ACCOUNTS OF CHEMICAL RESEARCH

VOLUME 3 NUMBER 3 MARCH, 1970

Structure and Mechanism in the Enzymatic Activity of Carboxypeptidase A and Relations to Chemical Sequence

WILLIAM N. LIPSCOMB

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

Received August 13, 1969

Carboxypeptidase A (CPA) is, as its name implies, an enzyme which cleaves peptides at the carboxyl end of a polypeptide substrate. Its precursor or proenzyme occurs in the pancreas of species at least from the spiny dogfish to man. This proenzyme can be enzymatically activated to at least four closely related CPA's, which differ in the number of amino acids at the nitrogen terminus.

Physiologically, CPA functions in the intestine where it releases C-terminal amino acids as proteins are digested. However, rates of release of the different amino acids are quite varied.[1] For bovine CPA, the material of the studies described here, rates are high for release of Tyr,[2] Phe, Trp, Leu, Ile, Met, Thr, Gln, His, Ala, and Val, slow for Asn, Ser, and Lys, very slow for Gly, Asp, and Glu, and almost zero for Pro and Arg. Furthermore these rates depend in varying amount upon which amino acid residues immediately precede the C-terminal residue in the sequence of the substrate. The three-dimensional basis for this specificity and the detailed nature of probable mechanisms of cleavage of substrates by the bovine enzyme are the subjects of this Account.

The early history of carboxypeptidase goes back at least to Waldschmidt-Leitz and Purr,[3] who tabulated pH–activity data for the substrate chloroacetyl-L-tyrosine. Reasonable rates occurred from pH 5.6 to 9.0, and an optimal rate was found at just over pH 7. The purity of early preparations was questionable, but in 1935 Anson[4] first reported crystalline CPA.

A series of studies then occurred on the details of the specificity of CPA, as follows. (1) The peptide bond which is hydrolyzed (Figure 1) must be adjacent to a terminal free carboxyl group.[5,6] (2) The rate of hydrolysis is enhanced if the C-terminal residue of the substrate is aromatic or branched aliphatic.[7] (3) Dipeptides having a free amino group are hydrolyzed slowly, but if this group is blocked by N-acylation the hydrolysis is rapid.[6] (4) The carboxyl-terminal residue must be in the L configuration,[8–10] but peptides having C-terminal Gly and D-Ala[11] are split slowly. (5) Substitution of a methyl group (sarcosine) or a methylene group (proline) for the H atom of the susceptible peptide bond prohibits or greatly reduces hydrolysis rates.[7,12] (6) The rate of hydrolysis of N-acyl dipeptides is greatly decreased by substitution of sarcosine[13] or β-alanine[14] for the penultimate amino acid. (7) At least five C-terminal residues of the substrate influence[15] K_M and, to a lesser extent, k_{cat}. In the X-ray diffraction study to be described below an electron-density map at 2.0-Å resolution has been obtained for the complex of CPA with the dipeptide glycyl-L-tyrosine (Gly-Tyr). The relationship of this complex to the probable active binding mode for substrates and catalytic mechanisms will be examined briefly.

Carboxypeptidase A is the first metalloenzyme for which the high-resolution structure and sequence are known. The first suggestion[16] that CPA contains a metal ion at the active site was supported, in part, by cysteine inhibition, but the functional metal ion was later identified[17] as Zn^{2+}. Peptidase activity has been shown[18,19] for the Co^{2+}, Ni^{2+}, Zn^{2+}, Mn^{2+}, and Fe^{2+} CPA's, but both Cu^{2+} CPA and apoCPA (which is defined by removal of Zn^{2+}) are inactive toward all substrates.

Esters are known[20] to be cleaved by CPA, and the

(1) For example, see R. P. Ambler, *Methods Enzymol.*, **11**, 155, 445 (1967).
(2) Refer to *Biochemistry*, **5**, 1465 (1966), for the standard abbreviations for amino acids.
(3) E. Waldschmidt-Leitz and A. Purr, *Ber. Deut. Chem. Ges.*, **62**, 2217 (1929).
(4) M. L. Anson, *Science*, **81**, 467 (1935); *J. Gen. Physiol.*, **20**, 663 (1937).
(5) E. Waldschmidt-Leitz, *Physiol. Rev.*, **11**, 358 (1934).
(6) K. Hoffman and M. Bergmann, *J. Biol. Chem.*, **134**, 225 (1940).
(7) M. Stahmann, J. S. Fruton, and M. Bergmann, *ibid.*, **164**, 753 (1946).
(8) M. Bergmann and J. S. Fruton, *ibid.*, **117**, 189 (1937).
(9) H. T. Hanson and E. L. Smith, *ibid.*, **179**, 815 (1949).
(10) C. A. Dekker, S. P. Taylor, Jr., and J. S. Fruton, *ibid.*, **180**, 155 (1949).
(11) I. Schechter and A. Berger, *Biochemistry*, **5**, 3371 (1966).
(12) E. L. Smith, *J. Biol. Chem.*, **175**, 39 (1948).
(13) J. E. Snoke and H. Neurath, *ibid.*, **181**, 789 (1949).
(14) H. T. Hanson and E. L. Smith, *ibid.*, **175**, 833 (1948).
(15) N. Abramowitz, I. Schechter, and A. Berger, *Biochem. Biophys. Res. Commun.*, **29**, 862 (1967).
(16) E. L. Smith and H. T. Hanson, *J. Biol. Chem.*, **179**, 802 (1949).
(17) B. L. Vallee and H. Neurath, *J. Am. Chem. Soc.*, **76**, 5006 (1954).
(18) J. E. Coleman and B. L. Vallee, *J. Biol. Chem.*, **235**, 390 (1960).
(19) J. E. Coleman and B. L. Vallee, *ibid.*, **236**, 2244 (1961).
(20) J. E. Snoke, G. W. Schwert, and H. Neurath, *ibid.*, **175**, 7 (1948).

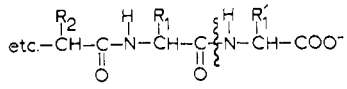

Figure 1. Position of cleavage, shown by wavy line, for peptide substrates of CPA.

substitution of Hg^{2+} or Cd^{2+} for Zn^{2+} retains the esterase activity but terminates peptidase activity.[19] A critique[21] of esterase and peptidase mechanisms appeared shortly before our structure study was completed. Inasmuch as no complex of CPA with an ester has been stable at 4° for the few hours required for X-ray study, our brief comments[22] on esters are based upon model building in conjunction with chemical evidence.

The four forms of CPA arise from enzymatic release of an N-terminal fragment of about 64 residues.[23] The X-ray study was made on CPA_α (predominantly N-terminal Ala, 307 residues[24]) prepared from pancreatic juice by procedures described elsewhere.[25] Earlier studies[26] established the differences in the N-terminal regions of these four forms; the other three are CPA_β (N-terminal Ser, 305 residues) and both CPA_γ and CPA_δ (N-terminal Asn, 300 residues). All forms are active, and the last two differ in their solubilities[27] and in reversibility of Zn^{2+} removal.[28]

Partial chemical sequence data published before computation of the X-ray diffraction maps at 2.8- and 2.0-Å resolution are an N-terminal fragment of 22 residues[26] (for CPA_α), a C-terminal fragment of 7 residues,[29] an "active-site" cysteinyl sequence of 14 residues,[30,31] and a "nonessential" cysteinyl sequence of 7 residues.[30,31] The X-ray study[32] indicates that these two cysteinyl sequences are joined by a disulfide bridge. The Zn^{2+} binding ligands, indicated from chemical evidence[33,34] to be a thiol group and the α-amino group (Asn) of CPA_γ, are shown below on the basis of combined X-ray and chemical sequence results to be His-69, Glu-72, and His-196.

One feature of substrate binding proposed earlier is in accord with the X-ray results, described below, on the complex of CPA_α with Gly-Tyr. Binding of the susceptible peptide bond's carbonyl oxygen to the metal ion was first suggested by Smith,[35] but his proposal that the substrate's terminal carboxyl group is also on the metal differs from the X-ray work. In a later more detailed proposal,[36,37] this carbonyl oxygen is also placed on Zn^{2+}, but the binding of the susceptible peptide bond's N atom to Zn^{2+} does not agree with the complex in the X-ray study.

Potential catalytic groups have also received earlier attention, especially in studies of the effect of chemical modification of the activities of CPA. The involvement of at least one tyrosine in hydrolysis of peptide substrates is well documented,[38] and the pH–rate profile and modifications have been used to implicate histidine[38,39] in the mechanism. Certainly one must expect a base B to attack the carbonyl, and an acid HA to donate a proton to the NH of the scissle peptide bond, and in general terms B and HA must be among the protein ligands or water. A base B, required for peptide hydrolysis but inhibitory for ester hydrolysis, was suggested[36,37] as that group which is iodinated, photooxidized, and protected from acetylation. These procedures modify tyrosine in CPA, although photooxidation may modify histidine. The X-ray study,[32] on the other hand, has cast a tyrosine residue (Tyr-248) in the role of a proton donor to the NH of the susceptible peptide bond and has identified group B as the carboxyl group of a glutamic acid residue (Glu-270). No experiments so far have distinguished whether B acts to form an acyl-enzyme intermediate (so far undiscovered) or to promote the attack by the oxygen of a water molecule at the substrate's carbonyl carbon.

Structure of CPA_α

The development of procedures,[25] with the help of Hirs, for isolation of pure CPA_α which gave large single crystals and of methods[25] for dialysis of heavy atom salts into a few crystals led to success in obtaining heavy atom derivatives which were isomorphous. The X-ray study was carried out successively at 6-, 2.8-, and 2.0-Å resolution by the method of isomorphous replacement.[40]

The 6-Å resolution study[25] yielded the molecular shape (about 50 × 42 × 38 Å), the absolute configuration from anomalous scattering data, an estimate of 25% of helix, the location of the Zn atom, and the associated active-site consisting of a pocket and groove for the substrate, the location of the C terminus, and

(21) F. W. Carson and E. T. Kaiser, *J. Am. Chem. Soc.*, **88**, 1212 (1966).
(22) W. N. Lipscomb, J. A. Hartsuck, G. N. Reeke, Jr., F. A. Quiocho, P. H. Bethge, M. L. Ludwig, T. A. Steitz, H. Muirhead, and J. C. Coppola, *Brookhaven Symp. Biol.*, **21**, 24 (1968).
(23) J. H. Freisheim, K. A. Walsh, and H. Neurath, *Biochemistry*, **6**, 3010, 3020 (1967).
(24) Both the number of amino acid residues and the detailed number identification of each in CPA are from the X-ray diffraction study.
(25) W. N. Lipscomb, J. C. Coppola, J. A. Hartsuck, M. L. Ludwig, H. Muirhead, J. Searl, and T. A. Steitz, *J. Mol. Biol.*, **19**, 423 (1966).
(26) K. S. V. Sampath Kumar, J. B. Clegg, and K. A. Walsh, *Biochemistry*, **3**, 1728 (1964).
(27) B. J. Allan, P. J. Keller, and H. Neurath, *ibid.*, **3**, 40 (1964).
(28) B. L. Vallee, T. L. Coombs, and F. L. Hoch, *J. Biol. Chem.*, **235**, PC45 (1960).
(29) J.-P. Bargetzi, E. O. P. Thompson, K. S. V. Sampath Kumar, K. A. Walsh, and H. Neurath, *ibid.*, **239**, 3767 (1964).
(30) K. S. V. Sampath Kumar, K. A. Walsh, J.-P. Bargetzi, and H. Neurath, in "Protein Structure and Crystallography," G. N. Ramachandran, Ed., Academic Press, New York, N. Y., 1963, p 319.
(31) H. Neurath, *Federation Proc.*, **23**, 1 (1964).
(32) G. N. Reeke, J. A. Hartsuck, M. L. Ludwig, F. A. Quiocho, T. A. Steitz, and W. N. Lipscomb, *Proc. Natl. Acad. Sci. U. S.*, **58**, 2220 (1967).
(33) B. L. Vallee, R. J. P. Williams, and J. E. Coleman, *Nature*, **190**, 633 (1961).
(34) T. L. Coombs, Y. Omote, and B. L. Vallee, *Biochemistry*, **3**, 653 (1964).
(35) E. L. Smith, *Proc. Natl. Acad. Sci. U. S.*, **35**, 80 (1949).
(36) B. L. Vallee, J. F. Riordan, and J. E. Coleman, *ibid.*, **49**, 109 (1963).
(37) B. L. Vallee, *Federation Proc.*, **23**, 8 (1964).
(38) M. Sokolovsky and B. L. Vallee, *Biochemistry*, **6**, 700 (1967).
(39) H. Neurath and G. W. Schwert, *Chem. Rev.*, **36**, 69 (1950).
(40) C. Bokhoven, J. C. Schoone, and J. M. Bijvoet, *Proc. Acad. Sci. Amsterdam*, **52**, 120 (1949); *Acta Cryst.*, **4**, 275 (1951).

the location of the N terminus. Previous estimates of the helix content were 80[41] and 70%.[42] Optical data examined in more detail, but after the low-resolution structure was known, have yielded values of 20–40%[25] and 21–23%.[43] The N terminus was located some 25 Å from Zn^{2+} by examination of a difference of electron density between CPA_α and CPA_δ, which are isomorphous and which differ by seven amino acids at the N terminus. The Zn^{2+} ion was located from a difference map between CPA_α and $apoCPA_\alpha$. Neither the detailed course of the polypeptide chain nor the large amount of pleated sheet (β) structure could be seen in this 6-Å map.

At 2.8-Å resolution, the excellent resolution of the carbonyl groups in the polypeptide chain and the detailed fitting at the atomic level of the N-terminal sequence of 22 residues led us to the conclusion[44] that our map of August 1966 would have been sufficient for a complete atomic structure if the chemical sequence were known. The few additional turns of helix raised the estimate of helix content to 30% and revealed some α_{II} helix.[45] The large amount of β structure, both parallel and antiparallel, was noted, and a complete trace of the polypeptide chain was made which upon later examination of the map at 2.0 Å was shown to be correct.

At 2.0-Å resolution[22,32] (Figure 2) the number of amino acid residues was established with certainty as 307, and the two cysteinyl sequence fragments were located and shown to be joined by a disulfide bond between Cys-138 and Cys-161. In further support, it was shown that neither sequence fragment could be located near the Zn^{2+}, nor anywhere else except in this region of the disulfide. Also the covalent bond distance, the dihedral angle of 100° between CH_2–S–S and S–S–CH_2 planes, and the integrated electron density of 47 (on a scale for which sulfurs of the three methionine residues were 23, 25, and 14) further supported this disulfide bond. No other cysteinyl residue could be found in the map.

The β structure is a pleated sheet (Figure 3) twisted by 120° from the bottom to the top of the molecule, containing both parallel and antiparallel pairs of extended chains. About 45 residues (15% of the 307) form hydrogen bonds within the sheet, but because of hydrophobic packing of side chains the general ordered structure extends somewhat beyond the hydrogen-bonded region. This twisted sheet separates the molecule into two main regions, one in which most of the helices occur in contact with solution and the other in which there are few hydrogen bonds (about ten) and the disulfide bond. Most of the conformational

Figure 2. Polypeptide segments between C_α atoms (circles) in CPA_α. The Zn^{2+} ion is near the center, where positions of the three protein ligands are shown by arrows. The N terminus is at the bottom, the disulfide bond at the right, and the C terminus at the left.

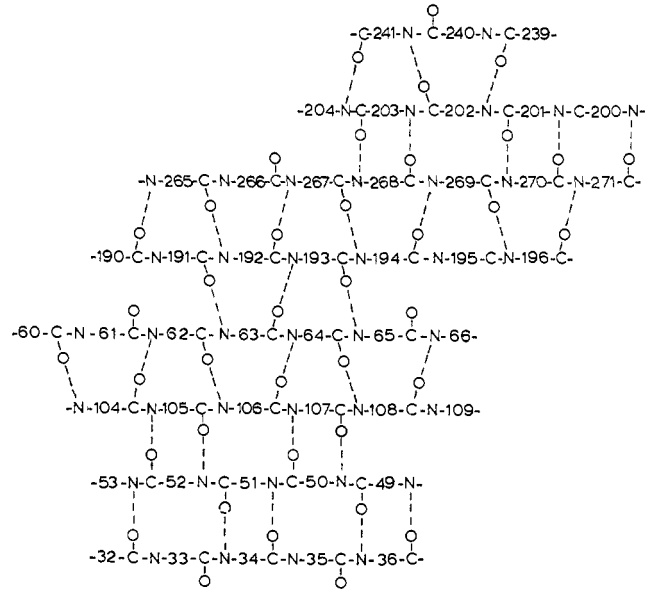

Figure 3. The residues forming hydrogen bonds (dashed lines) in the β structure. There are four adjacent pairs of parallel and three adjacent pairs of antiparallel extended chains in this pleated sheet, which in the three-dimensional structure is twisted by about 120° as one proceeds from bottom to top.

changes described below occur in this latter tortuous region.

Helix regions are 14–28, 72–80, 82–88, 94–103, 112–122, 173–187, 215–231, 254–262, and 285–306. Very little is nearly perfect α helix, and some regions do not fit any previous descriptions.

The conformation of proCPA is unknown, but some of its substrate binding characteristics resemble those of CPA.[46] If the CPA portion of procarboxypeptidase A is nearly in its final conformation in the zymogen, there are some problems of folding CPA from the N

(41) K. Imahori, H. Fujioka, and T. Ando, *J. Biochem.* (Tokyo), **52**, 167 (1962).
(42) J. A. Rupley and H. Neurath, *J. Biol. Chem.*, **235**, 609 (1960).
(43) F. A. Quiocho, W. H. Bishop, and F. M. Richards, *Proc. Natl. Acad. Sci. U. S.*, **57**, 525 (1967).
(44) W. N. Lipscomb, "Structural Chemistry and Molecular Biology," A. Rich and N. Davidson, Ed., W. H. Freeman, San Francisco, Calif., 1968, p 38.
(45) G. Némethy, D. C. Phillips, S. J. Leach, and H. A. Scheraga, *Nature*, **214**, 363 (1967).
(46) R. Piras and B. L. Vallee, *Biochemistry*, **6**, 348 (1967).

terminus. The sequential order of chains (Figure 3) numbered from N to C is 1, 2, 4, 3, 5, 8, 6, and 7 as one goes from the bottom to the top of the pleated sheet. Final hydrogen bonds of chains 3 (60–66) and 6 (200–204) cannot be formed until chains 4 (104–109) and 8 (265–271) are in place. Also the final positioning of residues 1–103 must occur after chain 4 (104–109) has passed between helices 14–28 and 72–88 and between Phe-52 and Phe-86. Chain 249–254 must pass between Gly-150 and Tyr-208, but these residues have a C_α–C_α distance of only 5.5 Å in the final structure. Also, the disulfide bond must be formed after residues 163–170 pass through the disulfide loop 138–161. In summary, some preliminary folding of at least two local regions before final self-assembly seems reasonable, but the detailed process remains obscure.

Some preference is observed for Arg, Gly, and Tyr in random coil, for Ala and Val in helix, for Asx, Glx, Gly, and Ile for ends of helix, and for Leu and Phe for β structure. However, it is to be recalled that the β structure is almost completely internal, and that nearly all helix is on the outer boundary of the molecule. Four of the ten prolines (94, 113, 214, and 288) terminate helices at the amino end, three (46, 60, and 205) are at ends of extended chains, and three (30, 160, and 282) are in random coil.

Finally, the probable occurrence of *cis* peptide between Ser 197 and Tyr 198 is an unusual feature of the configuration of the polypeptide chain.

Binding of Peptide Substrate to CPA_α

In X-ray diffraction studies[47] at 6-Å resolution, some large conformational changes were observed when CPA binds the poor substrate glycyltyrosine, the product phenylalanine, or the inhibitor *p*-iodo-β-phenylpropionate. Conformational changes were not observed at 6-Å resolution when lysyltyrosylamide was bound, but occurred[47] most clearly when hippurylphenylalanine was bound to acetyl CPA, a modified form of CPA showing no peptidase activity.[48]

All of these derivatives were prepared by diffusion of the substrate, product, or inhibitor into crystalline CPA, in which the region of the active site is open to the aqueous part of the crystal and free of intermolecular contacts.[22] The specific activity of CPA crystals, relative to that in solution, is reduced by a factor of several hundred.[49] While the crystalline material is enzymatically active, one must realize that an X-ray diffraction study of the complex of an enzyme with a poor substrate yields results for a static and possibly nonproductively bound complex. The study of one such complex, of CPA_α with Gly-Tyr, was carried to 2.0-Å resolution.[22,50]

The single binding mode of Gly-Tyr in CPA can be described by four interactions (Figure 4a). First, the C-terminal residue of the substrate inserts into the pocket of CPA, thereby displacing several water molecules. No specific interactions of the substrate's C-terminal side chain with residues of the protein are clearly dominant. Second, the essential C-terminal carboxyl group of the substrate forms a salt link with the guanidinium group of Arg-145. Third, the carbonyl oxygen of the substrate's susceptible peptide bond replaces an observable water molecule as the fourth ligand to Zn^{2+}. Fourth, in an interaction possible only with dipeptides, Glu-270 binds through a water molecule to the free amino (or ammonium) group of Gly-Tyr. It is possibly this last interaction which is responsible for the unusually slow rate of cleavage of Gly-Tyr by CPA. The first three of these interactions (Figure 4a) correlate well with model building of longer substrates as compared to binding and cleavage of these larger substrates. Hence, we believe that these three are characteristic of a productive complex, but that the fourth is not.

CPA undergoes several conformational changes when Gly-Tyr is bound, as shown by negative density at the native conformation and positive density in the modified conformation in the difference map. First, the guanidinium group of Arg-145 moves approximately 2 Å by rotation about the C_β–C_γ bond. Second, the carboxyl group of Glu-270 moves away from Zn^{2+} (Figures 5–7) about 2 Å by rotations about C_α–C_β and C_β–C_γ. Third, and most striking, the phenolic OH of Tyr-248 moves about 12 Å so that it moves to within roughly 3 Å of the NH group of the scissile peptide bond. This motion involves a rotation by 120° about the C_α–C_β bond and some limited motion of the peptide backbone. A system of four hydrogen bonds between Arg-145 and Tyr-248, involving intermediate groups and one water molecule, is broken by these conformational changes. Indeed, a coordinated movement of Arg-145 and Try-248 occurs as this Tyr approaches the substrate. Adaptation of a protein to other molecules has been suggested before,[51–53] but the conformational changes[54] in CPA are a clear example of the "induced-fit" theory.[55]

The probable productive mode of binding of longer substrates (Figure 8) has been extrapolated from these results on Gly-Tyr with the aid of results[15] on K_M and k_{cat} for substrates of various lengths. Correlation of these results with the three-dimensional model of CPA further clarifies the functions of a secondary binding

(47) T. A. Steitz, M. L. Ludwig, F. A. Quiocho, and W. N. Lipscomb, *J. Biol. Chem.*, **242**, 4662 (1967).
(48) J. F. Riordan and B. L. Vallee, *Biochemistry*, **2**, 1460 (1963).
(49) F. A. Quiocho, W. H. Bishop, and F. M. Richards, *Proc. Natl. Acad. Sci. U. S.*, **57**, 525 (1967).
(50) W. N. Lipscomb, G. N. Reeke, Jr., J. A. Hartsuck, F. A. Quiocho, and P. H. Bethge, *Phil. Trans.*, in press.
(51) F. Karush, *J. Am. Chem. Soc.*, **72**, 2705 (1950).
(52) R. Lumry and H. Eyring, *J. Phys. Chem.*, **58**, 110 (1954).
(53) F. Vaslow, *Compt. Rend. Trav. Lab. Carlsberg*, **31**, 29 (1958).
(54) Spectral changes have been noted when β-phenylpropionate binds to CPA [H. Fujioka and K. Imahori, *J. Biol. Chem.*, **237**, 2804 (1962)] or to nitro-CPA [J. F. Riordan, M. Sokolovsky, and B. L. Vallee, *Biochemistry*, **6**, 358 (1967)]. Before the structure of CPA became known, it was not possible to distinguish in these studies between a conformational change or association of β-phenylpropionate with the relevant aromatic groups of CPA.
(55) D. E. Koshland, Jr., *Cold Spring Harbor Symp. Quant. Biol.*, **28**, 473 (1963).

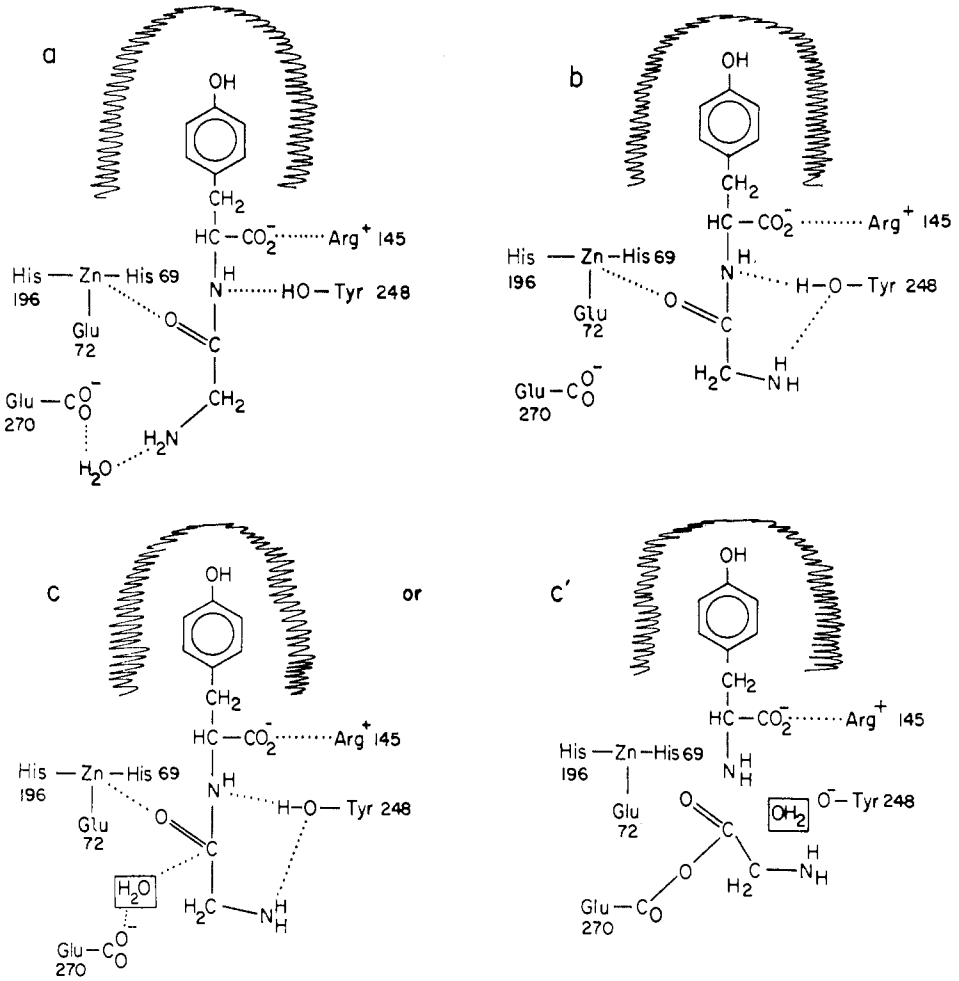

Figure 4. (a) Specific binding interactions of Gly-Tyr as deduced from the difference electron density. (b) Probable productive mode of binding of Gly-Tyr at the start of catalysis. (c) General base path in which Glu-270 promotes attack of a lone pair from H_2O at the carbonyl carbon, probably preceded by or concurrent with proton transfer from Tyr-248 to the NH group. (c') Anhydride intermediate pathway in which H_2O later attacks the acylenzyme intermediate. Further studies are required in order to resolve the ambiguity of mechanism.

site involving Arg-71, Tyr-198, and Phe-279 (as described below, the chemical sequence resolved the ambiguity in the two choices of the X-ray study for residue 279). However, none of these residues can move close enough to the susceptible peptide bond of the substrate to participate in the catalytic steps. In addition, the chemical results[13,14] which show that the presence of H on the NH group of the second peptide bond is important for rapid hydrolysis are interpreted by us as indicative of an additional hydrogen bond which stabilizes the final conformation of Tyr-248 in the presence of substrate (Figure 8).

Catalysis

The only parts of CPA which are near enough to the peptide bond to be directly involved in the catalysis are Glu-270, Zn^{2+}, and Tyr-248. The only other group of the protein within 3 Å of a functional group of the substrate is Arg-145, which is a binding group in the mechanisms proposed in the X-ray study.[22,32,50] Any mechanism must supply a proton to the NH of the susceptible peptide bond and either permit or promote attack of H_2O or some other general base at the carbon atom of the carbonyl group of the scissile peptide bond. Thus the three-dimensional structure severely limits possible mechanisms. The pH–rate profile[3] suggests a range of 5.6 to 9.0 for reasonable rates, with an optimum pH of just over 7. Ordinarily, one would expect Glu-270 to have a negative carboxyl ion, and Tyr-248 to retain an intact OH group in native CPA. Thus it is reasonable that the tyrosine residue, which is essential for peptidase activity but not for the esterase activity,[36] be assigned the role of a proton donor.[32] It is probable that Glu-270 either promotes the attack of a lone pair of electrons of water on the carbon of the substrate's susceptible carbonyl carbon or combines directly with this carbon to form an anhydride which is later split upon attack by a water molecule (Figure 4).

No acyl-enzyme intermediate has been isolated, and no transpeptidation[56] or transesterification[57] has been observed so far. Single turnover experiments with ^{18}O

(56) L. M. Ginodman, N. I. Mal'tsev, and V. N. Orekhovich, *Biokhimiya*, 31, 1073 (1966).
(57) P. L. Hall and E. T. Kaiser, *Biochem. Biophys. Res. Commun.*, 29, 205 (1967).

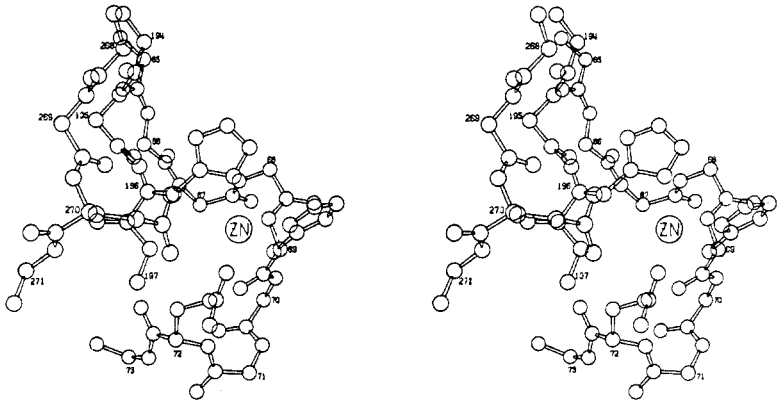

Figure 5. Stereoview of Zn^{2+} and its ligands, His-69, Glu-72, and His-196. The position of Glu-270 (closest to the reader) before its conformational change of about 2 Å is also shown. (Stereoviewers may be obtained, for example, from Wards Natural Science Establishment, Inc., Rochester, N. Y., Model 25W,2951.)

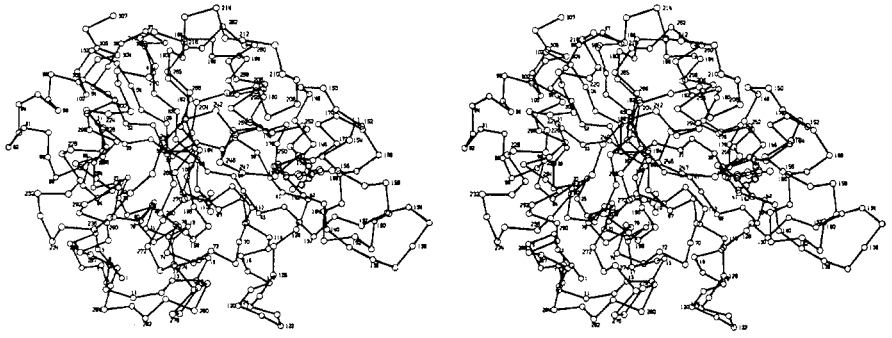

Figure 6. Stereoview of the polypeptide backbone, showing the conformations of Arg-145, Tyr-248, and Glu-270 before substrate enters the cavity.

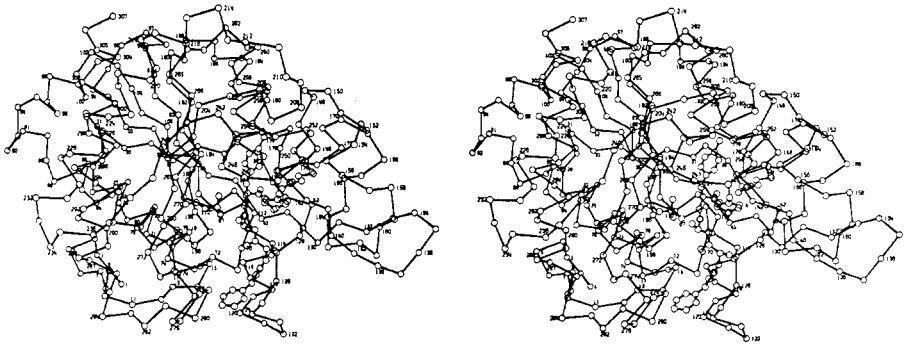

Figure 7. Stereoview of the polypeptide chain showing the substrate carbobenzoxy-Ala-Ala-Tyr extending into the cavity and the positions of Arg-145, Tyr-248, and Glu-270 after their conformational changes.

might prove valuable, but the presently available multiple turnover ^{18}O tracer experiments[56,58] do not distinguish these mechanisms. Deuterium tracer studies of CPA activity yield $k_{H_2O}/k_{D_2O} = 1.22$ for carbobenzoxy-Gly-Phe[59,60] and 1.1 for hippuryl-Phe,[61] while this ratio is very close to 2 for O-(*trans*-cinnamoyl)-L-β-phenyllactate,[62] and for O-(N-benzoylglycyl)-L-β-phenyllactate.[63] At least, these results suggest that a proton transfer is more important in the rate-determining step for these esters than for these peptides. Further studies of these types, fast kinetics, and comparative studies of a number of closely related substrates will probably be helpful in making the needed distinctions among mechanisms and for producing an ordered sequence of steps in the catalytic process.

CPA is known to show a variety of kinetic anomalies, such as activation and inhibition by certain substrates or their products produced upon hydrolysis. The three-dimensional structure yields specific models[22] for some of these anomalies and shows why the preponderance of these kinetic abberations has been shown by

(58) A. Kowalsky and P. D. Boyer, *J. Biol. Chem.*, **235**, 604 (1960).
(59) R. Lumry and E. L. Smith, *Discussions Faraday Soc.*, **20**, 105 (1955).
(60) R. Lumry, E. L. Smith, and R. R. Glantz, *J. Am. Chem. Soc.*, **73**, 4330 (1951).
(61) F. A. Quiocho, unpublished results.
(62) B. L. Kaiser and E. T. Kaiser, *Proc. Natl. Acad. Sci. U. S.*, **64**, 36 (1969).
(63) W. N. Lipscomb, J. A. Hartsuck, F. A. Quiocho, and G. N. Reeke, Jr., *ibid.*, **64**, 28 (1969).

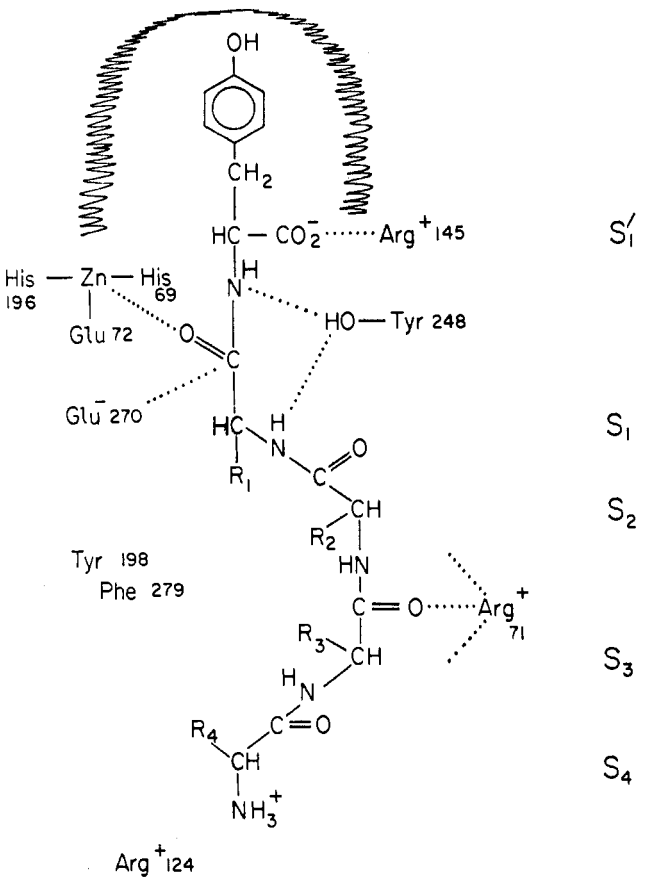

Figure 8. Further interactions, derived from experimental results on Gly-Tyr binding, model building, and results of binding and rate studies for longer substrates. The secondary (recognition) site Arg-71, Tyr-198, and Phe-279 is probably the region of enzyme structure which is associated with the substrate anomalies so prevalent for aromatic acyl dipeptides.

dipeptide or analogous ester substrates (or products) which have aromatic N-acyl groups. Although these proposals are based entirely on model building, they differ from all previous general models in that they are at the atomic level and hopefully subject to detailed structure study.

Relation between Structure and Sequence Studies

Although in theory the three-dimensional X-ray structure should also yield a sequence, in practice there are serious limitations owing to intrinsic disorder in the side chains and sometimes even in the polypeptide backbone. Also, several pairs of side chains are very similar in shape. Thus, it is unlikely that a sequence can be established by X-ray methods alone at any resolution. Chemical methods are therefore required in order to obtain reliable sequence. However, few chemical sequences (insulin!) have not required some later revision, and we therefore indicate here some roles that X-ray studies can play in facilitating sequence studies in collaboration with research groups using chemical approaches.

In myoglobin,[64,65] a partial determination of sequence and correlation with chemical results[66] on some 70% of the structure led to the conclusion[67] that about two-thirds of the residues could be identified by X-ray methods with some assurance. Distinctions were particularly difficult between Thr and Val, Glu and Gln, Asp and Asn, Leu and Asx, and His and Phe. The X-ray study did indicate that there were 151 residues. The occurrence of some 118 of these residues in α-helical regions played a role in placement of C_α atoms and hence in estimation of lengths and positions of forking (if present) of side chains.

In lysozyme,[68] there were five discrepancies between the two research groups[69,70] studying chemical sequence. The X-ray map indicated Val-92-Asn-93, Thr-40-Gln-41-Ala-42, and Ile-58-Asn-59 (this last a reversal[71] of the previous X-ray assignment[68]) for three of these regions, but gave no distinction between Asp and Asn for residues 65 and 66.

In papain, X-ray results[72] indicate that some large sections of the chemical sequence[73] need to be transposed, and that 13 residues (including two Ile residues assigned chemically as 29 and 30) must be inserted between Phe-28 and Arg-31. The transposition is the insertion of the peptide Thr-138 to Tyr-176 into the position between these 13 residues and Arg-31. The probable number of residues in papain is thus increased from 200 to 211. Also the sequence 160–164 given chemically as Asp-Tyr-Gly-Pro-Gly is probably, from the electron density map, Asp-Gly-Gly-Tyr-Pro. A more detailed study is promised.

Even at the earliest stages of X-ray work (unit cell, site symmetry, and molecular symmetry) and of sequence work (end group and later sequence analysis) there is a very close relationship between these techniques. An example is the conclusion[74,75] that aspartate transcarbamylase contains six regulatory and six catalytic subunits, instead of four of each of these subunits.

In carboxypeptidase A, the X-ray results indicate that the residues of interest are (a) the binding and catalytic groups 145 and 270 for peptides, (b) the additional catalytic group 248 for peptides, (c) the protein to Zn binding ligands 69, 72, and 196, (d) the substrate anomaly or recognition residues 71, 198, 297, and possibly 127, (e) the disulfide bond between 138 and 161, and (f) the carboxyl terminus 307 and the nearby residue 265. In addition, the numbering of the residues found

(64) J. C. Kendrew, R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips, and V. C. Shore, *Nature*, **185**, 422 (1960).

(65) J. C. Kendrew, H. C. Watson, B. E. Strandberg, R. E. Dickerson, D. C. Phillips, and V. C. Shore, *ibid.*, **190**, 668 (1961).

(66) A. B. Edmundson and C. H. W. Hirs, *ibid.*, **190**, 663 (1961).

(67) J. C. Kendrew, *Science*, **139**, 1259 (1963).

(68) C. C. F. Blake, D. F. Koenig, G. A., Mair, A. C. T. North, D. C. Phillips, and V. R. Sarma, *Nature*, **206**, 757 (1965).

(69) J. Jollès, J. Jauregui-Adell and P. Jollès, *Biochim. Biophys. Acta*, **78**, 68 (1963); *C. R. Acad. Sci. Paris*, **258**, 3926 (1964).

(70) R. E. Canfield, *J. Biol. Chem.*, **238**, 2698 (1963); R. E. Canfield and A. K. Liu, *ibid.*, **240**, 1997 (1965).

(71) C. C. F. Blake, G. A. Mair, A. C. T. North, D. C. Phillips, and V. R. Sarma, *Proc. Roy. Soc.* (London), **B167**, 365 (1967).

(72) J. Drenth, J. N. Jansonius, R. Kockoek, H. M. Swen, and B. G. Wolthers, *Nature*, **218**, 929 (1968).

(73) A. Light, R. Frater, J. R. Kimmel, and E. L. Smith, *Proc. Natl. Acad. Sci. U. S.*, **52**, 1276 (1967).

(74) D. C. Wiley and W. N. Lipscomb, *Nature*, **218**, 1119 (1968).

(75) K. Weber, *ibid.*, **218**, 1118 (1968).

from the sequence study to be replacement residues[76] of the two alleomorphs is, from early placement of peptide fragments in the X-ray study, 179, 228, and 305. The subsequent complete chemical sequence shows that all of the above residues were correctly numbered and that most were correctly identified in the X-ray investigation.[22,32]

The relationships of the X-ray and chemical sequence studies, with reference to these residues, occurred in five stages. (1) The four fragments published before the X-ray map was available were numbered in the X-ray study as 1–22 (N terminus[26]), 138–144 ("atypical Cys"[30,31]), 152–165 ("active site Cys"[30,31]), and 300–307 (C terminal,[29] including Ile-300). These residues include 138, 161, and 305 in the list above. (2) The tetrapeptide[77] associated with β-phenylpropionate inhibition was located in the X-ray map at 247–250 in July 1967. This fragment contains 248, and neighboring residues of some interest.[22] (3) The fragment 22–103 (October 1967[78,79]) contains residues 69, 72, and 71. (4) Sequence fragments containing residues 79 and 228 were received in January 1968.[78] (5) The remaining 93 residues[80] containing 145, 270, 248, 196, 198, 297, and 127 were communicated to us on June 23, 1969.[78] The complete X-ray sequence was communicated to the chemical sequence group (in Seattle) in 1967 and subsequently as revisions were made.

The specific binding and catalytic residues (for both peptides and esters) Arg-145 and Glu-270 were identified with no reasonable alternatives in the X-ray methods. The clear appearance of these residues both before and after conformational changes induced upon binding of CPA with Gly-Tyr confirmed Arg-145 and eliminated[22] two less likely alternatives for Glu-270. The clearly resolved contours of Tyr-248 established its identity uniquely also, and this residue was placed in the electron density when we located the peptide[77] associated with β-phenylpropionate inhibition.[81] The chemical evidence[38,82] for tyrosine involvement in peptide hydrolysis is very strong. Finally, the three-dimensional structure showed that 248 is the only Tyr which can approach the substrate's susceptible peptide bond. Thus, a case can be made that at 2.0-Å resolution the specific substrate binding and catalytic residues can be identified and that mechanisms can be elucidated, within certain ambiguities,[22] by X-ray methods before completion of the chemical sequence. At least, no change is required in the mechanistic conclusions of the X-ray study[22,32] as a result of the chemical sequence results.

The protein ligands to Zn^{2+} illustrate well the relationships between X-ray and sequence studies. Following X-ray identification of these ligands as 69, 72, and 196, residue 69 was established as His from the electron density map. However, in the tentative chemical sequence of residues 23–103, received by us in October 1967 from Neurath, residue 69 was Ile. Reversal of the identities 68 and 69 to yield Ile-68 and His-69 was indicated[22] by the electron density and was later confirmed in further study of the chemical sequence.[79] Residue 72 was readily identified as Glx from the X-ray results and established as Glu from the chemical sequence. Residue 196 was identified ambiguously and incorrectly as Glx or Lys in the X-ray study, but some preference for Lys was expressed[22] on the basis of the then available reaction[46] of dinitrofluorobenzene with Zn-free acetylcarboxypeptidase as compared with native acetylcarboxypeptidase. The chemical sequence study established that residue 196 is His, and thus there are two N ligands and one O ligand from the protein to Zn^{2+}. Divalent metal ions in the active site of carboxypeptidase have binding constants which are consistent with bonding to two ligands[17] if one is S and one is N, but which are also consistent[22] with three ligands if two are N and one is O. In summary, a combination of X-ray and chemical studies shows that His-69, Glu-72, and His-196 are the ligands from the protein to Zn.

The secondary binding (or recognition) site was identified as Arg-71, Tyr-198, and His(Phe)-279 in the X-ray study. The recent chemical sequence[80] resolves the ambiguity as Phe-279. An additional residue, 127, is 5 to 6 Å away from the susceptible peptide bond and hence cannot[22] be favored as an alternative for a catalytic residue. Its chemical identity[80] as Arg-127, rather than the X-ray identity as Glx-127, and its location in three dimensions, suggest that one function is to form a temporary binding site intermediate between Arg-71 and Arg-145 for the C-terminal carboxyl group of the substrate as it moves into the position from initial recognition (Arg-71) to the final position (Arg-145) for catalysis. Toward the end of the cleavage reaction, Arg-127 may salt-link to the new carboxyl group which is formed when a substrate is hydrolyzed.

The X-ray identification of a cystine between Cys-138 and Cys-161 has recently been confirmed by the sequence study,[80] but the sulfur chemistry needs further elucidation.

The side chain nearest the carboxyl terminus of CPA is 265 from the X-ray study, in which it was incorrectly identified as Arg. The chemical sequence[80] shows that this residue is Tyr-265. The shape of this side chain resembles Tyr in a new electron density map in which

(76) P. H. Pétra, R. A. Bradshaw, K. A. Walsh, and H. Neurath, *Biochemistry*, **8**, 2762 (1969).

(77) O. A. Roholt and D. Pressmann, *Proc. Natl. Acad. Sci. U. S.*, **58**, 280 (1967).

(78) H. Neurath, private communication.

(79) H. Neurath, R. A. Bradshaw, L. H. Ericsson, D. A. Babin, P. H. Pétra, and K. A. Walsh, *Brookhaven Symp. Biol.*, **21**, 1 (1968).

(80) R. A. Bradshaw, L. H. Ericsson, K. A. Walsh, and H. Neurath, *Proc. Natl. Acad. Sci. U. S.*, **63**, 1389 (1969). Clarifications of this reference are (a) the molecular weight of CPA_α (Val 305) is 34,472 and (b) residue 279 which is 7 Å from the active site and residue 127 which is 5 to 6 Å from the active site were regarded in the X-ray study[22] as unlikely or unfavorable candidates for alternative catalytic residues.

(81) The X-ray study indicates that *p*-iodo-β-phenylpropionate binds at four places in the CPA molecular crystal. This result suggests that protection experiments should be made in conjunction with X-ray results to be sure that the peptide fragment is actually near the active site.

(82) B. L. Vallee and J. F. Riordan, *Brookhaven Symp. Biol.*, **21**, 91 (1968).

calculated phases were used, omitting the contribution of this side chain.

Two allotypic forms of CPA_α exist,[76] either pure or as an equal mixture. The pure forms have Ile-179, Ala-228, and Val-305 *or* Val-179, Glu-228, and Leu-305. It is now possible to employ these chemical results to identify the material which was used in the X-ray work. All X-ray data of CPA_α were taken on material from a single cow (Alice), and the X-ray identities of these residues are unambiguously Ile-179, Ala-228, and Val-305.

The X-ray and chemical sequence identifications have not yet been completely reconciled for CPA_α. Questions remain at residues 93, 108, 151, and 245 for which X-ray preferences are Glx, Glu, Trp, and Ser, respectively. If 108 is Glu, then the H_2O nearby must be H_3O^+. Our preference of Trp-151 instead of Phe-151 may be a difference in sample or another allotypic site. Residue 256, identified as Asp in the chemical sequence, is preferred as Asn in the X-ray results in order to avoid a buried uncompensated charge as the substrate is bound. The residue of carboxypeptidase B, which is spatially equivalent to 256 of CPA_α, may indeed be Asp in order to accommodate the C-terminal Lys or Arg substrates preferred by carboxypeptidase B.

Finally, we comment on the use of X-ray sequence for discovery of homology between probably genetically related enzymes. The X-ray sequence, particularly in a well-ordered region critical for function, may have errors but is probably good enough for the recognition of homology. Before the chemical sequence of carboxypeptidase A was available, the X-ray sequence 246–257 (Ser-Ile-Tyr-Glx-Ala-Ser-Gly-Gly-Ile-Ser-Asx-His) was matched[83] against a corresponding sequence for carboxypeptidase B (Thr-Ile-Tyr-Pro-Ala-Ser-Gly-Gly-Ser-Asp-Asp-Trp). There are four errors in this fragment of X-ray sequence of carboxypeptidase A, three of which, when corrected by the chemical sequence, extend the homology. The chemical sequence is Thr-Ile-Tyr-Glu-Ala-Ser-Gly-Gly-Ser-Ile-Asp-Trp.

(83) T. H. Plummer, private communication, April 8, 1969; *J. Biol. Chem.*, **244**, 5246 (1969).

Nevertheless, the X-ray study of this and another sequence led to the conclusion[83] that carboxypeptidases A and B are homologous enzymes. The same conclusion was independently reached[84] on the basis of other regions of these molecules.

The general level of correctness of X-ray identifications in the CPA study is not encouraging. Assuming that this method does not distinguish Glu from Gln[85] or Asp from Asn, we find that 60% of the 93 residues for which no chemical information was available are correctly identified. In the somewhat more highly ordered region 23–103, 75% of the X-ray assignments were correct.[22] In spite of the possibilities of development of general environmental and conformational principles to assist these identifications, and of objective fits of electron density to residues by computer methods,[86] it seems extremely unlikely that any X-ray study, even at high resolution, will produce a correct, or nearly correct, sequence. Nevertheless, it is clear that the X-ray results can be of great value in conjunction with sequence studies as well as for the new three-dimensional structural information that becomes available for proteins and their complexes. One can look forward to opportunities in which close cooperation at all stages of studies by both methods will greatly facilitate placement of peptide fragments and reliable elucidation of sequence, structure, and function in enzymes and other proteins.

The X-ray studies were made in collaboration with J. A. Hartsuck, M. L. Ludwig, T. A. Steitz, G. N. Reeke, Jr., F. A. Quiocho, H. Muirhead, J. C. Coppola, and P. H. Bethge. We have been aided by discussions with C. H. W. Hirs, H. Neurath (who also communicated the sequence in advance of publication), B. L. Vallee, F. H. Westheimer, W. P. Jencks, and others. Special thanks are due to R. J. P. Williams for his efforts in promoting collaboration among research groups studying CPA.

(84) R. A. Bradshaw, H. Neurath, and K. A. Walsh, *Proc. Natl. Acad. Sci. U. S.*, **63**, 406 (1969).

(85) A systematic conformational difference is observed in CPA between Glu, which extends maximally into solution, and Gln, which tends strongly to be near the protein surface.

(86) R. Diamond, *Acta Cryst.*, **A25**, S189 (1969).