



**HAL**  
open science

## Crystal Structure of the Bifunctional Chorismate Synthase from *Saccharomyces cerevisiae*

Sophie Quevillon-Cheruel, Nicolas Leulliot, Philippe Meyer, Marc Graille, Michael Bremang, Karine Blondeau, Isabelle Sorel, Anne Poupon, Joël Janin, Herman van Tilbeurgh

► **To cite this version:**

Sophie Quevillon-Cheruel, Nicolas Leulliot, Philippe Meyer, Marc Graille, Michael Bremang, et al.. Crystal Structure of the Bifunctional Chorismate Synthase from *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 2003, 279, pp.619 - 625. 10.1074/jbc.m310380200 . hal-03299365

**HAL Id: hal-03299365**

**<https://hal.science/hal-03299365>**

Submitted on 26 Jul 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Crystal Structure of the Bifunctional Chorismate Synthase from *Saccharomyces cerevisiae*\*

Received for publication, September 18, 2003, and in revised form, October 14, 2003  
Published, JBC Papers in Press, October 21, 2003, DOI 10.1074/jbc.M310380200

Sophie Quevillon-Cheruel‡§, Nicolas Leulliot‡§, Philippe Meyer‡, Marc Graille¶, Michael Bremang‡, Karine Blondeau||, Isabelle Sorel¶, Anne Poupon¶, Joël Janin¶, and Herman van Tilbeurgh‡¶\*\*

From the ‡Institut de Biochimie et de Biophysique Moléculaire et Cellulaire (CNRS-UMR 8619), Université Paris-Sud, Bâtiment 430, 91405 Orsay, ¶Laboratoire d'Enzymologie et Biochimie Structurales (CNRS-UPR 9063), Bâtiment 34, 1 Avenue de la Terrasse, 91198 Gif sur Yvette, and the ||Institut de Génétique et Microbiologie (CNRS-UMR 8621), Université Paris-Sud, Bâtiment 360, 91405 Orsay, France

**Chorismate synthase (EC 4.2.3.5), the seventh enzyme in the shikimate pathway, catalyzes the transformation of 5-enolpyruvylshikimate 3-phosphate (EPSP) to chorismate, which is the last common precursor in the biosynthesis of numerous aromatic compounds in bacteria, fungi, and plants. The chorismate synthase reaction involves a 1,4-*trans*-elimination of phosphoric acid from EPSP and has an absolute requirement for reduced FMN as a cofactor. We have determined the three-dimensional x-ray structure of the yeast chorismate synthase from selenomethionine-labeled crystals at 2.2-Å resolution. The structure shows a novel  $\beta\alpha\beta\alpha$  fold consisting of an alternate tight packing of two  $\alpha$ -helical and two  $\beta$ -sheet layers, showing no resemblance to any documented protein structure. The molecule is arranged as a tight tetramer with D2 symmetry, in accordance with its quaternary structure in solution. Electron density is missing for 23% of the amino acids, spread over sequence regions that in the three-dimensional structure converge on the surface of the protein. Many totally conserved residues are contained within these regions, and they probably form a structured but mobile domain that closes over a cleft upon substrate binding and catalysis. This hypothesis is supported by previously published spectroscopic measurements implying that the enzyme undergoes considerable structural changes upon binding of both FMN and EPSP.**

In mammals, diet has to provide the major part of the so-called essential amino acids such as phenylalanine, tryptophan, and tyrosine. In bacteria, fungi, and plants, these aromatic amino acids are synthesized through the common complex shikimate pathway. Chorismate synthase (CS<sup>1</sup>; EC 4.2.3.5) is the seventh enzyme of this pathway, catalyzing the

conversion of 5-enolpyruvylshikimate-3-phosphate (EPSP) into chorismate (Fig. 1) (1). Chorismate is a building block for the synthesis of an array of aromatic compounds. It can, for instance, be converted into prephenate for the synthesis of phenylalanine and tyrosine or into anthranilate for the synthesis of tryptophan. In plants, chorismate is an essential substrate for the synthesis of *p*-aminobenzoate and folate. Because the shikimate pathway is only present in plants, fungi, and bacteria, enzymes of this pathway constitute attractive targets for antibiotics and herbicides. Recently enzymes of the shikimate pathway were also discovered in apicomplexan parasites, and herbicide inhibitors of this pathway limited *Toxoplasma gondii* infection in mice (2).

CS sequences are well conserved between fungi, plants, and bacteria. Most sequences code for proteins of about 370 residues, but some proteins are longer. Extensions are found at both N- and C-terminal regions depending on the organism. Chorismate synthases are soluble proteins that form homotetramers in solution (3). CD and Fourier transform-IR spectroscopic measurements showed that CS is composed of both  $\alpha$ -helices and  $\beta$ -strands. Based on secondary structure prediction, an  $(\alpha/\beta)_8$  barrel fold was proposed for the enzyme (4).

Formally, the reaction catalyzed by chorismate synthase is an *anti*-1,4-elimination of the 3-phosphate group and the C-6 pro-*R* hydrogen (Fig. 1) (5). Although this does not involve an overall change in redox state, the enzyme has an absolute requirement for reduced FMN (3, 6, 7). It was proposed that CS proceeds through the transient donation of an electron from FMN to EPSP to facilitate expulsion of the phosphate ion (7). For the *Escherichia coli* enzyme, experimental evidence for a radical mechanism was obtained. It was also shown that oxidized flavin and EPSP bind synergistically and induce major structural changes of the protein in solution (4).

Depending on their capacity to regenerate the reduced form of FMN, chorismate synthases are subdivided into two classes. Chorismate synthases from plants and eubacteria rely on an external system for the reduction of FMN, whereas the bifunctional enzymes from the fungi *Neurospora crassa* and *Saccharomyces cerevisiae* have an additional NADPH:FMN oxidoreductase activity (8). It was shown recently that the CS from *Plasmodium falciparum* is also monofunctional (9). Very little information is available on the interaction between the bifunctional enzymes and NADPH. One of the crucial issues regarding the catalysis of bifunctional chorismate synthases concerns the role and mechanism of action of NADPH in regeneration of enzyme activity. Fungal CSs have sequence extensions compared with their bacterial homologues. Truncated versions of the *Neurospora crassa* enzyme did not reveal a well

\* This work was supported by grants from the Ministère de la Recherche et de la Technologie (Programme Génopoles) and Association pour la Recherche Contre le Cancer (to M. G.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1R52 and 1R53) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

§ Both authors contributed equally to this work.

\*\* To whom correspondence should be addressed. E-mail: herman@lebs.cnrs-gif.fr.

<sup>1</sup> The abbreviations used are: CS, chorismate synthase; EPSP, 5-enolpyruvylshikimate 3-phosphate; SeMet, selenomethionine; r.m.s., root mean square.

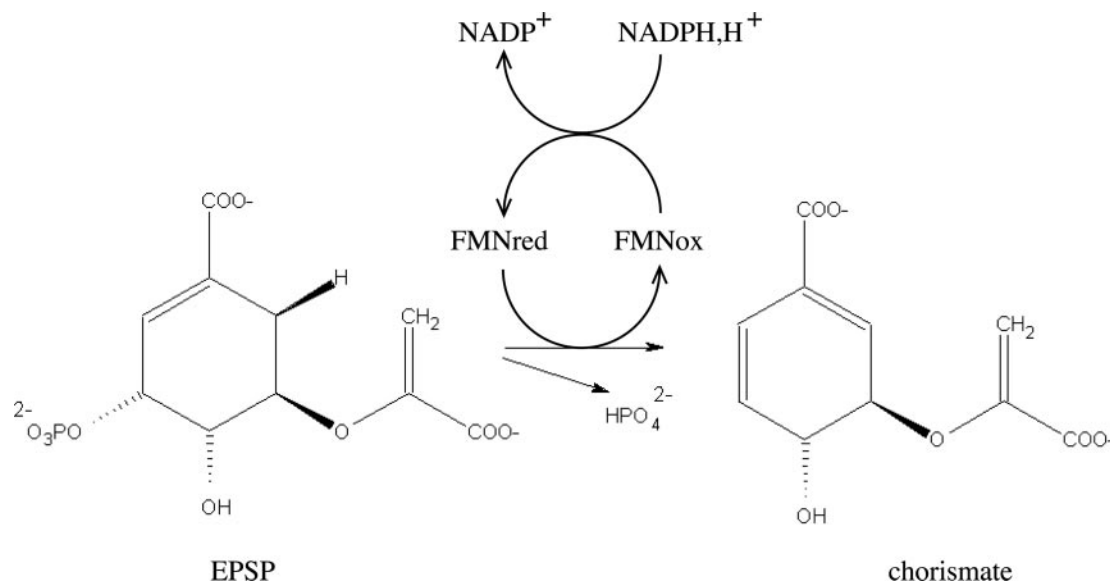


FIG. 1. Scheme illustrating the reaction catalyzed by CS. The bifunctional CS enzymes have an additional FMN oxidoreductase activity.

defined contiguous region in the sequence responsible for NADPH binding. Speculations about the involvement of these sequence extensions in interactions with NADPH were disproved by experiments (8).

Structural information on enzymes in the shikimate pathway is available for all but the seventh step. The crystallization and crystal characterization of a monofunctional chorismate synthase from *Helicobacter pylori* have been reported very recently (10). Structural information is needed (i) to obtain insights into the mechanism of this peculiar enzyme, (ii) to understand the distinction between the two classes (mono and bifunctional), and (iii) to investigate the potentiality of structure-based inhibitors. In this paper we describe the three-dimensional apo structure of a bifunctional enzyme from yeast at a resolution of 2.2 Å.

#### EXPERIMENTAL PROCEDURES

**Cloning, Expression, and Purification**—By using the sequenced *S. cerevisiae* S288C genomic DNA as a template for PCR, *YGL148w* was cloned between the NdeI and NotI restriction sites of a derivative of the pET9 vector (Stratagene), with an addition of His<sub>6</sub> codons at the 3'-end of the gene. Rosetta (DE3) pLysS (Novagen) was transformed with the construct and grown in 750 ml of 2× YT medium (Bio 101, Inc., Vista, CA) complemented with kanamycin, at 37 °C up to an A<sub>600 nm</sub> of 1. Expression of the recombinant protein was induced with 0.3 mM isopropyl-1-thio-β-D-galactopyranoside (Sigma) for a further 4 h. Cells were collected by centrifugation, suspended in 40 ml of 20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 5 mM β-SH, stored overnight at -20 °C, lysed by 2 cycles of freeze/thawing and sonication, and centrifuged at 13,000 × g. The His-tagged protein present in the supernatant was purified by affinity chromatography on nickel-nitrilotriacetic acid (Qiagen Inc.), followed by a gel filtration step on a Superdex 200 column (Amersham Biosciences) equilibrated in 20 mM sodium citrate, pH 5.6, 300 mM NaCl, 20 mM β-SH. The purity and integrity of the protein were checked by SDS-PAGE and mass spectrometry. The pure protein was subjected to an analytical size exclusion chromatography on a calibrated Superdex 200 column, in order to determine the oligomeric state in solution. The labeling of the protein with SeMet was conducted as described (11, 12).

**Crystallization and Structure Determination**—The His-tagged protein was crystallized at 293 K by the hanging drop vapor diffusion method in two different space groups (Table I). The crystals were difficult to obtain; their diffraction quality was very poor, and their transfer to a cryo-protecting solution was generally destructive. Crystals of the native protein were grown from 3:1 μl drops of protein (3 mg/ml), and precipitant solution containing 0.7 M sodium citrate, 0.1 M HEPES, pH 7.5, and belong to the P1 space group. Crystals of the SeMet protein grew in 1:1 μl drops of protein (8 mg/ml) and precipitant solution containing 0.7 M sodium citrate, 0.1 M sodium citrate, pH 5.6, 10 mM EDTA, 10 mM dithiothreitol. They belong to the I<sub>222</sub> space group.

The native crystals were transferred to 30% glycerol prior to flash-freezing in liquid nitrogen. The SeMet crystals were directly frozen in liquid nitrogen without the addition of cryo-protectant. The mother liquor was sufficient for protecting the crystals, and no ice rings were observed in the x-ray diffraction images. The structure was solved by single anomalous dispersion from 2.2-Å resolution data collected on crystals from the SeMet-labeled protein (11). A full data set was collected on the ID14-4 beam line at the ESRF (Grenoble, France). Data were reduced using the MOSFLM software, and further data treatment used the CCP4 software package (13). The SOLVE program retrieved 9 of 11 possible selenomethionine sites using the entire resolution range (14). Phases, calculated from the anomalous differences and from the SeMet positions, yielded an interpretable electronic density map. The quality of the map was considerably improved by applying solvent flattening as implemented in the program RESOLVE (14). This map allowed the building of 260 of 371 residues making use of the ARP/WARP software (15). The model could be partially completed and refined using the program REFMAC and the molecular graphics crystallographic software O (16, 17). The final model contains 286 of the 371 residues in the protein and 116 water molecules. As can be judged from Table I, the crystallographic refinement data are excellent. The structure refined to an *R* factor of 17% (*R*<sub>free</sub> 22%). As discussed, the residues missing in the structure are due to mobility.

Data on the native P1 crystal form were collected on the ID14 beam line at the ESRF to a resolution of 2.9 Å. The structure of the P<sub>1</sub> form contained four molecules in the asymmetric unit, and the structure was solved by molecular replacement (using the program AMORE (18)) with the I<sub>222</sub> crystal structure of the molecule as search model. A solution was found for the four copies of the subunit in the crystal, and the structure was refined without non-crystallographic symmetry constraints to a crystallographic *R*<sub>factor</sub> of 19% (*R*<sub>free</sub> 22%). Coordinates for both crystal forms have been deposited (Protein Data Bank codes 1R52 and 1R53).

#### RESULTS AND DISCUSSION

**Structure Determination**—In order to determine the crystal structure from the bifunctional CS from yeast, we cloned and overexpressed the protein in *E. coli*. The expression levels of the recombinant enzyme were good, but due to precipitation during the concentration steps the final purification yields were low. The integrity of the purified samples was tested by mass spectrometry and SDS-PAGE. Crystals of the native protein were generally more difficult to obtain than the SeMet-labeled protein and diffracted to lower resolution. The native and SeMet proteins, carrying a C-terminal His tag, were crystallized under two different sets of conditions in two different space groups (Table I). SeMet-labeled protein crystallized in space group I<sub>222</sub>, and the structure could be determined from single anomalous dispersion data at 2.2-Å resolution. As can be

TABLE I  
Data collection and refinement statistics

	Peak (SeMet labeled)	Native
Data		
Wavelength (Å)	0.979	0.934
Space group	I <sub>222</sub>	P <sub>1</sub>
Unit cell parameters		
a, b, c (Å)	57.5, 74.7, 155.0	57.8, 75.6, 91.3
α, β, γ	90, 90, 90	114.4, 108.4, 90
Resolution (Å) <sup>a</sup>	27–2.2	30–2.9
Total no. of reflections	63,849	11,0781
Total no. of unique reflections	17,154	26,514
R <sub>merge</sub> <sup>b</sup>	0.05 (0.15)	0.04 (0.3)
I/σ(I)	10.9 (4.9)	21.5 (9.4)
Overall completeness (%)	96.4 (77.4)	90 (62.3)
Phasing		
No. of sites (Se)	9/11	
FOM No. (before/after DM)	0.35/0.58	
Refinement		
Resolution (Å)	27.3–2.2	29.7–2.88
Reflections (working/test)	16,076/852	23,826/1344
R <sub>cryst</sub> /R <sub>free</sub> <sup>c</sup>	0.174/0.225	0.188/0.224
Non-hydrogen atoms	2236	8451
Water molecules	113	98
r.m.s. deviation		
Bonds (Å)	0.034	0.015
Angles (°)	1.72	1.46
Mean B factor (Å <sup>2</sup> )	28.1	32.3
Ramachandran analysis		
Most-favored (%)	91.4	86
Allowed (%)	8.2	13.1

<sup>a</sup> Numbers in parentheses are for highest resolution shell.

<sup>b</sup>  $R_{\text{merge}} = \sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i I_{hi}$ , where  $I_{hi}$  is the  $i$ th observation of the reflection  $h$ , while  $\langle I_h \rangle$  is the mean intensity of reflection  $h$ .

<sup>c</sup>  $R_{\text{factor}} = \sum |F_o| - |F_c| / |F_o|$ .  $R_{\text{free}}$  was calculated with a set of randomly selected reflections (5%).

judged from the crystallographic refinement statistics, the quality of the overall structure is good. Nevertheless, electron density was missing for significant portions (~23%) of the protein, and therefore part of the enzyme seems to be disordered in the crystal. The missing parts in our crystallographic model cover discontinuous segments of the sequence (Fig. 2). Inspection of the location of these mobile parts in the overall structure reveals that they are mainly clustered in the same region of the protein surface and could therefore form a sub-structure or domain within the protein (see below).

Native CS protein crystallized in the P<sub>1</sub> space group. Its structure could be solved by molecular replacement from the refined structure in the I<sub>222</sub> crystal form. Analysis of the crystal content revealed the presence of four subunits in the asymmetric unit. We used the structure of the monomer as a search model and found a molecular replacement solution for the four copies. The CS structures are identical between the two crystal forms and between the four subunits of the P<sub>1</sub> crystal form (r.m.s. ~0.44 Å for all Cα positions). The regions with missing electron density in the P<sub>1</sub> structure are the same for the four copies of the molecule in the asymmetric unit and those observed in the I<sub>222</sub> crystal form. Because the quality of the SeMet data was considerably better than those of the native P<sub>1</sub> crystals, we used the former structure for the discussion of our results.

**Overall Structure of Yeast Chorismate Synthase**—The yeast CS subunit has a flat compact shape with overall dimensions 45 × 55 × 35 Å<sup>3</sup>. The structure belongs to the α/β family and to our knowledge represents a novel fold. The topology diagram is illustrated in Fig. 3a. A search for structural homologues with the MSD server (www.ebi.ac.uk/msd/) did not reveal any similarity to previously determined protein structures. The core of the enzyme consists of two alternate β-sheet and α-helical layers with complex topology (Fig. 3b). These layers are packed

to form a compact structure. The first layer S1 in Fig. 3 consists of an anti-parallel β-sheet comprising five strands in the order β1-β2-β3-β8-β5. This sheet packs against the first layer of helices A1, containing α1, α2, α8, and α6. The α2-helix connects sheet S1 to the other β-sheet S2. In between, three short strands β4-β6-β7 are arranged in a small sheet that forms a cap to sheet S1. The A1 helical layer is sandwiched between S1 and the mixed parallel anti-parallel β-sheet S2, consisting of five strands in the following order β12-β15-β11-β9-β10. The second α-layer A2 covers the other side of the S2 sheet. Three helices (α3, α4, and α5) are inserted between strands β10 and β11 and cover part of the surface and the rim of the S2 sheet. Helix α7 of the A2 layer is part of the connection between strands β12 and β15. The same connection contains two small anti-parallel strands β13 and β14 that form a cap to the S2 sheet. The long C-terminal helix that is at the center of the A1 helical layer is situated at the exit of the last S2-strand (β15).

The structure could also be described as consisting of two intimately contacting sub-domains, consisting of anti-parallel sheets covered by helices: one contained between residues 1 and 147 (between β1 and α2 on the topology diagram) and the second domain from residue 148 to the C terminus (from β9 to α8). Search for structural homologues using these fragments yielded numerous protein fragments with weak structural analogy (typically Z scores of around 3 and high r.m.s. values for a limited number of superposed Cα positions). We therefore conclude that the molecule has a complex and novel topology.

**Analysis of the Disordered Regions**—Although the quality of the structure is satisfactory and electron density is very clear for the major part of the structure, 85 residues are missing from the final model. In Fig. 2, the missing regions from the model have been marked by stars below the sequence alignment. The missing residues are not contiguous in the sequence but are divided over a few large sequence blocks (named L1 to L4): between α1 and β5 (L1, residues 49–60), β8 and α2 (L2, residues 86–126), α7 and β13 (L3, residues 277–284), and β15 and α8 (L4, residues 314–336). No density was detected either for the five C-terminal residues. L1, L2, and L4 cover some of the best conserved regions of the molecule and are therefore probably important for function (Fig. 2). On Fig. 4a, we have schematically represented the missing regions as loops in dashed lines onto the protein structure. Inspection of the location of these disordered regions in the three-dimensional structure shows that the boundaries of L1, L2, and L4, the most conserved missing regions, all concentrate around a pocket situated at the center of the A1 helical layer. In both crystal forms, regions L1, L2, and L3 are pointing toward a crystal cavity, at a distance of 10 Å from the closest crystal contact. The N-terminal edge of region L3 is directly involved in such contact and could be affected by the crystal packing. In Fig. 4b, the accessible surface of the monomer is colored according to amino acid sequence conservation, showing that this pocket is significantly enriched in conserved residues highlighting its importance for CS function (see below).

**Quaternary Structure**—Chorismate synthases from other organisms were reported to form tetramers in solution (3). According to gel filtration experiments (results not shown), yeast CS also forms tetramers in solution and is present as identical tetramers in the two crystal forms obtained under different crystallization conditions, generated either by crystal symmetry (I<sub>222</sub> crystal form) or by local symmetry (P<sub>1</sub> crystal form). In both space groups, the inter-tetramer contacts involve similar regions around helices α3, α4, α5, and α7, all situated at the top and bottom edges of the tetramer. The r.m.s. deviation for all Cα positions between tetramers in the two crystal forms is 0.48 Å. The subunit packing in the tetramer is very tight, and the

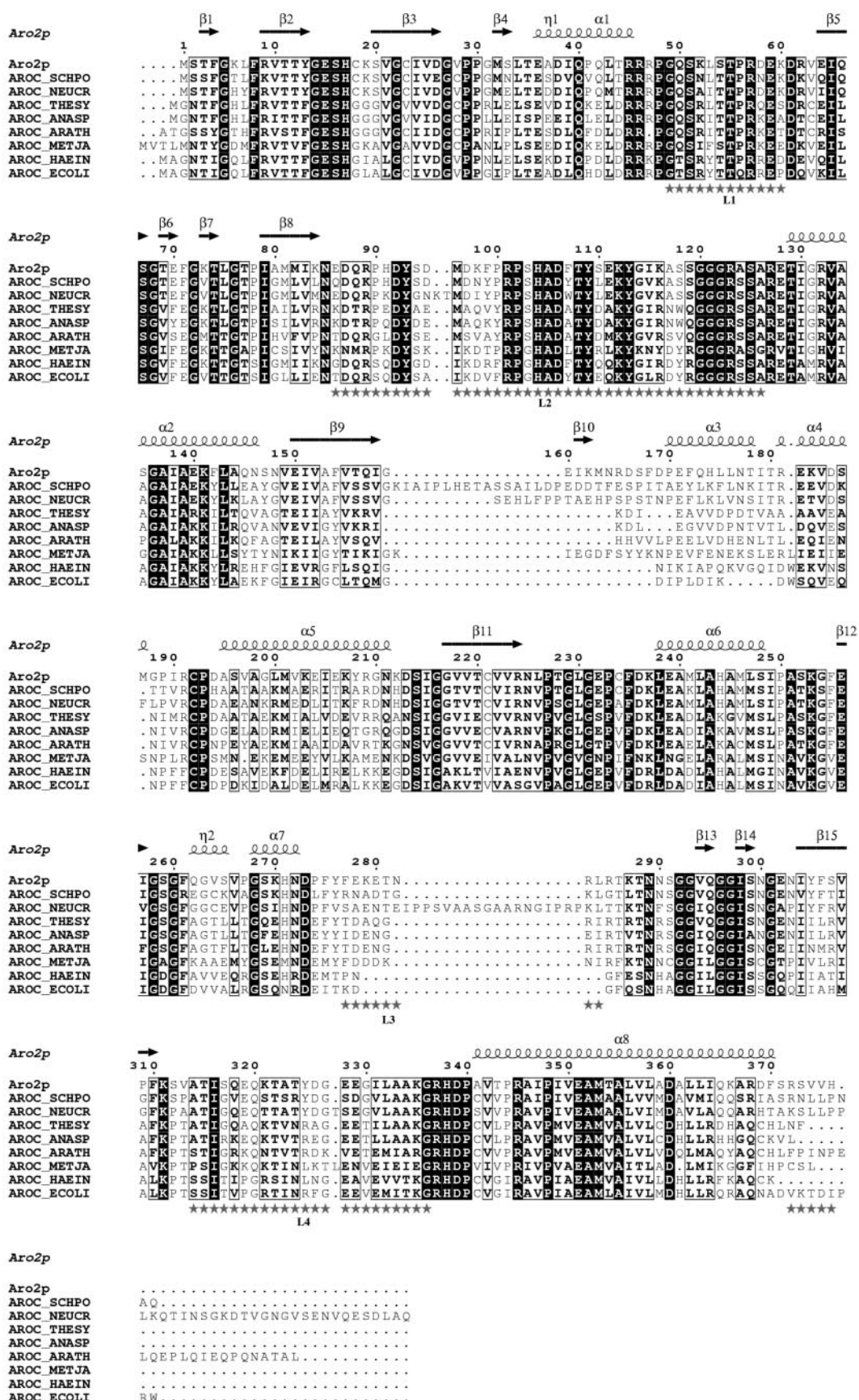


FIG. 2. Sequence alignment of some representative sequences of bi-functional and mono-functional CS enzymes. The secondary structure elements as derived from the present crystal structure of yeast CS are superposed. The figure is generated by ESPript (31). The parts of the sequence with unidentified electron density are marked by stars. Aro2p, *S. cerevisiae*; SCHPO, *S. pombe*; NEUCR, *N. crassa*; THESY, *Thermosynechococcus elongatus*; ANASP, *Nostoc* sp., ARATH, *Arabidopsis thaliana*; METJA, *Methanococcus jannaschii*; HAEIN, *Haemophilus influenzae*; ECOLI, *E. coli*. An N-terminal 50-residue extension in *A. thaliana* has been omitted from the figure.

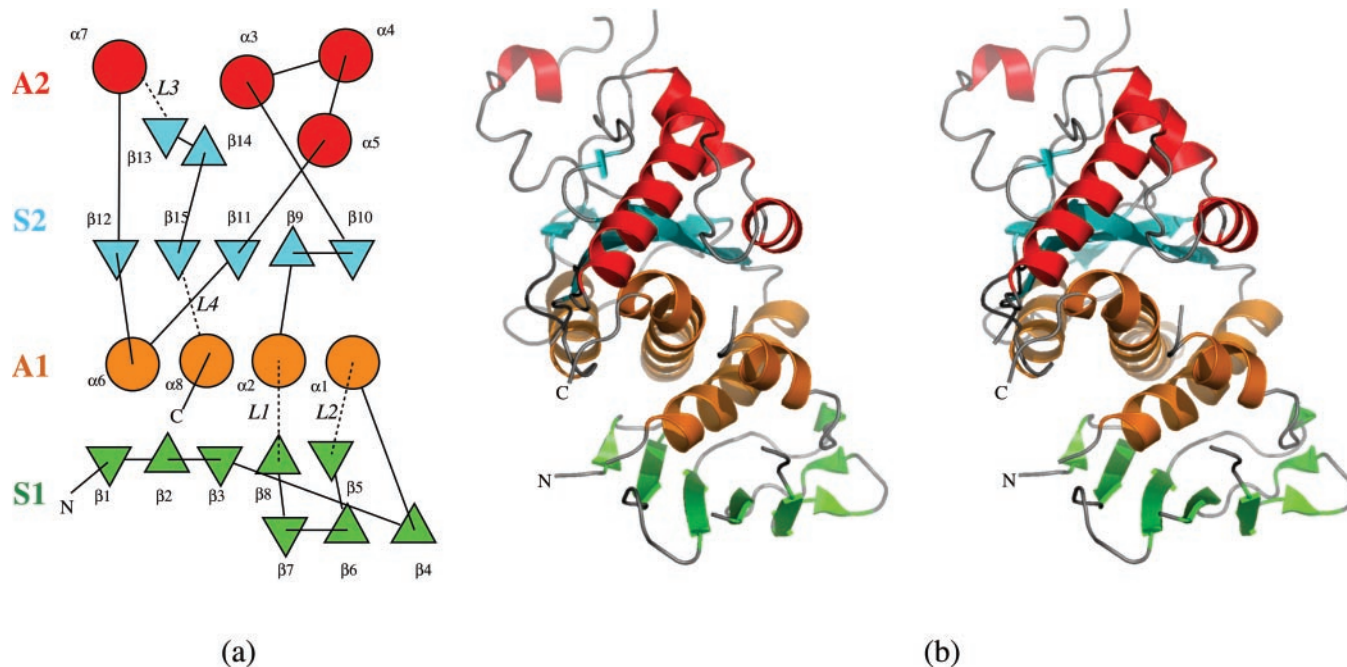


FIG. 3. *a*, topology representation of yeast CS (32). The four layers are color-coded *green* (S1), *orange* (A1), *blue* (S2), and *red* (A2). The orientation of the strands of the  $\beta$ -sheets and the  $\alpha$ -helices of the A1  $\alpha$ -helical layer are roughly perpendicular to the figure. The location of the L1, L2, L3, and L4 disordered regions are indicated. *b*, stereo ribbon presentation of the overall fold of the yeast CS structure in the same orientation as the topology diagram. The layers are colored as in *a*.

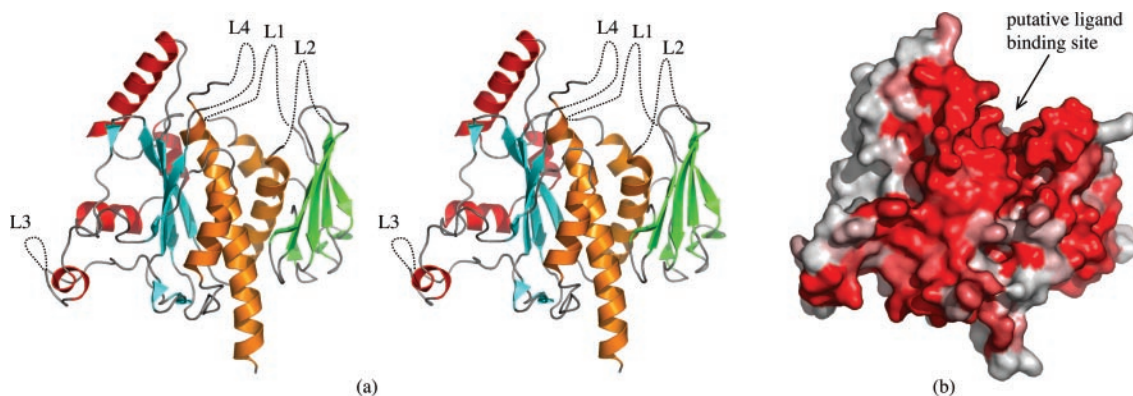
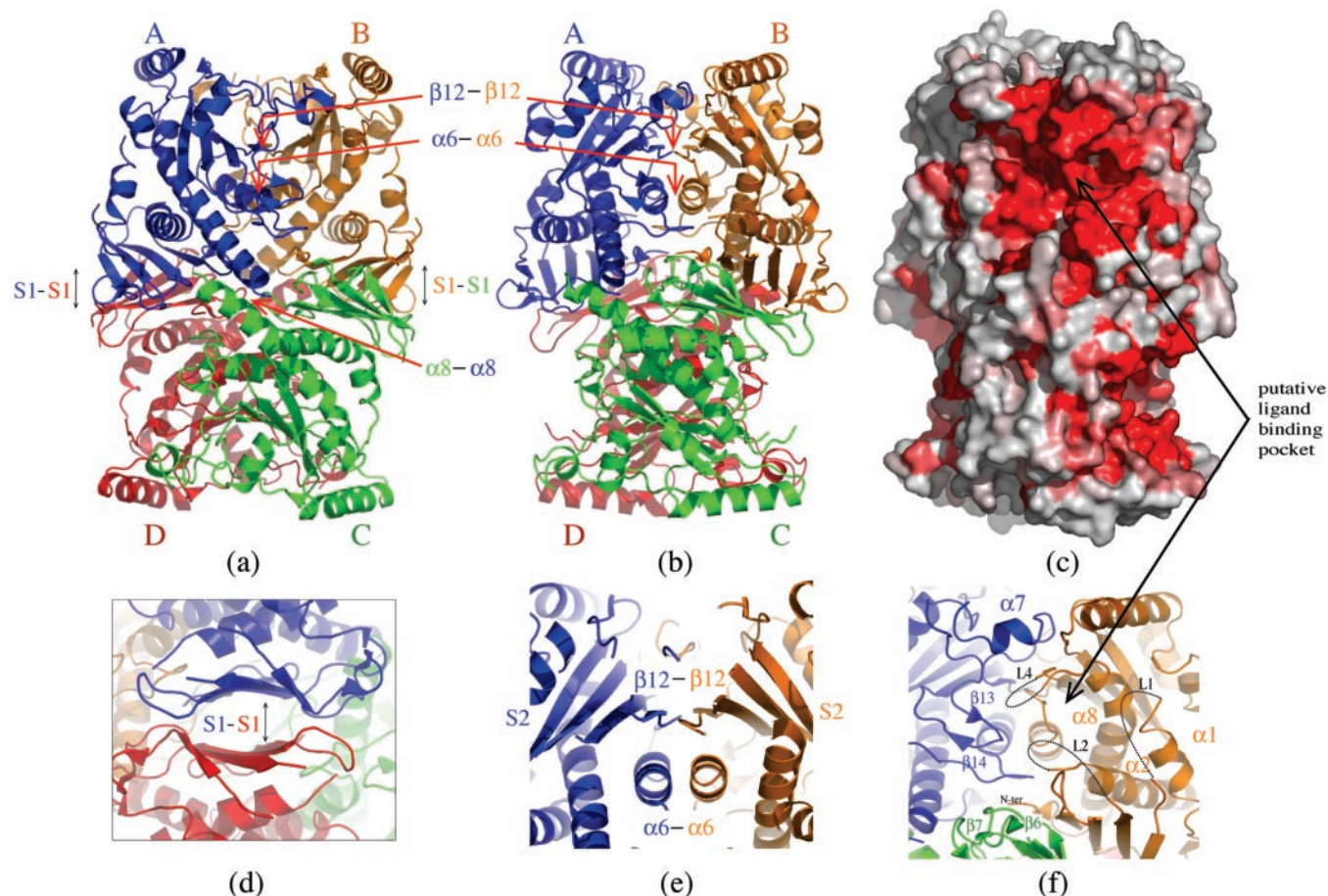


FIG. 4. *a*, stereo ribbon representation of the CS monomer, with the missing regions schematically represented as *black interrupted lines*: L1 (12 residues), L2 (41 residues), L3 (8 residues), and L4 (23 residues). See text and Fig. 2 for sequence identification of these segments. L1, L2, and L4 cluster in the same region of the molecule. *b*, surface representation of the CS monomer (same orientation as *a*). The surface is colored in *shades of red*, according to a score based on the conservation of the surface residue determined after alignment of 50 CS sequences (consurf.tau.ac.il/). The highly conserved region coincides with a cleft on the protein surface and could represent a putative ligand binding pocket surrounded by the flexible regions represented in *a*.

tetramer has a brick-like shape of dimensions  $86 \times 62 \times 43 \text{ \AA}^3$  (Fig. 5). The CS tetramer structure possesses D<sub>2</sub> symmetry, compatible with the I<sub>222</sub> crystal symmetry. Each subunit is in contact with the three others, creating an intricate packing arrangement. In total, the monomer buries an extensive part of its accessible surface area upon tetramer formation ( $3710 \text{ \AA}^2$  buried per monomer, *i.e.* 29% of the total accessible surface area). The extent of this surface and the number of residues implied in inter-molecular contacts convinced us that this tetramer corresponds to the biologically active tetramer (19). We now dissect the interactions between one monomer and its three partners in the tetramer. Two perpendicular views of the tetramer structure are represented in Fig. 5, *a* and *b*. The most extensive contact is between the A and B subunits ( $2170 \text{ \AA}^2$  or 17% accessible surface area buried). The contact surface involves the first and second  $\alpha$ -layers (A1 and A2) and sheet S2. Fig. 5 also illustrates two prominent features of this dimer contact. First, the two  $\beta$ 12-strands from two monomers form an

anti-parallel  $\beta$ -sheet, and therefore the S2 sheet forms a continuous 10-stranded  $\beta$ -sheet (Fig. 5, *b* and *e*). Second, the  $\alpha$ 6-helices of two molecules are packed over their whole length, forming an anti-parallel bundle (Fig. 5, *b* and *e*). Finally, there is an important contribution of the packing of loops from the  $\alpha$ -layer A2. The association of these two monomers is a mixture of hydrophobic packing and polar interactions (26 hydrogen bonds are counted for this part of the tetramer). The interaction of the A and D monomers consists entirely of a perpendicular stacking of the flat surface of the S1 sheet (Fig. 5, *a* and *d*). It buries  $935 \text{ \AA}^2$  (7% of the total accessible surface area) and is also stabilized both by hydrophobic and polar interactions (12 hydrogen bonds). Finally, there exists a contact between the A and C molecules, through the anti-parallel packing of the ends of their C-terminal  $\alpha$ 8-helices (burying  $734 \text{ \AA}^2$  or 5% of the total accessible surface area) (Fig. 5*a*).

*Putative Binding Pocket*—At present, nothing is known about the identity and even the nature of the active site resi-



**FIG. 5. Representation of the CS tetramer with the most prominent elements involved in packing indicated.** The four monomers A–D are colored *blue, orange, green, and red*, respectively. *a*, view showing the packing between the A/D (packing of the S1 sheet) and A/C monomers ( $\alpha 8/\alpha 8$  interaction). *b*, view looking down at the A/B dimer interface illustrating the  $\alpha 6/\alpha 6$  and  $\beta 12/\beta 12$  interactions. *c*, same view as *b* in surface representation color-coded according to residue conservation as in Fig. 4, showing the putative ligand-binding site at the interface of the A/B/C monomers. *d*, detailed view of the S1-S1  $\beta$ -sheet packing between monomers A and D. *e*, detailed view of the  $\alpha 6/\alpha 6$  and  $\beta 12/\beta 12$  interactions between monomers A and B. *f*, structural elements involved in the putative ligand binding pocket (same orientation as *c*). The L1, L2, and L4 disordered regions are indicated as *broken lines*.

dues in chorismate synthases. In the absence of biochemical data, it is difficult to identify with certainty the active site pocket. We have failed so far to obtain any diffracting crystals of complexes of the enzyme with its various substrates. However, from the analysis of the structure, we can make reasonable assumptions about the location of the active site.

Conserved surface residues are often good indicators of functionally important regions in proteins. In the CS monomer, the conserved residues are spread out over the total length of the protein (Fig. 4*b*). The tetrameric association of CS clusters the majority of these residues in a pronounced surface pocket (Fig. 5*c*), whereas other conserved surface patches are involved in tetramer packing. The walls of this pocket are formed by segments coming from three different molecules (A–C in Fig. 5, *c* and *f*): the A1  $\alpha$ -helical layer of monomer A ( $\alpha 6$ ,  $\alpha 8$ ,  $\alpha 2$ , and  $\alpha 1$ );  $\beta 12$ ,  $\alpha 7$ ,  $\beta 13$ , and  $\beta 14$  of monomer B; and the  $\beta 6/\beta 7$  region of monomer C (Fig. 5*f*). The boundaries of the L1, L2, and L4 disordered regions are positioned at the rim of this cavity. All these regions contain stretches of absolutely conserved residues as can be seen from Fig. 5*c*. We therefore consider this pocket as an excellent candidate for the active site.

It was reported that the presence of an N-terminal tag resulted in an inactive form of the enzyme (9, 20, 21). Inspection of the structure shows that the N-terminal residue is buried at the interface in a region where three molecules of the tetramer are in contact (Fig. 5*f*). Any terminal extension will therefore

sterically interfere with tetramer association, and perturbation of this process very likely explains the absence of enzymatic activity in N-terminal fusion constructs. Our genetic construct of yeast CS carries the His tag at the C terminus and does not interfere with subunit packing.

Although the reaction catalyzed by chorismate synthase does not change the redox state of the substrate, the activity of the enzyme has an absolute requirement for reduced FMN. It was shown that CS forms ternary complexes with EPSP and FMN, both substrates binding synergistically to the enzyme (22, 23). Physical and spectroscopic studies showed that CS undergoes a major structural change when both FMN and EPSP are bound. Native electrophoresis and small angle x-ray scattering data show that the complex has a more compact overall shape than the free enzyme (4). The oligomerization state of the enzyme, however, does not change. The enzyme is also more resistant to proteolytic attack when both FMN and EPSP are bound. Far-UV CD spectroscopy suggested that the secondary structure content does not change significantly upon substrate binding. All these observations suggest that there exists in the protein a disordered region that becomes more rigid upon substrate binding. Interpreting these observations in the light of our structural results leads to a coherent model of events during catalysis. The compact structure of the tetramer suggests that the conformational changes seen upon substrate binding do not involve a rearrangement of the quaternary

structure. The three missing regions (L1, L2, and L4) are ideally positioned to fold over a substrate that would bind into the aforementioned pocket (Fig. 5f). The high proportion of totally conserved residues in these regions indeed suggests that they could contribute to substrate binding and/or catalysis. The missing regions may fold over the substrate to exclude solvent from the reaction pocket. This would lead to a more compact structure as observed in the x-ray scattering experiments.

Oxidation of FMN causes inactivation of the enzyme, and activity can only be maintained by keeping FMN in the reduced form. Bifunctional chorismate synthases from the fungi *Neurospora crassa* and *Saccharomyces cerevisiae* have an additional NADPH:FMN oxidoreductase activity (8). Based on a sequence alignment with NADPH-binding modules, the amino acids involved in NADPH binding were predicted to be contained between residues 250 and 265 (24). This region contains a GXGXX(G/A) dinucleotide-binding fingerprint sequence, usually present in the first  $\beta$ - $\alpha$ - $\beta$  motif of dinucleotide binding domains (25, 26). CS, however, does not possess such a dinucleotide binding domain. The structural context supporting the fingerprint sequence in CS is totally different from that found in the canonical NAD domains (not shown). There is no experimental evidence for the moment indicating that this motif is involved in dinucleotide binding. In the case of the CS from *N. crassa*, these residues were also mapped to a predicted ( $\alpha/\beta$ )-barrel fold structure, but this prediction is not confirmed by the present crystal structure. In yeast CS the region containing the fingerprint sequence lines a groove formed at the contact between two monomers close to the  $\beta$ 10-strand. The floor of the groove is conferred by residues from the S2-sheet, and the walls are formed by loops and helices from the second  $\alpha$ -layer. This region, however, is not connected to the predicted active site pocket.

Although a large majority of NAD-binding proteins possesses the classical Rossmann fold or variants of this, others have totally different structures (26). Aldose reductase (27) has a ( $\alpha/\beta$ )<sub>8</sub> barrel structure and catalase (28), isocitrate dehydrogenase (29), and trichosantin (30) are all multidomain proteins in which NAD binds within an interdomain crevice. CS has no structural resemblance to any of these enzymes.

#### CONCLUSION

The crystal structure of chorismate synthase reveals a novel fold, consisting of a four-layered  $\beta\alpha\beta\alpha$  motif. The protein forms a very compact tetramer, confirming previous observations in solution. 85 residues, spread over three main sequence regions, have no defined electron density. These regions are at the circumference of an extended, well conserved pocket that may harbor part of the active site. The disordered regions most likely form a domain that closes above the substrate during the

catalytic cycle. Co-crystallization trials with substrates and inhibitors are underway to further define the active site and mechanism of this enzyme.

*Acknowledgments*—We thank the staff of the ESRF beam lines for help with data collection. We also thank Dr. Richard Buckingham for critical reading of the manuscript.

#### REFERENCES

- Bentley, R. (1990) *Crit. Rev. Biochem. Mol. Biol.* **25**, 307–384
- Roberts, F., Roberts, C. W., Johnson, J. J., Kyle, D. E., Krell, T., Coggins, J. R., Coombs, G. H., Milhous, W. K., Tzipori, S., Ferguson, D. J., Chakrabarti, D., and McLeod, R. (1998) *Nature* **393**, 801–805
- White, P. J., Millar, G., and Coggins, J. R. (1988) *Biochem. J.* **251**, 313–322
- Macheroux, P., Schonbrunn, E., Svergun, D. I., Volkov, V. V., Koch, M. H., Bornemann, S., and Thorneley, R. N. (1998) *Biochem. J.* **335**, 319–327
- Hill, R. K., and Newkome, G. R. (1969) *J. Am. Chem. Soc.* **91**, 5893–5894
- Welch, G. R., Cole, K. W., and Gaertner, F. H. (1974) *Arch. Biochem. Biophys.* **165**, 505–518
- Bornemann, S., Lowe, D. J., and Thorneley, R. N. (1996) *Biochemistry* **35**, 9907–9916
- Henstrand, J. M., Schaller, A., Braun, M., Amrhein, N., and Schmid, J. (1996) *Mol. Microbiol.* **22**, 859–866
- Fitzpatrick, T., Ricken, S., Lanzer, M., Amrhein, N., Macheroux, P., and Kappes, B. (2001) *Mol. Microbiol.* **40**, 65–75
- Ahn, H. J., Yang, J. K., Lee, B. I., Yoon, H. J., Kim, H. W., and Suh, S. W. (2003) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **59**, 569–571
- Hendrickson, W. A., Horton, J. R., and Lemaster, D. M. (1990) *EMBO J.* **9**, 1665–1672
- Van Duyne, P., Ray, M. V., Bertelsen, A. H., Jackson-Matthews, D. E., Sturmer, A. M., Merkler, D. J., Consalvo, A. P., Young, S. D., Gilligan, J. P., and Shields, P. P. (1993) *Bio/Technology* **11**, 64–70
- Collaborative Computational Project 4 (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50**, 760–763
- Terwilliger, T. C., and Berendzen, J. (1997) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55**, 849–861
- Perrakis, A., Morris, R. J., and Lamzin, V. S. (1999) *Nat. Struct. Biol.* **6**, 458–463
- Murshudov, G., Vagin, A., and Dodson, E. (1997) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53**, 240–255
- Jones, T. A., Zou, J. H., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47**, 110–119
- Navaza, J. (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53**, 157–163
- Lo Conte, L., Chothia, C., and Janin, J. (1999) *J. Mol. Biol.* **285**, 2177–2198
- Henstrand, J. M., Schmid, J., and Amrhein, N. (1995) *Plant Physiol.* **108**, 1127–1132
- Braun, M., Henstrand, J. M., Grolach, J., Amrhein, N., and Schmid, J. (1996) *Planta* **200**, 64–70
- Macheroux, P., Petersen, J., Bornemann, S., Lowe, D. J., and Thorneley, R. N. (1996) *Biochemistry* **35**, 1643–1652
- Macheroux, P., Bornemann, S., Ghisla, S., and Thorneley, R. N. (1996) *J. Biol. Chem.* **271**, 25850–25858
- Henstrand, J. M., Amrhein, N., and Schmid, J. (1995) *J. Biol. Chem.* **270**, 20447–20452
- Baker, P. J., Britton, K. L., Rice, D. W., Rob, A., and Stillman, T. J. (1992) *J. Mol. Biol.* **228**, 662–671
- Carugo, O., and Argos, P. (1997) *Proteins* **28**, 29–40
- Wilson, D. K., Bohren, K. M., Gabbay, K. H., and Quijoco, F. A. (1992) *Science* **257**, 81–84
- Gouet, P., Jouve, H. M., and Dideberg, O. (1995) *J. Mol. Biol.* **249**, 933–954
- Mesecar, A. D., Stoddard, B. L., and Koshland, D. E., Jr. (1997) *Science* **277**, 202–206
- Xiong, J. P., Xia, Z. X., and Wang, Y. (1994) *Nat. Struct. Biol.* **1**, 695–700
- Gouet, P., Courcelle, E., Stuart, D. I., and Metz, F. (1999) *Bioinformatics* **15**, 305–308
- Westhead, D. R., Hatton, D. C., and Thornton, J. M. (1998) *Trends Biochem. Sci.* **23**, 35–36