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

Shin-ichi Kawaguchi1,2, Jörg Müller, Dirk Linde, Seiki Kuramitsu1,3,4, Takehiko Shibata5, Yorinao Inoue1, Dmitry G.Vassylev1,6 and Shigeyuki Yokoyama1,4,6,7

Institute of Molecular Biology, Jena University, Winzerlaer Strasse 10, D-07745 Jena, Germany, 2RIKEN Harima Institute at SPring-8, 1-1-1 Kouto, Mikazuki-cho, Sayo-Gun, Hyogo 679-5148, 3Department of Biology, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, 4Genomic Sciences Center, RIKEN, Suehiro-cho, 1-7-22 Tsurumi, Yokohama, Kanagawa 223-0045, 5Laboratory of Cellular and Molecular Biology, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198 and 6Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Bunkyo-Ku, Tokyo 113-0033, Japan

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; Kumatomo 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
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$^2$ [$σ(B) = 2$ 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...
Bochkarev nucleic acid binding proteins (Subramanya forms a dimer in solution. These data strongly suggest revealed that, under physiological conditions, ttCsaA

The size exclusion chromatography (data not shown) that the dimer is the actual functional unit of the ttCsaA proteins of this family. Thus, the CsaA chaperone is the 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 ttCsaA 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 Arg60′ (L4′), respectively. As all CsaA sequence homologues (Trbp and MetRS-C) have been shown to be dimers (Kohda et al., 1987; Morales et al.,

<table>
<thead>
<tr>
<th>Table I. Summary of ttCsaA structure determination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection Native NaAuCl₄ (0.1 mM) HgCl₂ (100 mM) PIP (2 mM)</td>
</tr>
<tr>
<td>Resolution (Å)</td>
</tr>
<tr>
<td>Refinements and statistics</td>
</tr>
<tr>
<td>Reflections total (n)</td>
</tr>
<tr>
<td>Reflections unique (n)</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt;</td>
</tr>
<tr>
<td>Completeness (%)</td>
</tr>
<tr>
<td>Derivatives R&lt;sub&gt;merge&lt;/sub&gt; (%)</td>
</tr>
<tr>
<td>Refinement statistics</td>
</tr>
<tr>
<td>Resolution (Å)</td>
</tr>
<tr>
<td>Number of reflections</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt; (%)</td>
</tr>
<tr>
<td>R&lt;sub&gt;refine&lt;/sub&gt; (%)</td>
</tr>
<tr>
<td>Number of protein atoms</td>
</tr>
<tr>
<td>Number of water molecules</td>
</tr>
</tbody>
</table>

1The data for NaAuCl₄ heavy-atom derivative were used as the native data in the initial phasing (see Materials and methods).
2<sup>1</sup>R<sub>merge</sub> = Σ<sub|hkl|</sub> | F<sub>calc(hkl)</sub> | − | F<sub>obs(hkl)</sub> | / | Σ<sub|hkl|</sub> | F<sub>calc(hkl)</sub> |, where | F<sub>obs(hkl)</sub>| and | F<sub>calc(hkl)</sub>| are the structure factors of the heavy-atom derivative and the native for the reflections with indices hkl, respectively. The NaAuCl₄ data were used as a native.
3Phasing power = <F<sub>obs</sub>/E>, where | F<sub>obs</sub>| is the root mean square heavy-atom structure factor and E is the residual lack of closure error. The NaAuCl₄ data were used as a native.
4<sup>4</sup>R<sub>cryst</sub> = Σ<sub|hkl|</sub> | F<sub>calc(hkl)</sub> | − | F<sub>obs(hkl)</sub> | / | Σ<sub|hkl|</sub> | F<sub>calc(hkl)</sub>, where the crystallographic R-factor is calculated including and excluding refinement reflections. The free reflections constituted 5% of the total number of reflections.

(OF) fold (Murzin, 1993), and has the greatest structural similarity to the B2-domain of phenylalanyl-tRNA synthetase (Mосyak 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 Cα 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 protrudes 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; Pet 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...
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 Å², 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; Mechulum et al., 1999; Sugiuara 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 chromatography 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.
Materials and methods

Cloning, expression and purification

ttcsA from T. thermophilus HB8 (the genomic sequence will be reported elsewhere) was amplified by PCR using the primers (5'-GCGCCGGAAGGCAGCATATGACCTGGAAGC-3'), containing a NdeI site, and (5'-GCGCCGGAAGGCAGATCTAGAAGACGCTTTCCC-3'), containing a BamHI site. The amplified fragment was ligated to the pGEM-T Easy vector (Promega). In order to overproduce and purify the ttCsaA protein, the NdeI-BamHI ttcsA fragment was subcloned into the pET15b vector (Novagen). The resulting plasmid, pET15b-ttcsA, was transformed into E.coli strain JM109(DE3).

The E.coli strain JM109(DE3) (pET15b-ttcsA) 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 15000 r.p.m. for 30 min (Hitachi 46) and was loaded onto a Ni²⁺-nitrilotriacetic agarose (NTA) column (Invitrogen) (3 ml bed volume). His₉-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 Ni²⁺-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 ttCsA was performed using the hanging drop vapour diffusion method. The drops were composed of 2 μl of ttCsA 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 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 P4₁, with unit cell dimensions a = b = 55.7 Å, c = 167.1 Å.

Three heavy-atom derivatives were prepared by soaking experiments. The ttCsA 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 HgCl₂, di-μ-iodobis(ethylendiamine)diplatium (PIP) and NaAuCl₄, 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 HgCl₂ 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 Σ|Fₐₐₙ|/Σ|Fₐₐₙ| distribution (CCP4 program SCALIEIT) indicated essential non-isomorphism of both heavy-atom derivative crystals as compared with the native. On the other hand, when the NaAuCl₄ structure factors were substituted for the native

![Fig. 3.](image)

Fig. 3. ‘Open’ hydrophobic groove of ttCsA. The ttCsA molecular surface was calculated for the putative open conformation of the (α₁, β₁)’ helical pair. All of the exposed hydrophobic residues of ttCsA are coloured in yellow, whereas all of the other residues are white. The putative hydrophobic α-helix in the ttCsA groove is shown as a magenta ribbon diagram.
data in the SCALEIT calculations, the same statistics showed almost perfect isomorphism of the PIP and HgCl₂ 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 NaAuCl₄ 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 HgCl₂ 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 HgCl₂ derivatives. The MIRAS phases, using the NaAuCl₄ data as native, were calculated 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 NaAuCl₄ 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 tCsaA molecules using the O program (Jones et al., 1991). The refinement was carried out with the CNS program (Brüner 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=cv = 27.5%) at 2.0 Å resolution. The number of residues in the most favourable region of the Ramachandran plot was 88.8%, as indicated by the program PROCHECK (Laskowski et al., 1993).

Accession number
The coordinates and structure factors of the native tCsaA protein have been deposited with the Protein Data Bank (accession code 1G7D). The nucleotide sequence of tCsaA 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
Peat,T.S., Newman,J., Waldo,G.S., Berendzen,J. and Terwilliger,T.C.


