The EMBO Journal Vol. 20 No. 3 pp. 562–569, 2001

The crystal structure of the ttCsaA protein: an export-related chaperone from *Thermus thermophilus*

Shin-ichi Kawaguchi[1,2], Jörg Müller, Dirk Linde, Seiki Kuramitsu[1,3,4], Takehiko Shibata[5], Yorinao Inoue[1], Dmitry G.Vassylyev[1,6] and Shigeyuki Yokoyama[1,4,6,7]

Institute of Molecular Biology, Jena University, Winzerlaer Straße 10, D-07745 Jena, Germany, [1]RIKEN Harima Institute at SPring-8, 1-1-1 Kouto, Mikazuki-cho, Sayo-Gun, Hyogo 679-5148, [3]Department of Biology, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, [4]Genomic Sciences Center, RIKEN, Suehiro-cho, 1-7-22 Tsurumi, Yokohama, Kanagawa 230-0045, [5]Laboratory of Cellular and Molecular Biology, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198 and [7]Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Bunkyo-Ku, Tokyo 113-0033, Japan

[2]Present address: Department of Embryology, Carnegie Institution of Washington, 115 West University Parkway, Baltimore, MD 21210, USA

[6]Corresponding authors
e-mail: dmitry@yumiyoshi.harima.riken.go.jp or yokoyama@y-sun.biochem.s.u-tokyo.ac.jp

The CsaA protein was first characterized in *Bacillus subtilis* as a molecular chaperone with export-related activities. Here we report the 2.0 Å-resolution crystal structure of the *Thermus thermophilus* CsaA protein, designated ttCsaA. Atomic structure and experiments in solution revealed a homodimer as the functional unit. The structure of the ttCsaA monomer is reminiscent of the well known oligonucleotide-binding fold, with the addition of extensions at the N- and C-termini that form an extensive dimer interface. The two identical, large, hydrophobic cavities on the protein surface are likely to constitute the substrate binding sites. The CsaA proteins share essential sequence similarity with the tRNA-binding protein Trbp111. Structure-based sequence analysis suggests a close structural resemblance between these proteins, which may extend to the architecture of the binding sites at the atomic level. These results raise the intriguing possibility that CsaA proteins possess a second, tRNA-binding activity in addition to their export-related function.

Keywords: chaperone/crystal structure/protein export/tRNA

Introduction

For Sec-dependent translocation in bacteria, extracytosolic proteins are synthesized as pre-proteins containing short N-terminal extensions, the signal peptides (SPs), which are recognized by targeting factors to direct them to the translocase (SecA-SecYEG) (reviewed in Fekkes and Driessen, 1999). The SPs, which lack sequence similarity, preserve the extremely hydrophobic segments, which are believed to be a major factor in precursor recognition and translocation fidelity. In *Escherichia coli*, the export-dedicated cytosolic chaperone SecB is one of the targeting factors that binds to the mature regions of pre-proteins (Randall *et al*., 1990, 1998; Kumamoto and Francetic, 1993; Randall and Hardy, 1995; Knoblauch *et al*., 1999). However, a number of conflicting results argue that SecB also recognizes the SPs of pre-proteins (Watanabe and Blobel, 1989, 1995). The signal recognition particle (SRP) is a second targeting factor of *E.coli*, which is homologous to its eukaryotic counterparts (Phillips and Silhavy, 1992; de Gier *et al*., 1997; Valent *et al*., 1998). SRP specifically binds to the SPs of nascent precursor polypeptide chains in a co-translational manner (Rapoport *et al*., 1996). The crystal structures of the SP binding M-domains of the bacterial (Keenan *et al*., 1998; Batey *et al*., 2000) and human (Clemons *et al*., 1999) SRPs provided the structural basis for SP recognition, in which hydrophobic interactions are of central importance.

The CsaA protein from *Bacillus subtilis* was recently characterized as a chaperone with export-related activities (Müller *et al*., 1992, 2000a,b). The CsaA and SecB proteins have overlapping functions, although they share no sequence similarity. CsaA possesses affinities to a number of pre-proteins and the SecA subunit of the membrane translocase (Woodbury *et al*., 2000). Depletion of CsaA resulted in reduced export of a subset of secreted proteins. In addition, CsaA was shown to prevent the aggregation of the unfolded cytoplasmic protein, luciferase, which raises the possibility of its general chaperone activity. The CsaA counterparts (overall sequence identity and homology of 50 and 70%, respectively, which is a strong indication of functional identity) were found in the genomes of several other bacteria, including *Thermus thermophilus* (Figure 1). Interestingly, thus far, proteins homologous to SecB have only been identified in Gram-negative bacteria (de Cock and Tommasson, 1991). In particular, the complete sequences of the *B.subtilis* (Kunst *et al*., 1997) and *T.thermophilus* HB8 (S.-i.Kawaguchi, T.Shibata, Y.Inoue and S.Yokoyama, unpublished results) genomes revealed no SecB-like proteins. Thus, in these organisms, CsaA may compensate for the absence of SecB in Sec-dependent translocation.

According to sequence alignment, the CsaA and Trbp111 (Trbp) proteins and the C-domains of methionyl-tRNA synthetases (MetRS-C) belong to the same CsaA/Trbp/MetRS-C protein family (Figure 1). Whereas the biological function of MetRS-Cs remains elusive (Kohda *et al*., 1987), Trbp was recently shown to be a structure-specific tRNA-binding protein (Morales *et al*., 1999). Trbp binds as a dimer to any single intact tRNA molecule through interactions that are not sequence-specific. On the basis of *in vitro* studies, the protein may act as a

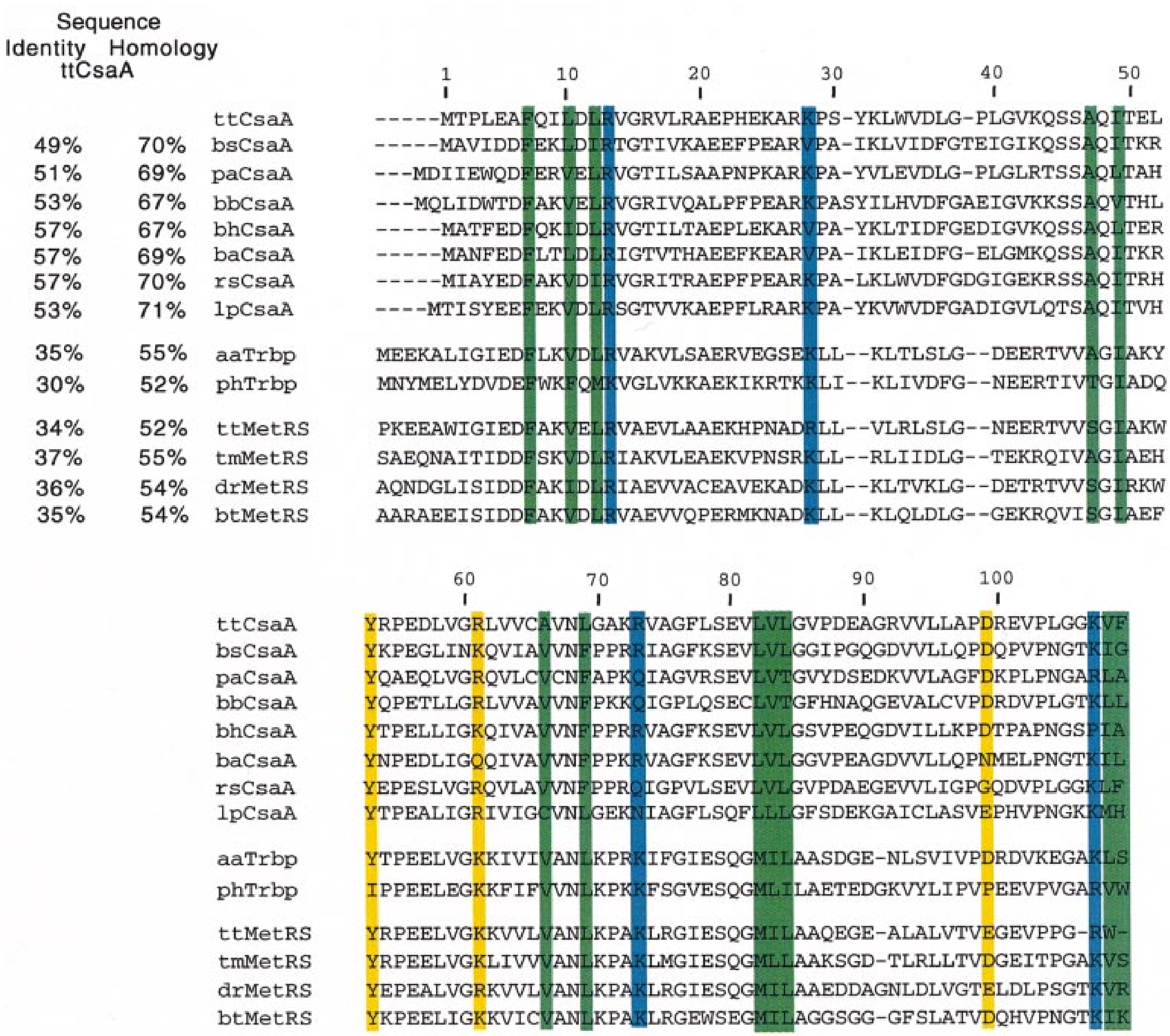


Fig. 1. Sequence alignment of the proteins/domains belonging to the CsaA/Trbp/MetRS-C family (Thompson *et al*., 1994; Altschul *et al*., 1997). For simplicity, only MetRS-Cs, which exhibit the highest sequence identities with the aaTrbp protein, are shown. The amino acid sequence identities with ttCsaA are shown on the left side of the sequences. The conserved basic and hydrophobic residues in the vicinity of the ttCsaA major hydrophobic cavity are coloured in blue and green, respectively. The conserved residues, which stabilize the ttCsaA dimer, are shown in yellow. aa, *Aquifex aeolicus* (Deckert *et al*., 1998); ba, *Bacillus anthracis*; bb, *Bordetella bronchiseptica*; bh, *Bacillus halodurans*; bs, *Bacillus subtilis* (Kunst *et al*., 1997); bt, *Bacillus stearothermophilus* (Mechulam *et al*., 1991); dr, *Deinococcus radiodurans* (White *et al*., 1999); lp, *Legionella pneumophila*; pa, *Pseudomonas aeruginosa*; ph, *Pyrococcus horikoshii* (Kawarabayasi *et al*., 1998); rs, *Rhodobacter sphaeroides*; tm, *Thermotoga maritima* (Nelson *et al*., 1999); tt, *Thermus thermophilus*. Preliminary sequence data for *B.anthracis*, *B.bronchiseptica*, *B.halodurans*, *L.pneumophila*, *P.aeruginosa* and *R.sphaeroides* CsaA proteins were obtained from The Institute for Genomic Research (TIGR) website at http://www.tigr.org.

tRNA-specific chaperone, which maintains the tRNA native fold (Morales *et al*., 1999).

We have determined the crystal structure of ttCsaA at 2.0 Å resolution. This structure provides a clue as to why ttCsaA binds specifically to the unfolded proteins and as to the mechanism by which ttCsaA may recognize the pre-protein SPs during translocation. Moreover, the structure-based sequence analysis suggests a second, Trbp-like tRNA-binding activity of ttCsaA.

Results

Crystal structure

The crystal structure of ttCsaA has been refined at 2.0 Å resolution to a final *R*-factor of 21.7% (R_{free} = 27.7%) (Table I). The asymmetric unit contains four protein molecules, which are arranged as a dimer of dimers (Figure 2A and B). The refined model includes 109 residues in each of the four protein molecules and 328 water molecules. The model lacks density for the first three residues (Gly-Ser-His), which originated from the cleavage of the His-tag. The first Met residue of the protein model is thus referred to as Met1. The average *B*-factor of the model is 38 Å^2 [$\sigma(B)$ = 2 Å^2].

Subunit structure

With the exception of the two short α-helices, the structure of the ttCsaA monomer is composed of β-strands. It can be divided into two major domains (Figure 2A and B). The first, which is a β-barrel domain (comprised of β1–β4, β7 and α2), constitutes the central core of the protein. It is reminiscent of the well known oligonucleotide binding

Table I. Summary of ttCsaA structure determination

Data collection	Native	$NaAuCl_4$[a] (0.1 mM)	$HgCl_2$ (100 mM)	PIP (2 mM)
Resolution (Å)	50.0–2.0	50.0–3.0	50.0–3.0	50.0–3.0
Reflections				
total (*n*)	135 225	21 624	25 490	32 788
unique (*n*)	33 368	8980	9681	10 013
R_{sym}[b] (%)	4.3 (19.1)[c]	6.5 (16.5)	7.3 (20.3)	7.4 (15.2)
completeness (%)	97.5 (94.5)	87.4 (66.2)	93.3 (71.3)	96.7 (94.0)
Derivatives				
R_{iso}[d] (%)	18.7	–	14.7	14.1
phasing power[e]	–	–	1.03	2.10

Refinement statistics		Stereochemistry	
Resolution (Å)	50.0–2.0	r.m.s.d. bond length (Å)	0.007
Number of reflections	33 368	r.m.s.d. bond angles (°)	1.54
R_{cryst}[f] (%)	21.7	r.m.s.d. improper angles (°)	0.90
R_{free}[f] (%)	27.7		
Number of protein atoms	3356		
Number of water molecules	328		

[a]The data for $NaAuCl_4$ heavy-atom derivative were used as the native data in the initial phasing (see Materials and methods).
[b]$R_{sym} = \Sigma_{hkl}\Sigma_j \mid I_j(hkl) - \langle I(hkl)\rangle \mid / \Sigma_{hkl}\Sigma_j \langle I(hkl)\rangle$, where $I_j(hkl)$ and $\langle I(hkl)\rangle$ are the intensity of measurement *j*, and the mean intensity for the reflection with indices *hkl*, respectively.
[c]The statistics for the highest resolution shell are shown in parentheses.
[d]$R_{iso} = \Sigma_{hkl} \| F_{der}(hkl) \| - \| F_{nati}(hkl) \| / \Sigma_{hkl} \mid F_{nati}(hkl) \mid$, where $F_{der}(hkl)$ and $F_{nati}(hkl)$ are the structure factors of the heavy-atom derivative and the native for the reflections with indices *hkl*, respectively. The $NaAuCl_4$ data were used as a native.
[e]Phasing power = $\langle F_h\rangle/E$, where $\langle F_h\rangle$ is the root mean square heavy-atom structure factor and *E* is the residual lack of closure error. The $NaAuCl_4$ data were used as a native.
[f]$R_{cryst,\ free} = \Sigma_{hkl} \| F_{calc}(hkl) \| - \| F_{obs}(hkl) \| / \Sigma_{hkl} \mid F_{obs} \mid$, where the crystallographic *R*-factor is calculated including and excluding refinement reflections. The free reflections constituted 5% of the total number of reflections.

(OB) fold (Murzin, 1993), and has the greatest structural similarity to the B2-domain of phenylalanyl-tRNA synthetase (Mosyak *et al*., 1995) and the cold shock protein, CspA (Schindelin *et al*., 1994), with a root mean square deviation (r.m.s.d.) of 2.3 and 2.5 Å over 86 and 63 C_α atoms, respectively. The second domain has no structural similarity with the other proteins. It consists of relatively short N- and C-terminal extensions of the OB fold domain (α1 helix and β8–β9 β-hairpin), which protrude from the central β-barrel on the same side of the molecule (Figure 2A and B). These unique structural elements are probably required to form the stable dimer.

Thus far, the OB fold has been found in a variety of nucleic acid binding proteins (Subramanya *et al*., 1996; Bochkarev *et al*., 1997; Horvath *et al*., 1998; Peat *et al*., 1998). There are no examples of peptide substrates for the proteins of this family. Thus, the CsaA chaperone is the first protein with an OB fold that possesses a different ligand specificity than those of the other members of the OB fold family.

The dimer is a functional unit of ttCsaA

In the crystal, there are two identical homodimers of ttCsaA (r.m.s.d. = 1.126 Å over all atoms in the dimer). The size exclusion chromatography (data not shown) revealed that, under physiological conditions, ttCsaA forms a dimer in solution. These data strongly suggest that the dimer is the actual functional unit of the ttCsaA protein.

There are two striking structural features of the dimeric ttCsaA molecule (Figure 2A). First, there are two deep, symmetrical cavities, with dimensions of ~7 × 13 × 15 Å, on the protein surface, which are likely to be the substrate binding sites (Figure 2C). Each cavity is composed of structural elements from both protein subunits (α1′, β1, β3, β4, β7, β9, L3, L5 and L8′), which emphasizes the significance of the dimerization for the function (Figure 2A and B). Secondly, the cavity interiors are essentially hydrophobic. In each cavity, there are 13 hydrophobic residues, the side chains of which are exposed to the solvent (Figure 2C). Each cavity is rimmed by seven basic residues (Arg13′, Arg27, Lys28, Lys72, Arg73, Arg92 and Lys107′), four of which are conserved within the CsaA/Trbp/MetRS-C family (Figures 1 and 2C).

In contrast to the ttCsaA monomer, which belongs to the well known family of OB fold proteins, the dimer conformation is unique. Upon dimer formation, the ttCsaaA subunits form an extensive intermolecular interface, which can be divided into two spatially distinct N- and C-terminal regions (Figure 2A and B). The C-terminal regions of the two monomers interact with each other through a mixture of multiple polar and hydrophobic contacts, and probably play the major structural role in dimer formation and stability. First, the β-strands, β9 and β9′, form an intermolecular anti-parallel β-sheet. Secondly, the L8 loop and β8′ strand (β8 and L8′) make several hydrogen bonds in a symmetrical fashion. Thirdly, a number of hydrophobic residues from both monomers constitute the intermolecular hydrophobic core. In total, 20 residues from each molecule make 44 van der Waals contacts, with distances <3.8 Å. Finally, Asp99 (at the end of β9) of one monomer forms a hydrogen bond and a salt bridge to the core residues of another molecule, Tyr53′ (L3′) and Arg61′ (L4′), respectively. As all CsaA sequence homologues (Trbp and MetRS-C) have been shown to be dimers (Kohda *et al*., 1987; Morales *et al*., 1999), Asp99, Tyr53′ and Arg61′ are likely to be crucial for the dimerization because they are well conserved in all sequences (Figure 1). In addition, the conservation of these residues suggests that the dimerization mode may be similar for all of the CsaA/Trbp/MetRS-C family members.

At the N-terminal region of the dimer interface, the N-terminal α-helices (α1, α1′) protrude from the protein surface and interact with each other in an anti-parallel manner (Figure 2A and D). As there are no hydrogen bonds and only a few weak van der Waals interactions (11 in total) between these α-helices and the protein core, the α-helices seem to be flexible and to contribute little to the dimer stability. The flexibility of the α1 and α1′ helices is also reflected in the poor final electron density and the abnormally high average *B*-value of 58 $Å^2$, which is ~10 σ(*B*-factor) larger than the overall average protein *B*-factor. On the other hand, this pair of α-helices is likely to play a functional role, as it mediates the interface between the two symmetrical major cavities (Figure 2A).

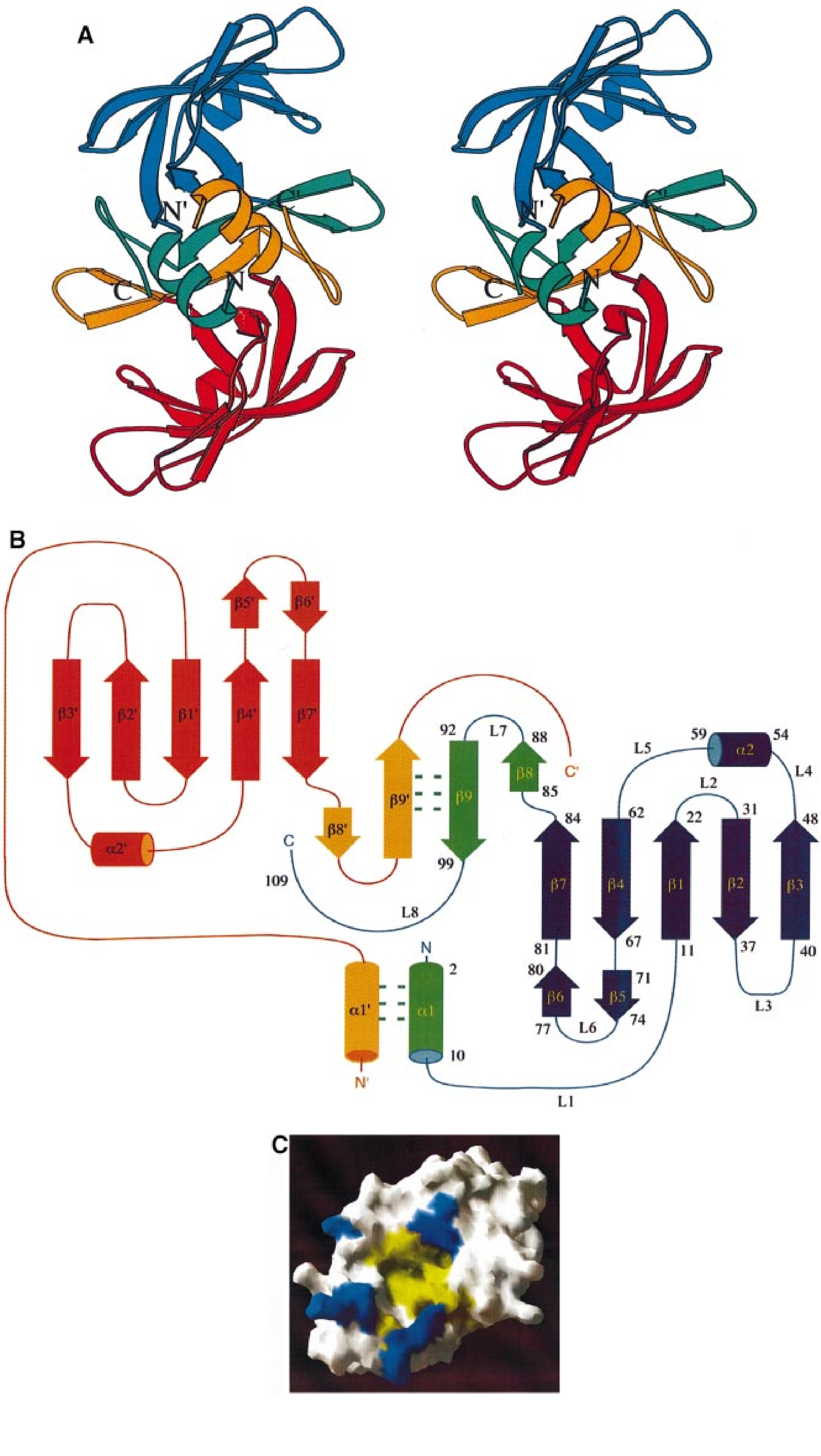
Discussion

The presence of the deep hydrophobic cleft on the protein surface implies that the ttCsaA peptide substrate should have a hydrophobic or amphiphilic nature. This implication is consistent with the chaperone-like functions of CsaA (Müller *et al*., 2000a). On the one hand, CsaA may bury the exposed hydrophobic segments of denatured cytoplasmic proteins to prevent their aggregation. The similar mechanism of binding might be used by CsaA to target the unfolded pre-proteins to the membrane translocase (Randall *et al*., 1990, 1998; Kumamoto and Francetic, 1993; Randall and Hardy, 1995; Knoblauch *et al*., 1999). On the other hand, in the translocation pathway, CsaA may bind specifically to the long hydrophobic segments of the pre-protein SPs. In that case, the unification of two hydrophobic cavities of ttCsaA would probably be required to accommodate the 21- to 26-residue SPs. This possibility is enhanced by the fact that the interface between the two cavities is 'closed' by only a pair of flexible N-terminal α-helices, which are loosely bound to the protein (Figure 2A and D). Upon the 'opening' of the α-helices, the cavities would be fused in a long hydrophobic groove, which resembles the SP binding sites of the SRPs (Keenan *et al*., 1998; Clemons *et al*., 1999). Given the plausible flexibility of the N-terminal α-helices, their opening may be easily induced by initial substrate binding and/or thermal motions. Additional support for this hypothesis comes from the observation that the N-terminal α-helices of ttCsaA are amphiphilic (Figures 1 and 2D). Thus, in the putative 'open' conformation, the helices may complement the hydrophobic SP binding site (Figure 3). Although co-immunoprecipitation and circular dichroism experiments showed a strong affinity of ttCsaA to the secreted pre-protein and ttCsaA/SP complex formation, respectively (data not shown), more systematic studies are required to elucidate the exact role of ttCsaA in translocation.

CsaA, Trbp and MetRS-C share ~38% sequence identity on average, and can therefore be grouped into a CsaA/Trbp/MetRS-C family (Figure 1). This level of sequence homology suggests an essential similarity in their three-dimensional structures. All three family members were shown to form dimers (Kohda *et al*., 1987; Morales *et al*., 1999). Moreover, the ttCsaA residues that are crucial for dimerization are conserved in all of the homologues (Figure 1). Thus, the structural similarity may be extended to the quaternary architecture of the proteins. We further analysed the sequences around the ttCsaA major binding cavity, which is likely to be structurally conserved. The mapping of the Trbp and MetRS-C sequences on the atomic structure of ttCsaA shows that the potentially important residues (basic for the tRNA-binding and hydrophobic for the chaperone activity) inside and around the proposed substrate binding cleft are well conserved (Figures 1 and 2C). These results suggest that CsaAs and other family members may possess dual substrate specificity. In this case, as the majority of the conserved basic residues (four out of six) (Figure 1) are located in the vicinity of the major ttCsaA hydrophobic cavity (Figure 2C), the cavity is likely to constitute the binding site for both the protein and RNA substrates.

In addition, the sequence alignment shows limited homology (~25% identity) between ttCsaA and two eukaryotic proteins: tyrosyl-tRNA synthetase (Kleeman *et al*., 1997) and the N-domain of endothelial monocyte-activating polypeptide II (EMAPII) (data not shown). The recently determined crystal structures of human EMAPII consistently revealed the presence of an OB fold in its N-domain (Kim *et al*., 2000; Renault *et al*., 2001). Surprisingly, the EMAPII C-domain, which has no sequence similarity to either its C-domain or the CsaA/Trbp/MetRS-C proteins, also adopts the truncated OB fold conformation. Moreover, the spatial arrangement of the EMAPII domains closely resembles that of the subunits in the ttCsaA dimer (Renault *et al*., 2001). It is worth mentioning that EMAPII has the potential to possess a dual function, comprising tRNA binding and cytokine activity.

The putative tRNA-binding activity of ttCsaA would not be surprising, as the protein structure exhibits the OB fold, which is known to be favourable for tRNA binding. (Cavarelli *et al*., 1994; Mosyak *et al*., 1995; Kleeman *et al*., 1997; Mechulam *et al*., 1999; Sugiura *et al*., 2000). However, the ttCsaA putative tRNA binding mode seems to be unique compared with the known structures of protein–tRNA complexes. A structure-based sequence analysis suggested that tRNA binding is likely to occur in the major hydrophobic cavities of the ttCsaA dimer and, thus, should be mediated by multiple hydrophobic interactions. It is possible that some tRNA bases exposed on the surface would be trapped in the protein cavity through van der Waals interactions, without making any polar base-specific contacts. At the entrance of the cavity, the exposed basic side chains may interact with the tRNA phosphate backbone to enhance protein–tRNA affinity. This view is consistent with the proposal for the Trbp structure-specific recognition of tRNAs (Morales *et al*., 1999). Although our preliminary experiments in solution (size exclusion chromotography and analytical ultracentrifugation) showed ttCsaA–tRNA binding (data not shown), the specificity of this binding should be investigated further. The detailed analysis of the potential ttCsaA bifunctionality is now underway.


A
N'
C'
C
N
B
β3'
β2'
β1'
β4'
β7'
β5'
β6'
α2'
β8'
β9'
β9
β8
92
L7
88
C'
85
84
99
C
109
L8
β7
β4
β1
β2
β3
62
22
31
48
59
54
α2
L5
L4
L2
81
67
11
37
40
80
71
β6
β5
77
L6
74
L3
α1'
α1
N
2
10
N'
L1
C

Materials and methods

Cloning, expression and purification

ttcsaA from *T.thermophilus* HB8 (the genomic sequence will be reported elsewhere) was amplified by PCR using the primers (5′-GGCGGGGAA-AGCGCATATGACCCCTCTGGAAGC-3′), containing a *Nde*I site, and (5′-GCCCCGCCGCGGGATCCTCATTAGAACACCTTCCC-3′), containing a *Bam*HI site. The amplified fragment was ligated to the pGEM-T Easy vector (Promega). In order to overproduce and purify the ttCsaA protein, the *Nde*I–*Bam*HI *ttcsaA* fragment was subcloned into the pET15b vector (Novagen). The resulting plasmid, pET15b-*ttcsaA*, was transformed into *E.coli* strain JM109(DE3).

The *E.coli* strain JM109(DE3) (pET15b-*ttcsaA*) was grown at 37°C in Luria–Bertani (LB) medium for 15 h. Cells were harvested, suspended in solution A (20 mM potassium phosphate, 30 mM borate, 0.5 M NaCl pH 8.5) and stored at –80°C. Cells were disrupted by sonication and were heated at 65°C for 10 min. The cleared cell extract was obtained by centrifugation at 15 000 r.p.m. for 30 min (Hitachi 46) and was loaded onto a Ni^{2+}-nitrilotriacetic agarose (NTA) column (Invitrogen) (3 ml bed volume). His_6-ttCsaA was eluted with a 0–250 mM imidazole gradient in solution A. Fractions containing ttCsaA were pooled. After thrombin (100 units) was added, in order to cleave the His-tag, the pooled fraction was dialysed against solution A. The resulting fraction was again loaded onto the Ni^{2+}-NTA column to remove the cleaved His-tag, and was then loaded onto a benzamidine Sepharose 6B (Amersham Pharmacia Biotech.) column to remove the thrombin. Finally, the ttCsaA protein was concentrated to ~14 mg/ml by CentriPrep ultrafiltration (Millipore).

Crystallization and data collection

Crystallization of ttCsaA was performed using the hanging drop vapour diffusion method. The drops were composed of 2 µl of ttCsaA solution and 1 µl of reservoir solution containing 3.8 M NaCl and 0.1 M sodium acetate pH 5.0. The crystals were grown up to 0.5 × 0.05 × 0.05 mm for 1 week at 20°C. The 2.0 Å resolution native data set was collected with a MAR165 CCD detector at the beamline BL44B2 (SPring-8, Japan). The diffraction data were processed with the programs DENZO and SCALEPACK (Otwinowski and Minor, 1996). The space group was $P4_3$, with unit cell dimensions $a = b = 55.7$ Å, $c = 167.1$ Å.

Three heavy-atom derivatives were prepared by soaking experiments. The ttCsaA crystals were transferred into 3 µl of a harvest solution containing 4.0 M NaCl, 0.1 M sodium acetate pH 5.0, and a heavy-atom compound. For $HgCl_2$, di-µ-iodobis(ethylenediamine)diplatinum (PIP) and $NaAuCl_4$, crystals were soaked for 6 h, 11 h and 20 min, respectively. The data sets were collected at 100 K with a Rigaku R-AXIS IV imaging plate detector, using CuKα radiation.

Structure determination and refinement

The structure was determined by a combination of multiple isomorphous replacement (MIR), anomalous scattering (AS) and non-crystallographic symmetry (NCS) averaging. The heavy-atom positions for the PIP and $HgCl_2$ heavy-atom derivatives were localized in the difference Patterson map using the CCP4 program (Collaborative Computational Project 4, 1994) RSPS and were confirmed by the difference cross-Fourier technique. However, the analysis of the $\Sigma|F_{der}–F_{nat}|/\Sigma F_{nat}$ distribution (CCP4 program SCALEIT) indicated essential non-isomorphism of both heavy-atom derivative crystals as compared with the native. On the other hand, when the $NaAuCl_4$ structure factors were substituted for the native

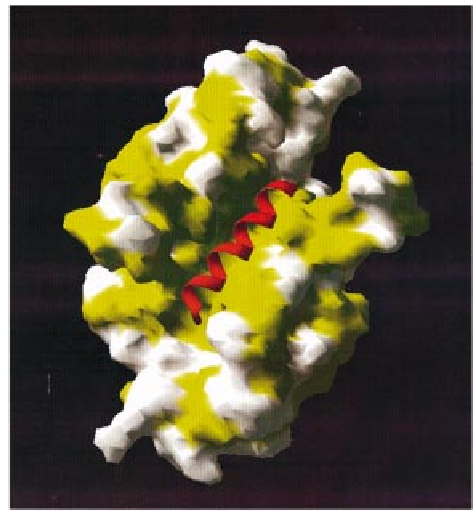


Fig. 3. 'Open' hydrophobic groove of ttCsaA. The ttCsaA molecular surface was calculated for the putative open conformation of the (α1,α1′) helical pair. All of the exposed hydrophobic residues of ttCsaA are coloured in yellow, whereas all of the other residues are white. The putative hydrophobic α-helix in the ttCsaA groove is shown as a magenta ribbon diagram.

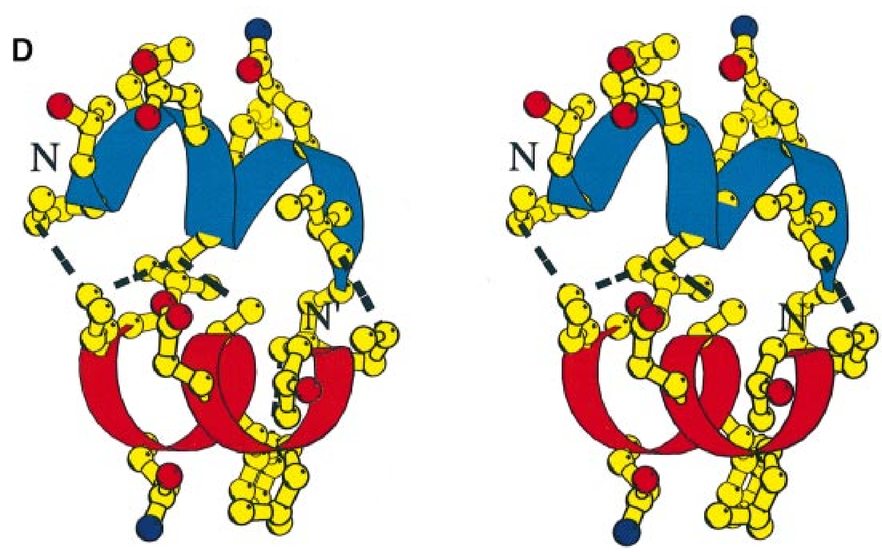


Fig. 2. Crystal structure of the ttCsaA protein. (**A**) Stereoview of the ribbon diagram of the ttCsaA dimer. The OB fold and the N- and C-terminal extensions are shown in blue and green for one subunit, respectively, and in red and orange for the other. (**B**) Topology diagram of the ttCsaA protein. The α-helices and β-strands are represented by cylinders and arrows, respectively. The colour scheme corresponds to that of (A). The two CsaA monomers are shown. (**C**) Molecular surface of the ttCsaA dimer. All residues are shown in white, except for the basic residues (blue) in the vicinity of the cleft and the hydrophobic residues (yellow) lining the interior of the major ttCsaA cavity. (**D**) Stereoview of the (α1,α1′) pair of the ttCsaA α-helices. The helices from two subunits (residues 2–10) are shown as blue and red ribbon diagrams with ball-and-stick side chains drawn with atom-dependent colours (C, yellow; O, red). The van der Waals inter-helical interactions are shown as black dashed lines.

data in the SCALEIT calculations, the same statistics showed almost perfect isomorphism of the PIP and $HgCl_2$ derivatives. It is worth mentioning that none of the techniques (isomorphous and anomalous difference Patterson search, and cross-Fourier) detected any binding of Au atoms to the protein. Finally, the replacement of the native with the $NaAuCl_4$ data resulted in an ~2.5-fold increase in the signal-to-noise ratio of the difference Patterson and cross-Fourier heavy-atom peaks of the PIP and $HgCl_2$ derivatives. Thus, we concluded that the Au atoms were not fixed in the crystal, but altered the crystal packing to produce non-isomorphism toward native crystals in a similar manner to that of the PIP and $HgCl_2$ derivatives. The MIRAS phases, using the $NaAuCl_4$ data as native, were calculated by the MLPHARE program (Otwinowski, 1991), and were improved by solvent flattening and histogram matching at 3.0 Å resolution with the DM program (Cowtan, 1994). In the resulting protein electron density map, four protein molecules in an asymmetric unit of the crystal were isolated. Then, the $NaAuCl_4$ 3.0 Å data set was replaced with the 2.0 Å native data, and 400 cycles of NCS averaging with phase extension from 4.0 to 2.0 Å resolution were carried out using the DM program. The final 2.0 Å resolution electron density map allowed unambiguous modelling of all ttCsaA molecules using the O program (Jones *et al*., 1991). The refinement was carried out with the CNS program (Brünger *et al*., 1998), which indicated a slight merohedral twinning of the native crystal, with a twinning fraction of 0.05. The *R*-factor of the initial model was 46.0% at 2.0 Å resolution. After several rounds of CNS refinement and manual rebuilding of the model, the *R*-factor dropped to the final value of 21.7% (R_{free} = 27.7%) at 2.0 Å resolution. The number of residues in the most favourable region of the Ramachadran plot was 88.8%, as indicated by the program PROCHECK (Laskowski *et al*., 1993).

Accession number

The coordinates and structure factors of the native ttCsaA protein have been deposited with the Protein Data Bank (accession code 1GD7). The nucleotide sequence of ttCsaA has been deposited with the DDBJ/EMBL/GenBank databases (accession No. AB052886).

Acknowledgements

The authors thank Drs K.Morikawa, D.Kohda, S.Borukhov and C.Cohen for helpful comments and discussions during the preparation of this manuscript.

References

Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Batey,R.T., Rambo,R.P., Lucast,L., Rha,B. and Doudna,J.A. (2000) Crystal structure of the ribonucleoprotein core of the signal recognition particle. *Science*, **287**, 1232–1239.

Bochkarev,A., Pfuetzner,R.A., Edwards,A.M. and Frappier,L. (1997) Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature*, **385**, 176–181.

Brünger,A.T. *et al*. (1998) Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.

Cavarelli,J. *et al.* (1994) The active site of yeast aspartyl-tRNA synthetase: structural and functional aspects of the aminoacylation reaction. *EMBO J.*, **13**, 327–337.

Clemons,W.M.,Jr, Gowda,K., Black,S.D., Zwieb,C. and Ramakrishnan,V. (1999) Crystal structure of the conserved subdomain of human protein SRP54M at 2.1 Å resolution: Evidence for the mechanism of signal peptide binding. *J. Mol. Biol.*, **292**, 697–705.

Collaborative Computational Project No. 4 (1994) The CCP4 Suite: Programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.

Cowtan,K. (1994) DM: an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallogr.*, **31**, 34–38.

Deckert,G. *et al*. (1998) The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature*, **392**, 353–358.

de Cock,H. and Tommassen,J. (1991) Conservation of components of the *Escherichia coli* export machinery in prokaryotes. *FEMS Microbiol. Lett.*, **64**, 195–199.

de Gier,J.W., Valent,Q.A., von Heijne,G. and Luirink,J. (1997) The *E. coli* SRP: preferences of a targeting factor. *FEBS Lett.*, **408**, 1–4.

Fekkes,P. and Driessen,A.J. (1999) Protein targeting to the bacterial cytoplasmic membrane. *Microbiol. Mol. Biol. Rev.*, **63**, 161–173.

Horvath,M.P., Schweiker,V.L., Bevilacqua,J.M., Ruggles,J.A. and Schultz,S.C. (1998) Crystal structure of the *Oxytricha nova* telomere end binding protein complexed with single strand DNA. *Cell*, **95**, 963–974.

Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, **47**, 110–119.

Kawarabayasi,Y. *et al*. (1998) Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. *DNA Res.*, **5**, 55–76.

Keenan,R.J., Freymann,D.M., Walter,P. and Stroud,R.M. (1998) Crystal structure of the signal sequence binding subunit of the signal recognition particle. *Cell*, **94**, 181–191.

Kim,Y., Shin,J., Li,R., Cheong,C., Kim,K. and Kim,S. (2000) A novel anti-tumor cytokine contains an RNA binding motif present in aminoacyl-tRNA synthetases. *J. Biol. Chem.*, **275**, 27062–27068.

Kleeman,T.A., Wei,D., Simpson,K.L. and First,E.A. (1997) Human tyrosyl-tRNA synthetase shares amino acid sequence homology with a putative cytokine. *J. Biol. Chem.*, **272**, 14420–14425.

Knoblauch,N.T., Rudiger,S., Schonfeld,H.J., Driessen,A.J., Schneider-Mergener,J. and Bukau,B. (1999) Substrate specificity of the SecB chaperone. *J. Biol. Chem.*, **274**, 34219–34225.

Kohda,D., Yokoyama,S. and Miyazawa,T. (1987) Functions of isolated domains of methionyl-tRNA synthetase from an extreme thermophile, *Thermus thermophilus* HB8. *J. Biol. Chem.*, **262**, 558–563.

Kumamoto,C.A. and Francetic,O. (1993) Highly selective binding of nascent polypeptides by an *Escherichia coli* chaperone protein *in vivo*. *J. Bacteriol.*, **175**, 2184–2188.

Kunst,F. *et al*. (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature*, **390**, 249–256.

Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

Mechulam,Y., Schmitt,E., Panvert,M., Schmitter,J.M., Lapadat-Tapolsky,M., Meinnel,T., Dessen,P., Blanquet,S. and Fayat,G. (1991) Methionyl-tRNA synthetase from *Bacillus stearothermophilus*: Structural and functional identities with the *Escherichia coli* enzyme. *Nucleic Acids Res.*, **19**, 3673–3681.

Mechulam,Y., Schmitt,E., Maveyraud,L., Zelwer,C., Nureki,O., Yokoyama,S., Konno,M. and Blanquet,S. (1999) Crystal structure of *Escherichia coli* methionyl-tRNA synthetase highlights species-specific features. *J. Mol. Biol.*, **294**, 1287–1297.

Morales,A.J., Swairjo,M.A. and Schimmel,P. (1999) Structure-specific tRNA-binding protein from the extreme thermophile *Aquifex aeolicus*. *EMBO J.*, **18**, 3475–3483.

Mosyak,L., Reshetnikova,L., Goldgur,Y., Delarue,M. and Safro,M.G. (1995) Structure of phenylalanyl-tRNA synthetase from *Thermus thermophilus*. *Nature Struct. Biol.*, **2**, 537–547.

Müller,J., Walter,F., van Dijl,J.M. and Behnke,D. (1992) Suppression of the growth and export defects of an *Escherichia coli secA*(Ts) mutant by a gene cloned from *Bacillus subtilis*. *Mol. Gen. Genet.*, **235**, 89–96.

Müller,J.P., Bron,S., Venema,G. and van Dijl,J.M. (2000a) Chaperone-like activities of the CsaA protein of *Bacillus subtilis*. *Microbiology*, **146**, 77–88.

Müller,J.P., Ozegowski,J., Vettermann,S., Swaving,J., Van Wely,K.H. and Driessen,A.J. (2000b) Interaction of *Bacillus subtilis* CsaA with SecA and precursor proteins. *Biochem. J.*, **348**, 367–373.

Murzin,A.G. (1993) OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J.*, **12**, 861–867.

Nelson,K.E. *et al*. (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature*, **399**, 323–329.

Otwinowski,Z. (1991) Maximum likelihood refinement of heavy atom parameters. In Wolf,W., Evans,P.R. and Leslie,A.G.W. (eds), *Isomorphous Replacement and Anomalous Scattering*. Science and Engineering Research Council, Warrington, UK, pp. 80–86.

Otwinowski,Z. and Minor,W. (1996) Processing X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.

Peat,T.S., Newman,J., Waldo,G.S., Berendzen,J. and Terwilliger,T.C.

(1998) Structure of translation initiation factor 5A from *Pyrobaculum aerophilum* at 1.75 Å resolution. *Structure*, **6**, 1207–1214.
Phillips,G.J. and Silhavy,T.J. (1992) The *E.coli ffh* gene is necessary for viability and efficient protein export. *Nature*, **359**, 744–746.
Randall,L.L. and Hardy,S.J. (1995) High selectivity with low specificity: how SecB has solved the paradox of chaperone binding. *Trends Biochem. Sci.*, **20**, 65–69.
Randall,L.L., Topping,T.B. and Hardy,S.J. (1990) No specific recognition of leader peptide by SecB, a chaperone involved in protein export. *Science*, **248**, 860–863.
Randall,L.L., Hardy,S.J., Topping,T.B., Smith,V.F., Bruce,J.E. and Smith,R.D. (1998) The interaction between the chaperone SecB and its ligands: evidence for multiple subsites for binding. *Protein Sci.*, **7**, 2384–2390.
Rapoport,T.A., Jungnickel,B. and Kutay,U. (1996) Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. *Annu. Rev. Biochem.*, **65**, 271–303.
Renault,L. *et al*. (2001) Structure of the EMAPII domain of human aminoacyl-tRNA synthetase complex reveals evolutionary dimer mimicry. *EMBO J.*, **20**, 570–578.
Schindelin,H., Jiang,W., Inouye,M. and Heinemann,U. (1994) Crystal structure of CspA, the major cold shock protein of *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **91**, 5119–5123.
Subramanya,H.S., Doherty,A.J., Ashford,S.R. and Wigley,D.B. (1996) Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7. *Cell*, **85**, 607–615.
Sugiura,I. *et al*. (2000) The 2.0 Å crystal structure of *Thermus thermophilus* methionyl-tRNA synthetase reveals two RNA-binding modules. *Struct. Fold Des.*, **8**, 197–208.
Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL_W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
Valent,Q.A., Scotti,P.A., High,S., De Gier,J.W., von Heijne,G., Lentzen,G., Wintermeyer,W., Oudega,B. and Luirink,J. (1998) The *Escherichia coli* SRP and SecB targeting pathways converge at the translocon. *EMBO J.*, **17**, 2504–2512.
Watanabe,M. and Blobel,G. (1989) SecB functions as a cytosolic signal recognition factor for protein export in *E. coli*. *Cell*, **58**, 695–705.
Watanabe,M. and Blobel,G. (1995) High-affinity binding of *Escherichia coli* SecB to the signal sequence. *Proc. Natl Acad. Sci. USA*, **92**, 10133–10136.
White,O. *et al*. (1999) Genome sequence of the radioresistant bacterium *Deinococcus radiodurans*. *Science*, **286**, 1571–1577.
Woodbury,R.L., Topping,T.B., Diamond,D.L., Suciu,D., Kumamoto, C.A., Hardy,S.J. and Randall,L.L. (2000) Complexes between protein export chaperone SecB and SecA: Evidence for separate sites on SecA providing binding energy and regulatory interactions. *J. Biol. Chem.*, **275**, 24191–24198.

Received September 20, 2000; revised December 7, 2000; accepted December 8, 2000