Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The 1.30 Å resolution structure of the *Bacillus subtilis* chorismate mutase catalytic homotrimer

Jane E. Ladner,[a] Prasad Reddy,[b] Andrew Davis,[b] Maria Tordova,[a] Andrew J. Howard[c] and Gary L. Gilliland[a]*

[a]Center for Advanced Research in Biotechnology, National Institute of Standards and Technology and the University of Maryland Biotechnology Institute, 9600 Gudelsky Drive, Rockville, MD 20850, USA, [b]Biotechnology Division of the National Institute of Standards and Technology, Bureau Drive, Gaithersburg, MD 20899, USA, and [c]Biological, Chemical, and Physical Sciences Department, Illinois Institute of Technology, Chicago, IL 60616, USA

Correspondence e-mail: gary.gilliland@nist.gov

Received 10 November 1999
Accepted 24 March 2000

PDB Reference: chorismate mutase, 1dbf.

The crystal structure of the *Bacillus subtilis* chorismate mutase, an enzyme of the aromatic amino acids biosynthetic pathway, was determined to 1.30 Å resolution. The structure of the homotrimer was determined by molecular replacement using orthorhombic crystals of space group $P2_12_12_1$ with unit-cell parameters a = 52.2, b = 83.8, c = 86.0 Å. The *ABC* trimer of the monoclinic crystal structure [Chook *et al.* (1994), *J. Mol. Biol.* **240**, 476–500] was used as the starting model. The final coordinates are composed of three complete polypeptide chains of 127 amino-acid residues. In addition, there are nine sulfate ions, five glycerol molecules and 424 water molecules clearly visible in the structure. This structure was refined with aniosotropic temperature factors, has excellent geometry and a crystallographic R factor of 0.169 with an $R_{\rm free}$ of 0.236. The three active sites of the macromolecule are at the subunit interfaces, with residues from two subunits contributing to each site. This orthorhombic crystal form was grown using ammonium sulfate as the precipitant; glycerol was used as a cryoprotectant during data collection. A glycerol molecule and sulfate ion in each of the active sites was found mimicking a transition-state analog. In this structure, the C-terminal tails of the subunits of the trimer are hydrogen bonded to residues of the active site of neighboring trimers in the crystal and thus cross-link the molecules in the crystal lattice.

1. Introduction

The shikimate pathway is responsible for the production of the aromatic amino acids and exists in bacteria, fungi and higher plants but not in animals. Chorismate mutase (E.C. 5.4.99.5), an important enzyme in this pathway, catalyzes the rearrangement of chorismate to prephenate, which can subsequently be converted to aromatic products such as tyrosine or phenylalanine. The enzyme provides a rate acceleration of greater than 10^6 over the uncatalyzed rearrangement at 298 K (Andrews *et al.*, 1977; Görisch, 1978). The mechanism of the enzyme-catalyzed rearrangement is not known. In aqueous solution, the chorismate molecule exists preferentially in the extended pseudo-diequatorial conformation and must be converted to the higher energy pseudo-diaxial conformation before conversion can occur (see Fig. 1; Copley & Knowles, 1985; Sogo *et al.*, 1984). An *endo*-oxabicyclic transition-state analog that mimics this proposed transition state has been synthesized (Bartlett & Johnson, 1985) and used in many studies.

Crystal structures of the enzyme from three different organisms (Lee *et al.*, 1995; Xue *et al.*, 1994; Chook *et al.*, 1993) and of an antibody elicited against the putative transition-state analog that has chorismate mutase activity (Haynes *et al.*, 1994) have been determined. The structures of the enzymes

are quite diverse. In *Escherichia coli*, the N-terminal 109 residues of the bifunctional P-protein make a fully functional chorismate mutase domain. The crystal structure of this domain has been reported to 2.2 Å resolution. It is a dimer and each of the monomers consists of three α-helices connected by two loops (Lee *et al.*, 1995). The structure was solved with the inhibitor in the active site; the inhibitor is completely buried with no solvent-accessible surface area. In *Saccharomyces cerevisiae*, chorismate mutase is an allosteric monofunctional dimer of 30 kDa monomers (Xue *et al.*, 1994). The shape of the molecule is described as bipyramidal and each monomer has 12 helices and essentially no β-sheets. The helices are arranged into two domains, one catalytic and one regulatory. In contrast, the chorismate mutase from *B. subtilis* is a homotrimer of polypeptide chains of 127 amino acids. After assembly of the trimer, the β-sheets are packed on the inside surrounded by the helices. The β-sheet core has both parallel and antiparallel strands. The structure has been solved with the active site containing only solvent (Chook *et al.*, 1993), prephenate (Chook *et al.*, 1994) and a transition-state analog (Chook *et al.*, 1994). All of these structures were solved with isomorphous crystals of space group $P2_1$ with four trimers in the asymmetric unit. The resolutions for these structures are 1.9, 2.2 and 2.2 Å, respectively. Most of the C-termini of the polypeptide chains end at residue 115; one extends as far as residue 120. In the active-site region, there is either no electron density for the side chain of Arg63 or it is modeled with very high *B* factors.

The fold of the monomer is described (Chook *et al.*, 1994) as a five-stranded mixed β-sheet flanked on one side by a 19-residue α-helix and a seven-residue 3_{10}-helix and on the other side by the C-terminal tail, which consists of a five-residue 3_{10}-helix. The monomer fold is classified as an $\alpha\beta_2$ layer sandwich with $\alpha\beta$ plaits (Orengo *et al.*, 1997). The fold of the monomer has been found to be similar to the C-terminal domain of protein FtsZ [PDB (Berman *et al.*, 2000) entry 1fsz; Löwe & Amos, 1998]. FtsZ has a four-stranded central β-sheet supported by two helices on one side. FtsZ has also been found to be similar to the α- and β-chains of tubulin (Nogales *et al.*, 1998; Löwe & Amos, 1998).

In the *B. subtilis* chorismate mutase structure, the three active sites of the macromolecule are located at the interfaces of adjacent monomers within the trimer (Chook *et al.*, 1994).

Figure 1
Ball-and-stick models of chorismate in the pseudo-diaxial conformation (left), of the proposed *endo*-oxabicyclic transition-state analog (center) and of the end product prephenate (right).

In the description of the active site, it is noted that there is very little change in the structure upon binding either the transition-state analog or prephenate. In the structure with the transition-state analog, all 12 active sites in the asymmetric unit contain the ligand. The residues that interact with the transition-state analog include Phe57, Ala59, Lys60, Val73, Thr74 and Cys75 from one subunit, and residues Arg7, Glu78, Arg90, Tyr108 and Leu115 from the adjacent subunit. Ionic contacts include those between the side-chain carboxylate O atoms (O3 and O4) of the inhibitor and the side chains of residues Arg7, Arg90 and Tyr108. Strong polar contacts are found between the inhibitor ether O atom (O7) and the side chain of Arg90, between the inhibitor hydroxyl group (O5) and the main-chain N atom of Cys75, the sulfhydryl group of Cys75, the carboxylate O atoms of Glu78 and the hydroxyl group of Thr74. Chook *et al.* (1994) noted that one side of the active site containing the transition-state analog is exposed to the solvent; the closest contacts listed for O1 and O2 are Lys60 N at 3.3 Å and Arg59 O at 3.7 Å. From these structures, it was concluded that the absence of active-site groups capable of protonating the ether O atom of chorismate and the absence of nucleophile groups capable of generating a covalent intermediate almost rules out the possibilities of the enzyme reaction proceeding by an electrophilic or nucleophilic mechanism. It is pointed out that the active site is electrostatically appropriate for stabilizing the polar transition state of the pericyclic reaction and thus it is this stabilization that contributes to the rate acceleration of the enzymatic reaction over the uncatalyzed reaction.

In the active site in the *E. coli* chorismate mutase structure, when compared with the active site found in the *B. subtilis* enzyme, the transition-state analog (see Fig. 1 for the atom labeling) has similar interactions for the hydroxyl group O5 (Arg and Glu), the carbonyl O atoms O3 and O4 (Arg and Lys) and the ether O atom O7 (Lys) (Lee *et al.*, 1995). It is also noted that the *E. coli* structure has polar interactions for the carboxylate O atoms O1 and O2 (Arg and Ser) and additional interactions for the ether O atom O7 (Gln). Lee *et al.* (1995) hypothesize that the additional hydrogen-bond interaction of the ether O atom with Gln stabilizes the negative charge on the ether O atom of the transition state. They also point out that the interaction of the Arg with the carboxylate O atoms O1 and O2 would stabilize the cyclohexadienyl cation fragment and in doing so enhance the rearrangement rate. The chorismate mutase from *S. cerevisiae* has been solved with the transition-state analog in the active site (PDB code 4csm; Sträter *et al.*, 1997) and the active site has been compared with the other active sites. The active site of the yeast structure has equivalent interactions to those found in *E. coli*. The hydroxyl O5 interacts with a Glu; the O3 and O4 carbonyl O atoms interact with an Arg; there is an Arg positioned to interact with the O atoms O4 and O5 and a Glu and a Lys positioned to interact with the ether O7. The authors point out that the interaction of the ether O atom with Gln in yeast and with Glu in *E. coli* would help stabilize the transition state. The importance of electrostatic interactions in the stabilization of the transition state is an important component of the reaction

mechanism that involves the pericyclic transition state (Gray & Knowles, 1994). The lack of defined interactions for the carbonyl O atoms O1 and O2 and the second interaction for the ether O atom are major differences between these structures and the structure for chorismate mutase from *B. subtilis* (Lee *et al.*, 1995; Sträter *et al.*, 1997).

Reported here is the structure of chorismate mutase from *B. subtilis* at 1.30 Å in the $P2_12_12_1$ space group with one trimer in the asymmetric unit. The structure, determined from crystals grown in high salt, is found to have sulfate and glycerol bound at the active site in a manner that mimics a transition-state analog.

2. Materials and methods

2.1. Materials

Crystallization conditions were surveyed using Crystal Screen Kit I from Hampton Research.[1] Chemicals used in the final crystallizations were purchased from Sigma and Fluka.

2.2. PCR amplification of *aroQ* gene

The *aroQ* gene was amplified from the plasmid pBSCM2, a gift from Drs D. Hilvert and P. Hast, ETH-Zentrum, Switzerland. Forward (CM1) and reverse (CM2) primers designed for amplification contained *Nde*I and *Bam*HI restriction endonuclease sites (bold), respectively, with the following sequences:CM1,5′-dGGAATTC**CATATG**ATGATTCGCGGA-3′; CM2, 5′-dCG**GGATCC**TTACAATTCAGTATT-3′. Final concentrations for polymerase chain reaction (PCR) amplification were as follows: 100 ng pBSCM2, 100 ng of each primer, 1× Pfu DNA polymerase buffer, 20 m*M* dNTPs, 2.5 units Pfu DNA polymerase I (Stratagene) and water to a final volume of 50 µl. Amplification conditions were: initial melting at 369 K for 60 s, 369 K for 45 s, 335 K for 45 s, 345 K for 60 s for 30 cycles, followed by a 343 K extension for 10 min.

2.3. Subcloning of *aroQ* gene into pRE expression vector

The PCR product was run on and purified from a 1.2% low-melting agarose. The insert was then digested with *Nde*I and *Bam*HI and purified from a 1.2% low-melting agarose. The product was then ligated into pREI expression vector (Reddy *et al.*, 1989) previously digested with *Nde*I and *Bam*HI. *E. coli* C6001λ was transformed with the ligation products by electroporation. A recombinant with *aroQ* gene was identified by restriction analysis and DNA sequencing.

2.4. Expression and purification of chorismate mutase

E. coli strain MZ1 (cI857ts) (Zuber *et al.*, 1987), carrying the pRE1-*aroQ* recombinant, was grown in 4 l of LB medium containing 50 µg ml^{-1} ampicillin at 305 K to an absorbance of 0.4 at 650 nm. An equal volume of preheated (333 K) LB medium containing 50 µg ml^{-1} ampicillin was then added, the temperature of the culture was shifted to 315 K to induce the λ P_L directed transcription and the culture was allowed to grow for 2.5 h in a water-bath shaker. Cells were suspended in 80 ml of ice-cold buffer *A* [50 m*M* Tris, 1 m*M* dithiothreitol (DTT), 1 m*M* EDTA, 0.1 m*M* PMSF at pH 7.5]. Cells were disrupted in a French pressure cell at 69 MPa and spun for 90 min at 100 000*g*. The supernatant was loaded onto an 80 ml DEAE anion-exchange column prequilibrated with buffer *A*. The column was then washed with buffer *A* and eluted with a linear NaCl gradient generated with 350 ml of buffer *A* and 350 ml of buffer *A* containing 0.4 *M* NaCl. Fractions were analyzed on a 15% SDS–PAGE and appropriate fractions were concentrated. Chorismate mutase was further purified by size-exclusion chromatography on a 500 ml Sephadex-75 (superfine) column equilibrated and run with buffer *A* containing 0.1 *M* NaCl at pH 7.5. Fractions were again analyzed as above. Fractions containing chorismate mutase were concentrated. The enzymatic activity of chorismate mutase was determined as described by Davidson & Hudson (1987).

Table 1
Data and refinement statistics.

Diffraction data	
Space group	$P2_12_12_1$
Unit-cell parameters (*a*, *b*, *c*) (Å)	52.2, 83.8, 86.0
Content of asymmetric unit	Trimer
V_m (Å^3 Da^{-1})	1.87
Measured intensities	255732
Unique reflections	89868
R_{merge}	0.092
R_{merge}, 1.34–1.30 Å	0.336
Completeness (%), 100–1.30 Å	92
Completeness (%), 1.34–1.30 Å	68
Redundancy, 100–1.30 Å	2.8
Redundancy, 1.34–1.30 Å	1.7
Average I/σ, 100–1.30 Å	11.9
Average I/σ, 1.34–1.30 Å	1.7
Refinement	
R factor, 90% of data	0.168
R_{free}, 10% of data	0.236
R factor with all data	0.169
Number of residues	381
Number of sulfate ions	9
Number of glycerol molecules	5
Number of water molecules	424
Double conformations	29
Geometry: r.m.s. deviations	
Bond length (Å)	0.012
Angle distance (Å)	0.031

2.5. Crystallization

The hanging-drop vapor-diffusion method was used at room temperature. After preliminary conditions were identified, they were optimized using customized screens close to the conditions that were identified with the kit. The final conditions employed 10 µl drops containing 5 µl of protein solution and 5 µl of reservoir solution. The protein solution contained 13 mg ml^{-1} protein in buffer *A* containing 100 m*M* sodium chloride. The 1 ml reservoir solution contained 2.2 *M* ammo-

[1] Certain commercial materials, instruments and equipment are identified in this manuscript in order to specify the experimental procedure as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that the materials, instruments or equipment identified are necessarily the best available for the purpose.

Table 2
R and R_{free} history.

	R	R_{free}
TNT beginning	0.376	0.405
TNT end	0.214	0.267
SHELXL isotropic *B* factors	0.224	0.259
SHELXL anisotropic *B* factors	0.185	0.241
SHELXL end	0.168	0.236
SHELXL with all data	0.169	

nium sulfate and 100 m*M* sodium acetate at pH 3.0. The crystals typically grew to 0.3 × 0.2 × 0.2 mm within 15 d.

2.6. Synchrotron data collection

Diffraction data were collected using a four-element charge-coupled device detector with a κ goniostat (Brüker AXS, Madison, WI), positioned 100 mm from the crystal at the IMCA-CAT beamline 17-ID of the Advanced Photon Source (Argonne, IL) using a 1.0 Å wavelength. During data collection, the crystals were maintained near 115 K in a nitrogen coldstream using an Oxford Cryosystems Cryostream (Oxford Cryosystems, UK). Diffraction data were processed with *X-GEN* (Molecular Simulations, 1997). Table 1 provides a summary of the diffraction data statistics.

2.7. Structure solution

The structure was solved by molecular replacement using the *AMoRe* program (Navaza, 1994) from the *CCP*4 software suite (Collaborative Computational Project, Number 4, 1994; Dodson *et al.*, 1997). The *ABC* trimer from the *B. subtilis* chorismate mutase structure 2chs (Chook *et al.*, 1994) was used as the search structure. The coordinates, after applying the final rigid-body solution from *AMoRe*, had an *R* value of 0.37 over the resolution range 10–4 Å. *TNT* (Tronrud *et al.*, 1987; Tronrud, 1997) was then used to refine the atomic positions and isotropic thermal factors. The final refinement used the *SHELXL* program (Sheldrick, 1997; Sheldrick & Schneider, 1997). This included the refinement of anisotropic thermal factors. During all of the refinement, the same 10% of the data was used for calculation of R_{free}. In the final refinement cycles, all data were included. The history of the crystallographic *R* and R_{free} is shown in Table 2. The molecular-display and map-fitting program *TURBO-FRODO* (Roussel *et al.*, 1996) was used to examine and adjust the structure during the refinement process. Positions were selected for water molecules visually and using *SHELXL*. The program *PROCHECK* (Laskowski *et al.*, 1993) was used to examine the stereochemical parameters of the final structure. Three more programs from the *CCP*4 suite were used: *SORTWATER* and *WATERTIDY* were used to examine the water structure and *ACT* was used to look at the inter- and intra-bonding structure. Coordinate files were aligned with one another using *TURBO-FRODO* and the program *ALIGN* (Cohen, 1997).

3. Results

3.1. Crystal form

The orthorhombic crystals of *B. subtilis* chorimate mutase reported here were grown in vapor-diffusion experiments equilibrated at high ionic strength (2.2 *M* ammonium sulfate at pH 3.0). The space group of these crystals is $P2_12_12_1$, with a single homotrimer in the asymmetric unit. The previously reported *B. subtilis* chorismate mutase crystals were grown by dialysis in low ionic strength buffer with 11–12% polyethylene glycol at pH 5.3 (Chook *et al.*, 1993). These crystals are of space group $P2_1$ and the asymmetric unit contains 12 monomers (four trimers).

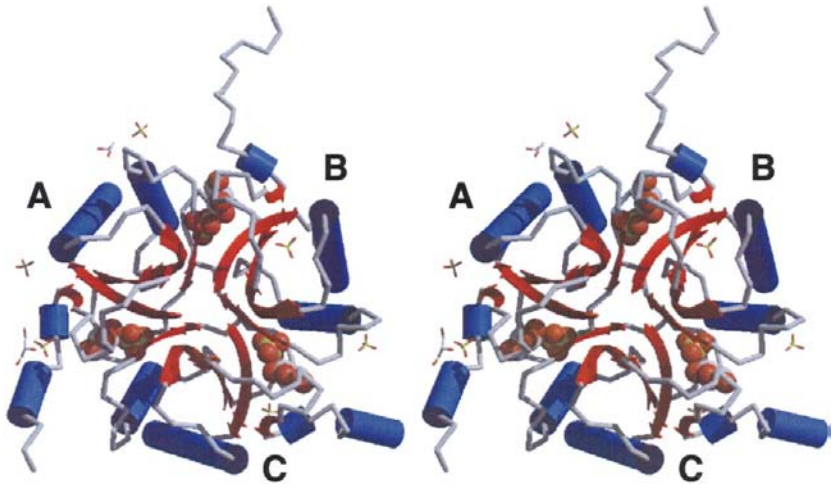

Figure 2
The chorismate mutase trimer is shown in stereo with its secondary structure. Space-filling models represent the sulfate and glycerol molecules in the active sites. Stick models represent the sulfate and glycerol molecules outside the active sites. This figure was generated using *TkRaster3D* (Bacon & Anderson, 1988; Merritt & Murphy, 1994). *TkRaster3D* is a graphical user interface developed by Hillary S. R. Gilson at the National Institute of Standards and Technology for the molecular-rendering package *Raster3D*.

3.2. Three-dimensional structure

A stereoview illustrating the threefold symmetry of the orthorhombic *B. subtilis* chorismate mutase trimer is shown in Fig. 2. The coordinates include 381 amino acids (127 per monomer), nine sulfate ions, five glycerol molecules and 424 water molecules. 29 of the amino acids are modeled in two conformations. The overall quality of the electron-density map is excellent, except for the region corresponding to residues 118–121 of the C-terminal tails of chains *B* and *C*. The model was refined with anisotropic temperature factors; however, the equivalent isotropic *B* factors are used for ease of analysis. As shown in Fig. 3, the values of the average isotropic main-chain *B* factors for residues 118–121 in chains *B* and *C* are the highest in the molecule, consistent with the final electron-density map. In a Ramachandran plot (Ramachandran & Sasisekharan, 1968), 92.7% of the non-glycine

Table 3
Secondary structure of monomers.

	Residues	Structural element
B1	2–12	β-strand
	12–15	β-turn type VIII
H1	16–36	α-helix
	38–40/38–41	3_{10}-helix/β-turn type I
B2	41–50	β-strand
	50–53	β-turn type IV
H2	58–65	α-helix
	65–68	β-turn type II
	68–71	β-turn type I
B3	72–78	β-strand
	81–84	β-turn type II
	85–88	β-turn type IV
B4	87–97	β-strand
	100–104/100–103	3_{10}-helix/β-turn type I
B5	107–109	β-strand
	109–112	β-turn type II′
H3	110–116	3_{10}-helix
	116–119*BC*	β-turn type I
H4	120–125*A*	α-helix
H5	121–126*C*	3_{10}-helix

and non-proline residues are in the most favored regions (Fig. 4), 6.4% are in the additional allowed regions and 0.9% (three residues) are in the generously allowed regions. These three residues are Glu110 from chains *A*, *B* and *C*; these were also reported as outliers in the previous structures (Chook *et al.*, 1994).

The three-dimensional structure of the enzyme in the orthorhombic crystals is very similar to the monoclinic structure reported earlier (Chook *et al.*, 1993, 1994). When one of the four trimers (*ABC*) of 2chs is superposed on the orthorhombic crystal structure, the r.m.s. deviation for the CA positions of 326 residues is 0.32 Å. The side-chain conformations of most interior residues are the same but, as expected, residues on the surface show more variation. The 2chs monomers includes residues 2–115 for each chain in the trimer, omitting the last 12 residues of the C-terminus. In contrast, all 127 residues in each monomer are included for the structure reported here.

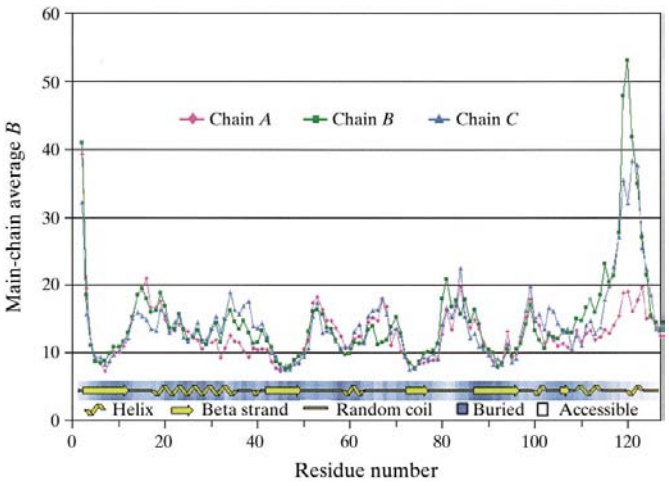

Figure 3
The average main-chain equivalent isotropic *B* values are plotted for each residue. The secondary-structure analysis from *PROCHECK* (Laskowski *et al.*, 1993) for chain *A* is superposed across the bottom of the chart. The secondary-structure analyses are the same for residues 1–116 of the *B* and *C* monomers.

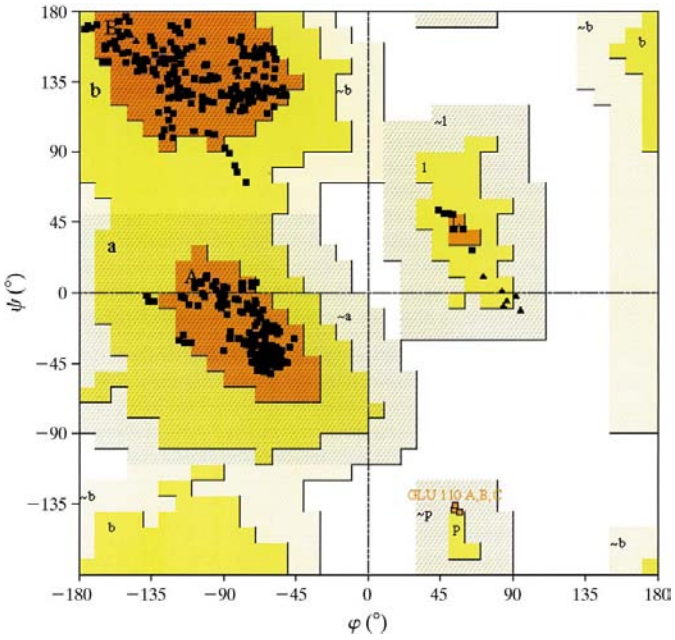

Figure 4
Ramachandran plot of the chorismate mutase trimer. Glycine residues are shown as triangles.

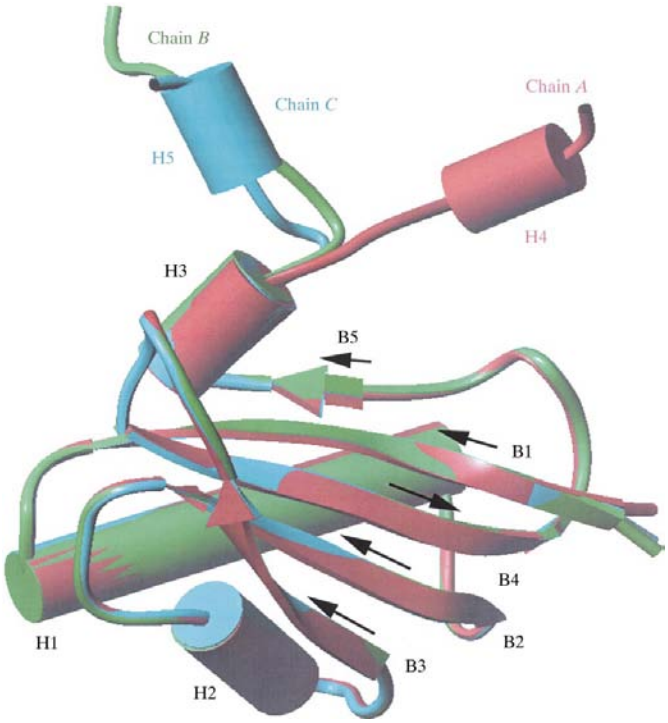

Figure 5
Superposition of chains *A*, *B* and *C* showing the divergence of the C-terminal tail regions. This figure was generated using *TURBO-FRODO* (Roussel *et al.*, 1996).

The extended Kabsch & Sander secondary-structural components and β-turns of the subunit are shown in Table 3 (Kabsch & Sander, 1983; Richardson, 1981; Wilmot & Thornton, 1990). The core of the structure is made up of five β-strands and three well defined helices, H1, H2 and H3. Residues 38–40 and 100–104 can be defined as short pieces of 3_{10} helix or as type I β-turns. In general, these structural elements agree with those reported (Chook *et al.*, 1994), with the exception of the helix H2 which is a 3_{10}-helix in the previous structures and is an α-helix in the orthorhombic crystal form. The orthorhombic form also adds the turn between the first strand and first helix, residues 12–15, as a type VIII β-turn. In the monoclinic forms, the distance between the CA atoms of residues 12 and 15 is greater than the 7.0 Å that has been set as the outer limit for the definition of a turn (Wilmot & Thornton, 1990). The distance in the orthorhombic form is 7.0 Å averaged over subunits *A*, *B* and *C*. Likewise, an additional type IV β-turn is observed in the orthorhombic form for residues 85–88.

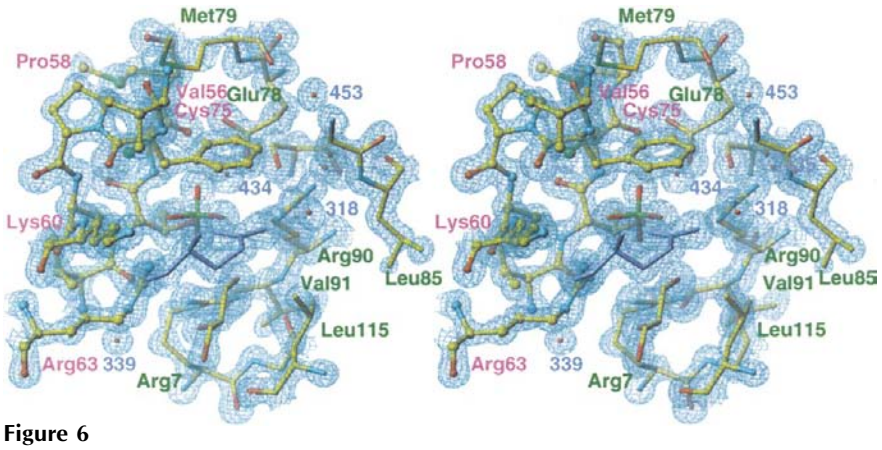

Figure 6
The active site of chorismate mutase shown with the electron density from a $2F_o - F_c$ map plotted at 2σ. The sulfate ion is in the center of the plot with glycerol below it. Glu126 from the crystallographic neighbor is shown in purple. The active site includes resides from adjacent chains; in this site, chain *A* residues are shown as ball-and-stick models and the chain *B* residues are drawn as just stick figures. This figure was generated in *TURBO-FRODO* (Roussel *et al.*, 1996).

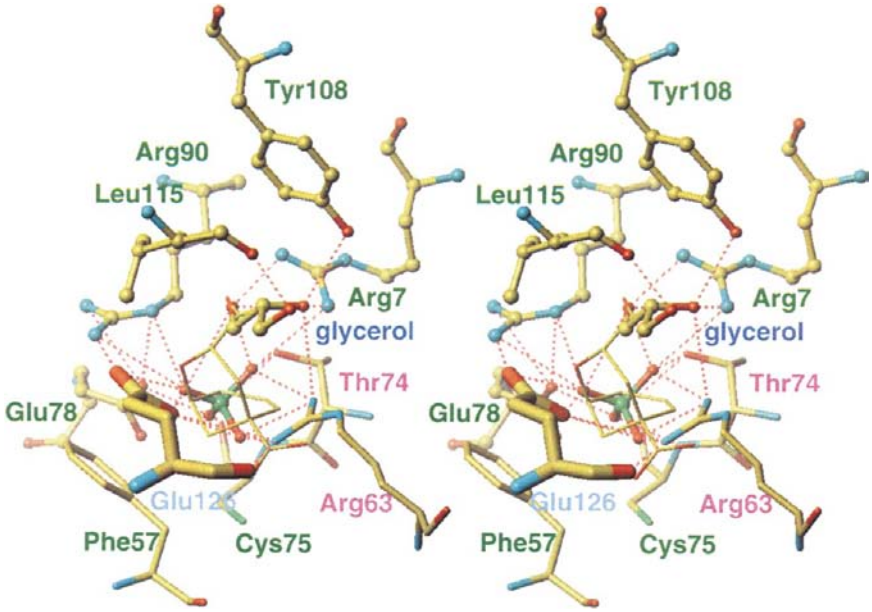

Figure 7
Superposition of the protein from 2cht (Chook *et al.*, 1994) onto our structure gives an r.m.s.d. of C^{α} positions of 0.31 Å for 324 pairs. The agreement is shown between the transition-state analog from 2cht (thin bonds) and the sulfate ion and glycerol molecule from the orthorhombic structure that results from this superposition. The principal active-site residues and potential hydrogen-bonding interactions are shown. This is the *AB* active site. The residues from chain *A* are shown as stick figures; the residues from chain *B* are shown as ball-and-stick models and the Glu126 from a neighboring molecule is show with very thick bonds. This figure was generated in *TURBO-FRODO* (Roussel *et al.*, 1996).

The non-crystallographic threefold symmetry is broken by the N-terminal Met, the C-terminal tails (residues 117–127) and by the conformations of some of the surface side chains. When the *A* chain is compared with the *B* and *C* chains, the r.m.s. deviation of the CA positions is 0.27 Å for 116 pairs and 0.22 Å for 112 pairs, respectively. The divergence of the C-terminal tails is, however, quite large after residue 118, as shown in Fig. 5. In the monoclinic structures, the tails are disordered and the model ends at residue 115 with the 3_{10} helix H3 for the unliganded form. In the crystals with transition-state analog and prephenate, the tails are reported as far as 119 in two monomers and as far as 120 in one monomer, respectively.

The *B* factors are well behaved except for four residues in the C-terminal tail of chain *B*, three residues in the C-terminal tail of chain *C* and the N-terminal Met in all three chains (Fig. 3). The average isotropic main-chain *B* factors for the three monomers are 12.9, 14.7 and 14.3 Å^2, respectively, with r.m.s. deviations of 1.4, 1.5 and 1.4 Å^2, respectively. The average isotropic side-chain *B* factors are 17.9, 20.6 and 19.2 Å^2, respectively, with r.m.s. deviations of 3.4, 3.8 and 3.5 Å^2, respectively.

3.3. Active site

The three active sites of the trimeric enzyme each contain one well defined sulfate ion and one glycerol molecule (see Fig. 6). These heterogroups form an extensive hydrogen-bond network with the protein (Fig. 7), mimicking the transition-state analog and prephenate that are bound to the structures 2cht and 1com (Chook *et al.*, 1994), which compared with one another show very little variation. The

Table 4
Potential hydrogen-bond forming interactions in the active site.

The distances are the average of the distances found in the trimer; the relative angle of the atoms involved was not used as a criteria for inclusion or rejection. The residues in bold type belong to the 'left' chain in the active site and the residues in normal type belong to the 'right' chain. The residues in italics belong to the C-terminal tail of a crystallographic symmetry mate. For clarity the interactions between the sulfate O atoms and Glu126 of the C-terminal tails are repeated.

Atom on active-site ligand	Residue	Atom on residue	Average distance (Å)
Sulfate O1	**Thr74**	OG1	3.2
	Cys75	N	2.9
	Glu78	OE2	2.6
Sulfate O2	Glycerol	O3	2.7
	Arg63	NH2	3.1
	Thr74	OG_1	2.7
	Arg7	NH1	3.3
Sulfate O3	**Arg63**	NH1	3.0
	Arg63	NH2	3.0
	Glu126	OE1	3.2
Sulfate O4	Glycerol	O3	3.2
	Glu78	OE2	3.4
	Arg90	NE	3.0
	Arg90	NH2	3.4
	Glu126	OE1	2.5
	Glu126	OE2	3.2
Glycerol O1	Leu115	O	2.7
Glycerol O2	**Arg63**	NH2	3.4
	Arg7	NH1	2.8
	Tyr108	OH	2.8
Glycerol O3	Arg7	NH1	3.2
	Arg7	NH2	2.9
	Arg90	NE	3.4
Glu125 O	**Phe57**	N	3.2
Glu126 $O^{\varepsilon 1}$	**Arg63**	NH1	3.4
	Sulfate	O3	3.2
	Sulfate	O4	2.5
Glu126 $O^{\varepsilon 2}$	Sulfate	O4	3.2
	Arg90	NH2	2.8
Glu126 O	**Arg63**	NH1	2.9
Glu127 O	**Lys60**	NZ	3.6
Glu127 O^{XT}	**Glu64**	OE2	2.8

Table 5
Potential hydrogen-bond forming interactions of waters that are conserved in the three active sites.

The distances are the average of the three distances found in the trimer. The residues in bold type belong to the 'left' chain in the active site and the residues in normal type belong to the 'right' chain. The residues in italics belong to the C-terminal tail of a crystallographic symmetry mate.

Waters	Residue	Atom on residue	Average distance (Å)	Average isotropic B (Å^2)
301, 440, 568	Ser88	OG	2.9	10.8
	Cys88	O	2.8	
	Arg90	NH1	2.9	
307, 318, 443	Leu85	N	2.8	15.2
	Glu126	OE2	2.8	
305, 435, 573	Leu85	O	2.7	13.3
	Arg90	NH1	3.0	
	Arg90	NH2	2.9	
572, 434, 438	Ser48	OG	2.6	11.3
	Thr74	OG1	2.8	
	Cys75	O	2.6	

superposition of the unligated structure on that containing the transition-state analog gives an r.m.s. deviation of 0.84 Å for the CA positions. Similar r.m.s. deviations are obtained for the other possible pairwise comparisons (Chook *et al.*, 1994). In contrast, the orthorhombic structure reported here has no transition-state analog or prephenate present. The crystals are grown at pH 3.0 in high ammonium sulfate. The active-site residues are well defined (see Fig. 6) and are in agreement with the previous structures, with the exception of residue Arg63. In all three monomers, Arg63 is turned inward toward the active site and is very well resolved. In the 12 monomers of the monoclinic crystal form, the Arg63 side chains are not visible in the electron-density map (Chook *et al.*, 1994). In the 1dbf structure with bound sulfate, the average values for isotropic main-chain B factors for residues 63 are 14.6, 13.4 and 15.5 Å^2 for the A, B and C chains, respectively, and the average side-chain isotropic B values are 12.5, 11.8 and 11.4 Å^2, respectively. The lower B-factor values for the side chain are indicative of the strong interactions of the side chain (see Table 4). The NH2 atom also interacts *via* a 2.8–2.9 Å hydrogen bond with the main-chain O atom of Val73. The comparison of 1dbf and the structures contained in the file 2chs shows that although the r.m.s. deviation of most of the structure is 0.32 Å, the r.m.s. deviation of the C^{α} positions for residue 63 and 64 is 2–3 Å. When 1dbf is superposed on each of the four trimers contained in 2chs, water molecules in the 2chs structure are found to be within 0.8 Å of the S atom of the sulfate ions in the orthorhombic structure. The sulfate ions are very strongly held and well resolved (Fig. 6); even when the high-salt crystallization solution is slowly exchanged for one containing no ammonium sulfate and 4K polyethylene glycol, the sulfate remains in the active site (Ladner & Gilliland, unpublished results).

When the 2cht model is superposed onto 1dbf, the transition-state analog lines up with the sulfate and glycerol molecules (Fig. 7). O5 (see Fig. 1) of the transition-state analog is less than 0.5 Å from the sulfate O1; O4 is less than 0.5 Å from the glycerol O3. The distances between potential hydrogen-bonding atoms and the sulfate and glycerol are shown in Table 4. The sulfate O1 makes its closest contacts with Cys75 N and Glu78 OE2. In the interactions of the transition-state analog these are the same atoms that make the closest contacts with O5. The glycerol O3 makes closest contact with Arg7 NH2, which is also in close contact with O4 in the transition-state analog complex. In addition, Arg63 is in the active site and interacts with the sulfate O3.

3.4. C-terminal tails

The monomers from residues 2 to 116 form a compact structure of β-strands and helices. The tails, residues 117–127, appear to be very flexible and in the orthorhombic model the tail of each monomer has a different conformation. These tails are important for the crystal packing of the orthorhombic crystal form. The tail of the A chain is very well defined. As seen in Fig. 3, all of the isotropic main-chain B factors of the tail are in the same range as those of the rest of the chain. Residues 123–127 of chains B and C are also well resolved and have B values comparable to the rest of the molecule. In each

Table 6
Residues modeled in double conformations and the occupancies of the conformers.

The location of the residue is specified as to the secondary-structural element to which the residue belongs; it is categorized as S for on the surface of the molecule, B for buried and T for part of the C-terminal tail. M means that the residue is involved in intermolecular contacts.

Residue	Occ. *A*	Occ. *B*	Location
Chain *A*			
Arg14	0.55	0.45	Turn S, M
Lys23	0.71	0.29	H1 S
Leu28	0.43	0.57	H1 B
Glu34	0.64	0.36	H1 S
Met45	0.67	0.33	B2 B
Gln77	0.61	0.39	B3 S
Val91	0.62	0.38	B4 B
Met54	0.54	0.46	B4 B
Val113	0.49	0.51	H3 S, M
Val114	0.37	0.63	H3 S
Thr122	0.47	0.53	H4 T
Chain *B*			
Glu18	0.66	0.34	H1 S
Ile31	0.50	0.50	H1 B
His36	0.66	0.34	H1 S
Glu40	0.52	0.48	Turn S
Met45	0.68	0.32	B2 B
Val73	0.42	0.58	B3 B
Asp98	0.63	0.37	Turn S
Val107	0.50	0.50	B5 B
Chain *C*			
Ile6	0.83	0.17	B1 S
Glu13	0.70	0.30	Turn S
Arg14	0.64	0.36	Turn S, M
Gln26	0.56	0.44	H1 S
Leu28	0.60	0.40	H1 B
Lys38	0.55	0.45	Turn S, M
Met45	0.65	0.35	B2 B
Met76	0.58	0.42	B3 B
Ile89	0.63	0.37	B4 B
Leu127	0.64	0.36	Tail T, M

chain, residue Glu126 interacts with the active site of a neighboring crystallographic symmetry-related molecule and forms hydrogen bonds with the sulfate and residues of the active site (Table 4).

3.5. Solvent structure

There are 424 water molecules in the structure. 322 water molecules are in the first hydration sphere, 89 are in the second hydration sphere and 13 are further from the protein. When the waters are divided based on their closest contact, 143 are associated with chain *A*, 136 with chain *B* and 145 with chain *C*. By superposing the monomer chains with their associated water molecules, 38 solvent positions are found to be conserved (within 0.5 Å in the overlaid view); if the agreement is relaxed to 1.0 Å, 14 more are found, giving a total of 52 conserved positions. There are four conserved water molecules found in the active sites. The potential hydrogen-bonding interactions of these are given in Table 5.

There are also buried conserved water molecules found associated with the β-sheet core. One site, water molecules 320, 446 and 569, was also found in the $P2_1$ crystal form (Chook *et al.*, 1994). The water molecule forms hydrogen bonds with Thr94 OG1 and Gln44 OE1. In the orthorhombic form, the average isotropic *B* of these water molecules is 15.8 Å^2. In two monomers, there is a second buried water, waters 555 and 657, close to this water that are less well defined and have *B* values of 54 and 32 Å^2, respectively.

3.6. Sulfate sites

There are six sulfate ions in the structure in addition to the three contained in the active sites (Fig. 2). These sites are all on the exterior of the molecule and all except one form hydrogen bonds between crystallographic neighbors. The one exception forms hydrogen bonds with the *A* chain C-terminal tail as it leaves the core of the molecule.

3.7. Glycerol sites

There are two well defined glycerol molecules in addition to the three that are found in the active sites (Fig. 2). The O atoms of one of these interact with NZ of Lys25 and O and OE1 of Glu64 of the *A* chain and neighboring water molecules. In chain *B*, the terminus of Lys25 is in a different conformation and the distance between its NZ and OE1 of Glu64 is increased by 1 Å. In chain *C*, there are water molecules in this region. The local environment is perturbed from

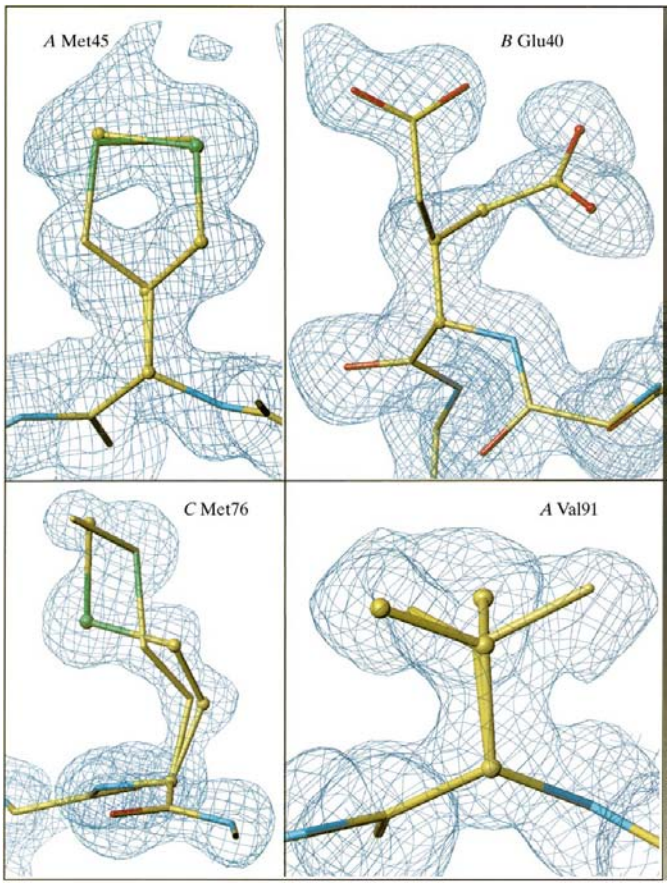

Figure 8
Four examples of the double conformations included in the model. The *A* conformation is depicted with ball-and-stick models and the *B* conformation with only stick models. The electron density is contoured at 0.7σ for chain *A* Met45, at 0.8σ for chain *B* Glu40, at 2.0σ for chain *C* Met76 and at 1.0σ for chain *A* Val91.

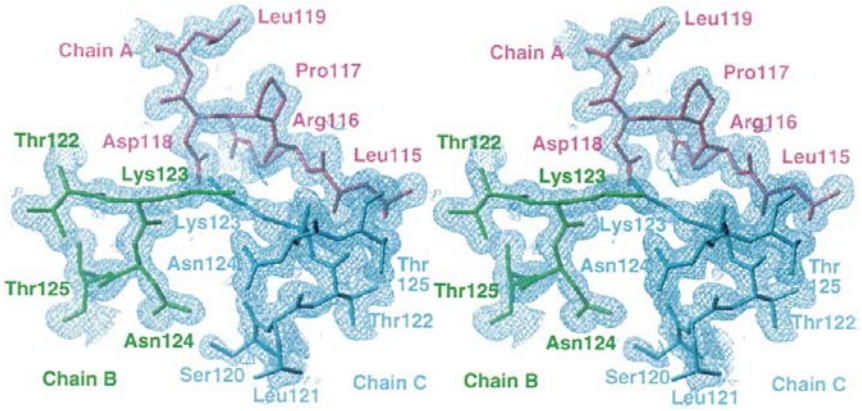

Figure 9
Portions of the C-terminal tails are shown with their electron density. Chain *A* is shown in pink and belongs to the primary molecule; chain *B* is shown in green and belongs to symmetry neighbor 14; chain *C* is shown in blue and belongs to symmetry neighbor 5. The electron density is contoured at 1σ and for clarity is shown only within 2 Å of the included atoms. This figure was generated in *TURBO-FRODO* (Roussel *et al.*, 1996).

that found in chain *A* owing to the presence of one of the crystal packing sulfate ions, which reorients the side chain of the neighboring Ser66. The other glycerol occurs between crystallographically related molecules O1 and O2 of the glycerol contact Gln103 NE2 of chain *A* and Gly83 O of chain *B* of different trimers. The third oxygen of the glycerol, O3, forms hydrogen bonds with water molecules. The conformation of the side chain of Gln103 is different for the three chains. Gln103 NE2 of chain *B* is involved in a crystal contact with Asp15 O of chain *C* and a water molecule. Gln103 NE2 of chain *C* forms a hydrogen bond with a water molecule that is also in contact with Asp102 OD1. In addition, Gly83 O atoms of chains *A* and *C* are hydrogen bonded to water molecules.

3.8. Double conformations

There are 29 residues modeled with two conformations or which have discrete disorder. The occupancies of the two conformations are given in Table 6. The residues are scattered throughout the molecule; 11 of these residues are in chain *A*, eight are in chain *B* and ten are in chain *C*. The only residue that is modeled in a double conformation in all three chains is Met45. Four examples are shown in Fig. 8. In the β-core, several disordered residues occur together. In chain *A*, there is a cluster of residues 28, 45, 91 and 93; in chain *B*, there are two close together, 45 and 73, and in chain *C* there is the pair 28 and 45.

3.9. Crystal packing

Each trimer in the orthorhombic crystal form interacts with ten different symmetry neighbors.[2] Since the active site is at the interface of two monomers, it can be labeled so that the first letter is the designation for the chain of the Arg63 and the second letter is the chain that contributes Arg90. The penultimate residue of each monomer, Glu126, is found in the active site of a different neighboring trimer. Residue 126 of the chain *A* interacts with neighbor (2 0 1 0) active site *AB*; residue 126 of the *B* chain interacts with symmetry neighbor (4 0 0 0) active site *BC* and residue 126 of the *C* chain interacts with symmetry neighbor (3 0 −1 0) active site *CA*. Another aspect of this interaction of the C-terminal tails is that the active sites of symmetry molecules are within 30 Å of each other and the tails are quite close to each other (Fig. 9). Three other symmetry neighbors account for the reciprocal interactions of the incoming tails of the neighboring molecules (2 0 1 −1; 4 −1 0 0; 3 0 0 0). The remaining four symmetry neighbors involve the end of the H1 helix and the subsequent turn of chain *A* interacting with the turn after helix H2 and the end of H2 of chain *B* (symmetries 3 1 0 0 and 3 1 −1 0) and H1 of chain *B* interacting with the turn between B1 and H1 of chain C (symmetries 1 −1 0 0 and 1 1 0 0). The specific interactions are given in Table 7. Arg116 is the last residue in the C-terminal 3_{10}-helix. In the orthorhombic crystal form, it participates in contacts with the crystallographic neighbors in all three chains (see Table 7). Sulfate ions 204 and 208 and water molecule 712 bridge with chain *A* Arg116 NH2 on one molecule and the main-chain N of Lys86 on the neighboring molecule. Chain *B* Arg116 makes contacts with Asp15 of chain *B* of a neighboring molecule. The distance between Arg NH2 and Asp O is 2.9 Å. There is also a bridging water molecule, water 566, that interacts with Arg116 NH1 and Asp OD2. Chain *C* Arg116 forms a bridge to chain *A* Lys111 and Glu13 of a neighboring molecule *via* the O atoms of sulfate 209.

4. Discussion

The orthorhombic crystal structure of *B. subtilis* chorismate mutase has been determined at 1.3 Å resolution. This high-resolution structure provides new details of the molecular structure of this enzyme. The higher resolution of this structure is a consequence of the crystal packing, which places only one trimer in the asymmetric unit and links the trimers together through interactions of the C-terminal tails.

The three well defined glycerol molecules found in the active sites provide additional examples where glycerol being used as a cryoprotectant becomes part of the active site. In a recent structure of *E. coli* uracil DNA glycosylase (Xiao *et al.*, 1999), the three hydroxyl groups of glycerol were found to mimic the interaction of uracil O2, O4 and N3 with the enzyme. Here, two of the hydroxyl groups of glycerol appear to mimic two hydroxyl groups of the transition-state analog.

In their study of chorismate mutase with ligands in the active site, Chook *et al.* (1994) point out that the most significant difference between the unliganded enzyme and the enzyme ligated with either prephenate or a transition-state

[2] The standard designation of symmetry neighbors gives the symmetry operator as the first number (for $P2_12_12_1$: 1 = x, y, z; 2 = $-x+\frac{1}{2}, -y, z+\frac{1}{2}$; 3 = $-x, y+\frac{1}{2}, z+\frac{1}{2}$; 4 = $x+\frac{1}{2}, -y+\frac{1}{2}, -z$) followed by three numbers indicating the cell translations in x, y and z.

Table 7
Strong packing interactions.

Structural region 1	Structural region 2
Symmetries 2 0 1 0 and 2 0 1 −1	
N-terminal region of B2 of chain *C*	Turn between B1 and H1 of chain *A* (one conformation of Arg14)
C-terminal tail of chain *A*	Active site *AB*
Sulfate 205 and water molecules	H2 side chains of chain *A*
Glycerol 255 and water molecules	Turn between B3 and B4 of chain *B*
Symmetries 4 0 0 0 and 4 −1 0 0	
C-terminal tail of chain *B*	Active site *BC*, sulfate 206 and C-terminal tail of chain *C*
Symmetries 3 0 −1 0 and 3 0 0 0	
C-terminal tail of *C*	Active site *CA*, sulfate 209
Sulfate 208 and water molecules	Side chains of H3 and turn between B3 and B4 of chain *A*
Symmetries 3 1 0 0 and 3 1 −1 0	
H1 and subsequent turn of chain *A*	H2 and subsequent turn of chain *B*
Symmetries 1 1 0 0 and 1 −1 0 0	
H1 of chain *B*, water molecules and sulfate 207	Turn before H1 of chain *C* (both conformers of Arg14)

analog is the appearance of more density for the C-terminal tails of the complexes. With prephenate, nine of the 12 active sites are occupied and the last residue modeled varies from Leu115 to Ser120. With the transition-state analog, all the active sites are occupied and the observed tails vary from Leu115 to Leu119 for the 12 monomers in the asymmetric unit. In the orthorhombic crystals in our study, the molecules are linked together in the crystal by the next to last residue of the C-terminal tail, Glu126, forming hydrogen bonds with the active-site residues of neighboring molecules and the very tightly bound sulfate ion which is found deep in the active site. This arrangement is indicative of the flexibility of the C-terminal tail and also of the very ionic environment of the active site. Residue Arg63 is also pulled into the active site of the orthorhombic crystals and occupies some of the space that is taken up by the ligands in the monoclinic form. However, this rather large movement of Arg63 causes only a local difference in the backbone of the molecules in the two crystal forms. Chook *et al.* (1994) hypothesize that the low $k_{\rm off}$ (Gray *et al.*, 1990) of prephenate is a consequence of the very tight fit of the ligand in the active site and that for it to leave the active site the C-terminal tail would need to move. The orthorhombic structure supports the flexibility of the C-terminal tail and the possibility of some movement of the residues at the entrance of the active site without disturbing the rest of the structure.

It would appear that it would be possible for the C-terminal tail to actually participate in the reaction mechanism. The three very different conformations found for the three tails show that the distance from CA of residue 116 to CA of residues 127 varies from 21.7 Å for chain *A* to 23.3 Å for chain *B* and 18.2 Å for chain *C*. The interaction of the crystallographic neighbor tail demonstrates that there are residues on the surface around the active site that can hydrogen bond with residues of the C-terminal tail and position Glu126 in the entrance of the site within contact distance of the ether O atom of the transition-state analog. If Glu126 does not fulfill this position, there are other residues in the tail that could, for instance Arg116 or Lys123. In the orthorhombic crystals, Arg116 is turned away from the body of the structure and is involved in crystal contacts in all three subunits, but a slight reorientation of the final 3_{10}-helix could rotate the residue to participate in the active site. The participation of these additional residues would increase the structural homology of the active site with that observed in the *E. coli* and yeast chorismate mutase structures.

The question then remains why were these interactions are not seen in the monoclinic structures of the *B. subtilis* enzyme. This difference could be related to the low ionic strength of the crystallization medium. In addition, the transition-state analog is, of course, not the exact transition state and the ionic character of the true transition state could be different. As for the structure with the prephenate, this is the end product and the opening up and disorder of the C-terminal tail may be indicative of the release mechanism of the enzyme. Further studies of this crystal form and others will be needed to further elucidate the reaction mechanism of this central enzyme in the aromatic amino-acid biosynthetic pathway.

Use of the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Energy Research, under Contract No. W-31-109-Eng-39. Data were collected at beamline 17-ID. The facilities of IMCA-CAT are supported by the companies of the Industrial Macromolecular Crystallography Association through a contract with Illinois Institute of Technology (IIT), executed through the IIT's Center for Synchrotron Radiation Research and Instrumentation.

References

Andrews, P. R., Cain, E. N., Rizzardo, E. & Smith, G. D. (1977). *Biochemistry*, **16**, 4848–4852.
Bacon, D. J. & Anderson, W. F. (1988). *J. Mol. Graph.* **6**, 219–220.
Bartlett, P. A. & Johnson, C. R. (1985). *J. Amer. Chem. Soc.* **107**, 7792–7793.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G. L., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res*. **28**, 235–242.
Chook, Y. M., Gray, J. V., Ke, H. & Lipscomb, W. N. (1994). *J. Mol. Biol.* **240**, 476–500.
Chook, Y. M., Ke, H. & Lipscomb, W. N (1993). *Proc. Natl Acad. Sci. USA*, **90**, 8600–8603.
Cohen, G. H. (1997). *J. Appl. Cryst.* **30**, 1160–1161.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Copley, S. D. & Knowles, J. R. (1985). *J. Amer. Chem. Soc.* **107**, 5306–5308.
Davidson, B. E. & Hudson, G. S. (1987). *Methods Enzymol.* **142**, 440–450.
Dodson, E. J., Winn, M. & Ralph, A. (1997). *Methods Enzymol.* **277**, 620–633.
Görish, H. (1978). *Biochemistry*, **17**, 3700–3705.
Gray, J. V., Eren, D. & Knowles, J. R. (1990). *Biochemistry*, **29**, 8872–8878.
Gray, J. V. & Knowles, J. R. (1994). *Biochemistry*, **33**, 9953–9959.

Haynes, M. R., Stura, E. A., Hilvert, D. & Wilson, I. A. (1994). *Science*, **263**, 646–652.
Kabsch, W. & Sander, C. (1983). *Bioploymers*, **22**, 2577–2637.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lee, A. Y., Karplus, P. A., Ganem, B. & Clardy, J. (1995). *J. Am. Chem. Soc.* **117**, 3627–3628.
Löwe, J. & Amos, L. A. (1998). *Nature* (*London*), **391**, 203–206.
Merritt, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* D**50**, 869–873.
Molecular Simulations (1997). *Crystallographic Workbench: X-GEN*. San Diego: Molecular Simulations.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Nogales, E., Wolf, S. G. & Downing, K. H. (1998). *Nature (London)*, **391**, 199–202.
Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B. & Thornton, J. M. (1997). *Structure*, **5**, 1093–1108.
Ramachandran, G. N. & Sasisekharan, V. (1968). *Adv. Protein Chem.* **23**, 283–438.
Reddy, P., Peterkofsky, A. & McKenney, K. (1989). *Nucleic Acids Res.* **17**, 10473–10488.
Richardson, J. S. (1981). *Adv. Protein Chem.* **34**, 167–339.
Roussel, A., Inisan, A.-G. & Cambillau, C. (1996). *AFMB and Bio Graphics.* Marseille, France.
Sheldrick, G. M. (1997). *SHELXL97. Crystal Structure Refinement Program*. University of Göttingen, Germany.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Sogo, S. G., Widlanski, T. S., Hoare, J. H., Grimshaw, C. E., Berchtold, G. A. & Knowles, J. R. (1984). *J. Amer. Chem. Soc.* **106**, 2701–2703.
Sträter, N., Schnappauf, G., Braus, G. & Lipscomb, W. N. (1997). *Structure*, **5**, 1437–1452.
Tronrud, D. (1997). *Methods Enzymol.* **277**, 306–319.
Tronrud, D., Ten Eyck, L. & Matthews, B. W. (1987). *Acta Cryst.* A**43**, 489–501.
Wilmot, C. M. & Thornton, J. M. (1990). *Protein Eng.* **3**, 479–493.
Xiao, G., Tordova, M., Jagadeesh, J., Drohat, A. C., Stivers, J. T. & Gilliland, G. L. (1999). *Proteins*, **35**, 13–24.
Xue, Y., Lipscomb, W. N., Gray, R., Schnappauf, G. & Braus, G. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 10814–10818.
Zuber, M., Patterson, T. A. & Court, D. L. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 4514–4518.