Structure of Dimeric SecA, the *Escherichia coli* Preprotein Translocase Motor

Yannis Papanikolau¹, Maria Papadovasilaki¹, Raimond B. G. Ravelli²
Andrew A. McCarthy², Stephen Cusack², Anastassios Economou¹,³* and Kyriacos Petratos¹*  

¹Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology, PO Box 1385, 71110 Heraklion, Greece  
²European Molecular Biology Laboratory, Grenoble outstation, 6 Rue Jules Horowitz, 38042 Grenoble Cedex 9, France  
³Department of Biology, University of Crete, PO Box 2208, 71409 Heraklion, Greece

SecA is the preprotein translocase ATPase subunit and a superfamily 2 (SF2) RNA helicase. Here we present the 2 Å crystal structures of the *Escherichia coli* SecA homodimer in the apo form and in complex with ATP, ADP and adenosine 5′-[β,γ-imido]triphosphate (AMP-PNP). Each monomer contains the SF2 ATPase core ( DEAD motor) built of two domains (nucleotide binding domain, NBD and intramolecular regulator of ATPase 2, IRA2), the preprotein binding domain (PBD), which is inserted in NBD and a carboxy-terminal domain (C-domain) linked to IRA2. The structures of the nucleotide complexes of SecA identify an interfacial nucleotide-binding cleft located between the two DEAD motor domains and residues critical for ATP catalysis. The dimer comprises two virtually identical protomers associating in an antiparallel fashion. Dimerization is mediated solely through extensive contacts of the DEAD motor domains leaving the C-domain facing outwards from the dimerization core. This dimerization mode explains the effect of functionally important mutations and is completely different from the dimerization models proposed for other SecA structures. The repercussion of these findings on translocase assembly and catalysis is discussed.

*Corresponding authors

Keywords: ATPase; crystal structure; DNA–RNA helicase; protein translocation; SecA

Introduction

Most extracytoplasmic proteins are translocated into and across the bacterial inner membrane by a complex membrane transporter called the translocon or translocase.¹,² The translocase is comprised of the core heterotrimeric SecY/SecE/SecG transmembrane protein conducting channel and a peripheral ATPase motor called SecA. The *Escherichia coli* translocase has been fully reconstituted in a functional state in *vitro*³ and the high resolution structures of an archaeal SecYEG⁴, the SecA from *Bacillus subtilis* (bsSecA)⁵–⁷ and a SecA from *Mycobacterium tuberculosis* (mtSecA)⁸ have been reported.

Secreted proteins are synthesized with N-terminal signal peptides and interact with export-specific piloting factors and chaperones. These interactions target the preprotein to the membrane where they bind to SecA.³ Preprotein binding triggers ATP-driven cycles of conformational changes in SecA bound at SecYEG, described as insertion/de-insertion cycles.⁹,¹⁰ During SecA membrane cycling, 20–30 residue segments of the preprotein are consecutively moved into the membrane through SecYEG, described as insertion/de-insertion cycles.¹² These results suggest that SecA acts as a processive machine and that specific co-ordinated motions in SecYEG take place.⁷
The functional oligomeric state of SecYEG and of SecA is a matter of controversy. *E. coli* SecA (ecSecA) forms a stable dimer in solution at concentrations comparable to those in the cell cytoplasm.\(^{14-16}\) ecSecA is functional as a homodimer during protein translocation.\(^{17-19}\) However, other studies propose that SecA is functional as a monomer.\(^{20,21}\) since SecA can bind to SecYEG as a monomer or dimer\(^{2,22,23}\) and appears to monomerize upon binding to lipids\(^{21}\) or to detergent-solubilized SecYEG.\(^{24}\)

SecA is large (102 kDa) and elongated.\(^ {25}\) Each protomer comprises four structural domains (see Results). Two of these comprise a RecA-like fold and associate to build a structure widespread in nucleic acid helicases.\(^ {20,27}\) Inserted in each of the SecA DEAD domain places SecA in the DExD/H helicase superfamily 2 (SF2).\(^ {28}\) Inserted in each of the SecA DEAD and derivatives thereof. Sequence homology demonstrates that SecA is a member of the vast body of mutagenic and biochemical data, which parts of SecA are involved remains unresolved.

To gain an insight into the molecular basis of SecA-mediated translocation and oligomerization we determined the structure of ecSecA at 2 Å resolution in the apo state and complexed with the nucleotides ADP, ATP and its non-hydrolyzable analogue AMP-PNP. Our data reveal that: (a) ecSecA crystallizes as an antiparallel dimer; (b) dimerization occurs exclusively between the DEAD motors; (c) the extreme N terminus of ecSecA does not participate in dimerization; (d) the DEAD motor and the two “specificity domains” can undergo significant rigid body motions; (e) several residues identified by genetic and biochemical studies to be important for either preprotein translocation or catalysis in nucleic acid helicases participate in or are located in the immediate vicinity of the ecSecA dimerization interface.

### Results

#### Structure determination

Here, the recombinant ecSecA[9–861] protein was used. ecSecA[9–861] is practically indistinguishable from the wild-type SecA (901 residues) as far as its *in vitro* and *in vivo* activity and dimerization are concerned.\(^ {20}\) For simplicity ecSecA[9–861] will be referred hereinafter as ecSecA. The full-length protein only ever produced crystals that diffracted to less than 8 Å resolution. These results and the steps taken to obtain suitable crystals for structure analysis of ecSecA, will be reported elsewhere (Y. P., unpublished results).

The structure of ecSecA was determined in the apoprotein and nucleotide-bound states from crystals soaked with ATP, ADP.Mg\(^ {2+}\) and AMP-PNP. Mg\(^ {2+}\). In all cases, two protein molecules were identified in the asymmetric unit. The structure of ecSecA apoprotein was built into electron density maps calculated to 2.0 Å. The structures of the ATP, ADP and AMP-PNP complexes were determined to 2.2 Å, 2.1 Å and 2.0 Å resolution, respectively. The

### Table 1. Data collection and processing statistics for the synchrotron data sets

<table>
<thead>
<tr>
<th>ecSecA (Se-Met)</th>
<th>ecSecA</th>
<th>ecSecA:ATP (Se-Met)</th>
<th>ecSecA:AMP-PNP</th>
<th>ecSecA:ADP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit cell parameters (Å, °)</td>
<td>a = 75.6</td>
<td>a = 75.0</td>
<td>a = 75.4</td>
<td>a = 75.0</td>
</tr>
<tr>
<td>b = 89.4</td>
<td>b = 89.2</td>
<td>b = 89.5</td>
<td>b = 90.0</td>
<td>b = 90.4</td>
</tr>
<tr>
<td>c = 163.7</td>
<td>c = 163.0</td>
<td>c = 163.4</td>
<td>c = 163.0</td>
<td>c = 163.3</td>
</tr>
<tr>
<td>β = 101.2</td>
<td>β = 100.5</td>
<td>β = 100.7</td>
<td>β = 100.5</td>
<td>β = 100.8</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.97550</td>
<td>0.93930</td>
<td>0.97889</td>
<td>0.97620</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>20–2.4(2.53–2.4)</td>
<td>20–2.0(2.10–2.0)</td>
<td>20–2.2(2.32–2.2)</td>
<td>19.9–2.0(2.10–2.0)</td>
</tr>
<tr>
<td>No. of observations</td>
<td>307,966(45,040)</td>
<td>448,157(32,034)</td>
<td>371,706(49,335)</td>
<td>297,234(20,883)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.8(100.0)</td>
<td>97.1(86.7)</td>
<td>96.9(92.0)</td>
<td>90.5(61.5)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>3.5(3.06)</td>
<td>3.2(2.3)</td>
<td>3.5(3.4)</td>
<td>2.3(1.7)</td>
</tr>
<tr>
<td>l/&gt;≤l&lt;/&gt;</td>
<td>13.4(2.3)</td>
<td>12.5(2.3)</td>
<td>8.5(1.1)</td>
<td>11.1(2.3)</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;</td>
<td>0.056</td>
<td>0.077</td>
<td>0.081</td>
<td>0.061</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;</td>
<td>0.072(0.392)</td>
<td>0.042(0.352)</td>
<td>0.077(0.644)</td>
<td>0.051(0.331)</td>
</tr>
</tbody>
</table>

Values in parentheses refer to the highest resolution shell. Se-Met, selenomethionylated.

\(^{a}\) R<sub>merge</sub> = \[ \sum | l<–l> | \sum | l<–l> \], where l<+> and l<–> are the integrated intensities of Bijvoet pairs of reflections.

\(^{b}\) R<sub>merge</sub> = \[ I<–I> \sum I<–I> \], where I is the integrated intensity of a given reflection, I<+> and I<–> are scaled separately.
structure was solved using multiple and single wavelength anomalous diffraction data collected from crystals of selenomethionylated protein (Table 1). There was insufficient electron density to confidently build some residues at the amino and carboxy termini and most of the preprotein binding domain (PBD; residues 221–376; Supplementary Data, Figure S1). The missing residues from the final refined model are the chain \( \alpha \) residues: 9–12, 229–364 and 835–861 and the chain \( \beta \) residues: 9–11, 233–279, 314–364 and 833–861. The determined structure of \( \varepsilon \)SecA represents \( \sim 84\% \) of the crystallized protein. Nevertheless, the experimentally determined part of the PBD structure was enough to determine its overall orientation. Modelling of the unresolved parts of \( \varepsilon \)SecA PBD was based on \( \beta \)SecA(1TF5)-PBD\(^6\) as described in Materials and Methods The modeled PBD contains in addition helices \( \alpha 10 \), \( \alpha 13 \) and strands \( \beta 7–\beta 11 \) (Supplementary Data, Figure S1).

Structure of the \( \varepsilon \)SecA protomer

\( \varepsilon \)SecA consists of 29 \( \alpha \)-helices and 23 \( \beta \)-strands organized into four structural domains (Figure 1 Supplementary Data, Figure S1). Two of these, termed the nucleotide binding domain (NBD; residues 1–220 and 377–416; dark blue) and the intramolecular regulator of ATPase 2 (IRA2; residues 417–621; light blue), are \( \alpha / \beta \) domains that assemble to form an SF2 DEAD motor.

NBD consists of a \( \beta \)-sheet made from seven parallel \( \beta \)-strands (strands \( \beta 1–\beta 5 \) and \( \beta 13–\beta 14 \); Supplementary Data, Figure S1A) that is surrounded by 11 \( \alpha \)-helices (\( \alpha 1–\alpha 9 \) and \( \alpha 14–\alpha 15 \)). PBD is inserted between helices \( \alpha 9 \) and \( \alpha 14 \) of NBD without disturbing the structural integrity of the RecA fold. PBD comprises characteristic Stem and Bulb sub-structures. The Stem is formed by two antiparallel \( \beta \)-strands (\( \beta 6 \) and \( \beta 12 \); Supplementary Data, Figure S1) and inserts in NBD. The Bulb comprises two lobes (Bulb1, extending to the end of helix \( \alpha 12 \) and Bulb2). In some helicases, e.g. UvrB,\(^32\) an unrelated “specificity domain” is inserted in exactly the same location, while in others (e.g. the Drosophila development protein Vasa)\(^33\) the corresponding \( \alpha 9 \) and \( \alpha 14 \) are fused. The position and length of secondary structural elements in different PBDs varies (Supplementary Data, Figure S1B).

IRA2 is composed of a central \( \beta \)-sheet made from seven parallel \( \beta \)-strands (\( \beta 15–\beta 19 \) and \( \beta 21–\beta 22 \); Supplementary Data, Figure S1A) and an antiparallel \( \beta \)-strand (\( \beta 20 \)) and, is surrounded by seven \( \alpha \)-helices (\( \alpha 16–\alpha 22 \)). IRA2 contains two characteristic sub-structures: the variable region (VAR; a hair pin of helices \( \alpha 19–\alpha 20 \) comprising residues Gly516 to Gly553; Supplementary Data, Figure S1A, red) that shows significant length and sequence variability in different SecA proteins (Figure 2) and the “Joint” comprising residues 591–621 (Supplementary Data, Figure S1A, grey) that connects IRA2 to the C-domain (see below). A similar, albeit differently oriented, structure is also seen in other helicases such as Vasa (Figure 3).\(^33\)

\( \varepsilon \)SecA has a carboxy-domain (C-domain)\(^34\) which consists of residues 622–901 (Figure 1b, green) and can be sub-divided into four sub-domains (Figure 1a; Supplementary Data, Figure S1A): the scaffold domain (SD) that comprises the longest helix (\( \alpha 23 \)) of the structure (residues 622–668); the wing domain (WD; residues 669–755) that comprises four helices (\( \alpha 24–\alpha 27 \)) and one short \( \beta \)-strand (\( \beta 23 \)); the IRA1 domain (residues 756–829) that comprises helices \( \alpha 28 \) and \( \alpha 29 \) connected by an extended loop; the extreme C-terminal domain (CTD) (residues 830–901) of which only residues 830–834 can be modelled (Figure 1a, orange).

The PBD Stem crosses loosely over SD/IRA1 bundle, forming only minor contacts. In contrast, significant contacts are made by helix \( \alpha 12 \) from the Bulb, whose N-terminal residues bind the concave surface underneath the tip of the IRA1 hairpin.
IRA1 residues that are important for SecA catalysis\textsuperscript{35} line the α\textsubscript{12} binding site (Figure 4).

Average residue temperature factors (Supplementary Data, Figure S2C) reveal that the most ordered domains of \textit{ec}SecA is NBD followed by IRA2, while the two specificity domains (C-domain and PBD) show increasing flexibility.

Structure of the \textit{ec}SecA dimer

In the asymmetric unit of the \textit{ec}SecA crystals there is a molecular dimer (Figure 5) of approximate 2-fold symmetry (179.4°). The area of the dimerization interface for \textit{ec}SecA is 3292 Å\textsuperscript{2} (7.9% of the monomer surface). The two antiparallel molecules of \textit{ec}SecA are virtually identical (rms deviation of 684 C\textalpha\ atoms 1.3 Å) with minor conformational differences (Supplementary Data, Figure S2A-B). Nonetheless, the relative flexibility between the two protomers differs (Supplementary Data, Figure S2C).

Dimerization is mediated exclusively by DEAD motor residues of the two juxtaposed protomers (Figure 6(a)). Most of the interfacial contacts are provided by the same residue from each protomer located in the NBD (N132, Y134, Q137, P159 and G160), IRA2 (T470, K471, H476, N477, F483, H484, N486 and A489) (Figure 6(b)) and its VAR region (Q520, A524, E527, N528 and K538) (Supplementary Data, Figure S4B). Additional contacts are provided only by protomer α (K475, W519 and P529) and protomer β (E141, M161, A525 and L526) residues. The contacts are primarily hydrophobic (Supplementary Data, Figure S3A) and further stabilised by a total of 15 hydrogen bonds (Supplementary Data, Figure S4 and Table S1). Residues N132, Y134, Q137 and E141 of helix α\textsubscript{6} and its preceding loop all belong to the helicase motif Ia of SecA (see below).

Additionally, residues F263–V265 of the PBD are in hydrophobic contact between the two protomers and could contribute to dimerization. In the dimer, residues N132, Y134, Q137 and E141 are highlighted in yellow.

---

\textbf{Figure 3.} The \textit{Drosophila} development protein Vasa on the DEAD motor of \textit{bs}SecA. Top view of the superimposed DEAD-box helicase Vasa (PDB accession code; 2DB3, beige, residues 253–616)\textsuperscript{33} and \textit{B. subtilis} SecA DEAD motor (PDB accession code; 1M6N, NBD; dark blue, residues 59–218 and 357–395, IRA2; light blue, residues 396–567).\textsuperscript{5} \textit{bs}SecA was chosen because, like Vasa, its DEAD motor assumes a closed conformation. Beside the nine conserved SF2 motifs, two regions of conservation in some DEAD-box helicases are α\textsubscript{16} and α\textsubscript{22} corresponding to \textit{ec}SecA residues, K137 and P159, respectively (coloured on both structures). The respective helices near or at these motifs of \textit{ec}SecA/Vasa structures as well as helix α\textsubscript{6} of motif Ia are drawn in ribbon presentation. The N and C-terminal helices of SecA are indicated (\textit{ec}SecA numbering). The red surface shows the RNA heptamer 5′-p-(Up)\textsubscript{6}-U bound to Vasa.

\textbf{Figure 4.} IRA1, SD and helix α\textsubscript{12} of PBD of \textit{ec}SecA. The N-terminal residues of PBD Bulb helix α\textsubscript{12} bind to the concave surface underneath the tip of the IRA1 hairpin. Five IRA1 residues that are important for SecA catalysis\textsuperscript{35} and line the α12 binding site are labelled. In order to make I789 visible, only the surface of the backbone atoms of R792 is shown.
helix $\alpha^{18}$ from IRA2 of one protomer penetrates into the ATP cleft of the DEAD motor of the other protomer (Figure 6(b); Supplementary Data, Figure S1A). $\alpha^{18}$ lies $\sim 10$ Å from the nucleotide, at the opposite side of the cleft (Supplementary Data, Figure S3C), and hence it does not sterically interfere with nucleotide binding.

Structure of ecSecA in complex with nucleotides

Nucleotides were soaked into the protein crystals and a single nucleotide molecule was observed in each protomer (Figure 7(a)–(d)), corroborating previous evidence that the NBD and IRA2 form a mononucleotide binding cleft in common with other helicases. Interactions of the bound adenine nucleotides with the DEAD motor (Figure 7(b)–(d)) are almost exclusively located on the NBD. The nucleotide binding sites are solvent-exposed on the outer rim of the dimer. The adenine ring is sandwiched between two hydrophobic residues of the Q motif (M81, F84) and additional polar contacts with Q motif residues (R82 and Q87; Supplementary Data, Figures S5 and S6), ensure a tight association. These interactions suggest that the Q motif is of importance in SecA catalysis, as shown in other SF2 RNA helicases.

Several NBD residues in motif I (T104, G105, E106, G107, K108, T109 and L110) are involved in direct contact with the nucleotides (Figure 7(b)–(d); Supplementary Data, Figures S5 and S6). Additional residues from motifs Ia (R138, D139, N142 and N143) and II (D209) make water-mediated interactions with the nucleotide (Figure 7(b); Supplementary Data, Figures S5 and S6). Only one residue from motif V (R509) of IRA2 makes a direct contact (Figure 7(c)–(d); Supplementary Data, Figures S5 and S6) and this is the first structural evidence that R509 of the helicase motif V interacts directly with ATP, underlining its essential contribution to SecA ATP hydrolysis. Motif VI residues are not involved in nucleotide binding, making instead specific contacts (Supplementary Data, Table S2) to motif V residues located in IRA2 (e.g. R574 with M506, R577 with R509 and D512) and to motif II residues located in the NBD (R566 with D212). Two of the motif VI residues that are essential for catalysis, R574 being the presumed “arginine finger” residue.

Figure 5. Crystal structure of the ecSecA dimer. Surface of protomer $\alpha$ is coloured as for Figure 1(b). Protomer $\beta$ is shown uniformly coloured beige. Front (a) and top view (b) of the dimer. The direction of the near 2-fold axis relating the protomers is clearly marked.

Figure 6. Map of interfacial residues. (a) Close up view of the dimerization interface. For simplicity, the ribbon diagram shows only the DEAD motors of protomers $\alpha$ and $\beta$. Colouring is as for Figure 5. The approximate direction of the near 2-fold axis is clearly marked. (b) Dimerization interface as in (a) with the IRA2-VAR sub-domain removed. Residues involved in contacts are shown labelled in the $\alpha$-protomer. The ordered solvent molecules participating in inter-subunit interactions are shown as violet dots with green broken lines indicating the polar interactions.
and R577 face away from the nucleotide, their side-chains being rotated by at least 45°. We must therefore assume that the \textit{ec} SecA in the crystal structure is catalytically inactive.

While soaking of nucleotides into the protein crystals does not result in major structural changes (rms deviation of 2716 main-chain atoms $\sim 0.5$ Å) (Supplementary Data, Table S3), there are distinct conformational changes in the side-chains of residues that line the nucleotide cleft. For example, R509, which contacts E210 in the apoprotein, reorients to interact with ADP or AMP-PNP in both protomers, but only with ATP in protomer $\beta$ (Supplementary Data, Figures S5 and S6). Moreover, R509 in protomer $\alpha$, its neighbours in the motif V loop (G508 and G510) and residues in VAR (A531 and E532) show a significant conformational difference (rms deviation of C$^\alpha$ atoms $\sim 1$ Å) between the ATP and ADP-bound structures. In protomer $\beta$, R509 does not undergo a significant structural change but residues in VAR (Q520, A521, A524, A525 and A531) do (rms deviation of C$^\alpha$ atoms $> 1$ Å). Moreover, some motif VI residues also become rearranged in the presence of ATP (Supplementary Data, Table S2), forming contacts with motif I (Q578 with E106) and motif III (N569 with T393).

**Comparison of \textit{ec} SecA with the other SecA structures**

Superimposition of \textit{ec} SecA with \textit{bs} SecA (PDB codes 1M6N and 1TF5) and mtSecA (PDB code 1NL3) indicates that they share a similar overall fold (Figure 8(a)–(c); Supplementary Data, Table S4). Significant sequence and structural variation is only seen in the VAR regions (Figures 2 and 8).

A comparison of the four structures reveals the remarkable repertoire of large-scale and localized domain motions possible, mainly involving the DEAD motor and specificity domains. The most...
extensive conformational difference involves a rigid body movement of the PBD. In the bsSecA(1M6N) and mtSecA structures the PBD is almost attached to WD via its Bulb1 lobe (Figure 9(a)). In the ecSecA and bsSecA(1TF5) structures the PBD Bulb swivels around its Stem by ~90°, such that Bulb1 moves away from WD to interact, from below, with the IRA1 sub-domain (Figures 4 and 9(a)). This motion creates a considerable solvent-accessible space between the PBD and WD. It is clear that this rigid body movement is accomplished by some significant secondary structure rearrangements. For example, helix α12 of Bulb1 undergoes a significant reduction in secondary structure and shortens by six residues (Figure 4; Supplementary Data, Figure S1B).

Another striking difference is that the DEAD motor is more “open” in ecSecA, allowing the second protomer to dock in an antiparallel fashion (see below). This occurs because IRA2 undergoes a significant lateral rigid body movement away from the NBD (Figure 9(b)) and was not observed in the other SecA structures. Nevertheless, IRA2 is not completely detached from the NBD and remains bound through a salt-bridge between D217 and R566. The rigid body motion of IRA2 away from NBD in a “clam-shell” fashion that mediates “opening” of the DEAD motor is possible because of the linker connecting the NBD to IRA2 (residues 413–420; Supplementary Data, Figure S1A). This linker was shown to be very flexible.30,34,38 Lastly, there is significant variation in the orientation of the SD and WD sub-domains between ecSecA and the other SecAs (Figure 8).

While the protomer folds are similar, the quaternary organization of the four SecA structures is widely different. These differences are readily visualized when the most conserved and rigid NBD domain of the ecSecA protomer α is superimposed pairwise with the corresponding domain from the other structures. This analysis reveals that protomer β (Figure 10, pink) assumes a completely different orientation in the three structures. The bsSecA (PDB code 1TF5) structure exists as a monomer in the crystals and was proposed to represent a monomeric form.6 ecSecA makes use of its VAR sub-structure in dimer formation. The VAR region is essentially missing in bsSecA (Figure 2) and the mtSecA DEAD motor assumes a “closed” conformation (with the VAR region coming close to the NBD), prohibiting the monomers from dimersing in a fashion similar to the ecSecA structure. bsSecA (PDB code 1M6N) crystallized as a monomer in the asymmetric unit and was proposed to form a dimer with a symmetry related molecule (Figure 10(a)). This proposed dimerization interface is mediated through associations between the extreme N-terminal NBD residues (G3, I4 and L5 of bsSecA) and the C-domain IRA1 hairpin of the symmetry-related chain, and buries 5432 Å² or 14.7% of the mono-

**Figure 8.** Comparison of one ecSecA protomer with other SecA structures. Ribbon diagrams showing superimpositions based on the structural alignments of the NBD domains of ecSecA with the other known SecA structures. ecSecA is coloured as for Figure 1(b). The VAR sub-domains of the other SecAs are coloured yellow. (a) bsSecA (PDB code 1M6N) is coloured beige. (b) bsSecA (PDB code 1TF5) is coloured pink. (c) mtSecA (PDB code 1NL3) is coloured grey. mtSecA corresponds to SECA1_MYCTU, one of the two SecAs in this bacterium.
mer surface. The resulting crystal-contact dimer was proposed to also exist in solution and to be physiological. The mt SecA also exists as a dimer in the asymmetric unit with a buried surface area of 2739 Å². The dimer interface involves PBD, IRA2 and C-domain (Figure 10(b)).

Several PrlD mutations line the ecSecA dimer interface

Protein localization (PrlD) mutant SecAs can secrete preproteins with defective or deleted signal peptides. Most of the PrlD mutations lie within the dimerization interface of the ecSecA structure. prlD2,22,23 and azi-17 involve mutations of residue Y134 that participates in both polar and apolar interactions at the dimerization interface of ecSecA (Supplementary Data, Figure S3B and Table S1), or NBD-IRA2 contacts in the other SecA structures (Figure 9(b)–(c)). prlD43 affects H484 and prlD2,3
mutated alteration A488. These residues participate in apolar interactions between the protomers in ecSecA (Supplementary Data, Figure S3B and Table S1) or are involved in NBD-IRA2 contacts in the other SecA structures (Figure 9(c)). Interestingly, the helicase motif V (Supplementary Data, Figure S1A) that is essential for catalysis lies directly behind H484 and A488 near the dimerization interface of ecSecA (Figure 9(c)) and A507 of motif V is a hotspot for prlD mutations.

Discussion

We present the high resolution structure of SecA, the dimeric prepore translocase motor, from E. coli. The overall protein fold and domain assembly of the monomer is basically conserved among the previously determined structures of SecA proteins, as expected from their high sequence identity. The fundamental organization of the protein involves a core helicase DEAD motor, from which protrude the PBD and C-domain or specificity domains. A single mononucleotide cleft is formed between the DEAD motor domains. A comparison of the SecA structures reveals that the DEAD motor and the two specificity domains participate in a repertoire of significant rigid body motions (Figures 8 and 9) that are discussed below.

The DEAD motor of ecSecA assumes a more "open" conformation (Figure 9(b)-(c)) when compared to the other SecA structures. When detached from the NBD, IRA2 can undergo a remarkable unfolding transition, allowing the release of ADP from the cleft. IRA2 restructuring, involving motifs V and VI, is believed to be necessary for ATP hydrolysis and presumably allows optimal positioning of IRA2 residues essential for catalysis. These include R509 of the helicase motif V, and R574/ R577 of motif VI. IRA2-like domains in SF2 helicases are very mobile and have even been crystallized in complete dissociation from the corresponding NBD-like domains. An important contributor to enhanced flexibility of IRA2 domains is the reduced number of α-helices that enfold the inherently unstable parallel β-sheet in their core. However, while complete detachment of IRA2-like domains may be an essential adaptation for the unwinding of nucleic acids, the IRA2 mobility is clearly under tighter control in SecA. Thus, SecA is unique among SF2 proteins in having acquired the C-domain that crosses the DEAD motor longitudinally from below and latches onto both IRA2 and NBD through the SD helix (Figure 1(a)). This effectively allows SD to act as a molecular staple that controls the IRA2 detachment from the NBD via the Joint region. In other helicases, the role of a weaker and transient molecular staple may be played by the nucleic acid, clearly seen in Vasa, to bind to both DEAD motor domains (Figure 3). Additional factors may act directly on IRA2 to regulate its structure and mobility, as is the case with eIF4G acting on eIF4A. Obviously, tightly regulated cycles of IRA2 detachment and rebinding to the NBD are a common and fundamental feature of SF2 catalysis.

IRA2 motions and nucleotide occupancy of the DEAD motor are transmitted to the PBD, causing both conformational and dynamical residue changes. The position of the PBD Stem region is similar in all four SecA structures, in contrast with the Bulb moiety that occupies distinctly altered states (Figure 9(a)), of presumably similar energy values. The "open" PBD state in bsSecA (PDB code 1TF5) is postulated to be coincident with the active monomeric state of SecA. However, our observation of an identical conformational state in the PBD of dimeric ecSecA excludes the possibility that this PBD conformation is related to monomerization. PBD conformation is probably dependent on crystallization conditions because it can adopt either the open or closed conformation in the crystal lattice without the presence of any translocation ligands. Nevertheless, we anticipate that the open PBD state is physiologically relevant, since SecA derivatives with mutations of highly conserved IRA1 residues in the binding region to open PBD (Figure 4) are non-functional. It is well known that the PBD is prone to conformational alterations that are regulated by the preprotein and nucleotide. These PBD motions, i.e. swivelling around its Stem that leads to opening and closing of the PBD–WD interface, could provide a cog-wheel mechanism that pushes extended preprotein segments into the membrane and the two adjacent PBVs in the dimer would allow a hand-over-hand mechanism for such a preprotein chain relay. The alternative possibility is that a PBD movement allows for the optimal docking of SecA to SecY where the opening between WD and PBD is wide enough (~16 Å) to accommodate SecY preprotein secondary structure elements in the open state, thereby facilitating SecA docking.

ecSecA forms a dimer exclusively via DEAD motor residues and is distinct from the dimers of bsSecA and mtSecA (Figure 10). We anticipate that the ecSecA dimer is physiologically relevant, since it is in agreement with extensive hydrodynamic analyses that reveal the propensity of the isolated SecAΔC to dimerize and tetramerize. More importantly, the mode of ecSecA dimerization is consistent with the front-to-front SecYEG dimer. The dimensions of this SecYEG dimer are in good agreement with the dimensions of ecSecA DEAD motor, the region of ecSecA that binds to SecYEG. Only the SecA dimer presented here places both DEAD motor domains on the same side of the structure for SecYEG binding. Another exciting possibility is offered by the ecSecA dimer and the differences observed between nucleotide interactions in the two protomers. This structural arrangement actually allows a direct nucleotide-regulated allosteric communication between the two protomers. Perhaps this could explain the existence of a second, kinetically determined but structurally elusive, low-affinity nucleotide binding site in SecA.
Extreme N-terminal residues were proposed to be important for eeSecA function and dimerization. However, these residues are clearly important for eeSecA dimerization interface and can be removed in eeSecA without affecting dimerization or function in vitro and in vivo. Similar N-terminal truncations also result in functional eeSecA proteins. Several SecA sequences are naturally devoid of these residues.

The mode of SecA oligomerization still remains a conundrum. All the available SecA structures have a dimerization interface of conserved in SecA. If preproteins and RNA bind to the adjacent interface, we can remove in vitro without affecting dimerization or function. However, these residues are clearly important for eeSecA function and dimerization and can be removed in vitro without affecting dimerization or function. Similar N-terminal truncations also result in functional eeSecA proteins. Several SecA sequences are naturally devoid of these residues.

It has not escaped our attention that the eeSecA structure allows a significant body of genetic and biochemical evidence to now be rationalized. Several of the isolated SecA PrlD mutants, whose molecular basis remains elusive, map to the dimerization interface of eeSecA. Some residues altered by PrlD mutations (e.g., Y134 from NBD and H484 and A488 from α18 of IRA2) are closely associated when the DEAD motor is in the closed state (e.g., bsSecA and mtSecA; Figure 9(c)). These residues dissociate and form new correct interfaces in the eeSecA dimer (Figures 6(b), Supplementary Data, Figures S3 and S4). Obviously, PrlD mutations could either pry open the dimer interface by introducing slightly bulkier side-chains (e.g., A488V) or weaken it by rendering large residues shorter, e.g., Y134C/S/N. PrlD mutations occurring at the dimer interface promote SecA monomerization, which would then allow the preprotein to be exported without a signal peptide. Preproteins alter the dimer–monomer equilibrium and PrlD mutations could mimic preprotein-driven “loosening” of the physiological SecA dimer during translocation, thus rendering functional signal peptides redundant. We consider two possibilities: (a) preproteins could act as allosteric modulators of the dimerization interface through their determined binding to the adjacent stem region of the PBD. (b) PrlD residues that affect dimerization may reveal a previously unsuspected pathway along the DEAD motor, where preproteins displace the interfacial helix, α18, thus possibly leading to monomerization. Regions in the Vasa helicase that correspond to the eeSecA dimerization interface (e.g., helix α6 and the β3/α7 connecting loop of NBD and helix α18 of IRA2) form an extended ridge that mediates RNA substrate binding (Figure 3). The binding surface is lined with 15 RNA binding residues, only three of which are conserved in SecA. If preproteins and RNA bind to similar surfaces this lack of conservation could be an adaptation that underpins substrate specificity.

Further understanding of how SecA catalyzes preprotein translocation and how it adapted nucleic acid helicase chemistry will require the determination of co-crystallized complexes with preproteins and translocase subunits as well as detailed in vitro dissection of the reaction using mutagenesis and biochemical assays. The structure presented here allows us to formulate testable hypotheses to achieve this dissection.

Materials and Methods

Bacterial strains and protein purification

Strains and DNA manipulation were as described. Transformed E. coli BL21/pLysS cells with plasmid pET3a-pMBB272 (N9-861) were grown in 2 l of LB growth medium at 37 °C. The culture was induced for expression of eeSecA at 30 °C with 0.3 mM IPTG when the absorbance A600=0.8. It was centrifuged after 5 h of induction yielding 3–8 g of cell paste. All following steps were performed at 23 °C unless mentioned otherwise. The 3 g of cell paste were thawed in 20 ml of buffer A (25 mM Tris–HCl (pH 7.6), 10% (v/v) glycerol, 50 mM NaCl). Then 1 mM phenyl-methyl-sulfonyl fluoride (PMSF) was added. The cells were disrupted by passing twice through a French press cell. In the resulting lysate 0.5 mM PMSF was added and centrifuged for 1 h at 18,000g at 4 °C. The supernatant was loaded onto an in-house prepared Cibacron-Blue Sepharose™ CL-6B (Pharmacia), equilibrated with buffer A. A linear gradient was applied from buffer A to 25 mM Tris–HCl (pH 7.6), 10% glycerol, 2.6 M KCl. eeSecA eluted at ~0.6 M KCl. The protein containing fractions were dialyzed 1:1 with water and loaded onto Hydroxyapatite™ (BioRad) equilibrated with buffer B (1 mM Tris–HCl (pH 7.6), 100 mM KCl). A linear gradient was applied from buffer B to 0.1 M potassium phosphate buffer (pH 6.8). eeSecA eluted at 20 mM potassium phosphate. The protein was dialysed against 50 mM potassium phosphate buffer (pH 7.5), 10% glycerol, concentrated with an Ultrafree-15 centrifugal filter device (Millipore) and loaded onto Sephacryl™ S-300 (Pharma- cia). Protein-containing fractions were passed once more through the latter column with 25 mM Tris–HCl (pH 7.6), 100 mM KCl. Protein purity was assessed by denaturing gel electrophoresis and protein amounts were determined by the Bradford photometric method. About 3 g of cell paste yields 20–38 mg of chemically homogeneous protein suitable for crystallization. The same purification protocol was applied for both native and selenomethionylated eeSecA.

Crystallization, data collection and structure determination

Crystals of both native and selenomethionylated eeSecA were obtained at 18 °C employing the hanging drop technique of vapour diffusion. Each hanging droplet contained a 6 μl aliquot of 25 mg/ml protein in water and 3 μl of reservoir solution (50 mM sodium citrate (pH 5.8), 6–9% (v/v) polyethylene glycol 35000, 6–10% glycerol and 50 mM ammonium sulphate) and was equilibrated against 1 ml of reservoir solution. The diffraction of the crystals was improved from 3.5 Å to 2 Å by resolution by the method of dehydration, with 2 M KCl, which reduced the crystal solvent content by ~9% (v/v). The resulting monoclinic crystals (space group P21) used for the structure determination, contained an estimated 56% (v/v) solvent corresponding to one molecular dimer per asymmetric unit. The molecular...
packing of these crystals is shown in Supplementary Data, Figure S7-A.B. The buried surface area between the closest dimeric neighbours varied from 2.1% to 5.6% of their solvent accessible surface. High-resolution native, multi-wavelength anomalous dispersion (MAD) and single wavelength anomalous dispersion (SAD) data were collected from single cryo-cooled crystals at the European Synchrotron Radiation Facility (ESRF) beamline ID14-4 using a Quantum 4R (ADSC) detector. The MAD phasing method was used with data processed to 2.4 Å resolution and Bijvoet differences due to Se atoms were maximized (Table 1). A combination of computational methods, including SHELXD, SOLVE and SHARP were used to locate, refine 58 of the 64 Se sites and phase calculations. An initial protein model was constructed using ARP/wARP and refined against the MAD (0.97939 Å wavelength) data using REFMAC as implemented in CCP4. The final models converged after several cycles of model building using XtalView with REFMAC5. The structures of the nucleotide complexes were refined with the respective data starting from the native coordinates. The results of all refinements are summarised in Table 2.

### Table 2. Refinement and stereochemical statistics

<table>
<thead>
<tr>
<th></th>
<th>ΔSecA</th>
<th>ΔSecA:ATP</th>
<th>ΔSecA:AMP-PNP</th>
<th>ΔSecA:ADP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>20–2.0(2.05–2.0)</td>
<td>19.6–2.2(2.26–2.2)</td>
<td>20–2.0(2.05–2.0)</td>
<td>19.9–2.1(2.16–2.1)</td>
</tr>
<tr>
<td>No. of protein Non-hydrogen atoms</td>
<td>11,203</td>
<td>11,316</td>
<td>11,235</td>
<td>11,448</td>
</tr>
<tr>
<td>No. of total unique reflections</td>
<td>140,049(10,078)</td>
<td>104,943(7205)</td>
<td>130,193(5843)</td>
<td>112,860(3313)</td>
</tr>
<tr>
<td>No. of ordered solvent molecules</td>
<td>434</td>
<td>431</td>
<td>431</td>
<td>434</td>
</tr>
<tr>
<td>Average coordinate error (Å)</td>
<td>0.16</td>
<td>0.21</td>
<td>0.16</td>
<td>0.20</td>
</tr>
<tr>
<td>Rms deviation of:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.029</td>
<td>0.034</td>
<td>0.026</td>
<td>0.030</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>2.21</td>
<td>2.40</td>
<td>2.07</td>
<td>2.28</td>
</tr>
<tr>
<td>Chiral volumes (Å³)</td>
<td>0.146</td>
<td>0.143</td>
<td>0.126</td>
<td>0.153</td>
</tr>
<tr>
<td>Average protein B-factor (Å²)</td>
<td>46.2</td>
<td>18.0</td>
<td>43.5</td>
<td>40.0</td>
</tr>
<tr>
<td>Most favoured regions (%)</td>
<td>90.4</td>
<td>89.9</td>
<td>90.8</td>
<td>89.5</td>
</tr>
<tr>
<td>Allowed regions (%)</td>
<td>6.8</td>
<td>7.9</td>
<td>6.8</td>
<td>7.5</td>
</tr>
<tr>
<td>Generously allowed regions (%)</td>
<td>1.3</td>
<td>1.2</td>
<td>1.2</td>
<td>1.8</td>
</tr>
<tr>
<td>Disallowed regions (%)</td>
<td>1.5</td>
<td>1.0</td>
<td>1.2</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Values in parentheses refer to the highest resolution shell.

* As defined by PROCHECK.

### Protein Data Bank accession codes

The coordinates of the reported structures have been deposited in the RCSB Protein Data Bank with accession codes: 2FSF, ΔSecA apoprotein; 2FSG, ΔSecA:ATP; 2FSH, ΔSecA:AMP-PNP; 2FSI, ΔSecA:ADP.

### Acknowledgements

We are grateful to L. Karamanou, G. Sianidis and C. Pozidis for initial purification protocols, biochemical and biophysical assays and molecular cloning; V. Bouriotis for resins; D. Dialektakis for fermentations; A. Kuhn, C. Kalodimos and B. Shilton for comments. Our research was supported by the European Union (TMR-ERBFMRXCT-960355, Biotech2-BIO4-CT98-0051, RTN1-1999-00149, QLK3-CT-2000-00082 and QLRF-2000-00122), Greek Secretariat of Research (AKMON) and Pfizer grants (to A.E.). Y.P. and K.P. thank the EMBL Grenoble Outstation, for providing support for measurements at the ESRF under the European Community-Research Infrastructure Action FP6 program.

### Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2006.12.049

### References


*Edited by J. Doudna*

(Received 20 October 2006; received in revised form 12 December 2006; accepted 17 December 2006) Available online 23 December 2006