Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Insight into cofactor recognition in arylamine *N*-acetyltransferase enzymes: structure of *Mesorhizobium loti* arylamine *N*-acetyltransferase in complex with coenzyme A

Ximing Xu,[a] Inés Li de la Sierra-Gallay,[b] Xavier Kubiak,[a,c] Romain Duval,[a] Alain F. Chaffotte,[d] Jean-Marie Dupret,[a,e] Ahmed Haouz[f] and Fernando Rodrigues-Lima[a,e]*

[a]Université Paris Diderot, Sorbonne Paris Cité, Unité BFA, CNRS UMR 8251, 75013 Paris, France, [b]Université Paris-Sud, Institut de Biochimie et Biophysique Moléculaire et Cellulaire, CNRS UMR 8619, 91405 Orsay, France, [c]Molecular Neuropharmacology and Genetics Laboratory, University of Copenhagen, DK-2200 Copenhagen, Denmark, [d]Institut Pasteur, Unité de Résonance Magnétique Nucléaire des Biomolécules, 75015 Paris, France, [e]UFR Sciences du Vivant, Université Paris Diderot, 75013 Paris, France, and [f]Institut Pasteur, Plateforme de Cristallographie, CNRS UMR 3528, 75015 Paris, France

Correspondence e-mail: fernando.rodrigues-lima@univ-paris-diderot.fr

Received 11 September 2014
Accepted 17 November 2014

PDB references: arylamine *N*-acetyltransferase 1, F42W mutant, 4nv8; complex with CoA, 4nv7

Arylamine *N*-acetyltransferases (NATs) are xenobiotic metabolizing enzymes that catalyze the acetyl-CoA-dependent acetylation of arylamines. To better understand the mode of binding of the cofactor by this family of enzymes, the structure of *Mesorhizobium loti* NAT1 [(RHILO)NAT1] was determined in complex with CoA. The F42W mutant of (RHILO)NAT1 was used as it is well expressed in *Escherichia coli* and displays enzymatic properties similar to those of the wild type. The apo and holo structures of (RHILO)NAT1 F42W were solved at 1.8 and 2 Å resolution, respectively. As observed in the *Mycobacterium marinum* NAT1–CoA complex, in (RHILO)NAT1 CoA binding induces slight structural rearrangements that are mostly confined to certain residues of its 'P-loop'. Importantly, it was found that the mode of binding of CoA is highly similar to that of *M. marinum* NAT1 but different from the modes reported for *Bacillus anthracis* NAT1 and *Homo sapiens* NAT2. Therefore, in contrast to previous data, this study shows that different orthologous NATs can bind their cofactors in a similar way, suggesting that the mode of binding CoA in this family of enzymes is less diverse than previously thought. Moreover, it supports the notion that the presence of the 'mammalian/eukaryotic insertion loop' in certain NAT enzymes impacts the mode of binding CoA by imposing structural constraints.

1. Introduction

Arylamine *N*-acetyltransferases (NATs; EC 2.3.1.5) are a family of enzymes that catalyze the acetyl coenzyme A (AcCoA)-dependent acetylation of arylamines (Riddle & Jencks, 1971; Sim, Walters *et al.*, 2008). NAT enzymes are found in a range of eukaryotic and prokaryotic species, where they have diverse functions (Grant *et al.*, 1991; Rodrigues-Lima & Dupret, 2002; Sim, Walters *et al.*, 2008; Glenn *et al.*, 2011). In humans, NATs play a key role in the detoxification and/or bioactivation of aromatic amine drugs and xenobiotics (Hein, 2002). Although the role of NATs in prokaryotes remains unclear, these enzymes may contribute to adaptive and/or defence mechanisms towards environmental toxins present in the habitats of bacteria (Rodrigues-Lima *et al.*, 2006; Sim, Walters *et al.*, 2008). More importantly, certain bacterial NATs have been shown to acetylate and inactivate different antibiotics (Payton *et al.*, 1999; Pluvinage *et al.*, 2007; Sim, Sandy *et al.*, 2008). Indeed, *Mycobacterium tuberculosis* NAT [(MYCTU)NAT1] is known to acetylate isoniazid (INH), and increased expression of this enzyme results in increased INH resistance (Payton *et al.*, 1999). Moreover, deletion of the *nat* gene in *M. smegmatis* and *M. bovis* leads to strains that are more sensitive to INH (Payton *et al.*, 2001; Bhakta *et al.*, 2004). In addition, deletion of *nat* in *M. bovis* also affects cell-wall

composition and the biosynthesis of mycolic acids, thus increasing susceptibility to antibiotics that permeate the cell wall (Bhakta *et al.*, 2004). In *Bacillus anthracis*, the *B. anthracis* NAT1 [(BACAN)NAT1] isoenzyme efficiently acetylates sulfamethoxazole and affords higher than normal resistance to this antimicrobial when expressed in *Escherichia coli* (Pluvinage *et al.*, 2007). Certain prokaryotic NAT enzymes therefore appear to be attractive therapeutic targets for the development of antibacterial compounds (Sim *et al.*, 2012, 2014). Recently, chemical drugs aimed at inhibiting mycobacterial NAT enzymes (including *M. tuberculosis* and *M. marinum* NATs) have been identified (Sim *et al.*, 2012, 2014). Structural approaches, in particular X-ray crystallography, have been important to better understand the structure, the catalytic mechanisms and the functions of this family of enzymes (Sim, Walters *et al.*, 2008; Kubiak, Dairou *et al.*, 2013). These structures may also be useful for drug design (Sim *et al.*, 2012, 2014). To date, the X-ray crystal structures of the two human NAT isoenzymes and of 11 prokaryotic NATs have been reported in the literature or in the Protein Data Bank (PDB) (Sinclair *et al.*, 2000; Sandy *et al.*, 2002; Westwood *et al.*, 2005; Holton *et al.*, 2005; Wu *et al.*, 2007; Fullam *et al.*, 2008; Martins *et al.*, 2008; Pluvinage *et al.*, 2011; Kubiak, Li de la Sierra-Gallay *et al.*, 2013; Abuhammad *et al.*, 2013; Cocaign *et al.*, 2014). In addition, the NMR structure of Syrian hamster NAT2 [(MESAU)NAT2] has been reported (Zhang *et al.*, 2006). All of these structural studies identified a common fold that comprises three domains and a cysteine protease-like catalytic triad in the active site (Sim, Walters *et al.*, 2008; Grant, 2008; Kubiak, Dairou *et al.*, 2013). However, although eukaryotic and prokaryotic NATs share the same fold, differences at the amino-acid level are known to have some structural and functional consequences (Westwood *et al.*, 2005; Zhang *et al.*, 2006; Wu *et al.*, 2007; Fullam *et al.*, 2008; Sim, Walters *et al.*, 2008).

To date, only the structures of *Homo sapiens* NAT2 [(HUMAN)NAT2], *M. marinum* NAT1 [(MYCMR)NAT1] and (BACAN)NAT1 have been obtained in complex with the enzyme cofactor CoA (for a review, see Kubiak, Dairou *et al.*, 2013). Analysis of these three complexes demonstrated marked differences in the mode of recognition and the location of the cofactor between these three NAT orthologues (Kubiak, Dairou *et al.*, 2013). In particular, the presence of an insertion of about 15 amino acids (known as the 'mammalian insertion loop' or the 'eukaryotic insertion loop') in (HUMAN)NAT2 and (BACAN)NAT1 was found to contribute to the mode of recognition of the cofactor by these two NAT isoforms (Kubiak, Dairou *et al.*, 2013).

To further understand the mode of binding of the cofactor by this family of enzymes, we determined the structure of *Mesorhizobium loti* NAT1 [(RHILO)NAT1] in complex with CoA. Like (MYCMR)NAT1 and in contrast to (HUMAN)NAT2 and (BACAN)NAT1, (RHILO)NAT1 lacks the 'mammalian/eukaryotic insertion loop'. We obtained the apo and holo structures of (RHILO)NAT1 at resolutions of 1.8 and 2 Å, respectively. We found that CoA binding to (RHILO)NAT1 induces slight structural rearrangements that are mostly confined to certain residues of the 'P-loop', which is in agreement with the data obtained for the (MYCMR)NAT1–CoA complex. More importantly, we found that the mode of binding CoA was highly similar (with the same locations/orientations and geometries) to that observed for (MYCMR)NAT1 but different from the modes reported for (BACAN)NAT1 and (HUMAN)NAT2. Our findings suggest that the mode of CoA binding by NAT enzymes is less diverse than previously supposed. It further emphasizes the notion that the presence of the 'mammalian/eukaryotic insertion loop' in certain NAT enzymes impacts the mode of binding CoA by imposing structural constraints. By demonstrating both similarities and an important divergence in cofactor binding among different NAT enzymes, we provide a better understanding of the structures and functions of members within this important family of xenobiotic metabolizing enzymes.

2. Materials and methods

2.1. Materials

Unless otherwise stated, all reagents were purchased from Sigma–Aldrich.

2.2. Protein expression, purification and crystallization of the (RHILO)NAT1 (F42W) enzyme

Recombinant (RHILO)NAT1 (F42W mutant) protein was expressed in *E. coli* BL21 (DE3) cells and purified as a 6×His-tagged protein as described previously (Rodrigues-Lima *et al.*, 2006). Briefly, transformed bacterial cells were grown at 16°C for 15 h in the presence of 0.5 m*M* isopropyl β-D-1-thio-galactopyranoside (IPTG). The recombinant (RHILO)NAT1 enzyme present in bacterial extracts was purified by Ni–NTA resin affinity chromatography. The purified enzyme was reduced with 10 m*M* dithiothreitol (DTT) prior to overnight dialysis against 25 m*M* Tris–HCl pH 7.5, 1 m*M* EDTA and was concentrated to 15 mg ml^{-1} using ultracentrifugation concentrators (Amicon).

Protein crystallization screening was carried out by the sitting-drop vapour-diffusion method using a Mosquito (TTP Labtech) automated crystallization system. Different crystallization solutions (576) were tested with (RHILO)NAT1 (F42W mutant) protein at 5 mg ml^{-1}. Crystals were reproduced manually according to hit conditions using the hanging-drop vapour-diffusion technique (at 18°C). For the apo-enzyme, crystals were grown in 20%(*w*/*v*) PEG 4000, 0.6 *M* NaCl, 0.1 *M* MES buffer pH 6.5. Crystallization of the enzyme in the presence of AcCoA (10 m*M* final concentration) was carried out in 30%(*w*/*v*) PEG 5000, 0.2 *M* ammonium sulfate, 0.1 *M* MES buffer pH 6.5. All crystals were flash-cooled in liquid nitrogen (100 K) using Paratone and paraffin oils (in a 1:1 ratio) as a cryoprotectant.

2.3. Data collection, structure determination and refinement

X-ray diffraction data were collected at 1.07169 Å (apo-enzyme) and 0.98011 Å (holoenzyme) on the PROXIMA1

beamline at Synchrotron SOLEIL, St Aubin, France using a DECTRIS PILATUS 6M detector. Diffraction data were processed by *xia*2 (Winter *et al.*, 2010) using *XDS* and *XSCALE* (Kabsch, 2010). Both the apo (PDB entry 4nv8) and holo (PDB entry 4nv7) protein structures were solved by the molecular-replacement method using *Phaser* from *CCP*4 (McCoy *et al.*, 2007) with the wild-type (RHILO)NAT1 structure (PDB entry 2bsz; Holton *et al.*, 2005) as a template model. The CoA conformation was determined by *LigandFit* in *PHENIX* (Adams *et al.*, 2010). Structure building and refinement were carried out manually with *Coot* (Emsley *et al.*, 2010) and *REFMAC*5 (Murshudov *et al.*, 2011) using local NCS restraints on atom positions. The quality of the final protein models was validated by *MolProbity* (Chen *et al.*, 2010). Data-collection, refinement and validation statistics are summarized in Table 1. Structure representations and models were generated with *UCSF Chimera* (Pettersen *et al.*, 2004).

2.4. Circular-dichroism spectroscopy

Prior to circular-dichroism (CD) spectroscopy, wild-type (RHILO)NAT1 and the F42W mutant were dialyzed against 20 m*M* potassium phosphate buffer pH 7.5. All CD spectra were measured with an Aviv 215 spectropolarimeter (Aviv Biomedical). Far-UV CD spectra were acquired from 260 to 180 nm (final protein concentration of 0.5 mg ml^{-1}) using a cylindrical cell with 0.02 cm path length. Ellipticity was measured every 0.5 nm (2.0 nm bandwidth) with an average integration time of 1.0 s at 25°C. Each scan was repeated in quadruplicate, and the baseline of buffer only was then subtracted from the average value. Spectra were normalized to the protein concentration, and ellipticity was converted to mean molar differential coefficient per residue ($\Delta\varepsilon$). Near-UV (250–350 nm) spectra were measured at 25°C with a 1 nm step and 1 s averaging time for every step, using a rectangular cell with 1.0 cm path length. The protein concentration of each sample had an absorbance at 280 nm of over 1.50. Each spectrum was the average value of four consecutive scans after subtraction of the baseline of the buffer. The final spectral data were normalized to the molar concentration of the peptide chain.

2.5. Enzyme assays

Enzymatic reactions were carried out using the DMAB method (Coroneos *et al.*, 1991). Wild-type (RHILO)NAT1 or (RHILO)NAT1 F42W (5 µg ml^{-1}) were mixed with *para*-aminosalicylate (PAS; ranging from 12.5 to 250 µ*M* final concentration) and the reaction was started by adding AcCoA (ranging from 25 to 100 µ*M* final concentration). The reactions were carried out at 25°C in 25 m*M* Tris–HCl buffer pH 7.5 in a 50 µl volume. The reactions were quenched by the addition of 40 µl cold 40% trichloroacetic acid. Finally, 100 µl 4-(dimethylamino)benzaldehyde (DMAB; 2% in 9:1 acetonitrile:water) was added to the mixture and the absorbance at 450 nm was measured. Kinetic assays were carried out in triplicate.

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	Holoenzyme (PDB entry 4nv7)	Apoenzyme (PDB entry 4nv8)
Data collection		
Beamline	PROXIMA1, SOLEIL	PROXIMA1, SOLEIL
Wavelength (Å)	0.98011	1.07169
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters		
a (Å)	53.50	52.86
b (Å)	114.86	115.08
c (Å)	115.61	115.57
$\alpha = \beta = \gamma$ (°)	90.00	90.00
Resolution (Å)	38.54–2.02 (2.07–2.02)	32.01–1.84 (1.89–1.84)
Completeness (%)	99.9 (99.9)	99.7 (99.9)
Multiplicity	6.5 (6.4)	5.1 (4.4)
$\langle I/\sigma(I)\rangle$	12.0 (2.8)	14.1 (2.2)
$R_{\rm merge}$†	0.095 (0.636)	0.064 (0.601)
No. of reflections	310329 (22026)	317414 (20228)
No. of unique reflections	47549 (3461)	61856 (4546)
Wilson B factor (Å)	28.511	25.942
Refinement		
Resolution (Å)	38.54–2.02 (2.07–2.02)	32.01–1.84 (1.89–1.84)
$R_{\rm work}$	0.1931 (0.2605)	0.2036 (0.2828)
$R_{\rm free}$	0.2208 (0.2717)	0.2281 (0.3173)
No. of non-H atoms		
Total	4535	4449
Protein	4223	4232
Ligands	96	
Water	216	217
No. of protein residues	540	538
R.m.s.d., bonds (Å)	0.008	0.012
R.m.s.d., angles (°)	1.33	1.45
Ramachandran plot		
Favoured (%)	97	97
Outliers (%)	0	0
Average B factor (Å^2)		
Overall	36.4	33.8
Protein	36.2	33.7
Ligands	45.7	
Solvent	35.8	35.3

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the mean intensity of reflection hkl.

3. Results and discussion

3.1. Overall structure of apo and CoA-bound (RHILO)NAT1

NAT enzymes catalyze the acetylation of aromatic amines using AcCoA as a cofactor. Whereas the structure of neither a prokaryotic nor a eukaryotic NAT structure has been obtained with AcCoA, three crystal structures of three different NAT enzymes in complex with CoA [(HUMAN)NAT2, PDB entry 2pfr; (MYCMR)NAT, PDB entry 2vfc; (BACAN)NAT1, PDB entry 3lnb] have been reported (Wu *et al.*, 2007; Fullam *et al.*, 2008; Pluvinage *et al.*, 2011). These three structures provided the first molecular details on the mode of binding and location of the cofactor in this family of enzymes. More importantly, the NAT enzymes co-crystallized with CoA suggested that both the mode of binding and the geometry of the cofactor were different in each of the three enzymes. In particular, the presence of a 'mammalian/eukaryotic insertion loop' in (HUMAN)NAT2 (a 17-residue insertion between the β10 and β11 strands) and in (BACAN)NAT1 (a 15-residue insertion between the β9 and β10 strands) but not in (MYCMR)NAT1

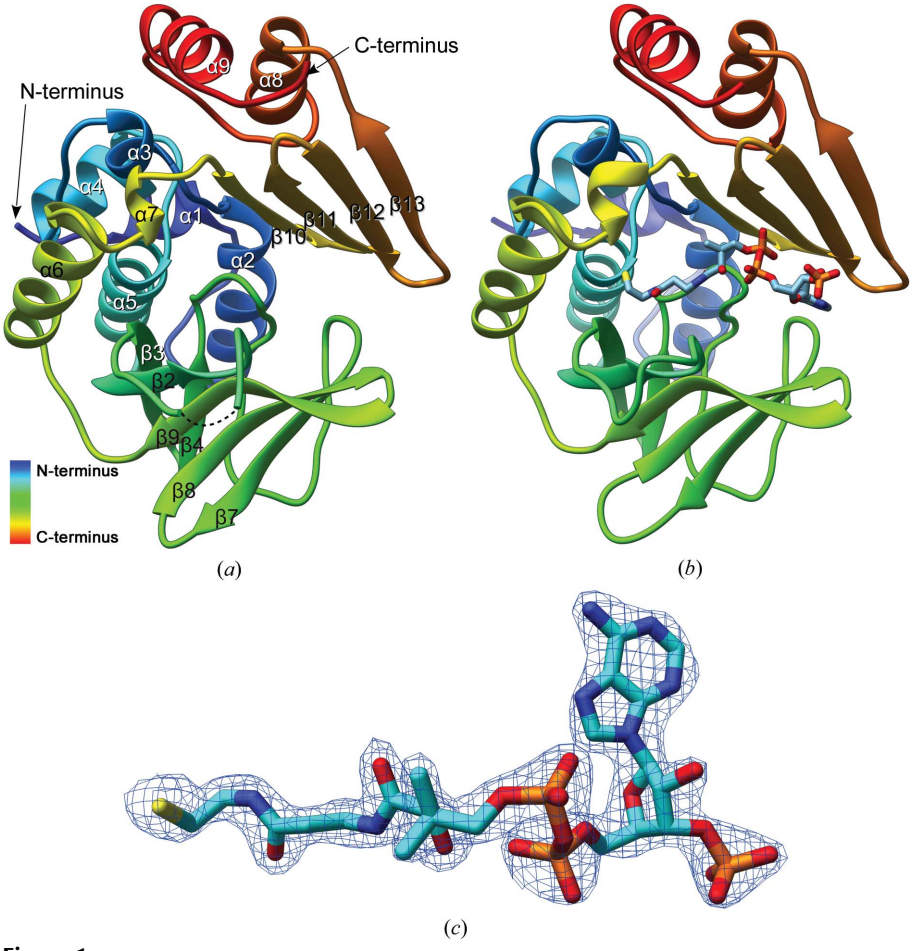


Figure 1
X-ray structure of the (RHILO)NAT1 apo form and of (RHILO)NAT1 complexed with CoA. (*a*) Ribbon diagrams of (RHILO)NAT1 (rainbow colouring represents residues from the N-terminus to the C-terminus). Black dots indicate the position of the unresolved residue 102. (*b*) Structure of (RHILO)NAT1 complexed with CoA (in sticks). (*c*) $2F_o - F_c$ (1.5σ) electron-density map of CoA.

was suggested to contribute to the different mode of binding of CoA (Wu *et al.*, 2007; Fullam *et al.*, 2008; Pluvinage *et al.*, 2011).

In contrast to eukaryotic NAT enzymes, the great majority of the known bacterial NATs are devoid of the 'mammalian/eukaryotic insertion loop' (Walraven *et al.*, 2007; Sim, Lack *et al.*, 2008; Pluvinage *et al.*, 2011; Kubiak, Dairou *et al.*, 2013). Like (MYCMR)NAT1, (RHILO)NAT1 is a well characterized bacterial NAT enzyme which is devoid of the 'insertion loop' (the two equivalent β-strands are connected by only two residues; Holton *et al.*, 2005). To further understand the mode of binding of CoA to bacterial NAT enzymes that lack the 'insertion loop', we have determined the structure of (RHILO)NAT1 in its apo form and in its holo form with CoA. In this study, we used the (RHILO)NAT1 F42W mutant in the crystallization experiments as it was found to be very well expressed as a soluble protein in *E. coli*. Moreover, circular-dichroism analysis supported similar secondary and tertiary structures for the wild type and the (RHILO)NAT1 F42W mutant (Supplementary Fig. S1*a*). As expected, the presence of an additional Trp residue in the (RHILO)NAT1 F42W mutant slightly impacts the near-UV spectrum (Supplementary Fig. S1*b*). In addition, steady-state kinetics supported similar catalytic efficiencies for both enzymatic forms [$k_{cat}/K_{m(AcCoA)}$ close to $15 \times 10^3\ M^{-1}\ s^{-1}$ and $k_{cat}/K_{m(PAS)}$ close to $9 \times 10^3\ M^{-1}\ s^{-1}$; Supplementary Fig. S2].

The crystal structure of the (RHILO)NAT1 F42W mutant in the apo form was solved by molecular replacement and refined to 1.8 Å resolution using the wild-type (RHILO)NAT1 structure (2 Å resolution) reported previously (Holton *et al.*, 2005). As expected, the two structures were highly similar (r.m.s.d. of 0.4 Å over 267 residues) and consisted of the three classical domains (domain I, α-helical bundle, amino acids 1–86; domain II, β-barrel, amino acids 87–180; domain III, α/β lid, amino acids 181–278) reported in all NAT structures (Kubiak, Dairou *et al.*, 2013; Fig. 1 and Supplementary Fig. S3). This is in agreement with the CD data (Supplementary Fig. S1). More importantly, we obtained a crystal structure at 2 Å resolution of the holo form of (RHILO)NAT1 (F42W mutant) in complex with CoA (Fig. 1). Both the apo and the holo crystals belonged to space group $P2_12_12_1$ (Table 1). Clear electron density was attributable to residues 5–275 (except for residue 102) and 4–275 in the apo and holo forms, respectively. The positive $2F_o - F_c$ electron density observed in the active site of the holo crystal was consistently modelled as a CoA molecule (Figs. 1 and 2). Although AcCoA was used in the co-crystallization experiments, only the electron density of a CoA molecule was clearly seen in the active site of the enzyme (Fig. 1). This is not surprising as in its reaction with AcCoA the enzyme is transiently acetylated on its catalytic cysteine to form an acetylcysteine which can be hydrolyzed back to cysteine. In addition, although CoA is a product of the NAT-catalyzed reaction, it is known to bind to NAT enzymes and to reduce their activity through competitive inhibition (Sim, Walters *et al.*, 2008).

3.2. Binding interactions and orientation of CoA in (RHILO)NAT1

CoA was principally found bound between a β-sheet in domain II (made up of strands β7, β8 and β9) and a β-sheet in domain III (made up of strands β11, β12 and β13). A structural alignment of the apo and holo structures reveals highly superimposable C^{α} backbones, with an r.m.s.d. of only 0.2 Å over 267 residues. However, CoA binding induces subtle structural rearrangements that are mostly confined to the 'P-loop' (residues 129–134; Fig. 2*a*). This loop moves by 1.8 Å

towards CoA to allow the formation of hydrogen bonds between the N atom of Gly132 and the carbonyl O atom of the pantothenic acid moiety of CoA, and between the carbonyl O atom of Phe130 and the N atom of the cysteamine group of CoA (Figs. 2*a* and 2*b*). Interestingly, in (MYCMR)NAT1 these amino acids are conserved and make the same interactions with CoA (Fullam *et al.*, 2008). Conversely, in (BACAN)NAT1 these 'P-loop' residues do not interact with the cysteamine group of CoA and but interact in a different manner (only one hydrogen bond and mainly van der Waals interactions) with the pantothenic acid moiety (Pluvinage *et al.*, 2011). In the case of (HUMAN)NAT2, no 'P-loop' residues are involved in CoA binding (Wu *et al.*, 2007). As found in the structure of (MYCMR)NAT1 in complex with CoA, several equivalent aromatic residues of (RHILO)NAT1 such as Trp42 [Phe42 in wild-type (RHILO)NAT1], Tyr72, Phe74, Trp100, Phe130 and

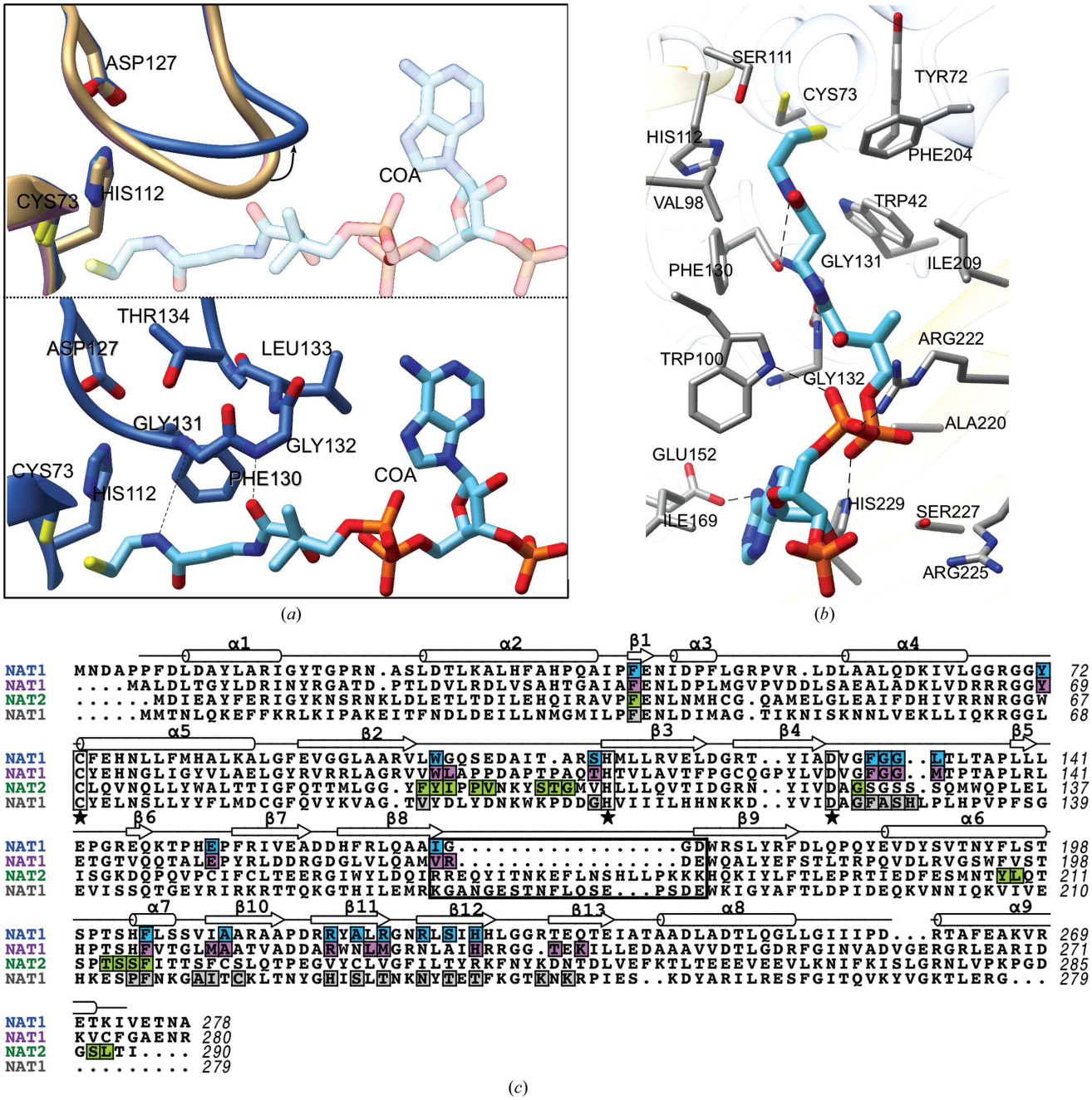


Figure 2
Interactions between CoA and NAT enzymes. (*a*) Upper panel, alignment of the active sites of wild-type apo (RHILO)NAT1 (pink; PDB entry 2bsz) and the apo and holo (RHILO)NAT1 F42W mutant (brown and blue, respectively). Lower panel, interactions between the 'P-loop' residues of the holo (RHILO)NAT1 F42W mutant and bound CoA. Hydrogen bonds are shown as dashed lines. (*b*) Details of the interaction between CoA (cyan) and the holo (RHILO)NAT1 F42W mutant (grey). Hydrogen bonds are shown as dashed lines. (*c*) Amino-acid sequence alignment of (RHILO)NAT1 (blue), (MYCMR)NAT1 (purple), (HUMAN)NAT2 (green) and (BACAN)NAT1 (grey). Residues interacting with CoA are shown in background colours. The catalytic triad Cys–His–Asp is indicated by stars. Boxed residues correspond to the 'mammalian/eukaryotic insertion loop' between β8 and β9 [with the secondary structure of (RHILO)NAT1 as a reference].

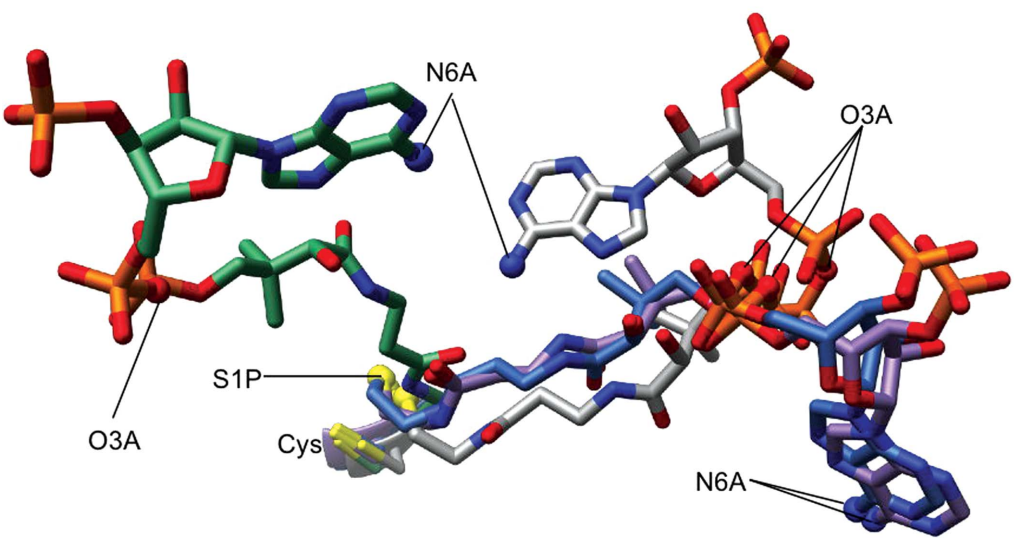
Phe204 make an extensive set of van der Waals contacts and hydrogen bonds to the pyrophosphate, pantothenic acid and cysteamine moieties of CoA (Figs. 2*a* and 2*b*; Fullam *et al.*, 2008). In addition to the above-mentioned residues, we found that the binding of CoA to (RHILO)NAT1 also relies on important interactions with Glu152, Ile169 and His229 (Fig. 2*b*). Indeed, Glu152 was found to make a hydrogen bond to the exocyclic N atom of the adenine moiety of CoA, whilst the side chain of Ile169 packs against the adenine moiety. His229 makes a salt bridge with an O atom of the pyrophosphate group. Interestingly, equivalent residues and interactions were found in the structure of (MYCMR)NAT1 in complex with CoA (involving Glu152, Val169 and His229) but not in the structures of (HUMAN)NAT2 and (BACAN)NAT1 (Fig. 2*c*). These data further show that (MYCMR)NAT1 and (RHILO)NAT1 share a similar mode of recognition of the cofactor. However, multiple alignment analysis indicates that most of the residues that interact with CoA are not conserved across the prokaryotic NAT enzymes (Supplementary Fig. S4).

3.3. Geometry of CoA in (RHILO)NAT1

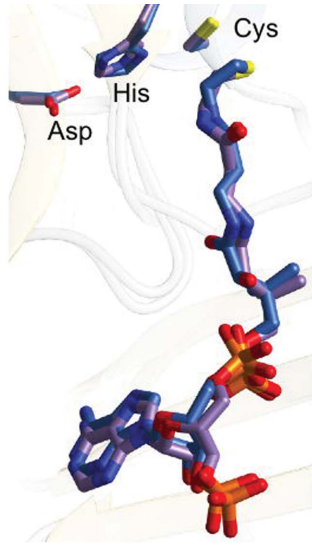
As stated above, the structures of (HUMAN)NAT2, (MYCMR)NAT1 and (BACAN)NAT1 in complex with CoA have shown that the distal regions of CoA (in particular the pyrophosphate and adenosine 3′-phosphate moieties) interact differently with the respective NATs, with marked differences in the geometry of CoA bound to the different enzymes (Kubiak, Dairou *et al.*, 2013; Fig. 2*d*). Nonetheless, as shown in Fig. 2(*d*), the structures of CoA in the active sites of the four NAT isoforms converge at the 2-mercaptoethylamino terminal portion of CoA, close to the catalytic cysteine (Kubiak, Dairou *et al.*, 2013; Fig. 2*d*). This is not surprising as in the first step of catalysis the acetyl group of AcCoA (which is covalently linked to the S atom of the 2-mercaptoethylamine moiety) must be close to the catalytic cysteine to permit acetyl transfer onto this residue (Fullam *et al.*, 2008; Kubiak, Dairou *et al.*, 2013). In addition, the four structures in complex with CoA have shown that the distal regions of the cofactor, in particular the pyrophosphate and adenosine 3′-phosphate moieties, interact differently with the respective NATs, including marked differences in the geometry of CoA (Kubiak, Dairou *et al.*, 2013; Fig. 2*d*). However, a comparison of the structure of the CoA molecules in (RHILO)NAT1 and (MYCMR)NAT1 clearly shows that the whole cofactor adopts the same geometry in the two enzymes (Figs. 2*d*, 2*e* and Table 2). In both the (RHILO)NAT1 and (MYCMR)NAT1 enzymes the cofactor adopts a similar extended conformation that extends for ∼17 Å from the S atom of CoA (S1P) to N6 of the adenine moiety (N6A) (Fig. 2*d* and Table 2). In contrast, in (HUMAN)NAT2 and (BACAN)NAT1 CoA has a much less extended structure, with S1P–N6A distances of 12.7 and 6.8 Å, respectively (Fig. 2*d* and Table 2). In addition, the angle between the S1P, O3A and N6A atoms (S1P–O3A–N6A) which are present in three different moieties of the CoA (2-mercaptoethylamine, pyrophosphate and adenine portions of CoA, respectively) was found to be identical in both (RHILO)NAT1 and (MYCMR)NAT1, with a value of 93°. In contrast, the same S1P–O3A–N6A angle in the CoA structure

Table 2
Geometry of CoA in the four NAT–CoA complex structures.

NAT–CoA complex	S1P–N6A distance (Å)	S1P–O3A–N6A angle (°)
(RHILO)NAT1	16.8	93
(MYCMR)NAT1	17.1	93
(HUMAN)NAT2	12.8	71
(BACAN)NAT1	6.8	23



(*d*)



(*e*)

Figure 2 (continued)
(*d*) CoA conformations from cofactor-bound structures of (RHILO)NAT1 (blue), (MYCMR)NAT1 (purple; PDB entry 2vfc), (HUMAN)NAT2 (green; PDB entry 2pfr) and (BACAN)NAT1 (grey; PDB entry 3lnb). The four structures were aligned and only the catalytic cysteine residue is shown (labelled Cys). (*e*) Superposition of holo (RHILO)NAT1 F42W (blue) and (MYCMR)NAT1 (purple) showing the alignment of CoA molecules (r.m.s.d. 1.14 Å). The catalytic triad residues Cys, His and Asp are shown.

of (HUMAN)NAT2 and (BACAN)NAT1 is very different, with values of 71 and 23°, respectively (Fig. 2*d* and Table 2). Prior to our study, the analyses of the three orthologous NAT structures in complex with CoA [(HUMAN)NAT2, (MYCMR)NAT1 and (BACAN)NAT1] suggested that the mode of CoA binding among NAT enzymes was likely to be diverse as these three isoforms were found to bind CoA in different manners (Wu *et al.*, 2007; Fullam *et al.*, 2008; Pluvinage *et al.*, 2011; Kubiak, Dairou *et al.*, 2013). The (RHILO)NAT1–CoA structure overturns this view and illustrates that two different orthologous NAT enzymes can bind their cofactors in a similar way. In contrast to (RHILO)NAT1 and (MYCMR)NAT1, the two other CoA-bound NATs crystallized possess an additional loop between domain II and domain III corresponding to an 'insertion loop' known as the 'mammalian/eukaryotic insertion loop' in (HUMAN)NAT2 (a 17-residue insertion between the β10 and β11 strands) and in (BACAN)NAT1 (a 15-residue insertion between the β9 and β10 strands) (Fig. 3). This insertion is highly mobile and in the case of (BACAN)NAT1 these amino acids are not seen in the electron-density map (Pluvinage *et al.*, 2011; Kubiak, Li de la Sierra-Gallay *et al.*, 2013; Fig. 3). The reasons for the presence of this insertion loop in the great majority of eukaryotic (especially mammalian) sequences and in certain prokaryotic NATs (especially NATs from *Bacillus* species) is not understood (Kubiak, Dairou *et al.*, 2013), a study has however suggested that it could contribute to protein stability (Walraven *et al.*, 2007). Although it has been shown in (HUMAN)NAT2 and (BACAN)NAT1 that the 'mammalian/eukaryotic insertion' does not play a direct role in CoA binding through molecular contacts, this insertion contributes to the shape of the cleft in which the cofactor binds. In (HUMAN)NAT2 and (BACAN)NAT1 the 'insertion' fills the part of the (MYCMR)NAT1 and (RHILO)NAT1 clefts that becomes occupied by the cofactor in the (MYCMR)NAT1–CoA and (RHILO)NAT1–CoA complexes (Fullam *et al.*, 2008; Fig. 3). Thus, the 'mammalian/eukaryotic insertion' in (HUMAN)NAT2 and (BACAN)NAT1 narrows the cleft in which CoA binds and impacts its geometry and location (Kubiak, Dairou *et al.*, 2013; Fig. 3). Indeed, whereas in (MYCMR)NAT1 and (RHILO)NAT1 CoA fits into an extended cleft that extends for ∼21 Å from the active-site cysteine to the adenine moiety, in (HUMAN)NAT2 and (BACAN)NAT1 the cofactor fits into equivalent but narrower clefts that extend for 16 and 9 Å, respectively (Fullam *et al.*, 2008; Pluvinage *et al.*, 2011; Fig. 3). The structure of the (RHILO)NAT1–CoA complex further supports the notion that the presence of the 'mammalian/eukaryotic insertion' precludes a shared mechanism of recognition of the nucleoside phosphate portion of the cofactor in NAT enzymes

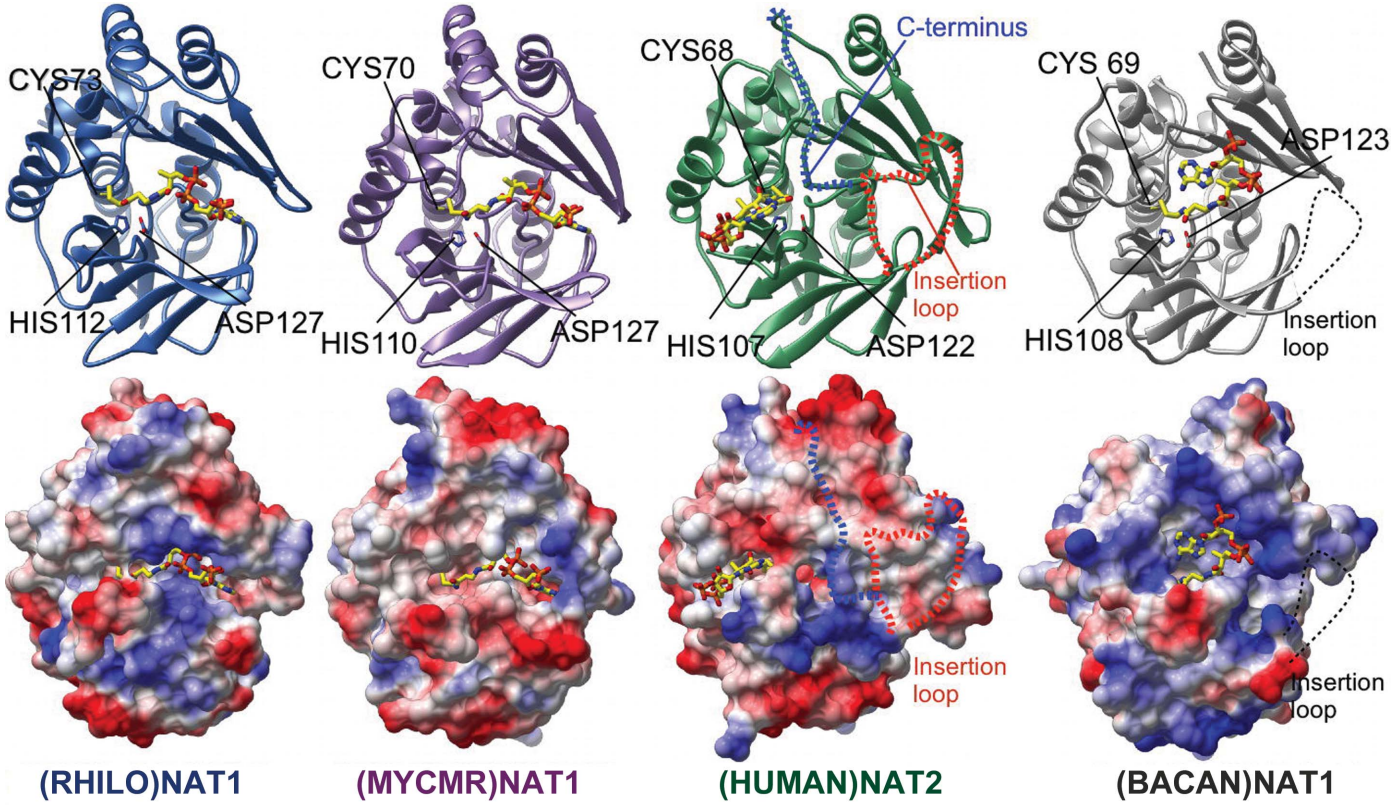


Figure 3
Structural comparison of different CoA-binding conformations in NATs. Overall structures (top) and electrostatic potential representations (bottom) of (RHILO)NAT1 (blue), (MYCMR)NAT1 (purple; PDB entry 2vfc), (HUMAN)NAT2 (green; PDB entry 2pfr) and (BACAN)NAT1 (grey; PDB entry 3lnb) complexed with CoA (yellow sticks) shown as ribbon diagrams. The 'mammalian/eukaryotic insertion loop' of (HUMAN)NAT2 is highlighted by a red dashed line. The 15 amino acids of the insertion loop of (BACAN)NAT1 remain unresolved owing to low electron density and are indicated by a black dashed line.

(Fullam *et al.*, 2008; Pluvinage *et al.*, 2011; Fig. 3). These data suggest that the mode of CoA binding by NAT enzymes is less diverse than previously proposed (Fullam *et al.*, 2008; Pluvinage *et al.*, 2011; Kubiak, Dairou *et al.*, 2013). Our results suggest that NAT isoforms lacking the 'mammalian/eukaryotic insertion loop' (*i.e.* the great majority of prokaryotic NAT enzymes) are likely to bind the cofactor in a similar manner. More structures of NATs in complex with CoA are needed to fully ascertain this point. In addition, as the cofactor-binding site of NATs has been suggested to be a potential site of interest for drug design, our data may be of importance for the identification of small-molecule inhibitors that target the cofactor site in NATs.

4. Related literature

The following references are cited in the Supporting Information for this article: Bond & Schüttelkopf (2009) and Edgar (2004).

This work was funded by grants from the Ministère de l'Enseignement Supérieur et de la Recherche (Université Paris Diderot-Paris 7), the Centre National de la Recherche Scientifique (CNRS) and Caisse d'Assurance Maladie des Professions Libérales-Provinces (CAMPLP). XX was supported by a PhD fellowship from the China Scholarship Council. XK was supported by a PhD fellowship from the Université Paris Diderot-Paris 7. RD was supported by a PhD fellowship from Région Ile-de-France (DIM Sent). We thank Patrick Weber (PF6, Institut Pasteur) for performing robot-driven crystallization trials and Professor Edith Sim for helpful discussions. We acknowledge Synchrotron SOLEIL, St Aubin, France for the provision of synchrotron-radiation facilities and the staff of the PROXIMA1 beamline for assistance during data collections.

References

Abuhammad, A., Lowe, E. D., McDonough, M. A., Shaw Stewart, P. D., Kolek, S. A., Sim, E. & Garman, E. F. (2013). *Acta Cryst.* D**69**, 1433–1446.
Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
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.
Bond, C. S. & Schüttelkopf, A. W. (2009). *Acta Cryst.* D**65**, 510–512.
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.
Cocaign, A., Kubiak, X., Xu, X., Garnier, G., Li de la Sierra-Gallay, I., Chi-Bui, L., Dairou, J., Busi, F., Abuhammad, A., Haouz, A., Dupret, J.-M., Herrmann, J.-L. & Rodrigues-Lima, F. (2014). *Acta Cryst.* D**70**, 3066–3079.
Coroneos, E., Gordon, J. W., Kelly, S. L., Wang, P. D. & Sim, E. (1991). *Biochim. Biophys. Acta*, **1073**, 593–599.
Edgar, R. C. (2004). *Nucleic Acids Res.* **32**, 1792–1797.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Fullam, E., Westwood, I. M., Anderton, M. C., Lowe, E. D., Sim, E. & Noble, M. E. M. (2008). *J. Mol. Biol.* **375**, 178–191.
Glenn, A. E., Karagianni, E. P., Ulndreaj, A. & Boukouvala, S. (2011). *FEBS Lett.* **584**, 3158–3164.
Grant, D. M. (2008). *Curr. Drug Metab.* **9**, 465–470.
Grant, D. M., Blum, M., Beer, M. & Meyer, U. A. (1991). *Mol. Pharmacol.* **39**, 184–191.
Hein, D. (2002). *Mutat. Res.* **506–507**, 65–77.
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.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kubiak, X., Dairou, J., Dupret, J.-M. & Rodrigues-Lima, F. (2013). *Expert Opin. Drug Metab. Toxicol.* **9**, 349–362.
Kubiak, X., Li de la Sierra-Gallay, I., Chaffotte, A. F., Pluvinage, B., Weber, P., Haouz, A., Dupret, J.-M. & Rodrigues-Lima, F. (2013). *J. Biol. Chem.* **288**, 22493–22505.
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.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Payton, M., Auty, R., Delgoda, R., Everett, M. & Sim, E. (1999). *J. Bacteriol.* **181**, 1343–1347.
Payton, M., Gifford, C., Schartau, P., Hagemeier, C., Mushtaq, A., Lucas, S., Pinter, K. & Sim, E. (2001). *Microbiology*, **147**, 3295–3302.
Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
Pluvinage, B., Dairou, J., Possot, O. M., Martins, M., Fouet, A., Dupret, J.-M. & Rodrigues-Lima, F. (2007). *Biochemistry*, **46**, 7069–7078.
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.
Riddle, B. & Jencks, W. P. (1971). *J. Biol. Chem.* **246**, 3250–3258.
Rodrigues-Lima, F., Dairou, J., Diaz, C. L., Rubio, M. C., Sim, E., Spaink, H. P. & Dupret, J.-M. (2006). *Mol. Microbiol.* **60**, 505–512.
Rodrigues-Lima, F. & Dupret, J.-M. (2002). *Biochem. Biophys. Res. Commun.* **293**, 783–792.
Sandy, J., Mushtaq, A., Kawamura, A., Sinclair, J., Sim, E. & Noble, M. (2002). *J. Mol. Biol.* **318**, 1071–1083.
Sim, E., Abuhammad, A. & Ryan, A. (2014). *Br. J. Pharmacol.* **171**, 2705–2725.
Sim, E., Fakis, G., Laurieri, N. & Boukouvala, S. (2012). *Adv. Pharmacol.* **63**, 169–205.
Sim, E., Lack, N., Wang, C.-J., Long, H., Westwood, I., Fullam, E. & Kawamura, A. (2008). *Toxicology*, **254**, 170–183.
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.
Sim, E., Walters, K. & Boukouvala, S. (2008). *Drug Metab. Rev.* **40**, 479–510.
Sinclair, J. C., Sandy, J., Delgoda, R, Sim, E. & Noble, M. E. M., (2000). *Nature Struct. Biol.* **7**, 560–564.
Walraven, J. M., Trent, J. O. & Hein, D. W. (2007). *Drug. Metab. Dispos.* **35**, 1001–1007.
Westwood, I. M., Holton, S. J., Rodrigues-Lima, F., Dupret, J.-M., Bhakta, S., Noble, M. E. M. & Sim, E. (2005). *Biochem. J.* **385**, 605–612.
Winter, G., Dökel, S., Jones, A. K., Scheerer, P., Krauss, N., Höhne, W. & Friedrich, B. (2010). *Acta Cryst.* F**66**, 452–455.
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.
Zhang, N., Liu, L., Liu, F., Wagner, C. R., Hanna, P. E. & Walters, K. J. (2006). *J. Mol. Biol.* **363**, 188–200.