Crystal structure of an mRNA-binding fragment of *Moorella thermoacetica* elongation factor SelB

M.Selmer\textsuperscript{1,2} and X.-D.Su\textsuperscript{1}

Department of Molecular Biophysics, Lund University, PO Box 124, S-221 00 Lund, Sweden
\textsuperscript{2}Present address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK
\textsuperscript{1}Corresponding authors
e-mail: maria@mrc-lmb.cam.ac.uk or xiao-dong.su@mphys.lu.se

SelB is an elongation factor needed for the co-translational incorporation of selenocysteine. Selenocysteine is coded by a UGA stop codon in combination with a specific downstream mRNA hairpin. In bacteria, the C-terminal part of SelB recognizes this hairpin, while the N-terminal part binds GTP and tRNA in analogy with elongation factor Tu (EF-Tu). We present the crystal structure of a C-terminal fragment of SelB (SelB-C) from *Moorella thermoacetica* at 2.12 Å resolution, solved by a combination of selenium and yttrium multiwavelength anomalous dispersion. This 264 amino acid fragment contains the entire C-terminal extension beginning after the EF-Tu-homologous domains. SelB-C consists of four similar winged-helix domains arranged into the shape of an L. This is the first example of winged-helix domains involved in RNA binding. The location of conserved basic amino acids, together with data from the literature, define the position of the mRNA-binding site. Steric requirements indicate that a conformational change may occur upon ribosome interaction. Structural observations and data in the literature suggest that this change happens upon mRNA binding.

**Keywords:** crystallography/elongation factor/
RNA binding/SelB/selenocysteine

Introduction

Selenocysteine (Sec), the 21st amino acid, exists in organisms from all three kingdoms. It is the main biological form of selenium, and is used in the active site of redox enzymes such as formate dehydrogenase and glutathione peroxidase. Selenocysteine is the only genetically coded non-standard amino acid. The selenocysteine codon consists of a UGA stop codon in combination with a specific hairpin structure on the mRNA. This selenocysteine insertion sequence (SECIS) allows the stop codon to be read through instead of promoting termination.

SelB is a specialized elongation factor for delivery of selenocysteylated tRNA\textsuperscript{Sec} to the ribosomal A-site when the two-component selenocysteine codon is present. It was first identified in *Escherichia coli* (Forchhammer et al., 1989). It forms a ternary complex with GTP and selenocysteylated tRNA\textsuperscript{Sec}, in analogy with elongation factor Tu (EF-Tu), but it also binds to the SECIS, which in bacteria is located nearby, downstream of the selenocysteine UGA codon. Thus, in addition to the requirement of correct codon–anticodon interaction, SelB needs to recognize a SECIS structure to deliver Sec-tRNA\textsuperscript{Sec} to the ribosomal A-site upon GTP hydrolysis.

The different tasks of SelB have been linked to different parts of the protein sequence (Kromayer et al., 1996). The N-terminal 342 amino acids display sequence homology to the three domains of EF-Tu, and contain the binding sites for tRNA and GTP. The C-terminal part is responsible for mRNA binding, and can be minimized to amino acids 472–634 (*E.coli* numbering) without losing binding affinity for the SECIS.

In *E.coli*, the determinants for RNA binding have been carefully dissected. SelB binds to the mRNA hairpin with a dissociation constant of ~1 nM (Thanbichler et al., 2000). The recognition element of RNA can be minimized to a 17 nucleotide hairpin (Figure 1A; Kromayer et al., 1996). Extensive mutagenesis studies have shown that all the loop nucleotides as well as the closing base pairs and the bulged U are required for selenocysteine insertion (Heider et al., 1992). Another important feature is the distance between the UGA codon and the minimal hairpin, which in the *E.coli* selenoprotein mRNAs is 11 nucleotides. The hairpin sequence is not very well conserved between different bacterial species (Figure 1).

Here we describe the crystal structure of the C-terminal fragment 370–634 of *Moorella thermoacetica* (previously called *Clostridium thermoaceticum*) SelB (SelB-C) determined to a resolution of 2.12 Å. This fragment contains the complete C-terminal extension as compared with EF-Tu. It is the first structure of a component of the selenocysteine insertion machinery. In the light of functional data from the literature and structural constraints imposed by the ribosome, the SelB-C structure allows us to propose a functional model for SelB in selenocysteine incorporation.

Results and discussion

**Crystal structure determination**

The preparation and crystallization of *M.thermoacetica* SelB-C will be published elsewhere (M.Selmer, R.Wilting, D.Holmlund and X.-D.Su, submitted). The crystals belong to space group *P*\textsubscript{2}1\textsubscript{2}1\textsubscript{2}1 and contain one molecule in the asymmetric unit. The structure was determined by multiwavelength anomalous dispersion (MAD) to 2.12 Å resolution using a single crystal of selenomethionine-substituted protein grown in the presence of yttrium chloride (Table I). Since both Se and Y were expected to be present in the crystal, wavelengths were chosen to optimize the anomalous signal from both elements. In total, six wavelengths were collected on the same crystal (Table I). Two yttrium sites and one selenium site could be found using the respective peak wavelength
data, and these sites subsequently were used for phasing. As seen in Table II, selenium made a slightly greater contribution to the phasing than did yttrium. Seventy-five percent of the residues were built into the experimental map. After phase combination, the rest of the chain could be traced. The structure was refined to an $R_{\text{work}}$ of 21.5% and $R_{\text{free}}$ of 25.8%, respectively, with good geometry (Table III). Two side chains are in the disallowed region of the Ramachandran diagram, Ala499 in the only poorly defined loop and Arg530 in a well-defined region of the map. The final model includes residues 380–634, one sulfate ion, two yttrium ions and 109 water molecules. The final $2F_o - F_c$ map around the sulfate ion is shown in Figure 2. Both yttrium ions are coordinated to residues from two neighbouring molecules, contributing to the crystal packing and explaining why the yttrium chloride additive was crucial for obtaining high quality crystals.

**Overall structure**

The SelB-C structure is L-shaped and has the overall dimensions $70 \times 50 \times 20$ Å (Figure 3). Each arm of the L consists of two globular winged-helix domains. The N-terminal part of SelB presumably consists of three domains (I–III) in analogy to EF-Tu. The SelB-C domains are therefore numbered IV–VII. Domain IV consists of amino acids Gly377–Ser436, domain V of Thr437–Phe511, domain VI of Ser512–His573 and domain VII of Arg574–Asn634.

The winged-helix domain is an $\alpha/\beta$ structure consisting of three $\alpha$-helices and a twisted three-stranded antiparallel $\beta$-sheet. The connectivity is $\alpha_1-\beta_1-\alpha_2-\beta_2-\beta_3$, and the secondary structure elements are indicated in Figure 4. The domains are arranged in a consecutive manner so that domains IV and V as well as domains VI and VII have approximately the same orientation. In the surfaces between domains IV and V and between domains VI and VII, the $\beta$-hairpin from one domain packs against the third $\alpha$-helix of the following domain. In the interface between domains V and VI, the turn between H4 and S4 in domain V contacts H9 in domain VI as well as the hinge between domains V and VI (between S6 and H7).

**Comparison of domains**

Structural alignment of domains IV–V with domains VI–VII using TOP (Lu, 2000) gives a root mean square deviation (r.m.s.d.) of 1.5 Å for 98 residues. These 98 residues (392–426, 433–454, 463–471, 473–476 and 478–505) include all secondary structure elements except the N-terminal helix. The sequence identity over this stretch is 11% and the sequence similarity 32%. Multiple structural alignment of all four domains using the MAPS program (http://bioinfo1.imbys.lu.se/TOP/maps.html) results in the superpositioning shown in Figure 5A. The domains are remarkably similar, the major differences

![Fig. 1. (A) Escherichia coli fdhF mRNA hairpin (Hüttenhofer et al., 1996). Essential nucleotides are indicated by shading. (B) Predicted mRNA hairpin from the *M.thermoacetica* fdhA gene (GenBank accession No. U73807). Nucleotides identical to the essential nucleotides in *E.coli* are indicated by shading. (C) Predicted hairpin from the *Aquifex aeolicus* SelD gene (GenBank accession No. NC_000918). Shading as in (B). Hairpin structures were predicted using the mfold program (Zuker et al., 1999).](image)

<table>
<thead>
<tr>
<th>Table I. Data collection statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>0.7250 ($\lambda_1$)</td>
</tr>
<tr>
<td>0.7255 ($\lambda_2$)</td>
</tr>
<tr>
<td>0.9773 ($\lambda_3$)</td>
</tr>
<tr>
<td>0.9783 ($\lambda_4$)</td>
</tr>
<tr>
<td>0.9840 ($\lambda_5$)</td>
</tr>
<tr>
<td>0.9905 ($\lambda_6$)</td>
</tr>
</tbody>
</table>

$r_{\text{sym}} = (\Sigma I - <I>/<\Sigma I>$, where $I$ is the observed intensity and $<I>$ is the average intensity of symmetry-related reflections.

<table>
<thead>
<tr>
<th>Table II. Phasing statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ytrrium phasing</td>
</tr>
<tr>
<td>peak ($\lambda_1$)</td>
</tr>
<tr>
<td>$R_{\text{Calf}}$</td>
</tr>
<tr>
<td>Phasing power</td>
</tr>
</tbody>
</table>

All values are for centric/acentric phasing.

$R_{\text{Calf}} = (\Sigma |F_{\text{calc}}| - |F_{\text{calc}}|)/|F_{\text{calc}}|$, where $F_{\text{calc}}$ is the calculated heavy-atom structure factor.

Phasing power $= |F_{\text{calc}}|/E$, where E is the estimated lack-of-closure error.
Table III. Refinement statistics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>19.2–2.12</td>
</tr>
<tr>
<td>$R_{work}$ (%)</td>
<td>21.5 (24.7)</td>
</tr>
<tr>
<td>$R_{free}$ (%)</td>
<td>25.8 (28.4)</td>
</tr>
<tr>
<td>No. of reflections</td>
<td></td>
</tr>
<tr>
<td>Working set</td>
<td>25,755</td>
</tr>
<tr>
<td>Test set</td>
<td>2,755</td>
</tr>
<tr>
<td>No. of atoms</td>
<td>2,073</td>
</tr>
<tr>
<td>Solvent</td>
<td>109</td>
</tr>
<tr>
<td>Hetero</td>
<td>7</td>
</tr>
<tr>
<td>Deviations from ideal geometry (r.m.s.d.)</td>
<td></td>
</tr>
<tr>
<td>Bond lengths</td>
<td>0.006 Å</td>
</tr>
<tr>
<td>Bond angles</td>
<td>1.2°</td>
</tr>
<tr>
<td>Dihedral angles</td>
<td>20.7°</td>
</tr>
<tr>
<td>$B$-factors (Å$^2$)</td>
<td></td>
</tr>
<tr>
<td>Minimum/maximum</td>
<td>12.5/53.2 Å$^2$</td>
</tr>
<tr>
<td>Average</td>
<td>29.9 Å$^2$</td>
</tr>
<tr>
<td>Ramachandran plot</td>
<td></td>
</tr>
<tr>
<td>Most favoured region</td>
<td>92.3%</td>
</tr>
<tr>
<td>Additionally allowed</td>
<td>6.4%</td>
</tr>
<tr>
<td>Generously allowed</td>
<td>0.4%</td>
</tr>
<tr>
<td>Disallowed</td>
<td>0.9%</td>
</tr>
</tbody>
</table>

$R_{work} = \sum |F_{o}\| - |F_{e}\|/\sum |F_{o}|$, where $F_{o}$ and $F_{e}$ are calculated and observed structure factors; $R_{free} = R_{work}$ for a 10% subset of the reflections that were not included in the refinement.

Numbers in parentheses represent the value for the highest resolution shell.

![Fig. 2. Section of the final 2$F_o$ – $F_e$ map showing the coordination of the sulfate ion by Arg599, Ser605 and Arg606, prepared using O (Jones et al., 1991).](image)

being the length of the helices corresponding to H1 and of the β-hairpin corresponding to S2–S3. Domain IV is structurally most similar to domain VI, while domain V is most similar to domain VII. There are three equivalent structural segments present in all four domains (Figure 5B). These segments are (with numbering from domain IV): amino acids 397–401 (central part of H2), 405–426 (H3–S2) and 433–436 (S3). Within these 31 residues from each of the four domains that superimpose with an r.m.s.d. of 1.6 Å, only two leucine residues in H3 are conserved throughout the four domains. In the hydrophobic core of the domain, Leu413 packs against H1, and Leu414 packs against H2 and S3. The low sequence homology between the four domains explains why their common fold was not detected earlier. It also shows that this is a fold where a large sequence diversity can be tolerated.

**Sequence conservation of SelB**

A sequence alignment of the 12 bacterial SelB sequences present in the databases is shown in Figure 4. The rather low sequence similarity between SelB from different bacterial species may reflect the same kind of divergence as the low sequence similarity between the four domains. The conserved amino acids are mapped to the structure in Figure 6A and B. The conserved exposed residues are localized in two regions of the protein. Most of the conserved residues are located in the C-terminal part of domain VII, which is rich in basic amino acids, and 11 out of 24 residues are conserved. In domain V, there is a small charged patch next to a small hydrophobic area (H5–H6).

**Structural comparison with other proteins**

The winged-helix motif is a subfamily within the helix–turn–helix family (HTH), which is used extensively in DNA-binding proteins (Gajiwala and Burley, 2000). In searches for structurally similar proteins with DALI (Holm and Sander, 1993) and TOP (Lu, 2000) using all four domains or two domains as search model, no hit is found that displays similarity for more than a single domain. Thus, the double winged-helix with the conserved packing between domains IV and V and domains VI and VII is not found elsewhere in the database. There is only one previous example in the literature of a protein containing two consecutive winged-helix domains, the replication initiator protein Rep54 (Komori et al., 1999). This protein has domain–domain packing with a pseudo...
2-fold axis between the domains instead of the translational repeat seen in SelB-C.

Searching with domain VII, the most similar proteins found by DALI are the transcriptional repressor Smbb (Cook et al., 1998) with an r.m.s.d. of 1.6 Å for 60 Cα-atoms, the Z-DNA-binding domain of double-stranded RNA-specific adenosine deaminase (Schwartz et al., 1999) with an r.m.s.d. of 2.0 Å for 58 residues, and human replication protein A (Mer et al., 2000) with an r.m.s.d. of 1.7 Å for 60 residues. There is a large number of winged-

Fig. 4. Sequence alignment of the bacterial SelB sequences present in the Swissprot, Trembl and GenBank databases. White, light grey and dark grey boxes indicate increasing degrees of conservation. The positions of compensatory mutations in response to loop mutations are shown by filled arrows (Kromayer et al., 1999). The positions of compensatory mutations in response to bulge mutations are shown by open arrows (Li et al., 2000). Abbreviations of species names (accession Nos): MOOTH, Moorella thermoacetica (Swissprot Q46455); ECOLI, Escherichia coli (Swissprot P14081); SALEN, Salmonella enterica serovar typhi (GenBank NP_458244); YERPE, Yersinia pestis (GenBank 16120353); HAENI, Haemophilus influenzae (Swissprot P43927); PASMU, Pasteurella multocida (Trembl Q0CK68); PSEAE, Pseudomonas aeruginosa (Trembl Q9HV02); RHIHE, Rhizobium meliloti (Trembl Q031DS); DESBA, Desulfdomonas baciliform (Swissprot Q46497); EUBAC, Eubacterium acidaminophilum (Trembl Q93K2); AQUAE, Aquifex aeolicus (Trembl O67141); CAMIE, Campylobacter jejuni (Trembl Q9PMS1).
**Fig. 5.** Superpositioning of domains IV–VII using MAPS (http://bioinfo1.mbfys.lu.se/TOP/maps.html). (A) Ribbon model. Colour coding as in Figure 3. The image was prepared using Molmol (Koradi et al., 1996). (B) Aligned sequences for the equivalent structural segments found in all four domains.

**Fig. 6.** (A) Location of conserved residues in SelB-C (see Figure 5). SelB-C is seen from below domains VI–VII, in parallel with helix H12, which is surrounded by conserved residues. (B) A 90° rotated view. (C) Positions of single compensatory mutations isolated in response to loop changes in the mRNA hairpin (Kromayer et al., 1999). (D) Positions of single compensatory mutations isolated in response to bulge changes in the mRNA hairpin (Li et al., 2000). The image was prepared using Rasmol (Bernstein, 2000).
helix proteins where >50 residues match, including many DNA-binding proteins but no RNA-binding protein. However, in the large family of HTH proteins, there are examples of RNA-binding proteins.

**Location of the RNA-binding site**

By deletion mutagenesis, it was shown that the C-terminal fragment of 163 amino acids of *E.coli* SelB binds to the mRNA hairpin with the same affinity as intact SelB (Kromayer *et al.*, 1996). This was the smallest construct that was assayed for RNA binding. In *M.tHERmoaCetica* SelB, this fragment corresponds to domains VI–VII with 17 extra N-terminal amino acids. The short N-terminal tail will most probably be unstructured, and domains VI–VII thus constitute the mRNA-binding part of SelB.

The most conserved part of the SelB-C sequence is domain VII (Figures 4 and 6). There are five conserved basic residues (Arg599, Arg606, Lys607, Arg624 and Arg629) and one acidic residue (Asp617). These residues are all located on one side of the domain, defining a surface between helices H11 and H12 and the β-hairpin S11–S12 (Figure 6A and B). The size of this surface, which probably interacts with RNA, is ~20 × 10 Å along H12. Since the predicted mRNA hairpins are quite diverse in different bacterial species (Figure 1), we predict these conserved amino acids to make mostly backbone contacts with the RNA. Similarly to most of the contacts between ribosomal proteins and rRNA (Brodersen *et al.*, 2002), SelB may rely on surface shape and charge complementarity for RNA recognition.

**Sulfate ion suggests a mode of RNA binding**

In our structure, Arg599, Ser605 and Arg606 coordinate a sulfate ion (Figure 2). Both arginine residues are conserved. We believe that the sulfate mimics an RNA backbone phosphate and that this is a conserved phosphate-binding site. In SelB from other species, other residues can probably substitute for Ser605. Sulfate ions present in the crystallization buffer often go into phosphate-binding sites (Su *et al.*, 1994) since the two ions are similar. In this case, no sulfate was present intentionally in the crystallization, and the ion must have stayed bound to SelB-C when the sulfate-containing buffer was exchanged for storage buffer (M.Selmer, R.Wilting, D.Holmlund and X.-D.Su, submitted). This phosphate is probably located in the vicinity of the loop or the bulge of the hairpin structure, both of which have been shown by chemical probing to be in contact with SelB (Hüttenhofer *et al.*, 1996). Since the bulge is not conserved throughout SECIS sequences (Figure 1), our best guess is that a phosphate available for interaction would be located in the loop region, and that the loop will be directed towards the sulfate side of domain VII.

The yttrium ions are also bound in this region of the protein, and could mimic cations involved in RNA binding. The yttrium ions are coordinated with five oxygen atoms each, in magnesium-like geometry. Some of the amino acids involved in coordination of the yttrium ions (Asp600 and Asp627) are in close proximity to the conserved residues in domain VII.

**Mapping of mutational data on the structure**

RNA recognition by SelB also involves sequence-specific interactions. Mutations of the loop nucleotides are detrimental to binding (Heider *et al.*, 1992), while these nucleotides are not conserved in SECIS from many other bacteria (Figure 1). Also, protection from chemical probing has shown that the bulged U as well as the G3–U4 of the loop is involved directly in binding to *E.coli* SelB (Hüttenhofer *et al.*, 1996). Two different genetic studies looking for compensatory mutations in *E.coli* SelB when essential elements of the SECIS are mutated have been performed. When the bulged U (Li *et al.*, 2000) or the loop nucleotides (Kromayer *et al.*, 1999) are mutated, most compensatory single mutations are located in domain VII (Figures 4, 6C and D). One exception is E437K, which will be discussed below. None of the mutations involve conserved residues, and only two involve surface-exposed residues. The mutation A569V in *E.coli* (corresponding to Leu595 in *M.tHERmoaCetica*) compensates a C5A change in the loop rather specifically, and the mutation M556I (*M.tHERmoaCetica* Val583) compensates the same change, as well as a U to C change of the bulged nucleotide. Thus, out of these mutations, the only amino acid that could potentially involve a side chain contact with the mutated base is Ala569 in *E.coli* (*M.tHERmoaCetica* Leu595). The conclusion of mapping these results on our structure is that there are probably few sequence-specific contacts between protein and RNA.

Some of the mutations, e.g. V578A (*M.tHERmoaCetica* Ser604) and F572Y (*M.tHERmoaCetica* Ala598) make subtle changes to the hydrophobic core that may slightly change the orientation of the exposed amino acids contributing to the RNA-binding site, or change the stability of the hydrophobic core. They may adjust the shape complementarity to a mutated RNA fragment. Notably, several mutations are found to be the same in both these studies, despite the fact that different parts of the RNA hairpin are changed (Figure 4), suggesting that these changes make the binding site more tolerant or less specific, unless the different mutations in the RNA change the shape of the RNA similarly.

**Winged-helix interaction with RNA**

To our knowledge, there are no previous examples in the literature of winged-helix domains that bind RNA. However, there are protein–DNA complex structures available for a number of winged-helix proteins. In the canonical mode of DNA binding, the recognition helix, the third helix in the winged-helix domain, interacts with the major groove of a DNA double-helix (as in the structure of Sap1, Figure 7). These domains can also interact with DNA in different ways, but the recognition helix is always involved (Gajiwala and Burley, 2000). Helix H12 is the corresponding helix in domain VII of SelB (similar view in Figure 6A). This helix is surrounded by conserved basic residues, suggesting that SelB uses the same part of the winged-helix structure for RNA binding as is normally used for DNA binding. However, a SelB–SECIS complex structure is needed to clarify how similar these interactions are.
Steric requirements imposed by the ribosome

The distance between the UGA codon and the minimal SelB-binding site in the E.coli selenoprotein mRNAs is 11 nucleotides, and in the *M. thermoacética* *fdhA* mRNA, the corresponding distance is 12 nucleotides. In the recent structure of a 70S ribosome bound to mRNA, the length of the mRNA tunnel between the entrance and A-site codon is ~7–9 nucleotides (Yusupova et al., 2001). Thus, we envisage that SelB can stay bound to the recognition site on the mRNA until Sec-tRNA^Sec^ is delivered under GTP hydrolysis to the A-site.

When SelB delivers Sec-tRNA^Sec^ to the ribosomal A-site, the orientation of its N-terminal part or G-domain before GTP hydrolysis should be very similar to the corresponding state of EF-Tu. The kirromycin-stalled EF-Tu complex, which represents the location of the factor after codon recognition and GTP hydrolysis, but supposedly is similar to the GTP state, has been localized on the 70S subunit by single crystal cryo-electron microscopy (Stark et al., 1997). If the ternary complex structure (Nissen et al., 1995) is docked to the A-site of the 70S ribosome (Yusupov et al., 2001) in this position, the distance from the C-terminus of EF-Tu to the entrance of the mRNA tunnel described above is ~90–100 Å. This corresponds to the distance from the C-terminal end of the EF-Tu homologous part of SelB to the start of the mRNA hairpin when a selenocysteine codon is to be decoded. Thus, SelB-C and the hairpin together have to span a distance of 90–100 Å from domain III of SelB to the mRNA tunnel entrance.

The length of a 17 nucleotide RNA hairpin would be ~25–30 Å, as judged from other hairpin structures in the Protein Data Bank. Despite the substantial length of the RNA hairpin, SelB-C has to bridge most of the distance, since a direct contact between SelB and the bulged U (Figure 1A) has been shown by chemical footprinting (Hüttenhofer et al., 1996). Thus, the maximum contribution of RNA will be two single-stranded nucleotides and two base pairs of RNA spacer, or ~10 Å. On the protein side, the EF-Tu-homologous sequence is followed by a stretch of 2–5 basic residues, and a predicted helix before the start of domain IV. The diagonal of the SelB structure is ~75 Å. Thus, the L-shaped molecule may need to open up to bridge this 90–100 Å distance in complex with the RNA hairpin.

Indications of communication between the tRNA- and mRNA-binding sites of SelB

There are several previous indications of interdomain communication or conformational changes in SelB. In analogy with EF-Tu, SelB has a low intrinsic GTPase activity that is stimulated upon addition of ribosomes even when tRNA is absent (Parmeggiani and Sander, 1981; Hüttenhofer and Böck, 1998). In the presence of ribosomes, but still without tRNA, the addition of a 17 nucleotide SECIS stimulates the GTP hydrolysis of SelB further (Hüttenhofer and Böck, 1998), suggesting that mRNA binding to SelB-C leads to a more active SelB GTPase or a more favourable interaction with the ribosome.

Interdomain communication in the opposite direction, between the tRNA- and mRNA-binding sites, was shown in another study. Isolated domains VI–VII bind more tightly to SECIS than does intact SelB (dissociation constants of 0.14 and 1.26 nM, respectively). When Sec-tRNA^Sec^ is bound to SelB, the affinity for the SECIS increases to the same level as for the isolated mRNA-binding domains (Thanbichler et al., 2000).

Thus, the binding of Sec-tRNA^Sec^ to the N-terminal part and SECIS to the C-terminal part of SelB affect each other. One possibility is that the two parts interfere with each other’s function in the absence of RNA. The other possibility is that RNA binding induces conformational changes in one part that contribute favourably to the activity of the other part in terms of GTPase activity or mRNA binding, respectively.

In a SELEX (systematic evolution of ligands by exponential enrichment) study selecting RNA fragments that would bind to SelB (Klug et al., 1997), several sequences that bound with similar affinity to SelB, and as judged by chemical and enzymatic probing data in a similar binding mode, did not promote selenocysteine read through. There was no clear connection between affinity and function. Rather, a specific SelB-mRNA interaction seemed to be needed to trigger a conformational change necessary to achieve UGA read through, or to orientate SelB in a proper way for functional interaction with the ribosome. In agreement with this, it was shown that overproduction of SelB and the other components of the selenocysteine insertion machinery fails to induce any detectable selenocysteine incorporation in the absence of the proper hairpin structure on the mRNA (Suppmann et al., 1999). Thus, the function of the mRNA hairpin is more than increasing the local concentration of SelB in proximity to the stop codon. Some kind of conformational change seems to be induced by the proper SelB–SECIS interaction.

Structural indication of switch

What kind of conformational change can happen in SelB upon mRNA binding? The most likely site for a conformational change is the elbow of the L-shaped structure. It is possibly flexible, as indicated by the small contact area between domains V and VI. The surface area
that is buried in this interaction is 383 Å² per domain as compared with the contact areas between domains IV and VI, and VI and VII, where 655 Å² and 706 Å² of the domain area is buried, respectively (calculated using areaimol; CCP4, 1994).

In our structure, there is a salt bridge between Nε of Arg461 and a carboxyl oxygen atom of Glu552 across the cleft between domains V and VI. These two residues are not conserved throughout the bacterial sequences, but they display a pattern of co-variation so that the salt bridge is conserved. It can be seen in Figure 4 that when residue 460 or 461 is negatively charged, 551 or 552 is positively charged, and vice versa. In genetic studies (Kromayer et al., 1999), it has been shown that a mutation altering the charge of residue 461 (E437K in E.coli, corresponding to M.tHERMOACETICA Arg461) makes SelB less stringent, allowing near-perfect SECS in to work in vivo. This may suggest that mRNA binding leads to a hinge movement, which in turn may signal to the N-terminal part of SelB that mRNA is bound. A hinge movement would cost less energy when the salt bridge is absent (mutational data), and therefore a near-perfect mRNA–protein interaction could be tolerated to promote read through.

Functional model

Summarizing our interpretation of the SelB-C structure in the light of the available literature data, we can conclude with a functional model (Figure 8). As depicted in this figure, SelB binds to GTP and Sec-tRNA^Sec in solution. It binds further to the SECS element with subnanomolar affinity (Thanbichler et al., 2000), using mainly domain VII. This induces a conformational change of SelB (Klug et al., 1997; Hüttenerofer and Böck, 1998) that may involve an opening of the hinge between domains V and VI. This conformational change is indicated by the large distance that SelB-C has to bridge from the tRNA-binding domains to the mRNA hairpin, and by the effect of breaking a salt bridge across the interdomain cleft. When the selenocysteine UGA codon reaches the A-site of a translating ribosome, the quaternary complex of SelB with GTP and Sec-tRNA^Sec is in the proper conformation and position to achieve codon–anticodon interaction. When a correct codon–anticodon match occurs, GTP is hydrolysed and tRNA is released. In the absence of tRNA, SelB has lower, but still nanomolar, affinity for the mRNA hairpin (Thanbichler et al., 2000), and the combined translocation and unwinding of the mRNA may be needed to release it from SelB.

In conclusion, the structure of SelB-C shows that winged-helix domains can be used for RNA binding and that the mRNA-binding site is located in the part most distant from the tRNA-binding site. Furthermore, selenocysteine incorporation most probably involves a conformational change between domains V and VI.

Materials and methods

Protein preparation and crystallization

Selenomethionine-substituted SelB was prepared using the methionine pathway inhibition method (Van Duyne et al., 1993). Purification and crystallization were performed as described (M.Selmer, R.Wütling, D.Holmlund and X.-D.Su, submitted). The crystals belong to space group P2_12_1 with the cell dimensions a = 37.84 Å, b = 67.01 Å, c = 105.36 Å, and contain one molecule per asymmetric unit.

X-ray data collection and structure solution

X-ray data were collected under cryo conditions (M.Selmer, R.Wütling, D.Holmlund and X.-D.Su, submitted) on a single crystal at BW7A, EMBL outstation, DESY, Hamburg using a 165 mm marCCD detector. Fluorescence scans were used to find the absorption edges for yttrium and selenium. Data were collected at peak and inflection point wavelengths of both elements plus at a low energy remote wavelength and a remote wavelength between the two peaks. Data were indexed and integrated using Denzo and scaled using Scalepack (Otwinski and Minor, 1997). Data collection statistics are shown in Table I. Two yttrium sites and one selenium site were found using Solve (Terwilliger and Berendzen, 1999). Phase calculation to 3.5 Å, where good anomalous signal existed, and cross checking of the sites in anomalous Fourier maps were performed in CNS (Brünger et al., 1998). Using the same program, yttrium and selenium phases were combined, and solvent flipping and phase extension to 2.12 Å were performed to improve the quality of the experimental map. Phasing statistics are summarized in Table II.

Model building and refinement

Model building was performed with the program O (Jones et al., 1991). About 75% of the residues could be built in the experimental map. After cycles of simulated annealing and combination of model and experimental phases, the full chain except the first nine amino acids could be built. The coordinates and individual restrained B-factors were refined against the mhf target (maximum-likelihood with Hendrickson–Lattman coefficients) with CNS (Brünger et al., 1998) and, finally, water molecules were added. Refinement statistics are summarized in Table III. Difference densities possibly indicate double conformations of some side chains but, due to the limited resolution, no double conformations were modelled.

The atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank with accession number 1lv1.

Sequence alignment

Sequence alignment was performed with Clustal X (Thompson et al., 1997) and presented with Aclscript (Barton, 1993). The alignment was edited manually to avoid gaps inside the conserved structural elements (Figure 5B).

Acknowledgements

We thank Christopher Enroth and Emklek Pohl, EMBL, Hamburg for help during data collection, and Professor Anders Liljas for stimulating discussions and valuable comments on the manuscript. M.S. and X.-D.S. are recipients of financial support from the Swedish Foundation for Strategic research through SMBNet. This work was supported by grants from the Swedish Research Council (NFR) to Anders Liljas.