

PART I
THE THREE DIMENSIONAL STRUCTURE
OF α -CHYMOTRYPSIN AT pH 4.2

PART II
THE STRUCTURE OF
 α -CHYMOTRYPSIN AT pH 6.7

Thesis for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
RICHARD LEE VANDLEN
1972



This is to certify that the
thesis entitled

PART 1: THE THREE DIMENSIONAL STRUCTURE OF
 α -CHYMOTRYPSIN AT pH 4.2

PART 11: THE STRUCTURE OF α -CHYMOTRYPSIN AT pH 6.7
presented by

RICHARD LEE VANDLEN

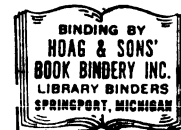
has been accepted towards fulfillment
of the requirements for

PH.D. degree in CHEMISTRY

A. Tulinsky
Major professor

Date Sept. 27, 1972.

Q-7639



ABSTRACT

PART I: THE THREE DIMENSIONAL STRUCTURE OF
 α -CHYMOTRYPSIN AT pH 4.2

PART II: THE STRUCTURE OF α -CHYMOTRYPSIN AT pH 6.7

By

Richard Lee Vandlen

The three dimensional structure of the proteolytic enzyme α -chymotrypsin was determined to 2.8A resolution by X-ray crystallographic techniques. Crystals of α -chymotrypsin with space group $P2_1$ are grown from ammonium sulfate solutions at pH 4.2 and have four molecules in the unit cell or two molecules in the asymmetric unit. The cell dimensions are $a = 49.24$, $b = 67.20$, $c = 65.94$ Å, and $\beta = 101.79^\circ$. A change of the pH from 4.2 to 6.7 of the ammonium sulfate solution in which the crystals are stored produced substantial changes in the X-ray diffraction pattern of the native enzyme. The structure of α -chymotrypsin at pH 6.7 was studied by difference electron density techniques.

The structures of the two independent molecules of α -chymotrypsin in the asymmetric unit were solved by the use of six heavy atom isomorphous derivatives. Intensity data were collected by a computer controlled diffractometer. The electron density of the enzyme was interpreted in terms of

the kn

a-chyme

The pat

is esse

sheets,

locatio

establi

the fit

bution.

The

related

Various

made in

differen

as shown

differen

Many of t

region an

enzyme mo

Diff

densities

of molecu

molecular

ment of th

terminal c

action wit

molecule i

the known amino acid sequence of 241 residues which comprise α -chymotrypsin. A model of the enzyme molecule was constructed. The path of the polypeptide chain through the globular molecule is essentially random with the formation of α -helices, pleated sheets, β -bends, nonpolar regions and aromatic clusters. The location of several sulfate molecules bound to the enzyme was established. The atomic coordinates were refined by optimizing the fit of the model to the observed electron density distribution.

The two independent molecules in the asymmetric unit are related by a local two-fold axis to form a 'dimeric' molecule. Various reciprocal interactions between the two molecules are made in the dimer interface region. There are significant differences in structure between the two independent molecules as shown by variations in the heavy atom derivatives and by a difference electron density map between the two molecules. Many of the observed deviations occur around the interface region and in the vicinity of the active sites of the two enzyme molecules.

Difference electron density maps between the electron densities of the enzyme at pH 4.2 and pH 6.7 revealed a number of molecular changes which are both intramolecular and intermolecular in nature. The largest changes involved the movement of the side chain of histidine 40, the rotation of the terminal carboxyl group of tyrosine 146 away from an interaction with the imidazole ring of histidine 57 of the second molecule in the asymmetric unit, and complicated movements

near the

changes

regions

adjustm

molecul

Richard Lee Vandlen

near the favored uranyl binding site. Somewhat smaller changes in the structure are observed around the active site regions of both molecules and other places. In addition, adjustments of the solvent structure on the surface of the molecule were revealed.

PART I: THE THREE DIMENSIONAL STRUCTURE OF
 α -CHYMOTRYPSIN AT pH 4.2

PART II: THE STRUCTURE OF α -CHYMOTRYPSIN AT pH 6.7

By

Richard Lee Vandlen

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry

1972

G788 76

To my parents, Mr. and Mrs. V. D. Vandlen, for
their constant encouragement and support through-
out the many years of my education.

For

ship thr

his sinc

The

Morimoto

ulating

structur

To

Liebman,

Mr. Step

many hel

Sup

Science

fully ac

Tha

Laborato

related

ACKNOWLEDGMENTS

For his constant guidance, support, enthusiasm and friendship throughout all aspects of this study, the author extends his sincere appreciation and thanks to Dr. Alexander Tulinsky.

The author wishes to thank Dr. N. V. Mani, Dr. Carl Morimoto, Dr. L. H. Wright and Dr. William Brinigar for stimulating discussions and for their many contributions to the structure determination of α -chymotrypsin.

To his fellow students, Dr. Penelope Coddling, Mr. Michael Liebman, Miss Irene Moustakali, Mr. Aristides Mavridis and Mr. Stephen Kraft, the author appreciates their assistance and many helpful discussions.

Support of the Molecular Biology Section of the National Science Foundation during the course of this study is gratefully acknowledged.

Thanks are also due to Mr. John Fox of the Computer Laboratory for his generous assistance with many computer-related problems.

GENERAL

PART I.

I.

II.

III. D

IV. M

V. ST

TABLE OF CONTENTS

GENERAL INTRODUCTION.	1
PART I: THE THREE DIMENSIONAL STRUCTURE OF α -CHYMOTRYPSIN AT pH 4.2	
I. INTRODUCTION	3
1. The Chymotrypsin Family of Proteins.	3
2. Physical and Chemical Properties of Chymotrypsin	5
3. Structural Studies	7
II. EXPERIMENTAL	9
1. Methods.	9
2. Three Dimensional Data Collection.	11
3. Heavy Atom Derivatives and Phase Angle Calculations	15
4. Electron Density Map Calculation	23
III. DESCRIPTION OF THE MOLECULE.	25
1. Folding of the Polypeptide Chain	25
2. β -Bends.	31
3. The Active Site Region	32
4. Non-polar Cavity	38
5. Aromatic Clusters.	41
IV. MOLECULAR INTERACTIONS	46
1. The Structure of the α -CHT Dimer	46
2. Local Two-fold Interactions.	53
3. Location of Bound Sulfate Molecules.	60
4. Differences Between the Two Molecules.	62
5. Uranyl Binding Interface	69
V. STRUCTURE REFINEMENT	76
1. Refinement of the Coordinates.	76
2. Real Space Refinement.	79
3. Comparison With the MRC Structure of "Average" α -CHT.	84

PART I.

VI.

VII.

VIII.

REFERENC

APPENDI

PART II: THE STRUCTURE OF α -CHYMOTRYPSIN AT pH 6.7

VI. INTRODUCTION.	87
1. Effect of pH on Protein Structure	87
2. Effect of pH on α -CHT	90
VII. EXPERIMENTAL.	95
1. Crystal Preparation	95
2. Data Collection and Processing.	98
3. Difference Electron Density and Error Assessment.	100
VIII. DISCUSSION OF THE RESULTS	104
1. General Features of the Difference Map.	104
2. Histidine 40.	107
3. Dimer Interactions.	112
4. Active Site	114
5. Uranyl Interface.	118
6. Other pH Changes.	118
7. Final Comments.	120
REFERENCES.	124
APPENDIX. Side Chain Environment and Clarity in The Electron Density of α -CHT.	129

Table

I.

II.

III.

IV.

V.

VI.

VII.

VIII.

IX.

No

X.

LIST OF TABLES

Table

I.	Heavy Atom Parameters at 2.8Å Resolution	17
II.	Summary of Side Chain Environment.	29
III.	Best Defined β -Bends in α -CHT.	33
IV.	Non-Polar Cavity, α -CHT.	42
V.	Aromatic Clusters.	45
VI.	Transformation Refinement Summary.	50
VII.	Bound Sulfate Positions.	63
VIII.	Ionizable Groups in Globular Proteins.	88
IX.	Cell Dimensions for pH 6.7 α -CHT Crystals and Native Crystals.	99
X.	Summary of Changes of α -CHT Upon Change of pH. . .	121

Figure

1. A
2. F
3. F
4. A
5. (r
o
t
6. Sc
ar
7. Sc
8. St
ar
9. Sch
uni
pla
10. Int
11. Loc
mole
12. Elec
of c
13. Elec
MET
14. Deta.
urany

LIST OF FIGURES

Figure

1.	Axial diffraction patterns for native α -CHT and for the Pt heavy atom derivative.	12
2.	Heavy atom refinement of α -CHT.	21
3.	Figure of merit distribution curve.	22
4.	Approximate shape of 'dimeric' molecule of α -CHT. .	27
5.	(a) Stereo drawing of a Type I β -Bend formed by residues SER 115 to VAL 118. (b) Stereo drawing of a Type II β -Bend formed by residues ASP 194 to GLY 197.	34
6.	Schematic drawing of the important residues around the active site.	37
7.	Schematic representation of non-polar cavity. . . .	40
8.	Stereoscopic drawing of the tryptophan aromatic cluster.	44
9.	Schematic packing diagrams of α -CHT in the unit cell; (a) projection on yz, (b) on xz plane	52
10.	Intermolecular interactions across dyad A	54
11.	Local two-fold interactions between the two molecules, (a) PHE 39, (b) peptide chain.	57
12.	Electron density of the active site region of one molecule	58
13.	Electron density of the helical region near MET 180	68
14.	Detail of the peptide residues near the uranyl binding site	73

Figur

15

16.

17.

18.

19.

20.

21.

22.

Figure

15.	(a) Typical pH-rate profile for chymotrypsin catalyzed reactions; (b) Functional relationship of pH on dimer formation.	92
16.	Axial scan along c^* direction for Tosyl, native and pH 6.7 crystals of α -CHT	97
17.	Radial distribution curves for pH 6.7 and native crystals.	101
18.	Gaussian curves showing effect of a small shift on the difference density.	106
19.	Schematic drawing of important residues around HIS 40.	108
20.	Electron density around HIS 40	111
21.	Electron density in vicinity of TYR' 146 and HIS 57	111
22.	Electron density map around SER 195.	115

GENER

inter

in the

cesses

techni

molecu

more i

sis ha

obtain

of an

detail

ture a

Couple

reason

rechan

tions

Refine

experim

ships

Th

above

ferent

acid se

GENERAL INTRODUCTION

The past decade has seen the growth of a tremendous interest in the chemistry and physics of biological systems, in the control and adaptation of mechanisms of life processes, and in the very essence of life. Of all of the new techniques that have been developed to study biological macromolecules, X-ray crystallography has emerged as one of the more important tools. As intensity data collection and analysis have improved and matured, it has now become possible to obtain, under the right conditions, the tertiary structure of an enzyme or protein to sufficient resolution to allow detailed conclusions concerning relationships between structure and functionality to be made with some certainty. Coupled with extensive chemical and biochemical studies, reasonable, intelligent proposals can be made concerning mechanisms of reactivity, catalysis, and molecular interactions and the physical forces which are intimately involved. Refinement of the proposals can then be achieved from new experiments designed on the knowledge of the spatial relationships of the amino acids involved.

The harvests of many years of dedicated work using the above principles are beginning to appear on a variety of different biological systems. From a knowledge of the amino acid sequences of a large number of cytochrome molecules

from

tion

propo

evolu

cowor

hemog

led t

logica

out th

tionsh

dispos

units

but re

elucid

sickle

globins

from various species, coupled with the structure determination of the oxidized and reduced forms, some very interesting proposals concerning the course and rate of the biological evolution of proteins have been made by Dickerson and coworkers (1,2). Likewise, years of patient study of the hemoglobin-myoglobin family of proteins by M. F. Perutz have led to some very important insights pertaining to the physiological processes involved in the transport of oxygen throughout the body (3). Not only has Perutz established the relationship among the coordination state of the heme iron, the disposition and structural changes associated with the subunits and the affinity of the molecule for various ligands, but recent work has also made significant advances in the elucidation of the causes of many blood disorders, such as sickle cell anemia, which are due to abnormal mutant hemoglobins (4).

PART I

THE THREE DIMENSIONAL STRUCTURE OF

α -CHYMOTRYPSIN AT pH 4.2

1.

of

ju

po

ta

er

ot

ar

sp

re

by

co

o.

to

s

a

s

i

r

r

a

I. INTRODUCTION

1. The Chymotrypsin Family of Proteins

The chymotrypsin family of proteins is one of a series of homologous enzyme families, found in mammalian digestive juices, whose principal task is the degradation of protein polypeptide chains. The active species of each family contains a unique serine residue, an absolute requirement for enzymatic activity. In addition to the chymotrypsin family, other 'serine' proteinases are trypsin, elastase, subtilisin, and thrombin. Although each enzyme has widely differing specificities, it has been suspected for some time that all react by similar mechanisms. This proposal is strengthened by noting that each enzyme also has a histidine residue as a component in the active site region and that the sequences of amino acid residues around the serine are very similar to one another (5).

The regulation and control of these families are also similar. Each family has an inactive precursor or 'zymogen', and one or more active forms. The enzyme is synthesized and stored in the body in the zymogen form, becoming activated when needed by the enzymatic cleavage of specific peptide bonds or fragments of the polypeptide chain. Such a control mechanism is needed since the enzymes, especially trypsin and chymotrypsin, are autocatalytic in nature.

0.
trypsin
This w
and pur
isolate
shows a
dition
dried
chymot
inacti
simila
each o
trypsi
an ove
specie
bonds
but po
and de
bond i
alpha
chymo
remov
are i
resid
tions
(abou

Of all biological protein systems, that of the chymotrypsin family is probably the most extensively studied system. This wide popularity is due to the relative ease of isolation and purification of the enzyme from bovine pancreas. First isolated by Northrup, Kunitz, and Herriott (6), the family shows a fairly high degree of stability under reasonable conditions and can be stored almost indefinitely as a freeze-dried powder without loss of enzymatic activity. The chymotrypsin family in bovine pancreas consists of two inactive precursors, chymotrypsinogen A and B, and four very similar, but characteristically different, active species of each of the zymogens: alpha, gamma, delta, and pi chymotrypsin A and B. The zymogens appear to be 'isozymes' with an overall amino acid similarity of 86% (5). The active species are formed by enzymatic cleavage of one or two peptide bonds or fragments from chymotrypsinogen followed by specific, but poorly understood, conformational changes (7). The pi and delta forms result from the cleavage of a single peptide bond and fragment respectively (SER 14 - ARG 15), while the alpha and gamma forms are the end products of activation of chymotrypsinogen, with an additional dipeptide fragment removed (THR 147 - ASN 148). Chemically, α and γ chymotrypsin are identical; operationally, the existence of the two forms resides in the formation of different crystalline modifications. The α species forms monoclinic crystals at acid pH (about pH 4) and the γ species forms tetragonal crystals at

ne

th

2.

pr

co

A,

res

24:

Har

a m

gen

iso

res

dis

conf

made

entr

othe

stat

conf

stud

extre

round

subtl

neutral or alkaline pH. The physical differences between these two species are not definitively known (7).

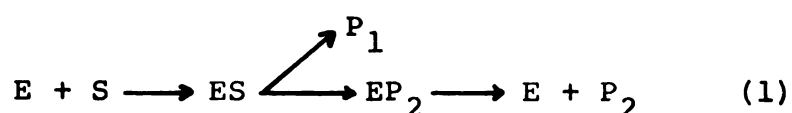
2. Physical and Chemical Properties of Chymotrypsin

Alpha chymotrypsin A (hereafter denoted as α -CHT) is a proteolytic enzyme of chemical composition $C_{1113}N_{300}O_{349}H_{1752}S_{12}$ consisting of 241 amino acids arranged in three peptide chains A, B, C of 13, 131, and 97 residues, respectively. The resultant molecular weight is approximately 25,300 amu. The 245 amino acid sequence of chymotrypsinogen was determined by Hartley (8,9) and Meloun, et al. (10) in 1964 and revised in a minor way in 1969 (11). α -CHT is formed from chymotrypsinogen by the removal of the dipeptides between leucine 13 and isoleucine 16 and between tyrosine 146 and alanine 149. The resultant three peptide chains are held together by five disulfide bridges between cysteine residues.

A study of the thermodynamic relationships between the conformation of α -CHT and its physiological function has been made by Lumry and others (12,13). By monitoring enthalpy and entropy changes as a function of temperature, pH, salt, and other experimental variables, a number of different folded states are found to exist. The magnitude of the changes in conformation can not be explicitly determined by solution studies because the thermodynamic variables observed are extremely sensitive to changes in the water structure surrounding the enzyme. Nonetheless, the existence of some subtle conformation changes during catalysis has been proposed

but the nature of these changes is currently not well understood (12).

The reactivity of α -CHT resides in its ability to cleave amide bonds adjacent to the carboxyl group of large, bulky aromatic amino acids. The enzyme catalyzed reaction is many orders of magnitude faster than comparable non-enzymatic digestion of proteins. The proposed mechanism for hydrolysis involves a three-step reaction



where E is the enzyme and S is the substrate, ES is an enzyme-substrate complex, EP_2 is a second enzyme complex involving the active site serine and P_1 and P_2 are products of the enzyme catalyzed reaction. Recent studies have shown that the hydrolysis reaction is much more complex than that represented in equation 1 with a number of ionization and conformational equilibria also involved (14). The activity of the enzyme displays a bell-shaped pH rate profile with a maximum at $pH \sim 7.2$, and is interpreted as evidence that two ionizing groups control the activity (15).

Unlike other enzymes, the specificity requirements of α -CHT are not very stringent and its affinity for binding substrates and inhibitors is not as large as might be expected, especially when compared to the binding affinities of other enzymes toward their inhibitors. In this sense, α -CHT can be considered to be a 'poor' enzyme. Although amino acid residues with aromatic rings are most frequently

the natural substrates, there is no absolute requirement for aromatic rings, as shown by the fact that saturated hydrocarbon rings such as cyclohexane bind as effectively as aromatic groups. The reasons for the relatively poor specificity are not completely understood; however, the products of chymotryptic catalysis are also inhibitors of α -CHT. Thus, to be reasonably efficient but yet not become completely inhibited by its products, the specificities for inhibitors and substrates should be relatively lower than in other enzymatic systems, where substrates and inhibitors often have completely different chemical identities.

3. Structural Studies

Crystallographic studies have been reported for nearly all members of the chymotrypsin family. The parent structure, chymotrypsinogen, has been determined by Kraut and coworkers to 2.5A resolution (16). The crystal structure of the alpha form was determined by Blow and coworkers at Cambridge (17,18) and the structure of the gamma form has been reported by Davies and coworkers at the National Institutes of Health (19). The π and δ -CHT forms are crystalline forms isomorphous with the γ species.

Our crystallographic studies on α -CHT were begun in 1962. In the intervening years, Blow and coworkers have reported a number of papers relating to the structure of α -CHT. Since there were substantial differences in the mode of operation and in the philosophy of approach between the work of

Blow and ours, this encouraged us to continue with the initial work. In addition to collection of X-ray intensity data by diffractometer methods as contrasted to photographic means, there were also a number of other experimental and procedural differences between the two efforts which have enabled us to elucidate some structural properties heretofore unreported that are important in the functionality of the enzyme. For instance, a systematic study of classes of inhibitors and substrate-like molecules has been made in order to ascertain modes of binding and resultant conformational changes in the enzyme. A detailed study concerning the physical and chemical properties of the enzyme molecules was undertaken also. It is known that environmental forces play a substantial role in governing the activity of the enzyme--such as changes in pH, ionic strength, buffer components, and temperature. Other physical probes and techniques for the study of protein structure and conformation can only show that changes are occurring; x-ray crystallographic studies that have been undertaken can show definitively what conformational changes occur when an environmental change is made. These results lead to new insights concerning conformational properties of enzymes and other biological macromolecules. In addition to the above, the occurrence of two molecules of α -CHT in the crystallographic asymmetric unit gives a unique opportunity to study conformational variability of a protein molecule in a systematic and detailed manner.

1.

wa.

us.

Cr

si

wa

H₂

22

sa

sc

of

Th

m.

1.

to

ti

t

b

r

r

r

c

II. EXPERIMENTAL

1. Methods

Three times crystallized, salt free, lyophilized α -CHT was obtained from Worthington Biochemical Corporation and used without further purification or characterization. Crystals suitable for x-ray work were obtained in a manner similar to that described by Northrup, et al. (6). The α -CHT was dissolved in distilled water adjusted to pH 3.5 with H_2SO_4 to produce a protein concentration of approximately 22 milligrams of protein per millimeter of solution. A 75% saturated ammonium sulfate solution was added to the protein solution until a slight cloudiness appeared. A small amount of water at pH 3.5 was added to just clear the turbidity. The final protein concentration is approximately 6 mg per ml with a pH of 4.2. Large, diamond shaped crystals up to 1.5 mm by 0.75 mm by 0.5 mm grow from this solution in three to four weeks at room temperature. The solution from which the crystals are grown is replaced by a 75% $(\text{NH}_4)_2\text{SO}_4$ solution at pH 4.2 and the tubes of crystals stored in a water bath maintained at 13°C. Crystals stored in this manner retain their excellent quality for many years.

Crystals of α -CHT belong to the monoclinic space group P2_1 as determined from precession photographs. The cell dimensions for the native enzyme determined from diffractometer

measuremen

$\alpha = 49.$

$\beta = 101$

Four enzym

two molecu

tains appr

characteri

crystals,

Since the

material,

in X-ray m

All c

use of a P

trolled by

storage an

raw intens

with an Ar

further p

Crys

glass cap

zing (20)

approxima

the wall

drop of n

a drop of

the capil

stabilize

measurements are

$$\underline{a} = 49.24 \text{ (7) } \text{\AA}, \underline{b} = 67.20 \text{ (10) } \text{\AA}, \underline{c} = 65.94 \text{ (9) } \text{\AA}, \\ \beta = 101.79 \text{ (8)}^\circ$$

Four enzyme molecules are contained in the unit cell, with two molecules in each asymmetric unit. The unit cell contains approximately 60% protein and 40% mother liquor. A characteristic of α -CHT is the tendency to form twinned crystals, with the twinning occurring along the c^* direction. Since the crystals contained varying amounts of twinned material, only crystals with small twins were generally used in X-ray measurements.

All of the X-ray work to be described was performed by use of a Picker Four Circle Automatic Diffractometer controlled by a Digital Equipment Corp. PDP-8 computer. Program storage and retrieval utilized a 32K DEC Disk File, with the raw intensity measurements being recorded on magnetic tape with an Ampex TMZ Tape Transport unit. The tapes were then further processed at the main computer center.

Crystals selected for X-ray examination were mounted in glass capillaries in a manner similar to that described by King (20). Generally, the crystals used in this work were approximately 1.0 x 0.6 x 0.4 mm. The crystals are held to the wall of the capillary by the surface tension of a small drop of mother liquor. To prevent the crystals from drying, a drop of liquid was placed above and below the crystal before the capillaries were sealed. The crystals were allowed to stabilize positionally for at least twenty-four hours in the

capilla

allowed

of the

of the

Pr

the reco

principa

diffract

to indic

changes

ticular

scans an

are show

Hea

a soluti

so as to

of approx

the heavy

case of t

work was

2. Three

Thre

Picker di

cedure (2

balanced

background

capillary before use. The distinct morphology of the crystals allowed the unique b axis to be mounted along the long axis of the capillary which becomes coincident with the phi axis of the diffractometer when the crystal is aligned.

Preliminary X-ray examination of every crystal included the recording of the diffraction pattern along the three principal axes of the crystal. A visual comparison of the diffraction pattern to that of the native crystals sufficed to indicate the quality of the crystal and whether significant changes in the diffraction pattern had occurred for the particular derivative of interest. Typical native crystal axial scans and the corresponding scans for a heavy atom derivative are shown in Figure 1.

Heavy atom derivatives were usually prepared by adding a solution of the heavy atom reagent to a tube of crystals so as to obtain a molar ratio of heavy atom reagent to protein of approximately 10 or 20 to 1. The crystals were soaked with the heavy atom derivative for at least two weeks, and in the case of the uranyl derivative, one month, before any X-ray work was attempted.

2. Three Dimensional Data Collection

Three dimensional intensity data were collected on the Picker diffractometer using a limited omega step scan procedure (21) with automatic realignment capabilities and balanced filters for the measurement of intensities and background (22). The normal mode of data collection consisted

u

u

40

Figure

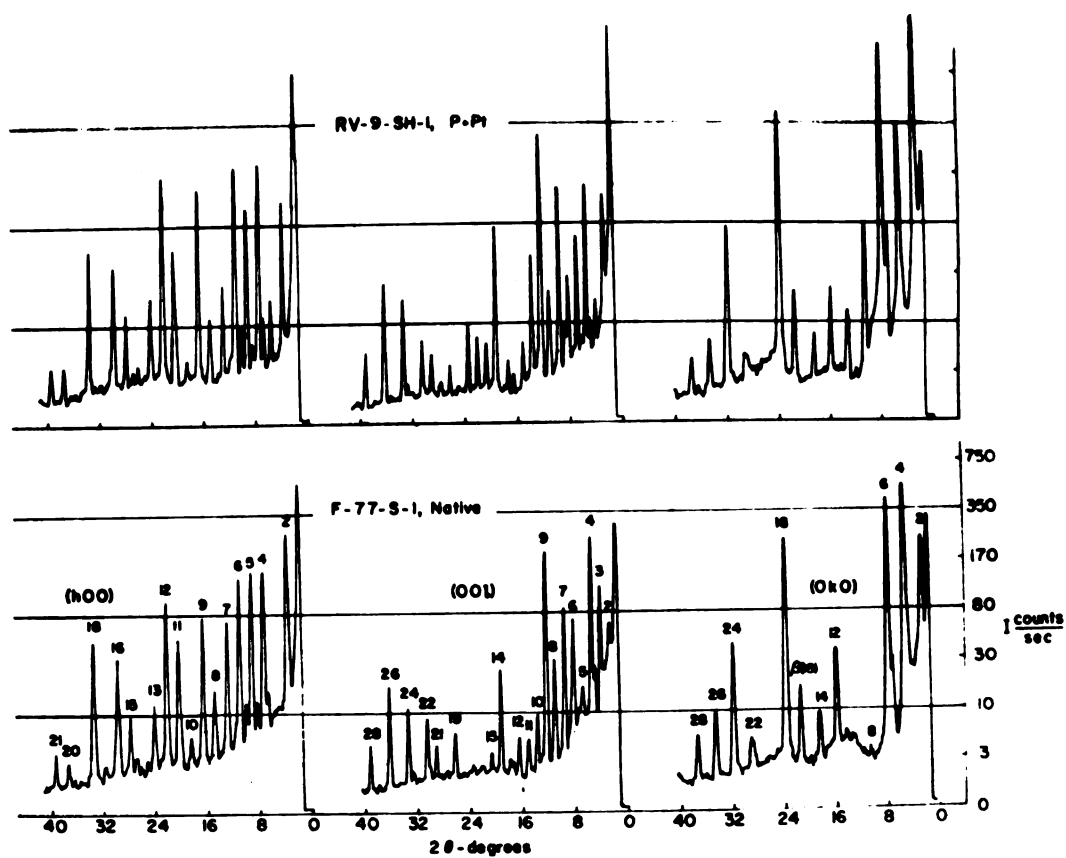


Figure 1. Axial diffraction patterns for native α -CHT and for the Pt heavy atom derivative.

of recor

omega ac

largest

sity. T

the step

filter.

rical pe

would oc

The data

tional s

reflectio

Anot

is the ab

crystal k

automatic

Because t

only a d

the capi

orientat

monitori

to small

lection.

certain s

tatic re

of refle

then cent

coordinat

of recording the diffracted intensity at each of six steps in ω across the center ($\omega = 0^\circ$) of the reflection. The four largest steps of the six were summed to form the raw intensity. The background for each reflection was measured at the step with the largest intensity with a balanced cobalt filter. Occasionally, due to crystal movement or asymmetrical peak shapes, the maximum intensity of the reflection would occur at an ω step on either end of the scan range. The data collection program would then make one or two additional steps to find the peak maximum. Approximately 2000 reflections per day could be measured in this fashion.

Another powerful feature of the data collection program is the ability to detect crystal movement and to 'realign' the crystal by obtaining a new least squares orientation matrix automatically before continuing with data collection. Because the crystals are mounted in glass capillaries with only a drop of moisture adhering the crystal to the wall of the capillary, most crystals tend to undergo some small orientational motions. These movements are detected by monitoring the intensities of three reflections sensitive to small angular displacements every hour or so of data collection. When the intensity of any monitor drops below a certain specified fraction of its initial intensity, the automatic realignment program takes control. A certain number of reflections, well distributed in reciprocal space, are then centered in a procedure which locates the best angular coordinates of each reflection with respect to the

diffi

desci

to th

autom

resun

are h

ment

the q

where

the r

respe

$R \sim 0$

diffe

quart

Since

expos

2.8A

the n

The i

able

measu

and 1

diffractometer angles ω , χ , and 2θ . A new matrix which describes the orientation of the crystal lattice with respect to the diffractometer axial system is then obtained by an automatic least squares procedure (23) before data collection resumes.

Generally, intensity measurements made in this manner are highly reliable and show good reproducibility. Assessment of the quality of the data can be obtained by evaluating the quantity R , as defined as

$$R = \frac{\sum_i^N \left| |F_i| - \langle F \rangle \right|}{\sum_i^N \langle |F| \rangle}$$

where F_i and $\langle F \rangle$ are the observed structure amplitudes of the reflections being compared and their average value, respectively. This factor is nearly always less than $R \sim 0.025$ for redundant data from the same crystal or even different crystals of the same derivative (22,24).

There are approximately 10,500 reflections within one quarter of the limiting sphere defined to 2.8Å resolution. Since the intensities of protein crystals tend to decay when exposed to the X-ray beam for extended periods of time, the 2.8Å set of reflections for each heavy atom derivative and the native enzyme was generally collected from five crystals. The intensity data were measured in shells of 2θ and a suitable number of overlap reflections between each shell were measured to check scaling procedures. Excluding unobserved and low order reflections with d-spacings greater than 15Å,

9044 ref

angle ca.

3. Heavy

Poss

basis of

(h0l) pha

some 12 d

best five

After app

ties, app

were obta

ence Patt

culatation

$|\Delta$

where $|F_p$

derivativ

crystal f

Beca

space gro

individua

the same

origin fo

functions

computed.

the type

9044 reflections (or 86% of the total) were used for phase angle calculations and electron density maps.

3. Heavy Atom Derivatives and Phase Angle Calculations

Possible heavy atom derivatives were screened on the basis of two dimensional projection work by using the centric (h0l) phases determined for the native enzyme (25). Although some 12 different heavy atom derivatives were prepared, the best five derivatives were chosen for three dimensional work. After appropriate data reduction and scaling of the intensities, approximate heavy atom positions for each derivative were obtained from the individual three dimensional difference Patterson maps (26). The coefficients used in the calculation of the Patterson maps were of the form

$$|\Delta(hkl)|^2 = \left| |F_{p+H}| - |F_p| \right|^2 \quad (2)$$

where $|F_{p+H}|$ is the structure amplitude of the heavy atom derivative and $|F_p|$ is the corresponding value of the native crystal for the reflection (hkl).

Because the origin along the y axis is arbitrary in space group $P2_1$, the heavy atom positions obtained from the individual Patterson maps did not necessarily correspond to the same origin. To relate all derivatives to the same origin for phase calculations (27), two types of correlation functions between the various heavy atom derivatives were computed. In the first case, Patterson-like coefficients of the type (28)

were u

deriva

gives

as pos:

tives :

as the

of the

(

which g

two dif

of all

the sam

positio

Ph

using a

(31) an

of phas

the hea

low res

data to

derivat

of the

be obta

distribu

$$\left| |F_{p+H_1}| - |F_{p+H_2}| \right|^2 \quad (3)$$

were used, where $|F_{p+H_1}|$ and $|F_{p+H_2}|$ refer to the heavy atom derivatives 1 and 2 used in the correlation. This function gives the self-vectors of the heavy atoms in each derivative as positive peaks and the mixed cross-vectors between derivatives as negative peaks. This function was not as useful as the second correlation map computed by using coefficients of the type (29,30)

$$(|F_{p+H_1}| - |F_p|) (|F_{p+H_2}| - |F_p|) \quad (4)$$

which gives only the vectors between the heavy atoms in the two different derivatives. Subsequently, the y coordinates of all heavy atom derivatives were adjusted to relate to the same origin, arbitrarily assigned to be the major uranyl position in the uranium derivative (Table I).

Phase angles for all observed reflections were computed using a program originally written by Hilary Muirhead, et al. (31) and modified to our particular needs. Alternate cycles of phase angle calculation and least squares refinement of the heavy atom parameters (32) were performed, first using low resolution data (6A and 3.5A) and later using all the data to 2.8A resolution. The analysis of each heavy atom derivative yields a probability distribution for the value of the phase of the native enzyme. Two types of phases can be obtained from the product of these individual probability distributions, the most probable phase angle, $\alpha_{\max}$ and the

Watch
for
LHP's

Table I. Continued.

Derivative	Site	δ_2	Z	x	y	z	B	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}	R ₁	R ₂	E
U(+)	U	-	83	0.953	-0.003	0.404	25.0*							0.09	0.48	47
	U'	-	42	0.082	0.100	0.363	25.0*									
	C	-	11	0.029	0.102	0.475	25.0*									
	(3)	0.65	17	0.759	-0.686	0.453	30.0*									
	(3')		15	0.764	-0.726	0.521	30.0*									
	(D)	1.62	16	0.969	-0.713	0.002	35.0*									
	(D')		14	0.952	-0.690	0.057	35.0*									
	(F)	0.52	12	0.729	-0.024	0.281	30.0*									
	(F')		16	0.719	-0.391	0.687	30.0*									
U	U	-	67	0.952	0.000	0.399		2.6	1.3	1.8	0.6	0.9	0.0	0.09	0.73	46
	U'	-	17	0.083	0.089	0.394		4.4	1.7	3.0	1.2	-0.3	0.3			
	UA	-	14	0.957	0.010	0.452	34.6									
	C	-	13	0.043	0.097	0.476	42.2									
	(D)	1.08	10	0.062	-0.214	0.014	35.0*									
	(D')		7	0.942	-0.714	0.044	30.0*									

δ - difference in positions (A) after 2-fold symmetry operation on one position, dash (-) indicates no 2-fold counterpart observed.

Z - occupancy in electrons.

B - isotropic temperature factor (A²).

$\beta_{11}, \beta_{22}, \beta_{33}$. . . - anisotropic temperature factors ($\times 10^3$) of form

$$\exp - (h^2\beta_{11} + k^2\beta_{22} + l^2\beta_{33} + hk\beta_{12} + hl\beta_{13} + kl\beta_{23})$$

$$R_1 = \frac{\sum |F_{PH}| - \sum |F_P| \exp(i\alpha_P) + f_H}{\sum |F_{PH}|}$$

$$R_2 = \frac{\sum |F_H - f_H|}{\sum |F_{PH}| - \sum |F_P|}$$

$|F_{PH}|$ and $|F_P|$ are the structure amplitudes for the heavy atom derivative and native enzyme, f_H is the calculated heavy atom contribution, α_P is the phase angle for the native enzyme.

$$F_H = |F_{PH}| \exp(i\alpha_{PH}) - |F_P| \exp(i\alpha_P)$$

E - root mean square lack of closure error in electrons/unit cell.

* - isotropic thermal parameters not refined.

best p

gravity

Be

maps we

tional

(

and

[

where a

phase, c

the cal

by using

ponding

minor o

with pe

lations

The

paramete

atom pos

the ther

allowed

was near

a sixth

was incl

the mole

used in

best phase angle, α_{best} , which is the phase of the center of gravity of the probability distribution.

Between cycles, three dimensional difference Fourier maps were computed for each derivative to locate any additional minor heavy atom sites utilizing coefficients

$$(|F_{p+H}| - |F_p|) \exp [i\alpha_p] \quad (5)$$

and

$$[|F_{p+H}| \exp (i\alpha_{pH}) - |F_p| \exp (i\alpha_p)] - f_H \quad (6)$$

where α_p is the current estimate of the most probable native phase, α_{pH} is a similar quantity for the heavy atom and f_H is the calculated heavy atom structure factor. Maps prepared by using coefficients of the first type have peaks corresponding to all atoms included in the calculation and other minor occupancies. The second map is a true difference map with peaks only for substitutions not included in the calculations.

The relative scale factors, occupancies, positional parameters and thermal parameters were refined for each heavy atom position. Upon inclusion of the high resolution data, the thermal parameters of the major occupancy sites were allowed to vary in an anisotropic manner. When convergence was nearly completed in the least squares refinement process, a sixth derivative with measurements of anomalous dispersion was included in order to fix the absolute configuration of the molecule (33). A summary of the heavy atom derivatives used in the phase determination process and other pertinent

refinement parameters are given in Table I and Figures 2 and 3. The final figure of merit, the mean of the cosine of the error in the phase angle, for all observed reflections to 2.8Å resolution is 0.76.

The six heavy atom derivatives used to determine the phases of the native enzyme by the multiple isomorphous replacement method were obtained from the following chemical reagents: K_2PtI_4 to form the derivatives denoted as Pt and Pt(+), phenyl mercuric acetate to make the Hg derivative, $UO_2(NO_3)_2$ to make the U and U(+) derivatives and $KAuI_4$ with uranyl nitrate to form the mixed derivative U+Au.

Except for the uranyl derivatives, the various heavy atom sites and occupancies in the derivatives were very sensitive to the concentration of the reagent used in the soaking process, and to a somewhat lesser degree, to the time the crystals were allowed to soak in the presence of the heavy metal solution. By use of the reagent K_2PtI_4 , for example, it was possible to prepare two different, isomorphous derivatives Pt and Pt(+). For these two derivatives, Table I shows that enhanced binding occurs at some of the sites, especially 1 and 1', in the Pt(+) derivative. In contrast to the two or three weeks usually needed to prepare a derivative, the U(+) derivative was formed after prolonged soaking (over one year at high concentrations of reagent). To minimize the effects of the concentration dependence, crystals from the same tube were used in the collection of the intensity data whenever possible.

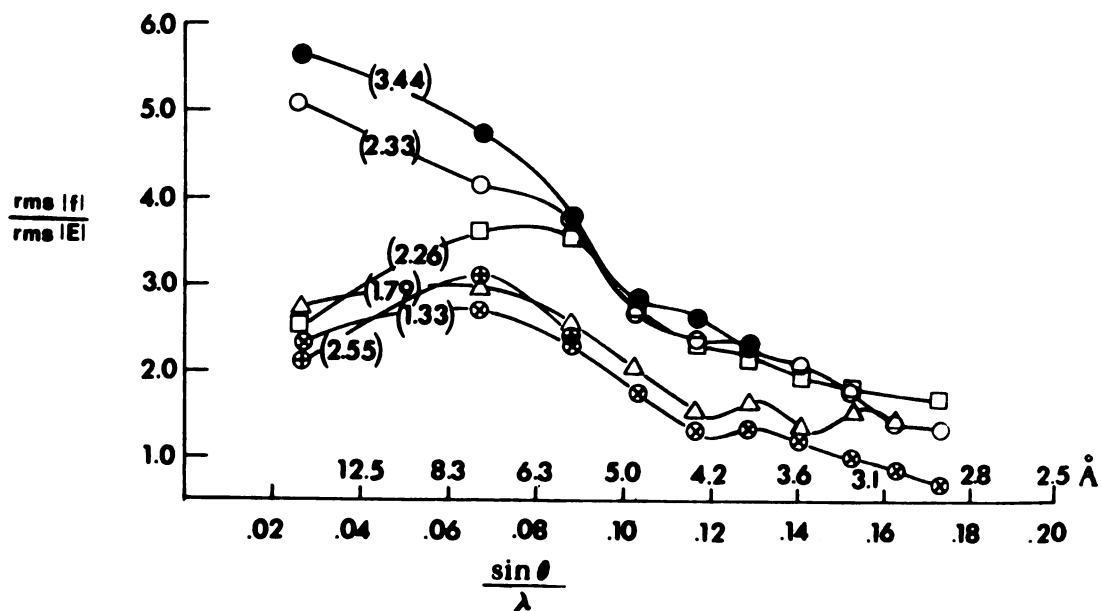


Figure 2. Heavy atom refinement of α -CHT. Curve shows ratio of rms calculated heavy atom contribution, f , to rms lack of closure error as a function of $\sin \theta / \lambda$. The overall ratio for each derivative is given in parentheses. The larger the ratio, the better the phase determining power of the derivative. Only for the last point of the U derivative are the calculated errors larger than the heavy atom contributions.

● - Pt(+); ○ - Pt; Δ - Hg; ⊗ - U; □ - U+Au;
 ● - U(+).

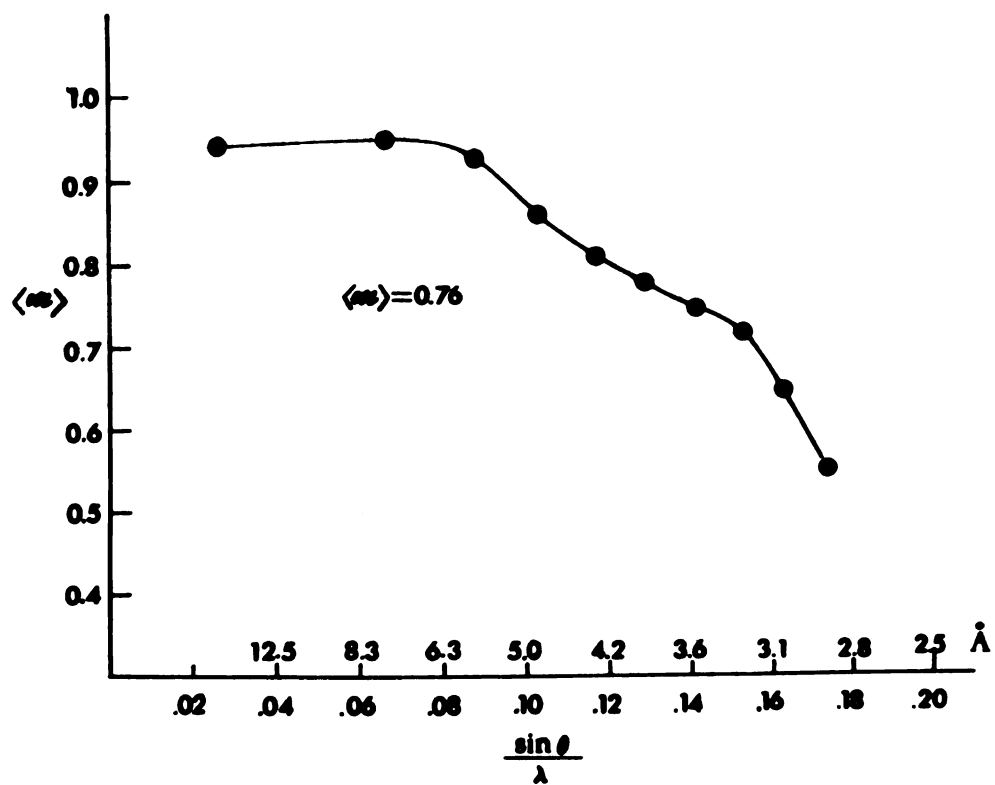


Figure 3. Figure of merit distribution curve. Plot of the average value of the figure of merit, m , as a function of $\sin \theta / \lambda$.

Thru
the heav
two-fold
of α -CHT
differenc
metry op
of sites
the two-
tion. In
have any
in δ_2 .
show part
Hg deriva
derivativ
metry is
occupanc
on the p
from the

4. Elec

The
map of n
with fig
phase ca
calculat
division
coeffici

Through this work, it has been observed that some of the heavy atom derivatives did not show the non-crystallographic two-fold symmetry known to be present between the two molecules of α -CHT in the asymmetric unit. Column δ_2 contains the differences in positions in angstroms after a two-fold symmetry operation is applied to one of the members of each pair of sites grouped by brackets in Table I. The derivation of the two-fold axis and its equation are given in a later section. In each derivative, there are some sites which do not have any two-fold related positions, denoted by the dash in δ_2 . The heavy atom sites in the uranyl derivative (UO_2^{++}) show particularly poor two-fold symmetry, while those in the Hg derivative are particularly good. For all heavy atom derivatives, the standard deviation from exact two-fold symmetry is approximately 0.5Å. The standard errors of the occupancies (Z) are approximately 2 electrons and the errors on the positional coordinates are 0.1 to 0.2Å, as calculated from the phase refinement.

4. Electron Density Map Calculation

The 2.8Å resolution three dimensional electron density map of native α -CHT was computed using the 8900 reflections with figures of merit greater than 0.30 as determined during phase calculations. The 'best' Fourier synthesis (34) was calculated in sections perpendicular to the x axis in grid divisions of a/76, b/106 and c/130. The 'best' Fourier uses coefficients of the type

where r
angle.
undistc
were dr
 0.5 eA^{-1}
plots w
and mou
stacked
purpose
the mod
the two
using K
Repetit

$$m |F_p| \exp (i\alpha_{\text{best}}) \quad (7)$$

where m is the figure of merit and α_{best} is the best phase angle. The resultant electron density was printed in an undistorted way on a scale of 2cm per Angstrom. Contours were drawn by a Calcomp plotter with the lowest contour at 0.5 eA^{-3} and with subsequent intervals of 0.25 eA^{-3} . The plots were traced onto transparent cellulose acetate sheets and mounted on plexiglass sheets. The sheets were then stacked into a "Richards' Optical Box" for model building purposes (35). A half-silvered mirror arrangement allowed the model to be fitted to the electron density. Models of the two α -CHT molecules in the asymmetric unit were built using Kendrew atomic skeletal units produced by the Cambridge Repetition Engineers, Ltd.

III. DESCRIPTION OF THE MOLECULE

1. Folding of the Polypeptide Chain

Molecules of α -CHT are packed four per unit cell, two per asymmetric unit. The two molecules in the asymmetric unit come into close physical contact and are related to each other by a non-crystallographic two-fold axis running nearly parallel to the crystallographic a^* direction. Schwert (36) and Schwert and Kaufman (37) observed that the sedimentation coefficient of α -CHT was concentration and pH dependent, suggestive of the formation of a monomer-dimer equilibrium. Molecular weight measurements verified the existence of the dimeric unit. Because of the intimate association in solution and in the crystal, the two molecules in the asymmetric unit can be properly termed a dimer. Many close contacts obscure the molecular boundaries in the dimer interfacial region. However, the boundaries elsewhere are generally well defined, with solvent regions of low electron density between adjacent dimeric molecules. Some small peaks present in these channels and near the protein molecule suggest the presence of ordered solvent molecules.

The dimeric unit formed from two molecules of α -CHT is roughly an elongated ellipsoid with slight cusps in the surface of the ellipsoid near the interface region. The ellipsoid is approximately 45A by 33A by 67A with the axes of

the ell

tions a

molecul

individ

spheric

and an

the mol

presents

cut out

rough cr

resides

The

clearly

whole mo

anchored

are unob

are at t

actions

with ne

occur on

group of

the poly

overlap

to hydro

of resol

Nea

chain or

the ellipsoid corresponding to the crystallographic directions $\underline{a^*}$, $\underline{b}$ and $\underline{c}$. Figure 4 shows the shape of the dimeric molecule obtained by considering the Kendrew model. The individual molecules in the asymmetric unit are more nearly spherical except along the side where the dimer is formed and an indentation is made in the sphere. While the rest of the molecular surface is relatively smooth, the contact region presents a rough appearing surface, as though a swath was cut out of the otherwise spherical molecule. It is in this rough crevice-like region that the active site of the enzyme resides.

The path of the polypeptide chain of the enzyme can be clearly discerned from the electron density for nearly the whole molecule. Only the last four residues of the A chain, anchored with a disulfide bridge between residues 1 and 122, are unobserved in the electron density map. These residues are at the top of the molecule and make no specific interactions to stabilize them with the rest of the molecule or with neighboring molecules. In many places, bumps which occur on the backbone density correspond to the carbonyl group of the amino acid residues. In those regions where the polypeptide chains are in close proximity, there is some overlap of the electron densities from the chains, due either to hydrogen bonding between the chains or possibly to a lack of resolution.

Nearly all of the side chains are resolved from the main chain or form obvious protuberances from the main chain

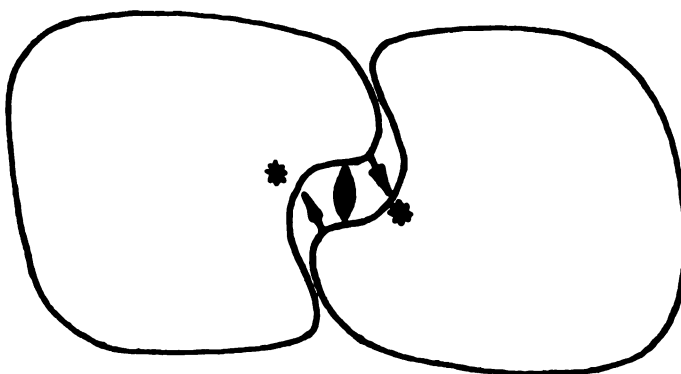


Figure 4. Approximate shape of 'dimeric' molecule of α -CHT. A local two-fold axis between the two molecules is indicated by the $\circ$ symbol. Interactions between the two molecules represented by arrows. The asterisk shows approximate location of the active site.

elect

ially

defin

found

electr

the no

dues a

the si

molecu

less w

side c

into t

T

the re

chymot

to whe

the su

an ind

densit

densit

observ

althou

others

is sun

occurr

Also,

ments

electron density. The large bulky aromatic groups, especially tryptophanes and phenylalanines, are very clearly defined. The histidine and tyrosine side chains, frequently found near the surface of the molecule, have characteristic electron densities but are somewhat less well defined than the nonpolar aromatic groups. In many cases, branched residues appear as regions of forked electron density. Although the side chain definition is best within the interior of the molecule, the exterior side chain positions are only somewhat less well defined. Surprisingly, nearly all of the lysine side chains are observed, even though most of them extend into the solvent.

The appendix gives the amino acid sequence of α -CHT, the residues being numbered according to the sequence of chymotrypsinogen. Each side chain is classified according to whether it is in the interior of the molecule (I), on the surface (S) or exposed to the solvent (E). In addition, an indication of the clarity of each side chain in the electron density map is given. This table is based on the electron density of one of the molecules in the asymmetric unit; the observations are generally applicable to the other molecule, although the resolution of some side chains is better and for others is worse in the other molecule of the dimer. The appendix is summarized in Table II. The distribution of the commonly occurring amino acids with side chains in α -CHT is given. Also, the distribution of each residue in the three environments specified above and the clarity of each residue is

Table II. Summary of Side Chain Environments and Clarity in the Electron Density of α -CHT.

Residue	Total	Environment*			Clarity#		
		I	S	E	Good	Fair	Poor
ALA	22	7	10	5	10	10	2
ARG	3	0	1	2	1	2	0
ASN	13	0	8	5	7	5	1
ASP	9	2	3	4	4	3	2
CYS	10	2	8	0	10	0	0
GLN	10	1	4	5	5	3	2
GLU	5	0	1	4	3	0	2
HIS	2	0	2	0	2	0	0
ILE	10	4	6	0	5	5	0
LEU	19	13	3	3	11	6	2
LYS	14	0	3	11	7	7	0
MET	2	1	0	1	1	1	0
PHE	6	1	4	1	3	3	0
PRO	9	3	6	0	8	1	0
SER	27	5	7	15	17	5	5
THR	22	4	11	7	12	7	3
TRP	8	4	4	0	6	2	0
TYR	4	1	2	1	2	2	0
VAL	22	13	8	2	8	12	3

*Environment: I - Interior, not readily accessible to solvent;
 S - On surface of molecule, partially accessible to solvent;
 E - Extending into solvent.

Clarity: Good - Complete and definitive electron density for side chain.
 Fair - Reasonable density for side chain, configuration less well defined.
 Poor - Little or no electron density for side chain.

tabulated. It can be seen that only a few residues are poorly described in the electron density map. Overall, about 85-90% of the amino acids have fair to good side chain densities.

While the backbone of the protein was generally clearly defined, the model building process depended heavily upon the knowledge of the chemical sequence for the identification of the side chains. The five disulfide bridges are readily apparent as cylinders of relatively high electron density (approximately $2-3 \text{ electrons/\AA}^3$) forming a bridge between two adjacent polypeptide chains.

Three regions of the polypeptide chain posed substantial problems when attempting to fit the sequence to the electron density: residues 10-13 had no apparent electron density, implying no preferred conformation; residues 67-70 on the edge of the molecule in a solvent region; and residues 34-37 were difficult to fit to the observed density in one molecule. In the latter case, the conformation was determined as implied from the electron density of the second molecule for this region.

The path of the polypeptide chain through the globular shaped molecule consists of stretches of fully extended peptide chain punctuated with abrupt changes in the direction of the peptide chain, with the chain doubling back upon itself. The path of the chain appears to be random, producing the resultant globular shaped molecule. Approximately

one half of the molecule is formed by the folding of the B chain upon itself with the other half formed by the C chain. The two chains occasionally intermingle for short stretches of the polypeptide chain. Short regions of the classical types of protein chain conformation--helix and pleated sheet--are observed in the molecule. There are only two short regions of the peptide chain which form helices; (1) a well defined α -helix of approximately 3 turns involving the residues LEU 234 to ASN 245 at the carboxyl terminal of the C chain on the surface of the molecule away from the dimer interface and (2) residues SER 164 to GLY 173, also on the surface of the molecule, forming a short, distorted helix of about 1.5 turns in length. This helix appears to be a hybrid of an α -helix and a 3_{10} helix. Several regions of extended peptide chain are found on the surface and through the interior of the molecule. Three of the longest extended chain regions consist of residues 80 to 93, 100 to 115 and 154 to 164. In several places, the extended chains are arranged in an antiparallel fashion with interchain hydrogen bonding to form short pieces of pleated sheet. One such pleated sheet showing a slightly curved surface is formed by the chain segments 85 to 95, 105 to 110 and 52 to 55.

2. β -Bends

The peptide chain can change direction abruptly by forming ' β -bends' as described by Venkatachalam (38). The β -bends, similar to 3_{10} helical bends, are produced by a

tight hydrogen bonding system between the carbonyl oxygen of the first residue and the amide nitrogen atom of the fourth residue in a tetrapeptide sequence. In α -CHT, there are twelve well defined β -bends which are listed in Table III, with several more peptide segments in a configuration similar to a β -bend but not quite so well defined. In agreement with the calculations of Venkatachalam, those bends where the amide plane between the second and third residues is flipped 180° (type II) with the carbonyl oxygen pointing towards the side chains have a glycine as the third residue. Similar peptide conformations have been observed in nearly all proteins whose structure have been reported, the most recent being carbonic anhydrase (39) and ferricytochrome C (1). It has been suggested that β -bends might function as 'directing' sections to provide proper orientation of distant segments of the polypeptide chain during the initial protein folding process (40). Figures 5a and 5b show stereoscopically the two types of bends which occur in α -CHT.

3. The Active Site Region

The active site of the enzyme is found along the side of the molecule which forms the interface between the two molecules. It is situated at the bottom of a 'jaw like' depression formed by a junction of the two halves of the folded molecule just off the surface of the molecule. The site is partially screened on two sides from the solvent by several amino acid residues arranged as a partial cover to the

Table III. Best Defined β -Bends in α -CHT.

Residues	Sequence	Type*
23-26	VAL-PRO-GLY-SER	II
27-30	TRP-PRO-TRP-GLN	I
35-38	ASP-LYS-THR-GLY	I
48-51	ILE-GLU-ASN-TRP	I
56-59	ALA-HIS-CYS-GLY	I
115-118	SER-GLN-THR-VAL	I
177-180	LYS-ASP-ALA-MET	I
191-194	CYS-MET-GLY-ASP	II
194-197	ASP-SER-GLY-GLY	II
203-206	LYS-ASN-GLY-ALA	II
217-220	SER-SER-THR-CYS	I
221-224	SER-THR-SER-THR	I

*See reference 38.

depression. On the top are the residues 190 to 193 and 215 to 219, while the bottom is closed by residues 94 to 97. The amino acid residues which form the active site components have been unambiguously identified by chemical and crystallographic studies on α -CHT and other homologous enzymes such as trypsin (41) and elastase (42).

Reaction of α -CHT with di-isopropylfluorophosphate was found to selectively phosphorylate a specific serine residue, SER 195, completely blocking all activity of the enzyme (43). Also, the sulphonyl fluoride compounds of Fahreny and Gold (44) were found to form ester linkages with SER 195. Studies of the catalysis of p-nitrophenyl-acetate (45) indicated that the reaction proceeds through an acyl-serine

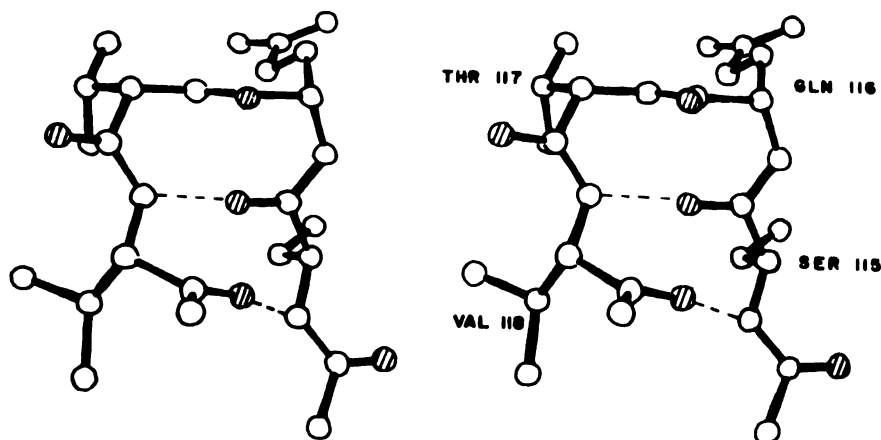


Figure 5a. Stereo drawing of a Type I β -Bend formed by residues SER 115 to VAL 118. Main chain carbonyl oxygen atoms indicated by lines through the atoms. Hydrogen bonds formed between carbonyl oxygen and amide hydrogen indicated by dotted lines. The carbonyl group of GLN 116 at the top of the loop is directed out of the page, while the adjacent side chains are directed into the page.

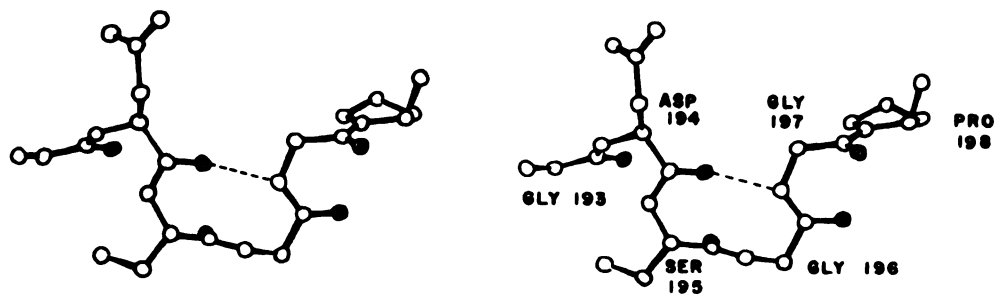


Figure 5b. Stereo drawing of a Type II β -Bend formed by residues ASP 194 to GLY 197. SER 195 carbonyl directed in same direction as side chains.

intermediate. Direct evidence for the involvement of a histidine residue (HIS 57) was provided by Schoellman and Shaw (46) with the specific alkylation of the imidazole ring of the histidine with loss of enzyme activity by tosylamino-2-phenylethylchloromethylketone.

In addition to the primary requirement of serine and histidine for activity, there are several secondary requirements also. It has been shown that the activity is dependent upon the protonation of the N-terminal amino group of isoleucine 16 (14,47). The carboxylate ion of ASP 194, directly adjacent to SER 195, forms an internal salt bridge with the former. This salt bridge in some way controls the conformation of the active site and its maintenance is crucial to the activity of the enzyme. One of the changes which has been observed when chymotrypsinogen, the inactive precursor, is converted to the active α -CHT by tryptic cleavage of the bond adjacent to ILE 16, is the repositioning of the side chain of ASP 194. In chymotrypsinogen, this side chain is said to be hydrogen bonding to HIS 40 and, upon conversion to the active form, the side chain swings up to form a buried salt bridge with the newly created N-terminal amino group of ILE 16 (16). The formation of this bridge may be a controlling step that converts the inactive form to the active form of the enzyme.

The enzyme becomes inactive at pH values higher than 8 to 9. Various experiments have shown that ILE 16 becomes deprotonated at these pH values leading, presumably, to the

disruption of the stabilizing ion pair and to changes of the active site conformation (49). By analyzing stop flow kinetic data of the binding of proflavin to α -CHT as a function of pH, Fersht was able to estimate the energy of the salt bridge between ASP 194 and ILE 16 to be approximately 2.9 kcal/mole (50).

The other residue in the active site region which has been implicated as necessary for activity is ASP 102 (11). ASP 102 is located somewhat removed from the surface of the enzyme and relatively inaccessible to solvent molecules larger than water. Depending upon its ionization state, and that of HIS 57, the carboxylate ion can either form a salt bridge to HIS 57 or hydrogen bond with it. Alternatively, the carboxylate ion can also form hydrogen bonds to the hydroxyl group of SER 214 or the main chain amide of HIS 57. Chemical attempts to block the carboxyl groups of α -CHT with glycine methyl ester result in 15 of the 17 carboxyls of α -CHT being modified. One of the unmodified carboxyl groups has been shown to be ASP 194, involved in the ion pair with ILE 16 (51). The other, unmodified, carboxyl group is presumably ASP 102 since it is located in such a position that it is relatively inaccessible to the reagent.

A schematic drawing of the approximate orientations of the catalytically important groups around the active site and a possible hydrogen bonding network linking the various residues together is shown in Figure 6. The conformation of the HIS 57 and SER 195 side chains is such that, directionally,

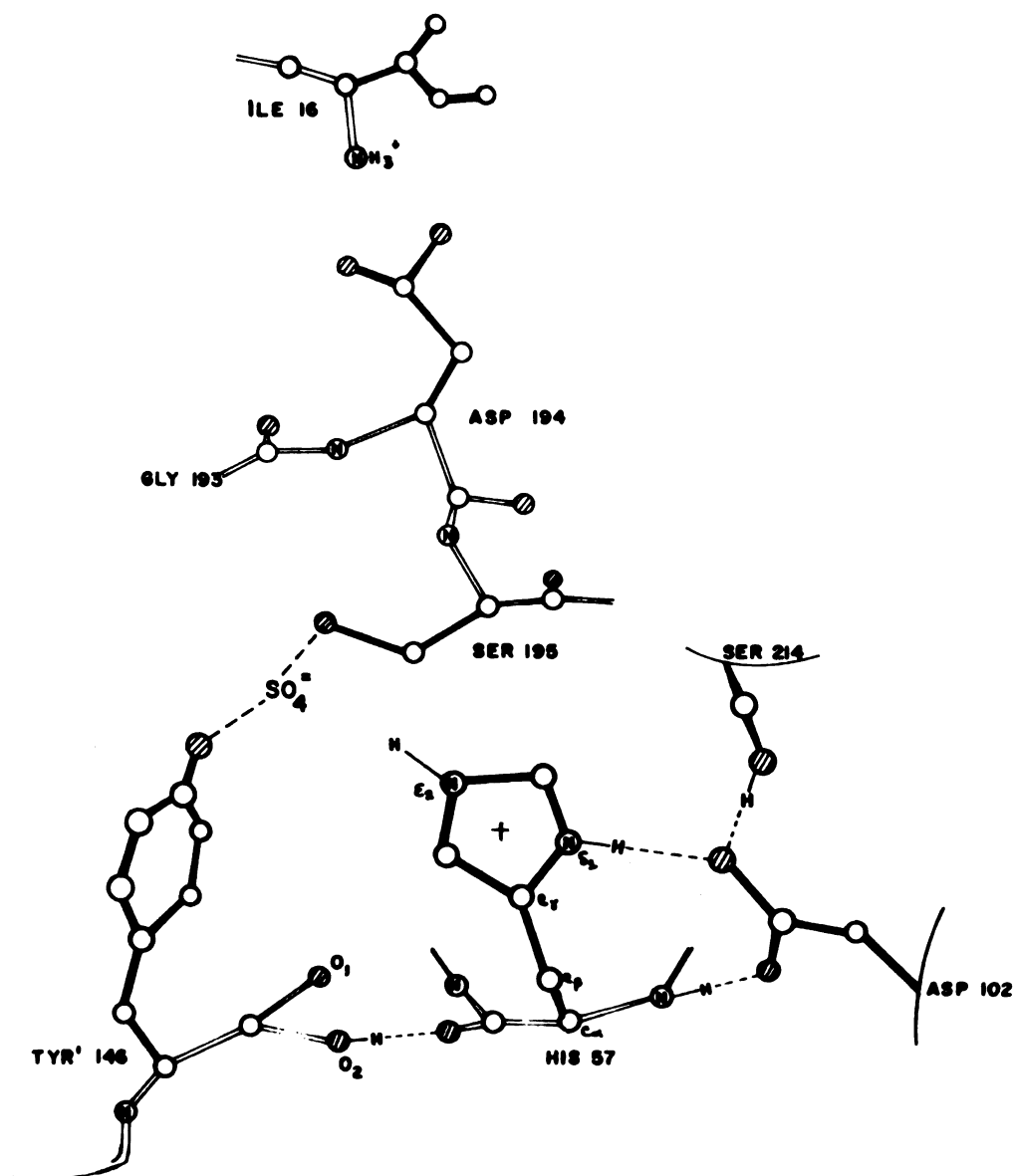


Figure 6. Schematic drawing of the important residues around the active site. The side chain atoms are connected with solid bonds, main chain atoms by open bonds. Bridging $\text{SO}_4^{=}$ molecule shown between SER 195 and TYR' 146. ⁴Possible hydrogen bonds indicated by dashed lines. Oxygen atoms - ; nitrogen atoms - ; carbon atoms - .

only a 'poor' hydrogen bond can be formed between histidine and serine. As shown, the carboxylate group of ASP 102 can form hydrogen bonds with the N_{δ2} nitrogen atom of HIS 57, the main chain amide of HIS 57, or to the hydrogen atom of the hydroxyl of SER 214. A sulfate molecule is also found in a position to hydrogen bond to SER 195. While these residues are directly involved with the chemistry of activity, there are undoubtedly many other residues in this region which are important in the alignment and binding of substrates to be hydrolyzed. Excellent review articles concerning the activity and specificity of α-CHT have been written by Hess (14,52).

4. Non-Polar Cavity

A very interesting structural feature became apparent when the model of the enzyme was built. Some seven different chain segments from the sequence come together to form the surface of a large, somewhat irregular, sphere with the side chains from 18 non-polar residues extending toward the center of the sphere. The interior of this sphere is an empty cavity approximately 8 to 9Å in diameter. All of the residues which participate in the formation of this 'oily' cavity have non-polar side chains consisting mainly of isoleucine, leucine, and valine groups. Three tryptophan residues are also involved, one on either end of the cavity, which seem to act as lids to the cavity, and one near the center, somewhat removed from the cavity surface. One side of the cavity

is bounded by the inner side of the α -helix at the carboxyl terminal of the C chain involving residues 231 to 242. Other chain segments which are involved are residues 47 to 53, 103 to 105, 121 to 124, and 209 to 212.

The non-polar cavity can be schematically represented as four shells or levels of residues as shown in Figure 7. Levels 1 and 4 form lids at either end of the cavity with levels 2 and 3 around the center of the cavity. The arrangement of residue names in each level approximates the side chain positions around the cavity. The cross-hatched region is empty space. Possible entrances to the cavity by solvent molecules are blocked by THR 241 and VAL 238.

The occurrence of such a large structural feature at some distance from the active site raises some interesting questions concerning the physical forces which interact during the folding of large biological molecules. Normally, the business end or active site of a molecule is located at a specific region on the enzyme molecule, involving the interaction of, usually, only a few critical amino acid residues. It is believed that the function of the rest of the molecule with its pleated sheets, helix, 3_{10} bends and other protein substructures is to hold the critical residues in the correct conformation for activity. In addition, these protein substructures may play an important role in directing the folding of the chain into its proper conformation of lowest free energy.

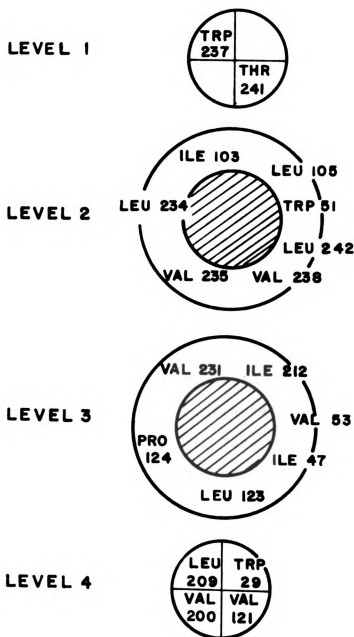


Figure 7. Schematic representation of non-polar cavity. Level 1 is near the surface of the protein and level 4 is in the interior. The cross-hatched regions correspond to the vacant space of the cavity.

The
the conf
the stru
evolutio
residues
polar ca
bin and
invarian
cavity.
servativ
shows th
with som
fied as
of these
a way th
while th
surface
parisons
because
served t
an impor
protein

5. Arom

Ano
an obviou
lar to th

The importance of this non-polar cavity to stabilize the conformation of α -CHT can be inferred by a comparison of the structures of other proteolytic enzymes thought to be evolutionarily related. A comparison of the amino acid residues of α -CHT involved in the formation of the non-polar cavity with those of various species of trypsin, thrombin and elastase (45) in Table IV shows a high degree of invariance in the amino acid residues concerned with the cavity. The substitutions which do occur are usually conservative, e.g., an isoleucine for leucine, etc. LEU 234 shows the greatest variability among the various species with some side chain substitutions which are usually classified as polar groups (lysine and tyrosine). The side chains of these polar residues could be positioned, however, in such a way that the polar regions are directed toward the solvent, while the non-polar regions of the side chain lay along the surface of the cavity. It becomes obvious from these comparisons that the hydrophobic stabilization which arises because of this arrangement is important enough to be conserved through evolutionary processes, and as such, must be an important factor in the tertiary conformation of these protein molecules.

5. Aromatic Clusters

Another protein substructure that manifests itself in an obvious manner is the clustering of aromatic groups similar to that observed in other protein structures (53,1). One

Table IV. Comparison of Residues Around Non-Polar Cavity in α -CHT to Several Homologous Enzymes.

Level	α -CHT	BT	DFT	Th	BCA	BCB	PE
1	TRP 237	TRP	--	TRP	TRP	TRP	TRP
	THR 241	THR	THR	VAL	THR	THR	VAL
2	TRP 51	TRP	--	TRP	TRP	TRP	TRP
	ILE 103	ILE	ILE	ILE	ILE	ILE	ILE
	LEU 105	LEU	LEU	LEU	LEU	LEU	LEU
	LEU 234	TYR	--	LYS	LEU	LEU	TYR
	VAL 235	VAL	--	LEU	VAL	MET	ILE
	VAL 238	ILE	ILE	ILE	VAL	VAL	ILE
	LEU 242	ILE	ILE	ILE	LEU	LEU	ILE
3	ILE 47	ILE	ILE	ILE	ILE	ILE	ILE
	VAL 53	VAL	--	--	VAL	VAL	MET
	LEU 123	LEU	LEU	--	LEU	LEU	LEU
	PRO 124	PRO	PRO	--	PRO	PRO	PRO
	ILE 212	ILE	ILE	ILE	ILE	ILE	VAL
	VAL 231	VAL	--	VAL	VAL	VAL	VAL
4	TRP 29	TYR	--	TRP	TRP	TRP	SER
	VAL 121	ILE	ILE	--	VAL	VAL	GLY
	VAL 200	VAL	VAL	VAL	VAL	VAL	HIS
	LEU 209	LEU	--	GLN	LEU	LEU	VAL

BT - Bovine Trypsin
DFT - Dogfish Trypsin
Th - Thrombin

BCA - Bovine CHT-A
BCB - Bovine CHT-B
PE - Pig Elastase

of the most pronounced groupings is that formed by the three tryptophan residues 27, 29 and 207. These residues form a nest-like arrangement with the proline residues 4, 8, and 28, acting as additional boundaries to the nest. Figure 8 shows this tryptophan arrangement.

A number of crystallographic experiments have been performed to study the interaction of aromatic substrate-like molecules and competitive inhibitors with α -CHT. It has been observed that, in addition to binding of these small molecules in the active site region, many of these aromatic molecules also are localized at another specific site on the enzyme (54). This site has been identified to be near this tryptophan nest.

Other aromatic clusters are found with HIS 40 and TRP 141. Their side chains are nearly parallel to one another with a distance of approximately 5Å between the two rings. Somewhat more distant, but nearly coplanar to the above residues is PHE 41. Near one end of the non-polar cavity, TRP 51, PHE 89 and TRP 215 are found clustered together near the active site region. As observed with the non-polar cavity, nearly all of these aromatic residues are invariant throughout the related proteolytic enzymes compared, with any substitutions being, again, conservative in nature. A comparison of these residues for several related enzymes is given in Table V.

If one considers the dimeric molecule of α -CHT, then two other aromatic clusters are formed. Near the active site,

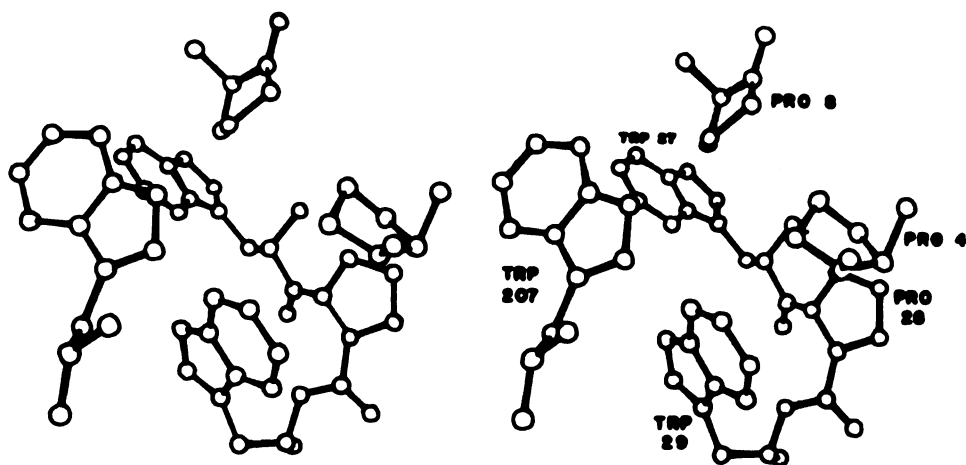


Figure 8. Stereoscopic drawing of the tryptophan aromatic cluster.

the aromatic side chains of HIS 57 from one molecule and TYR 146 from the other come into close proximity with their rings lying nearly in the same plane. Also in the interface region, there are two phenylalanine 39 residues, one from each molecule, which must interact with each other.

Table V. Comparison of Residues in Aromatic Clusters.*

Cluster	α -CHT	BT	DFT	Th	BCA	BCB	PE
1	PRO 28	PRO	--	PRO	PRO	PRO	PRO
	TRP 27	VAL	--	SER	TRP	TRP	TRP
	TRP 29	TYR	--	TRP	TRP	TRP	SER
	TRP 207	--	--	TRP	TRP	TRP	TYR
2	HIS 40	HIS	HIS	LEU	HIS	HIS	HIS
	PHE 41	PHE	PHE	GLN	PHE	PHE	PHE
	TRP 141	TRP	TRP	TRP	TRP	TRP	TRP
3	TRP 51	TRP	TRP	TRP	TRP	TRP	TRP
	PHE 89	ILE	ILE	ILE	PHE	PHE	VAL
	TRP 237	TRP	TRP	TRP	TRP	TRP	TRP
4	TYR 171	ALA	ALA	TYR	TYR	TYR	TYR
	TRP 172	TYR	TYR	PHE	TRP	TRP	TRP
	TRP 215	TRP	--	TRP	TRP	TRP	PHE

*Enzyme abbreviations same as in Table IV.

IV. MOLECULAR INTERACTIONS

1. The Structure of the α -CHT Dimer

The crystallographic asymmetric unit of α -CHT contains two enzyme molecules in intimate association with one another. The two molecules have been found to be related to one another by an approximate two-fold rotation axis which is nearly parallel to the a^* direction. The location of this non-crystallographic symmetry element with respect to the crystallographic axes was obtained by a comparison of the electron densities of the two molecules.

The transformation between the two molecules can be expressed by

$$\vec{x}_2 = C \vec{x}_1 + \vec{d} \quad (9)$$

where each point x_1 in one molecule is related to the point x_2 in the second molecule by a rotation matrix C and a translation vector $\vec{d}$. Values of C and $\vec{d}$ are sought which minimize

$$\sum (\rho_1(x_1) - \rho_2(x_2))^2 \quad (10)$$

where ρ_1 and ρ_2 are the electron density values at the points x_1 and x_2 , respectively. The rotation matrix C contains coefficients which are functions of the three independent Eulerian angles ϕ , ψ , θ and d is a column vector containing the translation components d_1 , d_2 , and d_3 with respect to a set of reference axes. Given an initial set of these variables,

refined values can be obtained from the least squares minimization of equation 10.

Initial estimates for these variables were obtained by taking portions of the electron density corresponding to a peptide chain from one molecule, superimposing it on the second molecule and recording the necessary transformation parameters. Because the rotation axis is perpendicular to the yz planes of the electron density, the two molecules can be compared section by section in this direction, and approximate transformation parameters for each section obtained.

Refinement of the transformation parameters between the two molecules of α -CHT in the asymmetric unit was performed using a computer program written by Dr. Carl Morimoto. The program was based on one used by Cohen, et al. (56), which was used in a low resolution comparison of the electron densities of γ -CHT and α -CHT. The 2.8Å resolution electron density map was stored in planes perpendicular to the a^* direction. A molecular boundary was defined so that during the refinement, only those portions of the electron density which corresponded to the two molecules of interest were used. This molecular boundary excluded intermolecular solvent regions and portions of the electron density associated with neighboring molecules. The final, refined values of the transformation parameters were

$$\phi = 0.0^\circ$$

$$\psi = 179.60^\circ \pm 0.02^\circ$$

$$\theta = 0.0^\circ$$

$$\begin{aligned}
 d_1 &= -0.207 \pm 0.009\text{\AA} \\
 d_2 &= 39.87 \pm 0.01\text{\AA} \\
 d_3 &= 49.39 \pm 0.01\text{\AA}
 \end{aligned}
 \tag{11}$$

The position of a local two-fold axis can be derived from these parameters. For a given x coordinate, the two-fold axis passes through the points y and z at

$$z = 0.1565 (\pm 0.0009) x + 0.3683 (\pm 0.0019) \tag{12}$$

and $y = 0.2957 (\pm 0.0018)$

where x, y and z are expressed in terms of fractional units of the crystallographic unit cell. Because of the choice of reference axes for the transformation parameters, it was not possible to refine the angles ϕ and θ meaningfully. The estimated standard deviations were obtained from the variance/covariance matrix of the last cycle of refinement. The values for d_1 , d_2 , and d_3 are in Angstrom units with respect to the a^* , b , and c directions of the crystal.

The assessment of the quality of agreement between the two molecules is expressed in terms of the root mean square difference between them.

$$[\sum (\rho_1(x) - \rho_2(x))^2]^{1/2} \tag{13}$$

The rms difference using 102,468 electron density grid points within the molecular boundary was 0.27 electrons per cubic Angstrom as compared to an rms electron density of $0.407 \text{ e\AA}^{-3}$. The rms difference is due to structural differences between the two molecules and to errors in the electron density map. It was observed that there were a substantial number of

regions within the molecular boundary for which the difference in electron densities was significantly greater than the rms difference. Another least squares calculation was performed excluding regions where the differences in electron density between the two molecules ($|\rho_1(x) - \rho_2(x)|$) was greater than 0.70 eA^{-3} . About 13% of the electron density points within one molecule were removed by this criterion. The calculated position of the two-fold axis remained essentially the same, but the rms discrepancy factor decreased to 0.24 eA^{-3} while the rms electron density remained nearly the same (0.40 eA^{-3}).

As a measure of the agreement between the electron densities of the two molecules, a correlation coefficient was computed during the least squares refinement procedure to indicate the progress of the refinement. The correlation coefficient c_{12} was computed as

$$c_{12} = \frac{\sum (\rho_1 - \bar{\rho}_1) (\rho_2 - \bar{\rho}_2)}{[\sum (\rho_1 - \bar{\rho}_1)^2 \sum (\rho_2 - \bar{\rho}_2)^2]^{1/2}} \quad (14)$$

where ρ_1 and ρ_2 are the electron densities of molecule 1 and 2, $\bar{\rho}_1$ and $\bar{\rho}_2$ are the average electron densities, with the summations being taken over all points within the molecular boundaries. For the case where the electron densities are exactly the same, c_{12} would have a value of 1.0; no correlation between the electron densities would be represented by a value of 0.0. By its definition, this coefficient is independent of any relative scale factor between the electron densities

being compared (56). The correlation coefficient between the two molecules using the initial transformation parameters was 0.772. The coefficient increased to 0.82 after one cycle of refinement; this indicated an obvious improvement in the initial parameters. Further refinement did not improve this value. The final values for the various agreement factors are summarized in Table VI. In addition, values for the various indicators are given for those calculations which were made where electron density points less than 0.25 eA^{-3} were excluded. The indicators in this case behaved in a manner comparable to those obtained from calculations made using all points within the molecular volume.

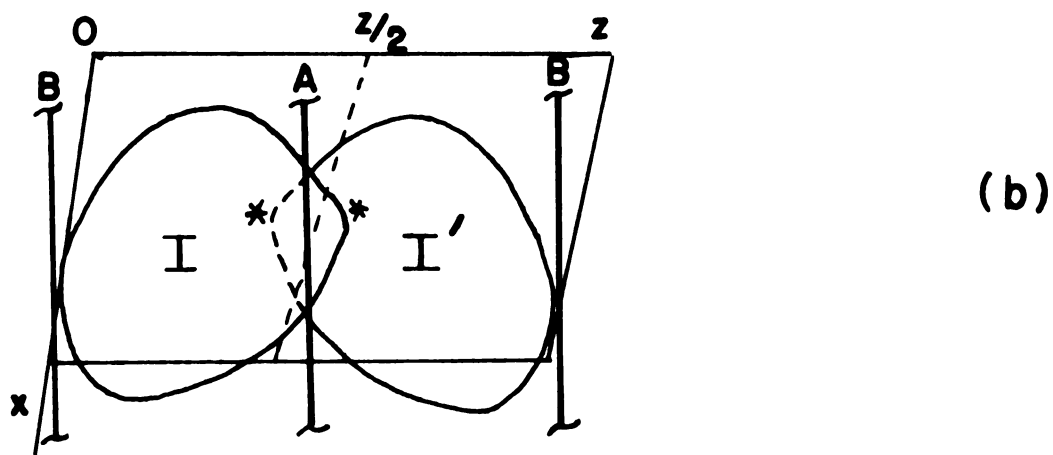
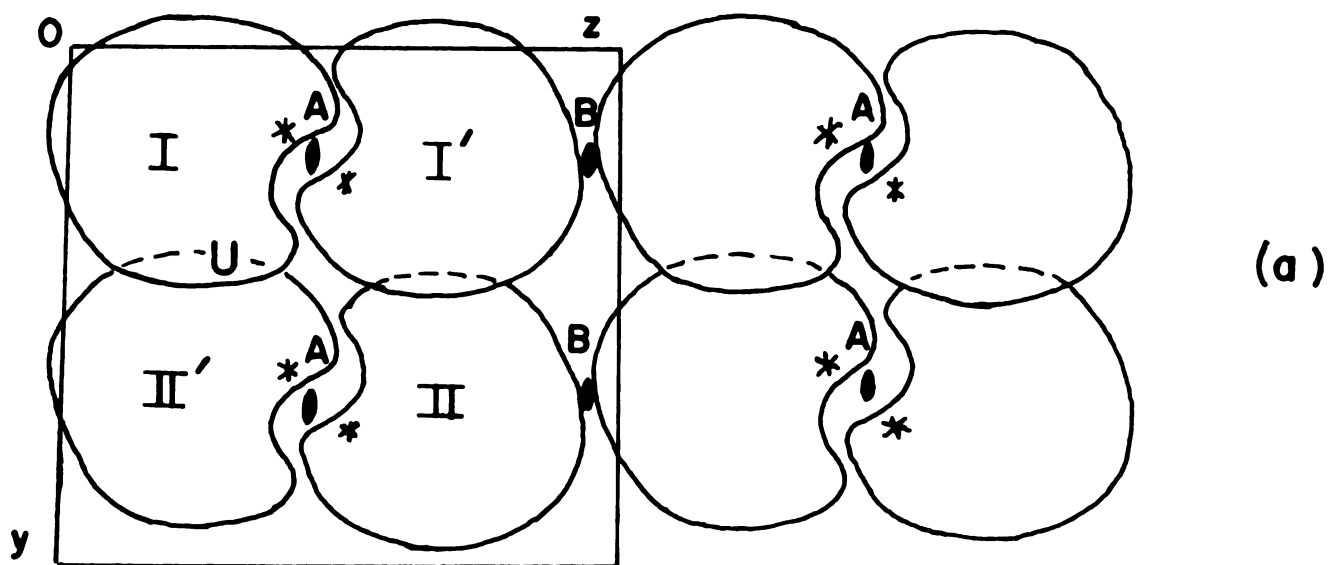
Table VI. Transformation Refinement Summary.

Calculation*	$\Delta \text{rms} (\text{eA}^{-3})$	$\bar{\rho} \text{rms} (\text{eA}^{-3})$	c_{12}	Number of Points
1.	0.269	0.407	0.782	102,468
2.	0.244	0.400	0.814	88,734
3.	0.300	0.499	0.820	63,780
4.	0.268	0.495	0.854	53,681

- *1. Calculations made including all points within the defined molecular boundary.
2. Same as 1 except excluding regions containing differences in electron density $>0.7 \text{ eA}^{-3}$.
3. Same as 1 but excluding electron density points $<0.25 \text{ eA}^{-3}$.
4. Same as 3 but excluding regions containing differences greater than 0.7 eA^{-3} .

The presence of one local two-fold axis along with the crystallographic two-fold screw axis generates another local two-fold axis, related to the first by a translation of one-half the unit cell along the z direction. Because these symmetry elements are not crystallographic elements, the two molecules related by the axis are not required to be identical. As will be seen later, the two independent molecules show many chemical and physical differences in reactivity, structure, sulfate binding and other effects. In addition, the environments of the two-fold axes are different.

Figures 9a and 9b show schematically the packing of the molecules projected on the yz and xz planes, respectively. Molecules I and I' form the asymmetric unit and are related by the local two-fold axis. Molecules II and II' belong to the second asymmetric unit and are related by the crystallographic two-fold screw axis to molecules I and I', respectively. The two dyad axes, A and B are indicated. It can be seen that dyad A, defined by equation 12, runs through the middle of the 'dimeric' molecule of α -CHT. The individual molecules, when related by a two-fold axis, are complementary in shape and fit together rather intimately. Between the close contacts at the top and bottom of the two monomeric molecules (Figure 9b), a large, partially open region is formed. The region is generally accessible to solvent and to smaller substrate and inhibitor molecules. It is generally hydrophilic in nature, with SER 195, ASP 35, GLN 73, HIS 57, TRY 146 and other polar residues within the region. In addition, several



* - ACTIVE SITES U - URANYL BINDING SITE A,B - DYADS

Figure 9. Schematic packing diagrams of α -CHT in the unit cell; (a) unit cell contents projected onto yz plane; (b) projection onto xz plane.

water molecules and sulfate ions are bound in the cavity between the two molecules.

2. Local Two-Fold Interactions

The molecular interactions which occur along this interface region around dyad A can be classified into two specific types: electrostatic and non-electrostatic, the latter including the effects of hydrophobic interactions and hydrogen bonding. The electrostatic interactions consist mainly of 'salt' bridges formed by a negatively charged group on one molecule and a positively charged group on the other. The position and orientation of the groups allow the opposing charges to be effectively neutralized, and results in a significant net attractive force between the ionic groups. It is expected that these salt bridge interactions would be dependent upon the pH of the surrounding medium; as the various residues involved become protonated or vice versa, the ionic interactions would be disrupted.

The various interactions which occur between the two molecules are diagrammed in Figure 10. The residue names are in an approximately correct position with respect to the two-fold axis position indicated by the horizontal line. The residue names with the prime symbol (Molecule I' in Figure 9a) belong to the second molecule in the asymmetric unit and all references to the second molecule hereafter will be denoted by this symbolism. The interactions which are purely electrostatic are those involving the carboxylate group of ASP' 64

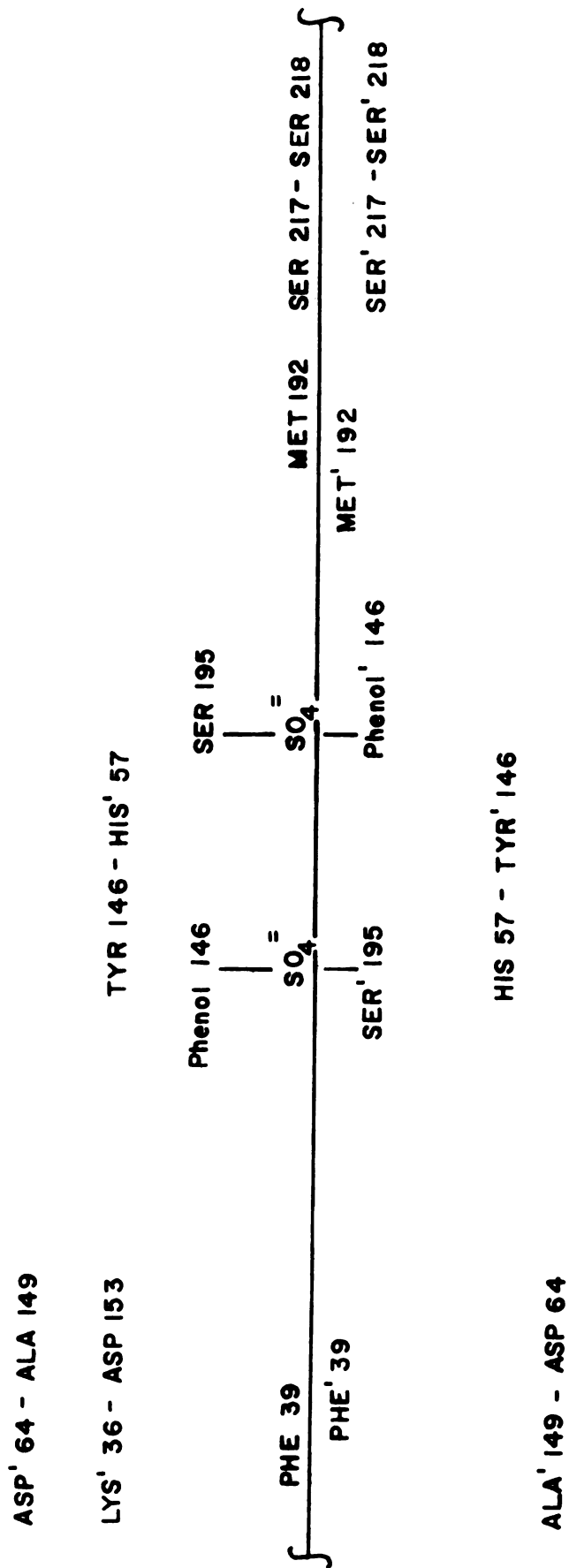


Figure 10. Intermolecular interactions across dyad A. Interactions are between molecule I (unprimed) and I' (primed residue names). The interaction between Phenol 146 and SER' 195 and the reciprocal interaction are shown displaced from one another for clarity.

and the amino terminal of ALA 149, the ϵ -amino group of LYS' 36 and the carboxylate of ASP 153, and the carboxyl terminal of TYR 146 and the imidazolium ion of HIS' 57, although the spatial orientation of the groups involved in the latter interaction is not the most favorable for complete neutralization of the respective charges. Each of the corresponding two-fold related electrostatic interactions is well defined also, except the LYS 36 - ASP' 153 interaction.

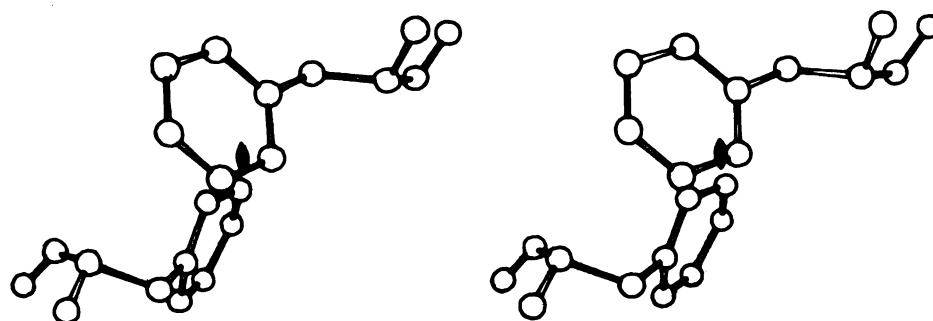
In addition to these direct ionic interactions, sulfate ions participate in forming a hydrogen bonding bridge between the phenolic hydroxyl of TYR' 146 and the hydroxyl group of SER 195. The location of the sulfate ion between these two groups, discussed in more detail in a subsequent section, is also close to the HIS 57 side chain and might play a role in neutralizing the charge which would be present on the imidazole ring of HIS 57 at pH 4, the pH at which the crystals were grown. Another $\text{SO}_4^{=}$ ion is found in the vicinity of LYS' 36, ALA 149 and ASP' 64, participating in a complicated charge interaction and hydrogen bonding scheme.

The non-electrostatic interactions are located close to the dyad axis and involve the same residues from both molecules. In the example of the PHE 39 self-interaction, the two-fold axis passes nearly through the center of the ring of one of the residues. If two exact monomers of α -CHT were brought together to form a dimer, the phenylalanine rings of the two PHE 39 residues would overlap one another. To avert this impossibility, the rotations of one or both of the

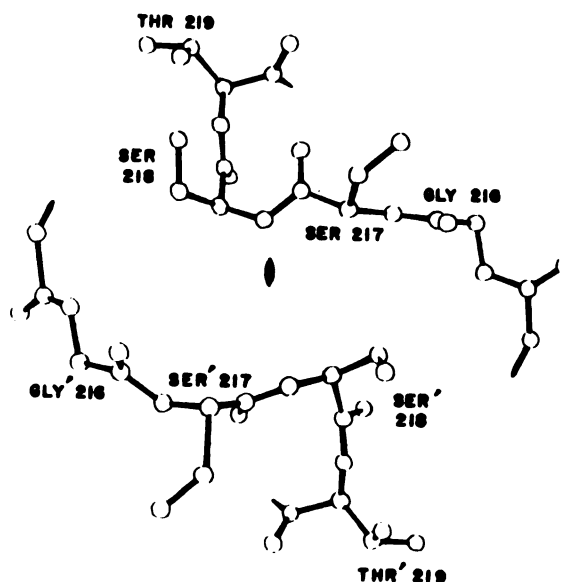
phenyl rings occur so that the final position of each ring is related to the other by a 180° rotation around the $C_\alpha-C_\beta$ bond, to locate the second ring at a sufficient distance from the first to preclude appreciable steric hindrance. This necessary rotation does not appear to affect the neighboring residues in the chain significantly, although HIS' 40 is in a slightly different conformation than its counterpart in the first molecule. Figure 11a shows the final arrangement of the two phenyl rings.

By comparing the electron densities around the active site, some substantial differences were observed. The largest differences occurred around the interacting methionine 192 residues (Figure 12). If the sulfur atoms showed exact two-fold symmetry, an extremely close S-S contact distance of approximately 3Å would be formed. To relieve this close interaction, one methionine residue is translated approximately 1.5Å with respect to the other along the a^* direction.

A similar rearrangement in the two molecules is observed around the residues GLY 216 - SER 217 - SER 218. The chain segments from the two molecules corresponding to these residues would approach too close to one another if exact two-fold symmetry were displayed. As a result, one chain becomes displaced in a perpendicular direction from the two-fold axis. This arrangement is shown in Figure 11b with the residues from the second molecule significantly displaced from the two-fold axis.



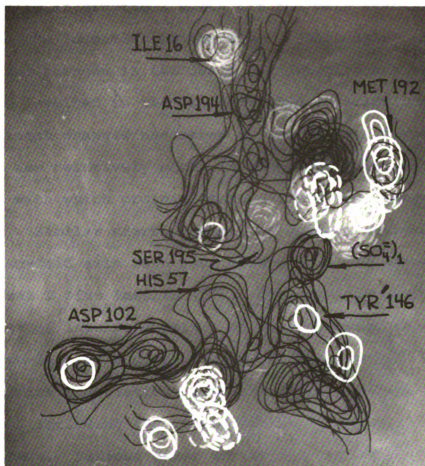
a



b

Figure 11. Local two-fold interactions between the two molecules. (a) Stereo drawing showing final arrangement of PHE 39 and PHE' 39. (b) Section of peptide chain near two-fold axis, ●. The lower peptide chain is obviously displaced from the two-fold axis and assumes a slightly different conformation.

Figure 12. Electron density of the active site region of one molecule. Contours at 0.25 eA^{-3} beginning at 0.50 eA^{-3} . Difference electron density between the two molecules shown in white; difference contours at 0.15 eA^{-3} beginning at 0.6 eA^{-3} . Solid white contours represent excess density in the molecule shown while broken contours represent excess density in the other molecule.



It is interesting to note that some of the interactions which are electrostatic in nature show, in general, precise local two-fold symmetry, while all of those involving non-ionic interactions do not show the same precise symmetry. It can be presumed from an energetic viewpoint that it is the electrostatic interactions which are the main stabilizing forces for the formation of the dimer. Obviously, the forces required for changes in the chain conformation are smaller than those involved in the electrostatic interactions. The conformational changes necessary to accommodate formation of the dimer are relatively minor in scope but point out the readiness with which protein conformations can change when necessary. Similar rearrangements of several residues around a local two-fold axis have been observed in the structure of rhombohedral 2 Zinc Insulin (57). The crystallographic asymmetric unit of insulin consists of two molecules related by a local two-fold axis perpendicular to a three-fold axis. Some significant deviations from the local two-fold symmetry were noted, including the reorientation of a phenyl ring as observed here. Various other deviations were observed which occurred at substantial distances from the rotation axis.

The other two-fold axis (B) relates dimeric molecules to other dimeric molecules in the crystal lattice. There are no specific, well-defined interactions across this two-fold axis as there are across dyad A. Only one interaction, possibly due to hydrogen bonding between the side chain amides

of GLN 204 and GLN' 240 and the corresponding two-fold interaction, could exist. The environment around this axis consists almost entirely of solvent.

3. Location of Bound Sulfate Ions

After a plausible model of α -CHT had been constructed, several regions of electron density for which no protein residues were assigned remained. It has been common practice in protein crystallography to label unidentified peaks as water molecules or, generally, solvent molecules with no attempt to differentiate between water and salt ions. Since many protein crystals are grown from relatively high ammonium sulfate solutions, it seems reasonable that some of the solvent peaks could well be localized sulfate ions.

To definitively identify those peaks which were due to sulfate ions, an attempt was made to exchange selenate ions for sulfate ions. Ammonium selenate and ammonium sulfate have similar molecular properties, including sizes, ionization constants and solubilities in water. Since the selenate ion has 18 more electrons than the sulfate ion, the difference in X-ray scattering power of the selenate and sulfate ions should be observable in the diffraction pattern of the protein if displacement occurs.

Ammonium selenate, purchased from K and K Laboratories, was purified by repeated crystallization from water before use. The crystal soaking solution of 75% saturated $(\text{NH}_4)_2\text{SO}_4$ was slowly exchanged with an equivalent $(\text{NH}_4)_2\text{SeO}_4$ solution.

Initially, the crystals were observed to crack, but upon sitting for several days, the cracks had disappeared and the crystals appeared to 'heal'. No apparent changes in the X-ray diffracting quality of the crystals as compared to that of the native crystals were observed. Three dimensional 2.8Å resolution data were first collected on crystals with a low $\text{SeO}_4^{=}$ / $\text{SO}_4^{=}$ molar ratio (approximately 5 to 1). After data reduction and scaling in the usual manner, a three dimensional difference map was computed. The map contained only two pre-dominant peaks which superimposed upon electron density regions thought to be due to solvent in the interface between the two independent molecules. The two peaks were of the same height (0.5 eÅ^{-3}) and were related by the two-fold rotation axis. The position coincided with the bridging sulfate ion between the phenolic hydroxyl of TYR' 146 and SER 195, near HIS 57. Since this was the only dominant position substituted by $\text{SeO}_4^{=}$, and since the crystals cracked when the $\text{SO}_4^{=}$ was being replaced by $\text{SeO}_4^{=}$ ions, it can be concluded that this bridging sulfate ion is an important factor in the formation of the crystal lattice and the stability of the dimeric molecule.

The soaking solution ratio was then further increased until the solution was nearly completely $(\text{NH}_4)_2\text{SeO}_4$. This was effectively achieved by removing the crystals and placing them in 75% $\text{SeO}_4^{=}$ solution. The crystals floated in this solution, since the crystal density was now less than the solvent density. (The density of the crystal is about 1.24 g/ml and that of the selenate about 1.47 g/ml.) Substantial changes

again occurred in the diffraction pattern in comparison to that obtained with the 5 to 1 selenate/sulfate molar ratio crystals.

Another set of three dimensional data was collected and processed. The difference map revealed six additional pairs of sites which corresponded to suspected sulfate positions in the two molecules, at somewhat lower occupancies ($\sim 0.22 \text{ eA}^{-3}$) than the initial substitution site. One additional set of data was collected on crystals grown entirely from ammonium selenate solutions. No additional sulfate positions were located and the occupancies were approximately the same as those obtained from the former experiment. This suggests that all $\text{SO}_4^{=}$ ions present are exchangeable. A summary of the sulfate positions and their interactions with the enzyme molecule is given in Table VII (58). Some of this work was carried out in conjunction with Dr. L. H. Wright of this laboratory.

4. Differences Between the Two Molecules

The fact that the two molecules in the asymmetric unit are not identical in structure has been indicated in a number of ways. The first suggestion of the non-equivalence of the two molecules was in their differing reactivity to the heavy atom reagents. In the event that the chemical reactivities were the same, each member of the heavy atom pairs related by the two-fold axis (equation 12), should have the same position and occupancy. For the major occupancy sites of the

Table VII. Bound Sulfate Positions.

$\text{SO}_4^{=}$	Peak Heights (eA^{-3})	Protein Contacts
1	0.57	Bridge between SER 195 and TYR' 146
1'	0.50	
2	0.28	ASN 95, LYS 177
2'	...	
3	0.24	Near THR 224
3'	0.27	
4	0.23	Adjacent to guanidinium group of ARG 154
4'	0.17	
5	0.23	Near NH_3^+ of ALA 149 chain terminus
5'	0.16	
6	0.17	ASN 245 side chain
6'	0.21	
7	0.18	Near ASN 236
7'	0.13	

Pt(+), Pt and Hg derivatives (Table I) reasonable agreement is observed for the two-fold related positions. Consideration of the minor occupancy sites shows a large discrepancy between occupancies, if indeed there is a two-fold related substitution at all.

The heavy atom derivatives based on the UO_2^{++} ion presented a different picture. Here, the major substitution site shows only a minor site 2A from the two-fold equivalent site. It will be shown in the next section why the large discrepancy in occupancies and positions exists. It is apparent from the heavy atom occupancies that there are some substantial differences between the two molecules.

The results from the selenate exchange work also indicated a similar difference in structure. In general, the occupancies between related sulfate ions were fairly consistent, but the positions varied substantially more from exact two-fold symmetry than did the heavy atom positions (58). In addition, $\text{SO}_4^{=}$ ion #2 in Table VII had no two-fold counterpart. The electron density of the native crystals grown from sulfate solutions indicated that excess electron density ($\sim 0.96 \text{ eA}^{-3}$ and 0.71 eA^{-3}) was positioned in these regions, but only one position became substituted with selenate ion, even in crystals grown in selenate solutions. From this, it can be surmised that one position contains an ordered sulfate molecule while the other position probably contains an ordered water molecule. The residues involved are on the surface of the molecule, and the electron density map suggests that the local environments around each position are different. The side chain of LYS 177 which interacts with the $\text{SO}_4^{=}$ in one molecule shows a different configuration in the molecule which does not bind sulfate ion.

The transformation parameters between the two molecules obtained by the least squares refinement were used to compute a difference map between the electron densities of the two molecules in the asymmetric unit. Positive and negative contours were drawn when the difference electron density exceeded $\pm 0.6 \text{ eA}^{-3}$, about 2.5 times the standard error of the difference density. Successive contours were drawn at intervals of 0.15 eA^{-3} . Figure 12 shows the electron density in the region

around the active site with the difference electron density superimposed in white. Solid contours represent excess density in the molecule shown, while broken contours represent excess density in the second molecule. Solid and broken contours adjacent to one another represent a shift or motion of the particular residue in one molecule with respect to the other. Isolated solid or broken contours may represent a solvent molecule in one enzyme molecule which is not in the other. Alternatively, single solid or broken contours superimposed upon the electron density peaks could represent the possibility that one enzyme is more ordered than the other. The net effect of this would be increased electron density peak heights in one molecule compared to those of the second.

The largest difference observed around the active site is the shift of MET 192 which was described earlier. The positive and negative contours are clearly visible in Figure 12. A positive and negative region on the alpha carbon of SER 195 indicates a slightly different position for this atom in the two molecules. Also, differences are noted in the position of the HIS 57 main chain and in the main chain of ASP 102. Of equal importance is the fact that the side chains of the catalytically important side chains SER 195, HIS 57, and ASP 102 possess no significant differences in orientation between the two molecules. A priori, this observation would suggest that the two molecules should have the same reactivity, if reactivity is only a function of the proper orientation of the important residues. However, close examination of the

individual electron densities shows small, and perhaps significant differences. In addition, the active site region of the second molecule contains more solvent molecules than the first molecule.

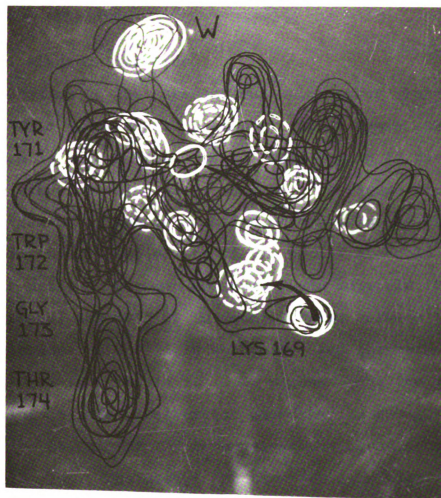
Further evidence for differences in the active site region comes from studies of the binding of inhibitors and substrate analogues to crystals of α -CHT. In some cases, the substitutions occurred to an equal extent in both active sites, with identical changes in the native structure. However, derivatives formed with N-pipsyl-L-Phenylalanyl chloromethyl ketone and N-formyl-L-Phenylalanine, among others, showed substantial differences in occupancies and modes of binding (59,25).

Recent crystallographic and chemical work with γ -chymotrypsin has elucidated the proposed mode of binding of oligopeptide inhibitors to α -CHT (60,61). It has been observed that a polypeptide inhibitor forms an extended, antiparallel configuration with the main chain of residues SER 214, TRP 215, and GLY 216. Presumably, these residues are necessary for substrate alignment before hydrolysis of the peptide bond near SER 195. As discussed earlier (see Figure 10 and associated text), these residues also show significant deviations from two-fold symmetry. If the position of these residues is critical, then the two molecules might be expected to show differing behavior toward substrates and inhibitors due to the two different active site conformations.

Another region of the electron density which shows both large deviations from two-fold symmetry and also rather exact two-fold symmetry in the same region is contained in the conformation of residues 165 to 174. Residues 165 - 170 are involved in a distorted helix near the surface of the molecule. Figure 13 shows the differences between the two molecules in this region. Many of the differences involve the main chain amides and carbonyl groups of the peptide linkages and correspond to different orientations of the planar peptide bonds. Also shown in the figure is the very obvious reorientation of the lysine side chain of residue 169. The solid contours represent its position in the first molecule and the broken contours represent its position in the second molecule. The change in conformation can be obtained by an approximate 45° rotation around the $C_\delta - C_\epsilon$ bond of the ϵ -amino group of lysine as shown by the arrow. Similar side chain reorientations are observed throughout the map for many residues, especially the side chains of lysine, glutamine and asparagine on the surface of the molecule.

The structure determination of insulin showed that many of the tyrosine residues had water molecules hydrogen bonded to the phenolic hydroxyl group (57). A similar observation was made in the structure of α -CHT, the results of which are amplified by the difference map. The peak marked W in Figure 13 corresponds to a water molecule hydrogen bonded to TYR 171. The large solid white contours indicate that there is no water molecule on the tyrosine of the second molecule.

Figure 13. Electron density of the helical region near
MET 180. Same contours as in Figure 12.



The reason for this observed difference is not obvious structurally but this example shows in a clear and direct manner the variability of the tertiary structure of α -CHT. In contrast, the absence of any differences in the residues TRP 172 to THR 174 shows the high degree of two-fold correspondence which can be obtained.

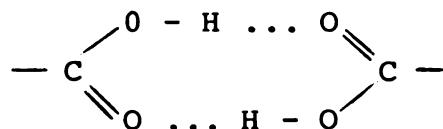
5. Uranyl Binding Interface

In addition to the interactions which occur along the local two-fold axes, there is another region where substantial intermolecular interactions occur. This region is referred to as the uranyl binding site, denoted U in Figure 9a. The region is formed by the interaction of one molecule and the two-fold screw axis equivalent of the second molecule in the asymmetric unit, or between molecule I and molecule II'.

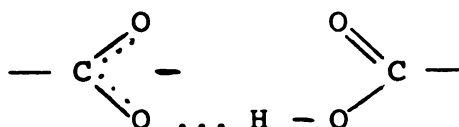
The uranyl ion (UO_2^{++}) is found located between the two carboxylate ions of GLU 21 in molecule I and the carboxylate ion of ASP 153 in molecule II'. It appears that the uranium atom obtains an octahedral oxygen coordination sphere, with the negative charges of the carboxyl groups being neutralized by the UO_2^{++} ion. The uranyl ion then forms an electrostatic interaction with the protein as compared to the covalent interaction with the other heavy atom derivatives. (The platinum and mercury derivatives generally form covalent complexes with the sulfur containing amino acids, methionine and cysteine.) This unusual binding arrangement for the uranyl derivative is probably responsible for the low thermal

parameters and high occupancies observed for the uranium derivative during the phase calculation and refinement process.

In the native structure, the above carboxyl groups might be forming a very interesting complex. Kirkwood and Schumaker (62) have shown that the electrostatic force caused by the fluctuations of protons between similar ionizable groups on the surface of a protein may result in a nonspecific attraction between two identical molecules. Timasheff has subsequently shown that with a certain geometrical configuration of identical ionizable groups, a specific attraction may be established as a result of the proton fluctuations (63). The largest effect will occur in the pH region of the pK of the ionizable group. He suggested that in a dielectric medium similar to that found on the surface of proteins, a net attractive energy of one to two kilocalories per mole may be available due to such an interaction. For neutral carboxyl groups, a model which might represent this type of interaction is



while if one of the carboxyl groups is deprotonated the model would be



where the lone hydrogen could be found between either pair of opposite oxygen atoms, or somewhere in between, with equal probabilities. Calculations have shown that the attractive interaction would be strongest at the pH corresponding to the pK of the particular group, or the pH at which the carboxyl group is protonated 50% of the time.

The pK's of the two carboxylic acid groups involved in the interaction, aspartate and glutamate, have been reported to be 4.1 and 4.5, respectively (64). The pH at which the native α -CHT crystals are grown is approximately 4.2, intermediate between the pK's of these groups. It is likely that this interaction promotes or enhances the association of asymmetric units during the crystallization process.

Supporting this possibility is the observation that best growth of crystals of CHT occurs around pH 4 and that crystals of α -CHT fail to grow, or grow very poorly, at pH values above or below the range of hydrogen ion concentrations which correspond to the pK's of these residues.

This intermolecular interface region is in striking contrast to the dimeric interfacial region between the two molecules I and I' in the asymmetric unit. In the latter, an interaction between two groups on the two molecules, i.e. TYR' 146 with HIS 57, was accompanied by a corresponding reciprocal interaction between TYR 146 and HIS' 57. Such is not the case here. The UO_2^{++} heavy atom binding studies showed there was only one site which became appreciably occupied. To produce this interaction there is an accommodation or

conformation change on the part of one or both of the molecules involved (different orientations of the side chains of ASP 153 and GLU 21). In addition, the location of the molecules in the unit cell precludes the reciprocal interaction. Since the proton fluctuation site is a very special site and is formed only when the residues possess a certain geometrical relationship, the possibility of a reciprocal site is diminished. This fact can be seen in Figure 14. The alpha carbon atoms of the peptide chain for molecule I are connected with solid lines, while the equivalent atoms belonging to molecule II' are connected with dotted lines. The side chains of GLU 21 and ASP' 153 are shown, with the uranyl ion binding at position U. It can readily be seen that the reciprocal interaction, ASP 153 - GLU' 21 could not occur due to the large distance (about 8A) between their respective side chains. It should also be noted that the orientation of ASP 153 is different in the two molecules. The different orientation brings the carboxyl group into a position such that it can form a salt linkage with LYS' 36 of molecule I'. This is an interaction across the dimeric interface with no two-fold equivalent interaction!

The electron density around the uranyl binding site has a large number of peaks, unaccounted for by the protein chain. Due to the polar nature of the interacting site, it is suspected that the extra peaks are due to localized, partially ordered solvent molecules. It has been shown by the selenate exchange work that there is only one sulfate ion in this

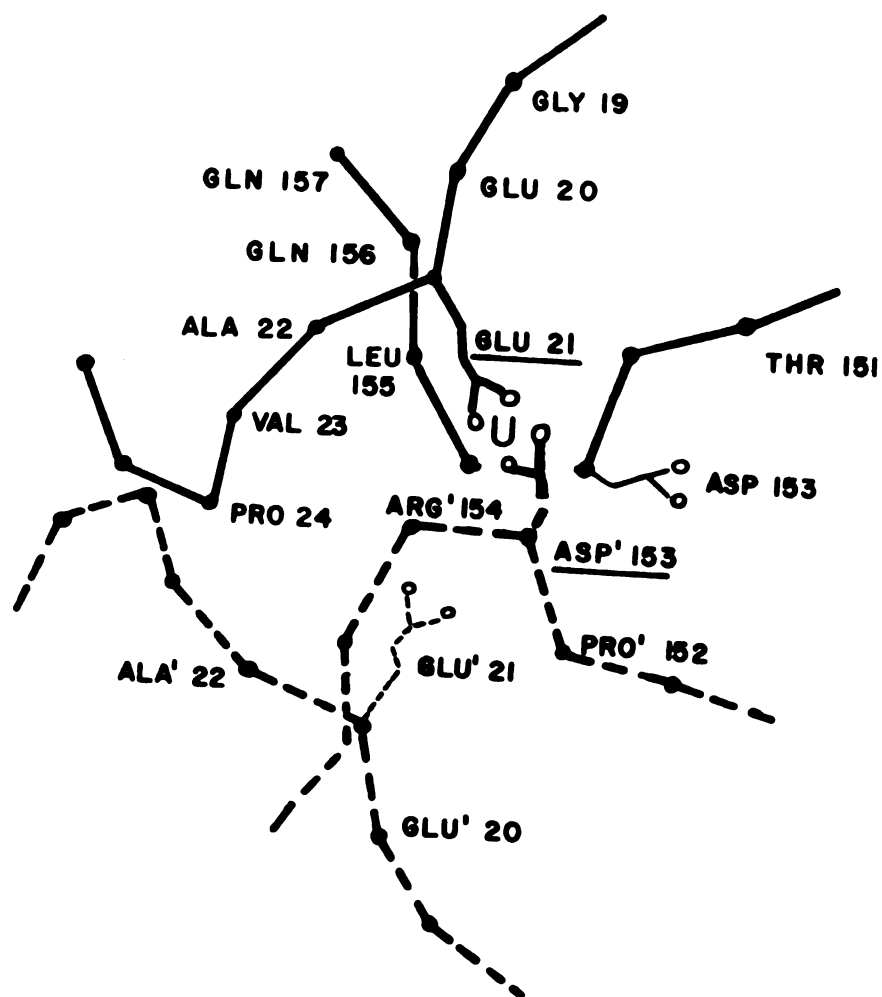


Figure 14. Detail of the peptide residues near the uranyl binding site. Solid circles represent the α -carbon positions of the peptide chains. Open circles are oxygen atoms. The solid lines connect residues of molecule I while the dashed lines are for molecule II' (see Figure 9a). The UO_2^{++} site is designated U.

region. The lone SO_4^- ion is adjacent to the guanidinium group of arginine 154. Since the guanidinium group carries a positive charge at pH 4, an ion pair is probably formed with the sulfate ion with charge neutralization. The remaining peaks in this region are water molecules or NH_4^+ ions from the ammonium sulfate solutions.

In general, the electron densities corresponding to the various side chains in this region have approximately the peak heights observed for side chains protruding into the solvent elsewhere. The peak heights for exterior side chains range from 0.5 to 0.8 eA^{-3} compared to 0.75 to 2.0 eA^{-3} for interior side chains. Before the amino acid sequence was traced in the electron density, the large peak heights of the two residues, ASP' 153 and GLU 21, was puzzling. The peaks now assigned to the two carboxyl groups always had the largest peak heights in the electron density maps, with heights approaching 3 eA^{-3} in some of the low resolution maps. This relatively high density can be possibly explained in two ways: (1) The proton-charge fluctuation interaction stabilizes these residues to such an extent that their disorder is very small, and hence, their peak heights high; or, less likely, (2) Errors are induced in the phases by the uranium heavy atom substitution. It is known that the sites occupied by heavy atoms used in the phase calculations are occasionally distorted in the electron density map of the native structure. This distortion is usually manifested as a residual negative or positive density at the heavy atom position. In the present case, the

electron density in the native map at the UO_2^{++} binding site is nearly zero; likewise, other heavy atom binding sites in the native electron density do not appear to have any particularly distorted features. While the effects of the second alternative can not be ruled out entirely, the extra rigidity induced by the proton fluctuation interaction might be sufficient to explain the large peak heights of these two residues.

In contrast to the large peak heights observed for GLU 21 and ASP' 153, the local two-fold counterparts, GLU' 21 and ASP 153 have low peak heights (0.35 eA^{-3} for GLU' 21 and 0.6 eA^{-3} for ASP 153). There are no particular forces stabilizing the side chain position of GLU' 21 except hydrogen bonding to the nearby ARG' 154 side chain.

V. STRUCTURE REFINEMENT

1. Refinement of the Coordinates

The fitting of the amino acid sequence with models to the observed electron density is subject to the interpretation of the builder and other difficulties. The model of α -CHT was built over a two-month period with the aid of Dr. William Brinigar of Temple University and Dr. Carl Morimoto of this laboratory. After the three dimensional model had been completely assembled, adjustments were made to eliminate close contacts and questionable arrangements of side chains. The relative x, y, z coordinates of all non-hydrogen atoms were measured. Two of the coordinates (y and z) were measured by sliding a 'grid' sheet, ruled at the scale of 2 cm to the Angstrom, into the Richard's box at the approximate x value of the atom to be recorded. Using the half-silvered mirror, the image of the atom was projected onto the grid sheet and the y and z coordinates were recorded for every atom. The relative x coordinates for each atom were likewise measured after the model was rotated with respect to the mirror by 90°. The reproducibility of the coordinate measurements was approximately $\pm 0.25\text{\AA}$ for each coordinate. The relative coordinates were then transformed into a coordinate system based on the crystal axes.

The refinement of the coordinates with respect to the electron density map can be objectively performed by using a series of elegant computer programs written by R. Diamond of the Medical Research Council, Cambridge, England. The first program adjusts the measured or guide coordinates to correspond to the known geometries of amino acid residues and the second program optimizes the fit of the adjusted protein model to the observed electron density distribution by least squares procedures.

The mathematical model building procedure (65) builds a polypeptide chain structure using the geometries of individual amino acids and peptide bonds as determined by X-ray crystallographic studies on individual residues. By keeping interatomic distances fixed, but allowing rotations about single bonds, the model polypeptide chain is folded in such a manner that the resulting chain and side chain positions approach the measured coordinates as closely as possible. By minimizing the sum of the squares of the distances between the guide and the model coordinates, rotational and translational shifts are computed and applied to the model coordinates. Because second and higher order terms are ignored when the derivatives for the least squares refinement are computed, several iterations are necessary to attain convergence.

Given the sequence of amino acids in the chain and the measured guide coordinates, the program starts at one end of the chain and builds a model from the sequence. After a number of residues have been built, adjustment of the model

coordinates of the individual amino acids to the guide coordinates begins at the first residue and moves sequentially along the peptide chain. A 'molten zone' is defined which contains a fixed number of residues. All of the residues are refined in the zone subject to constraints imposed by the fixed residues outside the molten zone. The molten zone then moves by one residue along the sequence, and the refinement of all residues currently in the zone performed. This sliding process continues until the zone reaches the other end of the chain. At this point, a list of the initial and refined coordinates for the input sequence are listed, along with the various interbond angles.

The model building procedure for the 241 amino acid residues in α -CHT was performed in several pieces due to storage limitations of the computer program. For each piece, the initial and refined coordinates were scrutinized for gross errors in the initial coordinates due to measurement or recording errors. In a couple of cases, the presence of a 'd' - amino acid or the wrong stereochemistry for a threonine or isoleucine side chain was detected. The necessary corrections were made to the input coordinates and that section of residues refined again. No attempt was made to readjust the model or the model coordinates based on the results of the building procedure except when a substantial error was clearly indicated. At the end of the refinement, the root mean square deviation between the input coordinates and the computed coordinates, using all atoms including the side chains, was 0.39 Angstrom

units. This number compares well with the value of 0.35Å reported by Diamond for the myoglobin structure, which he considered had extremely good guide coordinates (65). If one assumes the rms error to be equally distributed between the three coordinates x , y , z , the rms error for any one coordinate is then approximately 0.21Å. This is comparable to the estimated reproducibility of 0.25Å for the raw coordinates.

During the model building process, the backbone angle τ ($C_{\text{carbonyl}}-C_{\alpha}-N$) was allowed to vary, with any backbone distortions being accommodated by deviations of the angle from the expected value of 109.6° for tetrahedral carbon atoms. The average value of τ for the whole molecule was $112 \pm 10^{\circ}$. This value agrees with the average τ of 112° observed by Lipscomb for carboxypeptidase (66). The difference from the tetrahedral value was attributed to the packing of side chains in a large globular molecule which may require the peptides to be spread slightly at the alpha carbon atoms.

2. Real Space Refinement

The results from the model building program can then be input to the real space refinement program of Diamond (68) along with the observed electron density map. The model is then adjusted to fit the electron density by minimizing the integral

$$\int (\rho_o - \rho_m)^2 dv \quad (15)$$

where ρ_o and ρ_m are the observed and model electron densities, respectively. The model electron density is calculated as

$$\rho_m(\vec{r}) = \sum_i K Z_i G(a_i, \vec{r}-\vec{r}_i) + d \quad (16)$$

in which r is a general position vector, $\vec{r}_i$ is the vector position of the i th atom and $G(a_i, \vec{r}-\vec{r}_i)$ is a normalized, spherical Gaussian function representing the shape (a_i) of an atom's electron density distribution. The Z_i is the number of electrons in the i th atom, K is a local scaling factor for each residue and d is the correction for the level of the background of the electron density. The refinement procedure operates on a 'molten zone' of peptide residues, similar to that in the model building routine. All the residues within the zone are refined with respect to the observed electron density. Then the zone slides along the peptide chain by one residue and all of the residues currently in the zone are refined again. This process continues until the whole chain has been refined. The real space refinement program minimizes equation 15 by allowing K , d , a_i , Z_i , translational and rotational parameters to be varied. The refinement procedure preserves bond lengths and maintains chain continuity with the fixed parts of the molecule outside the molten zone.

At the end of each cycle of refinement of the whole molecule, the calculated electron density or difference density can be computed from the refined structure. In this way the correctness of the refinement can be ascertained. In an ideal situation, the difference map should be displayed on a visual display device along with the computed model or the observed electron density. One could then easily determine

those regions for which additional refinement is needed or for which other forms of help are required. In a refined structure, the difference map could also facilitate the location of solvent molecules which are binding to the protein molecule or could indicate the possibility of alternate conformations for side chains. In practice, comparison of a difference or calculated map to the observed map or model in a manual way is a very laborious task but should be done.

Because the real space refinement program is relatively new and detailed results from other protein structure refinement attempts are not yet available, it was necessary to experiment with the program to determine what types of refinement were best and to become familiar with its operation. A short section of the peptide chain from VAL 23 to GLN 30 was chosen for experimentation. The electron density corresponding to this chain was well resolved and included two tryptophan residues, 27 and 29. Initially, the overall scale factor K and the background parameter d were allowed to vary in addition to translational and rotational parameters. The radius parameter a_i was fixed at 1.2A, a value used in preliminary work with the 2.0A resolution myoglobin structure. The root mean square movement of the atoms in the refinement was 0.44 on the first cycle. Some larger shifts were observed for PRO 28, especially in the z direction. Examination of the observed electron density with the model coordinates revealed that the proline was placed on the edge of the electron density, and that the observed shifts from the refinement were moving the model to

higher electron density. The calculated translational shifts from a second cycle were in the same direction as in the first cycle, with an rms movement for all atoms of 0.27A. The rms movement of the third cycle was 0.19A with only one of the 68 shifts greater than 0.5A. By the fourth cycle, only one shift was greater than 0.3A with an overall rms movement of 0.12A.

Alternatively, a cycle of refinement was performed on this short segment starting from the initial coordinates but allowing the radius, a_i , to vary along with K and d and the translational-rotational parameters. The shifts observed were generally larger than the first cycle where a_i was not refined. The overall rms movement was 0.53A. A comparison of some of these shifts to the total movement recorded after four cycles of the previous calculations indicated that with a_i active convergence could be obtained with fewer cycles of least squares. During the refinement of a_i , most of the values increased from 1.2A. Since the resolution of the electron density map determines the average radius of the observed peaks, a value of 1.6 was used to begin refinement of the full molecule.

Translational and rotational refinements were performed after refinement of the overall scale factor and background level. The atomic radius a_i was also allowed to vary. The root mean square movement recorded for all atoms during the first cycle was 0.77A; for main chain atoms only (C_α , C, O, and N), the movement was 0.57A. Furthermore, it became apparent that the starting coordinates had a systematic displacement with

respect to the electron density along the x direction. The average shift recorded in the x direction during the first cycle for all atoms was -0.33A while that for the y and z directions was 0.04 and -0.13A, respectively.

A second cycle was completed starting with the coordinates from the first cycle and allowing the same parameters to vary as in the first cycle. The rms movement recorded was 0.35A for all atoms and 0.27A for the main chain atoms only. The movements on the second cycle were generally one half those of the first cycle. The average refined radius for all atoms was 1.57A. The average background level computed for each amino acid residue was 0.28 eA^{-3} with extremes from 0.12 to 0.44 eA^{-3} . The background level usually varied smoothly from residue to residue and did not appear to be a function of whether the residue was located in the interior of the molecule or on the exterior.

In general, the refinement program worked well and made some substantial improvements in the starting coordinates. In cases where the electron density was not fully resolved, the side chains sometimes approached one another closer than expected for van der Waals contacts. With several residues, the carbonyl group of the peptide chain was flipped over when the electron density was apparently stronger in the new orientation. One such flip occurred for a carbonyl group in the β -bend section of ILE 48 to TRP 51. The model was built with the carbonyl of GLU 49 arranged in a type II β -bend (38) but the bend was actually a type I bend. The refinement program

changed the orientation of the peptide bond by nearly 180° on the first cycle into a type I configuration with a supplementary adjustment on the second cycle.

An inherent problem with the real space refinement procedure is the complete neglect of non-bonded interactions due to adjacent peptide chains or side groups. As a result, some orientations of the side chains and main chain peptide linkages seem to be in an energetically unfavorable conformation, but no attempt was made to take these effects into consideration during the refinement. A program written by Levitt and Lifson allows a refinement of protein conformation using a macromolecular energy minimization procedure (69). Such a refinement procedure, alternating with the real space refinement procedure, would probably insure the best fit of the coordinates of the protein molecule to the electron density. Such a procedure will be carried out in the future by others.

3. Comparison with the MRC Structure of "Average" α -CHT

A preliminary comparison of the present structure of α -CHT with that of the average structure of Blow and coworkers at the Medical Research Council has been made using our partially refined coordinates and their refined coordinates (70). The comparison has been made between our molecule I and their averaged molecule, derived by averaging the electron densities of the two molecules in the asymmetric unit.

A number of significant differences exist between the two structures although, in general, the conformations of the two

independently determined structures are essentially the same. Considering the main chain coordinates, the rms deviation of atomic positions is approximately 1.3Å overall. One region of the molecule which shows close correspondence between the two determinations is the α -helix at the end of the C chain where the rms deviation is approximately 0.8Å. Other short chain segments throughout the molecule show similarly good correspondence, while other segments show significant differences, the largest of which are in the vicinity of the dimer interface region where it has already been noted that significant variations exist between the two molecules in the asymmetric unit. These regions include the residues GLY 216 to THR 219, TYR 146, ALA 149, ASP 35 and others with rms deviations which range from 2.6 to 3.5Å. The region GLU 70 to SER 77 shows the largest discrepancies between the two structure determinations; however, these are probably due to differing interpretations of a somewhat unresolved, diffuse electron density region. Another distinction between the two structures is the identification of sulfate ions bound on the surface of the molecule and in the active site region. Birktoff and Blow have indicated that only water molecules are in the active site (70). However, our selenate experiments provide irrevocable evidence for the presence of a sulfate ion hydrogen bonding to the active site serine in the crystal. Other discrepancies between the two structures probably arise because of slightly different crystallization conditions. This is especially true in the active site region;

Blow and coworkers grew crystals of α -CHT from ammonium sulfate solutions containing small amounts of dioxane, whereas our crystals were grown without dioxane. There is another complication in the direct comparison of the active sites of the two structures: their 'native' structure is that of the tosyl-inhibited enzyme rather than that of the active form of the enzyme used in our work. The fine differences in the conformation of the active site and elsewhere in the enzyme molecule will be revealed upon a closer and more detailed study of the two structures.

PART II

THE STRUCTURE OF α -CHYMOTRYPSIN AT pH 6.7

VI. INTRODUCTION

1. Effect of pH on Protein Structure

The physical chemistry of protein molecules is an extremely complicated subject and is poorly understood. The structure, stability, and activity of biological macromolecules are dependent on a large number of variables and attempts to correlate any one variable with a particular structural feature are almost certainly doomed to failure except in the best of cases. As more and more research is conducted, it becomes obvious that even the seemingly simplest physical concepts and systems are understood poorly. One current area of intensive research is the study of pure water and how its structure is modified by the presence of folded protein molecules consisting of hydrophobic cores and hydrophilic surfaces.

One of the experimental variables which plays an important role in the proper functioning of biological macromolecules is the hydrogen ion concentration in the surrounding medium. Of the 20 amino acids which commonly make up protein molecules, seven have side chains which can ionize in the pH range of 2 to 13. In addition to the side chains, the amino and carboxyl groups at either end of a protein chain are also ionizable. The state of ionization of these groups is a strong function of the medium and is sensitive to pH, dielectric constant, ionic strength, and temperature. Table VIII

lists some of the ionizable groups which commonly occur in proteins and their respective pK values.

Table VIII. Ionizable Groups in Globular Proteins.*

	<u>Structure</u> [#]	<u>pK</u>
A. Acidic Residues		
Terminal carboxyl	$\begin{array}{c} \text{H} \\ \\ \text{N} - \text{C} - \text{COOH} \\ \\ \text{H} \quad \text{R} \end{array}$	3.6 - 3.8
Aspartic Acid	R = -CH ₂ COOH	4.1
Glutamic Acid	R = -CH ₂ CH ₂ COOH	4.5
Tyrosine	R = -CH ₂ -  - OH	9.5 - 10.5
B. Basic Residues		
Histidine	$\text{R} = -\text{CH}_2 - \begin{array}{c} \text{C} = \text{C} \\ \diagup \quad \diagdown \\ \text{+HN} \quad \text{NH} \\ \diagdown \quad \diagup \\ \text{C} \end{array}$	6.3 - 7.8
Terminal amino	$\begin{array}{c} \text{H} \\ \\ - \text{C} - \text{C} - \text{NH}_3^+ \\ \quad \\ \text{O} \quad \text{R} \end{array}$	7.5 - 8.5
Lysine	R = -(CH ₂) ₄ - NH ₃ ⁺	9.6 - 10.4
Arginine	$\text{R} = -(\text{CH}_2)_3 - \text{NH} - \text{C} \begin{array}{l} \diagup \text{NH}_2 \\ \diagdown \text{NH}_2^+ \end{array}$	>13

*Compiled from references 64 and 71.

[#]R indicates side chain.

A protein molecule is typically composed of a large number of ionizable groups, a portion of which are positively or negatively charged at any given pH value. Proteins are often characterized by their isoelectric point, at which the number of positive charges equals the number of negative charges. With so many titratable groups, the titration curves of protein molecules are extremely complex. Additionally, several groups in a typical protein molecule will usually display abnormal characteristics by titrating above or below the expected values for the groups. This effect can arise from interactions of various groups with others of opposite charge, ionic groups buried or partially buried within the hydrophobic core of the protein, strong hydrogen bonding with ionizable residues, or the particular electronic environment around a certain residue. In some cases, there are groups in the protein that cannot be titrated at all as long as the native conformation of the protein is retained; this happens most often with tyrosine and histidine side chains which can be buried in the molecule when both groups are uncharged. Such is the case in hemoglobin, where 22 of 38 imidazole groups can not be titrated with the molecule in the native state (71). Similarly, in papain, 6 of 17 tyrosine residues do not titrate normally. The study of the effect of hydrogen ion concentration of protein conformation or activity must deal with these considerations and also with the many other electrostatic forces created by the charged groups and the various interactions which subsequently arise.

2. Effect of pH on α -CHT

The conformation and activity of α -CHT shows a complex dependence upon the pH of the surrounding medium. There are some 30 ionizable residues in α -CHT not including the three terminal α -carboxyl and α -amino groups at the ends of the three peptide chains. The interactions of several of these charged groups have been observed in the dimer interface and at the uranyl binding site. In addition, the presence of both histidine and aspartic acid in the active site suggests that the activity of the enzyme might be controlled, to some extent, by the pH of the medium.

A number of different thermodynamic states representing various stages of folding or unfolding of the peptide chains are accessible by changes in pH. The various stages have been identified by a number of thermodynamic measurements and are relatively stable states separated by small free energy differences (12). The transition from one state to another is generally a 'slow' process, but is somewhat dependent upon the prevailing conditions. The slowness of the transition is demonstrated by the conversion of α -CHT to the γ form: α -CHT is stable at acid pH values but slowly converts (hours) to γ -CHT when the pH is raised to alkaline values. The reverse transition also occurs, but requires a much longer time span (days).

The activity of α -CHT has been shown to be mainly dependent upon the ionization states of two amino acids in the active site region. In the one case, the deprotonation of

HIS 57 as the pH is increased coincides with the start of chymotryptic activity, as measured by the hydrolysis of neutral substrates such as acetyl-L-tryptophanamide (15). The rate of hydrolysis becomes a maximum at approximately pH 7.6 to pH 8.0, depending upon the substrate and the conditions of the solvent. As the pH is increased further, the salt bridge between ILE 16 and ASP 194 is destroyed by the deprotonation of the N-terminal amino group of ILE 16 and enzymatic activity begins to decrease. Above pH 10, only a small residual activity remains. The resultant bell-shaped dependence of the activity on pH is shown in Figure 15a (14).

Another gross effect that pH causes in the physical properties of α -CHT concerns the aggregation state of the enzyme molecules. Under the proper conditions (pH $\sim$ 4, concentrated protein solutions, high ionic strength), α -CHT exists primarily as dimeric molecules. The participation of charged electrostatic groups in the dimerization process was proposed since the extent of dimerization is dependent upon pH, ionic strength, and salt composition. Figure 15b shows the functional relationship of pH on dimer formation. Egan, et al. observed that the sedimentation constant reached a maximum at pH 4.0 and that the photo-oxidization of a single histidine residue caused the disappearance of any tendency for α -CHT molecules to associate (72). He suggested that a carboxylate ion and a charged imidazole group were responsible for the observed behavior. Similar conclusions were drawn by Aune and Timasheff after analysis of dimerization equilibrium

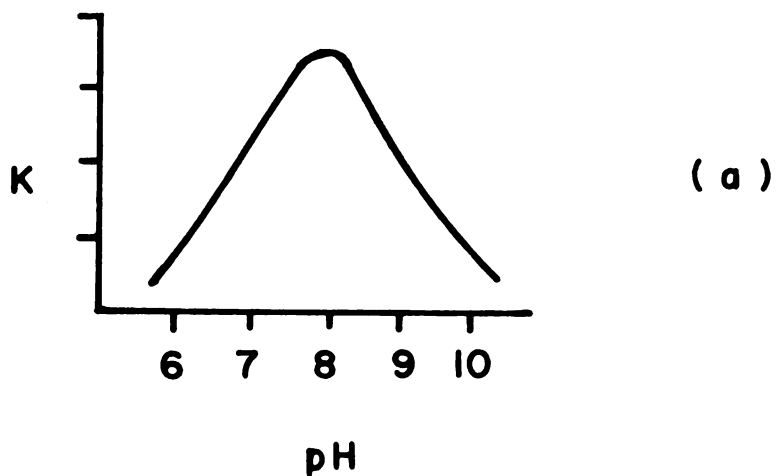


Figure 15a. Typical pH-rate profile for chymotrypsin catalyzed reactions showing rate of reaction K as a function of pH . (From Hess, et al., ref. 14)

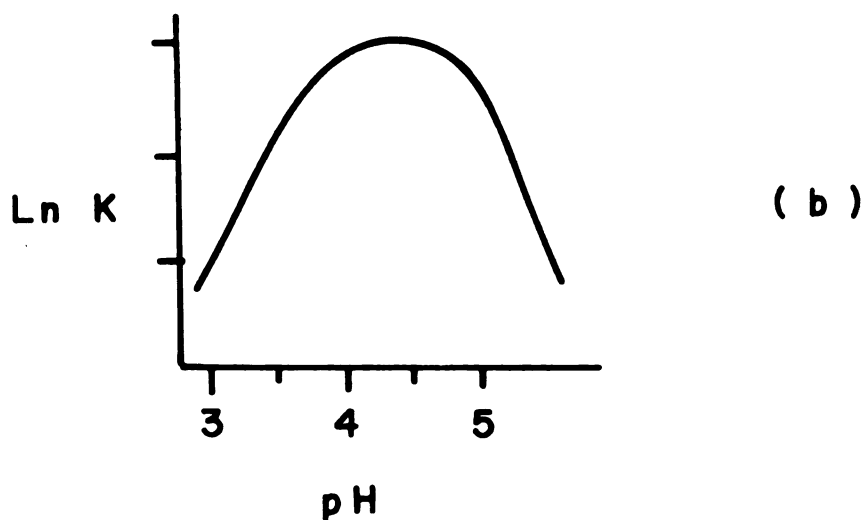


Figure 15b. Functional relationship of pH on dimer formation. The natural logarithm of the dimerization equilibrium constant versus pH for 0.1 M NaCl solutions at 25°C . Maximum dimer formation occurs at pH 4.4. (From Aune and Timasheff, ref. 73)

constants by the method of sedimentation equilibrium as a function of pH (73). Reference to Figure 10 shows that such an interaction occurs between TYR' 146 and HIS 57 along with the reciprocal interaction across the local two-fold axis.

The dimer which occurs in solution may then be the same dimer as, or very similar to, that which is observed in the crystalline state at pH 4.2. Support for this possibility was given by experiments of Gladner and Neurath (74) and Hexter and Westheimer (75). In the first case, the removal of TYR 146 by the action of carboxypeptidase resulted in a molecule which was no longer able to associate to form dimers. In the latter case, an intermolecular reaction was observed to occur between SER 195 and TYR' 146 during photolysis of diazoacetyl- α -CHT, made by reaction of α -CHT with p-nitrophenyldiazoacetate ($\text{O}_2\text{N} - \text{C}_6\text{H}_4 - \text{O} - \overset{\text{O}}{\underset{\text{O}}{\text{C}}} - \text{CHN}_2$) with release of p-nitrophenol.

The effect of a change of pH from 4 to higher values on the crystals of α -CHT was first observed in our laboratories by Mr. R. Timkovich (76) in a flow cell experiment in which the x-ray diffraction pattern of a crystal was monitored as various substances are flowed over and through the crystal (21). Substantial changes in the diffracted intensity of some reflections were observed when the pH of the 75% $(\text{NH}_4)_2\text{SO}_4$ soaking solution was changed from 4 to 7. The observed intensity changes suggested that significant protein conformational changes were occurring in the crystal and that these changes may be related to the dimerization process or to other

conformational changes. To delineate the changes that were occurring, a full three dimensional X-ray diffraction study of crystals of α -CHT whose pH was increased to 6.7 was undertaken.

VII. EXPERIMENTAL

1. Crystal Preparation

Crystals of α -CHT were grown from 50% ammonium sulfate solutions at pH 4.2 as described earlier. After sufficient growth had been attained, the solution was replaced with a 75% saturated ammonium sulfate solution adjusted to pH 4.2. The pH of the solutions was measured by a Beckman Zeromatic pH Meter with a combination glass electrode. Corrections for the errors in the hydrogen ion activity due to the high salt concentration ($\sim 3M$) were not made; relative pH changes were considered more important than the absolute pH values.

The hydrogen ion concentration of the ammonium sulfate solution was changed by the addition of concentrated ammonium hydroxide to the test tubes containing crystals. A sudden change in pH from 4.2 to 6.7 or higher caused all of the crystals to crack severely. When small amounts of NH_4OH were added periodically over a two-week period, most of the crystals developed cracks, but not as severely as before. However, a few crystals either did not crack or possessed only minor cracks; these crystals were used for the diffraction experiments. A similar experiment conducted over the period of a month produced comparable results. The latter experiment was performed because it was thought that a slow change of pH would allow for better accommodation by the crystal lattice of

any reorganization in the protein structure which might be occurring.

The diffraction patterns along the principal axial directions were recorded for a series of crystals at different pH values during the period of time that the pH was being changed. Changes in the diffraction pattern for crystals with a pH between 4.2 and 5.0 were small, although the b axis became progressively longer as the pH was increased. At pH values between 5.0 and 5.6, the crystals generally became cracked and substantial changes in the diffraction pattern occurred. Above pH 6 until approximately pH 8, only minor additional changes in the pattern were observed, with more changes occurring again above pH 8. Crystals at a pH of 6.7, intermediate between the two observed transition pH values, were chosen for three dimensional work. The stability of the diffraction pattern to X-rays was comparable to that of the native crystals. In addition, there was no appreciable difference in the X-ray diffracting quality of the crystals at pH 6.7 as compared to the native crystals at pH 4.2. Crystals from several different tubes and crystallization batches showed excellent reproducibility of the diffraction pattern at pH 6.7.

Typical axial scans along the c* direction for crystals at pH 4.2 (native) and pH 6.7 and for an inhibited derivative of α -CHT, tosyl- α -CHT (p-toluenesulphonyl- α CHT) are shown for comparison in Figure 16. The diffracted intensities are plotted on a log scale as a function of the Bragg scattering angle 2θ . Each division along the ordinate represents about

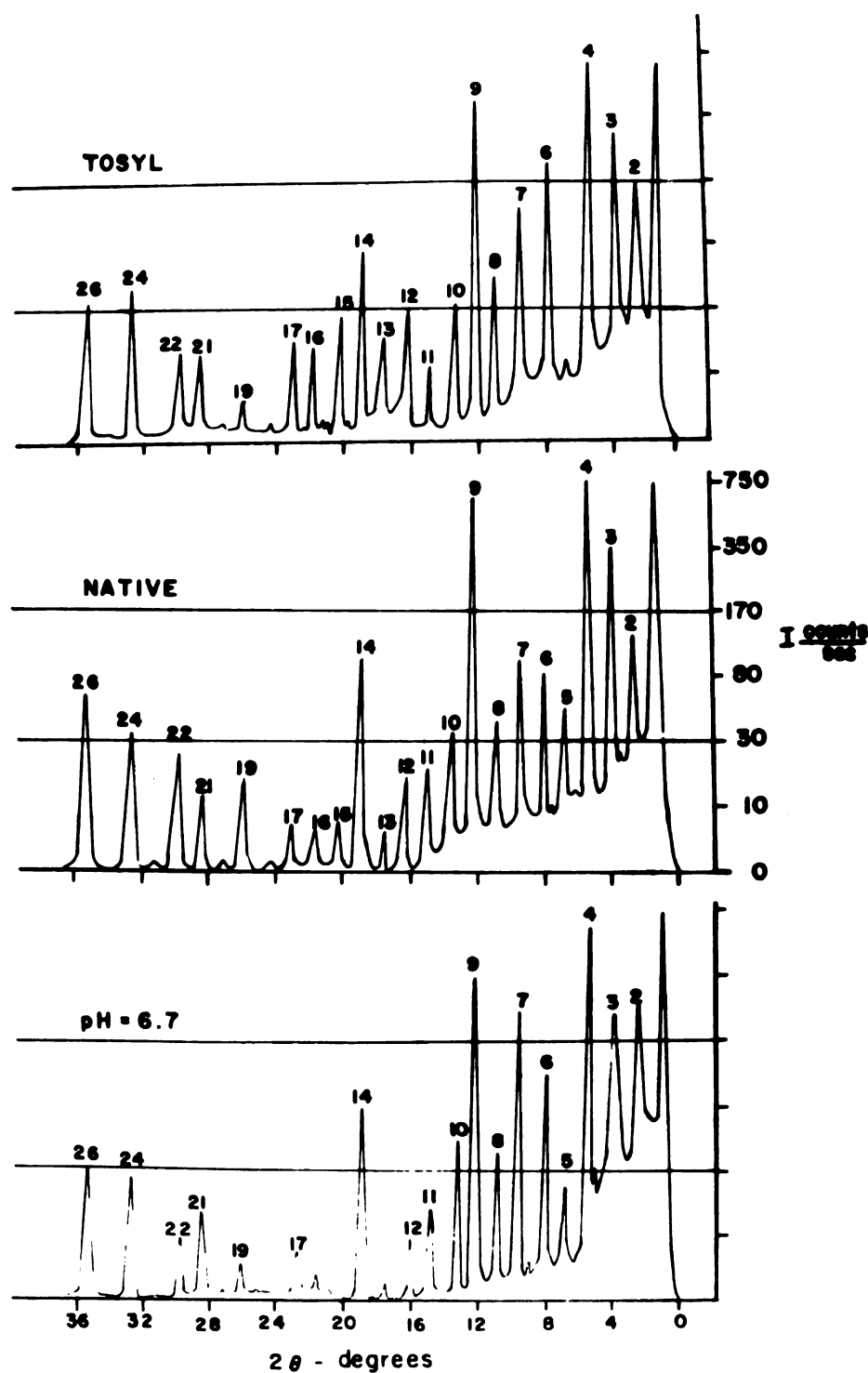


Figure 16. Axial scans along c^* direction for Tosyl, native and pH 6.7 crystals of α -CHT. The diffracted intensity is plotted as a function of 2θ . The order of diffraction is indicated above each peak.

a factor of two in the diffracted intensity. The three crystals were of comparable size and diffracting power. An appreciation of the changes that occurred in the diffraction pattern can be obtained by comparing the relative intensities of the four reflections 5, 6, 7, and 8 in the pH 6.7 crystal to the corresponding reflections in the native crystal. The magnitude of the changes is comparable to that produced by the inhibition of α -CHT by tosylation. The changes in tosyl diffraction pattern are due to the nearly 100% tosylation of SER 195 in both molecules in the asymmetric unit and include various protein movements which have been observed to occur upon tosylation. Comparable changes were observed on the other axial intensity distributions and in the three dimensional intensity distribution as compared to that of the native crystal.

The average cell dimensions measured from four different pH 6.7 crystals are compared to the native cell dimensions in Table IX. The standard error of each dimension is given in parentheses. The a and c directions decreased in length by about 0.1Å each, while the b axis length increased very significantly by 0.6Å. A small increase in the angle (β) between the a and c axes was also observed. The net increase in the cell volume upon change of pH is 1%.

2. Data Collection and Processing

The three dimensional intensity data to 2.8Å resolution were collected in a manner similar to that previously

Table IX. Cell Dimensions for pH 6.7 α -CHT Crystals and Native Crystals.

	<u>pH 6.7</u>	<u>Native</u>
a	49.13(5)A	49.24(7)A
b	67.83(7)A	67.20(10)A
c	65.81(7)A	65.94(9)A
β	101.92(6)°	101.79(6)°
Cell Volume	214500 $\pm$ 700 A ³	213400 $\pm$ 900A ³

described (24,25) for the heavy atom and inhibited derivatives with the following modifications. The data collection program was changed to measure only the observed reflections which had a figure of merit greater than 0.7 as determined from the heavy atom phase refinement procedure. The indices of these observed reflections (6200) were stored on the DEC 32K Disk File and retrieved sequentially. By this arrangement, and a reduction of the power of the incident X-ray beam, all the data to 2.8A resolution were collected from a single crystal in four days time, instead of the usual five or more crystals and several weeks of time.

The monitor reflections with high 2θ values, measured after every 100 reflections during the data collection, showed a total decay of approximately 30% from their initial intensities after 73 hours of exposure of the crystal to the X-ray beam. The decay of the diffracted intensities is a function of 2θ , with the greatest decay occurring at the largest 2θ

values. To decrease the effect of decay on the higher order reflections and the errors associated with its correction, the reflections from 3.5Å to 2.8Å resolution were measured first (2500 reflections with a maximum decay of less than 10%). Measurement of the lower resolution data followed with an overall decay substantially less than the 30% total decay recorded for the monitor reflections.

Before conversion to structure amplitudes, the intensities of the pH 6.7 crystal were corrected for twinning, absorption, and decay factors as described elsewhere (24). The resultant structure amplitudes were adjusted to an absolute scale by fitting the radial distribution curves (average F^2 versus 2θ) for the derivative and the native data. Figure 17 shows the radial distribution for the pH 6.7 crystal (dotted line) superimposed upon that of the native (solid line) after proper scaling of the structure amplitudes.

3. Difference Electron Density and Error Assessment

A 'best' difference electron density map between the structure at pH 6.7 and that at pH 4.2 was computed using coefficients

$$m (|F_{pH}| - |F_p|) \exp [i\alpha_p] \quad (17)$$

where $|F_{pH}|$ and $|F_p|$ are the structure amplitudes of the higher pH crystal and the native crystal, respectively, m is the figure of merit of the particular reflection used as a weight, and α_p is the 'best' protein phase. The mean square error in a difference electron density map of this kind is

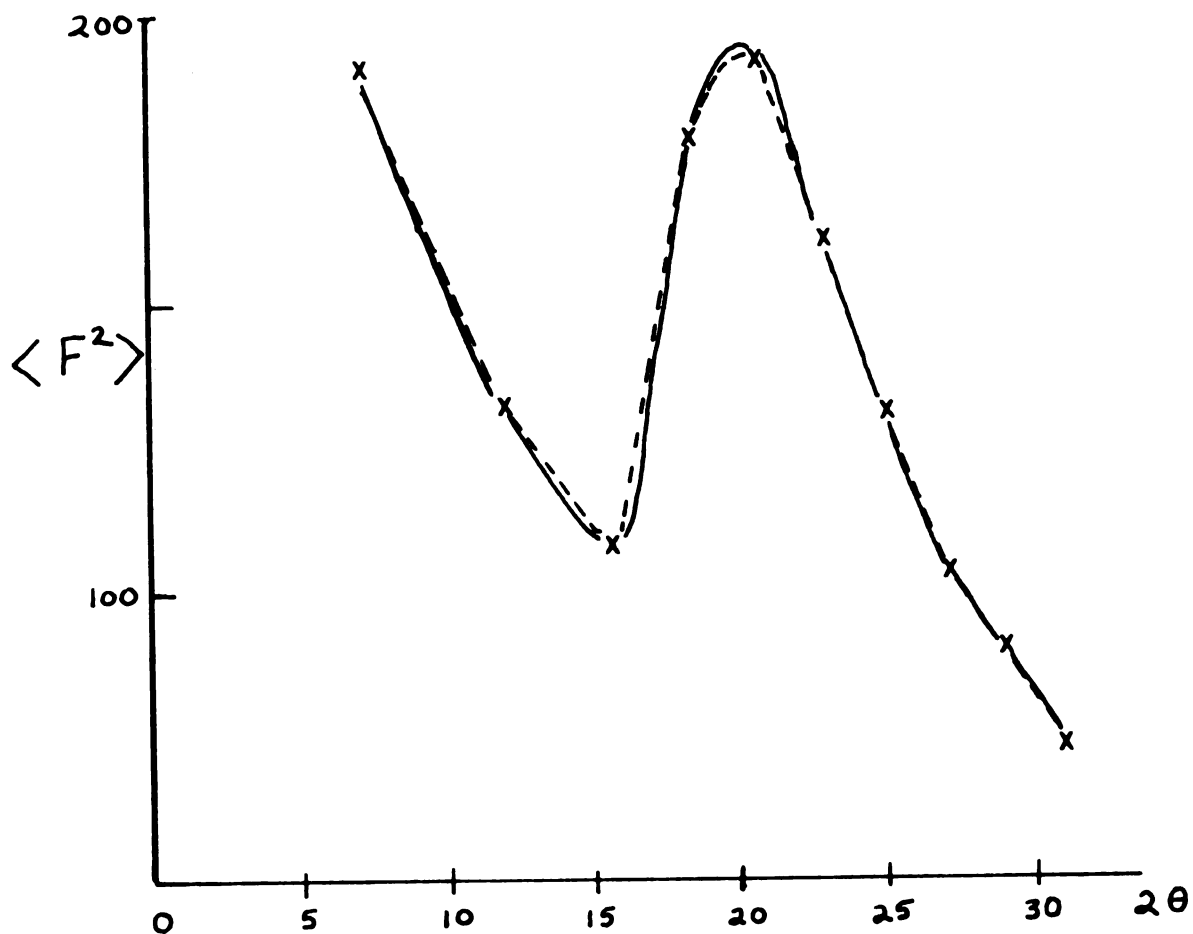


Figure 17. Radial distribution curves for pH 6.7 and native crystals. The average value of F^2 is plotted as a function of 2θ . Solid curve - native distribution, dashed curve - pH 6.7 distribution after scaling.

given by Henderson and Moffat (77) as

$$\langle \Delta \rho^2 \rangle = \frac{2}{V^2} \sum \Delta F^2 (1 - m^2) + \delta^2 + F^2 \quad (18)$$

where ΔF is the difference in structure amplitudes of the derivative and native crystals ($|F_{pH}| - |F_p|$). The first two terms in the summation are the contributions to the rms error caused by errors in the determination of the phases and by experimental errors in the ΔF values, respectively. The third term gives an estimate of the errors in the assumption that the phase angles for the native and derivative crystals are the same when the ΔF values are computed. The root mean square error for the ΔpH difference map was estimated by equation 18 to be 0.02 eA^{-3} . The difference electron density map was computed on the same scale as the native electron density map and was contoured and traced onto the plexiglas sheets carrying the native map.

Errors due to a lack of isomorphism in the phases and amplitudes caused by the changes in cell dimensions are difficult to assess quantitatively. Crick and Magdoff (78) have shown in a hypothetical case that a 'breathing' motion similar to that observed here (increased cell volume) could produce some significant changes in the diffracted intensities and phases. Lack of isomorphism errors could add uniformly to increase the background of the map or could appear as spurious peaks, distributed randomly throughout the electron density map. Of the many peaks observed in the difference density, only a few were unexplainable. If spurious peaks

were to be generated, a proportionate number of peaks should be observed in the regions between molecules. The few small peaks observed in these solvent regions usually had corresponding peaks in the native electron density. Situated near the protein molecule, these peaks might represent partially ordered salt or water molecules which have moved upon change of pH. The observed overall background of the difference electron density is approximately $\pm 0.05 \text{ eA}^{-3}$, which is somewhat higher than that predicted by equation 18.

VIII. DISCUSSION OF THE RESULTS

1. General Features of the Difference Map

A substantial number of positive and negative peaks, representing structural differences between the two structures at the different pH values, are observed throughout the difference electron density map with the majority of the peaks concentrated near the dimer interface and associated protein regions. The differences are primarily located on the exterior of the molecule, especially around residues whose side chains are polar. However, a few differences are also observed within the interior of the molecule and correspond to movements of segments of the polypeptide chain. As was mentioned earlier, there were essentially no significant difference density peaks in the solvent region. Furthermore, as might be expected, there were no significant peaks in or around the non-polar cavity which was described earlier.

In a number of places in the difference electron density map, a positive region will be accompanied by a negative region of approximately the same magnitude located on either side of a peak in the native electron density map. These positive-negative contours represent an electron density gradient or a shift or movement of the atoms or residues which correspond to the electron density. The direction of the shift is from the negative position in the native to the

positive position in the higher pH structure. Consider the electron density of an atom at a position x_1 to be represented by a Gaussian function whose height and width are functions of the number of electrons, the thermal motion associated with the atom or residue, and the resolution of the electron density map (curve ρ_1 , Figure 18). If the atom moves to a new position x_2 with electron density ρ_2 , the resultant difference electron density, $\rho_2 - \rho_1$, can be obtained by a point by point subtraction of curve ρ_2 from curve ρ_1 . The resultant curve has negative troughs and positive peaks, the depth and height of which are related to the amount of movement and the number of electrons involved. The positional shifts can be estimated by considering the slope of the difference electron density and the curvature of the electron density at the original position, x_1 , in a manner similar to that of Booth (79).

$$\Delta x = x_2 - x_1 = - \frac{\text{slope}}{\text{curvature}} = - \frac{(\partial \Delta \rho / \partial x)_{x=x_1}}{(\partial^2 \rho / \partial x^2)_{x=x_1}} \quad (19)$$

While this expression was derived for the position of a single atom, estimates for group movements can be made also. The largest difference density slopes or gradients were 0.4 to 0.6 eA⁻⁴, while the curvatures in the electron density at the peak positions were on the order of 0.7 to 0.9 eA⁻⁵, and correspond to group positional movements on the order of 0.4 to 0.8 Å.

One final feature of the difference electron density map which strengthened confidence in its correctness and its

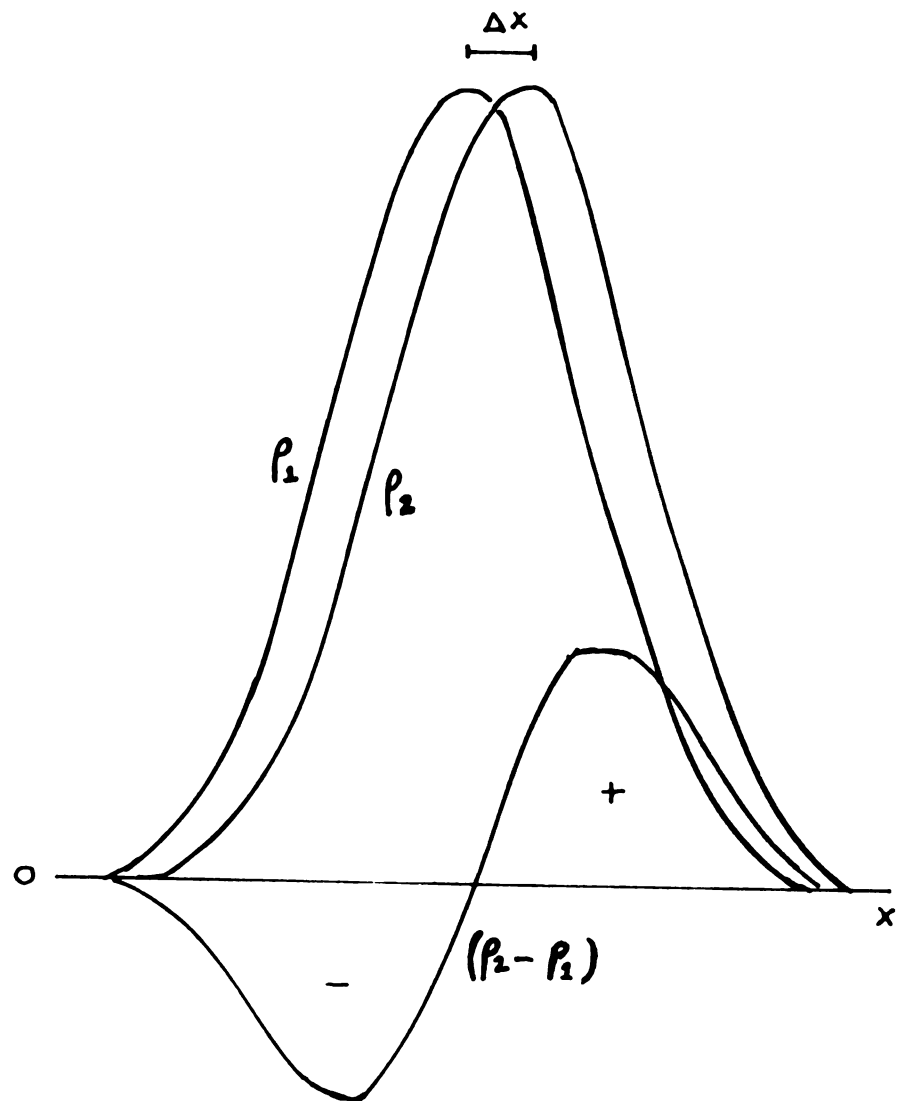


Figure 18. Gaussian curves showing effect of a small shift on the difference density.

interpretation was the existence of many two-fold related counterparts for many of the positive and negative peaks. If spurious peaks were generated by a lack of isomorphism between the native and pH 6.7 crystals, it seems improbable that equivalent peaks related by a non-crystallographic symmetry element would be generated as well.

2. Histidine 40

One of the largest structural changes in the pH difference density is associated with HIS 40. The peptide chain from residues 35 to 41 lies on the surface of the α -CHT molecule in the dimer interface region. While PHE 39 extends into the interface, HIS 40 points toward the interior of the molecule. The imidazole ring is positioned so that one side of the ring is relatively accessible to the solvent. The peptide chain 192 to 194 is found on the other side of the ring, with the plane of the ring roughly parallel to the direction of this peptide chain. The side chain of SER 32 is in the interior of the molecule with its hydroxyl group directed toward the imidazole ring. A schematic drawing of the location of HIS 40 with respect to these residues is shown in Figure 19.

The imidazole ring is positioned so as to allow the formation of a hydrogen bond with the hydroxyl group of SER 32 or to the carbonyl oxygen of GLY 193 through the $N_{\epsilon 2}$ atom of the ring, although the latter bond is less likely to occur, due to the long distance ($\sim 4.5\text{\AA}$) between the oxygen and nitrogen atoms involved. The orientation of the ring precludes the

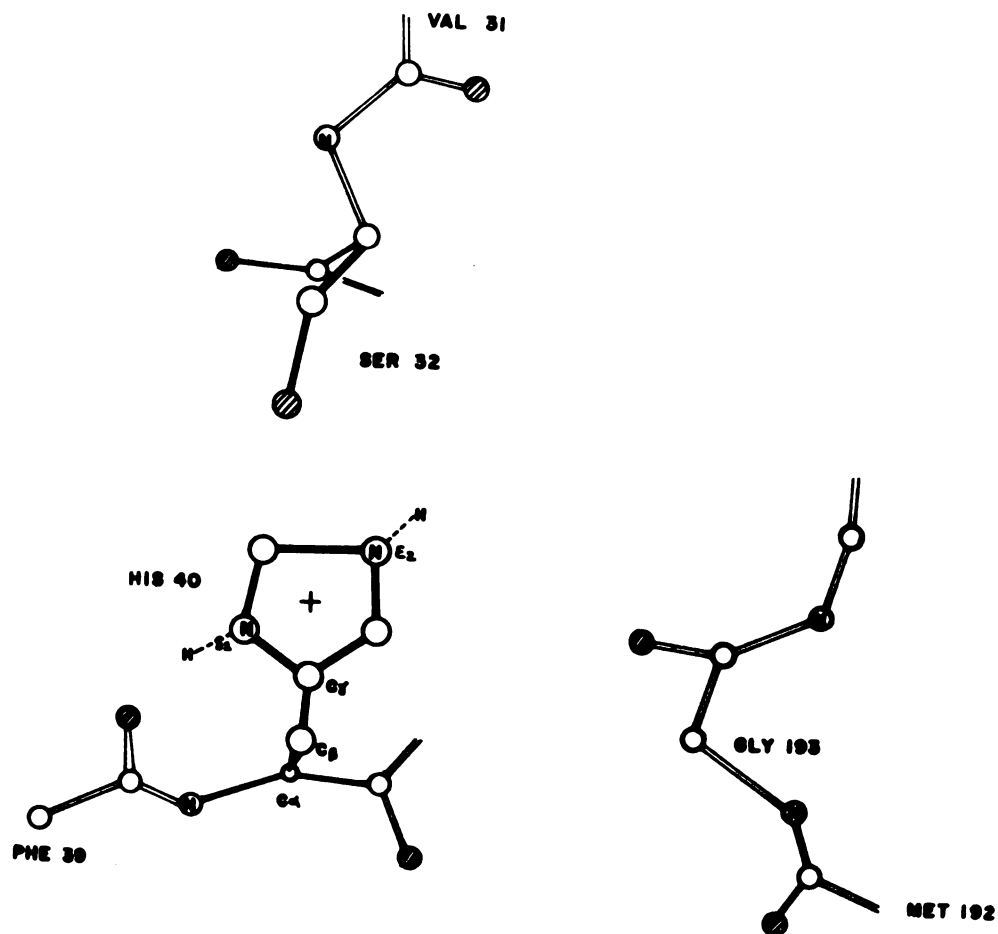


Figure 19. Schematic drawing of important residues around HIS 40. Main chain atoms connected by open bonds while side chain atoms are connected by solid bonds. Oxygen atoms -  ; nitrogen atoms -  ; carbon atoms - .

formation of simultaneous hydrogen bonds through both the $N_{\delta 1}$ and $N_{\epsilon 2}$ nitrogen atoms. The hydrogen bond formed with either possibility would probably utilize the $N_{\epsilon 2}$ nitrogen atom and would depend upon whether the ring is flipped around the $Ca-C\beta$ bond so that $N_{\delta 1}$ is pointing toward the solvent or toward GLY 193. In chymotrypsinogen, the inactive precursor of α -CHT, HIS 40 is found in a hydrogen bond complex with the side chain of ASP 194. Upon activation to α -CHT, it has been proposed that the ASP 194 side chain swings away from HIS 40 which then forms a hydrogen bond with SER 32 or GLY 193 (16).

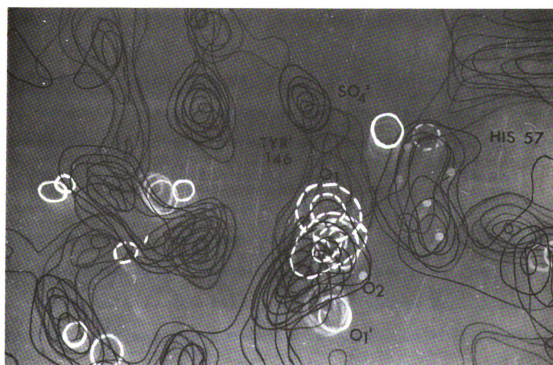
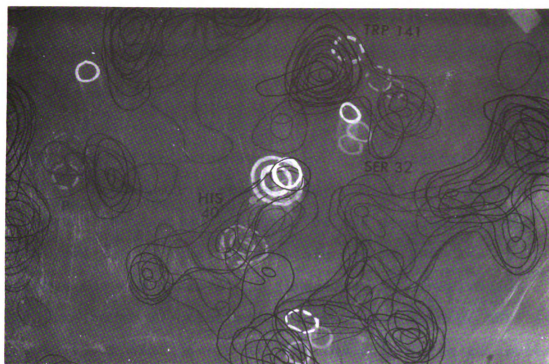
Since histidine residues typically have a pK around pH 6.5, it is expected that in crystals at pH 4.2 the HIS 40 imidazole ring would be protonated and positively charged, with both ring nitrogen atoms carrying hydrogen atoms and the positive charge delocalized over the ring. When the pH of the crystals is changed to 6.7, the difference electron density shows a peak of large positive density on the right side of the electron density corresponding to the ring and negative density on the left as depicted in Figure 19. These peaks imply a movement of approximately 0.5Å of the imidazole ring of HIS 40 towards GLY 193 to a position more favorable for the formation of a hydrogen bond through $N_{\epsilon 2}$ to the carbonyl oxygen of 193. The difference electron density also shows a positive density on and near the carbonyl oxygen of GLY 193, again indicating the presence of increased electron density between the carbonyl group and $N_{\epsilon 2}$ of HIS 40.

Coupled with these observed changes, a negative region with no accompanying positive region is observed adjacent to the $N_{\delta 1}$ position of the HIS 40 ring, corresponding to a small peak in the native electron density. The loss of a water molecule bound to the $N_{\delta 1}$ nitrogen atom at low pH could explain the negative peak. With this assumption, the orientation of the imidazole ring in the native structure would then be as shown in Figure 19. At pH 4.2, the conformations of HIS 40 and HIS' 40 are somewhat different; the pH change causes slightly different motions of the side chains with the final positions of both imidazole rings in approximately the same conformation. Figure 20 shows the difference electron density map superimposed on the electron density map of HIS 40 and neighboring regions of one molecule.

The aromatic side chain of TRP 141 is located approximately 5Å above the side chain of HIS 40 with the two rings being nearly coplanar. The fluorescence of α -CHT increases with increasing pH above pH 6 when an ionizing group of $pK \sim 6$ loses a proton (14). Model studies by Katchalski and coworkers (80) have shown that the fluorescence of tryptophan is quenched when a 'complex' is formed by the interaction of a charged imidazole ring of a histidine residue with the indole ring of tryptophan. An increase of fluorescence is observed when the imidazole ionizes and disrupts this complex. From the experiments on CHT mentioned above, it is probable that such a complex is formed between TRP 141 and HIS 40 at low pH values.

Figure 20. Electron density around HIS 40. Difference electron density between pH 6.7 and native crystals shown in white; contours at 0.15 eA^{-3} starting at 0.15^{-3} . The ring of HIS 40 is seen edge on from a position corresponding to GLY 193 in Figure 19. Peak marked P belongs to the alpha carbon of PHE' 39.

Figure 21. Electron density in vicinity of TYR' 146 and HIS 57. Difference electron density shown in white as described in Figure 20. The carboxylate oxygen moves from its position, O₁, in the native electron density to position O₁' when the pH is changed. The bridging $\text{SO}_4^{=}$ between TYR' 146 and SER 195 (not shown) is indicated. The negative-positive density on the HIS 57 ring indicates a small rotation.



When the pH is raised to 6.7, the deprotonation of the HIS 40 ring breaks the complex and increased fluorescence results.

3. Dimer Interactions

Aune and Timasheff have made a careful study of the dimerization-equilibrium constants by ultracentrifuge experiments (73). From their analysis, they postulated that the electrostatic interaction (at pH $\sim$ 4) of the protonated ring of HIS 57 and the terminal carboxylate ion of TYR' 146 is responsible for the observed pH dependence of dimerization and that the other electrostatic interactions between ASP' 64 - α -NH₃ ALA 149 or LYS' 36 - α -NH₃ ALA 149 only serve to increase the free energy of formation of the dimer.

Because of the number of electrostatic interactions which occur around the two-fold axis, one might expect to see changes in this region when the hydrogen ion concentration is changed. Not surprisingly then, the pH difference map revealed a number of molecular rearrangements in the interface region between the two molecules in the dimer. The largest interfacial change was associated with the carboxylate groups of TYR 146 and TYR' 146. As mentioned earlier, the carboxylate groups in the dimer at pH 4.2 were arranged so that they can interact with the charged imidazole rings of HIS 57 and HIS' 57. When the pH is raised to 6.7, the imidazole rings become deprotonated and neutral, and any electrostatic interactions which might have existed are disrupted. The difference electron density map showed a large positive peak ($\sim 0.35 \text{ eA}^{-3}$) near the

carboxylate ion on both tyrosine residues in the interface with a negative peak (-0.35 eA^{-3}) on the carboxylate electron density. Figure 6 shows the arrangement of residues in the region of the HIS 57 - TYR' 146 interaction at pH 4.2 and Figure 21 shows the electron density map around these residues with the pH difference map superimposed. The net result of the observed difference peaks is a rotation of a carboxylate oxygen (O_1) away from the HIS 57 side chain and corresponds to an out of the page movement, into the solvent. There is apparently no movement or reorientation of the phenolic group of the tyrosine residue or in the bridging $\text{SO}_4^{=}$ ion in either two-fold interaction.

In addition to the large peaks observed in the pH difference electron density around TYR' 146 there are smaller positive and negative peaks ($\pm 0.15 \text{ eA}^{-3}$) associated with the main chain and side chains of residues ARG 145 to GLY 140. The magnitude of the difference density peaks decreases from residue 145 to 140; a large movement at TYR 146 results in a series of smaller changes in the several residues away from 146, with the changes decreasing as the distance from 146 increases along the chain. The net result of these positive and negative peaks is to produce a small, complex movement of the whole chain segment away from HIS 57 and the interface region. This movement is presumably initiated by the reorientation of the carboxylate ion of TYR' 146. Similar movements are also observed for the two-fold related chain segment.

Other differences, somewhat lesser in magnitude, are observed in the dimer interface around residues ALA 149 and ASP' 64. The N-terminal amino group of ALA 149 forms a salt bridge with ASP' 64 of the two-fold related molecule and vice versa. The electron density around the ALA' 149 - ASP 64 interaction is strong and generally resolved. However, the two-fold related interaction is not so obvious or distinct. One reason for this unclarity is the formation of a complicated arrangement involving the side chains of ASP 35, ASP' 64, ALA 149 and a SO_4^- ion (see Table VII). At the higher pH, a negative peak in the difference map coincides with the electron density found to be a sulfate ion; this correspondence suggests the loss or movement of the sulfate ion from its original position. A similar change also occurs in the two-fold related molecule. A detailed interpretation of all of the observed differences in this region is very difficult and not without considerable uncertainty. The important conclusion is that some protein rearrangements take place at this interaction site upon the change of pH.

4. Active Site

A considerable number of changes are also observed in the active site region of the molecule and with residues associated with the active site. Figure 22 shows the electron density of residues 194 to 198 of molecule I with the pH difference electron density superimposed. Peaks in the difference density in the active site are relatively small (0.1 to

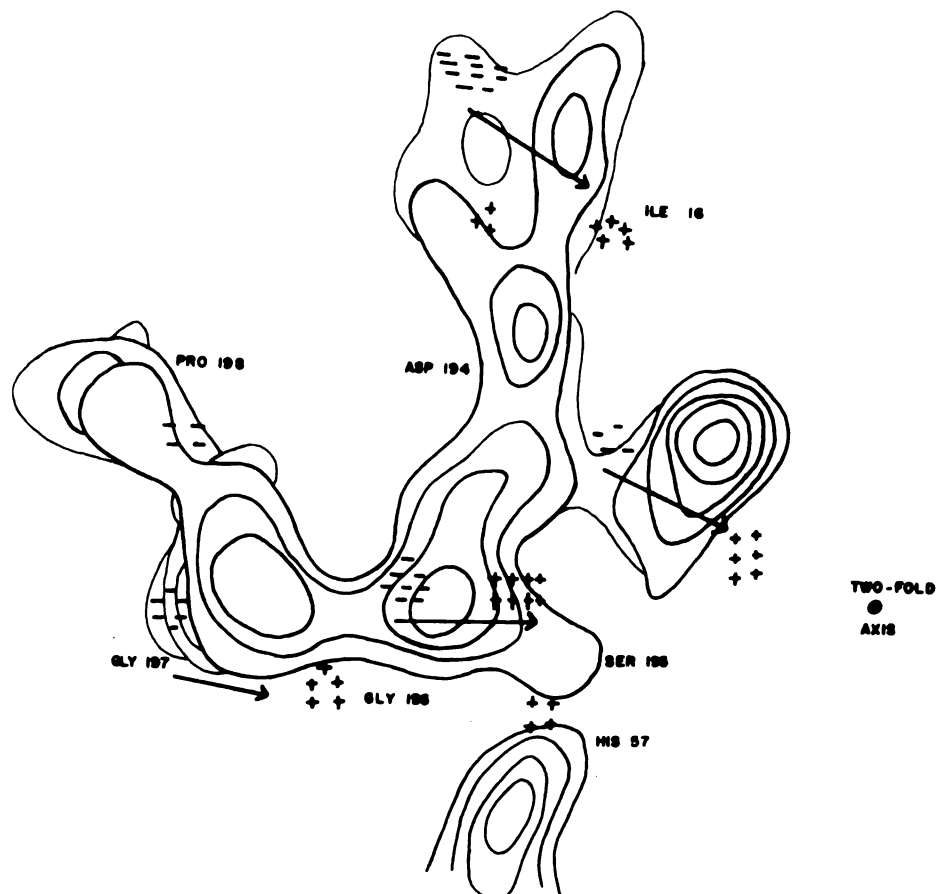


Figure 22. Electron density map around SER 195. Positive regions in the pH difference map indicated by plus signs, negative regions by minus signs. The arrows indicate direction of movement when the pH is changed to 6.7.

0.15 eA^{-3}) and numerous, but taken as a whole, are significant. In molecule I, the observed changes combine to imply a movement of the chain from 194 to 198 towards the dimer two-fold axis as indicated by the arrows in the figure. The negative-positive region on the main chain of SER 195 shows this movement clearly; also, the same configuration of peaks on ILE 16 is observed. Molecule I' shows an opposite behavior upon change of pH around its active site. The negative regions in molecule I are generally replaced by positive regions and vice versa. Again, taken as a whole, the chain from 194 to 198 in molecule I' appears to move away from the two-fold axis.

The reasons for these movements are rather difficult to rationalize until the differences between the two independent molecules at pH 4.2 are considered. At pH 4.2, the difference map computed between the electron densities of the two molecules in the asymmetric unit showed that the same residues in molecule I were further from the two-fold axis than the corresponding residues in molecule I'. The pH changes are in the direction as to make the active sites of both molecules more equidistant from the two-fold axis and, presumably, more alike in the resultant configuration of their active sites. In addition, molecule I' undergoes other small rearrangements upon change of pH. In both molecules, there are a number of positive and negative peaks around the CYS 191 - CYS 220 disulfide bridges and are probably also associated with the movements of the chain from 194 to 198.

In addition to the movements of SER 195, there are also pH induced differences in the other components in the active site. The imidazole ring of HIS 57 in molecule I undergoes a small rotation around the $C_\alpha - C_\beta$ bond which points the nitrogen $N_{\delta 2}$ toward the oxygen of the SER 195 side chain. In molecule I', a different behavior is observed for HIS' 57. Here, the imidazole ring does not move toward SER' 195 but moves toward the two-fold axis in the plane of the ring. While some pH changes appear to make the two active sites more similar, other pH changes occur which oppose the similarity. In any event, the exact stereochemistry of the important residues in the active site during catalysis remains uncertain.

Other pH changes are observed in the side chain of ASP 102 and in several residues on either side of ASP 102. When the pH is increased, the side chain carboxyl group exhibits a slight movement away from the imidazole ring of HIS 57. At pH 4.2, the carboxylate ion of ASP 102 is presumably negatively charged and the imidazole ring of HIS 57 positively charged. Upon a change to more neutral conditions, the electrostatic interaction between these groups would decrease as HIS 57 becomes deprotonated and uncharged. A slight repulsion of the carboxylate of ASP 102 would then be expected. Such a movement is implied further by a series of positive and negative pH differences along the main chain from ASN 101 to LEU 106, in addition to the differences observed on the side chain of ASP 102 itself.

5. Uranyl Interface

As described in section IV-5, the uranyl binding interface appears to be formed by a pair of carboxyl groups on different molecules coming together to possibly form a special complex stabilized by proton fluctuations. Such a complex is strongest at the pH which corresponds to the pK of the interacting groups. If the pH is changed to a value outside the pK ranges, disruption of this complex would be expected and is observed to occur with α -CHT.

As noted earlier, the electron density around the uranyl binding site is complex; so are the pH differences. The largest positive-negative peaks are associated with a movement of the guanidinium side chain (positively charged) of ARG 154 toward the carboxylate group of GLU 21 and a reciprocal movement of GLU 21 toward ARG 154. The result of these movements is to change the environment around the UO_2^{++} binding site substantially. Based on these observations of this region, it is probable that UO_2^{++} ions would not bind as strongly, if at all, at higher pH values.

6. Other pH Changes

Upon examination of some of the non-electrostatic interactions which occur in the interface region during dimerization, it becomes apparent that the change in pH, by loosening the salt-like interactions between the two molecules, disturbs the hydrophobic interactions as well. Two of the non-electrostatic interactions which must undergo some mutual

accommodation in order to form the dimer (PHE 39 - PHE' 39 and SER 217 - SER' 217) show some relatively substantial changes. The pH difference density of the two PHE 39 residues shows large but differing motions. The main chain of PHE 39 moves toward the two-fold axis while the main chain of PHE' 39 moves along the two-fold axis, maintaining somewhat respectable van der Waals contacts with each other. There are also movements of the main chain of SER 217 to SER 218 but no observed changes in the two-fold related segment SER' 217 - SER' 218. It is possible that the movement of these residues is not related to the change in pH directly. When the dimer is formed, strong electrostatic interactions might force these residues to adopt configurations which are of a higher energy than normal. When the pH is raised, and the molecules tend to separate, the hydrophobic residues might relax, possibly, to lower energy configurations.

Many of the other changes observed when the pH is changed are observed to be changes in the water or ion structure on the surface of the molecule. Most of the observed peaks occur near serine or threonine side chains and the charged side chains, both acidic and basic. For instance, the sulfate molecule bound near ASN 95 moves some 3A away to take a position near ASN 100. SER 217, SER' 223, ASP 70 and other residues appear to have attracted a water molecule with pH change, while SER 186, CYS 1, HIS 40, ASP 35, and TYR 94, among others, seem to show a loss or movement of solvent molecules. For instance, positive difference density peaks are also observed

near the main chain carbonyl groups of GLY 211, SER 159, and LEU 243, with negative densities near ILE 99, GLY 211, GLN 18 and TRP' 215.

Apart from an occasional peak of low magnitude which seems to be spurious, a curious feature was observed near the tryptophan residues. On the edge of each indole ring of five tryptophan residues, TRP 27, 29, 51, 172 and 215, there is a positive difference density contour, occurring in almost every case near the six membered end of the indole ring. The reason for these positive peaks at these positions is unknown; they do not seem to be due to motions of the tryptophan rings as there are no associated negative contours. A solvent molecule associating with the non-polar aromatic ring is equally unlikely. The explanation for these peaks probably awaits future developments in the understanding of protein physical chemistry.

7. Final Comments

A summary of the changes in the molecular structure and their probable interpretations of crystalline α -CHT upon change of pH of the surrounding medium from 4.2 to 6.7 is given in Table X. The observed pH changes are both intermolecular and intramolecular; the former concern the processes of dissociation of a dimeric molecule and the latter are probably related to pH induced conformational changes of the enzyme. A large number of the pH induced intramolecular changes involve active site components; similar changes may

Table X. Summary of Changes of α -CHT Upon Change of pH.

- | | |
|---|--|
| 1. HIS 40 and HIS' 40 | Rotation of ring 0.5A toward carbonyl of 193 with loss of H_2O off $N_{\delta 1}$ - breaks complex with TRP 141. |
| 2. TYR 146 - HIS' 57
TYR' 146 - HIS 57 | Movement of carboxyl terminal away from interaction with HIS 57. Causes movement of residues 141 to 146 in the process. |
| 3. ALA 149 - ASP' 64 | Complicated set of movements around salt bridge, displacement of interacting $SO_4^{=}$ molecule. |
| 4. Active Site | Movement of residues 192-198 in both molecules to become more nearly alike. Slight movement of ASP 102 away from HIS 57. Different movements for HIS 57 and HIS' 57. |
| 5. Uranyl Binding Site | Movement of ARG 154 toward GLU 21, with subsequent movement of GLU 21. Movements occur in both molecules although only molecule I has uranyl binding site. |
| 6. PHE 39 - PHE' 39 | Complicated motions occurring around the hydrophobic interaction of these two groups. |
| 7. SER 217 - SER 218 | Movement of main chain in molecule I but not in molecule I'. Residues are involved in non-electrostatic interaction with two-fold related residues. |

Table X. Continued.

8. Solvent Changes	Loss of H ₂ O from carbonyl of TRP' 215 and SER 186, CYS 1, ASP 35, TYR 94 and others. Gain water at SER 217, SER' 223, ASP 70. Movement of SO ₄ ⁼ from ASN 95 to ASN 100.
--------------------	---

also occur in the active site in solution. The ramifications of all the observed changes have yet to be unraveled.

For instance, to be studied in detail are the effects of lowering the pH below 4.2. It has been proposed that the salt bridge between ILE 16 and ASP 194 will be disrupted below pH 3.0 due to the protonation of the carboxylate group of ASP 194 (12). Preliminary data indicate that the magnitude of the changes in the diffraction pattern at pH 2.7 are comparable to those observed at pH 6.7. Also to be investigated in detail is the structural transition which also occurs above pH 8, again presumably due to the rupture of the salt bridge 16 - 194 with deprotonation of ILE 16. Preliminary diffraction data also suggest that the changes observed in the transition from pH 4.2 to 6.7 are reversible when the pH is returned to 4.2.

The crystalline state possesses many advantages for the study of the effects of pH and other experimental variables on the structure of proteins and enzymes. The crystalline forces prevent excessive movements and changes from occurring,

but the large spaces between molecules in protein crystals and the large proportion of solvent probably allows the molecule to assume a conformation similar to that in solution. By careful interpretation and study of the results, X-ray diffraction can provide meaningful information and give new insights to be used in unraveling some of the complexities of protein physical chemistry.

REFERENCES

References

1. R. E. Dickerson, T. Takano, D. Eisenberg, O. B. Kallai, L. Samson, A. Cooper and E. Margolaish, J. Biol. Chem., 246, 1511 (1971).
2. R. E. Dickerson, J. Mol. Evol., 1, 26 (1971).
3. M. F. Perutz, Nature, 228, 726 (1970).
4. M. F. Perutz, Nature, 237, 495 (1972).
5. B. S. Hartley, Phil. Trans. Roy. Soc. Lond., B257, 77 (1970).
6. J. H. Northrup, M. Kunitz and R. Herriott, "Crystalline Enzymes," Columbia University Press, New York, 1939.
7. H. T. Wright, J. Kraut and P. E. Wilcox, J. Mol. Biol., 37, 363-366 (1968).
8. B. S. Hartley, in "Structure and Activity of Enzymes," T. W. Goodwin, J. T. Harris and B. S. Hartley, Eds., Academic Press, London, 1964, p. 47.
9. B. S. Hartley and D. L. Kauffman, Biochem. J., 101, 229 (1966).
10. B. Meloun, I. Kluh, V. Kostka, L. Moravek, Z. Prusik, J. Vanecek, B. Keil and F. Sorm, Biochem. Biophys. Acta, 130, 543 (1966).
11. D. M. Blow, J. J. Birktoff and B. S. Hartley, Nature, 221, 337 (1969).
12. R. Lumry and R. Biltonen in "Structure and Stability of Biological Molecules," S. N. Timasheff and G. D. Fasman, Eds., Marcel Dekker, Inc., New York, 1969, p. 65.
13. R. Lumry and S. Rajender, Biopolymers, 9, 1125 (1970).
14. G. P. Hess, J. McConn, E. Ku and G. McConkey, Phil. Trans. Roy. Soc. Lond., B257, 89 (1970).
15. M. L. Bender and F. J. Kezdy, Ann. Rev. Biochem., 34, 49 (1965).

16. S. T. Freer, J. Kraut, J. D. Robertus, H. T. Wright and Ng. H. Xuong, Biochem., 9, 1997 (1970).
17. B. W. Matthews, P. B. Sigler, R. Henderson and D. M. Blow, Nature, 214, 652 (1967).
18. P. B. Sigler, D. M. Blow, B. W. Matthews and R. Henderson, J. Mol. Biol., 35, 143 (1968).
19. G. H. Cohen, E. W. Silverton, B. W. Matthews, H. Braxton and D. R. Davies, J. Mol. Biol., 44 (1969).
20. M. V. King, Acta Cryst., 7, 601 (1954).
21. H. W. Wyckoff, M. Doscher, D. Tsernoglou, T. Inagami, L. N. Johnson, K. D. Hardman, N. M. Allewell, D. M. Kelly and F. M. Richards, J. Mol. Biol., 27, 363 (1967).
22. R. L. Vandlen and A. Tulinsky, Acta Cryst., B27, 437 (1971).
23. W. R. Busing and H. A. Levy, Acta Cryst., 22, 457 (1967).
24. P. W. Coddling, Ph.D. Thesis, Michigan State University, E. Lansing, Mich., 1971.
25. A. Tulinsky, N. V. Mani, C. N. Morimoto and R. L. Vandlen, to be published.
26. D. M. Blow, Proc. Roy. Soc., 247, 302 (1958).
27. D. Harker, Acta Cryst., 9, 1 (1956).
28. M. G. Rossman, Acta Cryst., 13, 221 (1960).
29. G. Kartha, J. Bello, D. Harker and F. E. DeJarnette in "Aspects of Protein Structure," G. N. Ramachandran, Ed., Academic Press, London, 1963, p. 13.
30. L. K. Steinrauf, Acta Cryst., 16, 317 (1963).
31. H. Muirhead, J. M. Cox, L. Mazzarella and M. F. Perutz, J. Mol. Biol., 28, 117 (1967).
32. R. E. Dickerson, J. S. Kendrew and R. E. Strandberg, Acta Cryst., 14, 1188 (1961).
33. A. C. North, Acta Cryst., 18, 212 (1965).
34. D. M. Blow and F. H. C. Crick, Acta Cryst., 12, 794 (1959).

35. F. M. Richards, J. Mol. Biol., 37, 225 (1968).
36. G. W. Schwert, J. Biol. Chem., 179, 655 (1949).
37. G. W. Schwert and S. Kaufman, J. Biol. Chem., 190, 807 (1951).
38. C. M. Venkatachalam, Biopolymers, 6, 1425 (1968).
39. A. Liljas, Ph.D. Thesis, Uppsala University, Uppsala, Sweden, 1971.
40. P. N. Lewis, F. A. Momany, and H. A. Scheraga, Proc. Natl. Acad. Sci., USA, 68, 2293 (1971).
41. R. M. Stroud, L. M. Kay and R. E. Dickerson, "Cold Spring Harbor Symposia on Quantitative Biology," Vol. XXXVI, 125 (1971).
42. D. M. Shotton and H. C. Watson, Phil. Trans. Roy. Soc. Lond., B257, 111 (1970).
43. E. F. Jansen, M. D. F. Nutting and A. K. Balls, J. Biol. Chem., 179, 201 (1949).
44. D. E. Fahreney and A. M. Gold, J. Amer. Chem. Soc., 85 997 (1963).
45. B. S. Hartley and B. A. Kilby, Biochem. J., 50, 672 (1952).
46. G. Schoellman and E. Shaw, Biochem., 2, 252 (1963).
47. J. R. Garel and B. Labouesse, J. Mol. Biol., 47, 41 (1970).
48. B. H. J. Hofstee, Biochem. Biophys. Research Comm., 41, 1141 (1970).
49. A. R. Fersht and Y. Requena, J. Mol. Biol., 60, 279 (1971).
50. A. R. Fersht, J. Mol. Biol., 64, 497 (1972).
51. K. L. Carraway, P. Spoerl and D. E. Koshland, Jr., J. Mol. Biol., 42, 133 (1969).
52. G. P. Hess in "The Enzymes," Vol. III, Paul Boyer, Ed., Academic Press, London, 1971.
53. A. Liljas, K. K. Kannan, P.-C. Bergsten, I. Waara, K. Fridborg, B. Strandberg, U. Carlbom, L. Jarup, S. Lovgren and M. Petef, Nature, 235, 131 (1972).

54. A. Tulinsky, P. W. Coddington and R. L. Vandlen, Abstracts, Am. Cryst. Assoc., Columbia, S. C., Feb. 1972, No. K3.
55. J. M. Cox, J. Mol. Biol., 28, 151 (1967).
56. G. H. Cohen, B. W. Matthews and D. R. Davies, Acta Cryst., B26, 1062 (1970).
57. T. L. Blundell, J. F. Cutfield, E. J. Dodson, G. G. Dodson, D. C. Hodgkin and D. A. Mercola, "Cold Spring Harbor Symposia on Quantitative Biology," Vol. XXXVI, 233 (1971).
58. L. H. Wright, Michigan State University, personal communication, 1972.
59. A. Tulinsky, R. L. Vandlen, C. N. Morimoto, N. V. Mani and L. H. Wright, submitted for publication in Nature, New Biology.
60. D. M. Segal, G. H. Cohen, D. R. Davies, J. C. Powers and P. E. Wilcox, "Cold Spring Harbor Symposia on Quantitative Biology," Vol. XXXVI, 85 (1971).
61. D. M. Segal, J. C. Powers, G. H. Cohen, D. R. Davies and P. E. Wilcox, Biochemistry, 10, 3728 (1971).
62. J. G. Kirkwood and J. B. Shumaker, Proc. Natl. Acad. Sci. U.S., 38, 855 and 863 (1952).
63. S. N. Timasheff, Biopolymers, 4, 107 (1966).
64. V. Nozaki and C. Tanford, J. Biol. Chem., 242, 4731 (1967).
65. R. Diamond, Acta Cryst., 21, 253 (1966).
66. W. N. Lipscomb, G. N. Reeke, J. A. Hartsuck, F. A. Quiocho, and P. H. Bethage, Phil. Trans. Roy. Soc. Lond., B257, 177 (1970).
67. C. Ramakrishnan and G. N. Ramachandran, Biophys. J., 5, 909 (1965).
68. R. Diamond, Acta Cryst., A27, 436 (1971).
69. M. Levitt and S. Lifson, J. Mol. Biol., 46, 269 (1969).
70. J. J. Birktoff and D. M. Blow, J. Mol. Biol., 68, 187 (1972).
71. C. Tanford, Advan. Protein Chem., 17, 69 (1962).

72. R. Egan, H. O. Michael, R. Schlueter and B. J. Jandorf, Arch. Biochem. Biophys. Acta, 94, 535 (1957).
73. K. C. Aune and S. N. Timasheff, Biochemistry, 10, 1609 (1971).
74. J. A. Gladner and H. Neurath, J. Biol. Chem., 206, 911 (1954).
75. C. S. Hexter and F. H. Westheimer, J. Biol. Chem., 246, 3928 (1971).
76. R. Timkovich, Michigan State University, unpublished results.
77. R. Henderson and J. K. Moffat, Acta Cryst., B27, 1414 (1971).
78. F. H. C. Crick and B. S. Magdoff, Acta Cryst., 9, 901 (1956).
79. A. D. Booth, Trans. Faraday Soc., 42, 444 (1946).
80. M. Shinitzky, E. Katchalski, V. Grisaro and N. Sharon, Arch. Biochem. Biophys., 116, 332 (1966).
81. IUPAC-IUB Commission on Biological Nomenclature, J. Mol. Biol., 52, 1 (1970).

APPENDIX

SIDE CHAIN ENVIRONMENT AND CLARITY IN THE
ELECTRON DENSITY OF α -CHT

<u>Residue</u>	<u>Environment</u>	<u>Comments*</u>
A Chain		
1 CYS	S	Good density in bridge to 122
2 GLY		
3 VAL	E	Weak density for side chain
4 PRO	S	Well defined density for ring
5 ALA	E	Some CB density
6 ILE	S	Weak but defined density
7 GLN	E	Some density for side chain
8 PRO	S	Weak density
9 VAL	E	Weak density
10 LEU		No definitive density
11 SER		for remainder of A chain
12 GLY		
13 LEU		
B Chain		
16 ILE	I	Forked side chain density, salt bridge to ASP 194
17 VAL	I	No density for CG2
18 ASN	E	No density beyond CG
19 GLY		
20 GLU	E	Weak density for side chain
21 GLU	E	Strong density
22 ALA	I	Good density
23 VAL	S	No density beyond CB
24 PRO	S	Good density
25 GLY		
26 SER	S	Side chain on edge of density
27 TRP	I	Very strong density
28 PRO	I	Good density
29 TRP	I	Very good side chain density
30 GLN	I	Good side chain density

<u>Residue</u>	<u>Environment</u>	<u>Comments*</u>
31 VAL	I	Good density
32 SER	I	CB on edge of density-- H bond to HIS 40
33 LEU	I	Weak density for CD1
34 GLN	S	Well defined side chain density
35 ASP	S	Density o.k.
36 LYS	E	Weak density to CD
37 THR	S	Strong density for side chain
38 GLY		
39 PHE	E	Good side chain density
40 HIS	S	Partially exposed to solvent well defined
41 PHE	I	Good density
42 CYS	S	Strong disulfide bridge to CYS 58
43 GLY		\C=O on edge of density
44 GLY		
45 SER	I	Well defined density
46 LEU	I	Forked density for side chain
47 ILE	I	Good definition
48 ASN	E	Good to CB, less to end of chain
49 GLU	E	Strong density for whole side chain
50 ASN	E	Complete, defined density
51 TRP	S	Good density
52 VAL	I	Slightly forked density
53 VAL	I	No density past CB
54 THR	I	Density for one branch only
55 ALA	I	CB on edge of density
56 ALA	I	Well defined
57 HIS	S	Well resolved density
58 CYS	S	Good density for bridge to CYS 42
59 GLY		
60 VAL	I	No density for CG2

<u>Residue</u>	<u>Environment</u>	<u>Comments*</u>
61 THR	S	Sufficient density
62 THR	E	Good density
63 SER	E	Good density
64 ASP	S	Salt bridge to ALA' 149
65 VAL	S	Density not resolved from main chain
66 VAL	I	Resolved density
67 VAL	S	Side chain on edge of density
68 ALA	S	CB in bump on main chain
69 GLY	E	
70 GLU	E	Poor density for main and side chain
71 PHE	S	Good density for ring
72 ASP	E	Hard to fit side chain to density
73 GLN	E	Poor density for side chain
74 GLY		
75 SER	E	Side chain in good density
76 SER	E	Poor density--solvent
77 SER	E	Poor density
78 GLU	S	Good side chain density
79 LYS	E	Weak side chain density
80 ILE	S	No density for one branch
81 GLN	E	Good density for whole side chain
82 LYS	E	Good density to CG, less to NZ
83 LEU	S	Good density for CG
84 LYS	S	Continuous density for side chain
85 ILE	S	No density for CD1
86 ALA	S	CB in bump on main chain
87 LYS	E	Weak density
88 VAL	S	Forked density
89 PHE	S	Density somewhat less than other phenyls
90 LYS	E	No density past CE

<u>Residue</u>	<u>Environment</u>	<u>Comments*</u>
91 ASN	S	Good density for whole side chain
92 SER	E	No density for OG
93 LYS	E	Complete, resolved density for side chain
94 TYR	S	Somewhat diffuse, typical of TYR residues
95 ASN	S	Good density--H bonding to $\text{SO}_4=$
96 SER	E	Good density
97 LEU	E	Different orientations in two molecules
98 THR	S	Complete side chain density
99 ILE	S	Forked density, good
100 ASN	S	Good density for side chain
101 ASN	S	Complete side chain density
102 ASP	I	Good density for carboxyl group
103 ILE	I	Slightly forked density
104 THR	I	Side chain on edge of solid density
105 LEU	I	Good to CG, weaker density for rest
106 LEU	I	Well defined density
107 LYS	E	Density weak beyond CG
108 LEU	I	Good density
109 SER	S	Side chain with H_2O molecule
110 THR	E	Density not resolved from main chain
111 ALA	E	No density for CB
112 ALA	S	Strong CB density
113 SER	E	Good density, with H_2O
114 PHE	S	Funny-shaped density for phenyl ring
115 SER	S	Density elongated, H_2O
116 GLN	E	No density past CA
117 THR	S	Side chain density not resolved from main chain

<u>Residue</u>	<u>Environment</u>	<u>Comments*</u>
118 VAL	S	Density to CB only
119 SER	S	Good density
120 ALA	S	Weak density for CB
121 VAL	I	Weak density past CB
122 CYS	S	Good bridge density to CYS 1
123 LEU	I	Slightly forked density
124 PRO	I	Well defined
125 SER	S	Well defined
126 ALA	S	More than enough density for CB
127 SER	E	Side chain good
128 ASP	S	No density for CB, good for COO ⁻
129 ASP	E	Weak density for carboxylate
130 PHE	S	Well defined
131 ALA	E	No CB density
132 ALA	E	Weak density for CB
133 GLY		Good main chain density
134 THR	E	No density for CG2
135 THR	S	Good density
136 CYS	S	Strong density--bridge to CYS 201
137 VAL	S	CG1 on edge of density
138 THR	I	Good density
139 THR	I	Poor density for side chain
140 GLY		Main chain density weak
141 TRP	I	Weak side chain density
142 GLY		
143 LEU	I	Density good to CG
144 THR	S	Good density with H ₂ O
145 ARG	E	Density good except for CB and CG
146 TYR	E	Density somewhat irregular for ring; carboxyl terminal density good

<u>Residue</u>		<u>Environment</u>	<u>Comments*</u>
C Chain			
149	ALA	E	Molecule I density weak, I' very good
150	ASN	S	Strong to CB, weaker for rest of side chain
151	THR	S	Diffuse density
152	PRO	S	Somewhat weaker than for other prolines
153	ASP	E	Sufficient density for side chain
154	ARG	E	Weak density, not definitive
155	LEU	I	Fair density
156	GLN	S	Good density
157	GLN	S	Weak but sufficient
158	ALA	S	CB on edge of density
159	SER	S	Good density
160	LEU	I	Branched density
161	PRO	S	Good definition
162	LEU	I	Excellent, forked density
163	LEU	S	Acceptable density
164	SER	E	Fair density
165	ASN	S	Fair density for side chain
166	THR	E	Sufficient density
167	ASN	S	Good, different orientation in I'
168	CYS	I	Good
169	LYS	E	Complete, orientation differ- ent in two molecules
170	LYS	E	Density for whole side chain
171	TYR	S	Good, molecule I has H ₂ O but not I'
172	TRP	S	Good, flat density
173	GLY		
174	THR	E	Density for side chain
175	LYS	S	No density for only CB
176	ILE	S	No density for CG2

<u>Residue</u>	<u>Environment</u>	<u>Comments*</u>
177 LYS	S	Weak density for NZ
178 ASP	E	Two possible orientations
179 ALA	S	CB on edge of density
180 MET	I	Strong density on sulfur
181 ILE	S	Density for one branch only
182 CYS	I	Strong density
183 ALA	I	Good CB definition
184 GLY		
185 ALA	I	
186 SER	E	Good density
187 GLY		
188 VAL	S	CB density only
189 SER	I	Sufficient for side chain
190 SER	I	Not very definitive
191 CYS	S	Strong
192 MET	E	Weak on I, strong density on I'
193 GLY		
194 ASP	I	Good, salt bridge to ILE 16
195 SER	S	Well defined
196 GLY		
197 GLY		
198 PRO	I	Flattened density
199 LEU	I	Forked density
200 VAL	I	Slightly forked
201 CYS	S	Strong density
202 LYS	E	Good density for whole side chain
203 LYS	E	Density extends to CB
204 ASN	E	o.k.
205 GLY		
206 ALA	S	No CB density
207 TRP	S	Excellent shape
208 THR	S	Density well resolved
209 LEU	I	Slightly forked

<u>Residue</u>	<u>Environment</u>	<u>Comments*</u>
210 VAL	I	Density for one branch only
211 GLY		
212 ILE	I	Good
213 VAL	I	Weaker density with two branches
214 SER	I	Good density
215 TRP	I	Less density than other TRP residues
216 GLY		
217 SER	E	Poor density for side chain
218 SER	E	Density fairly good
219 THR	E	Density good for one branch
220 CYS	S	Strong density
221 SER	E	Good density--with $\text{SO}_4^{=}$ ion
222 THR	S	Little density for side chain
223 SER	E	Good density
224 THR	E	Complete density--with $\text{SO}_4^{=}$ ion of SER 221
225 PRO	S	Strong density
226 GLY	I	
227 VAL	I	Branched density
228 TYR	I	Good definition, somewhat weak
229 ALA	I	
230 ARG	S	Good density for whole side chain
231 VAL	I	Adequate density
232 THR	S	Complete density for side chain, H_2O
233 ALA	S	
234 LEU	I	Forked density
235 VAL	S	No density for CG2
236 ASN	S	No density for CB, good for rest of side chain
237 TRP	S	Excellent density
238 VAL	I	Good definition
239 GLN	S	CB on weak density

<u>Residue</u>		<u>Environment</u>	<u>Comments*</u>
240	GLN	E	Very good definition
241	THR	S	Slightly forked
242	LEU	S	Forked, somewhat weak density
243	ALA	E	
244	ALA	S	Good density for CB
245	ASN	E	Strong density at end of side chain

*IUPAC-IUB conventions for peptide nomenclature (81) are used with Latin equivalents of the Greek superscripts.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 03177 3405