Determinants of Protein Thermostability Observed in the 1.9-Å Crystal Structure of Malate Dehydrogenase from the Thermophilic Bacterium *Thermus flavus*†

Clyde A. Kelly,‡ Makoto Nishiyama,§ Yasuo Ohnishi,§ Teruhiko Beppu,§ and Jens J. Birktoft*,‡

Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110, and Department of Agricultural Chemistry, Faculty of Agriculture, University of Tokyo, Bunkyo-ku, Tokyo, 113 Japan

Received October 27, 1992; Revised Manuscript Received January 26, 1993

ABSTRACT: A binary complex of malate dehydrogenase from the thermophilic bacterium *Thermus flavus* (tMDH) with NADH has been crystallized from poly(ethylene glycol) 3500, pH 8.5, yielding diffraction-quality crystals in space group $P2_12_12_1$. The structure was solved at 1.9-Å resolution using molecular replacement and refined to an *R* factor of 15.8% with good geometry. The primary sequence of tMDH is 55% identical to that of cytoplasmic malate dehydrogenase (cMDH) [Birktoft, J. J., Rhodes, G., & Banaszak, L. J. (1989) *Biochemistry 28*, 6065–6081], and overall their three-dimensional structures are very similar. Like cMDH, tMDH crystallized as a dimer with one coenzyme bound per subunit. The coenzyme binds in the extended conformation, and most of the interactions with enzyme are similar to those in cMDH. In tMDH, small local conformational changes are caused by the replacement of a glutamic acid for the aspartic acid involved in hydrogen bonding to the adenine ribose of NADH. Comparison of tMDH with cMDH reveals that both tMDH subunits more closely resemble the B subunit of cMDH which therefore is the more likely representative of the solution conformation. While cMDH is inactivated at temperatures above about 50 °C, tMDH is fully active at 90 °C. On the basis of the X-ray crystal structure, a number of factors have been identified which are likely to contribute to the relative thermostability of tMDH compared to cMDH. The most striking of the differences involves the introduction of four ion pairs per monomer. All of these ion pairs are solvent-accessible. Three of these ion pairs are located in the dimer interface, Glu27–Lys31, Glu57–Lys168, and Glu57–Arg229, and one ion pair, Glu275–Arg149, is at the domain interface within each subunit. Additionally, we observe incorporation of additional alanines into α-helices of tMDH and, in one instance, incorporation of an aspartate that functions as a countercharge to an α-helix dipole. The possible contributions of these and other factors to protein thermostability in tMDH are discussed.

Malate dehydrogenases (MDH)[1] are part of a well-studied family of NAD-dependent dehydrogenases (Banaszak & Bradshaw, 1975). This family includes the enzymes lactate dehydrogenase (LDH), liver alcohol dehydrogenase, and glyceraldehyde-3-phosphate dehydrogenase (Rossmann et al., 1975). Characterization of these enzymes has included isolation from a large number of sources, sequence determination, and functional characterization. In addition, the three-dimensional X-ray crystal structure of at least one form of each of these enzymes has been determined. Structural similarities include the fact that each enzyme is composed of two domains, a catalytic and an NAD-binding domain. The two most closely related enzymes in this family are the MDHs and the well-studied LDHs. Because the substrate specificities of MDH and LDH both involve converting 2-hydroxy acids to the corresponding 2-keto acids and because of extensive structural similarities, these two enzymes together form the 2-hydroxy acid dehydrogenase family (Birktoft et al., 1982). As is true with the other enzymes of the NAD-dependent dehydrogenases, the NAD-binding domains of MDH and LDH are very similar. Unlike the other enzymes, significant primary, secondary, and tertiary structural similarity is also seen in the catalytic domain (Birktoft & Banaszak, 1983; Birktoft et al., 1982).

Three-dimensional X-ray crystal structures have been reported for three MDHs. These include two eukaryotic forms of MDH from porcine heart, mitochondrial (mMDH) (Roderick & Banaszak, 1986) and cytoplasmic (cMDH) (Birktoft et al., 1989), which share about 20% sequence identity, and one prokaryotic form of MDH from *Escherichia coli* (Hall et al., 1992). The *E. coli* MDH is most similar to mMDH which has 59% sequence identity and a similar tertiary structure. These MDHs are homodimers of approximately 67 000 daltons molecular mass. Each subunit is composed of 326–333 amino acids and binds 1 coenzyme molecule.

Iijima et al. (1980) have isolated and characterized the physicochemical and catalytic properties of malate dehydrogenase from the thermophilic bacterium *Thermus flavus* AT-62. Nishiyama et al. (1986) have cloned the gene for the *Thermus flavus* malate dehydrogenase (tMDH) into *E. coli* and reported the sequence (see Figure 1). Interestingly, the prokaryotic tMDH sequence is more similar to the eukaryotic cMDH, 55% identical, than to mMDH, about 20% identical. Despite its obvious similarities to cMDH, it is distinctive in that *Thermus flavus* has an optimal growth temperature in

† C.A.K. was a fellow of the American Heart Association, Missouri affiliate. This research was supported by a grant to J.J.B. from the Lucille P. Markey Charitable Trust supporting The Markey Center for Research in Molecular Biology of Human Disease at Washington University.

* Address correspondence to this author at the Department of Biochemistry and Molecular Biophysics, Box 8231, Washington University School of Medicine, 660 S. Euclid Ave., St. Louis, MO 63110. Telephone: (314)-362-3341. FAX: (314)-362-7183.

‡ Washington University School of Medicine.

§ University of Tokyo.

[1] Abbreviations: MDH, malate dehydrogenase; tMDH, *Thermus flavus* malate dehydrogenase; cMDH, cytoplasmic malate dehydrogenase; mMDH, mitochondrial malate dehydrogenase; LDH, lactate dehydrogenase; rms, root mean square; PEG, poly(ethylene glycol); NADH, reduced nicotinamide adenine dinucleotide; NAD^+, oxidized NADH.

0006-2960/93/0432-3913$04.00/0

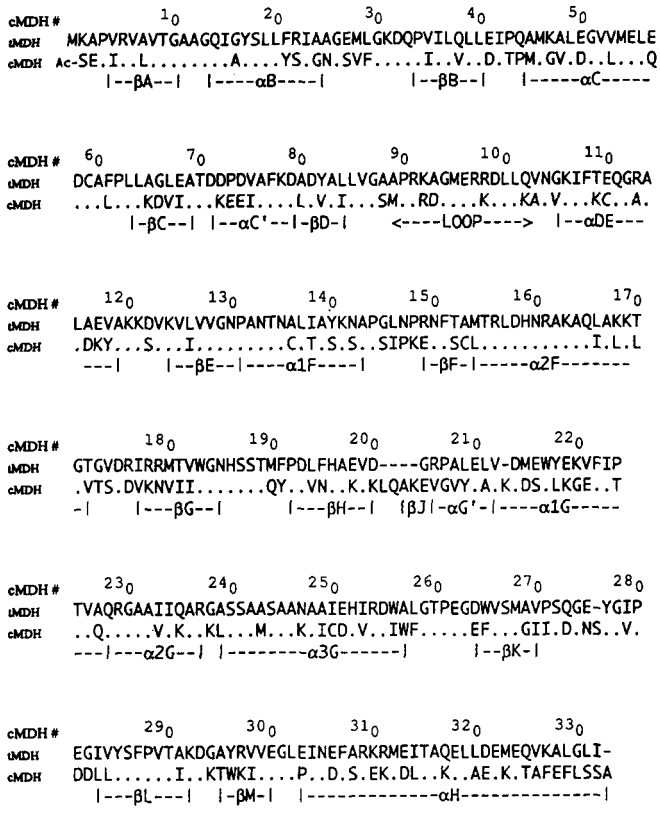

FIGURE 1: Amino acid sequence alignment of cMDH and tMDH. The indicated nomenclature for the secondary structure elements is the same as that used in describing cMDH. In the cMDH sequence line, amino acids which are identical to tMDH are indicated with a black dot. cMDH is N-acetylated at the amino terminus (indicated as Ac-) while the genetically derived tMDH contains an N-terminal methionine. These have been designated residue 0.

excess of 70 °C and tMDH is fully active after 60-min incubation at 90 °C.

The study of protein thermostability has implications both to the protein folding problem and also to the future of protein design and engineering. Historically, two paths have been followed in studying protein thermostability. The newer involves using molecular biology to empirically alter potentially significant single amino acids. The mutant protein is then characterized to evaluate the significance of the alteration. The other method involves comparative study of mesophilic and thermophilic protein variants. Both methods strive to identify a property or properties that may play the most significant role(s) in determining protein thermostability. Among the most cited works on protein thermostability are those by Menendez-Arias and Argos (1989), Zuber (Zuber, 1988; Zulli et al., 1991), Matthews [for example, see Matthews et al. (1990)], and Perutz and Raidt (1975). At some stage, it is anticipated that works such as these will be useful in engineering protein thermostability. However, despite years of active research in studying protein folding and stability, our understanding is far from complete. It is still safe to say that a reasonable goal is to identify structural determinants which strongly correlate with thermostability.

In the context of the cMDH/tMDH comparison, we have attempted to evaluate on the basis of perceived significance to thermostability and predictive power, i.e., correlation with the phenomenon of thermostability, those determinants which have been identified in the literature. Among these works, probably the most consistently stated observation is that protein thermostability tends to result from many amino acid substitutions in which each provides a small increment of thermostability. Are there certain determinants which are more significant than others? If so, are these applicable to most proteins or can they only be identified for a given protein or protein family? Questions such as these remain to be answered.

Table I: Crystallographic Statistics

Data Collection	
space group	$P2_12_12_1$
unit cell	
a (Å)	71.51
b (Å)	87.59
c (Å)	118.98
nominal resolution (Å)	1.9
total no. of measurements	376446
no. of independent reflections	63904
R_{merge} (%)	7.1
Refinement[a]	
resolution limit (Å)	1.9
R factor (%)	15.8
rms bond length deviation (Å)	0.016
rms bond angle deviation (deg)	2.90

[a] The refined structure includes 4976 protein atoms, 88 coenzyme atoms, and 223 water molecules per dimer.

We report here the solution of the three-dimensional X-ray crystal structure of tMDH complexed with NADH to 1.9-Å resolution. In the analysis of the structure, particular attention has been paid to structural comparisons with the binary cMDH–NAD complex (Birktoft et al., 1989) as well as with LDH. These structural comparisons are the basis for the identification of possible determinants of thermostability in tMDH.

MATERIALS AND METHODS

Structure Determination. The tMDH used in these studies was obtained from an *E. coli* expression system (Iijima et al., 1986). Diffraction-quality crystals were first obtained for a single-site mutant (Thr189Ile) of tMDH (Kelly et al., 1991). The molecular replacement technique was used to solve the phases for the mutant tMDH structure. This work is described elsewhere.[2]

During the initial stages of refinement of the mutant tMDH structure, crystals of the wild-type tMDH were obtained. These crystals grew under a narrow range of conditions using the sitting drop vapor diffusion method. Using a protein concentration of 13–15 mg/mL, large crystals grew using 0.1 M Tris, pH 8.5, 20–24% PEG 3350, 200–800 mM NaCl, 20 mM oxaloacetate, and 0.2–0.5 mM NADH. The crystals were determined to be isomorphous with the tMDH–T189I crystals which had unit-cell dimensions of a = 71.79 Å, b = 88.47 Å, and c = 118.89 Å. A full data set to 1.9-Å resolution was collected on a Xoung-Hamlin-type area detector (Table I). Refinement based on the partially refined Ile mutant structure was begun using the program package TNT (Tronrud et al., 1987) and brought to an R factor of 23.0% for data between 6.0- and 2.3-Å resolution, rms bond length deviation of 0.045 Å, and rms bond angle deviation of 5.0°.

Final Refinement. At this stage, X-PLOR-simulated annealing (Brunger, 1990) was used with an initial temperature of 3000 °C and a 0.0005-ns timestep. This protocol resulted in an R factor of 23.7% with an rms bond length deviation of 0.024 Å and an rms bond angle deviation of 3.9°. OMIT maps were calculated over the entire molecule in 20 amino

[2] C. A. Kelly, M. Nishiyama, T. Beppu, and J. J. Birktoft, unpublished results.

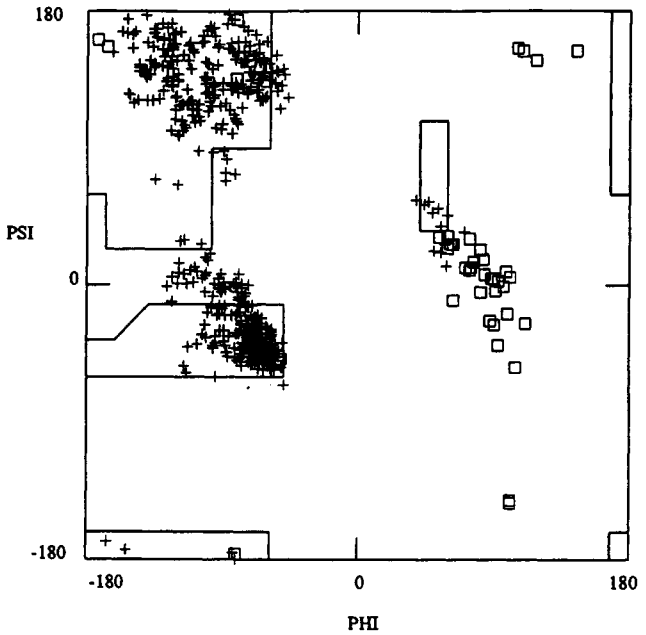

FIGURE 2: Ramanchandran diagram of the crystallographic model of tMDH. The main-chain torsional angle Phi (N–Cα bond) is plotted versus Psi (Cα–C′ bond). (+) represents non-glycine residues while (□) represents glycine residues. The plot includes both subunits of the dimer.

acid segments. In 1 cycle of this procedure, the coordinates for the 20 amino acids were removed from the model. Several rounds of positional refinement were then used to remove bias from the omitted amino acids in the calculated phases. $F_o - F_c$ OMIT maps were then calculated and contoured at 3σ for display on a Silicon Graphics IRIS graphics workstation using the TURBO-FRODO program (Rousell, 1992). The conformation of the model was verified or altered in the omitted region. The modeling of the NAD-binding loop was based upon an $F_o - F_c$ OMIT map calculated using data between 1.9 to 10 Å, leaving out residues 91–99,[3] and contoured between 2σ and 3σ. After this, a typical cycle consisted of 60 steps of positional refinement in X-PLOR followed by calculation of $F_o - F_c$ and $2F_o - F_c$ maps and identification of the peaks and pits in the $F_o - F_c$ map with magnitudes greater than 4σ. These regions were then investigated on a molecular graphics station and appropriate changes made to the model or waters added if the geometry was satisfactory. At this stage, the R factor was 18.0% with an rms bond length deviation of 0.014 Å and an rms bond angle of 2.78°. Final refinement was done using TNT.

The calculations for TNT were run on VAXstations 3200 and 4100. Most X-PLOR runs were done on a Silicon Graphics Power Series 340. Some of the smallest X-PLOR runs were done on the VAXstation 4100. Solvent accessibilities were calculated using the program LESK-ACCESS (Lesk, 1985).

RESULTS

Refinement. As tabulated in Table I, the final R factor for tMDH is 15.8% for data between 6.0 and 1.9 Å with rms bond length deviation of 0.016 Å and rms bond angle deviation of 2.90°. The quality of the refined structure is reflected in the Ramanchandran plot shown in Figure 2. Values from both

[3] The numbering system used is that based upon the cMDH structure (Birktoft et al., 1989). Where relevant, a suffix is added to the residue numbers to distinguish between subunits A and B.

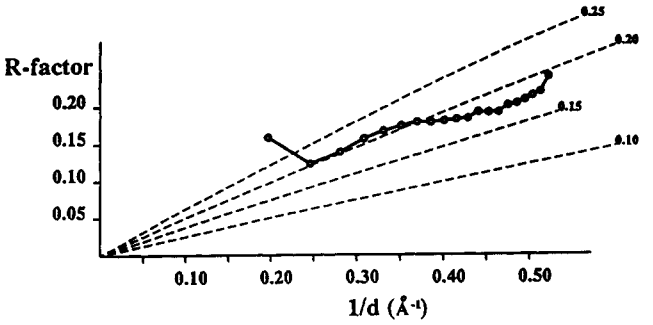

FIGURE 3: R factor plotted as a function of resolution. The theoretical estimate of the mean positional error in the atomic coordinates according to Luzatti (1952) superimposed upon the curve for the tMDH diffraction data.

subunits are included in the figure. It should be noted that no non-glycine residues are located outside the allowed regions. The plot obtained by the method of Luzatti (1952) indicates a 0.18–0.2 Å upper limit for error in atomic coordinates (Figure 3).

Electron Density Quality. Overall the quality of the electron density maps is very good. A sample region of $2F_o - F_c$ electron density is shown in Figure 4. Contouring of $2F_o - F_c$ maps at 1.5σ yields solid, continuous density throughout most of the protein. The most significant exceptions are summarized here. The electron density for the nicotinamide ring of the NADH is somewhat weak in both subunits. Importantly, although no NADH was included in the starting models used for refinement of tMDH, positive electron density was clearly present in the electron density maps at all stages. With the exception of the nicotinamide ring, the geometry and interactions of NADH with tMDH are clearly identifiable.

The electron density of the "NAD binding loop" between βD and αDE, residues 91–100, is very weak in both subunits of tMDH. It should be noted that this region in cMDH is disordered (Birktoft et al., 1989) and in LDH undergoes a significant conformational change upon substrate binding (Holbrook et al., 1975) which will be discussed later in the context of tMDH. While the wild type was crystallized in the presence of the substrate oxaloacetate, it could not be identified in the substrate-binding region of the electron density maps. Other regions of weak electron density include some of the solvent-exposed amino acid side chains such as Lys121, Arg178, and Arg297.

The tMDH expressed in *E. coli* contained an *N*-formylmethionine at the N-terminus. This residue is clearly recognizable in both tMDH subunits.

Structural Comparison of tMDH A and B Subunits. Within the tMDH dimer, the two subunits of tMDH are related by a molecular 2-fold axis. The A and B subunits of tMDH are very similar with an rms deviation of 0.22 Å for all Cα atoms based on superposition of all backbone atoms. Throughout most of the structure, the differences between the subunits are within the error limit of the coordinates as determined by the Luzatti plot. Differences which exceed this error limit are all located in loop regions on the protein surface. Further structural comparisons to other systems will be made using the B subunit of tMDH. Results which involve a significant deviation of the A and B subunits of tMDH will be presented.

Structural Comparison of tMDH with cMDH. (A) Monomer. The primary, secondary, and tertiary structure of tMDH is very similar to that of cMDH. Since the binary cMDH–NAD complex structure has already been reported

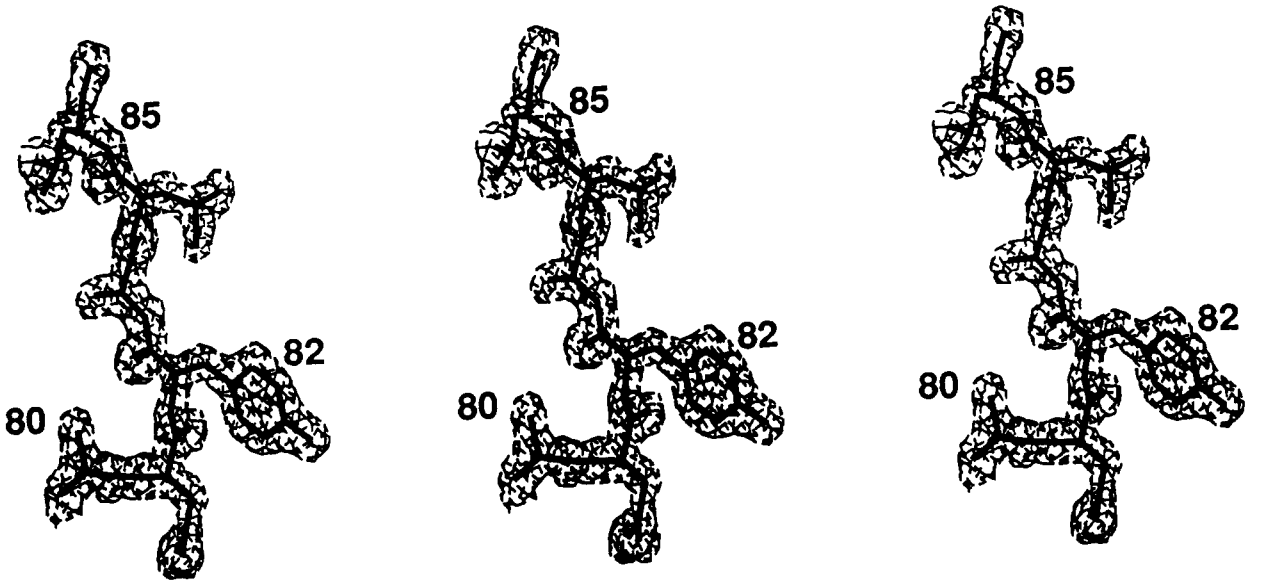

FIGURE 4: Sample of the $2F_o - F_c$ electron density map for the fully refined tMDH structure at 1.9-Å resolution. The map was contoured at 2.0σ. Stereo diagrams in this paper are presented for both cross-eye (left pair) and parallel (right pair) viewing.

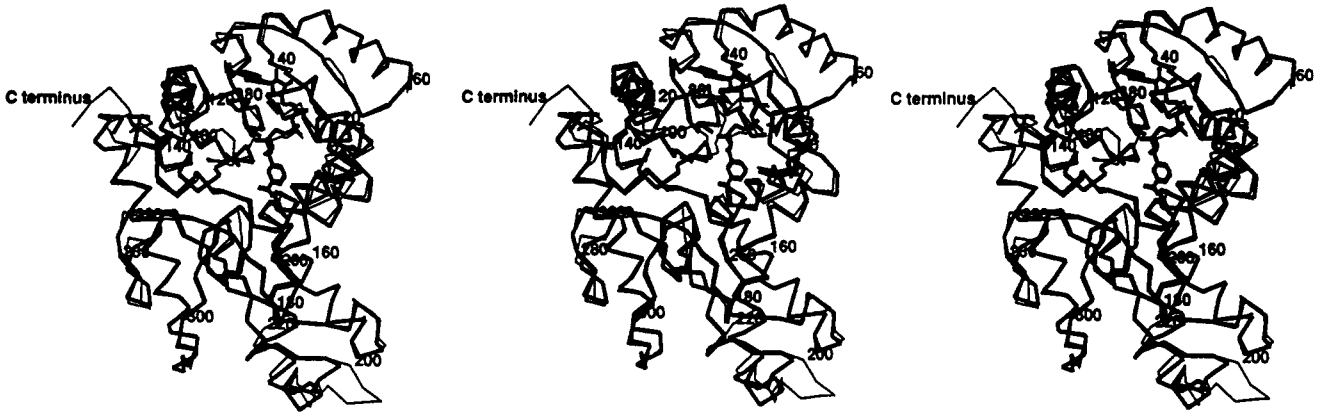

FIGURE 5: Superposition of the B subunit of tMDH with that of cMDH. The stereodiagram shows Cα models of cMDH (thinner lines) superimposed upon tMDH (thicker lines). Their superposition was done based upon all Cα atoms except those in the NAD-binding loops and the five C-terminal residues. The relative positions of the coenzyme are also included.

(Birktoft et al., 1989), the tMDH structure is reported here in the context of comparison/contrast with the cMDH structure.

The secondary structure elements within each monomer are all maintained with only the precise starting and ending residues differing in the two MDH forms. The four deletion regions in tMDH, residues 201–204, 213, 275, and 333, are all located in surface loops of cMDH (cf. Figure 1).

A comparison of the tertiary structure of one subunit of tMDH with one of cMDH shows a high degree of similarity (Figure 5). The tMDH subunit is more similar to the B subunit of cMDH as evidenced both by the rms deviation for all α-carbons based on superposition of all backbone atoms, 1.78 Å for the cMDH B subunit and 2.21 Å for the cMDH A subunit, and by a visual inspection of the active-site region. Excluding residues 91–100 from the comparison, which represent the NAD-binding loop of each subunit, significantly improves the fit, yielding rms deviations of 1.07 and 1.42 Å, respectively. On the basis of these comparisons, it appears that the effects of the crystal packing constraints identified in the cMDH structure (Birktoft et al., 1989) are localized to the A subunit. Therefore, all structural comparisons which follow refer to the B subunit of cMDH.

As shown in Figures 5 and 6, the largest displacements in tertiary structure involve solvent-exposed surface regions including the NAD-binding loop and the carboxy-terminal five residues which fold back onto the protein in tMDH. The electron density is very clear for the carboxy-terminal residues and less clear for the NAD-binding loop.

The conformation of the NAD-binding loop of tMDH was modeled without comparison to any other structure. The resulting three-dimensional conformation of the loop itself is very similar to that seen in the dogfish M-4 apo-LDH (Abad-Zapatero et al., 1987).

(*B*) *NADH Binding in tMDH.* NADH binds to tMDH in the same extended conformation seen for other members of this enzyme family (Webb et al., 1973; Rossmann et al., 1975). The torsion angles as defined in Eklund et al. (1984) are in good agreement with those measured from other structures of this family. The adenine ring binds in a hydrophobic crevice defined by Leu40, Ile42, Ile107, Thr9, Gly10, Gly87, and the backbone of Gly87 and Ala88. These side chains are all conserved between cMDH and tMDH as is the backbone conformation at residues 87 and 88. Several weak hydrogen bonds are possible between the protein and the adenine ring nitrogen atoms. Gln111, which is out of hydrogen-bonding range in cMDH, may bond weakly through Oε1 to N1A (3.72 Å) and to N10A (3.74 Å). Glu41 which is Asp41 in cMDH can hydrogen-bond through Oε1 to N3A (3.46 Å).

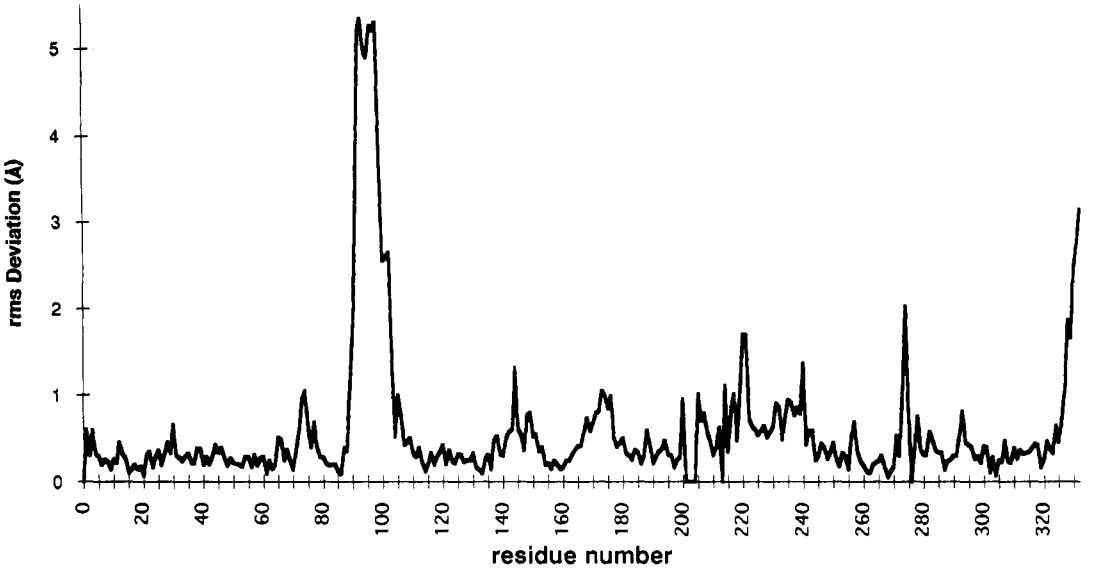

FIGURE 6: Plot of the root mean square deviation in the Cα position between the B subunits of cMDH and tMDH. The transformation matrix was calculated in X-PLOR based upon all non-hydrogen backbone atoms.

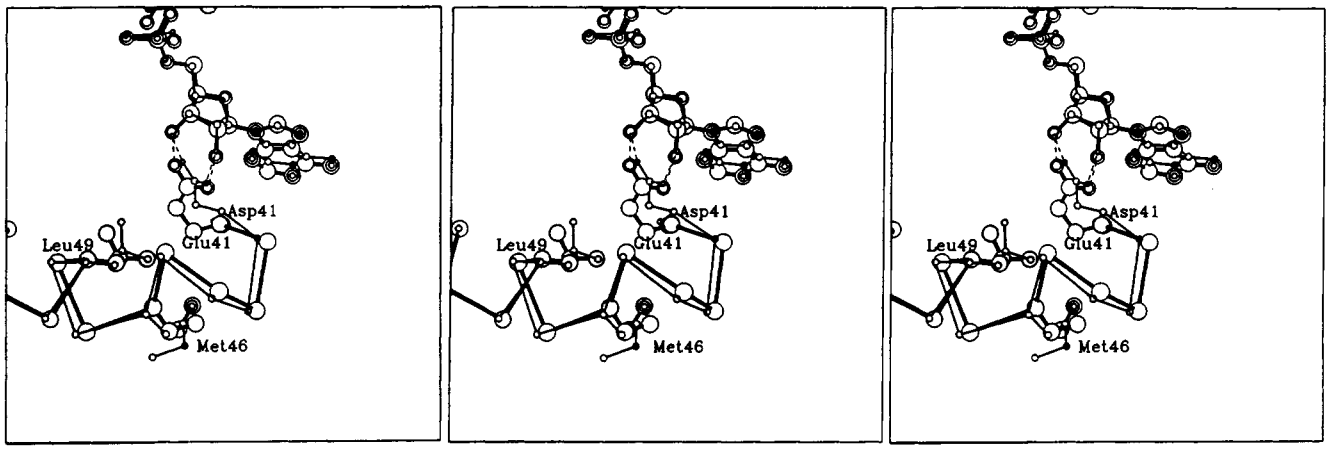

FIGURE 7: Comparison of tMDH with cMDH in the region of the coenzyme-binding residue 41. tMDH is shown as the larger ball and stick model while cMDH is the smaller. Cα models are shown between residues 40 and 52 along with the non-main-chain atoms described in the text. Hydrogen bonding between residue 41 and the adenine ribose of the coenzyme is shown as dashed lines. Carbon atoms are drawn as single circles, oxygen atoms are drawn as two concentric circles, and nitrogen and sulfur atoms are drawn as three concentric circles. Figures 7–10 were prepared using the program ARPLOT (Lesk & Hardman, 1982, 1985).

tMDH is unique among the NAD-dependent dehydrogenases in that residue 41 is a glutamic acid instead of the otherwise conserved aspartic acid which hydrogen-bonds to the adenine ribose (Rossmann et al., 1975; Birktoft & Banaszak, 1984). In a comparison with cMDH, shown in Figure 7, the strong hydrogen bonding to the adenine ribose is maintained with hydrogen-bonding distances between Oε1 of Glu41 to 02R of 2.57 Å and between Oε2 of Glu41 to O3R of 2.39 Å compared to 2.62 and 2.67 Å in cMDH. On the other side, Cβ of Glu41 moves closer to the αC helix, residues 45–61. The side chain of residue 41 contacts the side chains of Met46 and Leu49 which are present in both cMDH and tMDH. In tMDH, the side-chain crowding on this side is compounded by Met46 which is more solvent-exposed in cMDH and assumes a conformation in tMDH in which Sγ moves into the protein to crowd the Glu41 side chain. This movement is reflected in the decrease in solvent accessibility of Met46 from 53.6 Å^2 in the cMDH B subunit to 44.4 Å^2 in the tMDH B subunit. This is accompanied by a small change in χ_1 of Leu49 from −154° in cMDH to −176° in tMDH. Cα of residue 41 of tMDH is further from the coenzyme than in cMDH; the nearest atoms of the coenzyme are N3A and C2A of the adenine ring which are both 3.83 Å distant in tMDH and 3.57 Å distant in cMDH. Counterintuitively, the loop defined by the peptide backbone between residues 41 and 49 contracts in tMDH rather than expanding to accommodate the side-chain crowding in this area. The distances between the Cα atom of residue 41 and those of residues 46 and 49, located on the αC helix, decrease from 7.02 and 7.50 Å, respectively, in cMDH to 6.06 and 7.10 Å in tMDH.

The NADH pyrophosphate is hydrogen-bonded to the backbone of residues 14 and 15 and to the side chain of residue Gln14. These interactions are virtually identical to those found in cMDH which also contains Gln14.

The nicotinamide ribose is hydrogen-bonded to the side chain and backbone of residue Asn130 through both the 2′- and 3′-hydroxyls. These interactions are also present in cMDH, but the hydrogen bonds are shorter in tMDH.

In cMDH, the nicotinamide ring sits in a pocket composed on the protein side (the B side) of the side chains of residues Ile15, Val86, Val128, Leu154, Leu157, and Ala245. These residues are all conserved in tMDH with the exception of Leu154, which is methionine in tMDH. The presence of this methionine is accompanied by an Ile249Ala substitution which serves to provide space for the bulkier methionine side chain.

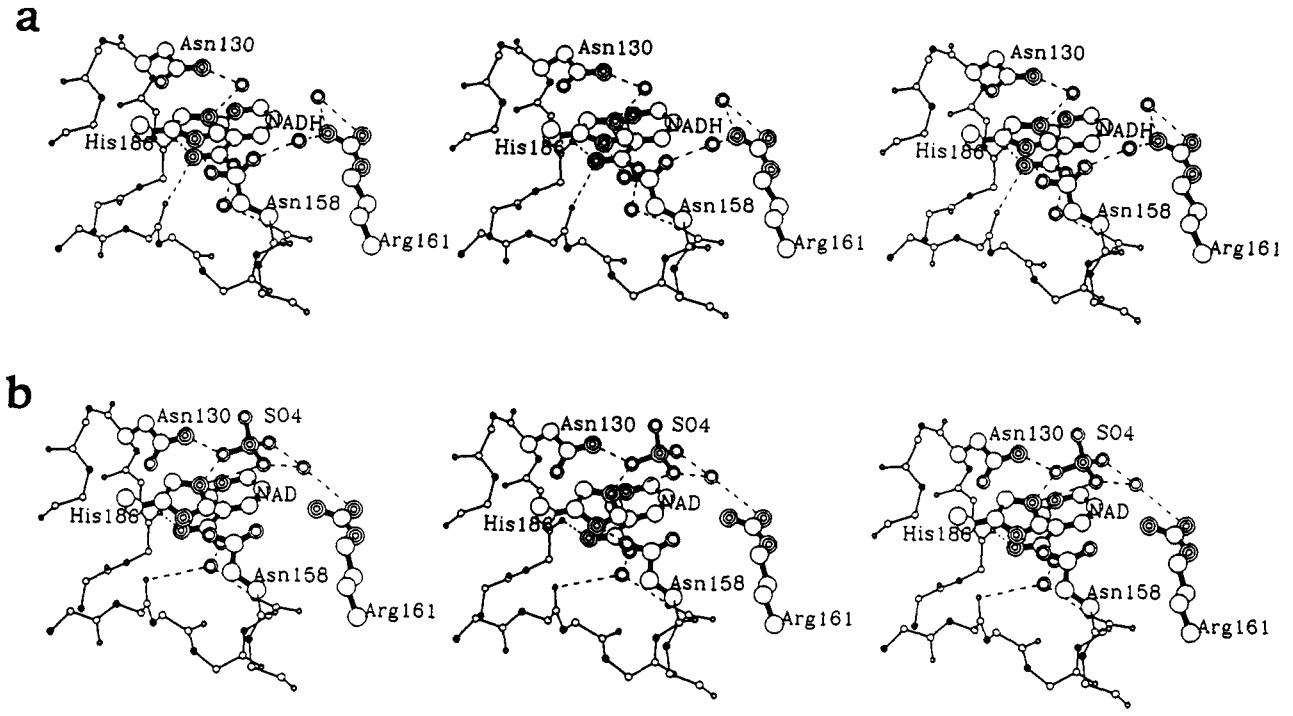

FIGURE 8: Active-site regions of tMDH and cMDH. (a) tMDH B subunit active site. The smaller ball and stick portions represent the conformation of the peptide main chain between residues 127–132 and 153–159. The larger ball and stick portions represent the nicotinamide ring of the coenzyme as well as the waters and side chains described in the text. Hydrogen bonding is indicated by dashed lines. The current model for the nicotinamide portion of NADH is shown. (b) cMDH B subunit active site. The representation is as described for tMDH with the addition of the sulfate ion in the large ball and stick portion.

While the overall electron density of the NADH is strong, the density of the nicotinamide ring is rather weak. Continuous density around the ring can only be obtained when contouring at the 0.7σ level. Comparison of the nicotinamide ring of the A and B subunits revealed that the carboxamide group was in strong density in both subunits and that the pattern of density around the ring differed. In the A subunit, the weak density occurred near C2, and in the B subunit, it occurred near C5–C6. Assignment of the ring as dihydronicotinamide in the A subunit currently appears conclusive. However, unexplained electron density is seen near the C5 nicotinamide atom in the B subunit, which may represent an alternate binding mode of the substrate. The location precludes it from being oxaloacetate bound in the substrate-binding site identified by homology with LDH (Holbrook et al., 1975) and with cMDH (Birktoft et al., 1989). In both subunits, the carboxamide nitrogen, N2N, is hydrogen-bonded to the peptide carbonyl oxygens of Met154 and Val128. The carboxamide oxygen, O1N, has a water-mediated hydrogen bond to the backbone nitrogen of Leu157.

The nucleotide-binding loop of tMDH, approximately residues 91–100, was modeled on the basis of the relatively weak electron density. However, the ends of the loop, residues 87–90 and 99–101, are clearly traceable and differ significantly in conformation from cMDH. Residues 87–90 of cMDH lie much closer to the coenzyme than do residues 87–90 of tMDH. In cMDH, the backbone nitrogen of Pro90 is 3.98 Å from the C3 atom of the nicotinamide ribose whereas in tMDH this distance is 7.18 Å.

In the cMDH–NAD complex, the main-chain carbonyl oxygen of residue 89 of the NAD-binding loop is attached to the main-chain carbonyl oxygen of residue Asn130 through hydrogen bonds mediated by the side chain of Asn104. This Asn104 has moved out of hydrogen-bonding distance in tMDH, but the connection between the peptide backbones is maintained by the inclusion of a water molecule to mediate the hydrogen bond.

(C) Active-Site Region. As shown in Figure 8, the important active-site residues, Asn130, His186, Asp158, and Arg161, are virtually superimposable with those of the B subunit of cMDH. The following comparisons refer to the B subunits both of cMDH and of tMDH after superposition of the two subunits based on all non-hydrogen backbone atoms. While cMDH contains a sulfate ion in the substrate-binding site, tMDH contains a water molecule at the position of the cMDH sulfate oxygen atom (O1) nearest to the Nϵ2 atom of His186. The His186-Nϵ2 to sulfate-O1 distance of cMDH is 2.71 Å while the His186-Nϵ2 to water oxygen distance of tMDH is 2.98 Å. This water oxygen also has a 2.51-Å hydrogen bond to Asn130-Nδ2 while the cMDH sulfate-O1 to Asn130-Nδ2 distance is 2.40 Å. The imidazole ring of His186 of tMDH is twisted by approximately −31° about the Cβ–Cγ bond, compared with cMDH, to be more nearly parallel to the plane of the nicotinamide ring. The Asp158–His186 charge relayed system is intact with an Asp158-Oδ2 to His186-Nδ1 distance of 2.80 Å. The corresponding distance in cMDH is 2.65 Å. Oδ1 of Asp158 is linked via a water-mediated hydrogen bond to Arg161 and also to the backbone carbonyl of residue 262. cMDH does not contain a water molecule at this position. The substrate-binding residue, Arg161, is hydrogen-bonded through its NH1 and NH2 atoms to a water molecule which is also present in cMDH.

tMDH/cMDH: Differences Affecting Thermostability. (A) Subunit–Subunit Interface. The subunit–subunit interface involves two sets of interactions of α-helices per monomer, αB of one subunit with αB of the other, and αC of one subunit with helices α2F, α2G, and α3G of the other. In tMDH, 40 residues per monomer are involved in the subunit–subunit interaction. Of these 40 interactions, 25 are conserved between cMDH and tMDH. The most striking observation among

Table II: Hydrogen Bond Distances between Subunits A and B in the tMDH Dimer[a]

	subunit A to B	subunit B to A
OH Tyr17···N Ala243	2.9	2.9
OH Tyr17···N Ala242	3.3	3.3
OH Tyr17···NH2 Arg237	3.2	3.4
OE1 Glu27···Nz Lys31[b]	2.9	2.9
O Met54···NH1 Arg229	3.5	3.3
OE2 Glu55···N Ser244	3.0	3.0
OE2 Glu55···OG Ser244	2.8	2.7
OE1 Glu57···NH1 Arg229[b]	2.9	3.1
OE1 Glu57···NZ Lys168[b]	3.2	2.6
OD1 Asp58···NH1 Arg229	3.1	3.0
OD2 Asp58···NE Arg229	2.9	2.9
OD2 Asp58···NE Arg161	3.6	3.7
O Cys59···ND2 Asn247	3.4	3.4

[a] Distances in Ångstroms. [b] Indicates an interaction at the subunit interface which is not common to the cMDH dimer.

those that change is that while the net charge remains the same the number of charged residues increases from 3 in cMDH to 7 in tMDH. The four additional charges per monomer in tMDH represent two unique sets of ion pairs, Glu27–Lys31 and Glu57–Lys168, which form between the subunits. One additional ion pair is revealed when hydrogen bonding between the subunits is tabulated as in Table II. A comparison with the intersubunit hydrogen bonding in cMDH [see Table II of Birktoft et al. (1989)] reveals that most of these contacts are conserved between cMDH and tMDH. The exceptions represent the two unique sets of ion pairs mentioned above as well as one in which a hydrogen-bonding pair, Gln57–Arg229, obtains some ionic character, Glu57–Arg229, in tMDH.

The intersubunit ionic interactions which are unique to tMDH are presented in Figure 9. These divide roughly into two ionic clusters. The cluster of hydrogen bonds and ionic interactions, seen in the upper part of Figure 9, involves substitution of Leu168 on α2F to lysine and of Gln57 on αC of the opposite subunit to glutamic acid. The resulting Lys168-Nζ to Glu57-Oϵ1 distance is 2.4 Å. In addition, Oϵ1 of Gln57 in cMDH lies within hydrogen-bonding distance of Arg229 of the opposite subunit. This Arg229 is retained in tMDH, and its interaction with Glu57-Oϵ2 (3.0 Å) must also be partly ionic in nature. The substitution of Leu168 to lysine is further favored in that the hydrophobic Leu168 of cMDH is largely solvent-exposed and resides in a polar environment near Gln57 and -165.

The other cluster, seen in the lower part of Figure 9, is centered around two substitutions on the αB helix, Ser22 of cMDH to arginine and Ser27 on the same αB helix to glutamic acid. The cluster involves ionic interactions of Glu27 with Arg22 on the same subunit and of Glu27 with Lys31 on the opposite subunit. In one subunit, residues Arg22, Glu27, and Lys31 lie in a line along the interface side of αB and of the N-terminus of βB. Across the subunit interface, Arg22A sits opposite Arg22B, Glu27A and Glu27B face one another with a Cγ to Cγ distance of 4.70 Å, and Lys31A and Lys31B are aligned with an Nζ to Nζ distance of 4.41 Å. Both Glu27A and Glu27B have dual conformations. In one arrangement, Glu27A is interacting with Arg22A, while Glu27B is interacting with both Lys31A and Lys31B. In the other arrangement, Glu27A breaks away from Arg22A to interact with Lys31A and Lys31B while Glu27B changes conformation to interact with Arg22B. An alternate arrangement in which both Glu27A and Glu27B interact simultaneously with the arginines or with the Lys31's is excluded on the basis of close contacts between the glutamic acid side chains. Attempts to model in waters rather than the multiple conformations of Glu27 consistently gave less satisfactory electron density maps. In cMDH, Lys31 is present but is pointing out into the solvent. Residues Ser22 and Ser27 of cMDH both reside in the solvent-filled interface between the subunits. They are hydrogen-bonded to solvent except for one bond between Ser27-Oγ and Asn25-Oδ1 of the same subunit (3.19 Å).

(B) Domain–Domain Interactions. A significant domain–domain interaction which is present in tMDH but not in cMDH involves substitution of Glu149 in cMDH to an arginine at the N-terminus of βF, the last strand of the nucleotide-binding domain, and is shown in Figure 10. This substitution is matched to an Asn275 to glutamic acid substitution in the catalytic domain. While position 149 is on the surface of the protein and Glu149 of cMDH is solvent-exposed, the side chain of Arg149 of tMDH is folded into a surface crevice in the catalytic domain with NH1 forming a strong ion pair with the Oϵ1 atom of Glu275 (2.76 Å). Further, Arg149 is fully hydrogen-bonded within the catalytic domain, forming direct hydrogen bonds to the peptide carbonyl oxygen of Pro271 and water-mediated hydrogen bonds to the main chain of residues Tyr296 and Val298. In cMDH, Asn275 points away from residue 149. The formation of the ion pair in tMDH is made possible by the deletion of residue 276. This causes a conformational change in this loop region in tMDH compared to cMDH which points Glu275 toward the nucleotide-binding domain.

DISCUSSION

tMDH/cMDH: Structural Comparison. (A) Dimer and Monomer. In the crystal structure of cMDH, it was found that the subunits were related by a 172° rotation (Birktoft et al., 1989). The deviation from an exact 2-fold rotation axis was attributed to crystal packing constraints and, as a result, there were found to be structural differences between the A and B subunits in cMDH. This nonequivalence was especially significant in the active-site region. Because of these structural modifications caused by crystal packing constraints, the analysis of the cMDH structure left in doubt whether it was the A or B subunit which better represented the solution conformation. The tMDH monomers are related by a 180° rotation. No such packing constraints and effects are evidenced in the tMDH structure. Comparison of cMDH with tMDH reveals that both subunits of tMDH are more similar to the B subunit of cMDH. This is therefore the more likely representation of the solution conformation. It is the A subunit of cMDH that is more affected by crystal packing.

(B) Active-Site Region. In the tMDH substrate-binding site identified by homology with LDH (Holbrook et al., 1975) and with cMDH (Birktoft et al., 1989), no substrate is seen in tMDH even though the crystallization was carried out in the presence of 20 mM oxaloacetate which is well in excess of the 3 μM K_m for oxaloacetate (Nishiyama et al., 1986). This could well be due to packing in the tMDH crystals. Within the 2-hydroxy acid dehydrogenase family, binding of substrate is accompanied by a large conformational change of the NAD-binding loop (Holbrook et al., 1975). In order to test whether this movement is possible within the tMDH lattice, two LDH structures were used as models, M-4 apo-LDH (dogfish) (Abad-Zapatero et al., 1987) and the ternary complex of M-4 LDH (dogfish) with NAD and pyruvate (White et al., 1976). The loop of apo-LDH is in the "open" conformation while that of the ternary complex M-4 LDH is in the "closed" conformation. The closing movement is very pronounced, resulting, for example, in a shift of the Cα atom of Gly106

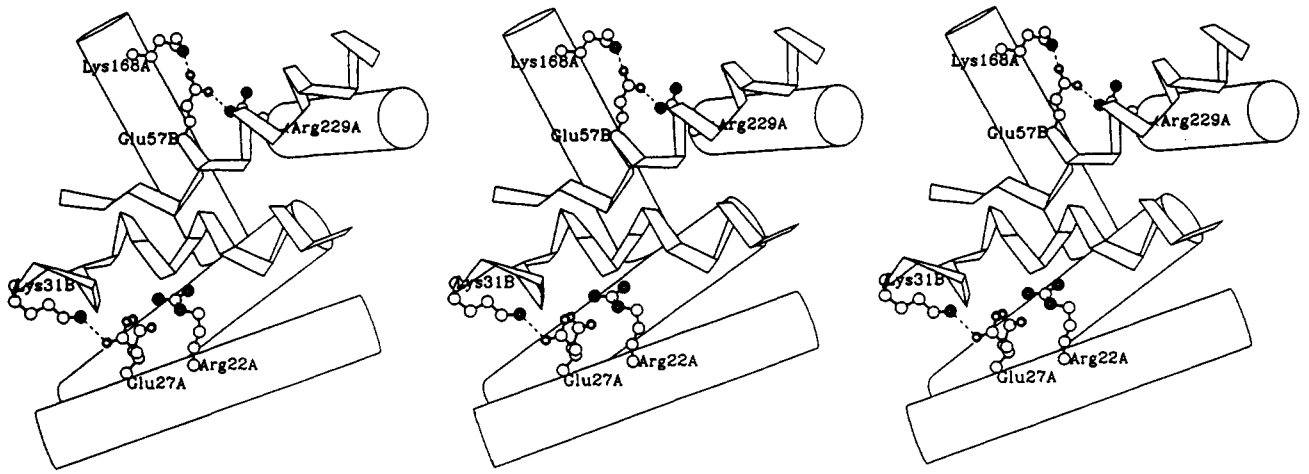

FIGURE 9: Subunit interface region of the tMDH dimer. The interface is composed of six α-helices per monomer. Four of these are from one subunit (represented as cylinders) while two are from the other subunit (represented as ribbons). The side chains involved in the three ionic interations at the subunit interface are shown as ball and stick models. The ionic interactions are indicated as dashed lines. The dual conformations of Glu27 are shown.

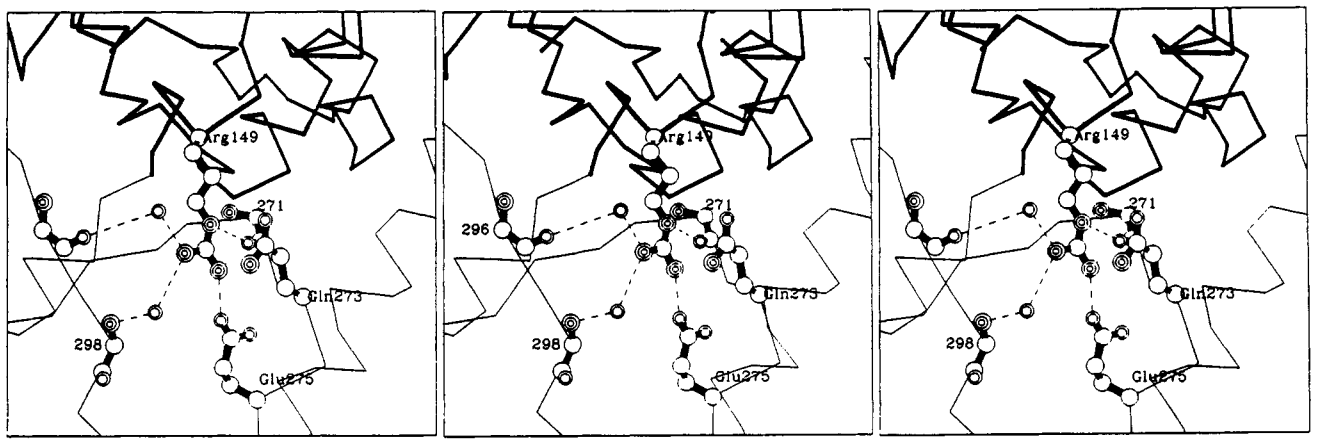

FIGURE 10: Unique ionic interaction in the domain interface region of the tMDH monomer. The NAD-binding domain is the darker stick model in the upper portion of the figure while the catalytic domain is the lighter stick model in the lower portion of the figure. Arg149 of the NAD-binding domain is shown in the ball and stick model interacting with residues of the catalytic domain. Only the main-chain atoms are shown for those residues whose side chains are not involved in interacting with Arg149. Hydrogen bonds are represented as dashed lines.

by 12.0 Å. Superimposing the apo-LDH structure on that of tMDH reveals that the NAD-binding loop of tMDH is even more open than that of apo-LDH. Closing of the loop to the position seen in the LDH ternary complex would be accompanied by a movement of 14 Å at the tMDH Cα atom of Gly94 (equivalent to Gly106 in the LDH ternary complex). However, in the tMDH crystal lattice, this movement appears to be prohibited by close contacts which would arise between residues 92–96 of the NAD-binding loop and helix α1G (residues 216–221) of a symmetry-related molecule in the crystal. Thus, substrate binding with its accompanying loop closure and crystal packing as seen in this tMDH form seem to be mutually exclusive. The substrate-bound form of tMDH would have to pack in a different crystal lattice.

The position of the NAD-binding loop relative to the body of the protein in tMDH was compared to that in cMDH and various LDH structures based upon superposition of Cα's excluding the NAD-binding loop. The relative position in the tMDH–NADH binary complex is more similar to that in apo-LDH than to that in the cMDH–NAD binary complex. In the cMDH–NAD complex, this loop lies in a partially closed position between that of the tMDH structure and the fully closed position seen in the ternary LDH complex. While the cMDH–NAD complex does not have substrate bound, a sulfate ion was found to occupy the substrate-binding site (Birktoft et al., 1989). This ion mimics some but not all of the substrate interactions which correspond to the loop closure event identified by Holbrook et al. (1975).

tMDH/cMDH: Differences Affecting Thermostability. (A) Evaluation of Proposed Determinants. On the basis of the three-dimensional structures of ferrodoxin and of hemoglobin and from available sequences from thermophiles and mesophiles, Perutz and Raidt (1975) identified surface ion pairs as significant determinants of thermostability. Walker et al. (1980) have also identified as important thermostabilizing elements additional ion pairs found in the three-dimensional structure of the thermostable *Bacillus stearothermophilus* D-glyceraldehyde-3-phosphate dehydrogenase. These results are in good agreement with the identification of four unique, solvent-accessible ion pairs at structural interface regions in tMDH. On the basis of studies of reversible denaturation of tMDH by guanidine hydrochloride and by acid, Iijima et al. (1984) propose that strong subunit interactions make a large contribution to the thermostability of tMDH compared to mesophilic counterparts. The contributions of these ion pair interactions at structural interface regions to thermostability in tMDH could be evaluated through a site-directed mutagenesis study. Unfortunately, if it were to turn out that introduction of surface ion pairs was a general method of protein thermostabilization, the task of engineering these

changes would remain nontrivial. In the case of tMDH, at least two of the four ion pair interactions would not have been predicted on the basis of modeling. In the Arg22–Glu27–Lys31 interaction, the Glu27 might have been modeled interacting with Arg22. However, Lys31 of cMDH is pointing in the opposite direction, into the solvent. There would seem to be no compelling reason to flip it toward the opposite subunit and to allow Glu27 to take up two conformations. Similarly, in modeling the Arg149–Glu275 interaction at the domain–domain interface, Glu149 of cMDH is solvent-exposed and pointing away from the catalytic domain. Further, there is a deletion in tMDH at position 274 that makes Glu275 available for interaction with Arg149. It seems unlikely that the results of all of these changes could have been successfully predicted a priori or, more importantly, that these types of interactions could be rationally designed into another protein.

On the basis of primary sequence comparisons of LDHs from thermophiles and from mesophiles, Zuber (1988) states that amino acids which are found more often in thermophiles are "preferred" and that these are primarily responsible for specific characteristics and properties of the temperature variants. For example, in substituting one hydrophobic amino acid for another hydrophobic amino acid, "we find a clear splitting up between preferentially thermophilic Phe, Val, and Ile residues and preferentially mesophilic or psychrophilic hydrophobic residues Leu, Ala, and Met". According to this study, polar amino acids in mesophilic enzymes are exchanged for hydrophobic and charged residues in thermophilic enzymes. Within the cMDH/tMDH comparison, it is possible to test the underlying premise that the ratio of thermostabilizing to thermolabilizing amino acids is correlated with thermostability. There are 146 amino acid substitutions per monomer between cMDH and tMDH. Unfortunately, of these substitutions, only 79 are groupable, and 72 are scalable by the available data (Zuber, 1988). The ratio of thermostabilizing to thermolabilizing amino acids in cMDH is found to be 1.0 while it is only 0.8 in tMDH. Thus, in this instance, the available criteria would predict greater thermostability in cMDH than in tMDH. On the basis of the poor prediction of thermostability in tMDH compared to cMDH from the available data, it is hard to justify use of this method as it stands in evaluating thermostability.

In a related study, Zulli et al. (1991) suggest that there appear to be regions of the three-dimensional structure in which these characteristic amino acid substitutions may cluster and which may be of particular importance in temperature adaptation. In this study, the αB region of LDH was identified as an important region based upon enhanced accumulation of amino acid exchanges between mesophiles and thermophiles. The thermostability of a mesophilic LDH was successfully enhanced by making amino acid substitutions at positions 29 and/or 39 in the αB region (LDH numbering). Interestingly, position 39 corresponds to the location of Arg22 in tMDH, i.e., a participant in one of the unique ion clusters seen in tMDH. While the αB region was found to be important in thermostabilization both in the present study and also in the work of Zuber, Zulli et al. (1991) point out that these important regions are likely to vary among different enzymes.

On the basis of a statistical survey of mesophilic and thermophilic enzymes of various families, Menendez-Arias and Argos (1989) report that there are two regions, α-helices and to a lesser extent domain interfaces, at which two properties, decreased flexibility and increased hydrophobicity, were well correlated with temperature adaptation. In addition, they ranked the 10 most often observed amino acid substitutions in comparing mesophilic with thermophilic enzymes. Using the same statistical methods to evaluate cMDH and tMDH, it was found that the top amino acid replacement reported by Argos [Table 3 of Menendez-Arias and Argos (1989)] was also the top replacement between cMDH and tMDH, i.e., lysine in the mesophile to arginine in the thermophile. Four other substitutions which Argos ranked in the top 10 were also identified from the cMDH/tMDH comparison, i.e., Ser to Ala, Ile to Val, Lys to Ala, and Lys to Glu. However, these were all at the rank of the fifth most observed cMDH/tMDH substitution (note the marked clustering of the cMDH/tMDH data due to the small data set size). Argos points out that the top amino acid replacements differ for different proteins (perhaps reflecting the fact that different proteins might use different strategies to achieve thermostability). Using the same methods employed by Argos, the differences in hydrophobicity (Rose et al., 1985) and in flexibility (Karplus & Schulz, 1985) were calculated between cMDH and tMDH in the α-helical and the domain interface regions. In the α-helical regions, there are 62 substitutions per monomer. In these regions, a 2% increase in hydrophobicity was calculated along with a 0.7% decrease in flexibility. Likewise, in the domain interface, there were 20 amino acid changes in the NAD-binding domain and 27 amino acid changes in the catalytic domain. Here a 2% increase in hydrophobicity and a 1% decrease in flexibility were found. The hydrophobicity over the entire subunit was also compared for cMDH and tMDH. There is a 0.4% increase in hydrophobicity in tMDH. The results for the two regions are in agreement that there is an apparent correlation in the α-helical and domain interface regions between thermostability and decreased flexibility and increased hydrophobicity. The correlation appears stronger for these regions than for the entire molecule. However, the magnitude of the effect measured in the cMDH/tMDH pair seems very small. It appears that for this pair, Argos' criteria are indeed predictive of thermostability but, on the basis of the small size of the effects, the source of the dramatic increase in thermostability of tMDH would appear to lie elsewhere.

Using the molecular biology approach, Matthews and coworkers [for example, see Matthews (1990)] have chosen to systematically isolate one property at one location that may affect thermostability and to modify that property through site-directed mutagenesis. Rather than ranking the significance of properties, they have identified those which do or do not appear to play roles in protein thermostabilization for T4 lysozyme. Property changes such as stabilization of helix dipole moments by introduction of appropriately charged residues (Nicholson et al., 1991) and addition of disulfide bridges (given appropriate constraints) have been shown to be significant in these studies. Several of these properties were investigated in the cMDH/tMDH system. A search for helix dipole stabilization yielded one substitution, Lys72Asp, which is well situated in tMDH to stabilize the small helix α-C′ which links the first Rossman fold (nucleotide-binding domain) with the second. Interestingly, no other member of the 2-hydroxyacid dehydrogenase family contains a negatively charged amino acid at this position. Additionally, much attention has been given by Matthews (Zhang et al., 1991) as well as by others [for example, see Menendez-Arias and Argos (1989)] to the stabilization of α-helices by introduction of alanines, particularly at the nonburied positions, in order to enhance protein thermostability. There are 19 more alanines in tMDH than in cMDH, which is the sum of 24 unique alanines in tMDH minus 5 unique alanines in cMDH (i.e.,

reverse substitutions). Of the 24 unique tMDH alanines, 13 are in α-helices, and all of these either are solvent-accessible (12 of 13) or are next to a solvent-accessible residue (1 of 13). These are likely to cumulatively have a thermostabilizing effect. Of the five reverse substitions, four are in α-helices, and these either are accessible (three of four) or are next to an accessible residue (one of four). To a first approximation, these substitutions should be destabilizing in tMDH.

(B) Determinants of Thermostability in tMDH. At this stage of our understanding of protein thermostabilization, we perceive two distinct goals. One is identification of important determinants within a given three-dimensional protein structure. The second, and broader, goal would be identification of determinants which may in future investigations prove predictive of thermostabilization. This also implies that such determinants could be used in designing thermostability.

Four ion pairs have been identified in tMDH which may play significant roles in thermostabilization. All four are located in positions where some additional intramolecular "glue" may enhance protein stability. Three of these ion pairs are located at the interface between monomers, and one ion pair is located at the interface between domains. The magnitude of the effects of these changes on thermostability remains to be measured. Assuming that they are significant contributors to thermostability, we could say that we have accomplished the first goal that we have set for ourselves. However, on the basis of our inability to predict these types of interactions, we were unsuccessful in our second goal.

Two other features, one putative helix dipole charge stabilization and the putative stabilization of helices by incorporation of alanines, may have accomplished both goals. Again, the measurement of the thermostabilization caused by these effects remains to be made in tMDH, but these would both be identifiable features based upon modeling and therefore predictive of thermostabilization. Two other features, increased hydrophobicity and decreased flexibility in α-helical and interdomain regions, appear to be predictive of thermostability. However, because of the small changes in these properties, their net contribution to thermostabilization may be small.

ACKNOWLEDGMENT

We thank the research group of Dr. F. Scott Mathews, especially Dr. R. Durley and Dr. F. Scott Mathews, for their assistance.

REFERENCES

Abad-Zapatero, C., Griffith, J. P., Sussman, J. L., & Rossmann, M. G. (1987) *J. Mol. Biol. 198*, 445–467.

Banaszak, L. J., & Bradshaw, R. A. (1975) *Enzymes (3rd Ed.) 11A*, 369–396.

Birktoft, J. J., & Banaszak, L. J. (1983) *J. Biol. Chem. 258(1)*, 472–482.

Birktoft, J. J., & Banaszak, L. J. (1984) *Peptide and Protein Reviews*, Vol. 4, pp 1–46, Marcel Dekker, New York.

Birktoft, J. J., Fernley, R. T., Bradshaw, R. A., & Banaszak, L. J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6166–6170.

Birktoft, J. J., Rhodes, G., & Banaszak, L. J. (1989) *Biochemistry 28*, 6065–6081.

Brunger, A. (1990) X-PLOR Manual (version 2.1).

Eklund, H., Samana, J. P., & Jones, T. A. (1984) *Biochemistry 23*, 5982–5996.

Fitzgerald, P. M. D. (1988) *J. Appl. Crystallogr. 21*, 273–278.

Hall, M. D., Levitt, D. G., & Banaszak, L. J. (1992) *J. Mol. Biol. 226*, 867–882.

Holbrook, J. J., Liljas, A., Steindel, S. J., & Rossmann, M. G. (1975) *Enzymes (3rd Ed.) 11A*, 191–292.

Iijima, S., Saiki, T., & Beppu, T. (1980) *Biochim. Biophys. Acta 613*, 1–9.

Iijima, S., Saiki, T., & Beppu, T. (1984) *J. Biochem. 95*, 1273–1281.

Iijima, S., Uozumi, T., & Beppu, T. (1986) *Agric. Biol. Chem. 50(3)*, 589–592.

Karplus, P. A., & Schulz, G. E. (1985) *Naturwissenschaften 72*, 558–563.

Kelly, C. A., Sarfaty, S., Nishiyama, M., Beppu, T., & Birktoft, J. J. (1991) *J. Mol. Biol. 221*, 383–385.

Lesk, A. M. (1985) *J. Mol. Biol. 183(2)*, 267–270.

Lesk, A. M., & Hardman, K. D. (1982) *Science 216*, 539–540.

Lesk, A. M., & Hardman, K. D. (1985) *Methods Enzymol. 115*, 381–390.

Luzatti, V. (1952) *Acta Crystallogr. 5*, 808–810.

Matthews, B. (1990) in *Crystallographic and Modeling Methods in Molecular Design* (Bugg, G. E., & Ealick, S. E., Eds.) pp 80–94, Springer-Verlag, New York.

Menendez-Arias & Argos (1989) *J. Mol. Biol. 206*, 397–406.

Nicholson, H., Anderson, D. E., Dao-pin, S., & Matthews, B. W. (1991) *Biochemistry 30*, 9816–9828.

Nishiyama, M., Matsubara, N., Yamamoto, K., Iijima, S., Uozumi, T., & Beppu, T. (1986) *J. Biol. Chem. 261(30)*, 14178–14183.

Nishiyama, M., Shimada, K., Horinouchi, S., & Beppu, T. (1991) *J. Biol. Chem. 266(22)*, 14292–14299.

Perutz, M. F., & Raidt, H. (1975) *Nature 255*, 256–259.

Roderick, S. L., & Banaszak, L. J. (1986) *J. Biol. Chem. 261(20)*, 9461–9464.

Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., & Zehfus, M. H. (1985) *Science 229*, 834–838.

Rossman, M. G., Liljas, A., Branden, C. I., & Banaszak, L. J. (1975) *Enzymes (3rd Ed.) 11A*, 61–102.

Roussel, A. TURBO-FRODO version 3.0 release a, Jan 1992, Bio-Graphics.

Tronrud, D., TenEyck, L., & Matthews, B. W. (1987) *Acta Crystallogr. 43A*, 489–501.

Vihinen, M. (1987) *Protein Eng. 1(6)*, 477–480.

Walker, J. E., Wonacott, A. J., & Harris, J. I. (1980) *Eur. J. Biochem. 108*, 581–586.

Webb, J. J., Hill, E. J., & Banaszak, L. J. (1973) *Biochemistry 121*, 5101–5108.

White, J. L., Hackert, M. L., Buehner, M., Adams, M. J., Ford, G. C., Lentz, P. J., Jr., Smiley, I. E., Steindel, S. J., & Rossmann, M. G. (1976) *J. Mol. Biol. 102*, 759–779.

Zhang, X.-J., Baase, W. A., & Matthews, B. W. (1991) *Biochemistry 30*, 2012–2017.

Zuber, H. (1988) *Biophys. Chem. 29*, 171–179.

Zulli, F., Schneiter, R., Urfer, R., & Zuber, H. (1991) *J. Biol. Chem. Hoppe-Seyler 372*, 363–372.