Crystal structure and active site location of N-(1-D-carboxylethyl)-L-norvaline dehydrogenase

K.L. Britton[1], Y. Asano[2] and D.W. Rice[1]

Opine dehydrogenases catalyze the NAD(P)H-dependent reversible reaction to form opines that contain two asymmetric centers exhibiting either (L,L) or (D,L) stereochemistry. The first structure of a (D,L) superfamily member, N-(1-D-carboxylethyl)-L-norvaline dehydrogenase (CENDH) from *Arthrobacter* sp. strain 1C, has been determined at 1.8 Å resolution and the location of the bound nucleotide coenzyme has been identified. Six conserved residues cluster in the cleft between the enzyme's two domains, close to the nucleotide binding site, and are presumed to define the enzyme's catalytic machinery. Conservation of a His-Asp pair as part of this cluster suggests that the enzyme mechanism is related to the 2-hydroxy acid dehydrogenases. The pattern of sequence conservation and substitution between members of this enzyme family has permitted the tentative location of the residues that define their differential substrate specificities.

Opines are the products of the NAD(P)H-dependent reductive condensation between an α-keto acid and the α- or ω-NH_2 group of an amino acid in a reaction catalyzed by a family of enzymes generically referred to as opine dehydrogenases. The products of these reactions have two asymmetric centers and, in nature, opines may exhibit either (L,L) or (D,L) stereochemistry. Compounds belonging to this class have been isolated from eukaryotic cells, plant tumors, bacteria and muscle tissue of marine invertebrates (reviewed in ref. 1).

Amino acid sequences are currently available for five opine dehydrogenases. Three of these enzymes are responsible for the synthesis of opines with (D,L) stereochemistry (nopaline, lysopine and N-(1-D-carboxylethyl)-L-norvaline dehydrogenase)[2–4] whilst the remaining two catalyze the synthesis of opines with (L,L) stereochemistry (saccharopine dehydrogenase and carboxylethyl ornithine synthase)[5,6]. Analysis of the amino acid sequences and biochemical studies have established that the enzymes that catalyze the (D,L) and (L,L) chemistry belong to two apparently distinct superfamilies with differential specificity for the keto acid and amino acid partners. The enzymes within one superfamily display differential substrate specificity but share sequence similarities in the range of 20–30% identities, whereas there is no detectable sequence similarity between the two superfamilies. This implies that the three-dimensional structure within a superfamily but not between superfamilies will be similar.

The lysopine and nopaline families of opines found in crown gall tumors are the most extensively studied of the naturally occurring opines[7]. Such tumors are widespread on many dicotlodenous plants[8] and these hyperplasias were first described by Aristotle[8]. However, it was not until 1907 that the causal relationship between infection of wounded plants by agrobacteria and tumor etiology was established[9]. It is now known that the enzymes catalyzing the synthesis of opines are found in transformed plant cells. However, the structural genes for these enzymes are encoded on large plasmids resident in virulent strains of agrobacterium[10]. These tumor inducing plasmids are required for crown gall induction, and tumorigenesis involves the excision of a segment of the plasmid DNA on which the opine dehydrogenase gene is located. Following integration of this DNA into the plant genome[11], the plant cell machinery is hijacked to divert resources to the synthesis of opines which are transported to the gall to allow growth of the tumor. Further studies have shown that opines play a major role at several stages of the intricate prokaryote-eukaryote inter-relationship between virulent strains of *Agrobacterium tumefaciens* and plants including increasing the transcription of the Ti plasmid-encoded virulence (*vir*) genes[12].

A major area of interest in this enzyme superfamily lies in the field of enzyme-based chiral synthesis. Opine-type secondary amine dicarboxylic acids are useful chiral intermediates for a number of important pharmaceuticals including, for example, the antihypertensive agents enalaprilat and lysinopril. Given that such compounds contain multiple chiral centers, their synthesis by enzyme-based technology, rather than classical chemistry, has enormous potential. Indeed, the 'one pot' synthesis of the opine N-[1-(R)-(carboxyl)ethyl]-(S)-phenylalanine from phenylpyruvate and pyruvate using CENDH, formate dehydrogenase and phenylalanine dehydrogenase has been achieved[13]. In the long term this may require the rational development of opine dehydrogenases with novel specificities having important applications in areas of drug production and discovery. This would in turn require the determination of the molecular structure of representative members of the opine dehydrogenase family.

The N-(1-D-carboxylethyl)-L-norvaline dehydrogenase (CENDH) from the soil isolate *Arthrobacter* sp. strain 1C belongs to the (D,L) opine dehydrogenase family and exists as a homodimer with a subunit molecular weight of ~36,000 M_r[4]. In the oxidative deamination reaction, the enzyme is active towards opines such as *N*-[1-D-(carboxyl)ethyl]-L-methionine (methiopine) and *N*-[1-D-(carboxyl)ethyl]-L-phenylalanine. In the reductive secondary amine forming reactions with NADH as cofactor, the enzyme utilizes hydrophobic L-amino acids such as L-methionine, L-isoleucine, L-valine, L-phenylalanine and

[1]Krebs Institute for Biomolecular Research, Department of Molecular Biology and Biotechnology, The University of Sheffield, Sheffield S10 2TN, UK. [2]Biotechnology Research Center, Faculty of Engineering, Toyama Prefectural University, 5180 Kurokawa, Kosugi, Toyama 939-03, Japan.

Correspondence should be addressed to K.L.B. *email: kbritton@sheffield.ac.uk*

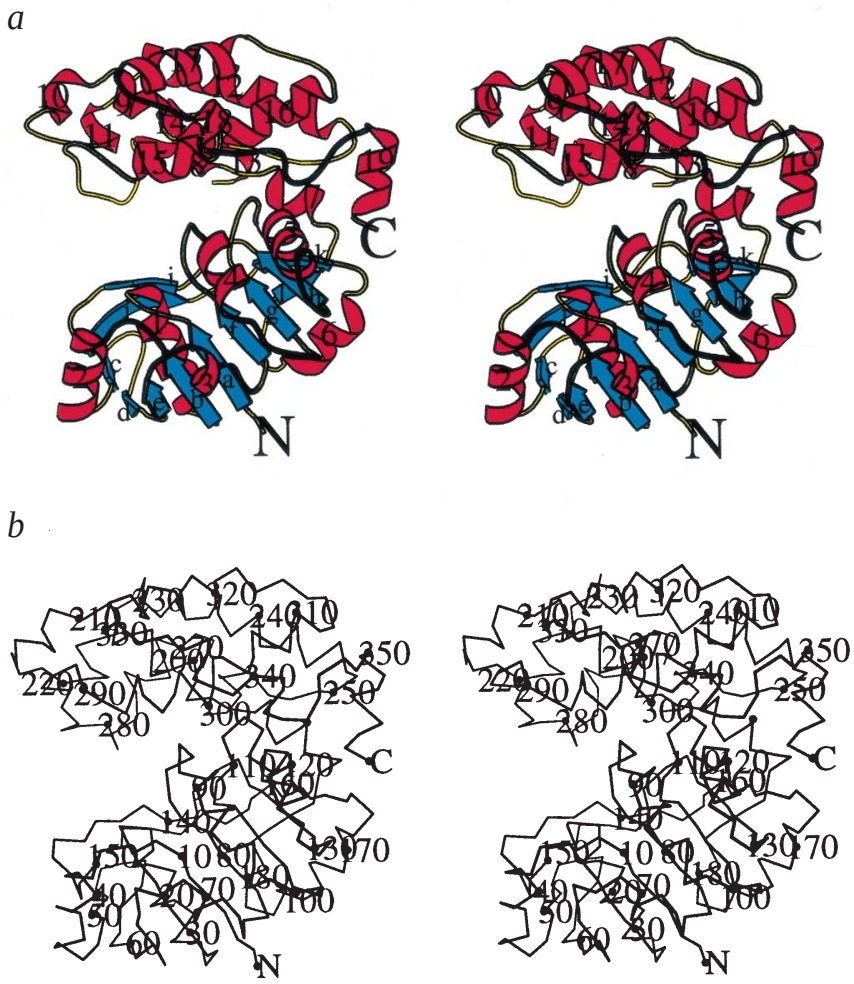

Fig. 1 Stereo diagrams, prepared using MOLSCRIPT[49], of a single subunit of CENDH. The organization of the subunit into two domains, separated by a deep cleft, can be seen. In this view the two-fold axis of the CENDH dimer runs horizontally. ***a***, Schematic representation with the strands and helices labeled. ***b***, Cα trace with every tenth residue indicated by a black dot.

L-leucine, as amino donors and α-keto acids, such as pyruvate and oxaloacetate, glyoxylate and α-ketobutyrate, as amino acceptors[14].

In order to understand the molecular basis of the differential substrate specificity in this superfamily we have initiated a structural study of CENDH. Here, we report the initial results from the structure determination by X-ray crystallography.

Structure determination

The structure of the binary complex of CENDH with NAD^+ was determined initially to a resolution of 2.6 Å by multiple isomorphous replacement (MIR) and density modification. Subsequently the structure was refined using the program TNT, gradually increasing the resolution to include a data set to 1.8 Å. Further cycles of model building and refinement were performed to produce a final model comprising 349 out of a total of 359 residues and including a total of 113 solvent molecules. No clear density could be found for the remaining ten residues of the model, which are accounted for in two places: the first three residues at the N-terminus and a loop region between α11 and α12 of domain II, comprising residues Lys 260–Ser 266. The current model includes two *cis*-prolines (residues 109 and 165 in CENDH) both of which occur in β-turn regions of secondary structure. However, neither of these prolines appears to be conserved across this enzyme superfamily. The NAD^+ moiety, though present in the crystal, has not been included in the model that was refined using the high resolution data because of its apparent low occupancy.

Quality of the model

The electron density is of a high quality throughout the model but ends abruptly at the boundary of the missing residues. SDS gel electrophoresis of the crystals confirms that the protein has not been cleaved during purification, hence these regions appear to be genuinely disordered in this crystal form. The high quality of the model is confirmed by the average B-values of 29.1 $Å^2$ for the protein atoms (for main chain atoms: 24.5 $Å^2$ overall, 24.0 $Å^2$ and 25.2 $Å^2$ for domains I and II respectively, with 34.5 $Å^2$ overall for side chain atoms). For the water molecules the average B-factor is 41.8 $Å^2$. A Ramachandran plot of the model showed that all residues lie either in the most favorable or additional allowed regions and examination of the χ1-χ2 plots for all residue types showed no side chains in unfavorable conformations[15].

Overall structure

The secondary structure assignments of CENDH, analyzed using PROCHECK show the subunit to be constructed from nineteen α-helices, twelve β-strands and four 3_{10}-helices (62% of the polypeptide chain) which fold to form two domains separated by a deep cleft (Fig. 1). Domain I comprises residues 1–195, which form a central eight-stranded β-sheet of mixed parallel and anti-parallel character (β1, β2, β5, β6, β7, β8, β11 and β12), flanked by a number of α-helices and also by a smaller four stranded β-sheet that contains both parallel and anti-parallel strands. Part of the eight stranded β-sheet is organized as a six-stranded parallel sheet which resembles the classical Rossmann fold commonly found in dinucleotide-binding enzymes[16]. This domain contains a classic βαβ unit that includes a dinucleotide binding helix (α1) that separates the first two β-strands in the structure. The amino acid sequence in the loop joining the first β-strand to α1 includes a glycine rich region with a sequence GLGNGG fitting the consensus GXGXXG/A which often delineates the nucleotide binding site in dehydrogenases[17]. Domain II is entered by a short helix (α8) and is composed of the remaining residues to the C-terminus (196–359). This domain is constructed from twelve α-helices connected by loops of varying length. The C-terminal helix (α19) finishes close to the first domain, forming largely water-mediated interactions between the last turn of this helix to residues carried on α5, βk and α7 of domain I. The remaining contacts across the domain interface are not extensive and arise

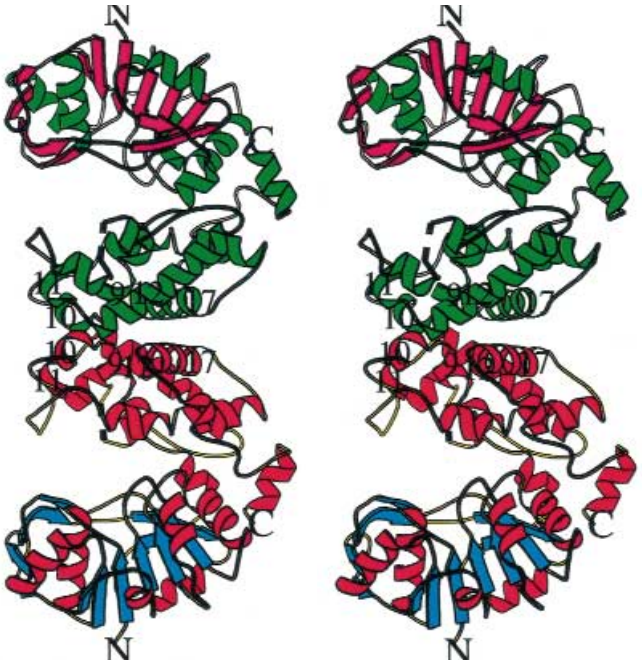

Fig. 2 Stereo representation of a CENDH dimer viewed with the two-fold axis horizontal and prepared using MOLSCRIPT. The N- and C-termini for each of the two subunits are labeled as are the secondary structure elements involved at the subunit interface. The dashed line represents the region of the seven disordered residues in the CENDH model for which the path of the polypeptide chain has not been determined.

primarily from loop regions rather than from specific secondary structural elements. These include interactions involving the start of α4, the β7 to α5 loop, the middle of α5, and the loops between α7 and α8; α14 and α15; and α17 and α18.

Subunit interface

Gel filtration experiments suggest that CENDH is a dimer[4] and this is supported by the packing in the crystal, in which an extended interface is formed between two subunits related by a crystallographic two-fold about c. This creates a dimer with overall dimensions 100 Å × 50 Å × 35 Å (Fig. 2) where each monomer makes contacts to its two-fold related partner through residues contained in domain II alone. Accessible surface area (asa) calculations[18] show that a single subunit of CENDH possesses an asa of 14,300 Å^2. On forming the dimer, 7.7% (1,100 Å^2) of the subunit asa is buried with this interface comprising 53% non-polar atoms and 47% polar atoms. This is both more polar and less extensive than that observed for other dimeric proteins, in which the average percentage of polar residues is 32% and the proportion of the asa buried on dimer formation commonly ranges from 6.5–29.4%[19]. The overall surface of the dimer interface is fairly flat with no deep protrusions from either subunit. In total, 31 residues are involved at the two-fold interface from each subunit, of which only one residue, Val 322, is conserved across the sequences of this opine dehydrogenase superfamily. Specifically, the interactions made across the dimer interface involve contacts between the helices α9, α10, α11, α12 and the loops between them; α17 and part of the loop region between α14 and α15. The dimer interface can be separated into two roughly equal halves, one comprising a major hydrophobic region with the other being hydrophilic in nature (Fig. 3). The hydrophobic part of the interface is made up of residues from two helices from each subunit (the N-terminal half of α12 and α17) which pack against each other, whereas the hydrophilic region is comprised of interactions between the C- and/or N-termini of the remainder of the participating helices together with loop regions. This latter region can also be further divided into three clusters, each of which involve water mediated interactions at their centers, one of which is completely buried (Fig. 3).

Nucleotide-binding site

The NAD$^+$ binding site was initially located through the identification of the electron density for the cofactor in the preliminary 2.6 Å map (Fig. 4*a*). Difference Fourier analysis against a data set collected on a crystal grown in the absence of the cofactor (NA1; Table 1) showed a single strong electron density feature associated with the CENDH with no evidence for a major conformational change on NAD$^+$ binding. Following preliminary refinement of the 2.6 Å data, no significant improvement was observed with good density apparent for the adenine ring, its associated ribose and the pyrophosphate moiety. However, the nicotinamide ribose moiety was only of moderate quality and that for the nicotinamide ring was very weak. Despite the fact that the crystals used for the high resolution CENDH data set were grown in the presence of NAD$^+$, the density was weak for the adenine ring and pyrophosphate moieties of the NAD$^+$ and very poor for the nicotinamide moiety. This has been attributed to low NAD$^+$ occupancy in this crystal, hence the cofactor has been omitted from the high resolution refinement.

The X-ray analysis of the refined binary complex of CENDH with NAD$^+$ at 2.6 Å has allowed us to determine some of the interactions between the nucleotide and the enzyme involving residues from domain I which contains the nucleotide binding fold. The protein provides a pocket for the adenine ring lined by the side chains of Ile 35 and His 87, and the main chain nitrogen of Leu 12 forms a hydrogen bond to the ring oxygen of the adenine ribose. There is no evidence of direct hydrogen bonding between the adenine ring and the protein. However, we cannot preclude the involvement of water-mediated interactions since the assignment of solvent is somewhat ambiguous at this resolution. As has been observed in many dehydrogenases[20], the adenine ribose hydroxyl groups make specific hydrogen bonds to the side chain carboxyl group of an acidic amino acid at the C-terminus of βb (Asp 34). In turn the side chain of this residue forms a hydrogen bond to the main chain amide nitrogen of a residue in the glycine rich turn that lies between the first β-strand of the domain (βa) and the dinucleotide binding helix (α1). Thus the glycine rich turn is involved indirectly in the recognition of the adenine ribose[21]. The pyrophosphate moiety is located above the end of the dinucleotide binding helix forming hydrogen bonds between the main chain nitrogen atoms of Asn 14 and Gly 15 in the glycine rich turn and the P1 and P2 oxygens respectively, of the phosphate group adjacent to the adenine ribose ring. Thus, the interactions in this region of the structure are similar to that seen in other dehydrogenases[17]. On the basis of the structure, a conserved residue Arg 292 must lie close to the pyrophosphate moiety although, in the present electron density map, the density for the side chain of this residue is weak implying that it does not play a crucial role in cofactor recognition.

The interactions between the enzyme and the nicotinamide ribose are uncertain due to the ambiguities in the locations of the ribose hydroxyl groups. However, we note that the side chain of

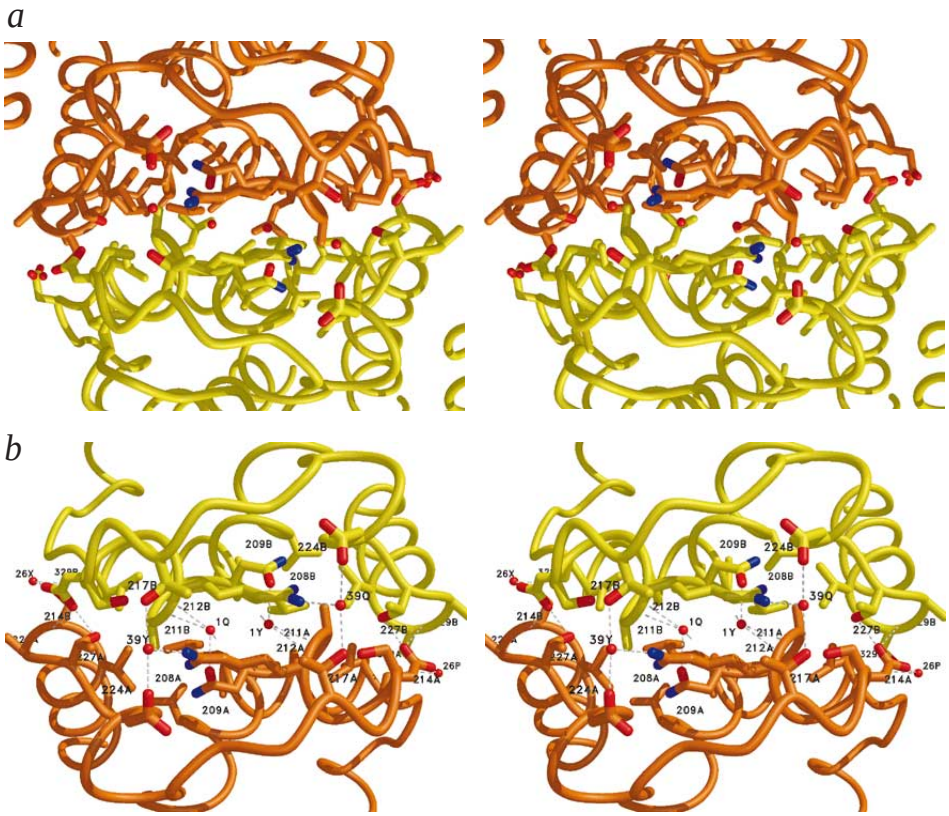

Fig. 3 The intersubunit interface of the CENDH dimer, viewed across the two-fold axis drawn using the program MIDASPLUS[50]. ***a***, The backbone traces together with the side chains of the 31 residues in each subunit (one shown in yellow, the other in orange) whose accessible surface areas change upon dimer formation are shown. Within these residues, all the side chain oxygen (red) and nitrogen (blue) atoms are further highlighted. Water molecules buried at this interface are shown as red spheres. ***b***, A close up of (*a*) highlighting the central water mediated interactions between the two subunits. The hydrogen bonds made between the protein and water molecules are indicated by dashed white lines.

the conserved glutamate, Glu 296, lies very close to density associated with the nicotinamide ribose hydroxyl groups and we presume it is involved in hydrogen bonding to these groups. The weak nature of the density for the nicotinamide ring prevents the conformation of the glycosidic bond from being assigned and therefore, at this stage, this precludes us from commenting on the stereochemistry of the hydride transfer reaction.

Similarity of CENDH to other proteins

The program PROTEP[22] was used to compare the three-dimensional coordinates of CENDH with those of all other proteins in the July, 1997 release of the PDB[23]. Most of the similarities detected by the program were between the nucleotide binding domain of CENDH and that of numerous dehydrogenases, including lactate dehydrogenase (LDH) and malate dehydrogenase (MDH). The most extensive similarity revealed by the PROTEP search was between CENDH and 6-phosphogluconate dehydrogenase (6-PGDH) from sheep liver[24] (1PGN). 6-PGDH is a homodimeric enzyme, composed of two subunits, each of 482 amino acids, which catalyzes a two-step oxidative decarboxylation whereby 6-phosphogluconate is converted to D-ribulose 5-phosphate with the concomitant reduction of $NADP^+$ to NADPH and the liberation of carbon dioxide[25]. Whilst this chemistry is completely different from that catalyzed by CENDH, the similarity between the two enzymes detected by PROTEP extends not only over the nucleotide binding domain (domain I) but also involves three helices from domain II. As shown in Fig. 4*b*, the region of common structure between 6-PGDH and CENDH comprises eight β-strands and 10 α-helices (for the purposes of the comparison search, helices α8 and α9 in CENDH were combined by the program and were equivalent to helix αh in 6-PGDH) with identical topology and sequence order. However, the active site of 6-PGDH involves contributions by elements from one subunit together with residues from its two-fold related partner. Therefore, since the mode of association of the dimer in CENDH is completely different, those symmetry related elements of 6-PGDH that contribute to the active site in this enzyme have no equivalents in CENDH. Furthermore, none of the catalytic residues of 6-PGDH arise from any of the common elements between the two enzymes. We therefore attach no particular significance to the similarity between these two proteins.

Location of the active site

In a previous analysis of the amino acid sequence and chemistry of enzymes belonging to the opine dehydrogenases family, analogies have been drawn with the chemistry carried out by the family of 2-hydroxy acid dehydrogenases[26]. Indeed, this similarity is further supported by the report that opine dehydrogenases in marine molluscs have a physiological role analogous to LDH of vertebrates in the maintenance of the cytoplasmic redox balance during anaerobic glycolysis[27]. The 2-hydroxy acid dehydrogenases, LDH and MDH, contain an Asp-Xaa-Xaa-Arg motif which lies at the active site. In LDH and MDH the role of the conserved arginine is to stabilize the α-keto acid by providing a two point interaction with the carboxylate group[28]. In these enzymes the conserved aspartate residue is thought to be involved in promoting catalysis, forming an Asp-His pair that has been shown to be part of a proton relay system[29]. Sequence analysis of both lysopine dehydrogenase (DHLO) and nopaline dehydrogenase (DHNO) highlighted the apparent conservation of an equivalent motif (residues 284–287 and 235–238 respectively) in conserved segments of their C-terminal regions. This has led to the suggestion that a similar triad-like arrangement plays a role in the

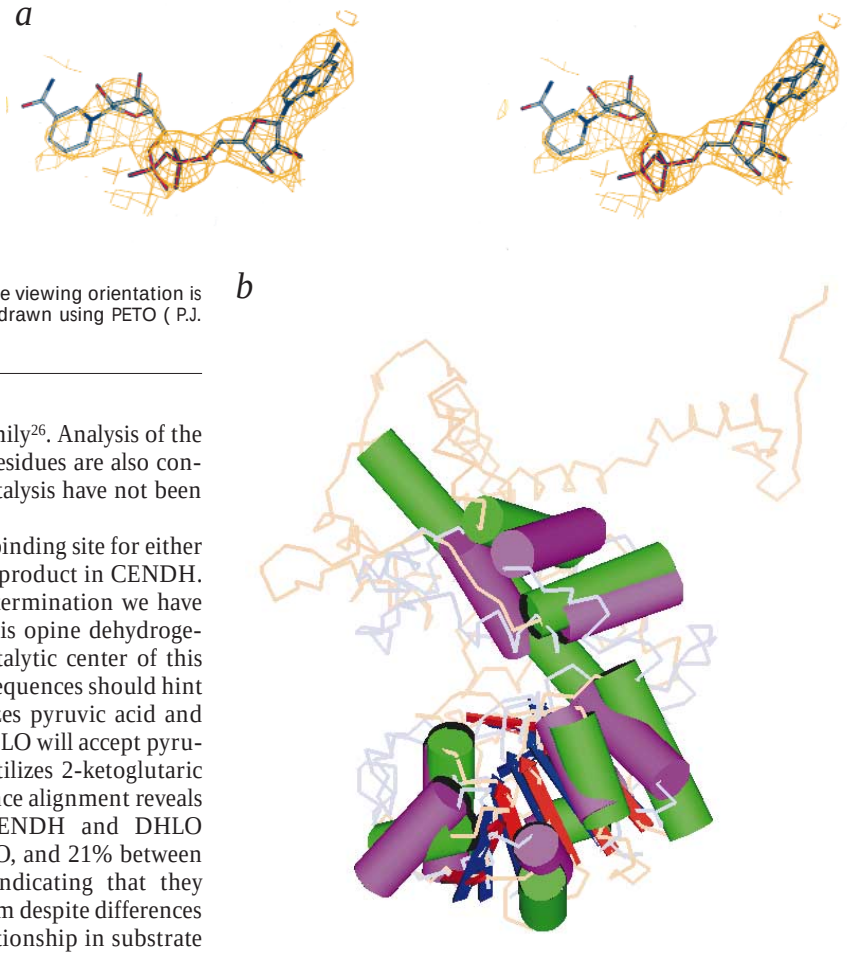

Fig. 4 ***a***, A BOBSCRIPT[51] representation of the positive difference electron density calculated with coefficients $(|F_{ND1}| - |F_{NA1}|)\alpha_{iso}$, contoured at 1σ for the binary complex of CENDH with NAD^+. Although not included in the high resolution model the initial fit for the NAD^+ moiety is shown in atom colors as in Fig. 3. ***b***, The structural overlap between CENDH and 6-PGDH detected using PROTEP[22]. The proteins are shown superimposed with only those helices and strands equivalenced by PROTEP displayed as solid cylinders and arrows respectively. Helices, strands and Cα backbone traces are pink, blue and mauve for CENDH, green, red and orange for 6-PGDH respectively. The viewing orientation is approximately the same as that in Fig. 1. Figure drawn using PETO (P.J. Artymiuk, unpublished program).

mechanism of the opine dehydrogenase family[26]. Analysis of the sequence of CENDH confirms that these residues are also conserved in this enzyme, but their roles in catalysis have not been confirmed.

As yet we have been unable to locate the binding site for either the keto acid, the amino acid or the opine product in CENDH. However, in the light of the structure determination we have analyzed the sequences for members of this opine dehydrogenase superfamily to try to identify the catalytic center of this enzyme with the rationale that conserved sequences should hint at the active site. CENDH primarily utilizes pyruvic acid and hydrophobic amino acids as substrates; DHLO will accept pyruvic acid and basic amino acids; DHNO utilizes 2-ketoglutaric acid and arginine. A structure based sequence alignment reveals sequence identities of 25% between CENDH and DHLO (Fig. 5), 22% between CENDH and DHNO, and 21% between lysopine and nopaline dehydrogenases indicating that they probably share the same catalytic mechanism despite differences in substrate specificity with the closer relationship in substrate specificity between CENDH and DHLO (both enzymes utilize pyruvate) being mirrored at the amino acid sequence level by a higher level of similarity and by a more similar pattern of insertions and deletions.

The sequence alignment highlights a total of 35 absolutely conserved residues across all three sequences. When the locations of these residues are plotted on the three-dimensional structure of CENDH, a cluster of six conserved residues is highlighted (Fig. 6*a*) — Asn 198, His 202, Arg 292, Tyr 293, Glu 296 and Asp 297, which lie in the cleft between the enzyme's two domains adjacent to the nucleotide binding site. In studies of other dehydrogenases crystallized from ammonium sulphate, it has frequently been observed that sulphate ions bind at or close to the active site[30]. Analysis of the refined density map of CENDH shows an unidentified electron density feature, stronger than that normally assigned for a water molecule, adjacent to the side chains of Tyr 293 and His 202, of the above conserved cluster, and Tyr 222 (Fig. 6*b*). An obvious candidate for this feature is a sulphate ion. However, refinement of the model produced very high B-factors for this sulphate and proved inconclusive. Nevertheless, this observation may indicate the general location of the enzyme's catalytic center.

The cluster of six conserved residues are all contained within domain I, whereas the residues involved in the binding of the nucleotide moiety are located in domain II. This proposed difference in roles of the two domains is similar to that seen in L-lactate dehydrogenase and is in contrast to the case of the D-specific 2-hydroxy acid dehydrogenases, such as D-glycerate dehydrogenase[31], which catalyzes a related reaction, but where the majority of residues concerned with binding both the nucleotide and keto acid are associated with the NAD^+-binding domain only. The conserved cluster of residues in CENDH also includes a His-Asp pair (His 202 and Asp 297) that is located adjacent to a conserved arginine (Arg 292; Fig. 6*c*). If this is indeed the enzyme's catalytic center, then there is clearly a strong parallel to the other keto acid dehydrogenases whose chemistry is dependent on a His-Asp-Arg triad. These similarities suggest that the carboxyl of the keto acid could interact with the Arg 292 and His 202 and may play a role as an acid/base catalyst by donating a proton to the substrate carbonyl group[29]. However, we note that the relative orientation of these three residues in CENDH is somewhat different from their arrangement in LDH and MDH and a detailed comparison of all three residues is currently precluded by the disorder of Arg 292 in CENDH. Moreover, neither Asp 297 nor Arg 292 in CENDH belongs to the Asp-Xaa-Xaa-Arg motif that had been identified by Monneuse and Rouzé[26] as a possible candidate for the catalytic center of this enzyme family. Rather, the latter motif involves Asp 238 and Arg 241 and is located on the back face of the molecule, behind the proposed active site, where these two residues are involved in a partially buried salt bridge. Therefore, in CENDH the residues of the Asp-Xaa-Xaa-Arg motif are unlikely to be directly involved with the enzyme chemistry, although their conservation behind is suggestive of a possible structural role. Furthermore, in CENDH the conserved aspartate and arginine residues that lie at the proposed active site do not form a motif of this type. Rather their order along the polypep-

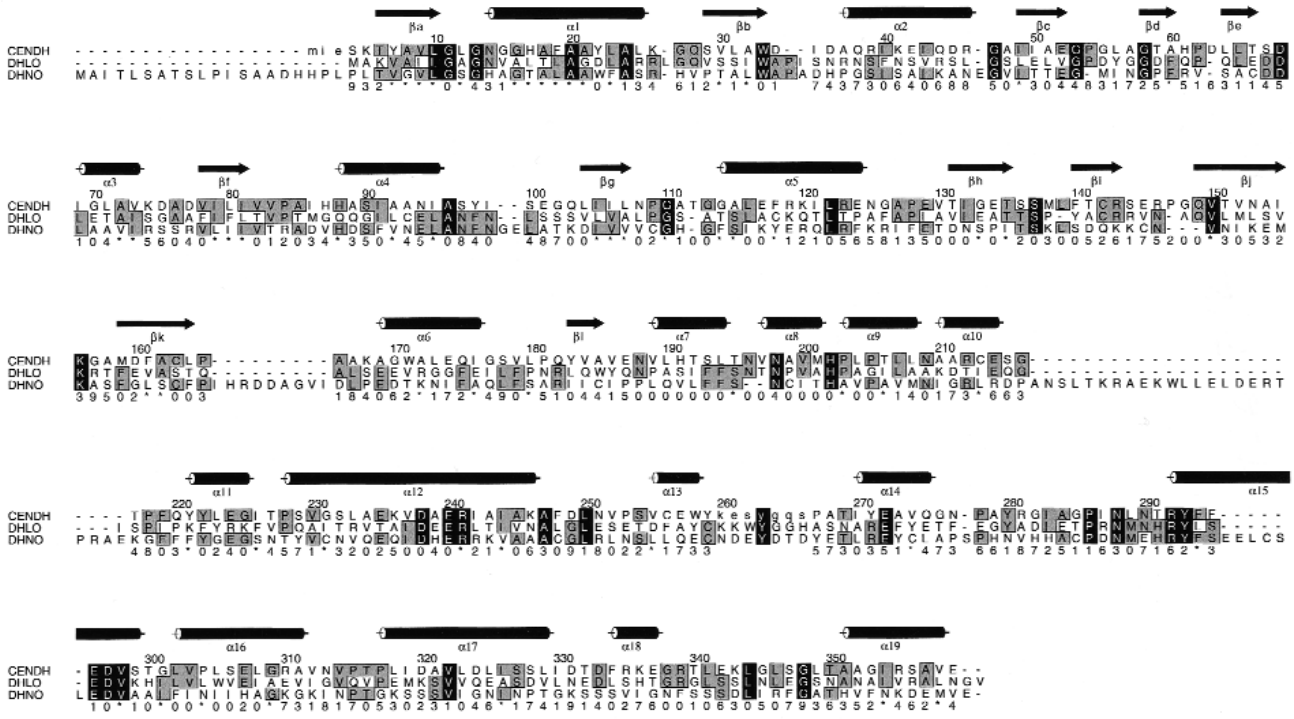

Fig. 5 Structure-based alignment of the sequence for the CENDH from *Arthrobacter* sp. strain 1C together with the other two sequenced (D,L) opine dehydrogenases: nopaline dehydrogenase (DHNO) and lysopine dehydrogenase (DHLO) both from *Agrobacterium tumefaciens*. Residues conserved among all three sequences are highlighted by a black background, whereas residues conserved in two out of the three sequences are shaded grey. Lower case has been used to denote the residues of the CENDH that are disordered and therefore absent from the model. The secondary structural elements of the three-dimensional structure of CENDH are shown above the sequence alignment with helices shown as cylinders and strands represented by arrows. The sequence numbering of every tenth residue in the CENDH sequence is also shown. Positions of deletions are indicated by dashes.

tide chain is reversed and the spacing between them is different. At this stage it is premature to try to define the mechanism of CENDH on the basis of this analysis alone. Nevertheless, the striking parallels with the other 2-hydroxy acid dehydrogenases and the location of the additional electron density feature clearly make the residues in this region strong candidates for site-directed mutagenesis.

In addition, we note that the disordered region of seven residues in the CENDH model may also be associated with the substrate binding site for this enzyme. The two extremes of this missing loop are located on the underside of domain II which is close to the cleft region of the subunit structure. This region of disordered residues also contains one conserved residue, Tyr 263, whose function is as yet uncertain.

While DHLO primarily utilizes basic amino acids, CENDH shows a broad specificity for a range of hydrophobic amino acids. Therefore, in order to identify the amino acid binding site, we have analyzed the positions in the aligned sequences where a change occurs from a hydrophobic amino acid in CENDH to a negatively charged or polar amino acid in DHLO. This analysis reveals that there are nine such substitutions within 12 Å of the putative catalytic center (defined as the center of the unidentified electron density feature). The most striking change is the replacement in DHLO of Val 197 and Val 255 in CENDH by Thr and Asp respectively. If this analysis is an indication of the location of the binding site for the side chain of the amino acid substrate, then the observed sequence differences are consistent with the change in specificity. This would then indicate that the cluster of hydrophobic residues centered on Val 197 and including Val 252, Pro 253, and Val 255 define the pocket for the amino acid substrate in CENDH. Analysis of the sequence of DHNO is complicated by the deletion of two residues between Thr 195 and Asn 198 (Fig. 6*d*) which are adjacent to the active site and therefore will necessarily result in significant structural changes in this part of the molecule. However, the nearby replacement of Trp 258 by glutamate provides a possible candidate for recognition of the arginine substrate for this enzyme.

The kinetic and chemical mechanisms of CENDH remain to be elucidated. However, the chemistry of the opine dehydrogenase family is analogous to that of the amino acid dehydrogenases with the differences between them corresponding to the reductive amination of a keto acid by an amino acid and ammonia respectively. Structural and biochemical data support the view that the chemical mechanisms of the family of amino acid dehydrogenases, which include glutamate dehydrogenase and leucine dehydrogenase, and the unrelated amino acid dehydrogenase alanine dehydrogenase, all proceed through carbinolamine and imino acid intermediates[32]. Given the analogies to the chemistry in CENDH this may suggest that the reaction proceeds through a carbinolamine intermediate following nucleophilic attack on the keto acid by the amino group of the amino acid. This is further supported by the recent observation from the structure of AlaDH[33] that the active site of this enzyme includes the conservation of a His, Glu and Arg cluster with clear relationships to the

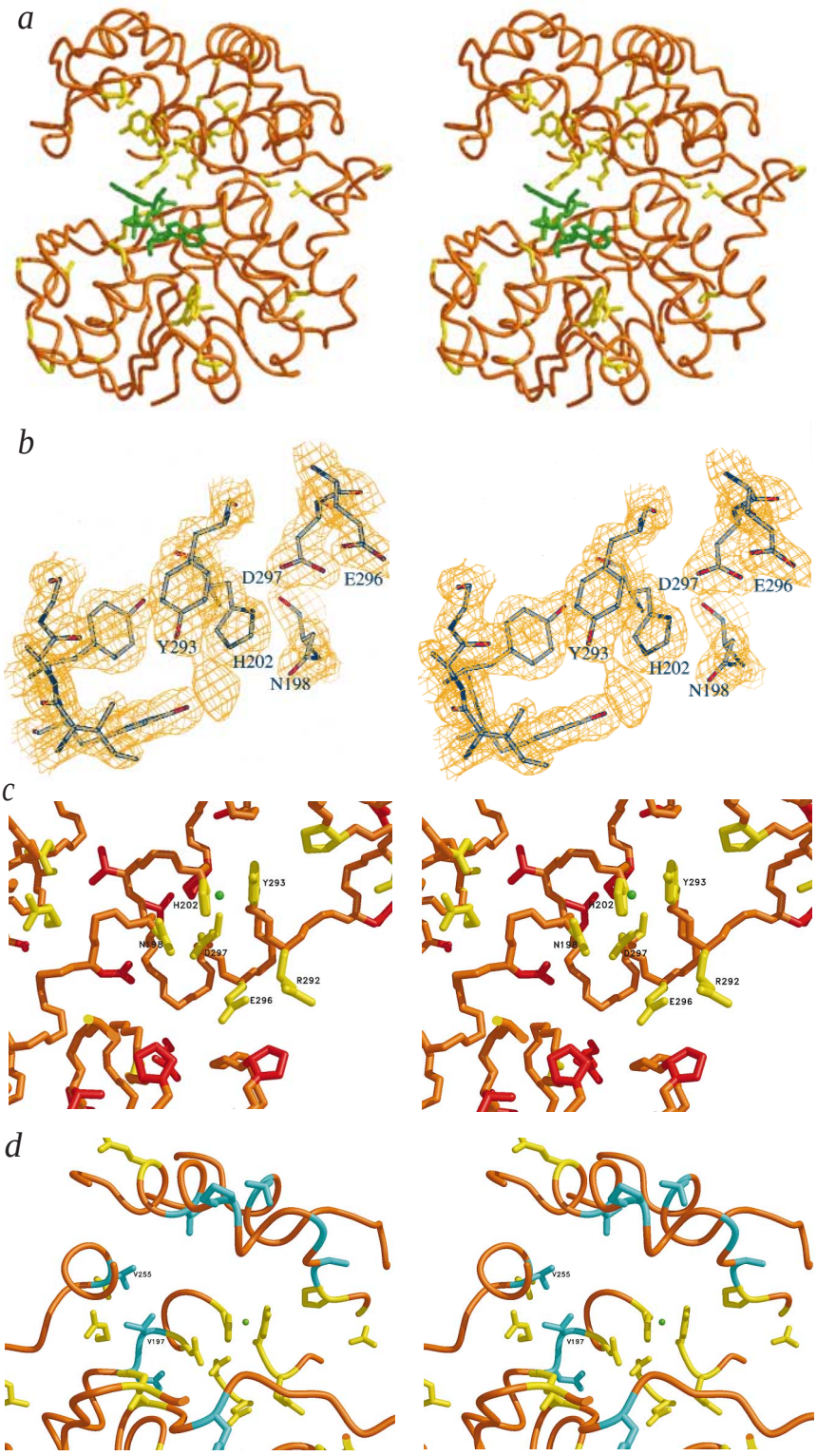

Fig. 6 ***a***, Stereoview of the backbone trace of a CENDH subunit (orange) together with the location of the 35 residues (yellow) conserved between the sequences of CENDH, DHLO and DHNO. Drawn using the program MIDASPLUS[50], the clustering of these conserved residues in the cleft region of the subunit is evident, particularly in the vicinity of the bound nucleotide cofactor, NAD^+ (green). ***b***, The portion of the $(2|F_{obs}| - |F_{calc}|)\alpha_{calc}$ electron density map calculated at 1.8 Å resolution drawn using the program BOBSCRIPT[51] with the CENDH model around the cluster of six conserved residues (labeled). The as yet unidentified electron density found adjacent to the conserved side chains of Tyr 293 and His 202 can be seen and may indicate the active site region of the molecule. ***c***, A MIDASPLUS[50] stereoview of the backbone trace of CENDH (orange) around the cluster of six residues conserved across the (D,L) family of opine dehydrogenases (yellow). Also marked are the positions of conserved residues between CENDH and DHLO alone (red). The location of the unexplained electron density is marked as a green dot. ***d***, A MIDAPLUS[50] stereoview of the backbone trace of CENDH (orange) around the proposed region for the amino acid binding site. As in (*a*) the picture again highlights the residues conserved across the family of opine dehydrogenases (yellow) together with the residues which switch from those of hydrophobic character in CENDH to polar or negatively charged in DHLO (cyan). The location of the unexplained electron density is marked as a green dot.

structure of CENDH. A further aspect of the mechanism of CENDH that merits consideration is the possible need for a conformational change as part of the cycle of catalytic activity. Clearly the enzyme has to prevent the abortive reduction of pyruvate to lactate and thus a conformational change may well allow the NADH to be remote from the pyruvate until such time as the nucleophilic attack, which leads to the production of a carbinolamine, has been completed. This could explain the apparent separation of the NAD^+ in domain I from the cluster of conserved residues in domain II. Certainly, if the mode of NAD^+ binding on the enzyme surface is similar to that found in the amino acid dehydrogenases then the current distance between the nicotinamide ring and the conserved residues in the model is approximately 8 Å, too far for hydride transfer to occur. This would imply that a conformational change is required for catalysis to occur.

The structure of CENDH represents the first to be reported for a member of either the (D,L) or (L,L) opine dehydrogenase family. With the extension of this analysis to include information on the determinants for substrate recognition it is hoped that a programme of site directed mutagenesis can be initiated, aimed at the rational modification of the specificity of this enzyme. This will not only underpin our basic understanding of the manipulation of enzyme specificity but also allow us to explore the biotechnological potential of this enzyme family.

Methods

Crystallization. Isomorphous crystals of CENDH in the presence and absence of NAD^+ were grown by the hanging drop vapor diffusion method from buffered ammonium sulphate solutions as described[34]. The enzyme crystallizes in the space group $P2_12_12$ with cell dimensions a = 104.9 Å, b = 80.0 Å and c = 45.5 Å. It has one monomer in the asymmetric unit and a V_m of 2.65 $Å^3\,Da^{-1}$.

Table 1 X-ray data collection and phasing statistics

Data set[1]	ND1	ND2	Hg1	Au3	Pt4	NA1
Resolution (Å)	2.6	1.8	2.6	2.4	2.4	2.4
No. observed reflections	38,132	66,518	84,226	79,114	77,706	39,802
No. unique reflections	9,796	30,854	19,868	23,690	20,584	12,473
Completeness (%)	82.4	89.0	91.0	89.0	77.0	86.0
R_{merge}[2]	0.03 (0.07)	0.04 (0.15)	0.06	0.04	0.04	0.04
Mean fractional isomorphous difference[3]	-	-	0.106	0.187	0.127	-
Heavy atom concentration (mM)	-	-	1	2	10	-
Soak time (h)	-	-	1	24	2	-
No. heavy atom sites	-	-	1	3	2	-
Phasing power (acentric/centric)[4]	-	-	1.25/0.93	1.16/0.91	1.25/1.02	-
R_{cullis} (acentric/centric)[5]	-	-	0.78/0.77	0.80/0.80	0.78/0.73	-
Detector	SDMS	MAR	SDMS	SDMS	SDMS	SDMS

[1]See the Methods section for descriptions.
[2]$R_{merge} = \Sigma_{hkl}|I_i - I_m|/\Sigma_{hkl}I_m$, where I_i and I_m are the observed intensity and the mean intensity of related reflections, respectively. Value in parentheses for the 5% of data at the limits of the resolution.
[3]Mean fractional isomorphous difference = $\Sigma||F_{PH}| - |F_P||/\Sigma|F_P|$, where F_{PH} and F_P are the structure factor amplitudes for derivative and native crystals respectively.
[4]Phasing power = $<F_H/$lack of closure$>$.
[5]R_{cullis} = <lack of closure>/<isomorphous difference>.

Data collection and processing. Initially, medium resolution native and derivative X-ray diffraction data on the crystals grown in the presence of NAD^+ were collected at room temperature on a twin San Diego multiwire systems (SDMS) area detector with graphite monochromated X-rays generated by a Rigaku RU-200 rotating copper anode using the Xuong-Hamlin method of data collection[35,36]. The data were processed and merged using SDMS software[37]. Data collection statistics are shown in Table 1. Subsequently a high resolution data set (ND2; Table 1) was collected at room temperature on a MAR image plate detector on station 9.5 at the CCLRC SRS Daresbury Laboratory, Warrington, UK. This data set was obtained from one crystal exposed as two halves and processed using the MOSFLM suite of programs[38]. The resulting data were merged and scaled using the programs ROTAVATA and AGROVATA[39].

Three isomorphous derivatives were prepared by soaking crystals grown in the presence of NAD^+ in solutions of 1 mM ethyl mercury phosphate (Hg1; Table 1); 2 mM $KAu(CN)_2$ (Au3; Table 1) and 10 mM $K_2Pt(CN)_4$ (Pt4; Table 1) respectively, and data collected on each using the SDMS area detector as above. In addition, data sets were collected on the SDMS area detector system from a crystal grown in the absence of NAD^+ (NA1; Table 1).

MIR phasing. The Patterson function of the mercury soaked crystal was readily interpretable and consistent with a single site derivative. The heavy atom parameters were refined and the phases calculated using the program MLPHARE[40]. These preliminary phases were used to produce difference Fouriers for the other two derivatives. The heavy atom parameters of all three derivatives and the phases calculated from them gave an overall figure of merit to 2.6 Å of 0.60 (acentric 0.59, centric 0.70). In each derivative, one site was in common and a second gold site was shared with one of the platinum sites, though different patterns of relative occupancy were displayed.

The MIR map was improved by solvent flattening (20 cycles) using the program DM[41], with the solvent content of the unit cell set to 45%. An initial map calculated from these phases showed clear molecular boundaries and several regions of readily identifiable secondary structure. The solvent flattened map was then skeletonized using the BONES procedure in the program O[42]. The Cα atoms were then positioned into the DM map for 318 residues of the CENDH subunit with the model from BONES used as a guide, using the program FRODO[43] on an Evans and Sutherland ESV workstation. The Cα atoms were assigned a preliminary amino acid sequence and the automatic addition of main chain and side chain atoms in the preferred rotamers was undertaken using the LEGO procedure in O[42]. Phases were calculated from this model using the program SFALL[39] and they were combined with the MIR phases using the program SIGMAA[39]. The combined phases were used to calculate a map which resulted in the chain trace being extended. This resulted in a model where the positions of 94% of the residues had been determined.

Refinement. The medium resolution data set used to solve the structure thus far was superceded at this point with a data set composed of the high resolution 1.8 Å set, collected at the CCLRC Synchrotron Radiation Source, Daresbury Laboratory, supplemented by reflections present in the initial medium resolution data set but absent in the high resolution data. A refinement protocol was set up to allow the use of R_{free} calculations by performing the refinement based on the random selection of 92% of the data. The initial model was submitted to restrained least-squares refinement using the program TNT[44,45] with a fixed value for an overall B-factor and a correction for the solvent continuum applied[46]. Initially, refinement was carried out at 2.5 Å followed by a stepped increase in resolution over a number of refinement cycles to gradually include the complete 1.8 Å data set. Only when all data to 1.8 Å were incorporated in the refinement process were individual B-factors refined. The model was then rebuilt to the observed density using $(2|F_{obs}| - |F_{calc}|)\alpha_{calc}$ maps. During iterative cycles of model building and refinement additional residues were identified and corrections made to the structure where necessary. Solvent molecules were added to features above 3σ in a $(|F_{obs}| - |F_{calc}|)\alpha_{calc}$ map, only where they made appropriate interactions to the protein. These waters were only retained in the model if their individual B-factors remained below 70 $Å^2$ following refinement. The refinement and rebuilding process led to a final model of 349 residues (2,589 atoms) together with 113 waters, an average B-factor of 29.8 $Å^2$ and a crystallographic R-factor of 0.188 (R_{free} 26.3%) based on all data between 20 and 1.8 Å (28,922 reflections). The r.m.s. bond deviation was 0.017 Å, r.m.s. angle deviation 1.7°, r.m.s. trigonal atom non-planarity 0.016 Å and r.m.s. planar groups deviation 0.021 Å. The statistics of the structure determination are presented in Table 1.

In parallel, refinement using TNT was conducted on the 2.6 Å data set to allow a more detailed determination of the cofactor binding site. This resulted in a model with crystallographic R-factor of 0.134 (R_{free} 26.4%) based on all data between 20 and 2.6 Å (8,942 reflections).

Structure comparison. A comparison of the structure of CENDH with those of all entries in the July, 1997 release of the Protein Data Bank was carried out using the program PROTEP[22]. In this procedure, CENDH and the proteins in the PDB are represented in

a simplified manner as linear helices and strands in three-dimensional space. These secondary structure elements and their interrelationships are represented as a mathematical graph[47] which is then searched using a maximal common subgraph algorithm[48]. The strongest similarity with CENDH was observed for 6-phosphogluconate dehydrogenase (PDB entry 1PGN) for which 18 secondary structure elements (8 β strands and 10 α helices) were detected as equivalent in the two structures.

Coordinates. The coordinates for N-(1-D-Carboxylethyl)-L-norvaline dehydrogenase have been submitted to the Brookhaven Data Bank (accession number 1bg6).

Acknowledgments

We thank the support staff at the Synchrotron Radiation Source at CCLRC Daresbury Laboratory for assistance with station alignment. This work was supported by grants from The Wellcome Trust, EU, New Energy and Industrial Development Organisation and International Scientific Research (Joint Research), supported by The Ministry of Education, Science, Sports and Culture of Japan. We acknowledge an award of a travel grant under The British Council/The Royal Society/JSPS Anglo-Japanese Scientific Exchange Scheme. The Krebs Institute is a designated BBSRC Biomolecular Sciences Centre and a member of the North of England Structural Biology Centre (NESBIC).

Received by 27 February 1998; accepted 27 May, 1998.

1. Thompson, J. & Donkersloot, J.A. N-(carboxylalkyl)amino acids: occurrence, synthesis and functions. *Ann. Rev. Biochem*. **61**, 517–557 (1992).
2. Bevan, M. Barnes, W.M. & Chilton, M.D. Structure and transcription of the nopaline synthase gene region of T-DNA. *Nucleic Acids Res.* **11**, 369–385 (1983).
3. Barker, R.F., Idler, K.B., Thompson, D.V. & Chilton, M.D. Nucleotide sequence of the T-DNA region from the *Agrobacterium tumefaciens* octopine Ti plasmid PTI15955. *Plant Mol. Biol.* **2**, 335–350 (1983).
4. Asano, Y., Yamaguchi, K. & Kondo, K. A new NAD^+-dependent opine dehydrogenase from *Arthrobacter* sp. Strain 1C. *J. Bacteriol.* **171**, 4466–4471 (1989).
5. Xuan, J. -W., Fournier, P., DeClerck, N., Chasles, M. & Gaillardin, C. Overlapping reading frames at the Lys5 locus in the yeast *Yarrowia lipolytica*. *Mol. Cell Biol.* **10**, 4795–4806 (1990).
6. Donkersloot, J.A. & Thompson, J. Cloning, expression, sequence-analysis and site-directed-mutagenesis of the TN5306-encoded N-5(carboxylethyl)ornithine synthase from *Lactococcus lactis* K1. *J. Biol. Chem.* **270**, 12226–12234 (1995).
7. Thompson, J. & Miller, S.P.F. N-5-(1-carboxylethyl)ornithine and related [N-carboxylalkyl]-amino acids - structure, biosynthesis and function. *Adv. Enz.* **64**, 317–399 (1991).
8. Drummond, M. Crown gall disease. *Nature* **281**, 343–347 (1979).
9. Smith**,** E.F. & Townsend, C.O. A plant-tumour of bacterial origin. *Science* **25**, 671–673 (1907).
10. DeGreve, H., Decraemer, H., Seurinck, J., Vanmontagu, M. & Schell, J. The functional organisation of the octopine *Agrobacterium tumefaciens* plasmid PTIB6S3. *Plasmid* **6**, 235–248 (1981).
11. Chilton, M.D. *et al.* Stable incorporation of plasmid DNA into higher plant cells: the molecular basis of crown gall tumorigenesis. *Cell* **11**, 263–271 (1977).
12. Veluthambi**,** K., Krishnan, M., Gould, J.H., Smith, R.H. & Gelvin, S.B. Opines stimulate induction of the *vir* genes of the *Agrobacterium tumefaciens* Ti plasmid. *J. Bacteriol*. **171**, 3696–3703 (1989).
13. Kato, Y., Yamada, H. & Asano, Y. Stereoselective synthesis of opine-type secondary amine carboxylic acids by a new enzyme opine dehydrogenase. Use of recombinant enzymes. *J. Mol. Catal.* **B1** 151–160 (1996).
14. Dairi**,** T. & Asano, Y. Cloning, nucleotide sequencing, and expression of an opine dehydrogenase gene from *Arthrobacter* sp. Strain 1C. *Appl. Environ. Microbiol.* **61**, 3169–3171 (1995).
15. Laskowski, R.A., MacArtur, M.W., Moss, D.S. & Thorton, J.M.J. PROCHECK — A program to check the stereochemical quality of protein structures. *Appl. Crystallogr.* **26**, 283–291 (1993).
16. Rossmann, M.G., Moras, D. & Olsen, K.W. Chemical and biological evolution of a nucleotide-binding protein. *Nature* **250**, 194–199 (1974).
17. Wierenga, R.K., De Maeyer, M.C.H. & Hol, W.G.J. Interactions of pyrophosphate moieties with α-helices in dinucleotide binding proteins. *Biochemistry* **24**, 1346–1357 (1985).
18. Lee, B & Richards, F.M. The interpretation of protein structures; estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400 (1971).
19. Jones, S. & Thornton, J.M. Protein–protein interactions: a review of protein dimer structures. *Prog. Biophys. Mol. Biol.* **63**, 31–65 (1995).
20. Abad-Zapatero, C., Griffith, J.P., Sussman, J.L. & Rossmann, M.G. Refined crystal structure of dogfish M = 4 = apo-lactate dehydrogenase. *J. Mol. Biol.* **198**, 445–467 (1987).
21. Baker, P.J., Britton, K.L., Rice, D.W., Rob, A. & Stillman, T.J. Structural consequences of sequence patterns in the fingerprint region of the nucleotide binding fold. Implications for nucleotide specificity. *J. Mol. Biol.* **228**, 662–671 (1992).
22. Grindley, H.M., Artymiuk, P.J., Rice, D.W. & Willett, P. Identification of tertiary structure resemblance in proteins using a maximal common subgraph isomorphism algorithm. *J. Mol. Biol*. **229**, 707–721 (1993).
23. Bernstein, F.C. *et al.* The protein data bank: a computer-based archival file for molecular structures. *J. Mol. Biol.* **112**, 535–542 (1977).
24. Adams M.J., Ellis, G.H., Gover, S., Naylor, C.E., & Phillips, C. Crystallographic study of coenzyme, coenzyme analog and substrate-binding in 6-phosphogluconate dehydrogenase — implications for NADP specificity and the enzyme mechanism. *Structure* **2**, 651–668 (1994).
25. Rosemeyer, M.A. The biochemistry of glucose-6-phosphate dehydrogenase, 6-phosphogluconate dehydrogenase and glutathione reductase. *Cell Biochem. Funct*. **5**, 79–95 (1987).
26. Monneuse, M.-O. & Rouzé, P. Sequence comparisons between *Agrobacterium tumefaciens* T-DNA-encoded octopine and nopaline dehydrogenases and other nucleotide-requiring enzymes: structural and evolutionary implications. *J. Mol. Evol.* **25**, 46–57 (1987).
27. Gäde, G. & Grieshaber, M.K. Pyruvate reductases catalyse the formation of lactate and opines in anaerobic invertebrates. *Comp. Biochem. Physiol.* **83B**, 255–272 (1986).
28. Hart, K.W. *et al.* The importance of arginine 171 in substrate binding by *Bacillus stearothermophilus* lactate dehydrogenase. *Biochem. Biophys. Res. Commun.* **146**, 346–353 (1987).
29. Clarke, A.R. *et al.* An investigation of the contribution made by the carboxylate group of an active site histidine-aspartate couple to binding and catalysis in lactate dehydrogenase. *Biochemistry* **27**, 1617–1622 (1988).
30. Phillips, C., Gover, S. & Adams, M.J. Structure of 6-phosphogluconate dehydrogenase refined at 2Å resolution. *Acta Crystallogr. D* **51**, 290–304 (1995).
31. Goldberg, J.D., Yoyokazu, T. & Brick**,** P. Crystal structure of a NAD-dependent D-glycerate dehydrogenase at 2.4 Å resolution. *J. Mol. Biol.*, **236**, 1123–1140 (1994).
32. Stillman, T.J., Baker, P.J., Britton, K.L. & Rice, D.W. Conformational flexibility in glutamate dehydrogenase. Role of water in substrate recognition and catalysis. *J. Mol. Biol.* **234**, 1131–1139 (1993).
33. Baker, P.J. *et al.* Analysis of the structure and substrate binding of *Phormidium lapideum* alanine dehydrogenase. *Nature Struct. Biol.* **5**, 561–567 (1998).
34. Britton, K.L. *et al.* Crystallization of *Arthrobacter* sp. Strain 1C opine dehydrogenase and its complex with NAD^+. *Acta Crystallogr.* **D54** 124–126 (1998).
35. Hamlin, R. Multiwire area X-ray diffractometers. *Meth. Enz.* **114**, 416–452 (1985).
36. Xuong, N.H., Nielsen, C., Hamlin, R. & Anderson, D. Strategy for data collection from protein crystals using a multiwire counter area detector diffractometer. *J. Appl. Crystallogr.* **18**, 342–350 (1985).
37. Howard, A.J., Nielsen, C. & Young, N.H. Software for a diffractometer with multiwire area detector. *Meth. Enz.* **114**, 452–472 (1985).
38. Leslie, A.G.W. Recent changes to the MOSFLM package for processing film and image data. In *Joint CCP4 and ESF-EADBM Newsletter on Protein Crystallography, no. 26.* (SERC Daresbury Laboratory, Warrington, UK; 1992).
39. Collaborative Computing Project No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
40. Otwinowski, Z. Maximum likelihood refinement of heavy atom parameters in isomorphous replacement and anomalous scattering. *In Proceedings of the CCP4 Study Weekend.* (eds Wolf, W., Evans, P.R. and Leslie, A.G.W.) 80–86 (SERC Daresbury Laboratory, Warrington, UK; 1991).
41. Cowtan, K. DM an automated procedure for phase improvement by density modification. In *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography no. 31.* 34–38 (SERC Laboratory, Daresbury, Warrington, WA4 4AD, UK; 1994).
42. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for the building of protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.* **A47**, 110–119 (1991).
43. Jones, T.A. Interactive computer graphics: FRODO. *Meth. Enz.* **115**, 157–171 (1985).
44. Tronrud, D.E., Ten Eyck, L.F. & Matthews, B.W. An efficient general-purpose least squares refinement program for macromolecular structures. *Acta Crystallogr.* **A43**, 489–501 (1987).
45. Tronrud, D.E. Conjugate-direction minimisation: an improved method for the refinement macromolecules. *Acta Crystallogr.* **A48**, 912–916 (1992).
46. Moews, P.C. & Kretsinger, R.H. Refinement of the structure of carp muscle calcium-binding parvalbumin by model building and difference Fourier analysis. *J. Mol. Biol.* **91**, 201–228 (1975).
47. Mitchell, E.M., Artymiuk, P.J., Rice, D.W. & Willett, P. Use of techniques derived from graph theory to compare secondary structure motifs in proteins. *J. Mol. Biol.* **212**, 151–166 (1990).
48. Bron, C. & Kerbosch, J. Algorithm 457—finding all cliques of an undirected graph. *Commun. ACM* **16**, 575–577 (1973).
49. Kraulis, P.J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
50. Ferrin, T.E., Huang, C.C., Jarvis, L.E. & Langridge, R. The MIDAS display system. *J. Mol. Graphics* **6**, 13–27 (1988).
51. Esnouf, R.M. An extensively modified version of Molscript that includes greatly enhanced colouring capabilities. *J. Mol. Graphics* **15**, 113–138 (1997).