LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-JRNL-209836

The crystal structure of Rv1347c, a putative antibiotic resistance protein from Mycobacterium tuberculosis, reveals a GCN5-related fold and suggests an alternative function in siderophore biosynthesis

G. L. Card, N. A. Peterson, C. A. Smith, B. Rupp, B. M. Schick, E. N. Baker

February 18, 2005

Journal of Biological Chemistry

The crystal structure of Rv1347c, a putative antibiotic resistance protein from *Mycobacterium tuberculosis*, reveals a GCN5-related fold and suggests an alternative function in siderophore biosynthesis

Graeme L. Card[1,4], Neil A. Peterson[1,2], Clyde A. Smith[1,5], Bernhard Rupp[3], Brian M. Schick[3] and Edward N. Baker [1,2*]

[1]School of Biological Sciences and [2]Centre for Molecular Biodiscovery, University Of Auckland, Auckland, New Zealand.

[3]Macromolecular Crystallography and Structural Genomics Group, University of California-LLNL, 7000 East Ave., Livermore, CA 94550-9234, USA

[4]Present address: Plexxikon Inc., 91 Bolivar Drive, Berkeley, CA 94710-2210, USA

[5]Present address: Stanford Synchrotron Radiation Laboratory, Menlo Park, CA 94025, USA

* Corresponding author: Phone: +64-9-373-7599

Fax: +64-9-373-7619

Email: ted.baker@auckland.ac.nz

Running title: Crystal structure of a mycobacterial acyltransferase

Keywords: Acyltransferase, crystal structure, Mycobacterium tuberculosis, siderophore biosynthesis, GNAT family

SUMMARY

Mycobacterium tuberculosis, the cause of TB, is a devastating human pathogen. The emergence of multi-drug resistance in recent years has prompted a search for new drug targets and for a better understanding of mechanisms of resistance. Here we focus on the gene product of an open reading frame from *M. tuberculosis*, Rv1347c, which is annotated as a putative aminoglycoside N-acetyltransferase. The Rv1347c protein does not show this activity, however, and we show from its crystal structure, coupled with functional and bioinformatic data, that its most likely role is in the biosynthesis of mycobactin, the *M. tuberculosis* siderophore. The crystal structure of Rv1347c was determined by MAD phasing from selenomethionine-substituted protein and refined at 2.2 Å resolution ($R = 0.227$, $R_{free} = 0.257$). The protein is monomeric, with a fold that places it in the GCN5-related N-acetyltransferase (GNAT) family of acyltransferases. Features of the structure are an acylCoA binding site that is shared with other GNAT family members, and an adjacent hydrophobic channel leading to the surface that could accommodate long-chain acyl groups. Modeling the postulated substrate, the N^{ε}-hydroxylysine side chain of mycobactin, into the acceptor substrate binding groove identifies two residues at the active site, His130 and Asp168, that have putative roles in substrate binding and catalysis.

INTRODUCTION

Mycobacterium tuberculosis, the causative agent of tuberculosis (TB), is the world's most devastating pathogen, responsible for 2-3 million deaths annually (1). Two features of this very slow-growing organism make it particularly difficult to combat. Firstly, it has a thick, waxy cell wall, rich in unusual lipids (2), that makes it impermeable to many drugs. Secondly, when engulfed by macrophages, it can switch its metabolism and remain in a latent or persistent state inside granulomas in the lung (3,4). This latent state can last for many years (5), until the organism is reactivated, for example when the immune system becomes compromised. Current estimates are that one-third of the world's population is infected, and that the incidence of active TB is rising, in particular as a result of synergy with the HIV/AIDS pandemic. Although effective anti-TB drugs exist, treatment regimes require a cocktail of 2-3 drugs administered for at least 6 months.

The emergence in recent years of strains of *M. tuberculosis* that are resistant to all of the current front-line drugs (6) presents a new threat, paralleling the rise in resistance to antibiotics across the whole spectrum of infectious disease (7). The publication of the complete genome sequence for the H37Rv strain of *M. tuberculosis* (8) presents new opportunities, both for understanding, at a molecular level, the factors that contribute to antibiotic resistance, and for identifying genes whose protein products have potential importance as targets for the design of new anti-TB drugs. At the same time, the problems of functional annotation are considerable; a large proportion of the gene products of the *M. tuberculosis* genome are still of unknown or uncertain function, some anticipated pathways cannot be traced in their entirety, and many presently unknown pathways are likely to exist.

Several open reading frames (ORFs) in the *M. tuberculosis* genome have been annotated as antibiotic resistance genes. These include two putative aminoglycoside 3'-phosphotransferases (APHs) (Rv3225c and Rv3817) and three putative aminoglycoside N-acetyltransferases (AACs) (Rv0133, Rv0262c and Rv1347c). The aminoglycosides, which include streptomycin, the first chemotherapeutic agent to be effective against *M. tuberculosis*, typically have a three-ring structure comprising one highly substituted aminocyclitol ring linked to a modified ribose, which is in turn linked to N-acetylglucosamine. The APH enzymes inactivate aminoglycoside antibiotics by ATP-dependent phosphorylation of a target oxygen atom, and the AAC enzymes by CoA-dependent acetylation of an amino group (9).

A confounding factor in the annotation of these ORFs, however, is that both the APHs and the AACs belong to wider families of enzymes with diverse functions. The APHs are structurally homologous with protein kinases (10) and possess weak protein kinase activity (11). The AACs belong to a large family of N-acetyltransferases that includes enzymes that acetylate histones and other amino-containing substrates, as well as aminoglycosides (12,13). Sequence identity within these families is generally low, making definitive identification difficult, and substrate specificity can be relatively broad. Thus, for example, the gene product of Rv0262c has been shown to be able to carry out the *in vitro* acetylation of aminoglycosides with either 2'-amino or 2'-hydroxyl substituents, but biochemical data and inferences from the crystal structure suggest that the "true" physiological substrate could be a substituted glucosamine derivative such as mycothiol (14,15).

Here we report the three-dimensional structure of the product of the *M. tuberculosis* ORF Rv1347c, determined by X-ray crystallography at 2.2 Å resolution. This gene product has been annotated as a possible aminoglycoside 6'-N-acetyltransferase, although recent *in vitro* biochemical assays have failed to demonstrate this activity (16). Intriguingly Rv1347c has been

found to be essential for the growth of *M. tuberculosis* in a genome-scale transposon mutagenesis analysis (17), suggesting that it could have some other, essential, function. The structure determined here shows that Rv1347c is a member of the GCN5 related N-acetyltransferase (GNAT) family of enzymes (13), which includes the AATs (18). Detailed analysis of the structure, however, combined with bioinformatic analysis and modeling, leads us to suggest an alternative role in siderophore biosynthesis. We propose that Rv1347c functions in the acylation of one or both of the N^{ε}-hydroxylysine arms of mycobactin, the essential iron chelator produced by *M. tuberculosis*, and further identify several key residues at the active site and a hydrophobic channel that can accommodate a long-chain acyl group.

EXPERIMENTAL PROCEDURES

Overexpression and purification

The gene coding for Rv1347c was amplified by PCR from genomic DNA, and cloned into a modified pET42a vector (Novagen), with an rTEV cleavage site incorporated. Rv1347c was expressed in the *E. coli* strain BL21(DE3) as a C-terminal GST-fusion protein. In order to increase the yield of soluble fusion protein, each 1 L culture was grown at 37^{o}C until an OD_{600} of 0.7 was reached, at which point the temperature was reduced to 25^{o}C. The temperature was further reduced to 18^{o}C once an OD_{600} of 1.2 was attained, after which expression was induced with 0.1 mM isopropyl-β-D-thiogalactopyranoside at an OD_{600} of 1.5. Expression was allowed to continue overnight.

Cells were harvested at 6,000x g for 15 min at 4^{o}C and resuspended in pre-cooled lysis buffer; 20 mM HEPES pH 8.0, 300 mM NaCl, 5 mM BME, containing 10 mM benzamidine and 1 mM

PMSF. The fusion protein was extracted from the cells using ultrasonication, and batch-purified on pre-equilibrated glutathione-Sepharose 4B resin (Pharmacia) at 4^{o}C for 1.5 h. Following three washes in cold lysis buffer (minus benzamidine and PMSF), the resin was resuspended in 10 mL of the same buffer and EDTA added to a final concentration of 0.5 mM. The resin was then incubated at 4^{o}C overnight with 0.2 mg of rTEV (Gibco), leaving Rv1347c in the soluble fraction. The resin was removed using a 0.2 μm filter and the Rv1347c separated from the polyHis-tagged rTEV by passing over a HiTrap chelating ion exchange column (Pharmacia) charged with nickel and pre-equilibrated in lysis buffer. Soluble Rv1347c was present in the flow-through fraction.

The flow-through fraction was concentrated using an Amicon stirred cell (membrane cut-off 10 kDa) and loaded on to a Superdex 75 FPLC column (Pharmacia), pre-equilibrated in running buffer; 20 mM HEPES pH 8.0, 300 mM NaCl, 5 mM βME, 0.01% NaN_3. Rv1347c was isolated in an elution volume consistent with a molecular weight of ~24 kDa. Dynamic light scattering (DynaPro, Protein Solutions) showed a monodisperse sample with a ratio Cp/R_H of 18.2% and a molecular weight of 28 kDa, consistent with a monomeric species in solution. The protein was concentrated to ~15 mg mL^{-1} and stored frozen in small aliquots at –80^{o}C.

Selenomethionine-incorporated (SeMet) protein was produced *via* inhibition of the methionine metabolism pathway (19) and purified as above. After concentration to ~15 mg mL^{-1}, TCEP (adjusted to pH 8.0) was added to a final concentration of 2 mM to help prevent oxidation of the SeMet. Incorporation of SeMet was assayed using ion-spray mass spectrometry, which confirmed the incorporation of five SeMet residues.

Crystallization and data collection

Initial crystallization conditions were identified by the Crystallization Facility of the Mycobacterium Tuberculosis Structural Genomics Consortium (http://www.doe-mbi.ucla.edu/TB/) at Lawrence Livermore National Laboratory using CRYSTOOL random screening (20), and were readily reproducible in our laboratory. Crystals were grown by vapor diffusion from hanging drops by mixing equal volumes of the native or SeMet incorporated protein with crystallization buffer; 25-27% MethoxyPEG 5000, 0.1 M Tris-HCl pH 6.5, 0.8 % BOG. The crystals belong to space group $P2_12_12_1$ with cell dimensions a = 75.85 Å, b = 77.39 Å, c = 297.60 Å, with 8 molecules per asymmetric unit giving a Matthews coefficient V_M of 2.4 $Å^3$/Da (51% solvent).

For data collection, the crystals were flash-cooled in a nitrogen stream at 100 K after stepwise addition of PEG 400 to the crystal drops to a final concentration of 5%. A native data set to 2.15 Å resolution was collected in-house, using Cu-Kα radiation from a Rigaku RU300 rotating anode generator equipped with Osmic mirror optics and a Mar345 image plate detector. MAD data to 2.25 Å resolution were collected at three wavelengths using a Quantum 4 ADSC CCD detector on Beamline 9-1 at the Stanford Synchrotron Radiation Laboratory (SSRL). All data were reduced and scaled using DENZO and SCALEPACK (21). Details of data collection and processing statistics are in Table 1.

Structure determination and refinement

The Se sites were found using SOLVE (22) which located 30 of the expected 32 sites, (excluding the N-terminal Met residues, which are disordered in all 8 molecules). The initial phases, which gave a figure of merit of 0.66 for data to 2.5 Å resolution, were improved using solvent flattening and 8-fold non-crystallographic symmetry averaging as in RESOLVE (23). The final figure of

merit was 0.72. Initial tracing of the polypeptide chain was performed using MAID (24), and the side chains were placed manually using the program O (25). Refinement was carried out using CNS (26) incorporating R_{free} validation to monitor the progress of refinement. The final model contained 1590 residues out of the 1680 expected in the asymmetric unit, together with 650 water molecules and four BOG molecules, which were found in equivalent positions in four of the eight molecules (D, E and G) in the asymmetric unit. The final R and R_{free} values are 0.227 and 0.257 respectively, with a root-mean-square (rms) deviation from standard geometry of 0.009 Å for bond lengths and 1.7^{o} for angles. The residues that are modeled in the eight molecules of the asymmetric unit are: A, 10-207; B, 12-207; C, 11-209; D, 10-209; E, 10-207; F, 8-209; G, 11-210; H, 10-206. The main chain torsion angles conform well with standard values, with 89% of non-glycine residues falling in the most favored regions of the Ramachandran plot, as defined in PROCHECK (27), and only 2 residues (0.1% of total) in disallowed regions.

RESULTS

Crystal structure determination

The ORF Rv1347c, which encodes a poplypeptide of 210 amino acid residues, was cloned into *E. coli*, overexpressed, purified and crystallized. The crystal structure was then solved, in its apo form, by multiwavelength anomalous diffraction (MAD) methods (28) using selenomethionine-substituted protein and was refined at 2.2 Å resolution to a final R factor of 0.227 (R_{free} = 0.257). The asymmetric unit of the crystal contains eight independent molecules. To investigate possible oligomerisation we analysed the interfaces between neighbouring molecules using the Protein-Protein Interaction Server (http://www.biochem.ucl.ac.uk/bsm/PP/server), based on principles described by Jones and Thornton (29). This analysis showed that the largest interface buries only 477 $Å^2$ (4.5%) of the total accessible surface of the molecule, typical for intermolecular crystal

contacts, strongly suggesting that the protein is monomeric in solution. This is consistent with gel filtration and dynamic light scattering data (not shown) which also indicate a monmeric species.

Molecular structure

The Rv1347c monomer is folded into a single domain based on a central β-sheet with helices packed against both faces of the sheet (Figure 1). The most striking feature of the structure, which is characteristic of all acyltransferases of the GNAT family (see below), is that the β-sheet is divided into two halves which diverge in the centre to create a cleft that serves as a conserved binding site for the acyl-CoA cofactor (13). The N-terminal four strands, β1-β4, form an antiparallel β-sheet that abuts a C-terminal 3-stranded antiparallel β-sheet comprising strands β5-β7. Strands β4 and β5 run parallel, joined by hydrogen bonding at their N-terminal ends but diverging half-way along. In other GNAT family members, this divergence has been attributed to the presence of a conserved β-bulge in strand β4 that gives an accentuated twist to this strand. Rv1347c does not have this bulge, however, yet β4 from Rv1347c aligns perfectly with the β4 strands of the other family members apart from the deletion of one residue from the middle of the strand. This suggests that the β-bulge may have more to do with the details of substrate binding and catalysis than the polypeptide conformation of the GNAT scaffold. The single residue, His130, that replaces the two β-bulge residues of the other proteins is invariant in all the closest homologs of Rv1347c, and seems likely to have a key active site role (see below).

One face of the central β-sheet has three helices packed against it, α1, α2 and α3, following the nomenclature of Modis and Wieringa (30). These three helices form the connection between strands β1 and β2, and together with the 16-residue β3-β4 loop (residues 110-125), the short β6-β7 loop, and part of a long N-terminal extension, enclose a cavity above the central β-sheet that

we propose to be the acceptor substrate binding site. The other face of the β-sheet has packed against it two helices, the long α4 helix connecting strands β4 and β5, which is a conserved feature of all GCN5 family enzymes, and the shorter α5 helix joining strands β5 and β6. Prior to strand β1, the N-terminal portion of the polypeptide wraps around the periphery of the molecule, largely in extended form except for a short 3_{10}-helix, and contributes a loop, residues 16-20, that helps enclose the proposed binding site for the acceptor substrate.

The overall topology resembles a left-handed glove, with the N-terminal half of the β-sheet and helices α1-α3 representing the palm and fingers of the hand and the C-terminal half of the β-sheet representing the thumb. The cleft formed between the "thumb" and the "palm" is the putative binding site for AcCoA and adjacent to this, in the hollow of the "palm" is the proposed substrate binding site.

The crystal structure contains several additional pieces of continuous density that cannot be accounted for by the polypeptide chain (Figure 2). The most prominent of these is an extended ribbon of density that is found associated with each of the eight molecules in the asymmetric unit. Although not equally well defined in each case, it can be modeled as a molecule of the detergent β-octylglucoside (BOG), which was used in crystallization, and has been fully refined as such in four of the eight molecules. The extended octyl chain inserts into a hydrophobic cleft in the "back" side of the molecule, in contact with residues Gly96, Trp98, Leu106, Ile133, Phe143, Leu147 and Ile151 from strands β2, β3 and β4, and helix α4. The glucosyl ring resides on the surface between the β2-β3 and β4-α4 loops, in contact with Trp98, Thr101, Asp135 and Lys138. A second, but less well-defined, ribbon of density follows another hydrophobic channel leading to the acyl-CoA binding site (discussed later).

Sequence and structural comparisons

Searches of the current sequence databases with BLAST (31) reveals many homologous sequences, reflecting the widespread occurrence of proteins from the GNAT family. The top BLAST hits, which are almost exclusively from bacteria, apart from a few fungal representatives, include many proteins that are annotated as "conserved hypotheticals". Perhaps significantly, however, more than half of the top 30 hits are to various bacterial homologues of IucB, which is a CoA-dependent N^{ε}-hydroxylysine N-acetyltransferase from the biosynthetic pathway for the siderophore aerobactin (32). Sequence identity between Rv1347c and these proteins is ~25% on a pairwise basis, with only nine residues completely conserved (Figure 3). Four of these conserved residues, Asp126, Gly128, His130 and Pro169 in Rv1347c, together with Asp168, which changes only to Glu, map to strands β4 and β5, around the region of the proposed active site.

A search of the Protein Data Bank (PDB) using the secondary structure matching program SSM (ref. 33; see also http://www.ebi.ac.uk/msd-srv/ssm/) aligns Rv1347c with a number of other GNAT family enzymes. The closest structural homologues are two aminoglycoside 6'-N-acetyltransferases, from *Salmonella enteritidis* (AAC6'-Iy) (34) and *Enterococcus faecium* (AAC6'-Ii) (18); the tabtoxin resistance protein (TTR) (35), serotonin N-acetyltransferase (AANAT) (36), yeast glucosamine-phosphate N-acetyltransferase (GNA1) (37), the histone acetyltransferase HPA2 (38), and three putative N-acetyltransferases from *Bacillus subtilis* (unpublished; PDB accession codes 1VHS, 1NSL and 1TIQ). Each of these proteins matches Rv1347c to a similar degree, with typically 130-140 Cα atoms aligning with root-mean-square (rms) differences of 2.6-3.0 Å and sequence identities over the aligned portions of 10-15%. The

most closely matching portion comprises strands β2-β6 of the central β-sheet, together with helices α4 and α5.

Cofactor binding

All structurally-characterised acyltransferases of the GNAT family share the same binding site for the acyl-CoA cofactor, in which the pantotheine moiety is wedged between the diverging β-strands β4 and β5, with the thioacyl group close to the point of divergence. Binding depends on three main features; hydrogen bonding of the pantotheine amide groups with main chain groups on strand β4, a hydrophobic pocket for the dimethyl moiety, and interactions of the diphosphate moiety with main chain NH groups from the β4-α4 connection and the N-terminus of helix α4. The latter interaction accounts for one of the most conserved sequence motifs in GNAT enzymes, designated motif A (12). The preponderance of main chain interactions accounts for the remarkably consistent CoA conformation found in GNAT enzymes (13), despite low sequence identity.

Attempts to co-crystallize Rv1347c with acetyl-CoA have not been successful but the pantotheine moiety can be modeled into the Rv1347c structure in a straightforward way, as in Figure 2, based on the other GNAT family members. The two pantotheine amide groups hydrogen bond to peptide C=O and NH groups from strand β4; these correspond in Rv1347c to 131 C=O and 133 NH. The side chain of Asn173 could also provide an additional hydrogen bond, either directly (after a small conformational adjustment) or *via* a bridging water; this is not a conserved interaction in GNAT enzymes, but is found in all cases where an Asn residue fills this position. As expected, the dimethyl group is adjacent to several hydrophobic side chains, from Ile133 and Val139. It is not possible to model the cofactor reliably beyond this point, however, as a result of

conformational and sequence differences in the β4-α4 connection. The sequence KVNRGFGPL (residues 138-146) approximates the GNAT motif A consensus Q/RxxGxGxxL which corresponds to the β4-α4 connection and diphosphate binding motif in other GNAT enzymes (13). However, a peptide flip between residues 139 and 140 removes one potential peptide NH interaction, and the presence of proline residues at Pro145 and Pro149 disrupts the beginning of helix α4. The closest model for Rv1347c in this region is probably the N-myristoyl transferase (39), which also has a proline (Pro184) at an analogous position to Pro145, and differs from the other GNAT enzymes in the way it binds the diphosphate and adenosyl groups. In Rv1347c, the β4-α4 loop region has high B-values (60-80 $Å^2$) in five of the eight molecules, suggesting that it may undergo conformational adjustment in response to cofactor binding. For these reasons we conclude that it is unrealistic to attempt to model the latter parts of the cofactor into the Rv1347c structure.

The thioacyl group of the cofactor sits at the point of divergence of the two β-strands, β4 and β5. Here, two features stand out. Firstly, the β-bulge in strand β4, which is conserved in all other structurally-characterised GNAT enzymes, is not found in Rv1347c. The effect of the β-bulge in most other GNAT enzymes is to direct two consecutive peptide NH groups towards the acyl oxygen of the acyl-CoA substrate; in Rv1347c, however, only the peptide NH of Ala131 can hydrogen bond to the acyl oxygen. Secondly, two highly conserved residues, His130 and Asp168, are located at this point; His130 is invariant in all homologues of Rv1347c, while Asp168 is only substituted by Glu residues. We conclude that these two residues are involved in catalysis and/or substrate specificity. His130 is located precisely where the middle of the β-bulge is in the other GNAT enzymes, and Asp168, which is hydrogen bonded to His130, is adjacent to another invariant residue, Pro169.

The acyl methyl group in GNAT acetyl-CoA complexes is directed into a hydrophobic pocket. In Rv1347c this pocket develops into two hydrophobic channels (Figure 2), starting at Phe167 and extending 10-12 Å towards the protein surface. Both channels contain ribbons of non-protein electron density that probably arise from the presence of partial-occupancy BOG molecules and indicate potential locations for a long-chain acyl group attached to CoA. The most favorably oriented of these channels (B) passes between helices α4 and α5, its walls formed by Val152, Pro149, Leu148 and Pro145 from α4, Leu179, Cys180 and Ala183 from α5, and Leu203. This channel could accommodate a CoA acyl chain of at least 8 carbons in length, and seems to be a specific feature of Rv1347c. It does not exist in the other GNAT enzymes, where helices α4 and α5 are 2-3 Å closer, and make contact through side chains, and it does not correspond to the hydrophobic groove that binds the 14-carbon acyl group in the myristoyl-CoA complex of N-myristoyl transferase (NMT) (39); the latter is blocked by side chains in Rv1347c.

Acceptor substrate binding site

By analogy with other GNAT family enzymes, the binding site for the acceptor substrate is predicted to be in a deep groove, about 7-8 Å wide, flanked by residues 68-73 (the α2-α3 loop) on one side, and residues 195-196 (from the C-terminal β6-β7 loop) and 18-19 (from the N-terminal region) on the other (Figure 4). This is on the opposite face of the β-sheet from the site of the proposed acyl channel, and is topologically equivalent to the acceptor substrate binding site in aminoglycoside complexes of AAC6'-Iy (34) and AAC2'-Ic (15), the substrate complex of GNA1 (37) and the bisubstrate analog complex of AANAT (40). Although the groove in Rv1347c would fit an aminoglycoside substrate, as judged by superposition of the AAC6'-Iy complex, and slight reorientation of the aminoglycoside, its chemical character appears

unfavorable. In AAC6'-Iy, where the groove is formed between two monomers, one sugar ring stacks between Trp22 from one monomer and Tyr66 from the other, but the predominant feature is the high negative potential from a number of acidic residues (34); the same is true in AAC2'-Ic (15). In Rv1347c, in contrast, the groove contains three arginine residues (from Arg19, Arg172 and Arg196) and only one acidic residue.

Catalytic site

Biochemical evidence suggests that acyl transfer in the GNAT family occurs by direct nucleophilic attack of the amino acceptor group on the thioacyl carbon, the weakness of the thioacyl linkage then leading to breakage of the S-C bond (13). For nucleophilic attack to occur, the amino nitrogen must be uncharged. Depending on the pKa of this group, a nearby general base may or may not be required for deprotonation, either directly or *via* intervening water molecules that link to a more distant base (34,40). In Rv1347c, the two conserved residues, His130 and Asp168, have their side chains oriented upwards into the acceptor substrate binding groove. Either of these residues would be well positioned to act as a general base. In many, but not all, GNAT enzymes a tyrosine residue is positioned to protonate the thiolate anion after collapse of the tetrahedral intermediate. No equivalent tyrosine is present in Rv1347c. The side chain of Thr176 is, however, well placed to play a similar role, either directly or through a bridging water molecule as is proposed for AAC6'-Iy (34). Our modeling of CoA binding suggests that helix α5 may move closer to the cofactor to enable Asn173 to hydrogen bond to a phosphate oxygen and if this happens Thr176 would be brought close (~ 3.5 Å) to the thiolate sulfur.

DISCUSSION

The crystal structure of Rv1347c shows clearly that it belongs to the GCN5-related family of N-acyl transferases known as the GNAT family. Enzymes of this family are functionally diverse, and share only low levels of sequence identity (12), making functional annotation difficult. Many use acetyl-CoA as their cofactor, transferring the acetyl group to a range of acceptor substrates, including lysine residues on histones, the amino groups on aminoglycoside antibiotics and a wide variety of small molecules such as serotonin. However, larger acyl groups than acetyl may also be transferred, such as the 14-carbon myristoyl group in the case of N-myristoyl transferase (39). *Mycobacterium tuberculosis* has particularly rich lipid chemistry, associated with the synthesis and processing of its complex, waxy, cell wall (2), and a great variety of different acyl-CoAs must be available as potential substrates.

Although Rv1347c was originally annotated as an aminoglycoside 6'-N-acetyltransferase, there seems little doubt that this annotation is wrong. The putative acceptor substrate binding groove appears unfavorable for binding aminoglycoside antibiotics, with their high positive charge; in known aminoglycoside-modifying enzymes the binding site is invariably marked by strong negative potential, whereas the groove in Rv1347c is much less so, and contains three arginine residues. Moreover, we note that (i) clinical resistance to aminoglycosides in *M. tuberculosis* has been shown to be due to mutations in the 16S rRNA gene or the gene encoding the S12 ribosomal protein (41), rather than to enzymatic inactivation, and (ii) assays of the *in vitro* activities of two putative APHs (Rv3225c and Rv3817) and two putative AATs (Rv1347c and Rv0262c) showed that only Rv0262c, the previously-characterised AAC(2')-Ic, had any significant aminoglycoside modifying activity (16). Even in the latter, it has been suggested that the *in vivo* function may instead be in mycothiol biosynthesis (15).

What, then, is the biological role of the gene product of Rv1347c? A number of very strong indications point to an involvement in the biosynthesis of mycobactin, the siderophore that is essential for iron acquisition by *M. tuberculosis*. Both bioinformatic analysis (42) and microarray experiments (43) show that expression of the Rv1347c gene product is under the regulatory control of the iron-dependent regulator IdeR; expression of Rv1347c is repressed by iron through IdeR. The closest amino acid sequence homologues of Rv1347c in other genomes are all either uncharacterized or code for the protein IucB, which functions in the biosynthesis of the siderophore aerobactin in strains of *E. coli* and many other bacteria (32). Moreover, the phylogenetic profile of Rv1347c, describing its distribution through 80 bacterial genomes, has all its closest matches among other siderophore biosynthesis proteins. This profile was determined using an improved version (Huang, Y-C., Riddle, P., Triggs, C., Arcus, V. L. and Lott, J. S., unpublished; see http://www.cs.auckland.ac.nz/~yhua033) of a method first proposed by Pellegrini *et al.* (44).

A role in mycobactin biosynthesis would explain why Rv1347c is one of the genes found to be essential for the growth of *M. tuberculosis* in a genome-wide mutational analysis (17); the mycobactin biosynthetic pathway has been previously shown to be essential for growth in macrophages (45). We propose that the specific biochemical function of Rv1347c parallels that of the IucB protein, its closest homologue. The latter catalyzes the CoA-dependent N-acylation of the N^{ε}-hydroxylysine arms of the siderophore aerobactin (32). Mycobactin (Figure 5) also possesses two N^{ε}-hydroxylysine moieties, one of which is cyclised after acetylation (46), and the other of which is acylated and can bear a range of acyl groups in different species (47,48). Variations of the latter acyl group result in two predominant forms of mycobactin, one with a longer, hydrophobic acyl arm, and another with a shorter, more soluble acyl arm (water-soluble

mycobactin, often referred to as carboxymycobactin) (46,48). Interestingly, although most of the genes implicated in mycobactin biosynthesis have been identified and associated with proposed biochemical steps in the pathway (46), neither the enzyme(s) involved in the acylation of the N^{ε}-hydroxylysine arms, nor the precise substrates involved, are known.

The Rv1347c gene is flanked, in the *M. tuberculosis* genome, by other iron-dependent genes that are under the control of the same regulator, IdeR, including genes for a putative acyl carrier protein (Rv1344), an acyl-CoA synthase (Rv1345, *fadD33*) and an acyl-CoA dehydrogenase (Rv1346, *fadE14*). Transposon mutagenesis in *Mycobacterium smegmatis* has indicated a direct role for the Rv1345 gene product in mycobactin biosynthesis and implied that this cluster acts, together with an unidentified acyltransferase, to generate the correct sidechains on the siderophore (49). We propose that it is the adjacent gene, Rv1347c, that codes for this acyltransferase.

In order to explore the hypothesis that the true substrate for the Rv1347c gene product is one of the N^{ε}-hydroxylysine (NHL) side chains of mycobactin, we placed an NHL moiety into the acceptor binding substrate groove and modeled the tetrahedral intermediate that would be formed when NHL attacks the thioacyl carbon (Figure 6). In this intermediate, the negatively charged oxygen is directed towards the main chain NH of residue 131, 2.9 Å away and the acyl group is directed towards Phe167 and the hydrophobic channel. Importantly the N^{ε}-hydroxyl group can form a hydrogen bond (2.5 Å) with the Nδ1 atom of His130, and the NHL nitrogen is about 3.3 Å from Asp168 Oδ1. This strongly indicates functional roles for these two residues, which are among the few residues that are conserved between Rv1347c and all IucB proteins; we conclude that His130 provides specific recognition of the hydroxyl group on NHL side chains, and Asp168

(which is replaced only by Glu) is the base that ensures deprotonation of the attacking nitrogen. Our proposed location for the NHL binding site, in which its α-carbon lies between Trp69 and Tyr71 on one side, and Arg19 and Arg196 on the other, also corresponds to the location of the aminoglycoside substrate in AAC6'-Iy (34) and of the substrate portion of the bisubstrate analog in AANAT (40), further validating this model.

The nature of the acyl group(s) that can be transferred by Rv1347c is unknown, given that *M. tuberculosis* can synthesize mycobactins with several different acyl groups attached to the NHL side chain. The predominant mycobactins are a membrane-associated form with a long, hydrophobic acyl chain of 18-20 carbons on the NHL arm, and a soluble form with 5-9 carbons (47). The hydrophobic tunnel adjacent to the CoA binding site could accommodate such chains nicely. Moreover, functional studies on the Rv1347c gene product further support the notion that acyl groups longer than acetyl are transferred. These functional studies failed to demonstrate any aminoglycoside N-acetyltransferase activity, but did demonstrate thioesterase activity with numerous acyl-CoAs, with a preference for longer acyl chains (16). Since one component of N-acyl transfer involves hydrolysis of the thioester bond, thioesterase activity is consistent with an N-acyl transfer function providing an appropriate acceptor substrate is bound. The fact that larger acyl-CoA substrates are hydrolysed is consistent with our structural and modeling results, and with the proposed role in mycobactin biosynthesis.

Acknowledgments

We gratefully acknowledge Chris Squire for help with data collection, Shaun Lott for Figure 5, and Katherine Kantardjeff for valuable discussions. The work was supported by the Health Research Council of New Zealand and was performed as part of the International TB Structural Genomics Consortium (http://www.doe-mbi.ucla.edu/TB/). The work at LLNL was funded under

the NIH P50 grant GM62410. LLNL is operated by the University of California for the US DOE under contract W-7405-ENG-48.

Deposited data

The atomic coordinates and structure amplitudes have been deposited with the Protein Data Bank (http://www.rcsb.org/) with the accession code xxxx.

References

1. Dye, C., Scheele, S., Dolin, P., Pathania, V., and Raviglione, M. C. (1999) *J. Am. Med. Assoc.* **282**, 677-686
2. Brennan, P. J., and Nikaido, H. (1995) *Annu. Rev. Biochem.* **64**, 29-63
3. Parrish, N. M., Dick, J. D., and Bishai, W. R. (1998) *Trends Microbiol.* **6**, 107-112
4. McKinney, J. D., Honer zu Bentrup, K., Munoz-Elias, E. J., Miczak, A., Chen, B., Chan, W. T., Swenson, D., Sacchettini, J. C., Jacobs, W. R., Jr., and Russell, D. G. (2000) *Nature* **406**, 735-738
5. O'Regan, A., and Joyce-Brady, M. (2001) *Brit. Med. J.* **323**, 635b-
6. Stokstad, E. (2000) *Science* **287**, 2391
7. Barker, K. F. (1999) *Brit. J. Clin. Pharmacol.* **48**, 109-124
8. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K. *et al*., and Barrell, B. G. (1998) *Nature* **393**, 537-544.
9. Smith, C. A., and Baker, E. N. (2002) *Curr. Drug Targets - Infect. Disord.* **2**, 143-160.

10. Hon, W. C., McKay, G. A., Thompson, P. R., Sweet, R. M., Yang, D. S. C., Wright, G. D., and Berghuis, A. M. (1997) *Cell* **89**, 887-895

11. Daigle, D. M., McKay, G. A., Thompson, P. R., and Wright, G. D. (1998) *Chem. Biol.* **6**, 11-18

12. Neuwald, A. F., and Landsman, D. (1997) *Trends Biochem. Sci.* **22**, 154-155

13. Dyda, F., Klein, D. C., and Hickman, A. B. (2000) *Annu. Rev. Biophys. Biomol. Struct.* **29**, 81-103

14. Hegde, S. S., Javid-Majd, F., and Blanchard, J. S. (2001) *J. Biol. Chem.* **276**, 45876-45881

15. Vetting, M. W., Hegde, S. S., Javid-Majd, F., Blanchard, J. S., and Roderick, S. L. (2002) *Nature Struct. Biol.* **9**, 653-658

16. Draker, K. A., Boehr, D. D., Elowe, N. H., Noga, T. J., and Wright, G. D. (2003) *J. Antibiot. (Tokyo)* **56**, 135-142

17. Sassetti, C. M., Boyd, D. H., and Rubin, E. J. (2003) *Mol. Microbiol.* **48**, 77-84

18. Wolf, E., Vassilev, A., Makino, Y., Sali, A., Nakatani, Y., and Burley, S. K. (1998) *Cell* **94**, 439-449

19. Doublie, S. (1997) *Methods Enzymol.* **276**, 523-530

20. Segelke, B. W. (2001) *J. Cryst. Growth* **232**, 553-562

21. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307-326

22. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallog.* **D55**, 849-861

23. Terwilliger, T. C. (1999) *Acta Crystallogr.* **D55**, 1863-1871

24. Levitt, D. G. (2001) *Acta Crystallogr.* **D57**, 1013-1039

25. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr.* **A47**, 110-119

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

27. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283-291

28. Hendrickson, W. A. (1991) *Science* **254**, 51-58

29. Jones, S., and Thornton, J. M. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 13-20

30. Modis, Y., and Wierenga, R. (1998) *Structure* **6**, 1345-1350

31. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403-410

32. De Lorenzo, V., Bindereif, A., Paw, B. H., and Neilands, J. B. (1986) *J. Bacteriol.* **165**, 570-578

33. Krissinel, E., and Henrick, K. (2004) *Acta Crystallogr*. **D60**, 2256-2268

34. Vetting, M. W., Magnet, S., Nieves, E., Roderick, S. L., and Blanchard, J. S. (2004) *Chem. Biol.* **11**, 565-573

35. He, H., Ding, Y., Bartlam, M., Sun, F., Le, Y., Qin, X., Tang, H., Zhang, R., Joachimiak, A., Liu, J., Zhao, N., and Rao, Z. (2003) *J. Mol. Biol.* **325**, 1019-1030

36. Hickman, A. B., Klein, D. C., and Dyda, F. (1999) *Mol. Cell* **3**, 23-32

37. Peneff, C., Mengin-Lecreulx, D., and Bourne, Y. (2001) *J. Biol. Chem.* **276**, 16328-16334.

38. Angus-Hill, M. L., Dutnall, R. N., Tafrov, S., Sternglanz, R., and Ramakrishnan, V. (1999) *J. Mol. Biol.* **294**, 1311-1325

39. Farazi, T. A., Waksman, G., and Gordon, J. I. (2001) *Biochemistry* **40**, 6335-6343

40. Hickman, A. B., Namboodiri, M. A. A., Klein, D. C., and Dyda, F. (1999) *Cell* **97**, 361-369

41. Blanchard, J. S. (1996) *Annu. Rev. Biochem.* **65**, 215-239

42. Makita, Y., Terai, G., Mitaku, S., Takagi, T., and Nakai, K. (2002) *Genome Informatics* **13**, 297-298

43. Rodriguez, G. M., Voskuil, M. I., Gold, B., Schoolnik, G. K., and Smith, I. (2002) *Infect. Immun.* **70**, 3371-3381

44. Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D., and Yeates, T. O. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 4285-4288

45. De Voss, J. J., Rutter, K., Schroeder, B. G., Su, H., Zhu, Y., and Barry, C. E. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 1252-1257

46. De Voss, J. J., Rutter, K., Schroeder, B. G., and Barry, C. E. (1999) *J. Bacteriol.* **181**, 4443-4451

47. Gobin, J., Moore, C. H., Reeve, J. R., Wong, D. K., Gibson, B. W., and Horwitz, M. A. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 5189-5193

48. Ratledge, C. (2004) *Tuberculosis* **84**, 110-130

49. LaMarca, B. B. D., Zhu, W., Arceneaux, J. E. L., Byers, B. R., and Lundrigan, M. D. (2004) *J. Bacteriol.* **186**, 374-382

50. Carson, M. (1991) *J. Appl. Crystallogr.* **24**, 961-985

51. Esnouf, R. M. (1999) *Acta Crystallogr. sect. D* **55**, 938-940

52. Nicholls, A., Sharp, K., and Honig, B. (1991) *Proteins* **11**, 281-296

Footnote

Abbreviations

ORF, open reading frame; AAC, aminoglycoside N-acetyltransferase; APH, aminoglycoside O-phosphotransferase; CoA, coenzyme-A; GNAT, GCN5-related N-acetyltransferase; TTR, tabtoxin resistance protein; GNA1, glucosamine-phosphate N-acetyltransferase; NMT, N-myristoyl transferase; rms, root-mean-square; PEG, polyethylene glycol; NHL, N^{ε}-hydroxylysine; SeMet, selenomethionine; BOG, β-octylglucoside; BME, β-mercaptoethanol; rTEV, recombinant tobacco etch virus protease.

Figure legends

Figure 1. Folding of Rv1347c. (A) Topology diagram showing the packing of helices (circles) on either face of the central β-sheet. 3_{10}-helices are labeled η1 and η2. The location of the active site is marked with a star. (B) Ribbon diagram, shown in stereo. Helices are shown in light blue, 3_{10}-helices in dark blue and β-strands in green. Two conserved residues, His130 and Asp168, mark the active site, where strands β4 and β5 diverge. The view looks down between the loops that enclose the proposed acceptor substrate binding site between the α2-α3 loop, the β6-β7 loop, and a loop from the extended N-terminal segment. Figure drawn with RIBBONS (48).

Figure 2. Non-protein electron density in the Rv1347c structure. The density, from a simulated annealing, NCS-averaged, F_o-F_c "omit" map, contoured at 3σ, follows two hydrophobic channels leading from the acyl-CoA binding site and is shown as a grey cloud. This density, which is present in all 8 molecules in the asymmetric unit, is attributed to bound β-octylglucoside molecules. BOG has been modeled into the left-hand channel (B), but it is the right-hand channel (A) that is the proposed site for the acyl chain of a substrate acyl-CoA molecule. A patch of density at the molecular surface (bottom left) represents the binding site for the BOG head group; the density leading from this into channel B is not continuous in this averaged map however, and BOG has only been modeled into 3 of the 8 molecules. In this stereo figure, the pantotheine arm of CoA is modeled into the conserved GNAT CoA binding site, as described in the text.

Figure 3. Multiple sequence alignment of *M. tuberculosis* Rv1347c and its closest homologs. The sequence numbering is that of Rv1347c, with its secondary structure elements shown above. Invariant residues (white on red background) and conservatively substituted residues (red) are indicated. Sequences shown are for hypothetical proteins from *B. halodurans*, *Anabaena sp*., *Halobacterium sp*., *R. leguminosarum* and *S. coelicolor*, and siderophore biosynthesis proteins

(IucB) from *R. meliloti*, *E. coli*, *B. bacteriovorus*, *S. flexneri*, *S. boydii*, *K. pneumoniae*, *V. mimicus*, and *Y. pestis*. In some cases, N-terminal extensions of 100-120 residues, not shared by Rv1347c, have been omitted from the figure. Figure prepared using ESPript (49).

Figure 4. Molecular surface of Rv1347c, showing proposed acceptor substrate binding cleft. The thioacyl group of CoA (shown in stick mode) can be seen at the bottom of this cleft. Below the thioacyl group is the hydrophobic tunnel that is one of the sites of β-octylglucoside binding in the crystal structure, and is the proposed binding site for long-chain acyl substituents; this tunnel leads away from the CoA to the opposite surface of the protein. Figure drawn with GRASP (50), and in a similar orientation to Figure 2.

Figure 5. Chemical structures of the proposed substrates for Rv1347c, showing their constituent parts. (A) An acyl-Coenzyme A molecule. (B) Mycobactin T, the major mycobactin siderophore produced by *M. tuberculosis*. The R group attached to the central N^{ε}-hydroxylysine side chain of mycobactin represents the long, hydrophobic acyl chain that is proposed to be transferred from the acyl-Coenzyme A by Rv1347c. Figures courtesy of Drs. Jodie Johnston and Shaun Lott.

Figure 6. Model for the proposed reaction intermediate. The pantetheine arm of the CoA molecule is shown making hydrogen bonds with main chain atoms from strand β4, as in other GNAT family enzymes. The N^{ε}-hydroxylysine moiety has been positioned such that its amino nitrogen completes a tetrahedral intermediate at the thioacyl carbon of the acyl-CoA, with the rest of the NHL corresponding in position to that of the substrate portion of the bisubstrate analog in AANAT (38). His130, proposed to function in substrate recognition, and Asp168, the proposed catalytic base, are shown interacting with the NHL moiety. A longer acyl chain, in place of the acetyl group shown here, would extend down into a hydrophobic channel, past Phe167 (see text).

Table 1. Data collection and processing statistics				
Data collection	Native	SeMet peak	SeMet inflection	SeMet remote
Wavelength (Å)	1.5418	0.9789	0.9795	0.9118
Resolution (Å) (outer shell)	40 - 2.15 (2.23 - 2.15)	50 – 2.25 (2.35 - 2.25)	50 – 2.25 (2.35 - 2.25)	50 – 2.25 (2.35 - 2.25)
Total reflections[a]	488508	592154	591959	585405
Unique reflections[a]	96453	157535	157591	157507
Completeness (outer shell) (%)	99.9 (100.0)	99.9 (99.8)	99.9 (99.8)	99.9 (99.7)
I/σ (outer shell)	21.2 (3.1)	30.1 (5.9)	30.8 (6.1)	28.0 (5.9)
R_{merge} (outer shell)	0.092 (0.551)	0.069 (0.316)	0.068 (0.305)	0.066 (0.294)

[a]Friedel pairs not merged for SeMet data

Table 2. Refinement statistics	
Resolution range (Å)	40.0 – 2.2
Number of reflections (working/test)	83972/1824
R factor/R_{free}	0.227/0.257
Number of atoms (non-hydrogen)	
Protein (8 molecules)	12812
Solvent	650
β-octylglucoside (4 mols)	80
Rms deviations from ideality	
Bonds (Å)	0.006
Angles (deg.)	1.33
Average B factors ($Å^2$)	
Protein atoms	36.0
Water molecules	33.1
β-octylglucoside	58.8
Residues in most favored region (%)	89.2

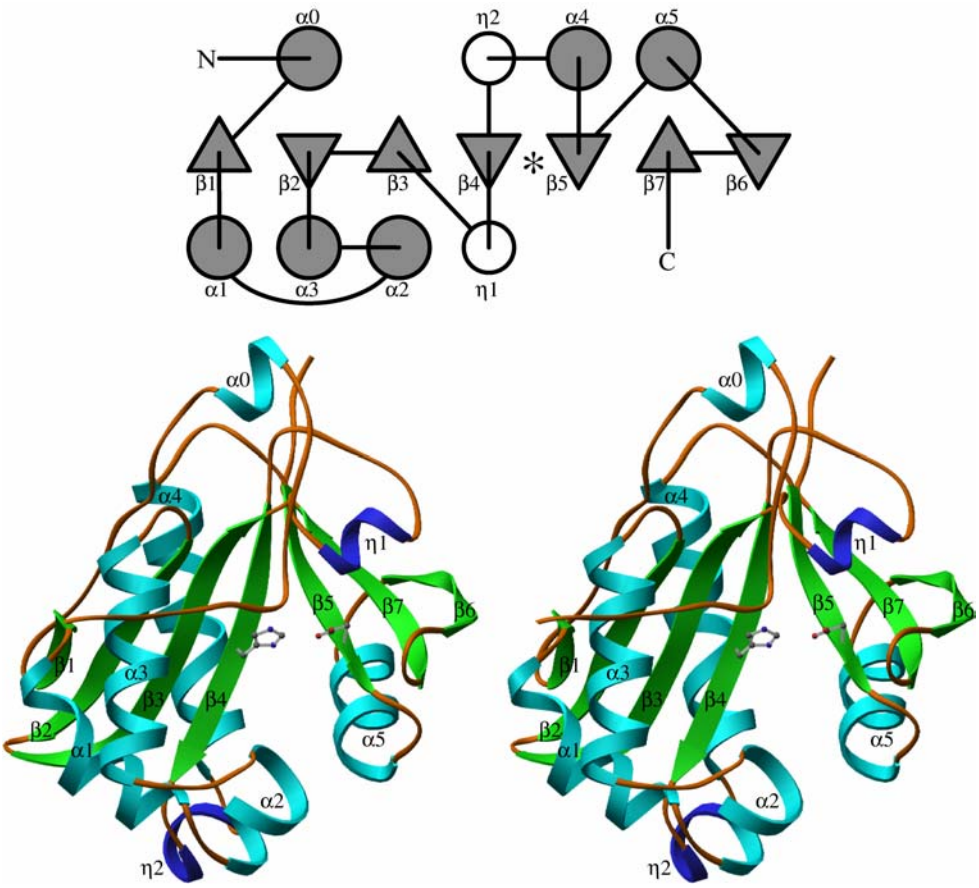

α0
η2
α4
α5
N
β1
β2
β3
β4
*
β5
β7
β6
C
α1
α3
α2
η1
α0
α4
η1
β5
β7
β6
β1
α3
β3
β4
β2
α1
α5
α2
η2
α0
α4
η1
β5
β7
β6
β1
α3
β3
β4
β2
α1
α5
α2
η2

Figure 1

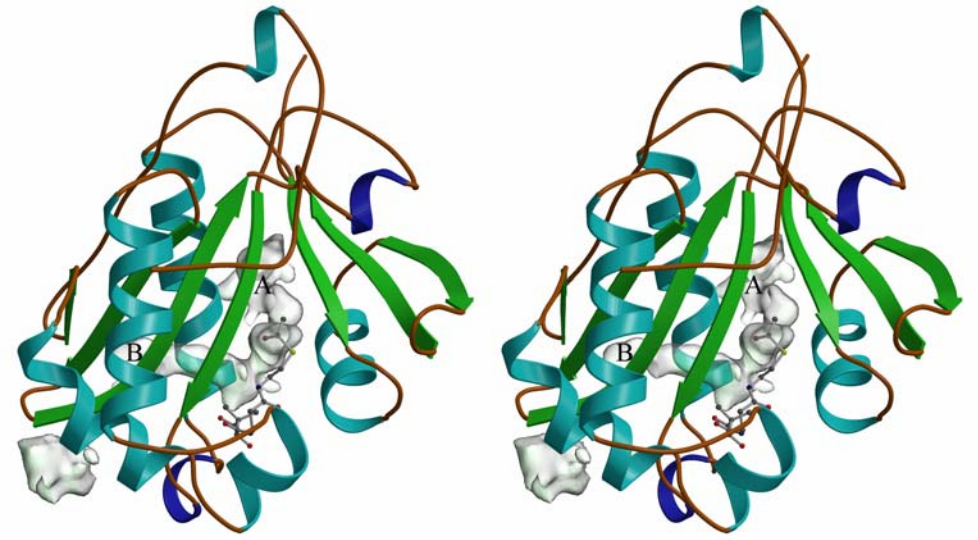
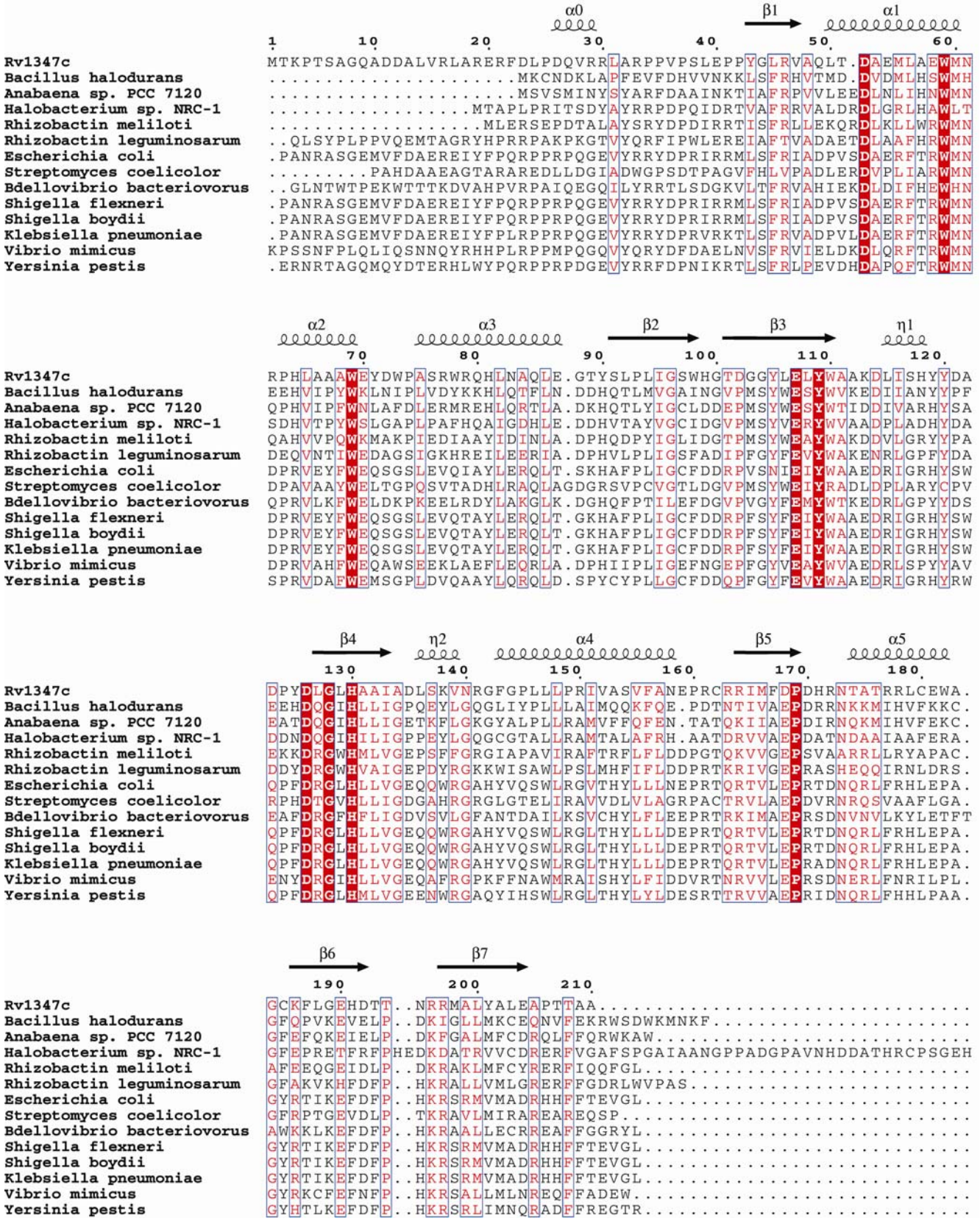


Figure 2


α0 β1 α1
1 10 20 30 40 50 60
Rv1347c MTKPTSAGQADDALVRLARERFDLPDQVRRLARPPVPSLEPPYGLRVAQLT.DAEMLAEWMN
Bacillus halodurans .....MKCNDKLAPFEVFDHVVNKKLSFRHVTMD.DVDMLHSWMH
Anabaena sp. PCC 7120 .....MSVSMINYSYARFDAAINKTIAFRPVVLEEDLNLIHNWMN
Halobacterium sp. NRC-1 .....MTAPLPRITSDYAYRRPDPQIDRTVAFRRVALDRDLGRLHAWLT
Rhizobactin meliloti .....MLERSEPDTALAYSRYDPDIRRTISFRLLEKQRDLKLLWRWMN
Rhizobactin leguminosarum ..QLSYPLPPVQEMTAGRYHPRRPAKPKGTVYQRFIPWLEREIAFTVADAETDLAAFHRWMN
Escherichia coli .PANRASGEMVFDAERELYFPQRPPRPQGEVYRRYDPRIRRMLSFRIADPVSDAERFTRWMN
Streptomyces coelicolor .....PAHDAAEAGTARAREDLLDGIADWGPSDTPAGVFHLVPADLERDVPLIARWMN
Bdellovibrio bacteriovorus ..GLNTWTPEKWTTTKDVAHPVRPAIQEGQILYRRTLSDGKVLTFRVAHIEKDLDIFHEWHN
Shigella flexneri .PANRASGEMVFDAERELYFPQRPPRPQGEVYRRYDPRIRRMLSFRIADPVSDAERFTRWMN
Shigella boydii .PANRASGEMVFDAERELYFPQRPPRPQGEVYRRYDPRIRRMLSFRIADPVSDAERFTRWMN
Klebsiella pneumoniae .PANRASGEMVFDAERELYFPLRPPRPQGEVYRRYDPRVRKTLSFRVADPVLDAERFTRWMN
Vibrio mimicus KPSSNFPLQLIQSNNQYRHHPLRPPMPQGQVYQRYDFDAELNVSFRVIELDKDLQRFTRWMN
Yersinia pestis .ERNRTAGQMQYDTERHLWYPQRPPRPDGEVYRRFDPNIKRTLSFRLPEVDHDAPQFTRWMN
α2 α3 β2 β3 η1
70 80 90 100 110 120
Rv1347c RPHLAAAWEYDWPASRWRQHLNAQLE.GTYSLPLIGSWHGTDGGYLELYWAAKDLISHYYDA
Bacillus halodurans EEHVIPYWKLNIPLVDYKKHLQTFLN.DDHQTLMVGAINGVPMSYWESYWVKEDIIANYYPF
Anabaena sp. PCC 7120 QPHVIPFWNLAFDLERMREHLQRTLA.DKHQTLYIGCLDDEPMSYWESYWTIDDIVARHYSA
Halobacterium sp. NRC-1 SDHVTPYWSLGAPLPAFHQAIGDHLE.DDHVTAYVGCIDGVPMSYVERYWVAADPLADHYDA
Rhizobactin meliloti QAHVVPQWKMAKPIEDIAAYIDINLA.DPHQDPYIGLIDGTPMSYWEAYWAKDDVLGRYYPA
Rhizobactin leguminosarum DEQVNTIWEDAGSIGKHREILEERIA.DPHVLPLIGSFADIPFGYFEVYWAKENRLGPFYDA
Escherichia coli DPRVEYFWEQSGSLEVQIAYLERQLT.SKHAFPLIGCFDDRPVSNIEIYWAAEDRIGRHYSW
Streptomyces coelicolor DPAVAAYWELTGPQSVTADHLRAQLAGDGRSVPCVGTLDGVPMSYWEIYRADLDPLARYCPV
Bdellovibrio bacteriovorus QPRVLKFWELDKPKEELRDYLAKGLK.DGHQFPTILEFDGVPVGYFEMYWTKEDRLGPYYDS
Shigella flexneri DPRVEYFWEQSGSLEVQTAYLERQLT.GKHAFPLIGCFDDRPFSYFEIYWAAEDRIGRHYSW
Shigella boydii DPRVEYFWEQSGSLEVQTAYLERQLT.GKHAFPLIGCFDDRPFSYFEIYWAAEDRIGRHYSW
Klebsiella pneumoniae DPRVEYFWEQSGSLEVQTAYLERQLT.GKHAFPLIGCFDDRPFSYFEIYWAAEDRIGRHYSW
Vibrio mimicus DPRVAHFWEQAWSEEKLAEFLEQRLA.DPHIIPLIGEFNGEPFGYVEAYWVAEDRLSPYYAV
Yersinia pestis SPRVDAFWEMSGPLDVQAAYLQRQLD.SPYCYPLLGCFDDQPFGYFEVYWAAEDRIGRHYRW
β4 η2 α4 β5 α5
130 140 150 160 170 180
Rv1347c DPYDLGLHAAIADLSKVNRGFGPLLLPRIVASVFANEPRCRRIMFDPDHRNTATRRLCEWA.
Bacillus halodurans EEHDQGIHLLIGPQEYLGQGLIYPLLLAIMQQKFQE.PDTNTIVAEPDRRNKKMIHVFKKC.
Anabaena sp. PCC 7120 EATDQGIHLLIGETKFLGKGYALPLLRAMVFFQFEN.TATQKIIAEPDIRNQKMIHVFEKC.
Halobacterium sp. NRC-1 DDNDQGIHILIGPPEYLGQGCGTALLRAMTALAFRH.AATDRVVAEPDATNDAAIAAFERA.
Rhizobactin meliloti EKKDRGWHMLVGEPSFFGRGIAPAVIRAFTRFLFLDDPGTQKVVGEPSVAARRLLRYAPAC.
Rhizobactin leguminosarum DDYDRGWHVAIGEPDYRGKKWISAWLPSLMHFIFLDDPRTKRIVGEPRASHEQQIRNLDRS.
Escherichia coli QPFDRGLHLLVGEQQWRGAHYVQSWLRGVTHYLLLNEPRTQRTVLEPRTDNQRLFRHLEPA.
Streptomyces coelicolor RPHDTGVHLLIGDGAHRGRGLGTELIRAVVDLVLAGRPACTRVLAEPDVRNRQSVAAFLGA.
Bdellovibrio bacteriovorus EAFDRGFHFLIGDVSVLGFANTDAILKSVCHYLFLEEPRTRKIMAEPRSDNVNVLKYLETFT
Shigella flexneri QPFDRGLHLLVGEQQWRGAHYVQSWLRGLTHYLLLDEPRTQRTVLEPRTDNQRLFRHLEPA.
Shigella boydii QPFDRGLHLLVGEQQWRGAHYVQSWLRGLTHYLLLDEPRTQRTVLEPRTDNQRLFRHLEPA.
Klebsiella pneumoniae QPFDRGLHLLVGEQQWRGAHYVQSWLRGLTHYLLLDEPRTQRTVLEPRADNQRLFRHLEPA.
Vibrio mimicus ENYDRGIHLLVGEQAFRGPKFFNAWMRAISHYLFIDDVRTNRVVLEPRSDNERLFNRILPL.
Yersinia pestis QPFDRGLHMLVGEENWRGAQYIHSWLRGLTHYLYLDESRTTRVVAEPRIDNQRLFHHLPAA.
β6 β7
190 200 210
Rv1347c GCKFLGEHDTT..NRRMALYALEAPTTAA.....
Bacillus halodurans GFQPVKEVELP..DKIGLLMKCEQNVFEKRWSDWKMNKF.....
Anabaena sp. PCC 7120 GFEFQKEIELP..DKFGALMFCDRQLFFQRWKAW.....
Halobacterium sp. NRC-1 GFEPRETFRFPHEDKDATRVVCDRERFVGAFSPGAIAANGPPADGPAVNHDDATHRCPSGEH
Rhizobactin meliloti AFEEQGEIDLP..DKRAKLMFCYRERFIQQFGL.....
Rhizobactin leguminosarum GFAKVKHFDFP..HKRALLVMLGRERFFGDRLWVPAS.....
Escherichia coli GYRTIKEFDFP..HKRSRMVMADRHHFFTEVGL.....
Streptomyces coelicolor GFRPTGEVDLP..TKRAVLMIRAREAREQSP.....
Bdellovibrio bacteriovorus AWKKLKEFDFP..HKRAALLECRREAFFGGRYL.....
Shigella flexneri GYRTIKEFDFP..HKRSRMVMADRHHFFTEVGL.....
Shigella boydii GYRTIKEFDFP..HKRSRMVMADRHHFFTEVGL.....
Klebsiella pneumoniae GYRTIKEFDFP..HKRSRMVMADRHHFFTEVGL.....
Vibrio mimicus GYRKCFEFNFP..HKRSALLMLNREQFFADEW.....
Yersinia pestis GYHTLKEFDFP..HKRSRLIMNQRADFFREGTR.....

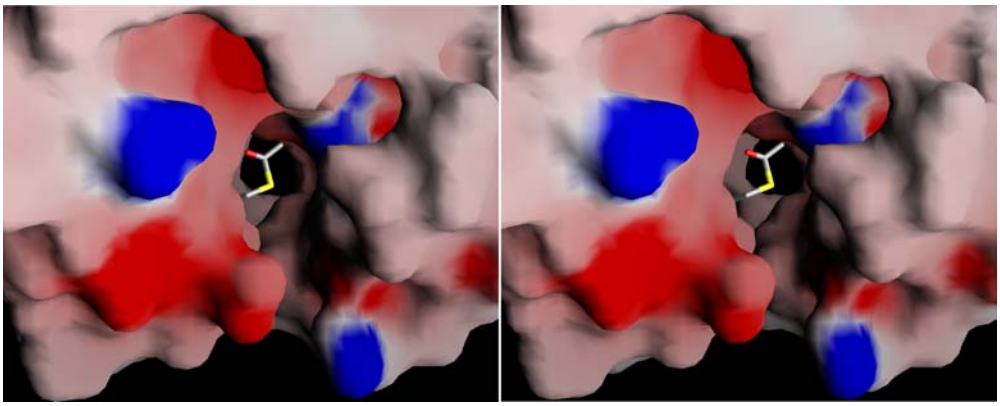


Figure 4

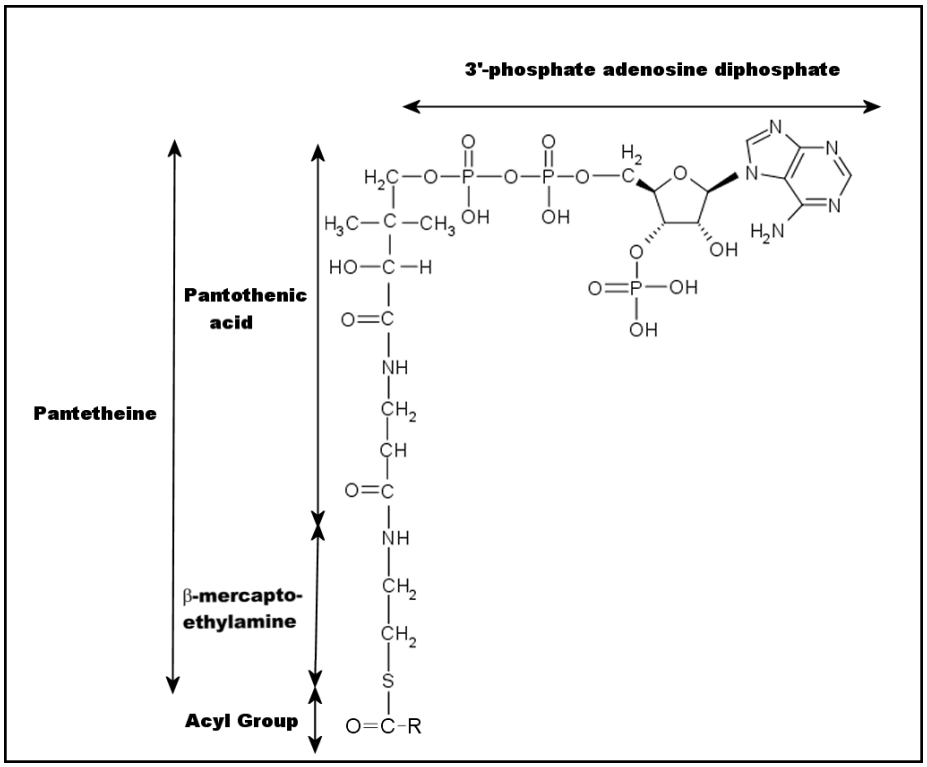


Figure 5A

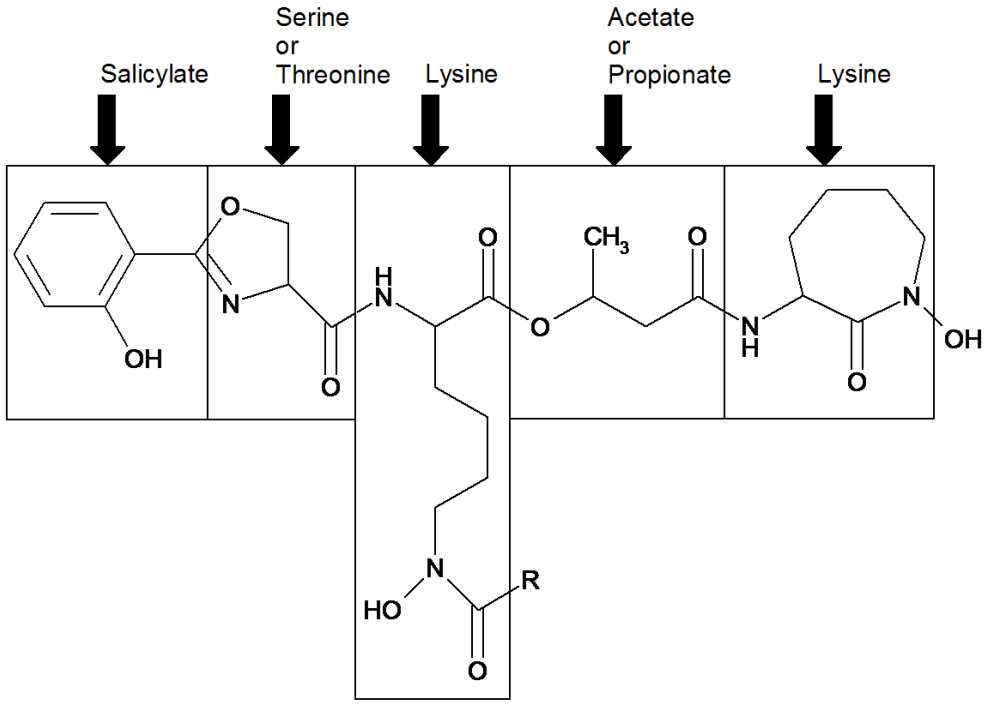


Figure 5B

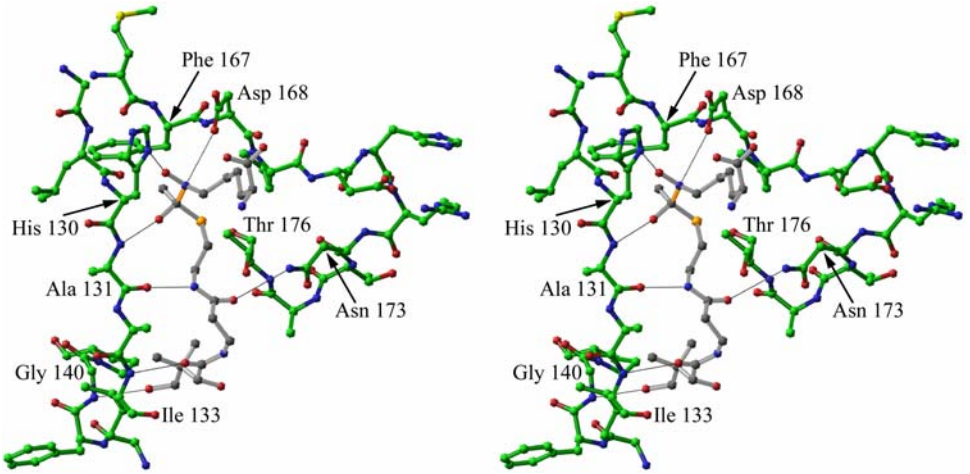

Phe 167
Asp 168
His 130
Thr 176
Ala 131
Asn 173
Gly 140
Ile 133
Phe 167
Asp 168
His 130
Thr 176
Ala 131
Asn 173
Gly 140
Ile 133

Figure 6