Thomson, A. J. (1985) in *Metalloproteins* (Harrison, P., Ed.) Part 1, pp 78–121, Verlag Chemie, Weinheim.
Tollin, G., Cheddar, G., Watkins, J. A., Meyer, T. E., & Cusanovich, M. A. (1984) *Biochemistry 23*, 6345–6349.
Tollin, G., Meyer, T. E., & Cusanovich, M. A. (1986) *Biochim. Biophys. Acta 853*, 29–41.
Watkins, J. A. (1986) Ph.D. Thesis, University of Arizona, Tucson, AZ.

Refined Crystal Structure of Cytoplasmic Malate Dehydrogenase at 2.5-Å Resolution†

Jens J. Birktoft,* Gale Rhodes,‡ and Leonard J. Banaszak§

Department of Biological Chemistry, Division of Biology and Biomedical Sciences, Washington University School of Medicine, St. Louis, Missouri 63110

Received December 15, 1988; Revised Manuscript Received March 21, 1989

ABSTRACT: The molecular structure of cytoplasmic malate dehydrogenase from pig heart has been refined by alternating rounds of restrained least-squares methods and model readjustment on an interactive graphics system. The resulting structure contains 333 amino acids in each of the two subunits, 2 NAD molecules, 471 solvent molecules, and 2 large noncovalently bound molecules that are assumed to be sulfate ions. The crystallographic study was done on one entire dimer without symmetry restraints. Analysis of the relative position of the two subunits shows that the dimer does not obey exact 2-fold rotational symmetry; instead, the subunits are related by a 173° rotation. The structure results in a R factor of 16.7% for diffraction data between 6.0 and 2.5 Å, and the rms deviations from ideal bond lengths and angles are 0.017 Å and 2.57°, respectively. The bound coenzyme in addition to hydrophobic interactions makes numerous hydrogen bonds that either are directly between NAD and the enzyme or are with solvent molecules, some of which in turn are hydrogen bonded to the enzyme. The carboxamide group of NAD is hydrogen bonded to the side chain of Asn-130 and via a water molecule to the backbone nitrogens of Leu-157 and Asp-158 and to the carbonyl oxygen of Leu-154. Asn-130 is one of the corner residues in a β-turn that contains the lone cis peptide bond in cytoplasmic malate dehydrogenase, situated between Asn-130 and Pro-131. The active site histidine, His-186, is hydrogen bonded from nitrogen ND1 to the carboxylate of Asp-158 and from its nitrogen NE2 to the sulfate ion bound in the putative substrate binding site. In addition to interacting with the active site histidine, this sulfate ion is also hydrogen bonded to the guanidinium group of Arg-161, to the carboxamide group of Asn-140, and to the hydroxyl group of Ser-241. It is speculated that the substrate, malate or oxaloacetate, is bound in the sulfate binding site with the substrate 1-carboxyl hydrogen bonded to the guanidinium group of Arg-161.

The malate dehydrogenases (malate dehydrogenase, L-malate:NAD oxidoreductase, EC 1.1.1.37)[1] are oxidoreductases that interconvert the substrates malate and oxaloacetate utilizing the NAD/NADH cofactor system. In eukaryotic cells, malate dehydrogenase occurs in two forms, both of which are synthesized in the cytoplasm. After synthesis the mitochondrial form is imported into the mitochondrial matrix (Chen & Freeman, 1984), a process that is facilitated by a translocation sequence (Grant et al., 1986). The cytoplasmic form, sometimes called soluble malate dehydrogenase, remains in the cytoplasm after synthesis as an N-terminal acetylated product. Both forms of malate dehydrogenase are homodimers with molecular masses of 2 × 35 and 2 × 33 kDa for the cytoplasmic and mitochondrial forms, respectively.

Cytoplasmic malate dehydrogenase was among the first oligomeric enzymes to have its three-dimensional structure determined at high resolution (Hill et al., 1972; Webb et al., 1973). Even though the amino acid sequence remained undetermined for a period of time, a number of significant conclusions could be reached regarding its structure–function relationships as well as evolutionary relationships to other nucleotide binding proteins (Webb et al., 1973; Banaszak & Webb, 1975; Birktoft & Banaszak, 1983; Birktoft et al., 1982a,b).

Recently the determination of the amino acid sequence of cMDH has been completed, by a combination of X-ray and chemical methods (Birktoft et al., 1987; Bradshaw et al., 1988).[2] With this information in hand, the restrained, least-squares refinement initiated in the absence of a complete amino acid sequence (Birktoft et al., 1982b) was continued. While this refinement was in progress, a partial cDNA nucleotide sequence for porcine cMDH was reported (Joh et al.,

† This work was supported by NSF Grant PCM-8208894 (L.J.B.) and National Institutes of Health biomedical research support grant to Washington University School of Medicine (J.J.B.).

* Address correspondence to this author at the Department of Biological Chemistry, Washington University School of Medicine, 660 South Euclid Ave., Box 8094, St. Louis, MO 63110.
‡ Present address: Department of Chemistry, University of Southern Maine, 96 Falmouth St., Portland, ME 04103.
§ Present address: Department of Biochemistry, University of Minnesota, Minneapolis, MN 55455.

[1] Abbreviations: rms, root mean square; cMDH, cytoplasmic malate dehydrogenase; mMDH, mitochondrial malate dehydrogenase.

[2] J.-K. Fang, M. J. Wade, B. E. Glatthaar, M. R. Sutton, G. R. Barbarash, G. A. Grant, and R. A. Bradshaw, unpublished results. R. A. Bradshaw, J.-K. Fang, R. T. Fernley, J. J. Birktoft, and L. J. Banaszak, unpublished results.

0006-2960/89/0428-6065$01.50/0

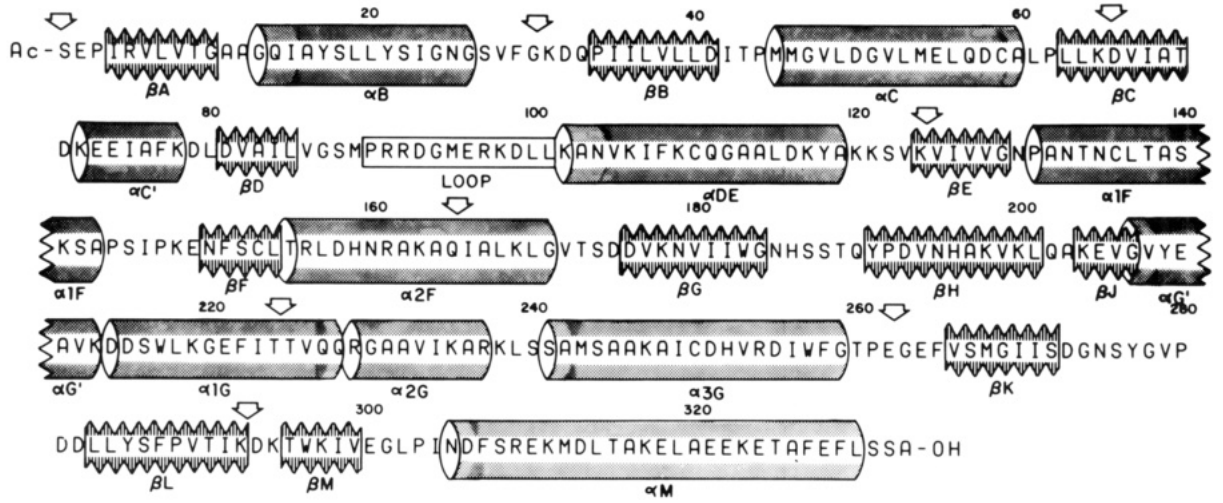

FIGURE 1: Amino acid sequence of porcine cytoplasmic malate dehydrogenase. The amino acid sequence amino acid sequence as shown was derived from a combination of chemical amino acid sequencing, cDNA nucleotide sequence, and X-ray methods as described in the text. Ac- is an abbreviation for the N-acetylated amino terminus. Every 20th residue is numbered above the sequence. Helical segments are enclosed in cylindrical tubes and are identified with an α followed by a number and/or a letter. Segments that form β-sheets are enclosed in the saw-toothed ribbons and are identified by a β followed by a letter. The large arrows mark the locations of the introns as determined in the genomic sequence of mouse cMDH (Setoyama et al., 1988).

1987), resulting in a few revisions of the amino acid sequence. This cDNA sequence and the combined X-ray/chemical sequence (Birktoft et al., 1987) were used to derive a consensus amino acid sequence, Figure 1, which was used in the further refinement.

Here we report the results of the restrained, least-squares refinement of porcine heart cMDH at 2.5-Å resolution. The refined models form the basis for the discussion of the molecular architecture of cMDH and of the interactions with the NAD, as well as a proposed description of the binding site for the substrates, malate and oxaloacetate. The structure of the apo form of cMDH is currently being refined (Carnahan et al., unpublished results) and a comparison between the apo and holo forms of cMDH will be reported elsewhere. Also appearing as separate publications are the detailed comparisons of the cMDH structure with that of mitochondrial malate dehydrogenase (Birktoft et al., 1988) and of the malate dehydrogenases with the lactate dehydrogenase structures (Birktoft & Banaszak, 1983).

Materials and Methods

Detailed information concerning crystal properties and the collection of X-ray diffraction data can be found in a previous publication (Hill et al., 1972) and will only be mentioned here in a summary form. The form of cMDH that has been refined and that is described in this paper was crystallized from 65% saturated ammonium sulfate buffered at pH 5.1 with acetate buffer, and with a 10-fold molar excess of NAD added to the crystallization medium. cMDH crystallizes in space group $P2_12_12$ with the following unit cell dimensions: a = 139.2 Å; b = 86.6 Å; c = 58.8 Å. The asymmetric unit contains one molecule of the cMDH homodimer with a molecular mass of 70 kDa. The X-ray diffraction data were collected by diffractometry from 32 different crystals. In the resolution range 6.0–2.5 Å, 22910 of the 23373 theoretically possible reflections were collected. All observed data were used throughout the crystallographic refinement.

Initially the refinement of cMDH utilized the restrained least-squares procedure of Hendrickson and Konnert (1980) and was later continued with a program called TNT developed by Tronrud et al. (1987). The dictionary of standard amino acid geometries as supplied by Tronrud et al. (1987) was used without modifications. In addition to the standard amino acids, cMDH contains an acetyl group covalently attached to the amino terminus of each subunit. The standard geometry of the acetyl group was derived from glycine by removing the nitrogen atom. Also bound to each cMDH subunit is one NAD molecule and one large anion, presumed to be a sulfate ion. The standard geometry for NAD was derived from the ideal structure of NADP (Bolin et al., 1982) and was provided by Dr. D. A. Matthews (University of California at San Diego). The NAD structure was obtained simply by removing the 2′-phosphate from the NADP structure. The sulfate standard group structure was already part of the TNT program package. On a VAX II microcomputer, one refinement cycle with about 5500 atoms in the asymmetric unit and 23000 reflections took about 2.33 h of cpu time.

At various intervals the refined model was inspected and adjusted as required on an IRIS 3030 graphics display system (Silicon Graphics Inc., Mountain View, CA 94043) using TOM, a version of the program FRODO (Jones, 1978) as modified by Dr. Christian Cambillau (University of Marseille, France). The models were checked and adjusted against electron density maps calculated with as coefficients either $|F_{obs}| - |F_{calc}|$ or $2|F_{obs}| - |F_{calc}|$, where $|F_{obs}|$ and $|F_{calc}|$ are the observed and calculated structure amplitudes, respectively. Phases were calculated from atomic coordinates, and Sim weighting was used in the phase calculation (Sim, 1960). The Fourier maps for model building and refinement were calculated on a grid corresponding to 0.74 Å or about one-third of the nominal resolution of the diffraction data. During the early stages of the refinement all residues were inspected on the graphics system. Later, only those residues that were suspected to be associated with errors were examined. The main criteria used for selecting a residue for inspection on the graphics system were as follows: (a) that the parameters for the main-chain dihedral angle placed the residue outside the acceptable regions in the Ramachandran diagram (see also Figure 5); (b) that the peptide bond deviated significantly from planarity; (c) that bond length and bond angles deviated from their canonical values by more than 0.1 Å and 4.0°, respectively; and (d) that an atom was located in low or negative electron density in the $(2F_{obs} - F_{calc})$ map. In addition, all peaks in the $(F_{obs} - F_{calc})$ map with an absolute value of more than 3σ were also inspected on the graphics system.

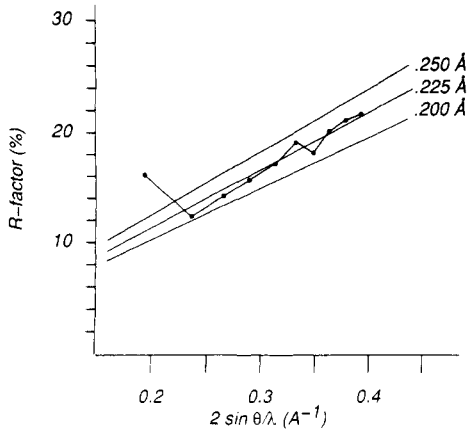

FIGURE 2: *R* factor plotted as a function of resolution. The theoretical estimates of the mean positional errors in atomic coordinates according to Luzzati (1952) are shown superimposed on the curve for the cMDH diffraction data.

Solvent molecules were identified as well-defined positive peaks in the electron density maps. They were included in the atomic list only if both the $(2F_{obs} - F_{calc})$ and the $(F_{obs} - F_{calc})$ maps had significant electron density and the solvent molecule could form hydrogen bonds with protein atoms or with other solvent atoms. Solvent atoms were assigned as water molecules and refined as oxygen atoms with individual temperature factors. Those that refined to positions too close to other atoms or ended up located in low electron density were removed from the solvent list. At infrequent intervals, individual temperature factors were refined isotropically. Prior to the start of each set of temperature factor refinement, average values were calculated for the four main-chain atoms and for the side-chain atoms of each residue. A single value for the temperature factor was calculated for each of the sulfate ions and NAD molecules. The occupancy for all atoms, including solvent atoms, was kept constant.

The coordinates that form the basis of the structure described in this paper have been deposited in the Protein Data Bank at the Brookhaven National Laboratory (Bernstein et al., 1977) with identification code 4MDH.

RESULTS

Refined Model of Cytoplasmic Malate Dehydrogenase. For the refined cMDH model, the conventional crystallographic *R* factor is 16.7% when all observed X-ray diffraction data (22 910 reflections) between 6.0 and 2.5 Å are included and is 16.1% when the weakest 5% of the diffraction data are excluded from the *R* factor calculation. In restrained crystallographic refinement, the relative weights given the geometrical terms and the diffraction data will influence the stereochemical quality and the *R* factor of the resulting model. The rms deviations from the canonical values used in the geometric terms in the refinement of cMDH are 0.017 Å for bond length and 2.57° for interbond angles. The mean value for the peptide bond torsional angle is 179.0°, and the rms deviation for the out-of-plane deviation for the five atoms in each peptide linkage is 0.02 Å.

An estimate of the upper limit of error in atomic coordinates can be obtained by the method of Luzzati (1952). The plot presented in Figure 2 suggests that the upper limit for the mean error of the refined cMDH coordinates is around 0.22 Å. An impression of the quality of the final model and its correspondence with a $2|F_{obs}| - |F_{calc}|$ electron density map is shown in Figure 3. This figure shows a representative segment of the final electron density map and includes parts of residues 175–179 as well as two solvent molecules, Wat-7 and Wat-9, that are labeled as 7-OH and 9-OH, respectively.

Amino Acid Sequence. The amino acid sequence of cMDH was derived from a combination of chemical and X-ray methods as described previously (Birktoft et al., 1987). This approach yielded a sequence of 331 amino acids, but it was also clear that for a few residues some uncertainty remained. In the course of the subsequent refinement and refitting, it became obvious that additional sequence revisions were necessary. Inspection of difference electron density maps suggested that at positions 2 and 3 the Pro-Glu sequence should be in inverted, i.e., Glu-Pro, and further that residue 205 should be changed from Ala to Glu. While the refinement was progressing, the nucleic acid sequence for a cDNA clone covering part of the coding region for porcine cMDH was published, confirming both observations (Joh et al., 1987).

The derived amino acid sequence, covering residues from position 45 to the carboxyl terminus, showed that residues 87 and 88 should be Gly-Ser rather than Ser-Gly, that residues 199–202 should be Val-Lys-Leu-Gln rather than Leu-Lys-Gln-Glu, and that the carboxyl terminus should be extended by two more residues, Ser-Ala. As noted above, the amino acid sequence for the first 44 residues was not reported by Joh

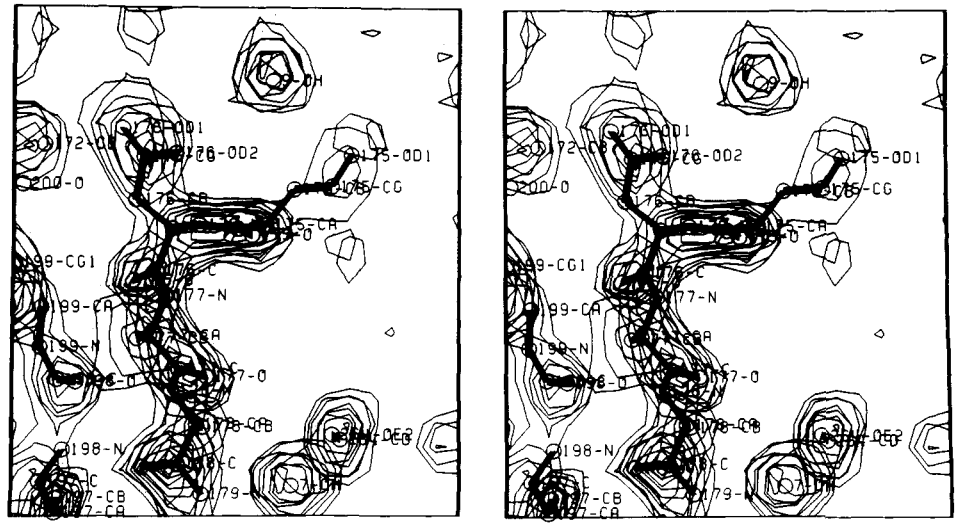

FIGURE 3: Stereoscopic view illustrating the final $2|F_{obs}| - |F_{calc}|$ map. Atoms of the molecular model that corresponds to this part of the map are shown as circles, and bonds are shown as solid, heavy lines. The atoms are identified by atom name and residue number. Contour levels are drawn at 5%, 10%, 20%, ... of the maximum height of the electron density map.

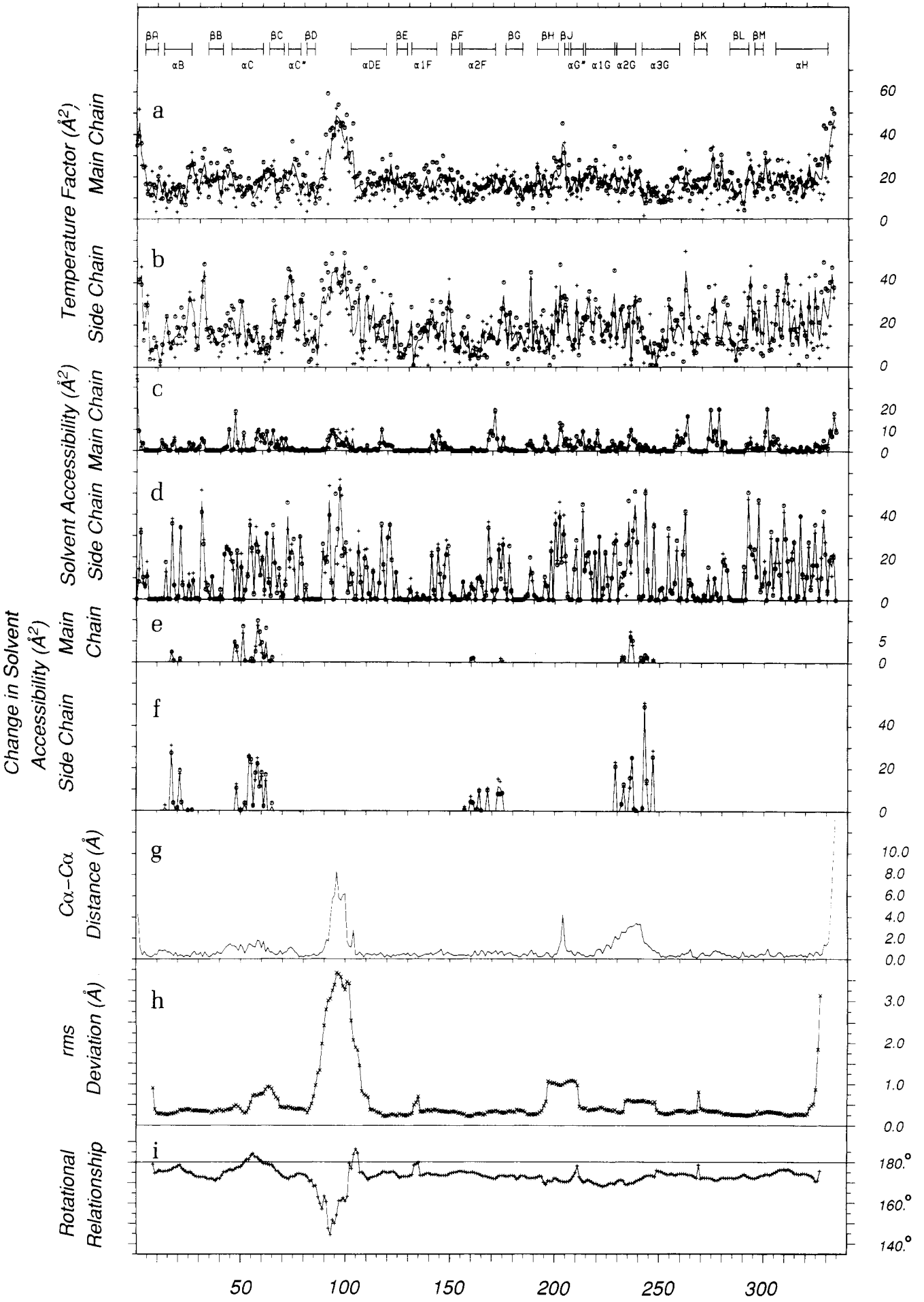
βA βB βC βD βE βF βG βH βJ βK βL βM
αB αC αC" αDE α1F α2F αG" α1G α2G α3G αH
a
b
c
d
e
f
g
h
i
Temperature Factor (Å²) Main Chain
Side Chain
Solvent Accessibility (Å²) Main Chain
Side Chain
Change in Solvent Accessibility (Å²) Main Chain
Side Chain
Cα–Cα Distance (Å)
rms Deviation (Å)
Rotational Relationship
60
40
20
0
40
20
0
20
10
0
40
20
0
5
0
40
20
0
10.0
8.0
6.0
4.0
2.0
0.0
3.0
2.0
1.0
0.0
180.°
160.°
140.°
50
100
150
200
250
300

FIGURE 4: Composite plot of several properties of cMDH as a function of residue number. The locations of the elements of secondary structure are indicated at the top of the figure. (a) Plot of mean temperature factor for the four main-chain atoms. The solid line represents the average for subunits A and B. The (+) and (O) marks are the values for subunits A and B, respectively. (b) The mean temperature factor for the side-chain atoms plotted in the same fashion as in (a). (c) Plot of the total solvent accessibility in Å^2 for the four main-chain atoms. The solid line and the (+) and (O) marks have the same meaning as in (a). (d) Plot of the total solvent accessibility in Å^2 for side-chain atoms. Static solvent accessibility was calculated with the method and program of Richmond (Richmond & Richards, 1978). (e) Plot of the change in solvent accessibility in Å^2 for the four main-chain atoms resulting from dissociating the cMDH dimer into monomer, assuming no conformational changes. The vertical scale is twice that in (c), (d), and (f). (f) Plot of the change in solvent accessibility in Å^2 for the side-chain atoms resulting from the hypothetical dissociation of the cMDH dimer, assuming no conformational changes. (g) plot of distances between equivalent Cα atoms after least-squares superpositioning using all Cα atom pairs in the two subunits. Application of the rotation-translation matrix in Table IV and comparison of the coordinates will yield this set of distances. (h) plot of the local symmetry relationship between cMDH subunits A and B. The relationship between subunits A and B was analyzed in blocks of 15 residues as described in the text. The curve shows the rms deviation for the Cα atoms in each 15-residue block plotted versus the residue number in the middle position. (i) Plot of the local relationship between subunits A and B. The curve shows the rotational relationship between each block of 15 residues. The value of the polar angle κ is plotted in the same fashion as in (h).

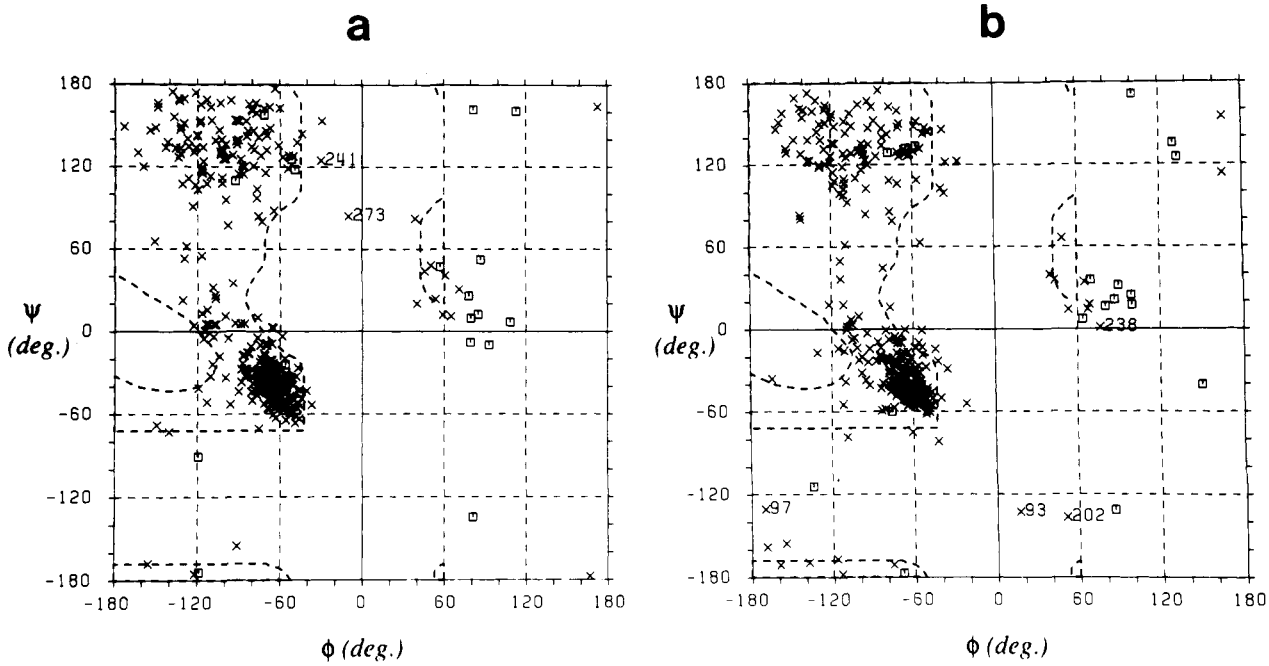

FIGURE 5: Ramachandran diagram of the crystallographic model of cMDH. The main-chain torsional angle Φ (N–Cα bond) is plotted versus Ψ (Cα–C′ bond). The following symbols are used: (×) non-glycine residues; (□) glycine residues. Non-glycine residues that are more than 15° outside the permissible regions are identified by residue numbers. (a) Subunit A; (b) subunit B.

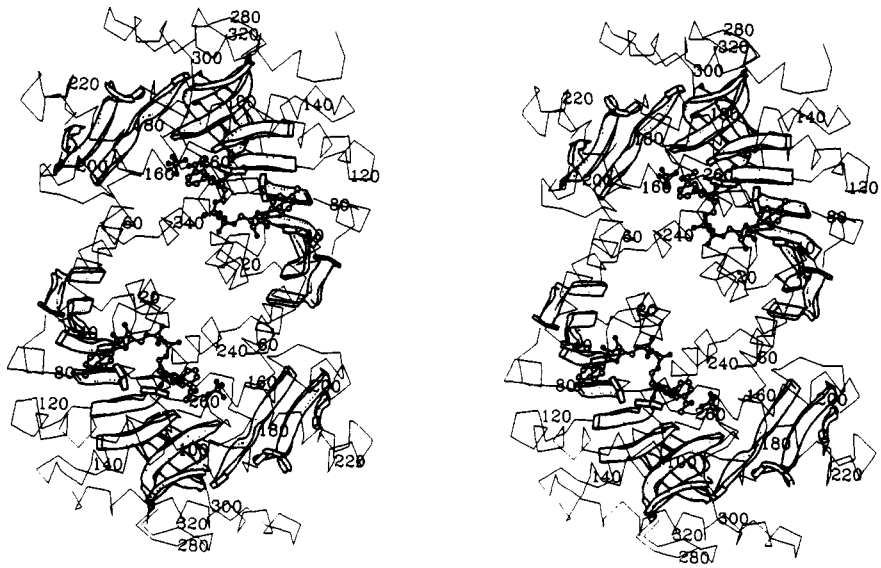

FIGURE 6: Stereo representation of dimeric cMDH. The view in this stereo cartoon is down the molecular symmetry axis, which is an approximate 2-fold axis located in the center of the drawing. The residue numbers for every 20th residue are marked. The bound NAD's and sulfate ions are drawn as solid ball and stick models. The stereo diagrams throughout this paper were produced by the computer program ARPLOT of Lesk and Hardman (1982, 1985).

et al. (1987). Of the remaining 289 residues in porcine cMDH, 93% are identical with those in mouse cMDH (Joh et al., 1987), and furthermore, with the exception of the inverted Glu-Pro sequence at positions 2 and 3, the first 44 residues in the previously proposed porcine cMDH sequence (Birktoft et al., 1987) are identical with those of the mouse amino acid

sequence (Joh et al., 1987). Consequently, it was assumed that porcine and mouse cMDH have identical amino acid sequences for residues 1–45.

The amino acid sequence of porcine cMDH that forms the basis for the structure described here is shown in Figure 1. Overall, the final X-ray maps are in agreement with the adopted amino acid sequence with the exception of a few places where only weak electron density is observed for side chains. Examples are lysines A102,[3] B72 and B109, glutamates A73 and A149, and glutamines B202 and B228. In a few regions, the entire polypeptide chain is located in weak or poorly defined electron density. This is the case for the following residues: A93–A100, A330–A333, B1 and B2, B93–B100, and B330–B333, all of which are located on the molecular surface.

The refined isotropic temperature factor provides another parameter that can illustrate the overall quality of the cMDH structure. The results for the two cMDH subunits are shown in Figure 4a,b. In these illustrations, separate average temperature factors were calculated for the main-chain atoms and for the side-chain atoms. Overall, the patterns for the two subunits are very similar, and the only significant differences are observed for residues that in the crystalline lattice are involved in different intermolecular contacts. Not surprisingly, the main-chain temperature factors, Figure 4a, are more similar in the two subunits than are those for the side-chain atoms, Figure 4b.

Other residues, and in particular Glu-A74, Ala-A203, and Ala-B203, all have relatively higher temperature factors and are located on the surface of the cMDH dimer. In fact, the overall pattern for the temperature factors is strongly correlated to that of solvent accessibility as can be appreciated by comparing the plots of temperature factors, Figure 4a,b, with the plots of solvent accessibility, Figure 4c,d.

General Properties of the Dimer. In Figure 5, the Ramachandran diagrams of the two cMDH subunits are presented separately with glycine residues represented by squares. The Φ and Ψ values were not restrained to any fixed values during the refinement but were adjusted manually to conform to the recognized limits at intermediate stages of the refinement. As is apparent from Figure 5, most backbone dihedral angles either fall within or are less than about 15° removed from the permissible regions. The few outliers are identified in Figure 5 with their residue numbers. Two of these, Asp-B93 and Arg-B97, are located in regions of poor density, and consequently, the conformation of these residues must be considered to be somewhat uncertain.

Of the remaining residues, Asp-A273 and Gln-B202 both form the second residue in hairpin turns, and in both instances, the conformations appear to be stabilized by hydrogen bonds. The main-chain conformations near Asp-273 are similar in subunits A and B, and the (Φ,Ψ) values are (−36,100) and (−10,84) for subunits A and B, respectively. For the A subunit this is less than 10° outside an acceptable region. The main-chain conformation in both subunits is stabilized by hydrogen bonds between O Ser-272 and N Asn-275 and between OD1 Asp-272 and N Asp-282. The orientation of the 272–273 peptide bond is further stabilized by a hydrogen bond between N Gly-273 and OD1 Asp-281.

The conformations at residues Gln-A202 and Gln-B202 are different. While in subunit A the (Φ,Ψ) values for Gln-202 are acceptable, in subunit B they place this residue in the lower-right quadrant of the Ramachandran plot. Although hydrogen bonds between N Leu-201 and O Lys-204 and between O Leu-201 and N Lys-204 are observed in both subunits, the main-chain conformations differ. A factor that is likely to contribute to this conformational difference is the different types of intermolecular interactions made by the two subunits. In both subunits, this part of the polypeptide chain is in direct contact with other molecules in the crystal lattice. However the intermolecular contacts are made with different symmetry-related molecules.

The (Φ,Ψ) values for Ser-A241 and Lys-B238 are both near an acceptable region and located near the corresponding residue in the other subunit. Both residues are located near the subunit interface and are in regions that differ in conformation in the two subunits, as is apparent in Figure 4g. In both subunits, O Ser-240 is hydrogen bonded to NH2 Arg-237. However, the orientation of the Arg-237 side chains differ in the two subunits, resulting in different orientations for the 240–241 peptide bond. In subunit A, O Ser-A241 is hydrogen bonded to a solvent molecule, Wat-443, and N Ala-A242 is hydrogen bonded to OE2 Glu-B55, i.e., a residue in the other subunit. In subunit B, O Ser-B241 is hydrogen bonded to N Ala-A245, while the hydrogen bond to OE2 Glu-A55 in the other subunit is donated by N Ser-B244. Lys-238 is the last residue in helix α2G, and while the main-chain hydrogen-bonding pattern is the same in the two subunits, the conformations differ slightly. In summary, all the residues mentioned above as having abnormal Ramachandran parameters are located on the molecular surface, and in most instances they appear to be involved in intermolecular or intersubunit interactions that differ in the two subunits.

The surface and interior properties of the cMDH molecule can be ascertained from X-ray coordinates by assessing the solvent accessibility. With a solvent probe with a radius of 1.4 Å, the variation in solvent accessibility is shown in Figure 4c,d. The reference or standard state to which all other results are best compared is the cMDH dimer in the presence of bound NAD and sulfate ion, but in the absence of all other solvent molecules. The accessibilities for the main-chain atoms, Figure 4c, are very similar for the two subunits with the biggest differences observed for residues 90–105 and somewhat smaller differences for residues 55–69. These are the same residues that differ most in their temperature factors, Figure 4a,b, and, as will be discussed in more detail later, appear to be conformationally different in the two subunits. As might be expected, the scatter for the side-chain values, Figure 4d, is somewhat larger than that for those for the main chain. When the solvent accessibility is related to the location of elements of secondary structure, as marked at the top of Figure 4, it can be seen that most of the β-strands are buried while at least part of every α-helix is in contact with solvent.

The overall pattern of side-chain exposure to solvent is what might be expected for a soluble, globular protein. Relatively few of the side chains with polar atoms are in internal positions, and in such instances the polar atoms participate in hydrogen bonding. A few ionizable side chains are removed from solvent contact. With one exception such residues either are involved in subunit–subunit interactions, Asp-58 and Glu-55, or are part of the active center of cMDH, Asp-158 and Arg-161. The exception is Asp-193, which does not appear to be of any obvious importance in the function of cMDH. This residue is located in an external surface loop that assumes slightly different conformations in the two subunits. The side chain itself is located just below the surface and is shielded from the solvent by several other side chains that are themselves in solvent contact.

Each cMDH subunit contains four histidines, two of which,

[3] When the sequence number refers to a residue in a specific subunit, it is prefixed with an A or B.

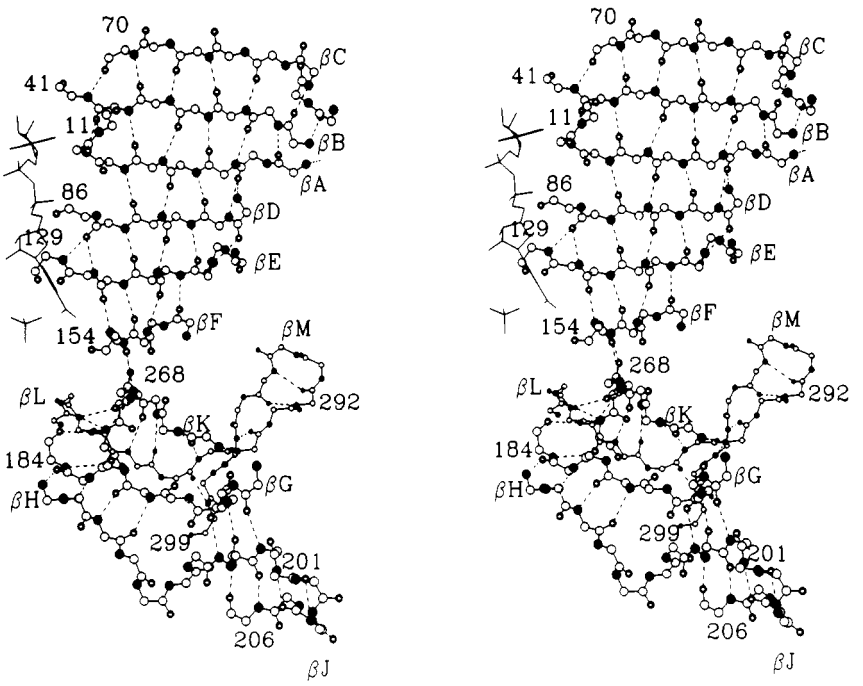

FIGURE 7: Eleven-stranded mixed β-sheet in cMDH. The stereo diagram shows the polypeptide backbone with interstrand hydrogen bonds indicated by stippled lines. The name of each β-strand is placed near the first residue in each strand, and the residue number of the last residue in each strand is also shown.

His-159 and His-252, are internal or inaccessible. In both the A and B subunits, the imidazole ring of His-159 has ND1 hydrogen bonded to OG Ser-267. Similarly, in both subunits, His-252 is involved in two hydrogen bonds: ND1 is hydrogen bonded to O Cys-153 and NE2 is hydrogen bonded to OG Ser-152. Considering the internal position of these two histidines and the absence of nearby negative charges, it seems reasonable to assume that both of these histidine residues are neutral in the structure analyzed here. Protonation of either imidazole ring would probably disrupt the hydrogen bonding, and the positive charge introduced should lead to a destabilization of the structure. His-186 is part of the catalytic site and will be discussed later.

The accessibility calculations can also be used to calculate changes in surface area accompanying dimerization and substrate or coenzyme binding. For example, the molecular surface area is decreased by about 11% upon association of the cMDH subunits to form the dimer. This does not appear to be a very dramatic change, especially since the accessible protein surface area is decreased by about 4% simply by the binding of the coenzyme to cMDH. Binding of NAD to cMDH causes a loss of 81% of the solvent-accessible area of the coenzyme, and when the bound sulfate is included, the decrease is 87%. These accessibility values, as well as changes resulting from NAD binding and subunit assembly, are similar to those calculated for lactate dehydrogenase and glyceraldehyde-3-phosphate dehydrogenase (Chothia & Janin, 1975; Janin & Chothia, 1978).

Secondary Structure of Cytoplasmic Malate Dehydrogenase. A stereoview of the structure of cMDH is presented in Figure 6 in the form of a cartoon where α-helices are presented as an α-carbon model and β-strands as highlighted curved arrows. The elements of secondary structure are also marked schematically in Figure 1. In the view given in Figure 6, one is looking down the molecular dyad axis and into the NAD binding site. For the purpose of classification, we have considered a residue to be part of a helix if at least one helical-type hydrogen bond, α-type or 3_{10}-type, is made and, further, if the torsional angles for the polypeptide backbone for that segment are approximately those of an α-helix. An exception from the criterion of torsional angles is made for the first or last residue in a helical segment. A similar set of criteria for torsional angles and hydrogen bonding was used for the classification of residues in β-strands.

Two distorted helices, αC′ and αG′, have not previously been recognized as elements of secondary structure in cMDH (Hill et al., 1972). Helix αC′, residues 72–78, does not have the usual helical hydrogen-bonding pattern. Helix αG′, residues 207–213, is slightly more than one helical turn long and contains four of the main-chain hydrogen bonds characteristic of a helix.

Helix αB has a rather pronounced bend or kink in its middle, and this helix has been divided into two segments, named α1B and α2B, at residue 20. The angle between the helical axes in the two halves of this helix is about 25°. αH also shows minor bending with a directional change of about 5–10°. The convex sides of the bent helices are toward the enzyme surface, so that any peptide nitrogens or oxygens that do not participate in intraprotein hydrogen bonding are in a position to interact with the solvent. No proline residues are observed at the bending points. Similar bending in α-helices in the absence of proline has also been observed in other proteins (Remington et al., 1982; Schirmer et al., 1986).

In the isolated cMDH monomer, all of the helices have at least one side exposed to the solvent on the exterior of the subunit, as can be seen in Figure 4c,d. Upon dimerization, the helices αB (15–26), αC (45–59), α2F (155–171), α2C (228–237), and α3G (242–259) have their solvent exposure reduced by the subunit–subunit interactions, Figure 4e,f. These helices can be found in the center of Figure 6.

In contrast to the helical secondary structure, most of the β-strands are located in the interior of the subunit with only the first and occasionally the last residue in contact with solvent; again see Figure 4e,f. One parallel β-strand, βC (63–70), and the antiparallel β-strands βH (191–201), βJ (204–208), and βM (295–299) have more substantial portions in contact with solvent. Overall, the β-strands are assembled into three well-defined β-sheets. The first six strands, βA

through βF, form sheet I. The next three strands, βG through βJ, form an antiparallel β-sheet, sheet II and the last three strands, βK through βM, a second antiparallel β-sheet, sheet III. The three β-sheet structures as they appear in a single subunit of cMDH are shown in Figure 7. A common feature of β-sheets is the characteristic right-handed twist[4] (Richardson, 1981). This feature is also observed in all three sheets in cMDH. Additionally, sheet III has a sharp kink in its middle caused by Pro-288, which is located in the middle of strand βL. The net effect is that the first half of sheet III runs nearly perpendicular to the second half. A similar folding pattern is observed in the intestinal fatty acid binding protein and related proteins (Sacchettini et al., 1988). In these proteins, the result is also that the same β-strand contributes component strands to two β-sheets that are orthogonal to each other.

Another deviation from normal β-sheet structure is the presence of β-bulges. In sheet I, a β-bulge is found in strand βC at position 64. This bulge can be seen in the upper-right corner of Figure 7 next to the βC label. In sheet II, another irregularity in the β-sheet structure is observed. The regular hydrogen-bonding pattern is disrupted by a three-residue insertion, residues 194–196, in strand βH, as shown in the lower part of Figure 7. This three-residue bulge in strand βH forms a 3_{10}-helix structure with a hydrogen bond between N Ala-197 and O Val-194. The interaction between strands βG and βH is further stabilized by hydrogen bonds between the side chain of Trp-183 and Asp-195 and by a solvent molecule that links O Gln-179 and N Asp-193 through hydrogen bonds.

Domain Structure of Cytoplasmic Malate Dehydrogenase. In terms of biochemical function and overall folding, cMDH is a characteristic double-domain structure (Rossmann et al., 1975). The first 153 residues constitute the "coenzyme" or "NAD binding" domain, and the catalytic domain begins with helix α2F. The structures of the two parts are largely self-contained, and with the exception of the C-terminal helix, αH, no elements of secondary structure are shared by both domains. Hence, αH appears to act as a stabilizing crosspiece between the two domains. Similar arrangements and functional roles of long C-terminal helices have also been observed in other multidomain structures such as glyceraldehyde-3-phosphate dehydrogenase (Buehner et al., 1974), alcohol dehydrogenase (Eklund et al., 1981), and papain (Drenth et al., 1971).

The folding pattern of the NAD binding domain is that of a typical parallel α/β structure (Richardson, 1981). In the case of cMDH, the centrally located β-sheet I is flanked by four α-helices, two on each side of the sheet. The five crossover connections between the β-strands are all helical structures and are all of the right-handed type (Richardson, 1977). The overall topography can be described in the following manner: $+1x, +1x, -3x, -1x, -1x$ (Richardson, 1977). As can be seen in Figure 6, several helices are packed against other helices. Theoretical considerations have led to suggestions for three classes of helix–helix interactions (Crick, 1953; Chothia et al., 1977), and the helical packing arrangements in cMDH fall within these three classes.

As noted previously (Rossmann et al., 1975), the NAD binding domain can be separated into two halves, each with similar topology, that can be represented as β–α–β–α–β. The connection between these two subdomains is formed by the irregular helical structure αC′. Although these two assemblies of secondary elements are structurally similar, no significant amino acid sequence homology is present. A similar phenomenon, homologous substructures–no sequence homology, has been described for other proteins such as rhodanase (Ploegman et al., 1978) and the serine proteases (McLachlan, 1979).

The catalytic domain contains the two antiparallel β-sheets, sheet II and sheet III, mentioned earlier. As can be seen in Figure 7, sheets II and III are in contact with each other, with strand βG of sheet II adjacent to strand βL of sheet III such that a five-stranded β-sheet is formed. The C-terminal end of βL is located between, and hydrogen bonded to, strands βG and βK, and βL runs antiparallel to both βG and βK. Only two main-chain hydrogen bonds, between residues Gly-184 and Ser-286, are found between βG and βL, but additional stabilization of this sheet structure is provided through hydrogen bonds involving side-chain and main-chain atoms (N Ile-182 to OG Ser-267 and O Ile-182 to OG Ser-286). Strand βM is hydrogen bonded to the C-terminal end of strand βL and is not part of this extended antiparallel β-sheet structure. The remaining part of the catalytic domain is made up of six α-helices: one is at the beginning of the domain, four consecutive helices are found in the middle situated between the two three-stranded β-sheets, and a long helix is present at the C-terminal end of the domain and the polypeptide chain.

The nonpolar contacts between the two domains involve several helix–helix interactions. The angles between the axes of the helices in the domain interface fall within the three expected ranges predicted for helix–helix interactions. Approximately 29 interdomain hydrogen bonds are formed, and most of these involve at least one side-chain atom. The parallel β-sheet in the NAD binding domain continues directly into the antiparallel β-sheet in the catalytic domain. Two backbone–backbone hydrogen bonds, between residues 153 and 269, mediate the link between the two sheets, linking strands βF and βK in an antiparallel fashion. The sheet structure is further stabilized by hydrogen bonds mediated by a solvent molecule that links N Ile-271 and O Phe-151. The result is the formation of an irregular 11-stranded mixed β-sheet as shown in Figure 7.

This 11-stranded mixed β-sheet is encompassed by both domains, runs through the middle of the entire subunit, and is largely inaccessible to the solvent. The overall right-handed twist between the strands leads to an angular separation of 180° between two extreme ends of the β-sheet, βC and βJ. The various helices are distributed on each side of the sheet, with the helical axes being approximately parallel to the plane of the sheet. Helices αC′, αDE, α1F, and αH are all on the same side of the sheet structure and on the side farthest from the subunit interface. In terms of describing the conformation of a cMDH subunit, there are two somewhat opposing views. On the one hand, we have described two domains, the nucleotide binding domain from the amino-terminal end to helix α2F and the catalytic domain from α2F to the carboxyl terminal. On the other hand, it has been shown that a nearly continuous β-sheet forms the core of the molecule transversing both domains. The conflict is simply due to the lack of a rigid definition for the term "domain", a problem that will not be resolved in this paper.

Symmetry Relationship between Subunits. The majority of dimeric proteins consisting of two chemically identical polypeptide chains have their subunits related by 2-fold rotational symmetry (Banaszak et al., 1981; Matthews & Bernhard, 1973). In the case of cMDH, all existing evidence indicates that the two subunits are chemically identical (Banaszak & Bradshaw, 1975; Bradshaw et al., unpublished re-

[4] The handedness is defined in terms of the hydrogen-bonding direction or of the peptide planes as viewed along a strand (Richardson, 1981).

FIGURE 8: Superposition of the two cMDH subunits. The stereo diagram shows Cα models of subunit B, drawn in thinner lines, superimposed on subunit A, drawn in heavier lines. The position of subunit B was computed with the transformation matrix in Table IV. Every tenth residue in subunit A is marked with its residue number while every fifth residue in subunit B is marked by an arrow. The NAD and sulfate ion bound to subunit A are shown as a ball and stick model.

sults),[2] yet several observations pertaining to the properties of cMDH in the crystalline state suggest that the two subunits are nonequivalent (Banaszak et al., 1971; Glatthaar et al., 1972; Hill et al., 1972; Weininger et al., 1977). First, in the crystalline state the two subunits do not bind NAD or heavy atom containing reagents with equal affinity. Second, comparisons of α-carbon coordinates using least-squares methods all suggest that the subunit relationship departs significantly from exact 2-fold rotational symmetry (Birktoft & Banaszak, 1983). On the other hand, no compelling evidence has been presented for the existence of nonequivalent binding of substrates and cofactors by cMDH in solution [reviewed in Weininger et al. (1977) and Frieden and Fernandez-Sousa (1975)]. Consequently, the observed departure from 2-fold symmetry in the crystalline state must have its basis in effects caused by molecular packing within the crystal lattice. This suggestion is also supported by measurements of the enzymatic activity of cMDH in the crystalline state (Zimmerle & Alter, 1983; Zimmerle et al., 1987). Activity and inhibition studies led these authors to suggest that asymmetry exists between the two subunits, but only when the enzyme is incorporated into a crystalline lattice.

It should be mentioned that simply having a dimeric molecule in the asymmetric unit will result in nonidentical environments for each subunit within the crystalline lattice. Analysis of the intermolecular contacts made within the crystal lattice of cMDH shows that homologous residues in the two subunits participate in very different noncovalent interactions [see Figure 1 in Birktoft and Banaszak (1983)]. In fact, all intermolecular interactions are unique for either subunit. Of particular interest are the close intermolecular contacts made with another molecule by the so-called "loop region", residues 90–103 in subunit A. The corresponding region in subunit B is in an open and unobstructed crystal environment. As will be discussed later, numerous residues that are involved in coenzyme binding are located in or adjacent to this loop. The observed nonequivalence in NAD binding by the two cMDH subunits probably has its structural basis in this difference in the crystal packing environments.

The symmetry relationship between the subunits of cMDH was explored with the method of Rao and Rossmann (1972) both before and after crystallographic refinement. It was found that the 334 α-carbon atoms (333 amino acids + 1 acetyl group) in subunit B could be rotated into the corresponding α-carbon atoms of subunit A within rms distance of 1.59 Å. The rotation axis relating the two subunits is defined by the spherical polar angles $\Phi = 29.8°$, $\Psi = -163.0°$, and $\kappa = 173.6°$. In the convention used here, Φ is the angle between the rotation axis and the crystallographic b axis, Ψ is the angle between the projection of the rotation axis on the a–c plane and the a axis, and κ is the rotation angle. The value for the rotation angle is significantly different from the value for a true 2-fold symmetry operation, i.e., 180°. If we exclude from the least-squares calculation the 37 atom pairs whose deviation exceeds the rms value of 1.59 Å, the rms value for the least-squares fit of the remaining 297 α-carbon atoms is 0.58 Å. However, the parameters for the position and direction of the rotation axis do not change appreciably, and κ is still equal to 173°, as is evident from the data in Table I. Furthermore, the parameters relating the two subunits of cMDH are not noticeably affected when all the atoms in the polypeptide backbone, or for that matter every atom in each subunit, is included in the least squares analysis.

Table I: Parameters Describing the Relationship between the Two Subunits of cMDH

Effect of Selected Coordinates upon the Calculated Subunit–Subunit Transformation[a]

atoms used	no. of atoms	rms	deg Φ	deg Ψ	deg κ	trans(A)
Cα (all)	334	1.59	29.8	−163.0	173.6	−0.62
Cα (partial)[b]	297	0.58	30.6	−163.2	173.5	−0.54
main-chain atoms	1335	1.62	29.7	−163.0	173.7	−0.59
whole subunit	2587	2.17	29.6	−163.3	173.8	−0.49

Rotation-Translation Matrix[c] for Rotating Subunit B into Subunit A

$$X_{rot} = -0.86554x + 0.46781x - 0.17888y + 55.214$$
$$Y_{rot} = 0.49979x + 0.82988y - 0.24802z - 1.799$$
$$Z_{rot} = 0.03242x - 0.30407y - 0.95210z + 89.133$$

[a] The values are for the rotation of subunit B into subunit A. [b] Atoms whose deviation exceeds the rms deviation calculated with all Cα atoms were removed in the calculation with Cα (partial). [c] The numbers are derived from the least-squares fit using all 334 Cα atom in each subunit. x, y, and z are the coordinates of subunit B. X_{rot}, Y_{rot}, and Z_{rot} are the coordinates of subunit B rotated into subunit A.

With the results given in Table I, the alignment of the 334 Cα atoms is illustrated in the stereo diagram in Figure 8. In addition, the deviations in angstroms between corresponding atom pairs are plotted as a function of residue number in Figure 4g. The conformational segments that appear to be different are located in isolated and separate segments of the primary sequence. However, these different segments are all located near the periphery of the molecule and seem to be

Table II: Hydrogen Bonds between Subunits A and B in the cMDH Dimer[a]

	subunit A to subunit B	subunit B to subunit A
OH Tyr-17···N Met-243	2.6	2.8
OH Tyr-17···N Ala-242	3.0	[3.6]
OH Tyr-17···NH1 Arg-237	3.4	[4.4]
O Met-54···NH2 Arg-229	2.8	3.2
OE1 Glu-55···N Ser-244	2.9	2.8
OE1 Glu-55···OG Ser-244	2.6	2.3
OE2 Glu-55···NE Arg-237	[4.6]	3.2
OE1 Gln-57···NH2 Arg-229	2.7	2.4
OD1 Asp-58···NH2 Arg-229	3.2	3.1
OD2 Asp-58···NE Arg-229	2.4	3.4
OD2 Asp-58···NE Arg-161	via W434	2.8
O Cys-59···NZ Lys-247	[4.5]	2.6

[a] Distances are given in angstroms.

clustered in one particular portion of the cMDH subunit. This area, which is located on the right side of Figure 8, contains many of the residues that participate in the binding of NAD and of the sulfate ion and, by implication, of the malate/oxaloacetate substrate. As will be discussed in more detail later, these structural differences result in different interactions between NAD and cMDH for the two subunits.

In order to see if one or more segments of the subunits more closely obeyed 2-fold symmetry, the least-squares analysis was carried out with a sliding block of the Cα atom coordinates of 15 residues. First, residues 1–15 from subunit A were compared with the same 15 residues in subunit B; then, residues 2–16 from the two subunits were compared and so on. The result of this analysis is shown in panels h and i of Figures 4, which present the rotation angle κ relating the structural segments as a function of residue number and the rms deviations for the segments being compared. Not shown are similar results with different segment sizes, which showed that the rms deviations and κ were not dependent on the size of the residue block.

Three observations emerge from this analysis. First, for most of the conformation, the overall rotational relationship is distinctly different from 180° or exact 2-fold symmetry. However, the overall structural similarity between the two subunits is quite good despite the departure from precise 2-fold symmetry. Second, the transformation relating atoms in subunit A to those in subunit B, seen as κ in Figure 4h, varies along the polypeptide chain. Third, several segments of the cMDH structure do, in fact, appear to be related to their counterparts by nearly 180° rotational symmetry. They include the helices $\alpha\beta$ and αC, which are involved in subunit–subunit interactions.

Subunit–Subunit Interactions. The contact region between the two subunits is formed by a bundle of six parallel helices, three from each subunit. Most prominent among the contacts are those made between the αB helices, residues 14–25, of each subunit and those between the αC, residues 47–60, and the α2G and α3G helices of opposite subunits. In addition, residues 157–175 in helix α2F are nearly perpendicular to and in contact with helix αC of the other subunit. The helical nature of the residues in the intersubunit contact region can be appreciated from Figure 4e,f. Here the pattern of change in solvent accessibility follows approximately the pattern characteristic for residues located in a helix. For a few residues, the solvent accessibility has changed as a result of dimerization even though they are not in direct contact across the dimer interface.

In addition to the intersubunit hydrogen bonds listed in Table II, a number of close contacts occur at the interface. In subunit A, a total of 123 atoms of which 70 are from side chains are within 4.25 Å of any atom in subunit B. In subunit B there are 125 atoms of which 70 are from side chains which are within 4.25 Å of any atom in subunit A. With regard to intersubunit hydrogen bonds, all that are listed in Table II have at least one atom participant from a side-chain atom. In addition to the hydrogen bonds that directly link the two subunits, polar interactions in the dimer interface are mediated through a number of solvent molecules. A consequence of the asymmetry in the subunit interface is that few of the solvent–protein interactions in this region are identical for the two subunits. The distances between several of the solvent molecules and the closest polar protein atoms are too long to be assigned as hydrogen bonds. If the solvent bridge, however, was provided by two rather than one solvent molecule, satisfactory hydrogen-bond dimensions would be obtainable. The electron density does, however, support only one and not two solvent molecules.

As can be seen in Table II, several of the side chains in the dimer interface would be expected to be charged at neutral pH. They are Glu-55, Asp-58, Arg-161, Arg-229, Arg-237, and Lys-247. With the exception of the arginines, they all would be expected to change their ionization state at pH values between 4 and 10. The two carboxylic acids are involved in ionic interactions with basic side chains, and protonation of these would be expected to cause destabilization of the dimer interface at low pH. At high pH, the deprotonation of Lys-247, as well as Tyr-17 and Tyr-21, should also tend to favor dissociation.

Dissociation of dimeric cMDH could have have effects on enzymatic activity. Arg-161, found in the subunit interface, has also been implicated in the binding of the substrates, malate and oxaloacetate (Birktoft & Banaszak, 1983). Figure 9 and Table II show that the contact between Arg-A161 and Asp-B58 is a direct hydrogen bond, while a solvent molecule links Arg-B161 and Asp-A58.

Other residues in or near the subunit interface are only indirectly involved in substrate binding and catalysis. Most prominent among these are the residues located in the two helical segments α2G and α3G, as well as the segment interconnecting them. In both of these helices, one side is in contact with the substrate, and the other side of the helix is in contact with the neighboring subunit. It seems reasonable to expect, therefore, that subunit dissociation could have a measurable effect on the enzymatic activity.

Solvent Structure. The refined coordinate list for cMDH includes 471 solvent molecules in addition to sulfate molecules placed in each subunit in two unusually large spheres of electron density located close to the nicotinamide ring (Webb et al., 1973). Recall that the crystals were grown in the presence of $(NH_4)_2SO_4$ (Webb et al., 1973). However, in order to test the possibility of other solutes being bound here, refinement cycles were executed where the putative sulfate ions either were omitted or were replaced with single water molecules. These refinements converged in ten cycles or less. In both instances high levels of electron density in the resulting $|F_{obs}| - |F_{calc}|$ maps suggested that the best interpretation was for sulfate ions rather than water in these two locations. The interactions between these two sulfate ions and the surrounding protein atoms, shown in the stereo diagrams in Figure 9, are also in agreement with an oxygen-containing anion being located in these positions. That is, both observed hydrogen-bonding and charge–charge interactions would accommodate sulfate ions.

In subunit A, the binding site for the sulfate ion is created

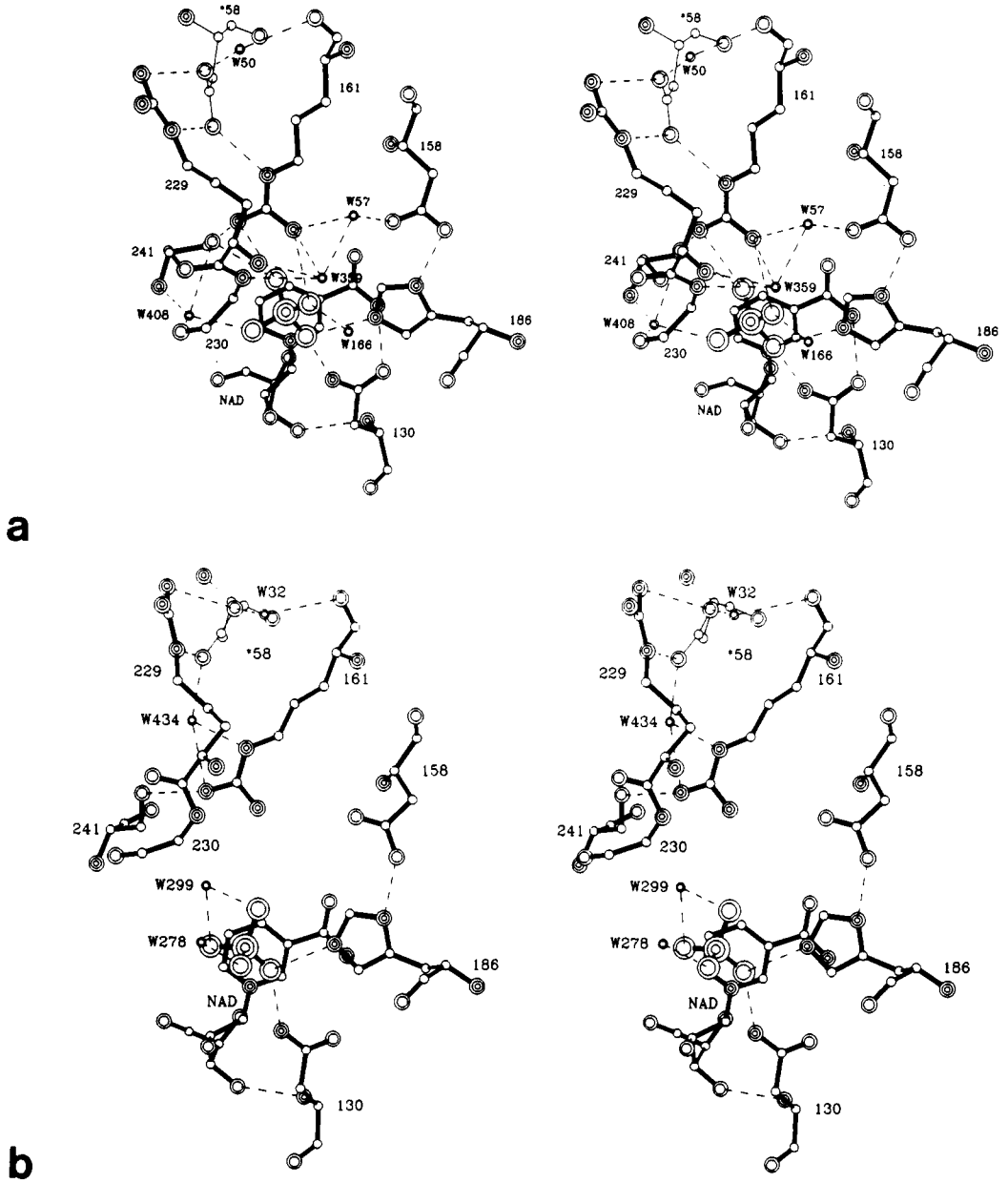

FIGURE 9: Stereo diagram showing the environment of the bound sulfate ion in cMDH. All residues are identified with their number. The residue with the thinner bonds, 58, belongs to the adjacent subunit. Solvent molecules are shown as isolated balls, and their residue numbers are preceded by a W. Hydrogen bonds are drawn as stippled lines. Non-carbon atoms are drawn with larger radii than carbon atoms. Oxygen atoms are drawn as two concentric circles, and nitrogen, sulfur and phosphorus atoms are drawn as three concentric circles. (a) Subunit A; (b) subunit B.

by the side chains of Asn-130, Arg-161, His-186, and Ser-241 and by the nicotinamide ring of NAD as shown in Figure 9a. Gly-A230 is close to the sulfate ion, and any other residue type at this location would interfere with sulfate or substrate binding. The four sulfur oxygens are all hydrogen bonded either to protein side-chain atoms or to other solvent molecules, and the positive charge on Arg-161 should neutralize one of the negative charges on the sulfate ion. The environment of this sulfate ion is shown in the stereo diagram in Figure 9a.

In subunit B, the exact nature of the sulfate environment is less clear because of poorly defined electron density. Furthermore, there is a possibility that in subunit B the sulfate ion may occupy more than one position. It is somewhat surprising that in its predominant position the sulfate ion does not interact directly with the guanidinium group of Arg-B161. Instead, solvent molecules seem to be mediating hydrogen bonds between the sulfate and Arg-B161 although convincing electron density is only seen for one solvent molecule here, Wat-299. The other hydrogen bonds formed between the sulfate oxygens and cMDH subunit B are similar to those observed in subunit A.

If the sulfate binding sites in the two subunits are superimposed, it is observed that Asn-130, Asp-152, His-186, and part of Arg-161 all align quite well, whereas Ser-241 and Gly-230 and the guanidinium group of Arg-161 deviate by about 1.5–2.0 Å from their counterparts in the other subunit. It is likely that the difference between the subunits has its basis in the lack of exact 2-fold molecular symmetry relating the cMDH subunits. Recall that residues 161, 230, and 241 all are in or near the dimer interface and that residues 230 and 241 are among those that show large differences when the two subunits are compared; see Figures 4g and 8. Thus the asymmetry in the subunit interface is transmitted into the sulfate binding or substrate binding site. It should be recalled that Arg-161, in addition to interacting with the bound anion, also interacts with the carboxylate group of Asp-58 in the other subunit, and this interaction differs in the subunit A to subunit B interaction from that in the subunit B to subunit A interaction, as can be observed in Figure 9.

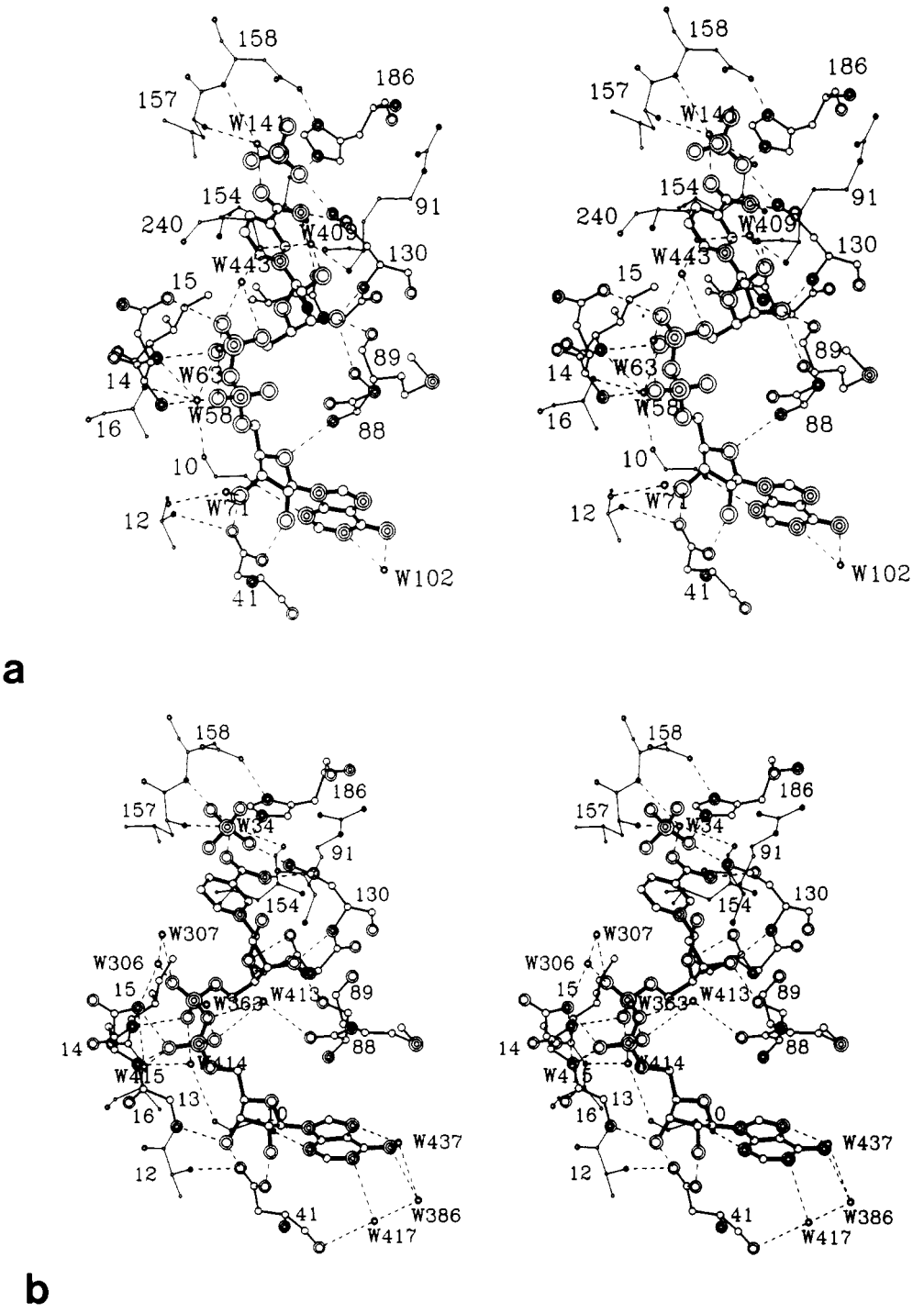

FIGURE 10: Stereo diagram showing the environment of the bound NAD. All residues are identified by their number. The NAD and the sulfate ion are shown with the largest atomic radii and chemical bonds. Residues drawn with the thinnest bonds are not hydrogen bonded directly to NAD. Rather, a solvent molecule forms hydrogen bonds to both the residues and NAD. Different atom types, solvent molecules, and hydrogen bonds are shown as in Figure 9. (a) Subunit A; (b) subunit B.

The remaining 471 solvent molecules have all been treated as if they were water molecules. The vast majority, 317 waters, interact directly with the protein or indirectly through other water molecules that themselves are bound to the protein. The latter category includes 61 waters. Most of the remaining 93 water molecules are located within 6 Å of the protein surface. About 25 waters are located entirely inside the protein structure; that is, they are not in direct contact with the bulk solvent. As mentioned earlier, waters are located in the subunit interface with 18 waters forming hydrogen-bond bridges between the subunits. These last two groups of solvent molecules should be considered integral parts of the dimeric structure.

To estimate the relative amounts of bound and free water, we assumed that 50% of the volume of the cMDH lattice is occupied by solvent. Using a volume for each water molecule of 30 Å^3 (Richards, 1974), we can then calculate that each asymmetric unit of the cMDH crystals contains about 2950 solvent molecules. For cMDH, the 471 water molecules identified at 2.5-Å resolution represent about 15% of the amount present in the crystalline protein.

NAD Binding and Conformation. The cofactor NAD, when bound to cMDH as shown in Figure 10, assumes the extended conformation that is typically observed when NAD is bound to an enzyme (Webb et al., 1973). Thus the adenine and nicotinamide rings are roughly perpendicular to each other with a distance between the centers of gravity of the two rings of 13.9 and 14.1 Å in subunits A and B, respectively. All ribose rings have identical E_2, $C_{(2')}$-endo puckering, and the

Table III: Torsional Angles (deg) for NAD^+ Bound to cMDH and Other Dehydrogenases[a]

protein	χ_a	γ_a	β_a	α_a	ζ_a	ζ_n	α_n	β_n	γ_n	χ_n
cMDH A	−103	−24	161	49	−148	79	76	167	20	−96
cMDH B	−89	−64	150	63	95	−164	78	−164	39	−92
LADH A	−96	−79	147	106	85	−153	59	−146	39	−102
LADH B	−114	−99	136	104	83	−151	76	−169	42	−104
GAPDH A	−103	−70	146	81	83	−161	67	165	67	77
LDH	−108	−78	−138	−76	115	−148	63	−179	−58	−106

[a] The nomenclature for the torsion angles in NAD is that adopted for alcohol dehydrogenase (LADH); see Figure 2 in Eklund et al. (1984). The values for LADH are from Eklund et al. (1984), those for one subunit of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) are from Skarzynski et al. (1987), and those for lactate dehydrogenase (LDH) are a personal communication from Rossmann. The letter following the protein abbreviation indicates different subunits.

conformations around the four glycosidic bonds are all anti.

The NAD dihedral angles calculated for each subunit are listed in Table III together with values for other NAD-dependent dehydrogenases. Overall, the values for the two cMDH subunits are very similar, the deviations being well within the error margin for the coordinates. Only the ζ_a[5] and ζ_n angles differ significantly. This is caused by a difference in position of the phosphodiester O3 oxygens in the two NAD molecules. Furthermore, as shown in Table III, the torsional values calculated for NAD bound to cMDH are remarkably similar to those for NADH bound to alcohol dehydrogenase and for NAD bound to glyceraldehyde-3-phosphate dehydrogenase and lactate dehydrogenase. This is particularly true for NAD bound to subunit B of cMDH.

The NAD molecules are contained entirely within the subunits to which they are bound. In other words, no interactions between the NAD and the adjacent subunit are observed. The contacts between NAD and cMDH are evenly distributed between hydrogen bonds and nonpolar interactions. The two aromatic rings make extensive nonpolar contacts with side chains in the enzyme, whereas the middle parts of the coenzyme interact with the protein nearly exclusively through hydrogen bonds which are visible in Figure 10. A number of these hydrogen bonds are mediated through solvent molecules rather than directly between enzyme and NAD, and these are also given in Figure 10. Although the NAD–cMDH interactions in the two subunits are similar, some differences are apparent, especially for those involving solvent molecules.

The adenine ring is located in a crevice that is formed by the side chains of Leu-40, Ile-42, and Met-45, by the hydrophobic part of Asp-41, and by the main chains and side chains of Thr-9, Gly-10, Ala-12, and Gly-13. Hydrogen bonds in subunit A are formed between N1A and Wat-102, between N7A and Wat-437, and between N6A and Wat-102. These water molecules are in turn in contact with the bulk solvent. In subunit B, N6A is hydrogen bonded to OE1 Glu-B73 of another molecule related by crystal symmetry. The other adenine ring nitrogens in NAD do not appear to be hydrogen bonded.

The two hydroxyl groups in the adenine ribose are hydrogen bonded to the carboxylate oxygen of Asp-41. An identical hydrogen-bonding pattern, including an aspartate interacting with the adenine ribose hydroxyl groups, has been described in most other dinucleotide binding oxidoreductases (Birktoft & Banaszak, 1984; Rossmann et al., 1975; Skarzynski et al., 1987). As is visible in Figure 10, the oxygen atoms in the pyrophosphate participate in numerous hydrogen bonds, either directly with cMDH or with solvent molecules, some of which in turn are hydrogen bonded to the enzyme. It is noticeable that there are no positively charged side chains in the immediate vicinity of the negatively charged pyrophosphate moiety. A similar lack of compensating charges near pyrophosphate groups has been described in glyceraldehyde-3-phosphate dehydrogenase (Skarzynski et al., 1987) and in glutathione reductase (Karplus & Schulz, 1987). Hydroxyl group O3N in the nicotinamide ribose is hydrogen bonded to the amido group of Asn-130. The side chain of this residue also interacts with the bound sulfate ion.

The nicotinamide group of NAD is located toward the center of the subunit in a pocket that, in addition to the cofactor, also must accommodate the malate/oxaloacetate substrate. One side of the nicotinamide ring is resting in a hydrophobic crevice circled by the side chains of residues Ile-15, Val-128, Asn-130, Leu-152, Leu-159, and Ala-245 and atoms in the main chains belonging to Val-128 and Gly-129. The other side of the pyridine ring faces the internal cavity that accommodates the substrate.

The refinement was performed without restraining the carboxamide group to be coplanar with the nicotinamide ring. In the final coordinates, the carboxamide oxygen and nitrogen atoms are about 20° out of the plane of the nicotinamide ring, although such deviations are barely significant. Both the oxygen and nitrogen atoms in the carboxamide group are involved in hydrogen bonds with the enzyme. Incidentally, the hydrogen-bonding pattern around the carboxamide group, as shown in Figure 10, allows a plausible identification of the atoms in the carboxamide group. cMDH is a so-called A-side-specific dehydrogenase, and it is the interactions between the carboxamide group and the enzyme that to a large extent determine the conformation around the nicotinamide glycosidic bond, causing the A-side of the pyrimidine ring to be facing the substrate. In both subunits, the oxygen of the carboxamide group, O1N, is hydrogen bonded to a buried solvent molecule, Wat-141 and Wat-34 in subunits A and B, respectively. These water molecules are in turn hydrogen bonded to three protein atoms, O Leu-154, N Leu-157, and N Asp-158, all visible in Figure 10.

Water molecules Wat-141 and Wat-34 are located at the N-terminal end of helix α2F. One might speculate that the positive charge generated by the the α-helical dipole and located at the N-terminal end of α-helix α2F might interact with the positively charged nicotinamide ring in NAD. This electrostatic interaction might explain why the reduced form of the coenzyme, NADH, containing a neutral nicotinamide ring is bound more tightly than NAD. The nitrogen of the carboxamide group, N2N, is hydrogen bonded to OD1 Asn-130 although the N–O distance is a bit long in subunit B.

Active Site. The active site of cMDH is located at the bottom of the internal cavity that must accommodate the cofactor NAD (NADH) and the malate/oxaloacetate substrate, in addition to the catalytic residues. Three active site residues were tentatively identified before the complete amino acid sequence was known (Birktoft & Banaszak, 1983), namely, an aspartate (Asp-158), an arginine (Arg-161), and a histidine (His-186). It was suggested then that the aspartate

[5] The ζ_a angle is defined by atoms O5′A, PA, OP3, and PN, and the ζ_n angle is defined by atoms PA, OP3, PN, and O5′N. The nomenclature for atoms and the other dihedral angles in NAD is the same as used for alcohol dehydrogenase. See Figure 2 in Eklund et al. (1984).

Table IV: Residues near Heavy Atom Binding Sites in cMDH[a]

heavy atom binding site	residues within 4.5 Å of heavy atom binding site	site related by molecular symmetry
Hg-1	Ser-A27, Val-A28, Cys-A250, Asp-A251, Arg-A254	Hg-2
Hg-2	Ser-B27, Cys-B250, Asp-B251, Arg-B254	Hg-1
Hg-3	Tyr-A21, Cys-A59, Leu-A61, Lys-B247	none
Pt-1	Glu-A73, Cys-A110	Pt-2
Pt-2	Glu-B73, Cys-B110	Pt-1
Pt-3	Met-A45	Pt-6
Pt-4	Asp-B50, Met-B54	Pt-5
Pt-5	Asp-A50, Met-A54	Pt-4
Pt-6	Ile-B42, Met-B45, NAD-B336, Gln-B202*	Pt-3
Pt-7	Met-B89, Lys-B98	Pt-8
Pt-8	Met-A89, Arg-A91, Arg-A92	Pt-7
UO2-1	Asp-A79, Glu-B324*	UO2-7
UO2-2	Gly-A94, Arg-A97, Lys-A98, Asp-A215*	none
UO2-3	Asp-A282, Glu-A328*	none
UO2-4	Arg-A306, Asp-A309	UO2-5
UO2-5	Asp-B306, Lys-A102*	UO2-4
UO2-6	Arg-A91, Gly-A94, Met-A95, Glu-A96, Arg-A97	none
UO2-7	Glu-A324, Glu-A324*	UO2-1

[a] Residues labeled with an (*) are located in another molecule related by crystal lattice symmetry.

and the histidine play a central role in the catalytic process, being involved in proton transfers critical to catalysis. The arginine side chain was suggested to be involved in binding the malate/oxaloacetate substrate.

In the refined model, the side chain of Asp-158 is hydrogen bonded to the imidazole ring of His-185. In subunit A, the OD2 Asp-158 to ND1 His-186 distance is 2.5 Å, and the OD2 atom is 0.3 Å out of the plane of the imidazole ring. In subunit B, the hydrogen-bond distance is 2.7 Å, and OD2 Asp-158 is 1.5 Å out of the plane of the imidazole ring. The aspartate carboxylate is shielded from solvent contact by the imidazole side chain and the side chains of residues Thr-155, Leu-157, Ile-182, and Gln-190. In subunit A the OD1 atom in Asp-158 is hydrogen bonded to a solvent molecule, Wat-57, which in turn is hydrogen bonded to NH1 Arg-161. In subunit B, a corresponding solvent molecule is not seen although there seems to be room enough near OD1 Asp-158. The edge of the imidazole ring opposite Asp158 is facing the bound sulfate anion and the nicotinamide ring of the NAD as discussed previously and presented in Figure 9. The refined structure therefore appears to agree with earlier observations that His-186–Asp-158 could serve as a shuttle for the removal of a proton from the –OH group of malate.

Heavy Atom Binding Sites. The method of multiple isomorphous replacement for phase determination in protein crystallography requires the preparation of heavy atom derivatives of the native protein crystals. In the case of cMDH, three different compounds containing heavy atoms were found to be suitable for this purpose (Hill et al., 1972). Each asymmetric unit containing a dimeric cMDH molecule binds *p*-(hydroxymercuri)benzenesulfonic acid at three sites, platinum ethylenediamine dichloride at eight sites, and uranoyl dioxygenate at seven sites.

The binding of heavy atom containing compounds frequently, but not always, results in conformational changes in the protein structure. If the effects are major, the resulting heavy atom derivative may not be useful for phase determination. The three derivatives used in the high-resolution phase determination of cMDH were all isomorphous with the native protein. While the existence of major conformational changes thus can be excluded, this does not eliminate the possibility of minor structural changes due to heavy atom binding. To examine the various heavy atom binding sites of cMDH, the local environment was defined as those residues that have at least one atom in the refined structure within 4.5 Å of each of the heavy atom positions. The results are given in Table IV.

Previously, the binding of the platinum derivative to cMDH was analyzed in some detail (Wade et al., 1973), without knowledge of the chemical sequence. At that stage, four platinum binding sites were associated with each subunit, and further, each binding site in one subunit was related by the pseudo-2-fold molecular symmetry to a binding site in the other subunit. This preliminary analysis has now been confirmed. Inspection of Table IV reveals that not only are the platinum binding sites related in a pairwise fashion but that two of the three mercury binding sites and four of the seven uranyl binding sites can be grouped in pairs as well. One mercury and three uranyl binding sites are unique in that these sites are only observed in one subunit.

The three mercury binding sites listed in Table IV are all located on the subunit surface near cysteine residues. The distance from the mercury positions to the cysteine sulfurs are somewhat longer, 3.4-4.8 Å, than the normal Hg–S distance of 2.5 Å. However, simple rotations about the Cα–Cβ bonds can bring the sulfur atoms to a position where covalent Hg–S bonds can be formed. When pig heart cMDH is reacted in solution with sulfhydryl reagents, two cysteine per molecule are initially modified, and only after prolonged treatment, or following prior incubation in 4 M urea, will additional cysteines react (Banaszak et al., 1971). In the crystalline state, Cys-250 participates in mercurial binding which obeys the local symmetry. Calculations show that it is accessible in the dimer.

However, the same calculations also show that the most exposed cysteine is Cys-110, but interestingly, this residue does not show any reaction with the mercurials in the crystalline state of cMDH. This lack of reactivity could have several possible explanations. Crystal lattice packing effects may block the access route for a mercurial to Cys-110. Alternatively, the reactivity of the sulfur atom could be altered by local steric hindrance, or local electrostatic effects which might impair binding and/or reactivity with the mercurial. We believe the latter to be the principle factor. At least four carboxyl groups are located within 12.5 Å of Cys-110, and when cMDH is incorporated into the crystal lattice, additional negative charges are brought into the vicinity of this residue. In total, there are eight and five negative charges within 12.5 Å of Cys-110 in subunits A and B, respectively. This accumulation of negative charges could possibly be the factor effecting the binding of *p*-(hydroxymercuri)benzenesulfonic acid, which itself is negatively charged.

The third unpaired mercurial site probably is related to Cys-59, which is exposed to solvent in the isolated monomers but, upon dimerization, becomes totally inaccessible. As discussed previously, some asymmetry is observed in the subunit interface, and Cys-59 is located in nonequivalent molecular environments in the two subunits. Although some minor adjustment of side chains near Cys-A59 in subunit A would be necessary for a mercurial to react with this residue, this apparently occurs. In subunit B, the side chain of Cys-B59 is located inside a cavity created by the side chains of Tyr-B21, Leu-B61, and Lys-A277 and the main chain of residues A243–A248 of the adjacent subunit. However, it appears that the necessary conformational changes needed for a mercurial to gain access to Cys-B59 cannot be permitted within the crystalline lattice.

All three mercury binding sites are removed from the active site, which is in agreement with the observation that cyste-

ine-directed reagents do not cause inactivation of cMDH (Banaszak et al., 1971). Cysteine residues 136 and 153 are both near the binding cavity for NAD, but both of these are inaccessible to solvent and thus also to the larger cysteine reagents. In order for these residues to react, partial unfolding of the subunit would be required, and an inactive enzyme would certainly result.

Curiously, Cys-136 and Cys-153 are located adjacent to each other in the cMDH structure. The distances between the pairs of Cα atoms are 6.3 and 6.0 Å in subunits A and B, respectively, which is close to the average Cα–Cα distance of 5.88 ± 0.49 Å observed for a number of disulfides in proteins (Katz & Kossiakoff, 1986). The distances between the pairs of Cβ atoms are 4.0 and 3.8 Å, respectively, for subunits A and B, and by simple rotations about the Cα–Cβ bonds, the two Sγ atoms can be brought to within 1.3 Å of each other. This distance is 0.7 Å shorter than the –S–S– distance observed in disulfides. To obtain the correct angular relationship for an internal disulfide, only minor readjustment of the adjacent protein structure would be necessary. Experimentally, reducing conditions were maintained throughout the purification and crystallization of cMDH, so a disulfide is chemically unlikely in the crystalline specimens of cMDH.

As was reported by Wade et al. (1973), the eight platinum binding sites can be divided into four pairs, all of which are related by the molecular symmetry. As is frequently observed in other proteins (Blundell & Johnson, 1976), all platinum binding sites in cMDH are near sulphur-containing side chains. Platinum sites 1 and 2 are near Cys-110, and the remainder are near methionine side chains. Wade et al. (1973) have discussed the relationship between platinum binding sites 7 and 8 and the inhibitory activity of platinum compounds. Each of these binding sites are close to the side chain of Met-89, one in each subunit. This residue is part of the so-called loop region and is near the bound NAD in the holo form of the enzyme. The importance of residue Met-89 is also demonstrated by chemical modification studies where reaction with iodoacetate results in a specific modification of a methionine residue and inactivation of the enzyme (Leskovac & Pfleiderer, 1969; Leskovac, 1971). Leskovac (1971) isolated a peptide from a chymotryptic digest of the modified enzyme. The reported amino acid composition of this peptide identifies the labeled methionine as Met-89.

Platinum binding sites 3 and 6 are near the side chain of Met-45 in subunits A and B, respectively. This residue is close to the adenine ring of bound NAD, and platinum binding to this residue could also have some influence on NAD binding and, consequently, on the enzymatic activity. Binding sites 4 and 5 are close to side chain of Met-54 in the respective subunits. This residue is located in the subunit interface, and both sulfur atoms are partially inaccessible. Reactions with this methionine could possibly influence the stability of the dimeric form of the enzyme.

All seven uranyl binding sites listed in Table IV are located on the surface of the molecular dimer, and all are near the carboxyl groups of either aspartic or glutamic acid residues. Four of the uranyl binding sites can be divided into two pairs of binding sites, with each member of a pair being related by the molecular symmetry. Site 3 is in an intermolecular location, which in subunit B is blocked by crystal lattice contacts. Sites 2 and 6 are only 6 Å apart and are located in the loop region in subunit A. This part of the structure assumes different conformations in the two subunits.

The symmetry relating the heavy atom binding sites could have been of help in establishing the molecular symmetry during the early stages of the structural analysis of cytoplasmic malate dehydrogenase. The rotation function analysis (Hill et al., 1973) was difficult to interpret, since the molecular "2-fold" symmetry axis is nearly parallel to the crystallographic *b* axis. In retrospect, the lack of exact 2-fold rotational symmetry in cMDH discussed earlier would have made the use of the molecular symmetry in phase determination (Bricogne, 1976) quite difficult, if not useless, because of the temptation to assume a 2-fold symmetric dimer. This departure from 2-fold symmetry would not have been recognized from analysis of either the heavy atom binding sites or the rotation function.

Discussion

The structure of cytoplasmic malate dehydrogenase discussed here is based on X-ray diffraction data measured to a nominal resolution of 2.5 Å. Restrained crystallographic refinement has yielded a molecular model that, for diffraction data in the 6.0–2.5-Å range, corresponds to a crystallographic *R* factor of 16.7%. The rms deviation for bond length is 0.017 Å, with the largest deviations being 0.10 Å from the values normally observed in precise, high-resolution crystal analysis. This is, however, better than the level of uncertainty that might be expected for structural analysis at this resolution (Luzzatti, 1952), which is shown in Figure 2 and is estimated to be 0.22 Å.

Another consideration at the present stage of refinement is the incomplete treatment given the solvent structure. In the final stages of the refinement procedure 471 solvent molecules were included, a number that represents some 9% of the total number of protein atoms. The 471 solvent molecules correspond to about 15% of the calculated theoretical number of solvent molecules in the asymmetric unit. Although this is only a crude estimate, it is obvious that a vast majority of the solvent structure is unaccounted for, and probably disordered.

The presence of additional solvent molecules is suggested in a number of locations. Specifically, these suggested water molecules would be positioned such that they might mediate hydrogen bonds between protein atoms. In several locations on the protein, potential hydrogen-bond donors and acceptors are poised for hydrogen bonding, but their interatomic distance is too long. If a solvent molecule was placed between such atoms, satisfactory hydrogen bonds could be formed. Such occurrences are typically seen in locations removed from the bulk solvent, and energetic considerations would suggest that a solvent molecule should be placed here.

As already noted, the two chemically identical subunits are related by only approximate 2-fold rotational symmetry. While this is a departure from what is expected for a dimeric protein comprised of two identical subunits, it is by no means without precedent. Insulin, hexokinase, and α-chymotrypsin are all well-documented instances of a dimeric protein displaying departure from exact 2-fold rotational symmetry [reviewed by Banaszak et al. (1981)]. Since biochemical data on cytoplasmic malate dehydrogenase obtained in solution do not suggest any type of molecular asymmetry, the differences in the X-ray study must have their basis in packing effects within the crystal lattice. This notion is further supported by the observation that those segments of the polypeptide chain that differ the most in the two subunits are without exception involved in different intermolecular contacts in the crystal lattice. The more extensive contacts observed near the coenzyme binding site in subunit A as compared to those in subunit B do not, by themselves, explain the increased NAD affinity by subunit A. It does, however, give a rationale for the observation that once NAD is bound to subunit A it cannot be removed without disruption of the crystal lattice. Reports describing the enzymatic activity in the crystalline state also

support the notion of nonequivalence of the subunits (Zimmerle & Alter, 1983).

The structure of each subunit has, by convention, been divided into two domains, a division that has both a structural and a functional rationale. However, many double-domain structures are separated by a noticeable cleft that often serves as a binding site for substrates and cofactors (Anderson et al., 1979). In cMDH, no such interdomain cleft is readily visible, and as previously pointed out, the elements of secondary structure in the two domains conform to a continuum across the domain interface. This is most noticeable in Figure 6 where the 11-stranded mixed β-sheet extends through the middle of an entire subunit. Depending on one's point of view, the structure of this enzyme and of other related 2-hydroxyacid dehydrogenases can then be considered as comprised of either an extended β-sheet surrounded by helices on either side or a double-domain structure. The latter is, of course, of more evolutionary purpose and is the description commonly used for this group of enzymes (Rossmann et al., 1975).

Earlier proposals about the active site of cMDH, particularly residues involved in substrate binding and the presence of a histidine–aspartate pair, have been reinforced by the incorporation of the know chemical sequence, crystallographic refinement, and the resultant molecular model. Unfortunately, the X-ray results are complicated by the lack of exact 2-fold symmetry in the cMDH dimer. Ample evidence indicates that the chemical properties of the two monomeric units in a solution dimer are identical. Furthermore, the refined crystal structure supports the idea that the conformational differences between the two subunits are the result of crystal packing phenomena. Although these structural dissimilarities are small, they do cause small differences in the active site. At the present time, there is no way of selecting subunit A or B as the structure closest to the solution form of cMDH, and further work on the catalytic mechanism will have to consider the details of both conformers of the cMDH subunit.

Acknowledgments

We thank Dr. Ralph Bradshaw for support and interest throughout all phases of this project and him and his coworkers for the amino acid sequence data that was vital to completion of this study. We gratefully acknowledge the excellent assistance of Suzanne Winkler and Sophie Silverman in the preparation of the manuscript. We are especially grateful to fellow crystallographers Paul Bethge, Dale Tronrud, Arthur Lesk, and Christian Cambillau, who by their generous contributions of highly developed software have enormously aided the studies described in this paper.

References

Anderson, C. M., Zucker, F. H., & Steitz, T. A. (1979) *Science 204*, 375–380.

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

Banaszak, L. J., & Webb, L. E. (1975) in *Structure and Conformation of Nucleic Acids and Protein-Nucleic Acid Interactions* (Sundaralingan, M., & Rao, S. T., Eds.) pp 375–386, University Park Press, Baltimore.

Banaszak, L. J., Tsernoglou, D., & Wade, M. J. (1971) in *Probes of Structure and Function of Macromolecules and Membranes* (Chance, B., Yonetani, T., & Mildvan, A., Eds.) Vol. 2, pp 71–80, Academic Press, New York.

Banaszak, L. J., Birktoft, J. J., & Barry, C. D. (1981) in *Protein-Protein Interactions* (Frieden, C., & Nichol, L. W., Eds.) pp 31–128, Wiley, New York.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535–542.

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

Birktoft, J. J., & Banaszak, L. J. (1984) *Pept. Protein Rev. 4*, 1–46.

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

Birktoft, J. J., Fernley, R. T., Bradshaw, R. A., & Banaszak, L. J. (1982b) in *Molecular Structure and Biological Activity* (Griffin, J. F., & Duax, W. L., Eds.) pp 37–55, Elsevier/North-Holland, New York.

Birktoft, J. J., Bradshaw, R. A., & Banaszak, L. J. (1987) *Biochemistry 26*, 2722–2734.

Birktoft, J. J., Fu, Z., Carnahan, G. E., Rhodes, G., Roderick, S. L., & Banaszak, L. J. (1989) *Biochem. Soc. Trans. 17*, 301–304.

Blundell, T. L., & Johnson, L. N. (1976) in *Protein Crystallography*, Academic Press, Orlando, FL.

Bolin, J., Filman, D., Matthews, D., Hamlin, R., & Kraut, J. (1982) *J. Biol. Chem. 257*, 13650–13662.

Bricogne, G. (1976) *Acta Crystallogr. A32*, 832–847.

Buehner, M., Ford, G. C., Moras, D., Olsen, K. W., & Rossmann, M. G. (1974) *J. Mol. Biol. 90*, 25–49.

Chien, S. M., & Freeman, K. B. (1984) *J. Biol. Chem. 259*, 3337–3343.

Chothia, C., & Janin, J. (1975) *Nature 256*, 705–708.

Chothia, C., Levitt, M., & Richardson, D. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4130–4134.

Crick, F. H. C. (1953) *Acta Crystallogr. 6*, 689–697.

Drenth, J., Jansonius, J. N., Koekoek, R., & Wolthers, B. G. (1971) *Adv. Protein Chem. 25*, 79–115.

Eklund, H., Samana, J.-P., Wallen, L., Branden, C.-I., Akeson, A., & Jones, T. A. (1981) *J. Mol. Biol. 146*, 561–587.

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

Frieden, C., & Fernandez-Souza, J. (1975) *J. Biol. Chem. 250*, 2106–2113.

Glatthaar, B. E., Banaszak, L. J., & Bradshaw, R. A. (1972) *Biochem. Biophys. Res. Commun. 46*, 757–765.

Grant, P. M., Tellam, J., May, V. L., & Strauss, A. W. (1986) *Nucleic Acids Res. 14*, 6053–6066.

Hendrickson, W. A., & Konnert, J. H. (1980) in *Biomolecular Structure, Function, Conformation and Evolution* (Srinivasan, R., Ed.) Vol. 1, pp 43–57, Pergamon, Oxford.

Hill, E., Tsernoglou, D., Webb, L., & Banaszak, L. (1972) *J. Mol. Biol. 72*, 577–591.

Hill, E. J., Tsernoglou, D., & Banaszak, L. J. (1973) *Acta Crystalllogr. B29*, 921–922.

Janin, J., & Chothia, C. (1978) *Biochemistry 17*, 2943–2948.

Joh, T., Takeshima, H., Tsuzuki, T., Setoyama, C., Shimada, K., Tanase, S., Kuramitsu, S., Kagamiyama, H., & Morino, Y. (1987a) *J. Biol. Chem. 262*, 15127–15131.

Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268–272.

Karplus, P. A., & Schulz, G. E. (1987) *J. Mol. Biol. 195*, 701–729.

Katz, B., & Kossiakoff, A. (1986) *J. Biol. Chem. 261*, 15480–15485.

Lee, B. K., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379–400.

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

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

Leskovac, V. (1971) *Croat. Chem. Acta 43*, 183–186.
Leskovac, V., & Pfleiderer, G. (1969) *Hoppe-Seyler's Z. Physiol. Chem. 350*, 484–492.
Luzzatti, V. (1952) *Acta Crystallogr. 5*, 808–810.
Matthews, B. W., & Bernhard, S. A. (1973) *Annu. Rev. Biophys. Bioeng. 2*, 257–317.
McLachlan, A. D. (1979) *J. Mol. Biol. 128*, 49–79.
Ploegman, J. H., Drenth, J., Kalk, K. H., & Hol, W. G. J. (1978) *J. Mol. Biol. 123*, 557–594.
Rao, S. T., & Rossmann, M. G. (1973) *J. Mol. Biol. 76*, 241–256.
Remington, S. J., Wiegand, G., & Huber, R. (1982) *J. Mol. Biol. 158*, 111–152.
Richards, F. M. (1974) *J. Mol. Biol. 82*, 1–14.
Richardson, J. S. (1977) *Nature 268*, 495–500.
Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167–339.
Richmond, T. J., & Richards, F. M. (1978) *J. Mol. Biol. 119*, 537–555.
Roderick, S. L., & Banaszak, L. J. (1986) *J. Biol. Chem. 261*, 9461–9464.
Rossmann, M. G., Liljas, A., Branden, C.-I., & Banaszak, L. J. (1975) *Enzymes (3rd Ed.) 11A*, 61–102.
Rudolph, R., Fuchs, I., & Jaenicke, R. (1986) *Biochemistry 25*, 1662–1669.
Sacchettini, J. C., Gordon, J. I., & Banaszak, L. J. (1988) *J. Biol. Chem. 263*, 5815–5819.
Schirmer, T., Huber, R., Schneider, M., Bode, W., Miller, M., & Hackert, M. L. (1986) *J. Mol. Biol. 188*, 651–676.
Setoyama, C., Joh, T., Tsuzuki, T., & Shimada, K. (1988) *J. Mol. Biol. 202*, 355–364.
Sim, G. A. (1960) *Acta Crystallogr. 12*, 511–512.
Skarzynski, T., Moody, P. C. E., & Wonacott, A. J. (1987) *J. Mol. Biol. 193*, 171–187.
Tronrud, D., TenEyck, L., & Matthews, B. W. (1987) *Acta Crystallogr. 43A*, 489–501.
Wade, M., Tsernoglou, D., Hill, E., Webb, L., & Banaszak, L. J. (1973) *Biochim. Biophys. Acta 322*, 124–132.
Webb, L. E., Hill, E. J., & Banaszak, L. J. (1973) *Biochemistry 12*, 5101–5108.
Weininger, M., Birktoft, J. J., & Banaszak, L. J. (1977) in *Pyridine Nucleotide Dependent Dehydrogenases* (Sund, H., Ed.) pp 87–100, de Gruyter, Berlin.
Zimmerle, C., & Alter, G. (1983) *Biochemistry 22*, 6273–6281.
Zimmerle, C. T., Yung, P. P., & Alter, G. M. (1987) *Biochemistry 26*, 8535–8541.

Geometry of Interaction of Metal Ions with Sulfur-Containing Ligands in Protein Structures†

P. Chakrabarti

Department of Chemistry and Biochemistry, Molecular Biology Institute, University of California, Los Angeles, California 90024

Received January 4, 1989; Revised Manuscript Received March 31, 1989

ABSTRACT: An analysis of the geometry of binding of metal ions by cysteine and methionine residues in protein structures has been made by using the Protein Data Bank. Metal ions have a distinct mode of binding to each of these residues, and this is independent of the nature of the metal center or the type of protein. Metal ions tend to approach the sulfur of Met roughly 38° from the perpendicular to the plane through atoms C_γ–S_δ–C_ϵ. For Cys, the approach direction is such that the M···S_γ–C_β–C_α torsional angle is about ±90 or 180°. The side-chain conformation of the cysteine residue is affected by the presence of the metal ion; there is a shift from the g^+ conformation toward g^- and mainly t conformations. When two Cys residues at positions $i - 3$ and i bind to the same metal center, there appears to be some restriction on the geometry of metal binding by the residue i; for such a residue χ_1 and M···S_γ–C_β–C_α angles are likely to be around 60° and 270°, respectively. Met and Cys residues coordinating to a metal ion are usually from coil or turn regions of the protein structure.

Many metalloproteins, especially those involved in electron transfer, have one or more sulfur ligands (Cys, Met) bound to metal ions. Many of these proteins have characteristic absorption spectra caused by S → Fe or S → Cu charge transfer. A knowledge of the geometry of the interaction of metal ions with sulfur ligands and their molecular orbitals is essential in understanding electronic structures and oxidation–reduction properties of these metal sites (Gewirth & Solomon, 1988; Noodleman et al., 1985). Although several studies have been carried out on directional preferences of various ligands toward metal ions in small molecule structures (Rosenfield et al., 1977; Chakrabarti & Dunitz, 1982; Einspahr & Bugg, 1981), a comparable analysis with macromolecular structures has not been reported. In this paper we present an analysis of the interaction of metal ions with Cys and Met ligands in protein structures.

Materials and Methods

The analysis is based on atomic coordinates from the Brookhaven Protein Data Bank, (PDB) (Bernstein et al., 1977). All the structures studied had been subjected to some sort of refinement. To compare the results with those of small-molecule structures, the Cambridge Crystallographic Database (CCDB) (Allen et al., 1979) was searched for metal···cysteine interactions where cysteine acts as an unidentate ligand binding metal ions only through the sulfur atom, as in the case of protein structures.

(*a*) *Met.* The relative position of the metal ion was expressed in terms of spherical polar coordinates (Figure 1): the M···S_δ distance; the angle θ between the S_δ···M direction and

† This work was supported by USPHS Grant GM 31299.

0006-2960/89/0428-6081$01.50/0