Cryo-microscopie électronique des complexes de l'adressage et de la translocation co-traductionnelle chez E. coli

Qiyang Jiang

UNIVERSITÉ DE GRENOBLE

EMBL

THÈSE / THESIS

Pour obtenir le grade de / To obtain the title of

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité / Discipline: **Chimie Physique Moléculaire et Structurale**

Arrêté ministériel : 7 août 2006

Présentée par / Presented by

Qiyang JIANG (姜启阳)

Thèse dirigée par / Thesis supervisor: **Prof. Dr. Christiane SCHAFFITZEL**

préparée au sein du / prepared at
European Molecular Biology Laboratory (EMBL), Grenoble Outstation
dans / in **l'École Doctorale de Chimie et Sciences du Vivant**

Cryo-microscopie électronique des complexes de l'adressage et de la translocation co-traductionnelle chez *E. coli*

Electron cryo-microscopy of complexes in *E. coli* co-translational targeting and translocation

Thèse soutenue publiquement le / Public defense on **18.06.2015**,
devant le jury composé de / jury members:

Prof. Dr. Valentin GORDELIY
Directeur de recherche, CEA, IBS, Grenoble. Président

Dr. Ariel BLOCKER
Reader, University of Bristol. Membre

Dr. Bruno MIROUX
Directeur de recherche, INSERM, IBPC, Paris. Membre

Dr. Guy SCHOEHN
Directeur de recherche, CNRS, IBS, UVHCI, Grenoble. Membre

Université Joseph Fourier / Université Pierre Mendès France /
Université Stendhal / Université de Savoie / Grenoble INP

RÉSUMÉ

La membrane cellulaire est la barrière qui sépare l'intérieur des cellules de l'environnement extérieur. Elle se compose de lipides et de protéines. Les gènes codant pour les protéines membranaires représentent environ 30% des génomes. Les protéines membranaires sont synthétisées dans le cytosol par les ribosomes, mais suivent des voies spécifiques pour s'intégrer dans la membrane cellulaire. Les ribosomes en cours de traduction de protéines membranaires sont reconnus dans le cytosol et adressés à la membrane. Par la suite, les chaînes naissantes de protéines membranaires sont insérées dans la bicouche lipidique puis repliées de façons appropriées, ce mécanisme s'appelle la translocation. Le processus d'adressage est médiée par la particule de reconnaissance du signal (SRP) et son récepteur, tandis que la translocation est effectuée par un certain nombre de complexes de protéines membranaires.

Cette thèse décrit deux des complexes impliqués dans cet adressage et translocation co-traductionnelle chez *Escherichia coli* : Le complexe ribosome-SRP-FtsY pour l'adressage en conformation «fermé» et le complexe dans lequel le ribosome est lié à l'holo-translocon (HTL) qui se compose de sept protéines membranaires. J'ai utilisé principalement la cryo-microscopie électronique pour caractériser ces complexes. La cryo-EM permet de déterminer la structure des échantillons biologiques à une résolution supérieure au nanomètre dans leur environnement natif, sans avoir à les cristalliser. Dans ce travail, j'ai bénéficié des améliorations récentes dans l'équipement et le traitement d'image.

A partir d'un ensemble de données de cryo-EM obtenu par les membres du groupe, j'ai déterminé la structure du complexe ribosome-SRP-FtsY en conformation «fermé» avec une résolution de 5.7 Å. Différentes stratégies de tri des calcul sont été appliquées pour identifier la partie la plus homogène de l'ensemble des données. La structure montre un domaine bien résolu SRP ARN et SRP M avec une séquence signal liée. L'interaction entre les SRP et le ribosome pourrait être modélisée avec une grande fidélité. Cette structure révèle également que les GTPases SRP-ftsY

sont détachées de la tétra-boucle de l'ARN et sont flexibles, libérant alors le site de sortie du ribosome permettant la liaison du translocons.

Dans le second projet, différentes approches ont été poursuivis pour résoudre la structure du complexe ribosome-HTL à haute résolution. Une structure initiale à 22 Å a été obtenue en mélangeant HTL solubilisés en détergent avec des ribosomes, démontrant la possibilité de préserver le complexe dans les conditions utilisées pour la préparation des grilles. J'ai ensuite exploré l'utilisation de nanodiscs et un nouveau détergent appelé LMNG pour stabiliser HTL dans des tampons sans détergent. Un deuxième ensemble de données a ensuite été recueilli à partir d'échantillon obtenu par gradient de fixation, la structure a été résolue à 17 Å. La préparation des échantillons a été optimisée utilisant entre autres les amphipoles. On a montré que deux types d'amphipole-HTL peuvent se lier au ribosome, et des structures de plus grande résolution devrait être obtenu à partir de ces échantillons.

MOTS CLÉS

cryo-microscopie électronique; protéine membranaire; ribosome; adressage des protéines; translocation des protéines

ABSTRACT

The cell membrane is the barrier that separates the interior of cells from the outside environment. It consists of lipids and proteins. Genes encoding membrane proteins make up about 30% of genomes. Membrane proteins are synthesized in the cytosol by ribosomes, but employ special pathways to integrate into the cell membrane. Ribosomes translating membrane proteins are recognized by special factors in the cytosol and targeted to the membrane. Subsequently, nascent chains of the membrane proteins are inserted into the lipid bilayer and are folded into their proper structures, a process termed translocation. The targeting process is mediated by the signal recognition particle (SRP) and its receptor, while the translocation is performed by a number of membrane protein complexes.

This thesis describes two of the complexes involved in co-translational targeting and translocation in *Escherichia coli*: The ribosome-SRP-FtsY targeting complex in the “closed” conformation and the complex of a ribosome with the holo-translocon (HTL) consisting of seven membrane proteins. I mainly used electron cryo-microscopy to characterize these complexes. Cryo-EM allows structural determination of biological samples at sub-nanometer resolution in their native environment, without the need to crystallize the specimen. In this work, I took advantage of the recent advances in both the hardware and the image processing.

Starting from a cryo-EM dataset obtained by group members, I have determined the structure of ribosome-SRP-FtsY complex in the “closed” conformation at 5.7 Å resolution. Different computational sorting strategies were applied to identify the most homogeneous sub-pool of the dataset. The structure shows a well-resolved SRP RNA and SRP M domain with a signal sequence bound. The interaction between SRP and ribosome could be modeled with high confidence. This structure also reveals that the SRP-FtsY GTPases are detached from the RNA tetraloop and are flexible, thus liberating the ribosomal exit site for binding of the translocation machinery.

In the second project, different approaches were pursued to solve the structure of the ribosome-HTL complex at high resolution. An initial structure at 22 Å was

obtained by mixing detergent-solubilized HTL with the ribosome, demonstrating that it is possible to preserve the complex under the conditions used for specimen preparation. I have then explored the use of nanodiscs and a new detergent called LMNG to stabilize HTL in detergent-free buffers. A second dataset was subsequently collected from a sample prepared by gradient-fixation, and the structure was solved at 17 Å. Sample preparation has been optimized further using amphipols. Two types of amphipol-HTL complexes were shown to bind to the ribosome, and higher resolution structures are expected to be obtained from these samples.

KEY WORDS

electron cryo-microscopy; membrane protein; ribosome; protein targeting; protein translocation

ACKNOWLEDGEMENTS

I first want to thank my supervisor, Prof. Dr. Christiane Schaffitzel, for the opportunity she gave me to work in her laboratory and for the enormous support and help I got from her during the thesis work.

I thank the members of my thesis advisory committee, Prof. Dr. Imre Berger, Dr. Edward Lemke, Dr. José Marquez, Prof. Dr. Rob Ruigrok and Prof. Dr. Guy Schoehn for their supports and advices.

I thank my jury members Dr. Ariel Blocker, Prof. Dr. Valentin Gordeliy, Prof. Dr. Bruno Miroux and Prof. Dr. Guy Schoehn for their time to evaluate my thesis.

I thank every current and former member of the Schaffitzel Team and the Berger Group in EMBL Grenoble for their supports and discussions. I give special thanks to Dr. Manikandan Karuppasamy for the help in electron microscopy, to Dr. Ottilie von Loeffelholz for the help in EM data processing and to Dr. Sarah Zorman for correcting the French abstracts of this thesis.

I would like to thank all the external people who contributed to the thesis work: Dr. Sophie Feuerstein for the help in nanodiscs; Dr. Wim Hagen and Prof. Dr. Guy Schoehn for the help in EM data collections; Prof. Dr. Shu-ou Shan for the collaboration in the ribosome-SRP project and Dr. Manuela Zoonens for the help in amphipols.

I could not have done the thesis work without Google, Wikipedia and unix.stackexchange.com. I thank all those anonymous people who contributed their knowledge to the public in this Information Age.

I give special thanks to my former supervisors Prof. Dr. Chunguang Wang and Prof. Dr. Marcel Knossow who brought me to the field of scientific research and who gave me continuous supports.

My motivation to do science is driven by curiosity, and my motivation to do science for living is because of you: my dear and beloved Ting and Lai!

TABLE OF CONTENTS / TABLE DES MATIERES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Å	**Å**ngströms, 10^{-10} m
CCD	**C**harge-**c**oupled **d**evice
CMC	**C**ritical **m**icelle **c**oncentration
Cryo-EM	**E**lectron **cryo-m**icroscopy
CTF	**C**ontrast **t**ransfer **f**unction
Da	**Da**lton, unified atomic mass unit
DDM	n-**D**o**d**ecyl-ß-D-**m**altoside
E. coli	***E**scherichia **coli***
EGTA	**E**thylene **g**lycol **t**etraacetic **a**cid
ER	**E**ndoplasmic **r**eticulum
GTP	**G**uanosine-5'-**t**ri**p**hosphate
HTL	**H**olo-**t**rans**l**ocon
LMNG	**L**auryl **m**altose **n**eopentyl **g**lycol
M	**M**olar concentration
MSP	**M**embrane **s**caffold **p**rotein
NAPol	**N**on-ionic **a**mphi**pol**
OD_{XXX}	**O**ptical **d**ensity, or absorbance at **XXX** nm
RCF	**R**elative **c**entrifugal **f**orce
RNA	**R**ibo**n**ucleic **a**cid
RNC	**R**ibosome-**n**ascent chain **c**omplex
RPM	**R**evolutions **p**er **m**inute
SAPol	**S**ulfonate **a**mphi**pol**
SDS-PAGE	**P**oly**a**crylamide **g**el **e**lectrophoresis with **s**odium **d**odecyl **s**ulfate
SRP	**S**ignal **r**ecognition **p**article

CHAPTER 1: INTRODUCTION

CHAPITRE 1: INTRODUCTION

Résumé

Trois sujets sont abordés ici: L'adressage et la translocation co-traductionnelle des procaryotes; Les approches et défis de l'étude des protéines membranaires; et les progrès récents en cryo-microscopie électronique.

La particule de reconnaissance du signal (SRP) et son récepteur FtsY dans E. coli, adressent les protéines membranaires en cours de traduction par les ribosomes à la membrane cellulaire, où les chaînes naissantes émergentes seront remises au translocon SecYEG. Ce dernier peut encore être associé à SecDF-YajC-YidC pour former l'holo-translocon (HTL). Dans cette thèse nous étudions le complexe ribosome-SRP-FtsY en conformation «fermé» et le complexe ribosome-HTL.

Pour la plupart des études structurales et fonctionnelles, les protéines membranaires doivent être extraites de bicouches lipidiques par l'utilisation de détergents, ce qui est difficile. D'autres approches ont été développées pour surmonter les problèmes de l'utilisation de détergents tels que les nanodiscs et les amphipols.

La cryo-microscopie électronique (cryo-EM) devient de plus en plus populaire dans la biologie structurale. Les progrès récents, notamment le développement de détecteurs électroniques directs pour l'acquisition des données, la possibilité de corriger le mouvement des particules induite par faisceau et de nouveaux logiciels avec des algorithmes améliorés permettent de déterminer la structure de protéines dans les résolutions quasi-atomiques. En parallèle des règles sont établies pour la validation des structures *via* la cryo-EM.

1.1 Prokaryotic co-translational targeting and translocation

1.1.1 Protein translation and the ribosome

In biological systems, translation is a process where individual amino acids are joined together into proteins using the genetic information as a blueprint. The genetic information is presented in form of a messenger RNA (mRNA), which binds to the ribosome. Correct amino acids entering the ribosome in form of amino acyl-transfer RNA will be fused into the growing polypeptide chain. How a polypeptide finally folds into an active protein and realizes its functions in the cell is still largely enigmatic.

Ribosomes are large and complex molecular machines. They were originally identified and purified by ultracentrifugation. This is why they are also named according to their sedimentation speed in Svedberg units. Prokaryotes have 70S ribosomes, each consisting of a small (30S) and a large (50S) subunit. 70S ribosomes are around 20 nm in diameter and are composed of 65% RNA and 35% proteins. The *Escherichia coli* ribosome comprises 56 ribosomal proteins.

Structures of ribosomes at near-atomic resolution have been first reported in 2000 (Wimberly *et al.*, 2000; Ban *et al.*, 2000; Schluenzen *et al.*, 2000). These ground-breaking structures allowed understanding the mechanism of protein synthesis in unprecedented detail. The correct amino acids are selected based on codon-anti codon pairing between mRNA and amino acyl-tRNA in the decoding center of the small ribosomal subunit. The peptidyl-transfer reaction, *i.e.* the addition of amino acids to the growing polypeptide chain is catalyzed in the active center of the large ribosomal subunit by the ribosomal RNA (rRNA). The newly synthesized polypeptide, also called the nascent chain passes through the ribosomal exit tunnel in the middle of the 50S subunit.

Although some low-order folding such as formation of secondary structure can occur inside the ribosomal exit tunnel, the majority of the protein folding occurs outside the ribosomal tunnel, usually while the translation is still on-going. In the case of membrane proteins, folding occurs virtually exclusively during the synthesis of the polypeptide.

1.1.2 Signal peptide and protein targeting

A signal peptide (also called signal sequence or leader peptide) is a short (8-30 amino acids) peptide at the N-terminus of the nascent chain. The function of the signal peptide is to target the newly synthesized peptide into particular destinations like the cell membrane or the extracellular space.

Signal peptides contain many hydrophobic amino acids that usually form a single alpha-helix. The first few amino acids are very often positively charged, which determines the topology of the membrane proteins by the positive-inside rule, *i.e.* each protein is orientated such that the side containing the higher number of positively charged residues faces the cytoplasm (von Heijne and Gavel, 1988). The C-terminus of the signal peptide contains typically a number of amino acids that can

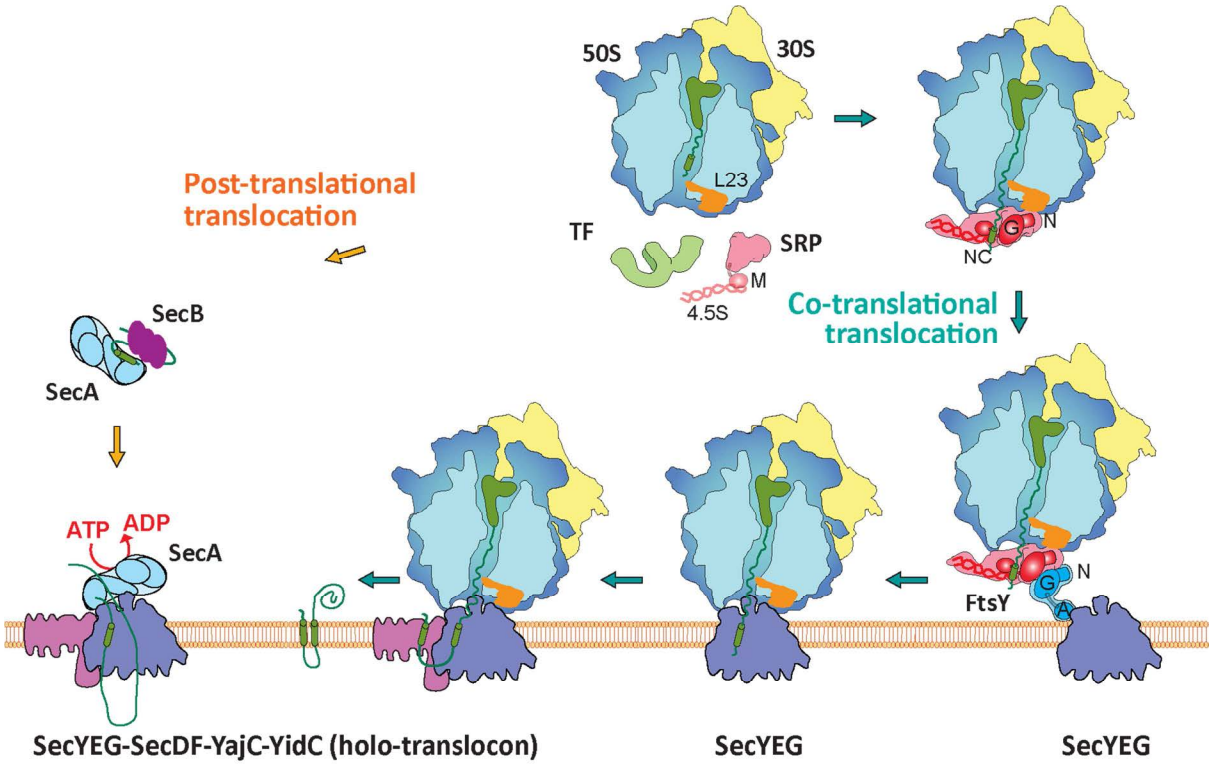


Figure 1-1: Protein targeting and translocation. In bacteria, secretory proteins are targeted post-translationally with the help of the SecB chaperone and the SecA ATPase to the translocation machinery in the membrane (left). SRP and its receptor (FtsY) target nascent membrane proteins co-translationally to the SecYEG protein-conducting channel (right). Additional proteins (SecD, SecF, YajC and YidC) can associate with SecYEG and assist in the translocation across or into the membrane. Adapted from von Loeffelholz et al., 2011.

be recognized and cleaved by the signal peptidase. In the case of membrane proteins the cleavage site is absent and therefore the signal peptide serves as the first transmembrane helix.

Signal peptides vary a lot among different proteins and species. The information carried by the signal peptide needs to be decoded by binding to corresponding factors responsible for targeting. In prokaryotes, the hydrophobic signal peptide of a nascent protein, usually belonging to a membrane protein, is recognized by the signal recognition particle (SRP), while the signal peptide of a secretory protein in the majority of cases is identified and targeted by the SecA protein (Figure 1-1). Secretory proteins are well solubilized and can be kept in an unfolded state by the chaperone SecB and exported after translation is terminated. This is not the case for membrane proteins, given their generally high hydrophobicity.

1.1.3 Co-translational translocation

During translation of a membrane protein, the nascent polypeptide emerging from the tunnel exit of the ribosome is screened by the SRP for the presence of a hydrophobic signal peptide. Once a signal peptide is recognized, the SRP binds the ribosome with high affinity (1 nM or less). In eukaryotes, this binding leads to the slow-down of protein synthesis. Such phenomena is absent in prokaryotes as the prokaryotic SRP lacks the corresponding Alu domain which interferes with translation.

SRP targets the ribosome-nascent chain complex (RNC) to the protein-conducting channel (the translocon) in the cell membrane. This is achieved by the binding of SRP to its receptor, FtsY in bacteria. FtsY is associated with the membrane and in close proximity to the translocon. After successful targeting by SRP and FtsY, the translating ribosome is docked onto the cytoplasmic face of the translocon and the signal peptide is inserted into the translocation pore. Subsequently, the newly-synthetized polypeptide is either translocated into the periplasm or inserted into the lipid bilayer depending on the nature of the polypeptide. This process is called co-translational translocation. It ensures the proper folding of membrane proteins, as the hydrophobic transmembrane helices are not exposed in the cytosol, but are directly inserted into the lipid bilayer.

In bacteria, the protein-conducting channel is a hetero-trimeric complex consisting of SecY, SecE and SecG, also called SecYEG (Rapoport, 2007). SecYEG is employed for post-translational translocation of secretory proteins as well as for co-translational protein translocation. The molecular mechanism of SecYEG has been revealed by the crystal structure, showing two groups of transmembrane segments in SecY along with a central plug that can either form a channel through the membrane to the outside, or open laterally into the membrane bilayer like a clam-shell (Van den Berg *et al.*, 2004).

In *Escherichia coli*, additional proteins are involved in the translocation of membrane proteins. Among them, YidC is an important membrane protein insertase that can perform the insertion of small polypeptides on its own (Samuelson *et al.*, 2001; Chen *et al.*, 2002; van der Laan *et al.*, 2004), but also acts together with the Sec translocon. Notably, YidC and SecYEG can associate through a subcomplex consisting of SecD, SecF, and YajC (Duong and Wickner, 1997; Nouwen and Driessen, 2002). Together, these proteins form a complex of seven subunits (Figure 1-2). This SecYEG-SecDF-YajC-YidC complex, named the holo-translocon (HTL), has been successfully isolated first by radioactive pull-downs from *E. coli* membrane (Duong and Wickner, 1997; Scotti *et al.*, 2000) and subsequently as a recombinantly produced protein complex (Bieniossek *et al.*, 2009; Schulze *et al.*, 2014). It has been shown that HTL is more effective in co-translational insertion of membrane proteins compared to the SecYEG core translocon (Schulze *et al.*, 2014). Koch and colleagues have localized YidC to the lateral gate of SecY by cross-linking experiments (Sachelaru *et al.*, 2013). Interactions between SecD and SecYEG are also identified (Schulze *et al.*, 2014).

The crystal structure of SecDF from *Thermus thermophilus* has been reported (Tsukazaki *et al*., 2011). It revealed two large periplasmic domains for SecD and SecF. The periplasmic domain of SecD consists of a head and a base, which can rotate about 120 degrees in relation to each other. The two conformations were confirmed by crosslinking experiments (Tsukazaki *et al*., 2011). Such a movement of the SecD periplasmic domain was suggested to facilitate post-translational translocation by coupling protein translocation to the proton-motive force (Tsukazaki *et al*., 2011), as suggested earlier (Arkowitz and Wickner, 1994).

More recently, the crystal structure of *E. coli* YidC has been solved (Kumazaki *et al.*, 2014). The structure revealed a conserved hydrophilic groove in the membrane formed by the YidC transmembrane helices. This hydrophilic groove with a conserved, positively charged arginine residue in the middle could be the primary

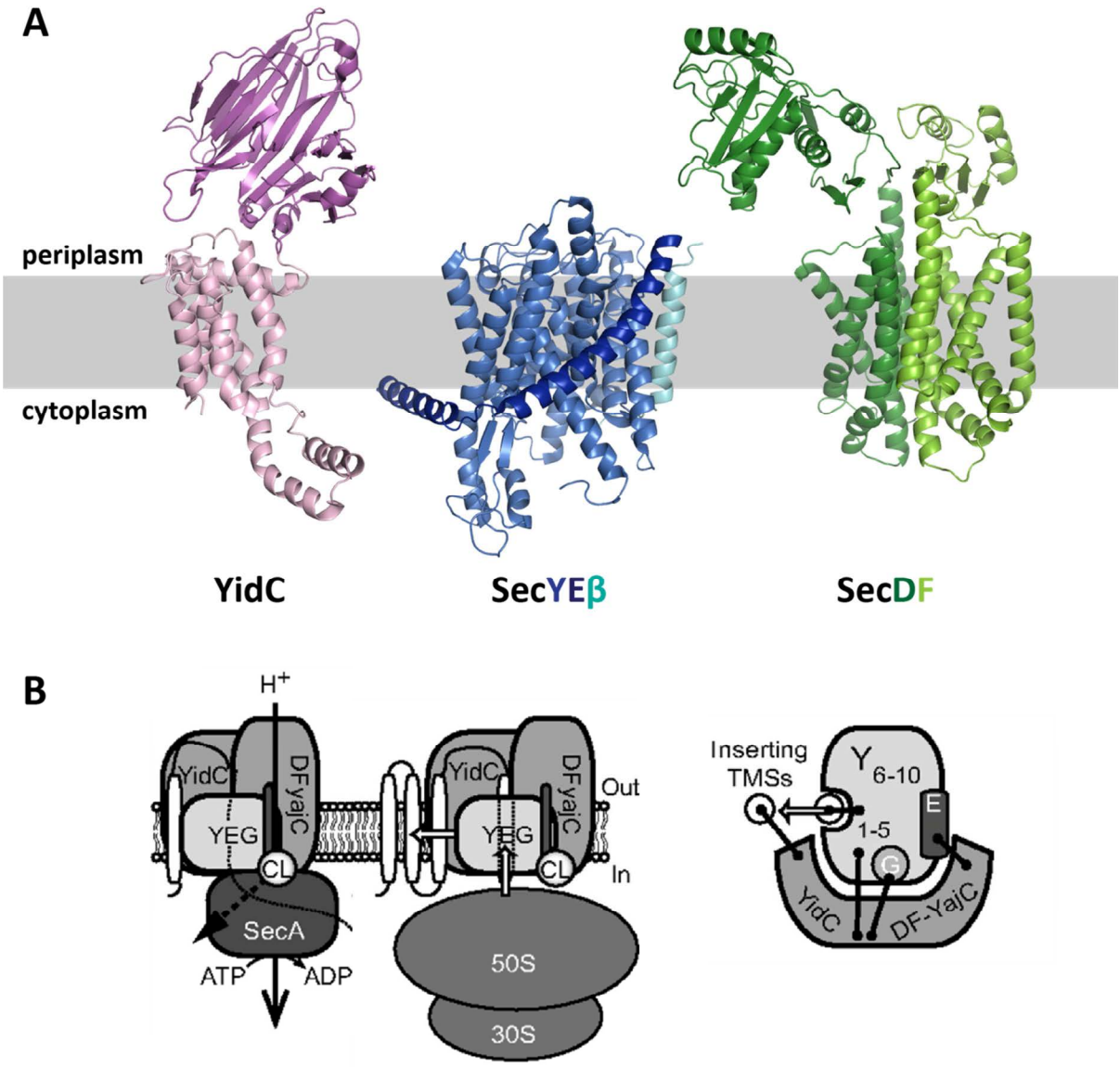


Figure 1-2: Components and structural organization of the HTL. (A) Crystal structures of the SecYEG homolog from *Methanococcus jannaschii* (PDB ID: 1RHZ, Van den Berg *et al.*, 2004); the SecDF homolog in *Thermus thermophilus* (PDB ID: 3AQP, Tsukazaki *et al.*, 2011) and the *E. coli* YidC (PDB ID: 3WVF, Kumazaki *et al.*, 2014) viewed in the membrane plane. (B) Model for structural organization and activity of the HTL in SecA-dependent post-translational translocation (left); in co-translational translocation (center) and the top view of proposed interactions for the HTL. Black bars represent interactions observed from in-membrane cross-linking experiments. Adapted from Schulze *et al.*, 2014.

binding site of translocation substrates and facilitate the insertion of transmembrane helices into the lipid bilayer. The periplasmic domain of YidC has been proposed to comprise a binding site for substrate proteins (Oliver and Paetzel, 2008; Ravaud *et al*., 2008).

The mechanism of co-translational targeting and translocation has been studied for many years. The general process is evolutionary conserved and comparatively well-understood. However, detailed understanding of the regulation of this process and of membrane protein insertion beyond partitioning of the transmembrane helices into the lipid bilayer is lacking to date. Moreover, although the biological functions of these proteins are probably conserved across species, their structure and conformation may vary between species and in the context of the holo-translocon complex. For example, the *Thermus thermophilus* SecDF that has an X-ray structure reported (Figure 1-2) is a single protein, while its equivalent in *E. coli* comprises a larger periplasmic SecD domain and is a tightly associated heterodimer composed of SecD and SecF. A high resolution crystal structure of *E. coli* SecYEG also has not been reported to date. That means that a high-resolution cryo-EM structure of HTL in which transmembrane helices can be visualized and modeled can provide information on the interactions among subunits and on the structure and function of the *E. coli* specific domains in the individual HTL subunits.

In this thesis work, we studied the ribosome-SRP-FtsY complex in the “closed” conformation, and the ribosome-HTL complex.

1.2 Approaches and challenges in studying membrane proteins

1.2.1 Membrane protein purification

Membrane proteins perform a variety of important functions, for example they serve as signal receptors and transporters. They are subject of many high-impact research studies, due to their biological significance, but also because of the difficulties to study them. Their hydrophobic surfaces, flexibility and lack of stability outside of the

biological membrane contribute to the technical barriers of current membrane protein research.

Although some studies can be performed *in situ*, *i.e.* in the original membrane, many biochemical and structural studies require the recombinant expression of membrane proteins and their extraction from the membrane. This procedure can be very challenging as problems often arise from the fact that the expression levels of functional proteins is rather low in comparison to cytosolic proteins and that the subsequent purification of membrane proteins involves a detergent solubilization step.

During the expression of membrane proteins, as mentioned in Section 1.1.3, specific translocation systems are required in the host cell. Some membrane proteins also need the presence of certain lipids. In these cases, the recombinant overexpression of membrane proteins has less chance of success if not in the original species. Therefore, most of the bacterial membrane proteins that are present in the Protein Data Bank (PDB) were expressed in *E. coli*, whereas eukaryotic membrane proteins are often purified from native sources (Carpenter *et al.*, 2008), or more recently, expressed in eukaryotic expression systems (see Section 1.3).

The purification of a membrane protein usually starts by separating the membrane in which the protein is located from the rest of the cell. This is usually achieved by (ultra)centrifugation steps. Sometimes, it is recommended to wash the membrane particles in order to remove membrane-associated proteins and keep only integral membrane proteins. The crucial step in membrane protein purification is to extract the membrane protein(s) of interest from the lipid bilayer using detergents.

1.2.2 Lipids, detergents and their physical properties

The cell membranes contain a variety of biological molecules, mainly lipids and proteins. They form a bilayer structure that can be considered as a two-dimensional liquid in which lipid and protein molecules diffuse more or less easily, as described by the fluid mosaic model (Singer and Nicolson, 1972). Phospholipids, glycolipids, and sterols are the three classes of lipids that form the cell membrane. The lipid

composition of cell membranes varies from one organism to another, and is even different spatially and temporally in a single cell.

In most cases phospholipids are most abundant in the cell membrane. Phospholipids are amphipathic. They have a hydrophilic head group containing the negatively charged phosphate group, and hydrophobic tails usually consisting of two hydrocarbon chains. Their structure determines the fundamental properties of the phospholipids in an aqueous system. The hydrophobic tails tend to cluster together while the head groups face the water, leading to the formation of vesicles (including liposomes) or membranes (Figure 1-3). The hydrocarbon tails of lipids are usually between 4 and 24 carbons long. The length and saturation of these fatty acids determines the transition temperature of the lipids and therefore the fluidity of the lipid bilayer. Most head groups are glyceride, with the exception of sphingomyelin which is derived from sphingosine. Different head groups are involved in the interactions with membrane proteins and membrane-associated proteins.

The fact that membrane proteins are naturally embedded in the lipid bilayer means that they have to be extracted from the bilayer before they can be purified and analyzed further as individually solubilized proteins, which is a prerequisite for most biochemical and structural studies. The extraction from the natural membrane is usually achieved by the use of detergents.

Detergents are amphipathic molecules, just as lipids are. They are composed of hydrophilic heads and hydrophobic tails. What distinguishes detergents from lipids is the concentration range for self-association and the assemblies they can form. Due to the larger head groups, detergent monomers do not self-associate into membranes, but structures called micelles (Figure 1-3). Formation of micelles occurs when the concentration of detergent rises beyond the critical micelle concentration (CMC). When more detergent is added only the concentration of micelles increases. Above the CMC, a detergent becomes capable of solubilizing hydrophobic and amphipathic molecules, such as lipids and membrane proteins, into mixed micelles. When more detergent is added, finally water-soluble detergent-protein complexes can be formed (Figure 1-3) (Garavito and Ferguson-Miller, 2001).

One of the problems of using detergents is their destabilizing effect on membrane proteins. As detergents are applied to break lipid-protein interactions, they may as

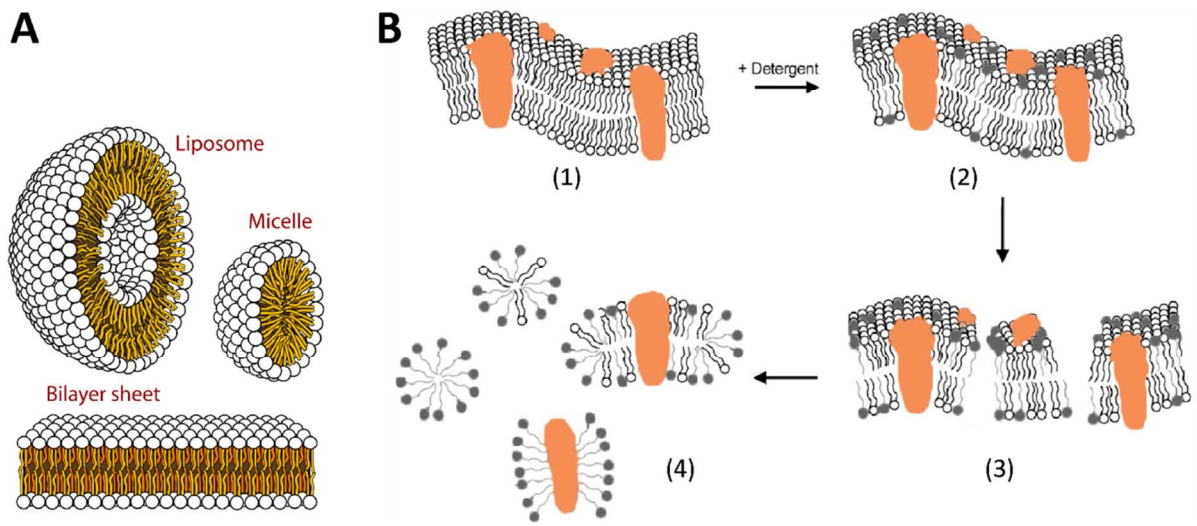


Figure 1-3: Lipid, detergent and membrane protein solubilization. (A) Self-organization of lipids and detergents into a bilayer, like the cell membrane; to a vesicle (liposome) which is double-layered and to a micelle. (Source: Wikipedia) (B) Scheme of membrane protein solubilization by detergents. When small amounts of detergents are added to membranes (1), the detergent (shown in gray with a single tail) penetrates into the membrane with minimal perturbation (2). As the concentration of detergent increases, the membrane bilayer is disrupted (3). With even higher detergent concentrations, mixed micelles begin to form which contain lipids, detergents and membrane proteins (4). Adapted from Kalipatnapu and Chattopadhyay, 2005.

well disrupt the inter- and intra-molecular interactions of membrane proteins, and ultimately denature the target proteins. The problem is more significant when studying membrane protein complexes, as the protein-protein interactions are generally much weaker compared to intra-protein interactions. Therefore, it is crucial to identify the best detergent that balances the solubilization and the stability of the target proteins.

1.2.3 Alternatives to detergents to solubilize membrane proteins

To overcome the problems caused by the use of detergents, alternative methods have been developed to solubilize (or to maintain the solubilization state of) membrane proteins. Here, I describe in more detail two systems that I used during in my thesis work, nanodiscs and amphipols.

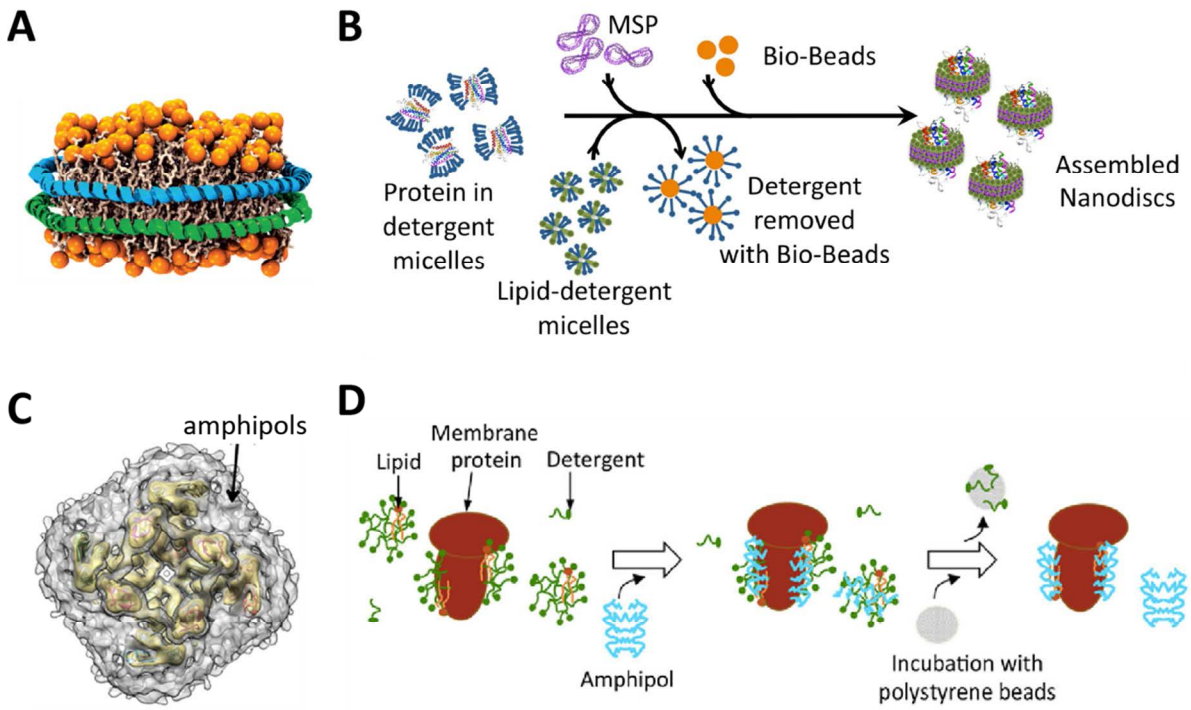


Figure 1-4: Nanodiscs and amphipols. (A) Schematic representation of a nanodisc. The two MSP proteins are shown in blue and green. Hydrophilic head groups of phospholipids are shown in orange. The hydrophobic tails are shown in light tan (Image taken from Shih *et al.*, 2007). (B) The process of nanodisc assembly. MSP and lipids are added to the membrane protein which is solubilized in detergent micelles. Nanodiscs are formed during the removal of detergents from solution (by addition of bio-beads adsorbent) (Image taken from MATER METHODS 2013;3:177). (C) Cryo-EM structure of the TRPV1 channel solubilized by amphipols (Liao *et al.*, 2013). The amphipol belt (in gray) appears as a low-density feature following the transmembrane surface. (D) Schematic representation of transferring a membrane protein from a detergent solution to amphipol. Polystyrene is used as an adsorbent for detergents (Image taken from Zoonens and Popot, 2014).

Nanodiscs are lipid bilayers surrounded and solubilized by membrane scaffold proteins (MSPs) (Nath, Atkins and Sligar, 2007). The engineered MSPs cover the hydrophobic side of the lipid patch, and thus stabilize the disc-shaped particle. Membrane proteins can be integrated into the bilayer, and studied in an environment mimicking the cell membrane. The radius of the nanodiscs can be manipulated by using MSPs of different lengths, and thus incorporate membrane proteins or protein complexes of different sizes. This can be used, for example, to compare the activity of a monomer and a dimer of the same protein (Dalal *et al.*, 2012). Another advantage of using nanodiscs is the integration of lipids which stabilize the

embedded membrane protein, and also enables the observation of protein-lipids interactions (Frauenfeld *et al.*, 2011).

Amphipols are amphipathic polymers. They are designed to bind to transmembrane helices of membrane proteins by multiple contacts, and therefore do not dissociate even at extreme dilutions (Zoonens and Popot, 2014). The multiple contacts also provide a stabilizing effect to the membrane proteins, and are therefore very helpful in studying membrane protein complexes. There are many successful applications of amphipols on structural study of membrane proteins, most notably the 3.4 Å cryo-EM structure of the TRPV1 ion channel (Liao *et al.*, 2013). Many deviates of amphipols exist adapted to different applications, *e.g.* containing biotin moieties or a fluorescent dye. Three types of amphipols, the canonical amphipol A8-35, a sulfonate amphipol (SAPol) and a non-ionic amphipol (NAPol) are explored in this thesis for solubilization of the holo-translocon complex (Section 3.5).

Neither nanodiscs nor amphipols have the ability to solubilize membrane proteins from lipid bilayers. They are usually applied after initially solubilizing membrane proteins by detergents. Subsequently the detergent micelle around the membrane protein(s) is replaced by these systems (Figure 1-4). The principle of the exchange is similar for nanodiscs and amphipols. First, the amphipols or the lipids and MSPs are added, allowing them to mix freely with the detergent and the proteins. Then the detergents are removed from the system, usually by addition of absorbents. With decreasing detergent concentration, interactions between membrane proteins and amphipols or nanodiscs are favored. Finally, proteins solubilized by amphipols or nanodiscs can be separated from other assemblies in the reaction mix by size-exclusion chromatography.

In this thesis, amphipols have shown great potential in stabilizing our holo-translocon membrane protein complex. They also allowed the concentration of our samples after purification, which was not possible in the presence of detergent.

1.3 Publication: Advances and challenges of membrane-protein complex production

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Advances and challenges of membrane–protein complex production

Sarah Zorman[1,2], Mathieu Botte[1,2], Qiyang Jiang[1,2], Ian Collinson[3] and Christiane Schaffitzel[1,2,3]

Most membrane–proteins exist in complexes rather than isolated entities. To fully understand their biological function it is essential to study the intact membrane–protein assemblies. The overexpression and purification of many essential membrane–protein complexes is still a considerable and often unsurmountable challenge. In these cases, extraction from source is the only option for many large multi-subunit cellular machines. Here, we describe recent advances in overexpression of multi-subunit membrane–protein complexes, the strategies to stabilize these complexes and highlight major achievements in membrane–protein structural research that were facilitated by the prospect of achieving subnanometer to near-atomic resolution by electron cryo-microscopy.

Addresses
[1] European Molecular Biology Laboratory, Grenoble Outstation, Grenoble 38042, France
[2] Unit for Virus Host-Cell Interactions, University Grenoble Alpes-EMBL-CNRS, Grenoble 38042 France
[3] School of Biochemistry, University of Bristol, Bristol BS8 1TD, United Kingdom

Corresponding authors: Collinson, Ian (Ian.Collinson@bristol.ac.uk) and Schaffitzel, Christiane (schaffitzel@embl.fr)

Current Opinion in Structural Biology 2015, **32**:123–130

This review comes from a themed issue on **New constructs and expressions of proteins**

Edited by **Imre Berger** and **Roslyn M Bill**

http://dx.doi.org/10.1016/j.sbi.2015.03.010

Introduction

About one third of the proteome of every cell is translocated into a membrane. Recent developments in high-throughput methods to detect and quantify protein–protein interactions described the interactome of soluble proteins in bacteria, yeast, and human [1–3]. However, complexes involving membrane–proteins are much more difficult to identify, produce and characterize due to their hydrophobic nature (Figure 1a). A genome-wide approach revealed the membrane–protein interaction landscape of *Saccharomyces cerevisiae* using affinity-purification in the presence of different non-denaturing detergents followed by mass spectrometry [4]. Remarkably, in the corresponding interactome almost two-thirds of the interactions relate hitherto unassigned functions. Evidently, yeast membrane–proteins have on average 2.1 interaction partners, around half of those prescribed to globular proteins [4,5], demonstrating that, as in the cytosol, most proteins occur and function in complexes rather than as isolated entities in the membrane.

While systems biology approaches help to catalogue the membrane content and membrane–protein interactions, significant progress is needed for the determination of their stoichiometry, structure and cellular function. Here, we review recent approaches to produce, purify and stabilize membrane–protein complexes for these analyses.

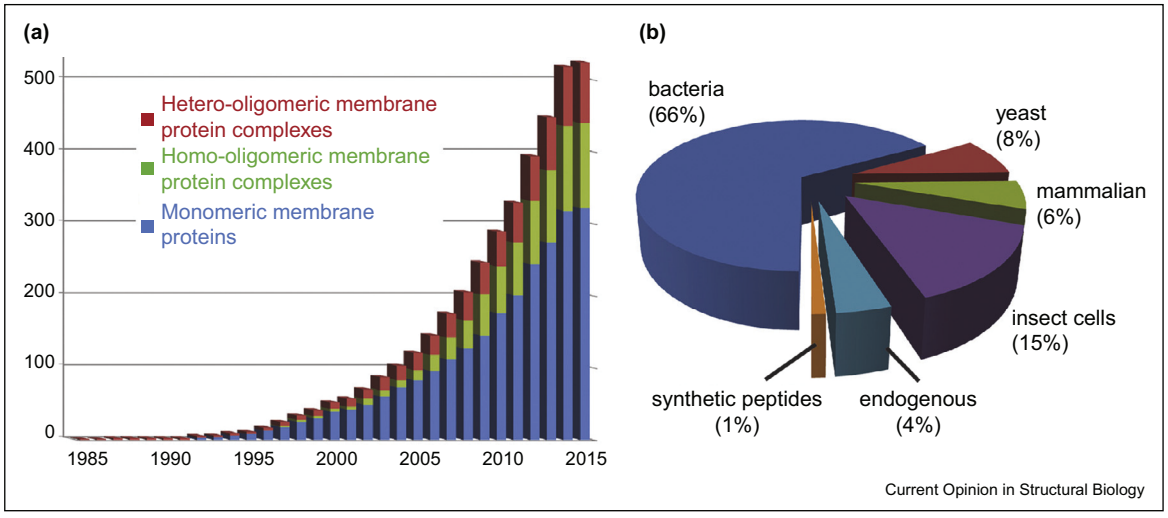
Recombinant membrane–protein complex production in prokaryotes

Escherichia coli has traditionally been used as expression host of choice for membrane–proteins, alongside Gram-positive alternatives [6] (Figure 1b). Recombinant production allows introducing truncations, mutations and tags for affinity-purification. In fact, the first crystal structure of a membrane–protein complex was the *E. coli* fumarate reductase respiratory complex [7,8].

Improved *E. coli* strains for membrane–protein expression

Significant effort has been invested to develop bacterial strains that are tailor-made for membrane–protein production. The T7 RNA polymerase-based expression hosts C41λ(DE3) and C43λ(DE3) have mutations in the *lac*UV5 promoter that governs T7 RNA polymerase expression, resulting in lower amounts of T7 RNA polymerase and consequently in slower transcription and translation rates of the proteins under the control of a T7 promoter [9,10]. Applying a similar strategy, the Lemo21(DE3) strain was designed to precisely control T7 RNA polymerase activity: expression levels of its natural inhibitor T7 lysozyme are titrated by rhamnose induction [10]. Slower membrane–protein production rates can reduce accumulation of misfolded, aggregated proteins in inclusion bodies. The C41λ(DE3) and C43λ(DE3) expression hosts are frequently used; successful examples include the AcrAB-TolC multidrug efflux pump [11•] and the LptD–LptE complex involved

Figure 1



Membrane–protein structures in the Protein Data Base and expression hosts used. **(a)** The number of membrane–proteins deposited in the Protein Data Bank (PDB, www.rcsb.org) is exponentially increasing since 1985, but still presents a small fraction compared to all the proteins in the PDB (currently ~98,770). The majority of the membrane–protein structures are monomeric or from homo-oligomers. Multi-subunit membrane–protein complexes are particularly challenging, and thus only 84 structures are currently available. **(b)** Expression hosts used for membrane–protein production for structures deposited in the PDB between 2010 and 2015. Membrane–proteins are traditionally expressed in bacteria (66%), the vast majority in *Escherichia coli*. However, eukaryotic expression systems are clearly becoming more important; in particular baculovirus-insect cell expression (15%) is successfully used for production of G protein-coupled receptors.

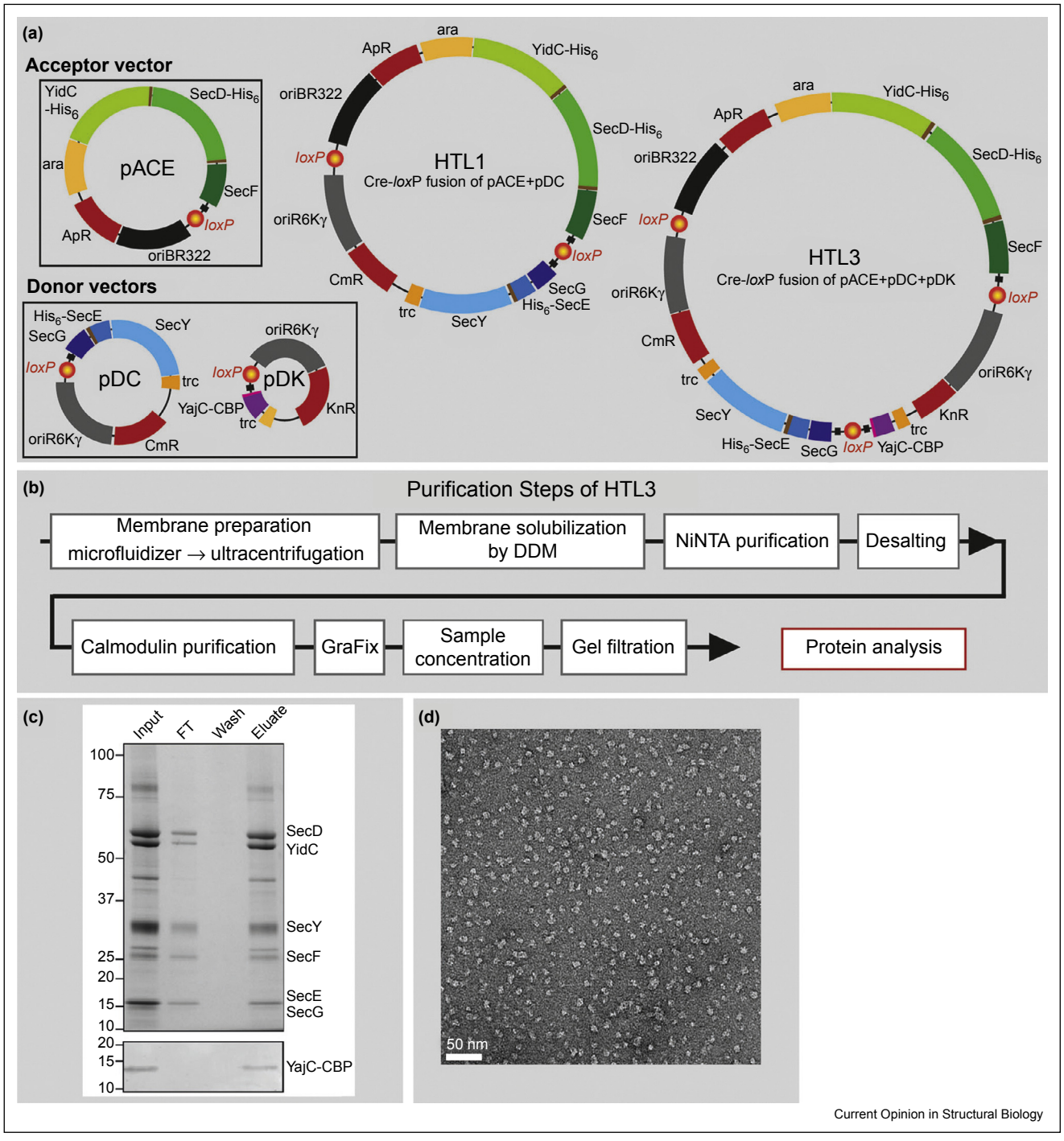
in lipopolysaccharide translocation into the outer membrane of Gram-negative bacteria [12] (depicted in Figure 3b). The twin-arginine translocase inner-membrane subunit (TatC) is one of many examples for successful membrane–protein expression in a Lemo(DE3) strain [13]. Additionally, auto-induction-based media which slow down protein expression are now commonly used to improve membrane–protein production [14].

Co-expression systems for protein complexes

Membrane–protein complexes usually cannot be reconstituted from the purified subunits or subcomplexes. This could be due to the fact that they need interaction partners for stable folding and function or because the protein–protein interactions formed by interfaces within the bilayer are masked by the detergent used for solubilisation. Co-expression of the complex subunits overcomes this problem. Co-transformation of several plasmids with single or multiple expression cassettes and with different origins of replication and antibiotic resistance [15,16] has the drawback that the copy numbers of the respective plasmids can vary significantly. This complicates the balanced expression of the subunits and adjusting the stoichiometry of the complex. Poly-cistronic expression systems using natural or artificial operons where the complex components are expressed from one messenger RNA often result in a more balanced protein production [17]. In the case of artificial operons the order of genes in the construct needs to be experimentally determined because expression levels depend on intrinsic properties of the particular coding sequences [18]. Moreover, larger operons with many genes often lead to low expression levels of the proteins encoded downstream. The trimeric Sec translocon complex was expressed from a synthetic poly-cistronic mRNA, purified and successfully crystallized with and without the partner ATPase SecA [19,20]. The approach is not restricted to *E. coli*; the Na^+-pumping NADH:quinone oxidoreductase complex was produced by homologous expression of the operon encoding its six subunits in *Vibrio cholerae* [21•]. The *E. coli* AcrABZ-TolC multidrug efflux pump recently has been elegantly produced in C43(DE3) using co-expression from a vector encoding an AcrA–AcrB fusion and a pETDUET vector [16] encoding an AcrA–AcrZ fusion-protein and TolC [11•]. By using two different AcrA fusion-proteins and thus providing two AcrA copies, the authors stabilized the AcrA:AcrB:AcrZ:TolC complex for characterization by cryo-EM and favoured the formation of a complex with a 6:3:3:3 stoichiometry [11•].

The ACEMBL system for multi-protein expression in *E. coli* uses small, designed acceptor and donor vectors [22]. Individual genes or poly-cistrons are inserted into the multiple-integration element of the acceptor and into one or several donor vectors with different antibiotic resistance. Donor vectors have a conditional origin of replication (oriR6Kγ in Figure 2a) and need to be propagated in strains encoding the phage R6Kγ *pir* gene. Incubation of

Figure 2



ACEMBL system for multi-protein complex production used for holo-translocon production. **(a)** The ACEMBL system consists of acceptor and donor vectors (left) with conventional BR322 (black) and conditional R6Kγ (grey) origin of replication respectively. For expression of the holo-translocon (HTL) an acceptor vector (pACE) carrying a poly-cistron encoding for YidC, SecD, and SecF was combined with a donor vector pDC containing a poly-cistron encoding for SecY, SecE, SecG. The position of hexa-histidine-tags in YidC, SecD and Sec E is indicated in brown. The resulting HTL1 construct was combined by Cre-*lox*P fusion with a second donor vector pDK encoding YajC with a C-terminal CBP-tag (indicated in pink) yielding HTL3. Arabinose (ara) and trc promoters are shown in orange, terminators as a black square, antibiotic resistance marker genes (ApR, ampicillin; KnR, kanamycin; CmR, chloramphenicol) in red. **(b)** Scheme for holo-translocon complex purification using the HTL3 construct. **(c,d)** Holo-translocon analysis by Coomassie-stained SDS/PAGE (c, above) and Western Blot using an antibody directed against the CBP-tag (c, below) demonstrating the presence of all seven membrane–proteins and by negative stain electron microscopy (d) showing the homogeneity of the complex.

donor and acceptor vectors with Cre recombinase and transformation into a *pir*$^{-}$ strain leads to generation of fusion plasmids that can be selected for by the appropriate choice of antibiotics (Figure 2a). The ACEMBL system was applied for the expression of the *E. coli* holo-translocon complex consisting of seven membrane–proteins (SecY, SecE, SecG, SecD, SecF, YajC and YidC), one of the largest recombinant membrane–protein complexes produced to date (Figure 2a) [22,23•]. The modular architecture of the individual plasmids allowed testing different promoter combinations and affinity tags to achieve a balanced expression of the subunits and to optimize complex purification (Figure 2b–d). It also enabled the expression of the heterotetrameric SecDF–YajC–YidC subcomplex and a subcomplex devoid of YidC.

Purification and stabilization of membrane–protein complexes

Affinity tags are usually attached to the C-terminus of membrane–proteins since N-terminal tagging may interfere with proper targeting and translocation. For large and less stable complexes the use of more than one affinity tag attached to different subunits facilitates tandem affinity purification; for instance, the holo-translocon could be efficiently enriched by Ni-NTA chromatography and subsequently separated from excess SecYEG subcomplexes by calmodulin-affinity chromatography using the CBP-tag fused to YajC (Figure 2a,c) [23•].

The choice of detergent is crucial because the solubilisation of the complex should be efficient without compromising the integrity of the membrane–protein complex (Figure 3a,b). Mild detergents such as digitonin are frequently used for cryo-EM studies such as the ribosome-Sec61 translocon complex for example [24,25]. More recently, new classes of detergents were developed to increase the protein stability, vary the size of the protein–detergent complex and enhance chances of obtaining crystals. Impressive results in stabilization of eukaryotic membrane–proteins, which often are less stable in solution than their prokaryotic homologues, were obtained with maltose-neopentyl glycol (MNG) characterized by two hydrophilic and two lipophilic subunits linked via a central quaternary carbon [26]. MNGs have a particular low critical micelle concentration allowing close to detergent-free buffers. For example, lauryl maltose-neopentyl glycol (LMNG) was used for crystallization of several G-protein-coupled receptors (GPCRs) such as the human β2 adrenergic receptor (β2AR)–β-arrestin-1 complex [27] and of the integral membrane–protein of the twin-arginine translocase (TatC) [13]. More recently, newly designed steroid-based facial amphiphiles were shown to form relatively small protein–detergent complexes and to stabilize flexible membrane–proteins, thus improving crystallisability [28].

Amphiphilic polymers (Amphipols) are particularly useful for membrane–protein stabilization in detergent-free solution (Figure 3a), because their binding is irreversible in the absence of competing detergents and lipids [29]. Amphipols form a compact layer around the transmembrane region, resulting in nearly monodisperse complexes. Moreover, they were reported to reduce protein dynamics. The first cryo-EM structure in presence of amphipols reported was the mitochondrial super-complex $I_1II_2IV_1$ [30]. Since then amphipols continue to be successfully used for cryo-EM studies. Recent subnanometer-resolution cryo-EM structures using amphipols include the mammalian transient receptor potential voltage-gated (TRPV1) channel [31••] and the human γ-secretase (see below, Figure 3c) [32••].

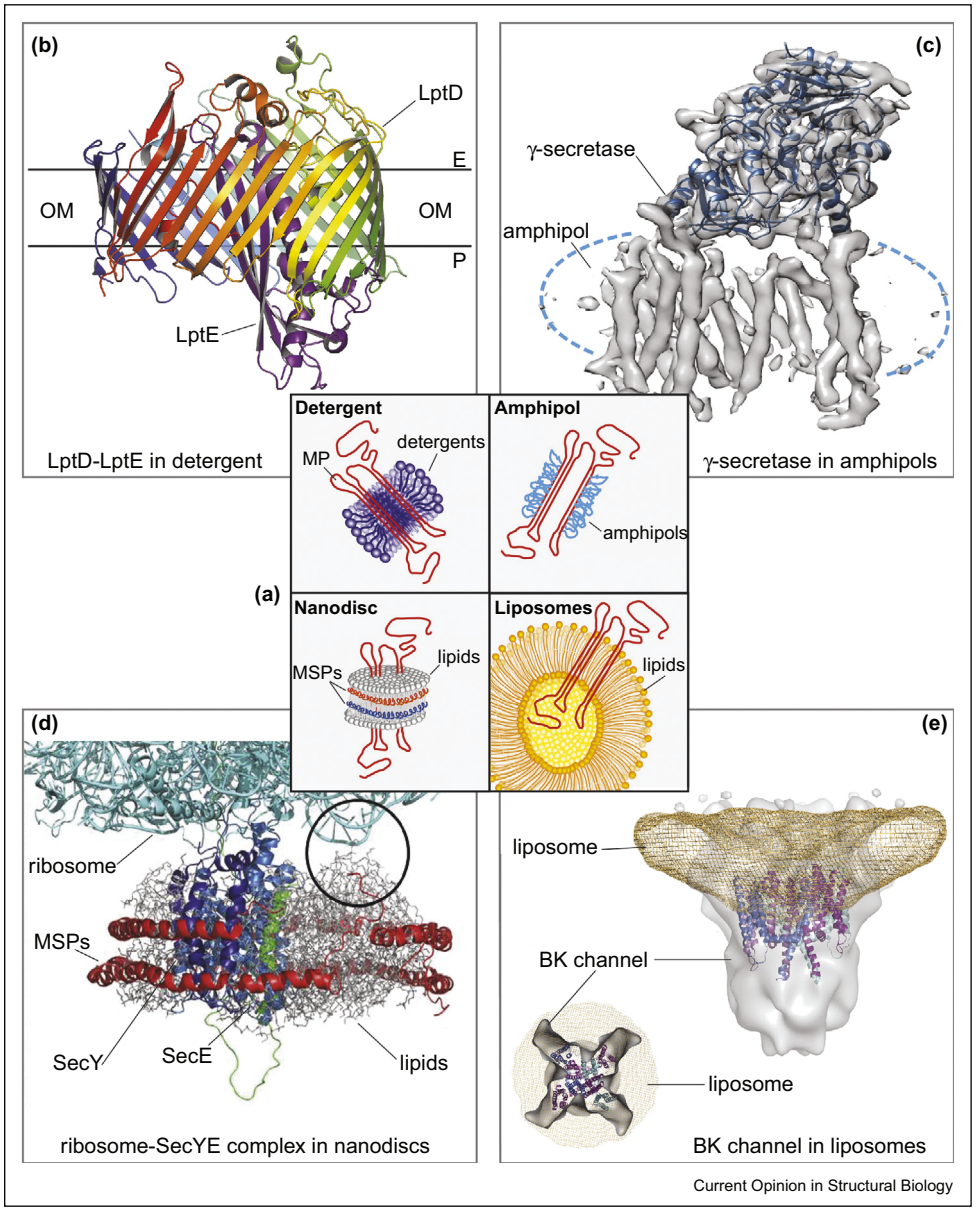
Nanodiscs are formed by the reconstitution of membrane–protein complexes within an annulus of phospholipids encased by an engineered membrane scaffold protein (MSP) (Figure 3a). This results in a disc-shaped lipid bilayer with embedded protein(s) surrounded by two MSPs which cover the hydrophobic moieties of the lipid leaflets [33]. Nanodiscs were used for structural characterization of ribosome-SecYE translocon complexes [34] (Figure 3d), and for the cryo-EM structure of the homotetrameric ryanodine receptor responsible for calcium release from the sarcoplasmic reticulum [35]. Nanodiscs have the advantage that membrane–proteins are stabilized in a near-native lipid environment, and detergent-free physiological buffers can be used, facilitating biochemical characterization [36,37]. Nanodiscs are tailored to accommodate proteins of different sizes by using MSPs of different lengths, however larger nanodiscs (diameter >15 nm) are often less stable.

Liposomes (Figure 3a) are commonly used for functional analyses of membrane–proteins; but can also be valuable for cryo-EM studies, for example, the structure of the human large-conductance calcium-activated and voltage-activated potassium channel [38] (Figure 3e). Other technologies to stabilize membrane–protein complexes, such as bicelles, lipidic cubic phases and high concentrations of lipid and detergent, as well as the use of mutagenesis-based screening and selection methods are reviewed for instance in Ref. [39].

Eukaryotic expression systems for membrane–protein complexes

In general, it is reasonable to choose an expression system closely related to the natural host of the protein, mimicking the natural membrane lipid composition and chaperone repertoire. The first crystal structures of recombinant mammalian membrane–proteins were obtained by overexpression in *S. cerevisiae* [40] and *Pichia pastoris* [41]. Today, yeast expression is still frequently used (Figure 1b) resulting for instance in recent crystal structures of the calcium-activated lipid scramblase TMEM16 [42], and of the calcium-activated chloride channel bestrophin1 [43].

Figure 3



Strategies for solubilisation and stabilization of membrane–protein complexes. **(a)** Scheme showing different approaches (described in the main text) to stabilize membrane–proteins and their complexes. **(b)** Crystal structure of the *Salmonella typhimurium* LptD–LptE complex solubilised in N-octyl-β-D-glucopyranoside, one of the most important detergents for purification of membrane–proteins. The complex is responsible for the lipopolysaccharide transport from the inner to the outer membrane ([12], PDB-ID: 2GUF). E, P and OM indicate the extracellular space, periplasm and outer membrane, respectively. **(c)** Cryo-EM structure and partial atomic model of the human γ-secretase complex consisting of four subunits in amphipols at 4.5 Å resolution ([32••], EMDB-ID: EMD-2677, PDB-ID: 4UPC). **(d)** Quasi-atomic model derived from the cryo-EM structure of a translating *E. coli* ribosome bound to the SecYE translocon reconstituted in a nanodisc ([34], PDB-ID: 3J00). The black circle highlights a previously unknown interaction between the ribosome and the lipid bilayer. MSP: membrane scaffold protein. **(e)** Cryo-EM structure and partial atomic model of the human BK potassium channel reconstituted in liposomes at 17–20 Å resolution ([38]; EMDB-ID: EMD-5114 and EMD-5121). A slice through the EM density in the plane of the membrane is shown on the bottom left, illustrating the fourfold symmetry.

The majority of the recent membrane–protein structures from eukaryotic expression hosts were obtained by baculovirus-insect cell expression (Figure 1b). In particular GPCR studies were hampered for many years by the difficulty in the production of these rather instable proteins with seven transmembrane-spanning regions [44]. The human β2AR–β-arrestin-1 complex was obtained by co-expression in insect cells [27]. The heterodimeric human γ-aminobutyric acid class B ($GABA_B$) GPCR involved in neurotransmission inhibition in the brain [45] was

produced using the MultiBac-insect cell expression system [46].

Transient expression of the human γ-secretase complex in HEK293F cells [32••] was achieved using the pMLink vector system [47] in which plasmids encoding the four subunits were combined by ligation-independent cloning. Subsequently, the structure of this small, 170 kDa complex was determined by single-particle cryo-EM at 4.5 Å resolution [32••].

Large membrane–protein complexes from endogenous source

In spite of remarkable progress in assembling membrane–protein complexes, the complexity of many large cellular machines is such that they cannot be produced recombinantly. Therefore, the only current option is to isolate them from their natural sources [48]. In the past this has restricted high-resolution structural analysis to naturally abundant material, such as the photosynthetic complexes [49] and the apparatus for respiration and oxidative phosphorylation of bacteria, chloroplasts and mitochondria [50•,51]. The approach relied predominantly on X-ray crystallography. Progress in this area over the past ten years has been remarkable, driven by the increased availability and quality of a wide range of detergents, the miniaturization and automation of crystallization screening, as well as ready access to intense micro-focus beam-lines, as the number of available membrane–protein structures testifies (Figure 1a).

The recent revolution in resolution in cryo-EM has provided new opportunities for the analysis of very large assemblies, especially of membrane–proteins. The amounts of protein required are very much reduced in cryo-EM, compared to X-ray crystallography. Moreover, the inevitable increase in the flexibility and heterogeneity in giant assemblies, which previously precluded their high-resolution analysis, can now be overcome through computational 'purification' of the sample and by classification into different conformations providing valuable insights into the dynamism of these machines. Examples already emerged and structures are approaching atomic resolution, including ribosome-Sec61 complexes [25] and bovine complex I [50•] amongst others. Finally, EM is extremely useful to screen for stabilizing factors and conditions that promote complex homogeneity for high-resolution studies; as exemplified for the β(2)-adrenoceptor-G-protein Gs complex EM structure [52].

Concluding remarks

With the recent advances in recombinant expression of membrane–proteins, in particular using eukaryotic expression hosts, the overexpression of rare, large and less stable multi-subunit membrane–protein complexes comes within reach. The development of new detergents, amphipols and nanodiscs which allow handling these complexes in virtually detergent-free physiological buffer is crucial for the stabilization of many complexes. Undoubtedly, the recent developments in cryo-EM and image processing will significantly speed up the structural characterization of larger membrane–protein complexes which are difficult to crystallize. Still, the production and characterization of membrane–protein complexes is extremely challenging, for instance due to our insufficient knowledge about specific interactions with lipids, sugars and ions which may be crucial for proper function of the proteins. Moreover, differences in lipid composition, glycosylation (which is absent in *E. coli*) and the repertoire of membrane–protein chaperones can have significant effects on the insertion, folding, complex assembly and function of the membrane–proteins when expressed in a heterologous system. Clearly, membrane–protein complexes will continue to be a rich field of discovery for the decades to come.

Conflicts of interest

The authors confirm that there are no known conflicts of interest.

Acknowledgements

CS acknowledges support by an ERC Starting Grant (project 281331), by a Sinergia grant from the Swiss National Science Foundation (project CRSII3_136254), and by contract research 'Methoden für die Lebenswissenschaften' of the Baden-Württemberg Stiftung.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Butland G, Peregrín-Alvarez JM, Li J, Yang W, Yang X, Canadien V, Starostine A, Richards D, Beattie B, Krogan N *et al.*: **Interaction network containing conserved and essential protein complexes in *Escherichia coli***. *Nature* 2005, **433**:531-537.

2. Gavin A-C, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dümpelfeld B *et al.*: **Proteome survey reveals modularity of the yeast cell machinery**. *Nature* 2006, **440**:631-636.

3. Havugimana PC, Hart GT, Nepusz T, Yang H, Turinsky AL, Li Z, Wang PI, Boutz DR, Fong V, Phanse S *et al.*: **A census of human soluble protein complexes**. *Cell* 2012, **150**:1068-1081.

4. Babu M, Vlasblom J, Pu S, Guo X, Graham C, Bean BDM, Burston HE, Vizeacoumar FJ, Snider J, Phanse S *et al.*: **Interaction landscape of membrane–protein complexes in *Saccharomyces cerevisiae***. *Nature* 2012, **489**:585-589.

5. Gavin A-C, Bösche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon A-M, Cruciat C-M *et al.*: **Functional organization of the yeast proteome by systematic analysis of protein complexes**. *Nature* 2002, **415**:141-147.

6. Schlegel S, Hjelm A, Baumgarten T, Vikström D, de Gier J-W: **Bacterial-based membrane protein production**. *Biochim Biophys Acta* 2014, **1843**:1739-1749.

7. Luna-Chavez C, Iverson TM, Rees DC, Cecchini G: **Overexpression, purification, and crystallization of the membrane-bound fumarate reductase from *Escherichia coli***. *Protein Expr Purif* 2000, **19**:188-196.

8. Iverson TM, Luna-Chavez C, Cecchini G, Rees DC: **Structure of the *Escherichia coli* fumarate reductase respiratory complex**. *Science* 1999, **284**:1961-1966.

9. Miroux B, Walker JE: **Over-production of proteins in *Escherichia coli:* mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels**. *J Mol Biol* 1996, **260**:289-298.

10. Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, Högbom M, van Wijk KJ, Slotboom DJ, Persson JO *et al.*: **Tuning *Escherichia coli* for membrane protein overexpression**. *Proc Natl Acad Sci U S A* 2008, **105**:14371-14376.

11. • Du D, Wang Z, James NR, Voss JE, Klimont E, Ohene-Agyei T, Venter H, Chiu W, Luisi BF: **Structure of the AcrAB-TolC multidrug efflux pump**. *Nature* 2014, **509**:512-515.
An elegant strategy was applied to stabilize the AcrABZ-TolC multiefflux pump which spans the inner and outer membrane of *Escherichia coli.* The authors generated fusion proteins of AcrA–AcrB and of AcrA–AcrZ which were co-expressed with TolC thereby favoring the correct stoichiometry of the complex.

12. Dong H, Xiang Q, Gu Y, Wang Z, Paterson NG, Stansfeld PJ, He C, Zhang Y, Wang W, Dong C: **Structural basis for outer membrane lipopolysaccharide insertion**. *Nature* 2014, **511**:52-56.

13. Rollauer SE, Tarry MJ, Graham JE, Jääskeläinen M, Jäger F, Johnson S, Krehenbrink M, Liu S-M, Lukey MJ, Marcoux J *et al.*: **Structure of the TatC core of the twin-arginine protein transport system**. *Nature* 2012, **492**:210-214.

14. Studier FW: **Protein production by auto-induction in high-density shaking cultures**. *Protein Expr Purif* 2005, **41**:207-234.

15. Chanda PK, Edris WA, Kennedy JD: **A set of ligation-independent expression vectors for co-expression of proteins in *Escherichia coli***. *Protein Expr Purif* 2006, **47**:217-224.

16. Tolia NH, Joshua-Tor L: **Strategies for protein coexpression in *Escherichia coli***. *Nat Methods* 2006, **3**:55-64.

17. Selleck W, Tan S: **Recombinant protein complex expression in *E. coli***. *Curr Protoc Protein Sci* 2008. Chapter 5:Unit 5.21.

18. Smolke CD, Keasling JD: **Effect of gene location, mRNA secondary structures, and RNase sites on expression of two genes in an engineered operon**. *Biotechnol Bioeng* 2002, **80**:762-776.

19. Van den Berg B, Clemons WM, Collinson I, Modis Y, Hartmann E, Harrison SC, Rapoport TA: **X-ray structure of a protein-conducting channel**. *Nature* 2004, **427**:36-44.

20. Zimmer J, Nam Y, Rapoport TA: **Structure of a complex of the ATPase SecA and the protein-translocation channel**. *Nature* 2008, **455**:936-943.

21. • Steuber J, Vohl G, Casutt MS, Vorburger T, Diederichs K, Fritz G: **Structure of the *V. cholerae* Na^+-pumping NADH:quinone oxidoreductase**. *Nature* 2014, **516**:62-67.
Important first insights into the molecular mechanism of redox-driven Na^+ translocation are provided by the X-ray structure of the Na^+-pumping NADH:quinone oxidoreductase from *Vibrio cholerae*. The six subunits were produced by homologous expression of the natural operon from a plasmid and by fusing a hexahistidine-tag to one of the subunits

22. Bieniossek C, Nie Y, Frey D, Olieric N, Schaffitzel C, Collinson I, Romier C, Berger P, Richmond TJ, Steinmetz MO *et al.*: **Automated unrestricted multigene recombineering for multiprotein complex production**. *Nat Methods* 2009, **6**:447-450.

23. • Schulze RJ, Komar J, Botte M, Allen WJ, Whitehouse S, Gold VAM, Lycklama A, Nijeholt JA, Huard K, Berger I, Schaffitzel C *et al.*: **Membrane protein insertion and proton-motive-force-dependent secretion through the bacterial holo-translocon SecYEG–SecDF–YajC–YidC**. *Proc Natl Acad Sci U S A* 2014, **111**:4844-4849.
The ACEMBL system is successfully applied for recombinant expression of the holo-translocon complex with seven integral membrane–protein subunits. The 250 kDa complex is active in co-translational and post-translational protein translocation.

24. Gogala M, Becker T, Beatrix B, Armache J-P, Barrio-Garcia C, Berninghausen O, Beckmann R: **Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion**. *Nature* 2014, **506**:107-110.

25. Voorhees RM, Fernández IS, Scheres SHW, Hegde RS: **Structure of the mammalian ribosome-Sec61 complex to 3.4 Å resolution**. *Cell* 2014, **157**:1632-1643.

26. Chae PS, Rasmussen SGF, Rana RR, Gotfryd K, Chandra R, Goren MA, Kruse AC, Nurva S, Loland CJ, Pierre Y *et al.*: **Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins**. *Nat Methods* 2010, **7**:1003-1008.

27. Shukla AK, Westfield GH, Xiao K, Reis RI, Huang L-Y, Tripathi-Shukla P, Qian J, Li S, Blanc A, Oleskie AN *et al.*: **Visualization of arrestin recruitment by a G-protein-coupled receptor**. *Nature* 2014, **512**:218-222.

28. Lee SC, Bennett BC, Hong W-X, Fu Y, Baker KA, Marcoux J, Robinson CV, Ward AB, Halpert JR, Stevens RC *et al.*: **Steroid-based facial amphiphiles for stabilization and crystallization of membrane proteins**. *Proc Natl Acad Sci U S A* 2013, **110**:E1203-E1211.

29. Zoonens M, Popot J-L: **Amphipols for each season**. *J Membr Biol* 2014, **247**:759-796.

30. Althoff T, Mills DJ, Popot J-L, Kühlbrandt W: **Arrangement of electron transport chain components in bovine mitochondrial supercomplex $I_1III_2IV_1$**. *EMBO J* 2011, **30**:4652-4664.

31. •• Liao M, Cao E, Julius D, Cheng Y: **Structure of the TRPV1 ion channel determined by electron cryo-microscopy**. *Nature* 2013, **504**:107-112.
A landmark paper describing the cryo-EM structure of a relatively small membrane-protein complex, an ion channel, at 3.4 Å resolution. At this resolution, the authors couldtrace the protein backbone and to recognize amino acid side chains.

32. •• Lu P, Bai X, Ma D, Xie T, Yan C, Sun L, Yang G, Zhao Y, Zhou R, Scheres SHW *et al.*: **Three-dimensional structure of human γ-secretase**. *Nature* 2014, **512**:166-170.
This study highlights the power of cryo-EM for small, non-symmetric membrane–protein complexes: the structure of human γ-secretase complex was solved at 4.5 Å resolution. The four subunits were transiently expressed in mammalian cells and the complex was stabilized using amphipols.

33. Alami M, Dalal K, Lelj-Garolla B, Sligar SG, Duong F: **Nanodiscs unravel the interaction between the SecYEG channel and its cytosolic partner SecA**. *EMBO J* 2007, **26**:1995-2004.

34. Frauenfeld J, Gumbart J, van der Sluis EO, Funes S, Gartmann M, Beatrix B, Mielke T, Berninghausen O, Becker T, Schulten K *et al.*: **Cryo-EM structure of the ribosome-SecYE complex in the membrane environment**. *Nat Struct Mol Biol* 2011, **18**:614-621.

35. Efremov RG, Leitner A, Aebersold R, Raunser S: **Architecture and conformational switch mechanism of the ryanodine receptor**. *Nature* 2015, **517**:39-43.

36. Bocquet N, Kohler J, Hug MN, Kusznir EA, Rufer AC, Dawson RJ, Hennig M, Ruf A, Huber W, Huber S: **Real-time monitoring of binding events on a thermostabilized human A_{2A} receptor embedded in a lipid bilayer by surface plasmon resonance**. *Biochim Biophys Acta* 2015 http://dx.doi.org/10.1016/j.bbamem.2015.02.014.

37. Adamson RJ, Watts A: **Kinetics of the early events of GPCR signalling**. *FEBS Lett* 2014, **588**:4701-4707.

38. Wang L, Sigworth FJ: **Structure of the BK potassium channel in a lipid membrane from electron cryomicroscopy**. *Nature* 2009, **461**:292-295.

39. Kang HJ, Lee C, Drew D: **Breaking the barriers in membrane protein crystallography**. *Int J Biochem Cell Biol* 2013, **45**:636-644.

40. Jidenko M, Nielsen RC, Sørensen TL-M, Møller JV, le Maire M, Nissen P, Jaxel C: **Crystallization of a mammalian membrane protein overexpressed in *Saccharomyces cerevisiae***. *Proc Natl Acad Sci U S A* 2005, **102**:11687-11691.

41. Long SB, Campbell EB, Mackinnon R: **Crystal structure of a mammalian voltage-dependent Shaker family K^+ channel**. *Science* 2005, **309**:897-903.

42. Brunner JD, Lim NK, Schenck S, Duerst A, Dutzler R: **X-ray structure of a calcium-activated TMEM16 lipid scramblase**. *Nature* 2014, **516**:207-212.

43. Kane Dickson V, Pedi L, Long SB: **Structure and insights into the function of a Ca(2+)-activated Cl(−) channel**. *Nature* 2014, **516**:213-218.

44. Maeda S, Schertler GFX: **Production of GPCR and GPCR complexes for structure determination**. *Curr Opin Struct Biol* 2013, **23**:381-392.

45. Geng Y, Bush M, Mosyak L, Wang F, Fan QR: **Structural mechanism of ligand activation in human GABA(B) receptor**. *Nature* 2013, **504**:254-259.

46. Vijayachandran LS, Viola C, Garzoni F, Trowitzsch S, Bieniossek C, Chaillet M, Schaffitzel C, Busso D, Romier C, Poterszman A *et al.*: **Robots, pipelines, polyproteins: enabling multiprotein expression in prokaryotic and eukaryotic cells**. *J Struct Biol* 2011, **175**:198-208.

47. Scheich C, Kümmel D, Soumailakakis D, Heinemann U, Büssow K: **Vectors for co-expression of an unrestricted number of proteins**. *Nucleic Acids Res* 2007, **35**:e43.

48. Mesa P, Deniaud A, Montoya G, Schaffitzel C: **Directly from the source: endogenous preparations of molecular machines**. *Curr Opin Struct Biol* 2013, **23**:319-325.

49. Kupitz C, Basu S, Grotjohann I, Fromme R, Zatsepin NA, Rendek KN, Hunter MS, Shoeman RL, White TA, Wang D *et al.*: **Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser**. *Nature* 2014, **513**:261-265.

50. • Vinothkumar KR, Zhu J, Hirst J: **Architecture of mammalian respiratory complex I**. *Nature* 2014, **515**:80-84.
State-of-the-art cryo-EM study of mammalian complex I at 5 Å revealing the architecture of 28 of the 45 subunits. This work is an example for the power of cryo-EM for large multi-subunit membrane–protein complexes extracted from source.

51. Allegretti M, Klusch N, Mills DJ, Vonck J, Kühlbrandt W, Davies KM: **Horizontal membrane-intrinsic α-helices in the stator α-subunit of an F-type ATP synthase**. *Nature* 2015. (advance online publication).

52. Westfield GH, Rasmussen SGF, Su M, Dutta S, DeVree BT, Chung KY, Calinski D, Velez-Ruiz G, Oleskie AN, Pardon E *et al.*: **Structural flexibility of the Gαs α-helical domain in the β2-adrenoceptor Gs complex**. *Proc Natl Acad Sci U S A* 2011, **108**:16086-16091.

1.4 Recent advances in electron cryo-microscopy

1.4.1 Three-dimensional electron microscopy

For many years, X-ray crystallography was the major technique to solve protein structures at high resolution. X-ray crystallography requires the crystallization of the sample which can be very difficult in particular for membrane proteins. Nuclear magnetic resonance (NMR) spectroscopy is another powerful tool in structural biology, but is generally limited to small proteins (< 35 kD) and requires large amount of the sample with high concentration. In contrast, electron cryo-microscopy (cryo-EM) allows the direct observation of specimens provided that the sample is sufficiently large (≥ 200 kD). In addition, cryo-EM eliminates the phase problem of X-ray crystallography, because the sample is visualized in real space. Moreover, the amount of purified sample required is significantly less for cryo-EM studies. Because of these advantages, cryo-EM has become increasingly popular in structural biology, particularly for membrane protein complexes and large biological assemblies.

Cryo-EM is a form of transmission electron microscopy, in which a beam of electrons is transmitted through the specimen. An image is formed from the interaction of the electrons with the specimen and is magnified and focused onto an imaging device. The principle is the same as for light microscopy, except that the small de Broglie wavelength of electrons allows imaging at a significantly higher resolution. The most advanced electron microscopy applications nowadays can achieve a resolution below 0.5 Ångströms (Erni *et al.*, 2009). However, the fragile nature of biological samples does not allow the application of the full strength of electron microscopy (EM). For the study of protein samples, the specimen is usually placed at cryogenic temperatures (generally liquid nitrogen temperature), with limited exposure to the electron beam to reduce the radiation damage caused by the high energy of the electrons. To date, the highest resolution archived for a biological sample is a ~2.9 Å structure for the *E. coli* ribosome in complex with elongation factor Tu (Fischer *et al.*, 2015). This cryo-EM structure provides comparable information to the best resolved X-ray structure of the *E. coli* 70S ribosome (Noeske *et al.*, 2014).

EM images are two dimensional. They represent the projections of the objects along the axis of the beam. There are two major ways of retrieving the information of the third dimension. Either by tilting the specimen to known degrees and therefore obtaining the three-dimensional (3D) structure of the same object; or by gathering a set of two-dimensional (2D) images of objects of the same kind but at different orientations. In both cases, the 3D structure is afterwards computationally reconstructed using the 2D images. The first strategy, called electron cryo-tomography, is currently mainly used to study large macromolecular structures and cellular organizations towards sub-nanometer resolution. The second approach, named single particle reconstruction, is suitable for obtaining structures of proteins at near-atomic resolution.

Next to the quality of the sample, the two critical issues in structure determination by cryo-EM are 1) the quality of the 2D images, and 2) how the 2D images are used to reconstruct the 3D model. Recent advances in microscope hardware and image processing are the driving force for being able to solve cryo-EM structures at high resolution.

1.4.2 Advances in microscope hardware

Radiation damage strictly limits the number of electrons that can be used per sample. This is the reason why any individual image is noisy and incomplete. Tens or hundreds of thousands of such images need to be aligned and averaged in order to retrieve the complete information with high-resolution details. The quality of the image plays a decisive role for at least two reasons: the high-resolution information needs to be recorded in the first place, such that it may be used in the final reconstruction. And second, the image has to be good enough for the software to determine its orientation in relation to the 3D model. The information from an image is helpful only if the image is correctly aligned. This is why a low quality of images cannot be easily compensated for by increasing the amount of images.

During the imaging process, major loss of information happens when the electrons are recorded by the detection devices. Traditionally, images were recorded using photographic films. Development and scanning of films is very tedious and became

impractical when datasets of cryo-EM projects exceed thousands of micrographs. Charge-coupled device (CCD) cameras were installed to allow automated cryo-EM data acquisition, but the quality of images is significantly worse than that of the film.

The availability of direct electron detectors (DEDs) brings a huge improvement in the quality of the EM images. DEDs are capable of measuring the event of a single electron hitting the device leading to a higher signal to noise ratio in the images. Apart from archiving detection efficiencies higher than that of photographic films, such detectors can capture images much faster than a CCD camera. This is very useful given that particles embedded in vitreous ice start to move when the specimen is hit by electrons. Prolonged exposure times thus leads to blurred images. Instead of a single, second-long exposure, a direct electron detector can record movies at a rate of many frames per second. With these movies, the movements of the particles during exposure are traceable and can be corrected for during the subsequent image processing, leading to increased resolution.

1.4.3 Advances in image processing

To take full advantage of the DEDs, programs were developed to perform movie processing. New algorithms and data collection strategies are emerging, aiming to preserve the precious structural information from noisy images in an optimal way.

The beam-induced movement of particles has long been observed, but was well characterized only after the invention of the DEDs. Realigning individual particles across movie frames to correct their movements has been successfully applied (Brilot *et al.*, 2012). Soon after, it was shown that correction of the beam-induced movement leads to increased resolution of the reconstructed structures (Bai *et al.*, 2013).

Another major improvement in image processing is the implementation of maximum-likelihood algorithms. Most of the cryo-EM samples are heterogeneous in composition and conformation, and different species need to be sorted computationally. Unlike traditionally cross-correlation based methods which rely heavily on the input model(s), maximum-likelihood-based (Sigworth, 1998) algorithms allow performing 3D classification without any prior knowledge of the

source of structural variability. This method was further improved by the integration of a Bayesian approach (Scheres, 2012-1).

In parallel, validation methods were introduced to the field, in response to the concerns about pseudo-high-resolution structures mainly caused by the over-fitting of noisy EM data. One way to prevent over-fitting is to refine two models entirely independently (one for each half of the data), a so-called gold-standard procedure (Henderson *et al.*, 2012). Gold-standard refinement is usually used together with the 0.143 criterion in the Fourier shell correlation (FSC) curve for resolution determination (Rosenthal and Henderson, 2003). Another method to validate a 3D structure is to use pairs of EM micrographs of the same sample recorded at a known tilt angle. If the structure is correct, the calculated angle resulting from the alignment of the tilted and untilted image to the 3D reconstruction should agree with the previously applied tilt angle. In practice, an agreement for 60% of the particles is considered acceptable (Henderson *et al.*, 2011).

Several new software packages were developed with the features mentioned above. Among them, RELION (Scheres, 2012-2) is widely used. The program is based on the maximum-likelihood method, and incorporates the gold-standard refinement procedure. Lately, it also includes programs to investigate the local resolution of a cryo-EM density map (Kucukelbir, Sigworth and Tagare, 2014).

Many of these new software packages are designed to work largely without user intervention. Often, they now have a graphic user interface to facilitate the use. However, it is worth mentioning that optimal results are rarely obtained using the default settings of software. Usually, parameters need to be optimized and different software programs tested. This can lead to significant improvements in the resulting 3D reconstructions as demonstrated in this thesis work.

CHAPTER 2: THE "CLOSED" STATE OF CO-TRANSLATIONAL TARGETING

CHAPITRE 2: LA CONFORMATION «FERMÉ» DE L'ADRESSAGE CO-TRADUCTIONNELLES

Résumé

Les réarrangements conformationnels de SRP et FtsY pendant l'adressage co-traductionnelle sont détaillés dans ce chapitre. La conformation «fermé» représente l'étape directement avant le déchargement et le transfert de la cargaison (*i.e*. Le ribosome avec une séquence signal hydrophobe) de la SRP au translocon.

Ici, la structure cryo-EM du ribosome-SRP-FtsY en conformation «fermé» a été résolue à un niveau de 5.7 Å. La structure a révélé que dans la conformation «fermé» les domaines SRP-FtsY GTPase s'écartent de la sortie du tunnel du ribosome, libérant les protéines ribosomales permettant l'interaction translocon-ribosome afin d'accomplir l'adressage co-traductionnelle. La base structurelle du ribosome et la séquence signal de liaison par le SRP M-domaine sont également décrits ici en détail sans précédent.

Les procédures de traitement des données et l'affinement structurel sont décrits en détail. Le jeu de données a été d'abord nettoyé par l'analyse 2D, puis par le tri hiérarchique des particules afin d'identifier la présence et l'absence de molécules SRP dans le complexe ainsi que son hétérogénéité conformationnelle. Les scripts et les commandes utilisés pour chaque étape de la classification et de l'affinement sont présentés, le raisonnement pour effectuer chacune de ces étapes est décrit.

2.1 Conformational rearrangements of SRP and FtsY during co-translational targeting

As briefly introduced in Section 1.1.2 and 1.1.3, the SRP mediates co-translational protein targeting by delivering translating ribosomes into the membrane. The *E. coli* SRP is a complex consisting of 4.5S RNA and a protein named Ffh (for Fifty-four-homologue). Ffh has three functional domains: an N-terminal domain adopting a four-helix bundle, a G domain which harbors GTPase activity, and a methionine-rich M domain that interacts with the 4.5S RNA (Batey *et al.*, 2000; Rosendal *et al.*, 2003) and with the signal sequence (Janda *et al.*, 2010; Hainzl *et al.*, 2011). The N and G domains form a compact structural and functional unit named the NG domain.

The SRP receptor, FtsY, comprises a homologous NG domain. *E. coli* FtsY contains in addition an acidic A-domain that is suggested to be responsible for membrane anchoring of FtsY. The A domain interacts with anionic phospholipids and the translocon (Weiche *et al.*, 2008). SRP and FtsY interact via their homologues NG domains. The crystal structure of the NG domain complex revealed a composite active site containing two nucleotides (Egea *et al.*, 2004; Focia *et al.*, 2004). This explains how the SRP and FtsY GTPases can activate each other without the need of a GTPase-activating protein.

The SRP pathway which is critical for the vitality of the cell is tightly regulated in time and space. SRP targeting involves a series of conformational changes with check-points in between them. Co-translational targeting includes the following steps: 1) the recognition of cargo (the translating ribosome synthetizing a nascent polypeptide with a signal sequence) by the SRP; 2) the delivery of cargo to the target membrane by complex formation between the SRP and FtsY (“early” state); 3) the unloading and transfer of cargo from the SRP to the translocon (“closed” state); and 4) GTP hydrolysis (“activated” state) and disassembly of the SRP-FtsY complex (Figure 2-1) (Zhang *et al.*, 2009). Structures of the first step, the SRP-bound ribosome, were first solved by cryo-EM at medium resolution (Schaffitzel *et al.*, 2006; Halic *et al.*, 2006-2). Subsequently, a structure of the RNC-SRP-FtsY complex in the absence of GTP, the so-called “early” state was revealed (Estrozi *et al.*, 2011). This cryo-EM structure demonstrated that the ribosome pre-positions the SRP to bind

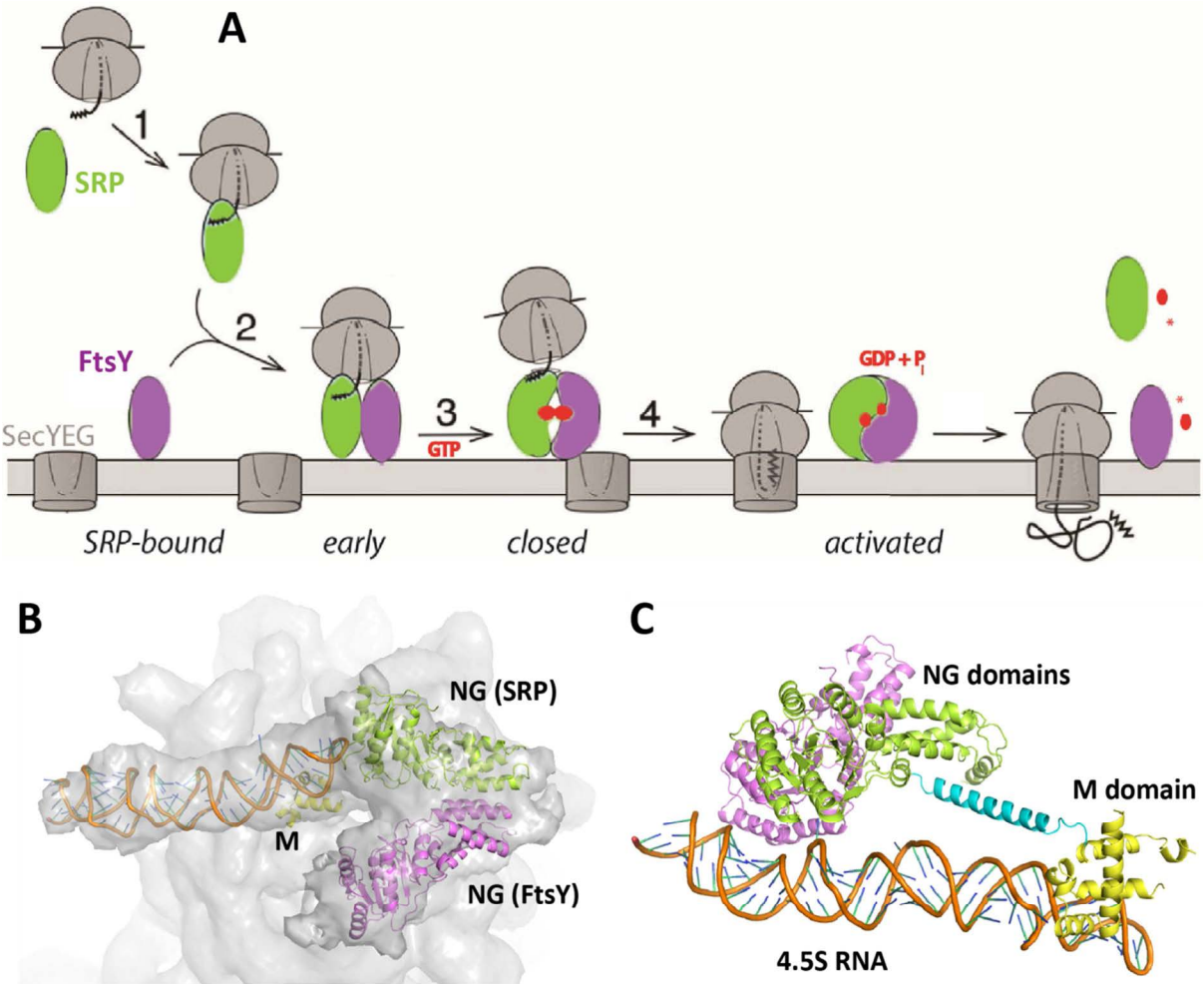


Figure 2-1: Conformational changes of SRP and FtsY during co-translational targeting. (A) Schematic representation of the states adopted by SRP and its receptor during co-translational targeting. See main text for details. Adapted from Zhang *et al.*, 2009. (B) Cryo-EM structure and atomic model of the "early" state complex (Estrozi *et al.*, 2011). (C) Crystal structure of the SRP-FtsY complex with non-hydrolysable GTP analogues (Ataide *et al.*, 2011).

FtsY. FtsY docks onto the ribosome-SRP complex and interacts with the SRP RNA while the NG-domains of FtsY and SRP interact only weakly; the composite active site is not formed in agreement with the fact that this complex can form in the absence of GTP. FtsY docking leads to an overall detachment of the SRP from the ribosome (Figure 2-1). More recently, another structure of the "early" state complex with a signal peptide representing an 'incorrect' cargo was solved (von Loeffelholz *et al.*, 2013). This structure showed a distorted NG-domain arrangement which hampers GTPase activation and favors dissociation of FtsY from the complex, *i.e.*

abortion of targeting. The cryo-EM structure with the incorrect cargo thus revealed mechanisms of quality control during co-translational targeting.

The next question was then the structure of the SRP-FtsY complex during the third step of protein targeting. GTP is required during this step to allow the SRP and FtsY to rearrange into the so-called “closed” state. Crystal structures were reported for the SRP-FtsY complex with non-hydrolysable GTP analogues which were suggested to represent the “closed” and “activated” states of the targeting complex (Ataide *et al.*, 2011; Voigts-Hoffmann *et al.*, 2013). Since co-translational targeting involves the ribosome and a signal sequence it is possible that the cargo can influence the interaction of the SRP and FtsY in the “closed” state. We therefore decided to study the closed state by cryo-EM.

2.2 Publication: Ribosome-SRP-FtsY cotranslational targeting complex in the closed state

Ribosome–SRP–FtsY cotranslational targeting complex in the closed state

Ottilie von Loeffelholz[a,b,1,2], Qiyang Jiang (姜启阳)[a,b,1], Aileen Ariosa[c], Manikandan Karuppasamy[a,b], Karine Huard[a,b], Imre Berger[a,b,d], Shu-ou Shan[c], and Christiane Schaffitzel[a,b,d,3]

[a]European Molecular Biology Laboratory, Grenoble 38042, France; [b]Unit for Virus Host–Cell Interactions, Université Grenoble Alpes, European Molecular Biology Laboratory, CNRS, Grenoble 38042, France; [c]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125; and [d]School of Biochemistry, University of Bristol, Bristol BS8 1TD, United Kingdom

Edited by Joachim Frank, Howard Hughes Medical Institute, Columbia University, New York, NY, and approved February 20, 2015 (received for review December 30, 2014)

The signal recognition particle (SRP)-dependent pathway is essential for correct targeting of proteins to the membrane and subsequent insertion in the membrane or secretion. In *Escherichia coli*, the SRP and its receptor FtsY bind to ribosome–nascent chain complexes with signal sequences and undergo a series of distinct conformational changes, which ensures accurate timing and fidelity of protein targeting. Initial recruitment of the SRP receptor FtsY to the SRP–RNC complex results in GTP-independent binding of the SRP–FtsY GTPases at the SRP RNA tetraloop. In the presence of GTP, a closed state is adopted by the SRP–FtsY complex. The cryo-EM structure of the closed state reveals an ordered SRP RNA and SRP M domain with a signal sequence-bound. Van der Waals interactions between the finger loop and ribosomal protein L24 lead to a constricted signal sequence-binding pocket possibly preventing premature release of the signal sequence. Conserved M-domain residues contact ribosomal RNA helices 24 and 59. The SRP–FtsY GTPases are detached from the RNA tetraloop and flexible, thus liberating the ribosomal exit site for binding of the translocation machinery.

protein targeting | signal recognition particle | signal sequence | ribosome | single-particle electron cryomicroscopy

The *Escherichia coli* signal recognition particle (SRP) is a complex consisting of the universally conserved protein Ffh and 4.5S RNA, which adopts a hairpin structure (1). Ffh is composed of the N-terminal domain, the G domain that harbors GTPase activity, and the C-terminal methionine-rich M domain that interacts with 4.5S RNA (2, 3) and with the signal sequence (4, 5). The N and G domains form a compact structural and functional unit termed "the NG domain." Targeting of ribosome-nascent chain complexes (RNC) containing a signal sequence depends on the interaction of the RNC–SRP complex with the SRP receptor FtsY, which is membrane associated (6–9). FtsY and Ffh interact via their homologous NG domains and form a composite GTPase active site (10, 11). Crystal structures of the M domain reveal a hydrophobic groove used to capture signal sequences (4, 5, 12).

Protein targeting is driven by highly regulated conformational rearrangements of SRP and FtsY as well as GTP hydrolysis. SRP recognizes and tightly binds to RNCs displaying a signal sequence (cargo). Next, RNC-bound SRP efficiently recruits FtsY to form a nucleotide-independent, transient early state that rearranges to a GTP-stabilized closed state (13). Ultimately, in the activated state, handover of the RNC to the Sec translocon takes place, followed by GTP hydrolysis and disassembly of the SRP–FtsY complex (14–16). These distinct conformational transitions are regulated by the ribosome and translocon in the membrane, leading to a switch from cargo recognition by SRP to cargo release (17, 18).

Cryo-EM structures of bacterial SRP-bound RNCs revealed a tight cargo-recognition complex (19, 20). In the SRP–FtsY early complex an overall detachment of SRP from the ribosome was observed (21). In this state, the G domain of FtsY contacts the conserved SRP RNA tetraloop, and Ffh and FtsY interact via their N domains (21) forming a pseudosymmetric V-shaped complex positioned above the ribosomal tunnel exit. The active sites of the GTPase domains are apart from each other, explaining why the early state is inactive in GTP hydrolysis (13, 21, 22).

GTP is required for SRP and FtsY to rearrange into the closed state. FRET experiments indicate that, in this state, the Ffh–FtsY NG domains adopt a conformation that resembles the intimate heterodimeric architecture observed in crystal structures (10, 11, 13). The complete SRP was crystallized in complex with the FtsY NG domain in the closed/activated state showing the NG domains docked at the distal end of the RNA hairpin (23, 24). Single-molecule total internal reflection fluorescence microscopy directly demonstrated that the Ffh–FtsY NG domains need to relocate from the tetraloop to the RNA distal end to become activated for GTP hydrolysis and to progress further in the targeting reaction (24).

Although the early, closed, and activated SRP–FtsY targeting complexes have been well-characterized biochemically, the generation of distinct, conformationally homogenous closed and activated ribosome–SRP–FtsY complexes for structural studies proved to be exceedingly difficult, because the ribosome stabilizes the early state (13). We overcame this challenge by developing a robust complex preparation strategy, and describe here the

Significance

The universally conserved signal recognition particle (SRP) and its receptor (FtsY) deliver ∼30% of the proteome to the proper cellular membrane. To ensure proper timing and fidelity of targeting, SRP and FtsY adopt multiple conformations in a GTP-dependent manner. We solved the cryo-EM structure of the SRP–FtsY complex with a GTP analogue in the presence of a ribosome translating a signal sequence (the closed state) at 5.7 Å resolution. We describe the structural basis of ribosome and signal sequence binding by the SRP M domain. We demonstrate that in the closed state the SRP–FtsY GTPase domains are moving away from the ribosomal tunnel exit, allowing for translocon–ribosome interactions to accomplish cotranslational targeting.

Author contributions: O.v.L., S.-o.S., and C.S. designed research; O.v.L., Q.J., A.A., M.K., and K.H. performed research; O.v.L., Q.J., A.A., M.K., and C.S. analyzed data; and O.v.L., Q.J., I.B., S.-o.S., and C.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The EM maps and atomic models have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 5AKA), and EMDataBank, www.emdatabank.org (accession no. EMD-2917).

[1]O.v.L. and Q.J. contributed equally to this work.

[2]Present address: Institute of Structural and Molecular Biology, Birkbeck College, London WC1E 7HX, United Kingdom.

[3]To whom correspondence should be addressed. Email: schaffitzel@embl.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1424453112/-/DCSupplemental.

cryo-EM structure of the closed state of the RNC–SRP–FtsY complex at a resolution of 5.7 Å.

Results

Closed Targeting Complex Formation. For the structural study of the closed targeting state we could not use the RNC^{FtsQ} complex displaying the N-terminal 108 amino acids of FtsQ with its transmembrane helix (19–21). RNC^{FtsQ}–SRP–FtsY complexes adopt a very stable early state, and rearrangement from the early to the closed state is slow and unfavorable (13). To obtain a more stable closed targeting complex, we used ribosomes displaying a nascent chain comprising the N-terminal 50 aa of leader peptidase (Lep50). RNC^{Lep50}–SRP–FtsY complexes were shown to exhibit significantly faster kinetics and equilibrium for rearrangement into the closed complex (25, 26).

To further enhance the formation of the closed targeting state, we used a truncated version of FtsY ($FtsY^{219}$). This variant lacks the N-terminal A domain and the first helix of the N domain, which was reported to inhibit FtsY–Ffh complex formation and GTPase activation (27); based on this, a single-chain construct comprising the $FtsY^{219}$ construct fused via a 31-aa glycine- and serine-rich linker to full-length Ffh ($scSRP^{219}$) was generated (Fig. 1*A*). A similar FtsY–SRP fusion composed of the full-length proteins was reported to be fully functional in vitro (21) and in vivo (17). Using ribosome-binding assays, we confirmed that $scSRP^{219}$ bound equally well to RNC^{Lep50} and RNC^{FtsQ} (Fig. 1*B*). We observe a twofold increase in GTPase activity of the $scSRP^{219}$ construct compared with the full-length FtsY–Ffh fusion protein with SRP RNA (Fig. 1*C*). A 10-fold increased GTPase activity was previously reported for the unlinked $FtsY^{219}$ and Ffh proteins in the absence of SRP RNA (27); the smaller stimulation we observed was likely due to the presence of SRP RNA in our experiment that also activates GTP hydrolysis for the wild-type Ffh–FtsY complex.

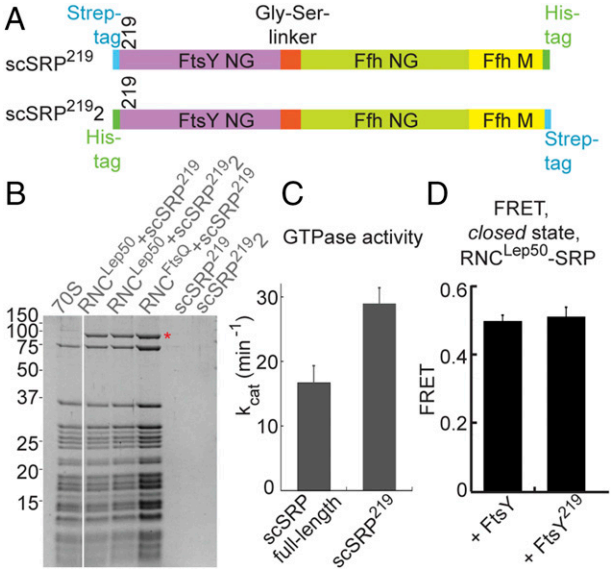


Fig. 1. Preparation of closed targeting complexes. (*A*) Schematic representation of single-chain SRP–$FtsY^{219}$ constructs ($scSRP^{219}$ and $scSRP^{219}2$). (*B*) Coomassie-stained SDS/PAGE section showing the binding of the $scSRP^{219}$ (added in fivefold molar excess) to RNC^{Lep50} and RNC^{FtsQ} analyzed by cosedimentation experiments. In the absence of ribosomes, scSRP constructs were not observed in the pellet fraction. The protein band corresponding to scSRP is marked with a red star. (*C*) GTPase assays comparing the observed GTP hydrolysis rate constant of the linked wild-type proteins (scSRP full-length) and $scSRP^{219}$. (*D*) FRET values between the Ffh and FtsY G domains in the RNC^{Lep50}–SRP–FtsY and RNC^{Lep50}–SRP–$FtsY^{219}$ complexes in the presence of GMPPNP (calculated from the data in *SI Appendix*, Fig. S1*A*).

Formation of the closed complex by SRP and $FtsY^{219}$ was tested using FRET experiments in the presence of RNC^{Lep50} (13). To this end, residue 345 of FtsY and residue 153 of Ffh (both located in the G domain) were labeled with BODIPY-FL and coumarin, respectively (13). The FRET signal of the RNC^{Lep50}–FtsY–Ffh complex in the presence of a nonhydrolysable nucleotide analog (GMPPNP) was comparable for $FtsY^{219}$ and full-length FtsY (both ∼0.5) (Fig. 1*D* and *SI Appendix*, Fig. S1*A*). For comparison, the FRET signal was 0.26 and 0.24 for the RNC^{Lep50}–FtsY–Ffh complex formed by full-length FtsY and $FtsY^{219}$, respectively, for the early complex formed in the absence of nucleotide (*SI Appendix*, Fig. S1*B*) (13, 21); this indicates that $FtsY^{219}$ and wild-type FtsY undergo similar conformational changes during targeting. Importantly, the higher FRET value in the presence of GMPPNP indicates that the G domains are closer to each other, confirming that the RNC^{Lep50}–Ffh–FtsY complex rearranges from the early to the closed state.

Cryo-EM of the Closed Complex. RNC^{Lep50} was incubated with a 10-fold excess of $scSRP^{219}$ in the presence of 1 mM nonhydrolysable GTP analog (GMPPCP) and used for cryo-EM. After computational sorting of the initial data set, which contained 132,390 particles (*SI Appendix*, Fig. S2) and subsequent refinement, the structure of the RNC^{Lep50}–SRP–FtsY closed complex was determined from 32,170 particles at 5.7 Å resolution ["gold-standard" method, i.e., refining independent halves of the data separately and the Fourier shell correlation 0.143 criterion] (28) (*SI Appendix*, Fig. S3*A*). The local resolution map indicates a resolution up to 4 Å in the rigid parts of the large ribosomal subunit (*SI Appendix*, Fig. S3*B*).

In the cryo-EM structure, we observe a distinct density at the exit of the ribosomal tunnel of the large subunit accounting for the single-chain SRP–FtsY complex (Fig. 2*A*). The ribosomal complex contains a P-site tRNA and nascent chain density reaching into the ribosomal tunnel (Fig. 1*B*). However, we could not detect a continuous density threading through the tunnel. The density at the exit of the ribosomal tunnel accounts for the SRP RNA and the Ffh M domain with the bound Lep signal sequence (Fig. 2*C*). The reconstruction of $scSRP^{219}$ close to the exit of the ribosomal tunnel has a local resolution of 6–7 Å (*SI Appendix*, Fig. S3*B*), revealing secondary structure elements. Further away from the exit of the ribosomal tunnel, the local resolution for the distal end of the 4.5S RNA decreases (10–12 Å), indicating flexibility.

Flexibility of the Ffh and FtsY NG Domains. We do not observe any density accounting for the Ffh and FtsY NG domains at the distal end of the 4.5S RNA or the RNA tetraloop (Fig. 2*A*). Our results suggest that during the rearrangement from the early to the closed state, the NG domains detach from the RNA tetraloop and become flexible with respect to the remaining complex, adopting multiple conformations. We confirmed the flexible nature of the NG domains by conjugating gold particles (1.8 nm diameter) to a His-tag fused to the N terminus of $FtsY^{219}$ in the single-chain construct ($scSRP^{219}2$ construct in Fig. 1*A*). Complexes RNC^{Lep50} with gold-labeled $scSRP^{219}2$ were analyzed by cryo-EM, followed by supervised classification (29) to remove vacant ribosomes and by multivariate statistical analysis and classification (30) (Fig. 2*D* and *SI Appendix*, Fig. S4). The nanogold particles could be clearly detected as black spots in the single images with scSRP bound to the ribosome, but were averaged out in the class averages (*SI Appendix*, Fig. S4 *A* and *B*).

The gold particles were most often found next to the tunnel exit of the large ribosomal subunit (Fig. 2*D* and *SI Appendix*, Fig. S4*C*) confirming the presence of the NG domains. We did not observe any accumulation of gold particles next to the RNA tetraloop or toward the distal end of SRP RNA (Fig. 2*D* and *SI Appendix*, Fig. S4*C*), confirming Ffh–FtsY NG domain flexibility. Our finding is consistent with the observation that targeting complexes formed

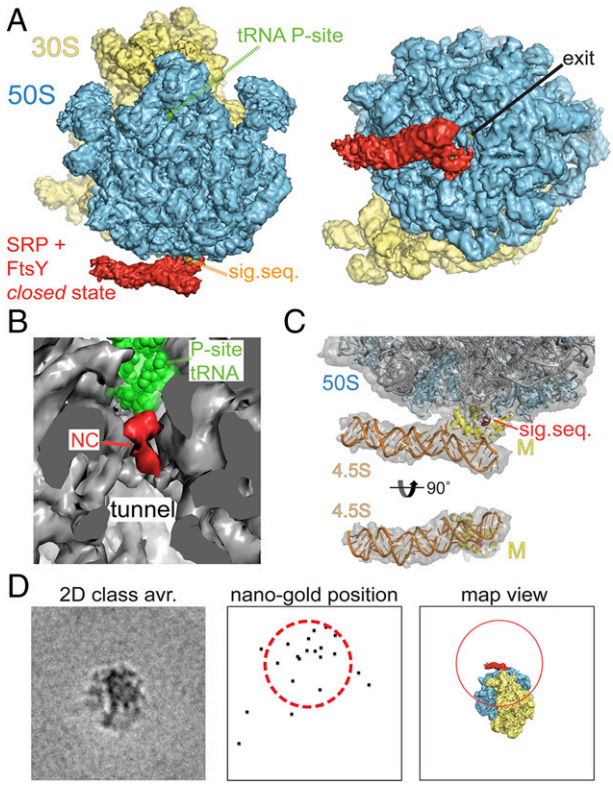


Fig. 2. Cryo-EM structure of RNCLep50–scSRP^{219} complex in the closed state. (*A*) Surface representation of the cryo-EM structure of RNC^{Lep50}–scSRP^{219} complex in the presence of GMPPCP. 30S, yellow; 50S, blue; scSRP, red; signal sequence, orange; P-site t-RNA, green. (*B*) Close-up of the experimental density at the beginning of the ribosomal tunnel. Density corresponding to 50S is shown in gray (contoured at 1σ), density corresponding to nascent chain in red (contoured at 0.5σ). The P-site t-RNA is depicted in space filling view (green). (*C*) Quasi-atomic model of the scSRP^{219} complex in a side view (*Upper*, close-up from *A*) and bottom view (*Lower*). The experimental density is shown in light gray. Ribosomal RNA is colored in gray, ribosomal proteins in cyan, SRP RNA in orange, the Ffh M domain in yellow, and the signal sequence in red. (*D*) Nanogold labeling of the FtsY–Ffh NG domains. (*Left*) 2D class average of gold-labeled RNC^{Lep50}–scSRP^{219}2 particles; (*Center*) plot showing the position of the nanogold in the individual images; (*Right*) surface view of the cryo-EM structure in the same orientation as the 2D class average. The red circles indicate the theoretical volume in which the nanogold could be found in the closed complex assuming flexible NG domains (*SI Appendix*, Fig. S4*C*), taking into account the length of the linker between the Ffh M domain and the NG domain, the dimensions of the NG domain complex (23), and the linker and size of the nanogold particle.

with mutant FtsY(A335W), which blocks the rearrangement from the closed to the activated state, are incapable of docking to the distal site of 4.5S RNA in single-molecule FRET experiments (31). Instead, the closed complex appears to adopt intermediate states where the NG domains are found in positions ranging from the RNA tetraloop to the RNA distal end (Fig. 2*D*) (31).

Ribosomal Contacts Involve Conserved M domain Residues. We placed the crystal structures of the *E. coli* large and small ribosomal subunits (32) and of the GMPPCP-bound SRP–FtsY complex (23) into the EM density to interpret the closed targeting complex at the atomic level (Fig. 2*C*). The well-resolved density of the SRP M domain allowed us to adjust the position of individual α-helices and to extend the atomic model by 4 aa to residue 436 at the C terminus of the M domain (Fig. 3*A*). In addition, we identified density corresponding to a α-helix of ~8 aa of the leader peptidase signal sequence (Fig. 3*A*).

The SRP engages the ribosome by forming three contacts in the closed targeting complex. All three contacts involve the M domain. One contact likely involves a hydrogen bond of Ffh glutamine 411 to the sugar phosphate backbone of rRNA helix 24 in the closed conformation (Fig. 3*B*). Gln411 is conserved in bacteria (*SI Appendix*, Fig. S5*A*), whereas in *Sulfolobus solfataricus* (archaea) a glutamic acid residue is found at this position (in which case, the contact could then possibly involve chelation of a magnesium ion). A second contact is formed between the C terminus of the M domain and rRNA helix 59 (Fig. 3*C*); it involves electrostatic interactions between lysines 431 and 432 and the sugar phosphate backbone of rRNA h59. Sequence comparison shows that all bacterial M domains contain positively charged amino acid residues in this region (*SI Appendix*, Fig. S5*A*). A third contact is observed when the map is contoured at a slightly lower level (0.8σ; Fig. 3*D*), which may indicate that this interaction is not overly strong. The contact involves van der Waals interactions between conserved hydrophobic residues in the finger loop (isoleucine 360 and possibly valine 364) and ribosomal protein L24 (alanine 51 and leucine 52; Fig. 3*D* and *SI Appendix*, Fig. S5 *A* and *B*). Interestingly, a similar contact could also be found in a previous

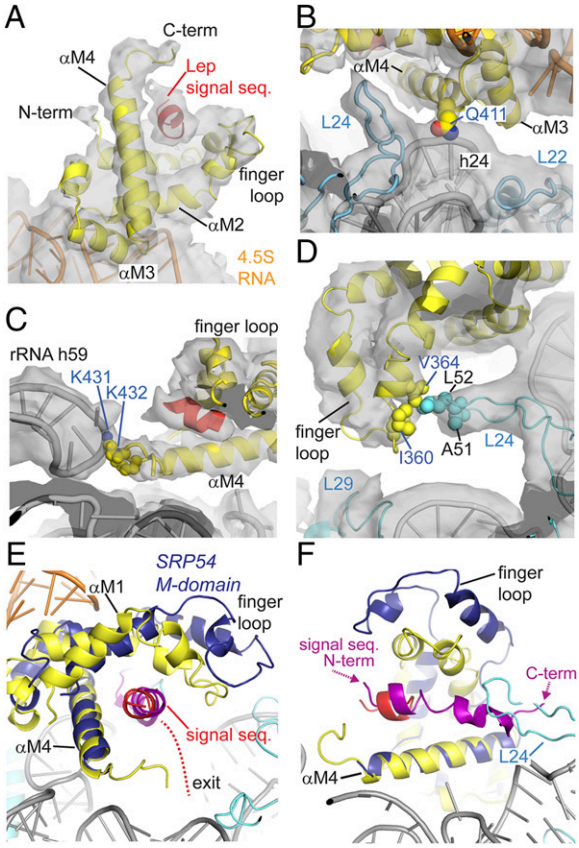


Fig. 3. Atomic model of the M domain, the bound signal sequence, and ribosomal contacts. (*A*) Model of the Ffh M domain and the Lep signal sequence bound in the hydrophobic groove and flanked by the finger loop. Color-coding as in Fig. 2C. (*B*) Interaction between rRNA helix 24 and the M-domain helix 4. (*C*) Two lysines (431 and 432) in the C-terminal part of the M domain bind the sugar-phosphate backbone of rRNA helix 59. (*D*) Contact formed between the M-domain finger loop and ribosomal protein L24. (*E*) Model of the M domain (yellow) with the bound Lep signal sequence (red) is superimposed on the crystal structure of the SRP54–signal sequence complex from *S. solfataricus* (adapted from ref. 5). SRP54 M domain, blue; yeast dipeptidyl aminopeptidase B-signal sequence, purple. The path of the nascent polypeptide chain exiting from the large ribosomal subunit is indicated by red dots. (*F*) Same as *E*, shown in a view from the ribosomal tunnel.

reconstruction of the *E. coli* RNC–SRP complex (33), albeit at lower resolution (~10 Å), which had ruled out interpretation at the amino acid residue level.

It is important to note that the interaction between the SRP and the ribosome is weaker in the closed targeting state compared with the early state. A K_d of 0.03 nM was determined for the SRP–FtsY binding to the RNC^{3A7L} in the early state, and a K_d of 0.9 nM in the closed state (34). Therefore, in different targeting states, it appears that additional amino acids could participate in mediating contacts to the ribosome.

The Finger Loop Stabilizes Signal Sequence Binding. Another interesting observation is the density connecting helix 1 (αM1) and helix 2 (αM2) of the M domain representing the finger loop (Fig. 3 *A* and *D*). Among numerous SRP crystal structures, only three structures visualized an ordered finger loop (3, 5, 12). In each of these structures, the finger loop adopts a different conformation. None of the crystal structures fitted to the EM density corresponding to the finger loop in the closed complex. We therefore modeled residues N343–D369 comprising the finger loop de novo using HHpred and flexibly fitted the resulting helix–loop–helix structure into the EM density (Fig. 3 *A* and *D*), providing the first insight into how the *E. coli* SRP finger loop is arranged in a ribosomal complex.

The Lep signal sequence is bound at virtually the same position as in the SRP54–signal peptide crystal structure from *S. solfataricus* (5) (Fig. 3 *E* and *F*). In our structure, which includes the RNC, we find that the finger loop is located closer to the signal sequence, giving rise to a sterically more restricted binding pocket. A 20-aa signal sequence was present in the crystal structure (5). We note that, in the presence of the ribosome, the path of the nascent chain containing the signal sequence is incompatible with the geometry observed in the crystal structure (5), because the C-terminal part of the signal peptide would clash with ribosomal protein L24 (Fig. 3*F*). Thus, it appears that in the closed targeting complex, only a smaller part of the signal sequence is bound by the M domain, and the nascent chain follows a different path when exiting from the ribosomal tunnel (Fig. 3 *E* and *F*).

Discussion

Here, we report the 5.7-Å cryo-EM structure of the RNC–SRP–FtsY complex in the presence of GMPPCP, representing the closed state. The quasi-atomic model of the SRP comprises the Ffh M domain and residues 14–87 of 4.5S RNA (Fig. 2*C*). Thus, the reported binding site of the NG domains at the distal end of the RNA (G83–C87) (23, 24) is included in our model. However, we could not detect any density corresponding to the Ffh and FtsY NG domains either at the distal end of the SRP RNA or close to the RNA tetraloop in the cryo-EM reconstruction (Fig. 2*A*). Instead, single-molecule FRET experiments (31) and our gold-labeling experiments of the FtsY NG domain indicate that, in the closed complex, the NG domains are flexible (Fig. 2*D* and *SI Appendix*, Fig. S4*C*). In contrast, in the early state the NG domains are bound to the RNA tetraloop (21, 35) (Fig. 4). In the activated state, the NG domains bind at the RNA distal end ~100 Å away from the tetraloop (23, 24); taken together, this suggests that the linker between the Ffh M domain and the NG domain must be very flexible to allow such dramatic conformational rearrangements (Fig. 4).

Whereas the RNC stabilizes the early state (13), the transition to the activated state is triggered by binding to the SecYEG translocon (18, 31). Indeed, L23 is accessible in the closed state for translocon binding (Fig. 4 and *SI Appendix*, Fig. S6), in contrast to the SRP-bound and early targeting complexes (19, 21, 33) where L23 is occluded by the Ffh NG domain. It is noteworthy in this context that a similar mobilization of the SRP–SRP receptor NG domain complex has been described for the eukaryotic cotranslational targeting complex (20), suggesting that the underlying molecular mechanisms are evolutionary conserved.

Our cryo-EM structure reveals a fully ordered M domain with a bound signal sequence. The signal sequence rests in a hydrophobic groove lined by methionine residues (methionine bristles) and flanked by the finger loop. The finger loop has been suggested to form a "lid" that closes over the signal sequence located in the groove (3–5, 12). The crystal structure of the *S. solfataricus* M domain in complex with a signal peptide (5) fits well into our M-domain density (Fig. 3 *E* and *F*), with exception of the finger loop. Eight amino acids of the N-terminal part of the signal sequence in the crystal structure (5) fit in our density, which corresponds to a signal sequence of minimal length. The remainder of the signal sequence is not detected, despite the fact that the Lep signal sequence comprises 16 hydrophobic amino acids. These eight amino acids may be stably bound in the core of the signal sequence binding groove of the M domain during the complete SRP targeting process. Possibly, the signal sequence is bound more tightly in earlier states of targeting and partially released during the closed state.

In the early state cryo-EM structure, we found that the signal sequence-binding part of the M domain was flexible as evidenced by loss of this density (21). It is not straightforward to draw conclusions about the M-domain conformation from these structures, because RNCs with a different nascent chain lengths and different signal sequences have been used in each study. Different signal sequence composition and nascent chain length has been reported to affect SRP binding to the ribosome (36). Therefore, it could be that the relatively short Lep50 nascent chain used in this study induces a less-flexible M domain, which thus can be observed in the EM density.

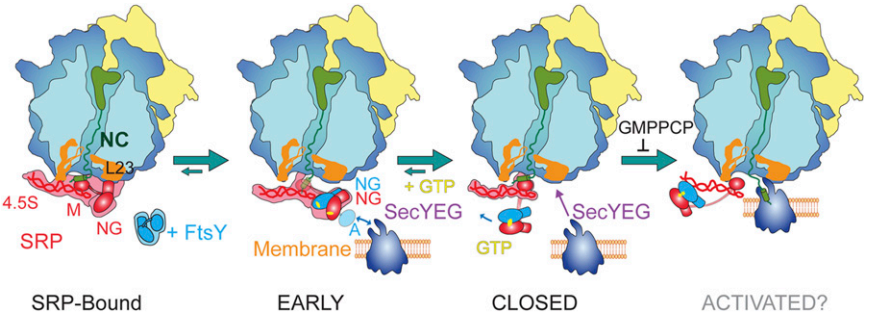


Fig. 4. Model of cotranslational targeting in *E. coli*. In the RNC–SRP complex, SRP is prepositioned to bind FtsY. FtsY binding leads to an early complex inactive in GTP hydrolysis. Rearrangement of the NG domains in the closed complex requires GTP and leads to detachment of the NG domains from the RNA tetraloop. SecYEG binding to L23 is suggested to lead to docking of the NG domains to the distal end of the RNA in the activated state and RNC handover to SecYEG. GTP hydrolysis results in SRP–FtsY disassembly.

We found three connections between the M domain and the large ribosomal subunit (Fig. 3 *B–D*). In our structure, none of the interactions is very strong, and this is in good agreement with the previously reported reduced affinity of the SRP–FtsY closed complex for RNCs (34). The Ffh M-domain helix 4 (αM4) seems to play a major role in ribosome binding of the closed state forming two contacts. A third contact involves the M-domain finger loop and L24 (Fig. 3*D*). Deletion of finger-loop residues led to a significant loss of SRP RNA-stimulated complex assembly by SRP and FtsY (4), indicating that the finger loop has a role beyond signal sequence binding. Intriguingly, the finger loop has been shown to have crucial functions as a sensor for signal sequence binding and facilitator of subsequent conformational changes in SRP and FtsY, including RNC unloading from the SRP (37). The finger loop contacts the signal sequence and ribosomal protein L24 in our structure (Fig. 3*D*). We speculate that L24 stabilizes signal sequence binding by inducing a finger-loop conformation that prevents premature release of the signal sequence and SRP dissociation.

In conclusion, the closed-complex cryo-EM structure reported here reveals the complex interactions of the SRP M domain with the signal sequence and the RNC. Our structure suggests that signal sequence binding to the M domain is stabilized by the ribosome, which induces the formation of a more constricted binding pocket by means of repositioning the finger loop. Furthermore, our results show that the NG domains are detached from the RNA tetraloop in the closed state but not yet bound to the distal end where they are fully activated for GTP hydrolysis. Additional high-resolution structures of cotranslational targeting complexes will be required to uncover in the future the intricate molecular mechanisms of communication among SRP GTPases, SRP RNA, signal sequences, and ribosomal complexes.

Material and Methods

Cloning, Expression, and Purification of scSRP^{219}, scSRP^{219}2, FtsY^{219}, and RNC^{Lep50}. Plasmids encoding the scSRP^{219} (pET24aFtsYΔNlinkFfh and pUC19Ffs) (19) were transformed in *E. coli* Bl21 Star cells. scSRP^{219} was expressed and purified by affinity chromatography using the hexahistidine- and the StrepII-tags, followed by anion exchange chromatography (MonoQ) as described (21). FtsY^{219}Q345C was expressed in *E. coli* Bl21 Star cells and purified by NiNTA, MonoQ, and size-exclusion chromatography (S200) in buffer A [50 mM Hepes–KOH, 100 mM KOAc, 8 mM $Mg(OAc)_2$ (pH 7.5), 500 μg/mL chloramphenicol]. RNCs were prepared by in vitro transcription and translation from pUC19StrepLep50 and purified by sucrose gradient centrifugation and affinity chromatography. See *SI Appendix, Methods* for cloning details.

Preparation and Characterization of RNC^{Lep50}nscSRP^{219}–GMPPCP Complexes. scSRP^{219} constructs were added in a 10- to 15-fold molar excess to 150–200 nM RNC^{Lep50} and incubated for at least 15 min on ice followed by addition of 1 mM GMPPCP (Sigma) and further incubation for 1 h at room temperature. The binding of scSRP^{219} to RNC^{Lep50} under these conditions was confirmed by cosedimentation experiments through a 0.5-M sucrose cushion in buffer A (Fig. 1*A*). The complexes were further characterized by GTPase assays, fluorescence measurements, and gold-labeling followed by cryo-EM (*SI Appendix, Methods*).

Electron Microscopy and Image Processing. Quantifoil grids (300 mesh; Agar Scientific) type R1.2/1.3 coated with a thin continuous carbon layer were glow-discharged for 30 s, and 2.5-μL sample (200 nM RNCs) was applied on the carbon side. The grids were blotted for 2 s inside the humidity chamber using a Vitrobot Mark IV (FEI) and plunge-frozen immediately in liquid ethane. Data were collected with FEI EPU software on a Titan Krios at 300 kV using −0.6 to −3 μm defocus in 100-nm steps and an electron dose of 24 eÅ^{-2}. Micrographs were recorded on a direct electron detector (Falcon II; FEI) with a calibrated pixel size of 1.08 Å per pixel.

The contrast transfer function was determined and corrected with bctf (Bsoft package) (38). A total of 132,390 particles were selected semiautomatically from 2,840 micrographs using e2boxer (EMAN2) (39). For initial alignments, the particles were resampled to 5 Å per pixel in Fourier space using Xmipp 3.0 (40), and were aligned against a 40-Å low-pass-filtered vacant ribosome (EMDB ID no. 1045) (41), yielding an initial reconstruction with relatively weak density of scSRP^{219} (*SI Appendix*, Fig. S2). The resulting map, together with another copy in which the scSRP^{219} density was deleted using UCSF Chimera (42), was used for supervised classification to sort scSRP^{219}-containing particles from vacant ribosomes. The remaining 67,236 particles were then reextracted with 2.5 Å per pixel. The average map from these particles was low-pass filtered to 40 Å and used to perform an unsupervised classification with four classes in RELION (43). We applied a spherical mask around scSRP^{219} and the 50S ribosomal subunit to guide the classification focusing on the different SRP conformations. The subset consisting of 32,170 particles demonstrates a well-defined M domain and 4.5S RNA. A final reconstruction for this subset was obtained using particles of 2 Å per pixel with a spherical mask applied around scSRP^{219} and 50S. We used autorefine in RELION, which is refining a separate model for two independent halves of the data and avoiding overfitting (43). The final map was b-factor sharpened using an automatically calculated b-factor value of −236.9.

Flexible Fitting and Generation of the Quasi-Atomic Model. A quasi-atomic model was generated by initial rigid body-fitting of crystal structures of the SRP (23), the 50S and the 30S ribosomal subunit (32), and the P-site tRNA into the EM density using Chimera. The signal sequence was modeled using a polyalanine sequence adopting a α-helix. The finger loop has been modeled in HHpred (44) using the sequence from N343 to D369 of the *E. coli* SRP. The quasi-atomic model of the Ffh M domain and the 4.5S RNA as well as rRNA Helix 59 along with other deviating parts of the ribosome were adapted to the density and energy-minimized using Coot (45), followed by energy minimization in CNS version 1.0 (46).

ACKNOWLEDGMENTS. We thank Sejeong Lee and Wolfgang Wintermeyer (Max Planck Institute for Biophysical Chemistry, Göttingen, Germany) for discussions and information about Lep50; Wim Hagen for data collection; the protein expression facility at EMBL Heidelberg and the Partnership for Structural Biology in Grenoble for support; and all members of C.S.'s group for discussion and advice with image processing. The Polara microscope is part of the Institut de Biologie Structurale Structural Biology and Dynamics GIS-IBISA–labeled platform. Support was provided by ERC Starting Grant Project 281331 (to C.S.).

1. Keenan RJ, Freymann DM, Stroud RM, Walter P (2001) The signal recognition particle. *Annu Rev Biochem* 70:755–775.
2. Batey RT, Rambo RP, Lucast L, Rha B, Doudna JA (2000) Crystal structure of the ribonucleoprotein core of the signal recognition particle. *Science* 287(5456):1232–1239.
3. Rosendal KR, Wild K, Montoya G, Sinning I (2003) Crystal structure of the complete core of archaeal signal recognition particle and implications for interdomain communication. *Proc Natl Acad Sci USA* 100(25):14701–14706.
4. Hainzl T, Huang S, Meriläinen G, Brännström K, Sauer-Eriksson AE (2011) Structural basis of signal-sequence recognition by the signal recognition particle. *Nat Struct Mol Biol* 18(3):389–391.
5. Janda CY, et al. (2010) Recognition of a signal peptide by the signal recognition particle. *Nature* 465(7297):507–510.
6. Angelini S, Boy D, Schiltz E, Koch HG (2006) Membrane binding of the bacterial signal recognition particle receptor involves two distinct binding sites. *J Cell Biol* 174(5):715–724.
7. Lam VQ, Akopian D, Rome M, Henningsen D, Shan SO (2010) Lipid activation of the signal recognition particle receptor provides spatial coordination of protein targeting. *J Cell Biol* 190(4):623–635.
8. Parlitz R, et al. (2007) *Escherichia coli* signal recognition particle receptor FtsY contains an essential and autonomous membrane-binding amphipathic helix. *J Biol Chem* 282(44):32176–32184.
9. Weiche B, et al. (2008) A cleavable N-terminal membrane anchor is involved in membrane binding of the *Escherichia coli* SRP receptor. *J Mol Biol* 377(3):761–773.
10. Egea PF, et al. (2004) Substrate twinning activates the signal recognition particle and its receptor. *Nature* 427(6971):215–221.
11. Focia PJ, Shepotinovskaya IV, Seidler JA, Freymann DM (2004) Heterodimeric GTPase core of the SRP targeting complex. *Science* 303(5656):373–377.
12. Keenan RJ, Freymann DM, Walter P, Stroud RM (1998) Crystal structure of the signal sequence binding subunit of the signal recognition particle. *Cell* 94(2):181–191.
13. Zhang X, Schaffitzel C, Ban N, Shan SO (2009) Multiple conformational switches in a GTPase complex control co-translational protein targeting. *Proc Natl Acad Sci USA* 106(6):1754–1759.
14. Connolly T, Rapiejko PJ, Gilmore R (1991) Requirement of GTP hydrolysis for dissociation of the signal recognition particle from its receptor. *Science* 252(5009):1171–1173.
15. Jiang Y, Cheng Z, Mandon EC, Gilmore R (2008) An interaction between the SRP receptor and the translocon is critical during cotranslational protein translocation. *J Cell Biol* 180(6):1149–1161.
16. Peluso P, Shan SO, Nock S, Herschlag D, Walter P (2001) Role of SRP RNA in the GTPase cycles of Ffh and FtsY. *Biochemistry* 40(50):15224–15233.
17. Braig D, et al. (2011) Signal sequence-independent SRP-SR complex formation at the membrane suggests an alternative targeting pathway within the SRP cycle. *Mol Biol Cell* 22(13):2309–2323.

18. Akopian D, Dalal K, Shen K, Duong F, Shan SO (2013) SecYEG activates GTPases to drive the completion of cotranslational protein targeting. *J Cell Biol* 200(4):397–405.
19. Schaffitzel C, et al. (2006) Structure of the *E. coli* signal recognition particle bound to a translating ribosome. *Nature* 444(7118):503–506.
20. Halic M, et al. (2006) Signal recognition particle receptor exposes the ribosomal translocon binding site. *Science* 312(5774):745–747.
21. Estrozi LF, Boehringer D, Shan SO, Ban N, Schaffitzel C (2011) Cryo-EM structure of the *E. coli* translating ribosome in complex with SRP and its receptor. *Nat Struct Mol Biol* 18(1):88–90.
22. von Loeffelholz O, et al. (2013) Structural basis of signal sequence surveillance and selection by the SRP-FtsY complex. *Nat Struct Mol Biol* 20(5):604–610.
23. Ataide SF, et al. (2011) The crystal structure of the signal recognition particle in complex with its receptor. *Science* 331(6019):881–886.
24. Voigts-Hoffmann F, et al. (2013) The structural basis of FtsY recruitment and GTPase activation by SRP RNA. *Mol Cell* 52(5):643–654.
25. Bornemann T, Jöckel J, Rodnina MV, Wintermeyer W (2008) Signal sequence-independent membrane targeting of ribosomes containing short nascent peptides within the exit tunnel. *Nat Struct Mol Biol* 15(5):494–499.
26. Holtkamp W, et al. (2012) Dynamic switch of the signal recognition particle from scanning to targeting. *Nat Struct Mol Biol* 19(12):1332–1337.
27. Neher SB, Bradshaw N, Floor SN, Gross JD, Walter P (2008) SRP RNA controls a conformational switch regulating the SRP-SRP receptor interaction. *Nat Struct Mol Biol* 15(9):916–923.
28. Rosenthal PB, Henderson R (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* 333(4):721–745.
29. Shaikh TR, et al. (2008) SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nat Protoc* 3(12):1941–1974.
30. van Heel M, Harauz G, Orlova EV, Schmidt R, Schatz M (1996) A new generation of the IMAGIC image processing system. *J Struct Biol* 116(1):17–24.
31. Shen K, Arslan S, Akopian D, Ha T, Shan SO (2012) Activated GTPase movement on an RNA scaffold drives co-translational protein targeting. *Nature* 492(7428):271–275.
32. Schuwirth BS, et al. (2005) Structures of the bacterial ribosome at 3.5 A resolution. *Science* 310(5749):827–834.
33. Halic M, et al. (2006) Following the signal sequence from ribosomal tunnel exit to signal recognition particle. *Nature* 444(7118):507–511.
34. Saraogi I, Akopian D, Shan SO (2014) Regulation of cargo recognition, commitment, and unloading drives cotranslational protein targeting. *J Cell Biol* 205(5):693–706.
35. Shen K, Shan SO (2010) Transient tether between the SRP RNA and SRP receptor ensures efficient cargo delivery during cotranslational protein targeting. *Proc Natl Acad Sci USA* 107(17):7698–7703.
36. Noriega TR, et al. (2014) Signal recognition particle-ribosome binding is sensitive to nascent chain length. *J Biol Chem* 289(28):19294–19305.
37. Ariosa AR, et al. (2013) Fingerloop activates cargo delivery and unloading during cotranslational protein targeting. *Mol Biol Cell* 24(2):63–73.
38. Heymann JB, Belnap DM (2007) Bsoft: Image processing and molecular modeling for electron microscopy. *J Struct Biol* 157(1):3–18.
39. Tang G, et al. (2007) EMAN2: An extensible image processing suite for electron microscopy. *J Struct Biol* 157(1):38–46.
40. de la Rosa-Trevín JM, et al. (2013) Xmipp 3.0: An improved software suite for image processing in electron microscopy. *J Struct Biol* 184(2):321–328.
41. Valle M, et al. (2002) Cryo-EM reveals an active role for aminoacyl-tRNA in the accommodation process. *EMBO J* 21(13):3557–3567.
42. Pettersen EF, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612.
43. Scheres SH (2012) RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* 180(3):519–530.
44. Söding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33(Web Server issue): W244–248.
45. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.
46. Brünger AT, et al. (1998) Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54(Pt 5): 905–921.

SUPPLEMENTARY INFORMATION

The SRP-SR co-translational targeting complex in the *closed* state displays flexible NG-domains and a defined M-domain with signal sequence

Ottilie von Loeffelholz[a,b,e,1], Qiyang Jiang[a,b,1], Aileen Ariosa[c], Manikandan Karuppasamy[a,b], Karine Huard[a,b], Imre Berger[a,b,d], Shu-ou Shan[c], and Christiane Schaffitzel[a,b,d,2]

Affiliations:

[a]European Molecular Biology Laboratory, Grenoble Outstation, Grenoble, 38042 France

[b]Unit for Virus Host-Cell Interactions, Univ. Grenoble Alpes-EMBL-CNRS, Grenoble, 38042 France

[c]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125

[d]School of Biochemistry, University of Bristol, Bristol, BS8 1TD, United Kingdom

[e]Current address: Institute of Structural and Molecular Biology, Birkbeck College, London, WC1E 7HX, United Kingdom

[1] These authors contributed equally to this work.

[2]To whom correspondence should be addressed:

Christiane Schaffitzel, EMBL, 71 avenue des Martyrs, CS 90181, 38042 Grenoble Cedex 9, France. Telephone: +33 4 76 20-7505. Email: schaffitzel@embl.fr

SI METHODS

Cloning of scSRP^{219}, scSRP^{219}2 and FtsY^{219}

To generate an expression plasmid for scSRP^{219} the truncated version of FtsY (FtsY^{219}) was amplified by PCR from a plasmid encoding for scSRP (1) with a forward primer containing part of the sequence coding for a Strep-tag II affinity tag (iba lifesciences) and annealing to the FtsY sequence from amino acid 220 on (5'-ccgcagttcgaaaaagccggcagcggcggtaaaaaaatcgacgatgatctg-3') and a reverse primer aligning to the C-terminal end of FtsY with an additional C-terminal *Pst*I restriction site (5'-tctagactgcaggcgtccggctgtatc-3'). The N-terminus of the PCR product was elongated by the remaining StrepII tag sequence via a second PCR another forward primer containing an N-terminal *Nde*I restriction site (5'-tctagacatatggctagctggagccac ccgcagttcgaaaaagccgg-3'). This PCR product digested with *Nde*I and *Pst*I and ligated into the *Nde*I/*Pst*I-digested pET24aFtsYlinkFfh plasmid (1) yielding pET24aFtsYΔNlinkFfh. pET24aFtsYΔNlinkFfh encodes a fusion of residues 220-498 of *E. coli* FtsY and the complete *E. coli* Ffh protein (residues 1-453) linked via a 31 amino acids glycine-serine rich linker.

A second plasmid encodoing scSRP^{219} with an N-teminal hexahistidine-tag and a C-terminal StrepII-tag (scSRP^{219}2) was generated from a PCR product using the pET24aFtsYΔNlinkFfh plasmid by PCR using as a template and (5'-aattaacatatgcatcatcaccatcaccatggtaaaaaa atcgacgatgatc-3' and 5'-attaataagcttagccgctacccgctttctcaaactgcggatggctccatgacgcagatctgcgaccag ggaag-3') as PCR primers. The PCR product was fused into pET24a by *Nde*I / *Hind*III restriction digest of vector and insert, followed by ligation yielding pET24a_scSRP^{219}2.

An N-terminal truncated version of FtsY with a Q345C mutation and an N-terminal penta-His-tag (FtsY^{219}Q345C) was PCR amplified and subcloned from a plasmid encoding the mutant FtsY Q345C (2) into a pET24a vector yielding pET24a_ FtsY^{219}Q345C.

Preparation of RNC^{Lep50} Complexes

A plasmid encoding the first 50 amino acids of the leader peptidase precursor with an N-terminal triple StrepII-tag followed by a TEV cleavage site was obtained from Genscript. This

Lep50-construct was subcloned into pUC19 yielding pUC19StrepLep50. The resulting plasmid was digested with *Hin*dIII, *in vitro* transcribed and translated using membrane-free cell extract (3). Sucrose gradient centrifugation and affinity purification were carried out as described (3) with the exception that all buffers contained 500 μg/ml chloramphenicol to stabilize the RNCs. After the final centrifugation step to concentrate the RNC^{Lep50} complexes, they were dissolved in buffer A (50 mM Hepes-KOH, 100 mM KOAc, 8 mM $Mg(OAc)_2$, pH 7.5, 500 μg/ml chloramphenicol).

GTPase Measurements

GTPase assays were carried out at 25°C in assay buffer [50 mM KHEPES, pH 7.5, 150 mM KOAc, 10 mM $Mg(OAc)_2$, 2 mM dithiothreitol (DTT), 0.5 μg/μL chloramphenicol and 0.01% Nikkol]. Hydrolysis of GTP was followed and analyzed by thin layer chromatography as described (4). GTPase reactions contain 50-100 nM wild-type or mutant scSRP construct, a two-fold excess of 4.5S RNA, and 100 μM GTP doped with γ-^{32}P-GTP (MP Biomedicals, Solon, OH).

Fluorescence Measurements

Fluorescence measurements were carried out at 25 °C on a FluoroLog-3-22 spectrofluorometer (Jobin-Yvon) in assay buffer. Where applicable, the buffer also contained 100-200 μM GMPPNP to induce formation of the closed complex. FRET between donor (DACM) and acceptor (BODIPY-FL)-labeled SRP(C153) and FtsY(C345) was measured using an excitation wavelength of 360 nm. Reactions contained 50–100 nM SRP, saturating concentrations of RNC^{Lep50} (100-200 nM), and saturating FtsY (1600 nM for closed complex, 500 nM for early complex) where applicable.

Nano-gold Labeling of scSRP ^{219}

Purified $scSRP^{219}2$ containing an N-terminal His-tag was incubated with a 5-fold molar excess of 1.8 nm Ni-NTA-Nanogold (Nanoprobes, INC) for 1h at 4C in binding buffer (50 mM Hepes pH 7.5, 100 mM KOAc, 8 mM MgOAc2). Excess of unbound Nanogold was removed

from the sample by purification of the labelled scSRP^{219} via its C-terminal StrepII tag followed by concentration and buffer exchange to buffer A using an Amicon Ultra-4 centrifugal filter unit (MWCO 50 kDA, Millipore). RNC^{Lep50}-scSRP^{219}-GMPPCP complexes were prepared for cryo-EM as described above with a 10-fold excess of scSRP^{219} and 1mM GMPPCP. Cryo-EM data were collected on a Tecnai G2 Polara (FEI) operating at 100 kV and a specimen level magnification of 76,000x with a Gatan 4k x 4k CCD camera in a defocus range between −0.7 µm and −5.7 µm with an initial pixel size of 1.86 Å on the object scale.

The contrast transfer function (CTF) was determined and corrected with bctf (Bsoft package). 19,699 particles were selected semi-automatically from 207 CCD frames using e2boxer (EMAN2). For alignments, the particles were re-sampled to 3.72 Å/pixel, inverted and low-pass-filtered to 12 Å. The particles were aligned using a low pass filtered vacant ribosome as initial reference (EMDB ID: 1045) in SPIDER (5). Classification according to presence of scSRP^{219}2 at the ribosomal tunnel exit in SPIDER resulted in a total number of 6,402 particles representing RNC^{Lep50}-scSRP^{219}2-GMPPCP. This particle pool was further classified into 100 classes by multivariate statistical analysis (MSA) and classification in IMAGIC-5 (6). Classes showing orientations of the RNC^{Lep50}-scSRP^{219}2-GMPPCP complexes where scSRP^{219} is clearly visible were identified by projection matching to a ribosomal map using SPIDER (5) and the corresponding 368 particles were submitted to a second round of MSA and classification into 20 classes using IMAGIC. Nine of these classes were identified to contain clearly visible SRP density using SPIDER. The calculated alignment and classification parameters were used to identify and align the corresponding not-inverted full-sized particles. The position of the nano-gold was marked manually in each aligned particle using boxer (EMAN, (7)) and plotted for the three most populated classes in which scSRP^{219}2 is distinguishable.

SI FIGURES

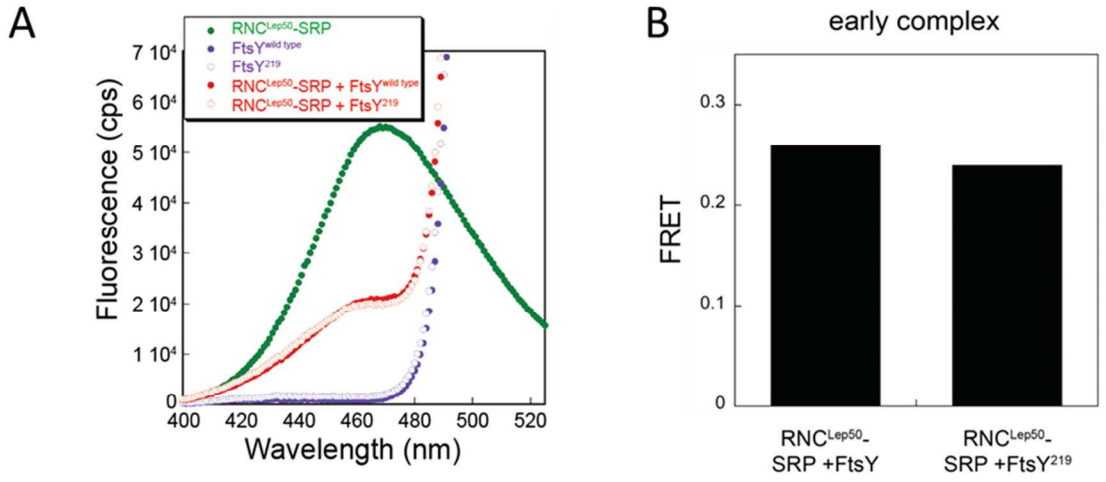


Figure S1: FRET measurements of the *early* complexes.

(A) Fluorescence emission spectra of DACM-labeled RNC^{Lep50} -SRP, BODIPY-FL labeled FtsY or $FtsY^{219}$, and the RNC^{Lep50} -SRP-FtsY (wild type) and RNC^{Lep50} -SRP-$FtsY^{219}$ complexes in the presence of GMPPNP. **(B)** FRET values calculated based on the fluorescence emission spectra of the fluorescence dye-labeled RNC^{Lep50}-SRP-FtsY (wild type) and RNC^{Lep50}-SRP-$FtsY^{219}$ complexes in the absence of nucleotide. Both targeting complexes are adopting the *early* state. The labeling and measurements were done exactly the same way as described for the *closed* complex.

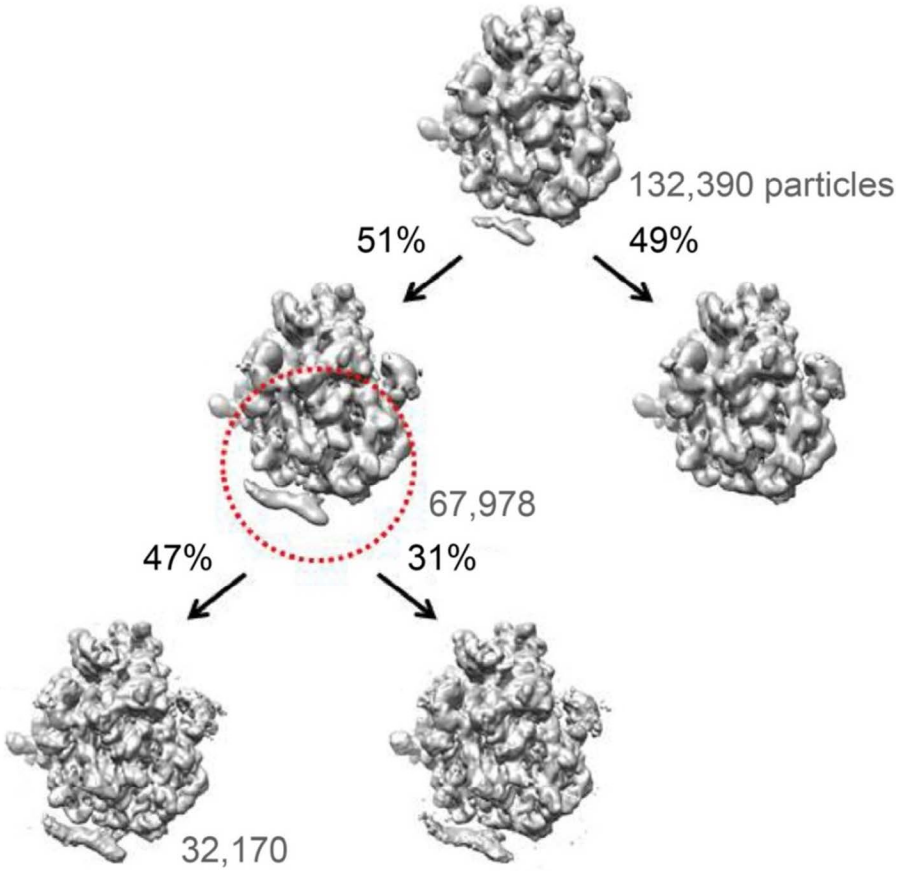


Figure S2: Computational sorting of the RNC^{Lep50}-scSRP^{219}-GMPPCP data set.

An initial map was generated using all 132,390 particles of the data set (first row). The particles were then separated according to whether they align better to the initial map, or to a modified map in which the density corresponding to scSRP^{219} was deleted in Chimera (8). This led to a SRP-containing pool and a vacant pool (second row). Next, a spherical mask centred near the ribosomal tunnel exit with a radius of 130 Å was applied to the SRP-containing pool for further un-supervised classification using Relion (9). The edge of the mask is indicated by red dots. In this round, ~22% particles were discarded based on low correlation coefficient. Using this sorting approach, we observed two similar ribosome-scSRP^{219} complex structures containing 32,170 particles and 20,203 particles, and two additional volumes with low particle numbers representing low-quality particles in the pool. The pool containing 32,170 particles (24.3% of the total data set) displayed a SRP density that was better defined compared to the second pool, and was therefore used for the final refinement.

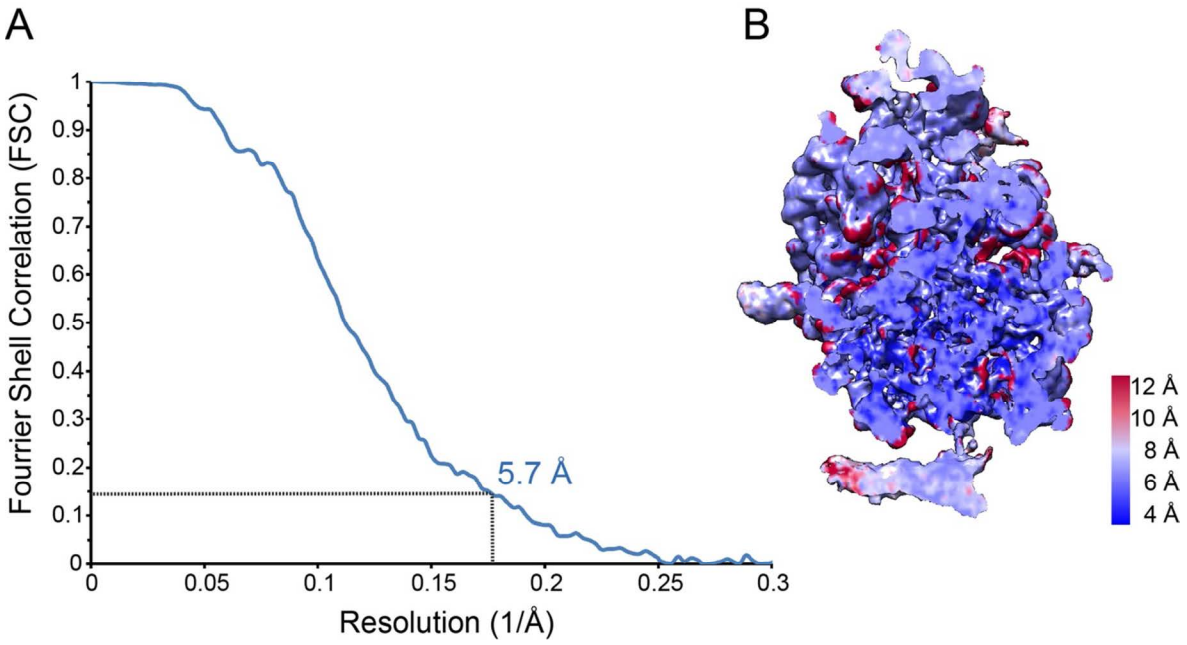


Figure S3: Resolution of the cryo-EM reconstruction of the *closed* RNC-SRP-FtsY complex. (A) Fourier Shell Correlation (FSC) curve of the cryo-EM reconstruction of the *closed* targeting complex, obtained by refining a separate model for two independent halves of the data in RELION (9). The resolution estimate is 5.7 Å according to the FSC = 0.143 criterion (10) and 9.0 Å according to the FSC = 0.5 criterion. **(B)** Local resolution estimation of the cryo-EM reconstruction. The unfiltered, non-sharpened cryo-EM map is shown in a sliced view. The local resolution has been measured using the ResMap program (11). The map is colored based on the calculated local resolution from 4 Å (dark blue) to 12 Å (red).

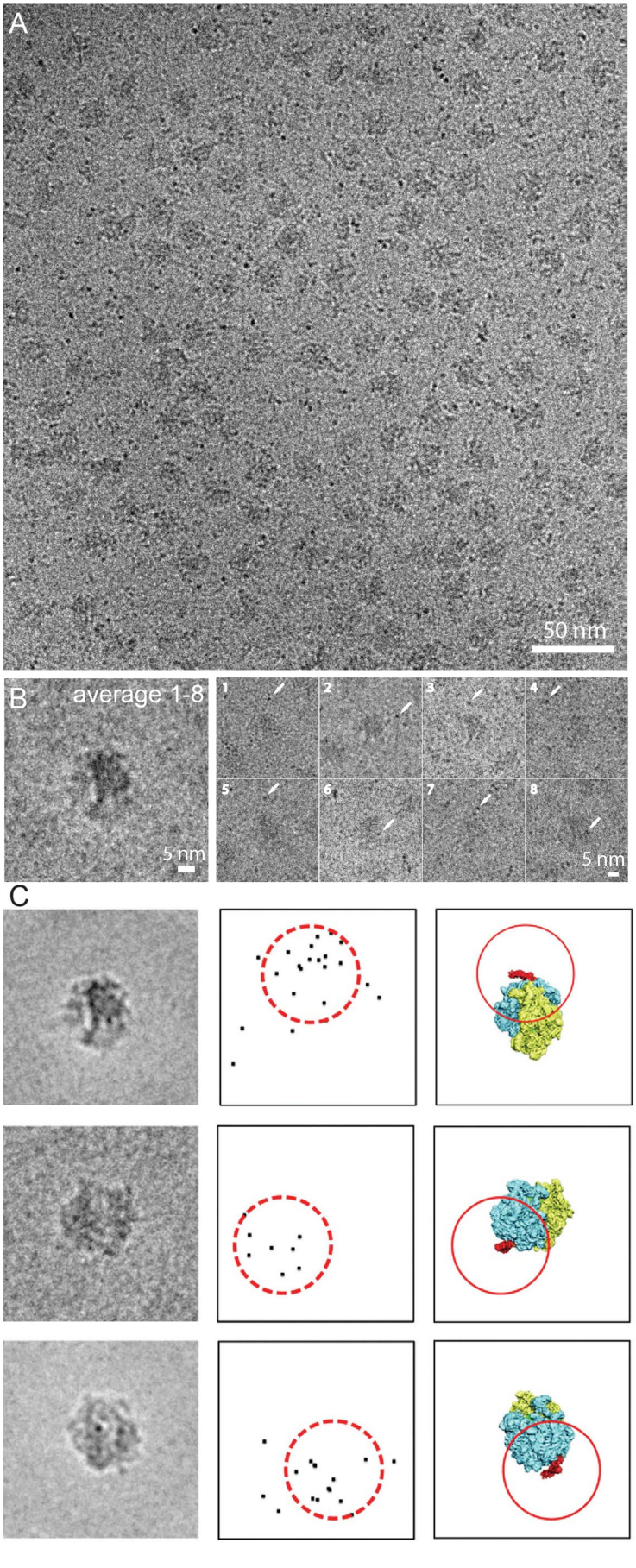

A
50 nm
B
average 1-8
5 nm
1
2
3
4
5
6
7
8
5 nm
C

Figure S4: Nano-gold-labeling of the FtsY-Ffh NG-domains and cryo-EM localization experiments. (A) A typical micrograph with RNC^{Lep50} and gold-labeled $scSRP^{219}2$ particles in the presence of GMPPCP. To shift the equilibrium towards complex formation, $scSRP^{219}2$ was added in ten-fold molar excess to ribosomes. **(B)** Eight individual particles used for localization of the FtsY-Ffh-NG domains and the average of these eight individual $scSRP^{219}2$-RNC particles calculated with SPIDER (AS R). Gold particles that were picked are indicated by a white arrow. Gold particles positioned at the edge of the box were not picked. A 50 nm scale bar and a 5 nm scale bar are shown for the micrograph and the particles, respectively. **(C)** 2D class averages of gold-labeled particles in three orientations in which SRP density can be identified (first column), plot of the position of the nano-gold in the individual images that were used to generate the 2D class averages (column 2) and surface view of the same orientation as the 2D class average (column 3). The red circle (dashed line in the plots of the gold position and solid line in the surface representation) indicates the theoretical volume in which the nano-gold could be theoretically found in the *closed* complex assuming flexible Ffh-FtsY NG-domains.

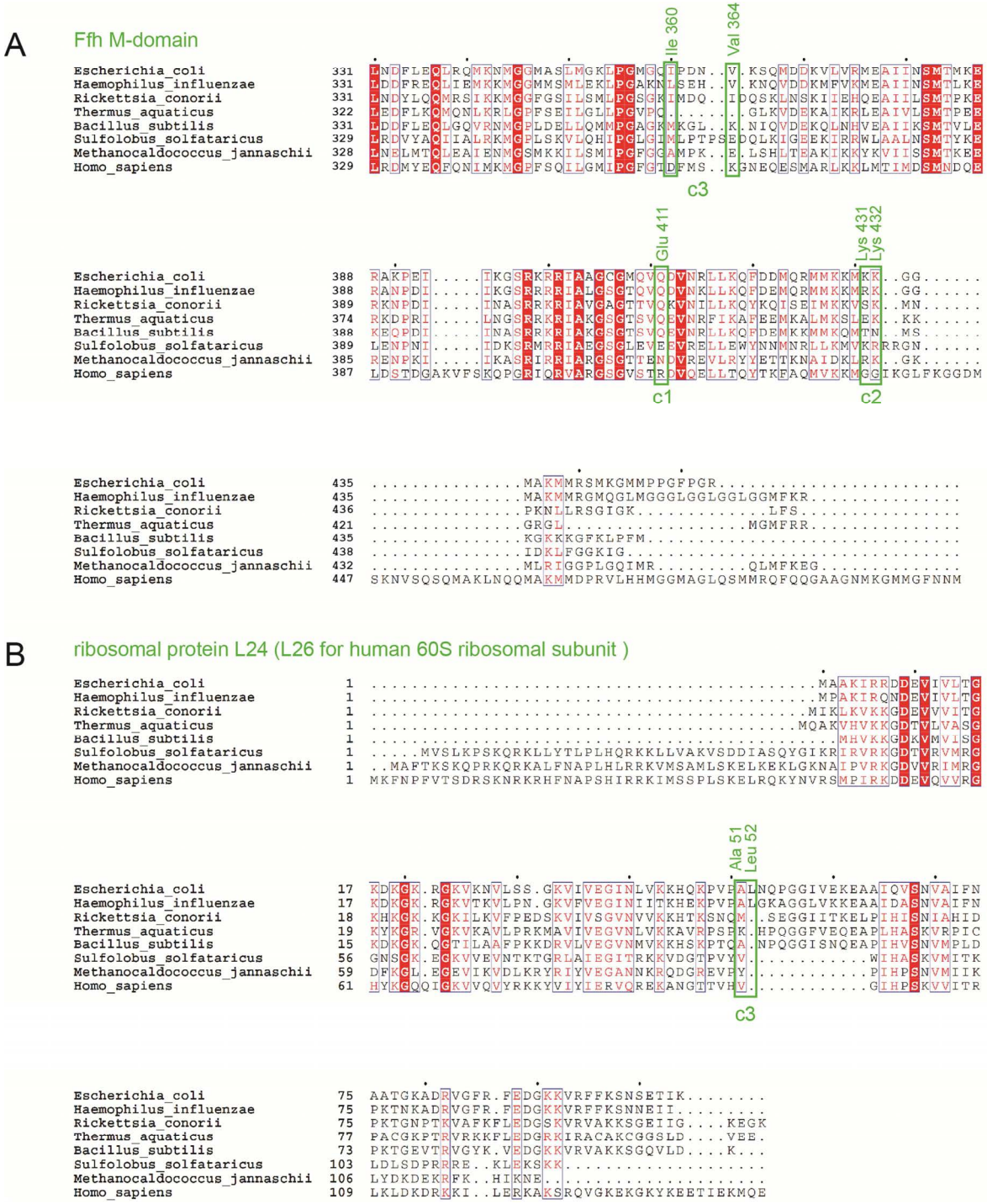


Figure S5: Conservation of M-domain residues involved in ribosomal contacts in the *closed* state.

(A) Alignment of M-domain sequences from Gram-negative and Gram-positive (*B. subtilis*) bacteria, archaea (*S. solfataricus* and *M. jannaschii*), and human. SRP crystal structures have been reported for *E. coli*, *T. aquaticus*, *S. solfataricus*, *M. jannaschii* and human. **(B)** Alignment of ribosomal protein L24 (L26 in human) sequences from the same species as in panel A. Conservation is indicated by color intensity. Among bacteria significant conservation is observed in the residues involved in ribosomal contacts (c1-c3) which are marked in green.

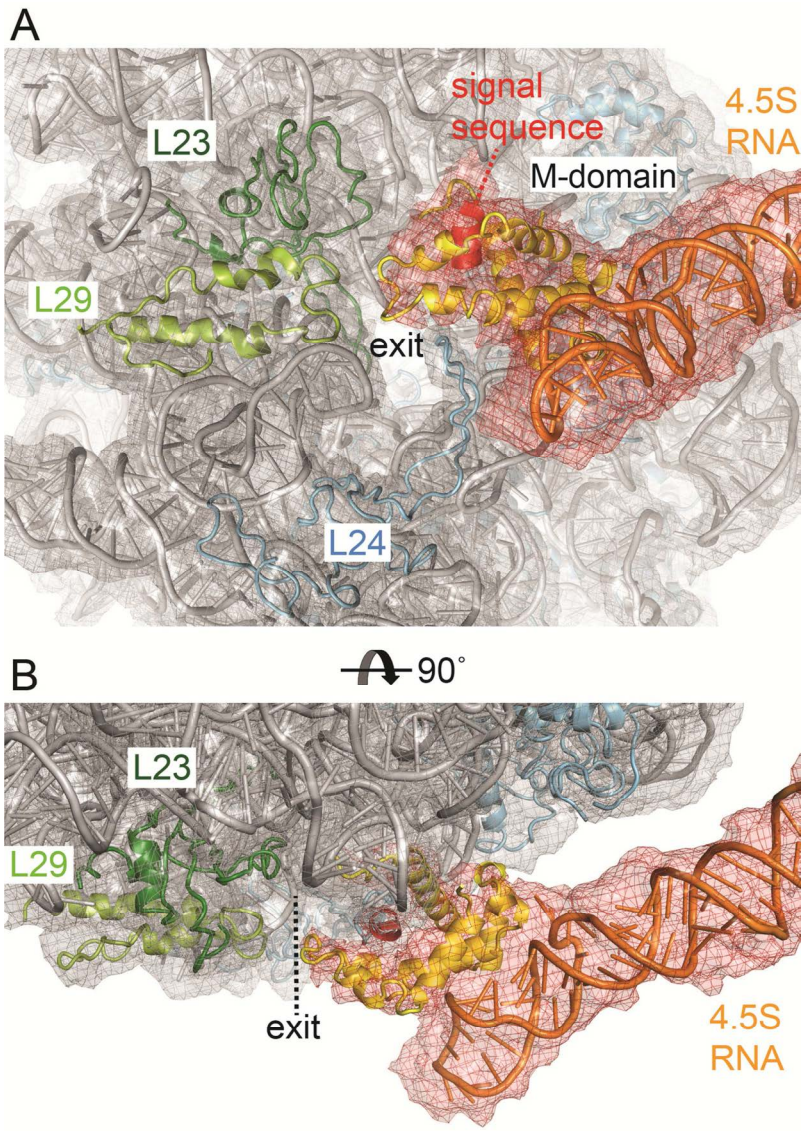


Figure S6: Ribosomal proteins L23 and L29 are accessible for translocon binding in the *closed* state. The structure and the quasi-atomic model are shown **(A)** in a view into the exit of the ribosomal tunnel and **(B)** in a side view. Density corresponding to the ribosome is depicted in transparent grey, density corresponding to the SRP in red. The SRP RNA is colored orange, the M-domain yellow and the signal sequence red. The ribosomal RNA is shown in grey, ribosomal proteins in cyan, L29 in green-yellow and L23 in forest. The position of the exit is marked.

SI REFERENCES:

1. Estrozi LF, Boehringer D, Shan SO, Ban N, Schaffitzel C (2011) Structure of the *E. coli* co-translational targeting complex in the stable *early* conformation. *Nat Struct Mol Biol* 18:88-90.

2. Zhang X, Kung S, Shan SO (2008) Demonstration of a multistep mechanism for assembly of the SRP x SRP receptor complex: implications for the catalytic role of SRP RNA. *J Mol Biol* 381(3):581-593.

3. Schaffitzel C, Ban N (2007) Generation of ribosome nascent chain complexes for structural and functional studies. *J Struct Biol* 158(3):463-471.

4. Peluso P, Shan SO, Nock S, Herschlag D, Walter P (2001) Role of SRP RNA in the GTPase cycles of Ffh and FtsY. *Biochemistry* 40(50):15224-15233.

5. Shaikh TR*, et al.* (2008) SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nat Protoc* 3(12):1941-1974.

6. van Heel M, Harauz G, Orlova EV, Schmidt R, Schatz M (1996) A new generation of the IMAGIC image processing system. *J Struct Biol.* 116(116):17-24.

7. Tang G*, et al.* (2007) EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* 157(1):38-46.

8. Pettersen EF*, et al.* (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605-1612.

9. Scheres SH (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* 180(3):519-530.

10. Rosenthal PB, Henderson R (2003) Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *J Mol Biol* 333(4):721-745.

11. Kucukelbir A, Sigworth FJ, Tagare HD (2014) Quantifying the local resolution of cryo-EM density maps. *Nature Methods* 11(1):63-65.

2.3 Detailed procedures of data processing and structural refinement

2.3.1 Particle picking and micrograph pre-processing

The dataset of RNC^{Lep50}-$scSRP^{219}$-GMPPCP was collected on a Titan Krios (FEI) operated at 300 kV, with a direct electron detector (Falcon II, FEI) of a calibrated pixel size of 1.08 Å per pixel. A total number of 2,840 contrast transfer function (CTF) -corrected micrographs were obtained.

I first performed a reference-based automated particle picking using the software package EMAN2 (Tang *et al.*, 2007) and a 70S ribosome as reference map (EMDataBank accession: EMD-1045) (Valle *et al.*, 2002). Afterwards, each micrograph was individually inspected to remove obvious “mistakes” by the software, and to pick manually some more ribosomal particles. Overall, 134,350 particles were selected though this semiautomatic method.

The cryo-EM micrographs were collected such that fine details could be preserved as much as possible. High-frequency signals which carry high-resolution information are by theory limited to half of the sampling rate, according to the Nyquist-Shannon theorem. This results in a theoretical resolution limit of twice of the pixel size. This is why micrographs are recorded with small pixel sizes. In practice however, high frequency signals are often lost before the Nyquist frequency due to factors such as defocus. The use of smaller pixel size (higher magnification) means more pixels are needed to record the same object, and fewer objects are present on one micrograph. Moreover, this results in bigger particle dimensions which cost more computational power. Since in the early stages of data processing I focused on large compositional and conformational differences like the presence or absence of the SRP and since I tested different sorting strategies, I initially down-sampled the images to increase the pixel size thus speeding up calculations. In fact, the high resolution information is not necessary at the early steps. I resampled the micrographs to different pixel sizes during the different stage of data processing.

Down-sampling of images used to be achieved by a method called binning, which is averaging neighboring pixels into one. A new method has been developed to

perform the down-sampling in Fourier space and has been shown to preserve the high-resolution information better (Sorzano *et al.*, 2009).

I first down-sampled all micrographs to 10 Å per pixel to perform an initial 2D analysis. The command in software Xmipp 3.0 (de la Rosa-Trevín *et al.*, 2013) is applied to individual micrographs. Thus, a simple bash script was written to facilitate the process.

```
for micrograph in *.mrc; do
xmipp_transform_downsample -i $micrograph -o ../SRP_sample10/$micrograph --step 9.259
done
```

All ribosomal particles were then extracted from these down-sampled micrographs. As the particles were picked in the original micrographs, the coordinate files needed to be adjusted to adapt them to the now smaller micrograph dimension. This was done by the following script. The operation is now automated with the newest version of RELION.

```
for boxfile in *.box; do
awk '{printf("%-7i %-7i %-7i %-7i\n",$1/9.259-3.8,$2/9.259-3.8,"'40'","'40'")}' $boxfile > ../SRP_sample10_boxfiles/$boxfile
done
```

2D class-averaging was performed using the software IMAGIC (van Heel *et al.*, 1996). The results were examined to identify “bad” groups of particles (noisy, distorted 2D class average). These particles were removed from the particle pool by removing the corresponding coordinate records from their box files. This procedure left 132,390 particles to be further analyzed in 3D.

2.3.2 Supervised classifications for SRP-bound particles

The 132,390 particles were extracted with pixel size of 5 Å, using similar scripts as mentioned above. The images were normalized and contrast-inverted using XMIPP (de la Rosa-Trevín *et al.*, 2013). To perform 3D refinements in RELION, an initial model (reference map) is required. The reference map (EMD-1045, Valle *et al.*,

2002) was resized to 5 Å per pixel in order to have corresponding dimensions using Bsoft (Heymann and Belnap, 2007).

```
bint -scale 0.538,0.538,0.538 -origin 0,0,0 -translate 3.4,3.4,3.4 -size 80,80,80
emd_1045.map sample5_emd_1045.mrc
```

After grey-scale correction of the input map (this procedure is also automated in the newest version of RELION), auto-refinement was performed in RELION using the following setting:

```
mpirun -np 144 -machinefile mpi_file_1 relion_refine_mpi --i IMAGIC_cleaned.star --o
Refine_0307/0307 --angpix 5 --iter 50 --tau2_fudge 2.5 --ref
GScorrct_sample5_emd_1045.mrc --particle_diameter 340 --ini_high 40 --flatten_solvent
--auto_refine --split_random_halves --offset_range 6 --offset_step 1 --norm --
max_memory 6 --pool 1
```

The resulting map had a resolution of 10.3 Å (Gold-standard refinements with FSC 0.143 criterion, same for the rest of the Chapter). This means the Nyquist limit was reached (as discussed above). The resulting map showed clearly the features of ribosomal RNA, and an extra density at the ribosomal tunnel exit (Figure 2-2, first row). This extra density which represents the scSRP[219] with the bound signal peptide will be referred as 'SRP' in the following texts. The connection between the ribosome and the SRP was relatively weak. This suggests the existence of both 'empty' ribosomes and SRP-bound ribosomes, which was expected based on previous experience with SRP-ribosome complexes (Schaffitzel *et al.*, 2006; Estrozi *et al.*, 2011, von Loeffelholz *et al.*, 2013). The difference between 'empty' ribosome and SRP-bound ribosomes is huge and thus sorting could be done even when the resolution was limited to 10 Å. Such sorting of known structures was usually performed in a supervised fashion, *i.e.* by providing two input maps corresponding to the SRP-bound ribosome and empty ribosome.

In the first attempt, the initial map calculated from all particles was used to represent the SRP-bound ribosome. Another copy of this map in which the SRP density was deleted using UCSF Chimera (Pettersen *et al.*, 2004) was used to mimic the empty ribosome, such that the only difference of the two maps was the presence or absence of the density at the ribosomal tunnel exit (Figure 2-2, second row).

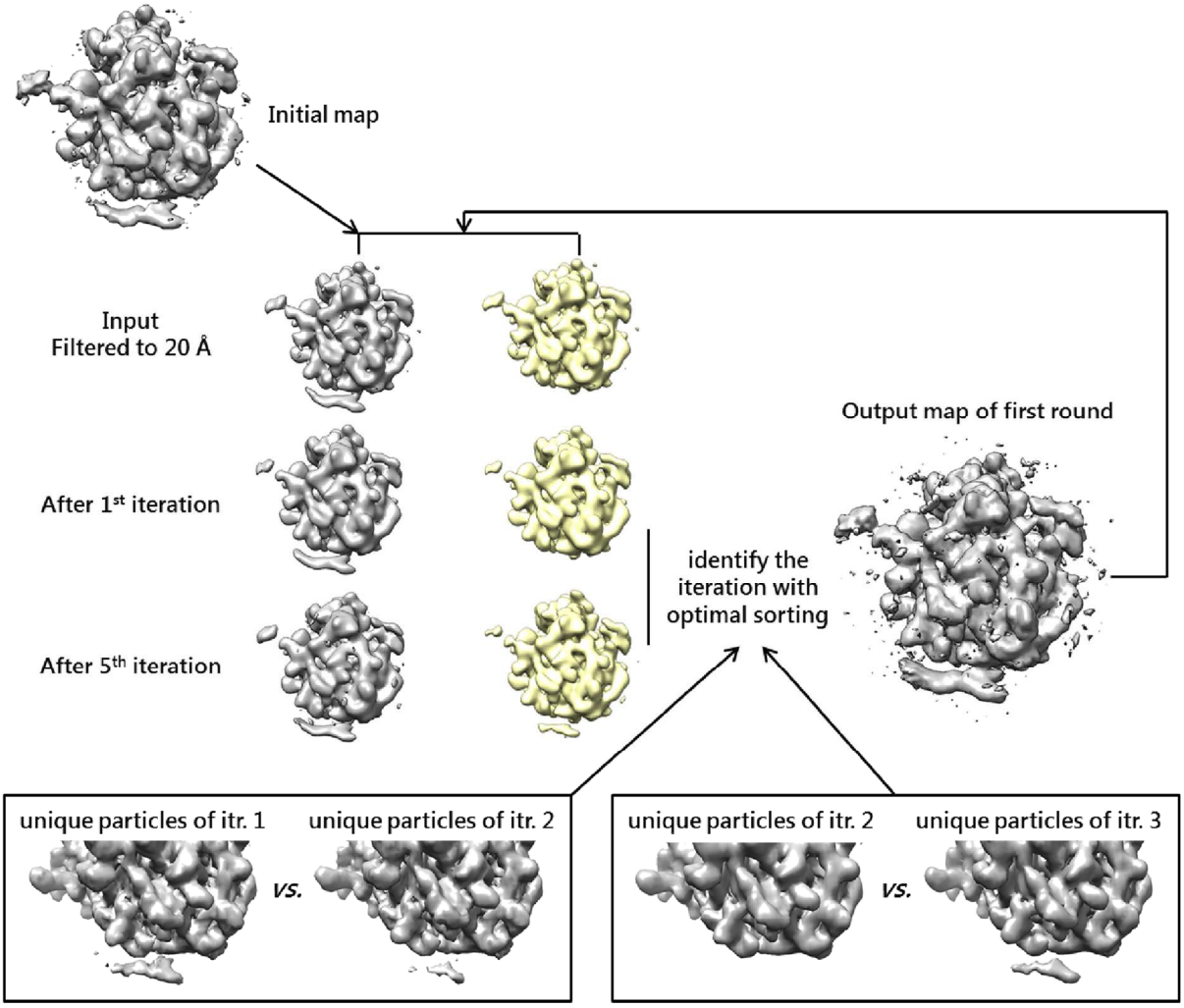


Figure 2-2: Sorting scheme for SRP-bound particles. First row: The initial map refined from all particles. This map was low-pass filtered to 20 Å to be used for the first round of sorting. Second row and below: Resulting maps from supervised classification. Maps in grey represent the SRP-bound pool and maps in yellow the vacant pool (empty ribosomes). The "jumper" particles aligning in different rounds to empty or SRP-bound ribosomes were examined further to compare the sorting results of iterations as described in the main text. The resulting SRP-bound map, in this case from iteration 2, was used for a second round of classification following the same scheme. The resulting map had a stronger SRP density compared to the initial map.

```
mpirun -np 144 -machinefile /user/qjiang/mpi_file_1 relion_refine_mpi --i
IMAGIC_cleaned.star --o Classify_0312/0312 --angpix 5 --iter 30 --tau2_fudge 2.5 --
healpix_order 2 --ref ref_0312.star --particle_diameter 340 --flatten_solvent --ini_high 20
--offset_range 6 --offset_step 1 --norm --max_memory 6 --pool 1
```

Supervised classification has been widely used before and usually involved several iterations (*e.g.* Schaffitzel *et al.*, 2006). The idea is that the input reference

for the empty ribosome has pixels of zero intensity around the ribosomal tunnel exit; this is due to the treatment (deletion of SRP density) in UCSF Chimera. This may introduce artifacts as the rest of the map where there is no protein/RNA presence is filled by noise. I expected that such artifacts could be corrected for during the evolution of the maps. However, the densities around ribosomal tunnel exit came back after roughly five iterations (Figure 2-2), and the two maps appeared to be identical after 15 iterations. This means the classification was subject to another criterion rather than the presence or absence of SRP (discussed in Section 4.1.2). Therefore, the question remained how to sort out particles corresponding to empty ribosomes.

I decided to compare the particle pools generated during the first three iterations of supervised classification. For example, 61,370 particles have been assigned to the empty ribosome pool in the first iteration, and 60,550 particles in the second iteration. The two pools had 38,232 particles in common, whereas 23,138 particles were unique in iteration one and 22,318 particles in iteration two. I decided to further analyze these ‘unique’ particle sub-pools. To do so, each pool of ‘unique’ particles was refined into a ribosome structure, and the two structures were visually checked (Figure 2-2, bottom row). In fact, we analyzed the ‘unique’ particles of the empty ribosome pools because the presence of SRP-bound ribosome particles in the empty ribosome pool can easily be detected at higher contour levels when displaying the map. Vice versa, contamination of the SRP-bound ribosome pool with empty ribosomes leads to maps with slightly weaker densities at the exit of the ribosomal tunnel which is more difficult to evaluate. The unique particles were identified using the following script:

```
awk '$8==1 {print $1}' Classify_0312/0312_it001_data.star > 0312_it1_class1.star
awk '$8==1 {print $1}' Classify_0312/0312_it002_data.star > 0312_it2_class1.star
comm -2 -3 0312_it1_class1.star 0312_it2_class1.star > it1_only.star
```

The results of this analysis indicated that iteration two had a more efficient separation between empty and SRP-bound ribosomes, as the corresponding map displayed almost no SRP density (Figure 2-2, bottom row). As mentioned in Section 1.3.3, the quality of input maps is critical for supervised classification. While the initial

input maps were based on the average map of all particles, it was contaminated by empty ribosomes. Therefore, a second round of such supervised classification was performed using the new SRP-bound ribosome structure that was generated from the sorted particle pool and showed enhanced EM density for the SRP. (Figure 2-2, right)

Thus, the 67,978 particles corresponding to the SRP-bound ribosome pool resulting from the second iteration were used for further unsupervised classification in order to sort for conformational varieties.

2.3.3 Unsupervised classifications for conformational varieties

Conformational variety always exists in cryo-EM datasets due to the absence of a crystal lattice. Heterogeneity happens at different levels, from thermodynamic movements to rearrangements of entire domains. Therefore, computational sorting is required to improve the quality of the map. In practice, sorting is often counter-balanced by a significant reduction of the number of particles in each pool.

SRP-bound particles were extracted with a pixel size of 2.5 Å as the high-resolution information in the images may be useful from this step on. Again, an average map of all particles was first generated using the RELION auto-refine routine. The resolution of the resulting EM map was 5.6 Å. The map showed a very well-defined ribosome, but the density accounting for the SRP was of relatively low quality. This suggested that further sorting step to improve the SRP density were required. As discussed in Section 1.3.3, such heterogeneities of unknown nature are better sorted by unsupervised classification. To focus the computational sorting on the SRP density, which we aimed to improve, a spherical, soft-edge mask was applied during the classification.

```
beditimg -create 160,160,160 -sampling 2.5.2.5.2.5 -sphere 105,80,80,52 -edgewidth 5 -
fill 1 -background 0 mask_X25_R52_G5.mrc
```

I tested the effects of numerous parameters during the multi-reference refinement, including the number of classes to be generated, the filtering of input maps, the regularization parameter (tau2_fudge in the command line), the filling of the masked

region and whether the mask is applied. Finally, the best separation of particles was achieved with the following settings:

```
mpirun -np 144 -machinefile /user/qjiang/mpi_file_1 relion_refine_mpi --i
Refine_0327/0327_data.star --o Classify_0421/0421 --K 4 --angpix 2.5 --iter 25 --
tau2_fudge 2.5 --ref Refine_0327/0327_class001.mrc --ini_high 40 --flatten_solvent --
solvent_mask mask_X25_R52_G5.mrc --healpix_order 5 --sigma_ang 1.66667 --
offset_range 3 --offset_step 1 --norm --max_memory 6
```

It is worth mentioning the local angular searching tool: During the calculations, 3D structures are represented by their 2D projections, and the angular sampling rate determines how finely different views of the structures are sampled. Small conformational changes will be noticed only with very fine sampling rate, but such sampling requires significantly more computing power. After global sampling, the idea of local angular searching is to firstly use large sampling intervals to roughly

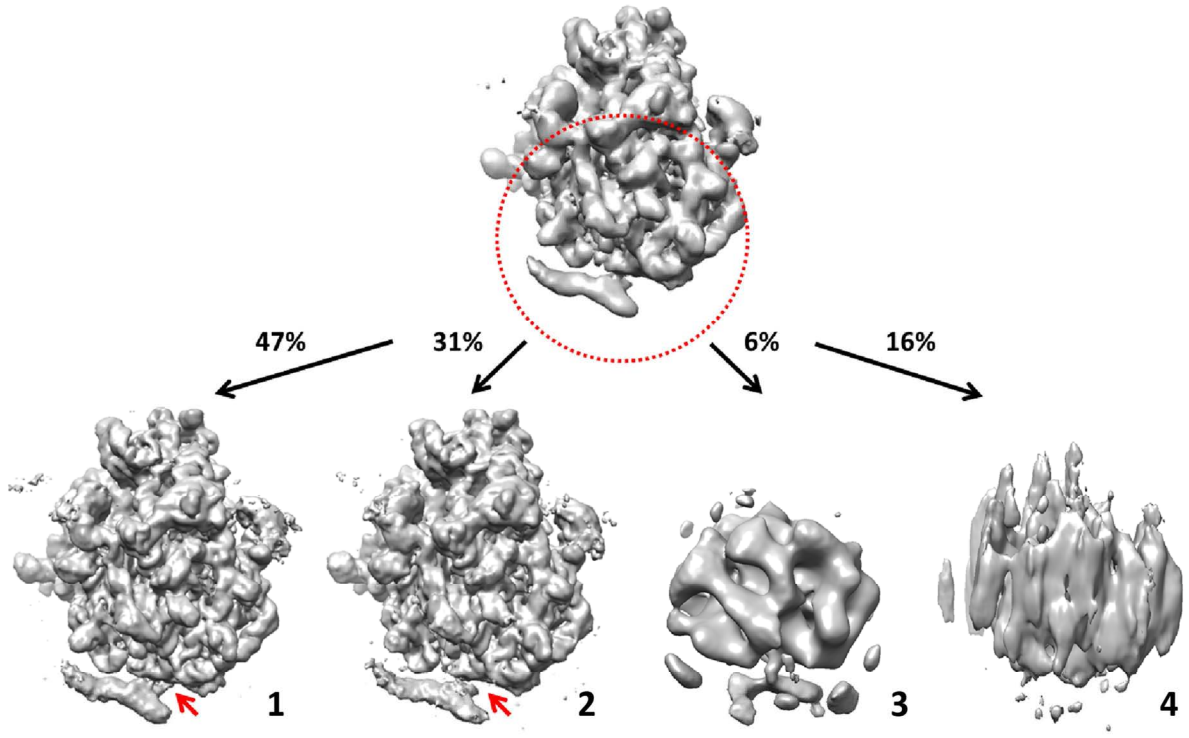


Figure 2-3: Sorting scheme for conformational variety. First row: The average map of SRP-bound particles was low-pass filtered to 20 Å to be used for unsupervised classification. The edge of the spherical mask used during the classification is illustrated with red dots. Second row: The four resulting maps from the 3D classification. The major difference between class 1 and class 2 is marked by red arrows.

determine the orientation of each particle, and then finer samplings are applied only around this angle to search for structural variation and to refine the Euler angles. This function is very useful for small conformational changes in a large molecule like the ribosome, however is not activated with the default setting of the software.

The sorting resulted in four classes of volumes. Two volumes were meaningless structures representing “junk” particles (dust, ice, aggregates). The other two volumes both represent SRP-bound ribosomes. In a close-up view, volume class 1 contained a better-resolved SRP while in volume class 2, the density for SRP RNA looked distorted, and the connection between the ribosome and the SRP was extremely weak (Figure 2-3). Volume class 1 was obviously better and therefore was further refined. I wondered what the difference is between class 2 and class 1? Does class 2 represent a distinct conformation compared to volume class 1, or did it arise from artifacts during the calculation?

While extracting particles from volume class 1, I noticed that some micrographs contributed no particles to this class. It was unexpected since each micrograph contained on average more than 45 particles, and therefore statistically more than 10 particles of each micrograph should belong class 1 (class 1 comprises 32,170 particles, comprising 24.3% of the total data set with 132,390 particles). I suspected therefore that the separation of particles into class 1 and class 2 is related to the defocus values of the micrographs. I wrote the following script to test this hypothesis: The resulting plot is shown in Figure 2-4.

```
string=1484
while [ $string -le 4383 ]; do
used=`wc -l SRP_sample2/CTF_140127_closedRiboSRP_0$string.box | awk '{print $1}'`
all=`wc -l SRP_sample2.5_shifted/CTF_140127_closedRiboSRP_0$string.box | awk '{print $1}'`
ratio=`echo "$used/$all" | bc -l `
defocus=`grep 140127_closedRiboSRP_0$string CTF_values.txt | awk '{print $2}'`
printf "%5f | %5f\n" $ratio $defocus >> versus.txt
string=$((string+1))
done
```

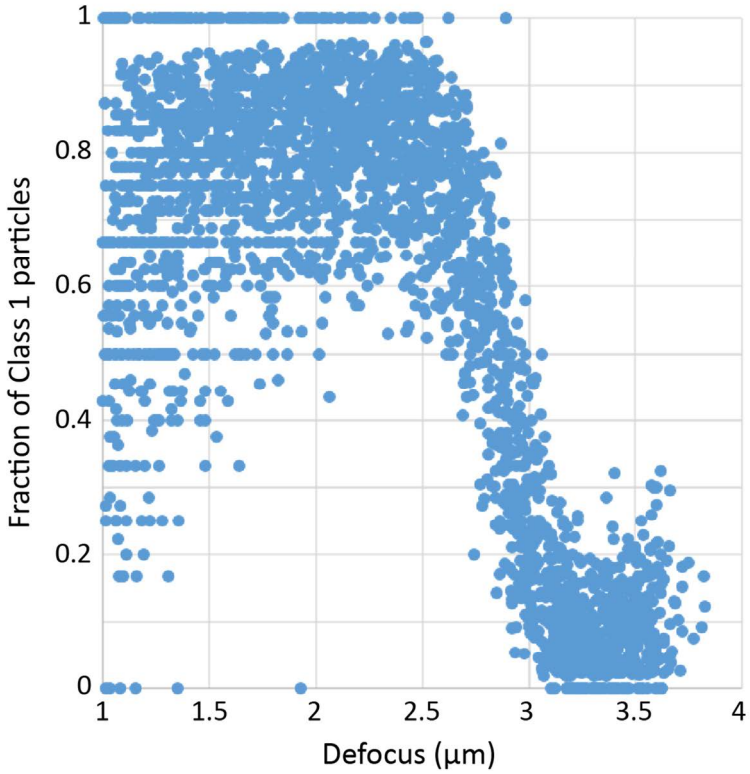


Figure 2-4: Correlation between sorting of particles into class 1 and micrograph defocus. The correlation between the fraction of particles belonging to class 1 and the defocus value of the corresponding micrograph has been plotted as described in the main text. Here, each spot represents one micrograph of the dataset. Its defocus value and the fraction of particles from this micrograph that were sorted to class 1 are marked. When the defocus of the micrograph exceeds 2.5 µm particles are less likely to be sorted into class 1.

The plot demonstrates a tendency of particles falling into class 2 when the defocus value of the micrograph exceeds 2.5 µm. Somehow the classification was led by certain information which was defocus related. The conclusion was that the class 1 and class 2 should represent biologically identical objects. In fact, it would be difficult to imagine that certain species exists in large amount on some micrographs but not on others. At this point, I decided to further refine the 32,170 particles from class 1 aiming for high-quality maps.

2.3.4 Refinement towards high resolution

The first refinement of the particles belonging to the class 1 resulted in a map of 6.58 Å resolution. Although the overall resolution by number was worse than previously before the unsupervised classification (5.6 Å), the SRP density was certainly better-resolved. Appling a soft-edge mask during the refinement further improved the resolution to 6.25 Å. Further optimization of the centering of the input map and the mask resulted a better structure with 5.88 Å resolution. Since the Nyquist limit was

again within reach, particles were re-extracted with a pixel size of 2 Å, and this led to a slightly improved resolution at 5.8 Å. In the last step, post-processing functions (B-factor sharpening) were applied, and the final map was generated at 5.7 Å resolution. The commands used are the follows.

```
bint -size 180,180,180 -origin 0,0,0 -translate -10,-10,0 -neworigin 0,0,0 -scale 2.5,2.5,2.5 ../archive/SRP_Relion_140307/GScorrct_sample5_emd_1045.mrc Gscorrct_sample2_emd_1045.mrc

beditimg -create 180,180,180 -sampling 2,2,2 -sphere 118,90,90,62 -edgewidth 5 -fill 1 -background 0 mask_X28_R62_G5.mrc

mpirun -np 160 -machinefile /user/qjiang/mpi_file_1 relion_refine_mpi --i all_ptcls.star --o Refine_0509/0509 --angpix 2 --tau2_fudge 2.5 --ref GScorrct_sample2_emd_1045.mrc --ini_high 40 --flatten_solvent --solvent_mask mask_X28_R62_G5.mrc --auto_refine --split_random_halves --norm --max_memory 4

relion_postprocess --i Refine_0509/0509 --angpix 2 --o Refine_0509/Post_autob --auto_bfac
```

CHAPTER 3: THE HOLO-TRANSLOCON AND THE RIBOSOME-HTL COMPLEX

CHAPITRE 3: L'HOLO-TRANSLOCON ET LE COMPLEXE RIBOSOME-HTL

Résumé

HTL est un complexe de protéine membranaire constitué de sept sous-unités différentes. L'expression et la purification du complexe de HTL est décrite en détail dans ce chapitre.

Plusieurs approches différentes ont été testées pour préparer des complexes ribosome-HTL et en résoudre la structure en cryo-EM à haute résolution. En utilisant HTL solubilisé dans du n-dodécyl-ß-D-maltoside (DDM), une première structure a été obtenue avec une résolution de 22 Å. Des efforts ont donc été faits afin d'insérer le complexe HTL dans un nanodisc, mais le rendement et la pureté de l'échantillon n'ont pas permis de préparer des complexes pour des études cryo-EM de haute qualité. L'utilisation d'un nouveau détergent le lauryl maltose néopentylglycol (LMNG) et une méthode de fixation par gradient ont été explorés. Un deuxième ensemble de donnés a été obtenu en utilisant HTL solubilisé par LMNG en complexe avec le ribosome en cours de traduction stabilisé par la réticulation glutaraldéhyde. La structure cryo-EM de ce complexe ribosome-HTL a été résolue à 17 Å donnant ainsi un premier aperçu de la base moléculaire de l'interaction entre le ribosome et le complexe de HTL. La préparation de l'échantillon d'HTL a été optimisée en utilisant les amphipoles. Les amphipoles ont permis la production d'échantillons d'HTL homogènes dans des tampons sans détergent. La liaison d'HTL solubilisé dans deux types d'amphipoles avec le ribosome est démontrée dans ce chapitre.

3.1 Expression and purification of HTL

3.1.1 The expression construct

The holo-translocon (HTL) is a membrane protein complex of seven subunits, with a total molecular weight of 240 kD. Expression of such large, multi-gene protein complexes especially in the case of membrane proteins can be very tedious and a good level of stoichiometric expression is usually difficult to archive. The Acembl co-expression system was designed to facilitate such multigene recombinant expression experiments in *E. coli* (Bieniossek *et al.*, 2009). In fact, HTL was among the first complexes that benefit from the Acembl system. The initial HTL1 construct did not comprise YajC. The HTL3 construct (Schultze *et al.*, 2014) which I used during my thesis (Figure 3-1) was developed based on the HTL1. The idea to include YajC which has a hitherto unknown function was to investigate whether it may stabilize the SecYEG-SecDF-YajC complex which was very aggregation-prone. Moreover, YajC was tagged with a calmodulin-binding peptide allowing for a second affinity purification step. In fact, much better protein yields could be achieved using the HTL3 expression construct compared to HTL1, but it was not possible to show a direct stabilizing effect of YajC on the complex. The expression and purification protocol for HTL3 which I initially used during my thesis was established by a former PhD student in the laboratory (Mathieu Botte). Here, I give an example of the expression and purification of HTL3 using this protocol.

3.1.2 HTL expression and purification

The plasmid pAcembl-HTL3 was transformed into the *E. coli* BL21 star (DE3) strain. The cells were grown in 2-YT Broth (16 g Tryptone, 10 g Yeast Extract and 5 g NaCl

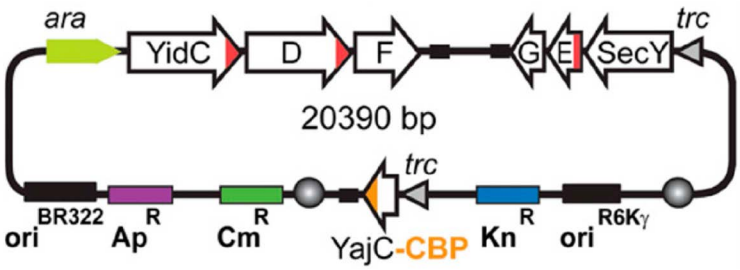


Figure 3-1: The pAcembl:HTL3 expression construct. The hexa-histidine tags and the calmodulin-binding peptide tag are marked in red and orange respectively (Image taken from Schulze *et al.*, 2014).

per one liter culture) in the presence of 50 µg/ml ampicillin, 20 µg/ml chloramphenicol and 30 µg/ml kanamycin. The cells were cultured at 37°C. At OD_{600} 0.7 the expression of proteins was induced by adding 0.5 mM IPTG and 0.2% L-arabinose. After an additional 3 hours of culturing at 37°C, the cells were harvested and fast-frozen in liquid nitrogen. The cell pellets were stored at -20°C for no more than one week.

The cell pellets were re-suspended in HSGM buffer (20 mM Hepes-KOH pH 8.0, 130 mM NaCl, 10% (w/v) glycerol, 5 mM $Mg(OAc)_2$,) supplied with protease inhibitors (Roche cOmplete EDTA-free, one tablet per 50 ml) to ~40% (w/v). The cells were disrupted using a microfluidizer operated at 18 kpsi, and 0.5 mM phenyl-methane-sulfonyl-fluoride (PMSF) was quickly added to the lysate. A first centrifugation at 16,000 RCF in a JA-25.50 rotor (Beckman Coulter) for 30 minutes should pellet cell debris. The supernatant was subjected to ultracentrifugation at 150,000 RCF in a 70 Ti rotor (Beckman Coulter) for 1 hour to isolate membranes. Membrane pellets were homogenized into ~30% (w/v) in the same HSGM buffer used for cell suspension. The membrane suspensions were fast-frozen in liquid nitrogen and then stored in the -80 °C freezer.

n-Dodecyl-ß-D-maltoside (DDM, from Glycon, Luckenwalde, Germany) was used to solubilize the membranes. 10% (w/v) DDM solution was added to the membrane suspension in a 15:85 (v/v) ratio, such that the final concentration of DDM is 1.5%. After 2 hours of gentle agitation at 4°C, the mixture was centrifuged at 150,000 RCF in a 70 Ti rotor (Beckman Coulter) for 1 hour at 4°C to remove insoluble particles. For immobilized metal ion affinity chromatography (IMAC) the supernatant was loaded onto a HisTrap HP column (GE Healthcare) equilibrated with HSGM buffer + 10 mM imidazol + 0.1% DDM. The column was first washed by HSGM buffer + 10 mM imidazol + 0.1% DDM for at least 10 column volumes, followed by a one-column-volume high-salt wash with HSGM buffer + 370 mM NaCl + 10 mM imidazol + 0.1% DDM. After further washing with HSGM buffer + 50 mM imidazol + 0.1% DDM for at least 5 column-volumes, the proteins were eluted by HSGM buffer + 300 mM imidazol + 0.1% DDM. The eluate was desalted to 50 mM Hepes-KOH pH 8.0, 130 mM NaCl, 10% (w/v) glycerol, 2 mM $CaCl_2$, 0.03% DDM using a HiTrap Desalting column (GE Healthcare). The proteins were then mixed with calmodulin

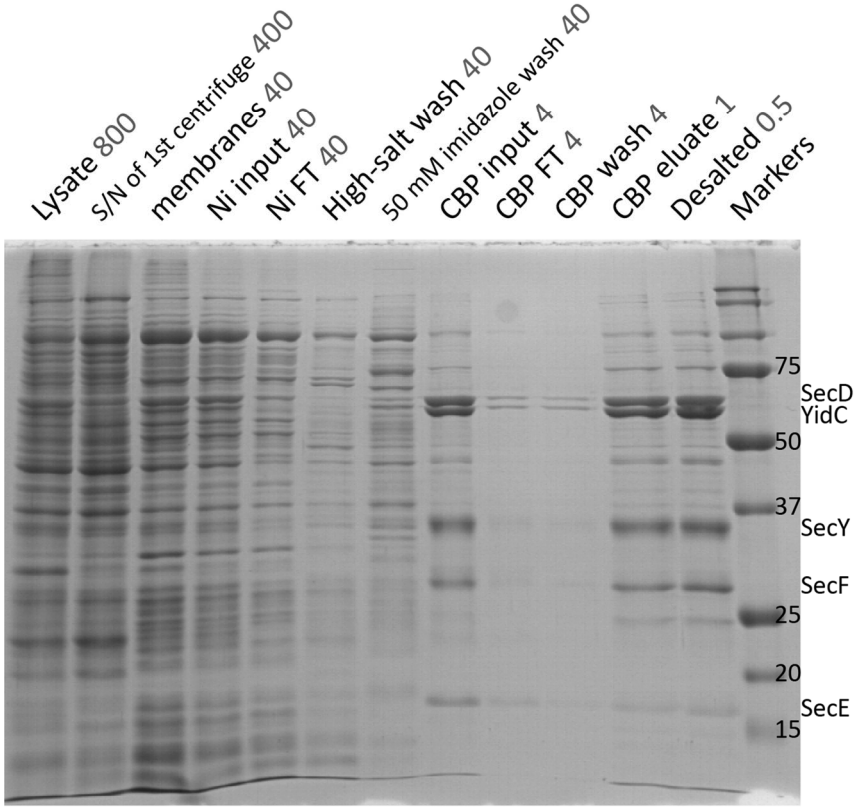


Figure 3-2: An example of HTL expression and purification. Coomassie stained SDS-PAGE gel of samples from an expression and purification of HTL. Abreviations used are: S/N for supernatant; FT for flow-through; Ni for Ni-NTA affinity purification step; CBP for calmodulin-binding peptide affinity purification step. The relative dilution factors of each sample compared to the CBP eluate are marked in grey. SecG and YajC are not seen due to their smaller sizes and weak Coomassie staining.

affinity resin (Stratagene/Agilent) (usually 1 ml resin per 5 mg protein). After incubation with gentle agitation (ranging from 2 hours to overnight incubation), the calmodulin resin was washed with 10 column volumes of 50 mM Hepes-KOH pH 8.0, 130 mM NaCl, 10% (w/v) glycerol, 0.2 mM $CaCl_2$, 0.03% DDM. Subsequently, the proteins were eluted by 50 mM Hepes-KOH pH 8.0, 400 mM NaCl, 3% (w/v) glycerol, 2 mM EGTA, 0.03% DDM. Finally the buffer of the purified proteins was exchanged according to the needs of the subsequent experiments. Figure 3-2 details each step of this purification procedure visualized by elution profiles and Coomassie-stained SDS-PAGE.

The yield of this purification is quite low. Typically from 12 L of cell culture, I obtained 1 mg of purified HTL complex. The eluted HTL from the calmodulin affinity column typically has a protein concentration of ~ 0.5 mg/ml, and there is still considerable heterogeneity as judged by negative-stain EM. Initial attempts have been made to concentrate the sample and to further purify the HTL by size-exclusion chromatography. However, this led to dissociation of the HTL complex into SecYEG and SecDF-YajC-YidC subcomplexes (Botte, thesis, 2014). During the course of my

thesis, different strategies have been pursued to overcome these problems, and the amphipol strategy was shown to be successful.

3.2 First cryo-EM dataset of RNC-HTL

3.2.1 Sample preparation and data collection

The aim of my thesis work was to obtain a cryo-EM structure of the HTL in complex with a translating ribosome. Such a ribosome-HTL complex structure should provide valuable information regarding the functional interaction between the ribosome and the active HTL. In addition, the image processing of such a large complex should be considerably easier compared to cryo-EM of the inactive HTL complex in isolation. The ribosome part of the complex should drive the alignment of the particles leading to higher accuracy of alignment and thus higher resolution of the entire map including the HTL part of the complex.

Previously it was shown that the HTL binds translating and non-translating ribosomes using sedimentation assays (Schultze *et al.*, 2014. The ribosome-HTL sample was prepared by former group member Mathieu Botte by mixing DDM-solubilized HTL (HTL^{DDM}) and ribosome displaying a 108-amino acid long FtsQ nascent chain with a signal sequence (RNC^{FtsQ}), *i.e.*, in a 25:1 molar ratio. The RNC^{FtsQ} complexes were prepared by *in vitro* translation, followed by ultracentrifugation and affinity purification as described previously (Schaffitzel and Ban, 2007).

Data was collected under low dose conditions on a Titan Krios (at EMBL Heidelberg) operated at 200 kV, with a 4k x 4k CCD camera (US4000, Gatan) of a calibrated pixel size of 1.4 Å. A total number of 4,937 micrographs were recorded using a defocus range of 1 to 3 µm. It has to be noticed that at the time this work was done (December, 2012), direct electron detectors were not available to most of the research groups, and the use of new software packages like RELION was not yet widely recognized.

3.2.2 Data processing of the RNC^{FtsQ}- HTL^{DDM}

CTF parameters of these micrographs were estimated followed by manual adjustment using Bctf (Bsoft, Heymann and Belnap, 2007). After sorting out micrographs with drift or strong astigmatism, 2,009 CTF-corrected micrographs were remaining. Reference-based automatic particle picking was performed in EMAN2 using the ribosome volume EMD-1045 as the reference map. The particles were extracted from micrographs which were binned to 5.6 Å per pixel. The particle pool was then subjected to 2D MSA and classification followed by multi-reference alignment in IMAGIC in order to clean-up the data set. This reduced the data set from 31,610 to 30,697 particles.

Subsequent supervised classification and refinement were performed in SPIDER (Shaikh *et al.*, 2008) using particles binned to 4.2 Å per pixel. An initial map was generated from all particles using an empty ribosome (EMD-1045) as initial reference. Extra densities at the ribosomal tunnel exit could be visualized after the first iteration, and the map did not change further after four iterations (Figure 3-3, first row). Supervised classification was employed to sort for HTL-bound particles and to sort out empty ribosomes. In the first round, the initial map and another copy of it where the HTL density was masked were used as reference maps, and the cross-correlations of each particle towards the two maps were compared. In this case, 15,790 particles aligned better to the HTL-bound ribosome and 14,970 particles aligned better to the vacant ribosome. The resulting maps were refined separately from these two pools. This yielded one map with strong density at the ribosomal tunnel exit while the other map displayed almost no density there (Figure 3-3, second row). These two maps were used as reference volumes for the HTL-bound pool containing 15,790 particles for a second round of supervised classification. 13,011 particles with HTL bound to the ribosome were obtained after this step, and this subpool of particles was used for the final refinement. To this end, the empty ribosome map (EMD-1045) was used as an input map to avoid model bias by the map generated from supervised classification. The refinement stopped after 6 iterations when the map was stable (Figure 3-3, third row).

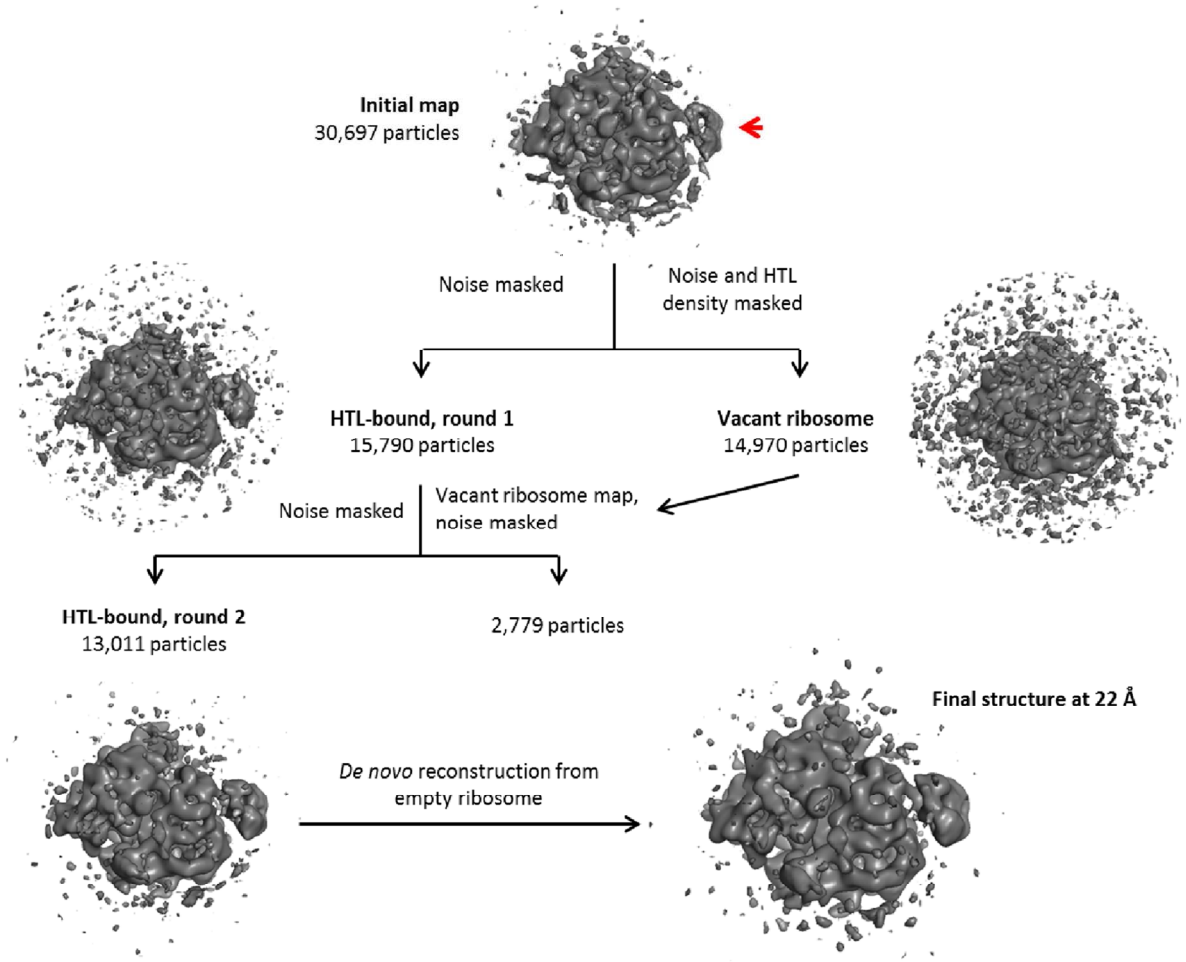


Figure 3-3: Sorting scheme of RNC^{FtsQ}-HTL^{DDM} dataset. First row: Initial map refined from all particles. The extra density which accounts for the HTL is marked with a red arrow. Second row: Maps resulting from the first round of supervised classification. Third row: the resulting HTL-bound map from the second round of classification (left). Particles from this map were used to reconstruct the final map (right).

3.2.3 Analysis of the RNC^{FtsQ}- HTL^{DDM} structure

The structure reached a resolution of 22 Å (with FSC 0.5 criterion) only. The EM structure showed a clear density at the ribosomal tunnel exit resolution (Figure 3-4). At this resolution it is not possible to interpret the HTL density in detail. This result demonstrated that some (in this case, 42%) of the complexes were preserved during the cryo-EM specimen preparation. Furthermore, compared to the cryo-EM structures available of SecYEG-ribosome complexes and Sec61-ribosome complexes (Mitra *et al.*, 2005; Ménétret *et al.*, 2007; Frauenfeld *et al.*, 2011), the

extra density in our map is larger and has a strong connection to the ribosome (Figure 3-4) which was not seen in the other maps, indicating the presence of extra components.

A more interpretable and higher-resolution structure may be obtained by collecting more images, using a direct electron detector for data collection and by refining in newer software packages. However, we decided to first improve the sample preparation protocol due to following reasons. First, a significant amount of aggregation was detected in the micrographs indicating a problem with the buffer conditions and/or the stability of the HTL. Second, the chosen strategy of using a large (25-fold) excess of HTL for grids preparation leads to a low number of ribosomal particles (~15) per micrograph due to the limited HTL concentration (DDM-solubilized HTL cannot be concentrated without complex dissociation). This means that with this type of grids we would need to collect ~10,000 micrographs in order to have a reasonable sized data set that allows multi-step/hierarchical sorting. This is impractical. Therefore, I had to either increase the concentration of HTL, or the binding of HTL to ribosomes has to be stabilized in order to avoid dissociation of complexes during grid preparation. Moreover, the sample quality would be improved if it would be subjected to a further step of purification. One strategy to solve these issues is the nanodiscs technology which allows to stabilize membrane protein complexes and to use detergent-free buffer conditions.

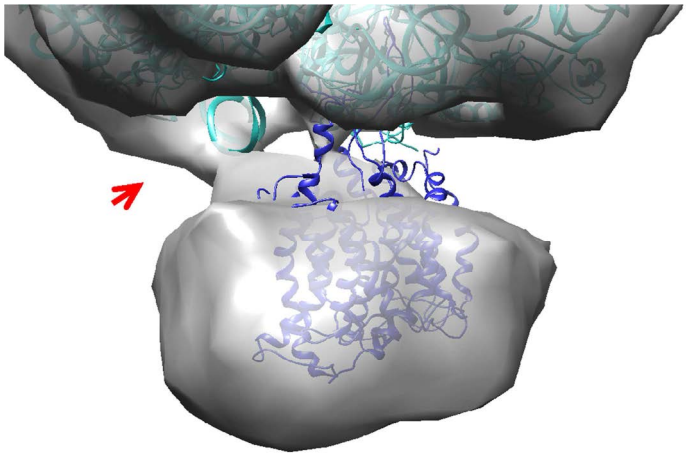


Figure 3-4: Detailed view of the RNC^{FtsQ}-HTL^{DDM} structure. Atomic model of the ribosome-SecYEG complex (Frauenfeld *et al.*, 2011) fitted into the cryo-EM structure of the RNC^{FtsQ}-HTL^{DDM} complex. The density of the HTL is larger than SecYEG alone, and an additional strong connection to the ribosome was identified (indicated by the red arrow).

3.3 Incorporating HTL into nanodiscs

3.3.1 Constructs and purification of the MSP proteins

As briefly introduced in Section 1.2.3, the molecular weight and corresponding length of the MSP proteins used for nanodisc reconstitution determines the radius of the reconstituted nanodisc. Therefore, the first question was to decide which MSP construct to use for incorporating HTL with its 34 transmembrane helices into a nanodisc.

Many derivatives of MSP proteins exist. The most widely used are MSP1D1 and MSP1E3D1, both developed by the Sligar Laboratory (Denisov *et al.*, 2004). The MSP1D1 (hereafter called MSP1) generates nanodiscs of ~9.7 nm in diameter, while the MSP1E3D1 (hereafter called MSP3) has three additional helices inserted to the MSP1 protein resulting in larger nanodiscs of ~12.1 nm in diameter (Figure 3-5). From negative stain EM images we estimated the diameter of the HTL between 10 and 15 nm. Duong and colleagues had reported the use of MSP1 to build nanodiscs containing only the monomer of SecYEG which has 15 transmembrane helices, and the MSP3 was used to generate nanodiscs that can incorporate a dimer of SecYEG (Dalal *et al.*, 2012). A cryo-EM structure of ribosome bound with SecYE was also published (Frauenfeld *et al.*, 2011) using a nanodisc reconstituted with the apo-A1 protein which is slightly longer than MSP1. Based on this cryo-EM structure, I estimated that it could be somewhat too small to accommodate the additional 19 transmembrane helices of the SecDF-YajC-YidC components of the HTL. Therefore, I used MSP3 protein for the initial nanodisc preparations.

After experiencing difficulties in reconstructing nanodiscs with HTL, a longer MSP protein was engineered based on MSP1 and MSP3. The purpose was to generate a nanodisc with ~15 nm in diameter which could accommodate the 34 transmembrane helices of HTL and have some additional space for lipids and allowing conformational changes in HTL which could occur during ribosome binding. The protein, named MSP1E6D1 (hereafter called MSP6) was designed such that the extra helix bundle of MSP3 compared to MSP1 (helix 4,5,6) was repeated (Figure 3-5. The construct pET28a_MSP1E6D1 was generated by substituting a fragment in MSP3 with a synthesized sequence, such that the three repeats of helices 4-5-6

have a different coding sequence to avoid homologous recombination. The substitution of the fragment was achieved by restriction enzyme digestion and ligation using the unique *EcoR*I restriction site in the helix bundle "123" and a *Sac*I site in the helix bundle "456". A complete enzymatic digestion followed by agarose gel purification ensured the retrieve of plasmid backbone with only one incomplete helix bundle "456" (Figure 3-5). The insert was gene synthesized, digested with *EcoR*I and *Sac*I before ligated to the backbone. The resulting MSP6 construct was confirmed by sequencing.

Expression and purification of MSP3 and MSP6 were performed as follows: Proteins were expressed in *E. coli* BL21 star (DE3) in 2-YT Broth with 30 µg/ml kanamycin. The cells were cultured at 37 °C until the OD_{600} reached 0.7, and then the temperature was lowered to 28 °C. Protein expression was induced by addition of 1 mM IPTG, 30 minutes after lowering the temperature. The cells were harvested 3 hours after induction by centrifugation, and were stored at -80 °C after flash-freezing in liquid nitrogen.

To purify the MSP proteins, the cells were resuspended in 20 mM sodium

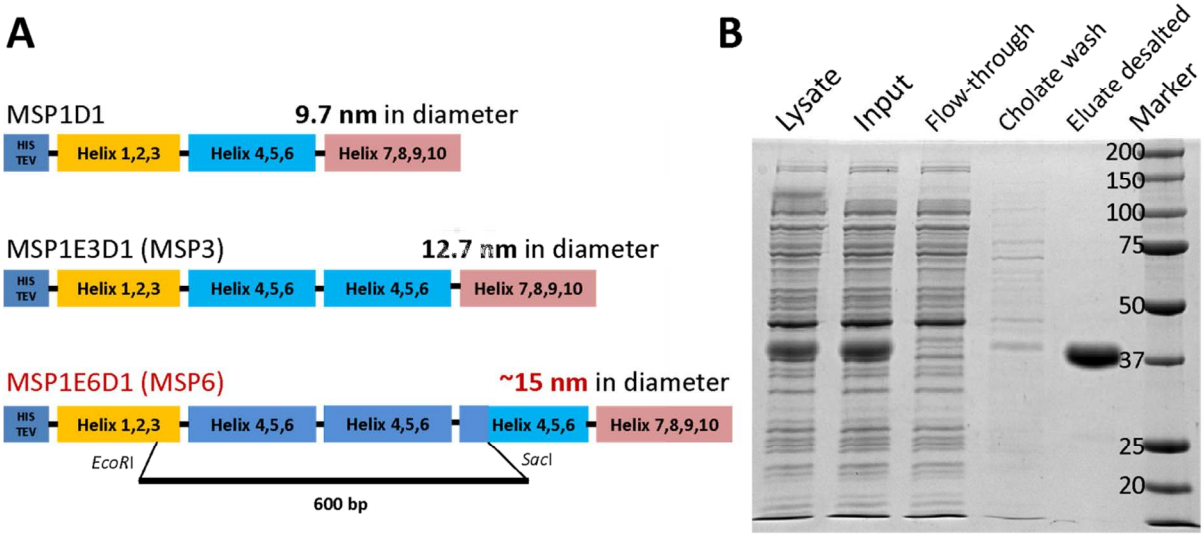


Figure 3-5: MSP protein constructs and purification of MSP6. (A) Schematic representation of MSP protein constructs. MSP proteins are composed of helix bundles. To engineer MSP6, synthesized DNA with one extra copy of helix bundle 4,5,6 was inserted in the gene encoding MSP3 by *EcoR*I/*Sac*I restriction enzyme digestion and ligation. (B) Coomassie-stained SDS-PAGE gel of samples from an expression and Ni-NTA affinity purification of MSP6 with a calculated molecular weight of 40 kD.

phosphate pH 7.4, 1% (v/v) Triton X-100 and protease inhibitors (Roche cOmplete EDTA-free). The cells were lysed by sonication (8 cycles of 15 s ON/30 s OFF, with amplitude 70). The cell lysate was cleared by centrifugation at 30,000 RCF in a JA-25.50 rotor for 30 minutes at 4°C. Subsequently, the supernatant was loaded onto a Histrap HP column equilibrated with 40 mM Tris-HCl pH 8.0, 300 mM NaCl, 1% Triton X-100. The column was washed with at least 5 column volumes of this buffer. Then, the Histrap column was washed with at least 10 column volumes of cholate buffer (40 mM Tris-HCl pH 8.0, 300 mM NaCl, 50 mM sodium cholate, 20 mM imidazole). The MSP proteins were subsequently eluted with 40 mM Tris-HCl pH 8.0, 300 mM NaCl, 500 mM imidazole, and desalted to 20 mM Hepes-KOH pH 8.0, 130 mM NaCl. Aliquots were fast-frozen and stored at -80 °C. An example of the affinity purification steps of the MSP6, visualized by Coomassie-stained SDS-PAGE, is shown in Figure 3-5.

3.3.2 Reconstructions of HTL-nanodiscs

The general idea of reconstructing nanodiscs was described in Section 1.2.3 and Figure 1-5. The critical factor to optimize during nanodisc reconstitution is the HTL:MSP:lipid ratio in the reaction mixture. The optimal ratio has to be explored for each sample and MSP protein. The ratio depends also on the buffer composition, on the concentration range of proteins, and to a large extent on the type of lipid or lipids used. Two problems rendered the nanodisc reconstitution experiments with HTL particularly difficult. First, the low concentration of purified HTL hampered efficient reconstitution. In fact, when the protein concentrations are lowered, the equilibrium of nanodisc formation may be changed and the optimal protein:MSP:lipid ratio may far away from the suggested starting condition. Second, the *E. coli* polar lipid extract (Avanti, Cat. 100600) I used for the experiments was a source of heterogeneity. This natural-sourced lipids mixture was chosen because it mimics better the natural environment of the HTL. In fact, many proteins involved in translocation are known to require acidic phospholipids (phosphatidylglycerol and cardiolipin) (de Vrije *et al.*, 1988; Lill, Dowhan and Wickner, 1990; Gold *et al.*, 2010). However, such natural lipid extract is also known to be heterogeneous and its physical properties are less characterized. Thus, most of the reported works of nanodiscs were carried out with

synthetic lipids, and the difficulty of generating nanodiscs with lipid extracts had been acknowledged (Roos *et al.*, 2012). Nonetheless, numerous conditions had been tested to optimize HTL reconstruction into nanodiscs. Here, I present the purification and reconstruction protocol for HTL-nanodisc which yielded the best results in my hands.

E. coli polar lipid extract (as lyophilized powder) was solubilized in 20 mM Hepes-KOH pH 7.5, 100 mM NaCl, 30 mg/ml sodium cholate, 1 mM dithiothreitol (DTT) to have a stock solution of 30 mM lipids in 70 mM cholate. HTL was purified following the protocol in Section 3.1.2 with the following modifications: The eluate from the Histrap HP column was desalted to 20 mM Hepes-KOH pH 7.5, 130 mM NaCl, 0.03% DDM, and was cleared by ultracentrifugation at 100,000 RCF in a TLA-55 rotor (Beckman Coulter) for 10 minutes at 4°C. The protein concentration was determined using an absorption coefficient of 250,000 at OD_{280}. HTL was then mixed with MSP6 and lipids to a final concentration of 2.1 µM, 8.4 µM and 420 µM respectively, corresponding to a molar ratio of 1:4:200. The mixture was rotated in the cold room for 1 hour, and was then mixed with equal amounts (1 g for 1 ml) of Bio-Beads SM-2 adsorbents (Bio-Rad). The removal of detergents was achieved overnight in the cold room with gentle agitation. The reaction mixture was separated from the settled Bio-Beads and subsequently was centrifuged at 100,000 RCF in a TLA-55 rotor for 10 minutes at 4°C. Typically, one-third of the HTL could be recovered as soluble protein according to the OD_{280} value. The HTL supernatant was mixed with calmodulin affinity resin, followed by the same procedure as described for the purification of DDM-solubilized HTL (Section 3.1.2), except that detergent was omitted from all buffers. Finally, the eluate fraction from the calmodulin resin was loaded onto a Superdex 200 PC 3.2/30 column (GE Healthcare) equilibrated with 20 mM Hepes-KOH pH 7.5, 200 mM NaCl, 3% glycerol, 0.5 mM EGTA for size-exclusion chromatography (Figure 3-6). This step served as an extra purification step and additionally demonstrated that the HTL was incorporated into nanodiscs since there was no more detergent present.

In the described procedure, the reconstruction of HTL-nanodisc was performed in-between two affinity purification steps. This was necessary because the reconstituted nanodiscs consisted of different species: nanodiscs with only lipids,

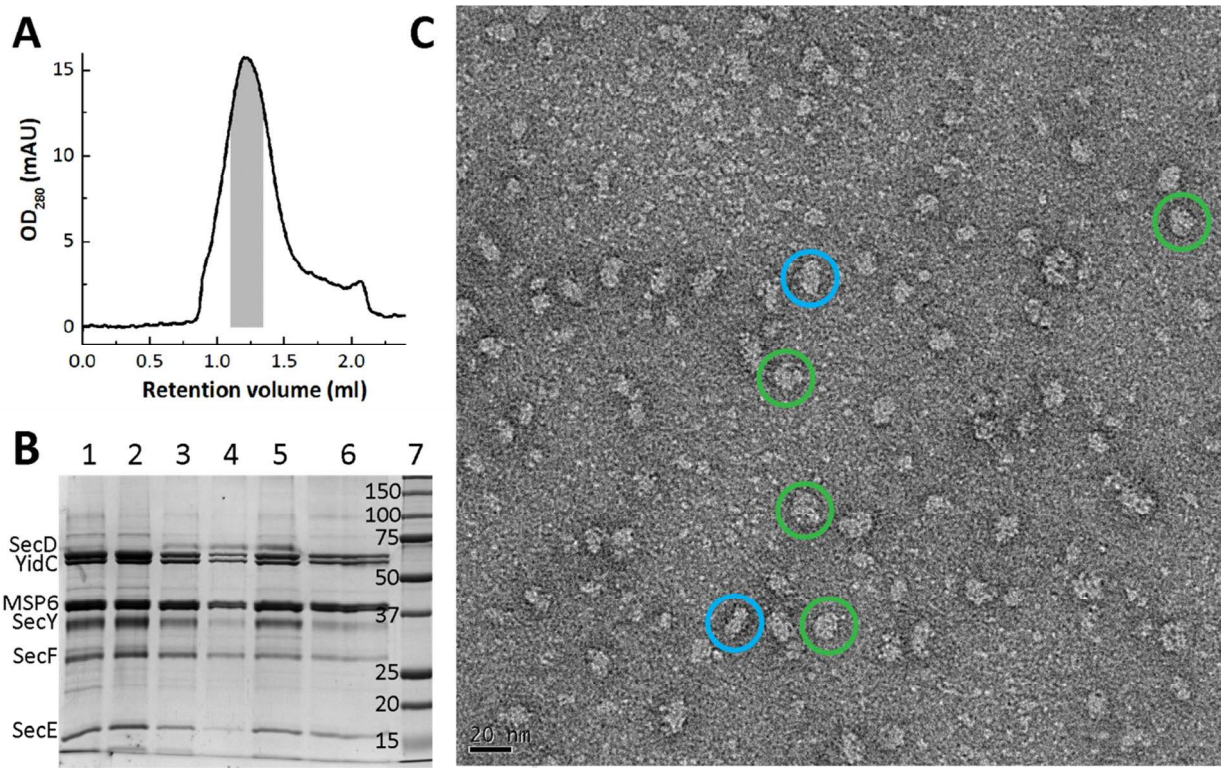


Figure 3-6: HTL^{nanodisc} purification and negative-stain EM image. (A) Size-exclusion chromatography (Superdex 200) of the HTL^{nanodisc}. The fraction used for EM is marked in grey. (B) Coomassie-stained SDS-PAGE gel of samples along the reconstruction of nanodisc. 1. Mixture before addition of Bio-Beads; 2. Pellet and 3. Supernatant fraction from the ultracentrifugation step after Bio-Beads treatment; 4. Flow-through and 5. Eluate fraction from the calmodulin affinity purification; 6. Peak fraction of the size-exclusion chromatography in panel A. (C) Negative-stain EM image of the same fraction showing top and side views of the HTL-nanodiscs (boxed in green and blue respectively).

nanodiscs with SecYEG, nanodiscs with SecDF-YajC-YidC and a small fraction of HTL-nanodiscs. All these species are similar in size, and thus the calmodulin affinity purification would be the best way to separate HTL-nanodiscs from empty and SecYEG-nanodiscs as the calmodulin-binding peptide tag is placed the YajC protein. Still we cannot exclude that the resulting nanodiscs consist of a mixture of SecDF-YajC-YidC and HTL. However, we decided that this would be acceptable because HTL has a higher affinity for the ribosome than the SecDF-YajC-YidC subcomplex. Thus, we expected to be able to enrich the HTL in complexes with the ribosome.

3.3.3 Electron microscopy

Peak fractions from the size-exclusion chromatography were pooled and were examined by negative stain EM. As shown in Figure 3-6, particles of ~15 nm in diameter could be found. The sample seems to be rather heterogeneous in shape and size. It is possible that the nanodiscs prepared from MSP6 may have a higher flexibility in their shapes given their area is ~2 times that of the original MSP1 nanodisc (resulting in ellipsoid shapes and bended rather than flat discs). Moreover, we observed a tendency of the HTL-nanodiscs to stick together.

Nevertheless, we decided to use this sample to prepare cryo-EM grids of ribosome bound with HTL-nanodisc. We reasoned that only functional HTL-nanodiscs would bind to the translating ribosome. Despite several attempts, however, cryo-EM grids of RNC^{FtsQ}-HTL-nanodisc were always found to have a very noisy background, likely due to the aggregation of HTL-nanodiscs which could be observed also in negative stain EM. I later discovered that pH and salt concentrations play a critical role in the monodispersity of the HTL samples (see Section 3.5.3). However, it was not possible to optimize these factors systematically with the nanodisc sample because of the low yields after each experiment. This approach has not been pursued further because the other two approaches (LMNG and amphipols) provided more promising results. However, it is still very interesting to pursue HTL-nanodisc preparations further, because having a lipid bilayer surrounding the HTL may provide extra information about HTL structure and function that the next two approaches cannot.

3.4 Second cryo-EM dataset of RNC-HTL using GraFix and LMNG

3.4.1 Rationale of using GraFix and LMNG

The cryo-EM work using detergent-solubilized HTL by other group members showed that HTL could be stabilized efficiently using GraFix (Botte, thesis, 2014). GraFix which stands for Gradient Fixation is a method to purify and stabilize macromolecular complexes (Stark, 2010). Samples are applied onto gradient with increasing concentration of chemical cross-linkers (0 - 0.15% glutaraldehyde in our

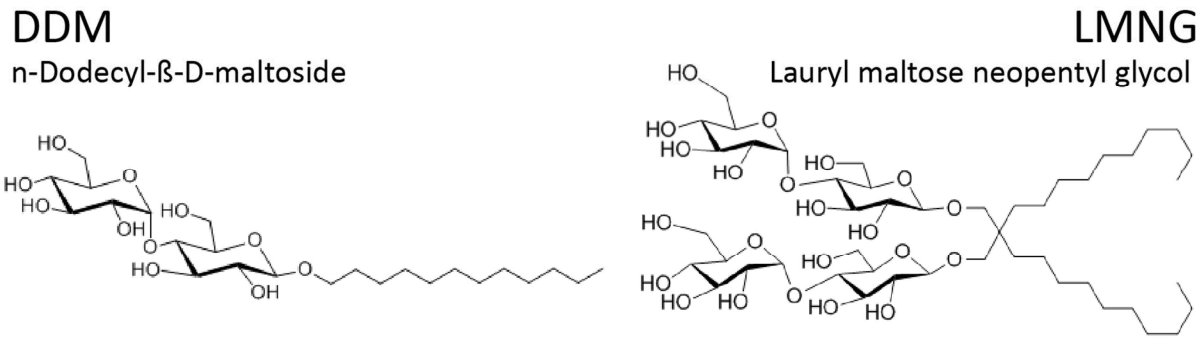


Figure 3-7: Molecular structures of DDM and LMNG.

case) and with an increasing density gradient. Thus, during ultracentrifugation the complexes are separated according to their size and exposed to the cross-linking agent. By following this protocol, the sample is diluted and exposed to a mild concentration of cross-linkers which both should minimize unspecific, inter-complex crosslinking leading to aggregation. The density gradient is adjusted such that the sample runs approximately into the middle of the gradient and can be efficiently separated from aggregates.

By applying GraFix to the RNC-HTL sample, it was expected that the binding of HTL to the RNC will be stabilized such that I can reduce the excess of HTL necessary for reconstruction of the complexes. This should lead to micrographs with less background noise, more particles per micrograph, and also a higher percentage of HTL-bound particles. Moreover, I expected that I could reduce the expected conformational variety of HTL, in particular that of the periplasmic domains of SecDF and YidC by cross-linking these domains to neighboring proteins. In fact, GraFix was shown to favor compact conformations of proteins.

In addition to GraFix, a new type of detergent was tested for HTL. Lauryl maltose neopentyl glycol (LMNG) is a newly developed detergent of the maltose-neopentyl glycol family. This type of detergent has a favorable behavior relative to conventional detergents (Chae *et al.*, 2010), most notably the extremely low critical micelle concentration (CMC; ~0.01 mM, or 0.001%, compared to DDM which is ~0.15 mM/0.008%). The high affinity of LMNG to membrane proteins can be explained by the fact that LMNG is virtually a dimer form of the DDM (Figure 3-7). LMNG is therefore expected to bind to membrane proteins very tightly such that proteins

remain to be soluble even when there is no more LMNG in the buffer (Chris Tate, personal communication). Therefore, LMNG appeared to be a good alternative to nanodiscs in order to obtain HTL in a detergent-free environment (*i.e*, no more detergent in the buffer, but only those bound to the membrane proteins). Moreover and importantly, MNG detergents were shown to stabilize membrane proteins (Chae *et al.*, 2010) which was a further incentive to switch from DDM to LMNG.

3.4.2 Sample preparation and data collection

The effect of LMNG and GraFix had to be tested step by step. First, I tested whether LMNG can be removed from buffers after the proteins are solubilized by it. I used SecYEG for this purpose as it is easier to produce (protocol described in Mitra *et al.*, 2005) and is much more stable than HTL. LMNG (Anatrace) was added to purified DDM-solubilized SecYEG (SecYEG^{DDM}), and the sample was further purified by size-exclusion chromatography (Superdex 200 10/300 GL) using LMNG-containing buffer first (20 mM Hepes-KOH pH 7.5, 100 mM KOAc, 6 mM $Mg(OAc)_2$, 0.01% LMNG).

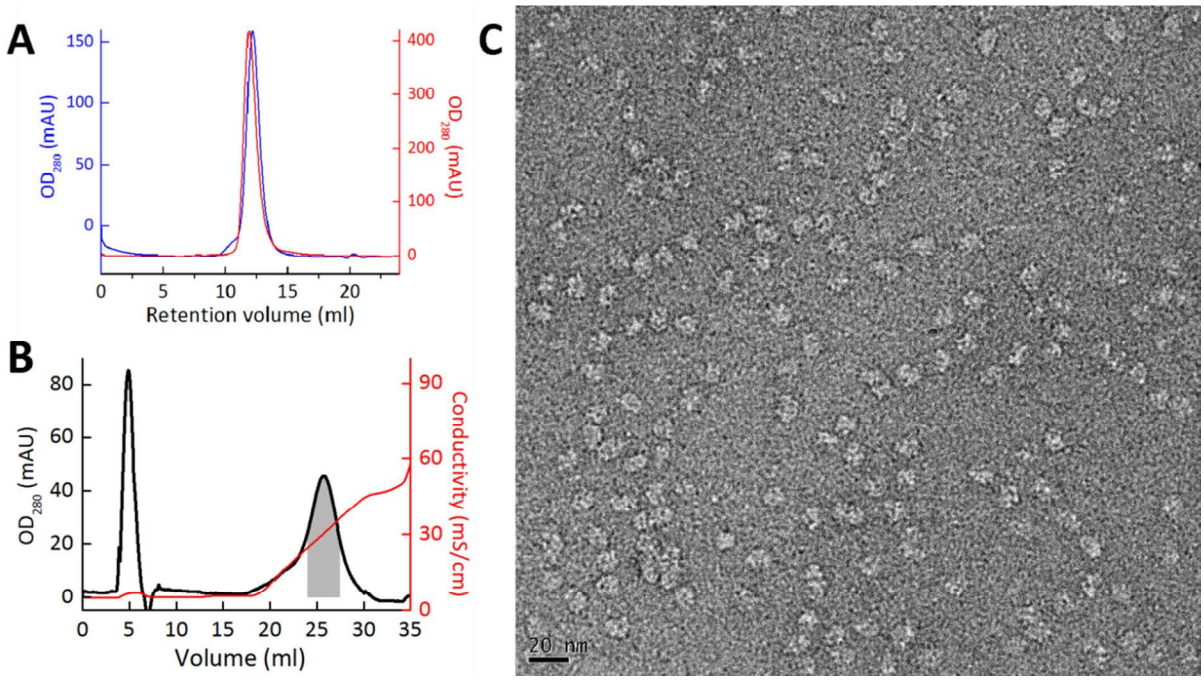


Figure 3-8: Purification of detergent-free SecYEG^{LMNG}. (A) Size-exclusion chromatography of the SecYEG^{DDM} (blue) and SecYEG^{LMNG} (red). (B) Detergent removal by ion-exchange chromatography as described in the main text. (C) Negative-stain EM image of the SecYEG^{LMNG} peak fraction marked in grey in panel B.

Re-chromatography of purified SecYEG samples showed that LMNG-solubilized SecYEG ($SecYEG^{LMNG}$) was homogenous and had a comparable size to $SecYEG^{DDM}$ (Figure 3-8). To assess the protein stabilities in detergent-free buffer, both samples were strongly diluted to reach a detergent concentration below their respective CMCs. After ultracentrifugation (10 minutes, 100,000 RCF in a TLA-55 rotor at 4°C), the remaining soluble protein was quantified by the Bradford method (Bio-Rad Protein Assay). The results indicated that the amount of soluble proteins was not affected for the $SecYEG^{LMNG}$ sample, but no soluble protein could be detected for $SecYEG^{DDM}$. Next, I tried to remove the excess LMNG from the buffer by absorbing $SecYEG^{LMNG}$ to an ion-exchange column (Hitrap SP HP, GE Healthcar) equilibrated in 50 mM sodium phosphate pH 7.9, 10% glycerol, 0.01% LMNG followed by washing and elution with detergent-free buffer and a NaCl gradient (Figure 3-8). Based on negative-stain EM images no aggregation was detected in this detergent-free $SecYEG^{LMNG}$ sample (Figure 3-8).

Next, I tried to assess the quality of HTL when solubilized by LMNG, and to test whether homogenous sample can be prepared by GraFix. First, I tested the exchange of DDM to LMNG during different steps of HTL purification. Protein precipitation was observed when purifying HTL^{LMNG} with calmodulin affinity resin. Therefore, the calmodulin affinity purification was performed as usual in DDM-containing buffers. The eluate from the calmodulin resin was desalted to 20 mM Hepes-KOH pH 8.0, 130 mM NaCl, 10 mM $Mg(OAc)_2$, 0.01% DDM, 0.03% LMNG, and was cleared by a ultracentrifugation before loading the protein onto the GraFix gradient. The GraFix gradient was produced from an upper buffer and a lower buffer mixed by a Gradient Master (Biocomp.) to create a liner gradient between the two buffers. In my experiment, the upper buffer comprised 20 mM Hepes-KOH pH 8.0, 130 mM NaCl, 10 mM $Mg(OAc)_2$, 10% glycerol, 0.01% LMNG, and the lower buffer was composed of 20 mM Hepes-KOH pH 8.0, 130 mM NaCl, 10 mM $Mg(OAc)_2$, 30% glycerol, 0.001% LMNG, 0.15% (w/v) glutaraldehyde. A control gradient was prepared without glutaraldehyde. I decided to not completely deplete the buffer from LMNG but kept a concentration of LMNG just at its CMC, as the fractions were used directly for electron microscopy without concentration steps. I speculated that the small amount of detergent could also help to obtain a random orientation of the

particles when adsorbing to the carbon foil on the grids. Same amount of HTL sample was loaded onto the two gradients. The gradients were centrifuged at 34,000 RPM for 18 hours in a SW 60 Ti rotor (Beckman Coulter). Fractions from the gradient without glutaraldehyde were analyzed by SDS-PAGE to track the migration of proteins in the gradient and to identify the best fraction for EM analysis (Figure 3-9). The corresponding fraction from the glutaraldehyde-containing gradient was then used to prepare negative-stain EM grids (Figure 3-9). To further assess the quality of the sample, 936 particles were manually picked from 50 micrographs to generate 2D class-averages (Figure 3-9). Based on the stoichiometry of the bands of the HTL subunits in the Coomassie-stained SDS-PAGE gel, on the rather homogenous particles in the negative stain micrograph, and on the quality of the 2D class-averages we decided to use HTL^{LMNG} for reconstitution of ribosomal complexes.

Finally, the RNC^{FtsQ}-HTL^{LMNG} complex was prepared following a similar procedure as described above: Purified HTL^{DDM} was buffer-exchanged into 20 mM Hepes-KOH pH 7.5, 100 mM NaCl, 10 mM $Mg(OAc)_2$, 0.01% DDM, 0.03% LMNG using a Hitrap Desalting column. The HTL^{LMNG} sample was mixed with RNC^{FtsQ} in a 10:1 molar ratio

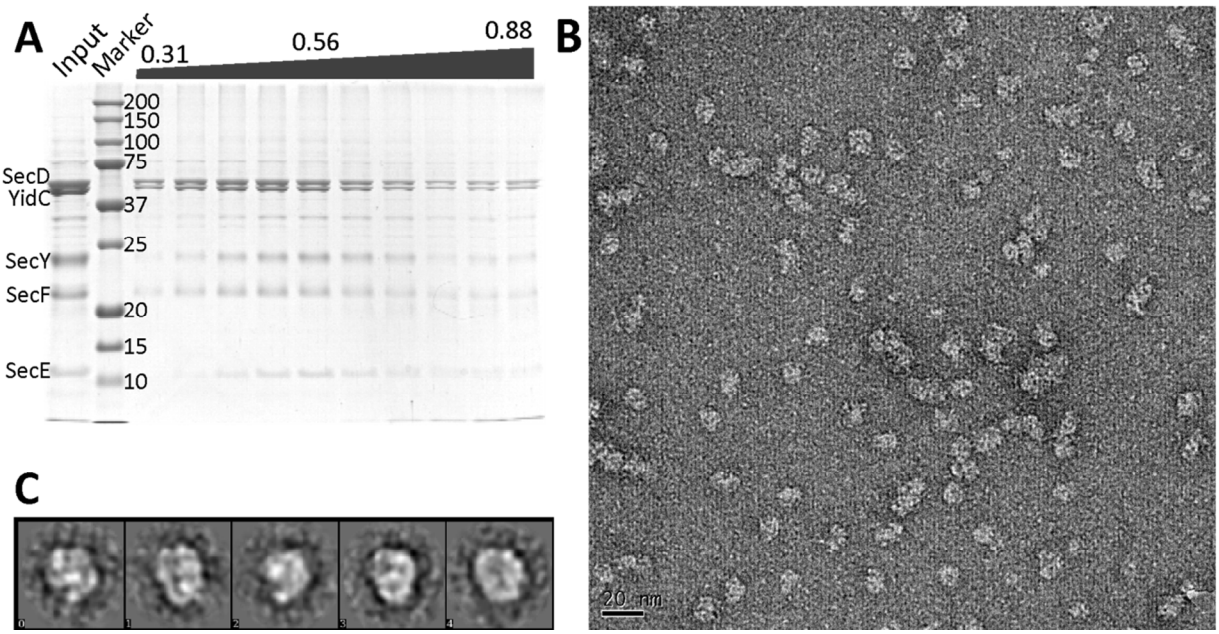


Figure 3-9: Preparation of HTL^{LMNG} by GraFix. (A) Coomassie-stained SDS-PAGE gel of the input and fractions of the GraFix experiment. Fractions are marked by their relative positions in the gradient. *i.e*, from 0 (no migration) to 1 (bottom of the tube). (B) A negative-stain EM micrograph of the fraction 0.56. (C) Selected 2D class-averages of crosslinked HTL^{LMNG} prepared as described in the main text.

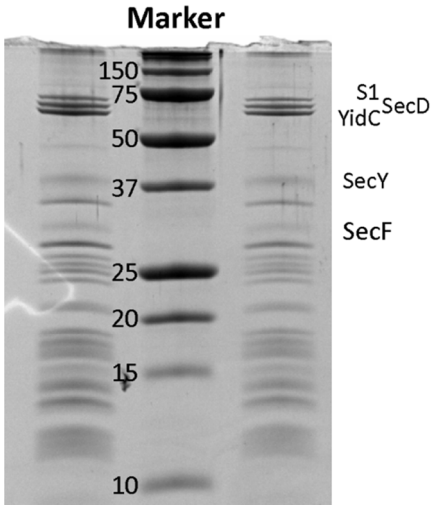


Figure 3-10: Fractions from RNC^{FtsQ}-HTL^{LMNG} gradient analyzed by Coomassie-stained SDS PAGE. The RNC^{FtsQ}-HTL^{LMNG} fractions from the non-crosslinking control gradient were loaded onto the SDS gel. The corresponding fractions of the crosslinking gradient (with glutaraldehyde) were used to prepare cryo-EM grids of RNC^{FtsQ}-HTL^{LMNG} and to collect cryo-EM data. S1 stands for ribosomal protein S1.

before loading the reaction mix onto the GraFix gradients. The composition of gradients and the parameters for centrifugation were adjusted as ribosomes are larger and denser. The upper buffer contained 20 mM Hepes-KOH pH 7.5, 100 mM NaCl, 10 mM $Mg(OAc)_2$, 10% (w/v) sucrose, 0.01% LMNG, and the lower buffer was 20 mM Hepes-KOH pH 7.5, 100 mM NaCl, 10 mM $Mg(OAc)_2$, 50% sucrose, 0.001% LMNG, 0.15% glutaraldehyde. Ultracentrifugation was carried out at 18,000 RPM for 15 hours in a SW 60 Ti rotor. Figure 3-10 shows two fractions which migrated through ~60% of the non-cross-linking gradient. Their corresponding cross-linked fractions were pooled and cryo-EM grids were prepared by incubating sample with Quantifoil grids coated with carbon films for one hour, followed by vitrification using a Vitrobot Mk5 (FEI) with 3 seconds blotting time and blotting force 5.

3.4.3 Data processing of the GraFix-RNC^{FtsQ}-HTL^{LMNG} complex

Data was collected on a Tecnai Polara (FEI) operated at 300 kV, with a 4k x4k CCD camera (US4000, Gatan) under low dose conditions at a calibrated pixel size of 1.93 Å. A total number of 4,150 micrographs were recorded using automated data acquisition software (FEI EPU). 168,219 particles were manually picked from 2,137 CTF-corrected micrographs. The data was initially processed in SPIDER for sorting of HTL-bound particles, a procedure described in Section 3.2.3. At the same time, new software packages including RELION, SPARX and Xmipp 3.0 were tested on this dataset, and the result from RELION was favored as it was less noisy and maps

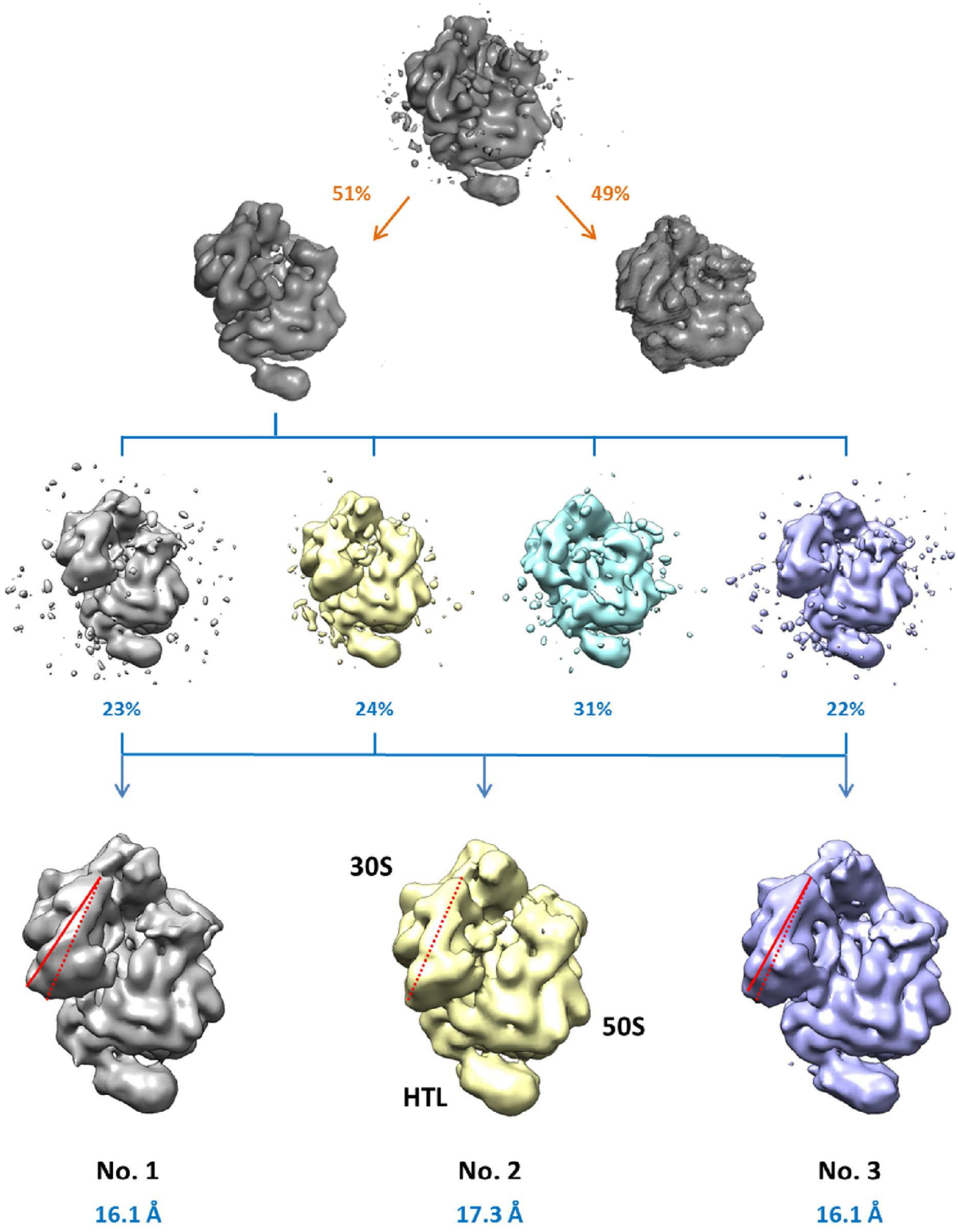


Figure 3-11: Sorting scheme of the GraFix RNC^{FtsQ}-HTL^{LMNG} dataset. See main text for detailed description. The different orientations of the 30S subunit in the three final maps (bottom row) are illustrated. Maps are aligned according the 50S subunit, and the red lines indicate the orientation of 30S in relation to volume No. 2.

had more features. Therefore, the 86,526 HTL-bound particles were processed further in RELION.

Four maps were generated from the first round of unsupervised classification (Figure 3-11). The most obvious difference among these maps was the movements of the small ribosomal subunit (30S subunit). One map showed an unusual 30S conformation which was possibly induced by glutaraldehyde crosslinking. The other three maps each represented a conformation of the 30S subunit that was well-documented by previous cryo-EM work (Fischer *et al.*, 2010). Particles from the first map were discarded, and the remaining particles were subjected to supervised classification using the other three maps as references. Finer angular sampling was used to allow more precise assignments of particles to the three conformations. Finally each subpool of particles (with 19,507; 22,382 and 18,566 particles respectively) was refined using RELION auto-refine (Figure 3-11).

3.4.5 Analysis of the RNC^{FtsQ}-HTL^{LMNG} EM structure

The three structures were refined to 16.1 Å, 17.3 Å and 16.1 Å respectively (Gold-standard refinements with FSC 0.143 criterion). Although volume 2 had a slightly lower resolution, it represents the conformation that is the most populated in non-cross-linked ribosomes (Fischer *et al.*, 2010). Moreover, the HTL density in this structure is slightly bigger compared to the other two volumes. Therefore, this structure was used for further analysis.

Crystal structures of the 70S ribosome (Protein Data Bank ID: 2AW4, 2AVY) (Schuwirth *et al.*, 2005) and a modelled structure of HTL (Botte, thesis, 2014) were fitted into the cryo-EM map (Figure 3-11). Dimensions of the density at the membrane plane agreed with the fitted quasi-atomic model of HTL, and interactions between HTL and ribosome can be identified for the first time. Positions of the connections from SecYEG to the ribosome and the placement of YidC were consistent with the published biochemical and structural data (Mitra *et al.*, 2005; Kohler *et al.*, 2009). This was an encouraging result, but the resolution hampered any further interpretation aiming for details. Although extra density is detected possibly corresponding to the periplasmic domains of HTL, this density was not large

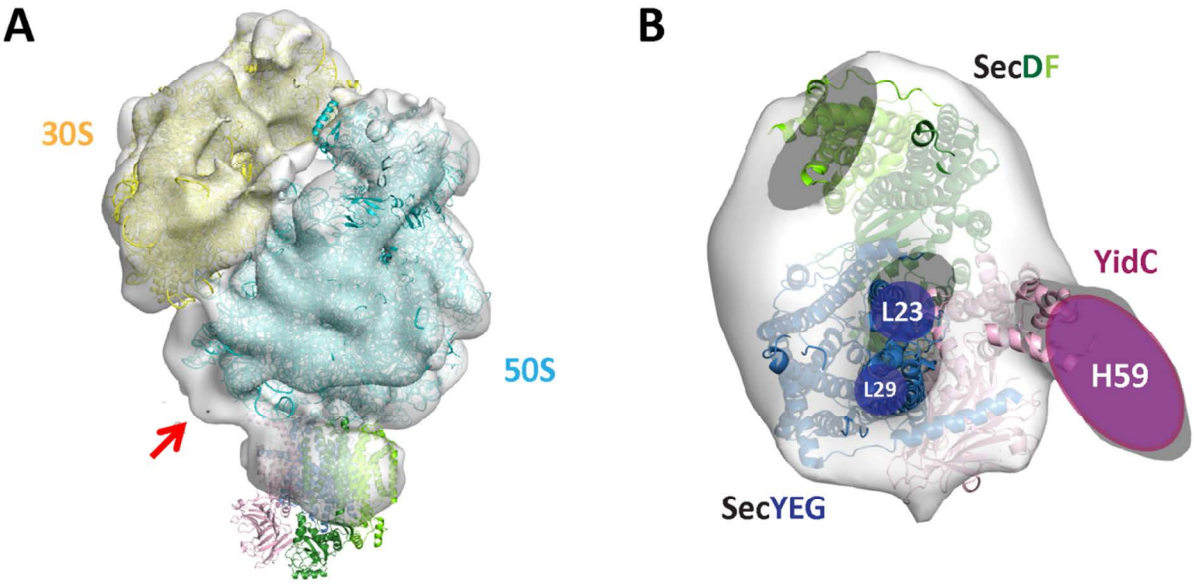


Figure 3-12: Fitting of crystal structures. (A) Crystal structures of large and small ribosomal subunits (PDB ID: 2AVY and 2AW4) and the components of HTL (as shown in Figure 1-2) were fitted into the cryo-EM structure of the GraFix-RNC^{FtsQ}-HTL^{LMNG}. The large extra density close to the helix 59 is marked by a red arrow. (B) Detailed view from the ribosomal tunnel exit to the HTL. The ribosomal proteins and rRNA helices involved in connections to the HTL are marked.

enough to accommodate the crystal structures of periplasmic domains of SecDF and YidC. Moreover, a large EM density was detected next to rRNA Helix 59 of the large ribosomal subunit (Figure 3-12) which has not been observed before. We speculate that the periplasmic domain of YidC was crosslinked to the Helix 59. If this is correct this is an artifact since in the cell the periplasmic domain is separated by the membrane from the ribosome. Initially, cross-linking was employed in order to reduce the conformational diversity of the complex, but in fact, it fixed the 30S subunit in several distinct conformations that had to be sorted for. This largely reduced the amount of particles in each particle pool, and additionally it may have introduced artifacts. We concluded that cross-linking should not be used for further experiments. Other means have to be found to reduce the conformational dynamics in the HTL complex.

3.5 Towards ribosome-HTL amphipol complexes

3.5.1 Initial experiments with amphipol A8-35

The search for detergent-free HTL samples continued with the application of amphipols. The use of amphipols has become popular in recent years, especially after several publications of high-resolution EM structures that benefited from the use of amphipols (Liao *et al.*, 2013; Lu *et al.*, 2014). In addition to being able to manipulate samples in detergent-free buffers, the polymer nature of amphipols is particular interesting to our research as the complex should be stabilized once the HTL is trapped in amphipols.

Amphipol A8-35 (Anatrace) (Figure 3-13) was used for initial experiments to test its ability to solubilize HTL and the properties of HTL in amphipols (HTL^{A8-35}). A8-35 was solubilized in H_2O and incubated on ice for several hours to ensure a good rehydration as recommended (Zoonens and Popot, 2014). Purified HTL^{DDM} was buffer-exchanged into 20 mM Hepes-KOH pH 7.5, 130 mM KOAc, 0.5 mM EGTA, 0.03% DDM. Different amounts of A8-35 were added to HTL^{DDM} to have final mass ratios of A8-35/HTL of 0:1, 2.5:1, 4:1 and 6:1 respectively. After 30 minutes of

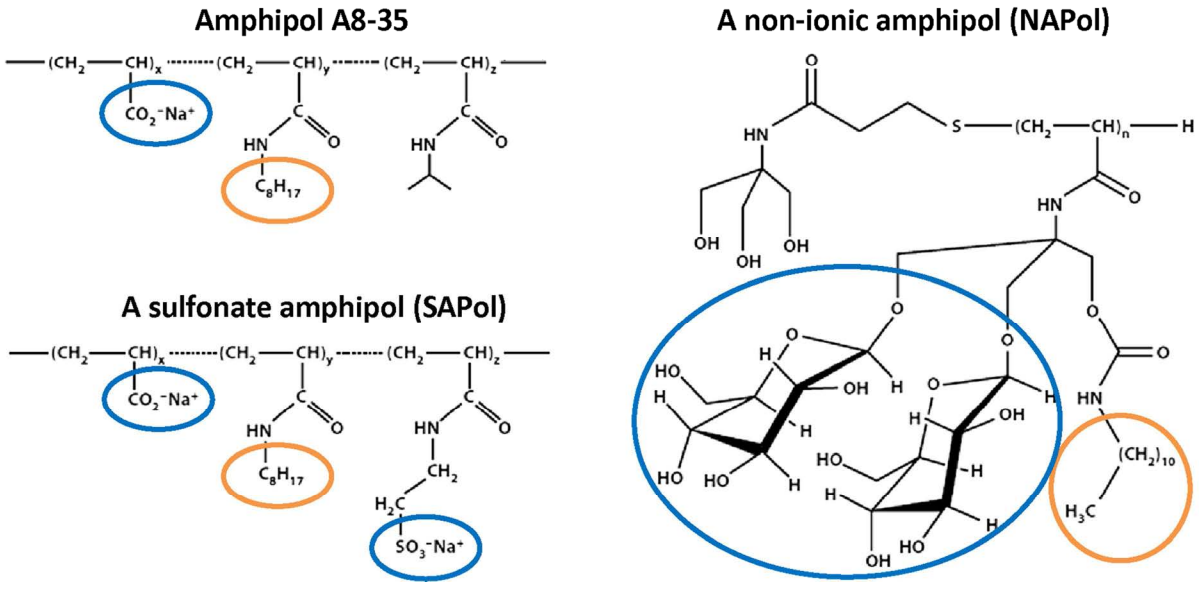


Figure 3-13: Different types of amphipols relevant to this work. Molecular structures of the amphipol A8-35, the SAPol and the NAPol. The chemical groups that provide aqueous solubility (hydrophilic groups) are marked in blue, and the chemical groups that bind to the transmembrane helices of the membrane proteins (hydrophobic groups) are marked in orange. Adapted from Zoonens and Popot, 2014.

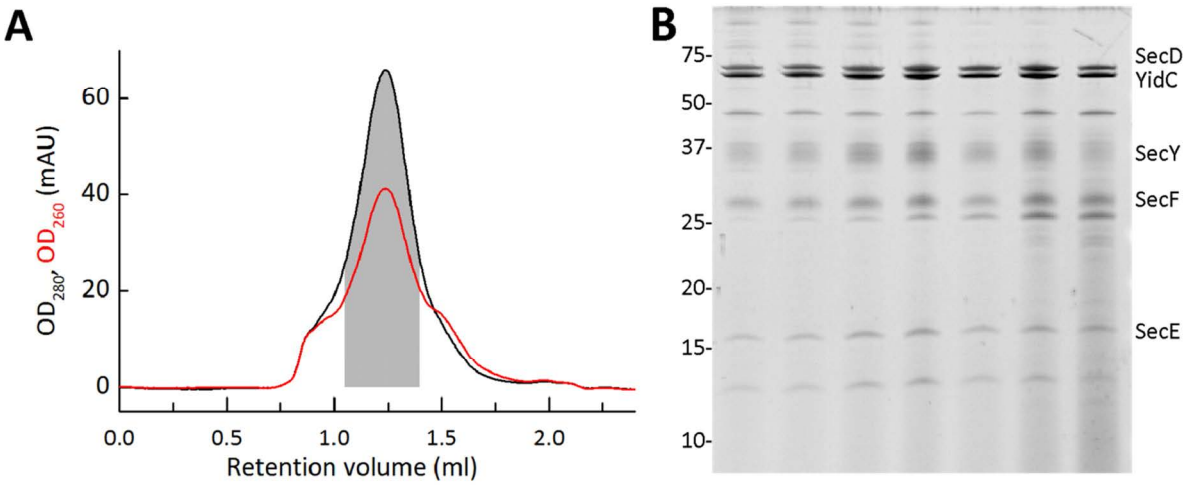


Figure 3-14: Purification of the HTL^{A8-35}. (A) Size-exclusion chromatography of the HTL^{A8-35}. (B) SYPRO Ruby (Life Technologies) stained SDS-PAGE gel of the fractions marked in grey in the panel A. Note that the stoichiometries of the subunits are constant across the fractions.

incubation on ice, Bio-Beads were added to the mixture (~ 20 mg Bio-Beads per 1 mg DDM; DDM bound to HTL was also taken to consideration assuming 1 mg DDM per 1 mg HTL). The reaction mixtures were rotated overnight in the cold room. The Bio-Beads-treated samples were ultra-centrifuged at 100,000 RCF for 10 minutes in a TLA-55 rotor. The OD_{280} was determined and compared to the value before Bio-Beads treatment. Almost no protein was detected on the 0:1 sample in which no amphipol was added. In contrast, 66%, 80% and 83% of protein sample was recovered from the 2.5:1, 4:1 and 6:1 samples respectively. This means that A8-35 indeed maintained HTL soluble in a detergent-free buffer, and indicates that a 4:1 mass ratio of A8-35 and HTL was sufficient for solubilization. HTL^{A8-35} from the 4:1 sample was analyzed by size-exclusion chromatography (Superdex 200 PC 3.2/30, buffer: 20 mM Hepes-KOH pH 7.0, 130 mM NaCl, 3% glycerol) and showed a good homogeneity with an expected elution volume (calculated molecular weight of 275 kD; the HTL is a 240 kD complex) (Figure 3-14).

3.5.2 Ribosome binding experiments with HTL^{SAPol} and HTL^{NAPol}

Although the results from HTL^{A8-35} were very encouraging, it was known that A8-35, as the most canonical amphipol, aggregates at the presence of divalent cations (Picard *et al.*, 2006). That prevented further experiments with ribosomes which

require magnesium ions in the buffer for their stability. The basis of the divalent ion-induced aggregation of A8-35 is that the carboxylate groups of the polymer chelate divalent cations. To overcome this problem, several non-commercial deviates have been developed, notably the sulfonate amphipols (SAPols) and the non-ionic amphipols (NAPols) (Figure 3-13). These amphipols were obtained from Manuela Zoonens from Institut de Biologie Physico-Chimique, Paris.

The SAPol was used first for HTL solubilization because compared to the A8-35 amphipol only the carboxylate groups were changed in the SAPols. The same 4:1 mass ratio of amphipol/HTL was applied as the average molecular mass and the molecular weight per hydrophobic chain of SAPol is similar to that of A8-35. Applying the same protocol for the exchange of DDM to amphipols, HTL^{SAPol} complexes can be purified in similar quality and yields as described for the HTL^{A8-35}. Notably, the size-exclusion chromatography of HTL^{SAPol} could be carried out in the presence of 10 mM $Mg(OAc)_2$.

Next, I tested the binding of HTL^{SAPol} to the ribosome. Ribosome binding was usually verified by ribosome co-sedimentation experiments: Ribosome and HTL were mixed in different ratios, incubated for 30 minutes on ice and subsequently were loaded onto a sucrose cushion (20 mM Hepes-KOH pH 8.0, 130 mM KOAc, 10 mM $Mg(OAc)_2$, 0.5 M sucrose). After ultracentrifugation (55,000 RPM, 3 hours in a TLA-55 rotor) ribosomes and ribosomal complexes should be found in the pellet fraction. HTL binding to ribosomes or ribosomal subunits can thus be detected by the

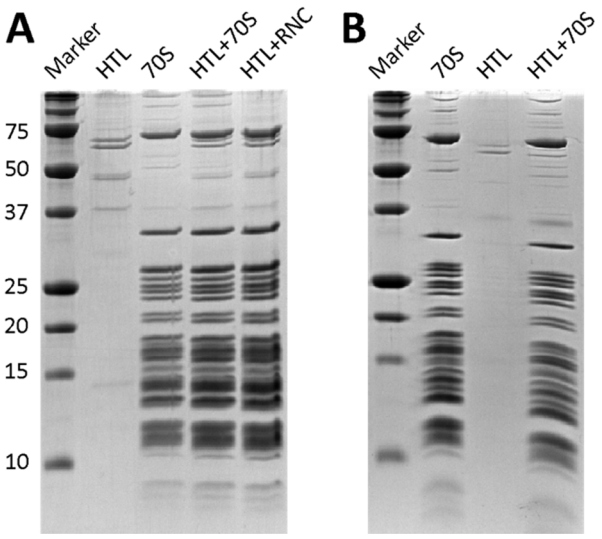


Figure 3-15: Ribosome co-sedimentation experiments. (A) Coomassie stained SDS-PAGE gel of the pellets from the co-sedimentation assay using HTL^{SAPol}. Similar amount of HTL appeared in the pellet when co-sedimented with ribosomes (70S or RNC) as in the HTL control, indicating no ribosome binding of HTL^{SAPol}. (B) Same experiment with HTL^{NAPol}.

presence of HTL in the ribosomal pellets. This method has been used to confirm the binding of HTL^{DDM}, HTL^{LMNG} and $HTL^{nanodisc}$ to ribosomes. Surprisingly, repeated experiments showed no sign of binding of HTL^{SAPol} to either non-translating ribosome or RNC^{FtsQ} (Figure 3-15). At that time, we suspected that the numerous negative charges on SAPol may interfere with the binding to the ribosome. In fact, positively charged cytoplasmic loops of SecY (Van den Berg *et al.*, 2004) and the positively charged C-terminus of YidC (Kohler *et al.*, 2009) are known to be responsible for ribosome binding. In HTL^{SAPol} complexes the positively charged, flexible cytoplasmic loops may interact with the SAPols rather than the ribosome.

Hoped to overcome this problem, one type of NAPols were used for HTL solubilization. The solubility of NAPols is provided by glucose-based groups and hence the molecule is non-ionic (Figure 3-12). HTL^{NAPol} was purified following a similar procedure. Its binding to the ribosome was tested by the co-sedimentation experiment, and surprisingly again, ribosome binding of HTL^{NAPol} could not be detected (Figure 3-15). The question was whether amphipols interfere with HTL binding to the ribosome or whether the ribosome-HTL complex does not survive the co-sedimentation assay when solubilized in amphipols. The sucrose cushion permits/rejects particles according to their densities. Amphipols were reported to form loose belts surrounding the solubilized proteins and may therefore reduce the overall density of the particle. It might be that the $HTL^{amphipols}$ has a strong tendency to float on sucrose cushions, causing $HTL^{amphipol}$ dissociation from ribosomes.

To test this hypothesis, new experiments had to be designed to identify binding of $HTL^{SAPol/NAPol}$ to ribosomes. To this end, we decided to use size-exclusion chromatography. Sample was loaded onto a Superdex 200 PC 3.2/30 column equilibrated in 20 mM Hepes-KOH pH 7.0, 130 mM KOAc, 10 mM $Mg(OAc)_2$. After gel filtration, ribosomes should appear in the void-volume, and any bound HTL should be detected in those fractions as well. The first chromatography experiment using 70 µg ribosomes (corresponded to 500 nM) confirmed that 70S ribosomes elute in the void volume, and that there is no HTL contamination in it. Second, gel filtration of 35 µg HTL^{SAPol} showed no aggregation of the sample and hence no presence of HTL proteins in the void-volume. Finally, a mixture of 35 µg HTL^{SAPol} and 70 µg ribosome (molar ratio of 5:1) was subjected to gel filtration. In this case

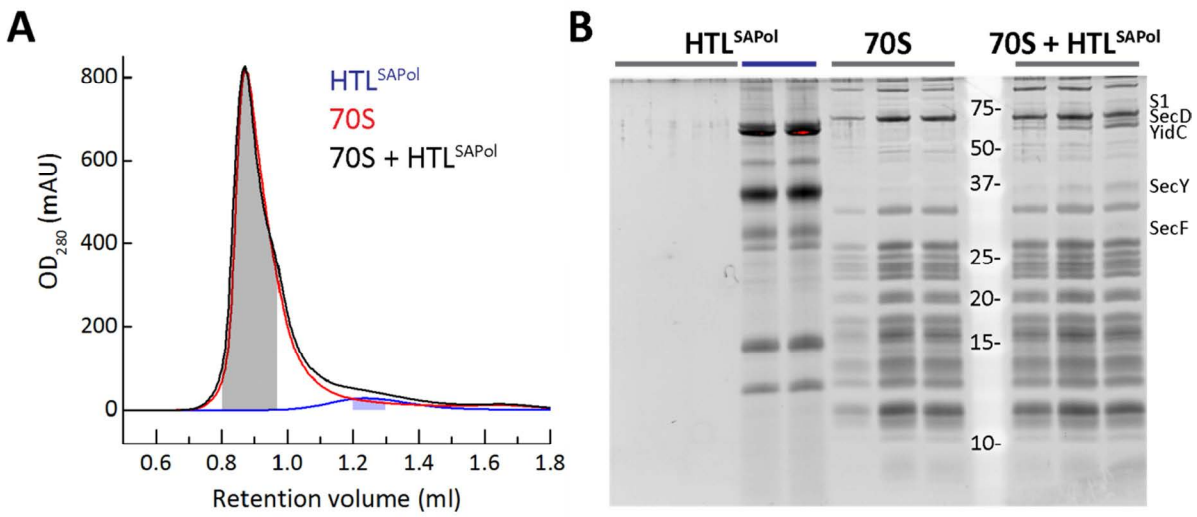


Figure 3-16: HTL^{SAPol}-ribosome binding analyzed by size-exclusion chromatography. (A) Elution profiles after gel filtration of 70S ribosomes, HTL^{SAPols} and 70S-HTL^{SAPol} samples. (B) SYPRO Ruby-stained SDS-PAGE gel of the fractions marked in panel A. See main text for details.

components of HTL can be clearly detected in the void-volume fraction, together with the ribosome (Figure 3-16). A slight shift of the void-volume peak towards higher molecular weight was also observed, with a shoulder at lower-molecular-weight side which may formed by dissociated HTL^{SAPol} particles during the chromatography experiment. The ribosome-HTL^{SAPol} complex was not stoichiometric as evaluated from the SDS-PAGE gel (Figure 3-16), but this can be explained by the fact that proteins are diluted during the gel filtration experiment and the disassociation of weak complexes cannot be reversed (*i.e.* this is a non-equilibrium experiment). Similar results were obtained using HTL^{NAPol}. We conclude that HTL$^{SAPol/NAPol}$ binds to ribosomes, however we still need to determine the affinity (for example by using thermophoresis) to find out whether the K_d is higher in the presence of amphipols. In conclusion, I am confident that it is possible to use this HTL$^{SAPol/NAPol}$-ribosome complex for cryo-EM.

3.5.3 Optimizing HTL^{NAPol} production

Because of the fact that the chemical structures of NAPols and SAPols are quite different, the optimal ratio of NAPol and HTL during the detergent exchange had to

be re-established. Mass ratios of 2:1, 4:1 and 8:1 were tested. Surprisingly, similar amounts of proteins were recovered after Bio-Beads treatment. After size-exclusion chromatography (Superdex 200 PC3.2/30, buffer: 20 mM Hepes-KOH pH 7.0, 130 mM NaCl, 3% glycerol), peaks of similar height were observed for the three samples at the expected elution volume. However, a large amount of protein elutes in the void volume. As a control, $HTL^{A8\text{-}35}$ that was purified at the same time with the same input HTL^{DDM} sample and procedure has shown good quality (Figure 3-17).

I found that the pH and salt concentrations during the detergent exchange step have a strong influence on the quality of the HTL^{NAPol} sample: In an experiment, purified HTL^{DDM} was buffer-exchanged into three different buffers comprising **(1)** 20 mM Hepes-KOH pH **7.0**, **300** mM NaCl, 0.01% DDM, **(2)** 20 mM Hepes-KOH pH **8.0, 130** mM NaCl, 0.01% DDM, and **(3)** 20 mM Hepes-KOH pH **8.0, 300** mM NaCl, 0.01% DDM. The same amount of NAPols was added in a mass ratio of 6:1 followed

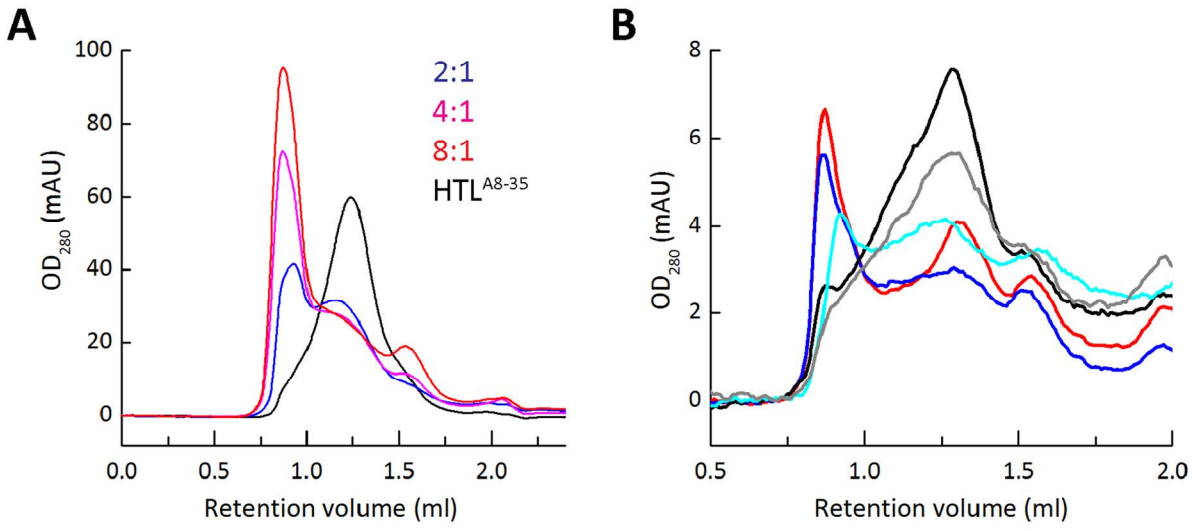


Figure 3-17: Optimization of the HTL^{NAPol} production. (A) Size-exclusion chromatography of $HTL^{A8\text{-}35}$ and HTL^{NAPol} complexes with different NAPol:HTL mass ratios, prepared from the same HTL sample following the same protocol. (B) Size-exclusion chromatography of the HTL^{NAPol} complexes. Black: HTL^{NAPol} prepared and chromatographed in high salt, high pH buffer (condition 3); Blue: HTL^{NAPol} prepared and chromatographed in low salt, high pH buffer (condition 2); Red: HTL^{NAPol} prepared and chromatographed in high salt, low pH buffer (condition 1); Cyan: HTL^{NAPol} prepared in low salt, high pH buffer (condition2) and chromatographed in high salt, high pH buffer (condition 3); Grey: HTL^{NAPol} prepared in high salt, high pH buffer (condition 3)and chromatographed in low salt, low pH buffer. See main text for more detail.

by Bio-Beads treatment. After ultracentrifugation, similar OD_{280} values were recorded from the three samples, but different elution profiles were obtained after size-exclusion chromatography using the corresponding buffers (with no DDM) (Figure 3-17). This suggested that either lower pH or lower salt concentration could cause the oligomerization of the HTL^{NAPol}. The fact that these larger particles could not be removed by the ultracentrifugation step indicated that these do not correspond to large aggregates.

Importantly, sample prepared in pH 8.0 and 300 mM NaCl maintained its solubility when chromatographed in low pH and low salt buffer (20 mM Hepes-KOH pH 7.0, 130 mM NaCl, 10 mM $Mg(OAC)_2$). However, sample prepared in pH 8.0 and 130 mM NaCl did not improve significantly when chromatographed afterwards in the presence of 300 mM NaCl, although the void-volume peak was diminishing in this case and a higher signal appeared at longer retention volumes. This implies a very slow disassociation of the oligomers (Figure 3-17).

Notably, the same HTL^{DDM} sample solubilized with A8-35 or SAPol was much more homogenous even in low pH and low salt buffer conditions. This may be explained by the fact that the HTL oligomers are likely to be formed by electrostatic forces and therefore could be disrupted by higher pH and higher salt concentration. Alternatively, the polyanionic charges of the A8-35/SAPols could help to dissolve the HTL oligomeric species. I speculate that once trapped by NAPol in low pH/ low salt, the oligomers were stabilized and therefore could not be reversed easily by higher pH/salt. Practically, the preparation of HTL^{NAPol} can be carried out in pH 8.0 with 300 mM NaCl during the detergent removal step. Formed HTL^{NAPol} can then be further purified by size-exclusion chromatography and be used for experiments in buffers with lower pH and/or salt concentration.

In conclusion, it was possible to improve the procedure of producing HTL^{NAPol} complexes. These RNC^{FtsQ}-HTL^{NAPol} complexes can now be analyzed cryo-EM and for other biophysical techniques (*e.g.* small angle X-ray scattering).

CHAPTER 4: DISCUSSION AND CONCLUDING REMARKS

CHAPITRE 4: DISCUSSION ET CONCLUSIONS

Résumé

Deux des principaux problèmes techniques que j'ai rencontré lors de mon travail de thèse sont abordés dans ce chapitre: les lipides associés lors de la purification de HTL et la précision de l'affectation angulaire dans le traitement des données cryo-EM.

En combinant les connaissances obtenues à partir de mon travail de thèse avec des résultats existants, je présente des modèles pour l'adressage co-traductionnelle médié par SRP ainsi que des modèles pour la translocation et le repliement de protéine membranaire médié par HTL.

Nous discuterons des différents complexes qui peuvent se former avec le ribosome servant à l'adressage et la translocation co-traductionelle. Ces complexes comprennent la conformation «activé» du complexe ribosome-SRP-FtsY; les complexes ribosome-SRP-FtsY avec différentes séquences de peptide signal; le complexe ribosome-HTL extrait directement à partir de la membrane cellulaire; et le complexe ribosome-HTL étudié *in situ* par tomographie cryo-électronique et sous-tomogramme moyenne. En outre, le repliement des protéines membranaires pourrait être étudié par la préparation de ribosomes en cours de traduction affichant plusieurs hélices transmembranaires. Pour finir, le complexe HTL-SecA lors de la translocation post-traductionnelle ainsi que l'adressage et la translocation co-traductionnelle chez les eucaryotes sont discutés.

4.1 Procedures and methods

4.1.1 Lipids and membrane protein purifications

Membrane proteins are naturally embedded in the lipid bilayers, and thus the interaction between lipids and membrane proteins can be very tight and specific. Many membrane proteins are known to require certain lipids for their activity and/or structural stability. However, membrane proteins are usually prepared by solubilizing and isolating them from the bilayers using detergents. This procedure does not ensure that the protein-lipids interactions are preserved. Therefore, particular attention has to be paid to the so-called “mildness” of the detergents. For example, non-ionic detergents such as DDM have been wildly used (Seddon, Curnow and Booth, 2004). As shown in Figure 1-3, the solubilization of membrane proteins by detergents is a continuous process. Under mild conditions, lipids can co-exist with detergents and proteins in a water-soluble form — the mixed micelle. The co-purification of lipids with the HTL complex was expected and was shown by small angle neutron scattering (unpublished work in the team). As discussed in Section 3.3.2, translocons are known to require acidic phospholipids (de Vrije *et al.*, 1988; Lill, Dowhan and Wickner, 1990; Gold *et al.*, 2010). The disassembly of HTL into sub-complexes with increasing concentrations of DDM has been shown, as well as an increase in HTL-stimulation of SecA ATPase in the presence of cardiolipin (Schulze *et al.*, 2014), indicating that lipids are important for integrity and activity of the HTL.

On the other hand, solubilizing membrane proteins under mild conditions also introduces the risk of inefficiently solubilizing the membrane proteins and obtaining very heterogeneous samples. In fact, it is very difficult to assess the type and amount of lipids bound to the membrane proteins and the role of these lipids. Lipids cannot be detected by traditional SDS-PAGE and therefore their existence in the sample is often ignored. Therefore, it is sometimes difficult to find a good balance between preserving the lipid-protein interactions and obtaining an efficiently solubilized sample for structural studies. Although the current purification procedures of HTL worked nicely to a large extent, it is still interesting to see if further improvement can be made with the use of another detergent such as octyl glucoside

or digitonin, both have been successfully applied in cryo-EM structural determination of membrane proteins.

4.1.2 Angular assignments and cryo-EM data processing

During the processing of single-particle cryo-EM data, the relative angles of the collected images have to be determined, for instance by using a defined data collection geometry as in the random-conical reconstruction approach (Radermacher *et al*., 1987) or by mathematical approaches, *i.e.* matching the images to a set of projections of a known 3D structure. In order to take full advantage of the high-resolution information contained in the images the projection matching has to be iterated and refined using more projections (angular refinement). A wrong angular information may not be easily noticed, partly because an EM structure is from the averaging of hundreds of thousands of images and therefore the contribution from an individual image cannot be easily assessed, and also because the resolution has long been the criterion to judge an EM map, while the importance of accurate angular assignments has been under-estimated. As briefly discussed in Section 1.4.2, the angular assignment of an image plays a crucial role during the reconstruction of a 3D model such that if the assignment is wrong, then the image is more likely to add noise to the model than to contribute “correct” information. One approach to overcome this problem is the use of maximum-likelihood algorithm. Instead of back-projecting images using their best-aligned orientations as in the cross-correlation based methods, each image is back-projected with all possible orientations, but its contribution in each orientation is corrected by its likelihood, a so-called probability-weighted averaging (Sigworth, *et al*., 2010). With this approach, noisy images will not contribute to the model because they can be aligned to many several projections. This approach is superior especially in the case of datasets at low signal to noise ratios.

Particular problems exist, for example in the experiments described in Section 2.3.2 and in Figure 2-2. In this case we knew the difference among the particles and wanted to sort for, namely the presence or absence of SRP. However, the supervised classification was led by other criteria. One explanation may be that the

probabilities of images are calculated globally so that even in the case where the difference is obvious in real space, it may not be the most significant difference, especially when calculations are performed in the reciprocal space. One thought to overcome this problem is that the angular assignment and the pixel contents of a given image can be treated separately. As in the case in Figure 2-2, the angular assignment of an image could be acquired by a global alignment to the reference, but then only the pixels at the known position of the ligand should be evaluated and used to determine whether the image corresponds to a ligand-bound ribosome or to a vacant ribosome.

In general, the information which is critical for angular assignment is different from the information critical for back-projection and hence model building: Low- to medium- frequencies are more useful for angular determination (Henderson *et al.*, 2011), while higher frequencies which contain more detailed information are needed for the reconstructions at higher resolution. Certain applications can take advantage by handling angular assignments and back projection separately, for example by using only low frequencies for angular assignment, a practice proven to be useful to prevent over-fitting of the cryo-EM structure (Henderson *et al.*, 2011). Furthermore, one may even use different images for angular assignment and for back projection. For example, a high-dose exposure after recording of a series of frames of a micrograph has become a popular strategy. These high-dose exposed images contain particles that are radiation-damaged, but they have a higher contrast and signal-to-noise ratio. In fact, the low frequencies are not affected by the radiation-damage in the high-dose exposed images. These images are usually used for particle picking, but their use for angular assignment is currently not explored. For back projection however, the “low-dose” particles should be used to preserve the high resolution information.

4.2 Current models of co-translational targeting and translocation

4.2.1 SRP-mediated protein targeting in *E. coli*

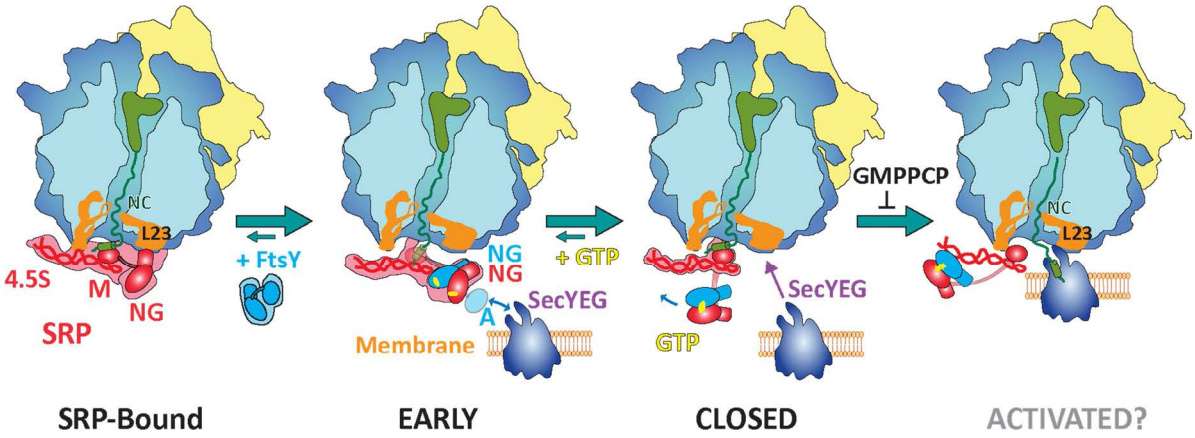


Figure 4-1: Current model of SRP-mediated co-translational targeting. In the RNC–SRP complex, SRP is prepositioned to bind FtsY. FtsY binding leads to an early complex inactive in GTP hydrolysis. Rearrangement of the NG domains in the closed complex requires GTP and leads to detachment of the NG domains from the RNA tetraloop. SecYEG binding to L23 is suggested to lead to docking of the NG domains to the distal end of the RNA in the activated state and RNC handover to SecYEG. GTP hydrolysis results in SRP–FtsY disassembly.

In this thesis work, I solved the cryo-EM structure of the ribosome-SRP-FtsY complex in the "closed" conformation. The structure revealed a conformation of the SRP and the FtsY that has not been reported before: a tightly bound and well solved SRP M domain, with the NG domains of the SRP and the FtsY remained flexible (Figure 4-1). This structure provides critical new insights into the SRP-mediated targeting of membrane proteins. Now, all the major conformational states which were originally identified by FRET measurements (Zhang *et al.*, 2009) were analyzed by cryo-EM, until the "closed" conformation. During co-translational targeting, the SRP tightly binds to the ribosome upon recognition of a signal sequence. The NG domain of the SRP is prepositioned to facilitate the binding of FtsY (Estrozi *et al.*, 2011). The FtsY G-domain first binds to the tetraloop of the 4.5S RNA, while the Ffh-FtsY NG-domains weakly interact via the N-domain, adopting a V-shaped arrangement which is inactive in GTP hydrolysis (Estrozi *et al.*, 2011). The activation of GTPases requires a rearrangement of the NG-domains to form an intimate Ffh-FtsY complex with a composite active site (Egea *et al.*, 2004; Focia *et al.*, 2004), and GTP hydrolysis is further stimulated by binding to the residues at the distal end of the 4.5S RNA (Voigts-Hoffmann *et al.*, 2013). Moreover, the displacement of the NG-domains from the tetraloop towards the distal end of the 4.5S RNA is required to provide access of the translocation machinery to the ribosomal proteins at the exit of the

ribosomal tunnel. Docking of NG-domains to the distal end may occur with additional conformational changes in the SRP and FtsY and may be triggered by the binding of a SecYEG translocon to the ribosome (Shen *et al.*, 2012), followed by the handover of the nascent chain to the translocation machinery. The hydrolysis of GTP then results in the disassembly of the SRP–FtsY complex (Figure 4-1). It is unclear at this moment whether the activation of the SRP-FtsY complex (*i.e*. the activated state) occurs on the ribosome or whether the SecYEG translocon displaces SRP from the ribosome first.

It is exciting to see that the steps of co-translational targeting by SRP and FtsY were first characterized biochemically using FRET and SRP mutants (Zhang *et al.*, 2009). The targeting model has then been steadily refined with the help of single-particle FRET (Shen *et al.*, 2012), crystal structures of SRP and SRP-FtsY complexes (Egea *et al*., 2004; Focia *et al*., 2004; Janda *et al*., 2010; Ataide *et al*., 2010; Voigts-Hoffman *et al*., 2013) and cryo-EM structures of ribosomal targeting complexes (Schaffitzel *et al*., 2006; Estrozi *et al*., 2011; von Loeffelholz *et al*., 2014; this work). SRP targeting is a good example of how different techniques are combined to help understanding a biological process. It is also a good illustration of how X-ray crystallography can provide the detailed information of individual proteins and smaller complexes while the cryo-EM is used to study large complexes, which are unlikely to crystallize, in an environment closer to the physiological conditions.

4.2.2 Membrane protein insertion by HTL

Prokaryotic translocation proteins have been studied intensely by crystallography and cryo-EM. Crystal structures of SecYEG (Zimmer *et al*., 2008), SecDF (Tsukazaki *et al*., 2011) and YidC (Kumazuki *et al*., 2014) have been described providing crucial mechanistic insights into the molecular mechanism of translocation. In parallel, cryo-EM structures of ribosome-SecYEG complexes (Park *et al.*, 2014; Gogala *et al*. 2014) and ribosome-YidC complexes (Kohler *et al.*, 2009; Wickles *et al.*, 2014) have been reported. However to understand how these translocation proteins cooperate, it will be essential to provide cryo-EM structures of ribosome-HTL complexes. Our cryo-EM structure of RNC^{FtsQ}-HTL^{LMNG} provides new insights into the HTL

architecture and interactions with the ribosome, albeit at low resolution. Combined with published results, we propose the current working model of HTL as shown in Figure 4-2.

The lateral gate of the SecYEG translocation pore is suggested to be in close proximity of YidC (Sachelaru *et al.*, 2013). Thus our current working model (Figure 4-2) suggests that SecY opens towards YidC to handover transmembrane helices of the nascent protein. The transmembrane helices bind to YidC until they can fold to a stable domain assisted by the YidC chaperone. Our cryo-EM structure of the ribosome-HTL complex could support this hypothesis. As shown in Figure 3-12, it is possible to fit SecYEG and YidC separately into the EM density at the exit of the ribosomal tunnel such that the lateral gate of SecYEG points towards YidC and such that the cytoplasmic loops of SecY contact ribosomal proteins L23 and L29 as previously reported (Park *et al.*, 2014; Gogala *et al*. 2014). Moreover, there is additional space in the cryo-EM density to further accommodate the transmembrane parts of SecD and SecF. Biochemical studies suggest that SecD and SecF facilitate the translocation of proteins by interacting with the translocation substrate on the periplasmic side. Binding of translocation substrates to the head domain of the SecD periplasmic domain (P1) may prevent backsliding of translocation substrates.

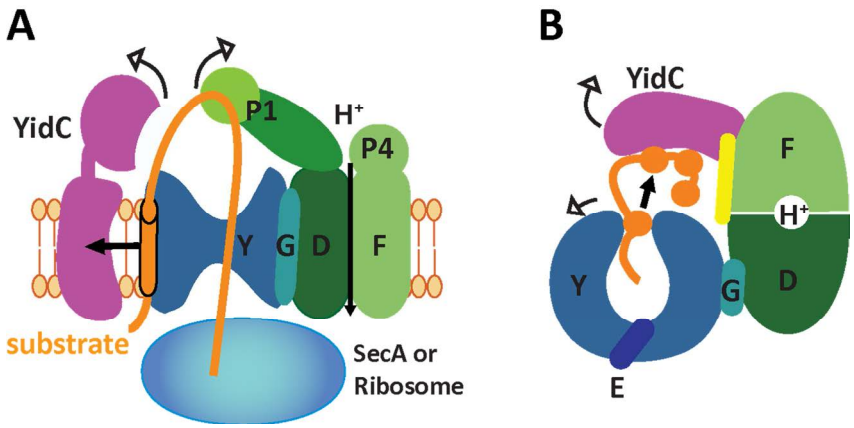


Figure 4-2: Current model of membrane protein insertion by HTL. Viewed in the membrane plane (A) and perpendicular to the plane of the membrane (B). Translocation of the nascent peptide through the SecYEG channel is facilitated by translocation substrate interactions with the periplasmic domains of YidC and SecD. Membrane protein integration and folding can occur at the interface between the SecY lateral gate and YidC in a protected lipid-holotranslocon environment where transmembrane helices accumulate until they can fold into a domain.

Moreover, conformational changes in SecD could couple protein translocation with proton transfer across the membrane, thus accelerating translocation (Tsukazaki *et al*, 2011). Notably, ribosomal RNA helix 59, which has a strong connection to the YidC in our ribosome-HTL cryo-EM structure, has been observed to interact with lipids in the cryo-EM structure of the ribosome-SecYEG complex with a nanodisc (Frauenfeld *et al.*, 2011). Indeed, the cytoplasmic side of YidC contains several positively charged residues which may interact with the negatively charged ribosomal RNA.

Taken together, the cryo-EM structure of RNC^{FtsQ}-HTL^{LMNG} provides new insights into the HTL architecture and interactions with the ribosome, albeit at low resolution. A better-resolved structure is highly desired to further interpret these interactions, if possible at near-atomic resolution and to provide testable hypotheses for co-translational translocation by the HTL complex.

4.3 Future Studies

4.3.1 Further works on ribosome-SRP-FtsY complex

With the “closed” ribosome-SRP-FtsY complex being solved, the collection of structures along the SRP-mediated targeting pathway seems to be close to complete. The obvious missing structure here is an “active” state complex. However, though the “active” state is described biochemically, it most likely represents a transient complex consisting of ribosome, SRP, FtsY and SecYEG. According to the current model, the binding of SecYEG induces the activation of the NG-domains’ GTPases, and at the same time, the nascent chain is released from the SRP M-domain and subsequently captured by the SecYEG. It may be possible to produce and stabilize a complex in which SecYEG binds to the ribosome while SRP remains its interaction with the ribosome and/or interact with the SecYEG, but so far there is no biochemical data indicating the existence of such complex.

In the future it will be interesting to study the “closed” state complex with different types of nascent chains. The work using EspP nascent chian in an “early” state

complex is a good demonstration of how RNCs displaying different signal peptides could influence the structural rearrangement of the SRP and FtsY, and thus provide a new structural basis for the SRP-mediated targeting pathway (von Loeffelholz *et al.*, 2013). Moreover, SRP could also target membrane proteins into YidC (Welte *et al.*, 2012). The choice of translocons during SRP targeting is very likely determined by signal peptide, and therefore the conformation of SRP and FtsY while carrying substrates for YidC and for SecYEG may also be different. However, the molecular details of this are enigmatic.

4.3.2 New approaches in studying ribosome-HTL complex

It has been shown that ribosome-translocon complexes can also be purified directly from the cell rather than being reconstituted. For example, to solve the cryo-EM structure of the mammalian ribosome-Sec61 complex (Voorhees *et al.*, 2014), the complex was extracted directly from porcine pancreatic microsomes where they are enriched using mild detergents. Despite the fact that we expect such sample to be less homogenous, it was possible to reach a resolution of 3.4 Å for the ribosome and ~5 Å for the translocon part (Voorhees *et al.*, 2014). A similar approach may be possible for ribosome-HTL complexes using cells that overexpress the HTL complex. In this case, cryo-EM specimen can be prepared by isolation of membranes, mild detergent solubilization followed by purification of the membrane-bound ribosomes that are expected to be bound by HTL. Given the advances on microscopy hardware and data processing, it should be possible (as in the case of ribosome-Sec61) to "purify" the ribosome-HTL complex by computational sorting.

At the same time, the new generation of direct electron detectors also impressively improved the resolution that can be achieved by electron cryo-tomography and rendered cryo-tomography and sub-tomogram averaging into a powerful tool in studying macromolecular complexes (Chen *et al.*, 2013). The tomography approach has several advantages over single-particle analysis: for instance, the possibility to image the sample *in situ*, *i.e.* without the need to extract and purify the complexes of interest. The subsequent data processing is facilitated as the angular relations between the images are known. Until recently, the resolution

of the EM structures obtained by this cryo-tomography was rather limited due to the relatively low signal-to-noise ratio of each individual image. For instance, a structure of an ER-bound ribosome has been solved at 31 Å using cryo-electron tomography (Pfeffer *et al.*, 2012). A more recent publication on membrane-associated mito-ribosomes using similar approaches has reached a resolution of 28 Å, allowing identification of a protein that mediates the binding of mitoribosomes to the translocon (Pfeffer *et al.*, 2014). With the tomography technique is still evolving, it is foreseeable that near-atomic resolution structures can also be solved by cryo-electron tomography provided that the sample is sufficiently homogenous and enriched in the cell such that many sub-tomograms can be obtained and used for averaging.

Besides HTL, the ribosome nascent-chain complex can also be manipulated. Just as the cases of ribosome-SRP-FtsY complexes, the nascent peptide displayed by the ribosome has a large influence on the structure adopted by the translocon. A study comparing cryo-EM structures of a non-translating ribosome-SecY complex and a RNC-SecY complex showed that major conformational changes of the SecY happen with the insertion of a nascent peptide (Park *et al.*, 2014). In fact, it is not difficult to imagine that the holo-translocon is a flexible machine that has to adopt different conformations to accommodate nascent peptides of different length, with different amounts of transmembrane helices and with different sizes of cytoplasmic and periplasmic loops and domains. As HTL was proposed to be required for the translocation of larger and complex membrane proteins, it is particularly interesting to obtain structural insights into the insertion of a nascent chain with multiple transmembrane helices captured by the HTL, and to observe how the transmembrane helices are handed over from SecYEG to YidC.

4.3.3 Post-translational translocation

The HTL complex has been shown to be active in the co-translational translocation of membrane proteins and as well in post-translation translocation of secretory proteins (Figure 1-1) (Schulze *et al*., 2014). During post-translocation, pre-proteins containing a signal sequence are held in an unfolded state by the SecB protein. ATP

hydrolysis by SecA is the driving force to push the pre-proteins through the SecYEG protein conducting channel. The crystal structure of the *Thermotoga maritime* SecA-SecYEG complex revealed that SecA binding causes an opening of the lateral gate of SecY and leads to a displacement of the plug domain of SecY, thus providing room to integrate the signal sequence into the lateral gate of the translocon leading to an opening of the central pore that allows the secretory protein to pass through the translocon in a mostly hydrophilic environment (Zimmer, Nam and Rapoport, 2008). Cross-linking experiments showed that the binding of YidC to the SecY lateral gate is preserved during SecA binding (Sachelaru *et al.*, 2013). In contrast, biochemical studies suggest that SecDF is involved in the regulation of the SecA-SecYEG interaction (Duong and Wickner, 1997). A structure of the SecA-HTL complex with a translocation substrate (*i.e*. a preprotein) will provide valuable information regarding the roles of SecDF-YajC and of YidC during post-translational translocation. Moreover, the different conformations of HTL subunits during co- and post-translational translocation could be revealed.

4.3.4 Eukaryotic co-translational targeting and translocation

Eukaryotic cells are more much more complex with many subcellular organelles. Targeting of proteins to the nucleus, the mitochondria *etc.* follows special pathways that do not exist in prokaryotic cells. The SRP-mediated targeting to the Sec translocation machinery however is evolutionary conserved. SRP targets proteins into and across the membrane of the endoplasmic reticulum (ER). In eukaryotic systems, nascent proteins with a signal peptide are recognized by the eukaryotic SRP, composed of six distinct polypeptide subunits bound to a SRP RNA (7S, 300 nucleotides) which is significantly larger than the *E. coli* SRP RNA (4.5 S, 114 nucleotides). Compared to its homologue in *E. coli*, the eukaryotic SRP comprises an Alu domain that is responsible for translational arrest or slow-down during SRP targeting.

For co-translational targeting, the RNC-SRP complex docks to the ER membrane via the SRP receptor which (in contrast to *E. coli* FtsY) consists of two subunits, one of them is an integral membrane protein (Nyathi, Wilkinson and Pool, 2013). Under

GTP hydrolysis, the RNC is transferred to the Sec translocon complex while the SRP and its receptor are released and translation resumes. The eukaryotic protein conducting channel is named Sec61, and is homologous to the bacterial SecYEG. Like SecYEG, Sec61 can also recruit further proteins to help in protein folding and processing, for example the signal peptidase and the oligosaccharyl transferase complex which adds N-linked glycans to the nascent chain were shown to bind to the Sec translocon (Figure 4-3).

High-resolution structural studies of protein complexes involved in co-translational targeting and translocation in eukaryotic cells are hampered by the large size of the complexes. However, it is likely that these complexes can now be studied by cryo-EM since they are abundant in the cell. Just as in prokaryotic systems, the SRP and its receptor undergo multiple conformational changes during the targeting process (Nyathi, Wilkinson and Pool, 2013). Currently, only one cryo-EM structure has been

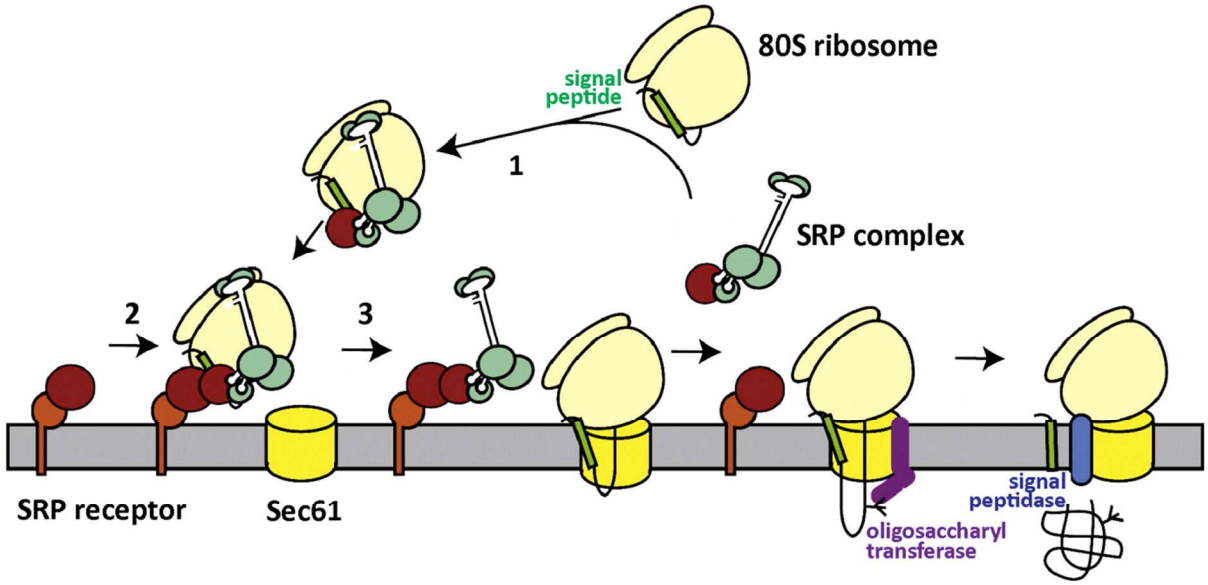


Figure 4-3: Eukaryotic co-translational targeting and translocation. Schematic representation of proteins involved in the co-translational targeting and translocation in eukaryotes. Ribosomes translating polypeptides with a signal peptide are recognized by the SRP, which contains an Alu domain responsible for translational arrest (1). The RNC-SRP complex is then recruited by the SRP receptor to the membrane of the endoplasmic reticulum (2). SRP and its receptor activate each other and the ribosome-nascent chain complex is handed over to the Sec61 translocon. The SRP and its receptor are recycled (3). Additional proteins may bind to the ribosome-Sec61 complex for instance the oligosaccharyl transferase (OST) complex and the translocon-associated protein (TRAP) complex (bottom right). Adapted from Nyathi, Wilkinson and Pool, 2013.

reported for a ribosome-SRP-SRP receptor complex and it was not possible to unambiguously interpret the SRP-SRP receptor part of this EM structure due to its limited resolution (Halic *et al*., 2006-1). The structure of the ribosome-Sec61 complex has been solved at high resolution (Voorhees *et al.*, 2014), but for the ribosomal complex with Sec61, TRAP and OST only a low-resolution EM structure is available (Pfeffer *et al.*, 2012). Therefore, to understand commonalities and differences in the molecular mechanisms, further structural and functional studies on eukaryotic co-translational targeting and translocation are highly demanded.

REFERENCES

ANGELINI, S., S. DEITERMANN AND H. G. KOCH FtsY, the bacterial signal-recognition particle receptor, interacts functionally and physically with the SecYEG translocon. EMBO Rep, May 2005, 6(5), 476-481.

ARKOWITZ, R. A. AND W. WICKNER SecD and SecF are required for the proton electrochemical gradient stimulation of preprotein translocation. EMBO J, Feb 15 1994, 13(4), 954-963.

ATAIDE, S. F., N. SCHMITZ, K. SHEN, A. KE, et al. The crystal structure of the signal recognition particle in complex with its receptor. Science, Feb 18 2011, 331(6019), 881-886.

BAI, X. C., I. S. FERNANDEZ, G. MCMULLAN AND S. H. SCHERES Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. Elife, 2013, 2, e00461.

BAN, N., P. NISSEN, J. HANSEN, P. B. MOORE, et al. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science, Aug 11 2000, 289(5481), 905-920.

BATEY, R. T., R. P. RAMBO, L. LUCAST, B. RHA, et al. Crystal structure of the ribonucleoprotein core of the signal recognition particle. Science, Feb 18 2000, 287(5456), 1232-1239.

BIENIOSSEK, C., Y. NIE, D. FREY, N. OLIERIC, et al. Automated unrestricted multigene recombineering for multiprotein complex production. Nat Methods, Jun 2009, 6(6), 447-450.

BOTTE, M. Architecture of the SecYEG-DF-YajC-YidC Holotranslocon. Th : Grenoble I : 2014.

BRILOT, A. F., J. Z. CHEN, A. CHENG, J. PAN, et al. Beam-induced motion of vitrified specimen on holey carbon film. J Struct Biol, Mar 2012, 177(3), 630-637.

CARPENTER, E. P., K. BEIS, A. D. CAMERON AND S. IWATA Overcoming the challenges of membrane protein crystallography. Curr Opin Struct Biol, Oct 2008, 18(5), 581-586.

CHAE, P. S., S. G. RASMUSSEN, R. R. RANA, K. GOTFRYD, et al. Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. Nat Methods, Dec 2010, 7(12), 1003-1008.

CHEN, M., J. C. SAMUELSON, F. JIANG, M. MULLER, et al. Direct interaction of YidC with the Sec-independent Pf3 coat protein during its membrane protein insertion. J Biol Chem, Mar 8 2002, 277(10), 7670-7675.

CHEN, Y., S. PFEFFER, T. HRABE, J. M. SCHULLER, et al. Fast and accurate reference-free alignment of subtomograms. J Struct Biol, Jun 2013, 182(3), 235-245.

DALAL, K., C. S. CHAN, S. G. SLIGAR AND F. DUONG Two copies of the SecY channel and acidic lipids are necessary to activate the SecA translocation ATPase. Proc Natl Acad Sci U S A, Mar 13 2012, 109(11), 4104-4109.

DE LA ROSA-TREVIN, J. M., J. OTON, R. MARABINI, A. ZALDIVAR, et al. Xmipp 3.0: an improved software suite for image processing in electron microscopy. J Struct Biol, Nov 2013, 184(2), 321-328.

DE VRIJE, T., R. L. DE SWART, W. DOWHAN, J. TOMMASSEN, et al. Phosphatidylglycerol is involved in protein translocation across Escherichia coli inner membranes. Nature, Jul 14 1988, 334(6178), 173-175.

DENISOV, I. G., Y. V. GRINKOVA, A. A. LAZARIDES AND S. G. SLIGAR Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size. J Am Chem Soc, Mar 24 2004, 126(11), 3477-3487.

DUONG, F. AND W. WICKNER The SecDFyajC domain of preprotein translocase controls preprotein movement by regulating SecA membrane cycling. EMBO J, Aug 15 1997, 16(16), 4871-4879.

EGEA, P. F., S. O. SHAN, J. NAPETSCHNIG, D. F. SAVAGE, et al. Substrate twinning activates the signal recognition particle and its receptor. Nature, Jan 15 2004, 427(6971), 215-221.

ERNI, R., M. D. ROSSELL, C. KISIELOWSKI AND U. DAHMEN Atomic-resolution imaging with a sub-50-pm electron probe. Phys Rev Lett, Mar 6 2009, 102(9), 096101.

ESTROZI, L. F., D. BOEHRINGER, S. O. SHAN, N. BAN, et al. Cryo-EM structure of the E. coli translating ribosome in complex with SRP and its receptor. Nat Struct Mol Biol, Jan 2011, 18(1), 88-90.

FISCHER, N., A. L. KONEVEGA, W. WINTERMEYER, M. V. RODNINA, et al. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. Nature, Jul 15 2010, 466(7304), 329-333.

FISCHER, N., P. NEUMANN, A. L. KONEVEGA, L. V. BOCK, et al. Structure of the E. coli ribosome-EF-Tu complex at <3 A resolution by C-corrected cryo-EM. Nature, Feb 23 2015.

FOCIA, P. J., I. V. SHEPOTINOVSKAYA, J. A. SEIDLER AND D. M. FREYMANN Heterodimeric GTPase core of the SRP targeting complex. Science, Jan 16 2004, 303(5656), 373-377.

FRAUENFELD, J., J. GUMBART, E. O. SLUIS, S. FUNES, et al. Cryo-EM structure of the ribosome-SecYE complex in the membrane environment. Nat Struct Mol Biol, May 2011, 18(5), 614-621.

GARAVITO, R. M. AND S. FERGUSON-MILLER Detergents as tools in membrane biochemistry. J Biol Chem, Aug 31 2001, 276(35), 32403-32406.

GOGALA, M., T. BECKER, B. BEATRIX, J. P. ARMACHE, et al. Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion. Nature, Feb 6 2014, 506(7486), 107-110.

GOLD, V. A., A. ROBSON, H. BAO, T. ROMANTSOV, et al. The action of cardiolipin on the bacterial translocon. Proc Natl Acad Sci U S A, Jun 1 2010, 107(22), 10044-10049.

HAINZL, T., S. HUANG, G. MERILAINEN, K. BRANNSTROM, et al. Structural basis of signal-sequence recognition by the signal recognition particle. Nat Struct Mol Biol, Mar 2011, 18(3), 389-391.

HALIC, M., M. BLAU, T. BECKER, T. MIELKE, et al. Following the signal sequence from ribosomal tunnel exit to signal recognition particle. Nature, Nov 23 2006, 444(7118), 507-511.

HALIC, M., M. GARTMANN, O. SCHLENKER, T. MIELKE, et al. Signal recognition particle receptor exposes the ribosomal translocon binding site. Science, May 5 2006, 312(5774), 745-747.

HENDERSON, R., S. CHEN, J. Z. CHEN, N. GRIGORIEFF, et al. Tilt-pair analysis of images from a range of different specimens in single-particle electron cryomicroscopy. J Mol Biol, Nov 11 2011, 413(5), 1028-1046.

HENDERSON, R., A. SALI, M. L. BAKER, B. CARRAGHER, et al. Outcome of the first electron microscopy validation task force meeting. Structure, Feb 8 2012, 20(2), 205-214.

HEYMANN, J. B. AND D. M. BELNAP Bsoft: image processing and molecular modeling for electron microscopy. J Struct Biol, Jan 2007, 157(1), 3-18.

JANDA, C. Y., J. LI, C. OUBRIDGE, H. HERNANDEZ, et al. Recognition of a signal peptide by the signal recognition particle. Nature, May 27 2010, 465(7297), 507-510.

KALIPATNAPU, S. AND A. CHATTOPADHYAY Membrane protein solubilization: recent advances and challenges in solubilization of serotonin1A receptors. IUBMB Life, Jul 2005, 57(7), 505-512.

KOHLER, R., D. BOEHRINGER, B. GREBER, R. BINGEL-ERLENMEYER, et al. YidC and Oxa1 form dimeric insertion pores on the translating ribosome. Mol Cell, May 15 2009, 34(3), 344-353.

KUCUKELBIR, A., F. J. SIGWORTH AND H. D. TAGARE Quantifying the local resolution of cryo-EM density maps. Nat Methods, Jan 2014, 11(1), 63-65.

KUMAZAKI, K., T. KISHIMOTO, A. FURUKAWA, H. MORI, et al. Crystal structure of Escherichia coli YidC, a membrane protein chaperone and insertase. Sci Rep, 2014, 4, 7299.

LIAO, M., E. CAO, D. JULIUS AND Y. CHENG Structure of the TRPV1 ion channel determined by electron cryo-microscopy. Nature, Dec 5 2013, 504(7478), 107-112.

LILL, R., W. DOWHAN AND W. WICKNER The ATPase activity of SecA is regulated by acidic phospholipids, SecY, and the leader and mature domains of precursor proteins. Cell, Jan 26 1990, 60(2), 271-280.

LU, P., X. C. BAI, D. MA, T. XIE, et al. Three-dimensional structure of human gamma-secretase. Nature, Aug 14 2014, 512(7513), 166-170.

MENETRET, J. F., J. SCHALETZKY, W. M. CLEMONS, JR., A. R. OSBORNE, et al. Ribosome binding of a single copy of the SecY complex: implications for protein translocation. Mol Cell, Dec 28 2007, 28(6), 1083-1092.

MITRA, K., C. SCHAFFITZEL, T. SHAIKH, F. TAMA, et al. Structure of the E. coli protein-conducting channel bound to a translating ribosome. Nature, Nov 17 2005, 438(7066), 318-324.

NATH, A., W. M. ATKINS AND S. G. SLIGAR Applications of phospholipid bilayer nanodiscs in the study of membranes and membrane proteins. Biochemistry, Feb 27 2007, 46(8), 2059-2069.

NOESKE, J., J. HUANG, N. B. OLIVIER, R. A. GIACOBBE, et al. Synergy of streptogramin antibiotics occurs independently of their effects on translation. Antimicrob Agents Chemother, Sep 2014, 58(9), 5269-5279.

NOUWEN, N. AND A. J. DRIESSEN SecDFyajC forms a heterotetrameric complex with YidC. Mol Microbiol, Jun 2002, 44(5), 1397-1405.

NYATHI, Y., B. M. WILKINSON AND M. R. POOL Co-translational targeting and translocation of proteins to the endoplasmic reticulum. Biochim Biophys Acta, Nov 2013, 1833(11), 2392-2402.

OLIVER, D. C. AND M. PAETZEL Crystal structure of the major periplasmic domain of the bacterial membrane protein assembly facilitator YidC. J Biol Chem, Feb 22 2008, 283(8), 5208-5216.

PARK, E., J. F. MENETRET, J. C. GUMBART, S. J. LUDTKE, et al. Structure of the SecY channel during initiation of protein translocation. Nature, Feb 6 2014, 506(7486), 102-106.

PARK, K. H., E. BILLON-DENIS, T. DAHMANE, F. LEBAUPAIN, et al. In the cauldron of cell-free synthesis of membrane proteins: playing with new surfactants. N Biotechnol, Apr 30 2011, 28(3), 255-261.

PETTERSEN, E. F., T. D. GODDARD, C. C. HUANG, G. S. COUCH, et al. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem, Oct 2004, 25(13), 1605-1612.

PFEFFER, S., F. BRANDT, T. HRABE, S. LANG, et al. Structure and 3D arrangement of endoplasmic reticulum membrane-associated ribosomes. Structure, Sep 5 2012, 20(9), 1508-1518.

PFEFFER, S., M. W. WOELLHAF, J. M. HERRMANN AND F. FORSTER Organization of the mitochondrial translation machinery studied in situ by cryoelectron tomography. Nat Commun, 2015, 6, 6019.

PICARD, M., T. DAHMANE, M. GARRIGOS, C. GAURON, et al. Protective and inhibitory effects of various types of amphipols on the Ca2+-ATPase from sarcoplasmic reticulum: a comparative study. Biochemistry, Feb 14 2006, 45(6), 1861-1869.

RADERMACHER, M., T. WAGENKNECHT, A. VERSCHOOR AND J. FRANK Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. J Microsc, May 1987, 146(Pt 2), 113-136.

RAPOPORT, T. A. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. Nature, Nov 29 2007, 450(7170), 663-669.

RAVAUD, S., G. STJEPANOVIC, K. WILD AND I. SINNING The crystal structure of the periplasmic domain of the Escherichia coli membrane protein insertase YidC contains a substrate binding cleft. J Biol Chem, Apr 4 2008, 283(14), 9350-9358.

ROOS, C., M. ZOCHER, D. MULLER, D. MUNCH, et al. Characterization of co-translationally formed nanodisc complexes with small multidrug transporters, proteorhodopsin and with the E. coli MraY translocase. Biochim Biophys Acta, Dec 2012, 1818(12), 3098-3106.

ROSENDAL, K. R., K. WILD, G. MONTOYA AND I. SINNING Crystal structure of the complete core of archaeal signal recognition particle and implications for interdomain communication. Proc Natl Acad Sci U S A, Dec 9 2003, 100(25), 14701-14706.

ROSENTHAL, P. B. AND R. HENDERSON Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J Mol Biol, Oct 31 2003, 333(4), 721-745.

SACHELARU, I., N. A. PETRIMAN, R. KUDVA, P. KUHN, et al. YidC occupies the lateral gate of the SecYEG translocon and is sequentially displaced by a nascent membrane protein. J Biol Chem, Jun 7 2013, 288(23), 16295-16307.

SAMBROOK, J. Molecular cloning : a laboratory manual. In D.W. RUSSELL. Cold Spring Harbor, N.Y. : : Cold Spring Harbor Laboratory Press, 2001.

SAMUELSON, J. C., F. JIANG, L. YI, M. CHEN, et al. Function of YidC for the insertion of M13 procoat protein in Escherichia coli: translocation of mutants that show differences in their membrane potential dependence and Sec requirement. J Biol Chem, Sep 14 2001, 276(37), 34847-34852.

SCHAFFITZEL, C. AND N. BAN Generation of ribosome nascent chain complexes for structural and functional studies. J Struct Biol, Jun 2007, 158(3), 463-471.

SCHAFFITZEL, C., M. OSWALD, I. BERGER, T. ISHIKAWA, et al. Structure of the E. coli signal recognition particle bound to a translating ribosome. Nature, Nov 23 2006, 444(7118), 503-506.

SCHERES, S. H. A Bayesian view on cryo-EM structure determination. J Mol Biol, Jan 13 2012, 415(2), 406-418.

SCHERES, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol, Dec 2012, 180(3), 519-530.

SCHLUENZEN, F., A. TOCILJ, R. ZARIVACH, J. HARMS, et al. Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. Cell, Sep 1 2000, 102(5), 615-623.

SCHULZE, R. J., J. KOMAR, M. BOTTE, W. J. ALLEN, et al. Membrane protein insertion and proton-motive-force-dependent secretion through the bacterial holo-translocon SecYEG-SecDF-YajC-YidC. Proc Natl Acad Sci U S A, Apr 1 2014, 111(13), 4844-4849.

SCHUWIRTH, B. S., M. A. BOROVINSKAYA, C. W. HAU, W. ZHANG, et al. Structures of the bacterial ribosome at 3.5 A resolution. Science, Nov 4 2005, 310(5749), 827-834.

SCOTTI, P. A., M. L. URBANUS, J. BRUNNER, J. W. DE GIER, et al. YidC, the Escherichia coli homologue of mitochondrial Oxa1p, is a component of the Sec translocase. EMBO J, Feb 15 2000, 19(4), 542-549.

SEDDON, A. M., P. CURNOW AND P. J. BOOTH Membrane proteins, lipids and detergents: not just a soap opera. Biochim Biophys Acta, Nov 3 2004, 1666(1-2), 105-117.

SHAIKH, T. R., H. GAO, W. T. BAXTER, F. J. ASTURIAS, et al. SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. Nat Protoc, 2008, 3(12), 1941-1974.

SHIH, A. Y., P. L. FREDDOLINO, S. G. SLIGAR AND K. SCHULTEN Disassembly of nanodiscs with cholate. Nano Lett, Jun 2007, 7(6), 1692-1696.

SIGWORTH, F. J. A maximum-likelihood approach to single-particle image refinement. J Struct Biol, 1998, 122(3), 328-339.

SIGWORTH, F. J., P. C. DOERSCHUK, J.-M. CARAZO AND S. H. W. SCHERES Maximum-likelihood methods in cryo-EM. Part I: theoretical basis and overview of existing approaches. Methods Enzymol, 2010, 482, 263-294.

SINGER, S. J. AND G. L. NICOLSON The fluid mosaic model of the structure of cell membranes. Science, Feb 18 1972, 175(4023), 720-731.

SORZANO, C. O. S., A. IRIARTE, R. MARABINI AND J. M. CARAZO. Effects of the downsampling scheme on three-dimensional electron microscopy of single particles. In Intelligent Signal Processing, 2009. WISP 2009. IEEE International Symposium on. 2009, p. 175-179.

STARK, H. GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. Methods Enzymol, 2010, 481, 109-126.

TANG, G., L. PENG, P. R. BALDWIN, D. S. MANN, et al. EMAN2: an extensible image processing suite for electron microscopy. J Struct Biol, Jan 2007, 157(1), 38-46.

TIAN, P. AND H. D. BERNSTEIN Identification of a Post-targeting Step Required for Efficient Cotranslational Translocation of Proteins across the Escherichia coli Inner Membrane. Journal of Biological Chemistry, Apr 24 2009, 284(17), 11396-11404.

TSUKAZAKI, T., H. MORI, Y. ECHIZEN, R. ISHITANI, et al. Structure and function of a membrane component SecDF that enhances protein export. Nature, Jun 9 2011, 474(7350), 235-238.

VALLE, M., J. SENGUPTA, N. K. SWAMI, R. A. GRASSUCCI, et al. Cryo-EM reveals an active role for aminoacyl-tRNA in the accommodation process. EMBO J, Jul 1 2002, 21(13), 3557-3567.

VAN DEN BERG, B., W. M. CLEMONS, JR., I. COLLINSON, Y. MODIS, et al. X-ray structure of a protein-conducting channel. Nature, Jan 1 2004, 427(6969), 36-44.

VAN DER LAAN, M., P. BECHTLUFT, S. KOL, N. NOUWEN, et al. F1F0 ATP synthase subunit c is a substrate of the novel YidC pathway for membrane protein biogenesis. J Cell Biol, Apr 26 2004, 165(2), 213-222.

VAN HEEL, M., G. HARAUZ, E. V. ORLOVA, R. SCHMIDT, et al. A new generation of the IMAGIC image processing system. J Struct Biol, Jan-Feb 1996, 116(1), 17-24.

VOIGTS-HOFFMANN, F., N. SCHMITZ, K. SHEN, S. O. SHAN, et al. The structural basis of FtsY recruitment and GTPase activation by SRP RNA. Mol Cell, Dec 12 2013, 52(5), 643-654.

VON HEIJNE, G. AND Y. GAVEL Topogenic signals in integral membrane proteins. Eur J Biochem, Jul 1 1988, 174(4), 671-678.

VON LOEFFELHOLZ, O., K. KNOOPS, A. ARIOSA, X. ZHANG, et al. Structural basis of signal sequence surveillance and selection by the SRP-FtsY complex. Nat Struct Mol Biol, May 2013, 20(5), 604-610.

VON LOEFFELHOLZ, O., M. BOTTE AND C. SCHAFFITZEL. Escherichia coli Cotranslational Targeting and Translocation. Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester. 2011.

VOORHEES, R. M., I. S. FERNANDEZ, S. H. SCHERES AND R. S. HEGDE Structure of the mammalian ribosome-Sec61 complex to 3.4 A resolution. Cell, Jun 19 2014, 157(7), 1632-1643.

WEICHE, B., J. BURK, S. ANGELINI, E. SCHILTZ, et al. A cleavable N-terminal membrane anchor is involved in membrane binding of the Escherichia coli SRP receptor. J Mol Biol, Mar 28 2008, 377(3), 761-773.

WELTE, T., R. KUDVA, P. KUHN, L. STURM, et al. Promiscuous targeting of polytopic membrane proteins to SecYEG or YidC by the Escherichia coli signal recognition particle. Mol Biol Cell, Feb 2012, 23(3), 464-479.

WICKLES, S., A. SINGHAROY, J. ANDREANI, S. SEEMAYER, et al. A structural model of the active ribosome-bound membrane protein insertase YidC. Elife, 2014, 3, e03035.

WIMBERLY, B. T., D. E. BRODERSEN, W. M. CLEMONS, JR., R. J. MORGAN-WARREN, et al. Structure of the 30S ribosomal subunit. Nature, Sep 21 2000, 407(6802), 327-339.

ZHANG, X., C. SCHAFFITZEL, N. BAN AND S. O. SHAN Multiple conformational switches in a GTPase complex control co-translational protein targeting. Proc Natl Acad Sci U S A, Feb 10 2009, 106(6), 1754-1759.

ZIMMER, J., Y. NAM AND T. A. RAPOPORT Structure of a complex of the ATPase SecA and the protein-translocation channel. Nature, Oct 16 2008, 455(7215), 936-943.

ZOONENS, M. AND J. L. POPOT Amphipols for each season. J Membr Biol, Oct 2014, 247(9-10), 759-796.