Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the putative DNA-binding protein SP_1288 from *Streptococcus pyogenes*

Vaheh Oganesyan,[a] Ramona Pufan,[a] Andrew DeGiovanni,[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

Correspondence e-mail: shkim@lbl.gov

Received 4 February 2004
Accepted 19 April 2004

PDB Reference: SP_1288, 1s7o, r1s7osf.

The crystal structure of the putative DNA-binding protein SP_1288 (gi|15675166, also listed as gi|28895954) from *Streptococcus pyogenes* has been determined by X-ray crystallography to a resolution of 2.3 Å using anomalous diffraction data at the Se peak wavelength. SP_1288 belongs to a family of proteins whose cellular function is associated with the signal recognition particle; no structural information has been available until now about the members of the family. Crystallographic analysis revealed that the overall fold of SP_1288 consists exclusively of α-helices and that 75% of the structure has good similarity to domain 4 of the σ subunit of RNA polymerase. This suggests its possible involvement in the biochemical function of transcription initiation, which includes interaction with DNA.

1. Introduction

SP_1288 is annotated as a putative DNA-binding protein in the genomic sequence database for *Streptococcus pyogenes* SSI-1. However, when the amino-acid sequence of SP_1288 was entered into the PFAM 11.0 database (November, 2003) of protein families, 28 similar sequences referred to as 'putative helix–turn–helix proteins' from the uncharacterized protein family UPF0122 (accession No. PF04297) were found. Genes from members of this family are often part of operons that encode components of the signal recognition particle (SRP), which in turn is involved with translation. These proteins may also regulate the expression of operons related to the SRP function. Of all the members of the UPF0122 family, to which SP_1288 belongs, a 13 kDa protein called p13 is probably the best characterized (Samuelsson *et al.*, 1997). It was identified as a *Mycoplasma mycoides* protein and its gene is located immediately downstream of a protein homologous to *Escherichia coli* FtsY, a GTPase involved in bacterial signal recognition particle (SRP) function. Sequence comparisons between UPF0122 family members (Fig. 1*a*) as well as between SP_1288 and p13 (Fig. 1*b*) suggest a functional similarity between these proteins. In many bacteria, p13 is encoded by the same operon as another protein homologous to the 54 kDa subunit of SRP, further supporting the idea that there may be a functional relationship between SP_1288 and SRP.

After the structure of SP_1288 had been solved, a structure-homology search using *DALI* (Holm & Sander, 1996) revealed that 75% of the structure comprising the N-terminal 80 residues had a good resemblance to domain 4 of the RNA polymerase σ subunit (PDB codes 1or7, 1ku3 and 1ku7), with a Z score of about 7.8 (Campbell *et al.*, 2002). This domain is

comprised of two extremely conserved regions, 4.1 and 4.2. Domain 4 of the σ subunit interacts with DNA during the initial step of specific RNA polymerase–promoter interaction in transcription initiation (Eichenberger *et al.*, 1997) and this interaction persists throughout complex formation. Since the domain of SP_1288 that has the highest sequence similarity with members of UPF0122 family resembles a portion of the σ subunit of RNA polymerase, the structure suggests that there may be a connection between SRP function and bacterial transcription initiation.

2. Materials and methods

2.1. Cloning, expression and purification

The DNA that encodes SP_1288 was cloned into pB3, an in-house expression vector under control of the T7 promoter, using the ligase-independent cloning (LIC) system (Aslanidis & de Jong, 1990). Six histidine residues were added to the N-terminus of the protein through a six-glycine linker for metal-affinity purification.

SP_1288 fusion protein was expressed in *Escherichia coli* strain BL21 (DE3) Star-pSJS1244 (Kim *et al.*, 1998) using an auto-inducible selenomethionyl (SeMet) medium (Dr William Studier, Brookhaven National Laboratory, personal communication). Cells were disrupted by microfluidization (Microfluidics, Newton, MA, USA) in 50 m*M* HEPES pH 7, 300 m*M* NaCl, 1.0 m*M* PMSF, 10 µg ml^{-1} DNase, 0.1 µg ml^{-1} antipain, 1 µg ml^{-1} chymostatin, 0.5 µg ml^{-1} leupeptin and 0.7 µg ml^{-1} pepstatin A. Cell debris was pelleted by centrifugation at 39 000*g* for 20 min at 277 K. The supernatant was then spun in an ultracentrifuge at 60 000*g* for 40 min at 277 K to remove membrane proteins. The fusion protein was affinity-purified using two 5 ml HiTrap Chelating HP columns in series on an ÄKTA Explorer chromatography system (Amersham Biosciences Corp., Piscataway, NJ, USA). The column was equilibrated in 50 m*M* HEPES pH 7.0, 300 m*M* NaCl. The SP_1288 fusion protein was eluted with a 10–400 m*M* linear gradient of imidazole in the same buffer over 13 column volumes. The protein was concentrated to 12 mg ml^{-1} 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, 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 Niquel, 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. Diffraction-quality crystals were obtained from a 1:1 mixture of the protein solution and a reservoir solution containing 0.1 *M* citric acid pH 3.2 and 2.0 *M* ammonium nitrate, which yielded 0.05 × 0.05 × 0.1 mm rhombohedral shaped crystals in 4 d. Prior to flash-cooling, the crystals were soaked for a fraction of a second in reservoir solution containing 10% glycerol. The peak-wavelength (λ = 0.9794 Å) 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 System Co. (ADSC, Poway, CA, USA) was used. The crystal-to-detector distance was set to 180 mm.

Table 1
X-ray diffraction data statistics.

Values in parentheses are for the outermost resolution shell.

Wavelength (Å)	0.9794
Resolution (Å)	38.35–2.31 (2.38–2.31)
Redundancy	3.64 (3.14)
Unique reflections (with Bijvoet pairs)	40607 (4066)
Completeness (%)	100 (99.8)
$I/\sigma(I)$	18.9 (4.5)
$R_{\rm sym}$† (%)	6.6 (33.6)

† $R_{\rm sym} = \sum_h \sum_i |I_{h,i} - \langle I \rangle| / \sum_h \sum_i |I_{h,i}|$.

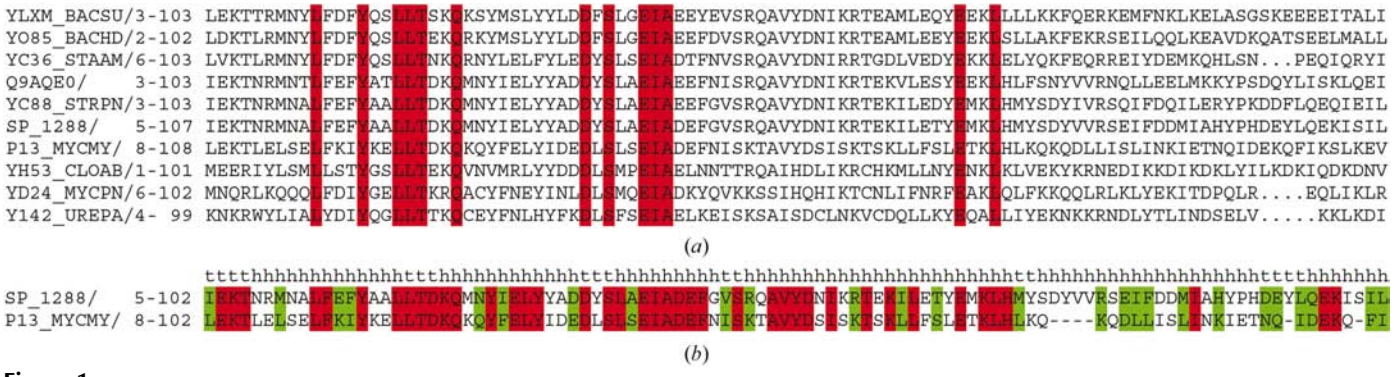

Figure 1
Amino-acid sequence alignment of several members of the UPF0122 family. The representatives of the family were chosen by PFAM 11.0 (http://pfam.wustl.edu; November 2003, 7255 families). (*a*) The 13 identical residues shown in red are conserved throughout the UPF0122 family. (*b*) Amino-acid sequence alignment between p13 and SP_1288. The pairwise alignment reveals many identical and similar residues (over 40% identity and over 70% similarity). Identical residues are shown in red and similar residues are shown in green. The alignment was generated using the program *CLUSTALW* (Altschul *et al.*, 1997). The secondary-structure elements (t for turns and h for helices) are shown on top of the SP_1288 sequence in (*b*).

In total, 360 images were collected (180 images for both direct and inverse beam) in 36 wedges with 1° oscillation range per image. 2.3 Å resolution data were processed using the program *HKL*2000 (Otwinowski & Minor, 1997). The crystal belongs to the monoclinic space group *C*2, with unit-cell parameters $a = 127.608$, $b = 69.685$, $c = 55.247$ Å, $\beta = 103.04°$. With three molecules in the asymmetric unit, the volume fraction of the unit cell occupied by protein atoms is 58% and the Matthews coefficient is 2.9 Å^3 Da^{-1}. A least-squares straight line from the Wilson plot approximates the *B* factor as around 39 Å^2. After being flash-cooled, the mosaicity of the crystal was 0.5°. The data statistics are shown in Table 1. The total number of reflections used for anomalous scaling was 156 703.

2.3. Structure determination and refinement

15 out of 21 possible Se-atom positions were located using *SOLVE* (Terwilliger, 2002), with a figure of merit (FOM) of 0.36 and a score of 70.39 at 2.3 Å resolution. The *RESOLVE* program found one twofold non-crystallographic symmetry (NCS) axis. The FOM after NCS averaging and density-modification procedures was 0.57. The resulting electron density allowed the tracing of 105 residues out of 113 per molecule, including side chains. The tracing and model building was performed using *O* (Jones *et al.*, 1991). The model was built for one molecule only. Since the expected number of molecules per asymmetric unit according to the Matthews coefficient was three, the coordinates of the built monomer were used as an input search model for the molecular-replacement program *AMoRe* (Navaza, 1994). Two further molecules were found. Refinement was carried out using *REFMAC*5 from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) at full resolution. Progressive improvement of the electron-density map using rounds of refinement and manual building was achieved (Fig. 2). The NCS constraints and restraints were completely released during the refinement. 5% of the data were randomly selected for free *R* factor cross-validation. The refinement statistics are shown in Table 2.

Table 2
Phasing and refinement statistics of the SAD X-ray diffraction data.

Phasing statistics	
Unit-cell parameters (Å, °)	$a = 127.61$, $b = 69.69$, $c = 55.25$, $\beta = 103.04$
Se atoms found per AU	15
FOM (figure of merit)	0.36
FOM after density modification	0.57
Refinement statistics	
Unique reflections (Bijvoet pairs merged)	19476
R factor (%)	20.5
R_{free}† (%)	23.9
Residues per monomer	*A*, 104; *B*, 105; *C*, 106
Residues per asymmetric unit	315
No. atoms per AU	2766
Protein	2670
Water	96
R.m.s. deviations	
Bond distances (Å)	0.016
Angle distances (°)	1.54
Ramachandran plot statistics	
Residues in most favored regions	292 (95.4%)
Residues in additional allowed regions	12 (3.9%)
Residues in disallowed regions	0

† R_{free} calculated as *R* factor but on 5% of data excluded from refinement.

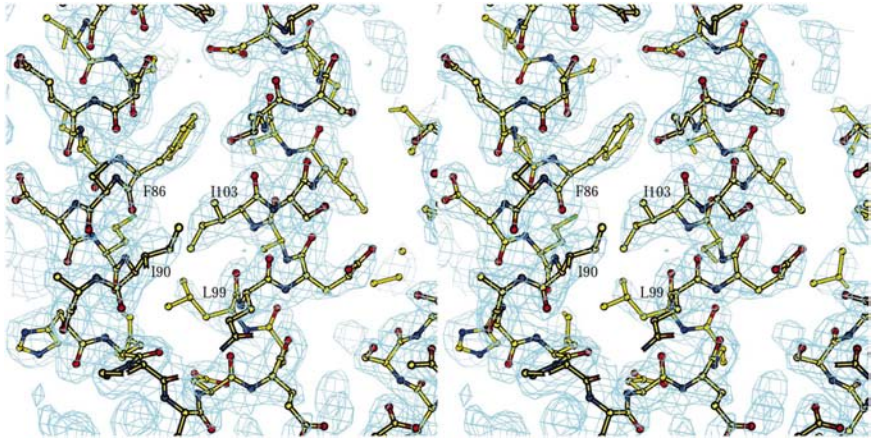

Figure 2
Electron-density map around the residues that are responsible for intramolecular hydrophobic interactions. The contour level corresponds to 1σ. Water molecules are not shown. Illustration prepared using *BOBSCRIPT* (Esnouf, 1997, 1999).

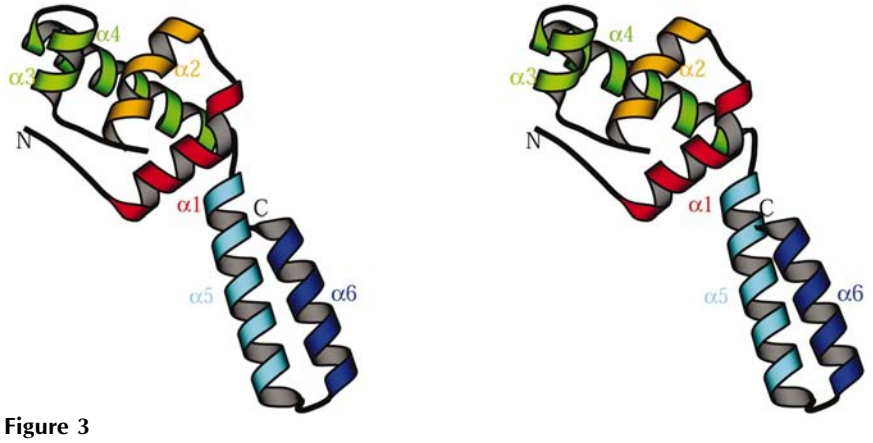

Figure 3
Overall fold of the SP_1288 monomer, with numbered helices. The domain that consists of helices $\alpha1$–$\alpha4$ has fold similarity with the $\sigma4$ subunit of RNA polymerase. The specific DNA molecule binds to $\sigma4$ through an interface that includes eight amino acids from helices $\alpha2$ and $\alpha4$. In this structure, helices $\alpha5$ and $\alpha6$ are responsible for primary crystal packing interactions and are absent in the σ-factor family of proteins. Illustration prepared using *MOLSCRIPT* (Kraulis, 1991).

3. Results and discussion

3.1. Overall structure

SP_1288 consists of 113 residues per molecule. Despite its small size, two domains with an overall helix–turn–helix topology can be distinguished. The N-terminal 74 residues comprise the first domain, which shows fold similarity to

domain 4 of the RNA polymerase σ subunit. The structural core of this domain consists of four compactly folded α-helices (α1, residues 10–20; α2, 24–38; α3, 40–49; α4, 51–73). The second domain consists of only two helices (α5, residues 75–94; α6, 97–111). The overall fold with numbered α-helices is shown in Fig. 3. The three molecules in the asymmetric unit are essentially identical, with the root-mean-square displacement of C^{α} positions for residues 7–111 (with well defined electron density) with respect to molecule *B* being 0.986 Å (maximum displacement 2.541 Å) for molecule *A* and 0.845 Å (maximum displacement 1.952 Å) for molecule *C* (Fig. 4). Those structural differences are most likely to be attributable to a different crystal-packing environment. The non-crystallography symmetry operators are two imperfect twofold axes with rotation angles 178.73 and 178.66°.

The primary interaction in crystal packing occurs between the second domains. The hydrophobic residues Ile85, Phe86, Met89, Ile90, Leu99, Ile103 and Leu106 in α5 and α6 are positioned in such a way that they not only hold those helices close to each other within a monomer, but also create an extensive hydrophobic core with mates from a symmetry-related molecule. The stacking interaction of symmetry-related Arg82 guanidyl groups, which are held in position by salt bridges with Asp110, contribute to both intramolecular and intermolecular interactions (Fig. 5).

Since the dynamic light scattering showed only one peak at a protein concentration of about 1 mg ml^{-1} with a polydispersity of 20%, we assume that the protein sample is monodisperse in its dimeric state. This is further supported by the extensive buried contact surface between two monomers observed in the crystal structure.

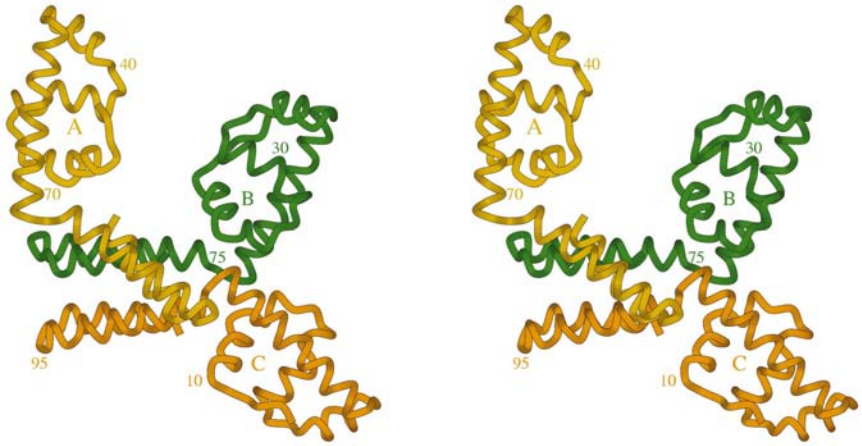

Figure 4
Asymmetric unit content for the SP_1288 crystal structure. *B* and *C* molecules interact through helices α5 and α6. In crystal packing, monomer *A* interacts with its symmetry mate through the same interface. Two imperfect twofold non-crystallographic axes describe this ensemble. Illustration prepared with *MOLSCRIPT* (Kraulis, 1991).

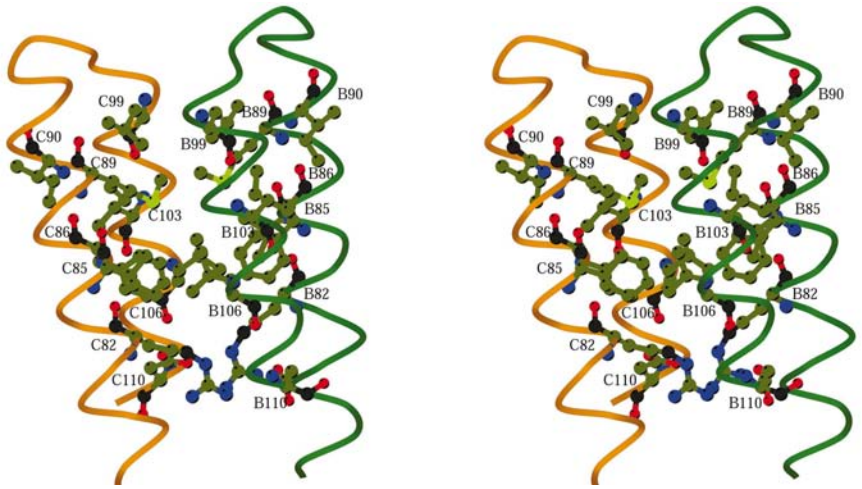

Figure 5
Detailed view of major crystal-packing interactions. The hydrophobic residues Ile85, Phe86, Met89, Ile90, Ile103 and Leu106 create both an intermolecular and intramolecular hydrophobic core. The salt bridge between Arg82 and Asp110 contributes to the α5–α6 interaction, whereas the stacking interaction of two Arg82 guanidyl groups from adjacent molecules contributes to the intermolecular interaction. Illustration prepared using *MOLSCRIPT* (Kraulis, 1991).

3.2. Comparison with similarly folded structures

After refinement of the model of the structure of SP_1288 had converged and the final model had been produced, a structure-homology search using *DALI* (Holm & Sander, 1996) produced a number of hits with a *Z* score ranging from 7.8 to 2. Several hits with the highest *Z* scores are presented in Table 3. The two highest *Z* scores of 7.8 and 7.7 were found with bacterial alternative and primary σ factors, respectively (PDB codes 1or7 and 1ku3). The σ factors are the key regulators of transcription initiation in bacteria. They bind the catalytically competent core RNA polymerase (RNAP) and direct the resulting holoenzyme to specific subsets of promoters. The central role of σ factors in transcription initiation includes interaction with DNA and proteins. The alternative σ factor, σ^{E}, the C-terminal domain of which was found to have a structure similar to SP_1288, responds to stress in the periplasmic compartment by activating the expression of proteins that mitigate the effects of unfolded proteins (Campbell *et al.*, 2003). The primary σ factor, σ^{A}, directs the bulk of transcription during exponential growth. Within σ^{A}, the structural domain σ_4, which consists of the conserved region 4.1–4.2, resembles the structure of SP_1288 (Campbell *et al.*, 2002). The superimposed structures of SP_1288, σ^{E} and σ_4 are shown in Figs. 6(*a*) and 6(*b*). The structure comparison of SP_1288 was also performed using coordinates of σ_4 complexed with −35 element DNA (PDB code 1ku7). The protein part of the complexed structure was

Table 3
First seven PDB entries found by *DALI* to be the most structurally similar to SP_1288.

No.	STRID2†	*Z*‡	RMSD§	LALI¶	LSEQ2††	%IDE‡‡	PROTEIN§§
1	1or7 *A*	7.8	2.4	59	181	19	RNA polymerase sigma factor
2	1ku3 *A*	7.7	2.0	55	61	18	Sigma factor mutant
3	1ofc *X*	6.0	2.3	64	265	5	Imitation swi protein
4	1a04 *A*	5.4	2.7	53	205	15	NARL, fragment
5	1fse *A*	4.7	3.0	49	67	27	Gere
6	1fc3 *A*	4.6	3.2	65	119	12	Sporulation response
7	1i1g *A*	4.3	5.5	69	140	20	Transcriptional regulator lrpa

† PDB identifier of aligned structure with chain identifier. ‡ *Z* score, *i.e.* strength of structural similarity in standard deviations above that expected. § Positional root-mean-square deviation of superimposed C^{α} atoms in Å. ¶ Total number of equivalent residues. †† Length of the entire chain of the equivalent structure. ‡‡ Percentage of sequence identity over equivalent positions. §§ COMPND record from the PDB file of the aligned structure.

virtually identical to its non-complexed counterpart. We noticed several conserved amino acids between SP_1288 and σ_4 in the interface region where σ_4 interacts with DNA, supporting the annotation of 'putative DNA binding' for SP_1288.

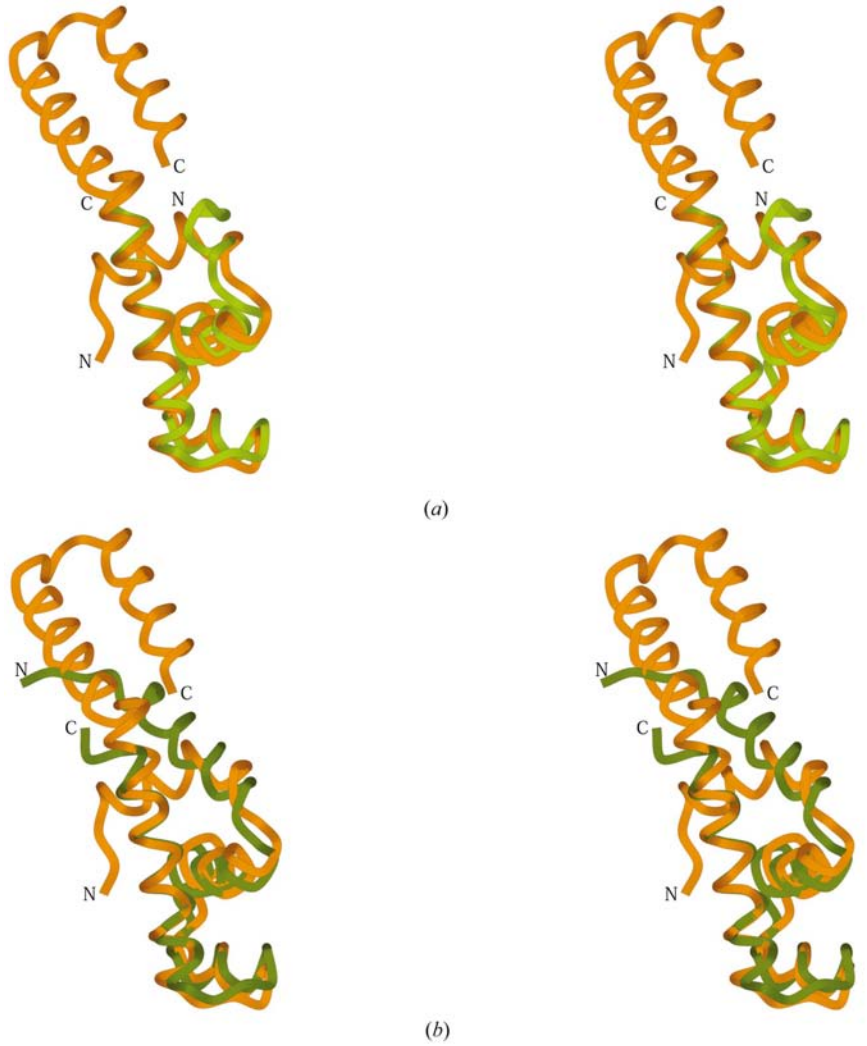

Figure 6
The SP_1288 structure superimposed with two σ factors found by *DALI* to be the closest structural homologues. Superimposition was performed with the *LSQKAB* program from the *CCP*4 suite using matrices that were suggested by *DALI* search engine. The SP_1288 model is shown in dark orange. (*a*) The C-terminal domain of alternative σ factor, σ^{E} (PDB code 1or7, shown in olive) has the highest homology with SP_1288. The superimposition was performed over C^{α} atoms. 59 equivalent residues from each structure yielded a *Z* score of 7.8 and an r.m.s.d. of 2.4 Å. (*b*) The structural domain σ4 of primary σ factor, σ^{A} (PDB code 1ku3, shown in yellow–green), has the next highest homology with SP_1288, yielding a *Z* score of 7.7 and an r.m.s.d. of 2.0 Å. The superimposition was performed over 55 C^{α} atoms. The illustration was prepared using *MOLSCRIPT* (Kraulis, 1991).

4. Conclusion

The crystal structure of the 'putative DNA-binding protein' SP_1288 from *S. pyogenes* is the first three-dimensional structure from the UPF0122 sequence family, some members of which are considered to be components of the SRP function. On the other hand, structural similarity to σ_4 and the C-terminal domain of σ^{E} of RNA polymerase strongly suggests the involvement of SP_1288 in bacterial transcription regulation. Based on these observations, we suggest that the cellular function of SP_1288 might be associated with the machinery of translation and transcription processes and that its molecular function may include interaction with a specific DNA sequence. This hypothesis is favored by the fact that the gene encoding SP_1288 in *S. pyogenes* is located upstream of the 'putative signal recognition particle' and downstream of the 'putative transcription regulator' (http://pedant.gsf.de; *S. pyogenes* SSI1). We further suggest that the association state of SP_1288 is a dimer.

We thank Dr Candice Huang for mass spectrometric analysis of the protein and the staff of Advanced Light Source beamline 5.0.2 at Lawrence Berkeley National Laboratory for help during diffraction data collection. We are grateful to J.-M. Chandonia for the bioinformatics search of the gene, B. Gold for cloning, and B. Martinez and M. Henriquez for technical help. This work was supported by Grant GM 62412 to the Berkeley Structural Genomics Center (http://www.strgen.org) from the

National Institute of General Medical Sciences, National Institutes of Health.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Aslanidis, C. & de Jong, P. J. (1990). *Nucleic Acids Res.* **20**, 6069–6074.
Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L. & Darst, S. A. (2002). *Mol. Cell*, **9**, 527–539.
Campbell, E. A., Tupy, J. L., Gruber, T. M., Wang, S., Sharp, M. M., Gross, C. A. & Darst, S. A. (2003). *Mol. Cell*, **11**, 1067–1078.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Eichenberger, P., Detiollaz, S., Buc, H. & Geiselmann, J. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 9022–9027.
Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–134.
Esnouf, R. M. (1999). *Acta Cryst.* D**55**, 938–940.
Holm, L. & Sander, C. (1996). *Science*, **273**, 595–602.
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., Sandler, S. J., Goldman, S., Yokota, H., Clark, A. J. & Kim, S.-H. (1998). *Biotechnol. Lett.* **20**, 207–210.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Samuelsson, T., Macao, B. & Bolske, G. (1997). *Biochem. Biophys. Res. Commun.* **231**, 839–843.
Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1937–1940.