Author for correspondence:
Daan M. F. van Aalten
e-mail: dmfvanaalten@dundee.ac.uk

†These authors contributed equally to this study.
‡Present address: Structural Biology Group, The Beatson Institute for Cancer Research, Garscube Estate, Switchback Road, Bearsden, Glasgow G61 1BD, UK.

Royal Society Publishing
Informing the science of the future

Structure of a bacterial putative acetyltransferase defines the fold of the human *O*-GlcNAcase C-terminal domain

Francesco V. Rao[1,†], Alexander W. Schüttelkopf[1,†,‡], Helge C. Dorfmueller[1], Andrew T. Ferenbach[1], Iva Navratilova[2] and Daan M. F. van Aalten[1,3]

[1]Division of Molecular Microbiology, [2]Division of Biological Chemistry and Drug Discovery, and [3]MRC Protein Phosphorylation and Ubiquitylation Unit, College of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, UK

1. Summary

The dynamic modification of proteins by *O*-linked *N*-acetylglucosamine (*O*-GlcNAc) is an essential posttranslational modification present in higher eukaryotes. Removal of *O*-GlcNAc is catalysed by *O*-GlcNAcase, a multi-domain enzyme that has been reported to be bifunctional, possessing both glycoside hydrolase and histone acetyltransferase (AT) activity. Insights into the mechanism, protein substrate recognition and inhibition of the hydrolase domain of human OGA (hOGA) have been obtained via the use of the structures of bacterial homologues. However, the molecular basis of AT activity of OGA, which has only been reported *in vitro*, is not presently understood. Here, we describe the crystal structure of a putative acetyltransferase (*Og*pAT) that we identified in the genome of the marine bacterium *Oceanicola granulosus*, showing homology to the hOGA C-terminal AT domain (hOGA-AT). The structure of *Og*pAT in complex with acetyl coenzyme A (AcCoA) reveals that, by homology modelling, hOGA-AT adopts a variant AT fold with a unique loop creating a deep tunnel. The structures, together with mutagenesis and surface plasmon resonance data, reveal that while the bacterial *Og*pAT binds AcCoA, the hOGA-AT does not, as explained by the lack of key residues normally required to bind AcCoA. Thus, the C-terminal domain of hOGA is a catalytically incompetent 'pseudo'-AT.

2. Introduction

Since its discovery three decades ago [1], it has become clear that modification of serine and threonine residues by a single *O*-linked *N*-acetylglucosamine (*O*-GlcNAc) is an essential, abundant and dynamic posttranslational process (reviewed in [2,3]). *O*-GlcNAc-modified proteins have been detected in both the nucleus and cytoplasm of higher eukaryotes [1,4] and it was shown that the levels of *O*-GlcNAc respond to nutrient levels and cellular stress [5]. *O*-GlcNAc shows extensive crosstalk with protein phosphorylation [6]. Modification by *O*-GlcNAc has been detected in hundreds of proteins [7], many of which play key roles in cellular processes. For instance, precise *O*-GlcNAc levels at specific sites on insulin receptor substrate 1 (IRS-1), protein kinase Bβ,

glycogen synthase kinase 3β and glycogen synthase are required for proper insulin sensitivity and response [8]. *O*-GlcNAc modification of transcription factors, such as c-Myc [9] and mSin3A [10], directly affects their activity [5]. Recently, it was shown that histones are dynamically modified with *O*-GlcNAc in the nucleosome core, suggesting that *O*-GlcNAc may be part of the histone code [11].

Dynamic protein *O*-GlcNAcylation is achieved by the interplay of two essential enzymes; *O*-GlcNAc transferase (OGT) and *O*-GlcNAcase (OGA) [12–14]. Both enzymes are required for life in the metazoan cell and are highly conserved from *Caenorhabditis elegans* to man [7,15]. The N-terminus of OGT possesses multiple tetratricopeptide repeats motifs that have been shown to be essential for recognition of large protein substrates [13,16,17]. Recent studies have reported the structure of a bacterial OGT homologue [18,19] and the structure of hOGT [20], and two different reaction mechanisms have recently been proposed [21,22].

Human OGA (hOGA) is a 92 kDa multi-domain protein, originally identified as an antigen expressed by meningiomas (MGEA5) [14,23,24]. Bioinformatic [25] and biochemical [26,27] studies have suggested that OGA possesses dual catalytic activity. The N-terminal portion of the enzyme recognizes and hydrolyses *O*-GlcNAc-modified peptides/ proteins [26,28] and belongs to the CAZy glycoside hydrolase family 84 (GH84) [29]. The use of structural and biochemical characterization of close bacterial homologues has helped our understanding of how the N-terminal domain of eukaryotic OGA would recognize and process *O*-GlcNAc substrates and has stimulated the development of a number of potent hOGA inhibitors [30–35]. Furthermore, bioinformatic analysis has suggested that the C-terminal domain of hOGA can adopt a GCN5 family acetyltransferase (AT)-like fold [28]. Toleman *et al.* [27] have reported *in vitro* histone AT activity for the C-terminal hOGA domain (hOGA-AT) purified using a mammalian expression system; however, activity for protein purified from bacteria was only observed after incubation with mammalian cell lysate [27,36]. These data prompted the authors to rename hOGA to nuclear cytoplasmic *O*-GlcNAcase and acetyl transferase [27]. The OGA and AT activities have been suggested to act synergistically, opening up the chromatin and directly activating transcription factors [27]. However, a report by Butkinaree *et al.* [37] casts doubt on this, as the authors were unable to reproduce histone AT activity.

There is currently no crystal structure of any eukaryotic OGA. It is thus unknown how hOGA-AT recognizes acetyl-CoA (AcCoA) or possible protein substrates and how this domain is positioned relative to the glycoside hydrolase domain. In addition, the amino acids that are involved in catalysis have not been identified. These interactions are generally poorly conserved within the GCN5 family, which makes predictions from bioinformatics or related structures challenging. Here, using X-ray crystallography and structure-guided mutagenesis coupled with surface plasmon resonance (SPR), we provide the first molecular insights into the structure of hOGA-AT by the use of a close bacterial homologue. Our data reveal that while the bacterial homologue binds AcCoA, hOGA-AT does not bind AcCoA owing to amino acid substitutions in the binding site. These results suggest that the C-terminal domain of OGA is a pseudo-AT (pAT) not capable of catalysing acetyl transfer onto a histone substrate.

3. Results and discussion

3.1. The marine bacterium *Oceanicola granulosus* possesses an operon-containing OGA-AT-like proteins

Despite the efforts by several research groups, full-length and truncated metazoan OGA has so far resisted protein crystallization. Bacterial homologues have previously been used to gain insights into the structure, mechanism and substrate recognition of the metazoan OGA-GH84 catalytic domain [30,32,34]. Of particular interest is a GH84 from the marine bacterium *Oceanicola granulosus*, *Og*OGA, which was recently crystallized [34], as it shows higher sequence identity to hOGA when compared with other bacterial OGA homologues, with sequence conservation extending beyond the catalytic core revealing a conserved peptide-binding groove [34]. Strikingly, close inspection of the *Og*OGA genomic location reveals an open reading frame coding for a predicted AT [25,38] immediately downstream of the *Og*OGA gene (figure 1*a*). Sequence alignment of this predicted AT from *O. granulosus* (*Og*pAT) with the hOGA-AT domain (hOGA-AT) shows good similarity (30% sequence identity; figure 1*b*). Furthermore, secondary structure predictions for hOGA-AT and *Og*pAT using Jpred [40] support that these two domains are structurally similar (figure 1*b*). Overall, the genomic organization of *Og*OGA and *Og*pAT bears remarkable similarity to the domain arrangement in hOGA. The biological functions of *Og*OGA and *Og*pAT in *O. granulosus* are presently unknown and reversible intracellular *O*-GlcNAc modification of proteins has not been detected in bacteria.

3.2. *Og*pAT possesses an acetyltransferase-like fold with a conserved AcCoA-binding pocket

To gain insights into the structure and function of hOGA-AT, we selected the bacterial *Og*pAT for structural studies. The gene for *Og*pAT was cloned into pGEX-6P-1 for expression as a GST (glutathione transferase) fusion in *Escherichia coli*. The protein was purified using glutathione affinity and crystallized from ammonium sulfate solutions (see Material and methods). Crystals of *Og*pAT in complex with AcCoA formed in space group P3$_2$21 with one molecule in the asymmetric unit and synchrotron diffraction data were collected to 1.8 Å resolution (table 1). The *Og*pAT structure was solved using experimental phases from a tungsten derivative and refined to $R = 19.9\%/R_{\text{free}} = 23.4\%$ with good stereochemistry (table 1). The crystal packing suggests that *Og*pAT is a monomer in solution.

The structure of *Og*pAT consists of mixed α/β-fold, composed of a central seven-stranded twisted β-sheet (topology 1234576) sandwiched between six α-helices (two packed on one side of the β-core and four on the other; figure 1*c*). The β-sheet comprises mostly antiparallel strands except for strands four and five, which are parallel, resulting in a V-shaped wedge appearance, a feature shared with other acetyl transferases (figure 1*c*) [43,45]. The structure reveals a typical conserved AcCoA-binding region composed of one α-helix (α5) and five β-strands (β1–5) as shown in figure 1*c*. Despite low sequence homology, the overall fold identifies *Og*pAT as a member of the GCN5-related *N*-AT (GNAT)
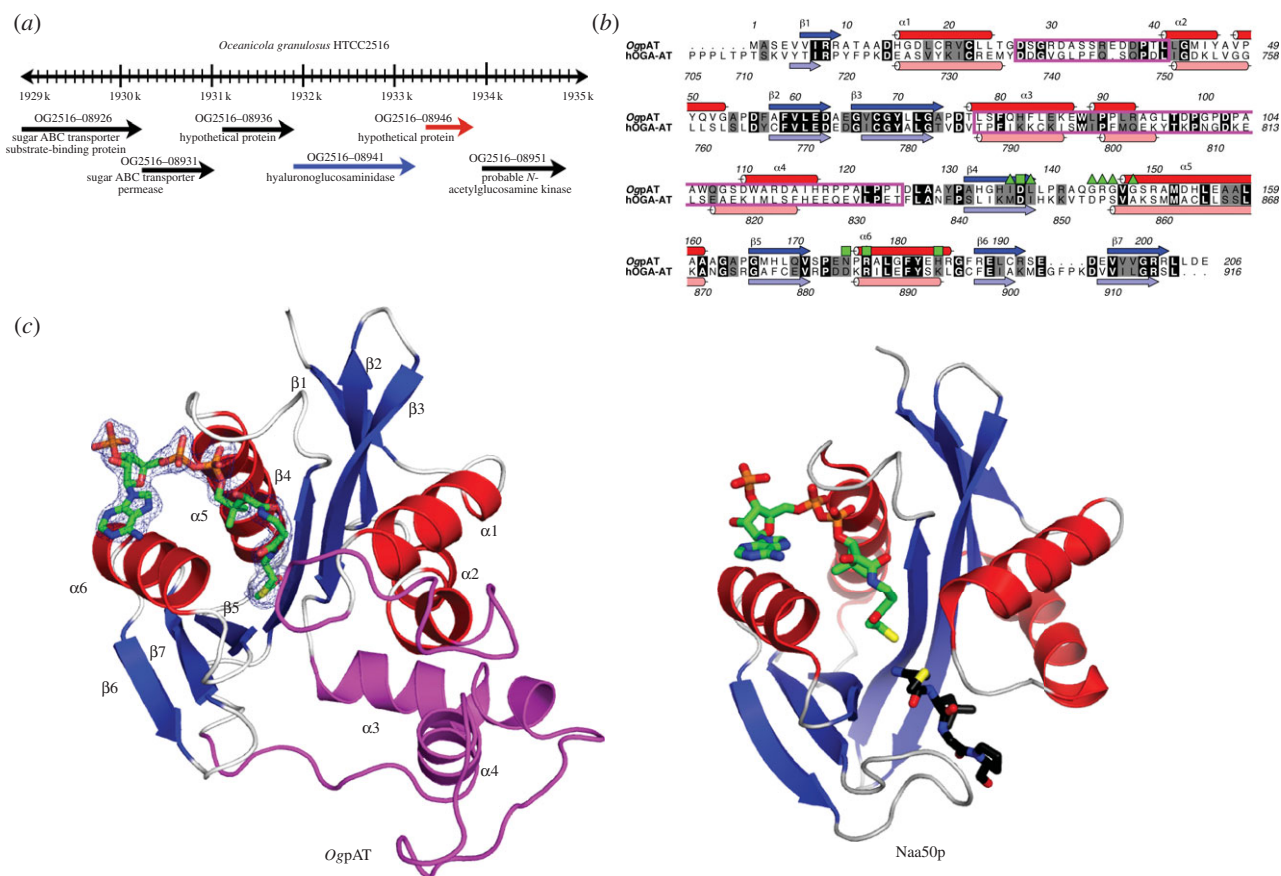
Figure 1. (a) Location of OgOGA (blue) and OgpAT (red) in the Oceanicola granulosus genome. (b) Sequence alignment of OgpAT and hOGA-AT. Identical residues are depicted in black. Secondary structure (calculated using DSSP [39]) for OgpAT is shown in blue and red for β-strands and α-helices, respectively. Predicted secondary structure elements (calculated using JPred [40] for hOGA-AT are shown in light blue and pink for β-strands and α-helices, respectively. AcCoA-interacting residues of OgpAT are indicated by green squares (interaction involves side chains) or green triangles (interaction involves backbone only). The two magenta boxes represent the two insertions when compared to sequences from other GNAT members. Numbering of the sequences are in accordance with their UniProt entries. Sequences were aligned with ClustaW [41] and annotated using the program ALINE [42]. (c) Cartoon view (colour based on secondary structure) of OgpAT in complex with AcCoA and Naa50p, an N-terminal AT (PDB code 3TFY [43]) in complex with CoA (green carbon atoms). The Naa50P acceptor peptide is shown with black carbon atoms. The two insertion regions in OgpAT are depicted in magenta. The unbiased $|F_o|-|F_c|$, ϕ_{calc} electron density map for AcCoA is shown in cyan, contoured at 2.5σ.

superfamily [46]. Members of the GNAT family share a conserved mechanism of catalysis involving an active site carboxylate (reviewed in [47]). Structural homology searches using SSM [48] reveals several members of the GNAT family that share structural features with OgpAT. Two of the most similar structures are the amino terminal AT Naa50p and RimI [43,45], with a r.m.s. deviation of Cα atoms of 2.2 and 2.0 Å (for 137 and 132 residues), respectively. The structures of OgpAT and Naa50p are compared in figure 1c, showing that much of the α/β core is conserved. However, the greatest region of divergence between the two proteins is that OgpAT contains an elongation of the loop between α1 and α2 (residues 27–41) and two additional helices (α3 and α4) are inserted between strands β3 and β4 (residues 78–125; figure 1c). These insertions create a narrow tunnel-like structure above the presumed catalytic site of OgpAT, which is likely to determine the acceptor specificity of the protein. These two insertions are not present in any of the known homologous GNAT structures, and thus appear unique to the OgpAT structure. Sequence comparison with other eukaryotic OGA-AT domain shows that this insertion is also present (figure 1b).

Inspection of the putative AcCoA-binding site revealed well-defined $|F_o|-|F_c|$, ϕ_{calc} electron density for the ligand, allowing building and refinement of the complete AcCoA

molecule (figures 1c and 2). The interactions between AcCoA and OgpAT are similar to those observed throughout the GNAT superfamily [43]. The adenosine moiety of AcCoA is located on the OgpAT surface and stacks against helix α6, while the ribose and 3'-phosphate project into the solvent (figure 2). The 3'-phosphate forms a hydrogen bond interaction with the side chain of His184 located at the end of helix α6. The pyrophosphate and pantetheine moieties form a series of both direct and water-mediated hydrogen bonds to the protein (figure 2). The most conserved interactions between the protein and AcCoA involve the 'P-loop' motif [38], which resides at the beginning of helix α5 (figures 1c and 2). The 'P-loop', which is conserved within the GNAT AT 'motif A' [10,38], is crucial for the recognition of the AcCoA pyrophosphate group in all ATs and consists of a conserved sequence [Gln/Arg]-x-x-Gly-x-[Gly/Ala]. In the OgpAT-AcCoA complex, the 'P-loop' is located at the start of helix α5 (figures 1b,c and 2) and consists of residues 143–148 (sequence Gln-Gly-Arg-Gly-Val-Gly). These residues, together with water molecules, form a network of hydrogen bonds with the pyrophosphate group of AcCoA (figure 2).

The thioester group of AcCoA is sandwiched between strands β4 and β5. The side chain of Asn175, which is conserved among the GNAT members, donates a hydrogen

Table 1. *Og*pAT X-ray diffraction data collection and refinement statistics. Values in parentheses pertain to the highest resolution shell of about 0.1 Å. Ramachandran plot values were obtained from PROCHECK [44].

	*Og*pAT-AcCoA complex
space group	$P3_221$
cell dimensions	$a = b = 65.06$ Å, $c = 92.01$ Å
wavelength	1.85 Å
no. of reflections	21060
resolution (Å)	$20.00 - 1.80$
R_{merge}	0.043 (0.424)
$I/\sigma(I)$	16.8 (3.5)
completeness (%)	98.1 (97.9)
redundancy	4.9 (4.8)
R_{work}/R_{free} (%)	19.9/23.4
protein residues	202
ligand molecules	001
solvent molecules	139
$\langle B \rangle (\text{Å}^2)$:	
protein	29.9
ligand	39.1
solvent	40.2
r.m.s. deviations:	
bond lengths (Å)	0.014
bond angles (°)	1.54
Ramachandran fit:	
most favoured (%)	91.4
allowed (%)	08.6
disallowed (%)	00.0

bond to the amide oxygen of the CoA pantothenate/cysteamine link, while the NH of the same amide forms a hydrogen bond with the side chain of Asp137 (figure 2). The acetyl group of AcCoA forms a hydrogen bond with the backbone nitrogen of Ile136 (figure 2). This type of interaction, where the carbonyl of the acetyl group is hydrogen bonded to the main-chain amide nitrogen of a residue downstream of the β-bulge on β4, has been observed in complexes of other GNAT proteins such as Naa50p (figure 2). The interactions between β4/β5 and coenzyme A are a distinctive characteristic of GNAT proteins and are an essential structural feature in promoting acetyl transfer. Thus, *Og*pAT is capable of binding AcCoA and sequence conservation in the active site suggests this as an active enzyme with an as yet unidentified acceptor substrate.

3.3. hOGA-acetyltransferase is a pseudo-acetyltransferase lacking key residues for catalytic activity and AcCoA binding

Circular dichroism experiments revealed that the secondary structure composition of hOGA-AT, purified by overexpression in *E. coli*, is very similar to that calculated from the *Og*pAT structure (see electronic supplementary material,

figure S1). This, combined with the sequence similarity between *Og*pAT and hOGA-AT (30% identity and 45% similarity, figure 1*b*), allowed construction of a homology model of hOGA-AT using SWISS-MODEL [49]. By sequence and structural similarity to *Og*pAT, hOGA-AT is likely to possess the two unique insertions, thus creating a similar deep-binding pocket (figures 1*c* and 2).

Close inspection of the hOGA-AT model reveals a number of key differences from the *Og*pAT-AcCoA complex and other members of the GNAT family (figure 2). Most members of the GNAT family contain a 'P-loop' [38], located at the N-terminal end of an α-helix such that the helix dipole supports binding of the negatively charged pyrophosphate of AcCoA. The sequence alignment and model of hOGA-AT reveal that the AT domain does not possess the 'P-loop' consensus sequence (figure 1*b*), instead a negatively charged aspartic acid (Asp853) is placed in close proximity to the pyrophosphate-binding site of AcCoA (figure 2). Furthermore, the steric constraints imposed by a proline residue (Pro854) are likely to impede the interaction between the hOGA-AT 'P-loop' and the AcCoA pyrophosphate (figure 2). This amino acid substitution is not present in any known active member of the GNAT family [38]. In addition, hOGA-AT Met735 (Thr25 in *Og*pAT) may clash with the pantothenate moiety of AcCoA. Taken together, the structural data and model suggest that hOGA-AT is unlikely to bind AcCoA.

Two distinct reaction mechanisms have been proposed for enzymes of the GNAT family [38,50,51]. The first mechanism involves an active site base that deprotonates the substrate amino group resulting in nucleophilic attack on AcCoA. The acetyl group is transferred from the thioester of AcCoA to the target amino group and an active site acid then donates a proton to the sulfur atom of CoA to bring the reaction to completion [52]. The second catalytic mechanism involves a covalent acetyl-enzyme intermediate via a conserved cysteine [53,54]. Analysis of the *Og*pAT active site and comparison with the ternary complex of Naa50p [43] reveals that His135 is in a position to function in catalysis (figure 2). Consistent with the first reaction mechanism, the imidazole group of this residue may participate in proton extraction from the acceptor amine group of the substrate. Subsequently, the uncharged amino group can perform a nucleophilic attack on the carbonyl carbon of the thioester group of AcCoA (figure 2). The residue corresponding to His135 of *Og*pAT, Lys844 in OGA-AT, could possibly act as a base in catalysis, though in the absence of an appropriately activating environment this seems unlikely, this lysine residue appears to be only conserved in OGA-AT of higher eukaryotes. However, inspection of the hOGA-AT model reveals that Glu879 may be a structural equivalent of His112, the general base in the Naa50p structure (figure 2).

It has previously been proposed that Tyr891 acts as a catalytic base and Asp853 or Asp884 as the catalytic acid for the hOGA-AT [27]. From the structure of *Og*pAT, the equivalent residues (Tyr182, Gly144 and Asn175, respectively) are not in a position to participate in catalysis. Tyr182 is more than 9 Å away from any potential acceptor substrate, Gly144 forms part of the conserved 'P-loop' and Asn175 forms a hydrogen bond with the pantothenate moiety of AcCoA (figure 2).

SPR was used to investigate *Og*pAT and hOGA-AT interactions with AcCoA. As expected from the crystallographic complex, wild-type *Og*pAT binds AcCoA ($K_d = 8.7$ μM, figure 3 and table 2). By contrast, wild-type hOGA-AT did
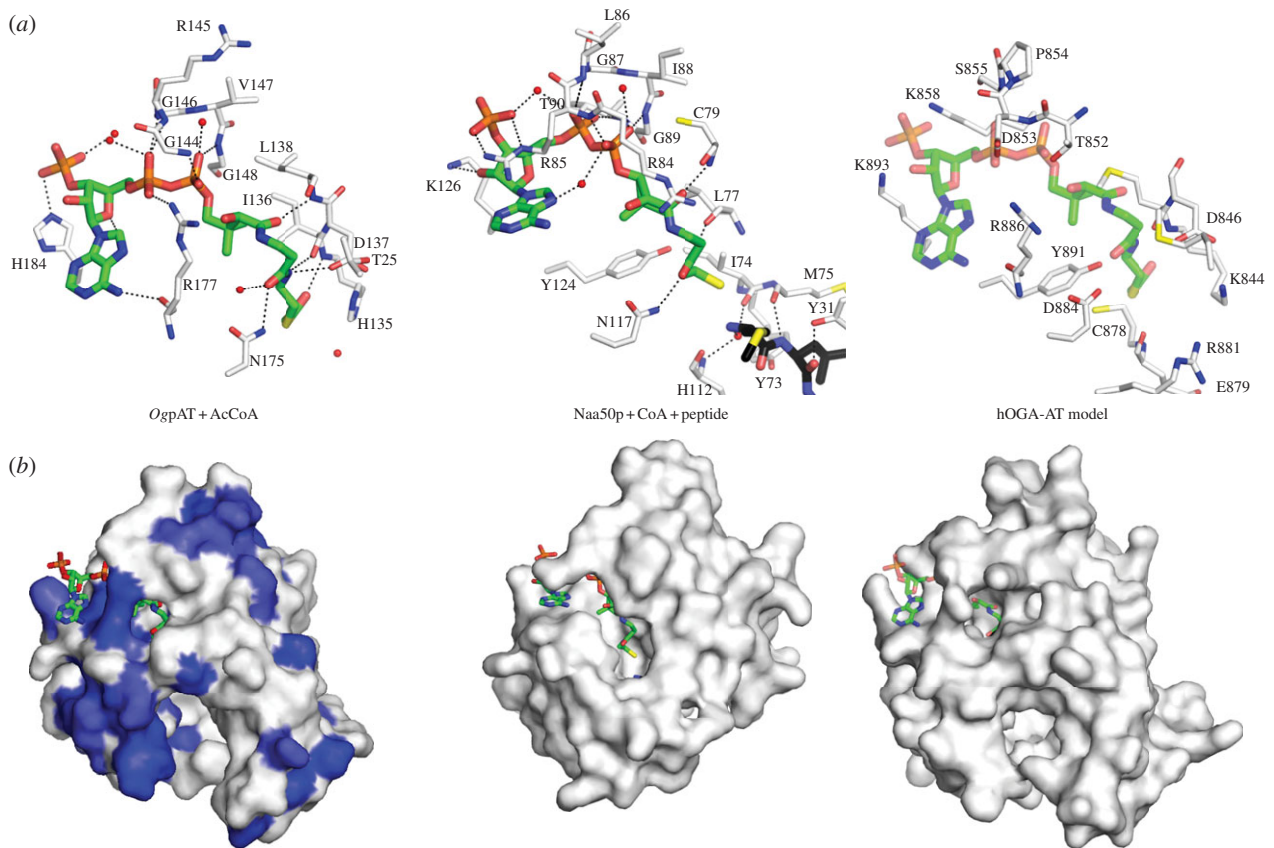
Figure 2. The *Og*pAT-AcCoA complex is compared with the Naa50p-CoA-peptide complex [43] and a model of hOGA-AT in complex with a superimposed AcCoA, using stick model views of the active site (*a*) and molecular surfaces (*b*). CoA/AcCoA are shown as sticks with green carbon atoms. Red spheres represent water molecules. Hydrogen bonds are shown as black dotted lines. The *Og*pAT surface is coloured by similarity to hOGA-AT (identical residues = dark blue; chemically similar residues = light blue, figure 1*b*).

not show any detectable binding of AcCoA. Additional SPR experiments with butyryl-CoA, decanoyl-CoA and CoA again revealed no detectable binding to hOGA-AT (table 2) in agreement with the structural analysis of the hOGA-AT model. Furthermore, mass spectrometric experiments showed that hOGA-AT is not purified as a complex with AcCoA from *E. coli*, which would have precluded the detection of AcCoA binding by SPR (see electronic supplementary material, figure S2). To investigate whether the unusual hOGA-AT 'P-loop' is compatible with AcCoA binding, key P-loop residues in *Og*pAT (Gly144, Arg145, Gly146) were mutated to the corresponding residues in hOGA-AT (Asp, Pro, Ser) (figures 1*b* and 2). No AcCoA binding could be detected with this *Og*pAT mutant (table 2). The inverse experiment of mutating Asp853, Pro854, Ser855 in hOGA-AT to the *Og*pAT equivalent (Gly, Arg, Gly) resulted in insoluble protein.

acids such as Met735 that may impede binding of AcCoA. In addition, the catalytic residues proposed by Kudlow *et al.* [55] do not appear to be in a suitable location to participate in catalysis. The data presented here are not compatible with the original report attributing histone AT activity to this hOGA domain. It is possible that the authors have purified a different protein from BSC-40 cells or a contaminant that had AcCoA transferase activity [27]. This may explain why bacterially expressed hOGA-AT, purified by the same authors, does not show enzymatic activity, which they could only show upon incubation with mammalian cell lysate. The function of the hOGA-AT domain remains unresolved, therefore this domain can be classified as a pseudo-histone AT. The AT domain of OGA is conserved from *Drosophila* to human, suggesting a functional role. Such roles might be simply protein stability of the OGA domain or binding to peptide target sequences to aid localization.

4. Conclusion

The data presented here give the first detailed structural insights into a putative bacterial AT (*Og*pAT) with significant sequence homology to hOGA-AT. The crystal structure of *Og*pAT in complex with AcCoA reveals key amino acids necessary for cofactor binding and gives insights into catalytic residues conserved in other active members of the GNAT family. Based on the *Og*pAT structure, a model of hOGA-AT was constructed and key amino acids were identified. The hOGA-AT model reveals a missing 'P-loop' and key amino

5. Material and methods

5.1. Cloning, expression and crystallization

A construct encoding hOGA-AT (residues 698–916) was amplified by PCR using gDNA (Sanger Institute, Cambridgeshire). The PCR-product was ligated into the pCR-Blunt II-TOPO (Invitrogen) and subcloned into a modified pET15b plasmid (encoding a PreScission protease (PP) cleavage site instead of the original thrombin site) using the *Nde*I and *Xho*I restriction sites. Site-directed mutagenesis for the
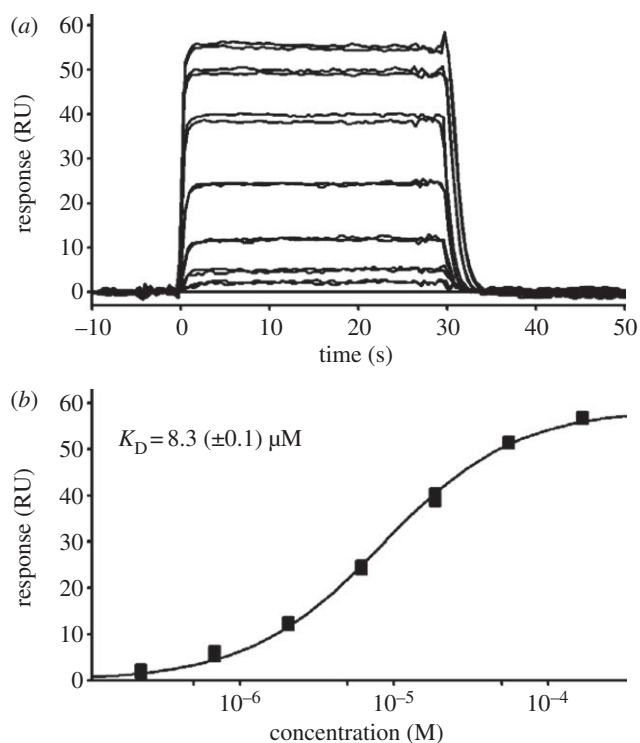
Figure 3. Sensorgram for binding of AcCoA to wild-type *Og*pAT. AcCoA was injected in duplicates at 7 concentrations (0.2, 0.7, 2.1, 6.2, 18.5, 55.5 and 166.7 μM). Equilibrium affinity fit is shown in (*b*) of the figure. RU, relative units.

Table 2. K_D determination by SPR. n.b denotes no detectable binding.

	*Og*pAT, wt	*Og*pAT, mutant	*h*OGA-AT
AcCoA	8.3 μM	n.b	n.b
butyryl-CoA	—	—	n.b
decanoyl-CoA	—	—	n.b
CoA	—	—	n.b

DNA encoding full-length *Og*pAT was PCR-amplified from genomic DNA using KOD DNA polymerase and then ligated into pGEX-6P-1 cut with *Bam*HI and *Eco*RI. For protein expression, the resulting plasmid was transformed into *E. coli* BL21(DE3) pLysS. Cultures were grown to an OD_{600} of approximately 0.7 at 37°C; after induction with 0.25 mM IPTG they were grown at 20°C overnight before harvesting by centrifugation. Cells were resuspended in lysis buffer (25 mM Tris, pH 7.5; 250 mM NaCl) supplemented with lysozyme and DNAse and lysed by sonication; cell debris and unbroken cells were removed by centrifugation, then the cell lysate was incubated with glutathione-sepharose beads for 1.5 h. After extensive washing with lysis buffer, the fusion protein was cleaved on the beads by incubation with GST-tagged PreScission protease at 4°C for 48 h. Cleaved *Og*pAT was eluted with lysis buffer and concentrated before loading onto a Superdex 75, 26/60 column equilibrated in lysis buffer. Appropriate gel filtration fractions were pooled; the protein was then assessed for purity by SDS-PAGE and concentrated to 3.8 mg ml^{-1}.

hOGA-AT mutants was performed using the QuikChange method (Stratagene) using standard protocols. DNA constructs were verified by DNA sequencing (The Sequencing Service, College of Life Sciences, University of Dundee, Scotland, UK). hOGA-AT-pET15bPP constructs were transformed into *E. coli* ArcticExpress competent cells (Stratagene). Cells were grown overnight at 37°C in Luria-Bertani (LB) medium containing 50 μg ml^{-1} ampicillin. Ten millilitres of the overnight culture was used for inoculation of 1 l LB autoinduction medium and grown at 30°C to reach an OD_{600} of 0.6. The temperature was then reduced to 12°C and cells were grown for 96 h.

The culture harvested by centrifugation for 30 min at 3500 r.p.m. (4°C) and the pellet from 1 l culture was resuspended in lysis buffer A (25 mM Tris–HCl, pH 8.5, 200 mM NaCl, 5 mM DTT) supplemented with protease inhibitors (PMSF (1 mM), leupeptin (0.2 μM) and benzamidine (1 mM)), lysozyme and DNase. The cell pellets were lysed with a constant cell disrupter (three passes at 20 kpsi) and the lysate was cleared by centrifugation (30 min, 20 000 r.p.m., 4°C). The resulting supernatant was passed through a 0.45 μm filter, and loaded onto a 5 ml His-Trap HP column (GE Healthcare) charged with $NiSO_4$. The column was washed with 10 column volumes of the same buffer, and subsequently the recombinant protein was eluted applying a linear imidazole gradient (0–500 mM imidazole) over 20 column volumes. Late elution fractions were pooled and buffer exchanged by dialysis into buffer A. The N-terminal His$_6$ tag was cleaved overnight by PreScission protease followed by a second round of nickel affinity purification. The resulting solution was concentrated to 5 ml and loaded onto a Superdex 75, 26/60 gel filtration column pre-equilibrated in buffer A. Pure fractions were verified by SDS-PAGE, pooled and spin concentrated using a 10 000 MWCO concentrator.

5.2. *Og*pAT data collection, structure solution and refinement

Crystals of *Og*pAT in complex with AcCoA (1 mM) were grown by hanging drop vapour diffusion over a reservoir of 0.1 M Tris, pH 8.5; 2.15 M $(NH_4)SO_4$; 0.24% HCl_{conc} at 20°C. For phasing, apo-*Og*pAT crystals were soaked by addition of 20 mM $(NH_4)WS_4$ to an equilibrated microdrop.

Crystals were cryoprotected by soaking in mother liquor supplemented with 20% ethylene glycol before being flash-frozen for data collection at 100 K. All data were processed/scaled with Denzo/Scalepack [56], and further handled with CCP4 software [57]. Data for a $WS_4{}^{2-}$ derivative were collected to 1.85 Å and phased using SHELXC/D/E [58] through HKL2MAP [59]. Automated model building with warpNtrace [60] generated a nearly complete model covering 200 of 206 residues. After rebuilding in Coot [61] and refinement with REFMAC5 [62], this model was used for molecular replacement with the AcCoA complex data, which were collected to 1.80 Å. Refinement was initiated immediately revealing well-defined $|F_o|-|F_c|$, ϕ_{calc} electron density for the ligand, which was built in with the help of PRODRG-generated coordinates and topology [63]. Through rounds of model building in Coot and refinement with REFMAC5 the model was improved to a final R_{free} value of 23.4% and validated in Coot and PROCHECK [44]. For complete data and model statistics, see table 1. All figures were produced with PyMOL [64].

5.3. Surface plasmon resonance

SPR measurements were collected using Biacore 3000 instrument (GE Healthcare). *Og*pAT (WT, M1) was biotinylated by mixing of the protein with amine-binding biotin (Pierce) in 1 : 1 molar ratio [65]. Streptavidin was immobilized on a CM5 sensor chip surface by amine coupling. A total of 10 mM Hepes, 150 mM NaCl, pH 7.4 was used as a running buffer for immobilization. The surface was activated by 15 min injection of NHS/EDC followed by injection of SA in 10 mM acetate, pH 4.5 until the required density (approx. 9000 relative units (RU)) was achieved and blocked by 4 min ethanolamine injection at 10 µl min^{-1} at 25°C. Biotinylated *Og*pAT was captured on the streptavidin surface at approximately 2500−4000 RU in running buffer containing 25 mM Tris (pH 7.5), 150 mM NaCl, 1 mM DTT and 0.005% Tween 20. AcCoA was injected in duplicates at threefold concentration series in a range of concentrations 0.2−166.6 µM. Association was measured for 30 s and dissociation for 1 min. All experiments were run at 50 µl min^{-1} at 25°C. All data were referenced for blocked streptavidin surface and blank injections of buffer. Scrubber 2 (BioLogic Software) was used to process and analyse data. Affinities were calculated using 1 : 1 equilibrium-binding fit.

References

1. Torres CR, Hart GW. 1984 Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *J. Biol. Chem.* **259**, 3308–3317.

2. Love DC, Hanover JA. 2005 The hexosamine signaling pathway: deciphering the O-GlcNAc code. *Science's STKE* **2005**, re13.

3. Dias WB, Hart GW. 2007 O-GlcNAc modification in diabetes and Alzheimer's disease. *Mol. BioSyst.* **3**, 766–772. (doi:10.1039/b704905f)

4. Holt GD, Hart GW. 1986 The subcellular distribution of terminal N-acetylglucosamine moieties. Localization of a novel protein-saccharide linkage, O-linked GlcNAc. *J. Biol. Chem.* **261**, 8049–8057.

5. Slawson C et al. 2005 Perturbations in O-linked beta-N-acetylglucosamine protein modification cause severe defects in mitotic progression and cytokinesis. *J. Biol. Chem.* **280**, 32 944–32 956. (doi:10.1074/jbc.M503396200)

6. Wang Z, Gucek M, Hart GW. 2008 Cross-talk between GlcNAcylation and phosphorylation: site-specific phosphorylation dynamics in response to globally elevated O-GlcNAc. *Proc. Natl Acad. Sci. USA* **105**, 13 793–13 798. (doi:10.1073/pnas.0806216105)

7. Zachara NE, Hart GW. 2004 O-GlcNAc a sensor of cellular state: the role of nucleocytoplasmic glycosylation in modulating cellular function in response to nutrition and stress. *Biochim. Biophys. Acta* **1673**, 13–28. (doi:10.1016/j.bbagen.2004.03.016)

8. Parker GJ et al. 2003 Insulin resistance of glycogen synthase mediated by o-linked N-acetylglucosamine. *J. Biol. Chem.* **278**, 10 022–10 027. (doi:10.1074/jbc.M207787200)

9. Chou TY, Hart GW, Dang CV. 1995 c-Myc is glycosylated at threonine 58, a known phosphorylation site and a mutational hot spot in lymphomas. *J. Biol. Chem.* **270**, 18 961–18 965. (doi:10.1074/jbc.270.32.18961)

10. Yang X, Zhang F, Kudlow JE. 2002 Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein O-GlcNAcylation to transcriptional repression. *Cell* **110**, 69–80. (doi:10.1016/S0092-8674(02)00810-3)

11. Sakabe K, Wang Z, Hart GW. 2010 Beta-N-acetylglucosamine (O-GlcNAc) is part of the histone code. *Proc. Natl Acad. Sci. USA* **107**, 19 915–19 920. (doi:10.1073/pnas.1009023107)

12. Kreppel LK et al. 1997 Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. *J. Biol. Chem.* **272**, 9308–9315. (doi:10.1074/jbc.272.14.9308)

13. Lubas WA, Hanover JA. 2000 Functional expression of O-linked GlcNAc transferase. Domain structure and substrate specificity. *J. Biol. Chem.* **275**, 10 983–10 988. (doi:10.1074/jbc.275.15.10983)

14. Gao Y et al. 2001 Dynamic O-glycosylation of nuclear and cytosolic proteins: cloning and characterization of a neutral, cytosolic beta-N-acetylglucosaminidase from human brain. *J. Biol. Chem.* **276**, 9838–9845. (doi:10.1074/jbc.M010420200)

15. Shafi R, Iyer SP, Ellies LG, O'Donnell N, Marek KW, Chui D, Hart GW, Marth JD. 2000 The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. *Proc. Natl Acad. Sci. USA* **97**, 5735–5739. (doi:10.1073/pnas.100471497)

16. Kreppel LK, Hart GW. 1999 Regulation of a cytosolic and nuclear O-GlcNAc transferase. Role of the tetratricopeptide repeats. *J. Biol. Chem.* **274**, 32 015–32 022. (doi:10.1074/jbc.274.45.32015)

17. Clark RJ, McDonough PM, Swanson E, Trost SU, Suzuki M, Fukuda M, Dillmann WH. 2003 Diabetes and the accompanying hyperglycemia impairs cardiomyocyte calcium cycling through increased nuclear O-GlcNAcylation. *J. Biol. Chem.* **278**, 44 230–44 237. (doi:10.1074/jbc.M303810200)

18. Clarke AJ, Hurtado-Guerrero R, Pathak S, Schuttelkopf AW, Borodkin V, Shepherd SM, Ibrahmin AF, van Aalten DM. 2008 Structural insights into mechanism and specificity of O-GlcNAc transferase. *EMBO J.* **27**, 2780–2788. (doi:10.1038/emboj.2008.186)

19. Martinez-Fleites C, Macauley MS, He Y, Shen DL, Vocadlo DJ, Davies GJ. 2008 Structure of an O-GlcNAc transferase homolog provides insight into intracellular glycosylation. *Nat. Struct. Mol. Biol.* **15**, 764–765. (doi:10.1038/nsmb.1443)

20. Lazarus MB, Nam Y, Jiang J, Sliz P, Walker S. 2011 Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **469**, 564–567. (doi:10.1038/nature09638)

21. Lazarus MB, Jiang J, Gloster TM, Zandberg WF, Whitworth GE, Vocadlo DJ, Walker S. 2012 Structural snapshots of the reaction coordinate for O-GlcNAc transferase. *Nat. Chem. Biol.* **8**, 966–968. (doi:10.1038/nchembio.1109)

22. Schimpl M et al. 2012 O-GlcNAc transferase invokes nucleotide sugar pyrophosphate participation in catalysis. *Nat. Chem. Biol.* **8**, 969–974. (doi:10.1038/nchembio.1108)

23. Dong DL, Hart GW. 1994 Purification and characterization of an O-GlcNAc selective N-acetyl-beta-D-glucosaminidase from rat spleen cytosol. *J. Biol. Chem.* **269**, 19 321–19 330.

24. Heckel D, Comtesse N, Brass N, Blin N, Zang KD, Meese E. 1998 Novel immunogenic antigen homologous to hyaluronidase in meningioma. *Hum. Mol. Genetics* **7**, 1859–1872. (doi:10.1093/hmg/7.12.1859)

25. Schultz J, Milpetz F, Bork P, Ponting CP. 1998 SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl Acad. Sci. USA* **95**, 5857–5864. (doi:10.1073/pnas.95.11.5857)

26. Wells L *et al*. 2002 Dynamic O-glycosylation of nuclear and cytosolic proteins: further characterization of the nucleocytoplasmic beta-N-acetylglucosaminidase, O-GlcNAcase. *J. Biol. Chem.* **277**, 1755 – 1761. (doi:10.1074/jbc.M105274200)

27. Toleman C, Paterson AJ, Whisenhunt TR, Kudlow JE. 2004 Characterization of the histone acetyltransferase (HAT) domain of a bifunctional protein with activable O-GlcNAcase and HAT activities. *J. Biol. Chem.* **279**, 53 665 – 53 673. (doi:10.1074/jbc.M410406200)

28. Schultz J, Pils B. 2002 Prediction of structure and functional residues for O-GlcNAcase, a divergent homologue of acetyltransferases. *FEBS Lett.* **529**, 179 – 182. (doi:10.1016/S0014-5793(02)03322-7)

29. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009 The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* **37**(Database issue), D233 – D238. (doi:10.1093/nar/gkn663)

30. Dennis RJ, Taylor EJ, Macauley MS, Stubbs KA, Turkenburg JP, Hart SJ, Black GN, Vocadlo DJ, Davies GJ. 2006 Structure and mechanism of a bacterial beta-glucosaminidase having O-GlcNAcase activity. *Nat. Struct. Mol. Biol.* **13**, 365 – 371. (doi:10.1038/nsmb1079)

31. Dorfmueller HC, Borodkin VS, Schimpl M, Shepherd SM, Shpiro NA, van Aalten DMF. 2006 GlcNAcstatin: a picomolar, selective O-GlcNAcase inhibitor that modulates intracellular O-glcNAcylation levels. *J. Am. Chem. Soc.* **128**, 16 484 – 16 485. (doi:10.1021/ja066743n)

32. Rao FV, Dorfmueller HC, Villa F, Allwood M, Eggleston IM, van Aalten DMF. 2006 Structural insights into the mechanism and inhibition of eukaryotic O-GlcNAc hydrolysis. *EMBO J.* **25**, 1569 – 1578. (doi:10.1038/sj.emboj.7601026)

33. Yuzwa SA *et al*. 2008 A potent mechanism-inspired O-GlcNAcase inhibitor that blocks phosphorylation of tau *in vivo*. *Nat. Chem. Biol.* **4**, 483 – 490. (doi:10.1038/nchembio.96)

34. Schimpl M, Schuttelkopf AW, Borodkin VS, van Aalten D. 2010 Human OGA binds substrates in a conserved peptide recognition groove. *Biochem. J.* **432**, 1 – 7. (doi:10.1042/BJ20101338)

35. Schimpl M, Borodkin VS, Gray LAJ, vanAalten DMF. 2012 Synergy of peptide and sugar in O-GlcNAcase substrate recognition. *Chem. Biol.* **19**, 173 – 178. (doi:10.1016/j.chembiol.2012.01.011)

36. Toleman CA *et al*. 2006 The histone acetyltransferase NCOAT contains a zinc finger-like motif involved in substrate recognition. *J. Biol. Chem.* **281**, 3918 – 3925. (doi:10.1074/jbc.M510485200)

37. Butkinaree C, Cheung WD, Park S, Park K, Barber M, Hart GWI. 2008 Characterization of beta-N-acetylglucosaminidase cleavage by caspase-3 during apoptosis. *J. Biol. Chem.* **283**, 23 557 – 23 566. (doi:10.1074/jbc.M804116200)

38. Dyda F, Klein DC, Hickman AB. 2000 GCN5-related N-acetyltransferases: a structural overview. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 81 – 103. (doi:10.1146/annurev.biophys.29.1.81)

39. Kabsch W, Sander C. 1983 Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577 – 2637. (doi:10.1002/bip.360221211)

40. Cole C, Barber JD, Barton GJ. 2008 The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* **36**(Web Server issue), W197 – W201. (doi:10.1093/nar/gkn238)

41. Larkin MA *et al*. 2007 Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947 – 2948. (doi:10.1093/bioinformatics/btm404)

42. Bond CS, Schuttelkopf AW. 2009 ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **65**, 510 – 512. (doi:10.1107/S0907444909007835)

43. Liszczak G, Arnesen T, Marmorstein R. 2011 Structure of a ternary Naa50p (NAT5/SAN) N-terminal acetyltransferase complex reveals the molecular basis for substrate-specific acetylation. *J. Biol. Chem.* **286**, 37 002 – 37 010. (doi:10.1074/jbc.M111.282863)

44. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. 1993 Procheck—a program to check the stereochemical quality of protein structures. *J. Appl Crystallogr.* **26**, 283 – 291. (doi:10.1107/S0021889892009944)

45. Vetting MW, Bareich DC, Yu M, Blanchard JS. 2008 Crystal structure of RimI from Salmonella typhimurium LT2, the GNAT responsible for N(alpha)-acetylation of ribosomal protein S18. *Protein Sci. Publ. Protein Soc.* **17**, 1781 – 1790. (doi:10.1110/ps.035899.108)

46. Neuwald AF, Landsman D. 1997 GCN5-related histone N-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein. *Trends Biochem. Sci.* **22**, 154 – 155. (doi:10.1016/S0968-0004(97)01034-7)

47. Marmorstein R. 2001 Structure of histone acetyltransferases. *J. Mol. Biol.* **311**, 433 – 444. (doi:10.1006/jmbi.2001.4859)

48. Krissinel E, Henrick K. 2004 Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2256 – 2268. (doi:10.1107/S0907444904026460)

49. Arnold K *et al*. 2006 The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195 – 201. (doi:10.1093/bioinformatics/bti770)

50. Lau OD *et al*. 2000 p300/CBP-associated factor histone acetyltransferase processing of a peptide substrate. Kinetic analysis of the catalytic mechanism. *J. Biol. Chem.* **275**, 21 953 – 21 959. (doi:10.1074/jbc.M003219200)

51. Peneff C *et al*. 2001 The crystal structures of Apo and complexed Saccharomyces cerevisiae GNA1 shed light on the catalytic mechanism of an amino-sugar N-acetyltransferase. *J. Biol. Chem.* **276**, 16 328 – 16 334. (doi:10.1074/jbc.M009988200)

52. Tanner KG *et al*. 1999 Catalytic mechanism and function of invariant glutamic acid 173 from the histone acetyltransferase GCN5 transcriptional coactivator. *J. Biol. Chem.* **274**, 18 157 – 18 160. (doi:10.1074/jbc.274.26.18157)

53. Corfield AP et al. 1984 Purification and properties of Acetyl-Coa - glucosamine 6-phosphate N-acetyltransferase from rat-liver. *Biochem. Soc. Trans.* **12**, 656.

54. Vessal M, Jaberi-Pour M. 1998 Partial purification and kinetic properties of three different D-glucosamine 6-P : N-acetyltransferase forms from human placenta. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **121**, 379 – 384. (doi:10.1016/S0305-0491(98)10119-0)

55. Toleman C, Paterson AJ, Whisenhunt TR, Kudlow JE. 2004 Characterization of the histone acetyltransferase (HAT) domain of a bifunctional protein with activable *O*-GlcNAcase and HAT activities. *J. Biol. Chem.* **279**, 53 665 – 53 673. (doi:10.1074/jbc.M410406200)

56. Otwinowski Z, Minor W. 1997 Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307 – 326. (doi:10.1016/S0076-6879(97)76066-X)

57. Winn MD *et al*. 2011 Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235 – 242. (doi:10.1107/S0907444910045749)

58. Sheldrick GM. 2010 Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 479 – 485. (doi:10.1107/S0907444909038360)

59. Pape T, Schneider TR. 2004 HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Crystallogr.* **37**, 843 – 844. (doi:10.1107/S0021889804018047)

60. Langer G, Cohen SX, Lamzin VS, Perrakis A. 2008 Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171 – 1179. (doi:10.1038/nprot.2008.91)

61. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010 Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486 – 501. (doi:10.1107/S0907444910007493)

62. Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F, Murshudov GN. 2004 REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2184 – 2195. (doi:10.1107/S0907444904023510)

63. Schuttelkopf AW, van Aalten DMF. 2004 PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 1355 – 1363. (doi:10.1107/S0907444904011679)

64. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC. See http://www.pymol.org.

65. Papalia G, Myszka D. 2010 Exploring minimal biotinylation conditions for biosensor analysis using capture chips. *Anal. Biochem.* **403**, 30 – 35. (doi:10.1016/j.ab.2010.03.044)