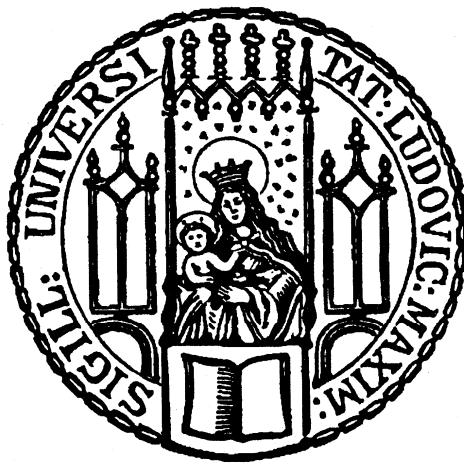


Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

A structural model of the active ribosome-bound membrane protein insertase YidC



Stephan Walter Wickles
aus
Pegnitz, Deutschland

2015

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Roland Beckmann betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 20.01.2015

Stephan Wickles

Dissertation eingereicht am 20.01.2015

1. Gutachter: Prof. Dr. Roland Beckmann

2. Gutachter: Prof. Dr. Mario Halic

Mündliche Prüfung am 10.03.2015

Parts of this thesis have been published in

Noroussi, R.* , **Wickles, S*** *et al.* Automatic post-picking using MAPPOS improves particle image detection from cryo-EM micrographs. *Journal of Structural Biology* **182**, 59–66 (2013)

* These authors contributed equally to the work

Wickles, S. *et al.* A structural model of the active ribosome-bound membrane protein insertase YidC. *eLife Sciences* **3**, e03035 (2014).

Parts of this thesis have been presented at international conferences

Gordon Conference: Three Dimensional Electron Microscopy
Les Diablerets, Switzerland 2012

Stephan Wickles, Ramin Noroussi, Roland Beckmann and Achim Tresch
MAPPOS: A Machine learning algorithm for particle post picking

Gordon Conference: Protein Transport Across Cell Membranes
Galveston, USA 2014

Stephan Wickles, Lukas Bischoff, Eli van der Sluis, Stefan Seemayer, Singharoy Abhishek, Klaus Schulten, Johannes Söding and Roland Beckmann
Ab initio modelling of YidC using novel bioinformatic tools

Gordon Conference: Three Dimensional Electron Microscopy
Girona, Spain 2014

Stephan Wickles, Lukas Bischoff, Eli van der Sluis, Stefan Seemayer, Singharoy Abhishek, Klaus Schulten, Johannes Söding and Roland Beckmann
A structural model of the active ribosome-bound membrane protein insertase YidC

Table of Content

Abstract.....	1
1 Introduction.....	3
1.1 Translation and insertion of membrane proteins	3
1.2 Features of α -helical membrane proteins	7
1.3 Co-evolution on a structural level	10
1.4 Cryo-electron microscopy.....	13
1.5 Aims of the study	15
2 Materials and Methods.....	16
2.1 Molecular Cloning	16
2.1.1 Vectors and Organisms.....	16
2.1.2 Media and Supplements.....	16
2.1.3 Polymerase Chain Reaction.....	17
2.1.4 Enzymatic Digestion of DNA	18
2.1.5 Agarose Gel Electrophoresis.....	18
2.1.6 Gel Extraction.....	19
2.1.7 DNA Ligation.....	19
2.1.8 Transformation of Calcium Competent <i>E.coli</i> Cells.....	19
2.1.9 Plasmid Preparation.....	20
2.1.10 Sequencing of DNA	20
2.1.11 Gene synthesis	20
2.2 Protein biochemistry.....	21
2.2.1 Expression of <i>E.coli</i> YidC	21
2.2.2 Purification of <i>E.coli</i> YidC	21
2.2.3 Expression of F ₀ C-RNCs.....	23
2.2.4 Purification of F ₀ C-RNCs.....	23
2.2.5 Disulphide Crosslinking	24
2.2.6 Preparation of Lipid/Cholat Mix	25
2.2.7 Reconstitution of YidC into Nanodiscs	25
2.3 Protein Analysis.....	27
2.3.1 SDS-Polyacrylamide Electrophoresis	27
2.3.2 NuPAGE® Gel Electrophoresis / Semi-Wet Western Blotting.....	27
2.3.3 Staining of Protein Gels.....	27
2.3.4 Western Blotting and Antibody Detection	28
2.4 <i>In vivo</i> Assay	29

2.4.1	Complementation Assay.....	29
2.5	Structural model of YidC.....	30
2.5.1	Calculation of Evolutionary Coupling Scores.....	30
2.5.2	Lipid Exposure Prediction.....	30
2.5.3	Building a Molecular Model of YidC.....	30
2.6	Molecular Dynamics Simulation.....	32
2.6.1	MD Simulation of the Structural Model of YidC.....	32
2.7	Structure Determination.....	33
2.7.1	Sample Preparation of the F ₀ c-RNC:YidC Complex.....	33
2.7.2	Cryo-Electron Microscopy.....	33
2.7.3	Data Import and Quality Check.....	33
2.7.4	Automated Particle Picking using SIGNATURE.....	34
2.7.5	MAPPOS.....	35
2.7.6	Simulation of Cryo-EM Images.....	35
2.7.7	Single Particle Analysis.....	36
2.7.8	Structure Refinement.....	36
2.7.9	Generation of a Non-Ribosomal Reference (Edge-volume) for Sorting.....	36
2.7.10	Sorting.....	37
2.7.11	Modelling the F ₀ c-RNC:YidC Complex.....	37
2.7.12	Figures.....	37
3	Results.....	38
3.1	Generation of an Automated Pre-Processing Pipeline.....	38
3.1.1	Automated Particle Picking using SIGNATURE.....	39
3.1.2	Machine Learning Algorithm for Particle Post Picking.....	39
3.1.3	Simulation of Cryo-EM Images.....	41
3.1.4	Comparison of MAPPOS to Human Experts on Simulated Data.....	42
3.1.5	Performance of MAPPOS on Real Cryo-EM Data.....	43
3.2	<i>In vitro</i> Reconstitution of a YidC Dependent Insertion Intermediate.....	45
3.2.1	Purification of F ₀ c-RNCs.....	45
3.2.2	Purification of YidC.....	47
3.2.3	Crosslinking Studies of Reconstituted YidC:F ₀ c-RNCs Complexes.....	48
3.3	Building a Structural Model of YidC.....	50
3.3.1	Evolutionary Covariance Matrix of the Conserved Core of YidC.....	51
3.3.2	Lipid-Protein Interface and Helix Arrangement.....	54
3.3.3	Three Dimensional Model of YidC.....	55
3.4	Validation of the Model.....	57
3.4.1	Molecular Dynamics Simulation of YidC.....	57
3.5	Cryo-EM Structure of the Active YidC bound to F ₀ c-RNCs.....	62
3.5.1	Structural Model of YidC Bound to the Translating Ribosome.....	66
3.5.2	Interaction of YidC with the Ribosome.....	68

3.6	Reconstitution of YidC into Nanodiscs.....	72
4	Discussion	75
4.1	A fast and accurate Pre-Processing Workflow for Single Particle Analysis.....	75
4.2	Hybrid Method Approach for Membrane Protein Modelling	77
4.3	Co-translational Protein Insertion by YidC	82
5	Summary and Outlook	86
6	References	88
7	Acknowledgements	96
8	Curriculum Vitae	97

Abstract

Cryo-electron microscopy (cryo-EM) studies using single particle reconstruction are extensively used to reveal structural information on macromolecular complexes. Aiming at the highest achievable resolution, a new generation of electron microscopes automatically acquire thousands of high-quality micrographs. Particles are detected on and boxed out from each micrograph using fully- or semi-automated approaches. However, the obtained particles still require laborious manual post-picking classification, which is one major bottleneck in single particle analysis of large datasets. In this study, a supervised post-picking strategy for the classification of boxed particle images was developed. The Machine learning Algorithm for Particle POST-picking (MAPPOS) employs machine learning techniques to train a robust classifier from a small number of characteristic image features. Comparisons between MAPPOS and manual post-picking classification by several human experts demonstrated that a trainings dataset of a few hundred sample images is sufficient to classify an entire dataset with a human-like performance. MAPPOS was shown to greatly accelerate the throughput of large datasets by reducing the manual workload by orders of magnitude while maintaining a reliable identification of non-particle images.

The integration of most membrane proteins into the cytoplasmic membrane of bacteria occurs co-translationally. The universally conserved YidC protein mediates this process either individually as a membrane protein insertase, or in concert with the SecY complex. In this study, a structural model of YidC was build, based on evolutionary co-variation analysis, lipid-versus-protein-exposure and molecular dynamics simulations. The model suggests a distinctive arrangement of the conserved five transmembrane domains and a helical hairpin between transmembrane segment 2 (TM2) and TM3 on the cytoplasmic membrane surface. The transmembrane domain generates a hydrophilic cavity within the lipid bilayer which is sealed towards the periplasm by strong hydrophobic stacking interactions. Hydrophobic mismatch between short TM helices induce a thinning of the membrane close to TM3 and TM5. The structural model was docked into a cryo-electron microscopy reconstruction of a YidC:ribosome nascent chain (RNC) complex displaying the YidC substrate F_{0c} . This structure reveals the interaction sites of a single copy of YidC with the ribosome at the ribosomal tunnel exit and locates the inserted helix close to TM3.

Taken together, the data suggests a catalytic function of YidC during membrane insertion by lowering the energy barrier for the translocation of hydrophilic moieties across the membrane.

1 Introduction

1.1 Translation and insertion of membrane proteins

All living cells use lipid bilayers to separate cellular processes from the environment and to form specialized compartments within the cell. Essential metabolic pathways, such as oxidative phosphorylation and photosynthesis, as well as the transport of soluble molecules and signal transduction are facilitated by membrane embedded proteins. These proteins are generally inserted co-translationally into the membrane by dedicated proteins known as translocons or insertases (Figure 1).

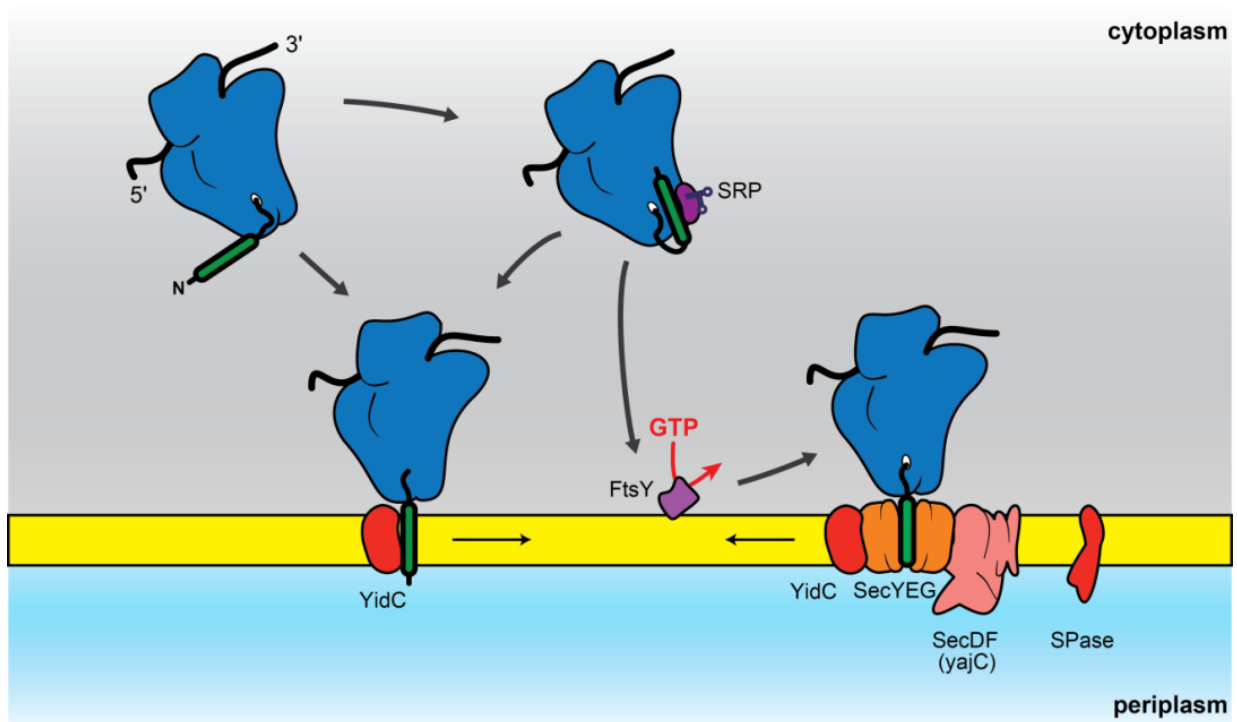


Figure 1: Co-translational membrane protein targeting and insertion in bacteria

Ribosome nascent chain (RNC) complexes (blue) translating proteins for Sec-dependent membrane insertion are recognized by SRP (purple) and targeted by its membrane bound receptor FtsY (purple) to the cytoplasmic membrane (yellow). IMPs are translated by SecYEG (orange) bound ribosomes and insertion of TM segments (green) is facilitated by SecY. Some membrane proteins need the interaction with YidC (red) and SecDF(yajC) (pink) for proper insertion. RNCs translating YidC-only substrates can be targeted to the membrane in a SRP-dependent or independent pathway. Figure adapted from (Driessen and Nouwen, 2008)

Most nascent inner membrane proteins (IMPs) are targeted by the signal recognition particle (SRP) in an evolutionary conserved pathway to the membrane.

SRP binds to hydrophobic stretches of nascent chain emerging from a ribosome and facilitates targeting of the RNC to the inner membrane by recognition and interaction with the membrane-associated SRP receptor. After proper targeting, SRP dissociates from the ribosome and the nascent chain is transferred to the translocon (for review see (Herskovits et al., 2000; Cross et al., 2009)).

In bacteria, co-translational membrane protein insertion is facilitated predominantly by the trimeric SecYEG translocon. The translocation channel is formed by the SecY subunit. It consists of two lobes which could open like a clam, thereby forming a lateral gate towards the lipid bilayer (Clemons Jr et al., 2004; Van den Berg et al., 2004; Bostina et al., 2005). Structural (Becker et al., 2009; Frauenfeld et al., 2011; Gogala et al., 2014) and biochemical data (Sato et al., 1997; Duong and Wickner, 1998) support a model for nascent TM helix insertion into the membrane by an opening of the lateral gate of SecY whereas periplasmic loops cross the membrane via a central hydrophilic pore (for review see (Park and Rapoport, 2012)).

In addition to SecYEG, YidC was identified to act as membrane protein insertase in combination or independent of SecYEG (Bonney et al., 1994; Samuelson et al., 2000; Scotti et al., 2000; Lührink et al., 2001). Members of the YidC/Oxa1/Alb3 family (Figure 2) are conserved in all kingdoms of life and fulfil many physiological functions (for review see (Dalbey et al., 2014)).

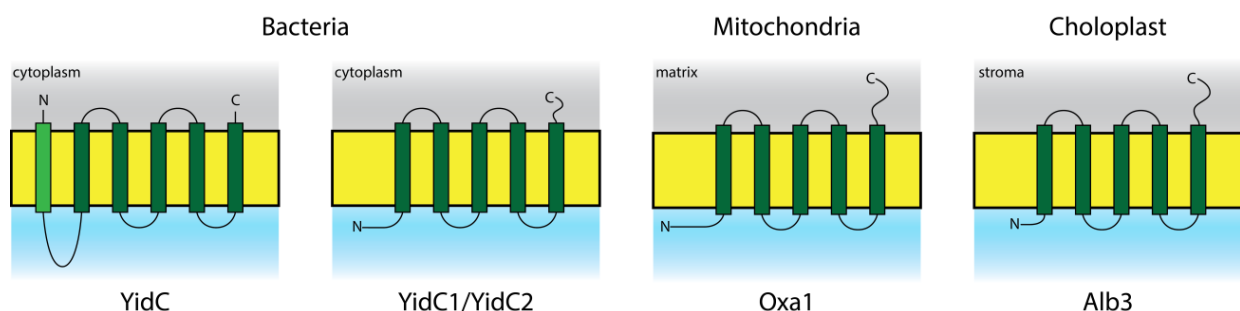


Figure 2: The highly conserved YidC/Oxa1/Alb3 family

Membrane topology of the different YidC homologues. The conserved TM segments (dark green) are located in the lipid bilayer (yellow). Proteobacteria (e.g. *E.coli*) contain only a single copy of YidC which has as additional N-terminal TM helix (light green) connected by a large periplasmic domain (P1 domain) to the conserved membrane core. Other bacteria (e.g. *B.subtilis*) harbor two versions of YidC (YidC1/YidC2) in their genome. Both of them consist of the conserved five TM helices but vary in length of their C-terminal tail. In mitochondria, the extended and positively charged C-terminus of Oxa1 is necessary for ribosomal binding. Alb3 in chloroplasts has also an extended C-terminus of unknown function. (adapted from (Kuhn et al., 2003))

They have been described to participate in membrane protein folding, assembly and quality control (Samuelson et al., 2000; Beck et al., 2001; Nagamori et al., 2004; van Bloois et al., 2008).

In the Sec-dependent pathway, YidC might be located in front of the lateral gate of SecY (Sachelaru et al., 2013), thereby mediating the partitioning of nascent TM segments from SecY into the lipid bilayer (Urbanus et al., 2001; van der Laan et al., 2001). The interaction with SecYEG is thought to be mediated via the accessory SecDFYajC complex (Nouwen and Driessen, 2002). The proper insertion and folding of some polytopic membrane proteins depend on YidC. In this case, YidC is proposed to function as an assembly site for the packing of TM segments of multi-spanning membrane proteins (Nagamori et al., 2004; Wagner et al., 2008; Zhu et al., 2013). Furthermore, YidC is involved in the assembly of oligomeric membrane complexes (van der Laan et al., 2003; Wickström et al., 2011). The topology of polytopic membrane proteins correlates with the distribution of positive charged residues in the loops which lead to the positive-inside rule (von Heijne, 1989). Membrane proteins with TM segments not following this rule were shown to be dependent on YidC for proper membrane insertion and folding (Gray et al., 2011). Thus, the exact contribution of YidC remains to be unclear.

In the Sec-independent pathway, YidC alone is sufficient for membrane insertion of a small but essential subset of membrane proteins (for review see (Dalbey et al., 2011)). In this function it resembles Oxa1 which is the only translocon for mitochondrially translated membrane proteins (Glick and Heijne, 1996). Substrates of the YidC-only insertion pathway are single- and double-spanning proteins possessing only a short hydrophilic moiety that has to be translocated across the membrane (Figure 3) (van der Laan et al., 2003; Serek et al., 2004; van der Laan et al., 2004; Facey et al., 2007). It is not completely established whether the insertion of YidC-only substrates is exclusively co-translational and whether SRP is needed for membrane targeting (Kiefer and Kuhn, 2007; Seitz et al., 2014).

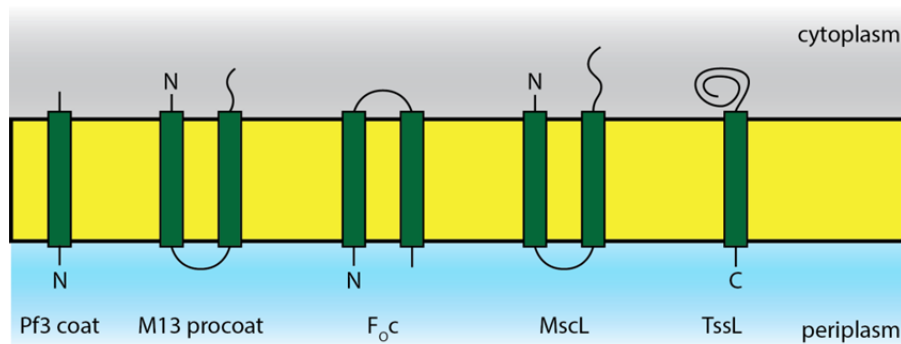


Figure 3: Substrates of the YidC-only pathway

Topology of proteins inserted via the YidC-only pathway. The TM segments (green) are located in the lipid bilayer (yellow) and termini of the proteins are indicated (N, N-terminus; C, C-terminus). The single-spanning Pf3 coat protein, the double-spanning M13 coat protein, the subunit c (F_0c) of the ATP synthase, the mechanosensor protein MscL and the C-tailed protein TssL are shown. (adapted from (Kuhn et al., 2003))

Members of the YidC/Oxa1/Alb3 family share a conserved core of 5 TM helices, which define the insertase function (Jiang et al., 2003). In *E.coli*, YidC contains an N-terminal extension to the core consisting of a large periplasmic domain (P1 domain) and an additional TM helix (Sääf et al., 1998). High resolution structural data is restricted to the non-essential periplasmic P1 domain (Oliver and Paetzel, 2008; Ravaud et al., 2008) and a 10 Å projection map of a 2D crystal of the full length YidC (Lotz et al., 2008). Cryo-EM reconstruction of YidC/Oxa1 bound to a ribosome displaying nascent chains of the YidC-only pathway could only provide contradictory hints concerning the oligomeric state of ribosome bound YidC (Kohler et al., 2009; Seitz et al., 2014). Deletion and mutagenesis studies based on the proposed topology of YidC were carried out to define functionally important regions/residues. A variety of deletion constructs revealed that the core domain of the last 5 TM helices is sufficient for function (Jiang et al., 2003). Surprisingly, most of the residues within this conserved core are tolerant to substitutions of single amino acids or longer amino acid stretches (Jiang et al., 2003). Residues involved in substrate binding are located in the TM core of YidC (Chen et al., 2002; Yu et al., 2008). In particular, residues in TM 3 were shown to contact nascent TM segments (Klenner et al., 2008; Yu et al., 2008), as well as residues in TM 4 and TM 5 (Klenner and Kuhn, 2012).

1.2 Features of α -helical membrane proteins

After insertion into the membrane, the stability of membrane protein depends on various physicochemical interactions with itself and the surrounding lipid bilayer (Figure 4 **A**). Lipid membranes are natural boundaries for cellular processes with defined biological and chemical conditions. The integrity of the lipid bilayer is achieved by the biphasic nature of lipid molecules. Aliphatic chains of the lipids generate a hydrophobic environment which is confined by the polar or charged head groups that interact with the aqueous environment on both sites of the membrane. The requirements for insertion and folding of membrane proteins are met by a defined amino acid distribution along transmembrane helices (Figure 4 **B**) (for review see (Cymer et al.)).

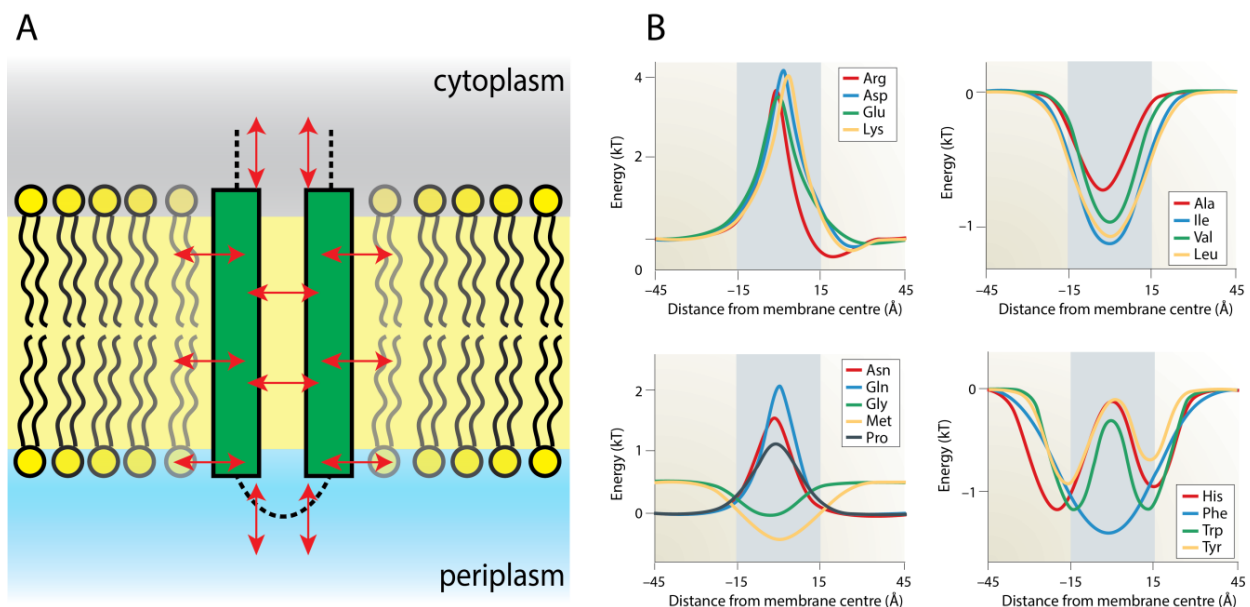


Figure 4: Interactions of membrane proteins with the lipid bilayer

A. Stabilizing interactions of membrane proteins. The TM segments and lipid bilayer adjust structurally to minimize the free energy. Important interactions are formed between (i) hydrophobic amino acids of the TM segment to the apolar lipid tails, (ii) residues at the ends of the TM segment to the lipid/water interface and (iii) direct TM-TM interactions. (adapted from (White et al., 2001)) **B.** Statistical free energy of insertion for amino acids calculated from a set of high-resolution X-ray structures (Ulmschneider et al., 2005). The higher the energy for an amino acid at a given position, the lower the probability to find it at this position in a protein structure.

The TM helix can be divided in three segments: (i) flanking regions that interact with the aqueous environment (ii) the interface of polar head groups and hydrophobic membrane core and (iii) the aliphatic lipid core. Certain amino acids have a preferred position along these segments. Hydrophobic side chains tend to

locate in the core of the TM helix, thereby facilitating the interaction with the apolar lipid tails. At the lipid–water interface, aromatic residues define the boundaries of the TM segment towards the aqueous solution. Charged and polar residues are predominantly located outside of the TM segment, but they can nevertheless be found in special cases within TM helices facilitating TM–TM interaction, sensing voltage or disrupting the lipid–water interface (Freites et al., 2005; Meindl–Beinker et al., 2006; Hristova and Wimley, 2011; Li et al., 2013). This knowledge can be exploited to predict TM segments based on the hydrophobicity distribution of the primary amino acid sequence (Hessa et al., 2005) which will be located in the lipid bilayer. Upon folding of polytopic or dimerization of single–spanning membrane proteins these α -helical TM segments have to specifically interact.

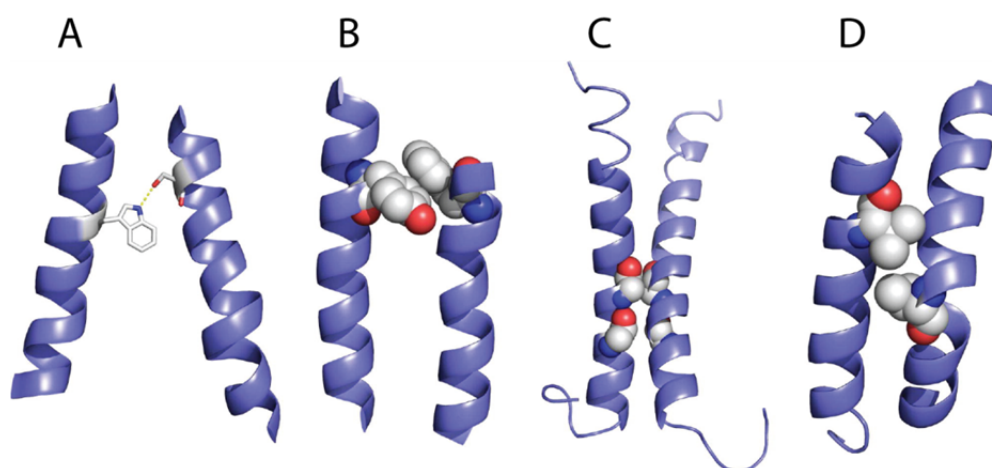


Figure 5: Common helix-helix interaction types in transmembrane helices

A. Hydrogen bond between tryptophan and serine residues in helices from the β -adrenergic GPCR (PDB code 2rh1). **B.** Aromatic stacking interactions between tryptophan residues in sensory rhodopsin (PDB code 1xio). **C.** GXXXG motif interaction in glycophorin A (PDB code 1afo). **D.** Valine residues form a knob in contact patch interaction in the mitochondrial ADP/ATP carrier (PDB code 1okc). Figure adopted from (Nugent and Jones, 2012b).

The interaction of TM helices is an important determinant for membrane protein folding, function, dynamics and structural classification. Analysis of crystal structures and clustering them according to their three-dimensional similarity revealed five classes of specific interhelical interactions – two types of packing motif, hydrogen bonds, salt bridges and aromatic interaction (Harrington and Ben–Tal, 2009).

TM residues involved in hydrogen bonding (Figure 5 **A**) are highly conserved (Hildebrand et al., 2008) and have a central role for folding, stabilization and function of helical membrane proteins (Senes et al., 2001; Adamian and Liang, 2002). A network of weak hydrogen bonds with alternative binding partners within

the TM domain or water filled cavities might be important to provide flexibility for the membrane protein (Joh et al., 2008; Bondar and White, 2012).

Aromatic interactions include stacking interactions of two residues with aromatic rings (Figure 5 **B**) as well as cation- π interactions (Shi et al., 2002; Johnson et al., 2007). They contribute to interactions of TM helices close to the lipid/water interface (Sal-Man et al., 2007).

Two classes consist of specific knob-in-hole type packing motifs (Figure 5 **C**, **D**). Strong van der Waals interactions are formed by side chains with limited conformational flexibility (knobs) filling a cavity (hole) of the interacting helix. These interactions allow tight packing of TM helices which are important for protein stability. Frequently studied examples are the GXXXG motif (Lemmon et al., 1992) or the heptad motif of leucine residues (Gurezka et al., 1999).

Salt bridges between residues of TM helices are very strong (Honig and Hubbell, 1984) and can be found in acid-sensing or voltage-gated ion channels (Palczewski et al., 2000; Abramson et al., 2003).

1.3 Co-evolution on a structural level

Co-evolution can be defined as interdependent evolutionary changes of two entities which play an important role in biological systems from ecosystems to molecules (Ochoa and Pazos, 2014). Co-evolution was first described on species level (Ehrlich and Raven, 1969) where two species show related changes within their interacting features. The growing number of genome sequences covering all kingdoms of life from bacteria to human and the accompanying opportunity to generate a diverse multiple sequence alignment (MSA) was the key for applying the concept of co-evolution to study molecular interaction solely based on sequence data (Marks et al., 2011). The co-evolutionary signals between proteins in larger complexes are very strong in cases where the evolutionary pressure is higher only for some parts of the complex. Good examples are complexes where the individual proteins are encoded in different cellular compartments (nucleus – mitochondria; nucleus – chloroplast). The components of the NADH-ubiquinone reductase complex and the rRNA and r-proteins of the mitochondrial ribosome are nuclear- as well as mitochondria-encoded. The nuclear-encoded proteins show an elevated evolutionary rate to compensate for the intrinsically higher evolution rate of their mitochondria-encoded counterparts (Gershoni et al., 2010; Barreto and Burton, 2013). The same principle was discovered for the proteins of the RuBisCo complex which are nuclear- and chloroplast-encoded (Pei et al., 2013). Another example for evolutionary linked mutations is the interaction of transcription factors to their respective DNA-binding site (Kuo et al., 2010; Yang et al., 2011). The general concept of co-evolutionary coupling can also be transferred to a single residue level of proteins and protein complexes (Marks et al., 2011; Hopf et al., 2014; Ovchinnikov et al., 2014). Assuming that evolutionary coupled residues are in close spatial proximity, then this provides structural restraints that might be sufficient for *ab initio* protein folding without prior structural knowledge (Marks et al., 2011).

Upon folding of a linear amino acid chain into its final three dimensional conformation, specific interactions of the protein backbone for building secondary structure elements, as well as direct tertiary contacts of side chains which might be far away in the linear sequence must be established. The direct interaction of these residue pairs is crucial for protein function and therefore shows evolutionary correlations at these positions (Hopf et al., 2012). The information about co-evolutionary interactions can be retrieved from large MSAs which are based on the rapidly rising number of genome sequences (Marks et al., 2011). The MSA

contains additionally a variety of information about important positions of the protein that can be extracted (Figure 6). Certain residues are highly conserved among different species and therefore show no evolutionary coupling to other residues.

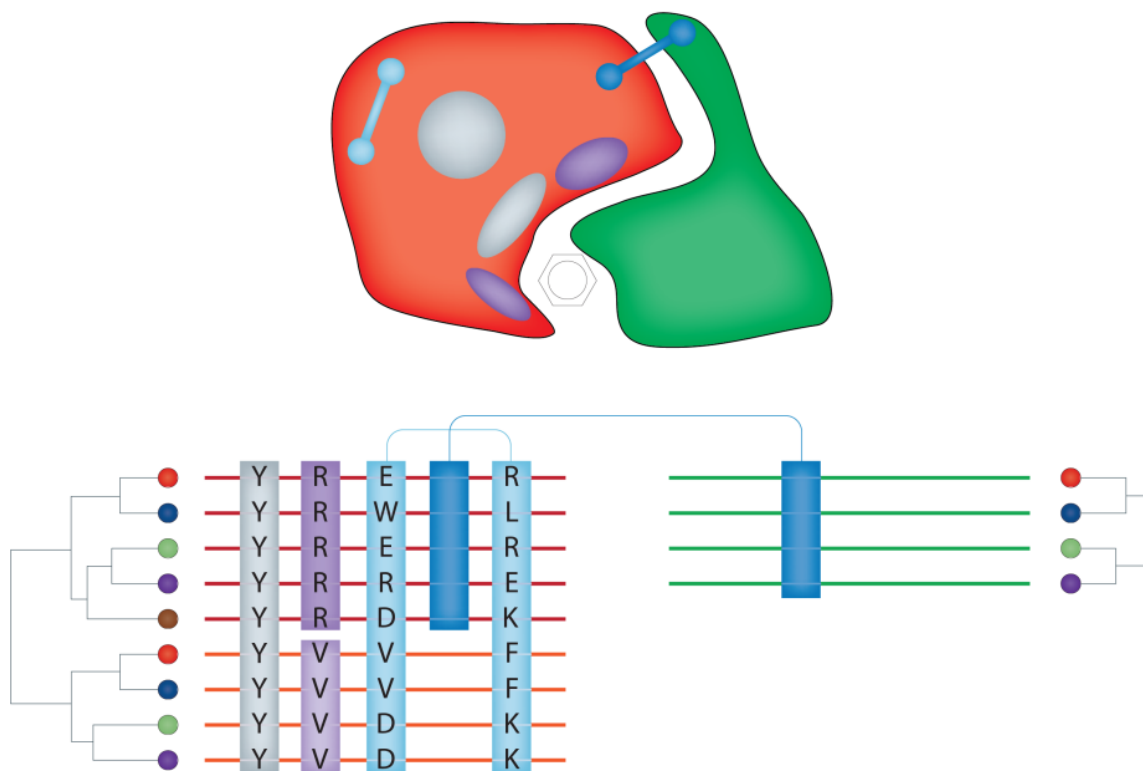


Figure 6: Features extracted from MSA

Two interacting proteins (red and green) and their MSA are schematized. Conserved positions (grey) are located in the protein core and functional regions (interaction sites, catalytic sites). Specificity-determining positions (SPDs, purple) tend to be close to functional site. Correlated mutations within a protein (light blue) reflect residue pairs in close special proximity and co-evolution of position between two proteins indicate potential interaction sites (dark blue). Adapted from (de Juan et al., 2013)

These conserved positions are indicative of an important role in protein stability or function at interaction or catalytic sites. More important for co-evolutionary analysis are positions in the MSA that show covariance in different species. Using global statistical methods (Marks et al., 2011; Nugent and Jones, 2012a) it is possible to distinguish direct (A-B, B-C) from transitive (A-C, linked by B) correlations and thereby eliminating false positive predictions. MSAs consisting of two different proteins can be used to retrieve evolutionary coupled residue pairs within the proteins that are crucial for protein interaction (Hopf et al., 2014; Ovchinnikov et al., 2014). The coupling score for every possible residue pair of the MSA is calculated and visualized in a contact map (see Figure 7). High coupling scores

can be due to (i) direct interaction within a monomeric protein, (ii) interacting residues of homo-oligomers and (iii) interacting residues in different conformational states. Diagonal or anti-diagonal patterns of higher coupling scores are indicative of interacting secondary structure elements (α -helices, β -sheets).

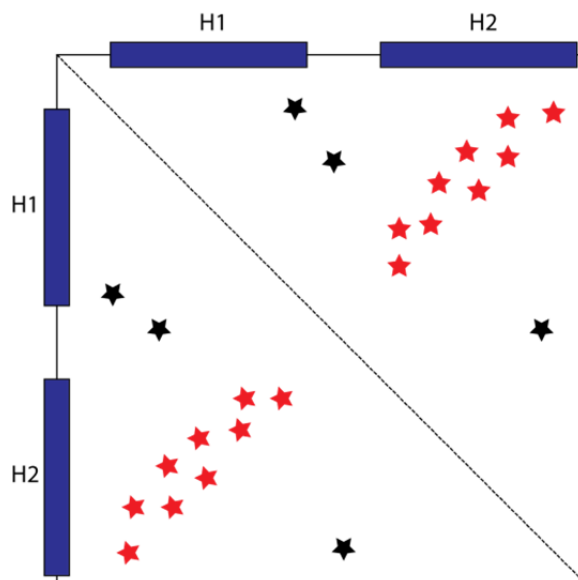


Figure 7: Contact map of coupling scores

The axis consists of the consensus sequence of the MSA. Residue pairs of high coupling scores are indicated as stars. The red coupling scores are indicating an anti-parallel conformation of H1 and H2. The black stars are violating the topology and are most probably false positive predictions of weak evolutionary coupling (adapted from (Hopf et al., 2012)).

Despite the enormous data that can already be retrieved from sequence information, there are still some limitations for the method. The quality of the MSA is the most crucial as it serves as input for all downstream analysis. There have to be enough sequences which are sufficiently diverse for sampling the complete protein sequence. This already excludes most of the eukaryotic specific proteins where the number of sequenced genomes is limited. Here, co-evolution analysis can give hints which new genome might give the most additional information and therefore help to guide high-throughput sequencing.

1.4 Cryo-electron microscopy

Macro-molecular complexes are involved in essential cellular processes. Detailed structural information allows revealing the molecular mechanisms of these complexes. This information can then be used for structural based drug design or the development of new biochemical tools. The determination of structural information of macromolecular complexes made enormous progress during the past decade which is directly linked to recent developments in cryo electron microscopy (cryo-EM). The strength of cryo-EM compared to other structural methods like X-ray crystallography or NMR is that the structure of the biological complex can be determined in its native environment, based on a few μg of purified complex. The samples for cryo-EM are vitrified on holey carbon grids before micrographs are recorded on a transmission electron microscope. Single particle analysis is then used to calculate the three dimensional structure of the complex.

For preparation of cryo-EM samples, the purified complex is directly applied to EM-grids. Blotting of the grid ensures a monolayer of randomly oriented complexes. The aqueous sample is then flash frozen in liquid ethane at high cooling rates to form amorphous ice. The formation of ice crystals which are opaque for the electron beam is prevented by keeping the samples below 130 K during all subsequent steps. Images are recorded using a cryo electron microscope under low dose conditions ($\sim 20 \text{ e}/\text{\AA}^2$) to minimize radiation damage. The small density difference between complex and water in combination with the low dose leads to a low signal-to-noise ratio (SNR) in the final image. To improve the signal, images are aligned and projections of the same orientation are averaged. The electron detection device strongly influences the quality and processing of the data. Datasets recorded on film are manually collected and digitized before further processing using single particle software. This limits the amount of available particles and therefore restricts the resolution of the final density maps. Only automated data acquisition using CCD-cameras enables the collection of large datasets ($>10,000$ micrographs; $>1,500,000$ particles) within a few days which is needed for high resolution reconstructions (Anger et al., 2013). The real breakthrough for high resolution was the recent introduction of the direct detection device (DDD) which records incoming electrons directly without intermediate conversion of electrons to photons using a CCD-camera (Kuhlbrandt, 2014).

The micrographs collected on the electron microscope contain the two-dimensional (2D) projections of randomly oriented particles within the amorphous

ice. In order to calculate the three-dimensional (3D) structure of the complex, the individual particles have to be detected, windowed out from the micrograph and subjected to single particle analysis. During this analysis the original 3D orientation (Euler angles) of the 2D projection has to be determined. The strategy for calculation depends on the structural knowledge of the sample.

Structures of unknown complexes can be determined using sophisticated 2D analysis including random conical tilt (Radermacher et al., 1987), classification, clustering and backprojection of stable classes into 3D (van Heel and Keegstra, 1981; Hohn et al., 2007; Elmlund and Elmlund, 2012). The correctness of the 3D volume can then be assessed using tilt-pair analysis (Henderson et al., 2011).

Single particle analysis of complexes with at least partially known structures can use this initial information and start directly with projection matching. Thereby, the 3D density map is projected into 2D in all possible orientations resulting into a reference set of projections of known orientation. The particles from the micrographs are aligned to these reference projections and assigned based on their cross-correlation (CC). Each particle is then back-projected using these alignment parameters (x-,y-shift, Euler angles) resulting into a 3D density map. The accuracy of the alignment parameters and therefore the resolution of the 3D map can be improved by iterating projection, alignment and back-projection using smaller sampling of the reference projections.

Cryo-EM captures the structure of a complex in its native conformations in contrast to the crystal packing of X-ray crystallography. Therefore, structural heterogeneity and flexibility of parts of the complex are also present in the 2D projections. 3D reconstructions using all projections will therefore be limited in resolution in exactly these parts. Classification of the dataset into stable sub-groups with defined stoichiometry or showing a stable conformation can restore this information. This can be achieved by sorting of the dataset using competitive projection matching or maximum likelihood approaches (Scheres, 2012).

Co-translational events such as membrane protein targeting and insertion or regulatory mechanisms on the ribosome are bona fide complexes for cryo-EM investigations.

1.5 Aims of the study

Membrane proteins perform diverse functions from exchange of metabolites to interaction of the cell with the environment. Almost one third of the proteome from bacteria to human consist of membrane proteins which have to be inserted into the lipid bilayer by translocons. The conserved Sec translocon facilitates translocation of secretory proteins across and insertion of membrane proteins into the membrane. In addition, the membrane protein insertase of the YidC/Oxa1/Alb3 family found in bacteria, mitochondria and thylakoidal membranes is involved in folding and insertion of membrane proteins. Despite its importance, little is known about the mechanism of co-translational membrane protein insertion by YidC. Biochemical studies identified functionally important sites of the protein but the available structural data did not allow the development of a comprehensive model. The aim of this study was to reveal structural information of an active YidC:ribosome complex engaged in co-translational insertion. Providing a structural model of YidC-mediated protein insertion would improve the understanding of existing biochemical data and allow structure-guided analysis of the insertion pathway.

2 Materials and Methods

2.1 Molecular Cloning

2.1.1 Vectors and Organisms

Constructs were cloned into standard vectors (Table 1) and *E.coli* strains were used for expression and molecular cloning (Table 2).

Table 1: Vectors

Name	Company	Selection marker	Vector used for
pET-16b	Novagen	Ampicillin	YidC purification (see 2.2.2)
pTrc-99a	Pharmacia	Ampicillin	<i>In vivo</i> complementation assay (see 2.4.1)
pBad	Invitrogen	Ampicillin	YidC purification (see 2.2.5)
pBad	Invitrogen	Ampicillin	F ₀ C-RNC purification (see 2.2.4)

Table 2: Organisms

Strain	Generated by	Strain used for
<i>E.coli</i> C43 (DE3)	Miroux and Walker 1996	YidC expression (see 2.2.1)
<i>E.coli</i> FTL10	Hatzixanthis et al. 2003	<i>In vivo</i> complementation assay (see 2.4.1)
<i>E.coli</i> KC6	Seidelt et al. 2009	F ₀ C-RNC expression (see 2.2.2)
<i>E.coli</i> DH5 α		Molecular cloning

2.1.2 Media and Supplements

Bacteria were grown in liquid LB medium (1% NaCl, 1% tryptone, 0.5% yeast extract) or on LB agarose plates supplement with 1% agarose. All media were supplemented with the appropriate antibiotic(s) for plasmid selection (ampicillin 100 μ g/ml; kanamycin 50 μ g/ml) during cloning and expression.

2.1.3 Polymerase Chain Reaction

The polymerase chain reaction (PCR) was used to amplify DNA fragments for cloning and site-directed mutagenesis. Therefore, KOD Hot Start DNA Polymerase (Merck, Millipore) was mixed with primers, template DNA and water according to the manufacturer's protocol. A standard touch-down PCR program was used to amplify the DNA independent of actual T_M of the primers. The elongation time was adjusted according to the length of the desired product (0.5 min/kbp). The amplified product was purified using the QIAquick® PCR Purification Kit (Qiagen) according to the manufacturer's manual.

The PCR product for blunt-end ligation using primers with phosphorylated 5' end was further purified by agarose gel electrophoresis (see 2.1.5) and gel extraction (see 2.1.6).

Table 3: Primers

Name	Sequence
5'Apal-F ₀ c	TTT AAA GGT ACC ATG GAA AAC CTG
3'KpnI-F ₀ c	CAC AGG GCC CAG CGT AAT CTG G
3'F ₀ c(G23C)	TGC CGC CAG ACC C
5' F ₀ c(G23C)	ATC GGT TGC GCG ATC GG
5'YidC_3C	ATG GAT TCG CAA CGC AAT CTT TTA GTC ATC GCT TTG C
3'YidC_3C	GCC ATA TCG AAG GTC GTC ATC TGG AAG TTC TGT TCC AGG GGC CC
5'YidC(M430C)	ATC CAG TGC CCA ATC TTC CTG GCG
3'YidC(C423S)	CAG CAG CGG GAA ACT GCC GCC CAG CGG G
5'YidC(Y377A)	AAA GCG CAG GCA ACC TCG ATG GCG
3'YidC(Y370A)	GGT CAG CGG TGC CAT GAT GCC
5'YidC(Y377F)	AAA GCG CAG TTT ACC TCG ATG GCG
3'YidC(Y370F)	GGT CAG CGG AAA CAT GAT GCC
5'YidC_TM2	GGC ATC ATG TAC CCG CTG ACC
3'YidC(R366A)	CGC AAC GAT AAA GGT GAT GAT GAT AAT GG
3'YidC(T362A)	ACG AAC GAT AAA CGC GAT GAT GAT AAT GG
5'YidC(F433A)	GCG CTG GCG TTG TAC TAC ATG C
5'YidC(Y438A)	TTC CTG GCG TTG TAC GCG ATG CTG ATG G

3'YidC_TM3	GCA GCC GCC CAG CGG G
5'YidC (M471A)	GCG GGC GTA ACG ATG TTC TTC ATT CAG
5'YidC (G472A)	ATG GCG GTA ACG ATG TTC TTC ATT CAG
5'YidC (T474A)	ATG GGC GTA GCG ATG TTC TTC ATT CAG
5'YidC (M475A)	ATG GGC GTA ACG GCG TTC TTC ATT CAG
3'YidC_TM4	CAG GAT CGG CAG GAT GTA GTA CG
5'YidC (F505A)	GCG TTC CTG TGG TTC CCG
5'YidC (F506A)	TTC GCG CTG TGG TTC CCG
5'YidC (F509A)	TTC TTC CTG TGG GCG CCG TCA GG
3'YidC_TM5	CAC GGT GAA GAT GAC C
5'YidC (Y516A)	GCG TAT ATC GTC AGC AAC CTG G
5'YidC (Y517A)	TAC GCG ATC GTC AGC AAC CTG G
3'YidC_TM6	CAG CAC CAG ACC TGA CGG
5'YidC (S520A)	ATC GTC GCG AAC CTG GTA ACC
5'YidC (N521A)	ATC GTC AGC GCG CTG GTA ACC
5'YidC (T524A)	CCT GGT A GCG AT TAT TCA GC
3'YidC_TM6II	ATA GTA CAG CAC CAG ACC TGA CGG

2.1.4 Enzymatic Digestion of DNA

Restriction endonucleases were used to digest DNA at specific positions for cloning. To that end, restriction enzymes (NEB) were mixed with DNA, buffer and water and incubated at the appropriate temperature and time according to the manufacturer's protocol. The digested DNA was separated using agarose gel electrophoresis (see 2.1.5) and the desired fragment was extracted from the gel (see 2.1.6).

2.1.5 Agarose Gel Electrophoresis

DNA samples of PCR products or after restriction digestion were separated using agarose gel electrophoresis. Agarose gels with varying agarose concentration (0.8 % – 1.2 %) depending on the size of the desired DNA product were prepared by dissolving UltraPure™ Agarose (Invitrogen) in TAE buffer. DNA samples were mixed with 6 x loading dye (Fermentas) and stained with SybrSafe according to

the manufacturer's protocol. The gels were run at 100 V in TAE buffer for 30 – 45 min and visualized using the INTAS UV system.

TAE buffer

40 mM Tris Base

20 mM acetic acid

2 mM EDTA pH 8.0

2.1.6 Gel Extraction

The bands corresponding to the desired DNA fragments were cut out from agarose gels under UV light illumination. A QIAquick® gel Extraction Kit (Qiagen) was used to purify the DNA according to the enclosed manual and the product was eluted in 30 µl nuclease free water.

2.1.7 DNA Ligation

The T4 DNA ligase (New England BioLabs) was used to conjugate either restricted insert DNA with restricted vector backbone or PCR products with phosphorylated 5' ends. To that end, vector and insert DNA or gel-purified PCR product were incubated with T4 DNA ligase according to the enclosed manual. The ligated product was used to transform calcium competent *E.coli* DH5α cells (see 2.1.8).

2.1.8 Transformation of Calcium Competent *E.coli* Cells

For transformation, 50 µl of competent cells were incubated with either 50 – 100 ng of plasmid DNA or 20 µl of ligation reaction for 5 min on ice. Cells were heat shocked for 45 s at 42 °C and immediately cooled on ice for an additional 5 min. After adding 700 µl of LB medium, the mix was incubated for 45 min at 37 °C. Cells were pelleted gently in a table top centrifuge (Eppendorf 5417R) for 1 min at 6,000 rpm, 600 µl of the supernatant were discarded and the cell pellet was resuspended in the remaining medium. 100 µl of resuspended cells were spread on LB plates containing the appropriate antibiotic(s) to select for the desired plasmid(s) and plates were incubated at 37 °C over night.

2.1.9 Plasmid Preparation

A single bacterial colony was picked from an agarose plate to inoculate 5 ml of LB medium containing the appropriate antibiotic(s). The culture was incubated shaking at 37 °C over night. Cells were pelleted and plasmid DNA was extracted using the QIAprep[®] Spin miniprep Kit (Qiagen) according to the manufacturer's manual.

2.1.10 Sequencing of DNA

For sequencing, plasmid DNA was sent to Eurofins MWG Operon (Ebersberg, germany) according to the company's instruction.

2.1.11 Gene synthesis

DNA sequences were synthesized by Eurofins MWG Operon (Ebersberg, Germany) and shipped in a vector harbouring an antibiotic resistance for selection.

2.2 Protein biochemistry

2.2.1 Expression of *E.coli* YidC

YidC and YidC variants were expressed from either pET-16 or pBAD plasmids in *E.coli* C43 cells and LB medium supplemented with 100 µg/ml ampicillin.

To that end, single colonies from agarose plates were used to inoculate pre-cultures of appropriate volumes. After growing over night at 37 °C and 150 rpm, the pre-cultures were used to inoculate pre-warmed expression medium to an OD₆₀₀ = 0.05 and incubated at 37 °C and 125 rpm for several hours to an OD₆₀₀ = 0.6. Depending on the plasmid, protein expression was induced by either 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) in the case of pET-16 or 0.2 % arabinose for pBAD plasmid. IPTG induced cultures were grown at 18 °C over night. In case of arabinose, protein was expressed at 30 °C for 3 h.

Cells were pelleted by centrifugation using a SLC-6000 rotor (Sorvall) for 10 min at 6,000 g and 4 °C. Cell pellets were resuspended in LB medium, transferred to 15/50 ml tubes and re-pelleted using a ROTANTA 46 R centrifuge (Hettich) for 15 min at 4,460 g and 4 °C. Supernatants were discarded, the remaining cell pellets were frozen in liquid nitrogen and stored at -80 °C until further purification (see 2.2.2).

2.2.2 Purification of *E.coli* YidC

YidC was expressed in the *E.coli* C43 strain, cells were harvested, frozen in liquid nitrogen and stored at -80 °C until purification (see 2.2.1). A frozen cell pellet was thawed on ice in lysis buffer. Cells were mechanically disrupted by three passages through a microfluidizer (Microfluidics) set to 17 kpsi and centrifuged for 20 min at 30,000 g and 4 °C in a SS-34 rotor (Sorvall) to remove cell debris. The cleared lysate was centrifuged in a Ti45 rotor (Beckmann-Coulter) for 30 min at 40,000 rpm and 4 °C to pellet the membrane fraction. The membranes were resuspended using a douncer and membrane proteins were solubilized by adding solubilisation buffer. Non-solubilized material was removed by centrifugation for 30 min at 40,000 rpm and 4 °C using the Ti45 rotor (Beckmann-Coulter). The solubilized membrane proteins were incubated with 0.25 ml cobalt-chelating matrix (Talon[®], Clontech) per liter culture for 1 h at 4 °C. The solution was loaded onto

a column and washed with 20 CV of washing buffer to remove non-specifically bound proteins. The poly-histidine-tagged YidC was bound to the matrix and eluted by two consecutive steps of adding 1 CV of elution buffer and incubation for 30 min at 4 °C on a turning wheel. The pooled fractions were dialysed over night at 4 °C against the dialysis buffer (0,1 L/liter culture). The N-terminal His-tag was removed by adding 3C protease according to the manufacturer's protocol directly into the elution fraction. Uncleaved protein was removed by rebinding to the metal affinity matrix. The protein without His-tag was concentrated, loaded on a Superdex S200 gel filtration column at 4 °C (GE Healthcare), peak fractions were pooled and concentrated to ~4 µm. The purified protein was immediately used for further biochemical or structural studies.

Lysis buffer

20 mM NaPO₄ pH 6.8
10 % glycerol
1 mM PMSF

Solubilisation buffer

20 mM NaPO₄ pH 6.8
100 mM NaCl
10 % glycerol
1 mM PMSF
1 % Cymal6

Washing buffer

20 mM NaPO₄ pH 6.8
100 mM NaCl
10 % glycerol
0.05 % Cymal6
25 mM imidazole

Elution buffer

20 mM NaPO₄ pH 6.8
100 mM NaCl
10 % glycerol
0.05 % Cymal6
250 mM imidazole

Dialysis buffer

20 mM NaPO₄ pH 6.8
100 mM NaCl
10 % glycerol
0.05 % Cymal6
1 mM β-mercaptoethanol

2.2.3 Expression of F₀c-RNCs

Nascent chains coding for the first transmembrane segment of F₀c were expressed in *E.coli* KC6 cells harbouring the pBAD plasmid with TnaC stalling sequence (Seidelt et al., 2009; Bischoff et al., 2014a).

Nascent chain sequence:

Poly-Histidine Tag – 3C Cleavage Site – F₀c – HA-Tag – TnaC stalling sequence

MGHHHHHHHDYDIPTTLEVLFGQPGT**MENLNMDLLYMAAAVMMGLAAIGAAIGIGILGGKFLEGAARQPDLIYPY**
 DVDPDYAGPNILHIS VTSKWFNIDNKIVDHRP

LB medium supplemented with 100 µg/ml ampicillin was inoculated with a single colony from an agarose plate and incubated over night at 37 °C and 150 rpm. The pre-warmed expression culture was inoculated to an OD₆₀₀ = 0.05 and incubated at 37 °C and 125 rpm. RNC expression was induced at an OD₆₀₀ = 0.5 by adding 0.2 % arabinose directly to the medium. Cells were harvested after 1 h by centrifugation using a SLC-6000 rotor (Sorvall) for 10 min at 6,000 g and 4 °C. Cell pellets were resuspended in LB-medium, transferred to 15/50 ml tubes and re-pelleted using a ROTANTA 46 R centrifuge (Hettich) for 15 min at 4,460 g and 4 °C. After discarding the supernatants, Cell pellets were frozen in liquid nitrogen and stored at -80 °C until purification (see 2.2.4).

2.2.4 Purification of F₀c-RNCs

RNCs were expressed *in vivo* in an *E.coli* KC6 strain, cells were harvested and cell pellets were frozen in liquid nitrogen before storage at -80 °C (see 2.2.2). Cell pellets were thawed on ice in 250 buffer supplemented with 1 mM tryptophane, 1 % DDM and 2 % (v/v) protease inhibitor (Roche). Cells were mechanically disrupted by three passages through a microfluidizer (Microfluidics) set to 17 kpsi and cell debris was removed by centrifugation for 20 min at 30,000 g and 4 °C in a SS-34 rotor (Sorval). The ribosomal fraction was pelleted (40,000 rpm, 30 min, 4 °C, Ti45) from the cleared lysate and resuspended in 250 buffer (1 ml/l expression culture) supplemented with 1 mM tryptophan. Metal affinity matrix (TalonR, Clontech) was equilibrated in 250 buffer and incubated for 1 h at 4 °C with the ribosomal fraction. The mixture was loaded on a column and washed with

10 CV 250 buffer supplemented with 1 mM tryptophan and 5 CV 500 buffer. RNCs were eluted in 1 CV elution buffer. The monosomal fraction was separated by applying the elution to a linear 10 % – 40 % sucrose gradient (16,000 rpm, 17,5 h, 4 °C, SW32), pelleted (40,000 rpm, 4 h, 4 °C, Ti45) and resuspended in 250 buffer. The presence of aminoacyl-tRNA was tested by western blotting using anti-HA antibodies (see 2.3.4) and RNCs were flash frozen in liquid nitrogen before storage at -80 °C.

250 buffer		500 buffer	
50 mM	HEPES pH 7.2	50 mM	HEPES pH 7.2
250 mM	KOAc	500 mM	KOAc
25 mM	Mg(OAc) ₂	25 mM	Mg(OAc) ₂
250 mM	sucrose	250 mM	sucrose
0.1 %	DDM	0.1 %	DDM

Elution buffer	
50 mM	HEPES pH 7.2
250 mM	KOAc
25 mM	Mg(OAc) ₂
100 mM	imidazole
0.1 %	DDM

2.2.5 Disulphide Crosslinking

The interaction of YidC with the nascent chain was tested by disulphide crosslinking of specific cysteine mutants.

F₀c(G23C)-RNCs with a single cysteine in the middle of the transmembrane helix were generated (see 2.1.3) and purified (see 2.2.2 and 2.2.4). Cysteine-free YidC(C423S) was used to create various YidC variants with single cysteins at specific positions. The YidC variants were purified (see 2.2.2) and various F₀c(G23C)-RNCs / YidC variant complexes were reconstituted by incubating 500 pmol of YidC variant with 100 pmol of F₀c(G23C)-RNCs for 30 min at 37 °C. Disulphide crosslinks were introduced by adding 1 mM 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB) for 10 min at 4 °C. The reaction was quenched by adding 20 mM

N-Ethylmaleimide (NEM) for 20 min at 4 °C. Crosslinked F₀C-RNC-YidC complexes were separated from unspecific crosslink products using a 10 % – 40 % linear sucrose gradient and subsequent harvesting of the 70S peak. The samples were analysed by NuPAGE® (see 2.3.2) and antibody detection against YidC and the nascent chain (HA-tag).

2.2.6 Preparation of Lipid/Cholat Mix

E.coli total lipid fraction (Avanti Polar Lipids, *E.coli* total extract in choloform) was dried under a stream of nitrogen to remove the organic solvent. Lipids were resuspended in ND-buffer containing cholate (1:1; lipids:cholate) to a final concentration of 20 mg/ml lipids. The suspension was vortexed and sonified until a clear solution was obtained. The mixture were frozen in liquid nitrogen and stored at -80 °C in aliquots á 20 µl.

ND-buffer

20 mM HEPES pH 7.2

100 mM KOAc

6 mM Mg(OAc)₂

1 mM DTT

2.2.7 Reconstitution of YidC into Nanodiscs

Freshly purified YidC (see 2.2.2) was reconstituted into nanodiscs by mixing 100 µg YidC, 600 µg Apo-A1 (scaffold protein) and 300 µg lipid/cholate mix (molar ratio 1:12:250) in reconstitution buffer. After incubating for 1 h at 37 °C, nanodisc formation was induced by adding Biobeads SM2 sorbent (Bio-Rad) and additional incubation for 4 h at room temperature. The mixture was subjected to gel filtration chromatography using a Superdex S200 10/30 column (GE Healthcare) and eluted in NDG-Buffer. Fractions were analysed by SDS-PAGE (see 2.3.1), Nd-YidC fractions were pooled and applied to a metal affinity column for enriching nanodisc-incorporated YidC.

Reconstitution buffer

20 mM HEPES pH 7.2

100 mM KOAc

6 mM Mg(OAc)₂

1 mM DTT

0.1 % Cymal 6

NDG-Buffer

20 mM HEPES pH 7.2

100 mM KOAc

6 mM Mg(OAc)₂

1 mM DTT

10 % glycerol

2.3 Protein Analysis

2.3.1 SDS-Polyacrylamide Electrophoresis

Protein samples were analysed by SDS-Polyacrylamide electrophoresis (SDS-PAGE) using discontinuous 15 % gels, featuring a stacking and separation gel. Samples were mixed with sample buffer, loaded on the gel and electrophoresis was performed at constant voltage of 60 – 220 V in running buffer using a Bio-Rad mini-Protean II Electrophoresis chamber (Biorad, Munich).

Sample buffer	Running buffer
50 mM Tris base pH 6.8	25 mM Tris base
2 % SDS	192 mM glycine
0.1 % bromophenol blue	0.1 % SDS
10 % glycerol	
100 mM DTT	

2.3.2 NuPAGE® Gel Electrophoresis / Semi-Wet Western Blotting

Protein samples of crosslinking experiments were analysed using the NuPAGE® electrophoresis system (Invitrogen). Proteins were separated on NuPAGE® Novex® 4 – 12 % Bis-Tris gradient gels (Invitrogen). Buffers and electrophoresis conditions were set up according to the enclosed manual.

After performing gel electrophoresis, proteins were transferred to PVDF membrane using semi-wet blotting in the XCell II™ Blot Module (Invitrogen). Buffers and transfer conditions were set up according to the enclosed manual. Antibody detection was performed as described in 2.3.4.

2.3.3 Staining of Protein Gels

Polyacrylamide gels were stained using SimplyBlue™ SafeStain (Invitrogen). Gels were boiled three times in water for 30 s in a microwave oven. The water was renewed each time. Afterwards gels were boiled in 15 ml staining solution for 30 s and kept at RT for imaging.

2.3.4 Western Blotting and Antibody Detection

Western blotting and antibody detection was used to detect untagged proteins (YidC, see 2.4.1), tagged proteins (His-tag) and nascent chains (HA-tag).

Proteins were separated using SDS-PAGE (see 2.3.1) and transferred on a PVDF membrane (Roth) using a standard semi-dry blotting apparatus (BioRad, Munich) for 50 min at 75 mA per gel.

The membrane was blocked for 1 h at RT with 5 % milk powder in TBS buffer. The 1st antibody (mouse anti-HA; 1:500 or mouse anti-His; 1:2000) was added to the solution and incubated over night at 4 °C. After washing 1x with TBS-T and 2 x with TBS for 10 min, the membrane was incubated with the 2nd antibody (goat anti mouse HRP; 1:2,000 in 5 % milk in TBS) for 1 h at RT. After washing 1x with TBS-T and 2 x with TBS buffer, the antibody was detected using the Chemiluminescent Detection Kit (AppliChem) together with films (Amerskam Hyperfilm ECL) or CCD camera (LAS 3000 mini, GE) according to the manufacturer's protocol.

TBS/TBS-T buffer

20 mM Tris-Cl pH 7.5

150 mM NaCl

1 % Tween-20 (only for TBS-T)

2.4 In vivo Assay

The effects of single amino acid substitutions of YidC were tested *in vivo*. To that end, *E.coli* cells having the genomically encoded YidC under the arabinose-inducible promotor were transformed with a plasmid of the YidC variants and tested for growth.

2.4.1 Complementation Assay

For *in vivo* complementation assays, untagged wild type *E.coli* YidC was cloned into a pTrc-99a vector (Pharmacia). All mutants were generated using touch-down PCR followed by blunt end ligation of the 5' end phosphorylated primers (see 2.1.3). *E.coli* FTL10 cells were transformed (see 2.1.8) with the resulting YidC variant plasmids, LB medium supplemented with 100 µg/ml ampicillin, 50 µg/ml kanamycin and 0.2 % arabinose was inoculated with single colonies from the transformation and grown shaking over night at 37 °C. All cultures were adjusted to an OD₆₀₀ of 0.1 and YidC depletion was induced by changing to LB medium supplemented with 100 µg/ml ampicillin, 50 µg/ml kanamycin, 0.2 % glucose. Subsequently, the cultures were incubated for 3 h at 37 °C. Stable expression of the variants was tested by western blotting (see 2.3.1 and 2.3.4) using anti-YidC antibody (Prof. Driessen, University of Groningen; 1st antibody 1:20,000 in 5 % milk in TBS buffer; 2nd antibody 1:20,000 in 5 % milk in TBS buffer). Dilution series of all constructs starting from OD₆₀₀ = 10⁻¹ to OD₆₀₀ = 10⁻⁵ were generated and each dilution was spotted on two LB agarose plate (see 2.1.2) supplement with 100 µg/ml ampicillin, 50 µg/ml kanamycin and either 0.2 % arabinose or 0.2 % glucose. The plates were incubated over night at 37 °C and imaged using a standard flatbed scanner.

2.5 Structural model of YidC

2.5.1 Calculation of Evolutionary Coupling Scores

All calculations were performed in collaboration with Jessica Andreani and Stefan Seemayer from the lab of Johannes Söding (Gene Center Munich & MPI Göttingen).

A multiple sequence alignment based on the conserved core of *E.coli* YidC was generated using HHblits (Remmert et al., 2012). Further post-processing using HHfilter yielded a non-redundant alignment at 90 % sequence identity. Based on the 2366 resulting sequences direct evolutionary coupling between pairs of YidC residues were calculated (Kamisetty et al., 2013).

Helix-helix interaction probabilities were calculated by aggregating strong coupling coefficients over the expected interaction patterns of helix-helix contacts. This was achieved by taking the expected periodicity of ~ 3.5 residues per alpha helix turn into account. Validation of this approach was performed on a dataset consisting of mainly alpha helical proteins from the CATH database (Sillitoe et al., 2013).

2.5.2 Lipid Exposure Prediction

First, the topology of *E.coli* YidC was predicted using the TOPCONS algorithm (Bernsel et al., 2009). Next, lipid exposure prediction (Lai et al., 2013) of the TM helices was calculated based on the predicted topology. This web-based tool can predict the relative accessible area (rASA) of residues in the lipid environment and is based on high resolution x-ray structures of membrane proteins.

2.5.3 Building a Molecular Model of YidC

The conserved TM helices of *E.coli* YidC were manually positioned according to the covariance based prediction of helix-helix interaction (see 2.5.1) and rotated according to the lipid exposure prediction (see 2.5.2). Additional information from a secondary structure prediction by JPred3 (Cole et al., 2008) and the direct coupling of residue pairs deducted from the covariance analysis were used as structural and spatial restraints to generate molecular models using MODELLER (Eswar et al., 2008). The resulting models were evaluated by measuring the

mean distances within evolutionary coupled residue pairs and the top ranking model was used for further validation using MD simulation (see 2.6.1).

2.6 Molecular Dynamics Simulation

Molecular dynamics simulations of the structural model of YidC (see 2.5.3) were performed in collaboration with Abhishek Singharoy and Prof. Klaus Schulten from the University of Illinois at Urbana–Champaign.

2.6.1 MD Simulation of the Structural Model of YidC

The simulations were performed with NAMD 2.9 using the CHARMM36 force field for proteins and lipids (Klauda et al., 2010). To that end, a lipid bilayer of 110 Å x 110 Å consisting of POPE and POPG in 3:1 ratio was modelled and the YidC model was manually placed into the membrane. After solvating with water, the system was minimized and equilibrated. MD simulation at 300 K was performed for 500 ns and the final 100 ns were repeated three times to examine the statistical significance of the results.

The positional variance of helix residues was quantified as measure of their flexibility. This was achieved by comparing the positions of each helix residues for each trajectory relative to their average positions.

A detailed analysis of interaction energy, hydrogen bonds and membrane thinning was performed to further characterize the structure. This was carried out using standard tools of VMD based on the MD trajectories. The thickness of the membrane at any given point was calculated by determining the minimum distance between phosphates of two lipid head groups on opposite sides of the membrane.

2.7 Structure Determination

2.7.1 Sample Preparation of the F₀C-RNC:YidC Complex

The complex was reconstituted by incubating 10 pmol F₀C-RNCs (see 2.2.4) with 50 pmol freshly purified YidC (see 2.2.2; with C-terminus from *R.baltica*) in a final volume of 50 μ l grid buffer at 37 °C for 20 min. Subsequently, the sample was kept on ice until vitrification.

Grid buffer

20 mM Hepes pH 7.2

100mM KOAc

10 mM Mg(OAc)₂

0.1 % Cymal 6

2.7.2 Cryo-Electron Microscopy

Grid preparation and data acquisition were performed by Charlotte Ungewickell and Otto Berninghausen.

The reconstituted complex (see 2.7.1) was applied to 2 nm pre-coated holey carbon grids (Quantifoil R3/3) and vitrified using a Vitrobot Mark IV (FEI Company) according to the standard protocol (Wagenknecht et al., 1988). Automated image acquisition was performed on a FEI TITAN KRIOS operating at 200 kV under low-dose conditions ($\sim 20 \text{ e}^-/\text{Å}^2$) using a 4 K x 4 K TemCam-F416 CMOS camera (TVIPS gmbH, Germany). The magnification was adjusted to a final pixel size of 1.035 Å at the object scale and images were recorded in a defocus range of $-3.5 \mu\text{m}$ to $-1.2 \mu\text{m}$.

2.7.3 Data Import and Quality Check

A semi-automated pre-processing pipeline was developed to manage the large amount of cryo-EM images acquired by the automated software (EM-Tools; TVIPS GmbH). To that end, existing software packages (Spider, SIGNATURE) (Frank et al., 1996; Chen, 2007) were combined with in-house developed programs

(MAPPOS). All processing steps were parallelized and calculation was performed on the in-house Linux cluster.

The raw micrographs from the camera were converted to Spider files using the 'CP FROM RAW' command omitting the image header. The defocus value of each micrograph was determined using the 'TF ED' command with spherical aberration set to 2.7 mM, an electron wavelength of $\lambda = 0.02508 \text{ \AA}$ and amplitude contrast set to 0.07. The quality of the power spectra was assessed by determining their rotational symmetry. To that end, the cross correlation under a masked power spectra with itself turned by 90° was calculated (Figure 8). Only micrographs in a defocus range of $-3.4 \mu\text{m}$ to $-1.3 \mu\text{m}$ and with highly symmetric power spectra were used for further analysis.

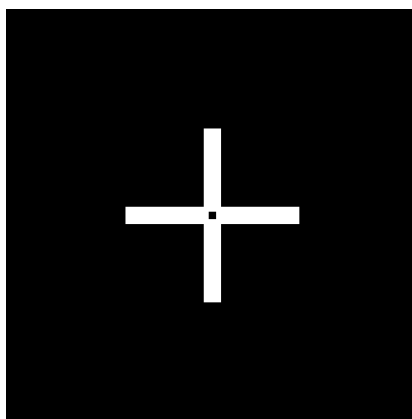


Figure 8: Mask for power spectra validation

The mask was used to determine the rotational symmetry of the power spectra by calculating the cross correlation of the spectra with itself turned by 90° .

2.7.4 Automated Particle Picking using SIGNATURE

Automated particle picking was performed on 4x decimated and band-pass filtered ($20 \text{ \AA} - 80 \text{ \AA}$) micrographs. A modified version of SIGNATURE (Chen, 2007) which can be executed from command line was used to detect particles based on the cross-correlation to reference projections of a 70S ribosome. The coordinates at cross-correlation peaks were used to window out these regions as single particles.

2.7.5 MAPPOS

A Machine learning Algorithm for Particle POST-picking (MAPPOS) was developed in collaboration with Ramin Narousi (LMU Munich).

Particles detected by SIGNATURE (see 2.7.4) contain, depending on the grid quality, a certain amount of false positive particles which can lead to artefacts in the 3D reconstruction. A further classification step using the in-house developed MAPPOS program was used to automatically clean the particle set.

To that end, a subset of automatically picked particles (see 2.7.4) was manually inspected, classified and grouped in two sub-datasets of 500 particles and 500 non-particles (contaminations). An ensemble of classifier was trained using these subsets. The trained classifier was then used to automatically sort contaminations out of the dataset.

2.7.6 Simulation of Cryo-EM Images

A simulated cryo-EM dataset was generated to test the performance of MAPPOS. To that end, projections of 3D volumes were modified to resemble real cryo-EM images with respect to their SNR and image contrast as described in (Baxter et al., 2009).

Briefly, flexibility of real biological samples was accounted for by adding random noise with zero-mean Gaussian distribution to the 2D projections to a SNR of 1.4. The image formation of a bright field electron microscope operating at 300 kV and at a defocus of $-2.0\ \mu\text{m}$ was simulated by modulating the projections with the corresponding contrast transfer function (CTF). For the simulation of low-dose data acquisition and data collection on a CCD camera additional random noise was added to the images to a SNR of 0.05. Then, particle images were band-pass filtered ($20\ \text{\AA} - 80\ \text{\AA}$) to resemble the particles on the micrographs used for automated particle picking (see 2.7.4). Ribosomal projections were generated based on a 70S *E.coli* crystal structure (pdb: 2QAL, 2QAM). Images of non-particles were simulated based on different 3D volumes (plate, cylinder, sphere, void) to account for all types of contaminations in real datasets.

2.7.7 Single Particle Analysis

Initial single particle analysis was performed on a 3x decimated dataset. The Euler angles and x/y – shifts were determined using projection matching to 83 reference projections of an empty 70S ribosome. To that end, each particle of a micrograph was aligned to reference projections which were CTF-distorted to match the contrast of the micrograph. Particles from micrographs collected at similar defocus values were combined into groups with a defocus spread of ~ 200 nm for further processing. A 3D density map of each defocus group was generated by back-projecting the particles according their alignment parameters (Euler angles, x/y – shifts). The resulting maps were CTF-corrected and combined using the Wiener Filter. The resolution of the reconstruction was determined using the $FSC_{0.5}$ criterion. To that end, two half-sets were back-projected and the correlation of the Fourier transformed maps was calculated. Local resolution of the 3D map was calculated using ResMap (Kucukelbir et al., 2014).

2.7.8 Structure Refinement

The resolution of the reconstruction was improved by iterations of alignment and back-projection with decreasing angular increments and decimation thereby increasing the accuracy of Euler angles and x/y – shifts of each particle. The reference volumes were low-pass filtered at 8 Å to avoid over-fitting of the dataset. For smaller angular increments, reference projections were generated separately for each particle based on its orientation. Smooth ribosomal masks including the region of the ligand were used to avoid the alignment of noise.

2.7.9 Generation of a Non-Ribosomal Reference (Edge-volume) for Sorting

The “Edge-volume” was created by assigning random Euler angles to non-ribosomal particle images (carbon-edges, methan blobs) and back-projecting them into 3D. This volume was used as additional cleaning step in the first round of sorting.

2.7.10 Sorting

The dataset was sorted into homogeneous subsets using competitive projection matching. Particles were aligned to reference projections of two volumes and assigned to the one with higher cross-correlation. Different masks were applied to sort for empty or ligand-free ribosomes. Sorting was stopped when particle numbers converged. Selective back-projection using the alignment parameters before sorting was performed to avoid sorting artefacts.

Focused sorting for a more stable ligand conformation was achieved by comparing only areas that correspond to the region of the ligand in 3D (Leidig et al., 2013). The flexibility of the ligand conformation was assessed by calculating the local resolution of the resulting maps using ResMap. Applying a cross-correlation limit based on projection groups of the most homogeneous subset yielded the final dataset.

2.7.11 Modelling the F₀C-RNC:YidC Complex

Molecular models of the ribosome (pdb: 2QAL, 2QAM) and the stable YidC conformation after MD simulation were placed into the cryo-EM density.

The orientation of the YidC model in the cryo-EM map was determined by calculating the cross-correlation of model and map at different relative positions. Models for extra density, assigned to TM1 of YidC and TM1 of the nascent chain based on crosslinking studies (see 2.2.5), were built in UCSF Chimera software using the 'Build Structure' procedure.

2.7.12 Figures

Figures including cryo-EM maps or models were generated using UCSF Chimera 1.8.

3 Results

3.1 Generation of an Automated Pre-Processing Pipeline

The most time-consuming step in pre-processing of single particle datasets is the detection of real particles on micrographs. This can be achieved either manually or using automated software which detects potential particles based on certain features (Zhu et al., 2004). Manual particle picking is not feasible for automatically collected ribosomal datasets consisting of $\sim 10,000$ micrographs and millions of particles. A pre-processing workflow with minimal user interaction was established using automated particle picking based on ribosomal projections followed by a post-picking classification step using the in-house developed MAPPOS program.

3.1.1 Automated Particle Picking using SIGNATURE

Micrographs were decimated to reduce processing time and band-pass filtered to minimize the contrast difference of micrographs collected at different defoci. Then, particles were detected based on reference projections of ribosomes using SIGNATURE (Chen, 2007). The automated particle picking is susceptible of detecting false positives in areas of high contrast (Figure 9 A). Such contaminations (marked with an asterisk in Figure 9 B) can lead to artefacts in the 3D reconstruction and therefore have to be sorted out of the dataset. To this end, a machine learning algorithm for particle post picking (MAPPOS) was developed.

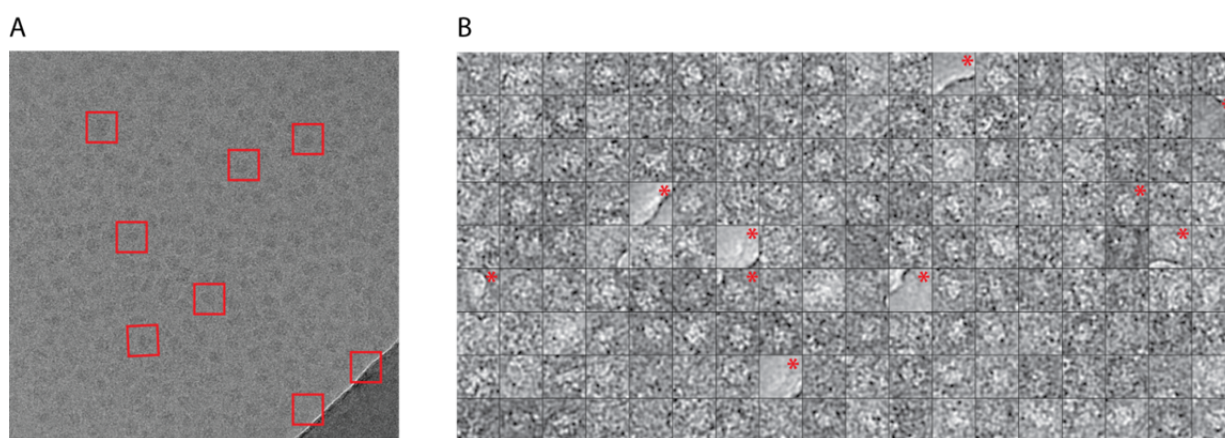


Figure 9: Particle picking from cryo-EM micrographs

A. Cryo-EM micrograph of 70S ribosomes. Red squares indicate a subset of particles detected by SIGNATURE (Chen, 2007) during automated particle picking. **B.** Particle gallery showing projections of 70S ribosomes (particles) as well as non-particles.

3.1.2 Machine Learning Algorithm for Particle Post Picking

MAPPOS was developed in collaboration with Ramin Narousi and Achim Tresch.

The major aim was to develop a fast algorithm which classifies large datasets ($> 1,000,000$ particles) with human-like accuracy. To achieve this, a machine learning approach was used to create an ensemble classifier based on a provided training set. An ensemble classifier consists of multiple classifiers and the final prediction made by the majority of the individual classifiers is considered the final prediction. This ensemble classifier is then used to categorize the complete dataset (Figure 10).

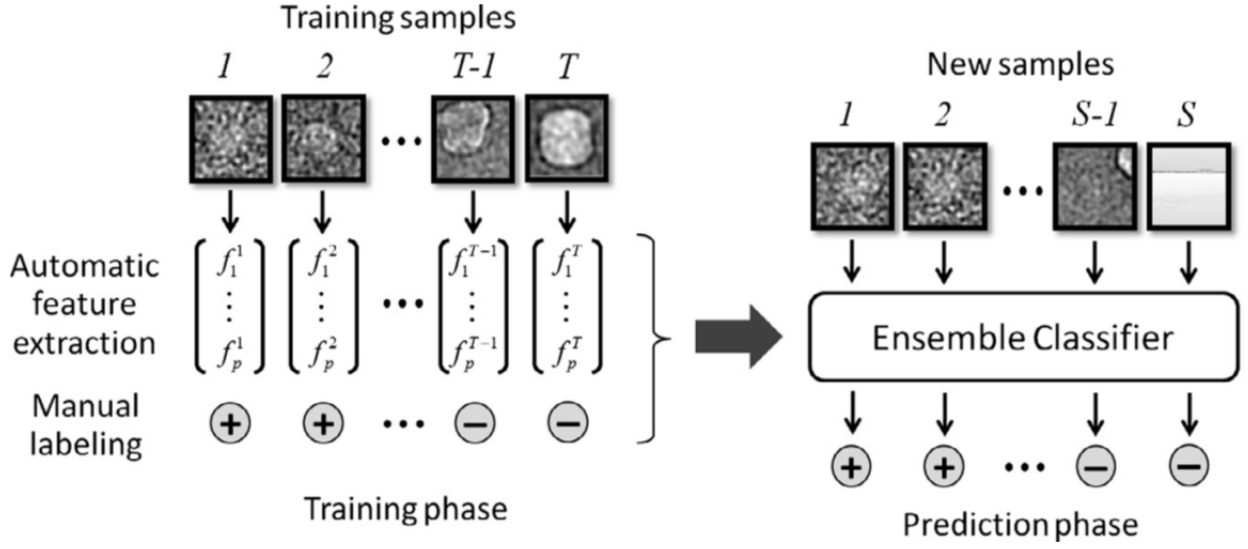


Figure 10: Workflow of MAPPOS

A manually classified trainings dataset is used to train an ensemble classifier. During the training phase, features of the individual images are extracted and the ensemble classifier is trained based on the manual labels. During the prediction phase, the trained ensemble classifier is automatically grouping the complete dataset. Adapted from (Norouzi et al., 2013)

It was crucial to find image features which are fast to calculate and have a high discriminatory power. These features were extracted from each image and were combined with the manual labels (particle, non-particle) during training phase to train a candidate classifier based on a trainings dataset of ~ 500 manually identified particles and non-particles. The parameters of the classifier were optimized and the best performing candidate classifier was added to the ensemble by a bootstrap aggregating approach until the addition of further classifiers did not improve the accuracy of the ensemble. The resulting ensemble classifier was used to automatically categorize the complete dataset.

3.1.3 Simulation of Cryo-EM Images

The performance of MAPPOS was assessed on a simulated dataset consisting of particles and contaminations (see 2.7.6). Ribosomal particles were simulated by projecting a 3D ribosomal density map based on crystal structure into 2D (Figure 11 **A**). These projections were modified to resemble real cryo-EM images with respect to SNR and image contrast (Figure 11 **B**). Contaminations in real datasets after automated picking were manually inspected and could be grouped into 4 classes (Figure 11 **D**). These non-particles were simulated by making projections of different 3D volumes (Figure 11 **C**).

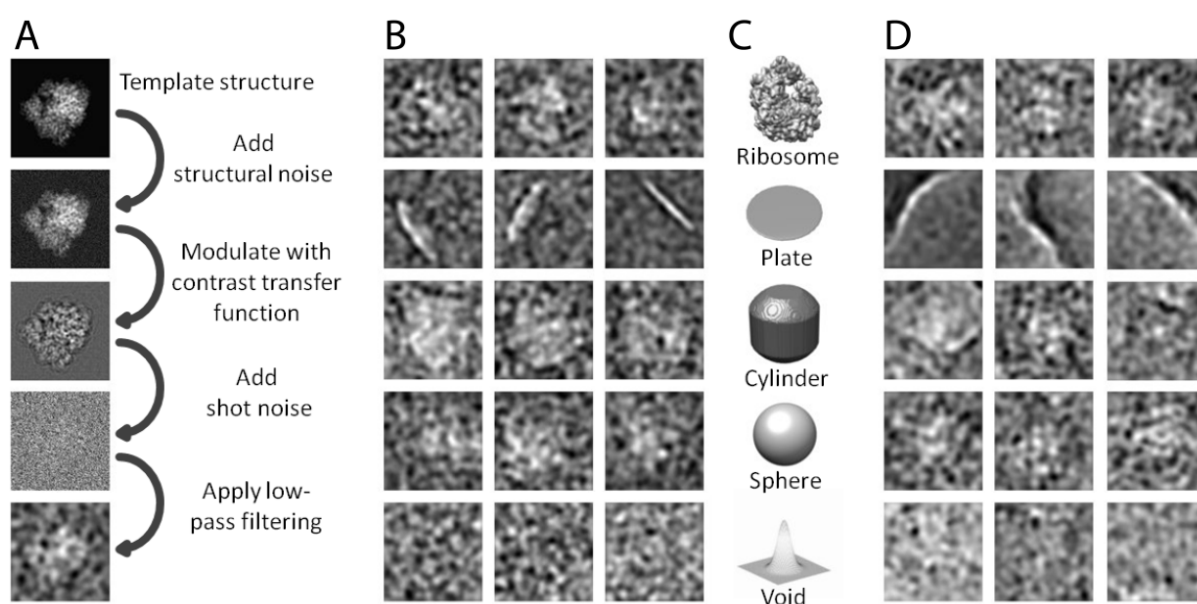
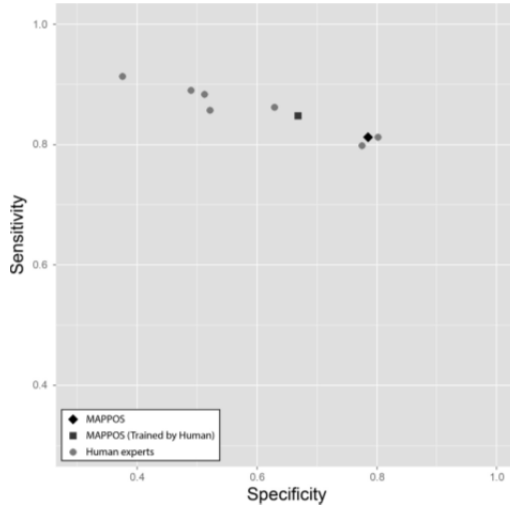


Figure 11: Simulation of a cryo-EM dataset

A. Workflow for the simulation of cryo-EM images. A crystal structure based 3D density map of the 70S ribosome (pdb: 2QAL, 2QAM) was projected into 2D using different orientations. Noise was added to account for structural heterogeneity. Next, the images were CTF-distorted to simulate the image formation of a bright field electron microscope at a negative defocus value. The SNR of the images was set to 0.05 to reflect the low-dose conditions of the data acquisition. A low-pass filter was applied to improve the contrast of the images for further processing steps. Examples of ribosomal projections and 4 types of contaminations of experimental cryo-EM images (**D**) and corresponding simulated images (**B**) based on 3D volumes (**C**). Adapted from (Norouzi et al., 2013)

3.1.4 Comparison of MAPPOS to Human Experts on Simulated Data

The performance of automatic classification using MAPPOS was compared to manual classification of seven human experts on a small simulated test dataset (see 2.7.6) consisting of 1638 particles and 410 contaminations. The major advantage of using a simulated dataset compared to a real dataset is the definite knowledge of the correct classification of each image (particle vs. non-particle). Classification of the test dataset with MAPPOS was performed with two different training datasets. The first training set consisted of only particles and non-particles (Figure 12, rhombus). The second training set contained randomly chosen particles and non-particles based on the classification of the best performing human experts (Figure 12, squares). The results of the classification were analysed by comparison to the known (correct) labels and determining the correctly classified particles (true positive, TP), correctly classified non-particles (true negative, TN), incorrectly classified particles (false negative, FN) and incorrectly classified non-particles (false positive, FP). Sensitivity and specificity of each classification were calculated as performance score (Langlois and Frank, 2011).



$$specificity = \frac{TN}{TN + FP}$$

$$sensitivity = \frac{TP}{TP + FN}$$

Figure 12: Classification with MAPPOS is comparable to human experts

Classification score of MAPPOS and human experts. Sensitivity and specificity values of seven human experts (circles) and MAPPOS trained with true particles and non-particles (rhombus) or trained by best performing human experts (square). Modified from (Norouzi et al., 2013)

This analysis shows that automated classification using MAPPOS is comparable to manual classification of human experts already in a small test dataset.

3.1.5 Performance of MAPPOS on Real Cryo-EM Data

The influence of different post-picking strategies was tested using a real cryo-EM dataset. To that end, cryo-EM images of *E.coli* 70S ribosomes were collected on an electron microscope (see 2.7.2). Automated particle picking using SIGNATURE (Chen, 2007) (see 2.7.4) resulted in a set of 85,726 particles. These particles were classified either manually or automatically (MAPPOS). The resulting three sub-datasets (without post-picking, manual post-picking, MAPPOS) were processed (see 2.7.7) identically and the 3D reconstructions (see 2.7.8) were compared after refinement (Figure 13).

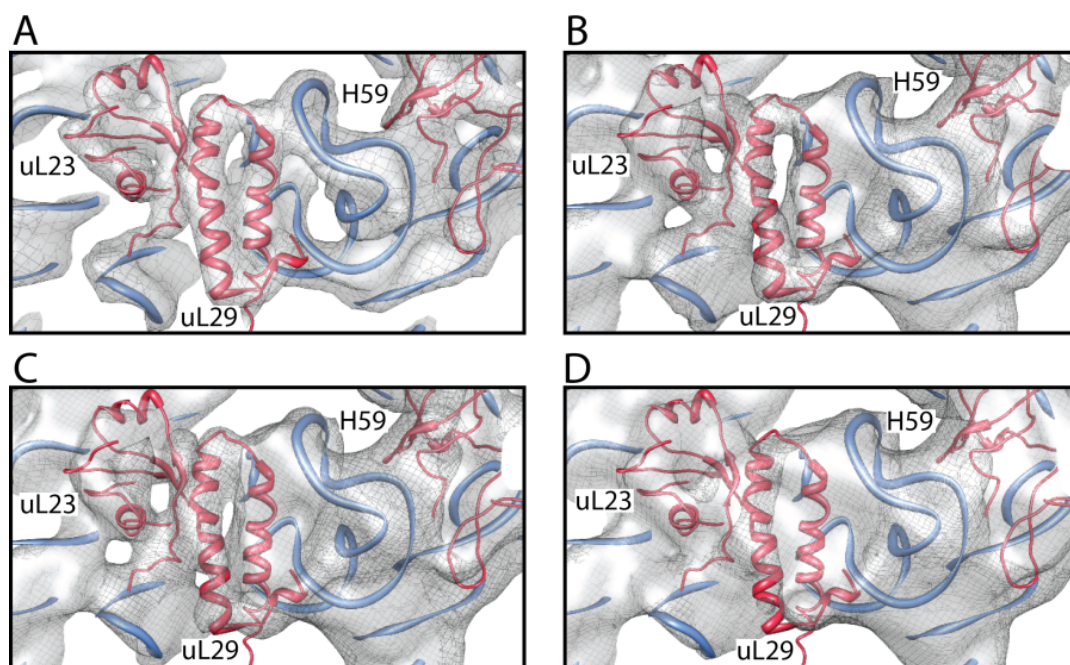


Figure 13: Cryo-EM reconstructions based on different post-picking strategies

The 3D density maps resulting from three different particle sub-sets of the same dataset were compared to the simulated electron density (filtered to 10 Å) of a crystal structure of an *E.coli* ribosome (pdb: 2QAL, 2QAM). The same crystal structure was fitted into all 3D reconstructions. Ribosomal RNA was colored in blue, ribosomal proteins in red and electron density is shown in grey. Secondary structure elements are resolved in the simulated map (A) as well as the manually (C) and automatically (B) post-picked reconstruction. The 3D map without post-picking shows no features for secondary structure (D). Modified from (Norousi et al., 2013)

The ribosomal density map of an *E.coli* crystal structure filtered to 10 Å is shown as reference (Figure 13 A). At this resolution, secondary structure elements e.g. α -helices start to be resolved as rod like densities (uL29). Single particle reconstructions of automatically (Figure 13 B) and manually (Figure 13 C) post-picked datasets showed features comparable to those of the simulated map. In

contrast, none of these features were resolved in the density map without post-picking (Figure 13 **D**).

These results show, that particle post-picking is essential for calculating well-resolved density maps and MAPPOS performs comparable to human experts also on real cryo-EM datasets.

3.2 *In vitro* Reconstitution of a YidC Dependent Insertion Intermediate

Structural investigation of co-translational insertion of a YidC-only substrate required the purification and biochemical characterisation of the *in vitro* reconstituted YidC:ribosome complex. To that end, ribosomes with defined nascent chain and the membrane protein insertase YidC were purified in sufficient amounts for biochemical and structural studies.

3.2.1 Purification of F_{OC}-RNCs

In order to observe structural snapshots of co-translational membrane protein insertion it is essential to generate stalled ribosomes with a defined nascent chain displayed outside the ribosomal tunnel. Using the TnaC-stalling sequence in an appropriate *E.coli* strain (see 2.1.1) it is possible to generate RNCs *in vivo* in large amounts (Bischoff et al., 2014b). The construct (Figure 14) was designed with a 3C cleavage site to remove N-terminal tags leaving only the native sequence of F_{OC}. The nascent chain construct was designed so that the C-terminal end of the first TM helix of F_{OC} aligns with the ribosomal tunnel exit. Therefore only amino acids coding for F_{OC} are displayed outside the ribosome.

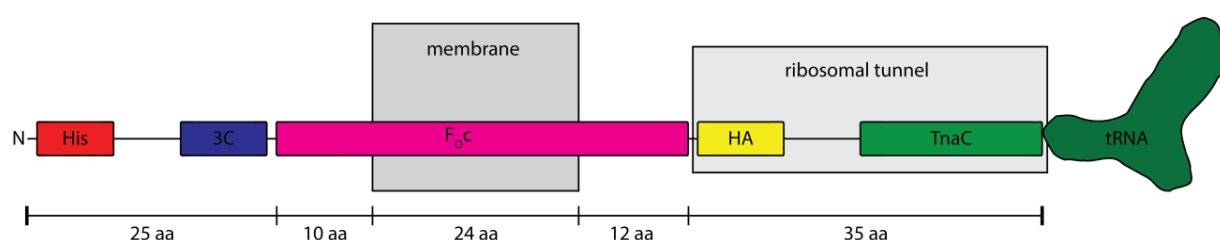


Figure 14: Linear representation of the F_{OC} construct

The N-terminal His-tag for purification is followed by the 3C cleavage site. The coding sequence of the first TM helix of F_{OC} is displayed directly outside the ribosomal tunnel which harbors an HA-tag for western blot detection and the TnaC sequence for stalling. Length is defined in amino acids (aa).

RNCs were purified as described in the material and methods chapter (see 2.2.4). Briefly, the construct described in Figure 14 was expressed in *E.coli* KC6 cells. The ribosomal fraction was loaded on a metal affinity column to enrich for ribosomes that are stalled with the TnaC sequence. Next, the eluate was loaded on a linear sucrose gradient and the 70S peak fraction was isolated and concentrated resulting into the final RNCs which were used for further biochemical and structural

studies. The western blot of the RNC fraction (Figure 15) showed a clear signal at a molecular weight of ~ 30 kDa corresponding to the nascent amino acid chain still attached to the tRNA and an additional band at ~ 10 kDa representing the free peptide.

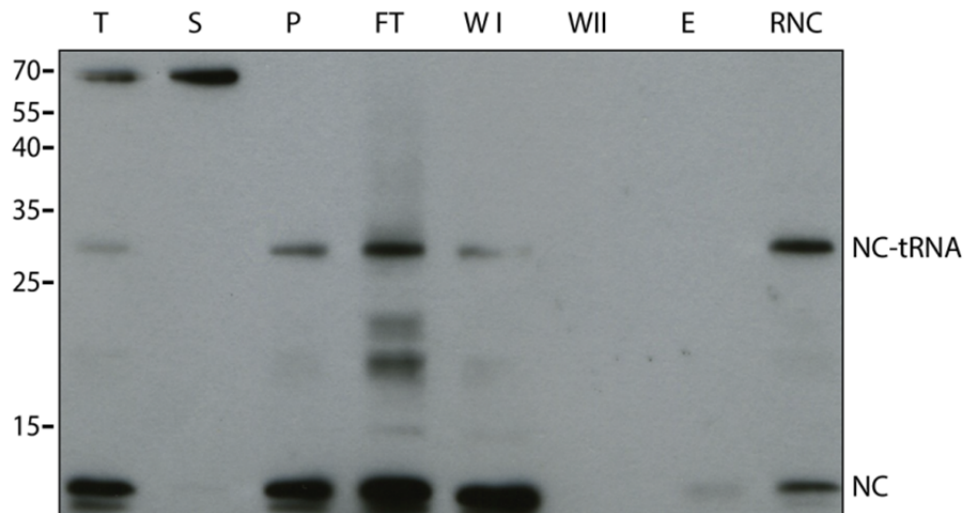


Figure 15: RNC purification

Fractions of each step of the purification were analyzed using western blot with antibody detection against HA-tags. The total cell lysate (T) was separated into ribosomal fraction (P) and soluble fraction (S) using ultra centrifugation. The ribosomes were resuspended and RNCs were enriched by metal ion affinity (FT, flow-through; WI, low salt wash; WII, high salt wash; E, elution). The elution fraction was loaded on a linear 10 % – 40 % sucrose gradient; the 70S peak was collected and concentrated (RNC).

Taken together, ribosomes stalled with the TnaC sequence and displaying a specific amino acid sequence outside the ribosomal tunnel can be expressed *in vivo* and purified in large amounts (~ 4 OD/l LB culture).

3.2.2 Purification of YidC

For the biochemical and structural studies of Sec-independent membrane protein insertion it was essential to purify YidC in reasonable amounts. The purification procedure as described in the material and methods chapter (see 2.2.2) was based on the structural study of YidC using 2D crystals (Lotz et al., 2008). To avoid unspecific binding of the affinity tag to the ribosome (Kedrov et al., 2013) the 3C cleavage site was used to cleave it during dialysis.

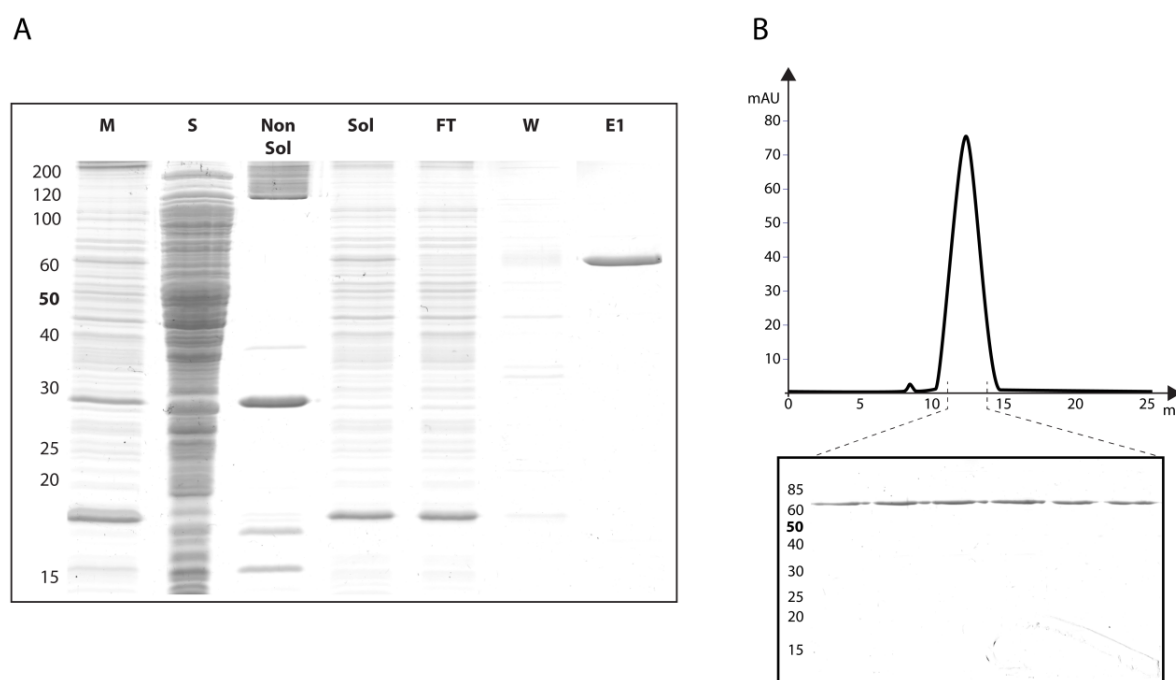


Figure 16: Purification of YidC

A. Coomassie stained SDS-PAGE of the YidC purification. The total membrane fraction (M) was solubilized with Cymal 6 and non-solubilized material (No Sol) was pelleted using ultra centrifugation. The solubilized membrane proteins (Sol) were applied to a metal affinity column. Non-tagged proteins (FT) and non-specifically bound proteins were washed (W) from the column. The His-tagged YidC protein was eluted using imidazole (E1). **B.** Elution profile of size SEC showing a distinct peak for YidC. Fractions of the YidC peak were analyzed using SDS-PAGE.

For purifying YidC, the membrane fraction was solubilized with Cymal 6 and the non-solubilized material was pelleted using ultra-centrifugation (Figure 16). The solubilized membrane proteins were loaded on a metal affinity column and the elution showed a very prominent band at ~60 kDa corresponding to YidC. After 3C protease treatment, the sample was loaded on a gel filtration column. A distinct and symmetric peak at 12 ml elution volume indicated a homogeneous sample which was concentrated to ~4 μ M and used directly for further biochemical or structural studies.

3.2.3 Crosslinking Studies of Reconstituted YidC:F₀C-RNCs Complexes

To test whether the *in vitro* reconstituted YidC:F₀C-RNC complex resembles an *in vivo* state, crosslinking studies of single cysteine mutants were performed. A single cysteine was introduced into cysteine-less YidC (C423S) at position M430. This construct was shown to crosslink co-translationally to transmembrane residues of nascent chains *in vivo* (Yu et al., 2008). Another single cysteine was introduced into the middle of the transmembrane helix of F₀C (G23C) to test for direct interaction of the YidC (C423S; M430C) and F₀C (G23C) cysteine residues after *in vitro* reconstitution.

Two different YidC:F₀C-RNC complex were reconstituted *in vitro* (see 2.7.1) and cysteine crosslinks were introduced (see 2.2.5). RNCs with and without single cysteine in the first TM helix of F₀C were incubated with detergent-solubilized YidC (C423S; M430C) to form the complex. Successively, cysteins were oxidized using 5,5'-dithiobis-2-nitrobenzoic acid (DTNB) to form the crosslinks. The samples were loaded on a 10 % – 40 % linear sucrose gradient and the ribosomal complexes in the 70S peak were harvested. To verify the crosslink formation, DTT was added to the oxidized sample to reduce the crosslink. Western blot analysis with antibodies against HA-tag (RNC) and against YidC was performed on all samples.

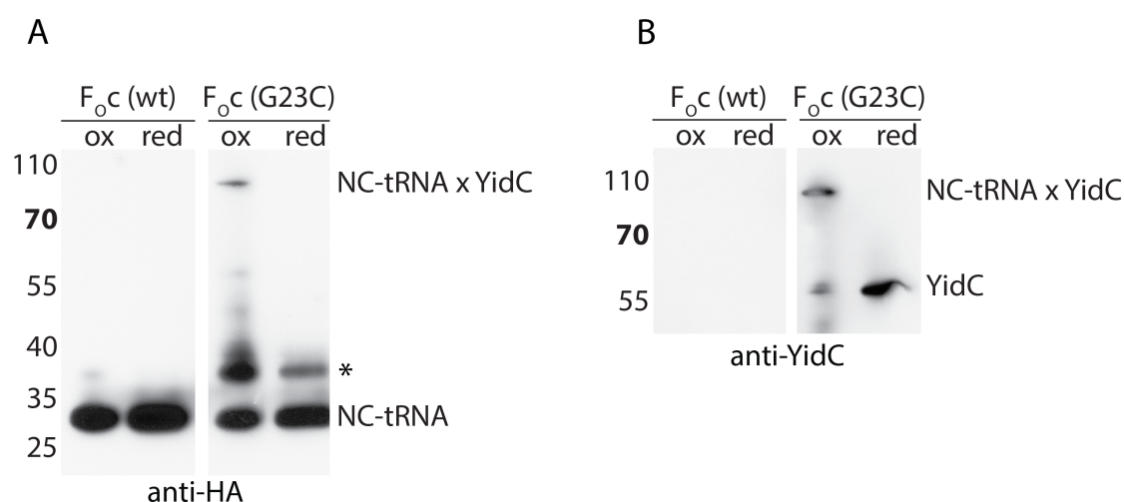


Figure 17: Nascent chain of F₀C-RNCs interacts directly with YidC

The *in vitro* reconstituted YidC:F₀C-RNC complexes were analyzed by western blotting using antibodies against the HA-tag located in the nascent chain (A) and YidC (B). A DTT sensitive additional band at ~90 kDa could be detected by both antibodies in the oxidized fraction of the YidC (C423S, M430C):F₀C (G23C)-RNC complex.

When YidC(C423S, M430C):F_{OC}(G23C)-RNCs-complexes were reconstituted and crosslinked *in vitro*, a clear band at the molecular weight of ~90 kDa was observed in the oxidized fractions of both western blots using antibodies against HA-tag (RNC) and YidC (Figure 17). After reduction this complex collapsed into a 60 kDa (YidC) and ~30 kDa (NC-tRNA) fragment, showing that the nascent chain was directly interacting with YidC. The western blot against the HA-tag showed an additional band (marked with *) corresponding to an unspecific and poorly reproducible crosslink of the nascent chain at ~37 kDa which is only partially reduced. The control experiment with cysteine-less F_{OC}-RNCs showed only bands for the non-crosslinked NC-tRNA (~30 kDa) in the HA blot (Figure 17) and no bands of higher molecular weight. The lack of signal in the western blot against YidC indicates that no YidC was co-purified together with the RNCs under these conditions.

Taken together, these crosslink experiments provide evidence that the *in vitro* reconstituted YidC:F_{OC}-RNC complex resembles the insertion intermediate of the nascent chain which can also be detected *in vivo*. This complex is therefore well-suited for further structural studies of this intermediate using cryo-EM and single particle reconstruction.

3.3 Building a Structural Model of YidC

High resolution structural information of YidC is only available for the non-essential periplasmic P1 domain (Oliver and Paetzel, 2008; Ravaud et al., 2008). The transmembrane part of *E.coli* YidC consists of six transmembrane helices with the overall topology shown in Figure 18 (Sääf et al., 1998). The N- and C-terminus are both located in the cytoplasm, whereas the large soluble P1 domain is in the periplasm. *E.coli* YidC and its homologs in Gram-positive bacteria, mitochondria and chloroplasts share a conserved core of the last five transmembrane helices (Dalbey et al., 2014). Only YidC of Gram-negative bacteria contains the additional N-terminal helix and the large soluble periplasmic P1 domain.

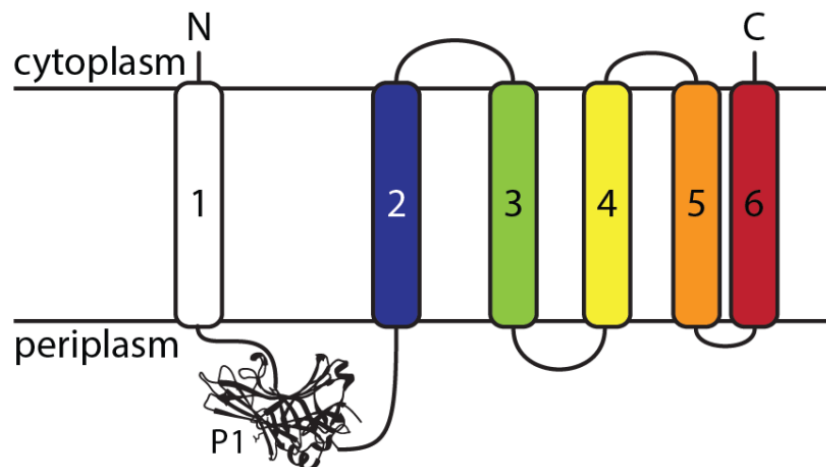


Figure 18: Topology of *E.coli* YidC

Membrane topology of YidC in Gram-negative bacteria (e.g. *E.coli*). TM1 (1) and P1 are missing in Gram-positive bacteria and eukaryotes. The conserved core of the last 5 transmembrane helices (TM2–TM6) is consistently colored in all following figures.

3.3.1 Evolutionary Covariance Matrix of the Conserved Core of YidC

The available biochemical and bioinformatics data provide a basic understanding about the overall topology of YidC and can give some hints on helices and residues that may possibly interact (Yuan et al., 2007; Klenner and Kuhn, 2012). Yet, these spatial restraints are not sufficient and accurate enough to build a structural model of the membrane domain of YidC. Recently developed algorithms (Marks et al., 2011; Hopf et al., 2012; Marks et al., 2012) are capable of retrieving structural restraints based solely on multiple sequence alignments. This new approach was applied to the conserved membrane part of YidC.

It can be assumed that some residue pairs, which are essential for protein stability and function, might be far away in the primary amino acid sequence. This suggests a higher evolutionary pressure on these positions to mutate dependently on each other. Thus, residue pairs that form direct contacts should have a higher covariance which can be calculated based on MSA of the protein family.

Direct evolutionary couplings between pairs of YidC residues were calculated (see 2.5.1) in collaboration with Jessica Andreani and Stefan Seemayer from the group of Johannes Söding (Gene Center Munich).

Briefly, a MSA consisting of 2366 sequences was build based on the conserved core of YidC (TM2–TM6). The direct evolutionary coupling between residues was calculated and the scores were stored in a covariance matrix (see 2.5.1).

The evolutionary covariance matrix (Figure 19 **A**) contains the coupling scores of each possible residue pair with the consensus sequence of the MSA on the x- and y-axis. Patterns of higher scores that are diagonal or anti-diagonal indicate residues that are possibly involved in helix-helix interactions. Most of these patterns coincide with predicted secondary structure element such as α -helices, β -sheets or transmembrane segments, although biophysical properties of the amino acids were not taken into account. Probabilities for helix-helix interaction were calculated based on the assumption that interacting residues of two parallel helices would face each other and therefore have a regular spacing of one helical turn (see 2.5.1). A distinct subset of interacting helices could be determined by comparison with helix interaction of known crystal structures (see 2.5.1) and marked with black boxes. Each transmembrane helix is coupled to at least two other helices, thereby providing a network of evolutionary coupled residues sufficient for initial TM positioning. In addition to the transmembrane helices there was strong evidence for two cytoplasmic helices between TM2 and TM3 to form an anti-parallel hairpin. The

50 residues pairs showing the highest coupling score are highlighted with red crosses of corresponding size. Most of them coincide with interacting transmembrane helices, but some also predict interaction within loop residues (Loop3-4; Loop5-6). A more schematic representation of the interaction network (Figure 19 **B**) illustrates a strong connection of the last three TMH (TM4 – TM6). TM4 is strongly coupled to TM5 and TM6. There is only weaker coupling on the cytoplasmic side between TM5 and TM6 (Figure 19 **A**). Linked via TM6, TM2 and TM3 form the second interaction core.

Taken together, evolutionary covariance between residue pairs of the conserved part of YidC was calculated. The patterns in the covariance matrix strongly suggest an interaction network of the TMs consisting of two distinct groups linked by TM6 (TM4 – TM6; TM2, 3, 6). This information allows the relative positioning of the conserved TMH of YidC.

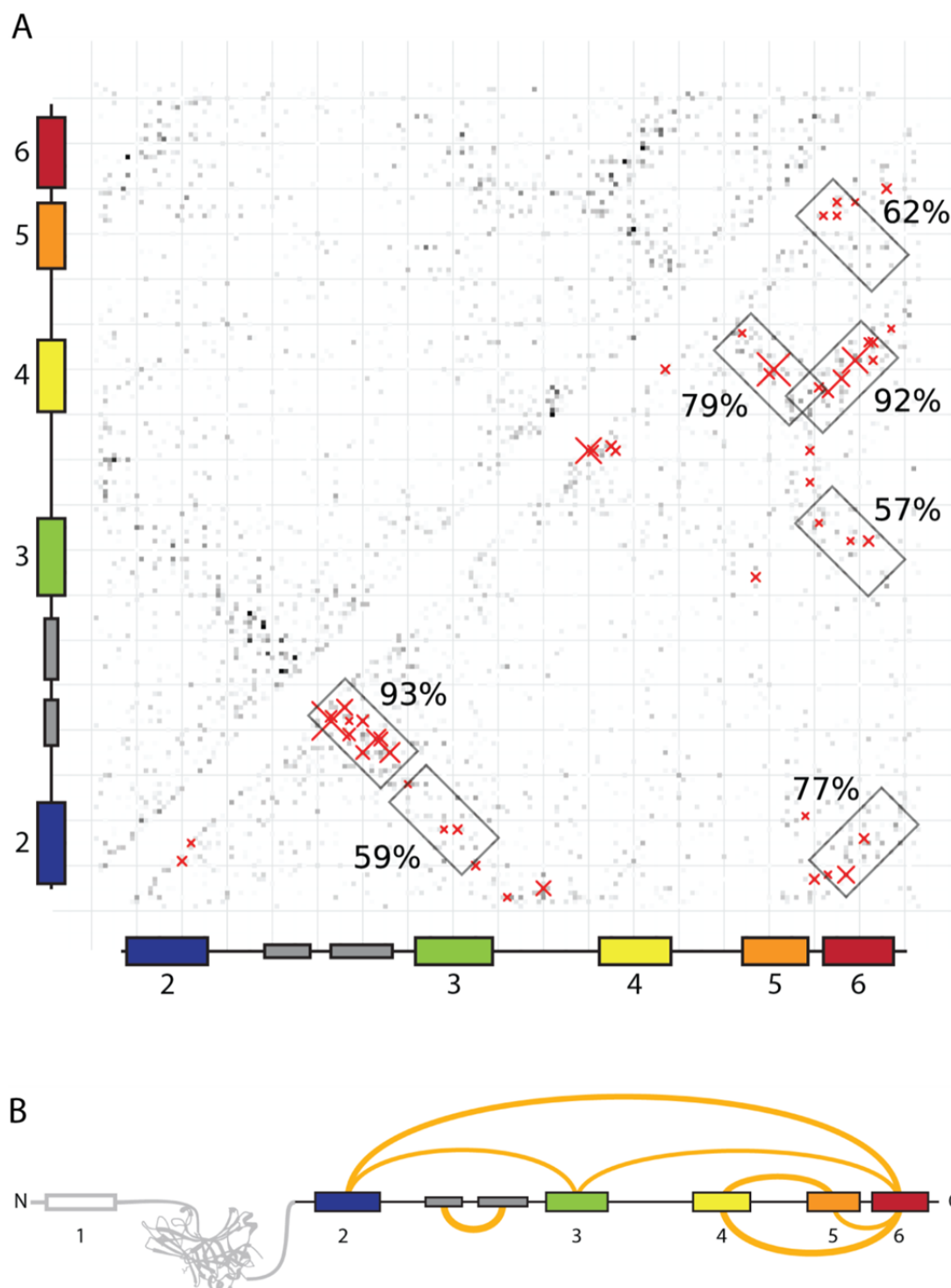


Figure 19: Evolutionary covariance matrix of YidC residues

A. Matrix of covariance coupling strength within the conserved part of YidC. The greyscale of the dots reflects coupling strength. The 50 residues with the highest coupling score are marked with a red cross. The size of the cross is proportional to the coupling strength value. The helix pairs resulting from the calculation of direct helix interaction probability (percent value) are marked with black boxes. **B.** Linear representation of *E. coli* YidC. Interacting transmembrane helices are connected by arches with the thickness corresponding to the interaction probability. Adapted from (Wickles et al., 2014)

3.3.2 Lipid-Protein Interface and Helix Arrangement

The transmembrane segments of polytopic α -helical membrane proteins can be predicted based on their hydrophobicity (Hessa et al., 2005). Hydrophobic residues are necessary for transmembrane segments to form an interaction surface with the surrounding aliphatic lipid tails. Additionally, the individual helices have to establish direct protein-protein contacts upon folding into their final three dimensional conformation. These ambivalent properties needed for the different interaction partners can be used to generate predictions of lipid vs. protein exposed sites within TMH (Lai et al., 2013).

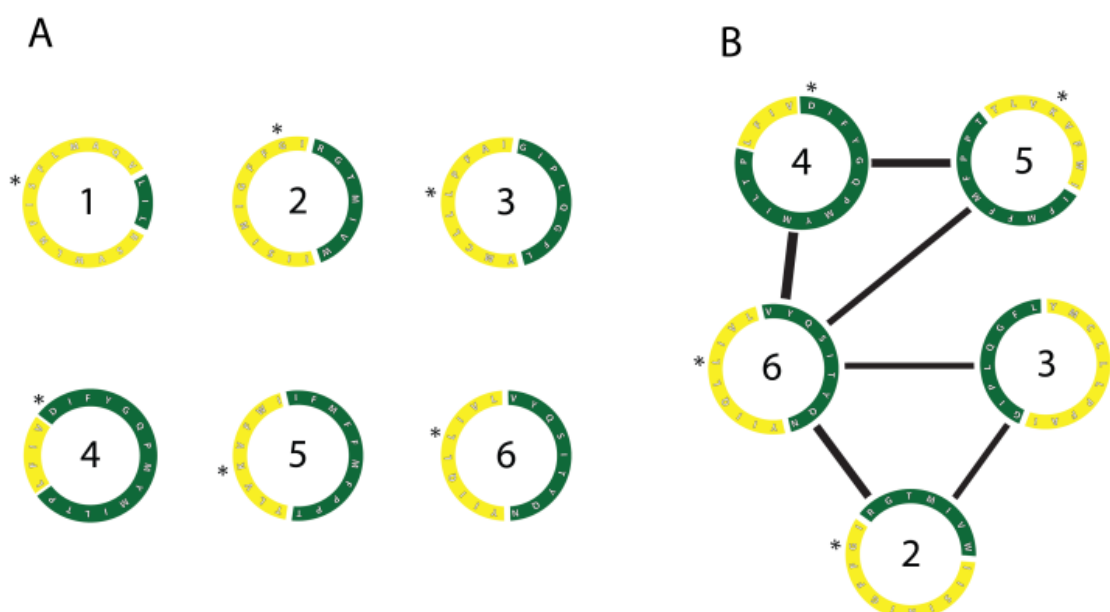


Figure 20: Lipid exposure prediction and helix arrangement

A. Wheel model of the six TMs of *E.coli* YidC. The first residue of each helix is marked with (*). The surfaces of the helices are colored according to predicted interaction with lipids (yellow) or protein (green). **B.** TMHs (colored as in **A**) of the conserved TMs are positioned according to their interaction probabilities calculated by evolutionary covariance and rotated based on the lipid exposure prediction. Modified from (Wickles et al., 2014)

Lipid exposure prediction was performed for the five conserved TM helices of YidC (see 2.5.2). Each of the TM helices exhibits a biphasic surface distribution (Figure 20 A). The surfaces involved in direct protein interaction (green) and lipid interaction (yellow) are unevenly distributed within the TM helices. The surface of the first TM helix, which is not a part of the conserved core, shows almost exclusive lipid interaction properties. In contrast, TM4 shows only a small lipid interacting surface which suggests strong helix-helix interactions to more than one

other TM helices. TM2, TM3, TM5 and TM6 are predicted to have equally exposed surfaces towards lipids and protein.

Taken together, the results of the lipid-exposure prediction and the helix-helix interaction based on the covariance analysis (see 3.3.1) can be combined to assemble an initial arrangement of the five conserved TMHs (Figure 20 **B**). This model is consistent with both independent analysis and illustrates the special role of TM6 which is the linking TM4 and 5 to TM2 and TM3.

3.3.3 Three Dimensional Model of YidC

Building molecular models of proteins sharing sequence homology to proteins with known crystal structures is widely used in structural biology. Without a 3D template, however, model building requires structural restraints that can be applied to the primary sequence for *in silico* folding. Bioinformatics tools can extract valuable information about secondary structure, topology and modification sites of a protein based on the primary amino acid sequence. New algorithms (Marks et al., 2011; Hopf et al., 2012; Marks et al., 2012) detect evolutionary coupled residue pairs that can be used as direct structural restraints in model building.

A structural model of YidC was built in several steps. First, the transmembrane segments of the protein were predicted (see 2.5.2). Next, the secondary structure was identified by JPred3 and applied to the primary sequence. Then, transmembrane helices were arranged according to covariance analysis and lipid exposure prediction (see 3.3.2) parallel to each other. The core model was then connected by loops, incorporating structural restraints of covariance analysis, using MODELLER (see 2.5.3). The distance between evolutionary coupled residue pairs was used to rank these models.

In the final model, the edges of the TMHs of the conserved membrane part of YidC are arranged in pentagonal shape (Figure 21, top view). The cytosolic loop connecting TM2 and TM3 forms a helix hairpin that is aligned parallel to the membrane. This 'helical paddle domain' (HPD) is highly conserved in length and amino acid distribution. All evolutionary predicted interactions are fulfilled within this monomeric model of YidC.

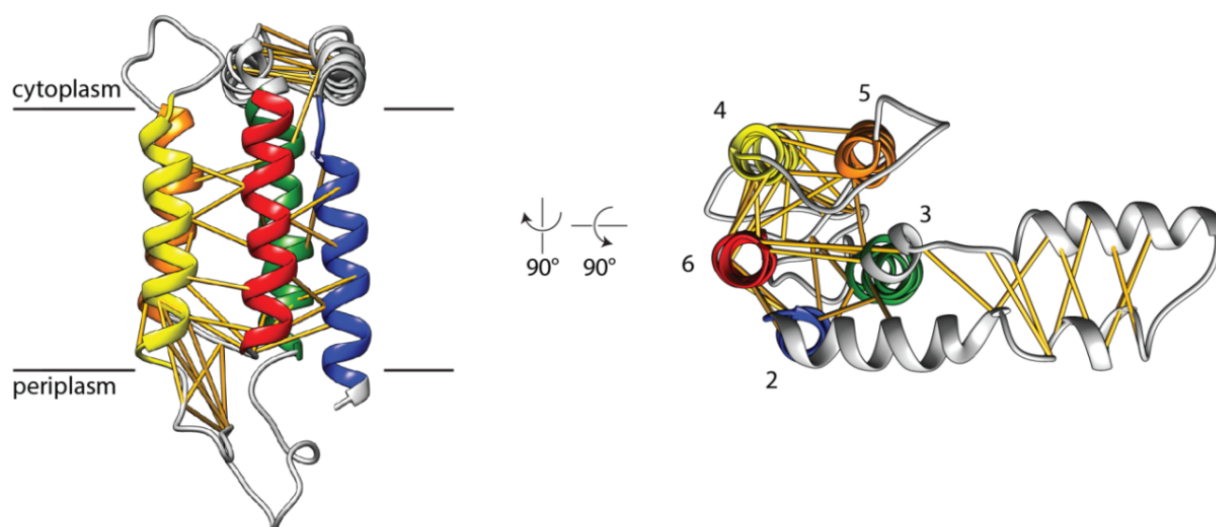


Figure 21: Three dimensional model of YidC with covariance restraints

Side view and top view of a structural model of the conserved core of YidC. Contacts predicted by covariance analysis are depicted as yellow pseudo-bonds and the TMHs are colored as in Figure 18. The model was generated by arranging the individual TMHs based on helix-helix interactions (see 3.3.1) and lipid-exposure (see 3.3.2). A full-length model was built using MODELLER applying distance restraints between residue pairs derived from the covariance analysis. The evolutionary coupled residues in the cytoplasmic loop connecting TM2 and TM3 suggest the formation of an α -helical hairpin. Adapted from (Wickles et al., 2014)

Taken together, the conserved part of YidC could be modelled *ab initio* (without 3D template) using only bioinformatics tools based on large MSA.

3.4 Validation of the Model

Structures of membrane proteins determined by X-ray crystallography provide a static picture at high resolution. The dynamic behaviour of a protein and possible conformational changes need to be addressed either by additional x-ray structures or by *in silico* simulations based on known structures. Molecular dynamics simulations offer tools for analysing, testing and validating membrane protein structures embedded in their natural lipid environment (Gumbart et al., 2005; Lindahl and Sansom, 2008; Khalili-Araghi et al., 2009). Therefore, MD simulations of the structural model of YidC were carried out in collaboration with Abhishek Singharoy from the lab of Klaus Schulten at the Beckman Institute for Advanced Science and Technology in Illinois.

3.4.1 Molecular Dynamics Simulation of YidC

MD simulation was performed to test the stability of the structural model in a membrane environment (see 2.6.1). To this end, the YidC model was placed into a lipid bilayer modelled of POPE:POGE in a 3:1 ratio mimicking the lipid composition of the *E.coli* inner membrane. After equilibration of the system, a 500 ns simulation was performed and trajectories were analysed.

Overall, the structural model was found to be very stable in the context of a lipid bilayer during simulation. The stability of the TM core is based on specific interactions within the TMHs and towards the apolar lipid tails (Figure 22). In contrast, the HPD shows a certain degree of flexibility. It aligns with the lipid surface stabilized by interactions of polar residues with charged lipid head groups.

The TM helices are mainly stabilized by different interactions within three distinct regions. First, hydrophobic interactions between apolar residues on the surface of the protein and the apolar lipid tails anchor the helices in the membrane. Second, the cytoplasmic side of the TM core is stabilized by strong charge-dipole interactions of specific polar residues within the helices. Third, residues at the periplasmic site cluster together by stacking interaction of aromatic residues.

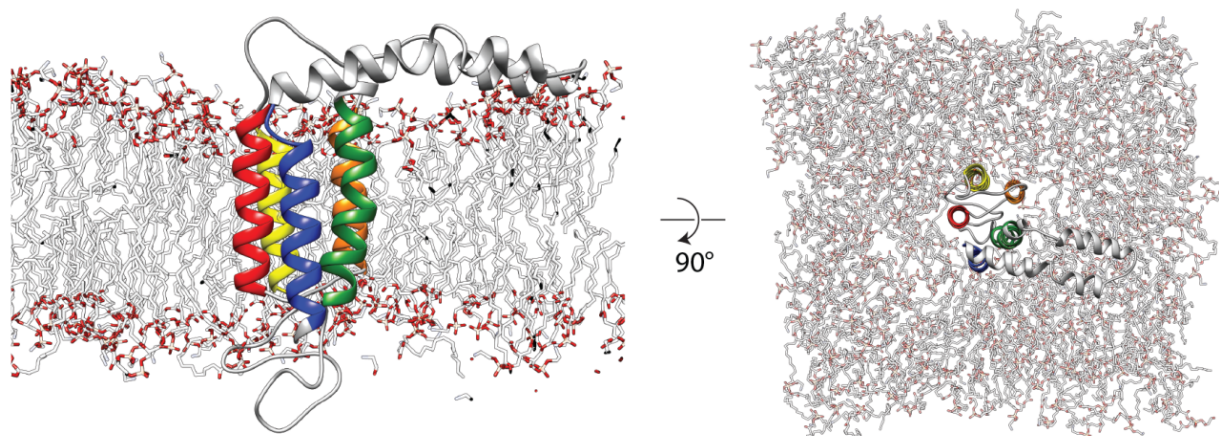


Figure 22: Stable YidC conformation in lipid bilayer

Side view (left) and top view (right) of a stable YidC conformation after a 500 ns molecular dynamics simulation in a lipid bilayer composed of POPE:POPG in 3:1 ratio. Adapted from (Wickles et al., 2014)

Taken together, this analysis shows that residues at certain positions of YidC need to have specific biophysical properties in order to contribute to the global stability of the protein.

To assess the contribution of individual residues, interaction energies for each trajectory were calculated. The analysis of inter-helix energy (see 2.6.1) revealed that most of the residues within the interior of the five TM helices are involved in direct helix-helix contacts (Figure 23 **A**). Residues having the highest interaction energies are T362 in TM2 and Y517 in TM6. The least interactions were observed for residues in TM3, explaining the high flexibility (Figure 23 **B**) observed during MD simulation. Outside the membrane, all loops including the structured HPD showed high flexibility.

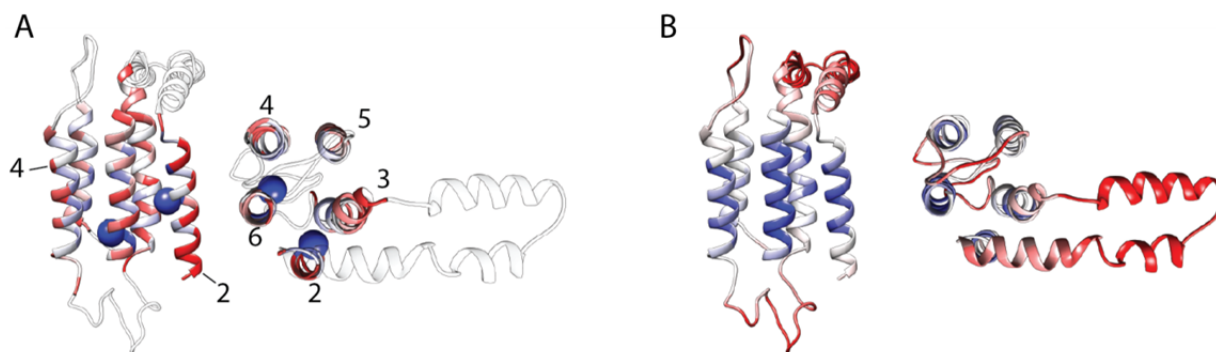


Figure 23: Residue properties based on MD simulation

A. Stable YidC conformation colored by inter-helix energy (in kcal/mol; blue: -7.5 to -1 ; white: -1 to -0.002 ; red: ≥ -0.002). The residues showing the strongest interactions are located within TM2 and TM6 (blue spheres). **B.** Same view as in **A.** Coloring according to flexibility during MD simulation (in \AA^2 ; blue: 0.04 to 0.09 ; white: 0.09 to 1 ; red: ≥ 1.0). The HPD between TM2 and TM3 shows the highest flexibility of the structured protein part. Within the membrane, TM3 is the most flexible. Adapted from (Wickles et al., 2014)

To validate the results of the MD simulation, the top ranking residues involved in helix-helix interactions were subjected to an *in vivo* complementation assay. The MD simulation suggested that these residues are involved in either charge-dipole or stacking interactions. Therefore, mutating them to alanine should destabilize these interactions without changing the overall hydrophobicity needed for a membrane helix.

The *in vivo* complementation assay was performed (see 2.4.1) in *E.coli* FTL10 cells with YidC under an arabinose-inducible promoter. The assay provides a quick way of testing the contribution of individual residues to the global activity of an essential protein.

Although, the TM core of YidC was shown to be very tolerant towards mutations and even replacement of complete TM helices (Jiang et al., 2003), two single mutations in TM2 (T362A) or TM6 (Y517A) completely inactivated the protein (Figure 24 A). These mutants were stably expressed (Figure 24 C) but could not fulfil their function. Interestingly, these residues are located at the same height within the membrane (Figure 24 B) and mutations showing decreased activity (F333, M471 and F505) were found to cluster around them.

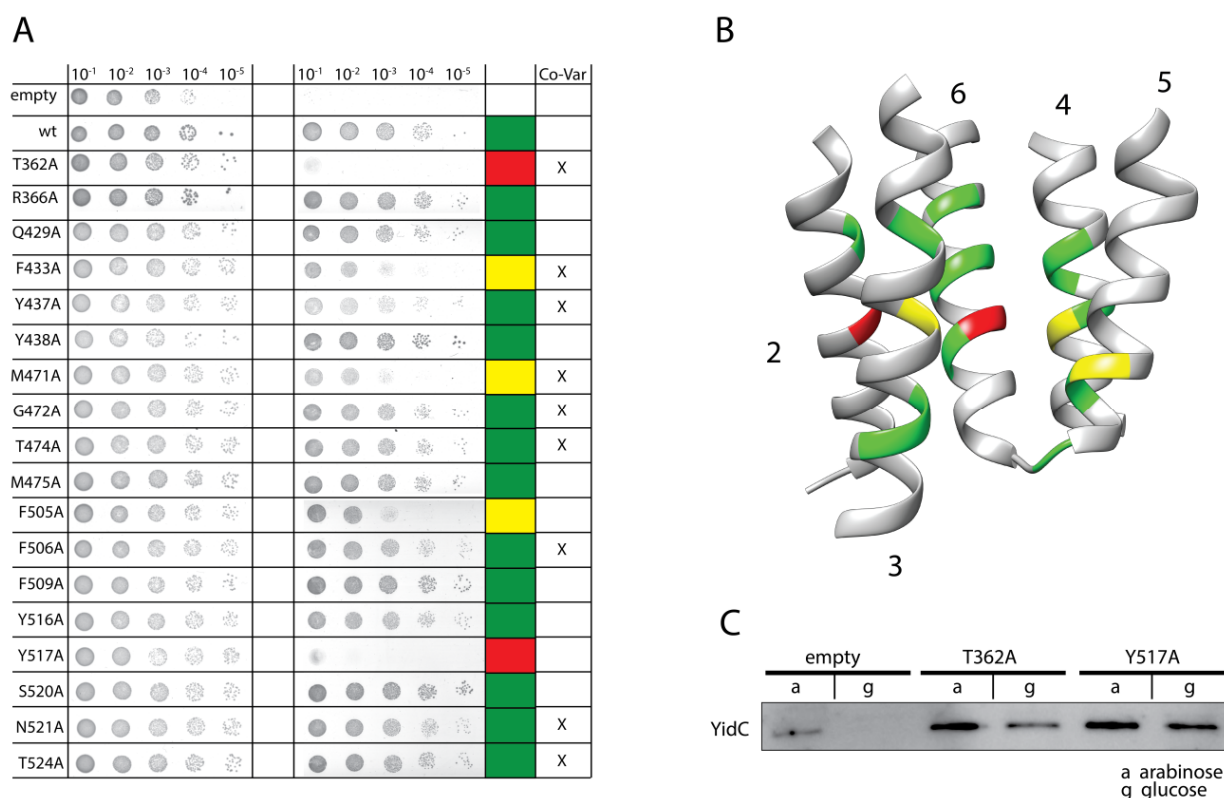


Figure 24: Structurally important TM residues of YidC

A. In vivo complementation assay of structurally important residues based on the MD simulation. Evolutionary coupled residues are marked with (X) in the Co-Var column. **B.** Model of the membrane part of YidC with positions tested in the complementation assay in green (no effect), yellow (reduced activity) and red (inactive). Numbers indicate the TMH. **C.** Mutations of residues in TM2 (T362A) or TM6 (Y517A) are stably expressed and sufficient to completely inactivate YidC. Residues located in the proximity (F333, M471, F505) show an reduced activity level. Modified from (Wickles et al., 2014)

These findings show that the results based on the MD simulations on the structural model of YidC are consistent with *in vivo* experiments. Surprisingly, the function of YidC can be inactivated by single amino acids substitutions of strongly interacting transmembrane residues. However, these inactivating mutations are restricted to specific positions indicating the importance of this region for protein activity.

After evaluating the contribution of individual residues, the global amino acid distribution was analysed based on the stable conformation derived after MD simulation. To that end, cross-sections at different heights within the membrane were coloured according to the hydrophobicity of the residues (Figure 25).

YidC forms a hydrophilic cavity in the membrane which is open towards the cytoplasm. It is built of polar and charged residues located in TM4, 6, 2 and TM3.

This cavity narrows towards the periplasm and is sealed by tight stacking interactions of aromatic residues. The ability of forming a hydrophilic environment in the membrane might play a relevant role in substrate insertion.

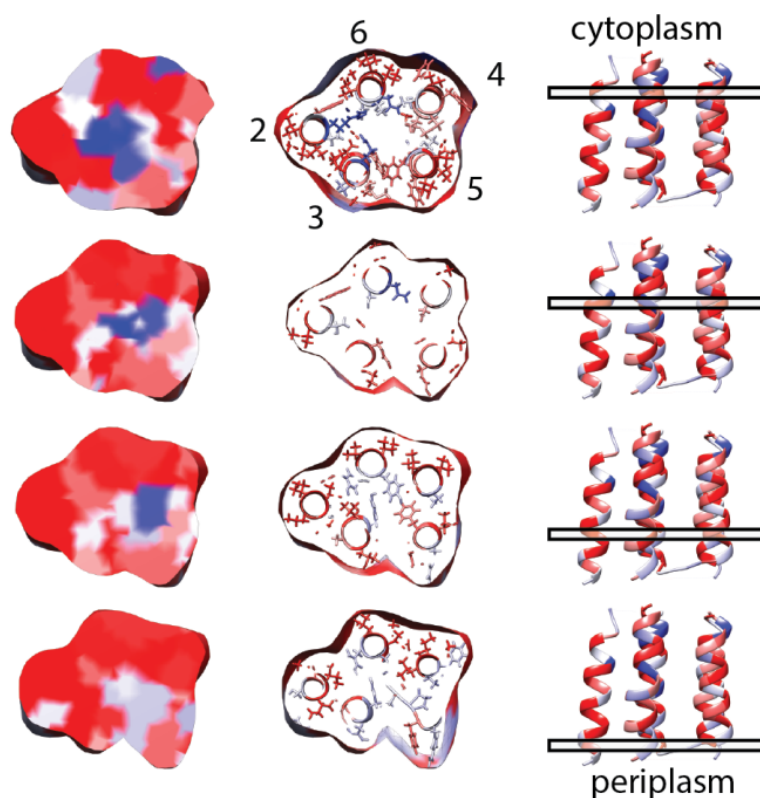


Figure 25: YidC forms a polar environment in the membrane

Distribution of hydrophilic (blue) and hydrophobic (red) residues at four sections at different height within the membrane. A hydrophilic cavity which is open towards the cytoplasm is formed by polar residues in the center of the TM core. On the periplasmic side apolar residues seal this cavity by tight stacking interactions of aromatic residues. Adapted from (Wickles et al., 2014)

Taken together, molecular dynamics simulation of the covariance based structural model provided useful insights into the stability of YidC in its native membrane environment. The model was found to be stable throughout the simulation, suggesting that it represents a possible physiological conformation of YidC. Furthermore, structurally important residues based on simulation also proved to be relevant for protein activity *in vivo*. The ability of YidC to create a hydrophilic environment within the membrane might play a central role in substrate insertion.

3.5 Cryo-EM Structure of the Active YidC bound to F₀c-RNCs

Structural insights into co-translational events can be accessed by cryo-electron microscopy (Becker et al., 2009; Frauenfeld et al., 2011; Gogala et al., 2014; Seitz et al., 2014). Studies addressing SecY-dependent membrane insertion/translocation have shed light on these essential cellular events on a molecular scale. For Sec-independent substrates using YidC as membrane insertase the functional state (monomeric/oligomeric), ribosome contacts and substrate binding sites are controversially discussed (Kohler et al., 2009; Kedrov et al., 2013; Seitz et al., 2014). Here, single-particle cryo-EM on the biochemically well-defined YidC:F₀c-RNC complex (see 3.2.3) was carried out to gain novel structural insights into co-translational membrane insertion by YidC.

The YidC:F₀c-RNC complex was reconstituted (see 2.7.1), vitrified on holey carbon grids and data was acquired under low-dose cryo conditions using the FEI Titan Krios electron microscope (see 2.7.2). Three dimensional density maps were calculated using single particle reconstruction (see 2.7.7). The data set (1151955 particles) was sorted (see 2.7.10) using competitive projection matching and focused sorting (Leidig et al., 2013). An additional particle cleaning step was achieved by initial sorting using the Edge-volume (see 2.7.9). Therefore, only particles which can be aligned well to ribosomal projections remain in the dataset.

Table 4 : Sorting scheme of the cryo-EM dataset

windowed particles	1151954		
ribosomal particles	876375	non alignable particles	275579
RNC + weak ligand density	500407	empty 70S ribosome	375968
RNC + ligand density	224757	RNC + no ligand	275650
RNC + stable ligand conformation	61760	RNC + flexible ligand	162997
final dataset (after CC-cutoff)	58960		

Only particles of sub-data sets having additional, non-ribosomal density at the ribosomal tunnel exit were subjected to further sorting rounds. The reliability of the additional density observed after sorting was verified by selective back-projection of the subset of particles with the orientation parameter obtained before sorting, thereby

eliminating any bias acquired during sorting. The sorting was continued until the local resolution of the ligand density did not improve any more. The cryo-EM reconstruction (Figure 26) was refined (see 2.7.8) to an overall resolution of $\sim 8 \text{ \AA}$ ($\text{FSC}_{0.5}$) (Figure 27) and a local resolution map was calculated using RESMAP (Figure 28).

The density map clearly showed features of a tRNA in the ribosomal P-site indicating a stalled ribosome nascent chain complex. Additional, non-ribosomal density was observed at the tunnel exit site of the ribosome. The protein conducting channel SecY (Frauenfeld et al., 2011; Bischoff et al., 2014b) as well as factors for nascent chain processing like trigger factor, SRP and MAP were shown to bind to the ribosome at this site (Kramer et al., 2009). The resolution of this density is high enough to discriminate the strong ligand protein electron density from the weaker electron density of the detergent micelle. Parts of the nascent chain (Figure 26 **B**) were traced from the tRNA through the ribosomal tunnel into the ligand density.

For more detailed interpretation of interaction and binding sites, atomic structures for the ribosome and the structural model of YidC were docked into the electron density map.

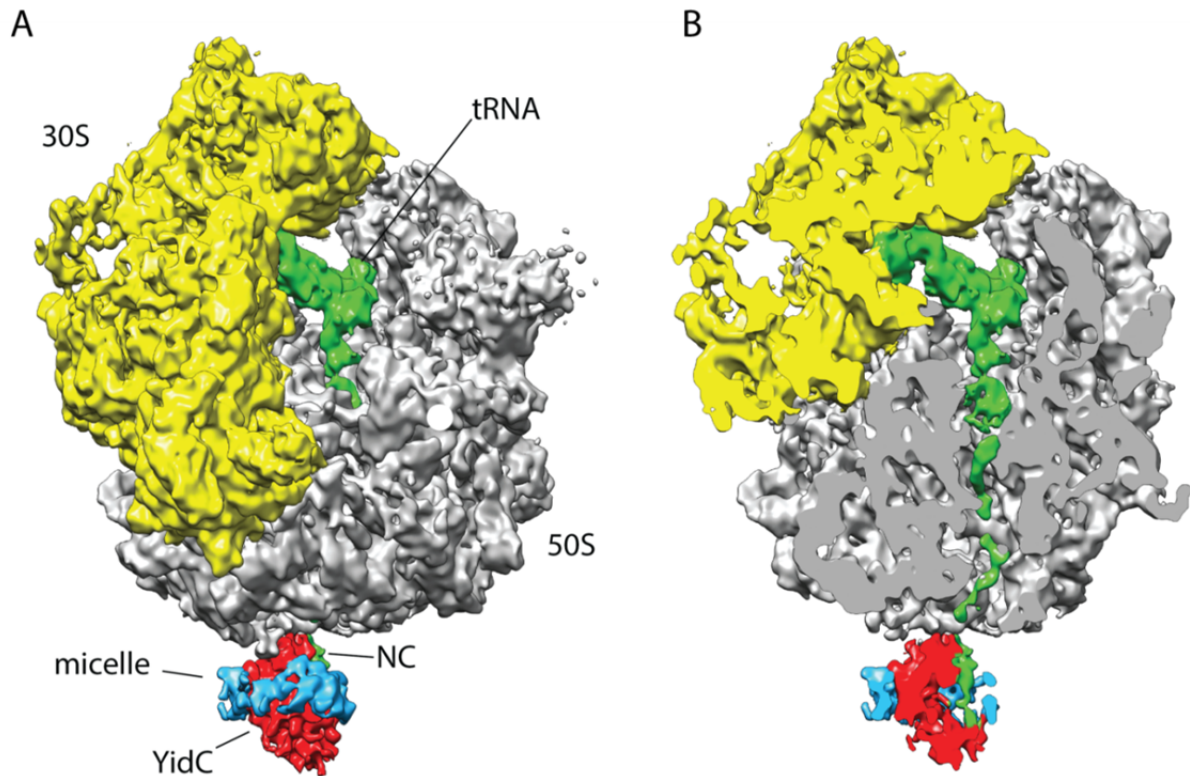


Figure 26: Cryo-EM reconstruction of YidC bound to F₀C-RNCs

Three dimensional density map of the YidC:RNC complex at $\sim 8 \text{ \AA}$ resolution. **A.** The clear density for tRNA (green) in the ribosomal P-site in the interface of the small (yellow) and large (grey) ribosomal subunit indicates a stalled ribosome nascent chain complex. Extra density at the tunnel exit can be assigned to the membrane protein YidC (red) surrounded by a micelle (blue). **B.** view and coloring as in **A**, cut through the ribosomal tunnel reveals density for the nascent chain (green) which can be partially traced from the tRNA into the transmembrane domain YidC. Adapted from (Wickles et al., 2014)

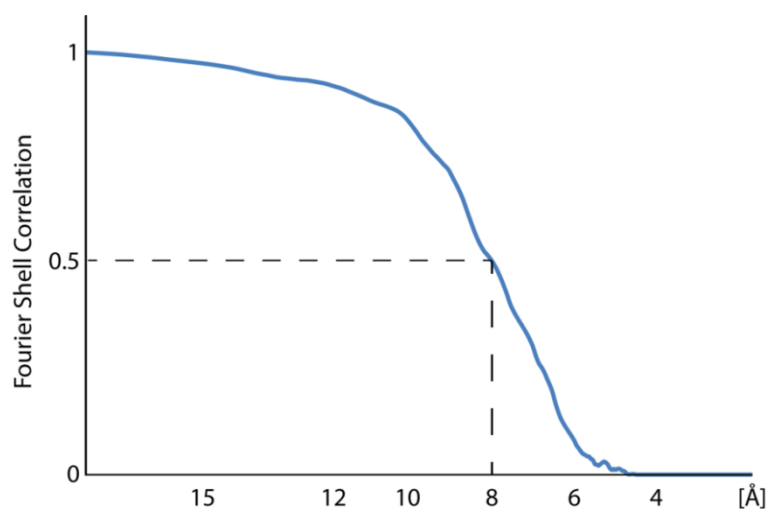


Figure 27: Fourier shell correlation plot

The resolution of the reconstruction was determined from the FSC plot at a correlation of 0.5.

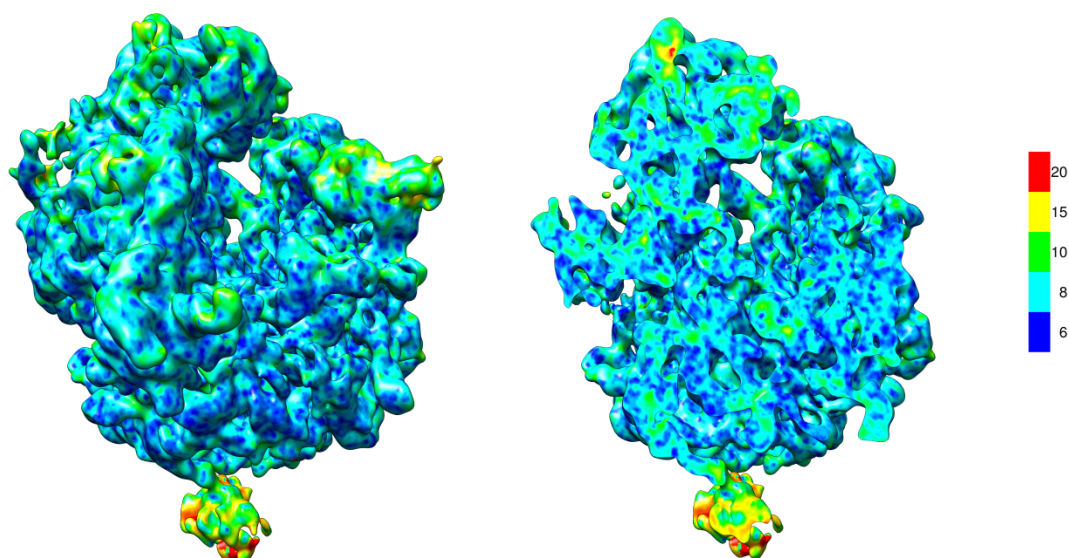


Figure 28: Local resolution of the reconstruction

Local resolution was calculated using ResMap based on unfiltered reconstructions of halfsets. The ribosomal core shows higher resolved features compared to the resolution of the ligand density.

3.5.1 Structural Model of YidC Bound to the Translating Ribosome

The stable conformation of the conserved core of YidC (see 3.4) was used for initial docking (Figure 29). Taking the presence of the substrate helix (F₀C) and TM1 of YidC into account the extra density at the tunnel exit could be accounted for only in one possible orientation. This docking allowed the placement of the F₀C-helix right next to TM3 of YidC where it was shown to form direct disulphide crosslinks with M430 (see 3.2.3).

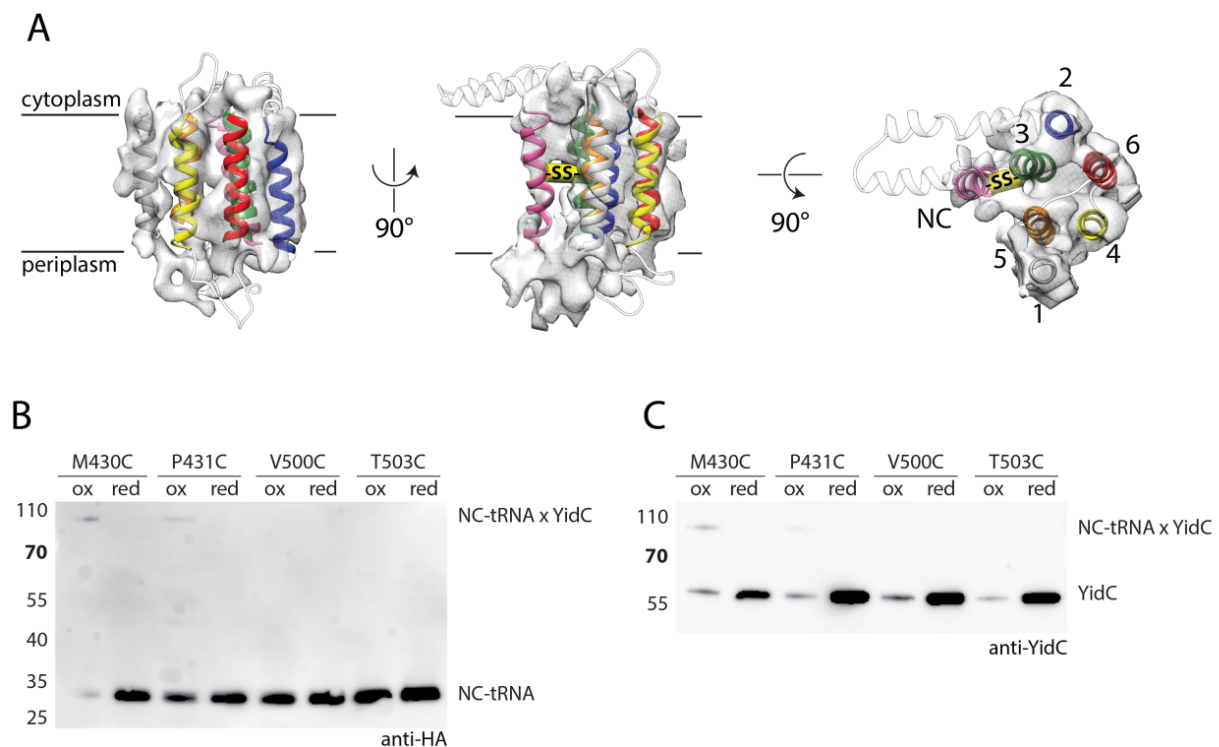


Figure 29: YidC substrate is located close to TM3

A. The structural model of YidC in complex with the first TMH of F₀C bound to the translating ribosome. Two views from within the membrane and one from the cytoplasm show the TM helices of YidC and the nascent chain (magenta). The previously identified position of the disulfide crosslink of substrate and YidC is depicted as yellow bond (-SS-). The density of the detergent micelle was removed for clarity. An additional density close to TM5 was assigned to TM1 of YidC. **B, C.** In-vitro cross-link experiments of F₀C(G23C)-RNCs with depicted single cysteines in YidC analysed by western blotting and antibody detection against nascent chain (HA-tag) and YidC. The nascent chain can cross-link to residues in TM3 (M430, P431) but not to TM5 (V500, T503). Modified from (Wickles et al., 2014)

An additional helical density close to TM5 and TM4 could be assigned to TM1 of YidC which was not part of the structural model. No density could be observed for the HPD, probably due to its high flexibility (Figure 23). The periplasmic P1 domain of YidC is connected to the transmembrane helices TM1 and TM2 with flexible linkers and could therefore not be visualized completely. All extra density at

the tunnel exit could be explained by docking the model of a single YidC monomer.

Taken together, the structural model of YidC was docked into the experimental electron density map and additional density could be assigned to substrate helix (F₀C) and TM1 of YidC. The size of the extra density reflects an active monomeric YidC in complex with its substrate helix.

3.5.2 Interaction of YidC with the Ribosome

The contact sites of YidC to the ribosome were analysed after docking the YidC model into the extra density. Structural information about the TnaC-stalled ribosome was taken from PDB: 2QAL and 2QAM. The density for H59 shows a different conformation in the map compared to the model where no ligand is bound at the ribosomal tunnel. The helix is moved towards the extra density and shows direct contacts to the cytoplasmic loop 2-3 of YidC (Figure 30 **A**). The bioinformatics analysis of residues in this loop revealed two positions with highly conserved aromatic residues at YidC₃₇₀ and YidC₃₇₇ (Figure 31).

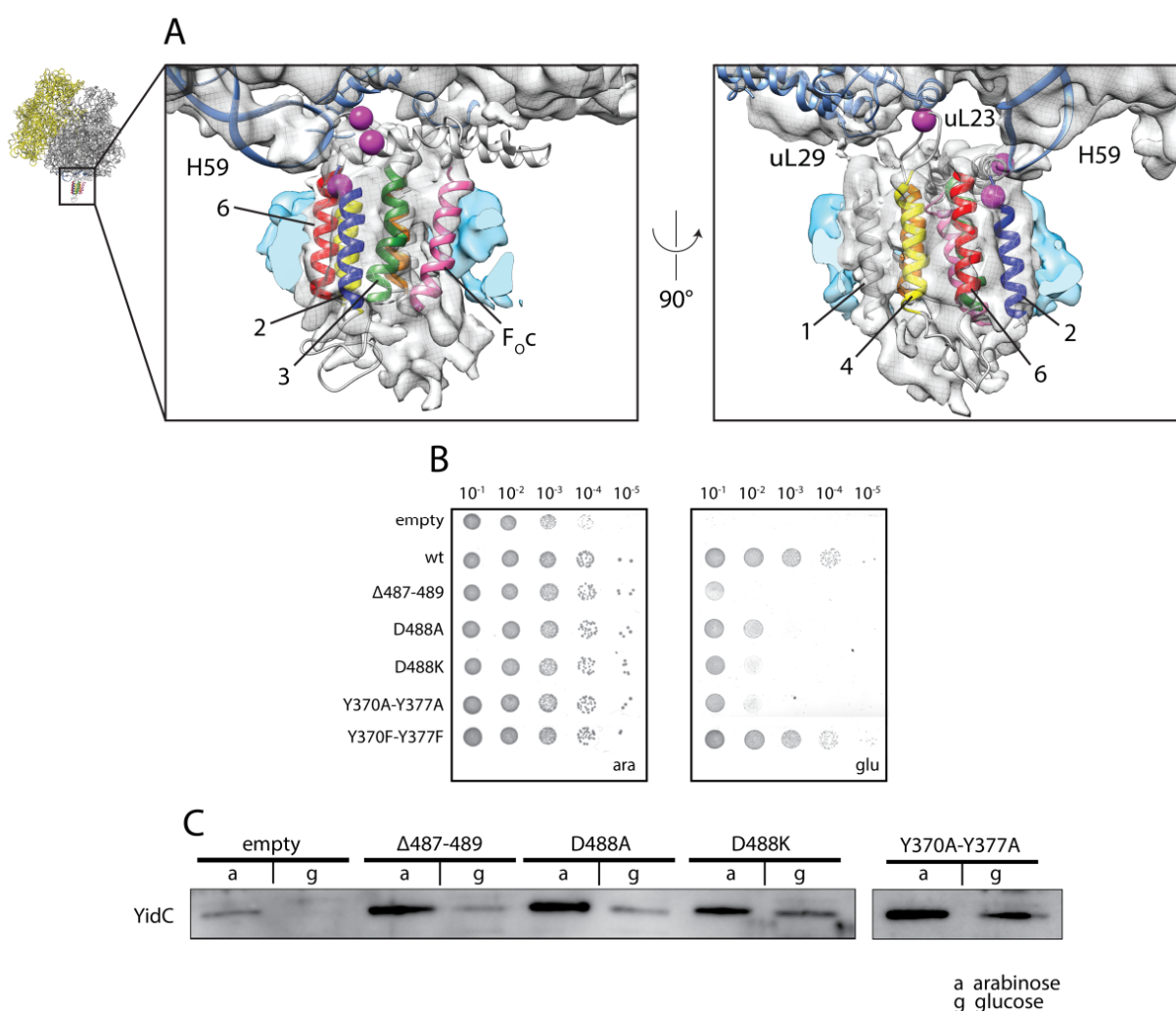


Figure 30: YidC interacts with H59 and uL23 of the ribosome

A. Contact sites of YidC to the ribosome. Ribosomal RNA helix H59 is contacting YidC at the cytoplasmic loop 2-3 near TM2 (magenta spheres). A second contact site is located in loop 4-5 of YidC close to ribosomal protein uL23. **B.** Complementation assay of possible ribosome interacting residues. **C.** Western blot of YidC inactivating mutants. Modified from (Wickles et al., 2014)

Mutations of these residues to alanine showed complete inactivation of the protein in an *in vivo* complementation assay whereas mutations to phenylalanine showed no effect (Figure 30 B). This indicates that aromatic residues at these positions are necessary for the activity of YidC.

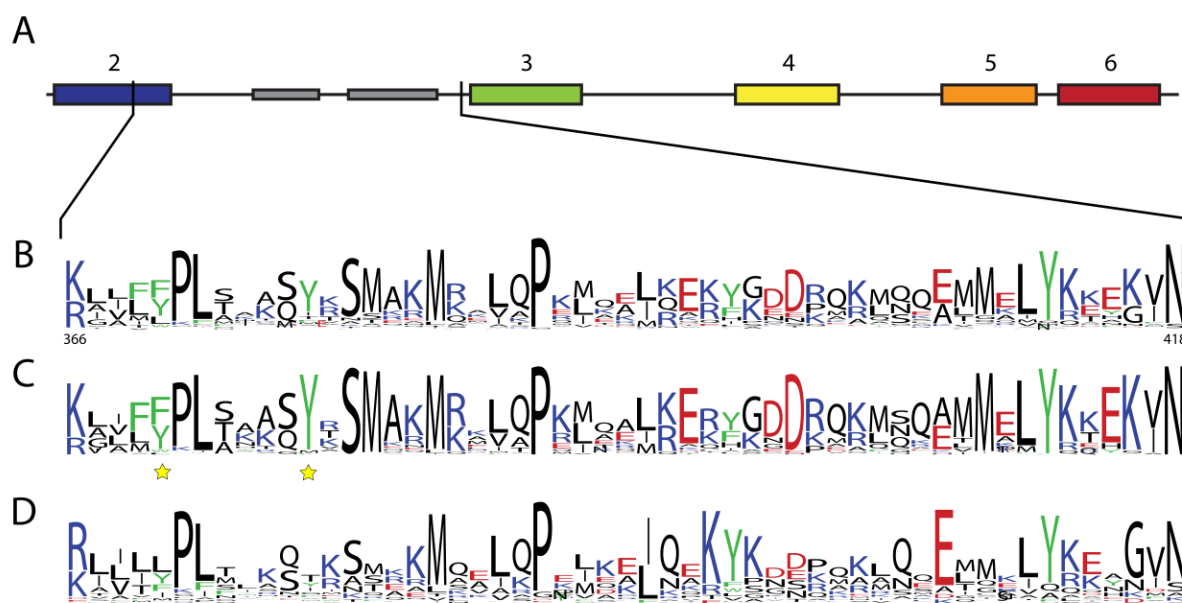


Figure 31: The HPD is enriched with conserved residues

Sequence logo of the helical paddle domain. **A.** Schematic representation of the conserved domain of YidC. Region used for sequence logo generation is indicated. Sequence logos of all sequences (**B**), only proteobacteria (**C**) and non-proteobacteria (**D**) show high conservation in length and at specific positions. The conserved aromatic residues in proteobacteria are marked with a star.

In addition, residues in the cytoplasmic loop 4–5 of YidC can contribute to the interaction with the ribosome. Analysis of the loop residues revealed the conserved negatively charged amino acid D488 close to the middle of this loop. Deletions of this region ± 1 ($\Delta 487$ –489) inactivated YidC (Figure 30 B). Single point mutations (D488A, D488K) were also sufficient to severely reduce its activity. This suggests an important interaction between the negatively charged residue in YidC and a positively charged residue of a ribosomal protein.

Taken together, potential interaction sites were found by a crystal structure of the ribosome and the molecular model of YidC into the cryo-EM density. Three positions within cytoplasmic loops of YidC were shown by *in vivo* complementation assays to be very important for its activity. These mutants were stably expressed (Figure 30 C) but inactivated the protein.

YidC is positioned by its specific interactions with the ribosome in a defined orientation close to the tunnel exit, thereby defining a potential path for substrate insertion. The density for the F_{OC} substrate helix was found to be located between TM3 of YidC and the tunnel exit of the ribosome (Figure 32).

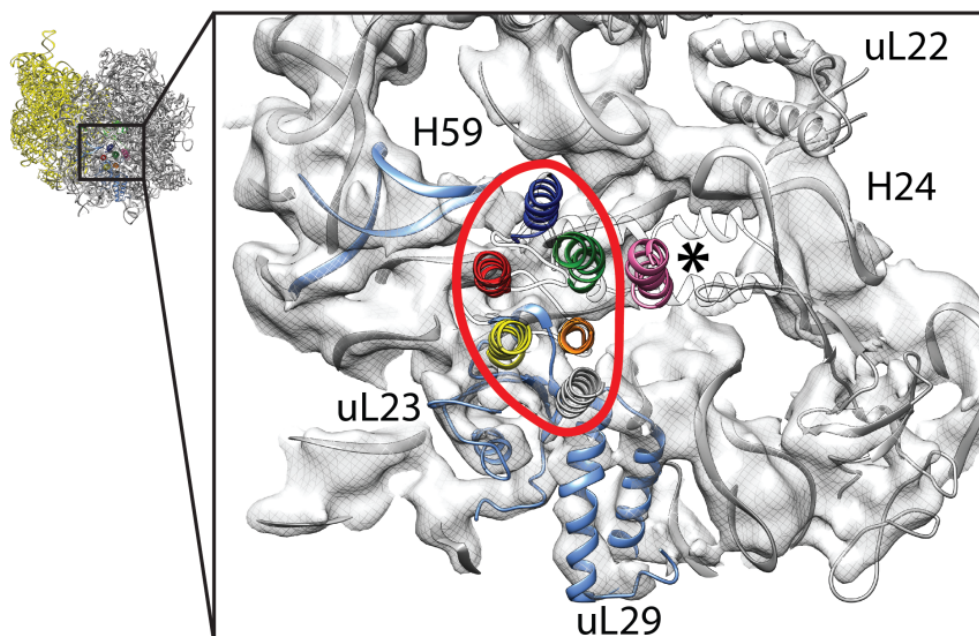


Figure 32: Tunnel exit view of the YidC binding site

YidC binds close to the tunnel exit of the ribosome. The conserved TM core of YidC (red outline) interacts with H59 and uL23 of the ribosome and is thereby positioned closely to the tunnel exit (marked with *). The substrate helix F_{OC} (pink) is located between the tunnel exit and YidC. Adapted from (Wickles et al., 2014)

The influence of YidC on the lipid bilayer was tested by measuring the minimal distance between the polar head groups of lipid molecules in different sides of the membrane during the MD simulation. YidC distorts the lipid bilayer by inducing membrane thinning close to TM3 and TM5 (Figure 33). This is mainly caused by hydrophobic mismatch of the short TM helices and the aliphatic lipid tails. The lipids contacting YidC have to adopt their tail conformation and are therefore more disordered compared to a native lipid bilayer. This might be an important feature of YidC for facilitating membrane protein insertion.

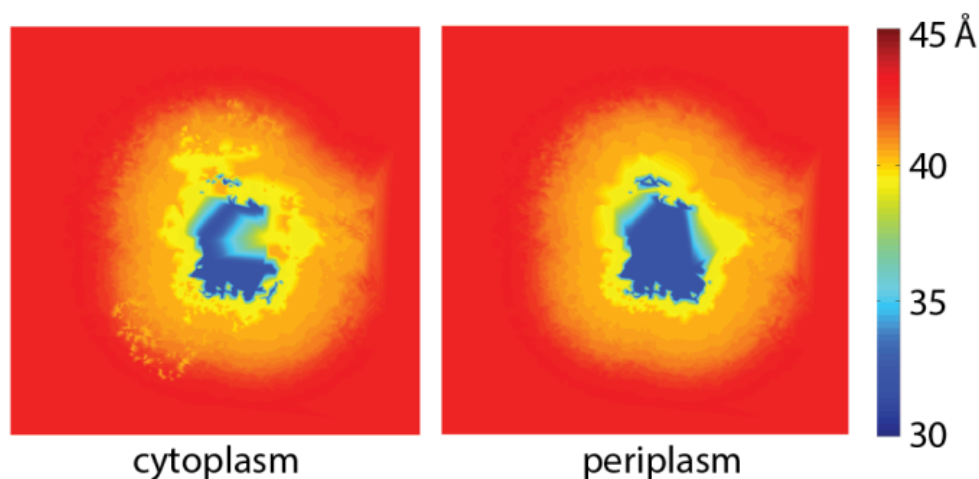


Figure 33: YidC causes membrane thinning at TM3/TM5

Thickness of the lipid bilayer averaged over the MD simulation shown as contour plot. YidC is positioned as in Figure 32 with TM3/TM5 being on the right side in both plots. The cytoplasmic and periplasmic sides were analyzed separately. The thickness was measured as the distance between the polar lipid head groups on either side of the membrane and shows thinning close to TM3/TM5 of YidC due to hydrophobic mismatch caused by the short TM helices. Adapted from (Wickles et al., 2014)

In summary, it can be stated that novel structural insights were gained using cryo-EM together with bioinformatics analysis and *in vivo* experiments on the biochemically well-defined YidC:F_{OC}-RNC complex. These findings suggest that nascent chains are inserted into the membrane at the protein-lipid interface of a monomeric YidC which is bound to ribosomal helix 59 close to the tunnel exit.

3.6 Reconstitution of YidC into Nanodiscs

The structural data obtained in this thesis was based on a detergent-solubilized membrane protein (see 2.2.2). Studies have shown that membrane proteins can be reconstituted into their native lipid environment for biochemical and structural investigations (Frauenfeld et al., 2011). Therefore, YidC was reconstituted into nanodiscs containing native *E.coli* lipids (see 2.2.7) using an adopted method from (Frauenfeld et al., 2011). Briefly, detergent solubilized YidC with His-tag (see 2.2.2), scaffold protein Apo-A1 and lipids (see 2.2.6) were mixed in the presence of cholate. Nanodisc formation was induced by depleting detergent from the mix using the Biobeads sorbent. Nanodiscs with and without incorporated YidC were separated by size exclusion chromatography (Figure 34) followed by a metal affinity column (Figure 35).

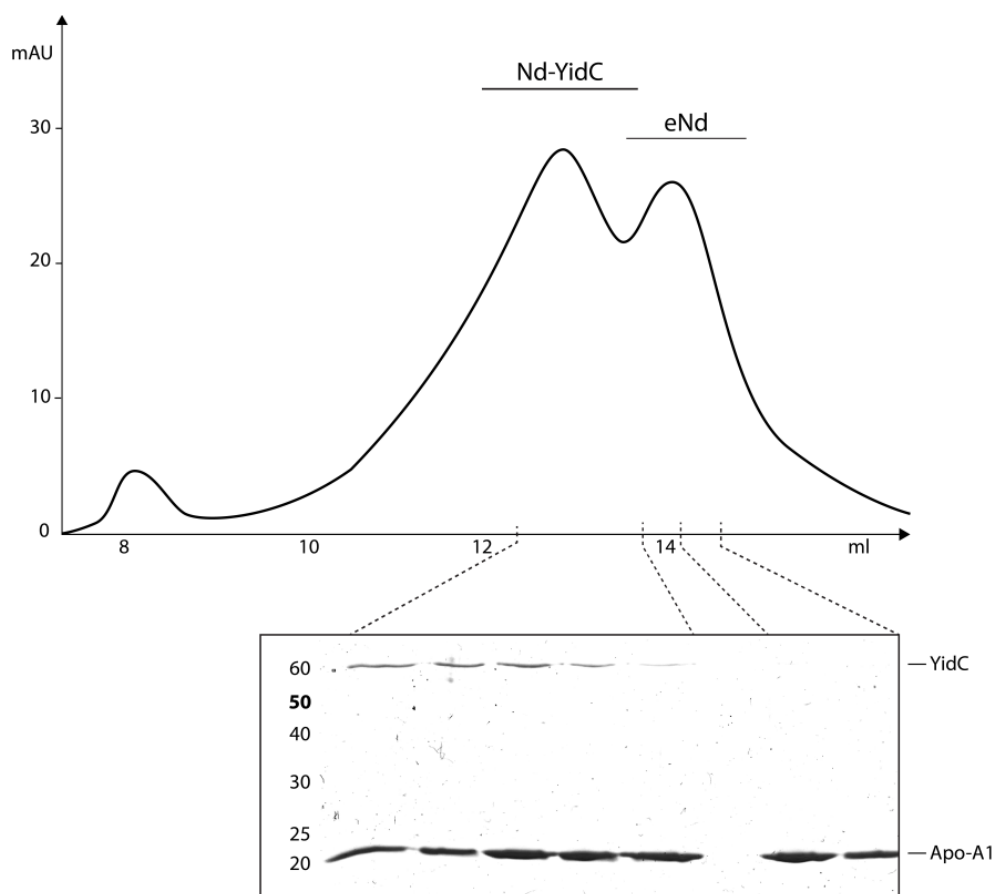


Figure 34: Size exclusion profile of YidC nanodisc reconstitution

The elution profile of the YidC–nanodisc preparation using size exclusion chromatography shows two peaks at ~ 12.5 ml and ~ 14 ml. Peak fractions were analyzed by SDS–PAGE followed by Coomassie staining of the gel. The 1st peak contains YidC in complex with the scaffold protein and the 2nd peak contains only scaffold protein.

Fractions of the two peaks after size exclusion chromatography were analysed by SDS–PAGE and Coomassie staining. Only fractions of the 1st peak showed presents of YidC and scaffold protein Apo–A1 indicating the formation of a water soluble YidC–Nanodisc complex. The 2nd peak fractions only contain scaffold protein corresponding to empty nanodiscs. The ratio between YidC and Apo–A1 suggests that these factions still contain empty nanodiscs which could be separated using the His–tag still attached to YidC. Therefore, YidC incorporated into nanodiscs was bound to metal affinity matrix. Empty nanodiscs were washed away and the YidC–nanodisc complex was eluted in stoichiometric ratio of 1:1 from the column (Figure 35).

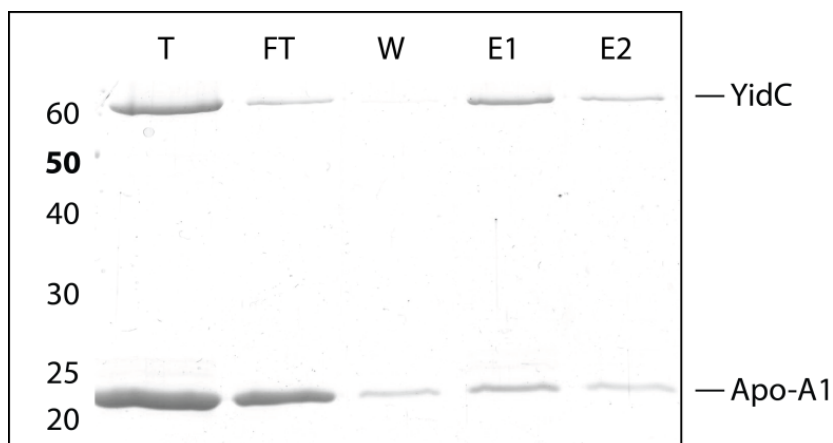


Figure 35: Purification of YidC incorporated nanodiscs

YidC-nanodisc fraction after size exclusion chromatography (T) was loaded on a metal affinity column. Empty nanodisc were washed (W) from the column and YidC-nanodisc complex was eluted (E1, 2).

Taken together, detergent solubilized YidC could be reconstituted into its natural lipid environment using the Apo-A1 scaffold protein.

4 Discussion

4.1 A fast and accurate Pre-Processing Workflow for Single Particle Analysis

Existing pre-processing steps of single particle analysis require intensive user involvement to assess the quality of the dataset. Although ribosomal structures can be obtained at near atomic resolution from a few thousand particles (Bai et al., 2013), this can only be achieved having a direct electron detector and more important, a highly homogeneous dataset. In contrast to ribosomal subunits (Amunts et al., 2014; Brown et al., 2014; Greber et al., 2014a; Greber et al., 2014b), ribosomal complexes involved in physiological processes show a higher degree of structural heterogeneity (Frauenfeld et al., 2011; Hashem et al., 2013; Leidig et al., 2013; Bischoff et al., 2014b; Gogala et al., 2014). Therefore, extensive sorting of the dataset is required to find the most homogeneous subset of particles yielding the highest resolved reconstruction. In addition to structural heterogeneity, flexibility of some parts of the complex also demands further sorting into stable sub-populations for getting structural information. Thus, achieving high resolution structure of a biological complex requires the efficient processing of large datasets and sorting of the particles into stable classes.

To that end, a fast and accurate pre-processing workflow for single particle analysis was build which can process large dataset with minimal user interaction. Processing steps requiring time-consuming user interaction were examined and automated alternatives were evaluated. Power spectra are representing the information content of a cryo-EM micrograph. For high resolution reconstruction only those showing signal at high special frequencies and without aberrations are desired for processing. To automatically evaluate the quality of the power spectra, its rotational symmetry of a small region under a mask was calculated. This symmetry value is quickly calculated and sufficient to rank the quality of the power spectra. The threshold was set to discard power spectra showing drift or astigmatism and to include only power spectra with high information content.

Next, the particle selection method from cryo-EM micrographs was improved using a machine learning-based approach. The detection of particles from micrographs is always a trade-off between minimizing the number of false positive particles and thereby missing a large fraction of potential particles or maximizing the

number of particles with the risk of including contaminations into the dataset. In our approach, the particle detection phase was therefore separated from the particle classification phase. This ensures a maximal particle yield from the micrographs in the initial detection phase with only a small fraction of contaminations left in the dataset after classification. A machine learning algorithm was developed which classifies the particles based on a small test dataset provided by the user. The test dataset consisting of particles and non-particles is representative for the complete dataset. Existing methods focus on the simultaneous detection and classification of particles from micrographs (Nicholson and Glaeser, 2001). This is achieved by providing similar references, selecting template particles or statistical image manipulation. Thresholds in these approaches can only be set from one direction. The classifier of MAPPOS can be trained with both, particles and non-particles, which allows a more accurate adjustment of the parameters. The performance of MAPPOS was approved on simulated as well as real cryo-EM data and results showed the human-like performance. This automated workflow enables the quick processing of datasets with minimal user involvement and allows therefore the parallel screening of multiple dataset or the fast evaluation of large datasets.

Taken together, this new pre-processing pipeline will significantly reduce the manual interaction in single particle processing and finally contribute to a competitive workflow of high resolution cryo-EM.

4.2 Hybrid Method Approach for Membrane Protein Modelling

In this study, a novel approach was taken for modelling a membrane protein structure without prior structural knowledge. Using recently developed bioinformatics tools it was possible to define sufficient structural restraints for *ab initio* model building. Well-established MD simulations were combined with *in vivo* functional complementation studies to validate the derived model. This structural model of the conserved part of YidC was further used for interpretation of a cryo-EM map of the functional ribosome:YidC complex.

During the course of this study two crystal structures (3WO6, 3WO7) of *Bacillus halodurans* YidC2 (BhYidC2) were published (Kumazaki et al., 2014). This allowed for the direct validation of the structural model by comparison to crystal structures resolved to 2.4/3.2 Å. To that end, an *E.coli* homology model based on pdb:3WO6 (34% sequence identity to *E.coli* YidC) was generated using HHPred (Hildebrand et al., 2009). Residues of transmembrane helices were coloured according to the structural model and covariance based interaction pairs were indicated by yellow pseudo-bonds.

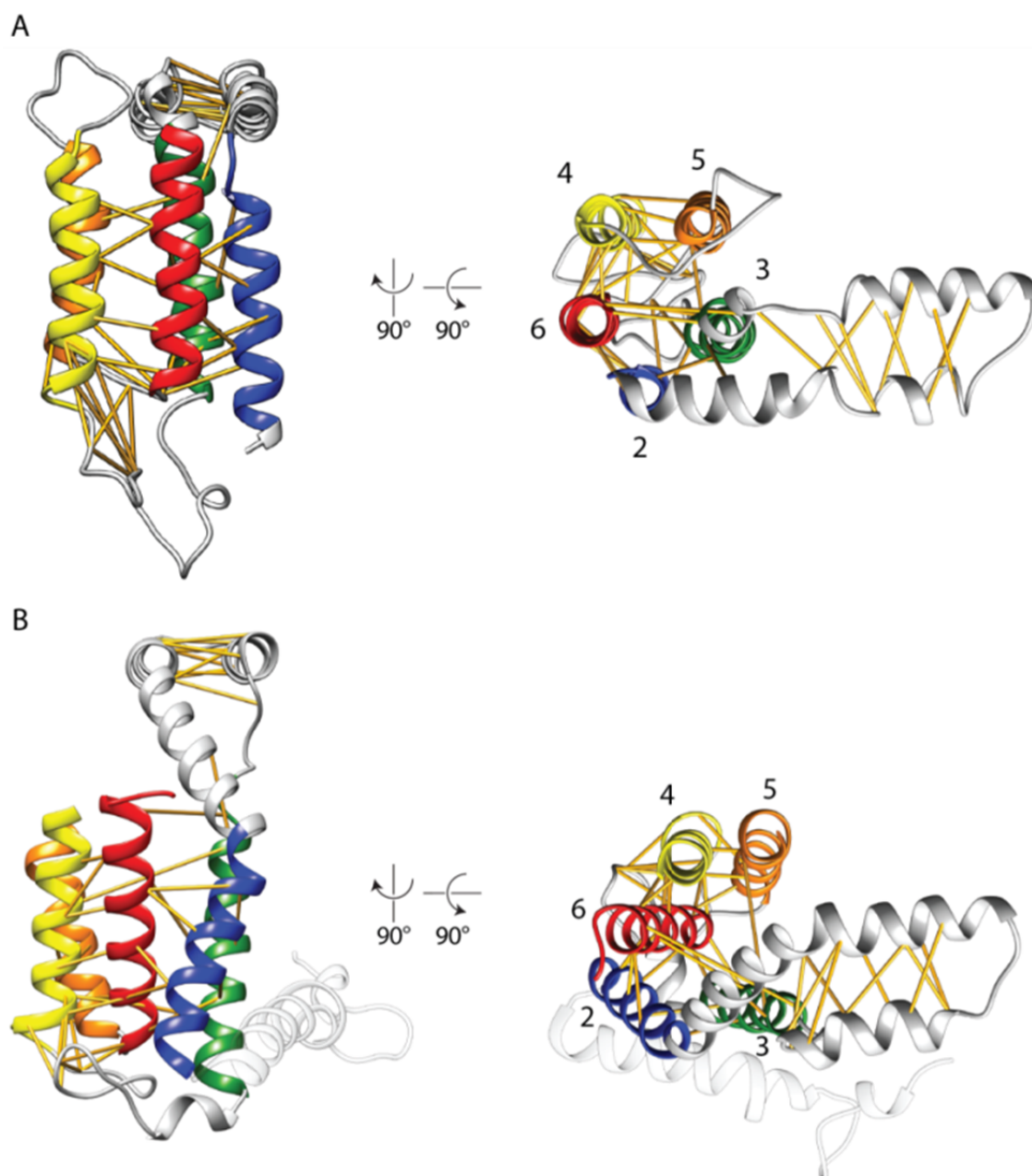


Figure 36: Comparison of the structural model and the crystal structure

The covariance-based structural model of YidC (**A**) is compared to an *E. coli* homology model based on the BhYidC2 crystal structure (3WO6) (**B**). Covariance-based interactions are depicted as yellow pseudo-bonds in both structures. Adapted from (Wickles et al., 2014)

The overall arrangement of helices is very similar in both structures. The TM helices are positioned in the same order (TM 5, 3, 2, 6, 4) and the root mean square deviation (RMSD) between c_{α} atoms of model and crystal structure of 7.5 Å is within the resolution limits of the covariance-based modelling. Individual helices are tilted in the crystal structure with respect to the membrane whereas no tilt was

implied for the structural model. The HPD is pointing in the same direction but shows a different angle relative to the membrane surface.

The spatial alignment of all three models allowed a detailed analysis of the structural differences in distinct parts of the protein (Figure 37). The two crystal structures show an almost identical fold of the C-terminal helices (TM4–6). Yet, significant differences can be observed in the conformation of the cytoplasmic loop region connecting TM 2 and TM 3 as well as TM 3 itself (Figure 37 red curve).

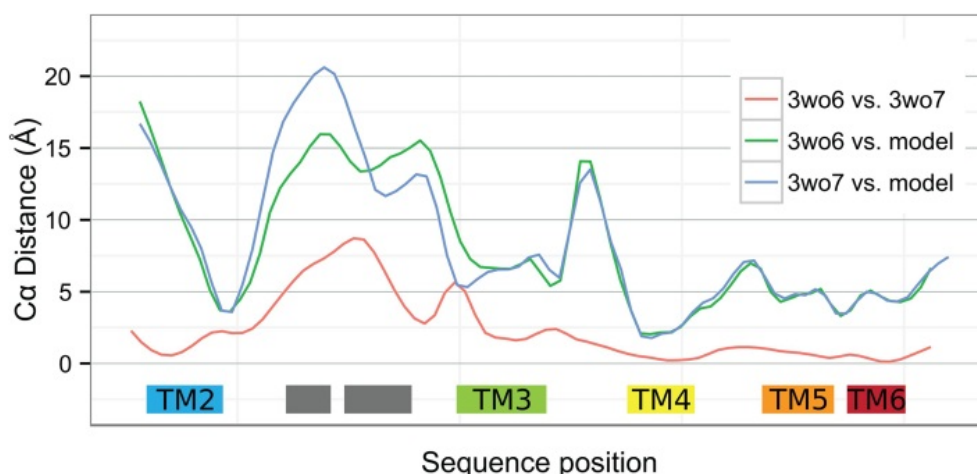


Figure 37: Backbone distances of YidC crystals and structural model

The distance of c_{α} -atoms between the crystal structures (red curve) and between the structural model and the crystal structures (green, blue curve) is plotted for each residue of the conserved core. Adapted from (Wickles et al., 2014)

The comparison of the structural model to the crystal structures revealed high similarity of the C-terminal region (TM4–6). Their relative position is almost identical given that helix tilt was not applied for the model. Also in good agreement with the fold observed in the crystal structure is the cytoplasmic end of TM 2 which shows increasing difference towards the periplasm. This difference might be explained by the missing amphipathic helix N-terminal of TM 2 which is present in both crystal structures but not in the model. The movement of the periplasmic end of TM 2 in the model is therefore less restricted resulting in a higher deviation from the crystal structure. The periplasmic loop connecting TM 3 and TM 4 is completely resolved in the crystal structures due to the existence of partial secondary structure and interactions with the periplasmic ends of the TM helices. Covariance restraints existed only for some residues close to TM 4 which were not sufficient to model this loop as precise as the conformation observed in the crystal structure. Furthermore, the covariance restraints used for modelling showed only a few coupled

residue pairs in the Loop 3–4 which were apparently also not sufficient to build the complete loop with high accuracy. The structured cytoplasmic loop between TM2 and TM3 is present in two different conformations in the crystal structures. Comparison with the structural model showed the highest deviation in this region of the protein. This is consistent with the high B-factor of the loop in both X-ray structures and the flexibility analysis of the structural model (see 3.4.1) which are suggesting large movements relative to the membrane surface. The high degree of flexibility might be relevant for activity, although contradictory results have been published about the importance of the loop (Jiang et al., 2003; Chen et al., 2014).

The differences in the relative position of TM3 to TM5 in crystal structures and structural model might be of mechanistic importance. A hydrophobic cavity observed for the structural model (see Figure 25) as well as the crystal structures is open towards the cytoplasm and its accessibility towards the lipid environment is restricted by TM3 and TM5. This opening differs in the two crystal structures due to a shift of the cytoplasmic end of TM3, which is in a more closed conformation in 3WO7 compared to 3WO6. Interestingly, in the structural model this cavity is even more closed, because structural restraints based on covariance analysis between TM3 and TM5 (Pro⁴²⁵–Pro⁴⁹⁹) and between TM3 and TM6 (Cys⁴²³–Gln⁵²⁸; Phe⁴³³–Thr⁵²⁴) were applied during modelling. Remarkably, only these three residue pairs are outliers when measuring the distance between c_{β} -atoms of evolutionary coupled residues in the crystal structure (see Table 5). Therefore, it is tempting to speculate that the movement of TM3 and the accompanied accessibility of the hydrophobic cavity is of functional importance and might play a crucial role during substrate insertion.

Taken together, a structural model of YidC was built based on novel bioinformatics tools. The quality of the model could be assessed by comparison to the recently published crystal structure. The overall fold of the structural model agrees remarkably well with the BhYidC2 crystal structures. The differences observed between the structures might have mechanistic implications that need further biochemical and structural characterisation. Finally, these results further strengthen the importance of evolutionary analysis and might help to guide future structural work.

Table 5: Top 50 scoring residue pairs in covariance analysis

Residue 1	# Res 1	Region		Residue 2	# Res 2	Region	d _{model} (Å)	d _{3WO6} (Å)	Reason for exclusion
TRP	354	TM2		-	-	-			indel
GLY	355	TM2		-	-	-			indel
PHE	356	TM2		-	-	-			indel
PHE	356	TM2		ARG	533	c-term			topology violation
ILE	358	TM2	<->	GLY	512	Loop5-6	9,1	6,1	
ILE	359	TM2	<->	VAL	519	TM6	6,5	5,2	
ILE	359	TM2	<->	LEU	515	TM6	8,5	7,9	
ILE	359	TM2		-	-	-			indel
ILE	361	TM2	<->	LEU	436	TM3	7,9	8,2	
THR	362	TM2		PRO	371	TM2			intrahelical
PHE	363	TM2	<->	VAL	523	TM6	5,2	6,1	
GLY	367	TM2	<->	VAL	523	TM6	6,0	8,2	
MET	369	TM2	<->	ILE	432	TM3	9,9	8,4	
Leu	372	Loop2-3		PRO	510	Loop5-6			topology violation
SER	379	Loop2-3	<->	PRO	425	TM3	10,2	9,9	
LEU	386	Loop2-3	<->	VAL	417	Loop2-3	7,5	7,1	
LEU	386	Loop2-3	<->	LEU	411	Loop2-3	6,2	6,1	
PRO	388	Loop2-3		GLN	429	TM3			topology violation
LYS	389	Loop2-3	<->	ALA	414	Loop2-3	10,5	9,8	
LYS	389	Loop2-3	<->	GLU	415	Loop2-3	11,2	10,0	
ILE	390	Loop2-3	<->	MET	408	Loop2-3	6,8	6,2	
MET	393	Loop2-3	<->	ILE	404	Loop2-3	7,9	7,4	
MET	393	Loop2-3	<->	LEU	411	Loop2-3	8,2	7,7	
ARG	394	Loop2-3	<->	ILE	404	Loop2-3	8,5	8,1	
ARG	396	Loop2-3	<->	GLU	407	Loop2-3	8,9	8,4	
CYS	423	TM3	<->	GLN	528	TM6	16,2	20,9	
PRO	425	TM3	<->	PRO	499	TM5	10,2	20,5	
PHE	433	TM3	<->	THR	524	TM6	11,0	14,9	
LEU	436	TM3	<->	GLY	512	Loop5-6	7,6	8,3	
TYR	437	TM3	<->	LEU	513	Loop5-6	9,8	6,4	
TRP	454	Loop3-4	<->	ASP	462	Loop3-4	6,6	7,0	
TRP	454	Loop3-4	<->	PRO	468	TM4	16,0	11,5	
TRP	454	Loop3-4	<->	SER	511	Loop5-6	9,8	8,3	
ILE	455	Loop3-4	<->	LEU	467	TM4	9,8	10,1	
ILE	455	Loop3-4	<->	ILE	466	TM4	11,0	8,0	
ASP	462	Loop3-4	<->	PRO	468	TM4	12,5	6,8	
ASP	462	Loop3-4	<->	SER	511	Loop5-6	11,1	4,2	
TYR	465	TM4	<->	LEU	507	TM5	10,4	8,7	
LEU	467	TM4	<->	LEU	515	TM6	11,6	6,6	
PRO	468	TM4	<->	LEU	513	TM6	14,5	8,8	
LEU	470	TM4	<->	ILE	518	TM6	6,3	5,4	
MET	471	TM4	<->	PHE	502	TM5	8,8	4,9	
GLY	472	TM4	<->	THR	503	TM5	6,7	5,3	
GLY	472	TM4		GLN	479	TM4			intrahelical
THR	474	TM4	<->	ASN	521	TM6	4,7	3,7	
THR	474	TM4	<->	ILE	525	TM6	6,7	7,8	
ILE	478	TM4	<->	ILE	525	TM6	9,0	5,0	
THR	485	Loop4-5		-	-	-			indel
PHE	506	TM5	<->	VAL	514	TM6	14,4	4,2	
GLY	512	Loop5-6		GLN	532	TM6			topology violation
Ø							9,3	8,1	

4.3 Co-translational Protein Insertion by YidC

The data obtained in this study allowed for a detailed structural interpretation of co-translational protein insertion by the YidC insertase. The quick purification and *in vitro* reconstitution of the complex as well as the sophisticated sorting of the cryo-EM data were essential for success. Crosslinking studies and *in vivo* complementation assays were combined with the structural data for molecular interpretation of the complex.

The cryo-EM map of the YidC:F₀C-RNC complex revealed a monomeric YidC bound to the translating ribosome. Although 2D projection maps of YidC (Lotz et al., 2008) as well as low resolution cryo-EM reconstructions of YidC bound to RNCs (Kohler et al., 2009) were interpreted as dimers, recent fluorescent data using detergent-solubilized and nanodisc-reconstituted YidC (Kedrov et al., 2013) showed that YidC can be fully active as monomer.

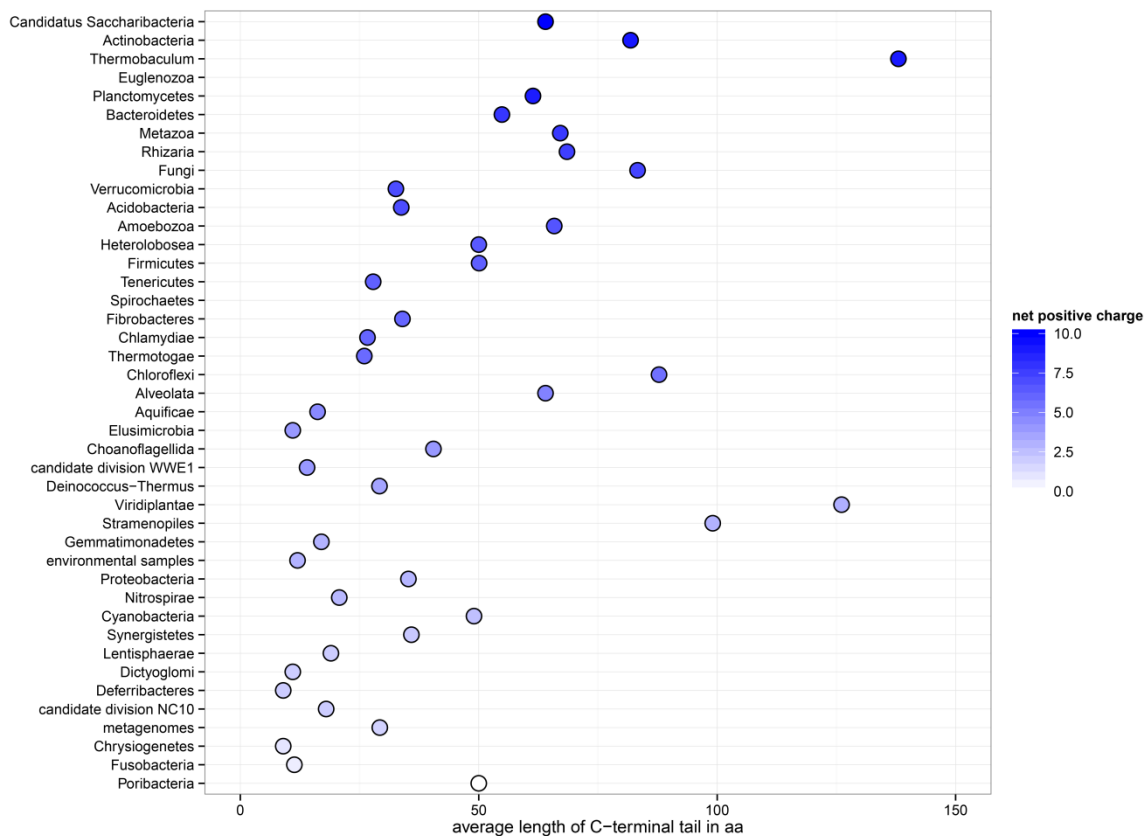


Figure 38: Average tail length and positive net charge in different organisms

The average length of the c-terminal tail for each branch was calculated based on the MSA and the net charge was determined.

Ribosomal binding of YidC is mediated by its positively charged C-terminal tail (Jia et al., 2003). The length and net-charge of the tail vary drastically among

different species (Figure 38) indicating a rather unspecific binding interface. TM6 of YidC is positioned relatively close to ribosomal RNA helix H59 at the tunnel exit. Therefore, it is tempting to speculate that the negatively charged phosphate backbone of the ribosomal RNA acts as binding platform for the positively charged N-terminal tail.

The cryo-EM reconstruction in combination with *in vivo* complementation assays revealed the importance of two additional regions of YidC which could explain the residual binding of the C-terminal truncated mutant (Kedrov et al., 2013). One strong contact is formed by interaction of the tetraloop at the tip of H59 and parts at the cytoplasmic end of TM2. Conserved aromatic residues in this region are essential for cell viability and might be involved in RNA binding.

The second ribosomal binding site is located within the cytoplasmic loop connecting TM4 and TM5 of YidC and is positioned close to ribosomal protein uL23. This ribosomal contact site is consistent with an earlier cryo-EM reconstruction (Kohler et al., 2009). Due to the high flexibility of the cytoplasmic loop, an *in vivo* complementation assay was used to determine the functionally important residues of the loop. One conserved negatively charged residue (D488) was found to be crucial for function which suggests an involvement in ribosomal binding. Replacement of TM4 and TM5 with a helical hairpin of LepB was shown to complement for function *in vivo* (Jiang et al., 2003). Analysis of the cytoplasmic loop connecting the two TM helices of LepB revealed also a negatively charged residue located in a similar position as in the YidC loop. Taken together, these interactions position a monomer of YidC at a certain area and orientation close to the ribosomal tunnel exit.

Interestingly, within the YidC:ribosome complex the position of the YidC-inserted nascent chains is similar to that of nascent chains translocated by SecY (Figure 39). But there are nevertheless important functional differences between SecY and YidC. SecY mediates translocation of hydrophilic loops across the membrane as well as insertion of TM helices into the lipid bilayer. Hydrophilic peptides use the water-filled channel within SecY to translocate to the periplasm passing a central constriction site which seals the inactive translocon. The exact path of a nascent TM helix from the ribosome via SecY into the membrane is still a field of ongoing investigation and not completely understood.

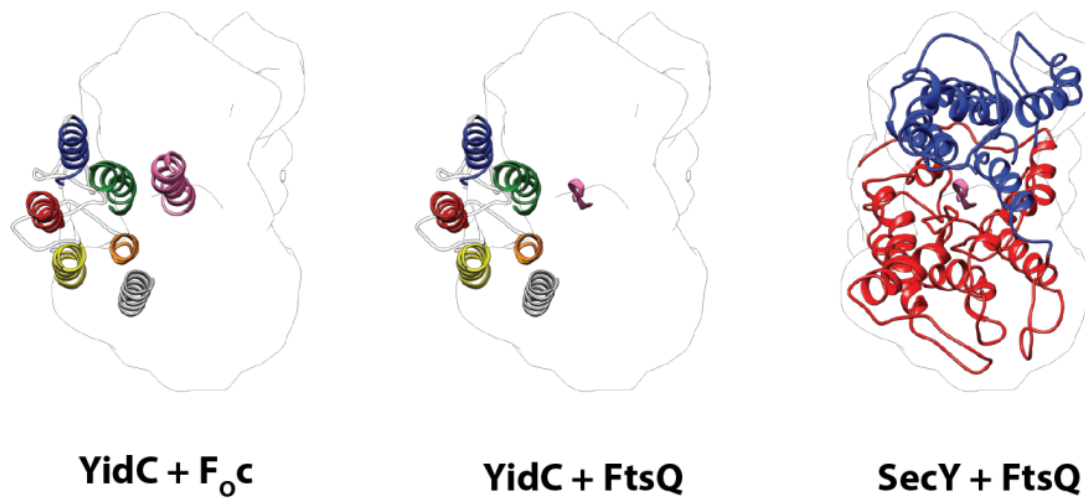


Figure 39: Comparison of active ribosome bound YidC and Sec

Model for active YidC and SecY were aligned to the ribosome. The structure of SecY (C-terminus: red, N-terminus: blue) in the native membrane environment is shown with the translocating nascent chain of FtsQ (NC: pink, right). The outline of SecY is imposed to the model of active YidC in complex with substrate helix F_{0c} (left). The path of the SecY mediated translocated nascent chain aligns with the position of YidC-mediated substrate insertion (middle). Adapted from (Wickles et al., 2014)

The latest model is based on the molecular dynamics simulation and biophysical considerations. The peptide must acquire secondary structure by forming hydrogen bonds (H-bonds) between the amino acid backbone to reduce the cost of insertion. This has direct implications on the path of nascent TM helix insertion. The central pore of SecY has (i) a hydrophilic environment and (ii) a limited space for acquiring secondary structure. Therefore, folding and insertion of nascent TM segments might occur already at the cytoplasmic site of SecY. This minimizes the exposure of the TM segment to the hydrophilic environment and facilitates interaction of the TM residues with the membrane core. The central pore of SecY facilitates translocation of the hydrophilic loops connecting TM segments whereas the TM helix inserts without entering the hydrophilic core. The clam-like structure of SecY can open laterally towards the lipid bilayer, thereby making the hydrocarbon core accessible to the hydrophobic amino acids of the TM segment. This interaction might drive the partitioning from the aqueous cytoplasmic site of SecY into the apolar lipid environment (Van den Berg et al., 2004; Frauenfeld et al., 2011; Bischoff et al., 2014b; Gogala et al., 2014) (Figure 40).

Unlike SecY, YidC mediates only insertion of TM segments into the lipid bilayer. The known YidC-only substrates share either a charged N-terminal tail or a short charged loops which have to be translocated (see Figure 3). In both cases polar

residues have to cross the hydrophobic membrane bilayer before the TM helix can be inserted. The partitioning of a TM helix into the lipid bilayer is mainly driven by the hydrophobic interaction of apolar residues of the inserted TM segment with the apolar lipid environment. Distinct properties of YidC might facilitate insertion by lowering the energy barrier for the translocation of polar residues across the membrane.

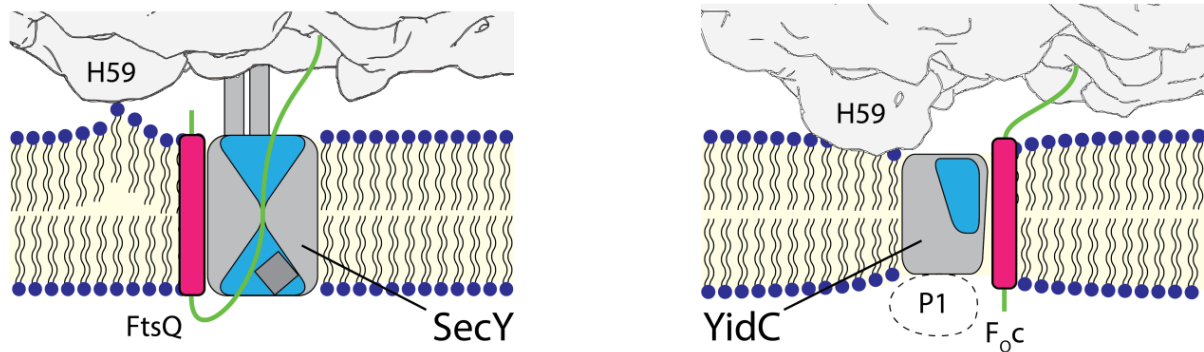


Figure 40: Comparison of protein insertion by SecY and YidC

SecY mediates TM helix insertion via opening of the lateral gate into the membrane whereas hydrophilic loops are translocated through a central pore across the membrane. YidC catalyzes insertion of substrate helices by lowering the energy barrier for the translocation of small hydrophilic moieties by providing a hydrophilic cavity within the membrane. Adapted from (Wickles et al., 2014)

The hydrophobic cavity formed by polar residues within the transmembrane helices of YidC is open towards the cytoplasm. This cavity is large enough to accommodate the polar moieties of the substrate nascent chains, thereby translocating them already half way across the membrane. Simultaneously, hydrophobic residues of the substrate start to interact with the apolar lipid tails, thus gaining energy for insertion. In combination with the thinning of the lipid bilayer induced by YidC, this might be sufficient to translocate the polar residues across the remaining membrane.

5 Summary and Outlook

In this work a new particle classification algorithm together with an automated workflow for single particle pre-processing of large datasets was developed. This allows fast screening and processing of automatically acquired cryo-EM images. The new pre-processing pipeline was used to determine the cryo-EM structure of an active YidC:ribosome complex engaged in co-translational membrane protein insertion. Molecular interpretation of the electron density was possible by building an *ab initio* structural model of YidC. This was achieved by combining several recently developed bioinformatics tools. Evolutionary co-variance analysis in combination with lipid-exposure prediction resulted into a set of structural restraints which could be satisfied in a monomeric model of YidC. The stability of the model in its native membrane environment was assessed by MD simulations and results could be validated by *in vivo* assays. The MD simulation revealed two important properties of YidC which might be important for function. First, a hydrophilic cavity within the membrane which is sealed towards the periplasm is formed by the transmembrane domain of YidC. Second, membrane thinning is induced by hydrophobic mismatch of the short TM helices next to TM3 and TM5. The model was docked into the non-ribosomal density of the cryo-EM map revealing the position of the inserted transmembrane helix of the F₀C substrate.

Taken together, the cryo-EM reconstruction of the YidC:F₀C-RNC complex in combination with MD simulations suggests an insertion site for nascent transmembrane helices at the protein-lipid interface. Membrane proteins inserted by the YidC-only pathway share a short hydrophilic moiety which has to be translocated across the membrane. YidC might act as a catalyst by lowering the energy of insertion. This is achieved by (i) shielding the hydrophilic parts of the substrate in the hydrophilic cavity, (ii) reducing the path for translocation by thinning of the membrane and (iii) allowing the interaction of hydrophobic substrate residues with the apolar membrane environment.

The structural insights gained in this study provide a good starting point for further analysis of YidC mediated membrane protein insertion. Trapped insertion intermediates based on the now available structural information will help to reveal the exact path of the nascent TM helix during insertion. In addition, structural data

of YidC in its native membrane environment will help to understand the influence and role of lipids for protein insertion. Furthermore, structural investigation of YidC in concert with SecY-mediated insertion is eagerly awaited to shed light on the membrane chaperon function of YidC.

6 References

- Abramson, J., I. Smirnova, V. Kasho, G. Verner, H.R. Kaback, and S. Iwata. 2003. Structure and mechanism of the lactose permease of *Escherichia coli*. *Science*. 301:610–615.
- Adamian, L., and J. Liang. 2002. Interhelical hydrogen bonds and spatial motifs in membrane proteins: Polar clamps and serine zippers. *Proteins: Structure, Function, and Bioinformatics*. 47:209–218.
- Amunts, A., A. Brown, X.C. Bai, J.L. Llacer, T. Hussain, P. Emsley, F. Long, G. Murshudov, S.H. Scheres, and V. Ramakrishnan. 2014. Structure of the yeast mitochondrial large ribosomal subunit. *Science*. 343:1485–1489.
- Anger, A.M., J.P. Armache, O. Berninghausen, M. Habeck, M. Subklewe, D.N. Wilson, and R. Beckmann. 2013. Structures of the human and *Drosophila* 80S ribosome. *Nature*. 497:80–85.
- Bai, X.-c., I.S. Fernandez, G. McMullan, and S.H.W. Scheres. 2013. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife Sciences*. 2.
- Barreto, F.S., and R.S. Burton. 2013. Evidence for compensatory evolution of ribosomal proteins in response to rapid divergence of mitochondrial rRNA. *Molecular biology and evolution*. 30:310–314.
- Baxter, W.T., R.A. Grassucci, H. Gao, and J. Frank. 2009. Determination of signal-to-noise ratios and spectral SNRs in cryo-EM low-dose imaging of molecules. *Journal of Structural Biology*. 166:126–132.
- Beck, K., G. Eisner, D. Trescher, R.E. Dalbey, J. Brunner, and M. Müller. 2001. YidC, an assembly site for polytopic *Escherichia coli* membrane proteins located in immediate proximity to the SecYE translocon and lipids. *EMBO Rep*. 2:709–714.
- Becker, T., S. Bhushan, A. Jarasch, J.-P. Armache, S. Funes, F. Jossinet, J. Gumbart, T. Mielke, O. Berninghausen, K. Schulten, E. Westhof, R. Gilmore, E.C. Mandon, and R. Beckmann. 2009. Structure of Monomeric Yeast and Mammalian Sec61 Complexes Interacting with the Translating Ribosome. *Science*. 326:1369–1373.
- Bernsel, A., H. Viklund, A. Hennerdal, and A. Elofsson. 2009. TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res*. 37:W465–468.
- Bischoff, L., O. Berninghausen, and R. Beckmann. 2014a. Molecular basis for the ribosome functioning as an L-tryptophan sensor. *Cell Rep*. 9:469–475.
- Bischoff, L., S. Wickles, O. Berninghausen, E.O. van der Sluis, and R. Beckmann. 2014b. Visualization of a polytopic membrane protein during SecY-mediated membrane insertion. *Nature Communications*. 5.
- Bondar, A.-N., and S.H. White. 2012. Hydrogen bond dynamics in membrane protein function. *Biochimica et Biophysica Acta (BBA) – Biomembranes*. 1818:942–950.
- Bonnefoy, N., F. Chalvet, P. Hamel, P.P. Slonimski, and G. Dujardin. 1994. OXA1, a *Saccharomyces cerevisiae* Nuclear Gene whose Sequence is Conserved from Prokaryotes to Eukaryotes Controls Cytochrome Oxidase Biogenesis. *Journal of Molecular Biology*. 239:201–212.
- Bostina, M., B. Mohsin, W. Kühlbrandt, and I. Collinson. 2005. Atomic Model of the *E. coli* Membrane-bound Protein Translocation Complex SecYEG. *Journal of Molecular Biology*. 352:1035–1043.

-
- Brown, A., A. Amunts, X.C. Bai, Y. Sugimoto, P.C. Edwards, G. Murshudov, S.H. Scheres, and V. Ramakrishnan. 2014. Structure of the large ribosomal subunit from human mitochondria. *Science*. 346:718–722.
- Chen. 2007. SIGNATURE: a single-particle selection system for molecular electron microscopy. *J Struct Biol*. 157.
- Chen, M.Y., J.C. Samuelson, F.L. Jiang, M. Muller, A. Kuhn, and R.E. Dalbey. 2002. Direct interaction of YidC with the Sec-independent Pf3 coat protein during its membrane protein insertion. *Journal of Biological Chemistry*. 277:7670–7675.
- Chen, Y., R. Soman, S.K. Shanmugam, A. Kuhn, and R.E. Dalbey. 2014. The Role of the Strictly Conserved Positively Charged Residue Differs among the Gram-positive, Gram-negative, and Chloroplast YidC Homologs. *J Biol Chem*. 289:35656–35667.
- Clemons Jr, W.M., J.-F. Ménétrét, C.W. Akey, and T.A. Rapoport. 2004. Structural insight into the protein translocation channel. *Current Opinion in Structural Biology*. 14:390–396.
- Cole, C., J.D. Barber, and G.J. Barton. 2008. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res*. 36:W197–201.
- Cross, B.C.S., I. Sinning, J. Lührink, and S. High. 2009. Delivering proteins for export from the cytosol. *Nat Rev Mol Cell Biol*. 10:255–264.
- Cymer, F., G. von Heijne, and S.H. White. 2014. Mechanisms of Integral Membrane Protein Insertion and Folding. *J Mol Biol*.
- Dalbey, R.E., A. Kuhn, L. Zhu, and D. Kiefer. 2014. The membrane insertase YidC. *Biochimica et Biophysica Acta (BBA) – Molecular Cell Research*. 1843:1489–1496.
- Dalbey, R.E., P. Wang, and A. Kuhn. 2011. Assembly of Bacterial Inner Membrane Proteins. In *Annual Review of Biochemistry*, Vol. 80. Vol. 80. R.D. Kornberg, C.R.H. Raetz, J.E. Rothman, and J.W. Thorner, editors. 161–187.
- de Juan, D., F. Pazos, and A. Valencia. 2013. Emerging methods in protein co-evolution. *Nat Rev Genet*. 14:249–261.
- Driessen, A.J.M., and N. Nouwen. 2008. Protein Translocation Across the Bacterial Cytoplasmic Membrane. *Annual Review of Biochemistry*. 77:643–667.
- Duong, F., and W. Wickner. 1998. Sec-dependent membrane protein biogenesis: SecYEG, preprotein hydrophobicity and translocation kinetics control the stop-transfer function. *EMBO J*. 17:696–705.
- Ehrlich, P.R., and P.H. Raven. 1969. Differentiation of populations. *Science*. 165:1228–1232.
- Elmlund, D., and H. Elmlund. 2012. SIMPLE: Software for ab initio reconstruction of heterogeneous single-particles. *J Struct Biol*. 180:420–427.
- Eswar, N., D. Eramian, B. Webb, M.Y. Shen, and A. Sali. 2008. Protein structure modeling with MODELLER. *Methods Mol Biol*. 426:145–159.
- Facey, S.J., S.A. Neugebauer, S. Krauss, and A. Kuhn. 2007. The Mechanosensitive Channel Protein MscL Is Targeted by the SRP to The Novel YidC Membrane Insertion Pathway of Escherichia coli. *Journal of Molecular Biology*. 365:995–1004.
- Frank, J., M. Radermacher, P. Penczek, J. Zhu, Y. Li, M. Ladjadj, and A. Leith. 1996. SPIDER and WEB: Processing and Visualization of Images in 3D Electron Microscopy and Related Fields. *Journal of Structural Biology*. 116:190–199.
- Frauenfeld, J., J. Gumbart, E.O.v.d. Sluis, S. Funes, M. Gartmann, B. Beatrix, T. Mielke, O. Berninghausen, T. Becker, K. Schulten, and R. Beckmann. 2011. Cryo-EM structure of the ribosome–SecYE complex in the membrane environment. *Nature Structural & Molecular Biology*. 18:614–621.
-

-
- Freites, J.A., D.J. Tobias, G.v. Heijne, and S.H. White. 2005. Interface connections of a transmembrane voltage sensor. *Proceedings of the National Academy of Sciences of the United States of America*. 102:15059–15064.
- Gershoni, M., A. Fuchs, N. Shani, Y. Fridman, M. Corral-Debrinski, A. Aharoni, D. Frishman, and D. Mishmar. 2010. Coevolution predicts direct interactions between mtDNA-encoded and nDNA-encoded subunits of oxidative phosphorylation complex i. *J Mol Biol*. 404:158–171.
- Glick, B.S., and G.V. Heijne. 1996. *Saccharomyces cerevisiae* mitochondria lack a bacterial-type Sec machinery. *Protein Science*. 5:2651–2652.
- Gogala, M., T. Becker, B. Beatrix, J.-P. Armache, C. Barrio-Garcia, O. Berninghausen, and R. Beckmann. 2014. Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion. *Nature*. 506:107–110.
- Gray, A.N., J.M. Henderson-Frost, D. Boyd, S. Shirafi, H. Niki, and M.B. Goldberg. 2011. Unbalanced Charge Distribution as a Determinant for Dependence of a Subset of *Escherichia coli* Membrane Proteins on the Membrane Insertase YidC. *mBio*. 2:e00238–00211.
- Greber, B.J., D. Boehringer, M. Leibundgut, P. Bieri, A. Leitner, N. Schmitz, R. Aebersold, and N. Ban. 2014a. The complete structure of the large subunit of the mammalian mitochondrial ribosome. *Nature*. 515:283–286.
- Greber, B.J., D. Boehringer, A. Leitner, P. Bieri, F. Voigts-Hoffmann, J.P. Erzberger, M. Leibundgut, R. Aebersold, and N. Ban. 2014b. Architecture of the large subunit of the mammalian mitochondrial ribosome. *Nature*. 505:515–519.
- Gumbart, J., Y. Wang, A. Aksimentiev, E. Tajkhorshid, and K. Schulten. 2005. Molecular dynamics simulations of proteins in lipid bilayers. *Current Opinion in Structural Biology*. 15:423–431.
- Gurezka, R., R. Laage, B. Brosig, and D. Langosch. 1999. A Heptad Motif of Leucine Residues Found in Membrane Proteins Can Drive Self-assembly of Artificial Transmembrane Segments. *Journal of Biological Chemistry*. 274:9265–9270.
- Harrington, S.E., and N. Ben-Tal. 2009. Structural Determinants of Transmembrane Helical Proteins. *Structure*. 17:1092–1103.
- Hashem, Y., A. des Georges, V. Dhote, R. Langlois, H.Y. Liao, R.A. Grassucci, C.U. Hellen, T.V. Pestova, and J. Frank. 2013. Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. *Cell*. 153:1108–1119.
- Henderson, R., S. Chen, J.Z. Chen, N. Grigorieff, L.A. Passmore, L. Ciccarelli, J.L. Rubinstein, R.A. Crowther, P.L. Stewart, and P.B. Rosenthal. 2011. Tilt-pair analysis of images from a range of different specimens in single-particle electron cryomicroscopy. *J Mol Biol*. 413:1028–1046.
- Herskovits, A.A., E.S. Bochkareva, and E. Bibi. 2000. New prospects in studying the bacterial signal recognition particle pathway. *Molecular Microbiology*. 38:927–939.
- Hessa, T., H. Kim, K. Bihlmaier, C. Lundin, J. Boekel, H. Andersson, I. Nilsson, S.H. White, and G. von Heijne. 2005. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature*. 433:377–381.
- Hildebrand, A., M. Remmert, A. Biegert, and J. Soding. 2009. Fast and accurate automatic structure prediction with HHpred. *Proteins*. 77 Suppl 9:128–132.
- Hildebrand, P.W., S. Günther, A. Goede, L. Forrest, C. Frömmel, and R. Preissner. 2008. Hydrogen-Bonding and Packing Features of Membrane Proteins: Functional Implications. *Biophysical Journal*. 94:1945–1953.
- Hohn, M., G. Tang, G. Goodyear, P.R. Baldwin, Z. Huang, P.A. Penczek, C. Yang, R.M. Glaeser, P.D. Adams, and S.J. Ludtke. 2007. SPARX, a new environment for Cryo-EM image processing. *Journal of Structural Biology*. 157:47–55.
-

-
- Honig, B.H., and W.L. Hubbell. 1984. Stability of "salt bridges" in membrane proteins. *Proc Natl Acad Sci U S A*. 81:5412-5416.
- Hopf, T.A., L.J. Colwell, R. Sheridan, B. Rost, C. Sander, and D.S. Marks. 2012. Three-Dimensional Structures of Membrane Proteins from Genomic Sequencing. *Cell*. 149:1607-1621.
- Hopf, T.A., C.P.I. Schärfe, J.P.G.L.M. Rodrigues, A.G. Green, O. Kohlbacher, C. Sander, A.M.J.J. Bonvin, and D.S. Marks. 2014. Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife Sciences*. 3.
- Hristova, K., and W.C. Wimley. 2011. A look at arginine in membranes. *J Membr Biol*. 239:49-56.
- Jia, L.X., M. Dienhart, M. Schramp, M. McCauley, K. Hell, and R.A. Stuart. 2003. Yeast Oxa1 interacts with mitochondrial ribosomes: the importance of the C-terminal region of Oxa1. *Embo Journal*. 22:6438-6447.
- Jiang, F., M. Chen, L. Yi, J.-W. de Gier, A. Kuhn, and R.E. Dalbey. 2003. Defining the Regions of Escherichia coli YidC That Contribute to Activity. *Journal of Biological Chemistry*. 278:48965-48972.
- Joh, N.H., A. Min, S. Faham, J.P. Whitelegge, D. Yang, V.L. Woods, and J.U. Bowie. 2008. Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature*. 453:1266-1270.
- Johnson, R.M., K. Hecht, and C.M. Deber. 2007. Aromatic and cation- π interactions enhance helix-helix association in a membrane environment. *Biochemistry*. 46:9208-9214.
- Kamisetty, H., S. Ovchinnikov, and D. Baker. 2013. Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. *Proceedings of the National Academy of Sciences*. 110:15674-15679.
- Kedrov, A., M. Sustarsic, J. de Keyser, J.J. Caumanns, Z.C. Wu, and A.J.M. Driessen. 2013. Elucidating the Native Architecture of the YidC: Ribosome Complex. *Journal of Molecular Biology*.
- Khalili-Araghi, F., J. Gumbart, P.C. Wen, M. Sotomayor, E. Tajkhorshid, and K. Schulten. 2009. Molecular dynamics simulations of membrane channels and transporters. *Curr Opin Struct Biol*. 19:128-137.
- Kiefer, D., and A. Kuhn. 2007. YidC as an essential and multifunctional component in membrane protein assembly. *In International Review of Cytology - a Survey of Cell Biology*, Vol 259. Vol. 259. K.W. Jeon, editor. 113-+.
- Klauda, J.B., R.M. Venable, J.A. Freites, J.W. O'Connor, D.J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A.D. MacKerell, Jr., and R.W. Pastor. 2010. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *The journal of physical chemistry. B*. 114:7830-7843.
- Klenner, C., and A. Kuhn. 2012. Dynamic Disulfide Scanning of the Membrane-inserting Pf3 Coat Protein Reveals Multiple YidC Substrate Contacts. *Journal of Biological Chemistry*. 287:3769-3776.
- Klenner, C., J. Yuan, R.E. Dalbey, and A. Kuhn. 2008. The Pf3 coat protein contacts TM1 and TM3 of YidC during membrane biogenesis. *Febs Letters*. 582:3967-3972.
- Kohler, R., D. Boehringer, B. Greber, R. Bingel-Erlenmeyer, I. Collinson, C. Schaffitzel, and N. Ban. 2009. YidC and Oxa1 Form Dimeric Insertion Pores on the Translating Ribosome. *Molecular Cell*. 34:344-353.
-

-
- Kramer, G., D. Boehringer, N. Ban, and B. Bukau. 2009. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nature Structural & Molecular Biology*. 16:589–597.
- Kucukelbir, A., F.J. Sigworth, and H.D. Tagare. 2014. Quantifying the local resolution of cryo-EM density maps. *Nat Methods*. 11:63–65.
- Kuhlbrandt, W. 2014. Biochemistry. The resolution revolution. *Science*. 343:1443–1444.
- Kuhn, A., R. Stuart, R. Henry, and R.E. Dalbey. 2003. The Alb3/Oxa1/YidC protein family: membrane-localized chaperones facilitating membrane protein insertion? *Trends in Cell Biology*. 13:510–516.
- Kumazaki, K., S. Chiba, M. Takemoto, A. Furukawa, K.-i. Nishiyama, Y. Sugano, T. Mori, N. Dohmae, K. Hirata, Y. Nakada-Nakura, A.D. Maturana, Y. Tanaka, H. Mori, Y. Sugita, F. Arisaka, K. Ito, R. Ishitani, T. Tsukazaki, and O. Nureki. 2014. Structural basis of Sec-independent membrane protein insertion by YidC. *Nature*. advance online publication.
- Kuo, D., K. Licon, S. Bandyopadhyay, R. Chuang, C. Luo, J. Catalana, T. Ravasi, K. Tan, and T. Ideker. 2010. Coevolution within a transcriptional network by compensatory trans and cis mutations. *Genome research*. 20:1672–1678.
- Lai, J.-S., C.-W. Cheng, A. Lo, T.-Y. Sung, and W.-L. Hsu. 2013. Lipid exposure prediction enhances the inference of rotational angles of transmembrane helices. *BMC Bioinformatics*. 14.
- Langlois, R., and J. Frank. 2011. A clarification of the terms used in comparing semi-automated particle selection algorithms in cryo-EM. *J Struct Biol*. 175:348–352.
- Leidig, C., G. Bange, J. Kopp, S. Amlacher, A. Aravind, S. Wickles, G. Witte, E. Hurt, R. Beckmann, and I. Sinning. 2013. Structural characterization of a eukaryotic chaperone--the ribosome-associated complex. *Nat Struct Mol Biol*. 20:23–28.
- Lemmon, M.A., J.M. Flanagan, H.R. Treutlein, J. Zhang, and D.M. Engelman. 1992. Sequence specificity in the dimerization of transmembrane alpha-helices. *Biochemistry*. 31:12719–12725.
- Li, L., I. Vorobyov, and T.W. Allen. 2013. The Different Interactions of Lysine and Arginine Side Chains with Lipid Membranes. *J. Phys. Chem. B*. 117:11906–11920.
- Lindahl, E., and M.S. Sansom. 2008. Membrane proteins: molecular dynamics simulations. *Curr Opin Struct Biol*. 18:425–431.
- Lotz, M., W. Haase, W. Kuhlbrandt, and I. Collinson. 2008. Projection Structure of yidC: A Conserved Mediator of Membrane Protein Assembly. *Journal of Molecular Biology*. 375:901–907.
- Luirink, J., T. Samuelsson, and J.-W. de Gier. 2001. YidC/Oxa1p/Alb3: evolutionarily conserved mediators of membrane protein assembly. *FEBS Letters*. 501:1–5.
- Marks, D.S., L.J. Colwell, R. Sheridan, T.A. Hopf, A. Pagnani, R. Zecchina, and C. Sander. 2011. Protein 3D Structure Computed from Evolutionary Sequence Variation. *PLoS ONE*. 6.
- Marks, D.S., T.A. Hopf, and C. Sander. 2012. Protein structure prediction from sequence variation. *Nature Biotechnology*. 30:1072–1080.
- Meindl-Beinker, N.M., C. Lundin, I. Nilsson, S.H. White, and G. von Heijne. 2006. Asn- and Asp-mediated interactions between transmembrane helices during translocon-mediated membrane protein assembly. *EMBO Rep*. 7:1111–1116.
- Nagamori, S., I.N. Smirnova, and H.R. Kaback. 2004. Role of YidC in folding of polytopic membrane proteins. *Journal of Cell Biology*. 165:53–62.
- Nicholson, W.V., and R.M. Glaeser. 2001. Review: automatic particle detection in electron microscopy. *J Struct Biol*. 133:90–101.
-

-
- Norouzi, R., S. Wickles, C. Leidig, T. Becker, V.J. Schmid, R. Beckmann, and A. Tresch. 2013. Automatic post-picking using MAPPOS improves particle image detection from cryo-EM micrographs. *Journal of Structural Biology*. 182:59–66.
- Nouwen, N., and A.J.M. Driessen. 2002. SecDFyajC forms a heterotetrameric complex with YidC. *Molecular Microbiology*. 44:1397–1405.
- Nugent, T., and D.T. Jones. 2012a. Accurate de novo structure prediction of large transmembrane protein domains using fragment-assembly and correlated mutation analysis. *Proceedings of the National Academy of Sciences*. 109:E1540–E1547.
- Nugent, T., and D.T. Jones. 2012b. Membrane protein structural bioinformatics. *Journal of Structural Biology*. 179:327–337.
- Ochoa, D., and F. Pazos. 2014. Practical aspects of protein co-evolution. *Front. Cell Dev. Biol.* 2.
- Oliver, D.C., and M. Paetzel. 2008. Crystal Structure of the Major Periplasmic Domain of the Bacterial Membrane Protein Assembly Facilitator YidC. *Journal of Biological Chemistry*. 283:5208–5216.
- Ovchinnikov, S., H. Kamisetty, and D. Baker. 2014. Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information. *eLife Sciences*. 3.
- Palczewski, K., T. Kumasaka, T. Hori, C.A. Behnke, H. Motoshima, B.A. Fox, I. Le Trong, D.C. Teller, T. Okada, R.E. Stenkamp, M. Yamamoto, and M. Miyano. 2000. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science*. 289:739–745.
- Park, E., and T.A. Rapoport. 2012. Mechanisms of Sec61/SecY-Mediated Protein Translocation Across Membranes. *Annual Review of Biophysics*. 41:21–40.
- Pei, Z.-Y., G.-L. Mu, J. Pan, and D.-M. Zhang. 2013. Codon usage and coevolution of the large and small subunits of ribulose-1, 5-bisphosphate carboxylase/oxygenase. *Journal of Systematics and Evolution*. 51:511–521.
- Radermacher, M., T. Wagenknecht, A. Verschoor, and J. Frank. 1987. Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. *J Microsc.* 146:113–136.
- Ravaud, S., G. Stjepanovic, K. Wild, and I. Sinning. 2008. The crystal structure of the periplasmic domain of the Escherichia coli membrane protein insertase YidC contains a substrate binding cleft. *Journal of Biological Chemistry*. 283:9350–9358.
- Remmert, M., A. Biegert, A. Hauser, and J. Soding. 2012. HHblits: lightning-fast iterative protein sequence searching by HMM–HMM alignment. *Nat Methods*. 9:173–175.
- Sääf, A., M. Monné, J.-W.d. Gier, and G.v. Heijne. 1998. Membrane Topology of the 60-kDa Oxa1p Homologue from Escherichia coli. *Journal of Biological Chemistry*. 273:30415–30418.
- Sachelaru, I., N.A. Petriman, R. Kudva, P. Kuhn, T. Welte, B. Knapp, F. Drepper, B. Warscheid, and H.-G. Koch. 2013. YidC Occupies the Lateral Gate of the SecYEG Translocon and Is Sequentially Displaced by a Nascent Membrane Protein. *Journal of Biological Chemistry*. 288:16295–16307.
- Sal-Man, N., D. Gerber, I. Bloch, and Y. Shai. 2007. Specificity in Transmembrane Helix–Helix Interactions Mediated by Aromatic Residues. *Journal of Biological Chemistry*. 282:19753–19761.
- Samuelson, J.C., M. Chen, F. Jiang, I. Möller, M. Wiedmann, A. Kuhn, G.J. Phillips, and R.E. Dalbey. 2000. YidC mediates membrane protein insertion in bacteria. *Nature*. 406:637–641.
- Sato, K., H. Mori, M. Yoshida, M. Tagaya, and S. Mizushima. 1997. Short Hydrophobic Segments in the Mature Domain of ProOmpA Determine Its Stepwise Movement during Translocation across the Cytoplasmic Membrane of Escherichia coli. *Journal of Biological Chemistry*. 272:5880–5886.
-

-
- Scheres, S.H. 2012. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol.* 180:519–530.
- Scotti, P.A., M.L. Urbanus, J. Brunner, J.-W.L. de Gier, G. von Heijne, C. van der Does, A.J.M. Driessen, B. Oudega, and J. Luirink. 2000. YidC, the Escherichia coli homologue of mitochondrial Oxa1p, is a component of the Sec translocase. *EMBO J.* 19:542–549.
- Seidelt, B., C.A. Innis, D.N. Wilson, M. Gartmann, J.-P. Armache, E. Villa, L.G. Trabuco, T. Becker, T. Mielke, K. Schulten, T.A. Steitz, and R. Beckmann. 2009. Structural Insight into Nascent Polypeptide Chain-Mediated Translational Stalling. *Science.* 326:1412–1415.
- Seitl, I., S. Wickles, R. Beckmann, A. Kuhn, and D. Kiefer. 2014. The C-terminal regions of YidC from Rhodopirellula baltica and Oceanicaulis alexandrii bind to ribosomes and partially substitute for SRP receptor function in Escherichia coli. *Mol Microbiol.* 91:408–421.
- Senes, A., I. Ubarretxena-Belandia, and D.M. Engelman. 2001. The Calpha ---H...O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. *Proc Natl Acad Sci U S A.* 98:9056–9061.
- Serek, J., G. Bauer-Manz, G. Struhalla, L. van den Berg, D. Kiefer, R. Dalbey, and A. Kuhn. 2004. Escherichia coli YidC is a membrane insertase for Sec-independent proteins. *EMBO J.* 23:294–301.
- Shi, Z., C.A. Olson, and N.R. Kallenbach. 2002. Cation-pi interaction in model alpha-helical peptides. *J Am Chem Soc.* 124:3284–3291.
- Sillitoe, I., A.L. Cuff, B.H. Dessailly, N.L. Dawson, N. Furnham, D. Lee, J.G. Lees, T.E. Lewis, R.A. Studer, R. Rentzsch, C. Yeats, J.M. Thornton, and C.A. Orengo. 2013. New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. *Nucl. Acids Res.* 41:D490–D498.
- Ulmschneider, M.B., M.S.P. Sansom, and A. Di Nola. 2005. Properties of integral membrane protein structures: Derivation of an implicit membrane potential. *Proteins: Structure, Function, and Bioinformatics.* 59:252–265.
- Urbanus, M.L., P.A. Scotti, L. Fröderberg, A. Sääf, J.-W.L.d. Gier, J. Brunner, J.C. Samuelson, R.E. Dalbey, B. Oudega, and J. Luirink. 2001. Sec-dependent membrane protein insertion: sequential interaction of nascent FtsQ with SecY and YidC. *EMBO Rep.* 2:524–529.
- van Bloois, E., H.L. Dekker, L. Fröderberg, E.N.G. Houben, M.L. Urbanus, C.G. de Koster, J.-W. de Gier, and J. Luirink. 2008. Detection of cross-links between FtsH, YidC, HflK/C suggests a linked role for these proteins in quality control upon insertion of bacterial inner membrane proteins. *FEBS Letters.* 582:1419–1424.
- Van den Berg, B., W.M. Clemons, Jr., I. Collinson, Y. Modis, E. Hartmann, S.C. Harrison, and T.A. Rapoport. 2004. X-ray structure of a protein-conducting channel. *Nature.* 427:36–44.
- van der Laan, M., P. Bechtluft, S. Kol, N. Nouwen, and A.J.M. Driessen. 2004. F1F0 ATP synthase subunit c is a substrate of the novel YidC pathway for membrane protein biogenesis. *Journal of Cell Biology.* 165:213–222.
- van der Laan, M., E.N.G. Houben, N. Nouwen, J. Luirink, and A.J.M. Driessen. 2001. Reconstitution of Sec-dependent membrane protein insertion: nascent FtsQ interacts with YidC in a SecYEG-dependent manner. *EMBO Rep.* 2:519–523.
- van der Laan, M., M.L. Urbanus, C.M. ten Hagen-Jongman, N. Nouwen, B. Oudega, N. Harms, A.J.M. Driessen, and J. Luirink. 2003. A conserved function of YidC in the biogenesis of respiratory chain complexes. *Proceedings of the National Academy of Sciences of the United States of America.* 100:5801–5806.
- van Heel, M., and W. Keegstra. 1981. IMAGIC: a fast, flexible and friendly image analysis software system. *Ultramicroscopy.* 7:113–129.
-

-
- von Heijne, G. 1989. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature*. 341:456–458.
- Wagenknecht, T., R. Grassucci, and J. Frank. 1988. Electron microscopy and computer image averaging of ice-embedded large ribosomal subunits from *Escherichia coli*. *J Mol Biol*. 199:137–147.
- Wagner, S., O. Pop, G.-J. Haan, L. Baars, G. Koningstein, M.M. Klepsch, P. Genevaux, J. Luirink, and J.-W. de Gier. 2008. Biogenesis of MalF and the MalFGK(2) maltose transport complex in *Escherichia coli* requires YidC. *Journal of Biological Chemistry*. 283:17881–17890.
- White, S.H., A.S. Ladokhin, S. Jayasinghe, and K. Hristova. 2001. How Membranes Shape Protein Structure. *Journal of Biological Chemistry*. 276:32395–32398.
- Wickles, S., A. Singharoy, J. Andreani, S. Seemayer, L. Bischoff, O. Berninghausen, J. Soeding, K. Schulten, E.O.v.d. Sluis, and R. Beckmann. 2014. A structural model of the active ribosome-bound membrane protein insertase YidC. *eLife Sciences*. 3.
- Wickström, D., S. Wagner, P. Simonsson, O. Pop, L. Baars, A.J. Ytterberg, K.J. van Wijk, J. Luirink, and J.-W.L. de Gier. 2011. Characterization of the Consequences of YidC Depletion on the Inner Membrane Proteome of *E. coli* Using 2D Blue Native/SDS-PAGE. *Journal of Molecular Biology*. 409:124–135.
- Yang, S., H.K. Yalamanchili, X. Li, K.-M. Yao, P.C. Sham, M.Q. Zhang, and J. Wang. 2011. Correlated evolution of transcription factors and their binding sites. *Bioinformatics*. 27:2972–2978.
- Yu, Z., G. Koningstein, A. Pop, and J. Luirink. 2008. The Conserved Third Transmembrane Segment of YidC Contacts Nascent *Escherichia coli* Inner Membrane Proteins. *Journal of Biological Chemistry*. 283:34635–34642.
- Yuan, J., G.J. Phillips, and R.E. Dalbey. 2007. Isolation of Cold-Sensitive yidC Mutants Provides Insights into the Substrate Profile of the YidC Insertase and the Importance of Transmembrane 3 in YidC Function. *Journal of Bacteriology*. 189:8961–8972.
- Zhu, L., H.R. Kaback, and R.E. Dalbey. 2013. YidC Protein, a Molecular Chaperone for LacY Protein Folding via the SecYEG Protein Machinery. *Journal of Biological Chemistry*. 288:28180–28194.
- Zhu, Y., B. Carragher, R.M. Glaeser, D. Fellmann, C. Bajaj, M. Bern, F. Mouche, F. de Haas, R.J. Hall, D.J. Kriegman, S.J. Ludtke, S.P. Mallick, P.A. Penczek, A.M. Roseman, F.J. Sigworth, N. Volkman, and C.S. Potter. 2004. Automatic particle selection: results of a comparative study. *J Struct Biol*. 145:3–14.

7 Acknowledgements

First of all, I would like to thank my Ph.D. supervisor Prof. Dr. Roland Beckmann for giving me the opportunity to work in his lab at the Gene Center, for the unique working conditions, atmosphere and his constant support.

I wish to express my gratitude to a number of people I had the pleasure of working with. Thomas, for the save guidance of my first steps into Fourier space and back; Christoph and Eli for their excellent support from cloning to gel filtration; Otto and Charlotte for making beautiful grids and collecting shiny micrographs; Lukas, Bertrand and Eli for sharing their profound experience and knowledge about membrane protein biogenesis; Julian, Sibylle, Lukas and Sarah for making the IMPRS retreats and seminar unforgettable; Alexej for showing me that it's not all about FSC but FCS; Lukas for brilliant illustrator advise; Martin and Andi for excellent help in data processing and cluster management; Abhi for his amazing motivation throughout our collaboration; Jessica, Stefan and Susann for their brilliant bioinformatics analysis; Ines for the excellent collaboration and sharing constructs; Ramin for developing great software; Joanna Musial, Heidi Sieber and Andrea Gilmozzi for their unbroken effort to keep the lab running; the friends of the tunnel exit; all volunteers of the G&T experiment; the beef team. I deeply appreciate the critical reading of this thesis by Christoph and Alexej and their excellent suggestions for corrections.

I want to thank all past and present members of the Beckmann and Wilson lab for the great and supportive working atmosphere during my time in the lab.

Furthermore, I would like to thank IMPRS-LS for financial support and providing an inspiring environment.

Special thanks go to my parents for their constant support throughout my life.

Finally, my biggest thanks go to my wife Sarah for her patience, trust and encouragement and to Nele and Ida. You are the greatest happiness of my life.

8 Curriculum Vitae

Personal details

Name	Stephan Wickles
Date of birth	1982/07/03
Place of birth	Pegnitz, Germany
Nationality	German

Education

2015/03 – 2010/03	Ph.D. student at Gene Center (LMU), Prof. Dr. Roland Beckmann
2015/03 – 2009/03	Fast Track PhD program at The International Max Planck Research School for Molecular and Cellular Life Sciences (IMPRS-LS)
2009/01	Bachelor of Science (B.Sc.) in Life Science
2008/12 – 2008/09	Research exchange at the University College Dublin (UCD) in the group of Prof. Dr. Oscarson
2009/01 – 2005/10	Studies of Life Science at the Leibniz University of Hannover
2004/07	Abitur at the Berufsoberschule (BOS) Bayreuth

Publications

- Wickles S**, Singharoy A, Andreani J, Seemayer S, Bischoff L, Berninghausen O, Soeding J, Schulten K, van der Sluis EO, Beckmann R: A structural model of the active ribosome-bound membrane protein insertase YidC. *Elife* 2014, 3:e03035.
- Bischoff L, **Wickles S**, Berninghausen O, van der Sluis EO, Beckmann R: Visualization of a polytopic membrane protein during SecY-mediated membrane insertion. *Nat Commun* 2014, 5:4103.
- Seitl I, **Wickles S**, Beckmann R, Kuhn A, Kiefer D: The C-terminal regions of YidC from *Rhodospirillum rubrum* and *Oceanicaulis alexandrii* bind to ribosomes and partially substitute for SRP receptor function in *Escherichia coli*. *Mol. Microbiol.* 2014, 91:408–421.
- Norouzi R, **Wickles S**, Leidig C, Becker T, Schmid VJ, Beckmann R, Tresch A: Automatic post-picking using MAPPOS improves particle image detection from cryo-EM micrographs. *J. Struct. Biol.* 2013, 182:59–66.
- Leidig C, Bange G, Kopp J, Amlacher S, Aravind A, **Wickles S**, Witte G, Hurt E, Beckmann R, Sinning I: Structural characterization of a eukaryotic chaperone--the ribosome-associated complex. *Nat. Struct. Mol. Biol.* 2013, 20:23–28.
- Dönhöfer A, Franckenberg S, **Wickles S**, Berninghausen O, Beckmann R, Wilson DN: Structural basis for TetM-mediated tetracycline resistance. *Proc. Natl. Acad. Sci. U.S.A.* 2012, 109:16900–16905.
- Becker T, Franckenberg S, **Wickles S**, Shoemaker CJ, Anger AM, Armache J-P, Sieber H, Ungewickell C, Berninghausen O, Daberkow I, et al.: Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. *Nature* 2012, 482:501–506.