Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of serine acetyltransferase from *Haemophilus influenzae* Rd

Jason Gorman[a] and Lawrence Shapiro[b,c]*

[a]Department of Ophthalmology, Columbia University College of Physicians and Surgeons, 630 West 168th Street, New York, NY 10032, USA, [b]Department of Biochemistry and Molecular Biophysics, Columbia University College of Physicians and Surgeons, 630 West 168th Street, New York, NY 10032, USA, and [c]Naomi Berrie Diabetes Center, Columbia University College of Physicians and Surgeons, 630 West 168th Street, New York, NY 10032, USA

Correspondence e-mail: shapiro@convex.hhmi.columbia.edu

Received 2 February 2004
Accepted 22 June 2004

PDB Reference: serine acetyltransferase, 1s80, r1s80sf.

The crystal structure of serine acetyltransferase (SAT) from *Haemophilus influenzae* Rd determined at 2.7 Å resolution is presented. SAT is a member of a family of hexapeptide-containing transferases that contain six-residue tandem repeats (LIV)-G-X_4 that have been shown to form left-handed parallel β-helices. In the current structure, each protomer is comprised of two domains: an N-terminal α-helical domain and a C-terminal left-handed parallel β-helix domain. Although other members of this protein family are known to form trimeric structures, SAT forms a dimer of trimers in which the trimer interface is mediated through interactions between both the β-helix domains and N-terminal domains; these trimers dimerize through contacts in the N-terminal domain. All dimer-of-trimer interactions are mediated through amino acids within an N-terminal extension common only to a subset of SATs, suggesting that members of this subfamily may also adopt hexameric structures. Putative active sites are formed by crevices between adjacent protomers in a trimer. Thus, six independent active sites exist in the hexameric enzyme complex.

1. Introduction

Serine acetyltransferase (SAT; EC 2.3.1.30), encoded by the *cysE* gene, is an enzyme that has been shown to catalyze the conversion of acetyl-CoA and L-serine to CoA and *O*-acetyl-L-serine in microorganisms and higher plants (Kredich & Tomkins, 1966; Smith & Thompson, 1971). This reaction is the first in a two-step process of sulfur assimilation performed in the biosynthesis of cysteine. Both steps of the process are believed to be carried out by a cysteine synthase multienzyme complex consisting of SAT and *O*-acetylserine thiol-lyase (OAS-TL; Kredich *et al.*, 1969). SAT is a low-abundance enzyme relative to OAS-TL and has been shown to be the rate-limiting component in this reaction (Kredich *et al.*, 1969; Smith & Thompson, 1971). Cysteine, the process end-product and a negative-feedback inhibitor of SAT, regulates the flow of substrates through this pathway (Kredich *et al.*, 1969).

SAT belongs to a family of proteins known as hexapeptide transferases. This family, annotated PF00132 in Pfam (Bateman *et al.*, 2002), is known to contain tandem hexapeptide repeats with consensus sequence (LIV)-G-X_4; these repeat regions have been shown to adopt a left-handed parallel β-helix conformation (Raetz & Roderick, 1995). The β-helical domain of SAT is highly conserved among plants and microorganisms and is believed to be responsible for hetero-oligomerization with OAS-TL (Wirtz *et al.*, 2001; Bogdanova & Hell, 1997). Most enzymes containing the β-helix fold are reported to be trimeric (Beaman, Blanchard *et al.*, 1998; Beaman, Sugantino *et al.*, 1998; Wang *et al.*, 2002). However,

© 2004 International Union of Crystallography
Printed in Denmark – all rights reserved

biophysical studies of the quaternary structure of SAT from *Escherichia coli* suggest that this enzyme forms a hexamer with 32 symmetry consisting of a dimer of trimers (Hindson *et al.*, 2000). SAT from *E. coli* shares 71% sequence identity with SAT from *Haemophilus influenzae*, the crystal structure of which is presented here. The structure indeed reveals a hexameric dimer of trimers with 32 symmetry. Each trimer is similar to previously determined structures from this enzyme family, with the most highly related structures according to a *DALI* search (Holm & Sander, 1993) being 3tdt (tetrahydrodipicolinate *N*-succinyltransferase), 2xat (xenobiotic acetyltransferase) and 1kqa (galactoside acetyltransferase) (Beaman, Blanchard *et al.*, 1998; Beaman, Sugantino *et al.*, 1998; Wang *et al.*, 2002). The r.m.s. differences reported by *DALI* for these structures were 2.8 Å (with 118 C^{α} atoms superimposed), 2.9 Å (103 C^{α} atoms superimposed) and 1.7 Å (101 C^{α} atoms superimposed), respectively.

A conserved N-terminal region in SAT (SATase_N) from plants and bacteria is also annotated in Pfam (PF06426; Bateman *et al.*, 2002). The dimer of trimers is formed by interactions between α-helices within the ~110-amino-acid SATase_N domain. Interestingly, all dimer-of-trimer contacts are within an N-terminal extension of about 80 amino acids found only in a subset of SATs. Comparison with previously determined structures (Beaman, Blanchard *et al.*, 1998; Beaman, Sugantino *et al.*, 1998; Wang *et al.*, 2002) suggests that the active site is formed by a cleft between trimer-related protomers, thus forming a hexameric enzyme complex of approximate dimensions 104 Å along the threefold axis and 55 Å along the twofold axis, with six independent active sites.

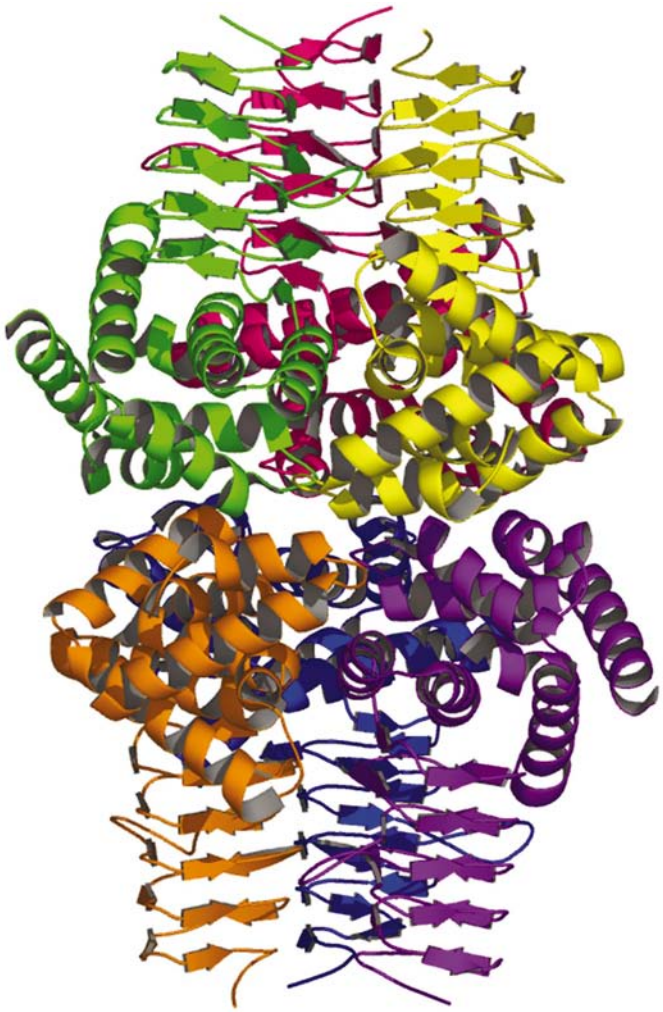

Figure 1
Ribbon diagram of the *H. influenzae* SAT hexameric enzyme complex, with each subunit colored differently. In this orientation, the non-crystallographic threefold axis is vertical and the non-crystallographic twofold axis is horizontal, projecting out of the page.

Table 1
Data-collection, solution and refinement statistics.

Data for the highest resolution shell are given in parentheses.

Collection date	11/02/2003
Collection site, beamline	APS, ID-31/SGX-CAT
Detector type and model	CCD, MAR 165
Temperature (K)	100
Exposure time (s)	3
No. images	197
Oscillation angle (°)	1
Wavelength (Å)	0.9793
Data-processing software	*DENZO*/*SCALEPACK*
SAD data and structure-solution statistics	
No. unique reflections	51228
Redundancy	8.2
$I/\sigma(I)$	18.0
Resolution range (Å)	20–2.7 (2.8–2.7)
Completeness (%)	100 (100)
$R_{\rm merge}$ (%)	0.102 (0.331)
Phasing power after solvent flattening (acentric, centric)	0.70, 0.67
Figure of merit after solvent flattening	0.70
No. molecules/asymmetric unit	6
Solvent content (%)	51.4
Structure-solution, solvent-flattening software	*SOLVE*, *RESOLVE*
Structure refinement	
Resolution range (Å)	20–2.7 (2.78–2.7)
R	0.174 (0.234)
$R_{\rm free}$	0.219 (0.290)
Completeness (%)	99.75
No. reflections in test set for $R_{\rm free}$	2601 (194)
No. reflections used in refinement (working set)	48592 (3372)
Overall average B factor (Å^2)	31.0
R.m.s. deviations, bond lengths (Å)	0.013
R.m.s. deviations, bond angles (°)	1.35
No. protein atoms	11094
No. solvent atoms	418
No. other atoms	0
Refinement software	*REFMAC* 5.1.9999
Ramachandran plot analysis	
Most favored regions (%)	89.6
Additionally allowed regions (%)	9.6
Generously allowed regions (%)	0.8
Disallowed regions (%)	0.0

2. Materials and methods

The coding sequence from the *H. influenzae cysE* gene was cloned into a modified pET26b vector that encodes a C-terminal six-His tag that was not removed by proteolysis. Density for the His tag is not observed. In addition, owing to cloning artifacts, the N-terminus begins with the sequence MSLT rather than the native MT (with the threonine position

being correct in both cases). Selenomethionine (SeMet) substituted protein was expressed in *E. coli* BL21 (DE3) cells and purified to homogeneity by affinity chromatography on nickel-chelating Sepharose and subsequent gel filtration on a preparative Superdex-200 column. The protein was concentrated to 7.5 mg ml^{-1} in 10 m*M* HEPES pH 7.5, 150 m*M* NaCl, 10 m*M* methionine and 10% glycerol. Crystals were obtained by vapor diffusion in 1.2 µl hanging drops containing 0.6 µl protein solution and 0.6 µl well solution. The well solution was 2 *M* NaCl, 0.16 *M* $MgCl_2$, 3% PEG 400 and 0.1 *M* sodium citrate pH 5.6 at 293 K. The crystals belong to space group $C222_1$, with unit-cell parameters $a = 130.0$, $b = 209.7$, $c = 136.6$ Å and six molecules per asymmetric unit. The crystals were flash-frozen at 100 K with the addition of cryoprotectant consisting of well solution supplemented with 30% glycerol. Data were collected at the ID-31 beamline (SGX-CAT) at the Advanced Photon Source.

The data were processed and merged with *DENZO* (Otwinowski & Minor, 1997). Se positions were located using *SOLVE* (Terwilliger & Berendzen, 1999) and the first model was built with the program *RESOLVE* (Terwilliger, 2000). Electron density was manually traced using *O* (Jones *et al.*, 1991). Refinement was performed with *REFMAC* 5.1.9999 (Murshudov *et al.*, 1997) using the *CCP4i* program suite (Potterton *et al.*, 2003; Collaborative Computational Project, Number 4, 1994). Tight NCS restraints were used in the initial rounds of refinement and gradually loosened in subsequent rounds. None were used in the refinement of the final model. Water molecules were added using *ARP*/*wARP* 6.0 (Perrakis *et al.*, 1999). Molecular-structure figures were produced using the program *PyMol* (DeLano, 2001) and the sequence-alignment figure was produced using *JALVIEW* (Clamp *et al.*, 2004).

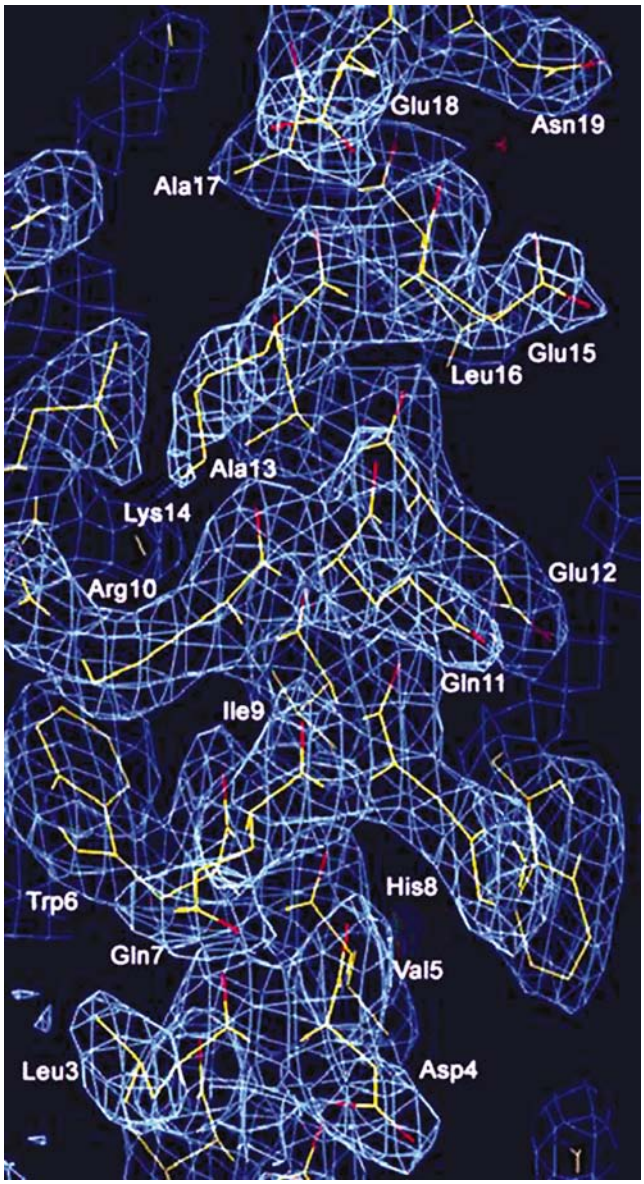

Figure 2
$2F_o - F_c$ electron-density map, contoured at 1.0σ, showing a segment of helix α1. The initial model was built from selenomethionine SAD-phased maps of excellent quality, yielding an unambiguous chain trace. The refined model is shown in this figure.

3. Results and discussion

We determined the structure of the free form of SAT from *H. influenzae* using phases derived from a selenium SAD experiment performed on SeMet-substituted protein. The final model was refined to 2.7 Å with an *R* value of 0.174 and

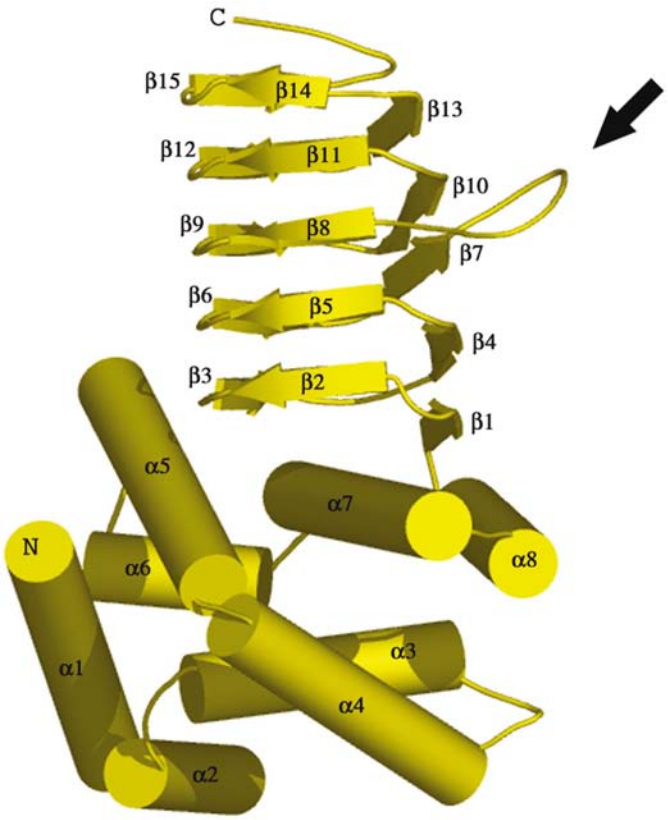

Figure 3
Ribbon diagram of the SAT protomer. Helices of the SATase_N domain are shown as cylinders and numbered α1–α8. β-Strands, all within the left-handed β-helix domain, are numbered β1–β15. The positions of the NH_2- and COOH-termini are indicated. The arrow indicates the eight-residue extension between β7 and β8 discussed in the text.

an R_{free} of 0.219 (Table 1). The model consists of 1446 residues (241 residues in each of the six chains) and 418 water molecules. The overall structure of the protein was found to be a hexamer comprised of a dimer of trimers (Fig. 1). The maximum r.m.s. coordinate difference between individual protomers was found to be 0.304 Å, with an average of 0.27 Å.

3.1. Structure of the protomer

The crystallized protein includes the native residues 2–266, three additional residues at the N-terminus (MSL) and ten additional residues at the C-terminus including the His tag (EGGSHHHHHH). Electron density is not observed for the N-terminal amino acid or for C-terminal residues after 240. Each protomer consists of two domains: a SATase_N domain and a left-handed parallel β-helix domain that includes the series of hexapeptide repeats. The SATase_N domain consists of eight helices that wind over and cap the β-helix. The helices are numbered α1 (residues 1–20), α2 (23–31), α3 (36–50), α4 (55–69), α5 (71–88), α6 (94–101), α7 (102–119) and α8 (121–136). Electron density for a segment of α1 is shown in Fig. 2.

Following the SATase_N domain is a five-coil β-helix, spanning residues 137–240 (Fig. 3). In the third coil, residues 180–188 extend out of the β-helix in a loop that has poorly defined density and the model in this region exhibits high B factors. In the related structures 2xat and 1kqa, analogous extensions of greater length (18 and 36 residues, respectively) form structures that cover the active site (Beaman, Sugantino *et al.*, 1998; Wang *et al.*, 2002). However, in *H. influenzae* SAT this extension comprises only eight residues. In another related structure, 3tdt (Beaman, Blanchard *et al.*, 1998), the corresponding loop is nine residues in length, similar to the structure reported here. In 3tdt, for which structures have been determined in both the free and a CoA/substrate-analog-bound form, this short loop extension is found to change conformation substantially upon ligand binding (Beaman, Blanchard *et al.*, 1998). Notably, a glutamic acid side chain of the 3tdt extension (Glu169, the third residue of the extension) forms hydrogen bonds to substrate atoms. In the *H. influenzae* SAT extension, Glu184, the fourth residue of this extension, may serve a similar function.

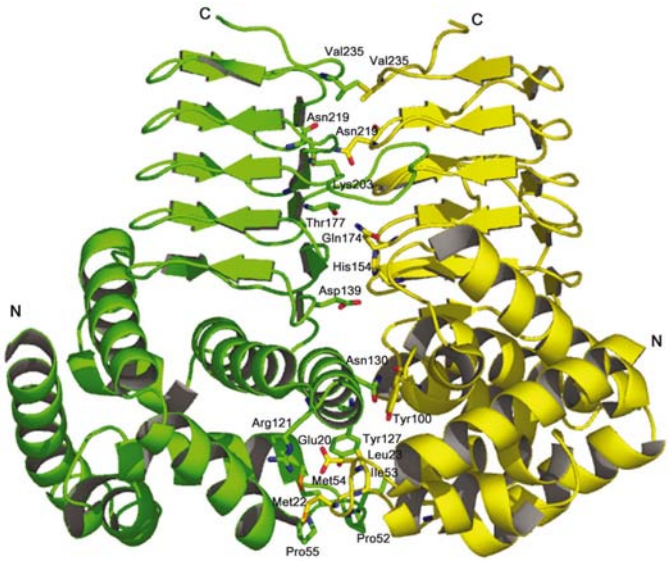

Figure 4
Details of the threefold-symmetric interface. For clarity, only two protomers are shown, colored as in Fig. 1. A salt bridge is formed between the side chains of Glu20 of one protomer and Arg121 of its mate. Hydrophobic interactions are formed between Met22, Leu23 and Tyr100 of from one subunit with Pro52, Ile53, Met54, Pro55 and Tyr127 of the other. In the β-helix, the side chain from Asn130 of one subunit forms a hydrogen bond with the main-chain O atom of Tyr100 from the other. An interchain salt bridge is found between the side chains of Asp139 and His154. Interchain hydrogen bonds are observed between the side chains of Gln174 and Thr177 and the side chains of Lys203 and Asn219. Additionally, the side-chain N atom from Asn219 forms a hydrogen bond with the main-chain carboxyl of Asn219 of the adjacent chain. A symmetrical hydrophobic interaction of the Val235 side chains, near the C-terminal end of each ordered chain, caps the end of the β-helix.

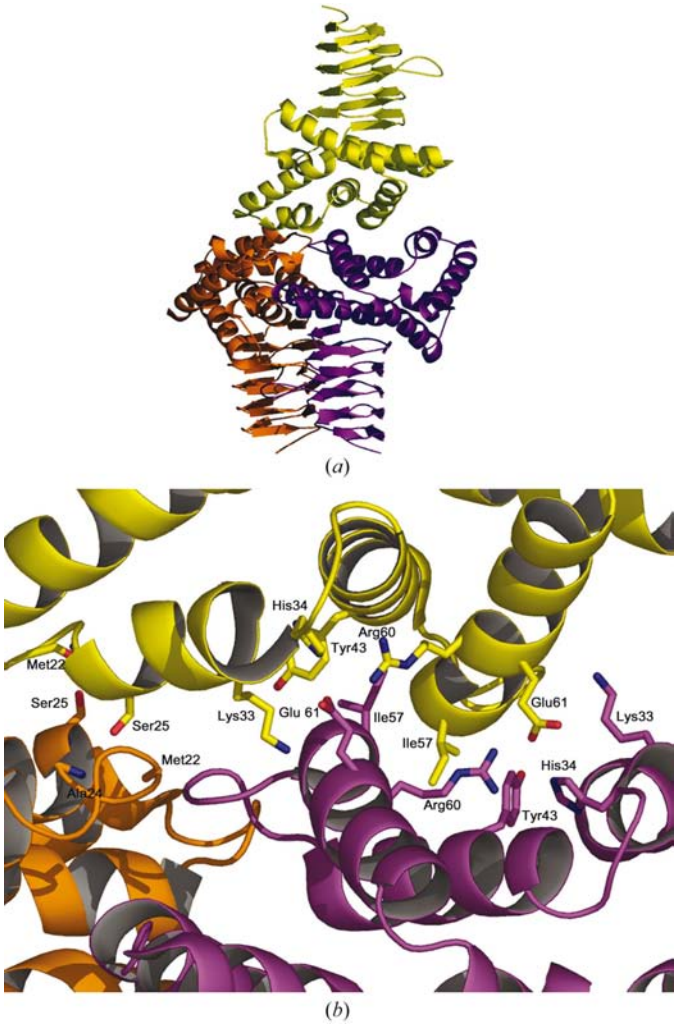

Figure 5
Twofold dimer-of-trimer interactions. (*a*) Three subunits are displayed according to the same coloring scheme as Fig. 1. The two trimers associate in a 'staggered' manner such that a single subunit from each trimer assembly contacts two subunits from the adjacent trimer. (*b*) A detailed view of this interface reveals hydrogen-bonding interactions between the side chain of Ser25 and the main-chain atoms of Pro21 (not shown), Met22, Ala24 and Ser25 of the adjacent trimer. Interchain hydrophobic interactions include the side chains of Ile57 and Tyr43. In addition, a pocket of complementary charges is formed by the association of side chains from Arg60, His34 and Lys33 from one chain with Glu61 from the other.

3.2. Threefold interface

Previously determined structures of related enzymes containing the β-helix fold often reveal trimeric assemblies (Beaman, Blanchard *et al.*, 1998; Beaman, Sugantino *et al.*, 1998; Wang *et al.*, 2002). The *H. influenzae* SAT reveals a similar trimer structure formed around a non-crystallographic threefold axis. This threefold axis combines with a perpendicular non-crystallographic twofold axis to form a dimer of trimers. The threefold interface is assembled through interactions among the β-helices as well as interactions between SATase_N domains (Fig. 4). Interactions between threefold-related SATase_N domains include a large region of primarily hydrophobic interactions that includes Met22, Leu23, Phe26 and Tyr100 of one chain, and Pro52, Ile53, Met54, Pro55 and Tyr127 of the related chain. In contrast, interactions among β-helices are mainly comprised of polar interactions. Furthermore, each β-helix turn donates only a single side chain in the formation of the threefold interface (Fig. 4).

3.3. Twofold interface

An approximately twofold symmetric non-crystallographic interface positions the trimers in a 'head-to-head' orientation to form the hexamer. This twofold interface is mediated by interaction of the SATase_N domains. The trimers, which are related by a non-crystallographic twofold axis perpendicular to the threefold, are arranged in a 'staggered' fashion such that

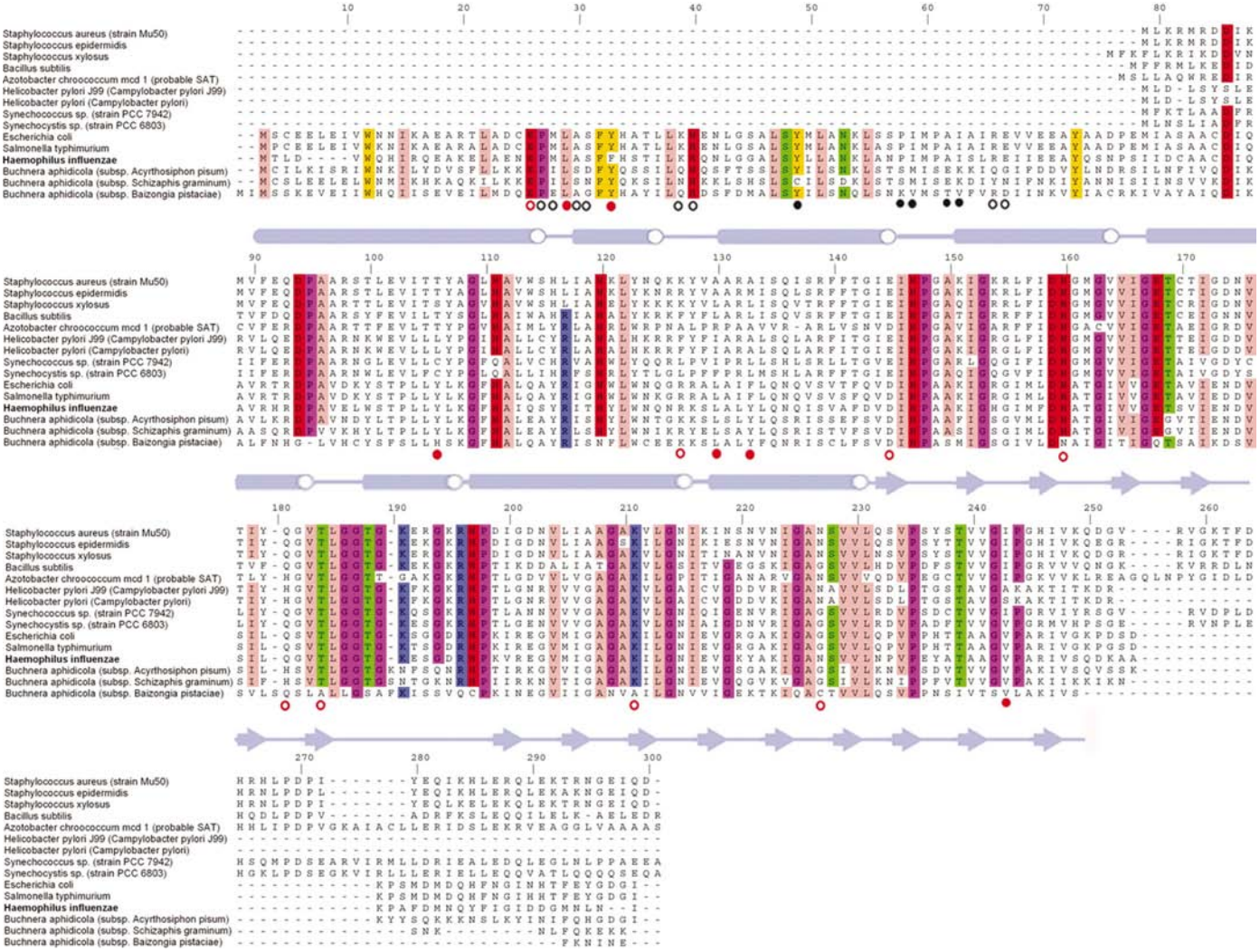

Figure 6
ClustalW (Thompson *et al.*, 1994) alignment of 15 serine acetyltransferases from the EC 2.3.1.30 entries in the Swiss-Prot database (Boeckmann *et al.*, 2003). Residues with greater than 80% sequence identity are colored according to the Zappo coloring scheme of the *JALVIEW* program (Clamp *et al.*, 2004). Secondary-structure elements based on *DSSP* analysis (Kabsch & Sander, 1983) of the *H. influenzae* SAT structure reported here are shown below the alignment as arrows for β-strands and cylinders for α-helices. Residues involved in formation of the trimer interface are marked in red, with open circles representing polar interactions and closed circles representing hydrophobic interactions. Residues involved in the dimer-of-trimers interface are shown in black using the same convention. Note that all dimer-of-trimer interactions are found within an N-terminal extension that is common to a subfamily of serine acetyltransferases. Swiss-Prot accession numbers, in the order of appearance in the alignment, are as follows: CYSE_STAAM, Q99W74; CYSE_STAEP, Q8CTU2; CYSE_STAXY, P77985; CYSE_BACSU, Q06750; NIFP_AZOCH, P23145; CYSE_HELPJ, Q9ZK14; CYSE_HELPY, P71405; CYSE_SYNP7, Q56002; CYSE_SYNY3, P74089; CYSE_ECOLI, P05796; CYSE_SALTY, P29847; CYSE_HAEIN, P43886; CYSE_BUCAI, P57162; CYSE_BUCAP, P32003; CYSE_BUCBP, Q89B11.

one protomer of each trimer interacts with two protomers of the related trimer (Fig. 5*a*). This interface includes ionic, hydrophobic and hydrogen-bonding interactions. The solvent-accessible surface area buried upon dimerization of the trimers is approximately 13.5% per trimer. An ionic interaction is observed between Glu61 from one protomer and a highly electropositive pocket on its protomer mate formed by side chains His34, Lys33 and Arg60 (Fig. 5*b*). There is a hydrophobic pocket formed closer to the threefold axis that includes interactions between the side chains of Tyr43 and Ile57. The hexameric assembly features a cavity near its center with approximate dimensions of 40 Å from top to bottom (parallel to the threefold axis) and 18 Å across (perpendicular to the threefold axis).

All of the residues involved in the twofold dimer-of-trimers interface are found within an N-terminal extension of approximately 80 amino acids that is found only in a subset of serine acetyltransferases, but not other acetyltransferases (Fig. 6). A *BLASTP* (Altschul *et al.*, 1997) search of the NCBI protein database using the entire sequence of *H. influenzae* SAT yields 782 sequences with expectation values of <0.01 primarily annotated as either serine acetyltransferases or 'unknown'. A similar search using only the N-terminal extension yields 105 sequences also annotated as either serine acetyltransferases or 'unknown'. Thus, if all of the proteins of 'unknown' function are indeed serine acetyltransferases, they are likely to fall into two categories: those that include this N-terminal extension and may form hexamers (13% of likely SATs) and those that do not include this N-terminal extension and are likely to be trimeric (87% of likely SATs).

3.4. Putative active site

Comparisons of the *H. influenzae* SAT structure with related enzymes suggest by analogy that the active sites in *H. influenzae* SAT are likely to be formed by a crevice between adjacent threefold-related protomers. The long axis of this crevice, which is seen in the unliganded structure presented here, runs approximately parallel to the threefold axis. It has been shown for the structurally similar proteins 3tdt, 2xat and 1kqa that CoA binds to this site through interactions with residues from two adjacent threefold-related subunits (Beaman, Blanchard *et al.*, 1998; Beaman, Sugantino *et al.*, 1998; Wang *et al.*, 2002). For each of these structures, the binding of CoA is reported to depend primarily on interactions with β-helix regions, whereas substrate binding is accomplished by interactions with residues from the β-helix regions and residues in the flexible loop insertions emanating from these β-helix regions.

Here, we have shown that *H. influenzae* SAT adopts a protomer structure similar to that of related acetyltransferases, which includes an N-terminal SATase_N domain closely associated with a left-handed β-helix domain formed by a series of hexapeptide repeats. Like the related proteins, a trimeric assembly is observed, but an additional perpendicular non-crystallographic twofold interface gives rise to a hexameric enzyme complex. Sequence database searches suggest that this additional twofold interface may be common to a subset of SATs.

Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. W-31-109-Eng-38. Use of the SGX Collaborative Access Team (SGX-CAT) beamline facilities at Sector 31 of the Advanced Photon Source was provided by Structural GenomiX Inc., who constructed and operate the facility. This is a contribution of the New York Structural Genomics Consortium, NIH Structural Genomics Pilot Center grant 1P50 GM62529.

References

Altschul, S., Madden, T., Schäffer, A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M. & Sonnhammer, E. L. L. (2002). *Nucleic Acids Res.* **30**, 276–280.
Beaman, T. W., Blanchard, J. S. & Roderick, S. L. (1998). *Biochemistry*, **37**, 10363–10369.
Beaman, T. W., Sugantino, M. & Roderick, S. L. (1998). *Biochemistry*, **37**, 6689–6696.
Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S. & Schneider, M. (2003). *Nucleic Acids Res.* **31**, 365–370.
Bogdanova, N. & Hell, R. (1997). *Plant J.* **11**, 251–262.
Clamp, M., Cuff, J., Searle, S. M. & Barton, G. J. (2004). *Bioinformatics*, **20**, 426–427.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
DeLano, W. L. (2001). *The PyMOL User's Manual.* San Carlos, CA, USA: DeLano Scientific.
Hindson, V. J., Moody, P. C., Rowe, A. J. & Shaw, W. V. (2000). *J. Biol. Chem.* **275**, 461–466.
Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–118.
Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
Kredich, N. M., Becker, M. A. & Tomkins, G. M. (1969). *J. Biol. Chem.* **244**, 2428–2433.
Kredich, N. M. & Tomkins, G. M. (1966). *J. Biol. Chem.* **241**, 4955–4965.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. M. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). *Acta Cryst.* D**59**, 1131–1137.
Raetz C. R. & Roderick S. L. (1995). *Science*, **270**, 997–1000.
Smith, I. K. & Thompson, J. F. (1971). *Biochim. Biophys. Acta*, **227**, 288–295.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Wang, X. G., Olsen, L. R. & Roderick, S. L. (2002). *Structure*, **10**, 581–588.
Wirtz, M., Berkowitz, O., Droux, M. & Hell, R. (2001). *Eur. J. Biochem.* **268**, 686–693.