Proc. Natl. Acad. Sci. USA
Vol. 77, No. 1 pp. 277–280, January 1980
Biochemistry

Structure of potato inhibitor complex of carboxypeptidase A at 5.5-Å resolution

(enzyme complex/protease inhibitors/protein crystallography)

DOUGLAS C. REES AND WILLIAM N. LIPSCOMB*

Gibbs Chemical Laboratory, Harvard University, Cambridge, Massachusetts 02138

Contributed by William N. Lipscomb, October 18, 1979

ABSTRACT **The complex of the 39-amino acid inhibitor (potato) of bovine carboxypeptidase A (carboxypolypeptidase; peptidyl-L-amino-acid hydrolase, EC 3.4.12.2) was crystallized in space group $P3_2$. There are two protein–inhibitor complexes in the asymmetric unit. These crystals exhibited pseudo-$P3_221$ symmetry due to twinning about the a_3 axis. Heavy atom difference Patterson maps and rotation functions indicated, however, that the noncrystallographic twofold axis that relates these two complexes is nearly coincident with the a_3 axis. Consequently, to a good approximation at low resolution, the space group of the complex is $P3_221$ and the effects of twinning may be ignored. The structure was solved by using multiple isomorphous replacement and molecular replacement techniques. At 5.5-Å resolution, the multiple isomorphous replacement map was readily interpretable in terms of the known native carboxypeptidase A structure plus extra density around the active site. The position of this extra density is consistent with the binding mode for extended substrates proposed from earlier model building studies with the native enzyme (Lipscomb, W. N., Hartsuck, J. A., Reeke, G. N., Quiocho, F. A., Bethge, P. H., Ludwig, M. L., Steitz, T. A., Muirhead, H. & Coppola, J. C. (1968) *Brookhaven Symp. Biol.* 21, 24–90).**

Crystallographic studies on the complexes of small peptides and inhibitors with bovine carboxypeptidase A (CPase A) have led to specific structural proposals concerning the interactions between substrates and CPase A (1–3). Although not an allosteric enzyme, CPase A exhibits a range of kinetic anomalies with various substrates, including substrate inhibition and activation and product inhibition and activation (2). Such behavior implies that wide variations in the binding modes of different molecules to CPase A are permitted. Apparently, substrate interactions well removed from the active site can influence the catalytic mechanism, because at least five carboxy-terminal residues of the substrate affect the binding and kinetic parameters (4). In order to investigate these effects, it is necessary to determine the structure of a complex of an extended substrate with CPase A.

We report here the crystallization and low resolution structure of the complex of CPase A with the CPase A inhibitor from potatoes (PCI), which competitively inhibits CPase A with an inhibition constant of 5×10^{-9} M (5). The sequence of the 39 amino acids in PCI, including the cysteine pairings of the three disulfides, has been established (6, 7). Although *a priori* there is no need for the interaction between CPase A and PCI to resemble the binding mode of substrates, experience with the naturally occurring protease inhibitors of the serine protease family suggests this mode as a strong possibility (8). Chemical modification experiments indicating that the carboxy-terminal glycine of PCI is involved in the inhibition are also consistent with this viewpoint (9).

Because of the exopeptidase activity of CPase A, the CPase A–PCI complex has interesting implications for the reactive site model of protease inhibitor action (10). For many serine protease inhibitors, incubation of the inhibitor with the appropriate protease produces a modified inhibitor species. Although the reactive site peptide bond has been cleaved in the modified inhibitor, this form still retains inhibitory activity. Because the two forms of inhibitor are in equilibrium in the presence of protease, the protease–inhibitor complex must represent an intermediate in the cleavage of the scissile peptide bond. For the modified form to remain inhibitory, however, the two constituent peptide chains must be prevented from any dissociation that would destroy the modified inhibitor (10). In the serine protease inhibitors for which structures have been determined, the two peptides are joined together by disulfide bonds. By assuming that the carboxy-terminal glycine is bound to the active site of CPase A, however, no disulfide bond is possible to this residue, and it is unlikely that noncovalent interactions will prevent dissociation of the glycine from the remainder of the inhibitor. Consequently, the CPase A–PCI complex is of interest not only for providing details of substrate interactions with CPase A but also for understanding the nature of protease inhibitor action.

METHODS

Bovine CPase A (Cox) was purchased from Sigma and used without further purification. PCI was prepared from Russett-Burbank potatoes by the method of Ryan *et al.* (5). The complex was formed by adding a 30% molar excess of PCI to CPase A and dialyzing against 0.25 M LiCl/0.02 M Tris·HCl, pH 7.5. The complex was crystallized as hexagonal plates by the method of hanging drop vapor diffusion against 8% polyethylene glycol-6000 (11). Systematic absences identified the space group as $P3_2$, (or the enantiomeric space group $P3_1$), with cell constants $a = 53.45$ Å, $c = 218.5$ Å. By assuming the same crystal density as native CPase A crystals, two CPase A–PCI complexes are contained in the crystallographic asymmetric unit. The crystals also exhibit strong pseudo $P3_221$ symmetry, suggesting that the noncrystallographic twofold axis is near the a_3 axis, parallel to the hexagonal u axis.

The data were collected at room temperature on a Syntex $P2_1$ diffractometer with Cu$K\alpha$ radiation and a graphite monochromator. A 50-cm helium beam tunnel was placed between the crystal and detector to improve counting statistics (12). Five check reflections were monitored every 300 reflections, and data collection was terminated when the average of these reflections had decayed more than 15%. The crystal lifetime was approximately 100 hr for native crystals and 60 hr for derivative crystals. In addition to the standard Lorentz and polarization

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: CPase A, carboxypeptidase A; PCI, CPase A inhibitor from potatoes; MIR, multiple isomorphous replacement; MR, molecular replacement.

* To whom reprint requests should be addressed.

Table 1. Preparation of heavy atom derivatives

Compound	Concentration, mM	Soaking time, days	Buffer*
$HgCl_2$	1	5	LiCl/Tris
$AgNO_3$	5	14	Acetate
$K_2Pt(NO_2)_4$	2	5	LiCl/Tris
K_2PtCl_4	1	5	LiCl/Tris

* LiCl/Tris, 0.10 M LiCl/0.02 M Tris·HCl, pH 7.5/4% polyethylene glycol-6000; acetate, 0.1 M Na acetate, pH 7.5/4% polyethylene glycol-6000.

factors, a linear decay correction and a semiempirical absorption correction (13) were applied during initial data processing.

Additional data sets were also collected by using the oscillation film method. Data collection and processing procedures are described in ref. 14.

Conditions for the preparation of heavy atom derivatives are given in Table 1. All derivatives used in the native CPase A study produced significant intensity changes, except for lead citrate. This derivative presumably does not bind due to blockage of the active site by PCI. Lattice constants for all derivatives differed from the native by less than 0.2%, except for the K_2PtCl_4 derivative, in which the a and b axes had expanded by 1.5%. Data sets for the native protein and for the $HgCl_2$, $AgNO_3$, and $K_2Pt(NO_2)_4$ derivatives were collected to 2.8 Å, and those for the K_2PtCl_4 derivative, to 5.5 Å. Values of $R = \Sigma|I_{hi} - I_{hj}|/\Sigma|I_{hi} + I_{hj}|$ summed over all symmetry-related reflections h, collected from crystals i and j, were 0.045, 0.049, 0.060, 0.053, and 0.048 for the native, $HgCl_2$, $AgNO_3$, $K_2Pt(NO_2)_4$, and K_2PtCl_4 data sets, collected from 5, 10, 9, 1, and 1 crystal(s), respectively.

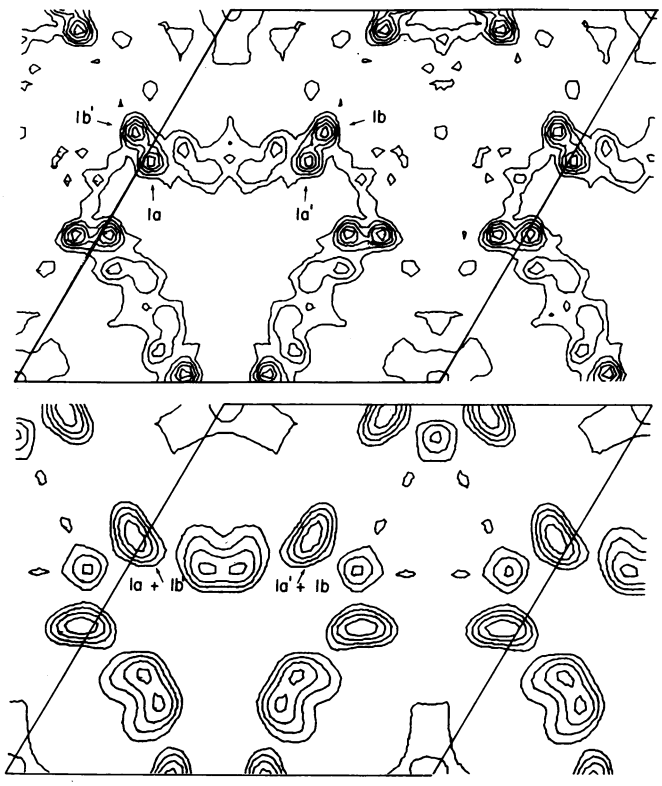

FIG. 1. $w = 2/3$ Harker section of the $HgCl_2$ difference Patterson function at 5.5 Å (*A*) and 2.8 Å (*B*) resolution. A full trigonal unit cell along the u and v axes is illustrated. The position of the self vector from the major $HgCl_2$ site is designated 1a. Site 1b is related to site 1a by the noncrystallographic twofold axis. Sites 1a′ and 1b′ are generated from sites 1a and 1b, respectively, by the twinning operation.

Location and Refinement of Heavy Atom Parameters. The locations of the first mercury sites were determined from a 5.5-Å *h0l* difference Patterson function by using the method of vector superposition (15). Subsequent sites were located by using three-dimensional difference Patterson and difference Fourier techniques (16). A number of peaks that were unresolved in the 5.5-Å difference Patterson maps split at 2.8-Å resolution into two peaks whose corresponding heavy atom positions were related by a twofold axis approximately parallel to the a_3 axis (Figs. 1 and 2). The important consequence of this observation is that, to a good approximation at low resolution, the noncrystallographic twofold axis is along the crystallographic a_3 axis and the corresponding space group becomes $P3_221$. From the heavy atom positions determined from high-resolution Patterson functions, the noncrystallographic twofold axis was more accurately positioned 1.8° from the a_3 axis in the xy plane and 0.7° below the xy plane, failing to intersect the c axis by 2.3 Å.

In addition to pairs of sites related by a noncrystallographic twofold axis, peaks were also present in the difference Patterson functions corresponding to pairs of sites exactly related by a twofold axis along the a_3 axis. This behavior suggests that the crystals suffer from twinning by merohedry about the a_3 axis (17). In this type of twinning, an apparently single crystal is actually composed of domains that are related by a symmetry element occurring in the lattice but not in the crystal. If this twinning operation were perfect, it would create a strict twofold relationship in the data by superimposing the *hkl* and *khl̄* reflections from the two domains. Patterson functions of heavy atom differences calculated from these intensities would consequently reveal pairs of sites related by an apparent twofold

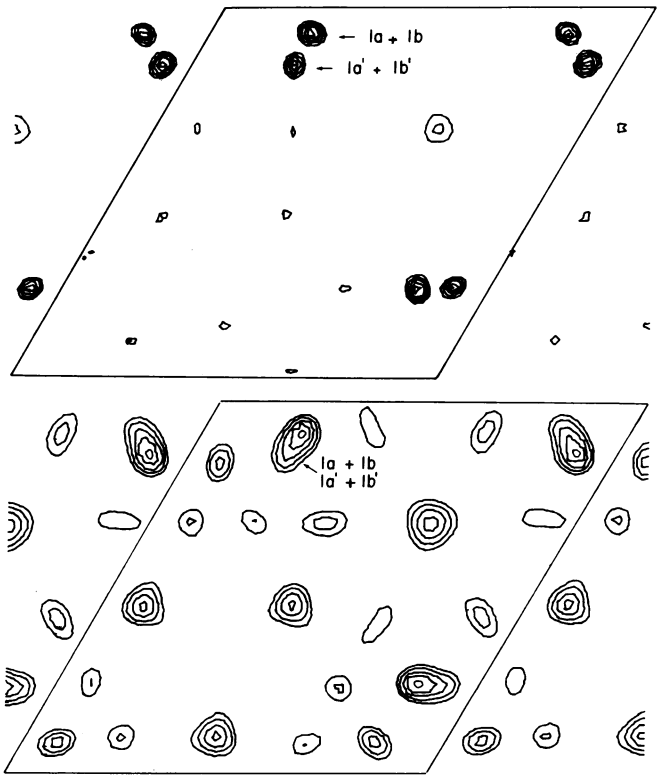

FIG. 2. $w = 2/3$ Harker section of the $K_2Pt(NO_2)_4$ difference Patterson function at 5.5 Å (*A*) and 2.8 Å (*B*) resolution. The relationships among the self vectors generated from the major $K_2Pt(NO_2)_4$ site by the noncrystallographic twofold and twinning operations are indicated with the same labels as in Fig. 1.

axis along the a_3 axis. The relative peak heights of these sites would be determined by the twinning fraction of the crystals and would be equal only in the case of perfect twinning. Subsequent application of several tests for twinning based on correlations between intensities of the *hkl* and *khl̄* reflections (18, 19) and development of a quantitative twinning test based on intensity statistics (ref. 20; unpublished results) supported the twinning interpretation and indicated that various crystals had twinning fractions of 30–40%. The twinning fraction did not vary much from crystal to crystal, because the R values for merging the 2.8-Å resolution data sets were under 6%, well within the normally observed range.

The effect of noncrystallographic symmetry and twinning on the difference Patterson functions are illustrated in Figs. 1 and 2 for the Harker section at $w = 2/3$ of the $HgCl_2$ and $K_2Pt(NO_2)_4$ derivatives. The $HgCl_2$ data represent merged diffractometer data from 10 crystals. In order to emphasize the twinning effects, these data were perfectly twinned by averaging the intensities of the *hkl* and *khl̄* reflections. The $K_2Pt(NO_2)_4$ data were collected from a single crystal whose twinning fraction was estimated to be 43% by using the methods in refs. 19 and 20.

Due to proximity of the noncrystallographic twofold axis to the a_3 axis (twinning axis), we decided to refine the heavy atom parameters in space group $P3_221$ at low resolution. Because this space group corresponds to a perfectly twinned data set, the intensities of the twin-related reflections were averaged. Initial heavy atom positions and occupancies were refined by using cycles of multiple isomorphous replacement (MIR) centroid phasing and least squares (21). Temperature factors for the heavy atom sites were fixed at 30 Å^2 (22). The final heavy atom parameters appear in Table 2. The overall figure of merit at 5.5 Å was 0.77. Phasing statistics from the last refinement cycle are given in Table 3. When the heavy atom refinements were extended to 2.8 Å (in the space group $P3_221$), the overall figure of merit was 0.60. These indices of agreement are surprisingly good for a crystal structure that is twinned and that is assigned the wrong space group.

Table 2. Heavy atom parameters

Site	Occupancy	Axes x	y	z	Native CPase A site*
$HgCl_2$					
1	0.59	−0.461	−0.926	−0.0360	Hg,s_3
2	0.42	−0.382	−0.905	−0.0576	Hg,s_2
$AgNO_3$					
1	0.54	−0.482	−0.915	−0.0349	Ag_3
2	0.14	−0.706	−0.954	−0.0358	
3	0.49	−0.321	−0.193	−0.1311	Ag_1
4	0.64	−0.634	−0.193	−0.1306	
5	0.62	−0.576	−0.924	−0.1292	
6	0.27	−0.910	−0.920	−0.0030	Ag_2
7	0.42	−0.586	−0.709	−0.0639	
8	0.21	−0.754	−0.930	−0.0017	
9	0.33	−0.812	−0.436	−0.1245	
10	0.30	−0.449	−0.952	−0.0419	
$Pt(NO_2)_4$					
1	0.35	−0.798	−0.459	−0.1085	Pt_4
2	0.10	−0.625	−0.201	−0.1272	Pt_2
K_2PtCl_4					
1	0.32	−0.309	−0.001	−0.0670	Pt_1
2	0.21	−0.793	−0.461	−0.1075	Pt_4
3	0.20	−0.623	−0.201	−0.1293	Pt_2
4	0.22	−0.651	−0.201	−0.1095	
5	0.15	−0.016	−0.513	−0.0987	

* For a description of the native CPase A sites, see refs. 1, 2, and 3.

Table 3. Statistics of phasing at 5.5-Å resolution*

Derivative	rms F_c†	rms E_j‡	R_K§	R_c¶
$HgCl_2$	111.5	62.2	0.10	0.48
$AgNO_3$	127.1	62.6	0.11	0.45
$K_2Pt(NO_2)_4$	64.1	45.2	0.08	0.54
K_2PtCl_4	80.9	62.7	0.11	0.58

* The overall figure of merit was 0.77.

† rms, Root mean square. rms $F_c = (\Sigma f_{hj}^2/n)^{1/2}$, where f_{hj} is the heavy atom scattering amplitude for reflection h of derivative j. The root mean square structure factor for the native protein was 494.

‡ rms $E_j = (\Sigma e_{hj}^2/n)^{1/2}$, where e_{hj} is the lack of closure for reflection h of derivative j.

§ $R_K = \Sigma(||F_{pH}^0| - |F_p^c + f_h||)/\Sigma(F_{pH}^0)$ summed over all reflections, where F_{pH}^0 is the absolute observed structure factor for the derivative and F_p^c and f_h are the calculated structure factors for the native protein and heavy atom, respectively.

¶ $R_c = \Sigma(||F_{pH}^0 - F_p^0| - |f_h||)/\Sigma|F_{pH}^0 - F_p^0|$ calculated for centric reflections only.

Molecular Replacement. Because the CPase A molecule contributes approximately 90% of the mass of the complex, phases for the structure were calculated by using molecular replacement (MR) methods with the known CPase A structure (23). Rotational and translational parameters for placing the native molecule in the CPase A–PCI unit cell were determined by using Crowthers' fast rotation function (24) and Crowther and Blow's translation function (25). The location of the intermolecular vector between two molecules related by the threefold screw axis in the $w = 2/3$ section of the translation function indicates that a 3_2 screw axis is present in the unit cell, which uniquely establishes the space group of the complex as $P3_2$. Solutions corresponding to twinned and noncrystallographically related molecules were present in the high-resolution calculations, although the peaks were not resolved at low resolution. The relative heights of the translation function peaks were found to give an estimate of the twinning fraction consistent with results obtained from other methods (18, 19).

For the low-resolution study, a single set of rotation and translation parameters was determined in space group $P3_221$. By using the orthogonal coordinate system reported for the native CPase A coordinates (2) and an orthogonal coordinate system oriented along the y and z axes of the trigonal CPase A–PCI lattice, the rotational Euler angles and translation vectors required to place the native molecule in the complex unit cell are 266.7°, 23.7°, and 324.7° and 14.40 Å, 11.95 Å, and −11.75 Å, respectively. With these coordinates, the R factor between calculated and observed structure factors was 0.40 at 5.5-Å resolution and 0.47 at 2.8-Å resolution. The average phase shift between the MR and MIR phases was 53° at 5.5-Å resolution.

By using the MR results, it is possible to correlate the CPase A–PCI heavy atom sites with sites located in the native structure work. This comparison appears in Table 2. As expected, no substitution of the zinc in these derivatives was found. The $K_2Pt(NO_2)_4$ derivative differs from the K_2PtCl_4 derivative by failing to bind to cysteine-161 (native site Pt_1). The location of this residue near an intermolecular contact may explain the increased lattice dimensions observed for the K_2PtCl_4 derivative. The root mean square deviation between the calculated and observed positions of sites common to the two structures was 1.1 Å.

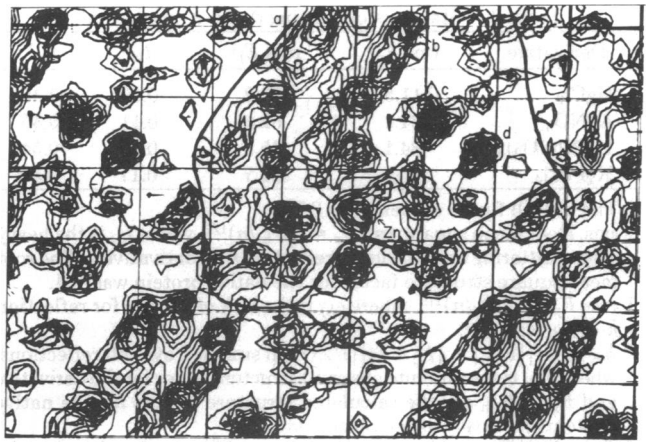

FIG. 3. Superposition of the sections from $z = -0.08$ to -0.03 of the 5.5-Å resolution MIR electron density map of the CPase A–PCI complex. The black outlines indicate the boundaries of the component proteins. The position of the zinc site in the native enzyme and the α-helices with residues 14–28 (a), 72–88 (b), 285–306 (c), and 215–231 (d) are labeled. The grid spacings are at 10.7-Å intervals.

RESULTS AND DISCUSSION

The electron density map of the CPase A–PCI complex was calculated at 5.5 Å with figure of merit weighted centroid MIR phases (26). By using the MR parameters, a 5.5-Å map of the native molecule in the CPase A–PCI unit cell was also calculated as an aid in identifying inhibitor electron density.

The major features of the native molecule are apparent in the map of the complex (Fig. 3). Especially prominent are the α-helices containing residues 14–28 and 72–88 (in the plane of the sections in Fig. 3) and residues 285–306 and 215–231 (perpendicular to the sections). Of the four or five sets of intermolecular contacts present in the native and complex structures, only the contact site involving the amino-terminal residues and residues 232–234 is common to both structures.

Comparison of maps of the complex and of the native enzyme established the presence of strong non-CPase A density near the active site. The position of this density at this resolution is consistent with the binding mode for extended substrates proposed from earlier model building studies with the native enzyme (1–3). In particular, the density runs near the expected positions of arginine-71, tyrosine-198, and phenylalanine-279. Consistent with a carboxy-terminal glycine, little density is found in the specificity pocket of the enzyme, where earlier studies had located the tyrosine ring of the glycyl-tyrosine dipeptide complex with CPase A. Attempts to define these interactions at higher resolution by using MIR and combined MIR–MR phase sets (27) in space group $P3_221$ have, unfortunately, produced maps with increased noise levels due to breakdown of the twofold axis approximation and the effects of twinning.

A fundamental assumption behind the low-resolution work is that the entire CPase A–PCI complex has pseudo $P3_221$ symmetry at low resolution. The MR results have established that the native molecule obeys this pseudosymmetry when considered as a rigid body. Small, local deviations could not be detected by this technique or at this resolution, although they are, in fact, anticipated to exist (28). There is no direct evidence for the symmetry of the inhibitor binding. The highest inhibitor electron density in the active site pocket is 80% of the maximum density present in the entire map (found in the α-helix containing residues 14–28), indicating it is unlikely that only one of the noncrystallographically related active sites is occupied. Outside of the binding pocket, however, the inhibitor density is more diffuse, and definite statements concerning the conservation of symmetry cannot be made.

At 5.5-Å resolution, it is not possible to trace the inhibitor peptide chain. The presence of three disulfide bonds in a total of 39 amino acids places tight constraints on the chain trace of PCI. We note that the size and disulfide pairings in a domain of wheat germ agglutinin (29) are very similar to those in PCI. This information may help while attempting to locate the disulfide bonds in the inhibitor. Determining the position of these sites should fix the general path of the peptide chain. Details of the interaction of the inhibitor with CPase A will require, however, completion of the high resolution structure.

We thank H. Wyckoff for advice on data collection and the National Institutes of Health (Grant GM 06920) for support.

1. Lipscomb, W. N., Hartsuck, J. A., Reeke, G. N., Quiocho, F. A., Bethge, P. H., Ludwig, M. L., Steitz, T. A., Muirhead, H. & Coppola, J. C. (1968) *Brookhaven Symp. Biol.* **21**, 24–90.
2. Quiocho, F. A. & Lipscomb, W. N. (1971) *Adv. Protein Chem.* **25**, 1–78.
3. Hartsuck, J. A. & Lipscomb, W. N. (1971) in *The Enzymes*, ed. Boyer, P. (Academic, New York), Vol. 3, pp. 1–56.
4. Abramowitz, N., Schechter, I. & Berger, A. (1967) *Biochem. Biophys. Res. Commun.* **29**, 862–867.
5. Ryan, C. A., Hass, G. M. & Kuhn, R. W. (1974) *J. Biol. Chem.* **249**, 5495–5499.
6. Hass, G. M., Nau, H., Biemann, K., Grahn, D. T., Ericsson, L. H. & Neurath, H. (1975) *Biochemistry* **14**, 1334–1342.
7. Leary, T. R., Grahn, D. T., Neurath, H. & Hass, G. M. (1979) *Biochemistry* **18**, 2252–2256.
8. Huber, R. & Bode, W. (1978) *Acc. Chem. Res.* **11**, 114–122.
9. Hass, G. M., Ako, H., Grahn, D. T. & Neurath, H. (1976) *Biochemistry* **15**, 93–100.
10. Laskowski, M. & Sealock, R. (1971) in *The Enzymes*, ed. Boyer, P. (Academic, New York), Vol. 3, pp. 376–473.
11. McPherson, A. (1976) *J. Biol. Chem.* **251**, 6300–6303.
12. Wyckoff, H. W., Doscher, M., Tsernoglu, D., Inagami, T., Johnson, L. N., Hardman, K. D., Allewell, N. M., Kelly, D. M. & Richards, F. M. (1967) *J. Mol. Biol.* **27**, 563–578.
13. North, A. C. T., Phillips, D. C. & Mathews, F. S. (1968) *Acta Crystallogr. Sect. A* **24**, 351–359.
14. Crawford, J. L. (1977) Dissertation (Harvard Univ., Cambridge, MA).
15. Lipscomb, W. N. & Jacobson, R. A. (1972) in *Physical Methods of Chemistry*, eds. Weissberger, A. & Rossiter, B. (Wiley, New York), Vol. 1, Part IIID, pp. 1–123.
16. Blundell, T. & Johnson, L. N. (1976) *Protein Crystallography* (Academic, New York).
17. Buerger, M. (1960) *Crystal Structure Analysis* (Wiley, New York), pp. 53–68.
18. Murray-Rust, P. (1973) *Acta Crystallogr. Sect. B* **29**, 2559–2566.
19. Britton, D. (1972) *Acta Crystallogr. Sect. A* **28**, 296–297.
20. Stanley, E. (1972) *J. Appl. Crystallogr.* **5**, 191–194.
21. Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1961) *Acta Crystallogr.* **14**, 1188–1195.
22. Warren, S. G., Edwards, B. F. P., Evans, D. R., Wiley, D. C. & Lipscomb, W. N. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 1117–1121.
23. Rossmann, M. G., ed. (1972) *The Molecular Replacement Method* (Gordon & Breach, New York).
24. Crowther, R. A. (1972) in *The Molecular Replacement Method*, ed. Rossmann, M. G. (Gordon & Breach, New York), pp. 173–178.
25. Crowther, R. A. & Blow, D. M. (1967) *Acta Crystallogr.* **23**, 544–548.
26. Blow, D. M. & Crick, F. H. C. (1959) *Acta Crystallogr.* **12**, 794–802.
27. Bricogne, G. (1976) *Acta Crystallogr. Sect. A* **32**, 832–847.
28. Tulinsky, A., Vandlen, R. L., Morimoto, C. N., Mani, N. V. & Wright, L. H. (1973) *Biochemistry* **12**, 4185–4200.
29. Wright, C. S. (1977) *J. Mol. Biol.* **111**, 439–457.