Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of a bifunctional alcohol dehydrogenase involved in bioethanol generation in *Geobacillus thermoglucosidasius*

Jonathan Extance,[a] Susan J. Crennell,[a]* Kirstin Eley,[b] Roger Cripps,[b] David W. Hough[a] and Michael J. Danson[a]

[a]Centre for Extremophile Research, Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, England, and [b]TMO Renewables Ltd, 40 Alan Turing Road, Surrey Research Park, Guildford, Surrey GU2 7YF, England

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

Received 2 May 2013
Accepted 23 July 2013

PDB Reference: ADH domain of ADHE, 3zdr

Bifunctional alcohol/aldehyde dehydrogenase (ADHE) enzymes are found within many fermentative microorganisms. They catalyse the conversion of an acyl-coenzyme A to an alcohol *via* an aldehyde intermediate; this is coupled to the oxidation of two NADH molecules to maintain the NAD^+ pool during fermentative metabolism. The structure of the alcohol dehydrogenase (ADH) domain of an ADHE protein from the ethanol-producing thermophile *Geobacillus thermoglucosidasius* has been determined to 2.5 Å resolution. This is the first structure to be reported for such a domain. *In silico* modelling has been carried out to generate a homology model of the aldehyde dehydrogenase domain, and this was subsequently docked with the ADH-domain structure to model the structure of the complete ADHE protein. This model suggests, for the first time, a structural mechanism for the formation of the large multimeric assemblies or 'spirosomes' that are observed for this ADHE protein and which have previously been reported for ADHEs from other organisms.

1. Introduction

The thermophilic microorganism *Geobacillus thermoglucosidasius* NCIMB 11955 has been selected by the second-generation biofuel company TMO Renewables Ltd (Guildford, England) as a suitable candidate for sustainable bioethanol production. This Gram-positive bacterium grows optimally between 333 and 338 K and is able to utilize C_5 and C_6 sugars, including some oligomeric forms, derived from biomass feedstocks to produce lactate, formate, acetate and ethanol. The strain TM242 has been metabolically optimized to increase ethanol production through a series of gene knockouts and up-regulations, and these have been described previously by Cripps *et al.* (2009).

The bifunctional dehydrogenase thought to catalyse the conversion of acetyl-CoA to ethanol in TM242 is known as ADHE. ADHE proteins are common amongst fermentative microorganisms, in which acetyl-CoA is fermentatively converted to ethanol *via* an acetaldehyde intermediate with the essential concomitant regeneration of NAD^+ from NADH (Arnau *et al.*, 1998; Asanuma *et al.*, 2004; Atteia *et al.*, 2003; Bruchhaus & Tannich, 1994; Dan & Wang, 2000; Fontaine *et al.*, 2002; Koo *et al.*, 2005; Membrillo-Hernandez *et al.*, 2000; Peng *et al.*, 2008; Shaw *et al.*, 2008). ADHE proteins consist of an N-terminal acetylating aldehyde dehydrogenase domain (AldDH) and a C-terminal alcohol dehydrogenase domain (ADH) that is normally a member of the Fe-containing ADH superfamily. Together they catalyse the reactions

$$\text{acetyl-CoA} + \text{NADH} + \text{H}^+ \rightarrow \text{acetaldehyde} + \text{NAD}^+ + \text{CoASH},$$

$$\text{acetaldehyde} + \text{NADH} + \text{H}^+ \rightarrow \text{ethanol} + \text{NAD}^+.$$

Evaluation of an ADHE knockout strain of TM242 showed the complete abolition of ethanol production and an inability of the strain to survive under anaerobic conditions (TMO Renewables Ltd, unpublished data), strongly supporting its role in ethanol production.

An unusual property of ADHE proteins is the commonly observed formation of multimeric assemblies known as spirosomes (Bruchhaus & Tannich, 1994; Espinosa, 2001; Kessler *et al.*, 1992). These multimers are helical in nature and appear to be made up of more than 20 copies of the ADHE protein. The reason for the formation of these complexes remains unclear, but they may enhance catalytic efficiency through substrate channelling of the reactive acetaldehyde intermediate or provide some stabilization of the protein.

No high-resolution structure of a complete ADHE protein is currently available. Such a structure would allow the elucidation of possible channelling mechanisms between the AldDH and ADH domains of the protein, as well as illuminating the mechanisms of spirosome formation. One currently unpublished structure exists for the AldDH domain of *Vibrio parahaemolyticus* ADHE (PDB entry 3my7; Midwest Center for Structural Genomics, unpublished work), but no structural information is available for a corresponding ADH domain.

Here, we report the characterization and X-ray crystal structure of the ADH domain of the *G. thermoglucosidasius* ADHE protein and an *in silico* homology model of the AldDH domain based on similar proteins within the PDB. Furthermore, *in silico* docking of the two domains predicts their interactions within ADHE and suggests for the first time a structural mechanism for the multimeric assembly into spirosomes.

2. Materials and methods

Unless otherwise stated, crystallization-specific chemicals and consumables used were supplied by Molecular Dimensions (Newmarket, England); general laboratory reagents and bacteriological media were supplied by Sigma–Aldrich Co. Ltd (Poole, England) or Fisher Scientific Ltd (Loughborough, England).

2.1. TM242 cultures and cell-extract preparation

G. thermoglucosidasius TM242 was grown in a 2.5 l Braun Biostat B fermenter in 1.5 l modified urea sulfate medium (USM) containing 60 g l^{-1} sucrose, 2%(*w*/*v*) yeast extract, 25 m*M* NaH_2PO_4, 50 m*M* urea, 25 m*M* K_2SO_4, 5 m*M* citric acid, 3 m*M* $MgSO_4$, 50 µ*M* $CaCl_2$, 0.3 µ*M* biotin and 12.5 ml l^{-1} trace-element solution (60 m*M* H_2SO_4, 1.44 g l^{-1} $MnSO_4{\cdot}7H_2O$, 5.56 g l^{-1} $FeSO_4{\cdot}7H_2O$, 1.69 g l^{-1} $MnSO_4{\cdot}H_2O$, 0.25 g l^{-1} $CuSO_4{\cdot}5H_2O$, 0.56 g l^{-1} $CoSO_4{\cdot}7H_2O$, 0.06 g l^{-1} H_3BO_3, 0.89 g l^{-1} $NiSO_4{\cdot}6H_2O$). The system was run at 333 K and pH 6.7 with stirring at 600 rev min^{-1} and was inoculated with 100 ml of a mid-log phase seed culture (OD_{600} = 2). Once the dissolved oxygen levels had decreased to 0%, the fermentation was allowed to continue for a further 2.5 h prior to ending the run.

The cells were harvested by centrifugation at 5300*g* for 20 min and resuspended at 0.3 g of cells per millilitre of 50 m*M* EPPS buffer pH 8.0 containing 1 m*M* 4-(2-aminoethyl)-benzenesulfonyl fluoride hydrochloride (AEBSF); 1.5 µl of Benzonase nuclease (250 U μl^{-1}) per millilitre of resuspended cells was also added. The cells were lysed by four 30 s bursts of sonication using a 150 W Ultrasonic Disintegrator (MSE Scientific Instruments, Crawley, England) and the soluble fraction of the cell extract was obtained by centrifugation at 16 000*g* for 18 min at 277 K.

2.2. Purification of native ADHE protein

Anion-exchange chromatography and gel filtration were carried out at 298 K on an ÄKTAexplorer FPLC system (GE Healthcare, Buckinghamshire, England). A TM242 soluble cell extract in 50 m*M* EPPS buffer pH 8.7 containing 0.1 m*M* zinc acetate and 5 m*M* reduced glutathione was loaded onto two pre-equilibrated GE HiTrap 5 ml Q-Sepharose HP columns run in series at a flow rate of 1 ml min^{-1}. Proteins were eluted using a 0–1 *M* NaCl gradient over a 60 min period. The fractions with the highest enzyme activity were pooled, concentrated using a Vivaspin 5K MWCO centrifugal filter device (Sartorius, Epsom, England) and loaded onto a pre-equilibrated GE Superdex 200 10/300 GL column run at a flow rate of 0.3 ml min^{-1} in 50 m*M* EPPS pH 8.0, 0.1 m*M* zinc acetate, 5 m*M* reduced glutathione, 10%(*v*/*v*) glycerol. Fractions were assessed by SDS–PAGE and enzyme assays and those judged to be >95% pure were pooled.

2.3. Investigation of multimeric assembly

Assembly of the purified ADHE was investigated by gel filtration on a Superdex 200 10/300 GL column calibrated with molecular-weight standards (ribonuclease, 13 700; ovalbumin, 43 000; conalbumin, 75 000; aldolase, 158 000; ferritin, 440 000; GE Healthcare). Dynamic light-scattering analysis was carried out using a Nano-S Zetasizer (Malvern Instruments, Malvern, England). All readings were taken over an 80 s period at 298 K in a low-volume quartz cuvette containing 50 µl sample.

Samples of purified protein were also subjected to analysis using a NanoSight LM10 instrument (NanoSight, Amesbury, England; Filipe *et al.*, 2010). Protein samples were injected into the sample chamber and subjected to 635 nm laser light; images were collected at a rate of 30 frames per second. The data were then analysed using the associated analytical software, which generated the particle-size distribution plots.

2.4. Determination of protein concentration

Where possible, the concentration of protein present in the purified protein samples was determined using A_{280} measurements, with absorption coefficients determined for the protein of interest using the *ProtParam* tool (ExPASy). For

impure samples, protein concentrations were determined using a dye-binding assay (Bradford, 1976).

2.5. Cloning of the ADH gene fragment

From a vector containing the full-length *G. thermoglucosidasius adhe* gene, the fragment encoding the ADH domain (amino acids 459–869) of ADHE was PCR-amplified using Phusion DNA polymerase (Finnzymes, Vantaa, Finland) and the oligonucleotides CCGCATATGATTTACATGAATATGCAATGGTTTAAAG (forward primer) and CCCTCGAGGCTTAAACTCCTTTAAACGCT (reverse primer). The amplified fragment was ligated into the pET28a vector between the *Nde*I/*Xho*I sites of the multiple cloning site of this vector to incorporate an N-terminal His tag and the sequence was confirmed.

2.6. Recombinant protein expression and purification

The pET28a-ADH plasmid construct was transformed into the *Escherichia coli* protein-expression strain BL21 (DE3). Cultures were grown in LB medium containing kanamycin (30 µg ml^{-1}) at 310 K 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 m*M* IPTG for 4.5 h. The cells were harvested by centrifugation at 5300*g* for 20 min at 277 K and were resuspended in HIS-BIND buffer (300 m*M* NaCl, 50 m*M* Tris buffer pH 8.0, 20 m*M* imidazole; Acros Organics, Geel, Belgium). EDTA-free protease-inhibitor tablets (Roche, Welwyn Garden City, England) were added at a minimum concentration of one tablet per 10 ml sample volume and soluble cell extracts were obtained as described for the native enzyme.

The protein was purified by affinity chromatography on metal-chelating cellulose (Bioline, London, England) charged with $NiSO_4$. After loading and washing in HIS-BIND buffer, the protein was eluted with 30%(*v*/*v*) HIS-ELUTE buffer (300 m*M* NaCl, 50 m*M* Tris buffer pH 8, 0.3 *M* imidazole) and pooled fractions containing enzyme activity were dialysed into 50 m*M* EPPS buffer pH 8.0, 0.1 m*M* zinc acetate.

2.7. Enzyme assays

The ADHE protein possesses two separate enzymatic activities, both of which were measured spectrophotometrically at 333 K using a Varian Cary 50 Bio UV–visible light spectrophotometer with a Peltier temperature controller (Varian, California, USA). Where enzyme activities obeyed Michaelis–Menten kinetics, kinetic parameters were determined using the direct linear method (Eisenthal & Cornish-Bowden, 1974) using the *EnzPack* software program (Biosoft, Cambridge, England). Where substrate inhibition appeared to be operating, the data were fitted to the substrate-inhibition equation

$$v = \frac{V_{\max}[S]}{K_{\mathrm{m}} + [S] + \left(\dfrac{[S]^2}{K_{\mathrm{i}}}\right)}$$

using the *Origin* computer program (OriginLab, Massachusetts, USA), where K_{i} is the dissociation constant of the nonproductive enzyme–substrate complex.

Standard assay conditions for AldDH consisted of 50 m*M* citrate buffer pH 6.0, 0.1 m*M* zinc acetate, 0.24 m*M* NADH, 0.14 m*M* acetyl-CoA. Assays were started by the addition of enzyme and the decrease in NADH concentration with time was followed at 340 nm. ADH was similarly assayed using 200 m*M* acetaldehyde in the place of acetyl-CoA. Substrate-independent background rates were measured for all assays.

2.8. Stoichiometry of the ADHE-catalysed reaction

As the product of the AldDH domain is the substrate of the ADH domain of ADHE, assays were designed to measure the metabolic flux to ethanol. That is, an AldDH assay was carried out for 30 s at 333 K and the reaction was stopped by a shift in pH to pH 8.5 through the addition of 1 *M* Tris pH 8.8. The reduction in the NADH concentration was measured at 340 nm and the CoASH released was estimated using 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB) at 412 nm (DTNB ε_{412} = 13 600 M^{-1} cm^{-1}).

2.9. Metal-ion analysis

A PerkinElmer AAnalyst 100 atomic absorption spectrometer (PerkinElmer, Massachusetts, USA) was used to investigate the presence of zinc and iron within samples of the ADH-domain protein. The instrument was calibrated as required with zinc standard solutions at 0.2, 0.4, 1 and 2 p.p.m. or iron standard solutions at 2.5, 5 and 7.5 p.p.m. Protein samples were exhaustively dialysed at 277 K into 50 m*M* EPPS pH 8.0.

Samples of protein were also analysed by scanning electron microscope–energy dispersive X-ray spectroscopy (SEM–EDS) using a JSM6480LV scanning electron microscope (Jeol, Welwyn Garden City, England) fitted with an INCA X-ray analyser (Oxford Instruments, Abingdon, England). Protein samples, purified and dialysed as above, were dried onto graphene-coated sample mounts and the residue was analysed. The instrument was optimized using a copper standard at the same energy settings as the samples.

2.10. Crystallization conditions

Prior to crystallization-condition screening, protein samples were concentrated using a Vivaspin 5K MWCO centrifugal filter device (Sartorius) and were then filtered using an Ultrafree centrifugal filter device (0.45 µm; Millipore, Massachusetts, USA).

Purified ADH-domain protein crystals were grown at a protein concentration of 5.4 mg ml^{-1} diluted 1:1 with well solution [0.1 *M* sodium acetate pH 5.0, 0.1 *M* ammonium sulfate, 0.3 *M* sodium formate, 11.5%(*v*/*v*) PEG 2K MME, 3–3.5%(*v*/*v*) PGA-LM]. The crystals produced were flat and diamond-shaped in morphology and were approximately 0.2 × 0.2 × 0.05 mm in size. The crystals took one week to appear at 289 K and growth stopped after two weeks. Glycerol

[30%(*v*/*v*) diluted in well solution] was used as a cryoprotectant when the crystals were cooled for data collection.

2.11. X-ray data collection and structure solution

Data were collected on beamline I03 at the Diamond Light Source at a wavelength of 0.98 Å with an oscillation angle of 0.5° at 100 K. Data from two crystals were processed and scaled using the *HKL*-2000 software package (Otwinowski & Minor, 1997), with one crystal giving complete data to 3 Å resolution and the other giving data that were less complete but that extended to higher resolution. The data were submitted to *BALBES* (Long *et al.*, 2008), which used the structure of *E. coli* lactaldehyde reductase (PDB entry 1rrm; 32% identity to the ADH sequence; New York SGX Research Center for Structural Genomics, unpublished work) as a model to solve the structure by molecular replacement. This solution was improved automatically using *ARP*/*wARP* (Langer *et al.*, 2008) before manual model building with *Coot* (Emsley *et al.*, 2010), refinement using *REFMAC*5 in *CCP*4*i* (Murshudov *et al.*, 2011; Potterton *et al.*, 2003; Winn *et al.*, 2011) and final refinement and model evaluation using *PHENIX* (Adams *et al.*, 2010) and *MolProbity* (Chen *et al.*, 2010). In this final refinement the 'Optimize X-ray/Stereochemistry weight' option was selected as using the default parameters (lowering the weight on the stereochemistry) resulted in a significantly lower *R* factor and insufficiently constrained geometry but had little effect on $R_{\rm free}$. Structure-based sequence alignment of ADH with other ADHE sequences and those ADH structures (downloaded from the PDB) sharing more than 30% sequence identity was carried out using *SALIGN* (Braberg *et al.*, 2012).

2.12. Creation of a homology-modelled structure of the AldDH domain

A homology model of the AldDH domain of the *G. thermoglucosidasius* ADHE protein (amino acids 1–458) was generated using *MODELLER* (Sali & Blundell, 1993) based on two homologous structures: PDB entries 3my7 (an AldDH domain of an ADHE from *V. parahaemolyticus*; 46% identity; Midwest Center for Structural Genomics, unpublished work) and 3k9d (a probable AldDH from *Listeria monocytogenes*; 41% identity; New York SGX Research Center for Structural Genomics, unpublished work). The nine N-terminal amino acids of the modelled AldDH domain of ADHE appeared to be unstructured when compared with the homologous structures and therefore the protein was truncated to Val10 for the subsequent docking work. The protein was modelled as a dimer, this being observed for other proteins of a similar fold, with the intersubunit interaction involving the C-terminal face of each polypeptide.

2.13. Creation of a complete ADHE model using docking programs

The modelled AldDH dimer was 'docked' with the structure of the dimeric ADH domain using three different protein-interaction prediction programs: *ClusPro* (Comeau *et al.*, 2004), *HEX* (Ritchie & Kemp, 1999) and *ZDOCK* (Sali & Blundell, 1993). To elucidate any biologically relevant interactions without bias, no restrictions were imposed on the modelled interactions in terms of termini locations or interaction face.

3. Results

3.1. Native ADHE enzyme

AldDH and ADH activities were assayed in unfractionated TM242 cell extracts and their kinetic parameters were determined: AldDH, specific activity = 1.2 ± 0.1 U mg^{-1}, $K_{\rm m(acetyl\text{-}CoA)} = 19 \pm 3$ μ*M*, $K_{\rm m(NADH)} = 0.16 \pm 0.01$ m*M*; ADH, specific activity = 2.0 ± 0.1 U mg^{-1}, $K_{\rm m(acetaldehyde)}$ = 34 ± 2 m*M*, $K_{\rm m(NADH)} = 38 \pm 3$ μ*M*. It should be noted that AldDH appeared to exhibit substrate inhibition with respect to acetyl-CoA and therefore the data were fitted to the substrate-inhibition equation (see §2); the determined $K_{\rm i}$ was found to be 0.44 ± 0.08 m*M*. The stoichiometry of the ADHE reaction in these cell extracts was determined to be 1.8 ± 0.1 mol NADH oxidized per mole of CoASH produced. Finally, it was found that a range of different divalent metal ions were capable of stimulating the ADH activity of ADHE in cell extracts to varying degrees. These included zinc, magnesium, copper, cobalt, manganese, iron(II) and, to a limited extent, nickel. No such stimulation was observed for the AldDH activity of ADHE.

The ADHE protein was purified to ∼97% homogeneity from a TM242 cell extract by anion-exchange chromatography and gel filtration. The dominant protein species, as assessed by SDS–PAGE, was a 96 kDa protein corresponding to the predicted protein product of the *adhe* gene. It was noticed that the AldDH activity (6.2 ± 0.4 U mg^{-1}) was preferentially lost during purification compared with the ADH activity. This loss did not appear to be due to the removal of other AldDH activities from the cell extract; rather, it was probably due to thiol oxidation, as the presence of reducing agents partially protected the enzyme activity. A catalytic thiol residue in the active site of the AldDH domain may help to explain the apparent oxygen-sensitivity of this protein, with Cys257 being highly conserved between the TM242 ADHE and several other ADHE proteins discussed in the literature (Chen *et al.*, 2004; Espinosa, 2001). The specific activity of the ADH activity of the purified ADHE was 51 ± 2 U mg^{-1} and the $K_{\rm m(acetaldehyde)}$ was 80 ± 7 m*M*. The latter value is twice that observed in unfractionated extracts; this may be due to the fact that the AldDH domain has lost substantial activity or that there was activity from an additional ADH in the cell extracts (the *G. thermoglucosidasius* genome contains 13 genes annotated as alcohol dehydrogenases).

3.2. Multimeric assembly of ADHE

Both unfractionated and purified ADHE eluted from a Superdex 200 gel-filtration column in the void volume, suggesting a molecular weight of greater than 1.3×10^6. DLS analysis showed that the purified protein sample was virtually

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the outer shell.

Space group	$P2_12_12$
Unit-cell parameters (Å, °)	a = 73.721, b = 96.588, c = 58.200, $\alpha = \beta = \gamma$ = 90.00
Resolution (Å)	50.00–2.50 (2.54–2.50)
Total No. of reflections	567074
No. of unique reflections	13897
Completeness (%)	93.5 (77.0)
Multiplicity	6.3 (3.1)
$\langle I/\sigma(I)\rangle$	14.7 (2.1)
$R_{\rm merge}$	0.090 (0.393)
Overall B factor from Wilson plot (Å^2)	49.0
Structure refinement	
Resolution range (Å)	48.29–2.50
No. of reflections, working set	13863
Reflections in test set (%)	5
Final $R_{\rm cryst}$	0.1744
Final $R_{\rm free}$	0.2396
No. of non-H atoms	
Protein	3560
Water	27
Ions, ligands	12
R.m.s. deviations	
Bonds (Å)	0.004
Angles (°)	0.713
B factors (Å^2)	
Average	51.16
Protein main chain	48.63
Protein side chain	53.76
Water	48.03
Ramachandran plot	
Favoured regions (%)	94.22
Additionally allowed regions (%)	99.75
Outliers (%)	0.25
MolProbity score	1.83 [98th percentile, n = 6723, 2.504 ± 0.25 Å]
Clashscore	2.02 [100th percentile, n = 262, 2.504 ± 0.25 Å]

monodisperse and NanoSight analysis detected large protein particles in the 50–300 nm range, with the peak at approximately 200 nm. Several attempts were made to crystallize the purified native ADHE, but despite testing a range of crystallization conditions no protein crystals were obtained. At this point, therefore, it was decided to work with the recombinant protein.

3.3. Separation of ADHE into AldDH and ADH domains

The ADHE protein was expressed recombinantly in various *E. coli* protein-expression strains using the pET28a expression vector to incorporate an N-terminal His tag to facilitate purification. When attempts to investigate the kinetic parameters of the partially purified recombinant protein were undertaken, it became apparent that the AldDH domain of the protein was poorly active (<2% of the activity of the ADH domain of the protein). Given our inability to produce a recombinant ADHE with full AldDH activity, it was decided to resolve the enzyme into its two respective activities for further investigation. A variety of gene fragments corresponding to what was predicted to be the AldDH and ADH domains of ADHE were subcloned for expression in *G. thermoglucosidasius* and characterization of the recombinant proteins. Despite a range of AldDH fragments being produced, none were produced in an active form. However, several active ADH fragments were produced and a minimal functional unit corresponding to amino acids 459–869 was generated; this protein was then expressed recombinantly in *E. coli* with an N-terminal His tag to aid its purification.

Kinetic parameters were determined for the purified ADH domain: $V_{\rm max}$ = 430 ± 10 U mg^{-1}, $K_{\rm m(acetaldehyde)}$ = 121 ± 5 m*M* and $K_{\rm m(NADH)}$ = 62 ± 1 µ*M*. The differences observed in the $K_{\rm m}$ and $V_{\rm max}$ values between the purified native ADHE protein and the recombinantly produced ADH domain are not totally unexpected due to the removal of the N-terminal AldDH domain and the addition of the His tag in the recombinant enzyme. However, the increase in $V_{\rm max}$ is considerable, with the $k_{\rm cat}$ values for the two proteins being 83 s^{-1} for the purified ADHE and 350 s^{-1} for the recombinant ADH fragment.

The assembly of this recombinant ADH domain was investigated using gel-filtration and light-scattering techniques, giving molecular-weight values of 86 000 and 99 000, respectively. From the subunit molecular weight of 48 600, these values correspond to 2.0 and 1.8 polypeptides, showing that the protein is dimeric and that, unlike the native ADHE, it does not assemble into spirosome structures.

As found for the native ADHE, the activity of the recombinant ADH fragment was increased upon the addition of 0.1 m*M* zinc acetate. Iron(II) also stimulated this fragment, but to a lesser extent than zinc, even when the concentration was tenfold higher (1 m*M*). However, modulation of activity was not observed post metal-affinity purification, suggesting that the protein is capable of scavenging divalent metal ions from the metal column.

Prior to the analysis of the metal-ion content, thrombin protease was used to cleave the His tag from the ADH domain as it has been shown that His tags on recombinant proteins can be involved in the binding of zinc ions (Evers *et al.*, 2008). No loss of activity was found on removal of the tag. Atomic absorption spectroscopy analysis of this protein revealed a ratio of 0.43 Zn atoms per protein monomer. No iron was detected within either the thrombin-cleaved or uncleaved protein samples using atomic absorption spectroscopy or SEM–EDS.

3.4. Solution and refinement of the ADH-fragment crystal structure

X-ray diffracting crystals were obtained for the ADH fragment and a 2.50 Å resolution data set was collected at the Diamond Light Source, Oxford, England. Data-collection and subsequent refinement statistics are described in Table 1. The model was subjected to several rounds of refinement and validation, and 27 water molecules, a sulfate ion, a Zn^{2+} ion with 0.69 occupancy and a glycerol molecule were all added. The final model excluded part of one loop in the structure (Phe766–Glu773) as there was not sufficient density for the conformation to be determined. Density for the N-terminal His tag was also not observed in the data; the last residue with missing density is immediately prior to the first coding residue

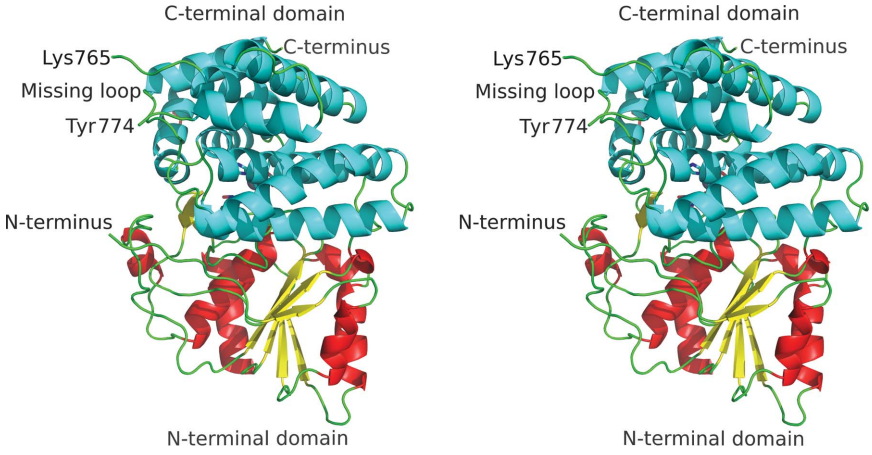

Figure 1
A cartoon stereoview of the ADH-domain crystal structure drawn using *PyMOL* (v.1.2r3pre; Schrödinger). Spirals represent α-helices (red for the N-terminal domain and cyan for the C-terminal domain) and yellow arrows represent β-strands. Termini and the missing C-terminal domain loop region are indicated. A zinc ion (grey sphere), glycerol (blue) and a sulfate molecule (yellow) are also shown in the structure.

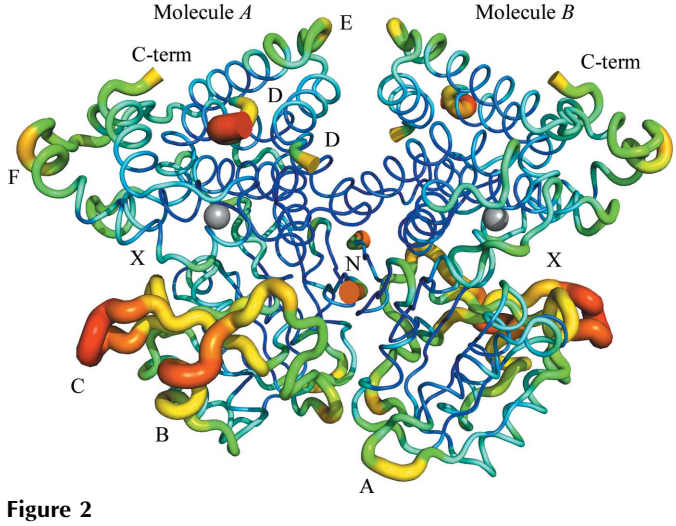

Figure 2
Cartoon diagram for the main interface between ADH-domain molecules within the dimer visualized by temperature factor. Wider red regions indicate increased mobility compared with thinner blue regions, which indicate limited mobility. The labels A–F correspond to loop identities (the termini of the missing loop are both labelled C). Visible termini (N and C-term) are indicated in the image. Molecules *A* and *B* are indicated and the active-site metal ion is shown in grey. The predicted active-site clefts are labelled X.

of the ADH domain of ADHE. The crystallographic data and model have been deposited in the PDB as entry 3zdr.

3.5. Overview of the ADH-domain structure

The ADH domain is composed of two structural domains (Fig. 1). The N-terminal domain has a three-layer ($\alpha\beta\alpha$) sandwich or Rossmann-fold architecture (CATH code 3.40.50.1970) typical of NAD(P)$^+$ cofactor-binding domains (Rossmann *et al.*, 1974). The C-terminal domain is α-helical, with an up–down bundle architecture known as a dehydroquinate synthase-like α-domain (CATH code 1.20.1090.10).

The dimeric structure observed by gel filtration and DLS was also present within the unit cell of the crystals, generated from the monomer in the asymmetric unit by the twofold axis parallel to *Z* (space group $P2_12_12$). On coming together in the dimer, the first few amino acids at the N-terminus of this domain form a β-strand that lies parallel to and extends the β-sheet formed from six strands in the other monomer. According to *PISA* analysis (Krissinel & Henrick, 2007), a significant interaction was observed along the N-terminal side of the protein (shown in Fig. 2) that is coordinated by six salt bridges and seven hydrogen bonds; the interface area is 1777.2 Å^2 and the Δ^iG is -26.7 kcal mol^{-1}.

Fig. 2 shows that the region around the predicted N-terminal dimerization interface has limited mobility, whereas some of the outer loop regions are significantly more mobile, and the mobile loops congregate on one surface of each monomer. The most mobile region of the protein is the visible part of loop D (Lys762–Lys765 and Tyr774–Lys776), which has an average temperature factor of 77.1 Å^2, followed by loop C (Ile619–Pro629, the central part of which is present, but poorly defined; it contains Thr624, the single Ramachandran plot outlier in the structure) with 76.0 Å^2 and loop B (Thr571–Lys599) with a mean of 71.1 Å^2; the average of the whole molecule is 51.2 Å^2. The relatively large difference observed between *R* and $R_{\rm free}$ for the structure may in part be due to the limited resolution, but is also likely to be due to flexibility in these loop regions of the protein as indicated by the high temperature factors observed. The only significant difference map peak is in the vicinity of His730, one of the metal-coordinating residues, suggesting that this should be in multiple conformations; however, refinement as such resulted in an increase in $R_{\rm free}$, so this was not retained in the final model.

A search of the Protein Data Bank with the ADH-domain sequence revealed a group of structures of between 33 and 35% sequence identity and a clearly distinct group of less than 25% identity. To identify features unique to an ADH domain within an ADHE structure that therefore might be involved in the interface between the two components of the bifunctional enzyme, a structure-based sequence alignment of ADH with other ADHE ADH domains and the more similar group of structures was carried out (Fig. 3). Across the alignment there is clear conservation of large areas of the sequences and structures (secondary structure shown for a representative ADH structure). However, the regions corresponding to the missing part of loop D and the mobile loop B contain major insertions found only in the ADHs that form part of the ADHE dual functional proteins (gaps boxed in Fig. 3). Indeed, the 'missing residues' (Phe766–Glu773; in bold and outlined in Fig. 3), which are too mobile to be seen in this single-domain structure and show some conservation in other ADHE

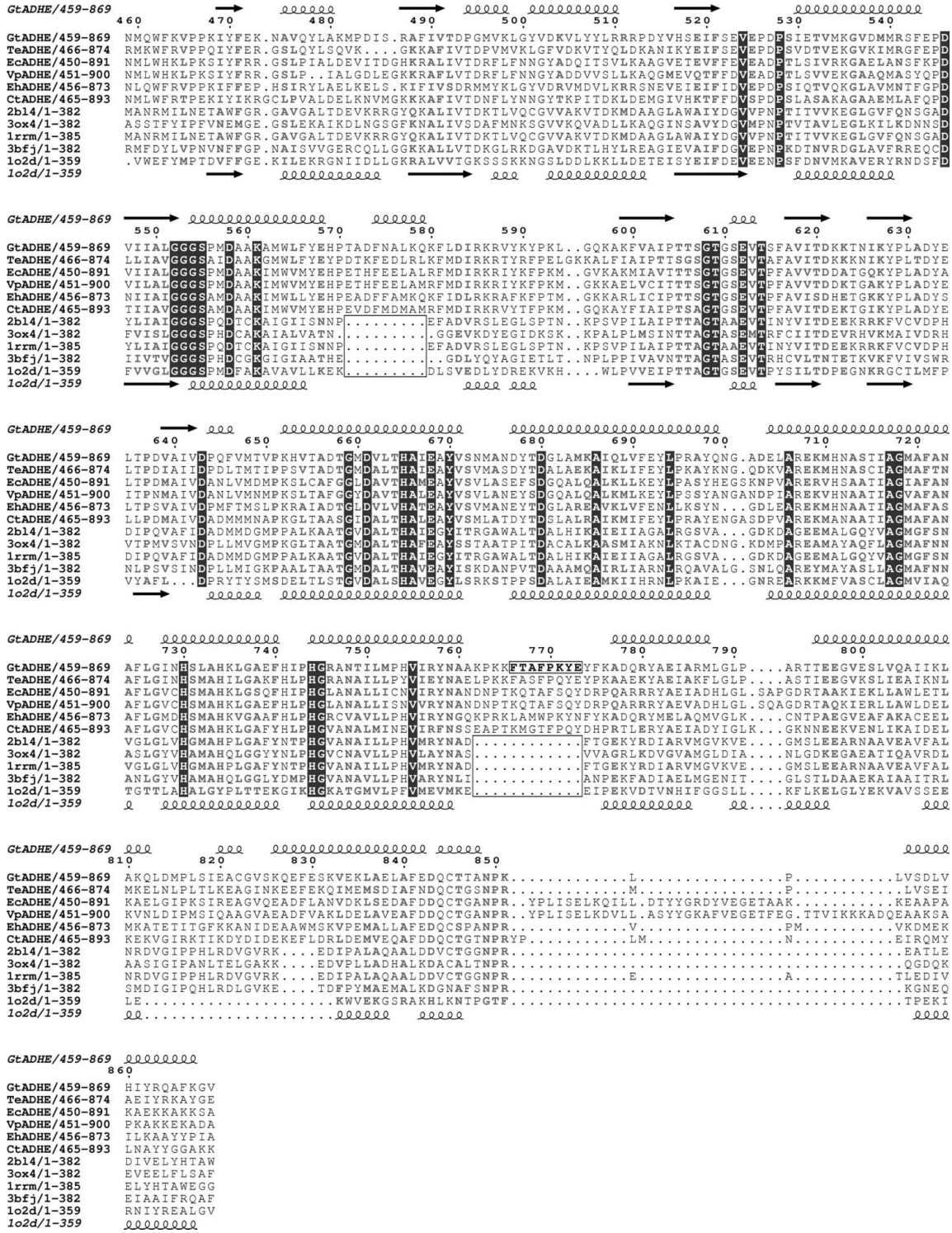

Figure 3
Structure-based sequence alignment carried out using *SALIGN* (Braberg *et al.*, 2012) and displayed with *ESPript* (Gouet *et al.*, 1999) of the ADH domains of ADHE from *G. thermoglucosidasius* C56-YS93 (AEH49709.1; GtADHE), *Thermoanaerobacter ethanolicus* (ABH06551.1; TeADHE), *Escherichia coli* (NP_415757.1; EcADHE), *Vibrio parahaemolyticus* RIMD 2210633 (NP_798500.1; VPADHE), *Entamoeba histolytica* (Q24803; EhAHE) and *Clostridium thermocellum* ATCC 27405 (YP_001036854.1; CtADHE) with the sequences of other ADHs from the PDB sharing more than 30% identity with *Geobacillus* ADHE ADH: *E. coli* FucO (PDB entry 2bl4; Montella *et al.*, 2005), *Zymomonas mobilis* ADH (PDB entry 3ox4; Moon *et al.*, 2011), *E. coli* lactaldehyde reductase (PDB entry 1rrm; New York SGX Research Center for Structural Genomics, unpublished work), *Klebsiella pneumoniae* 1,3-propanediol dehydrogenase (PDB entry 3bfj; Marçal *et al.*, 2009), *Thermotoga maritima* ADH (PDB entry 1o2d; Schwarzenbacher *et al.*, 2004). Above the alignment is a schematic representation of the secondary structure of PDB entry 3zdr (ADHE ADH) and below the alignment a similar representation of the structure of *T. maritima* ADH (PDB entry 1o2d) representing the single-domain ADH structures. Identical residues across all sequences are white on a black background, those not visible in the ADH-domain structure are in bold and boxed and the gaps in the two loops that are shorter in the single-domain ADH structures are boxed.

sequences, are missing entirely in single-domain structures, suggesting that they are important for interactions with the other AldDH domain of ADHE proteins.

3.6. Metal-ion coordination in the ADH domain

The probable location of the active site of ADH is between the Rossmann fold and the α-helical domain. Strong positive difference electron density for an octahedrally coordinated metal ion was observed between an aspartic acid (Asp661), three histidine residues (His665, His730 and His744) that are all part of the α-helical domain of the protein and a glycerol molecule (Fig. 4).

As Zn^{2+} was detected by ion analysis but Fe^{2+} was not, Zn^{2+} was modelled into the X-ray structure of the protein rather than the Fe^{2+} that is more commonly found in deposited ADH structures. Also, in agreement with the ion analysis, fully occupying the site with a Zn atom appeared to overaccount for the observed density. On refinement, the Zn^{2+} occupancy dropped to 0.69 with a temperature factor comparable with those of the metal-coordinating atoms and this lowered the negative difference density peak significantly (Fig. 4). Partial occupancy of the metal-binding site is also suggested by the indication of low-occupancy multiple conformations of His730 and His738, since in the absence of a metal such close proximity of His side chains would be unfavourable. The sixth metal-coordination site was occupied by a species larger than water and was modelled as the cryoprotectant glycerol (Fig. 4). The terminal alcohol group in glycerol is similar to that of ethanol and may be a product mimic for this ADH domain.

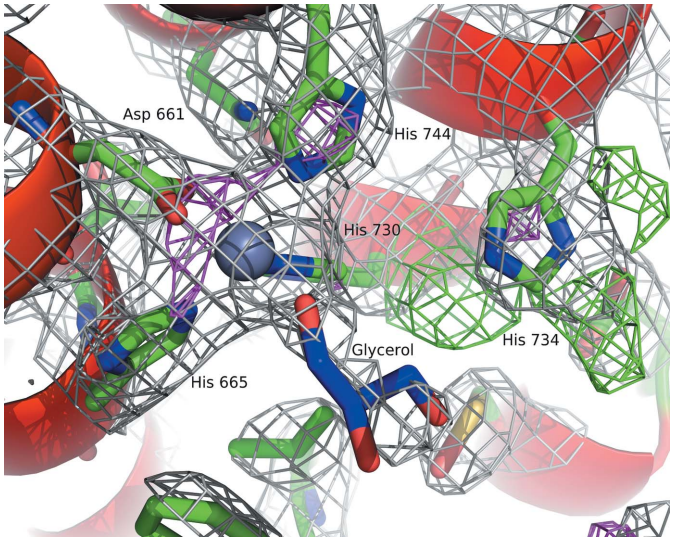

Figure 4
The metal ion-binding site in the ADH-domain crystal structure, overlaid with electron density. Coordinating amino acids are represented in stick form with green C atoms; the zinc ion (grey sphere) and glycerol molecule (blue C atoms) are indicated with α-helices shown as red spirals. Grey mesh represents the electron density from the final $2F_o - F_c$ map contoured at 1σ; green mesh represents the positive difference $(F_o - F_c)$ density and magenta mesh the negative difference density contoured at $+3\sigma$ and -3σ, respectively. Some parts of the density and structure are excluded for clarity.

3.7. Generation of a homology model of the AldDH domain

Attempts to crystallize the complete ADHE protein and the AldDH domain were unsuccessful. *In silico* modelling was carried out to predict the possible interactions between the AldDH and the ADH domains of the ADHE protein. A homology model of the AldDH domain (amino acids 1–458 of ADHE) of the *G. thermoglucosidasius* ADHE protein was generated using *MODELLER* (Sali & Blundell, 1993). The protein was modelled as a dimer as observed for the proteins of similar fold, with the C-termini, which would be joined to the ADH domain in ADHE, on the same face (Fig. 5).

As in the two homologous structures on which it was based, two structural domains are present in the modelled AldDH domain of ADHE. Both of the domains are from superfamilies with three-layer ($\alpha\beta\alpha$) sandwich topologies (N-terminal domain, CATH code 3.40.605.10; C-terminal domain, CATH code 3.40.309.10). Evaluation of the model in comparison to the two homologous proteins used to create it showed few significant differences in terms of the energy of the residues within the protein. A similar conclusion was reached following analysis using *MolProbity*.

3.8. Docking of the AldDH model and the ADH domain

The AldDH model dimer was 'docked' with the ADH-domain dimer using three different protein-interaction prediction programs. The top-ranked results for each program are shown in Fig. 6. All the programs predicted a similar interaction face between the AldDH and ADH domains, although the AldDH domain was placed in different positions in each case. These modelling results indicate that this face of the ADH domain of the protein is the one that is most likely to form physiologically relevant interactions with an AldDH domain in the native ADHE protein.

The modelled N-terminal AldDH domain (amino acids 1–458) and the structure of the C-terminal ADH domain (amino acids 459–869) account for the whole sequence of ADHE. Therefore, in the docked structure, the correct positioning of

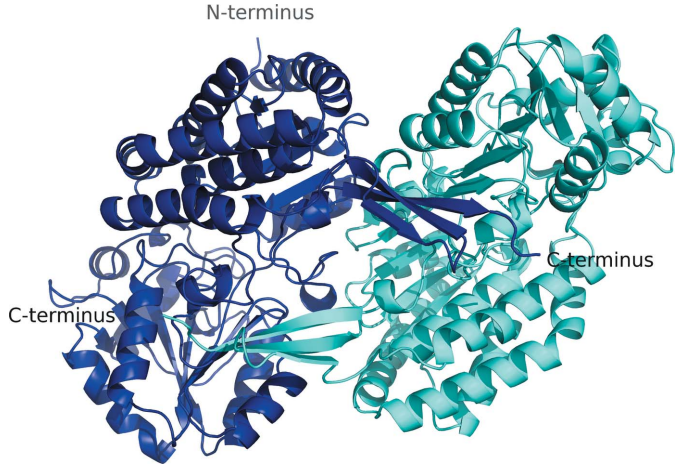

Figure 5
Cartoon diagram of the modelled dimeric AldDH. Dark blue, molecule *A*; cyan, molecule *B*. Termini that are visible in the figure are indicated.

Table 2
Summary of *PISA* analysis.

Δ^iG is the solvation free-energy gain upon interface formation (where negative indicates a hydrophobic interface), with the Δ^iG *P*-value of the solvation free-energy gain being a measure of the interface specificity, where $P > 0.5$ indicates nonspecific interfaces and $P < 0.5$ indicates interfaces with higher than average hydrophobicity that may be considered interaction-specific. HB, number of hydrogen bonds formed; SB, number of salt bridges formed. 1 cal = 4.184 J.

Model generated with	Interface area (Å^2)	Δ^iG (kcal mol^{-1})	Δ^iG *P*-value	HB	SB
ZDOCK	2026.5	−16.7	0.452	13	0
HEX	1798.7	−1.9	0.623	9	6
ClusPro	1704.4	−6.2	0.990	0	21

the two domains should result in the C-terminal residue of AldDH being close to the N-terminus of ADH. The *ZDOCK* model positioned the AldDH C-terminus 7.4 Å away from the ADH N-terminus, while the other two, although on the same face, were more distant.

PISA analysis was carried out on the top-rated model from each program used to evaluate the predicted interfaces between the two domains of ADHE (Table 2). Although none of the models had a complex-formation significance score high enough to show an interface relevant to complex formation (data not shown), this analysis implies that the *ZDOCK* model is the most likely of the models to be interaction-specific for the ADHE protein (the lowest Δ^iG *P*-value). Although the interaction surface identified by *ClusPro* includes more salt bridges, there are no hydrogen bonds in this interface, and the solvation energy *P*-value suggests an interface that is no better than random. As all three prediction methods identified an interaction between the same faces of the domain, but the *ZDOCK* model brings the termini closest together and has the better *PISA* scores (although the low significance of the interaction suggests this model is not sufficiently accurate to predict interactions at the amino-acid level), subsequent evaluation of the structure of ADHE has been based on the *ZDOCK* model.

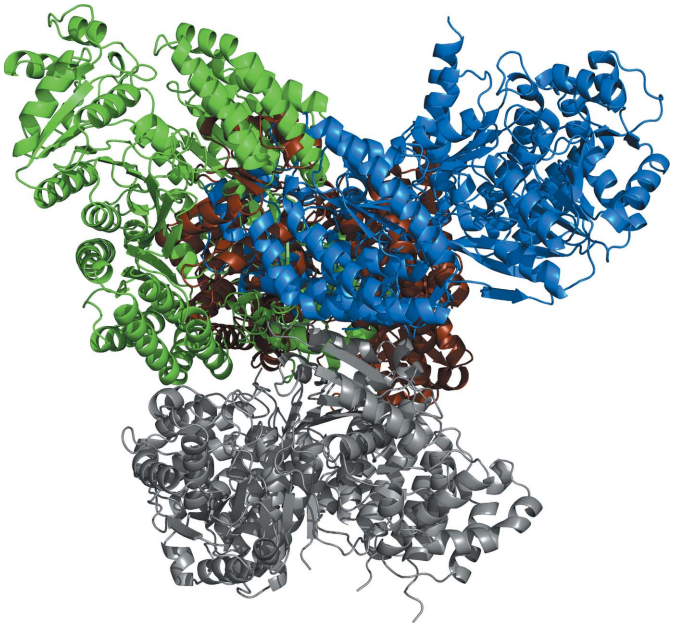

Figure 6
Cartoon overview of the predicted interactions between the modelled AldDH and the crystal structure of the ADH-domain dimers. Grey, ADH domain; brown, modelled AldDH (top *HEX* result); blue, modelled AldDH (top *ClusPro* result); green, modelled AldDH (top *ZDOCK* result).

3.9. Interactions between AldDH and ADH domains in the 'ADHE' model

The *ZDOCK* model of the complete ADHE structure shows interactions between both ADH-domain loop D (15 amino acids, the most mobile loop, in which the central part is too mobile to be observed) and loop B (24 amino acids, the third most mobile loop) and the AldDH domain. Although the second most mobile loop observed in the ADH-domain crystal structure (loop C; eight amino acids) appears to be relatively distant from the proposed interface, the mobility of this loop may be sufficient to allow interaction. As ADHE is a thermostable protein, the complete structure is unlikely to contain long highly mobile loops as these would initiate unfolding, so our hypothesis is that loops B and D and possibly loop C, the three most mobile loops, are located in the proposed interface and that these loops are stabilized through interaction with the AldDH domain in the intact ADHE, the mobility observed being an artefact of analysis of the ADH domain in isolation. If correct, the proximity of the docked AldDH domain to these loops could be used as another indication of having identified the correct docking interface.

The sequence of the missing loop in the ADH-domain structure is KPKK**FTAFPKYE**YFK (the eight missing residues in the structure are shown in bold and are highlighted in Fig. 3) and can be seen to be positively charged and enriched

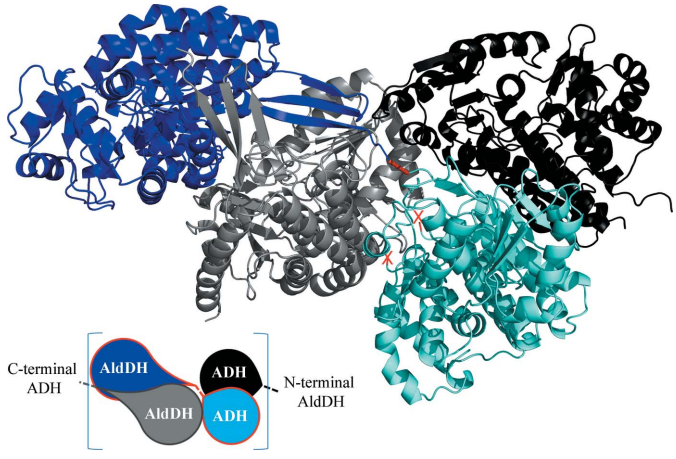

Figure 7
Cartoon overview diagram of the top result for the predicted interaction between modelled AldDH-domain and ADH-domain dimers using *ZDOCK*. Dark blue, modelled AldDH of ADHE 1; light blue, ADH domain of ADHE 1; grey, modelled AldDH monomer; black, ADH-domain monomer. The hypothesized link between the C-terminus of the AldDH and the N-terminus of ADH domain is shown in red. Red crosses indicate the termini of the truncated loop in the ADH-domain structure. A schematic diagram in the same colour scheme is shown in which one ADHE monomer is outlined in red.

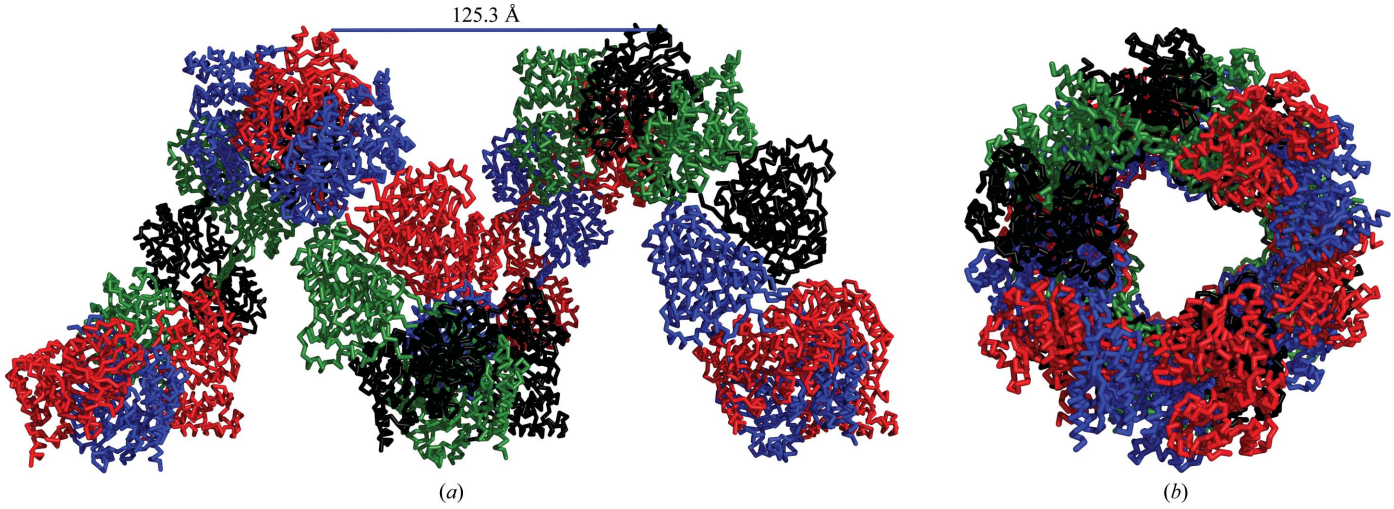

Figure 8
C^{α} trace of ADHE assembly based on the *ZDOCK* model in (*a*) a side-on view and (*b*) an end-on view. The colours of the predicted ADHE monomers alternate through the figure (right to left: red–blue–black–green . . .). Approximately seven monomers make up a whole turn, taking 125 Å to complete.

in aromatic residues. To identify possible stabilizing residues at the predicted interface of the two domains, the solvent-accessible areas for residues of the modelled AldDH dimer were calculated using *AREAIMOL* in *CCP*4*i* (Lee & Richards, 1971; Winn *et al.*, 2011). Negatively charged amino acids and aromatic amino acids that had an accessible solvent area greater than 20 Å were mapped onto the protein structure. Several negative amino acids and aromatic residues are present close to the modelled interface between the missing loop in the ADH domain and the interacting AldDH monomer. This indicates that the positively charged aromatic-enriched loop could feasibly be stabilized by residues in the interaction area proposed by the model.

It was not possible to determine the likelihood of substrate channelling between the two domains of ADHE using the model produced here; this must await a high-resolution crystal structure of the intact ADHE protein.

3.10. Domain-dimerization modes suggest a method of spirosome formation

Although both the AldDH-domain and the ADH-domain fragments form dimers, in the former the termini lie on the same face of the protein (*i.e.* they are related by a twofold axis perpendicular to the interface) while they are on opposite faces in ADH (the twofold axis is parallel to the interface). Other AldDH and ADH structures in the PDB retain this difference in dimerization modes. This suggests that the AldDH and ADH domains of a single ADHE may form dimers with different ADHE monomers rather than both with the same molecule (Fig. 7). When the interactions predicted between the two domains of ADHE shown in Fig. 7 are extrapolated, a right-handed helical assembly of ADHE monomers can be formed with seven subunits per turn (Fig. 8). This is consistent with the large multimeric assemblies of ADHE monomers that were observed during characterization of the native *G. thermoglucosidasius* ADHE. Without a high-resolution structure of the ADHE protein coupled with an electron-density map of the spirosome assemblies, the models described here remain purely speculative; however, this is the first structural hypothesis for the spirosome structure.

4. Discussion

The structure of the ADH domain of the *G. thermoglucosidasius* ADHE protein has been determined using X-ray crystallography to 2.5 Å resolution. This is the first reported ADH-domain structure of an ADHE enzyme. The protein has been shown to consist of an NAD^{+}-binding domain (Rossmann fold) and an α-helical domain containing residues that are coordinating a metal ion. This metal ion is likely to be catalytic in nature due to its positioning at the interface between the two domains and the increased rate of catalysis observed for the ADH domain in the presence of various divalent metal ions. The structure of the N-terminal AldDH domain of ADHE was not determined, thus preventing structural analysis of this protein. However, a model of the AldDH domain of ADHE has been created to allow *in silico* prediction of interactions between the two domains of ADHE.

In contrast to our data for the *Geobacillus* ADHE and the ADH domain, the ADH activity of *E. coli* ADHE is only stimulated by the presence of Fe^{2+} and not by other metal ions (Kessler *et al.*, 1991). This is also the case for the enzymes from *Entamoeba histolytica* and *Streptococcus bovis* (Asanuma *et al.*, 2004; Espinosa *et al.*, 2009); moreover, in these cases the dehydrogenase activities of ADHE were inhibited by the presence of other divalent metal ions, whereas in the *Geobacillus* ADHE divalent metal ions stimulated the enzyme. Other investigations into ADHE enzymes have not described the Fe^{2+}-dependence, and significant activity is found in the absence of added metal ions (Fontaine *et al.*, 2002; Koo *et al.*, 2005; Pei *et al.*, 2010; Sánchez, 1998). By contrast, the majority of single-domain ADH structures that have been deposited in the PDB, and which share more than 30% sequence identity

with the ADH domain of ADHE, contain Fe. However, the structures of *Thermotoga maritima* TM0920 and *E. coli* lactaldehyde:1,2-propanediol oxidoreductase (FucO) have been deposited twice, each with the metal ion being Zn in the first structure and then changing to Fe in the second without comment. Of all these structures, only in *E. coli* FucO has the metal been shown experimentally to be iron. Even in this case, the enzyme was shown to have a higher affinity for Zn^{2+}, which displaced the bound Fe and inactivated the enzyme (Montella *et al.*, 2005). This weak binding of the metal is also suggested by the 1.3 Å resolution *T. maritima* ADH structure, where the occupancy of the Fe site was refined to 60% (PDB entry 1o2d; Schwarzenbacher *et al.*, 2004). Analysis of the temperature factors of the metal in these similar structures shows that where the identity of the metal is known, for instance in *E. coli* FucO, the metal *B* factors are slightly below the mean for all of the protein atoms. However, in the *Klebsiella pneumoniae* and *Zymomonas mobilis* ADHs, where the identity of the metal was not determined, the temperature factors vary by up to a factor of four, suggesting that although all metal ions are modelled as fully occupied Fe, in fact either a variety of metal ions are present or not all Fe sites are fully occupied. This, combined with the partial occupancy found in the *E. coli* ADH protein by spectroscopy and in *Thermotoga* ADH by structure solution, suggests that a variety of metals can be accommodated in this structure and that they can be exchanged after folding. In the ADHE ADH structure, modelling the metal as a partially occupied Zn^{2+} ion resulted in a temperature factor comparable to the mean, whereas Fe had to be fully occupied to approach a comparable value. This and the absence of Fe by both spectroscopy and microscopy suggest that partially occupied Zn, with possible low-level substitution by other M^{2+} ions, is the most accurate description of the metal ion in the ADHE ADH domain. We suggest that this class of ADH structures should be described as metal-ion-dependent rather than rigidly either iron-specific or zinc-specific.

The role of the divalent metal ion in the active site of the protein is probably to aid polarization of the acetaldehyde carbonyl O atom, allowing reduction by NADH to proceed. The charge density of the metal ion may have an effect on the rate of catalysis due to differences in the strength of the polarization of the carbonyl group. However, several different metal ions may be able to perform this role; the physiologically relevant metal ion present in the ADH domain of *G. thermoglucosidasius* ADHE has not been unambiguously identified.

The unusually high temperature factors associated with several of the loop regions of the ADH-domain protein, coupled with the limited thermostability observed during biochemical characterization (data not reported), indicate that some stabilizing interactions may exist between the two domains of ADHE. Exposed flexible loop regions within proteins can be susceptible to degradation and play a role in the instability of a protein at high temperatures (Nagi & Regan, 1997). It is therefore common to observe that loops in thermophilic proteins are shorter than those in their mesophilic homologues, thereby limiting the flexibility of these regions and thus enhancing stability. This may be through loop shortening and stabilizing electrostatic interactions within a loop (Russell *et al.*, 1997) or by stabilization of the loops through oligomerization (Vieille & Zeikus, 2001). In the case of the *Geobacillus* ADHE, the most flexible loops are also the sites of sequence insertions relative to single-domain ADH molecules (Fig. 3). The availability of many single-domain ADH structures in the PDB, all of which have shorter loops, has helped to emphasize the conservation and hence importance of these longer loops across ADHE sequences and to corroborate the *in silico* modelling that suggests some of these mobile loop regions may indeed be stabilized through interactions with an AldDH domain. Interestingly, this work suggests that such interactions may be intermolecular, rather than intramolecular, *i.e.* between different ADHE monomers. This in turn leads to an explanation for the formation of the spirosome assemblies that have been observed for ADHE proteins.

That the *Geobacillus* ADHE forms large assemblies, probably spirosomes, is shown by gel filtration and DLS. The predicted helical assembly of spirosomes with a right-handed helix with seven ADHE units per turn differs significantly from that reported for the *E. coli* ADHE by Kessler *et al.* (1992), which comprised a left-handed helix of four molecules per turn. Nonetheless, the potential for ADHE proteins to interact to form helical assemblies is an intriguing observation and one which may effect a degree of substrate channelling (Zhang, 2011) that would protect the cell from the reactive ADHE intermediate acetaldehyde. Moreover, substrate channelling between the two domains of ADHE might result in a much lower *in vivo* K_m value of ADH for this acetaldehyde. Clearly, a high-resolution structure of the whole ADHE remains a priority with respect to these possibilities.

The authors wish to thank Professor David Leak and Dr Jeremy Bartosiak-Jentys (University of Bath) and Dr Alex Pudney (TMO Renewables Ltd) for assistance with the manipulation of *G. thermoglucosidasius* strains and for supplying the appropriate vectors for protein expression in this organism. We also thank Mr Alan Carver and Dr John Mitchels (University of Bath) for the atomic absorption spectroscopy and SEM–EDS analyses, respectively, and Dr Bob Carr and Mr Matthew Wright (NanoSight Ltd, Amesbury, England) for the use of the NanoSight instrumentation and assistance with sample analysis. Preliminary modelling work carried out by David Corcoran (University of Bath) is also acknowledged. This work was funded by TMO Renewables Ltd and the BBSRC through a CASE studentship for JE. We thank Diamond Light Source for access to beamline I03 (in project mx1226-23) that contributed to the results presented here.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Arnau, J., Jørgensen, F., Madsen, S. M., Vrang, A. & Israelsen, H. (1998). *J. Bacteriol.* **180**, 3049–3055.
Asanuma, N., Yoshii, T. & Hino, T. (2004). *Arch. Microbiol.* **181**, 122–128.

Atteia, A., van Lis, R., Mendoza-Hernández, G., Henze, K., Martin, W., Riveros-Rosas, H. & González-Halphen, D. (2003). *Plant Mol. Biol.* **53**, 175–188.
Braberg, H., Webb, B. M., Tjioe, E., Pieper, U., Sali, A. & Madhusudhan, M. S. (2012). *Bioinformatics*, **28**, 2072–2073.
Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
Bruchhaus, I. & Tannich, E. (1994). *Biochem. J.* **303**, 743–748.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Chen, M., Li, E. & Stanley, S. L. (2004). *Mol. Biochem. Parasitol.* **137**, 201–205.
Comeau, S. R., Gatchell, D. W., Vajda, S. & Camacho, C. J. (2004). *Nucleic Acids Res.* **32**, W96–W99.
Cripps, R. E., Eley, K., Leak, D. J., Rudd, B., Taylor, M., Todd, M., Boakes, S., Martin, S. & Atkinson, T. (2009). *Metab. Eng.* **11**, 398–408.
Dan, M. & Wang, C. C. (2000). *Mol. Biochem. Parasitol.* **109**, 25–36.
Eisenthal, R. & Cornish-Bowden, A. (1974). *Biochem. J.* **139**, 715–720.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Espinosa, A. (2001). *J. Biol. Chem.* **276**, 20136–20143.
Espinosa, A., Perdrizet, G., Paz-y-Mino C., G., Lanfranchi, R. & Phay, M. (2009). *J. Antimicrob. Chemother.* **63**, 675–678.
Evers, T. H., Appelhof, M. A., Meijer, E. W. & Merkx, M. (2008). *Protein Eng. Des. Sel.* **21**, 529–536.
Filipe, V., Hawe, A. & Jiskoot, W. (2010). *Pharm. Res.* **27**, 796–810.
Fontaine, L., Meynial-Salles, I., Girbal, L., Yang, X., Croux, C. & Soucaille, P. (2002). *J. Bacteriol.* **184**, 821–830.
Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Kessler, D., Herth, W. & Knappe, J. (1992). *J. Biol. Chem.* **267**, 18073–18079.
Kessler, D., Leibrecht, I. & Knappe, J. (1991). *FEBS Lett.* **281**, 59–63.
Koo, O. K., Jeong, D.-W., Lee, J. M., Kim, M. J., Lee, J.-H., Chang, H. C., Kim, J. H. & Lee, H. J. (2005). *Biotechnol. Lett.* **27**, 505–510.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
Lee, B. & Richards, F. (1971). *J. Mol. Biol.* **55**, 379–400.
Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* D**64**, 125–132.
Marçal, D., Rêgo, A. T., Carrondo, M. A. & Enguita, F. J. (2009). *J. Bacteriol.* **191**, 1143–1151.
Membrillo-Hernandez, J., Echave, P., Cabiscol, E., Tamarit, J., Ros, J. & Lin, E. C. (2000). *J. Biol. Chem.* **275**, 33869–33875.
Montella, C., Bellsolell, L., Pérez-Luque, R., Badía, J., Baldoma, L., Coll, M. & Aguilar, J. (2005). *J. Bacteriol.* **187**, 4957–4966.
Moon, J.-H., Lee, H.-J., Park, S.-Y., Song, J.-M., Park, M.-Y., Park, H.-M., Sun, J., Park, J.-H., Kim, B. Y. & Kim, J.-S. (2011). *J. Mol. Biol.* **407**, 413–424.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Nagi, A. D. & Regan, L. (1997). *Fold. Des.* **2**, 67–75.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pei, J., Zhou, Q., Jiang, Y., Le, Y., Li, H., Shao, W. & Wiegel, J. (2010). *Metab. Eng.* **12**, 420–428.
Peng, H., Wu, G. & Shao, W. (2008). *Anaerobe*, **14**, 125–127.
Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). *Acta Cryst.* D**59**, 1131–1137.
Ritchie, D. W. & Kemp, G. J. L. (1999). *J. Comput. Chem.* **20**, 383–395.
Rossmann, M. G., Moras, D. & Olsen, K. W. (1974). *Nature (London)*, **250**, 194–199.
Russell, R. J., Ferguson, J. M., Hough, D. W., Danson, M. J. & Taylor, G. L. (1997). *Biochemistry*, **36**, 9983–9994.
Sali, A. & Blundell, T. L. (1993). *J. Mol. Biol.* **234**, 779–815.
Sánchez, L. B. (1998). *Arch. Biochem. Biophys.* **354**, 57–64.
Schwarzenbacher, R. *et al.* (2004). *Proteins*, **54**, 174–177.
Shaw, A. J., Jenney, F. E. Jr, Adams, M. W. W. & Lynd, L. R. (2008). *Enzyme. Microb. Tech.* **42**, 453–458.
Vieille, C. & Zeikus, G. J. (2001). *Microbiol. Mol. Biol. Rev.* **65**, 1–43.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Zhang, Y.-H. P. (2011). *Biotechnol. Adv.* **29**, 715–725.