Published in final edited form as:
Proteins. 2007 November 1; 69(2): 422–427. doi:10.1002/prot.21417.

Crystal structure of an acetyltransferase protein from *Vibrio cholerae* strain N16961

M.E. Cuff[1,2], **H. Li**[1], **S. Moy**[1], **J. Watson**[3], **A. Cipriani**[1], and **A. Joachimiak**[1,2,*]

[1]Midwest Center for Structural Genomics, Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439

[2]Structural Biology Center, Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439

[3]Midwest Center for Structural Genomics, Biomolecular Structure and Modeling Group, European Bioinformatics Institute, Birkbeck College, Cambridge, CB10 1SD, United Kingdom

INTRODUCTION

The genus *Vibrio* consists of gram-negative, motile, rod-like bacterium capable of both respiratory and fermentative metabolism.[1] Commonly found in the surface waters of the world, Vibrios are found in both marine and freshwater sources. *V. cholerae* is an important human pathogen transmitted by water or food that affects the small intestine through enterotoxin secretion. Although cholera has been virtually eliminated in the U.S. due to modern water and sewage treatment systems, it still remains a serious health threat in developing countries with poor sanitation and limited health care. The complete genome sequence of *V. cholerae* pathogenic strain N16961 has been determined to consist of two circular chromosomes with 3890 open reading frames (ORFs).[2] Genes that perform essential cell functions such as transcription, translation, cell-wall biosynthesis, DNA replication, and pathogenicity are located on the larger chromosome. The genomic sequence of *V. cholerae* provides an important starting point in understanding how this widespread environmental microorganism remains a persistent human pathogen.[2] Applying a structural genomics survey to this genome can provide new potential targets for drug and therapeutics development. The posttranslational N^{α}-terminal acetylation of proteins is very common in eukaryotes, but so far has been only reported for few bacterial proteins.[3] In this article, we report the crystal structure of VC1889 protein, a putative acetyltransferase from *V. cholerae*, that was determined at 1.7 Å resolution and discuss its potential function.

MATERIALS AND METHODS

Protein cloning, expression and purification

The ORF of this protein from *V. cholerae* was amplified from genomic DNA with *KOD* DNA polymerase using conditions and reagents provided by the vendor (Novagen, Madison, WI). The gene was cloned into the pMCSG7 vector[4] using a modified ligation independent cloning

*Correspondence to: Dr. Andrzej Joachimiak, Biosciences Division, Midwest Center for Structural Genomics and Structural Biology Center, Argonne National Laboratory, 9700 S. Cass Ave. Argonne, IL 60439. andrzejj@anl.gov.

protocol.5 This process generated an expression clone producing a fusion protein with an N-terminal His_6 tag and a TEV protease recognition site (ENLYFQ↓S). The fusion protein was over-produced in an *E. coli* BL21-derivative harboring a plasmid encoding three rare *E. coli* tRNAs: Arg [AGG/AGA] and Ile [ATA]. A selenomethionine (SeMet) derivative of the expressed protein was prepared as described by Walsh et al.[6] The protein was purified from a resuspension of IPTG-induced bacterial cells in binding buffer (500 m*M* NaCl, 5% glycerol, 50 m*M* HEPES, pH 8.0, 10 m*M* imidazole, and 10 m*M* β-mercaptoethanol). The cells were lysed by the addition of lysozyme at 1 mg/mL in the presence of a protease inhibitor cocktail (Sigma P8849) (0.25 mL/5 g cells) and sonicated on ice for 3 min. After clarification by centrifugation (30 min at 30,000g) and passage through a 0.2 μm filter, the lysate was applied to Ni-HiTrap Sepharose HP resin (Amersham) and unbound proteins were removed by washing with 10 volumes of binding buffer. The fusion protein was eluted from the column with 250 m*M* imidazole in buffer (50 m*M* HEPES, 500 m*M* NaCl, 5% glycerol, 10 m*M* β-mercaptoethanol). The fusion tag was then cleaved with recombinant His_6-tagged TEV protease.[7] The cleaved protein was purified from the His_6-tag, undigested protein and His_6-tagged TEV protease by application of the solution to a Ni-NTA column (Qiagen). The purified protein was dialyzed in 20 m*M* HEPES pH 8.0, 250 m*M* NaCl, 2 m*M* DTT and concentrated using Centricon Plus-20 (Millipore, Bedford, MA).

PROTEIN CRYSTALLIZATION

The protein was crystallized by vapor diffusion in sitting drops by mixing 1 μL of the protein solution (17 mg/mL) with 1 μL of reservoir solution (0.1*M* Bis-Tris propane pH 7.0, 4*M* NaNO3) and equilibrated at 289 K over 500 μL of this solution. Crystals appeared after 24 h. Prior to data collection, crystals were flashfrozen in liquid nitrogen with reservoir solution supplemented with 28% sucrose as a cryoprotectant. The crystals belong to the orthorhombic space group $P2_12_12$ with unit cell dimensions of a = 53.05 Å, b = 103.07 Å, c = 37.86 Å, and diffracted beyond 1.7 Å Bragg spacing.

DATA COLLECTION

Two wavelength anomalous diffraction (MAD) data were collected at 100 K at the 19ID beamline of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory. The data were recorded on an ADSC Quantum 315 detector. Peak and inflection point energies were determined from the X-ray absorption spectrum of the SeMet-labeled crystal. All data were collected from a single crystal using inverse beam geometry at the peak and inflection point energies. The crystal was exposed for 3 s per 1.0° rotation of Ω with a crystal to detector distance of 220 mm. Data collection strategy, integration, and scaling were performed with the HKL3000 package.[8] A summary of crystallographic information can be found in Table I.

STRUCTURE DETERMINATION AND REFINEMENT

All three expected selenium sites were located, and initial phasing and density modification were calculated with an early version of HKL3000. This largely automated structure solution program package includes SHELXC, SHELXD, SHELXE, MLPHARE, DM, and SOLVE/RESOLVE.[9–12] Final phasing and model building were accomplished with the autoSHARP suite[13] including ARP/wARP.[14] Model completion and adjustments to the structure were accomplished with COOT.[15] A total of 174 out of 178 residues were modeled with residues 42–45 missing due to a lack of electron density. The model was subjected to iterations of refinement, manual correction, and solvent addition, achieving a final R factor of 21.1% and R free of 24.4%. TLS and restrained refinement in Refmac516 was performed against peak data to 1.7 Å using Hendrickson-Lattman coefficients from autoSHARP prior to density modification. For refinement the peak data were used with Bijvoet pairs averaged. The

stereochemistry of the structure was analyzed with PROCHECK17 and MolProbity.[18] All residues were in allowed regions of the MolProbity Ramachandran plot, with 97.62% of these in favored regions. Atomic coordinates and experimental structure factors have been deposited in the Protein Data Bank (PDB) and are accessible under the code 2FCK. Refinement details are provided in Table I.

DISCUSSION

VC1889 forms a compact globular α/β protein with a twisted S-shaped, 8-stranded β-sheet at its core and five α-helices and three 3_{10} helices arranged in three layers [Fig. 1(a)]. The sheet can be separated into two antiparallel sections: β1–β5 and β6–β8. The sections are joined by hydrogen bonding between parallel strands β5 and β6, which veer apart at their C-terminal ends. Helices η1, α1, η2 and α2 lay above the sheet while helices α3, α4 and α5 are below. Short 3_{10} helix η3 is in the loop between strands β4 and β5. A disordered loop occurs between η2 and α2. Also found in the electron density maps were 12 nitrate ions scattered through the structure and a glycerol molecule. A cavity exists between the β sheet and helices η1, α1, η2 and α2 with several conserved residues lining its site [Fig. 2(a)]. Interestingly, only one side of the cavity is lined with conserved residues while the other side is sequence diverged including potential catalytic Cys139 residue that is not conserved. This may suggest that these proteins use a common cofactor, with a divergent set of substrates.

The structure was submitted to ProFunc,[19] a server that performs a number of sequence and structure based analyses, in order to identify putative functions. Sequence comparison provides strong supporting evidence for acetyltransferase activity. A BLAST[20] search for this protein sequence resulted in no identical sequences. However, the sequence shares over 50% identity with a protein from *V. vulnificus* assigned as ribosomal-protein-serine acetyltransferase (Uniprot entry Q8DAW1). A search against the superfamily library of Hidden Markov Models21,22 derived from the SCOP (Structural Classification of Proteins) database23 found one motif matching the sequence. The match occurred between residues 1–4, 6–74, and 77–178 and identifies the characteristic motif of the Acyl-CoA *N*-acyltransferases (Nat) superfamily (no. 55729). An InterProScan search,[24] which attempts to find sequence motifs in a database of protein families, domains, and functional sites, produced five matching sequence motifs to the protein. The matching motifs were: PD451850 and PD338839 (Q9KQV9_VIBCH_Q9KQV9), PF00583 (Acetyltransf_1), SSF55729 (Acyl-CoA N-acyltransferases Nat), and motif G3D.3.40.630.30 (CATH classification: transferase).

Structural analysis using SSM[25] revealed an identical fold match to a number of ribosomal N^{α}-protein acetyl-transferases with good Z-scores (between 6 and 8), although the sequence similarity between structures is less than 25%. A multiple structure superposition identified a putative coenzyme A binding site [Fig. 1(b)]. We have compared the structure of VC1889 with the structure of RimL N^{α}-acetyltransferase (PDB ID = 1s7n)[26] [Fig. 2(b)]. There are some differences in the VC1889 structure including helix α3. This region is a loop in RimL and is involved in cofactor A binding [Fig. 2(c)]. There is also significant shift of the hairpin (β7 and β8) and the loop region (corresponding to residues 160–170) that appears to be close to the active site where the acetyl group attaches to coenzyme A, therefore its exact conformation has bearing on the substrate specificity. It was reported that upon binding of coenzyme A acetyltransferases undergoes some conformational changes, therefore these differences may correspond to differences between apo and liganded forms of the protein. On the other hand, they may also represent an adaptation to a different acetylation target.

The RimL protein is a dimer. VC1889 shares a similar dimer interface with a symmetry mate within the orthorhombic crystal. Whereas the different monomers superpose with an approximate RMSD of 2.1 Å, a 30–40° twist between members of the dimer brings the β4–

β5 hairpin into the proposed RimL active site trough[26] and raises the overall RMSD to 3.4 Å. An analysis of the VC1889 protein dimer interface with the PQS server,[28] however, suggests that this protein is monomeric.

To identify possible ligand specificities, the templatebased approaches of ProFunc were investigated. A number of ligand binding template matches to proteins binding coenzyme A and acetyl coenzyme A were found, but the superposition of the structures is poor. Examination of the reverse template results from ProFunc[12] reveals that almost all hits are to acetyltransferase proteins. Of particular interest are the top two matches to the apo form of RimL-ribosomal L7/L12 N^{α}-protein acetyltransferase (PDB ID = 1s7f and 1s7k). Although RimL shares only 21.3% sequence identity (38% similarity) with VC1889, they have a high structural similarity. The residues used to create a template from the query structure are identified as Tyr100, Trp101, and Ala147. These residues bind the coenzyme A ligand in the N^{α}-protein acetyltransferases close to the point where the acetyl group is connected [Fig. 2(c)]. Further investigation of the local environment around this template match shows 18 identical and 15 similar residues within 10 Å of the template match giving a local sequence identity of 40%. This is a prime illustration showing that although the global sequence has significantly diverged, the overall fold is retained and the local similarity at the functionally important site remains high. The actual substrate for the acetyltransferase is not known, however, the most likely function is that of a ribosomal N^{α}-protein acetyltransferase.

Acknowledgments

We wish to thank all members of the Structural Biology Center at Argonne National Laboratory for their help in conducting the experiments.

Grant sponsor: National Institutes of Health; Grant numbers: GM62414 and GM074942; Grant sponsor: U.S. Department of Energy, Office of Biological and Environmental Research; Grant number: DE-AC02-06CH11357.

REFERENCES

1. Todar, K. University of Wisconsin: Internet; 2005. *Vibrio cholerae* and asiatic cholera. Todar's online textbook of bacteriology. (cited 2005, Dec 1); Available at: http://www.textbookofbacteriology.net/
2. Heidelburg JF, Elsen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, Gill SR, Nelson KE, Read TD, Tettelin H, Richardson D, Ermolaeva MD, Vamathevan J, Bass S, Qin H, Dragoi I, Sellers P, McDonald L, Utterback T, Fleishmann RD, Nierman WC, White O, Salzberg SL, Smith HO, Colwell RR, Mekalanos JJ, Venter JC, Fraser CM. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature 2000;406:477–483. [PubMed: 10952301]
3. Polevoda B, Sherman F. Composition and function of the eukaryotic N-terminal acetyltransferase subunits. Biochem Biophys Res Commun 2003;308(1):1–11. [PubMed: 12890471]
4. Stols L, Gu M, Dieckman L, Raffen R, Collart FR, Donnelly MI. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. Protein Expr Purif 2002;25:8–15. [PubMed: 12071693]
5. Kim Y, Dementieva I, Zhou M, Wu R, Lezondra L, Quartey P, Joachimiak G, Korolev O, Li H, Joachimiak A. Automation of protein purification for structural genomics. J Struct Funct Genomics 2004;5:111–118. [PubMed: 15263850]
6. Walsh MA, Dementieva I, Evans G, Sanishvili R, Joachimiak A. Taking MAD to the extreme: ultrafast protein structure determination. Acta Crystallogr D Biol Crystallogr 1999;55(6):1168–1173. [PubMed: 10329779]
7. Kapust RB, Waugh DS. Controlled intracellular processing of fusion proteins by TEV protease. Protein Expr Purif 2000;19:312–318. [PubMed: 10873547]
8. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. HKL-3000: the integration of data reduction and structure solution— from diffraction images to an initial model in minutes. Acta Crystallogr D Biol Crystallogr 2006;62:859–866. [PubMed: 16855301]

9. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Red RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54:905–921. [PubMed: 9757107]
10. Terwilliger TC. SOLVE and RESOLVE: automated structure solution and density modification. Methods Enzymol 2003;374:22–37. [PubMed: 14696367]
11. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50:760–763. [PubMed: 15299374]
12. Laskowski RA, Watson JD, Thornton JM. Protein function prediction using local 3D templates. J Mol Biol 2005;351(3):614–626. [PubMed: 16019027]
13. de La Fortelle E, Bricogne G. Maximum-likelihood heavy-atom parameter refinement for the multiple isomorphous replacement and multiwavelength anomalous diffraction methods. Methods Enzymol 1997;276:472–494.
14. Morris RJ, Perrakis A, Lamzin VS. ARP/wARP and automatic interpretation of protein electron density maps. Methods Enzymol 2003;374:229–244. [PubMed: 14696376]
15. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 2004;60:2126–2132. [PubMed: 15572765]
16. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 1997;53:240–255. [PubMed: 15299926]
17. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 1993;26:283–291.
18. Lovell SC, Davis IW, Arendall WB III, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by C-alpha geometry: phi, psi, and C-beta deviation. Proteins: Struct Funct Gen 2003;50:437–450.
19. Laskowski RA, Watson JD, Thornton JM. ProFunc: a server for predicting protein function from 3D structure. Nucleic Acids Res 2005;33((Web Server issue)):W89–W93. [PubMed: 15980588]
20. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25 (17):3389–3402. [PubMed: 9254694]
21. Gough J, Karplus K, Hughey R, Chothia C. Assignment of homology to genome sequences using a library of Hidden Markov Models that represent all proteins of known structure. J Mol Biol 2001;313:903–919. [PubMed: 11697912]
22. Madera M, Vogel C, Kummerfeld SK, Chothia C, Gough J. The SUPERFAMILY database in 2004: additions and improvements. Nucl Acids Res 2004;32:D235–D239. [PubMed: 14681402]
23. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540. [PubMed: 7723011]
24. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. InterProScan: protein domains identifier. Nucleic Acids Res 2005;33:W116–W120. [PubMed: 15980438]
25. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr 2004;60:2256–2268. [PubMed: 15572779]
26. Vetting MW, De Carvalho LP, Roderick SL, Blanchard JS. A novel dimeric structure of the RimL Nalpha-acetyltransferase from *Salmonella typhimurium*. J Biol Chem 2005;280(23):22108–22114. [PubMed: 15817456]
27. DeLano, WL. The PyMOL molecular graphics system (internet). 2002. (cited 2006, Sept 1). Available at: http://www.pymol.org
28. Henrick K, Thornton JM. PQS: a protein quaternary structure file server. Trends Biochem Sci 1998;23 (9):358–361. [PubMed: 9787643]
29. Gouet P, Courcelle E, Stuart DI, Metoz F. Bioinformatics 1999;15:305–308. [PubMed: 10320398]
30. Corpet F. Nucleic Acids Res 1988;16:10881–10890. [PubMed: 2849754]

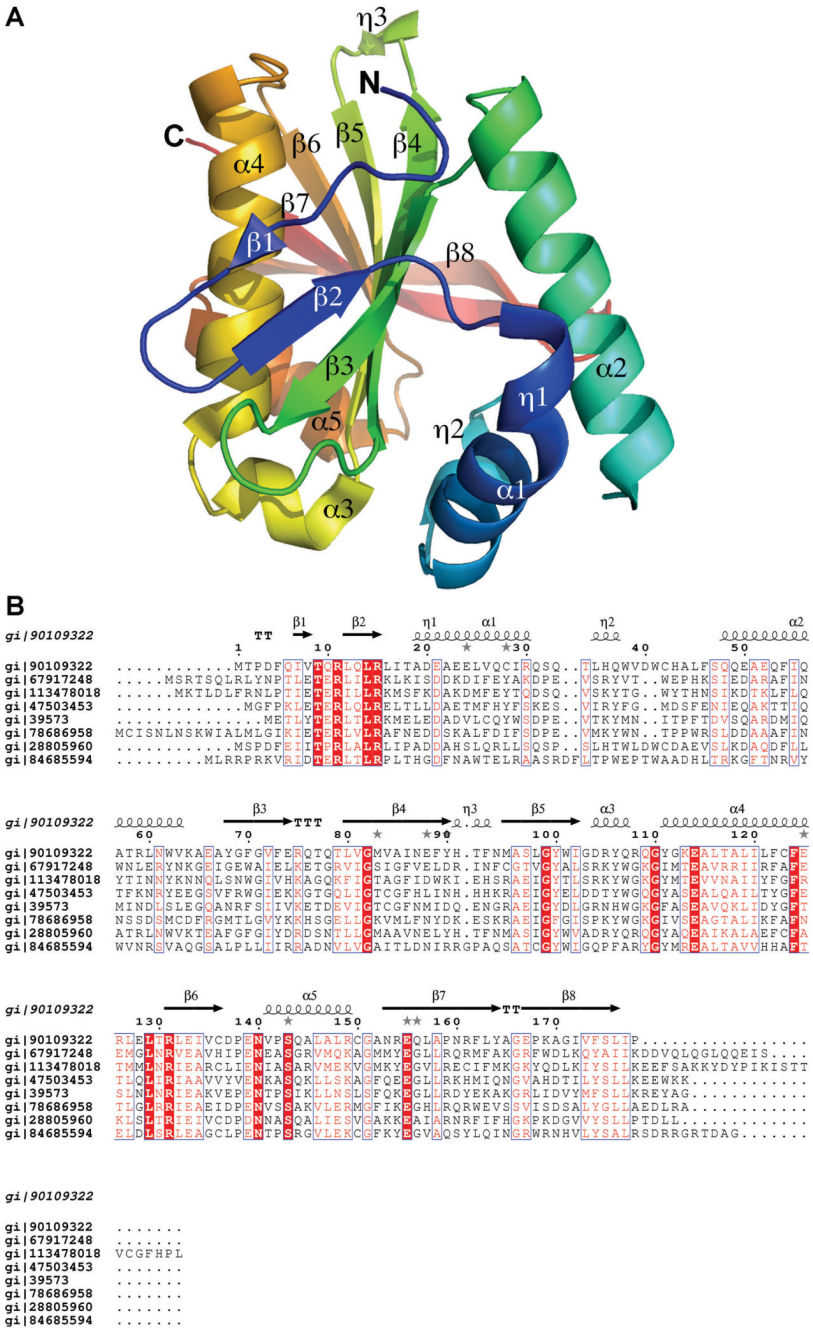

Figure 1.

(**a**) Structure of VC1889. The ribbon schematic representation is colored in a gradient from amino to carboxy-terminus, blue to red. The secondary structure elements are labeled. All ribbon drawings were generated with PyMOL.[27] (**b**) Multiple sequence alignment of VC1889 protein homologs from several bacterial species:**gi|90109322/ PDB ID** = **2FCK** (VC1889) from V. cholerae **gi|47503453** from Bacillus anthracis **gi|39573** from Bacillus licheniformis **gi|67917248** from Clostridium thermocellum **gi|28805960** from Vibrio parahaemolyticus **gi|113478018** from Trichodesmium erythraeum **gi|84685594** from Rhodobacterales bacterium, and **gi|78686958** from Shewanella sp. Invariant residues are shown in red boxes, while positions with conservative replacements are shown in blue boxes. Secondary structural

elements of the VC1889 protein from V. cholerae are shown at the top of the sequence alignment. Sequences were aligned using MultAlin/ESPrint programs.[29,30]

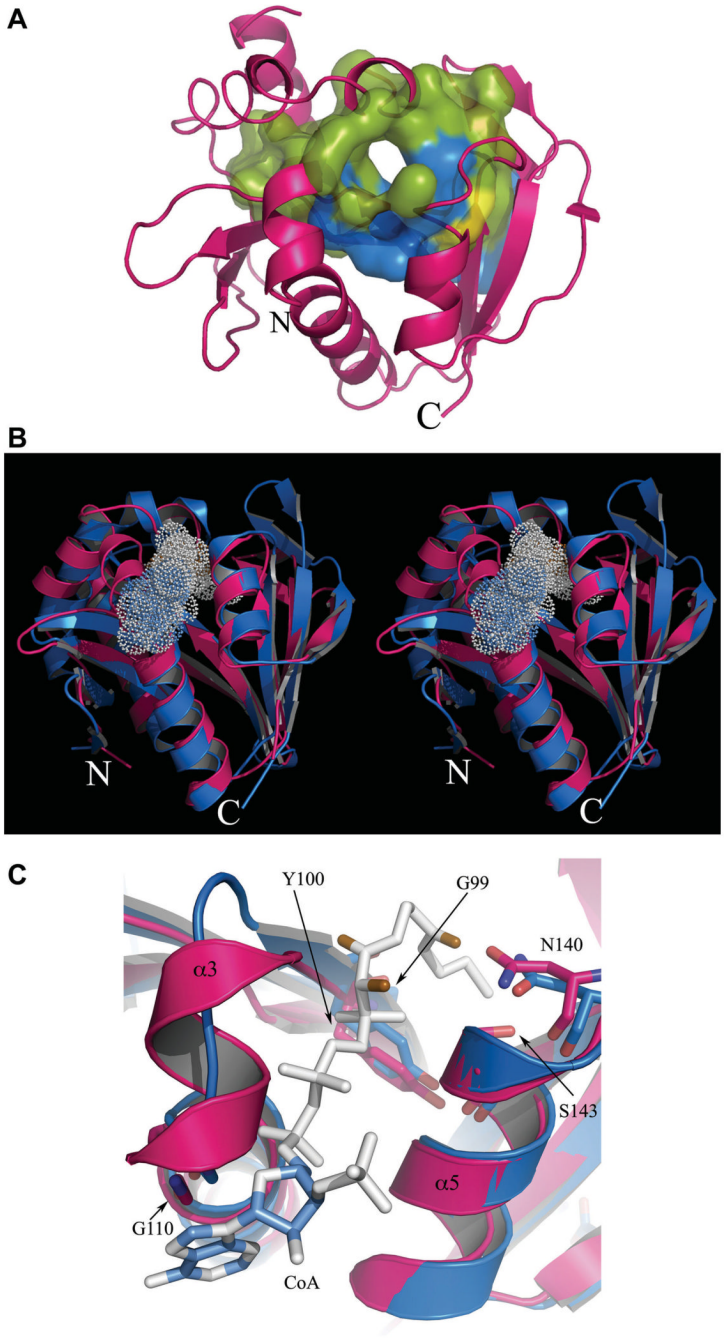

Figure 2.

(**a**) The surface of the VC1889 calculated using GPSS server with CoA site on the right and the active site surface on the left. Surface regions for conserved residues are colored blue and that for the unconserved Cys139 is in yellow. (**b**) A stereoview of the superposition of VC1889 (pink) with RimL N^{α}-acetyltransferase (PDB ID = 1S7N) (blue) with cofactor A (white) in dot representation. Note that in place of helix α3 in the case of VC1889, the RimL protein has a loop, with resides that hydrogen bond to and make room for cofactor A. (**c**) A close-up pf the superposition at the coenzyme A binding site. The residues in stick form are invariant among VC1889 (pink) and close bacterial homologs, and RimL N^{α}-acetyltransferase from S. typhimurium (blue). Residue numbers are for VC1889: Gly 99, Tyr 100 (sometimes Phe), Gly

110, Asn 140, and Ser 143. Gly 110 appears to be instrumental in the bend between a α3 and α4, and small enough for the adenine ring to pack nearby.

Table I

Summary of Crystallographic Information

MAD data[a]		
Unit cell	a = 53.05 Å, b = 103.07 Å, c = 37.86 Å	
Space group	P 2_1 2_1 2	
M_W (Da)	20510 [178][b]	
Mol (AU)	1	
SeMet (AU)	3	
	Peak	*Remote*
Wavelength (Å)	0.97915	0.97900
Resolution range (Å)	50–1.65	50–1.65
I/σ	11.9	12.1
Average redundancy[c]	4.7 (29.7)	4.7 (33.8)
Completeness (%)	94.1 (75.3)	92.2 (64.5)
Refinement statistics		
Resolution range (Å)	30.82–1.70	
No of reflections	22627	
Cutoff	0.0	
R-value (%)	21.0	
Free *R*-value (%)	24.0	
Rms deviations from ideal geometry		
Bond length (Å)	0.017	
Angle (degree)	1.661	
Chiral (degree)	0.119	
No of atoms		
Protein	1494	
Water	238	
Mean B-factor (Å^2)	45.94	
Ramachandran plot statistics (%)		
Residues in most favored regions		93.5%
Residues in additional allowed regions		5.8%
Residues in disallowed region		0.0%

[a] Numbers in parentheses are for the highest resolution bin, 1.71–1.65 Å.

[b] Value in square brackets indicate no of residues.

[c] Bijvoets not merged.