Structure of an acetylating aldehyde dehydrogenase from the thermophilic ethanologen *Geobacillus thermoglucosidasius*

Jonathan Extance, Michael J. Danson and Susan J. Crennell

Centre for Extremophile Research, Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, England

Correspondence e-mail: s.j.crennell@bath.ac.uk

Correspondence Telephone number: +44 1225 384203

Correspondence FAX number +44 1225 386779

Running Title (50 characters or less)

G. thermoglucosidasius AcAldDH

Competing interests: All authors declare that they have no competing interests in relation to the publication of this paper, that no funding source will gain or lose financially through publication of this paper, and that they have no personal financial interests that would be affected by this publication

Total number of Manuscript pages: 15

Supplementary material pages: 5

Figures: 6

Tables: 1

Description of supplementary material

Figure 1: Michaelis-Menten plot of AcAldDH enzymic activity (μmol NADH oxidised/min/mg of protein) against concentration of acetyl-CoA (mM) at a fixed concentration of NADH (0.21 mM) for the purified *Gt*AcAldDH enzyme. (MM_AcCoAvarying.tif)

Figure 2: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of AcAldDH activity (μmol NADH oxidised/min/mg protein) with respect to concentration of acetyl-CoA (mM) for the purified *Gt*AcAldDH enzyme. (HW_AcCoAvarying.tif).

Figure 3: Michaelis-Menten plot of AcAldDH enzymic activity (μmol NADH oxidised/min/mg of protein) against concentration of NADH (mM) at a fixed concentration of acetyl-CoA (0.34 mM) for the purified *Gt*AcAldDH enzyme. (MM_NADHvarying.tif)

Figure 4: Hanes-Woolf plot ([S]/v vs. [S]) for the variation of aldDH activity (μmol NADH oxidised/min/mg of protein) with respect to concentration of NADH (mM) for the purified *Gt*AcAldDH enzyme. (HW_NADHvarying.tif)

Figure 5: Summary chromatograph of a *Gt*AcAldDH gel filtration run on a Superdex 200 10/300 GL column. The blue chromatograph line corresponds to A_{280nm} reading (mAU). SDS-PAGE analysis of peak fractions is shown (L = load, M = Biorad Broad Range markers [M_r/1000], 1-13 = protein fractions from gel filtration A10-B7). Peak 2, eluted at a volume of 12.71ml, corresponds to a relative molecular mass (M_r) of 191,000. Since the expected M_r of *Gt*AcAldDH is 52,174, this suggests that the majority of *Gt*AcAldDH is in a tetrameric form. (GF3.tif)

Abstract (no more than 250 words)

Acetylating aldehyde dehydrogenases (AcAldDH) catalyse the acetylation of Coenzyme-A (CoA), or in reverse generate acetaldehyde from Acetyl-CoA using NADH as a co-factor. This paper reports the expression, purification, enzyme assay and X-ray crystal structures of an AcAldDH from *Geobacillus thermoglucosidasius* (*Gt*AcAldDH) to 2.1Å and in complex with CoA and NAD^+ to 4.0Å. In the structure, the AcAldDH forms a close-knit dimer, similar to that seen in other Alcohol Dehydrogenase (ADH) structures. In *Gt*AcAldDH these dimers

associate via their N-termini to form weakly-interacting tetramers. This mode of tetrameric association is also seen in an unpublished AcAldDH deposited in the PDB, but is in contrast to all other ADH structures, (including the one other published AcAldDH found in a bacterial microcompartment), in which the dimers bury a large surface area including the C-termini. This novel mode of association sequesters the active sites and potentially-reactive acyl-enzyme intermediates in the centre of the tetramer. In other respects the structure is very similar to the other AcAldDH, binding the cofactors in a corresponding fashion. This similarity enabled the identification of a shortened substrate cavity in *G. thermoglucosidasius* AcAldDH, explaining the limitations on the length of substrate accepted by the enzyme.

Keywords

Geobacillus thermoglucosidasius, acetylating aldehyde dehydrogenase, X-ray crystal structure, tetramer, substrate specificity

Importance/ Impact of this work (50-75 word statement)

Although in all other respects a similar structure to that of other ADH, the acetylating alcohol dehydrogenase from *Geobacillus thermoglucosidasius* has a novel mode of tetrameric association that may prevent inappropriate reactions by sequestering the active sites from all but the smallest molecules. From the structure, the limited range of substrates accepted can be explained by a shortened substrate cavity.

1. Introduction

The thermophilic microorganism *Geobacillus thermoglucosidasius* NCIMB 11955 is a Gram-positive bacterium that grows optimally between 333 and 338 K and is able to ferment biomass-derived C6 and C5 sugars to lactate, acetate, formate and ethanol. The strain TM242 has been metabolically engineered to produce optimal levels of ethanol [1] and therefore is of potential biotechnological importance in second-generation biofuel production.

In *G. thermoglucosidasius* the bifunctional dehydrogenase, ADHE, catalyses the production of ethanol from acetyl-CoA in a two-step reaction:

(a) Acetyl-CoA + NADH + H^+ ⟶ Acetaldehyde + NAD^+ + CoASH

(b) Acetaldehyde + NADH + H^+ ⟶ Ethanol + NAD^+

ADHE consists of an N-terminal acetylating aldehyde dehydrogenase domain (AcAldDH), catalysing reaction (a) and a C-terminal alcohol dehydrogenase domain (ADH) catalysing reaction (b). We have previously reported the catalytic properties and the structure of the *G. thermoglucosidasius* ADH domain to 2.5 Å resolution [2]. However, no catalytically-active AcAldDH domain could be obtained, either expressed heterologously in *Escherichia coli* or homologously in *G. thermoglucosidasius* [3]. Moreover, the AcAldDH activity in native and recombinant ADHE was found to be low compared to that of the ADH domain and was relatively unstable, especially in terms of oxygen sensitivity. This raises the possibility of taking another more active and stable AcAldDH to fuse with the ADH domain to improve the rate of bioethanol production.

Therefore, in the current paper, we report the identification of another AcAldDH-encoding gene in *G. thermoglucosidasius*, confirm the catalytic activity of the recombinant enzyme and report its X-ray crystal structure to 2.1 Å resolution. A lower resolution structure (4 Å) has also been obtained in the presence of its ligands, acetyl-CoA and NAD^+.

2. Results

2.1 Characterisation of recombinant AcAldDH

The purified recombinant enzyme catalysed the conversion of acetyl-CoA and NADH to acetaldehyde and NAD^+ with the following kinetic parameters: V_{max} = 36 (± 4) μmol min^{-1} mg^{-1}, K_M (acetyl-CoA) = 24 (± 1) μM, K_M (NADH) = 42 (± 2) μM (Michaelis-Menten and Hanes-Woolf plots of the data are shown in Supplementary Figures 1-4). The temperature optimum in the standard assay was 60°C, at which temperature the enzyme retained >80% of its activity in 60 min. Gel filtration indicated that the enzyme is a homo-tetramer (Supplementary Figure 5).

2.2 Overall structure

The structure of AcAldDH was refined to 2.1 Å resolution with an R-factor of 14.3% and R_{free} of 19.6%, containing four protein monomers and ordered solvent molecules including 4 Mg^{2+} ions (one per protein monomer), and 1269 water, 8 acetate and 36 glycerol molecules. Analysis by Molprobity indicates that 97.8% of the residues are in the most favoured regions of the Ramachandran plot, with no outliers, and the overall score for the protein geometry was at the 100th percentile for this resolution. Data collection and processing statistics are given in Table 1.

Each AcAldDH monomer has 3 domains: two αβα domains with the active site cleft lying between them, and a dimerization domain of three β-strands that wraps around its partner monomer. The structures of the monomers are very similar, with superposition of any pair having an RMSD of 0.99 Å over the full length of the structure. The N-terminal 3 residues of each monomer, interacting at the tetramer interface, are disordered, taking up at least two conformations in the structure. This may be an artefact of the N-terminal His-tag (not visible).

PISA analysis [4] suggests that the 4 protein molecules in the asymmetric unit interact to form a dimer of dimers; within the dimers the extensive interactions of the dimerisation domains bury a surface area of 2975Å^2, while the interaction of N-terminal helices that provides the tetrameric interface only buries 453Å^2. Each monomer contains one

Mg^{2+} ion, bound to D215. Although at the dimer interface, the ion does not appear to play a significant role in dimerization; interactions with the second monomer are distant and water-mediated. Similarly, the Mg^{2+} ions are unlikely to play a role in catalysis, being more than 20 Å away from the nearest nucleotide-binding site.

Related structures were identified in the PDB using the *PDBeFold* at the European Bioinformatics Institute [5]. The most similar were *Listeria monocytogenes* aldehyde dehydrogenase (*Lm*AcAldDH, 3K9D, [6], used as molecular replacement model) with an RMSD of 1.0 Å over 450 C_α positions, a propionaldehyde dehydrogenase from *Lachnoclostridium phytofermentans* (C*p*AcAldDH 4C3S, [7]) with an RMSD of 1.3 Å°over 402 C_α positions, and the ACDH domain of an ADHE from *Vibrio parahaemolyticus* (*Vp*AcAldDH 3MY7, [8]) with an RMSD of 2.0 Å over 390 C_α positions. Other identified structures had lower sequence and structural similarity and belong to different ALDH families.

Obtaining a co-crystal structure was particularly challenging due to the sensitivity of the crystals to both cryoprotectants and radiation damage; however, a room temperature structure to relatively low resolution (4Å) was obtained (data in Table 1). The tetrameric arrangement of the enzyme is the same in both structures. Analysis by DYNDOM [9] of any pair of monomers in the apo and cofactor-bound AcAldDH structures showed no internal domain motions on binding substrates; however, a small (7.6°) rotation of the dimers relative to one another was detected. No significant movements of side chains were detected on binding substrate, although this would be limited by the resolution of the co-crystal structure.

2.3 Catalytic site.

Through sequence alignment with other AcAldDH, Cys 273 can be predicted to be the catalytic residue involved in acetyl-transfer. This prediction is corroborated by the particular sensitivity of Cys 273 to radiation damage; that is, the AcAldDH sequence contains 4 Cys residues yet this is the only one that is oxidised to cysteine-sulphinic acid in the apo

structure. Electron-rich sulphur atoms are susceptible to radiation damage, including the breaking of S-H bonds, and a cysteine involved in catalysis would be particularly susceptible, as the active site will have evolved to stabilise the charged residue. The oxidation is particularly likely in AcAldDH as the active site is required to interact with an acetyl-group bound to the cysteine during the catalytic cycle, so will readily accommodate oxidised forms such as cysteine-sulphinic acid. Mutation of the corresponding amino acid, Cys 269, in CpAcAldDH was shown to abolish activity [7]. The rationale for growing co-crystals in the presence of acetyl-CoA and NAD^+ was to capture the acetyl-enzyme intermediate, and as the shape of the moiety attached to Cys 273 in the co-crystals was appropriate and significantly different from sulphinic acid, it was modelled as acetyl-cysteine. (Figure 1).

The 4 moieties bound in the AcAldDH active sites in the co-crystal structure were modelled as ADP, the core structure shared by CoA and NAD (Figure 1). Some extra density was seen extending towards Cys273 in some monomers, but it was not interpretable due to the limited resolution of the structure.

3. Discussion

3.1 Purified enzyme

The purified *Gt*AcAldDH had a similar K_M (acetyl-CoA) as the AcAldDH from ADHE [2], but its K_M (NADH) is 4-fold lower; however, it is not possible to compare their V_{max} values as the AcAldDH component of ADHE lost activity during purification [2]. The *Gt*AcAldDH was active and stable at the growth temperature of *G. thermoglucosidasius*, and no sensitivity to oxygen was observed for this enzyme, in contrast to the AcAldDH domain of ADHE [2]. Transcriptomic analysis of *G. thermoglucosidasius* grown aerobically on glucose minimal medium indicates that the enzyme is expressed but at a level <5% of that observed for ADHE; moreover, the level of *Gt*AcAldDH expression did not change significantly when cultures were grown fermentatively under anaerobic conditions, whereas the level of ADHE transcription increased markedly [D.J. Leak, University of Bath, personal communication].

Therefore, it appears that the *Gt*AcAldDH does not contribute significantly to fermentative ethanol production under the conditions investigated.

3.2 Tetrameric structure

In these structures, the *Gt*AcAldDH has been shown to form a dimer-of-dimers tetramer, this tetrameric nature being consistent with the gel filtration results. Three other AcAldDH structures are available in the PDB as described in Results. All four of these enzymes form very similar C2 dimers, with the *Vp*AcAldDH alone not forming higher oligomers (probably because it is a fragment of a larger molecule), so it was not considered further. The remaining 3 enzymes form tetramers, but with different modes of association. Both AcAldDH and *Lm*AcAldDH interact in the same way via their N-termini (interface area 453Å^2) to form a loose tetramer, while the *Cp*AcAldDH dimers have a significantly larger tetrameric interface (1148 Å^2, [7]) (Figure 2).

This larger interface of *Cp*AcAldDH uses patches of the structure that are structurally conserved with the other enzymes, yet less conserved in sequence than the majority, in particular the first two β-strands and the C-terminus (which together form a three-stranded sheet described in *Cp*AcAldDH as an 'oligomerisation domain,' [7]), together with a helix within the catalytic domain at the other side of the structure (Figure 2, 3). The oligomerisation domain corresponds to the dimerization domain identified above in *Gt*AcAldDH, and it is also involved in dimerization in *Cp*AcAldDH, the main difference being the conformation of the C-terminal extension from the sheet, which in AcAldDH wraps around the dimer partner, and in *Cp*AcAldDH forms a large part of the tetramer interface. This is reflected in the larger buried surface area in the *Gt*AcAldDH dimer (2975Å^2) compared to *Cp*AcAldDH (2258 Å^2). Structural alignment of the three sequences is shown in Figure 3 with amino acids involved in the *Cp*AcAldDH interfaces in brown font. The *Cp*AcAldDH E123 that forms a salt bridge with the C-terminus (amino acids involved in salt bridges highlighted with a brown dot in figure 3) is replaced in the two non-BMC enzymes by

a hydrophobic residue that forms part of the dimer interface. Thus the *Cp*AcAldDH mode of tetrameric association is destabilised in the non-BMC enzymes.

The sequence alignment also shows that, although truncated to remove a microcompartment localisation signal, the remaining N-terminus of C*p*AcAldDH in the structure 4C3S is at least as long as the visible portion of *Gt*AcAldDH (Figure 3), yet instead of forming the extended interface helix seen in the centre of the *Gt*AcAldDH and *Lm*AcAldDH tetramers, the N-termini of *Cp*AcAldDH folds back on the surface of the dimer, extending a β-sheet on the outer surface that would be a suitable location for the BMC signal (Figure 2). In *Cp*AcAldDH the 4 catalytic sites are also on the outer surface, where they are more accessible to solvent, while in the other two AcAldDHs they face into the tetrameric central cavity (indicated by bound nucleotides in Figure 2).

Although these are the first structures of acetylating aldehyde dehydrogenases to be published, numerous other aldehyde dehydrogenases have an oligomeric structure similar to that of *Cp*AldDH ([7], [10]); thus, it is the structures of the AcAldDH and *Lm*AcAldDH enzymes that are unusual. This association by the N-termini allows the active sites to shelter inside the molecule and therefore, in the absence of a microcompartment, this organisation may help to sequester the reactive acyl intermediate from the rest of the cell.

3.3 Active sites and co-factor binding

As predicted by Shone *et al*. [11] and shown in *Cp*AcAldDH, NAD (4C3S) and CoA (5DBV) occupy similar but not identical positions in the active site, the difference being particularly marked at the adenine ring. In the AcAldDH active site, the phosphate groups of the co-factor interact with His 166 that is part of a highly-conserved PHP motif. Superposition of the NAD- and CoA- bound *Cp*AcAldDH structures indicates that the cofactor conformation seen in AcAldDH is much more similar to that of CoA than NAD (Figure 4).

Although the pyrophosphates all overlap with the CoA conformation, in only three of the ADP moieties in AcAldDH is the adenine ring in a similar conformation to that of CoA in

5DBV, while the fourth has the adenine ring in a more NAD-like conformation (data not shown). More accurate modelling of the cofactors is not possible due to the resolution.

A mechanism for acylating aldehyde dehydrogenases was proposed following mutagenesis studies of *Cp*AcAldDH.[7] In this enzyme, His387 acts as a base to activate the catalytic Cys269 and to stabilise the acyl-transfer intermediate. Following the departure of NADH, the arriving CoA is deprotonated by Glu357 and attacks the acyl-enzyme intermediate to form the thioester product. These amino acids are conserved in sequence (highlighted with blue symbols in Figure 3) and position in the structure across the other enzymes (Figure 5); thus the acylation mechanism is likely to be conserved.

*Cp*AldDH was shown to be active with aldehydes from C2 to C6, [7] while the activities of AcAldDH with propionyl and butyryl-CoA were only 28% and 9%, respectively, of the activity of acetyl-CoA. In *Cp*AldDH a considerable substrate-binding tunnel was observed in which acetate molecules were bound (Figure 6A). In *Gt*AcAldDH, this is much shortened by a ring of aliphatic amino acids, consistent with the preference for shorter substrates (Figure 6B). These residues Ile272, 429 and Leu 439 are marked with blue squares in Figure 3. Most significantly, in *Gt*AcAldDH the amino acid immediately N-terminal of the catalytic Cys, is an isoleucine (Ile 272) which truncates the pocket, while the corresponding amino acid in *Cp*AcAldDH is a proline allowing larger substrates to be admitted (Figures 6 and 3). Due to the conservation of the Ile in *Lm*AcAldDH, we would predict that this enzyme too has a preference for shorter substrates.

4. Materials and Methods

4.1 Cloning of the AcAldDH gene

To identify possible AcAldDH protein-coding genes in the *G. thermoglucosidasius* NCIMB 11955 genome, a protein BLAST search (of hypothesised gene products) was carried out based on the protein sequence of the predicted AcAldDH domain of ADHE protein using the ERGO (integrated genomics) tool. Two genes encoding proteins annotated to belong to EC

1.2.1.10 (acetylating aldehyde dehydrogenases) were identified; gene *acAldDH* (corresponding to the gene with locus tag *Geoth_3234* in the C56-YS93 strain) was chosen for all the work reported in this paper as the encoded protein possesses 43% identity and 59% similarity to the AcAldDH domain of ADHE.

The AcAldDH gene was PCR amplified using the forward primer 5'-GCT AGC ATG TTG CGT GAC ATC GAT TTG C-3' and the reverse primer 5'-CCG CTC GAG TTA TTT AGT TAA TGC AGG TTC TTT TTG-3'. The primers contained *Nhe*I and *Xho*I restriction sites, respectively. The PCR product was A-tailed, ligated into the pGEM-T easy vector, ethanol precipitated and transformed into *E. coli* JM109 cells for blue/white screening. Successful cloning of the protein coding sequence was confirmed by DNA sequencing. The gene in pGEM-T was then digested with *Nhe*I and *Xho*I, gel purified and ligated into the previously digested pET28a vector for expression of the AcAldDH with an N-terminal His-tag. Ligations were transformed into JM109 cells that were subject to PCR colony screening. Positive plasmids from the PCR screen were screened by restriction digestion with *Nhe*I and *Xho*I, prior to positive constructs being sent for DNA sequencing with vector-specific primers.

4.2 Recombinant protein expression and purification

The pET28a-AcAldDH plasmid construct was transformed into *E. coli* protein expression strain BL21 (DE3). Cultures were grown in LB medium containing kanamycin (30 µg/ml) at 37°C with shaking at 225 rev min^{-1}, and protein expression was induced at an OD_{600} of between 0.8 and 1.0 by supplementation with 1 mM IPTG for 4.5 h. Cells were harvested by centrifugation at 5,300*g* for 20 min at 277 K, and were resuspended in HIS-BIND buffer (300 mM NaCl, 50 mM Tris buffer pH 8.0, 20 mM imidazole; Acros Organics, Geel, Belgium). EDTA-free protease-inhibitor tablets (Roche, Welwyn Garden City, UK) were added at a minimum concentration of one tablet to 10 ml sample volume, and the cells were lysed by four 30-second bursts of sonication using a 150-watt Ultrasonic Disintegrator (MSE Scientific

Instruments, Crawley, UK) and the soluble fraction of the cell extract obtained by centrifugation at 16,000*g* for 18 min at 277 K.

The protein was purified by affinity chromatography on metal-chelating cellulose (Bioline, London, UK) charged with $NiSO_4$. After loading and washing in HIS-BIND buffer, protein was eluted with 20% (v/v) HIS-ELUTE buffer (300 mM NaCl, 50 mM Tris buffer pH 8, 0.2 M imidazole) and pooled fractions containing enzyme activity were dialysed into 50 mM Tris buffer pH 8.0, 150 mM NaCl.

4.3 Enzyme assay

AcAldDH was routinely assayed in 50 mM citrate buffer, pH 6.0, 0.1 mM zinc acetate, 0.24 mM NADH and 0.14 mM acetyl-CoA. Assays were started by the addition of enzyme, and the decrease in NADH concentration with time was followed at 340 nm. Substrate-independent background rates were measured. Enzyme activity is expressed as µmol NADH oxidised/min/mg protein.

4.4 Crystallisation, X-ray data collection and structure solution

Prior to crystallisation-condition screening, purified AcAldDH samples were concentrated using a Vivaspin 5K MWCO centrifugal filter device (Sartorius, Stonehouse, UK) and then filtered using an Ultrafree centrifugal filter device (0.45 µm) (Millipore, MA, USA). Initial screening of crystallisation conditions was carried out using various crystallisation screens, with best results obtained with the PACT screen (Molecular Dimensions Ltd., Newmarket, UK). Optimised crystals were grown using the hanging-drop method at a protein concentration of 9.3 mg ml^{-1} diluted 2:1 with well solution (0.1 M sodium acetate pH 4.8, 0.2 M $MgCl_2$, 15% (v/v) PEG 6000). Crystals took approximately 2-4 weeks to appear at 289 K; 30% (v/v) glycerol was used as a cryo-protectant when the crystals were cooled for data collection.

To obtain crystals containing ligands, 0.5 mM acetyl-CoA and 5 mM NAD^+ were added to the protein solution immediately prior to the screens. Crystals producing the best

diffraction data, optimised from PACT screens, were grown using protein at 10.7 mg ml^{-1} diluted 1:1 with well solution (0.1 M MMT buffer pH 5.0 (Molecular Dimensions Ltd), 23% (v/v) PEG 1500), taking approximately 4 weeks to appear at 16°C.

Data were collected from apo crystals at 100 K on the Diamond Light Source beamline I04 at a wavelength of 0.98 Å, and processed using HKL2000. Initial phases were obtained by molecular replacement using BALBES [12], using as a model the *Listeria monocytogenes* aldehyde dehydrogenase (PDB code 3K9D, [6]), which shares 53% identity with AcAldDH. This model underwent cycles of manual rebuilding in COOT [13] and refinement in REFMAC5 [14], with final refinement in PHENIX [15]. Model validation was performed with MOLPROBITY [16]. Although extensive efforts were made to find a cryoprotectant for the co-crystals, every treatment destroyed diffraction quality, so data were collected at room temperature. Radiation damage, even on the home source (Rigaku MicroMax 007HF with a Saturn 944+ CCD detector), was considerable, limiting the data collected to a maximum of 20° from each crystal; the final data set was derived from the 6 most isomorphous crystals, restricted to 4 Å although diffraction was seen beyond 3.2 Å in the first images. An initial model was obtained by rigid body refinement of the 4 chains of the apo structure, followed by similar rebuilding and refinement in COOT and PHENIX, with validation using MOLPROBITY. All figures were drawn using PyMOL [17] unless otherwise stated.

Acknowledgements

This work was funded by TMO Renewables Ltd and the Biotechnology and Biological Sciences Research Council (UK) through a CASE studentship for JE. We thank Diamond Light Source for access to beamline I04 (in project mx7131-1) that contributed to the results presented here.

References

1. Cripps RE, Eley K, Leak DJ, Rudd B, Taylor M, Todd M, Boakes S, Martin S, Atkinson T (2009). Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. Metab. Eng. 11: 398-408.
2. Extance J, Crennell SJ, Eley K, Cripps R, Hough DW, Danson MJ (2013). Structure of a bifunctional alcohol dehydrogenase involved in bioethanol generation in *Geobacillus thermoglucosidasius*. Acta Crystalllogr. D69:2104-2115.
3. Extance J, (2012) Bioethanol Production: Characterisation of a bifunctional alcohol dehydrogenase from *Geobacillus thermoglucosidasius* Thesis (PhD) University of Bath.
4. Krissinel E, Henrick K (2007). Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774-797.
5. Krissinel E, Henrick K (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. D60, 2256-2268.
6. Patskovsky Y, Toro R, Freeman J, Miller S, Sauder JM, Almo SC, Burley SK, Crystal structure of probable aldehyde dehydrogenase from Listeria monocytogenes EGD-e, PDB
7. Tuck LR, Altenbach K, Ang TF, Crawshaw AD, Campopiano DJ, Clarke DJ, Marles-Wright J, (2016). Insight into Coenzyme A cofactor binding and the mechanism of

acyl transfer in an acylating aldehyde dehydrogenase from *Clostridium phytofermentans*. *Scientific Reports*, 6, 22108; doi:10.1038/srep22108.

8. Stein AJ, Weger A, Volkart L, Gu M, Joachimiak A. The crystal structure of the ACDH domain of an alcohol dehydrogenase from *Vibrio parahaemolyticus* to 2.5Å, PDB
9. Hayward S, Berendsen HJC (1998). Systematic analysis of domain motions in proteins from conformational change; New results on citrate synthase and T4 lysozyme. Prot. Struct. Func. Genet., 30: 144-154.
10. Di Costanzo L, Gomez GA, Christianson DW (2007). Crystal structure of lactaldehyde dehydrogenase from *Escherichia coli* and inferences regarding substrate and cofactor specificity. J. Mol. Biol. 366:481-93.
11. Shone CC, Fromm HJ (1981) Steady state and pre-steady-state kinetics of coenzyme A linked aldehyde dehydrogenase from *Escherichia coli* . Biochemistry 20: 7494-7501.
12. Long F, Vagin AA, Young P, Murshudov GN (2008) BALBES: a molecular replacement pipeline. Acta Crystallogr. D64,125-132.
13. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallogr. D66:486-501.
14. Murshudov GN, Skubák, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D67:355-367.
15. Adams PD, Afonine PV, Bunkoczi G, Chen VB, David IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Crosse-Kunstleve RW, McCoy AJ, Moriaty NW, Oeffner R, Read RJ, Richardson DC, Richardson JA, Terwilliger TC, Zwart PH (2010).PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D66: 213-221.
16. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS & Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D66:12-21.

17. Schrodinger, LLC (2016). The PyMOL Molecular Graphics System, Version 1.8.
18. Armougom F, Moretti S, Poirot O, Audic S, Dumas P, Schaeli B, Keduas V, Notredame C (2006). Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. Nucl. Acids Res. 34, W604-8.
19. Robert X, Gouet P (2014). Deciphering key features in protein structures with the new ENDscript server. Nucl Acids Res. 42, W320-W324.

Figure legends

Figure 1. A cartoon view of the secondary structure around the active site of *G. thermoglucosidasius* AcAldDH, with active site residues and ADP (representing the conserved core of the two compounds, coenzyme A and NAD^+, soaked into the crystals) shown in stick form and surrounded by the co-crystal 2FoFc electron density contoured at 1σ. Cys273 in the base of the active site has reacted to form acetyl-cysteine (SCY), a hypothesis supported by the electron density.

Figure 2. The tetrameric structure of (A) *G. thermoglucosidasius* *Gt*AcAldDH and (B) *Lachnoclostridium phytofermentans* C*p*AcAldDH in cartoon form with three monomers in subdued colours and the fourth in rainbow colour from blue at the N-terminus to red at the C-terminus. Cofactors bound into the active sites of each monomer are shown in CPK form. Inter-dimer interactions can be seen to be mediated by the N-termini in *Gt*AcAldDH, and by both the 3-stranded sheet incorporating the C-terminus and a helix (orange) at the other side of the *Cp*AldDH monomer.

Figure 3. Structure-based sequence alignment carried out using EXPRESSO [18] and annotated using ESPript3 [19] of *Gt*AcAldDH, the *Gt*AcAldDH His-tag construct crystallised (5J78, *Lm*AcAldDH (Q8Y701_LISMO), the *Lm*AcaldDH His-tagged construct crystallised (3K9D), C*p*AcAldDH (A9KN57_CLOPH), and the crystallised N-terminally-truncated 4C3S, with secondary structures for the 3 pdb files at the top, conserved amino acids white on red background, and similar amino acids in red text and boxed. His tags are shaded in orange, the bacterial micro-compartment localisation signal in a horizontal blue box at the N-terminus, the tetrameric interaction areas in *Cp*AcAldDH (4C3S) in brown text with the salt bridges in the tetramer interface shown by cyan dots. Annotation of residues important for the mechanism or substrate selection are shown by blue symbols under the alignment, the catalytic Cys (*Cp*AcAldDH 273) by a star, the Glu (*Cp*AcAldDH 357) and His (*Cp*AcAldDH

387) implicated in the mechanism by triangles, and those lining the substrate tunnel by rectangles.

Figure 4. Cartoon structures of the active sites of *Gt*AcAldDH (green) with that of *Cp*AcAldDH (grey) superimposed, centred on the cofactor-binding Histidine with other active-site residues in the background. (A) superposition of *Cp*AcAldDH structure containing CoA (5DBV) (light grey) and (B) NAD (4C3S, darker grey).

Figure 5. Superposition of the *Cp*AldDH (grey) and *Gt*AcaldDH (green) active site residues showing the similarity in position of Glu 357/362, His 387/394 and Cys 268/273 (mutated to Ala in *Cp*AldDH, and the acety-lcysteine labelled SCY in *Gt*AcAldDH). The NADH seen in *Cp*AldDH overlaps with the acetyl-cysteine in *Gt*AcAldDH in which structure the moiety has been truncated to ADP (orange) due to the resolution.

Figure 6. Surface slices through the active site regions of (A) *Cp*AcAldDH and (B) *Gt*AcAldDH in approximately the same orientation to show the extended ligand/substrate tunnel in *Cp*AcAldDH and the truncated tunnel in *Gt*AcAldDH. In *Cp*AcAldDH the tunnel is lined by Phe431 and Ile421 and gated by Phe423, while in *Gt*AcAldDH a ring of residues including the corresponding Leu439 and Ile 429 and the Ile272 terminate the tunnel. CoA is included in (A) to show the catalytic site and acetyl-cysteine (SCY-273) in (B).