THE STRUCTURE OF CARBOXYPEPTIDASE A, IX. THE X-RAY DIFFRACTION RESULTS IN THE LIGHT OF THE CHEMICAL SEQUENCE

By William N. Lipscomb, Jean A. Hartsuck, Florante A. Quiocho, and George N. Reeke, Jr.

DEPARTMENT OF CHEMISTRY, HARVARD UNIVERSITY

Communicated July 7, 1969

Abstract.—Several features of carboxypeptidase A (CPA) which were previously established by the X-ray diffraction structure studies have now been confirmed by the chemical sequence analysis. These results include the number (307) of amino acid residues in CPA_α, the identities of the residues (Arg 145, Glu 270, and Tyr 248) shown by the X-ray study to be involved in substrate binding and catalysis, and the existence of a disulfide bond.

The Zn ligands, shown by the X-ray study to be residues 69, 72, and 196 and identified as His, Glx, and either Glx or Lys, are proved by the chemical sequence to be His, Glu, and His, respectively.

No change is required in our previous mechanistic deductions, which are here extended to include a specific mechanism of activation of the substrate by a net charge on the metal ion, which suffers a change in local dielectric constant when it is covered by a substrate.

Carboxypeptidase A (CPA) is the first enzyme for which the X-ray structure became available before more than fragmentary portions of chemical sequence were known. The electron density map of 2.0-Å resolution[1] has therefore been employed to reach conclusions[2, 3] concerning the chemical identity of the various residues in addition to establishing the structure and specifying the residues involved in binding of substrates and catalysis. Many of these conclusions can now be evaluated in terms of the recently completed chemical sequence.[4]

Following some general comments on the number of residues and on the amino acid sequences at various stages of this study, we consider the identity and function of a number of the more important amino acid residues. The relation to binding and to probable catalytic steps is then followed by a discussion of a mechanism for activation of the substrate by a charge which is buried during the formation of the enzyme-substrate complex.

Amino Acid Sequences.—The number of residues in CPA_α, established as 307 in the X-ray studies,[1, 2] has been confirmed by the chemical sequence. No revision is required of the sequential numbering or of the detailed trace of the polypeptide chain (Fig. 1). The chemical formula of CPA_α (Val) derived from the sequence[4] is $C_{1561}H_{2352}N_{406}O_{465}S_5Zn$, resulting in a molecular weight of 34,472.

At the time that our 2.8-Å electron density map[5] was calculated (August 1966), the known sequence fragments were 22 residues at the N-terminus of CPA_α,[6] a cysteinyl sequence of 7 residues,[7, 8] a cysteinyl sequence of 14 residues,[7, 8] and a C-terminal fragment of 7 residues.[9] After computation of our 2.0-Å map (June 1967), the sequence of a tetrapeptide related to inhibition by β-phenylpropionate

Fig. 1.—The peptide chain, showing (near the center) Zn with its three ligands from the protein, the disulfide bond (at the right), the N-terminus (at the bottom), and the C-terminus (at the left).

was published.[10] Residues 23–103, communicated[11] to us by Professor Neurath as tentative sequence (October 1967) and as final sequence (January 1968), have been compared to X-ray results in some detail.[2, 12] The remaining residues, 104–307, as tentatively identified from the electron density map, were communicated by us to Professor Neurath in October 1967, and subsequently as revisions were made. Within residues 104–307, all available fragments of chemical sequence were placed and numbered, and X-ray identifications were made for the 93 remaining residues. The breakdown of these residues is: 104–125 (segment A[13]), 126–137 (X-ray only), 138–144 ("atypical" Cys[7, 8]), 145–151 (X-ray only), 152–165 ("active site Cys" [7, 8]), 166–177 (X-ray only), 178–184 (replacement[13]), 185–197 (X-ray only), 198–201 (segment E_2[13]), 202–239 (segment B[13]), 240–246 (X-ray only), 247–250 (β-phenylpropionate protection[10]), 251–257 (X-ray only), 258–264 (segment E_1[13]), 265–299 (X-ray only), and 300–307 (C-terminal[9]). The resulting complete tentative sequence is reported elsewhere.[3] We are now able to comment on the general level of correctness of the X-ray identifications of the 93 residues for which no chemical sequence information was available. Assuming that the X-ray method cannot distinguish Glu from Gln or Asp from Asn, 56

amino acids (60%) were correctly identified. By contrast, X-ray identifications were correct for some 75 per cent of the 81 residues in the more highly ordered region 23–103. 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, it seems extremely unlikely that any X-ray study alone will produce a correct, or even a nearly correct, sequence. This method is intrinsically limited by disorder, especially of side chains which extend into solution in the protein crystal. In addition, some nearly isosteric pairs, for example, Val and Thr, are especially difficult to distinguish.

The complete chemical sequence is inconsistent at a few residues with the electron density maps. Bradshaw *et al.*[4] identify residue 151 as Phe, but the electron density of this side chain is too large for Phe. We have integrated the densities of all the tryptophan and phenylalanine side chains in the CPA molecule. On an arbitrary scale, Trp ranges from 6.10 to 9.15 with an average of 7.14 ± 1.14; while Phe ranges from 4.24 to 6.73 with an average of 4.92 ± 0.90. On the same scale residue 151 is 8.07, which is more than three standard deviations from the average Phe density. Thus, residue 151 is Trp with a high statistical confidence. Residue 93 is identified chemically to be Asn,[4] but Asn does not fit the electron density of maps synthesized using either the isomorphous replacement phases or the phases calculated from the rest of the structure. Glx 93 is consistent with both maps. The methyl carbon atom of Thr 245 is entirely missing in both kinds of maps, and therefore the X-ray choice for 245 is Ser. If residue 108 is Glu, then the water molecule adjacent to it must be H_3O^+ because the surrounding medium is entirely hydrophobic.

Functional Residues.—With this background, we now consider individual functional residues in CPA_α, discussing first the more important ones.

The three residues most intimately associated with binding and catalysis are Arg 145, Glu 270, and Tyr 248. The first two of these residues were correctly identified in the X-ray study in the absence of chemical sequence fragments. Tyr 248 was readily identified after location[1, 2] of the tetrapeptide[10] associated with inhibition. An important factor in the uniqueness of these identifications was the behavior of each of these three residues in large conformational changes when the substrate Gly-Tyr was bound to CPA. In particular, the conformational change of Glu 270 eliminated the only two reasonable alternatives for this residue.

The Zn, which is also important in binding and catalysis, has three protein ligands, residues 69, 72, and 196, identified from the X-ray study. Residue 69 was shown in the X-ray study to be His. This result was confirmed by the chemical sequence after the X-ray study showed that the tentative chemically established order of residues 68 and 69 should be reversed to yield Ile 68 and His 69. Residue 72 was identified independently by the X-ray study and the chemical sequence analysis. Residue 196 is shown by the chemical sequence study to be His; no unique X-ray identification was made. The two choices of the X-ray study[2] were Lys or Glx, but His fits the electron density well[14] (Fig. 2). We should not have missed this one!

The N-terminal and C-terminal fragments[6, 9] were readily located in both the

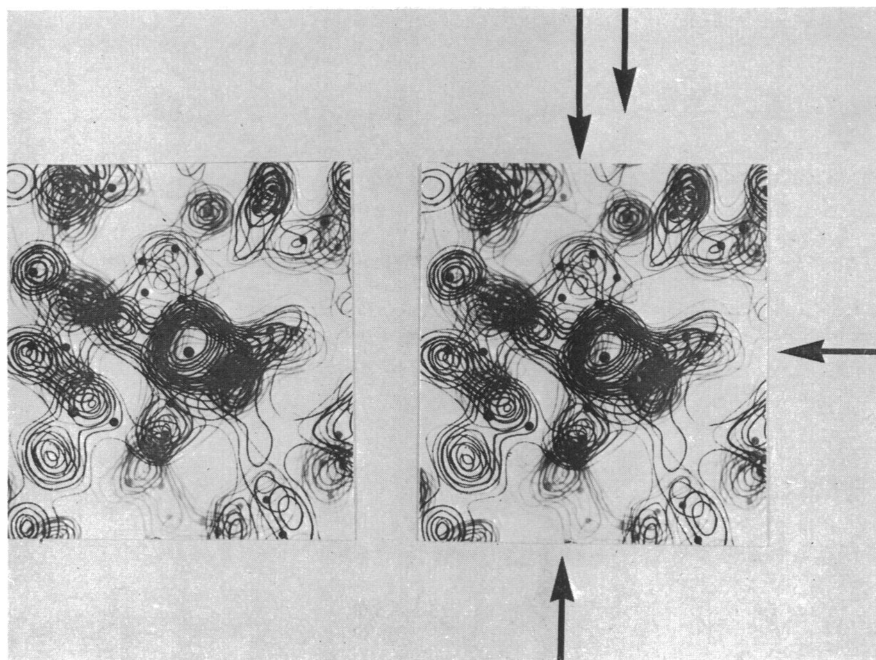

FIG. 2.—Stereoview of electron density in the region of the Zn atom. His 69 is at the right of Zn, the carboxyl group of Glu 72 is below the Zn, and the imidazole group of His 196 is above and slightly to the left of Zn, as shown by the arrows. A water molecule which is hydrogen-bonded to His 196 is indicated by the shorter arrow at the top of the figure.

2.8-Å and 2.0-Å electron density maps. The N-terminus is some 25 Å from Zn and the C-terminal Asn residue is close to the side chain of residue 265. The chemical sequence indicates that 265 is Tyr, and thus the X-ray identification of 265 as Arg is incorrect. Either the electron density map is distorted here, or Tyr 265 is disordered (Fig. 3).

The joining of the only two Cys residues of CPA_{α} by a disulfide bond (Fig. 3) some 20 Å from Zn was clearly proved in our X-ray study[1, 2] by the heavy electron density, the C_{α} to C_{α} distance of 5 Å, the characteristic C-S-S-C dihedral angle of 100°, and the unique location of the two cysteinyl sequences of 14 and 7 residues. The existence of this disulfide bond has now been confirmed by the chemical sequence study,[4] but the sulfur chemistry[12, 15] needs further elucidation.

A secondary, or recognition, binding site some 6 to 8 Å from the Zn is probably responsible for many of the binding and kinetic anomalies shown most prominently by aromatic N-acyl dipeptides.[2] The X-ray identifications of the three residues in this region are Arg 71, Tyr 198, and His(Phe) 279. The chemical sequence study has confirmed the first two identifications and resolved the ambiguity of the last as Phe 279.

The two allotypic forms of CPA_{α} have either Ile 179, Ala 228, and Val 305 *or* Val 179, Glu 228, and Leu 305 at the three positions where replacements have been found.[4] Some samples have an equal mixture of these two forms. All

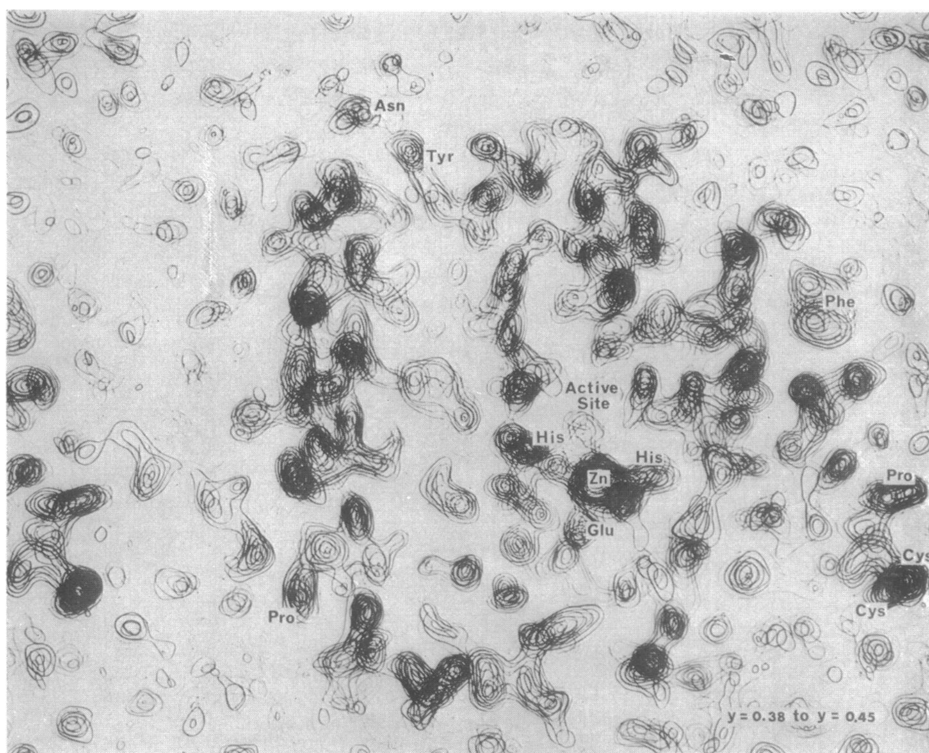

FIG. 3.—Electron density sections $y = 0.38–0.45$, showing Pro 288, Asn, 307 and Tyr 265 at the left, His 69 (right of Zn), Glu 72 and His 196 (upper left of Zn) near the center, and the disulfide bond between Cys 138 and Cys 161, along with Pro 160 and Phe 151 at the extreme right.

X-ray data for native CPA_α were taken on material from a single cow. Our X-ray identifications for our sample are Ile 179, Ala 228, and Val 305.[16]

Binding and Cleavage of Substrates.—No essential changes in our concepts[2, 3] of binding of substrates (Figs. 4, 5*a* and *b*) or of the cleavage mechanism (Fig. 5*c* and *c'*) as deduced from the complex of CPA with Gly-Tyr are required in the light of the sequence study. The revised residues in these two figures (His 196 and Phe 279) do not participate in the catalytic steps, and hence the conclusions[2] from the X-ray study concerning the roles of Arg 145, Tyr 248, and Glu 270 remain intact. Arg 145 is the binding site of the substrate's C-terminal carboxylate group. Tyr 248 is not the nucleophile but probably is the proton donor[1, 2] to the NH of the susceptible peptide bond, and Glu 270 is either the promoter of general base catalysis (Fig. 5*c*) or the nucleophile (Fig. 5*c'*). Further chemical studies are required to resolve this ambiguity of mechanism and to deduce the order of events.

No complexes between CPA and ester substrates have yet proved suitable for X-ray study, but a study of the cleavage of *O*-(*N*-benzoylglycyl)-L-β-phenyllactate by CPA_α in H_2O and in D_2O has yielded a $(k_{cat})_{H_2O}/(k_{cat})_{D_2O}$ ratio of 2.0. An independent study by Kaiser and Kaiser[17] of the CPA-catalyzed hydrolysis of

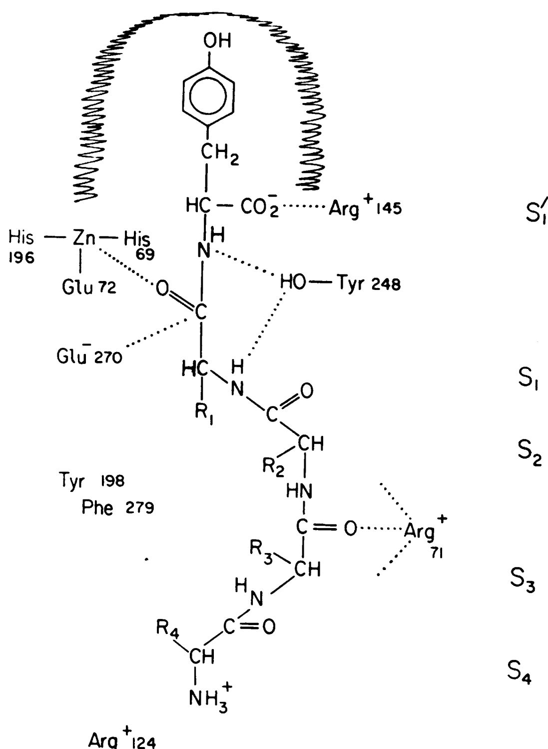

FIG. 4.—Probable binding of polypeptide substrates to CPA.

O-(*trans*-cinnamoyl)-L-β-phenyllactate also yields a similar value for this ratio. (Such large ratios have not been found for peptide substrates.) As discussed earlier,[2] ratios of this magnitude have been used in other systems to support a general base step in the catalysis. However, it is not clear whether attack of H_2O (or D_2O) occurs on the substrate or on the acyl-enzyme intermediate (if formed), and it seems safe only to conclude that a proton transfer seems more important in the rate-determining step for ester hydrolysis than for peptide hydrolysis.

Charge Covering Mechanisms.—The functions of strain, polarization, and electron localization in the substrate have often been discussed in enzyme mechanisms, for example in CPA hydrolysis.[2] In addition, we now describe a process by which a charged group of an enzyme, for example, a metal ion, is activated during the catalytic step. There are three requirements: a net charge, water (solution) displacement, and the covering of the net charge by a polarizable group of importance in the catalytic step. The Zn^{+2} with its three protein ligands (Glu^-, 2 His) has a net charge of +1. The fourth ligand to Zn in native CPA is H_2O (or OH^-, depending on pH), and its neighborhood is in contact with aqueous solution of high dielectric constant K_1. Consequently, the electrostatic lines of force of this net charge will be concentrated preferentially toward this region of high K_1, rather than toward the interior of the protein where a low dielectric constant K_2 prevails. When an organic substrate, for example, a small polypeptide, displaces the fourth Zn ligand and water molecules in the active site region, the net charge of +1 is more nearly entirely surrounded by a medium of low dielectric constant. However, within this surrounding medium of low dielectric constant lies the readily polarizable carbonyl group of the substrate, beyond which are polarizable water molecules and the negative charge of Glu 270.

Induction of a dipole in the substrate by the charged Zn atom could be enhanced in two ways by virtue of this complex existing on the surface of the enzyme. First, when the Zn ion is covered by a substrate molecule which, like the

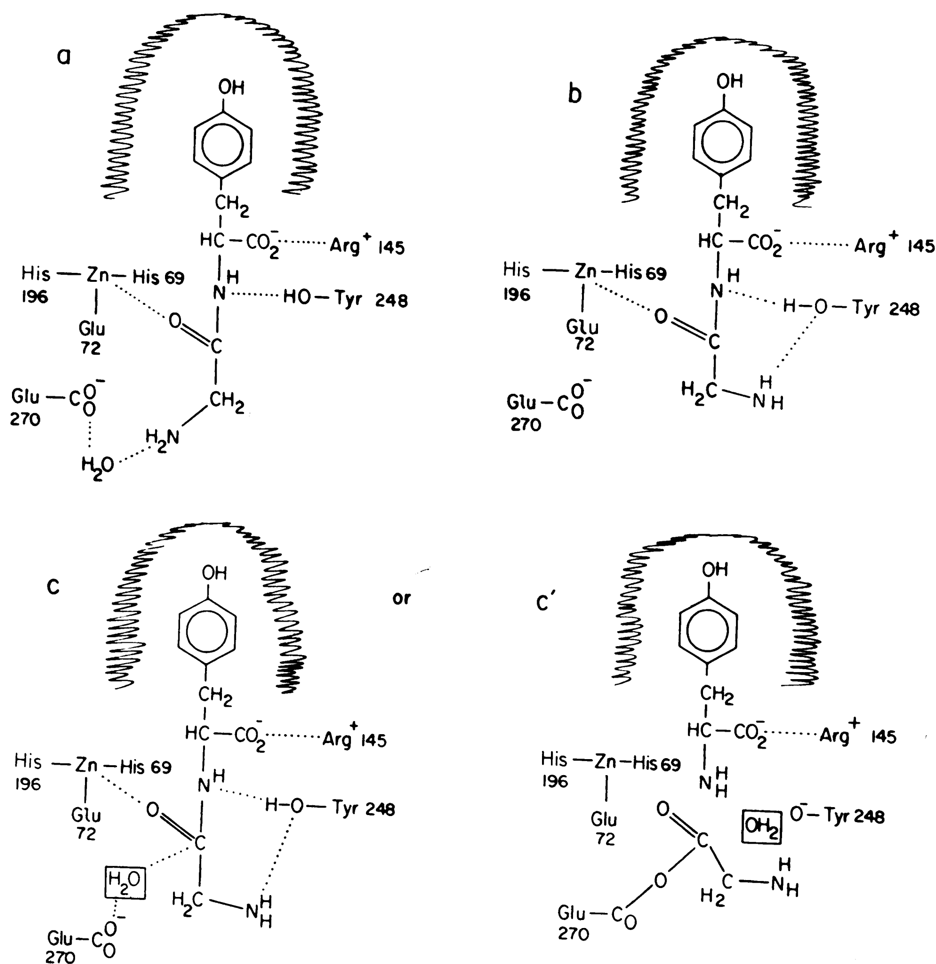

FIG. 5.—(*a*) Binding of Gly-Tyr as deduced from the 2.0-Å resolution study of the enzyme-substrate complex.

(*b*) Pròbable binding of Gly-Tyr in the configuration for cleavage.

(*c*) General base path, in which Glu 270 promotes attack of H_2O on the susceptible peptide's carbonyl carbon.

(*c*′) Acyl enzyme path in which Glu 270 forms an anhydride which is then cleaved by H_2O.

We are not certain in these diagrams whether the N terminus of the Gly-Tyr is NH_2 or NH_3^+ in the complex with CPA.

enzyme molecule on the other side of the Zn, is not easily polarized, the effective charge of the ion in a medium of lower K is increased. We realize that the energy change associated with covering the ion with substrate opposes the binding of substrate; however, we expect that this energy is small compared to other factors involved in substrate binding, for example, neutralization of the charge on Arg 145, hydrophobic bonding forces, and entropy changes. Second, when the Zn ion is surrounded by an apolar medium, the lines of force from the charge tend to be localized on the carbonyl group of the substrate. Polarization of this group is further enhanced by the presence of the negative charge of Glu 270 and of

polarizable water molecules adjacent to the carbonyl carbon atom and away from the Zn.

It is, of course, well known that uncompensated charges inside proteins are quite unstable, and that charge transfer (or proton transfer) may be induced by such buried charges. However, the idea that the substrate can be induced into an unusually high state of polarization when it covers a charge attached to a protein may be new. This effect is more unambiguous in an enzyme in which a metal ion is functional, because the charge distribution is less dependent upon possible proton transfer than in enzymes which have no functional metal ion.

We wish to thank Professor H. Neurath for sending us the chemical sequence of CPA_α prior to publication, and we thank the National Institutes of Health, the Advanced Research Projects Agency, and the Eli Lilly Company for support of this research.

[1] Reeke, G. N., J. A. Hartsuck, M. L. Ludwig, F. A. Quiocho, T. A. Steitz, and W. N. Lipscomb, these PROCEEDINGS, **58**, 2220 (1967).

[2] Lipscomb, W. N., 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, in *Structure, Function and Evolution in Proteins*, Brookhaven Symposia in Biology, No. 21 (1968), p. 24.

[3] Lipscomb, W. N., G. N. Reeke, Jr., J. A. Hartsuck, F. A. Quiocho, and P. H. Bethge, *Phil. Trans. Roy. Soc. London*, in press.

[4] Bradshaw, R. A., L. H. Ericsson, K. A. Walsh, and H. Neurath, these PROCEEDINGS, **63**, 1389 (1969), communicated to us on June 23, 1969.

[5] Lipscomb, W. N., in *Structural Chemistry and Molecular Biology*, ed. A. Rich and N. Davidson (San Francisco: W. H. Freeman and Co., 1968), p. 38.

[6] Sampath-Kumar, K. S. V., J. B. Clegg, and K. A. Walsh, *Biochemistry*, **3**, 1728 (1964).

[7] Neurath, H., *Federation Proc.*, **23**, 1 (1964).

[8] Sampath-Kumar, K. S. V., K. A. Walsh, J.-P. Bargetzi, and H. Neurath, in *Aspects of Protein Structure*, ed. G. N. Ramachandran (New York: Academic Press, 1963), p. 319.

[9] Bargetzi, J.-P., E. O. P. Thompson, K. S. V. Sampath-Kumar, K. A. Walsh, and H. Neurath, *J. Biol. Chem.*, **239**, 3767 (1964).

[10] Roholt, O. A., and D. Pressman, these PROCEEDINGS, **58**, 280 (1967).

[11] Neurath, H., private communication.

[12] Neurath, H., R. A. Bradshaw, L. H. Ericsson, D. R. Babin, P. H. Petra, and K. A. Walsh, in *Structure, Function and Evolution in Proteins*, Brookhaven Symposia in Biology, No. 21 (1968), p. 1.

[13] Tentative sequence received in January 1968 from Professor H. Neurath.

[14] Placement of the α-carbon of residue His 196, while acceptably in density, is quite different (by about 1.6 Å) than it was when 196 was assumed to be Lys or Glx. This new placement of the α-carbon requires strain in the backbone which seems to be best accommodated by distorting the peptide between 197 and 198. Our currently preferred interpretation of this region makes this a *cis*-peptide.

[15] Vallee, B. L., and J. F. Riordan, in *Structure, Function and Evolution in Proteins*, Brookhaven Symposia in Biology, No. 21 (1968), p. 91.

[16] Residue 256, which is identified in the chemical sequence as Asp, is located at the back of the pocket, where the C-terminal side chain of the substrate binds. When an aromatic side chain is inserted in the pocket, displacing water, residue 256 must undergo a conformational change in order that the charge not be buried.

[17] Kaiser, B. L., and E. T. Kaiser, these PROCEEDINGS, **64**, 36 (1969).