Structure of the hypothetical *Mycoplasma* protein, MPN555, suggests a chaperone function

Ursula Schulze-Gahmen[a], Shelly Aono[a,c], Shengfeng Chen[a], Hisao Yokota[a], Rosalind Kim[a] and Sung-Hou Kim[a,b]

[a] Berkeley Structural Genomics Center, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA and [b]Department of Chemistry, University of California, Berkeley, California, 94720, USA. [c]Current address: Dept. Anatomy, Physiology & Pharmacology, Auburn University, Auburn, AL 36849.

Correspondence e-mail: shkim@lbl.gov

Synopsis

The crystal structure of a hypothetical protein (MPN555) from *Mycoplasma pneumoniae* revealed a tri-lobal molecule with at least one central binding pocket or channel. The molecule has structural homology with two bacterial chaperone proteins, which suggests a similar chaperone function for MPN555.

Abstract

The crystal structure of the hypothetical protein MPN555 from Mycoplasma pneumoniae (gi| 1673958) has been determined to a resolution of 2.8 Å using anomalous diffraction data at the Se peak wavelength. Structure determination revealed a mostly α-helical protein with a three-lobed shape. The three lobes or fingers delineate a central binding groove and additional grooves between lobes 1 and 3, and between lobes 2 and 3. For one of the molecules in the asymmetric unit, the central binding pocket was filled with a peptide from the uncleaved N-terminal affinity tag. The MPN555 structure has structural homology to two bacterial chaperone proteins, SurA and trigger factor from *Escherichia coli*. The structural data and the homology to other chaperone proteins suggests an involvement in protein folding as a molecular chaperone for MPN555.

Keywords

Structural Genomics, gi 1673958, *Mycoplasma pneumoniae*, molecular chaperone, X-ray crystallography

1. Introduction

MPN555 is a hypothetical protein from *Mycoplasma pneumoniae*. Sequence similarity searches using Psi-Blast (Altschul et al. 1997) identified no sequences with strong homologies, and very few weak sequence homologs, most of which occur in *Mycoplasma* or *Ureaplasma* species. Only one very remote sequence homolog was identified: the C-terminal domain of a trigger factor from *Pseudomonas putida* (accession number PF05698 in Pfam database (Bateman et al. 2000)). Trigger factors are molecular chaperones with peptidyl-prolyl-cis-trans-isomerase (PPIase) activities. The C-terminal domain of the *E.coli* trigger factor appears to be critical for the chaperone activity of this protein (Kramer *et al*., 2004) and its structure has been determined recently (Ferbitz *et a*l., 2004). However, the sequence similarity between MPN555 and the C-terminal domain of *E.coli* trigger was not even recognized in Psi-Blast searches. The three-dimensional structures of other MPN555 homologous proteins are unknown. We determined the crystal structure of MPN555 to obtain some clues about the possible molecular function of this protein family.

2. Material and methods

2.1 Cloning, expression and purification

Primers (Life Tech, Bioneer, MWG or Operon) for PCR amplification from genomic DNA contained an NdeI restriction site in the forward primer (5'-CATATGGCTACAAATCTTAAATCAACCG) and a BamHI site in the reverse primer (5'-GGATCCTTAGTTTGGTAATTGTCCAGTTAAGTTAAA). PCR was performed using Deep Vent Polymerase (New England Biolabs, Inc., Beverly, MA) and genomic

DNA. The PCR product was cloned into pCR-BluntII-TOPO vector (Invitrogen Corp., Carlsbad, CA) and the gene insert was confirmed by DNA sequencing. The amplified TOPO vector was restricted with NdeI and BamHI and the gene insert was purified by agarose gel electrophoresis extraction. This insert was ligated into pSKB3 (gift from Steve Burley, Rockefeller University, New York), digested with NdeI and BamHI, and transformed into DH5α. A plasmid containing the gene insert was confirmed and then transformed into BL21(DE3)pSJS1244 (Kim *et al*., 1998).

MPN555 fusion protein was expressed in *Escherichia coli* strain BL21(DE3)pSJS1244 using an auto-inducible selenomethionyl (SEMet) medium (Dr. William Studier, Brookhaven National Laboratory, personal communication). For cell lysis, 3.5 g cell pellet was resuspended in 20 ml 50 mM Tris/HCl pH 7.5, 300 mM NaCl, 5 mM imidazole with 1 mM PMSF, 1 μg/ml leupeptin, 2.5 μg/ml pepstatin, 0.2 μg/ml antipain, 1 μg/ml chymostatin, and 10 μg/ml DNAse I. Cells were disrupted by sonication, followed by centrifugation at 15,000 g to remove cell debris. The supernatant was then spun at 60,000 g for 30 mins at 277 K to remove membrane proteins. The His-tagged fusion protein was then purified over a 7 ml cobalt affinity column (Talon resin; Clontech, Palo Alto, CA) and eluted with 100 mM and 300 mM imidazole. Protein containing fractions were combined and buffer exchanged against 50 mM Tris/HCl pH 9.0, 1 mM EDTA, 5% glycerol prior to anion exchange chromatography on a HQ20 Poros column (Applied Biosystems, Foster City, CA), which was equilibrated in the same buffer and developed with a NaCl gradient. MPN555 eluted in a single peak with a small shoulder at about 200 mM NaCl from the anion exchange column. Fractions from the main peak were combined and concentrated to 30 mg/ml in 20 mM Tris pH 8.0, 1 mM

EDTA, 2 mM DTT using an Ultrafree 10 kDa cutoff unit (Millipore Corp., Bedford, Massachusetts, USA). The purity of the expressed protein was determined by SDS-PAGE to be ~99%. Dynamic light scattering (DynaPro 99, proterion Corporation, Piscataway, NJ, USA) showed a single monodisperse peak (23% polydispersity), indicating homogeneity of the protein.

2.2. Crystallization and data collection

Screening for crystallization conditions was performed using the sparse-matrix method (Jancarik & Kim, 1991) with several screens from Hampton Research (Laguna Niguel, CA, USA). Several 96-well plates were used to set up the screens with the 'Hydra Plus One' crystallization robot (Matrix Technologies, Hudson, NH, USA) using the sitting drop vapor-diffusion method at room temperature. The best initial crystals were grown in 50 m*M* Bis/Tris pH 6.5, 50 m*M* ammonium sulfate, 30% Pentaerthritol Ethoxylate (PEE). Crystals grew very large but with multiple crystals intertwined. Addition of detergent eventually improved the crystals sufficiently to provide small single crystals for data collection. The best crystals were grown in 50 m*M* Bis/Tris pH 6.5, 60 m*M* ammonium sulfate, 35% Pentaerythritol Ethoxylate, 0.4 mM Zwittergent 3-14 at 22ºC in sitting drops. The precipitant of the MPN555 crystallization is an effective cryoprotectant, so that crystals could be flash-cooled directly out of the crystallization drop without any further soaking. The Se peak-wavelength (λ=0.9795) SAD data set was collected at the Macromolecular Crystallography facility beamline 5.0.2 at the Advanced Light Source (Lawrence Berkeley National Laboratory, Berkeley, CA, USA). A Quantum 4 CCD detector from Area Detector Systems Co. (ADSC, Poway, CA, USA) was used. The crystal-to-detector distance was set to 250 mm. In total, 360 images were collected (180

images for both direct and inverse beam) in 18 wedges with 1º oscillation range per image. A data set of 2.8 Å resolution was processed using the program Mosflm (Collaborative Computational Project, Number 4, 1994). The crystal belongs to the monoclinic space group C2, with unit cell parameters a = 132.37, b = 45.59, c = 153.93, β = 111.40º. With four molecules in the asymmetric unit (AU), the solvent content of the crystal is 40% with a Matthews coefficient of 2.13 $Å^3$ Da^{-1}. A least squares straight line from the Wilson plot approximates the *B* factor as around 65 $Å^2$. The data statistics are shown in Table 1.

2.3. Structure determination and refinement

Twelve out of 20 possible Se-atom positions were located using the program *HySS* from the *PHENIX* package (Grosse-Kunstleve & Adams, 2003), followed by phasing in *CNS* (Brunger *et al*., 1998). Diffraction data and phases were imported into *RESOLVE 2.07* (Terwilliger, 2002) for density modification and model building. *RESOLVE* found non-crystallographic translational symmetry (NCS) relating two pairs of MPN555, which agrees with a strong peak at 0.5, 0.175, 0.5 in the native Patterson map. The partial model built by RESOLVE included 536 residues and resulted in an R_{free} of 46.5%. Further tracing and rebuilding was performed using the program *O* (Jones *et al*., 1991) alternating with refinement in *CNS* and *REFMAC 5.2* (Murshudov *et al*., 1997). The model was built for two molecules and the remaining two molecules in the asymmetric unit were created using NCS operators. NCS restraints were included in the early stages of refinement but completely released in the later stages. The electron density for most parts of the molecules is well connected (Fig. 1). However, electron density for the N-terminal tag

and linker sequence varies enormously among the molecules, from almost invisible in molecule B, to untraceable fragments in molecules A, and C, to clearly traceable in molecule D. The untraceable densities in some of the molecules may contribute to the relatively high R-factor of this structure (Table 2). The final structure includes the following residues: t12-t15, t22-A96, A102-A190, B5-B93, B103-B145, B149-B187, t14-t18, t23-C2, C5-C97, C102-C190, t14-t118, D100-D144, and D150-D192 (Fig. 2). The atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) under the accession code 1ZXJ.

3. Results and discussion

3.1. Structure description

The crystallized form of MPN555 includes the 193 amino acids of the complete protein and an additional 25 residues for the uncleaved N-terminal His-tag (Fig. 2). Most of the protein residues are clearly defined in all four molecules in the asymmetric unit. Only residues in the loops between α3 and α4, and between α5 and α6 (Fig. 3A) are missing. The quality of the density for the N-terminal His-tag varies enormously among the different molecules. Molecule B has no density for the tag, molecule A and C have density fragments that are difficult to interpret, and molecule D shows quite clear density for residues t14 to t25 of the His-tag. Two pairs of the four molecules in the asymmetric unit are very similar in structure with an r.m.s. value on main-chain atoms of 0.6Å for molecule A and C, and an r.m.s. value of 0.65Å for molecule B and D. The other possible pairs (A-B, A-D, B-C, and D-C) have larger structural differences in residues 22 to 27 and 42 to 50, leading to r.m.s. values around 1.6Å for all main-chain atoms. The

structural differences are mostly located in loop regions, and in a translational shift of the N-terminal segment of α2. MPN555 is a mostly α-helical structure with two very small β-sheets (Fig. 3A). It has a tri-lobal shape where the three lobes or fingers delineate a pocket or groove in the center of the molecule. The three lobes are made up of α1 and α2 (lobe 1), α3, α4, β1, and β3 (lobe 2), and α5, α6, and part of α7 (lobe 3). A small β-sheet extends away from the main part of the molecule. The central pocket of the protein is filled by residues from the N-terminal His-tag in molecule D. There are two additional empty grooves between lobe 1 and 3 and between lobe 2 and 3. The shape of the molecule resembles a hand with thumb, pointer, and middle finger extended and slightly curved.

3.2. Comparison with similarly folded structures

Despite the low sequence homologies, the search for proteins with structural similarity using *DALI* (Holm & Sander, 1996) revealed two proteins with significant structural homology. The *E.coli* protein SurA (PDB code 1m5y) and the *E.coli* trigger factor (PDB code 1w26) both contain domains with structural similarity to MPN555 with Dali Z-scores of 6.2 and 9.2, respectively. SurA protein facilitates correct folding of outer membrane proteins in gram-negative bacteria. It accomplishes this in the absence of ATP or other sources of chemical energy. The protein sequence is divided in four segments, an N-terminal segment, two peptidyl-prolyl isomerase segments (PPIase), and a C-terminal segment. The amino-terminal and carboxy-terminal segments together with the first PPIase segment form a core structural module, and the second PPIase segment is a satellite domain tethered 30Å from the core structure (Bitto & McKay, 2002). The

MPN555 structure is homologous to the part of the core structure of SurA that is composed of the N-terminal and C-terminal segments (Fig. 4A). Interestingly, the crystal contacts in the SurA structure show that peptides bind within a crevice of the core module that is equivalent to the binding pocket in MPN555 that is filled with the N-terminal tag in some of the molecules. In addition, it was found that the N and C domains of SurA alone can be expressed *in vivo* and complement wild-type *surA* activity. The fusion construct can be expressed as a recombinant protein in *E.coli*, is stable, and can be purified. Although the SurA sequence includes two PPIase domains, the PPIase activity is apparently dispensable for SurA function *in vivo* (or is encoded by redundant PPIase genes). SurA has a chaperone mechanism that is largely encoded in the non-PPIase domains of its sequence, and this is the part that is structurally homologous to MPN555.

The second protein with structural homology to MPN555 is the *E.coli* ribosome-associated trigger factor (TF). TF binds to nascent protein chains on the ribosome. It is ATP independent and does not require a co-chaperone. TF is composed of an N-terminal ribosome-binding domain, a PPIase domain, and a C-terminal domain that shows structural homology to MPN555 and SurA (Ferbitz *et al.*, 2004). Similarly to SurA, PPIase acitivity is not essential for nascent chain binding and chaperone activity of trigger factor (Kramer *et al.*, 2004). The fused N-terminal and C-terminal domains provide trigger factor with almost wild-type-like chaperone activity *in vivo* and *in vitro*. The two domains form a cradle with a back and two 'arms' that has the capacity to bind peptide fragments. Analysis of the crystal packing reveals that a neighbouring molecule inserts an α-helix and β-strand between the 'arms' of the C-terminal domain (Ferbitz *et al.*, 2004). The structures of MPN555 and the C-terminal domain of trigger factor show

clear similarity (Fig. 4B). Lobe 1 in MPN555 corresponds to the ‘back’ of the trigger factor domain, and lobe 2 and 3 of MPN555 correspond to the ‘arms’ of the trigger factor.

The structural homologies of MPN555 with SurA and *E.coli* trigger factor, and the observed binding of the N-terminal tag in the MPN555 binding pocket all suggest a chaperone function for MPN555. This would also explain the low sequence conservation. Overall shape and surface distribution of hydrophobic patches are more likely to be critical for substrate binding than particular side chains, because molecular chaperones need to be able to bind a diverse set of peptides. The structural results and analysis from this study can direct future biochemical studies to further explore and possibly confirm the protein’s function.

Table 1

X-ray diffraction data statistics

Wavelength (Å)	0.9795
Resolution (Å)	42.76-2.8 (2.95-2.80)
Redundancy	3.3 (3.1)
Unique reflections	40298
Completeness	96.5 (91.8)
I/σ(I)	6.7 (1.8)
R_{sym}† (%)	7.9 (47.2)

Values in parenthesis are for the outermost resolution shell.

† $R_{sym} = \sum_{hkl}\sum_{I}|I_i - \langle I\rangle|/\sum\langle I\rangle$, where I_i is the intensity of the *i*th measurement of reflection *hkl* and $\langle I\rangle$ is the average intensity of the reflection.

Table 2

Phasing and refinement statistics of the SAD X-ray diffraction data.

Phasing statistics	
Unit-cell parameters (Å, °)	a = 132.37, b = 45.59, c = 153.93, β = 111.40
Se atoms found per AU	12
FOM (figure of merit)	0.24
FOM after density modification	0.55
Refinement statistics	
Unique reflections	38935
R factor (%)	24.9
R_{free}† (%)	32.3
Residues per monomer	A, 192; B, 171; C, 192; D, 193
Residues per asymmetric unit	748
No. atoms per asymmetric unit	5919
Protein	5905
Water	14
R.m.s. deviations	
Bond distances (Å)	0.008
Angle deviations (º)	1.265
$B_{average}$ (Å^2)	40.0
Ramachandran plot statistics	
Residues in most favored regions (%)	88.5
Residues in additional allowed regions (%)	11.4
Residues in disallowed regions (%)	0

†R_{free} calculated as R factor but on 7.5% of data excluded from refinement.

Fig. 1

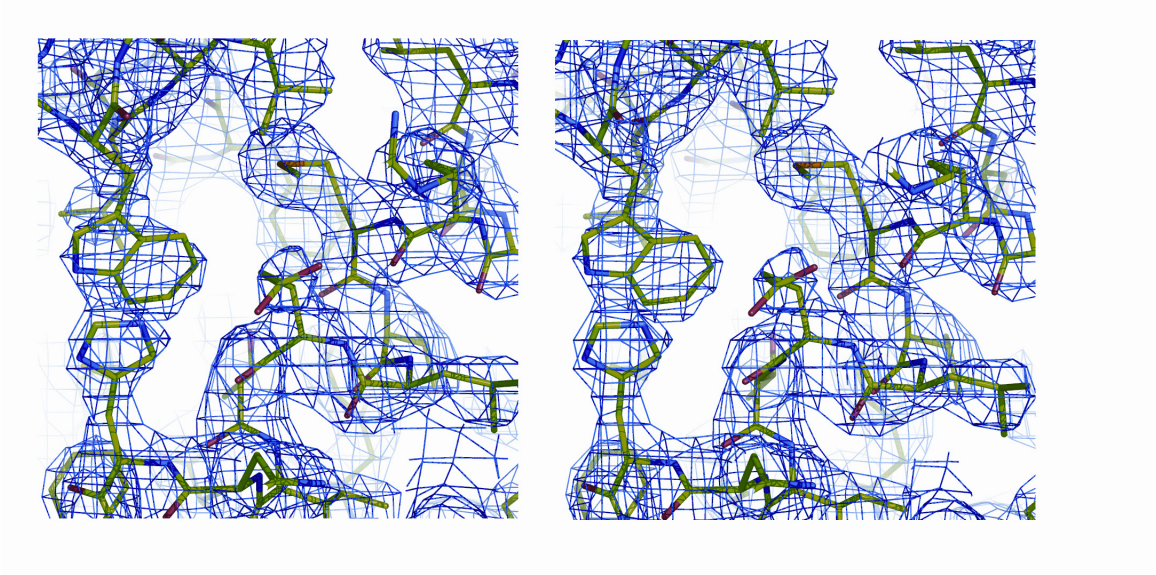

Fig. 2

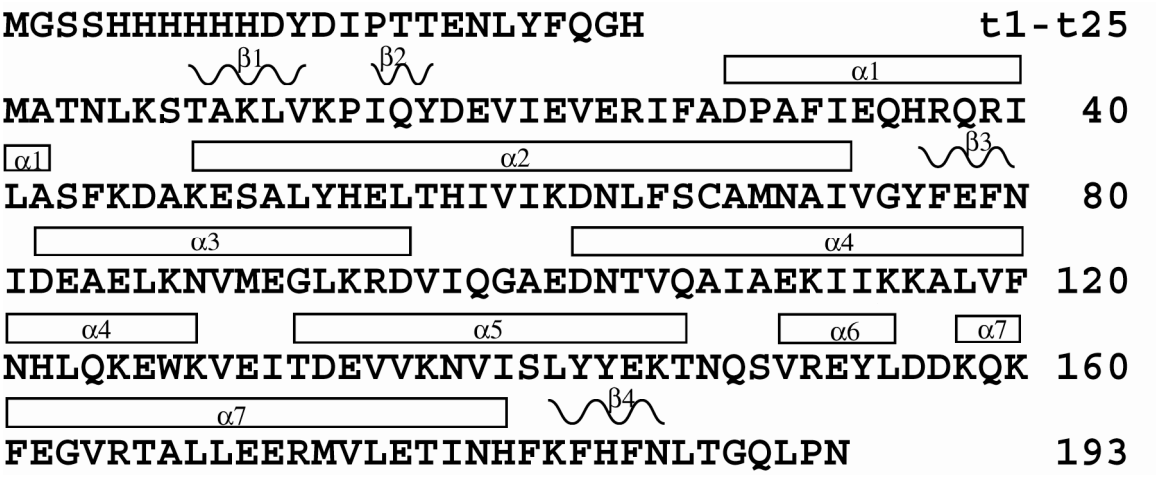
MGSSHHHHHHDYDIPTTENLYFQGH t1-t25
β1
β2
α1
MATNLKSTAKLVKPIQYDEVIEVERIFADPAFIEQHRQRI 40
α1
α2
β3
LASFKDAKESALYHELTHIVIKDNLFSCAMNAIVGYFEFN 80
α3
α4
IDEAELKNVMEGLKRDVIQGAEDNTVQAIAEKIIKKALVF 120
α4
α5
α6
α7
NHLQKEWKVEITDEVVKNVISLYYEKTNQSVREYLDDKQK 160
α7
β4
FEGVRTALLEERMVLETINHFKFHFNLTGQLPN 193

Fig. 3A

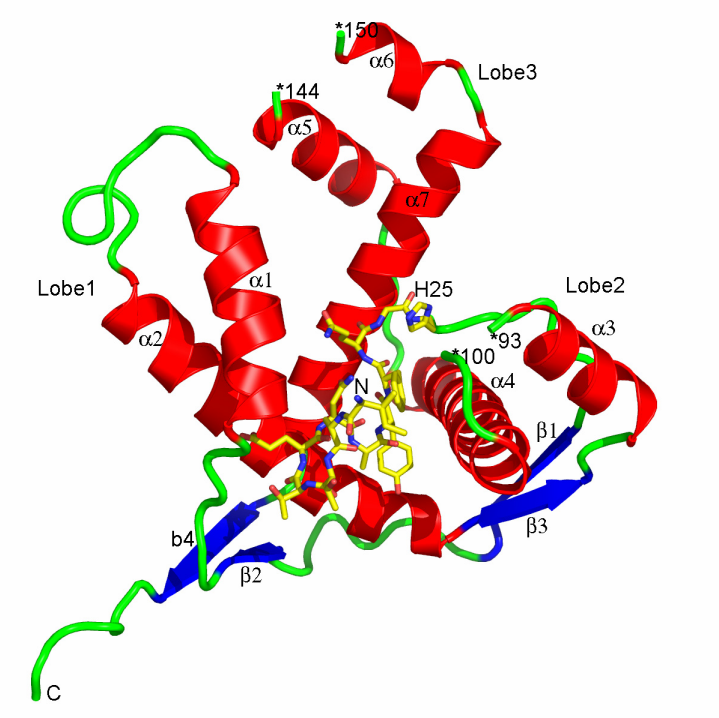
*150
α6
Lobe3
*144
α5
α7
Lobe1
α1
H25
Lobe2
α2
α3
*93
*100
α4
N
β1
b4
β3
β2
C

Fig. 3B

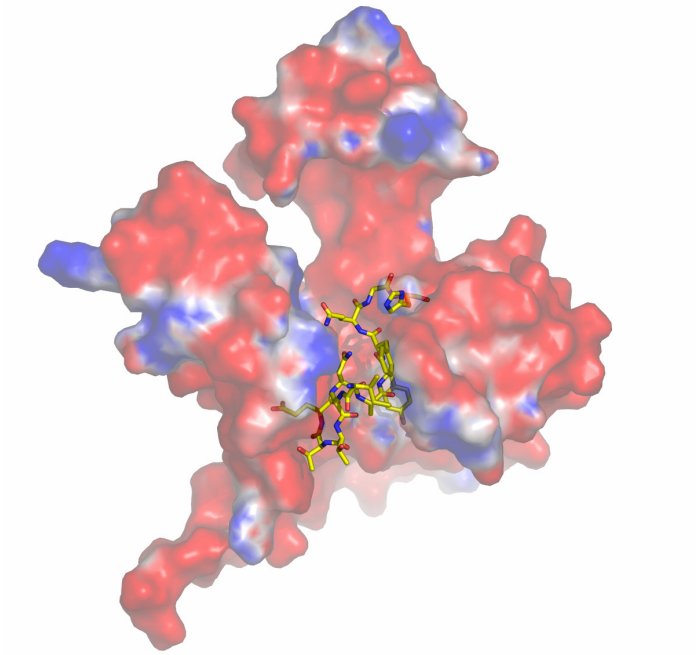

Fig. 4

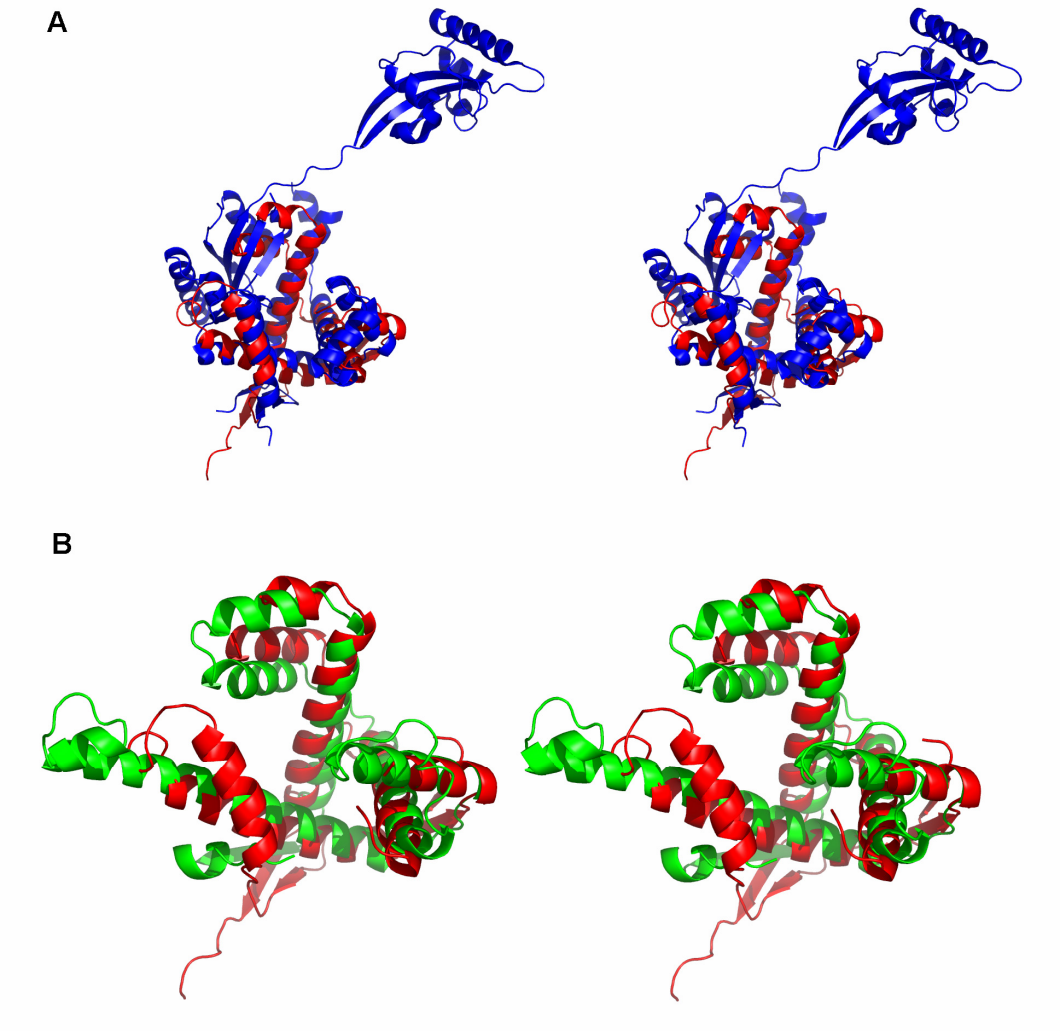

Figure legends

Fig. 1

Typical electron density map centered on E176. The 2Fo-Fc map was contoured at 1.2σ. This illustration and all other ribbon and surface drawings were prepared using Pymol (DeLano, 2002).

Fig. 2

Secondary-structure annotated protein sequence of MPN555. The protein sequence of MPN555 is numbered from 1 to 193, the N-terminal His-tag from t1 to t25. α-helices are drawn as rectangles, β-sheets as zig-zag lines above the sequence.

Fig. 3

Overall fold and surface representation of MPN555. Fig. 3A shows the arrangement of secondary structures into three lobes in the MPN555 fold. A peptide from the N-terminal His-tag is bound in the central binding pocket. Fig 3B shows a surface representation of the same molecules. Negatively charged surfaces are colored in red, and positively charged surfaces in blue. Electrostatic surfaces were calculated using the program *APBS* (Baker *et al*., 2001) in *PYMOL*.

Fig. 4

Superposition of two structurally homologous proteins onto MPN555. Parts of the N-terminal and C-terminal segment of SurA (PDB code 1M5Y) were superimposed onto MPN555 with the program PDBSET from the CCP4 suite using matrices suggested by DALI search engine (Fig. 4A). The r.m.s.d. for 139 aligned residues is 4.7 Å. MPN555 is shown in red. The complete SurA protein is shown in blue. Fig 4B shows a similar superposition of the C-terminal domain of *E.coli* trigger factor (PDB code 1W26) onto

MPN555. The r.m.s.d. for 129 aligned residues is 3.5 Å. MPN555 is shown in red, the C-terminal domain of TF is shown in green.

Acknowledgements

We thank Dr. P. Adams for helpful discussions concerning the structure determination and refinement of the protein. The work described here was supported by the National Institute of Health grant GM 62412.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997). *Nucleic Acids Res*. **25**:3389-3402.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J. & McCammon, J.A. (2001). *Proc. Natl. Acad. Sci. USA* **98**, 10037-10041.

Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L. & Sonnhammer, E.L.L. (2000). Nucleic Acid Res. **28**, 263-266.

Bitto, E. & McKay, D.B. (2002). Structure 10, 1489-1498.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. & Warren, G.L. (1998). Acta Cryst. D54, 905-921.

Collaborative Computational Project, Nr. 4, (1994). *Acta Cryst*. D**50**, 760-763.

DeLano, W.L. The PyMOL Molecular Graphics System (2002). DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org

Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E. & Ban, N. (2004). Nature 431, 590-596.

Holm & Sander (1996). Methods Enzymol. 266, 653-62.

Grosse-Kunstleve & Adams, P.D. (2003). *Acta Cryst*. D**59**, 1966-1973.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst*. **24**, 409-411.

Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). *Acta Cryst*. A**47**, 110-119.

Kim, R., S.J. Sandler, S. Goldman, H. Yokota, A.J.Clark, and S-H. Kim (1998). *Biotechnol. Lett*. **20**: 207-210.

Kramer, G., Patzelt, H., Rauch, T., Kurz, T.A., Vorderwulbecke, S., Bukau, B. & Deuerling, E. (2004). *J. Bacteriol*. **186**, 3777-3784.

Murshudov,G.N., Vagin, A.A. & Dodson, E.J. (1997). *Acta Cryst*. **D53**, 240-255.

Terwilliger, T.C. (2002). *Acta Cryst.* D**58**, 1937-1940