THE STRUCTURE OF CARBOXYPEPTIDASE A, VI. SOME RESULTS AT 2.0-Å RESOLUTION, AND THE COMPLEX WITH GLYCYL-TYROSINE AT 2.8-Å RESOLUTION

BY G. N. REEKE, J. A. HARTSUCK, M. L. LUDWIG, F. A. QUIOCHO, T. A. STEITZ, AND W. N. LIPSCOMB

DEPARTMENT OF CHEMISTRY, HARVARD UNIVERSITY

Communicated October 25, 1967

Bovine carboxypeptidase A_α (CPA) is a zinc-containing enzyme of mol wt 34,600 which catalyzes the hydrolysis of polypeptides at the C-terminal peptide bond, especially if the terminal residue of the substrate is aromatic. We have extended our X-ray diffraction study of CPA[1-4] to 2.0-Å resolution and the study of its complex with glycyl-L-tyrosine to 2.8 Å. A detailed examination of the course of the polypeptide chain shows that the molecule is composed of several helical regions, a large central twisted sheet of eight extended chains with considerable β-structure, and a tortuously folded coil. We have located published portions of the amino acid sequence at the N-terminus, the C-terminus, the two cysteinyl residues, and the active tyrosyl residue. We also describe the protein side chains which are near the bound Gly-Tyr and the large conformational changes which occur in the presence of this substrate.

Experimental.—CPA crystallizes in the monoclinic space group $P2_1$, with unit cell dimensions $a = 51.41$ Å, $b = 59.89$ Å, $c = 47.19$ Å, $\beta = 97.58°$. On the native enzyme we have measured 20,600 reflections to 2.0-Å resolution. We also have measured complete data to 2.8-Å resolution on four heavy atom derivatives:[1, 3] Pb (2 sites), Hg (3 sites), Hg (1 site), and Pt (4 sites). On two of these, Pb(2) and Hg(3), we have also measured those 6,000 of the (14,000) reflections between 3.0 and 2.0 Å which gave the largest intensities in the native data set. All data were recorded at 4°C, using a Supper-Pace automated diffractometer and copper radiation. The quality of these data is indicated by the quantities $R = \sum_{Hs} |F_{Hs}^2 - F_H^2| / \sum_{Hs} F_H^2 = 0.074$ for Pb(2) and 0.070 for Hg(3), obtained in scaling together all of the reflections, H, from the several crystals, s, common to each derivative.

Refinement of heavy atom positional, thermal, and occupancy parameters yielded substantially the same values as were obtained at 2.8 Å. The figure of merit (the average value of the cosine of the error in the phase angle) was 0.63 for all reflections, and 0.52 for the 2.5- to 2.0-Å range. Details of the refinements will be presented in a later paper of this series. The electron density map was calculated at points of an orthogonal grid at a scale of 2 cm per Å, in sections at intervals of 0.01 of the b axis. This map was interpolated and presented by the computer in symbolic form so that the contour lines could be traced on Mylar sheets directly from the printed output.

The procedures used for the preparation of the enzyme-substrate complex and the calculation of the difference electron-density map have been published previously.[4]

Structural Features of the CPA Molecule.—A detailed atomic interpretation of the 2.0-Å map was accomplished by placing colored markers at proposed atomic

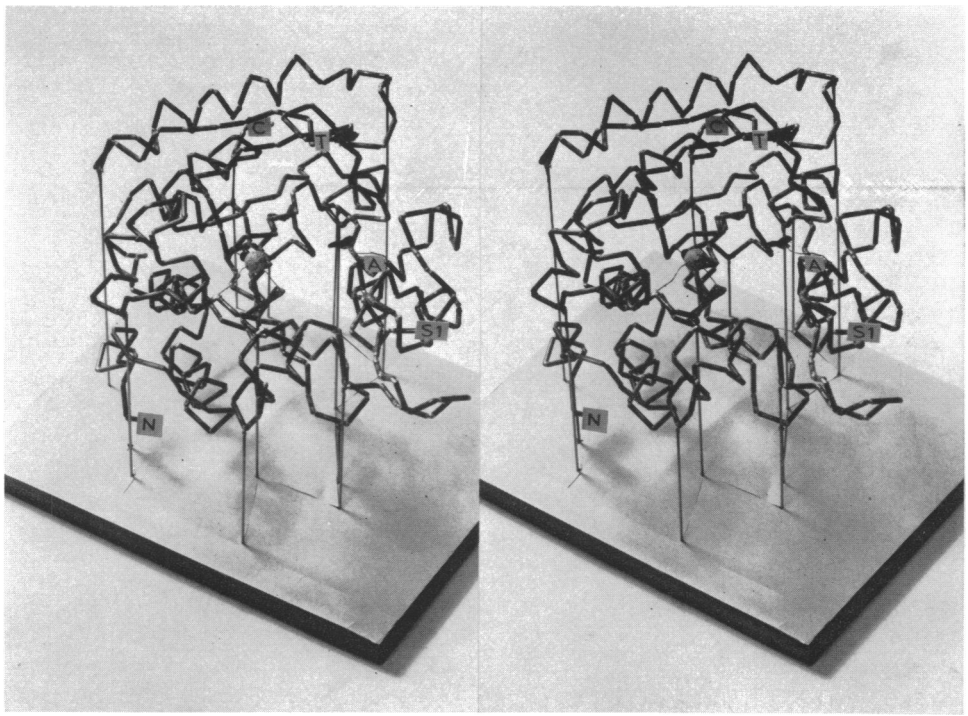

FIG. 1.—Stereoscopic view of the peptide unit structure of carboxypeptidase A. The C_α atoms are at the joints, and each short length of rod is a single peptide unit. The larger ball is the Zn atom. Symbols are $N = NH_3^+$ terminus, $C = CO_2^-$ terminus, *S1* = Cys, *A* = Arg (binding site of CO_2^- of substrate), and *T* = Tyr (center of ring moves 8 Å when substrate is bound). The direction of crystallographic *b* axis is along the vertical support rods. Stereoviewers may be obtained, for example, from Wards Natural Science Establishment, Inc., Rochester, New York, model 25W 2951.

Note: S1, displaced accidentally during photography, should be moved one residue to the left and towards the reader and covalently bonded to the residue on the part of the chain just below

positions for the entire backbone and a substantial fraction of the side-chain atoms. Following the course of the polypeptide chain was aided by the unusually clear appearance of the carbonyl groups in almost all instances. Backbone atoms could be placed so that their bonds deviated within ±0.5 Å of expected bond distances.

Side chains of several residues could be easily identified. Among the aromatics tryptophan is unique, and in most instances the tyrosine hydroxyl is visible. Also, prolines, methionines, cysteines, and ordered arginines, lysines, and isoleucines can be distinguished. Many of the other side chains can be identified if their environment in addition to their shape is taken into account. However, ambiguities often arise in distinguishing sterically similar groups such as His, Phe; Leu, Asp, Asn; Glu, Gln; and Val, Thr. In addition, disordering of the side chains and bound water molecules frequently make identification difficult on the surface of the protein molecule.

We have constructed a wire model (Fig. 1) of the backbone in which α carbons are represented by the junctions between straight peptide units. This model of CPA_α

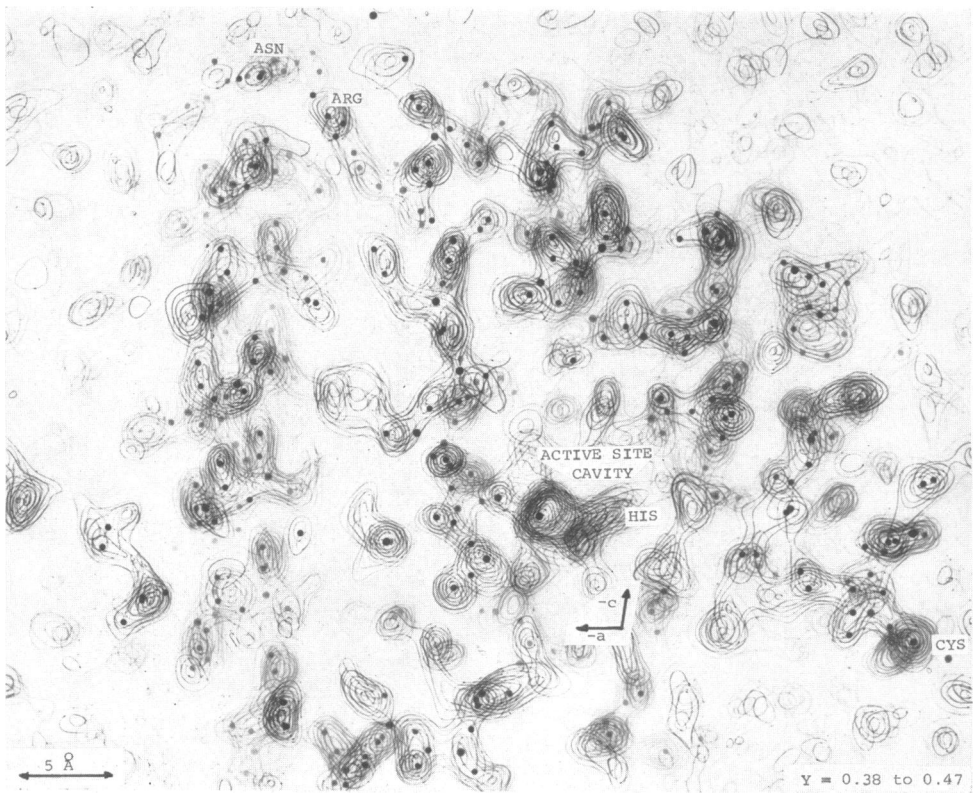

FIG. 2.—The C-terminal helix extends along the left side of the figure and ends at the top at Asn 307, which is linked to an Arg. The densest feature on the map is the Zn atom, which is linked to His. The second Zn ligand extends upward to the left. The third Zn ligand extends downward from the Zn atom and slightly toward the left. The S atom (S1) of a Cys at the far right is just to the left of its label.

contains 307 residues in agreement with the chemical studies.[5] Comparison of this backbone model with that constructed earlier from our 6-Å map[1] shows that, although the latter is not accurate in detail (for reasons cited at that time), it does reflect very well many of the important structural features of CPA, such as the distribution of helix and coil, the presence of a pocket, and the location of the N- and C-termini.

The helical regions, which represent about 30 per cent of the molecule, are mostly α-helix,[6] and are nearly all on the outside of the molecule. There are four major helical regions involving residues 14–29, 72–88, 215–233, and 288–305, four other regions, each containing a few turns of helix of various degrees of perfection, and also a few isolated turns. Also, αII helix[7] is observed at the end of two helices.

About 20 per cent of the residues are contained in eight extended chains, forming a twisted pleated sheet[8] in which the top chain is rotated 120 degrees with respect to the bottom chain. This sheet is made of four pairs of parallel and three pairs of antiparallel chains, and forms the lining for one side of the pocket which extends into the molecular interior. Within this sheet about 55 residues are involved in about 42 β-structure hydrogen bonds. We have assumed at this stage of refine-

ment that hydrogen bonds between chains are formed if the distance between the carbonyl oxygen of one chain and an amide nitrogen of the adjacent chain is 2.5–3.5 Å. Most of the helices are on one side of the pleated sheet and are approximately parallel to the adjacent extended chains. These structural features are no doubt a source of great molecular stability.

The coil at the right of the pleated sheet (Fig. 1) is tortuous and possesses no apparent secondary structure. This 20 per cent of the molecule (residues 122–174 and 247–254) forms the primary region involved in the structural changes described below.

The zinc atom appears to have three protein ligands, one of which is an imidazole nitrogen of histidine (Fig. 2). In addition to the three ligands there is density adjacent to the zinc which may form one or more other ligands. However, in our present interpretation, this density is not a part of the protein molecule.

Known sequences: The N-terminal sequence of 22 residues[9] fits the contour maps very well indeed as exemplified in Figure 3. These results confirm with certainty

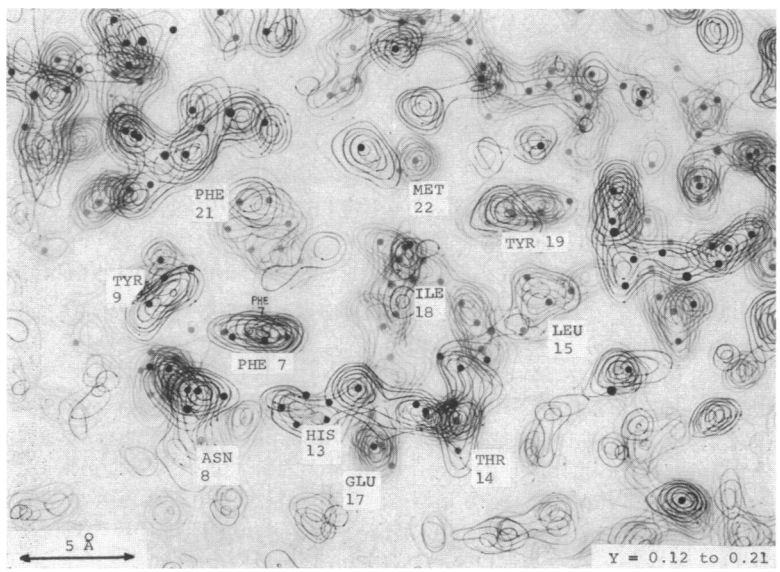

FIG. 3.—Residues in the general region of the N terminus.

our earlier[1] location of the N-terminal region of the CPA molecule. Our location[1] of the C-terminus at the top of the helix in Figure 2 is also confirmed by fitting of the seven residues (301–307) established by chemical methods.[10] The C-terminal Asn is probably hydrogen-bonded to the guanidinium group of an Arg (Fig. 2), helping to make Asn 307 unavailable for self-digestion.[10]

There are two cysteinyl residues per molecule of CPA. One was presumed to be a Zn ligand[11] and the other was described as unreactive.[12] The residues assigned previously[13, 14] to the active site cysteinyl sequence are Gly-Lys-Ala-Gly-Ala-Ser-Ser-Pro-Cys-Ser-Glu-Thr-Tyr, and those assigned to the atypical cysteinyl sequence are Cys-Val-Gly-(Val, Asp)-(Asn, Ala). These two sequences have been located, and appear to be joined by a covalent link at S1 (Fig. 1).

Vallee and co-workers[15] have shown that two tyrosines are involved in the active site of CPA. In Figure 4 we present a region whose density is consistent with the sequence Ile-Tyr-Glx-Ala, which was shown to include one of these active site tyrosines.[16] The tyrosyl side chain (*T* in Fig. 1) is projecting into the solution in native CPA.

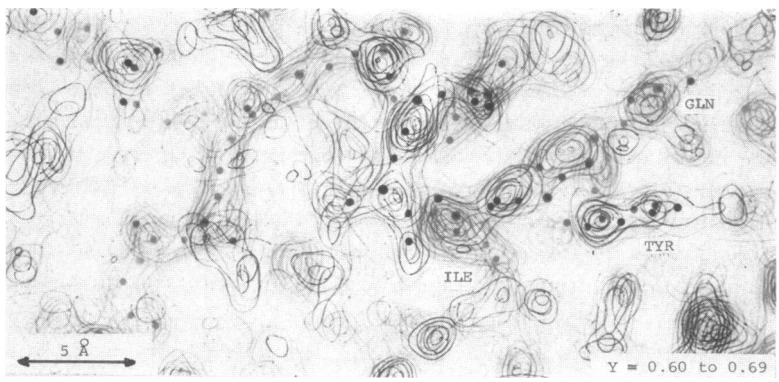

FIG. 4.—Location of the active tyrosine sequence in the native enzyme. An extended polypeptide chain runs from the lower left corner to the upper center of the figure.

Enzyme-Substrate (Gly-Tyr) Complex.—The 2.8 difference electron-density synthesis of the unusually stable (perhaps nonproductive) Gly-Tyr CPA complex clearly shows the orientation of the substrate tyrosyl side chain and the location of its carboxylate group (Fig. 5). This binding of the substrate to CPA involves a number of interactions, several of which we can recognize. First, there is no specific binding group, e.g., an aromatic side chain, for the C-terminal side chain of the substrate. Furthermore, the enzyme pocket which binds the substrate is large and will accommodate a tryptophan side chain. These observations are in agreement with the absence of high specificity for this side chain. Second, the terminal carboxylate group of the substrate interacts with the guanidinium of an Arg (*A* in Fig. 1), which moves 2 Å towards the carboxylate group (Fig. 5). Third, although we cannot recognize the substrate peptide bond at 2.8-Å resolution, our placement of the tyrosyl side chain and the carboxylate group, along with the fact that only L-amino acids are hydrolyzed, limits the possible locations of the peptide bond. The most attractive of the possibilities is one in which the carbonyl oxygen of the peptide bond serves as a zinc ligand. Fourth, a second tyrosine of the enzyme is near iodine peak 2 in our iodophenylpropionate difference map,[4] and thus may be one of the tyrosines protected by phenylpropionate from acetylation.[15] Model building suggests that this tyrosine is so placed that it may interact with a substrate having a side chain on the penultimate amino acid of a substrate.

Most of the conformational changes produced when Gly-Tyr is bound involve residues in the right-hand coil, but only two motions have been interpreted. First, the phenolic hydroxyl of Tyr T (Fig. 1) moves about 14 Å so that the OH group is in the vicinity of the peptide bond of the substrate. This motion involves a rotation of the side chain about the C_α—C_β bond as well as a motion of the peptide

backbone. The proposal that this tyrosine moves when CPA binds substrates or inhibitors is consistent with the protection of this tyrosine by the inhibitor β-phenyl propionate from iodination,[17] acetylation,[15] and nitration.[18] Second, the guanidinium group of Arg *A* (Figs. 1 and 5) moves about 2 Å by means of a rotation about the C_β—C_γ bond. The other motions in the coil may provide a relationship between the movement of Tyr *T* and Arg *A*. We hope that higher resolution studies and the examination of other complexes will clarify the nature and function of the structural change.

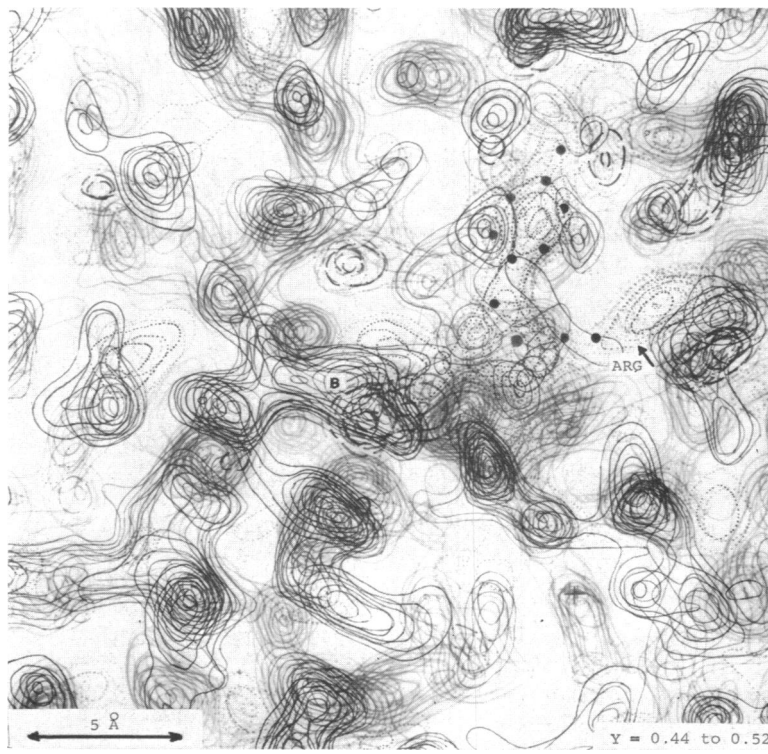

FIG. 5.—Superposition of the Gly-Tyr difference map contours on the native enzyme electron density. Negative difference map contours are dashed lines and positive contours are dotted. The phenol and carboxylate groups of the substrate Gly-Tyr are shown as dots. The arrow indicates the direction of motion of guanidinium of Arg *A* toward the carboxylate group of Gly-Tyr. Residue *B* is indicated.

In addition to the tyrosine and the zinc, at least one other group, the carboxylate of an Asp or Glu (*B* in Fig. 5), is near the peptide bond of the substrate. Although our current picture of the enzyme-substrate complex is a static one, and certainly incomplete, it is tempting to speculate on a mechanism of hydrolysis involving these three groups. One mechanism which is consistent with our data, most of the available chemical data on CPA, model compound studies, and the general mechanistic scheme which has been proposed,[19] is one in which the carboxylate of residue *B* participates as a nucleophile to form a transient anhydride with the substrate carbonyl group. The tyrosine *T* could donate its hydrogen to the peptide

amide nitrogen. The zinc functions both to orient the substrate and to polarize the carbonyl group, facilitating both the formation of an anhydride and its subsequent hydrolysis at the carbonyl of the substrate.

Details of these and further studies will be published in conjunction with related studies from the laboratories of H. Neurath and B. L. Vallee, both of whom we thank for helpful comments.

We wish to thank the National Institutes of Health and the Advanced Research Projects Agency for support of this research. We also acknowledge a National Science Foundation predoctoral fellowship to G. N. R. We are also grateful to Dr. C. H. W. Hirs for supplying us with pancreatic juice, and most indebted to our former colleagues H. Muirhead and J. C. Coppola, whose participation in the earlier research made a substantial contribution to the present results.

[1] Lipscomb, W. N., J. C. Coppola, J. A. Hartsuck, M. L. Ludwig, H. Miurhead, J. Searl, and T. A. Steitz, *J. Mol. Biol.*, **19**, 423 (1966).

[2] Lipscomb, W. N., M. L. Ludwig, J. A. Hartsuck, T. A. Steitz, H. Muirhead, J. C. Coppola, G. N. Reeke, and F. A. Quiocho, *Federation Proc.*, **26**, 385 (1967).

[3] Ludwig, M. L., J. A. Hartsuck, T. A. Steitz, H. Muirhead, J. C. Coppola, G. N. Reeke, and W. N. Lipscomb, these PROCEEDINGS, **57**, 511 (1967).

[4] Steitz, T. A., M. L. Ludwig, F. A. Quiocho, and W. N. Lipscomb, *J. Biol. Chem.*, **242**, 4662 (1967).

[5] Bargetzi, J.-P., K. S. V. Sampath Kumar, D. J. Cox, K. A. Walsh, and H. Neurath, *Biochemistry*, **2**, 1468 (1963).

[6] Pauling, L., R. B. Corey, and H. R. Branson, these PROCEEDINGS, **37**, 205 (1951).

[7] Nemethy, G., D. C. Phillips, S. J. Leach, and H. A. Sheraga, *Nature*, **214**, 363 (1967).

[8] Pauling, L., and R. B. Corey, these PROCEEDINGS, **37**, 729 (1951).

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

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

[11] Vallee, B. L., T. L. Coombs, and F. L. Hoch, *J. Biol. Chem.*, **235**, PC45 (1960).

[12] Walsh, K. A., K. S. V. Sampath-Kumar, J.-P. Bargetzi, and H. Neurath, these PROCEEDINGS, **48**, 1443 (1962).

[13] 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.

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

[15] Simpson, R. T., J. F. Riordan, and B. L. Vallee, *Biochemistry*, **2**, 616 (1963).

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

[17] Simpson, R. T., and B. L. Vallee, *Biochemistry*, **5**, 1760 (1966).

[18] Riordan, J. F., M. Sokolovsky, and B. L. Vallee, *Biochemistry*, **6**, 358 (1967).

[19] Vallee, B. L., *Federation Proc.*, **23**, 8 (1964).