The structure of the bacterial protein translocation complex SecYEG

I. Collinson

Department of Biochemistry, University of Bristol, Bristol BS8 1TD, U.K.

Abstract
Proteins destined for secretion, membrane insertion or organelle import contain signal sequences that direct them to the membrane. Once there, transport machines receive and translocate them appropriately across or into the membrane. The related SecY and Sec61 protein translocation complexes are ubiquitous components of machines that are essential for protein transport. They co-operate with various partners such that the substrate polypeptide is pulled or pushed through the membrane by post- or co-translational mechanisms. In bacteria and archaea, the SecY complex (SecYEG) is a heterotrimer, which associates with ribosomes so that the polypeptide is threaded through the channel during its synthesis. Bacteria possess an additional pathway, whereby the newly synthesized substrate protein is maintained in an unfolded state and is engaged by the ATPase SecA and delivered to the translocon. Recent medium- (cryo-electron microscopy) and high-resolution (X-ray) structures of the Sec complex have dramatically increased our understanding about how proteins pass through membranes, but have posed a number of new questions. The Sec complex is active as an oligomer, but the structure indicates that the protein-conducting channel is formed by a monomer of SecYEG. Structures of the membrane-bound dimer of Escherichia coli SecYEG and the detergent-solubilized monomer of Methanococcus jannaschii SecYEG will be described and discussed in the context of the mechanism that underlies protein secretion and membrane insertion.

Key words: membrane protein, protein translocation, secretion, SecYEG, translocon.

Abbreviations used: BiP, immunoglobulin heavy-chain binding protein; hsp70, heat-shock protein 70; SRP, signal recognition particle; TM, transmembrane.

1email: ian.collinson@bristol.ac.uk

Introduction
Polypeptide transport and import usually occurs when they are in an unfolded state. In rare cases, proteins that necessarily fold before their transport across a membrane are translocated by the so-called TAT (twin arginine translocase), which translocates fully folded and often large proteins across the membrane (reviewed in [1]). This poorly understood pathway probably arose to translocate proteins that have folded in the cytosol following incorporation of a cofactor required for activity. In contrast, like the import complexes of chloroplasts and mitochondria, the Sec complex receives the substrate polypeptide in an unfolded state before passage through or into the membrane. The core of the Sec-dependent pre-protein export pathway in all cells is formed by the membrane-bound Sec complex [2], which retains sequence homology throughout [3]. The nature of the protein channel must therefore be the same from bacteria to humans.

The Sec complex provides a pathway for the passage of large and diverse objects through the hydrophobic membrane barrier, without loss of small molecules such as ions and valuable metabolites. Even more remarkable is that the channel possesses the capability to appropriately send proteins through or into the membrane, for respective secretory and membrane proteins. This dual-gating mechanism is an essential feature of the translocation complex and is also unique among membrane transporters; in some cases, both pathways need to act on single substrates, which have both membrane and large soluble domains. Polytopic membrane proteins have both hydrophobic and hydrophilic segments, which define their membrane topology. The translocon needs to recognize these elements and act to pass hydrophobic TM (transmembrane) domains into the membrane phase, and translocate certain domains through the membrane. Other domains need not be translocated and remain on the inside. How this occurs has, until recently, been a mystery.

Co-translational translocation
While the movement of proteins through the complex occurs by the same mechanism, the manner in which they are driven can differ considerably according to the species and the nature of the polypeptide substrate. Co-translational translocation probably occurs in every cell. The N-terminus of the protein substrate emerges from the exit channel of the ribosome exit site and is recognized by the SRP (signal recognition particle), which is also highly
Post-translational translocation

Eukaryotes

Another mechanism used to drive protein translocation through the channel is post-translational, and eukaryotes and bacteria deploy different mechanisms. This version of translocation is thought to be the favoured pathway in yeast, and this is also where it has been best characterized. The core trimeric Sec61 complex associates with additional components to form a seven-component complex that is capable of protein translocation [16]. The basic mechanism of transport in this case is quite well understood. The protein channel can be considered as a passive conduit, capable of accommodating unfolded polypeptides and allowing the passage of specific segments through the membrane. Bound secretory proteins are free to diffuse in both directions with respect to the endoplasmic reticulum lumen. An hsp70 (heat-shock protein 70) homologue, BiP (immunoglobulin heavy-chain binding protein), is a resident of the lumen and acts as a molecular ratchet to grab the emerging polypeptide, which in turn prevents backsliding; this effect is achieved by the action of several BiP molecules [14]. One component of the seven-component post-translational Sec complex contains a DNA J-domain on its luminal face, which serves to activate BiP [15]. The result is the hydrolysis of the ATP bound to BiP and an increased affinity for unfolded polypeptide [15]. The interaction with the J-domain is transient, so that the BiP polypeptide substrate is free to diffuse away, leaving enough space for another BiP to access the J-domain, become activated and grab the next stretch of polypeptide, effecting a molecular ratchet. The cycle is completed as the BiP-bound ADP is exchanged for ATP, the hsp70 is released, and the transported protein is then free to fold. Similar processes also occur in the mitochondrial import of precursor proteins [16]. The net result of the reaction is the loss of ATP as the polypeptide is pulled through the protein channel and folds.

Prokaryotes

Protein translocation through the Sec complex in bacteria occurs on the plasma membrane, and therefore cannot be driven by this type of ‘pulling’ reaction, as there is no ATP outside of the cell. Instead, another mechanism has evolved to push the substrate through the membrane. Substrates for the secretory pathway after their synthesis are held in an unfolded and translocation-component conformation by the chaperone SecB [17,18]. An ATPase, SecA, receives the pre-protein from this complex and engages the Sec complex (SecYEG in Escherichia coli) to initiate the translocation reaction [17]. Protein translocation proceeds as ATP is hydrolysed, a reaction that has been reconstituted in vivo [19]. A wealth of data has been published on the genetics and biochemistry of this enzyme, and, more recently, three structures have been published [20–22]. The structures reveal two RecA-like folds, sandwiching a bound nucleotide; they vary with respect to their oligomeric state and by the presence of an opened groove, proposed to be occupied by polypeptides during protein translocation [20]. In spite of the published work, the following key points remain to be resolved. The oligomeric state that SecA adopts during the reaction has yet to be clarified; dimers have been implicated [23–25], as well as monomers [25,26]. The nature of the interaction with the channel complex SecYEG has also not been defined in any detail; there have been suggestions that it is able to penetrate the entire depth of the membrane, and in this way drive the translocation reaction [27,28]. How the binding and hydrolysis of ATP conduct the conformational changes through SecA and SecYEG to drive the processive movement of polypeptide through the membrane is also yet to be resolved.

Structure determination of the protein pore

Concerning the channel itself, thanks to recent structural insights, our level of understanding is at a reasonable advanced stage. In order to obtain structural information of sufficient detail to understand the mechanism and to relate to previous biochemical and genetic data, crystallographic approaches are required. The SecYEG complex found in E. coli has been overexpressed and purified in yields sufficient to grow crystals for structural analysis [29].

One approach is to grow two-dimensional crystals of the complex. This involves the incorporation of membrane proteins into phospholipid vesicles to such a density that they associate with one another and form crystalline patches within the plane of the membrane, and hence two-dimensional crystals [30]. The crystals are then visualized by cryo-electron microscopy to collect the data to calculate a three-dimensional model. Structures determined in this way are limited by the crystal in respect of the resolution attainable in the final map. Potentially, atomic detail can be achieved, as in the case of bacteriorhodopsin [31], the light-harvesting complex II [32] and aquaporin [33,34]. Typically, however, the resolution is between 6–8 Å (1 Å = 0.1 nm) [35–38], sufficient...
to resolve secondary-structural elements such as TM α-helices, but not to assign them or to resolve amino-acid side chains. Neither can the loops be seen at this resolution, as they are usually less well defined and disordered.

In spite of this disadvantage, there remain several good reasons to pursue this avenue of research. In comparison with three-dimensional crystallography, approx. ten times less material is required for the growth of crystals and data collection. Moreover, the crystals form with the protein in their native environment, embedded in the lipid bilayer. This confers a large stabilization effect and increases the chances that the protein will be crystallized in its native state. The fact that these structures appear as they do in the membrane is particularly helpful when studying protein translocation, since the reaction relies on lipids and the existence of an intact membrane; being in a native-like state, it also adds the potential to generate structures with alternative conformations and co-complexes with substrates and partner proteins. The low-to-medium-resolution structure determined by electron microscopy of complex assemblies may be impossible to approach by X-ray crystallography, as crystals are generally much more difficult to attain. Therefore, if the simplified components of larger complexes can be crystallized in three dimensions, to generate high-resolution data, this information can be used together with the electron-microscopy data. The idea is that a medium-resolution scaffold of a complex machine can be used as a template to ‘jigsaw’ in more detailed information of the individual components that it contains. There are many examples where this has been used with great success, and it is the approach that we hope to use to study the large translocation assemblies which act as the proteins pass through the membrane.

Medium-resolution structure determination of the SecY complex by cryo-electron microscopy

Two-dimensional crystals of the E. coli SecYEG complex were grown by adding lipids to the purified sample in the detergent C₁₂E₉ (dodecyl nonaoxyethylene ether). When the correct cocktail of detergent, lipids, protein, pH and salt concentration were incubated at 23°C, and the detergent was removed slowly by dialysis, membranes formed which contained two-dimensional crystals [29]. These crystals were grown by essentially the same procedure used to produce active vesicles incorporating SecYEG, except that the protein/lipid ratio needed to be quite a bit higher to drive lattice formation [29]. Therefore there is every reason to believe that the complex found in the crystals is in its native state, capable of protein translocation. Crystals were used to collect data, and a three-dimensional structure was calculated [39].

The first detailed view of a protein translocation complex

A dimer of the SecYEG complex was resolved in the membrane to 8 Å resolution, which was good enough to see all of the 15 TM α-helices in each of the two complexes. The loops connecting the helices could not be seen, except for a few faint densities outside of the membrane domain (see below). At this resolution, neither the α-helical segments nor the amino-acid side chains could be assigned. Each monomeric complex consisted of 13 core TM domains, with an additional two peripheral segments (Figure 1). An interesting feature was a highly tilted TM helix on the monomer–monomer interface which was determined to be the essential third TM helix of SecE. Because of the limits in the detail of the map, further interpretation of the structure with respect to function required a higher-resolution map. This was achieved by employing X-ray crystallography.

High-resolution structure determination of the SecY complex by X-ray crystallography

In the first place, it was quickly realized that the E. coli SecYEG complex proved refractory to three-dimensional crystallization attempts. Although the trials showed initial promise, the crystals that were grown were very small, soft and only diffracted X-rays to a low order. This was probably because of the instability of the complex in high detergent concentrations, as well as the long incubation times at room temperature required to grow crystals. A long and intensive effort finally produced crystals of sufficient quality and order required to calculate a workable structure. The key finding was that the homologue from Methanococcus jannaschii seemed to crystallize more easily. This was probably a result of its increased stability, as the organism is active at temperatures exceeding 80°C. The first crystals of the SecYE heterodimeric complex obtained diffracted to approx. 7 Å, which seemingly could not be improved upon. After the identification of a third subunit, Secβ, from within the genome,
it was incorporated into the expression plasmid, from where the SecYEβ heterotrimeric complex could be produced in very high levels [40]. When dissolved in the detergent DHPC (1,2-diheptanoyl-sn-glycero-3-phosphocholine), it was found to diffract crystals beyond 3.5 Å, enough to collect the data and solve the phase problem from selenomethionine-derivatized complexes [40]. The structure of the detergent-solubilized monomer was calculated to an atomic resolution, and it revealed several interesting and surprising results [40].

The architecture of the SecY complex
An animation showing the SecYEG complex can be seen at http://www.biochemsoctrans.org/bst/033/bst0331225add.htm. The most obvious finding, in contrast with the E. coli structure, was that the complex had crystallized as a monomer. This is consistent with the fact that high detergent concentrations bring about dissociation of the dimeric form of the E. coli SecYEG [41], although it should be stated that the active oligomeric form of the archaeal complex has yet to be determined. Indeed, there have been no functional analyses of the SecYEβ complex, the majority of studies having been performed on the E. coli, yeast and mammalian systems. Therefore it was quite important that we had the E. coli structure for a comparison to validate the high-resolution X-ray structure. The architecture of each SecYEG and SecYEβ were, reassuringly, very similar (Figure 1), except that the E. coli complex has an extra three TM helices [39,40]; two are the peripheral pair, belonging to SecE, and the other forms part of SecG, which has two TM domains compared with the single helix in Secβ.

The complex comprises 12 TM domains connected by loops on each face of the membrane plane [40] (Figure 1). The largest are on the cytosolic face, between TM6 and TM7 and TM8 and TM9, which presumably interacts with the ribosome during co-translational translocation. Their counterparts in E. coli also interact with the partner SecA. SecE has a single highly tilted TM domain, which tracks across one face of the complex, making extensive contacts with the larger SecY subunit. In addition, there is an amphipathic helix, which runs parallel to the membrane on the cytosolic face. This helix was also visible in the electron-microscopy map, and was extended by the two extra peripheral TM helices on the other side of the complex (Figure 1). It appears that SecE forms a clamp across the SecY subunit, perhaps acting as a brace to hold it in a compact and closed conformation. The tilted helix of SecE has been cross-linked to the corresponding TM domain in another SecYEG complex, and these cross-links were shown to modulate the translocation activity [42]. Therefore, in an active oligomeric assembly of the Sec complex, this tilted TM domain must be involved in intermolecular interactions, exactly what we observe in the membrane-bound dimeric form [39]. Secβ was also identified as a single helix on the outside of the complex, consistent with the fact that it is not essential for the major activity of the protein channel.

SecY is formed from two domains that share an inverted pseudo-symmetry; each half, TM1–5 and TM6–10, is arranged in the fashion of a clamshell, and these are connected by a hinge between TM5 and TM6, which is also visible in the electron-microscopy map. The arrangement is such that a lateral opening could only occur from one side of the complex. The protein-conduction pore can be identified in the centre of the complex, which is formed between two halves of the SecY subunit. The channel, found in the centre of a monomeric complex, is blocked by a short plug domain and a ring of hydrophobic residues [40] (Figures 1 and 2). The plug domain is an unusual structure formed by helix 2a and resides in the centre of the channel on the periplasmic side of the complex. This structural element could also be seen in the electron-microscopy structure [39].

Three lines of evidence support the fact that the centrally blocked pore forms the path for translocating polypeptide. First, the pore lining exhibits considerably higher sequence conservation with other SecY sequences, compared with the peripheral areas. Secondly, mutations that influence the activity of the channel involve residues that line the putative channel. Finally, the most convincing argument comes from the fact that numerous cross-links between translocating polypeptide and SecY were found along the pore [43]. The structure of the SecY complex also revealed the location of the signal-sequence-binding site, identified by cross-linking to be between TM2b and TM7 of SecY [44]. It is situated adjacent to the channel and between the two lobes of each half of the SecY, at the opening points of the clam-shell-shaped subunit [40] (Figures 1 and 3). An analysis of the size of the protein pore in the active mammalian translocon revealed that it might open up to a diameter of 60 Å [45].

Mechanism of channel gating
The mechanism of channel gating and the rearrangements required to accommodate and translocate proteins are not clear. The channel might open by the movement of the two N- and C-terminal domains of SecY about the hinge between the two halves. In addition, it appears that the displacement
of the plug to a new position in the structure would be required (Figures 2 and 3). This hypothesis has support from an earlier in vivo cross-linking experiment [46]. It was shown that cysteine residues introduced in the domain of SecY(F67C) forming the plug and at the C-terminal end of SecE(S120C) can form disulphide bridges. These two cysteine residues are 20 Å apart in the closed channel structure, so that the observed cross-links can now be understood by the movement of the plug away from the channel to this new position. Accordingly, the disulphide bridge formation was found to have a dominant lethal effect, by locking the channel into a permanently open state [46]. Further evidence supports the fact that the plug might move during protein transport, as the plug is in a slightly different position in the active membrane-bound dimeric complex [40]. Paradoxically, several lines of evidence suggest that the active form of the SecY complex is oligomeric [9,41,47,48]. The membrane-bound form of the complex was visualized as a dimer [39]. The dimer seems to be an active unit of the complex, remaining so when it is associated with SecA [25]. Likewise, two SecY molecules, linked genetically, were shown to be active and to form a high-affinity binding site for SecA [24]. However, the formation of higher oligomeric states, such as tetramers [47], at some stage in the post-translational reaction cycle cannot be ruled out. During the co-translational mode of translocation in eukaryotes, the active channel bound to ribosomes consist of four copies of Sec61, probably arranged as a dimer of dimers [49]. The reasons for the apparent requirement for either two or four copies of the SecY complex in the active translocon remain unknown. Perhaps a larger assembly might be needed to provide enough room for the association of the large partners. The interaction between SecYEG monomers at the interface may bring about conformational changes that are essential for partner binding and for subsequent translocation reactions. The observed variability in the oligomeric state is also puzzling, but might be a result of species- and substrate-dependent differences, or perhaps the arrangement may vary over different stages of the translocation cycle.

Scope for the future
The structural work described above, together with the previous genetic and biochemical findings, have brought us to a position where we can begin to understand the underlying mechanisms of protein translocation through a biological membrane. However, as is often the case, a new finding arises to pose a set of new and unanswered questions. The structures as we find them are obviously in their closed...
state. How the channel opens to accommodate the substrate needs to be addressed, to verify that the models propose are indeed accurate. In addition, the nature of the interaction with the partner proteins, and the mechanisms that transduce the energy of nucleotide hydrolysis to the vectorial movement of protein across and into the membrane, will form the basis for future experiments. The internal workings of the translocon, with respect to the conformational changes that occur during the reaction cycle, must now be addressed through further biophysical investigation. The key will be to determine structures of the complex in an open state, and also together with substrate and partner proteins. The combination of electron microscopy of large translocation complexes with the existing X-ray structures of SecA and the SecY complex might prove to be useful in this sense. Equally important experiments must also be tackled to look at the dynamics and kinetics of the reaction resulting as ATP drives proteins through the SecYE–SecA complex. These outstanding problems will be helpfully guided by the new novel and difficult strategies, as well as some clever experimental design.

This review describes work carried out by a large number of people over several years, and, in particular, I mention Professor Tom Rapoport, Professor Steve Harrison and Professor Werner Kühlbrandt, with whom it has been a pleasure to work. I have enjoyed generous support from EMBO, Howard Hughes Medical Institute, Human Frontiers Research Program, Max-Planck-Institute for Biophysics, and BBSRC (BB/C503538/1). Thanks also to Dr Mihnea Bostina and the partner proteins, and the mechanisms that transduce the energy of nucleotide hydrolysis to the vectorial movement of protein across and into the membrane, will form the basis for future experiments. The internal workings of the translocon, with respect to the conformational changes that occur during the reaction cycle, must now be addressed through further biophysical investigation. The key will be to determine structures of the complex in an open state, and also together with substrate and partner proteins. The combination of electron microscopy of large translocation complexes with the existing X-ray structures of SecA and the SecY complex might prove to be useful in this sense. Equally important experiments must also be tackled to look at the dynamics and kinetics of the reaction resulting as ATP drives proteins through the SecYE–SecA complex. These outstanding problems will be helpfully guided by the new novel and difficult strategies, as well as some clever experimental design.

This review describes work carried out by a large number of people over several years, and, in particular, I mention Professor Tom Rapoport, Professor Steve Harrison and Professor Werner Kühlbrandt, with whom it has been a pleasure to work. I have enjoyed generous support from EMBO, Howard Hughes Medical Institute, Human Frontiers Research Program, Max-Planck-Institute for Biophysics, and BBSRC (BB/C503538/1). Thanks also to Dr Mihnea Bostina and Ms Brigitte Mohsin for helping with one of the Figures.

References