

near to the last carbon of LDAO (Fig. 7) where the protein offers no suitable residue. Because the reported S-cyclase shows a side reaction (10) resulting in about 10% diplopterol (hopan-22-ol), we suggest that B<sub>2</sub> is a water molecule polarized by other waters that are in contact with the hydrogen-bonding network of Gln<sup>262</sup>:Glu<sup>45</sup>:Glu<sup>93</sup>:Arg<sup>127</sup>, which could store a proton (Fig. 7). Diplopterol is formed if the front water adds as hydroxyl to the last carbocation instead of accepting the proton.

Taken together, squalene enters the central cavity through the constriction, which acts as a gate, and is forced into the unfavorable conformation necessary for catalysis. The scale of the required actions corresponds to the low turnover number of 0.3 s<sup>-1</sup> (12, 16). Hopene formation releases ~200 kJ/mol at the bottom of the α<sub>6</sub>-α<sub>6</sub> barrel, exceeding by far the usual protein stabilization energy of ~50 kJ/mol. The barrel does not disintegrate because its α helices at the surface are connected and thus stabilized by the QW-motifs, characteristic for this enzyme family. The resulting excitation then facilitates the return of hopene to the membrane.

## REFERENCES AND NOTES

- E. J. Corey, S. P. T. Matsuda, B. Bartel, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2211 (1994).
- A. Eschenmoser, L. Ruzicka, O. Jeger, D. Arigoni, *Helv. Chim. Acta* **38**, 1890 (1955).
- G. Ourisson and M. Rohmer, *Acc. Chem. Res.* **25**, 403 (1992).
- R. B. Woodward and K. Bloch, *J. Am. Chem. Soc.* **75**, 2023 (1953); J. W. Cornforth *et al.*, *ibid.* **87**, 3224 (1965); I. Abe, M. Rohmer, G. D. Prestwich, *Chem. Rev.* **93**, 2189 (1993).
- E. J. Corey *et al.*, *Biochem. Biophys. Res. Commun.* **219**, 327 (1996); C. H. Baker, S. P. T. Matsuda, D. R. Liu, E. J. Corey, *ibid.* **213**, 154 (1995); I. Abe and G. D. Prestwich, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 9274 (1995); Z. Shi, C. J. Buntel, J. H. Griffin, *ibid.* **91**, 7370 (1994); E. J. Corey, S. P. T. Matsuda, B. Bartel, *ibid.* **90**, 11628 (1993); C. A. Roessner *et al.*, *Gene* **127**, 149 (1993); T. Kaneko *et al.*, *DNA Res.* **3**, 109 (1996); I. G. Reipen, K. Poralla, H. Sahm, G. A. Sprenger, *Microbiology* **141**, 155 (1995); D. Ochs *et al.*, *J. Bacteriol.* **174**, 298 (1992).
- E. E. van Tamelen, *J. Am. Chem. Soc.* **104**, 6480 (1982); R. B. Boar, L. A. Couchman, A. J. Jacques, M. J. Perkins, *ibid.* **106**, 2476 (1984).
- W. S. Johnson *et al.*, *ibid.* **109**, 5852 (1987); C. J. Buntel and J. H. Griffin, *ibid.* **114**, 9711 (1992).
- Y. F. Zheng *et al.*, *J. Am. Chem. Soc.* **117**, 670 (1995).
- G. Balliano, F. Viola, M. Ceruti, L. Cattel, *Arch. Biochem. Biophys.* **293**, 122 (1992).
- B. Seckler and K. Poralla, *Biochim. Biophys. Acta* **881**, 356 (1986).
- Crystal forms A, A', and A'' grew under identical conditions in hanging drops starting at a protein concentration of 8 mg/ml, 0.3% (w/v) C<sub>8</sub>E<sub>4</sub> (*n*-octyl-tetraoxyethylene), 50 mM sodium citrate, pH 4.8, with a reservoir at 100 mM sodium citrate, pH 4.8 [K.-U. Wendt *et al.*, *Protein Sci.* **6**, 722 (1997)]. For the refinement we used a form A crystal that grew with 0.01% (w/v) LDAO as an additive in the starting drop. Diffraction data were collected with an area detector (Siemens X-1000) on a rotating anode (Rigaku RU200B) and processed with XDS [W. Kabsch, *J. Appl. Crystallogr.* **21**, 916 (1988)].
- S. Neumann and H. Simon, *Biol. Chem. Hoppe Seyler* **367**, 723 (1986).
- P. M. Alzari, H. Souchon, R. Dominguez, *Structure* **4**, 265 (1996); M. Juy *et al.*, *Nature* **357**, 89 (1992); A. Aleshin *et al.*, *J. Biol. Chem.* **267**, 19291 (1992); H.-W. Park *et al.*, *Science* **275**, 1800 (1997).
- Cavity volume (probe included) and solvent-accessible surface areas were calculated with a 1.4 Å radius probe [M. L. Connolly, *Science* **221**, 709 (1983)].
- K. Poralla *et al.*, *Trends Biochem. Sci.* **19**, 157 (1994).
- D. Ochs *et al.*, *Eur. J. Biochem.* **194**, 75 (1990).
- I. Abe and G. D. Prestwich, *J. Biol. Chem.* **269**, 802 (1994).
- E. J. Corey *et al.*, *J. Am. Chem. Soc.* **119**, 1277 (1997); E. J. Corey *et al.*, *ibid.*, p. 1289.
- C. Feil, R. Süßmuth, G. Jung, K. Poralla, *Eur. J. Biochem.* **242**, 51 (1996).
- G. Blobel, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1496 (1980).
- S. A. Simon *et al.*, *Biophys. J.* **19**, 83 (1977).
- D. Picot *et al.*, *Nature* **367**, 243 (1994).
- Mobilities are equated with crystallographic *B* factors.
- L. C. Tarshis, M. Yan, C. D. Poulter, J. C. Sacchettini, *Biochemistry* **33**, 10871 (1994).
- CCP4, Collaborative Computing Project Number 4, *Acta Crystallogr.* **D50**, 760 (1994).
- T. A. Jones *et al.*, *ibid.* **A47**, 110 (1991).
- A. T. Brünger *et al.*, *Science* **235**, 458 (1987).
- R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **26**, 283 (1993).
- D. Frishman and P. Argos, *Proteins Struct. Funct. Genet.* **23**, 566 (1995).
- A. Nicholls *et al.*, *ibid.* **11**, 281 (1991).
- We thank A. Lenhart for several measurements, C. Feil for the clone of mutant D376C, and R. Schwesinger for discussions. Supported by the Deutsche Forschungsgemeinschaft under contracts SFB-323 and SFB-388. The atomic model and the structure factors are deposited in the Protein Data Bank under the accession code 1SQC.

8 April 1997; accepted 13 August 1997

## Structural Basis for Cyclic Terpene Biosynthesis by Tobacco 5-Epi-Aristolochene Synthase

Courtney M. Starks, Kyoungwhan Back, Joseph Chappell, Joseph P. Noel\*

Terpene cyclases catalyze the synthesis of cyclic terpenes with 10-, 15-, and 20-carbon acyclic isoprenoid diphosphates as substrates. Plants have been a source of these natural products by providing a homologous set of terpene synthases. The crystal structures of 5-epi-aristolochene synthase, a sesquiterpene cyclase from tobacco, alone and complexed separately with two farnesyl diphosphate analogs were analyzed. These structures reveal an unexpected enzymatic mechanism for the synthesis of the bicyclic product, 5-epi-aristolochene, and provide a basis for understanding the stereochemical selectivity displayed by other cyclases in the biosynthesis of pharmacologically important cyclic terpenes. As such, these structures provide templates for the engineering of novel terpene cyclases.

Terpene cyclases control the synthesis of cyclic terpenoids including flavors and fragrances such as menthol and camphor, plant defense chemicals like capsidiol and lubimin (1), and more common compounds like steroids and lipid-soluble vitamins. Several cyclic terpenoids have pharmacological activity; for example, limonene can inhibit tumorigenesis induced in mice by particular carcinogens (2), and the diterpenoid taxol has antitumor activity (3). Numerous terpene cyclases from plant and microbial sources have been characterized (4, 5). Although the plant cyclases exhibit a significant degree of similarity in amino acid sequence, very

little similarity is observed between the bacterial, fungal, and plant terpene cyclases (6). These soluble enzymes convert the acyclic isoprenoid diphosphates geranyl diphosphate (GPP, 10 carbon), farnesyl diphosphate (FPP, 15 carbon), and geranylgeranyl diphosphate (GGPP, 20 carbon) into cyclic monoterpenes, sesquiterpenes, and diterpenes, respectively. In most cases, loss of diphosphate (pyrophosphate) from the enzyme-bound acyclic substrate results in an allylic carbocation that electrophilically attacks a double bond further down the terpene chain to effect the first ring closure. Additional rearrangements involving transient carbocations can include proton abstractions, hydride and alkyl migrations, and additional electrophilic attacks.

TEAS (tobacco 5-epi-aristolochene synthase) (7) from *Nicotiana tabacum* converts farnesyl diphosphate (FPP) to 5-epi-aristolochene (Fig. 1) (8), a precursor of the antifungal phytoalexin capsidiol. TEAS shares 77% amino acid identity with *Hyoscyamus muticus* vetispiradiene

C. M. Starks and J. P. Noel, Structural Biology Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA, and Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0301, USA.

K. Back and J. Chappell, Plant Physiology, Biochemistry, and Molecular Biology Program, University of Kentucky, Lexington, KY 40546-0091, USA.

\*To whom correspondence should be addressed. E-mail: noel@sbl.salk.edu

synthase (HVS) (9). Vetispiradiene (Fig. 1) is a precursor to the phytoalexins solavetivone and lubimin. Both enzymes have similar reaction mechanisms (9); in fact, several TEAS-HVS chimeras produce mixtures of the natural reaction products 5-epi-aristolochene and vetispiradiene (10).

In order to better understand terpene cyclase mechanisms, we have determined x-ray crystal structures at 2.2 to 2.8 Å resolution of TEAS alone and in complexes with two different substrate analogs. Recombinant TEAS was expressed in *Escherichia coli*, purified and crystallized (11). We determined the initial structure for a TEAS-HVS chimeric cyclase (CH3) (10) by multiple isomorphous replacement with anomalous scattering (MIRAS) with data to 2.8 Å resolution (Table 1). After den-

synthesis modification, a nearly complete CH3 model was built and refined against the CH3 diffraction data to 2.8 Å. This model was then used for construction of the initial TEAS model with diffraction data for native TEAS crystals extending to 2.25 Å (Table 1). The TEAS model was refined to an R factor of 19.9% with good stereochemistry (12).

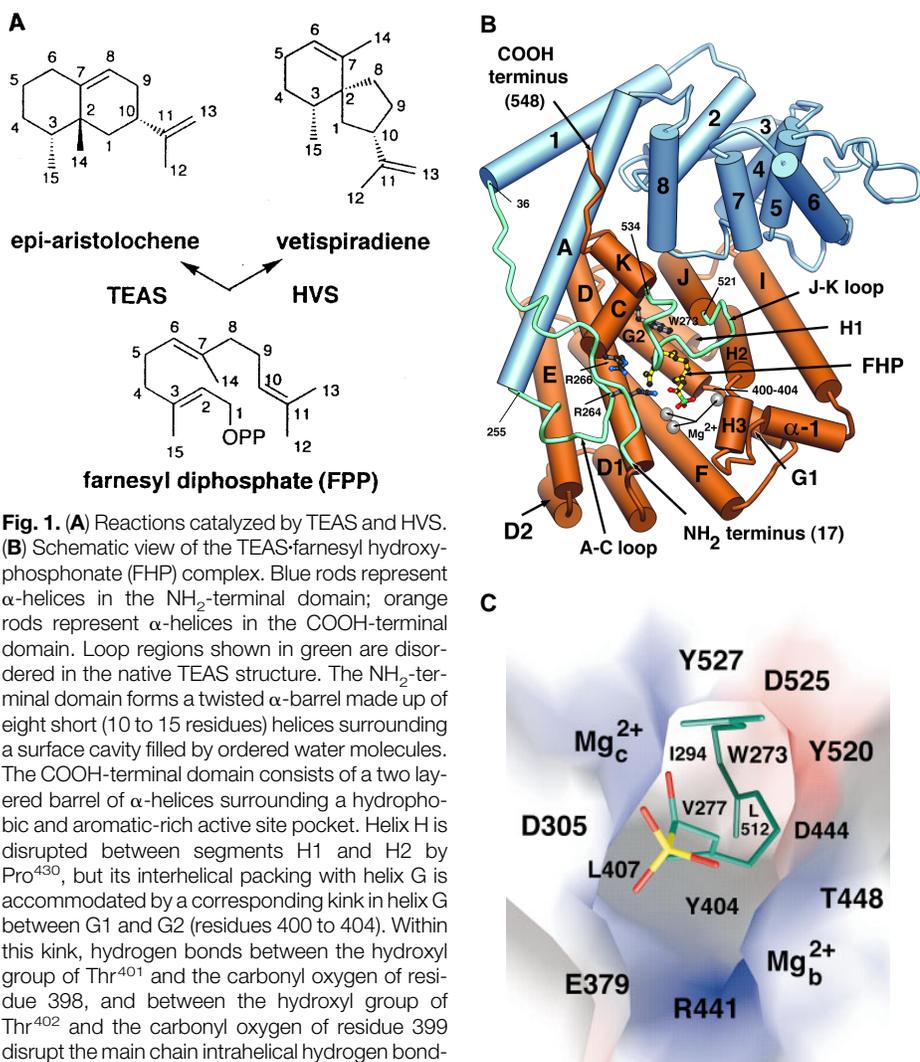
TEAS consists entirely of  $\alpha$ -helices and short connecting loops and turns, and is organized into two structural domains (Fig. 1B). The backbone of the NH<sub>2</sub>-terminal domain (residues 36 to 230) aligns structurally with the catalytic cores of two glycosyl hydrolases: glucoamylase (PDB code 3GLY) from *Aspergillus awamori* (13), and endoglucanase CelD (PDB code 1CLC) from *Clostridium thermocellum*. (14). The function of this domain in

TEAS is unknown. The COOH-terminal domain aligns structurally with avian FPP synthase (FPS), despite their lack of sequence similarity (15).

The enzyme active site was identified in the hydrophobic pocket of the COOH-terminal domain by the location of two Mg<sup>2+</sup> coordination sites at its opening, by the location of bound substrate analogs, and by comparison of the structure with the published FPS structures (15). Two Mg<sup>2+</sup> ions are coordinated on opposite sides of the entrance to the active site pocket in positions analogous to those in FPS, and constitute a diphosphate binding site. Asp<sup>301</sup> coordinates Mg<sub>a</sub><sup>2+</sup> in the native TEAS structure, and the side chain carboxyl of Glu<sup>379</sup> provides a longer range interaction. Asp<sup>305</sup> provides an additional coordination bond in the enzyme with substrate analogs bound. Asp<sup>301</sup> and Asp<sup>305</sup> are part of a -DDXXD- sequence found in terpene cyclases and were previously assumed to constitute a binding site for the required divalent metal (16). Asp<sup>301</sup> directly contacts Mg<sub>a</sub><sup>2+</sup>, whereas Asp<sup>302</sup> demonstrates no direct metal coordination. The side chains of Asp<sup>444</sup>, Thr<sup>448</sup>, Glu<sup>452</sup>, and one water molecule coordinate Mg<sub>b</sub><sup>2+</sup>. In the native TEAS structure the A-C and J-K loops and the residues NH<sub>2</sub>-terminal of residue 36 are disordered (Fig. 1B). Given their location near the entrance to the active site pocket, the flexibility of these regions results in an open active site that displays a number of solvent accessible hydrophobic residues poised to bind the hydrophobic farnesyl chain of FPP.

In order to pinpoint the contributions of particular side chains to substrate and product specificity and to elucidate the stereochemical control exerted by TEAS during farnesyl cyclization, we examined the complexes between TEAS and two FPP analogs: a farnesyl hydroxyphosphonate (FHP), and a trifluoro-farnesyl diphosphate (F<sub>3</sub>-FPP). Complexes with each were obtained both by incubation of TEAS with analog prior to crystallization and by soaking native crystals in solutions containing the analogs.

The native TEAS model, lacking the J-K loop and the residues NH<sub>2</sub>-terminal of residue 36, was refined against data from the TEAS-FHP complex (17). Difference electron density maps calculated with phases derived from this partial model revealed new electron density for FHP, and the previously disordered NH<sub>2</sub>-terminal segment, A-C loop, and J-K loop (Fig. 2, top panel). The ordering of these loops upon FHP binding results in a closed, solvent-inaccessible active site pocket, as expected for binding the hydrophobic sub-



**Fig. 1.** (A) Reactions catalyzed by TEAS and HVS. (B) Schematic view of the TEAS-farnesyl hydroxyphosphonate (FHP) complex. Blue rods represent  $\alpha$ -helices in the NH<sub>2</sub>-terminal domain; orange rods represent  $\alpha$ -helices in the COOH-terminal domain. Loop regions shown in green are disordered in the native TEAS structure. The NH<sub>2</sub>-terminal domain forms a twisted  $\alpha$ -barrel made up of eight short (10 to 15 residues) helices surrounding a surface cavity filled by ordered water molecules. The COOH-terminal domain consists of a two layered barrel of  $\alpha$ -helices surrounding a hydrophobic and aromatic-rich active site pocket. Helix H is disrupted between segments H1 and H2 by Pro<sup>430</sup>, but its interhelical packing with helix G is accommodated by a corresponding kink in helix G between G1 and G2 (residues 400 to 404). Within this kink, hydrogen bonds between the hydroxyl group of Thr<sup>401</sup> and the carbonyl oxygen of residue 398, and between the hydroxyl group of Thr<sup>402</sup> and the carbonyl oxygen of residue 399 disrupt the main chain intrahelical hydrogen bonding of helix G. Naming of helices in the COOH-terminal domain corresponds to the convention used for FPP synthase (FPS) (15). (C) Surface electrostatic representation of the hydrophobic active site pocket of TEAS in the closed, FHP-bound structure. The surface is colored by electrostatic potential, with blue corresponding to positive, red to negative, and white to neutral regions. FHP is shown in stick representation with carbons in green. Panel (B) was generated with RIBBONS (24) and panel (C) was prepared with GRASP (24).

strate and for protecting reactive carbocation intermediates from attack by water (Fig. 1C). As the J-K loop (residues 521 to 534) becomes ordered, it forms a lid that clamps down over the active site entrance in the presence of FHP. This clamp places Tyr<sup>527</sup> next to Trp<sup>273</sup>, forming an extended aromatic box deep within the active site pocket. The A-C loop, which contains Arg<sup>264</sup> and Arg<sup>266</sup>, translates inward toward the active site on FHP binding, positioning the side chain of Arg<sup>264</sup> in close proximity to the C1 hydroxyl group of FHP. Arg<sup>266</sup> hydrogen bonds with residues on both the J-K loop and the NH<sub>2</sub>-terminal segment in the FHP complex. Both Arg<sup>264</sup> and Arg<sup>266</sup> are conserved among many terpene cyclases. On FHP binding, a third Mg<sup>2+</sup> site, designated Mg<sub>c</sub><sup>2+</sup>, forms. The single phosphonate group of FHP binds between Mg<sub>b</sub><sup>2+</sup> and Mg<sub>c</sub><sup>2+</sup>. Much of the FHP farnesyl tail exhibits strong electron density as it extends back into the hydrophobic active site.

In the TEAS-F<sub>3</sub>-FPP complex (18), we observe a well-ordered diphosphate binding pocket; the A-C loop and the NH<sub>2</sub>-terminal segment exhibit well-defined electron density, the A-C loop translates toward the active site, and there is strong electron density for the F<sub>3</sub>-FPP diphosphate moiety. The

hydrophobic pocket, however, remains flexible; the J-K loop and the F<sub>3</sub>-FPP farnesyl chain are disordered (Fig. 2, bottom panel).

The positions of the two analogs in the TEAS active site suggest a binding mode for the natural substrate FPP that is consistent with the known product specificity of TEAS. Both the phosphonate of FHP and the β-phosphate of F<sub>3</sub>-FPP reside in nearly identical positions between Mg<sub>b</sub><sup>2+</sup> and Mg<sub>c</sub><sup>2+</sup>; the C1 hydroxyl of FHP also lies near the position of the α-phosphate of F<sub>3</sub>-FPP, near Mg<sub>c</sub><sup>2+</sup>. Utilizing these structural constraints, we modeled the binding mode for FPP (Fig. 3A). This model of TEAS complexed with its natural substrate defines our proposed catalytic mechanisms for 5-epi-aristolochene biosynthesis in TEAS, and vetispiradiene biosynthesis in the related enzyme HVS, that are consistent with the cyclization pathways proposed on the basis of biochemical data (9).

In our proposed mechanism, as diphosphate is generated (Fig. 3A), its additional negative charge would be offset by interactions with the three Mg<sup>2+</sup> ions, Arg<sup>264</sup>, and Arg<sup>441</sup>. This concentrated region of positive charge may serve to direct the diphosphate away from the hydrophobic pocket where a number of highly reactive carbocations form throughout the reaction. The released

diphosphate, if not constrained, would likely capture the newly formed allylic carbocation at C1 or C3, thereby regenerating FPP (C1 attack) or its tertiary allylic isomer nerolidyl diphosphate (C3 attack). The allylic carbocation of the intermediate is positioned near the main chain carbonyl oxygens of residues 401 and 402 at the kink in helix G. These carbonyls are not involved in hydrogen bonding within the helix; instead, their dipole moments are directed toward the positive charge delocalized over C1, C2, and C3. Movement of the allylic carbocation deeper into the active site would improve this dipolar stabilization by positioning the carbocation closer to the peptide carbonyls of residues 401 and 402 as well as closer to the dipole of Thr<sup>403</sup> hydroxyl group (Fig. 3B). Just as the Mg<sup>2+</sup> and arginines directed the diphosphate away from the active site, these dipoles would direct the cationic end of the farnesyl chain into the hydrophobic active site, poised for attack on C10.

Once C1 has been positioned near the π orbitals of the C10–C11 bond, electrophilic attack at C10 would create a C1–C10 bond and a tertiary carbocation on C11 (Fig. 3C). The quadrupole of Tyr<sup>527</sup> is nicely positioned to stabilize the positive charge on C11. In turn, the newly

**Table 1.** Data collection, structure determination, and refinement. Diffraction data for chimeric and derivative crystals were collected on a MacScience imaging plate detector DIP2020K (MacScience Corp.) with double focusing Pt/Ni coated mirrors and CuKα x-rays. All other data sets were collected at the Stanford Synchrotron Radiation Laboratory, beamline 7-1 (λ = 1.08 Å) on a MAR imaging plate system. Crystals were stabilized by soaking in synthetic mother liquor containing 20% ethylene glycol, and subsequently frozen in a 100 K nitrogen gas stream. Data sets were

processed with DENZO (25) and scaled with SCALEPACK (25). The structure was solved by MIRAS; heavy atom sites were located with difference Patterson and difference Fourier maps. Sites were refined and initial phases were calculated with ML-PHARE (26), giving a figure of merit of 0.39. This initial phase set was improved and the phases extended to 2.8 Å by solvent flattening and histogram matching with the program DM (27), assuming a solvent content of 65%. All models were constructed with program O (28) and refined with XPLOR (29).

Crystal	Resolution (Å)	Reflections measured (N)	Data collection statistics						Sites (N)	Phasing power‡	Anomalous phasing power‡
			Data completeness (%)		R <sub>sym</sub> † (%)		⟨I/σ⟩				
			All	Outer shell	All	Outer shell	All	Outer shell			
CH3	2.8	121,533	97.7	98.5	11.0	46.3	7	4			
K <sub>2</sub> PtCl <sub>4</sub>	3.0	60,109	56.2	60.5	8.0	41.2	9	2	5	0.88	0.25
CH <sub>3</sub> HgCl	3.5	43,166	89.0	88.0	10.3	41.6	7	2	3	1.10	0.27
Lu (acetate) <sub>3</sub>	3.5	18,493	58.3	59.0	15.4	43.9	5	2	1	0.71	
TEAS	2.25	272,509	95.2	96.7	6.3	27.5	11	6			
FHP	2.8	136,072	98.3	96.3	18.7	53.1	5	2			
F <sub>3</sub> -FPP	2.15	268,076	91.5	70.3	7.6	46.9	8	3			

Structure	Resolution range (Å)	R factor§ (all data, %)	R <sub>free</sub> § (all data, %)	Bond lengths (rmsd, °)	Bond angles (rmsd, °)
TEAS	20–2.25	19.9	24.2	0.005	1.21
FHP	100–2.8	23.4	28.1	0.011	1.29
F <sub>3</sub> -FPP	40–2.25	25.0	29.4	0.019	1.42

\*Figure of merit = ∫ P(φ) exp(iφ)dφ / ∫ P(φ)dφ where P is the probability distribution of φ, the phase angle. †R<sub>sym</sub> = Σ<sub>h</sub> |I<sub>h</sub> - ⟨I<sub>h</sub>⟩| / Σ<sub>h</sub> I<sub>h</sub> where ⟨I<sub>h</sub>⟩ is the average intensity of reflection h for its symmetry and Friedel equivalents. ‡R<sub>phasing</sub> = (Σ |F<sub>PH</sub> ± F<sub>P</sub> - F<sub>H(calc)</sub>|) / (Σ |F<sub>PH</sub> ± F<sub>P</sub>|); phasing power = Σ |F<sub>H(calc)</sub>| / Σ ||F<sub>P</sub>| exp(iφ) + F<sub>H(calc)</sub> - |F<sub>PH</sub>||; anomalous phasing power = Σ |F<sub>H</sub>(0)| / Σ ||F<sub>PH</sub>(+) - F<sub>PH</sub>(-) - 2F<sub>H(calc)</sub> sinφ<sup>c</sup>| where F<sub>H(calc)</sub> is the calculated heavy atom structure factor, |F<sub>P</sub>| and |F<sub>PH</sub>| are the observed amplitudes for the protein and heavy atom derivatives respectively, φ<sup>c</sup> is the experimental phase, F<sub>H</sub>(0) is the imaginary component of the calculated heavy atom structure factor, and F<sub>PH</sub>(+) and F<sub>PH</sub>(-) are observed amplitudes for Bijvoet pairs in heavy atom derivatives. §R factor = (Σ |F<sub>obs</sub> - F<sub>calc</sub>|) / Σ (F<sub>obs</sub>) where F<sub>obs</sub> and F<sub>calc</sub> are the observed and calculated structure factors, respectively. R<sub>free</sub> is calculated in an analogous manner for 5% of the data that has never been used for refinement. Both values were calculated with no sigma cutoff.

formed carbocation at C11 will substantially increase the acidity of the C12 and C13 methyl protons. As modeled, the carboxyl group of Asp<sup>525</sup> would abstract a proton from C13 (cis methyl group) leading to the formation of a neutral germacrene intermediate (Fig. 3D).

Subsequently, the reaction requires proton addition at C6. In the closed conformation observed in the TEAS-FHP complex, an Asp<sup>444</sup>-Tyr<sup>520</sup>-Asp<sup>525</sup> triad organizes on the interior surface of the active site as the J/K loop, containing Tyr<sup>520</sup> and Asp<sup>525</sup>, clamps over the active site. A carboxyl oxygen from each Asp side chain is within hydrogen bonding distance of Tyr<sup>520</sup>'s hydroxyl group (Fig. 2, top panel). In the proposed TEAS-germacrene complex, the Tyr<sup>520</sup> hydroxyl group also resides near C6 of germacrene (Fig. 3D). In our proposed reaction scheme, the carboxyl group of Asp<sup>444</sup> would remove the hydroxyl proton from Tyr<sup>520</sup>. At the same time, the phenolic oxygen at Tyr<sup>520</sup> would accept a proton from Asp<sup>525</sup>. Tyr<sup>520</sup> then could donate this newly positioned proton to the double bond of germacrene at C6 (Fig. 3E). In a concerted fashion, Tyr<sup>520</sup> would re-accept the proton currently on the Asp<sup>444</sup> carboxyl group. Concomitant with the C6 protonation by Tyr<sup>520</sup>, germacrene would undergo a second

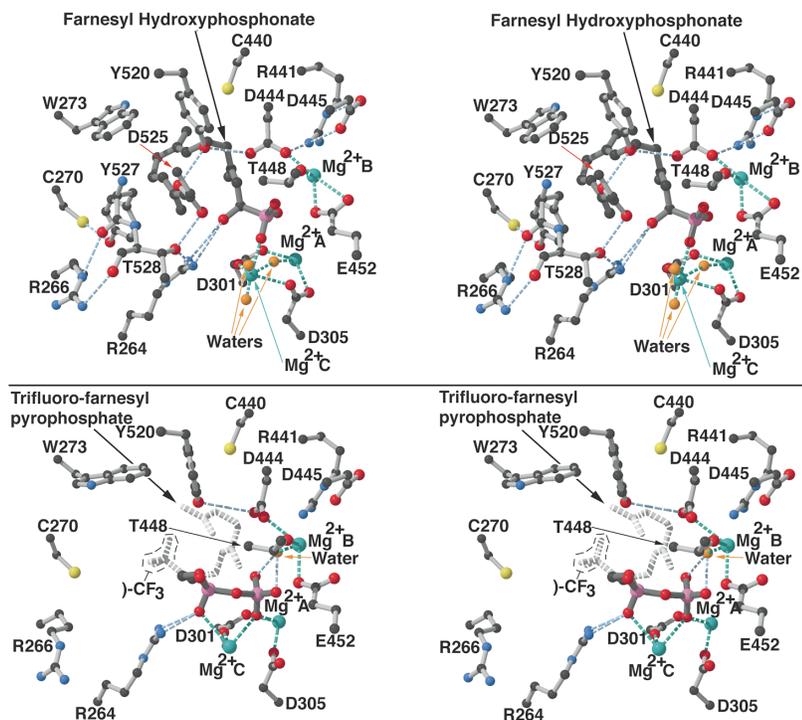
ring closure as the  $\pi$  orbitals on C2 and C7 line up for the formation of the C2–C7 sigma bond. The resulting bicyclic eudesmane carbocation intermediate (Fig. 3F) with the positive charge on C3 would be stabilized by the summed dipoles of the main chain carbonyls of residues 401 and 402 and the Thr<sup>403</sup> hydroxyl group.

In order to ensure either 5-epi-aristolochene (TEAS) or vetispiradiene (HVS) synthesis, both enzymes subsequently stabilize a carbocation centered on C7. In both cases, the C2 hydride first migrates to the planar C3 carbocation along the top of the bicyclic ring system leaving a tertiary carbocation with planar geometry at C2 (Fig. 3F). C7 abuts the aromatic face of Trp<sup>273</sup>, and stabilization afforded by this aromatic quadrupole would selectively position the electrophilic center at C7. Shifting the carbocation to C7 from C2 occurs via two different routes that distinguish the TEAS catalyzed synthesis of 5-epi-aristolochene (Fig. 3, G<sub>1</sub> to I<sub>1</sub>) and the HVS catalyzed synthesis of vetispiradiene (Fig. 3, G<sub>2</sub> to I<sub>2</sub>). In TEAS, the eudesmane carbocation intermediate with the positive charge centered at C2 would exist in a conformation that orients the C7–C14 sigma bond parallel to the empty *p* orbital on C2. This conformer would fa-

litate the migration of the C14 methyl group from C7 to C2 (Fig. 3G<sub>1</sub>). The conformation of the resulting carbocation with a positive charge positioned on C7 directs a proton on C8 toward the indole ring of Trp<sup>273</sup>. The presence of the carbocation center on C7 greatly increases the acidity of the proton at C8 which may now be removed by Trp<sup>273</sup> (Fig. 3H<sub>1</sub>), giving rise to a positive arenium ion (TrpH<sup>+</sup> at residue 273) and the final reaction product, 5-epi-aristolochene (Fig. 3I<sub>1</sub>).

In vetispiradiene synthesis by HVS, an energetically allowed 60° rotation around the C2–C7 bond of the eudesmane carbocation intermediate with the positive charge centered on C2 would result in a conformation in which the C7–C8 bond is now parallel to the empty *p* orbital of C2 (Fig. 3G<sub>2</sub>). In this conformation, the methylene group C8 migrates from C7 to C2, resulting in the collapse of the original six-membered ring to a five-membered ring. The resulting carbocation intermediate directs a proton of C6 toward the indole ring of Trp<sup>273</sup>. As described for TEAS, the greatly enhanced acidity of this proton could allow facile deprotonation by Trp<sup>273</sup> (Fig. 3H<sub>2</sub>), giving rise to a positive arenium ion (TrpH<sup>+</sup> residue 273), and the reaction product vetispiradiene (Fig. 3I<sub>2</sub>). In both enzymes, the active site would again become solvent-accessible following product dissociation as the A-C and J-K loops and residues at the NH<sub>2</sub>-terminus become mobile. The TrpH<sup>+</sup> cation would then lose a proton to the surrounding aqueous environment.

Trp<sup>273</sup> is postulated to act as a proton acceptor during the final deprotonation step for two reasons. First, the size and positioning of the indole ring allows it to accept an acidic proton from either C8 (Fig. 3H<sub>1</sub>) or C6 (Fig. 3H<sub>2</sub>) as would be required in the chimeric TEAS or HVS enzymes that catalyze the synthesis of both 5-epi-aristolochene and vetispiradiene (10). No other proton acceptors are positioned to abstract a proton from C6 or C8 without a substantial reorientation of one or more of the cyclic intermediates. Mutation of Trp<sup>273</sup> to either nonaromatic or aromatic residues results in cessation of epi-aristolochene or vetispiradiene synthesis (19). Second, the greatly reduced nucleophilicity of the indole ring, compared to general bases such as histidine, aspartate, or glutamate, greatly diminishes the likelihood of inadvertent enzyme alkylation by the highly reactive carbocation intermediate. While the enzymatic roles proposed here for Trp<sup>273</sup> are novel, there is chemical precedence for both quadrupole mediated cation interactions (20)



**Fig. 2.** Stereo views of the TEAS active site. Blue dashed lines are hydrogen bonds; green dashed lines are coordination bonds. The top panel illustrates the TEAS-FHP complex. FHP is emphasized with dark gray bonds. The lower panel illustrates the TEAS-F<sub>3</sub>-FPP complex. The diphosphate moiety is emphasized with dark bonds. The rest of the F<sub>3</sub>-FPP molecule, which exhibits weak electron density, is indicated with dashed bonds. Figure was prepared with MOLSCRIPT (24) and RASTER3D (24).

and for electrophilic aromatic substitutions in indole ring systems (21). In the solvent shielded active site of TEAS or HVS, the arenium cation formed upon protonation of the indole ring of Trp<sup>273</sup> would be more energetically stable than the eudesmane carbocations, due to the presence of the nitrogen atom and the extended conjugation of the indole ring.

Modeling the structure of HVS based on our crystallographic model of TEAS suggests that the residues which directly line the active site are conserved between the two enzymes; product selectivity,

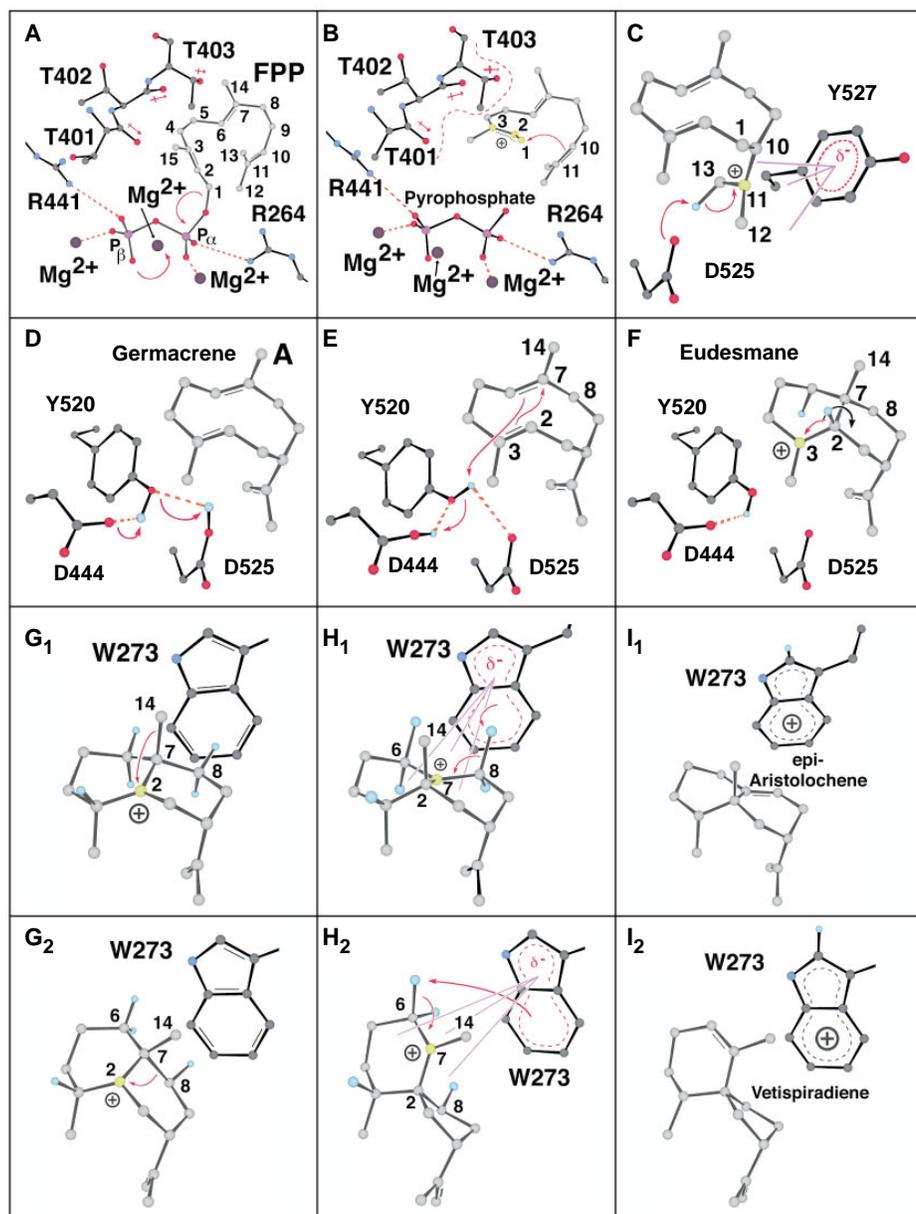
therefore, likely derives from the amino acid differences in layers surrounding the active site. This pattern of conservation suggests that the catalytic residues required to form the two products are the same, and that specificity depends on having a particular active site conformation determined by the surrounding layers. This model is consistent with the observation that TEAS and HVS chimeric enzymes can produce mixtures of 5-epi-aristolochene and vetispiradiene. These multiproduct chimeric enzymes, in which the active site topology is influenced by a

combination of interactions from the two wild-type enzymes, still catalyze terpene cyclization, but have lost the ability to select a single conformation of the eudesmane intermediate that would give rise to a single product.

Sequence alignments between TEAS and other plant cyclases viewed in the context of the TEAS structure suggest that all plant cyclases share similar three-dimensional structures. For example, the large sequence insertions in taxadiene synthase (TS) (22) with respect to TEAS occur at the NH<sub>2</sub>-terminus, or within solvent exposed loops in the NH<sub>2</sub>-terminal domain. TS converts geranylgeranyl diphosphate to the tricyclic taxol precursor, taxadiene. Modeling the TS structure on the basis of the three-dimensional fold of TEAS suggests that the TS active site is more spacious than that of TEAS, consistent with the larger substrate of TS. The TS active site is also rich in aromatic residues, including Phe<sup>602</sup> (Leu<sup>290</sup> in TEAS), Trp<sup>753</sup> (Cys<sup>440</sup> in TEAS), Phe<sup>835</sup> (Thr<sup>519</sup> in TEAS), and Tyr<sup>841</sup> (Tyr<sup>527</sup> in TEAS). Given the likely involvement of aromatic quadrupoles in carbocation stabilization, this extensive aromatic surface may reflect the greater spatial distribution of carbocations on various intermediates in the proposed TS reaction mechanism (23). In closing, the high resolution structures and comparative models that we describe provide a structural foundation for efforts to create novel terpene cyclases.

## REFERENCES AND NOTES

1. M. E. M. Guedes, J. Kuc, R. Hammerschmidt, R. Bostock, *Phytochemistry* **21**, 2987 (1982).
2. M. A. Morse and A. L. Toburen, *Cancer Lett.* **104**, 211 (1996).
3. S. Borman, *Chem. Eng. News* (1 July 1996), pp. 27–29.
4. D. E. Cane, *Chem. Rev.* **90**, 1089 (1990).
5. R. Croteau, *ibid.* **87**, 929 (1987).
6. J. Chappell, *Plant Physiol.* **107**, 1 (1995).
7. Compounds are referred to here by their standard names, but for simplicity in discussions of mechanisms, atoms are numbered according to their position in the acyclic isoprenoid diphosphate precursor.
8. K. Back, S. Yin, J. Chappell, *Arch. Biochem. Biophys.* **315**, 527 (1994).
9. K. Back and J. Chappell, *J. Biol. Chem.* **270**, 7375 (1995).
10. K. Back and J. Chappell, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6841 (1996).
11. TEAS and CH3 were expressed in *Escherichia coli* with NH<sub>2</sub>- or COOH-terminal His<sub>6</sub>-tags, purified by metal chelation, anion exchange, and gel filtration chromatography, and crystallized in hanging drops with 15% PEG 8000, 200 mM magnesium acetate or magnesium chloride, 100 mM 3-(*N*-morpholino)-2-hydroxypropane sulfonic acid (pH 7), 1 mM dithiothreitol as a precipitant. The crystals belong to the tetragonal space group *P*4<sub>1</sub>2<sub>1</sub>2; the unit cell dimensions vary by a few angstroms between crystals, but on average *a* = 126 Å, *c* = 122 Å.
12. The CH3 model was initially refined with the TEAS diffraction data set with AMORE [J. Navaza, *Acta Crystallogr.* **A50**, 157 (1994)]. Subsequent rounds of positional and restrained B-factor refinement with bulk solvent



**Fig. 3.** Proposed catalytic mechanisms of TEAS and HVS based on the TEAS crystal structures. (A) through (F) are reaction steps common to both TEAS and HVS. (G<sub>1</sub>) through (I<sub>1</sub>) are specific to TEAS, and (G<sub>2</sub>) through (I<sub>2</sub>) are specific to HVS. Carbocations are yellow, protons are light blue, and aromatic quadrupoles are indicated with purple lines. Intermediate structures are numbered according to the acyclic precursor, FPP (A). Figure was generated with Chem3D (CambridgeSoft).

# Crystal Structure of Pentalenene Synthase: Mechanistic Insights on Terpenoid Cyclization Reactions in Biology

Charles A. Lesburg, Guangzhi Zhai, David E. Cane, David W. Christianson\*

The crystal structure of pentalenene synthase at 2.6 angstrom resolution reveals critical active site features responsible for the cyclization of farnesyl diphosphate into the tricyclic hydrocarbon pentalenene. Metal-triggered substrate ionization initiates catalysis, and the  $\alpha$ -barrel active site serves as a template to channel and stabilize the conformations of reactive carbocation intermediates through a complex cyclization cascade. The core active site structure of the enzyme may be preserved among the greater family of terpenoid synthases, possibly implying divergence from a common ancestral synthase to satisfy biological requirements for increasingly diverse natural products.

Sesquiterpenes comprise a group of natural products secreted by marine and terrestrial plants, fungi, and certain microorganisms. The structural diversity and stereochemical complexity of the  $C_{15}$ -isoprenoid skeletons of these metabolites are remarkable. Indeed, of the more than 300 cyclic sesquiterpenes that have been characterized to date, each is derived from a common acyclic precursor, farnesyl diphosphate (1), in a reaction catalyzed by a sesquiterpene cyclase (2). Many cyclic sesquiterpenes exhibit useful medicinal properties and have been essential components of the pharmacopoeia since times of antiquity. For example, the sesquiterpenes furanoeudesma-1,3-diene and curzarene are responsible for the analgesic effects of myrrh by interacting with brain opioid receptors, thereby explaining the prescription of myrrh

for use as a pain killer by Pliny the Elder, Hippocrates, and their predecessors (3).

In addition to the 15-carbon sesquiterpenes, the greater family of terpenoids includes the 10-carbon monoterpenes derived from geranyl diphosphate, the 20-carbon diterpenes derived from geranylgeranyl diphosphate, the 25-carbon sesterterpenes derived from geranylarnesyl diphosphate, and the 30-carbon triterpenes (sterols) derived from two farnesyl diphosphate molecules. Terpenoid cyclases (also known as terpenoid synthases) perform critical biosynthetic tasks in metabolic pathways as diverse as cholesterol biosynthesis in mam-



**Fig. 1.** Ribbon plot (25) of pentalenene synthase. The mouth of the active site cavity opens toward the top of the figure, and the aspartate-rich segment beginning with Asp<sup>80</sup> is red. The dotted line is the disordered Phe<sup>158</sup>-Asp<sup>164</sup> loop, and the Cys<sup>128</sup>-Cys<sup>136</sup> disulfide linkage is yellow.



**Fig. 2.** Ribbon plot (25) of farnesyl diphosphate synthase (8) (PDB accession code 1FPS). The core terpenoid synthase structure shared with pentalenene synthase is blue; the two aspartate-rich segments [beginning with Asp<sup>257</sup> (left) and Asp<sup>117</sup> (right)] are red. The Asp<sup>117</sup> segment of farnesyl diphosphate synthase aligns with the Asp<sup>80</sup> segment of pentalenene synthase (Fig. 1). Both substrates of this enzyme, isopentenyl diphosphate and dimethylallyl diphosphate, are believed to bind to the two aspartate-rich segments through bridging magnesium ions.

- modeling were carried out against all data between 20.0 and 2.25 Å. Simulated-annealing omit maps were used to confirm model building. Water molecules were added with ARP [V. X. Lamzin and K. S. Wilson, *Acta Crystallogr.* **D49**, 127 (1993)], and X-PLOR (29). Our model consists of residues 36 to 521 and 533 to 548, two Mg<sup>2+</sup> ions, and 270 water molecules; 95% of the residues are in the most favored regions of the Ramachandran plot, and none are in disallowed regions.
13. A. E. Aleshin, C. Hoffman, L. M. Firsov, R. B. Honzatzko, *J. Mol. Biol.* **238**, 575 (1994).
  14. M. Juy *et al.*, *Nature* **357**, 89 (1992).
  15. L. C. Tarshis, M. Yan, C. D. Poulter, J. C. Sacchettini, *Biochemistry* **33**, 10871 (1994); L. C. Tarshis, P. J. Proteau, B. A. Kellogg, J. C. Sacchettini, C. D. Poulter, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 15018 (1996).
  16. D. E. Cane, Q. Xue, B. C. Fitzsimons, *Biochemistry* **35**, 12369 (1996).
  17. The uncomplexed TEAS structure was initially refined to 2.8 Å against data collected from a crystal grown in the presence of 2 mM FHP. Electron density at the active site allowed unambiguous modeling of FHP, the A-C and J-K loops, and nine additional residues at the NH<sub>2</sub>-terminus. The refined TEAS·FHP model consists of residues 17 to 548, three Mg<sup>2+</sup> ions, 150 water molecules, and one FHP molecule.
  18. The uncomplexed TEAS structure was refined against data collected from a crystal grown in the presence of 1 mM F<sub>3</sub>-FPP. Strong electron density initially appeared for the diphosphate moiety of F<sub>3</sub>-FPP; as refinement proceeded, density also appeared for additional residues at the NH<sub>2</sub>-terminus and in the J-K loop. The refined TEAS·F<sub>3</sub>-FPP model consists of residues 21 to 523 and 529 to 548, 150 water molecules, three Mg<sup>2+</sup> ions, and one F<sub>3</sub>-FPP molecule.
  19. K. Back, C. M. Starks, J. P. Noel, J. Chappell, in preparation.
  20. D. A. Dougherty, *Science* **271**, 163 (1996).
  21. K. M. Biswas and A. H. Jackson, *J. Chem. Soc. Perkin Trans. I* **11**, 1981 (1989) and M. I. Abdullah, A. H. Jackson, P. P. Lynch, K. A. F. Record, *Heterocycles* **30**, 317 (1990).
  22. M. R. Wildung and R. Croteau, *J. Biol. Chem.* **271**, 9201 (1996).
  23. X. Lin, M. Hezari, A. E. Koepp, H. G. Floss, R. Croteau, *Biochemistry* **35**, 2968 (1996).
  24. RIBBONS, M. Carson and C. E. Bugg, *J. Mol. Graphics* **4**, 121 (1986); MOLSCRIPT, P. Kraulis, *J. App. Crystallogr.* **24**, 946 (1991); RASTER3D, D. Bacon and W. F. Anderson, *J. Mol. Graphics* **6**, 219 (1988); GRASP, A. Nicholls, K. A. Sharp, B. Honig, *Proteins Struct. Funct. Genet.* **11**, 281 (1991).
  25. Z. Otwinowski, in *Data Collection and Processing*, L. Sawyer, N. Isaacs, S. Bailey, Eds. (CCP4 Study Weekend, SERC Daresbury Laboratory, Warrington, UK, 1993), pp. 56–62.
  26. ———, *ML-PHARE* (CCP4, SERC Daresbury Laboratory, Warrington, UK, 1991).
  27. K. Cowtan, *DM* (Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography) **31**, 34 (1994).
  28. T. A. Jones, J. Y. Zou, S. W. Cowan, M. Kjeldgaard, *Acta Crystallogr.* **A47**, 110 (1991).
  29. A. T. Brünger, *X-PLOR Version 3.1—A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, 1992), pp. 187–207.
  30. We thank D. W. Christianson for communicating results prior to publication; C. D. Poulter for providing the FHP and F<sub>3</sub>-FPP; J. Kyte for discussion; M. Bowman, M. Harrington, and D. Wild for technical assistance; A. Bilwes, M. Bowman, R. Ranganathan, and S. Redford for assistance during data collection at beamline 7-1 at the Stanford Synchrotron Radiation Laboratory (SSRL); the staff at SSRL for assistance and advice. Supported in part by the Lucille P. Markey Charitable trust and NIH grant GM54029 (J.P.N.); NSF grant IBW-9408152 (J.C.); the Chapman Foundation and NIH training grant GM07240, administered through the University of California, San Diego (C.M.S.). Work performed at SSRL was supported by the NIH and DOE. Coordinates for uncomplexed TEAS have been deposited in the Brookhaven Protein Data Bank, code 5EAS.

11 April 1997; accepted 26 August 1997