Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 10814–10818, November 1994
Biochemistry

The crystal structure of allosteric chorismate mutase at 2.2-Å resolution

(allosteric protein/effector binding/Greek key helix bundles)

YAFENG XUE*, WILLIAM N. LIPSCOMB*†, RONEY GRAF‡, GEORG SCHNAPPAUF‡, AND GERHARD BRAUS‡

*Gibbs Chemical Laboratory, Harvard University, 12 Oxford Street, Cambridge, MA 02138; and ‡Institut für Mikrobiologie, Eidgenössische Technische Hochschule, Schmelzbergstrasse 7, CH-8092 Zürich, Switzerland

Contributed by William N. Lipscomb, July 5, 1994

ABSTRACT **The crystal structure of an allosteric chorismate mutase, the Thr-226 → Ile mutant, from yeast *Saccharomyces cerevisiae* has been determined to 2.2-Å resolution by using the multiple isomorphous replacement method. Solvent-flattening and electron-density modification were applied for phase improvement. The current crystallographic *R* factor is 0.196. The final model includes 504 of the 512 residues and 97 water molecules. In addition, two tryptophan molecules were identified in the interface between monomers. The overall structure is completely different from the reported structure of chorismate mutase from *Bacillus subtilis*. This structure showed 71% helices with essentially no β-sheet structures.**

Chorismic acid is the last common intermediate in the biosynthetic pathway of the three aromatic amino acids in archaebacteria, eubacteria, plants, and fungi (1, 2). The tryptophan branch begins with anthranilate synthase (EC 4.1.3.27), which converts chorismate to anthranilate. In the other branch, chorismate mutase (chorismate pyruvatemutase, EC 5.4.99.5) catalyzes the intramolecular rearrangement of chorismate to prephenate, which is the last common intermediate before the pathway branches again toward either phenylalanine or tyrosine. In yeast *Saccharomyces cerevisiae*, the regulation of enzyme activities plays a major role in the control of the pathway for synthesis of aromatic amino acids. Among the chorismate mutases from different organisms, the yeast enzyme is an allosteric protein with no sequence homology with the *Bacillus subtilis* enzyme (2). The pool of chorismate is regulated through chorismate mutase and anthranilate synthase: tryptophan acts as an activator of chorismate mutase and as a feedback inhibitor of anthranilate synthase. In addition, chorismate mutase is subject to feedback inhibition by tyrosine.

It has been proposed that both enzymatic and nonenzymatic rearrangement of chorismate to prephenate is a concerted reaction that proceeds via a transition state with chair-like geometry (3–6). The enzymatic reaction is distinctive among known enzymes in that it catalyzes a pericyclic process. The recently published crystal structure of monofunctional chorismate mutase from *B. subtilis* shows that no functional group from the protein is available for direct involvement in the catalysis (7). The two-million-fold rate enhancement is thus achieved by selection of the active conformation and stabilization of the transition state by the enzyme.

The structural data on the catalytic antibody with low chorismate mutase activity (8) also supports the conclusion that the catalysis of the isomerization of chorismate proceeds by stabilizing the same pericyclic transition state that occurs in the uncatalyzed thermal reaction, rather than by providing alternative pathways to the product.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Yeast chorismate mutase is a monofunctional dimer of two 30-kDa polypeptides encoded by the *ARO7* gene (9). Biochemical data showed that in the dimer there are two substrate-binding sites, two activator-binding sites, and two inhibitor-binding sites that may or may not be distinct from the activator sites. Kinetic data of the wild-type chorismate mutase show positive cooperativity toward the substrate. This cooperativity is lost in the presence of tryptophan. The product of one of the *ARO7* mutant alleles, with a single substitution of Thr-226 → Ile-226 (T226I), has been characterized as a constitutively activated chorismate mutase that does not respond to activation by tryptophan or inhibition by tyrosine. As a relatively small protein in the allosteric family, yeast chorismate mutase provides an ideal model system for exploring the detailed mechanisms of allosteric regulation as well as to examine the unique catalysis. We describe here the crystal structure of the mutant chorismate mutase, T226I, from yeast *S. cerevisiae*.§ In contrast to the available structure of chorismate mutase from *B. subtilis* which exists as a homotrimer, the yeast protein is a dimer that has a completely different folding topology. In addition, the binding sites for tryptophan in the structure were identified.

MATERIALS AND METHODS

The mutant enzyme was isolated, purified, and characterized by R.G., G.S., and G.B., and the structure determination was made by Y.X. and W.N.L., as described below.

The crystallization of the protein was done by using the hanging-drop method, as described earlier (10). The space group was established as $P6_1$ or $P6_5$, and the unit cell parameters are $a = b = 95.8$ Å, $c = 157.9$ Å, $\alpha = \beta = 90°$, and $\gamma = 120°$. Heavy-atom derivatives were prepared by soaking the crystals in 75% (wt/vol) saturated ammonium sulfate at pH 8.0 (50 mM Tris·HCl/2 mM dithiothreitol/0.1 mM EDTA) containing heavy-atom reagents. The derivative data were collected on a Siemens area detector with CuKα radiation (Rigaku RU200BEH operating at 50 kV and 80 mA) and processed using the program XDS (11). The CCP4 program (12) was used to scale derivative and native data, to refine heavy-atom parameters, and to calculate electron-density maps. Among nine different heavy-atom compounds that were tested in soaking experiments, three of them, uranyl acetate, trimethyllead acetate, and methylmercury acetate gave usable derivatives (Table 1). Four uranyl sites were located by manual inspection, and their positions were confirmed by use of the program PATSOL (13). Two mercury sites (methylmercury acetate) and one lead site (trimethyllead

Abbreviations: MIR, multiple isomorphous replacement; T226I, Thr-226 → Ile mutant chorismate mutase.

†To whom reprint requests should be addressed.

§The atomic coordinates and structure factors have been deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973 (references 1CSM and 1CSM-SF).

Table 1. Heavy-atom derivative preparation

Heavy-atom compound	Concentration, mM	pH	Time
Uranyl acetate	10	6.4	12 hr
Trimethyllead acetate	25	6.0	10 days
Methylmercury acetate	0.5	7.6	16 hr

The soaking was done at 4°C.

acetate) were found by using PATSOL. The origins were correlated by use of the "cross phase" difference Fourier method. Heavy-atom parameters were then refined using the program MLPHARE (14). In Table 2 we record the statistics of the refinement. Improvement of the multiple isomorphous replacement (MIR) phases, which are poor, as indicated by the low phasing power (Table 2), was primarily achieved by phase refinement. The diffraction phases obtained from the MIR method at 3.0 Å were improved and extended to 2.5-Å resolution by solvent-flattening and density averaging using the program PHASES (15), which employs Wang's method (16) for automatic protein–solvent boundary determination. Although the solvent content in the crystal was calculated as 62% per asymmetric unit, a more conservative value of 55% was used in the solvent-flattening procedure to minimize possible truncation of protein density. To begin further refinement using the 2-fold noncrystallographic symmetry, the initial parameters for location of this molecular 2-fold axis were obtained from the two mercury sites and the four uranyl sites in the dimer using the program O (17). These parameters were refined by using the program PHASES. When the electron-density map was then averaged over this noncrystallographic 2-fold symmetry (iterative density averaging in real space), a significant improvement was seen in the electron-density map, which clearly showed the carbonyl groups within the α-helices. The handedness of these helices established the space group as $P6_1$.

Model building was carried out by using the program O. The skeletonized electron-density map was edited to show only the main-chain traces of the fragments of the protein. Then the polypeptide backbone was built, and all of the fragments were fitted to the amino acid sequence. However, at this stage, residues for which no density was visible in the map contoured at 1.25 σ were residues 212–224 and the last two residues in the C-terminal region. Amino acid side chains were built in by choosing one of the most frequently observed conformations of rotamers consistent with the electron density. Then the model was fitted to the density by orienting each residue as a rigid body. Crystallographic refinement was then done by using the program X-PLOR (18), starting from the initial R factor of 0.41. The first round of simulated annealing refinement (slow-cooling procedure starting from temperature of 4000 K) produced an R factor of 0.243. Manual

Table 2. MIR data

	Native	UO	HG	PB
Parameter				
Resolution, Å	2.0 (2.2)*	3.2	3.2	3.0
Reflections, observed	83,251	17,824	13,744	22,085
Reflections, unique	38,840	9,450	9,931	10,963
Completeness, %	70.2 (80.6)*	69.6	73.1	66.5
R_{sym},† %	8.6	9.5	5.6	9.1
MIR statistics‡				
Sites, no.		4	2	1
R_{iso},§ %		31.5	18.3	15.2
R_{cullis},¶		0.78	0.89	0.94
Phasing power‖		1.29	0.80	0.50

UO, uranyl acetate; HG, methylmercury acetate; PB, trimethyllead acetate.

*Data in the 2.0- to 2.2-Å shell are <50% complete but are still included in the refinement. Values in parentheses refer to data at 2.2-Å resolution.

†R_{sym} is the R factor on F between symmetry-related reflections.

‡MIR statistics were calculated by using program MLPHARE (14).

§$R_{iso} = \Sigma(|F_{der}| - |F_{nat}|)/\Sigma|F_{nat}|$.

¶R_{cullis} = (lack-of-closure error/isomorphous difference).

‖Phasing power is the mean value of the heavy-atom structure factor amplitude divided by the residual lack-of-closure error; figure of merit (centric reflections) = 0.55.

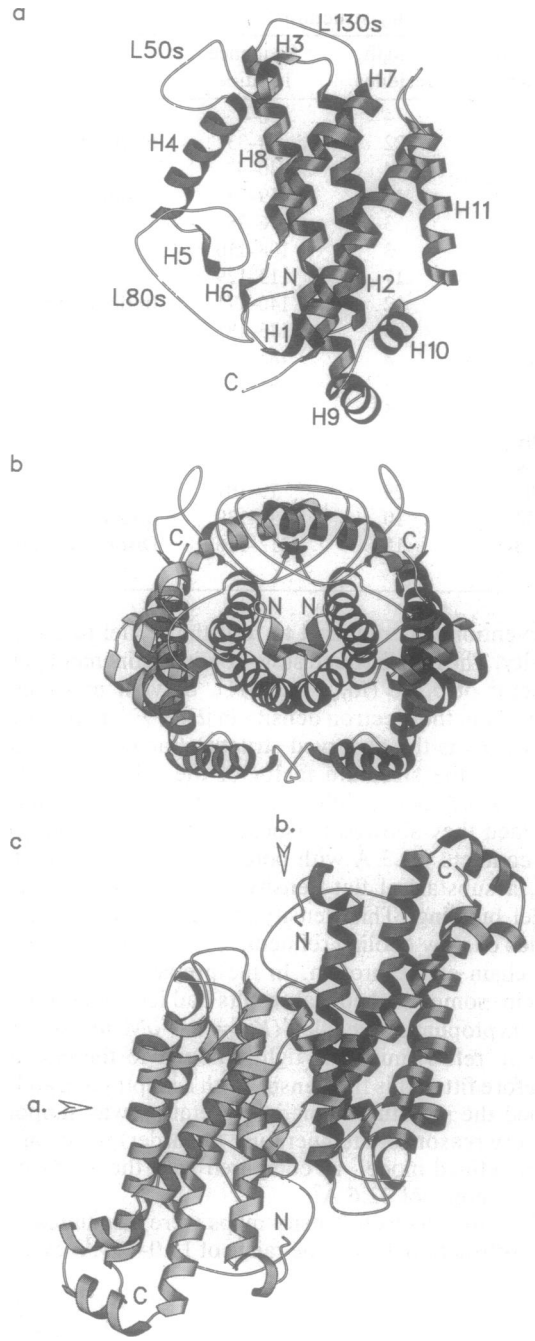

FIG. 1. Different views of the structure drawn in ribbon style. (*a*) One monomer is shown. The N termini and the C termini are labeled as N and C, respectively. There are 12 α-helices, of which 11 are labeled. The last helix, H12, is the one close to the C termini. (*b*) View of the dimer showing the hydrophilic channel through the molecule. The letters N and C mark the N termini and the C termini, respectively, for the dimer. One monomer is shown lightly shaded. (*c*) View along the approximate 2-fold noncrystallographic axis. The viewing directions in *a* and *b* are indicated by the arrowheads.

Table 3. Secondary structure

Structure element	Length, residues	Sequence location	Note
H1	4	6–9	
H2	22	14–33	In dimer interface
H3	4	40–43	
H4	15	59–73	In dimer interface
H5	3	76–78	3_{10} helix
H6	3	108–110	3_{10} helix
H7	17	113–129	
H8	32	140–171	In dimer interface
H9	9	173–181	
H10	8	185–192	
H11	17	195–211	
H12	25	227–251	
L30s	6	34–39	
L50s	15	44–58	Strand 48–50
L80s	29	79–107	
L130s	10	130–139	Strand 131–134
L220s	15	212–226	Disordered, no density for residues 218–221

intervention was required to refit the model to the electron density. The fourth round slow-cooling refinement yielded an R factor of 0.230 (R_{free} = 0.311). Solvent molecules were added from the electron density in $|F_o - F_c|$ maps (contoured at 3 σ) (F_o is the observed structure factor of the complex, and F_c is the structure factor of the refined model of the complex without its inhibitors) and $|2F_o - F_c|$ maps (1 σ), provided they showed reasonable hydrogen bond geometry and contacts <3.3 Å with potential protein atoms. In addition, a substantial flat density was observed in the initial model building. This density, which remained after many rounds of slow-cooling refinement, could not be fitted with a side chain of the protein. In the preparation of the mutant protein, some biochemical results had led to the speculation that tryptophan molecules (K_a = 1.5 μM to the wild-type protein, ref. 9) might be tightly bound to the enzyme. We therefore fitted this flat density with a tryptophan and further refined the structure. This interpretation was supported by the very reasonable temperature (disorder) factor of ≈20 Å^2 in the refined model, as compared with the average overall protein atoms of 32.6 Å^2.

All of the electron-density maps were calculated by using x-ray diffraction data in the range of 15.0–2.0 Å, even though only 40% of the data are available in the 2.0- to 2.2-Å shell. After rebuilding the model we made five rounds of slow-cooling refinement using data in the range of 5.5–2.0 Å. The 50s loops (44–58) and the 220s loops were rebuilt based on an omit map that had been annealed. The 220s loop showed disorder including no density for residues 218–221, which were left out in the final model. The two tryptophan molecules were built into the dimer in the fourth round. Another five rounds of refinement including positional refinement (200–300 cycles) and individual temperature refinement (20 cycles) were carried out as solvent molecules were introduced into the structure gradually. Those water molecules that showed temperature factor >60 Å^2 after one round of refinement were deleted in the next round. The refinement was completed at an R value of 0.196 (8.0–2.2 Å) in a model that had 504 residues and 97 water molecules in the asymmetric unit. Root-mean-square deviations from ideal values are 0.010 Å for bond lengths and 1.60° for bond angles. The overall quality of the electron-density maps was excellent. Analysis with the program PROCHECK (19) indicates that the overall stereochemical quality of the structure is good in comparison with other well-refined structures at 2.0-Å resolution. The Ramachandran plot showed 99.5% of the non-glycine residues in allowed regions (89.3% in "most-favoured" regions). The coordinate error of the model is 0.25–0.30 Å, as estimated from Luzzati analysis (20).

Table 4. Hydrophobic clusters in dimer interface

Structure element(s)	Residues	Sites, no.
H2 and H12	Phe-28A, Ile-31A, Val-211A, Tyr-212A, Phe-28B, Ile-31B, Val-211B, Tyr-212B	1
H4, L50s, L80s	Leu-50A, Ile-52A, Pro-53A, Phe-59A, Trp-62A, Ile-74B, Phe-94B, Leu-95B	2 (2-fold symmetry related)
H4	Leu-67A, Leu-67B	1

RESULTS AND DISCUSSION

Monomer Structure. The topology of the monomer can be described as an antiparallel α domain of Greek key helix bundles (21), where adjacent fragments in the sequence are

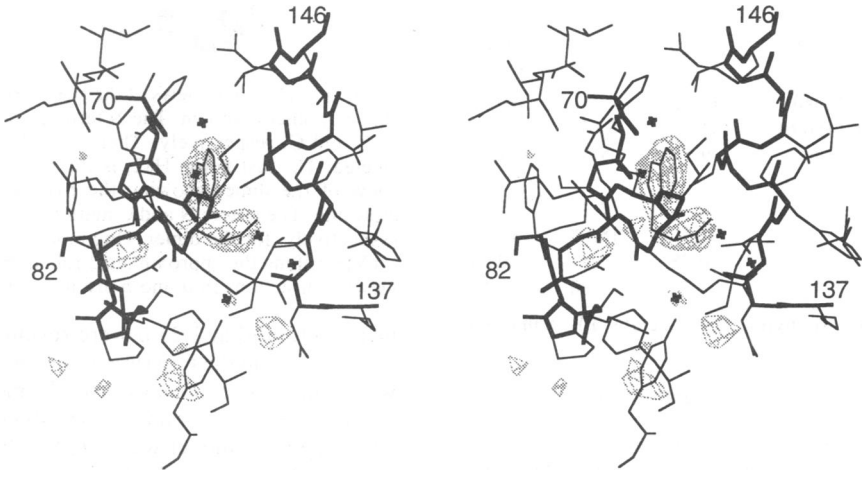

FIG. 2. Stereoview showing one of the two tryptophan-binding sites. Dashed lines show density in $|F_o - F_c|$ map (contoured at 3 σ) before adding the tryptophan into the model. The thick lines show the main-chain atoms from residues 70–82 (monomer A) and 137–146 (monomer B). Water molecules are indicated as crosses. The electron density in the other ligand tryptophan site is almost indistinguishable from that shown here.

not in immediate contact in the tertiary structure (Fig. 1*a*). There is essentially no β-sheet structure. The helical structure accounts for ≈71% of the whole sequence. There are 12 helices including two one-turn 3_{10} helices (Table 3). The major feature of the monomer is a twisted two-layer structure in which H2 and H8 are in one layer, and H7, H11, and H12 are in another layer (Fig. 1*a*). The packing angle between the helical axes from one layer and those from the other one is ≈60°. There are 32 residues in the H8 helix, the longest helix in the structure. The longest loop (L80s) involves residues 79–107. This loop is flanked by the two 3_{10} helices, H5 and H6 (Fig. 1*a*; Table 3). The helix H4 (residues 59–73) tilts away from the core bundle, forming a big patch together with H5, L80s, and H6. Helix H9 is located in the axial end of the molecule in an orientation nearly normal to the plane of the layers (Fig. 1 *a* and *c*).

Dimer Structure. The overall shape of the dimer is a bipyramid that has approximate dimensions of 90 Å × 55 Å × 55 Å. Four helices are involved in dimerization: H2, H4, H8, and H11. While H2 and H4 are packed against their counterparts mainly by hydrophobic forces, the interface between H8 dimers appears to be loose. Here the closest distance between C^{α} from opposite monomers in this region is >10 Å. A hydrophilic channel through the center of the molecule is formed from hydrophilic side chains from H8 (including Ser-142, Thr-145, Arg-146, Glu-149, Gln-152, Ser-153, and Arg-156), side chains from H2 (Glu-18, Arg-21, Lys-29, and Glu-32), H4 (His-71), H5 (Arg-76), and L80s (Glu-82) along with intervening solvent molecules (Fig. 1*b*). The residues Val-211 and Tyr-212 (from helix H11) are part of the hydrophobic patch formed together with H2 from both monomers (Table 4). In addition to the interaction between helices in the dimer interface, numerous interactions involving residues from loops are also observed. The loop L80s interacts with a 15-residue loop, L50s (residues 44–58), from the other monomer mainly through hydrophobic interactions (Table 4). A view down the hydrophilic channel shows that the dimer structure is better described as a 4-helix bundle of (2 × H2 + 2 × H8) wrapped by a twisted layer of the remaining α-helical structures. The strong hydrophobic association between two monomers observed here provides an explanation for the fact that ≈4 M guanidine hydrochloride is needed to dissociate the dimer, a process that occurs only upon unfolding of the monomers (9). The N-terminal region is close to the dimer interface and is somewhat buried in the vicinity of a hydrophobic cluster formed by residues Met-1, Phe-3, and residues from H1 (Val-9), H2 (Ile-15), L80s (Pro-80, Ile-103, Leu-104) and H8 (Phe-160 and Phe-163) in the same monomer. The central hydrophilic channel is partially blocked by the N termini (Fig. 1 *b* and *c*).

Table 5. Close contacts (hydrogen bonds) of ligand tryptophan

Potential hydrogen bonds		
Trp atom	Protein atom	Distance,* Å
N	Water-574 O^1	2.93
	Asn-138A O	3.26
	Asn-139A $O^{\delta 1}$	3.21
	Ser-142A O^{γ}	2.93
$N^{\varepsilon 1}$	Water-563 O^1	3.04
O	Water-521 O^1	2.79
	Water-527 O^1	2.87
	Water-536 O^1	2.92
O′	Gly-141A N	2.81
	Ser-142A N	2.98
	(Asn-139A C)	(3.27)
	(Asn-139A C^{α})	(3.27)
	(Gly-141A C^{α})	(3.25)

*Only distances <3.35 Å are listed.

Table 6. Close contacts (van der Waals) of ligand tryptophan

van der Waals contacts*		
Trp atom	Protein atom	Distance, Å
$C^{\delta 2}$	Arg-76B C^{γ}	3.55
	Arg-76B C^{α}	3.58
	Arg-76B C^{β}	3.56
$C^{\varepsilon 2}$	Gly-141A C	3.61
	Arg-76B C^{γ}	3.62
$C^{\varepsilon 3}$	Arg-76B C^{α}	3.23
	Arg-76B C^{β}	3.38
	Arg-76B N	3.49
	Ile-74B O	3.63
$C^{\zeta 2}$	Gly-141A C	3.62
	Gly-141A O	3.53
	Thr-145A $C^{\gamma 2}$	3.62
$C^{\zeta 3}$	Gly-141A C^{α}	3.43
	Arg-76B C^{β}	3.59
	Ile-74B O	3.45
$C^{\eta 2}$	Gly-141A C^{α}	3.58

*Only distances <3.65 Å are listed.

Tryptophan Binding Sites. The B values for ligand tryptophans are 18.6 $Å^2$ and 20.8 $Å^2$ for the two sites, respectively, in the refined model. The two binding sites are essentially identical, although there are slight variations in the distances between ligand tryptophan and its neighboring protein atoms when these two sites are compared. Here the binding site which has the lower average B value for the bound tryptophan is described. The ligand molecule is located in the dimer interface, where a cleft is formed between the pairs H4B and H5B, and L130sA and H8A (Figs. 1*a* and 2). Its amino nitrogen is hydrogen-bonded to side chains of Asn-139B and Ser-142B, and a solvent molecule Water-574. One carboxyl oxygen of tryptophan is hydrogen-bonded to amide nitrogens from Gly-141B and Ser-142B, respectively. The other carboxyl oxygen is within hydrogen bond distances to three water molecules. Another water molecule Water-563 is 3.04 Å from the ring nitrogen atom $N^{\varepsilon 1}$. Strong van der Waals interactions between the ring atoms and protein are observed: all of the six phenyl atoms have close nonpolar contacts with the protein (Tables 5 and 6). One ring carbon atom $C^{\varepsilon 3}$ is 3.23 Å from the C^{α} of Arg-76B.

The fact that liganded tryptophan interacts so strongly with the longest helix (H8) in the structure indicates that the binding of the tryptophan is capable of supporting long-range effects through the whole protein molecule.

The mutated residue 226 is located in the connection between L220s and H12. The lack of response of the Thr-226 → Ile mutant to the allosteric effector tyrosine suggests that tyrosine binding in the wild-type enzyme is affected by the loop L220s. However, a direct disturbance by disrupting hydrogen bonds in the vicinity of the tyrosine-binding site is also possible.

The fact that yeast chorismate mutase has a completely different tertiary fold from the chorismate mutase from *B. subtilis* raises the question as to whether the catalysis of isomerization of chorismate to prephenate requires only a local environment that is capable of selecting a favorably oriented rotamer of chorismate and stabilizing the transition state. Further studies of this enzyme are required to establish whether this mechanism or some more specific interaction of the substrate with this enzyme occurs during catalysis.

Note Added in Proof. A stereoview of the polypeptide chain of one monomer is shown in Fig. 3.

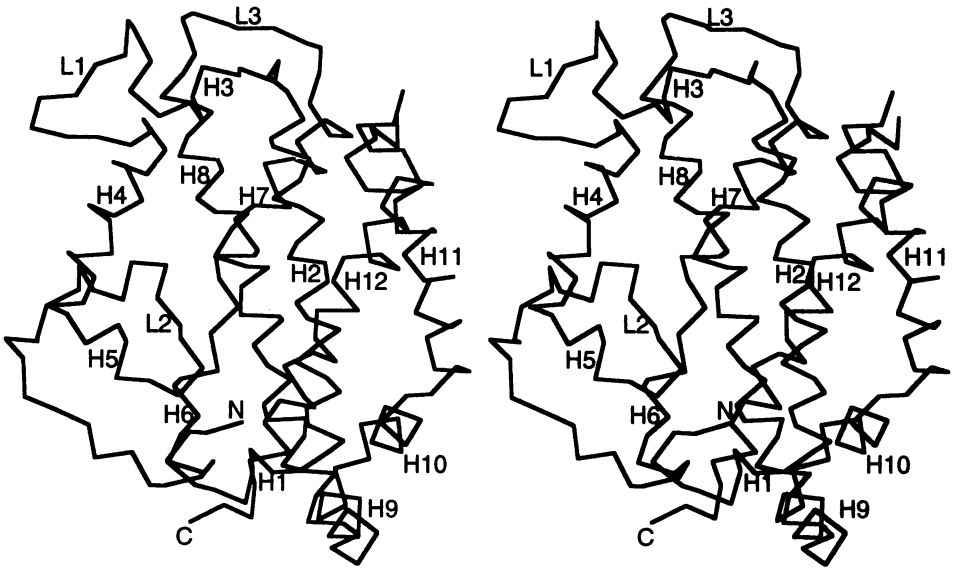

FIG. 3. The tertiary structural fold of a monomer of yeast chorismate mutase. The direction of view is the same as that shown in Fig. 1*a*. The loops L1, L2, and L3 in Fig. 3 correspond to L50s, L80s, and L130s, respectively, in Fig. 1*a*.

Special thanks go to Dr. Y. Zhang and Mr. S. Huang for help in data collection. This research is supported by National Institutes of Health Grant GM06920.

1. Weiss, U. & Edwards, J. M. (1980) *The Biosynthesis of Aromatic Amino Compounds* (Wiley, New York), pp. 134–184.
2. Braus, G. H. (1991) *Microbiol. Rev.* **55**, 349–370.
3. Sogo, S. G., Widlanski, T. S., Hoare, J. H., Grimshaw, C. E., Berchtold, G. A. & Knowles, J. R. (1984) *J. Am. Chem. Soc.* **106**, 2701–2703.
4. Copley, S. D. & Knowles, J. R. (1985) *J. Am. Chem. Soc.* **107**, 5306–5308.
5. Hilvert, D., Carpenter, S. H., Nared, K. D. & Auditor, M.-T. M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4953–4955.
6. Jackson, D. Y., Jacobs, J. W., Sugaswara, R., Reich, S. H., Bartlett, P. A. & Schultz, P. G. (1988) *J. Am. Chem. Soc.* **119**, 4841–4842.
7. Chook, Y. M., Ke, H. & Lipscomb, W. N. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 8600–8603.
8. Haynes, M. R., Stura, E. A., Hilvert, D. & Wilson, I. A. (1994) *Science* **263**, 646–652.
9. Schmidheini, T., Mosch, H.-U., Evans, J. N. S. & Braus, G. (1990) *Biochemistry* **29**, 3660–3668.
10. Xue, Y. & Lipscomb, W. N. (1994) *J. Mol. Biol.* **241**, 273–274.
11. Kabsch, W. (1988) *J. Appl. Crystallogr.* **21**, 916–924.
12. Science and Engineering Research Council (1979) CCP4: *A Suite of Programs for Protein Crystallography*, Collaborative Computing Project No. 4, Daresbury Laboratory, Warrington (SERC, Warrington, U.K.).
13. Tong, L. & Rossmann, M. G. (1993) *J. Appl. Crystallogr.* **26**, 15–21.
14. Otwinowski, Z. (1991) in *Isomorphous Replacement and Anomalous Scattering*, eds. Wolf, W., Evans, P. R. & Leslie, A. G. W. (SERC, Warrington, U.K.), pp. 80–86.
15. Furey, W. & Swaminathan, S. (1990) in *Abstracts of the Annual Meeting of the American Crystallographic Association*, ACA, Buffalo, NY PA33, Vol. 18, p. 73.
16. Wang, B. C. (1985) *Methods Enzymol.* **115**, 90–112.
17. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47**, 110–119.
18. Brünger, A. T., Kuriyan, J. & Karplus, M. (1987) *Science* **235**, 458–460.
19. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283–291.
20. Luzzati, V. (1952) *Acta Crystallogr. Sect. A* **5**, 801–810.
21. Richardson, J. S. (1981) *Adv. Protein Chem.* **34**, 166–339.