Biophysical Journal
Article

Biophysical Society

Structure and Dynamics of the Central Lipid Pool and Proteins of the Bacterial Holo-Translocon

Remy Martin,[1] Andreas Haahr Larsen,[2] Robin Adam Corey,[1] Søren Roi Midtgaard,[2] Henrich Frielinghaus,[3] Christiane Schaffitzel,[1] Lise Arleth,[2,*] and Ian Collinson[1,*]
[1]School of Biochemistry, University of Bristol, Bristol, United Kingdom; [2]Niels Bohr Institute, University of Copenhagen, Universitetsparken 5, Copenhagen, Denmark; and [3]Jülich Centre for Neutron Science at Heinz Maier-Leibnitz Zentrum, Garching, Germany

ABSTRACT The bacterial Sec translocon, SecYEG, associates with accessory proteins YidC and the SecDF-YajC subcomplex to form the bacterial holo-translocon (HTL). The HTL is a dynamic and flexible protein transport machine capable of coordinating protein secretion across the membrane and efficient lateral insertion of nascent membrane proteins. It has been hypothesized that a central lipid core facilitates the controlled passage of membrane proteins into the bilayer, ensuring the efficient formation of their native state. By performing small-angle neutron scattering on protein solubilized in "match-out" deuterated detergent, we have been able to interrogate a "naked" HTL complex, with the scattering contribution of the surrounding detergent micelle rendered invisible. Such an approach has allowed the confirmation of a lipid core within the HTL, which accommodates between 8 and 29 lipids. Coarse-grained molecular dynamics simulations of the HTL also demonstrate a dynamic, central pool of lipids. An opening at this lipid-rich region between YidC and the SecY lateral gate may provide an exit gateway for newly synthesized, correctly oriented, membrane protein helices, or even small bundles of helices, to emerge from the HTL.

INTRODUCTION

The general process of protein secretion and membrane protein insertion is achieved by the conserved secretory, or Sec, machinery at the plasma membrane of bacteria and archaea, and the endoplasmic reticulum of eukaryotes. The protein-conducting channel is formed by a core heterotrimeric assembly—the SecY complex of bacteria and archaea, and the Sec61 complex of eukaryotes (1,2)—through which secretory and membrane proteins are driven, respectively, across and into the membrane. This process occurs either during protein synthesis, involving the direct binding of co-translating ribosomes to the Sec complex, or post-translationally, powered by associated energy-transducing factors, such as the ATPase SecA in bacteria (3,4).

Additional components combine with the core complex to facilitate the lateral passage of transmembrane α-helices into the bilayer or for the implementation of specific modifications, like glycosylation in eukaryotes. Indeed, the structure of the eukaryotic holo-complex engaged with the ribosome illustrates how the core complex and accessory factors could streamline the efficient translocation and glycosylation of proteins at the endoplasmic reticular membrane (5).

The bacterial core translocon SecYEG associates with the ancillary subcomplex SecDF-YajC (6) and YidC (7) to form a 7 protein supercomplex, also known as the holo-translocon (HTL) (8). Generally, secretion through the translocon occurs post-translationally, whereas membrane protein insertion is co-translational (9). The HTL ensures efficient translocation, folding, and the assembly of secretory and membrane proteins and can be produced in sufficient quantities for structural and functional analyses (8,10,11). Its availability enabled a preliminary structural analysis combining electron cryomicroscopy (cryo-EM) and small-angle neutron scattering (SANS) (12) (Fig. 1). Interestingly, the proteins are arranged around a central cavity, most likely constituted of lipids, which we proposed to form a protected environment for the co-translational insertion of transmembrane α-helical bundles. The encapsulation of nascent unfolded membrane proteins would prevent catastrophic proteolysis or aggregation and thus promote efficient protein

Submitted December 10, 2018, and accepted for publication April 2, 2019.
*Correspondence: arleth@nbi.ku.dk or ian.collinson@bristol.ac.uk
Remy Martin and Andreas Haahr Larsen contributed equally to this work.
Robin Adam Corey's present address is Department of Biochemistry, University of Oxford, OX1 3QU, United Kingdom
Editor: Michael Sattler.
https://doi.org/10.1016/j.bpj.2019.04.002

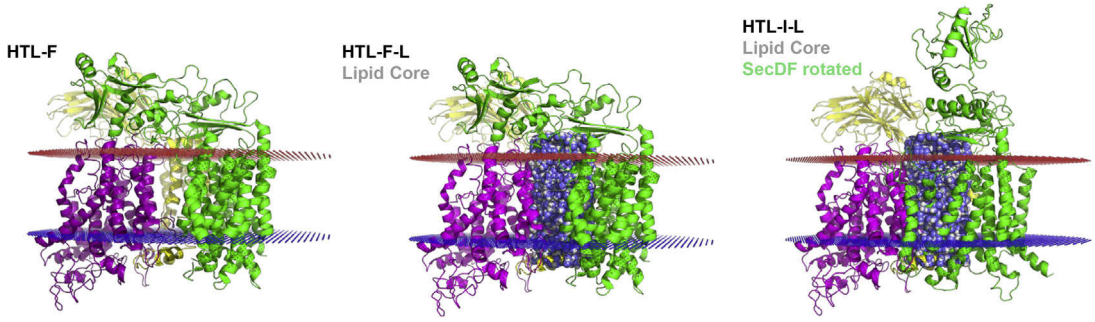

FIGURE 1 Atomistic models of HTL with and without a lipidic core positioned "by hand." Three HTL structures are used to fit the experimental SANS data. HTL-F is the starting structure (HTL with SecDF in the F-form) and is based on the electron microscopy-fitted structure from Botte et al. (12). SecYEG is shown in magenta, SecDF in green and YidC in yellow. HTL-F-L is the same protein arrangement with the addition of a lipid core. HTL-I-L is the same structure as HTL-F-L (i.e., containing lipids) but has had the SecDF P1 domain rotated (HTL with SecDF in the I-form). Lipid bilayer planes are marked in red (periplasmic side) and blue (cytoplasmic side). To see this figure in color, go online.

folding, much in the same way that GroEL facilitates the folding of globular proteins within a secluded hydrophilic chamber (13).

High-resolution structures of the individual components of the HTL are known (14–16), and they could be fitted into the low-resolution cryo-EM structure to create a preliminary atomic model of the HTL, supported also by biochemical data (12). In this model, the lateral gate of SecY, through which nascent transmembrane helices enter the membrane (14), faces the central lipid cavity. YidC is located on the opposite side of the cavity, with its putative binding site for inserting transmembrane helices (15) also facing the lipid pool. The juxtaposition of these regions at the proposed central lipid core of the HTL provides a compelling case for their concerted action in membrane protein insertion.

To explore further the structure and dynamics of the central lipid pool, we conducted an analysis of the HTL, combining SANS and Martini coarse-grained (CG) molecular dynamics (MD) simulations. For the SANS experiments, the HTL was solubilized in match-out deuterated detergent n-dodecyl-β-D-maltoside (d-DDM) previously developed in the Arleth lab (17). This d-DDM was deuterated separately in the head and tail group to fully match out the neutron contrast of the detergent in a 100% D_2O-based buffer. In this way, the detergents become invisible in the SANS experiment, thus allowing the distinction and description of the lipid component of the translocon. The Martini CG MD simulations support the notion of a stable and persistent lipid-filled cavity within the center of the HTL. Beyond this, we discuss the role of such a lipid pool in the insertion and folding of membrane proteins via the Sec machinery.

MATERIALS AND METHODS

HTL preparation and d-DDM exchange

HTL was purified as described previously (8). Purified HTL in hydrogenated n-dodecyl-β-D-maltoside (DDM) was exchanged into a 100% D_2O buffer containing deuterated DDM. Detergent exchange was performed on a Superose 6 (10/300) column equilibrated in a simple TS buffer (20 mM Tris (pH 7.5) and 100 mM $NaCl_2$), made with 100% D_2O, and 0.02% deuterated DDM.

SANS data collection for deuterated detergent

Samples were prepared and measured in 2 mm quartz cuvettes (Hellma, Müllheim, Germany) and temperature controlled at 10°C. SANS data were collected on KWS-1 at Forschungsreaktor München II at Heinz Maier-Leibnitz Zentrum (Garching, Germany) at a wavelength of $\lambda = 5$ Å and a wavelength spread of $\Delta\lambda/\lambda = 10\%$ (full width at half maximum). Sample detector/collimation distances of 1.5 m/4 m and 8 m/8 m were used to obtain a q range of 0.006–0.44 $Å^{-1}$, with a good overlap between the settings. The wave vector, q, has the usual magnitude of $4\pi\sin(\theta)/\lambda$, where 2θ is the scattering angle. Transmissions were measured at a 4 m/4 m setting with 3 min exposure time. Data were calibrated using plexiglass as a calibrant to yield the absolute scaled scattering intensity, $I(q)$, in units of cm^{-1}.

Correction and averaging was performed using QtiKWS (version 10; www.qtikws.de), and the buffer measurement was subtracted subsequently. The sample was measured for ~4 h (1.75 h at the 8 m/8 m setting and 2 h at the 1.5 m/4 m setting) to obtain a sufficient signal over the background. 15 min measurement windows were used to monitor the change in scattering over time. No change was observed, meaning that the sample was stable during the measurements.

SANS data analysis

A combination of the home-written software CaPP (v. 3.9; github.com/Niels-Bohr-Institute-XNS-StructBiophys/CaPP) and WillItFit (18) were used to fit the data. A 3 Å thick water layer with 6% higher scattering length than bulk D_2O was added (19) but was excluded from the hydrophobic transmembrane region. The thickness of this was set to 30.6 Å in accordance with the hydrophobic bilayer thickness reported in the orientations of proteins in membranes database (20). Resolution effects were included using the resolution width, $\Delta q(q)$, present in the fourth column of the data files provided by the SANS beamline. CaPP was also used to calculate the theoretical pair distance distribution functions, $p(r)$, for the atomistic structures. Experimental (r) were calculated using BayesApp (21), including a constant background in the fit and truncation of data at $q = 0.3$ $Å^{-1}$. The fit to obtain the $p(r)$ had a χ_r^2 value of 2.7. As the model is generic and thus true for this data set as well, this value was expected to be close to unity. There were 112 data points in the fitted range, and

the degrees of freedom of the model was estimated as $N_g = 8.3$ (22). The probability of obtaining a χ_r^2 of 2.7 given 112 points and $N_g = 8.3$ is only $\sim 10^{-16}$ for a true model. We could therefore conclude that the experimental errors were underestimated. These wrongly estimated errors are problematic when using χ^2 statistics and should ideally be corrected. Unfortunately, wrongly estimated errors are a general problem at many small-angle scattering facilities as discussed (e.g., by Franke et al. (23) and Rambo and Tainer (24)), and there are ongoing efforts to solve it in the general case. As a work around for this specific data set, the errors were renormalized by $\sigma_{new} = \beta\ \sigma_{old}$, where $\beta = \sqrt{\chi_r^2}$, and χ_r^2 is the value obtained from BayesApp, using $N_g = 8.3$.

The forward scattering, as determined by Guinier analysis (Fig. S1), can be used to calculate a model-free estimation of the number of lipids in the lipid core. The protein concentration of the sample was calculated from a measurement of the UV280 absorption of 0.65 cm^{-1} and an extinction coefficient of 234,600 $cm^{-1}\ M^{-1}$ as calculated from the protein sequence, using ExPASy ProtParam (https://web.expasy.org/protparam). The forward scattering from the protein-lipid complex (HTL + lipid core) is given as:

$$I(0) = c|\Delta\rho_{HTL}V_{HTL} + \Delta\rho_{LIP}V_{LIP}|^2,$$

where c is the concentration (number of complexes per cm^3), $\Delta\rho_{HTL}$ and $\Delta\rho_{LIP}$ are the excess scattering length densities (scattering contrasts) of the protein and lipid, respectively, and V_{HTL} and V_{LIP} are the corresponding volumes. The sample was purified with a total *Escherichia coli* lipid extract of known composition (67% phosphatidylethanolamine, 23.2% phosphatidylglycerol, and 9.8% cardiolipin), so $\Delta\rho_{LIP}$ could be estimated. The only unknown was, therefore, V_{LIP}, the volume of the lipid core, which can be expressed as follows:

$$V_{LIP} = \frac{\sqrt{I(0)/c} - |\Delta\rho_{HTL}| \cdot V_{HTL}}{|\Delta\rho_{LIP}|}.$$

V_{LIP} is calculated by the subtraction of two numbers, $\sqrt{I(0)/c}$ and $(|\Delta\rho_{HTL}| \cdot V_{HTL})$, equal in magnitude and each with an associated uncertainty, which result in a relatively large error on the calculated result. The major contributions to the uncertainty stems from the absorption measurement used to estimate the molar concentration. We assumed a 15% uncertainty on the concentration measurement, 10% on the estimation of $I(0)$, and 2% on the estimated volumes of HTL and the lipids. The number of lipids could then be found by dividing V_{LIP} by the mean volume of the *E. coli* lipids (1216 $Å^3$), which was calculated from the lipid composition (https://avantilipids.com/product/100600) using a known volume for the different lipid components (25).

A fit was made using two forms of HTL. The first form was HTL with SecDF in the F-form (Fig. 1), in which the P1 domain of SecDF is close to the rest of the protein (12). The other form was HTL with SecDF in the I-form, with the P1 domain rotated away from the rest of the protein. Both forms contained a cylindrical lipid core in the central cavity of HTL (12), representing a lipid bilayer, and the forms were denoted, respectively, HTL-F-L and HTL-I-L. The fitting algorithm was allowed to mix HTL-F-L and HTL-I-L to obtain the optimal fit (Fig. 2). The intensity of the mix was given as:

$$I(q) = k \cdot I(0) \cdot [A \cdot P_{HTL-F-L}(q) + (1 - A) \cdot P_{HTL-I-L}(q)] + B,$$

where A is the fraction of the sample in the HTL-F-L form, k is a scaling constant of the forward scattering (close to unity), and B is a constant background (close to zero).

The goodness of the fits was evaluated using the reduced χ^2, given as $\chi_r^2 = \chi^2/(N - K)$, where N is the number of data points, and K is the number of fitting parameters. The χ^2 is defined in terms of the measured experimental intensities I_i^{exp} and corresponding uncertainties σ_i and the fitted theoretical intensities I_i^{fit}:

$$\chi^2 = \sum_{i=1}^{N} \frac{\left(I_i^{exp} - I_i^{fit}\right)^2}{\sigma_i^2}.$$

There was a minor contribution of aggregates in the sample as indicated from the Guinier plot (Fig. S1). The presence of a small amount of aggregates was also clear from the "tail" of the $p(r)$ with a large maximal intraparticular distance, D_{max}, of $\sim$200 Å (Fig. 3 *B*). The aggregate contribution was taken into account in the fits (Fig. 3 *A*) by including a fractal structure factor, $S_{frac}(q)$, in the model, as previously described (26). Shortly, a fractal aggregate description was used (27) in combination with the decoupling approximation (28) and the form factor of the complex, $P(q)$:

$$I_{frac}(q) = k \cdot I(0) \cdot [A \cdot P_{HTL-F-L}(q) S_{frac}(q) + (1 - A) \cdot P_{HTL-I-L}(q) S_{frac}(q)] + B,$$

where $S_{frac}(q)$ is the effective form factor after the decoupling approximation was applied. A mean radius of $R = 42.1$ Å was used for HTL,

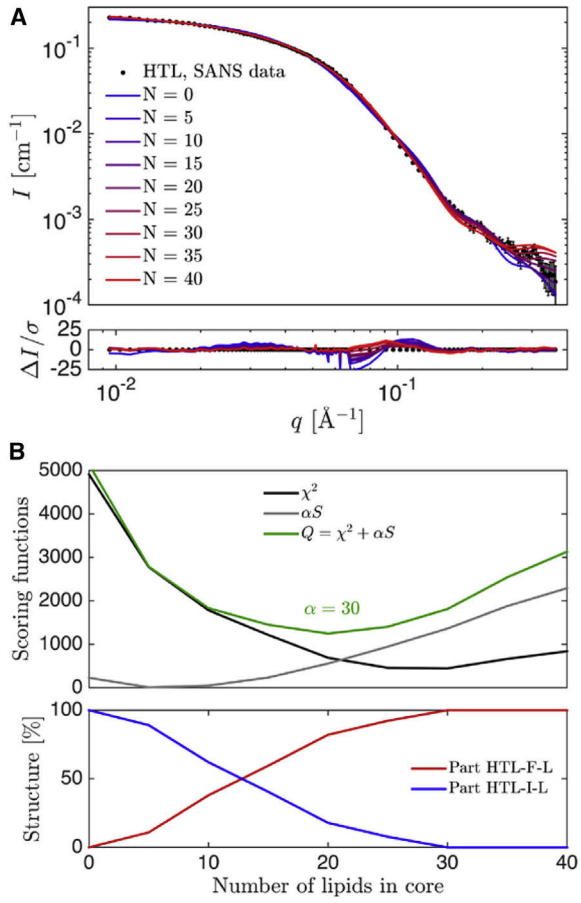

FIGURE 2 Model fitting of theoretical scattering to experimental SANS data of HTL in d-DDM. (*A*) Shown is the theoretical scattering of a linear combination of model HTL-F-L and HTL-I-L in which the number of central lipids, N, is varied from 0 to 40 (*red* and *blue*), plotted against experimental HTL data (*black dots*). (*B*) The upper panel shows χ^2, αS and Q (see text) for the fits, and the lower panel shows the amount of HTL-F-L and HTL-I-L in the fits. To see this figure in color, go online.

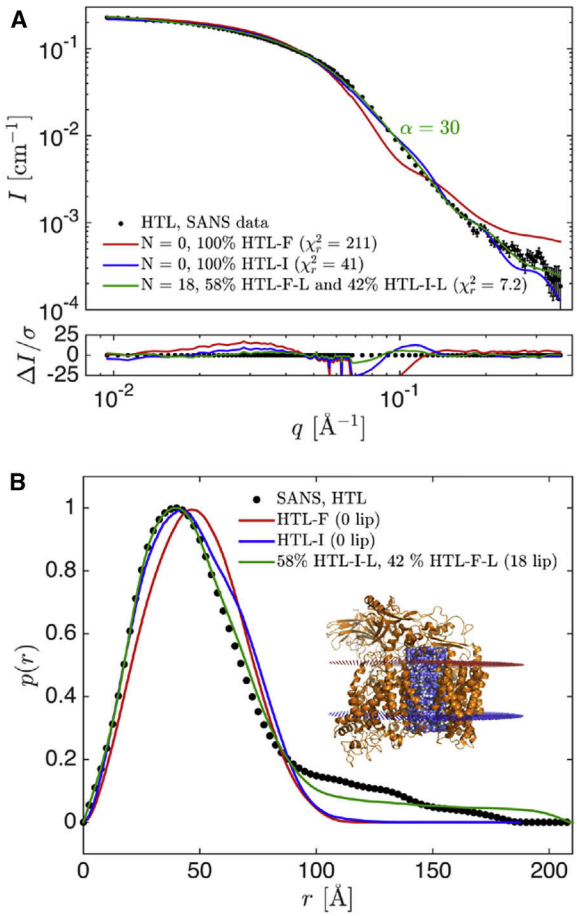

FIGURE 3 Experimental SANS data of HTL in DDM, fits to data, and pair distance distribution functions *p(r)* for data and models. (*A*) Shown is a fit to data with HTL-F (no lipids) in red and HTL-I (no lipids) in blue and the most probable fit in green for $\alpha = 30$ (see text), with 18 lipids in the core, 58% in the HTL-F-L form and 42% in the HTL-I-L form, and with ~1% of the total protein being aggregated. (*B*) Shown is the *p(r)* plot of data and models with the same colors. Inset shows the HTL-F-L in cartoon representation (*orange*) with a lipid core representative of 18 lipids (*blue* and *white*). Lipid bilayer planes are marked in red (periplasmic side) and blue (cytoplasmic side). To see this figure in color, go online.

corresponding to the radius of a sphere with the volume equal to the sum of Van der Waals volumes of the atoms in the protein (29). The models were implemented in WillItFit (18).

Molecular dynamics simulations

CG MD simulations were built according to the MemProtMD protocol (30), using Protein Data Bank (PDB): 5MG3 as an input. Briefly, PDB structure files were converted to a Martini model description using the Martinize script (31) with the secondary structure defined using Define Secondary Structure of Proteins (DSSP) (32). The protein was placed in the center of a ~15 × ~15 × ~15 nm simulation box with 414 palmitoyl-2-oleoyl-phosphatidylethanolamine (POPE), 129 palmitoyl-2-oleoylphosphatidyl-glycerol (POPG), and 16 cardiolipin lipids placed in random orientations around it. The systems were solvated in ~17,500 Martini water and neutralized with ions to 0.15 M.

Simulations were run in the NPT ensemble at 310 K with the V-rescale thermostat (33) and semi-isotropic Parrinello-Rahman pressure coupling (34) for 350 ns with elastic network restraints of 1000 kJ mol^{-1} mn^{-2} between all protein beads within a cut-off distance of 1 nm, at 310 K using 20 fs time steps. These simulations were then extended to 3 μs with elastic networks only applied to beads within 1 nm and on the same protein chain. This was done to allow the central lipid pore to change size without the restriction of inter-subunit elastic networks. Postsimulation snapshots were converted to an atomistic description (35) in the Charmm36 force field (36) for comparison with the SANS data.

Simulations were run on Phase 3 of BlueCrystal, the University of Bristol's High Performance Computer. Images of proteins were made in PyMOL and VMD, and data were plotted with gnuplot or matplotlib.

RESULTS

SANS confirms a central lipid core within the HTL

Previous studies show that the purified HTL complex is composed of its constituent protein subunits and significant proportions of lipid and detergent (12). In the purified complex, the majority of this lipid and/or detergent component is localized at the center of the complex with the protein at the periphery. Because of the relatively close contrast match points of standard hydrogenated DDM (21.7% D_2O) and *E. coli* lipids (13.1% D_2O), it is difficult to distinguish and separate the scattering contributions from the lipid and the solubilizing detergents. Therefore, it was not possible to confidently attribute the central scattering contribution to lipid or detergent in our previous study. To address this, SANS experiments for this study were performed on the HTL using partially deuterated DDM (d-DDM) to mask the scattering signal associated with the detergent. The DDM sugar headgroup and tail moieties were chemically deuterated independently, to different degrees, such that the scattering length densities of both the head and tail group of the d-DDM are equivalent to that of 100% D_2O (17).

Purified HTL was detergent exchanged into d-DDM buffer by gel filtration chromatography (see Materials and Methods), wherein the d-DDM buffer was made up with 100% D_2O. So, the recorded SANS measurements were conducted at the d-DDM contrast match point and with minimal incoherent scattering background from the buffer. Thus, only scattering contributions of the protein and lipid components were measured and the detergent rendered effectively invisible.

Guinier analysis of the collected data indicates a radius of gyration (R_g) for the HTL in the absence of the DDM scattering contribution as 41.1 ± 0.3 Å (Fig. S1), slightly higher than the calculated theoretical R_g of 37.0 Å. The forward scattering, $I(0)$, determined by Guinier analysis (37), can be used to calculate a model-free estimation of the lipid volume of the HTL (see Materials and Methods). From an $I(0)$ value of 0.23 ± 0.2 cm^{-1} and a protein concentration of 0.7 ± 0.1 mg/mL as well as the calculated scattering length densities of protein and lipids, respectively, the volume of the lipid pool (V_{LIP}) can be estimated to be 12000 ± 17000 Å^3. Assuming a mean volume of an *E. coli* lipid as 1216 ± 24 Å^3 (calculated from (25)), the model-free estimation indicates the presence of a lipid pool

consisting of 10 lipids, supporting a significant lipid-based scattering contribution in the core of the complex (Fig. 1).

However, because of the nature of the determination of $I(0)$ and the cumulative uncertainties involved in calculating $V_{LIP}V_{LIP}$ (see Material and Methods), the estimated error on the result was ± 14 as obtained by error propagation. Moreover, the small amount of aggregation of the sample, as visible from the Guinier analysis (Fig. S1) and the "tail" of the pair distance distribution function (Fig. 3 *B*), would result in an overestimated value for the forward scattering and, thus, of the estimated number of lipids.

Therefore, we sought an alternative to the calculations described above, with a new data-fitting approach. A more precise estimation of the lipid content could be obtained by model-based analysis, in which the full q-range was investigated. Moreover, this alternative approach would enable the unequivocal modeling and exclusion of the contribution of the minor aggregated component (Fig. S2).

Protein-lipid complex model building for SANS data fitting

A model lipid core of this volume was created, with the height corresponding to a typical lipid bilayer (50 Å). A simple cylindrical shape was assumed for the lipid core (Fig. 1), which spanned the height of the transmembrane part of HTL. The lipid core consisted of beads homogeneously filling the cylindrical volume with an average scattering length density corresponding to the average scattering length density of the lipid extract used in the sample purification and preparation. The cylindrical volume was filled with beads, each representing dummy CH_2 groups. The beads were placed in the cylindrical volume by a Monte Carlo approach and the resulting lipid core was positioned in the central cavity of the preliminary electron microscopy-fitted HTL structure (PDB: 5MG3 (12)). The structure without lipids was termed HTL-F (HTL with SecDF in the F-form) and the structure with lipids HTL-F-L (Fig. 1).

To assess the effects of known domain flexibility, the model structure with lipids was also modified to reflect the distinct known conformations of the periplasmic domain of SecD (P1): the F- and I-forms, with an approximate 100° rotation of P1 between the two structures (16,38). The model was created by taking the lipid containing the HTL-F-L structure and replacing SecDF in the F-form with SecDF in the I-form (PDB: 5XAM) and termed HTL-I-L (Fig. 1). The models with and without lipids, and with the alternate positions of SecD-P1, were utilized in the subsequent data fitting to the scattering data below.

Determination of the number of centrally bound lipids and structural flexibility of the HTL

Theoretical scattering was calculated for a series of HTL-F-L and HTL-I-L structures containing a varying number of lipid molecules in the core (0–40) to find the most probable model (Fig. 2 *A*). The model fit to the experimental data improved as the number of lipids increases up to a count of 29 lipids, as assessed by the calculated χ_r^2 values (Materials and Methods), and worsened at numbers above this value (Fig. 2 *B*). However, the data were fitted on an absolute scale with a scaling parameter, k, for the forward scattering. Optimally, k should be unity, and deviation from unity indicates a less probable model. But some deviation from unity was expected because of the uncertainty of the estimation of $I(0)$ (see Materials and Methods). We, therefore, introduced a penalty function $S(k)$:

$$S(k) = \frac{(k-1)^2}{\sigma_k^2},$$

that increases as k differs from unity. σ_k^2 is the estimated uncertainty of k, which was set to 0.05. The most probable solution could then be found where the function $Q = \chi^2 + \alpha S$ was minimized (39), where α determines the weight given to χ^2 and S, respectively. Q, χ^2, and αS are plotted on Fig. 2 *B* for $\alpha = 30$. With this regularized expression, a lower number of eight bound lipids was estimated because both S and χ^2 increase as N decreases below 8. Similarly, an upper limit of 29 lipids was determined because both scoring functions increase as N rises above this value.

The optimal value depends on the choice of α, which is not trivial to determine and which we will not go into detail with here. Details can be found in Larsen et al. (39). In Fig. 3 *A*, the best fit for $\alpha = 30$ is given, with $N = 18$, in the middle of the "allowed" range. Of the whole range (between 8 and 29 lipids), the best fit is a linear combination of the HTL-F-L and the HTL-I-L structures (e.g., for $N = 18$, the model that best fits the data has 58% of HTL-F-L and 42% of HTL-I-L). As lipid numbers increase, the calculated proportion of HTL-I-L increases as well. The number of lipids and the structural conformation of HTL are thus highly correlated parameters, but the most probable model has some structural flexibility of the domain, in addition to a significant scattering contribution from a lipid core.

All models included a small portion of aggregates, varying from below 1% for the model with eight lipids and above 5% for the model with 29 lipids in the core. An increase in the amount of aggregates results in increased forward scattering, $I(0)$. An increase in the amount of lipids likewise increases $I(0)$. Therefore, it is striking that both quantities increase together. This is possible because the scale parameter, k, decreases as the amount of aggregations and lipids increase, from $k = 1.0$ for the model with eight lipids to $k = 0.68$ for the model with 29 lipids. The model with 18 lipids in the core included around 1% of aggregates. The relative contributions of the aggregates to the total scattering was minor, as shown visually in Fig. S2, but had to be taken into account, in particular for the low q part of the data.

In summary, the SANS analysis suggests that HTL has a lipid core of between 8 and 29 lipids and exhibits some flexibility of the periplasmic part of the SecDF domain.

CG simulation supports the existence of a lipid core

To assess the stability of the HTL complex and begin the characterization of a central lipid core to the complex as indicated by SANS, a CG MD study was performed. An atomic model of HTL was constructed using *E. coli* YidC (40), SecYEG (41), and *E. coli* homology models from *Thermus thermophilus* SecDF (12,42). These structures were arranged to fit the experimental cryo-EM density of the HTL (PDB: 5MG3). The atomic structures were converted to CG models using the Martini force field (31,43), inserted into a simulation box filled with randomly oriented Martini lipids (72% POPE, 22% POPG, and 6% cardiolipin), and solvated with Martini water and ions. The system was allowed to self-assemble, forming a clear lipid bilayer around the HTL. After bilayer formation and an initial settling period, *the* R_g of the protein complex settles at ~38 Å (Fig. 4), in good agreement with the scattering data (~41 Å). The HTL was simulated for a total duration of 3 μs (Fig. 4 *B*) and remained stable as determined using root mean-square deviation and R_g analysis (Fig. 4 *C*).

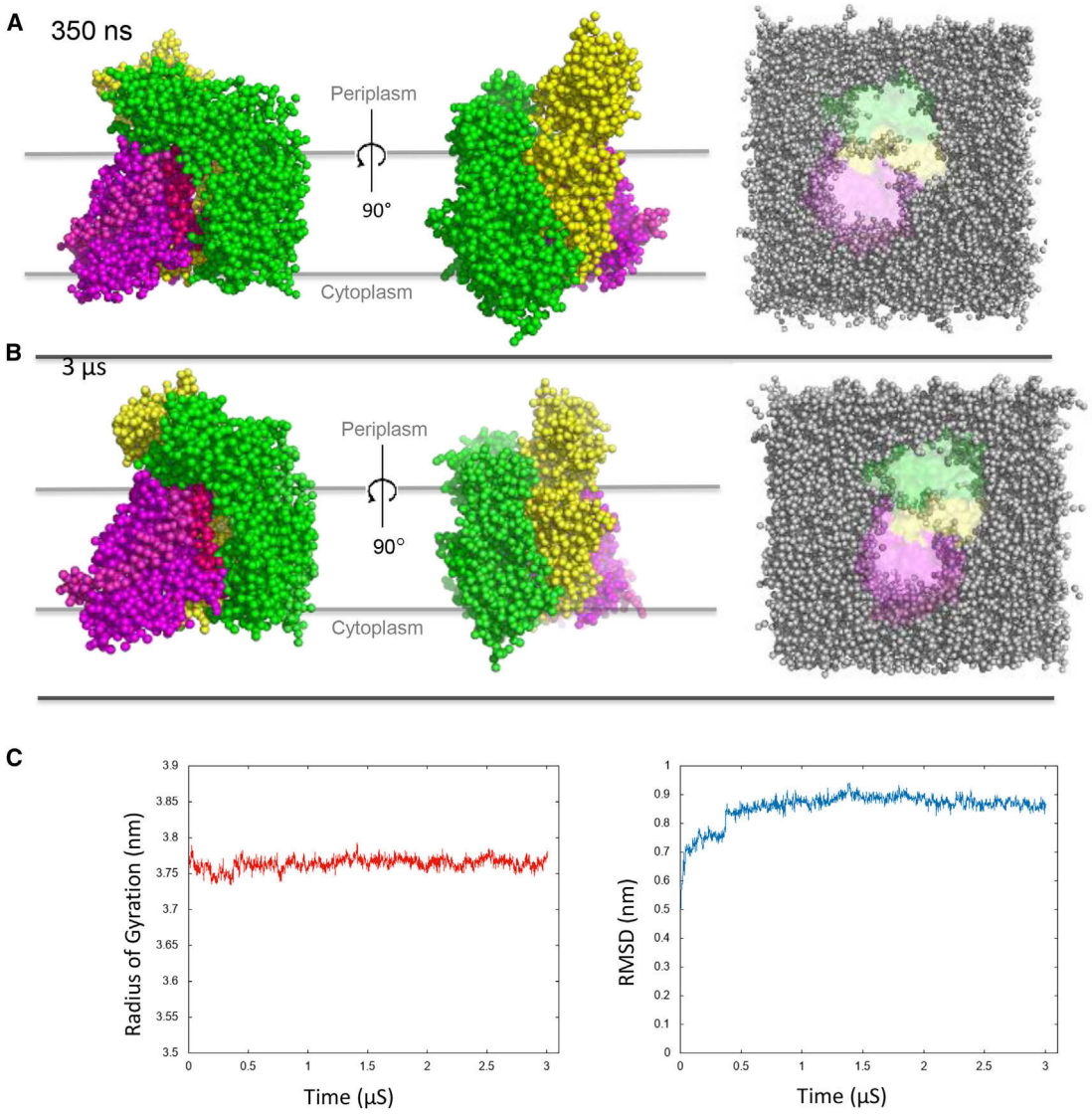

FIGURE 4 Coarse-grained HTL model, pre- and post-simulation. Shown is the coarse-grained HTL after 350 ns and 3 μs simulation. (*A*) HTL is shown after 350 ns simulation, viewed transversely through the membrane from two orientations and from cytoplasmic face, showing the lipid arrangement within the complex. SecYEG is shown in magenta, with SecDF in green and YidC in yellow. (*B*) The same thing is shown as previous, but after 3 μs of simulation. (*C*) Graphs show the stability of the structure over 3μs, both from the radius of gyration (*left*) and root mean-square deviation (*right*). To see this figure in color, go online.

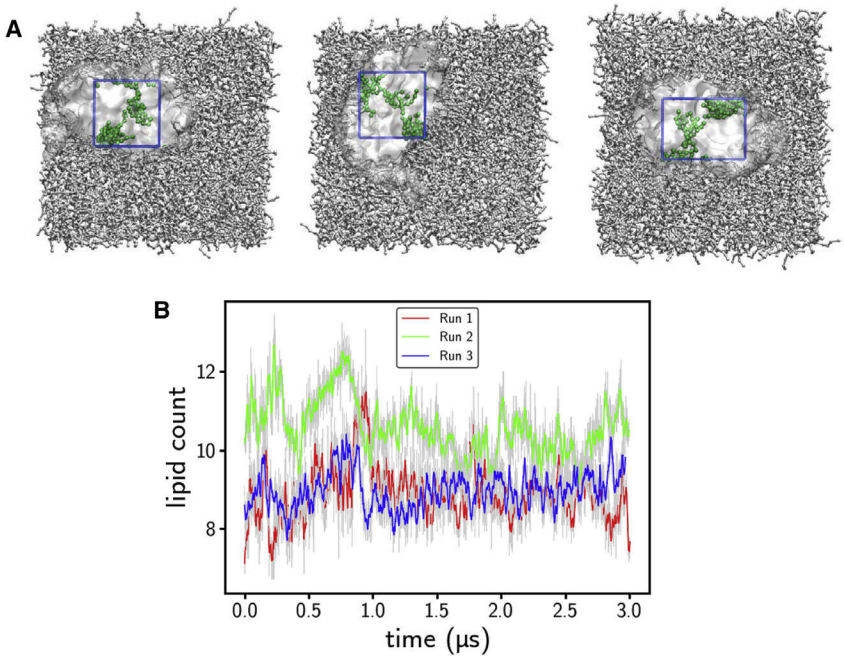

FIGURE 5 Localization and number of lipids within the HTL during MD. The localization of the lipids within the HTL during simulations is shown. (*A*) Shown are the shots of three independent CG simulations of HTL in a mixed lipid bilayer. In each image, the lipids present in the center after a 3 μs simulation are highlighted green. A boundary box was created for each simulation, and the lipid presence within the area was quantified. (*B*) Graph shows the number of lipids within the core of the HTL, as defined by the boundary box, over the course of the simulation time. To see this figure in color, go online.

CG modeling of the HTL complex shows the presence of a stable lipid pool at the interface between the transmembrane domains of all of the components of the HTL at the center of the complex (Fig. 5). The number of lipids within this island remains between 7 and 13 for the entire 3 μs duration of the simulation (Fig. 5 *B*). Furthermore, the average number of lipids remaining in the center of the HTL complex is 9.4 $\pm$ 0.8 lipids for the final 2 μs of a 3 μs simulation. Lipids are seen to diffuse in and out of the pool, predominantly through the gap between the SecY lateral gate and YidC, which may act as an opening point of the complex. Because of lipid diffusion, the lipid pool fluctuates in shape throughout the simulation but remains between ~20 and 40 Å in diameter depending on the number of lipids present.

An HTL model was extracted from the last frame of the simulation and converted to an atomistic model with seven POPE and two POPG in the core. This model had an overall structure similar to HTL-F-L. A model of HTL-I-L with the lipids from the simulations was also generated. The scattering from these models was calculated and compared with the data (Fig. 6). The HTL-I-L with the core of Martini lipids fitted the data with a χ_r^2 of 31.4, which was significantly better than the fit of HTL-F-L with Martini lipids (χ_r^2 = 164). Less than 1% of aggregates were included in the two fits. A model with a mix of the two structures with Martini lipids was tested, and the best fit was achieved with all the protein in the HTL-I-L form with Martini lipids (χ_r^2= 31.4). Thus, the simulated structures can to some extent explain the data but are not in full agreement (see also, the residuals on Fig. 6).

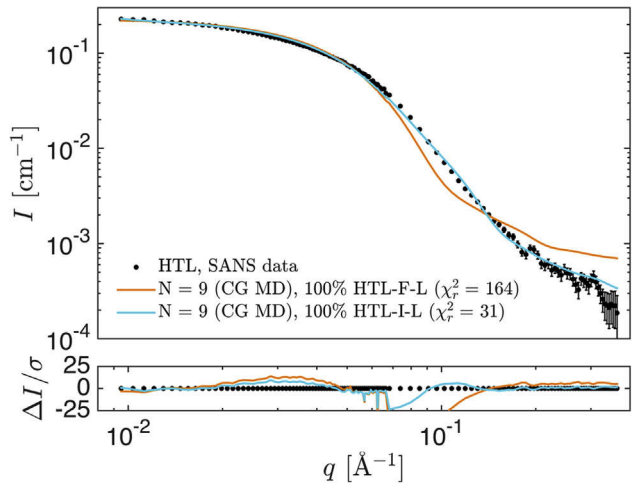

FIGURE 6 Model fit of HTL with postsimulated lipids to experimental SANS data. Experimental HTL SANS data (*black dots*) fitted with the HTL-F-L (*orange*) or HTL-I-L (*cyan*) structure, in each case with the nine lipids from the CG MD simulation. To see this figure in color, go online.

TABLE 1 The Estimated Number of Lipids in the Central Cavity of HTL, Estimated by Three Different Methods

Method	Estimated Number of Lipids in the Core	Shape of the Lipid Core (see Fig. 7)
CG MD	9.4 $\pm$ 0.8	lipid monolayer
SANS, I(0)	10 $\pm$ 14[a]	no shape assumed
SANS, fitting	8–29	lipid bilayer with cylindrical shape

[a]The number of lipids determined from the *I*(0) alone was affected by protein aggregation (see main text). These aggregations were taken into account in the fitting process.

DISCUSSION

The results presented show the HTL to be a dynamic complex, unequivocally demonstrating that the individual subunits are arranged around a central lipid core. The SANS data supports a model of the HTL containing a pool of between 8 and 29 lipids at its center. The fit is improved, accounting for the flexibility of SecDF, indicating that a significant part of the proteins have a rotated SecD periplasmic domain.

The lipid pool at the center of the HTL complex was observed to be stable during the CG MD simulations. This correlates well with both the SANS data in this study and previous structural studies of the HTL, indicating that the protein is located toward the periphery of the particle in solution with lipid and/or detergent material located toward the center (12). The number of lipids observed in the central core during the simulation remained between 7 and 13 (Fig. 5), toward the lower range of the estimate provided by SANS. Results are compared in Table 1. The lipids were observed to diffuse in and out of the core during the simulations, suggesting that there is a natural fluctuation of the lipid core volume in bilayer conditions.

In detergent-solubilized conditions, it has been shown that prolonged exposure to detergent can remove associated annular lipids from membrane proteins (44,45), though the integral lipids within the center of the HTL are not likely to be as easily dissociated. However, it is feasible that DDM exposure may destabilize the HTL, potentially facilitating the exchange of some central lipids into the surrounding detergent micelles, raising the possibility that the lipid numbers estimated by the SANS models may be lower than the true physiological value.

There was a difference between the applied lipid cores obtained by MD simulations and used in the SANS analysis. The lipid core used in the SANS analysis, and which fitted the data best, was a small cylindrical bilayer restrained to the center of the complex and spanning the entire transmembrane domain. The lipids in the CG MD simulation, on the other hand, formed a monolayer that penetrated the complex in the bilayer plane (Fig. 7). None of the two lipid core models are perfect descriptions. The cylindrical model is very simplified, with a homogeneous scattering contrast, a fully symmetric morphology. Moreover, the model was not checked for steric clashes between the lipid and protein and should therefore not be considered a fully physical

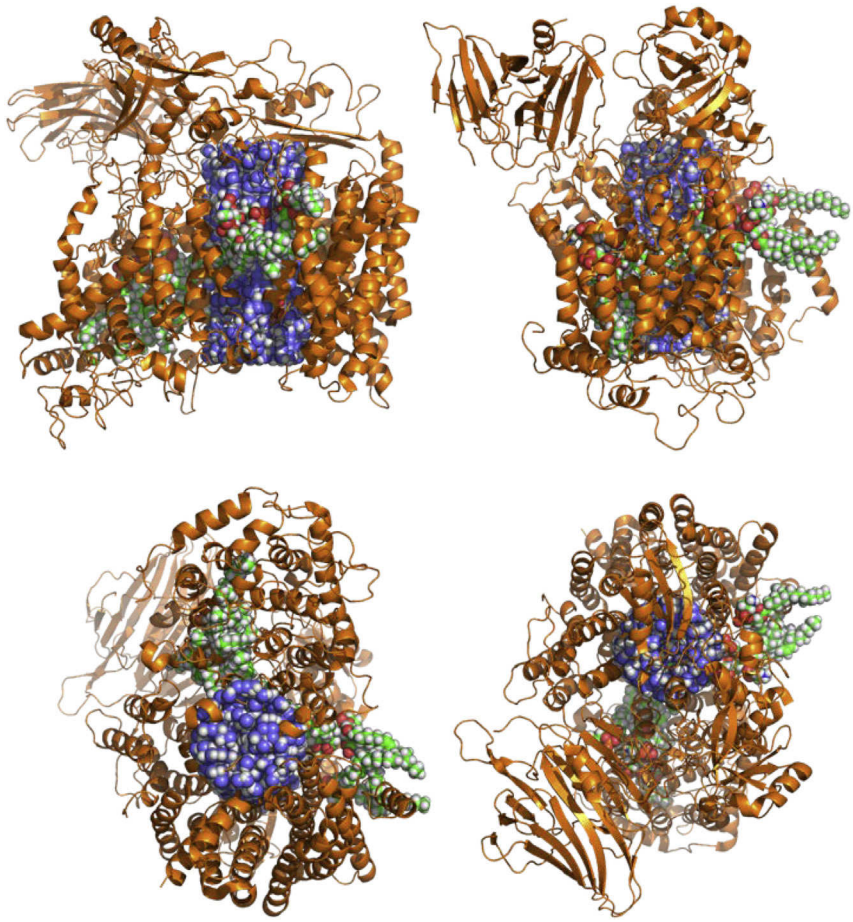

FIGURE 7 Atomistic comparison of HTL with positioned lipids positioned "by hand" or as simulated. The HTL complex (PDB: 5MG3, *orange*) with two different models for the lipid core is shown. The simple bilayer model was used to fit the SANS data (*blue*/*white*, corresponding to 18 lipids) and the monolayer of lipids from the CG MD simulations (*red*/*green*/*white*, nine lipids). Shown as orthogonal views, perpendicular to the plane of the membrane (*top*), and planar membrane views from the cytoplasm (*bottom left*) and periplasmic (*bottom right*). To see this figure in color, go online.

model. For that reason, it is not surprising that even the most probable models did not fit the data perfectly (Fig. 3). Some systematic discrepancies were apparent, especially around 0.1 $Å^{-1}$ (as clearly seen in the residual plot) and a χ_r^2 of 7.2 with 18 lipids in the core (i.e., still relatively far from unity).

The Martini lipid model is likely more plausible because it is constrained with force fields and has no steric clashes. The lipids are also better models for real lipids than a simple, homogeneous cylindrical lipid core. However, the Martini lipid model did not fit the data as well as the simpler cylindrical model (Fig. 6), as reflected in the χ_r^2 value of 31.4 for the best fit with HTL in the I-form. This suggests that there is room for the improvements of the simulations and that these data could possibly be used in future studies to benchmark and improve the applied force fields used for such challenging protein-lipid systems.

It should also be mentioned that an ensemble of structures might well be a better representation for the dynamic complex than a single static structure because of the diffusion of the lipids and the structural dynamic of the protein periplasmic parts. However, both lipid models, in their present form, provide valuable complementary information about the HTL proteo-lipid complex.

The MD simulation points to a potential gateway in the cytoplasmic membrane face between SecY and YidC, through which lipids are able to diffuse in and out of the lipid pore (Fig. 5). This gateway is capped at both ends by SecY residues previously identified as acidic lipid contact sites (46). During the insertion process, YidC is known to function in concert with SecY, performing chaperone activities that facilitate the correct folding of transmembrane helices as they sequentially exit the lateral gate of SecY (47,48). In this context, the encapsulated lipidic microenvironment between this lateral gate and YidC would prevent aggregation and help achieve the native state of membrane proteins during the co-translational insertion process. This mechanism could operate for assisted folding of small membrane proteins or for successively emerging helical bundles of larger polytopic substrates. The subsequent release of membrane proteins or the sequential release of their domains from the holocomplex would then be facilitated by the observed flexibility of the HTL (12), presumably by the partitioning of SecYEG and SecDF-YajC subcomplexes. This deployment of lipids for encapsulated membrane protein folding and controlled release presumably offers an effective means to efficiently achieve native states of monomeric proteins and assembly-competent states for multisubunit complexes.

SUPPORTING MATERIAL

Supporting Material can be found online at https://doi.org/10.1016/j.bpj.2019.04.002.

AUTHOR CONTRIBUTIONS

R.M. and I.C. designed protein preparations for SANS, executed by R.M. C.S. and L.A. conceived the SANS experiments. R.M., S.R.M., A.H.L., and L.A. designed the SANS experiment. R.M., S.R.M., A.H.L., and H.F. conducted the SANS experiments and performed the initial analysis. A.H.L. and L.A. designed the SANS data analysis, which was then carried out by A.H.L. with inputs from L.A., R.M., and S.R.M. R.M. and R.A.C. designed and performed the MD experiments. R.M. and R.A.C. performed analysis of the MD data. R.M., A.H.L., L.A., and I.C. co-wrote the manuscript with input from all co-authors.

ACKNOWLEDGMENTS

Part of this work is based upon experiments performed at the KWS-1 instrument and the authors would like to thank for the awarded beamtime. We would also thank Frederik Tidemand and Nicolai T. Johansen for great support during the SANS beamtime. We thank for the deuterated DDM (d-DDM) that was synthesised by Dr. Tamim Darwish (Australia's Nuclear Science and Technology Organisation, NSW, Australia).

R.M. received a University of Bristol Postgraduate Scholarship. Additional support was gratefully received by I.C. from the Biotechnology and Biological Sciences Research Council (BB/M003604/1 and BB/I008675/1). L.A. and A.H.L. thank the CoNeXT project and University of Copenhagen for co-funding the project. The authors gratefully acknowledge the financial support provided by Jülich Centre for Neutron Science to perform the neutron scattering measurements at the Heinz Maier-Leibnitz Zentrum (Garching, Germany).

REFERENCES

1. Brundage, L., J. P. Hendrick, …, W. Wickner. 1990. The purified E. coli integral membrane protein SecY/E is sufficient for reconstitution of SecA-dependent precursor protein translocation. *Cell.* 62:649–657.
2. Görlich, D., and T. A. Rapoport. 1993. Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. *Cell.* 75:615–630.
3. Rapoport, T. A., L. Li, and E. Park. 2017. Structural and Mechanistic Insights into Protein Translocation. *Annu. Rev. Cell Dev. Biol.* 33:369–390.
4. Cranford-Smith, T., and D. Huber. 2018. The way is the goal: how SecA transports proteins across the cytoplasmic membrane in bacteria. *FEMS Microbiol. Lett.* 365.
5. Pfeffer, S., L. Burbaum, …, F. Förster. 2015. Structure of the native Sec61 protein-conducting channel. *Nat. Commun.* 6:8403.
6. Duong, F., and W. Wickner. 1997. The SecDFyajC domain of preprotein translocase controls preprotein movement by regulating SecA membrane cycling. *EMBO J.* 16:4871–4879.
7. Scotti, P. A., M. L. Urbanus, …, J. Luirink. 2000. YidC, the Escherichia coli homologue of mitochondrial Oxa1p, is a component of the Sec translocase. *EMBO J.* 19:542–549.
8. Schulze, R. J., J. Komar, …, I. Collinson. 2014. Membrane protein insertion and proton-motive-force-dependent secretion through the bacterial holo-translocon SecYEG-SecDF-YajC-YidC. *Proc. Natl. Acad. Sci. USA.* 111:4844–4849.
9. Müller, M., H. G. Koch, …, U. Schäfer. 2001. Protein traffic in bacteria: multiple routes from the ribosome to and across the membrane. *Prog. Nucleic Acid Res. Mol. Biol.* 66:107–157.
10. Bieniossek, C., Y. Nie, …, I. Berger. 2009. Automated unrestricted multigene recombineering for multiprotein complex production. *Nat. Methods.* 6:447–450.

11. Komar, J., S. Alvira, …, I. Collinson. 2016. Membrane protein insertion and assembly by the bacterial holo-translocon SecYEG-SecDF-YajC-YidC. *Biochem. J.* 473:3341–3354.
12. Botte, M., N. R. Zaccai, …, C. Schaffitzel. 2016. A central cavity within the holo-translocon suggests a mechanism for membrane protein insertion. *Sci. Rep.* 6:38399.
13. Xu, Z., A. L. Horwich, and P. B. Sigler. 1997. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature.* 388:741–750.
14. Van den Berg, B., W. M. Clemons, Jr., …, T. A. Rapoport. 2004. X-ray structure of a protein-conducting channel. *Nature.* 427:36–44.
15. Kumazaki, K., S. Chiba, …, O. Nureki. 2014. Structural basis of Sec-independent membrane protein insertion by YidC. *Nature.* 509:516–520.
16. Tsukazaki, T., H. Mori, …, O. Nureki. 2011. Structure and function of a membrane component SecDF that enhances protein export. *Nature.* 474:235–238.
17. Midtgaard, S. R., T. A. Darwish, …, L. Arleth. 2018. Invisible detergents for structure determination of membrane proteins by small-angle neutron scattering. *FEBS J.* 285:357–371.
18. Pedersen, M. C., L. Arleth, and K. Mortensen. 2013. WillItFit: a framework for fitting of constrained models to small-angle scattering data. *J. Appl. Cryst.* 46:1894–1898.
19. Persson, F., P. Söderhjelm, and B. Halle. 2018. The geometry of protein hydration. *J. Chem. Phys.* 148:215101.
20. Lomize, M. A., I. D. Pogozheva, …, A. L. Lomize. 2012. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* 40:D370–D376.
21. Hansen, S. 2012. BayesApp: a web site for indirect transformation of small-angle scattering data. *J. Appl. Cryst.* 45:566–567.
22. Vestergaard, B., and S. Hansen. 2006. Application of Bayesian analysis to indirect Fourier transformation in small-angle scattering. *J. Appl. Cryst.* 39:797–804.
23. Franke, D., C. M. Jeffries, and D. I. Svergun. 2015. Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. *Nat. Methods.* 12:419–422.
24. Rambo, R. P., and J. A. Tainer. 2013. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature.* 496:477–481.
25. Armen, R. S., O. D. Uitto, and S. E. Feller. 1998. Phospholipid component volumes: determination and application to bilayer structure calculations. *Biophys. J.* 75:734–744.
26. Larsen, A. H., J. Dorosz, …, J. S. Kastrup. 2018. Small-angle neutron scattering studies on the AMPA receptor GluA2 in the resting, AMPA-bound and GYKI-53655-bound states. *IUCrJ.* 5:780–793.
27. Teixeira, J. 1988. Small-angle scattering by fractal systems. *J. Appl. Cryst.* 21:781–785.
28. Kotlarchyk, M., and S. Chen. 1983. Analysis of small angle neutron scattering spectra from polydisperse interacting colloids. *J. Chem. Phys.* 79:2461–2469.
29. Svergun, D., C. Barberato, and M. H. Koch. 1995. CRYSOL– A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* 28:768–773.
30. Stansfeld, P. J., J. E. Goose, …, M. S. Sansom. 2015. MemProtMD: automated insertion of membrane protein structures into explicit lipid membranes. *Structure.* 23:1350–1361.
31. Monticelli, L., S. K. Kandasamy, …, S. J. Marrink. 2008. The MARTINI coarse-grained force field: extension to proteins. *J. Chem. Theory Comput.* 4:819–834.
32. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 22:2577–2637.
33. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101.
34. Parrinello, M., and A. Rahman. 1981. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52:7182–7190.
35. Stansfeld, P. J., and M. S. Sansom. 2011. From coarse grained to atomistic: a serial multiscale approach to membrane protein simulations. *J. Chem. Theory Comput.* 7:1157–1166.
36. Best, R. B., X. Zhu, …, A. D. Mackerell, Jr. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles. *J. Chem. Theory Comput.* 8:3257–3273.
37. Guinier, A., and G. Fournet. 1955. Small-Angle Scattering of X-Rays. Wiley, New York.
38. Furukawa, A., K. Yoshikaie, …, T. Tsukazaki. 2017. Tunnel formation inferred from the I-form structures of the proton-driven protein secretion motor SecDF. *Cell Reports.* 19:895–901.
39. Larsen, A. H., L. Arleth, and S. Hansen. 2018. Analysis of small-angle scattering data using model fitting and Bayesian regularization. *J. Appl. Cryst.* 51:1151–1161.
40. Kumazaki, K., T. Kishimoto, …, O. Nureki. 2014. Crystal structure of Escherichia coli YidC, a membrane protein chaperone and insertase. *Sci. Rep.* 4:7299.
41. Tanaka, Y., Y. Sugano, …, T. Tsukazaki. 2015. Crystal structures of SecYEG in lipidic cubic phase elucidate a precise resting and a peptide-bound state. *Cell Reports.* 13:1561–1568.
42. Eswar, N., B. Webb, …, A. Sali. 2007. Comparative protein structure modeling using MODELLER. *In* Current Protocols in Protein Science. John Wiley & Sons, Inc., Hoboken, NJ, pp. 2.9.1–2.9.31.
43. Marrink, S. J., H. J. Risselada, …, A. H. de Vries. 2007. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B.* 111:7812–7824.
44. Bechara, C., A. Nöll, …, C. V. Robinson. 2015. A subset of annular lipids is linked to the flippase activity of an ABC transporter. *Nat. Chem.* 7:255–262.
45. Laganowsky, A., E. Reading, …, C. V. Robinson. 2014. Membrane proteins bind lipids selectively to modulate their structure and function. *Nature.* 510:172–175.
46. Corey, R. A., E. Pyle, …, I. Collinson. 2018. Specific cardiolipin-SecY interactions are required for proton-motive force stimulation of protein secretion. *Proc. Natl. Acad. Sci. USA.* 115:7967–7972.
47. Nagamori, S., I. N. Smirnova, and H. R. Kaback. 2004. Role of YidC in folding of polytopic membrane proteins. *J. Cell Biol.* 165:53–62.
48. Zhu, L., H. R. Kaback, and R. E. Dalbey. 2013. YidC protein, a molecular chaperone for LacY protein folding via the SecYEG protein machinery. *J. Biol. Chem.* 288:28180–28194.