ADVANCES IN PROTEIN CHEMISTRY

EDITED BY

C. B. ANFINSEN, JR.

National Institute of Arthritis and Metabolic Diseases
Bethesda, Maryland

JOHN T. EDSALL

Biological Laboratories
Harvard University
Cambridge, Massachusetts

FREDERIC M. RICHARDS

Department of Molecular Biophysics
Yale University
New Haven, Connecticut

VOLUME 25

1971

ACADEMIC PRESS • *New York and London*

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
Berkeley Square House, London W1X 6BA

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 448853

PRINTED IN THE UNITED STATES OF AMERICA

ADVANCES IN PROTEIN CHEMISTRY

Volume 25

CONTRIBUTORS TO THIS VOLUME

G. A. O. Alleyne

J. Drenth

J. N. Jansonius

R. Koekoek

William N. Lipscomb

Karl A. Piez

Florante A. Quiocho

Wolfie Traub

J. C. Waterlow

B. G. Wolthers

CONTRIBUTORS TO VOLUME 25

Numbers in parentheses indicate the pages on which the authors' contributions begin.

G. A. O. ALLEYNE, *Medical Research Council of Great Britain, Tropical Metabolism Research Unit, University of the West Indies, Jamaica* (117)

J. DRENTH, *Laboratory for Structural Chemistry, University of Groningen, Groningen, The Netherlands* (79)

J. N. JANSONIUS, *Laboratory for Structural Chemistry, University of Groningen, Groningen, The Netherlands* (79)

R. KOEKOEK,† *Laboratory for Structural Chemistry, University of Groningen, Groningen, The Netherlands* (79)

WILLIAM N. LIPSCOMB, *Department of Chemistry, Harvard University, Cambridge, Massachusetts* (1)

KARL A. PIEZ, *National Institute of Dental Research, National Institutes of Health, Bethesda, Maryland* (243)

FLORANTE A. QUIOCHO, *Department of Chemistry, Harvard University, Cambridge, Massachusetts* (1)

WOLFIE TRAUB, *Department of Chemistry, The Weizmann Institute of Science, Rehovot, Israel* (243)

J. C. WATERLOW,* *Medical Research Council of Great Britain, Tropical Metabolism Research Unit, University of the West Indies, Jamaica* (117)

B. G. WOLTHERS, *Laboratory for Structural Chemistry, University of Groningen, Groningen, The Netherlands*

† Deceased.

* Present address: Department of Human Nutrition, London School of Hygiene and Tropical Medicine, London, England.

PREFACE

The present volume of Advances in Protein Chemistry begins with two chapters that describe the application of X-ray crystallography to the study of the structural basis of enzyme action. The power of this technique, and its broad application since the pioneering studies of Perutz and Kendrew and their colleagues on hemoglobin and myoglobin, has led to a veritable flood of protein structures over the past few years. It will clearly be impossible to present, within the covers of this series, an adequate treatment of this mass of data and its significance for the understanding of protein function. However, the current discussions of carboxypeptidase A and papain are excellent examples of the correlations that can be made between three-dimensional structure and the details of enzyme catalysis as revealed by studies of enzymes in solution. In fairness to the complexity of the general problem of the mechanism of enzyme action it must be emphasized that, with all the insight afforded by the experiments outlined in the chapters by Quiocho and Lipscomb and by Drenth, Jansonius, Koekoek, and Wolthers, the precise elucidation of enzyme catalysis still eludes us. As was pointed out by L. H. Jensen and his colleagues at the most recent Cold Spring Harbor Symposium, the degree of resolution required for the ultimate description of such reactions may well have to be at the level of 1.5 Å or better to clearly define, critically, both bond angles and lengths. We hope in future issues of these Advances to present further reviews dealing with the crystallographic approach, including chapters on insulin, ribonuclease, and other well-studied proteins. The application of current refinements not only in crystallography but in the study of enzymes in solution should give us a complete interpretation of enzyme catalysis in one or more cases in the near future.

The second half of this volume consists of two chapters which describe the advances that have been made in two topics that were reviewed about 10 years ago in this series. Waterlow and Alleyne have written an extensive summary of protein malnutrition in children, which carries on from the chapter on protein malnutrition in man by Waterlow, Cravioto, and Stephen in Volume 15. The present chapter emphasizes the outgrowth of an important concept in the understanding of nutritional problems—the role of adaptation of the organism to nutritional situations.

The extensive review on the structure of collagen and gelatin by Harrington and von Hippel that appeared in Volume 16 is succeeded, in the present volume, by an equally thorough and authoritative summary of

the present status of the chemistry and structure of collagen by Traub and Piez. The application of modern techniques of sequence analysis together with the crystallographic and physicochemical study of polytripeptide models of collagen have recently provided a fairly unambiguous picture of the conformation of this protein, and the chapter describes what is probably the final solution to a puzzle in protein structure that has been under intensive investigation since the earliest days of X-ray crystallography.

We wish to express our sincere appreciation to the staff of Academic Press for their invaluable help in the preparation of this Volume and in the arduous task of assembling the index.

C. B. Anfinsen
J. T. Edsall
F. M. Richards

July 1971

ADVANCES IN PROTEIN CHEMISTRY

Volume 25

CONTENTS

The Chemistry and Structure of Collagen

WOLFIE TRAUB AND KARL A. PIEZ

CARBOXYPEPTIDASE A: A PROTEIN AND AN ENZYME

By FLORANTE A. QUIOCHO and WILLIAM N. LIPSCOMB

Department of Chemistry, Harvard University, Cambridge, Massachusetts

I. Introduction

The emphasis of this chapter is on the relationship of the three-dimensional structures of bovine carboxypeptidase A (CPA),[1] and of its complexes with substrates and inhibitors, to the functional behavior of this enzyme. In particular, we describe the basis for substrate specificity, modes of binding, and the possible mechanisms of hydrolytic cleavage of substrates for this enzyme. Also, where clear relationships exist, the many chemical studies of CPA and its activities are interpreted in terms of the structural results.

Carboxypeptidase A is a zinc-containing proteolytic enzyme, which catalyzes the hydrolysis of carboxy-terminal peptide bonds in protein and peptide substrates. It is secreted by the acinar cells of the pancreas (Siekevitz and Palade, 1958) as an inactive zymogen, procarboxypeptidase A (ProCPA) (Anson, 1935, 1937), which is distributed among species ranging at least from the spiny dogfish to man. Most studies, including the crystallographic work, have been made on the enzyme obtained from bovine pancreas. Some of its physicochemical properties are collected in Table I.

[1] Abbreviations used in this article: CPA, carboxypeptidase A; proCPA, procarboxypeptidase A; (apoCPA), apocarboxypeptidase A; Z, carbobenzoxy.

TABLE I
Some Physicochemical Properties of Carboxypeptidase A[a]

Property		References[c]
Formula: $C_{1561}H_{2352}O_{465}N_{406}S_5Zn$[b]		(1, 2)
Amino acid composition: $Asp(n)_{29}$, Thr_{26}, Ser_{32}, $Glu(n)_{25}$, Pro_{10}, Gly_{23}, Ala_{21}, Val_{16}, Met_3, Ile_{21}, Leu_{23}, Tyr_{19}, Phe_{16}, His_8, Lys_{15}, Arg_{11}, Trp_7, Cys_2		(1)
Molecular weight: 34,472		(1, 2)
$s_{20,w}$(S), at pH 7.0: 3.06 (CPA_δ)		(3)
Isoelectric point pH: (at ionic strength 0.2)	6.0 (CPA_δ)	(4)
$D_{20,w}$ (cm^2 sec^{-1}):	8.86×10^{-7} (CPA_δ)	(4)
$\epsilon_{270\,nm}$ (liter $mole^{-1}$ cm^{-1}):	6.42×10^4	(5)
b_0, native enzyme (degrees):	−125 (CPA_δ)	(6)
b_0, CPA in 8 *M* urea (degrees):	−20 (CPA_δ)	(6)
Crystallographic data		(7, 8)
Space group	$P2_1$	
Molecules/unit cell	2	
Unit cell parameters	$a = 51.41$ Å $b = 59.89$ Å $c = 47.19$ Å $\beta = 97.58°$	
Molecular dimensions	50 Å × 42 Å × 38 Å	(9)
Zinc coordination number	4	(8, 10)
Zinc ligands	H_2O, His 69, Glu 72, His 196	(1, 2, 10)
Helix content (%)	38	(9)
β-Structure (%)	17	(9)

[a] Unless otherwise specified reference is to carboxypeptidase A_α (Val 307).

[b] For CPA_α (Val 307), assuming charged groups for Arg, Lys, Glu, Asp, N-terminus, and C-terminus, and assuming that each His has a proton on one N only.

[c] Key to references: (1) Bradshaw *et al.* (1969b); (2) Lipscomb *et al.* (1969); (3) Smith *et al.* (1949); (4) Putnam and Neurath (1946); (5) Simpson *et al.* (1963); (6) Quiocho *et al.* (1967); (7) Ludwig *et al.* (1963); (8) Reeke *et al.* (1967); (9) Lipscomb *et al.* (1970); (10) Lipscomb *et al.* (1968).

The discovery of carboxypeptidase and its peptidase activity by Waldschmidt-Leitz (1931) and co-workers led to details of its behavior. For example, pH-activity data (Waldschmidt-Leitz and Purr, 1929) for the substrate chloroacetyl-L-tyrosine gave reasonable rates from pH 5.6 to 9.0, with an optimal rate at just over pH 7. The C-terminal specificity gave rise to a suggestion (Waldschmidt-Leitz, 1931) of ionic binding between the substrate's carboxylate group and some basic group on the enzyme. Also, the side chain specificity was studied, but questions remained as to the purity of these early preparations. Hence, it was a considerable advance when Anson (1935, 1937) first isolated monodispersed crystalline CPA.

As a result, details of the specificity of CPA became established firmly. (1) The peptide bond which is hydrolyzed (Fig. 1) must be adjacent to

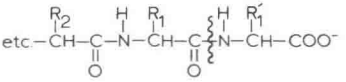

FIG. 1. Peptide substrate for CPA, showing position of cleavage at the wavy line.

a C-terminal free carboxylate ion (Waldschmidt-Leitz, 1931; Hofmann and Bergmann, 1940). For example, amidation of this carboxylate ion prevents cleavage of the peptide bond. (2) The rate of hydrolysis is enhanced if the side chain of the C-terminal residue is aromatic or branched aliphatic (Stahmann *et al.*, 1946). A somewhat simplified summary is that, at about pH 8, 25°, and ionic strength of 0.2, rates are generally high for C-terminal Tyr, 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 (Ambler, 1967). (3) Dipeptides having a free amino (or NH_3^+) group are hydrolyzed slowly, but if this group is blocked by N-acylation the hydrolysis is rapid (Hofmann and Bergmann, 1940). (4) Although peptides having C-terminal Gly or D-Ala (Schechter and Berger, 1966) are hydrolyzed very slowly, other side chains at this position must be in the L-configuration (Bergmann and Fruton, 1937; Hanson and Smith, 1949; Dekker *et al.*, 1949). (5) Substitution of a methyl group (in sarcosine) or a methylene group (in proline) for the H atom of the susceptible peptide bond prohibits or greatly reduces hydrolysis of this C-terminal residue of the substrate (Stahmann *et al.*, 1946; Smith, 1948). (6) The rate of hydrolysis of the C-terminal peptide bond in N-acyl dipeptides is greatly decreased by the substitution of β-alanine (Hanson and Smith, 1948) or sarcosine (Snoke and Neurath, 1949) for the penultimate amino acid of the substrate. (7) At least five C-terminal residues of the substrate influence K_m, and, to a somewhat lesser extent, k_{cat} (Abramowitz *et al.*, 1967). Thus the binding region is about 5 residues, or 18 Å, in length. In a later section we shall correlate these chemical results with the three-dimensional structure, at 2.0 Å resolution, for CPA and its complex with the dipeptide glycyl-L-tyrosine (Gly-Tyr). In particular, proposals are introduced for binding, for catalysis, and for anomalies in the interactions of CPA with its substrates and modifiers.

Carboxypeptidase A is a metalloenzyme, the first for which the structure is known. Inhibition by a number of metal-ion combining substances, including cysteine, led to the proposal by Smith and Hanson (1949) that the enzyme contains a metal ion at the active site. They suggested that the metal ion was Mg^{2+}, but it was later identified as Zn^{2+} (Vallee and Neurath, 1954). Removal of Zn^{2+}, either by lowering the pH below 5.5 or by use at neutral pH of a variety of chelating agents, yields an inactive enzyme, apocarboxypeptidase A (Vallee *et al.*, 1958). Pep-

tidase activity is known (Coleman and Vallee, 1960, 1961) for Co^{2+}, Ni^{2+}, Mn^{2+} and Fe^{2+} in place of Zn^{2+}, but substitution of Cu^{2+} for Zn^{2+} yields an enzyme which is inactive toward all substrates. Esters are also cleaved by CPA (Snoke *et al.*, 1949), and substitution of Hg^{2+}, Cd^{2+}, or Pb^{2+} retains esterase activity, although these heavy metal derivatives are not peptidases in solution (Coleman and Vallee, 1961). Crystals of the mercury derivative, however, have shown some peptidase activity (Bishop *et al.*, 1966). The crystalline metal-free apoenzyme is stable, enzymatically inactive, having physical characteristics like those of the native enzyme, although crystals of apoCPA grown from solution have been shown by Kraut (Rupley and Neurath, 1960) to possess unit cell parameters different from those of CPA. On the other hand, direct removal of Zn from crystals of CPA yields crystals of apoCPA which are isomorphous with those of CPA (Lipscomb *et al.*, 1966).

Four forms of CPA arise from enzymatic release, in a complex series of reactions, of an N-terminal fragment of some 64 residues (Freisheim *et al.*, 1967) from one of the three subunits of bovine proCPA. The CPA_α form (Cox *et al.*, 1964) (Table II), prepared by methods described elsewhere (Lipscomb *et al.*, 1966), was shown by the X-ray diffraction

TABLE II
Chemical Forms of Carboxypeptidase A

Form	Common name	N-terminus (No. of residues)	Crystal cell constants for space group $P2_1$	Isolation	Reference for isolation[a]
α	Cox	Ala (307)	51.41, 59.89, 47.19 97°35′	Chromatographic (DEAE) purification of proCPA, then tryptic activation	(1, 2)
β	—	Ser (305)	Not isolated	A contaminant	(3)
γ	Anson[b]	Asn	50.9, 57.9, 45.0 94°40′	Selective precipitation of autolyzates from frozen pancreas glands	(4)
δ	Allan	Asn (300)	Same as α	Selective precipitation of dissolved pancreatic acetone powders, after trypsin activation	(5)

[a] Key to references: (1) Cox *et al.* (1964); (2) Lipscomb *et al.* (1966); (3) Sampath Kumar *et al.* (1964); (4) Anson (1935, 1937); (5) Allan *et al.* (1964).

[b] Commercial product.

study to contain 307 residues (Reeke *et al.*, 1967). Earlier sequence studies of the N-terminal region (Sampath Kumar *et al.*, 1964) had established the differences among these four forms (Table II), but all have C-terminal Asn. All these forms show comparable specific enzymatic activity. Conformational differences may exist between CPA_γ and CPA_δ, both of which have 300 residues. These two forms differ in solubilities (Allan *et al.*, 1964), and in reversibility of removal of Zn^{2+}. Thus $apoCPA_\gamma$ is more readily reactivated than is $apoCPA_\delta$ (Vallee *et al.*, 1960). Nevertheless the γ and δ forms have the same sedimentation coefficient and electrophoretic mobility.

Partial chemical sequence data published before computation of X-ray diffraction maps at atomic resolution [2.8 Å (Lipscomb, 1968) in August 1966, 2.0 Å (Reeke *et al.*, 1967) in June 1967] are the 22-residue N-terminal fragment (Sampath Kumar *et al.*, 1964) of CPA_α, the 7-residue C-terminal fragment (Bargetzi *et al.*, 1964), a 14-residue "active-site" cysteinyl sequence and a 7-residue "nonessential" cysteinyl sequence (Sampath Kumar *et al.*, 1964; Neurath, 1964). The X-ray study (Reeke *et al.*, 1967) showed that these two cysteines were covalently linked in a disulfide bond some 20 Å from Zn, and that the N-terminus is about 25 Å from Zn (Lipscomb *et al.*, 1966; Reeke *et al.*, 1967). The Zn^{2+} binding ligands, which appeared earlier from chemical evidence (Vallee *et al.*, 1961; Coombs *et al.*, 1964) to be a thiol group of Cys and the α-amino group (Asn) of CPA, proved actually to be His 69, Glu 72, and His 196 as shown by a combination of X-ray (Lipscomb *et al.*, 1968) and sequence studies (Bradshaw *et al.*, 1969a). The complete chemical sequence of CPA_α (Fig. 2) was established in June 1969 (Bradshaw *et al.*, 1969a). Methods included initial cleavage of CPA_δ by CNBr at methionines 22, 103, and 301 (Nomoto *et al.*, 1969; Bradshaw *et al.*, 1969b; Bradshaw, 1969) followed by their proper ordering (Neurath *et al.*, 1970). All numbering refers to CPA_α in the present literature. The larger size and the great difficulty of obtaining soluble peptides from the large 104–301 fragment necessitated the use of five different proteolytic enzymes, and maleation of lysine groups. Two allotypic forms were identified, one having Ile 179, Ala 228, and Val 305 (CPA_α^{Val}), and the other having Val 179, Glu 228, and Leu 305 (CPA_α^{Leu}) (Petra *et al.*, 1969). The X-ray identifications for these residues were uniquely those of CPA_α^{Val}. Comparison (Lipscomb *et al.*, 1968, 1969; Lipscomb, 1970) of X-ray sequence (Lipscomb *et al.*, 1970) and chemically established sequence (Bradshaw *et al.*, 1969a) showed that only 60–85% of the side chains were identified correctly by X-ray methods, depending upon the clarity in various regions of the maps at 2.0 Å. Of the few discrepancies which remain, only one is related to function: Asp 256 of the chemical sequence

ALA ARG SER THR ASN THR PHE ASN TYR ALA 10 THR TYR HIS THR LEU ASP GLU ILE TYR ASP 20
PHE MET ASP LEU LEU VAL ALA GLN HIS PRO 30 GLU LEU VAL SER LYS LEU GLN ILE GLY ARG 40
SER TYR GLU GLY ARG PRO ILE TYR VAL LEU 50 LYS PHE SER THR GLY GLY SER ASN ARG PRO 60
ALA ILE TRP ILE ASP LEU GLY ILE HIS SER 70 ARG GLU TRP ILE THR GLN ALA THR GLY VAL 80
TRP PHE ALA LYS LYS PHE THR GLU ASN TYR 90 GLY GLN ASN PRO SER PHE THR ALA ILE LEU 100
ASP SER MET ASP ILE PHE LEU GLU ILE VAL 110 THR ASN PRO ASN GLY PHE ALA PHE THR HIS 120
SER GLU ASN ARG LEU TRP ARG LYS THR ARG 130 SER VAL THR SER SER SER LEU CYS VAL GLY 140
VAL ASP ALA ASN ARG ASN TRP ASP ALA GLY 150 PHE GLY LYS ALA GLY ALA SER SER SER PRO 160
CYS SER GLU THR TYR HIS GLY LYS TYR ALA 170 ASN SER GLU VAL GLU VAL LYS SER ILE VAL 180
ASP PHE VAL LYS ASN HIS GLY ASN PHE LYS 190 ALA PHE LEU SER ILE HIS SER TYR SER GLN 200
LEU LEU LEU TYR PRO TYR GLY TYR THR THR 210 GLN SER ILE PRO ASP LYS THR GLU LEU ASN 220
GLN VAL ALA LYS SER ALA VAL ALA ALA LEU 230 LYS SER LEU TYR GLY THR SER TYR LYS TYR 240
GLY SER ILE ILE THR THR ILE TYR GLN ALA 250 SER GLY GLY SER ILE ASP TRP SER TYR ASN 260
GLN GLY ILE LYS TYR SER PHE THR PHE GLU 270 LEU ARG ASP THR GLY ARG TYR GLY PHE LEU 280
LEU PRO ALA SER GLN ILE ILE PRO THR ALA 290 GLN GLU THR TRP LEU GLY VAL LEU THR ILE 300
MET GLU HIS THR VAL ASN ASN 307

FIG. 2. The amino acid sequence for the 307 amino acids in CPA_{α}^{Val}, i.e., Val 305 (Bradshaw *et al.*, 1969a).

is identified in the X-ray study as Asn 256 on the basis of its environment, in order to avoid a buried uncompensated negative charge deep in the active-site pocket when the ES complex is formed.

In addition to elucidation of the enzyme structure at 2.0 Å resolution, the difference Fourier technique has been used to examine substrates and inhibitors bound to CPA or to modified CPA (Steitz *et al.*, 1967; Lipscomb *et al.*, 1968, 1969). From several such studies at 6.0 Å resolution, from the complex of Gly-Tyr with CPA at 2.0 Å resolution and from the complex of Phe-Gly-Phe-Gly at 2.8 Å resolution (Lipscomb *et al.*, 1971), we have identified the active site for peptide hydrolysis, and have given some detailed structural features of the modes of binding of substrates. Combining the detailed structural information from these two poor substrates with the positions of the Zn atom and amino acid side chains of CPA, and with structural features of substrates which are hydrolyzed most rapidly, we have derived a probable structure for a productive enzyme–substrate complex (Lipscomb *et al.*, 1968, 1969). Those features of binding of Gly-Tyr and of Phe-Gly-Phe-Gly which may be characteristic of productive binding are (1) the substrate's C-terminal side chain inserts into a "dead-end" pocket, (2) the C-terminal carboxylate forms a salt-link with the guanidinium group of Arg 145, and (3) the carbonyl oxygen of the susceptible peptide bond displaces the single water molecule bound to Zn, and binds to Zn in place of this water molecule. This binding of substrates is accompanied by very large conformational changes,