The Structure of Retinal Dehydrogenase Type II at 2.7 Å Resolution: Implications for Retinal Specificity[†,‡]

Audrey L. Lamb and Marcia E. Newcomer*

Department of Biochemistry, Vanderbilt University School of Medicine, Nashville Tennessee 37232

Received January 7, 1999; Revised Manuscript Received March 11, 1999

ABSTRACT: Retinoic acid, a hormonally active form of vitamin A, is produced in vivo in a two step process: retinol is oxidized to retinal and retinal is oxidized to retinoic acid. Retinal dehydrogenase type II (RalDH2) catalyzes this last step in the production of retinoic acid in the early embryo, possibly producing this putative morphogen to initiate pattern formation. The enzyme is also found in the adult animal, where it is expressed in the testis, lung, and brain among other tissues. The crystal structure of retinal dehydrogenase type II cocrystallized with nicotinamide adenine dinucleotide (NAD) has been determined at 2.7 Å resolution. The structure was solved by molecular replacement using the crystal structure of a mitochondrial aldehyde dehydrogenase (ALDH2) as a model. Unlike what has been described for the structures of two aldehyde dehydrogenases involved in the metabolism of acetaldehyde, the substrate access channel is not a preformed cavity into which acetaldehyde can readily diffuse. Retinal dehydrogenase appears to utilize a disordered loop in the substrate access channel to discriminate between retinaldehyde and short-chain aldehydes.

Vitamin A is a fat-soluble, essential dietary vitamin also known as retinol. Two derivatives of retinol are biologically active: retinal and retinoic acid. Retinal is the chromophore of the photoreceptors in the visual system (*1*), while retinoic acid is a hormonal modulator of gene expression. Retinoic acid signaling is tightly regulated for proper functioning of the reproductive systems, growth and differentiation of many cell types, and pattern formation in embryonic development. Vitamin A deficiency during embryonic development causes a characteristic spectrum of congenital malformations including defects of the cardiovascular system, respiratory tract, diaphragm, urogenital tract, and eye (*2*−*3*). Similarly, excess Vitamin A during pregnancy can lead to teratological effects including brain and spinal cord deformities, abnormal craniofacial development, limb and organ defects, and altered rostrocaudal pattern formation (*4*−*7*).

The first suggestion that retinoic acid could act as a morphogen came from experiments on amphibian limb regeneration (*8*−*10*) and chick limb bud development (*11*−*15*). The discovery of members of the steroid/thyroid hormone nuclear receptor super family that modify transcriptional activity upon specific binding of retinoic acid furthered the idea that retinoids could be morphogens during development. It was originally proposed that the distribution of these retinoic acid responsive transcription factors would delineate areas of retinoic acid action in the embryo (*16*−*19*). However, the results of retinoic acid receptor (RAR) and retinoid X receptor (RXR) gene targeting experiments revealed that knowledge of these expression patterns is not sufficient to understand the function of the receptors during embryogenesis (*20*−*22*). Therefore, consideration should be made for the effects that cellular retinoid-binding proteins may have on retinoid availability for signaling. Also, the sites of retinal formation by retinol dehydrogenases and retinoic acid formation by retinal dehydrogenases, indicating the apex of the morphogenic gradient, play a role in the spatio-temporal signaling (*23*).

It has been shown for rat embryos post-gastrulation that cellular retinol-binding protein type I, retinol dehydrogenases, retinal dehydrogenase type II (RalDH2),[1] and the retinol-binding protein receptor all co-localize in tissues which require vitamin A for normal development (ref *24*; Napoli, J. L., personal communication). The area in the gastrulating embryo with the highest concentration of retinoic acid is the primitive node (*25*). The enzyme that produces retinoic acid in the mouse embryo during gastrulation is retinal dehydrogenase type II (*23*). Retinal dehydrogenase type II (RalDH2) has been shown to be in the tissue surrounding the primitive node. It has been suggested that retinol is converted to retinal in the primitive streak and retinal is then converted to retinoic acid in the primitive node. Therefore, the reactions for the generation of retinoic acid in the gastrulating embryo may take place sequentially as cells migrate from the primitive streak and through the primitive node (*23*).

Retinal dehydrogenase type II belongs to the super family of aldehyde dehydrogenases, enzymes which convert an aldehyde to a carboxylic acid for use in metabolic pathways or for elimination. The aldehyde retinal is required for the

[†] This work was supported by an NIH Grant to M.E.N. (GM55420). A.L.L. was supported by an NIH training Grant (GM0820).

[‡] Atomic coordinates have been deposited in the PDB with the accession code 1BI9.

* Corresponding author. E-mail: newcomer@lhmrba.hh.vanderbilt.edu. Phone: (615) 343-7333. Fax: (615) 343-1898.

[1] Abbreviations: RalDH2, retinaldehyde dehydrogenase type II; ALDH1, class I aldehyde dehydrogenase; ALDH2, class II aldehyde dehydrogenase; ALDH3, class III aldehyde dehydrogenase.

10.1021/bi9900471 CCC: $18.00
Published on Web 04/22/1999

visual system, and as a precursor to retinoic acid. The conversion to the latter is presumably a function of specific aldehyde dehydrogenases. However, aldehydes have been shown to produce the deleterious effects of carcinogenicity, cytotoxicity, mutagenicity, and genotoxicity (*26*, *27*), and aldehyde dehydrogenases are also detoxification enzymes. In general, aldehyde dehydrogenases selectively eliminate the aldehydes generated as physiological intermediates in the metabolism of other compounds. One example is the conversion of acetaldehyde to acetic acid in ethanol metabolism.

The aldehyde dehydrogenase super family is subdivided into three classes. Classes I and II are comprised of the tetrameric isozymes (subunit ~55 kD). The isozymes of class I are cytosolic while those of class II are mitochondrial (*28*). Class III aldehyde dehydrogenases are smaller, dimeric, and primarily tumor-specific, inducible, or found in the endoplasmic reticulum (*29*). All three classes require either nicotinamide adenine dinucleotide or nicotinamide adenine dinucleotide phosphate as a cofactor. Classes I and II share approximately 70% sequence identity. However, within a class, these enzymes are approximately 95% identical (*29*). X-ray structures have been determined for class III (*30*) and class II (*31*) enzymes, as well as the related betaine aldehyde dehydrogenase from cod liver (*32*). More recently, the structure of a retinaldehyde metabolizing class I aldehyde dehydrogenase has been reported (*34*).

RalDH2 is a homotetramer of 55 kD monomers which is approximately 70% identical to either the class I or class II enzymes (*34*, *35*). In vitro, RalDH2 efficiently processes a number of aliphatic aldehydes including octanal, decanal, and retinal, but not acetaldehyde (*33*).

We report here the crystal structure of retinal dehydrogenase type II complexed with nicotinamide adenine dinucleotide at 2.7 Å resolution as determined by X-ray crystallography. The enzyme appears to utilize a disordered stretch of approximately 20 amino acids to limit access to the catalytic site by smaller aldehydes.

MATERIALS AND METHODS

Crystallization. The protein was purified as previously described (*34*, *36*). The crystallization was based on the previously published report (*36*), but the conditions were modified slightly to enhance crystal size. Utilizing the sitting drop method and solutions that had been degassed by He sparging, a reservoir solution was prepared that contained 1.0−1.3 M ammonium sulfate, 0.1 M MES, pH 7.1, 5% dioxane, and 9−12% ethylene glycol. The protein solution was 5 mg/mL (as determined by Bradford assay) in a buffer system containing 20 mM hepes, pH 8.5, 150 mM KCl, 1 mM EDTA, 1 mM β-mercaptoethanol, 2 mM NAD, and 2 mM β-ionone. β-Ionone was added as a nonmetabolizable substrate analogue. Although at this concentration it does not inhibit the enzyme in our standard activity assay, its addition to the drops did enhance crystal size. Drops were made using 10 μL of the protein solution and 10 μL of the well solution over a reservoir of 1.4 mL. The trays were setup on ice and then moved to a 25 °C incubator.

Cryo-protection. Crystals were prepared for data collection by serial soaks in mother liquor containing increasing cryoprotectant concentration. Three mother liquors were generated containing the reservoir solution above with increasing amounts of ethylene glycol: 15%, 20%, 25%. The crystal was harvested from the drop with a loop and soaked sequentially in each mother liquor for 5 min. The crystal was then flash cooled in the N_2 stream of the cold flow apparatus.

Table 1.

Data Collection Statistics	
space group	$P2_12_12$
cell axes (Å)	
a	149.9
b	167.3
c	107.5
maximum resolution (Å)	2.7
no. of crystals	1
temperature (°C)	−150
reflections measures	864923
unique reflections	75572
completeness (%)	99.7(99.9)
Rmerge (%)[a]	7.5(41.2)
I/σ	21.5(3.6)
V_m (Å^3/Da)	3.09
Refinement Statistics	
model	
no. of protein atoms	14826
no. of NAD atoms	108
no. of water molecules	224
no. of chloride atoms	2
data	
resolution (Å)	30−2.7
sigma cutoff	none
no. of reflections	72113
R_{factor} (%)	23.4(34.3)
R_{free} (%)[b]	28.9(39.9)
stereochemistry (rms deviation from ideality)	
bond length (Å)	0.008
bond angle (deg)	1.3
dihedral angles (deg)	23.2
improper twist angles (deg)	0.86
average *B* factor	
nucleotide-binding domain (Å^2)	37.2
catalytic domain (Å^2)	63.6
tetramerization domain (Å^2)	40.7
NAD (Å^2)	45.7
solvent (Å^2)	37.1
overall	48.1
B from Wilson plot	53.2
Ramachandran	
outliers (%)	0.3
most favored regions (%)	85.4

[a] The values in parentheses are for the highest-resolution shell. [b] The percentage of the reflections used to calculate R_{free} was 10.3%.

Data Collection and Processing. Diffraction data were collected at −150 °C on an RaxisII image plate system mounted on a Rigaku RU-200 rotating anode. The exposure time per 0.5° frame was 45 min with a crystal to detector distance of 165 mm. All data were processed using DENZO (*37*, *38*). Data processing statistics are given in Table 1.

Structure Determination and Refinement. Molecular replacement calculations were done with AMORE (*39*), as implemented in the CCP4 suite of programs (*40*). Mitochondrial aldehyde dehydrogenase (*31*) served as the search model. The coordinates for this model were kindly supplied by the authors prior to publication in the Brookhaven Protein Data Bank. The full tetramer with side chains intact served as the search model. Data from 8 to 4 Å resolution were used with the space group as originally defined ($P2_12_12_1$; ref *36*). The rotation and translation functions revealed a clear

solution with a correlation coefficient of 0.505 and an R_{factor} of 0.465. While the map drawn with this solution did look remarkably good in the N-terminal nucleotide-binding domain, the map in the C-terminal catalytic domain was remarkably poor and frequently uninterpretable and the free R_{factor} upon initial refinement in X-PLOR (*41*) remained unrealistically high (*42*). A new molecular replacement solution was sought. A number of other translation solutions that were less statistically plausible were checked for validity by viewing electron density maps. None was better than the initial and most statistically correct model. Finally, close inspection of the rejected reflections of the native diffraction data and the packing diagrams of the molecular replacement solution led to a questioning of the space group assignment.

To test the hypothesis that the space group had been defined improperly, we generated pseudo-precession images using the Precess program in the PHASES package (*43*). The data did appear to have systematic absences characteristic of 2-fold screw axes along *x*, *y*, and *z*: thus, the previous space group assignment of $P2_12_12_1$. However, very close inspection of the reflections believed to be absent showed weak reflections. These very weak reflections were included in a new processing of the data set, and related space groups were investigated. The best translation function solution obtained was in space group $P2_12_12$ with the axes permuted from $a = 108$ Å, $b = 150$ Å, and $c = 167$ Å of the original indexing to $a = 150$ Å, $b = 167$ Å, and $c = 108$ Å. The correlation coefficient for this solution was 0.450, and the R_{factor} was 0.456. The molecular replacement solution previously found for space group $P2_12_12_1$ has the same packing as the solution for space group $P2_12_12$ at the tetramer level. The difference comes from the inspection of the monomer placement after the generation of symmetry mates; that is, the packing of the molecules is pseudo $P2_12_12_1$ as a result of the fact that a 2-fold axis of the tetramer is parallel to the crystallographic 2-fold.

The positioned model was subjected to rigid body refinement using the program package X-PLOR, allowing for further refinement of the orientation and position of the model. This refinement was done at the tetramer, monomer, and domain levels to account for any relative domain shifts between RalDH2 and the original model ALDH2 structure. The new RalDH2 model was modified to make it more accurate by changing the side chains from the side chains of ALDH2 to the side chains of RalDH2 using the O (*44*).

The first two cycles of simulated annealing refinement with X-PLOR were done without noncrystallographic restraints. The density of the four monomers was then averaged with the program DM in the CCP4 package. It was immediately obvious that there was no density for the loop comprised of residues 462−477 in any of the monomers, in either the averaged or unaveraged maps, and this portion of the original model was removed. Successive cycles of model building in O and simulated annealing and positional refinement in X-PLOR using 4-fold noncrystallographic restraints for all but the N-terminal 20 amino acids improved the structure progressively. The ADP moieties of the four NAD molecules (one per monomer) were added to the structure as the corresponding densities became well-defined. The nicotinamide ring and ribose portions of the cofactor could not be modeled as there was no electron density for them.

For subsequent refinement and rebuilding cycles, torsion angle dynamics as implemented in CNS were utilized (*45*). Data between 30.0 and 2.7 Å resolution with no sigma cut off were included in the refinement, and a bulk solvent correction was applied. At this stage additional amino acids at the amino terminus of monomer A were apparent in the electron density maps, as well as solvent molecules. Two chloride atoms were added upon visual inspection of $|F_o - F_c|$ and $|2F_o - F_c|$ maps in high sigma electron density peaks at sites which might bind anions. Noncrystallographic restraints were removed for the final annealing and *B* factor (one main-chain and one side-chain B per residue) refinement cycle. The final model has an R_{factor} of 23.4% and R_{free} of 28.9%. The rms deviations of main-chain atoms of the monomers are ∼0.4 Å. Ramachandran plots, generated with the program PROCHECK (*46*), revealed 5 amino acids in the disallowed regions. Four of these (Lys 377 of monomers B, C, and D and Phe B425) are found in turns. The fifth (Asn D457) is amino-terminal to the missing loop.

Although there are isolated patches of electron density in the cavities of monomers C and D which could be due to polypeptide, this density is ambiguous and we were unable to model the missing loop in this area. The addition of solvent molecules in the area did not help to resolve the ambiguities. The final refinement statistics are given in Table 1. Representative electron density maps are illustrated in Figure 1.

RESULTS AND DISCUSSION

The retinal dehydrogenase type II monomer is comprised of 3 domains: a nucleotide-binding domain (1−136, 161−270), a catalytic domain (271−484), and a tetramerization domain (137−160, 485−500) (see Figure 2). The nucleotide-binding domain and the catalytic domain are both α−β in structure, while the tetramerization domain, which is significantly smaller than the other two domains, is formed by three β-strands. Tetrameric RalDH2 is a dimer of dimers. One dimerization interface, located between nucleotide-binding domains, involves antiparallel contacts of the αI helix (residues 247−260) with the same helix of its dimer mate. In addition, the β11 strand onto which αI packs (residues 264−269) runs antiparallel to the β11 strand of the dimer pair, and although the two strands are not close enough to form β-sheet contacts, ordered water molecules are found between them. A second dimerization interface is provided by β18 (residues 450−454) of the catalytic domain β-sheet of one monomer and β19 (residues 486-495) from the tetramerization domain of its dimer partner. In this interaction, monomers "embrace" via the formation of antiparallel β-sheet contacts. This "embrace" results in the formation of a channel for substrate access to the active site (Figure 3a).

Tetramerization is via the β5 strands (residues 143−144) which form an antiparallel β-sheet between monomers of opposing dimers (Figure 3). One might envision the tetramer as an "X", with the dimers forming the upper and lower halves of the "X". The tetramerization domains, as well as the substrate access channels, are located equatorially on the "X". The nucleotide-binding sites are at the periphery, or tips, of the "X".

As mentioned above, X-ray structures for class II (ALDH2) and class III (ALDH3) and recently class I (ALDH1) aldehyde dehydrogenases have been reported. The ALDH2

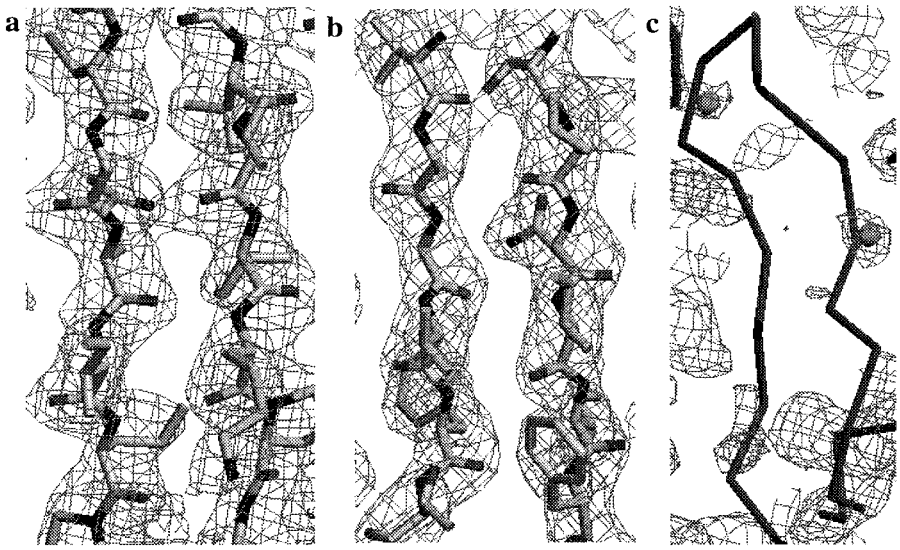

FIGURE 1: Representative portions of the final $|2F_o - F_c|$ map. (a) A region of the nucleotide-binding domain contoured at 1σ. (b) A region in the catalytic domain contoured at 1σ. (c) The area where the substrate access channel loop is positioned in the highly homologous ALDH2, contoured at 0.8σ. The electron density which surrounds the loop area is from neighboring side chains.

enzyme structure served as the search model in our molecular replacement solution of RalDH2. Aldehyde dehydrogenases employ a catalytic Cys which forms an acyl intermediate in the course of the oxidation of substrate. The dimeric (class III) and tetrameric (class I and II) enzymes share a common structural motif, as would be expected for the level of observed sequence identity (roughly 30% (III) and 70% (I and II) with RalDH2, respectively). From the results of the previous X-ray crystallographic studies, it is clear that some of the amino acids which may impart specificity in this enzyme family are at the carboxy-terminal end of the catalytic domain. B. C. Wang and colleagues, in their description of the ALDH3 structure, indicated that the residues which correspond to 455−465 in RalDH2 line part of the substrate-binding cavity in that enzyme. These amino acids are found near the tetramerization interface of RalDH2 where there is a tunnel roughly 12 Å deep which would allow access to the catalytic cysteine (Cys 302 in RalDH2). On the other side of the catalytic Cys is the NAD. The walls of the channel include amino acids from both the nucleotide-binding and catalytic domains as well.

The structure of RalDH2 was determined in the presence of NAD and the absence of substrate or substrate analogue. Amino acids 462−478 are not visible in the electron density maps (Figure 1) and, unlike their counterparts in the previous ALDH structures which were also determined in the presence of cofactor but absence of substrate, are disordered. (Alternatively, these regions may have more than one possible conformation, and these conformations could be randomly distributed in the tetramer.) In contrast to what is observed for ALDH2 and ALDH3, retinaldehyde dehydrogenase type 2 shows preference for aldehydes with long hydrocarbon chains in vitro and does not efficiently oxidize acetaldehyde. The K_m and V_{max} for acetaldehyde are 645 μM and 139 nm/min/mg but substantially improved for octanal and decanal (5 and 3 μM, 759 and 653 nm/min/mg, respectively). Furthermore, the putative biological substrate all-trans-retinal has a K_m of 0.4 μM and V_{max} of 20 nm/min/mg (*34*).

Unlike the previously described aldehyde dehydrogenases which have preformed substrate access channels into which acetaldehyde can readily diffuse, the substrate access channel of RalDH2 appears to be disordered. If retinal binds in the cavity with a hand-in-glove-like fit, as retinol does in the retinol-binding protein (*47*) or retinoic acid in the retinoic acid receptor ligand-binding domain (*48*), the binding of the hydrophobic substrate could stabilize the loop as well as the catalytic machinery. By this mechanism, the enzyme can discriminate between long- and short-chain aldehydes, as the short-chain aldehydes do not have the large hydrophobic surface necessary to bury inside the channel and as a consequence stabilize the catalytic groups. RalDH2 would fall into the collection of enzymes in which substrate binding promotes an "induced fit" as a prerequisite for catalysis. In this case "induced fit" may include a disorder to order transition and is not simply a highly localized conformational change. Substantial precedent exists for this proposal. For example, a disordered loop of 20 amino acids in length has been proposed to be part of the mechanism for substrate specificity of a deubiquitinating enzyme (*49*). Furthermore, retinoic acid binding to the retinoic acid receptor induces a marked conformational change, albeit one which does not appear to involve the ordering of extensively mobile regions (*48*).

What is not readily inferred from the comparison of the structure of the highly homologous tetrameric ALDH2 with RalDH2 is the structural basis for the order in the former enzyme and "disorder" in the latter. The most significant differences in sequences in the region in question are the substitution of a Phe (459)-Gly (460) in ALDH2 for Leu-Asn in RalDH2. The Phe in ALDH2 participates in an aromatic cluster formed with Trp 177, Phe 170, and Phe 296. This cluster could contribute to the stability of this region in the ALDH2 structure. In addition, the substitution of Gly with non-Gly residues may prevent placement of the RalDH2 loop in a conformation equivalent to that seen for the loop in ALDH2. Although the Gly in ALDH2 has main-chain torsion angles accessible to non-Gly residues, the addition of the side chain would appear to be sterically prohibitive (Figure 4a,b). In terms of sequence changes in the neighboring residues which may be of consequence, there is a salt

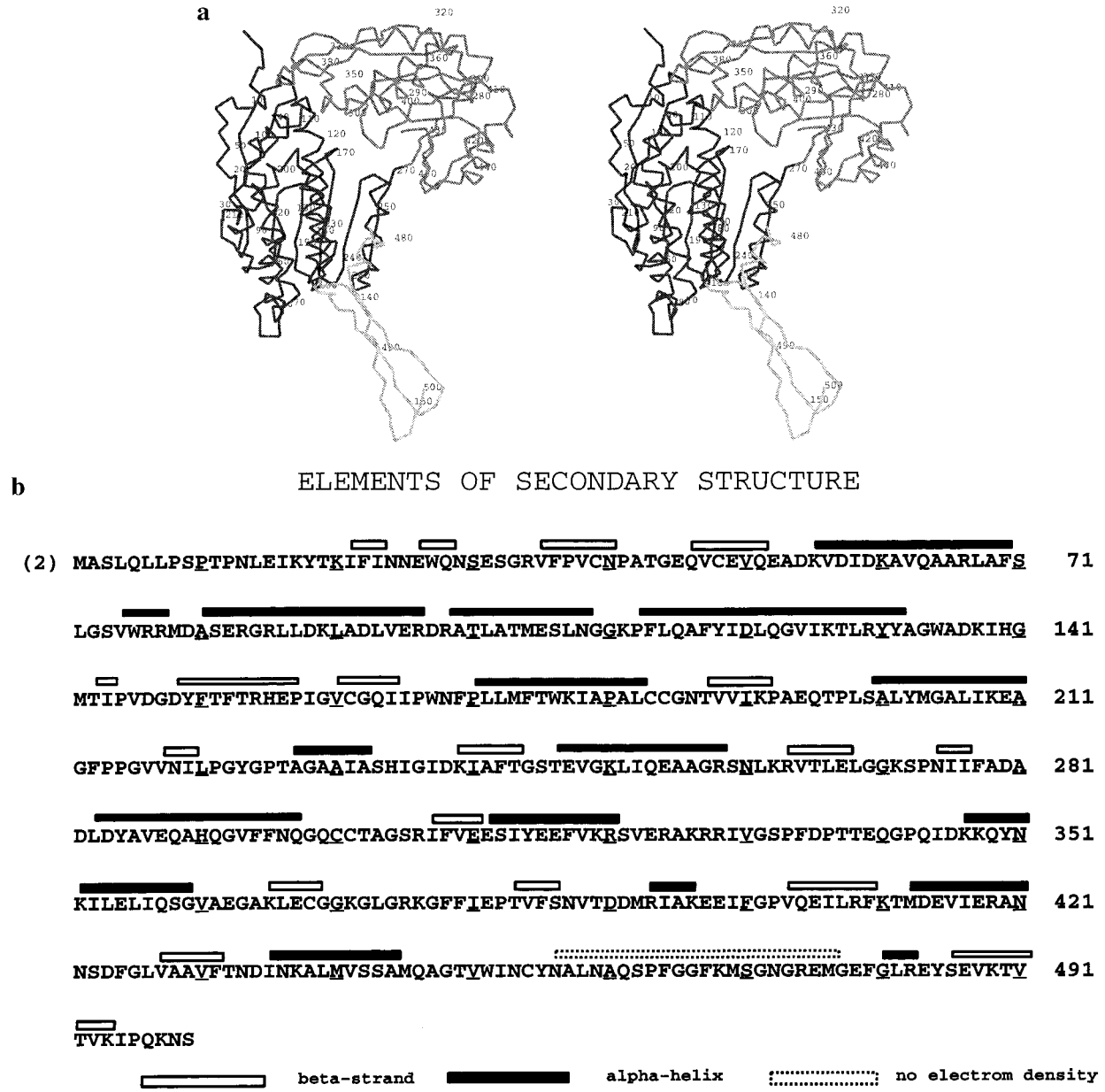

FIGURE 2: (a) Stereo Cα trace of a monomer of Retinal dehydrogenase type II. The nucleotide-binding domain is shown in black, the catalytic domain is gray, and the tetramerization domain is light gray. Every tenth residue is marked with a sequence number. (b) The RalDH2 sequence with elements of secondary structure indicated.

link in ALDH2 but not RalDH2. This intermonomer link, which involves Lys 424 and Glu 236 of the monomer mate, bridges the catalytic domain of A (C) with the nucleotide-binding domain of B (D). A more distal part of the mobile loop hairpins in the vicinity of Lys-424 in ALDH2. The main chains of ALDH2 and RalDH2 have larger structural differences in this region. Views of the substrate access channel can be seen in Figure 4a,b. The undetected loop has been modeled as it appears in ALDH2 in the figures. Note that the loop is located between the nucleotide-binding and catalytic domains and could represent a flexible linkage between the two domains, which may allow for rigid body domain motions.

Recently, the 2.3 Å resolution structure of a retinaldehyde metabolizing class I aldehyde dehydrogenase (ALDH1) has been reported (*33*). The loop region which is disordered in RalDH2 is ordered in ALDH1. In ALDH1 a larger substrate access channel relative to that of ALDH2, the consequence of substitutions to smaller side chains, reorientation of conserved side chains, and subtle rigid body movements of elements of secondary structure, is proposed to be the reason that ALDH1 can metabolize retinaldehyde. Why then, is the loop ordered in ALDH1 but not RalDH2? Again, key changes in the loop region of ALDH1 relative to RalDH2 which may allow for the positioning of the loop in ALDH1 can be identified. Tyr 132 in RalDH2 is a Cys in ALDH1. This amino acid is found where the loop hairpins and substitution to a much smaller side chain may allow the loop to be more readily accommodated. Substitution of the neighboring amino acid Leu-482 with a Phe in retinal metabolizing ALDH1 may also contribute to structural differences in the loop area.

As mentioned above, the binding cavity and substrate access channel are formed by amino acids that are contributed by all three domains: the nucleotide-binding, the catalytic, and the tetramerization domains. The residues that line the

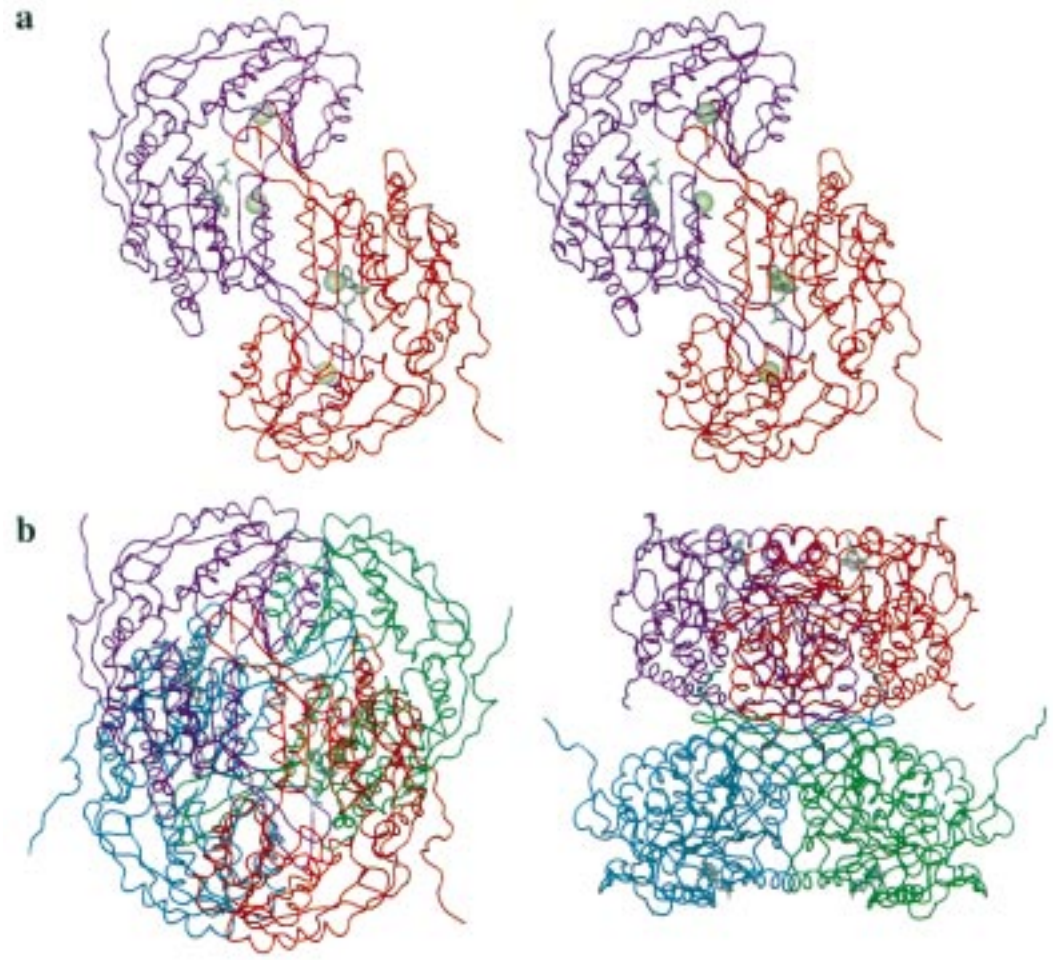

FIGURE 3: (a) The dimer of RalDH2. The green spheres mark the amino and carboxy termini of the substrate access channel loop. The ADP of the cofactor is rendered in ball-and-stick format. (b) A tetramer of RalDH2 in the same orientation as part (a). (c) Rotation of the tetramer 90° about the horizontal axis.

substrate-binding cavity for the dimeric ALDH3 as described by Wang and colleagues are marked with an asterisk (*) in Table 2. For a simple approximation of the difference in size of the substrate-binding cavity between RalDH2 and the class I, II, and III ALDH's, one could count the difference in nonhydrogen side-chain atoms in the residues that line the cavity. By this crude approximation, retinaldehyde metabolizing ALDH1 has by far the largest cavity, and that of RalDH2 is only slightly larger than that of ALDH2. Hence the Gly (460) to Asn substitution in RalDH2 may have more striking structural consequences than the substitution of Gly → Ser in ALDH1, as the larger cavity of the latter will readily accommodate the loop. The substrate access channel and binding cavity of ALDH3 are the smallest of the isozymes. The difference in chemical nature between the cavities of these proteins can also be identified. Between ALDH2 and RalDH2, there is one amino acid difference that causes a change in chemical nature: a charged amino acid in ALDH2 is a polar amino acid in RalDH2. The difference is more pronounced when one compares ALDH3 and RalDH2. Four polar residues and one charged residue in the class III enzyme are hydrophobic in RalDH2. As RalDH2 binds a larger and more hydrophobic ligand compared to the class II and III ALDH's, these differences in the substrate-binding cavity are what would be predicted.

The structures of the monomers of the aldehyde and betaine aldehyde dehydrogenases are superimposed in Figure 4c. As would be expected from the level of sequence identity, RalDH2 is more structurally similar to the tetrameric ALDH2 than ALDH3. The rms deviation for RalDH2 with the former is 0.94 Å (459 atoms) and for the latter 1.8 Å (340 atoms). The alignment of RalDH2 with ALDH2 is somewhat better in the nucleotide-binding domain than the catalytic domain (0.69 Å for 262 atoms vs 0.87 Å for 187 atoms). The converse is true for the alignment of RalDH2 with the class III enzyme (1.6 Å for 165 atoms of the catalytic domain vs 1.8 Å for 178 atoms for the nucleotide-binding domain). The related betaine aldehyde dehydrogenase, which has 40% sequence identity to RalDH2, has rms deviations of 1.2 Å overall (445 atoms) vs 1.3 Å for the nucleotide-binding domain (244 atoms) and 0.97 Å for the catalytic domain (186 atoms). The apparent better agreement for the catalytic domains of ALDH3 and betaine aldehyde dehydrogenase with RalDH2 is most likely simply a reflection of the fact that the number of core atoms compared is smaller for the catalytic domains.

The electron density for the catalytically active cysteine and the other residues in the vicinity is remarkably poor for RalDH2. Also, we are unable to see the catalytic portion of the NAD. It appears that all of the catalytic machinery is mobile, possibly awaiting substrate binding to induce a stable conformation. The fact that the *B* factors for the catalytic domains of RalDH2 are higher than those of the nucleotide-binding domains (average $\approx$ 64 Å^2 versus 37 Å^2) suggests

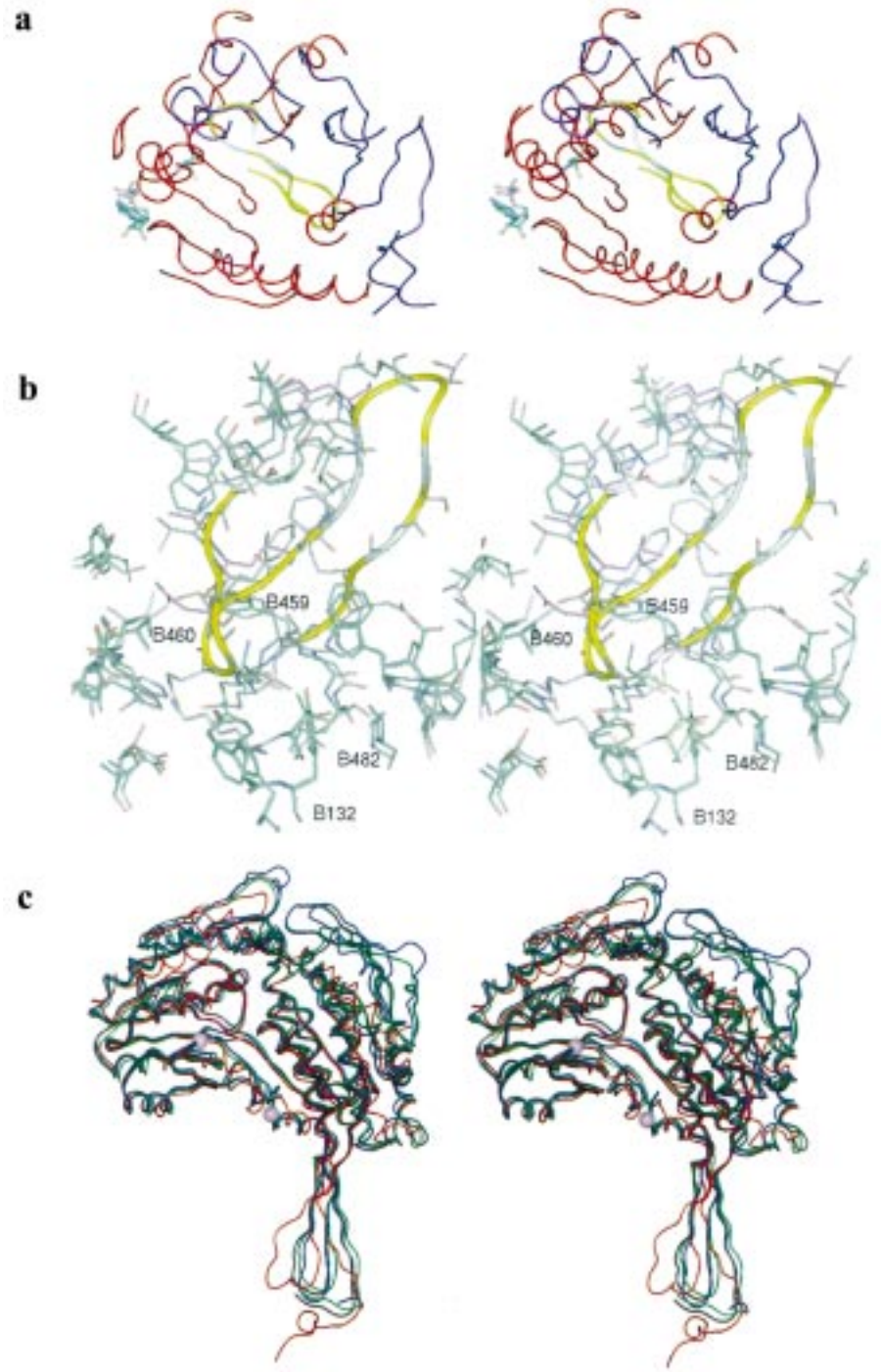

FIGURE 4: (a) The loop as positioned in ALDH2 (yellow, white indicates positions of glycines in RalDH2 sequence) in the context of a RalDH2 dimer (purple and red). The catalytic Cys and ADP of the NAD cofactor are depicted. (b) Detail of the vicinity of the loop. The structures of ALDH2 (gray, loop in yellow) and RalDH2 (green) are superimposed. The residues in the loop which are different in RalDH2 are modeled in pink. Two important differences are at positions 459 (ALDH2, Phe; RalDH2, Leu) and 460 (ALDH2, Gly; RalDH2, Asn). Key differences between RalDH2 and retinal metabolizing ALDH1 are at positions B132 and B482. (c) An overlay of monomers of RalDH2 (black), betaine dehydrogenase (blue), ALDH2 (green), and ALDH3 (red). The termini of the missing loop in RalDH2 are marked with spheres.

that the entire catalytic domain is more mobile than the rest of the structure. This mobility may be apparent because the domain does not participate in extensive inter-monomer contacts and there are limited crystal-packing contacts to anchor the domain in the lattice as well.

Steinmetz and colleagues (*31*) indicated in their discussion of the mitochondrial aldehyde dehydrogenase structure that electron densities for the nicotinamide ring and ribose were absent from electron density maps calculated with their native data as well. Similarly, Johanson et al. (*32*) noted that in the betaine aldehyde dehydrogenase structure the cofactor at these same regions had less electron density than the ADP portion of the cofactor, an observation consistent with flexibility of the nicotinamide ring and ribose. Furthermore, two possible orientations for the nicotinamide moiety of the cofactor were revealed in the ALDH1 structure (*33*). For ALDH2, only with a samarium derivative data set was it possible to see electron density for the nicotinamide ribose

Table 2. Alignment of the Sequences for Residues Which Line the Catalytic Channel

144	FSNAYLMDLGGCIKT	128	ALDH1
114	FLQAFYIDLQGVIKT	128	RalDH2
114	YIISYLVDLDMVLKC	128	ALDH2
56	EWTSYYEEVAHVLEE	69	ALDH3
	* *** ** **		
169	NFPLLMFLW	177	ALDH1
169	NFPLLMFTW	177	RalDH2
169	NFPLLMQAW	177	ALDH2
114	NYPFNLTIQ	122	ALDH3
	* ** *		
291	QGVFYAQGQCCI...L	426	ALDH1
291	QGVFFNQGQCCT...L	426	RALDH
291	FALFFNQGQCCC...L	426	ALDH2
233	WGKFMNSGQTCV...L	361	ALDH3
	* * * * *		
454	NCYSVVSAQCPF...EL	476	ALDH1
454	NCYNALNAQSPF...EL	476	RalDH2
454	NCYDVFGAQSPF...EL	476	ALDH2
390	NIVHITVPTLPF...EH	412	ALDH3
	** **** *		

[a] Asterisks (*) indicate the residues lining the channel in ALDH3.

and ring and to identify a possible orientation for this portion of the molecule (the bound samarium interacts directly with the phosphate groups of the NAD). In their discussion of the ALDH2 structure, the authors commented that the conformation of the nicotinamide ring as seen is not possible for the entire catalytic cycle, as it sterically hinders the functioning of the general base, Glu268 (or a water molecule activated by Glu268). It was suggested that the nicotinamide must change positions during the catalytic cycle. This suggestion would be consistent with the fact that the NAD is in a different conformation in ALDH3, or that the nicotinamide ring is mobile in RalDH2 in the absence of substrate. Furthermore, it was suggested that, since identical ALDH2 crystals grew in the presence or absence of NAD, the conformational changes of the protein which occur during the catalytic cycle are very minor. This may not be the case for RalDH2, as no crystals grew in any of the screened conditions in the absence of NAD. Steinmetz and colleagues (*31*) also suggested that the stabilization of the nicotinamide ring and ribose by the addition of samarium could reflect a physiologic interaction with the divalent cation as the concentration of Mg^{2+} is sufficiently high in the mitochondria to produce the transition from half-of-the-sites to full-sites activity seen in vitro. In addition, crystal growth of retinal metabolizing ALDH1 required the presence of Mg^{2+}. In contrast, activity assays conducted in the presence of divalent cation (Mg^{2+}) reduced the specific activity of RalDH2 by half (ALL, data not shown). The contrasting effects of Mg^{2+} on RalDH2 activity versus that of ALDH1 and ALDH2 suggest discrete differences in the catalytic mechanisms of these enzymes.

Retinal dehydrogenase type II is one of several dehydrogenases identified as enzymes which produce retinoic acid in vivo for regulation of gene expression in the developing embryo. Unlike in the homologous dehydrogenases which metabolize acetaldehyde, the catalytic machinery and substrate access channel in RalDH2 appear to be mobile. This mobility could provide a mechanism for discrimination between long-chain and short-chain aldehydes. Further studies to test the hypothesis that the binding of retinal is accompanied by an ordering of the substrate access channel and catalytic amino acids have been initiated.

ACKNOWLEDGMENT

The authors thank J. L. Napoli for his generous gift of the RalDH2 clone and his continued interest in the work.

REFERENCES

1. Wald, G. (1968) *Nature 219*, 800−807.
2. Wilson, J. G., and Warkany, J. (1949) *Am. J. Anat. 85*, 113−155.
3. Wilson, J. G., Roth, C. B., and Warkany, J. (1953) *Am. J. Anat. 92*, 189−217.
4. Lammer, E. J., and Shyan, C. S. (1985) *N. Engl. J. Med. 313*, 837−841.
5. Cohlan, S. Q. (1953) *Science 117*, 535−536.
6. Morriss, G. (1973) *J. Embryol. Exp. Morphol. 30*, 219−242.
7. Shenefelt, R. E. (1972) *Teratology 5*, 103−118.
8. Keeble, S., and Maden, M. (1989) *Dev. Biol. 132* (1), 26−34.
9. Maden, M. (1983) *Dev. Biol. 98*, 409−416.
10. Maden, M. (1983) *J. Embryol. Exp. Morphol. 77*, 273−295.
11. Tickle, C., Crawley, A., and Farrar, J. (1989) *Development (Cambridge, U.K.) 106*, 691−705.
12. Tickle, C., Alberts, B., Wolpert, L., and Lee, J. (1982) *Nature 296*, 564−566.
13. Tickle, C., Lee, J., and Eichele, G. (1985) *Dev. Biol. 109*, 82−95.
14. Thaller, C., and Eichele, G. (1987) *Nature 327*, 625−628.
15. Eichele, G., Tickle, C., and Alberts, B. M. (1985) *J. Cell Biol. 101* (5 Pt 1), 1913−1920.
16. Ruberte, E., Dolle, P., Chambon, P., and Morriss-Kay, G. (1991) *Development (Cambridge, U.K.) 111*, 45−60.
17. Dolle, P., Fraulob, V., Kastner, P., and Chambon, P. (1994) *Mech. Dev. 45*, 91−104.
18. Dolle, P., Ruberte, E., Kastner, P., Petkovich, M., Stoner, C. M., Gudas, L. J., and Chambon, P. (1989) *Nature 342*, 702−705.
19. Dolle, P., Ruberte, E., Leroy, P., Morriss-Kay, G., and Chambon, P. (1990) *Development (Cambridge, U.K.) 110*, 1133−1157.
20. Lohnes, D., Mark, M., Mendelsohn, C., Dolle, P., Dierich, A., Gorry, P., Gransmuller, A., and Chambon, P. (1994) *Development (Cambridge, U.K.) 120*, 2723−2749.
21. Mendelsohn, C., Lohnes, D., Decino, D., Lufkin, T., Lemeur, M., Chambon, P., and Mark, M. (1994) *Development (Cambridge, U.K.) 120*, 2749−2771.
22. Kastner, P., Grondona, J. M., Mark, M., Gransmuller, A., LeMeur, M., Decimo, D., Vonesch, J. I., Dolle, P., and Chambon, P. (1994) *Cell 78*, 987−1003.

23. Niederreither, K., McCaffery, P., Drager, U. C., Chambon, P., and Dolle, P. (1997) *Mech. Dev. 62* (1), 67−78.
24. Bavik, C., Ward, S. J., and Ong, D. E. (1997) *Mech. Dev. 69*, 155−167.
25. Hogan, B. L. M., Thaller, C., and Eichele, G. (1992) *Nature 358*, 237−241.
26. Lindahl, R. (1992) *Crit. Rev. Biochem. Mol. Biol. 27*, 283−335.
27. Jacoby, W. B., and Ziegler, D. M. (1990) *J. Biol. Chem. 265*, 20715−20718.
28. Lindahl, R., and Hempel, J. (1991) *Enzymol. Mol. Biol. Carbonyl Metab. 3*, 1−8.
29. Hempel, J., and Lindahl, R. (1989) *Enzymol. Mol. Biol. Carbonyl Metab. 2*, 3−17.
30. Liu, Z.-J., Sun, Y.-J., Rose, J., Chung, V.-J., Hsiao, C.-D., Chang, W. R., Kuo, I., Perozich, J., Lindahl, R., Hempel, J., and Wang, B. C. (1997) *Nat. Struct. Biol. 4* (4), 317−326.
31. Steinmetz, C. G., Xie, P., Weiner, H., and Hurley, T. D. (1997) *Structure 5* (5), 701−711.
32. Johansson, K., Mustapha, E.-A., Ramaswamy, S., Hjelmqvist, L., Jörnvall, H., and Eklund, H. (1998) *Protein Sci. 7*, 2106−2117.
33. Moore, S. A., Baker, H. M., Blythe, T., Kitson, K. E., Kitson, T. M., and Baker, E. N. (1998) *Structure 6*, 1541−1551.
34. Wang, X., Penzes, P., and Napoli, J. L. (1996) *J. Biol. Chem. 271* (27), 16288−16293.
35. Zhao, D., McCaffery, P., Ivins, K. J., Neve, R. L., Hogan, P., Chin, W. W., and Drager, U. C. (1996) *Eur. J. Biochem. 240* (1), 15−22.
36. Lamb, A. L., Wang, X., Napoli, J. L., and Newcomer, M. E. (1998) *Acta Crystallogr., Sect. D 54*, 639−642.
37. Minor, W. (1993) XDISPLAYF Program. Purdue University, West Lafayette, Indiana.
38. Otwinoswki, Z. (1993) in *CCP4 Study Weekend: Data Collection and Processing* (Sawyer, L., Issacs, N., Bailey, S., Eds.) pp 56−62, SERC Daresbury Laboratory, Warrington, U.K.
39. Navaza, J. (1994) *Acta Crystallogr., Sect. D 50*, 157−163.
40. Collaborative Computing Project, No. 4 (1994) *Acta Crystallogr., Sect. D 50*, 760−763.
41. Brunger, A. T. (1992) X-PLOR: A system for X-ray and NMR. Yale University Press, New Haven, CT.
42. Brunger, A. T. (1992) *Nature 355*, 472−475.
43. Furey, W., and Swaminathan, S. (1998) *Methods Enzymol. 227* (part B), 590−620.
44. Jones, T. A., Cowan, S., Zou, J.-Y., and Kjelgaard, M. (1991) *Acta Crystallogr., Sect. A 47*, 110−119.
45. Rice, L. M., and Brunger, A. T. (1994) *Proteins: Struct., Funct., Genet. 19*, 277−290.
46. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *Appl. Crystallogr. 26*, 283−291.
47. Cowan, S. W., Newcomer, M. E., and Jones, T. A. (1990) *Proteins 8*, 44−61.
48. Renaud, J.-P., Rochel, N., Ruff, M., Vivat, V., Chambon, P., Gronemeyer, H., and Moras, D. (1995) *Nature 378*, 681−689.
49. Johnston, S. C., Larsen, C. N., Cook, W. J., Wilkinson, K. D., and Hill, C. P. (1997) *EMBO J. 16*, 3787−3796.

BI9900471