X-ray crystal structure of ornithine acetyltransferase from the clavulanic acid biosynthesis gene cluster

Jonathan M. ELKINS¹, Nadia J. KERSHAW¹ and Christopher J. SCHOFIELD²

Department of Chemistry, Chemistry Research Laboratory, University of Oxford, Mansfield Road, Oxford OX1 3TA, U.K.

The *orf6* gene from the clavulanic acid biosynthesis gene cluster encodes an OAT (ornithine acetyltransferase). Similar to other OATs the enzyme has been shown to catalyse the reversible transfer of an acetyl group from *N*-acetylornithine to glutamate. OATs are Ntn (N-terminal nucleophile) enzymes, but are distinct from the better-characterized Ntn hydrolase enzymes as they catalyse acetyl transfer rather than a hydrolysis reaction. In the present study, we describe the X-ray crystal structure of the OAT, corresponding to the *orf6* gene product, to 2.8 Å (1 Å = 0.1 nm) resolution. The larger domain of the structure consists of an $\alpha\beta\beta\alpha$ sandwich as in the structures of Ntn hydrolase enzymes. However, differences in the connectivity reveal that OATs belong to a structural family different from that of other structurally characterized Ntn enzymes, with one exception: unexpectedly, the

$\alpha\beta\beta\alpha$ sandwich of ORF6 (where ORF stands for open reading frame) displays the same fold as an DmpA (L-aminopeptidase D-ala-esterase/amidase from *Ochrobactrum anthropi*), and so the OATs and DmpA form a new structural subfamily of Ntn enzymes. The structure reveals an $\alpha_2\beta_2$ -heterotetrameric oligomerization state in which the intermolecular interface partly defines the active site. Models of the enzyme–substrate complexes suggest a probable oxyanion stabilization mechanism as well as providing insight into how the enzyme binds its two differently charged substrates.

Key words: arginine biosynthesis, clavulanic acid, β -lactam antibiotic, N-terminal nucleophile hydrolase, ornithine acetyltransferase.

INTRODUCTION

Clavulanic acid is a potent inhibitor of class A serine β -lactamases, and is administered clinically in combination with penicillin antibiotics [1]. It is produced for clinical use by fermentation of *Streptomyces clavuligerus*, hence its biosynthesis is of interest. Currently, the operon for clavulanic acid biosynthesis is believed to comprise 18 ORFs (open reading frames) [2–5]. The first step in the biosynthetic pathway is the reaction of two primary metabolites, L-arginine and glyceraldehyde-3-phosphate, to form *N*²-(2-carboxyethyl)arginine [6,7]. This reaction is catalysed by *N*²-(2-carboxyethyl)arginine synthase, the product of the first ORF in the clavulanic acid gene cluster, *orf2* [7], the crystal structure of which has recently been reported [8]. The actions of the gene products of *orf3*–*orf5*, whose structures have also been elucidated [9–11], then convert *N*²-(2-carboxyethyl)arginine into clavaminic acid (Figure 1). Clavaminic acid is converted into clavaldehyde, the immediate precursor of clavulanic acid, by undefined processes (see [12–14] for reviews).

Orf6 from the clavulanic acid biosynthetic gene cluster of *S. clavuligerus* encodes a protein with 40% identity to an OAT (ornithine acetyltransferase) from *Thermus thermophilus* and 30–36% identity to OATs from other species. In most prokaryotes, OAT (also known as ArgJ) is one of the enzymes involved in arginine biosynthesis. This arginine biosynthesis pathway is initiated by the acetylation of glutamate [15], which is converted through three *N*-acetylated intermediates into *N*-acetylornithine. OAT then transfers an acetyl group from *N*-acetylornithine to glutamate, producing ornithine as well as allowing recycling of the acetyl group (Figure 2). *N*-Acetylglutamate, the first intermediate,

is originally produced from glutamate and acetyl-CoA; some OATs have been shown to be bifunctional, catalysing this acetyl transfer as well as the one from *N*-acetylornithine. ORF6 has been shown to act as a monofunctional OAT, catalysing solely the reaction from *N*-acetylornithine. It is capable of using a number of other acetyl acceptors (lysine, glutamine and arginine), although none function as well as glutamate. It is proposed that ORF6 is involved in increasing the amount of arginine directed specifically towards clavulanic acid biosynthesis rather than primary metabolism [16], as there is already an OAT coded in the arginine biosynthesis gene cluster of *S. clavuligerus* [17]. Furthermore, an *orf6* homologue (named *oat1*) has been identified in *S. clavuligerus* clustered with homologues of *orf2*–*orf4* of the clavulanic acid biosynthetic gene cluster, although it has not been shown to encode an active OAT [18].

OATs are synthesized as precursor proteins that undergo autoproteolytic cleavage between the alanine and threonine residues of a conserved KGXGMXXPX–[M/L]AT[M/L]L motif to form the mature enzyme [19,20]. It has been shown that the catalytic nucleophile involved in acetyl transfer is on the β -subunit [16,20] and it is supposed that the threonine at the N-terminus of the newly created β -subunit is responsible. Site-directed mutagenesis has shown that this threonine is essential for autoproteolysis and acetyltransferase activity [21].

Therefore OATs belong to the family of Ntn (N-terminal nucleophile) enzymes, whose catalytic mechanism involves nucleophilic attack on the carbonyl group of an amide bond using the side-chain of an N-terminal threonine, serine or cysteine. In some cases, the nucleophile is at the N-terminus of the protein ‘as translated’, whereas in other cases the nucleophile is created

Abbreviations used: CNS, crystallography and NMR system; DmpA, L-aminopeptidase D-ala-esterase/amidase; MAD, multiple anomalous dispersion; NCS, non-crystallographic symmetry; Ntn, N-terminal nucleophile; OAT, ornithine acetyltransferase; ORF, open reading frame; PA, penicillin G acylase; SeMet, selenomethionine.

¹ These authors have contributed equally to this work.

² To whom correspondence should be addressed (email christopher.schofield@chem.ox.ac.uk).

Figure 1 The role of arginine in the biosynthesis of clavulanic acid

CEAS, N^2 -(2-carboxyethyl)arginine synthase; ThP_2 , thiamin diphosphate.

by autoproteolysis. One structurally characterized example of an as-translated Ntn enzyme is asparagine synthetase B [22]. Structurally characterized examples of autoproteolytically activated Ntn enzymes include PA (penicillin G acylase from *E. coli*) [23], cephalosporin acylase [24], glycosyl asparaginase [25], glutamine phosphoribosylpyrophosphate amidotransferase [26] and the proteasome [27]. The structures of all these enzymes show a similar, characteristic fold and have been reviewed [28].

To address the question of how an OAT achieves acetyl transfer between amino acids with acidic and basic side-chains the structure of ORF6 was determined. The structure provides the first structural information on an OAT and shows that OATs belong to a family of Ntn enzymes different from those with the characteristic fold previously identified. Unexpectedly, this fold is the same as that of an unusual DmpA (L-aminopeptidase D-alal-esterase/amidase from *Ochrobactrum anthropi*) that has been identified recently [29,30].

MATERIALS AND METHODS

Protein expression, purification and crystallization

ORF6 was prepared as described in [16]. SeMet (selenomethionine)-substituted ORF6 was produced using a metabolic inhibition method and LeMaster media supplemented with 50 mg/l L-SeMet. Complete (i.e. > 95 %) SeMet incorporation was achieved, as observed by electrospray ionization MS.

Crystallization of ORF6 (at 12 mg/ml) was accomplished by hanging-drop vapour diffusion at 17 °C. The mother liquor con-

sisted of 1.1–1.3 M ammonium sulphate, 0.04 M ammonium phosphate, 0.1 M Tris/HCl (pH 8.0) and 5–8 % (v/v) glycerol. Crystallization of SeMet-substituted ORF6 was accomplished under similar conditions but with the addition of 5 mM dithiothreitol.

Preparation of ORF6 T181A (Thr¹⁸¹ → Ala) mutant

pTYB11/orf6T181A was prepared by mutagenesis of the pTYB11/orf6 vector using the Quik Change[®] method (Stratagene). The primers used were: forward primer, 5'-CCCACATGGC-GGCGCTGCTGACCTTCTTCGCC-3'; reverse primer, 5'-AGAAGGTCAGCAGCGCCCATGTCTGGGCTC-3'. Conditions for expression and purification of the mutant enzyme were identical with that of the wild-type protein.

Crystallographic data collection and structure solution

Crystals were cryocooled by plunging into liquid nitrogen and X-ray data were collected at 100 K using a nitrogen stream. Cryo-protection was accomplished by transferring into a solution containing 2.0 M ammonium sulphate and 25 % glycerol. A three-wavelength MAD (multiple anomalous dispersion) dataset was collected to 2.8 Å resolution on beamline BM14 of the ESRF (European Synchrotron Radiation Facility, Grenoble, France) using a MarCCD detector. The Native 1 dataset was collected on beamline 14.1 of the Synchrotron Radiation Source (Daresbury, U.K.) and the Native 2 dataset on beamline ID14-2 of the ESRF, both using an ADSC Quantum 4 detector.

All data were processed with MOSFLM [31] and the CCP4 suite [32]. The crystals belonged to space group $P2_1$ or $P2_12_12_1$. The MAD data were collected on a crystal from the $P2_1$ space group. Using SHELXD [33] 28 selenium positions were located, corresponding to four molecules in the asymmetric unit. Phases were calculated with SHARP [34] and density modification, including 4-fold NCS (non-crystallographic symmetry) averaging, was performed using DM [35]. The overall figure of merit increased from 0.43 to 0.78 after density modification.

An initial model was built using O [36] and refinement against the MAD data (remote wavelength) was performed using CNS (crystallography and NMR system) [37]. One cycle of simulated annealing followed by grouped *B*-factor refinement brought the R_{free} to 37 %. The reflections (5 %) were excluded for the calculation of R_{free} . Initial refinements used a strict NCS relationship, then changing to NCS restraints, which were applied throughout the subsequent refinements. The Native 2 dataset was solved by molecular replacement. Merging statistics and the

Figure 2 The reaction catalysed by ORF6

The proposed acetyl-enzyme intermediate is shown boxed.

Table 1 Data collection and refinement statistics

Rmsd, root-mean-square deviation.

Dataset	Native 1	Se λ 1	Se λ 2	Se λ 3	Native 2
Data collection					
Spacegroup	P2 ₁ 2 ₁ 2 ₁	P2 ₁	P2 ₁	P2 ₁	P2 ₁
Wavelength (Å)	1.488	0.978363	0.978825	0.885596	0.933
Unit cell [<i>a</i> , <i>b</i> , <i>c</i> (Å), β (°)]	63.7, 94.4, 180.4, 90.0	60.7, 179.9, 73.3, 89.8	60.8, 180.3, 73.5, 89.9	60.8, 180.2, 73.3, 90.1	61.0, 180.3, 73.0, 89.9
Resolution range* (Å)	32.11–2.75 (2.90–2.75)	23.64–2.90 (3.05–2.90)	23.70–2.90 (3.06–2.90)	20.50–2.75 (2.90–2.75)	67.42–3.00 (3.16–3.00)
No. of unique observations*	29 114 (4190)	34 648 (4860)	34 357 (4576)	40 839 (5983)	29 519 (4590)
No. of total observations*	203 839 (29 743)	243 455 (20 780)	121 710 (10 748)	152 083 (22 179)	95 109 (15 709)
Completeness* (%)	99.9 (100.0)	99.1 (95.3)	98.3 (89.9)	99.8 (100.0)	93.6 (99.2)
Multiplicity*	7.0 (7.1)	7.0 (4.3)	3.5 (2.3)	3.7 (3.7)	3.2 (3.4)
<i>R</i> _{merge} * (%)	11.4 (33.5)	10.2 (28.2)	8.3 (26.3)	9.8 (31.6)	10.4 (26.9)
$\langle I/\sigma(I) \rangle$ *	5.5 (2.2)	6.8 (2.5)	8.2 (2.7)	6.9 (2.3)	4.0 (2.9)
Refinement					
<i>R</i> factor (%)	21.2	–	–	23.5	22.3
<i>R</i> _{free} (%)	22.9	–	–	28.8	31.3
Rmsd bond length (Å) [angle (°)]	0.007 (1.3)	–	–	0.008 (1.5)	0.007 (1.3)
PDB no.	1VZ6	–	–	1VZ8	1VZ7

* Numbers in parentheses represent the value in the highest resolution shell.

similar locations of disordered sections of individual molecules suggested that the MAD and Native 2 datasets were indexed comparably.

The P2₁2₁2₁ crystal form was solved by molecular replacement using CNS and the co-ordinates from the P2₁ form. There were two molecules in the asymmetric unit. Further refinement with CNS brought the conventional *R*-factor to 21.2 % and the *R*_{free} to 22.9 %. Figures were created using MolScript [38] or PyMol (http://www.pymol.org).

RESULTS AND DISCUSSION

Crystallization, structure solution and crystal packing

Rod-shaped crystals of ORF6 grew to a maximum size of 0.1 mm × 0.02 mm × 0.01 mm after 6 weeks and diffracted X-rays to 3.0 Å resolution or better. Although the presence of the ammonium phosphate was not essential for crystal growth, its presence substantially increased the reliability of the crystallization procedure.

The structure of ORF6 was solved from these crystals at 2.8 Å resolution in two different space groups (Table 1). Since all of the crystals had the same morphology, there was no visual indication of the space group. In space group P2₁2₁2₁ there were two molecules in the asymmetric unit, forming a dimer. In P2₁ there were four molecules in the asymmetric unit, arranged as two dimers related by a pure translation. Each of the dimers had an arrangement of monomers identical with that in the P2₁2₁2₁ space group. The electron density for each molecule in space group P2₁2₁2₁ was of similar quality. However, in P2₁, one dimer was of significantly better quality than the other, a difference reflected in the average *B*-factors of 26 and 24 for the monomers of one dimer and 42 and 46 for the other dimer (Se λ 3). At the end of the refinements, only residues 1–8 and 337–343 were not resolved in any model. Unless otherwise indicated, the structural discussions refer to the P2₁2₁2₁ Native 1 dataset.

Structure of ORF6

After autoproteolytic processing, each ORF6 molecule consists of two chains, A and B, which are made up of residues 1–180 and 181–393 respectively [16], and thus ORF6 is formally an $\alpha\beta$ -

Figure 3 The ORF6 $\alpha\beta$ -heterodimer

Chain A is shown in yellow and chain B in green. The Ntn is at the end of strand β 8. Blue colouring indicates the positions of the amino acids separated by autoproteolysis.

heterodimer where α corresponds to chain A and β to chain B. Both in solution and in the crystalline form, the $\alpha\beta$ -heterodimer dimerises to form $\alpha_2\beta_2$ -heterotetramers. The overall $\alpha\beta$ -heterodimer structure is in two domains: residues 1–261 form domain 1 and 262–393 form domain 2. Thus chains A and B do not define the separate physical domains. Domain 1 is the larger and consists of a four- (β 1, β 2, β 4 and β 5) and five- (β 3, β 6– β 9) membered mixed β -sheet, which are stacked on top of each other and surrounded by six α -helices (α 1– α 6) (Figure 3). The overall arrangement of domain 1 is of a stacked $\alpha\beta\alpha$ system, with two

Figure 4 Structural comparison of ORF6 (on the left) and DmpA (on the right)

For each protein, chain A is shown in blue and chain B in red.

α -helices above and two below the central stacked β -sheets, and an additional flanking α -helix on each edge of the β -sheets. Domain 2 is composed of a four-membered mixed β -sheet and five α -helices. Three of the α -helices ($\alpha 7$ – $\alpha 9$) are located on one face of the β -sheet, whereas $\alpha 10$ projects into space and is largely solvent exposed. The C-terminal α -helix, $\alpha 11$, points back towards the active site in domain 1 (see below).

Fold analysis and comparison with other Ntn hydrolases

Analysis of the fold of ORF6 using Dali [39] revealed that domain 1 was of the same fold as that of the DmpA [30]. Although the topologies of domain 1 of ORF6 and DmpA are essentially identical, it was only possible to three-dimensionally align the structures of ORF6 and DmpA such that 68 C α atoms were superimposed with an rmsd (root-mean-square deviation) of 1.44 Å. The low level of superimposition (68 residues out of 216 in domain 1) was due to substantial differences in the lengths and positions of loops connecting secondary structural elements, and relative shifts of secondary structural elements that were otherwise similar in position and direction. Visual comparison of the structures (DmpA versus domain 1 of ORF6) reveals greater similarity than suggested by the three-dimensional-superimposition statistics (Figure 4).

The greatest difference between the structures of ORF6 and DmpA is the presence of domain 2 in ORF6. The C-terminal 130 residues of ORF6 form domain 2, which begins after helix $\alpha 6$, but there are only 35 residues after the equivalent α -helix of DmpA. Whereas in ORF6 these residues are capable of forming a discrete domain, in DmpA they form a two-membered antiparallel β -sheet and a single α -helix. The loop between $\beta 7$ and $\beta 8$ (residues 171–180) in ORF6, which is broken during the auto-proteolysis (see below), is much shorter than the equivalent loop in DmpA (residues 220–245). The C-terminus of this loop in DmpA occupies a position equivalent to the overall C-terminus of ORF6. Other significant differences include a loop connecting $\beta 9$ and $\alpha 6$ in ORF6 much shorter than the equivalent loop in DmpA and a loop connecting $\beta 1$ and $\beta 2$ much shorter than the equivalent in DmpA.

Figure 5 Topology comparison of ORF6 (on the left) and PA (on the right)

○, α -helices; $\triangle$ and ∇ , β -strands. Light grey shading for ORF6 indicates chain B. Dark grey shading for PA indicates the consensus Ntn elements [28].

The similarity of the folds of ORF6 and DmpA therefore means that ORF6 and the other OATs have a fold different from that of the consensus Ntn hydrolase. As was observed for DmpA [30], although it was possible to superimpose the structures of ORF6 and the Ntn hydrolases, with both the secondary structural elements of the $\alpha\beta\alpha$ sandwich overlaid and considerable agreement between active-site residues, the directions and connectivity of the secondary structural elements differed. A topology diagram comparing ORF6 with PA as an example Ntn hydrolase illustrates the differences (Figure 5).

Owing to the differences in direction and connectivity of the secondary structural elements, it is not possible to readily align the sequences of ORF6 and the Ntn hydrolases. However, the super-impossibility of secondary structural elements, similar active-site locations and oxyanion stabilization (see below), together with similar amide-bond cleavage during both auto-proteolysis and substrate turnover nonetheless indicate a divergent evolutionary relationship.

All of the Ntn enzymes structurally characterized so far are hydrolases, whereas OATs, such as ORF6, are acyl-transferases. Although DmpA can act as a hydrolase, none of the substrates identified for DmpA were efficiently turned over [29]. It is possible that DmpA is in fact an acyl-transferase for an as yet unidentified substrate. If this were the case then, as for the OAT from *Chlamydomonas reinhardtii* [40], the hydrolysis reactions catalysed by DmpA would be an alternative or side-reaction. It is

Figure 6 Structure of the ORF6 $\alpha_2\beta_2$ -heterotetramer

One $\alpha\beta$ -heterodimer is shown as ribbons and the other as a surface. For each $\alpha\beta$ -heterodimer, domain 1 is coloured in blue and domain 2 in green.

possible that the two structural families of Ntn-enzyme now observed correspond to the two different activities of hydrolase or acyl-transferase.

ORF6 oligomer structure

The presence of α and β chains of each molecule means that the observed dimeric form of ORF6 (Figure 6) is formally an $\alpha_2\beta_2$ -heterotetramer, in agreement with studies on solutions of this enzyme [16] and on other OATs [20,41]. The intermolecular interface buries a surface area of 3980 Å² and is stabilized by a large number of hydrogen bonds. The interface is largely composed of interactions between $\alpha 8$ and $\alpha 9$ of one molecule with the β -sheet connecting loops of the other, and the interactions of the C-terminal α -helix $\alpha 11$ with $\alpha 8$ and $\alpha 11$ of the other molecule. Overall, therefore, the interface is mostly between domain 2 of one monomer and domain 1 of the other (~2800 Å²), since although the two domain 2s are close in space there is a sizeable cavity between them and a significantly lower area of interaction (~1200 Å²). The large area of interaction suggests that the dimer may be the form of ORF6 both pre and postautoproteolysis, as was shown to be the case for glycosylasparaginase [42,43]. Although a rearrangement after auto-proteolysis that would affect only residues 171–180 (the loop that would connect $\beta 7$ and $\beta 8$) is conceivable, these residues are threaded beneath the loop connecting $\alpha 5$ and $\beta 9$ (residues 208–221), which is part of the intermolecular interface, and so it is possible that a permanent or temporary change in the oligomerization interactions does occur during autoproteolysis. The CD spectra of a T181A mutant, which does not undergo the autoproteolysis reaction, and the wild-type ORF6 were essentially identical, indicating that no large-scale change in folding occurs on autoproteolysis. If the β -strand containing Thr-181 extended in its N-terminal direction, it would intersect with the intermolecular interface, although the direction is not towards the centre of gravity of the heterotetramer. This places the active site of ORF6 close to the intermolecular interface, which is thus important in defining the shape of the active site for the catalysis reaction as well as for autoproteolysis; residues from both ORF6 molecules in the heterotetramer line the sides of the active-site cavity. The position of Thr-181 close to the dimer interface contrasts with DmpA, which exists in a tetrameric (formally $\alpha_4\beta_4$ -hetero-octameric) form [30]. In DmpA, the equivalent Ser-250 is further from the intermolecular interface

and the β -strand containing Ser-250 is not directed towards the interface. However, in DmpA also, a loop from another molecule forms part of the active site, in this case the loop equivalent to that connecting $\alpha 2$ – $\beta 5$ in ORF6 (DmpA residues 132–139).

Substrate binding and self-activation

Thr-181 is at the centre of a small, deep, pocket (Figures 7a and 7b). This pocket is defined by the disconnected loop formed by residues 171–180, residues 148–149 (loop $\alpha 4$ – $\beta 6$), the side chain of Glu-260 (loop $\alpha 6$ – $\beta 10$), Asn-388 ($\alpha 11$), residues 219–220 (loop $\alpha 5$ – $\beta 9$), residues 111–114 (loop $\beta 5$ – $\alpha 3$) and also by Thr-181 itself and potentially the protein C-terminus. The side-chains of Leu-291 ($\alpha 8$) and Arg-305 ($\alpha 9$), two residues from the other molecule in the heterotetramer, complete the active-site definition.

In eight out of the ten protein chains observed, the three C-terminal residues of ORF6 (Tyr-391 to Thr-393) were not resolved. However, in two of the chains in the Se $\lambda 3$ dataset they were resolved, and observed in a position defining one side of the active-site pocket (Figure 7b). The size and shape of the active site are thus affected significantly by the presence or absence of these C-terminal residues (Figures 7c and 7d). It is possible that the movement of the C-terminus in and out of the active site enables ORF6 to accept two different substrates by altering the substrate specificity of the binding pocket (see below).

In all of the molecules examined in each structure a sulphate or phosphate anion was observed in the active site. Although the improvement in reliability of crystallization with phosphate may indicate a preferential binding site for phosphate over sulphate, this anion was modelled as sulphate due to its substantially greater concentration in the crystallization conditions. The resolution of the structures is low; nevertheless, the best fit of a sulphate to this electron density results in a very reasonable binding environment. In this model, the sulphate is bound by one oxygen atom to the backbone oxygen of Gly-112 and the side-chain oxygen of Thr-111, by another to the N-terminal nitrogen of Thr-181 and the side-chain of Lys-170 and by a third to the side-chain of Thr-148.

It is known that OATs operate by a 'ping-pong' mechanism [20], whereby one substrate binds to the enzyme, to which it transfers its acetyl group. This substrate then dissociates from the enzyme, allowing the second substrate to bind and pick up the acetyl group from the enzyme. In this mechanism, only one substrate is bound to the enzyme at a given time. Given the probable nature of the acetylated enzyme intermediate it is unlikely that there would be two entirely distinct binding sites for the two substrates. Instead, most of the binding site is probably shared and in particular a common acetyl binding position seems probable.

Structural alignment of ORF6 with DmpA (PDB no. 1B65) and PA (PDB no. 1GM9) reveals a possible component of the oxyanion stabilization that would define the probable position of the substrate acetyl group. The backbone NH of Gly-112 is similarly positioned to that of Tyr-146 of DmpA and Ala-B69 of PA. In the structure of PA with penicillin G sulphoxide [44] this backbone NH is well-positioned to stabilize the oxyanion formed during hydrolysis of the penicillin side-chain. Asn-B241 of PA, also a proposed component of the oxyanion stabilization, is functionally conserved as Asn-218 in DmpA; however there is no equivalent residue in ORF6 and indeed the volume occupied by the side-chain of Asn-B241 is empty in ORF6. Instead, in ORF6 Thr-111 probably fulfils this role.

This position for the oxyanion is supported by the position of the active-site sulphate anion bound to Gly-112, which is in a region well conserved in other OATs (Figure 8). It is reasonable that the

Figure 7 An active-site-binding pocket

(a) The active site as in chain B of the SeMet structure (PDB no. 1VZ8), where the C-terminal three residues (Tyr-391–Thr-393) are resolved. A sulphate molecule is shown in the active site. (b) Same as (a) but with a protein surface displayed. (c) A model of the substrate *N*-acetyl-ornithine in the active site. (d) A model of the substrate *N*-acetyl-glutamate in the active site when the three C-terminal residues are omitted.

position of the sulphate oxygen atom bound to Gly-112 could also be the position of the substrate acetyl oxygen atom. It was possible to model the tetrahedral transition state formed by an L-configured *N*-acetyl amino acid in the active site so that the acetyl oxygen atom overlaid the sulphate oxygen position and also so that one substrate α -carboxylate oxygen atom overlaid another sulphate oxygen position (Figure 7c). A model of the substrate in this position has also the benefit of favourable interactions for the other α -carboxylate oxygen atom with the backbone oxygen of Gly-173 and the N-terminal nitrogen of Thr-181.

ORF6 accepts alternately two substrates with differently charged but similar length side-chains, ornithine and glutamate. Assuming a common conformation for the substrate acetyl and α -carboxylate groups there are three possibilities for accommodating the different side-chains: (i) the side-chains bind in different conformations, probably accompanied by conformational changes of the protein; (ii) the side-chains bind in similar conformations but the protein undergoes some conformational change to allow different residues to bind the different substrates; and (iii) the side-chains bind in similar conformations to the same, probably predominantly neutral, protein residues which must be capable of accommodating the different charges.

Given the proposed binding positions of the substrate acetyl and α -carboxylate groups, the positive side-chain of *N*-acetyl-ornithine could be conveniently bound to the side-chains of Glu-260 and Asn-388 (Figure 7c). Alternatively, the side-chain could be bound to the C-terminal carboxylate (Thr-393).

For the *N*-acetyl-glutamate substrate neither of these interactions appears favourable. The nearest positively charged residue that could bind the glutamate would be Arg-305 (from the other monomer of the dimer), which would require only a small adjustment to be involved in substrate binding. The C-terminus of ORF6, which may be involved in binding the *N*-acetyl-ornithine substrate, may hinder Arg-305 from binding to the *N*-acetyl-glutamate substrate. Whereas the ordered/disordered state of the C-terminus may be a crystallographic artifact, it is possible that its presence/absence allows the binding of the two different substrates, by allowing access to Arg-305 for the negatively charged substrate (Figure 7d). It is therefore proposed that the side-chain of the glutamate binds in a different conformation to that of ornithine, accompanied by a change in the conformation of the protein C-terminus. This conclusion is consistent with the conservation of the C-terminal residues Tyr-391 and Thr-393 in most of the OATs (Figure 8).

Although a comparison of ORF6 with DmpA does not reveal any significant similarity in the potential ornithine side-chain-binding region (DmpA has hydrophobic residues in a similar position to ORF6 Glu-260 and Asn-388), interestingly, despite the differences in oligomerization, the hydrophobic side-chain of Phe-135 from another monomer of DmpA is in a similar position to Arg-305 of ORF6. In DmpA, Tyr-146 occupies a similar position to that of the C-terminus in ORF6. These similarities in the oligomerization determined shape of the active site might hint at a substrate-binding role for these residues in DmpA. A

Figure 8 Sequence alignment of ORF6 with seven other OATs (with accession nos)

The seven OATs were obtained from the organisms: *T. thermophilus* (P96137), *Methanococcus jannaschii* (Q57645), *Lactobacillus plantarum* (Q08319), *Thermotoga neapolitana* (Q57645), *Neisseria gonorrhoeae* (P38434), *Bacillus stearothermophilus* (Q07908), *Saccharomyces cerevisiae* (Q04728), as well as OAT1 from *S. clavuligerus*. The ORF6 secondary structure is shown above the alignment with α -helices as grey cylinders and β -sheets as black arrows. Similar amino acid residues in all of the aligned proteins are highlighted in black and those present in seven or more sequences are highlighted in grey. The conserved threonine (at position 23) that becomes the catalytic nucleophile after autoproteolysis is marked by a black triangle. The alignment was produced using ClustalW [46] and GeneDoc (<http://www.psc.edu/biomed/genedoc>).

hydrophobic cluster containing these DmpA residues (Tyr-146A, Phe-135D, Leu-136D and Trp-137D) was previously proposed to be involved in binding a hydrophobic side-chain of the substrate peptide [30], which would be consistent with kinetic studies that showed an increase in activity for a peptide with such a hydrophobic residue [29].

As discussed, the observed quaternary structure of ORF6 is most probably the form of both pre and postautoproteolysis proteins. Cleavage of the peptide bond between Ala-180 and Thr-181

breaks the loop connecting strands $\beta 7$ and $\beta 8$ (Figure 3). Given the probable binding position of the substrate acetyl group during catalysis, it is probable that the oxyanion stabilization during autoproteolysis is achieved by a mechanism somewhat different from that during catalysis. The probable position of the backbone carbonyl group of Ala-180 in the precursor peptide is occupied by the side-chain of Met-174, from the loop that is disconnected during auto-proteolysis. It is therefore a strong possibility that residues from the loop connecting $\alpha 5$ and $\beta 9$, under which the

newly created C-terminus (residues 171–180) of the α -subunit is threaded, are involved in the autoproteolysis.

The crystal structure of the precursor protein of glycosyl-asparaginase reveals that the scissile peptide bond (between Asp-151 and Thr-152) is in a strained conformation, offset 20° from an idealized *trans* conformation [42]. In glycosylasparaginase, this raised energy conformation is caused by a tight turn of the protein chain, which is assisted by the interactions made by the side-chain of Asp-151. In ORF6, the equivalent residue is Ala-180, which would not be capable of forming such interactions.

Sequence alignment of ORF6 against other OATs

A database search using three-dimensional-PSSM [45] and an ORF6 search-sequence identified more than 140 potential OATs. Figure 8 shows a sequence alignment of ORF6 with some of the better-characterized examples. The conserved autoproteolysis motif of the OATs was just one part of the generally high level of similarity throughout the length of these sequences. As mentioned above, the region containing Thr-111 and Gly-112, that may be involved in oxyanion stabilization, is highly conserved, as are the C-terminal residues. Furthermore, Lys-170, which may bind the substrate α -carboxylate, is strictly conserved. The active-site residues that may bind the substrate side-chains (Glu-260, Asn-388 and Arg-305) are also conserved, supporting their proposed role.

It is interesting that the recently identified ORF6 homologue encoded by *oat1* from *S. clavuligerus*, that is clustered with homologues of other clavulanic acid biosynthesis genes (*orf2*–*orf4*), [18] differs in key regions of the amino acid sequence compared with ORF6. OAT1 possesses a modified autoproteolysis motif (Figure 8). However, due to the conservative substitution of serine for threonine, autoproteolysis may still occur. The proposed acetyl-oxygen-binding residues are conserved in OAT1, as is Lys-170, but the proposed substrate side-chain-binding residues are not. The three C-terminal residues of ORF6 are also not conserved in OAT1 although this might not prevent their functioning in a similar role. Taken together, these observations suggest that OAT1 is either an unusual OAT or that it has an alternative function, possibly directly associated with clavam biosynthesis.

We thank the Biotechnology and Biological Sciences Research Council, the Engineering and Physical Sciences Research Council and the European Union for funding. We also thank Dr M. Walsh at BM14 of the European Synchrotron Radiation Facility for expert assistance.

REFERENCES

- Matagne, A., Dubus, A., Galleni, M. and Frère, J.-M. (1999) The β -lactamase cycle: a tale of selective pressure and bacterial ingenuity. *Nat. Prod. Rep.* **16**, 1–19
- Jensen, S. E., Alexander, D. C., Paradkar, A. S. and Aidoo, K. A. (1993) Extending the beta-lactam gene cluster in *Streptomyces clavuligerus*. In *Industrial Microorganisms: Basic and Applied Molecular Genetics* (Baltz, R. H., Hegeman, G. D. and Skatrud, P. L., eds.), pp. 169–176, American Society for Microbiology, Washington, DC, U.S.A.
- Li, R., Khaleeli, N. and Townsend, C. A. (2000) Expansion of the clavulanic acid gene cluster: identification and *in vivo* functional analysis of three new genes required for biosynthesis of clavulanic acid by *Streptomyces clavuligerus*. *J. Bacteriol.* **182**, 4087–4095
- Mellado, E., Lorenzana, L. M., Rodríguez-Sáiz, M., Díez, B., Liras, P. and Barredo, J. L. (2002) The clavulanic acid biosynthetic cluster of *Streptomyces clavuligerus*: genetic organization of the region upstream of the *car* gene. *Microbiology (Reading, U.K.)* **148**, 1427–1438
- Hodgson, J. E., Fosberry, A., Rawlinson, N. S., Ross, H. N. M., Neal, R. J., Arnell, J. C., Earl, A. J. and Lawlor, E. J. (1995) Clavulanic acid biosynthesis in *Streptomyces clavuligerus*: gene cloning and characterization. *Gene* **166**, 49–55
- Elson, S. W., Baggaley, K. H., Davidson, M., Fulston, M., Nicholson, N. H. and Risbridger, G. D. (1993) The identification of three new biosynthetic intermediates and one further enzyme in the clavulanic acid pathway. *J. Chem. Soc., Chem. Commun.* (issue 15), 1212–1214
- Khaleeli, N., Li, R. and Townsend, C. A. (1999) Origin of the β -lactam carbons in clavulanic acid from an unusual thiamine pyrophosphate-mediated reaction. *J. Am. Chem. Soc.* **121**, 9223–9224
- Caines, M. E. C., Elkins, J. M., Hewitson, K. S. and Schofield, C. J. (2004) Crystal structure and mechanistic implications of *N*²-(2-carboxyethyl)arginine synthase, the first enzyme in the clavulanic acid biosynthesis pathway. *J. Biol. Chem.* **279**, 5685–5692
- Miller, M. T., Bachmann, B. O., Townsend, C. A. and Rosenzweig, A. C. (2001) Structure of β -lactam synthetase reveals how to synthesize antibiotics instead of asparagine. *Nat. Struct. Biol.* **8**, 684–689
- Elkins, J. M., Clifton, I. J., Hernández, H., Doan, L. X., Robinson, C. V., Schofield, C. J. and Hewitson, K. S. (2002) Oligomeric structure of proclavaminic amidino hydrolase – evolution of an enzyme catalyzing intermediate engineering during antibiotic biosynthesis. *Biochem. J.* **366**, 423–434
- Zhang, Z., Ren, J., Stammers, D. K., Baldwin, J. E., Harlos, K. and Schofield, C. J. (2000) Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase. *Nat. Struct. Biol.* **7**, 127–133
- Jensen, S. E. and Paradkar, A. S. (1999) Biosynthesis and molecular genetics of clavulanic acid. *Anton. Leeuwen. Int. J.* **75**, 125–133
- Baggaley, K. H., Brown, A. G. and Schofield, C. J. (1997) Chemistry and biosynthesis of clavulanic acid and other clavams. *Nat. Prod. Rep.* **14**, 309–333
- Townsend, C. A. (2002) New reactions in clavulanic acid biosynthesis. *Curr. Opin. Chem. Biol.* **6**, 583–589
- Cunin, R., Glandsdorff, N., Piérard, A. and Stalon, V. (1986) Biosynthesis and metabolism of arginine in bacteria. *Microbiol. Rev.* **50**, 314–352
- Kershaw, N. J., McNaughton, H. J., Hewitson, K. S., Hernández, H., Griffin, J., Hughes, C., Greaves, P., Barton, B., Robinson, C. V. and Schofield, C. J. (2002) ORF6 from the clavulanic acid gene cluster of *Streptomyces clavuligerus* has ornithine acetyltransferase activity. *Eur. J. Biochem.* **269**, 2052–2059
- Rodríguez-García, A., de la Fuente, A., Pérez-Redondo, R., Martín, J. F. and Liras, P. (2000) Characterization and expression of the arginine biosynthesis gene cluster of *Streptomyces clavuligerus*. *J. Mol. Microbiol. Biotechnol.* **2**, 543–550
- Tahlan, K., Park, H. U., Wong, A., Beatty, P. H. and Jensen, S. E. (2004) Two sets of paralogous genes encode the enzymes involved in the early stages of clavulanic acid and clavam metabolite biosynthesis in *Streptomyces clavuligerus*. *Antimicrob. Agents Chemother.* **48**, 930–939
- Liu, Y. S., Heeswijck, R. V., Hoj, P. and Hooengraad, N. (1995) Purification and characterisation of ornithine acetyltransferase from *Saccharomyces cerevisiae*. *Eur. J. Biochem.* **228**, 291–296
- Marc, F., Weigel, P., Legrain, C., Almeras, Y., Santrot, M., Glandsdorff, N. and Sakanyan, Y. (2000) Characterisation and kinetic mechanism of mono- and bifunctional ornithine acetyltransferases from thermophilic microorganisms. *Eur. J. Biochem.* **267**, 5217–5226
- Marc, F., Weigel, P., Legrain, C., Glandsdorff, N. and Sakanyan, V. (2001) An invariant threonine is involved in self-catalyzed cleavage of the precursor protein for ornithine acetyltransferase. *J. Biol. Chem.* **276**, 25404–25410
- Larsen, T. M., Boehlein, S. K., Schuster, S. M., Richards, N. G. J., Thoden, J. B., Holden, H. M. and Rayment, I. (1999) Three-dimensional structure of *Escherichia coli* asparagine synthetase B: a short journey from substrate to product. *Biochemistry* **38**, 16146–16157
- Duggleby, H. J., Tolley, S. P., Hill, C. P., Dodson, E. J., Dodson, G. G. and Moody, P. C. E. (1995) Penicillin acylase has a single-amino-acid catalytic centre. *Nature (London)* **373**, 264–268
- Kim, Y., Yoon, K.-H., Khang, Y., Turley, S. and Hol, W. G. J. (2000) The 2.0 Å crystal structure of cephalosporin acylase. *Structure* **8**, 1059–1068
- Oinonen, C., Tikkanen, R., Rouvinen, J. and Peltonen, L. (1995) 3-Dimensional structure of human lysosomal aspartylglucosaminidase. *Nat. Struct. Biol.* **2**, 1102–1108
- Smith, J., Zaluzec, E., Wery, J., Niu, L., Switzer, R., Zalkin, H. and Satow, Y. (1994) Structure of the allosteric regulatory enzyme of purine biosynthesis. *Science* **264**, 1427–1433
- Seemüller, E., Lupas, A. and Baumeister, W. (1996) Autocatalytic processing of the 20 S proteasome. *Nature (London)* **382**, 468–470
- Oinonen, C. and Rouvinen, J. (2000) Structural comparison of Ntn-hydrolases. *Protein Sci.* **9**, 2329–2337
- Fanuel, L., Goffin, C., Cheggour, A., Devreese, B., Van Driessche, G., Joris, B., Van Beeumen, J. and Frère, J.-M. (1999) The DmpA aminopeptidase from *Ochrobactrum anthropi* LMG7991 is the prototype of a new terminal nucleophile hydrolase family. *Biochem. J.* **341**, 147–155

- 30 Bompard-Gilles, C., Villeret, V., Davies, G. J., Fanuel, L., Joris, B., Frère, J.-M. and Van Beeumen, J. (2000) A new variant of the Ntn hydrolase fold revealed by the crystal structure of L-aminopeptidase D-ala-esterase/amidase from *Ochrobactrum anthropi*. *Structure* **8**, 153–162
- 31 Leslie, A. G. W. (1999) Integration of macromolecular diffraction data. *Acta Crystallogr. D* **55**, 1696–1702
- 32 Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763
- 33 Schneider, T. R. and Sheldrick, G. M. (2002) Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779
- 34 De La Fortelle, E. and Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494
- 35 Cowtan, K. (1994) An automated procedure for phase improvement by density modification. In Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography, vol. 31, pp. 34–38
- 36 Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119
- 37 Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S. et al. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921
- 38 Kraulis, P. J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950
- 39 Holm, L. and Sander, C. (1997) Touring protein fold space with Dali/FSSP. *Nucleic Acids Res.* **25**, 316–319
- 40 Staub, M. and Dénes, G. (1966) Mechanism of arginine biosynthesis in *Chlamydomonas reinhardtii*. Purification and properties of ornithine acetyltransferase. *Biochim. Biophys. Acta* **128**, 82–91
- 41 Baetens, M., Legrain, C., Boyen, A. and Glansdorff, N. (1998) Genes and enzymes of the acetyl cycle of arginine biosynthesis in the extreme thermophilic bacterium *Thermus thermophilus* HB27. *Microbiology (Reading, U.K.)* **144**, 479–492
- 42 Xu, Q., Buckley, D., Guan, C. and Guo, H.-C. (1999) Structural insights into the mechanism of intramolecular proteolysis. *Cell (Cambridge, Mass.)* **98**, 651–661
- 43 Wang, Y. and Guo, H.-C. (2003) Two-step dimerization for autoproteolysis to activate glycosylasparaginase. *J. Biol. Chem.* **278**, 3210–3219
- 44 McVey, C. E., Walsh, M. A., Dodson, G. G., Wilson, K. S. and Brannigan, J. A. (2001) Crystal structures of penicillin acylase enzyme-substrate complexes: structural insights into the catalytic mechanism. *J. Mol. Biol.* **313**, 139–150
- 45 Kelley, L. A., MacCallum, R. M. and Sternberg, M. J. E. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.* **299**, 499–520
- 46 Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680

Received 17 May 2004/17 August 2004; accepted 7 September 2004

Published as BJ Immediate Publication 7 September 2004, DOI 10.1042/BJ20040814