Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : Physique pour les Sciences du Vivant

Arrêté ministériel : 25 mai 2016

Présentée par

Benjamin BROCCO

Thèse dirigée par **Trevor (phys) FORSYTH**, Institut Laue Langevin
et codirigée par **Michael HAERTLEIN**
préparée au sein du **Laboratoire Institut LAUE LANGEVIN**
dans **l'École Doctorale Physique**

Neutron scattering and methodological developments for the study of the bacterial translocation machinery

Diffraction de neutron et développements méthodologiques pour l'étude de la machinerie de translocation bactérienne

Thèse soutenue publiquement le **4 juillet 2018**
devant le jury composé de :

Monsieur Trevor FORSYTH
Professeur, Université de Keele, Directeur de thèse
Monsieur Jean-Michel Jault
Directeur de Recherche, Université de Lyon, Rapporteur
Monsieur Patrice Gouet
Professeur, Université de Lyon, Rapporteur
Madame Cecile Breyton
Directeur de Recherche, Université Grenoble Alpes, Examinateur
Monsieur Giuseppe Zaccai
Directeur de Recherche Emérite, Université Grenoble Alpes, Examinateur
Monsieur Pierre-Emmanuel Milhiet
Directeur de Recherche, Université de Montpellier, Examinateur
Madame Selma Maric
Chargé de Recherche, Karolinska Institutet, Examinateur

"*There are only three pre-requisite to a successful SANS experiment: the sample, the sample and the sample*"

Giuseppe Zaccai

UNIVERSITÉ GRENOBLE-ALPES

Abstract

École doctorale de physique

Institut Laue Langevin

Doctor of Philosophy

Neutron scattering studies and methodological developments for the study of the bacterial translocation machinery

by Benjamin Brocco

Résumé en Français

La diffusion de neutrons à petits angles (SANS) est une méthode utilisée pour l'étude d'une large variété de particules en solution. Combinée à un marquage isotopique et à la variation de contraste, elle permet d'obtenir des informations structurales uniques sur des complexes biologiques impliquant plusieurs partenaires, la rendant particulièrement intéressante pour l'étude des mécanismes de translocation. Dans les trois grands domaines de la vie, jusqu'à 30% des protéines doivent être sécrétées hors de la cellule ou intégrées dans sa membrane. Ces mécanismes complexes impliquent deux grands chemins de translocation qui dépendent de multiples protéines cytoplasmiques et membranaires.

L'Holotranslocon (HTL) est un large complexe protéique membranaire composé de sept sous-unités : le translocon triméric SecYEG et les sous-unités accessoires SecD-SecF-YidC-YajC. Cet assemblage peut sécréter des protéines vers l'extérieur de la cellule et en intégrer dans la bicouche lipidique. Nous avons utilisé une sélection de méthodes biophysiques pour analyser différentes stratégies de préparation et la diffusion de neutrons pour analyser les caractéristiques structurales de ce complexe. Nous avons étudié l'efficacité des protocoles actuels pour la production d'un échantillon suffisamment concentré et homogène mais aussi la possibilité d'utiliser des Amphipols comme substitut aux détergents classiquement utilisés pour la purification de ce complexe. Nous avons également tenter de deutérer HTL pour étendre les possibilités offertes par le SANS dans ce contexte. Alors que les caractérisations biophysiques utilisées n'ont pas permis d'améliorer les méthodes de préparation actuelles, la diffusion de neutron nous a permis de confirmer la présence de lipides au centre de la structure. Nous avons assuré la fiabilité de cette stratégie pour étudier des complexes bien définis, formés par plusieurs composants. Nous proposons des méthodes alternatives, basées sur la fluorescence ou la génération de lipodisques, pour l'étude de ce système complexe.

Nous avons également étudié la protéine tétramérique SecB, une chaperonne moléculaire qui est impliquée dans l'adressage des substrats de translocation aux translocons membranaires. Puisque SecB interagit avec des partenaires cytosoliques et ses propres substrats, nous avons utilisé la diffusion de neutrons ainsi que le marquage au deutérium pour analyser des complexes pertinents dans le processus de translocation et impliquant jusqu'à trois partenaires : SecB, un substrat déplié et l'ATPase SecA. Nous avons utilisé les valeurs de "contrast match point" mesurées pour obtenir des informations sur la stœchiométrie et l'affinité des différents assemblages. L'analyse des rayons de girations a permis de localiser les différents composants au sein de ces complexes. De plus, nous avons montré que SecB s'étend lorsqu'elle reconnaît son substrat et nous proposons un modèle fonctionnel, basé sur nos données, pour l'adressage des protéines *via* le chemin post-traductionnel.

Dans l'optique de diversifier les applications du SANS pour les systèmes biologiques, nous avons étendus les protocoles utilisés pour la production de protéines partiellement deutérées *in vivo* aux lipides et acides nucléiques bactériens. Les niveaux de deutération ont été analysés grâce à la diffusion de neutrons, spectrométrie de masse ou résonance magnétique nucléaire. Sur la base de ces données, nous avons extrapolé les niveaux de deutération et "contrast match point" pour chaque type de biomolécule afin de pouvoir prédire les conditions de cultures optimales requises pour atteindre un marquage spécifique. Nous avons, de plus, développé une nouvelle stratégie permettant de marquer sélectivement les protéines tout en conservant un marquage minimal des autres types de molécules.

Abstract in English

Small Angle Neutron Scattering (SANS) is a method that is used for the study of a wide range of particles in solution. Combined with isotope labelling and contrast variation approaches, it allows the extraction of unique structural information on biological complexes involving multiple partner molecules, making it a method of interest for the study of protein translocation mechanisms. In all three kingdoms of life, up to 30% of all proteins need to be secreted out of the cell or integrated into its membrane. These complex mechanisms involve two pathways of translocation that rely on multiple cytoplasmic and membrane proteins.

The Holotranslocon (HTL) is a large membrane protein complex composed of seven subunits: the core trimeric translocon SecYEG and the accessory subunits SecD-SecF-YidC-YajC. This assembly can secrete proteins out of the cell or integrate them into the lipid bilayer. In this project, a range of biophysical methods and sample preparation strategies have been used to analyse the structural features of this complex by SANS. The efficiency of the current protocols to produce a concentrated and homogeneous sample have been investigated, as well as the use of an amphipol molecules as a substitute to classical detergents for the purification of this complex. The deuteration of HTL was attempted to expand the capabilities of SANS in this context. While the biophysical characterization methods we used did not allow us to further improve the current preparation protocols, a key result was the fact that SANS confirmed the presence of lipids at the centre of the structure and the reliability of this analysis method for the study of the well-defined multi-component complexes was tested. Alternative methods for the analyses of this complicated system are proposed, including fluorescence-based assays or lipodisc formation.

The tetrameric protein SecB was also studied; SecB is a molecular chaperone that is involved in the delivery of translocation substrates to the translocon. Since SecB interacts with cytosolic partners and its own substrate, SANS and deuterium labelling were used to analyse translocation-relevant complexes involving up to three partners: SecB, an unfolded substrate and the SecA ATPase. The measured contrast match point was used to obtain information on the stoichiometry and affinity of the various assemblies, and the radii of gyration was used to localize the different components within the complexes. It has also been shown that SecB expands upon substrate binding; a new working model for the post-translational targeting pathway has been proposed, based on these data.

As a corollary to the work described above, protocols for *in vivo* protein deuterium labelling to the deuteration of *E. coli* lipids and nucleic acids have been developed and it is hoped that these may be of general value in widening the scope of SANS applications for the study of biological systems. These approaches were characterized and evaluated by SANS, nuclear magnetic resonance and mass spectrometry. Based on these data, the relevant deuteration levels and contrast match points were extracted for each class of biomolecule so that the optimal culture conditions can be used to achieve a specific level of deuteration. In addition, a new strategy has been developed that allows the selective labeling of proteins while keeping the labelling of other classes of molecules to a minimum within a single culture.

Acknowledgements

In first place, I would like to thank **Trevor Forsyth**, **Michael Haertlein** and the **Institute Laue-Langevin** for giving me the opportunity to prepare my thesis in such a stimulating environment and for exposing me to a wide range of scientifically driven challenges, encounters and travels. I equally thank **Nathan Zaccai** for designing the project.

I would like to thank **Jean-Michel Jault**, **Patrice Gouet**, **Pierre Emmanuel Milhiet**, **Cécile Breyton**, **Selma Maric** and **Giuseppe Zaccai** for taking on their time to read, comment and review my thesis work. A special thank you to **Patrice Gouet**: in 2012, after your fascinating intervention at the ENS Lyon about structural virology, the advices you gave me led me, after a fortunate chain of event, to applying for this PhD project.

My sincere acknowledgements to the neutron facilities and platforms of the Partnership for Structural Biology of Grenoble that I got the chance to use and which played a decisive role in my work. I thank **Anne Martel** and **Lionel Porcar** from the SANS instrument D22 at ILL, **Marie-Sousai Appavou** from the SANS instruments at FRMII, **Elisabetta Erba Boeri** and **Luca Signor** for the mass spectrometry, **Alicia Vallet** and **Adrien Favier** for the nuclear magnetic resonance, **Aline Leroy** and **Christine Ebel** for the analytical ultra-centrifugation, **Caroline Mas** for the multi angle laser scattering and finally **Anne-Marie Villard** for the detergent screening.

I express my gratitude to my colleagues for their advices, support and friendliness, who have made these three and a half years enjoyable, both scientifically and personally. I warmly thank **Trevor Forsyth** for his cheerfulness, unconditional support when it was most necessary. I thank **Michael Haertlein** for the numerous discussions, valuable scientific input and suggestions on my work and thesis. My most sincere thanks to **Nathan Zaccai** for his endless investment in my neutron experiments, biochemistry, data analysis, interpretation, writing... A significant part of this work may not have seen the light without his ideas and advices. A special thank to **Joe Zaccai** who has, with a pencil, a sheet of paper and occasionally a napkin from the canteen, held my hand and walked me through the maze of neutron scattering. I would also like to thank **Christiane Schaffitzel** for introducing me to the world of bacterial translocation, her advices and for commenting my introduction. I thank **Ian Collinson** and his team for their help, plasmids and protocols.

Je remercie **Juliette Devos** pour m'avoir chaperonné ainsi que mon travail, pour m'avoir enseigné toutes ces subtilités empiriques, avoir assuré ma sécurité ainsi que celle des équipements que j'ai approché et pour sa vigilance sans faille quant à la propreté de la paillasse des balances. Un grand merci à **Martine Moulin** pour ses innombrables conseils, "time saving tips", pour avoir sû masquer son agacement en me voyant quadrupliquer mes cultures et surtout pour m'avoir appris que toute la théorie du monde ne remplace parfois pas l'expérience. Je remercie tout également **Valérie Laux** pour les nombreuses discussions et conseils, qu'ils soient d'ordre scientifique, personnel ou même politique, mais aussi de ne pas avoir succombé à la rage engendrée par le "boom-boom" incessant de nos basses qui se sont prouvées plus pénétrantes que les murs fussent isolants...

A special thank to **Martin Nors Pedersen** and **Stuart Fisher** whose programming skills I have abused in order to save significant amounts of time with data analysis. Even if I did not find absolution in your relentless "boots n' cats", I thank **Ashley Jordan**, a.k.a Dr. Mujere, who has been my bench mate for 3 years and has therefore helped me to tackle the days-to-days small lab issues while entertaining me during the column washes. A thanks to my fellow PhD students: **Jennifer Channell** for sharing our ups and down (and making me feel good about my 0.5 mg of proteins), **Vinesh Jugnarain** for keeping me company and convincing me to have food delivered at the lab at very random and unlikely time, **Alycia Yee** for joining me in the mountains and making sure I don't miss out on a good conference, **Samuel Freeman** for riding the trails alongside with me and teaching me about the wonders of Latex, **Lindsay McGregor** for head-banging with me when everyone else has deserted the labs, **Sarah Waldie** for keeping our sugar levels high enough to feed our scientific ideas (or at least make it through the day), **Lukas Gajdos** who will certainly forgive me for not knowing where to put the accents on his name and **Maria Pedrosa** for reminding us that neutron are not only good at being scattered. I shall not forget our beloved trainees who never fail to bring novelty and freshness in the lab : **Amy** (I felt really important being called "boss"), **Katherine F** (yes, practising the flute in the culture room is acceptable), **Alice**, **Katherin M**, **Abi**, **Dany** and **Freya**. Je remercie **Florent** pour nos discussions, bons moments passés et pour savoir maintenir le bâtiment et PSB en un seul morceau !

Je remercie ensuite toutes les personnes qui ne font pas partie du laboratoire mais dont les conseils et le support ont été important. Ma **mère** qui s'inquiète suffisament pour que je n'aie pas à le faire moi-même et mon **père** avec qui je m'échappe dans les collines pour y trouver air frais, conseils et soutien. Ma **sœur**, **Chouchou** et **Léon** qui sont comme un camp de retranchement où on y trouve discussion, conseils, bonne chaire et aussi un bel exemple à suivre! **Yaya** qui, malgré que le mot "Holotranslocon" ne soit pas encore accepté en championnat de Scrabble, n'a pas manqué de s'en souvenir pour me demander régulièrement comment il va. Je n'oublierais certainement pas mes amis, qu'ils soient à Toulouse, Lyon, Paris ou Nice; Italiens, Macédoniens ou Français; scientifiques, photographes, danseurs ou maître crêpiers; compagnons de montagne, de voyage, de table ou de jeu; merci à vous tous pour votre soutien **P****s**, **Rompela**, **Spasevski**, **Perrinette**, **Plonplon**, **Pipou**, **Roubass**, **Moral**...

Last (but certainly not least), **Serena**. You have your own paragraph, firstly because I did not know in which section to put you in as you would qualify to be part of all of them, but also because these are my acknowledgements and I might as well do as I please. I thank you for your constant support, the laughs, the travels, the adventures, the discussions, the curries, the foodies... I am grateful for the time we have spent together (even remotely) and I am sure I can thank you in advance for the great times to come ! Oh.. and thank you for spell-checking everything so willingly !

And thanks to thee, unknown stranger, for you are about to read my work in the hope of finding the answer to your own dismay.

Contents

List of Figures

List of Tables

List of Abbreviations

Apol	**A**mphipol A8-35
AUC	**A**nalytical **U**ltra **C**entrifugation
CMC	**C**ritical **M**icellar **C**oncentration
CMP	**C**ontrast **M**atch **P**oint
DDM	n-**D**o**D**ecyl-β-D-**M**altopyranoside
DFYY	Sec**D** Sec**F** **Y**idC **Y**ajC
DLS	**D**ynamic **L**ight **S**cattering
DPG	**D**i**P**hosphatidyl**G**lycerol
DTT	**D**i**T**hio**T**hreitol
EDTA	**E**thylene**D**iamine**T**etraacetic **A**cid
ESI	**E**lectro**S**pray **I**onization
FRMII	**F**orschungs**R**eaktor **M**ünchen **II**
GF	**G**el **F**iltration
HTL	**H**olo**T**rans**L**ocon (SecYEG-SecDF-YidC-YajC)
ILL	**I**nstitut **L**aue **L**angevin
IM	**I**nner **M**embrane
IPTG	**I**so**P**ropyl β-D-1-**T**hio**G**alactopyranoside
IMAC	**I**mmobilized **M**etal **A**ffinity **C**hromatography
LC	**L**iquid **C**hormatography
MALDI	**M**atrix **A**ssisted **L**aser **D**esorption **I**onization
MALS	**M**ulti **A**ngle **L**aser **S**cattering
MS	**M**ass **S**pectrometry
MSP	**M**embrane **S**caffold **P**rotein
OM	**O**uter **M**embrane
PAGE	**P**oly**A**crilamide **G**el **E**lectrophoresis
PCC	**P**rotein **C**onducting **C**hannel
PE	**P**hosphatidyl**E**thanolamine
PG	**P**hosphatidyl**G**lycerol
POPG	1-**P**almitoyl-2-**O**leoyl-sn-glycero-3-**P**hospho-rac-(1 **G**lycerol)
POPE	1-**P**almitoyl-2-**O**leoyl-sn-glycero-3-**P**hospho**E**thanolamine
RBD	**R**ibosome **B**inding **D**omain
RNC	**R**ibosome **N**ascent **C**hain
RPM	**R**otations **P**er **M**inute
SANS	**S**mall **A**ngle **N**eutron **S**cattering
SAS	**S**mall **A**ngle **S**cattering
SEC	**S**ize **E**xclusion **C**hromatography
SDS	**S**odium **D**odecyl **S**ulfate
SLS	**S**tatic **L**ight **S**cattering

SMA	**S**tyrene **M**aleic **A**cid
TOF	**T**ime **O**f **F**light
TFA	**T**ri**F**luoro **A**cetic acid
TLC	**T**hin **L**ayer **C**hromatography

Chapter 1

Introduction

Résumé en Français

La translocation bactérienne

Les membranes biologiques jouent un rôle primordial, notamment en délimitant la cellule pour permettre la mise en place et le maintien de conditions favorables aux réactions biochimiques nécessaires à son fonctionnement. Cependant, la survie des cellules dépend de leur capacité à interagir avec leur environnement de façon contrôlée. Ce dialogue est orchestré par des protéines qui peuvent être sécrétées hors de la cellules ou intégrées dans les membranes pour assurer l'interface entre le cytoplasme et le milieu extra-cellulaire. Ces macromolécules jouent un rôle majeur et représentent jusqu'à un tiers des protéines codées par les génomes des trois grands domaines de la vie. Dans la majorité des cas, ces protéines sont incapables de traverser la membrane hautement hydrophobe par elles-même et doivent utiliser des machineries moléculaires spécialisées.

Chemins de translocation: Les mécanismes de sécrétion ou d'insertion des protéines dans les membranes sont regroupés sous le nom de mécanismes de translocation. Chez la bactérie, la translocation est principalement prise en charge par le système "Sec" dans lequel le complexe triméric SecYEG occupe une place centrale. SecYEG, également appelé translocon, est composé de trois sous-unités (SecY, SecE et SecG) dont l'architecture caractéristique est directement liée à sa fonction :

- Quand il est inactif, la présence d'un anneau de constriction et d'un "bouchon" permet d'assurer que le translocon ne perturbe pas l'étanchéité de la membrane.
- Lorsqu'une protéine doit emprunter le translocon, l'anneau de constriction est écarté pour permettre le passage de la protéine dépliée, tout en maintenant l'étanchéité de la membrane
- Dans le cas des protéines membranaires, le translocon peut s'ouvrir latéralement afin de donner un accès direct à la membrane pour l'insertion de protéines membranaires.

Cependant, le translocon est un complexe passif qui doit s'associer avec des partenaires fournissant l'énergie nécessaire à la translocation. Dans cette optique, deux

principaux chemins peuvent être utilisés: la translocation co-traductionnelle ou celui de la translocation post-traductionnelle.

L'adressage au chemin de translocation co-traductionnelle se fait généralement grâce à une séquence signal suffisamment longue et hydrophobe. Quand la traduction est initiée et que la séquence signal émerge du ribosome pour être reconnue par la signal recognition particle, la traduction est arrêtée et le super-complexe amené au niveau de la membrane. Le ribosome pourra alors interagir avec le translocon qui s'ouvre pour accueillir la protéine, qui est alors traduite et transloquée en même temps. L'énergie déployée par le ribosome pour l'élongation de la protéine en cours de traduction permet de pousser la protéine à travers le translocon.

L'adressage au chemin de translocation post-traductionnel se fait généralement par des séquences signal plus petites et moins hydrophobes. Durant la traduction de la nouvelle protéine, des chaperonnes, et notamment SecB, empêchent son repliement. La séquence signal et SecB sont reconnues par la protéine moteur SecA qui se lie alors au translocon. Ce dernier est une fois de plus ouvert et SecA hydrolyse de l'ATP pour pousser la protéine dépliée au travers du translocon.

Un super-complexe pour la translocation: SecYEG est aussi capable d'interagir avec d'autres partenaires membranaires qui supportent son activité de translocation.

Parmi eux, la protéine YidC est une "insertase" capable de transloquer des protéines membranaires par elle-même. Elle peut également s'associer avec le translocon pour favoriser l'insertion membranaire de protéines peu hydrophobe ou portant des charges inhabituelles. Elle favorise également l'assemblage de gros complexes membranaires..

Le translocon peut également interagir avec un groupe de protéines exprimé dans un même opéron, SecD-SecF et YajC. SecDF est capable d'utiliser la force proton-motrice pour favoriser l'intégration de protéines membranaires, particulièrement quand celles-ci ont de larges boucles périplasmiques.

De nombreuses études ont rapporté que ces partenaires pouvaient se lier au translocon mais la preuve d'une interaction simultanée n'a été que récemment apportée, grâce au développement de nouvelles méthodes d'expressions spécialisées pour la production de gros complexes composés de nombreuses sous-unités. Cet "Holotranslocon" est plus efficace que le translocon classique pour l'intégration de protéines membranaires et peut également interagir avec le ribosome ou SecA. Une première caractérisation structurale en cryo-microscopie électronique a étudié l'organisation des sous-unités entre elles. L'utilisation de la diffusion de neutrons à petits angles a de plus mis en évidence une cavité lipidique au sein de ce complexe qui pourrait participer au processus de translocation en attribuant une fonction de "chaperonne" membranaire à l'Holotranslocon.

Les neutrons en biologie

Les neutrons sont des objets quantiques pouvant être définis comme une onde avec une longueur d'onde donnée ou comme une particule avec une masse et vitesse données.

Les neutrons peuvent interagir avec les atomes ; là où les photons (typiquement dans des rayons-X) interagissent avec le nuage électronique, les neutrons interagissent avec le noyau et l'ampleur de cette interaction, contrairement aux photons, n'est donc pas dépendante du numéro atomique mais plutôt de facteurs associés aux caractéristiques magnétiques du noyau.

Notion de contraste : La longueur de diffusion, qui caractérise la façon dont le rayonnement "voit" l'atome, dépend du type de noyau rencontré. Pour les molécules, on parle de densité de longueur de diffusion, laquelle correspond à la somme des longueurs de diffusion de chaque atome composant la molécule divisé par son volume. Dans une expérience structurale en solution, le signal diffusé provient de la molécule d'intérêt et de son solvant. De ce fait, la différence entre ces deux composants nous permet d'avoir accès au signal provenant de la molécule, signal d'autant plus important que la différence est grande; on parle de contraste. Un cas particulier émerge lorsque la densité de longueur de diffusion d'un solvant est égale à celle de la molécule ; les deux deviennent indiscernables et l'on parle alors de "contrast match point". Il est important de noter que le "contrast match point" d'un système est dépendant de sa composition atomique et donc de sa composition moléculaire.

Substitution isotopique : Dans une expérience de diffusion de neutrons biologique, le contraste est le seul facteur auquel nous avons accès (indépendamment du design de l'instrument) et il peut être modulé pour augmenter ou réduire l'intensité du signal obtenu. Puisque les neutrons interagissent avec les noyaux, ils sont sensibles aux isotopes qui auront des caractéristiques nucléaires différentes. C'est notamment le cas pour le couple (^{1}H) / (^{2}H) qui possède des longueurs de diffusion très différentes, phénomène qui peut ainsi être utilisé pour faire varier le contraste d'une expérience.

La première méthode consiste à substituer le solvant aqueux par un solvant où les H_2O sont remplacés en tout ou partie par du D_2O. Ainsi, en variant la compositions du solvant, il est possible de moduler le contraste d'une expérience. On parle d'expérience de variation de contraste.

La seconde méthode consiste à modifier la composition isotopique d'une molécule en substituant les hydrogènes par des deutérium. Il devient alors important de différencier deux types d'hydrogène dans une molécule : les hydrogènes labiles (liés à des azotes ou oxygènes) et les hydrogènes non-échangeables (liés aux carbones). Si la première catégorie peut être simplement substituée par échange passif avec le solvant, les hydrogène non-échangeables doivent être intégrés aux molécules lors de leur biosynthèse. On parle alors de deutération biomoléculaire, généralement réalisée en cultivant des organismes unicellulaires *in vivo* dans un milieu contenant du deutérium.

Objectif de l'étude

La translocation bactérienne : Cette machinerie implique de nombreux partenaires. L'holotranslocon est un super-complexe composé de sept sous-unités, qui pourrait jouer un rôle central dans la sécrétion, mais surtout l'intégration de protéines

membranaires dans la bicouche lipidique. Dans cette étude, nous utilisons la diffusion de neutrons à petits angles, ainsi que des méthodes biophysiques afin d'analyser différentes stratégies de purification, caractériser le complexe d'un point de vue fonctionnel et structural.

La chaperonne SecB est impliquée dans le chemin de translocation post-traductionnel et peut, de ce fait, interagir avec les substrats ou partenaires impliqués dans ce mécanisme. Nous caractérisons ces différentes interactions, utilisons la diffusion de neutrons à petits angles et la deutération biomoléculaire afin d'analyser différents complexes impliquant jusqu'à trois partenaires.

Deutération de biomolécules : La deutération des biomolécules est souvent une étape primordiale pour le succès d'expériences neutroniques. Nous définissons, dans cette études, les conditions de cultures permettant d'obtenir *in vivo* un niveau de deutération cible de façon prédictible pour les protéines, lipides et acides nucléiques.

Nous proposons également une nouvelle méthode de deutération qui permet de sélectivement marquer *in vivo* les protéines sans affecter le marquages des autres classes de biomolécules. Cette méthode pourrait être utilisée dans des cas particulier où des complexes à plusieurs composants (protéines membranaires, complexe protéine/acide nucléique) ne pourrait pas être reconstitués *in vitro*.

1.1 Translocation mechanisms

1.1.1 Secreted and membrane protein in cells

The most current scientific definition of life relies on seven characteristics; homeostasis, organization, metabolism, growth, adaptation, response to stimuli and reproduction. Koshland, 2002 offered an alternative definition based on seven "pillars" that are necessary to define a living organism; program, improvisation, energy, regeneration, adaptability, seclusion and compartmentalization. The two last terms describe the necessity for an organism to be confined in a defined volume, so its internal composition can be controlled and maintained to allow for the reactions that are required to maintain life. Compartmentalization therefore relies on the presence of a physical barrier that needs to be sufficiently impermeable to keep the organism's content inside and resistant enough to protect it from outside deleterious factors. However, it is also primordial to have a certain permeability and flexibility so that the organism can exchange chemicals with its surrounding, withstand changes of physical parameters or even move.

These characteristics have led to the universal selection of lipid assemblies as the core component of biological membranes. Their structure (hydrophilic head group and hydrophobic tails) gives them amphiphilic properties that allow their self-assembly into bilayers that expose the hydrophilic heads to the solvent and packs the hydrophobic tails together (see Figure 1.1). The lipid bilayer is therefore impermeable to big, hydrophilic molecules and ions but will allow the unrestricted intake of small neutral molecules, thus making it a selective barrier secluding the living organism from its environment (Voet *et al.*, 2008).

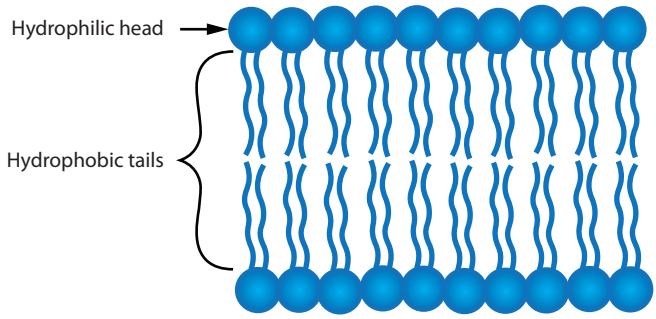


FIGURE 1.1: Schematic representation of a lipid bilayer

Permeability to small non-charged molecules is not sufficient to maintain the homeostasis and does not account for the adaptability of a living organism that will need to exchange ions (osmotic balance, reaction cofactors, synthesis of biomolecules etc.) or big molecules (energy, biomolecules precursors, metabolic wastes, etc.). Furthermore, lipids by themselves do not allow the organism to sense its surrounding (chemicals and physical parameters) and membranes therefore need alternative mechanisms to serve as the interface between the organism and its environment.

The capabilities of lipid bilayers are improved by the presence of a large array of membrane proteins which can be integral (through both leaflets) or peripheral (anchored to

a leaflet or to the polar heads). This class of biomolecule accounts for up to a third of all expressed proteins in living organisms and their roles range from ion transport to signal transduction. Due to their key roles in cell function and their accessibility at the surface of cells, about 60% of pharmaceutical drug targets are directed against membrane proteins (Rawlings, 2016). Their importance is undeniable and it is therefore of prime interest to have a broad understanding of their functions and biogenesis.

One key aspect of membrane proteins is their integration into the membrane; since lipids are highly hydrophobic, membrane proteins must have hydrophobic residues in order to be integrated into the membrane. They usually also have hydrophilic residues in order to interact with the aqueous environment. Hydrophilic amino acids may need to cross the central hydrophobic part of the bilayer, which is a process that is thermodynamically unfavoured. This is also the case for secretory proteins that go through the membrane without necessarily having hydrophobic segments, which would destabilize it in its target environment. In some cases, these processes can occur spontaneously, but the vast majority of these events rely on the use of a translocation machinery.

1.1.2 Translocation in bacteria

The integration of membrane proteins into the lipid bilayer and secretion of proteins to the outside of the cell are processes that are grouped under the name translocation. It includes the mechanisms involved in the initial steps of the translation of membrane or secreted proteins and how the proteins are targeted to their final destinations. Bacteria have evolved sophisticated mechanisms that allow secreted and membrane proteins to be adequately handled depending on their fate; 96% of them (among which 60% are transmembrane proteins) are translocated by the universally conserved Sec translocon (Orfanoudaki and Economou, 2014).

The SecYEG core translocon

The SecYEG Protein Conducting Channel (PCC), or translocon, is a complex which has an architecture and dynamic conformation that forms a channel through the membrane, while maintaining its impermeability to ions. It allows for vectorial and lateral movement of proteins for secretion and membrane integration respectively.

The SecYEG translocon is composed of three proteins, SecY, SecE and SecG (see Figure 1.2). SecY possesses ten transmembrane helices, four of them forming the hydrophilic pore which gains flexibility from the periplasmic loop that act like hinges. The pore itself is hourglass shaped, the central part being occupied by six hydrophobic residues that form a 5 Å ring (van der Berg *et al.*, 2003). Together with the plug, a small helix that rests on top off the pore on the periplasmic side, the ring acts as a channel gate preventing ion leakage in the resting closed state and forming a gasket around the substrate protein during the translocation (Park and Rapoport, 2011). A striking feature of the SecYEG translocon is the flexible arrangement of four transmembrane

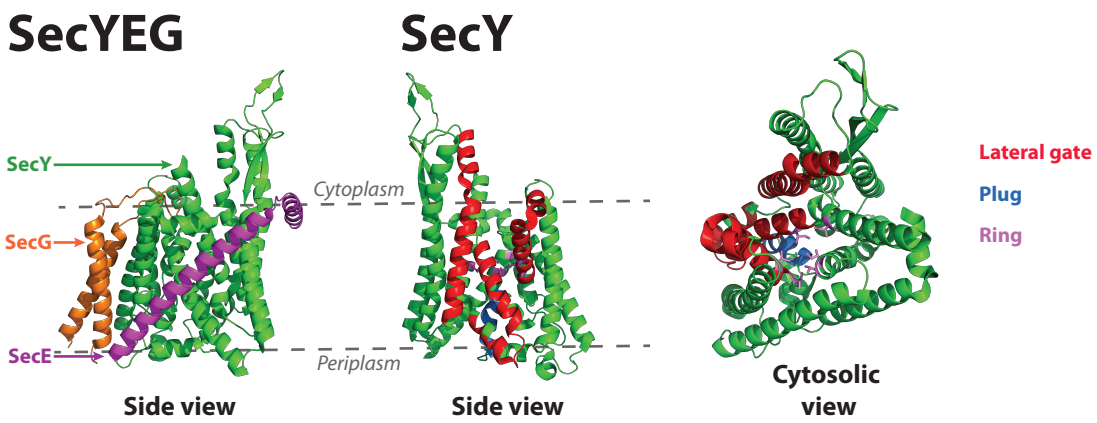


FIGURE 1.2: Cartoon representation of the crystal structure of *T.termophilus* **SecYEG** translocon (pdb:5AWW, Tanaka *et al.*, 2015). SecYEG: Arrangement of the subunits, SecY (green), SecE (magenta) and SecG (orange). **SecY**: the internal structure of SecY is represented, the lateral gate (red), ring (pink with side chains) and the plug helix (blue). The side view faces the lateral gate's opening and the cytosolic view shows the organization of the ring's side chains and helix plug that prevent leakages.

helices that create a lateral gate in order to give direct access to the lipid bilayer for transmembrane helices of translocation substrates.

SecE binds to the opposite side of the lateral gate and stabilizes the conformation of SecY by engaging its transmembrane helices around the pore. The small, non-essential SecG, contacts SecY laterally and seals the cytoplasmic funnel with a cytoplasmic loop covering the channel (Tanaka *et al.*, 2015).

The oligomeric state of the translocon has been widely debated as the detergent solubilized complex has been characterized biochemically and structurally as monomeric (van der Berg *et al.*, 2003) or dimeric. The latter has been described as a back to back dimer, only one channel being engaged with a pre-protein and leaving the lateral gate free for membrane protein insertion (Hizlan *et al.*, 2012), or front to front dimer with the second copy of SecY forming a support conduction channel (Mitra, Schaffitzel, *et al.*, 2005). If the monomeric form is sufficient to observe translocation activity, it is believed that the second copy of SecYEG may stabilize the other and may serve as a platform for favouring the binding of partners, therefore enhancing translocation efficiency (Park and Rapoport, 2012).

Since SecYEG is a passive complex, the energy required for the translocation process is provided by its main partners: (1) the ribosome which uses the energy developed by translation as a driving force for unidirectional movement of pre-proteins through the ribosomal tunnel and translocation pore, (2) SecA, which powers the process with its ATPase activity (Rapoport *et al.*, 2017). The conformational changes occurring during the translocation and binding to the partners are further discussed in section 1.1.2.

Sorting of proteins

Secreted and membrane proteins strongly differ in their biochemical and biophysical properties and cannot necessarily be handled in the same way. In most cases, a key point is to keep the protein in a translocation competent state. They need to remain unfolded as translocating a fully folded protein through the membrane would imply having a machinery that is capable to provide a pore at least as wide as the protein itself. Furthermore, it is essential for proteins with hydrophobic regions that they remain soluble to prevent aggregation (Tsirigotaki *et al.*, 2017). Two distinct pathways have been evolved by bacteria in order to handle and target these proteins to the SecYEG translocon:

- Co-translational: all processes are done at the same time as the protein is being translated by the ribosome.
- Post-translational: the translocation and targeting of the newly synthesised protein happens after it has been produced by the ribosome.

The sorting of the proteins relies on the *signal sequence*, an amino-terminal short peptide. Although no consensus motif has been clearly identified, shorter and less hydrophobic signal sequences tend to target proteins to the post-translational pathway. Longer, highly hydrophobic signal sequences will on the other side target the nascent protein to the co-translational pathway. In many cases of membrane proteins, the first transmembrane segments are extremely hydrophobic and will not only serve as a signal sequence, but also control the topology of the membrane protein by dictating the orientation of the first segment following the "positive-inside rule" (the positively charged end is oriented to the cytoplasm, Junne *et al.*, 2007). In frequent cases, internal transmembrane segments replace the first one to ensure accurate sorting of membrane proteins (Schibich *et al.*, 2016).

The dominant actor of protein sorting is the *Signal Recognition Particle (SRP)* which, because of its strong preference for transmembrane domains (Noriega *et al.*, 2014), routes 87% of all inner membrane proteins to the co-translational pathway (Schibich *et al.*, 2016). The *Trigger Factor (TF)* does not actively select its substrates and handles periplasmic or cytoplasmic proteins by binding the less hydrophobic and shorter signal sequences (von Loeffelholz *et al.*, 2011) that have been ignored by SRP (Schibich *et al.*, 2016).

The SecA ATPase has also been reported in a ribosome bound state and shown to take part in protein sorting by interacting with nascent protein at the ribosome exit tunnel (Mitra, Frank, *et al.*, 2006), even though the trigger factor outnumbers SecA and is therefore more likely to handle post-translocational targeting (Singh *et al.*, 2014). This sorting strategy cannot however be considered as universal since SecA has for example been reported to cooperate with SRP for the translocation of secretory proteins (Y. Zhou *et al.*, 2014).

Ultimately, the signal sequence may be cleaved off by the signal peptidase after translocation as been completed (Driessen and Nouwen, 2008).

Co-translational translocation

The addressing of nascent protein to the co-translational pathway is handled by the SRP (Rapoport, 2007; Shan and Walter, 2005), a universally conserved machinery that delivers nascent membrane or secreted proteins to the plasma membrane by recognizing the signal sequence. The SRP is an RNA-protein complex which in *E. coli* consists of a hairpin structured 4.5S RNA and Ffh, a 48kDA protein (see Figure 1.3). The N-terminal GTPase domain of Ffh docks on empty or translating ribosome while a methionine rich domain is inserted in the exit tunnel for the signal peptide recognition (Jomaa *et al.*, 2016). The Ribosome Nascent Chain (RNC)/SRP complex is stabilized if an appropriate signal sequence (or transmembrane helix) is recognized (X. Zhang, Rashid, *et al.*, 2010). Substrate specificity is further enhanced by FtsY, a membrane associated protein, which binds to SRP through an interaction of their N-termini thus initiating the targeting of the large assembly to the plasma membrane (Saraogi *et al.*, 2014). The membrane anchored RNC can then interact with the SecYEG translocon, which induces the release of SRP and FtsY from the complex through a GTP hydrolysis mediated SRP/FtsY conformational rearrangement (X. Zhang, Schaffitzel, *et al.*, 2009).

The inactive translocon is in a closed state, the lateral gate being shut, the channel constricted by the the ring's six hydrophobic amino acids and obstructed by the plug. Upon binding of the Ribosome Nascent Chain complex to the cytosolic protrusions of SecY, long range conformational changes initiate the displacement of the plug, ring and the opening of the lateral gate (Ge *et al.*, 2014). Once partially opened, the lateral gate can accommodate the hydrophobic signal sequence or transmembrane segment which are recognized by the hydrophobic residues of the lateral gate but also by the lipid partitioning mechanisms (Rapoport *et al.*, 2017; Hessa *et al.*, 2005). The channel is then further opened by coordinated movement of the lateral gate, plug and constriction ring that is formed by only four amino acid residues at this point so that the nascent chain can slide through as a loop. The hourglass shape minimizes the risks of non-specific interaction between SecY and the peptide that could impair the movement of the translocated protein (L. Li *et al.*, 2016). The folding is initiated outside of the translocon on the periplasmic side (Kadokura and Beckwith, 2009).

In the case of secreted proteins, the signal sequence is cleaved by the signal peptidase towards the end or after full translocation of the protein so it can be released into the periplasm (Auclair *et al.*, 2012). In the case of membrane protein, the transmembrane segment will intercalate into the lateral gate and slide sideways out to the lipid bilayer (Rapoport *et al.*, 2017; Hessa *et al.*, 2005), the channel switching from a conformation allowing vectorial movement for hydrophilic segments to a conformation allowing lateral movement of the transmembrane segments (Gouridis *et al.*, 2013). For polytopic membrane proteins, several transmembrane helices can be retained at or in the vicinity of the pore to be released in bulk into the membrane (Hou *et al.*, 2012). The large majority of membrane proteins are handled by the co-translational pathway as it has been speculated the slower translocation kinetics facilitates the progress of the critical steps required for proper membrane protein integration (Schibich *et al.*, 2016).

The main steps of co-translational translocation are illustrated in Figure 1.5.

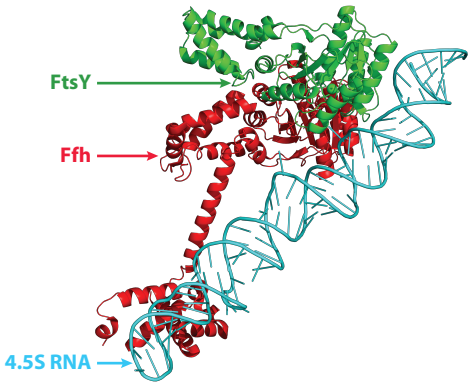


FIGURE 1.3: Representation of the crystal structure of *E.coli* **Signal Recognition Particle** (Ffh + 4.5s RNA) with its receptor (**FtsY**) (pdb:2XXA, Ataide *et al.*, 2011). Ffh (red), 4.5S RNA (cyan) and FtsY (green).

Post-translational translocation in bacteria

The choice of post-translational pathway is initiated by a shorter, less hydrophobic signal sequence that is ignored by SRP and recognized by the Trigger Factor (TF) (Schibich *et al.*, 2016). This ATP independent chaperone composed by a Ribosome Binding Domain (RBD), a peptidyl-prolyl *cis-trans* isomerase (PPIase) that is essential to its chaperone activity, and a C-terminal domain whose function remains mainly unknown (see Figure 1.4, Ludlam *et al.*, 2004). If TF is ubiquitous in bacteria, it is however dispensable for protein export and viability (Lee and Bernstein, 2002). Is is found in the cytoplasm as a dimer but binds to the ribosome as a monomer (Merz *et al.*, 2008) where it exposes its surface to the nascent protein for docking onto the signal sequence through hydrophobic and polar interactions (Kaiser *et al.*, 2006). TF can bind to N-terminal and distal domains of the nascent chain, implying that multiple TF can be found on one translating ribosome nascent chain, allowing aggregation prone domains to be screened from the solvent (Saio *et al.*, 2014), and keeping the protein in an unfolded translocation competent state even after it is released from the ribosome.

Another actor of the post-translational pathway is SecB, a protein that tetramerizes (Xu *et al.*, 2000). It handles only 4% of the bacterial secretome but is a representative example of a secretion specific chaperone. SecB binds to nascent or fully translated proteins mainly through hydrophobic contacts (see Figure 1.4, Bechtluft, Kedrov, *et al.*, 2010) so that the target protein ends up wrapped around the chaperone (Huang *et al.*, 2016) thus avoiding exposure of their hydrophobic segments to the solvent. The binding of SecB does not necessarily require a signal peptide; it is sequence independent, but favours unfolded proteins and structural rearrangements allow SecB to accommodate a wide range of protein sizes (Huang *et al.*, 2016). In a similar way as for TF, SecB prevents pre-proteins from folding so that they can be handed over to SecYEG bound or cytoplasmic SecA for which SecB has a high affinity (Hartl *et al.*, 1990).

The translocase motor protein SecA is an essential and conserved ATPase in bacteria. It can be found as a dimer in the cytosol (Zimmer *et al.*, 2008), bound to the ribosome (Singh *et al.*, 2014) or to SecY (Gouridis *et al.*, 2013). SecA (see Figure 1.4) binds

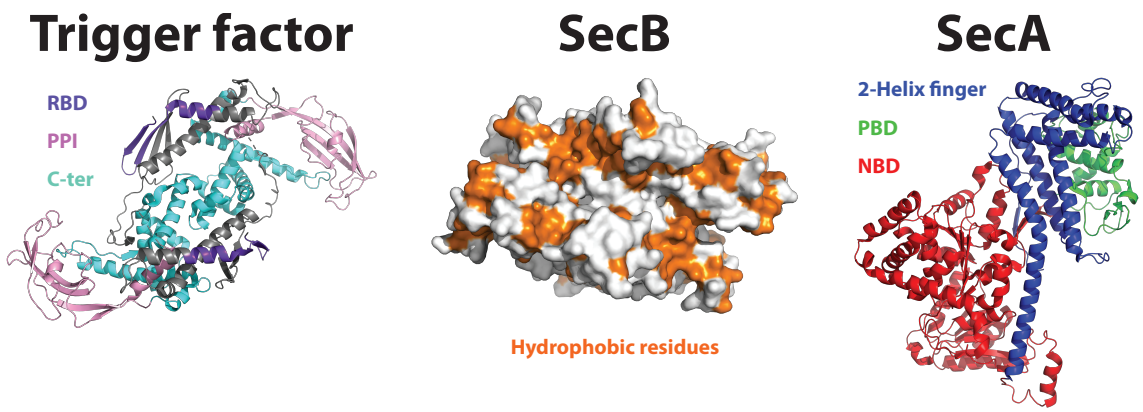


FIGURE 1.4: Representation of the *V.cholerae* **Trigger Factor** (TF) (pdb: 1T11,Ludlam *et al.*, 2004 with the Ribosome Binding Domain (RDB, purple), the Peptidyl-Prolyl *cis-trans* Isomerase (PPIase, pink) and the C-terminal domain (cyan). The surface of the crystal structure of **SecB** tetramer (pdb: 1QYN, Dekker *et al.*, 2003) is represented with the hydrophobic residues in orange. The crystal structure of **SecA** is represented (pdb: 1M6N, Hunt *et al.*, 2002) with the 2-Helix finger domain (blue), Pre-protein Binding Domain (PBD, green) and nucleotide binding domain (NBD, red).

to the signal sequences with its peptide binding groove, the specificity being insured by acidic residues recognizing positive charges. The specificity is further insured by a "proof reading" mechanism involving the C-terminal tail of SecA that occludes the binding grove and that can only be displaced by bare or SecB-escorted signal sequences (Gelis *et al.*, 2007). Additionally, it has been proposed that mature domain targeting signals, which are hydrophobic domains in mature proteins, could also play a major role in the targeting process by addressing pre-proteins to the post-translational pathway. This is done through an interaction of those domains with a hydrophobic platform that neighbours the peptide binding groove of SecA (Chatzi, Sardis, Tsirigotaki, *et al.*, 2017) but also enhances the stability of the translocation competent complex (Gouridis *et al.*, 2013). SecA can exist in complex with SecB in 1:4 or the more translocation-proficient 4:2 stoichiometry (Suo *et al.*, 2015), though SecB is dispensable for the translocation of many proteins.

SecA itself undergoes dramatic conformational changes during the translocation process. These are regulated by nucleotides (Karamanou *et al.*, 2007; Hunt *et al.*, 2002), pre-proteins (Gold *et al.*, 2013), SecYEG, lipids and chaperones (Chatzi, Sardis, Economou, *et al.*, 2014). The oligomeric state of SecA when bound to SecYEG has been widely debated and studies indicate both monomeric and dimeric forms are necessary intermediates of the translocation process. The binding of SecA to SecYEG occurs when SecA is a dimer; it then undergoes quaternary conformation transitions prior to pre-protein induced ATP hydrolysis and monomerization (Gouridis *et al.*, 2013). In a similar way as occurs for the ribosome, one protomer of the SecA dimer interacts with cytoplasmic protrusions of SecY which initiate, in a nucleotide dependent manner, the destabilization of the plug, partial opening of the ring and gate of SecYEG (Zimmer *et al.*, 2008; Allen *et al.*, 2016). This binding stimulates the ATP hydrolysis activity (Lill *et al.*,

1990), and the presence of a pre-protein elongates and loosens the SecA dimer (Gouridis *et al.*, 2013) while also promoting its ATPase activity (Karamanou *et al.*, 2007). The pre-protein establishes superficial contacts with SecY (Schiebel *et al.*, 1991), and the SecA dimer disassembles (Gouridis *et al.*, 2013) so SecA can associate with the funnel of the translocon. The signal peptide relocates to the lateral gate, with the N-terminus oriented toward the cytoplasm. The plug is fully displaced and the pre-protein moves through the central ring formed by SecY, forming a loop in the periplasm (L. Li *et al.*, 2016). This movement is powered by cycles of ATP hydrolysis (Schiebel *et al.*, 1991) during which the two-helix finger domain contacts the pre-protein, pushes it through the pore and releases it to initiate a new cycle (Zimmer *et al.*, 2008; Erlandson *et al.*, 2008; Bauer *et al.*, 2014). Alternative models suggest that SecA does not actually push the pre-protein through the pore but biases its diffusion in the forward direction (Allen *et al.*, 2016). The post-translational translocation event is terminated by the action of the signal peptidase (Auclair *et al.*, 2012) to release the protein into the periplasm, a process possibly helped by the periplasmic chaperone Ppid (Sachelaru, Petriman, Kudva, and H.-G. Koch, 2014).

While SecA is the main actor of post-translational translocation, it is also able to interact with the ribosome (Mitra, Frank, *et al.*, 2006) and is implicated in the co-translational pathway for the translocation of large periplasmic loops of membrane proteins (Neumann-Haefelin *et al.*, 2000). These considerations may indicate, together with the fact that no clear sorting rule has been identified, that both pathways cooperate for efficient protein translocation.

The main steps of post-translational translocation are illustrated in Figure 1.5.

1.1.3 Translocation in eukaryotes

In eukaryotic cells, protein translocation analogous to the bacterial translocation occurs at the surface of the endoplasmic reticulum and involves a complex homologous to SecYEG composed of the three subunits: Sec61 $\alpha\beta\gamma$. Sec61 has a very similar structure to SecYEG, with two halves composed of transmembrane helices that are connected by a luminal loop. This hour-glass shaped complex narrows on six aliphatic amino acid projecting their hydrophobic side chains radially to form a constriction ring that delimits the pore (R. M. Voorhees and R. S. Hegde, 2016). The complex can also open laterally and the luminal side of the channel is occupied by a plug domain that seals the pore. The structure and function of the core translocon is universally conserved from bacteria to mammals (Rapoport *et al.*, 2017). Sec61 can be engaged in co-translational translocation events; membrane and secreted proteins are being targeted through a Signal Recognition Particle dependent mechanism (R. Voorhees and R. Hegde, 2018). This allows for the delivery of the translating ribosome to the Sec complex. For the post-translational pathways, cytosolic chaperones hold the protein in an unfolded translocation-competent form and address the pre-protein to the translocation machinery (Ast *et al.*, 2013), though this targeting mechanism, that likely involves the Sec62/62 complex, remains unclear (Rapoport *et al.*, 2017).

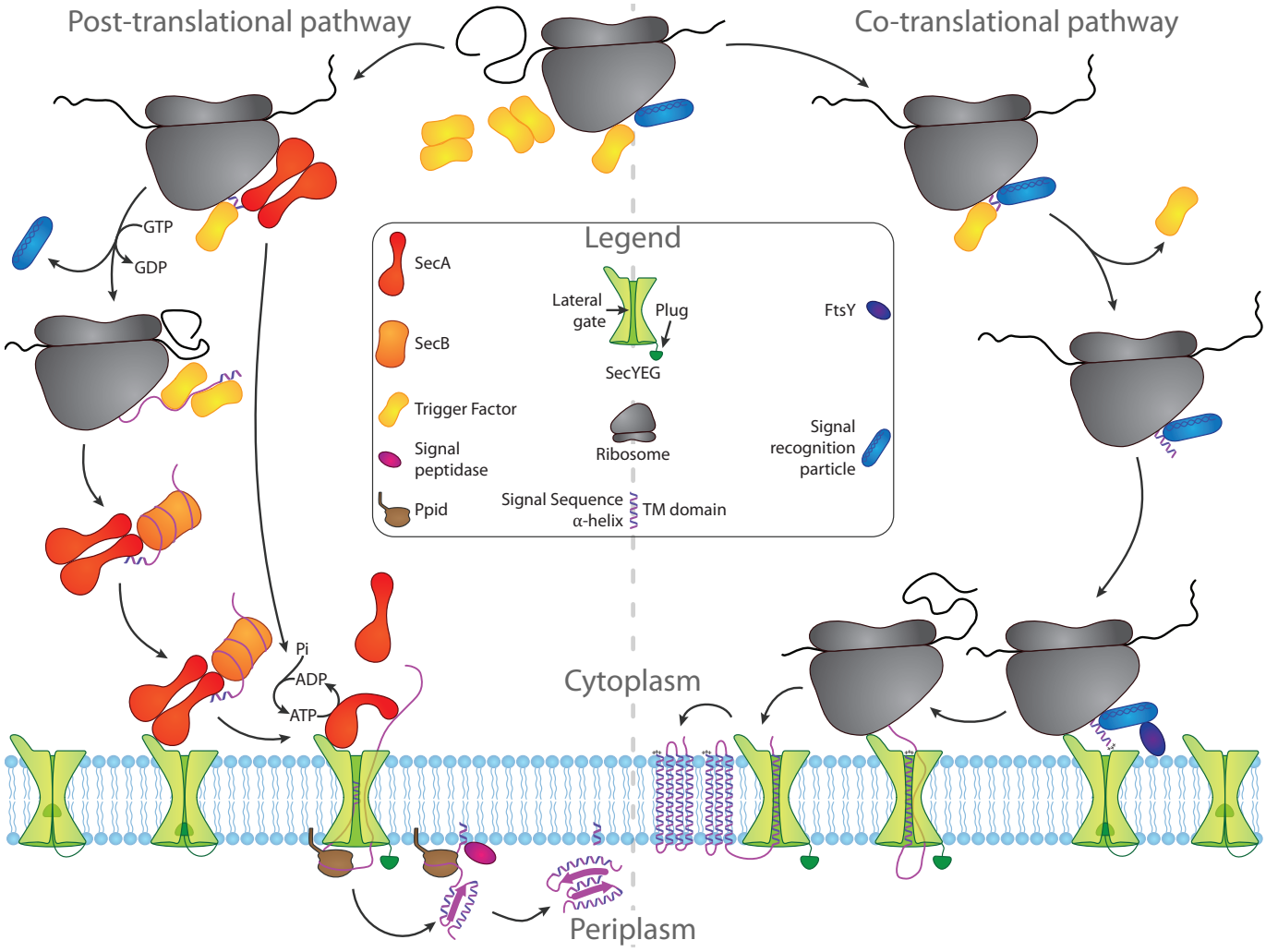


FIGURE 1.5: Schematic representation of the translocation pathways. TF and SRP wait for a signal sequence to emerge from the ribosome exit tunnel. **Co-translational pathway**: SRP binds the signal sequence and TF is released. The translation is arrested until the RNC-SRP-FtsY complex binds SecYEG that switches to the pre-open conformation. The first transmembrane domain is inserted into the lateral gate, inducing the open conformation of SecYEG. The transmembrane segments are sequentially released into the membrane through the lateral gate. **Post-translational pathway**: TF binds the signal sequence and additional TFs help keeping the pre-protein unfolded. It is then handed over to SecB-SecA complex, the pre-protein wrapping around SecB to be delivered to SecYEG. Alternatively, SecA takes the protein directly to SecYEG. The binding of SecA to the translocon induces the pre-open conformation. ATP hydrolysis cycles release one protomer of SecA, initiates the translocation and the signal sequence is inserted in the lateral gate to induce the open conformation. Ppid prevents the pre-protein from sliding backward and help the folding on the periplasmic side. The signal peptidase cleaves the signal sequence to release the folded protein into the periplasm.

The binding of translocation partner (e.g ribosome, Sec62/63) primes the channel, triggering the partial opening of the lateral gate (Panzner *et al.*, 1995; Gogala *et al.*, 2014) in a similar way as described for prokaryotes (Rapoport *et al.*, 2017). The channel fully opens when the lateral gate is occupied by a signal sequence (or transmembrane domain) and allows the protein to slide through toward the lumen of the endoplasmic reticulum. During co-translational translocation, the protein elongation by the ribosome provides sufficient energy to drive unidirectional motion of the protein through the pore. In both co- and post-translational translocation, the Brownian diffusion is biased by the luminal BiP ATPase (Matlack *et al.*, 1999). This ATP-bound chaperone waits for a polypeptide to emerge from Sec61; the interaction with the membrane protein Sec63 triggers ATP hydrolysis and subsequent unspecific binding of Bip to the translocated region Misselwitz *et al.*, 1998, thus preventing the peptide from sliding backward (Matlack *et al.*, 1999). As the peptide moves forward, another BiP binds and the process is repeated until the peptide is fully translocated and the ADP is exchanged for ATP to open Bip and release the protein in the lumen.

1.1.4 A larger translocation competent complex

The accessory subunits

The SecYEG translocon, together with the ribosome or SecA, is sufficient to allow protein translocation; these factors alone cannot explain how translocation is affected by the Proton Motive Force (PMF) (Schiebel *et al.*, 1991) or how SecYEG would control the folding of large transmembrane helices bundles (Nagamori *et al.*, 2004). Other membrane proteins can therefore join SecYEG to promote translocation, assist membrane insertion, folding and complex assembly (Duong and Wickner, 1997a; Sachelaru, Petriman, Kudva, P. Kuhn, *et al.*, 2013; Schulze *et al.*, 2014; Komar, Botte, *et al.*, 2015; Berger, Q. Jiang, *et al.*, 2017).

YidC: YidC is a membrane protein that is present in similar stoichiometry as SecYEG in *E.coli* (G.-W. Li *et al.*, 2014). It is described as an insertase that catalyses by itself the transmembrane insertion of newly synthesised membrane proteins. Among its 6 transmembrane helices (Sääf *et al.*, 1998), the C-terminal ones are of particular importance for the insertase activity (F. Jiang *et al.*, 2003), while the large periplasmic domain (Sääf *et al.*, 1998) may contribute to the folding of periplasmic regions of substrate proteins (Dalbey *et al.*, 2014) to which it interacts through a cleft in the structure (see Figure 1.6).

While free YidC have been found as oligomers, it was shown it operates and interact as a monomer in the context of Ribosome Nascent Chain (RNC) complexes (Kedrov *et al.*, 2013; Seitl *et al.*, 2014). YidC has been linked to the SRP targeting pathway (Facey *et al.*, 2007). Interestingly, *E.coli* YidC, as opposed to its homologues that have an extended C-terminal domain, has lost the ability to directly bind to ribosomes (Jia *et al.*, 2003; Kedrov *et al.*, 2013), explaining its dependency on the SRP pathway. Some rare proteins can also be targeted to YidC independently of the SRP pathway

(M. Chen *et al.*, 2002) but the mechanism remains elusive and the number of substrates using the YidC-only membrane integration pathway, mostly short translocated regions, are limited (Serek *et al.*, 2004; van der Laan *et al.*, 2004).

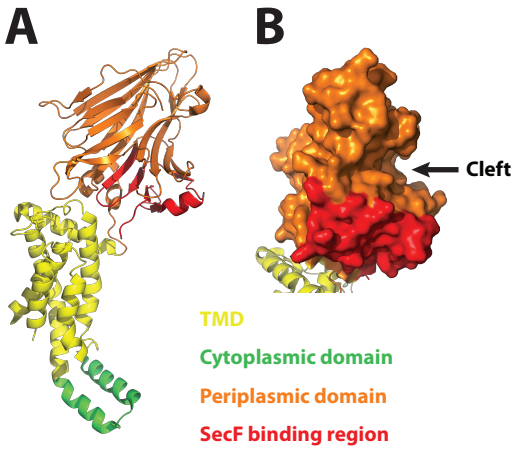


FIGURE 1.6: Crystal structure of *E.coli* YidC (pdb: 3WVF, Kumazaki *et al.*, 2014). **A**: Full length YidC with its transmembrane domains (TMD, yellow), the short cytoplasmic domain (green), periplasmic domain (red) and its region that binds to SecF (orange). **B**: Surface representation of the periplasmic domain showing the cleft that binds preprotein.

For other proteins, such as subunits of the ATP synthase, YidC interaction with the core Sec translocon is mandatory (Yi *et al.*, 2003) and happens through a physical binding of YidC to SecYEG (Scotti *et al.*, 2000). This interaction has been mapped, revealing YidC contacts the lateral gate of SecY with its hydrophobic domains facing the lateral gate. This suggests that YidC would prevent uncontrolled exit of hydrophobic peptides into the membrane (Sachelaru, Petriman, Kudva, P. Kuhn, *et al.*, 2013). Upon RNC binding, YidC is partially displaced and allows transmembrane domains to exit through the lateral gate while restricting premature release of transmembrane domains to favour the packing that is required for consistent folding (Sachelaru, Petriman, Kudva, P. Kuhn, *et al.*, 2013).

The dependence on YidC for the translocation of membrane protein has been linked to the presence of negatively charged periplasmic regions (Kiefer and A. Kuhn, 1999) and imbalanced charge distribution in the transmembrane segments of YidC substrates (Gray *et al.*, 2011). Most importantly, YidC is required for the insertion of domains with low hydrophobicity indexes (Ernst *et al.*, 2011) and proteins with segments who do not follow the "positive inside" rule (Gray *et al.*, 2011), which corroborate the observation that YidC assists the SecYEG translocon to support the lateral exit of transmembrane domains whose insertion is usually not favoured. Finally, the correct assembly of certain multicomponent membrane complexes is dependent on YidC (Wagner *et al.*, 2009), making it a versatile protein acting at various levels and in different ways to promote and ensure correct membrane protein integration and folding.

A model for the function of YidC is represented in Figure 1.8, panel B.

SecDF-YajC: SecD, SecF and YajC are membrane proteins that are encoded by the same operon (Gardel *et al.*, 1990; K. Pogliano and Beckwith, 1994), whose depletion leads to severe defects of translocation (J. Pogliano and Beckwith, 1994). Paradoxically, these proteins do not enhance *in vitro* translocation (Matsuyama *et al.*, 1992) raising the question of their function in the translocation process.

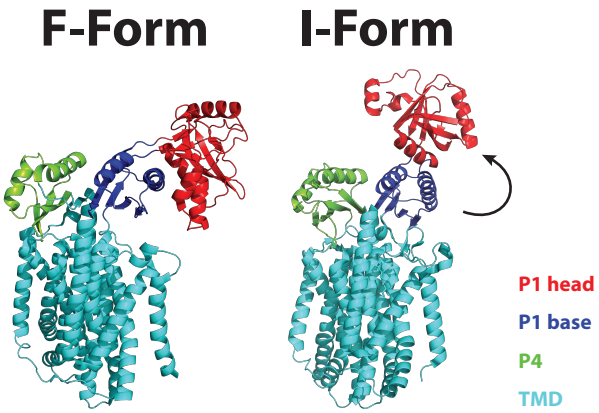


FIGURE 1.7: Cartoon representation of the crystal structure of *E.coli* SecDF homologues. **A**: SecDF homologue from *T.thermophilus* (pdb: 3AQP, Tsukazaki *et al.*, 2011) representing the F-form. The P1 head (red) is linked to the p1 base (blue) by a flexible hinge. The P4 domain (green) is linked to the transmembrane domains (TMD, cyan). **B**: For the I form (pdb: 5XAP, Furukawa *et al.*, 2017), the P1 domains rotates upward (black arrow) to pull the pre-protein from the protein conducting channel.

SecDF-YajC do not allow translocation by themselves and their activity is directly linked to the core translocon and YidC, with which they interact (Scotti *et al.*, 2000; Duong and Wickner, 1997b). A suggested role of this heterotrimeric complex is to promote the release of translocated protein through interactions between a periplasmic domain (P1) and the translocated protein (Nouwen, Piwowarek, *et al.*, 2005). The P1 domain, which can rotate around the flexible hinges attached to the P1 base, transits from a bent-pore facing conformation (named the *F*-form) to a straight conformation (named the *I*-from) in a pulling motion (Tsukazaki *et al.*, 2011, see Figure 1.7). This complex does not have ATP hydrolysis activity and ATP is not present in the periplasm. Another source of energy is therefore required to perform those motions. SecDF have been linked to the electrochemical dependency of translocation events (Arkowitz and Wickner, 1994; Duong and Wickner, 1997b). Studies showed the translocation efficiency of SecYEG-SecDF-YajC-YidC containing proteoliposomes is enhanced (as compared to SecYEG alone) only in the presence of a electrochemical gradient (Schulze *et al.*, 2014), supporting the hypothesis that SecDF-YajC favours the late stages of protein translocation using the proton-motive force.

The current working model, suggests that SecA or the ribosome push the pre-proteins through the SecYEG pore while SecDF in *F*-form (Tsukazaki *et al.*, 2011) interacts with the emerging protein on the periplasmic side before switching to the *I*-form (Furukawa *et al.*, 2017), pulling the pre-protein. This transition is accompanied by the de-protonation in the cytoplasm of an aspartic acid residue (Asp340) and an opening

of the proton channel, therefore using the proton-motive force as a source of energy. The pre-protein is then released, the proton channel closed and the P1 head adopts the *F*-form.

Finally, while YidC has been shown to interact directly with SecY (Scotti *et al.*, 2000), a stronger physical interaction with SecF has also been identified (Kumazaki *et al.*, 2014; Nouwen and Driessen, 2002). It has therefore been proposed that SecDF-YajC could act as a intermediate that promotes the interaction of YidC with SecYEG, that is otherwise relatively weak. This explains the strongly deleterious effects of SecDF-YajC depletion as the binding of YidC to SecYEG is essential for efficient translocation (Samuelson *et al.*, 2000).

A summary of SecDF functioning is shown in Figure 1.8, panel A.

The translocation holoenzyme

As discussed in subsection 1.1.2 and showed biochemically (Osborne and Rapoport, 2007; Schulze *et al.*, 2014; L. Li *et al.*, 2016), the SecYEG core translocon is capable of performing both protein secretion and membrane protein integration, provided it interacts with its cytoplasmic partners. Several experiments have indicated that the translocation can be enhanced by the association of accessory proteins to the core SecYEG translocon, such as the SecDF-YajC complex (J. Pogliano and Beckwith, 1994; Duong and Wickner, 1997a; Nouwen and Driessen, 2002) or YidC (Scotti *et al.*, 2000; Urbanus, Scotti, *et al.*, 2001; Nouwen and Driessen, 2002). However, it remained unclear whether these accessory subunits cooperate (Xie *et al.*, 2006) by joining with the core translocon to form a holoenzyme competent for protein translocation : the SecYEG-SecDF-YajC-YidC Holotranslocon (HTL).

This question remained unanswered mostly because of the difficulty in over-expressing, purifying and characterising this complex. Over-expression of membrane proteins can be difficult because of toxicity and restricted access to the membrane (Miroux and Walker, 1996). In the case of multicomponent systems, a major limitation is also the stoichiometry as the copy number of plasmids in cells cannot be fully controlled, therefore introducing a bias at the genetic level. The latter issue has been overcome by the development of the ACEMBL system (Bieniossek *et al.*, 2009), a breakthrough in the field of multicomponent system bacterial expression, followed by the multiBac eukaryotic expression system for insect cells (Berger, Garzoni, *et al.*, 2013). ACEMBL makes use of Sequence and Ligation Independent Cloning (SLIC) that allows reliable integration of the genes of interest into donor and acceptor plasmids that are then fused using LoxP sites and Cre recombinase. The selection of plasmids is carried out by using a combination of conditional origins of replication and antibiotic resistance genes, ensuring that only fused plasmids are selected. A plasmid (see Figure 1.9) that contains all seven subunits, with a set of purification tags and inducible promoters has been generated by fusing three plasmids (containing SecYEG, SecDF-YidC and YajC respectively). This plasmid allows the over-expression of the Holotranslocon (Bieniossek *et al.*, 2009; Schulze *et al.*, 2014).

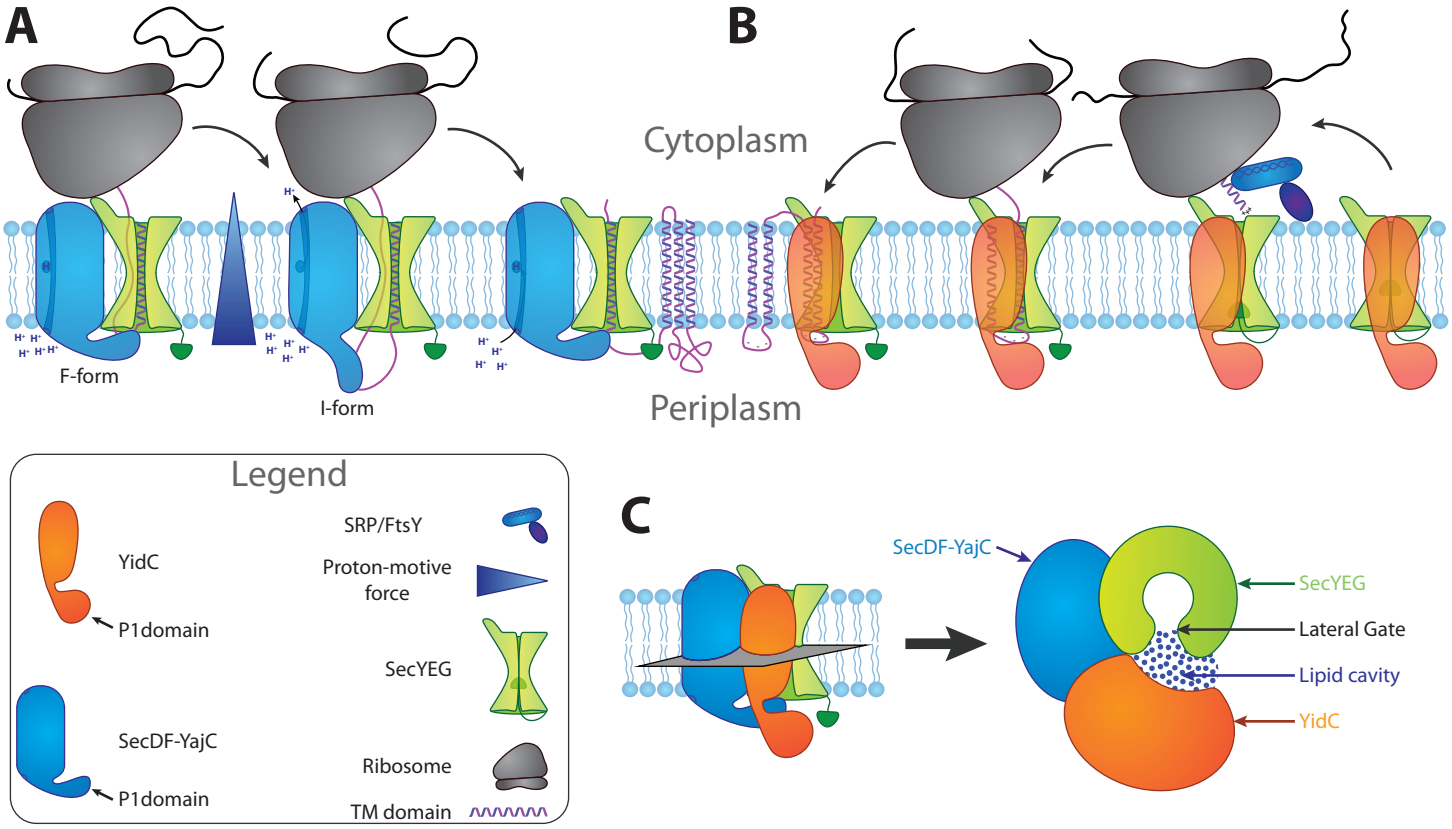


FIGURE 1.8: Schematic representation of the accessory subunits and the holoenzyme. **A**: SecDF-YajC bind to SecYEG in F-form, its P1 domain sitting on top of the protein conducting channel exit. It interacts with the nascent protein, the aspartic acid (Asp340) is deprotonated toward the cytoplasm and the complex switches to the I-form, the head pulling on the substrate and helping the translocation of the large periplasmic loop. A proton is captured from the periplasm by the aspartate residue to return in F-form and perform another cycle. **B**: YidC is bound to SecYEG and occludes the lateral gate. Upon ribosome nascent-chain binding, YidC is displaced so transmembrane domain can leave the protein conducting channel. The periplasmic domain of YidC assists the translocation (typically for negatively charged periplasmic loops) and the transmembrane helices can be retained by YidC before being released in the membrane. **C**: SecDF-YajC and YidC can both bind SecYEG to form the Holotranslocon. The grey plane represents the transverse section diplayed on the right. YidC contacts SecDF-YajC and is positioned so that its hydrophobic domains face the lateral gate. A cavity filled with lipid may help the folding of membrane proteins.

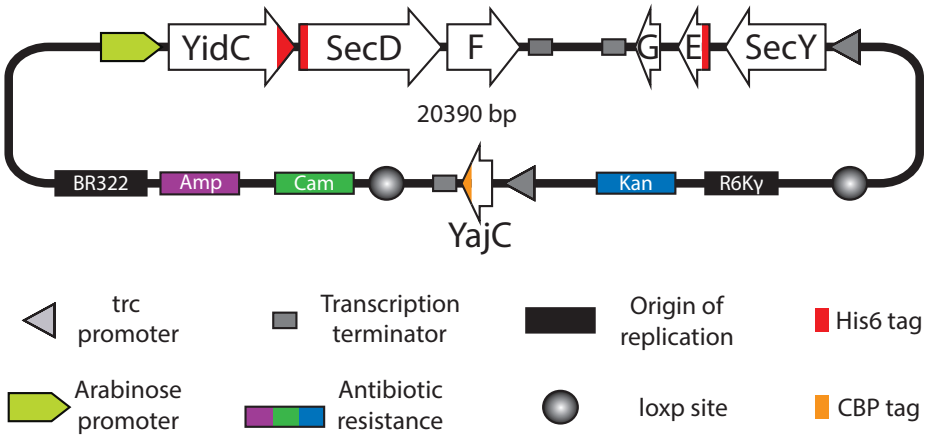


FIGURE 1.9: Schematic representation of the Holotranslocon plasmid map, adapted from Bieniossek *et al.*, 2009.

As a result of this new tool, it was shown that HTL can be purified (Schulze *et al.*, 2014) and that the complex contains one copy of each subunit (SecYEG-SecDF-YajC-YidC) (Botte *et al.*, 2016). Biochemical approaches have demonstrated that HTL is capable of translocating proteins across and into the membrane (Schulze *et al.*, 2014). However, when compared to SecYEG, the SecA-driven ATP energy conversion into transport is less efficient for HTL, despite similar affinities. HTL has a higher dependency on the proton-motive force, an observation that can probably be attributed to the presence of the PMF-dependant subunits SecDF-YajC (Schulze *et al.*, 2014). On the other hand, HTL shows a 6-fold higher affinity for the ribosome and is more efficient at integrating many membrane protein substrates into the membrane (Schulze *et al.*, 2014), an activity that could be attributed to the presence of the YidC insertase. Further structural studies proposed a quasi-atomic model where YidC and SecD are arranged so that a flexible cavity filled with lipid is formed in vicinity of the lateral gate (see Figure 1.10). It has also been suggested that SecD and YidC periplasmic heads are ideally positioned to prevent the backsliding of translocating proteins (Botte *et al.*, 2016). These authors therefore proposed a model where HTL would be a complex specialized for membrane protein integration. It would act as a chaperone by (1) providing nascent proteins with a protected hydrophobic environment where the accessory subunits could assist the folding, (2) help retain transmembrane domains close to the lateral gate to favour efficient packing of subsequently translocated domains, before being freely released in the lipid bilayer. HTL's reduced efficiency for secretion, enhanced membrane protein integration activity together with the analysis of protein copy number per cell (4 SecYEG for 1 SecDF and 3 YidC, G.-W. Li *et al.*, 2014), suggest that the translocation machinery can dynamically be rearranged depending on the type of substrate, SecYEG dimers being preferentially used for secretion while membrane proteins would be handled by SecYEG-YidC (short translocated domain) or the HTL (larger proteins).

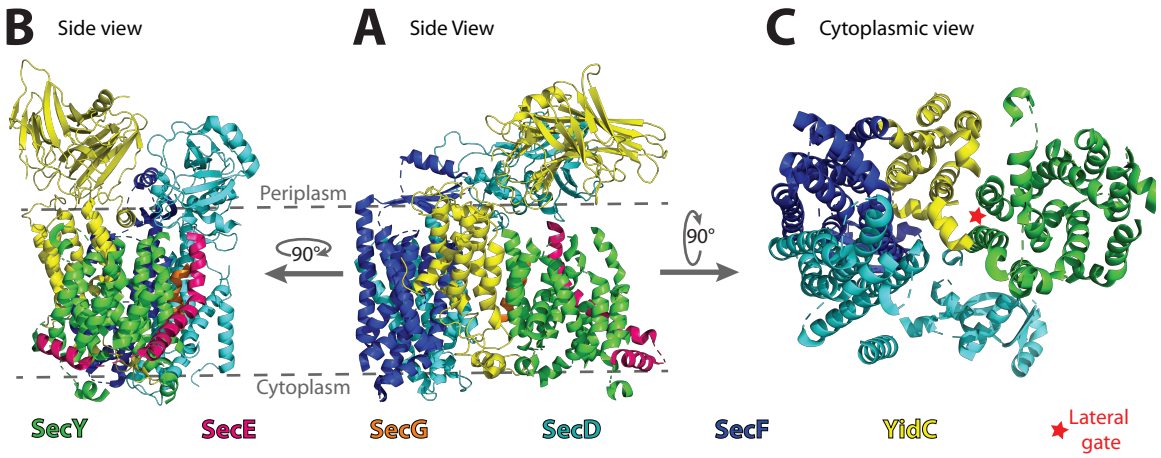


FIGURE 1.10: Quasi-atomic structure model of the Holotranslocon proposed by Botte *et al.*, 2016 from cryo-electron microscopy medium resolution envelopes and individual crystal structure docking (except YajC that could not be located)(PDB:5MG3). The unstructured loops of SecY, SecE, SecG and SecF are not represented for clarity. **A**: Side view of HTL showing the relative arrangements of the subunits. **B**: Side view of HTL showing the ideal disposition of SecD and YidC's periplasmic domains above the exit of the protein conducting channel. **C**: Cytoplasmic view of HTL showing YidC faces the lateral gate of SecY. SecE-G are not showed.

1.1.5 Structural analysis of the translocation machinery in bacteria

Over the past years, the structural analysis of the translocation machinery has relied on the use of crystallography (Xu *et al.*, 2000; Tanaka *et al.*, 2015; Breyton *et al.*, 2002; Tsukazaki *et al.*, 2011; Kumazaki *et al.*, 2014; L. Li *et al.*, 2016), nuclear magnetic resonance (Gelis *et al.*, 2007; Huang *et al.*, 2016) and cryo-electron microscopy (Frauenfeld *et al.*, 2011; Singh *et al.*, 2014; Q. Zhou *et al.*, 2012; Botte *et al.*, 2016; Jomaa *et al.*, 2016). Recently, small-angle neutron scattering approaches in parallel with electron cryo-microscopy, have allowed the structural characterization of the Holotranslocon complex (Botte *et al.*, 2016). The implementation of SANS in this study has also provided unique information and shed light on the presence of a central cavity filled with lipids, information that could not be accessed due to the relatively low resolution achieved by cryo-electron microscopy (14 Å).

While the resolution achieved by small-angle neutron scattering (SANS) cannot compare with that of crystallographic, electron microscopy or nuclear magnetic resonance studies, it comes with its own unique advantages. In addition to not requiring a crystal, being compatible with large particles and measured in close to physiological conditions, SANS can uniquely exploit contrast variation and isotope labelling to answer distinct questions on multicomponent complexes (further discussed in section 1.2). In the case of the translocation machinery, it therefore allows multi-protein complexes that are differently labelled (see chapter 3) to be investigated or the large multi-component transmembrane complex solubilized with surfactants (see chapter 2).

1.2 Small Angle Neutron Scattering for biology

Neutrons are quantum objects and can therefore be defined as a wave with a given wavelength (λ) or a particle with a given mass (m) and velocity (v). These properties are related by De Broglie's equation:

$$\lambda = \frac{h}{mv} \tag{1.1}$$

where h is Planck's constant.

When a neutron encounters an atom, it may or may not interact with the nucleus, the probability of which being defined as the nucleus-dependent cross-section. The interaction can be either an absorption, nuclear reaction characterized by the integration of the neutron within a nucleus or a scattering event which can occur in different ways:

- **Elastic**: the incoming neutron will see its trajectory deviated but its energy (wavelength and velocity) is conserved.
- **Inelastic**: the deviation is accompanied by a transfer of energy to the scattered neutron.
- **Coherent**: when considering a group of atoms, neutrons scattered from different nuclei interfere constructively, therefore providing information on their relative position at a given time.
- **Incoherent**: A neutron scattered by an atom does not interfere constructively with any other scattered neutron.

Incoherent scattering will therefore give information on individual atoms, such as their energy state, while coherent scattering will give information about the relative position (elastic) or energy state (inelastic) about a group of atoms. Inelastic scattering is used for the study of the dynamic states of a system while elastic coherent scattering is used for structural studies (N. Zaccai *et al.*, 2017). Only the latter case will be considered in this chapter.

1.2.1 Scattering length density and contrast

Neutrons can be scattered by all atomic nuclei and the amplitude of the neutron waves scattered by a given nucleus is defined by the scattering length. This physical value has a coherent component (noted b_c) that describes the capacity of a scattered neutron to interfere constructively with neutrons scattered by other nuclei and an incoherent component that essentially generates an isotropic background in coherent experiments (Jacrot and Zaccai, 1981). In the case of molecules with a defined volume (V), its scattering power is described by the coherent scattering length density (ρ) as:

$$\rho = \frac{\sum b_c}{V} \tag{1.2}$$

and is therefore directly linked to its composition.

In structural solution studies, the signal arises from both the particle of interest and the solvent, which is considered as noise to be subtracted from the scattering pattern. The amount of information coming from the particle that can be extracted from these experiments therefore depends on the difference of scattering intensity between the solvent and the particle, which is defined as:

$$\Delta\rho = \rho_{particle} - \rho_{solvent} \tag{1.3}$$

and is referred to as the contrast (N. Zaccai *et al.*, 2017). It is an important consideration in coherent elastic neutron experiments, as it is the only factor that can be modulated experimentally, independently of the instrument set-up and particles of interest (see Equation 5.19).

1.2.2 Importance of deuterium for biological studies with neutrons

Neutron diffraction and reflection have been extensively used in biology to address biological problems. Neutrons are difficult to produce and sometimes underused because of the relatively low flux achieved by reactors and spallation sources in comparisons with synchrotron X-rays. However, neutrons have unique advantages for the study of biological systems because of the way they interact with matter. Neutrons are scattered by nuclei and are not linearly affected by the atomic number but by the spin states of the nucleus; all biologically relevant atoms have comparable scattering lengths. This characteristic allows neutrons to detect light atoms but also to be sensitive to different isotopes (Jacrot and Zaccai, 1981). This last point is of crucial importance for the study of biological systems since it is possible to substitute hydrogen by its heavier isotope deuterium, with significant benefits:

- Hydrogen has a negative coherent scattering length ($b_{coh} = -3.741 \times 10^{-15}$ m), while it is positive for deuterium ($b_{coh} = 6.671 \times 10^{-15}$ m, see Figure 1.11). This has major consequences for neutrons experiments, for example in the case of neutron crystallography, deuterium is clearly visible while hydrogen appears as a negative density and tends to cancel the signal from surrounding atoms.
- When the scattering of hydrogen is dominated by its incoherent contribution ($b_{inc} = 25.274 \times 10^{-15}$ m), deuterium has a significantly lower incoherent scattering length ($b_{inc} = 4.040 \times 10^{-15}$ m). In coherent elastic experiments, using deuterium therefore allows the incoherent contribution to be significantly reduced and consequently improve the signal to noise ratio for coherent studies.
- Deuterium has the same chemical properties as hydrogen and can therefore replace it with minor effects on the function of the molecule. However, several cases reported slower kinetics for deuterated protein (Jasnin *et al.*, 2008; Pinchuk and Lichtenberg, 2017; Trimmer *et al.*, 2017); because deuterium is twice as heavy as

hydrogen, the covalent bond involving a deuterium has a lower zero-point vibrational energy, making it harder to break (Scheiner and Čuma, 1996). Furthermore, some proteins will have a higher propensity for aggregation when deuterated or exposed to a deuterated solvent; this observation is attributed to the hydrophobic effect (Jasnin *et al.*, 2008).

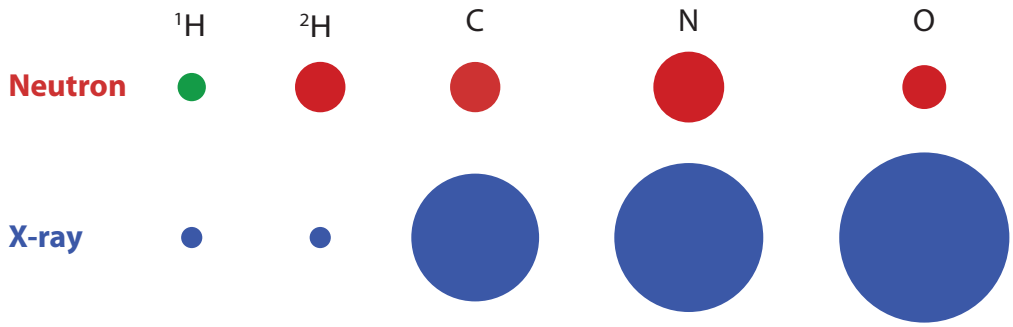


FIGURE 1.11: Representation of the coherent neutron (in red) and X-ray (in blue) scattering lengths of most commons biological atoms. The negative neutron scattering length of hydrogen is represented in green.

Contrast variation and contrast match point

As discussed in subsection 1.2.1, contrast is a key factor in the success of a scattering experiment. Since it is described as the difference between the scattering length density of the solvent and particle of interest (noted $\Delta\rho$), a straightforward way to modulate it is to make use of isotopic substitution on the solvent. Since hydrogen has a negative coherent scattering length, H_2O generates an overall negative signal while D_2O has a positive signal; the scattering length density of a mixed solvent is the weighed sum of both scattering length densities. Most biological systems are measured in aqueous solvent, where the scattering length density can be modulated by mixing water (H_2O) with heavy water (D_2O); it is therefore possible to finely tune the contrast of an individual experiment.

A special case of contrast occurs when the scattering length density of the solvent and particle of interest are equal. It is called the contrast match point (CMP) and is defined as :

$$CMP = \rho_{solvent} - \rho_{particle} = \Delta\rho = 0 \tag{1.4}$$

Experimentally, the buffer and particle are indistinguishable in this condition. Because the scattering length density depends on the atomic composition of the particle (see subsection 1.2.1), the contrast match point of a particle can directly be related to its composition.

Biological molecules are mostly composed of hydrogen, carbon, oxygen, sulphur and phosphate, all together making up more than 99% of their masses. The distribution of these atoms however varies widely (between protein, lipids, DNA...), thus implying the scattering length density of biomolecules, and therefore their contrast match point, is different (Figure 1.12). Contrast variation experiments are therefore of fundamental

interest for the study of complexes as it is possible to selectively make one of the component invisible if its atomic composition is different enough. For example, most proteins are matched out at 40-42% D_2O, while DNA matches out around 65% D_2O so that when studying a DNA-protein complex, a measurement at 0% allows both component to be observed, but at 42% D_2O the protein is invisible and at 65% D_2O the DNA is matched out. Conversely, in the case of a composite particle of unknown composition (for example a membrane protein solubilized with detergents), by knowing the scattering length density of each component and measuring the contrast match point of the complex, it is possible to decipher its relative volume composition.

Deuteration of Biomolecules

The other way of modulating the contrast is to alter the scattering length density of the particle of interest by using deuterium substitution. It is then important to make the difference between two classes of hydrogen atoms :

- **Exchangeable labile hydrogen atoms**: or labile hydrogen atoms make low energy covalent bonds, typically with atoms that are electronegative such as oxygen and nitrogen. These atoms can be easily substituted by simply exposing them to labile deuterium such as the D_2O in a solvents. The fact that scattering length density of a biomolecules is dependent on the fraction of D_2O, as seen in Figure 1.12, illustrate this point.
- **Non-exchangeable hydrogen atoms**: are not labile, when involved in covalent bonds with carbon atoms for example. It is important to note that some labile hydrogen atoms cannot be exchanged as the tertiary structure may involve that some of them are buried in the protein and therefore not exposed to the solvent. These hydrogen atoms therefore need to be substituted during the biogenesis of the molecule. This class of hydrogen is considered when we refer to biomolecular deuteration.

In order to use deuteration in biology, it is important to understand how the deuterium are integrated into the biomolecules. To date, the *in vivo* partial deuteration of proteins has been studied in detail and it is possible to modulate the growth conditions of bacteria to achieve a specific level of protein deuteration (Leiting *et al.*, 1998). It is therefore possible to tune the culture conditions to achieve different regime : (1) match out labelling which deuterates the protein so that it can be matched out in 100% D_2O, (2) perdeuteration which correspond to a full deuteration of the protein. The first regime is often used in SANS experiments for studying protein complexes where each differently deuterated protein can be matched out independently. The second is used for crystallography where it allows the coherent signal to be maximized and facilitates analysis; this allows the use of smaller crystals (Haertlein *et al.*, 2016).

However, other classes of biomolecules (such as lipids and nucleic acids) have not been characterized in the same way. A few studies based on the production of fully deuterated synthetic lipids have been described (de Ghellinck *et al.*, 2014), nucleic acids (Shotton *et al.*, 1997; Jünemann *et al.*, 1996), or even selective deuteration of lipid moieties (Maric *et al.*, 2015).

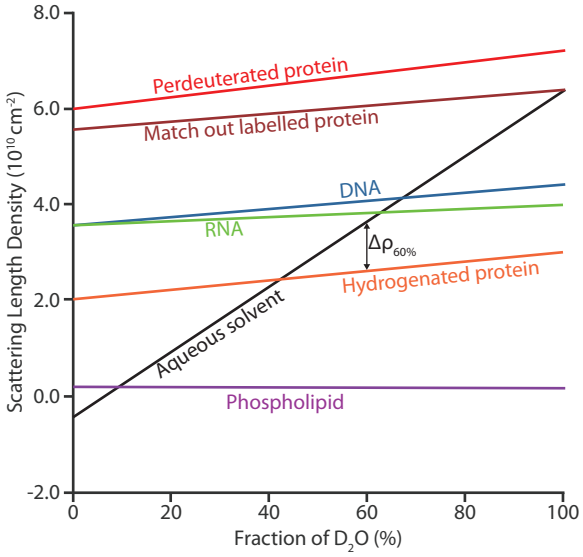


FIGURE 1.12: Scattering length densities of biomolecules at different D_2O solvent fraction. The black line represent the scattering length density of a aqueous solvent composed of an increasing fraction of D_2O. The intersection of the solvent line with the individual curves of biomolecules represent its contrast match point. The difference between the solvent line and any other line represent the contrast (the double arrow represents the contrast of hydrogenated protein with 60%D_2O solvent).

1.3 Aim of the study

1.3.1 Investigation of the bacterial translocation mechanisms

Translocation is a complicated process that occurs following various pathways and involves many partners. Recent structural and biochemical studies have shed the light on major steps of the translocation process (Tsirigotaki *et al.*, 2017; Rapoport *et al.*, 2017); however many questions still remain to be answered before secretion and membrane protein integration can be fully understood.

Investigating the structural and dynamic functions of the Holotranslocon

The notion of a super complex that is larger than just SecYEG has long been considered in translocation studies since it has been shown that other membrane partners were of major importance for the survival of bacteria (Arkowitz and Wickner, 1994; Duong and Wickner, 1997b; Sääf *et al.*, 1998). These ideas have further been enhanced by several independent reports of physical interaction between the core translocon and accessory proteins (J. Pogliano and Beckwith, 1994; Duong and Wickner, 1997a; Scotti *et al.*, 2000; Urbanus, Scotti, *et al.*, 2001) and were subsequently demonstrated in the work of Botte *et al.*, 2016.

If the existence of a translocation holoenzyme, the Holotranslocon, is proven, its exact role within the cellular translocation machinery remains to be clearly determined.

Given the disparity of subunit copy number *in vivo* (K. Pogliano and Beckwith, 1994; Urbanus, Fröderberg, *et al.*, 2002; G.-W. Li *et al.*, 2014), it is possible that this complex may only assemble transiently. Functionally, it is not understood why the Holotranslocon is less efficient at promoting SecA-dependent translocation than SecYEG. Making use of a wide range of biochemical and biophysical characterization methods, deuterium labeling and SANS, the structural features of the Holotranslocon have been investigated and the impact of an interaction with its cytosolic partner SecA evaluated.

The Sec proteins and their role in translocation

The post-translational pathway requires that the newly synthesised protein remains in an unfolded state so it can be handled by the translocation machinery. A representative actor of this mechanism is the SecB protein that is capable of binding unfolded proteins and protects their hydrophobic segments from the aqueous solvent (Huang *et al.*, 2016). This "holdase" has also the capacity to bind to the SecA ATPase (Hartl *et al.*, 1990) and therefore contributes to the targeting of pre-protein to the translocation complex. The stoichiometry of this complex as well as its structure are not fully understood. Through biochemical and biophysical characterization coupled to deuterium labeling and SANS, a translocation competent complex containing SecB, SecA and the pre-protein have been analysed.

1.3.2 Development of *in vitro* labelling strategies

Comprehensive study of protein, lipid and nucleic acid *in vivo* deuteration

Although it is possible to produce fully deuterated lipids and nucleic acids, at the moment, no study reports the correlation between the growth medium and the level of deuteration of these biomolecules. Even though the natural contrast arising from the different chemical composition of biomolecules can be sufficient for certain studies, a higher contrast would facilitate data interpretation, and experiments could be envisaged in which use is made of differential deuteration levels to distinguish several biomolecules within a single complex, or matching out several components biochemically different at once.

This issue is addressed in this PhD work and the correlation between the condition of culture and the deuteration level achieved has been analysed. As described before (Leiting *et al.*, 1998), not only the nature of the solvent is important but also the carbon source, which is, in prototrophic organism such as *E.coli*, the main source of matter intake. The effect of both factors were evaluated (see section 4.1).

In order to investigate the deuterium/hydrogen content of various biomolecules produced in different deuterated conditions, different physical parameters were considered:

- The difference in neutron scattering length (see section 1.2.2) allows to indirectly quantify the $^{1}H/^{2}H$ by measuring the neutron contrast match and relating it to the atomic composition.

- The difference in mass due to the extra neutron in deuterium. It is therefore possible to measure the mass difference between hydrogenated and deuterated sample and directly relate it to the level of deuteration.
- ^{1}H and ^{2}H do not have the same spin value (+1/2 and +1 respectively). 1D ^{1}H NMR allows to measure the relative quantity of hydrogen atoms in each sample and therefore convert it to deuteration levels.

Development of selective deuterium labelling strategies

Having total control on the labeling of each type of biomolecule independently is a substantial step forward for the use of neutrons in biology; however it implies that the different components are produced independently and reassembled afterwards, which is not always possible for some systems. We can for example consider membrane proteins, which stability may depend on the presence of tightly bond lipids (Palsdottir and Hunte, 2004). We can also consider nucleic acid-bound proteins whose stability or structure may depend on the binding of their nucleic acid partner. In order to address this issue, we investigated the possibility to achieve different levels of deuteration within the same culture.

At the moment, the deuteration protocols mostly rely on the use of a deuterated solvent (typically D_2O) and carbon source (glycerol, methanol, etc.). In both cases, the deuterium provided by these means will be found in all biomolecules as they are the primary sources of metabolites supplied to micro-organisms. To introduce an isotopic label in one type of biomolecule, the specific synthesis pathways need to be targeted. A major limitation for this strategy is the availability of deuterated precursors as well as their price; as a proof of principle, the reverse experiment was done. We adapt the currently available deuteration protocols (see subsection 5.1.2) to achieve specific hydrogen labeling of protein in fully deuterated conditions. This study focused on deuterium/hydrogen substitution, but could be adapted to other atoms relevant for biophysical studies.

Chapter 2

Preparation and analysis of the Holotranslocon

Résumé en français

Sur-expression et purification de l'Holotranslocon

L'objectif de l'étude étant d'analyser l'Holotranlocon par des méthodes biophysiques, et notamment le SANS, la préparation de l'échantillon est une étape primordiale. L'expression et la purification d'un tel complexe membranaire (composé de 7 sous-unités et 34 hélices trans-membranaires) est une tâche difficile. Nous avons analysé les étapes de production de l'échantillon afin de les optimiser pour obtenir une solution protéique de qualité suffisante pour être analysée par des méthodes biophysiques.

Expression de l'Holotranslocon : Nous avons montré que la surexpression de l'Holotranslocon en cellules *E. coli* est une première étape qui peut s'avérer décisive pour le succès de sa production. La complexité de ce système le rend sensible aux changements de conditions de culture et aux types de milieux utilisés. Nous avons notamment montré que l'utilisation de cultures à haute densité affecte de façon dramatique la stœchiométrie des sous-unités, posant un frein significatif au marquage isotopique par le deutérium. Nous avons également comparé une souche d'expression classique d'*E. coli* avec une souche spécialisée dans l'expression de protéine membranaire.

Méthodes biophysiques de caractérisation : Nous avons appliqué des techniques biophysiques variées pour évaluer la qualité d'un échantillon : "dynamic light scattering" (DLS), "size exclusion chromatography" (SEC), "Multi angle laser scattering coupled to size exclusion chromatography (SEC-MALS)", ultracentrifugation analytique (AUC), spectrométrie de masse à temps de vol avec ionisation assistée par matrice (MALDI-TOF). Ces techniques ont été utilisées pour vérifier la qualité de l'échantillon lors des optimisations.

Méthode du purification : Dans l'optique d'optimiser les rendements et la qualité de l'échantillon, nous avons comparé deux stratégies de purification publiées. La première se base sur deux étapes successives de purification par affinité mais nous avons montré que la quantité de protéine produite est souvent insuffisante pour permettre des analyses ultérieures. La seconde méthode se base sur une seule étape de purification par affinité suivie d'une étape de chromatographie à exclusion de taille. Si ce protocole permet de produire de plus grandes quantités d'échantillons, son manque de spécificité génère une grande hétérogénéité dans l'échantillon final.

Solubilisation du complexe : Nous discutons du détergent utilisé (DDM) pour l'extraction du complexe des membranes biologiques et avons mis en évidence une série de nouveaux détergents qui pourraient se substituer efficacement à celui utilisé dans cette étude. Nous avons également tenté de substituer ce détergent par un polymère amphipatique (amphipol A8-35) mais les résultats mitigés n'ont pas clairement justifié le basculement vers ce nouveau surfactant. Nous discutons également des opportunités d'utiliser des méthodes basées sur la génération de nanodisques d'origine polymérique ("Styrene Maleic Acids") ou protéique ("Membrane Scaffold Protein").

Interaction avec SecA : Finalement, nous avons étudié l'effet de la formation d'un complexe entre l'Holotranslocon et la protéine moteur SecA et mis en évidence le fait que le complexe ne peut être retrouvé intact après interaction. Il n'apparait cependant pas clairement, si la dissociation du complexe est induite par la liaison de SecA ou si le complexe n'est en premier lieu pas intact.

Diffusion de neutrons à petits angles : Nous avons analysé les différentes préparations d'Holotranslocon en utilisant la diffusion de neutrons à petits angles et en le comparant à une référence (SecYEG). Dans le cas du complexe non marqué, il semble que les amphipols puissent être utilisés mais l'intégrité du complexe n'a pas pu être confirmée avec certitude. Par rapport aux résultats publiés en DDM, le complexe en amphipol contient significativement moins de surfactant, observation associée à un faible rayon de giration. Ces indications semblent suggérer que le complexe en amphipol n'est pas intact.

Nous avons également effectué des expériences de substitution isotopique afin de générer un Holotranslocon marqué au deutérium. Les analyses de diffusion de neutrons à petits angles ont confirmé que le protocole de marquages a un effet sur l'intégrité du complexe. L'optimisation du protocole d'expression nous a permis d'obtenir un complexe marqué que nous avons solubilisé avec du DDM. L'analyse de cet échantillon a révélé une forte quantité de détergents liés mais également une faible quantité de lipides. Le faible rayon de giration semble indiquer que le complexe est dissocié.

L'Holotranslocon est un complexe instable et difficile à produire

Ces résultats montrent que la mise en place d'un protocole de purification permettant la production d'Holotranslocon en quantité suffisante et homogène pour être analysé par des techniques biophysiques est difficile.

La nature de cette assemblage protéique le rend intrinsèquement difficile à exprimer et purifier. Étant un complexe multi-protéique, l'Holotranslocon est sensible à une expression déséquilibrée des sous-unités, ce qui peut induire une hétérogénéité dès les premières étapes de production. Puisqu'il est membranaire, ce complexe doit être purifié en présence de surfactants, qui sont nécessaire à son maintien en solution. La présence de surfactants peut cependant interférer avec les expériences biophysiques, rendant leur analyse et interprétation compliquées.

Nous discutons de la possibilité d'utiliser des techniques alternatives pour favoriser la solubilisation et stabilité de l'Holotranlocon. Un axe d'étude intéressant serait d'introduire le complexe dans des lipodisques, ce qui permettrait de l'analyser dans des conditions les plus proches possible de physiologiques. Là où la taille de ce complexe peut être limitante avec des lipodisques classiques, un nouveau système permet de reconstituer des disques plus grands et plus stable en circularisant la ceinture protéine de façon covalente. D'autres méthodes se basent sur l'utilisation de polymères ("Styrene maleic acid"), qui sont de plus capables de remplacer les détergents pour l'extraction de protéines membranaires.

De façon générale, nous discutons des avantages et inconvénients liés aux différentes méthodes utilisées, confrontons nos résultats aux travaux précédemment publiés et proposons de nouvelles lignes de recherches qui pourraient, à l'avenir, permettre d'obtenir un système de travail fiable pour des études biochimiques et structurales.

Les résultats de diffusion de neutrons à petits angles obtenus pour les complexe hydrogénés semblent indiquer que l'utilisation d'amphipol favorise une réduction de taille, observation pouvant être corrélée aux résultats publiés indiquant que ce surfactant favorise une forme "fermée" de l'Holotranslocon. Quant au complexe deutéré, outre les difficultés liées à la production de l'échantillon, les résultats indiquent que seule une très faible quantité de lipides est liée à la protéine.

La validité et fiabilité des méthodes de diffusion de neutrons à petits angles sont indéniablement prouvées par l'analyse de SecYEG puisque les résultats sont robustes et en accords avec notre connaissance actuelle du complexe. Cependant, dans le cas de l'Holotranslocon, nous estimons que la qualité de l'échantillon mesuré est à l'origine des incohérences parfois observées. Dans cette optique, nous discutons des approximations utilisées pour la mesure de la concentration de l'échantillon, l'évaluation de la fraction d'hydrogène labiles qui est échangeable en présence de deutérium dans le solvant et leur effet sur l'analyse du complexe. Nous discutons finalement de l'impact de l'hétérogénéité de l'échantillon sur l'évaluation de la composition du complexe (protéine:surfactant:lipide) par l'analyse de son "neutron contrast match point".

Nous proposons finalement de nouvelles pistes de travail afin d'améliorer la qualité de l'échantillon : optimisation du système d'expression, mise en place de systèmes de

FRET pour suivre l'intégrité du complexe, recherche de nouveaux surfactants pour l'extraction et solubilisation de l'Holotranslocon, mise en place de systèmes basé sur la génération de nanodisques ("Styrene Maleic Acid", "Membrane Scaffold Protein") ou même l'utilisation de "minicells" pour permettre l'analyse du complexes en conditions strictement physiologique.

2.1 Preparation strategy

2.1.1 Expression strategy

The successful over-expression of the Holotranslocon is a challenging task on different levels.

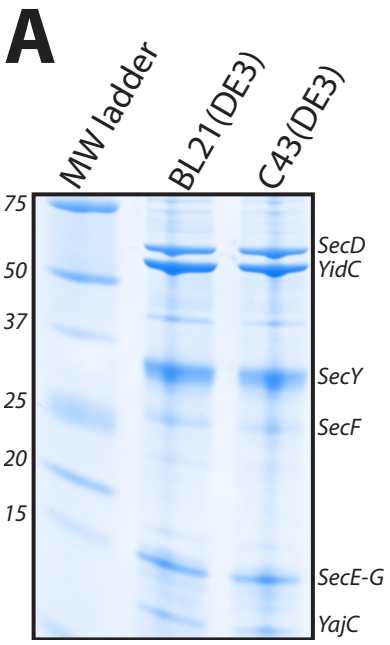
- The Holotranslocon is a membrane protein which makes it intrinsically difficult to over-express efficiently. When cytosolic proteins are expressed they are stored in the cytosol of bacteria, whereas membrane proteins can only remain stable and folded if they are integrated in the plasma membrane. Therefore, the available space for the storage of over-expressed membrane protein is significantly lower than for cytosolic proteins, which implies that the number of molecules that can be produced per bacteria is several orders of magnitude lower for membrane proteins. Furthermore, membrane protein over expression is often toxic for the bacteria (Dumon-Seignovert *et al.*, 2004), which further limits the amount produced.
- The complex is composed of seven subunits, for a final molecular weight of about 250 kDa. Classical cloning methods do not allow reliable sequential insertion of several proteins sequences from a single vector.
- To reconstitute the complex, the subunits should be expressed in the right proportions. Co-transformation of bacteria with several plasmids may lead to issues, among which the copy-number of each plasmid that may introduce a bias at the replicational and transcriptional level as under-represented plasmids are less likely to be replicated or transcribed. This will in the end affect the expression level of the protein of interest.

The ACEMBL technology (Bieniossek *et al.*, 2009) tackles the previously mentioned issues and allows consistent over-expression of all seven subunits, with a set of purification tags, from a single plasmid (see Figure 1.9).

Choice of the expression host

Miroux and Walker, 1996 have described two *E. coli* mutant strains derived from the strain BL21(DE3) : C41(DE3) and C43(DE3). They have been successful in overcoming severe over-expression toxicity of membrane protein while also improving the expression levels of model membrane proteins compared to what can be achieved with classical BL21 (DE3) (Dumon-Seignovert *et al.*, 2004). Reduced toxicity and increased expression efficiency can be attributed to a reduced transcription rate, thus preventing the toxic accumulation of mRNA which saturates the translation and translocation machinery (Miroux and Walker, 1996).

In the case of the Holotranslocon, no significant toxicity could be observed during the culture, regardless of the type of medium used and the culture methods (flask and high cell density). However, the yield of expression and purification of HTL, at a suitable



	BL21(DE3)	C43(DE3)
Purification step		
IMAC elution	0.28 mg per g_{cells}	0.27 mg per g_{cells}
CBP elution	47.3 µg per g_{cells}	37.9 µg per g_{cells}
Relative abundance		
Sec D	1	1
Yid C	1.72	1.19
Sec Y	2.62	1.62
Sec F	0.62	0.35
Sec E-G	1.24	0.79
Yaj C	0.34	0.23

TABLE 2.1: Protein produced and relative abundance of subunits

FIGURE 2.1: Comparison of the expression of the Holotranslocon in BL21(DE3) and C43(DE3). **A** : SDS-PAGE of purified HTL produced with the two affinity steps protocol. The table summarizes the quantity of HTL produced after each purification step and the relative abundance of the subunits as quantified from panel **A**.

subunit ratio has been a major limitation for successful biophysical and structural studies.

We therefore decided to compare the efficiency of C43(DE3) expression strain, which has been shown to be more efficient for membrane protein expression than the classical BL21(DE3) strain. We compared these strains to the situation where the protein was expressed in high cell density culture as these conditions are the most likely to suffer from toxicity or imbalanced protein expression due to the long induction time. To compare the yield of C43 or BL21 expression for the entire HTL, we performed the expression in parallel and under the same conditions. Purification were done with the same initial mass of cell paste and using the two affinity steps protocol (section 5.2.1), which ensures the co-purification of all seven subunits. We compared the concentration and subunit ratios after purification and analysed the Coomassie stained SDS-PAGE data.

In both cases, no significant differences could be observed during the different steps by SDS-PAGE (data not shown). Figure 2.1 panel A shows that the samples were very similar after purification ; no significant difference was observed and the profiles were similar to published results (Schulze *et al.*, 2014). The concentration was measured by micro-spectrophotometry after IMAC affinity elution and the yield was considered equal as the variation in quantity of protein was below 4%. After the Calmodulin elution step, we estimated that 25% more protein was purified from the BL21(DE3) culture, but given the small quantities achieved, these differences could be explained by experimental variations that are difficult to avoid over a three-day protocol. The

relative intensities of the subunits were quantified and expressed as a function of the intensity of SecD. In the case of BL21(DE3), YidC and SecY seemed in excess as compared with C43(DE3), while SecF seemed to be depleted in C43(DE3) cells. However, repeated experiments showed that SecF did not retain the stain as well as other subunits, SecE and SecG were not always separated efficiently and YajC was not always detected on gel. Therefore, these results were interpreted with care as the quantification on gel is dependent on the molecular weight of the subunit, depends on the nature of the proteins and is highly sensitive to experimental variation (homogeneity of the staining and destaining, quality of the gel and buffers etc.).

These results did not reveal significant differences between the strains C43(DE3) and BL21(DE3). In order to be consistent with previously published work (Botte *et al.*, 2016; Schulze *et al.*, 2014), the C43(DE3) strain was used in further work.

High cell density bacterial culture

High cell density bacterial cultures maximize bacterial growth by allowing them to reach higher cell density than classical cultures in flasks. This is achieved by tuning the oxygenation of the medium (through agitation and controlled airflow), balancing the pH to compensate for metabolic acidification, controlling the input of carbon source and the temperature. These cultures were performed in fermenters operated with control sequences that allow the above-mentioned culture conditions to be maintained (see subsection 5.1.1).

This method was applied to the culture of the Holotranslocon using ENFORS medium. The yield was 45 g of cell paste per litre of culture, as compared to flask culture that produced on average 2 g of cell paste per litre of culture. Hence, the fermenter used with minimal medium was equivalent to 36L of flask culture in rich medium. This increased yield significantly lowers the cost of deuteration approaches by reducing the amount of D_2O and deuterated carbon source required.

Figure 2.2 shows the growth profile for HTL cultured to high cell density in hydrogenated ENFORS minimal medium. The three phases described in subsection 5.1.1 are observed here. During the batch phase, bacteria used the available carbon source until it runs out at 11h. The metabolism of bacteria reduces, inducing an increase of the pO_2, which in turn triggers the fed batch phase. The carbon source was provided in small batches to avoid accumulation of metabolic by-products until the OD_{600} was sufficient for the induction. In this case, the induction was started at $OD_{600} = 21$ with 1 mM IPTG and 0,2% l-arabinose (as indicated in subsection 5.2.1) for a duration of 18 hours. This fermenter reached an optical density of $OD_{600} = 27$ and produced 88 g of cell paste for a volume of 1.7 l.

To ensure that the expression of the different subunits of Holotranslocon was not affected by high cell density culture, we performed purification using one affinity step (see section 5.2.1) and analysed the different steps by SDS-PAGE. Figure 2.3 panel A shows the resulting purification on a high cell density culture performed in 85% D_2O with a induction time of 18 hours. It was striking to observe that only SecYEG was clearly observed on this gel, indicating that there was an important imbalance in

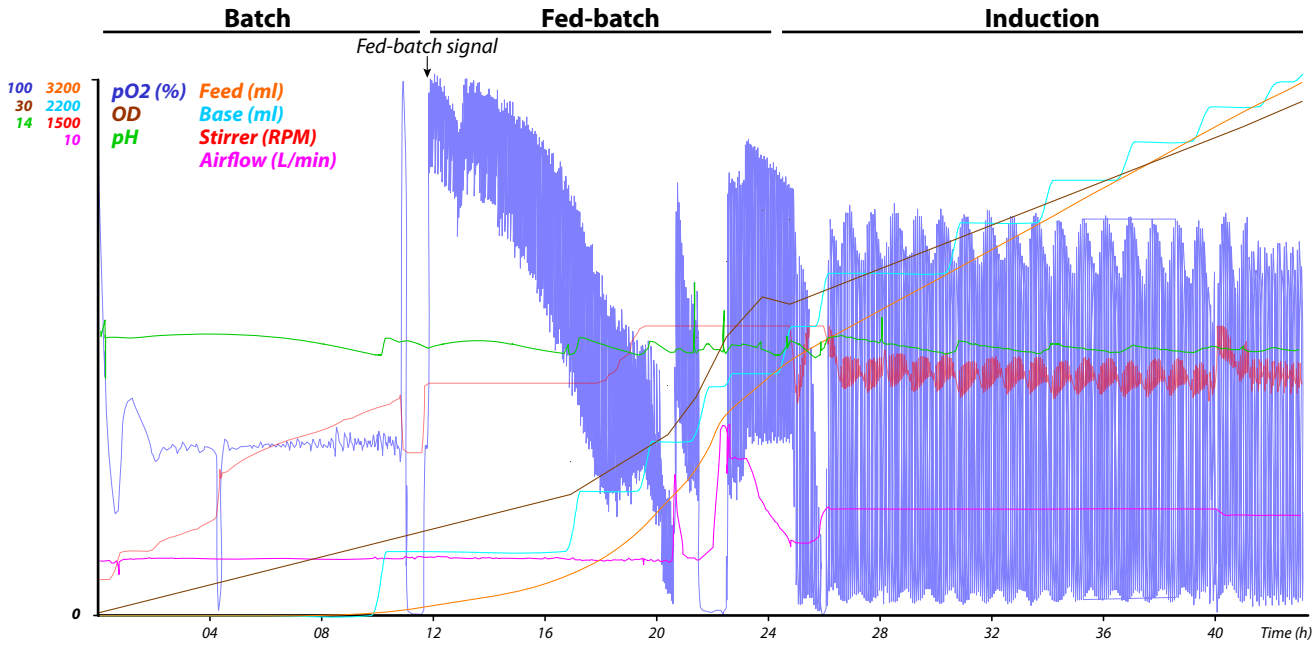


FIGURE 2.2: Typical high cell density culture profile of *E. coli* culture for the production of Holotranslocon in hydrogenated medium. Hydrogenated ENFORS minimal medium was used for this culture. The pO_2 (in blue) is controlled by the stirring speed (in red) and the air flow (in magenta). The pH (in green) is controlled by the addition of base (in cyan) and the carbon source is added by the feed pump (in orange). The growth is monitored off-line by regular measurements of the OD_{600} (in brown).

the expression of the subunits under these conditions. It was also noted that using two affinity purification steps (see section 5.2.1), all subunits were observed (data not shown) as the Calmodulin affinity steps selects out the SecYEG complexes which were not bound to the other subunits; the yield was however very low.

The band observed around 60 kDa could be attributed to YidC as it is naturally present in relatively high copy-number in the cell. Since SecDF and YidC are under the control of the arabinose promoter, we hypothesized that this loss of expression could be attributed to the fact that, during the induction, the bacteria may activate their arabinose operon. The sugar would therefore be metabolised, as it is preferred over glycerol, therefore impairing the expression of the genes under the arabinose operon. Similar results were observed for high cell density culture performed in hydrogenated conditions or 85% deuterated media.

To test this hypothesis, we performed another high cell density culture under the same conditions, but stopped the experiment after 6 hours of induction. Figure 2.3 panel B clearly shows that this allowed at least the partial recovery of the expression of SecD and YidC. A very faint band at 30 kDa could be attributed to SecF. The imbalance observed between SecD and YidC could again be due to the natural high copy-number of YidC in *E. coli*. When compared to Figure 2.3 panel C, which shows a reference purification of hydrogenated HTL produced in flask, we observed that SecYEG was still significantly over-represented.

These results suggest that the expression of the Holotranslocon by high cell density

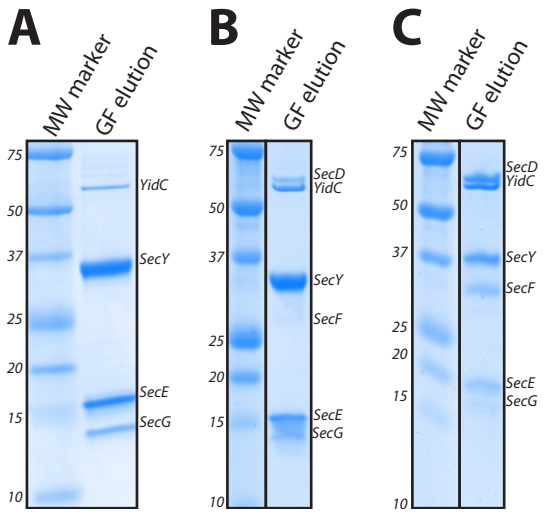


FIGURE 2.3: Comparison of the Holotranslocon expression under different conditions. The gel filtration elution obtained after a single step of affinity purification are analysed. **A**: High cell density culture in 85% D_2O induced for 18 hours. **B**: High cell density culture in 85% D_2O induced for 6 hours. **C**: Flask culture in hydrogenated rich medium with 3 hours induction.

culture tended to alter the expression ratio of the subunits, making this strategy unsuitable for further experiments. Furthermore, the lack of reliable methods to ascertain a suitable ratio of the subunits (see subsection 2.2.1) makes the optimization of this strategy tedious and uncertain.

In vivo deuterium labeling

The expression of proteins in deuterated medium required the adaptation of the bacteria to these new conditions by cultivating them for several days in minimal hydrogenated medium then minimal deuterated medium. As mentioned in section 2.1.1, the Holotranslocon could be expressed in deuterated conditions, though with mitigated results. The deuteration itself did not seem to affect the expression of HTL but the major drawback for the labeling of HTL was the need to use high cell density cultures, which generates imbalanced expression of the subunits.

Match-out labelled HTL was produced and analysed by small-angle neutron scattering (see section 2.4).

2.1.2 Solubilization

The ability of integral membrane proteins to insert and remain stable in a lipid bilayer makes them, by their nature, intrinsically hydrophobic and insoluble in an aqueous environment. To circumvent this, the choice of the solubilization method is of key importance for the extraction of the protein from the membrane, its purification and

stability. Most commonly, the solubilization is achieved by using surfactants; their amphiphilic structure allows the binding of their hydrophobic tails to the transmembrane segments of the protein while the hydrophilic heads interact with the aqueous solvent (D. Hardy *et al.*, 2016).

It is commonly admitted that high concentration of detergents often leads to disruption of the quaternary or tertiary structures of proteins, as they interfere with the electrostatic and hydrophobic forces that drive proteins folding. On the other hand, in order to be efficient, a detergent should be used at a concentration superior to its Critical Micellar Concentration (CMC). Therefore, detergents which have low CMC are often chosen as they allow the extraction of the protein from its membrane while minimizing the disturbances of its structure (Moraes *et al.*, 2014).

Detergents can be classified in different categories depending on the nature of their head group. Ionic detergents are usually harsh and used when a protein is hard to solubilize. Non-ionic detergents are milder and tend to preserve protein/protein interactions; they are often used in membrane protein research for solubilization and crystallization (short chain detergent) or purification and stabilization. Finally, zwitterionic detergents are commonly used for crystallization and Nuclear Magnetic Resonance studies (Moraes *et al.*, 2014).

Many detergents are commercially available and the afore-mentioned characteristics may help with this choice, but selecting the appropriate detergent is still an empirical and time consuming process.

Detergent Screening

Previous publications (Schulze *et al.*, 2014; Botte *et al.*, 2016) and collaborators have chosen the commonly used n-dodecyl-β-d-maltoside (DDM), a non-ionic detergent with a sugar-based head group and long acyl-chains that have a low CMC and is considered as a mild detergent. To the best of our knowledge, this choice has not been supported by any comparative study.

The detergent screening platform, Robiomol, has been used to test a wide range of detergents:

- **Ionic detergents**: NLS (N-Lauroylsarcosinate),
- **Nonionic detergents**: Cymal 2, Cymal 3, Cymal 4, Cymal 5, b-HG (n-Heptyl-β-D-glucopyranoside), B-OG (n-Octyll-β-D-glucopyranoside), DDM (n-Dodecyl-β-D-maltopyranoside), DM (n-Decyl-β-D-maltopyranoside), TX100 (Triton-X-100), UM (n-undecyl-β-D-maltopyranoside), LDAO (n-dodecyl-N,N'- dimethylamine-N-oxide), LAPAO (3-Laurylamido-N,N'-dimethylpropylamine oxide), MNG3 (Lauryl-Maltose-neopentyl glycol), C8E4 (Octyl Tetraethylene Glycol Ether), C12E8 (Octaethylene Glycol Monododecyl Ether), LTM (N-Dodecyl-β-D-Thiomaltopyranoside), C12E9 (Polyoxyethylene(9)dodecyl Ether), T3DM (Tridecyl-β-D-maltoside), OGNG (Octyl-Glucose-Neopentyl-Glycol), β-HG (n-Nonyl-β-D- Glucopyranoside).
- **Zwitterionic detergents**: CHAPS, Foc-Choline 16, Fos-choline 12

E. coli total membrane extracts were prepared as described in section 5.2.1 from a culture expressing the Holotranslocon. The total protein concentration within the membrane fraction was evaluated with a bicinchoninic acid assay kit (Biorad) to 7 mg ml^{-1} and provided to the platform. The membranes were solubilized at a detergent concentration of $CMC + 20mM$ and incubated for 2h at 4 °C. The solution was applied to 50 µl of nickel IMAC resin equilibrated with Tris 20mM pH8, 130mM NaCl with the detergent tested. The resin was washed twice with 50mM imidazole and eluted with 300mM imidazole. The samples were loaded on SDS-PAGE and analysed by Coomassie staining. The bands were manually defined on the gel and their intensity relatively quantified with background subtraction (using Imagelab software, BioRad).

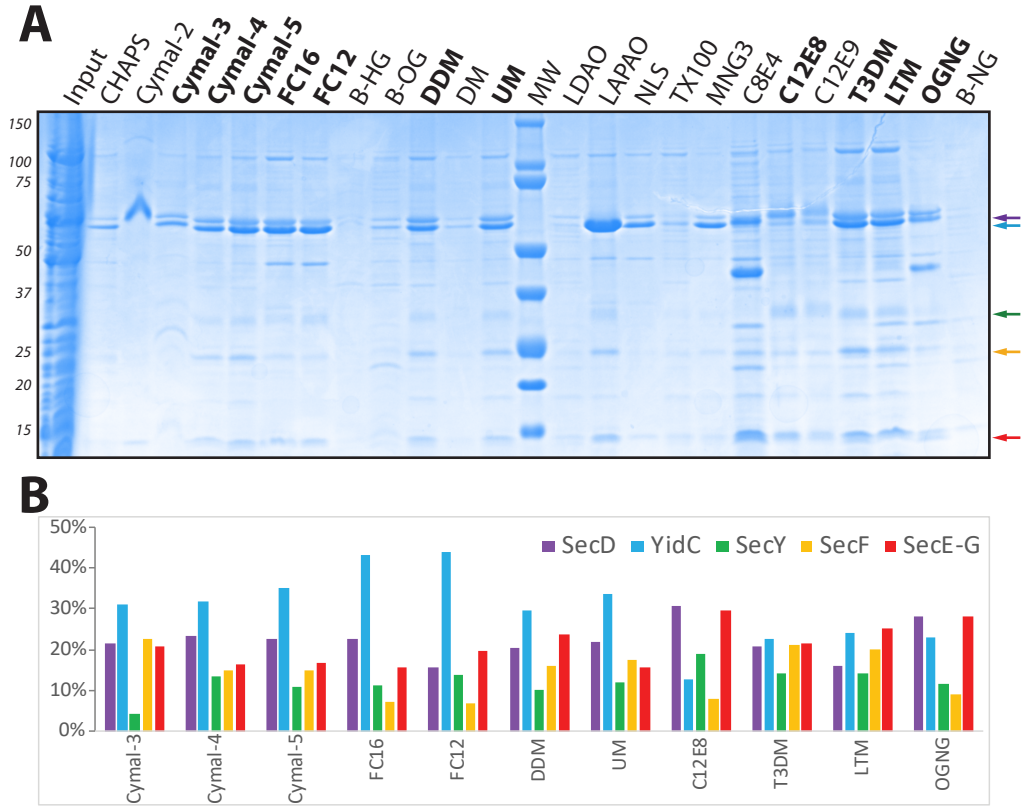


FIGURE 2.4: Detergent screening for HTL. **A**: SDS-PAGE of the elution for each detergent. Arrows indicate the bands for SecD (in purple), YidC (in cyan), SecY (in green), SecF (in orange), SecE-G (in red). The detergent in bold are used for the relative quantification of the bands. **B**: Relative abundance of the bands. Each band is expressed as a percentage of the total intensity of the 5 bands quantified

.

Figure 2.4 panel A shows that some detergent seemed to yield very little, incomplete, or no Holotranslocon (CHAPS, Cymal-2, B-HG, DM, LDAO, TX100, B-NG). C8E4 seems to be highly contaminated (band at 40 kDa) and was therefore not taken into account. Among the remaining detergents, the bands corresponding to the subunits were quantified. The detergents for which a band was representing more than 50% of the total quantified intensity are not included in panel B (B-OG, LAPAO, NLS, MNG3, C12E9).

Panel B shows the results obtained with the remaining detergents. For DDM, the detergent commonly used for the purification of HTL, we observed a significant imbalance between the expression of SecD and YidC, which should have appeared equivalent as their molecular weights are similar. This imbalance in favour of YidC could be explained by the fact that YidC is natively more expressed than SecD (3 YidC for 1 SecD, G.-W. Li *et al.*, 2014). SecY also seems to be under-represented. However, since SecE and SecG could not be clearly distinguished and were quantified together in a single band, we could consider that each subunit represent 50% of the band intensity measured, therefore bringing the intensity of SecE and SecG to a similar value as SecY.

Cymal-4, Cymal-5 and UM displayed subunit distribution similar to that observed with DDM. For the zwitterionic detergents FC12 and FC16, YidC was still over-represented, which may indicate that the extraction of HTL for the membrane was less efficient than the extraction of monomeric YidC. In the case of Cymal-3, SecY seemed to be dramatically under-represented while SecE-G was still significantly present. This was surprising since SecE-G was expected to be bound to SecY to make the SecYEG translocon but could be explained as previous work (Botte, 2013) reported the existence of a SecE-G complex by native mass spectrometry. C12E8 and OGNG displayed an imbalance of subunit purifications with an excess of SecD, SecEG and a deficiency of SecF. Finally, T3DM and LTM yielded a homogeneous distribution of subunits but also included more contaminants during the purification.

Even though quantification on SDS-PAGE gel is not sufficient to confidently assess the subunit ratio achieved during the purification, these results suggested that Cymal-4, Cymal-5 and UM could be used to extract and purify HTL with similar efficiency as DDM. T3DM and LTM could be interesting substitutes as they seemed to yield more homogeneous subunit distribution during Nickel affinity purification. All these detergents are non-inonic with a maltopyranoside head group. UM, DDM, LTM and T3DM have linear acyl chains (11, 12, 12, and 13 carbons respectively) while Cymal-4 and 5 have cyclic carbon chains. The CMC (in H_2O) of DDM is 0.17 mM, which is in the same range for UM (0.59 mM), an order of magnitude lower than Cymal-4 and 5 (7,6 and 5 mM respectively) and higher than LTM and T3DM (0,05 and 0.03 mM respectively).

The concentration of detergents and their CMC seemed to be critical. If the cyclic nature of Cymal-4 and 5 carbon chains seem to be partially compensating for their high CMC, it remains a major limiting factor, as Cymal-3, which has a high CMC of 34.5 mM, highly affects the subunit distribution of HTL. On the other hand, T3DM and LTM, which have 10-fold lower CMC than DDM, have shown more homogeneous subunits distribution. The Holotranslocon therefore seems to be efficiently purified in non-ionic detergents, maltoside-based with acyl chains of 10 to 13 carbons, with the CMC playing a decisive role in the balance of subunits.

This experiment was performed at a late stage of the presented work. Therefore, due to the lack of a reliable method to unequivocally assess the quality of the sample (especially the subunit distribution and size characterization), the variety of potential detergents candidate and limited time available, this aspect of the work could not be pursued.

This experiment used the platforms of the Grenoble Instruct centre (ISBG ; UMS3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB).

Styrene Maleic Acid Lipodiscs (SMALPS)

If detergents are successful at extracting proteins from their native membrane, they do not reliably mimic the natural environment of membrane proteins. Styrene Maleic Acids (SMA) are a class of copolymer arranged in a semi-randomly alternating pattern forming a linear chain. At a critical concentration, they self-assemble into a nano-scale discs ranging from 6 to 100 nm diameter, before entering into the membrane and excising a working section of lipid bilayer (Esmaili and Overduin, 2018). This method was successful in producing soluble membrane proteins that have shown higher stability than in detergents (Jamshad, Charlton, *et al.*, 2015), or could even solubilize complexes containing up to 17 subunits for a size of 18 nm (Bell *et al.*, 2015).

The use of SMALPS for the direct solubilization of native Holotranslocon from *E. coli* membrane has been first introduced in the work of Komar, Alvira, *et al.*, 2016, showing intact HTL could be extracted and immuno-precipitated from wild type *E. coli* membranes. We therefore initiated a collaboration with Oxford University in order to apply such a system to over-expressed Holotranslocon in order to achieve direct extraction from *E. coli* membrane. Preliminary results showed that a large complex could be purified by nickel IMAC affinity and comparison with a strain over-expressing SecYEG suggested that this complex may be the complete Holotranslocon (data not shown). However, due to lack of time and given the extensive optimization process required in order to produce SMALPS solubilized Holotranslocon at a scale compatible with structural studies, this work could not be further pursued.

2.1.3 Purification

The plasmid generated from the ACEMBL technology (Bieniossek *et al.*, 2009) contains all seven subunits under the control of two different types of promoters. Four of the subunits have been fused to purification tags: YidC has a C-terminal poly-histidine tag, Sec D and SecE have a N-terminal poly-histidine tag, and YajC a Calmodulin Binding Peptide (CBP) tag on its C-terminal (see Figure 1.9).

SecYEG and SecDF-YidC-YajC are complexes that have been described as stable by themselves (Botte, 2013), as well as YidC that can also operate as a monomer. Therefore, if the multiplicity of the hexa-histidine tags allows for increased affinity of the complex for the resin, it introduces uncertainties as the Nickel affinity step can theoretically purify complete Holotranslocon, as well as SecYEG, SecDF-YidC-YajC complexes or YidC alone.

Two purification strategies were compared. The first makes use of the unique CBP tag affinity tag, which allows the assessment for the integrity of the complex by simple SDS-PAGE (Botte *et al.*, 2016), the other exploits the inevitable difference in size

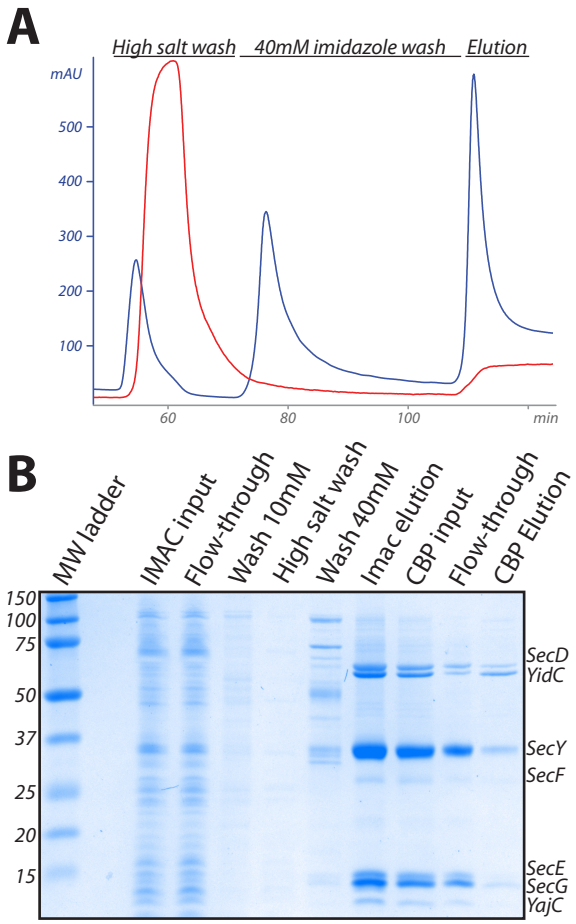


FIGURE 2.5: Purification of the Holotranslocon using two affinity purification steps. **A**: purification profile of the IMAC nickel affinity step performed on an Aktä prime (GE Healthcare). The high salt wash is done in 500 mM NaCl, the washes with imidazole and the elution in 330 mM imidazole. The A_{280} (in blue) and conductivity (in red) are monitored. **B**: Coomassie stained SDS-PAGE of the purification procedure.

between the different complexes by relying on size exclusion chromatography (Schulze *et al.*, 2014).

Two affinity purification steps

This strategy combines two affinity purification step : nickel IMC affinity and Calmodulin affinity (see section 5.2.1). The procedure was done in Hepes buffer, and included extensive washes : 10 mM imidazole, 500 mM NaCl, 40 mM imidazole for the nickel affinity and 0.2 mM $CaCl_2$ for the CBP affinity.

Figure 2.5 panel B shows the SDS-PAGE analysis of the purification procedure. In the total membrane fraction (corresponds to the nickel input), the HTL subunits were not clearly identifiable: a band below 37 kDa could correspond to SecY, though the flow-through fraction also contains this band. The washes at 10 mM had the same profile as the flow-through as it washed out the remaining membrane fraction. In the high salt

wash fraction, close to no protein was detected but Figure 2.5 panel A clearly shows a signal at 280 nm, indicating that contaminants were stripped off the column. The 40 mM imidazole wash got rid of high molecular weight contaminants, but also seemed to strip off some SecYEG from the column. In the IMAC elution, we could clearly identify all seven subunits, with the band of SecY and SecG being particularly intense. The solution was buffer exchange for the Calmodulin Binding Peptide buffer and loaded on the calmodulin resin. In the CBP flow-through, we could observe all seven subunits, but with a clear excess of SecYEG. The same profile was observed during the washing of the calmodulin resin with reduced concentration of $CaCl_2$. Finally, SecD, YidC and SecYEG were clearly identified in the CBP elution fractions. SecF and YajC are hardly observable due to the low concentration of the elution fraction (around 0.1 mg ml^{-1}), but their presence was confirmed by concentrating the fraction and performing a anti-CBP blot (data not shown). The solution could be further purified by size exclusion chromatography (see Figure 2.7 and Figure 2.12).

We therefore observed that this two-affinity-step protocol allowed for the purification of the Holotranslocon by using both the histidine tags and the CBP tag. The CBP affinity purification selected out the SecYEG that were not bound to YajC. However, SecDF, YidC and YajC could also be observed in the flow-through. This may indicate that despite an incubation overnight, the binding to the Calmodulin resin had a rather low efficiency, which has been confirmed by the measurement of HTL's concentration; less than 1 mg is usually purified from 10 g of cell paste using this protocol. This observation could also indicate that the complex was partially dissociated.

Single affinity purification step

This strategy relies on the use of the histidine tags and Size Exclusion Chromatography (SEC, or noted GF for gel filtration) to further refine the purification. The HEPES buffer was substituted for Tris buffer, the pH was kept at 8 throughout the whole procedure and the wash strategy was milder: the membrane fraction was loaded without imidazole and a single wash was performed at 30 mM imidazole. The quantity of DDM used for stabilization after Nickel purification was reduced from 0,03% to 0,02%, but remained above the critical micellar concentration (0.009%).

Figure 2.6 panel A shows the nickel IMAC affinity purification profile, omitting the loading and first 0 mM imidazole wash. The wash at 30 mM stripped away some contaminants and the protein is efficiently eluted at 330 mM imidazole. The corresponding samples were analysed by SDS-PAGE on panel B, showing all the subunits in the IMAC elution, with a large excess of SecYEG, as described before. The SecF and YajC bands were hardly observed, but further experiments confirmed their presence and showed these bands tend to disappear during the SDS-PAGE destaining procedure. After nickel IMAC purification, the elution was loaded as it is to a Superdex 200 GF column and a typical profile is represented in panel C. A first minor peak was observed at 9 ml retention volume, which corresponded to the void volume of the column (probably aggregates). A second peak was observed at 14.3 ml, which would correspond to a molecular weight of 180 kDa, as calculated from calibration curves (note that the calibration curves have been calculated in a buffer without glycerol, which could in this

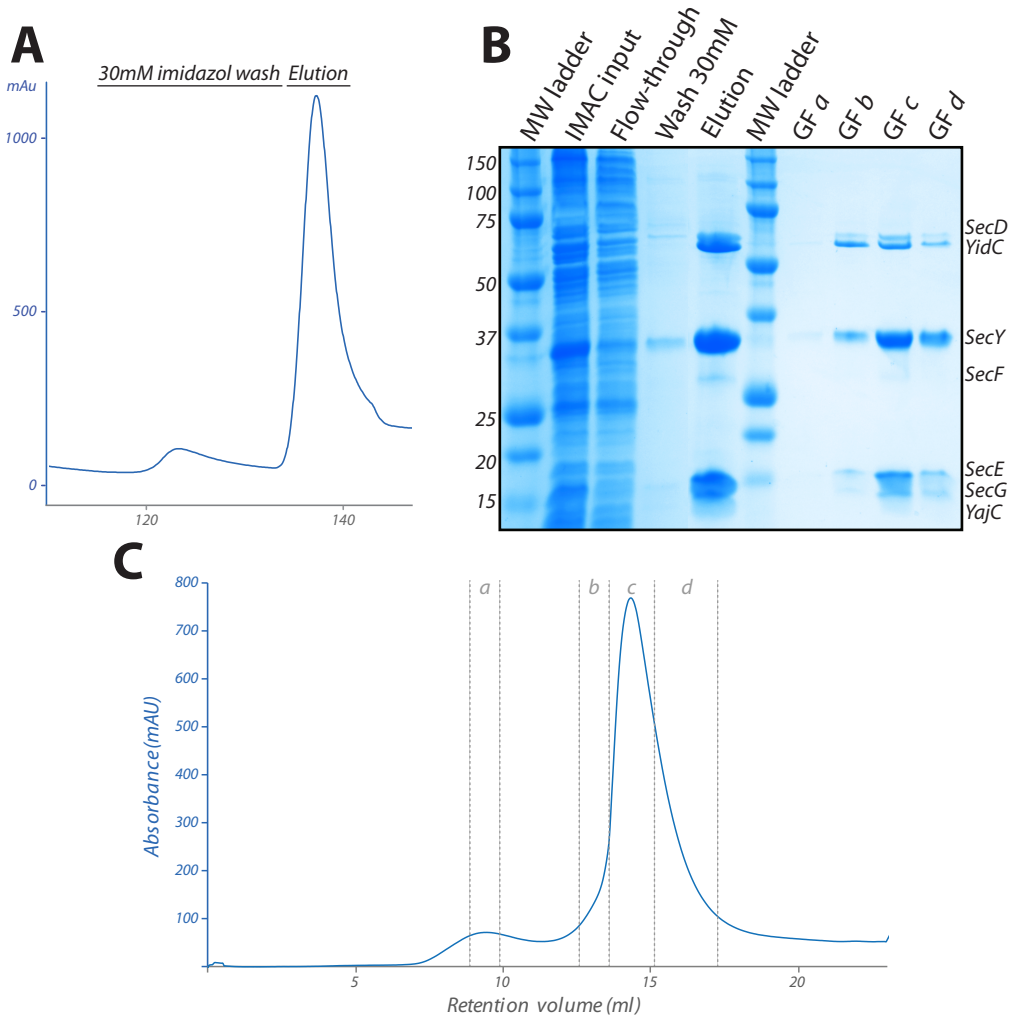


FIGURE 2.6: Purification of the Holotranslocon using a single affinity purification step. **A**: purification profile of the IMAC nickel affinity chromatography performed on an ÄKTA prime (GE Healthcare). The elution is done in 330 mM imidazole. **B**: Coomassie stained SDS-PAGE of the purification procedure. **C**: Size exclusion chromatography profile obtained on a Superdex 200 column on a ÄKTA purifier. a,b,c and d represent the gel filtration (GF) fractions that have been analysed in panel B.

case lead to a slight underestimation of the molecular weight). The latter peak was slightly asymmetrical, with fractions corresponding to the beginning (b), middle (c) and end (d) of the peak were analysed separately; panel B shows that the composition seemed to be similar, with incidentally a higher intensity in fraction ***c*** (see panel C). It is important to note that the intensity of the subunit bands seem to suggest that the ratio of SecYEG to DFYY is imbalanced.

This protocol was first described in (Schulze *et al.*, 2014), where they combine the GF with in-line anion exchange column, so that the size chromatography elution goes directly through the anion exchange column and the flow-through is collected. Among the individual subunits, only SecG, SecF, and less notably YidC carry a negative charge at pH 8 and should therefore be retained by the anion exchange column. All the other subunits and most importantly the other sub-complexes (DFYY, SecYEG) would carry a positive charge. Most importantly such a set-up would imply that the size exclusion fractions are diluted and partially mixed when passing through the anion exchange column; this set-up would therefore strongly impair the resolution of the gel filtration.

We found that, when comparing a sample analysed with and without the anion exchange column in-line, the second case led to a reduced intensity of the signal by a factor of 2,3 and the width of the peak increased by a similar factor (data not shown). The flow-though of the anion exchange column contained all seven subunits while only faint bands for SecD and YidC were detected after eluting the Q column. Despite the fact that only YidC, SecG and SecF carry negative charges at this pH, SecD could also be detected, which would suggest that DFYY is retained, despite its theoretical positive charges. It was also interesting to note that the total amount of protein retained by the anion exchange column, when it is in line with the GF column, was much higher than when the peak gel filtration fractions were loaded separately on this column (data not shown). This suggested that most of the contaminants actually retained by the anion exchange did not come from the peak fractions.

This protocol allowed the Holotranslocon to be purified with a significantly higher yield than the two affinity steps protocol (usually 5 to 10 times more). However, because the CBP tag on YajC is not being used and as suggested by Figure 2.6 panel B, this protocol was more sensitive to subunit ratio imbalance (especially in the cases where SecYEG is over-represented, as observed in section 2.1.1). The size exclusion chromatography step following the IMAC affinity purification may be efficient at separating the complete Holotranslocon (255 kDa) from other sub-complexes (SecYEG : 75-150 kDa, DFYY : 180 kDa). However, the present experiment suggested that, if such sub-complexes existed, they could not be observed by GF. The addition of an anion exchange column in-line with the SEC significantly reduced the resolution and therefore the possibility of separating sub-complexes based on their size. The use of GF for the characterisation of HTL is further discussed in section 2.2.

2.2 Stabilization of the complex

For the purposes of extracting the Holotranslocon from the membrane and in order to keep it soluble during the purification procedure, surfactants must be used. However, as discussed in section 2.1.2, detergents are known to disrupt the structure of proteins. We therefore used a combination of biophysical methods to assess the stability and homogeneity of the Holotranslocon in detergents. In parallel, we investigated the possibility of using amphipol A8-35 as a substitute for DDM.

2.2.1 Detergents

The use of DDM certainly allows the purification of all subunits but the integrity and homogeneity of the complex can be questioned. During the purification process and depending on the expression conditions, imbalances of subunits ratio could be observed. It is also important to note that HTL has been shown to be highly unstable in DDM, unless chemically cross-linked, when purified with two affinity steps (see section 5.2.1) (Botte, 2013). In order to investigate these issues, a range of biophysical methods was used.

Size Exclusion Chromatography (SEC)

Size exclusion chromatography (gel filtration, noted GF ; column volume noted CV) is used as part of the purification protocol and the principal way to ensure that the complex is homogeneous and monodisperse (Schulze *et al.*, 2014). In order to assess the stability and size of the complex, 85% deuterated HTL was run on SEC under different conditions and compared to SecYEG by GF.

When analysing Figure 2.7 panel A, it was striking to see that SecYEG, with an expected molecular weight of 75 kDa, eluted before HTL(255 kDa), even if account is taken for the fact that SecYEG is known to dimerize (Hizlan *et al.*, 2012). These results were surprising as numerous attempts proved that all seven subunits composing the Holotranslocon could be detected on SDS-PAGE after size exclusion (see Figure 2.6). Considering the fact that the GF was carried out in a buffer containing glycerol (3 to 10%) and that detergents are bound to the complex, therefore affecting its hydrodynamic properties, the molecular weight cannot be reasonably inferred from the elution volume as the calibration curves use non viscous solvent and globular soluble proteins. These first data could suggest that the Holotranslocon may have a significantly smaller hydrodynamic radius than implied by its molecular weight, or that the Holotranslocon somehow dissociates with a co-migration of the dissociation products (see Figure 2.6, panel B).

Panel B represents another GF profile obtained from a large scale purification where it was observed that the Holotranslocon eluted as a double peak, suggesting that two populations of complexes, with distinct molecular weights, were present. SDS-PAGE showed that HTL subunits were detected in both peaks (data not shown). When reloading the first peak (60.3 ml, 0.50 CV), we observed that the peak shifts to 69 ml

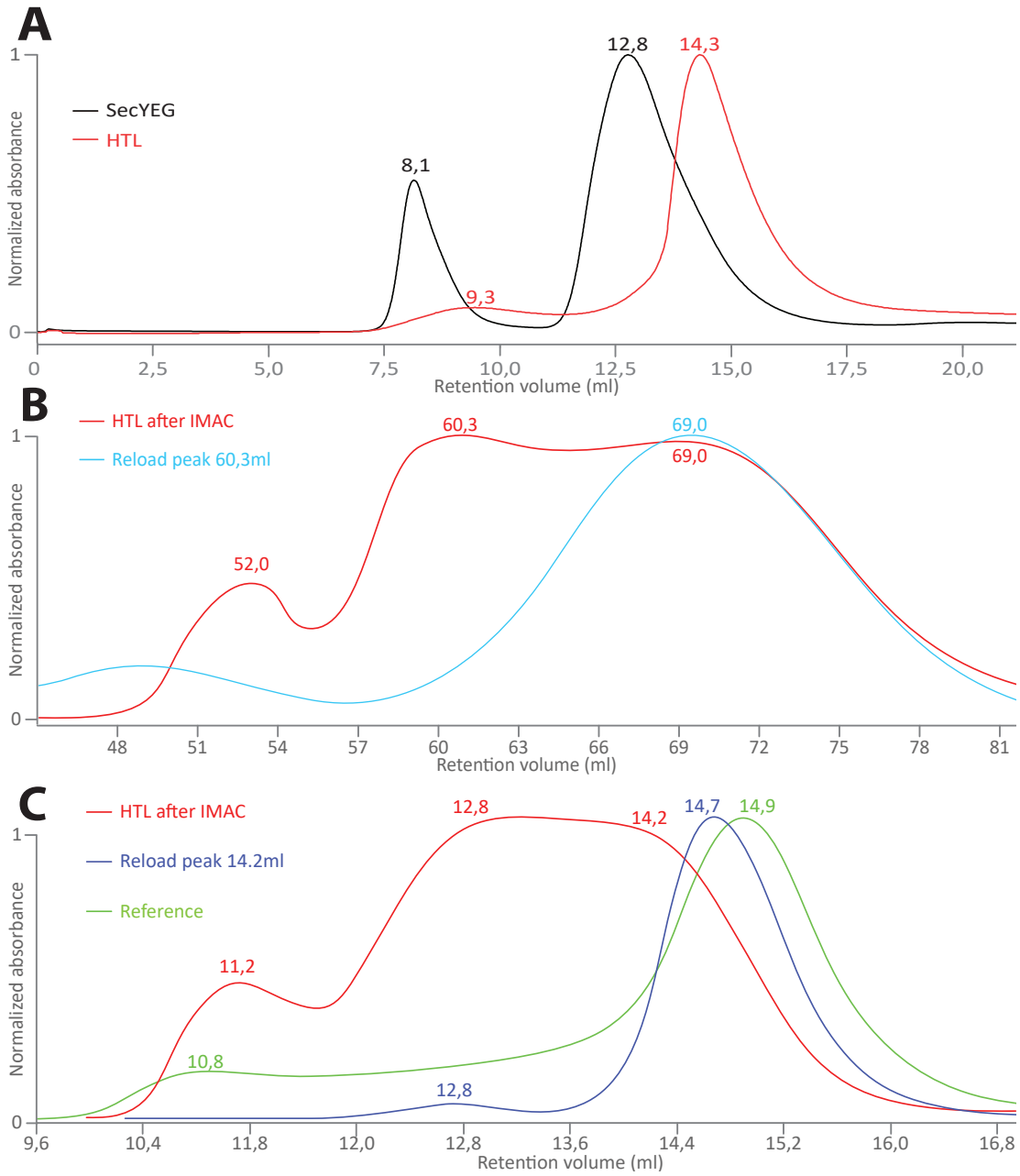


FIGURE 2.7: Size exclusion chromatography (Gel filtration, GF) analysis of the Holotranslocon solubilized in DDM. **A**: Comparison of purified DDM solubilized SecYEG (in black) at 2.5 mg ml^{-1}(33.3 µM) and d-HTL (in red) at 0.9 mg ml^{-1}(3.6 µM) profiles on a Superdex 200 10/300. **B**: GF of d-HTL on a Superdex 200 10/16 after nickel affinity at 1.8 mg ml^{-1} (7.2 µM). The elution corresponding to the peak at 60.3 ml (0.50 CV) is reloaded (in cyan) on the same column at 1 mg ml^{-1}(4.0 µM. **C**: GF of d-HTL on a Superdex 200 10/300 after nickel affinity. The elution corresponding to the second peak at 14.2 ml (0.59 CV) is reloaded (in blue) on the same column. The reference (in green) corresponds to a independent h-HTL purification at 1.1 mg ml^{-1} (4.4 µM).

(0.58 CV), which corresponds to the second peak observed in the first GF. In order to further investigate this effect, panel C shows that additional GFs have been performed on an analytical column and we observe that when reloading the second peak (here at 14.2 ml, 0.59 CV, the particles elutes in the same region, with a minor peak at 12.8 ml (0.53 CV) corresponding to the first peak. These profiles were compared to a reference sample (hydrogenated HTL produced independently) and we observed an elution volume of 14.9 ml (0.62 CV) was observed. Independent experiments on hydrogenated HTL gave similar results, though the concentration effect seemed to be less pronounced and the early elution peak usually less intense.

This observation can be attributed to a concentration effect: when eluted from the nickel affinity column, the concentration reach 1.8 mg ml^{-1} (7.2 µM) but did not go above 1 mg ml^{-1}(4.0 µM) after GF. In the case of the reference sample, the concentration was 1.1 mg ml^{-1} (4.4 µM), but we could not clearly identify a peak around 12.8 ml (0.53 CV). However, we can observe the absorbance is higher than the baseline, suggesting the presence of particles. These experiments suggest that at a critical concentration (above 1.1 mg ml^{-1}), a higher molecular weight particle was observed. Based on these data, it is hard to unequivocally assess whether this peak contains all of HTL's subunits. Given the molecular weight range that can be analysed using a Superdex 200 columns, it is rather unlikely that this peak would correspond to a multimeric HTL (which would have a molecular weight of more than 500 kDa). It could be hypothesized that the integrity of the Holotranslocon is concentration dependent and that as it is lowered, HTL dissociates into sub-complexes that cannot be efficiently separated by gel filtration. It is also possible that the SEC, which inevitably dilutes the sample, favours the dissociation of the complex. It is however important to note that even at high concentration, we could not obtain a single peak that would correspond to the intact HTL. Ultimately, SEC did not allow us to unequivocally confirm the integrity of the sample and might not reliably produce a homogeneous solution of intact HTL.

Dynamic Light Scattering (DLS)

Dynamic Light Scattering (DLS) can also be used to evaluate the size of the complex based on their mobility in solution and could therefore prove to be useful at assessing the quality of Holotranslocon samples in solution.

It is important to note that for both diluted and concentrated sample, Figure 2.8 the calculated polydispersity was significantly higher than 20%, which is the commonly admitted threshold that describe a polydisperse sample (Malvern, Common term defined). Considering the polydispersity of the sample, it was observed that the maximum intensities overlap for a calculated size of 50.75 nm. The concentrated samples appeared to be slightly more polydisperse, with a shift toward higher size particles resulting in an increased average size.

The calculated sizes (and therefore molecular weights) were an order of magnitude higher than what is expected: the Holotranslocon should weigh 255 kDa. This surprising result could be attributed to the presence of DDM micelles which also scatters

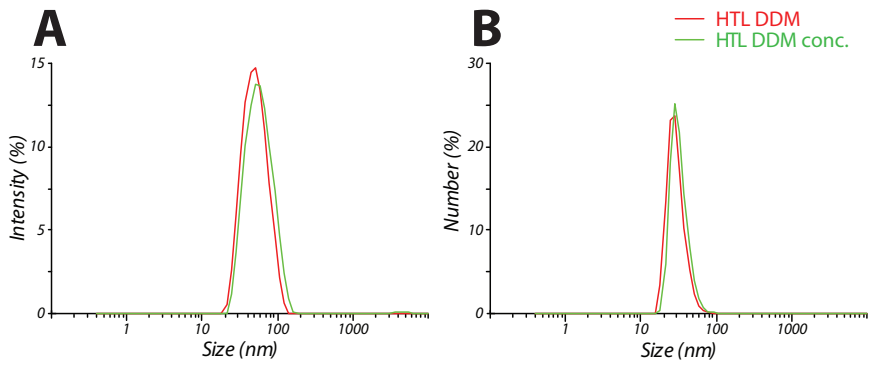


TABLE 2.2

	Size_{peak}(nm)	$\text{MW}_{peak}(kDa)$	$\text{Size}_{sol.}$(nm)	$\text{MW}_{sol.}(kDa)$	$\%\text{PD}_{sol.}$
HTL	$50,75 \pm 19,63$	6500 ± 2660	$46,74 \pm 16,27$	5360 ± 1870	$34,8\%$
HTL conc.	$50,75 \pm 24,02$	6500 ± 3900	$52,92 \pm 19,70$	7170 ± 2670	$37,2\%$

FIGURE 2.8: Dynamic Light Scattering analysis of HTL (in red) purified with one affinity step at $0.50\,\text{mg}\,\text{ml}^{-1}$ ($2.00\,\mu\text{M}$) and HTL concentrated 4 times (in green) solubilized in DDM. **A**: Intensity (in %) distribution as a function of the size of the particles. **B**: Number of particle distribution as a function of their size. The table summarizes the size and polydispersity analyses results of the main peak and the solution (sol).

light. The molecular weight of DDM micelles has been evaluated at 75 kDa (Strop and Brunger, 2005), they could therefore contribute to the polydispersity of the sample. However, the presence of DDM micelles or DDM bound to HTL cannot explain such a high calculated molecular weight. It could arise from aggregation but this has not been observed by size exclusion chromatography (see Figure 2.7) or other methods. These results are therefore to be interpreted with care and only seem to indicate that DLS would not be a suitable option for the quality assessment of HTL solubilized in DDM.

Size Exclusion Chromatography - Multi Angle Laser Scattering (SEC-MALS)

Another way to make use of light scattering method is to use Size Exclusion Chromatography coupled with Multi Angle Laser Scattering (MALS). While DLS uses dynamic light scattering to evaluate the size of a particle based on its diffusion coefficient in solution, MALS relates the total scattering intensity of a particle to its molecular weight and concentration, the latter being measured by refractometry.

Hydrogenated HTL produced with one affinity step (see section 5.2.1) has been analysed previously (data not shown) and after gel filtration with similar results except for the aggregation products observed in the first case. The effect of concentration and time is investigated here. It is interesting to notice in Figure 2.9 that in all cases, the main peak of absorbance at 280 nm elutes at 12 ml, suggesting the same protein species. A peak of refractive index and light scattering is observed at 14.5 ml, corresponding to empty DDM micelles, as confirmed by measuring the buffer alone (data not shown). When the sample is concentrated, we observed the appearance of a light scattering peak at 8 ml corresponding to aggregation but most importantly, the main peak extends to lower

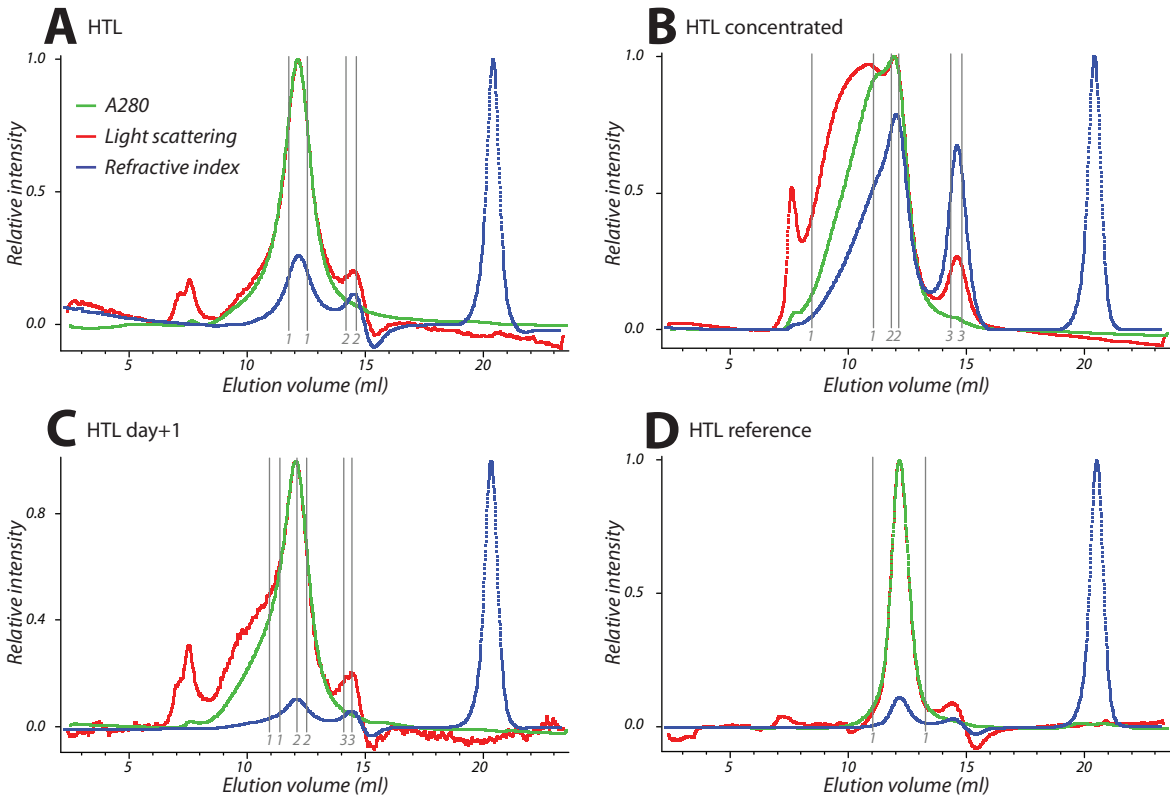


TABLE 2.3

Sample	MW_1 (kDa)	MW_2 (kDa)	MW_3 (kDa)
HTL	193,4 (± 3,1%)	90,6 (± 13,7%)	
HTL conc.	500,0 (±1,2%)	194,9 (± 2,0%)	63,7 (± 8,1%)
HTL day+1	251,6 (±4,5%)	168,1 (±4,4%)	77,0 (±13,0%)
HTL ref	165,2 (±5,5%)		

FIGURE 2.9: HTL solubilized in DDM and purified with one affinity step analysed by SEC-MALS on a Superdex 200 column while monitoring the A_{280} (in green), light scattering (in red), refractive index (in blue). **A**: HTL after size exclusion chromatography at a concentration of $0.13\,\mathrm{mg\,ml^{-1}}$ ($0.52\,\mu\mathrm{M}$). **B**: HTL concentrated ($1.39\,\mathrm{mg\,ml^{-1}}$, $5.6\,\mu\mathrm{M}$) after size exclusion chromatography. **C**: Same sample as A measured after one day at 4°C. **D**: Reference sample purified independently. The table summarizes the calculated molecular weight from the SEC-MALS data.

elution volume, suggesting a bigger particle is generated by concentration. However, considering that the relatively low UV absorbance, we concluded that these particles were probably mostly composed of DDM or aggregates. This increased concentration of DDM in the solution is confirmed by the observed increase of the DDM micelles peak at 15 ml. It is important to note that the window chosen covers a wide range of elution volume and that it probably contains a wide range of particle size, leading to an average of 500 kDa. After one day conservation at 4 °C (see panel C), an aggregation peak is also observed and the shoulder has a calculated molecular weight of 251.6 kDa. If this increased observed molecular weight could be partially attributed to the way the peak have been defined, we can also observe that the main peak has a relatively lower refractive index peak, suggesting a loss of DDM (or potentially lipids) over time, which could therefore explain the appearance of higher order particles or aggregates. Finally the reference displays a similar profile as panel A with incidentally a similar calculated molecular weight, even though the relative amount of DDM bound could be lower (low refractive index).

Taken together, these results suggest that, despite the use of a 100 kDa molecular cut-off concentrator, the concentration of DDM increased in the sample, which may lead to dissociation of the complex, formation of detergent particles with a high hydrodynamic radius and therefore strongly affect the homogeneity of the sample. The conservation of protein solution for a day at 4 °C seemed to generate aggregates as well as affecting the composition of the complex. Finally, it is important to note that even in the case of the reference, the calculated molecular weight is significantly lower (165.2 kDa) than the expected molecular weight of the Holotranslocon (255 kDa), raising questions on the integrity of the purified complex. This is further questioned by the fact that, in these analyses, we took both protein and detergents into account: our attempts to separate these contributions using the "three detector method" have led to even lower calculated protein molecular weights.

A similar experiment was done on SecYEG alone (data not shown) in order to identify whether the calculated molecular weight could correspond to a dissociation of HTL into its subcomponents. SecYEG eluted as two peaks with molecular weight calculated at 351.8 kDa and 150.4 kDa respectively, the latter being very close to what is calculated for HTL and probably corresponds to a SecYEG dimer (expected at 150 kDa). It is also to be considered that the two molecular weights measured for SecYEG could correspond to a monomer with DDM bound or dimer with DDM bound: this observation would seriously question the integrity of HTL as the DDM could account for the same intensity of signal as the proteins.

It is also important to note that the mass concentration of SecYEG sample is 2,5 times higher than for HTL, which could explain the appearance of higher order multimers that were not clearly observed as a peak in HTL sample. Finally, the same experiment has been done in a buffer containing 10% glycerol instead of 3%, which yielded sensibly identical calculated molecular weight for both HTL and SecYEG (data not shown). Similar profiles have been observed on deuterated HTL, with a calculated molecular weight of 190 kDa (data not shown).

This work used the platforms of the Grenoble Instruct-ERIC Center (ISBG : UMS

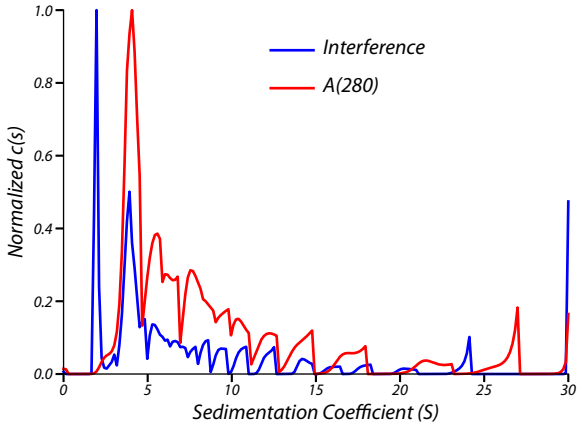


FIGURE 2.10: Analytical Ultracentrifugation analysis of HTL solubilized in DDM produced with two affinity steps. The sedimentation coefficient (S) of the sample diluted to $A_{280} = 1.2$ is analysed at 4°C in a Anti-50 rotor spinning at 40000 RPM. The absorbance at 280 nm (in red) and interference (in blue) are monitored.

3518 CNRS-CEA-UGA-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). We thank Caroline Mas and/or Marc Jamin, for assistance and/or access to the Biophysical platform.

Analytical Ultra Centrifugation (AUC)

Analytical Ultra Centrifugation (AUC) calculates sedimentation coefficient (S) that are directly related to the size of particles and therefore can be used to further assess the stability and homogeneity of the purified complex solution. We performed sedimentation velocity experiments on samples prepared with two affinity steps (seesection 5.2.1).

Figure 2.10 show the sedimentation coefficient calculated from absorbance and interference signals. The measurement have been done at three different concentrations ($A_{280} = 0.4, 1.2$ and 2.3 corresponding to 0.2, 0.6 and 1.2 mg ml^{-1}) with similar results. A contribution at 1.9S ($s_{20w} = 3.6 \pm 0.1S$) is detected only in interference and would correspond to DDM micelles, in agreement with published data (Salvay *et al.*, 2007). A large contribution of interference and absorbance between 3 and 18S with a maximum at 4S correspond to larger complexes. Considering the peak at 4S, the interference / absorbance ratio suggests a complex composed by 1.4 g of DDM per gram of protein (considering only a protein/DDM complex) while 0.9 g per gram of protein is calculated for the range 3-18S. Data modelling indicate, from the protein sequences and considering a compact globular shape, that the sedimentation coefficient of HTL should be of 6.2 (or higher if bound to DDM). On the other hand, considering the calculated sedimentation coefficient and given an intact complex, modelling suggests a highly asymmetrical and elongated particle. The binding of lipids to HTL would slightly reduce the sedimentation coefficient and considering 2 g of bound lipids per gram of protein would reduce S to 5.6 (considering a globular particle and 0.9 g of DDM per gram of protein). Finally, the value of $S = 4$ therefore suggests a globular

particle with a contribution of 77 kDa of protein complex and 1.4 g of DDM per gram of protein. Aggregates with $S > 18$ are also detected and represent 7% of the total signal.

These results therefore suggest that the complex is dissociated with a maximum at 77 kDa that could correspond to SecYEG complex. It is interesting to note that no smaller complexes can be clearly observed, suggesting that dissociation of HTL would not yield individual proteins. The discrete peaks observed between 4S and 18S could therefore comprise DFYY (180.5 kDa), SecYEG dimers (150.1 kDa) and probably a small portion of intact HTL as well as higher order multimers.

A similar experiment was performed on HTL samples prepared with one affinity purification steps (see section 5.2.1), but the data suggested a continuum of sedimentation coefficient that could not be fitted to models in agreement with HTL, therefore suggesting a high polydispersity of the sample. Taken together, these data showed that, regardless of the purification protocol used, no homogeneous Holotranslocon sample could be analysed and that HTL is likely to be dissociated.

This work used the platforms of the Grenoble Instruct centre (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB).

Mass spectrometry (MS)

Mass spectrometry (MS) can be used for different purposes in the case of protein complexes. Matrix Assisted Laser Desorption Ionization Time Of Flight (MALDI-TOF) mass spectrometry can identify the individual subunits and therefore confirm their presence in the solution. Native Electro Spray Ionization Time Of Flight (ESI-TOF) mass spectrometry gives information on the molecular weight of the intact complex and can therefore confirm the stoichiometry of the complex.

MALDI-TOF experiments were performed on HTL purified as described using the one affinity step purification protocol (see section 5.2.1). Alternatively, the last gel filtration is done in 250 mM ammonium acetate, which facilitates the detection of subunits as opposed to classical TSG buffer (data not shown). Dilution series were prepared in formic acid solution and loaded on the MALDI target with the appropriate matrix as described in subsection 5.3.4.

Figure 2.11 panel A shows low molecular mass range and one major peak at 11.341 could be attributed to SecG (expected mass: 11.365). The minor peak at 15.197 is thought to be SecE (expected mass : 15.227). Repeated independent experiments and different dilution or matrix conditions did not allow the clear identification of YajC (expected at 14.706), unless a shift of more than 300 m/z happened for both SecE and YajC (then attributed to the peak at 15.197 and 15.597). The peak at 17.581 could correspond to SecF carrying 2 charges (expected at 17.691), and the peak at 11.704 could be SecF with 3 charges (expected 11.794).

Panel B displays a wider range of m/z for high molecular weight proteins. The main peaks at 35.191, 48.234, 63.543 and 67.307 could be attributed respectively to SecF

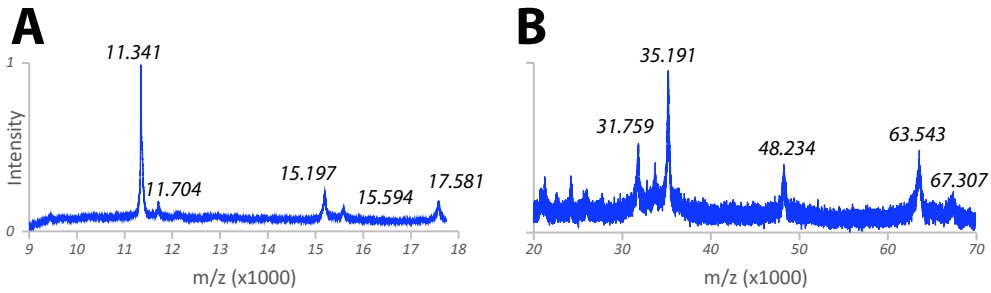


FIGURE 2.11: MALDI-TOF analysis of HTL solubilized in DDM and diluted 4 times in formic acid. Two ranges of mass/charge (m/z) have been analysed and the main peaks labelled with their m/z value. **A**:8000-18000 (represented from 9000). **B**:18000-70000 (represented from 20000)

(expected : 35.382), SecY (expected : 48.511), YidC (expected : 62.923) and SecD (expected : 67.540). The peak at 31.759 could correspond to YidC carrying two charges (expected at 31.461). While the measurement of the molecular weight is accurate for the small m/z range (0.1% off the expected value), it was noticed that this variation increased by up to 0.6% for the high molecular weight range. This could be a consequence of the high laser intensity (<50%) that was required to obtain a reasonable signal/noise ratio.

Samples for the ESI-TOF experiments were prepared by performing the last gel filtration in 250 mM ammonium acetate. The gel filtration profiles were highly distorted in these conditions but the HTL subunits could still be identified on a gel. The fractions were concentrated and submitted to ESI. However, the electro spray ionization process was impaired by recurrent aggregation in the needle and no significant signal could be acquired.

Native mass spectrometry could not be used in order to investigate the integrity and subunit ratio of the complex. MALDI-TOF confirmed the presence of six subunits but failed to identify YajC. Since the data are highly susceptible to subunit inherent capacity to be ionized by MALDI, they cannot be analysed quantitatively and can therefore only give binary information on the presence of the subunits.

2.2.2 Amphipols (Apol)

The study of membrane protein is importantly limited by the difficulty to produce and keep them soluble, the latter issue being partially solved through the use of detergents. However, because of their nature, they may also induce instability and lead to polydisperse sample. The community has therefore invested efforts in the development of alternative molecules to replace detergents while favouring the solubility and stability of membrane proteins. One class of molecule, whose development was initiated in 1996 , has proved to be an interesting substitute to detergents and has been named amphipol because of the amphiphatic and polymeric natures (Tribet *et al.*, 1996). The most commonly used amphipols, A8-35, correspond to a polyacrylate chain of about 35 residues onto which are grafted octylamine chains (typically 9) and isopropylamine

units (typically 14). The ungrafted acrylates give negative charges to the backbone, the octylamine chains have hydrophobic properties and the isopropylamine reduces the charge density of the backbone (Gohon, Pavlov, *et al.*, 2004). Because of their high density of alkyl chains, amphipols have a very high affinity for hydrophobic protein domains with a displaced equilibrium in favour of bound amphipol so that no, or traces, of free molecule suffice to keep the protein soluble, as opposed to detergent who need be used above their critical micellar concentration.

The study of the Holotranslocon solubilized in detergent has suffered from poor stability and high sample heterogeneity (see subsection 2.2.1, Botte, 2013) and amphipols have been used for low resolution cryo-EM reconstruction (Botte *et al.*, 2016). We therefore further investigated, characterized and tried to optimize the production of amphipol A8-35 solubilized Holotranslocon, in sufficient quality and quantity for small-angle scattering experiments.

Size Exclusion Chromatography (SEC)

To characterize the size distribution and homogeneity of amphipol solubilized HTL solution, we performed Size Exclusion Chromatography on samples purified with two affinity steps (see section 5.2.1), as published (Botte *et al.*, 2016). We also attempted to optimize the amphipol to protein mass ratio in the case of Holotranslocon produced with one affinity step (seesection 5.2.1) by applying increasing amounts of amphipol to constant protein concentration.

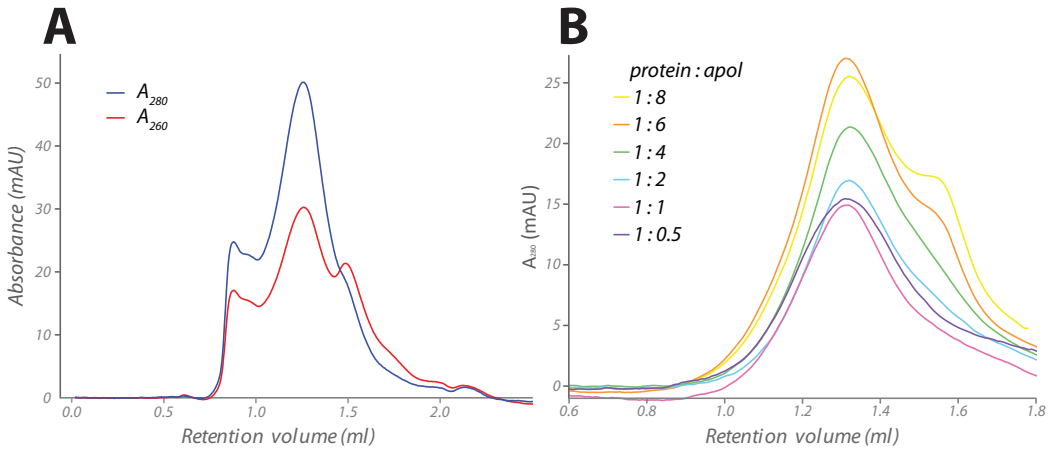


FIGURE 2.12: Size exclusion chromatography (Superdex 200 3.2/30) analysis of amphipol solubilized HTL. **A**: typical GF profile obtained from section 5.2.1 at a protein to amphipol ratio of 1:4. The absorbance at A_{280} (in blue) and A_{260} (in red) have been recorded. **B**: Test of protein:apol (1:0.5, 1:1, 1:2, 1:4, 1:6, 1:8) mass ratio for the solubilization of HTL obtained from one affinity step purification.

Panel A shows a typical profile obtained when preparing HTL with the two affinity-step purification protocol. We observe some aggregation at 0.6-0.7 ml, which corresponds to the void volume of this GF column. The majority peak eluted between 1.1 and 1.4 ml and observed a shoulder around 1.5 ml. Coomassie SDS-PAGE analysis confirmed the

presence of HTL subunits in the main peak (data not shown) and the high A_{260} suggests that the shoulder at higher elution volume is not composed by pure protein and could be n due to amphipol micelles contamination. It is however important to note that GF allows the purification of intact a Holotranslocon (Botte *et al.*, 2016) although the yield and concentration are incompatible with small-angle scattering experiments (a few micro litres at less than $1\,\mathrm{mg\,ml^{-1}}$).

Panel B show the optimization procedure attempted on a sample prepared as described in the single step affinity procedure. The DDM solubilized protein sample has been split and submitted to amphipol exchange so that the amount of protein is constant between the conditions and can therefore be analysed in a quantitative way. It is important to note that the samples had been ultracentrifugated at 150000g before being applied to gel filtration in order to get rid of potential aggregates. It was observed that, from ratio of 1:0.5 to 1:6, the more amphipols is added, the more soluble protein is recovered. However, when the protein to amphipol ratio is raised to 1:8, it was observed that the shoulder previously described in panel A became more prominent and the absorbance of the main peak is slightly reduced as compared to the 1:6 ratio. We therefore concluded that in these conditions and after the one affinity step protocol, a 1:6 HTL to amphipol ratio favoured the solubility of HTL and could be a suitable alternative to produce more soluble HTL as compared to the previously described protocol.

Dynamic Light Scattering (DLS)

As described previously (subsection 2.2.1), the homogeneity of the solution is investigated through the size distribution of particles detected by Dynamic Light Scattering. The protein was purified using two (panels A-D) or one affinity purification step section 5.2.1 (panel E) and the effect of deuteration or solvent labeling was investigated.

Figure 2.13 panel A shows purified hydrogenated HTL solubilized with amphipol and we observed the dimensions were five times smaller than what was measured in DDM (see Figure 2.8). This size (12.31 nm) corresponded to a molecular weight of 236 kDa, which would be compatible with the expected size of HTL (255kDA). It is however important to notice that the intensity and number distribution did not overlap, indicating that the solution is dominated by smaller particles and therefore was not homogeneous ($\%_{polydispersity} = 39.7\%$). The same observation is valid for other panels. Figure 2.13 Panel B corresponds to the same experiment with the last gel filtration being carried out in deuterated buffer. We calculated that the size and molecular weight were higher than in hydrogenated buffer. If the presence of moderate aggregation cannot be clearly dismissed (a small amount of large aggregates are detected at 300 nm), it is important to note that D_2O has a superior viscosity than H_2O (25 to 30% higher, R. C. Hardy and Cottington, 1949); particles would therefore have a slower diffusion coefficient and appear bigger than they actually are; this could explain why larger molecular weight were measured in deuterated buffers.

Panels C and D correspond to the same experiment performed on deuterated HTL. We observed that the detected particle size was smaller than before, suggesting dissociation or the presence of sub-complexes (SecYEG, DFYY). The higher polydispersity

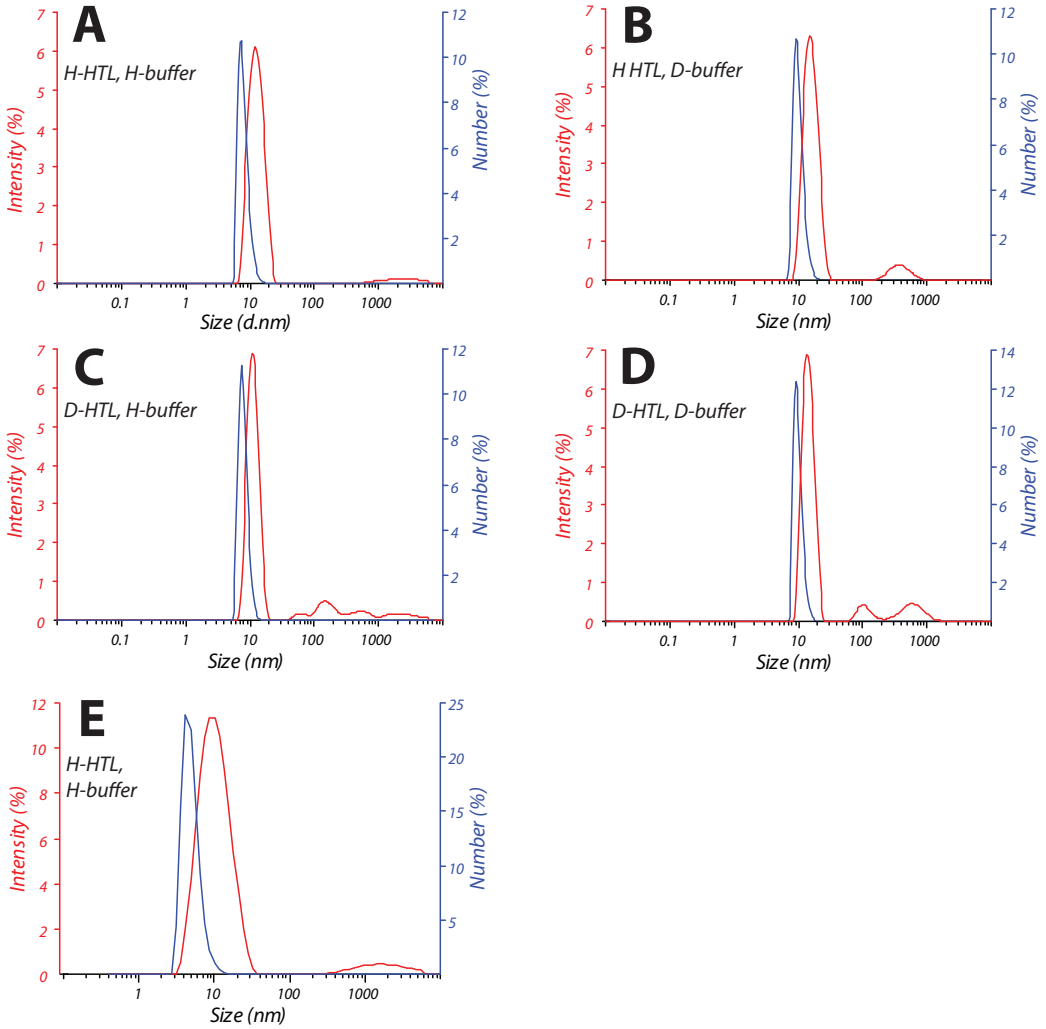


TABLE 2.4

	Size_{peak}(nm)	$\text{MW}_{peak}(kDA)$	$\text{Size}_{sol.}$(nm)	$\text{MW}_{sol.}(kDA)$	$\%\text{PD}_{sol.}$
A H-HTL, H-buffer	$12,3 \pm 3,4$	$236,5 \pm 68,4$	$12,6 \pm 5,0$	$247,0 \pm 98,3$	$39,7\%$
B H-HTL, D-buffer	$14,8 \pm 4,0$	$364,5 \pm 109,2$	$16,5 \pm 7,5$	$467,0 \pm 212,7$	$45,6\%$
C D-HTL, H-buffer	$10,7 \pm 2,3$	$171,0 \pm 39,3$	$13,8 \pm 8,2$	$307,0 \pm 182,7$	$59,5\%$
D D-HTL, D-buffer	$13,5 \pm 3,0$	$293,6 \pm 71,3$	$17,1 \pm 9,7$	$507,0 \pm 289,4$	$57,1\%$
E H-HTL, H-buffer	$8,7 \pm 5,1$	$105,6 \pm 84,6$	$9,7 \pm 4,6$	$136,0 \pm 64,8$	$47,7\%$

FIGURE 2.13: Dynamic Light Scattering analysis of H and D-HTL solubilized in amphipols and prepared from the two affinity steps protocol in labelled buffers. The intensity per size (in red) and number of particle per size (in blue) are represented. **A**: Hydrogenated HTL in hydrogenated buffer at $0.86\,\text{mg}\,\text{ml}^{-1}$ ($3.44\,\mu\text{M}$). **B**: H-HTL in deuterated buffer at $0.96\,\text{mg}\,\text{ml}^{-1}$ ($3.84\,\mu\text{M}$). **C**: Deuterated HTL in H buffer at $0.80\,\text{mg}\,\text{ml}^{-1}$ ($3.20\,\mu\text{M}$). **D**: D-HTL in D-buffer at $0.98\,\text{mg}\,\text{ml}^{-1}$ ($3.92\,\mu\text{M}$). **E**: H-HTL in H-buffer purified by the one step affinity protocol at $0.5\,\text{mg}\,\text{ml}^{-1}$ ($2.0\,\mu\text{M}$). The table summarizes the size (particle's diameter) and polydispersity analyses results of the main peak (*peak*) and the solution (*sol.*).

indexes (57,1 and 59,5%) supported this hypothesis. We observed the same behaviour as for the hydrogenated complex in D_2O with a slight increase in measured molecular weight and the presence of large aggregates, though also detected in hydrogenated buffer. It is also important to notice that deuterated HTL was produced in high cell density culture which was shown to distort subunit expression balance. The presence of smaller sub-complexes could explain both smaller size observed as well as the increased polydispersity as compared to hydrogenated samples.

In the case of the sample prepared with one affinity step (see section 5.2.1) at a protein amphipol mass ratio of 1:6, we observed a significantly lower size and hence molecular weight, suggesting dissociation or imbalanced subunit expression leading to prevalence of SecYEG in solution (see Figure 2.6).

If the solubilization by amphipol of the Holotranslocon gave significantly better results by DLS than in DDM, the calculated polydispersity of the solution remained too high ($> 35\%$), probably because of the presence of high molecular weight aggregates. The quality of these preparation could be improved by the classical approaches (GF, centrifugation) and concentration of the sample by ultra filtration further increased the polydispersity of the solution (data not shown). Taken together this data tend to suggest that the two affinity steps protocol produces samples of higher quality but the yield and tendency to aggregate remain major drawbacks.

Analytical Ultra Centrifugation (AUC)

Analytical Ultra Centrifugation (AUC) was done together with DDM solubilized HTL in order to compare the samples. The proteins were purified with two affinity steps (see section 5.2.1). The measurement were done at three different concentrations ($A_{280} = 0,4,\ 1,2$ and $1,7$ corresponding to 0,2, 0,6 and $0.9\,\mathrm{mg\,ml^{-1}}$).

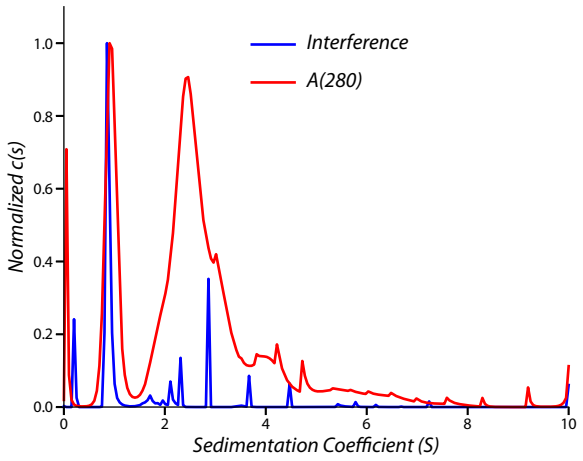


FIGURE 2.14: AUC analysis of HTL solubilized in amphipol A8-35 produced with two affinity steps. The sedimentation speed of the sample diluted to $A_{280} = 1.2$ is analysed at 4°C in a Anti-50 rotor spinning at 4000 RPM. The absorbance at 280 nm and interference are recorded.

Figure 2.14 shows an interference peak at 0.9S that represents 69% of the total interference signal, the corresponding s_20, w of $1.7 \pm 0.2S$ correspond to what was expected for amphipol micelles (Gohon, Giusti, *et al.*, 2006). This observation was further confirmed by the analysis of interference and absorbance ratios that suggested 13.5 g of amphipols are bound per gram of protein, a value largely superior to published results (Gohon, Giusti, *et al.*, 2006). The large contribution between 1.5 and 8S showed discrete peaks in interference (higher concentrations showed the same behaviour in absorbance, data not shown) which are most likely not relevant. The signal is therefore considered over the whole range instead. This region corresponding to 78% of the absorbance contribution, indicated the presence of protein. The maximum peak at 2.55S ($s_20, w = 4.8 \pm 0.2S$) showed shoulders that could correspond to multimers. The analysis of absorbance and interference data suggests 0.6 g of amphipols were bound per gram of protein. Considering a monomeric globular complex, the expected sedimentation coefficient should be of 6.2S (or superior if bound to amphipols). On the other hand, considering the calculated sedimentation coefficient and a monomeric complex, the data suggested an unreasonably elongated and asymmetrical shape. If the potential binding of lipids would slightly reduce the sedimentation coefficient (down to 4.5S if we consider 2 g of bound lipid per gram of protein), this value of S=2.55 suggested that the particles were smaller than expected (evaluated at 60 kDa for a globular protein bound to amphipols). Minor aggregation at 8S was also observed.

These result suggested that the amphipol solubilized HTL dissociated into smaller complexes (the peak at 2.55S could correspond to SecYEG monomer bound to some lipids) or that the solution was polydisperse. It is to be considered that the high forces to which the particles are submitted during ultra centrifugation could favour the dissociation of the inherently unstable HTL into smaller complexes.

This work used the platforms of the Grenoble Instruct centre (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB).

Mass spectrometry (MS)

Holotranslocon produced using one affinity step (see section 5.2.1) was solubilized with amphipols at a protein to amphipol mass ratio of 1:6 and concentrated. The solution was diluted in formic acid and deposited on the Matrix Assisted Laser Desorption Ionization (MALDI) target coated with matrix. Several dilution factors, laser intensity and matrix composition were tried.

Figure 2.15 shows the spectrum obtained from a 4 fold dilution and deposition on dried matrix. Only one peak at 11.383 is clearly seen. This peak most likely corresponded to SecG with its expected molecular weight of 11.365 kDa and is similar to that found in Figure 2.11. A minor peak, but consistently observed during other experiments was also detected at 15.210 and could correspond to SecE (expected at 15.227 kDa). YajC, which is meant to be in the same m/z range, could not be detected. A higher m/z range (18000-70000) was analysed but no clear signal could be identified, regardless of the matrix strategy, dilution and laser intensity used.

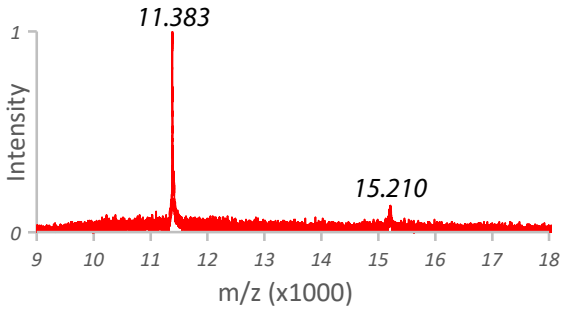


FIGURE 2.15: MALDI-TOF analysis of HTL solubilized in amphipols and diluted four times in formic acid. The 8000-18000 (represented from 9000) ranges of mass/charge (m/z) has been analysed and the main peaks labelled with their m/z value.

These data could only confirm the presence of SecG and since the subunits could all be detected on SDS-PAGE (data not shown), we believe that amphipols could have interfered with the ionization of large subunits. It is also important to note that amphipol A8-35 are to be used in alkaline environment so that their backbone remain negatively charged ; the dilution in formic acid could have disrupted the amphipol belt around protein, therefore leading to aggregation of the proteins into large particles that could not be efficiently ionized or detected.

2.3 Interaction with SecA

The SecA ATPase has been reported to interact with both SecYEG and the Holotranslocon (Schulze *et al.*, 2014) and provides the energy required for the translocation of proteins through the membrane. How SecA interacts with HTL and most importantly how it affects the structure and composition of the translocons is not fully understood (Schulze *et al.*, 2014). In order to investigate these aspects, notably by small-angle neutron scattering, we produced and purified both hydrogenated and deuterated SecA. The interaction of SecA with HTL has also been investigated by SEC.

2.3.1 Production of hydrogenated and deuterated SecA

SecA-His6 was successfully expressed from BL21 DE3 cells in DYT medium and purified as described in subsection 5.2.3. High cell density culture is required for the deuteration of biomolecules. However, since this process is much longer than flask expression, the stability of the plasmid becomes a major concern. In the case of high cell density culture, beta lactamase, the enzyme used to provide bacteria with ampicillin resistance, leaks into the medium therefore hydrolysing the β-lactam ring of the antibiotic, leading to the loss of environmental pressure that constrain bacteria to keep their heterologous plasmid (Haertlein *et al.*, 2016). The use of kanamycin resistance is therefore preferred.

Kanamycin *in vitro* transposition was therefore attempted on the SecA-His6 plasmid as described in section 5.1. Since the insertion of the transposon is random, we used

a double selection procedure to ensure that kanamycin resistance was gained at the expense of the loss of ampicillin resistance. By doing so, we confirmed that the transposon was integrated in the region containing the ampicillin resistance gene, therefore limiting the chances of disrupting the sequence of interest for expression (origin of replication, insert, promoter...). Despite repeated attempts, we could not isolate a colony that satisfied the aforementioned conditions.

It was decided to make use of another plasmid, the SecAΔ_{cys} (Appendix A) which corresponds to the same sequence with the cysteines replaced by alanines (Erlandson *et al.*, 2008) that already has a kanamycin resistance. The SecA protein was therefore successfully expressed in flasks or by high cell density culture in 85% deuterated medium (section 2.1.1).

2.3.2 Purification

The purification of SecA was performed as described in subsection 5.2.3 by a combination of nickel IMAC, anion exchange and gel filtration purification. No significant differences could be observed between deuterated and hydrogenated proteins.

Figure 2.16 shows a typical purification of SecA. Panel A represents the nickel Immobilized Metal Affinity Chromatography (IMAC) step where we see that a large amount of contaminants was lost while washing at 30 mM imidazole. The elution fractions obtained at 330 mM imidazole were dominated by SecA at 100 kDa, as seen on panel D. Panel B shows the anion exchange chromatography profile; on panel D, a it is observed that small contaminant is not retained by the anion exchange column and that a fraction of SecA did not bind. A large peak (at 300 mM KCl) is followed by a smaller peak (600 mM KCl), the first being confirmed as SecA on panel D. Finally, the SEC results (panel C) show that SecA eluted as a single symmetrical peak between 60 and 80 ml (0,5-0,6 CV). A minor contribution of higher molecular weight particles can be observed between 40 and 60 ml (0,3-0,5 CV); panel B suggests that these could correspond to SecA multimers (even though dimerization should have been prevented by cysteine deletions and the 1 mM DTT) or aggregates. SecA was nonetheless purified with high efficiency and with minor contaminations (less than 5% of the signal).

2.3.3 Size exclusion analysis of the interaction

The potential binding of HTL to SecA was investigated by SEC. HTL was purified with one affinity step (see section 5.2.1) in DDM and obtained at a concentration of 3 µM. SecA was then added in equimolar quantities to HTL solution together with 2mM $MgCl_2$ and incubated on ice before loading of Superdex 200 10/300.

Figure 2.17 panel A shows the GF profile of the different solutions. All experiments were done under the same concentration conditions so that the profiles are on an absolute scale. For HTL and SecA alone, we observed similar elution volumes (14.2 - 14.5 ml) but with a long shoulder for HTL, suggesting again inhomogeneity of the peak.

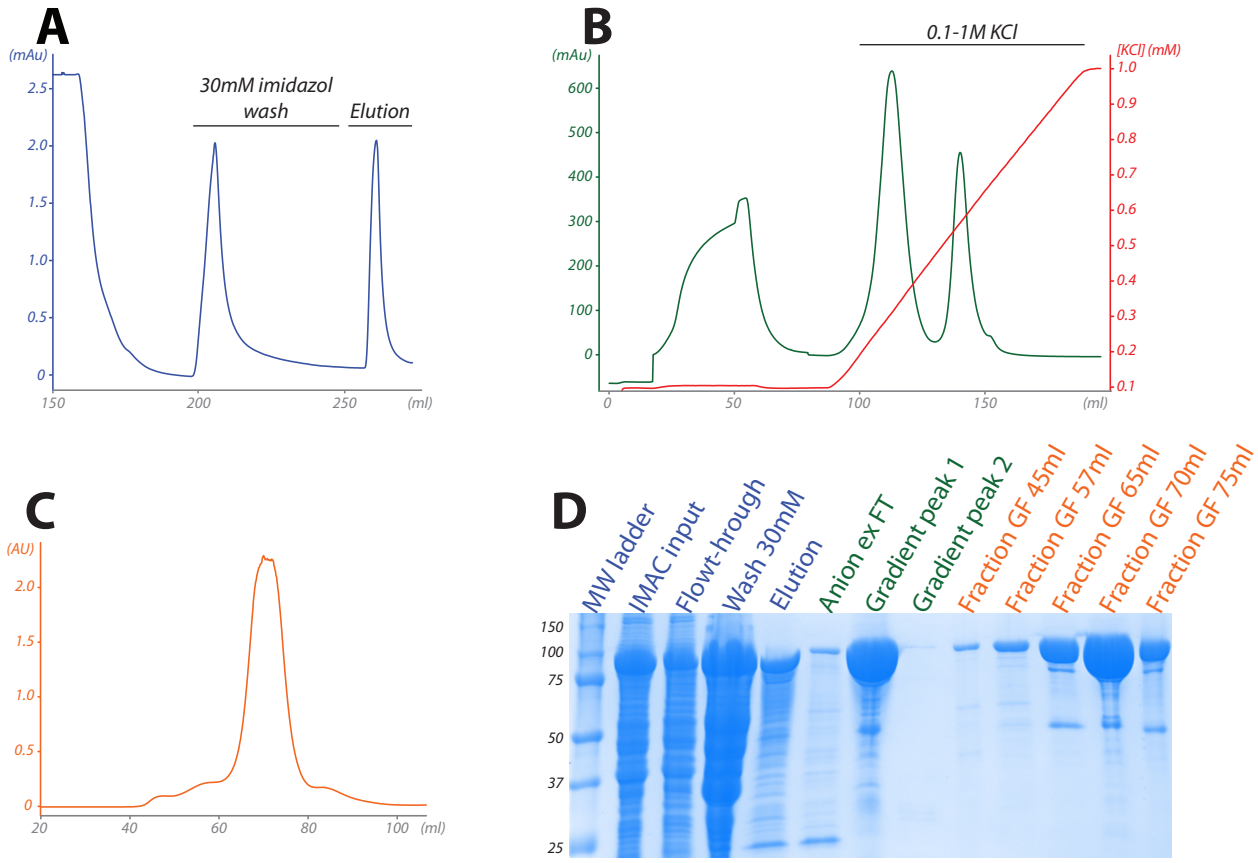


FIGURE 2.16: Purification of 85% deuterated Δ_{cys}-SecA. **A**: Nickel affinity purification profile (in blue). **B**: Anion exchange (Q column) purification profile with the A_{280} (in green) and salt gradient (in red). **C**: Size exclusion chromatography profile (in orange) performed on a Superdex 200 16/60 . **D**: SDS-PAGE Coomassie stained PAGE of the purification procedure. The wash of the IMAC column is carried out with imidazole and the elution with 330 mM imidazole.

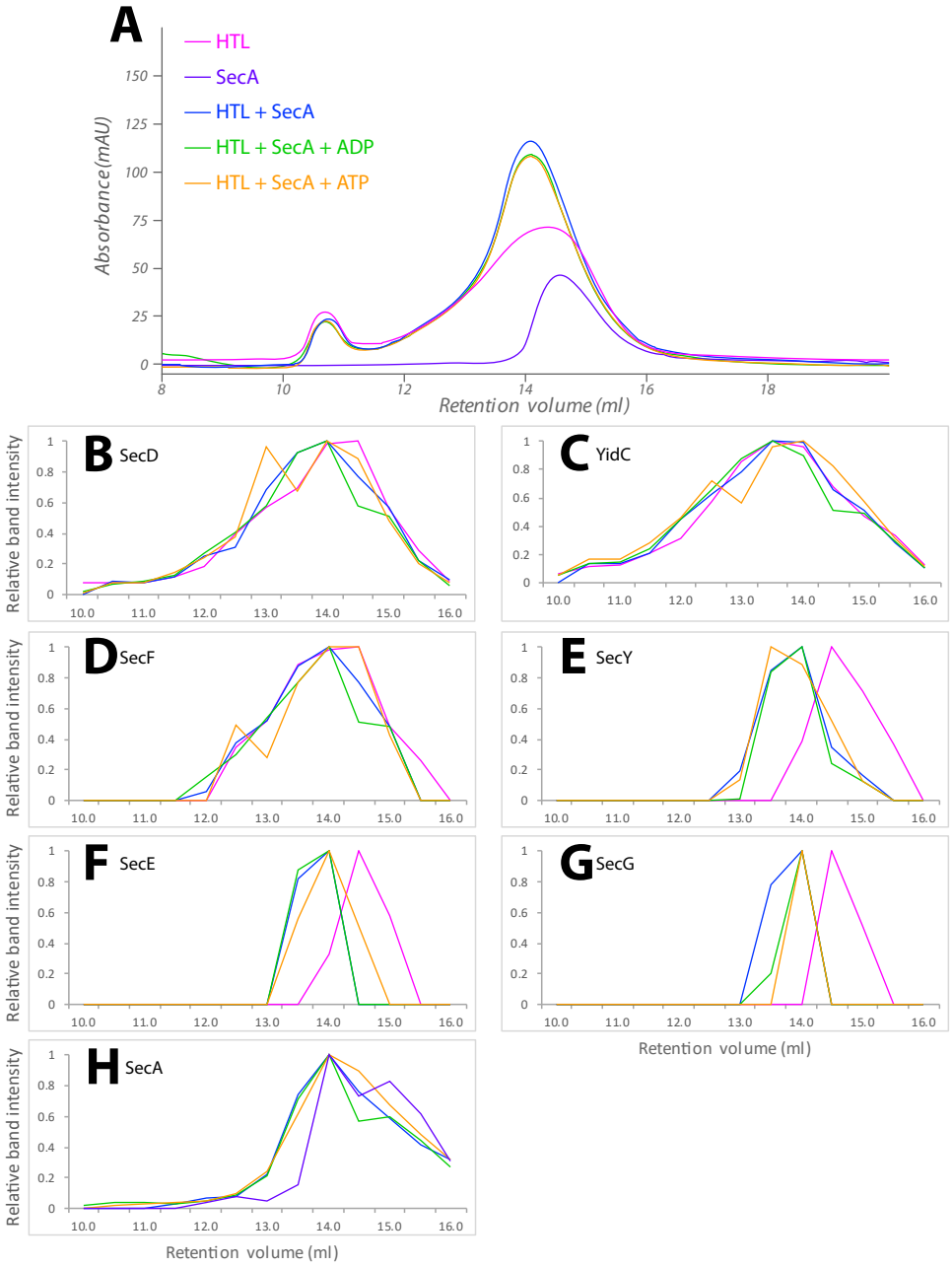


FIGURE 2.17: Interaction of SecA with DDM-solubilized HTL analysed by size exclusion chromatography. HTL alone (in pink), SecA alone (in purple), HTL + SecA (in blue), HTL + SecA + 1mM ADP (in green), HTL + SecA + 1mM ATP (in orange). A: Size exclusion chromatography profiles. B-H: Relative band intensity per subunit quantified on Coomassie stained SDS-PAGE from the GF fractions of panel A.

When SecA was added to HTL, we observed a shift of the peak to smaller fractions (14 ml), suggesting an interaction of SecA with HTL.

Figure 2.17 panels B-H correspond to the Coomassie stained SDS-PAGE analysis of the size exclusion chromatography fractions obtained in panel A. Each panel represent the relative quantification on gel of the subunit in the elution fractions, per interaction condition. Each band intensity was normalized with regard to the highest intensity measure on the gel so that the results coming from different gel could be compared.

For panels B, C and D, which correspond to SecD, YidC and SecF,it was observed that, regardless of the condition, the maximum band intensity is located in the 14 ml fraction. It may be interesting to note that for all three subunits, in the case of an interaction with SecA in presence of ATP, there seem to be a increased representation of these subunits at 12.5 ml or 13 ml. The peak remains however for all three proteins, in the 14 ml elution fraction. In the case of panel E, F and G (which correspond to SecY, E and G respectively), the subunit was found in the 14.5 ml fraction for HTL alone. When SecA was added, the maximum band intensity shifted to lower elution volumes (mostly 14 ml but also 13.5 ml). No difference could be observed when supplementing the medium with nucleotides.

These data therefore showed that when purified SecA is added to purified HTL, SecD, YidC and SecF do not bind to SecA while SecYEG shifted toward lower elution volume fractions; this suggested that an interaction had occurred. However when analysing the intensity of the bands corresponding to SecA (see panel H), the maximum of intensity is located in the 14 ml fraction, regardless of the interaction conditions. Upon addition of nucleotides, bands corresponding to 80% of the maximum signal were observed in the fraction at 15 ml, suggesting that a small portion of SecA interacted with HTL/SecYEG.

It is also interesting to notice that SecDF and YidC shifted toward lower elution volumes upon addition of SecA + ATP (in orange). However, these results are difficult to interpret since for SecF and YidC, the signal in the earlier fractions represent only 50 and 70% of the highest intensity recorded for these proteins. In the case of SecYEG, the shift observed undoubtedly corresponds to the maximum intensity recorded. Furthermore, while YidC and SecF peaked at 12.5 ml, SecYEG was completely absent from this fraction and peaked at 13-14 ml. If the effect of SecA on the migration of SecYEG was clear and the majority of the proteins shifted toward higher elution volumes, its effect on SecDF-YidC was moderated and only observable in presence of ATP.

Even though these results were not striking for SecA, these experiments seem to indicate that when added to HTL solution, the ATPase interacted mainly with SecYEG. These observations can be interpreted in three different ways :

- HTL is dissociated before the experiment and SecA binds mainly to SecYEG. This interpretation would be in agreement with previous biophysical result indicating that HTL is not stable in DDM solution.
- As it binds to HTL, SecA could induce dissociation of the complex into SecYEG + HTL. The expected size of the different sub-complexes (SecYEG+SecA : 175 kDa, DFYY : 175 kDa) would explain that they could not be separated by SEC.

- SecA stabilizes HTL and would prevent the dissociation of SecYEG. On gel filtration, this effect would induce a shift of SecYEG toward higher elution fractions when SecA is present.

The current data do not allow a definitive interpretation, although the data reported in section 2.2 tend to suggest that HTL is probably not intact in the first place. The fact that only SecYEG shifted upon addition of SecA + ADP and that it is absent from the fractions were SecDF-YidC shifted upon addition of SecA + ATP seems to indicate that the sub-complexes SecYEG and SecDF-YidC migrated independently in this experiment.

2.4 Small Angle Neutron Scattering (SANS)

Small-angle neutron scattering (SANS) has been previously used in order to characterize the Holotranslocon (Botte *et al.*, 2016). Most importantly, this method provided new information on a unique structural feature of this large complex. Cryo-electron microscopy has revealed that a cavity could be observed within the structure of the complex but insufficient resolution did not allow its composition to be identified. The analysis of scattering length densities and radius of gyration measured by SANS showed that the cavity was filled with lipids. These combined results led to the establishment of a working model that proposed the lipid cavity plays an important role in the function of the Holotranslocon.

One of the problems associated with the analyses of such a complex macromolecular assembly is that it is composed of three types of components: proteins, lipids and surfactants. SANS, uniquely, has the capacity to distinguish between the different components as long as their scattering length densities are sufficiently different. Proteins have a well-defined scattering length density that allows them to be matched out in 40-42% D_2O; lipids and surfactants have scattering length densities that are similar. The surfactant used ins this study have a contrast match point of 21,7% for DDM (Botte *et al.*, 2016) and 23,5% for amphipol A8-35 (Gohon, Pavlov, *et al.*, 2004) while the lipids have a contrast match point estimated to 14,2% from the composition of the inner *E. coli* membrane (see section 4.1.2). The contrast between protein and the other components is sufficient (20-30%), but small between lipids and surfactants (less than 10%).

A better contrast between the different components can be achieved by deuterating the sample, therefore bringing the contrast match point of both proteins and lipids up. The data described in chapter 4 showed that contrast between protein and lipids, remained mainly unchanged (superior to 20%). Therefore, deuterating the sample would allow the different components to be distinguished thanks to the dramatic increase of the contrast between lipids and surfactants (superior to 40%). The deuteration was carried out in a medium containing 85% of D_2O in order to achieve match-out labelling of the protein.

SANS was exploited to analyse the composition and structural features of HTL samples prepared in different conditions:

- Hydrogenated HTL stabilized with amphipols and purified with two affinity steps (see section 2.1.3)
- Hydrogenated HTL solubilized with amphipols and purified with a single affinity steps (see section 2.1.3)
- Hydrogenated HTL solubilized with DDM and purified with a single affinity steps (see section 2.1.3)
- Deuterated HTL solubilized with amphipols and purified with two affinity steps (see section 2.1.3 and section 2.1.1)
- Deuterated HTL solubilized with amphipols and purified with a single affinity steps (see section 2.1.3)
- Deuterated HTL solubilized with DDM and purified with a single affinity steps from a high cell density culture induced for six hours (see section 2.1.1)
- Hydrogenated SecYEG solubilized with DDM, used as a reference

The data were collected on the instrument D22 at the Institut Laue Langevin, as described in section 5.4. The buffer was the same as the last buffer used in the purification process, depending on the purification method and surfactant used. Samples were measured in 1 mm quartz cuvettes at 6 °C. The detector distances were of 1.4 m and 8.6 m and the neutron wavelength was 6 Å.

The Guinier fits were performed on buffer or constant-subtracted data using the Primus software (ATSAS suite, Svergun, Richard, *et al.*, 1998) in order to access forward scattering values (I_0) and radii of gyration (Rg). These data were treated in order to access specific information:

- The molecular weight of the sample (see Equation 5.26)
- The scattering length (see Equation 5.25) and the contrast match point
- The composition of the particle (see Equation 5.28)
- An analysis of the components sizes (see section 5.4.6)

2.4.1 Hydrogenated HTL and SecYEG (DDM)

The data collected for SecYEG and hydrogenated Holotranslocon stabilized with amphipol A8-35 were studied first. A summary of the Guinier fit results can be found in Table 2.5, the Guiniers fits are represented in Appendix C and the data shown on Figure 2.18.

The data acquired for HTL purified with two steps (see section 2.1.3) are not shown in the figures and tables. We firstly showed, biochemically and biophysically, that the sample could not be reliably purified and yielded a polydisperse solution (see section 2.1.3 and subsection 2.2.2). In addition to the low yield associated with this method of purification, it was shown by AUC that the sample was highly prone to dissociation. In Table 2.5, it is seen that the abnormally small radii of gyration measured

TABLE 2.5: Summary of SANS data for hydrogenated HTL and SecYEG (DDM)

Guinier Analysis

Sample	Conc.(mg ml^{-1})	%D_2O	I_0	$\sigma(I_0)$	Rg	$\sigma(Rg)$
h-SecYEG	1,83	0,0%	0,560	0,006	41,0	0,7
	1,74	20,0%	0,055	0,008	44,0	2,8
	1,56	60,0%	0,900	0,004	36,1	0,3
	1,46	80,0%	2,150	0,004	37,2	0,1
	1,33	96,0%	3,400	0,006	37,4	0,1
h-$HTL_{1StepApol}$	0,86	0,0%	0,200	0,002	40,0	0,4
	0,87	10,0%	0,079	0,002	44,1	0,9
	0,88	20,0%	0,019	0,002	38,4	3,8
	0,90	40,0%	0,033	0,001	31,3	0,9
	0,92	60,0%	0,280	0,001	37,4	0,2
	0,96	94,3%	1,090	0,004	37,6	0,1
h-$HTL_{1StepDDM}$	0,70	0,0%	0,320	0,004	59,1	1,0
	0,70	10,0%	0,130	0,003	58,8	1,7
	0,69	24,4%	0,008	0,002	12,9	0,9
	0,70	35,0%	0,042	0,001	31,0	1,4
	0,51	60,0%	0,410	0,004	47,8	0,7
	0,20	75,0%	0,300	0,002	44,6	0,5

Theoretical values

Complex	*Stoichiometry*	$CMP_{protein}$(%)	$MW(kDa)$	Rg(Å)
SecYEG	Monomer	37,6%	75,1	$25,7_{pdb:5AWW}$
SecYEG	Dimer	37,6%	150,1	$36,4_{pdb:5AWW}$
HTL	Monomer	38,8%	255,6	44,0 (Botte *et al.*, 2016)

Analysis of the forward scattering

Complex	CMP(%)	$CMP_{nonprot}$(%)	$\%_{protein}$	$\%_{surfactant}$	$\%_{lipid}$
h-SecYEG	25,4 ±2,0	20,8	27,2	63,8	9,0
h-$HTL_{(1stepApol)}$	28,3 ±1,6	16,0	55,0	7,3	37,7
h-$HTL_{(1stepDDM)}$	26,2 ±1,8	18,7	37,5	35,7	26,7

Analysis of the radii of gyration

Complex	$\alpha_{Stuhrmann}$	Rm(Å)	$Rg_{nonprot}$(Å)	Rg_{prot}(Å)
h-SecYEG	>0	38,2	33,1	47,9
h-$HTL_{(1stepApol)}$	>0	39,0	32,3	44,4
h-$HTL_{(1stepDDM)}$	>0	51,2	41,3	68,9

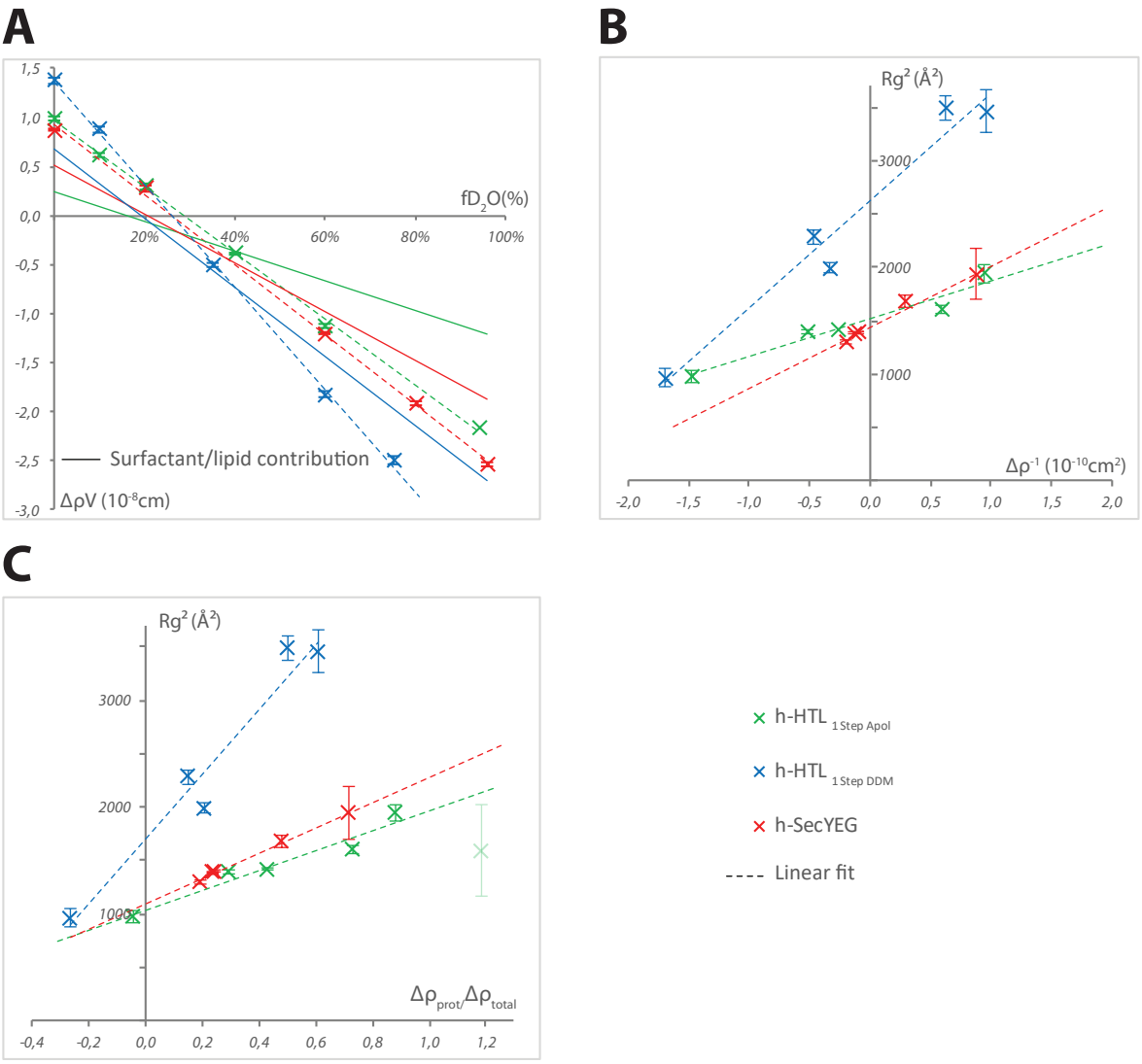


FIGURE 2.18: Small-angle neutron scattering results of hydrogenated Holotranslocon solubilized in amphipols or DMM and SecYEG in DDM. The HTL sample prepared with two affinity purification step was omitted. HTL's data are represented in green for amphipol (Apol), blue for DDM and SecYEG in red. The dashed lines represent linear fits. **A** Excess scattering length ($\Delta\rho V$) plotted against the fraction of D_2O allows to calculate the contrast match point of each complex. The solid lines represent the scattering length of surfactants and lipids after subtraction of the protein's contribution. **B** Stuhrmann analysis of the radii of gyrations plots the squared Rg against the inverse of the contrast $\Delta\rho V$ and gives indications on the relative organisation of the scattering density phases. **C** Parallel axis analysis plots Rg^2 against the ratio of calculated contrast over the protein's contrast in order to get information on the radii of gyration of each phase. The Rg calculated from the 20% D_2O contrast for HTL in amphipols was excluded from the Rg analysis due to its high error. The data point measured at 24,4% for HTL solubilized in DDM was omitted due to its high errors and inconsistency with the other data.

at high concentration of D_2O supported this observation. The slope of the parallel axis analysis also suggest a negative radius of gyration for the non-protein component, consequence of the important polydispersity of the sample. If the calculated molecular weight seemed to indicate a large complex was present, it is important to note that this value had to be extrapolated from the data measured at 0% D_2O, implying that it reflects the molecular weight of the protein but also of the surfactant and lipids and may therefore be largely over-estimated. Taking into account these factors, we did not further consider this sample.

Analysis of SecYEG

In order to have a reference point, SANS experiment on SecYEG solubilized with DDM was performed. The analysis of the scattering length density is shown in panel A of Figure 2.18. The analysis of the radii of gyration are represented in panels B and C, and the results summarized in Table 2.5.

Molecular weight: To evaluate the molecular weight of the protein itself, we used the forward scattering measured at 20% D_2O as it was very close to the theoretical molecular weight of the detergent (21.7%). Using the theoretical contrast match point of SecYEG (calculated with MULCh, Whitten *et al.*, 2008) and the value measured when the detergents were matched out, we extrapolated the forward scattering value (I_0) that would have been obtained in the absence of detergents. The molecular weight calculated was 152.6 kDa and corresponded to a dimer of SecYEG, which has been described as the predominant form (Hizlan *et al.*, 2012; Park and Rapoport, 2012).

Contrast match point: The scattering analysis showed a contrast match point of $25,4 \pm 2,0\%$, which was significantly lower than the theoretical value for SecYEG ($37,6\%$, calculated from its sequence). This indicated that the particle measured was not composed of protein alone, but also of other components with a lower contrast match point. By subtracting the contribution of the protein from the calculated scattering length density, it was possible to determine that the contrast match point of the "non-protein" components was $20,8\%$. Considering that SecYEG is a membrane protein purified using detergents, it was reasonable to assume that this "non-protein" component was composed of lipids and detergents. Considering that the remaining density has a contrast match point of $20,8\%$, it is also reasonable to assume, within the experimental errors, it is essentially composed of detergents ($CMP = 21,7\%$). Considering the calculated values, it was estimated that the SecYEG particles were composed of 27,2% protein, 63,8% DDM ($\approx$ 607 molecules) and 9,0% lipids ($\approx$ 52 molecules), which corresponds to a 1:2,3:0,3 protein:DDM:lipid volume ratio.

Radius of gyration: The radius of gyration of the SecYEG monomer was calculated to be 25.7 Å from the crystal structure of a monomer (PDB:5AWW, Tanaka *et al.*, 2015) or 36.4 Å for the dimer (generated from PDB:5AWW) using CRYSON (v2.7) (Svergun, Richard, *et al.*, 1998) and considering a hydrogenated protein in 0% D_2O

buffer. The Stuhrmann analysis allowed the calculation of a radius of gyration at infinite contrast (Rm) of 38.2 Å. The parallel axis analysis allowed the evaluation of the radius of gyration of the protein to be of 47.9 Å while the non protein component had a radius of gyration of 33.1 Å. Finally, the positive slope ($\alpha_{Stuhrmann} > 0$) observed from the linear fit applied to the radii of gyration suggest that the two centre of masses were aligned but also that the protein phase was arranged around the detergents. This last observations was rather unexpected as it indicated most of the detergent was not disposed around the protein. However, it is conceivable that detergents may be accumulated along the interface of the two copies of SecYEG. It appeared as a dominant compact mass of detergent in the particle while the surrounding detergents are forming a thin layer around the rest of the protein.

Holotranslocon solubilized with amphipols (one step affinity purification)

The Holotranslocon prepared as described in section 2.1.3 and stabilized with amphipols was measured. The scattering length density and radii of gyration were analysed.

Molecular Weight: The molecular weight of the particle was calculated using the forward scattering value measured in 20% D_2O. In the same way as for SecYEG, the theoretical contrast match point value and the value that was the closest to amphipol contrast match point (23,5%) was used to extrapolate the forward scattering value at 0% D_2O. This extrapolation corresponds to the value that would have been measured if no surfactant was present on the particles. A value of 409.8 kDa was calculated, which is too small to correspond to an oligomer but significantly larger than what we would expect from a complex that should weigh 255 kDa (see Table 2.5).

This observation could be attributed to different effects. Firstly, the data were collected at 20% D_2O, solvent fraction that is the closest to the match point of the particle. The error associated with that data point is large (10,0%) while the other points have error ranging from 0,4% to 2,5%. Additionally, the contrast match point of the surfactant was further from 20% than for SecYEG, adding uncertainty to the extrapolated values. As previously published (Botte *et al.*, 2016) and further investigated later on, the complex might contain lipids which would therefore increase the calculated molecular weight beyond the expected value. The observed high molecular weight calculated could then be a consequence of the uncertainties linked to the contrast used and the potential presence of lipids.

Contrast match point: We analysed the scattering length density of the sample and calculated the contrast match point (CMP). In a similar way as for SecYEG, the measured value (28,3±1.6%) was far from the expected theoretical CMP for protein alone, an observation attributed to the presence of surfactants (here amphipol). The contribution of protein was subtracted from the sample and the contrast variation curve extrapolated for the other components (see Equation 5.29). Strikingly, in the case of HTL, the calculated CMP of the "non-protein" fraction (16,0%) was significantly lower

than that of amphipol (23,5%). This has been attributed to the presence of lipids (Botte *et al.*, 2016). Considering that the contrast match point of *E. coli* inner membrane lipids is evaluated at 14,2% (see section 4.1.2), it was calculated that the particles analysed were composed of 55,0% protein, 7,3% amphipol (≈ 3 molecules) and 37,7% lipids (≈ 182 molecules).

In order to compare these results to published values (33%:26%:41% protein:DDM:lipids) and with SecYEG (27,2%:63,8%:9,0% protein:DDM:lipid), these ratio were normalized with regard to the protein fraction. Volume ratios of protein:surfactant:lipid of 1:0,1:0;7, 1:2,3:0,3 for h-HTL in amphipols, SecYEG in DDM were therefore calculated. The ratio was of 1:0,8:1,2 for published results (HTL in DDM). Striking difference were observed as our sample had relatively lower levels surfactants and fewer lipids bound as compared with published results. It is known that less amphipol is required as compared to DDM to keep a protein soluble (Tribet *et al.*, 1996) but they have also been reported to preserve membrane protein/lipid interaction (Zoonens and Popot, 2014). On the other hand, medium resolution analysis of HTL by cryo-EM has reported similarities in terms of the dimensions and shape between amphipol-solubilized HTL and DDM-stabilized cross-linked HTL, the latter having been associated to lower quantities of bound lipids (Botte *et al.*, 2016). Indeed, the cross-linked HTL was analysed to have a protein:lipid ratio of 1:0,6, in agreement with our results.

It was however surprising to observe that the scattering length densitiy values for this sample was similar to that of SecYEG (solubilized in DDM) but also significantly different than HTL solubilized in DDM. The fact that the slope of the latter sample is higher indicate a bigger complex. This observation could be associated with two effects. Firstly, when DDM is used to solubilize membrane proteins, a significantly larger amount of detergent is attached to the protein than in the case of amphipol solubilized HTL; this implies that the scattering particle solubilized with DDM will had a larger volume, a higher total molecular weight and incidentally a higher scattering length ($\Delta\rho V$). Secondly, it is possible that HTL in amphipols was not intact and that the solution measured was composed of a mixture of HTL and dissociation products; if the subtraction of HTL's contribution gave reasonable results it does not rule out the possibility that a fraction of the complex fell apart and that the protein's contribution was slightly over-subtracted. In this case, over-subtraction would lead to an overestimation of lipids bound to the complex and underestimation of the amount of amphipol bound. The fact that we calculated only 3 molecules of amphipols were bound to this presumably 255 kDa complex (while published results suggest that 6 amphipols bind a 25 kDa protein, Gohon, Dahmane, *et al.*, 2008) further supported our suspicion of over-subtraction as it seemed unlikely HTL requires ten times less amphipol to be solubilized than the bacteriorhodopsin. Consequently, the amount of lipid would be over-estimated. The implications of over-subtraction are further discussed in section 2.5.2.

Radius of gyration: Since the atomic model available on the PDB (5MG3) is missing sections of the protein, the CRYSON program could not be used, hence the value of 44 Å that has been calculated from the EM envelope was used (Botte *et al.*, 2016).

First, it was observed that the values measured were in the same range as the theoretical value but also that they were rather similar to what was calculated for SecYEG. The Rm (radius of gyration at infinite contrast) was evaluated to be 39.0 Å, value very similar to that measured for SecYEG.

The decomposition of the radii of gyration by the parallel axis analysis showed a radius of gyration of 32.3 Å for the "non protein" component and 44.4 Å for HTL, the latter being in good agreement with the published values. Together with the positive $\alpha_{Stuhrmann}$ data, the protein seem to be mostly on the outside of the non-protein phase. Surprisingly, the radii of gyration of HTL and that of the non-protein contributions were both smaller than what was calculated for SecYEG. If this observation seemed counter intuitive, given that the molecular weight of HTL is significantly higher than that of dimeric SecYEG, it could be explained by the organisation of the complexes. The cryo-EM structure (Botte *et al.*, 2016) indicated that the additional subunits are closely associated to the SecYEG channel, forming a globular shape. In the case of SecYEG, the two channels are simply interacting and form a bilobal shape. This observation was supported by analysing the distance distribution function ($P(R)$, at 0% D_2O) that showed the maximum distance within the particle was $D_{max} = 108.2$ Å for SecYEG and $D_{max} = 84.5$ Å for HTL.

Holotranslocon solubilized with DDM (one step affinity purification)

The Holotranslocon prepared as described in section 2.1.3 and stabilized with DDM was measured. The scattering length density and radii of gyration were analysed.

Molecular Weight: The molecular weight of the particle was calculated using the forward scattering value measured in 24,4% D_2O. In the same way as described previously, the theoretical contrast match point value of HTL and the value that was the closest to DDM contrast match point (21,7%) were used to extrapolate the forward scattering value at 0% D_2O. This extrapolation would correspond to the value we would have measured if no surfactant was attached to the particles. A value of 487.1 kDa was calculated, which is smaller than an oligomer but significantly larger than what would be expected from a complex that should weigh 255 kDa (see Table 2.5).

This observation may again be attributed to the fact that data was collected at 24,4% D_2O, solvent fraction that was the closest to the match point of the particles (26,2%), therefore having large observational error. The contrast match point of the surfactant (21,7%) is rather far from the data point used (24,4%), therefore adding uncertainty to the extrapolated values. As previously mentioned, and further investigated later on, the complex contained lipids which will tend to increase the calculated molecular weight beyond that expected. The observed high molecular weight calculated may therefore be a consequence of the uncertainties linked to the SANS parameters.

Contrast match point: The scattering length density of the sample was analysed and the contrast match point (CMP) calculated. Again, the measured value

(26,2±1,8%) was far from the expected theoretical CMP for protein alone, an observation attributed to the presence of surfactants (here DDM). The contribution of protein was subtracted from the sample and the contrast variation curve of the other components extrapolated (see Equation 5.29). As for the previous HTL sample, the calculated CMP of the "non-protein" fraction (18,7%) was significantly lower than that of DDM (21,7%). This was attributed to the presence of lipids (Botte *et al.*, 2016). Considering that the contrast match point of *E. coli* inner membrane lipids was evaluated at 14,2%, it was calculated that the particles analysed were composed of 37,5% protein, 35,7% DDM (≈ 414 molecules) and 26,7% lipids (≈ 182 molecules).

The relative protein:DDM:lipid ratio was therefore of 1:1,1:0,6. As compared to the published ratio that correspond to a similar sample, this sample contained less lipids but more detergents per protein. As compared to the HTL sample solubilized in amphipols, a similar protein:lipid ratio was calculated but consequent differences were noted for the amount of surfactant bound. This further confirms that less amphipol bound to membrane protein as compared to the situation for DDM. Similar protein:lipid ratio as that of amphipol solubilized HTL but also of a cross-linked HTL solubilized with DDM were calculated (Botte *et al.*, 2016).

In any case, the presence of lipids within HTL solubilized with DDM was uniquely supported by the scattering length density analyses.

Radius of gyration: It was observed that the radii of gyration measured were significantly larger than for amphipol solubilized HTL. The values were closer to that of the uncross-linked HTL solubilized in DDM published (Botte *et al.*, 2016). A Rm of 51.2 Å was calculated, significantly larger than amphipol solubilized HTL (39.0 Å). The decomposition of the Rg into its protein and non-protein components by the parallel axis analysis gave values of 68.9 Å and 41.3 Å respectively (published values : 75 Å and 49 Å, Botte *et al.*, 2016). In agreement with published cryo-EM data, it was observed that the HTL solubilized in DDM appears significantly larger than the amphipol solubilized complex. However, this observation could not be solely attributed to the contribution of the DDM corona since it was calculated that the Rg of the protein was in itself higher. Furthermore, it was noted that the Rg of the non-protein component was smaller than that of the protein, an observation that is puzzling since it is known that the detergent corona is in the periphery of the protein and should therefore appear larger. This is firstly due to the fact that the corona's shape would corresponds to a ring; therefore its radius of gyration will be dominated by interference of scattering events occurring between neighbouring molecules rather than interference arising from across the ring (a ring will have a smaller Rg than a globular shape of the same dimension). Secondly, and as a consequence of the previous observation, the presence of lipid density of smaller dimensions than the protein contributes to lowering the Rg of the "non protein" component as its globular shape dominates the signal of the large DDM ring shaped corona.

Considering that the slope of the linear fit in the Stuhrmann analysis is positive and that the complex was composed of a large protein (Rg=68.9 Å) and a smaller "lipid/DDM"

density, it was concluded, in agreement with published data (Botte *et al.*, 2016), that the protein surround a density of lipids.

2.4.2 Deuterated HTL

The data were compared to that collected for deuterated Holotranslocon solubilized in amphipols A8-35 or Dodecylmaltoside. A summary of the data and Guinier fit results can be found in Table 2.6, the guinier fits represented in Appendix C and the contrast match point plots are represented on Figure 2.19.

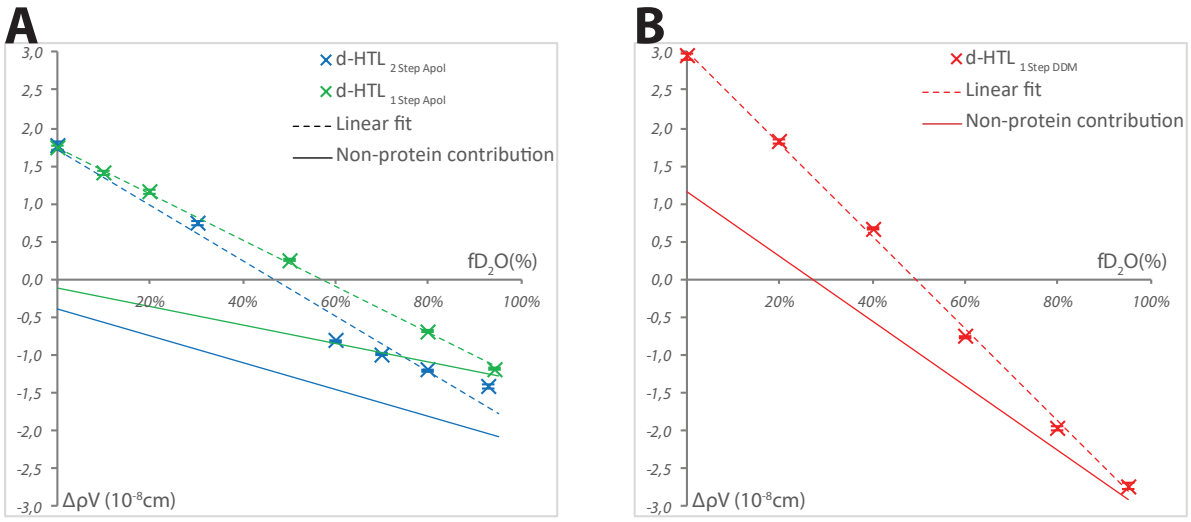


FIGURE 2.19: Small-angle neutron scattering results of 75% deuterated Holotranslocon. Excess scattering length ($\Delta\rho V$) plotted against the fraction of D_2O allows to calculate the contrast match point of the complex. **A:** HTL solubilized in amphipols and prepared as described in the two steps (blue) or one step (green) affinity protocols. **B:** d-HTL sample solubilized with DDM and purified with the one step affinity protocol (red). The dashed lines represent linear fits and the solid lines represent the contribution of surfactants and lipids after subtraction of the protein's contribution.

For the deuterated HTL purified with two affinity steps, the calculated molecular weight was surprisingly low (59.3 kDa) as compared to the expected value (266.5 kDa). The measured radii of gyration were closer to that expected for SecYEG monomer (25.7 kDa), suggesting that the sample was dissociated. The subtraction of HTL's contribution to the measured scattering length density should yield a scattering length density (and therefore contrast match point) compatible with a mixture of lipids and surfactants. In this case, the excess scattering length after subtraction was negative at all contrasts, which is strictly incompatible with the known presence of surfactants in the sample.

The same observation could be made in the case of deuterated HTL purified with one affinity step. The analysis of the scattering length density data indicate that the subtraction of HTL's contribution leads to a large over-subtraction and that the excess scattering length was not compatible with the presence of surfactants. The calculated molecular weight (202.1 kDa) was smaller than that expected (266.5 kDa).

TABLE 2.6: Summary of SANS data for deuterated HTL solubilized with amphipols or DDM

Guinier Analysis

Sample	Conc.(mg ml^{-1})	%D_2O	I_0	$\sigma(I_0)$	Rg	$\sigma(Rg)$
d-$HTL_{2StepApol}$	2,60	0,0%	1,960	0,031	50,7	0,6
	2,38	30,0%	0,310	0,009	35,5	0,6
	1,68	60,0%	0,270	0,002	15,5	0,2
	1,51	70,0%	0,360	0,003	15,7	0,2
	1,55	80,0%	0,540	0,003	14,3	0,2
	1,85	93,0%	0,900	0,007	16,6	0,2
d-$HTL_{1StepApol}$	0,80	0,0%	0,580	0,003	34,7	0,2
	0,82	10,0%	0,390	0,003	35,4	0,3
	0,84	20,0%	0,270	0,002	34,1	0,2
	0,89	50,0%	0,014	0,001	15,6	1,0
	0,95	80,0%	0,110	0,001	37,5	0,3
	0,98	94,3%	0,330	0,001	37,1	0,1
d-$HTL_{1StepDDM}$	0,30	0,0%	0,630	0,004	43,0	0,3
	0,30	20,0%	0,240	0,001	42,1	0,2
	0,30	40,0%	0,033	0,001	34,6	0,9
	0,30	60,0%	0,042	0,001	45,0	0,6
	0,30	80,0%	0,280	0,001	46,8	0,2
	0,30	95,0%	0,540	0,003	46,2	0,2

Theoretical values

Complex	*Stoichiometry*	$CMP_{protein}$(%)	$MW(kDa)$	Rg(Å)
$HTL_{(75\%d)}$	Monomer	100,1	266,5	44,0 (Botte *et al.*, 2016)

Analysis of the forward scattering

Complex	CMP(%)	$CMP_{nonprot}$(%)	$\%_{protein}$	$\%_{surfactant}$	$\%_{lipid}$
d-$HTL_{2StepApol}$	42,4 ± 2,3	∅	∅	∅	∅
d-$HTL_{1StepApol}$	57,1 ±1,0	∅	∅	∅	∅
d-$HTL_{1StepDDM}$	49,9 ±1,1	27,0	31,4	62,4	6,2

Analysis of the radii of gyration

Complex	$\alpha_{Stuhrmann}$	Rm(Å)	$Rg_{nonprot}$(Å)	Rg_{HTL}(Å)
d-$HTL_{2StepApol}$	>0	25,5	22,4	27,8
d-$HTL_{1StepApol}$	<0	35,3	37,3	33,6
d-$HTL_{1StepDDM}$	>0	44,4	46,0	40,1

Furthermore, the radius of gyration of the protein (33.6 Å) was too small as compared to that expected from HTL (44.0 Å).

These observations were further complemented by the biochemistry data (see section 2.1.1) that showed the production of deuterated HTL in high cell density cultures generated important imbalances of the subunits. The analysis therefore focused on a sample that was expressed in high cell density conditions but with a short induction time to limit imbalanced expression (see section 2.1.1).

Deuterated HTL in DDM

The sample was purified using one affinity step (section 5.2.1) and solubilized with DDM. The scattering length densities and radii of gyration have been analysed and given in Figure 2.20.

Molecular weight: In a similar manner as described previously, the values of I_0 measured at 20% (closest to the match point of DDM, 21,7%) were used. By using the theoretical contrast match point of HTL deuterated at 75% (CMP=100,1%), the I_0 value at 0% D_2O were extrapolated to provide an approximation the value in the absence of detergents. By doing so, a molecular weight of 312.8 kDa was calculated. As noted before, this value may be over-evaluated as a result of the presence of deuterated lipids that also contribue to the total scattering at 20% D_2O. The fact that the difference between measured and theoretical value was smaller than that with hydrogenated HTL may arise from the fact that the surfactant, here DDM (CMP=21,7%), was almost perfectly matched out as opposed to amphipol which has a higher contrast match point (CMP=23,5%). It is then noteworthy to observe that the molecular weight of DDM-solubilized SecYEG was perfectly evaluated using this method.

Contrast match points: The measured contrast match point (49,9%) was significantly different from the theoretical value (100,1%) (see Table 2.6). As noted previously, this difference can be explained by the presence of detergents (CMP=21,7%) and deuterated lipids (CMP=80,2%, calculated from section 4.1.2) that reduced the match point of the particles. The scattering length density of the protein complex alone were calculated and subtracted from the data. The excess scattering length density yielded a contrast match point of 27,0%, a value that could not reflect deuterated lipids or hydrogenated detergents alone. It was therefore calculated that the particle is composed of 31,4% protein, 62,4% DDM (≈ 860 molecules) and 6,2% lipids (≈ 52 molecules), which corresponds to a protein:DDM:lipid ratio of 1:2,0:0,2. Similar protein:DDM:lipid ratios as for hydrogenated SecYEG (1:2,3:0,3) were obtained but significantly less lipids and more detergents than for the equivalent hydrogenated HTL (1:1,1:0,6).

Again, it has to be recognised that the Holotranslocon is not intact or the solution not homogeneous. Therefore, the subtraction of the contribution of a fully deuterated HTL may lead to errors in the evaluation of lipid and DDM content. In this case, because the lipids are deuterated, over-subtracting would lead to an over-estimation of DDM volume and underestimation of lipids. Since the calculated ratios are significantly

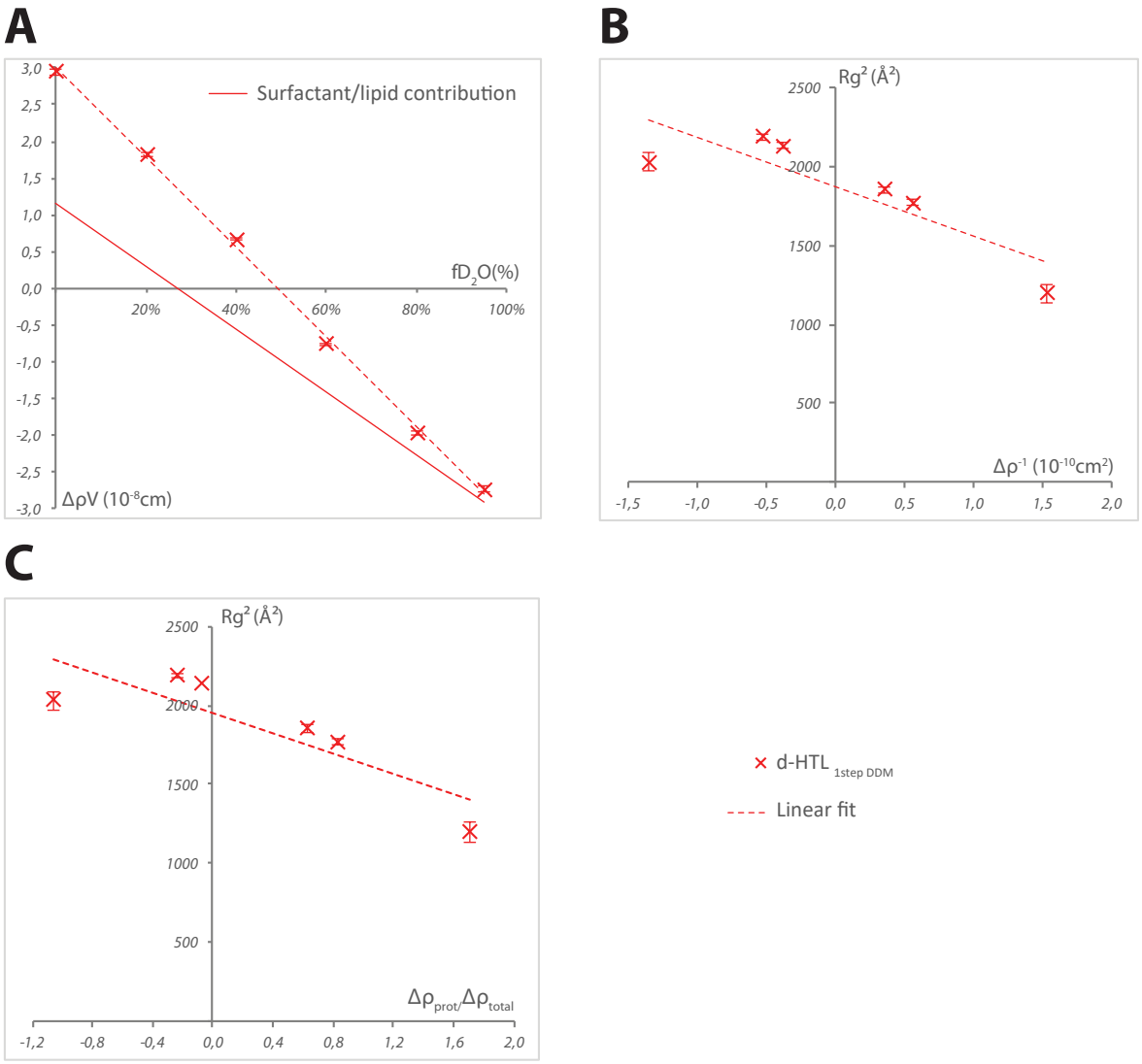


FIGURE 2.20: Small-angle neutron scattering results of 75% deuterated Holotranslocon in DDM. The d-HTL samples solubilized with amphipols are not represented. The dashed line represent linear fits. **A :** Excess scattering length ($\Delta\rho V$) plotted against the fraction of D_2O allows to calculate the contrast match point of the complex. The solid line represent the contribution of surfactants and lipids after subtraction of the protein's contribution. **B :** Stuhrmann analysis of the radii of gyrations plots the squared Rg against the inverse of the contrast $\Delta\rho V$ and gives indication on the relative organisation of the scattering density phases. **C :** Parallel axis analysis plots the squared Rg against the ratio of the hydrogenated phase contrast and calculated contrast in order to give information on the radii of gyration of each phase.

TABLE 2.7: Summary of SANS analysis

Sample	prot.:surf.:lip.	surf. bound	lip. bound	Rg_{prot}(Å)
h-$HTL_{uncrosslinked}$	1:0,8:1,2		330-390	75
h-$HTL_{crosslinked}$	1:0,8:0,6		160-200	52
h-$SecYEG_{DDM}$	1:2,3:0,3	607_{DDM}	52	48
h-$HTL_{1StepApol}$	1:0,1:0;7	3_{Apol}	182	44
h-$HTL_{1StepDDM}$	1:1,1:0,6	414_{DDM}	182	69
d-$HTL_{1StepDDM}$	1:2,0:0,2	860_{DDM}	52	40

different than for hydrogenated HTL, with an over-representation of DDM and low quantities of lipids, it was suspected that the initial solution was not homogeneous. The effect and implication of over-subtraction are further discussed in section 2.5.2.

Radius of gyration: The measured radii of gyration were in the same range as previously observed for SecYEG and hydrogenated HTL (see Table 2.6). The Stuhrmann analysis gave a radius of gyration at infinite contrast (Rm) of 44.4 Å, which was slightly higher that for hydrogenated HTL. Most importantly, in this case we observed that the radius of gyration of the protein (40.1 Å) was lower than that found previously (44.4 Å). This observation may be correlated with the fact that relatively little lipid was detected, an effect that has been attributed to a "closed form" of the HTL (Botte *et al.*, 2016). In agreement with the fact that the $\alpha_{Stuhrmann}$ is negative, the radius of gyration of the non-protein part appears bigger than that of the protein, indicating that the "non-protein" component was mostly arranged around the protein. In this case, the "non-protein" value would be dominated by detergents and not by lipid since their scattering length density was very different from than that of the protein (which is not the case for deuterated lipids) but also because the amount of lipid was significantly lower than the amount of detergents. These data were in agreement with expectations and indicated that the deuterated components (proteins and lipids) were lying on the inside with the hydrogenated component (DDM) surrounding them.

Concluding comments on the SANS results

As it was expected given the biochemical and biophysical data that were collected, some datasets indicated that the sample was dissociated or inhomogeneous. Data collected for the Holotranslocon were compared, either in hydrogenated form or deuterated form solubilized either with amphipols or DDM, with data collected for the hydrogenated SecYEG and published results (Botte, 2013).

The analysis of the scattering length density and contrast match points of hydrogenated complexes allowed to confirm that the scattering particles were composed of protein, but also a significant amount of lipid and surfactant.

In the case of HTL solubilized with amphipols, it was calculated that these was similar amount of bound lipids as compared to the published results for DDM-solubilized

HTL (Botte *et al.*, 2016) but the number of amphipol bound to the complex was unaccountably low by comparison with published results (Gohon, Dahmane, *et al.*, 2008). It has been proposed that, when solubilized with amphipol molecules, HTL has similar dimensions as a DDM-crosslinked-HTL that is itself significantly smaller than "native" DDM-HTL. Amphipols would therefore favour a more compact form of HTL, similar to the cross-linked HTL that has been associated to a lower quantity of bound lipids. The compact state corresponds to a Holotranslocon in which the accessory subunits (SecDF-YidC-YajC) are closely associated to SecYEG, thus reducing the size of the lipid cavity (Botte *et al.*, 2016). The analysis of the radii of gyration would be in agreement with this observation, and support the observation that the lipids would mainly be located at the interior of a Holotranslocon. These results are to be considered with care since the low scattering length observed for this sample as compared to DDM solubilized samples, together with the low Rg and low amount of bound amphipol would tend to suggest the sample was inhomogeneous or dissociated. The relative ratio of components calculated would therefore be inaccurate and could be a consequence of over subtraction (see section 2.5.2).

In the case of hydrogenated HTL solubilized with DDM, the data suggested that less lipids were retained as compared with published results (Botte *et al.*, 2016). The radius of gyration measured for the protein was also smaller than published results; this calculation however relied on data that are usually highly prone to error and the variations that were observed (about 10%) may be a consequence of high measured error on data points collected close to the contrast match point of the complex and low number of measured contrasts (Botte, 2013). The Stuhrmann and parallel axis analyses clearly indicated that the protein surrounds the low scattering phase. The presence of lipids at a central position of the structure, though in more limited quantity, was therefore strongly supported by these results.

Finally, the analysis of deuterated Holotranslocon showed only a modest amount of lipid was detected, which appeared in similar proportions as observed for SecYEG. This observation may however be a consequence of sample heterogeneity. In this case, and because of contrast considerations, the lipids could not be localized as the hydrogenated detergents dominated the signal. We therefore observed that the DDM was located in periphery of the protein. The dimensions of the protein (40.1 Å) was not in agreement with previously published results for a DDM solubilized complex (44.0 Å) (Botte *et al.*, 2016). These data, together with biophysical results, seem to indicate the sample was not homogeneous or possibly dissociated and that the measured values would correspond to a smaller assembly than that expected of HTL.

The limitations of SANS for this approach are discussed in subsection 2.5.2.

2.5 Discussion

2.5.1 The production of HTL leads to unstable and polydisperse sample

Imbalanced expression of the subunits

Before the development of the ACEMBL strategy (Bieniossek *et al.*, 2009), the Holotranslocon could not be efficiently over-expressed and purified. This was partly due to the difficulties faced to generate a plasmid that allows the balanced expression of all seven subunits. The control of the copy-number of a protein is a highly multi-factorial process that involves regulations at different levels:

- **Pre-transcriptional**: some plasmids have lower copy numbers than other depending on the negative feedback mechanism used (Del Solar and Espinosa, 2000). In addition to this, two plasmids that are regulated in similar ways might replicate unevenly by chance. Therefore, co-transformation is bound to generate imbalanced copy numbers in between plasmids.
- **Transcriptional**: One of the main factor that control the level of expression of a protein is the strength of its promoter (J. Li and Y. Zhang, 2014), which will consequently lead to a higher number of mRNA and hence protein. Different promoters may therefore lead to imbalanced expression.
- **Post-transcriptional**: mRNA levels can also be regulated post-transcriptionally by small RNAs (sRNA). RNA binding proteins can also modulate the stability of a mRNA, the accessibility of the Ribosome Binding Site (RBS) or the transcription termination (Van Assche *et al.*, 2015).
- **Translational**: the translation efficiency of a given protein is a highly multi-factorial process. The strength of the Shine-Dalgarno sequence and the structure of the mRNA are two important factors, that however do not suffice to predict the efficiency of translation (G.-W. Li *et al.*, 2014). When considering an operon, the sequences that are transcribed first are exposed to the ribosome earlier and therefore translated to higher degrees (Lim *et al.*, 2011). Since all codons are not equally represented in bacteria (Ermolaeva, 2001), the sequence of a given protein will directly affect the translation efficiency as certain tRNA might have a limited availability, therefore slowing down the translation.
- **Post-translational**: numerous examples (Batchelor and Goulian, 2003; Barkai and Shilo, 2007) report that the level of proteins that can be regulated negatively by bacteria through degradation. This process is protein-dependent and is not fully understood.

If the ACEMBL technology successfully tackled the copy-number issues associated with multiple transformations, it cannot compensate for the other factors mentioned. In particular, the presence of both arabinose and Trc promoters on the HTL plasmid generated by Bieniossek *et al.*, 2009 could partially explain the imbalanced subunits expression that was important in the case of high cell density culture (see section 2.1.1).

Therefore, despite a significant improvement by using the ACEMBL generated plasmid, the balanced expression cannot be insured and might affect the expression efficiency, assembly of the subunits into the full Holotranslocon (and not sub-complexes such as SecYEG) as well as the final stoichiometry of the subunits within the HTL. Alternative promoter combinations have been attempted but the expression/purification efficiencies did not compare to those obtained with the plasmid used : HTL3 (Botte, 2013).

It is also interesting to note that on this plasmid, SecY, SecE and SecG are on the same operon, which is not the case *in vivo* (Mulder *et al.*, 2013). Similarly, YidC is not on the same operon as SecD and SecF *in vivo*, but YajC is (K. Pogliano and Beckwith, 1994). This genetic organization is therefore strongly modified during the expression process, which may help to explain the difficulties faced for the purification of intact complexes.

Stability, integrity and stoichiometry of the complex

The achievement of a reliable expression of all seven subunits at equimolar ratio would be the first step to favour the production of the Holotranslocon and allow better biophysical measurements. The next challenge lies in the development of a successful and reliable purification strategy for the production of a sample that is compatible with structural studies such as small-angle scattering or electron microscopy.

Solubilization and stabilization of the complex As discussed in subsection 2.1.2, the choice of the solubilization method is of key importance. In this study, we used a common detergent for the purification of membrane protein : the DoDecyl Maltopyranoside. This detergents is considered as "mild" because of its non ionic nature and because of its low critical micellar concentration. However, present results (see subsection 2.2.1) have suggested the Holotranslocon was not intact after purification and that particles smaller than expected could be detected. These results could indicate that either DDM failed to extract intact HTL from the membrane, or that HTL is destabilized and dissembled during the purification procedure. These considerations were supported by previous work (Botte, 2013; Botte *et al.*, 2016) where chemical cross-linking was required in order to keep the Holotranslocon intact for electron microscopy studies.

It is firstly important to note that the presence of detergents often interferes with the biophysical methods. In the case of methods that are based on the measurement of the hydrodynamic radius (such as size exclusion chromatography, analytical ultracentrifugation, dynamic light scattering), it is assumed that the detergents, that were inevitably bound to the transmembrane domains of the proteins, would make the final particle appear bigger. By using the three detector method with SEC-MALS or interference measurements with AUC, it is possible to evaluate the quantities of detergent bound and therefore make these methods better suited for the characterisation of the size of the particles. Static and dynamic light scattering, which are solely based on scattering intensity, were strongly affected by the presence of detergent assemblies. They

therefore provided results that were largely dominated by micelles since the concentration of detergent was one to three orders of magnitude higher than that of the protein. Finally, detergent may have interfered with the ionization process (matrix assisted or electro-spray) that is a key step for the success of mass spectrometry analyses.

To date, most of the data published on HTL (Schulze *et al.*, 2014; Komar, Alvira, *et al.*, 2016) are mainly based on biochemical assays. The existence of the Holotranslocon, both *in vivo* as shown by Komar, Alvira, *et al.*, 2016 with the immunoprecipitation of all subunits of HTL solubilized with Styrene Maleic Acid, and *in vitro* as shown by Botte, 2013; Botte *et al.*, 2016 cannot be denied. However, no data, present or published, clearly supports a durable interaction between HTL's subunits in the absence of a chemical cross-link; therefore it has not been unequivocally proved that the Holotranslocon can be kept soluble and intact in DDM without chemical cross-linking. Schulze *et al.*, 2014 and Komar, Alvira, *et al.*, 2016 report that proteoliposomes containing HTL are, at least in some cases, more efficient at promoting membrane protein integration than SecYEG or YidC, or that the binding of ribosome nascent chain and SecA is more efficient on HTL. It is undeniable that proteoliposomes containing all seven subunits composing HTL display different biological activity than SecYEG or YidC; it is however not clear whether these effects can be fully imputed to HTL or to a combined effect of SecYEG, SecDF and YidC that do not interact durably in the membrane.

To the best of our knowledge, this detergent has been used for the Holotranslocon similarly to the way it has been used for the purification of SecYEG. However, Botte, 2013 reported that the purification of the sub-complex containing SecDF-YidC-YajC with DDM led to significantly lower yields as compared to SecYEG and even HTL. This suggested that the difficulties faced during the purification of the Holotranslocon might be a consequence of the presence of SecDF-YidC-YajC.

Taken together, these information seem to indicate that DDM is a poor detergent for the purification and stabilization of this large transmembrane complex. Our preliminary work on detergent screening (see section 2.1.2) opens possibilities for finding an efficient substitute of DDM for the purification of the Holotranslocon as a few detergent (T3DM, LTM, UM...) seem to give similar, if not better, results than DDM in terms of solubilization efficiency. However, using a new detergent would require a thorough optimization process that would require the use of biophysical methods to reliably ensure, not only that the solubilization is efficient, but also that the detergent does not disrupt the quaternary structure of the complex and can keep it soluble for extended periods of time.

One of the limitations for the optimization of membrane protein expression, solubilization and purification is that it is difficult to unequivocally identify the particle of interest. This usually involves methods (such as SEC-MALS, AUC, DLS, SAS, etc.) that are labour intensive and not adapted for high throughput screening, which ideally involves automation. Bird *et al.*, 2016 and Moraes, 2016 developed a platform dedicated to the optimization of expression and solubilization for membrane protein by using cleavable fluorescent reporters fusion. This allows for easier monitoring of expression levels, detergent extraction efficiency and solubility of the protein of interested.

The streamlined application of circular dichroism, size exclusion chromatography, thermofluor shift assay and SEC-MALS allows reliable evaluation of the size, stability and secondary structure of the membrane protein, thus allowing solubilization and purification strategies to be confidently identified. This strategy could be applied to the Holotranslocon as a way to screen more detergents and purification methods while also reliably assessing its integrity.

Another significant problem inherent in the use of detergent is that the composition of the buffer and especially its concentration in detergents must be carefully assessed and controlled. The purification process, regardless of the method used, usually implies several steps which are accompanied by a reduction of the protein's concentration. In the case of structural studies, the concentration of the protein of interest is a primordial factor, making it mandatory to concentrate the sample, usually by ultra-filtration. Even if the molecular weight cut-off used is higher than that of the DDM micelle, it has appeared clearly (see Figure 2.9) that detergent was also concentrated in the process. Therefore, by trying to reach concentration compatible with structural studies (typically around $1\,\mathrm{mg\,ml^{-1}}$), the concentration of the detergent will have been increased, which may lead to an increased instability of the complex.

In addition to detergent induced instability, it was demonstrated that using D_2O as a solvent may lead to increased polydispersity of the sample (see Figure 2.13). In the case of partially deuterated HTL, a reduced size of particles together with a higher polydispersity was also observed, which suggested dissociation.

A solution that was further investigated for overcoming detergent related issues is the substitution of DDM for amphipols at later stages of the purification. In a similar way as described previously (Botte *et al.*, 2016), the optimal amphipol to protein ratio was investigated by size exclusion chromatography and analysed the sample biophysically. While some results are promising in terms of measured size of the complex (see Figure 2.13), the homogeneity of the sample is still insufficient as compared to usual standards for structural biology; other experiments (see Figure 2.14) suggested that the complex solubilized in amphipols was not stable.

Taken together this information shows that the current strategies that have been used over the last few years are not capable to reliably deliver a Holotranslocon that is homogeneous, stable and concentrated enough for structural biology studies. The detergent is probably one of the main factor responsible for this observed instability and heterogeneity.

It is also important to note that both SecYEG and YidC can perform translocation by themselves *in vivo* and *in vitro* (Tsirigotaki *et al.*, 2017; Rapoport, 2007; F. Jiang *et al.*, 2003; Kedrov *et al.*, 2013; Seitl *et al.*, 2014) and that proteomic analyses reported a copy-number diversity with ratio of 4 SecYEG : 3 YidC : 1 SecDF*in vivo* (G.-W. Li *et al.*, 2014). Therefore, HTL would represent at the best 25% of the total SecYEG population and a third of the YidC population. It is therefore reasonable to think that the translocation machinery is likely to be highly dynamic and rearrange into either SecYEG dimer, YidC monomer or full HTL depending on the substrate requirements (secretion, membrane integration). This dynamic view of the translocation machinery

could be one of the reason explaining why the Holotranslocon has remained elusive as it would have a natural tendency to disassemble into its sub-complexes.

Purification strategy Imbalanced subunit expression ratios and detergent mediated heterogeneity of the complex are issues that may be partly compensated by appropriate purification processes. In this work, two different published strategies were compared.

The first strategy (using one affinity purification step, see section 5.2.1) only makes use of the three histidine tags that are present on YidC, SecD and SecE. It is therefore undeniable that the nickel affinity step will not be able to separate integral complexes that contains all seven subunits from sub-complexes (DFYY and SecYEG) or even free YidC as they could all contain at least one histidine tag. In order to refine this purification, the size exclusion chromatography is employed. However, and as discussed previously, the presence of detergents strongly affects the efficiency and reliability of size exclusion chromatography because the size of the detergent belt cannot be anticipated and therefore unpredictably affects the hydrodynamic radius of the protein. This was particularly well illustrated in Figure 2.7 where SecYEG (150 kDa) appeared bigger than HTL (255 kDa). Further experimentation (SEC-MALS, AUC) did not allow us to prove that this protocol could yield an homogeneous and integer complex. Following the protocol published by Schulze *et al.*, 2014, we also tried to combine the size exclusion chromatography to an in-line anion exchange so that the elution is stripped off potential negatively charged contaminants. In theory, this resin should not be able to help discriminate between complete and dissociated HTL (SecYEG, DFYY and HTL all carry positive charges at pH 8) except for YidC and SecG alone which should be retained. However, the main issue was that connecting an additional affinity column directly after the gel filtration column dramatically reduces the resolution of the size exclusion chromatography. Indeed, the fractions exiting the gel filtration column will inevitably be diluted and mixed together, therefore defeating the purpose of size exclusion chromatography. As expected, attempts yielded very wide peaks with low intensity and the proteins retained by the resin seemed to correspond to non-specific binding. In conclusion, it could not be shown that this protocol could yield stable and homogeneous sample suitable for structural biology, although the yield is satisfactory.

The second strategy, two affinity purification steps (see section 5.2.1), makes use of both histidine tags (on SecE, SecD and YidC) and the Calmodulin binding peptide tag (on YajC). This strategy therefore combines a tag that has a high recovery efficiency with a tag that is present on a single subunit. To the best of our knowledge, the first undeniable proof of existence of the SecYEG-SecDF-YidC-YajC complex (by Botte, 2013; Botte *et al.*, 2016) relied on the co-purification of all seven subunits using the CBP tag. However, structural studies were not possible as the Holotranslocon displayed high heterogeneity and was shown to dissociate (Botte, 2013). Analytical ultra-centrifugation showed that HTL required chemical cross-linking in order to be stable enough for cryo-EM, though the resolution did not exceed 15 Å. However, it is evident that cross-linking significantly alters the native structure and the function of a protein and as such cannot be used for biochemical assays. In order to solve this issue, Botte *et al.*, 2016 implemented the use of amphipols A8-35 as a substitute for cross-linking since amphipols have been associated with higher stability (Popot, 2010).

However, the resolution could not be improved (Botte *et al.*, 2016) and the HTL seemed to lose its ability to bind to SecA. It could not be shown, either by DLS, SEC-MALS or AUC, that a stable and homogeneous complex was produced. Furthermore, the yield achieved by this protocol were drastically lower (more than 30 L of culture required per SANS experiment). This can probably be attributed to the initial heterogeneity of the sample which implies that the fraction of HTL actually present in solution is lower than what can be expected. We experienced up to 90% of sample loss when using this approach. Furthermore, any additional purification step is bound to come with a reduction of yield as no purification procedure can be achieve with a 100% efficiency. It is also important to note that many steps of purification gives rise to longer exposure to detergents and may therefore lead to delipidation of the detergent solubilized protein, which may in turn lead to reduced stability.

These two purification procedures come with their own advantages and disadvantages. The protocol involving only one step of affinity purification (section 5.2.1) allows a maximized yield as the number of steps is kept to a minimum. However this protocol, by its nature, will be highly susceptible to subunit imbalanced expression, dissociation of the complex and heterogeneity. On the other hand, the protocol involving an additional affinity purification step (section 5.2.1) is highly selective and allows intact HTL to be extracted from a polydisperse solution. However the yield observed is drastically reduced, an observation that can be partially related to initial sample polydispersity but also relatively low binding capacities of the Calmodulin resin in these conditions.

SecA - HTL interaction

To investigate the potential binding of SecA to the Holotranslocon, size exclusion chromatography was used. As discussed previously and because of the presence of detergents in the solution, the resolution of the size exclusion chromatography together with the polydispersity of HTL solution is insufficient to unequivocally separate SecA (expected molecular weight of 95 kDa and HTL or sub-complexes between 75 to 255 kDa). If all of the elution peaks overlapped, it would however be possible to identify that adding SecA to HTL induced a shift toward lower elution volumes, indicating the existence of a bigger particle and therefore of binding. However, the careful analysis of the composition of the peaks by SDS-PAGE seemed to indicate that only the subunits of SecYEG shifted toward lower elution volumes when exposed to SecA. On the other side, for SecD, SecF and YidC, the addition of SecA seemed to reduce their elution volume only if combined with ATP and toward fractions that completely lacked SecYEG. We concluded that the addition of SecA to HTL solution mostly affected SecYEG and less significantly SecDF-Yid and that SecA did not bind to a full Holotranslocon as the shift of SecYEG and SecDF-YidC happen independently and in different conditions. This observation was further strengthened by the analysis of SecA/HTL equimolar mixture by SANS: the bimodal character of the scattering length data suggested that HTL and SecA remained mostly unbound (data not shown), observation supported by the radii of gyration that did not suggest binding of SecA (Rm similar to HTL alone.)

Schulze *et al.*, 2014 report an interaction between SecA and HTL. However, these interaction assays are based on SecA mediated ATP hydrolysis and SecA fluorescence

quenching. This increased binding could arise from the combined effect of a binding to SecYEG and SecDF-YidC independently since the integrity of HTL during the interaction is not supported by any independent data. Furthermore, while Komar, Alvira, *et al.*, 2016 seem to indicate that HTL is more efficient for the translocation of all the substrates tested, in all cases, the reunion of SecYEG-only and YidC-only translocation efficiencies would give similar efficiency than for HTL. It is also important to note that a control consisting of proteoliposomes that have been reconstituted with both SecYEG and YidC would have allowed the effect of intact HTL and combined effect of the subunits to be distinguished. If the existence of a Holotranslocon super-complex cannot be refuted (co-purification of all subunits using the CBP tag by Botte, 2013; Botte *et al.*, 2016 and styrene maleic acid extraction from native membrane by Komar, Alvira, *et al.*, 2016), the present data together with published biochemical assay cannot unequivocally prove that the Holotranslocon remains intact, stable and homogeneous, either in proteoliposomes or in surfactants.

Concluding remarks on the homogeneity of the sample

The biophysical data seemed to indicate difficulties in the purification of the Holotranslocon as well as maintaining required levels of concentration, homogeneity and stability. The work of Botte, 2013 has concluded that HTL could not be analysed by electron cryo-microscopy, unless the complex was stabilized by a chemical cross-link. The analysis also required *in silico* sorting (47% of the particles were rejected); the heterogeneity of the sample, but also flexibility of the protein, prevented them from achieving high resolution (Botte *et al.*, 2016). In the case of cryo-EM reconstruction of the non cross-linked amphipol solubilized HTL, 82% of the selected particles were rejected, further indication of the high polydispersity of the sample. These observations raise the question of whether the current protocols (as published by Schulze *et al.*, 2014; Botte *et al.*, 2016 and tested biophysically in this work) are optimized for the preparation and analysis of the Holotranslocon by structural methods.

The result in section 2.2.1 have brought forward the possibility that the stability of the complex was strongly dependent on its concentration. It is important to note that this effect could only clearly be observed in the case of deuterium labelled HTL solubilized in DDM and we estimated that the higher molecular weight particle (that is likely to correspond to an intact HTL) was observed at a concentration of higher than 1.1 mg ml^{-1}-4.4 µM, typically right after the IMAC elution. When the hydrogenated HTL was analysed by SEC-MALS (see Figure 2.9), a particle that eluted earlier was detected after concentration of the sample to 1.4 mg ml^{-1}-5.6 µM but its molecular mass was evaluated to 50 kDa. In the context of AUC analyses, the complex was dissociated at a concentration of 1.2 mg ml^{-1}-4.8 µM (see Figure 2.10). When bound to SecA, HTL was at a concentration of 1.4 mg ml^{-1}-5.6 µM and dissociation was suggested (see Figure 2.17). In the case of SANS with DDM solubilized d-HTL, the concentration was of only 0.3 mg ml^{-1}-1.2 µM. In the case of amphipol solubilized HTL, DLS suggested the particle's molecular weight was close to what was expected for HTL at a concentration of 0.9 mg ml^{-1}-3.6 µM (see Figure 2.13). The amphipol solubilized HTL was suggested to be dissociated at 0.9 mg ml^{-1}-3.6 µM by AUC (see Figure 2.14) but intact at the

same concentration by SANS (see Figure 2.18). A summary of the concentration and method used to analyse each sample is shown in Table 2.8.

It could be hypothesized that the concentration indeed plays an important role in the integrity of HTL and that the K_d of the complex is close to the concentration achieved in these experiments. The dissociation was most often observed when using methods that involve dilution of the complex during the measurement (size exclusion chromatography-multi angle laser scattering, analytical ultra-centrifugation...). In addition to the fact that no clear trend could be identified, it is important to note that most of the time, the complex had to be concentrated after purification (even for SANS experiments) as the concentration was often bellow 0.5 mg ml^{-1}-2.0 µM at the latest stages. The K_d of the SecYEG-(SecDF-YidC-YajC) has not been established, but in light of these results, it could be expected to be in the micro molar range. It would therefore be likely that along the purification procedure, the complex may fall apart when its concentration is reduced (notably after gel filtration). Considering the stability issues and in agreement with previous observations (Botte, 2013), it seems unlikely that HTL would spontaneously re-associate upon concentration, a process that would involve extensive reorganization of the detergent coronas. Furthermore, is has been shown that the concentration of the sample by ultra-filtration was accompanied by an increase of the detergent concentration (see Figure 2.9), which may further destabilize the complex.

The quality of the sample is of prime importance for small-angle scattering studies since the technique averages the signal from all particles in solution. The heterogeneity of the particles analysed makes the interpretation of the data complicated (typically preventing the identification of the Guinier region), put the modelling out of reach and affect the accuracy of the analysis performed.

TABLE 2.8: Summarization of biophysical characterization of HTL

	SANS			DLS			AUC			SEC-MALS			GF	
Sample	[C]	Rg_{prot}	concl.	[C]	MW	concl.	[C]	MW	concl.	[C]	MW	concl.	[C]	concl.
$SecYEG_{DDM}$	1,5	48	dimer	-	-	-	-	-	-	-	-	-	2,5	>HTL
$h\text{-}HTL_{1StepApol}$	0,9	44	dissoc.	0,5	105	dissoc.	-	-	-	-	-	-	-	-
$h\text{-}HTL_{2StepApol}$	1,5	43	dissoc.	0,9	237	dissoc.	1,2	60	dissoc.	-	-	-	-	-
$h\text{-}HTL_{1StepDDM}$	0,7	69	Intact	0,5	6500	micelle	-	-	-	1,4	195	dissoc.	0,75	dissoc.
$h\text{-}HTL_{2StepDDM}$	-	-	-	-	-	-	1,2	77	dissoc.	-	-	-	-	-
$d\text{-}HTL_{1StepApol}$	0,9	34	dissoc.	-	-	-	-	-	-	-	-	-	-	-
$d\text{-}HTL_{2StepApol}$	2,0	28	dissoc.	0,8	171	dissoc.	-	-	-	-	-	-	-	-
$d\text{-}HTL_{1StepDDM}$	0,3	40	dissoc.	-	-	-	-	-	-	-	-	-	0.9	<SecYEG

[C] in $\mathrm{mg\,ml^{-1}}$*, Rg in* Å*, MW in* kDa
concl. : conclusion
dissoc. : dissociated

2.5.2 Small-angle scattering, approximations and importance of the sample

Approximation of the concentration

The concentration is measured by spectrophotometry as an absorbance and needs to be converted to a mass concentration using the extinction coefficient and molecular weight of the protein. These values are calculated from the sequence of the protein by assuming that the solution is composed only by these exact molecules or protein assembly, which is, as we have seen in section 2.2, not necessarily the case. In the case of SecYEG and HTL, the extinction coefficients were measured in the past by collaborators using amino acid analysis, which also relies on the assumption that the sample's composition is known and controlled. This particularly applies for the measurement of complexes concentration, as their composition needs to be assumed in order to access the concentration. This is especially true if complexes are subjected to further purification (see purification of SecB:MBP: SecA complexes subsection 3.2.2) where there is therefore no control on the composition of the final solution.

This has implications in the calculation of the contrast match point of a protein and the calculation of the molecular weight, which both rely on the concentration (see Equation 5.25 and Equation 5.26). By assuming the complexes are intact or of a precise composition, we inevitably bias the calculated values. This bias is even more pronounced for the calculation of the molecular weight as this value is extrapolated from a single data point (while the contrast match point is supported by several points), that is itself extrapolated from the Guinier approximation.

Furthermore, when proteins are deuterated, the molecular weight is not adjusted with regard to the level of deuterium integrated in the protein (which is itself approximated) therefore introducing another bias in the calculation of the concentration. This observation is also valid with deuterated solvent, where deuterium will exchange with the protein to an extent that cannot be fully anticipated.

Hydrogen/deuterium exchange

As described in section 1.2, the hydrogen involved in covalent bound with oxygen and nitrogen are labile and can exchange with the solvent. The extent of that exchange cannot be fully anticipated as it will depend on the exposure of the exchangeable hydrogen to the solvent. This will on one side depend on the structure of the protein (buried hydrogen are unlikely to be exchanged) and Jacrot and Zaccai, 1981 estimate about 80% of the hydrogen can be exchanged in proteins. This value is an estimation aimed at typical globular proteins but might not be well-suited for membrane protein which are, to an extent, even more susceptible to this uncertainty as a large proportion of their surface is screened from the solution by detergents. In our case, because the calculation of particle composition depends on theoretical contrast match point (see Equation 5.28), the uncertainty on the fraction of exchangeable hydrogen ultimately leads to uncertainties on the particle's composition. It is however noteworthy to temper this statement as three different levels of exchangeable hydrogen accessibility were

tested (70, 80 and 95%) and led to variations that were smaller than the experimental error.

Quantification of detergents and lipids

The relative calculation of scattering length density phases relies in first place on an accurate evaluation of the protein's contribution, which is already affected by the factors mentioned previously. The quantification of the lipids and detergent is therefore solely based on the excess scattering length observed after the subtraction of HTL's contribution.

Since the contribution of a full Holotranslocon is subtracted, this implies a susceptibility to over-subtraction; as the biophysical experiments have shown, the solutions were not necessarily homogeneous and HTL may have dissociated. This was particularly striking for deuterated HTL where the imbalanced expression induced by the use of fermentation in deuterated conditions led to obvious over-subtraction (see Table 2.6). In the two first cases (d-HTL in amphipols), the contrast match point clearly indicated that the particles were not only composed by deuterated proteins. However, when the contribution of the deuterated protein was subtracted, the contrast match point of the remaining components was negative, which was not possible in principle since detergents must be present in the particles. In the case of hydrogenated HTL purified with amphipols (see Figure 2.18), the scattering length was significantly lower than that of DDM solubilized HTL and close to DDM solubilized SecYEG. While this effect could be attributed to the fact that more DDM is bound to the protein than amphipol, making the scattering length appear higher, the data did not allow to certainty in ruling out the possibility that the amphipol sample partially dissociated. This would therefore lead to over-subtraction and an erroneous evaluation of the lipid/detergent bound to the sample.

If over-subtraction makes the contrast match point of the "non-protein" component appear smaller than it is, the consequences are different for hydrogenated and deuterated samples. In the first case, the "non-protein" contrast match point should be in between lipid (14,2%) and surfactant (21,7-23,5%) ; over-subtraction will lead to a smaller CMP of the lipid/surfactant components and therefore to an overestimation of the contribution of lipid which have the lowest scattering length density. With a deuterated sample, the contrast match point of the "non-protein" component should be between the surfactant (21,7-23,5%) and deuterated lipids (80,2%), the over-subtraction therefore leading to an over-estimation of the surfactant content which has a lower contrast match point. Therefore, inhomogeneity or dissociation of the sample may lead to inaccurate subtraction of the protein's contribution, which in turns leads to erroneous evaluation of the particle's composition.

Another aspect to be considered is that, in order to calculate the relative quantity of each component, it is assumed that both detergent and lipid phases are homogeneous, which might not be the case for the lipid phase. Indeed, the lipids present in the lipid cavity have been identified as phosphatidyl-ethanolamine and phosphatidyl-glycerol (Botte *et al.*, 2016), but it is not to be excluded that other species are present

and could not be detected: for example cardiolipin which has been shown to play a major role in HTL's function (Schulze *et al.*, 2014). The polydispersity of the lipid tails, which length and saturation may vary, can also affect the calculations. Both factors will affect the contrast match point of the lipid phase therefore adding an uncertainty on the evaluation of the lipid and detergent quantities (Botte *et al.*, 2016). To limit the approximation related to this quantification, a first step could be to quantify the amount of detergent bound to the complex independently by mass spectrometry (Chaptal *et al.*, 2017) or Multi Angle Laser Scattering coupled to size exclusion chromatography(Slotboom *et al.*, 2008). Ultimately, the precise identification of the lipids bound to the complex and their quantification would strengthen the neutron analyses of this system.

2.5.3 Future work and development

Plasmid design

The current plasmid (see Figure 1.9, Bieniossek *et al.*, 2009) uses two different promoters (trc and arabinose) and expresses seven proteins among which 3 are fused to a histidine tag (YidC, SecD and SecE) and one with a CBP tag (YajC). It is important to note that this organization is a consequence of the fusion of existing plasmids that were used independently. Therefore, the plasmid is neither optimized for the balanced expression of all subunits nor for the reliable purification of an intact complex. Optimization of the plasmid could be approached in different ways.

First, a different combination of promoter may favour more homogeneous expression and therefore lead to more homogeneous sample. Since the current plasmid uses two different promoters, the easiest approach would be to substitute the arabinose promoter for a trc promoter. All subunits should therefore be induced with similar efficiencies.

The combination of purification tags may also allow to separate sub-complexes more efficiently. Currently, the His-tag is represented three times in the complex. The multiplication of identical tags may favour purification efficiency. However, since the histidine tags are present on two sub-complexes (SecYEG, DFYY) and on a protein (YidC), that are known to be stable by themselves, this strategy may favour the unwanted co-purification of complexes with various compositions, therefore leading to solution heterogeneity.

Ideally, and in order to ensure the composition of a complex, the tags that are used should be unique. A single histidine tag on SecYEG and the CBP tag on YajC should allow purification of complexes that contain SecYEG with a IMAC resin and also the elimination of free SecYEG by purifying the isolated complexes over a Calmodulin resin.

If this method could in theory favour the purification of an homogeneous sample, it cannot compensate for possible disassembly of the complex during or after the purification process.

Fluorescence based integrity assay

This work has shown that it is difficult to assess the integrity of the Holotranslocon. Various biophysical techniques have been used that proved to be time and sample consuming. Furthermore, the presence of detergents often interferes with the experiments and increases the complexity of the analysis.

To reliably assess for the integrity of HTL, both *in vivo* and *in vitro*, developing a Förster Resonance Energy Transfer (FRET) based assay may be considered. By labelling for example the lateral gate of SecY and a transmembrane segment of YidC with FRET pairs, it should be possible to easily follow the integrity of the complex at any step of the experiment. Such a tool could therefore be used to monitor HTL and identify the sensitive preparation steps for a more directed optimization of the process.

In section 2.3, it was reported that SecA could only bind to SecYEG but not to the HTL's accessory subunits. However, it could not be proved whether SecA binds only to SecYEG that are already present in solution, or induces the dissociation of SecDF-YidC-YajC from the HTL upon binding. By analysing the FRET signal, it could be investigated whether the addition of SecA is correlated to a diminution of the FRET efficiency, which would confirm the second hypothesis.

Alternative detergents

The current efforts (Botte, 2013; Schulze *et al.*, 2014; Komar, Alvira, *et al.*, 2016; Botte *et al.*, 2016) have focused on the use of DoDecyl Maltoside detergent to extract HTL from the membrane, purify it and keep it stable over time. The biophysical data described here, together with previous work (Botte, 2013), suggests that this detergent is not optimal to work with the Holotranslocon.

The results presented in this work (see Figure 2.4) have shown that the nature of the detergent (non ionic, maltoside chains, 10-13 carbon long acyl chains) and the critical micellar concentration (low CMC gave better results) play key roles in the efficiency of extraction and the relative balance of subunits.

Most importantly, it is shown that alternative detergents could be used to extract HTL efficiently. This opens many possibilities for exploring new purification strategies, which may lead to the production of higher quality of samples. However, the success of this empirical approach would strongly rely on the ability to assess the stability, integrity and ideally functionality of HTL at any stage of the purification procedure. It would be of high interest to be able to clearly identify the limiting steps to tackle them rationally. Such a procedure could imply exchanging the detergent during the purification process as some detergents may be efficient at extracting HTL from the membrane while others would be ideal for keeping it intact during the purification procedure (D. Hardy *et al.*, 2016).

Membrane Scaffold Protein nanodiscs

If using surfactants is mandatory to allow for the extraction of the target membrane protein from its native membrane and for keeping them soluble in a aqueous environment, they may have a negative impact on the protein. These deleterious effect (stability, misfolding, aggregation...) can be attributed to the fact that detergents may replace lipids that play a key role in the stability and function of the protein (Palsdottir and Hunte, 2004), interfere with hydrophobic interactions that contribute to the folding of the protein (Dyson *et al.*, 2006) or generally destabilize the protein due to the lack of lateral pressure from the membrane (D. Hardy *et al.*, 2016). Therefore, being capable of keeping the protein in its native membrane environment would be of high interest for studying structurally the Holotranslocon.

Nanodiscs are generated by the self-assembly of a belt proteins, Membrane Scaffold Protein (MSP) derived from a human apolipoprotein (ApoA1), with lipids in order to reconstitute a soluble nano-sized patch of membrane (Denisov, Grinkova, *et al.*, 2004). The size of the disc can be modified by duplication of 22-mer helical segments, allowing the size to be modulated from 7 (Hagn *et al.*, 2013) to 17 nm (Grinkova *et al.*, 2010). This system has allowed the SecYEG translocon to be visualized in its native membrane environment and in complex with the ribosome by single particle cryo-electron microscopy (Frauenfeld *et al.*, 2011).

Until recent developments, the application of this system to the Holotranslocon has been partly limited by the available sizes of nanodiscs. The largest membrane protein that was successfully integrated into nanodisc was composed by 24 transmembrane helices (Schuler *et al.*, 2013) while HTL has 34. If the size measured by cryo-EM indicates HTL reache 11.4 nm at its widest (Botte *et al.*, 2016), it is important to note it corresponds to the cross-linked version of the protein. The native complex may be significantly wider (up to 45% wider as suggested by radius of gyration analysis, Botte *et al.*, 2016).

The reconstitution of HTL in nanodiscs has been attempted previously, making use of a home-designed MSP containing enough duplication of the helical segment to reach a size compatible with the target protein. These attempts were not successful and led to highly heterogeneous solutions, probably because these larger nanodiscs were not stable enough (data not shown, obtained as part of a master project under the supervision of Christiane Schaffitzel).

If in theory, nanodiscs larger than 17 nm could be generated, it has been shown that an increase of the size of MSP may lead to instability and formation of spherical aggregates (Denisov and Sligar, 2017). Recently, an alternative MSP-based nanodisc has been developed, making use of a single ApoA1 derived scaffold protein which can be circularized by Sortase mediated protein fusion (Nasr *et al.*, 2016). The covalent circularization strongly stabilizes the protein scaffold, therefore allowing for increased size (up to 50 nm) of nanodisc and limits the heterogeneity of the particles formed. This new system could be a suitable alternative to conventional nanodisc fo the production of nanodisc solubilized HTL. This approach has been initiated and circularized nanodisc

have been be produced (data not shown) but the project could not be further pursued due to lack of time.

It is however important to mention that, as discussed previously, one of the bottleneck for successful HTL purification and analysis is the use of detergents that strongly destabilize the quaternary structure of the Holotranslocon and may promote dissociation of the complex (Botte *et al.*, 2016). To date, detergents are still necessary to allow for solubilization and purification of the membrane protein prior to nanodisc reconstitution.

Styrene Maleic Acids LipoDiscs

As discussed in section 2.1.2, Styrene Maleic Acids lipodiscs (SMALPS) appear as an attractive strategy for the production of Holotranslocon. Their size, ranging from 5 to 100 nm diameter, would make them a particularly versatile tool that could accommodate a very wide range of membrane proteins. Current developments attempt to limit the effect of the polymer on the function of the target protein, favour the homogeneity of the discs produced and allow for functionalization of polymer's side chains to help purification and imaging of the system (Esmaili and Overduin, 2018).

The most striking feature of these systems is their ability to solubilize the proteins directly from the native membrane (Jamshad, Charlton, *et al.*, 2015; Jamshad, Grimard, *et al.*, 2015) without using any detergents. This approach has efficiently been used on the Holotranslocon and contributed to proving that this complex exists *in vivo* (Komar, Alvira, *et al.*, 2016). The work initiated in section 2.1.2 is a first step toward the development of a strategy that may allow the Holotranslocon to be characterized biochemically, biophysically and structurally in its native environment.

Minicells for membrane protein expression

The latest advances in hardware and image processing software has propelled cryo-electron microscopy at the forefront of structural biology. Cryo-electron tomography (cryo-ET) also benefited from these advancements and may be used for *in vivo* structural characterization of large complexes.

Farley *et al.*, 2016 has described the emergence of minicells as a potential system that may be suitable for the observation of proteins "*in vivo*" by cryo-ET. Minicells arise from aberrant cell division of rod-shaped bacteria that produce a small "cell-like" structures. Although they lack the bacterial chromosome and therefore cannot initiate cell division, they still contain a complete and functional cellular machinery that is capable of performing all other cellular processes such as transcription and translation. Mutant *E. coli* strains can favour the production of minicells and could therefore be used as a toolbox for the over-expression of proteins that could then be analysed by cryo-ET.

In the case of HTL, the use of mutant *E.coli* stains to produce minicells that contain the HTL plasmid could be imagined. Since they do not contain the genomic chromosome,

but may have the plasmid of interest, only HTL would be expressed, therefore generating mini-cells with HTL enriched membranes. The complex could then be analysed without involving any detergent or even purification step, in its native membrane-embedded state.

Such a system would open major opportunities for structural analysis, notably by cryo-ET, (as described in Farley *et al.*, 2016; Liu *et al.*, 2011). It is also possible to imagine combining this system to selective deuteration methods (see chapter 4) in order to produce samples that would contain specifically labelled HTL for SANS or even dynamic studies. If electron-microscopy may allowed for *in silico* purification, neutrons studies would however be sensitive to contamination by other proteins that are initially present in the minicell. The method would also require the redesign of the current plasmid, as the the promoters rely on the inducible T7 polymerase that is present on the chromosome.

Cell-free expression

Cell-free expression system are biotechnological tools that allow protein expressing *in vitro*. This is usually done by mixing a DNA sequence of interest with cell extracts supplemented with nucleotides and amino acids. Since the cell extract contains the transcription and translation machinery, the DNA can be transcribed and the newly synthesised RNA translated into the protein of interest (Swartz, 2006). In the past few years, the available systems have been optimized and are now commercially available for research and industrial purposes (Kuruma and Ueda, 2015). It has been shown that this type of system could be used for the expression of membrane protein by supplementing the reaction mix with liposomes (Kuruma and Ueda, 2015). Interestingly, SecYEG has been expressed in such conditions and has been shown to self-integrate into the liposomes while conserving its function (Matsubayashi *et al.*, 2014a; Matsubayashi *et al.*, 2014b; Ohta *et al.*, 2016).

The cell free expression strategy could have significant advantages for the production of the Holotranslocon. Rather than using a single plasmid containing all seven subunits, the genes could be added individually as described by (Matsubayashi *et al.*, 2014a); since the yield for each protein depends on the amount of DNA provided in the system, the ratio of subunits could be carefully controlled so that they are all stoichiometrically represented in the liposomes. If sufficient concentration of the proteins could be achieved, their equimolar distribution would favour the assembly of the Holotranslocon.

Cell free expression may also be used as a unique biological tool by providing a way of obtaining proteoliposome highly enriched in HTL, without requiring the use of any surfactant. HTL-enriched liposomes could be use to perform functional assays but could also be analysed structurally by cryo-EM or cryo-ET. In the case of neutron methods, liposomes that are designed to matched-out may be generated so that HTL could be characterized in its native environment using techniques such as neutron reflectrometry and SANS.

Chapter 3

SecB forms complexes with an unfolded substrate and SecA

Résumé en français

La chaperonne SecB joue un rôle majeur pour la translocation post-traductionnelle en interagissant notamment avec la protéine substrat pour la garder dans un état-déplié compatible avec la translocation, mais également avec la protéine SecA avec qui elle contribue à l'adressage des substrats vers les translocons.

SecB intéragit avec ses partenaires pour la translocation

Production des protéines: La protéine SecB a pu être produite en quantités suffisantes, suivant les protocoles publiés. Sa forme homotétramérique a été confirmée biochimiquement et biophysiquement. Le substrat modèle ("Maltose Binding Protein") a été produit sous forme hydrogénée ou partiellement deutérée et a pu être déplié par addition d'un agent chaotropique. La protéine moteur SecA a également été purifiée sous forme hydrogénée ou deutérée avec une efficacité similaires.

Interaction des partenaires : Nous avons montré que SecB est capable d'interagir avec une protéine dépliée ou avec la protéine moteur SecA. L'interaction de SecB avec son substrat déplié est relativement instable et une dissociation a pu être observée. Cette dissociation est cependant considérablement réduite en présence de SecA qui semble donc impliquée dans le maintien du substrat sous-forme déplié.

Caractérisation des complexes par SANS : Nous avons ensuite utilisé la diffusion de neutrons à petits angles pour analyser les complexes formés d'un point de vue structural. Nous avons combiné cette méthode avec des méthodes de deutération biomoléculaires afin de rendre possible l'identification de chaque composant au sein des complexes. Le substrat (protéine modèle) et SecA ont ainsi été deutérés avec succès et purifiés en quantité suffisantes pour effectuer les expériences requises.

Les résultats de diffusion de neutrons ont été analysés afin de calculer le "neutron contrast match point" et rayons de giration. Il est rapidement apparu que les complexes formés n'étaient pas homogènes en termes de stœchiométrie, rendant la modélisation structural inadéquate. Nous avons utilisé les valeurs de "contrast match point" et poids moléculaire calculées pour proposer des stœchiométries d'interaction qui expliquent les résultats.

Modèle d'interaction

L'analyse des rayons de giration nous a permis de confirmer que le substrat déplié est disposé en périphérie de la chaperonne et que SecA se localise de façon asymétrique. Les données suggèrent également que le tétramère de SecB peut augmenter ses dimensions significativement en présence d'un substrat.

L'étude des "contrast match point" nous a permis d'analyser la composition des assemblages moléculaires, confirmant l'instabilité du complexe SecB:substrat et indiquant que le substrat favorise la liaison d'un monomère de SecA quand l'absence de substrat favorise la présence de dimères.

Ces résultats ont été confrontés aux modèles actuels pour émettre des hypothèse concernant les mécanismes moléculaires ayant lieu pré-translocation. Nos données nous permettent de proposer un modèle où la protéine SecA semble stabiliser l'interaction entre SecB et son substrat. La présence du substrat semble également favoriser l'interaction avec le monomère de SecA, là où le dimère est favorisé en l'absence de substrat. Ces interactions coopératives pourraient contribuer à comprendre les mécanismes d'adressage des protéines aux système de translocation via le chemin post-traductionnel.

Ces expériences ont mis en évidence l'utilité de la diffusion de neutrons et marquage isotopiques pour répondre à des questions impliquant des assemblages de protéines complexes et étendre les connaissances liées au chemin de translocation post-traductionnel. Nous proposons de nouveaux axes de recherches pour mieux caractériser ces interactions d'un point de vu biochimique et structural, mais aussi étudier l'impact de la taille du substrat sur la conformation de SecB.

3.1 Preparation of SecB

The proteins used for these assays (SecB, SecA and MBP) have been produced according to the protocols described in subsection 5.2.4, subsection 5.2.3 and subsection 5.2.5. For MBP and SecA purification results, refer to subsection 4.1.1 and subsection 2.3.1.

BL21(DE3) cells were transformed with SecB plasmid, cultivated in DYT medium, protein expression induced with 1 mM IPTG and the cell lyzed by cell disruption at 1.1 kbar. The supernatant was cleared by centrifugation at 24000G and applied to anion exchange column since SecB has no tag. The resin was washed at 0 mM NaCl, 250 mM NaCl and the elution was performed using a gradient from 250 to 600 mM NaCl. Several peaks eluted between 250 mM and 450 mM NaCl (corresponds to fractions 38-50) and subsequently between 450 and 1 mM NaCl (corresponds to fractions 62-73). The resin was washed with 1 M NaCl (corresponds to fraction 83).

Figure 3.1 panel A shows the SDS-PAGE analysis, which allowed the identification of SecB, which was mainly found in fraction 50 (corresponding to 450 mM NaCl). The SecB containing fractions were pooled and loaded on a Superdex 75 gel filtration (GF) column for further purification. Figure 3.1 panel B shows the sizing profile and the main peak eluting at 9.90 ml, clearly distinguishable from the void volume. Figure 3.1 panel C shows the fractions analysed on SDS-PAGE, which confirmed that the peak fraction obtained by GF corresponded to SecB. Using the calibration standards associated to this GF column, we calculated that a retention volume of 9.9 ml indicated a molecular weight of 70 kDa, which corresponds to the expected size of the well defined homotetrameric SecB protein (Sala *et al.*, 2014). Earlier anion exchange elution fractions (around 300 mM) were also subjected to SEC and eluted at 13 ml, therefore corresponding to the size the SecB monomer. The stability of the tetramer was assessed by applying the GF elution on the same column again, which showed that despite the decreased concentration, the tetramer was still intact and eluted at 9.90 ml again (data not shown).

3.2 Size characterisation of the subunits and complexes

The interactions were characterized by comparing the SEC profile of the complexes with the proteins alone and analysing the fractions on SDS-PAGE. The individual proteins were analysed independently to determine their retention volume. In first instance, a Superdex 200 column was used but given the theoretical size of the complexes (up to 210 kDa), the retention volumes were too close to the void volume. Higher resolution and easier peak discrimination was achieved with a Superose 6 column(data not shown). Figure 3.2 panel A shows a retention volume of 16.4 ml for SecB and 15.9 ml for SecA. In the case of MBP, two peaks were observed at 16.4 ml and 17.9 ml, corresponding respectively to the unfolded or partially folded MBP and to the refolded MBP as confirmed by GF performed on the folded MBP (data shown). We also observed that a small fraction of MBP had aggregated and found in the void volume (8 ml, not shown).

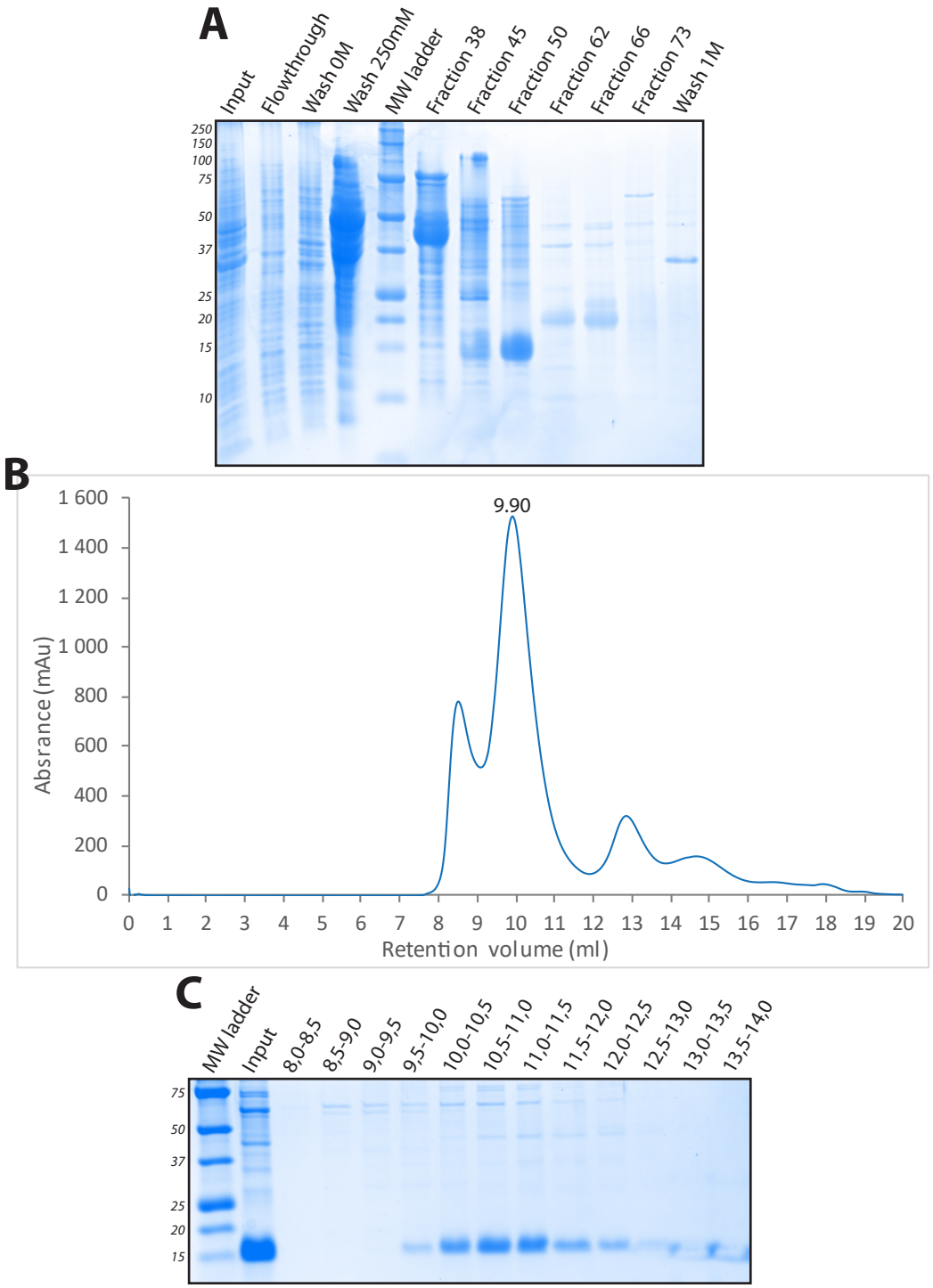


FIGURE 3.1: Purification of SecB **A:** SDS-PAGE analysis of the anion exchange performed on 5ml Q-sepharose column. The washes are done with NaCl, the elution with a NaCl gradient (250-600mM) and the column is washed with 1M NaCl. **B:** Size exclusion chromatography performed on a Superdex 75. The peak fraction elutes at 9.90 ml. **C:** The SEC fractions are analysed on SDS-PAGE.

Figure 3.2 panel B confirmed the identity of the proteins and quantities observed on gel are in agreement with the SEC profiles. The retention volumes of SecB, folded MBP, and SecB:MBP complexes were sufficiently different to be clearly distinguished.

3.2.1 Interaction with an unfolded substrate

The binding of MBP with SecB was performed by unfolding MBP in 6 M urea and by sequential quick dilution into the SecB solution at a final SecB:MBP molar ratio of 4:1.2 (considering monomers of SecB for the calculation of its concentration, thus providing MBP in slight excess). Despite the fact that the experiment is in principle rather trivial, the final urea concentration proved to be a decisive factor in the stability of the complexes (SecA tends to aggregate in high urea). To keep the final urea concentration below 2 M, MBP was unfolded in 6 M urea (instead of 8M urea initially) and SecB diluted.

Figure 3.2 panel A shows that the SecB:MBP complex (in theory 111.6 kDa) had a higher retention volume (15.6 ml) than its individual components (16.6 ml and 16.3 ml) and was close to what was calculated for SecA alone (15.9 ml for 100 kDa). Figure 3.2 panel B confirms that SecB and MBP co-eluted earlier (14.5-16.5 ml), which clearly indicated that the SecB tetramer and MBP formed a complex. However, when the complex was concentrated and applied again on GF (sample SecB:MBP R), the appearance of a shoulder at 16.0 ml and a peak at 18.2 ml suggested that the complex had dissociated; this was confirmed on panel B as MBP is mostly found in the 17,0-18.5 ml fractions. These results indicated that SecB interacts with unfolded MBP but that the complex tends to dissociate when it is further concentrated or submitted to a subsequent gel filtration.

3.2.2 Interaction of the complex with the ATPase

The binding of the SecB:MBP complex to SecA was carried out by adding SecA to the already formed complex at a SecB:MBP:SecA molar ratio of 4:2:1 while ensuring the urea concentration stayed bellow 2 M, since it induced aggregation of SecA (data not shown). Figure 3.2 shows the full complex was characterized by a retention volume of 15.0 ml and that a fraction of refolded MBP could be observed at 18.2 ml due to the excess of MBP provided. The presence of all three proteins was assessed by SDS-PAGE (data not shown). The major peak was collected, concentrated and submitted to a second gel filtration in order to assess the stability of the complex (sample SecB:MBP:SecA R). Surprisingly, the shoulder and dissociation that was observed with the SecB:MBP complex did not occur. Panel B confirms that the majority peak was composed of the three proteins in equal proportions. The integrity of the complex was further confirmed by the presence of all three proteins in the fraction at 14.0 ml, which was empty in the other experiments. These data suggest that in addition to interacting with the SecB:MBP complex, SecA seemed to stabilize it. A dataset acquired independently confirmed these observations and also suggested that SecA can interact with the SecB tetramer alone. However, the interaction of SecA to SecB alone seemed

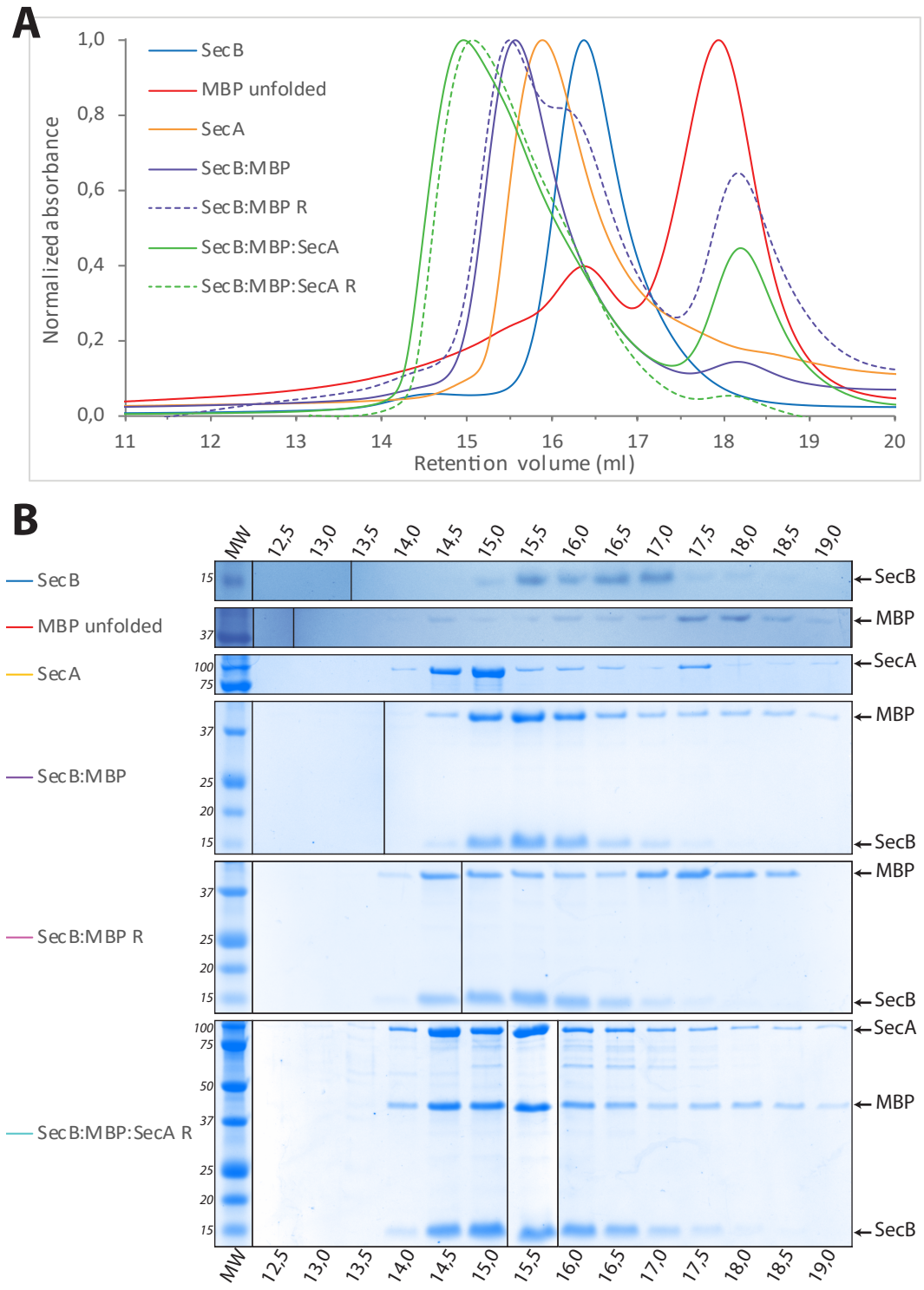


FIGURE 3.2: Characterization of the interaction of SecB with unfolded MBP and SecA by gel filtration performed on Superose6 column. **A:** scaled GF profiles of SecB (blue), unfolded MBP (red) or SecA Alone (orange), the complex SecB:MBP (purple) and the complex SecB:MBP:SecA (green). The peak fractions of the complexes were run again on gel filtration to test the stability of the complexes (noted R and represented as dashed lines). **B:** corresponding SDS page of the elution fractions. Lanes originating from a same gel have been reordered and aligned to match the gel filtration results.

less strong than the one reported with SecB:MBP as the GF results suggest that SecB alone was detected (data not shown). These results were consistent with published work (Hartl *et al.*, 1990).

These data suggested that the SecB tetramer can interact with both unfolded MBP and SecA to form a stable complex.

3.3 Small-angle scattering analysis

In order to further investigate the complexes involving SecB, we performed small-angle neutron scattering (SANS) experiments using deuterium labelled proteins. We therefore measured :

- hydrogenated SecB alone in 0% or 99% D_2O
- hydrogenated SecB in complex with unfolded match-out labelled Maltose Binding Protein (noted hSecB:dMBP)
- hydrogenated SecB in complex with match-out labelled SecA (noted hSecB:dSecA)
- hydrogenated SecB in complex with unfolded hydrogenated MBP and match-out labelled SecA (noted hSecB:hMBP:dSecA)

The interaction was prepared in similar conditions as before, with MBP being unfolded in 6 M urea then added sequentially to SecB by rapid dilution. SecA was added to the SecB or SecB:MBP complex at equimolar ratios. The different complexes were processed by size exclusion chromatography on a Superose 6 column in hydrogenated or deuterated buffers (no differences were observed between the two buffers). The early peak fractions were collected in order to favour the recovery of larger intact complexes and H_2O and D_2O elutions were mixed to generate the contrast variation series.

The sample were measured without prior concentration on the instrument KWS1 (FRMII, Garching) at 6 °C. The data were reduced using the in-house software, correcting for transmission, electronic background, and flux. Due to time constraints, data collected at 17 m and 10 Å could not be measured for long enough to obtain reliable results and were therefore omitted from the analysis. The low quality of the data recorded for the buffers led us to subtract a constant, defined at high angle for each sample, to the scattering data (data not shown). It is important to note that despite the limited statistics, the signal of the buffers was strictly dominated by a flat incoherent background (see Figure 3.3), which legitimated the subtraction of a constant. The Guiniers fits are represented in Appendix C.

The Guinier approximation was used and the fits carried out using ATSA (v2.8.3) to extract the forward scattering intensity (I_0) and radius of gyration (Rg), with $Rg * q < 1.3$ (see example of fit on Figure 3.3). The results are summarized in Table 3.1. $\Delta\rho V$ was calculated according to Equation 5.25 and the contrast match point was calculated from the linear fit applied to the data (using IgorPro v6.3.7.2). The molecular weight was calculated from data collected in 0% D_2O using Equation 5.26 and values of $\Delta\rho$ computed with MULCh (Whitten *et al.*, 2008) from the sequences of the proteins,

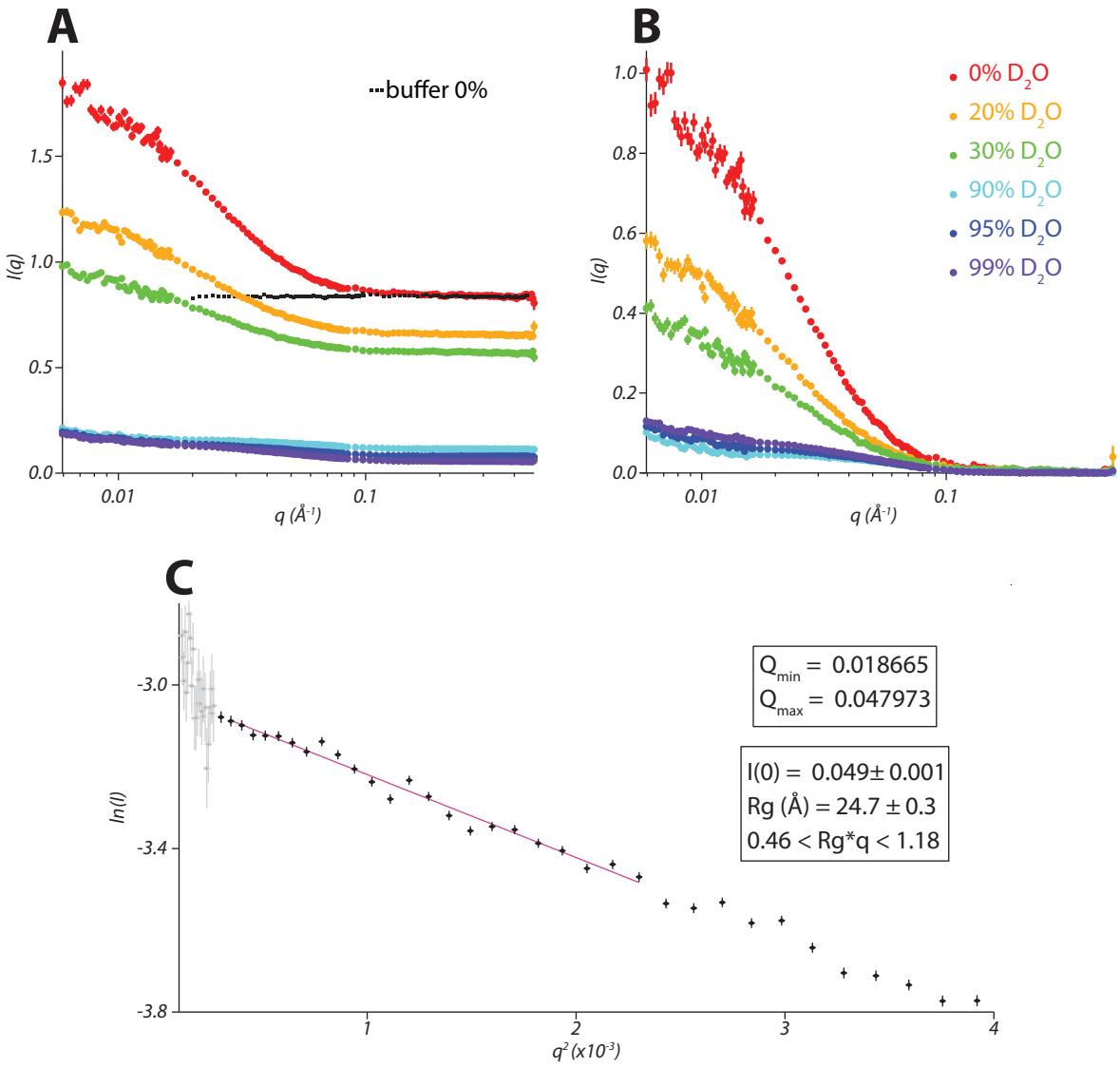


FIGURE 3.3: Example of SANS data collected on KWS1, FRMII. The dataset shown corresponds to the hSecB:hMBP:dSecA complex measured at 0% (red), 20% (orange), 30% (green), 90% (cyan), 95% (blue) and 99% (purple) D_2O. **A**: Data before solvent subtraction. The black squares represent the buffer at 0% D_2O. **B**: Same dataset after constant subtraction. All the samples reached I=0 at high angle. **C** typical Guinier approximation fit (in magenta) on the 90% D_2O sample. The data points in grey were collected at larger detector distance (lower neutron flux) and were too noisy to be taken into account.

TABLE 3.1: Guinier analysis data for the SecB complexes

Sample	$\%D_2O$	$[C]_{prot,(\mathrm{mg\,ml^{-1}})}$	$I_{0,(\mathrm{cm^{-1}})}$	$\sigma(I_0)$	$Rg_{\mathrm{Å}}$	$\sigma(Rg)$
SecB	0%	4,8	0.171	0.001	25.2	0.4
	99%	4,8	0,280	0.001	25,4	0.1
hSecB:dMBP	0%	4,0	0,400	0,007	39,3	0,7
	20%	4,0	0,190	0,003	41,3	0,7
	30%	3,2	0,087	0,002	40,5	0,7
	42%	3,3	0,025	0,002	22,9	1,2
	90%	2,5	0,073	0,001	30,0	0,6
	95%	3,9	0,160	0,001	34,1	0,5
	99%	2,8	0,140	0,001	33,8	0,4
hSecB:dSecA	0%	2,5	1,150	0,008	51,1	1,9
	30%	2,4	0,520	0,005	53,1	1,0
	95%	2,3	0,029	0,001	22,1	0,8
	99%	2,3	0,034	0,000	23,1	0,6
hSecB:hMBP:dSecA	0%	2,1	0,840	0,011	55,2	1,3
	20%	2,1	0,460	0,008	54,3	1,2
	30%	2,1	0,290	0,005	48,8	2,5
	90%	2,1	0,049	0,001	24,7	0,3
	95%	2,1	0,066	0,001	30,9	0,5
	99%	2,1	0,086	0,001	34,5	0,3

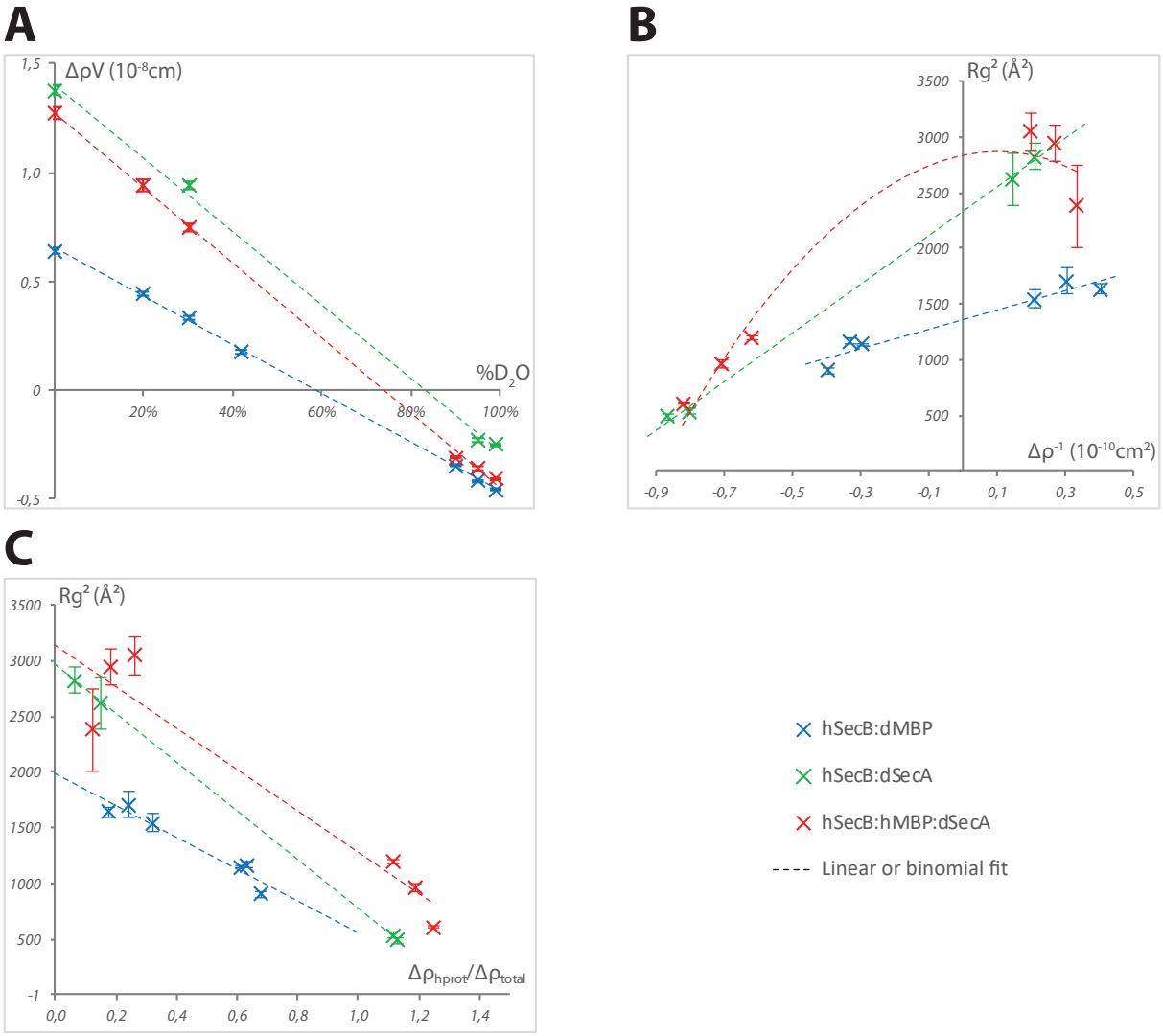


FIGURE 3.4: Analysis of the small-angle neutron scattering results obtained for the hSecB:dMBP, hSecB:dSecA and hSecB:hMBP:dSecA complexes. **A** Excess scattering length ($\Delta\rho V$) plotted against the fraction of D_2O allows to calculate the contrast match point of each complex. **B** Stuhrmann analysis of the radii of gyration (Rg) squared against the inverse of the contrast $\Delta\rho V$, giving an indication of the relative organisation of the scattering density phases. **C** Parallel axis analysis plots the squared *Rg* against the ratio of the hydrogenated phase contrast and calculated contrast in order to yield information on the radii of gyration of each phase.

the composition of the buffer and assuming 80% of the labile hydrogen are exchanged (Jacrot and Zaccai, 1981). The radius of gyration at infinite contrast (Rm, $\Delta\rho^{-1} = 0$) is either calculated from linear (hSecB:dMBP and hSecB:dSecA) or second order polynomial (hSecB:hMBP:dSecA) fit on the Stuhrmann plot.

Due to the high disparity of the errors associated to the radii of gyration, the fit was performed without error weighting. The Rg of the hydrogenated proteins (Rg_{hprot}) was calculated from the linear fit applied using the parallel axis method at $\Delta\rho_{hprot}/\Delta\rho_{total} = 1$ and that of deuterated protein (Rg_{dphase}) at $\Delta\rho_{hprot}/\Delta\rho_{total} = 0$.

Figure 3.4 panel A shows that a straight line could effectively be fitted to the data. The contrast match point analysis and calculation of the molecular weight of the particles (see Table 3.2) showed that, in all cases, the experimental data differed from the theoretical values.

3.3.1 SecB

The SecB tetramer was only measured on its own in 0% and 99% D_2O in order to verify the molecular weight and radius of gyration. The contrast match point was not considered since only 2 contrasts were measured. It is important to note that the measured Rg (25,2 ±0,4 at 0% and 25,4 ± 0.1 Å at 99% D_2O, see Table 3.1) was very similar to the Rg inferred from the crystal structure of SecB tetramer calculated with CRYSON (v2.7, Svergun, Richard, *et al.*, 1998) at 24.1 Å(pdb:1QYN). A monomer of SecB would have a Rg of 17.5 Å and the dimer a Rg of 20,5 to 22.0 Å depending on the geometry of the dimer. The calculated molecular weight (64,6 ±1.2 kDa) was in reasonable agreement with the theoretical molecular weight (69.0 kDa). When measured from the 99% D_2O contrast, the value calculated was significantly smaller than expected (58.1 kDa), an observation that could be attributed to the fact that the calculation of the molecular weight depends on the contrast (Equation 5.26), which itself depends on the unpredictable fraction of labile hydrogen that are actually exchanged for deuterium coming from the buffer. Jacrot and Zaccai, 1981 described errors in the calculation of the molecular weight that could reach 10% error that could explain the underestimation of the molecular weight. This does not apply to the determination of the molecular weight calculated from a sample in 0% D_2O since only hydrogen is present.

The reliable purification process together with a consistent radius of gyration and molecular weight, led us to believe SecB was reliably produced and analyzed as the tetramer we are expecting.

3.3.2 SecB in complex with deuterated MBP (hSecB:dMBP)

Analysis of the Contrast match point and molecular weight

Compared to the theoretical values of a 4:1 complex, we observe that the measured contrast match point (58,8%) was lower than the expected value (64,4%, see Table 3.2).

TABLE 3.2: Contrast match point, molecular weight and radii of gyration

Theoretical values

Protein/complex	*Stoichiometry*	$CMP(\%)$	$MW(kDa)$	Rg_{cryson}(Å)
hSecB	4	41,9	69,0	$24{,}1_{pdb:1QYN}$
hMBP	1	41,1	42,5	undefined
dMBP	1	100,8	44,2	undefined
dSecA	1	101,1	98,8	$33{,}2_{pdb:1M6N}$
dSecA	2	101,1	197,7	$42{,}1_{pdb:2FSF}$
hSecB:dMBP	4:1	64,4	113,3	
hSecB:dSecA$_1$	4:1	76,1	167,9	
hSecB:dSecA$_2$	4:2	85,2	266,7	
hSecB:hMBP:dSecA$_1$	4:1:1	68,8	210,4	
hSecB:hMBP:dSecA$_2$	4:1:2	78,9	309,2	

Measured values

Complex	$CMP(\%)$	$MW(kDa)$	Rm	Rg_{hprot}	Rg_{dprot}
hSecB	*43,2%	64,6 ±1,2			
hSecB:dMBP	58,8 ±1,1	77,3 ±2,0	39,9	23,2	44,5
hSecB:dSecA	83,3 ±1,7	259,2 ±4,5	48,2	27,9	54,3
hSecB:hMBP:dSecA	73,6 ±1,7	270,6 ±6,3	53,2	35,7	55,9

**CMP (Contrast Match Point) calculated from only 2 data points*
Similar scattering objects are coloured

This result may indicate that the solution contained an excess of the hydrogenated SecB, reducing the CMP. Consequently, the measured molecular weight (77.3 kDa) is also smaller than expected (113.3 kDa).

It is important to note that small-angle neutron scattering signal arises from the averaging of all particles in solution. It was possible to explain these results by considering that the solution was not solely composed of intact SecB:MBP complexes. Indeed, when considering the molecular weight, the reduced value could be explained by a dissociation of the complex into SecB tetramer and partially or totally refolded MBP. This consideration is strongly supported by our size exclusion chromatography analysis since we clearly observed in Figure 3.2 that this complex was not stable and dissociation could be clearly observed less than an hour after gel filtration. As for the contrast match point and given the time scale considered for the dissociation of the complex, it is likely that the complex partially dissociated during the size exclusion chromatography. Our experimental design favours larger complexes (by collecting the early fractions of the peak) and it is reasonable to suggest that our final sample may be enriched in free SecB tetramer since its hydrodynamic radius is bound to be higher than that of refolded MBP, let alone partially unfolded MBP.

We ran data simulations of mixtures having different components stoichiometry in order to explain the measured values (see Figure 3.5). In the case of the molecular weight, the ratio of each component weights its contribution to the final molecular weight. In the case of the contrast match point, the contribution takes into account the ratio and fraction of its molecular weight as compared to the calculated molecular weight. We simulated different mixtures, considering that the solution was composed of hSecB:dMBP complexes, free SecB tetramer and MBP (with their respective calculated molecular weight and CMP); we found that a mixture of 30% complex, 50% SecB tetramer and 20% MBP gave the same contrast match and molecular weight as experimentally measured. The comparison of theoretical contrast match point and molecular weight is represented on Figure 3.5 and provides a visual support for an understanding of the mixtures.

Analysis of the radius of gyration

The parallel axis data analysis indicated that the radius of gyration of the hydrogenated protein (in this case SecB) was 23.2 Å (see Table 3.2), in good agreement with the theoretical Rg of SecB alone, as calculated from the crystal structure. The Sturhmann plot indicated that the linear fit had a positive slope, which indicated that the component with high scattering length density (here deuterated MBP) was located in periphery of the lower density (here SecB). This observation was supported by the higher Rg calculated for the deuterated phase by the parallel axis analysis (44.5 Å), and the comparison with the theoretical Rg of a folded MBP (cryson, pdb:1ANF) that is calculated at 20.6 Å indicate that MBP is mostly not found as a folded protein. The radius of gyration at infinite contrast (RM) therefore corresponds to an intermediate value between that of SecB and that of unfolded MBP.

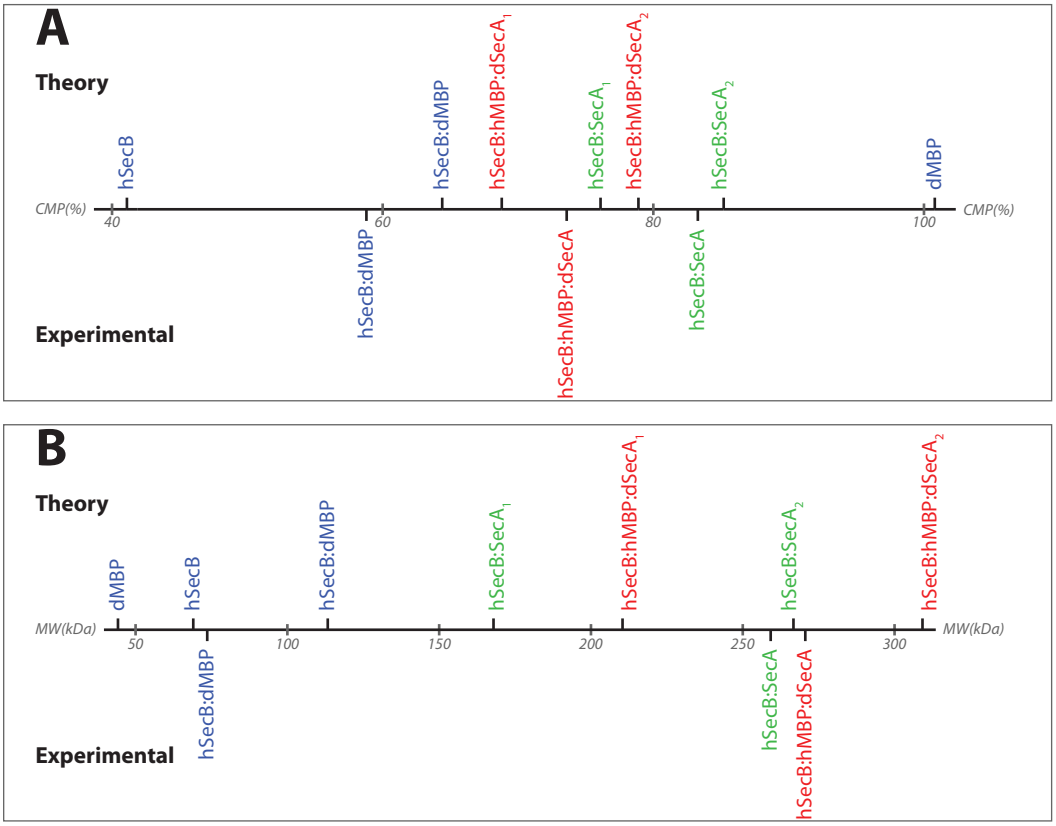


FIGURE 3.5: Representation of the contrast match point and molecular weights of the complexes. **A**: The theoretical CMP are represented above the axis and the measured values bellow. **B**: The theoretical MW are represented above the axis and the measured values bellow. The relative positioning of the measured complexes with regards to the relevant theoretical complexes represent the mixture that explain the data.

It is also interesting to note that, as compared to SecB alone, the measurement of the complex in 99% D_2O (so with MBP matched out) gives a Rg of 33 Å. This increased Rg upon MBP binding could be a consequence of a substrate mediated reorganization and expansion of the SecB tetramer.

3.3.3 SecB in complex with deuterated SecA (hSecB:dSecA)

Contrast match point of deuterated SecA

We independently measured the contrast match point of 75% deuterated SecA in order to confirm that the theoretical value computed from the sequence and SecB interaction conditions were consistent with experimental values. We therefore used SecA purified according to subsection 5.2.3 and measured the sample on the instrument D22 at the ILL at a wavelength of 6 Å.

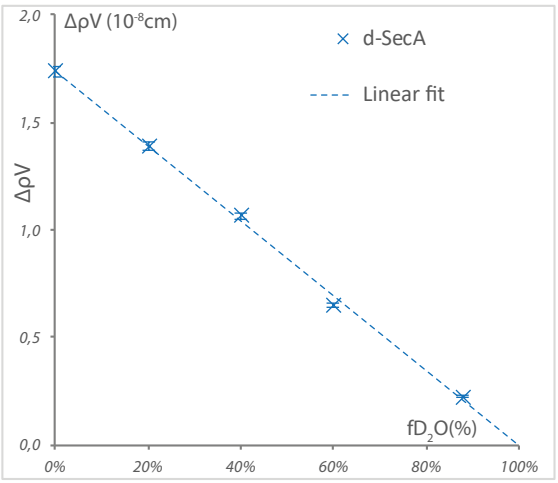


FIGURE 3.6: Determination of the contrast match point of 75% deuterated SecA. Excess scattering length ($\Delta\rho V$) plotted against the fraction of D_2O allows to calculate the contrast match point of the protein.

The contrast match point of the protein was calculated at 99,5% by fitting a linear curve to the data (see Figure 3.6), which was in good agreement with the theoretical values (99,1% in SecA's buffer). In the case of SecB's buffer, the high concentration of potassium acetate (300 mM) brought the contrast match point of deuterated SecA to 101,1%. In this case, weighing the data by their error generated an artefact as the point measured at 87,7% (closest to the match point) was surprisingly associated to a low error, and led to an overestimation of the contrast match point.

This experiment confirmed that our evaluation of the contrast match point was reasonable and that we could reliably use the computed value for SecB's buffer (101,1%) to evaluate the composition of SecA containing complexes.

Analysis of the contrast match point and molecular weight

The measured contrast match point (83,3%) was higher than the value theoretically calculated from a 4:1 hSecB:dSecA$_1$ complex (76,1%, see Table 3.2). This result indicated that the solution contained an excess of the deuterated SecA, that consequently raised the CMP. Consistently, the measured molecular weight (259.2 kDa was higher than expected from a 4:1 (167.9 kDa).

Following the same reasoning as used for hSecB:dMBP complex, an increase of both the CMP and MW could not be explained by an excess of free SecA alone. Interestingly, SecA is known to be found as a dimer (Zimmer *et al.*, 2008) in the cytosol but also when bound to SecB (Suo *et al.*, 2015). Even though the SecA was purified in reducing conditions and corresponded a truncation that should limit is capacity to dimerize (Erlandson *et al.*, 2008), the fact that SecA dimerizes in physiological conditions and may bind SecB as a dimer (Suo *et al.*, 2015) proved to be a reasonable explanation of the data.

Simulations based on theoretical CMP and MW of both hSecB:dSecA$_1$ and hSecB:dSecA$_2$ complexes (see Figure 3.5) indicated that our data could be explained as a mixture of 30% 4:1 and 85% 4:2 complexes. In this case, we put the emphasis on the CMP as this measured value was calculated on the basis of a linear fit involving four independent data points as opposed to the MW that was calculated from a single data point and therefore more error prone. By fixing the CMP of the simulation at 83,3%, we calculated a MW_{sim} of 237.1 kDa. A simulation of a 10% 4:1 and 90% 4:2 mixture satisfied both parameters ($CMP_{sim} = 84,6\%$, $MW_{sim} = 256.8$ kDa), within the experimental error ($CMP = 83,3 \pm 1.7\%$, $MW = 259,2 \pm 4,5\%$). A mixture of 15% free SecB with a 85% hSecB:dSecA$_2$ complex (5% and 95% within the experimental error) gives exactly the same results. The latter scenario is however less likely as our gel filtration strategy favours big complexes and the MW difference (almost 200 kDa) would have been clearly separated.

These results may seem surprising if we consider the fact that SecA was added to the solution in equimolar proportions to SecB tetramer and should therefore favour the 4:1 complex. However, the large dominance of hSecB:dSecA$_2$ may simply be explained by the gel filtration strategy which favours lower elution fractions, and therefore bigger complexes. It is important to note that the hydrodynamic radii of the considered complexes may have been too close to be discriminated by gel filtration.

Analysis of the radius of gyration

The parallel axis analysis (see Table 3.2) indicates that the calculated molecular weight of SecB is higher than expected but it is interesting to note that the Rg measured close to the match point of SecA (e.g at 95 and 99% D_2O) was similar to that expected (22-23 Å, see Table 3.1). The error observed when referring to the parallel axis analysis may be a consequence of the low number of points (4) in this analysis, which may therefore reduce its accuracy.

The calculated Rg of SecA (54.3 Å) was significantly higher than that obtained from the crystal structure of a dimer (42 Å). This could also be a consequence of the low number of points used for the analysis. This result could either indicate that the dimer may not be as compact in solution as observed in a crystalline form, an hypothesis that was supported by the fact that this truncated form of SecA Erlandson *et al.*, 2008 may have reduced dimerization capacities and therefore be more loosely associated.

Finally, the straight fit applied to the Sturhmann analysis would suggest that SecA is surrounding SecB, with for example one copy of SecA on each side of SecB. However, the number of points available for the fit was insufficient to deny the possibility that the fit could be parabolic, which would indicate two misaligned centre of masses, a model in better agreement with our understanding of SecB:SecA complexes.

3.3.4 Hydrogenated SecB and MBP in complex with deuterated SecA (hSecB:hMBP:dSecA)

Analysis of the Contrast match point and molecular weight

We observed that the measured contrast match point (73,6%) was higher than the value theoretically calculated from the hSecB:hMBP:dSecA$_1$ complex (68,8%, see Table 3.2). This result may indicate that the solution contains an excess of deuterated protein that cannot be SecA alone since the measured molecular weight (270.6 kDa was also higher than expected from the hSecB:hMBP:dSecA$_1$ (167.9 kDa).

In a similar way as described previously, and since we did not observe dissociation by gel filtration for this complex, simulations were performed in order to explain the measured results by mixtures of complexes (see Figure 3.5). Consistent with the previous observations, the CMP could be explained by a mixture of 60% hSecB:hMBP:dSecA$_1$ with 40% hSecB:hMBP:dSecA$_2$, which results in a theoretical MW of 249.9 kDa. A mixture of 45% 4:1:1 with 55% 4:1:2 complex gave $CMP_{sim} = 75,3\%$ and $MW_{sim} = 264.7$ kDa, which satisfied both parameters within the experimental error ($CMP = 73,6 \pm 1.7\%$, $MW = 270,6 \pm 6,3$). If the occurrence of a complex without SecA is unlikely (because of the gel filtration strategy), the simulation considering 35% of hSecB:hMBP and 65% of hSecB:hMBP:dSecA$_2$ gives a MW_{sim} of 240.0 kDa. In this case, both parameters could not be satisfied within experimental error.

This result was in agreement with the previous complexes (hSecB:dSecA) and showed that it was necessary to consider the presence of the hSecB:hMBP:dSecA$_2$ complexes containing a dimer of SecA to explain both the contrast match point and molecular weight measured.

Analysis of the radius of gyration

For this complex, the Rg calculated (see Table 3.2) for the hydrogenated phase (here SecB and MBP) was of 35.7 Å, which was in reasonable agreement with the Rm (radius of gyration at infinite contrast) calculated for the hSecB:dMBP complex (39.9 Å). The

Rg calculated for SecA was similar to that calculated for SecA within the hSecB:dSecA complex. The Rm was unsurprisingly higher than the Rm of other complexes measured.

Most interestingly in this case, a linear fit appeared particularly inappropriate for the Stuhrmann plot and the data were best explained using a binomial fit. The parabolic aspect of the fit indicated that the complex was composed of two volumes having different scattering length and with misaligned centres of mass. These data would be in agreement with the current models that report an asymmetrical complex (Suo *et al.*, 2015). With regards to the previous hSecB:dSecA Rg data, these results strengthens our suspicion that a linear fit was not optimal and that a larger number of points may have allowed the binomial trend to be more clearly established.

3.4 Discussion

SecB is a unique example of a chaperone that is dedicated to translocation. This tetrameric protein (Xu *et al.*, 2000) helps maintain the pre-proteins in a translocation-competent state by preventing their folding and contribute to the targeting and delivery to the SecYEG translocation machinery (Hartl *et al.*, 1990; Bechtluft, Nouwen, *et al.*, 2010). The interaction of SecB with its substrate is mostly permitted by hydrophobic interactions (Bechtluft, Kedrov, *et al.*, 2010).

Recent studies (Huang *et al.*, 2016) further investigated these interaction and interestingly showed that the substrate protein wraps around the tetramer so that the contacts between the two proteins are maximized, and buries its hydrophobic side chains into SecB's binding grooves. This organization not only prevents the formation of tertiary structures within the substrate but also the formation of secondary structure. Most significantly, this study show that SecB has the capacity to adjust its structure so that the binding grooves can accommodate 25 to 40 residues of the substrate (Huang *et al.*, 2016). Interestingly, the affinity of SecB for a substrate is positively correlated to its size since the surface of contact is increased with longer proteins.

The capacity of SecB to help the targeting of pre-preotein to the peptide conducting channel is intimately linked to its capacity to interact with the SecA ATPase (Suo *et al.*, 2015). It is also reported that SecA can be bound to SecB as a monomer or as a dimer, the latter being more efficient at promoting translocation.

In this work, we investigated the capacity of SecB to form a complex with its substrate but also with SecA. We made use of the unique opportunities offered by the combination of contrast variation and SANS with protein labelling to analyse the individual components within the various complexes.

3.4.1 SecB forms a complex with an unfolded substrate and SecA

SecB is known to be able to bind unfolded protein non specifically and we showed SecB could bind the unfolded maltose binding protein we used as a model. This complex

proved to be rather unstable both by size exclusion chromatography (Figure 3.2) and SANS (Figure 3.4). However, addition of SecA to the mixture not only allowed bigger complex to form, but also stabilized the binding of unfolded MBP to SecB.

It has also been shown that SecA could formed a complex with SecB, either as a monomer but most significantly as a dimer. However, when SecA was bound to SecB that already had its substrate wrapped around it, the monomeric form of the ATPase was favoured.

These observations raise the exciting possibility that the network of interaction between SecB, its substrate, and SecA may be cooperative. Indeed, SecA would stabilize the interaction of the substrate with SecB but the substrate would also stabilize the complex where a single SecA is bound to SecB, a form that is usually less frequent (Suo *et al.*, 2015). This hypothesis is supported by the previously reported asymmetry of the SecB:substrate:SecA with the two copies of SecA having different affinity for the complex, one of them intimately contacting both SecA and the substrate while the other remains in a closed state, its peptide binding domain being folded over the C-terminal domain (Tang *et al.*, 2011). Together with our results, we can hypothesize that the additional contact that SecA can establish with substrate are sufficient to further stabilize the SecB:substrate complex but also stabilize the SecA monomer on SecB. Because our experiment have been done in stoichiometrical conditions that favour the SecA monomer, we could describe a difference of stoichiometry with or without substrate, observation that might not have been possible if SecA was provided at higher concentration, therefore naturally favouring the dimeric form.

3.4.2 SecB undergoes substrate driven expansion

Huang *et al.*, 2016 reports that SecB modulates its structure locally in order to accommodate various sizes of substrate, allowing it to insert different lengths of unfolded hydrophobic regions in the binding grooves of SecB. However, this NMR-based study could not account for large scale reorganization that may occur in SecB.

Here we observed that when SecB was not bound to any substrate, its radius of gyration was the same as predicted from the crystal structure (24-25 Å), and SecA did not seem to have any effect on SecB. However, when the substrate was bound to SecB, we noted a dramatic increase (up to 50%) of the radius of gyration, an observation that was possible because the substrate is matched out, and therefore invisible.

This observation could indicate that, in addition to remodelling its peptide binding grooves, SecB may undergo a large-scale transition to modulate its size. This would be of particular interest as it would account for SecB's capacity to chaperon a wide range of substrate sizes. The substrate wraps around the tetramer; therefore, if SecB has the capacity to modulate its dimension, the surface available for the substrate binding would vary accordingly. This variation of size could therefore favour the binding of bigger substrate (as reported in Huang *et al.*, 2016) by offering more binding opportunities, but may also help guiding the target sequences into the peptide binding grooves.

SecB tetramer has a disc-like structure and it is known that SecA contacts negatively charged β-sheets forming the "flat" side of the tetramer (Diao *et al.*, 2012; J. Zhou and Xu, 2003). It is therefore tempting to speculate that this increase of dimensions might also expand the SecA binding site on SecB, which could therefore account for the higher affinity for SecA monomer previously discussed.

3.4.3 Working model

SecB tetramer (Xu *et al.*, 2000) and SecA dimer (Zimmer *et al.*, 2008) are present in the cytoplasm and may interact. It has been proposed that when the pre-protein is being translated, SecA interacts with the signal sequence (Gelis *et al.*, 2007) and SecB with the body of the protein, wrapping it around the tetramer (Huang *et al.*, 2016). The proximity of SecB and SecA mediated by their binding to the substrate favour their interaction. The expansion of SecB triggered by the substrate may also increase the surface of the SecA interaction site, further enhancing the affinity of SecA for SecB and thereby stabilizing the complex (current work). This stable complex permits the substrate of the transmembrane translocation machinery to be targeted (Suo *et al.*, 2015). Upon binding of the complex to the translocon, the SecA ATP cycles are initiated (Lill *et al.*, 1990), the presence of the substrate favours the loosening of SecA dimer (Gouridis *et al.*, 2013), and one protomer of SecA is released (Gouridis *et al.*, 2013). The release of a SecA protomer would therefore preferentially happen upon initiation of the translocation, event favoured by the presence of the substrate on SecB (present data). This ATP hydrolysis-mediated release is favoured by an asymmetric binding of SecA with SecB that may also favour the transfer of the substrate from SecB to the translocon (Suo *et al.*, 2015). SecA undergoes ATP catalytic cycles to move the pre-protein through the translocon (Allen *et al.*, 2016; Zimmer *et al.*, 2008; Erlandson *et al.*, 2008).

3.4.4 Future work

The present results contribute to a better understanding of the post-translational translocation pathway but also that small-angle neutron scattering may prove to be of great use for further investigation of large multi-protein complexes relevant to the translocation mechanisms.

The use of contrast variation together with differential deuteration of the components may prove to be of great use as it allows the stoichiometry of the complexes to be investigate. By successively analysing a specific complex at the contrast match point of its different components, the effect of the interaction on the structural features of each component independently could be assessed, provided that an homogeneous solution could be prepared.

In the current work, we suggest that the presence of a substrate on SecB is stabilized by the binding of SecA but also that the binding of SecA is favoured by the presence of the substrate. In order to further investigate these points, it would be interesting to determine the coefficient of dissociation of SecA (monomer or dimer) with SecB

in the presence and absence of substrate. This could be done by isothermal titration calorimetry or surface plasmon resonance. It may also be interesting to assess the effect of SecB/substrate on the ATPase activity of SecA and inversely whether the SecA ATP hydrolysis affects the SecB/substrate interaction with SecA. It could also be envisaged to add the SecYEG translocon to the system in order to investigate the structural and dynamic characteristic of this complex.

More generally, obtaining additional structural information on these complexes would be of high interest. However, given the proven polydispersity of the solution produced, *de novo* modelling could not be performed. Therefore, further structural analysis of these complex by SANS will required optimization of the conditions in order to have a careful control of the complexes stoichiometry.

In order to investigate the potential rationale behind the observed expansion of SecB upon MBP binding, it would be interesting to investigate the effect of various sizes of substrate on the radius of gyration of SecB. The same strategy could therefore be applied and a range of deuterated substrate sizes may allow us to investigate whether this expansion can be reproduced with other substrate and whether it is substrate dependent. These different substrates could also be tested for their affinity for SecB. We could also investigate whether the stability of other SecB/substrate complexes depend on SecA to be stabilized and inversely whether they favour the interaction of SecA with SecB.

Chapter 4

Differential *in vivo* deuteration of biomolecules

Résumé en français

La deutération des molécules ouvre des possibilités uniques pour l'étude des complexes biologiques. Par exemple, dans un complexe « protéine-protéine », une deutération sélective de l'un des composants, permet de les différencier lors des analyses et d'obtenir des informations structurales uniques. Dans le cas des protéines, la relation entre les conditions de culture a été précédemment établie et il devient possible de les deutérer à n'importe quel niveau et de façon prédictible. En revanche, si les acides nucléiques et les lipides ont été deutérés par le passé, le marquage partiel n'a pas été étudié pour permettre de le réaliser de façon prédictible et maîtrisée.

Deutération des molécules biologiques

Nous avons cultivé des bactéries dans des conditions de deutération croissantes afin d'étudier la relation entre les conditions de culture et l'intégration de deutérium dans les molécules biologiques. Pour mesurer le niveau de deutération des biomolécules, nous exploité les différences de propriétés physiques entre l'hydrogène et le deutérium.

Nous avons utilisé la diffusion de neutrons à petits angles pour la détermination du "contrast match point" des protéines et lipides ; cette valeur a été reliée au niveau de deutération. La spectrométrie de masse a permis le calcul direct du niveau de deutération mais seulement pour les protéines. Après l'échec de la diffusion de neutrons à petits angles et de la spectrométrie de masse pour analyser les acides nucléiques, la résonance magnétique nucléaire a permis de relier le niveau d'hydrogénation avec celui de deutération. Ces données nous ont permis d'établir un lien entre les conditions de culture des bactéries *E.coli* et le niveau de deutération des molécules biologiques, nous permettant ainsi de prédire les conditions requises pour atteindre une deutération cible.

L'utilisation de la spectrométrie de masse, pour les protéines, nous a également permis de mettre en évidence un effet indésirable de la deutération partielle aléatoire, laquelle génère des niveaux de marquages variables de façon stochastique.

Deutération spécifique *in vivo*

Nous avons développé une méthode permettant de produire, dans une même culture bactérienne, des classes de molécules avec des marquages différents. Nous avons supplémenté un milieu de cultures totalement deutéré avec des acides aminés hydrogénés, nous permettant ainsi de produire à la fois des lipides et acides nucléiques totalement deutérés tout en maintenant un bas niveau de marquage dans les protéines.

Nous avons testé différentes conditions de cultures dans le but d'améliorer le marquage spécifique des protéines. Nous avons ainsi pu mettre en évidence, le fait que la concentration d'acides aminés hydrogénés joue un rôle primordial dans l'efficacité du marquage.

Perspectives

Ce travail a pour but d'étendre les capacités de deutération biomoléculaires *in vivo* dans le but d'optimiser les expériences biologiques impliquant la diffusion ou réflexion de neutrons.

La déterminations des conditions de cultures permettant de deutérer partiellement les molécules biologique permettra dans le futur de mettre en place des expériences complexes exploitant au maximum les possibilités offertes par la variation de contraste.

Nous avons également mis en place une nouvelle méthode qui pourrait avoir des applications variées et uniques pour l'utilisation de la diffusion de neutrons à petits angles, notamment pour les systèmes complexes à plusieurs composants. Nous proposons une série d'expérience qui pourrait contribuer à améliorer la spécificité de marquage. Nous suggérons notamment d'utiliser des mélange d'acides aminés spécifiques, d'analyser plus en profondeur l'aspect temporel du marquage ou encore de modifier certains mécanismes ou voies de biosynthèse biologiques qui pourraient être responsable de l'intégration indésirable d'acides aminés non marqués. Cette étude pourrait également être étendue à d'autres organismes d'expression (levures, micro-algues...) ou à d'autre biomolécules en utilisant d'autres précurseurs.

In this chapter we investigated the correlation between the deuterated culture conditions and the level of deuteration achieved in proteins, lipids and nucleic acids. To do so, we analysed either the contrast match of molecules by using Small Angle Neutron Scattering (SANS), quantified the amount of hydrogen in the molecules by 1D Nuclear Magnetic Resonance (NMR) or directly measured the mass by Mass Spectrometry (MS). The possibility to achieve selective *in vivo* deuteration has been investigated so that proteins can be specifically labelled while keeping the labelling of nucleic acids and lipids to a minimum.

4.1 Partial deuteration of biomolecules

4.1.1 Protein deuteration

The maltose-binding protein has been expressed in BL21(DE3) cultivated in ENFORS medium at different deuteration regimes:

1. 0% D_2O
2. 25% D_2O
3. 50% D_2O
4. 75% D_2O
5. 100% D_2O + 100% hydrogenated glycerol (later referred to as 100% D_2O)
6. 100% D_2O + 50% hydrogenated + 50% deuterated glycerol (d_8-gly, later referred to as "100%D_2O 50% deuterated glycerol"
7. 100% D_2O + 100% deuterated glycerol (d_8-gly)

The protein has been produced and purified as described in subsection 5.2.5. The purity was assessed by SDS-PAGE (Figure 4.1). The different purification steps required to purify MBP are shown on Figure 4.1 panel A. MBP (42 kDa) migrated between the 37 and 50 kDa band of the molecular weight ladder on a 12% SDS-PAGE. In the input we see MBP was efficiently over-expressed as it was the majority band. The flow-through showed that most of the bacterial protein are not retained by the cobalt resin but that MBP is efficiently bound as it cannot be detected. Non-specific binding was further minimized by extensively washing the resin with 5 mM imidazole, and contaminants could not be detected by PAGE. The elution was performed at 500 mM imidazole and about two resin volume were sufficient to elute about 90% of the protein.

Figure 4.1 Panel B shows that the same procedure could be efficiently applied to MBP expressed in partially or fully deuterated conditions. When high quantities of pure protein were loaded on PAGE (such as for MBP3, 4, 5, 6 and 7), degradation products of MBP were detected. For MBP 4 and 5, the degradation products represented a significant fraction of the total protein content. The purification was repeated from the same cell paste and Panel C shows that these degradation products were mostly avoided by ensuring proper cooling during the sonication steps.

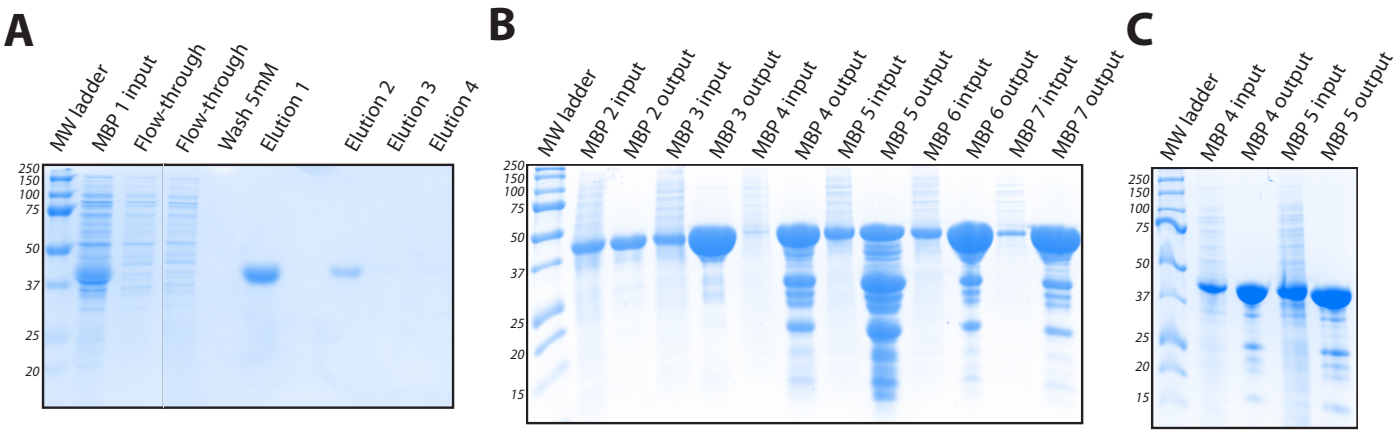


FIGURE 4.1: Coomassie stained 12% PAGE. **A**: Purification steps of MBP (1=0% D_2O) by cobalt IMAC **B**: Purification of MBP grown in different culture conditions (2=25% D_2O, 3=50% D_2O, 4=75% D_2O, 5=100% D_2O, 6=100% D_2O + 50% deuterated glycerol (d_8-gly), 7=100% D_2O + 100% deuterated glycerol (d_8-gly)). **C**: Purification of MBP 4 and 5 repeated to overcome the observed degradation

MBP was therefore consistently and efficiently produced from *E. coli* cells, regardless of the deuterium content of the culture media.Tens of milligrams were produced from 250 ml cultures for all deuteration conditions.

Determination of the neutron contrast match point

Part of the purified protein was buffer exchanged against a 100% deuterated buffer (20 mM MES pH 6.2) using a PD-10 (Biorad) column and the concentration measured with a nano-scale spectrophotometer at 280 nm. Contrast variation series were prepared by mixing MBP in hydrogenated and deuterated buffer in the proportion required to achieve 0%, 25%, 50%, 75% and 100% deuteration of the solvent. The samples were measured on the D22 instrument at the Institut Laue-Langevin using 1 mm quartz cuvettes at detector distances of 2.8 and 7.8 m and using a 6 Å neutron wavelength.

Figure 4.2 shows the contrast match point plot for all different growth conditions. In all cases, the data could be fitted with a linear regression and a coefficient of determination (R^2) superior to 99%. The errors were inferior to 6%, except for the data points that were close to the calculated contrast match point of the protein. The data were normalized by the concentration of the sample and the similarity of the slopes indicates that the concentration has been accurately measured so that the $\sqrt{I_{sub}/C}$ data could be interpreted confidently. In the case of the "100% D_2O + 100% d_8-gly" sample condition, the slope is lower than for the other samples, suggesting that the concentration was overestimated. However, since this overestimation was the same for all of the samples in this contrast series, the contrast match point could still be interpreted and is in agreement with that published previously(Leiting *et al.*, 1998) ; this condition predicts

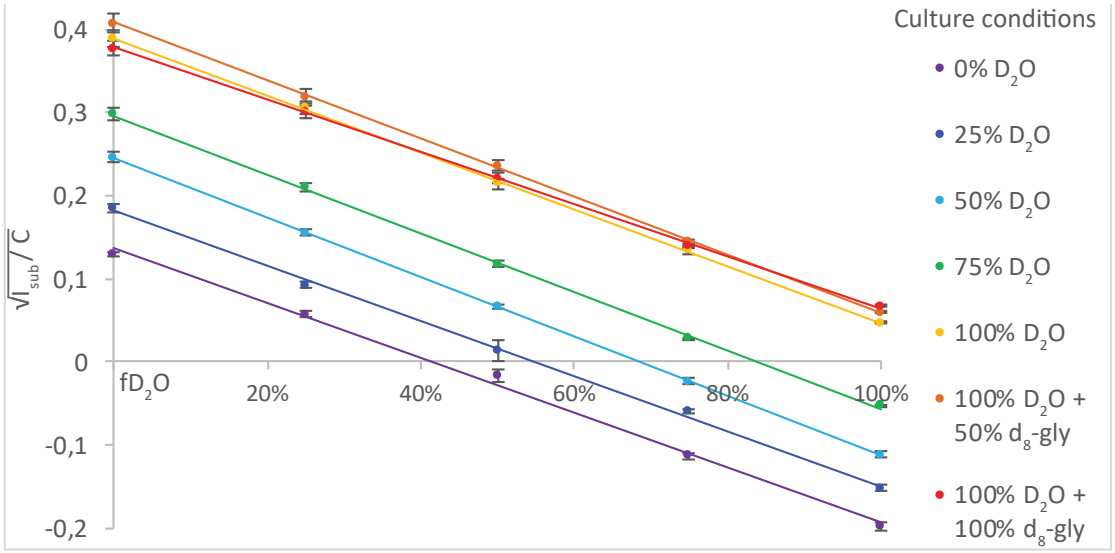


FIGURE 4.2: Contrast variation analysis of MBP produced in different growth deuterated conditions. $\sqrt{I_{sub}/C}$ plotted against the D_2O fraction in the buffer. The errors are propagated from the radial averaging errors.

100% deuteration, which corresponds to a contrast match point of around 120%. The data are summarized in Table 4.1. The calculation of deuteration levels associated with these data is given in section 4.2. The contrast match point (CMP) and errors are also summarized in Table 4.1.

Determination of the non-exchangeable deuterium content by ESI-TOF mass spectrometry

In parallel with the determination of the contrast match point previously discussed, the protein produced under each culture condition solubilized in H_2O solvent was analysed by ElectroSpray Ionisation-Time of flight Mass Spectrometry coupled to liquid chromatography (ESI-TOF LC-MS) by diluting them to 5 µM in TriFluoro Acetic acid (TFA).

Figure 4.3 panel A shows the mass spectra collected for each deuterated growth condition and displays an increasing mass of the model protein as the deuterium level in the medium increases. The peaks were integrated and the highest intensity peak was used for the calculation of the molecular weight of the protein. The measured molecular weight is converted into deuteration level (see Equation 5.14), the width of the peak ($\Delta mass$) converted into variation of deuterium content ($\Delta\%$ *deuteration*) and the mass distribution factor calculated to quantify the "sharpness" of the first peak. The results are summarized in Table 4.1.

The protein produced in 0% D_2O medium has a mass of 42360.9 Da which corresponds exactly to the theoretical molecular weight of MBP calculated from the sequence using average isotopic distribution (takes into account the natural abundance of hydrogen, carbon and nitrogen isotopes) and considering the N-terminal methionine has been cleaved off. The latter consideration is a well known effect that is imputed to the

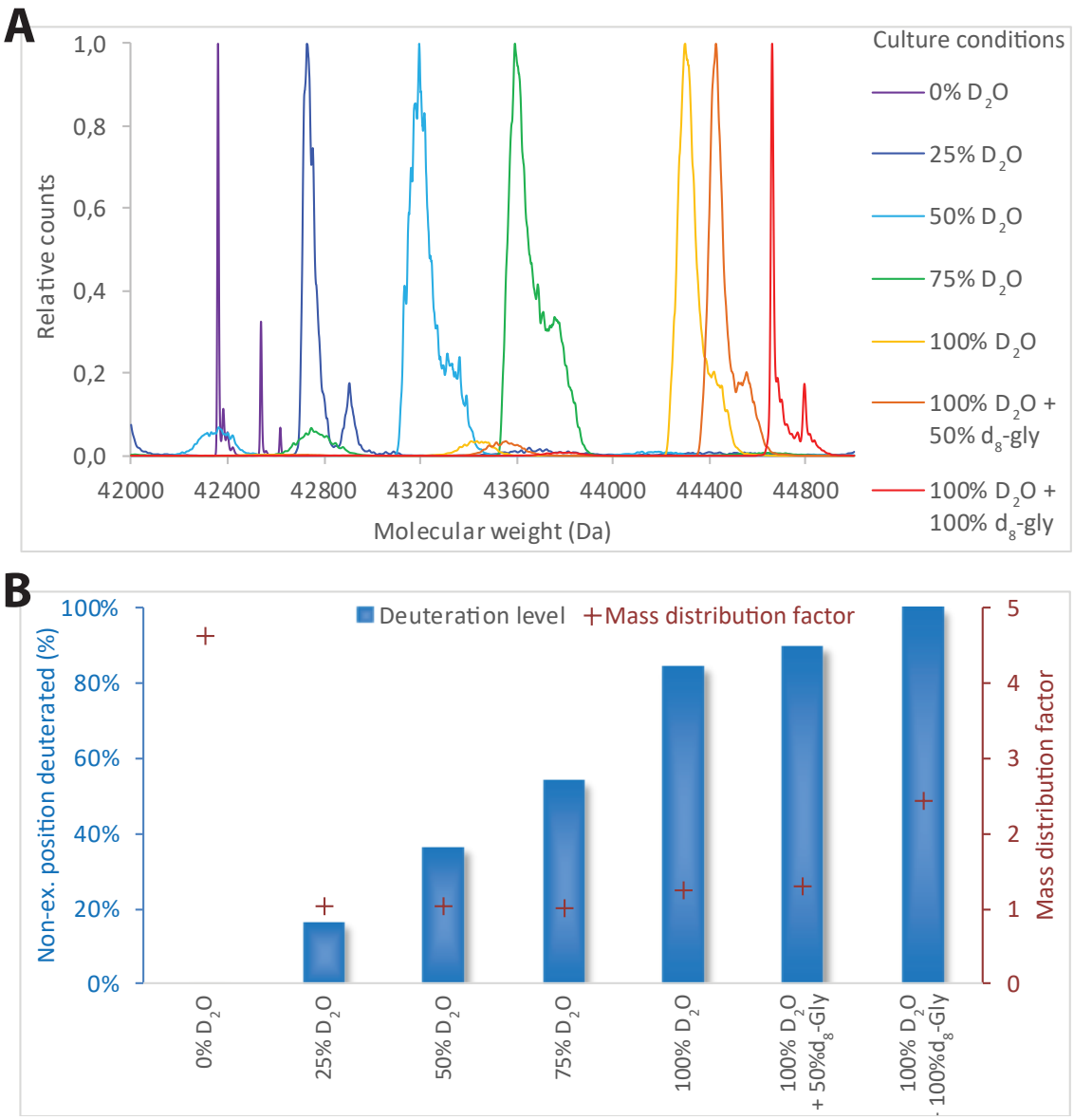


FIGURE 4.3: LC-MS (ESI-TOF) performed on MBP produced in different deuterated conditions. A: Normalized mass spectra. B: Calculated molecular weight converted to deuterium content and associated mass distribution factors.

methionine aminopeptidase and which is known to be particularly important when the second amino acid is a glycine, as is the case for the MBP construct. It is interesting to note that another peak is observable at higher mass, that could be identified as the uncleaved methionine MBP. However, the measured mass (42539.0 Da) for this is slightly higher than the expected isotope averaged and uncleaved MBP (42491.16). In all cases, the spectra displayed the same two peaks and the calculated molecular masses of the main peak were linked to their deuteration level using Equation 5.14.

Figure 4.3 panel B shows the calculated deuteration levels and the mass distribution factors. As expected, the deuteration level of the protein follows the deuteration content of the medium. However, this relation is not linear, which suggests that the hydrogen is integrated more efficiently in proteins than deuterium. We also see that glycerol (the carbon source at $5\,\mathrm{g\,l^{-1}}$) provides about 15% of the hydrogen as only 84.3% deuteration could be achieved using D_2O solvent alone. In a similar way noted for the solvent, it seems that there is a selectivity for hydrogenated glycerol as the relation between the percentage of deuterated glycerol and the deuteration of the protein is not linear. Additional data points would be required to unequivocally confirm this effect.

With regards to the shape of the peaks, we observed a clear and significant trend that can be quantified by the mass distribution factor. If the "0% D_2O" (mass distribution factor = 4,62) and "100% D_2O + 100% D_2O d_8-glycerol" (mass distribution factor = 2,45) media displayed rather sharp peaks, all the other peaks were significantly wider and their mass distribution factors did not exceed 1,31. The sharpness of the peak is correlated to the homogeneity of the samples and a wider peak indicates a wider distribution of masses. In our case, the wideness of the peaks was attributed to a wider distribution of deuteration level that arose from random fractional deuteration and stochastic drift. Since both deuterium and hydrogen were available in the medium, amino acids that are produced are likely to integrate H/D in ratio that did not necessarily represent the medium's composition. For example, since glycine has only 2 non-exchangeable hydrogen atoms, if cultured in a medium containing 25% deuterium, it would have a 56,25% chance to be fully hydrogenated, a 37,50% chance to be 50% deuterated and a 6,25% chances to be fully deuterated. Glycine will not represent the deuteration level of the solvent and will then introduce a bias in the deuteration level of a protein. When taken into account for all amino acids, this explained that the sample that have not been cultivated in fully hydrogen or deuterium conditions displayed wider peaks with a Δmass and hence Δdeuteration that represent the extent of stochastic drift of random fractional deuteration.

Finally, using on-line calculators (MULCh, Whitten *et al.*, 2008), the percentage of deuteration calculated by mass spectrometry was used to calculate the corresponding contrast match point, considering 75% of exchangeable hydrogen are accessible to the solvent Figure 4.8. The results were in reasonable agreement with the experimentally determined contrast match point except for the sample prepared in 100% D_2O and 100% D_2O supplemented with 50% deuterated glycerol. These samples displayed a high heterogeneity (random fractional deuteration) and rather asymmetric peaks, probably contributing to the observed difference that nevertheless stays within the $\Delta\%deuteretaion$ (see Table 4.1).

TABLE 4.1: Summary of Small Angle Neutron Scattering (SANS) and mass spectrometry (ESI-TOF) data for MBP

Culture conditions	**0% D_2O**	**25% D_2O**	**50% D_2O**	**75% D_2O**
Measured CMP$_{SANS}$	41,0%	55,3%	68,7%	84,4%
% Error	2,1%	2,9%	2,9%	3,2%
Measured mass$_{MS}$	42360,9	42729,8	43195,7	43600,0
% deuteration$_{(non-ex)}$	0,2%	16,1%	36,3%	53,8%
Δ mass	93,5	121,3	182,5	222
Δ % deuteration	4,0%	5,3%	7.9%	9,6%
Mass distrib factor	4,62	1,04	1,05	1,02
Theoretical CMP	41,2%	53,7%	69,5%	83,2%

Culture conditions	**100% D_2O**	**100% D_2O + 50% d_8-glycerol**	**100% D_2O + 100% d_8-glycerol**
Measured CMP$_{SANS}$	113,8%	117,0%	121,7%
% Error	5,3%	5,6%	6,2%
Measured mass$_{MS}$	44302,9	44425,4	44659,6
% deuteration$_{(non-ex)}$	84,3%	89,6%	99,7%
Δ mass	247,5	196	111,5
Δ % deuteration	10,7%	8,5%	4,8%
Mass distrib factor	1,25	1,31	2,45
Theoretical CMP	107,0%	111,2%	119,1%

4.1.2 Lipid deuteration

Lipids have been extracted from *E. coli* total membranes prepared from the same culture as the one used in subsection 4.1.1 :

1. 0% D_2O
2. 25% D_2O
3. 50% D_2O
4. 75% D_2O
5. 100% D_2O + 100% hydrogenated glycerol (later referred to as 100% D_2O)
6. 100% D_2O + 50% hydrogenated + 50% deuterated glycerol (d_8-gly, later referred to as "100%D_2O 50% deuterated glycerol"
7. 100% D_2O + 100% deuterated glycerol (d_8-gly)

The lipids were purified and extracted as described in subsection 5.2.7 then analysed by thin layer chromatography.

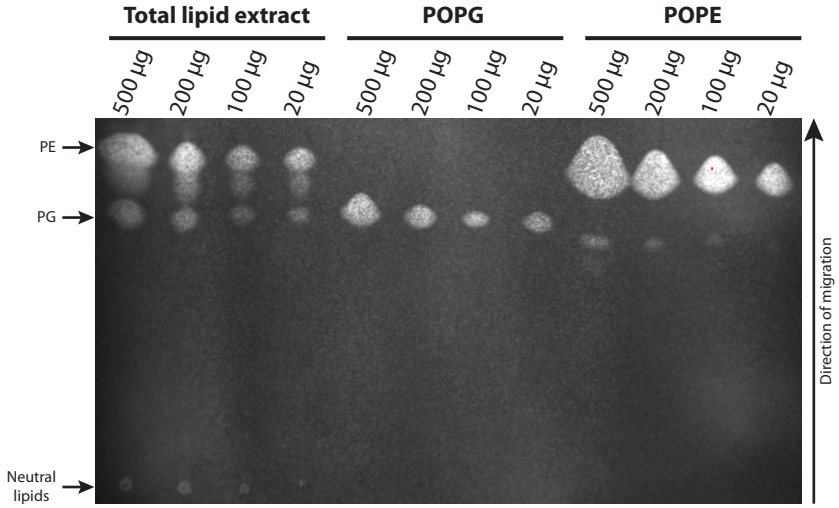


FIGURE 4.4: Thin Layer Chromatography of total *E. coli* lipids using commercial POPE and POPG as standards.

Figure 4.4 shows that, as expected from the typical composition of bacterial membranes, we extracted the two major classes of bacterial lipids : phophatidylglycerol (PG) which migrated in a similar way as POPG and phosphatidylethanolamine (PE) which migrated in a similar way as POPE. Even though this analysis was not aimed at being quantitative, the observed ratio of PE:PG was 72:20, which was in good agreement with the typical ratio usually observed in *E. coli* (75:20 Morein *et al.*, 1996). We also observed that, for the total lipid extract, a small fraction of the lipids did not migrate and could be attributed to neutral *E. coli* lipids that do not migrate with this mobile phase.

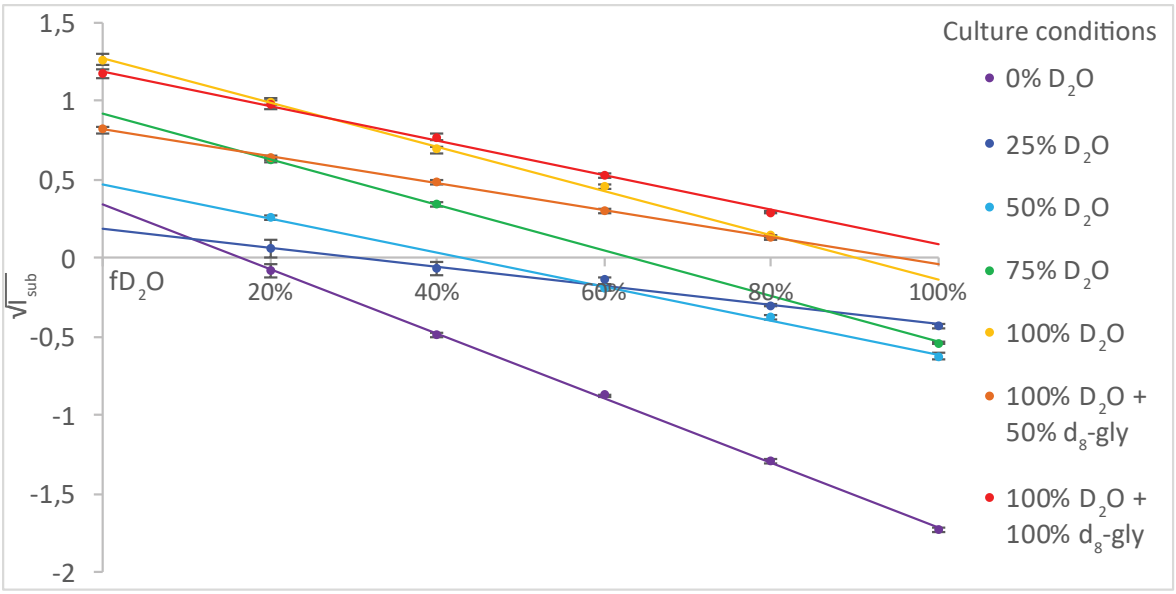


FIGURE 4.5: Contrast variation experiment on *E. coli* total lipid extract produced in different deuterated conditions. $\sqrt{I_{sub}}$ is plotted against the D_2O fraction in the buffer. The errors are propagated from the radial averaging errors.

Determination of the neutron contrast match point

The total *E. coli* lipid extracts were resuspended in either hydrogenated buffer (for the sample produced in 100% D_2O with or without deuterated glycerol) or deuterated buffers (produce in 0 to 75% D_2O) and made into liposome by extrusion. For lipids, the concentration could not be accurately measured thus the data were not normalized. The samples to be used in contrast variation studies were prepared by dilution of the liposome solutions with the relevant H_2O/D_2O buffer to achieve the contrast variation series at 0%, 20%, 40%, 60%, 80% D_2O (for lipids produced in 100% D_2O with or without deuterated glycerol) or 20%, 40%, 60%, 80%, 100% (for other samples). Data were collected on the D22 instrument at the Institut Laue-Langevin using 1 mm quartz cuvettes at detector distances of 2.8, 7.8 and 17 m, and a 6 Å neutron wavelength.

Figure 4.5 shows the contrast match point plots of total lipid extract for all different growth conditions. In the case of the sample prepared by culturing in 50% D_2O condition, the buffer subtraction generated a negative value for the point collected at 40% D_2O. Due to its proximity to the calculated contrast match point and the high errors, we omitted this data point. The sample grown in 75% D_2O medium suffered from the same problem, the data measured at 60 and 80% (which are also close to the later calculated contrast match point) were therefore omitted.

Taking into account the previous remarks, linear regressions fitted the data with a coefficient of determination (R^2) superior to 98%. The propagated errors for each point are overall higher than those of proteins; this is explained by the fact that these data were based on measurements performed at 17 m with incidentally a lower neutron count. When the contrast is higher than 20%, the errors were bellow 3%.

TABLE 4.2: Summary of SANS data for lipids

Culture conditions	0% D_2O	25% D_2O	50% D_2O	75% D_2O
Calculated CMP	16,8%	34,8%	43,4%	63,4%
% Error	1,8%	11,6%	4,3%	3,3%

Culture conditions	100% D_2O	100% D_2O + 50% d_8-glycerol	100% D_2O + 100% d_8-glycerol
Calculated CMP	90,7%	96,0%	106,2%
% Error	5,2%	6,8%	6,8%

When compared to the results obtained for proteins in Figure 4.2, we note that the slopes were significantly different between the samples. This is due to the fact that in the case of lipids, data could not be normalized for concentration. The contrast match point could however be confidently calculated for each sample as the experiment design ensured identical concentrations for all of the samples within a given contrast variation series. The contrast match points and the associated errors are summarized in Table 4.2. The calculation of deuteration from this analysis is treated in section 4.2.

Theoretical computation of *E. coli* lipids contrast match point

The values of contrast match measured by SANS have been recorded using total *E.coli* lipid extract, the signal corresponding therefore to an averaged value of all lipids present in the *E.coli* inner and outer membrane.

In order to evaluate the possible deuteration levels, and therefore contrast match point, of each type of lipid, we used our calculated values of deuteration (see Table 4.7) in order to compute the contrast match point of different lipids. Morein *et al.*, 1996 report that the inner and outer *E. coli* membrane are composed essentially of Phosphatidylethanolamine (PE), Phosphatidylglycerol (PG) and Diphosphatidylglycerol (DPG, or cardiolipin). The paper also reports the distribution of lipid tails in terms of length and unsaturations (the proportion were normalized in order to exclude the unidentified tails) (see Table 4.3).

Using MULCh online calculator (Whitten *et al.*, 2008), we evaluated the contrast match point and molecular volume of each head and tail. We considered that the amine functional (in PE) groups were carrying three hydrogen and alcohol group (in PG and DPG) one hydrogen since their pKa (9 and 15 respectively) is above the experimental pH (7.4), but that the phosphate groups were not hydrogenated (pKa=2,12). All carbon bound hydrogen were considered unexchangeable, the nitrogen and oxygen bound hydrogen considered exchangeable and fully accessible to the solvent (see Table 4.4).

We used the published distribution of lipid tails (Morein *et al.*, 1996) in order to calculate the contrast match point for each class of lipid with an averaged lipid tail, for both inner and outer membrane. We then calculated the average match point of inner and outer membranes, considering the distribution of each class of lipid. These

calculation were reproduced for each deuteration level. The results are summarized in Table 4.5.

It is important to not that these calculations were dependent on three assumptions. We firstly had to assume that the deuteration was homogeneous between the head and tails, which might not be the case as their biosynthesis pathways are significantly different. Secondly, we assumed that the type of medium used (minimal medium) and the deuteration had no effect on the relative distribution of head groups and tails in the outer and inner membrane. It has been shown that using a deuterated medium for the growth of yeasts (*P. pastoris*) mostly affects the distribution of lipid tails, with an accumulation of saturated C18 chains and diminution of C16 and unsaturated C18 chains (de Ghellinck *et al.*, 2014). Similar experiments have not been done on *E. coli* and we could therefore not correct for deuterium effect on lipid tail distribution for our calculations. Finally, we assumed that the lipid extract were only composed of this three lipid classes, however, the lipid extraction protocol (Bligh and Dyer, 1959) does not allow to separate them from wild-type lipopolysaccharides (Raetz, 1978).

Since our extract correspond to a total lipid extract, it contains both the lipids of the inner and outer membrane, but may also contain lipopolysaccharides (LPS). This could explain why our measured value of total membrane contrast match point (16,8% D_2O) differs from the computed value (14,6% D_2O). The contrast match point of Ra-LPS has been evaluated to 27,0% (Arunmanee *et al.*, 2016) ; considering the outer membrane corresponds to about 50% of the total lipid extract, we evaluated that the contrast match point of the outer membrane would need to be of 19,0% to explain the experimental value. This would corresponds to a volume contribution of LPS of 35,5% within the OM ($\approx$ 18% of the volume total membrane extract) while Gmeiner and Schlecht, 1980 reports only 7% of B strain *E. coli* lipids are LPS. However, LPS are highly heterogeneous and the contrast match point used by Arunmanee *et al.*, 2016 corresponds to a mutant strain version of LPS. This consideration, together with the observation the Bligh and Dyer method is prone to protein contamination (imperfect phase separation), could explain why our measured value for the hydrogenated membranes differs from the computed value.

Surprisingly, when the deuteration level increases, the measured value becomes smaller than the computed value. To explain this observation, it is to be considered that the purification of lipids have been done with hydrogenated buffer and that the drying has not been done under anoxic conditions. This may have led to uncontrolled oxidation of fatty acids, desaturation or partial degradation of the lipids which would have been accompanied by an uncontrolled hydrogenation of the molecules and therefore diminution of the contrast match point. By considering that the unsaturated lipids may have been saturated with hydrogen, the computed contrast match point could be reduced by 2%, which did not fully explain the difference. It is however important to notice that the calculation of the deuteration levels have been done by considering the sample cultivated in 0% deuterium and 100% deuterium were defined has fully hydrogenated or deuterated respectively. Therefore, if the measurement of the contrast match point might be affected by uncontrolled incorporation of hydrogen in the molecules, our calculation of the deuteration level, being relative, would correct for this effect.

TABLE 4.3: Distribution of head and tails of *E. coli* lipids in inner membrane (IM) and outer membrane (OM)

	Formula	**nb. H exch.**	**V(Å^3)**	**% IM**	**% OM**
PE	C7H12O8NP	3	255,8	75,0	79,0
PG	C8H12O10P	2	288	19,0	17,0
DPG	C13O17P2H16	1	458,1	6,0	4,0
14:00	C14H29	0	379,5	3,9	3,9
14:1c7	C14H27	0	369,2	0,3	0,3
15:0	C15H31	0	406,2	0,4	0,4
16:0	C16H33	0	433	44,9	47,4
16:1c9	C16H31	0	422	33,4	31,3
18:0	C18H37	0	486,5	1,7	1,2
18:1c11	C18H35	0	476,2	15,4	15,6

TABLE 4.4: Contrast match point (in % D_2O) of *E. coli* lipid head and tails

%D	**PE**	**PG**	**DPG**	**14:0**	**14:1c$_7$**	**15:0**	**16:0**	**16:1c$_9$**	**18:0**	**18:1c$_{11}$**
0,0	52,6	48,9	52,4	2,1	4,9	2,2	2,3	4,7	2,4	4,6
20,2	65,5	60,7	62,7	25,3	27,1	25,4	25,4	27,0	25,5	26,9
29,9	71,8	66,3	67,6	36,5	37,8	36,5	36,6	37,7	36,6	37,6
52,1	86,0	79,3	78,9	62,0	62,2	62,0	62,0	62,1	62,0	62,1
82,6	105,6	97,0	94,4	97,0	95,7	97,0	96,9	95,8	96,8	95,8
88,6	109,4	100,5	97,5	103,9	102,3	103,8	103,8	102,4	103,7	102,5
100	116,8	107,2	103,3	117,0	114,8	116,9	116,8	114,9	116,7	115,0

TABLE 4.5: Computed contrast match point (in % D_2O) of *E. coli* averaged lipids and membranes

	PE		**PG**		**DPG**		**Membrane**		
%D	IM	OM	IM	OM	IM	OM	IM	OM	Measured
0	14,6	14,6	14,8	14,7	13,7	13,6	14,6	14,6	16,8±1,8
20,2	35,1	35,1	34,8	34,7	33,8	33,8	34,9	34,9	34,8±11,6
29,9	45,0	45,0	44,4	44,4	43,5	43,5	44,7	44,8	43,4± 4,3
52,1	67,5	67,5	66,3	66,3	65,6	65,6	67,1	67,2	63,4±3,3
82,6	98,5	98,5	96,5	96,5	95,9	96,0	97,8	98,0	90,7±5,2
88,6	104,6	104,6	102,5	102,5	102,0	102,0	103,9	104,0	96,0±6,8
100	116,1	116,1	113,7	113,7	113,3	113,3	115,3	115,5	106,0±6,8

Finally, the high errors associated to these measurements (between 1,8 to 11,6% D_2O fraction) are to be considered and could also partly explain the differences observed with computed values.

4.1.3 Nucleic acid deuteration

The determination of the deuteration level of nucleic acids *in vivo* has been attempted by different approaches :

- Small Angle Neutron Scattering performed on total *E. coli* nucleic acid prepared by phenol-chloroform extraction as described in subsection 5.2.6. However, the sample's scattering was insufficient to perform contrast variation experiments, probably due to the low solubility of the sample that required dilution, and therefore implied low concentration (maximum 0.2 mg).
- MALDI-TOF mass spectrometry performed on enzymatically digested pHSG298 plasmids (see subsection 5.3.4). A combination of two restriction enzymes has proved to be efficient at digesting the plasmid into fragments with a size compatible with MALDI-TOF analyses (see Figure 4.6). However, no signal could be obtained, probably due to the carry-over of sodium during the precipitation which interferes with the ionization process. The test of several matrix and extensive washes after the precipitation of DNA did not resolve the issue.
- 1D-Nuclear Magnetic Resonance measurements were performed on the same samples as used for SANS. The nucleic acids were digested into nucleotides by acid hydrolysis (see section 5.2.6), dried and resuspended in 100% D_2O for the measurement of the hydrogenation levels at the non-exchangeable positions (the exchangeable positions will be deuterated by the solvent).

Determination of the hydrogen content by 1D-Nuclear Magnetic resonance

To assess the reproducibility of the acid hydrolysis a first NMR experiment was performed on two identical samples processed independently. The spectra could be consistently superimposed, validating the hydrolysis protocol (data not shown).

The following deuteration conditions were tested :

1. 0% D_2O
2. 25% D_2O
3. 50% D_2O
4. 75% D_2O
5. 100% D_2O + 100% hydrogenated glycerol (later referred to as 100% D_2O)
6. 100% D_2O + 50% hydrogenated + 50% deuterated glycerol (d_8-gly, later referred to as "100%D_2O 50% deuterated glycerol"

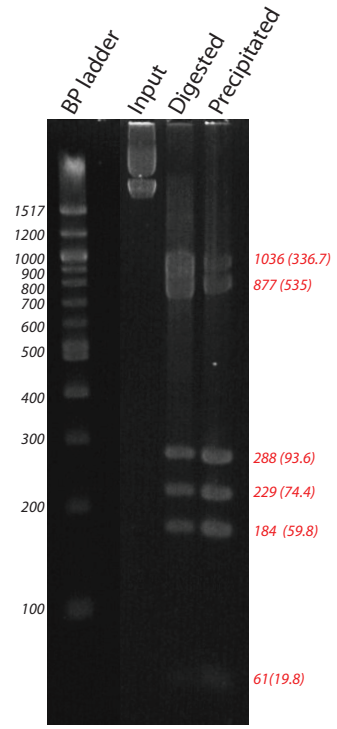


FIGURE 4.6: Enzymatic digestion of pHSG298 with APOI-HF and BTGI monitored on 3% agarose gel stained with SYBR safe 1/7500. The first lane corresponds to the base pair ladder with the bands annotated on the left. The second lane corresponds to undigested plasmid, the third to digested plasmid and the last to digested and precipitated DNA. The red legends correspond to the expected size in base pairs (and kDa) calculated from the sequence of the plasmid.

7. 100% D_2O + 100% deuterated glycerol (d_8-gly)

The 1D hydrogen spectra were acquired for each sample. It is important to note that the data were collected in D_2O to avoid interference from the buffer: this ultimately implies that the hydrogen atoms observed are not exchangeable. The data were normalized for sample concentration (measured at 260 nm).

Figure 4.7 panel A shows the NMR spectra collected for the hydrolysed DNA over the 9.0-5.0 ppm range, which corresponds to the hydrogen atoms of the nitrogenous bases. Because of a strong buffer contribution, hydrogen atoms of the deoxyribose (expected bellow 5.6 ppm) were not consistently detected and were therefore omitted from the analysis. Figure 4.7 panel B and C show that, even if some hydrogen atoms are consistently detected at a defined chemical shift value with superimposed peaks across samples, some peaks are clearly not superimposed and sometimes not even detected (panel B, 0% D_2O). These variations were imputed to the preparation strategy; the HCl concentration was kept constant therefore inducing variation in nucleic acid and buffer concentrations. The peak widening observed for the 25% D_2O sample at 8.1 or 6.3 ppm may indicated incomplete lysis. Depurination of adenosine and guanosine has also been reported when the hydrolysis is performed at high temperature (Adams *et al.*, 2013).

To reduce the sensitivity of the analysis to chemical shift variations, we processed the data in terms of integrated peaks over a range of ppm, ultimately giving us information on the amount of hydrogen detected for given range of chemical shift. Figure 4.7 panel D shows the integrated intensity over the various ranges of ppm that correspond to the identified groups of peaks. Overall, it was observed that for each region the intensity of the signal reduced as the amount of hydrogen provided in the medium was reduced. Unexpectedly however, the sample prepared in 0% D_2O did not follow this trend. Figure 4.7 panel E shows the sum of the calculated integration, ultimately representing the amount of hydrogen detected per sample. With the exception of the 0% D_2O sample, the integrated intensity was consistent with the expected trend.

These values were plotted, the effect of D_2O and deuterated glycerol considered independently by fitting two second order polynomial functions to the data (sample 2-5 and 5-7 respectively). The first allowed the extrapolation of the values for any $\%D_2O$ level in the culture medium and the second allowed the extrapolation of the values for any fraction of deuterated glycerol in 100% D_2O. The 0% D_2O sample was omitted and its theoretical value was estimated from the fit of the other points.

In order to relate the NMR integrated intensity to the percentage of hydrogen in the sample, we considered that the sample prepared at 100% D_2O + 100% d_8-glycerol was fully deuterated and therefore had 0% hydrogen and that the sample prepared in hydrogenated conditions had 0% deuterium. The NMR data are summarized in Table 4.6. The calculation of the deuteration level from these data is treated in section 4.2 and the results summarized in Table 4.7.

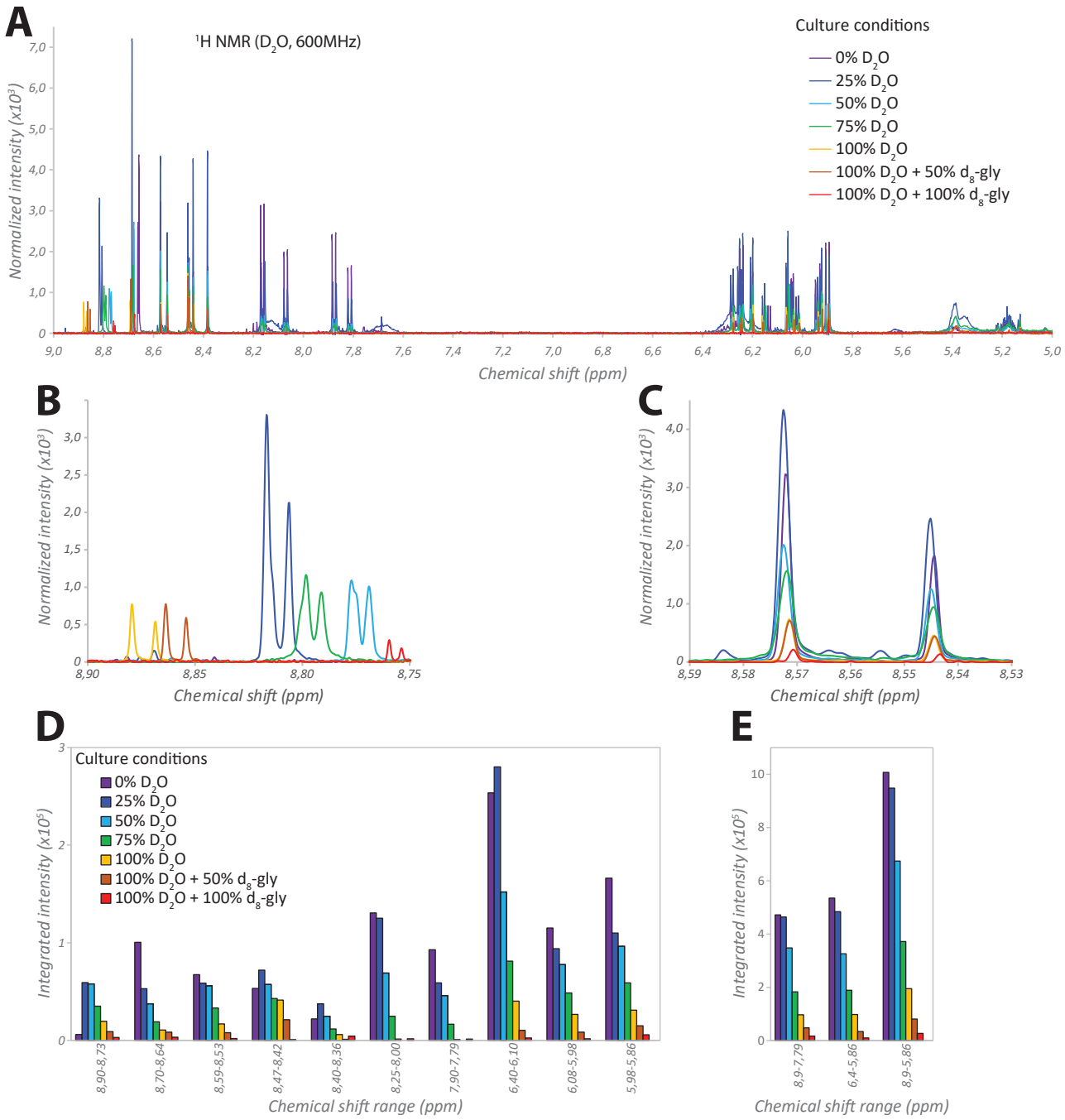


FIGURE 4.7: 1D ^{1}H NMR analysis of hydrolysed total nucleic acid extracted from *E. coli* cells grown in different deuterated conditions. **A:** all spectra represented for the 9-5 ppm range. **B:** Zoom on a smaller range (variation of chemical shift between samples). **C:** Zoom on a small range of ppm (superimposition of spectra for all samples). **D:** Integration of the peaks by group of neighbouring ppm. **E:** Sum of the previously integrated peaks over different ranges of ppm.

TABLE 4.6: Summary of NMR data for nucleic acids

Culture conditions	**0% D_2O**	**0% D_2O** *extrapolated*	**50% D_2O**	**75% D_2O**
Total intensity	1007103	1308900	948251	674284

75% D_2O	**100% D_2O**	**100% D_2O + 50% d_8-glycerol**	**100% D_2O + 100% d_8-glycerol**
371821	194383	79572	26310

TABLE 4.7: Summary of deuteration level as calculated from SANS, MS and NMR results.

Culture	**0% D_2O**	**25% D_2O**	**50% D_2O**	**75% D_2O**
MBP_{SANS}	*0,0±0,7%	17,8±0,9%	34,5±1,5%	53,9±2,0%
MBP_{MS}	0,2%	16,1%	36,3%	53,8%
$Lipid_{SANS}$	*0±1,0%	20,2±6,7%	29,9±3,0%	52,1±2,7%
DNA_{NMR}	*0,0%	28,1%	49,5%	73,1%

Culture	**100% D_2O**	**100% D_2O +50% d_8-glycerol**	**100% D_2O + 100% d_8-glycerol**
MBP_{SANS}	90,2±4,2 %	94,3±4,5%	*100,0±5,1%
MBP_{MS}	84,3%	89,6%	99,7%
$Lipid_{SANS}$	82,6%±4,7%	88,6%±6,2 %	*100,0%±6,3%
DNA_{NMR}	86,9%	95,8%	*100,0%

** : Values fixed to 0 or 100%*

4.2 Differential *in vivo* deuteration

To achieve selective deuteration, we acted on the biosynthesis pathway of proteins and bypassed the amino acid biogenesis by supplementing the medium with amino acids. This strategy was tested and the effect on proteins, lipids and nucleic acids monitored by LC-MS, SANS and 1D-NMR.

The cells used were grown in 100% D_2O + 100% d_8-gly + 5 g l^{-1} of hydrogenated amino acids and pre-incubated for 20 minutes before induction. The analysis was performed on protein (SANS, MS), lipids (SANS) and nucleic acid (1D-NMR), and the contrast match or deuteration content calculated for these sample in the same way as described previously (see subsection 4.1.1, subsection 4.1.2 and section 4.1.3).

To link the previous SANS results back to the deuteration level we considered that the culture condition at 0% D_2O corresponds to 0% deuteration and that the one at 100% D_2O + 100% d_8-glycerol corresponds to 100% deuteration. We fitted two second order polynomial curves to the data, one representing the effect of D_2O (allows to extrapolate the deuteration levels at any percentage of D_2O in the medium) and one representing the effect of deuterated glycerol (allowing the extrapolationof the deuteration levels at any percentage of deuterated glycerol in 100% D_2O medium). The error bars represent the propagated error from the SANS data. The results are plotted in Figure 4.8.

For the 1D NMR results, the same approach was used and it was considered that the sample prepared from 0% D_2O culture was 0% deuterated and that the one cultivated in 100% D_2O + 100% d_8-glycerol corresponds to 100% deuteration. A second order polynomial curve was fitted to extrapolate the deuteration level for any culture condition. The contrast match point associated with these deuteration levels were calculated using the online calculator (MULCh, Whitten *et al.*, 2008), considering DNA or RNA sequences where all nucleotides were equally distributed and that all exchangeable hydrogen atoms were accessible to the solvent. Results are plotted on Figure 4.8.

For the mass spectrometry results, the deuteration level was plotted directly and fitted with a second order polynomial curve to allow for extrapolation of deuteration level at any culture condition. The error bars correspond to the calculated Δ% (see subsection 5.3.5). The contrast match point corresponding to this level of deuteration was calculated online (MULCh, Whitten *et al.*, 2008) considering the sequence of MBP and that 75% of the exchangeable hydrogen atoms are accessible to the solvent (Svergun, M. H. J. Koch, *et al.*, 2013). The results are plotted on Figure 4.8.

For each selective deuteration condition, the contrast match points or deuteration levels were calculated by SANS, mass spectrometry or 1D-NMR, converted to deuteration level (for SANS data) or contrast match point (for MS and NMR data) and the results are summarized in the table of Figure 4.8. It was observed that while the supplementation of fully deuterated medium with hydrogenated amino acids barely affected the level of deuterations of lipids (from 100% deuteration to 95,5%) and nucleic acids (from 100% deuteration to 98,6%) it had a dramatic effect on the deuteration level of proteins (from 100% deuteration to 24,4-29%). As an indication, using the polynomial fit determined for each dataset, we related those selective deuteration values to culture conditions. In the case of lipids and nucleic acids, the selective deuteration conditions

were approximately equivalent to cultivating bacteria in 100% D_2O + 82,5% or 77% of d_8-glycerol, respectively. On the other hand, proteins with such a level of deuteration would be obtained by cultivating bacteria in 37,3-44,7% of D_2O.

It was interesting to note that the deuteration levels calculated from SANS or MS data for proteins differ by 4,6%. This could be attributed to the fact that SANS measures an average signal and is therefore sensitive to contamination. In this case, contaminants could be coming from the purification protocol and are likely to be fully deuterated (as only MBP is over expressed when the hydrogenated amino acid are provided in the medium). If we consider that MBP is 24,4% ($\pm 8,9\%$) deuterated (as calculated by MS) and that the corresponding match point would be 59,1% (see Figure 4.8), a contamination by fully deuterated proteins of 8,5% (in particle's volume fraction) would be enough to raise the measured contrast match point to 64,4% (measured by SANS).

These results show that we were able to selectively label a protein by providing labelled amino acid during its over-expression. $5\,g\,l^{-1}$ of amino acids allowed isotope labelling to a level of 75%, with moderate effect on nucleic acids and lipids (less than 5%).

4.2.1 Optimization of protein selective hydrogenation

We investigated both the effect of pre-incubation time before induction and of concentration on the deuterium content of the model protein MBP. We decided to use deuterated medium (100%D_2O + 100% d_8-glycerol note 100% D later refered to as "100% D") supplemented with hydrogenated casaminoacids (noted aa) under the following conditions :

1. 100% D + $5\,g\,l^{-1}$ aa , 0' pre-incubation
2. 100% D + $5\,g\,l^{-1}$ aa , 20' pre-incubation
3. 100% D + $5\,g\,l^{-1}$ aa , 60' pre-incubation
4. 100% D + $10\,g\,l^{-1}$ aa , 0' pre-incubation
5. 100% D + $10\,g\,l^{-1}$ aa , 20' pre-incubation
6. 100% D + $20\,g\,l^{-1}$ aa , 0' pre-incubation
7. 100% D + $50\,g\,l^{-1}$ aa , 0' pre-incubation

MBP was prepared and measured in the same way as before by LC-MS (ESI-TOF). The data were analyzed in the same way in order to calculate the deuteration level of the protein and to evaluate the extent of random fractional deuteration through the mass distribution factor.

Figure 4.9 panel A shows the zoomed mass spectroscopy data collected. This clearly shows that some culture conditions (especially the culture with $5\,g\,l^{-1}$ aa + 20' pre-incubation) had a very wide distribution of masses and hence deuteration levels. Panel B summarizes the results in which we concluded that both the pre-incubation time of

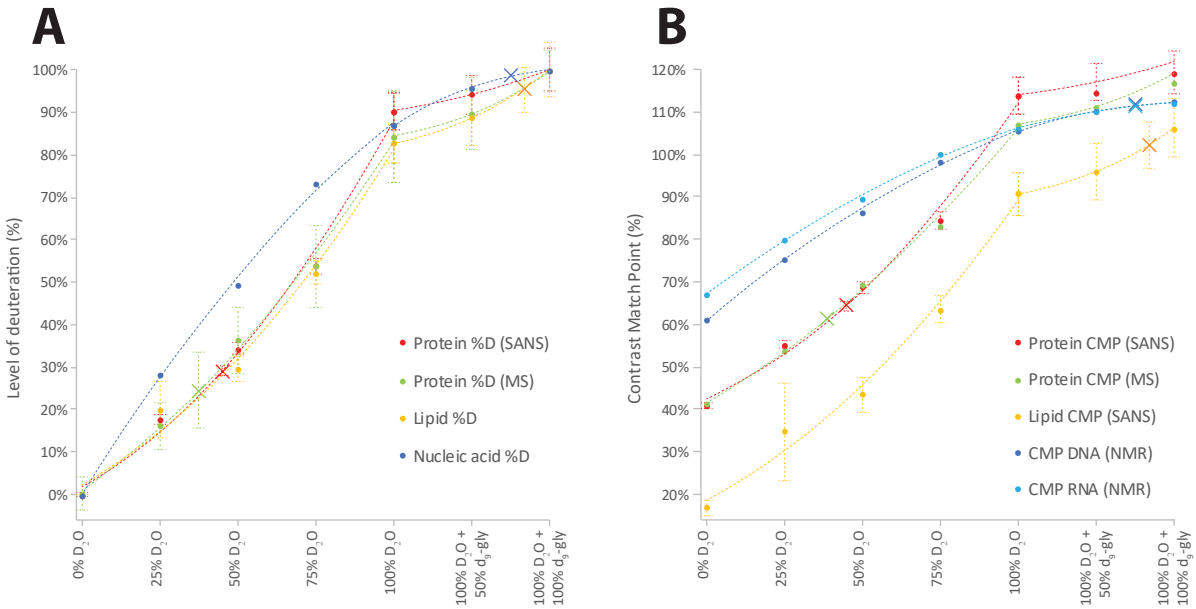


TABLE 4.8

	Deuteration level	CMP	Culture condition
Protein_{SANS}	29,0%±1,1%	**64,4**±2,5%	44,7% D_2O
Protein_{MS}	**24,4%** (8,9Δ%)	61,3	37,3% D_2O
Lipid	95,5±5,3%	**102,1**±5,7%	100% D_2O+82,5% d_8-gly
DNA	**98,6%**	111,5	100% D_2O+77% d_8-gly
RNA	**98,6%**	111,4	100% D_2O+77% d_8-gly

FIGURE 4.8: Deuteration of biomolecules depending on culture conditions. A: Deuteration level was calculated from SANS for protein (in red) and lipids (in orange) and measured directly by mass spectrometry for protein (in green) and 1D NMR for nucleic acids (in blue). B: The contrast match point was measured directly by SANS for protein (in red) and lipids (in orange) or calculated from deuteration level for protein (MS) (in green), DNA (in blue) and RNA (in cyan). The dashed lines represent the second order polynomial fits. The error bars represent propagated errors for SANS data and Δ% for MS data. The crosses represent, for each biomolecule, the culture condition supplemented with hydrogenated amino acids for selective deuteration. The table summarizes the selective deuteration results and links the deuteration levels, contrast match points and corredponding culture conditions that would have yielded the same result in normal deuteration regime. The values measured experimentally are in bold.

amino acids and their concentration had an effect on the deuteration level and distribution of the deuteration levels. By comparing the conditions using $5\,g\,l^{-1}$ of amino acids, we saw that without pre-incubation, the deuteration level was lower (21.6% versus 24.4% and 24%) and the mass distribution factor higher(0.97 versus 0.82 and 0.84). These first data suggested a higher and more homogeneous integration of hydrogen was achieved without pre-incubation of the cultures with the hydrogenated amino acids. However, this trend was not followed when the concentration of amino acids was raised to $10\,g\,l^{-1}$, suggesting that concentration and pre-incubation time effects interacted.

The concentration effect was further investigated by increasing the concentration of hydrogenated amino acid during the culture up to $50\,g\,l^{-1}$ and we observed a slight decrease of the deuteration (from 21.6% with $5\,g\,l^{-1}$ to 17.4% with $50\,g\,l^{-1}$) to reach a level that was very close to what was achieved by culturing in 25%D_2O (16.1%). We also observed a dramatic improvement in the distribution of the deuteration levels as the mass distribution factor went from 0.97 to 1.21, a value that was even higher than that obtained in 25%D_2O (1.04). The theoretical contrast match point was calculated using the online calculator (Biomolecular Scattering Length Density Calculator, ISIS) considering that 75% of the labile hydrogen positions were exchanged (in agreement with the previous SANS data, within the error range). These results are summarized in Table 4.9.

Table 4.9: Summary of LC-MS data for MBP selective deuteration

Culture conditions	**25% D_2O**	**100% D + 5g aa 0'**	**100% D + 5g aa 20'**	**100% D + 5g aa 60'**
Measured mass	42729.82	42854,87	42919,68	42911,7
% deuteration	16,1%	21,6%	24,4%	24,0%
Δ mass	121,3	220	206	227,5
Δ % deuteration	5,3%	9,5%	8,9%	9,9%
Mass distrib factor	1,04	0,97	0,82	0,84
Theoretical CMP	55,3%*	57,3%	59,4%	59,1%

Culture conditions	**100% D + 10g aa 20'**	**100% D + 10g aa 0'**	**100% D + 20g aa 0'**	**100% D + 50g aa 0'**
Measured mass	42801,36	42850,54	42794,56	42758,17
% deuteration	19,2%	21,4%	18,9%	17,4%
Δ mass	143	192,5	181	186
Δ % deuteration	6,2%	8,3%	7,8%	8,1%
Mass distrib factor	0,83	0,84	0,97	1,21
Theoretical CMP	55,5	57,15%	55,3%	54,2%

**Measured in subsection 4.1.1*

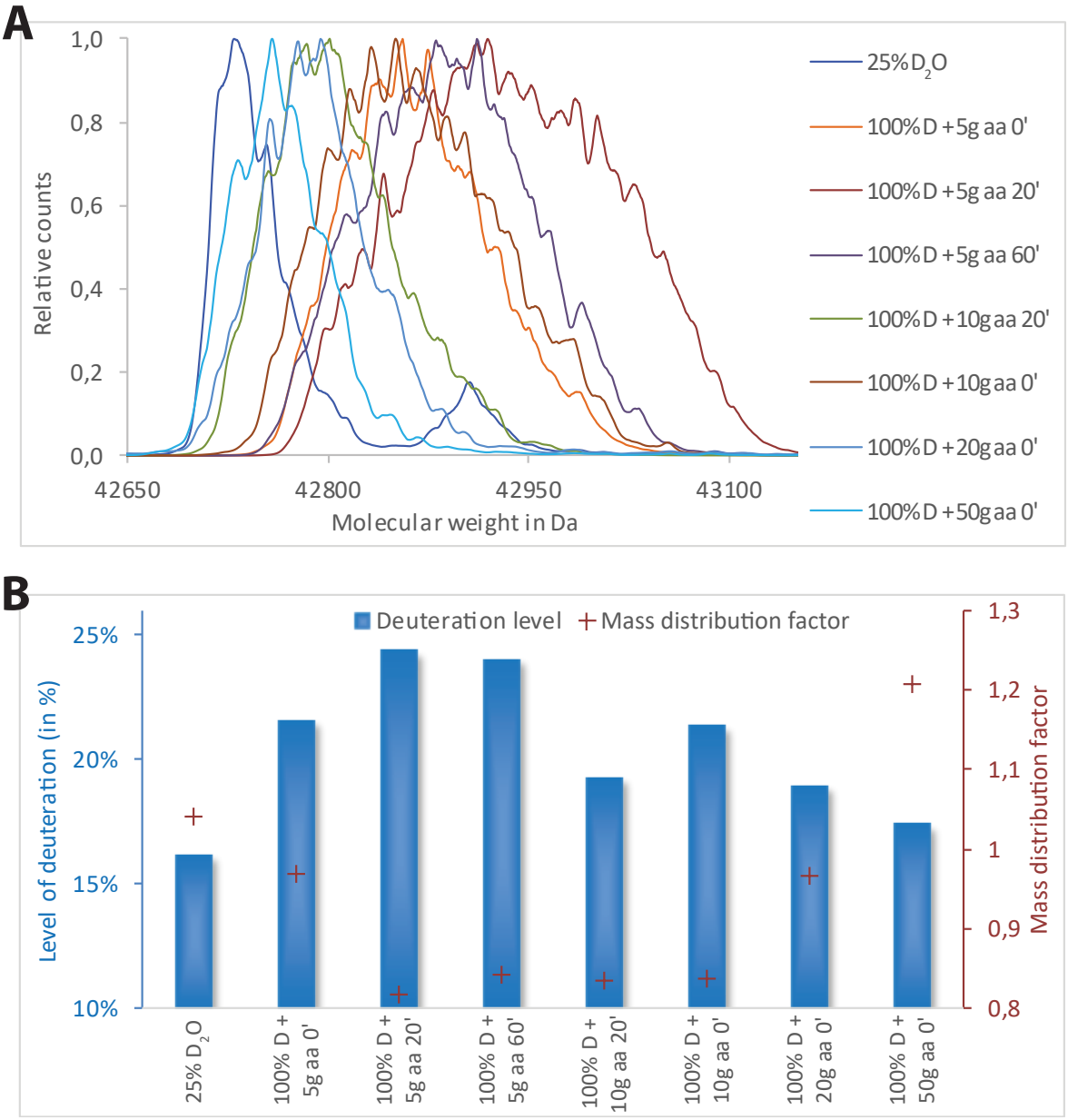


FIGURE 4.9: LC-MS (ESI-TOF) performed on MBP produced in different deuterated conditions. A: Normalized mass spectra. B: Calculated molecular weight converted into deuterium content and associated mass distribution factors. The sample produced in 25%D_2O measured previously is plotted as a comparison.

4.3 Discussion

The need to exploit isotope labeling with biological molecules has became increasingly important as the accessibility of nuclear magnetic resonance and neutron methods for biological systems has developed. The significant difference of neutron scattering length between hydrogen and its heavier isotope deuterium permits a wide range of opportunities such as the use of contrast variation with Small Angle Neutron Scattering, neutron crystallography or dynamic studies. Apart from the reported difference of covalent bond lengths and angles (Artero *et al.*, 2005) which have only moderate effects on protein structure (Fisher *et al.*, 2008) or stability of proteins (Brockwell *et al.*, 2001), hydrogen and deuterium have very similar physical and chemical properties. If higher eukaryotes cannot withstand high levels of deuterium, unicellular organism such as *E. coli* or *S. cerevisae*, which are the most common models used for the production of biological material *in vivo*, can be cultivated in fully deuterated conditions and have therefore been extensively used over the past 20 years (Haertlein *et al.*, 2016) for the isotope labelling of proteins.

Dynamic or crystallographic studies may require perdeuteration (typically for proteins) but contrast variation experiments for the study of multi-components systems are best done where subunits can be partially deuterated separately. This can currently be achieved for proteins as the relation between culture conditions and deuteration levels of proteins have been established (Leiting *et al.*, 1998), making it possible to reach any level of deuteration. However, in the case of other major biological molecules (lipids and nucleic acids), this relation had yet to be established in details.

4.3.1 The prediction of biomolecule deuteration is not linear

In first place, we repeated the work of Leiting *et al.*, 1998 and analysed the level of deuteration that can be achieved in different culture conditions by mass-spectrometry and small-angle neutron scattering. We therefore obtained direct access to both the deuteration level of non-exchangeable hydrogen and also the experimental contrast match point. We obtained very similar results to those previously published and observe that only 84-90% of non-exchangeable position can be deuterated through the use of deuterated solvent alone. Supplementing the medium with a deuterated carbon source allowed us to achieve 99.7% deuteration. It was also particularly interesting to notice that, as previously observed, the deuteration of protein did not evolve linearly with the deuteration of the medium. This observation was attributed to kinetic isotope effect that has been observed and discussed before (Leiting *et al.*, 1998; Pinchuk and Lichtenberg, 2017; Hochuli *et al.*, 2000).

The same experiment was performed with total *E.coli* lipid extract and we were able to establish a direct relationship between the culture condition and the deuteration level. The contrast match point was converted to percentage of deuterium by assuming the extreme values corresponded to 0% and 100% deuteration. This relation was very similar to that observed for proteins where only 82.6% of deuteration was achieved by the solvent alone. Unsurprisingly, a deuterated carbon source (here glycerol) was required

to achieve full deuteration as acetyl-CoA is the main precursor of lipid biosynthesis and is produced from carbohydrates oxidation. In a similar way to proteins, the relation between the supply of deuterium and its integration into lipids was not linear, which again indicates that hydrogen is preferred over deuterium during the biosynthesis of lipids. Using these deuteration levels, we calculated the theoretical contrast match point of each type of lipid head, tail, as well as an averaged value for each class of lipid and inner or outer *E. coli* membrane. It is important to note that this study takes into account total *E. coli* lipid extract ; the results obtained came from the average of all classes of lipids, regardless of their fatty acid chain length and unsaturations. As discussed previously (see section 4.1.2), the differences observed between computed and measured values could be attributed to SANS measurement errors on the CMP, the presence of lipopolysaccharides or protein contaminants and isotopic effect on the distribution of lipids. It would therefore be of great interest to repeat the experiments and analyse the lipid extracts by gas chromatography coupled to mass spectrometry : this approach would not only allow to calculate accurately the level of deuteration of each class of lipids (including labelling of head and tail), but also study the isotopic effect of deuterium on the distribution of lipid heads and tails in *E. coli*, as previously analyzed in yeast (de Ghellinck *et al.*, 2014).

Finally, the relation was investigated for nucleic acids. After unsuccessful attempts with small-angle neutron scattering and mass spectrometry, we made use of nuclear magnetic resonance in order to evaluate the number of non-exchangeable hydrogen atoms that were present in hydrolysed nucleic acids solubilized in D_2O. We were able to infer the deuteration level and observed that full deuteration was also dependent on the presence of deuterated carbon source. In order to convert these values to contrast match point, different factors were to be considered. First, the nucleic acid purification protocol used was meant to recover both DNA and RNA. However, the experiment was not done under RNAse-free conditions so we did not expect RNA in our final sample. Furthermore, the acid hydrolysis protocol has been reported to favour depurination reactions (Adams *et al.*, 2013), which might have affected the final composition of the sample. Since those factors could not be controlled for this experiment, we decided to consider the final sample was solely composed by DNA with an equimolar distribution of the four nucleotides, an approximation that we considered reasonable as the variations that may arise from false assumptions would be within experimental error. Finally, we were surprised to observe that the relation between deuterium in the medium and in the molecule is not linear, but would in this case indicate that deuterium is preferred over hydrogen during nucleic acid biogenesis.

Considering proteins and lipids, the preference for hydrogen over deuterium during the biosynthesis could be explained by kinetic effects: the shorter length of the carbon-deuterium covalent bond as compared to that of carbon-hydrogen, tends to reduce the reaction's rate creating a bias toward hydrogen, which is therefore slightly over-represented in lipids and proteins by comparison with the culture medium. In the case of nucleic acids, the inverse trend was noted. We could not assess confidently whether this effect was actually due to a biochemical preference during synthesis (which seemed unlikely), or if it was an artefactual result that may be attributed to the experimental procedure. The nucleic acids have been submitted to highly acidic conditions in order

to trigger their degradation to single nucleotides. Since the lysis was performed in hydrogenated conditions, we would actually expect an over-representation of hydrogen. In the same way as the carbon-hydrogen/deuterium carbon bond affects the kinetics of a biochemical reaction, it is possible that the efficiency of the lysis of the phosphodiester bond might somehow be negatively affected by the presence of deuterium as this isotope is also know to generally affect kinetics of reactions (Westheimer, 1961). Since NMR signals depend on our capacity to efficiently hydrolyse the nucleic acids, a negative effect of deuterium on the kinetic of the hydrolysis should also favour hydrogenated nucleotides. This surprising result would need to be further investigated in order to confirm a potential preference for deuterium uptake in nucleic acids. Improving the hydrolysis method in order to obtain reliable and consistent measurement would also be of great interest as a higher homogeneity and sample quality would allow us to clearly identify each hydrogen position. Such an improvement would allow us to further understand the effect of deuteration at the atomistic level and may give further insight into the way deuterium is integrated in the biogenesis pathway of nucleic acids.

4.3.2 Random fractional deuteration affects labeling homogeneity

These deuteration protocols were based on the use of fully or partially deuterated media. In the latter case, mass-spectrometry allowed us to determine that proteins produced in such conditions display a certain mass distribution around the maximum peak. In the most extreme case (in 100% D_2O), the differential mass reaches 0.6% of the total protein mass. This observation was a consequence of random fractional deuteration, as explained in Figure 4.1.1. In principle, since both hydrogen and deuterium are present in the medium, they are randomly incorporated into the protein. However, some residues may integrate deuterium at a level that does not reflect the medium's composition.

It is important to note that this effect was only clearly identified for when the molecule was analysed by mass spectrometry since this method is based on the signal of single molecules. NMR and SANS measure an averaged signal that arises from all molecules. If such an effect was observed with protein, it is therefore likely to be applicable to the other biomolecules and two considerations need to be accounted for. First, the size of the molecule is of certain importance when considering this random fractional deuteration associated effect. Indeed, the small molecules, with incidentally less hydrogen, will be statistically more susceptible to stochastic drift and their deuteration level is therefore less likely to represent the deuteration level of the medium. We could then expect lipids to be more affected. Secondly and as a consequence of the first point, the sensitivity of the molecule to random fractional deuteration may be, for protein and DNA, directly linked to their sequence. In the case of protein, the diversity of the amino acids may partly compensate, but in the case of nucleic acids, pyrimidines have an additional non-exchangeable hydrogen positions. We would therefore expect that a sequence that is rich in thymine (or uracilin RNA) and cytosine might be less affected by stochastic drift as their additional hydrogen position reduces the chance of stochastically increasing the representation of one isotope. The reasoning is the same

for proteins as the sequence directly affects the number of non-exchangeable positions and therefore susceptibility to stochastic drift.

To evaluate the extent of these effects, it would be required to adapt the current protocols so that the three types of biomolecules could be studied my mass spectrometry. We could therefore imagine studying the sequence dependency on the random fractional deuteration stochastic drift by analysing model proteins or nucleic acids with sequences containing an extreme number of non-exchangeable positions. If this consideration is conceptually interesting, the implications for small-angle scattering are relatively minor since the signal is averaged over all molecules. Limiting the stochastic drift of deuteration level of a given molecule would not change the contrast match point *per se* but may improve our capacity to produce molecules that can be better matched out.

4.3.3 A new method for selective labelling of biomolecules *in vivo*

In this work, we described a protocol that allowed the differential deuteration of biomolecules *in vivo* ot be carried out. We were able to produce, using the same cultures, lipids and nucleic acids that are close to fully deuterated while the protein integrates only 24,4% of deuterium. This was achieved by supplementing a fully deuterated medium with hydrogenated amino acids. We investigated the effect of concentration, but also pre-incubation time with hydrogenated amino-acids, on the final deuteration level of the protein and managed to maximise the hydrogenation of the protein so that only 17,4% deuterium are found in the protein, while minimizing the effect of random fractional deuteration stochastic drift. In order to further improve these results, several directions could be followed.

Firstly it would important to assess whether the addition of large amounts of hydrogenated amino acids affect the deuteration level of the other classes of biomolecules. We showed that increasing the amount of labelled amino acid improves the efficiency of labelling in proteins. However, it is important to note that by multiplying the concentration of hydrogenated amino acids by ten, we reduced the deuterium by only 7%. Therefore, an adequate balance need to be found in order to maximize the hydrogenation of proteins while limiting the cross-hydrogenation of the other molecules.

The addition of hydrogenated amino acid was sufficient to obtain proteins that are largely hydrogenated. However, we estimated that about 20% of the amino acids in the protein were still deuterated. It is important to mention that the amino acids provided to the cells came from the degradation of casein ; the mix was therefore lacking tryptophane, asparagine, cystein and glutamine, amino acids that represent 9,8% of MBP's sequence. Therefore, we expect that a mix containing all 20 amino acids would partially compensate for this integration of deuterated amino acids. As for the remaining fraction of deuterated amino acids, these may be coming from a pool of free amino acids (typically maintained in the cell during exponential growth that is readily available for protein production or as a source of energy Clark *et al.*, 1972), from a turnover of amino acids obtained by recycling existing proteins (Visick and Clarke,

1995) or simply because deuterated amino acids are still being *de novo* synthesized by the cell due to imperfect feedback inhibition (Berg *et al.*, 2002).

We could therefore design strategies that may prevent the previously mentioned mechanism to occur. With regard to the pool of available protein, we could imagine using a double induction system, with a promoter allowing for the expression of a random protein and our protein of interest under the control of another promoter. Therefore, the first protein could be induced shortly before the protein of interest to deplete the pool of free amino acids. We could also consider inducing starvation before induction to force bacteria to use the free amino acid pool as a source of energy. In both cases however, a reduced efficiency of expression would be expected as in the first case the cell resources would be split between the production of two proteins (unless the expression of the first protein can be turned off) or in the second case the starvation might simply reduce the protein expression efficiency since the bacteria would not be in exponential phase any more.

The recycling of proteins could be limited by using either mutant strains with impaired protein degradation pathways, inducible proteolytic systems (Cameron and Collins, 2014) or by using protease inhibitors (Prouty and Goldberg, 1972). These strategies might reduce the importance of protein recycling but pleiotropic effects (such as toxicity due to other protein accumulation) may impair the efficiency of protein expression.

When supplementing the medium with amino acids, a delay is expected before they can actually be used by the cells. Indeed, amino acids first need to be imported in the cell, activated and loaded on tRNA and a lag phase is to be expected. On the other end, a certain delay is also to be expected when the protein is being induced as the T7 polymerase needs to be produced before transcribing the target gene. Therefore, in order to optimize the system, a complete understanding of these factors could help to prevent the accumulation of deuterated proteins before the hydrogenated amino acids are available.

Finally, the *de novo* synthesis of amino acids could be limited by ensuring that the inhibiting feedback is optimized. We started to investigate this factor by playing on the pre-incubation of the culture with hydrogenated amino acids before induction in order to ensure that the feedback had enough time to be established, but observed that better results were obtained without pre-incubation. It is possible that, due to insufficient amino acid supplies implying they were completely consumed before the arrest of the culture, the amino acid biogenesis pathways were reactivated. Further investigation would be required to address this issue and could also be extended to the use of an amino acid mix whose composition would be controlled according to the sequence of the target protein in order to ensure that all amino acids are provided in sufficient quantities. Alternatively, auxotrophic strains could be used so that only the amino acids that are provided can be used. We would need to make sure that a constant supplementation of the culture with hydrogenated amino acids would not impact negatively the deuteration levels of other biomolecules. Ultimately, this strategy could prove to be particularly costly if deuterated amino acids were to be used.

4.3.4 Prospective applications

First of all, this work established the link between the culture conditions and the level of deuteration of lipids and nucleic acids so that any contrast match point can be achieved. These results will therefore increase the capabilities of neutron scattering experiments by allowing the study of a wider range of biological complexes using optimized contrast. Concerning the random fractional deuteration of lipids, further characterization may complete the current picture, for example by analysing the different classes of lipids independently. Following the work of Maric *et al.*, 2015, we could also consider improving the protocol in order to make the scattering length density of lipids homogeneous (the head and tails have very different scattering length densities).

As compared to small-angle X-ray scattering, small-angle neutron scattering becomes interesting when used with complex systems where the components have different scattering length densities. However the natural contrast that exist between biological molecules may benefit from enhancement. This is usually done by producing the components separately and reassembling them, a strategy that may not be usable for difficult systems. The proposed strategy could therefore allow more contrast to be generated between the components of a systems directly *in vivo*, and purify a complex without requiring reassembly steps. We could for example consider the case of membrane proteins that may become unstable when extracted from the native membrane. Styrene Maleic Acid (SMA) copolymers have been shown to be able to extract proteins from their membrane while conserving surrounding lipids by making a nanodisc (Long *et al.*, 2013), a method that could be applicable to neutron scattering, though the three scattering length density phases (the protein, the lipids and the polymers) would complicate the experiment. It was shown (Maric *et al.*, 2015) that it is possible to generate nanodiscs that can be matched at 100% D_2O by combining a partially deuterated belt protein with selectively deuterated lipids. The approach described here could be used to adapt this type of strategy to generate differential deuteration between the protein and the lipids so that the lipids can be matched together with the copolymers while maximizing the contrast with the protein. This could create unique capabilities for the analysis membrane proteins in their native environment with a maximized efficiency using small-angle neutron scattering.

Deuteration is also of key interest for the study of the effect of protein crowding by neutron scattering. Current strategies rely on the use of a hydrogenated protein whose dynamics are being investigated and a crowding agent (usually a protein or mix of protein that is deuterated). If the highly concentrated crowding agent is supposed to simulate the conditions found in the cytoplasm, one can argue that these conditions do not reliably mimic the intracellular environment that also contain ions and various different proteins. The strategy described here could be used to design new experimental conditions to produce a hydrogenated protein (by short induction) in a fully deuterated environment. Whole cells could therefore be measured in deuterated solvent in order to analyse the dynamics of the protein of interest that would be the only hydrogenated component in the cell.

Finally, the strategy developed in this work could be translated to other organisms commonly used for the production of deuterated material such as other prokaryotes,

cyanobacteria *Anabaena* or halobacteria *halobacterium* and the eukaryotes *Saccharomyces cerevisae* or *Pichia pastoris* (Haertlein *et al.*, 2016).

Our protocol for differential deuteration *in vivo* together with the determination of the culture conditions permitting to reliably predict deuteration levels of biomolecules contributes to the enrichment of the toolbox provided by *in vivo* labeling strategies that have been developed during the past years (Haertlein *et al.*, 2016) and will contribute to expanding the capabilities offered by neutron scattering methods in biology.

Chapter 5

Material and methods

5.1 Molecular biology and cell culture

Culture media

The culture media were used and prepared as described in Appendix D.

Transformation

Chemically competent *E.coli* cells were incubated with up to 10% (volume:volume) of plasmid (typically 100 ng µl^{-1}) for 30 minutes at 4 °C. cells were heat shocked for 45 seconds at 42 °C and recovered on ice for 2 minutes. Rich medium (SOC) was added and the cells were incubated at 37 °C for 1-3 hours. Cells were plated on agar medium with relevant antibiotics and incubated at 37 °C.
Electrocompetent *E.coli* cells were mixed with up to 10% (volume:volume) of plasmid and electroporated twice at 1.8 kV in a 0.1 cm cuvette. Rich medium (SOC) was added and the cells were incubated for 1 hour under agitation at 37 °C. Cells were plated on agar medium with relevant antibiotics and incubated at 37 °C.

Transposition reaction for antibiotic resistance substitution

Using the commercial transposition kit (Ez-Tn5, Epicentre©), 6 µl of water, 1 µl of reaction buffer (X10), 1 µl of transposon diluted 1/10, 1 µl of target DNA and 1 µl of transposase enzyme were mixed. The solution was incubated for 2 hours at 37 °C then 1 µl of stop solution was added and the solution was incubated for 10 minutes at 70 °C. The modified DNA was transformed into *E.coli* Top10 competent bacteria and plated on 35 µg ml^{-1} of kanamycin. Single colonies were screened by re-plating each of them on plates supplemented with the new antibiotic and plates with the previous antibiotic. Only the colonies that lost the previous resistance were considered in order to limit the chances of having transposons inserted into the Open Reading Frame (ORF) of the protein of interest.

Strain conservation

Bacterial strains were conserved at −80 °C.
Glycerol Stock : 1 ml of cell culture at $OD_{600} = 1$ was mixed with 700 µl of 80 % glycerol and frozen at −80 °C.
Bead stock : Commercial Microbank© bacterial and fungal preservations system were used. Overnight cell cultures were mixed with the beads, incubated for 1 h at room temperature and the excess medium was discarded before freezing at −80 °C. For deuterated cultures, the conservation medium was partly replaced by deuterated medium.

Culture in flasks for protein expression

The plasmid of interest was transformed into the relevant strain of *E.coli* and plated on LB-agar plate supplemented with the relevant antibiotics. From a single colony, an overnight pre-culture was prepared in 10 ml of medium and was further used to prepare a medium scale pre-culture. Large scale cultures were prepared with multiple 1 l volume in 2 l baffled flasks, inoculated with up to 50ml of pre-culture and incubated under agitation (typically 130 RPM). The expression was typically induced at $OD_{600} = 0.8$. Cells were collected by centrifugation at 12000g for 20 minutes. The expression of protein was monitored by coomassie stained PAGE.

5.1.1 High cell density culture

ENFORS medium was used for *E.coli* high cell density cultures. Bacteria were adapted to minimal (and eventually deuterated) medium by performing up to 6 successive small scale cultures inoculated at $OD_{600} = 0.1$. The culture parameters were computer monitored and controlled: the dissolved O_2(pO_2) was adjusted by stirring and flowing dried air into the medium with pH kept at 6.9 to compensate for metabolism by-products by the addition of 30 % NaOH (NH_3 for hydrogenated cultures), temperature regulated and the carbon source added in batch (from ENFORS supplemented with 12 % glycerol). The control sequence was organized in 3 phases:

- **Batch**: Carbon source in limited concentration was present in the initial medium and the pO_2 was controlled by feedback loops on stirring and air flow.
- **Fed-batch**: When the carbon source was depleted, the reduction of bacterial respiration was detected as a sudden increase of pO_2, which triggers the initiation of the fed batch sequence. A carbon source was then added on a periodic basis (typically every 250 seconds), the quantity being controlled based on the generation time of the bacteria (2 hours in hydrogenated medium, and 4 hours in 85 % deuterated medium). This fed-batch strategy prevented the accumulation of by-products (such as acetate) as a result of a metabolic overflow in presence of large excess of the main carbon source (Haertlein *et al.*, 2016).

- **Induction**: The induction sequence was manually triggered by adding the induction solution to the culture. A negative feedback was implemented on the stirrer to help maintaining a constant pO_2 as the metabolism of the bacteria is reduced by the induction. The feeding strategy was similar to the previous step with a longer interval for the feeding (280-350 seconds) to favour protein expression over cell growth.

The bacteria were collected by centrifugation at 12000g, 4 °C and the cell paste stored at −80 °C.

5.1.2 Isotope labeling

The deuteration of biomolecules was achieved by growing *E.coli* cells in ENFORS medium prepared with the desired fraction of heavy water. In addition, d_8-glycerol was used as a carbon source in order to achieve higher deuteration level. The cells were cultivated for up to six passages in deuterated medium in order to ensure optimal adaptation. The deuterated culture was maintained in a flask or fermenter for high cell density culture.

5.2 Production of biomolecules

5.2.1 Holotranslocon (HTL)

Expression

HTL3 plasmid (Figure 1.9) was transformed into chemically competent BL21(DE3)* or C43(DE3) cells and cultured with 30 µg ml^{-1} kanamycin, 50 µg ml^{-1} of ampicillin and 20 µg ml^{-1} chloramphenicol. Cell cultures were maintained using DYT medium and expression induced with 1mM IPTG and 0.2% of L-arabinose for 3 hours at 37 °C. High cell density cultures were performed with cells adapted to ENFORS medium and eventually 85%D ENFORS medium. The high cell density cultures were performed at 30 °C and induced at $OD_{600} > 15$ for 6 to 18 hours.

Membrane preparation

All the steps were performed at 4-10°C. Cells were resuspended in 5ml of buffer (supplemented with complete EDTA-free© according to recommendation) and eventually 0.5mM of PMSF) per 1g of cell paste. The lysis was performed with either a Constant Systems Ltd Cell-D© at 2kBar or in a M-110L Microfluidizer© at 18kPsi. The broken cells were recovered and centrifuged at 24000g for 20 minutes. The membrane were pelleted at 150000g for 1.5 hour, resuspended in 40% of the previous volume and homogenized in a tissue homogeneizer.

Two affinity steps protocol

All the steps were performed at 4-10°C. Membranes were resuspended in HSGM and solubilized with 1.5% dodecylmaltoside (DDM) for 1.5 hour. The unsolubilized membrane fraction was pelleted at 150000g for at least 30 minutes. The soluble membrane fraction was loaded on a Ni-NTA IMAC column (HisTrap FF© GE Healthcare) and equilibrated with HSGM + 10mM imidazole + 0.1% DDM. Later steps were performed on an Akta Prime© (GE Healthcare) at maximum flow rate of $1\,\mathrm{ml\,min^{-1}}$ in order to monitor the A_{280}nm absorbance. A 10 column volume (CV) wash was carried out with the loading buffer, followed by a high salt wash with one CV of HSGM + 10mM imidazole + 0.1%DDM + 500mM NaCl and a wash for another 5 CV with loading buffer. Another wash with HSGM + 40mM imidazole + 0.1% DDM was performed and the elution was done with HSGM + 300mM imidazole + 0.1% DDM. Buffer was exchanged for Calmodulin Binding Peptide (CBP) binding buffer + 2mM $CaCl_2$ on a PD-10© buffer exchange column (Biorad©) and the solution incubated with CBP resin (Agilent©) overnight under gentle agitation. The resin was washed with CBP buffer + 0.2mM $CaCl_2$ for 10 column volume and the elution was carried out with CBP buffer + 2mM EGTA. At this point, the sample can either be applied directly onto a Size Exclusion Chromatography column (Superdex 200, GE Healthcare®) operated on an Akta Purifier (GE Healthcare©) or a Biorad DuoFlow (Biorad©), or be processed for Amphipol exchange.

Amphipol exchange : From the CBP elution, the buffer was exchanged on a PD10 column for SEC buffer and the concentration monitored. Amphipol A8:35 (Anatrace©) was added to the protein solution at a protein:Amphipol mass ratio of 1:4 and the solution incubated for 2 h under gentle agitation. The detergents were slowly removed by adding Biobeads SM2 (Biorad©, previously hydrated with methanol 15 min, and 3x5 min in water) at a DDM:Biobeads mass ratio of 1:20 and incubated under gentle agitation overnight. A aliquote with neither Biobeads nor Amphipol was kept as a control. The solution was recovered, centrifuged at 100000g and applied on SEC column in SEC buffer.

HSGM : 20mM HEPES-KOH pH 8, 130mM NaCl, 10% glycerol, 5mM $Mg(CH_3CO)_2$
CBP buffer: 50mM HEPES-KOH pH 8, 130mM NaCl, 10% glycerol, 0.03%DDM
SEC buffer: 20mM HEPES-KOH pH 7.5, 130mM NaCl, 10% glycerol, 0.03% DDM.

One step affinity protocol

All the steps were performed at 4-10 °C. Membranes were resuspended in TSG and solubilized with 1.5% dodecylmaltoside (DDM) for 1.5 hour. Unsolubilized membrane fractions were pelleted at 150000g for at least 30 minutes. The membranes were then applied on Ni-NTA IMAC resin equilibrated with TSG + 0.1% DDM. Wash for about 4 CV with TSG + 30mM imidazole + 0.1% DDM and elute with TSG + 330mM imidazole + 0.1% DDM. The sample was applied on Size Exclusion Chromatography column (Superdex 200, GE Healthcare©) operated on an Akta Purifier (GE Healthcare©) or DuoFlow (Biorad©) in TSG + 0.02%DDM.

TSG : 20mM Tris pH8, 130mM NaCl, glycerol 10%

5.2.2 SecYEG

The plasmid Pet99a-SecYEG was transformed into *E.coli* BL21(DE3)* cells, and cultured in DYT supplemented with 100 µg ml^{-1} at 37 °C and induced with 1mM of IPTG. Cells were collected and membrane prepared as described in the protocols. Purification was either done in HSGM in a similar way as HTL (1 Ni-NTA affinity purification step then size exclusion chromatography) or in TSG (following HTL purification, 1 affinity step protocol). SecYEG was concentrated in 100kDa molecular weight cut off concentrators (Amicon©)

5.2.3 SecA

Expression

BL21(DE3)* were transformed with pET5a-SecA-His6 and selected on 100 µg ml^{-1} ampicillin or pET30b-SecA-Δ*cys* and selected on 35 µg ml^{-1} kanamycin. Flask cultures were carried out in DYT medium at 37 °C and expression was induced with 1mM IPTG for 3 hours. High cell density cultures were achieved with cells adapted to 85%D ENFORS medium. The high cell density culture was performed at 30 °C and induced at $OD_{600} > 15$ for 6 to 18 hours. Cells were collected by centrifugation at 12000g for 30 minutes at 4 °C and stored at −80 °C.

Purification

Cells were resuspended in TKM buffer. The lysis was done either in a micro-fluidizer at 18 kψ or by sonication. The lysate was cleared by spinning at 24000g for 20 minutes and loaded on a IMAC Ni-NTA column (His-trap FF, GE HealthCare©) equilibrated with TKM. The column was washed with TKM + 30mM imidazole and the protein eluted at 330mM of imidazole. The peak fraction was loaded on an anion exchange column (Q sepharose, GE Health Cwere©) equilibrated with TKM + 1mM DTT and eluted with a linear gradient from 100mM KCl to 1M KCL over 6 column volumes. Size exclusion chromatography was performed on a Superdex 200 (GE HealthCare©) in TKM + 1mM DTT. Nucleotides can be stripped off from SecA by performing an additional size exclusion chromatography in TKM + 1mM DTT + 50mM EDTA + 20% glycerol (usually performed before size exclusion chromatography in TKM + 1mM DTT).

TKM buffer : 20mM Tris-HCl pH 7.5, 100mM KCl, 2mM $MgCl_2$

5.2.4 SecB

Expression

SecB plasmid was transformed into *E.coli* BL21(DE3) cells and cultured in DYT medium supplemented with100 µg ml^{-1} ampicillin and the expression was induced with 1mM IPTG at 30°C for 2 hours. Cells were collected by centrifugation at 12000g for 30 minutes at 4 °C and stored at −80 °C.

Purification

Cells were resuspended in 20mM Tris-HCl pH 7.6 supplemented with protease inhibitor cocktail (cOmplete EDTA-free©) and benzonase. The lysis was performed by passing the suspension three times through a constant Systems Ltd Cell-D© at 1.1 kbar and unbroken cells were pelleted at 24000g. The supernatant was loaded on an anion exchange column (Q sepharose, GE HealthCare©, 1 ml of resin for 2 g of cell paste) equilibrated with buffer A. The resin was rinsed with buffer A and buffer A + 250mM NaCl. The elution was performed with a gradient from 250 to 600mM NaCl over 15 column volumes. The peak fractions were identified by SDS-PAGE, collected, concentrated if necessary (10kDa molecular weight cut-off ultra filtration device, Centricon©) and applied to size exclusion chromatography column (S75, S200, or superose 6, GE HealthCare©). The peak fractions containing SecB can be concentrated.

Buffer A: 20mM Tris-HCl pH 7.6, 2mM DTT
Buffer B: 20mM Tris-HCl pH 7.6, 2mM DTT + 1M NaCl
SEC buffer: 20mM Tris-HCl pH 7.6, 300mM CH_3CO_2K, 2mM DTT

5.2.5 Maltose Binding Protein (MBP)

Expression

MBP was expressed in flasks with ENFORS medium supplemented with 35 µg ml^{-1} kanamycin at 37°C. Protein expression was induced with 1mM IPTG for two hours under agitation (typically 130 RPM).

Purification of MBP

Cells were resuspended in 30 ml of lysis buffer for 500 ml of culture. Lysis was performed by on-off cycles of sonication (2.5 minutes total on time) and the lysate centrifuged at 16000g for 30 minutes at 4 °C. If required, the cleared lysate was spun at 150000G to pellet the membranes for at least 45 minutes. The supernatant was applied to cobalt resin (Qiagen©) (8-10ml for 500ml of culture) equilibrated with lysis buffer. The flowthrough was reloaded once and the resin was washed with lysis buffer + 5mM imidazole until $A_{280} < 0.1$. Elution was done in lysis buffer + 500mM imidazole and the peak fraction was buffer exchanged on a PD-10 column against MES buffer. The

concentration was measured with a micro-volume spectrophotometer and the purity of the sample monitored by Coomassie stained PAGE and by monitoring the 280/260. Aliquots were flash frozen in liquid nitrogen for later use.

Lysis buffer: 10mM Tris-HCl pH 7.5, 100mM NaCl, 10% glycerol **MES buffer**: 10mM MES pH6.2

5.2.6 Nucleic acids

Production

BL21(DE3) *E.coli* cells were grown in ENFORS medium until they reached stationary phase.

Total Extraction

The cell pellet was resuspended in TE buffer, typically 45ml for 500ml of culture. The lysis was performed at 37 °C with 0.6% SDS and 0.12mg/ml of proteinase K (Sigma Aldrich) for 1h. The solution was homogenized by a quick sonication. An equal volume of phenol (equilibrated Tris-HCl pH 8, Sigma Aldrich©):chloroform (Sigma Aldrich©) was added and the solution homogeneized. The phase separation was triggered by centrifugation, the aqueous phase collected and re-extracted twice with phenol:chloroform 1:1. The leftover phenol was extracted from the aqueous phase three times by adding an equal volume of chloroform. DNA was precipitated by addition of pure ethanol at −20 °C. The DNA was spun at 4 °C and maximum speed then rinsed twice with 70% ethanol. The pellets were dried and the DNA stored at −20 °C. DNA was resuspended in TE buffer by vigorous agitation at 4 °C and the concentration monitored by micro spectrophotometry at 260 nm and the quality of the sample monitored by following the ratio 260/280 (detects phenol or protein contamination, considered satisfying above 1.8) and 260/230 (detects phenol, and carbohydrate, considered satisfying between 2.0 and 2.2). If necessary, the samples were diluted to favour solubility. This protocol was adapted from He, 2011.

TE buffer : 10mM Tris-HCl pH8, 1mM EDTA

Acid hydrolysis

The Feulgen hydrolysis was adapted from Tomonaga *et al.*, 1971. Nucleic acid were incubated with 8.5M of HCl for 30 minutes at 25 °C. The solvent was evaporated under vacuum using a SpeedVac^{TM} system. The nucleotides were resuspended in pure D_2O under agitation at 4 °C overnight and the concentration monitored with a micro volume spectrophotometer at 260 nm.

Plasmid purification

Cells were grown overnight and pelleted by centrifugation at a minimum of 10000g for at least 10 minutes. Plasmids were purified using Qiagen plasmid purification kits of the relevant scale : Mini for transformation (ID : 12123) or Giga for mass spectrometry analysis (ID : 12191).

5.2.7 Total *E.coli* lipids

Production

Lipids were produced together with MBP as mentioned in subsection 5.2.5.

Extraction

Cell paste was resuspended in lysis buffer and the cells disrupted by sonication. The lysate was cleared at 16000g for 30 minutes and membranes were pelleted at 150000g for at least 45 minutes. The membranes were resuspended and homogenized in lysis buffer (5 ml per gram of cell paste) with a homogenizer. One volume of chloroform and 2.2 volumes of methanol were added and steered for 30 minutes at room temperature. One more volume of chloroform and one volume of lysis buffer were added and the sample spun for 10 minutes at 800g, 4 °C to trigger the phase separation. The lower chloroform phase was collected and the aqueous phase extracted two more times with chloroform. All the chloroform fractions were pooled, dried with compressed air and the mass difference between the empty glass vial and the final step with dried lipids was measured on a precision scale. This protocol was adapted from Bligh and Dyer, 1959.

Liposome preparation

Dried lipids were dissolved and resuspended in buffer at a concentration of at least $10\,\mathrm{mg\,ml^{-1}}$. The solution was frozen in liquid nitrogen and thawed at least 6 times and the solubilization was helped by vortexing and pipetting the solution. The non soluble membranes were pelleted by spinning shortly. Unilamelar vesicles were generated by extrusion using the kit from Avanti Polar lipids©, with a 0.2 µm membrane at a temperature of 45 °C (above the gel-fluid lipid phase transition). The solution was passed 41 times through the membrane to ensure the homogeneity of the solution.

Thin Layer Chromatography (TLC)

After extraction and drying, lipids were resuspended in $CHCl_3$ to a final concentration of $100\,\mathrm{\mu g\,ml^{-1}}$. Commercial 2-Oleoyl-1-palmitoyl-sn-glycerol3-phospho-rac-(1 glycerol) (POPG) and 1-palmitoyl-2-oleoyl-sn-glycerol-3-phosphoethanolamine (POPE) were used

as references and resuspended in $CHCl_3$ at a concentration of 100 µg ml^{-1}. Decreasing amount of lipids (500, 200, 100 and 20 µg) were deposited on a Thin Layer Chromatography (TLC) plate (0.25 mm, silica matrix SIL G, Macherey Nagel), and migrated with $CHCl_3$:CH_3OH:H_2O at a volume ratio of 65:25:4. The plate was dried and the lipids revealed using Primuline (Sigma-Aldrich) solubilized in C_3H_6O/H_2O(8:2) to a final mass concentration of 0.005%. The plates were imaged under UV light.

Lysis buffer: 10mM Tris-HCl pH 7.5, 100mM NaCl, 10% glycerol
Liposome buffer: 20mM MOPS, pH7.4 0.5mM EDTA

5.2.8 Biochemistry

SecB/MBP/SecA interaction

MBP was unfolded in 6-8 M urea, 20mM Tris-HCl pH 7.6, 300mM CH_3CO_2K, 2mM DTT and concentrated. Interaction with SecB was triggered by rapid dilution of unfolded MBP into SecB containing solution. If necessary MBP was concentrated and SecB diluted so that the final Urea concentration in the mixture was below 2 M. SecA was added to the complex in equimolar proportion with SecB tetramer. The complexes were analysed by Superdex 200 or Superose 6 size exclusion chromatography (GE HealthCare©). The fractions were analyzed by SDS-PAGE and can be concentrated with 10sikDa cut-off ultra-filtration devices (Centricon©) if necessary.

5.3 Biophysical analyses

5.3.1 Dynamic Light Scattering (DLS) - Static Light Scattering (SLS)

The experiments were performed on a Zetasizer nanoseries S, Malvern instrument, operated in protein mode, at room temperature and considering the composition of the buffer. The measurement was done in quartz cuvettes of 10mm path length using a sample with a concentration of minimum 0.5 mg ml^{-1}. Each data point was measured three times and the results were averaged. The data were treated automatically by the Zetasizer software (v7.11).

Dynamic light scattering

In principle, a polarized laser beam is directed at the sample and the scattered light is recorded by a photomultiplier that projects the image onto the detectors. The time dependent fluctuations in the scattering intensity at a given point from the speckle patterns are integrated by the correlator and correlated to the translational diffusion coefficient (D). This is then related to the hydrodynamic radius $d(H)$ through the

Stokes-Einstein equation:

$$d(H) = \frac{kT}{3\pi\eta D} \tag{5.1}$$

with η the viscosity of the solvent (calculated from its composition), k the Boltzmann's constant and T the absolute temperature (in Kelvin). The fitting of a single exponential allows the estimation of the mean size (z-average diameter in nm) as well as the width of the distribution (considered polydisperse if $\%PD > 20\%$). A multiple exponential fitting gives the distribution of particle sizes and is displayed as a size-intensity plot, that can be converted to a size-number distribution via the Mie theory to give a more realistic view of the solution composition (Malvern, DLS technical note MRK656-01).

Static light scattering

In a similar way, the sample is illuminated with a polarized laser beam and the averaged scattered intensity is measured at different angles. The molecular weight (M_W) and 2^{nd} virial coefficient (A_2) are determined through the Rayleigh equation :

$$\frac{KC}{R_\theta} = \left(\frac{1}{M_W} + 2A_2C\right) P(\theta) \tag{5.2}$$

with C the concentration (in M), R_θ the Rayleigh ratio, $P(\theta)$ the angular dependence of the sample scattering intensity and K as defined by :

$$K = \frac{4\pi^2}{\lambda_o^4 N_A} \left(n_o \frac{dn}{dc}\right)^2 \tag{5.3}$$

with N_A the Avogadro's number, λ_o the laser wavelength, n_o the solvent refractive index and $\frac{dc}{dc}$ the differential refractive index increment with regard to the concentration variations. Both M_W and A_2 can be extracted from a Debye plot and are calculated from the extrapolation at $C = 0$ and from the gradient of the slope respectively (Malvern, DLS technical note MRK656-01).

5.3.2 Size Exclusion Chromatography - Multi Angle Laser Scattering (SEC-MALS)

Principle

The instrumental setup is composed of a pump (L2130, Elite LaChrom) connected to a size exclusion chromatography column, in line with a spectrophotometer (L2400, Elite LaChrom) to monitor the absorbance at 280nm, a multi-angle laser scattering device (Dawn Helios-u, Wyatt) and a refractometer (Optilab T-rEX, Wyatt). The software used for analysis was ASTRA6 (Wyatt Technology).

Given that the wavelength is significantly bigger than the particles under study, the scattering is considered isotropic, the intensity is averaged and related to the concentration and molecular weight of the particles.

$$\Delta LS = \left(\frac{I_\theta}{I_0}\right)_{solution} - \left(\frac{I_\theta}{I_0}\right)_{solvent} = K(dN/dC)^2 M_W C \tag{5.4}$$

The concentration was calculated by refractometry as:

$$C = \frac{\Delta R_I}{\frac{dn}{dc}} \tag{5.5}$$

where C the concentration (in $\mathrm{mg\,ml^{-1}}$), ΔRI the difference in refractive index between the molecule and the buffer and dN/dC isthe specific refractive index increment(in $\mathrm{ml\,g^{-1}}$). The latter is constant for molecules of given composition ($\frac{dn}{dc}_{protein} = 0.185\mathrm{ml\,g^{-1}}$). In the case of composite particles (membrane proteins, detergent solubilized proteins) $\frac{dn}{dc}$ is unknown and depends on the relative volume of each components. In this case, the protein molecular weight can be calculated using the three-detector method, which combine information from the MALS, refractometer and UV absorbance to deconvolute the relative ratio and molecular mass of each component (Slotboom *et al.*, 2008).

Application to membrane protein samples

A volume of 50 µl of sample was loaded at minimum $1\,\mathrm{mg\,ml^{-1}}$ on a gel filtration column (Superdex 200, 10/300) at a flow rate of $1\,\mathrm{ml\,min^{-1}}$. For the three-detector method, $\left(\frac{dN}{dC}\right)_{DDM} = 0.1430$ and $\left(\frac{dN}{dC}\right)_{protein} = 0.1850$ were assumed.

5.3.3 Analytical Ultra Centrifugation (AUC)

Principle

The experiment was done in XLI analytical ultracentrifuge with a Anti-50 rotor (Beckman Coulter) with 12 or 3mm pathlength cells equiped with Saphir windows (Nanolytics). During the experiment, the particles are spun and consequently submitted to the centrifugal, buoyant and frictional forces .

- The centrifugal force (F_c) :

$$F_c = m\omega^2 r = \frac{M}{N_A}\omega^2 r \tag{5.6}$$

  where m is the particle mass, ω the angular velocity, r the distance to the center of rotation, M the molecular mass of the particle and N_A Avogadro's number.

- The buoyant force (F_b) :

$$F_b = -m_0\omega^2 r = -\left(\frac{M}{N_A}\bar{v}\rho^0\right)\omega^2 r \tag{5.7}$$

where m_0 is the mass of displaced solvent, $\bar{v}$ the particle partial specific volume, and ρ^0 the density of the solvent.

- The frictional force (F_d) :

$$F_d = -fu \tag{5.8}$$

where f is the frictional coefficient and u the velocity of the particles.

- When the force equilibrium is reached :

$$F_c + F_b + F_d = 0 \iff m\omega^2 r(1 - m_0) = fu \tag{5.9}$$

The sedimentation coefficient (s) is then defined as :

$$s = \frac{u}{\omega^2 r} = \frac{M(1 - \rho^0\bar{v})}{N_A f} = \frac{M(1 - \rho^0\bar{v})}{N_A 6\pi\eta^0 R_H} \tag{5.10}$$

where η^0 is the solvent viscosity and R_H the hydrodynamic radius of the protein, thus linking the sedimentation coefficient to the hydrodynamic radius and molecular weight of the particles. In the case of biological molecules that are measured at a temperature different than 20 °C, in a buffer whose viscosity and density differ from those of water, the sedimentation coefficient are corrected as follow :

$$s_{20,w} = s_{exp}\frac{(1 - \bar{v}\rho_{20,w})\eta_{exp}}{(1 - \bar{v}\rho_{exp})\eta_{20,w}} \tag{5.11}$$

where $\rho_{20,w}$ and $\eta_{20,w}$ is the density and viscosity of water at 20 °C, ρ_{exp} and η_{exp} the experimental values.

Application to HTL

Sedimentation velocity experiments were carried out at 42 000 RPM, 4 °C with acquisition of data by absorbance at 278 nm (for the detection of proteins) and interference (for the detection of all type of particles). The data were treated with the software REDATE (v0.2.1) and the analysis done with SEDFIT (v15.01b) and GUSSI (v1.1.0).

Purified HTL was in 20mM Tris pH 8.0, 130mM NaCl and 3% (0.326 M) glycerol and the associated solvent parameters were calculated with Sednterp (Daemon version 20120828 beta): $\rho_{buffer} = 1.012\,77\,\mathrm{g\,ml^{-1}}$ and $\eta_{buffer} = 1.704\,\mathrm{mPa\,s}$. The biophysical parameters of DDM and Amphipol A8-35 can be found in page 195.

5.3.4 Matrix-Assisted Laser Desorption Ionization coupled to Time Of Flight (MALDI-TOF) mass spectrometry

Principle

Mass spectrometry (MS) is an analytical method used to determine the mass of molecules. It is based on the production of ions which can be separated according to their mass-over-charge ratio (m/z). this technique has a wide range of applications: from the analysis of small molecules to the study of large protein complexes. A mass spectrometer is composed of three main parts, which can differ according the goal of the analyses: an ion source, an analyser and a detector.

The molecules must be ionized and this step is performed by the ion source. Matrix Assisted Laser Desorption Ionization (MALDI) is a soft ionization method which permits biomolecules to stay intact and be detected without fragmentation. To induce this type of ionization a laser radiation is focused at high power to evaporate the sample adsorbed on the target. This process is assisted by mixing the sample with a volatile matrix that quickly sublimates when heated by the laser and promote the protonation of the sample. The analyser is the part where ions are separated according to their m/z ratio. A Time Of Flight (TOF) analyser is often coupled with MALDI. The m/z is then measured by a time of flight analyser by linking it to its velocity (u) through :

$$u = \left(\frac{2zV_{acc}}{m}\right)^{1/2} \tag{5.12}$$

which was calculated from the time (t) spent in the tube of length (L) :

$$t = \left(\frac{m}{2zV_{acc}}\right)^{1/2} L \tag{5.13}$$

with V_{acc} representing the potential difference within the acceleration region (N. Zaccai *et al.*, 2017).

The data were collected on a MALDI-TOF instrument (Autoflex, Bruker Daltonics, Bremen, Germany) with the software flexAnalysis 3.0 (Bruker Daltonics, Bremen, Germany), in linear positive mode (for proteins as they carry positive charges at low pH) or negative mode (for DNA whose phosphate groups are negatively charged), with a laser intensity of 30-70%. The calibration was done using standard proteins of known masses (protein standard II, Bruker Daltonics, Bremen).

Protein analyses

A thin layer of saturated α-cyano-4-hydroxycinnamic acid (α-CHCA) solution was deposited on stain steel target (Bruker) and left to dry. The purified sample, at a concentration of at least 1 µM was diluted up to 8 times in formic acid (HCOOH) and deposited on the α-CHCA thin layer. The sample was immediately mixed with

α-CHCA at 10 mg/ml), 2,5-dihydroxybenzoic acid (DHB) at 10 mg/ml) in 70% acetonitrile, 0.0015% trifluoric acid (TFA), 0.75% formic acid (N. Zaccai *et al.*, 2017; Signor and Boeri Erba, 2013).

Nucleic acid analyses

The pHSG298 (see Appendix A) purified plasmid (see subsection 5.2.6) was subjected to enzymatic restriction using ApoI-HF (New England Biolabs $^{®}_{inc}$ alone or in combination with Btg1. The digestion efficiency was checked on a 3% agarose gel run at 90V and stain with 1/7500 SYBR safe. The gel was imaged using a E-box (Vilber). The fragments were dialyzed against water and precipitated with 1M ammonium acetate and 70% ethanol at −20 °C for 2 hours and 30 minutes. The DNA was collected by spinning 30 minutes at 20000g and the pellet washed 4 times with 70% ethanol. The last wash was supplemented with 1M ammonium acetate to favour precipitation and the pelleted fragments resuspended in water at 4 °C overnight and concentrated to a minimum of 10 µM. The DNA fragments were diluted in 30% NH_3, denatured at 95 °C for a minute and mixed with a matrix before depositing on target. The matrices were:

- 0.2M of 2,6-dihydroxyacetophenone (DHAP), 0.16M 2,4,6-trihydroxyacetophenone (THAP), 0.1M ammonium citrate dibasic (AC) in 56% acetonitrile
- 0.25M DHAP, 0.5M AC in 50% aceteonitrile
- 3-hydroxypicolinic acid (3-HPA) : 0.5M of 3-HPA diluted in 30% ACN/H2O (Isola *et al.*, 2001)

5.3.5 Electro Spray Ionization Time Of Flight (LC-ESI-TOF) mass spectrometry

Principle

Electro Spray is a soft method to ionize a liquid sample. High voltage is applied to a metallic needle through which the sample is injected. The applied voltage generates an electric field which causes the exposed liquid surface to form an extended structure, known as a Taylor cone, from which charged droplets are emitted. Droplets are desolvated as they travel down pressure and potential difference toward the mass analyser. Through the ESI process, the molecules can acquire multiple charges, allowing to measure a shorter range of m/z and hence achieve higher resolution for a wider range of molecular masses.

Protein analyses

To analyse protein samples and their degree of labeling ESI-TOF instrument coupled with a liquid chromatography purification was used. Water (HPLC grade Plus, Sigma, cat. no. 34877) Acetonitrile (ACN), purity ≥99.9% (LC-MS Chromasolv, Fluka, cat.

no.34965), Methanol (HPLC), Isopropanol (HPLC), 0.03% Trifluoroacetic acid (TFA) were used. The mobile phase A was composed of 95% ACN, 0.03% TFA and the phase B by a calibration mixture (ESI-L Low Concentration Tuning Mix, Agilent Technologies, cat. no. G1969-85000, containing ten different molecules whose m/z signals range from 118.08 to 2721.89). Data acquisition was carried out in positive ion mode and mass spectra were recorded in the 300-3200 m/z range. The ESI-TOF was operated using the following parameters: ESI source temperature was set at 300 °C. Nitrogen was used as drying gas (with a flow rate of $7\,\mathrm{l\,min^{-1}}$) and as nebulizer gas (using a pressure of 10 psi). The capillary needle voltage was 4 kV. Voltages in the first part of the instrument were set as follows: the fragmentor voltage was 250 V and the skimmer one was 60 V. Spectra acquisition rate was 1 spectrum/s. Instrument pressure values were typically 45×10^{-3} psi (rough vacuum) and 89×10^{-10} psi (TOF vacuum). The calibration of the TOF was semi-automated and the mass error normally was lower than 1 ppm.

For the determination of deuteration level, the proteins were measured in hydrogenated buffer, the exchangeable hydrogen positions (O-H, N-H, S-H) were considered fully hydrogenated and the number of non exchangeable deuterium (D_{non-ex}) is reported in Table 5.1. The deuteration level ($\%D$) was then calculated from the experimental molecular weight (MW_{exp}), the sequence of the protein, assuming all amino acids were equally deuterated as follows:

$$\%D = 1 - \frac{D_{non-ex} - \Delta MW}{D_{non-ex}} \tag{5.14}$$

where Δ_{MW} represents the molecular weight difference between the experimental molecular weight and the theoretical MW for a hydrogenated protein and $\frac{(D_{non-ex} - \Delta MW)}{D_{non-ex}}$ the fraction of the mass that represent hydrogenated unexchangeable positions.

TABLE 5.1: Number of non exchangeable H/D position per amino acid for the calculation of deuteration levels

Amino acid	A	R	N	D	C	E	Q	G	H	I
Number of D_{non-ex}	4	7	3	3	3	5	5	2	5	10
Amino acid	L	K	M	F	P	S	T	W	Y	V
Number of D_{non-ex}	10	9	8	8	7	3	5	8	7	8

The spectra display two peaks: the protein that has been cleaved by the methionine amino-peptidase and the less intense intact protein peak. The peaks were integrated and the beginning of the second peak was defined by calculating the deuteration level of the beginning of the first peak and selecting the mass that corresponds to this deuteration level for the second peak. We therefore consider that the cleaved and uncleaved proteins will have the same extreme deuterations values. In order to quantify the "sharpness" of the first peak,the mass distribution factor was calculated as:

$$Massdistributionfactor = \frac{height}{area/width} \tag{5.15}$$

which can be used to compare the spectra in terms of shape and not only height or area.

5.3.6 1D Nuclear magnetic resonance

Nuclear magnetic resonance describes the capacity of isotopes with a non-zero spin to absorb electromagnetic waves in the radio frequency region (10 MHz-1 GHz) when submitted to a high magnetic field. In the case of molecules, the external magnetic field induces currents in the electron clouds, which in turn generate shielding magnetic fields. As a result, the resonance frequency of a given nuclei will be affected by the surrounding atoms and the difference in resonance frequency, *i.e* the chemical shift (δ), is expressed as:

$$\delta = (\upsilon_x - \upsilon_{ref})/\upsilon_0 \tag{5.16}$$

where υ_x is the resonance frequency of the nucleus of interest, υ_{ref} the resonance of a reference nucleus and υ_0 the frequency of the spectrometer. δ is expressed in part per million (ppm) (Marion, 2013; N. Zaccai *et al.*, 2017).

For the determination of nucleic acid deuteration level, the signal was integrated over a given range of chemical shift and normalized by the concentration of the sample. The integrated intensity was therefore related to the quantity of non null spins (fraction of ^{1}H). The average percentage of ^{1}H for a given ppm range, or over the whole spectrum, was directly related to the percentage of deuteration of the non-exchangeable position of the sample.

The instrument used was a 600 MHz NMR spectrometer (Bruker Avance III). The data were collected in ^{1}H mode, in 100% D_2O buffer with the following parameters: spectral width = 12 ppm, number of scan = 1500, measurement time = 1 h, acquisition time = 1.432 s, FID size = 20664 points.

5.4 Small Angle Neutron Scattering (SANS)

5.4.1 Small-angle scattering theory

The diffraction phenomenon arises from the interaction of matter with radiation waves that are scattered and interfere to give a observable pattern. The interference pattern is observable when the wavelength used is comparable to or smaller than the distance separating the atoms in matter. X-rays interact with the electron clouds of atoms while neutrons interact with the nucleus. As nuclei are much smaller than the wavelength, they behave like a point scatterer. Scattering by a point yields isotropic spherical waves and the intensity per unit angle (I) of the scattered wave is:

$$I = I_0 b^2 \tag{5.17}$$

with I_0 the incident flux and b the scattering length of the nucleus. For neutrons, and to the opposite of X-rays, the scattering length is not proportional to the atomic number

and is in the same range (although negative for hydrogen) for all nuclei, making it easier to detect light atoms with neutrons.

A neutron scattering experiment can be described as follows:

$$I(\lambda,\theta) = I_0(\lambda)\Delta\Omega\eta(\lambda)T_s V\frac{d\sigma}{d\Omega}(Q) \tag{5.18}$$

where $I(\lambda,\theta)$ is the intensity at a given wavelength and angle, I_0 the beam intensity at a given wavelength, $\Delta\Omega\eta(\lambda)$ describes the detector efficiency and geometrical factors inherent to the design of the instrument, T_s the transmssion of the sample (see Equation 5.21), V the effective sample volume and $\frac{d\sigma}{d\Omega}(Q)$ the differential cross section expressed as :

$$\frac{d\sigma}{d\Omega}(Q) = N_p V_p^2(\Delta\rho)^2 P(Q)S(Q) + b_{inc} \tag{5.19}$$

with N_P the number of particles, V_P their volume, $\Delta\rho$ the contrast, $P(Q)$ the form factor that describes the shape of the particle, $S(Q)$ the structure factor that describe the inter particle effects (considered negligible in a classical SANS experiment with a dilute solution) and b_{inc} the incoherent scattering component. The contrast is defined as the difference in scattering length density between the particle and the solvent as $\Delta\rho = \rho_{particle} - \rho_{solvent}$ where ρ is defined as:

$$\rho = \frac{\sum b_p}{V_p} \tag{5.20}$$

with b_p the scattering length of the particle's atoms and V_p the volume of the particle. (N. Zaccai *et al.*, 2017).

5.4.2 Instruments set-up

The small-angle neutron scattering experiments were performed with continuous neutron sources.

Institut Laue Langevin

The Institut Laue Langevin operates the high flux reactor that has a nominal power of 58.3 MW. The neutrons produced by the fission of ^{235}U are moderated in the reactor D_2O pool to 333 K. The energy is further decreased by the liquid deuterium cold source to reach 25 K. The neutrons are then extracted and taken to the instrument using neutron guides. The instrument used was the large dynamic range small-angle diffractometer D22 (see Figure 5.1). The neutron's wavelength is controlled by a velocity selector (Anatole), which select the neutrons based on their speed, at a resolution of $\Delta\lambda/\lambda = 10\%$ for a wavelength ranging from 4.5 to 40 Å. The dynamic collimation compensates for beam divergence and allows a source to sample distance from 1.4 to 17.6 m. The collimation can be replaced by a neutron guide to maximize the flux. The diaphragm shapes the beam to the sample, the sample position can be

adapted to multiple environment (22 position temperature controlled sample changer, online SEC, online fluorescence...) and the maximum flux at the sample position is $1.2 \times 10^8\,\mathrm{n\,cm^{-2}\,s^{-1}}$. The ^{3}He detector is inside an evacuated tube and the sample to detector distance can be modulated between 1.05 to 17.8 m to allow for a total Q range of 4×10^{-4} to $1.05 \times 10^{-1}\,\mathrm{Å}^{-1}$. The instrument is operated by the in-house software Nomad.

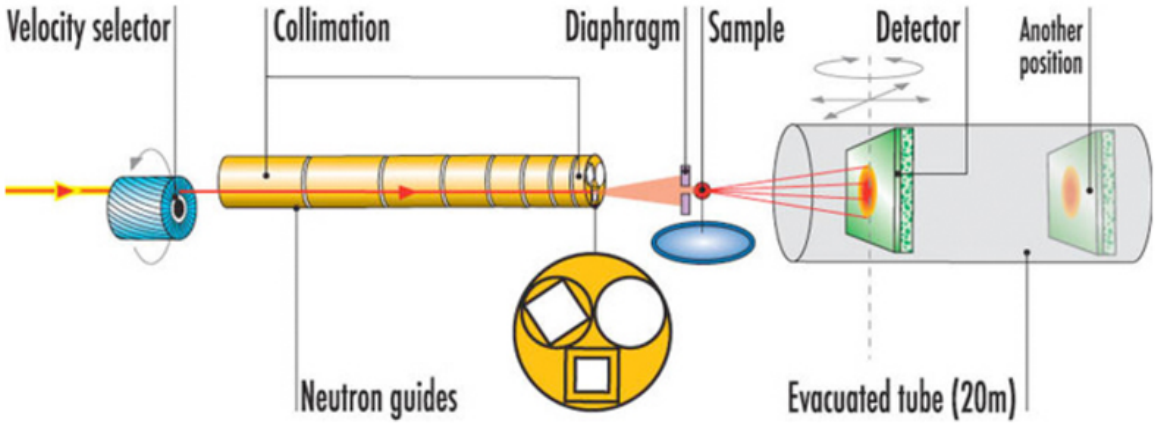


FIGURE 5.1: D22 dynamic range small-angle diffractometer layout. From left to the right : the neutron guide brings the neutron from the cold neutron source, the velocity selector selects a specific wavelength, collimation reduces the divergence of the beam or is replaced by a neutron guide to maximize the flux, the diaphragm shapes the beam to the sample, the sample positions is adaptable to multiple environment, the evacuated tube contains the mobile ^{3}He detector (ILL, Yellow Book).

Forschungsreaktor München II

The Forschungsreaktor München II operates a reactor at a nominal power of 20 MW. The small-angle scattering instrument (KWS-2) (see Figure 5.2) is connected to a liquid deuterium 20 K cold neutron source through a neutron guide. The wavelength is selected using an Astrium velocity selector operating at a resolution of $\Delta\lambda/\lambda = 10{-}20\%$ for wavelengths ranging from 3 to 20 Å and the resolution can be further enhanced with a high speed chopper to a resolution of $\Delta\lambda/\lambda = 1\%$. The guide/collimation section allows a maximum source to sample distance of 18 m and can be combined with a transmission polarizer. A set of MgF_2 focusing lenses allows for up to $2 \times 10^8\,\mathrm{n\,cm^{-2}\,s^{-1}}$ flux at the sample position. The evacuated tube hosts a ^{3}He detector that can move from 1 to 20 m from the sample and is supported by a high resolution ^{6}Li-scintillator at 17 m for a Q range of 1×10^{-4} to $1\,\mathrm{Å}^{-1}$.

5.4.3 Data collection

The samples were measured in quartz cuvettes (Hellma, QS100) in a temperature-controlled sample environment (typically 6 °C). On beamline D22, the Q range was

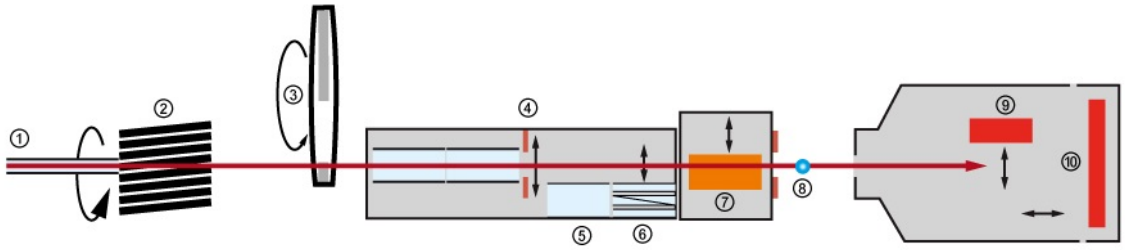


FIGURE 5.2: KWS-2 small-angle scattering instrument layout. 1: Neutron guide, 2: Velocity selector $\Delta\lambda/\lambda = 10-20\%$, 3: High-speed chopper $\Delta\lambda/\lambda = 1\%$, 4: pinhole aperture, 5: neutron guide section, 6: transmission polarizer, 7: MgF_2 focusing lenses, 8: sample, 9: high resolution ^{6}Li-scintillator, 10: ^{3}He detector. http://www.mlz-garching.de/kws-2.

typically achieved by measuring at a wavelength of 6 Å, with a detector distance of 1.4, 7.8 and 17 m and further extended at high Q by shifting the detector laterally at 1.4 m. On KWS2 beamline, data were collected at a wavelength of 7 Å for the detector distance of 2 and 8 m, and at 10 Å for the 17m detector distance. During the measurement, the direct beam was masked by a boron plate and replaced by a attenuator for the measurement of the sample's transmission. The measurement time was calculated so that at least 3×10^5 counts was reached, by evaluating the molecule's total scattering intensity ($countrate_{molecule} = countrate_{sample} - countrate_{buffer}$). The transmission was typically measured for 1 min.

5.4.4 Data correction and normalization on D22

The correction involves the measurement of cadmium plate which completely absorbs the beam (corresponds to the electronic noise of the detector, noted Cd), empty beam (used for the calculation of the transmission, noted eb), empty cuvette (corresponds to the contribution of the empty cuvette in the transmission and scattering, noted ec), water (used to calculate the efficiency of the detector and compensate for geometrical effects) and finally the sample (noted s).

The transmission (T) was calculated from the intensity I_0 as follow:

$$T_s = \frac{I_{0,s}}{I_{0,eb}} \tag{5.21}$$

where the sample was considered as the solution + the cuvette. The transmission of the empty cuvette was then :

$$T_s = \frac{I_{0,ec}}{I_{0,eb}} \tag{5.22}$$

Each pixel intensity (I) was corrected as follows:

$$I_{corrected} = \left(\frac{1}{T_s T_{ec}}(I_s - I_{Cd})\right) - \left(\frac{1}{T_{ec}}(I_{ec} - I_{Cd})\right) \tag{5.23}$$

The first bracket subtracts the electronic noise to the signal of the sample while taking into account the attenuation of the signal from the transmission of the cuvette and the transmission of the sample's solvent. The second bracket subtracts the contribution of the cuvette corrected by its transmission and the electronic noise.

Finally, an independent detector measures the incoming neutron flux at the sample position in order to normalize the data by compensating for the variations inherent to the operation of the reactor and the beam preparation devices. The intensity is divided by the thickness of the sample (typically 1 mm) to normalize the signal to the effective sample volume.

When two different detector distances are used, the intensity is also corrected to compensate for the spreading of scattered neutrons over a larger surface by considering that the counts per unit solid angle are identical, thus bringing the two detector distance intensity to the same scale.

When different collimation distances are used, the variations of flux at the sample are compensated by using relative factors previously calculated at given collimation length and given wavelength.

5.4.5 Data reduction

The centre of the beam was defined from a measurement of an attenuated direct beam. The isotropic scattering 2D data were then radially averaged to express the intensity as a function of the moment transfer ($I(q)$). If necessary, masks can be defined to select out dead pixels or misalignment artefacts.

5.4.6 Data analysis for the determination of particle composition

Guinier approximation

The data were treated using Igor (v6.36, WaveMetrics) with the SANS and USANS package (v7.5, National Institute of Standards and Technology). The detector distances were merged and the buffer curves were subtracted from the samples. The forward scattering ($I(0)$) and radius of gyration (R_G) were calculated using the Guinier approximation:

$$\begin{aligned} I(Q) &= I(0)e^{-(1/3)/R_G^2 Q^2} \\ \ln I(Q) &= \ln I(0) - (1/3)R_G^2 Q^2 \end{aligned} \tag{5.24}$$

For a globular particle, this approximation is considered valid if $R_G Q < 1.5$. The I(0) values were normalized by the mass concentration (C in mg ml^{-1}) of the sample and were expressed as scattering length densities (ρ) using the relation :

$$\frac{I(0)}{C} = \left(\frac{N_A \times 10^3}{M_W}\right)(\Delta\rho V)^2 \tag{5.25}$$

where N_A is Avogadro's number, M_W the molecular weight (Da), V the particle's volume (cm^3) and $\Delta\rho = \rho_{particle} - \rho_{solvent}$ represents the contrast in $Å^{-2}$. The scattering length was plotted against the volume fraction of D_2O and fitted to a linear function to calculate the contrast match point.

Determination of the molecular weight

The molecular weight of measured proteins and complexes was calculated from the SANS data (Jacrot and Zaccai, 1981; Rubinson *et al.*, 2008) as it can be related to the measured I_0 as:

$$I_0 = \frac{CM_W}{Na} * \Delta\rho^2 * \bar{v}^2 \tag{5.26}$$

where C is the concentration in $g\,cm^{-3}$, M_W the molecular weight in $g\,mol^{-1}$, NA Avogadro's number in mol^{-1}, $\Delta\rho$ the contrast in cm^{-2} and $\bar{v}$ the average partial molar volume of proteins (usually of $0.74\,cm^3\,g^{-1}$, Perkins, 1986). The molecular weight is therefore evaluated from data ideally collected at 0% D_2O; in D_2O, this calculation was more error prone as the percentage of exchanged labile hydrogen cannot be truthfully estimated (Jacrot and Zaccai, 1981).

Contrast match point analysis

In the case of a multicomponent complex, the total scattering length of the particle can be expressed with regards to the volumes and scattering length densities of its individual components :

$$(\Delta\rho V)_{particle} = \sum (\Delta\rho V)_{component} \tag{5.27}$$

For example a membrane protein solubilized with detergents, the scattering particle is composed by three scattering length phases : the protein, the detergent and the lipids. The expected scattering length density and volume of the protein were calculated from its amino acid sequence using MULCh online calculator (Whitten *et al.*, 2008) and can be subtracted from the particle's scattering to calculate the contribution of the "non-protein" components and hence its contrast match point. The Equation 5.27 could be expressed as a function of the contrast match points of the components as :

$$CMP_{particle}V_{particle} = CMP_{prot}V_{prot} + CMP_{non-prot}V_{non-prot} \tag{5.28}$$

By subtracting the contribution of the protein's scattering from the total scattering, we then have a direct access to the contrast match point of the "non-protein" component. Assuming the "non-protein" component is composed of DDM and lipids and their respective contrast match point being known (see Appendix D), the relative volume ratio could be calculated :

$$CMP_{non-prot}V_{non-prot} = CMP_{DDM}V_{DDM} + CMP_{lipid}V_{lipid} \tag{5.29}$$

thus allowing to calculate the volume of each components as $V_{particle} = V_{prot} + V_{DDM} + V_{lipid}$ (Botte *et al.*, 2016).

Densities relative localization

Guinier derived R_G values were used for the calculation of the Stuhrmann plot by plotting R_G^2 against the inverse of the solution's contrast ($\Delta\rho^{-1}$), the relation being :

$$R_G{}^2 = R_m + \frac{\alpha}{\Delta\rho} - \frac{\beta}{\Delta\rho^2} \tag{5.30}$$

where R_m is the radius of gyration at infinite contrast ($\Delta\rho^{-1} = 0$), α the internal density fluctuation within the scattering object and β the displacement of scattering length distribution with contrast (Ibel and Stuhrmann, 1975). A positive α indicates the highest scattering density is on average on the outside of the particle, a negative α indicates the opposite and $\beta = 0$ indicates the centre of densities are superimposed. A parabolic fit ($\beta \neq 0$) indicates two distinct centres of densities and the curvatures informs on the distance in between.

The parallel axes analysis also provides information on the distance (D) between two centres of densities by relating it to the R_G of the particle and the R_G of the centres of densities weighted by their scattering length :

$$R_G^2 = f_1 R_{G1}^2 + f_2 R_{G2}^2 + f_1 f_2 D^2 \tag{5.31}$$

where $f_i = \frac{\Delta\rho_i V_i}{\Delta\rho V}$ represents the contribution of a centre of density at a given contrast. The R_G^2 was plotted against f_i and the data fitted with a linear regression.

5.4.7 Guinier independent determination of the contrast match point

For the purpose of determining lipid and protein contrast match points (see subsection 4.1.1 and subsection 4.1.2), the radius of gyration was not required for this analysis and consequently the Guinier approximation was not required. For each sample, an average value at high angle (typically $0.4 < Q < 0.5$ Å^{-1}), which represents the contribution of the buffer, was subtracted to an average value at low angle ($0.02 < Q < 0.03$Å^{-1} for protein or $0.016 < Q < 0.0.19$ Å^{-1} for lipids), which represent the contribution of the particles of interest. The subtracted averaged intensity (I_{sub}) was used instead of the Guinier derived I_0 for the determination of the contrast match point of the particle by plotting $\sqrt{\frac{I_{sub}}{C}}$ (C the concentration in [mg/ml], if applicable) against the volume fraction of D_2O.

The errors(σ) were calculated as standard deviation during the radial averaging and were propagated as follows for averaging :

$$\sigma_{average} = \sqrt{\sum \sigma_{ni}^2}/n \tag{5.32}$$

which corresponds to the quadratic sum of uncertainties of all points, divided by the number of points averaged. For the subtraction :

$$\sigma_{sub} = \sqrt{\sum I_{HighQ}^2 + I_{LowQ}^2} \tag{5.33}$$

For the division by concentration (for protein only) :

$$\sigma_{division} = \sqrt{\left(\frac{\sigma_{sub}}{I_{sub}}\right)^2 + \sigma_{concentration}^2} \tag{5.34}$$

where $\sigma_{concentration}$ represent the 5% experimental and concentration measurement errors. For the error on $\sqrt{\frac{I_{sub}}{C}}$ for protein :

$$\sigma_{\sqrt{\frac{I_{sub}}{C}}} = 0.5\frac{\sigma_{division}}{I_{sub}/C} \tag{5.35}$$

and for the error on $\sqrt{I_{sub}}$ for lipids :

$$\sigma_{\sqrt{I_{sub}}} = 0.5\frac{\sigma_{sub}}{I_{sub}} \tag{5.36}$$

A linear function $y = a \pm \sigma_a \ + \ b \pm \sigma_b \ f_{D_2O}$ weighted by the errors was fitted to the data using Igor (v6.36, WaveMetrics) and the contrast match point was calculated from the equation of the fit as $\sqrt{(I)} = 0$ and the error calculated as :

$$\sigma_{CMP} = -\frac{a}{b}\sqrt{\left(\frac{2\sigma_a}{a}\right)^2 + \left(\frac{2\sigma_b}{b}\right)^2} \tag{5.37}$$

Chapter 6

Concluding remarks

Conclusion en Français

La translocation est un processus cellulaire qui est nécessaire à la survie des cellules. Elle requiert de nombreux partenaires qui effectuent les différentes étapes amenant à la sécrétion de protéines vers le périplasme ou l'insertion de protéines dans la membrane plasmique. Les deux étapes majeures nécessaires pour le transport de la protéine vers sa destination finale est l'adressage vers le canal de conduction des protéines et la translocation elle-même. Dans cette étude, nous avons analysé des complexes protéiques spécifiques qui contribuent à ces mécanismes.

La biologie structurale est devenue un acteur majeur pour la compréhension des fonctions cellulaires puisqu'elle permet d'analyser des processus spécifiques pour en comprendre leur fonctionnement à l'échelle moléculaire. Actuellement (avril 2018, voir https://www.rcsb.org/stats/summary), 89,5% des structures déposées sur la PDB ont été résolues en utilisant la cristallographie à rayons X. Cependant, cette méthode requiert la génération de cristaux, étape délicate qui constitue souvent une barrière importante pour certaines macro-molécules et particulièrement les gros assemblages protéiques ou les protéines membranaires. D'autres méthodes permettent de résoudre les structures à résolution atomique, tels que la RMN et microscopie électronique, mais leur contribution reste marginale (respectivement 8,7% et 1,5% des structures déposées sur la PDB).

Si les méthodes de diffusion en solution ne permettent pas d'atteindre la résolution atomique, elles ont des avantages qui leurs sont propres. Le cas de la diffusion de neutrons à petits angles est particulièrement intéressant pour l'étude de complexes macro-moléculaires puisqu'elle peut être combinée à la variation de contraste et à la deutération biomoléculaire pour permettre l'acquisition de données uniques. Puisque la composition atomique des types de biomolécules diffère, il est possible de les distinguer au sein d'un complexe composite. Dans le cas de complexes protéine-protéine, des protéines marquées par des isotopes peuvent être produites pour pouvoir être distinguées de leur partenaires non marqués. Ainsi, le succès et la portée d'une expérience de SANS peut dépendre de notre capacité à marquer les biomolécules de façon précise et contrôlée.

Les avantages de la diffusion de neutrons à petits angles ont pu être idéalement exploités dans cette étude de la machinerie de translocation bactérienne puisque les acteurs de ces processus cellulaires sont des complexes muti-protéiques ou des complexes à plusieurs composants tels que les protéines membranaires.

Nous avons combiné la biochimie et la deutération biomoléculaire avec des expériences de variation de contraste dans l'optique d'analyser l'organisation de complexes composés de SecB, SecA et un substrat de translocation. Cette première observation d'un complexe capable de promouvoir la translocation nous a donné des informations sur la disposition relative des différentes protéine, la stœchiométrie de ces complexes ainsi que sur les modifications structurales induites par le substrat de translocation. Ces données nous ont permis de proposer un modèle suggérant que SecB s'étend quand elle se lie à son substrat et que l'affinité de SecA pour SecB dépend de la présence de ce substrat sur la chaperonne.

Nous avons également analysés les caractéristiques structurales de l'Holotranslocon, un gros assemblage protéique impliqué dans la translocation des protéines sécrétées et membranaires. Puisque la structure de l'Holotranslocon forme une cavité occupée par des lipides, la diffusion de neutrons à petits angles et sa capacité à distinguer différent types de molécules sont particulièrement intéressants pour l'étude de ce complexe à plusieurs composants. Nous avons pu analyser HTL ayant été solubilisé avec un nouveau surfactant ainsi que le complexe deutéré.

Cependant, dans les deux cas, la qualité de l'échantillon ne nous a pas permis de poursuivre les analyses jusqu'à l'établissement d'un modèle structural. Dans le cas de SecB, les données de SANS nous ont clairement indiqué que les complexes formés n'étaient pas homogènes en terme de stœchiométrie et avaient tendance à se dissocier sous certaines conditions. Ces problèmes pourraient être résolus en améliorant les conditions de purification et le protocole d'interaction. Dans le cas de l'Holotranlocon, les caractérisations biophysiques ont montré que le complexe ne pouvait pas être efficacement purifié en tant qu'hétéro-heptamère et avait tendance à se dissocier. Nous avons attribué ces problèmes au fait que ce complexe s'assemble et se désassemble de façon dynamique *in vivo* mais aussi à l'utilisation de surfactants qui déstabilisent inévitablement HTL et interfèrent avec les méthodes d'analyse biophysique. Nous sommes convaincus que la stratégie de préparation doit être ré-évaluée afin d'effectuer de nouvelles analyses structurales et fonctionnelles. Puisque que cet échantillon est finalement composé de protéines, lipides et surfactants, cette étude pourrait bénéficier de contrastes améliorés pouvant être apportés par l'utilisation de marquages isotopiques spécifiques.

Alors que le marquage de protéines par le deutérium *in vivo* est une méthode établie, l'utilisation des neutrons en biologie pourrait être limitée par le manque d'informations quant au marquage isotopique des lipides et acides nucléiques. L'analyse de système complexes impliquant différents composants pourrait également être limitée par le faible contraste inhérent à leur composition et notre incapacité à effectuer des marquages spécifiques *in vivo*. Dans ce travail, nous nous sommes attaqués à ces deux problèmes. Nous avons d'abord établi la relation entre les conditions de cultures de bactérie et le niveau de marquage des lipides et acides nucléiques. Nous avons ensuite développé

un protocole permettant de marquer spécifiquement les protéines en supplémentant les milieux de culture avec des acides aminés marqués par des isotopes.

De ce fait, nous avons montré, dans cette étude, que les méthodes de diffusion de neutrons peuvent être utilisées pour répondre à des questions biologiques impliquant des complexes à plusieurs composants. Cependant, la production d'échantillon peut être un facteur limitant important puisque le SANS requiert d'utiliser une solution suffisamment concentrée et homogène, particulièrement si l'établissement de modèles structuraux est envisagé. Le marquage isotopique est aussi d'une importance capitale pour l'optimisation et la pertinence des études biologiques impliquant les neutrons. Les derniers développement en terme de marquage isotopique *in vivo*,instruments (intensité et préparation du faisceau) et les environnements d'échantillons tels que la chromatographie à exclusion de taille *in situ* (Jordan *et al.*, 2016) ou la micro-fluidique (Lopez *et al.*, 2015; Adamo, Poulos, Miller, *et al.*, 2017; Adamo, Poulos, G Lopez, *et al.*, 2018) vont assurer la pérennité des méthodes neutroniques pour la biologie.

Conclusion in english

Translocation is a cellular mechanism that is necessary for the survival of the cell. It requires numerous partners to perform the different steps that lead to the secretion of protein into the periplasm or insertion into the plasma membrane. The two major steps that are necessary to efficiently transport the newly synthesised protein to its final destination are the targeting to the protein conducting channel, and the translocation process itself. In this work, specific proteins complexes that take part in these mechanisms have been studied.

Structural biology has become a major actor in understanding of cellular functions since it allows the analysis of specific processes of importance to biological function a molecular level. To date (April 2018, https://www.rcsb.org/stats/summary), 89,5% of all deposited molecular structures in the PDB have been solved using X-ray crystallography. However, this method requires the ability to crystallise the biological system in question; this can be very problematic and in many cases is not possible – especially in the case of large protein assemblies or membrane proteins. Other structural methods allow the acquisition of atomic resolution data, such as NMR and Electron microscopy, although these make up relatively small fractions of the PDB database (8,7% and 1,5% of all deposited structures on the PDB respectively).

While solution scattering does not provide atomic resolution detail, it comes with special advantages. Small-angle neutron scattering (SANS) is of particular interest for the study of macromolecular complexes since the combined use of contrast variation and biomolecule deuteration enables the acquisition of unique data. Because the atomic composition of the different classes of biomolecules differ, it is possible to use SANS to distinguish them within a complex, and therefore to selectively collect information on the different components within a composite complex. In the case of protein-protein complexes, isotope labelled proteins can be produced and distinguished from unlabelled partners within a complex. The success and scope of SANS methods (as compared

to SAXS methods) depends crucially on maximising the exploitation of deuterium labelling in increasingly innovative ways.

Given these advantages of SANS as used in combination with deuteration, large complexes such as the bacterial translocation machinery may be accessible to these approaches since the actors of this cellular processes are multi-protein complexes or large multicomponent complexes such as membrane proteins.

Biomolecular deuteration methods have been used in conjunction with contrast variation to investigate the organisation of the complexes composed of SecB, SecA and a translocation substrate. This first observation of a "translocation competent complex" has given us information on the relative disposition of the different proteins, the stoichiometry of the complexes and the structural rearrangements that occur prior to the translocation of the protein substrate. Based on these data a model has been proposed that suggests that SecB expands when bound to its substrate; furthermore, it has been found that SecA affinity to SecB strongly depends on the presence of this substrate on the chaperone.

The structural features of the Holotranslocon (HTL) have also been investigated. HTL is a large membrane protein assembly involved in the translocation of secreted and membrane proteins. Since the HTL's structure forms a cavity that encloses lipids, SANS and its capacity to differentiate different type of biomolecules are particularly suited for the study this multicomponent complex. HTL solubilized in the presence a new surfactant, as well as in deuterated form, was analysed.

However, in both cases, the quality of the sample did not allow us to pursue the analysis with detailed structural modelling. In the case of SecB, the SANS data clearly indicated that the complexes formed were not homogeneous in terms of stoichiometry and tended to disassemble under certain conditions. For the future, these issues may be tackled by improving the purification and interaction protocol. In the case of HTL, extensive biophysical characterization showed that the complex could not be efficiently purified as the hetero-heptamer as anticipated and had a high tendency to disassemble. These issues were partly attributable to the inherent nature of HTL, which is likely to dynamically assemble and disassemble *in vivo*, but also to the use of surfactants that inevitably destabilize the complex and interfere with biophysical methods. It is believed that the preparation strategy may need to be further developed to enable further structural and functional characterization. Considering that the final sample is composed by protein, surfactant and lipids, the study of this multicomponent complex could benefit from improved contrasts that would be achieved through selective isotopic labelling.

While the *in vivo* isotope labelling methods for proteins are well established and continue to develop rapidly, the use of neutrons in biology is still limited by the lack of information on the isotope labelling of lipids and nucleic acids. The analysis of complex systems involving different components is also be limited by the inherent contrast of the components themselves and the special conditions required to reliably perform selective labelling *in vivo*. In this work, both issues were tackled. The relationship

between the bacterial culture conditions and the extent of labelling for lipids and nucleic acid was first established. A protocol was also developed that allows the selective labeling of a protein by supplementing the culture medium with labelled amino acids.

In this work it is demonstrated that neutron scattering methods can be used to answer complex biological questions that involve multi-components systems. However, the production of large highly complex samples emphasizes the need to pursue advanced methodological approaches that address sample stability as well as appropriate labeling regimes that optimize the exploitation of SANS. The latest developments in terms of *in vivo* labelling methods, instruments (increase of flux, preparation of the beam) and sample environments such as *in situ* size exclusion chromatography (Jordan *et al.*, 2016) or microfluidic (Lopez *et al.*, 2015; Adamo, Poulos, Miller, *et al.*, 2017; Adamo, Poulos, G Lopez, *et al.*, 2018), will further expand to portfolio of methods available for the use of neutrons in the study of biological systems.

Appendices

Appendix A

List of plasmids

Insert	Vector	Promoter	Resistance	Reference
HTL	Acembl	Trc, Ara	Amp, Kan, Cam	Bieniossek *et al.*, 2009
SecYEG	pTRC-99a	Trc	Amp	P. Kuhn *et al.*, 2011
SecDF YidC, YajC	Acembl	Trc, Ara	Amp, Kan	Bieniossek *et al.*, 2009
SecA-His6	pET-5a	Trc	Amp	Duong, 2003
SecAΔ_{cys}	pET-30b	Trc	Kan	Erlandson *et al.*, 2008
SecB	pT-7	Trc	Kan	Weiss *et al.*, 1988
MBP	pET-28a	Trc	Kan	Laux *et al.*, 2008
NW50	pET-28a	Trc	Kan	Nasr *et al.*, 2016
Sortase	pET-29	Trc	Kan	I. Chen *et al.*, 2011

Appendix B

Protein sequences and biophysical parameters

Maltose Binding Protein

Sequence:

```
MBP  MGSSHHHHHHSSGLVPRGSHMKTEEGKLVIWINGDKGYNGLAEVGKKFEK    50
MBP  DTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEI   100
MBP  TPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTW   150
MBP  EEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIK   200
MBP  DVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGP   250
MBP  WAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEF   300
MBP  LENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEI   350
MBP  MPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQT   387
```

Number of residues	MW ($g\,mol^{-1}$)	ϵ_{280} ($M^{-1}\,cm^{-1}$)	Isoelectric point	Tag
387	42491,16	66350	5,65	His-6

SecB

Sequence:

```
SecB  MSEQNNTEMTFQIQRIYTKDISFEAPNAPHVFQKDWQPEVKLDLDTASSQ   50
SecB  LADDVYEVVLRVTVTASLGEETAFLCEVQQGGIFSIAGIEGTQMAHCLGA  100
SecB  YCPNILFPYARECITSMVSRGTFPQLNLAPVNFDALFMNYLQQQAGEGTE  150
SecB  EHQDA  155
```

Number of residues	MW ($g\,mol^{-1}$)	ϵ_{280} ($M^{-1}\,cm^{-1}$)	Isoelectric point	Tag
155	17277,36	13200	4,26	none

SecY

Sequence:

```
SecY  MAKQPGLDFQSAKGGLGELKRRLLFVIGALIVFRIGSFIPIPGIDAAVLA   50
SecY  KLLEQQRGTIIEMFNMFSGGALSRASIFALGIMPYISASIIIQLLTVVHP  100
SecY  TLAEIKKEGESGRRKISQYTRYGTLVLAIFQSIGIATGLPNMPGMQGLVI  150
SecY  NPGFAFYFTAVVSLVTGTMFLMWLGEQITERGIGNGISIIIFAGIVAGLP  200
SecY  PAIAHTIEQARQGDLHFLVLLLVAVLVFAVTFFVVFVERGQRRIVVNYAK  250
SecY  RQQGRRVYAAQSTHLPLKVNMAGVIPAIFASSIILFPATIASWFGGGTGW  300
SecY  NWLTTISLYLQPGQPLYVLLYASAIIFFCFFYTALVFNPRETADNLKKSG  350
SecY  AFVPGIRPGEQTAKYIDKVMTRLTLVGALYITFICLIPEFMRDAMKVPFY  400
SecY  FGGTSLLIVVVVIMDFMAQVQTLMMSSQYESALKKANLKGYGR  443
```

Number of residues	MW ($g\,mol^{-1}$)	ϵ_{280} ($M^{-1}\,cm^{-1}$)	Isoelectric point	Tag
443	48511,69	44475	9,89	none

SecE

Sequence:

```
SecE  MHHHHHHDDDDKAMGANTEAQGSGRGLEAMKWVVVVALLLVAIVGNYLYR   50
SecE  DIMLPLRALAVVILIAAAGGVALLTTKGKATVAFAREARTEVRKVIWPTR  100
SecE  QETLHTTLIVAAVTAVMSLILWGLDGILVRLVSFITGLRF   140
```

Number of residues	MW ($g\,mol^{-1}$)	ϵ_{280} ($M^{-1}\,cm^{-1}$)	Isoelectric point	Tag
140	15227,03	19480	9,59	His-6

SecG

Sequence:

```
SecG  MYEALLVVFLIVAIGLVGLIMLQQGKGADMGASFGAGASATLFGSSGSGN   50
SecG  FMTRMTALLATLFFIISLVLGNINSNKTNKGSEWENLSAPAKTEQTQPAA  100
SecG  PAKPTSDIPN   110
```

Number of residues	MW ($g\,mol^{-1}$)	ϵ_{280} ($M^{-1}\,cm^{-1}$)	Isoelectric point	Tag
110	11365,16	6990	6,09	none

SecD

Sequence:

```
SecD  MHHHHHHGLNRYPLWKYVMLIVVIVIGLLYALPNLFGEDPAVQITGARGV   50
SecD  AASEQTLIQVQKTLQEEKITAKSVALEEGAILARFDSTDTQLRAREALMG  100
SecD  VMGDKYVVALNLAPATPRWLAAIHAEPMKLGLDLRGGVHFLMEVDMDTAL  150
SecD  GKLQEQNIDSLRSDLREKGIPYTTVRKENNYGLSITFRDAKARDEAIAYL  200
SecD  SKRHPDLVISSQGSNQLRAVMSDARLSEAREYAVQQNINILRNRVNQLGV  250
SecD  AEPVVQRQGADRIVVELPGIQDTARAKEILGATATLEFRLVNTNVDQAAA  300
SecD  ASGRVPGDSEVKQTREGQPVVLYKRVILTGDHITDSTSSQDEYNQPQVNI  350
SecD  SLDSAGGNIMSNFTKDNIGKPMATLFVEYKDSGKKDANGRAVLVKQEEVI  400
SecD  NIANIQSRLGNSFRITGINNPNEARQLSLLLRAGALIAPIQIVEERTIGP  450
SecD  TLGMQNIEQGLEACLAGLLVSILFMIIFYKKFGLIATSALIANLILIVGI  500
SecD  MSLLPGATLSMPGIAGIVLTLAVAVDANVLINERIKEELSNGRTVQQAID  550
SecD  EGYRGAFSSIFDANITTLIKVIILYAVGTGAIKGFAITTGIGVATSMFTA  600
SecD  IVGTRAIVNLLYGGKRVKKLSI   622
```

Number of residues	MW ($g\,mol^{-1}$)	ϵ_{280} ($M^{-1}\,cm^{-1}$)	Isoelectric point	Tag
622	67540,10	33350	8,62	His-6

SecF

Sequence:

```
SecF  MAQEYTVEQLNHGRKVYDFMRWDYWAFGISGLLLIAAIVIMGVRGFNWGL   50
SecF  DFTGGTVIEITLEKPAEIDVMRDALQKAGFEEPMLQNFGSSHDIMVRMPP  100
SecF  AEGETGGQVLGSQVLKVINESTNQNAAVKRIEFVGPSVGADLAQTGAMAL  150
SecF  MAALLSILVYVGFRFEWRLAAGVVIALAHDVIITLGILSLFHIEIDLTIV  200
SecF  ASLMSVIGYSLNDSIVVSDRIRENFRKIRRGTPYEIFNVSLTQTLHRTLI  250
SecF  TSGTTLMVILMLYLFGGPVLEGFSLTMLIGVSIGTASSIYVASALALKLG  300
SecF  MKREHMLQQKVEKEGADQPSILP  323
```

Number of residues	MW ($g\,mol^{-1}$)	ϵ_{280} ($M^{-1}\,cm^{-1}$)	Isoelectric point	Tag
323	35382,38	33920	5,57	none

YidC

Sequence:

```
YidC  MDPSSRDSQRNLLVIALLFVSFMIWQAWEQDKNPQPQAQQTTQTTTTAAG   50
YidC  SAADQGVPASGQGKLISVKTDVLDLTINTRGGDVEQALLPAYPKELNSTQ  100
YidC  PFQLLETSPQFIYQAQSGLTGRDGPDNPANGPRPMYNVEKDAYVLAEGQN  150
YidC  ELQVPMTYTDAAGNTFTKTFVLKRGDYAVNVNYNVQNAGEKPLEISTFGQ  200
YidC  LKQSITLPPHLDTGSSNFALHTFRGAAYSTPDEKYEKYKFDTIADNENLN  250
YidC  ISSKGGWVAMLQQYFATAWIPHNDGTNNFYTANLGNGIAAIGYKSQPVLV  300
YidC  QPGQTGAMNSTLWVGPEIQDKMAAVAPHLDLTVDYGWLWFISQPLFKLLK  350
YidC  WIHSFVGNWGFSIIIITFIVRGIMYPLTKAQYTSMAKMRMLQPKIQAMRE  400
YidC  RLGDDKQRISQEMMALYKAEKVNPLGGCFPLLIQMPIFLALYYMLMGSVE  450
YidC  LRQAPFALWIHDLSAQDPYYILPILMGVTMFFIQKMSPTTVTDPMQQKIM  500
YidC  TFMPVIFTVFFLWFPSGLVLYYIVSNLVTIIQQQLIYRGLEKRGLHSREK  550
YidC  KKSHHHHHH  559
```

Number of residues	MW ($g\,mol^{-1}$)	ϵ_{280} ($M^{-1}\,cm^{-1}$)	Isoelectric point	Tag
559	62923,38	96260	7,77	His-6

YajC

Sequence:

```
YajC  MSFFISDAVAATGAPAQGSPMSLILMLVVFGLIFYFMILRPQQKRTKEHK    50
YajC  KLMDSIAKGDEVLTNGGLVGRVTKVAENGYIAIALNDTTEVVIKRDFVAA   100
YajC  VLPKGTMKALSSGGKRRWKKNFIAVSAANRFKKIS   135
```

Number of residues	MW ($g\,mol^{-1}$)	ϵ_{280} ($M^{-1}\,cm^{-1}$)	Isoelectric point	Tag
135	14706,45	8480	10,32	CBP

SecYEG

Sequence: SecY + SecE + SecG

Number of residues	MW ($g\,mol^{-1}$)	$\epsilon_{280,measured}$ ($M^{-1}\,cm^{-1}$)	Isoelectric point	Tag
693	75067,85	139000	9,79	His-6

HTL

Sequence: SecY + SecE + SecG + SecD + SecF + YidC + YajC

Number of residues	MW ($g\,mol^{-1}$)	$\epsilon_{280,measured}$ ($M^{-1}\,cm^{-1}$)	Isoelectric point	Tag
2332	255520,05	172000	9,25	3 x His-6, CBP

DFYY

Sequence: SecD + SecF + YidC + YajC

Number of residues	MW ($g\,mol^{-1}$)	$\epsilon_{280,measured}$ ($M^{-1}\,cm^{-1}$)	Isoelectric point	Tag
1639	180470,22	497000	8.83	2 x His-6, CBP

SecA

Sequence:

```
SecA  MLIKLLTKVFGSRNDRTLRRMRKVVNIINAMEPEMEKLSDEELKGKTAEF   50
SecA  RARLEKGEVLENLIPEAFAVVREASKRVFGMRHFDVQLLGGMVLNERSIA  100
SecA  EMRTGEGKTLTATLPAYLNALTGKGVHVVTVNDYLAQRDAENNRPLFEFL  150
SecA  GLTVGINLPGMPAPAKREAYAADITYGTNNEYGFDYLRDNMAFSPEERVQ  200
SecA  RKLHYALVDEVDSILIDEARTPLIISGPAEDSSEMYKRVNKIIPHLIRQE  250
SecA  KEDSETFQGEGHFSVDEKSRQVNLTERGLVLIEELLVKEGIMDEGESLYS  300
SecA  PANIMLMHHVTAALRAHALFTRDVDYIVKDGEVIIVDEHTGRTMQGRRWS  350
SecA  DGLHQAVEAKEGVQIQNENQTLASITFQNYFRLYEKLAGMTGTADTEAFE  400
SecA  FSSIYKLDTVVVPTNRPMIRKDLPDLVYMTEAEKIQAIIEDIKERTAKGQ  450
SecA  PVLVGTISIEKSELVSNELTKAGIKHNVLNAKFHANEAAIVAQAGYPAAV  500
SecA  TIATNMAGRGTDIVLGGSWQAEVAALENPTAEQIEKIKADWQVRHDAVLE  550
SecA  AGGLHIIGTERHESRRIDNQLRGRSGRQGDAGSSRFYLSMEDALMRIFAS  600
SecA  DRVSGMMRKLGMKPGEAIEHPWVTKAIANAQRKVESRNFDIRKQLLEYDD  650
SecA  VANDQRRAIYSQRNELLDVSDVSETINSIREDVFKATIDAYIPPQSLEEM  700
SecA  WDIPGLQERLKNDFDLDLPIAEWLDKEPELHEETLRERILAQSIEVYQRK  750
SecA  EEVVGAEMMRHFEKGVMLQTLDSLWKEHLAAMDYLRQGIHLRGYAQKDPK  800
SecA  QEYKRESFSMFAAMLESLKYEVISTLSKVQVHHHHHH   837
```

Number of residues	MW ($g\,mol^{-1}$)	ϵ_{280} ($M^{-1}\,cm^{-1}$)	Isoelectric point	Tag
837	94980,04	74260	5,54	His-6

Appendix C

Guinier fits of the small-angle neutron scattering data

The next pages present the small-angle neutron scattering data after subtraction of the buffer's contribution.

The fits of the guinier region were performed with Atsas (v2.7, Svergun, Richard, *et al.*, 1998).

FIGURE C.1: Guinier fits of h-HTL and SecYEG

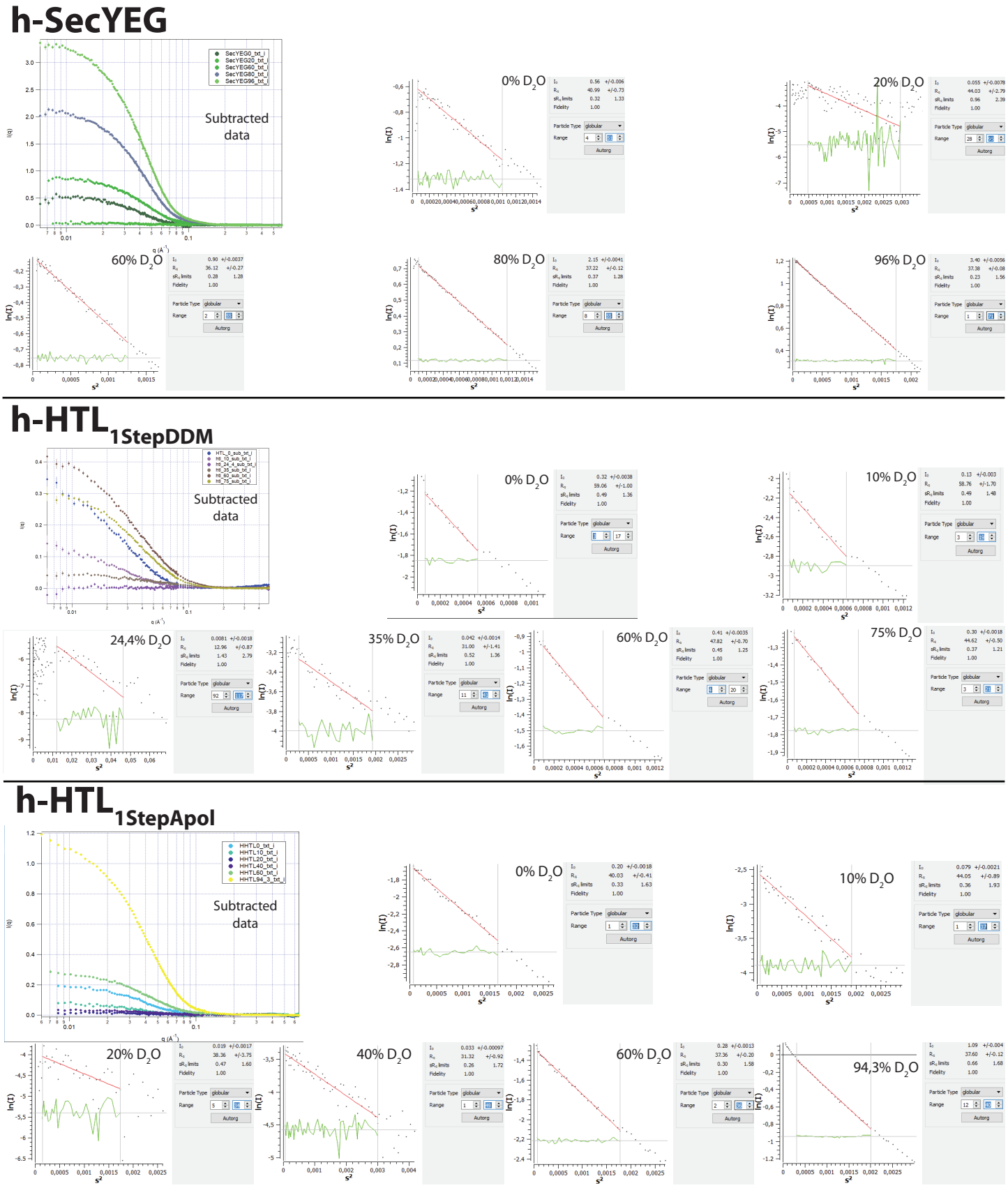


FIGURE C.2: Guinier fits of d-HTL

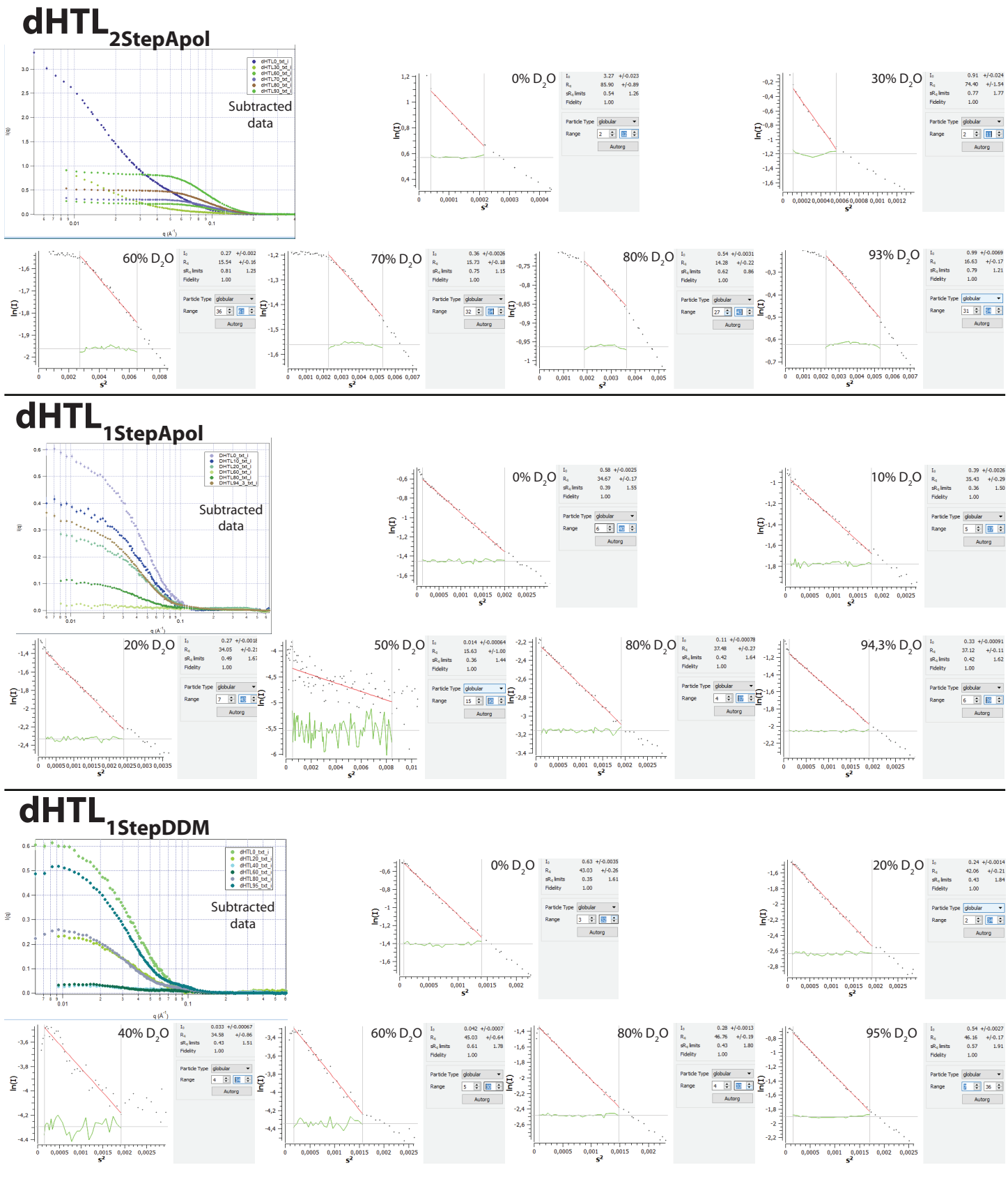
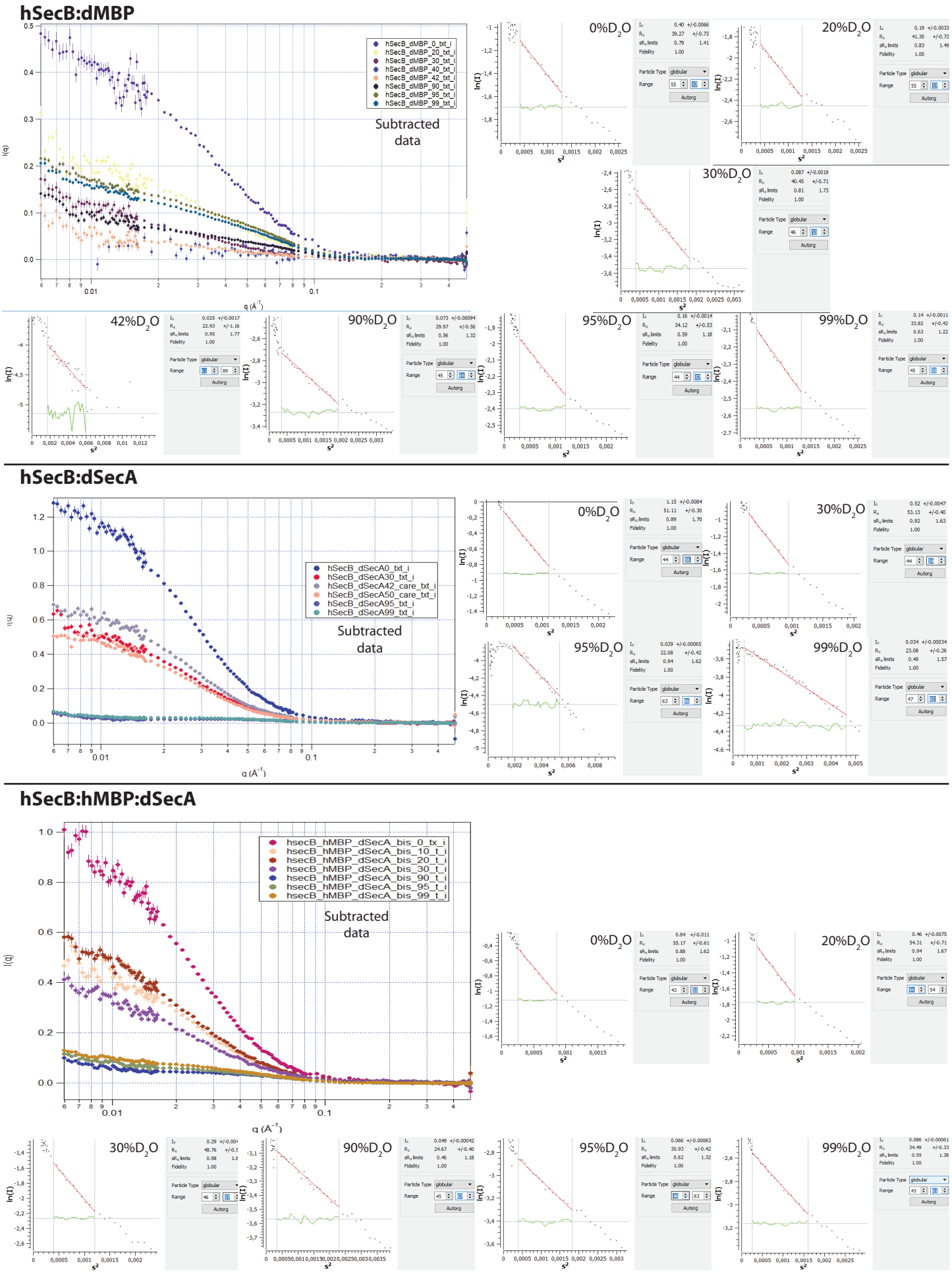


FIGURE C.3: Guinier fits of SecB, MBP and SecA complexes



Appendix D

Culture media composition

Luria Broth (LB): $10\,g\,l^{-1}$ tryptone, $5\,g\,l^{-1}$ yeast extract, $10\,g\,l^{-1}$ NaCl

2xYT (DYT) medium: $16\,g\,l^{-1}$ tryptone, $10\,g\,l^{-1}$ yeast extract, $5\,g\,l^{-1}$ NaCl

ENFORS (minimal medium): $6.86\,g\,l^{-1}$ $(NH_4)_2SO_4$, $1.56\,g\,l^{-1}$ KH_2PO_4, $6.01\,g\,l^{-1}$ Na_2HPO_4, $0.49\,g\,l^{-1}$ $(NH_4)_2HC_6H_6O_7$, $5\,g\,l^{-1}$ glycerol, 1 mM $MgSO_4$, $0.5\,mg\,l^{-1}$ $CaCl_2 \cdot 2\,H_2O$, $16.7\,mg/l$ $FeCl_3 \cdot 6\,H_2O$, $0.18\,mg\,l^{-1}$ $ZnSO_4 \cdot 7\,H_2O$, $0.16\,mg\,l^{-1}$ $CuSO_4 \cdot 5\,H_2O$, $0.15\,mg\,l^{-1}$ $MnSO_4 \cdot 4\,H_2O$, $0.18\,mg\,l^{-1}$ $CoCl_2 \cdot 6\,H_2O$, $20.1\,mg\,l^{-1}$ $Na \cdot EDTA$

SOC medium : $20\,g\,l^{-1}$ tryptone, $5\,g\,l^{-1}$ yeast extract, $0.5\,g\,l^{-1}$ NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$

Agar plates: culture medium + $15\,g\,l^{-1}$ of agar-agar

Appendix E

List of providers

All chemicals were bought from Sigma-Aldrich, unless cited in the list below :

- **n-DoDecyl-β-D-Maltopyranoside**: GLYCON Biochemicals GmbH
- **Amphipol A8-35**: Anatrace
- **Luria Broth**: Fischer, BD Difco^{TM}
- **Calmodulin resin**: Agilent Technologies
- **Chromatography columns**: GE Healthcare
- **Bacterial and fungal preservations system**: Microbank
- **Ultra-filtration devices**: Amicon©
- **Cobalt IMAC resin**: Qiagen©
- **PD10 desalting columns**: Biorad
- **Plasmid purification kits**: Qiagen
- **Extrusion kit**: Avanti Polar Lipids©
- **TLC plates**: Macherey Nagel

Bibliography

Adamo, M., A. S. Poulos, C. G Lopez, *et al.* (Mar. 7, 2018). "Droplet Microfluidic SANS". In: *Soft Matter* 14.10, pp. 1759–1770. ISSN: 1744-6848. pmid: 29355865.

Adamo, M., A. S. Poulos, R. M. Miller, *et al.* (May 2, 2017). "Rapid Contrast Matching by Microfluidic SANS". In: *Lab on a Chip* 17.9, pp. 1559–1569. ISSN: 1473-0189. URL: http://pubs.rsc.org/en/content/articlelanding/2017/lc/c7lc00179g (visited on 04/14/2018).

Adams, R. L. P., J. T. Knowler, and D. P. Leader (Apr. 18, 2013). *The Biochemistry of the Nucleic Acids*. Springer Science & Business Media. 689 pp. ISBN: 978-94-011-2290-0.

Allen, W. J., R. A. Corey, P. Oatley, *et al.* (May 16, 2016). "Two-Way Communication between SecY and SecA Suggests a Brownian Ratchet Mechanism for Protein Translocation". In: *eLife* 5, e15598. ISSN: 2050-084X. URL: https://elifesciences.org/content/5/e15598v1 (visited on 06/01/2016).

Arkowitz, R. and W. Wickner (Feb. 15, 1994). "SecD and SecF Are Required for the Proton Electrochemical Gradient Stimulation of Preprotein Translocation." In: *The EMBO Journal* 13.4, p. 954. pmid: 8112309. URL: /pmc/articles/PMC394897/?report=abstract (visited on 04/11/2014).

Artero, J.-B., M. Härtlein, S. McSweeney, *et al.* (Nov. 1, 2005). "A Comparison of Refined X-Ray Structures of Hydrogenated and Perdeuterated Rat γE-Crystallin in H2O and D2O". In: *Acta Crystallographica Section D: Biological Crystallography* 61.11, pp. 1541–1549. ISSN: 0907-4449. URL: http://scripts.iucr.org/cgi-bin/paper?wd5042 (visited on 03/11/2018).

Arunmanee, W., M. Pathania, A. S. Solovyova, *et al.* (Aug. 23, 2016). "Gram-Negative Trimeric Porins Have Specific LPS Binding Sites That Are Essential for Porin Biogenesis". In: *Proceedings of the National Academy of Sciences* 113.34, E5034–E5043. ISSN: 0027-8424, 1091-6490. pmid: 27493217. URL: http://www.pnas.org/content/113/34/E5034 (visited on 04/09/2018).

Ast, T., G. Cohen, and M. Schuldiner (Feb. 28, 2013). "A Network of Cytosolic Factors Targets SRP-Independent Proteins to the Endoplasmic Reticulum". In: *Cell* 152.5, pp. 1134–1145. ISSN: 1097-4172. pmid: 23452858.

Ataide, S. F., N. Schmitz, K. Shen, *et al.* (Feb. 18, 2011). "The Crystal Structure of the Signal Recognition Particle in Complex with Its Receptor". In: *Science (New York, N.Y.)* 331.6019, pp. 881–886. ISSN: 0036-8075. pmid: 21330537. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3758919/ (visited on 02/13/2018).

Auclair, S. M., M. K. Bhanu, and D. A. Kendall (Jan. 2012). "Signal Peptidase I: Cleaving the Way to Mature Proteins". In: *Protein Science : A Publication of the Protein Society* 21.1, pp. 13–25. ISSN: 0961-8368. pmid: 22031009. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3323777/ (visited on 01/17/2018).

Barkai, N. and B.-Z. Shilo (Dec. 14, 2007). "Variability and Robustness in Biomolecular Systems". In: *Molecular Cell* 28.5, pp. 755–760. ISSN: 1097-2765. pmid: 18082601.

Batchelor, E. and M. Goulian (Jan. 21, 2003). "Robustness and the Cycle of Phosphorylation and Dephosphorylation in a Two-Component Regulatory System". In: *Proceedings of the National Academy of Sciences of the United States of America* 100.2, pp. 691–696. ISSN: 0027-8424. pmid: 12522261. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC141058/ (visited on 02/21/2018).

Bauer, B. W., T. Shemesh, Y. Chen, *et al.* (June 5, 2014). "A "Push and Slide" Mechanism Allows Sequence-Insensitive Translocation of Secretory Proteins by the SecA ATPase". In: *Cell* 157.6, pp. 1416–1429. ISSN: 1097-4172. pmid: 24906156.

Bechtluft, P., A. Kedrov, D.-J. Slotboom, *et al.* (Mar. 23, 2010). "Tight Hydrophobic Contacts with the SecB Chaperone Prevent Folding of Substrate Proteins". In: *Biochemistry* 49.11, pp. 2380–2388. ISSN: 1520-4995. pmid: 20146530.

Bechtluft, P., N. Nouwen, S. J. Tans, *et al.* (2010). "SecB—A Chaperone Dedicated to Protein Translocation". In: *Mol. BioSyst.* 6.4, pp. 620–627. ISSN: 1742-206X, 1742-2051. URL: http://xlink.rsc.org/?DOI=B915435C (visited on 01/26/2017).

Bell, A. J., L. K. Frankel, and T. M. Bricker (July 24, 2015). "High Yield Non-Detergent Isolation of Photosystem I-Light-Harvesting Chlorophyll II Membranes from Spinach Thylakoids". In: *The Journal of Biological Chemistry* 290.30, pp. 18429–18437. ISSN: 0021-9258. pmid: 26055710. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4513103/ (visited on 11/16/2017).

Berg, J. M., J. L. Tymoczko, and L. Stryer (2002). "Amino Acid Biosynthesis Is Regulated by Feedback Inhibition". In: URL: https://www.ncbi.nlm.nih.gov/books/NBK22371/ (visited on 03/12/2018).

Berger, I., F. Garzoni, M. Chaillet, *et al.* (July 11, 2013). "The multiBac Protein Complex Production Platform at the EMBL". In: *Journal of Visualized Experiments: JoVE* 77, e50159. ISSN: 1940-087X. pmid: 23892976.

Berger, I., Q. Jiang, R. Schulze, *et al.* (2017). "Multiprotein Complex Production in E. Coli: The SecYEG-SecDFYajC-YidC Holotranslocon". In: *Methods in Molecular Biology (Clifton, N.J.)* 1586, pp. 279–290. ISSN: 1940-6029. pmid: 28470612.

Bieniossek, C., Y. Nie, D. Frey, *et al.* (June 2009). "Automated Unrestricted Multigene Recombineering for Multiprotein Complex Production". In: *Nature Methods* 6.6, pp. 447–450. ISSN: 1548-7091. URL: http://www.nature.com.gate1.inist.fr/nmeth/journal/v6/n6/full/nmeth.1326.html (visited on 10/09/2014).

Bird, L. E., J. E. Nettleship, V. Järvinen, *et al.* (2016). "Expression Screening of Integral Membrane Proteins by Fusion to Fluorescent Reporters". In: *Advances in Experimental Medicine and Biology* 922, pp. 1–11. ISSN: 0065-2598. pmid: 27553231.

Bligh, E. G. and W. J. Dyer (Aug. 1959). "A Rapid Method of Total Lipid Extraction and Purification". In: *Canadian Journal of Biochemistry and Physiology* 37.8, pp. 911–917. pmid: 13671378.

Botte, M. (Dec. 13, 2013). "Architechture of the SecYEG-DF-YajC-YidC Holotranslocon". Université de Grenoble. 177 pp.

Botte, M., N. R. Zaccai, J. L. à. Nijeholt, *et al.* (Dec. 7, 2016). "A Central Cavity within the Holo-Translocon Suggests a Mechanism for Membrane Protein Insertion". In: *Scientific Reports* 6. ISSN: 2045-2322. pmid: 27924919. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5141469/ (visited on 03/10/2017).

Breyton, C., W. Haase, T. A. Rapoport, *et al.* (Aug. 8, 2002). "Three-Dimensional Structure of the Bacterial Protein-Translocation Complex SecYEG". In: *Nature* 418.6898, pp. 662–665. ISSN: 0028-0836. URL: http://www.nature.com.gate1.inist.fr/nature/journal/v418/n6898/full/nature00827.html (visited on 03/28/2014).

Brockwell, D., L. Yu, S. Cooper, *et al.* (Mar. 2001). "Physicochemical Consequences of the Perdeuteriation of Glutathione S-Transferase from S. Japonicum". In: *Protein Science : A Publication of the Protein Society* 10.3, pp. 572–580. ISSN: 0961-8368. pmid: 11344325. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2374125/ (visited on 03/11/2018).

Cameron, D. E. and J. J. Collins (Dec. 2014). "Tunable Protein Degradation in Bacteria". In: *Nature biotechnology* 32.12, pp. 1276–1281. ISSN: 1087-0156. pmid: 25402616. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4262603/ (visited on 03/12/2018).

Chaptal, V., F. Delolme, A. Kilburg, *et al.* (Feb. 8, 2017). "Quantification of Detergents Complexed with Membrane Proteins". In: *Scientific Reports* 7. ISSN: 2045-2322. pmid: 28176812. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5297245/ (visited on 02/12/2018).

Chatzi, K., M. Sardis, A. Economou, *et al.* (Aug. 2014). "SecA-Mediated Targeting and Translocation of Secretory Proteins". In: *Biochimica Et Biophysica Acta* 1843.8, pp. 1466–1474. ISSN: 0006-3002. pmid: 24583121.

Chatzi, K., M. Sardis, A. Tsirigotaki, *et al.* (May 1, 2017). "Preprotein Mature Domains Contain Translocase Targeting Signals That Are Essential for Secretion". In: *The Journal of Cell Biology* 216.5, pp. 1357–1369. ISSN: 0021-9525, 1540-8140. URL: http://www.jcb.org/lookup/doi/10.1083/jcb.201609022 (visited on 02/09/2018).

Chen, I., B. M. Dorr, and D. R. Liu (July 12, 2011). "A General Strategy for the Evolution of Bond-Forming Enzymes Using Yeast Display". In: *Proceedings of the National Academy of Sciences of the United States of America* 108.28, pp. 11399–11404. ISSN: 0027-8424. pmid: 21697512. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3136257/ (visited on 02/21/2018).

Chen, M., J. Samuelson, F. Jiang, *et al.* (Mar. 8, 2002). "Direct Interaction of YidC with the Sec-Independent Pf3 Coat Protein during Its Membrane Protein Insertion". In: *The Journal of Biological Chemistry* 277.10, pp. 7670–7675. ISSN: 0021-9258. pmid: 11751917.

Clark, V. L., D. E. Peterson, and R. W. Bernlohr (Nov. 1972). "Changes in Free Amino Acid Production and Intracellular Amino Acid Pools of Bacillus Licheniformis as a Function of Culture Age and Growth Media". In: *Journal of Bacteriology* 112.2, pp. 715–725. ISSN: 0021-9193. pmid: 5086658. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC251479/ (visited on 03/11/2018).

Dalbey, R. E., A. Kuhn, L. Zhu, *et al.* (Aug. 2014). "The Membrane Insertase YidC". In: *Biochimica Et Biophysica Acta* 1843.8, pp. 1489–1496. ISSN: 0006-3002. pmid: 24418623.

De Ghellinck, A., H. Schaller, V. Laux, *et al.* (Apr. 18, 2014). "Production and Analysis of Perdeuterated Lipids from Pichia Pastoris Cells". In: *PLoS ONE* 9.4. ISSN: 1932-6203. pmid: 24747350. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3991571/ (visited on 08/07/2017).

Dekker, C., B. de Kruijff, and P. Gros (Dec. 1, 2003). "Crystal Structure of SecB from Escherichia Coli". In: *Journal of Structural Biology* 144.3, pp. 313–319. ISSN: 1047-8477. URL: http://www.sciencedirect.com/science/article/pii/S1047847703001679 (visited on 02/14/2018).

Del Solar, G. and M. Espinosa (Aug. 1, 2000). "Plasmid Copy Number Control: An Ever-growing Story". In: *Molecular Microbiology* 37.3, pp. 492–500. ISSN: 1365-2958. URL: http://onlinelibrary.wiley.com/doi/10.1046/j.1365-2958.2000.02005.x/full (visited on 02/21/2018).

Denisov, I. G., Y. V. Grinkova, A. A. Lazarides, *et al.* (Mar. 1, 2004). "Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size". In: *Journal of the American Chemical Society* 126.11, pp. 3477–3487. ISSN: 0002-7863. URL: http://dx.doi.org/10.1021/ja0393574 (visited on 04/18/2014).

Denisov, I. G. and S. G. Sligar (Mar. 22, 2017). "Nanodiscs in Membrane Biochemistry and Biophysics". In: *Chemical Reviews* 117.6, pp. 4669–4713. ISSN: 0009-2665. URL: http://dx.doi.org/10.1021/acs.chemrev.6b00690 (visited on 09/11/2017).

Diao, L., Q. Dong, Z. Xu, *et al.* (Jan. 2, 2012). "Functional Implementation of the Posttranslational SecB-SecA Protein-Targeting Pathway in Bacillus Subtilis". In: *Applied and Environmental Microbiology* 78.3, pp. 651–659. ISSN: 0099-2240, 1098-5336. pmid: 22113913. URL: http://aem.asm.org/content/78/3/651 (visited on 03/08/2018).

Driessen, A. and N. Nouwen (2008). "Protein Translocation Across the Bacterial Cytoplasmic Membrane". In: *Annual Review of Biochemistry* 77.1, pp. 643–667. pmid: 18078384. URL: http://dx.doi.org/10.1146/annurev.biochem.77.061606.160747 (visited on 01/27/2015).

Dumon-Seignovert, L., G. Cariot, and L. Vuillard (Sept. 2004). "The Toxicity of Recombinant Proteins in Escherichia Coli: A Comparison of Overexpression in BL21(DE3), C41(DE3), and C43(DE3)". In: *Protein Expression and Purification* 37.1, pp. 203–206. ISSN: 1046-5928. pmid: 15294299.

Duong, F. (Sept. 1, 2003). "Binding, Activation and Dissociation of the Dimeric SecA ATPase at the Dimeric SecYEG Translocase". In: *The EMBO Journal* 22.17, pp. 4375–4384. ISSN: 0261-4189. pmid: 12941690. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC202361/ (visited on 02/21/2018).

Duong, F. and W. Wickner (1997a). "Distinct Catalytic Roles of the SecYE, SecG and SecDFyajC Subunits of Preprotein Translocase Holoenzyme". In: *The EMBO journal* 16.10, pp. 2756–2768. URL: http://onlinelibrary.wiley.com/doi/10.1093/emboj/16.10.2756/full (visited on 04/02/2014).

– (Aug. 15, 1997b). "The SecDFyajC Domain of Preprotein Translocase Controls Preprotein Movement by Regulating SecA Membrane Cycling". In: *The EMBO Journal* 16.16, pp. 4871–4879. URL: http://emboj.embopress.org/content/16/16/4871 (visited on 04/11/2014).

Dyson, H. J., P. E. Wright, and H. A. Scheraga (Aug. 29, 2006). "The Role of Hydrophobic Interactions in Initiation and Propagation of Protein Folding". In: *Proceedings of the National Academy of Sciences of the United States of America* 103.35, pp. 13057–13061. ISSN: 0027-8424. pmid: 16916929. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1559752/ (visited on 04/25/2018).

Erlandson, K. J., S. B. M. Miller, Y. Nam, *et al.* (Oct. 16, 2008). "A Role for the Two-Helix Finger of the SecA ATPase in Protein Translocation". In: *Nature* 455.7215, pp. 984–987. ISSN: 0028-0836. URL: http://www.nature.com.gate1.inist.fr/nature/journal/v455/n7215/full/nature07439.html (visited on 03/28/2014).

Ermolaeva, M. D. (2001). "Synonymous Codon Usage in Bacteria". In: *Current issues in molecular biology* 3.4, pp. 91–97.

Ernst, S., A.-K. Schönbauer, G. Bär, *et al.* (Sept. 16, 2011). "YidC-Driven Membrane Insertion of Single Fluorescent Pf3 Coat Proteins". In: *Journal of Molecular Biology* 412.2, pp. 165–175. ISSN: 1089-8638. pmid: 21798266.

Esmaili, M. and M. Overduin (Feb. 2018). "Membrane Biology Visualized in Nanometer-Sized Discs Formed by Styrene Maleic Acid Polymers". In: *Biochimica Et Biophysica Acta* 1860.2, pp. 257–263. ISSN: 0006-3002. pmid: 29056560.

Facey, S. J., S. A. Neugebauer, S. Krauss, *et al.* (Jan. 26, 2007). "The Mechanosensitive Channel Protein MscL Is Targeted by the SRP to the Novel YidC Membrane Insertion Pathway of Escherichia Coli". In: *Journal of Molecular Biology* 365.4, pp. 995–1004. ISSN: 0022-2836. pmid: 17113597.

Farley, M. M., B. Hu, W. Margolin, *et al.* (Apr. 15, 2016). "Minicells, Back in Fashion". In: *Journal of Bacteriology* 198.8, pp. 1186–1195. ISSN: 0021-9193, 1098-5530. pmid: 26833418. URL: http://jb.asm.org/content/198/8/1186 (visited on 07/07/2017).

Fisher, S. J., J. R. Helliwell, S. Khurshid, *et al.* (June 1, 2008). "An Investigation into the Protonation States of the C1 Domain of Cardiac Myosin-Binding Protein C". In: *Acta Crystallographica Section D: Biological Crystallography* 64.6, pp. 658–664. ISSN: 0907-4449. URL: http://scripts.iucr.org/cgi-bin/paper?fw5170 (visited on 03/11/2018).

Frauenfeld, J., J. Gumbart, E. O. van der Sluis, *et al.* (Apr. 17, 2011). "Cryo-EM Structure of the Ribosome–SecYE Complex in the Membrane Environment". In: *Nature Structural & Molecular Biology* 18.5, pp. 614–621. ISSN: 1545-9993, 1545-9985. URL: http://www.nature.com/doifinder/10.1038/nsmb.2026 (visited on 01/18/2014).

Furukawa, A., K. Yoshikaie, T. Mori, *et al.* (May 2, 2017). "Tunnel Formation Inferred from the I-Form Structures of the Proton-Driven Protein Secretion Motor SecDF". In: *Cell Reports* 19.5, pp. 895–901. ISSN: 2211-1247. pmid: 28467902. URL: http://www.cell.com/cell-reports/abstract/S2211-1247(17)30524-7 (visited on 02/22/2018).

Gardel, C., K. Johnson, A. Jacq, *et al.* (Oct. 1990). "The secD Locus of E.Coli Codes for Two Membrane Proteins Required for Protein Export." In: *The EMBO Journal* 9.10, pp. 3209–3216. ISSN: 0261-4189. pmid: 2170107. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC552051/ (visited on 01/20/2018).

Ge, Y., A. Draycheva, T. Bornemann, *et al.* (Oct. 15, 2014). "Lateral Opening of the Bacterial Translocon on Ribosome Binding and Signal Peptide Insertion". In: *Nature Communications* 5, p. 5263. ISSN: 2041-1723. URL: http://www.nature.com/doifinder/10.1038/ncomms6263 (visited on 10/16/2014).

Gelis, I., A. M. J. J. Bonvin, D. Keramisanou, *et al.* (Nov. 16, 2007). "Structural Basis for Signal-Sequence Recognition by the Translocase Motor SecA as Determined by NMR". In: *Cell* 131.4, pp. 756–769. ISSN: 0092-8674. pmid: 18022369.

Gmeiner, J. and S. Schlecht (Sept. 1980). "Molecular Composition of the Outer Membrane of Escherichia Coli and the Importance of Protein-Lipopolysaccharide Interactions". In: *Archives of Microbiology* 127.2, pp. 81–86. ISSN: 0302-8933, 1432-072X. URL: http://link.springer.com/10.1007/BF00428010 (visited on 04/29/2018).

Gogala, M., T. Becker, B. Beatrix, *et al.* (Feb. 6, 2014). "Structures of the Sec61 Complex Engaged in Nascent Peptide Translocation or Membrane Insertion". In: *Nature* 506.7486, pp. 107–110. ISSN: 0028-0836. URL: http://www.nature.com.gate1.inist.fr/nature/journal/v506/n7486/full/nature12950.html (visited on 03/28/2014).

Gohon, Y., T. Dahmane, R. W. Ruigrok, *et al.* (May 2008). "Bacteriorhodopsin/Amphipol Complexes: Structural and Functional Properties". In: *Biophysical Journal* 94.9, pp. 3523–3537. ISSN: 00063495. URL: http://linkinghub.elsevier.com/retrieve/pii/S0006349508704302 (visited on 10/17/2014).

Gohon, Y., F. Giusti, C. Prata, *et al.* (Jan. 31, 2006). "Well-Defined Nanoparticles Formed by Hydrophobic Assembly of a Short and Polydisperse Random Terpolymer, Amphipol A8-35". In: *Langmuir: the ACS journal of surfaces and colloids* 22.3, pp. 1281–1290. ISSN: 0743-7463. pmid: 16430295.

Gohon, Y., G. Pavlov, P. Timmins, *et al.* (Nov. 15, 2004). "Partial Specific Volume and Solvent Interactions of Amphipol A8-35". In: *Analytical Biochemistry* 334.2, pp. 318–334. ISSN: 0003-2697. URL: http://www.sciencedirect.com/science/article/pii/S0003269704006451 (visited on 10/17/2014).

Gold, V. A. M., S. Whitehouse, A. Robson, *et al.* (Feb. 1, 2013). "The Dynamic Action of SecA during the Initiation of Protein Translocation". In: *Biochemical Journal* 449 (Pt 3), pp. 695–705. ISSN: 0264-6021. pmid: 23126322. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3685266/ (visited on 10/23/2017).

Gouridis, G., S. Karamanou, M. F. Sardis, *et al.* (Dec. 12, 2013). "Quaternary Dynamics of the SecA Motor Drive Translocase Catalysis". In: *Molecular Cell* 52.5, pp. 655–666. ISSN: 1097-2765. pmid: 24332176. URL: http://www.cell.com/molecular-cell/abstract/S1097-2765(13)00860-5 (visited on 01/17/2018).

Gray, A. N., J. M. Henderson-Frost, D. Boyd, *et al.* (2011). "Unbalanced Charge Distribution as a Determinant for Dependence of a Subset of Escherichia Coli Membrane Proteins on the Membrane Insertase YidC". In: *mBio* 2.6. ISSN: 2150-7511. pmid: 22108384.

Grinkova, Y. V., I. G. Denisov, and S. G. Sligar (Nov. 2010). "Engineering Extended Membrane Scaffold Proteins for Self-Assembly of Soluble Nanoscale Lipid Bilayers". In: *Protein Engineering, Design and Selection* 23.11, pp. 843–848. ISSN: 1741-0126. pmid: 20817758. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2953958/ (visited on 03/05/2018).

Haertlein, M., M. Moulin, J. M. Devos, *et al.* (2016). "Chapter Five - Biomolecular Deuteration for Neutron Structural Biology and Dynamics". In: *Methods in Enzymology*. Ed. by Z. Kelman. Vol. 566. Isotope Labeling of Biomolecules - Applications. Academic Press, pp. 113–157. URL: http://www.sciencedirect.com/science/article/pii/S0076687915006370 (visited on 03/10/2017).

Hagn, F., M. Etzkorn, T. Raschle, *et al.* (Feb. 6, 2013). "Optimized Phospholipid Bilayer Nanodiscs Facilitate High-Resolution Structure Determination of Membrane Proteins". In: *Journal of the American Chemical Society* 135.5, pp. 1919–1925. ISSN: 0002-7863. URL: https://doi.org/10.1021/ja310901f (visited on 03/05/2018).

Hardy, D., R. Bill, A. Jawhari, *et al.* (June 15, 2016). "Overcoming Bottlenecks in the Membrane Protein Structural Biology Pipeline". In: *Biochemical Society Transactions* 44.3, pp. 838–844. ISSN: 0300-5127, 1470-8752. pmid: 27284049. URL: http://www.biochemsoctrans.org/content/44/3/838 (visited on 11/16/2017).

Hardy, R. C. and R. L. Cottington (1949). "Viscosity of Deuterium Oxide and Water in the Range 5 to 125 Degree Celsius". In: URL: http://archive.org/details/jresv42n6p573 (visited on 12/19/2017).

Hartl, F. U., S. Lecker, E. Schiebel, *et al.* (Oct. 19, 1990). "The Binding Cascade of SecB to SecA to SecY/E Mediates Preprotein Targeting to the E. Coli Plasma Membrane". In: *Cell* 63.2, pp. 269–279. ISSN: 0092-8674. pmid: 2170023.

He, F. (2011). "E. Coli Genomic DNA Extraction". In: *BIO-PROTOCOL* 1.14. ISSN: 2331-8325. URL: http://www.bio-protocol.org/e97 (visited on 12/16/2016).

Hessa, T., H. Kim, K. Bihlmaier, *et al.* (Jan. 27, 2005). "Recognition of Transmembrane Helices by the Endoplasmic Reticulum Translocon". In: *Nature* 433.7024, pp. 377–381. ISSN: 1476-4687. pmid: 15674282.

Hizlan, D., A. Robson, S. Whitehouse, *et al.* (Jan. 26, 2012). "Structure of the SecY Complex Unlocked by a Preprotein Mimic". In: *Cell Reports* 1.1, pp. 21–28. ISSN: 2211-1247. pmid: 22576621. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3333808/ (visited on 01/16/2018).

Hochuli, M., T. Szyperski, and K. Wüthrich (May 1, 2000). "Deuterium Isotope Effects on the Central Carbon Metabolism of Escherichia Coli Cells Grown on a D2O-Containing Minimal Medium". In: *Journal of Biomolecular NMR* 17.1, pp. 33–42. ISSN: 0925-2738, 1573-5001. URL: https://link.springer.com/article/10.1023/A:1008329124672 (visited on 03/11/2018).

Hou, B., P.-J. Lin, and A. E. Johnson (Nov. 9, 2012). "Membrane Protein TM Segments Are Retained at the Translocon during Integration until the Nascent Chain Cues FRET-Detected Release into the Lipid Phase". In: *Molecular cell* 48.3, pp. 398–408. ISSN: 1097-2765. pmid: 23022384. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3496027/ (visited on 01/17/2018).

Huang, C., P. Rossi, T. Saio, *et al.* (Sept. 8, 2016). "Structural Basis for the Antifolding Activity of a Molecular Chaperone". In: *Nature* 537.7619, pp. 202–206. ISSN: 0028-0836. URL: http://www.nature.com/nature/journal/v537/n7619/full/nature18965.html (visited on 01/26/2017).

Hunt, J. F., S. Weinkauf, L. Henry, *et al.* (Sept. 20, 2002). "Nucleotide Control of Interdomain Interactions in the Conformational Reaction Cycle of SecA". In: *Science (New York, N.Y.)* 297.5589, pp. 2018–2026. ISSN: 1095-9203. pmid: 12242434.

Ibel, K. and H. B. Stuhrmann (Apr. 5, 1975). "Comparison of Neutron and X-Ray Scattering of Dilute Myoglobin Solutions". In: *Journal of Molecular Biology* 93.2, pp. 255–265. ISSN: 0022-2836. URL: http://www.sciencedirect.com/science/article/pii/002228367590131X (visited on 02/21/2018).

Isola, N. R., S. L. Allman, V. V. Golovlev, *et al.* (May 1, 2001). "MALDI-TOF Mass Spectrometric Method for Detection of Hybridized DNA Oligomers". In: *Analytical Chemistry* 73.9, pp. 2126–2131. ISSN: 0003-2700. pmid: 11354500.

Jacrot, B. and Zaccai (Nov. 1, 1981). "Determination of Molecular Weight by Neutron Scattering". In: *Biopolymers* 20.11, pp. 2413–2426. ISSN: 1097-0282. URL: http://onlinelibrary.wiley.com.gate1.inist.fr/doi/10.1002/bip.1981.360201110/abstract (visited on 01/13/2016).

Jamshad, M., J. Charlton, Y.-P. Lin, *et al.* (Apr. 16, 2015). "G-Protein Coupled Receptor Solubilization and Purification for Biophysical Analysis and Functional Studies, in the Total Absence of Detergent". In: *Bioscience Reports* 35.2. ISSN: 0144-8463. pmid: 25720391. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4400634/ (visited on 11/16/2017).

Jamshad, M., V. Grimard, I. Idini, *et al.* (Mar. 2015). "Structural Analysis of a Nanoparticle Containing a Lipid Bilayer Used for Detergent-Free Extraction of Membrane Proteins". In: *Nano Research* 8.3, pp. 774–789. ISSN: 1998-0124, 1998-0000. URL: http://link.springer.com/10.1007/s12274-014-0560-6 (visited on 09/17/2015).

Jasnin, M., M. Tehei, M. Moulin, *et al.* (June 1, 2008). "Solvent Isotope Effect on Macromolecular Dynamics in E. Coli". In: *European Biophysics Journal* 37.5, pp. 613–617. ISSN: 0175-7571, 1432-1017. URL: https://link.springer.com/article/10.1007/s00249-008-0281-4 (visited on 01/28/2018).

Jia, L., M. Dienhart, M. Schramp, *et al.* (Dec. 15, 2003). "Yeast Oxa1 Interacts with Mitochondrial Ribosomes: The Importance of the C-Terminal Region of Oxa1". In: *The EMBO journal* 22.24, pp. 6438–6447. ISSN: 0261-4189. pmid: 14657017.

Jiang, F., M. Chen, L. Yi, *et al.* (Dec. 5, 2003). "Defining the Regions of Escherichia Coli YidC That Contribute to Activity". In: *The Journal of Biological Chemistry* 278.49, pp. 48965–48972. ISSN: 0021-9258. pmid: 14506280.

Jomaa, A., D. Boehringer, M. Leibundgut, *et al.* (Jan. 25, 2016). "Structures of the E. Coli Translating Ribosome with SRP and Its Receptor and with the Translocon". In: *Nature Communications* 7. ISSN: 2041-1723. pmid: 26804923. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4737761/ (visited on 01/17/2018).

Jordan, A., M. Jacques, C. Merrick, *et al.* (Nov. 2, 2016). "SEC-SANS: Size Exclusion Chromatography Combined in Situ with Small-Angle Neutron Scattering". In: *Journal of Applied Crystallography* 49 (Pt 6), pp. 2015–2020. ISSN: 0021-8898. pmid: 27980509. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5139991/ (visited on 04/14/2018).

Jünemann, R., J. Wadzack, F. J. Triana-Alonso, *et al.* (Jan. 1, 1996). "In Vivo Deuteration of Transfer RNAs: Overexpression and Large-Scale Purification of Deuterated Specific tRNAs". In: *Nucleic Acids Research* 24.5, pp. 907–913. ISSN: 0305-1048. URL: https://academic.oup.com/nar/article/24/5/907/1046025 (visited on 01/28/2018).

Junne, T., T. Schwede, V. Goder, *et al.* (Sept. 11, 2007). "Mutations in the Sec61p Channel Affecting Signal Sequence Recognition and Membrane Protein Topology". In: *Journal of Biological Chemistry* 282.45, pp. 33201–33209. ISSN: 0021-9258, 1083-351X. pmid: 17893139. URL: http://www.jbc.org/content/282/45/33201 (visited on 01/14/2018).

Kadokura, H. and J. Beckwith (Sept. 18, 2009). "Detecting Folding Intermediates of a Protein as It Passes Through the Bacterial Translocation Channel". In: *Cell* 138.6, pp. 1164–1173. ISSN: 0092-8674. pmid: 19766568. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2750780/ (visited on 01/17/2018).

Kaiser, C. M., H.-C. Chang, V. R. Agashe, *et al.* (Nov. 23, 2006). "Real-Time Observation of Trigger Factor Function on Translating Ribosomes". In: *Nature* 444.7118, pp. 455–460. ISSN: 1476-4687. pmid: 17051157.

Karamanou, S., G. Gouridis, E. Papanikou, *et al.* (June 20, 2007). "Preprotein-Controlled Catalysis in the Helicase Motor of SecA". In: *The EMBO journal* 26.12, pp. 2904–2914. ISSN: 0261-4189. pmid: 17525736.

Kedrov, A., M. Sustarsic, J. de Keyzer, *et al.* (Nov. 15, 2013). "Elucidating the Native Architecture of the YidC: Ribosome Complex". In: *Journal of Molecular Biology* 425.22, pp. 4112–4124. ISSN: 1089-8638. pmid: 23933010.

Kiefer, D. and A. Kuhn (Nov. 15, 1999). "Hydrophobic Forces Drive Spontaneous Membrane Insertion of the Bacteriophage Pf3 Coat Protein without Topological Control". In: *The EMBO journal* 18.22, pp. 6299–6306. ISSN: 0261-4189. pmid: 10562542.

Komar, J., S. Alvira, R. J. Schulze, *et al.* (Oct. 1, 2016). "Membrane Protein Insertion and Assembly by the Bacterial Holo-Translocon SecYEG-SecDF-YajC-YidC". In: *Biochemical Journal* 473.19, pp. 3341–3354. ISSN: 0264-6021, 1470-8728. URL: http://biochemj.org/cgi/doi/10.1042/BCJ20160545 (visited on 09/29/2016).

Komar, J., M. Botte, I. Collinson, *et al.* (2015). "ACEMBLing a Multiprotein Transmembrane Complex: The Functional SecYEG-SecDF-YajC-YidC Holotranslocon Protein Secretase/Insertase". In: *Methods in Enzymology* 556, pp. 23–49. ISSN: 1557-7988. pmid: 25857776.

Koshland, D. E. (Mar. 22, 2002). "The Seven Pillars of Life". In: *Science* 295.5563, pp. 2215–2216. ISSN: 0036-8075, 1095-9203. pmid: 11910092. URL: http://science.sciencemag.org/content/295/5563/2215 (visited on 01/12/2018).

Kuhn, P., B. Weiche, L. Sturm, *et al.* (May 2011). "The Bacterial SRP Receptor, SecA and the Ribosome Use Overlapping Binding Sites on the SecY Translocon". In: *Traffic (Copenhagen, Denmark)* 12.5, pp. 563–578. ISSN: 1600-0854. pmid: 21255212.

Kumazaki, K., T. Kishimoto, A. Furukawa, *et al.* (Dec. 3, 2014). "Crystal Structure of Escherichia Coli YidC, a Membrane Protein Chaperone and Insertase". In: *Scientific Reports* 4. URL: http://www.nature.com.gate1.inist.fr/srep/2014/141203/srep07299/full/srep07299.html (visited on 01/12/2015).

Kuruma, Y. and T. Ueda (Aug. 13, 2015). "The PURE System for the Cell-Free Synthesis of Membrane Proteins". In: *Nature Protocols* 10.9, pp. 1328–1344. ISSN: 1754-2189, 1750-2799. URL: http://www.nature.com/doifinder/10.1038/nprot.2015.082 (visited on 08/24/2015).

Laux, V., P. Callow, D. I. Svergun, *et al.* (Feb. 15, 2008). "Selective Deuteration of Tryptophan and Methionine Residues in Maltose Binding Protein: A Model System for Neutron Scattering". In: *European Biophysics Journal* 37.6, pp. 815–822. ISSN: 0175-7571, 1432-1017. URL: http://link.springer.com.gate1.inist.fr/article/10.1007/s00249-008-0280-5 (visited on 02/08/2016).

Lee, H. C. and H. D. Bernstein (Aug. 11, 2002). "Trigger Factor Retards Protein Export in Escherichia Coli". In: *Journal of Biological Chemistry* 277.45, pp. 43527–43535. ISSN: 0021-9258, 1083-351X. pmid: 12205085. URL: http://www.jbc.org/content/277/45/43527 (visited on 01/17/2018).

Leiting, B., F. Marsilio, and J. F. O'Connell (Dec. 15, 1998). "Predictable Deuteration of Recombinant Proteins Expressed inEscherichia Coli". In: *Analytical Biochemistry* 265.2, pp. 351–355. ISSN: 0003-2697. URL: http://www.sciencedirect.com/science/article/pii/S0003269798929049 (visited on 11/12/2015).

Li, G.-W., D. Burkhardt, C. Gross, *et al.* (Apr. 24, 2014). "Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources". In: *Cell* 157.3, pp. 624–635. ISSN: 0092-8674. pmid: 24766808. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4006352/ (visited on 02/21/2018).

Li, J. and Y. Zhang (Sept. 2014). "Relationship between Promoter Sequence and Its Strength in Gene Expression". In: *The European Physical Journal. E, Soft Matter* 37.9, p. 44. ISSN: 1292-895X. pmid: 25260329.

Li, L., E. Park, J. Ling, *et al.* (Mar. 17, 2016). "Crystal Structure of a Substrate-Engaged SecY Protein-Translocation Channel". In: *Nature* 531.7594, pp. 395–399. ISSN: 1476-4687. pmid: 26950603.

Lill, R., W. Dowhan, and W. Wickner (Jan. 26, 1990). "The ATPase Activity of SecA Is Regulated by Acidic Phospholipids, SecY, and the Leader and Mature Domains of Precursor Proteins". In: *Cell* 60.2, pp. 271–280. ISSN: 0092-8674. pmid: 2153463.

Lim, H. N., Y. Lee, and R. Hussein (June 28, 2011). "Fundamental Relationship between Operon Organization and Gene Expression". In: *Proceedings of the National Academy of Sciences of the United States of America* 108.26, pp. 10626–10631. ISSN: 1091-6490. pmid: 21670266.

Liu, J., C.-Y. Chen, D. Shiomi, *et al.* (Sept. 1, 2011). "Visualization of Bacteriophage P1 Infection by Cryo-Electron Tomography of Tiny Escherichia Coli". In: *Virology* 417.2, pp. 304–311. ISSN: 0042-6822. pmid: 21745674. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3163801/ (visited on 07/07/2017).

Long, A. R., C. C. O'Brien, K. Malhotra, *et al.* (2013). "A Detergent-Free Strategy for the Reconstitution of Active Enzyme Complexes from Native Biological Membranes into Nanoscale Discs". In: *BMC biotechnology* 13.1, p. 41. URL: http://www.biomedcentral.com/1472-6750/13/41/ (visited on 07/07/2015).

Lopez, C. G., T. Watanabe, A. Martel, *et al.* (Jan. 12, 2015). "Microfluidic-SANS: Flow Processing of Complex Fluids". In: *Scientific Reports* 5, p. 7727. ISSN: 2045-2322. URL: https://www.nature.com/articles/srep07727 (visited on 04/14/2018).

Ludlam, A. V., B. A. Moore, and Z. Xu (Sept. 14, 2004). "The Crystal Structure of Ribosomal Chaperone Trigger Factor from Vibrio Cholerae". In: *Proceedings of the National Academy of Sciences of the United States of America* 101.37, pp. 13436–13441. ISSN: 0027-8424. pmid: 15353602. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC518775/ (visited on 02/14/2018).

Maric, S., M. B. Thygesen, J. Schiller, *et al.* (2015). "Biosynthetic Preparation of Selectively Deuterated Phosphatidylcholine in Genetically Modified Escherichia Coli". In: *Applied microbiology and biotechnology* 99.1, pp. 241–254. URL: http://link.springer.com/article/10.1007/s00253-014-6082-z (visited on 07/02/2015).

Marion, D. (Nov. 2013). "An Introduction to Biological NMR Spectroscopy". In: *Molecular & Cellular Proteomics : MCP* 12.11, pp. 3006–3025. ISSN: 1535-9476. pmid: 23831612. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3820920/ (visited on 02/21/2018).

Matlack, K. E., B. Misselwitz, K. Plath, *et al.* (May 28, 1999). "BiP Acts as a Molecular Ratchet during Posttranslational Transport of Prepro-Alpha Factor across the ER Membrane". In: *Cell* 97.5, pp. 553–564. ISSN: 0092-8674. pmid: 10367885.

Matsubayashi, H., Y. Kuruma, and T. Ueda (Dec. 2014a). "Cell-Free Synthesis of SecYEG Translocon as the Fundamental Protein Transport Machinery". In: *Origins of Life and Evolution of Biospheres* 44.4, pp. 331–334. ISSN: 0169-6149, 1573-0875. URL: http://link.springer.com/10.1007/s11084-014-9389-y (visited on 10/12/2016).

– (July 14, 2014b). "In Vitro Synthesis of the E. Coli Sec Translocon from DNA". In: *Angewandte Chemie International Edition* 53.29, pp. 7535–7538. ISSN: 14337851. URL: http://doi.wiley.com/10.1002/anie.201403929 (visited on 05/04/2018).

Matsuyama, S., Y. Fujita, K. Sagara, *et al.* (July 13, 1992). "Overproduction, Purification and Characterization of SecD and SecF, Integral Membrane Components of the Protein Translocation Machinery of Escherichia Coli". In: *Biochimica Et Biophysica Acta* 1122.1, pp. 77–84. ISSN: 0006-3002. pmid: 1633199.

Merz, F., D. Boehringer, C. Schaffitzel, *et al.* (June 4, 2008). "Molecular Mechanism and Structure of Trigger Factor Bound to the Translating Ribosome". In: *The EMBO Journal* 27.11, pp. 1622–1632. ISSN: 0261-4189. pmid: 18497744. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2426727/ (visited on 01/17/2018).

Miroux, B. and J. E. Walker (July 19, 1996). "Over-Production of Proteins in Escherichia Coli: Mutant Hosts That Allow Synthesis of Some Membrane Proteins and Globular Proteins at High Levels". In: *Journal of Molecular Biology* 260.3, pp. 289–298. ISSN: 0022-2836. pmid: 8757792.

Misselwitz, B., O. Staeck, and T. A. Rapoport (Nov. 1998). "J Proteins Catalytically Activate Hsp70 Molecules to Trap a Wide Range of Peptide Sequences". In: *Molecular Cell* 2.5, pp. 593–603. ISSN: 1097-2765. pmid: 9844632.

Mitra, K., J. Frank, and A. Driessen (Nov. 2006). "Co- and Post-Translational Translocation through the Protein-Conducting Channel: Analogous Mechanisms at Work?" In: *Nature Structural & Molecular Biology* 13.11, pp. 957–964. ISSN: 1545-9993, 1545-9985. URL: http://www.nature.com/articles/nsmb1166 (visited on 01/16/2018).

Mitra, K., C. Schaffitzel, T. Shaikh, *et al.* (Nov. 17, 2005). "Structure of the E. Coli Protein-Conducting Channel Bound to a Translating Ribosome". In: *Nature* 438.7066, pp. 318–324. ISSN: 0028-0836, 1476-4687. URL: http://www.ncbi.nlm.nih.gov.gate1.inist.fr/pmc/articles/PMC1351281/ (visited on 03/28/2014).

Moraes, I., ed. (2016). *The Next Generation in Membrane Protein Structure Determination*. Advances in Experimental Medicine and Biology. Springer International Publishing. ISBN: 978-3-319-35070-7. URL: //www.springer.com/gp/book/9783319350707 (visited on 03/01/2018).

Moraes, I., G. Evans, J. Sanchez-Weatherby, *et al.* (Jan. 2014). "Membrane Protein Structure Determination — The next Generation". In: *Biochimica et Biophysica Acta* 1838.1, pp. 78–87. ISSN: 0006-3002. pmid: 23860256. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3898769/ (visited on 11/16/2017).

Morein, S., A.-S. Andersson, L. Rilfors, *et al.* (Mar. 22, 1996). "Wild-Type Escherichia Coli Cells Regulate the Membrane Lipid Composition in a Window between Gel and Non-Lamellar Structures". In: *Journal of Biological Chemistry* 271.12, pp. 6801–6809. ISSN: 0021-9258, 1083-351X. pmid: 8636103. URL: http://www.jbc.org/content/271/12/6801 (visited on 02/21/2018).

Mulder, K. C. L., J. Bandola, and W. Schumann (May 2013). "Construction of an Artificial secYEG Operon Allowing High Level Secretion of α-Amylase". In: *Protein Expression and Purification* 89.1, pp. 92–96. ISSN: 1096-0279. pmid: 23473827.

Nagamori, S., I. N. Smirnova, and H. R. Kaback (Apr. 2004). "Role of YidC in Folding of Polytopic Membrane Proteins". In: *The Journal of Cell Biology* 165.1, pp. 53–62. ISSN: 0021-9525. pmid: 15067017.

Nasr, M. L., D. Baptista, M. Strauss, *et al.* (Nov. 21, 2016). "Covalently Circularized Nanodiscs for Studying Membrane Proteins and Viral Entry". In: *Nature Methods* 14.1, pp. 49–52. ISSN: 1548-7091, 1548-7105. URL: http://www.nature.com/doifinder/10.1038/nmeth.4079 (visited on 03/15/2017).

Neumann-Haefelin, C., U. Schäfer, M. Müller, *et al.* (Dec. 1, 2000). "SRP-Dependent Co-Translational Targeting and SecA-Dependent Translocation Analyzed as Individual Steps in the Export of a Bacterial Protein". In: *The EMBO Journal* 19.23, pp. 6419–6426. ISSN: 0261-4189. pmid: 11101515. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC305875/ (visited on 01/17/2018).

Noriega, T. R., J. Chen, P. Walter, *et al.* (Oct. 30, 2014). "Real-Time Observation of Signal Recognition Particle Binding to Actively Translating Ribosomes". In: *eLife* 3, e04418. ISSN: 2050-084X. URL: https://elifesciences.org/articles/04418 (visited on 01/16/2018).

Nouwen, N. and A. Driessen (June 1, 2002). "SecDFyajC Forms a Heterotetrameric Complex with YidC". In: *Molecular Microbiology* 44.5, pp. 1397–1405. ISSN: 1365-2958. URL: http://onlinelibrary.wiley.com.gate1.inist.fr/doi/10.1046/j.1365-2958.2002.02972.x/abstract (visited on 04/11/2014).

Nouwen, N., M. Piwowarek, G. Berrelkamp, *et al.* (Aug. 2005). "The Large First Periplasmic Loop of SecD and SecF Plays an Important Role in SecDF Functioning". In: *Journal of Bacteriology* 187.16, pp. 5857–5860. ISSN: 0021-9193. pmid: 16077136. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1196081/ (visited on 01/20/2018).

Ohta, N., Y. Kato, H. Watanabe, *et al.* (Nov. 3, 2016). "In Vitro Membrane Protein Synthesis inside Sec Translocon-Reconstituted Cell-Sized Liposomes". In: *Scientific Reports* 6, p. 36466. ISSN: 2045-2322. URL: http://www.nature.com/srep/2016/161103/srep36466/full/srep36466.html (visited on 11/14/2016).

Orfanoudaki, G. and A. Economou (Dec. 2014). "Proteome-Wide Subcellular Topologies of E. Coli Polypeptides Database (STEPdb)". In: *Molecular & Cellular Proteomics : MCP* 13.12, pp. 3674–3687. ISSN: 1535-9476. pmid: 25210196. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4256514/ (visited on 01/16/2018).

Osborne, A. R. and T. A. Rapoport (Apr. 6, 2007). "Protein Translocation Is Mediated by Oligomers of the SecY Complex with One SecY Copy Forming the Channel". In: *Cell* 129.1, pp. 97–110. ISSN: 0092-8674. URL: http://www.sciencedirect.com/science/article/pii/S0092867407003054 (visited on 10/09/2014).

Palsdottir, H. and C. Hunte (Nov. 3, 2004). "Lipids in Membrane Protein Structures". In: *Biochimica et Biophysica Acta (BBA) - Biomembranes*. Lipid-Protein Interactions 1666.1, pp. 2–18. ISSN: 0005-2736. URL: http://www.sciencedirect.com/science/article/pii/S0005273604001567 (visited on 01/28/2018).

Panzner, S., L. Dreier, E. Hartmann, *et al.* (May 19, 1995). "Posttranslational Protein Transport in Yeast Reconstituted with a Purified Complex of Sec Proteins and Kar2p". In: *Cell* 81.4, pp. 561–570. ISSN: 0092-8674. pmid: 7758110.

Park, E. and T. A. Rapoport (May 11, 2011). "Preserving the Membrane Barrier for Small Molecules during Bacterial Protein Translocation". In: *Nature* 473.7346, p. 239. ISSN: 1476-4687. URL: https://www.nature.com/articles/nature10014 (visited on 01/16/2018).

– (2012). "Mechanisms of Sec61/SecY-Mediated Protein Translocation Across Membranes". In: *Annual Review of Biophysics* 41.1, pp. 21–40. pmid: 22224601. URL: http://dx.doi.org/10.1146/annurev-biophys-050511-102312 (visited on 01/16/2015).

Perkins, S. J. (May 1, 1986). "Protein Volumes and Hydration Effects". In: *European Journal of Biochemistry* 157.1, pp. 169–180. ISSN: 1432-1033. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1432-1033.1986.tb09653.x/abstract (visited on 01/14/2018).

Pinchuk, I. and D. Lichtenberg (June 2017). "Deuterium Kinetic Isotope Effect (DKIE) in Copper-Induced LDL Peroxidation: Interrelated Effects of on Inhibition and Propagation". In: *Chemistry and Physics of Lipids* 205, pp. 42–47. ISSN: 1873-2941. pmid: 28455203.

Pogliano, J. and J. Beckwith (Feb. 1, 1994). "SecD and SecF Facilitate Protein Export in Escherichia Coli." In: *The EMBO Journal* 13.3, pp. 554–561. ISSN: 0261-4189. pmid: 8313900. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC394844/ (visited on 12/01/2017).

Pogliano, K. and J. Beckwith (Feb. 1994). "Genetic and Molecular Characterization of the Escherichia Coli secD Operon and Its Products." In: *Journal of Bacteriology* 176.3, pp. 804–814. ISSN: 0021-9193. pmid: 7507921. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC205118/ (visited on 01/23/2018).

Popot, J.-L. (2010). "Amphipols, Nanodiscs, and Fluorinated Surfactants: Three Nonconventional Approaches to Studying Membrane Proteins in Aqueous Solutions". In: *Annual Review of Biochemistry* 79.1, pp. 737–775. pmid: 20307193. URL: http://www.annualreviews.org/doi/abs/10.1146/annurev.biochem.052208.114057 (visited on 04/17/2014).

Prouty, W. F. and A. L. Goldberg (May 25, 1972). "Effects of Protease Inhibitors on Protein Breakdown in Escherichia Coli". In: *Journal of Biological Chemistry* 247.10, pp. 3341–3352. ISSN: 0021-9258, 1083-351X. pmid: 4554917. URL: http://www.jbc.org/content/247/10/3341 (visited on 03/12/2018).

Raetz, C. R. (Sept. 1978). "Enzymology, Genetics, and Regulation of Membrane Phospholipid Synthesis in Escherichia Coli." In: *Microbiological Reviews* 42.3, pp. 614–659. ISSN: 0146-0749. pmid: 362151. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC281446/ (visited on 04/29/2018).

Rapoport, T. A. (Nov. 29, 2007). "Protein Translocation across the Eukaryotic Endoplasmic Reticulum and Bacterial Plasma Membranes". In: *Nature* 450.7170, pp. 663–669. ISSN: 0028-0836. URL: http://www.nature.com.gate1.inist.fr/nature/journal/v450/n7170/full/nature06384.html (visited on 03/28/2014).

Rapoport, T. A., L. Li, and E. Park (2017). "Structural and Mechanistic Insights into Protein Translocation". In: *Annual Review of Cell and Developmental Biology* 33.1, null. pmid: 28564553. URL: https://doi.org/10.1146/annurev-cellbio-100616-060439 (visited on 06/14/2017).

Rawlings, A. E. (June 15, 2016). "Membrane Proteins: Always an Insoluble Problem?" In: *Biochemical Society Transactions* 44.3, pp. 790–795. ISSN: 0300-5127, 1470-8752. pmid: 27284043. URL: http://www.biochemsoctrans.org/content/44/3/790 (visited on 11/16/2017).

Rubinson, K. A., C. Stanley, and S. Krueger (Apr. 1, 2008). "Small-Angle Neutron Scattering and the Errors in Protein Structures That Arise from Uncorrected Background and Intermolecular Interactions". In: *Journal of Applied Crystallography* 41.2, pp. 456–465. ISSN: 0021-8898. URL: http://scripts.iucr.org/cgi-bin/paper?S0021889808004950 (visited on 01/14/2018).

Sääf, A., M. Monné, J. W. de Gier, *et al.* (Nov. 13, 1998). "Membrane Topology of the 60-kDa Oxa1p Homologue from Escherichia Coli". In: *The Journal of Biological Chemistry* 273.46, pp. 30415–30418. ISSN: 0021-9258. pmid: 9804807.

Sachelaru, I., N.-A. Petriman, R. Kudva, and H.-G. Koch (Jan. 8, 2014). "Dynamic Interaction of the Sec Translocon with the Chaperone PpiD". In: *Journal of Biological Chemistry* 289.31, pp. 21706–21715. ISSN: 0021-9258, 1083-351X. pmid: 24951590. URL: http://www.jbc.org/content/289/31/21706 (visited on 12/12/2016).

Sachelaru, I., N.-A. Petriman, R. Kudva, P. Kuhn, *et al.* (July 6, 2013). "YidC Occupies the Lateral Gate of the SecYEG Translocon and Is Sequentially Displaced by a Nascent Membrane Protein". In: *Journal of Biological Chemistry* 288.23, pp. 16295–16307. ISSN: 0021-9258, 1083-351X. pmid: 23609445. URL: http://www.jbc.org.gate1.inist.fr/content/288/23/16295 (visited on 04/11/2014).

Saio, T., X. Guan, P. Rossi, *et al.* (May 9, 2014). "Structural Basis for Protein Antiaggregation Activity of the Trigger Factor Chaperone". In: *Science (New York, N.Y.)* 344.6184, p. 1250494. ISSN: 1095-9203. pmid: 24812405.

Sala, A., P. Bordes, and P. Genevaux (2014). "Multitasking SecB Chaperones in Bacteria". In: *Frontiers in Microbiology* 5. ISSN: 1664-302X. URL: http://journal.frontiersin.org/article/10.3389/fmicb.2014.00666/full (visited on 01/26/2017).

Salvay, A. G., M. Santamaria, M. le Maire, *et al.* (Dec. 2007). "Analytical Ultracentrifugation Sedimentation Velocity for the Characterization of Detergent-Solubilized Membrane Proteins Ca++-ATPase and ExbB". In: *Journal of Biological Physics* 33.5-6, pp. 399–419. ISSN: 0092-0606. pmid: 19669527. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2565766/ (visited on 12/16/2017).

Samuelson, J. C., M. Chen, F. Jiang, *et al.* (Aug. 10, 2000). "YidC Mediates Membrane Protein Insertion in Bacteria". In: *Nature* 406.6796, pp. 637–641. ISSN: 0028-0836. URL: http://www.nature.com.gate1.inist.fr/nature/journal/v406/n6796/full/406637a0.html (visited on 04/11/2014).

Saraogi, I., D. Akopian, and S.-o. Shan (June 9, 2014). "Regulation of Cargo Recognition, Commitment, and Unloading Drives Cotranslational Protein Targeting". In: *The Journal of Cell Biology* 205.5, pp. 693–706. ISSN: 0021-9525. pmid: 24914238. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4050729/ (visited on 01/17/2018).

Scheiner, S. and M. Čuma (Jan. 1, 1996). "Relative Stability of Hydrogen and Deuterium Bonds". In: *Journal of the American Chemical Society* 118.6, pp. 1511–1521. ISSN: 0002-7863. URL: http://dx.doi.org/10.1021/ja9530376 (visited on 01/28/2018).

Schibich, D., F. Gloge, I. Pöhner, *et al.* (Aug. 2016). "Global Profiling of SRP Interaction with Nascent Polypeptides". In: *Nature* 536.7615, pp. 219–223. ISSN: 0028-0836, 1476-4687. URL: http://www.nature.com/articles/nature19070 (visited on 04/13/2018).

Schiebel, E., A. J. M. Driessen, F.-U. Hartl, *et al.* (Mar. 8, 1991). "ΔμH+ and ATP Function at Different Steps of the Catalytic Cycle of Preprotein Translocase". In: *Cell* 64.5, pp. 927–939. ISSN: 0092-8674. URL: http://www.sciencedirect.com/science/article/pii/009286749190317R (visited on 01/18/2018).

Schuler, M. A., I. G. Denisov, and S. G. Sligar (2013). "Nanodiscs as a New Tool to Examine Lipid-Protein Interactions". In: *Methods in molecular biology (Clifton, N.J.)* 974, pp. 415–433. ISSN: 1064-3745. pmid: 23404286. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4201044/ (visited on 03/05/2018).

Schulze, R. J., J. Komar, M. Botte, *et al.* (Jan. 4, 2014). "Membrane Protein Insertion and Proton-Motive-Force-Dependent Secretion through the Bacterial Holo-Translocon SecYEG–SecDF–YajC–YidC". In: *Proceedings of the National Academy of Sciences* 111.13, pp. 4844–4849. ISSN: 0027-8424, 1091-6490. pmid: 24550475. URL: http://www.pnas.org/content/111/13/4844 (visited on 04/11/2014).

Scotti, P. A., M. L. Urbanus, J. Brunner, *et al.* (Feb. 15, 2000). "YidC, the Escherichia Coli Homologue of Mitochondrial Oxa1p, Is a Component of the Sec Translocase". In: *The EMBO Journal* 19.4, pp. 542–549. URL: http://emboj.embopress.org/content/19/4/542 (visited on 04/08/2014).

Seitl, I., S. Wickles, R. Beckmann, *et al.* (Jan. 2014). "The C-Terminal Regions of YidC from Rhodopirellula Baltica and Oceanicaulis Alexandrii Bind to Ribosomes and Partially Substitute for SRP Receptor Function in Escherichia Coli". In: *Molecular Microbiology* 91.2, pp. 408–421. ISSN: 1365-2958. pmid: 24261830.

Serek, J., G. Bauer-Manz, G. Struhalla, *et al.* (Jan. 28, 2004). "Escherichia Coli YidC Is a Membrane Insertase for Sec-Independent Proteins". In: *The EMBO journal* 23.2, pp. 294–301. ISSN: 0261-4189. pmid: 14739936.

Shan, S.-o. and P. Walter (Feb. 7, 2005). "Co-Translational Protein Targeting by the Signal Recognition Particle". In: *FEBS Letters*. Molecular Mechanisms of Biological Systems. 130th Nobel Symposium 579.4, pp. 921–926. ISSN: 0014-5793. URL: http://www.sciencedirect.com/science/article/pii/S0014579304014322 (visited on 03/28/2014).

Shotton, M. W., L. H. Pope, T. Forsyth, *et al.* (Nov. 1, 1997). "A High-Angle Neutron Fibre Diffraction Study of the Hydration of Deuterated A-DNA". In: *Biophysical Chemistry* 69.1, pp. 85–96. ISSN: 0301-4622. URL: http://www.sciencedirect.com/science/article/pii/S0301462297000902 (visited on 01/28/2018).

Signor, L. and E. Boeri Erba (Sept. 9, 2013). "Matrix-Assisted Laser Desorption/Ionization Time of Flight (MALDI-TOF) Mass Spectrometric Analysis of Intact Proteins Larger than 100 kDa". In: *Journal of Visualized Experiments* 79. ISSN: 1940-087X. URL: http://www.jove.com/video/50635/matrix-assisted-laser-desorptionionization-time-flight-maldi-tof-mass (visited on 02/12/2016).

Singh, R., C. Kraft, R. Jaiswal, *et al.* (Mar. 7, 2014). "Cryo-Electron Microscopic Structure of SecA Protein Bound to the 70S Ribosome". In: *The Journal of Biological Chemistry* 289.10, pp. 7190–7199. ISSN: 1083-351X. pmid: 24443566.

Slotboom, D. J., R. H. Duurkens, K. Olieman, *et al.* (Oct. 2008). "Static Light Scattering to Characterize Membrane Proteins in Detergent Solution". In: *Methods* 46.2, pp. 73–82. ISSN: 10462023. URL: http://linkinghub.elsevier.com/retrieve/pii/S1046202308000996 (visited on 08/16/2016).

Strop, P. and A. T. Brunger (Aug. 2005). "Refractive Index-Based Determination of Detergent Concentration and Its Application to the Study of Membrane Proteins". In: *Protein Science : A Publication of the Protein Society* 14.8, pp. 2207–2211. ISSN: 0961-8368. pmid: 16046633. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2279333/ (visited on 12/06/2017).

Suo, Y., S. J. S. Hardy, and L. L. Randall (Feb. 27, 2015). "The Basis of Asymmetry in the SecA-SecB Complex". In: *Journal of molecular biology* 427.4, pp. 887–900. ISSN: 0022-2836. pmid: 25534082. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4370339/ (visited on 01/17/2018).

Svergun, D. I., S. Richard, M. H. J. Koch, *et al.* (Mar. 3, 1998). "Protein Hydration in Solution: Experimental Observation by x-Ray and Neutron Scattering". In: *Proceedings of the National Academy of Sciences* 95.5, pp. 2267–2272. ISSN: 0027-8424, 1091-6490. pmid: 9482874. URL: http://www.pnas.org/content/95/5/2267 (visited on 04/10/2018).

Svergun, D. I., M. H. J. Koch, P. A. Timmins, *et al.* (Aug. 8, 2013). *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*. OUP Oxford. 369 pp. ISBN: 978-0-19-150703-8.

Swartz, J. (July 2006). "Developing Cell-Free Biology for Industrial Applications". In: *Journal of Industrial Microbiology & Biotechnology* 33.7, pp. 476–485. ISSN: 1367-5435. pmid: 16761165.

Tanaka, Y., Y. Sugano, M. Takemoto, *et al.* (Nov. 24, 2015). "Crystal Structures of SecYEG in Lipidic Cubic Phase Elucidate a Precise Resting and a Peptide-Bound State". In: *Cell Reports* 13.8, pp. 1561–1568. ISSN: 2211-1247. pmid: 26586438. URL: http://www.cell.com/cell-reports/abstract/S2211-1247(15)01179-1 (visited on 01/16/2018).

Tang, Y., X. Pan, Y. Chen, *et al.* (Jan. 27, 2011). "Dimeric SecA Couples the Preprotein Translocation in an Asymmetric Manner". In: *PLoS ONE* 6.1. ISSN: 1932-6203. pmid: 21304597. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3029384/ (visited on 03/08/2018).

Tomonaga, S., A. Nomura, and K. Awaya (1971). "Optimal Feulgen Hydrolysis for Microspectrophotometry". In: *Acta Histochemica et Cytochemica* 4.3, pp. 166–172.

Tribet, C., R. Audebert, and J.-L. Popot (Dec. 24, 1996). "Amphipols: Polymers That Keep Membrane Proteins Soluble in Aqueous Solutions". In: *Proceedings of the National Academy of Sciences of the United States of America* 93.26, p. 15047. pmid: 8986761. URL: /pmc/articles/PMC26353/?report=abstract (visited on 05/26/2014).

Trimmer, E. E., U. S. Wanninayake, and P. F. Fitzpatrick (Nov. 4, 2017). "Mechanistic Studies of an Amine Oxidase Derived from D-Amino Acid Oxidase". In: *Biochemistry* 56.14, pp. 2024–2030. ISSN: 1520-4995. pmid: 28355481.

Tsirigotaki, A., J. D. Geyter, N. Šoštaric´, *et al.* (Jan. 2017). "Protein Export through the Bacterial Sec Pathway". In: *Nature Reviews Microbiology* 15.1, p. 21. ISSN: 1740-1534. URL: https://www.nature.com/articles/nrmicro.2016.161 (visited on 01/13/2018).

Tsukazaki, T., H. Mori, Y. Echizen, *et al.* (May 11, 2011). "Structure and Function of a Protein Export-Enhancing Membrane Component SecDF". In: *Nature* 474.7350, pp. 235–238. ISSN: 0028-0836. pmid: 21562494. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3697915/ (visited on 01/17/2018).

Urbanus, M., L. Fröderberg, D. Drew, *et al.* (Apr. 12, 2002). "Targeting, Insertion, and Localization of Escherichia Coli YidC". In: *The Journal of Biological Chemistry* 277.15, pp. 12718–12723. ISSN: 0021-9258. pmid: 11821429.

Urbanus, M., P. A. Scotti, L. Fröderberg, *et al.* (June 6, 2001). "Sec-Dependent Membrane Protein Insertion: Sequential Interaction of Nascent FtsQ with SecY and YidC". In: *EMBO Reports* 2.6, pp. 524–529. ISSN: 1469-221X. pmid: 11415986. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1083902/ (visited on 01/16/2015).

Van Assche, E., S. Van Puyvelde, J. Vanderleyden, *et al.* (Mar. 3, 2015). "RNA-Binding Proteins Involved in Post-Transcriptional Regulation in Bacteria". In: *Frontiers in Microbiology* 6. ISSN: 1664-302X. pmid: 25784899. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4347634/ (visited on 02/21/2018).

Van der Berg, B., W. M. Clemons, I. Collinson, *et al.* (Mar. 12, 2003). "X-Ray Structure of a Protein-Conducting Channel". In: *Nature* 427.6969, pp. 36–44. ISSN: 0028-0836. URL: http://www.nature.com.gate1.inist.fr/nature/journal/v427/n6969/full/nature02218.html (visited on 03/28/2014).

Van der Laan, M., P. Bechtluft, S. Kol, *et al.* (Apr. 26, 2004). "F1F0 ATP Synthase Subunit c Is a Substrate of the Novel YidC Pathway for Membrane Protein Biogenesis". In: *The Journal of Cell Biology* 165.2, pp. 213–222. ISSN: 0021-9525. pmid: 15096523.

Visick, J. E. and S. Clarke (June 1, 1995). "Repair, Refold, Recycle: How Bacteria Can Deal with Spontaneous and Environmental Damage to Proteins". In: *Molecular Microbiology* 16.5, pp. 835–845. ISSN: 1365-2958. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2958.1995.tb02311.x/abstract (visited on 03/12/2018).

Voet, D., J. G. Voet, and C. W. Pratt (2008). *Fundamentals of Biochemistry: Life at the Molecular Level*. 3rd ed. Hoboken, NJ: Wiley. 1099 pp. ISBN: 978-0-470-12930-2.

Von Loeffelholz, O., M. Botte, and C. Schaffitzel (2011). "Escherichia Coli Cotranslational Targeting and Translocation". In: *eLS*. John Wiley & Sons, Ltd. ISBN: 978-0-470-01590-2. URL: http://onlinelibrary.wiley.com.gate1.inist.fr/doi/10.1002/9780470015902.a0023170/abstract (visited on 10/03/2014).

Voorhees, R. M. and R. S. Hegde (Jan. 1, 2016). "Structure of the Sec61 Channel Opened by a Signal Sequence". In: *Science (New York, N.Y.)* 351.6268, pp. 88–91. ISSN: 1095-9203. pmid: 26721998.

Voorhees, R. and R. Hegde (2018). "Structures of the Scanning and Engaged States of the Mammalian SRP-Ribosome Complex". In: *eLife* 4 (). ISSN: 2050-084X. pmid: 26158507. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4497383/ (visited on 01/23/2018).

Wagner, S., O. Ioan Pop, G.-J. Haan, *et al.* (Dec. 25, 2009). "Biogenesis of MaIF and the MaIFGK(2) Maltose Transport Complex in Escherichia Coli Requires YidC. (Vol 283, Pg 17881, 2008)". In: *Journal of Biological Chemistry* 284, pp. 36720–36720.

Weiss, J. B., P. H. Ray, and P. J. Bassford (1988). "Purified SecB Protein of Escherichia Coli Retards Folding and Promotes Membrane Translocation of the Maltose-Binding Protein in Vitro". In: *Proceedings of the National Academy of Sciences of the United States of America* 85.23, pp. 8978–8982. ISSN: 0027-8424. JSTOR: 32869.

Westheimer, F. H. (June 1, 1961). "The Magnitude of the Primary Kinetic Isotope Effect for Compounds of Hydrogen and Deuterium." In: *Chemical Reviews* 61.3, pp. 265–273. ISSN: 0009-2665. URL: https://pubs.acs.org/doi/abs/10.1021/cr60211a004 (visited on 03/12/2018).

Whitten, A. E., S. Cai, and J. Trewhella (Feb. 1, 2008). "MULCh : Modules for the Analysis of Small-Angle Neutron Contrast Variation Data from Biomolecular Assemblies". In: *Journal of Applied Crystallography* 41.1, pp. 222–226. ISSN: 0021-8898. URL: http://scripts.iucr.org/cgi-bin/paper?S0021889807055136 (visited on 05/03/2018).

Xie, K., D. Kiefer, G. Nagler, *et al.* (Nov. 7, 2006). "Different Regions of the Nonconserved Large Periplasmic Domain of Escherichia Coli YidC Are Involved in the SecF Interaction and Membrane Insertase Activity". In: *Biochemistry* 45.44, pp. 13401–13408. ISSN: 0006-2960. pmid: 17073462.

Xu, Z., J. D. Knafels, and K. Yoshino (2000). "Crystal Structure of the Bacterial Protein Export Chaperone SecB". In: *Nature Structural & Molecular Biology* 7.12, pp. 1172–1177. URL: http://www.nature.com/nsmb/journal/v7/n12/abs/nsb1200_1172.html (visited on 01/26/2017).

Yi, L., F. Jiang, M. Chen, *et al.* (Sept. 9, 2003). "YidC Is Strictly Required for Membrane Insertion of Subunits a and c of the F(1)F(0)ATP Synthase and SecE of the SecYEG Translocase". In: *Biochemistry* 42.35, pp. 10537–10544. ISSN: 0006-2960. pmid: 12950181.

Zaccai, N., I. Serdyuk, and J. Zaccai (May 18, 2017). *Methods in Molecular Biophysics*. Cambridge University Press. 709 pp. ISBN: 978-1-107-05637-4.

Zhang, X., R. Rashid, K. Wang, *et al.* (May 7, 2010). "Sequential Checkpoints Govern Substrate Selection During Cotranslational Protein Targeting". In: *Science (New York, N.Y.)* 328.5979, pp. 757–760. ISSN: 0036-8075. pmid: 20448185. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3760334/ (visited on 01/17/2018).

Zhang, X., C. Schaffitzel, N. Ban, *et al.* (Oct. 2, 2009). "Multiple Conformational Switches in a GTPase Complex Control Co-Translational Protein Targeting". In: *Proceedings of the National Academy of Sciences* 106.6, pp. 1754–1759. ISSN: 0027-8424, 1091-6490. pmid: 19174514. URL: http://www.pnas.org/content/106/6/1754 (visited on 01/17/2018).

Zhou, J. and Z. Xu (Nov. 2003). "Structural Determinants of SecB Recognition by SecA in Bacterial Protein Translocation". In: *Nature Structural & Molecular Biology* 10.11, pp. 942–947. ISSN: 1545-9985. URL: https://www.nature.com/articles/nsb980 (visited on 03/08/2018).

Zhou, Q., S. Sun, P. Tai, *et al.* (Oct. 4, 2012). "Structural Characterization of the Complex of SecB and Metallothionein-Labeled proOmpA by Cryo-Electron Microscopy". In: *PLoS ONE* 7.10. ISSN: 1932-6203. pmid: 23056562. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3464278/ (visited on 12/12/2016).

Zhou, Y., T. Ueda, and M. Müller (Apr. 9, 2014). "Signal Recognition Particle and SecA Cooperate during Export of Secretory Proteins with Highly Hydrophobic Signal Sequences". In: *PLoS ONE* 9.4. ISSN: 1932-6203. pmid: 24717922. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3981701/ (visited on 01/16/2018).

Zimmer, J., Y. Nam, and T. A. Rapoport (Oct. 16, 2008). "Structure of a Complex of the ATPase SecA and the Protein-Translocation Channel". In: *Nature* 455.7215, pp. 936–943. ISSN: 0028-0836. URL: http://www.nature.com.gate1.inist.fr/nature/journal/v455/n7215/full/nature07335.html (visited on 01/27/2015).

Zoonens, M. and J.-L. Popot (Oct. 2014). "Amphipols for Each Season". In: *The Journal of membrane biology* 247.0, pp. 759–796. ISSN: 0022-2631. pmid: 24969706. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4282167/ (visited on 04/10/2018).