Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

The structure of SAV1646 from *Staphylococcus aureus* belonging to a new 'ribosome-associated' subfamily of bacterial proteins

Yuri N. Chirgadze,[a]* Teresa E. Clarke,[b] Vladimir Romanov,[b] Gera Kisselman,[b] Jean Wu-Brown,[b] Maria Soloveychik,[b] Tiffany S. Y. Chan,[b] Roni D. Gordon,[b] Kevin P. Battaile,[c] Emil F. Pai[b,d] and Nickolay Y. Chirgadze[b,e]*

[a]Institute of Protein Research, Russian Academy of Sciences, Puschino 142290, Moscow Region, Russian Federation, [b]Campbell Family Institute for Cancer research, Princess Margaret Cancer Center, University Health Network, Toronto, ON M5G 2C4, Canada, [c]Hauptman–Woodward Medical Research Institute, IMCA-CAT, Advanced Photon Source, Argonne National Laboratory, Argonne, IL 60439, USA, [d]Departments of Biochemistry, Molecular Genetics, and Medical Biophysics, University of Toronto, Toronto, ON M5S 1A8, Canada, and [e]Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON M5S 1A8, Canada

Correspondence e-mail: chir@vega.protres.ru, nchirgadze@gmail.com

Received 2 July 2014
Accepted 23 November 2014

PDB reference: SAV1646, 2p92

The crystal structure of the SAV1646 protein from the pathogenic microorganism *Staphylococcus aureus* has been determined at 1.7 Å resolution. The 106-amino-acid protein forms a two-layer sandwich with α/β topology. The protein molecules associate as dimers in the crystal and in solution, with the monomers related by a pseudo-twofold rotation axis. A sequence-homology search identified the protein as a member of a new subfamily of yet uncharacterized bacterial 'ribosome-associated' proteins with at least 13 members to date. A detailed analysis of the crystal protein structure along with the genomic structure of the operon containing the *sav1646* gene allowed a tentative functional model of this protein to be proposed. The SAV1646 dimer is assumed to form a complex with ribosomal proteins L21 and L27 which could help to complete the assembly of the large subunit of the ribosome.

1. Introduction

Large-scale genomic sequencing and structural proteomics projects have provided new and very valuable information (Montelione & Anderson, 1999). A result of these efforts is an increasing number of protein sequences and structures with unknown biological function. The number of uncharacterized proteins has now reached more than several thousand. One such database that contains proteins with well determined three-dimensional structures but unknown functions is the PDB-UF database (von Grotthuss *et al.*, 2006). Unfortunately, in most cases, putative functional information still cannot be obtained automatically, and a comprehensive analysis of all available data is required. The function of a newly determined protein structure may be inferred using different approaches. Among the most commonly used are the sequence-homology search tool *BLAST* (*Basic Local Alignment Search Tool*; Altschul *et al.*, 1997; http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) and a search of conserved structural motifs (Panchenko, 2004; Nimrod *et al.*, 2008). However, a decision on the protein function can face problems if the pairwise sequence residue identity is less than 20%. In fact, our recent experience shows that in such cases specific approaches are required (Clarke *et al.*, 2011; Gordon *et al.*, 2013). One of the most intriguing strategies is to establish the putative function of 'life's essential' proteins as identified by gene-knockout experiments. Here, we have studied such a protein from the pathogenic bacterium *Staphylococcus aureus*. Since this bacterium can cause problematic infections, the results may be very important not only for fundamental science but also to a greater extent for medicine and hospital practices.

Robust protein biosynthesis in the cell is determined by the accurate assembly of the ribosome particles (Wilson & Nierhaus, 2007). Although the main principles of this process are

well known for *Escherichia coli* and other bacteria (Spirin, 2011), some important but as yet still unknown details of the process remain. Numerous homologous 'ribosome-associated' sequences with uncharacterized functions are found in GenBank. One of them, the *sav1646* gene from *S. aureus*, is located in an operon with adjacent genes for ribosomal proteins L21 and L27, which are part of the large ribosomal subunit. Using the *BLAST* search tool several homologous bacterial sequences were found, which are assigned as 'ribosome-associated' although their biological functions remain unknown.

Here, the atomic structure of the SAV1646 protein from the pathogenic bacterium *S. aureus* was considered. Structural analysis of the crystal dimer together with the structure of the operon containing the *sav1646* gene infers a possible model of biological function which seems to be related to the process of assembly of the ribosome 50S particle. At present, there are three similar protein structures with uncharacterized function from *Streptococcus mutans*, *S. pneumoniae* and *Thermotoga maritima* (Midwest Center for Structural Genomics, unpublished work; H.-F. Hou, Z.-Q. Gao, L.-F. Li, Y.-H. Liang, X.-D. Su & Y.-H. Dong, unpublished work; Shin *et al.*, 2005), which have rather low average sequence identities of 28, 25 and 20% to SAV1646 from *S. aureus*, respectively. The current analysis of the new structure in this communication provides a basis for extending the proposed function to a group of uncharacterized related bacterial proteins that are similar in sequence and structure. As a result, a novel 'ribosome-associated' subfamily of bacterial proteins, with at least 13 members to date, has now been characterized.

2. Materials and methods

2.1. Cloning, expression and purification of recombinant protein

The *sav1646* gene from *S. aureus* subsp. *aureus* Mu50 (GenBank NCBI reference sequence NP_372170.2) is assigned to superfamily DUF464, where DUF simply means 'domain of unknown function' according to GenBank annotation. The *sav1646* gene was amplified from genomic DNA (ATCC) and cloned into a modified pET-15b vector (Novagen, Madison, USA) using standard molecular-biology techniques. The resulting vector was transformed into *E. coli* BL21-CodonPlus (DE3)-RIPL cells (Stratagene, La Jolla, USA). The cells were grown on selenomethionine medium (Medicilon, California, USA) in 1 l Tunair flasks at 37°C to an OD_{600} of 1.2, at which point the temperature was lowered to 16°C and IPTG was added to 0.5 m*M*. Expression was allowed to proceed overnight, after which the cells were harvested by centrifugation, flash-cooled in liquid nitrogen and stored at −80°C.

The cells were thawed on ice and resuspended in binding buffer consisting of 50 m*M* HEPES pH 7.5, 500 m*M* NaCl, 5% glycerol, 0.2 m*M* TCEP supplemented with 0.5% CHAPS, 0.25 m*M* phenylmethylsulfonyl fluoride, 0.5 m*M* benzamidine. After disruption by sonication and centrifugation at 40 000*g* for 60 min, the cell-free extracts were passed through a 2.6 × 7 cm DE-52 column that had been pre-equilibrated with the same buffer and were then loaded by gravity flow onto an Ni–nitrilotriacetic acid (Ni–NTA) column (Qiagen, Madison, USA). The column was washed with 30 volumes of wash buffer (50 m*M* HEPES pH 7.5, 500 m*M* NaCl, 5% glycerol, 20 m*M* imidazole, 0.2 m*M* TCEP) supplemented with 0.5% CHAPS followed by ten volumes of wash buffer. The His_6-tagged SAV1646 protein was eluted with the same buffer containing 250 m*M* imidazole. This sample was concentrated using a Vivaspin unit (Sartorius NA, New York, USA) and loaded onto a 2.6 × 60 cm Superdex 200 column (GE Healthcare) equilibrated with gel-filtration buffer (10 m*M* HEPES pH 7.5, 250 m*M* NaCl, 0.2 m*M* TCEP). The eluate apparently contained the protein in a dimeric form (data not shown). The final protein sample was concentrated, divided into 1.5 mg aliquots, flash-cooled and stored at −80°C.

2.2. Protein crystallization

The protein preparation was thawed on ice and diluted to 15 mg ml^{-1} with gel-filtration buffer. The solution was then centrifuged at 14 000 rev min^{-1} for 10 min at 20°C. Crystals were obtained after a few days by sitting-drop vapour diffusion against a reservoir solution consisting of 31%(*v*/*v*) PEG 400, 0.2 *M* magnesium chloride, 0.1 *M* HEPES pH 7.2 in CrystalClear strips (Hampton Research, California, USA) at 20°C. The crystals were 'flash-cryocooled' in a mixture of one part Paratone and one part mineral oil.

2.3. X-ray data collection and structure determination

Diffraction data were collected to 1.7 Å resolution at 100 K from one crystal on beamline 17-ID at the Advanced Photon Source, Argonne National Laboratory. X-ray diffraction data were processed with the *HKL*-2000 program suite (Otwinowski & Minor, 1997). Phases were determined by molecular replacement using *MOLREP* (Vagin & Teplyakov, 2010) from the *CCP*4 program suite (Winn *et al.*, 2011). The coordinates of a conserved uncharacterized protein from *S. pneumoniae* were used as a search model (PDB entry 2idl; Midwest Center for Structural Genomics, unpublished work). Although the protein was labelled with selenomethionine, it was not necessary to use the anomalous signal to phase the data. Following initial rigid-body refinement, interactive cycles of model building and refinement were carried out using *Coot* (Emsley *et al.*, 2010), *REFMAC* (Murshudov *et al.*, 2011) and *BUSTER-TNT* (Bricogne *et al.*, 2011). The electron density for the protein main chain was continuous, with the exception of residues 22–31 in chain *A* and residues 23–32 in chain *B*. These chain fragments seem to be assigned to freely moving loops outside the main compact body of the protein molecule. The electron density of the N-terminal His tag of the first chain was not visible and only several residues of the N-terminal His tag of the second chain could be modelled. Methionine residues were replaced with selenomethionine in the model. The data-collection and refinement statistics are presented in Table 1.

2.4. Sequence-homology search

The sequence homology of the SAV1646 protein was investigated using the protein *BLAST* search program (Altschul *et al.*, 1997).

3. Results and discussion

3.1. Protein structure

The protein monomer contains 106 amino-acid residues. The topology of the peptide-chain pathway is $\beta\beta\alpha\beta\beta\alpha\beta$. The molecule comprises two layers: one is an antiparallel β-sheet of five strands and the other includes two α-helices (Fig. 1*a*). The β-strands along the protein chain are $\beta1$ (residues 3–9), $\beta2$ (residues 13–30), $\beta3$ (residues 56–62), $\beta4$ (residues 66–77) and $\beta5$ (residues 101–105). The α-helical layer comprises two elongated helices $\alpha1$ (residues 32–52) and $\alpha2$ (residues 76–97) with lengths of 21 and 22 amino acids, respectively.

In the crystal, a protein dimer was found in the asymmetric part of the crystal unit cell. It consists of two protein monomers related by a pseudo-twofold rotation axis (Fig. 1*a*). Inside the dimer, the main contacts between monomers are formed through the α-helices. A significant part of the dimer interface is formed by a cluster of nonpolar residues. These nonpolar residues are located at the N-terminal end of helix $\alpha1$. They form nonpolar contacts inside molecule *A* (or *B*), as well as at the dimer interface between molecule *A* and molecule *B*. The external nonpolar contacting residues Ala35, Gly36, Ala39, Val40 and Gly43 are located inside a highly conserved sequence region (Fig. 1*b*).

An interesting feature of the protein surface is the strong polarization of the charged residues found on the two opposite sides of the monomer molecule. The major part of the α-helical side molecular surface is neutral, while the surface of the β-sheet side is occupied entirely by negatively charged carboxylic groups. There are six aspartic acid residues and one glutamic acid residue. Three of them, Asp20, Asp61 and Asp62, form a cluster. On the surface of the dimer structure, the negative charge is more evident (Fig. 1*a*). Both external opposite sides of the dimer, as viewed from the side of molecule *A* or *B*, are negatively charged. These sides are pseudo-symmetrical.

3.2. Sequence and structural homology

The first homology *BLAST* search among proteins of known spatial structure identified three proteins similar to SAV1646. Two of these structures had sequence identities of 28 and 25%, with gap values of 5.6 and 6.6%, respectively. These were bacterial proteins with uncharacterized function from *S. mutans* and *S. pneumoniae* (PDB entries 2g0i and 2idl; Midwest Center for Structural Genomics, unpublished work; H.-F. Hou, Z.-Q. Gao, L.-F. Li, Y.-H. Liang, X.-D. Su & Y.-H. Dong, unpublished work). The remainder of the protein structures had even lower sequence identities, ranging from 15 to 21%, which were also observed in a limited sequence range of 30–70 residues. The third structure, of the hypothetical protein TM1457 from *T. maritima* (PDB entry 1s12; Shin *et al.*, 2005), had a similar peptide-chain fold to the structures 2g0i and 2idl but had a sequence identity of only 19% to the sequence of SAV1646. We chose structure 2idl as the search model for the solving the crystal structure of SAV1646. Thus, during this search, no suitable homologous proteins with both known structure and function were found.

Table 1
Crystallographic data and refinement statistics.

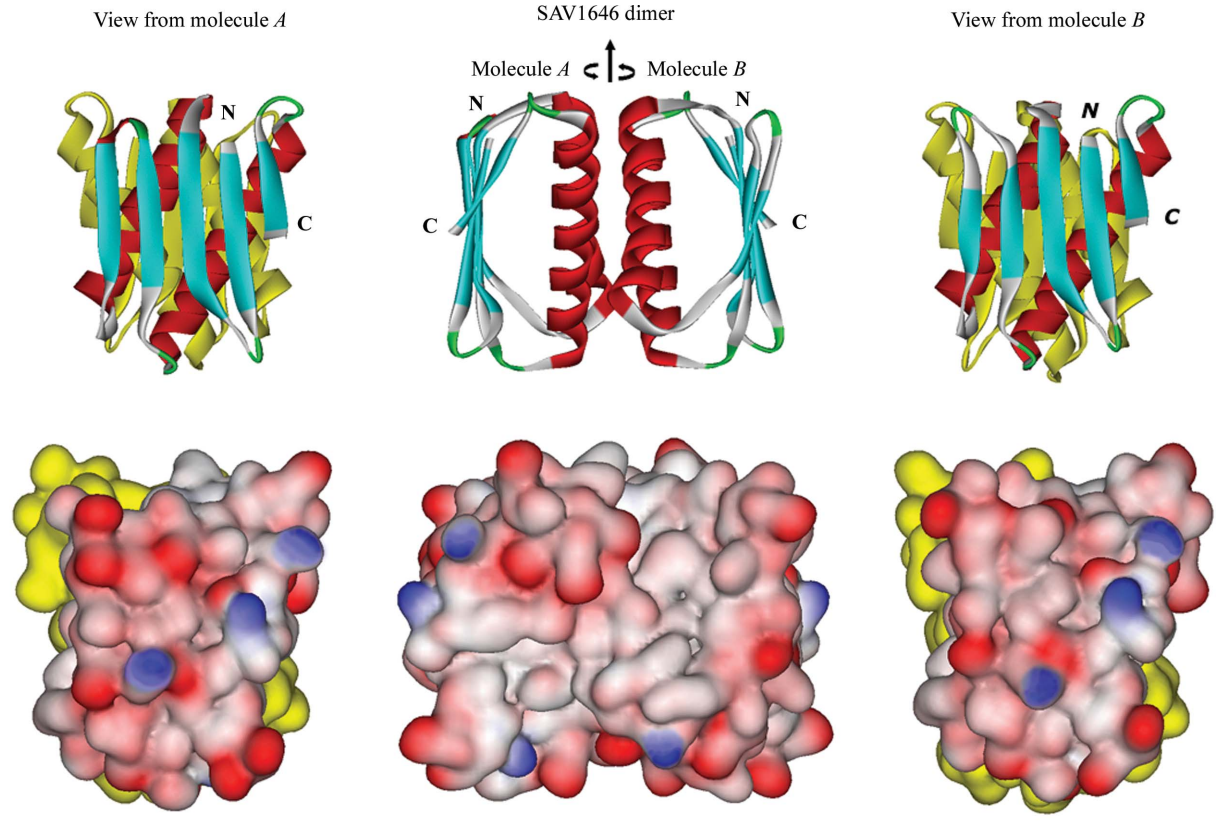
Values in parentheses are for the highest resolution shell.

Data collection	
Wavelength (Å)	1.00
Resolution (Å)	50.0–1.73 (1.79–1.73)
Space group	*H*3
Unit-cell parameters	
$a = b$ (Å)	67.2
c (Å)	127.3
Molecules per asymmetric unit	2
Unique reflections	22209 (2230)
Multiplicity	3.6 (3.3)
Average $I/\sigma(I)$	18.1 (3.4)
$R_{\rm merge}$	0.033 (0.359)
Completeness (%)	99.2 (99.5)
Refinement and structure statistics	
$R_{\rm work}$	0.186 (0.304)
$R_{\rm free}$	0.225 (0.336)
R.m.s.d. from ideal geometry	
Bond lengths (Å)	0.010
Bond angles (°)	1.060
No. of atoms	
Protein non-H atoms	1550
Water O atoms	140
Ramachandran plot statistics (%)	
Favoured regions	97.7
Allowed regions	2.3
Disallowed regions	0.0

Another homology search of all nonredundant sequences revealed 13 different bacterial sequences with residue identities from 25 to 90%. Nine of these sequences with a sequence homology to SAV1646 of greater than 25% are presented in Fig. 1(*b*). Here, five sequences, numbered from 2 to 6, were identified as 'ribosome-associated' but without any definite function. The most interesting one was the protein sequence from *S. lugdunensis*, which is annotated as 'ribosome-associated' like some others from the DUF464 superfamily and displays a very high sequence identity of 89% to SAV1646. This clearly indicates that SAV1646 also belongs to the new 'ribosome-associated' protein subfamily (Fig. 1*b*). It should be noted that only two proteins, SAV1646 and PDB entry 2idl in Fig. 1, have been characterized structurally at present. We selected the conserved residue positions with more than 78% average identity for all of the sequences: these are shown as white letters on a black background. It should be underlined that the common conserved region 30–40, numbered according to the SAV1646 sequence, belongs to the invariant dimer interface. This means that the biological function of the members of this subfamily is strongly determined by the dimeric molecule.

3.3. Hypothetical function of SAV1646

The operon containing the *sav1646* gene in the genome of *S. aureus* Mu50 (Kuroda *et al.*, 2001; Gill *et al.*, 2005) contains



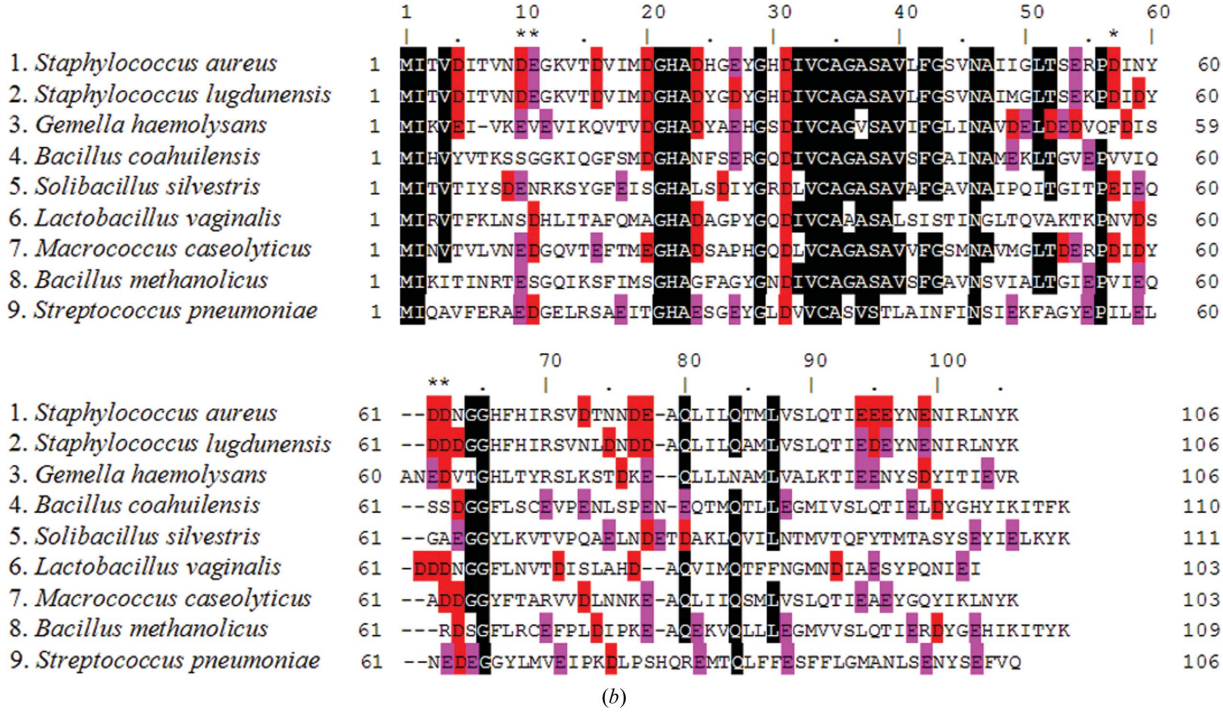


Figure 1
Crystal structure and sequence alignment of SAV1646 from *S. aureus*. (*a*) Crystal structure of the dimer of SAV1646. Upper row: pathway of the protein main chain with α-helices in red, β-strands in cyan and turns in green. The yellow colour corresponds to the second monomer, which is located further away from the viewer and is partially hidden by the first monomer. Lower row: electrostatic surfaces, with negative groups shown in red, positive groups in blue and neutral surfaces in white. (*b*) Sequence alignment of SAV1646 and other bacterial proteins with average residue identities from 90 to 25%. Conserved residue positions with identities of greater than 78% are shown as white letters on a black background. Negatively charged Asp and Glu residues are marked in red and magenta. The bacterial species numbered 2–6 were assigned as 'ribosome-associated' but without any definite function; the rest were also previously assigned to have an unknown function. With the exception of those from *S. aureus* and *S. pneumoniae*, all of the other bacterial proteins have no determined spatial structures.

two adjoining genes coding for the ribosomal proteins L21 and L27 of the large ribosomal subunit. In addition, the operon includes the *mreD* gene, a cell-shape determinant. This suggests that the SAV1646 protein belongs to the group of essential proteins related to formation of the cell. Since these genes are located in the operon together with SAV1646, this corroborates our assignment of this protein to the 'ribosome-associated' proteins.

The protein dimer was observed both in the crystal and in solution (see §2.1), suggesting high stability of the dimeric state. This means that the most probable biological unit is a dimer. The dominant negative charge on the dimeric surface suggests that at the physiological cell pH near 6–7 this dimer could form a stable complex with other proteins which have positively charged regions on their surfaces. It is possible that such a complex is formed between SAV1646 and the ribosomal proteins L21 and L27. Their genes are located inside the same operon just before and after the gene for SAV1646. This seems to help in the accurate assembly of the 50S ribosome subunit. Such a complex may shuttle the 'late-assembly' ribosomal proteins L21 and L27 to complete the assembly of the ribosome. If the structures of the ribosomal proteins from *S. aureus* were available, they could be used to confirm this hypothesis. Unfortunately, at present no exact data are available.

However, we can analyze the structures of the homologous ribosomal proteins L21 and L27 from *T. thermophilus* HB8 which are known from the crystal structure of the 70S ribosome in complex with mRNA and tRNA (Selmer *et al.*, 2006), and this could be used as a proxy. The possibility of charge interaction of two ribosomal proteins with SAV1646 is important for the proposed hypothesis of its function and we therefore consider this point in detail. The atomic coordinates of L21 (PDB entry 2j01, chain *V*) and L27 (PDB entry 2j01, chain *O*) were obtained and the pairwise sequence homologies of the ribosomal proteins L21 and L27 from *S. aureus* Mu50 and *T. thermophilus* HB8 were compared. The sequence-pair identities of these proteins with SAV1646 are 54 and 58% for L21 and L27, respectively. Note that the sequence positive similarity values, which also include the coincidence of polar and nonpolar residues, are much higher at 67 and 74%, respectively. This suggests homology of the structure of the ribosomal proteins from *T. thermophilus* and *S. aureus*.

An analysis of pairwise sequence homology for L21 and L27 from *S. aureus* and *T. thermophilus* shows similar clusters of positively charged arginine and lysine residues. The structures of both the L21 and L27 molecules are of a similar shape and consist of a large head and a long tail. The protein fold and surface, with the latter coloured according to electrostatic charge, are shown in Fig. 2. The L21 protein from *T. thermophilus* contains conserved clusters of positively charged amino acids that correspond to the positive clusters found in the *S. aureus* protein. Thus, the conserved charged residues which could bind SAV1646 are as follows: Lys70, Lys77, Arg79, Lys86 and Arg89 in *S. aureus* and Lys69, Lys76, Lys78, Lys85 and Arg 88 in *T. thermophilus*.

The structure of L27 is slightly different from that of L21. The N- and C-termini are not close together and there are no extended clusters of positively charged residues. The potential binding-contact residues of L27, however, could be as follows: Lys28 and Arg48 in *S. aureus* and Lys19 and Arg39 in *T. thermophilus*.

Owing to the clustering of oppositely charged residues, the SAV1646 protein dimer from *S. aureus* could form a complex with the ribosomal proteins L21 and L27. There seem to be two or more possible binding sites for L21 and L27 on the dimer surface.

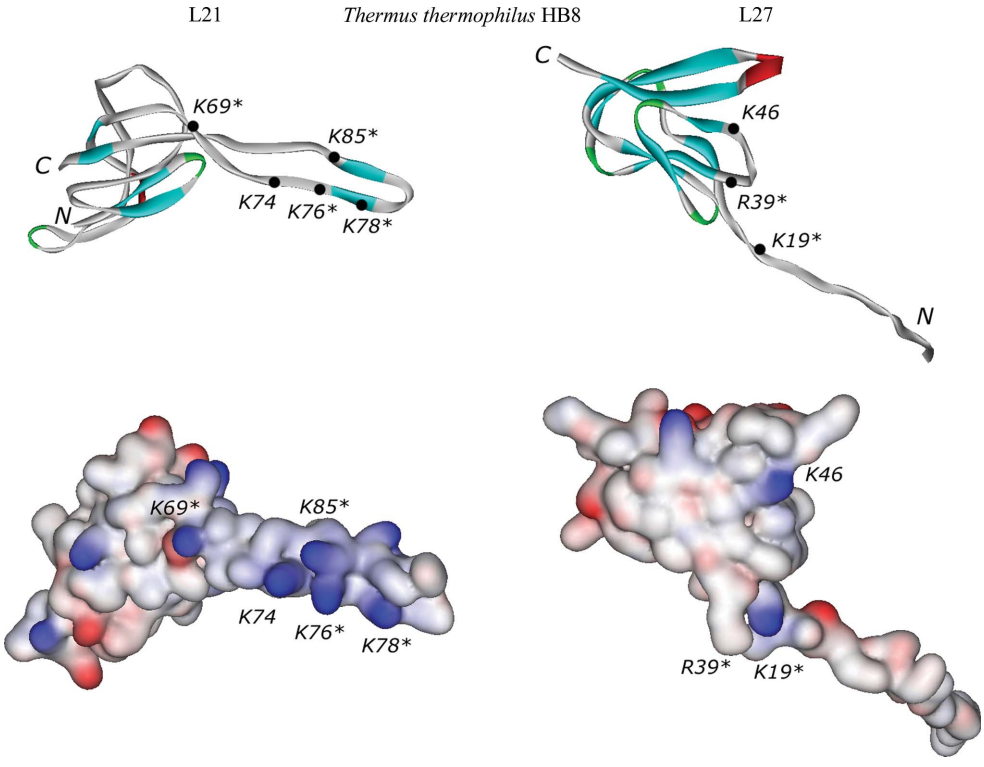


Figure 2
Protein main-chain folds and electrostatic surfaces of the positively charged ribosomal proteins L21 and L27 from *T. thermophilus* HB8. Positive and negative charges of side groups are marked in blue and red, respectively. Residues marked with an asterisk are proposed to be valuable for binding contacts with the SAV1646 protein dimer.

We thank Dr V. Ksenzenko, Dr V. Kolb, Dr V. Vasiliev and Dr S. Garbuzinsky from the Institute of Protein Research, Russian Academy of Sciences for valuable discussions. EFP acknowledges

support from the Canada Research Chairs program. YNC would like to express gratitude to his wife L. I. Chirgadze for her invaluable help and encouragement during this work. Contract grant sponsors: Ontario Research and Development Challenge Fund (99-SEP-0512), the Canada Research Chairs program, Hauptman–Woodward Medical Research Institute and the US Department of Energy, Office of Basic Energy Sciences (contract grant No. W-31-109-Eng38).

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.
Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O., Vonrhein, C. & Womack, T. (2011). *BUSTER* v.2.10.0. Global Phasing Ltd, Cambridge.
Clarke, T. E., Romanov, V., Chirgadze, Y. N., Klomsiri, C., Kisselman, G., Wu-Brown, J., Poole, L. B., Pai, E. F. & Chirgadze, N. Y. (2011). *BMC Struct. Biol.* **11**, 27.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Gill, S. R. *et al.* (2005). *J. Bacteriol.* **187**, 2426–2438.
Gordon, R. D., Qiu, W., Romanov, V., Lam, K., Soloveychik, M., Benetteraj, D., Battaile, K., Chirgadze, Y. N., Pai, E. F. & Chirgadze, N. Y. (2013). *J. Biomol. Struct. Dyn.* **31**, 1057–1065.
Grotthuss, M. von, Plewczynski, D., Ginalski, K., Rychlewski, L. & Shakhnovich, E. I. (2006). *BMC Bioinformatics*, **7**, 53.
Kuroda, M. *et al.* (2001). *Lancet*, **357**, 1225–1240.
Montelione, G. T. & Anderson, S. (1999). *Nature Struct. Biol.* **6**, 11–12.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Nimrod, G., Schushan, M., Steinberg, D. M. & Ben-Tal, N. (2008). *Structure*, **16**, 1755–1763.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Panchenko, A. R. (2004). *Protein Sci.* **13**, 884–892.
Selmer, M., Dunham, C. M., Murphy, F. V., Weixlbaumer, A., Petry, S., Kelley, A. C., Weir, J. R. & Ramakrishnan, V. (2006). *Science*, **313**, 1935–1942.
Shin, D. H., Lou, Y., Jancarik, J., Yokota, H., Kim, R. & Kim, S.-H. (2005). *J. Struct. Biol.* **152**, 113–117.
Spirin, A. S. (2011). *Molecular Biology: Ribosomes and Protein Biosynthesis*, pp. 133–159. Moscow: Akademia.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Wilson, D. N. & Nierhaus, K. H. (2007). *Crit. Rev. Biochem. Mol. Biol.* **42**, 187–219.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.