Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Structure of *Mesorhizobium loti* arylamine *N*-acetyltransferase 1

Simon J. Holton,[a] Julien Dairou,[b] James Sandy,[c] Fernando Rodrigues-Lima,[b,d] Jean-Marie Dupret,[b,d] Martin E. M. Noble[a] and Edith Sim[c]*

[a]Laboratory of Molecular Biophysics, Department of Biochemistry, Oxford University, South Parks Road, Oxford OX1 3QU, England, [b]CNRS-UMR 7000, Faculté de Médecine Pitié-Salpêtrière, 105 Boulevard de l'Hôpital, 75013 Paris, France, [c]Department of Pharmacology, University of Oxford, Mansfield Road, Oxford OX1 3QT, England, and [d]UFR de Biochimie, Université Denis Diderot-Paris 7, 75005 Paris, France

Correspondence e-mail: edith.sim@pharm.ox.ac.uk

Received 21 November 2004
Accepted 23 November 2004
Online 24 December 2004

PDB Reference: arylamine *N*-acetyltransferase 1, 1v09, r1v09sf.

The arylamine *N*-acetyltransferase (NAT) enzymes have been found in a broad range of both eukaryotic and prokaryotic organisms. The NAT enzymes catalyse the transfer of an acetyl group from acetyl Co-enzyme A onto the terminal nitrogen of a range of arylamine, hydrazine and arylhydrazine compounds. Recently, several NAT structures have been reported from different prokaryotic sources including *Salmonella typhimurium*, *Mycobacterium smegmatis* and *Pseudomonas aeruginosa*. Bioinformatics analysis of the *Mesorhizobium loti* genome revealed two NAT paralogues, the first example of multiple NAT isoenzymes in a eubacterial organism. The *M. loti* NAT 1 enzyme was recombinantly expressed and purified for X-ray crystallographic studies. The purified enzyme was crystallized in 0.5 *M* $Ca(OAc)_2$, 16% PEG 3350, 0.1 *M* Tris–HCl pH 8.5 using the sitting-drop vapour-diffusion method. A data set diffracting to 2.0 Å was collected from a single crystal at 100 K. The crystal belongs to the orthorhombic spacegroup $P2_12_12_1$, with unit-cell parameters $a = 53.2$, $b = 97.3$, $c = 114.3$ Å. The structure was refined to a final free-*R* factor of 24.8%. The structure reveals that despite low sequence homology, *M. loti* NAT1 shares the common fold as reported in previous NAT structures and exhibits the same catalytic triad of residues (Cys-His-Asp) in the active site.

1. Introduction

The arylamine *N*-acetyltransferase (NAT) family of phase II enzymes catalyse the acetylation of primary arylamines, arylhydrazines and *N*-hydroxylated metabolites (Riddle & Jencks, 1971; Weber & Hein, 1985; Evans, 1989; Delomenie *et al.*, 2001). Initially identified in humans through their ability to inactivate the anti-tuberculosis therapeutic isoniazid, NAT-encoding sequences have now been identified in a broad range of eukaryotic and prokaryotic species (Pompeo *et al.*, 2002). To date, X-ray crystallographic structures of three NAT paralogues have been reported, all of which possess a highly conserved three-domain fold whose active sites contain a triad of residues (Cys-His-Asp) proposed to catalyse substrate acetylation (Sinclair *et al.*, 2000).

Mesorhizobium loti is a nitrogen-fixing rhizobacterium that lives in symbiosis with leguminous plants (Kaneko *et al.*, 2000). Recent bioinformatics analysis of the *M. loti* genome revealed two NAT isoforms, NAT1 and NAT2, and was the first example of a eubacterial species with multiple NAT isoforms (Rodrigues-Lima & Dupret, 2002). In this paper, we report the 2 Å X-ray crystallographic structure of *M. loti* NAT1, a comparison with other NAT structures and a discussion of the implications for *M. loti* NAT2.

2. Experimental

The *M. loti* NAT1 open-reading frame was cloned, expressed and purified from *M. loti* MAFF303099 strain. Details of cloning, expression and purification will be published elsewhere.

Pure recombinant protein (in 20 m*M* Tris–HCl pH 7.5, 1 m*M* EDTA, 1 m*M* DTT) was concentrated to 10 mg ml^{-1} using Amicon ultracentrifugation concentrators (Millipore, Watford, Hertfordshire). The crystals described in this paper were grown at 292 K using the sitting-drop vapour-diffusion technique, with initial conditions

formed by mixing equal volumes (1 µl) of the concentrated *M. loti* NAT1 solution with mother liquor [0.5 *M* $Ca(OAc)_2$, 16% PEG 3350, 0.1 *M* Tris–HCl pH 8.5]. Typically, crystals appeared after 2–5 d.

Crystals cryoprotected in paraffin oil were frozen at 100 K and diffraction data were collected at beamline ID14-2 at the European Synchrotron Radiation Facility, Grenoble, France. Data were indexed, integrated and scaled using the programs *MOSFLM* and *SCALA* (Collaborative Computational Project, Number 4, 1994). Initial phases were determined by molecular replacement with the program *AMoRe* using NAT from *Mycobacterium smegmatis* (PDB code 1gx3) as a search model. Initial model building was carried out using a combination of automatic rebuilding programs (*ARP*/*wARP* and *FFFEAR*; Collaborative Computational Project, Number 4, 1994) and manual rebuilding using the program *O* (Jones *et al.*, 1991). Subsequent refinement involved iterative manual rebuilding and maximum-likelihood refinement using the program *REFMAC*5 (Collaborative Computational Project, Number 4, 1994). A final refinement step was to solvate the protein using the program *ARP* (Perrakis *et al.*, 1999). The stereochemical quality of the final model was verified using the program *PROCHECK* (Laskowski *et al.*, 1993). Refinement statistics are presented in Table 1.

Table 1
Crystallographic summary.

Values in parentheses are for the highest resolution shell.

Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 53.2, b = 97.3, c = 114.3
Maximum resolution (Å)	2.0 (2.00–2.11)
Observed reflections	185511
Unique reflections	40315
Completeness (%)	98.5 (91.3)
$I/\sigma(I)$	4.9 (2.9)
Mosaicity (°)	0.62
Redundancy	4.6 (2.7)
R_{merge}† (%)	9.8 (24.8)
V_M (Å^3 Da^{-1})	2.4
Refinement	
Protein atoms	4174
Protein residues	
Chain *A*	6–231, 235–275
Chain *B*	7–230, 235–274
Other atoms	256 water molecules
Resolution range (Å)	25.05–2.00
R_{conv}‡ (%)	19.5
R_{free}§ (%)	24.8
Mean temperature factor (Å^2)	
Main chain	26.8
Side chain	30.4
Water atoms	37.3
R.m.s.d. bond lengths (Å)	0.015
R.m.s.d. bond angles (°)	1.457

† $R_{merge} = \sum_h \sum_j |I_{h,j} - \langle I_h \rangle| / \sum_h \sum_j I_{h,j}$, where $I_{h,j}$ is the intensity of the jth observation of unique reflection h. ‡ $R_{conv} = \sum_h ||F_{oh}| - |F_{ch}|| / \sum_h |F_{oh}|$, where F_{oh} and F_{ch} are the observed and calculated structure-factor amplitudes for reflection h. § R_{free} is equivalent to R_{conv}, but is calculated using a 5% disjoint set of reflections excluded from the maximum-likelihood refinement stages.

3. Results and discussion

The *M. loti* NAT1 structure adopts the characteristic three-domain NAT fold in which the first two domains, a helical bundle and a β-barrel, are aligned to allow three residues (Cys73, His112 and Asp127, *M. loti* NAT1 numbering) to form a catalytic triad (Fig. 1). This three-residue structural motif is well documented in NATs and is similar to that found in the structurally related cysteine protease superfamily (Sinclair *et al.*, 2000). The relative geometry of these catalytic residues is absolutely conserved between *M. loti* NAT1 and other NAT structures. The sequence and structural homology within the active site of *M. loti* NAT1 and other NATs is significantly higher than the global sequence identity, which ranges between 34 and 41%, and suggests that the different NAT isoforms share a common enzymatic mechanism. The third domain is linked to the second *via* an interdomain helix (residues 202–207), whilst the interface formed between domains II and III and the lid region together form a substantial active-site cleft. This structure extends the phylogenetic

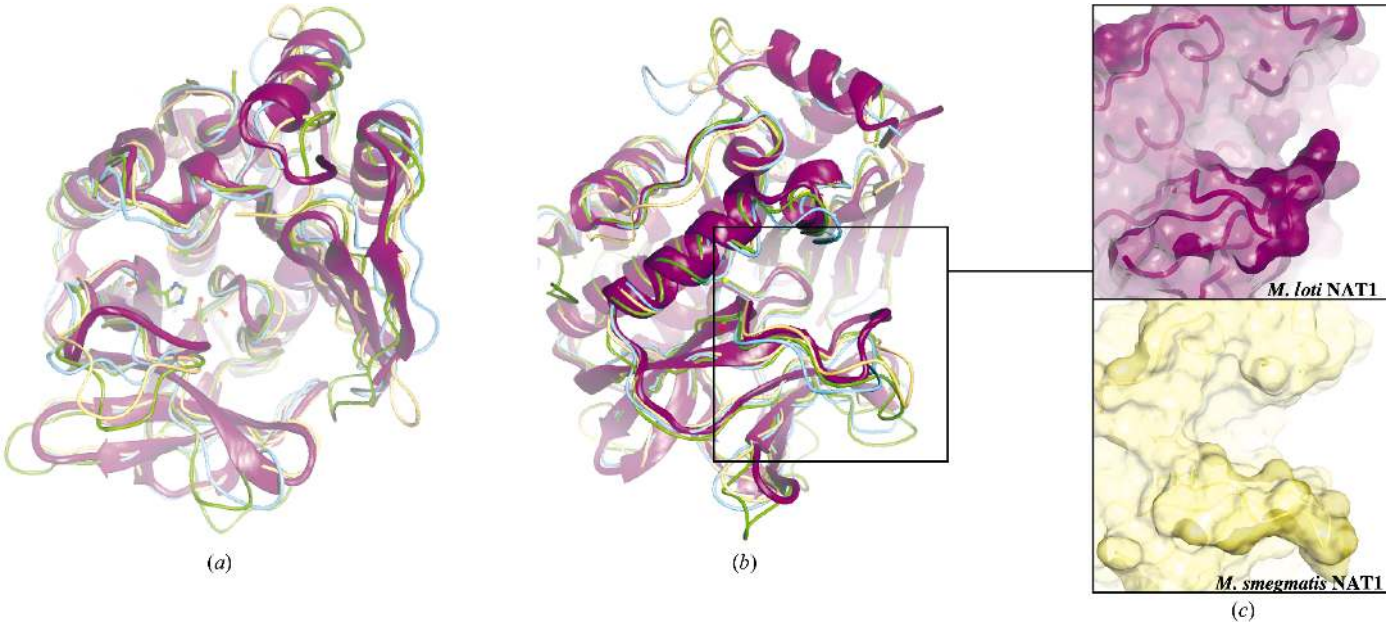

Figure 1
X-ray crystal structure of *M. loti* NAT1. (*a*) and (*b*) Orthogonal secondary-structure representation of *M. loti* NAT1 (mauve) superimposed with worm representations of the NAT structures from *Salmonella typhimurium* (cyan; PDB code 1e2t), *Mycobacterium smegmatis* (yellow; PDB code 1gx3) and *Pseudomonas aeruginosa* (green; PDB code 1w4t) using the program *O* (Jones *et al.*, 1991). The *M. loti* NAT1 active-site catalytic triad is also shown in ball-and-stick representation. (*c*) Surface representation of *M. smegmatis* NAT1 (yellow) and *M. loti* NAT1 (mauve) in the region of the active-site entrance. Underlying secondary-structural elements are drawn in a similar colour scheme. For clarity, the α6 helix, corresponding to residues 183–205 in the *M. loti* structure, is not shown in this figure. The active-site catalytic triad is also shown in ball-and-stick mode to aid orientation. All images were generated using *AESOP* (M. E. M. Noble, Oxford University, England).

range of species over which NAT folds have been observed and supports the prediction of a conserved fold for bacterial NAT orthologues (Payton *et al.*, 2001; Westwood *et al.*, 2005).

Comparisons between NAT structures have previously revealed minor conformational differences resulting from both single-residue insertions/deletions that are accommodated within loop regions of the structure and from mobile loop regions linking more stable secondary-structural elements (Sinclair *et al.*, 2000; Sandy *et al.*, 2002; Westwood *et al.*, 2005). These result in little perturbation to the protein fold and as such are not considered significant. The superposition of *M. loti* NAT1 with other NAT structures reveals a similar pattern of structural conservation (Fig. 1). The largest and potentially most significant deviation centres on the loop between β3 and β4 (approximately residue 103). Although the conformation of this loop differs between all published NAT structures, this is the only example where accessibility of the active site is potentially reduced owing to the conformation of the loop (Fig. 1). Little is known about the interactions between NATs and their substrates/cofactors and so the effect, if any, of such a conformation on the enzymatic character of *M. loti* NAT1 remains to be seen.

Despite sharing moderate to low levels of sequence identity, the high level of structural homology between NAT structures suggests that *M. loti* NAT2 will adopt a similar fold. Mapping sequence conservation between the two *M. loti* NAT paralogues onto the NAT1 structure reveals that with the exception of the active-site core, there are no extensive regions of residue conservation (data not shown). This may reflect a differentiation of cellular role and/or location from a common ancestral protein or simply a lack of evolutionary pressure on conservation away from the active site. In human cells, NAT isoforms exhibit overlapping but distinct substrate selectivity and cellular location profiles (Sim *et al.*, 2000). If the *M. loti* paralogues represent a similar scenario, then the unusually low levels of residue conservation may reflect more divergent enzymatic roles.

In this communication, we have presented the X-ray crystallographic structure of *M. loti* NAT1. Despite low sequence similarity, *M. loti* NAT1 shares the well conserved NAT fold and highlights the sensitive nature of the profile-based search methods which first identified the sequence as encoding a NAT (Rodrigues-Lima & Dupret, 2002). The *M. loti* NAT1 structure adds to the growing number of published NAT structures and will complement current work aimed at identifying bacterial NAT antagonists.

We are grateful to the Wellcome Trust for continued financial support. We also thank Association pour la Recherche sur le Cancer (ARC), Association Française contre les Myopathies and Rétina France for financial support. We gratefully acknowledge the help of the beamline scientists at beamline ID14-2 at the ESRF, Grenoble and also thank Isaac Westwood for helpful discussions. This work would not have been possible without the support of the Franco–British Alliance Partnership Programme (No. PN 03.066).

References

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Delomenie, C., Fouix, S., Longuemaux, S., Brahimi, N., Bizet, C., Picard, B., Denamur, E. & Dupret, J. M. (2001). *J. Bacteriol.* **183**, 3417–3427.
Evans, D. A. (1989). *Pharmacol. Ther.* **42**, 157–234.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kaneko, T. *et al.* (2000). *DNA Res.* **7**, 331–338.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Payton, M., Mushtaq, A., Yu, T. W., Wu, L. J., Sinclair, J. & Sim, E. (2001). *Microbiology*, **147**, 1137–1147.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Pompeo, F., Brooke, E., Kawamura, A., Mushtaq, A. & Sim, E. (2002). *Pharmacogenomics*, **3**, 19–30.
Riddle, B. & Jencks, W. P. (1971). *J. Biol. Chem.* **246**, 3250–3258.
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., Payton, M., Noble, M. & Minchin, R. (2000). *Hum. Mol. Genet.* **9**, 2435–2441.
Sinclair, J. C., Sandy, J., Delgoda, R., Sim, E. & Noble, M. E. (2000). *Nature Struct. Biol.* **7**, 560–564.
Weber, W. W. & Hein, D. W. (1985). *Pharmacol. Rev.* **37**, 25–79.
Westwood, I., Holton, S. J., Rodrigues-Lima, F., Dupret, J.-M., Bhakta, S., Noble, M. & Sim, E. (2005). In the press.