Three-Dimensional Structure of the Biotin Carboxylase Subunit of Acetyl-CoA Carboxylase†,‡

Grover L. Waldrop,* Ivan Rayment, and Hazel M. Holden*

Institute for Enzyme Research, Graduate School and Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53705

Received April 28, 1994; Revised Manuscript Received June 8, 1994®

ABSTRACT: Acetyl-CoA carboxylase is found in all animals, plants, and bacteria and catalyzes the first committed step in fatty acid synthesis. It is a multicomponent enzyme containing a biotin carboxylase activity, a biotin carboxyl carrier protein, and a carboxyltransferase functionality. Here we report the X-ray structure of the biotin carboxylase component from *Escherichia coli* determined to 2.4-Å resolution. The structure was solved by a combination of multiple isomorphous replacement and electron density modification procedures. The overall fold of the molecule may be described in terms of three structural domains. The N-terminal region, formed by Met 1–Ile 103, adopts a dinucleotide binding motif with five strands of parallel β-sheet flanked on either side by α-helices. The "B-domain" extends from the main body of the subunit where it folds into two α-helical regions and three strands of β-sheet. Following the excursion into the B-domain, the polypeptide chain folds back into the body of the protein where it forms an eight-stranded antiparallel β-sheet. In addition to this major secondary structural element, the C-terminal domain also contains a smaller three-stranded antiparallel β-sheet and seven α-helices. The active site of the enzyme has been identified tentatively by a difference Fourier map calculated between X-ray data from the native crystals and from crystals soaked in a Ag^+/biotin complex. Those amino acid residues believed to form part of the active site pocket include His 209–Glu 211, His 236–Glu 241, Glu 276, Ile 287–Glu 296, and Arg 338. The structure presented here represents the first X-ray model of a biotin-dependent carboxylase.

Acetyl-CoA carboxylase catalyzes the first committed and one of the regulated steps in the biosynthesis of long-chain fatty acids (Wakil *et al.*, 1983). The enzyme, found in all animals, plants, and bacteria, catalyzes the biotin-dependent carboxylation of acetyl-CoA to form malonyl-CoA in a two-step reaction mechanism (Lane *et al.*, 1974; Alberts & Vagelos, 1972):

$$\text{enzyme–biotin} + Mg^{2+}\text{-ATP} + HCO_3^- \overset{Mg^{2+}}{\rightleftharpoons} \text{enzyme–biotin–}CO_2^- + Mg^{2+}\text{-ADP} + P_i \quad (1)$$

$$\text{enzyme–biotin–}CO_2^- + \text{acetyl-CoA} \rightleftharpoons \text{malonyl-CoA} + \text{enzyme–biotin} \quad (2)$$

In *Escherichia coli*, acetyl-CoA carboxylase is composed of three subunits that are isolated separately and that display distinct functional properties (Guchhait *et al.*, 1974a). The biotin carboxylase subunit of acetyl-CoA carboxylase catalyzes the first half-reaction which involves the ATP-dependent carboxylation of the 1′ nitrogen of biotin to form carboxybiotin. As with all biotin-dependent carboxylases studied thus far, bicarbonate serves as the source of CO_2 (Knowles, 1989). The second half-reaction involves the transfer of the carboxyl group from biotin to acetyl-CoA and is catalyzed by the carboxyltransferase subunit. The third component of acetyl-CoA carboxylase is the biotin carboxyl carrier protein which contains biotin covalently attached to the ε-nitrogen of a lysine residue. The current hypothesis is that the biotin carboxyl carrier protein provides a "swinging arm" that moves between the biotin carboxylase and carboxyltransferase subunits of the complex.

The genes encoding the biotin carboxylase and the biotin carboxyl carrier protein have been cloned and overexpressed (Kondo *et al.*, 1991; Li & Cronan, 1992a). Both biotin carboxylase and the carrier protein are homodimers containing 449 and 156 amino acid residues per monomer, respectively. Likewise, the gene encoding the carboxyltransferase has been cloned (Li & Cronan, 1992b). The carboxyltransferase is an $\alpha_2\beta_2$ tetramer with the α- and β-subunits containing 319 and 304 amino acid residues, respectively (Li & Cronan, 1992b). In contrast to the bacterial enzyme, animal acetyl-CoA carboxylases have molecular weights of approximately 225 000 and contain all three functions on a single polypeptide chain (Lane *et al.*, 1974).

Most mechanistic studies of biotin-dependent carboxylation reactions have focused on the biotin carboxylase component of *E. coli* acetyl-CoA carboxylase because it retains enzymatic activity when isolated from the other components. Moreover, biotin carboxylase is one of only two known biotin-dependent carboxylases that will utilize free biotin as a substrate, the other being β-methylcrotonyl-CoA carboxylase (Guchhait *et al.*, 1974b). Several lines of evidence suggest that the chemical mechanism of biotin carboxylase involves the reaction of bicarbonate and ATP to form a carboxyphosphate intermediate (Polakis *et al.*, 1974; Climent & Rubio, 1986; Ogita & Knowles, 1988). The proposed carboxyphosphate intermediate either reacts directly with the enol (enolate) tautomer of biotin to form a tetrahedral adduct or collapses into CO_2 and phosphate with subsequent attack of biotin on CO_2.

† This research was supported in part by grants from the NIH (GM18938 to W. W. Cleland and HL42322 to H.M.H.). H.M.H. is an Established Investigator of the American Heart Association.

‡ The X-ray coordinates have been deposited in the Brookhaven Protein Data Bank (file name 1BNC).

* To whom correspondence should be addressed.

® Abstract published in *Advance ACS Abstracts,* August 1, 1994.

0006-2960/94/0433-10249$04.50/0

Table 1: Intensity Statistics for Native and Derivative Data

	resolution range (Å)						
	overall	100.00–4.00	3.17	2.77	2.52	2.34	2.20
native							
no. of measurements	93923	31832	26780	13092	11498	8685	2036
no. of independent reflections	41566	9233	8791	8073	7575	6302	1592
completeness of data (%)	75	96	94	88	84	69	17
av intensity	5577	10000	5890	2067	1123	744	639
av σ	379	430	419	301	295	309	330
R factor[a] (%)	3.9	3.0	4.7	6.8	11.3	17.5	21.5
$LuCl_3$/AMPPNP							
no. of measurements	171693	69285	53517	24808	19111	4972	0
no. of independent reflections	33956	9451	9051	7134	6308	2012	0
av intensity	6085	10000	5337	1810	953	628	0
av σ	397	420	396	367	364	371	0
R factor (%)	7.9	6.6	9.2	13.8	20.8	27.5	0
K_2PtCl_4							
no. of measurements	171273	70214	53258	24965	18746	4090	0
no. of independent reflections	36833	9301	9034	8611	7797	2090	0
av intensity	5911	10000	4783	1514	819	569	0
av σ	554	544	597	515	524	541	0
R factor	8.3	6.6	10.5	17.5	27.5	37.4	0
trimethyllead acetate							
no. of measurements	172743	69547	54430	25021	18839	4906	0
no. of independent reflections	35497	9066	8486	8206	7410	2329	0
av intensity	6185	10000	5545	1933	1032	672	0
av σ	380	429	396	300	284	280	0
R factor	6.4	5.3	7.5	10.8	15.0	19.7	0

[a] R factor = $(\Sigma|I - \bar{I}|/\Sigma I) \times 100$.

Mechanistic details of biotin carboxylase or any other biotin-dependent enzyme have been hindered, however, by the lack of structural information. For example, nothing is known about the active site residues or how they participate in catalysis. It is also unknown how the enzyme stabilizes carboxyphosphate, a relatively unstable molecule with an estimated half-life of 70 ms (Sauers *et al.*, 1975) or how biotin is converted from the keto tautomer to the more nucleophilic enol tautomer by the enzyme. As the first step toward addressing these issues, we have determined the three-dimensional structure of the apo form of biotin carboxylase from *E. coli* to 2.4-Å resolution as described in this report. The X-ray coordinates have been deposited in the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977).

MATERIALS AND METHODS

Crystallization and Search for Heavy-Atom Derivatives. Crystals of biotin carboxylase were grown by microdialysis against 10 mM potassium phosphate (pH 7.0) according to the procedure of Waldrop *et al.* (1994). They belonged to the space group $P2_12_12_1$ with unit cell dimensions of $a = 61.9$ Å, $b = 96.1$ Å, and $c = 180.6$ Å, and two subunits per asymmetric unit.

For the preparation of heavy-atom derivatives, crystals were transferred at 4 °C to a synthetic mother liquor containing 10% poly(ethylene glycol) 8000, 10 mM HEPES, pH 7.0, and various heavy metal reagents. Most of the compounds either did not react with the crystalline protein or produced nonisomorphous changes. However, it was possible to prepare a heavy-atom derivative by soaking the crystals in 6.0 mM $LuCl_3$ complexed with 150 μM AMPPNP (a nucleotide analog of ATP) for 24 h. Subsequently, two additional derivatives were prepared by soaking the crystals in 6.0 mM $LuCl_3$/150 μM AMPPNP for 3 h and then transferring them to either 0.15 mM K_2PtCl_4 or 1.0 mM trimethyllead acetate solutions for soak times of 24 or 60 h, respectively. It was not possible to prepare isomorphous heavy-atom derivatives with the platinum and lead compounds alone.

X-ray Data Collection and Processing. For X-ray data collection, crystals were mounted in quartz capillary tubes. A native X-ray data set was collected to 2.2-Å resolution from a single crystal at 4 °C with a Siemens X1000D area detector system. The X-ray source was nickel-filtered Cu Kα radiation from a Rigaku RU200 X-ray generator operated at 50 kV and 50 mA and equipped with a 200-μm focal cup. A crystal-to-detector distance of 21 cm was used together with a step size of 0.15° per frame. The heavy-atom-derivative X-ray data sets were collected in a similar manner. Friedel pairs were measured for all reflections in the heavy-atom-derivative data sets. These X-ray data were processed with the data reduction software package XDS (Kabsch, 1988a,b) and internally scaled according to a procedure developed in the laboratory by Dr. Gary Wesenberg (G. Wesenberg and I. Rayment, unpublished results). Relevant X-ray data collection statistics may be found in Table 1. The native X-ray data set was 88% and 75% complete to 2.4- and 2.2-Å resolution, respectively. Each heavy-atom-derivative X-ray data set was placed on the same scale as the native X-ray data set by a "local" scaling procedure developed by Drs. G. Wesenberg, W. Rypniewski, and I. Rayment.

Structure Determination and Least-Squares Refinement. The positions of the heavy-atom-binding sites for the three derivatives were determined by inspection of difference Patterson maps calculated with all X-ray data from 30- to 5-Å resolution. These heavy-atom sites were placed on a common origin by difference Fourier maps. The positions and occupancies were refined by the origin-removed Patterson-function correlation method and are listed in Table 2 (Rossmann, 1960; Terwilliger & Eisenberg, 1983). Anomalous difference Fourier maps calculated from 30- to 5.0-Å resolution were employed for determining the correct hand of the heavy-atom constellation. Protein phases were calculated with the program HEAVY (Terwilliger & Eisenberg, 1983), and relevant phase calculation statistics may be found in Table 3.

Except for the single platinum site, the heavy atoms bound in special positions with y-values near $^1/_4$ or $^3/_4$, as can be seen in Table 2. In addition, each derivative had two common

Table 2: Refined Heavy-Atom Parameters

derivative	site no.	rel occupancy	x	y	z	B^a (Å^2)	$R_{iso}{}^b$ (%)	location
$LuCl_3$/AMPPNP	1	1.76	0.173	0.251	0.126	5.0	13.4	Glu 288
	2	1.71	0.150	0.762	0.880	5.0		Glu 276
								subunits 1 and 2
K_2PtCl_4	1	2.37	0.167	0.247	0.126	5.0	29.0	Met 184
	2	1.77	0.135	0.767	0.882	5.0		subunit 1
	3	3.26	0.219	0.370	0.886	10.0		
trimethyllead acetate	1	1.79	0.178	0.251	0.125	5.0	19.9	Cys 50
	2	1.85	0.134	0.762	0.881	5.0		subunit 2
	3	2.28	0.148	0.247	0.466	10.0		

[a] Temperature factors were not refined. x, y, and z are the fractional atomic coordinates. [b] $R_{iso} = (\Sigma||F_N| - |F_H||/\Sigma|F_N|) \times 100$, where F_N is the native structure factor amplitude and F_H is the derivative structure factor amplitude.

Table 3: Phase Calculation Statistics

	resolution range (Å)							
	∞–8.86	5.64	4.42	3.76	3.32	3.01	2.77	2.58
no. of reflections	1800	3084	3880	4556	4826	4909	5036	4960
figure of merit	0.66	0.68	0.55	0.50	0.42	0.40	0.36	0.26
phasing power[a]								
$LuCl_3$/AMPPNP								
centric reflections	0.80	1.04	0.75	0.79	0.82	1.02	1.16	0.97
acentric reflections	1.01	1.40	1.13	1.13	1.21	1.59	1.59	1.46
K_2PtCl_4								
centric reflections	0.72	0.76	0.55	0.55	0.55	0.63	0.70	0.82
acentric reflections	0.99	0.96	0.70	0.71	0.82	0.90	0.98	1.05
trimethyllead acetate								
centric reflections	0.90	0.97	0.58	0.63	0.62	0.65	0.90	0.79
acentric reflections	1.16	1.20	0.90	0.89	1.02	1.15	1.25	1.24

[a] Phasing power is the ratio of the root-mean-square heavy-atom scattering factor amplitude to the root-mean-square lack of closure error.

sites, namely, the two $LuCl_3$ positions. Consequently, the protein phases, based on the multiple isomorphous derivatives alone, were not sufficient to solve the structure. The addition of phasing information from the anomalous scattering of the heavy-atom derivatives, however, improved the quality of the electron density map such that it was possible to locate the two subunits of biotin carboxylase in the asymmetric unit.

Since there were multiple copies in the asymmetric unit, it was possible to improve further the phases by molecular averaging and solvent flattening. The necessary rotational and translational matrices relating the two subunits in the asymmetric unit were determined and refined by the program MUNCHKINS [developed by Drs. G. Wesenberg and I. Rayment and described in Rypniewski *et al.* (1991)]. A "globally averaged" electron density map based on these matrices was calculated to 5.0-Å resolution, plotted onto transparencies, and stacked onto Plexiglas sheets. Visual examination of this electron density map revealed the protein–solvent boundaries for the "averaged" subunit. A molecular envelope was subsequently drawn around the molecule, digitized on a Calcomp digitizing board, and converted into a logical mask by a series of computer programs written by Drs. I. Rayment and G. Wesenberg. This molecular mask was used for the subsequent refinement of the protein phases by iterative molecular averaging and solvent flattening at 3.0-Å resolution for 15 cycles (Bricogne, 1976). The R factor between the calculated structure factor amplitudes from the averaged electron density map and the observed X-ray data was 19.9%.

The averaged electron density map contained well-defined areas of secondary structure although many loops and random coil regions were unclear. The building of the model began by first placing alanines into those clearly defined regions with the program FRODO (Jones, 1985). Approximately 350 alanines were positioned into the electron density map. This polyalanine model, built from the averaged electron density map, was placed back into the unit cell and subjected to 10 cycles of least-squares refinement with the program TNT at 3.0-Å resolution (Tronrud *et al.*, 1987). New rotational and translational matrices for a second round of averaging at 2.7-Å resolution were calculated from these refined polyalanine coordinates. Additionally, protein phases based on the polyalanine model were combined with the multiple isomorphous replacement phases to 2.7-Å resolution by the program SIGMAA (Read, 1986). Fifteen additional cycles of molecular averaging and solvent flattening were conducted with these "combined" phases serving as the starting point. After this round of cyclic averaging, the R factor between the calculated structure factor amplitudes from the averaged electron density map and the observed X-ray data was 20.6%.

The new averaged electron density map clearly revealed the position of the single tryptophan residue, Trp 367. It was then possible to build the polypeptide chain from Arg 208 to Asp 412. The electron density preceding Arg 208 was very weak. The N-terminus was also located, and it was possible to match the polyalanine model with the amino acid sequence from Met 1 to Cys 130. This partial model was subjected to further least-squares refinement at 2.7-Å resolution. New rotational and translational matrices and protein phases were calculated on the basis of these refined coordinates. The model phases were combined with the multiple isomorphous replacement phases, and fifteen additional cycles of averaging and solvent flattening were conducted. At the end of this cycle of density modification, the R factor between the calculated structure factor amplitudes from the averaged electron density map and the observed X-ray data was reduced to 15.2%. This final averaged electron density map was improved sufficiently such that it was possible to locate Met 200–Pro 207, Gly 413–Leu 446, and many of the residues between Val 131 and Tyr 199 that constitute the "B-domain".

The final averaged model was placed back into the unit cell and subjected to alternate cycles of least-squares refinement and manual model building at 2.4-Å resolution. It was possible to build into the electron density many of the missing residues. The present model, however, does contain several breaks in the polypeptide chain: Ala 160–Gly 166 and Ala 191–Asp 196 in subunit I and Pro 137–Gly 139, Ala 160–Met 169, and Glu 188–Val 198 in subunit II. Also, the last three C-terminal residues are missing in subunit I, and the last C-terminal residue is missing in subunit II. The current R factor is 18.3% for all measured X-ray data from 30.0 to 2.4 Å with root-mean-square deviations from "ideal" geometry of 0.014 Å for bond lengths, 2.3° for bond angles, and 0.010 Å for groups of atoms expected to be coplanar. A total of 116 solvent molecules was included in the model. In addition, two

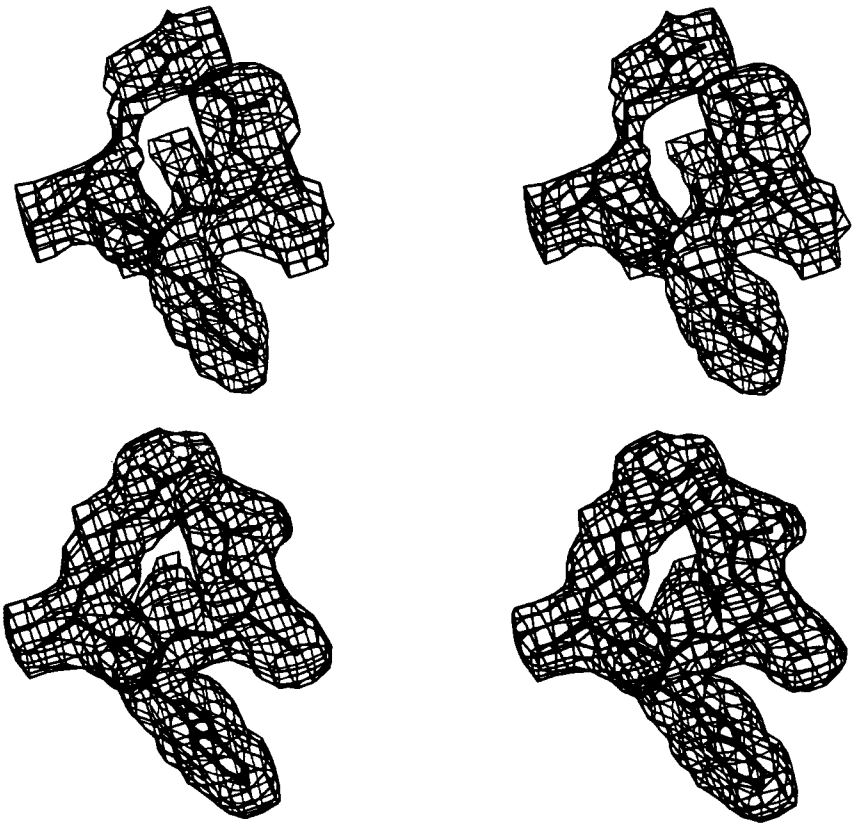

FIGURE 1: Representative portion of the electron density map. (a, top) Shown here is a portion of the electron density in the original map calculated to 3.0-Å resolution with MIR phases and corresponding to residues Arg 366, Trp 367, and Glu 368. It was this region that allowed for the initial correlation between the amino acid sequence and the electron density. (b, bottom) The same portion of the electron density map is shown but calculated with refined coefficients of the form $2F_o - F_c$, where F_o was the native structure factor amplitude and F_c was the calculated structure factor amplitude from the model refined at 2.4-Å resolution. Protein phases were calculated from the refined model. As can be seen, Arg 366 and Glu 368 form a salt bridge such that N^{ϵ} and $N^{\eta 2}$ of Arg 366 are 2.8 and 2.7 Å, respectively, from $O^{\epsilon 1}$ of Glu 368. These figures were prepared with the plotting package FROST, written by Dr. Gary Wesenberg.

phosphate anions, one per subunit, were positioned into the electron density map.

The amino acid sequence of biotin carboxylase was determined by both Kondo *et al.* (1991) and Li and Cronan (1992a). The two primary structures differed at amino acid residues 260 (Cys or Ser) and 261 (Ala or Arg). It was difficult to distinguish between a serine or cysteine residue in an electron density map calculated to 2.4-Å resolution. However, when residue 260 was modeled into the electron density as a cysteine, the refined *B*-values for the β-carbon and the sulfur of the side chain were 28.7 and 30.0 A^2, respectively, for subunit I and 14.9 and 26.6 A^2, respectively, for subunit II. These *B*-factors suggested that the residue was, most likely, a cysteine. At position 261, there was no electron density beyond the β-carbon. Consequently, the model presented here corresponds to the primary structure of Kondo *et al.* (1991).

RESULTS AND DISCUSSION

A representative portion of the electron density map for biotin carboxylase is displayed in Figure 1. Excluding the region from Val 131 to Pro 207 and several C-terminal amino acid residues, the electron density for both subunits (I and II) was of the quality shown in Figure 1. In both subunits, Pro 155 and Pro 244 were modeled in the electron density in the cis conformation. While most of the polypeptide chain backbone dihedral angles adopted values within the theoretically allowed regions, Ala 226, in both subunits, displayed ϕ, ψ angles of 53.0° and −156.7° and of 62.9° and −164.6°, respectively. These alanines reside in a stretch of β-sheet thereby producing a kink in the strand. The electron density for these residues was unambiguous.

A ribbon representation of subunit I of biotin carboxylase is shown in Figure 2a. The overall dimensions are approximately 67 Å × 52 Å × 48 Å, and as can be seen the protein is somewhat asymmetric with a small domain extending from the main body of the subunit. The molecular architecture of biotin carboxylase can be described in terms of three structural motifs. The N-terminal domain, delineated by residues Met 1–Ile 103, consists of five strands of parallel β-pleated sheet flanked on either side by a total of four α-helices as shown in Figure 2b. This type of structural architecture is similar to that observed in the NAD^+-dependent dehydrogenases and in UDP-galactose 4-epimerase and was unanticipitated since biotin carboxylase is ATP-dependent. A superposition of this N-terminal domain with the nucleotide-binding motif of UDP-galactose 4-epimerase is given in Figure 3. The α-carbon positions for 60 structurally equivalent amino acid residues in these two proteins superimpose with a root-mean-square deviation of 1.7 Å according to the algorithm of Rossmann and Argos (1975). In addition to the β-strands and α-helices, the N-terminal domain of biotin carboxylase contains six type I turns.

Following this dinucleotide-binding fold, the polypeptide chain adopts an α-helical conformation from Ala 107 to Ala 126. This helix is distorted by Asp 115, which adopts ϕ, ψ angles of −119° and 122°. The polypeptide chain extends from the main body of the protein outward to form the B-domain. This region of the electron density map, especially in subunit II, was not as well-defined, suggesting some type of conformational flexibility. In particular, the loops connecting the α-helices and β-strands in the B-domain were

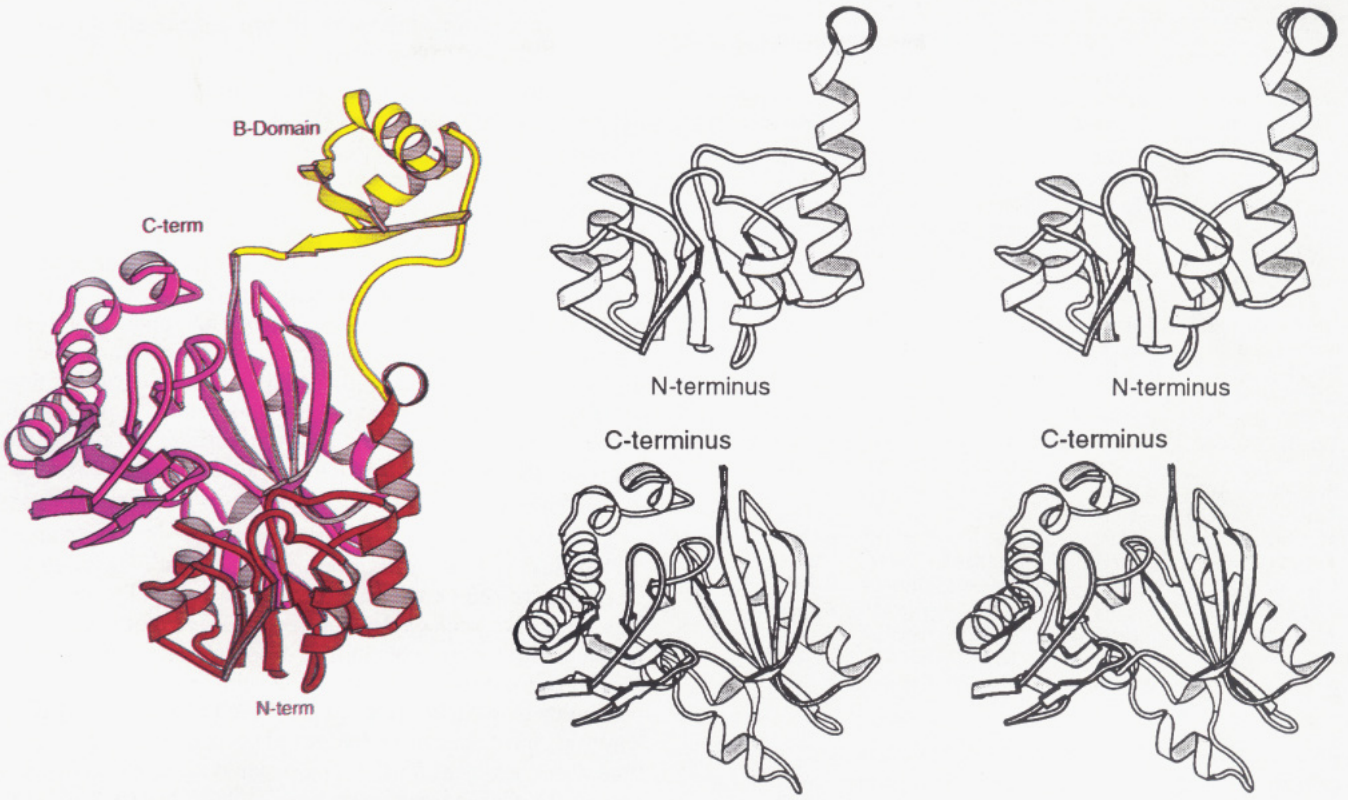

FIGURE 2: Ribbon representation of biotin carboxylase (subunit I). This figure was prepared with the software package MOLSCRIPT (Kraulis, 1991). (a, left) The entire subunit is displayed. The N-terminal domain, defined by amino acid residues Met 1–Pro 105, is shown in red, the B-domain, formed by Val 131–Tyr 203, is displayed in yellow, and the C-terminal domain, delineated by Arg 208 to the C-terminus, is displayed in magenta. (b, top right) The N-terminal domain adopts a dinucleotide binding motif. (c, bottom right) The C-terminal domain contains mostly antiparallel β-sheet. All figures are shown in the same orientation.

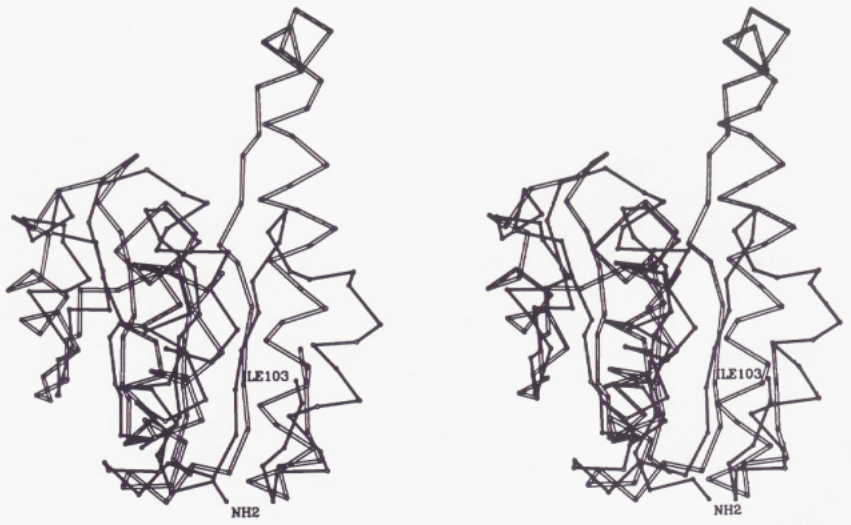

FIGURE 3: Superposition of the nucleotide-binding domains found in UDP-galactose 4-epimerase and biotin carboxylase (subunit I). The filled and open bonds correspond to biotin carboxylase and epimerase, respectively. This dinucleotide binding motif was unanticipated since biotin carboxylase is ATP-dependent. The so-called "P-loop" observed in various ATP-dependent enzymes appears not to be present in biotin carboxylase (Walker *et al.*, 1982).

ill-defined, and many of the side chains were disordered. The amino acid sequence in this region also indicates the potential for conformational flexibility. For example, there are five glycines in tandem (Gly 162–Gly 166), followed by Arg 167, and then another glycine residue (Kondo *et al.*, 1991; Li & Cronan, 1992a). Consequently, the present model contains various breaks in the polypeptide chain. Aside from these breaks, the B-domain contains two α-helices defined by Met 142–Ile 152 and Asp 175–Ala 187 and three strands of antiparallel β-sheet delineated by Tyr 154–Lys 159, Gly 168–Val 172, and Met 197–Leu 204.

After the B-domain, the polypeptide chain proceeds back to the main body of the subunit where it forms the C-terminal domain. The predominant feature of this region is an eight-stranded antiparallel β-sheet as can be seen in Figure 2c. Two of these β-strands, namely, those defined by Asn 221–Asp 229 and Phe 284–Ile 293, are disrupted by β-bulges. In addition to this major structural feature, the C-terminal domain also contains a smaller three-stranded antiparallel β-sheet, seven α-helices, and seven reverse turns. This domain forms a basket that cradles the N-terminal nucleotide-binding motif, and as can be seen in Figure 2a, these two domains

Table 4: List of Secondary Structural Elements

amino acid residue no.	secondary structural element	amino acid residue no.	secondary structural element
Lys 4–Ile 7	β-sheet	Val 240–Pro 244	β-sheet
Glu 12–Cys 21	α-helix	Ala 245–Ile 248	type II turn
Lys 22–Gly 25	type I turn	Pro 250–Asp 266	α-helix
Ile 26–His 32	β-sheet	Tyr 269–Phe 279	β-sheet
Ser 33–Asp 36	type I turn	Glu 280–Glu 283	~type I′ turn
Asp 36–Leu 39	type I turn	Phe 284–Ile 293	β-sheet
Lys 40–Leu 44	α-helix	Gln 294–His 297	type I turn
Glu 47–Gly 52	β-sheet	Pro 298–Thr 304	α-helix
Ser 59–Asn 62	~type I turn	Leu 308–Arg 314	α-helix
Ile 63–Ala 70	α-helix	Ile 315–Gly 318	type I turn
Glu 71–Gly 74	type I turn	Lys 324–Glu 327	type III turn
Ala 77–His 79	β-sheet	His 333–Ala 341	β-sheet
Leu 85–Asn 88	~type I turn	Asp 343–Thr 346	~type I turn
Ala 89–Ser 98	α-helix	Lys 353–Gly 361	β-sheet
Phe 100–Ile 103	β-sheet	Val 365–Ser 369	β-sheet
Ala 107–Ala 126	α-helix	Tyr 372–Tyr 375	type II turn
Met 142–Ile 152	α-helix	Thr 376–Pro 378	β-sheet
Tyr 154–Lys 159	β-sheet	Met 384–Gly 392	β-sheet
Gly 168–Val 172	β-sheet	Arg 395–Glu 408	α-helix
Asp 175–Ala 187	α-helix	Leu 409–Thr 416	β-sheet
Ala 197–Leu 204	β-sheet	Val 418–Asn 426	α-helix
Asn 206–Ala 216	β-sheet	Glu 428–His 432	α-helix
Asn 221–Asp 229	β-sheet	Tyr 439–Leu 446	α-helix

form a decided cleft in the biotin carboxylase structure. A complete listing of the secondary structural elements for subunit I is given in Table 4.

The subunits constituting the dimer have very similar structures except for the B-domains and for the surface loops defined by Ala 341–Pro 351, as indicated in Figure 4. Not surprisingly, the electron density for Ala 341–Pro 351 in the averaged map was very weak, and the position of the polypeptide chain in this region became clear only after least-squares refinement and model building in the unit cell. In subunit I, there are 14 intermolecular contacts within 4.0 Å between the Ala 341–Pro 351 loop and neighboring molecules in the crystalline lattice while in subunit II there are only 3. None of these contacts in subunit II, however, are closer than 3.6 Å, and none are involved in hydrogen-bonding interactions. Most likely, the differences between the two subunits observed in this loop region are due to crystal packing. Excluding the amino acid residues that do not adopt similar positions, the α-carbons for the two biotin carboxylase subunits in the asymmetric unit superimpose with a root-mean-square value of 0.22 Å (Rossmann & Argos, 1975).

The quaternary structure of biotin carboxylase is shown in Figure 5. The B-domains for each subunit are located on opposite edges of the dimer, nearly 90 Å from one another. The subunit–subunit interface for biotin carboxylase is formed by three α-helices (Glu 12–Cys 21, Lys 40–Leu 44, Arg 395–Glu 408) and the β-strand, reverse turn, β-strand motif delineated by Phe 357–Tyr 372. The only tryptophan residue in biotin carboxylase, Trp 367, resides in the β-strand, reverse turn, β-strand motif forming part of the subunit–subunit interface. This region of secondary structure serves as a bridge to connect the 8-stranded C-terminal β-sheets in each subunit, thereby forming an approximate 16-stranded antiparallel β-sheet. The surface area lost upon dimer formation is approximately 2600 Å^2, as calculated according to the method of Lee and Richards (1971) with a probe sphere of 1.4 Å.

A close-up view of the putative active site for biotin carboxylase is given in Figure 6. This region is believed to be the active site on the basis of the following observations. First, there is a peak of electron density located near Glu 296 which is clearly large enough to accommodate a phosphate anion. It is unusual to observe a phosphate ion lying in close proximity to a carboxylate group. However, at pH 7, this anion is most likely hydrogen phosphate such that the protonated oxygen of the anion could interact with the negatively charged carboxylate side chain of Glu 296. In addition to Glu 296, this hydrogen phosphate is surrounded by Lys 238, Arg 292, Arg 338, and Gln 294. Also, the backbone amide nitrogen of Val 295 is approximately 3.2 Å from one of the phosphate oxygens and has the proper geometry for hydrogen bonding. At this resolution, however, it is not possible to position the hydrogen phosphate into the electron density unambiguously, and therefore, it is not appropriate to quote the lengths of specific interactions between the anion and the protein. Furthermore, it is possible that the electron density represents something other than a hydrogen phosphate anion although the density is clearly too large for a water molecule. In fact, when a water molecule was modeled into the electron density, the *B*-value refined to an anomalously low value.

The location of the active site is further supported by the binding position of the Lu^{3+}/AMPPNP complex. A difference Fourier map calculated between the native and native/Lu^{3+}/

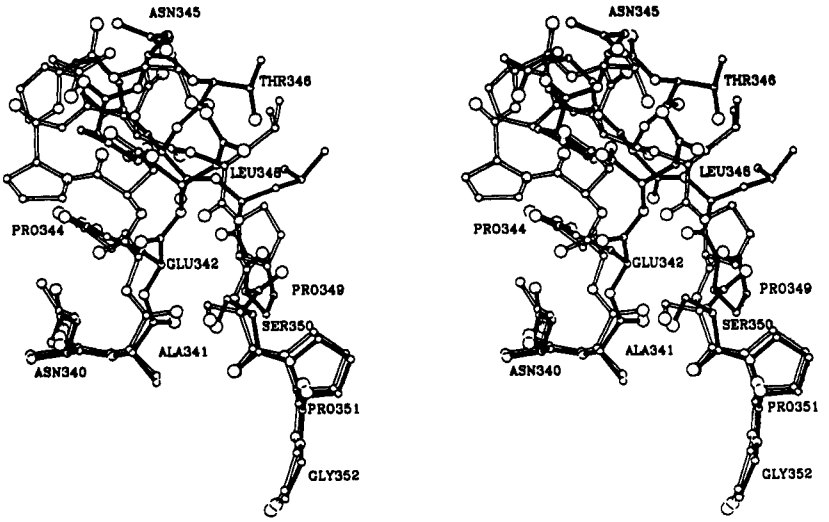

FIGURE 4: Superposition of the loops formed by Asn 340–Gly 352 in subunits I and II. Apart from the B-domains, the only other region where the two subunits constituting the dimer significantly differ is shown here. Subunits I and II are represented in solid and open bonds, respectively.

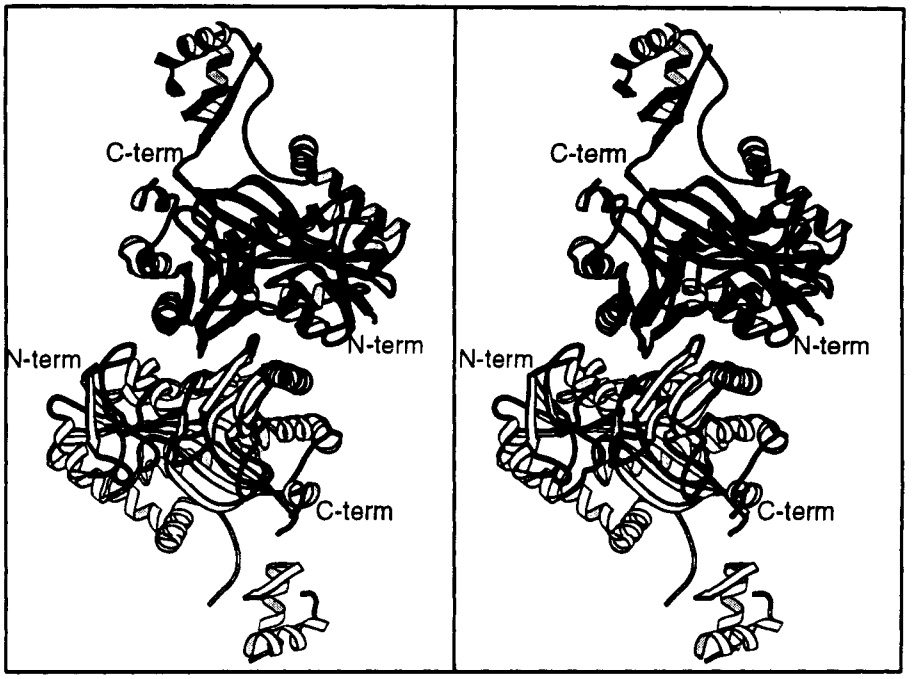

FIGURE 5: Ribbon representation of the quaternary structure of biotin carboxylase. The overall dimensions of the dimer are approximately 93 Å × 92 Å × 50 Å.

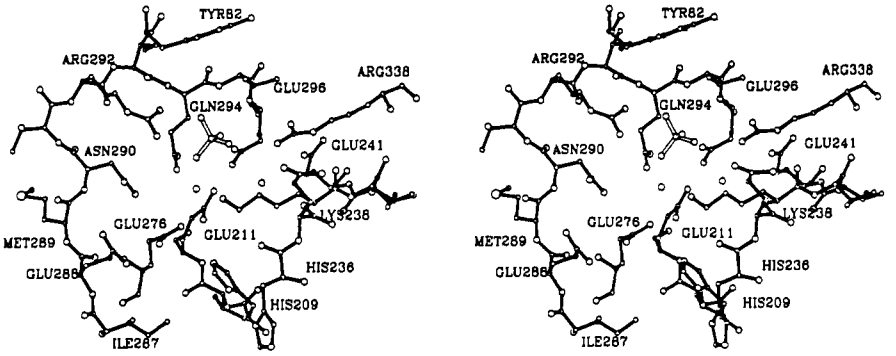

FIGURE 6: Close-up view of the active site (subunit I). Those amino acid residues located within approximately 5 Å of the phosphate binding site and the Lu^{3+} binding site are shown. The putative phosphate anion is depicted in an open-bond representation. In addition to the phosphate, there are two water molecules located within the active site and are displayed as spheres. These water molecules and the phosphate are displaced when biotin is bound in the active site.

AMPPNP X-ray data sets at 3.0-Å resolution clearly reveals that the Lu^{3+} is chelated by the carboxylate groups of Glu 276 and Glu 288. The electron density for the AMPPNP portion of the derivative, however, is not as clear, suggesting conformational flexibility. Since the Lu^{3+}/AMPPNP binding site is near the surface of the molecule, the nucleotide may, indeed, adopt various orientations.

The third piece of evidence for the active site is the location of the difference electron density in a map calculated between the native X-ray data set and the X-ray data set collected from a crystal soaked in a saturated solution of a Ag^{+}/biotin complex (Aoki & Saenger, 1983) as shown in Figure 7. This difference electron density, determined at 3.0-Å resolution, can easily accommodate a biotin molecule although the silver ion appears to be missing. The two ordered water molecules and the hydrogen phosphate in the active site are displaced when the Ag^{+}/biotin complex binds to the enzyme. Those amino acid residues surrounding the biotin molecule include His 236, Lys 238, Glu 276, Glu 288, Asn 290, Arg 292, Gln 294, Glu 296, and Arg 338. While a detailed description of the interactions between biotin and the enzyme must await the determination and refinement of the model to higher resolution, on the basis of the electron density map calculated at 3.0-Å resolution, however, it appears as if His 236, Lys 238, Glu 276, and Glu 288 interact with the ureido group of the cofactor.

There are two highly conserved regions of amino acid sequence identity between biotin carboxylase and carbamoyl phosphate synthetase which also catalyzes the formation of carboxyphosphate from Mg^{2+}-ATP and bicarbonate (Wimmer *et al.*, 1979). One of these regions is defined by Glu 288, Met 289, Asn 290, Thr 291, and Arg 292 (Kondo *et al.*, 1991). The glutamate, asparagine, and arginine residues are completely conserved in both homologous halves of carbamoyl phosphate synthetase isolated from *E. coli* and in the synthetases identified from seven other sources (Quinn *et al.*, 1991). Likewise, these amino acid residues are conserved in chicken acetyl-CoA carboxylase, *Pseudomonas aeruginosa* acetyl-CoA carboxylase, rat propionyl-CoA carboxylase, and yeast pyruvate carboxylase (Kondo *et al.*, 1991; Best & Knauf, 1993). Additionally, the Glu 841 to Lys 841 substitution in carbamoyl phosphate synthetase, which corresponds to Glu 288 in biotin carboxylase, abolishes the overall synthesis of carbamoyl phosphate with either glutamine or ammonia as the source of nitrogen (Guillou *et al.*, 1992). These amino acid sequence homologies further support the idea that the region shown in Figure 7 is, indeed, the active site for biotin carboxylase. Structural details of the active site, however,

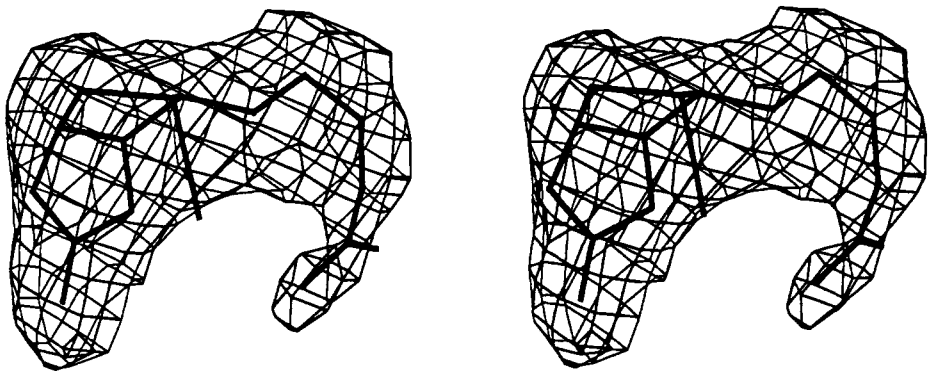

FIGURE 7: Difference electron density corresponding to the Ag^+/biotin complex. The electron density shown here was calculated to 3.0-Å resolution with coefficients of the form $F_{nb} - F_n$, where F_n was the observed structure factor amplitude for the native protein and F_{nb} was the observed structure factor amplitude for the biotin-containing protein. The difference map shown was contoured at 3σ with the highest peak of 13σ located near the tetrahydrothiophene ring of the cofactor. Protein phases were calculated from the native model lacking solvent molecules. The solid bonds correspond to the X-ray coordinates for the Ag^+/biotin complex as determined by Dr. Jim Thoden. It was expected that the highest peak in the electron density map would correspond to the heavy metal. As can be seen, however, the silver ion extends out of the electron density and has therefore most likely dissociated from the complex. Unlike that observed in strepavidin, avidin, and BirA, in biotin carboxylase the valeric acid side chain of the cofactor curls backward toward the bicyclic ring system rather than adopting an extended conformation (Weber *et al.*, 1989; Livnah *et al.*, 1993; Wilson *et al.*, 1992).

must await the results of the substrate analog binding studies presently underway.

Perhaps the most striking aspect of the biotin carboxylase structure is the extension of the B-domain away from the body of the enzyme. The role of the B-domain is presently not understood, but it may be involved in the interaction with the biotin carboxyl carrier protein. Alternatively, it may act as a "lid" that closes down on the active site when biotin, Mg^{2+}-ATP, and HCO_3^- are positioned for catalysis. In light of the intermolecular packing interactions displayed by the present crystal form, this type of large-scale domain movement would undoubtedly destroy the crystalline lattice. It should be noted that the temperature factors are significantly higher in the B-domain and that the crystals crack within 7 min upon addition of Mg^{2+}-ATP and bicarbonate. Thus, in order to fully understand the catalytic mechanism of biotin carboxylase, it may be necessary, for example, to cocrystallize the enzyme with the biotin carboxyl carrier protein or in the presence of Mg^{2+}-AMPPNP, HCO_3^-, and biotin. These experiments are also underway.

ACKNOWLEDGMENT

We thank Drs. Gary Wesenberg and Frank Raushel for helpful discussions and Dr. James Thoden for preparing the Ag^+/biotin complex and determining its three-dimensional structure. We gratefully acknowledge Dr. John Cronan at the University of Illinois for graciously proving the *E. coli* strain employed in this investigation and the reviewers of the manuscript who provided helpful suggestions.

REFERENCES

Alberts, A. W., & Vagelos, P. R. (1972) in *The Enzymes* (Boyer, P. D., Ed.) Vol. 6, pp 37–82, Academic Press, New York.
Aoki, K., & Saenger, W. (1983) *J. Inorg. Biochem. 19*, 269–273.
Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rogers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535–542.
Best, E. A., & Knauf, V. C. (1993) *J. Bacteriol. 175*, 6881–6889.
Bricogne, G. (1976) *Acta Crystallogr., Sect. A 32*, 832–847.
Climent, I., & Rubio, V. (1986) *Arch. Biochem. Biophys. 251*, 465–470.
Guchhait, R. B., Polakis, S. E., Dimroth, P., Stoll, E., Moss, J., & Lane, M. D. (1974a) *J. Biol. Chem. 249*, 6633–6645.
Guchhait, R. B., Polakis, S. E., Hollis, D., Fenselau, C., & Lane, M. D. (1974b) *J. Biol. Chem. 249*, 6646–6656.
Guillou, F., Liao, M., Garcia-Espana, A., & Lusty, C. J. (1992) *Biochemistry 31*, 1656–1664.
Jones, T. A. (1985) *Methods Enzymol. 115*, 157–171.
Kabsch, W. (1988a) *J. Appl. Crystallogr. 21*, 67–71.
Kabsch, W. (1988b) *J. Appl. Crystallogr. 21*, 916–924.
Knowles, J. R. (1989) *Annu. Rev. Biochem. 58*, 195–221.
Kondo, H., Shiratsuchi, K., Yoshimoto, T., Masuda, T., Kitazono, A. T., Tsuru, D., Anai, M., Sekiguchi, M., & Tanabe, T. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9730–9733.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946–950.
Lane, M. D., Moss, J., & Polakis, S. E. (1974) *Curr. Top. Cell. Regul. 8*, 139–195.
Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379–400.
Li, S.-J., & Cronan, J. E., Jr. (1992a) *J. Biol. Chem. 267*, 855–863.
Li, S.-J., & Cronan, J. E., Jr. (1992b) *J. Biol. Chem. 267*, 16841–16847.
Livnah, O., Bayer, E. A., Wilchek, M., & Sussman, J. L. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 5076–5080.
Ogita, T., & Knowles, J. R. (1988) *Biochemistry 27*, 8028–8033.
Polakis, S. E., Guchhait, R. B., Zwergel, E. E., & Lane, M. D. (1974) *J. Biol. Chem. 249*, 6657–6667.
Quinn, C. L., Stephenson, B. T., & Switzer, R. L. (1991) *J. Biol. Chem. 266*, 9113–9127.
Read, R. J. (1986) *Acta Crystallogr., Sect. A 42*, 140–149.
Rossmann, M. G. (1960) *Acta Crystallogr. 13*, 221–226.
Rossmann, M. G., & Argos, P. (1975) *J. Biol. Chem. 250*, 7525–7532.
Rypniewski, W. R., Breiter, D. R., Benning, M. M., Wesenberg, G., Oh, B.-H., Markley, J. L., Rayment, I., & Holden, H. M. (1991) *Biochemistry 30*, 4126–4131.
Sauers, C. K., Jencks, W. P., & Groh, S. (1975) *J. Am. Chem. Soc. 97*, 5546–5553.
Terwilliger, T. C., & Eisenberg, D. (1983) *Acta Crystallogr., Sect. A 39*, 813–817.
Tronrud, D. E., Ten Eyck, L. F., & Matthews, B. W. (1987) *Acta Crystallogr., Sect. A 43*, 489–501.
Wakil, S. J., Stoops, J. K., & Joshi, V. C. (1983) *Annu. Rev. Biochem. 52*, 537–579.
Waldrop, G., Holden, H. M., & Rayment, I. (1994) *J. Mol. Biol. 235*, 367–369.
Walker, J. E., Saraste, M., Runswick, M. J., & Gay, N. J. (1982) *EMBO J. 1*, 945–951.
Weber, P. C., Ohlendorf, D. H., Wendoloski, J. J., & Salemme, F. R. (1989) *Science 243*, 85–88.
Wilson, K. P., Shewschuk, L. M., Brennan, R. G., Otsuka, A. J., & Matthews, B. W. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9257–9261.
Wimmer, M. J., Rose, I. A., Powers, S. G., & Meister, A. (1979) *J. Biol. Chem. 254*, 1854–1859.