Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# Structure of arylamine *N*-acetyltransferase from *Mycobacterium tuberculosis* determined by cross-seeding with the homologous protein from *M. marinum*: triumph over adversity

**Areej Abuhammad,[a,b,c] Edward D. Lowe,[b] Michael A. McDonough,[d] Patrick D. Shaw Stewart,[e] Stefan A. Kolek,[e] Edith Sim[a,f] and Elspeth F. Garman[b]***

[a]Department of Pharmacology, University of Oxford, Mansfield Road, Oxford OX1 3QT, England, [b]Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, England, [c]Faculty of Pharmacy, University of Jordan, Queen Rania Street, Amman 11942, Jordan, [d]Department of Chemistry, University of Oxford, Chemistry Research Laboratory, Mansfield Road, Oxford OX1 3TA, England, [e]Douglas Instruments, Douglas House, East Garston, Hungerford, Berkshire RG17 7HD, England, and [f]Faculty of Science, Engineering and Computing, Kingston University, Penrhyn Road, Kingston KT1 2EE, England

Correspondence e-mail: elspeth.garman@bioch.ox.ac.uk

Received 16 April 2013
Accepted 31 May 2013

**PDB Reference:** arylamine *N*-acetyltransferase, 4bgf

Arylamine *N*-acetyltransferase from *Mycobacterium tuberculosis* (TBNAT) plays an important role in the intracellular survival of the microorganism inside macrophages. Medicinal chemistry efforts to optimize inhibitors of the TBNAT enzyme have been hampered by the lack of a three-dimensional structure of the enzyme. In this paper, the first structure of TBNAT, determined using a lone crystal produced using cross-seeding with the homologous protein from *M. marinum*, is reported. Despite the similarity between the two enzymes (74% sequence identity), they show distinct physical and biochemical characteristics. The structure elegantly reveals the characteristic features of the protein surface as well as details of the active site of TBNAT relevant to drug-discovery efforts. The crystallographic analysis of the diffraction data presented many challenges, since the crystal was twinned and the habit possessed pseudo-translational symmetry.

## 1. Introduction

*Mycobacterium tuberculosis* is still the leading cause of death worldwide by an infectious agent. Despite the recent reduction in cases of tuberculosis (TB), the global burden remains enormous. In 2011, there were an estimated 8.7 million new cases of TB and 13% of them were co-infections with HIV (World Health Organization, 2012).

*M. tuberculosis* grows slowly and is protected by its unique impenetrable cell wall, which consists of mycolic acids, glycoproteins and the complex polysaccharides arabinogalactan and lipoarabinomannan (Jankute *et al.*, 2012). It can survive within macrophages for many years (latent infection) further protected by granuloma formation. It may emerge as an active infection in immunocompromised individuals, for example. This has meant that the disease is particularly difficult to treat, requiring a regimen of usually four antibiotics for six months (World Health Organization, 2012). The emergence of multidrug-resistant (MDR), extensively drug-resistant (XDR) and the recently reported totally drug-resistant (TDR) TB further complicates the control and the eradication of the disease (Shenoi & Friedland, 2009; Udwadia *et al.*, 2011; Rowland, 2012).

Despite extensive research for more than 100 years, the battle against TB is still ongoing. The World Health Organization's Stop TB Strategy involves research funded by governments, charitable foundations and nongovernmental organizations (Squire *et al.*, 2006; Lienhardt *et al.*, 2012). The TB Alliance is one of these organizations and their pipeline of potential therapies 'has created and advanced the largest TB drug pipeline in history' according to Ginsberg (2011). Targets for latent TB treatment include energy metabolism (Andries

*et al.*, 2005), cell-wall synthesis and lipid metabolism (Zhang, 2005). In particular, the role of cholesterol catabolism has gained recent attention owing to its role in cell entry and persistence (Griffin *et al.*, 2012; Ouellet *et al.*, 2011; Kendall *et al.*, 2010; Van der Geize *et al.*, 2007).

Several completely novel drugs are now in clinical development, foremost of which is the ATPase inhibitor bedaquiline (TMC-207; Cole & Riccardi, 2011; Andries *et al.*, 2005; Grosset *et al.*, 2012). Several other compounds have also qualified for preclinical development; for example, the benzothiazinones, which block cell-wall formation through inhibition of arabinan synthesis (Makarov *et al.*, 2009; Batt *et al.*, 2012). Interestingly, cholesterol catabolism inhibitors have been shown to exhibit antimycobacterial activity against persistent and MDR strains of *M. tuberculosis* in a murine model (Ahmad *et al.*, 2006). However, none of the current compounds being supported through the TB Global Alliance (http://www.tballiance.org/) specifically target cholesterol catabolism in mycobacteria. Therefore, the development of novel inhibitors targeting this pathway could provide new strategies for the treatment and control of latent TB. A gene cluster involved in cholesterol catabolism is known to be essential for the survival of *M. bovis* BCG within the macrophage from whole-gene studies (Sassetti *et al.*, 2003) and targeted gene approaches (Yam *et al.*, 2009; Lack *et al.*, 2010; Bhakta *et al.*, 2004). This gene cluster is almost identical in *M. bovis* BCG and *M. tuberculosis*. The *nat* gene encoded on this cluster has the same sequence in *M. bovis* BCG and *M. tuberculosis*..

Deleting the *nat* gene in *M. bovis* BCG resulted in delayed growth and caused morphological changes of the BCG bacilli. Moreover, the *nat* gene knockout resulted in depleted mycolic acid and virulence-lipid content (phthiocerol dimycocerosate and cord factor), which rendered the mycobacteria sensitive to gentamicin (Bhakta *et al.*, 2004). Arylamine *N*-acetyltransferase is a 31 kDa cytosolic enzyme that is found in *M. tuberculosis* (TBNAT) as well as in many other organisms (Sim *et al.*, 2008). The NAT enzymes are known to catalyse the transfer of an acetyl group from acetyl-CoA (Ac-CoA) to an arylamine substrate (Fig. 1) through a conserved cysteine residue by a ping-pong bi-bi mechanism (Sinclair *et al.*, 2000). The ability of the enzyme to utilize propionyl-CoA (Pr-CoA) as well as Ac-CoA (Lack *et al.*, 2009), the intermediate cofactors in virulence-lipid synthesis, provides a good explanation of these findings as well as relating NAT to cholesterol catabolism. Interestingly, NAT inhibition resulted in similar changes in cell-wall lipid composition, morphology and intracellular survival to those observed upon deleting the gene (Westwood *et al.*, 2010). These findings make NAT an attractive target for TB drug therapy. Because of the ever-present problem of resistance, there is a requirement to explore many new targets to ensure the future of effective TB therapies.

The impact of protein structure determination in target identification and drug discovery has encouraged collaborative initiatives to solve the three-dimensional structures of *M. tuberculosis* proteins. The TB Structural Genomics Consortium (TBSGC; USA) is one of the organizations that has significantly contributed to structure determination and methods development (Terwilliger *et al.*, 2003; Arcus *et al.*, 2006; Murillo *et al.*, 2007; Chim *et al.*, 2009, 2011). The three-dimensional structures of the other proteins which are products of the same gene cluster as TBNAT (HsaAB, HsaC and HsaD) are now available and are expected to facilitate inhibitor design (Lack *et al.*, 2008, 2010; Yam *et al.*, 2009; Dresen *et al.*, 2010). NATs from closely related mycobacteria have been studied previously. NAT inhibitors with antimycobacterial activity have been identified by high-throughput screening (HTS) using homologous NAT enzymes (Westwood *et al.*, 2011). Although the protein fold is highly conserved within the NAT family (ten structures; see §3.7), the unique selectivity profile of NAT enzymes was clearly observed in this HTS (Laurieri *et al.*, 2010; Westwood *et al.*, 2010; Abuhammad *et al.*, 2012).

The high sequence similarity (74% sequence identity) of TBNAT to the NAT from *M. marinum* (MMNAT) has led to the use of MMNAT as a surrogate model for medicinal chemistry and structure–activity relationship studies (Fullam, 2007; Westwood *et al.*, 2010; Fullam *et al.*, 2011, 2013). The structure of MMNAT has been solved in the apo form as well as with a substrate, a cofactor and an inhibitor (Abuhammad *et al.*, 2012; Fullam *et al.*, 2008).

Despite the uncertainties that are associated with using a surrogate model, recent research has shown that this approach can be used successfully for the identification and optimization of lead compounds including ATPase inhibitors (Andries *et al.*, 2005) and has been used for drug design against other poorly soluble *M. tuberculosis* proteins (Björkelid *et al.*, 2011). The high yield of recombinant MMNAT and the ease with which it crystallizes have been highly advantageous for pursuing this enzyme as a model for drug design. MMNAT crystallizes readily in a range of crystallization conditions and in the presence of different substrates and inhibitors (Abuhammad *et al.*, 2010, 2012; Fullam *et al.*, 2008). The use of MMNAT as a model allowed the successful identification of inhibitors with novel mechanisms of action (Abuhammad *et al.*, 2012).

Despite its high sequence similarity to MMNAT, TBNAT has particular characteristic physical properties and shows a distinctive inhibition profile (Lack *et al.*, 2009; Fullam *et al.*, 2009). Therefore, the determination of the three-dimensional structure of TBNAT is required to pursue ligand optimization. Our efforts to subclone the *tbnat* gene into the pVLT31 vector provided a means of overcoming the low protein yield (Abuhammad *et al.*, 2011).

The TBNAT enzyme proved to be recalcitrant to crystallization despite prolonged and varied attempts. In this study, successful cross-seeding using microcrystals of the similar

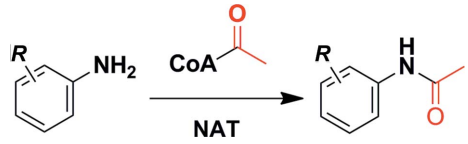


**Figure 1**
The arylamine acetylation reaction catalysed by the NAT enzyme.

MMNAT resulted in a crystal that allowed the determination of the structure of TBNAT. The structure obtained corresponds to the same enzyme in two mycobacterial species, *M. tuberculosis* and *M. bovis* BCG, since the sequences of these two NATs are identical.

## 2. Materials and methods

All chemicals and reagents were purchased from Sigma–Aldrich (Poole, Dorset, England) unless otherwise stated.

### 2.1. Protein expression and purification

The *nat* gene from *M. marinum* (*mmnat*) was expressed in *Escherichia coli* and purified as described previously (Fullam *et al.*, 2008; Abuhammad *et al.*, 2011).

The *nat* gene from *M. tuberculosis* H37Rv (*tbnat*) was expressed in competent *E. coli* NEB Express cells (New England BioLabs) transformed with the pVLT31-*tbnat* construct (Abuhammad *et al.*, 2011).

The recombinant TBNAT was purified by immobilized metal-ion affinity chromatography using cobalt TALON as described previously (Abuhammad *et al.*, 2011). The protein was either concentrated with the intact N-terminal hexahistidine tag (His tag) (partially purified His-TBNAT) or the His tag was cleaved using the CleanCleave kit as described previously (Abuhammad *et al.*, 2011).

A further purification step was employed to purify the His-TBNAT. The eluted fraction of partially purified His-TBNAT in 150 m*M* imidazole buffer was diluted threefold with buffer *A* (20 m*M* phosphate buffer pH 8, 75 m*M* NaCl) to lower the imidazole concentration to 50 m*M*. The enzyme solution was then incubated with Ni–NTA resin that had been equilibrated with buffer *A* containing 50 m*M* imidazole. The column was subsequently washed with buffer *A* containing 60 and 75 m*M* imidazole followed by elution with 150 m*M* imidazole solution in the same buffer. A summary description of the purification protocol is shown in Supplementary Fig. S1.[1] Each of the three batches of the protein (*i.e.* pure TBNAT, partially purified His-TBNAT and pure His-TBNAT) was buffer-exchanged by dialysis against 20 m*M* PIPES pH 7.0 buffer and concentrated separately using 10 kDa MWCO VivaSpin Protein Concentrators (Sartorius Stedim). Aliquots of 5–19 mg ml$^{-1}$ protein in 20 m*M* PIPES pH 7.0 buffer were then stored at 193 K with 1 m*M* dithiothreitol (DTT).

### 2.2. Acetyltransferase activity assay

The activity was assessed from the rate of CoA formation from acetyl-CoA catalysed in the presence of hydralazine (HLZ) as described previously (Brooke *et al.*, 2003). To measure the effect of pH on protein stability, samples of each enzyme (10 µl of 2 mg ml$^{-1}$ MMNAT or TBNAT) were incubated prior to the assay at different pH values for 30 min in an ice bath in a 100 m*M* equimolar mixture of MES, HEPES, Tricine and CHES over the pH range 4–11. Samples of each enzyme–pH combination were diluted 100-fold into the assay buffer (20 m*M* Tris–HCl pH 8.0) and pre-incubated with 500 µ*M* HLZ in the same buffer for 5 min in a 96-well plate. Acetyl-CoA was added to a final concentration of 400 µ*M* to start the reaction in a total volume of 100 µl 20 m*M* Tris–HCl pH 8.0 assay solution to establish the initial linear rate conditions at 297 K. The reaction was quenched with 25 µl 5 m*M* 5,5′-dithiobis-(2-nitrobenzoic acid) (DTNB) solution in 6.4 *M* guanidine–HCl, 0.1 *M* Tris–HCl pH 7.3 buffer after 10 min. The absorbance was measured within 2 min at 405 nm (Sunrise Plate Reader, Tecan). The investigation of the variation of enzyme activity with the pH at which it was incubated was designed to analyse the stability of the enzyme in different storage conditions rather than to measure the optimal pH for enzyme activity.

### 2.3. Size-exclusion chromatography and multi-angle light scattering (SEC-MALS) for determination of the oligomeric state

Size-exclusion experiments were performed on a Superdex 200 10/300 GL column (GE Healthcare) equilibrated with 25 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl at 0.4 ml min$^{-1}$ on an ÄKTA system (Pharmacia). The column was followed in-line using a Dawn HELEOS II light-scattering detector (Wyatt Technologies) and an Optilab rEX refractive-index monitor (Wyatt Technologies). Molecular-mass calculations were performed using *ASTRA* v.5.3.4.14 (Wyatt Technologies) assuming a refractive-index increment (d*n*/d*c* value) of 0.186 ml g$^{-1}$.

### 2.4. Protein stability: thermal shift assay

Thermal shift assays were performed using an Mx3005p QPCR System (Agilent Technologies) as described previously (Niesen *et al.*, 2007). The reactions were performed in a total well volume of 20 µl containing 4× SYPRO Orange dye (5000× stock solution; Invitrogen, catalogue No. S6650). The stability of TBNAT (100 µg ml$^{-1}$) was measured in a range of buffers and salts as described in §3. The 96-well PCR microplate was heated from 297 to 373 K at a heating rate of 0.5 K min$^{-1}$. The fluorescence intensity was measured every 1 K and was plotted as a function of temperature. Data analysis was performed using *GraphPad Prism* (http://www.graphpad.com/prism/).

### 2.5. Crystallization

Protein crystallization was carried out using the sitting-drop vapour-diffusion technique. Sitting drops were set up in 96-well plates containing commercially available sparse-matrix and systematic grid screen conditions.

Initial high-throughput screens to identify crystallization conditions were performed using a Tecan Genesis ProTeam 150 robot (Tecan). Equal volumes (100 nl) of mother liquor and protein solution were combined as sitting drops using a Mosquito crystallization robot (TTP LabTech).

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: WD5214). Services for accessing this material are described at the back of the journal.

High-throughput screens to identify crystallization conditions for TBNAT and His-TBNAT were performed using variable combinations of the available sparse-matrix screens (over 1000 conditions; Supplementary Table S1), protein concentration (5–19 mg ml$^{-1}$) and temperature (277, 293 and 308 K; Supplementary Table S2). Further screens were performed by including different additives in the conditions or by changing the buffer in which the protein was prepared prior to setting up crystallization trials. Following the failure of all of these trials, cross-seeding using microcrystals of MMNAT was attempted.

The MMNAT starter crystals were grown at 293 K using His-MMNAT solution (10 mg ml$^{-1}$ in 20 m*M* Tris–HCl pH 8.0, 1 m*M* DTT) by the sitting-drop vapour-diffusion method. High-quality crystals were obtained in conditions C10 [0.2 *M* $MgCl_2$, 0.1 *M* HEPES pH 7.0, 20%(*w*/*v*) PEG 6000] and C11 [0.2 *M* $CaCl_2$, 0.1 *M* HEPES pH 7.0, 20%(*w*/*v*) PEG 6000] of the PACT premier screen. The seed-stock preparation was expedited as described by Shaw Stewart *et al.* (2011). Briefly, the MMNAT crystals in each crystallization drop (C10 and C11) were thoroughly crushed in the corresponding drop well (not the reservoir well) of the crystallization plate using a glass probe with a 0.25 mm bead melted on the end. For each of the two wells, 2 µl reservoir solution was added and mixed with the entire contents of the drop well by withdrawing and dispensing from the pipette tip several times. The two mixtures were then combined in a single Eppendorf tube containing one Seed Bead (Hampton Research) and 10 µl of each of the reservoir solutions (C10 and C11). The mixture was transferred back and forth between the Eppendorf tube and the two drop wells several times to ensure that all of the crystalline material had been transferred from the drop wells. The tube contents were vortexed for 2 min and then kept on ice; the Seed Bead was removed and the seed stock was used immediately without further dilution.

**2.5.1. Microseeding**. Automated seeding with random screens (D'Arcy *et al.*, 2007) was performed using an Oryx8 crystallization robot (Douglas Instruments Ltd) using the MMNAT seed stock detailed above. Vapour-diffusion sitting-drop experiments with His-TBNAT (10 mg ml$^{-1}$) and the JCSG-*plus* and PACT premier screens (Molecular Dimensions Ltd; Newman *et al.*, 2005) were set up using this seed stock in 96-well two-drop plates (Swissci AG). The first drop in each position consisted of 0.3 µl protein solution, 0.27 µl reservoir solution and 0.03 µl seed stock. The second drop consisted of 0.3 µl protein solution and 0.3 µl reservoir solution only. The plates were sealed with clear tape and incubated at 296 K. Crystals were harvested after 7 d and suitable cryoprotectant agents were added depending on the particular mother liquor (Garman & Mitchell, 1996) for cryocooling into liquid nitrogen. Over 20 crystals were tested for diffraction at 100 K using an in-house Rigaku FR-E+ X-ray generator with Osmic HF optics and a Saturn 944+ CCD detector. Only one crystal gave a diffraction pattern corresponding to protein, and this crystal was subsequently used for structure determination without further optimization of the crystallization conditions.

### 2.6. X-ray diffraction and data processing

Native data sets were collected at Diamond Light Source, Didcot, England at 100 K to 2.1 Å resolution (beamline I04-1; $\Delta\varphi = 120^\circ$) and 2.9 Å resolution (beamline I03; $\Delta\varphi = 180^\circ$) with PILATUS 2M and PILATUS 6M detectors, respectively. Data were initially processed using *xia*2 (Winter, 2010) and the structure was solved by molecular replacement (MR) using a previously determined MMNAT crystal structure in complex with hydralazine (PDB entry 3ltw; 2.1 Å resolution; Abuhammad *et al.*, 2010), stripped of heteroatoms, as the search model in *Phaser* (McCoy *et al.*, 2007). Rigid-body refinement of the MR solution and the remaining cycles of restrained refinement were carried out using *PHENIX* (Adams *et al.*, 2010). The quality of the refinement was evaluated by the inspection of $R_{\rm free}$ (Brünger, 1992).

Model building was performed using *Coot* (Emsley *et al.*, 2010). The stereochemical properties and quality of the final model were assessed with the program *MolProbity* (Chen *et al.*, 2010). Structural figures and graphical renderings were made with either *PyMOL* (v.1.3r1; Schrödinger) or *Discovery Studio Visualizer* (*DS Visualizer*; v.3.5; Accelrys).

## 3. Results and discussion

### 3.1. Protein expression and purification

Previous studies have relied on improved yields of recombinant TBNAT without a His tag (Abuhammad *et al.*, 2011) to produce protein of sufficient quality for structural studies to be attempted. However, crystallization of the purified protein without a His tag proved to be challenging since initial trials using protein without a His tag gave no crystals. The protocol described (Abuhammad *et al.*, 2011) relied on the elution of the untagged protein after removal of the His tag. Therefore, an additional purification step was required in order to produce His-tagged protein of sufficient purity for structural studies. This step depends on the differential affinity of His-TBNAT for cobalt TALON resin and Ni–NTA resin as a means of removing contaminants whilst maintaining the yield (see Supplementary Material and Fig. S1).

The additional Ni–NTA resin step resulted in pure His-TBNAT suitable for protein crystallization trials. Interestingly, SDS–PAGE analysis of samples from the various purification steps showed diffuse low-molecular-weight bands (Supplementary Fig. S2). These bands were probably owing to lipid, since the association of TBNAT with membrane lipids has been observed previously in a whole-cell analysis (Målen *et al.*, 2010). Band smearing owing to the presence of lipids in protein samples has been observed and reported previously (Adamina *et al.*, 2007).

### 3.2. Pre-crystallization characterization of TBNAT and comparison with MMNAT

**3.2.1. Stability of TBNAT**. Protein stability, homogeneity and solubility are key factors that affect protein crystallizability (Ericsson *et al.*, 2006). Therefore, optimization of these properties is important in order to improve the success

rate of crystallization. The stability of TBNAT was studied by measuring its enzyme activity as a function of pH and carrying out thermal shift assays (TSAs) in the presence of various small-molecule ligands (Lavinder *et al.*, 2009). Parallel studies of MMNAT were performed to allow comparison of the two enzymes.

TBNAT was active over a wide range of pH values (pH 6–9), in agreement with previous stability studies on the effect of pH on protein folding (Abuhammad *et al.*, 2011; Fig. 2). The enzyme activity observed at pH 6 is consistent with the proposed role of the protein in intracellular survival and persistence inside macrophages of the dormant bacterium (Bhakta *et al.*, 2004). In the light of evidence that live mycobacteria disturb the acidification defence mechanism and retain the pH within macrophage vacuoles at between 6.1 and 6.5, it is likely that TBNAT remains active following infection.

The TBNAT activity gradually increased as a function of pH, in contrast to MMNAT which was active at a more constant level between pH 5 and 9 (see Fig. 2). Interestingly, *M. marinum* can grow over a wide range of pH values, with an optimum growth rate in acidic conditions (pH 5–6.5; Falkinham, 2002).

Thermal shift assays (TSAs) to determine the protein melting temperatures ($T_m$) were carried out as described previously (Lavinder *et al.*, 2009; Kemp *et al.*, 2012). The result obtained for TBNAT was ∼343 K, which is comparable to the ∼341 K measured previously using circular dichroism (CD; Lack *et al.*, 2009; Supplementary Figure S3). For MMNAT, the melting temperature obtained was ∼320 K, which was again very similar to the value of ∼319 K obtained previously using CD (Fullam, 2007; Supplementary Fig. S3). Evaluation of protein stability by TSAs has been reported to give similar results to those obtained from CD for other proteins where both measurements have been carried out (Vedadi *et al.*, 2006, 2010).

Small ligands have been shown to often improve the stabilization of proteins and hence to enhance crystallization (Yeh, 2009). To investigate this further, TBNAT with the His tag removed was screened against two types of small-molecule library. The first library comprised compounds commonly found to be useful in crystallization trials (provided by the SGC, Oxford and hence referred to as the SGC stability screen; Niesen *et al.*, 2007). The second library contained compounds specific for NAT and will be referred to as NAT ligands. The latter library was generated in our laboratory and comprised substrates, physiologically relevant small molecules (*e.g.* different acyl-CoA cofactors) and other molecules that were either identified as, or predicted to be, ligands of TBNAT (totalling 34 compounds; details of this library can be found in Abuhammad, 2013). Of the 130 compounds tested by TSAs, fewer than 20 were shown to bind and/or stabilize TBNAT. However, crystallization trials with these ligands (Supplementary Fig. S4 and Table S3) were no more successful than those without them.

**3.2.2. Multi-angle light scattering (SEC-MALS) and the influence of the His tag**. The hydrodynamic properties of proteins (a measure of their homogeneity) have an extremely strong influence on the success of crystallization (Price *et al.*, 2009). Hence, the oligomeric states of both TBNAT and MMNAT with and without a His tag were determined by size-exclusion chromatography coupled with MALS.

For TBNAT without a His tag, previous similar measurements had shown microheterogeneity, with a tetramer and hexamer being observed in addition to the main dimer peak (Abuhammad *et al.*, 2011; Fig. 3*a*). The presence of the enzyme in this dynamic oligomeric state is likely to hamper attempts to crystallize it. In comparison, the more easily crystallizable MMNAT exhibited higher homogeneity, with most of the protein being present as a dimer. MMNAT also showed a greater tendency to aggregate, as indicated by the large peak eluting in the void volume (Fig. 3*b*). This aggregation is very likely to be a result of nonspecific hydrophobic interactions. The presence of a His tag on each protein has a significant effect on their chromatographic behaviour. His-TBNAT showed a single peak corresponding to the mass of the dimer, and the MALS measurement showed a stable homogeneous species within the expected mass of the protein (Fig. 3*c*). For His-MMNAT, the presence of the His tag resulted in a significant reduction in the aggregation of the protein (Fig. 3*d*). An improvement of the tendency of MMNAT to crystallize has been observed when the His tag is retained on the protein (Abuhammad, 2013). These observations strongly suggest that the presence of the His tag on either enzyme would improve the likelihood of crystallization by preventing aggregation and improving microhomogeneity.

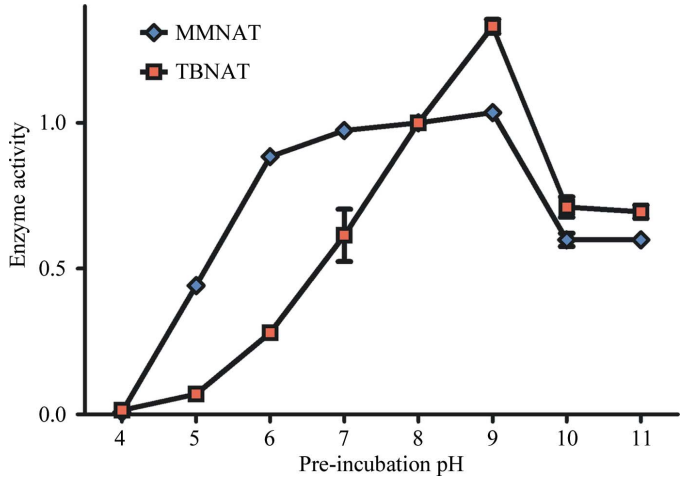


**Figure 2**
Effect of pH on the stability of both TBNAT and MMNAT as measured using an enzymic assay. The enzymes were incubated prior to the assay at different pH values for 30 min in an ice bath. The intended pH levels were obtained by adjusting the pH of a combined buffer [100 m*M* equimolar mixture of MES (p$K_a$ 6.16), HEPES (p$K_a$ 7.55), Tricine (p$K_a$ 8.16) and CHES (p$K_a$ 9.55)] by one pH unit over the pH range 4–11 (note: pH values of 4 and 11 are not within the optimal buffer range). The activity of each enzyme was then measured at 297 K and pH 8 by the NAT activity assay in the presence of 500 µ*M* HLZ. The reaction was performed in the linear range of the enzyme activity. The enzyme activities of each NAT are presented relative to the activity of that NAT at pH 8. Each point represents the mean ± standard deviation from triplicate determinations. The error bars are within the symbols.

### 3.3. Crystallization of TBNAT

Despite the thousands of different conditions screened and the methods employed, the TBNAT enzyme failed to crystallize. A wide variety of commercial sparse-matrix screens and the use of different temperatures, protein concentrations and buffers (pH values) were employed in trials using protein with and without a His tag, as well as in the presence or absence of ligands and small molecules. Other methods for crystallization such as hanging-drop vapour diffusion, microbatch crystallization and heteroseeding using human hair were also utilized but were unsuccessful in inducing crystallization.

Since no hits were observed throughout all of the trials, a different strategy was pursued: that of cross-seeding. Since MMNAT crystallized readily, MMNAT crystals were used as seeds in a TBNAT crystallization trial. A single crystal of TBNAT was successfully grown using a concentrated MMNAT seed stock prepared as described in §2. A cubic TBNAT crystal with sides of ∼20 µm grew in condition A4 [0.02 *M* $CaCl_2$, 0.1 *M* sodium acetate pH 4.6, 30%(*v*/*v*) MPD] from the JCSG-*plus* screen in a week (Supplementary Fig. S5). Since this condition included 30%(*v*/*v*) MPD, no additional cryoprotectant agent was necessary prior to cryocooling.

### 3.4. Diffraction data analysis and structure solution

Data were initially processed in space group $C222_1$, with unit-cell parameters $a$ = 135.1, $b$ = 137.9, $c$ = 139.1 Å, which differed from the MMNAT seed crystals ($P4_12_12$). The overall temperature factor of the diffraction data was 26.2 $Å^2$ according to the Wilson plot. The structure was solved by molecular replacement (MR) with four molecules in the asymmetric unit and 52.7% solvent content. Although the structure could be modelled well in the electron density, the refinement statistics remained poor ($R_{cryst}$ and $R_{free}$ of 0.26 and 0.30, respectively) and could not be further improved by rebuilding. This led to a re-examination of the original data processing and structure determination. Molecular replacement was carried out in space group $P1$ with a 2.9 Å resolution 180° data set and 16 molecules in the unit cell. From this, the space group was then found to be $P2_1$ (unit-cell parameters $a$ = 96.50, $b$ = 139.20, $c$ = 96.51 Å, $\beta$ = 91.2°) with a pseudo-merohedral twinning fraction of 0.43 (twin law $l$, $-k$, $h$). A comparison of the unit-cell parameters of the TBNAT crystal with those of the MMNAT seed crystal is shown in Table 1. In $P2_1$ there are eight molecules of TBNAT in the asymmetric unit related by NCS. These could be considered as four dimers which had presented as two crystallographically equivalent dimers when processed in $C222_1$. When treated correctly in

**Table 1**
A comparison of the unit-cell parameters of the TBNAT crystal and the MMNAT seed crystal.

The crystals grew in different space groups.

| | MMNAT ($P4_12_12$) | TBNAT ($P2_1$) |
|---|---|---|
| $a$ (Å) | 51.94 | 96.50 |
| $b$ (Å) | 51.94 | 139.20 |
| $c$ (Å) | 176.65 | 96.51 |
| $\alpha$ (°) | 90.00 | 90.00 |
| $\beta$ (°) | 90.00 | 91.18 |
| $\gamma$ (°) | 90.00 | 90.00 |

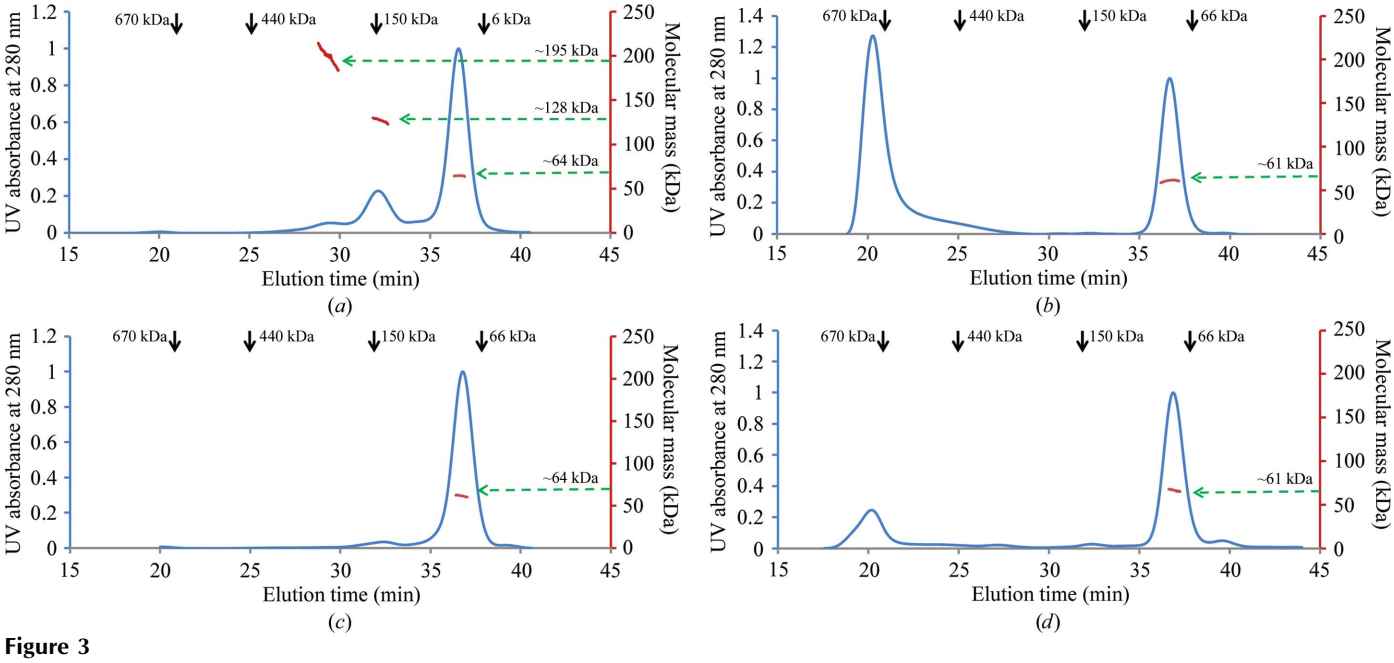


**Figure 3**
Representative multi-angle light-scattering analyses of TBNAT, MMNAT, His-TBNAT and His-MMNAT. Analyses of (*a*) TBNAT, (*b*) MMNAT, (*c*) His-TBNAT and (*d*) His-MMNAT by size-exclusion chromatography followed by in-line multi-angle light scattering. The right axis (red) refers to the MALS and represents the molecular mass at any given point in the elution profile. The segments of the curve within each peak are highlighted and labelled with the average measured molecular mass. The elution profile of TBNAT is shown and demonstrates an abundant apparent dimer (∼64 kDa) and lower quantities of apparent tetramer (∼128 kDa) and hexamer (∼196 kDa). The column volume is 23 ml (57 min), the void volume is 7.5 ml (18.75 min) and the peaks for protein standards for SEC are indicated by the black arrows at the top of the graph.

**Table 2**
Data-collection, processing and refinement statistics for the TBNAT structure determination (PDB entry 4bgf).

The TBNAT crystal was obtained by cross-seeding using MMNAT crystals as the source of the microseeds. Values in parentheses are for the highest resolution shell.

| | |
|---|---|
| Data collection | |
| Beamline | I04-1, DLS |
| Wavelength (Å) | 0.92 |
| Temperature (K) | 100 |
| Space group | $P2_1$ |
| $\varphi$ range (°) | 120† |
| $\Delta\varphi$ (°) | 0.2 |
| Time per exposure (s) | 0.3 |
| Transmission (%) | 100 |
| Crystal-to-detector distance (mm) | 251.9 |
| Data processing | |
| Unit-cell parameters (Å, °) | $a$ = 96.50, $b$ = 139.20, $c$ = 96.51, $\alpha = \gamma = 90$, $\beta = 91.18$ |
| Resolution range (Å) | 29.6–2.1 (2.2–2.1) |
| Measured reflections | 346636 (46791) |
| Unique reflections | 136366 (19483) |
| Multiplicity | 2.5 (2.4) |
| Mosaicity (°) | 0.29 |
| Overall $\langle I/\sigma(I)\rangle$ | 8.5 (1.8) |
| Overall completeness (%) | 92.1 (90.2) |
| $R_{\rm merge}$ | 0.08 (0.53) |
| $R_{\rm meas}$ | 0.10 (0.66) |
| $R_{\rm p.i.m.}$ | 0.059 (0.39) |
| Solvent content (%) | 52.7 |
| Mean $B$ factor (Å$^2$) | 39.9 |
| Refinement | |
| $R_{\rm cryst}$ | 0.16 |
| $R_{\rm free}$ | 0.19 |
| Refined residues | 2160 |
| Twin law | $l$, $-k$, $h$ |
| Twin fraction | 0.43 |
| Refined water molecules | 1156 |
| Average $B$ value for main chain (Å$^2$) | |
| Chain *A* | 36.5 |
| Chain *B* | 42.0 |
| Chain *C* | 35.7 |
| Chain *D* | 32.2 |
| Chain *E* | 36.8 |
| Chain *F* | 36.3 |
| Chain *G* | 31.9 |
| Chain *H* | 36.9 |
| R.m.s.d. for bond lengths (Å) | 0.014 |
| R.m.s.d. for bond angles (°) | 1.5 |
| Ramachandran plot | |
| Most favoured regions (%) | 95.7 |
| Allowed regions (%) | 4.16 |
| Disallowed region (%) | 0.14 |
| Rotamer outliers (%) | 1.86 |
| *MolProbity* clashscore | 6.5 |

† See further details in §§2.6 and 3.4.

$P2_1$ we observe two dimers which superpose exactly with the molecules which were present in the $C222_1$ crystal and two dimers which occupy roughly the same position in the lattice but are rotated by approximately 1°, thus explaining the poor refinement statistics when the data were assigned space group $C222_1$. Following refinement of the eight molecules using twin refinement in *PHENIX* (Adams *et al.*, 2010) and rounds of rebuilding in *Coot*, the final $R_{\rm cryst}$ and $R_{\rm free}$ were 0.16 and 0.19, respectively, with a concomitant improvement in the geometry statistics and the electron density (see Table 2). The eight molecules in the asymmetric unit are unlikely to reflect the tetrameric oligomerization state observed in solution. A few residues in each of the eight monomers of the TBNAT model (residue 3 in monomer *C*; residue 161 in monomer *F*) were omitted owing to weak electron density. The final model contained 2160 amino-acid residues and 1156 water molecules and was of good quality as validated by *MolProbity*. However, there remained a few rotamers (1.86%; 32 *R* groups) that were not in the optimum conformation. These were checked and found to be well supported by the electron density. In particular, the active-site residue Tyr69 was observed to be in an unusual conformation, as is also observed in the available MMNAT structures (PDB entries 2vfb, 2vfc, Fullam *et al.*, 2008; and 3ltw, Abuhammad *et al.*, 2010).

### 3.5. Structure of TBNAT

The overall structure and topology of the 2.1 Å resolution structure reveals the unique NAT fold and comprises three distinct domains of approximately equal size (Figs. 4*a* and 4*c*): domain I forms an $\alpha$-helical bundle, domain II forms a $\beta$-barrel and domain III forms an $\alpha/\beta$-lid over the active site of the protein. The catalytic triad of the active site is formed by residues Cys70, His110 and Asp127 (Fig. 4*d*).

A comparison between the C$^\alpha$ backbones of native TBNAT and MMNAT (Figs. 4*a* and 4*b*) shows that these backbones are almost superimposable and have an r.m.s.d. of only 0.65 Å over 252 equivalent amino acids. As for MMNAT, no density was observed in TBNAT for the N-terminus (residues 1–2, the His tag and four vector-encoded residues) and eight residues at the C-terminus (275–283).

Weak electron density was observed for residues 100–103 within the flexible loop formed by residues 96–107. In apo MMNAT (PDB entry 2vfb), these latter residues are also only supported by weak electron density (Fullam *et al.*, 2008). It has been speculated that this loop might undergo conformational changes in the NAT catalytic cycle owing to its generally observed flexibility in the NAT structures (Sandy *et al.*, 2002; Fullam *et al.*, 2008).

The sequence within this flexible loop (residues 96–107) comprises four proline residues with two being consecutive in each sequence, albeit at different positions (Supplementary Fig. S6). In addition to the role of proline as a secondary-structure modulator (*cis–trans* isomerization), the puckering conformation of the pyrrolidine ring plays an important role in protein folding and stabilization (Crespo & Rubini, 2011).

Analysis of the surface properties showed the spread of several areas of hydrophobic potential (hydrophobic patches) on the surface of TBNAT away from the active site, whereas for MMNAT the hydrophobic areas are more concentrated at areas surrounding and in the active site (Fig. 5). These dissimilarities in the surfaces properties of the two proteins could explain the varying physical characteristic and catalytic behaviour of TBNAT compared with the very similar MMNAT. The difference in protein surface residues between MMNAT and TBNAT has previously been reported to explain the differing thermal stabilities of the two enzymes (Lack *et al.*, 2009). A larger fraction of hydrophobic intermolecular

contacts is usually observed in proteins with elevated melting temperature (Murphy, 2001).

The difficulties associated with protein expression and purification (Abuhammad *et al.*, 2011) are very likely to be a consequence of this wide spread of hydrophobic patches in the TBNAT molecule. The fact that TBNAT was only attainable when co-expressed with chaperonins is a direct consequence of such surface properties (Fullam *et al.*, 2009; Sikora *et al.*, 2008; Abuhammad *et al.*, 2011). The hydrophobic regions of nascent protein chains bind the hydrophobic walls of the chaperone complex, which provides a stabilizing environment in which the protein can fold.

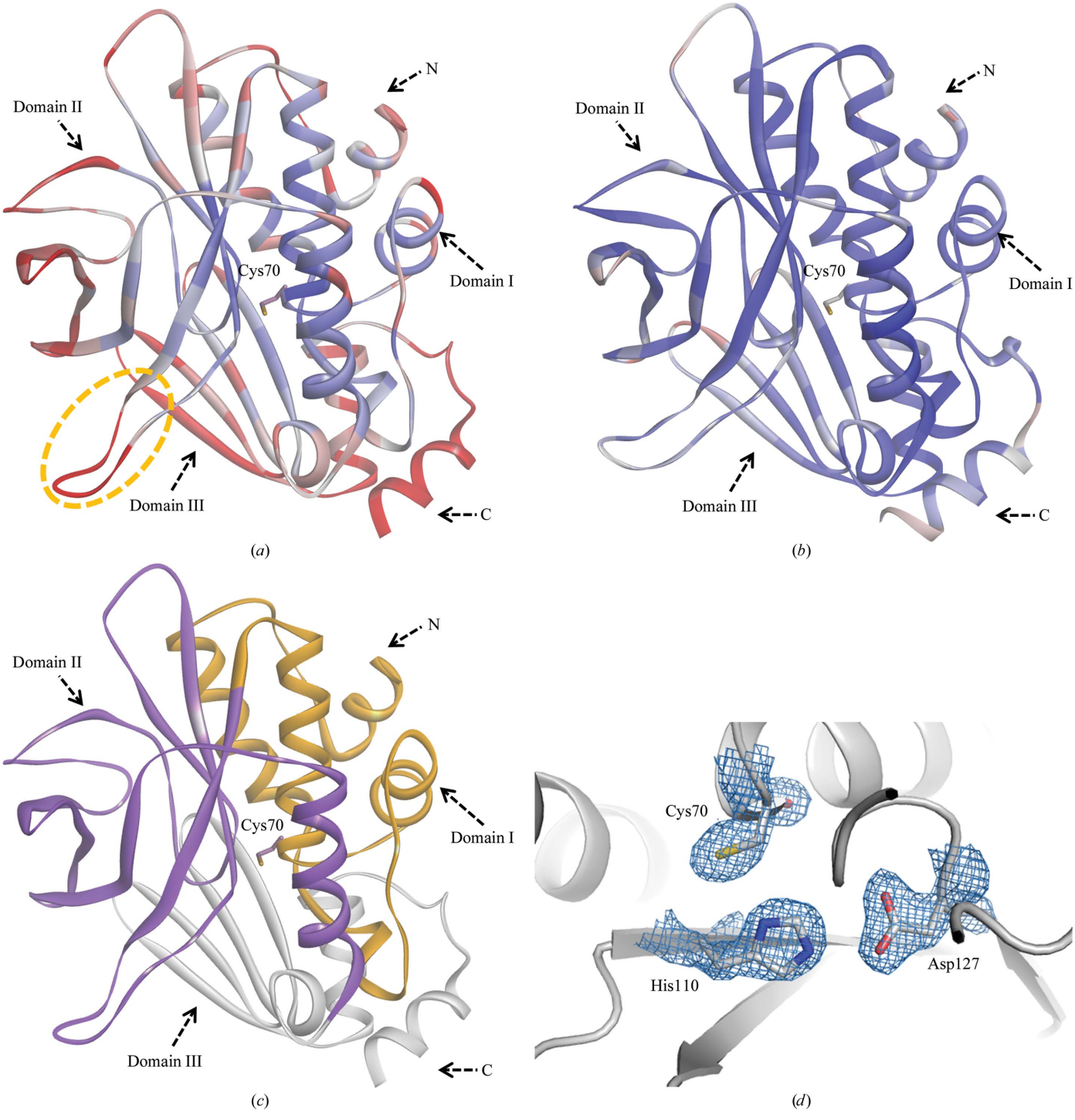


**Figure 4**
Comparison of the TBNAT and MMNAT structures. The ribbon representation shows the $C^{\alpha}$ trace of (*a*) TBNAT (chain *H*) and (*b*) MMNAT (chain *A*; PDB entry 2vfb; 2.1 Å resolution). The structures are coloured according to temperature factor (from blue to red with increasing temperature-factor value from 15 to 50 $Å^2$). The active-site cysteine is shown. The flexible loop in the TBNAT structure is circled in yellow. (*c*) TBNAT (chain *H*) colour-coded to show the three domains. (*d*) The catalytic triad within the TBNAT active site is shown with the electron-density contour level at 1.5$\sigma$.

Surface hydrophobicity is also crucial in protein function. It has been shown that hydrophobic patches play an important role in intramolecular and intermolecular recognition processes such as ligand binding (Li *et al.*, 2013), protein folding (Dill, 1990) and protein–protein association (Cheng *et al.*, 2004; Porter *et al.*, 2003; Cutruzzolà *et al.*, 2002). The association of TBNAT with membrane lipids (Målen *et al.*, 2010) highlights a possible role of surface hydrophobicity in the function of TBNAT, which is related to cell-wall lipid biosynthesis as well as cholesterol catabolism (Bhakta *et al.*, 2004; Anderton *et al.*, 2006; Lack *et al.*, 2009).

### 3.6. Active site

The TBNAT structure shows an almost identical configuration of the catalytic triad (Cys70, His110 and Asp127) to that of MMNAT. The orientation of the thiol group of Cys70 towards His110 is consistent with the previous observation of the existence of the cysteine-triad motif as a zwitterion (Asp$\cdots$His$^{+}\cdots$Cys$^{-}$; Fig. 4*d*), with the thiolate anion of the cysteine functioning as a nucleophile during the initial stages of catalysis (Sim *et al.*, 2012; Sinclair *et al.*, 2000).

The TBNAT structure has revealed some distinctive features of the binding pocket of this enzyme which are likely to affect drug-discovery efforts. The NAT active site has previously been studied using MMNAT ligand binding with arylamine substrates (Abuhammad *et al.*, 2010) and CoA (Fullam *et al.*, 2008) as well as the piperidinol NAT inhibitors (Abuhammad *et al.*, 2012). Despite the high similarity between MMNAT and TBNAT, enzymic studies have shown that MMNAT is more active than TBNAT. Subtle differences in the

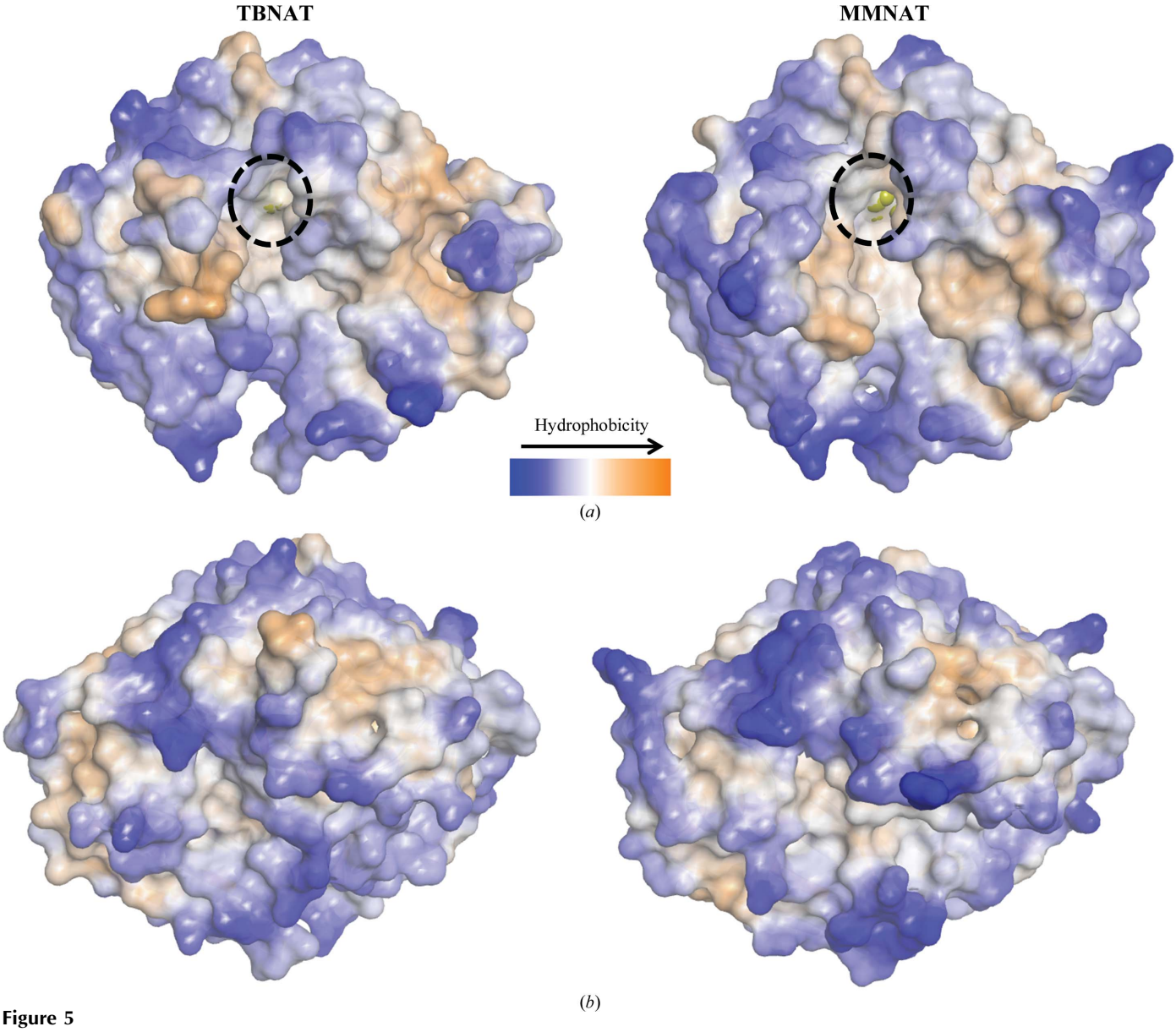


**Figure 5**
Hydrophobic potential surfaces of TBNAT and MMNAT. Comparison of the hydrophobic potential surfaces of TBNAT and MMNAT (PDB entry 2vfb): (*a*) is rotated +180° about the *y* axis compared with (*b*). The surface is coloured by hydrophobic potential (from blue to white to orange with increasing hydrophobicity as calculated by *DS Visualizer* using Kyte–Doolittle hydrophobicity values; Accelrys). The active-site Cys70 residues are circled in black.

selectivity of TBNAT for arylamine substrates and inhibitors have also been observed. Although MMNAT has successfully been used to identify and optimize inhibitors, some classes of NAT inhibitors exhibited selective activity against MMNAT whilst showing weak or no activity against TBNAT (Fullam *et al.*, 2011, 2013; Abuhammad *et al.*, 2012; Westwood *et al.*, 2010). In order to facilitate comparison between the two binding pockets of the two enzymes, the NAT pocket was divided into three different regions: the arylamine-binding pocket, the P-loop region and the adenosine diphosphate (ADP) binding pocket, in which the ADP part of CoA binds (Fig. 6).

The arylamine pocket is almost the same in residue identity and conformation in MMNAT and TBNAT (Fig. 6*a*), except for Tyr71, which is a Phe in TBNAT. This part of the active site accommodates the arylamine substrates and the acyl mercaptoethanolamine fragment of CoA cofactors as well as NAT inhibitors. The high similarity within this region of binding explains the comparable arylamine substrate profiles as well as the preference of both enzymes for acetyl-CoA and propionyl-CoA as acyl donors (Lack *et al.*, 2009; Fullam *et al.*, 2009). Site-directed mutagenesis studies showed a moderate contribution of Tyr71 of MMNAT to the substrate specificity (Fullam *et al.*, 2009). It has been shown that the fundamental cation–$\pi$ binding abilities of phenylalanine and tyrosine are similar. However, it is possible that the hydroxyl group of tyrosine can form hydrogen-bond interactions with the ligand to potentiate its cation–$\pi$ bonding ability (Dougherty, 2007). Hence, this amino-acid residue can contribute to the finely tuned selection of the arylamine substrate.

The second region within the NAT pocket is the P-loop region, which recognizes the pantothenic acid fragment of CoA and has been predicted to be complementary to the different classes of NAT inhibitors that have been identified to date (Fullam *et al.*, 2011, 2013; Abuhammad *et al.*, 2012; Westwood *et al.*, 2010). Binding of inhibitors to this pocket would interfere with Ac-CoA binding.

This region represents the major difference between the MMNAT and TBNAT pockets. Six out of the nine amino-acid residues that constitute this pocket and are involved in CoA binding are different between MMNAT and TBNAT (Fig. 6*b*). This site seems to converge in many NATs, and variation of the amino acids in this region can also modulate arylamine recognition and account for the observed variation in the arylamine substrate profile (Westwood *et al.*, 2006). Moreover, this pocket appears to be more hydrophilic in TBNAT than in MMNAT and provides sites for the formation of ionic and hydrogen-bond interactions in TBNAT (Fig. 6*b*). Substitution of the two methionine residues within this pocket by serine and threonine also affects its volume. Although the sulfur of methionine is known for its capability to form a weak hydrogen bond, recent studies have shown that the strength of the S—N bond can rank this interaction among strong hydrogen bonds in peptidases together with the classical N—H or O=C bonds of the secondary structures (Biswal *et al.*, 2012). These differences are likely to alter the affinity of TBNAT for the various NAT inhibitors. It could also affect the affinity for both Ac-CoA, the reaction cofactor, as well as the byproduct of the reaction, CoA. If so, this could explain the difference in the rate of the acylation reactions that take place with the different arylamine substrates. This region appears to be narrower in TBNAT, which might increase the affinity of the pocket for the CoA part of the cofactor. An increased retention of CoA could also explain the previously observed higher stability of the acetylated complex of TBNAT, which results in reduced substrate turnover. According to the mechanism of enzyme action, CoA seems to be the last reaction product to depart the binding pocket; thus, its effect

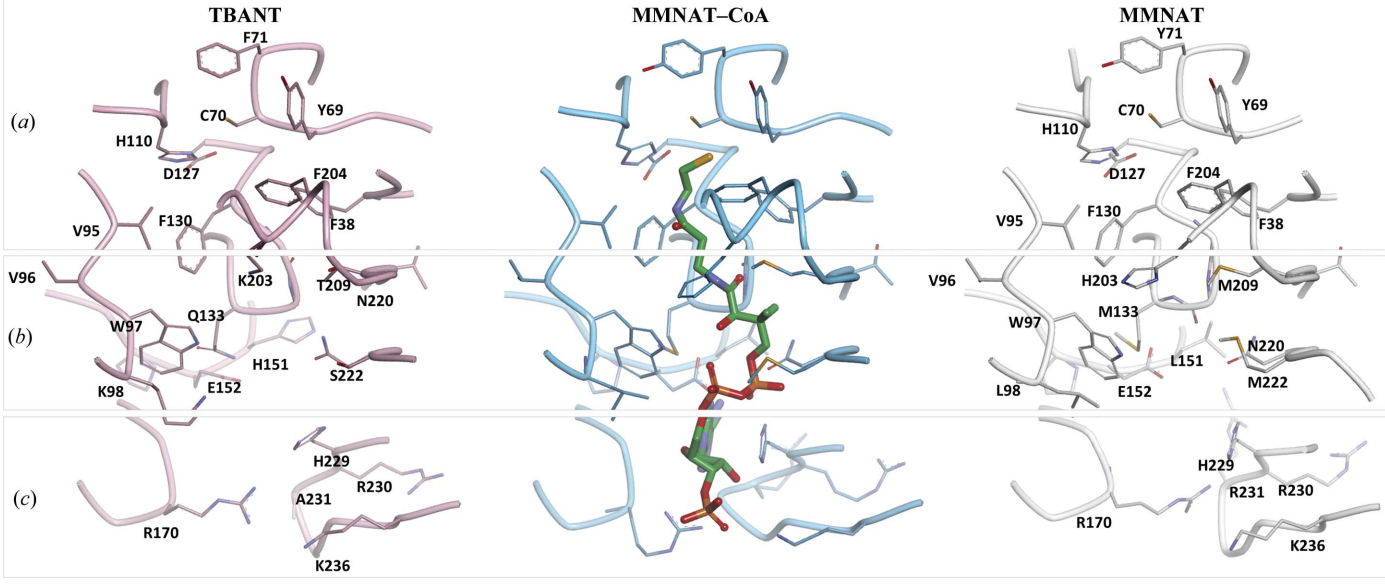


**Figure 6**
The CoA-binding pocket within TBNAT (pink), the MMNAT–CoA complex (PDB entry 2vfc; blue) and MMNAT (PDB entry 2vfb; grey). The main residues constituting (*a*) the arylamine-binding pocket, (*b*) the P-loop region and (*c*) the binding pocket for the adenosine diphosphate of CoA are shown. The figure was prepared using *DS Visualizer* v.3.5.

on the rate at which substrate binds or the product leaves could explain the lower acetylation activity of TBNAT when compared with MMNAT (Fullam *et al.*, 2009).

The adenine moiety of CoA is recognized by a mixture of hydrophobic and polar contacts with His229. The third region (where the 3′-phospho-adenosine diphosphate binds) is again similar in the two proteins, except for Arg231, which is an Ala in TBNAT and might be expected to contribute to the different orientation of the loop in this region (Fig. 6*c*).

Taking the various features of the three pockets described above, it can be seen that the MMNAT and TBNAT structures are remarkably similar, but that the differences revealed by the high-resolution X-ray structure of TBNAT could influence medicinal chemistry efforts.

### 3.7. Protein–protein crystal contacts in NATs

It has previously been observed that many prokaryotic NATs exist as dimers in gel-filtration analysis (Abuhammad *et al.*, 2011). Thus, in order both to examine the oligomerization state of TBNAT and to assess the contribution of common interfaces to the success of the cross-seeding method in inducing crystallization of TBNAT, a protein–protein contact analysis was carried out on the three-dimensional structures of MMNAT (PDB entry 2vfb) and TBNAT (the current structure; PDB entry 4bgf). This analysis was performed with the *PISA* web server (http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver) and enabled the identification of seven protein–protein interface regions for MNNAT and 21 for TBNAT. The interfaces that promoted the packing of MMNAT by utilizing several intermolecular interacting residues were also observed in TBNAT (Fig. 7). However, interfaces I and III show the lowest sequence similarity relative to those of the other interfaces between MMNAT and TBNAT (58 and ∼60%, respectively; Fig. 7). Intermolecular interactions within these interfaces are well maintained by hydrogen bonds and salt bridges (ten interactions) in both TBNAT and MMNAT.

Studies of three-dimensional protein structures have revealed that interfacial hydrogen bonds are less optimal than intrachain bonds, with their geometries demonstrating wider distributions (Xu *et al.*, 1997). Despite this, interfacial hydrogen bonds and salt bridges play an important role in protein–protein interactions (Xu *et al.*, 1997). When buried in the protein interface, an unfulfilled hydrogen-bond donor/acceptor, or an isolated charge which does not form a salt bridge, could substantially destabilize binding owing to the desolvation effect. Therefore, substitution of residues in these regions may well have a substantial impact on the forces that dominate potential crystal contacts.

The involvement of the relatively similar interfaces is to be expected, given the high similarity between the two enzymes and the limited degree of translational and rotational freedom (only six degrees of freedom) available to the chains of the fully folded proteins to achieve their most favourable packed configuration in a crystal. In most cases, only rather minor adjustments are generally observed to optimize the binding and most of these involve variation of the rotamer conformer (Norel *et al.*, 1994; Xu *et al.*, 1997).

Interestingly, TBNAT utilizes the same interfaces for interactions (data not shown). However, a different arrangement of contacts was observed owing to the substitution of some residues. In particular, interface I provides a rigid network of noncovalent interactions (hydrogen bonds and ionic interactions) that is mediated by $\alpha$6 from each molecule,

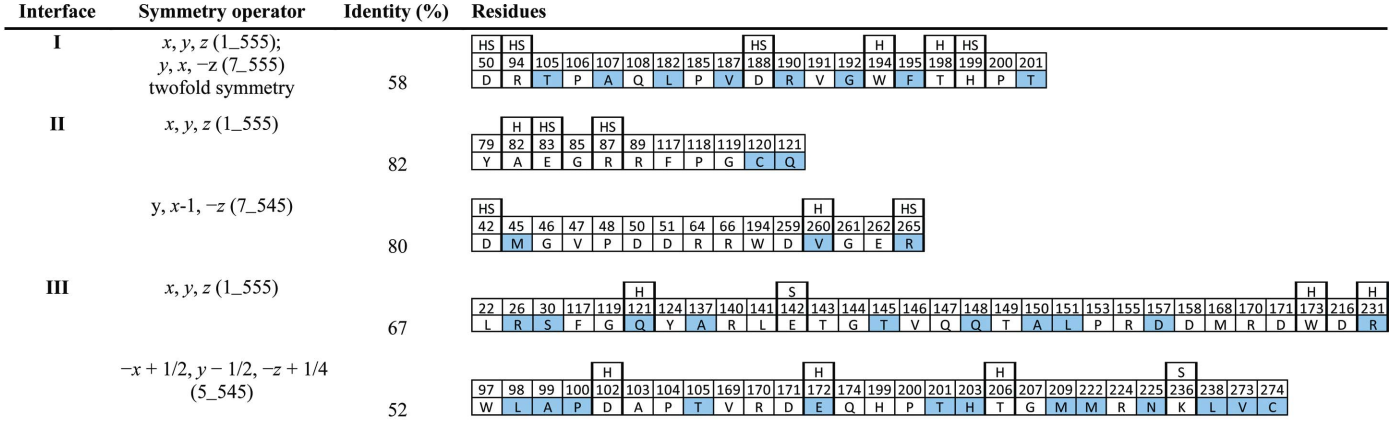


**Figure 7**
The common intermolecular interfaces between the TBNAT (PDB entry 4bgf) and MMNAT (PDB entry 2vfb) apoprotein crystal structures as calculated using the *PISA* web server (http://www.ebi.ac.uk/msd-srv/prot_int/). The symmetry operator indicates the index of crystallographic symmetry-related molecules in the MMNAT structure participating in protein–protein interface contacts. The number in parentheses shows the symmetry operator index with the translation vector in fractional units. Interactions participating in the interface region are marked H for hydrogen bond and/or S for salt bridge. In the software tool *PISA*, interfacing atoms (atoms within interfacing residues) are those atoms which are exposed to another molecule and not to the solvent and they may be involved in bonding across the interface. Not all of the atoms in an interfacing residue will be involved in interface formation or contact. Some may be exposed to the solvent and some may be inaccessible. The *PISA* bonding criteria are that a hydrogen bond is present if the distance between the donor and acceptor atoms is less than 3.89 Å and a salt bridge if the interacting atoms are within 4 Å. The listed residues are these that participate in the interface region in the MMNAT sequence. Residues that are different from those of the TBNAT sequence are highlighted in blue.

resulting in twofold symmetry between them (Figs. 8*a* and 8*b*). This interface is preserved in the MMNAT apo structure (PDB entry 2vfb) and in the MMNAT–CoA complex structure (PDB entry 2vfc; Fullam *et al.*, 2008), and remarkably also in the structures of NATs from *M. smegmatis* (PDB entry 1gx3; Sandy *et al.*, 2002), *Salmonella typhimurium* (PDB entry 1e2t; Sinclair *et al.*, 2000), *Nocardia farcinica* (PDB entry 3d9w; Martins *et al.*, 2008), *Mesorhizobium loti* (PDB entry 2bsz; Holton *et al.*, 2005) and *Pseudomonas aeruginosa* (PDB entry 1w4t; Westwood, 2005). This observation suggests that interface I may in fact be the as yet uncharacterized biologically significant dimerization interface, owing to its structural conservation yet high degree of sequence variation (56% identity) between MMNAT and TBNAT at this interface (Fig. 7).

Furthermore, the dimeric species of the two structures superimpose perfectly (Fig. 8*c*). Gel-filtration analysis of the purified recombinant TBNAT and MMNAT in this study and of NATs from other prokaryotic species in previous studies showed that active dimers are formed upon storage (Sandy *et al.*, 2002; Sinclair *et al.*, 2000; Fullam *et al.*, 2008). However, this conserved dimer interface is not present in human NATs (NAT1, PDB entry 2pqt; NAT2, PDB entry 2pfr; Wu *et al.*, 2007) or in the NAT from *Bacillus anthracis* (PDB entry 3lnb; Pluvinage *et al.*, 2011). Mammalian NATs do not appear to form dimers in solution as determined by ultracentrifugation and gel-filtration studies (Watson *et al.*, 1990). The eukaryotic enzymes are characterized by the presence of a unique insertion loop between the second and third domains and showed a different mode of CoA binding (Wu *et al.*, 2007;

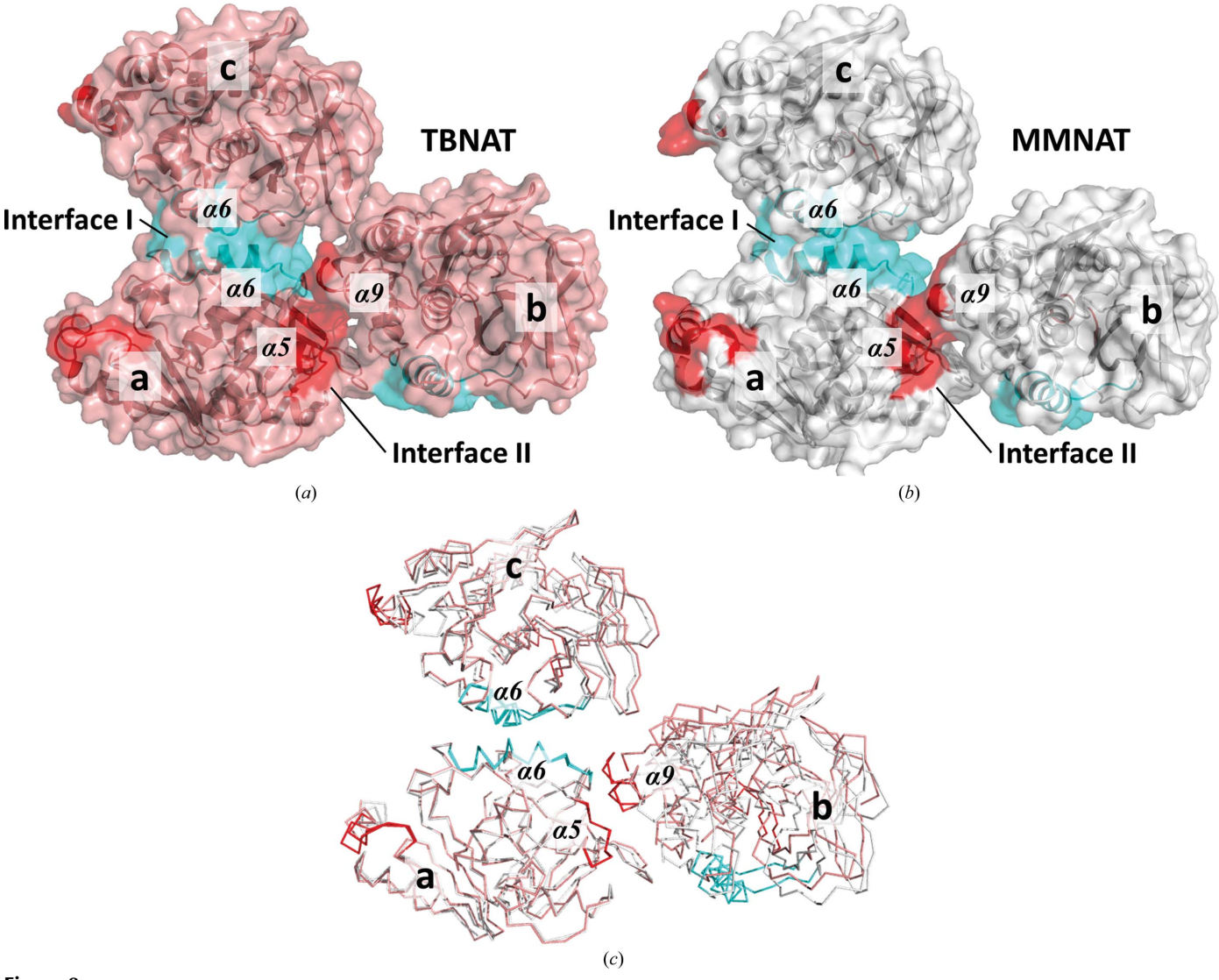


**Figure 8**
Preserved packing interactions across mycobacterial NATs. The blue surface represents a conserved proposed dimerization interface (I: molecules **a** and **c**) observed in all mycobacterial NAT structures determined thus far, as well as in those from *S. typhimurium*, *N. farcinica* and *P. aeruginosa*, whilst the red surface highlights another intermolecular interface (II) that is conserved only within known mycobacterial structures. (*a*) TBNAT. (*b*) MMNAT. (*c*) $C^{\alpha}$ traces of the superimposed TBNAT (pink) and MMNAT (grey) structures demonstrate substantial overlap in crystal packing involving the preserved interfaces despite the fact that they crystallize in different space groups.

Payton *et al.*, 2001). Interestingly, a similar insertion was observed in the structure of *Bacillus anthracis* NAT, which also showed marked differences in the mode of binding and the location of CoA when compared with other NATs (Pluvinage *et al.*, 2011).

However, further mutagenesis studies are needed to confirm the biological significance of this interface, although studies have been carried out on deletion of the 'eukaryotic loop' (Kawamura, 2005; Walraven *et al.*, 2007).

In addition to the preservation of interface I across the NAT structures, another smaller interface (interface II) appears to be shared between homologues. It involves $\alpha$5 from one molecule and $\alpha$9 from another molecule (Fig. 8).

Almost 50% of the residues comprising this interface lie within the third domain of the enzyme, which is the domain that shows the most variability between MMNAT and TBNAT (67% sequence identity; Fullam *et al.*, 2009).

## 4. Conclusions

TBNAT has been identified as a drug target for latent TB therapy development. The enzyme has been recalcitrant in crystallization trials. In this work, a cross-microseeding step using the more crystallizable MMNAT resulted in successful structure determination of TBNAT. Two conserved interfaces (interfaces I and II) have been observed in the crystal packing of NATs. The active oligomeric states of NATs in solution correlate with the presence of interface I. However, further investigation of interface I is required to confirm the classification into functionally relevant dimeric and monomeric NATs.

The fact that the two interfaces are conserved across most mycobacterial NATs may be an indication of why microseeding was successful in this case (74% sequence identity). It may be that a combination of conserved interfaces and sequence identity is necessary for success by this method, and we are currently investigating this question further. The overall structure is very similar to that of MMNAT, which supports the validity of MMNAT as a model for TBNAT, but it also elucidated subtle differences between the binding pockets of the two enzymes which are important for drug discovery against this target. With the determination of the structure of TBNAT, three-dimensional structures now exist for all of the proteins on this operon (HsaA, HsaB, HsaC and HsaD; Lack *et al.*, 2008, 2010; Yam *et al.*, 2009; Dresen *et al.*, 2010). The three-dimensional structure of TBNAT thus has potential to contribute to the understanding of the role of this operon.

We are extremely grateful to Lindsay Eltis (University of British Columbia, Canada) who generously donated the pVLT31 vector. We also thank Oleg Fedorov (Structural Genomic Consortium, Oxford) for help with the TSA analysis, Steven Johnson (Sir William Dunn School of Pathology, Oxford) for help with the MALS analysis, Luc Henry for providing the nonhydrolysable CoA derivatives, Simon Newstead for help in data collection and Federico Forneris and Paul Emsley for helpful discussions. Crystallographic discussions with Edward Snell and subsequent assistance in identifying the correct space group were especially appreciated. We thank Diamond Light Source for beamtime under the BAG scheme. We are also grateful to the University of Jordan for a studentship (AA). ES is a member of the MRC UK consortium TBDUK.

## References

Abuhammad, A. (2013). DPhil thesis. University of Oxford, England.
Abuhammad, A., Fullam, E., Lowe, E. D., Staunton, D., Kawamura, A., Westwood, I. M., Bhakta, S., Garner, A. C., Wilson, D. L., Seden, P. T., Davies, S. G., Russell, A. J., Garman, E. F. & Sim, E. (2012). *PLoS One*, **7**, e52790.
Abuhammad, A., Lack, N., Schweichler, J., Staunton, D., Sim, R. B. & Sim, E. (2011). *Protein Expr. Purif.* **80**, 246–252.
Abuhammad, A. M., Lowe, E. D., Fullam, E., Noble, M., Garman, E. F. & Sim, E. (2010). *Protein Cell*, **1**, 384–392.
Adamina, M., Schumacher, R. & Spagnoli, G. C. (2007). *Interactions of Liposomes with the Biological Milieu.* New York, London: Informa Healthcare.
Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Ahmad, Z., Sharma, S. & Khuller, G. K. (2006). *FEMS Microbiol. Lett.* **258**, 200–203.
Anderton, M. C., Bhakta, S., Besra, G. S., Jeavons, P., Eltis, L. D. & Sim, E. (2006). *Mol. Microbiol.* **59**, 181–192.
Andries, K. *et al.* (2005). *Science*, **307**, 223–227.
Arcus, V. L., Lott, J. S., Johnston, J. M. & Baker, E. N. (2006). *Drug Discov. Today*, **11**, 28–34.
Batt, S. M., Jabeen, T., Bhowruth, V., Quill, L., Lund, P. A., Eggeling, L., Alderwick, L. J., Fütterer, K. & Besra, G. S. (2012). *Proc. Natl Acad. Sci. USA*. **109**, 11354–11359.
Bhakta, S., Besra, G. S., Upton, A. M., Parish, T., Sholto-Douglas-Vernon, C., Gibson, K. J., Knutton, S., Gordon, S., DaSilva, R. P., Anderton, M. C. & Sim, E. (2004). *J. Exp. Med.* **199**, 1191–1199.
Biswal, H. S., Gloaguen, E., Loquais, Y., Tardivel, B. & Mons, M. (2012). *J. Phys. Chem. Lett.* **3**, 755–759.
Björkelid, C., Bergfors, T., Henriksson, L. M., Stern, A. L., Unge, T., Mowbray, S. L. & Jones, T. A. (2011). *Acta Cryst.* D**67**, 403–414.
Brooke, E. W., Davies, S. G., Mulvaney, A. W., Pompeo, F., Sim, E. & Vickers, R. J. (2003). *Bioorg. Med. Chem.* **11**, 1227–1234.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Cheng, Z., Liu, Y., Wang, C., Parker, R. & Song, H. (2004). *Protein Sci.* **13**, 2673–2684.
Chim, N. *et al.* (2011). *Tuberculosis*, **91**, 155–172.
Chim, N., McMath, L. M., Beeby, M. & Goulding, C. W. (2009). *Infect. Disord. Drug Targets*, **9**, 475–492.
Cole, S. T. & Riccardi, G. (2011). *Curr. Opin. Microbiol.* **14**, 570–576.
Crespo, M. D. & Rubini, M. (2011). *PLoS One*, **6**, e19425.
Cutruzzolà, F., Arese, M., Ranghino, G., van Pouderoyen, G., Canters, G. & Brunori, M. (2002). *J. Inorg. Biochem.* **88**, 353–361.
D'Arcy, A., Villard, F. & Marsh, M. (2007). *Acta Cryst.* D**63**, 550–554.
Dill, K. A. (1990). *Biochemistry*, **29**, 7133–7155.
Dougherty, D. A. (2007). *J. Nutr.* **137**, 1504S–1508S.
Dresen, C., Lin, L. Y.-C., D'Angelo, I., Tocheva, E. I., Strynadka, N. & Eltis, L. D. (2010). *J. Biol. Chem.* **285**, 22264–22275.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N. & Nordlund, P. (2006). *Anal. Biochem.* **357**, 289–298.
Falkinham, J. O. (2002). *Clin. Chest Med.* **23**, 529–551.
Fullam, E. (2007). DPhil thesis. University of Oxford.

Fullam, E., Abuhammad, A., Wilson, D. L., Anderton, M. C., Davies, S. G., Russell, A. J. & Sim, E. (2011). *Bioorg. Med. Chem. Lett.* **21**, 1185–1190.

Fullam, E., Kawamura, A., Wilkinson, H., Abuhammad, A., Westwood, I. & Sim, E. (2009). *Protein J.* **28**, 281–293.

Fullam, E., Talbot, J., Abuhammed, A., Westwood, I., Davies, S. G., Russell, A. J. & Sim, E. (2013). *Bioorg. Med. Chem. Lett.* **23**, 2759–2764.

Fullam, E., Westwood, I. M., Anderton, M. C., Lowe, E. D., Sim, E. & Noble, M. E. M. (2008). *J. Mol. Biol.* **375**, 178–191.

Garman, E. F. & Mitchell, E. P. (1996). *J. Appl. Cryst.* **29**, 584–587.

Ginsberg, A. (2011). *Future Med. Chem.* **3**, 1247–1252.

Griffin, J. E., Pandey, A. K., Gilmore, S. A., Mizrahi, V., McKinney, J. D., Bertozzi, C. R. & Sassetti, C. M. (2012). *Chem. Biol.* **19**, 218–227.

Grosset, J. H., Singer, T. G. & Bishai, W. R. (2012). *Int. J. Tuberc. Lung Dis.* **16**, 1005–1014.

Holton, S. J., Dairou, J., Sandy, J., Rodrigues-Lima, F., Dupret, J.-M., Noble, M. E. M. & Sim, E. (2005). *Acta Cryst.* F**61**, 14–16.

Jankute, M., Grover, S., Rana, A. K. & Besra, G. S. (2012). *Future Microbiol.* **7**, 129–147.

Kawamura, A. (2005). DPhil thesis. University of Oxford.

Kemp, M. M., Weïwer, M. & Koehler, A. N. (2012). *Bioorg. Med. Chem.* **20**, 1979–1989.

Kendall, S. L., Burgess, P., Balhana, R., Withers, M., Ten Bokum, A., Lott, J. S., Gao, C., Uhia-Castro, I. & Stoker, N. G. (2010). *Microbiology*, **156**, 1362–1371.

Lack, N. A., Kawamura, A., Fullam, E., Laurieri, N., Beard, S., Russell, A. J., Evangelopoulos, D., Westwood, I. & Sim, E. (2009). *Biochem. J.* **418**, 369–378.

Lack, N., Lowe, E. D., Liu, J., Eltis, L. D., Noble, M. E. M., Sim, E. & Westwood, I. M. (2008). *Acta Cryst.* F**64**, 2–7.

Lack, N. A., Yam, K. C., Lowe, E. D., Horsman, G. P., Owen, R. L., Sim, E. & Eltis, L. D. (2010). *J. Biol. Chem.* **285**, 434–443.

Laurieri, N., Crawford, M. H., Kawamura, A., Westwood, I. M., Robinson, J., Fletcher, A. M., Davies, S. G., Sim, E. & Russell, A. J. (2010). *J. Am. Chem. Soc.* **132**, 3238–3239.

Lavinder, J. J., Hari, S. B., Sullivan, B. J. & Magliery, T. J. (2009). *J. Am. Chem. Soc.* **131**, 3794–3795.

Li, X., Chen, Y., Lu, S., Huang, Z., Liu, X., Wang, Q., Shi, T. & Zhang, J. (2013). *J. Mol. Graph. Model.* **40**, 30–39.

Lienhardt, C., Glaziou, P., Uplekar, M., Lönnroth, K., Getahun, H. & Raviglione, M. (2012). *Nature Rev. Microbiol.* **10**, 407–416.

Makarov, V. *et al.* (2009). *Science*, **324**, 801–804.

Målen, H., Pathak, S., Søfteland, T., de Souza, G. A. & Wiker, H. G. (2010). *BMC Microbiol.* **10**, 132.

Martins, M., Pluvinage, B., Li de la Sierra-Gallay, I., Barbault, F., Dairou, J., Dupret, J.-M. & Rodrigues-Lima, F. (2008). *J. Mol. Biol.* **383**, 549–560.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Murillo, A. C. *et al.* (2007). *Infect. Disord. Drug Targets*, **7**, 127–139.

Murphy, K. P. (2001). *Protein Structure, Stability, and Folding.* Totowa: Humana Press.

Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J. L., Stuart, D. I. & Perrakis, A. (2005). *Acta Cryst.* D**61**, 1426–1431.

Niesen, F. H., Berglund, H. & Vedadi, M. (2007). *Nature Protoc.* **2**, 2212–2221.

Norel, R., Lin, S. L., Wolfson, H. J. & Nussinov, R. (1994). *Biopolymers*, **34**, 933–940.

Ouellet, H., Johnston, J. B. & de Montellano, P. R. (2011). *Trends Microbiol.* **19**, 530–539.

Payton, M., Mushtaq, A., Yu, T.-W., Wu, L.-J., Sinclair, J. & Sim, E. (2001). *Microbiology*, **147**, 1137–1147.

Pluvinage, B., Li de la Sierra-Gallay, I., Kubiak, X., Xu, X., Dairou, J., Dupret, J.-M. & Rodrigues-Lima, F. (2011). *FEBS Lett.* **585**, 3947–3952.

Porter, S. W., Xu, Q. & West, A. H. (2003). *Eukaryot. Cell*, **2**, 27–33.

Price, W. N. *et al.* (2009). *Nature Biotechnol.* **27**, 51–57.

Rowland, K. (2012). *Nature News*, doi:10.1038/nature.2012.9797.

Sandy, J., Mushtaq, A., Kawamura, A., Sinclair, J., Sim, E. & Noble, M. (2002). *J. Mol. Biol.* **318**, 1071–1083.

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

Shaw Stewart, P. D., Kolak, S. A., Briggs, R. A., Chayen, N. E. & Baldock, P. F. M. (2011). *Cryst. Growth Des.* **11**, 3432–3441.

Shenoi, S. & Friedland, G. (2009). *Annu. Rev. Med.* **60**, 307–320.

Sikora, A. L., Frankel, B. A. & Blanchard, J. S. (2008). *Biochemistry*, **47**, 10781–10789.

Sim, E., Fakis, G., Laurieri, N. & Boukouvala, S. (2012). *Adv. Pharmacol. Chemother.* **63**, 169–205.

Sim, E., Sandy, J., Evangelopoulos, D., Fullam, E., Bhakta, S., Westwood, I., Krylova, A., Lack, N. & Noble, M. (2008). *Curr. Drug Metab.* **9**, 510–519.

Sinclair, J. C., Sandy, J., Delgoda, R., Sim, E. & Noble, M. E. M. (2000). *Nature Struct. Biol.* **7**, 560–564.

Squire, S. B., Obasi, A. & Nhlema-Simwaka, B. (2006). *Lancet*, **367**, 955–957.

Terwilliger, T. C. *et al.* (2003). *Tuberculosis*, **83**, 223–249.

Udwadia, Z. F., Amale, R. A., Ajbani, K. K. & Rodrigues, C. (2011). *Clin. Infect. Dis.* **54**, 579–581.

Van der Geize, R., Yam, K., Heuser, T., Wilbrink, M. H., Hara, H., Anderton, M. C., Sim, E., Dijkhuizen, L., Davies, J. E., Mohn, W. W. & Eltis, L. D. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 1947–1952.

Vedadi, M., Arrowsmith, C. H., Allali-Hassani, A., Senisterra, G. & Wasney, G. A. (2010). *J. Struct. Biol.* **172**, 107–119.

Vedadi, M. *et al.* (2006). *Proc. Natl Acad. Sci. USA*, **103**, 15835–15840.

Walraven, J. M., Trent, J. O. & Hein, D. W. (2007). *Drug Metab. Dispos.* **35**, 1001–1007.

Watson, A. P., Wang, P. D. & Sim, E. (1990). *Biochem. Pharmacol.* **39**, 647–654.

Westwood, I. (2005). DPhil thesis. University of Oxford, England..

Westwood, I. M., Bhakta, S., Russell, A. J., Fullam, E., Anderton, M. C., Kawamura, A., Mulvaney, A. W., Vickers, R. J., Bhowruth, V., Besra, G. S., Lalvani, A., Davies, S. G. & Sim, E. (2010). *Protein Cell*, **1**, 82–95.

Westwood, I. M., Kawamura, A., Fullam, E., Russell, A. J., Davies, S. G. & Sim, E. (2006). *Curr. Top. Med. Chem.* **6**, 1641–1654.

Westwood, I. M., Kawamura, A., Russell, A. J., Sandy, J., Davies, S. G. & Sim, E. (2011). *Comb. Chem. High Throughput Screen.* **14**, 117–124.

World Health Organization (2012). *Global Tuberculosis Report 2012*. Geneva: World Health Organization. http://www.who.int/tb/publications/global_report/en/.

Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.

Wu, H., Dombrovsky, L., Tempel, W., Martin, F., Loppnau, P., Goodfellow, G. H., Grant, D. M. & Plotnikov, A. N. (2007). *J. Biol. Chem.* **282**, 30189–30197.

Xu, D., Tsai, C.-J. & Nussinov, R. (1997). *Protein Eng.* **10**, 999–1012.

Yam, K. C., D'Angelo, I., Kalscheuer, R., Zhu, H., Wang, J.-X., Snieckus, V., Ly, L. H., Converse, P. J., Jacobs, W. R. Jr, Strynadka, N. & Eltis, L. D. (2009). *PLoS Pathog.* **5**, e1000344.

Yeh, J. I. (2009). *Protein Crystallization*, 2nd ed., edited by T. M. Bergfors, pp. 331–349. La Jolla: International University Line.

Zhang, Y. (2005). *Annu. Rev. Pharmacol. Toxicol.* **45**, 529–564.