Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 1172–1176, February 1995
Biochemistry

A common fold for peptide synthetases cleaving ATP to ADP: Glutathione synthetase and D-alanine:D-alanine ligase of *Escherichia coli*

(protein homology/enzyme tertiary structure)

CHANG FAN*, PAUL C. MOEWS*, YIAN SHI†, CHRISTOPHER T. WALSH†, AND JAMES R. KNOX*‡

*Department of Molecular and Cell Biology, The University of Connecticut, Storrs, CT 06269-3125; and †Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115

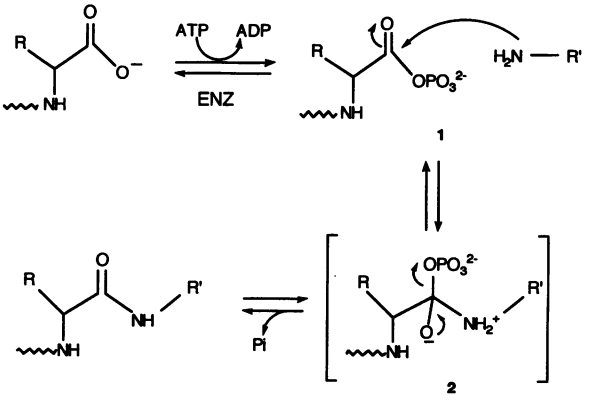
Contributed by Christopher T. Walsh, November 7, 1994

ABSTRACT **Examination of x-ray crystallographic structures shows the tertiary structure of D-alanine:D-alanine ligase (EC 6.3.2.4), a bacterial cell wall synthesizing enzyme, is similar to that of glutathione synthetase (EC 6.3.2.3) despite low sequence homology. Both *Escherichia coli* enzymes, which convert ATP to ADP during ligation to produce peptide products, are made of three domains, each folded around a 4-to 6-stranded β-sheet core. Sandwiched between the β-sheets of the C-terminal and central domains of each enzyme is a nonclassical ATP-binding site that contains a common set of spatially equivalent amino acids. In each enzyme, two loops are proposed to exhibit a required flexibility that allows entry of ATP and substrates, provides protection of the acylphosphate intermediate and tetrahedral adduct from hydrolysis during catalysis, and then permits release of products.**

Enzymes that activate acyl groups typically couple ATP hydrolysis to provide the thermodynamic driving force and exhibit two major cleavage patterns, (*i*) ATP to AMP and PP_i or (*ii*) ATP to ADP and P_i. In the case of peptide bond synthesis, both in ribosomes and in nonribosomal multienzyme complexes, the nucleotidyl transfer route *i* is used via aminoacyl-AMP intermediates. A few nonribosomal peptide synthetases use route *ii* via phosphoryl transfer to the carboxylate to be activated, yielding aminoacyl phosphate intermediates. These include such enzymes as glutamine synthetase (EC 6.3.1.2) (1), γ-Glu-Cys synthetase (EC 6.3.2.2) (2), glutathione synthetase (EC 6.3.2.3; GSHase) (3), D-Ala:D-Ala ligase (EC 6.3.2.4; DD-ligase) (4), and the MurC, -D, and -E bacterial peptidoglycan synthetic enzymes that sequentially convert UDP-muramic acid to UDP-muramyl-tripeptide via addition of L-alanine, γ-D-glutamic acid, and *meso*-diaminopimelate (5).

While it has been clear for some time that the ADP-producing peptide synthetases use the generic acylphosphate **1** → tetrahedral adduct **2** → peptide route of Scheme **I**, only recently have x-ray structures of various members of this class become available. The *Escherichia coli* glutamine synthetase, where $H_2NR' = NH_3$, is a homododecamer of 50-kDa subunits that is highly regulated by intersubunit interactions (1). For a long time, it was the sole representative whose structure was known (6). More recently the structure of GSHase from *E. coli*, a homotetramer of 36-kDa subunits, was solved to 2.0-Å resolution (7). Crystal soaking with ATP and the analog γ-Glu-L-aminobutyrate revealed approximate locations of nucleotide and dipeptide binding. Details of the binding interactions were not published, however, because two loops near the binding sites were disordered.

We recently reported the x-ray structure of DD-ligase of the *ddlB* gene of *E. coli* complexed with ADP and a phosphorylated phosphinate analog of tetrahedral adduct **2** (8). A dramatic similarity was noted in the three-dimensional structures of the DD-ligase and GSHase, in spite of only marginal sequence homology. The crystallographic structure of the DD-ligase complex may, therefore, provide significant structural details and mechanistic insights not available from the disordered structure of the homologous GSHase complex. Thus, the two crystal structures suggest the enzymes possess a signature fold and catalytic site that may prove characteristic of a family of ADP-forming peptide synthetases.



Scheme I

METHODS

Primary Sequence Comparison. For the protein data base search of primary amino acid sequences, the BLASTP program (9) was used. Sequence alignments were done with the CLUSTAL program (10) using a Dayhoff weighting scheme with PAM250 (11).

Tertiary Structures. The crystallographic structure of the ternary complex of *E. coli* DD-ligase with ADP and a phosphorylphosphinate, refined at 2.3-Å resolution to an *R* factor of 0.171 (8), is deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, entry 2DLN. The enzyme has 306 residues, all of which are visible in the electron density map. DD-Ligase is dimeric across the diad of space group $P2_12_12$.

The crystallographic structure of *E. coli* GSHase, refined at 2.0-Å resolution to an *R* factor of 0.186 (7), was taken from the Protein Data Bank, entry 1GLT. The enzyme has 315 amino acid residues but atomic coordinates of only 296 residues are available because two regions, Gly-164 to Gly-167 and Ile-226 to Arg-241, are invisible in the electron density map. Coordi-

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

Abbreviations: GSHase, glutathione synthetase; DD-ligase, D-Ala:D-Ala ligase.

‡To whom reprint requests should be addressed.

nates of ATP, ADP, or substrate analogs are not deposited. GSHase is a tetramer having exact 222 point group symmetry in space group $P6_222$.

Tertiary Structure Comparison. Three-dimensional searches were undertaken using the program DEJAVU (12), which compares the secondary structure array of DD-ligase to all structure arrays in a data base of 1600 crystallographic entries from the Protein Data Bank. The program finds three-dimensional motifs (i.e., collections of α-helices and β-strands) that have approximately the same number of residues per element, are separated by similar center-to-center distances, and have comparable directions and connectivity. Foldings were also compared by a second method (13) that overlays short successive segments of α-carbon atoms of one molecule upon those of the second by using a general fitting and minimization procedure.

RESULTS

Use of the primary amino acid sequence of the *E. coli* DD-ligase (14) in a search of 130,340 sequences with the BLASTP program (9) failed to detect a match with the GSHase sequence. Primary sequence alignment of the two enzymes by the automated CLUSTAL method (10) produced a Dayhoff-weighted homology of only 12%. Though this calculated homology between DD-ligase and GSHase is very low, the folding of the two ligases is found to be strongly similar (Fig. 1). In view of this discrepancy, the CLUSTAL alignment was more closely examined and found to have very little relation to an alignment, discussed below, based on overlay of the crystallographic structures. Simple counts of exact pairs and similar pairs in the two alignments are compared in Table 1 with the weighted value. It is seen that the CLUSTAL method produces as many, or more, matching pairs as observed in the crystallographic alignment, but the matches themselves are often different.

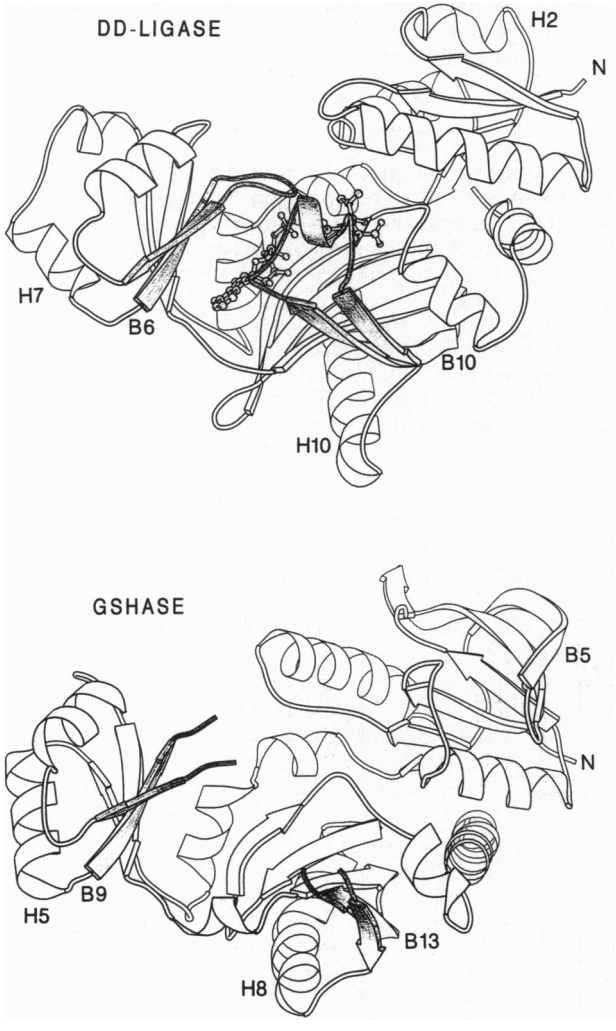


FIG. 1. Comparison of DD-ligase and GSHase, both from *E. coli*. ATP- and substrate-binding loops and their β-strands are shaded. GSHase is unknown from residues 164 to 167 and from residues 226 to 241 (7). Diagrams were drawn by Molscript (15).

In the three-dimensional search method with DEJAVU (12), several DD-ligase motifs of secondary structure elements, each containing a given number of residues, were compared with secondary structure motifs in a large crystallographic data base. For these searches DD-ligase was assigned its 28 secondary structural elements: 12 α-helices and 16 β-strands. Two principal parameters, the minimum number of residues in each element and the minimum number of elements to match, were varied as shown in Table 2. After the number of match elements was raised to 11, the search process identified GSHase as the sole match from the crystallographic data base and produced a rms separation between element centers of 3.00 Å. With a smaller motif of only 6 elements having 6 residues per element, the separation between elements is reduced to 2.25 Å. The glutamine synthetase structure (6) was not matched by DEJAVU, and it does not, in fact, resemble DD-ligase.

Another fitting procedure for three-dimensional structures (13) gave a correlation map (Fig. 2) that indicates a significant overlay of three domains centered approximately at residues 80, 160, and 260 (in DD-ligase numbering). Secondary structure in the three common domains is shown schematically in Fig. 3. Each domain forms around a 4- to 6-stranded β-sheet. Both N-terminal domains contain 4 parallel β-strands, though the domain of GSHase contains two additional antiparallel β-strands in place of helix H2 in DD-ligase. The antiparallel-stranded central and C-terminal domains of each enzyme have equivalent secondary and tertiary structure (Figs. 3 and 4). In the case of DD-ligase, these latter two domains are known from crystallography to form the ATP-binding site. The site of ATP binding in GSHase is also in this region, but the ATP position is not well-defined, possibly because reactants were diffused into pregrown native crystals (7) rather than cocrystallized with enzyme after reaction in solution (8).

Examination of the superposed tertiary structures shows that GSHase contains many amino acid side chains that are spatially equivalent to those surrounding ATP (or ADP) in DD-ligase. For example, phosphate-binding Lys-97 and Lys-144 of DD-ligase closely overlay Lys-125 and Lys-160 of GSHase (Table 3). Also in the ATP site, Glu-187 and probably the corresponding Asp-208 hydrogen bond to the ribose hydroxyl groups in DD-ligase and GSHase, respectively.

Within the β-sheet of the C-terminal domain of each molecule is a rare perturbation of the secondary structure, a so-called β-bulge (16). The β-bulge results from the insertion of Leu-269 into β-strand B13 of DD-ligase and Thr-280 into β-strand B16 of GSHase. In DD-ligase, and possibly in GSHase,

Table 1. Homologies between DD-ligase and GSHase

Alignment method	Weighted	Matching pairs, no.	
		Exact	Similar
CLUSTAL*	12	46 (15)	77 (25)
Crystallographic†	NA	33 (11)	76 (25)

Data in parentheses are the percentage of the total number of residues in DD-ligase (306 residues). Similar pairs are Ile = Leu = Val = Phe = Ala, Asp = Glu, and Lys = Arg. NA, not applicable.
*Ref. 10.
†From overlay of tertiary structures.

Table 2. Crystallographic data base search (12) using DD-ligase

Total no. of elements	Min. no. elements included	Min. no. residues per element	Separation (rms) of element centers, Å	No. matches other than GSHase
28 (12α, 16β)	11	3	3.00	0
	9	3	2.43	1
	8	3	2.17	6
24 (12α, 12β)	10	4	3.03	0
	8	4	2.39	0
	7	4	2.16	8
19 (10α, 9β)	7	5	2.44	0
	6	5	2.22	57
15 (10α, 5β)	6	6	2.25	0
	5	6	1.92	37

this alteration of the usual β-strand conformation places the side chain of the residue near the adenine and ribose rings of ATP.

In GSHase two loops made of residues 164–167 and 226–241 are thought to help form the catalytic site (17, 18) but were too disordered to be described crystallographically. We believe these loops correspond in DD-ligase to the loop of residues 148–151 and the longer loop of residues 208–224 (Fig. 5), both of which in the phosphinate transition-state complex are seen to cover the sites of ATP and inhibitor binding (8). This assumption of equivalence of the loops is supported by proteolysis experiments with both enzymes. Chymotrypsin cleaves the DD-ligase at Tyr-210–Asp-211 in the large loop, but only in the native enzyme (19). Complexation with ATP and a phosphinate prevented cleavage, presumably because the 208–224 loop, with the 148–151 loop, becomes less exposed as it folds inward to cover the reactants. In fact, in the crystal complex of DD-ligase, it appears that the 210–211 peptide linkage would be completely inaccessible to a proteinase. In similar fashion, the large 226–241 loop in GSHase is cleaved by arginyl endopeptidase at Arg-233–Gly-234, but the loop is protected from cleavage when the enzyme is complexed with ATP and/or the γ-Glu-Cys substrate (17).

Alignment of the cleavable loops in the two enzymes is not clear because of the structural disorder in GSHase and poor sequence homology in this region.

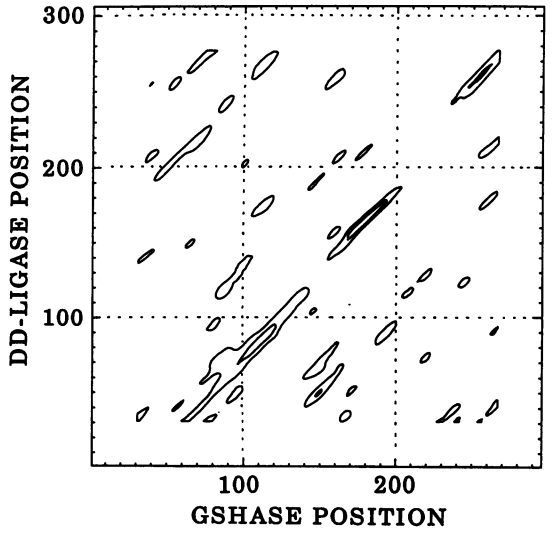


FIG. 2. Comparison of the structures in Fig. 1 by minimizing the fit of an α-carbon segment of length L onto consecutive segments of the second structure (13). A perfect match would be indicated by a continuous contour along the diagonal. With $L = 60$, the mean Cα–Cα separation is 12.8 Å with a standard deviation of 2.0 Å. Contours begin at 1.5 standard deviations below the mean, with an interval of 1.5 deviations.

```
-TFYDYEAK(215)YLSDETQY-  DD-ligase
-IPQGGETR(233)GNLAAGGR-  GSHase
```

In DD-ligase, Lys-215 interacts electrostatically with the ATP β- and γ-phosphate groups and is thought to facilitate the transfer of the γ-phosphate to the N-terminal D-alanine substrate to form the acylphosphate intermediate (8). The invisible Arg-233 of GSHase may be the counterpart of Lys-215. While site-directed mutagenesis of Arg-233 to alanine or lysine reduced relative activity to about 5%, the binding affinity of ATP to GSHase was reduced by only 5-fold (17), so the role of Arg-233 is uncertain. Based on the overlay of the two crystallographic structures, the shorter loops in each enzyme are more clearly aligned.

```
-EGSS(151)-  DD-ligase
-GMGG(167)-  GSHase
```

In the DD-ligase complex, the backbone amide groups of Ser-150 and Ser-151 form hydrogen bonds with the β- and γ-phosphate groups, and the two hydroxyl groups bind with the larger loop in the closed (complexed) configuration. In GSHase any interloop linkages would obviously involve only backbone groups.

At the quaternary structure level, the dimer interface in DD-ligase contains eight hydrophobic residues from helices H2, H3, H4, and H5. In tetrameric GSHase, the corresponding interface involves helices H2, H3, and H4 and β-strands B3 and B8. In both enzymes, the subunit interfaces are distant from the catalytic site.

DISCUSSION

After the chance discovery of tertiary structure similarity of DD-ligase and GSHase, a retroactive search and alignment of

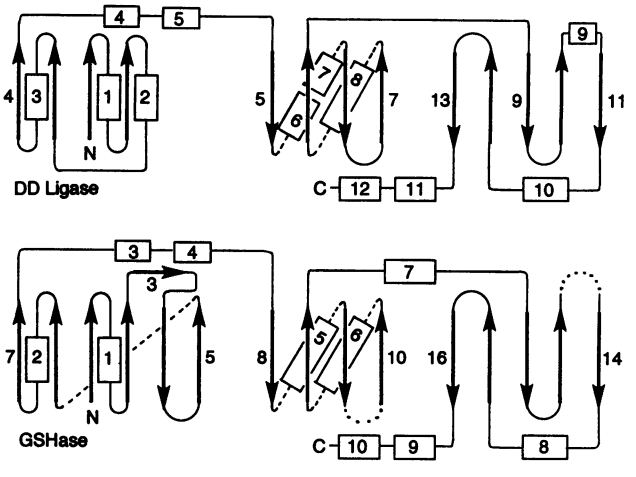


FIG. 3. Secondary structure elements in the three domains of DD-ligase and GSHase. Dotted segments are disordered in GSHase (7).

FIG. 4. Stereoview of α-carbon atom overlay of central and C-terminal domains of DD-ligase (residues 97–306) and GSHase (dotted chain, residues 125–316). Arrow points to loops of DD-ligase covering ADP and phosphoryl phosphinate. Equivalent GSHase chain is absent. View is as in Fig. 1.

the amino acid sequences of the two enzymes was undertaken. The search of more than 130,000 sequences failed to match the two enzymes, possibly because the correct alignment, that is, the structure-based crystallographic alignment, exhibits an exact identity of only 11% and an unimpressive similarity of 25%. An automated sequence alignment produced a Dayhoff-weighted homology of 12%, and it failed to align important structure-matched pairs in the ATP-binding domain (Table 3). Troublesome is the finding that the 15% identity and the 25% similarity based on this incorrect alignment are greater than and equal to the respective percentages based on the correct alignment. On the other hand, the automated search (12) of three-dimensional motifs in the crystallographic data base quickly found a unique match with GSHase.

Comparison of the three domains in DD-ligase with those in GSHase shows that the central and C-terminal domains (Fig. 4) are more similar than the two N-terminal domains, presumably because the former domains contain most of the binding site for ATP and substrates. The ATP site in both enzymes has a unique nonclassical "palmate" motif (7) made of two antiparallel β-sheets. With knowledge of only approximate crystallographic positions for ATP and substrate in GSHase, Tanaka *et al.* (17, 18) did not describe specific side-chain-reactant interactions. Overlaying the DD-ligase structure, we see spatial matching of six GSHase residues that would interact with ATP (Table 3). However, many residues likely to interact with amino acid substrates would be on the two disordered loops, if DD-ligase is accepted as a model (8). Site-directed mutagenesis of GSHase established that one of these loop residues, Arg-233, is required for activity, and its removal affects the binding of substrates more than the binding of ATP (17). The presumed counterpart residue in the DD-ligase, Lys-215, is thought to facilitate phosphoryl transfer by bridging ATP and substrate (8) and its mutation abrogates activity (Y.S. and C.T.W., unpublished data).

The unanticipated architectural homology between GSHase and DD-ligase suggesting a family relationship may reflect the functional commonality that each enzyme produces a small peptide by coupling two amino acids, cleaving ATP to ADP and proceeding via an acylphosphate intermediate. While the specificity for the amino acid substrates is completely distinct, γ-Glu-Cys with glycine vs. D-alanine with D-alanine, the relative positioning of ATP and substrates in GSHase and DD-ligase is similar. Both enzymes have polypeptide segments,

Table 3. Spatially equivalent residues in crystal structures of DD-ligase and GSHase

Residue		
DD-Ligase	GSHase	Location/function
Glu-15	Asp-19?	At N terminus of helix H1 in each enzyme; in DD-ligase, binds α-ammonium group of N-terminal substrate.
Lys-97	Lys-125	On helix H5(DD-ligase)/H4(GSHase); in DD-ligase, near α,β-phosphate groups of ATP.
Lys-144	Lys-160	On β-strand B6/B9; in DD-ligase, near nitrogen of adenine ring, oxygen of ribose ring, and α-phosphate group.
Ser-150	Gly-166?	On shorter flexible loop; invisible in GSHase.
Glu-180	Gln-198	On β-strand B8/B11; in DD-ligase, hydrogen-bonded to adenine amine group.
Trp-182	Tyr-200	Hydrophobic interaction with adenine ring in DD-ligase.
Leu-183	Leu-201	Hydrophobic interaction with adenine ring in DD-ligase.
Glu-187	Asp-208	On β-strand B9/B12; in DD-ligase, hydrogen-bonded to ribose OH groups.
Lys-215	Arg-233?	On larger loop; in DD-ligase, electrostatic binding to phosphate groups; invisible in GSHase.
Asp-257	Asp-273	On β-strand B12/B15; coordinated to Mg^{2+} in DD-ligase.
Leu-269	Thr-280	β-Bulge residue in β-strand B13/B16; in DD-ligase, near adenine and ribose rings.
Glu-270	Glu-281	On β-strand B13/B16; in DD-ligase, below ATP phosphate groups and coordinated to Mg^{2+}.
Asn-272	Asn-283	On β-strand B13/B16; in DD-ligase, below ATP phosphate groups and coordinated to Mg^{2+}.

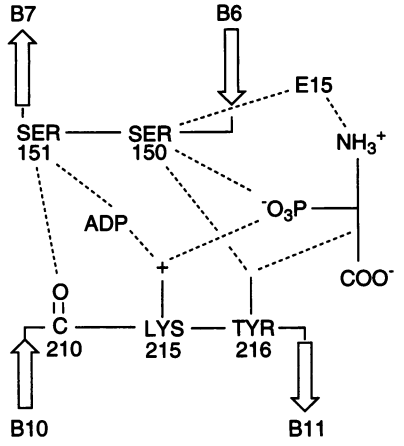


FIG. 5. Schematic of the two loops in DD-ligase found covering ADP and the phosphorylated phosphinate analog (8). Interactions involving side chain and backbone groups are indicated.

first detected by protease sensitivity, that swing inward to cover the substrates, acylphosphate intermediate, and tetrahedral adducts during catalysis. It may be that other nonribosomal peptide synthetases that cleave ATP to ADP and P_i will turn out to be members of this family, each member having flexible loops that protect a large catalytic cavity and function to exclude water and stabilize activated intermediates.

We thank S. J. Remington and G. J. Kleywegt for providing structure fitting programs and J. A. Kelly for helpful discussions. This work was supported by National Institutes of Health Grant 1RO1-AI-34330 (to J.R.K.) and GM-49338 (to C.T.W.).

1. Rhee, S. G., Chock, P. B. & Stadtman, E. R. (1985) *Methods Enzymol.* **113,** 213–241.
2. Allison, R. D. (1985) *Methods Enzymol.* **113,** 419–437.
3. Gushima, H., Miya, T., Murata, K. & Kimura, A. (1983) *J. Appl. Biochem.* **5,** 210–218.
4. Neuhaus, F. C. (1962) *J. Biol. Chem.* **237,** 778–786.
5. Mengin-Lecreulx, D., Flouret, B. & Heijenoort, J. V. (1982) *J. Bacteriol.* **151,** 1109–1117.
6. Yamashita, M. M., Almassy, R. J., Janson, C. A., Cascio, D. & Eisenberg, D. (1989) *J. Biol. Chem.* **264,** 17681–17687.
7. Yamaguchi, H., Kato, H., Hata, Y., Nishioka, T., Kimura, A., Oda, J. & Katsube, Y. (1993) *J. Mol. Biol.* **229,** 1083–1100.
8. Fan, C., Moews, P. C., Walsh, C. T. & Knox, J. R. (1994) *Science* **266,** 439–443.
9. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
10. Higgins, D. G. & Sharp, P. M. (1988) *Gene* **73,** 237–244.
11. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, pp. 345–352.
12. Kleywegt, G. J. & Jones, T. A. (1994) in *From First Map to Final Model*, ed. Bailey, S., Hubbard, R. & Waller, D. (SERC Daresbury Laboratory, Warrington, U.K.), pp. 59–66.
13. Remington, S. J. & Matthews, B. W. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 2180–2184.
14. Robinson, A. C., Kenan, D. L., Sweeney, J. & Donachie, W. (1986) *J. Bacteriol.* **167,** 809–817.
15. Kraulis, P. (1991) *J. Appl. Crystallogr.* **24,** 946–950.
16. Richardson, J. S. & Richardson, D. C. (1989) in *Prediction of Protein Structure and the Principles of Protein Conformation*, ed. Fasman, G. D. (Plenum, New York), pp. 1–99.
17. Tanaka, T., Kato, H., Nishioka, T. & Oda, J. (1992) *Biochemistry* **31,** 2259–2265.
18. Tanaka, T., Yamaguchi, H., Kato, H., Nishioka, T., Katsube, Y. & Oda, J. (1993) *Biochemistry* **32,** 12398–12404.
19. Wright, G. D. & Walsh, C. T. (1993) *Protein Sci.* **2,** 1765–1769.