Enzymatic Ketonization of 2-Hydroxymuconate: Specificity and Mechanism Investigated by the Crystal Structures of Two Isomerases†

H. S. Subramanya,‡,§ D. I. Roper,§,|| Z. Dauter,||,⊥ E. J. Dodson,|| G. J. Davies,|| K. S. Wilson,||,⊥ and D. B. Wigley*,‡

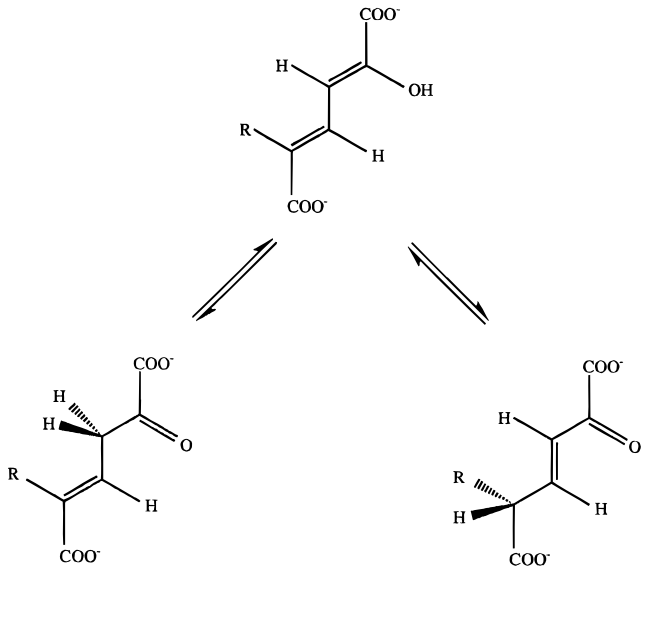
Laboratory of Molecular Biophysics, Rex Richards Building, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K., Department of Chemistry, University of York, Heslington, York YO1 5DD, U.K., and EMBL Outstation, c/o DESY, Notkestrasse 85, 22603 Hamburg, Germany

Received July 26, 1995; Revised Manuscript Received November 17, 1995⊗

ABSTRACT: 5-Carboxymethyl-2-hydroxymuconate isomerase (CHMI) and 4-oxalocrotonate tautomerase (4-OT) are enzymes that catalyze the isomerization of unsaturated ketones. They share a common enzyme mechanism, although they show a preference for different substrates. There is no apparent sequence homology between the enzymes. To investigate the molecular mechanism and the basis for their substrate specificity, we have determined the crystal structures of the two enzymes at high resolution. 4-OT is hexameric, with the subunits arranged with 32 symmetry. CHMI is trimeric and has extensive contacts between subunits, which include secondary structural elements. The central core of the CHMI monomer has a fold similar to a 4-OT dimer, but the secondary structural elements that form the subunit contacts around the 3-fold axis are different in the two enzymes. The region of greatest similarity between the two enzymes is a large pocket that is proposed to be the active site. The enzymes appear to operate via a "one-base" mechanism, and the possible role of residues in this pocket is discussed in view of this idea. Finally, the molecular basis for substrate specificity in the two enzymes is discussed.

Microbial degradation of aromatic carbon sources to the intermediates of the citric acid cycle proceeds by the cleavage of the aromatic nucleus followed by a series of reactions. Meta-fission pathways for the degradation of the aromatic carbon sources have been studied in a number of bacterial strains, but most extensively in the pseudomonad species. The chemistry and enzymology of several enzymes of this pathway are being studied by various methods. The entire pathway in the *Pseudomonas* sp CF600 strain is encoded by a naturally occurring megaplasmid analogous to that encoded by the TOL plasmid in *Pseudomonas putida* (Harayama & Rekik 1990). 4-Oxalocrotonate tautomerase (4-OT) is an enzyme in this pathway and has been shown to catalyze the ketonization of 2-oxo-4-hexene-1,6-dioate, via 2-hydroxymuconate, to yield an α,β-unsaturated ketone, 2-oxo-3-hexene-1,6-dioate (Whitman *et al.*, 1991) (Figure 1).

4-OT is a hexameric enzyme (Roper *et al*., 1994) consisting of subunits of only 62 amino acid residues, which is the smallest enzyme subunit reported to date (Chen *et al.*, 1992). It is a highly efficient catalyst with k_{cat}/K_M values approaching the diffusion-controlled limit. Recent kinetic studies have shed some light on the possible mechanism for the enzyme catalysis (Lian & Whitman, 1993). They have also shown that 4-OT catalyzes the isomerization of 2-hydroxymuconate stereoselectively and proceeds by a one-base mechanism.



R = H or CH_2COO^-

FIGURE 1: Preferred substrates and reactions catalyzed by 4-OT and CHMI. The side group, R, is H in the case of 4-OT and CH_2COO^- for CHMI. The "forward" direction of the reaction is from left to right.

However, the molecular details of the enzyme mechanism still need to be explored.

In the homoprotocatechuate pathway of *Escherichia coli* C, the necessary enzymes are chromosomally encoded, and the genes are closely linked in a single regulated operon (Roper *et al*., 1993). One of the enzymes involved with this pathway (5-carboxymethyl-2-hydroxymuconate isomerase, CHMI) catalyzes the isomerization of 5-(carboxymethyl)-2-hydroxymuconate (CHM) to 5-(carboxymethyl)-2-oxo-3-hexene-1,6-dioate (COHED). Initially, it was proposed that

† This work was partially supported by an EMBO short-term fellowship (D.B.W.). This project was initiated by D.I.R. as a BBSRC-funded student in the laboratory of R. A. Cooper. D.B.W. is a BBSRC Advanced Fellow.

* Author to whom correspondence should be addressed.
‡ University of Oxford.
§ These authors contributed equally to this work.
|| University of York.
⊥ EMBL Outstation.

⊗ Abstract published in *Advance ACS Abstracts,* January 1, 1996.

0006-2960/96/0435-792$12.00/0

Table 1: Crystal Forms

type	space group	cell dimensions (Å)	V_m (Å^3/Da)	subunits per asymm unit	diffraction limit (Å)
4-OT					
	triclinic, *P*1	$a = 39.6$ $b = 51.5$ $c = 51.6$ $\alpha = 60.0°$ $\beta = 81.4°$ $\gamma = 69.6°$		6	1.9
CHMI					
I	orthorhombic,	$a = 88$			
	$I\,2_12_12_1$	$b = 89$	2.82	3	3.0
	or $I\,222$	$c = 121$			
II	cubic, $P4_132$	$a = b = c = 137$	2.07	1	8.0
III	tetragonal, $P4_12_12$	$a = b = 90.3$			
		$c = 129.8$	3.15	3	2.1

Table 2: Data Collection and Processing

data	resolution (Å)	no. refs	R_{merge}[a] (%)	completeness (%)	anomalous measurements
4-OT					
native	1.9	24515	5.9	93.6	
K_2PtCl_4	2.8	7657	6.6	96.8	yes

data	resolution (Å)	R_{merge} (%)	no. refs	completeness (%)
CHMI				
native	2.1	4.9	28779	89.4
K_2PtCl_4	3.7	4.3	6063	98.6
$NaAuCl_4$	3.7	3.6	6045	98.3
$NaAuCl_4$	2.6	5.4	16757	98.2

[a] $R_{merge} = \sum\sum_j |I_j - \langle I \rangle| / \sum |\langle I \rangle|$, where I_j is the intensity measurement for the reflection j and $\langle I \rangle$ is the mean intensity for multiply recorded reflections.

the conversion of CHM to COHED was direct, but now it seems likely that the mechanism is more complex and also involves an initial conversion of CHM to 5-(carboxymethyl)-2-oxo-4-hexene-1,6-dioate that is not catalyzed by the enzyme. The enzyme then converts 5-(carboxymethyl)-2-oxo-4-hexene-1,6-dioate to COHED via CHM (Hajipour *et al*., 1993; Figure 1). The chemical mechanism of the reaction has been analyzed in solution in both the presence and absence of the enzyme and has been shown to follow essentially the same pathway. Thus, the enzyme appears to operate by stabilizing the same transition states as those that occur in the uncatalyzed reaction.

CHMI catalyzes a reaction analogous to 4-OT on a chemically similar substrate (Roper & Cooper, 1990). The two isomerases are specific and each shows a preference for its own substrate, although each enzyme is able to utilize both substrates (Whitman *et al*., 1992). Furthermore, the catalysis has been shown to proceed with the same stereospecificity, indicating a similarity between the mechanisms of the two enzymes (Whitman *et al*., 1992). This has led to speculation about the evolution of these two enzymes. However, CHMI is reported to be a dimer with subunits of 126 amino acids (Roper & Cooper, 1990), while 4-OT is reported to be either a pentamer (Davenport & Whitman, 1993) or a hexamer (Roper *et al*., 1994) with subunits of 62 amino acids. Furthermore, the two enzymes do not share any apparent sequence similarity. To investigate this apparent paradox, we have solved the structures of 4-OT and CHMI with the idea of comparing them to determine whether this will shed light on the mechanism and evolutionary lineage of the two enzymes. We are also seeking to discover the molecular basis for substrate specificity and, in particular, how the enzymes discriminate against the wrong substrate. This information will be useful for the design of enzymes with novel specificities.

MATERIALS AND METHODS

Structure Determination of 4-OT. Purification and crystallization conditions for the protein have been described previously (Roper *et al.*, 1994). Several crystal forms were obtained under similar conditions. However, the crystals used for the structure determination were of the *P*1 space group (Table 1). Crystals were harvested into 80% saturated ammonium sulfate solution containing 100 mM HEPES (pH 7.5) and 1% PEG 200 (harvesting solution) at 20 °C. The platinum derivative was obtained by soaking the crystals for 4 h in harvesting solution containing 2 mM K_2PtCl_4.

All X-ray diffraction data were collected by using a rotating anode X-ray generator and a MAR image plate. Native data were collected initially to a resolution of 2.4 Å using a single crystal. The same crystal was used for heavy atom soaking experiments and to collect a 2.8 Å resolution K_2PtCl_4 derivative data set. Finally, native data were collected to a resolution of 1.9 Å using another crystal.

Data sets were indexed and integrated by using the program DENZO (Otwinowski, 1993). Subsequent data reduction and processing were carried out using the programs from the CCP4 suite of programs (CCP4, 1994). Unless otherwise stated, the programs used for the structure determination were from the CCP4 package. Table 2 shows the results of data processing.

Native and derivative data were scaled together by using anisotropic scaling (Table 3). Difference Patterson maps computed in different resolution bins showed peaks significantly above the noise level. These peaks were found to be

Table 3: Phasing Statistics[a]

derivative	resolution (Å)	MFID[b] (%)	no. sites	occupancy (absolute scale)	phasing power
4OT					
K_2PtCl_4	2.8	20.7	6	0.451	1.3
				0.435	
				0.429	
				0.387	
				0.305	
				0.300	

derivative	resolution (Å)	MFID (%)	no. sites	Cullis *R* (%)	occupancy (absolute scale)	phasing power
CHMI						
K_2PtCl_4	3.7	26.6	3	73.0	0.80	1.10
					0.73	
					0.70	
$NaAuCl_4$	3.7	9.0	2	80.5	0.27	0.70
					0.19	
$NaAuCl_4$	2.6	7.0	2	97.0	0.08	0.30
					0.06	

[a] Heavy atom *B*-factors were fixed at 25 Å^2. [b] MFID $= \sum||F_{ph}| - |F_p||/\sum|F_p|$, where F_{ph} and F_p are the derivative and native structure factors, respectively.

related by the expected noncrystallographic symmetry (Roper *et al.*, 1994). Since the choice of origin is arbitrary in this space group, reflections were phased initially by using a single position at the origin. Five other positions with respect to the heavy atom position at the origin were determined by using difference Fourier methods. Electron density maps were calculated by using phases calculated from the heavy atom positions and their inverses. One of these maps clearly was superior and was used for the subsequent density manipulation procedures.

The heavy atom parameters were refined and the results are shown in Table 3. It is clear that the phasing power remained good to the resolution limit of the data. The refined heavy atom parameters were used along with the anomalous information to calculate a set of SIRAS phases to 2.8 Å, and an initial electron density map was calculated by using these phases. Although the map showed some secondary structural elements, it was uninterpretable at this stage. However, the electron density map was of sufficient quality to reveal the structural similarity between 4-OT and CHMI (see the following).

Solvent flattening was attempted as a first step toward phase improvement. A mask was calculated from a 6 Å electron density map and was used to solvent flatten the 2.8 Å map. However, the resulting map did not show any significant improvement. This is probably due to the relatively low solvent content of the crystals.

Since the self-rotation functions and the heavy atom positions clearly revealed the noncrystallographic symmetry of the molecule (Roper *et al.*, 1994), real space electron density averaging was carried out. The six noncrystallographic symmetry operators were derived from a least-squares fit of the heavy atom positions. Because of the structural similarity between 4-OT and CHMI (see the following), a mask was computed by using the Cα coordinates of the CHMI structure with a radius of 5 Å for the atoms. The RAVE program package (Jones, 1992) was used for averaging. The phases from the averaged map were combined with the initial SIRAS phases, and a new map was computed by using these phases. Convergence was achieved after several cycles of averaging. The resulting map showed significant improvements and allowed partial model building.

Table 4: Deviations of the Final Models from Ideal Geometry

parameter	4-OT	CHMI
distances (Å)		
bonds (1-2 neighbor)	0.014	0.015
angles (1-3 neighbor)	0.047	0.051
interplanar (1-4 neighbor)	0.050	0.060
planar groups (Å)	0.024	0.013
chiral centers (Å^3)	0.039	0.136
torsion angles (deg)		
planar (0, 180)	4.82	2.53
staggered (±60, 180)	20.39	21.46
orthonormal (±90)	42.83	22.92
nonbonded contacts (Å)		
single torsion	0.195	0.206
multiple torsion	0.272	0.225
possible H-bond	0.186	0.210
thermal factors (Å^2)		
main chain bond	1.49	0.91
main-chain angle	2.33	1.63
side-chain bond	2.57	3.62
side-chain angle	4.18	5.64

A partial model (comprising 50% of the protein) was built into the electron density of one of the subunits. The graphics program TURBO FRODO (Roussel & Cambillau, 1989) was used for model building. The noncrystallographic symmetry was used to generate the other five subunits. This initial model was refined by the restrained least-squares minimization procedure (Hendrickson & Konnert, 1985), applying noncrystallographic restraints. The refined model was used for phase combination with the initial SIRAS phases to generate a set of improved phases. Several subsequent models were built and refined, each containing more of the protein. The model at this stage contained the first 53 residues from each subunit. The density for the last few residues was rather poor in all subunits. At this point, the structure was refined further by using XPLOR (Brunger *et al.*, 1989). All data between 10 and 1.9 Å resolution were used for this procedure, and strict noncrystallographic symmetry constraints were applied. The structure was first energy minimized and then heated at 3000 K for 1000 steps of 1 fs, cooled at 300 K for 500 steps of 1 fs, and finally energy minimized again. Maps calculated by using the refined phases by this procedure showed great improvement. This allowed the model building of four more residues at the C-terminus. However, there was still no significant density for the last three residues. At this stage, water molecules were included in the structure. The final model was refined by using a least-squares refinement procedure by gradually reducing the noncrystallographic restraints. In the final stages of refinement, restrained *B*-factor refinement was also undertaken. The final model contained 60 residues from each subunit (360 residues in total) and 140 water molecules (Table 5). This model has an *R*-factor of 20.2%, using all data between 10 and 1.9 Å resolution. The deviations of the model from ideal geometry are presented in Table 4. A Ramachandran plot (Ramakrishnan & Ramachandran, 1965) reveals that there are no residues (other than glycines) in disallowed regions (data not shown).

Structure Determination of CHMI. Protein purification (Roper & Cooper, 1990) and crystallization conditions (Wigley *et al.*, 1989) have been described previously. A

Table 5: Refinement Statistics

parameter	4-OT	CHMI
resolution range (all data used) (Å)	10−1.9	10−2.1
R-factor (%)	20.2	17.9
no. residues	360	375
no. water molecules	140	201
other atoms included	none	7 sulfate ions
PDB entry code	1OTF	1OTG

number of different crystal forms were obtained under similar conditions (Table 1), but the tetragonal form proved to be the most suitable for X-ray analysis. Crystals were harvested into 40% saturated ammonium sulfate solution containing 100 mM Tris Cl (pH 7.5) (harvest solution) at 30 °C.

Derivatives were prepared by soaking harvested crystals in solutions of heavy atom compounds in harvest solution for 18 h at 30 °C. The platinum derivative was obtained by soaking crystals in K_2PtCl_4 solutions. After soaking in 10 mM platinum reagent, the crystals ceased to diffract. However, crystals survived soaking in 1 mM K_2PtCl_4 and still diffracted strongly. Screened precession photographs indicated large differences in the diffraction pattern compared to that produced by native crystals, but also indicated a change in the *c*-axis of about 1% (increasing from 129.8 to 131.0 Å). The gold derivative was prepared in a similar way. Soaking in 10 mM $NaAuCl_4$ again resulted in a total loss of diffraction, but diffraction was still observed after soaking the crystals in lower concentrations (1 and 2 mM).

All data were collected by using 1.0 Å wavelength synchrotron radiation and an image-plate scanner at the EMBL Outstation in Hamburg, Germany. Native data were collected from three crystals to 2.1 Å resolution, with one crystal being used to collect complete low-resolution (3.7 Å) data, which were merged with higher resolution (2.1 Å) data collected from the other two crystals. This strategy allowed the accurate determination of strong reflections that were overloaded in the high-resolution data. For data from the two heavy atom derivatives, the anomalous scattering measurements were retained. Data from the K_2PtCl_4 derivative were collected to 3.7 Å resolution from a single crystal. The $NaAuCl_4$ derivative proved to have two sites (see the following), both of low occupancy. Attempts to increase the occupancy of these sites by increasing the $NaAuCl_4$ concentration resulted in a loss of diffraction by the crystals. However, 2.6 Å data were collected for this derivative at a low concentration of $NaAuCl_4$ (from a single crystal, by combining 3.7 and 2.6 Å data sets; AU1) and 3.7 Å data from a single crystal at a higher concentration of the gold reagent (AU2).

Data were integrated with profile fitting using the MOSFLM package (Leslie, 1992) and subsequently processed by using programs from the CCP4 suite of programs (CCP4, 1994). Unless otherwise stated, all of the programs described herein were from the CCP4 package.

Derivative and native data were scaled together by using anisotropic scaling (Table 3). Two things were evident at this stage. Firstly, the overall mean fractional isomorphous difference (MFID) was much higher for the platinum derivative data than for either of the gold data sets. For the high-resolution gold data (AU1), the MFID was very low (<5%) even to medium resolution, but then gradually increased at higher resolution. Secondly, analysis of MFID with respect to indices for the platinum data showed that MFID was fairly constant with resolution for the *h* and *k* indices, but showed a rapid increase with respect to the *l* index beyond about 4 Å spacing, suggesting a serious lack of isomorphism. This is consistent with the 1% increase in the unit cell parameter, *c*, observed for the platinum derivative. Consequently, higher resolution data were not collected for this derivative.

Heavy atom positions for the two derivatives were determined by a combination of direct methods, difference and anomalous Patterson syntheses, and difference Fourier methods. Inspection of Harker sections of the anomalous Patterson map for the platinum derivative revealed three significant peaks corresponding to three sites per asymmetric unit. The 1% alteration in the *c*-axis indicated a lack of isomorphism, and this proved sufficient to make the difference Patterson uninterpretable. However, because of the high quality of the anomalous Patterson map, the anomalous measurements could be used to determine the heavy atom positions by using MULTAN (Wilson, 1978). A few terms with abnormally high *E* values were excluded from the calculations to reduce bias, and only data to 6 Å resolution were used.

Anomalous and difference Patterson maps calculated by using the gold derivative data (AU1 and AU2) also had clear peaks on Harker sections, and in this case the difference Patterson maps were more similar to the anomalous Patterson maps, suggesting that this derivative was more isomorphous. However, the peak heights were rather low, suggesting a low level of substitution. Difference Fourier maps, using phases calculated from the platinum derivative, indicated two sites per asymmetric unit.

The heavy atom parameters were refined, and these results are presented in Table 3. Parameters for the gold derivative data, at the two concentrations, were refined independently against the platinum and native data to reduce bias. In the final phasing cycle, the refined heavy atom parameters for all three derivative data sets were used to produce a set of MIR phases to 2.6 Å. It is clear from Table 3 that the phasing power for the platinum derivative was good in spite of the observed lack of isomorphism, probably due to the high level of substitution. It is also evident that the occupancy of the gold sites was considerably lower than that of the platinum sites, in both AU1 and AU2. In the case of AU1, the occupancies of the two sites correspond to only 5 and 4 electrons, respectively. Despite this, electron density maps were calculated at 3.7 and 2.6 Å resolution. Although some secondary structural elements were clearly visible, both maps were uninterpretable; however, the 2.6 Å map did show more detail than the 3.7 Å map, suggesting that some useful phase information could be obtained from the higher resolution derivative data (AU1).

Solvent flattening was used to extend the phasing from 3.7 to 2.6 Å resolution in steps of 0.1 Å, combining with the MIRAS phases to a resolution of 3.7 Å and the SIRAS gold phases (from AU1) at each stage. This resulted in a map that was interpretable and allowed a partial model (comprising ∼40% of the protein) to be built. Although the noncrystallographic symmetry was known approximately by this stage, it was not utilized to improve the phasing. Rather, the density for each of the three subunits was rotated to place each subunit in a similar orientation. Pieces of the model were built by inspection of all three maps. Cyclic averaging was not necessary.

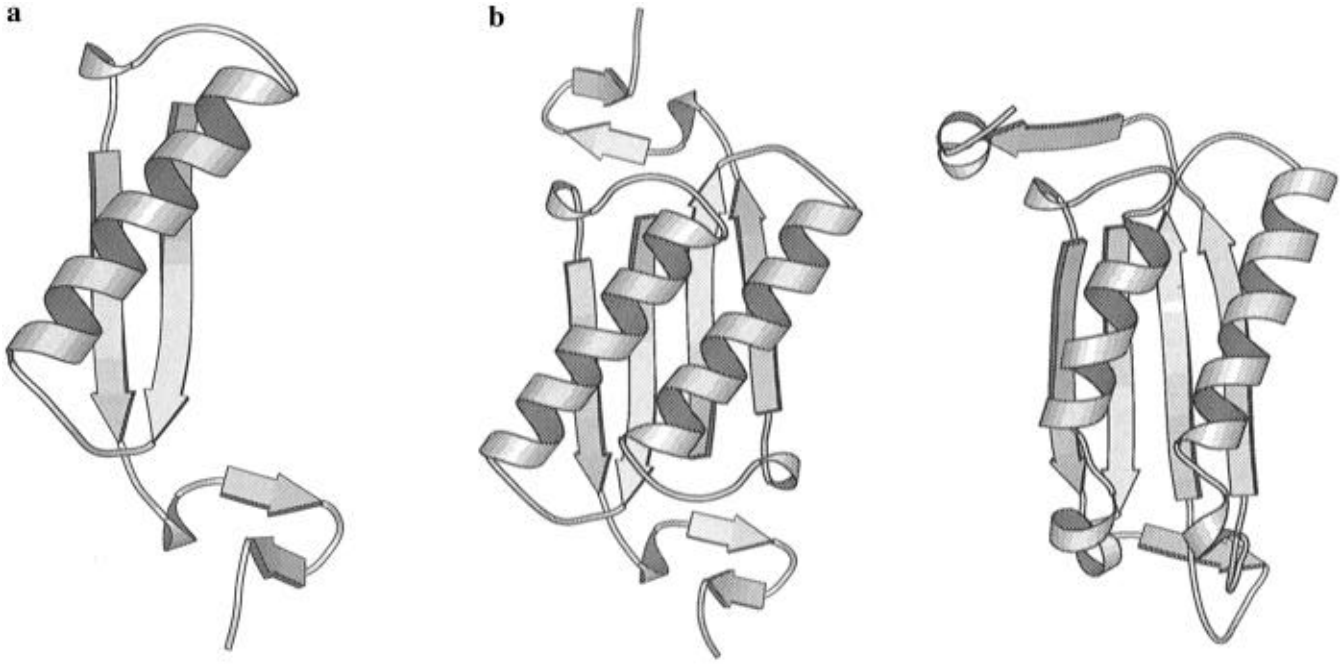


FIGURE 2: (a) Representation of the fold of the 4-OT monomer. (b) Representation of the fold of the 4-OT dimer (left) and the CHMI monomer (right).

Once each round of model building was completed, the local symmetry was applied to the model to generate the other subunits of the trimer. Initial cycles of refinement used the restrained least-squares minimization procedure (Hendrickson & Konnert, 1980), applying noncrystallographic restraints (strict NCS for the main chain, but less strict for side chains). This model (*R*-factor = 48%) was then used for phase combination at 2.6 Å resolution to improve the phases. Several subsequent models were built and refined, each containing more of the protein, until the entire chain had been built (at least as polyalanine) and chain connectivity was certain. As described earlier, initial rounds of refinement, using incomplete models, were carried out by restrained least-squares minimization at 2.6 Å resolution. Once the main chain tracing was complete and unambiguous, a round of refinement was carried out by using XPLOR (Brunger *et al*., 1989), with the resolution of the data increased to 2.1 Å. For this procedure, strict noncrystallographic symmetry constraints were applied. The structure was first energy minimized and then heated at 3000 K for 1000 steps of 1 fs, cooled at 300 K for 500 steps of 1 fs, and finally energy minimized again. The *R*-factor before the XPLOR procedure was 39% and only fell to 34%. However, both the geometry of the structure and the quality of the density maps were greatly improved. Refinement was completed by cycles of restrained least-squares minimization and model building by using $2F_o - F_c$ and $F_o - F_c$ electron density maps to fit the solvent and correct some side chain conformations. The final model contained all 375 residues of the CHMI trimer, 201 water molecules, and 7 sulfate ions (Table 5). This model has an *R*-factor of 17.9% using all data between 10 and 2.1 Å resolution. The geometry of the final model is given in Table 4. A Ramachandran plot (Ramakrishnan & Ramachandran, 1965) reveals that only one non-glycine residue (Ala-53) lies in an unfavorable region (data not shown). This residue is in a well-defined turn in the structure.

RESULTS

Overall Structure of 4-OT. The ribbon diagram of one of the subunits of 4-OT is shown in Figure 2a. The structure of the monomer comprises four β-strands and a single α-helix. The overall structure is essentially a $\beta\alpha\beta$-motif, with two additional β-strands at the C-terminus folding back on themselves. The first two β-strands are involved in extensive contacts with a 2-fold related subunit of the molecule, forming a four-stranded β-sheet structure across the subunits (Figure 2b). Despite previous suggestions that the enzyme is pentameric (Davenport & Whitman, 1993), it is in fact hexameric, consistent with preliminary crystallographic analysis (Roper *et al*., 1994). The enzyme dimers are arranged around a 3-fold rotation axis to form a hexamer with 32 symmetry (Figure 3). The contacts around the 3-fold axis are less extensive than those between the subunits related by the 2-fold axes and mainly involve the two β-strands at the C-terminus of each subunit. The structure is also stabilized by extensive molecular contacts within the hydrophobic core of the hexamer.

Overall Structure of CHMI. The structure of the monomer (Figure 2b) comprises a central core of a four-stranded β-sheet with two α-helices running over the top. The arrangement can be subdivided into two adjacent $\beta\alpha\beta$-units, which run in opposite directions. This is a relatively common motif, and comparison of the fold of the CHMI monomer with that of other known structures reveals similarities to a number of other proteins, but most strikingly with 4-OT, although the fold of the CHMI monomer is equivalent to that of a 4-OT dimer (Figure 2b). This is unexpected since there is no detectable sequence homology between CHMI and 4-OT. CHMI proved to be a trimer rather than a dimer, as suggested previously (Roper & Cooper, 1990). The subunit arrangement in the CHMI trimer is shown in Figure 3. The subunit contacts are quite extensive and involve secondary structural elements. Each monomer contributes a β-strand to both of its neighbors, such

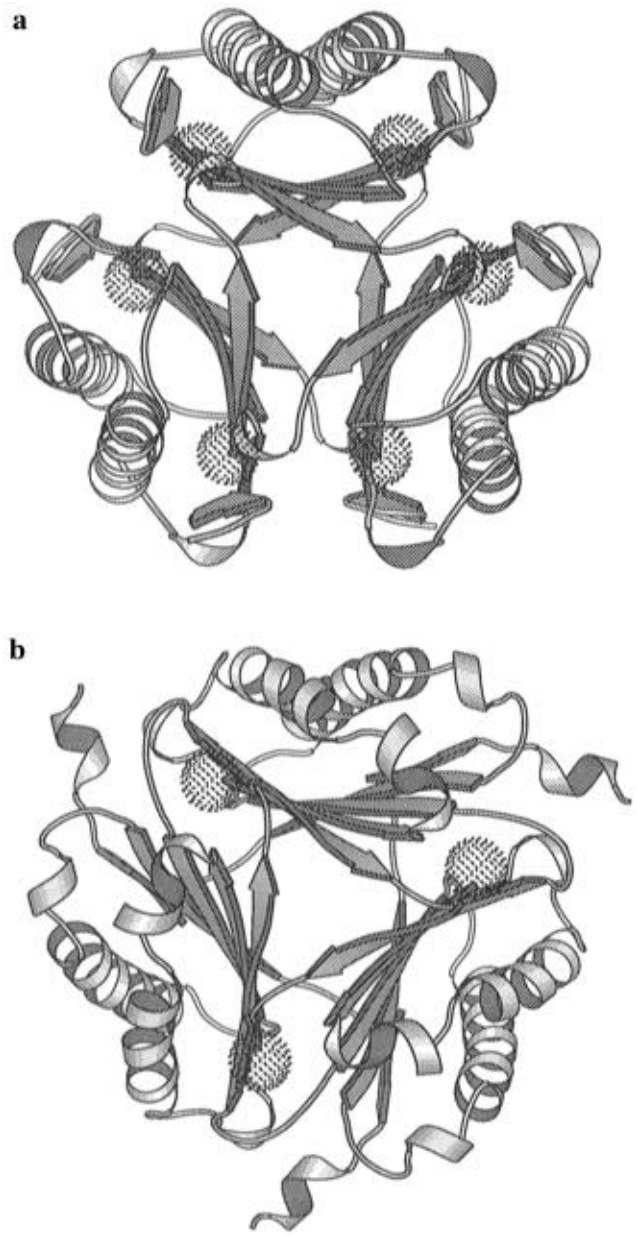


FIGURE 3: Subunit arrangement in the 4-OT hexamer (a) and the CHM isomerase trimer (b). The positions of the active sites are indicated by dotted spheres centered on the N-terminal proline residues of each protein subunit.

that each of the six-stranded β-sheets comprises β-strands from all three subunits.

Comparison of the Structures of 4-OT and CHMI. The overall structure of the CHMI trimer is remarkably similar, at first sight, to that of the 4-OT hexamer (Figure 3). The subunits in both molecules are arranged around a 3-fold axis with the active sites on the outside of the molecule. One interesting difference between the structures is that because of the higher symmetry of the 4-OT hexamer, there are potentially six active sites in 4-OT, yet only three of these are conserved in CHMI. The pseudosymmetry of the CHMI monomer is not sufficient to produce the additional three active sites, and there is no evidence for gene duplication in the CHMI monomer amino acid sequence (unpublished data). There is a cavity running up the 3-fold axis of the CHMI trimer, which indicates that the structure is stabilized largely through the interactions between the β-strands rather than other direct contacts between the subunits. This is in contrast to 4-OT, where the subunit contacts around the 3-fold axis are rather different. The contacts between the 4-OT subunits around the 2-fold axes are essentially the same as those within the central core of a CHMI monomer, mainly consisting of the interactions between adjacent antiparallel $\beta\alpha\beta$-units. However, the stabilization of the overall structure around the 3-fold axis in 4-OT apparently is mainly through hydrophobic interactions on the 3-fold axis, which comprises residues that are much more closely packed and nonpolar than the equivalent residues in CHMI. There is an additional small contact region between each of the subunits that, as in CHMI, involves the C-terminal residues, but this is very much less extensive in 4-OT than in CHMI. Whether the similarity between the structures of these enzymes is a consequence of convergent or divergent evolution is not clear. In spite of the oligomeric nature of both enzymes, there is no evidence of cooperativity in either case.

Location of the Active Site Region. The electron density in the CHMI structure has been interpreted to include seven sulfate ions bound to the trimer. One of these is positioned on the molecular 3-fold axis and is surrounded by histidine side chains from each of the three subunits. A similar arrangement of histidine residues has been observed around the 3-fold axis of insulin (Baker *et al*., 1988), but the site is occupied by a metal ion in the insulin crystal. Whether this site can be occupied by a metal ion in CHMI is not known, although the enzyme does not appear to have any metal requirements for activity (D. I. Roper, unpublished data).

The other sulfate ions were bound in three pairs, one pair to each subunit. In each case, the two ions were found in a pocket on the enzyme surface at one end of the β-strands and close to the N-terminus of the protein (Figure 4). The sites are surrounded by positively charged residues, consistent with the fact that the substrate (Figure 1) is highly negatively charged at pH 7.5. The pocket is certainly large enough to accommodate the substrate, and the distance between the sulfate ions (6.7 Å) is similar to that between two of the carboxylate groups of the substrate (6.0 Å). However, attempts to soak crystals in substrate have failed to yield a complex, possibly because of the high ionic strength required for stabilization of the crystals, and there are no substrate analogues available for cocrystallization experiments at present.

Examination of the molecular surface of the 4-OT structure reveals that there is only one cavity on the enzyme surface of each subunit, and this is in a location similar to the sulfate binding pocket in CHMI. There are a number of other reasons that lead us to conclude that this pocket is the active site of the enzyme. Firstly, enzyme active sites are usually in solvent accessible crevices or cavities on the surface of the protein. Secondly, the active site has to bind the substrate, which is highly negatively charged. To do this, the active site will have to carry a compensatory positive charge. Calculations using GRASP (Nicholls & Honig, 1991) reveal that the proposed active site does indeed carry a net positive charge. Thirdly, the reaction catalyzed by the enzyme involves proton transfer, implying that at least one suitable base has to be present to catalyze the reaction (see the following). Perhaps most convincing is the observation that the region of greatest similarity to the related enzyme, 5-carboxymethyl-2-hydroxymuconate isomerase, is in this pocket. Since there is practically no sequence homology between these enzymes, the degree of structural similarity

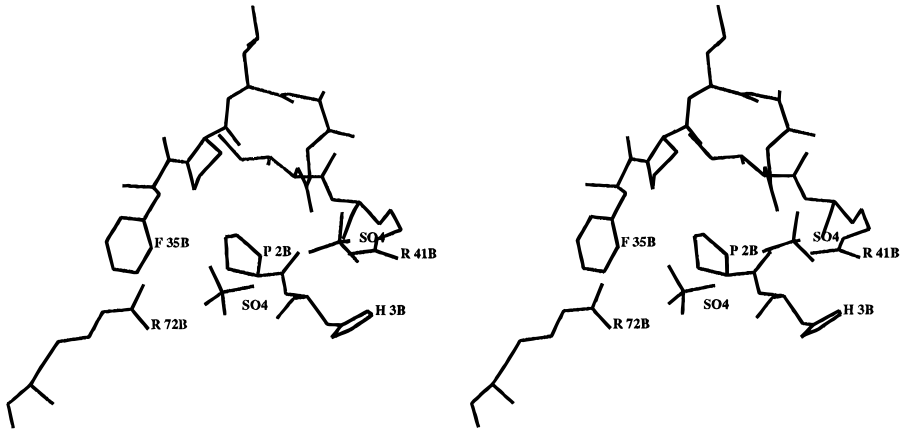


FIGURE 4: Location of the adjacent sulfate binding sites in a cleft on the surface of the CHMI molecule.

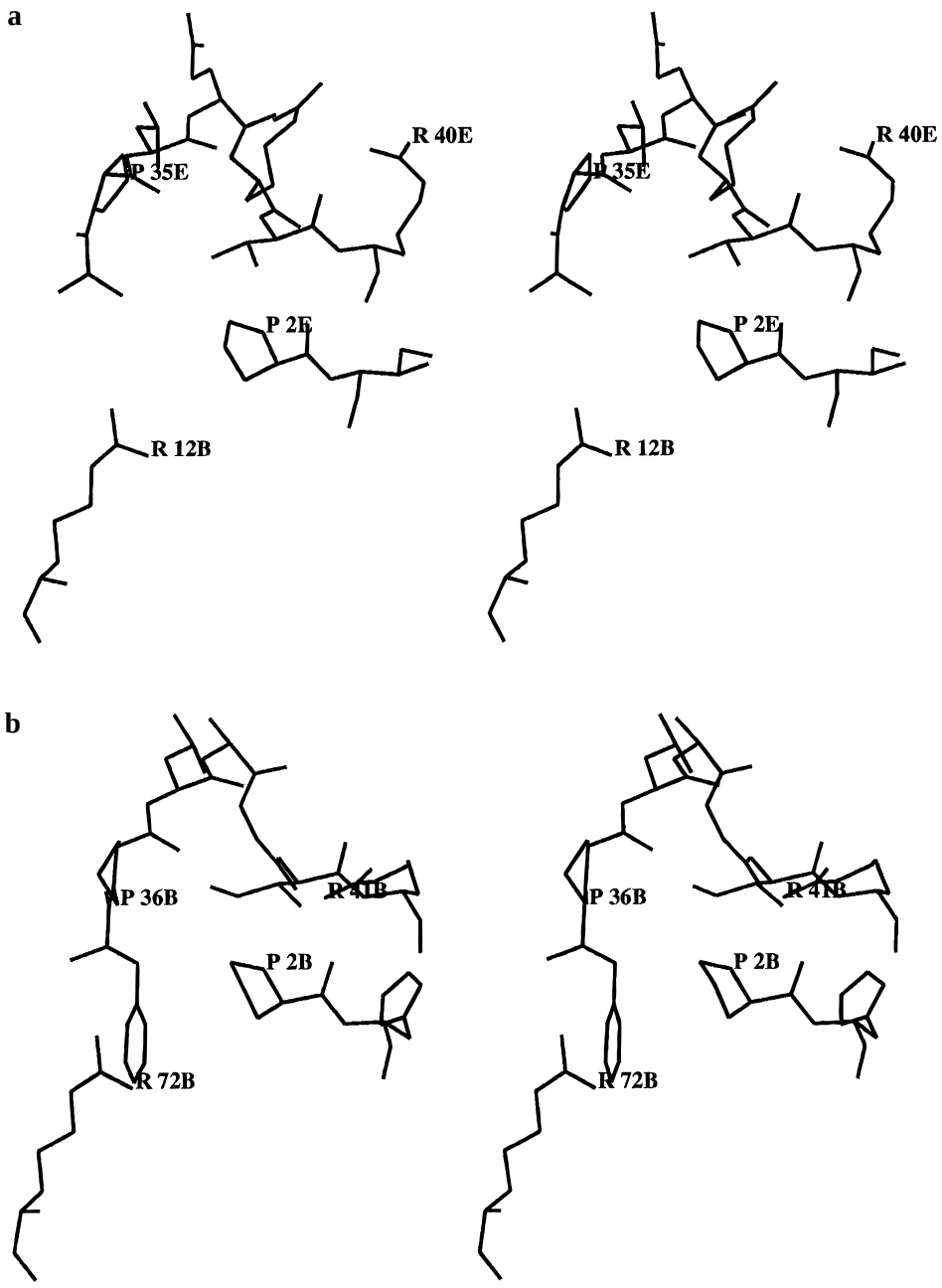


FIGURE 5: Comparison of the proposed active site regions in 4-OT (a) and CHMI (b). Only residues that are conserved in both active sites are labeled.

of this pocket in the two proteins is quite remarkable (Figure 5). While the two proteins do share a similar fold, the structures diverge significantly outside of the active site region.

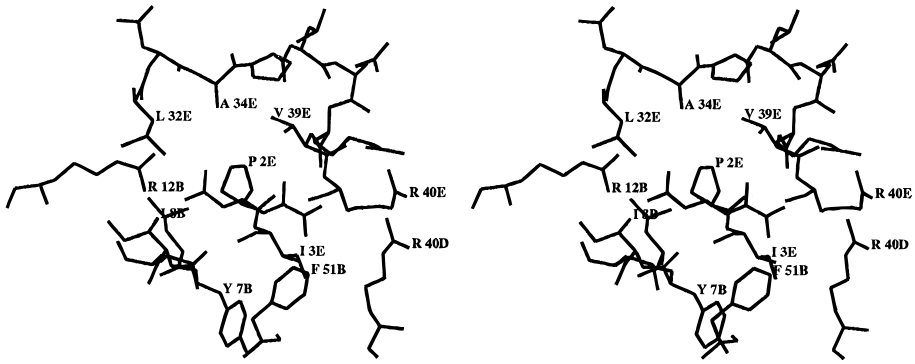


FIGURE 6: Model for the binding of substrate at the active site of 4-OT. The protein structure shown is that of the enzyme without substrate bound, but illustrates the relative position of the arginine side chains with respect to the substrate and the N-terminal proline residue.

DISCUSSION

By assuming that this pocket is indeed the active site, we can postulate how the enzyme might bind substrate and catalyze its isomerization. Previous studies (Whitman *et al.*, 1992) have revealed the stereochemistry of the 4-OT- and CHMI-catalyzed reactions. In addition, it has been shown that the isomerization proceeds by a suprafacial process, and thus it was concluded that the mechanism involves a single catalytic base (Lian & Whitman, 1993). In other words, the proton that is transferred between C-3 and C-5 leaves the substrate in the first step to protonate the base, and this same base then transfers the proton back to the substrate but at a different position. A similar process has been proposed for a number of enzymes, including triose phosphate isomerase (Rose, 1971). In the case of triose phosphate isomerase, the catalytic base is a glutamate residue that has a highly perturbed pK_a (Hartman, 1971). Examination of the active site region of 4-OT reveals that there are very few candidates for the catalytic base. In fact, without invoking any large conformational changes as a consequence of the binding of substrate, there is really only one potential candidate, a proline residue at the N-terminus of each subunit [the N-terminal methionine residue has been shown to be removed from both 4-OT and CHMI (D. I. Roper, unpublished data)]. However, the pK_a of the secondary amine group of a proline molecule in solution should be around 10.6, which is too high to act as an efficient catalyst at neutral pH. There are many examples of amino acid residues whose pK_a's are perturbed as a result of their protein environment; most notabe for the current discussion is triose phosphate isomerase in which a glutamate side chain is utilized as the catalytic base (Hartman, 1971). Pro-2 is completely surrounded by hydrophobic residues (Figure 6), an environment that will disfavor the charge introduced by protonation of the nitrogen and, thus, reduce the pK_a of this N-terminal proline residue. Because the environment of Pro-2 is so hydrophobic, it is not unreasonable to suggest that the pK_a of this residue will be sufficiently reduced to allow it to act as the catalytic base. This is facilitated further by the strong overall positive electrostatic potential over the active site pocket, which is revealed by calculations with the program GRASP (Nicholls & Honig, 1991) and is due to the proximity of the two arginine residues at either side of the active site pocket. In fact, Stivers *et al*. (1996) have shown that this residue has a pK_a of around 6.4. Consequently, we conclude that the N-terminal proline is the catalytic base in the active site. This is a highly unusual role for a proline residue and can only occur if the residue is at the N-terminus of the polypeptide chain. We believe this to be the first example of a proline residue acting as a catalytic base.

It has been observed that enzymes usually utilize arginine residues to bind to carboxylates on a substrate (Wigley *et al.*, 1987). Given this preference, we might expect at least two arginine residues to be present in the 4-OT active site to bind the carboxylates of the substrate. Examination of the active site reveals that just two arginine residues, at positions 12 and 40, are conserved in both of the known 4-OT sequences and in CHMI (Chen *et al*., 1992; Shingler *et al*., 1992; Roper & Cooper, 1990). The equivalent arginine residues are involved in binding the two sulfate ions that are found in each active site pocket in CHMI. These sulfate ions are 6.7 Å apart, compared to 6.0 Å for the distance between the carboxylates of the substrate in its extended form. If these arginine residues do indeed bind to the two carboxylate groups of the substrate, then they must do so in such a way that they position the substrate close to a suitable base that can catalyze the proton transfer reaction. One possible arrangement is shown in Figure 6. This would place the substrate in a position that would allow the N-terminal proline residue to act as the catalytic base.

On the basis of the structural evidence presented here, we can propose a mechanism for the reaction that is catalyzed by 4-OT and most likely also by the closely related enzyme CHMI (Figure 7). The initial substrate (2-oxo-4-hexene-1,6-dioate) binds to the active site region such that the carboxylate groups interact with two arginine side chains. This places the substrate in a position that allows the N-terminal proline residue to remove a proton from C-3. At the same time, an electronic rearrangement occurs in the substrate to shift the negative charge to the oxygen attached to C-2. It has been demonstrated that there is an acid group in the active site of 4-OT that is required for catalytic activity (Stivers *et al*., 1996). Examination of the 4-OT active site (and comparison with the CHMI active site) reveals that there are rather few possibilities for this acid. In fact, the only likely candidate is the side chain of Arg-40. However, the measured pK_a of 9.0 (Stivers *et al*., 1996) would be highly perturbed for an arginine side chain in solution, which one would typically expect to be closer to 12.5. The factors that give rise to the perturbation of the pK_a of Pro-2 may also apply to Arg-40, since they are in close proximity. Hence, Arg-40 in 4-OT (or Arg-41 in CHMI) may have two roles in the enzyme: firstly to interact with the carboxylate of the substrate and secondly to stabilize the buildup of charge

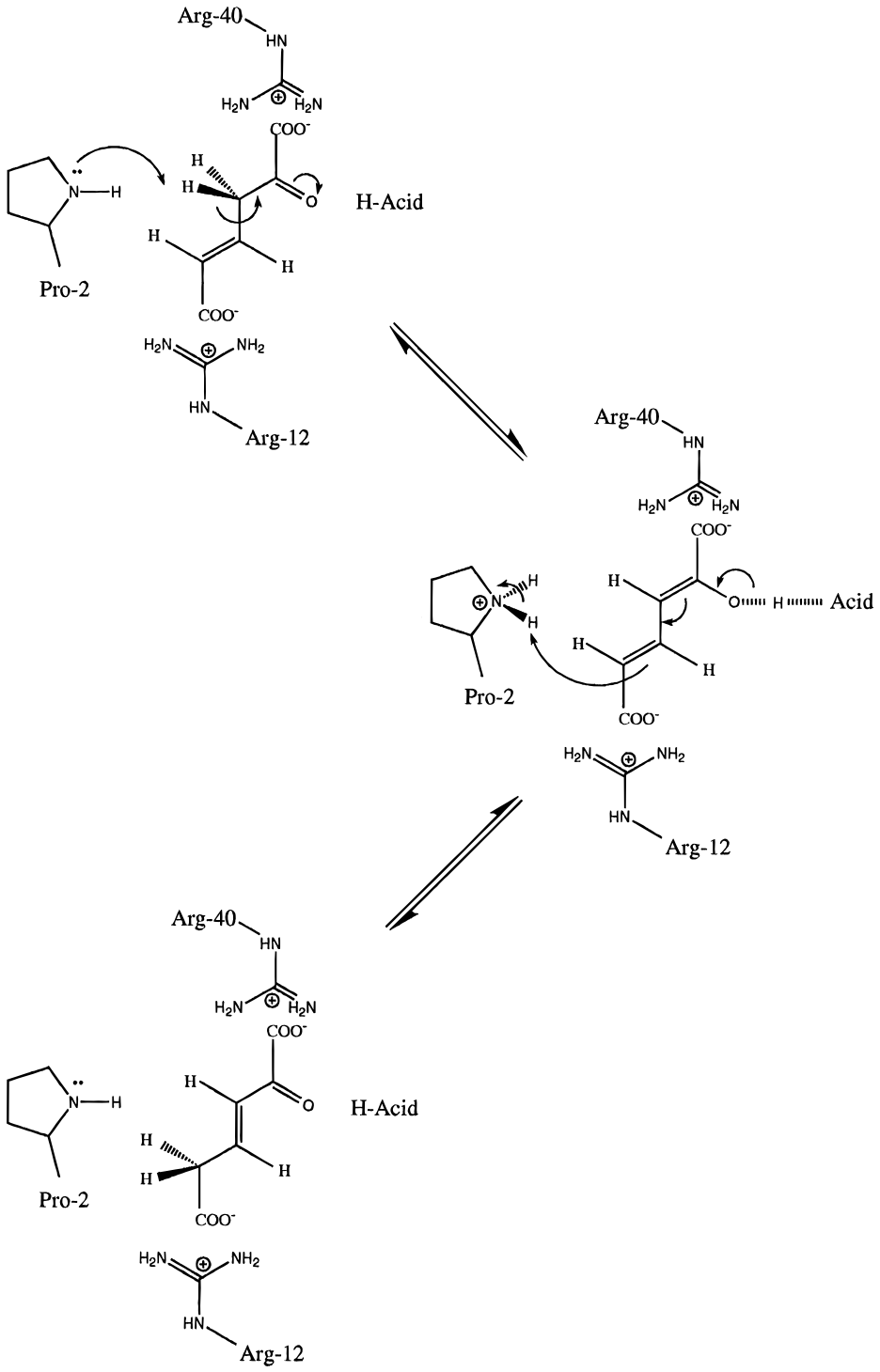


FIGURE 7: Plausible mechanism for the enzyme-catalyzed reaction in both 4-OT and CHMI.

on the intermediate (2-hydroxymuconate). This stabilization may be in the form of a shared hydrogen bond between the substrate oxyanion and the arginine. This sort of mechanism has been suggested for a number of other enzyme systems [see Gerlt (1994) for a review] and includes examples of histidine in triose phosphate isomerase (Lodi & Knowles, 1991) and lysine in RNase A (Gerlt & Gassman, 1993). If Arg-40 is not the active site acid group, then it would be difficult to suggest an alternative from the crystal structure without proposing a large conformational change during catalysis. Hopefully, future biochemical experiments will resolve this issue.

The enzyme has been shown to catalyze the isomerization in two stages. 2-Hydroxymuconate has a higher free energy than the initial substrate, as evidenced by the equilibrium between the compounds in solution. Thus, the initial role of the enzyme may be to stabilize 2-hydroxymuconate relative to the initial substrate, 2-oxo-4-hexene-1,6-dioate. In the second step of the reaction, the N-terminal proline residue could transfer a proton back to 2-hydroxymuconate, but this time to the carbon at position C-5. Another electronic rearrangement would accompany this protonation step to form the final product, 2-oxo-3-hexenedioate, which would then be released from the enzyme. Thus, the structural

evidence presented here confirms that the enzyme operates by a one-base mechanism, as suggested previously (Lian & Whitman, 1993), and now, surprisingly, identifies that base as the N-terminal proline residue.

One of the most interesting comparisons between 4-OT and CHMI is the question of substrate specificity. Whitman *et al.* (1992) reported the kinetic parameters for both enzymes when utilizing 2-hydroxymuconate (2-HM) or CHM as substrate. They report that while 4-OT shows a difference in the specificity constant (k_{cat}/K_M) of greater than 4 orders of magnitude in favor of 2-HM, CHMI is less discerning with only a 10-fold favoring of CHM as substrate. Nonetheless, there are significant differences between the enzymes, in spite of the close similarity of their substrates and their active sites, which convert a 46000-fold favoring of 2-HM as substrate to a 10-fold favoring of CHM, a swing of over 5 orders of magnitude. A comparison of the structures of 4-OT and CHMI provides clues about the manner in which this specificity has been achieved.

CHMI and 4-OT have evolved to catalyze very similar reactions, but upon different substrates. Consequently, one might expect the catalytic machinery of the enzymes to be similar, and because of the close structural similarity of the two substrates, one would probably also expect to see a good deal of similarity between the substrate binding sites. Indeed, there are strong similarities between the active sites of the enzymes, in particular the N-terminal proline residue sitting at the bottom of a hydrophobic pocket, but there are also differences that can explain their different substrate preferences. The difference between the two substrates is the replacement of a hydrogen atom at the C-5 position in 2-HM by a carboxymethyl group in CHM (Figure 1). Given the similarity between the enzymes and their substrates, as well as the similarity between their mechanisms, it is likely that the substrates bind in a similar fashion. The stereochemistry of the reaction (Whitman *et al.*, 1992) indicates that the carboxymethyl group must point out of the hydrophobic pocket of the active site. Since the difference in affinity of CHMI for 2-HM and CHM is only 2-fold, there can be few favorable interactions between this carboxymethyl group and the enzyme, or the difference in affinity for the two substrates would be greater. The difference in k_{cat} when utilizing the different substrates is 5-fold, suggesting that the favoring of CHM over 2-HM as substrate is due more to stabilization of the transition state than to affinity for the substrate, although there is no structural evidence to explain this suggestion.

The much greater substrate specificity shown by 4-OT is easier to explain. Examination of the molecular surface in the active site region of both enzymes reveals that the shapes of the active site pockets are rather different. In the case of 4-OT, the active site is elongated (Figure 8a), closely resembling the configuration of 2-HM that is implied from the stereochemistry of the reaction (Figure 1). Given that neither of the structures that we have determined has substrate bound, the molecular surfaces of the enzymes are remarkably complementary to their respective substrates, in terms of both shape and charge. However, the two substrates differ in size and shape, with CHM presumably adopting a more L-shaped configuration in the active site to be consistent with the stereochemistry. These differences are reflected in the shape of the CHMI active site (Figure 8b). The difference in the shape of the active site regions provides

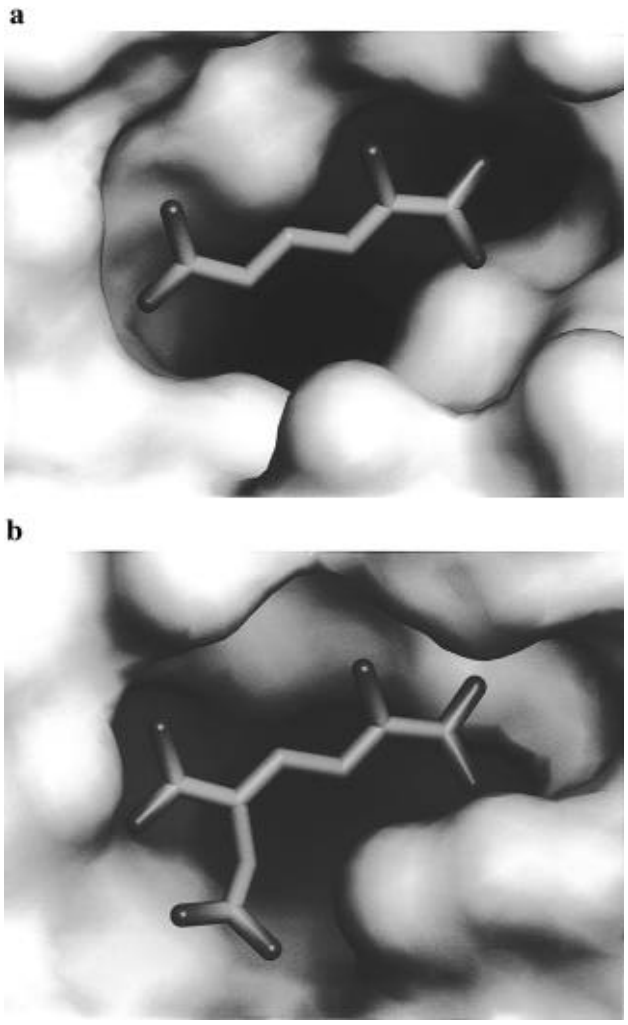


FIGURE 8: Molecular surface in the region of the active site, illustrating the shape of the active site cleft for (a) 4-OT and (b) CHMI.

a good explanation for the substrate preference of the two enzymes. 4-OT is able to discriminate between the substrates because its active site is too small to easily accommodate CHM. In the case of CHMI, because 2-HM represents a part of CHM, it is impossible for the active site to bind CHM and exclude 2-HM on steric grounds, which probably explains why CHMI is so poor at discriminating between the two substrates. This suggests that substrate specificity in these isomerase enzymes is determined by steric constraints, rather than other factors such as overall charge balance or charge complementarity. However, charge complementarity has been shown to be an important factor in converting the substrate specificity of lactate dehydrogenase to that of the closely related malate dehydrogenase (Wilks *et al.*, 1988). The difference in structure between lactate and malate is similar to the difference between 2-HM and CHM, but the mechanism that determines enzyme specificity appears to be different in the two systems. The additional charge resulting from the extra carboxylate group does not seem to be compensated for by CHMI. The reason for this remains unclear, but it will be interesting to test the effect on the affinity of the enzyme for its substrate by introducing an arginine at this position in the structure by site-specific mutagenesis.

ACKNOWLEDGMENT

We thank V. Shingler for providing the 4-OT overexpressing clone and C. Whitman for providing data prior to publication and for helpful discussions about the mechanism of the enzyme. We also thank EMBO for a short-term fellowship (D.B.W.) to allow data collection in Hamburg and M. Hartshorn for the preparation of figures using his program MOLVIEWER.

REFERENCES

Baker, E. N., Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. M., Hubbard, R. E., Isaacs, N. W., Reynolds, C. D., Sakabe, K., Sakabe, N., & Vajayan, N. M. (1988) *Philos. Trans. R. Soc. London 319*, 369−456.

Brunger, A. T., Karplus, M., & Petsko, G. A. (1989) *Acta Crystallogr. A45*, 50−61.

Chen, L. H., Kenyon, G. L., Curtin, F., Harayama, S., Bembenek, M. E., Hajipour, G., & Whitman, C. P. (1992) *J. Biol. Chem. 267*, 17716−17721.

Collaborative Computation Project No. 4 (1994) *Acta Crystallogr. D50*, 760−763.

Davenport, R. C., & Whitman, C. P. (1993) *J. Mol. Biol. 231*, 509−512.

Gerlt, J. A. (1994) *Curr. Opin. Struct. Biol. 4*, 593−600.

Gerlt, J. A., & Gassman, P. G. (1993) *Biochemistry 32*, 11943−11952.

Hajipour, G., Johnson, W. H., Dauben, P. D., Stolowich, N. J., & Whitman, C. P. (1993) *J. Am. Chem. Soc. 115*, 3533−3542.

Harayama, S., & Rekik, M. (1990) *Mol. Gen. Genet. 221*, 113−120.

Hartman, F. C. (1971) *Biochemistry 10*, 146−154.

Hendrickson, W. A., & Konnert, J. H. (1985) In *Biomolecular Structure, Function and Evolution* (Srinivisan, R., Ed.) Vol. 1, pp 43−57, Pergamon, Press, Oxford.

Jones, T. A. (1992) in *Molecular Replacement* (Dodson, E. J., Ed.) pp 87−90, SERC Daresbury Laboratory, U.K.

Leslie, A. G. W. (1992) CCP4 and ESF-EACMB Newsletter on Protein Crystallography, No. 26.

Lian, H., & Whitman, C. P. (1993) *J. Am. Chem. Soc. 115*, 7978−7984.

Lodi, P. J., and Knowles, J. R. (1991) *Biochemistry 30*, 6948−6956.

Nicholls, A., & Honig, B. (1991) *J. Comput. Chem. 12*, 435−445.

Otwinowski, Z. (1993) in *Data Collection and Processing*, SERC Daresbury DL/SC1/R34, pp 56−62, SERC, Daresbury, U.K.

Ramakrishnan, C., & Ramachandran, G. N. (1965) *Bio. Phys. J. 5*, 909.

Roper, D. I., & Cooper, R. A. (1990) *FEBS Lett. 266*, 63−66.

Roper, D. I., Fawcett, T., & Cooper R. A. (1993) *Mol. Gen. Genet. 23*, 241−250.

Roper, D. I., Subramanya, H. S., Shingler, V., & Wigley, D. B. (1994) *J. Mol. Biol. 243*, 799−801.

Rose, I. A. (1971) *Adv. Enzymol. 43*, 491−517.

Roussel, A., & Cambillau, C. (1989) TURBO-FRODO. in *Silicon Graphics Geometry Partner Directory*, pp 77−78, Silicon Graphics, Mountain View, CA.

Shingler, V., Powlowski, J., & Marklund, U. (1992) *J. Bacteriol. 174*, 711−724.

Stivers, J. T., Abeygunawardana, C., Mildvan, A. S., Hajipour, G., & Whitman, C. P. (1996) *Biochemistry 35*, 814−823.

Whitman, C. P., Hajipour, G., Watson, R. J., Johnson, W. H., Bembenek, M. E., & Stolowich, N. J. (1992) *J. Am. Chem. Soc. 114*, 10104−10110.

Wigley, D. B., Lyall, A., Hart, K. W., & Holbrook, J. J. (1987) *Biochem. Biophys. Res. Commun. 48*, 927−929.

Wigley, D. B., Roper, D. I., & Cooper, R. A. (1989) *J. Mol. Biol. 210*, 881−882.

Wilks, H. M., Hart, K. W., Feeney, R., Dunn, C. R., Muirhead, H., Chia, W. N., Barstow, D. A., Atkinson, T., Clarke, A. R., & Holbrook, J. J. (1988) *Science 242*, 1541−1544.

Wilson, K. S. (1978) *Acta Crystallogr. B34*, 1599−1608.

BI951732K