The Active Site Architecture of a Short-Chain Dehydrogenase Defined by Site-Directed Mutagenesis and Structure Modeling†

Lluís Ribas de Pouplana*,‡ and Linda A. Fothergill-Gilmore

Department of Biochemistry, University of Edinburgh, George Square, Edinburgh EH8 9XD, Scotland

Received January 26, 1994; Revised Manuscript Received April 13, 1994®

ABSTRACT: A high-resolution crystal structure is not currently available for *Drosophila* alcohol dehydrogenase. A detailed three-dimensional model for this enzyme, based on the structure of 3α,20β-hydroxysteroid dehydrogenase, has been generated by extensive computer modeling studies. Aspects of the model concerned with coenzyme binding have been tested by site-directed mutagenesis of residues Gly-14 to Ala, Gly-19 to Ala, Asp-38 to Ala, and Pro-214 to Ser. All enzymes have been characterized in terms of kinetic constants, relative stabilities to guanidinium chloride, and heat inactivation. The contribution of NAD binding to the stabilization of each of the enzymes was also measured. The results obtained with enzymes mutated at positions 14, 38, and 214 are in accordance with published data on *Drosophila* alcohol dehydrogenase and suggest interactions of these residues with the cofactor NAD. The introduction of a methyl group at residue Gly-19 abolished the ability of the enzyme to utilize NADP instead of NAD. This reflects a proximity of residue Gly-19 to the ribose ring of the bound cofactor. This result, coupled to the three-dimensional model built for *Drosophila* alcohol dehydrogenase, suggests a binding mechanism for the cofactor NAD different from that found for 3α,20β-dehydroxysteroid dehydrogenase and similar to that found in the crystal structure of rat liver dihydropteridine reductase. The model of *Drosophila* alcohol dehydrogenase also enables many previous observations from chemical modification, sequence comparisons, site-directed mutagenesis, and limited proteolysis experiments to be placed into a structural context. An active site architecture is proposed involving a loop closure mechanism similar to that of lactate dehydrogenase. It is envisaged that the initial binding of NAD in the ordered mechanism instigates a C-terminal loop to close over the active site.

Alcohol dehydrogenases (EC 1.1.1.1.–1.1.1.176) are a relatively diverse family of enzymes that catalyze the oxidation of a wide variety of primary and secondary alcohols to aldehydes and ketones. The reactions catalyzed by these enzymes involve the concomitant reduction of NAD or NADP [reviewed by Eklund and Brändén (1987)]. Sequence information has been used to classify alcohol dehydrogenases into three main groups: short chain, medium chain, and long chain (Persson et al., 1991).

More than 20 enzymes have been described that belong to the short-chain group. Of these, typical examples are *Drosophila* alcohol dehydrogenase (ADH)[1] and 3α,20β-hydroxysteroid dehydrogenase (HSD) from *Streptomyces hydrogenans*. Other members of the group include ribitol, sorbitol-6-phosphate, glucose, and hydroxyprostaglandin dehydrogenases. More recently, the class has been expanded because of the discovery of high structure similarity between HSD, dihydropteridine reductase (DHR), and UDP-4-galactose epimerase (UDP) (Krook et al., 1993b; Holm et al., 1994). Despite the high level of structural similarity, the sequence conservation between the class is limited to only five residues. These include three glycine residues in the glycine-rich loop common to Rossmann folds and the sequence YxxxK involved in the active site of the enzymes (Holm et al., 1994). It is apparent that there is considerable diversity of substrate structure among the family.

The short-chain alcohol dehydrogenases from *Drosophila* are dimers of approximately 250 residues per monomer and have a preference for secondary alcohols such as 2-propanol (Winberg et al., 1982). Their reaction follows an ordered mechanism, with the coenzyme binding first (Winberg et al., 1985, 1986; Winberg & McKinley-McKee, 1988). With the preferred substrates, the dissociation of the enzyme–NADH complex is the rate-limiting step, although hydrogen transfer may be rate determining when ethanol is the substrate (Heinstra et al., 1988). As a result, the *Drosophila* ADH isoenzymes that have the strongest coenzyme binding have the lowest specific activities. The two most commonly occurring isoenzymes, ADH "slow" (ADH-S) and ADH "fast" (ADH-F), differ with respect to coenzyme binding. These two isoenzymes have a single amino acid difference at position 192 (Lys in ADH-S and Thr in ADH-F) (Thatcher, 1980), and this has the consequence that ADH-S binds the coenzyme more tightly.

Drosophila ADH has been particularly well characterized in terms of its evolution and genetic and enzymatic properties [reviewed by Chambers (1988)]. Unfortunately, little is currently known of its detailed three-dimensional structure,

† This work was supported by the British Council Acciones Integradas Programme and The Wellcome Trust. Ll.R.P. was the recipient of a British Council/La Caixa Award and a studentship from the University of Edinburgh Faculty of Medicine.

* To whom correspondence should be addressed.

‡ Present address: Department of Biology, Room 68-222, MIT, Cambridge, MA 02139.

® Abstract published in *Advance ACS Abstracts*, June 1, 1994.

[1] Abbreviations: ADH, *Drosophila* alcohol dehydrogenase; ADH-S and ADH-F, "slow" and "fast" alleles of *Drosophila* alcohol dehydrogenase, respectively; ADH-FChD, the Chateau Douglas variant of ADH-F; ADH-W, "Waksman" (Ile-173 to Thr) variant of ADH-S; *Adh*, *Drosophila* ADH gene; *Adh-Waksman Drosophila* ADH-W gene; G14A, G19A, D38A, and P214S, ADH-W mutated at Gly-14 to Ala, Gly-19 to Ala, Asp-38 to Ala, and Pro-214 to Ser, respectively; HSD, 3α,20β-hydroxysteroid dehydrogenase; UDP, UDP-4-galactose epimerase; DHR, dihydropteridine reductase; GndHCl, guanidinium hydrochloride; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; rms, root mean square.

0006-2960/94/0433-7047$04.50/0

although crystals have been obtained and preliminary data concerning crystal dimensions have been reported (Gordon et al., 1992). Many residues have been implicated by chemical modification, limited proteolysis, site-directed mutagenesis, and sequence comparisons as being important for the activity of *Drosophila* ADH. Among these, the invariant residues Gly-14 and Asp-38 have been suggested to be important for coenzyme binding (Chen et al., 1990, 1991). Similarly, Tyr-152 and Lys-156 are conserved in all *Drosophila* ADHs and have been implicated as playing important roles in catalysis. The mutation of these two residues to a variety of other side chains invariably results in enzymes of highly decreased activity (Albalat et al., 1992; Krook et al., 1992; Cols et al., 1993; Chen et al., 1993). Although chemical modification of histidine and cysteine residues leads to loss of activity (Thatcher, 1981), site-directed mutagenesis of the two cysteine residues of ADH indicates that they are not required for catalysis (Chen et al., 1990).

A comparison of thermal stabilities and sequences of natural ADH variants has indicated that residue 214 is important for stability (Chambers, 1991). The substitution of a serine for a proline at this position confers resistance to heat inactivation and also leads to a decrease in specific activity because of tighter coenzyme binding.

Finally, it should be mentioned that limited proteolysis studies have revealed that an accessible peptide bond in the C-terminal half of the enzyme is rendered resistant to proteolysis and, hence, inactivation in the presence of ligands (Krook et al., 1992). Moreover, partial proteolytic cleavage very near the C-terminus at residue 244 also leads to loss of activity, thus implying an important role for the C-terminal 10 residues.

The determinations of the crystal structures of HSD (Ghosh et al., 1991), UDP (Bauer et al., 1992), and DHR (Varughese et al., 1992) are of obvious relevance to an understanding of the structure and activity of *Drosophila* ADH. The crystal structure of HSD has been used to build a three-dimensional model of 15-hydroxyprostaglandin dehydrogenase (Krook et al., 1993a). This work shows that the fold of HSD is probably conserved among the family of short-chain dehydrogenases.

Computer modeling by homology of protein structures is becoming increasingly common and reliable as the database of three-dimensional structures increases. Successful three-dimensional models have now been build for a number of proteins [see Bajorath et al. (1993) and references therein]. However, the results of a modeling exercise will always require experimental confirmation, of either a biochemical or crystallographic nature.

Because of higher sequence similarity between the two enzymes, we have undertaken detailed computer modeling of *Drosophila* ADH on the basis of the structure of HSD, coupled with site-directed mutagenesis experiments. Our results offer a structural framework for the interpretation of the wealth of information on *Drosophila* alcohol dehydrogenase, indicating that *Drosophila* ADH has an active site architecture more similar to that of DHR than to that of the proposed active site of HSD and providing a putative model for the conformation of the active site of the enzyme.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification. ADH-W is a variant of ADH-S that differs in a single amino acid (Ile-173 to Thr in ADH-W). This enzyme is expressed from a chimera gene constructed by William Sofer and colleagues (Clark, 1988) and is indistinguishable from the wild-type enzyme (see results below). The sequence of the gene has been submitted to GeneEMBL under the accession number X67811. Expression and purification of the *Adh* products (Atrian et al., 1990; Ribas de Pouplana & Fothergill-Gilmore, 1993) yielded pure enzyme preparations for the wild-type and mutant enzymes G14A, G19A, and P214S as judged by SDS–PAGE. However, mutant D38A could not be effectively separated by this procedure from a contaminant that constituted about 15% of the final sample.

Site-Directed Mutagenesis. Point mutations were introduced into the *Adh-Waksman* gene, previously subcloned into M13mp19, using the method of Kunkel (1985). Oligonucleotides were synthesized by Oswel DNA Service (Department of Chemistry, University of Edinburgh). The mutated genes were then subcloned into the shuttle vector pVTU (Vernet et al., 1987) for expression in yeast (Atrian et al., 1990). Each mutant gene was recovered from yeast by plasmid rescue techniques and sequenced fully to confirm its expected sequence.

Steady-State Kinetics. ADH activity was determined spectrophotometrically by monitoring NAD reduction at 340 nm. All assays were done in 100 mM Tris-HCl, pH 8.7, at 25 °C. The apparent K_m values for NAD and 2-propanol were determined by monitoring the enzyme activity with a range of concentrations for one substrate. NAD concentrations between 0.05 and 3.2 mM were used, keeping the alcohol concentration at 10 mM. Similarly, concentrations ranging from 0.25 to 16 mM 2-propanol were used, keeping the cofactor NAD at 1 mM. The apparent K_m for NADP was calculated, keeping the 2-propanol at 10 mM and using concentrations of NADP from 0.1 to 6.4 mM. All data were obtained in triplicate. Hanes plots were used for the determination of the apparent K_m and V_{max} values for each of the enzymes analyzed.

Guanidinium-HCl Denaturation and $\Delta G_{(water)}$ Calculations. The determination of the $\Delta G_{(water)}$ values for each of the enzymes and the calculation of the $\Delta\Delta G$ effect of NAD binding were done as described previously (Ribas de Pouplana et al., 1991). All enzyme concentrations were approximately 10 μg/mL.

Thermal Denaturation Studies. The conformational free energy of ADH-F has been determined (Ribas de Pouplana et al., 1991), but other *Drosophila* ADHs have not previously been analyzed. Therefore, the thermal stability of each enzyme was determined in order to be able to compare the stability effects of the introduced mutations to published reports on stability (Thatcher & Sheikh, 1981; Chen et al., 1990, 1991). Samples of each enzyme were incubated at three different temperatures (0, 25, and 40 °C) on ice or in a thermostatted water bath. Small volumes were removed at fixed time periods, and the ADH activity present was immediately analyzed. All activity assays were done in duplicate.

Computer Modeling. Sequence alignments were done using the programs GAP (UWGCG Package 7.0; Devereaux et al., 1984) and CLUSTAL (Higgins & Sharp, 1988). Hydrophobicity calculations were done with PEPPLOT (UWGCG package). Predictions of the secondary structure of *Drosophila* ADH and *Streptomyces* HSD were made from the amino acid sequences using a multiple-structure prediction program (PREDICT) which combines a number of published algorithms (Eliopoulos et al., 1982).

The computer modeling was done using the modeling package SYBYL version 5.4 (Tripos Associates Inc.), run on an Evans and Sutherland ESV workstation. Protein sequences were obtained from the Swissprot database, and the three-dimensional coordinates for the α-carbons of HSD (Ghosh et

Table 1: Kinetic Constants for ADH-W and Mutant Enzymes[a]

	NAD			NADP			2-propanol		
enzyme	$K_{m(app)}$ (mM)	k_{cat} (s^{-1})	k_{cat}/K_m	$K_{m(app)}$ (mM)	k_{cat} (s^{-1})	k_{cat}/K_m	$K_{m(app)}$ (mM)	k_{cat} (s^{-1})	k_{cat}/K_m
ADH-W	0.18	16.6	89.5	1.40	3.9	1.8	0.60	16.7	27.7
G14A	0.36	7.1	19.8	2.40	1.3	0.5	0.64	7.1	11.2
G19A	0.32	17.2	53.7	no detectable activity			0.66	17.2	26.1
D38A	0.45	9.0	20.0	4.27	1.0	0.2	1.99	9.0	4.5
P214S	0.09	3.0	33.3	1.90	0.5	0.3	0.47	30	6.7

[a] Values calculated from Hanes reciprocal plots. All standard errors were less than 5%. The assays were done in 100 mM Tris-HCl, pH 8.7, as described in the text.

al., 1991) were from the Brookhaven Protein Data Bank (Bernstein et al., 1977). Refinement by simulated annealing was done with X-PLOR (Brünger et al., 1987).

Comprehensive sequence alignments of all available short-chain dehydrogenases, secondary structure predictions, and hydrophobicity plots were used to identify the putative hydrophobic β-sheet core and helical regions of *Drosophila* ADH. The residues selected to be in each secondary structure were then substituted into the coordinates of the backbone structure of HSD, which was generated from the α-carbon coordinates using the algorithm of Claessens et al. (1989). Whenever possible, the amphiphilic nature of the helices constructed was used as the main parameter to determine their packing in the model. Where amphiphilicity was not clear, the sequence alignment results were used to decide the external and internal faces of helical regions.

All the loops were constructed using the loop search algorithm of SYBYL. This method searches a database of high-resolution structures to find suitable sets of loops selected by the following parameters: (1) relative three-dimensional orientation between N-terminal and C-terminal residues of loops compared to that of the two residues to be connected by the loop in the model; (2) same number of residues in the loop as that to be modeled; and (3) sequence similarity between the loop in the reference structure and the loop being constructed. Using these three parameters, the best loop was chosen for each different connection point.

Energy minimizations were done using the SYBYL ANNEAL procedure, which relies on a molecular mechanics approach and Kollman's force field. The minimization uses an initial atom to atom approach (20 iterations) to remove internal strains followed by simultaneous optimization by conjugate gradients (100 iterations).

The program X-PLOR (Brünger et al., 1987) was used for the molecular dynamics calculations. Typically, the structure was subjected to 200 iterations of energy minimization using conjugate gradient methods with backbone coordinates fixed to remove steric clashes before the molecular dynamics procedure. The dynamics analysis was started by increasing the energy of the system to 2000 K followed by 34 cooling steps of 1-ps duration to reach a final temperature of 300 K. Due to the presence of exposed hydrophobic residues on the surface of the monomeric model, presumably corresponding to the monomer–monomer binding interface, electrostatic values of charged residues were set to zero and the dynamic simulations were run in vacuum.

RESULTS AND DISCUSSION

Steady-State Kinetics. The kinetic constants for the ADH-W enzyme and four mutants are shown in Table 1. The values obtained for ADH-W and for mutant G14A are very similar to those published previously (Chambers, 1988; Chen et al., 1990) and are included for reference purposes.

Generally speaking, the effect of the mutations at positions 14, 19, and 38 confirm the role of the three residues in the recognition of NAD, mainly affecting the K_m for that substrate without altering the K_m for the alcohol. As mentioned previously, the mutant D38A could not be totally purified; hence, its kinetic values must be approached with caution. The difference in the k_{cat} effect between G14A and G19A suggests that the two mutations affect the binding of NAD in different fashions. G19A does not affect the catalytic rate of the enzyme, indicating that the loss in cofactor recognition is compensated by a faster rate of reaction turnover.

The most revealing of all the results obtained with the N-terminal mutants of ADH is the lack of activity of mutant G19A when the cofactor NADP is used. Since that mutation has only a moderate effect on the activity of the enzyme with NAD, it must be concluded that the introduction of the extra methyl group directly interferes with the binding of the phospho group of NADP. This observation provides crucial information on the positioning of the cofactor in the active site of the enzyme.

The P214S mutant enzyme shows a 2-fold lower apparent K_m for the cofactor when compared to the wild-type enzyme. As expected, this results in a slower reaction speed, with a k_{cat} reduced by a factor of 5.5, thereby making this enzyme the slowest of all those tested. The mutation has little effect on the apparent K_m for 2-propanol. Residue 214 is thus clearly affecting the interaction with the cofactor NAD, although it is distant in sequence from other residues known to be involved in cofactor binding. Its location in the C-terminal part of the enzyme suggests that it might be in the environment of the reaction center. However, our kinetic results indicate that it is not in direct contact with the substrate.

Guanidinium-HCl Denaturation and $\Delta G_{(water)}$ Calculations. Typical GndHCl denaturation plots for the ADH-W enzyme and two of the mutants (G19A and P214S) in the presence and absence of NAD are shown in Figure 1. Extrapolation plots to obtain $\Delta G_{(water)}$ values are also shown. The other two mutants (G14A and D38A) also manifested very similar two-state denaturation processes (data not shown). Table 2 gives the values of $\Delta G_{(water)}$ thus obtained, as well as the $\Delta\Delta G$ effect of NAD binding on each of the enzymes. The $\Delta G_{(water)}$ values give a measure of the stability of each of the enzymes, and the contribution of coenzyme to stabilization in terms of $\Delta\Delta G_{(NAD)}$ is given by comparisons of the $\Delta G_{(water)}$ values of the apo- and holoenzymes (Ribas de Pouplana et al., 1991).

It has been known for some time from thermal gradient electrophoresis studies that the ADH-S isoenzyme is more stable than ADH-F (Thatcher & Sheikh, 1981). This difference in stability is quantified by our studies, which show a difference in free energy of approximately 0.7 kcal/mol in favor of ADH-W (Table 2). A difference between their $\Delta\Delta G_{(NAD)}$ values of approximately 2.3 kcak/mol was found, in agreement with the known fact that ADH-S binds NAD more tightly than does ADH-F. This value is similar to the

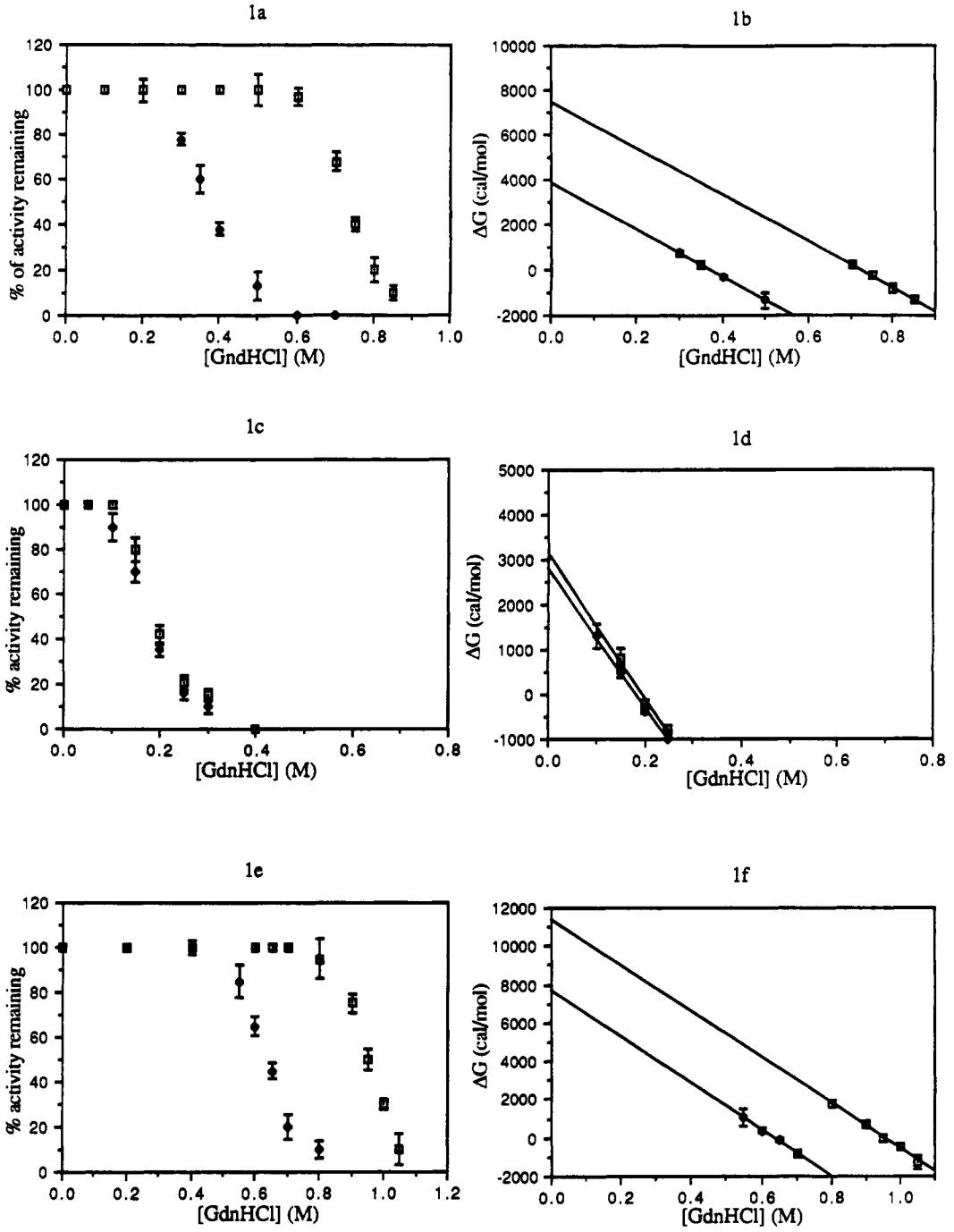

FIGURE 1: Guanidine inactivation of ADH-W and mutants G19A and P214S. The primary data are given in the left-hand plots and the corresponding $\Delta G_{(water)}$ extrapolations on the right. The measurements were done in the absence (◆) or presence of NAD (□). Plots: a and b, ADH-S; c and d, mutant G19A; e and f, mutant P214S. All the correlation coefficients for the equations used to fit the lines in plots b, d, and f were 0.99 or better.

enery of a hydrogen bond between uncharged donors and acceptors (0.5–1.8 kcal/mol) that was calculated from site-directed mutagenesis studies on tyrosyl tRNA synthetase (Fersht, 1987).

Interestingly, the apo forms of the two ADH isoenzymes differ by only 0.7 kcal/mol, whereas the two holo forms differ by 2.9 kcal/mol. This difference indicates that the substitution at position 192 has a larger effect on the binding of the cofactor than on the intrinsic stability of the enzyme.

The $\Delta G_{(water)}$ values of the apo forms of mutants G14A, G19A, and D38A are all reduced from ADH-W by similar amounts (from 1.1 to 1.5 kcal/mol). By contrast, the holo forms of the mutant enzymes show relatively large differences with regard to the energy that NAD binding confers. Mutant G14A is stabilized by NAD binding nearly to the same extent as is the ADH-W isoenzyme, whereas the other two mutants display only negligible stabilization.

These results are very interesting when compared to the apparent $K_{m(NAD)}$ and k_{cat} values (Table 1). Mutant G14A shows a 2-fold increase in apparent K_m and a 2-fold reduction in k_{cat}. The denaturation analysis suggests that the reason for the increase in the apparent K_m is probably due to changes in the catalytic constant. A possible structural explanation is that the mutation induces a small repositioning of NAD, maintaining its normal range of interactions but affecting the architecture of the transition state. The denaturation results and kinetic values for mutant G19A confirm that the interaction energy between enzyme and cofactor has been

Table 2: ΔG and $\Delta\Delta G$ Values for ADH-F, ADH-W, and Mutant Enzymes

	$GdnHCl_{50}$ (M)[a]		$\Delta G_{(water)}$ (kcal/mol)[b]		
enzyme	apoADH	+NAD	apoADH	+NAD	$\Delta\Delta G_{(NAD)}$[c]
ADH-F[d]	0.40	0.58	3.2	4.5	1.3
ADH-W	0.38	0.74	3.9	7.4	3.5
G14A	0.26	0.52	2.6	5.9	3.3
G19A	0.18	0.19	2.8	3.2	0.4
D38A	0.28	0.31	2.4	2.7	0.3
P214S	0.64	0.95	7.7	11.4	3.7

[a] $GdnHCl_{50}$, guanidinium hydrochloride concentration at which 50% loss of enzyme activity was achieved. [b] $\Delta G_{(water)}$ corresponds to the extrapolated $\Delta G_{(water)}$ value from the denaturation curve of each enzyme. [c] $\Delta\Delta G_{(NAD)}$ corresponds to the difference between the $\Delta G_{(water)}$ of each enzyme without cofactor (apoADH) and with cofactor (+NAD) present. [d] These values are from Ribas de Pouplana et al. (1991).

reduced while the reactivity of the enzyme has not been affected.

The results obtained with mutant P214S confirm that this substitution has the same stabilizing effect in ADH-W as the one naturally occurring in ADH-F isoenzymes (Chambers et al., 1981). The mutant enzyme has a $\Delta G_{(water)}$ value of 7.7 kcal/mol, an increase of 3.8 kcal/mol with respect to that of the original ADH-S molecule. The kinetic studies indicate that mutant P214S has a higher affinity for the cofactor than ADH-W, but this is not reflected in the $\Delta\Delta G_{(NAD)}$ values for both enzymes which are quite similar (only 0.15 kcal/mol higher for mutant P214S).

The monitoring of denaturation by circular dichroism shows that the secondary structure loss follows a much shallower slope than the activity loss (Ribas de Pouplana et al., 1991), and the possibility arises that the loss of activity is due to the monomerization of the enzyme. If this is the case, the positive effect of NAD on the Gibbs free energy of the enzyme would be due to a stabilization of the ADH dimer upon binding of NAD, either through long-range interactions between one monomer and the cofactor bound to the other or by structural rearrangements of the monomer–monomer interface.

Thermal Denaturation Studies. Thermal denaturation has been used regularly as a way to measure the stability of *Drosophila* ADH both in crude extracts and purified form (Chambers, 1984; Hernández et al., 1988; Chen et al., 1990, 1991). Normally, the results from this type of analysis are subject to too many different factors (enzyme concentration, purity, proteolysis) to be used reliably to determine small differences in stability, but large differences as in the case of the thermostable ADH-FChD can be consistently observed (Chambers, 1984).

The thermal denaturation of the four mutants as well as the wild-type ADH-W was studied to enable direct comparisons with previous reports. All the enzymes were stable when kept at 0 °C for 150 min. At 25 °C, the enzymes fell into two groups. ADH-W and mutant P214S were relatively stable over 150 min (Figure 2) and retained 80% and 95% of their respective activities. In contrast, mutants G14A, G19A, and D38A were much more labile and retained only 50%, 40% and 5% of their original activities under the same conditions. These results correlate quite well with the $\Delta G_{(water)}$ values obtained for all the enzymes (Table 2), thus confirming the general pattern of stability. At a higher temperature (40 °C), all the enzymes but mutant P214S were completely inactivated after 30 min of incubation. This is slightly faster than other reported experiments (Chambers, 1984; Chen et al., 1991). The P214S mutant enzyme retained 50% of its activity at 40 °C after 40 min.

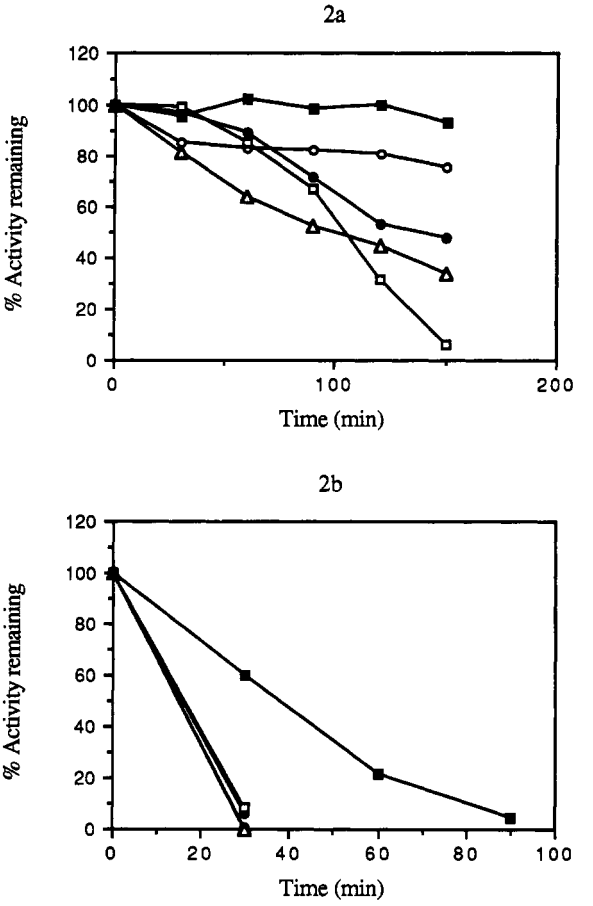

FIGURE 2: Thermal inactivation of wild-type and mutant ADHs. The incubations were done at 25 °C (a) or 40 °C (b). Each point was done in duplicate, and the mean value is plotted, ADH-W (○), mutant G14A (●), G19A (△), D38A (□), and P214S (■).

Computer Modeling. The portions of the ADH sequence that were selected to construct the main secondary structure elements of the model are shown in Table 3. The assignment of the lengths of each particular element in the ADH model depended upon parameters such as the presence of prolines, gaps in the alignment, or regions of poor sequence similarity. Thus, the lengths of the ADH elements were not necessarily the same at those in the hydroxysteroid dehydrogenase structure.

The crystal structure of HSD shows a large loop of 32 residues between helix-F and strand-G (Ghosh et al., 1991). The corresponding region of the ADH enzyme is virtually impossible to model because it has no sequence similarity to HSD and is far too long to use any standard database search algorithms reliably. Another part of the ADH enzyme that could not be modeled is the C-terminal region following strand-G. This region provides very low electron density in the HSD structure and has a very large temperature factor. It is also a region of poor sequence similarity between the two enzymes and, in fact, appears to be a very variable region among all short-chain dehydrogenases. These regions were simply substituted with the sequence of ADH and energy minimized without any further modification.

A schematic representation of the model can be seen in Figure 3. The backbone of the ADH model has a mean structural difference of 1.6 Å rms with respect to the backbone of HSD. Apart from the low electron density zones of HSD, the main differences between the HSD structure and the ADH model involve the loops between strand-D and helix-D, strand-E and helix-E, and strand-F and helix-F. This is not

Table 3: Alignments of the Sequences of the Secondary Elements of the HSD Structure (Ghosh et al., 1991) and the Parts of the ADH Sequence Used To Model the Same Elements[a]

strands		helices	
strand-A	7-KTVIIT-12	helix-A	16-ARLGAEAARQAVAA-29
	7-KNVIFV-12		18-IGLDTSKELLKR-29
strand-B	32-RVVLADV-38	helix-B	40-DEEGAATARELGDAA-54
	34-LVILDR-39		44-AAIAELKAIN-53
strand-C	58-HLDVTI-63	helix-C	67-WQRVVAYAREE-77
	62-PYDVTV-67		72-TTKLLKTIFAQ-82
strand-D	81-VDGLVNNAGIS-91	helix-D	100-SVERFRKVVDINLTGVFIGMKTVIP-124
	86-VDVLINGAGI-95		103-RTIAVNYTGLVNTTAILDFWDKRK-127
strand-E	132-GSIVNIS-138	helix-E	150-SSYGASKWGVRGLSKLAAVELGT-172
	132-GIICNIG-138		151-VYSGTKAAVVNFTSSLAKLA-170
strand-F	174-RIRVNSVH-181	helix-F	189-MTAETGIRQGEGNY-202
	175-VTAYTVN-181		189-LVHKFNSWLDVE-200
strand-G	234-GAELA-238		
	235-KLDLG-239		

[a] The HSD sequence is the top sequence of each aligned pair.

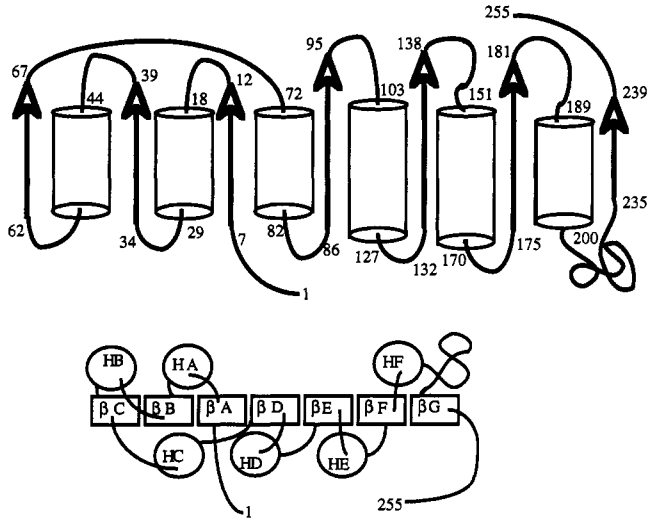

FIGURE 3: Schematic representations of the model of *Drosophila* ADH. In the upper diagram, arrows indicate β-strands and cylinders are α-helices. The residue numbers delimiting the elements of regular secondary structure are given. In the lower diagram, rectangles indicate β-strands A–G and circles represent α-helices A–F.

surprising as loops generally tend to be poorly conserved, and moreover, these regions were the larger ones constructed by the database search methods of loop sequences. Large structural differences were introduced by the dynamics procedure into the long loop between residues 200 and 230 and at the C-terminal tail of the enzyme.

An indication of the validity of the model comes from a consideration of the location within it of regions known to have high sequence variability. The variable regions identified from multiple sequence alignments of Drosophila ADHs (Villarroya & Juan, 1991) occur at positions far from the NAD-binding site, mainly in the loops following the helices of the structure (data not shown).

The Ramachandran plot of HSD shows six non-glycine residues in "nonallowed" regions of the plot (Ghosh et al., 1991), all belonging to loop areas or regions of poor electron density. The model of *Drosophila* ADH has 21 residues in nonallowed regions of the plot. Of these, 11 are located in the large loop preceding strand-G, which could not be modeled. Six are at the C-terminal tail, which was not modeled either, and the four remaining ones are distributed in three loop regions.

NAD-Binding Site and Putative Active Site. The positioning of the NAD molecule in ADH in a manner similar to that found in HSD leaves the nicotinamide part of the cofactor almost 10 Å away from any residue with possible catalytic activity. The position of NADH found in the structure of DHR (Varughese et al., 1992), by contrast, is more similar to the classical binding mechanism of dehydrogenases and places the nicotinamide ring of the cofactor in the proximity of the conserved residues Tyr-152 and Lys-156. Moreover, the mutation of Gly-19 to alanine selectively abolishes activity with NADP, indicating that the mode of coenzyme binding in ADH with respect to the poly(glycine) loop is more like that found in medium-chain dehydrogenases. Also, the NAD-binding position must take account for the results on the relative cofactor binding strengths of ADH-S and ADH-F, which suggest that residue 192 is in the proximity of the cofactor. It must be also kept in mind that the results of mutating Tyr-152 and Lys-156 (Ensor & Tai, 1991; Albalat et al., 1992; Cols et al., 1993; Chen et al., 1993) indicate that these residues are probably involved in catalysis.

A molecule of NAD, in a similar conformation to that described by Ghosh et al. (1991) when bound to HSD, was used to explore different positions in the ADH model that could agree with the structural proximity of all the mentioned residues to the cofactor molecule. A rotation and translation movement of the cofactor, without modifying its conformation, was enough to locate it in a sterically suitable position in which the carbamide atom of the nicotinamide ring finds an environment typical of a nonmetal dehydrogenase active site (Adams, 1987). The final position of the cofactor was improved by energy minimizing its conformation in the context of the active site of the model.

In the ADH model, strand-B, helix-C, and strand-D (Figure 3) form a cleft that can accommodate both the adenine ring and its adjacent ribose. Strand-A is located at the bottom of the cleft and immediately precedes the glycine residues that pack against the NAD molecule. It seems probable that the adenine ribose may pack against the hydrophobic residues Ile-40, Val-12, and Ala-93, which contribute to the bottom of the cleft. This type of environment is similar to those found in other dehydrogenases, such as malate dehydrogenase (Grau et al., 1981). Unfortunately, it is not possible to decide whether the adenine ribose may be hydrogen bonded to residue Asp-38 or to nearby Glu-41, as it would depend on the precise position of the NAD inside the cleft. However, residue Asp-64, implicated from sequence comparisons as possibly being responsible for hydrogen bonding to the ribose (Persson et al., 1991), is too far away in the model to carry out this function.

The ribose of the adenine moiety of the cofactor and the first atoms of the pyrophosphate are in close contact with the loop from residues 11 to 19. This is the poly(glycine) loop

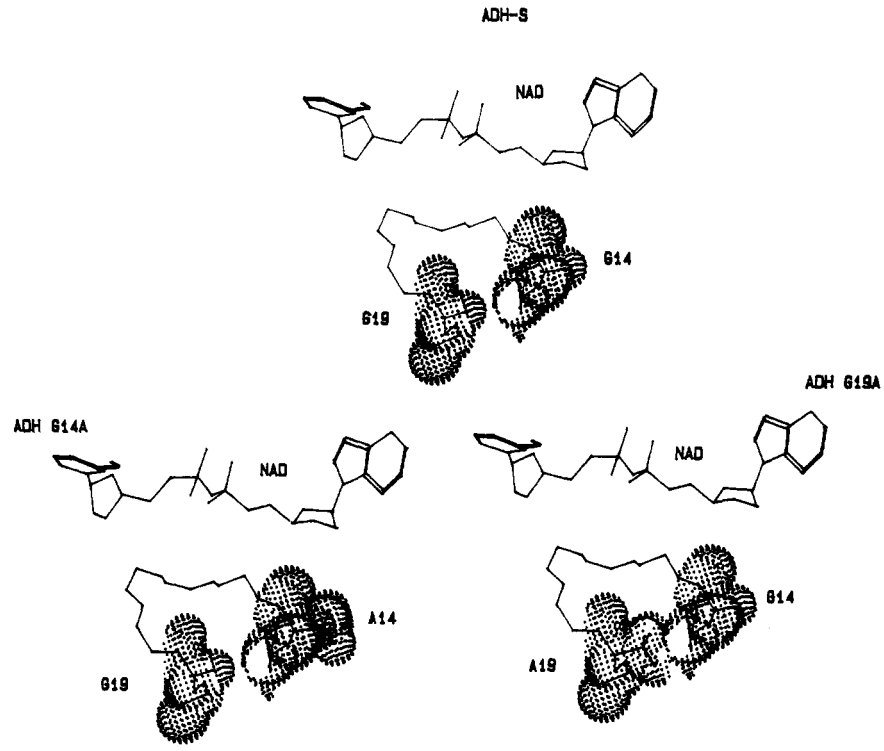

FIGURE 4: α-Carbon trace of the poly(glycine) loop of the models of ADH-W and of mutants G14A and G19A. The NAD molecule is shown at the top of each diagram, with the van der Waals surfaces of residues 14 and 19 indicated. Top diagram: wild-type ADH-W loop with Gly-14 and Gly-19. Bottom left: mutant G14A. The van der Waals surface representation of the new side chain shows that it does not affect the overall shape of the loop backbone. Bottom right: mutant G19A. In this case, the van der Waals surface representation shows that the methyl group of Ala-19 would clash with the opposite side of the loop backbone.

characteristic of all nucleotide-binding domains. Residues 12–15 of the loop bend inward, creating a pocket close to the ribose that is capable of accommodating a phospho group such as would be present in NADP. Figure 4 shows this loop with the cofactor in its proposed binding position. Residue Gly-14 is at the bottom of the pocket, with its hydrogen "side chain" pointing away from the loop. The incorporation of a methyl group in this position in mutant G14A may produce steric clashes with the neighboring loop at about residue 40, but it does not seem to affect the overall shape of its own loop. In contrast, Gly-19 is in very close packing to the backbone of residues 12 and 13, and the introduction of a methyl group here would be likely to distort the overall structure of the loop, including the pocket created by residues 12–15. This may be able to explain the inability of mutant G19A to use NADP because the perturbation of the loop may be enough to make the docking of the phospho group impossible. This may also be the reason for the difference found in $\Delta\Delta G$ of NAD binding between the two mutant enzymes. While mutant G14A has a $\Delta\Delta G$ value of NAD binding similar to that of the wild-type enzyme, in mutant G19A, this energy is reduced by 99% (Figure 1 and Table 2).

Another part of the ADH model in contact with the adenine moiety of the cofactor is the sequence Gly-Ala-Gly at positions 92–94 in strand-D. This sequence is conserved in 17 of 20 short-chain dehydrogenases (Persson et al., 1991), and the three residues are totally conserved among all the *Drosophila* enzymes. A similar sequence is also found in the region of lactate dehydrogenase that packs in a corresponding manner to the NAD molecule (Grau et al., 1981). The role of these residues is probably the same as that postulated for the Gly-rich loop of most NAD-binding structures, namely to provide enough space to allow the cofactor to enter the building pocket (Eklund & Brändén, 1987).

The nicotinamide ribose of the cofactor packs mainly against the loop which connects strand-F with helix-F (residues 182–188). This region is of interest because the only amino acid difference between the two main ADH isoenzymes in wild populations of *Drosophila melanogaster* is nearby at position 192 (threonine in ADH-S, lysine in ADH-F). This single substitution induces changes in the strength of cofactor binding, and many evolutionary studies have attempted to link this polymorphism to the differential distribution of fruit flies expressing one or the other of the two isoenzymes [reviewed by Chambers (1988)].

In the ADH model, residue 192 is not in direct contact with the cofactor but is positioned toward the exterior of the molecule at the first turn of helix-F. In this position, a direct interaction with any of the ligands is unlikely. Sequence comparisons between all the regions in contact with the cofactor in the ADH model and the corresponding regions in nucleotide-binding folds of other dehydrogenases consistently found a striking similarity between residues 182–192 of the ADH model and the portion of lactate dehydrogenase that forms the mobile loop involved in enzyme catalysis (Figure 5) (Gernstein & Chothia, 1991). This sequence similarity is highest with bacterial lactate dehydrogenases (67% identity), whose tertiary structures are almost identical to that of dogfish lactate dehydrogenase (Grau et al., 1981). The similarity between *Drosophila* ADH and lactate dehydrogenase suggests the presence of a comparable kind of loop in both enzymes. The idea of loop closure over the NAD molecule in ADH finds support from the observation that the cofactor has the capacity to protect ADH against proteolysis, chemical attack, or denaturation and to induce structural changes (Chen et al., 1990; Ribas de Pouplana et al., 1991; Krook et al., 1992). The effect of the substitution at position 192 might be related to changes in the dynamics of the structure upon cofactor binding.

```
B. stearo.   LDH      AGANQK PGETRLDLVDK NIAIFRS
B. mega.     LDH      AGANQA PGETRLDLVEK NVXIFEX
L. casei     LDH      AGAPKQ PGETRLDLVNK LNKILKS
                      *** :  *********:*   *:

                 176                              199
Drosophila   ADH      TAYTVN PGITRTTLVHK FNSWLDV
                      ::     ** ** :**:*  ::
```

FIGURE 5: Alignment of the sequence of residues 176-199 of ADH with the sequences of the mobile catalytic loops of bacterial lactate dehydrogenases [bacterial sequence alignments modified from Gernstein and Chothia (1991)]. The asterisks indicate identical residues and the colons similar residues. The boxed sequences show a strikingly high degree of identity. Abbreviations: *B. stearo., Bacillus stearothermophilus; B. mega., Bacillus megaterium; L. casei, Lactobacillus casei.*

In the proposed NAD-binding position, the nicotinamide ring itself lies over a set of hydrophobic residues, namely Ile-95, Leu-96, and Ile-137. Carbon-4 of the ring is in close proximity to residues Lys-156 and Ser-139, and residues Asp-97 and Tyr-152 are also in the vicinity.

Proposed Reaction Mechanism. A number of residues are good candidates to participate in the catalytic mechanism of ADH. Lys-156 and Tyr-152 are conserved among all short-chain dehydrogenases, and the results of mutating Tyr-152 and Lys-156 (Ensor & Tai, 1991; Albalat et al., 1992; Cols et al., 1993; Chen et al., 1993) suggest that these residues can be involved in catalysis. Also, the chemical modification of one of the four histidines of ADH results in the inactivation of the enzyme (Thatcher, 1981; Retzios, 1982). Of the four histidines, only three are totally conserved in *Drosophila* alcohol dehydrogenases (residues 191, 210, and 250). In the ADH model, His-191 and His-210 are not in the neighborhood of the active site of the enzyme and are thus unable to participate in catalysis. In contrast, several lines of evidence implicate the C-terminal loop of the enzyme as a contributor to the catalytic mechanism. This possibility has strong experimental support in the report by Krook et al. (1992) that shows that specific proteolysis at amino acid residue 244 of *Drosophila* ADH inactivates the enzyme. This proteolysis is inhibited by the presence of NAD, suggesting a change of conformation of this part of the enzyme upon NAD binding. This region of ADH corresponds to a mobile portion of HSD< as shown by X-ray crystallography (Ghosh et al., 1991). The C-terminal loop in our computer model is close to the proposed active site space delimited by the position of NAD (Figure 6).

On the basis of experimental data and the ADH model, we wish to propose a reaction mechanism involving the closure of the C-terminal loop. A round of catalysis would be initiated by the binding of the cofactor. This would provoke a modification to the environment of the active site, allowing it to receive the substrate molecule, and would also trigger the closure of the C-terminal loop over the reaction center. Possible reacting residues in the active site would be Tyr-152, Lys-156, and His-250. Once the dehydrogenase reaction is complete, the loop would open again, perhaps due to electrostatic repulsion with NADH, and product release would then ensure. The enzyme would now be ready to undertake another round of catalysis.

This reaction mechanism is compatible with many lines of experimental evidence and, for example, offers explanations for the effect of chemical modification of histidines. The proposed mechanism also provides an explanation for the rate-limiting effect of the release of NADH, as this would be a function of the opening of the C-terminal loop. Moreover, all the residues proposed to be involved are totally conserved in all the *Drosophila* ADHs sequenced so far.

In addition to the three active site residues which are proposed to be directly involved in the catalytic mechanism, the model shows that Val-140 could be intimately involved in substrate binding. One 2-propanol molecule could pack against the methyl groups of Val-140, a residue that is also conserved in all *Drosophila* ADHs. This role for Val-140 is consistent with the observations of Winberg et al. (1982), which indicated that each methyl group of the substrate interacts with a hydrophobic region of the enzyme.

Two Different Classes of Short-Chain Dehydrogenases. The determination of the crystal structure of 3α,20β-hydroxysteroid dehydrogenase (Ghosh et al., 1991) has provided long-awaited structural information needed to analyze the structure–function relationships of short-chain dehydrogenases. However, the structural features of this enzyme are in some respects quite different from those characteristic of other dehydrogenases, particularly with regard to the positions of the cofactor and substrate. It is apparent that ADH and HSD share very similar overall folds. As might be expected, the regions of protein structure that show the greatest differences are in the surface loops distant from the ligand-binding sites. However, HSD and ADH appear to be quite different with regard to cofactor recognition and architecture of the active site. In these respects, our results suggest that ADH binds NAD in a classical dehydrogenase manner, more closely resembling the active site architecture of dihydropteridine reductase (Varughese et al., 1992). Nevertheless, the very different structures of the substrates of these enzymes probably imply that there must be large

FIGURE 6: Stereo ribbon diagram of the ADH model with a skeletal representation of the NAD molecule in its proposed binding position. The side chains of four residues thought to be in the active site are from right to left: His-250, Val-140, Lys-156, and Tyr-152.

differences in the reaction center. Further work on the active site residues defined by our model is now required to determine the alcohol-binding mechanism of ADH.

ACKNOWLEDGMENT

We are very grateful to Dr. Steven K. Chapman (Department of Chemistry, University of Edinburgh) for help with the analysis of the active site of the ADH model and the reaction theory and to Dr. Nicholas C. Price (Department of Biological and Molecular Sciences, University of Stirling) and Prof. Paul Schimmel (Department of Biology, MIT) for helpful discussions. We thank Dr. William Sofer (Waksman Institute, Rutgers University) for plasmid pWX3008 containing an intronless version of the gene encoding *Drosophila* ADH-W. We are very grateful to Dra. Silvia Atrián and Prof. Roser Gonzàlez-Duarte (Department of Genetics, University of Barcelona) for helpful discussions and suggestions about the manuscript.

REFERENCES

Adams, M. J. (1987) in *Enzyme Mechanisms* (Page, M. I., & Williams, A., Eds.) pp 477–505, The Royal Society of Chemistry, Cambridge, U.K.

Albalat, R., Gonzàlez-Duarte, R., & Atrian, S. (1992) *FEBS Lett. 308*, 235–239.

Atrian, S., Gonzàlez-Duarte, R., & Fothergill-Gilmore, L. A. (1990) *Gene 93*, 205–12.

Bajorath, J., Stenkamp, R., & Aruffo, A. (1993) *Protein Sci. 2*, 1798–1810.

Bauer, A. J., Rayment, I., Frey, P. A., & Holden, H. M. (1992) *Proteins 12*, 372–381.

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

Brünger, A. T., Kuriyan, J., & Karplus, M. (1987) *Science 235*, 458–460.

Chambers, G. K. (1984) *Biochem. Genet. 22*, 529–549.

Chambers, G. K. (1988) *Adv. Genet. 25*, 39–107.

Chambers, G. K. (1991) *Comp. Biochem. Physiol. 99*, 723–730.

Chambers, G. K., Wilks, A. V., & Gibson, J. B. (1981) *Aust. J. Biol. Sci. 34*, 625–637.

Chen, Z., Lu, L., Shirley, M., Lee, W. R., & Chang, S. H. (1990) *Biochemistry 29*, 1112–1118.

Chen, Z., Lee, W. R., & Chang, S. H. (1991) *Eur. J. Biochem. 202*, 263–267.

Chen, Z., Jiang, J. C., Lin, Z.-G., Lee, W. R., Baker, M. E., & Chang, S. H. (1993) *Biochemistry 32*, 3342–3346.

Claessens, M., Van Cutsem, E., Lasters, I., & Wodak, S. (1989) *Protein Eng. 22*, 335–345.

Clark, W. (1988) Ph.D. Thesis, Rutgers University, Piscataway, NJ.

Cols, N., Marfany, G., Atrian, S., & Gonzàlez-Duarte, R. (1993) *FEBS Lett. 319*, 90–94.

Devereaux, J., Haeberli, P., & Smithies, O. A. (1984) *Nucleic Acids Res. 12*, 765–784.

Eklund, H., & Brändén, C.-I. (1987) in *Biological Macromolecules and Assemblies*, (Jurnak, F. A., & McPherson, A., Eds.) Vol. 3, pp 74–114, John Wiley & Sons, New York.

Eliopoulos, E. E., Geddes, A. J., Brett, M., Papin, D. J. C., & Findlay, J. B. C. (1982) *Int. J. Biol. Macromol. 2*, 263–268.

Ensor, C. M., & Tai, H. H. (1991) *Biochem. Biophys. Res. Commun. 176*, 840–845.

Fersht, A. (1987) *Trends Biochem. Sci. 12*, 301–309.

Gernstein, M., & Chothia, C. (1991) *J. Mol. Biol. 220*, 133–149.

Ghosh, D., Weeks, C. M., Grochulski, P., Duax, W. L., Erman, M., Rimsay, R. L., & Orr, J. C. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 10064–10068.

Gordon, E. J., Bury, S. M., Sawyer, L., Atrian, S., & Gonzàlez-Duarte, R. (1992) *J. Mol. Biol. 227*, 356–358.

Grau, U. M., Trommer, W. E., & Rossmann, M. G. (1981) *J. Mol. Biol. 151*, 289–297.

Heinstra, P. W. H., Scharloo, W., & Thorig, G. E. W. (1988) *J. Mol. Evol. 28*, 145–150.

Hernández, J. J., Vilageliu, Ll., & Gonzàlez-Duarte, R. (1988) *Genetica 77*, 15–24.

Higgins, P. G., & Sharp, P. M. (1988) *Gene 73*, 237–244.

Holm, L., Sander, C., & Murzin, A. (1994) *Nature Struct. Biol. 1*, 146–147.

Krook, M., Prozorovski, V., Atrian, S., Gonzàlez-Duarte, R., & Jörnvall, H. (1992) *Eur. J. Biochem. 209*, 233–239.

Krook, M., Ghosh, D., Daux, W., & Jörnvall, H. (1993a) *FEBS Lett. 322*, 139–142.

Krook, M., Ghosh, D., Strömberg, R., Carlquist, M., & Jörnvall, H. (1993b) *Proc. Natl. Acad. Sci. U.S.A. 90*, 502–506.

Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 488–492.

Persson, B., Krook, M., & Jörnvall, H. (1991) *Eur. J. Biochem. 200*, 537–543.

Retzios, A. (1982) Ph.D. Thesis, University of Edinburgh, Edinburgh, Scotland.

Ribas de Pouplana, Ll., Atrian, S., Gonzàlez-Duarte, R., Fothergill-Gilmore, L. A., Kelly, S. M., & Price, N. C. (1991) *Biochem. J. 276*, 433–438.

Thatcher, D. R. (1980) *Biochem. J. 187*, 875–883.

Thatcher, D. R. (1981) *Biochem. Soc. Trans. 9*, 299–300.

Thatcher, D. R., & Sawyer, L. (1980) *Biochem. J. 187*, 884–886.

Thatcher, D. R., & Sheikh, R. (1981) *Biochem. J. 197*, 111–117.

Varughese, K. I., Skinner, M. M., Whiteley, J. M., Matthews, D. A., & Xuong, N. H. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 6080–6084.

Vernet, T., Dignard, D., & Thomas, D. Y. (1987) *Gene 52*, 225–233.

Villarroya, A., & Juan, E. (1991) *J. Mol. Evol. 32*, 421–428.

Wierenga, R. K., Terpstra, P., & Hol, W. G. J. (1985) *J. Mol. Biol. 187*, 101–107.

Winberg, J.-O., & McKinley-McKee, J. S. (1988) *Biochem. J. 251*, 223–227.

Winberg, J.-O., Thatcher, D. R., & McKinley-McKee, J. S. (1982) *Biochim. Biophys. Acta 704*, 7–16.

Winberg, J.-O., Hovik, R., & McKinley-McKee, J. S. (1985) *Biochim. Genet. 23*, 205–216.

Winberg, J.-O., Hovik, R., Mckinley-McKee, J. S., Juan, E., & Gonzàlez-Duarte, R. (1986) *Biochem. J. 235*, 481–490.