Protein Science (1998), 7:2106–2117. Cambridge University Press. Printed in the USA.

Structure of betaine aldehyde dehydrogenase at 2.1 Å resolution

KENTH JOHANSSON,[1] MUSTAPHA EL-AHMAD,[1,2] S. RAMASWAMY,[1] LARS HJELMQVIST,[2] HANS JÖRNVALL,[2] AND HANS EKLUND[1]

[1]Department of Molecular Biology, Swedish University of Agricultural Sciences, S-751 24 Uppsala, Sweden
[2]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, S-171 77 Stockholm, Sweden

(RECEIVED February 9, 1998; ACCEPTED June 2, 1998)

Abstract

The three-dimensional structure of betaine aldehyde dehydrogenase, the most abundant aldehyde dehydrogenase (ALDH) of cod liver, has been determined at 2.1 Å resolution by the X-ray crystallographic method of molecular replacement. This enzyme represents a novel structure of the highly multiple ALDH, with at least 12 distinct classes in humans. This betaine ALDH of class 9 is different from the two recently determined ALDH structures (classes 2 and 3). Like these, the betaine ALDH structure has three domains, one coenzyme binding domain, one catalytic domain, and one oligomerization domain. Crystals grown in the presence or absence of NAD^+ have very similar structures and no significant conformational change occurs upon coenzyme binding. This is probably due to the tight interactions between domains within the subunit and between subunits in the tetramer. The oligomerization domains link the catalytic domains together into two 20-stranded pleated sheet structures. The overall structure is similar to that of the tetrameric bovine class 2 and dimeric rat class 3 ALDH, but the coenzyme binding with the nicotinamide in *anti* conformation, resembles that of class 2 rather than of class 3.

Keywords: active site cysteine, betaine aldehyde dehydrogenase, conserved residues, crystal structure, NAD^+-binding, substrate pocket

Aldehyde dehydrogenases (ALDHs) catalyze the irreversible oxidation of a broad range of aldehydes to the corresponding acids. The substrates that ALDH work on include aliphatic and aromatic aldehydes, but also 2-enoic, 2-hydroxy, and 2-halogenated aldehydes. Together with alcohol dehydrogenases (ADHs), ALDHs are important components of cellular pathways that metabolize alcohols and aldehydes and they have been ascribed important functions in cellular detoxification and defense systems (Hempel et al., 1993; Jörnvall, 1994). A common feature of both these enzyme families is a highly multiple nature, with, apparently, at least 12 classes (and corresponding genes) of ALDH in humans (Yoshida et al., 1998) and with at least six or seven classes of ADH in vertebrates (Jörnvall & Höög, 1995; Kedishvili et al., 1997).

The human liver aldehyde dehydrogenase classes ALDH1 (cytosolic) and ALDH2 (mitochondrial) have been studied extensively because of their involvement in the conversion of ethanol-derived acetaldehyde. A mutation in the Oriental form of the mitochondrial enzyme results in low activity which is associated with alcohol intolerance (Farrés et al., 1994). Recently, an abundant form of cod liver aldehyde dehydrogenase was purified and characterized (Hjelmqvist et al., unpubl. results). The enzymatic activity during preparation was measured with formaldehyde as substrate. Further characterization demonstrated that this enzyme has high specific activity for betaine aldehyde and is a betaine aldehyde dehydrogenase of the class 9 type (Hjelmqvist et al., unpubl. results). The sequence show 66% identity to the human ALDH9 enzyme compared to 16–35% identity to other aldehyde dehydrogenases. The betaine aldehyde ALDH9 has been thoroughly investigated in humans (Kurys et al., 1989; Chern & Pietruszko, 1995; Pietruszko et al., 1997). One postulated role for this enzyme is in the metabolism of putrescine to γ-aminobutyric acid which is a principal neurotransmitter in the central nervous system. A second metabolic function of ALDH9 is related to its conversion of betaine aldehyde to betaine. Of the major enzyme classes, only ALDH9 catalyzes this reaction at a significant level. Betaine can serve as a methyl donor for the biosynthesis of methionine and may also function as a defense against osmotic stress (Pietruszko et al., 1997). The amount of the ALDH9 enzyme present in human liver has been estimated to be about 4% of the total ALDHs but with high catalytic activity (Pietruszko et al., 1997).

ALDH subunits contain approximately 500 amino acid residues and are mostly tetrameric, compared to the dimeric mammalian ADHs with approximately 375-residue subunits (Persson et al., 1994). The first ALDH primary structure of a class 1 enzyme was

Reprint requests to: S. Ramaswamy, Department of Molecular Biology, Swedish University of Agricultural Sciences, Uppsala S-751-24, Sweden; e-mail: rams@xray.bmc.uu.se.

established in 1984 (Hempel et al., 1984), but three-dimensional structures of two ALDH classes were determined only recently, those of the bovine class 2 ALDH (ALDH2) tetrameric enzyme (Steinmetz et al., 1997) and the rat class 3 ALDH (ALDH3) dimeric enzyme (Liu et al., 1997). Apart from differing in quaternary structure, they also differ in subunit size but they were found to have a largely similar overall fold. In both cases there are three domains, a coenzyme binding domain, a catalytic domain, and an oligomerization domain. However, a major difference appears to exist in the mode of coenzyme binding, where NAD^+ binds deeper into the binding cleft of ALDH2 than of ALDH3. Furthermore, the nicotinamide conformation is *anti* in ALDH2 and *syn* in ALDH3. The latter difference has led to conflicting views of the catalytic mechanism. Because of these differences, the extreme multiplicity of the enzyme (Yoshida et al., 1998) and the differences in evolutionary properties between the ALDH classes (Yin et al., 1991), knowledge of additional structures of ALDHs is essential. We present here the three-dimensional structure of betaine ALDH (ALDH9) from cod liver in its apo and holo forms.

Results and discussion

Structure solution

Betaine aldehyde dehydrogenase was crystallized in the triclinic space group P1 and contains a tetramer in the cell. Two different data sets were collected, one at room temperature on a holoenzyme complex crystal with NAD^+, and a second at 100 K on a native crystal without NAD^+ (Table 1). The crystallization conditions were the same in both cases.

The structure was solved by molecular replacement using the native data set and the tetramer of the bovine enzyme as a search model. After one round of simulated annealing followed by solvent flattening, histogram matching, and fourfold molecular averaging, an interpretable electron density map at 2.5 Å resolution was obtained. The majority of the polypeptide chain was traced from this map. Subsequently, several rounds of model building and refinement were carried out, extending the phases to 2.1 Å, after which the entire polypeptide chain, including both the N- and the C-termini, could be traced, except for residues 31–35. The final refined model contains residues 1–503 in all four subunits and has an *R*-factor of 22.3% and and an R_{free} of 25.3% (Table 2). The coenzyme position was deduced from the holoenzyme crystal data set at 2.8 Å resolution. This data set was refined separately with the native data set as a starting model. Ramachandran plots show only one nonglycine residue, Asn478, in the disallowed regions.

Subunit structure of betaine aldehyde dehydrogenase

The structure of betaine aldehyde dehydrogenase is similar to the two previously solved structures, the bovine ALDH2 (Steinmetz et al., 1997) and rat ALDH3 enzymes (Liu et al., 1997), although the latter is a dimer while the two others are tetramers. Each subunit of betaine ALDH has three distinct domains: a coenzyme binding domain, a catalytic domain, and an oligomerization domain (Fig. 1A,B).

The coenzyme binding domain has an α/β structure of a Rossmann type but, like the other ALDH structures, deviates in several respects from that of the classical dehydrogenases. Alcohol dehydrogenase has the most symmetrically arranged domain with an $(\alpha\beta)_6$ structure (Eklund et al., 1984), which makes it a good partner for comparison. Compared to ADH, the ALDH coenzyme binding domain lacks the last $\alpha\beta$ unit. In this respect, it is more similar to the NAD(P) domain of the ferredoxin-$NADP^+$ reductase superfamily (Ingelman et al., 1997). The αA helix of ADH is not present in ALDH but helix α8 occupies the corresponding space. Similar arrangements are found in other dehydrogenases, e.g., lactate dehydrogenase (Eklund & Brändén, 1987). However, the main difference from a classical coenzyme binding domain, like that of ADH, is the presence of a long extension on the N-terminal side of the domain. The central strand of the domain, βA in ALDH, starts at residue 157. These extra residues mainly forms a cap around the first mononucleotide binding unit of the domain. It provides two extra strands in the central pleated sheet, on the βC side of the classical dehydrogenases, the sequentially first strand is antiparallel to the others and forms hydrogen bonds to βC. There are four extra helices on one side of the sheet, two of which participate in subunit interactions in the tetramer. The N-terminal helix is unique for betaine ALDH and is not present in any of the other ALDH structures. It is located at the surface and is folded back onto the

Table 1. *Data collection statistics*

Data set	Native	NAD complex
Temperature	Liquid nitrogen, 100 K	Room temperature
Space group	P1	P1
Cell parameters (Å)	a = 84.2, b = 86.2, c = 88.3,	a = 85.4, b = 86.6, c = 89.8
(°)	α = 105.2, β = 115.1, γ = 100.0	α = 105.3, β = 113.3, γ = 98.2
No. of unique reflections	90,475	49,803
Resolution (Å)	2.1	2.8
Mosaicity (°)	>1.0	0.5
Completeness (%)		
All data	75.2	92.6
Last shell	40.3 (2.23–2.1 Å)	88.5 (2.98–2.80 Å)
R-merge (%)	5.6	12.4
I/σ		
All data	13.1	5.0
Last shell	5.2 (2.18–2.10 Å)	1.6 (2.99–2.80 Å)

Table 2. *Refinement statistics*

Data set	Native, cryo	NAD complex
Resolution limits (Å)	8.0–2.1	8.0–2.8
R-factor	21.9	23.9
R-free	25.9	25.7
Average *B*-factor		
Protein atoms	16.8	38.0
Water molecules	18.1	29.9
NAD+	—	61.7
All atoms	17.4	38.1
Wilson *B*-factor	14.6	44.1
RMSD bond lengths (Å)	0.011	0.014
RMSD bond angles (°)	2.4	1.7

domain. The other side of the sheet is covered by one additional helix and a long β-hairpin (residues 36–53). This hairpin interacts with the catalytic domain. The residues just before the central βA strand form an extended β-hairpin, which is the main part of the oligomerization domain. The last strand at the C-terminus (residues 487–497) is the third component of this oligomerization domain.

The catalytic domain has a similar fold as the central parts of the coenzyme binding domain, with five central strands. The best superposition of the two domains covers five strands and three helices and superimposes 84 Cα-atoms with an RMS difference of 2.2 Å using the O program (Jones et al., 1991). In this respect, ALDH is similar to a group of D-specific dehydrogenases, including formate dehydrogenase, D-glycerate dehydrogenase, D-lactate dehydrogenase, and D-phosphoglycerate dehydrogenase, for which the similarities between the catalytic and coenzyme binding domains are even more pronounced (Kutzenko et al., 1998). There is no apparent sequence similarity between the structurally aligned domains of ALDH but similarity between the domains was suggested earlier based on calculated hydrophobicity plots (Hempel et al., 1984).

The catalytic domain has a long insertion in the regular domain structure between βB′ and βC′ containing a long β-hairpin, two long helices, and two less ordered regions, together about 100 residues. The β-hairpin and the less ordered regions form most of the contact surface toward the coenzyme binding domain.

There are two covalent connections between the coenzyme binding and the catalytic domains; one at residues 265–266 where the sheet in the coenzyme binding domain continues until the sheet of the catalytic domain starts, and the other after the last strand of the catalytic domain, residues 456–471, where a long β-hairpin fills one end of the cleft between the two domains. The first strand of the β-hairpin forms hydrogen bonds to the catalytic domain while the last strand forms a short antiparallel stretch with three hydrogen bonds to the last strand of the coenzyme binding domain. Subsequent to this β-hairpin, there is a short α-helix that interacts with the coenzyme binding domain. The chain ends with the C-terminal third strand of the oligomerization domain. The last 50 residues connects all three domains in the subunit.

In general, the interactions between the coenzyme binding and the catalytic domains are more pronounced than in the case of ADH and other related dehydrogenases. There are a number of hydrogen bonds on both sides of the cleft between the domains. The presence of these tight connections is probably one of the reasons why, unlike ADH, there are no conformational changes between the NAD^+ and NAD^+-free form of the enzyme. The RMS differences between the two structures is 0.31 Å for all Cα atoms. Similar results have been obtained for the other aldehyde dehydrogenase structures (Liu et al., 1997; Steinmetz et al., 1997).

Tetramer structure

The subunits of ALDH associate into dimers (Fig. 2A) such that the last strand of the oligomerization domain (residues 491–495) is hydrogen-bonded to the last strand of the catalytic domain in the second subunit. Together, the oligomerization domain of one subunit and the catalytic domain of another form a ten-stranded pleated sheet. The oligomerization domain can thus be considered to be an extension of the coenzyme binding domain. With the arrangement of the ten-stranded sheet, the domains are kept locked relative to each other in a much tighter manner than in, for example, ADH. On the other hand, the coenzyme binding domains do not come sufficiently close to the last strand of the symmetry related domain to form hydrogen bonds as in ADH. Instead, these strands are separated by 10 Å from each other in ALDH. Thus, the residues connecting the catalytic domain and the last strand of the oligomerization domain form a short antiparallel connection to parts of the last strand of the coenzyme binding domain. The side chain of the last residue in this connection, Phe474 from both subunits in the dimer, is located between the center of the last strands of the two coenzyme binding domains, keeping them separated from each other. The αE helices of the coenzyme binding domains have numerous van der Waals interactions to each other in the dimer.

The tetramer is formed by two dimers that interact back to back at the side opposite to where the coenzyme binds (Fig. 2B). Both of the two oligomerization domains of the dimer form hydrogen bonds to the corresponding parts of the other dimer such that two 20-stranded sheets are formed through the tetramer. Both tetramerization and dimerization buries huge accessible surface areas, 2,176 and 2,021 $Å^2$, respectively.

Comparisons with other ALDH structures

The cod betaine ALDH sequence is only 37 and 21% identical to those ALDHs for which the three-dimensional structures are known, the tetrameric bovine ALDH2 and the dimeric rat ALDH3, respectively (Fig. 3). In spite of this low sequence similarity, the overall subunit structures are very similar. In fact, 464 residues of the ALDH2 structure can be superimposed on the ALDH9 structure, with an RMS deviation (RMSD) of 1.1 Å, and 360 residues of the ALDH3 structure can be superimposed on ALDH9 with an RMSD of 1.7 Å using standard criteria in O (Jones et al., 1991). When dimers are superimposed, 927 residues of the ALDH2 structure can be superimposed on ALDH9 with an RMSD of 1.1 Å, while 655 Cα atoms of the ALDH3 structure can be superimposed on ALDH9, with an RMSD of 2.0 Å. This demonstrates that the dimer organization within the tetrameric enzymes is the same, and that the organization of the subunits in the dimeric ALDH deviates very little from that of the subunits in the tetrameric enzymes.

Coenzyme binding

We first determined the structure of ALDH in the absence of NAD^+. However, in the averaged map, traces of coenzyme were

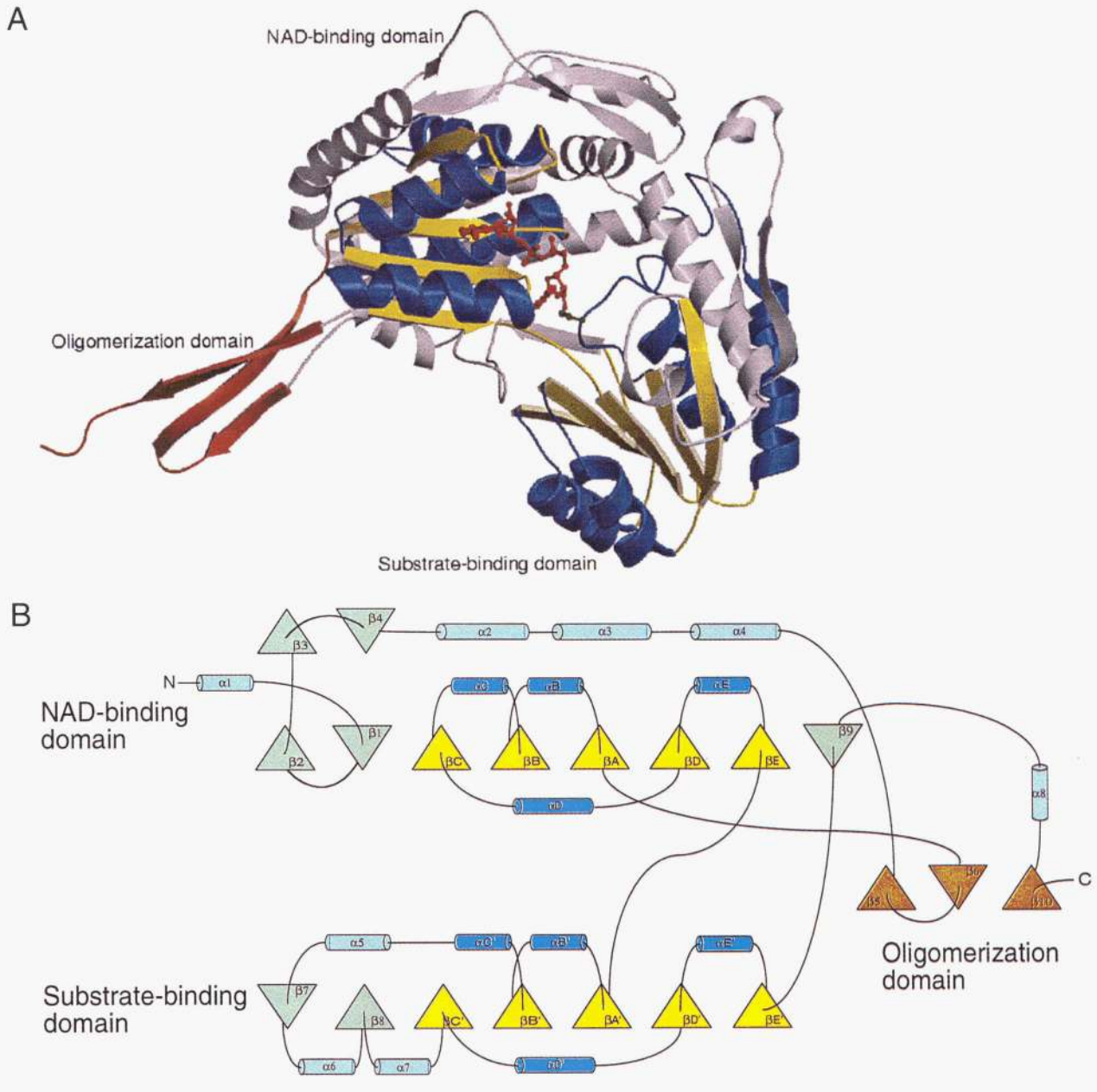

Fig. 1. Domain organization of the ALDH monomer. **A:** The structure of one subunit of betaine ALDH where the common strands and helices in the Rossmann type fold of the two domains are colored yellow and blue, respectively. The remaining helices and strands of these domains are colored gray, the oligomerization domain red. The bound NAD^+ is shown in a red ball and stick model and the active site cysteine in green. **B:** Schematic drawing of the secondary elements of the three domains colored as in **A**. Drawings were made by MOLSCRIPT (Kraulis, 1991) and rendered using Raster 3D (Bacon & Anderson, 1988; Merritt & Murphy, 1994).

still present. It was mainly the ADP part of NAD^+ that was visible (at about 25% occupancy). We then also crystallized ALDH in the presence of 1 mM NAD^+. In the maps from these crystals, there is density for the complete coenzyme but the ADP part has stronger density, indicating some flexibility of the nicotinamide and its ribose (Fig. 4A).

The overall binding of the coenzyme is significantly different between ALDHs on the one hand, and classical dehydrogenases on the other (Fig. 4B,C). An important difference between the coenzyme binding domains of ALDHs and ADH is the segment for the carboxyl end of βA to the beginning of αB. In ADH, this part contains the famous "fingerprint sequence" with glycine residues (Wierenga et al., 1986). In ALDHs, there is, instead, a protruding loop corresponding to where the coenzyme binding site is in ADH, and there is no glycine motif in ALDHs as observed in attempts at defining the coenzyme binding site when only the primary structure was known (Hempel et al., 1984). A proline residue (position 168 in ALDH9) prevents the formation of a helix turn. Conserved Trp165 and Asn166 are also located in this turn. The consequence of this difference is that NAD^+ is pushed out toward the surface compared to NAD^+ binding in ADH. The phosphates have moved out about 13 Å from the position in ADH, while the

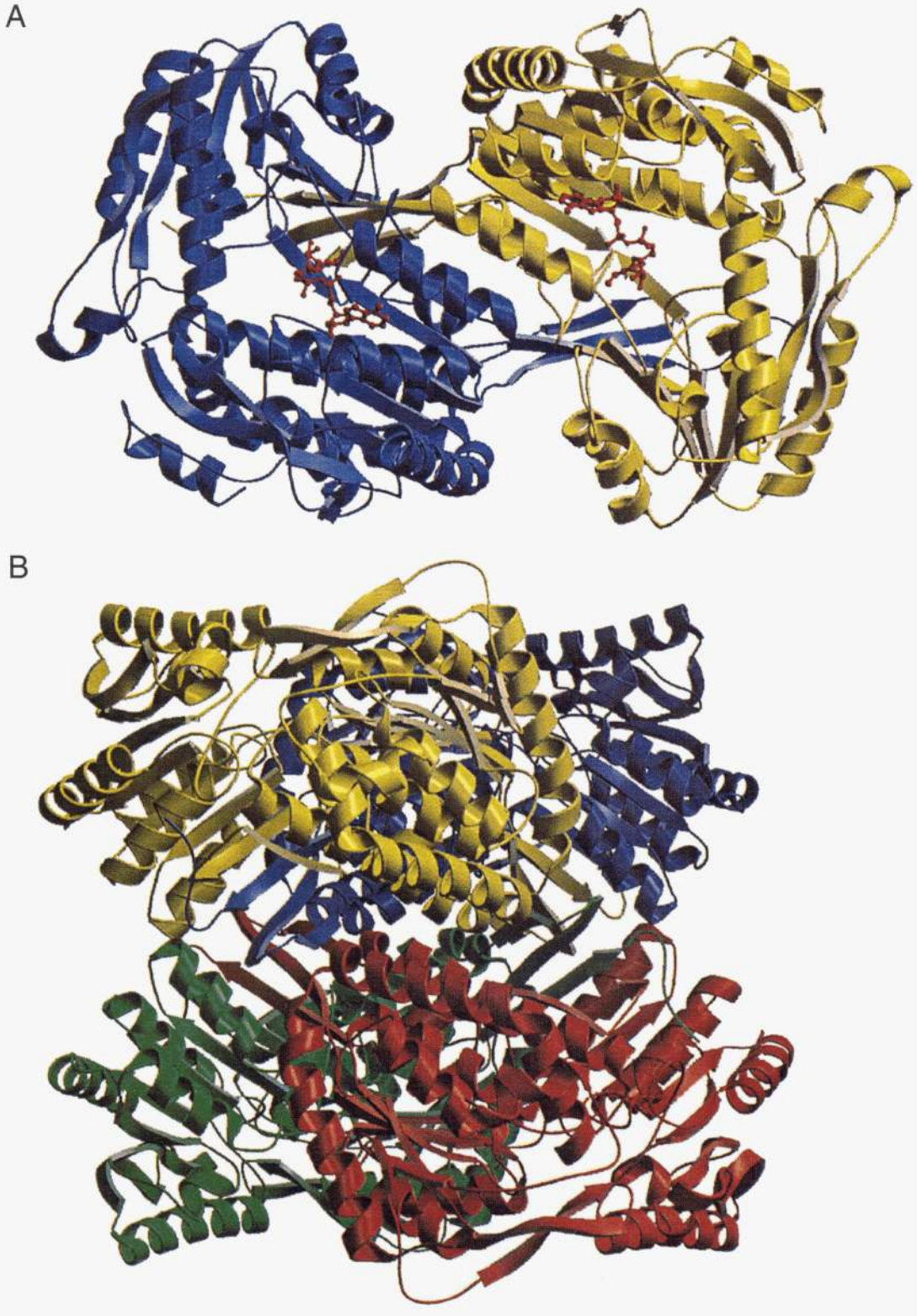

Fig. 2. Dimer and tetramer organization of ALDH. **A:** The ALDH subunits form dimers where the oligomerization domain pleated sheet is continued into the β sheet of the catalytic domain of the other subunit. The sheet of the coenzyme binding domain is not directly connected to its counterpart in the second subunit. **B:** The tetramer with the yellow-blue dimer interacting back to back with the red-green dimer.

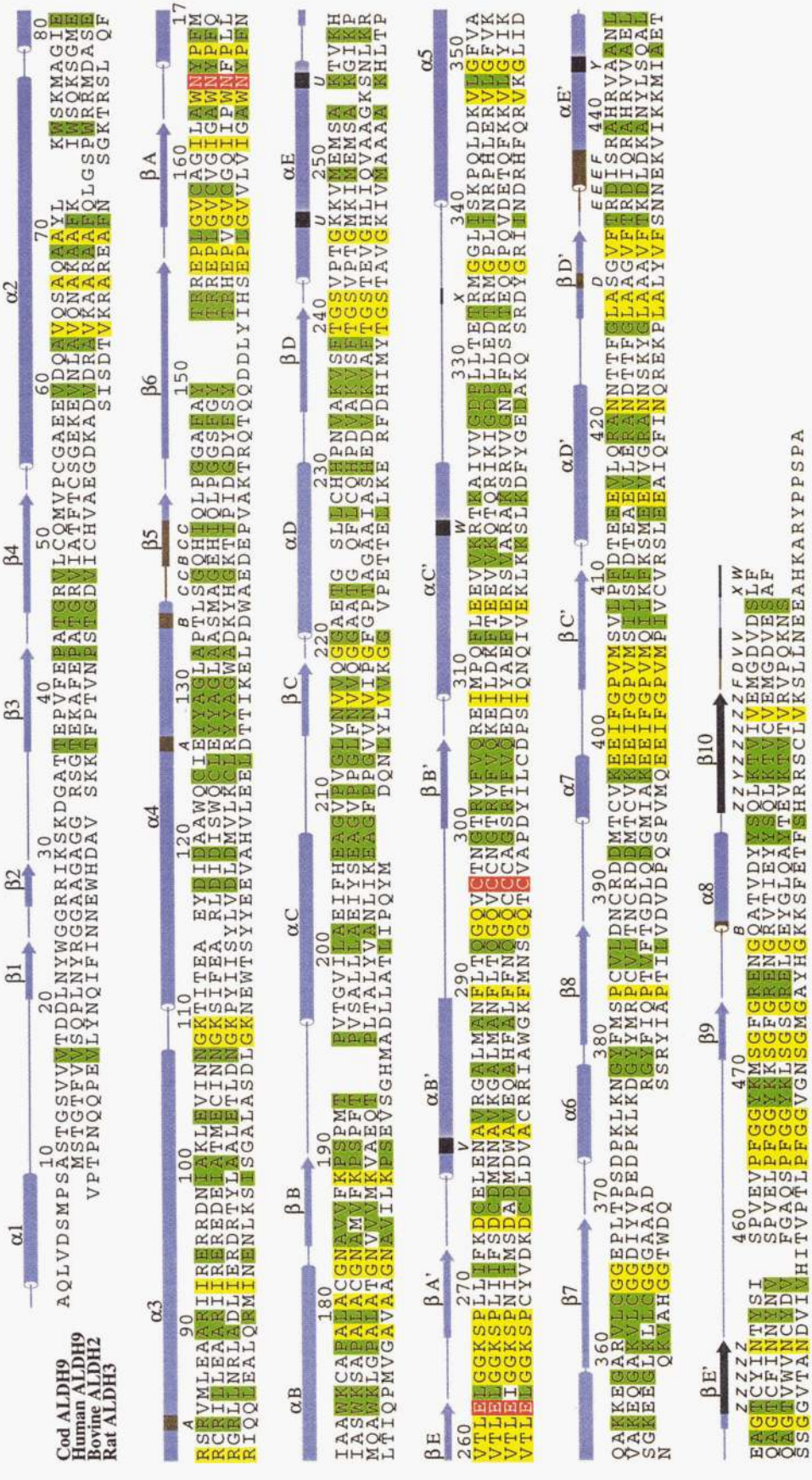

Fig. 3. Structural alignment of ALDH sequences. The sequences of betaine ALDH from cod liver and the corresponding human betaine ALDH 9, structurally aligned with the bovine ALDH2 and rat ALDH3. The secondary structure elements for betaine ALDH from cod liver are shown on top of the sequences. Active site residues Asn166, Glu263, and Cys297 are colored red, whereas other residues, which are the same in all four sequences, are colored yellow and those that are the same in three of the sequences are colored green. Secondary structure elements shown in blue or red indicate residues participating in dimer (Fig. 2A) or tetramer interactions, respectively. The accompanying letter below the blue/red residues indicates interacting residues, i.e., red A interacts with red A in tetramer formation, and blue U interacts with blue U in dimer formation, etc. The figure was made using ALSCRIPT (Barton, 1993).

adenine ring has moved about 9 Å. This also results in an insertion of the adenine ring deeper between helix αD and αE, previously shown (Liu et al., 1997). The shift also results in binding of the phosphates to the amino end of αE instead of to αB as in ADH. The loop between βD and αE contains the strictly conserved Gly residue in the sequence Thr-Gly-Ser and, after three additional residues, an additional, highly conserved Gly. This motif is reminiscent of the glycine motif in the classical dehydrogenases. It is, however, located in the second half of the coenzyme binding domain instead of in the first one, but at a roughly equivalent position. This mode of coenzyme binding is similar for ALDH9 and the other known ALDH structures (Liu et al., 1997; Steinmetz et al., 1997).

The interactions between the enzyme and NAD^+ are summarized in Figure 4D. The adenine ring of ALDH9 is inserted between Ala222–Gly225 of αD and Thr242–Leu245 of αE. The adenine pocket is limited on the N1 side by Ser226 and Met252 and on the N3 side by Ile162. The N6–N7 face is directed toward the solvent. The adenosine ribose ring has 2′-endo conformation and is hydrogen bonded by O2′ and O3′ to the conserved Lys189, and by O3′ to the carbonyl oxygen of Leu163. One oxygen of one of the phosphate groups is hydrogen bonded to both the main-chain nitrogen and the side chain of the conserved Ser242 and to the side chain of Thr245. One oxygen atom of the second phosphate makes a hydrogen bond to the conserved Trp165. The NMN-ribose oxy-

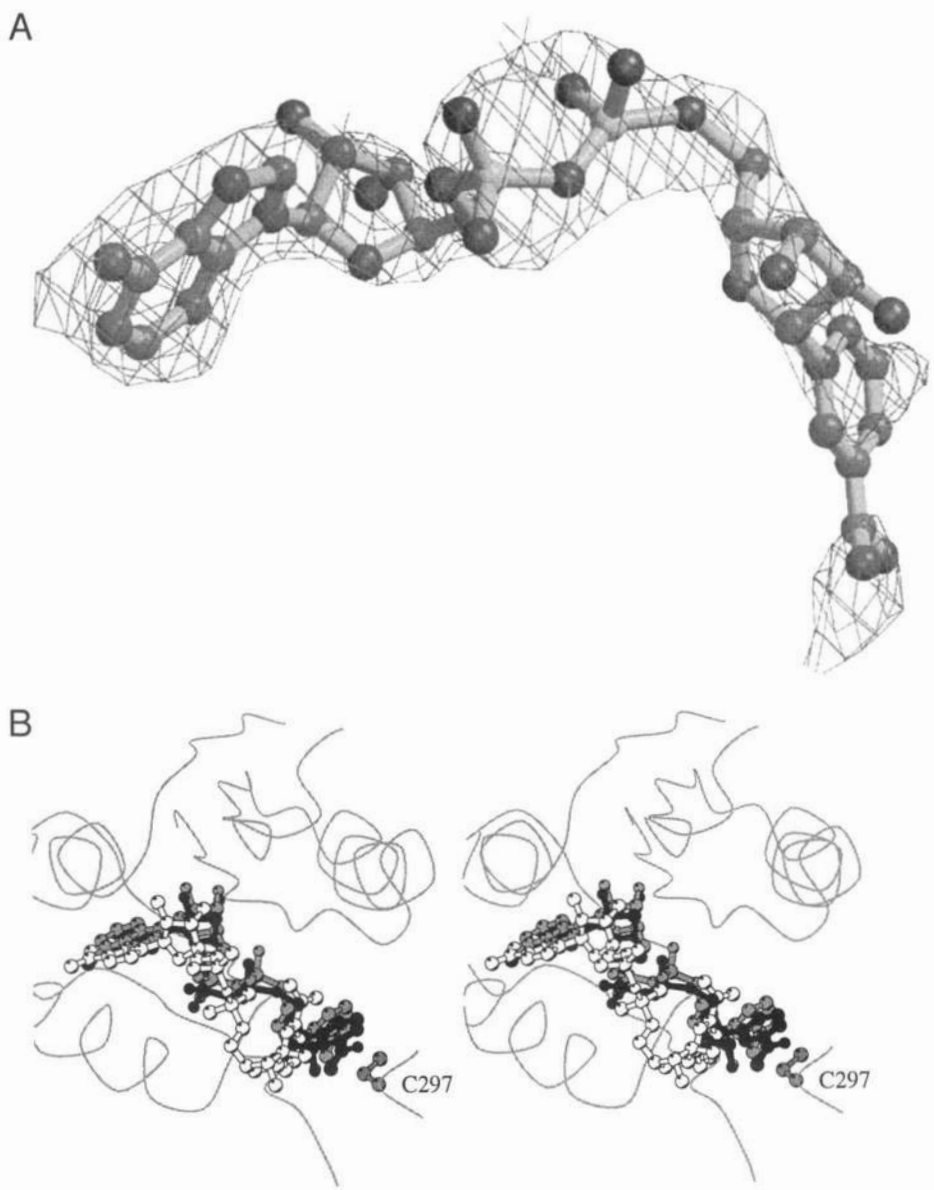

Fig. 4. Coenzyme binding. **A:** The omit $F_o - F_c$ difference density for bound NAD^+ in the coenzyme complex (contoured at 3σ). The ADP part has stronger density, indicating some flexibility of the nicotinamide and its ribose. **B:** Stereoview of the binding of the coenzyme in betaine aldehyde dehydrogenase (black), ALDH2 (gray), and ALDH3 (white) to the coenzyme binding domain. The active site cysteine is also shown. **C:** Binding of NADH to ADH. The overall binding of the coenzyme for ALDHs is significantly different from that of ADH and the other classical dehydrogenases. **D:** The interactions between NAD^+ and the enzyme. Hydrogen bonds are drawn as thin lines. (*Figure continues on facing page.*)

gens make hydrogen bonds to the conserved Glu400. The ribose and nicotinamide ring are less well defined in the crystal structure.

The coenzyme of ALDH9 is bound in a similar manner as in the tetrameric ALDH2, whereas the binding to the dimeric ALDH3 differs in several ways (Fig. 4B). For example, the coenzyme bound to ALDH3 is further out in the cleft and there is an indirect binding of the adenosine ribose via a water molecule to the conserved lysine, whereas in the two tetrameric ALDHs, there are hydrogen bonds to both O2′ and O3′ from the corresponding lysine. The adenosine ribose of both ALDH2 and ALDH3 hydrogen bonds to a glutamic acid (Glu195 in ALDH2, which is equivalent to Glu140 in ALDH3). At the corresponding position in ALDH9, there is no such interaction and the corresponding residue is a Pro. The absence of negative residues at the adenosine binding site and the presence of a lysine residue suggest that $NADP^+$ should work better with ALDH9 than with the other ALDHs.

ALDH9, like ALDH2, has no positively charged residue at the diphosphate binding site of NAD^+, the only residue in the neighborhood of the phosphates is the conserved Trp165 and the diphosphate site is remarkably open. ALDH3 has a His and an Arg in the neighborhood of the phosphates. The relative position of the coenzyme in the dimeric enzyme is significantly different from that of the tetrameric ones. The nicotinamide in ALDH3 is roughly at the position where the NMN-ribose is in both ALDH9 and ALDH2, and the nicotinamide rings occupy very different sites. This is surprising in view of the very similar location of the active

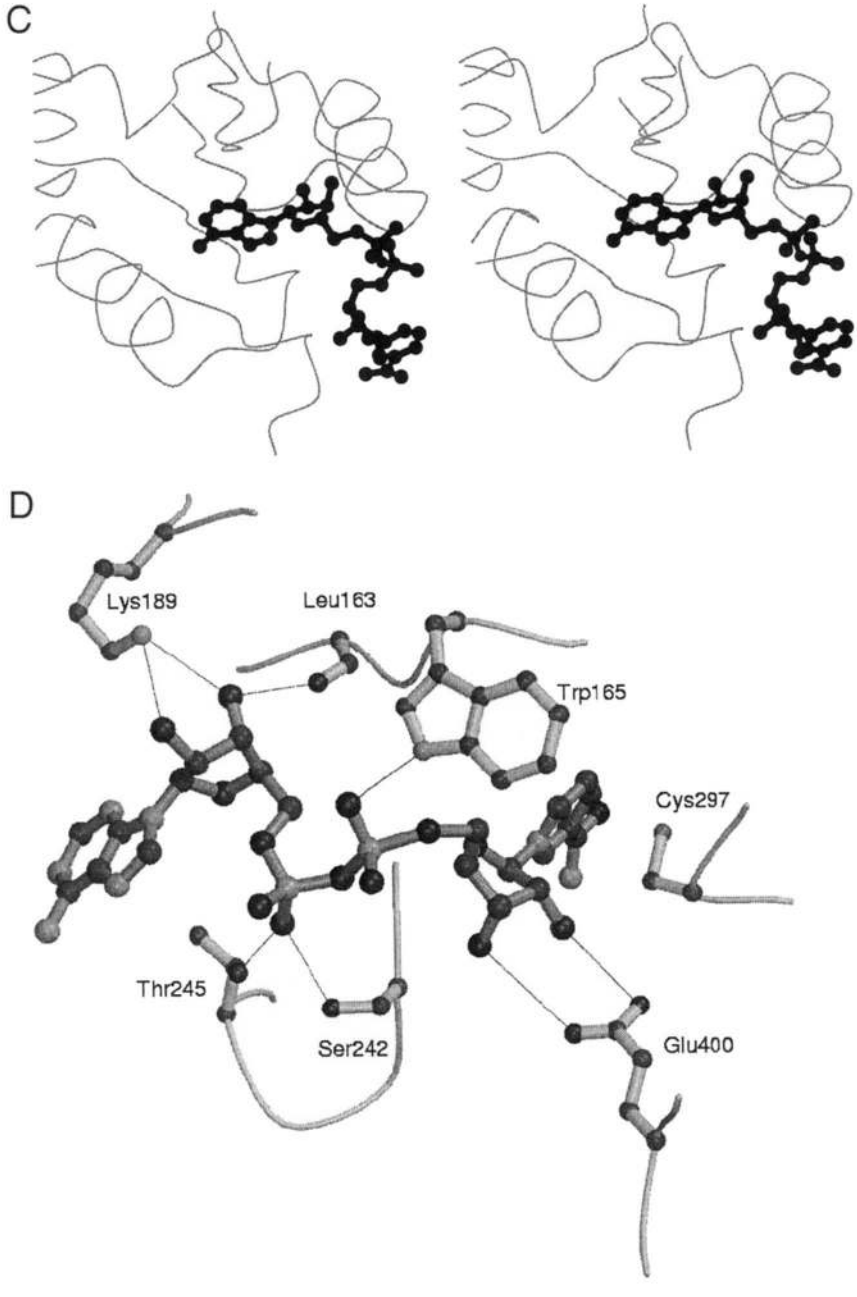

Fig. 4. *Continued.*

site in the two enzymes but has been suggested to reflect different oxidation states of the coenzyme (Steinmetz et al., 1997). The two earlier ALDH structures suggest very different bindings of the nicotinamide and ribose. In the dimeric enzyme, the nicotinamide is in *syn* conformation, whereas it is in *anti* conformation in the tetrameric enzymes. NAD^+ bound to betaine aldehyde dehydrogenase has an *anti* conformation of the nicotinamide very similar to that in ALDH2. The nicotinamide, as positioned in the difference density, forms hydrogen bonds to Glu263. Asn166 seems suitably positioned to bind to the substrate.

Active site

ALDH, like GAPDH, uses covalent intermediates to convert an aldehyde into an acid (Weiner, 1979). Removal of a hydride ion from an aldehyde is energetically costly because of the dipolar character of the carbonyl group (Stryer, 1995). Aldehydes bind to an active site cysteine residue and form a covalently attached alcohol intermediate that is readily oxidized to an aldehyde. The substrate then leaves as an acid by the attack of water. Cys297 is the active site cysteine residue in betaine ALDH, and there is a tunnel leading from the surface of the molecule to this residue. This channel is partly hydrophobic but also contains a number of hydrophilic residues: Asn166 and Thr298 are about 5 Å away from the cysteine, Glu263 is 6 Å away on the inside close to the carboxyamide of the nicotinamide, and both Tyr167 and Tyr453 are about 8 Å away. There is no positively charged residue in the neighborhood of Cys297. Instead, the nucleophilic character of the cysteine may be induced by the positioning of the positively charged nicotinamide ring close to the cysteine. The nicotinamide in our crystals is positioned close to the cysteine residue with its A-side toward the cysteine residue.

A betaine aldehyde was modeled covalently bound to Cys297 and the geometry of the nicotinamide and the substrate was optimized to resemble that of the hydride transfer step (Fig. 5). This required repositioning of the nicotinamide and ribose by about 1 Å further into the active site cleft than in the observed binding. The hydride transfer is then to the A-side of the nicotinamide ring in agreement with experimental evidence (Jones et al., 1987). The nicotinamide ring is then hydrogen bonded with its carboyxamide to Glu263, its C4 atom is close to Asn166 and the ring sandwiched between Thr240 and Cys297. The NMN-ribose is bound to Glu400. The substrate oxygen is hydrogen bonded to Asn166 and the main-chain nitrogen of residue 297. There is a large, essentially hydrophobic, pocket around the modeled betaine aldehyde, lined by Tyr167, Met170, Trp174, Val296, Thr298, Ser460, and Phe466. The active site is obviously suitable for large aldehyde substrates, in agreement with kinetic data (Hjelmqvist et al., unpubl. results). This pocket has several similarities to the other ALDH structures (Table 3) but some significant differences can be noted. The pocket is smaller in the other structures, mainly because of Met174 and Phe459 in ALDH2 and Leu119 and Ile394 in ALDH3, instead of Ile171 and Ser460 in ALDH9.

Our modeling of the transition state complex is very similar to that made with ALDH2 (Steinmetz et al., 1997) but differs in some important respects to that of ALDH3 (Liu et al., 1997) The feature that is common to all three models is, besides the binding of the substrate to the cysteine residue, a hydrogen bond between the substrate oxygen and the asparagine (residue 166 in ALDH9). Another hydrogen bond between the substrate oxygen and the main chain of the active site cysteine is also common between the ALDH2 and ALDH9 models, as well as the position of the nicotinamide ring. The ALDH3 model has the nicotinamide ring in *syn* conformation and, consequently, should accept the hydride at the B-side of the nicotinamide ring during the reaction, unless conformational changes specific for the dimeric enzyme would occur (Liu et al., 1997).

The cod ALDH9 enzyme has a low K_m and high k_{cat} for betaine like the human class 9 enzyme (Hjelmqvist et al., unpubl. results) in contrast to other classes (Kurys et al., 1989). There is no negatively charged residue in the substrate pocket that should be responsible for specific interactions with the trimethylamino group of the betaine aldehyde. This group, instead, may interact with Trp174 similar to acetylcholine esterase, where the quaternary ion of acetylcholine binds in an aromatic site, primarily with Trp84 (Sussman et al., 1991, 1993; Gilson et al., 1994; Silman et al., 1994). How-

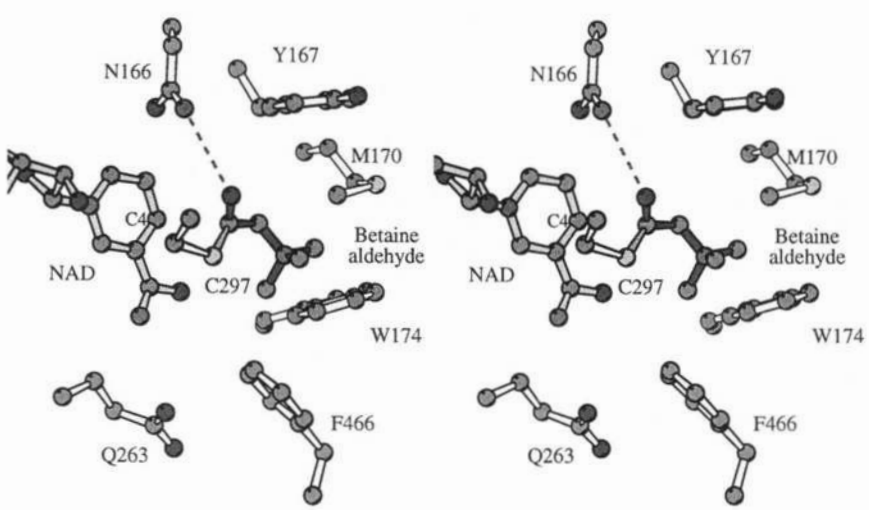

Fig. 5. Substrate model. A betaine aldehyde modeled covalently bound to Cys297 with the geometry of the nicotinamide and the substrate optimized to resemble the hydride transfer step. The surrounding of the modeled substrate is shown.

Table 3. *Substrate binding pocket*[a]

ALDH9	ALDH2	ALDH3
Interactions with the substrate oxygen		
Asn166	Asn169	Asn114
Cys297	Cys302	Cys243
Substrate binding pocket		
Tyr 167	Phe170	Tyr115
Met170	Leu173	Asn118
Trp174	Trp177	Gln122
Glu263	Glu268	Glu209
Ser460	Phe459	Ile394
Phe466	Phe465	Phe401

[a]Residues interacting with the substrate and residues lining the substrate pocket in the ALDH9. The corresponding residues in the other aldehyde dehydrogenase structures are listed.

ever, several other aldehyde dehydrogenase enzymes also have a Trp at a corresponding position to Trp174 in cod ALDH9. As judged from the ALDH2 structure, Met174 (corresponding to Ile171 in cod ALDH9) would prevent access to this site in ALDH2.

Conserved residues in ALDHs

When the sequence of betaine ALDH from cod liver is compared with those of other betaine ALDHs, and with all other known vertebrate forms of ALDH, the number of strictly conserved residues in the ALDH family is significantly reduced from that previously reported (Hempel et al., 1993). Of the 15 residues now strictly conserved (Table 4), many are clustered around the active site and coenzyme binding site, but approximately the same number of residues are conserved for structural reasons, such as internal hydrophobic cores and β-turns. A few residues at the subunit interaction area are also conserved.

Material and methods

Crystallization and data collection

Crystals of betaine aldehyde dehydrogenase were grown at 14 °C from solutions containing 0.1 M HEPES, pH 7.5, 9.5% isopropanol, 20% PEG4000, and 4 mg/mL protein (and 1 mM NAD^+ for the NAD^+ co-crystals). The enzyme was in 20 mM bis-tris buffer, pH 7.1, 150 mM NaCl, and 0.1 mM DTE. Crystallization was carried out by the hanging drop vapor diffusion method. The crystals belong to the triclinic space group P1, with cell dimensions of a = 84.2 Å, b = 86.2 Å, c = 88.3 Å, α = 105.2°, β = 115.1°, and γ = 100.0° for the native structure, and a = 85.4 Å, b = 86.6 Å, c = 89.8 Å, α = 105.3°, β = 113.3°, and γ = 98.2° for the NAD^+ structure. There are four molecules per asymmetric unit related by proper 222 noncrystallographic symmetry (NCS).

The native data set was collected on BM14 at ESRF (Grenoble) using a Princeton CCD camera. The crystals were flash frozen at 100 K after a short dip into a cryo solution consisting of a 50% isopropanol and 50% PEG4000 solution mixed 1:1 with mother liquor. Data for the NAD^+ co-crystals were collected at room

Table 4. *Conserved residues in aldehyde dehydrogenases of the 44 different sequences available in the sequence databases*

Residue	Comment
Strictly conserved	
Gly157	At a break-point just before a β-strand in the coenzyme binding domain
Gly183	Ending of αB, in a β-turn, main-chain torsion angles in Gly-area of the Ramachandran plot
Lys189	Forms hydrogen bonds to O2′ and O3′ in the adenosine ribose and to O190
Gly241	At the NAD^+ phosphate binding site, a side chain should interfere with NAD^+ main-chain torsion angles in Gly-area of the Ramachandran plot
Gly265	Very close to the ribose in the NMN part of NAD^+. A side-chain would disturb NAD^+ binding.
Gly294	In β-turn
Cys297	Nucleofile at the active site
Gly366	Hydrogen bond from the main-chain nitrogen to O383. Main-chain torsion angles in Gly-area of the Ramachandran plot.
Pro384	At the beginning of a β-strand, close to Pro404.
Glu400	Hydrogen bonds to NMN ribose
Phe402	Toward the nicotinamide ring
Pro404	At the beginning of a β-strand
Gly450	Hydrogen bond to N491 in another subunit
Asn455	Hydrogen bond to N434 and O434
Gly468	Close to Leu428 at the active site
Conserved in all but one of the sequences	
Arg82	Hydrogen bond to O77. Toward subunit interaction but no intermolecular hydrogen bonds
Pro155	At a breakpoint between the oligomerization domain and the coenzyme binding domain
Ile162	Packs against the adenosine part of NAD^+ (C2, 3A, and O′4) and against Lys189, which binds the ribose
Asn166	van der Waals contact to C5 in the nicotinamide ring, 5.3 Å from the nucleofile Cys297 at the active site
Pro168	In the beginning of αB, prevent that a phosphate binding site is formed
Gly246	Internal in the first turn of αE. Equivalent to the last Gly in the Gly-motif (in the first turn of αB) for classical dehydrogenases
Gln295	Hydrogen bond to N340. In contact with Trp165 at the NAD^+ binding site.
Glu399	At the surface
Gly403	Close to the active site
Asn422	At the surface. Hydrogen bond to the main chain O446 and N448.
Ser472	Hydrogen bond to N427 and O265
Gly475	Internal in a turn between strands. Main-chain torsion angles in Gly-area of the Ramachandran plot.
Conserved in all but two of the sequences	
Lys110	Hydrogen bond to O166 in the active site and to Asp118
Ala179	Internal in the end of αB
Thr240	At the active site, in contact with the nicotinamide C4 and Glu263. Forms a hydrogen bond to Lys175.
Glu263	At the active site. 4.2 Å from the nicotinamide. 6.2 Å from Cys297.
Lys267	Hydrogen bond to O396 (C-terminal capping of αK).
Gly365	Main-chain torsion angles in Gly-area of the Ramachandran plot.
Leu428	At the active site. In contact with Cys297 and Gly468 and the nicotinamide ring.
Gly467	Close to Glu263 at the active site
Gly473	Close to the main chain of Glu263. Main-chain torsion angles in Gly-area of the Ramachandran plot.
Gly479	Carbonyl hydrogen bond to Tyr129. Main-chain torsion angles in Gly-area of the Ramachandran plot.

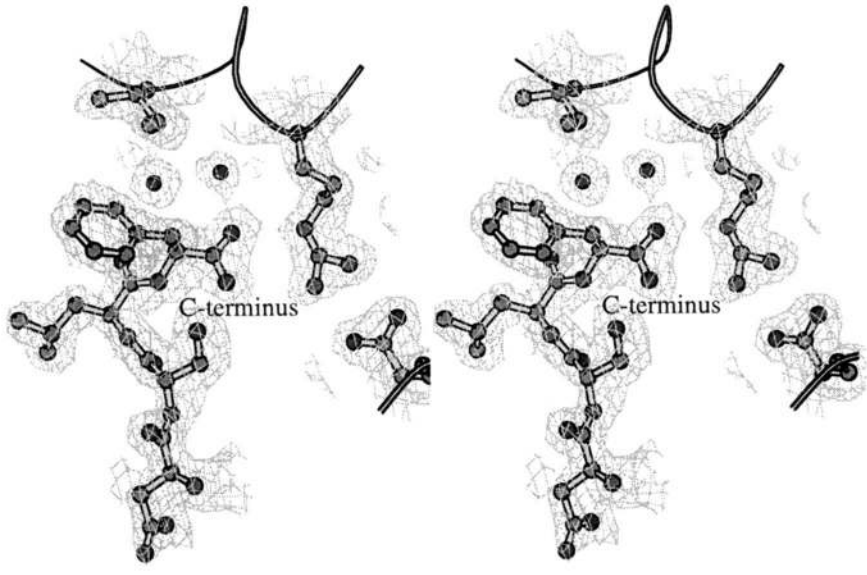

Fig. 6. Representative final electron density map. The final 2.1 Å resolution $2F_o - F_c$ map showing the C-terminus of the molecule and surroundings contoured at 1σ.

temperature on a RAXIS IIC area detector equipped with a Rigaku RU200 rotating anode. The data were indexed, integrated, merged, and scaled using the HKL package (Otwinowski, 1993). The native data set had a mosaicity larger than 1°, but during processing, it was set to 1°.

Phasing and structure refinement

The crystal structure of betaine aldehyde dehydrogenase was solved by molecular replacement with AMORE (Navaza, 1994) using the native cryo data set and a polyalanine model of the tetrameric bovine ALDH2 enzyme (Steinmetz et al., 1997) as search model. For creating an initial working model with all insertions, deletions, and amino acids mutated in the polyalanine model, the program MODELLER (Sali & Blundell, 1993) was used.

The initial model was refined from data between 8.0 and 2.5 Å, with strict fourfold NCS constraints by several rounds of simulated annealing using CNS (Rice & Brünger, 1994), and molecular averaging, solvent flattening, and histogram matching using DM (CCP4, 1994; Cowtan, 1994). The program O was used to build the model and examine the electron density maps (Jones et al., 1991). During the later stages of refinement, when the whole chain (except residues 31–35 in a loop) could be traced, phases were extended to 2.1 Å and NCS restraints were used to make differences possible between individual subunits. Though data are incomplete below 2.5 Å, all data were used as we have used fourfold NCS restraints, which compensates for the low completeness by introducing redundancy in the data. The minimization steps were made using the programs REFMAC (Murshudov et al., 1997) and SHELX (Sheldrick & Schneider, 1997). No significant differences in main-chain atoms between monomers were observed (average RMSD for Cα's is 0.077 Å). Representative density from the final 2.1 Å resolution map is shown in Figure 6.

Accessibility calculations were made by the program SURFACE (CCP4, 1994). A probe of radius 1.40 Å (i.e., water) was used.

Structure comparisons were made with tetrameric bovine ALDH2 enzyme (1ag8) and the dimeric rat ALDH2 enzyme (1ad3).

The final model contains 2,012 amino acids and 727 water molecules. The *R*-factor is 22.3% and R_{free} calculated from a test set with 2% randomly selected reflections is 25.3%. Ramachandran plots, generated by the program PROCHECK (Laskowski et al., 1993), show only one nonglycine residue, Asn478, in the disallowed regions. The coordinates and structure factors have been deposited at the Protein Data Bank (accession code 1a4s).

The coenzyme containing crystals were refined separately with the native structure as starting model. Simulated annealing refinement was carried out using the program CNS with strict NCS constraints followed by fourfold averaging using DM (CCP4, 1994; Cowtan, 1994). The NAD molecule was fitted in the averaged omit map, and the coordinates were included in the final positional refinement (Table 2). The coordinates have been deposited at the Protein Data Bank (accession code 1bpw).

Acknowledgments

We thank Thomas Hurley for early access to the coordinates of ALDH2 and Andy Thompson for valuable advice on data collection. This work was supported by grants from the Swedish Natural Science and Medical Research Councils and the Swedish Alcohol Research Fund.

References

Bacon DJ, Anderson WF. 1988. A fast algorithm for rendering space-filling molecule pictures. *J Mol Graphics 6:*219–220.

Barton GJ. 1993. ALSCRIPT: A tool to format multiple sequence alignments. *Protein Eng 6:*37–40.

CCP4. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Cryst D50:*760–763.

Chern MK, Pietruszko R. 1995. Human aldehyde dehydrogenase E3 isozyme is a betaine aldehyde dehydrogenase. *Biochem Biophys Res Commun 213:*561–568.

Cowtan K. 1994. DM. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography 31:*34–38.

Eklund H, Brändén C-I. 1987. Crystal structure, coenzyme conformations, and protein interactions. In: Dolphin D, Poulson R, Avramovic O, eds. *Pyridine nucleotide coenzymes*. New York: John Wiley & Sons. pp 51–98.

Eklund H, Samama J-P, Jones TA. 1984. Crystallographic investigations of nicotinamide adenine dinucleotide binding to horse liver alcohol dehydrogenase. *Biochemistry 23*:5982–5996.

Farrés J, Wang X, Takahashi K, Cunningham SJ, Wang TT, Weiner H. 1994. Effects of changing glutamate 487 to lysine in rat and human liver mitochondrial aldehyde dehydrogenase. *J Biol Chem 269*:13854–13860.

Gilson MK, Straatsma TP, Mccammon JA, Ripoll DR, Faerman CH, Axelsen PH, Silman I, Sussman JL. 1994. Open back door in a molecular dynamics simulation of acetylcholinesterase. *Science 263:*1276–1278.

Hempel J, Nicholas H, Lindahl R. 1993. Aldehyde dehydrogenases—Widespread structural and functional diversity within a shared framework. *Protein Sci 2*:1890–1900.

Hempel J, von Bahr-Lindström H, Jörnvall H. 1984. Aldehyde dehydrogenase from human liver. Primary structure of the cytoplasmic isoenzyme. *Eur J Biochem 141*:21–35.

Ingelman M, Bianchi V, Eklund H. 1997. The three-dimensional structure of flavodoxin reductase from *Escherichia coli* at 1.7 Å resolution. *J Mol Biol 268:*147–157.

Jones KH, Lindahl R, Baker DC, Timkovich R. 1987. Hydrid transfer stereospecificity of rat liver aldehyde dehydrogenase. *J Biol Chem 262*:10911–10913.

Jones TA, Zou JY, Cowan SW, Kjeldgaard M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr 47*:110–119.

Jörnvall H. 1994. The alcohol dehydrogenase system. In: Jansson B, Jörnvall H, Rydberg U, Terenius L, Vallee BL, eds. *Toward a molecular basis of alcohol use and abuse*. Basel, Switzerland: Birhäuser, pp 221–229.

Jörnvall H, Höög JO. 1995. Nomenclature of alcohol dehydrogenases. *Alcohol Alcohol 30*:153–161.

Kedishvili NY, Gough WH, Chernoff EAG, Hurley TD, Stone CL, Bowman KD, Popov KM, Bosron WF, Li TK. 1997. cDNA sequence and catalytic properties of a chick embryo alcohol dehydrogenase that oxidizes retinol and 3 beta,5 alpha-hydroxysteroids. *J Biol Chem 272*:7494–7500.

Kraulis P. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24:*946–950.

Kurys G, Ambroziak W, Pietruszko R. 1989. Human aldehyde dehydrogenase. Purification and characterization of a third isozyme with low K_m for gamma-aminobutyraldehyde. *J Biol Chem 264*:4715–4721.

Kutzenko AS, Lamzin VS, Popov VO. 1998. Conserved supersecondary structural motif in NAD-dependent dehydrogenases. *FEBS Lett 423*:105–109.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemistry quality of protein structures. *J Appl Crystallog 26*:283–291.

Liu Z, Sun Y, Rose J, Chung YJ, Hsiao CD, Chang WR, Kuo I, Perozich J, Lindahl R Hempel J, Wang BC. 1997. The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. *Nat Struct Biol 4*:317–326.

Merritt EA, Murphy MEP. 1994. Raster3D version 2.0—A program for photorealistic molecular graphics. *Acta Cryst D50:*869–873.

Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst D53*:240–255.

Navaza J. 1994. AMoRe: An automated package for molecular replacement. *Acta Cryst 50*:157–163.

Otwinowski Z. 1993. Data collection and processing. In: Sawyer L, Issacs N, Bailey S, eds. *Proceedings of the CCP4 study weekend*. Warrington, UK: Daresburry Laboratories. pp 56–62.

Persson B, Zigler JSJ, Jörnvall H. 1994. A super-family of medium-chain dehydrogenases/reductases (MDR). Sub-lines including zeta-crystallin, alcohol and polyol dehydrogenases, quinone oxidoreductase enoyl reductases, VAT-1 and other proteins. *Eur J Biochem 226*:15–22.

Pietruszko R, Kikonyogo A, Chern MK, Izaguirre G. 1997. Human aldehyde dehydrogenase E3. Further characterization. *Adv Exp Med Biol 414*:243–252.

Rice LM, Brünger AT. 1994. Torsion angle dynamics: Reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins Struct Funct Genet 19*:277–290.

Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol 234:*779–815.

Sheldrick GM, Schneider TR. 1997. SHELXL: High resolution refinement. *Methods Enzymol 277:*319–343.

Silman I, Harel M, Axelsen P, Raves M, Sussman JL. 1994. Three-dimensional structures of acetylcholinesterase and of its complexes with anticholinesterase agents. *Biochem Soc Trans 22:*745–749.

Steinmetz CG, Xie P, Weiner H, Hurley T. 1997. Structure of mitochondrial aldehyde dehydrogenase: The genetic component of ethanol aversion. *Structure 5*:701–711.

Stryer L. 1995. *Biochemistry*. New York: W.H. Freeman and Company.

Sussman JL, Harel M, Frolow F, Oefner C, Goldman A, Toker L, Silman I. 1991. Atomic structure of acetylcholinesterase from *Torpedo californica:* A prototypic acetylcholine-binding protein. *Science 253:*872–879.

Sussman JL, Harel M, Silman I. 1993. 3-Dimensional structure of acetylcholinesterase and of its complexes with anticholinesterase drugs. *Chem Biol Interact 87:*187–197.

Weiner H. 1979. Aldehyde dehydrogenase: Mechanism of action and possible physiological roles. In: Majchrowicz E, Noble EP, eds. *Biochemistry and pharmacology of ethanol, Vol. 1*. New York: Plenum Press. pp 107–124.

Wierenga RK, Terpstra P, Hol WG. 1986. Prediction of the occurrence of the ADP-binding beta alpha beta-fold in proteins, using an amino acid sequence fingerprint. *J Mol Biol 187:*101–107.

Yin S-J, Vagelopoulos N, Wang S-L, Jörnvall H. 1991. Structural features of stomach aldehyde dehydrogenase distinguish dimeric aldehyde dehydrogenase as a 'variable' enzyme. 'Variable' and 'constant' enzymes within the alcohol and aldehyde dehydrogenase families. *FEBS Lett 283:*85–88.

Yoshida A, Rzhetsky A, Hsu LC, Chang C. 1998. Human aldehyde dehydrogenase gene family. *Eur J Biochem 251*:549–557.