resonance frequency of 600 MHz. ^{15}N R_2 relaxation rates of N_{TAIL} in its free form and in the context of intact nucleocapsids were measured at a 1H frequency of 1,000 MHz. Standard pulse sequences were used and the spectra were recorded with a sweep width of 14 kHz and 1,024 complex points in the 1H dimension and a sweep width of 3 kHz and 100 complex points in the ^{15}N dimension (42).

Asteroids Description of the Molecular Recognition Element of N_{TAIL} from NMR Data. Experimental RDCs and $C\alpha$ chemical shifts were used in a combined approach to obtain an ensemble description of the MoRE of N_{TAIL} using the minimal ensemble approach (19). A representative ensemble description of the N_{TAIL} MoRE (defined between residues 485–502) was obtained by generating ensembles of N_{TAIL} each consisting of 10,000 conformers using *flexible-meccano* (14) with varying helix lengths and positions within the MoRE. One hundred and twenty different ensembles were created to cover the entire MoRE with helices with a minimum length of four residues and a maximum length of 18 residues. Furthermore, an ensemble without helices comprising 50,000 conformers was generated. The alignment tensor of each conformer in the ensembles was calculated using PALES (43, 44) and ensemble-averaged RDCs were obtained for each of the 121 ensembles. Ensemble-averaged chemical shifts were calculated using SPARTA (45) using 1,000 conformers, except for the completely unfolded ensemble where 5,000 conformers were used.

The number of helices, N , necessary to describe the experimental data, and the position and length of the helices, were determined by incrementing N . For each step, the genetic algorithm ASTEROIDS (16) was used to select N helical ensembles and their associated populations such that the predicted population weighted RDCs (Fig. 1B, Fig. S1) and chemical shifts (Fig. 1C) were in agreement with the experimental values using:

$$O_{CALC} = \sum_{k=1}^N p_k O_k + (1 - \sum_{k=1}^N p_k) O_U. \quad [1]$$

O_k and O_U are the simulated ensemble-averaged observables for the k th helical and unfolded ensemble, respectively, and p_k is the population associated with the k th ensemble. A χ^2 function is calculated over all residues of the MoRE. Model selection is achieved by optimization of the population and a scaling factor for the RDCs corresponding to the degree of alignment

(Table S2). Experimental $C\alpha$ chemical shift uncertainty used in the combined target function was estimated as 0.3 ppm. The ASTEROIDS selection used 2,000 successive generations and was repeated 10 times for each run to ensure a well defined solution for each value of N (16). Standard F-statistics were used to test the significance of one model over the other (Table S1).

Modelling of N_{TAIL} in the Context of the Capsid. N_{TAIL} was built onto the atomic resolution model of N_{CORE} derived from docking of the RSV N_{CORE} structure into the EM density of MeV N-RNA capsids. Disordered N_{TAIL} conformers were built using the *flexible-meccano* algorithm that sequentially constructs peptide chains by randomly sampling amino acid specific dihedral angle distributions (14). Steric clashes are avoided with the folded domains of all copies of N_{CORE} in the capsid. Angular order parameters relative to the capsid frame were calculated over 5,000 conformers in a single ensemble of independent copies of N_{TAIL} from the same N protein as described (46).

Small Angle X-Ray Scattering. SAXS experiments were carried out on intact and trypsin-digested nucleocapsids at concentrations of 0.35 mM (intact capsids) and 0.25 mM (digested capsids). All sample volumes were adjusted to 50 μ L, and were measured on the high brilliance diffracted angle distributions (14). Steric clashes are avoided with the folded domains of all copies of N_{CORE} in the capsid. Angular order parameters relative to the capsid frame were calculated over 5,000 conformers in a single ensemble of independent copies of N_{TAIL} from the same N protein as described (46).

Small Angle Neutron Scattering. Small Angle Neutron Scattering (SANS) experiments were carried out on intact and trypsin-digested nucleocapsids at concentrations of 0.35 mM (intact capsids) and 0.05 mM (digested capsids). All sample volumes were adjusted to 200 μ L and were measured on the instrument D22 at the Institute Laue-Langevin (ILL) (Grenoble, France) in Hellma® quartz cuvettes 100QS with 1 mm optical path length. Scattering data were recorded at instrumental configurations (collimator/detector) 2 m/2 m, 8 m/8 m, and 17.6 m/17.6 m at a neutron wavelength $\lambda = 6$ Å. At each configuration, the samples, the buffers, the empty beam, an empty quartz cuvette, as well as a boron sample (electronic background) were measured. Exposure times varied from 30 min to 3 h according to sample and collimator/detector setup. Transmissions were measured during 2 min for each sample. Raw data were reduced using a standard ILL software package (48), normalized to an absolute scale after the various detector corrections and azimuthally averaged to obtain the one-dimensional scattering curve.

Calculation of Cross-Sectional Radius of Gyration from Scattering Data. Assuming capsid structures (hollow cylinders with an overall length much larger than the diameter), scattering curves were analyzed in terms of rod-like shaped particles. Cross-sectional radii of gyration, R_C , were extracted from linear fits of SAXS and SANS data according to (49):

$$\ln[I(Q)Q] = \ln[I(0)Q] - \frac{1}{2}R_C^2Q^2. \quad [2]$$

$I(0)$ is the cross-sectional part of the scattering. The range of validity of the approximation was reasonably fulfilled in both cases (intact form: $0.92 \leq R_C Q \leq 1.43$; cleaved form: $0.79 \leq R_C Q \leq 1.22$).

The experimentally determined cross-sectional radii of gyration (Eq. 2) were compared to the ones calculated from the atomic resolution structure of RSV docked into the EM density map of MeV N-RNA using the radial coordinates r_i of the N atoms in a unit sectorial element around the cylindrical axis of symmetry:

$$R_C^2 = \frac{1}{N} \sum_i r_i^2. \quad [3]$$

ACKNOWLEDGMENTS. We thank Dr. Phil Callow (ILL) for providing local support on the instrument D22, Dr. Theyencheri Narayanan for support using the ID02 beamline at the ESRF, and Stéphanie Costanzo for preparation of the isolated N_{TAIL} domain and the C-terminal domain of P. We would like to thank the Partnership for Structural Biology for access to biochemical and biophysical platforms, and Irina Gutsche for discussions. This work was supported by the Agence Nationale de la Recherche through

ANR PCV 2007 ProteinMotion (M.B.), ANR "Physico-Chimie du Vivant" ANR-08-PCVI-0020-01 (S.L.), ANR "Jeunes Chercheuses Jeunes Chercheurs" 2010 ProteinDisorder (M.R.J.), and the Finovi foundation of Lyon (M.B. and R.W.H.R.). M.R.J. benefited from a long-term EMBO fellowship and support

from Lundbeckfonden, Denmark. The authors acknowledge use of the 1 GHz spectrometer installed at the French high field NMR center in Lyon through the Très Grandes Infrastructures de Recherche, and acknowledge access to CERM research infrastructure (EC contracts 026145 and 261863).

- Curran J, Kolakofsky D (1999) Replication of paramyxoviruses. *Adv Virus Res* 54:403–422.
- Kingston RL, Baese WA, Gay LS (2004) Characterization of nucleocapsid binding by the measles virus and mumps virus phosphoproteins. *J Virol* 78:8630–8640.
- Curran J, et al. (1993) The hypervariable C-terminal tail of the Sendai paramyxovirus nucleocapsid protein is required for template function but not for RNA encapsidation. *J Virol* 67:4358–4364.
- Johansson K, et al. (2003) Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein. *J Biol Chem* 278:44567–44573.
- Longhi S, et al. (2003) The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. *J Biol Chem* 278:18638–18648.
- Bourhis J, et al. (2004) The C-terminal domain of measles virus nucleoprotein belongs to the class of intrinsically disordered proteins that fold upon binding to their physiological partner. *Virus Res* 99:157–167.
- Dunker AK, et al. (2002) Intrinsic disorder and protein function. *Biochemistry* 41:6573–6582.
- Tompa P (2002) Intrinsically unstructured proteins. *Trends Biochem Sci* 27:527–533.
- Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6:197–208.
- Tompa P, Fuxreiter M (2008) Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci* 33:2–8.
- Sugase K, Dyson HJ, Wright PE (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447:1021–1025.
- Bracken C, Iakoucheva LM, Romero PR, Dunker AK (2004) Combining prediction, computation and experiment for the characterization of protein disorder. *Curr Opin Struct Biol* 14:570–576.
- Gely S, et al. (2010) Solution structure of the C-terminal X domain of the measles virus phosphoprotein and interaction with the intrinsically disordered C-terminal domain of the nucleoprotein. *J Mol Recognit* 23:435–447.
- Bernadó P, et al. (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle X-ray scattering. *Proc Natl Acad Sci USA* 102:17002–17007.
- Jensen MR, et al. (2009) Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. *Structure* 17:1169–1185.
- Nodet G, et al. (2009) Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *J Am Chem Soc* 131:17908–17918.
- Marsh JA, Singh VK, Jia Z, Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci* 15:2795–2804.
- Jensen MR, Salmon L, Nodet G, Blackledge M (2010) Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. *J Am Chem Soc* 132:1270–1272.
- Jensen MR, et al. (2008) Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. *J Am Chem Soc* 130:8055–8061.
- Jensen MR, Blackledge M (2008) On the origin of NMR dipolar waves in transient helical elements of partially folded proteins. *J Am Chem Soc* 130:11266–11267.
- Serrano L, Fersht AR (1989) Capping and alpha-helix stability. *Nature* 342:296–299.
- Serrano L, Sancho J, Hirshberg M, Fersht AR (1992) Alpha-helix stability in proteins. I. Empirical correlations concerning substitution of side-chains at the N and C-caps and the replacement of alanine by glycine or serine at solvent-exposed surfaces. *J Mol Biol* 227:544–559.
- Mukrasch MD, et al. (2007) Highly populated turn conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation. *J Am Chem Soc* 129:5235–5243.
- Wells M, et al. (2008) Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc Natl Acad Sci USA* 105:5762–5767.
- Luisi DL, Wu WJ, Raleigh DP (1999) Conformational analysis of a set of peptides corresponding to the entire primary sequence of the N-terminal domain of the ribosomal protein L9: evidence for stable native-like secondary structure in the unfolded state. *J Mol Biol* 287:395–407.
- Kingston RL, Hamel DJ, Gay LS, Dahlquist FW, Matthews BW (2004) Structural basis for the attachment of a paramyxoviral polymerase to its template. *Proc Natl Acad Sci USA* 101:8301–8306.
- Szymczynska B, Gan L, Johnson J, Williamson J (2007) Solution NMR studies of the maturation intermediates of a 13 MDa viral capsid. *J Am Chem Soc* 129:7867–7876.
- Finch JT, Gibbs AJ (1970) Observations on the structure of the nucleocapsids of some paramyxoviruses. *J Gen Virol* 6:141–150.
- Lund GA, Tyrrell DL, Bradley RD, Scraba DG (1984) The molecular length of measles virus RNA and the structural organization of measles nucleocapsids. *J Gen Virol* 65:1535–1542.
- Fooks AR, et al. (1993) Measles virus nucleocapsid protein expressed in insect cells assembles into nucleocapsid-like structures. *J Gen Virol* 74:1439–1444.
- Bhella D, Ralph A, Murphy LB, Yeo RP (2002) Significant differences in nucleocapsid morphology within the Paramyxoviridae. *J Gen Virol* 83:1831–1839.
- Schoehn G, et al. (2004) The 12 A structure of trypsin-treated measles virus N-RNA. *J Mol Biol* 339:301–312.
- Bhella D, Ralph A, Yeo RP (2004) Conformational flexibility in recombinant measles virus nucleocapsids visualised by cryo-negative stain electron microscopy and real-space helical reconstruction. *J Mol Biol* 340:319–331.
- Desfosses A, Goret G, Estrozi LF, Ruigrok RWH, Gutsche I (2011) Nucleoprotein-RNA orientation in the measles virus nucleocapsid by three-dimensional electron microscopy. *J Virol* 85:1391–1395.
- Tawar RG, et al. (2009) Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus. *Science* 326:1279–1283.
- Rückert M, Otting G (2000) Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J Am Chem Soc* 122:7793–7797.
- Lescop E, Schanda P, Brutscher B (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J Magn Reson* 187:163–169.
- Delaglio F, et al. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277–293.
- Goddard T, Kneller D SPARKY 3 (University of California, San Francisco).
- Vranken WF, et al. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59:687–696.
- Kern T, Schanda P, Brutscher B (2008) Sensitivity-enhanced IPAP-SOFAST-HMQC for fast-pulsing 2D NMR with reduced radiofrequency load. *J Magn Reson* 190:333–338.
- Farrow N, et al. (1994) Backbone dynamics of a free and a phosphopeptide-complexed SRC homology-2 domain studied by N-15 NMR relaxation. *Biochemistry* 33:5984–6003.
- Zweckstetter M (2008) NMR: prediction of molecular alignment from structure using the PALES software. *Nat Protoc* 3:679–690.
- Zweckstetter M, Bax A (2000) Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. *J Am Chem Soc* 122:3791–3792.
- Shen Y, Bax A (2007) Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *J Biomol NMR* 38:289–302.
- Markwick PRL, et al. (2009) Toward a unified representation of protein structural dynamics in solution. *J Am Chem Soc* 131:16968–16975.
- Sztucki M, Narayanan T (2007) Development of an ultra-small-angle X-ray scattering instrument for probing the microstructure and the dynamics of soft matter. *J App Crystallogr* 40:S459–S462.
- Gosh R, Egelhaaf S, Rennie A (2006) A computing guide for small-angle scattering experiments. Institute Laue Langevin internal report.
- Porod G (1982) *Small Angle X-ray Scattering*, eds O Glatter and O Kratky (Academic Press, New York).

Bibliographie

- [Albertini 2006] A. A. Albertini, A. K. Wernimont, T. Muziol, R. B. Ravelli, C. R. Clapier, G. Schoehn, W. Weissenhorn et R. W. Ruigrok. *Crystal structure of the rabies virus nucleoprotein-RNA complex*. *Science*, vol. 313, pages 360–363, Jul 2006. (Cité en pages 6 et 94.)
- [Andreani 1995] C. Andreani, A. Filabozzi, F. Menzinger, A. Desideri, A. Deriu et D. Di Cola. *Dynamics of hydrogen atoms in superoxide dismutase by quasielastic neutron scattering*. *Biophys. J.*, vol. 68, pages 2519–2523, Jun 1995. (Cité en page 89.)
- [Baltimore 1970] D. Baltimore, A. S. Huang et M. Stampfer. *Ribonucleic acid synthesis of vesicular stomatitis virus, II. An RNA polymerase in the virion*. *Proc. Natl. Acad. Sci. U.S.A.*, vol. 66, pages 572–576, Jun 1970. (Cité en page 93.)
- [Banerjee 1974] A. K. Banerjee, S. A. Moyer et D. P. Rhodes. *Studies on the in vitro adenylation of RNA by vesicular stomatitis virus*. *Virology*, vol. 61, pages 547–558, Oct 1974. (Cité en page 94.)
- [Banerjee 1987] A. K. Banerjee. *The transcription complex of vesicular stomatitis virus*. *Cell*, vol. 48, pages 363–364, Feb 1987. (Cité en page 8.)
- [Baud 2009] C. Baud, H. Hodak, E. Willery, H. Drobecq, C. Loch, M. Jamin et F. Jacob-Dubuisson. *Role of DegP for two-partner secretion in Bordetella*. *Mol. Microbiol.*, vol. 74, pages 315–329, Oct 2009. (Cité en pages 6 et 149.)
- [Baumann 1993] U. Baumann, S. Wu, K. M. Flaherty et D. B. McKay. *Three-dimensional structure of the alkaline protease of Pseudomonas aeruginosa : a two-domain protein with a calcium binding parallel beta roll motif*. *EMBO J.*, vol. 12, pages 3357–3364, Sep 1993. (Cité en page 55.)
- [Bée 1988] M. Bée. *Quasielastic neutron scattering : principles and applications in solid state chemistry, biology, and materials science*. Adam Hilger, 1988. (Cité en pages 11, 60 et 148.)
- [Benz 1992] I. Benz et M. A. Schmidt. *AIDA-I, the adhesin involved in diffuse adherence of the diarrhoeagenic Escherichia coli strain 2787 (O126 :H27), is synthesized via a precursor molecule*. *Mol. Microbiol.*, vol. 6, pages 1539–1546, Jun 1992. (Cité en page 2.)
- [Bernado 2007] E. Petoukhov M.V. Blackledge M. Svergun D.I. Bernado P. Mylonas. *Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering*. *Journal of the American Chemical Society*, vol. 129, no. 17, pages 5656–5664, 2007. (Cité en pages 17, 27, 30 et 96.)
- [Bernado 2009] Pau Bernado et Martin Blackledge. *A Self-Consistent Description of the Conformational Behavior of Chemically Denatured Proteins from NMR and Small Angle Scattering*. *Biophysical Journal*, vol. 97, no. 10, pages 2839 – 2845, 2009. (Cité en page 48.)

- [Bicout 2001] D. J. Bicout et G. Zaccai. *Protein flexibility from the dynamical transition : a force constant analysis*. Biophys. J., vol. 80, pages 1115–1123, Mar 2001. (Cit  en pages 88, 89 et 144.)
- [Blondel 2002] D. Blondel, T. Regad, N. Poisson, B. Pavie, F. Harper, P. P. Pandolfi, H. De The et M. K. Chelbi-Alix. *Rabies virus P and small P products interact directly with PML and reorganize PML nuclear bodies*. Oncogene, vol. 21, pages 7957–7970, Nov 2002. (Cit  en page 7.)
- [Brzozka 2005] K. Brzozka, S. Finke et K. K. Conzelmann. *Identification of the rabies virus alpha/beta interferon antagonist : phosphoprotein P interferes with phosphorylation of interferon regulatory factor 3*. J. Virol., vol. 79, pages 7673–7681, Jun 2005. (Cit  en page 94.)
- [Chenik 1994] M. Chenik, K. Chebli, Y. Gaudin et D. Blondel. *In vivo interaction of rabies virus phosphoprotein (P) and nucleoprotein (N) : existence of two N-binding sites on P protein*. J. Gen. Virol., vol. 75 (Pt 11), pages 2889–2896, Nov 1994. (Cit  en page 8.)
- [Chong 1993] L. D. Chong et J. K. Rose. *Membrane association of functional vesicular stomatitis virus matrix protein in vivo*. J. Virol., vol. 67, pages 407–414, Jan 1993. (Cit  en page 93.)
- [Clantin 2004] B. Clantin, H. Hodak, E. Willery, C. Locht, F. Jacob-Dubuisson et V. Villeret. *The crystal structure of filamentous hemagglutinin secretion domain and its implications for the two-partner secretion pathway*. Proc. Natl. Acad. Sci. U.S.A., vol. 101, pages 6194–6199, Apr 2004. (Cit  en pages 3, 4, 26, 45, 52, 57, 59, 62, 149 et 150.)
- [Clantin 2007] B. Clantin, A. S. Delattre, P. Rucktooa, N. Saint, A. C. Meli, C. Locht, F. Jacob-Dubuisson et V. Villeret. *Structure of the membrane protein FhaC : a member of the Omp85-TpsB transporter superfamily*. Science, vol. 317, pages 957–961, Aug 2007. (Cit  en pages 2, 5, 26, 62 et 149.)
- [Clantin 2010] B. Clantin, C. Leyrat, A. Wohlkonig, H. Hodak, E. d. e. A. Ribeiro, N. Martinez, C. Baud, C. Smet-Nocca, V. Villeret, F. Jacob-Dubuisson et M. Jamin. *Structure and plasticity of the peptidyl-prolyl isomerase Par27 of Bordetella pertussis revealed by X-ray diffraction and small-angle X-ray scattering*. J. Struct. Biol., vol. 169, pages 253–265, Mar 2010. (Cit  en pages 27 et 48.)
- [Clementi 2000a] C. Clementi, P. A. Jennings et J. N. Onuchic. *How native-state topology affects the folding of dihydrofolate reductase and interleukin-1beta*. Proc. Natl. Acad. Sci. U.S.A., vol. 97, pages 5871–5876, May 2000. (Cit  en page 22.)
- [Clementi 2000b] C. Clementi, H. Nymeyer et J. N. Onuchic. *Topological and energetic factors : what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding ? An investigation for small globular proteins*. J. Mol. Biol., vol. 298, pages 937–953, May 2000. (Cit  en pages 22, 62 et 96.)

- [David 1973] A. E. David. *Assembly of the vesicular stomatitis virus envelope : incorporation of viral polypeptides into the host plasma membrane*. J. Mol. Biol., vol. 76, pages 135–148, May 1973. (Cit  en page 93.)
- [Debye 1915] P. Debye. *Die konstitution des wasserstoff-molek ils*. Franz., 1915. (Cit  en page 13.)
- [Delattre 2011] A. S. Delattre, N. Saint, B. Clantin, E. Willery, G. Lippens, C. Loch, V. Villeret et F. Jacob-Dubuisson. *Substrate recognition by the POTRA domains of TpsB transporter PhaC*. Mol. Microbiol., vol. 81, pages 99–112, Jul 2011. (Cit  en page 5.)
- [Delmas 2010] O. Delmas, R. Assenberg, J. M. Grimes et H. Bourhy. *The structure of the nucleoprotein binding domain of lyssavirus phosphoprotein reveals a structural relationship between the N-RNA binding domains of Rhabdoviridae and Paramyxoviridae*. RNA Biol, vol. 7, pages 322–327, 2010. (Cit  en page 94.)
- [Ding 2004] H. Ding, T. J. Green et M. Luo. *Crystallization and preliminary X-ray analysis of a proteinase-K-resistant domain within the phosphoprotein of vesicular stomatitis virus (Indiana)*. Acta Crystallogr. D Biol. Crystallogr., vol. 60, pages 2087–2090, Nov 2004. (Cit  en page 8.)
- [Ding 2006] H. Ding, T. J. Green, S. Lu et M. Luo. *Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus*. J. Virol., vol. 80, pages 2808–2814, Mar 2006. (Cit  en pages 8, 94, 96, 139 et 150.)
- [Domenighini 1990] M. Domenighini, D. Relman, C. Capi u, S. Falkow, A. Prugnola, V. Scarlato et R. Rappuoli. *Genetic characterization of Bordetella pertussis filamentous haemagglutinin : a protein processed from an unusually large precursor*. Mol. Microbiol., vol. 4, pages 787–800, May 1990. (Cit  en page 3.)
- [Doster 1989] W. Doster, S. Cusack et W. Petry. *Dynamical transition of myoglobin revealed by inelastic neutron scattering*. Nature, vol. 337, pages 754–756, Feb 1989. (Cit  en pages 20 et 89.)
- [Dufrene 2002] Y. F. Dufrene. *Atomic force microscopy, a powerful tool in microbiology*. J. Bacteriol., vol. 184, no. 19, pages 5205–5213, Oct 2002. (Cit  en page 23.)
- [Ebert 2003] O. Ebert, K. Shinozaki, T. G. Huang, M. J. Savontaus, A. Garcia-Sastre et S. L. Woo. *Oncolytic vesicular stomatitis virus for treatment of orthotopic hepatocellular carcinoma in immune-competent rats*. Cancer Res., vol. 63, pages 3605–3611, Jul 2003. (Cit  en page 6.)
- [Emerson 1972] S. U. Emerson et R. R. Wagner. *Dissociation and reconstitution of the transcriptase and template activities of vesicular stomatitis B and T virions*. J. Virol., vol. 10, pages 297–309, Aug 1972. (Cit  en page 94.)
- [Emerson 1975] S. U. Emerson et Y. Yu. *Both NS and L proteins are required for in vitro RNA synthesis by vesicular stomatitis virus*. J. Virol., vol. 15, pages 1348–1356, Jun 1975. (Cit  en pages 6, 8, 94 et 140.)

- [Emerson 1987] S. U. Emerson et M. Schubert. *Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus*. Proc. Natl. Acad. Sci. U.S.A., vol. 84, pages 5655–5659, Aug 1987. (Cit  en page 8.)
- [Emsley 1996] P. Emsley, I. G. Charles, N. F. Fairweather et N. W. Isaacs. *Structure of Bordetella pertussis virulence factor P.69 pertactin*. Nature, vol. 381, pages 90–92, May 1996. (Cit  en page 56.)
- [Finke 2003] S. Finke, R. Mueller-Waldeck et K. K. Conzelmann. *Rabies virus matrix protein regulates the balance of virus transcription and replication*. J. Gen. Virol., vol. 84, pages 1613–1621, Jun 2003. (Cit  en page 93.)
- [Flory 1953] P.J. Flory. Principles of polymer chemistry. George Fisher Baker non-resident lectureship in chemistry at Cornell University. Cornell University Press, 1953. (Cit  en page 15.)
- [Franke 2009] Daniel Franke et Dmitri I. Svergun. *DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering*. Journal of Applied Crystallography, vol. 42, no. 2, pages 342–346, Apr 2009. (Cit  en pages 13, 30 et 48.)
- [Freddolino 2006] Peter L. Freddolino, Anton S. Arkhipov, Steven B. Larson, Alexander McPherson et Klaus Schulten. *Molecular Dynamics Simulations of the Complete Satellite Tobacco Mosaic Virus*. Structure, vol. 14, no. 3, pages 437 – 449, 2006. (Cit  en page 21.)
- [Gabel 2002] F. Gabel, D. Bicout, U. Lehnert, M. Tehei, M. Weik et G. Zaccai. *Protein dynamics studied by neutron scattering*. Q. Rev. Biophys., vol. 35, pages 327–367, Nov 2002. (Cit  en page 21.)
- [Gaspar 2008] A. M. Gaspar, M. S. Appavou, S. Busch, T. Unruh et W. Doster. *Dynamics of well-folded and natively disordered proteins in solution : a time-of-flight neutron scattering study*. Eur. Biophys. J., vol. 37, pages 573–582, Jun 2008. (Cit  en pages 139 et 144.)
- [Gerard 2007] F. C. Gerard, E. d. e. A. Ribeiro, A. A. Albertini, I. Gutsche, G. Zaccai, R. W. Ruigrok et M. Jamin. *Unphosphorylated rhabdoviridae phosphoproteins form elongated dimers in solution*. Biochemistry, vol. 46, pages 10328–10338, Sep 2007. (Cit  en page 140.)
- [Gerard 2009] F. C. Gerard, E. d. e. A. Ribeiro, C. Leyrat, I. Ivanov, D. Blondel, S. Longhi, R. W. Ruigrok et M. Jamin. *Modular organization of rabies virus phosphoprotein*. J. Mol. Biol., vol. 388, pages 978–996, May 2009. (Cit  en pages 8, 94, 102, 139, 140, 148 et 150.)
- [Go 1983] N. Go. *Theoretical studies of protein folding*. Annu. Rev. Biophys. Bioeng., vol. 12, pages 183–210, 1983. (Cit  en page 22.)
- [Grass 2000] S. Grass et J. W. St Geme. *Maturation and secretion of the non-typable Haemophilus influenzae HMW1 adhesin : roles of the N-terminal and C-terminal domains*. Mol. Microbiol., vol. 36, pages 55–67, Apr 2000. (Cit  en page 26.)

- [Green 2000] T. J. Green, S. Macpherson, S. Qiu, J. Lebowitz, G. W. Wertz et M. Luo. *Study of the assembly of vesicular stomatitis virus N protein : role of the P protein*. J. Virol., vol. 74, pages 9515–9524, Oct 2000. (Cit  en page 8.)
- [Green 2006] T. J. Green, X. Zhang, G. W. Wertz et M. Luo. *Structure of the vesicular stomatitis virus nucleoprotein-RNA complex*. Science, vol. 313, pages 357–360, Jul 2006. (Cit  en pages 8, 94 et 96.)
- [Green 2009] T. J. Green et M. Luo. *Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P*. Proc. Natl. Acad. Sci. U.S.A., vol. 106, pages 11713–11718, Jul 2009. (Cit  en page 140.)
- [Grubmuller 1996] H. Grubmuller, B. Heymann et P. Tavan. *Ligand binding : molecular mechanics calculation of the streptavidin-biotin rupture force*. Science, vol. 271, pages 997–999, Feb 1996. (Cit  en page 22.)
- [Guedin 1998] S. Guedin, E. Willery, C. Locht et F. Jacob-Dubuisson. *Evidence that a globular conformation is not compatible with FhaC-mediated secretion of the Bordetella pertussis filamentous haemagglutinin*. Mol. Microbiol., vol. 29, pages 763–774, Aug 1998. (Cit  en pages 26 et 91.)
- [Guinier 1939] A. Guinier. *La diffraction des rayons x aux tr s petits angles : application   l’ tude de ph nom nes ultramicroscopiques*. Masson, 1939. (Cit  en page 29.)
- [Guinier 1955] A. Guinier et G. Fournet. *Small-angle scattering of x-rays. Structure of matter series*. Wiley, 1955. (Cit  en page 15.)
- [Gunasekaran 2003] K. Gunasekaran, C. J. Tsai, S. Kumar, D. Zanuy et R. Nussinov. *Extended disordered proteins : targeting function with less scaffold*. Trends Biochem. Sci., vol. 28, pages 81–85, Feb 2003. (Cit  en page 139.)
- [Gupta 1997] A. K. Gupta et A. K. Banerjee. *Expression and purification of vesicular stomatitis virus N-P complex from Escherichia coli : role in genome RNA transcription and replication in vitro*. J. Virol., vol. 71, pages 4264–4271, Jun 1997. (Cit  en page 94.)
- [Halter 1984] R. Halter, J. Pohlner et T. F. Meyer. *IgA protease of Neisseria gonorrhoeae : isolation and characterization of the gene and its extracellular product*. EMBO J., vol. 3, pages 1595–1601, Jul 1984. (Cit  en pages 25 et 26.)
- [Hess 2008] Berk Hess, Carsten Kutzner, David van der Spoel et Erik Lindahl. *GROMACS 4 : Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation*. Journal of Chemical Theory and Computation, vol. 4, no. 3, pages 435–447, 2008. (Cit  en page 21.)
- [Hirono 1997] I. Hirono, N. Tange et T. Aoki. *Iron-regulated haemolysin gene from Edwardsiella tarda*. Mol. Microbiol., vol. 24, pages 851–856, May 1997. (Cit  en page 2.)

- [Hodak 2006] H. Hodak, B. Clantin, E. Willery, V. Villeret, C. Locht et F. Jacob-Dubuisson. *Secretion signal of the filamentous haemagglutinin, a model two-partner secretion substrate*. Mol. Microbiol., vol. 61, pages 368–382, Jul 2006. (Cité en page 62.)
- [Hodak 2008] H. Hodak, A. Wohlkonig, C. Smet-Nocca, H. Drobecq, J. M. Wieruszki, M. Senechal, I. Landrieu, C. Locht, M. Jamin et F. Jacob-Dubuisson. *The peptidyl-prolyl isomerase and chaperone Par27 of Bordetella pertussis as the prototype for a new group of parvulins*. J. Mol. Biol., vol. 376, pages 414–426, Feb 2008. (Cité en pages 6, 27, 28, 31, 48, 52 et 149.)
- [Howatson 1962] A. F. Howatson et G. F. Whitmore. *The development and structure of vesicular stomatitis virus*. Virology, vol. 16, pages 466–478, Apr 1962. (Cité en page 93.)
- [Hunt 1976] L. A. Hunt et D. F. Summers. *Association of vesicular stomatitis virus proteins with HeLa cell membranes and released virus*. J. Virol., vol. 20, pages 637–645, Dec 1976. (Cité en page 8.)
- [Ivanov 2010] I. Ivanov, T. Crepin, M. Jamin et R. W. Ruigrok. *Structure of the dimerization domain of the rabies virus phosphoprotein*. J. Virol., vol. 84, pages 3707–3710, Apr 2010. (Cité en page 94.)
- [Jackson 2005] A. O. Jackson, R. G. Dietzgen, M. M. Goodin, J. N. Bragg et M. Deng. *Biology of plant rhabdoviruses*. Annu Rev Phytopathol, vol. 43, pages 623–660, 2005. (Cité en page 102.)
- [Jacob-Dubuisson 1999] F. Jacob-Dubuisson, C. El-Hamel, N. Saint, S. Guedin, E. Willery, G. Molle et C. Locht. *Channel formation by FhaC, the outer membrane protein involved in the secretion of the Bordetella pertussis filamentous hemagglutinin*. J. Biol. Chem., vol. 274, pages 37731–37735, Dec 1999. (Cité en pages 5 et 26.)
- [Jacob-Dubuisson 2004] F. Jacob-Dubuisson, R. Fernandez et L. Coutte. *Protein secretion through autotransporter and two-partner pathways*. Biochim. Biophys. Acta, vol. 1694, pages 235–257, Nov 2004. (Cité en page 2.)
- [Jacob 2000] Y. Jacob, H. Badrane, P. E. Ceccaldi et N. Tordo. *Cytoplasmic dynein LC8 interacts with lyssavirus phosphoprotein*. J. Virol., vol. 74, pages 10217–10222, Nov 2000. (Cité en page 94.)
- [Jacrot 1981] Bernard Jacrot et Giuseppe Zaccai. *Determination of molecular weight by neutron scattering*. Biopolymers, vol. 20, no. 11, pages 2413–2426, 1981. (Cité en page 15.)
- [Jamros 2010] M. A. Jamros, L. C. Oliveira, P. C. Whitford, J. N. Onuchic, J. A. Adams, D. K. Blumenthal et P. A. Jennings. *Proteins at work : a combined small angle X-RAY scattering and theoretical determination of the multiple structures involved on the protein kinase functional landscape*. J. Biol. Chem., vol. 285, pages 36121–36128, Nov 2010. (Cité en page 96.)

- [Jasnin 2010] M. Jasnin, L. van Eijck, M. M. Koza, J. Peters, C. Laguri, H. Lortat-Jacob et G. Zaccai. *Dynamics of heparan sulfate explored by neutron scattering*. Phys Chem Chem Phys, vol. 12, pages 3360–3362, Apr 2010. (Cit  en page 91.)
- [Kajava 2001] A. V. Kajava, N. Cheng, R. Cleaver, M. Kessel, M. N. Simon, E. Willery, F. Jacob-Dubuisson, C. Loch et A. C. Steven. *Beta-helix model for the filamentous haemagglutinin adhesin of Bordetella pertussis and related bacterial secretory proteins*. Mol. Microbiol., vol. 42, pages 279–292, Oct 2001. (Cit  en pages 2, 3, 4 et 62.)
- [Kajava 2006] A. V. Kajava et A. C. Steven. *Beta-rolls, beta-helices, and other beta-solenoid proteins*. Adv. Protein Chem., vol. 73, pages 55–96, 2006. (Cit  en page 56.)
- [Klauser 1990] T. Klauser, J. Pohlner et T. F. Meyer. *Extracellular transport of cholera toxin B subunit using Neisseria IgA protease beta-domain : conformation-dependent outer membrane translocation*. EMBO J., vol. 9, pages 1991–1999, Jun 1990. (Cit  en page 2.)
- [Klauser 1992] T. Klauser, J. Pohlner et T. F. Meyer. *Selective extracellular release of cholera toxin B subunit by Escherichia coli : dissection of Neisseria Iga beta-mediated outer membrane transport*. EMBO J., vol. 11, pages 2327–2335, Jun 1992. (Cit  en page 26.)
- [Kobe 2000] B. Kobe et A. V. Kajava. *When protein folding is simplified to protein coiling : the continuum of solenoid protein structures*. Trends Biochem. Sci., vol. 25, pages 509–515, Oct 2000. (Cit  en page 55.)
- [Koch 2003] M. H. Koch, P. Vachette et D. I. Svergun. *Small-angle scattering : a view on the properties, structures and structural changes of biological macromolecules in solution*. Q. Rev. Biophys., vol. 36, pages 147–227, May 2003. (Cit  en page 16.)
- [Kozin 2001] M. B. Kozin et D. I. Svergun. *Automated matching of high- and low-resolution structural models*. Journal of Applied Crystallography, vol. 34, no. 1, pages 33–41, Feb 2001. (Cit  en pages 30 et 48.)
- [Kratky 1982] O. Kratky et O. Glatter. *Small angle x-ray scattering / edited by o. glatter and o. kratky*. Academic Press, London ; New York :, 1982. (Cit  en pages 15 et 48.)
- [Krippahl 2003] L. Krippahl, J. J. Moura et P. N. Palma. *Modeling protein complexes with BiGGER*. Proteins, vol. 52, pages 19–23, Jul 2003. (Cit  en pages 30, 51 et 52.)
- [Kuhlman 2001] B. Kuhlman, J. W. O’Neill, D. E. Kim, K. Y. Zhang et D. Baker. *Conversion of monomeric protein L to an obligate dimer by computational protein design*. Proc. Natl. Acad. Sci. U.S.A., vol. 98, pages 10687–10691, Sep 2001. (Cit  en page 58.)

- [Kuhlman 2003] B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard et D. Baker. *Design of a novel globular protein fold with atomic-level accuracy*. *Science*, vol. 302, pages 1364–1368, Nov 2003. (Cité en page 58.)
- [Lambert-Buisine 1998] C. Lambert-Buisine, E. Willery, C. Locht et F. Jacob-Dubuisson. *N-terminal characterization of the Bordetella pertussis filamentous haemagglutinin*. *Mol. Microbiol.*, vol. 28, pages 1283–1293, Jun 1998. (Cité en pages 3 et 26.)
- [Lee 2010] W. Lee, X. Zeng, H. X. Zhou, V. Bennett, W. Yang et P. E. Marszalek. *Full reconstruction of a vectorial protein folding pathway by atomic force microscopy and molecular dynamics simulations*. *J. Biol. Chem.*, vol. 285, pages 38167–38172, Dec 2010. (Cité en page 24.)
- [Leyrat 2011a] Cédric Leyrat, Malene Ringkjøbing Jensen, Euripedes A. Ribeiro, Francine C. A. Gérard, Rob W. H. Ruigrok, Martin Blackledge et Marc Jamin. *The NO-binding region of the vesicular stomatitis virus phosphoprotein is globally disordered but contains transient α -helices*. *Protein Science*, vol. 20, no. 3, pages 542–556, 2011. (Cité en pages 8 et 140.)
- [Leyrat 2011b] Cédric Leyrat, Filip Yabukarski, Nicolas Tarbouriech, Euripedes A. Ribeiro Jr., Malene Ringkjøbing Jensen, Martin Blackledge, Rob W. H. Ruigrok et Marc Jamin. *Structure of the Vesicular Stomatitis Virus N^o-P Complex*. *PLoS Pathog*, vol. 7, no. 9, page e1002248, 09 2011. (Cité en pages 8, 94, 95, 140 et 150.)
- [Li 2007] M. S. Li. *Secondary structure, mechanical stability, and location of transition state of proteins*. *Biophys. J.*, vol. 93, pages 2644–2654, Oct 2007. (Cité en page 24.)
- [Lichty 2004] B. D. Lichty, A. T. Power, D. F. Stojdl et J. C. Bell. *Vesicular stomatitis virus : re-inventing the bullet*. *Trends Mol Med*, vol. 10, pages 210–216, May 2004. (Cité en page 6.)
- [Liu 1997] D. F. Liu, E. Phillips, T. M. Wizemann, M. M. Siegel, K. Tabei, J. L. Cowell et E. Tuomanen. *Characterization of a recombinant fragment that contains a carbohydrate recognition domain of the filamentous hemagglutinin*. *Infect. Immun.*, vol. 65, pages 3465–3468, Aug 1997. (Cité en page 3.)
- [Liu 2006] Y. Liu et B. Kuhlman. *RosettaDesign server for protein design*. *Nucleic acids research*, vol. 34, no. Web Server issue, Juillet 2006. (Cité en pages 58 et 60.)
- [Locht 1993] C. Locht, P. Bertin, F. D. Menozzi et G. Renaud. *The filamentous haemagglutinin, a multifaceted adhesion produced by virulent Bordetella spp*. *Mol. Microbiol.*, vol. 9, pages 653–660, Aug 1993. (Cité en page 60.)
- [Locht 2001] C. Locht, R. Antoine et F. Jacob-Dubuisson. *Bordetella pertussis, molecular pathogenesis under multiple aspects*. *Curr. Opin. Microbiol.*, vol. 4, pages 82–89, Feb 2001. (Cité en page 2.)

- [Lu 1999] H. Lu et K. Schulten. *Steered molecular dynamics simulations of force-induced protein domain unfolding*. *Proteins*, vol. 35, pages 453–463, Jun 1999. (Cit  en page 23.)
- [Makhov 1994] A. M. Makhov, J. H. Hannah, M. J. Brennan, B. L. Trus, E. Kocsis, J. F. Conway, P. T. Wingfield, M. N. Simon et A. C. Steven. *Filamentous hemagglutinin of Bordetella pertussis. A bacterial adhesin formed as a 50-nm monomeric rigid rod based on a 19-residue repeat motif rich in beta strands and turns*. *J. Mol. Biol.*, vol. 241, pages 110–124, Aug 1994. (Cit  en page 3.)
- [Matlin 1982] K. S. Matlin, H. Reggio, A. Helenius et K. Simons. *The entry of enveloped viruses into an epithelial cell line*. *Prog. Clin. Biol. Res.*, vol. 91, pages 599–611, 1982. (Cit  en pages 7 et 93.)
- [Maupetit 2006] J. Maupetit, R. Gautier et P. Tuffery. *SABBAC : online Structural Alphabet-based protein BackBone reconstruction from Alpha-Carbon trace*. *Nucleic Acids Res.*, vol. 34, pages W147–151, Jul 2006. (Cit  en page 30.)
- [Maurer 1999a] J. Maurer, J. Jose et T. F. Meyer. *Characterization of the essential transport function of the AIDA-I autotransporter and evidence supporting structural predictions*. *J. Bacteriol.*, vol. 181, pages 7014–7020, Nov 1999. (Cit  en page 2.)
- [Maurer 1999b] J. Maurer, J. Jose et T. F. Meyer. *Characterization of the essential transport function of the AIDA-I autotransporter and evidence supporting structural predictions*. *J. Bacteriol.*, vol. 181, pages 7014–7020, Nov 1999. (Cit  en page 26.)
- [Mavrakis 2003] M. Mavrakis, F. Iseni, C. Mazza, G. Schoehn, C. Ebel, M. Gentzel, T. Franz et R. W. Ruigrok. *Isolation and characterisation of the rabies virus N degrees-P complex produced in insect cells*. *Virology*, vol. 305, pages 406–414, Jan 2003. (Cit  en pages 7, 8, 94 et 140.)
- [Mavrakis 2004] M. Mavrakis, A. A. McCarthy, S. Roche, D. Blondel et R. W. Ruigrok. *Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus*. *J. Mol. Biol.*, vol. 343, pages 819–831, Oct 2004. (Cit  en pages 94, 102 et 103.)
- [Mavrakis 2006] M. Mavrakis, S. Mehoulas, E. Real, F. Iseni, D. Blondel, N. Tordo et R. W. Ruigrok. *Rabies virus chaperone : identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA*. *Virology*, vol. 349, pages 422–429, Jun 2006. (Cit  en page 8.)
- [McCammon 1977] J. A. McCammon, B. R. Gelin et M. Karplus. *Dynamics of folded proteins*. *Nature*, vol. 267, pages 585–590, Jun 1977. (Cit  en page 21.)
- [Meli 2006] A. C. Meli, H. Hodak, B. Clantin, C. Lochter, G. Molle, F. Jacob-Dubuisson et N. Saint. *Channel properties of TpsB transporter FhaC point to two functional domains with a C-terminal protein-conducting pore*. *J. Biol. Chem.*, vol. 281, pages 158–166, Jan 2006. (Cit  en page 3.)
- [Mohan 2006] A. Mohan, C. J. Oldfield, P. Radivojac, V. Vacic, M. S. Cortese, A. K. Dunker et V. N. Uversky. *Analysis of molecular recognition features*

- (*MoRFs*). J. Mol. Biol., vol. 362, pages 1043–1059, Oct 2006. (Cité en page 140.)
- [Newcomb 1981] W. W. Newcomb et J. C. Brown. *Role of the vesicular stomatitis virus matrix protein in maintaining the viral nucleocapsid in the condensed form found in native virions*. J. Virol., vol. 39, pages 295–299, Jul 1981. (Cité en page 8.)
- [Noel 2010] J. K. Noel, P. C. Whitford, K. Y. Sanbonmatsu et J. N. Onuchic. *SMOG@ctbp : simplified deployment of structure-based models in GRO-MACS*. Nucleic Acids Res., vol. 38, pages W657–661, Jul 2010. (Cité en page 96.)
- [Ohnishi 1994] Y. Ohnishi, M. Nishiyama, S. Horinouchi et T. Beppu. *Involvement of the COOH-terminal pro-sequence of Serratia marcescens serine protease in the folding of the mature enzyme*. J. Biol. Chem., vol. 269, pages 32800–32806, Dec 1994. (Cité en page 2.)
- [Oomen 2004] C. J. Oomen, P. van Ulsen, P. van Gelder, M. Feijen, J. Tommassen et P. Gros. *Structure of the translocator domain of a bacterial autotransporter*. EMBO J., vol. 23, pages 1257–1266, Mar 2004. (Cité en page 2.)
- [Osborne 2005] Andrew R. Osborne, Tom A. Rapoport et Bert van den Berg. *PROTEIN TRANSLOCATION BY THE SEC61/SECY CHANNEL*. Annual Review of Cell and Developmental Biology, vol. 21, no. 1, pages 529–550, 2005. (Cité en page 26.)
- [Palma 2000] P. Nuno Palma, Ludwig Krippahl, John E. Wampler et José J.G. Moura. *BiGGER : A new (soft) docking algorithm for predicting protein interactions*. Proteins : Structure, Function, and Bioinformatics, vol. 39, no. 4, pages 372–384, 2000. (Cité en page 52.)
- [Patton 1984] J. T. Patton, N. L. Davis et G. W. Wertz. *N protein alone satisfies the requirement for protein synthesis during RNA replication of vesicular stomatitis virus*. J. Virol., vol. 49, pages 303–309, Feb 1984. (Cité en page 8.)
- [Paul 1988] P. R. Paul, D. Chattopadhyay et A. K. Banerjee. *The functional domains of the phosphoprotein (NS) of vesicular stomatitis virus (Indiana serotype)*. Virology, vol. 166, pages 350–357, Oct 1988. (Cité en pages 8 et 94.)
- [Pettersen 2004] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng et T. E. Ferrin. *UCSF Chimera—a visualization system for exploratory research and analysis*. J Comput Chem, vol. 25, pages 1605–1612, Oct 2004. (Cité en page 29.)
- [Phillips 2005] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kalé et Klaus Schulten. *Scalable molecular dynamics with NAMD*. Journal of Computational Chemistry, vol. 26, no. 16, pages 1781–1802, 2005. (Cité en pages 21, 30, 51 et 59.)

- [Pickersgill 1998] R. Pickersgill, D. Smith, K. Worboys et J. Jenkins. *Crystal structure of polygalacturonase from Erwinia carotovora ssp. carotovora*. J. Biol. Chem., vol. 273, pages 24660–24664, Sep 1998. (Cité en pages 57, 58 et 60.)
- [Pohlner 1987] J. Pohlner, R. Halter, K. Beyreuther et T. F. Meyer. *Gene structure and extracellular secretion of Neisseria gonorrhoeae IgA protease*. Nature, vol. 325, pages 458–462, 1987. (Cité en page 2.)
- [Pugsley 1993] A. P. Pugsley. *The complete general secretory pathway in gram-negative bacteria*. Microbiol. Rev., vol. 57, pages 50–108, Mar 1993. (Cité en pages 1 et 26.)
- [Putnam 2007] C. D. Putnam, M. Hammel, G. L. Hura et J. A. Tainer. *X-ray solution scattering (SAXS) combined with crystallography and computation : defining accurate macromolecular structures, conformations and assemblies in solution*. Q. Rev. Biophys., vol. 40, pages 191–285, Aug 2007. (Cité en page 48.)
- [P.V.Konarev 2003] A.V.Sokolova M.H.J.Koch P.V.Konarev V.V.Volkov et D. I. Svergun. *PRIMUS - a Windows-PC based system for small-angle scattering data analysis*. Journal of Applied Crystallography, vol. 36, pages 1277–1282, 2003. (Cité en page 29.)
- [Relman 1989] D. A. Relman, M. Domenighini, E. Tuomanen, R. Rappuoli et S. Falkow. *Filamentous hemagglutinin of Bordetella pertussis : nucleotide sequence and crucial role in adherence*. Proc. Natl. Acad. Sci. U.S.A., vol. 86, pages 2637–2641, Apr 1989. (Cité en page 3.)
- [Renauld-Mongenie 1996] G. Renauld-Mongenie, J. Cornette, N. Mielcarek, F. D. Menozzi et C. Locht. *Distinct roles of the N-terminal and C-terminal precursor domains in the biogenesis of the Bordetella pertussis filamentous hemagglutinin*. J. Bacteriol., vol. 178, pages 1053–1060, Feb 1996. (Cité en page 2.)
- [Rhodes 1974] D. P. Rhodes, S. A. Moyer et A. K. Banerjee. *In vitro synthesis of methylated messenger RNA by the virion-associated RNA polymerase of vesicular stomatitis virus*. Cell, vol. 3, pages 327–333, Dec 1974. (Cité en page 93.)
- [Ribeiro 2008] E. A. Ribeiro, A. Favier, F. C. Gerard, C. Leyrat, B. Brutscher, D. Blondel, R. W. Ruigrok, M. Blackledge et M. Jamin. *Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein*. J. Mol. Biol., vol. 382, pages 525–538, Oct 2008. (Cité en pages 8, 94, 96, 102, 139, 140 et 150.)
- [Rief 1997] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez et H. E. Gaub. *Reversible unfolding of individual titin immunoglobulin domains by AFM*. Science, vol. 276, pages 1109–1112, May 1997. (Cité en pages 22, 23 et 62.)
- [Roediger 2008] E. K. Roediger, K. Kugathasan, X. Zhang, B. D. Lichty et Z. Xing. *Heterologous boosting of recombinant adenoviral prime immunization with*

- a novel vesicular stomatitis virus-vectored tuberculosis vaccine*. Mol. Ther., vol. 16, pages 1161–1169, Jun 2008. (Cit  en page 6.)
- [Rojas 2002] C. M. Rojas, J. H. Ham, W. L. Deng, J. J. Doyle et A. Collmer. *HecA, a member of a class of adhesins produced by diverse pathogenic bacteria, contributes to the attachment, aggregation, epidermal cell killing, and virulence phenotypes of Erwinia chrysanthemi EC16 on Nicotiana clevelandii seedlings*. Proc. Natl. Acad. Sci. U.S.A., vol. 99, pages 13142–13147, Oct 2002. (Cit  en page 2.)
- [Ryckaert 1977] J. Ryckaert, G. Ciccotti et H. Berendsen. *Numerical integration of the cartesian equations of motion of a system with constraints : molecular dynamics of n-alkanes*. Journal of Computational Physics, vol. 23, no. 3, pages 327–341, Mars 1977. (Cit  en page 30.)
- [Sato 1984] H. Sato et Y. Sato. *Bordetella pertussis infection in mice : correlation of specific antibodies against two antigens, pertussis toxin, and filamentous hemagglutinin with mouse protectivity in an intracerebral or aerosol challenge system*. Infect. Immun., vol. 46, pages 415–421, Nov 1984. (Cit  en page 3.)
- [Saul 2004] F. A. Saul, J. P. Arie, B. Vulliez-le Normand, R. Kahn, J. M. Betton et G. A. Bentley. *Structural and functional studies of FkpA from Escherichia coli, a cis/trans peptidyl-prolyl isomerase with chaperone activity*. J. Mol. Biol., vol. 335, pages 595–608, Jan 2004. (Cit  en pages 31 et 149.)
- [Schellman 1958] John A. Schellman. *The Factors Affecting the Stability of Hydrogen-bonded Polypeptide Structures in Solution*. The Journal of Physical Chemistry, vol. 62, no. 12, pages 1485–1494, 1958. (Cit  en page 48.)
- [Serdyuk 2007] I.N. Serdyuk, N.R. Zaccai et G. Zaccai. *Methods in molecular biophysics : structure, dynamics, function*. Cambridge University Press, 2007. (Cit  en page 16.)
- [Shaw 2010] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan et W. Wriggers. *Atomic-level characterization of the structural dynamics of proteins*. Science, vol. 330, pages 341–346, Oct 2010. (Cit  en page 21.)
- [Shi 2009] W. Shi, Q. Tang, X. Chen, P. Cheng, P. Jiang, X. Jing, X. Chen, P. Chen, Y. Wang, Y. Wei et Y. Wen. *Antitumor and antimetastatic activities of vesicular stomatitis virus matrix protein in a murine model of breast cancer*. J. Mol. Med., vol. 87, pages 493–506, May 2009. (Cit  en page 6.)
- [Simpson 1966] R. W. Simpson et R. E. Hauser. *Structural components of vesicular stomatitis virus*. Virology, vol. 29, pages 654–667, Aug 1966. (Cit  en page 6.)
- [Sokol 1969] F. Sokol, H. D. Schlumberger, T. J. Wiktor et H. Koprowski. *Biochemical and biophysical studies on the nucleocapsid and on the RNA of rabies virus*. Virology, vol. 38, pages 651–665, Aug 1969. (Cit  en page 94.)
- [Soria 1974] M. Soria, S. P. Little et A. S. Huang. *Characterization of vesicular stomatitis virus nucleocapsids. I. Complementary 40 S RNA molecules in nucleocapsids*. Virology, vol. 61, pages 270–280, Sep 1974. (Cit  en page 94.)

- [Svergun D.I. 1995] Barberato C. Svergun D.I. et Koch M.H.J. *CRY SOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates*. Journal of Applied Crystallography, vol. 28, pages 768–773, 1995. (Cité en page 13.)
- [Svergun 1998] D. I. Svergun, S. Richard, M. H. Koch, Z. Sayers, S. Kuprin et G. Zaccai. *Protein hydration in solution : experimental observation by x-ray and neutron scattering*. Proc. Natl. Acad. Sci. U.S.A., vol. 95, pages 2267–2272, Mar 1998. (Cité en page 96.)
- [Taketomi 1975] Hiroshi Taketomi, Yuzo Ueda et Nobuhiro G? *STUDIES ON PROTEIN FOLDING, UNFOLDING AND FLUCTUATIONS BY COMPUTER SIMULATION*. International Journal of Peptide and Protein Research, vol. 7, no. 6, pages 445–459, 1975. (Cité en page 22.)
- [Tehei 2001] M. Tehei, D. Madern, C. Pfister et G. Zaccai. *Fast dynamics of halophilic malate dehydrogenase and BSA measured by neutron scattering under various solvent conditions influencing protein stability*. Proc. Natl. Acad. Sci. U.S.A., vol. 98, pages 14356–14361, Dec 2001. (Cité en pages 91 et 148.)
- [Tomba 2002] P. Tompa. *Intrinsically unstructured proteins*. Trends Biochem. Sci., vol. 27, pages 527–533, Oct 2002. (Cité en pages 8 et 139.)
- [Toukmaji 1996] Abdalnour Toukmaji, Daniel Paul, John A. Board et Jr. *Distributed Particle-Mesh Ewald : A Parallel Ewald Summation Method*, 1996. (Cité en page 30.)
- [Tskhovrebova 1997] L. Tskhovrebova, J. Trinick, J. A. Sleep et R. M. Simmons. *Elasticity and unfolding of single molecules of the giant muscle protein titin*. Nature, vol. 387, pages 308–312, May 1997. (Cité en page 22.)
- [Tuomanen 1985] E. Tuomanen, A. Weiss, R. Rich, F. Zak et O. Zak. *Filamentous hemagglutinin and pertussis toxin promote adherence of Bordetella pertussis to cilia*. Dev. Biol. Stand., vol. 61, pages 197–204, 1985. (Cité en page 3.)
- [Volkov 2003] Vladimir V. Volkov et Dmitri I. Svergun. *Uniqueness of ab initio shape determination in small-angle scattering*. Journal of Applied Crystallography, vol. 36, no. 3 Part 1, pages 860–864, Jun 2003. (Cité en pages 30 et 48.)
- [Wall 1997] M. E. Wall, S. E. Ealick et S. M. Gruner. *Three-dimensional diffuse x-ray scattering from crystals of Staphylococcal nuclease*. Proc. Natl. Acad. Sci. U.S.A., vol. 94, pages 6180–6184, Jun 1997. (Cité en page 31.)
- [Wang 1997] M. D. Wang, H. Yin, R. Landick, J. Gelles et S. M. Block. *Stretching DNA with optical tweezers*. Biophys. J., vol. 72, pages 1335–1346, Mar 1997. (Cité en page 22.)
- [Ward 2004] J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton et D. T. Jones. *Prediction and functional analysis of native disorder in proteins from the three kingdoms of life*. J. Mol. Biol., vol. 337, pages 635–645, Mar 2004. (Cité en page 139.)

- [White 1981] J. White, K. Matlin et A. Helenius. *Cell fusion by Semliki Forest, influenza, and vesicular stomatitis viruses*. J. Cell Biol., vol. 89, pages 674–679, Jun 1981. (Cité en page 7.)
- [Whitt 1989] M. A. Whitt, L. Chong et J. K. Rose. *Glycoprotein cytoplasmic domain sequences required for rescue of a vesicular stomatitis virus glycoprotein mutant*. J. Virol., vol. 63, pages 3569–3578, Sep 1989. (Cité en page 8.)
- [Wohlkonig 2008] A. Wohlkonig, H. Hodak, B. Clantin, M. Senechal, C. Bompard, F. Jacob-Dubuisson et V. Villeret. *Crystallization and preliminary X-ray diffraction analysis of the peptidylprolyl isomerase Par27 of Bordetella pertussis*. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun., vol. 64, pages 809–812, Sep 2008. (Cité en page 31.)
- [Wu 2007] S. Wu et Y. Zhang. *LOMETS : a local meta-threading-server for protein structure prediction*. Nucleic Acids Res., vol. 35, pages 3375–3382, 2007. (Cité en page 4.)
- [Xu 2007] X. Xu, S. Wang, Y. X. Hu et D. B. McKay. *The periplasmic bacterial molecular chaperone SurA adapts its structure to bind peptides in different conformations to assert a sequence preference for aromatic residues*. J. Mol. Biol., vol. 373, pages 367–381, Oct 2007. (Cité en pages 31 et 149.)
- [Yeager 1976] M. J. Yeager. *Data analysis of low-angle diffraction by one-dimensional crystals*. Brookhaven Symp. Biol., pages VII77–VII106, May 1976. (Cité en page 15.)
- [Yen 2002] M. R. Yen, C. R. Peabody, S. M. Partovi, Y. Zhai, Y. H. Tseng et M. H. Saier. *Protein-translocating outer membrane porins of Gram-negative bacteria*. Biochim. Biophys. Acta, vol. 1562, pages 6–31, May 2002. (Cité en pages 1 et 25.)
- [Yoder 1993] M. D. Yoder, S. E. Lietzke et F. Jurnak. *Unusual structural features in the parallel beta-helix in pectate lyases*. Structure, vol. 1, pages 241–251, Dec 1993. (Cité en page 55.)
- [Zaccai, G. 2000] Zaccai, G., Tehei, M., Scherbakova, I., Serdyuk, I., Gerez, C. et Pfister, C. *Incoherent elastic neutron scattering as a function of temperature : A fast way to characterise in-situ biological dynamics in complex solutions*. J. Phys. IV France, vol. 10, pages Pr7–283–Pr7–287, 2000. (Cité en pages 89 et 148.)
- [Zaccai 2000] G. Zaccai. *How soft is a protein ? A protein dynamics force constant measured by neutron scattering*. Science, vol. 288, pages 1604–1607, Jun 2000. (Cité en pages 20 et 88.)
- [Zagrovic 2002] Bojan Zagrovic, Christopher D. Snow, Michael R. Shirts et Vijay S. Pande. *Simulation of Folding of a Small Alpha-helical Protein in Atomistic Detail using Worldwide-distributed Computing*. Journal of Molecular Biology, vol. 323, no. 5, pages 927 – 937, 2002. (Cité en page 21.)

Liste des Abréviations

ARN	Acide ribonucléique
AT	Autotransporteur
ATP	Adénosine tri-phosphate
BPTI	inhibiteur de la trypsine pancréatique bovine
BSA	Albumine sérique bovine
DegP	Protéine périplasmique ayant des activités chaperonne et protéase
FHA	Hemagglutinine filamenteuse
G	Glycoprotéine
GTP	guanosine triphosphate
L	ARN polymérase ARN dépendante
M	Protéine Matrice
MBP	Maltose Binding Protein, protéine possédant une affinité pour le maltose
N	Nucleoprotéine
N-ARN	Complexe nucléoprotéine/ARN
P	Phosphoprotéine
Par27	Protéine chaperonne périplasmique ayant une activité PPIase de la familles des Parvulines, de 27 kDa
PDB	Protein data Bank, banque de données centralisée contenant les coordonnées atomiques de toutes les structures déjà résolues
POTRA	Polypeptide transport associated domain
RGD	motif Arginine-Glycine-Aspartate
RMN	Résonance magnétique nucléaire
SANS	Diffusion de neutrons aux petits angles
SAXS	Diffusion de rayons X aux petits angles
Sec	Secretory pathway
SEC-MALLS-RI	Chromatographie d'exclusion de taille couplée à le diffusion de lumière statique et réfractométrie
TEV	Tobacco Etch Virus protein, protéase virale clivant une sequence en acides aminés spécifique
TPS	Secretion à deux partenaires
VSV	Virus de la Stomatite vésiculaire

Résumé :

L'objectif de la thèse était d'étudier les propriétés structurales et dynamiques de deux systèmes biologiques, avec une approche centrée sur les techniques de diffusion de neutrons.

Les premier système biologique porte sur le système de transport FHA/FhaC de la bactérie à Gram négatif *Bordetella pertussis*. C'est un modèle pour le système de transport bactérien à deux partenaires (TPS), qui est largement utilisé par de nombreuses bactéries à Gram négatif, essentiellement pour excréter des facteurs de virulence. Cette thèse comporte une analyse structurale sur le complexe chaperonne-client Fha30/Par27 mais aussi une étude sur les propriétés mécaniques de Fha30, qui possède une structure particulière en hélice β .

Le deuxième système porte sur le complexe de réplication des virus de la famille des *Rhabdoviridae*. Outre son importance en termes de santé humaine, le complexe de réplication comporte des protéines aux propriétés qui, sur le plan structural, sont très intéressantes, comme par exemple la phosphoprotéine qui comporte des parties structurées reliées par des parties dépourvues de structure. On trouvera dans cette thèse une étude des propriétés dynamiques des différents domaines de la phosphoprotéine, mais aussi une étude structurale sur le complexe que celle-ci forme avec la nucleoprotéine.

Abstract :

The main goal of this thesis is to study the dynamical and structural properties of two biological systems, with an approach centered on neutron scattering techniques. The first system consists of the transport system FhaB/FhaC from Gram negative bacteria *Bordetella pertussis*. It is a model system for the so called Two Partner Secretion (TPS) system, which is widespread among Gram negative bacteria to export virulence factors to the medium. This thesis emphasizes on the complex Fha30, a N-terminal fragment of FHA, forms with the periplasmic chaperone Par27, and on the mechanical properties of FHA, which has an uncanny β helix structure.

The other system is the replication complex of the *Rhabdoviridae* virus family. Apart from its importance in human health issues, the replication complex is composed of proteins with interesting structural and dynamical properties, specially the phosphoprotein which is composed of structured parts linked by flexible linkers. This thesis emphasizes on the dynamical properties of different fragments of the phosphoprotein, but also on the structural aspects of the complex the latter forms with the nucleoprotein.

Keywords : Protein folding, protein-protein complexes, neutron small-angle scattering, molecular dynamics, neutron elastic incoherent scattering
