



RETURNING MATERIALS:

Place in book drop to
remove this checkout from
your record. FINES will
be charged if book is
returned after the date
stamped below.

Miss Jol. Book!
D. Dear

OCT 22 '88
1 00 A 24 0

PART I: CONFORMATIONAL ENERGY MINIMIZATIONS OF
γ-CHYMOTRYPSIN.
PART II: THE REFINEMENT AND STRUCTURE OF α-CHYMOTRYPSIN
AT 1.67 Å RESOLUTION.

By

Richard Alan Blevins

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Chemistry

1984

ABSTRACT

PART I: CONFORMATIONAL ENERGY MINIMIZATIONS OF
 γ -CHYMOTRYPSIN.

PART II: THE REFINEMENT AND STRUCTURE OF α -CHYMOTRYPSIN
AT 1.67 Å RESOLUTION.

By

Richard Alan Blevins

The three dimensional structure of the proteolytic enzyme γ -chymotrypsin has been studied with conformational energy minimization techniques. The studies addressed questions of structural stability, protein-solvent interactions, and protein-protein aggregation.

The largest perturbation of the crystallographically observed structure occurs upon energy minimization at the surface of the protein, specifically the dimer interface residues found in the alpha form of the enzyme. The larger changes in the interface residues are significant due to their implications in protein-protein and protein-solvent interactions.

The structure of γ -chymotrypsin surrounded by solvent has also been refined. The protein-solvent system was modeled using the Ferro and Hol "mobile solvation layer plus ice lattice model" of bulk solvent. A mechanical, non-thermodynamic surface tension was calculated for the γ -chymotrypsin monomer. Results indicate that the region of the monomeric γ -chymotrypsin protein corresponding to the interface of dimeric α -chymotrypsin possesses a

mechanical surface tension approximately twice that for the global protein exterior. These results suggest the possibility of predicting protein-protein interface sites when component structures are known but aggregates are not.

The structures of the two independent molecules of the α -chymotrypsin dimer have been refined using Hendrickson's PROLSQ refinement program. The refinement was initiated using an exact two-fold structure, coordinates of which were obtained from a model fitted to a 2-fold averaged electron density map. The trial structure calculated well at 3.0 Å resolution, the conventional R-factor being .364. Manual interventions were also performed using FRODO on an Evans and Sutherland PS300 interactive computer graphics system. A total of 247 probable water molecules were located and 97 cycles of least squares refinement were performed giving a final R-factor of 0.179.

The final structure of the α -chymotrypsin dimer has a root mean square asymmetry of 0.24 Å for the main chain atoms and 0.64 Å for the side chains. The total r.m.s. shift from the trial structure is 0.50 Å and 1.02 Å for main and side chain atoms, averaged between the two molecules. Most of the asymmetry resides in the configurations and conformations of the side chain atoms.

To my entire family for their support and
encouragement throughout my
many years of education.

ACKNOWLEDGMENTS

For their guidance, support and friendship during this study, I wish to extend my sincere gratitude to Dr. Alexander Tulinsky and Dr. Paul Hunt.

The author wishes to thank Dr. B.A. Karcher, Dr. C.E. Buck, Dr. M. Ragazzi and J.R. Zimmermann for stimulating discussions and contributions during all aspects of this work.

TABLE OF CONTENTS

CHAPTER		PAGE
	LIST OF TABLES.	vi
	LIST OF FIGURES.	viii
	PART I: CONFORMATIONAL ENERGY MINIMIZATIONS OF γ -CHYMOTRYPSIN.	
I	INTRODUCTION.	1
II	CONFORMATIONAL ENERGY CALCULATIONS.	7
	A. Refinement Strategies.	7
	B. γ -Chymotrypsin Refinements.	8
	C. The Extended Atom Implementation.	11
	D. Parameter Choices.	12
	E. Preparatory Steps Before Refinement.	18
III	THE MOBILE SOLVATION LAYER PLUS ICE LATTICE MODEL	20
	A. The Model.	20
	B. Generation of Trial Structures.	21
IV	CHYMOTRYPSIN REFINEMENT RESULTS.	28
	A. Series I-a, I-b, II-a.	28
	B. Crystallographic Analysis.	45
V	PROTEIN-PROTEIN ASSOCIATION AND MECHANICAL SURFACE TENSION.	51
	A. Mechanical Surface Tension Calculations.	53
	B. Surface Tension Results.	55
	PART II: THE REFINEMENT AND STRUCTURE OF α -CHYMOTRYPSIN AT 1.67 Å RESOLUTION	
VI	INTRODUCTION.	60
	A. Refinement Methods Based on X-Ray Data.	60
	B. PROLSQ — Restrained Least Squares Refinement.	63
	1. Structure Factors.	65
	2. Bond Distances.	66

CHAPTER	PAGE
3. Planar Groups.	66
4. Chiral Centers.	66
5. Non-bonded Contacts.	67
6. Torsion Angles.	67
C. Graphics Intervention and FRODO.	68
VII REFINEMENT OF THE α -CHYMOTRYPSIN DIMER.	76
A. Experimental.	76
B. Refinement Summary.	77
VIII RESULTS OF THE LEAST SQUARES REFINEMENT.	89
A. The Independent Molecules.	89
B. Solvent Structure.	106
C. Dimer Asymmetry.	113
D. The Active Site.	130
E. The ILE-16, ASP-194 Ion Pair.	136
F. The Specificity Site.	139
G. The TRP Cluster.	142
H. Side Chain Asymmetry.	143
IX ENERGETIC ANALYSIS.	146
X COMPARISON OF THE INDEPENDENT MOLECULES OF α -CHT WITH γ -CHT.	152
A. Active Sites.	155
B. Specificity Sites.	161
C. The TRP Clusters.	162
D. Hydrogen Bonding.	162
E. Solvent Structure.	164
F. Concluding Remarks.	169
LIST OF REFERENCES.	170
APPENDIX A: Extended Atoms and their Non- Bonded Parameters.	175
APPENDIX B: Residues found in Specific CHT Regions.	176
APPENDIX C: Variable Dihedral Angles in the α -CHT Dimer.	178
APPENDIX D: Hydrogen Bonds in the α -CHT Dimer.	187
APPENDIX E: Solvent Molecule Positions in the α -CHT Dimer.	194
APPENDIX F: Protein-Solvent Hydrogen Bonds in the α -CHT Dimer.	195
APPENDIX G: Polar Protein Atoms — Solvent Interactions in the α -CHT Dimer.	197

LIST OF TABLES

TABLE	PAGE
1 Fractional Charges in Active Site Region.	10
2 Sample Weighting Scheme Used in Energetic Refinements.	17
3 Final Energies of the Refined γ -CHT Structures (kcal/mole).	32
4 R.M.S. Deviations from the Observed γ -CHT Structure for the Energetic Refinements.	33
5 Tau-Angle Distributions for the Energetic Refinements of γ -CHT.	34
6 Omega-Angle Distributions (absolute value) for the Energetic Refinements of γ -CHT.	35
7 R.M.S. Deviations Between Energetically Refined γ -CHT Structures.	42
8 R-Factors for the γ -CHT Energy Refinements.	47
9 Surface Area and Mechanical Surface Tension Results.	56
10 Summary of Least Squares Parameters and Deviations.	85
11 Dihedral Angles of Disulfide Bridges of the Independent Molecules of α -CHT.	95
12 Asymmetry of Hydrogen Bonding in α -CHT.	99
13 Average Thermal Parameters for the α -CHT Dimer ($\text{\AA}^2$).	103
14 R.M.S. Asymmetry for the α -CHT Dimer.	114
15 Dimer Interface Interactions in α -CHT.	126

TABLE		PAGE
16	Hydrogen Bonds in the Catalytic Sites a.) Involving Protein Atoms, b.) Involving Solvent Atoms.	134
17	Two-Fold Symmetric Water Molecules in the α -CHT Dimer.	135
18	Asymmetry Classified by Residue Type in the α -CHT Dimer.	144
19	Calculated Energies and Surface Areas for the α -CHT Dimer.	150
20	Transformations Relating γ -CHT to α -CHT.	153
21	R.M.S. Differences between the Independent Molecules of α -CHT and γ -CHT.	154
22	Main Chain Hydrogen Bond Differences with Respect to γ -CHT.	163
23	Equivalent Water Molecules (within 1.0 Å) in both α -CHT and γ -CHT.	166

LIST OF FIGURES

FIGURE	PAGE
1 Extended Atom Hydrogen Bond Potential Energy. . .	16
2 Summary of Ice Lattice Generation and Final Results.	24
3 Slices through x-y Plane for the Cubic and Face Centered Cubic Protein Plus Ice Lattice Systems. Movable solvent shaded; cubic lattice, top; face centered cubic lattice, bottom.	26
4 Stereoview of CA Atoms of the γ -CHT Monomer. Dimer interface residues shaded.	30
5 Ramachandran Plot for the Final Structure Resulting from Refinement Series I-a. Glycines represented as circles.	37
6 Examples of FRODO Graphical Displays. (a) stick diagram, (b) stick diagram plus electron density, (c) van der Waals surface. . .	72
7 Progress of R-Factor During Refinement. The resolution stages and FRODO interventions are indicated.	80
8 Progress of Asymmetry Development and Shifts During Refinement. Diamonds and squares, main and side chain asymmetry; triangles and circles, main and side chain shifts with respect to the trial structure, respectively. . .	82
9 Variation of R-Factor with Scattering Angle. Triangles 3.0 Å, squares 2.5 Å, diamonds 2.0 Å, and inverted triangles 1.67 Å resolution; broken lines are theoretical curves for 0.15, 0.18 and 0.20 Å coordinate error.	86
10 Omega-Angle Distribution. Molecule 1, top and molecule 2, bottom.	91

FIGURE	PAGE
11	Ramachandran Plots of α -CHT. Molecule 1, top; molecule 2, bottom; GLY not included. 94
12	Histograms of α -CHT Hydrogen Bond Distances and Donor-Hydrogen-Acceptor Angles. Molecule 1 left; molecule 2, right. 98
13	Distribution of Some Side Chain Conformational Angles. (a) SER, (b) THR, (c) VAL, (d) ILE and LEU. 102
14	R.M.S. Thermal Parameters of α -CHT. Main chain, solid; side chain, broken; molecule 1, top; molecule 2, bottom. 105
15	Distribution of Occupancies (top) and Thermal Parameters (bottom). Occupancies greater than one were set to one during refinement. 110
16	Distribution of Solvent-Protein (top) and Solvent-Solvent (bottom). All solvent-solvent minimum distances >5.0 Å grouped together. 112
17	R.M.S. Asymmetry Between Individual Molecules of α -CHT. Only atoms with $B < 23.0$ Å ² are included; main chain, solid; side chain, broken; a-dimer interface regions, b-dyad B regions near noncrystallographic 2-fold axis between dimers, c-external turns. 118
18	Stereo CA Plots of the α -CHT Dimer. Top — view down XO (local 2-fold axis), bottom — view down YO (2-fold axis can be seen. 121
19	Stereoview of Representative Surface Asymmetry. Residues 172-179; molecule 2 bold. 123
20	Stereoview of Overall Asymmetry of α -CHT. Viewed down local 2-fold axis designated by asterisk; side chains shown only if r.m.s. asymmetry >0.5 Å and $\langle B \rangle < 23.0$ Å; main chain atoms corresponding to these residues are also shown. 125
21	Stereoview of Typical Dimer Interface Asymmetry. Residues 35-41; molecule 2, bold. 129

FIGURE

PAGE

- 22 Stereoview of Catalytic Residues of Independent Molecules of α -CHT. HIS 57, ASP-102, SER-195; molecule 2, bold.132
- 23 Stereoview of Active Site Regions of α -CHT. Included are solvent and TYR-146 of the other molecule. Molecule 1, top; molecule 2, bottom. Solvent common to both shaded. 138
- 24 Stereoview of Specificity Site Regions of α -CHT. Included are solvent molecule 1, top; molecule 2, bottom. Symmetric waters shaded.141
- 25 Progress of Dimerization Energy during Refinement. Resolution Stages are indicated. 149
- 26 R.M.S. Differences between α -CHT and γ -CHT. Only atoms with $B < 23.0 \text{ \AA}^2$ (α -CHT) and $< 15.0 \text{ \AA}^2$ (γ -CHT) included; molecule 1, top; molecule 2, bottom, main chain, solid, side chain, broken; intermolecular contacts in α - and γ -CHT.157
- 27 Stereoviews of Superpositions of the Catalytic Site Regions of α -CHT and γ -CHT. Molecule 1 — γ -CHT, top; molecule 2 — γ -CHT, bottom. γ -CHT, bold in each case.160

PART I: CONFORMATIONAL ENERGY MINIMIZATIONS OF
 γ -CHYMOTRYPSIN.

CHAPTER I

INTRODUCTION

The application of energy minimization techniques to the study of protein conformations was pioneered by Scheraga and coworkers¹ and Lifson and coworkers.² Energy minimization techniques have been used in theoretical studies of protein folding,³ intramolecular motion in proteins and nucleic acids⁴ and the energetics of activated processes in proteins.^{5,6}

Atomic coordinates derived from x-ray diffraction studies of protein crystals are subject to uncertainties which can be of the order of several tenths of an Angstrom. If the x-ray structure is to be used to study enzyme-substrate or enzyme-enzyme interactions, the protein must be in a low energy conformation so that these types of interactions may be examined. Conformational energy calculations are proving useful in elucidating how interatomic interactions dictate stable conformations of polypeptides and proteins, along with their intermolecular complexes.

Rigorous quantum mechanical methods, ab initio and semiempirical, have increased in power over the last few

decades. The size of problems of biological interest, however, requires the use of the most elementary model empirical energy functions. The basic assumption of elementary energy calculations is that one may replace the Born-Oppenheimer energy surface by a computationally convenient sum of analytical functions. In most cases, the potential energy function is chosen as a sum of approximate strain energies and non-bonded interactions. A basic framework has emerged from this early work. First, a suitable energy function that will describe the potential energy of the molecule accurately is found. Second, using the above energy function, the conformation of the molecule is adjusted to achieve a stable equilibrium structure. An assumption inherent in the above approach is that the biologically active protein conformation does possess a low potential energy.

Several effects are neglected in empirical energy functions that are routinely used. The first is the anharmonic form of the potentials far from the potential minima. Second, the non-bonded interactions employed neglect charge-induced dipole terms and three-body polarization effects.¹⁰ Most of the algorithms employed today reflect a balance between computational efficiency and accuracy.

Various types of molecular force fields have been proposed for use in the study of biological macromolecules. Most agree in general respects; however, there are differences in detail and the numerical values of the

parameters may show large variations. Although energy functions applied to smaller molecules have usually included hydrogen atoms explicitly, those for polypeptides sometimes exclude them in order to reduce the number of free variables. Some or all of the hydrogen atoms are merged with their attached atoms to form "extended atoms". Loss in precision is sure to result from this type of approximation, but a quantitative assessment of the "extended atom" method in energetic refinements has not yet been performed.

Various criteria may be used to estimate the agreement between calculated and observed equilibrium structures. The r.m.s. deviation of carbon alpha atoms measures the agreement of the backbone conformations. The deviations of side chain atoms can also be used, but in this case, these atoms are affected by solvent and other molecules in the crystal; care must be taken if these atoms are to be considered. Also, differences in the main chain hydrogen bonding can be used to measure whether certain interactions are important in both structures.

A question that is only recently being studied is the importance of solvent in energy calculations, and how its effect can be modeled in a simple and accurate fashion. Protein crystals contain anywhere from 20 to 80% solvent, often containing a high molarity of salt or organic precipitating agent.⁷ The majority of solvent molecules cannot be located as discrete maxima in electron density maps from x-ray studies. Since most of the solvent appears

to be very mobile and to possess a fluctuating structure, at present only a statistical description is possible. It is clearly desirable to understand and describe the structure of water near the protein surface in terms of protein-water and water-water interactions.

Chymotrypsin is an enzyme of considerable potential importance as a generalizable model for the study of protein-protein and protein-solvent interactions. Although there is general interest in the microscopic energetics and dynamics of protein molecules, attention is now centering on the aggregation of biomolecules, through dimer- or oligomerization or through the formation of heteromolecular complexes. Conformational energy calculations on the structure of γ -chymotrypsin may prove useful in addressing some basic questions concerning not only its structure, function and specificity but also protein-protein and protein-solvent interactions.

The structure of γ -chymotrypsin was chosen for study due to the fact that it has been refined to high resolution.¹¹ Additional and more extensive work may be done on the dimeric form of the enzyme now that it has been refined crystallographically.¹² γ -Chymotrypsin, hereafter denoted γ -CHT, is composed of 241 amino acid residues, arranged in three polypeptide chains of 13, 131 and 97 residues. γ -CHT exists as a monomer at neutral pH, whereas the alpha form (α -CHT) crystallizes as a dimer at pH 3.5.

The chymotrypsin enzyme is of particular interest as a model for protein aggregation since the alpha form is asymmetric^{12,13} and alpha crystals exposed to substrate analogs and irreversible inhibitors show asymmetric binding in the catalytic sites of the dimer.^{14,15}

The availability of a high resolution, crystallographically refined γ -chymotrypsin structure affords an excellent opportunity to qualitatively and quantitatively examine different types of energy minimization methods as applied to protein molecules. The reliability of the "extended atom" method may be shown in terms of agreement of energetically refined structures with the crystallographically observed structure. The inclusion of the crystallographic water molecules found in the γ -chymotrypsin molecule may be studied in terms of their effect on the structure (global energy and atomic forces) in conformational energy refinement. A study of the γ -chymotrypsin molecule, including bulk solvent, may further the understanding of protein-solvent interactions at the protein surface. Also, protein-protein and protein-solvent interactions at the corresponding dimer interface residues of the monomeric γ -chymotrypsin enzyme may explain the stability gained in the dimerization process by the dimeric α -chymotrypsin. Preliminary results will show that in the presence of bulk solvent, the dimer interface residues of the monomeric γ -chymotrypsin possess a local, non-thermodynamic molecular surface tension approximately

twice that of the exterior residues globally. As a consequence, this region would tend to internalize preferentially upon dimerization. The application of local molecular surface tension techniques may prove to be a valuable tool in the study of oligomeric systems where component structures are known but the aggregates are not.

CHAPTER II

CONFORMATIONAL ENERGY CALCULATIONS

A. Refinement Strategies

In the previous chapter, fundamental questions concerning protein conformational energy minimizations were outlined. Additional questions that have developed during the development of energy minimization techniques concern the problem of generating low-energy conformations of proteins which are acceptable by crystallographic standards, and the dilemma of whether energy minimization and crystallographic refinement should be used together in the refinement of protein structures. A test of this type of hybrid procedure has recently been performed on the bovine pancreatic trypsin inhibitor (BPTI) by Fitzwater and Scheraga.⁸ There, a potential energy-constrained real space refinement method was developed for use with diffraction data of medium to low resolution.

In real space refinements of protein molecules, the model is adjusted to minimize the following function:

$$\int (\rho_o - \rho_m)^2 dv \quad (1)$$

where ρ_o is the observed electron density and ρ_m is the density associated with the model. An objection to this type of method could be that by choosing real space, the new electron density map is biased toward the phasing model used to obtain it. As an alternative, a potential energy-constrained reciprocal space method may be employed.⁹ A comparison of the results of similar refinements on BPTI has been performed. The reciprocal space refinement resulted in a final structure with a lower R-factor, but the real space refinement method displayed a lower r.m.s. (root mean square) shift from the crystallographic structure.

B. γ -Chymotrypsin Refinements

In all energy refinements discussed below, the 1.9 Å resolution crystallographic structure of the globular serine protease γ -chymotrypsin reported by Cohen, Silverton and Davies¹¹ served as input in the conformational energy calculations.

A wide variety of conformational energy refinements were performed on the structure of γ -CHT, some including the crystallographically observed water molecules and two including bulk solvent.¹⁶ The refinements may be grouped into three series. In series I, with the exception of half-electron charges on the carboxylic group of the side chain of ASP-194 and the side chain of ILE-16, zero net charge was assigned to all ionizable groups. These

half-electron charges were used to represent a salt bridge, which has been proposed to be an integral factor in the enzyme's functionality.¹⁷ The basic criticism of the use of zero net charge is the fact that γ -CHT crystals are grown at or near pH 5. At this pH, many of the basic amino acid residues, the carboxylic terminal residues, and a large fraction of the glutamate and aspartate residues are ionized. Therefore, the series II refinements were performed using fractional charges for all ionizable side chains.¹⁸ Unlike ASP-194 and ILE-16, neither HIS-57 nor SER-195 carried a net charge in either series I or II, although the imidazole ring of HIS-57 is strongly polarized in the series I refinements. A list of side chain fractional charges employed for the active site region and the proposed salt bridge are listed in Table 1. The series III refinements attempted to model the effects of bulk solvent on the structure and energetics of γ -CHT. Also, series III results were used in the calculation of approximate local mechanical surface tensions, in an attempt to predict protein-protein interface sites. Two different refinements were performed in series III. In the first, a realistic diamond ice lattice surrounded γ -CHT. In the second, a simple cubic ice lattice was employed. Side chain fractional charges were not employed in either of the series III refinements. The generation of these bulk solvent structures will be discussed below.

Table 1: Fractional Charges in Active Site Region.

	Aspartic Acid #102		Aspartic Acid #194		Serine #195		Isoleucine #16		Histidine #57	
	I	II	I	II	I	II	I	II	I	II
N	-.077	-.110	-.077	-.110	-.070	-.070	-.080	-.080	-.080	-.080
CA	.070	.070	.070	.070	.070	.070	.085	.085	.080	.075
CB	0.	-.150	0.	-.150	0.	.118	.188	.018	.010	.010
CG	0.	.365	.270	.365					.050	.050
CG2							.192	.012		
CG1							.058	-.012		
CD2									-.015	-.015
CD1							.062	-.018		
OG					0.	-.118				
OD2	0.	-.575	-.385	-.575						
OD1	0.	-.575	-.385	-.575					.220	.017
ND1									.135	.145
CE1									-.40	-.190
NE2									.350	.348
C	.387	.355	.387	.355	.360	.360	.350	.350	.350	.360
O	-.380	-.380	-.380	-.380	-.360	-.360	-.355	-.355	-.350	-.360

C. The Extended Atom Implementation

In all of the chymotrypsin energy refinements, a locally modified "extended atom" method of representing a protein was employed using a standard dictionary of ideal bond lengths, bond angles, dihedral angles and force constants. The modification to the standard extended atom method⁵ provided an enhanced representation of hydrogen bonding; polar hydrogens were incorporated according to the following procedure. All polar hydrogens, including those of solvent, were added to the crystallographic structure in geometrically idealized positions. During the course of all the energy minimizations, the positions of the hydrogens were constrained to remain approximately ideal by assignment of very high force constants to the polar hydrogen to donor (or donor chain) bonds. The polar hydrogens were excluded from the set of calculated non-bonded interactions. Since no donor-hydrogen and hydrogen-acceptor parameters exist in the extended atom dictionary adopted, geometrical parameters from a refinement method incorporating all hydrogens¹⁹ were used for both polar hydrogens and for donor atoms.

The extended atom approximation has been shown to provide a satisfactory representation of the internal vibrations and bulk properties of small molecules and simple peptides.¹⁰ There are advantages and disadvantages to the extended atom approximations. Some of the

advantages are: a) their use significantly reduces the computational size of the problem, in most cases by a factor of two, b) fewer non-bonded interactions and internal degrees of freedom result and c) in most cases, hydrogen atom coordinates are unobserved and must be inferred from the non-hydrogenic coordinates obtained from the x-ray crystallographic study. Some possible disadvantages include: a) unless polar hydrogens are used, it is very difficult to represent hydrogen bonding, b) there is a loss of steric effects arising from hydrogens, as an extended atom is always spherical, c) hydrogen atom coordinates are necessary for some forms of analysis (e.g., proton and ^{13}C NMR phenomena). A list of extended atoms employed, along with their corresponding non-bonded and hydrogen bond parameters, is given in Appendix A.

D. Parameter Choices

Preparations and parameter choices for the chymotrypsin conformational energy refinements will now be described. All energetic refinements were performed using the program REFINE,²⁰ locally modified from a Univac version to run on a VAX 11-750.^{8,9} Unlike some energy minimization methods now in use, REFINE does not include electron density constraints (real space or reciprocal space). The energy is expressed as a sum of non-bonded and bonded contributions:

$$E = \frac{1}{2} w_{\ell} \sum_i \bar{k}_{\ell_i} (\ell_i - \ell_{o_i})^2 + \frac{1}{2} w_{\theta} \sum_j \bar{k}_{\theta_j} (\theta_j - \theta_{o_j})^2 + \frac{1}{2} w_{\rho} \sum_k \bar{k}_{\rho_k} (\rho_k - \rho_{o_k})^2 \\ + w_{NB} E_{NB} + \frac{1}{2} w_{\phi} \sum_p \bar{V}_{\phi_p} [1. - \cos(n(\phi_p - \phi_{o_p}))] \quad . \quad (2)$$

where $\bar{k}$ is a force constant, $\bar{V}$ is a potential minimum, and n represents the number of dihedral rotational minima. In Equation 2, the summation indices i, j, k, p run over all bond lengths (ℓ_i), bond angles (θ_j), "frozen" dihedrals (ρ_k), and free dihedrals (ϕ_p), respectively. In each case, the subscript zero denotes an ideal geometric value. The non-bonded contribution (van der Waals and electrostatic), E_{NB} , represents the sum over all the pairs of non-bonded atoms at less than 6 Å separation chosen as the cut-off distance. For the n 'th pair of atoms between which hydrogen bonding is impossible, the energy is computed as:

$$E_m^{NB} = A_m r^{-12} - C_m r^{-6} + w_{NB}^{-1} Q_m D^{-1} r^{-1} \quad . \quad (3)$$

For the interaction between possibly hydrogen bonded (non-hydrogenic) atoms, the angle η formed by them with the H atom at the vertex and the hydrogen to acceptor distance (d_{HA}) were computed: if $\eta > 90^\circ$ and $d_{HA} > 3.5 \text{ Å}$, the following expression was evaluated:

$$E_m^{HB} = E_m^{NB} \sin^2 \eta + (A_m^{HB} r^{-12} - C_m^{HB} r^{-10}) \cos^2 \eta \quad . \quad (4)$$

Thus Equation 4 provides a smooth transition from hydrogen bonding to a simple non-bonded interaction as η is decreased from 180° . A plot of a typical extended atom hydrogen bond potential energy curve is presented in Figure 1. In the electrostatic term, Q_m is the charge product of the m-th interacting pair.

The computationally convenient assumption $D=r$ was made, which is consistent with other work.^{10,18} The use of a distance dependent dielectric term introduces an approximate screening effect. Several additional methods of calculating the electrostatic energy are now being tested. These include the use of a constant dielectric, a shifted dielectric and "electrostatics by groups".¹⁰

In a conformational energy minimization, the function actually minimized is

$$f = \epsilon + \frac{1}{2} w_T \sum_t |\bar{x}_t - x_{ot}|^2 \quad (5)$$

In Equation 5, $\bar{x}_t$ represents the trial coordinate vector and $\bar{x}_{ot}$ the corresponding initial vector of the t-th atom. In Equations 2, 3 and 5, the w factors are weights that are varied during the energy refinement to accelerate convergence.¹⁹ A sample weighting scheme used in a typical refinement is given in Table 2. Strong geometrical similarity constraints (i.e. high w_T values) are usually imposed during the early stages of a refinement, thus maintaining structural ideality. As the refinement proceeds,

Figure 1. Extended Atom Hydrogen Bond Potential Energy Curve.

Hydrogen Bond Potential Energy Curve

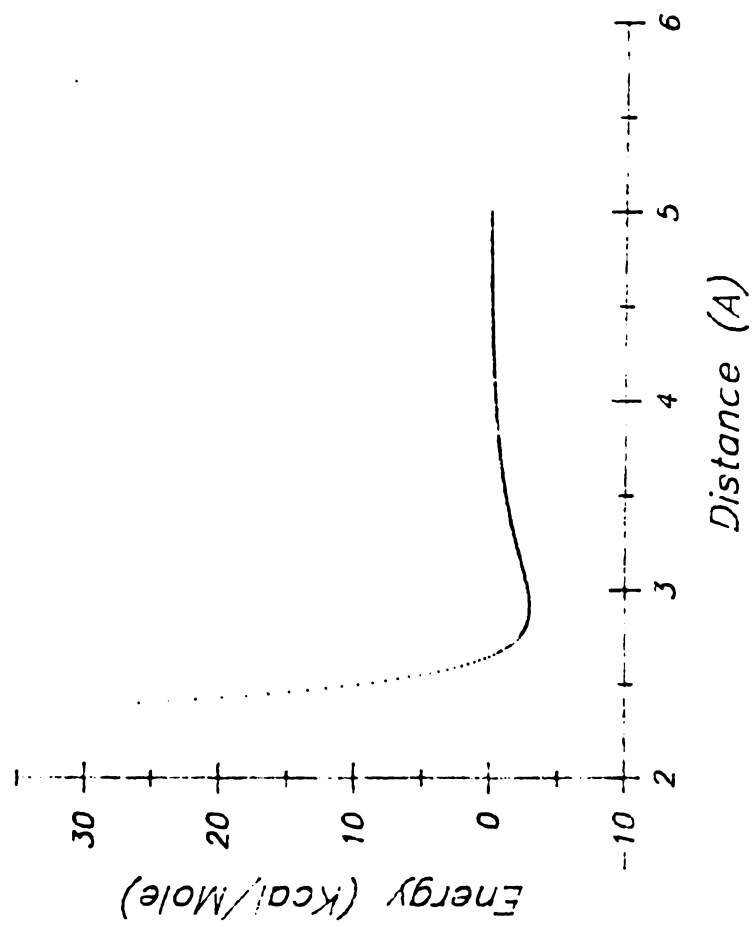


Table 2: Sample Weighting Scheme used in Energetic Refinements.^a

<u>Number of Cycles</u>		<u>w_t</u>	<u>w_l</u>	<u>w_Θ</u>	<u>w_{NB}</u>
1-20	Initial	1000.	.05	0.5	1.0
	Final	1.	.10	0.5	1.5
21-40	Initial	1.	.10	0.5	1.5
	Final	0.	.30	0.8	1.0
41-60	Initial	0.	.30	0.8	1.0
	Final	0.	1.0	1.0	1.0
61-80	Initial	0.	1.0	1.0	1.0
	Final	0.	1.0	1.0	1.0

^aWeights for torsional and electrostatic energy are unity throughout.

w_{NB} , is increased while the other weights are reduced. At the conclusion of the refinement, all weights are set to unity except w_T , which is zero. The effect of the w_{NB} term in Equation 3 is to give the electrostatic interactions unit weight during all cycles of the refinement. Similarly, $w_\rho = w_\phi = 1$ in all cycles.

E. Preparatory Steps Before Refinement

Prior to the actual energy refinement, other features of the REFIN program were used to optimize the γ -CHT protein structure. The ND1 and OD1 pair and the OE1 and NE2 pair were rotated by 90° about the CG and CD atoms of the ASN and GLN residues, respectively. The final conformation of the side chain was dictated by the lower conformational energy. This procedure is a way of correcting for the crystallographic indistinguishability of nitrogen and oxygen atoms.

During an energetic refinement, most of the effort at the beginning is concentrated on geometrical idealization. Considerable CPU time can be saved by the process of model building before the actual energy minimization is begun. Here, the time consuming process of calculating a list of non-bonded interactions is not needed. At the end of a model building state (usually about 40-50 cycles), all bond lengths are within $0.05 \text{ \AA}$, bond angles within 5° and dihedral angles 20° of the ideal values set in the

extended atom dictionary. This procedure almost always creates a few short non-bonded contacts. Because such short contacts produce anomalously large interaction energies, they are removed by selective local energy minimizations prior to the global energy refinement.

Depending on the size of the protein structure being studied, REFIN output is assumed to be effectively converged when the global energy change per cycle falls below 1.5-2.0 kcal/mole. This energy cut-off for convergence was chosen due to the steepest descent algorithm employed in REFIN. The rate of convergence is much slower than that of a conjugate gradient method.²¹ Additionally, the shifts of the atoms may be monitored, and convergence may be decided by an overall root mean square shift per cycle, as is done in other types of refinements.²²⁻²⁵

CHAPTER III

THE MOBILE SOLVATION LAYER PLUS ICE LATTICE MODEL

Crystallography has shown that a significant portion of the first shell of waters surrounding a protein is highly ordered.²⁶ Although a complete description of a protein-water system can only be achieved through statistical or dynamical methods, energy minimization may be useful in locating stable solvent. Low energy starting configurations may be generated and used as input for additional, more definitive work.

A. The Model

Ferro and Hol have developed a model for the study of protein-solvent interactions. Their model is particularly well suited for the REFIN system of programs and involves a mobile solvation layer plus ice lattice.¹⁰

An isolated protein may be described as surrounded by two layers of water molecules. The inner layer contains all water molecules that interact significantly with protein atoms, and is thick enough for protein and solvent atom rearrangement. All water molecules lying within a

chosen distance R_s from a non-hydrogen protein atom are taken to belong to the inner layer. They are free to move in the energy refinement. The bulk solvent surrounding the protein is represented by the outer layer. In order to bound the system, only those molecules within a chosen distance R_c from a non-hydrogen protein atom are included, and their main interactions are with other waters rather than the protein.

B. Generation of Trial Structures

Two types of ice lattices were generated in this work, a simple cubic ice lattice as suggested by Ferro and Hol and a more realistic face-centered cubic ice lattice. In both cases, the following steps were taken to generate the trial structure. First, a cubic box was generated such that when the protein was placed in the middle of the box no protein atom was closer than $R[\text{HOH}] + R[\text{VDW}]$, where $R[\text{HOH}]$ is a chosen van der Waals radius of water ($1.4 \text{ \AA}$) and $R[\text{VDW}]$ is the van der Waals radius of the protein atom under consideration. No hydrogen atoms were considered here and the dimensions of the box were chosen such that every protein atom was at least $10 \text{ \AA}$ from the edge of the box. Many different trial protein-solvent systems were generated by translating and rotating the lattice system with respect to the protein system and by thermally randomizing the positions of the lattice sites. The rotation was

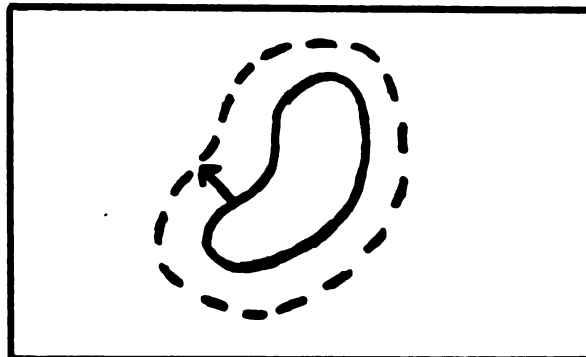
accomplished by generating a rotation matrix from a given set of Euler angles.²⁷ In thermally randomizing the coordinates of the lattice sites, random deviates from the surface of a unit sphere⁷⁴ were generated to fix the orientation and an appropriate random number chosen as a function of the unit cell edge of the lattice fixed the magnitude. Second, lattice sites that were within the cut-off distance R_s were classified as movable. Hydrogen atoms of all water molecules were added in geometrically ideal positions. Third, a low-energy model of the solvent was first created by minimizing the energy of the solvation shell (inner layer). Once this was accomplished, the entire protein-solvent system was refined.

A graphical representation of the mobile solvation plus ice lattice model is presented in Figure 2. The results of the model calculations were used as input for the series III refinements. Figure 2 indicates the number of free and fixed water molecules, the van der Waals radii employed and the final densities of the protein-bulk solvent systems. Figure 3 displays a section through each of the two types of ice lattices generated, indicating clearly the simple cubic and face centered cubic ice lattice structures.

Since no potentials exist for the protein-water and water-water interactions in the extended atom dictionary adopted, non-bonded and geometrical parameters for water oxygens were used from other work¹⁹ (see Appendix A). A

Figure 2. Summary of Ice Lattice Generation and Final Results.

<u>Cubic ice lattice</u>		<u>Face Centered Cubic ice lattice</u>	
Cell edge	3.1034 Å	Cell edge	6.38 Å
Density	1.000 G/CC	Density	.93 G/CC



All water molecules were removed if the oxygen was closer than $(R_i + R_w)$ to any protein atom.

R_i = vdW radius of the protein atom

R_w = effective vdW radius of water

Movable layer of waters at R_s

Results:

Cubic ice lattice:

$$R_s = 8 \text{ Å}$$

Density = .98 G/CC

Movable waters 1812

Total waters 14126

Diamond ice lattice:

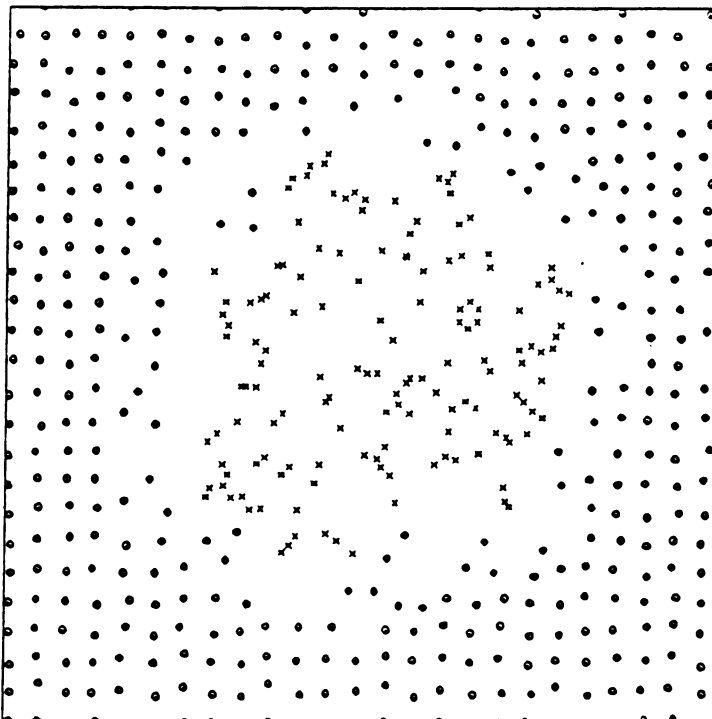
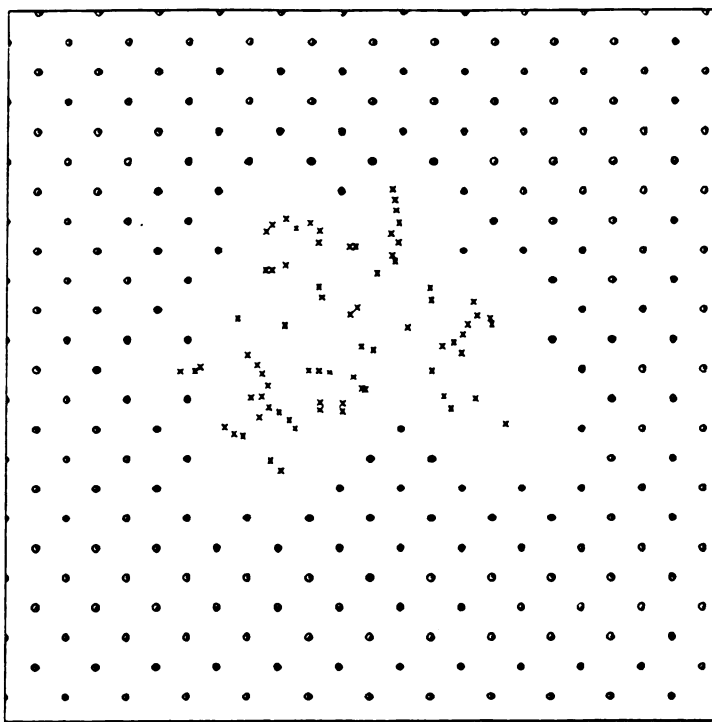
$$R_s = 8 \text{ Å}$$

Density = .94 G/CC

Movable waters 1484

Total waters 12634

Figure 3. Slices through x-y Plane for the Cubic and Face Centered Cubic Protein Plus Ice Lattice Systems. Movable solvent shaded; cubic lattice, top; face centered cubic lattice, bottom.



complete extended atom description of protein-solvent interactions is needed. Hermans and co-workers have developed a new solvent model for use in the extended atom approximation.^{7,28,29} Their "simple point charge" model is a first step at addressing this question.

CHAPTER IV

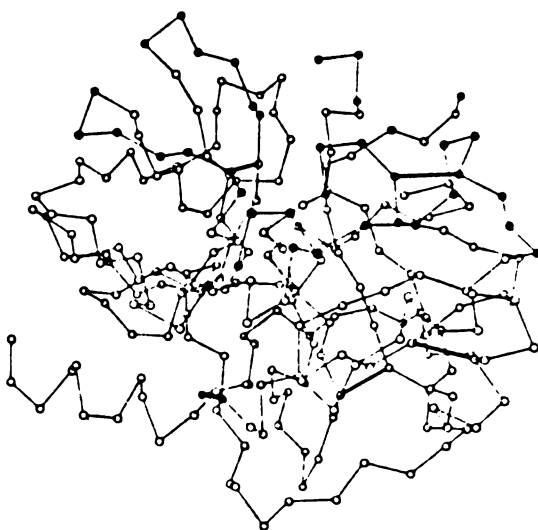
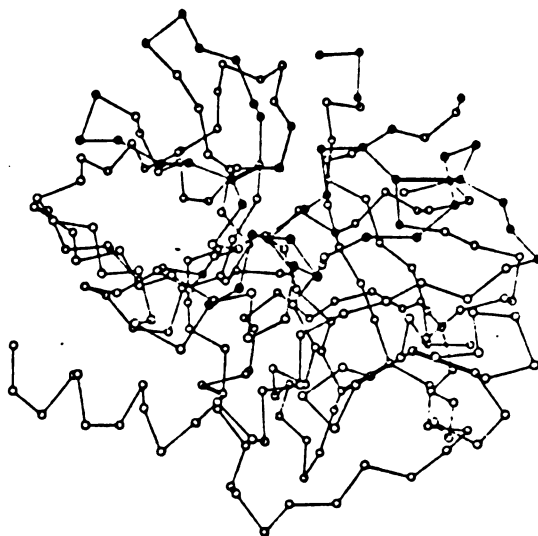
CHYMOTRYPSIN REFINEMENT RESULTS

A. Series I-a, I-b, II-a

To assist in the analysis of the refinements, the chymotrypsin protein was sub-divided into interior, interface and exterior regions. A complete list of the amino acid residues assigned to each region in CHT is given in Appendix B. A stereoview of the CA atoms of γ -CHT is displayed in Figure 4. The amino acid residues that constitute the dimer interface in the alpha form of CHT may easily be seen. In order to meaningfully compare the results of the final structures of γ -CHT obtained, a least squares procedure was used to rotate and translate one structure to another. The method was adopted from the REFIN system. In all cases, hydrogen atoms were removed from the coordinate lists before the rotation-translation was performed. In summary, in refinement I-a, the 150 crystallographically observed water molecules were subjected to energy minimization, but charged ionizable side chains were not employed. In refinement I-b, the effects of neglecting solvent molecules in an energetic refinement were investigated. The effects of neglecting

Figure 4. Stereoview of CA Atoms of the γ -CHT Monomer.
Dimer interface residues shaded.

(Top)



the charges on ionizable side chain residues were investigated in comparing refinements I-a and II-a. The effect of bulk solvent is shown in refinements III-a and III-b.

Tables 3 and 4 indicate the final energies and root mean square deviations respectively of the different series of refinements from the observed crystallographic structure. The greatest energetic improvement in every case is seen to arise from the non-bonded term, however, all terms show marked improvement, especially in the bond lengths and bond angles. In an energetic refinement, it is important to analyze the final geometry in terms of standard geometrical properties of amino acids. Table 5 shows that the refinements of γ -CHT narrowed the tau angle (N-CA-C) distribution around the ideal dictionary value of 110° . On the other hand, a degradation of the omega angle distribution is noted in Table 6. The worst of the post-refinement omega angles is seen to be near the carboxy terminal residues, especially at the end of the C chain. The Ramachandran plot for the I-a refined structure is shown in Figure 5. Very few of the residues show deviations from the allowed non-bonded contact zones.

Table 4 displays the r.m.s. deviations from the observed structure for the various classes of atoms shown in Appendix B. The r.m.s. movements of the main chain atoms and the CYS sulfur atoms are quite small compared with the overall r.m.s. deviations. The largest movements of the main chain

Table 3: Final Energies of the Refined γ -CHT Structures
(kcal/mole).^a

	<u>Observed</u>	<u>I-a</u>	<u>I-b</u>	<u>II-a</u>	<u>III-a</u>	<u>III-b</u>
Bond Length	901	15	19	110	1074	1276
Bond Angle	997	182	171	201	197	222
Dihedral	162	51	44	65	56	53
Non-Bonded	6398	-2837	-2264	-2763	-6759	-5977
Torsional	325	202	211	212	201	200
Electrostatic	-126	-163	-134	-294	-169	-162
Global	8658	-2490	-1953	-2498	-5600	-4388
Average Energy per Water	—	-2.79	—	-1.56	-1.44	-1.03

^aFor the series III refinements, results are for protein plus movable ice lattice waters.

Table 4: R.M.S. Deviations from the Observed γ -CHT Structure for the Energetic Refinements.

	Refinement Series				
	I-a	I-b	II-a	III-a	III-b
Main Chain	0.47	0.38	0.51	0.50	0.44
Side Chain	0.90	0.73	0.95	0.92	0.76
Carbonyl Oxygens	0.98	0.94	1.16	0.99	0.86
Sulfurs	0.60	0.47	0.65	0.56	0.50
Catalytic Site	0.76	0.74	1.01	0.78	0.70
TRP Cluster	0.46	0.32	0.49	0.47	0.31
Interior					
Main Chain	0.45	0.34	0.48	0.47	0.39
Side Chain	0.81	0.67	0.87	0.73	0.66
Exterior					
Main Chain	0.48	0.39	0.52	0.52	0.45
Side Chain	0.96	0.82	1.04	1.00	0.82
Interface					
Main Chain	0.48	0.43	0.53	0.54	0.50
Side Chain	1.05	0.96	1.05	1.14	0.91
Domain 1					
Main Chain	0.46	0.30	0.49	0.47	0.43
Side Chain	0.88	0.70	0.94	0.89	0.76
Domain 2					
Main Chain	0.49	0.44	0.54	0.54	0.45
Side Chain	0.91	0.75	0.95	0.95	0.75

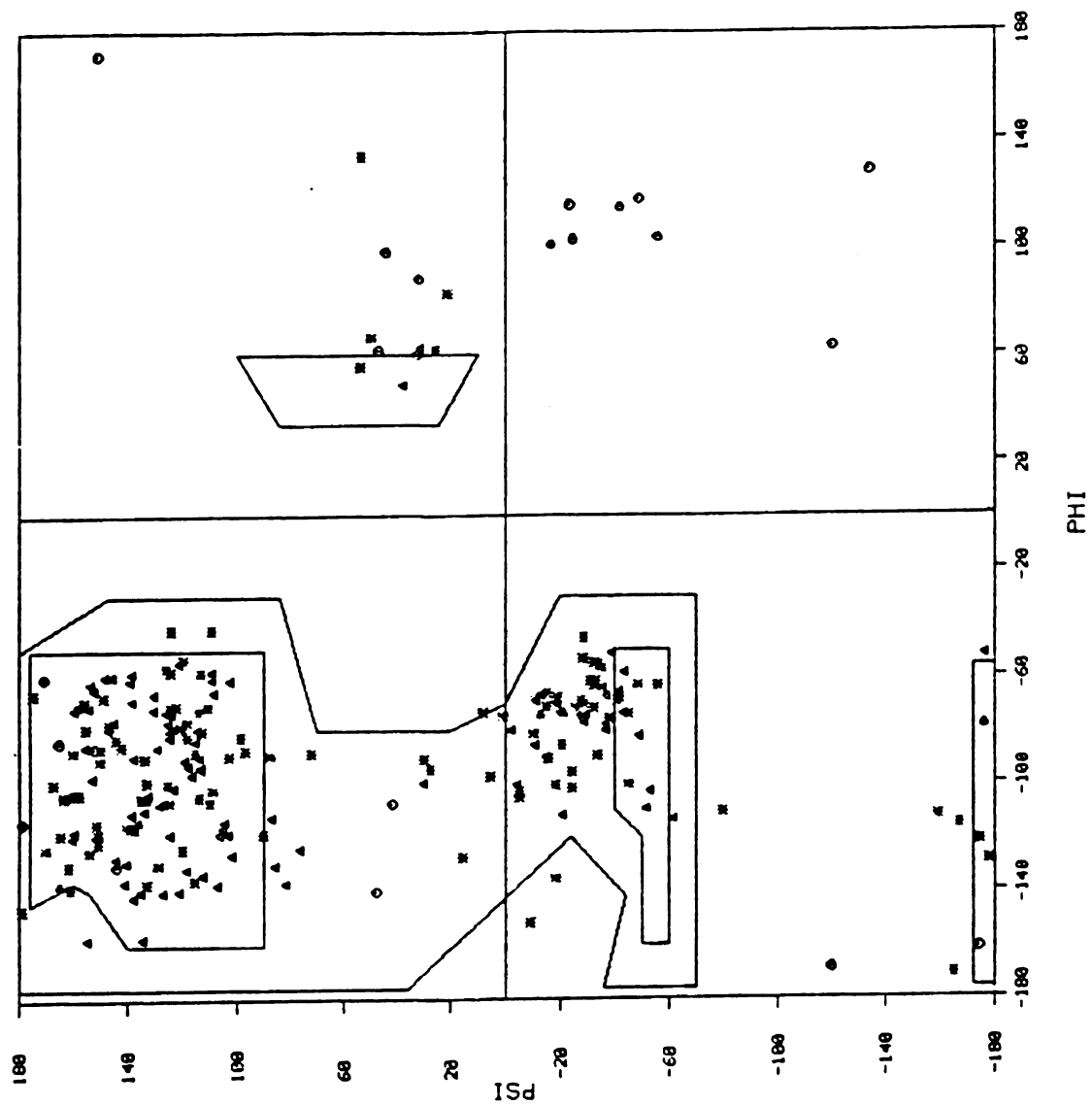
Table 5: Tau-Angle Distributions for the Energetic Refinements of γ -CHT.

<u>Region (deg)</u>	<u>I-a</u>	<u>I-b</u>	<u>II-a</u>	<u>III-a</u>	<u>III-b</u>	<u>Observed</u>
97.5-102.5						7
102.5-107.5	7	5	7	9	5	62
107.5-112.5	139	139	133	137	145	120
112.5-117.5	94	97	100	94	89	44
117.5-122.5	1		1	1	2	7
122.5-127.5						1
Average	111.9	111.9	112.0	109.7		

Table 6: Omega Angle Distributions (Absolute Value) for
the Energetic Refinements of γ -CHT.

<u>Region (deg)</u>	<u>I-a</u>	<u>I-b</u>	<u>II-a</u>	<u>III-a</u>	<u>III-b</u>	<u>Observed</u>
180-175	95	97	77	95	99	176
174-170	77	89	73	64	70	32
169-165	35	25	34	43	31	3
164-160	7	5	22	11	14	1
159-155	1	1	1	3	3	
154-150	1			1		
149-145			1			
144-140						

Figure 5. Ramachandran Plot for the Final Structure
Resulting from Refinement Series I-a.
Glycines represented as circles.



were found near the carboxy terminal residues, especially near ALA 149, which is also consistent with the omega angle distribution results.

The TRP cluster is a hemispherical cavity about 7 Å in diameter and 7 Å deep, bordered by TRP 27, PRO 28, TRP 29 and TRP 207, with PRO 4 and PRO 8 located approximately 4 Å above the opening. The TRP cluster found in the family of chymotrypsin enzymes is of interest because it has been suggested that it, along with the three other aromatic clusters found in chymotrypsin, lend stability to the folding of the molecule. Additionally, it can serve as a secondary binding site for aromatic substrate-like molecules¹⁴ and it is approximately symmetric with the active site across the protein center of mass. The center of mass of the TRP cluster residues is defined by a vector from the protein center of mass of length 11.0 Å. The active site center of mass (i.e., that of the catalytic triad) is defined by a similar vector of length 9.22 Å, the angle between the vectors is 172.5°. Quite spectacularly, the TRP cluster shows an overall r.m.s. deviation that is very small compared with the overall r.m.s. deviation. The magnitude of the PRO 4 and PRO 8 contribution to the r.m.s. movement of the TRP cluster may be reduced due to their smaller size. Nevertheless, the 0.46 Å r.m.s. change in the TRP cluster during refinement I-a is small relative to the 0.556 Å r.m.s. change for

the other six tryptophans and the 0.693 Å r.m.s. change of the other seven prolines in γ -CHT (not listed).

Table 4 indicates that the side chain atoms of the residues which constitute the dimer interface region of the α -CHT dimer exhibit an r.m.s. movement larger than the r.m.s. movement displayed by the side chain atoms of the exterior residues separately, or any other classes of atoms.

With the exception of the number of cycles performed, all other aspects of refinements I-a and I-b were similar. These refinements indicate the effects of omitting the crystallographically observed water molecules from energy minimization. Table 3 shows that the final energies are not changed, outside of the contributions arising from solvent-protein non-bonded interactions. Geometrical analysis shows equivalent distributions of tau and omega angles. However, the r.m.s. movements in refinement I-b are smaller than those reported for refinement I-a in Table 4, even in the interior of the protein. In the absence of solvent, the r.m.s. movement of the catalytic site residues is larger than the overall r.m.s. motion and is virtually the same size as in refinement I-a. Furthermore, in refinement I-b, the r.m.s. movements of the two domains of chymotrypsin are quite different, especially for the main chain atoms. This difference may be rationalized by the approximately 10% difference in the number of water molecules found in the two domains.

Once again, the TRP cluster and the interface atoms show small and large r.m.s. movements respectively.

The effects of including fractional charges on the side chain atoms of polar residues may be seen by comparing refinements I-a and II-a. Here, fractional charges are seen to reduce the electrostatic contribution but increase the non-bonded contribution to the total global energy of γ -CHT. The improvement in the electrostatic energy in refinement II-a suggests that charge localization such as that employed in refinement I-a may present difficulties in conformational energy minimization techniques. A consequence of this effect may be seen easier by examining the proposed salt bridge between ILE 16 and ASP 194. As was mentioned previously, half-electron charges were used in refinements I-a and I-b to simulate this proposed salt bridge. Initially, it was found that only one hydrogen bond existed between N of ILE 16 and OD1 of ASP 194. After refinement without side chain fractional charges, this hydrogen bond was lost. In refinement II-a, this initial hydrogen bond was not only preserved, but also improved, and an additional hydrogen bond was formed between ILE 16 and a solvent water molecule. The use of charged side chain atoms generally increased the r.m.s. movement during the energy refinements. Some deviations from this pattern were seen: e.g., the carboxy terminals were found to be positioned closer to the observed positions than in refinement I-a.

Another comparison of the r.m.s. deviations between the final energetically refined structures shows that the structures I-a and II-a are more similar to each other than either is to the observed crystallographic structure. These results can be seen in Table 7. Similar results are obtained for refinements I-a and I-b. However, when refinements I-b and II-a are compared with each other and with the observed structure, the refined structures are seen to differ more from each other than either does from the input structure. Some important exceptions include the catalytic site and the TRP cluster.

The least squares procedure used to fit the final γ -CHT structures to the observed crystallographic structure and with themselves may in itself be contributing an artificial effect. The translating and rotating algorithm employed treats all atoms (except hydrogens, which are not included) equally. Consequently, a sulfur atom is given the same weight as a carbon atom for example. The electron density will establish the position of the sulfur atom with a much greater accuracy than the carbon atom. Also, the thermal factors are not examined before the least squares fit. Atoms with larger thermal factors are treated exactly as those with smaller thermal parameters. In this way, a long side chain (probably possessing large thermal factors) such as LYS, will have the same weight as the side chain of an ALA, having only CB as it's side chain.

Table 7: R.M.S. Deviations Between Energetically
Refined γ -CHT Structures.

	<u>Ia-IIa</u>	<u>Ia-Ib</u>	<u>Ib-IIa</u>
Main Chain	0.40	0.33	0.41
Side Chain	0.67	0.83	0.99
Carbonyl Oxygens	0.38	0.83	0.72
Sulfurs	0.32	0.40	0.48
Catalytic Site	0.42	0.54	0.47
TRP Cluster	0.32	0.41	0.47
Interior			
Main Chain	0.38	0.34	0.41
Side Chain	0.67	0.64	0.74
Exterior			
Main Chain	0.41	0.33	0.42
Side Chain	0.73	0.73	0.81
Interface			
Main Chain	0.56	0.36	0.57
Side Chain	0.87	0.71	0.91
Domain 1			
Main Chain	0.36	0.32	0.39
Side Chain	0.62	0.64	0.69
Domain 2			
Main Chain	0.44	0.34	0.43
Side Chain	0.71	0.62	0.74

Additional constraints should be added to the least squares algorithm to include the effects of the two points mentioned above. Mass weighting and thermal factor cut-off criteria may be the answer. However, in the chymotrypsin energy refinements, the thermal factors of the final energetically refined structures were not available. If the above constraints are included, the results presented for the r.m.s. comparisons will likely show lower asymmetry for the main chain and greater asymmetry for the side chains, generally.

In both refinements I-a and II-a, no charges were placed on the solvent molecules. The average energy of a solvent molecule is -2.79 kcal/mole in refinement I-a and -1.56 in refinement II-a. The extended atom dictionary maxima for hydrogen bonds ranges from -2.5 to -3.5 kcal mole⁻¹. The average solvent molecule energetic contribution is consistent with approximately the energy of one hydrogen bond, while that of refinement II-a seems to be artificially small. The above results are also consistent with the difference in the mean solvent-protein closest approach distances of 3.01 Å in I-a and 3.15 Å in II-a. The observed structure showed a mean protein-solvent closest approach distance of 3.05 Å.

Overall, the r.m.s. magnitude of the force acting on the atoms in refinement I-b is 2.36 kcal/mole/Å. This value compares very well with the r.m.s. force reported for the structure of the bovine pancreatic trypsin

inhibitor (with 4 internal water molecules) used as input in a biomolecular dynamics simulation.^{18,28,29} In the presence of solvent, but not of charged side chains, an r.m.s. force of 6.19 kcal/mole/Å was calculated for refinement I-a. In refinement II-a, the r.m.s. force was 4.78 kcal/mole/Å. A simple rationalization of large r.m.s. movements during energy minimization being accompanied by small final r.m.s. forces fails; the effects of side chain charges and solvent must be taken into account.

The refinements performed in series III were identical to those in series I in that no side chain fractional charges were included in the refinements; however, bulk solvent was included. Also, the crystallographically observed waters were omitted. The final energies reported in Table 3 show similarity with the other series. Meaningful comparisons are made with refinement I-a. The striking feature is that the average energy per water molecule is drastically reduced in both of the series III refinements compared to that of series I. The bulk solvent present in the former has the effect of distributing the energy and forces throughout the solvent system much better than in I-a. This may be a consequence of the fact that interactions with internal protein atoms are minimal for the movable layer of water molecules in the bulk solvent model adopted and the larger number of movable waters in the series III refinements. The unit cell edge

of the ice lattice in each case is over 3.0 Å; water-water interactions do not contribute a significant amount to the global energy. The r.m.s. (Table 4) deviations of the series-III refinements clearly show that in employing a solvent model in protein energetic refinements, the final structures obtained possess a reasonable energy; at the same time, atomic positions do not deviate from the observed structure by an unreasonable amount.

B. Crystallographic Analysis

As was stated earlier, an important factor in determining the reliability or accuracy of the results of purely energetic refinements on protein molecules is the degree to which the procedure preserves agreement with the crystallographic observations. Some workers have found it advantageous to incorporate crystallographic restraints in their refinement programs.^{8,9} However, changes have also been made in crystallographic refinement routines to include potential energy terms.³⁰ It can be argued that including potential energy restraints into crystallographic refinement programs may destroy some of the information that results in high resolution refinements on protein molecules. A typical example of this effect can be seen in the refinement of the alpha form of chymotrypsin, which follows in part II. α -CHT crystallizes as a dimer at pH 3.5. A close examination of the final

structure of the dimer revealed four or five close contacts in the dimer interface region. The difference electron density maps indicated that within the accuracy of the method employed, the dimer interface had refined to the correct structure. No positive or negative peaks in the difference electron density maps were noted. The energy of the two monomers and that of the dimer of the final structure of α -CHT were calculated and about 10 to 15 kcal/mole of energy existed in these close contacts in the dimer interface region. If potential energy restraints had been included in this refinement, these close contacts would have been lost, and the difference electron density maps would probably have shown errors in this region. A more detailed examination of the dimer interface region in α -CHT will follow in part II.

For the purposes of this work, the reliability and accuracy of the final refined structures can be checked by calculating the crystallographic R-factors,³¹ defined by the following equation:

$$R = \frac{\sum (|F_o| - |F_c|)}{\sum (|F_o|)} \quad (6)$$

In Equation 6, $|F_o|$ and $|F_c|$ represent the amplitudes of the observed and calculated structure factors, respectively. The final structure of γ -CHT refined to an R-factor of .180, at a resolution of 7.0 to 1.90 Å. The results of these calculations are given in Table 8. In all cases, the

Table 8: R-Factors for the γ -CHT Energy Refinements.

<u>Structure</u>	<u>Resolution ($\text{\AA}$)</u>	<u>R-Factor</u>
γ -CHT(obs) R = .180	7.0	0.191
	3.0	0.174
	2.5	0.202
	2.0	0.191
	1.9	0.212
γ -CHT(obs) (no solvent) R = .231	7.0	0.284
	3.0	0.229
	2.5	0.252
	2.0	0.220
	1.9	0.235
I-a R = .317	7.0	0.297
	3.0	0.313
	2.5	0.354
	2.0	0.321
	1.9	0.347
I-b R = .345	7.0	0.287
	3.0	0.349
	2.5	0.377
	2.0	0.360
	1.9	0.359
II-a R = .352	7.0	0.297
	3.0	0.357
	2.5	0.381
	2.0	0.352
	1.9	0.363
III-a R = .345	7.0	0.320
	3.0	0.357
	2.5	0.362
	2.0	0.356
	1.9	0.364
III-b R = .332	7.0	0.311
	3.0	0.342
	2.5	0.358
	2.0	0.326
	1.9	0.351

energetically refined structures were translated and rotated to fit the structure of γ -CHT. This placed the energetically refined structure in the correct coordinate reference frame for the structure factor calculations, removing any drift that may have occurred in the coordinates without effecting the final energy. Since thermal factors and occupancies are not refined in potential energy minimization procedures, the B-factors and occupancies from the final crystallographic structure of γ -CHT were used for the protein and solvent atoms, respectively. Additionally, since the residues 10-13 and 149-150 were not seen in the observed electron density maps, these residues were not included in the structure factor calculations. Besides testing the validity of the refinement results obtained, this type of comparison will attest to the accuracy of the extended atom plus polar hydrogen method in general.

Table 8 displays the R-factors of the final, energetically refined γ -CHT protein structures, along with that of the observed structure. Since some of the energetically refined structures had no solvent included, a structure factor calculation omitting the solvent in the observed γ -CHT protein was also performed. The R-factor versus scattering angle is also shown in Table 8 at 7.0, 3.0, 2.5, 2.0 and 1.9 Å.

Interestingly, the solvent makes a strong contribution to the diffraction in the observed γ -CHT structure, about

5.0%. As expected, the greatest contribution of the solvent occurs in the low-angle data, where at 7.0 Å resolution, the contribution is over 9.0%. Examination of the energetically refined γ -CHT structures reveals that generally, there was an increase in the R-factor of about 10-15%. The largest increase in R-factor is found in the high angle data in all of the final structures. The structure refined with no solvent and no side chain fractional charges showed the best agreement with the observed crystallographic structure of γ -CHT. The R-factor from 7.0-1.90 Å increased about 13%, the agreement in the low angle data being about 10%. Examination of the series III refinements shows that the use of a more realistic face centered cubic ice lattice results in a final protein structure that agrees with the observed structure by 1.3% over using just a simple cubic ice lattice.

The results of the above structure factor calculations indicate the advantages and disadvantages of the use of the extended atom plus polar hydrogen method. However, it is known that the use of all atoms (including hydrogens) in an energetic refinement results in final structures that agree with the crystallographic structure to within 5-6%.¹⁹ Also, if a refinement is carried out with only extended atoms and no polar hydrogens, the increase in R-factor is well above 15%.³² The use of the extended atom method then seems to be a compromise between computational

speed and crystallographic accuracy. Better potentials may help to produce agreement with the observed structure but it should be realized that the coordinate uncertainty in most crystallographic studies approaches 0.2-0.3 Å, and may be even more for atoms with larger thermal parameters. In many cases, the r.m.s. movements of certain groups of atoms are very similar to the uncertainty in the coordinates. This fact shows that the R-factor can be drastically affected by slight movements in atomic positions. Better crystallographic agreement for the energetically refined γ -chymotrypsin structures could be obtained if two or three cycles of least squares refinement were performed, refining only thermal factors and occupancies of the solvent.

CHAPTER V

PROTEIN-PROTEIN ASSOCIATION AND MECHANICAL SURFACE TENSION

Electrostatic, hydrogen bonding and van der Waals interactions have all been shown to be important in the folding of a polypeptide chain into a three dimensional protein structure. The biological importance of protein-protein association, including dimer- and oligomerization, is widely recognized. Recently, hydrophobicity has been suggested to be a major force in the stabilization of protein-protein association.³³ Hydrophobicity can be assessed using the concept of accessible surface area.³¹ For a protein atom, this is the area of the surface over which the center of a water molecule can be placed while it is in contact with the atom and not penetrating any other protein atom. Each square Angstrom of surface area buried upon association gives a hydrophobic free energy of about 25 cal/mole.³³ Additional mechanisms have been proposed as rationalizations and origins of the free energy of association in protein-protein association. Architectural complementarity or "lock and key" descriptions have been proposed.³⁵ Kauzmann has suggested that hydrophobic energies arising from surface patches of non-polar side

chains may play a role.³⁶ Chothia and Janin^{33,37} have also stressed the effects upon solvent entropy (and hence upon the free energy of association) of excluding protein surface area from interaction with the solvent. Since it is desirable to be able to predict patterns of association when component structures are known but the aggregates are not, tests and improvements of the above mentioned association models are needed.

The conformational energy minimizations of γ -CHT in the presence of solvent discussed above may be of use in predicting protein dimerization and complex formation. During the analysis of the γ -CHT refinements, it was found that the interface region of the γ -CHT monomer displayed large movements. This may be seen in Table 4 (p. 33) where it is shown that the r.m.s. movements of the sets of all, main chain and side chain atoms in the interface region were uniformly larger than the corresponding supersets in the protein exterior. This observation motivated the calculation of r.m.s. forces, which led naturally to the suggestion that this region may possess a large local surface tension. Upon association, this region would then be expected to be internalized.

The REFIN system enables the manipulation of the final microscopic forces on the structure numerically to obtain a mechanical, non-thermodynamic surface tension, derived solely from a protein-solvent potential. This

work is conceptually related to that reported by Lee,³⁸ however, there, a thermodynamic surface tension was calculated. This type of investigation may prove to be a natural complement of the "excluded solvent-accessible area" model of Chothia and Janin.³³

A. Mechanical Surface Tension Calculations

In this section, the methods employed to approximate the effective molecular surface tension will be described. In the first method, the chymotrypsin center of mass was determined for use as the origin of a spherical polar coordinate system. Each atom's contribution, $\epsilon_i(\tilde{r}, \tilde{\theta}, \tilde{\phi})$ to the system energy was calculated. After all radial coordinates were multiplied by 1.001, the atomic energy contributions were recalculated. The microscopic surface tension was approximated with the numerical difference

$$\gamma^{\text{mech}} \cong (\Delta A)^{-1} \sum_i [\epsilon_i(1.001 \tilde{r}, \tilde{\theta}, \tilde{\phi}) - \epsilon_i(\tilde{r}, \tilde{\theta}, \tilde{\phi})] \quad (7)$$

where the summation was taken over the set of all exterior residue atoms to find $\gamma_{\text{ext}}^{\text{mech}}$ and the set of all interface residue atoms to find $\gamma_{\text{inter}}^{\text{mech}}$. In Equation 6, ΔA represents the difference in van der Waals surface areas associated with the summation set before and after the radial coordinate scaling. The surface area calculations were performed using the atomic van der Waals radii

$r(O) = 1.5 \text{ \AA}$, $r(N) = 1.6 \text{ \AA}$, $r(C) = 1.8 \text{ \AA}$, and $r(S) = 1.9 \text{ \AA}$. No hydrogen atoms were included in the calculations. All results were obtained using Connolly's implementation³⁹ of Lee and Richards' molecular surface area algorithm.³⁴

As a check on the effective surface tension calculations, a second method has been devised, in order to give a greater statistical sample of atomic energies and remove the problem of incommensurate changes, such as those of aromatic rings located parallel and perpendicular to the surface of the protein. Implementation of method two has recently been started. In method two, several different structures will be generated for the final γ -CHT structure including the 150 crystallographically observed waters.

The generation of the protein structures used in method 2 is as follows. The harmonic potential force constants for each atom are calculated. This need be done only once for the model. Once the force constants are calculated, trial structures are generated using the following method. An energetic contribution is assigned to each atom according to a Boltzmann distribution. A reasonable energetic cut-off is chosen at 5.0 kcal/mole. The energy is distributed over x, y, z randomly, subject to the constraint that $E_x + E_y + E_z = E_{\text{tot}}$. Using the force constants determined previously, a shift is calculated for each atom. For each component, a random direction is chosen on the classical phase space ellipse and the shift

computed as $\text{shift} = \sqrt{2E/K} \cos(\theta)$. θ is the random phase angle for each (x,y,z) direction. The shifts are applied and the energy of the resultant system is calculated. The energetic components (exterior and interface) and the molecular surface area are used to obtain the local macroscopic surface tension. These results are then averaged for the trial structures generated to find the final value of the local molecular surface tension for each region (exterior and interface).

B. Surface Tension Results

For each of the three refinements (I-b, III-a, III-b), the total protein van der Waals surface area, the ratios of the exterior to interface surface areas and regional energies and the exterior and interface γ^{mech} values are collected in Table 9. The total surface area values reveal that the presence of solvent in energy minimizations II and III tends to reduce the accessible surface area relative to the isolated enzyme calculation I. Lee³⁸ suggests that the average potential energy of a molecule can be represented as a linear function of its surface area. Comparison of the ratios of the exterior to interface surface areas and energies in the protein-solvent refinements II and II is consistent with this assumption.

In the presence of solvent, the $\gamma_{\text{inter}}^{\text{mech}}$ values are twice the size of the corresponding values for the molecular

Table 9: Surface Area and Mechanical Surface Tension Results.

	Refinement Series		
	I-a	III-a	III-b
Total Surface Area ($\text{\AA}^2$)	23226.2	22863.9	22922.1
Area(ext)/Area(inter)	4.540	4.580	4.560
Energy(ext)/Energy(inter)	4.410	4.720	4.320
$\gamma_{\text{ext}}^{\text{mech}}$ (kcal/mole- $\text{\AA}^2$)	0.145	0.351	0.131
$\gamma_{\text{inter}}^{\text{mech}}$ (kcal/mole- $\text{\AA}^2$)	0.063	0.838	0.250

exterior including the interface. Although the percentage changes in the exterior and interface regional surface areas were similar upon radial scaling, the percentage change in the interface energy is twice that of the exterior regional energy. The inconsistent results obtained for I-a are explicable on the grounds that the mechanical surface tensions represented protein-vacuum interfaces; the central role of a real solvent is artificially absent.

The values of the mechanical surface tension calculated for chymotrypsin in the preceding section differ in two important aspects from results reported by Lee³⁸ for a variety of protein molecules. The values above were derived from a mechanical potential function describing the protein-solvent system and hence do not reflect the contribution of entropy to a thermodynamic surface tension based upon free energy. In addition, the results are intended to exploit the microscopic detail of conformational energy minimizations by representing, at least to some level of approximation, variations in the mechanical surface tension between certain regions of the protein exterior instead of representing a single, global value of the surface tension. The first distinction largely precludes the comparison of the mechanical surface tensions in Table 9 with those obtained by Lee,³⁸ which are approximately an order of magnitude smaller (~ 35 cal/mole/Å).

The prospect of identifying possible protein-protein association sites on the basis of local variations in

mechanical surface tension is enticing, particularly since the conformational energy minimizations which would be utilized in such attempts are an increasingly routine part of protein structure refinement. However, few such conformational energy minimizations reported to date have attempted to model the interaction of proteins with bulk solvent. It is possible that the high mechanical surface tension found for the interface region of chymotrypsin is fortuitous, and examination of other associating molecules with well-characterized structures is clearly necessary; among the possible targets for study are hemoglobins, insulin, and the trypsin-trypsin inhibitor complex. Besides requiring analysis of several associating molecules, the validation of mechanical surface tension as a guide to possible interface sites will require study of several model-specific factors.

It has been argued that solvent entropy gains upon protein surface area reduction drive the association process,^{33,40-42} yet the "mobile solvation layer inside an ice lattice" model used in this work entails a well-ordered bulk solvent. Ideally, mechanical surface tension calculations would be attempted with several solvent models. It could prove that mechanical surface tension is tangential to the important aspects of protein aggregation, particularly since the application of macroscopic concepts of surface chemistry to single molecules is recognizably difficult.⁴³ Yet it is worth noting that

calculations utilizing mechanical potentials include contributions from van der Waals interactions and hydrogen bonds. Ross and Subramanian⁴⁴ have criticized the absence of such contributions from excluded volume theories^{33,37} and the possibility of obtaining complementary information from detailed conformational energy calculations is thus attractive.

Approximations which may influence mechanical surface tension calculations include the use of the modified "extended atom" representation of protein hydrogens, the particular potential dictionary used for interacting non-hydrogenic atoms, the representation of charged ionizable side chains employed, and the $D = r$ representation of the dielectric. The effects of the above approximations need further investigation. The generalizability to other associating proteins of the high local mechanical surface tension found in the dimer interface region of the isolated chymotrypsin monomer needs further study also.

PART II: THE REFINEMENT AND STRUCTURE OF α -CHYMOTRYPSIN
AT 1.67 Å RESOLUTION.

CHAPTER VI

INTRODUCTION

A. Refinement Methods Based on X-ray Data

In any structure analysis based on x-ray diffraction data, two main components exist. The first is to deduce a model or a set of phases which correspond to most, if not all of the atomic positions in the molecule. In protein structural studies, the phases are usually determined from several heavy atom derivatives whose crystals are isomorphous with those of the native protein⁴⁵; density modification procedures may then be employed as a method to extend the resolution without bias of a model.^{13,46} Second, the initial model can be adjusted so that the calculated structure factor amplitudes match the observed values as closely as possible. This process is termed refinement.

Early refinements of protein structures were performed almost exclusively using either the real space method⁴⁷ or difference Fourier methods followed by a conventional block diagonal least squares procedure.⁴⁸ Watenpaugh used the second procedure in refining the structure of rebredoxin. The results of this work indicated that much more structural information can be obtained, in particular, solvent

structure, than by using the least squares method or the real space method alone. With the latter, the model is not refined in the usual crystallographic sense since the model is fit to the electron density based on phases which do not change during the refinement. Deisenhofer and Steigemann have used a combination of real space and difference Fourier methods with a great deal of success.⁴⁹

In the unconstrained least squares refinement of atomic parameters, the function minimized takes the form:

$$P = \sum_{hkl} W(hkl) \left[|F_o(hkl)| - |F_c(hkl)| \right]^2 \quad (8)$$

where $W = 1/\sigma^2(hkl)$ is the weighting function. The atomic parameters are corrected using the matrix equation:

$$\Delta U = -H^{-1}G \quad (9)$$

Here H is the normal matrix and G is the gradient vector.^{50,51} One of the problems with this type of procedure is the immensity of the computational problem. The size of the normal matrix is $M \times M$, where M is the number of parameters. The length of the gradient vector is also M. Agarwal⁵² has developed a much faster least squares procedure for refining atomic parameters. His method is based on the fast Fourier transform method (FFT). For very large structures, the amount of computation is proportional to the size of the structure, making it

extremely attractive for the refinement of biological macromolecules.

In recent years, new algorithms have been developed for the crystallographic refinement of biological molecules. Hoard and Nordman⁵³ as well as Sussman et al.,⁵⁴ have introduced the concept of group constraints into reciprocal space refinements. The basic concept here is that certain peptide fragments, for example side chains, may possess geometries which are well established and should be preserved. In the case of full matrix refinements, the reduction in the number of parameters can substantially reduce computing time while at the same time provide accurate refinement results.^{50,53}

Hoard and Nordman have applied this rigid-group restraint method in developing a crystallographic refinement program based on the Gauss-Seidel least squares procedure. Here, the normal equations for each structural unit (rigid group) are solved and the new estimates for the group parameters are used to update the calculated structure factors. The procedure is basically block diagonal and considerable computation time is saved by calculating the contributions from one atom to all reflections at a time.⁵⁵ Some problems are associated with this method, one being that it is difficult to simultaneously impose restraints on chirality at asymmetric centers while at the same time restraining the planarity of certain groups of atoms.⁵⁶

The approach that is most commonly employed today in the refinement of biological macromolecules is the least squares refinement of structure factors coupled with simultaneous optimization of the stereochemistry. Two approaches to this problem have been developed. First, Hendrickson and Konnert⁵⁷ introduce the stereochemical data as additional observations in the least squares refinement. The second method, discussed in part I, was developed by Jack and Levitt⁹ where a potential energy function is included. Since the energy function used by Jack and Levitt is quadratic, the two methods are essentially the same.

B. PROLSQ — Restrained Least Squares Refinement

There are two major obstacles that need to be overcome in the refinement of large macromolecules. The first has already been mentioned, being the large computing time involved, even with block diagonal least squares programs. The second is the limited amount of diffraction data available from large molecules such as proteins. It is very rare to find a protein crystal that will scatter x-rays beyond the 2.0 Å limit. Generally, the diffraction data is reduced by the sheer size of the protein molecule and disorder associated with it.⁵⁰ In any least squares procedure, the reliability of the results is decreased as the ratio of the number of observations to the number

of parameters is reduced. Two ways of overcoming the obstacle of limited data is to either decrease the number of parameters or add additional observations. One method of reducing the number of parameters is to use rigid group constraints.^{53,54} Alternatively, the number of observations may be effectively increased by including information in the form of constraints or restraints on the known geometry of the molecule. These might include information about bond lengths, bond angles and torsion angles. From crystal structure analyses of amino acids, spectroscopic and chemical analyses, and theoretical studies, a great deal of information has been gathered concerning the geometry and stereochemistry of the components of proteins and nucleic acids. The program PROLSQ was developed with this information in mind. PROLSQ, a least squares, reciprocal space refinement program, employs restraints on the known geometry of proteins to "increase" the number of observations or more exactly, to effectively reduce the number of free variables. It is important to note that unlike constraints, restraints restrict the features of the model to a range of realistic values.

PROLSQ is a least squares procedure, where the best set of final parameters minimizes the weighted sum of the squared residuals. In PROLSQ, the weights chosen are always inversely proportional to the variances. There are many classes of "observations" employed in PROLSQ. Each

class is treated separately in the sum, the total function for minimization being the sum of all observational classes. Some of the various classes of "observations" that are treated in PROLSQ are outlined below.

1. Structure Factors

The observational function for structure factors takes the form:

$$\phi = \frac{\text{reflections}}{\Sigma} \frac{1}{\sigma_F^2} \left(|F_O| - |F_C| \right)^2 \quad (10)$$

The calculated structure factors are determined from the equation:

$$F_C = K \sum_j f_j(hkl) \exp(-B_j S_{(hkl)}^2) \exp[2\pi i(hx_j + ky_j + lz_j)] \quad (11)$$

where (h,k,l) are the reflection indices, K is a scale factor, f_j is the atomic scattering factor, B is the isotropic temperature factor, x,y,z are the atomic coordinates in fractions of the unit cell, S is $\sin \theta / \lambda$ and the summation is over all atoms in the asymmetric unit. It is also possible to include variable occupancy factors and the inclusion of six anisotropic temperature factors.

2. Bond Distances

Interatomic distances are restrained using the following "observational" function:

$$\phi = \sum_j^{\text{distances}} \frac{1}{\sigma_D^2(j)} \left(r_j^{\text{ideal}} - r_j^{\text{model}} \right)^2 \quad (12)$$

r being the distance between the atoms. By also restricting next nearest neighbor and 1-4 distances, bond angles and dihedral angles may be restrained.

3. Planar Groups

Certain groups of atoms are restricted within the least squares plane of the group of atoms. The "observational" equation takes the form:

$$\phi = \sum_k^{\text{planes}} \sum_i^{\text{coplanar atoms}} \frac{1}{\sigma_P^2(i,k)} \left(\bar{m}_k \cdot \bar{r}_{i,k} - d_k \right)^2 \quad (13)$$

where $\bar{m}_k$ and $\bar{d}_k$ are the parameters defining the least squares plane.

4. Chiral Centers

One of the best features of the PROLSQ program is its ability to restrain the stereochemistry at asymmetric centers, using the chiral volume as the "observational" equation, which takes the form:

$$\phi = \sum_{\ell}^{\text{chiral centers}} \frac{1}{\sigma_c^2(\ell)} \left(v_{\ell}^{\text{ideal}} - v_{\ell}^{\text{model}} \right)^2 \quad (14)$$

5. Non-bonded Contacts

Instead of employing a potential energy function in the least squares procedure, PROLSQ uses only the repulsive part of the standard Lennard-Jones potential in the "observational" function:

$$\phi = \sum_m^{\text{non-bonded contacts}} \frac{1}{\sigma_{(m)}^4} \left(d_m^{\text{min}} - d_m^{\text{model}} \right)^4 \quad (15)$$

The summation is taken only over repulsive contacts, i.e. $d^{\text{model}} < d^{\text{min}}$, the value of d^{min} depending on the type of contact being considered.

6. Torsion Angles

Flexible conformational torsion angles such as the Ramachandran angles, and rotations in peptide side chains are restrained with the "observational" function:

$$\phi = \sum_t^{\text{torsion angles}} \frac{1}{\sigma_T^2(t)} \left(\chi_t^{\text{ideal}} - \chi_t^{\text{model}} \right)^2 \quad (16)$$

PROSLQ contains additional "observational" equations that may be used in the refinement of a protein structure.

These include positional and thermal restraints, non-crystallographic symmetry, damping of excessive shifts and occupancy factors. These will not be discussed in detail here; the "observational" functions take on the usual least squares form and the pertinent equations may be found elsewhere.⁵⁷ Many of these additional features were not used in the refinement to be discussed in the following sections.

C. Graphics Intervention and FRODO

During the least squares refinement of a biological macromolecule, a point is reached when further refinement cycles will not produce any meaningful changes in the structure or in reduction of the R-factor. This is especially true in a restrained least squares refinement procedure. The restraints applied to the stereochemistry prohibit the refinement from moving atoms large distances, even when the diffraction data demands it. Also, it is very unlikely that least squares will move atoms meaningfully more than 1.0-1.5 Å. It becomes necessary to somehow examine and change the current structure. Over the last decade, several different molecular graphics programs have been designed to accomplish this on a wide variety of hardware systems and almost every crystallography laboratory had their own system. By testing and actually using these systems in refinements, one program has emerged as having

the greatest flexibility, ease of use and most of all, transportability. This program is called FRODO.

FRODO was originally written for use on a Vector General graphics system, with a PDP-11/40 as the host computer.⁵⁸ Several additional versions have now appeared^{58,59} and the system can be operated on VAX computers using Vector General or Evans and Sutherland graphics systems. Since graphics interventions must be performed during refinements, some of the features that are routinely used will be described. A complete description of the program may be found elsewhere.⁵⁸

Jim Pflugrath and Mark Saper, working with F. Quiocho at Rice University were kind enough to send a version of FRODO to Michigan State University that was updated for a VAX 11-750 computer using an Evans and Sutherland PS300 graphics system. All of the features available in the original version of FRODO were implemented, as well as some additional features that were added due to the graphics system hardware available. The greatest advantage of the PS300 system over others is the fact that all calculations may be performed on the PS300 instead of using the host computer. This means that any additional users that may be on the host system will not be aware of the PS300. This is not the case for other graphics configurations. For example, the Picture System 200 from Evans and Sutherland requires a dedicated host computer for efficient operation and even then, execution time is very

slow. Also, the PS300 package comes with a graphics tablet and control dials as well as an extremely versatile keyboard. This enables the user to quickly form and break bonds, move entire fragments of molecules, examine different types of electron density maps simultaneously, and efficiently rotate around bonds changing torsion angles. The two most important options available are the ability to rotate about certain bonds and the ability to examine different electron density maps. In most cases, the stereochemistry is very good when a graphics intervention is needed, so the need to manipulate bonds and move fragments is not necessary. What is desired is the ability to move certain atoms into the electron density. This is accomplished very quickly by the application of torsion angle rotations.

A diagram showing the menu of commands available on the PS300 screen is shown in Figure 6(a-c). Each menu is chosen by moving the pen connected to the graphics tablet to the location of the menu item. By pressing down, this option is selected. At any time, control of execution may be turned over to the host computer by selecting the CHAT option. There are additional commands on the host computer that may be performed, including such options as adding and deleting atoms, refinement, which only includes geometrical idealization, listing the current coordinates, etc. Once a particular region is chosen for examination, the user types GO, and control

Figure 6. Examples of FRODO Graphical Displays. (a) stick diagram, (b) stick diagram plus electron density, (c) van der Waals surface.

Figure 6a

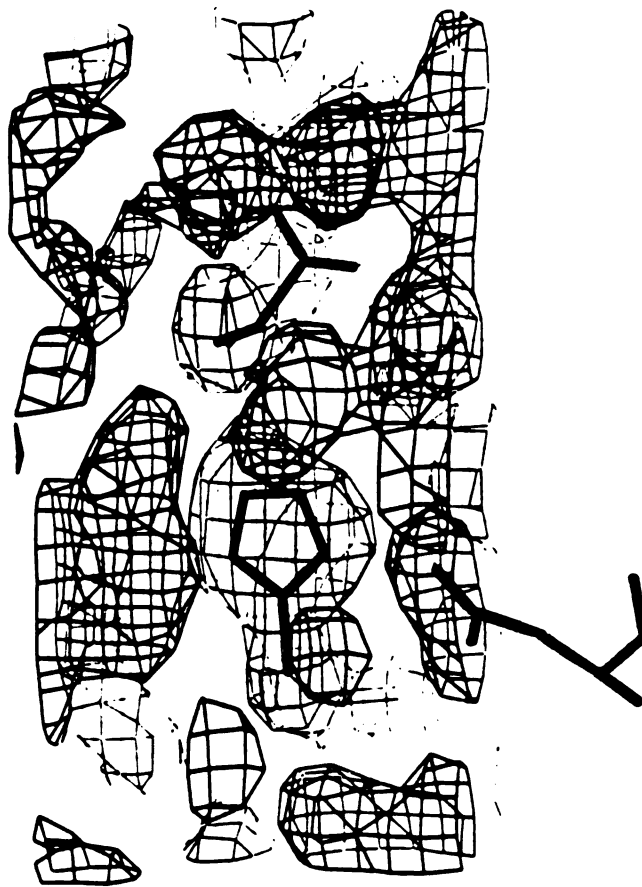


Figure 6b

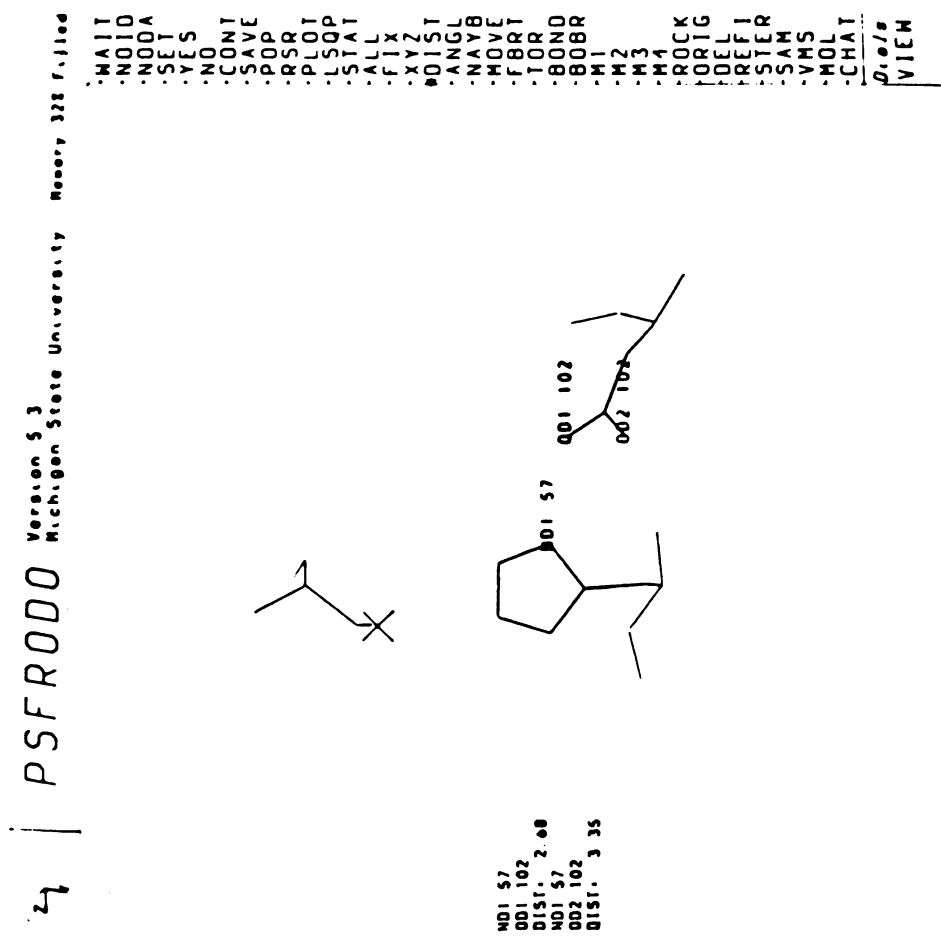
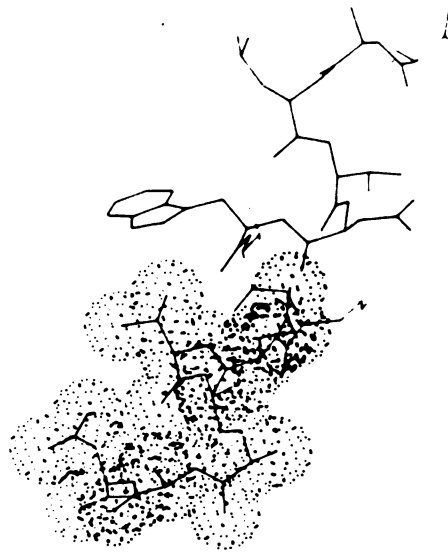


Figure 6c



PSFR000 Version 5.3
Michigan State University May 2002 Filed

MAIT
NOID
NODA
SET
YES
NO
CONT
SAVE
POP
RSR
PLOT
PLSQ
STAT
ALL
FIX
XYZ
DIST
ANGL
NAYB
MOVE
FBRT
TOR
BOND
MI
M2
M3
M4
ROCK
ORIG
DEL
REFI
STER
SAM
VMS
MOL
CHAT
D.o/a
VIEM

is then returned to the PS300. Figures 6(a-c) give some examples of the types of displays that may be obtained with FRODO. Figure 6-a shows a normal stick diagram of a peptide fragment, Figure 6-b presents a peptide fragment with the electron density superimposed upon the atoms and Figure 6-c indicates how a van der Waals surface may be displayed by FRODO.

It should be noted that FRODO as currently implemented, operates in full color, the user being able to change the color of any object on the screen simply by turning the control dials. This feature makes it extremely easy to see atoms in the electron density and if needed, change atomic positions. Many additional options are currently being added to the system, both at Rice University, here, and at other labs. These include real time monitoring of distances and angles, real space refinement of atomic fragments and hardcopy plotting options.

CHAPTER VII

REFINEMENT OF THE α -CHYMOTRYPSIN DIMER

A. Experimental

A short summary of some experimental aspects of the α -CHT dimer will be given at this point. A complete description can be found elsewhere.¹²

Crystals of α -CHT were grown from about 50% saturated ammonium sulfate solutions at pH 4.2. The crystals were stored in 70% saturated ammonium sulfate at pH 3.5 so that this study corresponds to the pH 3.5 conformer of α -CHT.⁶⁰ Crystals of the pH 3.5 conformer of α -CHT are monoclinic, space group P2₁ (B unique axis), $a = 49.29 \text{ \AA}$, $b = 67.48 \text{ \AA}$, $c = 65.94 \text{ \AA}$, $\beta = 102.02^\circ$. Intensity data were measured from one crystal specimen using a Nicolet P3/F diffractometer at 250 W power (5 mA) with a resultant intensity loss of about 13%. Total exposure time was 395 hours. The data collection proceeded in order of decreasing Bragg angle so that decay corrections were generally much less than 13%. A total of 27,534 reflections were observed, about 54% of the total possible.

B. Refinement Summary

Throughout the early work on the structure and function of α -CHT, a great deal of evidence was accumulated indicating that the structure of the dimer of α -CHT is asymmetric. Asymmetric binding of substrates and inhibitors as well as changes in electron and difference density maps upon changes in pH have been shown. Birktoft and Blow⁶¹ have reported the structure of α -CHT at 2.0 Å resolution. However, in their work, the electron density of the two molecules of the dimer of α -CHT was averaged around a local 2-fold axis, and a model was built to fit this average electron density. Rotation matrices and translation vectors were reported that enabled the construction of a dimer, but the structure obtained in this manner was symmetric. At the beginning of the present work, it was decided that the best rationale for a refinement of an asymmetric dimer of α -CHT was to begin the least squares refinement with a symmetrical structure and let the refinement search out the asymmetry. A natural starting structure was with the above coordinates of Birktoft and Blow. Thus, a symmetrical dimeric structure was created and an initial structure factor calculation performed. The R-factor³¹ of this symmetrical dimer was 0.37 Å using 5.0-3.0 Å data with only two different thermal factor values, 8.0 Å² for the main chain and 11.0 Å² for the side chains.

The least squares refinement program PROLSQ⁵⁷ was used for refining the structure of this symmetrical dimer. The refinement was performed in four distinct resolution stages, 5.0-3.0 Å, 5.0-2.5 Å, 5.0-2.0 Å, and 5.0-1.67 Å. A total of three interactive graphics interventions using FRODO were also performed. The course and progress of the R-factor for the refinement is shown in Figure 7.

At 5.0-3.0 Å resolution, the R-factor decreased from 0.37 to 0.227. Individual standard deviations for the structure factors based on intensity statistics were used during this refinement. An r.m.s. asymmetry of 0.28 Å and 0.69 Å for the main and side chain atoms developed during this stage. The first of three interactive graphics interventions was also performed using $(2|F_O| - |F_C|)$ and $(|F_O| - |F_C|)$ electron density maps. The r.m.s. asymmetry increased slightly to 0.30 Å and 0.91 Å for the main and side chain atoms due to the FRODO intervention (Figure 8). In extending the resolution to 2.5 Å, the R-factor increased to 0.320, but after 23 cycles of least squares refinement, the R-factor decreased to 0.203 (Figure 7). Up to this point in the refinement, an overall thermal factor had been refined. During the last three cycles at this resolution, restrained isotropic thermal factors were introduced. It may be seen in Figure 8 that the r.m.s. asymmetry remained essentially the same during the 2.5 Å refinement. However, there was a large decrease in the R-factor in the low-order data as higher order

Figure 7. Progress of R-Factor During Refinement. The resolution stages and FRODO interventions are indicated.

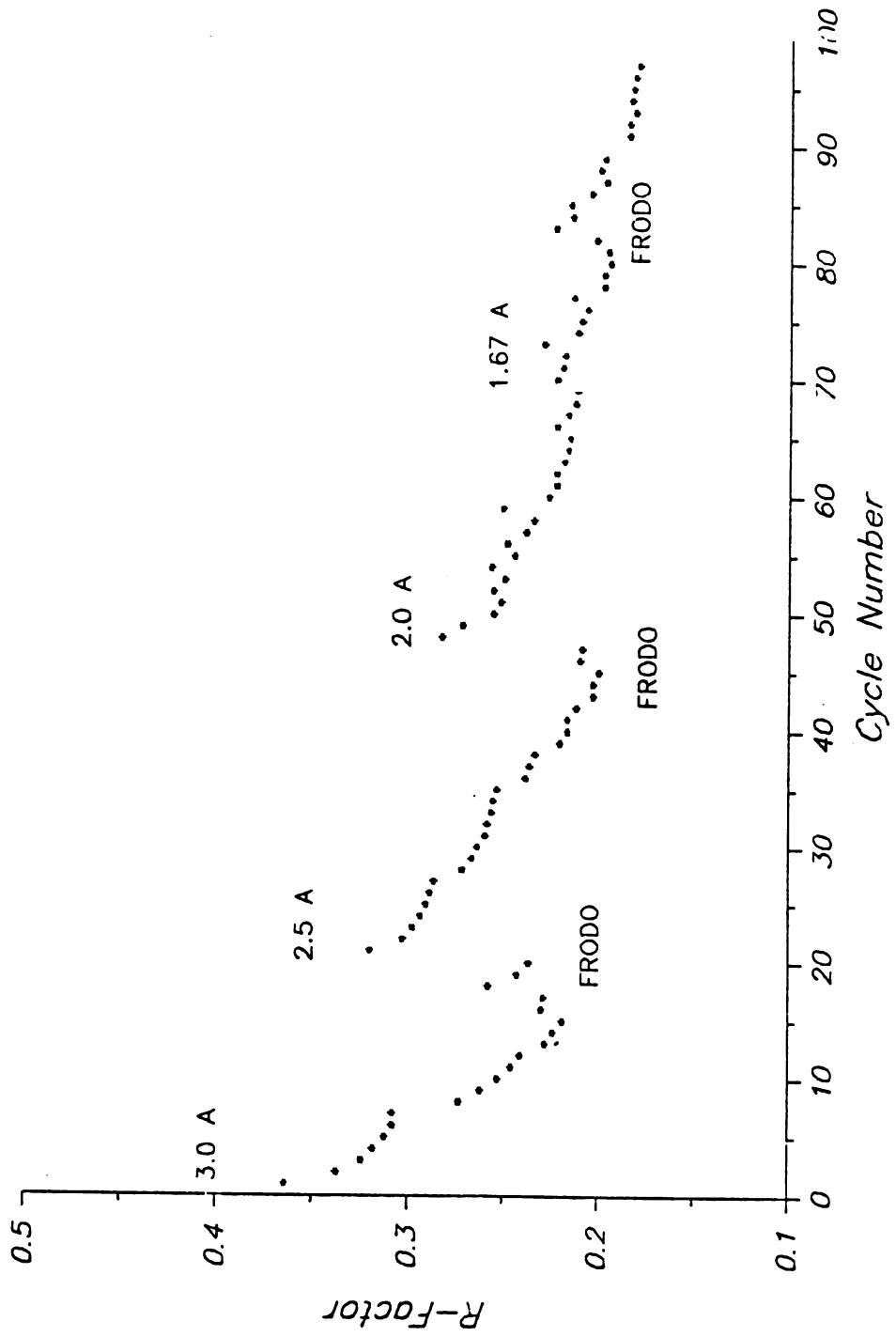
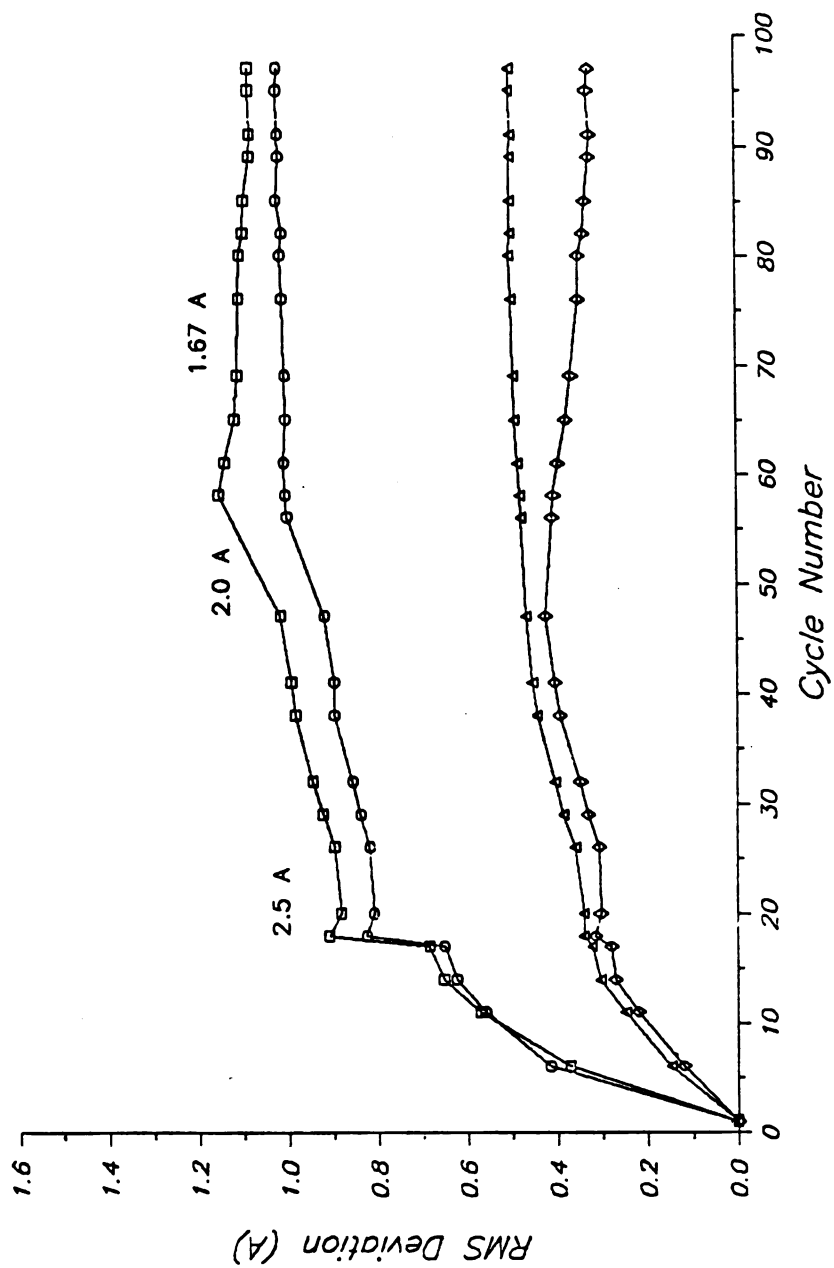


Figure 8. Progress of Asymmetry Development and Shifts During Refinement. Diamonds and squares, main and side chain asymmetry; triangles and circles, main and side chain shifts with respect to the trial structure, respectively.



reflections were included in the refinement. At the end of the 2.5 Å refinement, the R-factor had decreased to 0.186.

The second FRODO interactive graphics intervention was performed at this stage and the resolution extended to 2.0 Å and finally to 1.67 Å. During these latter stages, solvent structure was introduced as water molecules. This was accomplished by the careful examination of difference electron density maps, calculated with 8.0-2.0 Å and 5.0-2.0 Å data and their 1.67 Å resolution counterparts. Positive peaks greater than $3 \times \text{r.m.s. deviation}$ in both difference maps and within 1.0 Å of each other were included as solvent. Before any comparison of peaks was performed, all possible water positions were placed as close as possible to the protein by applying P21 symmetry operations to the coordinates. Also, solvent positions that did not appear to hydrogen bond to the protein or were too close to a protein atom were discarded. During the entire refinement, the positions of the solvent molecules were monitored and if at any time, the above conditions were not satisfied, the corresponding solvent was removed from the calculation. The coordinates and the occupancies of the solvent were refined along with the protein structure. Usually after every 3 or 4 cycles, the occupancies were kept constant and a cycle of refinement was carried out on the thermal parameters of the solvent. During the latter stages of the 2.0 Å refinement, the weights applied to the structure factors were changed to the form: $\sigma_0(|F_0|) - S(\sin \theta / \lambda - 1/6)$

where $\sigma_o(|F_o|)$ was taken to be $\sim 0.5 < |F_o| - |F_c| >$ and S was chosen such that the weighted squared discrepancies remained approximately constant over the scattering range (Table 10). A final FRODO intervention was carried out after the 82nd cycle of refinement.

A total of 97 cycles of restrained least squares refinement were carried out on the dimer of α -CHT. The ranges of restraints applied during the course of the refinement are listed in Table 10 along with the restraints applied on the final structure (first value listed) and the r.m.s. deviations from ideal geometry at cycle 97. The refinement of the α -CHT dimer corresponds to 3472 protein atoms, 25534 structure factor amplitudes, 570 chiral centers, 2198 torsion angles and 35598 possible van der Waals contacts. Close examination of Table 10 indicates that the final structure conforms superbly with the ideal geometry and van der Waals contacts. The final R-factor is 0.179, the weighted R-factor being 0.198. If the 247 solvent molecules are removed from the structure, the R-factor increases to 0.218, indicating the strong contribution the solvent makes to the observed diffraction. When the final dimeric structure of α -CHT has hydrogen atoms added in ideal geometrical positions, the R-factor remains essentially constant.

Examination of the R-factor of the final structure of the dimer versus scattering angle, (Figure 9), can be used to estimate the mean coordinate error.⁶² The value

Table 10: Summary of Least Squares Parameters and Deviations.

	<u>Target Sigma</u>	<u>R.M.S. Delta from Ideal</u>
Distances (Å)		
Bond Lengths	0.02 - 0.04	0.021
Bond Angles	0.04 - 0.06	0.057
Planar 1-4	0.05 - 0.08	0.061
Disulfides	0.02 - 0.04	0.030
Planar Groups (Å)		
Deviation from Plane	0.02 - 0.04	0.018
Chiral Centers (Å ³)		
Chiral Volume	0.15	0.210
Non-Bonded Contacts (Å)		
Single Torsion	0.50	0.210
Multiple Torsion	0.50	0.315
Possible (x,y) H-bond	0.50	0.350
Torsion Angles (deg)		
Planar	5.00	8.900
Staggered	15.00	22.000
Orthonormal	20.00	25.100

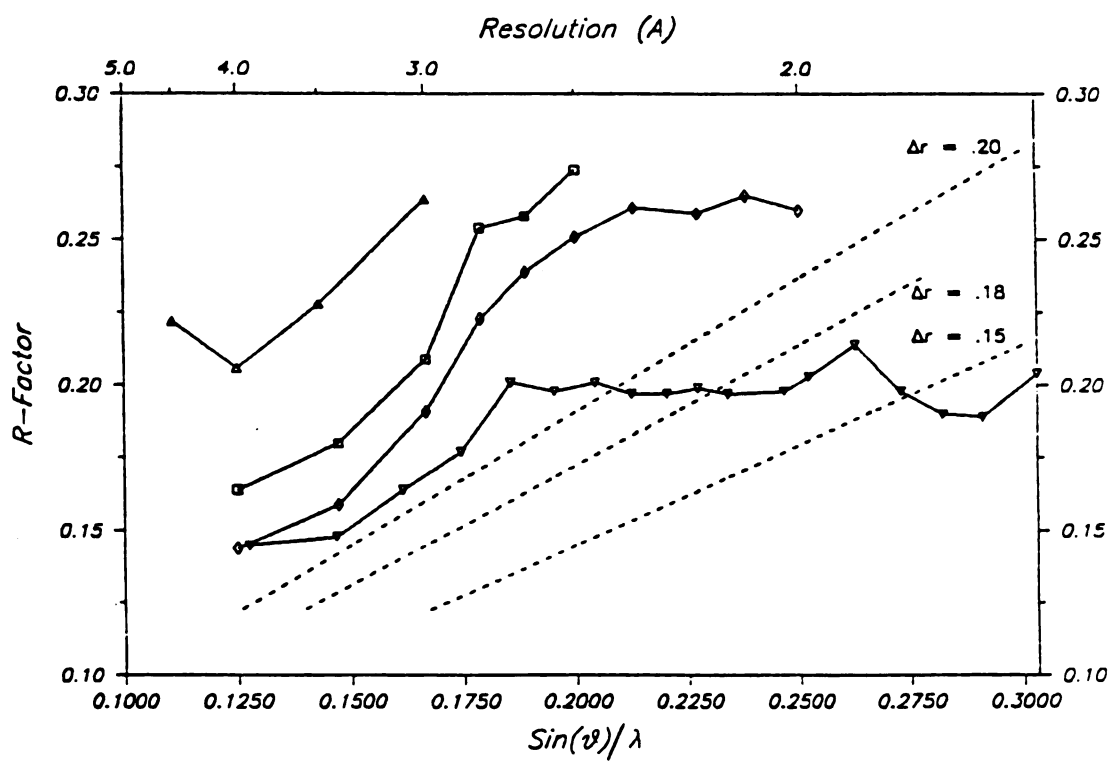
32.22 is the average $|F_C - F_C|$ discrepancy
 Sigmas for FQBS = (19.0) + (-70.0) + (S-1/6)

Isotropic Thermal Factor Restraints

<u>Type</u>	<u>Number</u>	<u>Sigma</u>	<u></u>	<u>r.m.s. B</u>
1	1964	1.0	0.77	0.98
2	2498	1.5	1.25	1.58
3	1586	1.0	0.77	1.02
4	2370	1.5	1.19	1.58

Type 1 =main chain bond, 2 =main chain angle,
 3 =side chain bond, 4 =side chain angle.

Figure 9. Variation of R-Factor with Scattering Angle. Triangles 3.0 Å, squares 2.5 Å, diamonds 2.0 Å, and inverted triangles 1.67 Å resolution; broken lines are theoretical curves for 0.15, 0.18 and 0.20 Å coordinate error.



indicated is ~ 0.18 to $0.20 \text{ \AA}$. These average values assume that all discrepancies between observed and calculated structure factors are due to positional errors. This is clearly not the case, so that some atoms are better positioned than $0.20 \text{ \AA}$ while atoms with large thermal parameters may have a value considerably larger than $0.20 \text{ \AA}$. Furthermore, the choice of weighting scheme applied to the structure factor amplitudes can have considerable effects on the R-factor, particularly the Bragg angle dependence. The mean error values indicated here are similar to those of other comparable refinements.^{11,63,64}

CHAPTER VIII

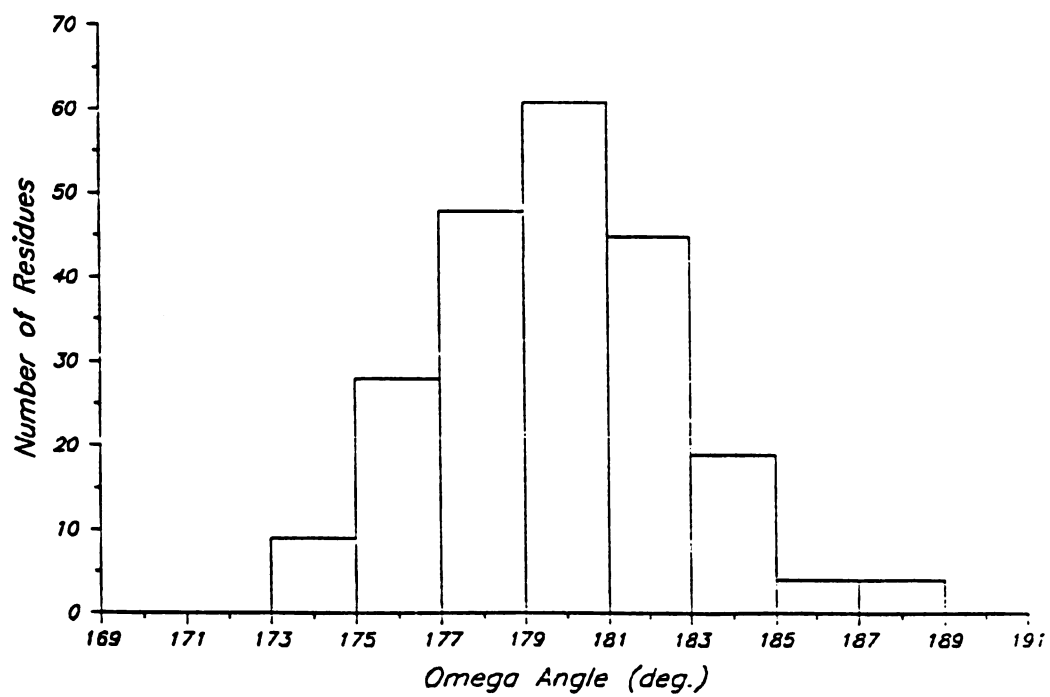
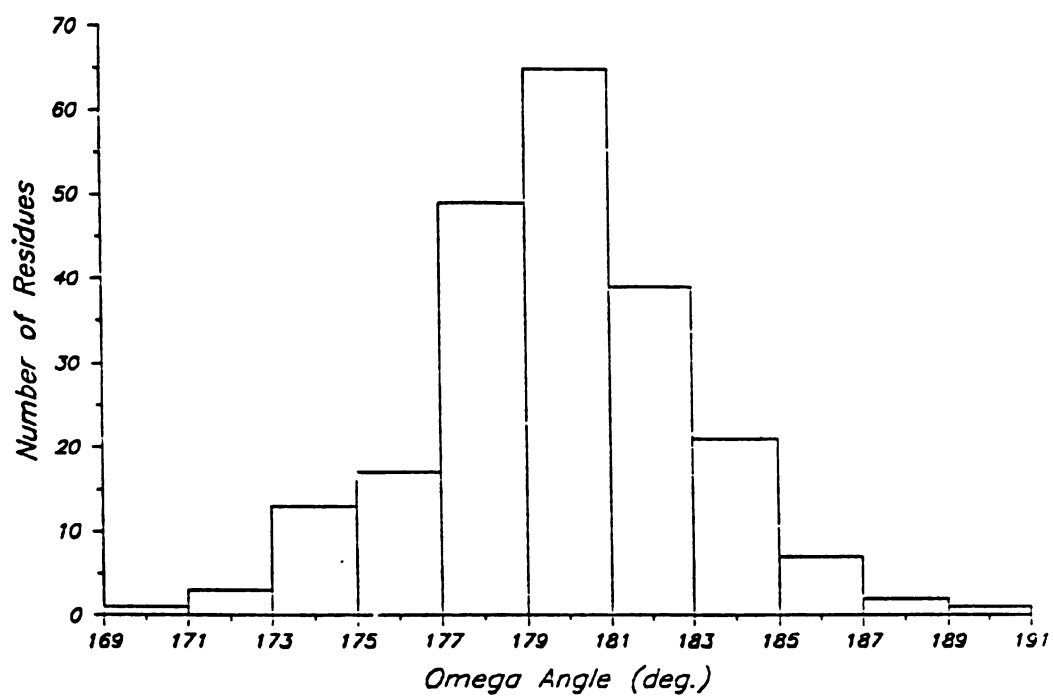
RESULTS OF THE LEAST SQUARES REFINEMENT

A. The Independent Molecules

The coordinates, thermal factors and occupancies of the solvent of the final dimeric structure of α -CHT have been deposited in the Protein Data Bank.⁶⁵

The r.m.s. deviations from ideal values listed in Table 10 correspond very closely in the independent molecules. The beauty of the program PROLSQ is that it is able to restrain geometrical and structural parameters. For instance, it is imperative that the omega angles, which describe the planarity of the peptide bond be close to 180° . A histogram of the omega angle distribution in both molecules of the α -CHT dimer is presented in Figure 10. Taken as a whole, planarity of the peptide units shows an r.m.s. deviation of $0.04 \text{ \AA}$ ($\pm 1.5^\circ$). Generally the angles are within $\pm 5^\circ$ of 180°). Also, the carbonyl carbons of the peptide units should be planar. Despite the fact that this restraint is not explicitly included in PROLSQ, the sum of the angles around the carbonyl carbon averages 359.9° ($\pm 0.16^\circ$). The tau angle (N-CA-C) should be close to 110° . Analysis shows that 90% of the residues in the dimer are

Figure 10. Omega-Angle Distribution. Molecule 1, top
and molecule 2, bottom.



within 7.5° of 110.0° . The Ramachandran plots of the individual molecules are presented in Figure 11. These figures clearly indicate that the non-bonded PHI-PSI contacts of the two molecules conform to the allowed regions.⁶⁶ The dihedral angles of the five disulfide bridges found in each monomer of α -CHT are listed in Table 11. Generally, the conformations of these disulfides are similar, at least within the experimental error of the coordinates. A list of all the torsion angles in the final structure may be found in Appendix C.

A complete list of the hydrogen bonds found in both molecules of α -CHT is given in Appendix D. In preparing this list of hydrogen bonds, polar hydrogens were added to the final structure of each monomer, hydrogen bonds were removed from the list if the hydrogen to acceptor distance was greater than $2.45 \text{ \AA}$ or if the donor to acceptor distance was greater than $3.30 \text{ \AA}$ and if the angle formed by the donor-hydrogen-acceptor was less than 120.0° . This is a very conservative approach of identifying hydrogen bonds, as was the calculation which examined both distances and angles by the extended atom plus polar hydrogen method. A total of 134 and 141 hydrogen bonds were found in molecules 1 and 2 respectively. In molecule 1, of the 134 total, 105 involve main chain donors and acceptors exclusively, 27 involve just one main chain donor and acceptor and 2 involve side chain donors and acceptors. In molecule 2, the respective numbers from the 141 total are

Figure 11. Ramachandran Plots of γ -CHT. Molecule 1, top; molecule 2, bottom, GLY not included.

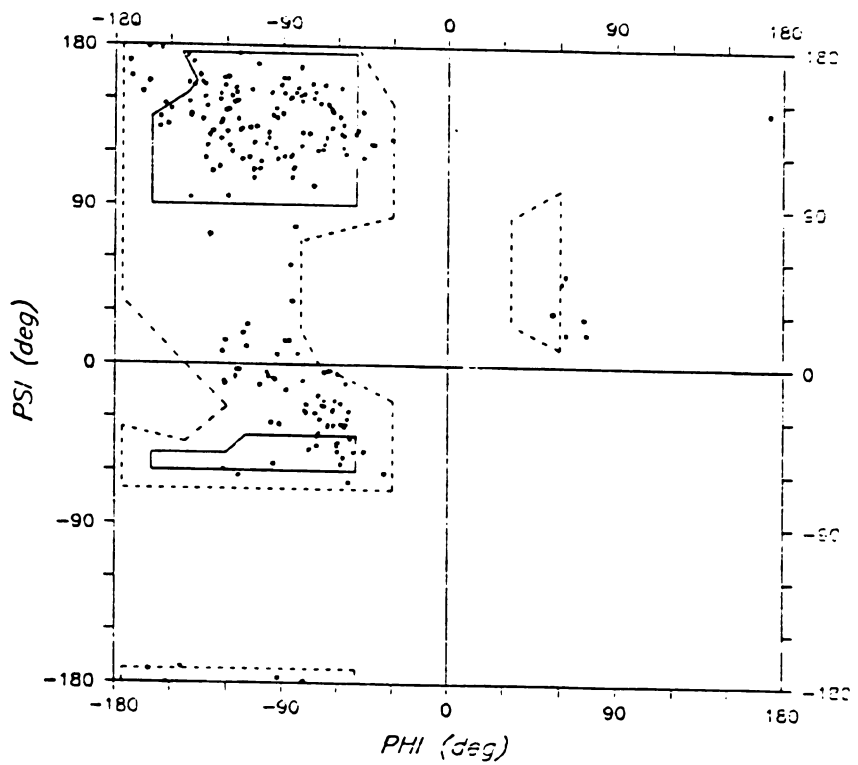
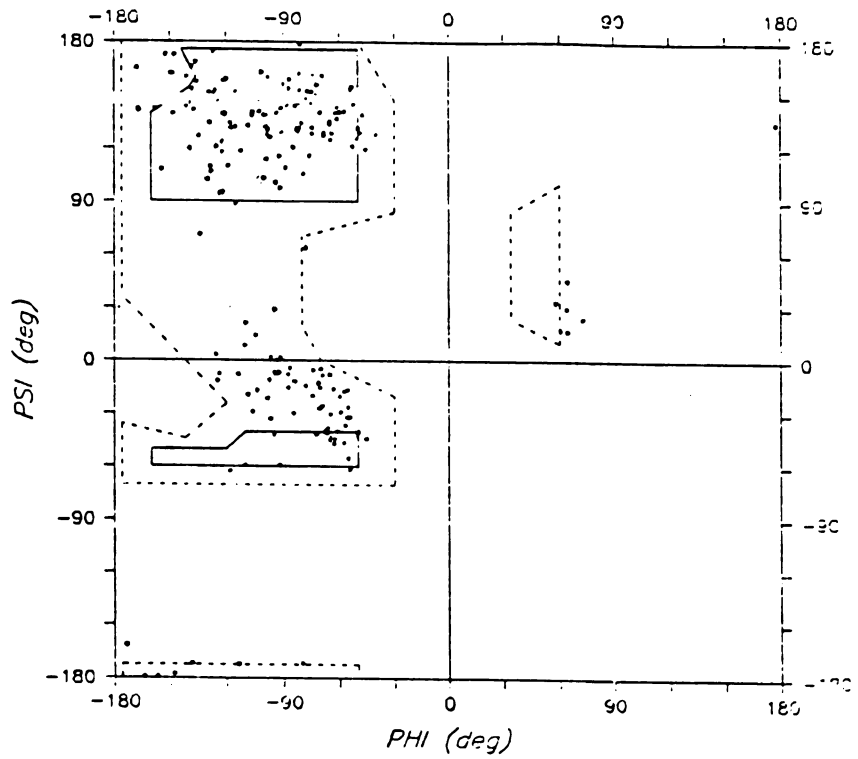


Table 11: Dihedral Angles of Disulfide Bridges of Independent Molecules of α -CHT.

Bridge (molecule)	χ_1	χ_2	χ_3	χ_2'	χ_1'
1-122(1)	64	78	97	-53	-68
(2)	67	72	108	-70	-51
42-58(1)	-106	-140	-86	-92	-69
(2)	-96	-149	-91	-91	-62
136-201(1)	-55	-137	99	-89	-43
(2)	-54	-126	107	-95	-43
168-182(1)	-164	167	-80	-166	-51
(2)	-165	175	-84	-172	-53
191-220(1)	-155	41	98	-168	-60
(2)	-155	44	89	-175	-51

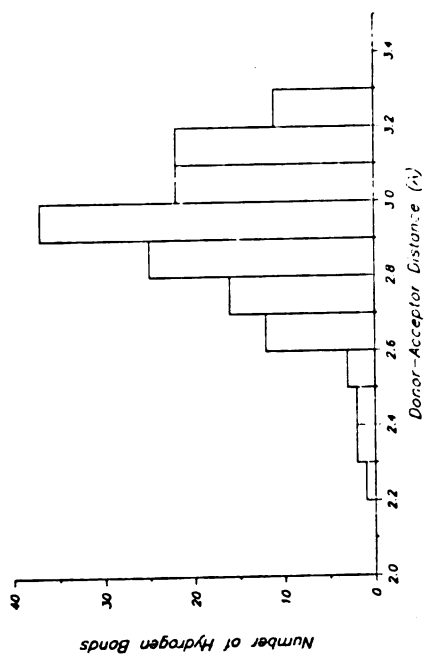
110, 28 and 3. Histograms of the distribution of donor-acceptor and theta angles are shown in Figure 12 for molecules 1 and 2. These results indicate that α -CHT possesses a strong hydrogen bonding pattern, even though stringent criteria were used with the extended atom plus polar hydrogen method to locate the hydrogen bonds. The average donor-acceptor distance is 2.91 Å for both molecules 1 and 2, the average angle between the donor, hydrogen and acceptor is 155.7° and 154.8° for molecule 1 and molecule 2, respectively. These average parameters of the possible hydrogen bonds in α -CHT are very reasonable at least compared to other similar refinements.^{11,63,64}

Evidence of asymmetry between the two molecules of the α -CHT dimer is shown in Table 12, where the hydrogen bonds found in one molecule but not the other are listed. In every case, the hydrogen bonds are found near the surface of the protein or in the dimer interface, reflecting the adaptability of surface residues. Interestingly, there is an additional hydrogen bond found in the catalytic site of molecule 1 that is not found in molecule 2 (56 N-102.0).

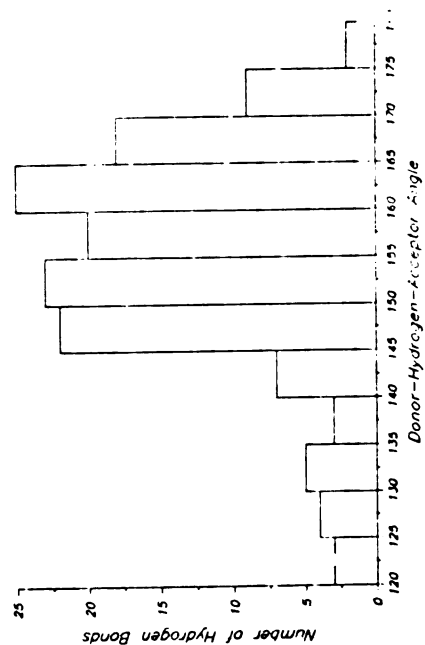
The overall distribution of χ -1 angles (N-CA-CB-CD) of the side chains in α -CHT agrees very well with the trimodal prediction of theoretical calculations and corresponds well to the observed distribution among g^- , t and g^+ positions of a large number of proteins.⁶⁷ Despite the fact that many of the side chain dihedral angles were restrained during the refinement, the observed distribution

Figure 12. Histograms of α -CHT Hydrogen Bond Distances and Donor-Hydrogen-Acceptor Angles. Molecule 1, left; molecule 2, right.

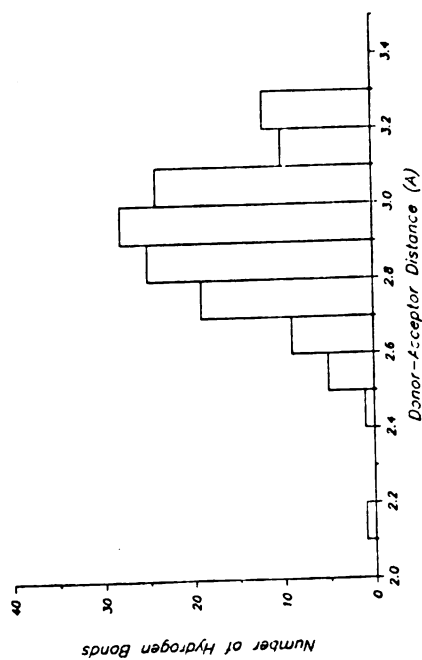
Distribution of Hydrogen Bond Distances (Molecule 2)



Distribution of Hydrogen Bond Theta Angles (Molecule 2)



Distribution of Hydrogen Bond Distances (Molecule 1)



Distribution of Hydrogen Bond Theta Angles (Molecule 1)

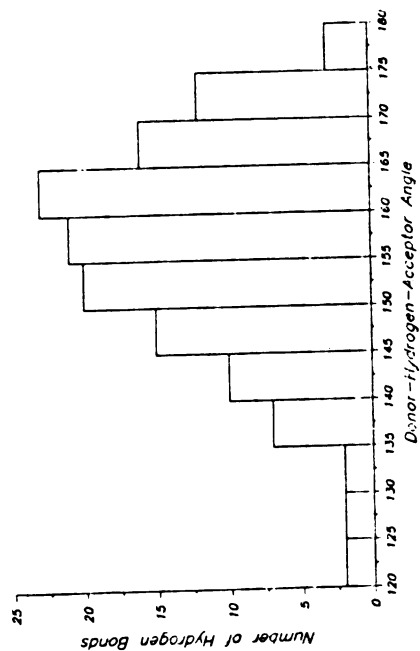


Table 12: Asymmetry of Hydrogen Bonding in α -CHT.

a.) Hydrogen Bonds Found Only in Molecule 1.

	<u>Donor</u>	<u>Acceptor</u>	<u>H-A^a</u>	<u>D-A^b</u>	<u>Theta (deg)^c</u>
157	NE2 HNE2	21 O	2.50	3.40	149.7
39	N HN	35 OD2	2.28	3.21	154.9
38	N HN	35 OD2	1.53	2.52	171.8
40	NE2 HE2	193 O	1.51	2.48	161.6
56	N HN	102 O	2.66	3.54	146.3
154	NH1 HH12	72 OD1	2.59	3.56	163.9
93	N HN	91 OD1	2.18	2.86	123.2
107	NZ HZ2	103 O	2.51	3.41	149.9
118	N HN	115 OG	2.43	3.26	139.3
203	NZ HZ2	128 OD2	2.35	3.35	177.6
139	N HN	198 O	2.73	3.60	145.4
145	NH2 HH22	150 OD1	1.96	2.95	172.7
230	NH1 HH12	165 OD1	2.49	3.41	158.4
175	NZ HZ3	172 O	2.58	3.50	153.3

b.) Hydrogen Bonds Found Only in Molecule 2.

	<u>Donor</u>	<u>Acceptor</u>	<u>H-A^a</u>	<u>D-A^b</u>	<u>Theta (deg)^c</u>
2	N HN	120 O	1.93	2.93	173.4
18	ND2 HND2	187 O	2.08	2.97	146.5
157	NE2 HNE2	20 OE2	2.28	3.24	159.4
39	N HN	35 OD1	2.53	3.48	158.8
37	N HN	35 OD1	1.93	2.85	151.4
75	N HN	72 O	2.49	3.48	170.6
98	N HN	95 OD1	2.04	2.99	158.8
118	N HN	115 O	2.41	3.25	141.3
125	N HN	128 OD2	1.58	2.40	135.6
127	N HN	125 OG	2.54	3.47	153.4
167	N HN	164 OG	2.23	3.14	149.8
224	N HN	221 O	2.61	3.57	163.0

^aHydrogen to acceptor distance.

^bDonor to acceptor distance.

^cDonor-hydrogen-acceptor angle.

generally reflects the starting angular conformations. Figure 13 presents a more detailed comparison of the χ -1 angles. There is no preferred conformation of the χ -1 angles in SER residues while THR residues prefer the g^- and g^+ positions. This probably results from the greater steric hinderance of the methyl group of THR. The residues VAL and ILE/LEU show a marked preference for the g^+ conformation. Here, one CG is in g^+ and the other in the t position. The behavior of other classes of residues is in general agreement with the observations of larger comparisons.

The behavior of the thermal parameters of the independent molecules is summarized in Table 13 and shown graphically in Figure 14. Restrained individual thermal parameters were introduced during the latter stages of the 5.0-2.5 Å resolution refinement, and refined thereafter. Examination of Table 13 and Figure 14 indicates that the thermal parameters of the independent molecules are fairly similar, this being especially true near residues 39, 110, 130, 160-180, 205 and 215-225. Since a symmetry restraint on the thermal parameters was not included in the refinement, the agreement between the two molecules is a reassuring result. The region from 70-80 is noticeable in both molecules. In molecule 1, this region was disordered and was not included in the refinement (occupancies were assigned a value of 0.01). However, this

Figure 13. Distribution of Some Side Chain Conformational Angles. (a) SER, (b) THR, (c) VAL, (d) ILE and LEU.

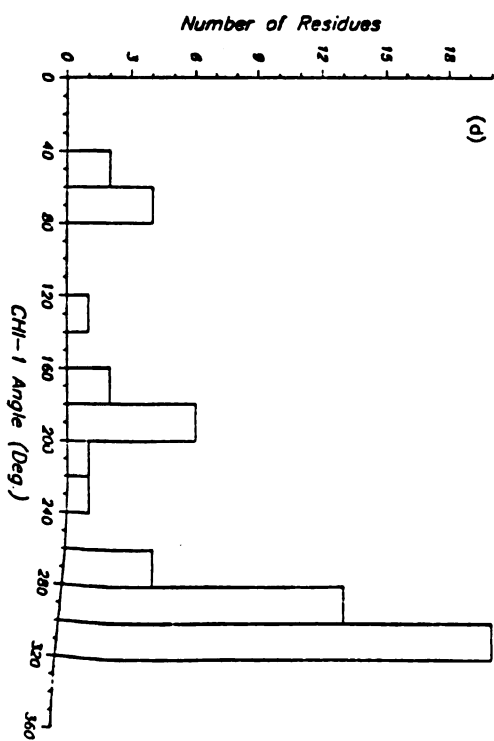
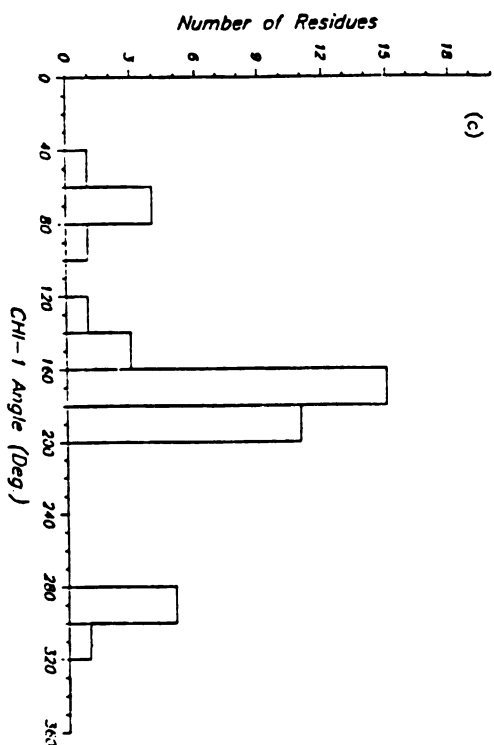
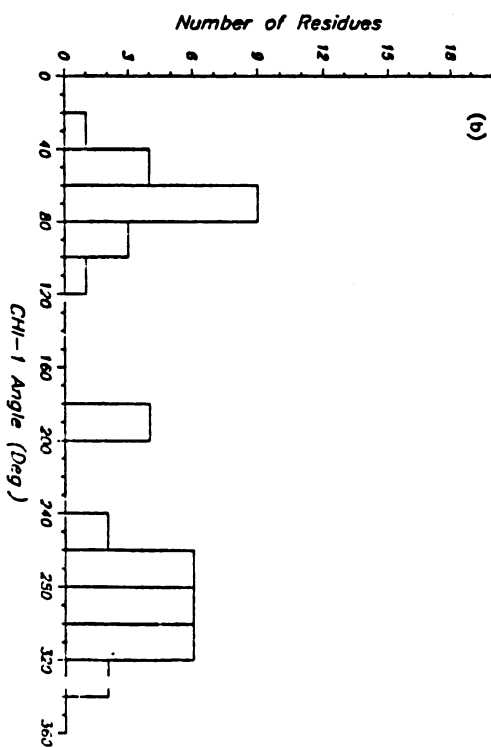
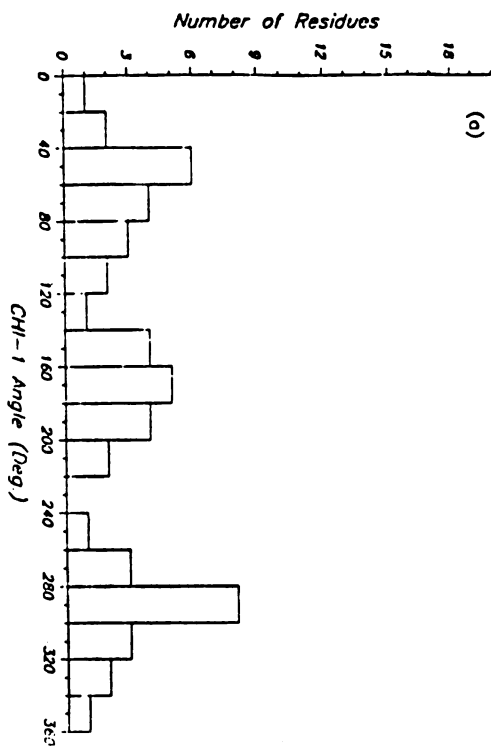


Table 13: Average Thermal Parameters of the α -CHT Dimer ($\text{\AA}^2$).^a

	<u>Molecule 1</u>	<u>Molecule 2</u>
Protein Atoms (3472) ^b	15.7	16.1
Main Chain (708)	15.0	15.2
Carbonyl Oxygens (239)	15.2	15.4
Side Chains (789)	16.4	17.0
Sulfurs (12)	12.5	13.2
Interior (480)		
Main Chain	10.5	10.3
Side Chain	10.8	10.9
Exterior (1256)		
Main Chain	16.8	17.1
Side Chain	18.5	19.2
Dimer Interface (280)		
Main Chain	15.5	16.2
Side Chain	16.2	17.2
Catalytic Site (24) ^c		
Main Chain	10.1	8.5
Side Chain	10.1	9.3
TRP Cluster (56) ^d		
Main Chain	9.2	10.6
Side Chain	9.9	11.3
Domain 1 (859) ^e		
Main Chain	15.7	15.1
Side Chain	16.7	16.6
Domain 2 (877) ^f		
Main Chain	14.0	14.0
Side Chain	15.6	15.6
Waters	22.2	
Sulfates	24.7	

^aThe overall average B-factor = 15.9

^bNumbers in parentheses indicate the number of atoms in each region.

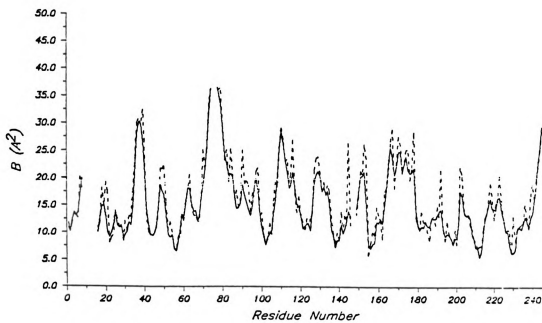
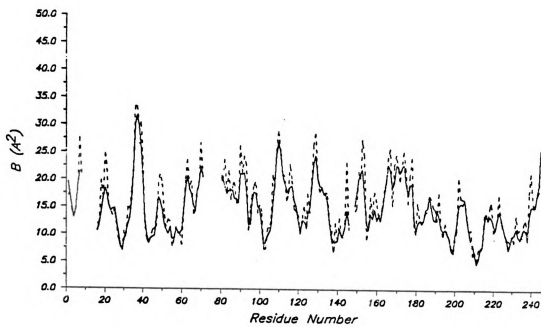
^cHis 57, Asp 102, Ser 195.

^dTrp 27, Pro 28, Trp 29, Trp 207.

^eResidues 1-122.

^fResidues 123-245.

Figure 14. R.M.S. Thermal Parameters of α -CHT. Main chain, solid; side chain, broken; molecule 1, top; molecule 2, bottom.



only applies to residues 74-76 of molecule 2, although this region shows large B-values also.

The average thermal parameter for the dimer of α -CHT is about $16 \text{ \AA}^2$, while that of the sulfur atoms of disulfide bonds is much smaller. As in the energetic refinements of γ -CHT presented in part I, the chymotrypsin monomer was divided into convenient structural regions for analysis (see Appendix B). The thermal parameters for the interior of the individual molecules are much smaller (average $B = 11.0 \text{ \AA}^2$) while the average parameter for the dimer interface, which is located in the interior of the dimer is comparable to that of the exterior atoms. This is probably related to the asymmetry of the dimer. Two other interesting regions are the catalytic site and the TRP cluster. Both of these regions show the smallest thermal parameters in the enzyme, despite the fact that the TRP cluster is located near the surface and the catalytic site is in the interface region of the dimer.

B. Solvent Structure

Spacegroup symmetry (P21) was used to place the solvent as close as possible to the protein while at the same time, removing short contacts between water molecules and protein atoms. In most cases, 10-15% of the possible water molecules were removed from consideration due to short contacts or the fact that no symmetry operation could place the water

within 8.7 Å of any protein atom. Once the selection was accomplished, the new waters were examined using FRODO. It is much easier to remove additional waters at this point, since the position of the water can be examined in terms of positioning and whether it is in positive difference density. At times, the position of a possible water could be close to disordered density of a side chain. In most of these cases, the water was removed from the calculation. If the distance to the side chain was at least 2.0 Å, the water was accepted, and monitored during the rest of the refinement. During the first solvent additions, the occupancies of the waters were set at 1.0 and the B-factors to 25.0 Å². Since the solvent usually refined to occupancies of 0.5-0.8, during the last few examinations of difference electron density maps, new water molecules were introduced with occupancies of 0.75 to speed convergence.

The same procedure was used throughout in refining the water structure. The B-factors of the protein were refined continuously from the 2.0 Å resolution stage to the end. However, the refinement of the solvent proceeded as follows. After the inclusion of new water in the refinement, 3 cycles of refinement on B-factors and coordinates of protein atoms and occupancies and coordinates of water were performed. This was followed by 1-2 cycles of refinement on B-factors and coordinates of protein atoms and water. At the conclusion of the refinement of the α-CHT dimer, a total of 247 solvent molecules had been introduced. This

number compares very well with the 151 waters found in the refinement of γ -CHT¹¹ and is generally a conservative estimate (~ 0.5 solvent molecules/residue). Examination of the final solvent structure of the α -CHT dimer revealed that 6.0% of the water is located in the interior of the enzyme, 85.5% is found in the exterior and 8.5% is located in the dimer interface, which is interior in the dimer. A list of the relative locations of the solvent molecules in the α -CHT dimer may be found in Appendix E.

The distributions of the final occupancies and thermal parameters of the solvent structure are shown in Figure 15. Both distributions are skewed toward greater significance and possess highly acceptable average values. In addition, the values of the occupancies suggest a well-defined absolute scale for the observed data. The average value for the thermal parameters of the solvent is $\sim 22.0 \text{ \AA}^2$. Despite the fact that this value is larger than the average for the protein atoms, it is still less than the value given to the solvent when introduced in the refinement so that the solvent makes a significant contribution to the low-order data. The overall R-factor is reduced by 0.039 by including solvent. The distance distribution of solvent molecules from protein atoms and themselves is shown in Figure 16 from which it may be seen that over half the solvent can potentially hydrogen bond to the protein. The solvent-solvent minimum distance distribution is featureless because in this refinement, the solvent structure basically

Figure 15. Distribution of Occupancies (top) and Thermal Parameters (bottom). Occupancies greater than one were set to one during refinement.

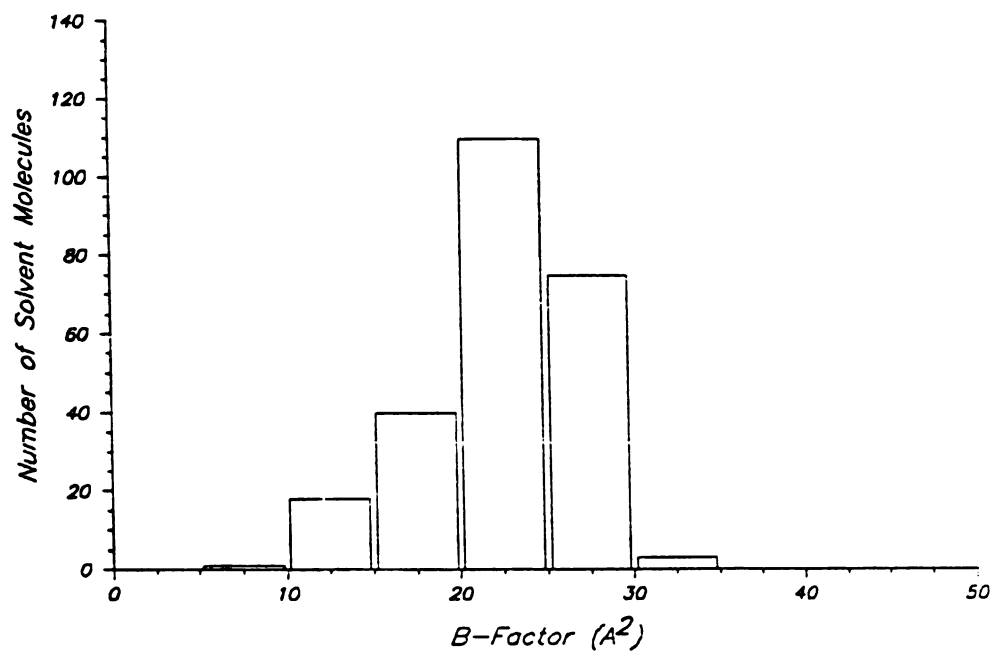
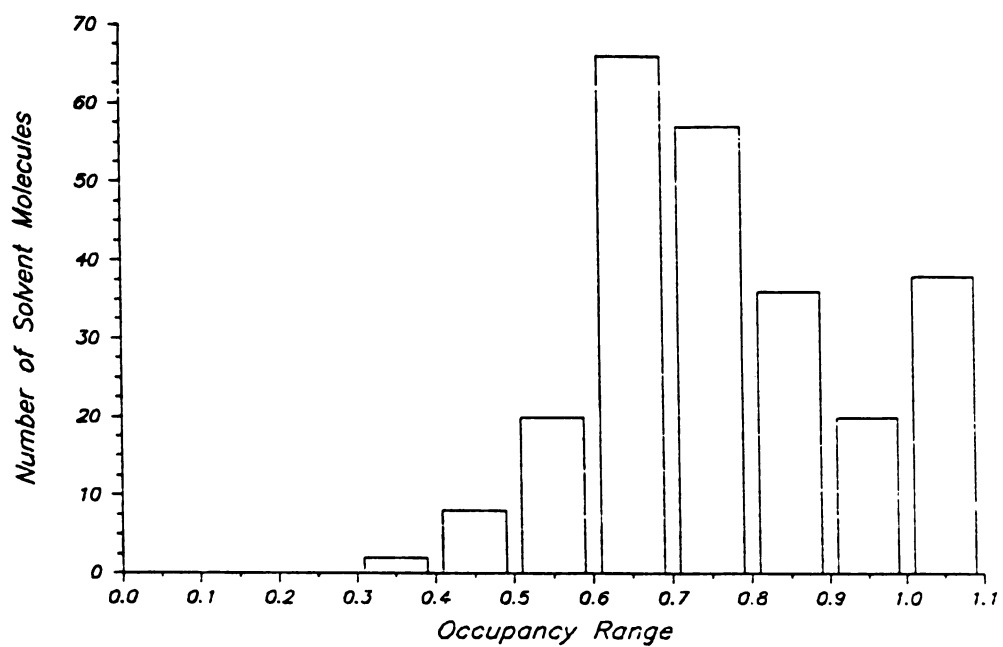
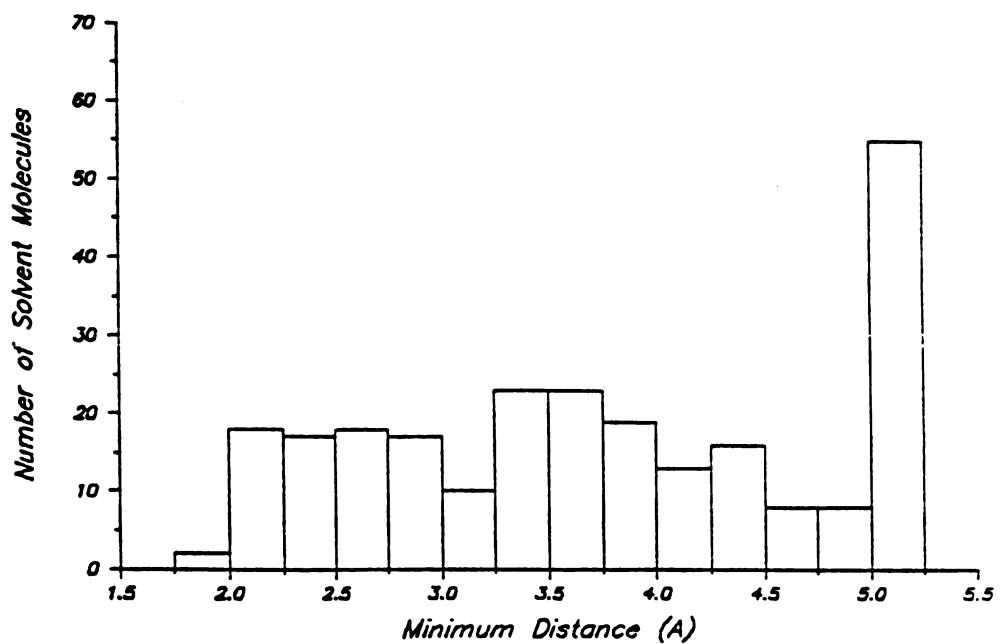
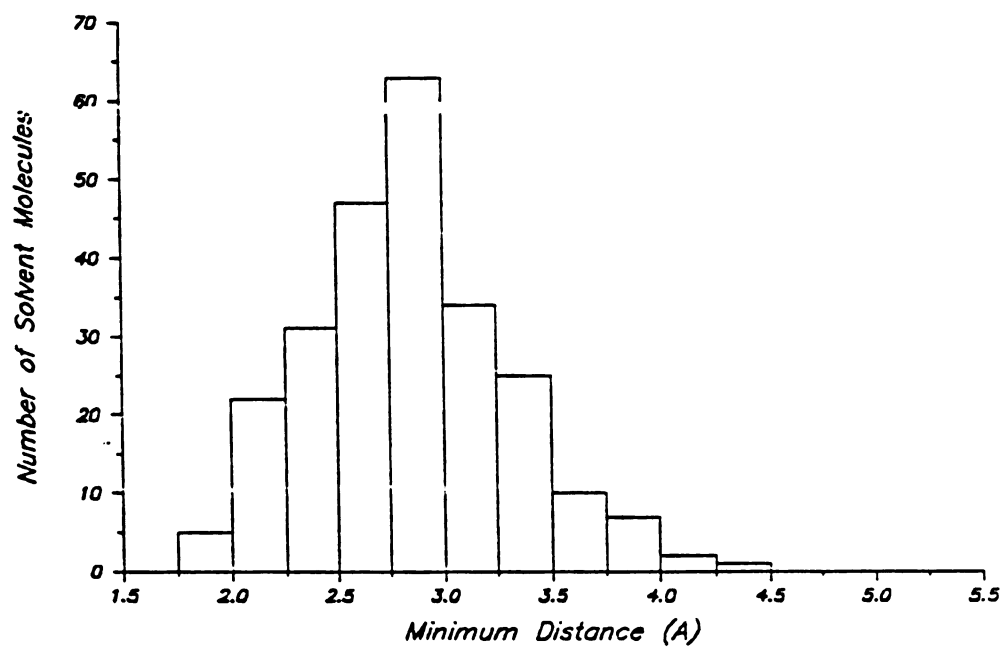


Figure 16. Distribution of Solvent-Protein (top) and Solvent-Solvent (bottom). All solvent-solvent minimum distances >5.0 Å grouped together.



consists of a 1-2 atom thickness shell around the protein which is necessarily less dense than liquid water. Weaker peaks in the difference electron density maps were not pursued beyond this layer. Thus, there is no clear indication of liquid water structure from the solvent-solvent distances.

A complete list of protein-solvent hydrogen bonds is presented in Appendix F. In molecule 1, 42 water molecules hydrogen bond to a protein atom and in molecule 2, there are 41 protein-water hydrogen bonds. In all cases, the donor-acceptor distances and the hydrogen bond angles possess highly acceptable average values. All hydrogens were added to the water molecule oxygens in idealized geometrical positions; however, the orientation of the hydrogen atom is completely random in space. Many possible protein water interactions were not included in this list of hydrogen bonds since it was unrealistic to use the water molecule as the hydrogen bond donor. A list of polar protein atom-water molecule interactions was therefore generated to include interactions where the water oxygen may have been the hydrogen bond donor. This list is presented in Appendix C.

C. Dimer Asymmetry

During the course of the refinement of the α -CHT dimer, deviations from the non-crystallographic 2-fold

symmetry were investigated by calculating the rotation matrix and translation vector that minimized the squares of the differences in the coordinates between the independent molecules. Although all atoms were used in these calculations, removal of large discrepancies did not alter the results for practical purposes indicating that the asymmetry is not systematic but basically random. The matrix-vector relating Cartesian Angstrom coordinates of molecule 2 to molecule 1 is:

$$\begin{pmatrix} .9138 & -.0066 & .4059 \\ -.0017 & -.9999 & -.0126 \\ .0406 & .0108 & -.9138 \end{pmatrix} \begin{pmatrix} -9.94 \\ 40.60 \\ 47.60 \end{pmatrix}$$

The development of main-side chain asymmetry was noted after the first few cycles of refinement at 5.0-3.0 Å resolution. There are discontinuities in the asymmetry at cycles 18, 47 and 82 which are related to the manual interactive graphics interventions using FRODO. These discontinuities decrease with extent of refinement indicating that the Fourier and least squares results finally converge to the same structure.

Closer examination of the average thermal parameters of the α-CHT dimer along with visual inspection of difference electron density maps revealed that generally, atoms with thermal factors greater than 23 Å² did not usually appear reliably so that their positions are somewhat uncertain. Therefore, in all the analyses of the asymmetry in the

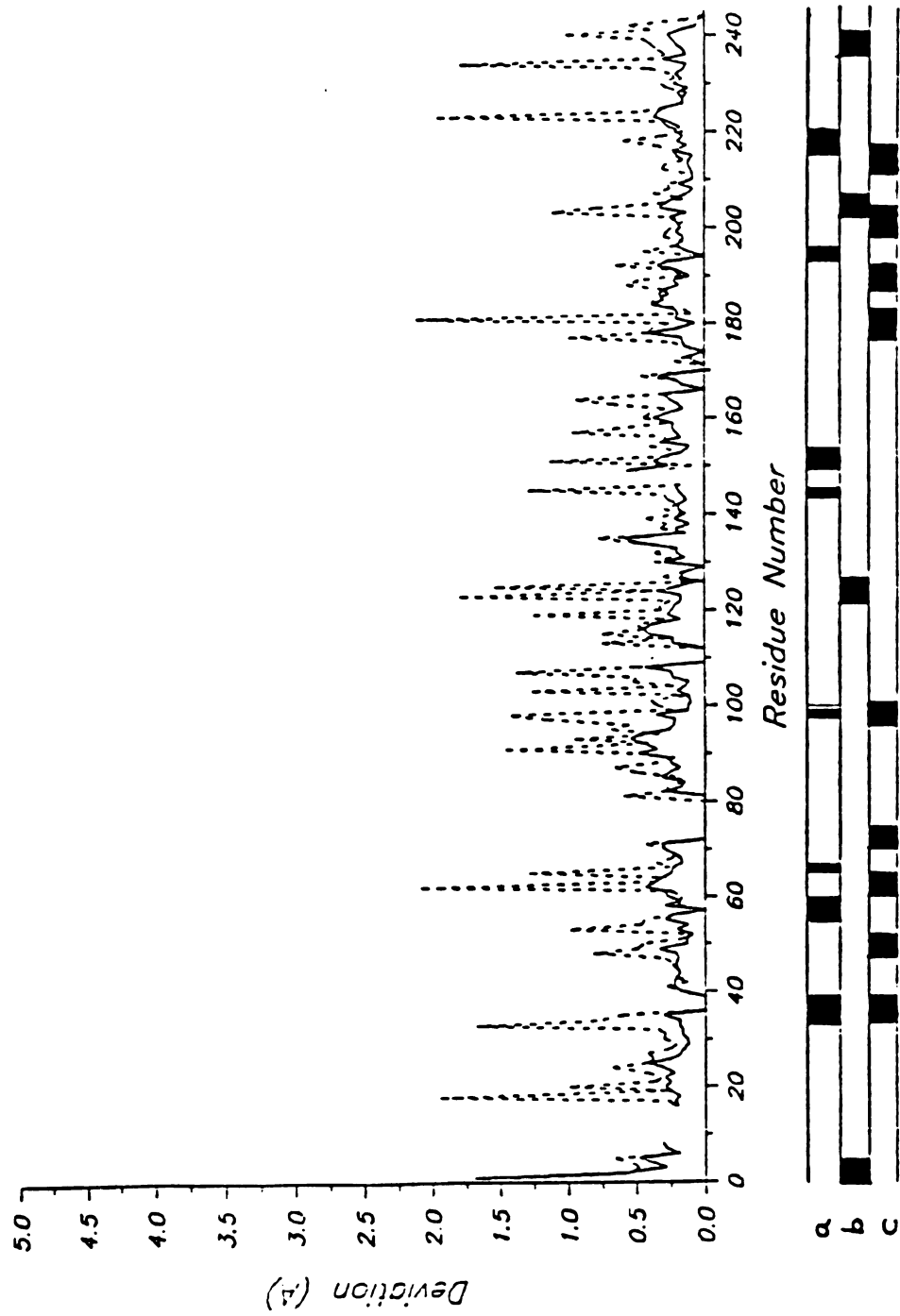
dimer, atoms whose B-factors which were greater than $23 \text{ \AA}^2$ were removed from consideration. Table 14 summarizes the results of the asymmetry present in the α -CHT dimer. The overall asymmetry for the main chain atoms is $0.24 \text{ \AA}$ while that for the side chains is $0.64 \text{ \AA}$. The interior displays the most symmetry while the exterior and dimer interface residues (which are internal in the dimer) are nearly equal in asymmetry. The catalytic site and the TRP cluster also show good symmetry. One would expect the atoms nearer the surface of a protein to be less well defined and this may be seen in Table 14. About a quarter of the atoms are removed from consideration of the surface and about a sixth are removed from the dimer interface.

The error in the coordinates has been determined to be between 0.18 and $0.20 \text{ \AA}$ (Figure 9). It is clear then that the main chain possesses a high degree of fidelity between the two molecules and the folding is essentially 2-fold like within experimental error. Only a few regions of the main chain approach $0.5 \text{ \AA}$ in asymmetry, two of these being terminal residues (PRO 8, TYR 146). The same does not apply to the side chains where there are highly significant deviations from 2-fold symmetry. Figure 17 shows the average asymmetry per residue, separated into main and side chain components, for atoms whose thermal factors are less than $23 \text{ \AA}^2$. The summary of Table 14 shows that 10-15% of the dimeric structure is asymmetric with almost all of it residing in the side chains (~25%).

Table 14: R.M.S. Asymmetry for α -CHT Dimer.

	<u>Asymmetry (Å)</u>	<u># Atoms Removed</u>		
Protein Atoms	0.47			
Main Chain	0.24			
Carbonyl Oxygens	0.35			
Side Chains	0.64			
Sulfurs	0.18			
Interior				
Main Chain	0.18	(0)		
Side Chain	0.49	(0)		
Exterior				
Main Chain	0.27	(90)		
Side Chain	0.68	(206)		
Dimer Interface				
Main Chain	0.29	(14)		
Side Chain	0.59	(29)		
Catalytic Site				
Main Chain	0.12	(0)		
Side Chain	0.29	(0)		
TRP Cluster				
Main Chain	0.19	(0)		
Side Chain	0.35	(0)		
Domain 1 (1-122)				
Main Chain	0.27	(60)		
Side Chain	0.67	(127)		
Domain 2 (123-245)				
Main Chain	0.22	(30)		
Side Chain	0.60	(79)		
+ Cutoff, B > 23.0 Å ² Removed 157 atoms Mol. 1 and 139 Mol. 2				
	<u>Main Chain</u>	<u>Carbonyl Oxygens</u>	<u>Side Chain</u>	<u>Overall</u>
0.0 - 0.25 Å	485	107	229	794
0.25 - 0.50 Å	142	87	262	491
0.50 - 0.75 Å	13	11	50	74
0.75 - 1.00 Å	3	1	26	30
1.00 - 1.50 Å	2	0	20	22
>1.50 Å	0	1	28	29

Figure 17. R.M.S. Asymmetry Between Individual Molecules of α -CHT. Only atoms with $B < 23.0^2$ are included; main chain, solid; side chain, broken; a-dimer interface regions, b-dyad B regions near noncrystallographic 2-fold axis between dimers, c-external turns.



Previous studies of 2.8 Å difference electron density maps between the two molecules have shown that ~16% of the density had differences that were greater than 0.7 eÅ^{-3} or $3\sigma(\Delta\rho) = 3(\sqrt{2}\sigma(\rho_o))$.¹⁴ Some of the surface asymmetry may be attributed to inter-dimer contacts (LYS 203 and ASN 204 and ASN 236-VAL 233)¹⁵ and external loops (60-65 and 95-99) in one of the two similar antiparallel β -sheet barrel domains. This accounts for only about 1/3 of the observed asymmetry. The remainder of the observed asymmetry must simply reflect the high degree of adaptability associated with tertiary surface structure.

A view of the final dimeric structure of α -CHT (CA atoms only) is shown in Figure 18, viewed down the crystallographic x and y axes. Although difficult to discern, some differences between the two molecules may be seen, especially in turns near the surface of the protein. A representative view of the surface asymmetry is shown in Figure 19 (residues 172-179), from which side chain asymmetry can easily be seen. An overall stereoview of the asymmetry is shown in Figure 20. Here, side chains possessing an average asymmetry greater than 0.5 Å were drawn. The lack of asymmetry in the interior as well as the total lack of aromatic residues is noticeable.

The dimer interface interactions are listed in Table 15. Included in this list are potential hydrogen bonds and ion pairs. It is clear from the list that half of the interface interactions display good symmetry relations.

Figure 18. Stereo CA Plots of the α -CHT Dimer. Top
— view down XO (local 2-fold axis), bottom
— view down YO (2-fold axis can be seen).

(Top)

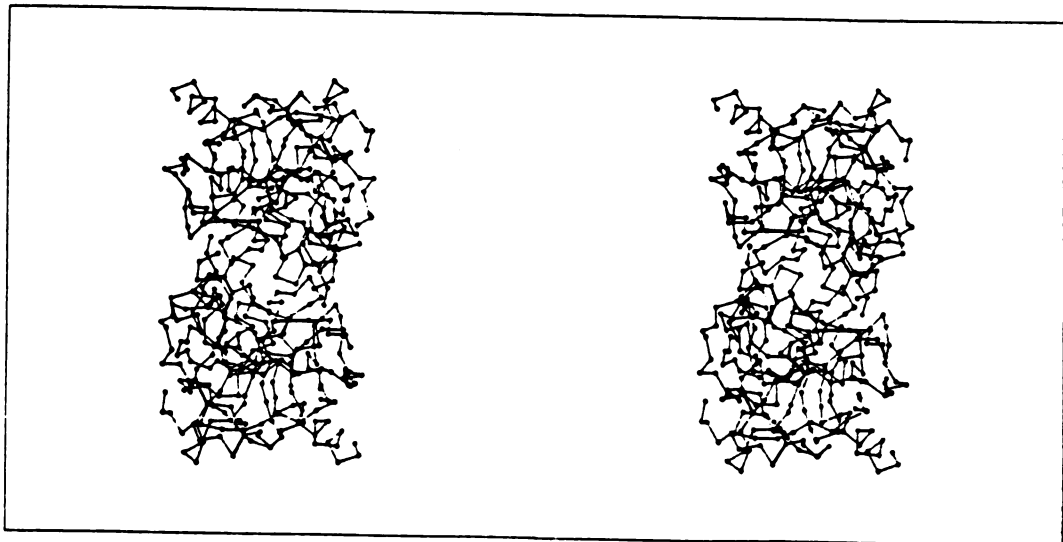
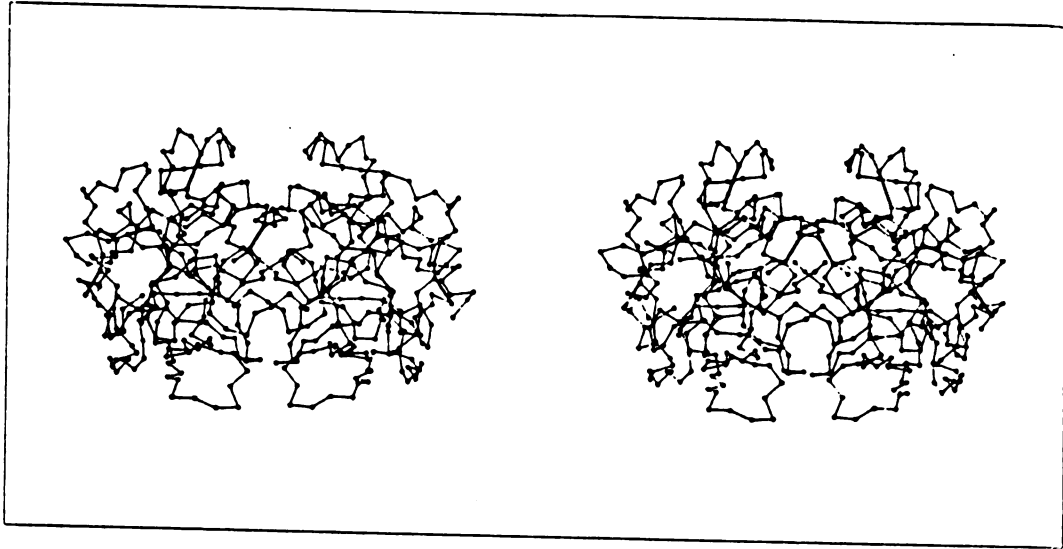


Figure 19. Stereoview of Representative Surface Asymmetry.
Residues 172-179; molecule 2 bold.

(Top)

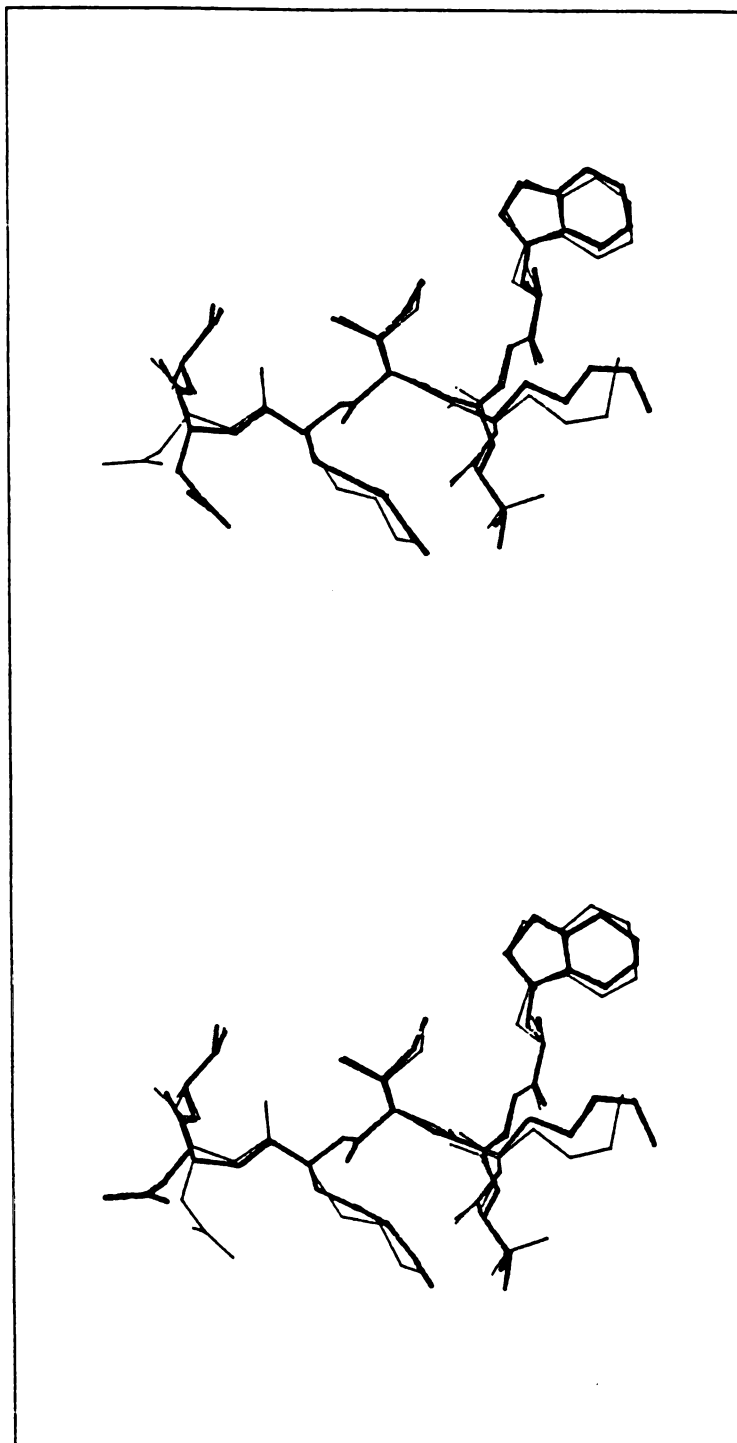


Figure 20. Stereoview of Overall Asymmetry of α -CHT. Viewed down local 2-fold axis designated by asterisk; side chains shown only if r.m.s. asymmetry $> 0.5 \text{ \AA}$ and $\langle B \rangle < 23.0 \text{ \AA}^2$; main chain atoms corresponding to these residues are also shown.

(Top)

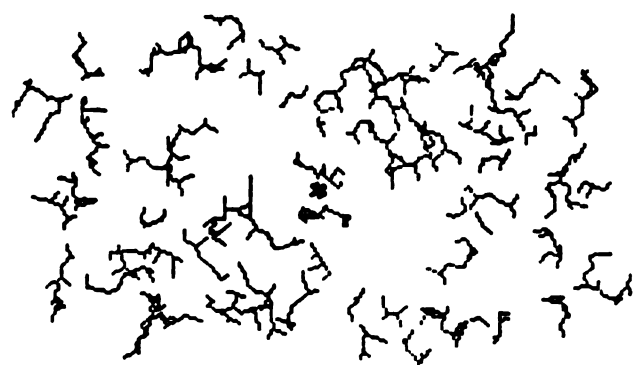
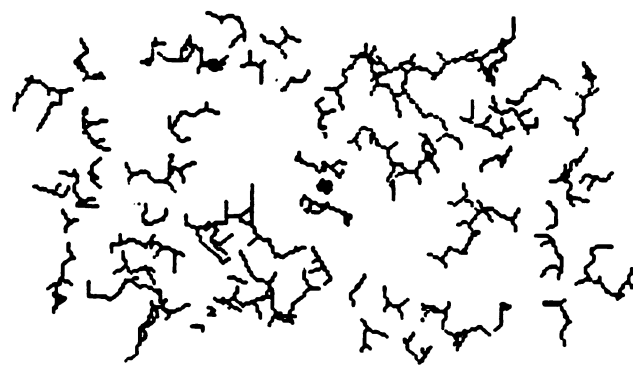


Table 15: Dimer Interface Interactions in α -CHT.

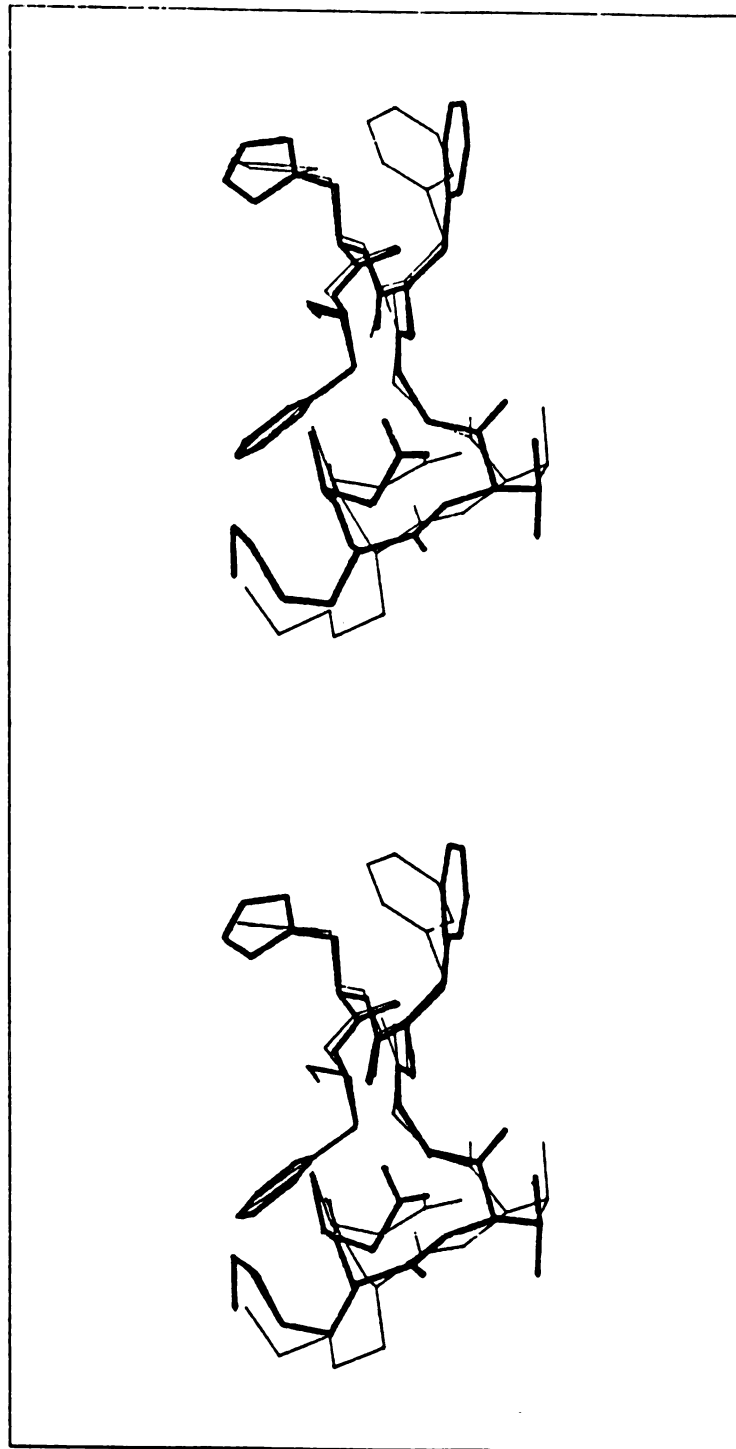
Molecule 1	Molecule 2	1-2 (Å)	2-1 (Å)	
ASP 35OD1	THR 151OG1	3.16		Hbond
THR 37OG1	PHE 39CE1	3.24		
HIS 57O	TYR 146O	3.37	3.34	
HIS 57O	TYR 146O	2.77	2.76	Hbond
HIS 57O	TYR 146OT	3.43	3.27	} Ion Pair
HIS 57CD2	TYR 146OT	3.30	3.47	
HIS 57O	ASN 150ND2	3.12	2.92	Hbond
GLY 59O	ALA 149N	2.79	2.61	Hbond
GLY 59O	ALA 149CA	3.33	3.19	
ASP 64CG	ALA 149N	3.37		} Ion Pair
ASP 64OD1	ALA 149N	2.42	3.06	
SER 96CB	ARG 145NH1	3.30		
ILE 99CG1	THR 219OG1	3.47		
TYR 146CE2	SER 214O	3.45		
TYR 146CZ	SER 214O	3.37		
TYR 146OH	SER 214O	2.52	2.66	Hbond
ASP 153CG	THR 37CG2	3.31		
ASP 153OD1	THR 37CG2	3.37		
ASP 153OD2	THR 37CG2	2.80		
GLY 216O	SER 218CB	2.78	2.63	
SER 218CG	GLY 216O	3.40		

Even in the interactions where symmetry is noted, some asymmetry is present in the van der Waals distances as well as the angles (not listed). Very prominent is the possible ion pair between the terminal carboxyl group of the B-chain (TYR 146) of one molecule and the protonated imidazole of the catalytic site of the other molecule. Both of these groups may be charged at pH 3.5, although a hydrogen bonding protonated TYR 146 carboxyl group is likely from previous change in pH studies.⁶⁰ Another possible ion pair occurs between ASP 64 and the amino terminal of the C-chain ALA 149. Asymmetry is evident in the two molecules; a water molecule is present in one molecule but not the other, complicating this interaction. There are about ten additional hydrogen bonds in this region as well as some additional close contacts, especially near the local 2-fold axis (GLY 216-SER 218). During the refinement, the position of OG of SER 218 was monitored, and near the end it was moved $\sim 40^\circ$ about $\chi-1$ away from GLY 216 of the other molecule. Examination of the difference electron density maps showed clearly that the position of the OG atom on SER 218 of the second molecule was positioned correctly. Other close contacts were not corrected due to a lack of electron density or difference density indications. A typical example of the asymmetry in the dimer interface region is shown in Figure 21 where residues 35-41 are superimposed. As in the example of surface asymmetry, most of the asymmetry is found at

.

Figure 21. Stereoview of Typical Dimer Interface
Asymmetry. Residues 35-41; molecule 2, bold.

(Top)



the side chain atoms; some symmetry is preserved even in the side chains.

D. The Active Site

The catalytic residues of the independent molecules (HIS-57, ASP-102, SER-195) display excellent 2-fold symmetry (Table 14), well within the estimated coordinate error (Figure 9), despite the fact that they are located in the dimer interface region of the molecule. A stereo-view of the residues of the active site superimposed upon one another is presented in Figure 22 where it may be seen that the differences between the two molecules may be 1) a slight displacement of the imidazole ring of HIS-57 and 2) the positions of CB and OG of SER-195. The χ -1 angles of SER-195 are -84° and -104° in molecule 1 and molecule 2, respectively. These angles are very similar to the angle of SER-195 found in the high resolution refinement of γ -CHT¹¹ and in other serine proteases.⁶⁸ The OG of SER 195 is nearly coplanar with the imidazole of HIS-57, the out-of-plane deviations being -0.26 and $+0.27$ Å, respectively. This may lead to hydrogen bonding between SER-195 OG and HIS-57 NE2, but when the hydrogen bond angle is examined, the hydrogen to acceptor distance is 2.3 and 2.1 Å with angles of 119° and 102° . Thus, a hydrogen bond is very unlikely here, in agreement with the results found for γ -CHT but for a different reason. In γ -CHT, a hydrogen bond between SER-195 OG and HIS-57 NE2

Figure 22. Stereoview of Catalytic Residues of Independent Molecules of α -CHT. HIS-57, ASP-102, SER-195; molecule 2, bold.

was not feasible because OG was over 0.7 Å out-of-plane of the imidazole ring, giving a distance of 3.8 Å. If the donor-acceptor roles are reversed in α-CHT, which is possible at pH 3.5 where the imidazole is protonated, the hydrogen bond angles are still unacceptable, being ~120°. A complete list of hydrogen bonds in the active site is given in Table 16, including interactions with solvent molecules. Most of the hydrogen bonds listed in Table 16 occur in γ-CHT, an exception being the lack of hydrogen bond between 214 OG and 102 OD1 in α-CHT due to a close inter-dimer contact near SER 214. A short contact also occurs between 56 N and 102 OD2 in α-CHT (~2.95 Å); however the hydrogen bond angles are small (108° and 116° respectively).

Examination of the list of water molecules which show 2-fold symmetry in α-CHT (Table 17) and in the active site (Table 16) reveals that the solvent molecule hydrogen bonding pattern in the active site residues is indeed asymmetric. Five hydrogen bonds are found in molecule 1 and six in molecule 2, but of these hydrogen bonds, only 3 are in both molecules. One of these involves a hydrogen bond to ASP 102 N, but the other two are hydrogen bonded to the TYR residue from the other molecule, and interacting with the imidazole ring of HIS 57. The average thermal factor for the solvent molecules in the active site regions are 21.1 and 22.2 Å² for molecules 1 and 2, respectively. The corresponding occupancies are 0.81 and 0.83. Averaging only those water molecules that are found in both active

Table 16: Hydrogen Bonds in the Catalytic Sites.

a.) Involving Protein Atoms.

Donor-Acceptor	Molecule 1		Molecule 2	
	Distance (Å)	Angle (deg)	Distance (Å)	Angle (deg)
195OG-57NE2	3.02	118.7	2.68	102.2
57ND1-102OD1	2.68	161.1	2.58	169.0
57N-102OD2	2.85	167.8	2.79	150.4
43N-195O	2.56	164.5	2.60	168.0

b.) Involving Solvent Atoms.

Molecule 1		Molecule 2	
Water 514	— 102 N 3.1 Å	<—>	Water 537 — 102 N 2.9 Å
Water 545	— 146*OH 3.3 Å	<—>	Water 630 — 146* OH 3.1 Å
Water 617	— 146* O 3.4 Å	<—>	Water 585 — 146* O 3.4 Å
Water 619	— 195 N 3.3 Å		Water 523 — 57 O 3.1 Å
	195 OG 3.5 Å		146* O 3.6 Å
Water 732	— 146*OH 2.2 Å		Water 541 — 146* OT 3.1 Å
			57NE2 3.3 Å
			Water 662 — 195 N 3.0 Å
			195 OG 3.4 Å

<—> Arrows indicate symmetry related water molecules.

* indicate residues from the other monomer.

Table 17: Two-Fold Symmetric Water Molecules in the α -CHT Dimer.^a

Water Molecule 1	Water Molecule 2	Water Molecule 1	Water Molecule 2
508	497	546	588
498	536	548	653
504	522	549	574
505	515	550	660
509	519	625	558
510	533	609	559
531	511	561	578
512	530	663	563
513	517	565	638
514	537	652	567
516	529	632	577
555	518	581	626
520	524	617	585
553	521	586	699
525	568	591	706
592	526	596	631
573	528	650	598
694	534	603	606
535	737	612	736
551	536	710	614
590	538	622	684
539	576	634	661
542	636	658	738
544	579	676	641
545	630		

^aTotal of 49 pairs.

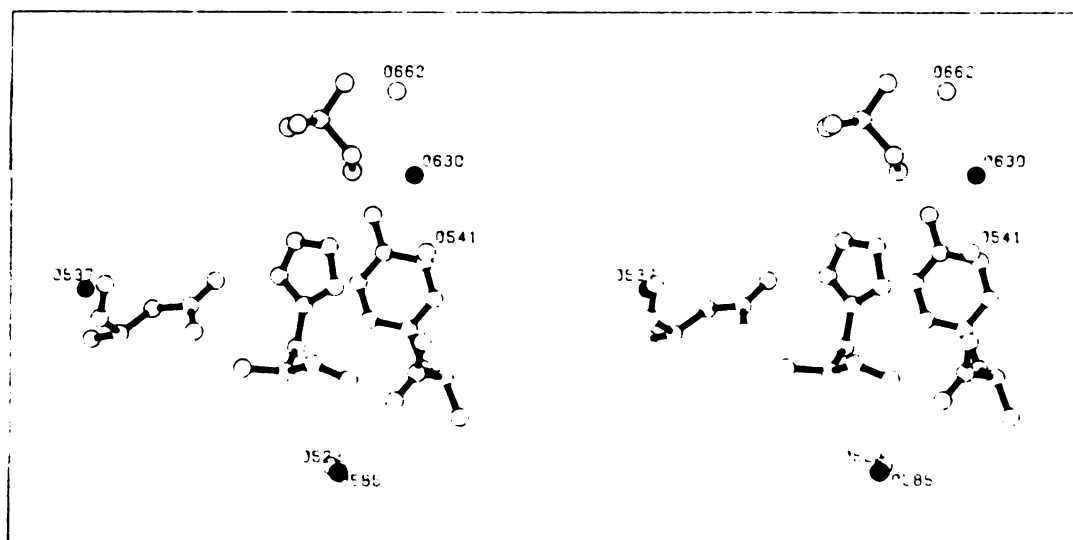
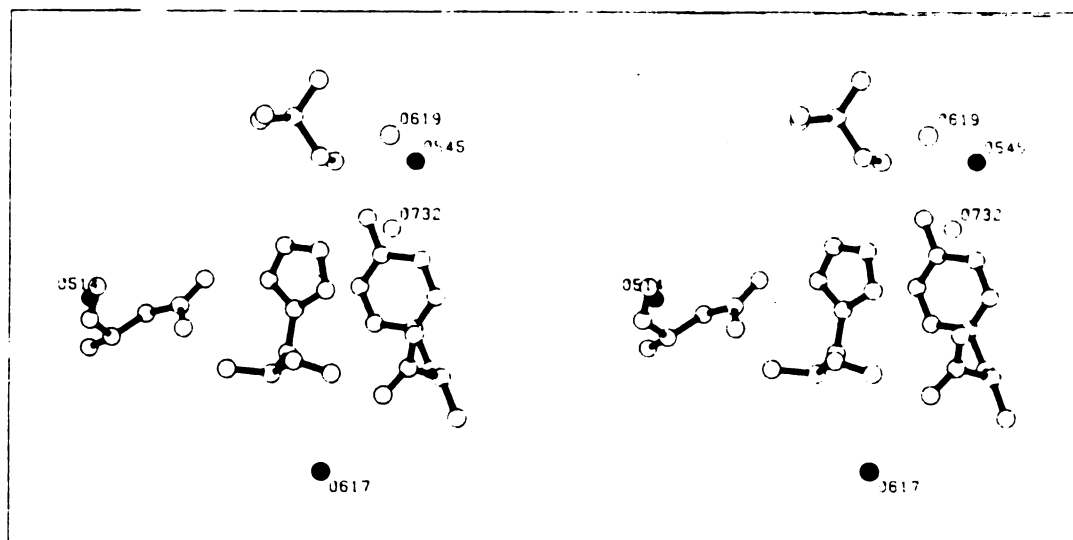
sites, the average thermal parameters are 19.0 and 21.7 Å² while the average occupancies change to 0.89 and 0.83, for molecules 1 and 2, respectively. In all cases, the average thermal factors are lower and the average occupancies are at least as great as the average parameters for the entire solvent. Stereoviews of the active site residues are presented in Figure 23, displaying the solvent molecules in the active site.

E. The ILE-16, ASP-194 Ion Pair

As in γ -CHT¹¹ and in other serine proteases,⁶⁸ there are five water molecules hydrogen bonding in the region of the salt bridge between ILE-16 and ASP-194. Four of the five water molecules (505, 516, 555, 634 and 619 in molecule 1 and 515, 529, 518, 661 and 662 in molecule 2) display symmetry between the two molecules of the α -CHT dimer. Only waters 619 and 662 do not show symmetry within 1.0 Å; however, both appear to interact strongly with ASP-194 N. Additionally, three of these four symmetric water molecules (505, 516, 555 in molecule 1 and 515, 529, 518 in molecule 2) are also found in γ -CHT. It may be that the water molecules help to dissipate the charge of the ILE-16, ASP-194 ion pair. The geometry of the region shows similarity between the two molecules, most of the differences occurring in the side chains. Both ion pairs indicate a hydrogen bond between the N of ILE-16

Figure 23. Stereoview of Active Site Regions of α -CHT. Included are solvent and TYR-146 of the other molecule. Molecule 1, top; molecule 2, bottom. Solvent common to both shaded.

(Top)



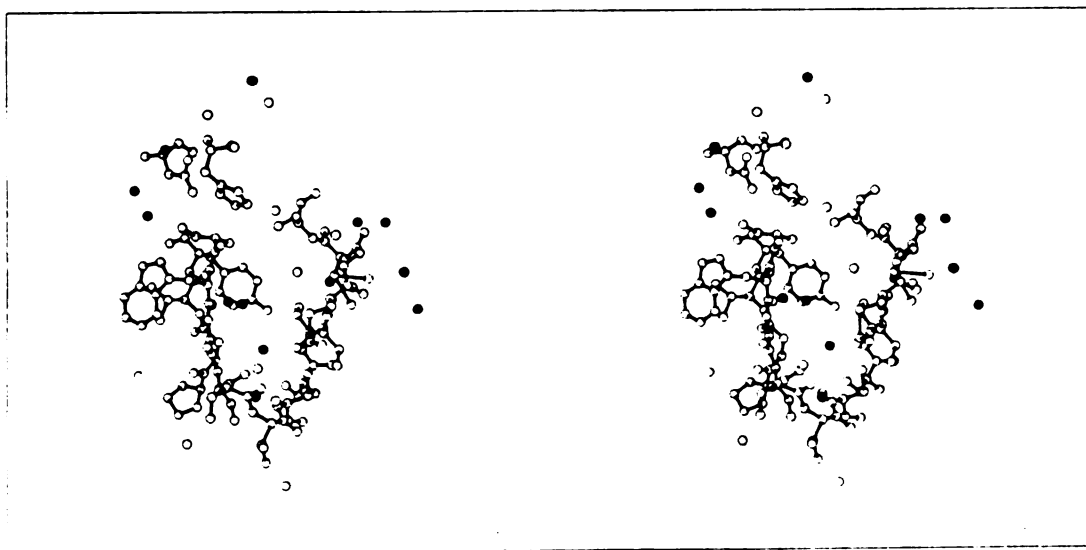
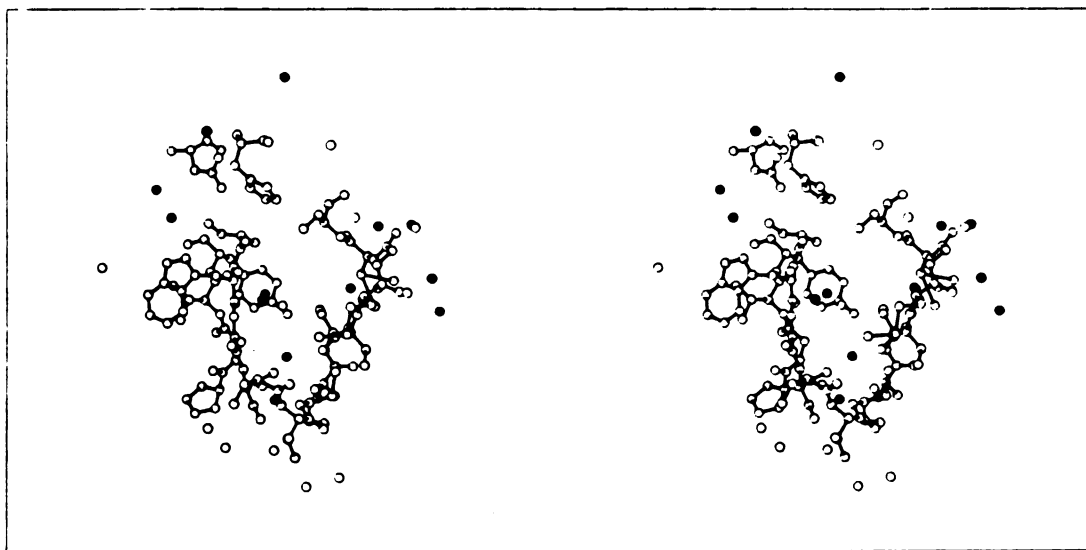
and OD1 of ASP-194, although the donor-acceptor distances seem to be quite small (2.6 and 2.3 Å in molecules 1 and 2, respectively).

F. The Specificity Site

The specificity site is defined by residues 189-195, 214-220 and 225-228. The catalytic triad is also included, located near one end of the site. The specificity site displays good symmetry (r.m.s. $\delta = 0.24$ Å), and also contains several water molecules displaying 2-fold symmetry. Figure 24 presents stereoviews of both specificity sites of the α -CHT dimer including water molecules. One of the symmetric water molecules makes a close contact with TYR-228 OH (2.9 and 3.1 Å) while the other is close to TRP 215 O (2.9 Å) and the main chain of VAL 227. With the transition state analog phenylethane boronic acid bound in the active site, specificity site water molecules are displaced upon binding.⁶⁹ A least squares refinement of this structure⁷⁰ shows that 2-3 of the specificity site water molecules are displaced while the others remain localized at the closed far-end of the specificity site most distant to the catalytic residues HIS-57, ASP-102 and SER-195. The specificity site does not possess any obvious characteristics that may lead to reasons for aromatic specificity. The size of the site is large enough to accommodate large side chains such as

Figure 24. Stereoview of Specificity Site Regions of α -CHT. Included are solvent molecule 1, top; molecule 2, bottom. Symmetric waters shaded.

(Top)



LYS and ARG (as in trypsin). As a consequence, the aromatic specificity of α -CHT may be due in part to the fact that the buried water molecules of the site are not displaced upon substrate binding and in this way aid in the positioning of the substrate for catalysis.

G. The TRP Cluster

The TRP cluster is a cavity 7.0 Å in diameter, containing the residues TRP-27, PRO-28, TRP-29 and TRP-207, with PRO-4 and PRO-8 being slightly above the cavity. In part I, it was shown that the TRP cluster of γ -CHT remained essentially positionally stationary during conformational energy calculations, in isolation, with the crystallographically observed water molecules, and in the presence of bulk solvent. Other reasons for interest in this region are a) there are three other aromatic clusters which have been suggested to lend stability to the protein, b) within the experimental coordinate error, this region displays excellent 2-fold symmetry, comparable to the catalytic site, c) it may serve as a secondary bonding site for aromatic substrate-like molecules,^{14,70} and d) the electron density seems to define the positions of the residues better than any other region in the α -CHT dimer. There are also 14 and 12 water molecules that surround or are within the TRP clusters of molecules 1 and 2, respectively. Again, 8 of these water molecules show symmetry within 1.0 Å between the two

monomers. The others are all within hydrogen bonding distance to some atom of the cluster. The presence of such a large number of solvent molecules may lend some stability to the region; however, the full significance and importance of this region still remains unclear.

H. Side Chain Asymmetry

Some of the surface asymmetry clearly arises from close inter-dimer contacts in the crystal structure as well as from asymmetrical stabilizing interactions in the dimer interface region. Most of the asymmetry probably arises from the flexibility and adaptability of side chains to distribute among equally probable configurations. To examine the question of side chain asymmetry, the side chains in the α -CHT dimer were examined as a function of residue type. The results are summarized in Table 18. In this case however, atoms with thermal factors greater than $20.0 \text{ \AA}^2$ were not included in the comparison. By reducing the thermal factor cut-off, a greater sample of each type of amino acid side chain was used, this being especially true for ASN residues in α -CHT.

The methyl groups of ALA residues show evidence of asymmetry. The only possible cause of such an observation is that there must be main chain difference associated with these residues. The asymmetry indicated for LYS and ARG

Table 18: Asymmetry Classified by Residue Type in the α -CHT Dimer.^{a,b}

Code	Total	Exterior	Interior	Interface
GLY	0.000(22)	0.000(15)	0.000(8)	0.000(3)
ALA	0.327(22)	0.335(16)	0.305(6)	0.000(0)
ARG	0.263(3)	0.263(3)	0.000(0)	0.000(0)
LYS	0.222(14)	0.222(14)	0.000(1)	0.000(1)
ASP	0.126(9)	0.120(7)	0.146(2)	0.152(1)
GLU	0.148(5)	0.148(5)	0.000(0)	0.000(0)
ASN	0.390(13)	0.390(13)	0.000(0)	0.444(2)
GLN	0.407(10)	0.418(9)	0.297(1)	0.000(0)
SER	0.527(26)	0.569(21)	0.293(5)	0.591(8)
THR	0.430(22)	0.448(18)	0.304(5)	0.560(7)
ILE	0.648(10)	0.733(6)	0.495(4)	0.231(1)
LEU	0.671(17)	0.301(6)	0.804(11)	0.175(2)
VAL	0.381(22)	0.445(9)	0.330(13)	0.158(1)
PHE	0.180(6)	0.157(5)	0.267(1)	0.189(2)
TRP	0.266(8)	0.253(4)	0.278(4)	0.185(1)
TYR	0.194(4)	0.213(3)	0.118(1)	0.261(2)
CYS	0.165(8)	0.166(6)	0.160(2)	0.191(3)
HIS	0.207(2)	0.207(2)	0.000(0)	0.207(2)
MET	0.483(2)	0.651(1)	0.209(1)	0.651(1)
PRO	0.352(9)	0.385(6)	0.276(3)	0.000(0)

^aB's < 20.00 Å²; Number of atoms removed is 238, Molecule 1 and 254, Molecule 2.

^bAsymmetry expressed in Å. Numbers in parentheses indicate number of residues.

is misleading, since in these cases, only 2-3 atoms were used in the comparison and in the case of ARG, there are only two residues in each monomer. Interestingly, there is considerably better symmetry displayed by the acid groups ASP and GLU. This was also noted in previous work.¹⁴ The better symmetry displayed by the carboxylic acids is probably due to stabilizing interactions such as ion pair formation with other cations or solvent and hydrogen bond formation. The smaller polar side chains of SER and THR display much better symmetry in the interior of the molecule, possibly due to the higher electron density in the interior. However, these residues show much greater asymmetry in the dimer interface, suggesting that these indications might well be real. The non-polar residues ILE, VAL and LEU show an unexpectedly large asymmetry, despite the fact that most are in the interior of the molecule. The main chain folding in the interior has been shown to possess the greatest symmetry so that the asymmetry displayed by the non-polar residues must be due to the size of the side chains and the free rotations they can effect.

CHAPTER IX

ENERGETIC ANALYSIS

Throughout the least squares refinement of the α -CHT dimer, the REFIN system was used to calculate the global energy of both monomers, as well as that of the dimer.²⁰ Many of the other features included in REFIN were also used to analyze geometry and stereochemistry. Due to the fact that REFIN and PROLSQ employ slightly different representations and dictionaries of ideal geometrical parameters for protein molecules, it was impossible to quantitatively examine the geometrical energy terms, since in most cases, large energetic contributions resulted. Therefore, only the non-bonded and the electrostatic contributions to the global energy were compared. This seems to be a good approximation since PROLSQ indicated from the very beginning that the α -CHT dimer possessed good geometry, at least with respect to the dictionary of ideal values. The r.m.s. delta values from ideal are very small and are listed in Table 10. As in part I, the extended atom plus polar hydrogen atom approximation was used in calculating the energetic contribution. All atoms (except residues 9-13) were included.

The dimerization energy of α -CHT was monitored throughout the refinement. This energy is calculated as

$$\Delta E = E_{\text{Dim}} - (E_1 + E_2) \quad , \quad (17)$$

where E_{Dim} is the energy of the dimer and E_1 and E_2 are the energies of molecules 1 and 2. The change in dimerization energy with refinement is shown graphically in Figure 25. Many large changes occurred in ΔE throughout the refinement, most due to the graphics changes in the structure using FRODO and when the resolution was extended, a large change in ΔE usually occurred. Using only non-bonded contributions (which includes hydrogen bonding), the dimerization energy of the final structure is +44.9 kcal/mole. When electrostatic interactions are included, ΔE increases to +50 kcal/mole. These results can be compared with the observed enthalpy of dimerization of 1-4 kcal at pH 4.1 between 15-20°C.^{71,72}

The separate energies of the monomers calculate extremely well considering the fact that only crystallographic coordinates were used and an energetic term was not included in the least squares refinement. The final energetic terms are listed in Table 19. The final global energy of the α -CHT dimer may be significantly reduced by energy refinement, but with an increase in R-factor.

The surface area buried in the native state of a folded protein has been suggested to be proportional to

Figure 25. Progress of Dimerization Energy during Refinement. Resolution states are indicated.

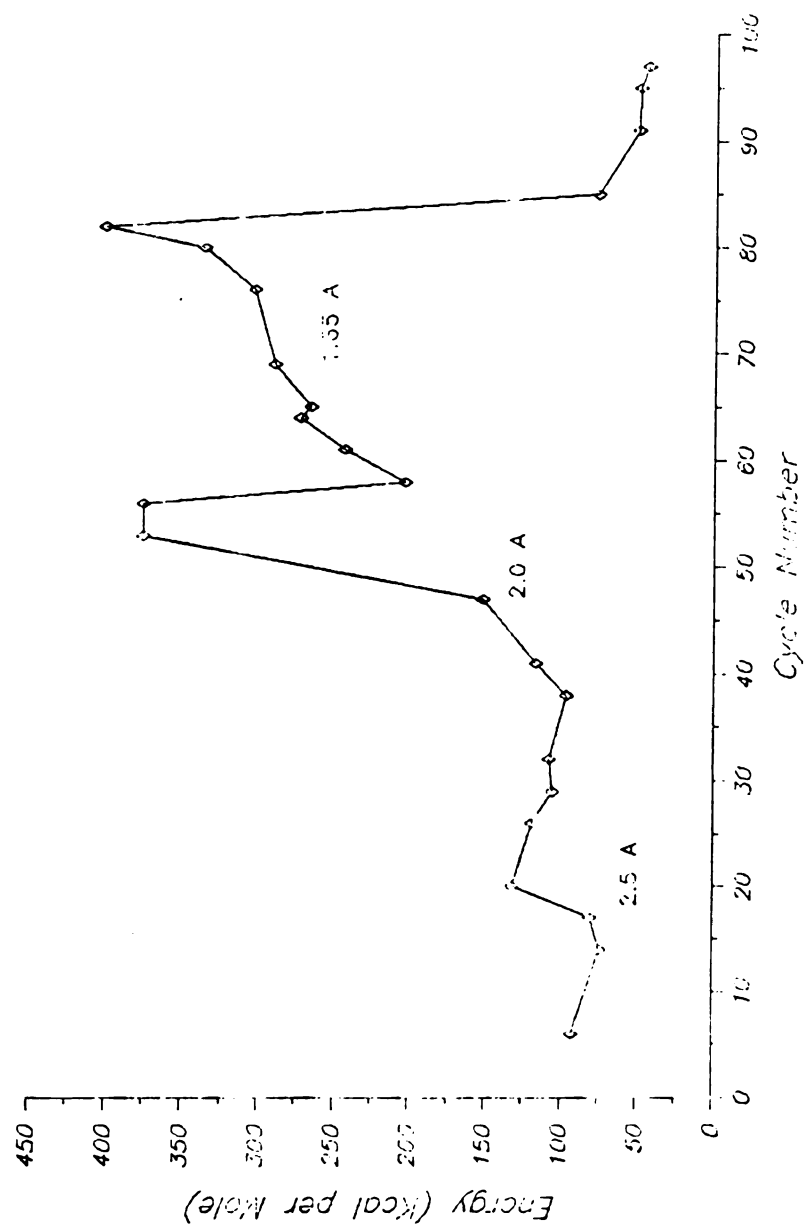


Table 19: Calculated Energies and Surface Areas for the γ -CHT Dimer.

	Surface Area ($\text{\AA}^2$)		Energies (kcal/mole)	
	van der Waals	Total Molecular	van der Waals H-Bonds	Electrostatic Included
Molecule 1	22479.91	5945.06	-1176.1	-1327.8
Molecule 2	22236.54	4090.05	-1122.1	-1265.2
Dimer	44648.66	10218.47	-2253.3	-2542.8
Buried	67.78	183.36		
Hydrophobic Free Energy (kcal/mole)	1.69	4.58		

the gain in free energy of dehydration and hydrogen bond formation due to folding.⁷³ Recently, analytical calculations of buried surface areas have been employed in locating stable domains in proteins as well as predicting folding pathways.³³ The buried surface area of the α -CHT dimer has been calculated and the contribution of hydrophobicity to the stability of the dimer ascertained.

The buried surface area is calculated as the sum of the surface area of the monomers minus the surface area of the dimer. The results of the surface area calculations for α -CHT are listed in Table 19. From studies on hydrocarbons and amino acids,⁴² it has been suggested that 1.0 Å² of surface area corresponds to about 25 cal/mole of hydrophobic free energy. For the α -CHT dimer, the gain in free energy corresponds to about 4.6 Kcal/mole, this value being significantly less (about a factor of 2) than similar results for insulin dimer, trypsin-PII complex and the hemoglobin α - β dimer.^{33,37} Thus, in the case of the α -CHT dimer, other contributions to the free energy of association must have a stronger effect than those present in these other complexes. Possibilities may include van der Waals interactions and hydrogen bonds, complementarity and the loss of translational and rotational entropy.^{33,37}

CHAPTER X

COMPARISON OF THE INDEPENDENT MOLECULES OF α -CHT WITH γ -CHT

The structures of the independent molecules of the dimer of α -CHT at 1.67 Å resolution have been compared with the refined structure of γ -CHT at 1.9 Å resolution. Each molecule of the α -CHT dimer was translated and rotated, in a similar manner in which the independent molecules of the alpha structure were compared to themselves. The final rotation matrices and translation vectors which relate the Cartesian coordinates of the monomers of α - and γ -chymotrypsin are listed in Table 20. As in the comparisons discussed previously, atoms whose thermal factors which were greater than a chosen cut-off in one or the other structure were excluded from the calculations. The thermal factor cut-offs were 23 Å² for α -CHT and 15 Å² for the γ -CHT structure, the average thermal factor for γ -CHT being less than that of α -CHT. A summary of the r.m.s. differences between the structures is given in Table 21, which also indicates that the atoms on the surface and in the dimer interface show the largest thermal factors. In addition, certain residues were not included in the r.m.s. calculations due to the fact that

Table 20: Transformations Relating γ -CHT to α -CHT.

	<u>Matrix Elements</u>			<u>Vector</u>
$\gamma\text{-}\alpha_1$	-.5041	-.3321	-.7972	72.3
	.8585	-.0916	-.5046	16.3
	.0946	-.9388	.3313	11.5
$\gamma\text{-}\alpha_2$	-.4249	-.6935	-.5936	60.5
	-.8533	.0845	.5137	23.4
	-.3010	.7250	-.6195	66.5

Table 21: R.M.S. Differences Between the Independent Molecules of α -CHT and γ -CHT.

	<u>Molecule 1</u>	<u># Atoms Received</u>	<u>Molecule 2</u>	<u># Atoms Received</u>
All Atoms	0.58		0.60	
Main Chain	0.37		0.39	
Carbonyl Oxygens	0.49		0.44	
Side Chain	0.77		0.80	
Sulfurs	0.47		0.35	
Interior				
Main Chain	0.25	(0)	0.26	(0)
Side Chain	0.53	(0)	0.47	(0)
Exterior				
Main Chain	0.41	(51)	0.44	(73)
Side Chain	0.86	(190)	0.93	(202)
Interface				
Main Chain	0.53	(15)	0.52	(17)
Side Chain	0.88	(38)	0.89	(39)
Catalytic Site				
Main Chain	0.29	(0)	0.28	(0)
Side Chain	0.31	(0)	0.52	(0)
TRP Cluster				
Main Chain	0.20	(0)	0.21	(0)
Side Chain	0.25	(0)	0.29	(0)
Domain 1				
Main Chain	0.34	(22)	0.38	(31)
Side Chain	0.78	(81)	0.89	(84)
Domain 2				
Main Chain	0.40	(29)	0.40	(42)
Side Chain	0.76	(109)	0.70	(118)

Summary Table of Deviations by Number of Atoms^a

	<u>Main Chain</u>	<u>Carbonyl Oxygens</u>	<u>Side Chain</u>	<u>Overall</u>
0.00 - 0.25 Å	277,268	66,71	159,131	502,470
0.25 - 0.50 Å	262,219	87,78	233,230	582,528
0.50 - 0.75 Å	68,92	40,36	101,114	209,242
0.75 - 1.00 Å	6,14	8,13	22,26	36,53
1.00 - 1.50 Å	2,1	1,0	22,26	25,27
>1.50 Å	3,2	3,1	38,40	43,43

^aFirst number, molecule 1- γ -CHT, second number, molecule 2- γ -CHT.

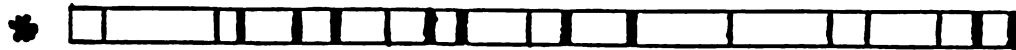
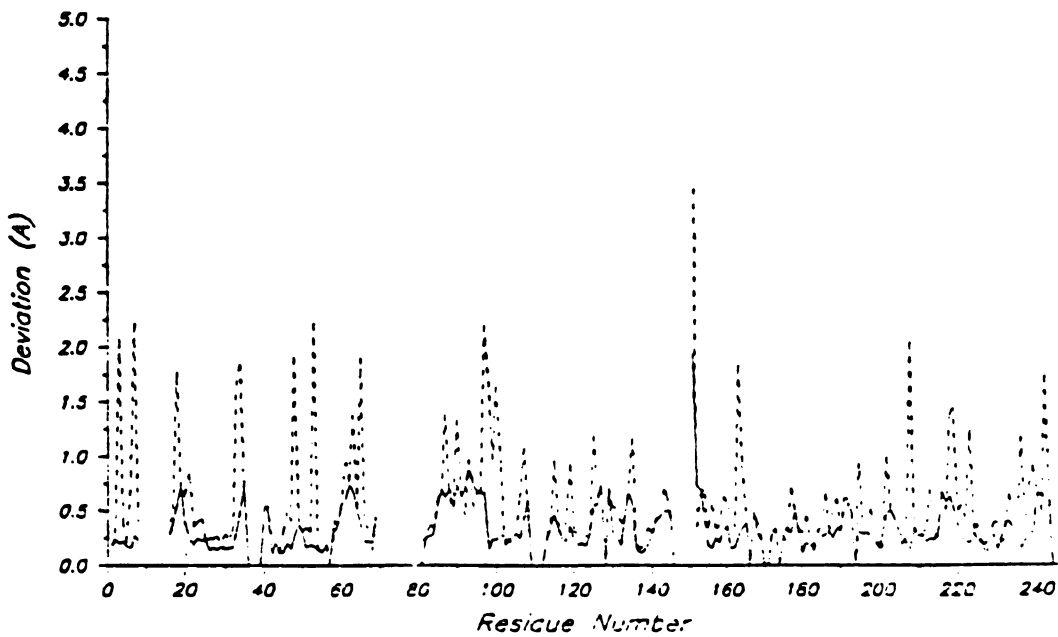
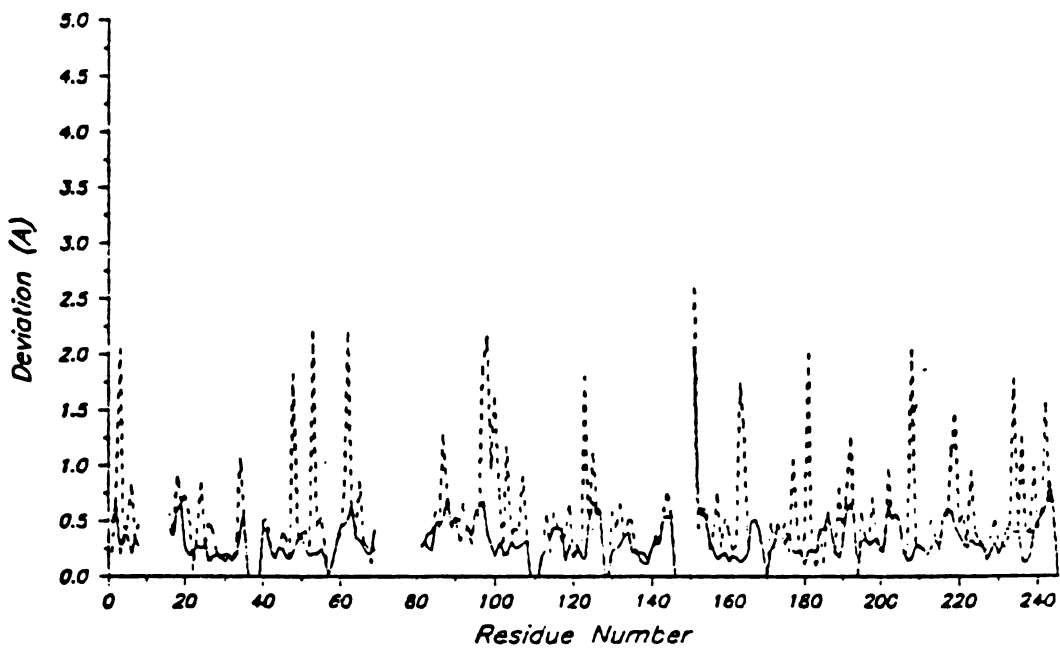
they were disordered in one or the other structures (9-13, 70-80 and 149-150). Examination of Tables 14 and 21 shows that the two monomers of α -CHT are more similar to each other than either is to the structure of γ -CHT. Figure 26 displays the r.m.s. deviations of the individual residues (main and side chains). The main chain is seen to possess good agreement although there are some departures in the regions between residues 80-100.

The largest differences between the structures of α -CHT and γ -CHT occur in the surface residues and in the dimer interface. This is not unexpected since these regions display asymmetry in the dimer molecule. The differences in these regions may be due in part to the differences in intermolecular contacts in the crystal forms (shown in Figure 17), but again, most is probably due to surface side chain adaptability among different positions.

A. The Active Sites

There are regions which show excellent agreement between the two molecules of α -CHT and γ -CHT. The active site residues are very similar, especially ASP-102. The r.m.s. deviations of the catalytic residues are: molecule 1, HIS-57 = 0.31 Å, ASP-102 = 0.17 Å, SER-195 = 0.63 Å and molecule 2, 0.52 Å, 0.20 Å and 0.98 Å, respectively. Superposition of the catalytic triad residues is shown

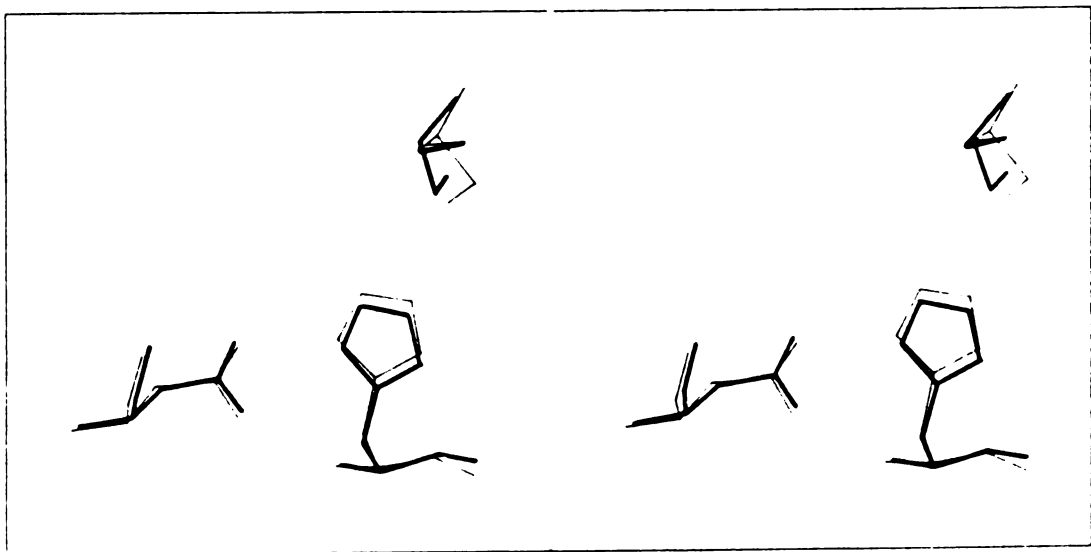
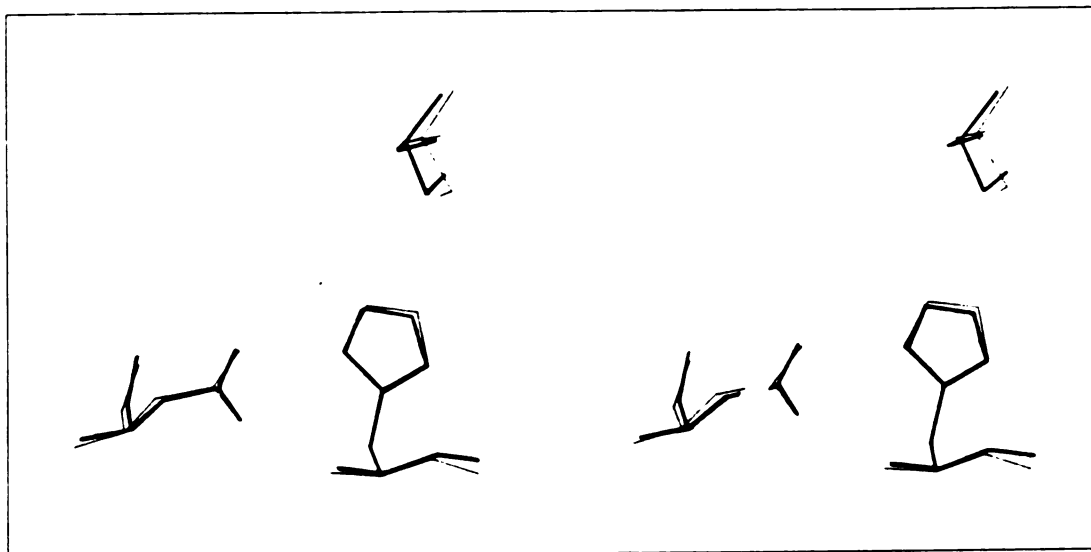
Figure 26. R.M.S. Differences Between α -CHT and γ -CHT. Only atoms with $B < 23.0 \text{ \AA}^2$ (α -CHT) and $< 15.0 \text{ \AA}^2$ (γ -CHT) are included; molecule 1, top; molecule 2, bottom; main chain, solid; side chain, broken; * intermolecular contacts in α - and γ -CHT.



in Figure 27. The imidazole groups of HIS-57 are slightly different between the two molecules; however, the major difference in conformation between the alpha and gamma structures seems to be in the orientation of the OG in SER-195. The position of OG differs by 0.67 Å in molecule 1 and 0.95 Å in molecule 2. The χ -1 angles differ by almost 50° in the alpha and gamma structures and there seems to be no indication of a hydrogen bond between the SER-195 OG and HIS-57 NE2 in either α -CHT or γ -CHT. The difference in conformation between the two alpha catalytic sites and that of γ -CHT might be the result of the difference in pH of the two crystal forms and the possibility that the imidazole might not be protonated in γ -CHT. In the alpha structures, molecules 1 and 2 contain 5 and 6 water molecules that interact strongly with the active site residues. In the case of α -CHT, TYR-146 of the other molecule is in close proximity to the catalytic residues. However, the active site, while showing a strong similarity in the protein positions involved, shows larger deviations with respect to the solvent structure. Only three water molecules are within hydrogen bonding distance to the catalytic triad (331, 390, and 464) in γ -CHT. Of these three, water 381 in γ -CHT is found in both α -CHT monomers (interacting with ASP-102) but water 464 is found only in molecule 2 in α -CHT. This interaction involves a hydrogen bond between the water and both TYR-146 OT and HIS-57 NE2.

Figure 27. Stereoview of Superpositions of the Catalytic Site Regions of α -CHT and γ -CHT. Molecule 1 — γ -CHT, top; molecule 2 — γ -CHT, bottom. γ -CHT, bold in each case.

(Top)



The list of close contacts in the dimer interface region (Figure 16) indicates that there could be an ion pair between these residues in both molecules of α -CHT. The asymmetry and differences between the two alpha structures and that of γ -CHT is shown most clearly here in differences in solvent structure.

B. The Specificity Sites

The specificity site of γ -CHT is also similar with those of α -CHT (r.m.s. deviation = 0.51 Å, averaged over both molecules of α -CHT). One exception are the residues from 216-218, which are in the dimer interface region of α -CHT located very near the local 2-fold axis. Structural changes must occur in this region upon dimerization to remove the very close contacts that would result. Conformational energy calculations reveal an extremely high non-bonded contribution to the global energy of the initial symmetrical dimer which was used as input for the least squares refinement. The final conformation of the region shows large deviations between the two alpha monomers and also with respect to the γ -CHT structure (r.m.s. deviations of 1.1-1.2 Å). The solvent structure also shows that only two of the waters found in the specificity site of γ -CHT are present in the α -CHT monomers.

C. The TRP Clusters

Interestingly, the positions of the atoms found in the TRP cluster in γ -CHT are almost identical to those found in both molecules 1 and 2 of α -CHT. Their positions are most certainly within the error of the two independent structure determinations and high resolution refinements. The solvent structure about the TRP clusters in α -CHT and γ -CHT is similar also. Of the eight water molecules that show symmetry between the two alpha structures, four are also found in the γ -CHT TRP cluster. The stability given by solvent interactions may be a partial explanation of the similarity between the two molecules in this region. The large rigid groups possess a smaller number of energetically favorable positions which may be assumed in the structures.

D. Hydrogen Bonding

Differences in the main chain hydrogen bonding pattern between the monomers of α -CHT and γ -CHT protein are listed in Table 22. A complete list of the main chain hydrogen bonds of γ -CHT may be found elsewhere¹¹ and the complete list for the α -CHT structures may be found in Appendix D. The hydrogen bonds in γ -CHT were chosen according to the following criteria: stereochemically reasonable and the donor-acceptor distance being less than 3.5 Å. In the case of α -CHT, the angle

Table 22: Main Chain Hydrogen Bond Differences with Respect to γ -CHT.

----- Molecule 1 -----

<u>Hydrogen Bonds in α-CHT Only ^a</u>	<u>Hydrogen Bonds in γ-CHT Only</u>
56N-1020	2N-1200
144N-1500 ^b	16N-1430
175N-1720	42N-330
245N-2420	100N-950
	119N-280
	169N-1640
	184N-1610

----- Molecule 2 -----

<u>Hydrogen Bonds in α-CHT Only ^a</u>	<u>Hydrogen Bonds in γ-CHT Only</u>
59N-560	16N-1430
144N-1500 ^b	42N-330
175N-1720	60N-560
245N-2420	121N-460

^aDistance <3.5 Å, angle >150°.

^bASN 150 is disordered in γ -CHT.

(donor-hydrogen-acceptor) was also examined. The donor-acceptor distance was kept less than 3.5 Å and the theta angle greater than 120°. Table 22 indicates a hydrogen bond involving ASN-150 in both molecules of α -CHT. This hydrogen bond is not present in the γ -CHT structure since ASN-150 is disordered. Also, an additional hydrogen bond is found in the terminal alpha-helix (245N-242O) in both α -CHT monomers.

E. Solvent Structure

Since both the structures of α -CHT and γ -CHT have now been refined at a high resolution, the opportunity presents itself to actually compare the solvent structures of the two and more specifically, to examine the differences in the dimer interface region. The amino acid residues that are part of the dimer interface region in α -CHT have already been presented in Appendix B. These residues were also used as the corresponding residues for the interface region comparisons in γ -CHT. In actually choosing which waters are part or near the dimer interface region in both α -CHT and γ -CHT, a combination of two methods were used. First, FRODO was used to examine the structures and the water molecules that seemed to be near the interface region were selected. Second, a nearest-neighbor calculation was performed on the interface residues with all the solvent. In this manner, solvent molecules selected

could be removed if they were initially incorrect and new solvent added if they were missed in the FRODO examination. After all solvent was selected, the rotation matrices and translation vectors used to best fit the two alpha structures with that of γ -CHT (Table 20) were also used to rotate the solvent structure of γ -CHT to the solvent structure surrounding α -CHT.

Table 23 presents a summary of the water molecules in α -CHT molecule 1 and molecule 2 that are within 1.0 Å of the water positions in γ -CHT. Once again, asymmetry in the final α -CHT monomers is apparent. There are 38 solvent molecules in α -CHT molecule 1 that are also found in γ -CHT. The number for the other monomer is 34. These results are enhanced by the fact that in the alpha structures, the average thermal factor of these waters is lower and the average occupancy is larger than the average for the solvent globally (22.3 Å^2 and 0.765). The same can be said for the waters of γ -CHT, although it is very difficult to compare the final thermal factors and occupancies of γ -CHT with those of α -CHT since there seems to be a difference in scaling between the two molecules. The average thermal factor in γ -CHT for waters occurring in α -CHT molecule 1 is about 7.4 Å^2 and 5.6 Å^2 for molecule 2. The corresponding average occupancies are 0.819 and 0.845. These numbers are much lower than the average thermal factor for the γ -CHT solvent (9.9 Å^2). In addition, 25 pairs of water molecules of the 38 and 34

Table 23: Equivalent Water Molecules (within 1.0 Å)
in both α -CHT and γ -CHT.

a.) Molecule 1 of:

α -CHT			γ -CHT			Deviation (Å)
Number	Occupancy	Thermal Factor	Number	Occupancy	Thermal Factor	
498 *	0.60	21.7	324	1.00 DI	3.0	0.624
499	0.65	26.1	383	0.83	14.0	0.707
503	0.53	26.1	254	1.00 DI	26.1	0.941
504 *	1.00	21.9	430	0.82	7.2	0.376
505 *	0.81	16.1	314	0.90	2.0	0.348
508 *	1.00	9.5	311	1.00 DI	3.5	0.573
509 *	1.00	11.0	323	0.84	8.1	0.250
510 *	1.00	12.1	313	0.99	2.0	0.493
512 *	1.00	13.2	309	1.00 DI	2.0	0.244
514 *	0.99	12.6	381	1.00	10.9	0.662
516 *	1.00	12.4	304	1.00 DI	2.0	0.272
520 *	1.00	17.5	302	0.94	7.2	0.063
531 *	0.83	14.3	307	1.00	2.0	0.156
542 *	0.85	14.8	322	0.98 DI	2.0	0.149
544 *	1.00	16.5	337	0.95	13.7	0.275
546 *	0.89	21.3	358	0.66	4.1	0.293
551 *	0.77	20.6	312	1.00	2.5	0.583
553 *	0.98	16.4	365	1.00	11.4	0.532
555 *	1.00	20.7	328	1.00 DI	7.9	0.394
562	0.86	19.5	445	0.50	5.0	0.473
580	0.65	21.9	466	0.30	7.3	0.790
584	0.67	27.0	363	0.47	2.0	0.883
587	0.81	21.3	353	0.76	9.6	0.441
592 *	0.87	26.3	377	0.86	15.8	0.811
593	0.70	17.3	336	0.89	14.7	0.421
604	0.70	16.4	315	0.91	7.9	0.632
608	0.96	29.9	316	1.00	11.6	0.489
609 *	0.80	27.0	366	0.64 DI	2.5	0.732
612 *	0.72	20.5	446	0.45	6.3	0.557
623	0.90	19.9	471	0.53	6.9	0.784
625 *	0.72	23.2	332	0.90	7.0	0.241
632 *	0.76	23.1	335	1.00 DI	5.6	0.646
634 *	0.82	20.4	413	0.80	7.2	0.202
644	0.81	22.4	360	0.85	5.5	0.614
658 *	0.71	25.5	463	0.63	8.0	0.553
674	0.54	20.2	474	0.44	10.2	0.861
694 *	0.74	24.8	419	0.71	2.0	0.810
739	0.61	26.0	425	0.58 DI	15.6	0.821
Average	0.82	19.9		0.82	7.4	0.518

Table 23 Continues.

Table 23 Continued.

b.) Molecule 2 of:

α -CHT			γ -CHT			Deviation (Å)
Number	Occupancy	Thermal Factor	Number	Occupancy	Thermal Factor	
497 *	0.48	17.6	311	1.00 DI	3.5	0.608
511 *	1.00	16.9	307	1.00	2.0	0.166
515 *	0.96	11.4	314	0.90	2.0	0.351
517 *	1.00	14.9	306	1.00	2.0	0.303
518 *	1.00	14.2	328	1.00 DI	7.8	0.536
519 *	1.00	12.7	323	0.84	8.1	0.231
521 *	0.92	17.8	365	1.00	11.4	0.822
524 *	0.91	15.2	302	0.94	7.2	0.243
526 *	1.00	23.2	377	0.86	15.8	0.538
529 *	0.83	13.6	304	1.00	2.0	0.385
530 *	0.97	14.1	309	1.00 DI	2.0	0.078
532	0.87	22.1	301	1.00	5.1	0.562
533 *	1.00	17.5	313	0.99	2.0	0.465
534 *	1.00	14.3	419	0.71	2.0	0.380
536 *	1.00	15.9	312	1.00	2.5	0.610
537 *	0.84	17.6	381	1.00	11.0	0.562
541	0.80	22.1	464	0.28 DI	2.0	0.923
556 *	1.00	24.5	324	1.00 DI	3.0	0.424
557	1.00	17.7	317	0.88	3.0	0.256
558 *	0.68	13.8	332	0.90	7.0	0.350
559 *	1.00	18.4	366	0.64 DI	2.5	0.891
577 *	0.88	16.9	301	1.00	5.0	0.590
579 *	0.64	14.4	337	0.95	13.7	0.366
583	1.00	16.1	371	0.74	3.4	0.220
588 *	0.77	18.2	358	0.66	4.1	0.730
600	0.84	24.7	424	1.00	9.7	0.712
618	0.82	23.9	475	0.46	3.0	0.612
636 *	0.79	25.0	322	0.98 DI	2.0	0.509
637	0.70	23.5	452	0.51	2.8	0.885
661 *	0.63	18.2	413	0.80	7.2	0.164
665	0.86	26.1	315	0.91	7.9	0.510
700	0.74	25.3	329	0.72 DI	15.5	0.877
736 *	0.69	26.0	446	0.45	6.3	0.546
738 *	0.64	26.0	463	0.63	8.0	0.749
Average	0.86	18.8		0.85	5.7	0.504

* — Symmetric Water Molecules in α -CHT.

DI — Dimer Interface Waters.

waters in the α -CHT monomers found in γ -CHT show 2-fold symmetry. This emphasizes that these may well be the best determined solvent molecules in the α -CHT structure.

A nearest-neighbor calculation performed on the final structure of γ -CHT along with an examination using FRODO showed that there were 33 solvent molecules in γ -CHT located near the corresponding residues of the dimer interface region of α -CHT. Using these waters, examination of the solvent occurring in both α -CHT and γ -CHT revealed which water molecules are excluded in dimerization, which are simply displaced by a certain distance and which do not change position during the process of dimerization. Of the 33 solvents given above, the change in position of the corresponding α -CHT waters was anywhere from 0.1 Å to over 4.0 Å. It is very difficult therefore to decide from this type of distribution of distances which solvent molecules may be displaced. It was decided that any solvent molecules within 1.0 Å of each other in both the α -CHT and γ -CHT structures would be considered the same. With this assumption, the following results emerge. Of the 33 waters in the γ -CHT interface, 12 (36%) are also found in the final structure of α -CHT. Of the remaining 21 water molecules, 11 (34%) are simply displaced by less than 2.8 Å (r.m.s. = 2.04 Å) and the remaining 10 (30%) are lost during the process of dimerization. In addition to the asymmetrical structural changes in the dimer interface side chain atoms that must

occur upon dimerization, the above results indicate that changes in the solvent structure in this region must also be important in stabilizing the dimerization process.

F. Concluding Remarks

The least squares refinement of α -CHT has focused on two molecules per asymmetric unit. The basic results that have emerged as a result of this work may be applied to structures containing more than two molecules per asymmetric unit. The folding of the main chain is the same within experimental error but this does not apply generally to the side chain stereochemistry. The deviations in the side chains may be due in part from inter-dimer contacts in the crystal but most of the differences are probably a result of the high adaptability associated with rotational degrees of freedom in the tertiary surface structure. The results of the refinement of the α -CHT dimer clearly show that the folding of a protein molecule is basically independent of most of the detailed stereochemistry of the side chain atoms. Since a large number of protein structures are being studied by averaging about an appropriate symmetry element, the results of this work indicate that care must be exercised in analyzing the side chain configurations. Ideally, the structure should be unaveraged for correct interpretation, especially near surface and inter-subunit regions.

LIST OF REFERENCES

LIST OF REFERENCES

1. F.D. Gibson and H.A. Scheraga, Proc. Nat'l. Acad. Sci. USA, 58, 420-428 (1967).
2. M. Levitt and S. Lifson, J. Mol. Biol., 46, 269-279 (1969).
3. M. Levitt, J. Mol. Biol., 104 59-107 (1976).
4. J.A. McCammon, B.R. Gelin, M. Karplus and P.G. Wolynes, Nature, 262, 325-326 (1976).
5. B.R. Gelin and M. Karplus, Biochemistry, 18, 1256-1268 (1979).
6. S. Harvey and J.A. McCammon, Comp. and Chem., 6, 173-179 (1982).
7. J. Hermans and M. Vacatello, Water in Polymers, American Chemical Society, pp. 199-214 (1980).
8. S. Fitzwater and H.A. Scheraga, Proc. Nat'l. Acad. Sci. USA, 79, 2133-2137 (1982).
9. A. Jack and M. Levitt, Acta Cryst., A34, 931-935 (1978).
10. B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan and M. Karplus, J. Comp. Chem., 4, 187-217 (1983).
11. C.H. Cohen, E.W. Silverton and D.R. Davies, J. Mol. Biol., 148, 449-479 (1981).
12. R.A. Blevins and A. Tulinsky, J. Biol. Chem., in press.
13. N.V. Raghavan and A. Tulinsky, Acta Cryst., B35, 1776-1785 (1979).
14. A. Tulinsky, R.L. Vandlen, C.N. Morimoto, N.V. Mani and L.H. Wright, Biochemistry, 12, 4185-4192 (1973).
15. L.D. Weber, A. Tulinsky, J.D. Johnson and M.A. El-Bayoumi, Biochemistry, 18, 1297-1303 (1979).

16. D.R. Ferro and W.C.J. Hol, Energy Minimizations of Proteins. An Attempt to Explicitly Account for Water-Protein Interactions, CECAM Workshop Project (1980).
17. A. Tulinsky, N.V. Mani, C.N. Morimoto and R.L. Vandlen, *Acta Cryst.*, B29, 1309-1322 (1973).
18. J.A. McCammon, P.C. Wolynes and M. Karplus, *Biochemistry*, 18, 927-942 (1979).
19. D.R. Ferro, J.E. McQueen Jr., J.T. McCown and J. Hermans, *J. Mol. Biol.*, 136, 1-18 (1980).
20. J. Hermans and J.E. McQueen Jr., *Acta Cryst.*, A30, 730-739 (1974).
21. U. Burkert and N.L. Allinger, *Molecular Mechanics*, American Chemical Society Monograph 177, pp. 59-71, (1982).
22. M. Levitt, *J. Mol. Biol.*, 82, 393-420 (1974).
23. M. Levitt, *Protein Folding*, ed. R. Jaenicke, pp. 17-39, Elsevier/North-Holland Biomedical Press (1980).
24. P.K. Warne and H.A. Scheraga, *Biochemistry*, 13, 757-767 (1974).
25. P.K. Weiner and P.A. Kollman, *J. Comp. Chem.*, 2, 287-303 (1981).
26. G. Nemethy, W.J. Peer and H.A. Scheraga, *Ann. Rev. Biophys. Bioeng.*, 10, 459-497 (1981).
27. E.B. Wilson, J.C. Decius and P.C. Cross, *Molecular Vibrations*, Dover, NY, p. 286 (1980).
28. W.F. van Gunsteren, H.J.C. Berendsen, J. Hermans, W.C.J. Hol and J.P.M. Postma, *Proc. Nat'l. Acad. Sci. USA*, 80, 4315-4319 (1983).
29. W.F. van Gunsteren and M. Karplus, *J. Comp. Chem.*, 1, 266-274 (1980).
30. M. Carson (private communication).
31. J.P. Glusker and K.N. Trueblood, *Crystal Structure Analysis: A Primer*, Oxford University Press, New York (1972).

32. R.A. Blevins, M. Ragazzi and P.M. Hunt (unpublished results).
33. C. Chothia and J. Janin, *Nature*, 256, 705-708 (1975).
34. B. Lee and F.M. Richards, *J. Mol. Biol.*, 55, 379-400 (1971).
35. L. Pauling and D. Pressman, *J. Am. Chem. Soc.*, 67, 1003-1012 (1945).
36. W. Kauzmann, *Adv. Protein Chem.*, 14, 1-63 (1959).
37. J. Janin and C. Chothia, *Biochemistry*, 17, 2943-2948 (1978).
38. B. Lee, *Proc. Nat'l. Acad. Sci. USA*, 80, 622-626 (1983).
39. M. Connolly, *QCPE Bull.*, 1, 75 (1981).
40. C. Chothia, *Ann. Rev. Biochem.*, 53, 537-572 (1984).
41. S.J. Wodak and J. Janin, *Biochemistry*, 20, 6544-6552 (1981).
42. C. Chothia, *J. Mol. Biol.*, 105, 1-14 (1976).
43. C. Tanford, *Proc. Nat'l. Acad. Sci. USA*, 9, 4175-4176 (1979).
44. P.D. Ross and S. Subramanian, *Biochemistry*, 11, 3096-3102 (1981).
45. L.H. Jensen, *Ann. Rev. Biophys. Bioeng.*, 3, 81-92 (1974).
46. A. Tulinsky, Methods of Enzymology. Diffraction Methods in Molecular Biology., in press.
47. R. Diamond, *Acta Cryst.*, A27, 436-452 (1971).
48. K.D. Watenpaugh, L.C. Sieker, J.R. Herriot and L.H. Jensen, *Acta Cryst.*, B29, 943-956 (1973).
49. J. Deisenhofer and W. Steigemann, *Acta Cryst.*, B31, 238-250 (1975).
50. N. Isaacs, Structural Studies of Molecules of Biological Interest, eds. G. Dodson, J.P. Clusker and D. Sayre, Clarendon Press, Oxford, 1981, pp. 274-287.

51. C.J. De Ranter, X-ray Crsytallography and Drug Action, eds. A.S. Horn and C.J. De Ranter, Oxford Science Publications, 1984, pp. 1-22.
52. R.C. Agarwal, Acta Cryst., A34, 791-809 (1978).
53. L.G. Hoard and C.E. Nordman, Acta Cryst., A35, 1010-1015 (1979).
54. J.L. Sussman, S.R. Holbrook, G.M. Church and S.H. Kim, Acta Cryst., A33, 800-804 (1977).
55. R.M. Burnett and C.E. Nordman, J. Appl. Cryst., 7, 625-627 (1974).
56. R.A. Blevins, C.E. Buck and A. Tulinsky (unpublished results).
57. W.A. Hendrickson and J.H. Konnert, Computing in Crystallography, eds. R. Diamond, S. Ramaseshan, K. Venkatesan, Indian Acad. Sci., Bangalore, pp. 13.01-13.26 (1980).
58. T.A. Jones, Compututational Crystallography, ed. D. Sayre, Clarendon Press, Oxford, pp. 303-317 (1982).
59. B. Bush, Comp. and Chem., 8, 1-11 (1984).
60. A. Mavridis, A. Tulinsky and M.N. Liebman, Biochemistry, 13, 3661-3666 (1974).
61. J.J. Birktoft and D.M. Blow, J. Mol. Biol., 68, 187-240 (1972).
62. V. Luzzati, Acta Cryst., 5, 802-810 (1952).
63. A.R. Sielecki, W.A. Hendrickson, C.G. Broughton, L.T.J. Delbaere, G.D. Brayer and K.N.C. James, J. Mol. Biol., 134, 781-804 (1979).
64. M.N.G. James and A.R. Sielecki, J. Mol. Biol., 163, 299-361 (1983).
65. Protein Data Bank, (1984) Brookhaven National Laboratory, Upton, NY.
66. G.M. Ramachandran, C. Ramakrishnam, and V. Sasisekharan, J. Mol. Biol., 7, 95-99 (1963).
67. J. Janin, S. Wodak, M. Levitt and B. Maigret, J. Mol. Biol., 125, 357-396 (1978).

68. D.A. Matthews, R.A. Alden, J.J. Birktoft, S.T. Freer and J. Kraut, J. Biol. Chem., 252, 8875-8883 (1977).
69. A. Tulinsky, I. Mavridis and R.F. Mann, J. Biol. Chem., 253, 1074-1078 (1971).
70. B.A. Karcher and A. Tulinsky (unpublished results).
71. K.C. Aune and S.N. Timasheff, Biochemistry, 10, 1609-1617 (1971).
72. K.C. Aune, L.C. Goldsmith and S.N. Timasheff, Biochemistry, 10, 1617-1622 (1971).
73. A.R. Rashin, Biopolymers, 23, 1605-1620 (1984).
74. IMSL, International Mathematical and Statistical Library, Inc., Subroutine CGSPH, Edition 9, 1982.

APPENDICES

APPENDIX A

Appendix A: Extended Atoms and their Non-Bonded Parameters.

Atom	α	N_{eff}	r_{vdw}	Groups Represented
O	0.84	6	1.60	Carbonyl Oxygen, Water Oxygen
OH	1.20	7	1.70	Alcoholic Hydroxyl
OM	2.14	6	1.60	Carboxyl Oxygen
NH	1.40	7	1.65	Peptidic Nitrogen
N(2)	1.70	8	1.70	-NH ₂ Terminals
N(3)	2.13	9	1.75	-NH ₃ ⁺ Terminals
OH	1.35	6	1.85	Aliphatic-CH
C(2)	1.77	7	1.90	Aliphatic-CH ₂
C(3)	2.17	8	1.95	Methyl Terminal
C	1.65	5	1.80	Aromatic/Carbonyl Carbon
CR	2.07	6	1.90	Aromatic-CH
S	0.34	16	1.90	Sulfur (Cys, Met)

Hydrogen Bond Potentials

Bond Type	E_{min}^{HB} (kcal/mole)	R_{min}^{HB} (Å)
OH-O	-3.5	2.80
OH-OH	-3.5	2.75
OH-OM	-3.5	2.85
NH-O	-3.0	2.95
NH-OH	-3.0	3.08
NH-OM	-2.5	3.10
N(2)-O	-2.5	2.87
N(2)-OH	-2.5	2.87
N(2)-OM	-2.5	2.87
N(3)-OH	-2.5	3.00

$$A^{NB} = \frac{1}{2} C(r_i + r_j)^6 \quad E_{MIN}^{HB} = -.067 (C^{HB})^6 / A^{HB5}$$

$$C^{NB} = (3e\hbar/2\sqrt{m}) \alpha_i \alpha_j / [(\alpha_i/N_i)^{\frac{1}{2}} + (\alpha_j/N_j)^{\frac{1}{2}}]$$

$$R_{MIN}^{HB} = (1.2 A^{HB} / C^{HB})^{\frac{1}{2}}$$

See Reference 10 and Equations 2 and 3.

APPENDIX B

Appendix B: Residues found in Specific CHT Regions.

	Interior	Exterior	Interface
ALA	22, 55, 56, 183, 185, 229	5, 68, 86, 111, 112, 120, 126, 131, 132, 149, 158, 179, 206, 233, 243, 244	149
ARG		145, 154, 230	
ASN		18, 48, 50, 91, 95, 100, 101, 150, 165, 167, 204, 236, 245	95, 150
ASP	194	35, 64, 72, 128, 129, 153, 178, 102	102, 194
CYS	182, 168	1, 42, 58, 122, 136, 191, 201, 220	58, 191, 220
GLN	30	7, 34, 73, 81, 116, 156, 157, 239, 240	
GLU		20, 21, 49, 70, 78	
HIS		40, 57	40, 57
ILE	16, 47, 103, 212	6, 80, 85, 99, 176, 181	99
LEU	33, 46, 105, 106, 108, 123, 155, 160, 199, 209, 234	83, 97, 163, 242, 143, 162	97, 143

Appendix B Continues.

Appendix B Continued.

	Interior	Exterior	Interface
LYS		36, 79, 82, 84, 87, 90, 93, 107, 169, 170, 175, 177, 202, 203	36
MET	180	192	192
PHE	41	39, 71, 89, 114, 130	39, 41
PRO	28, 124, 198	4, 8, 24, 152, 161, 225	
SER	32, 45, 189, 214	26, 63, 75, 76, 77, 92, 96, 109, 113, 115, 119, 190, 125, 127, 159, 164, 186, 195, 217, 218, 221, 223	63, 96, 190, 195, 217, 218, 211, 223
THR	54, 104, 138, 139	37, 61, 62, 98, 110, 117, 134, 135, 144, 151, 166, 174, 208, 219, 222, 224, 232, 241	37, 61, 62, 98, 151, 219, 222
TRP	27, 29, 141, 215	51, 172, 207, 237	215
TYR	228	94, 146, 171	94, 146
VAL	17, 31, 52, 53, 60, 66, 121, 200, 210, 213, 227, 231, 238	3, 23, 65, 67, 88, 118, 137, 188, 235	60
GLY	43, 44, 140, 142, 184, 197, 211	2, 19, 25, 38, 59, 69, 74, 133, 173, 187, 196, 205, 216, 226	38, 58, 216

APPENDIX C

Appendix C: Variable Dihedrals in the α -CHT Dimer

A.) Molecule 1

Residue	Phi	Psi	Omega	χ -1	χ -2	χ -3	χ -4	χ -5
C 1		-7	-177	139	-64	172	-54	
G 2	-170	38	-178					
V 3	-123	94	180	-178				
P 4	-61	140	172	30	-36	25		
A 5	-67	-36	174					
I 6	-102	103	-177	-66	143			
Q 7	-67	128	178	-47	134	71		
P 8	-56	150		-31	42	-37		
I 16		134	177	-54	169			
V 17	-92	129	177	-174				
N 18	74	24	-180	-112	-103			
G 19	-96	-167	-180					
E 20	-149	163	-178	24	89	168		
E 21	-65	135	177	-164	176	91		
A 22	-81	163	173					
V 23	-75	127	-178	-177				
P 24	-60	143	-180	-2	3	-3		
G 25	84	-15	-180					
S 26	-72	-13	178	81				
W 27	-133	72	-171	-64	110			
P 28	-71	-11	177	30	-38	30		
W 29	-89	-14	-178	33	95			
Q 30	-66	127	177	165	91	62		
V 31	-121	160	177	-68				
S 32	-118	133	176	158				
L 33	-100	128	179	-58	169			
Q 34	-119	136	-177	-70	-177	2		
D 35	-81	-172	180	67	-158			
K 36	-72	-42	179	-116	-51	-178	-123	
T 37	-45	-44	-179	117				
G 38	117	36	174					
F 39	-141	160	-173	-170	64			
H 40	-46	120	176	179	86			
F 41	-125	-11	-180	64	100			
C 42	-168	166	176	-106	-140	-86	-92	
G 43	-106	-176	176					
G 44	-166	179	179					
S 45	-136	142	178	-53				
L 46	-81	131	176	-82	157			
I 47	-108	-19	-179	62	167			
N 48	-153	174	175	-77	165			
E 49	-75	-5	-179	-89	-157	-20		
N 50	-123	-7	-171	-92	-77			
W 51	-143	145	180	-65	81			

Residue	Phi	Psi	Omega	X-1	X-2	X-3	X-4	X-5
V 52	-127	135	173	169				
V 53	-101	139	174	73				
T 54	-148	-177	179	-172				
A 55	-83	145	175					
A 56	-63	-43	-178					
H 57	-59	-17	-179	83	-100			
C 58	-63	-15	-177	-69	-92	-86	-140	
G 59	60	36	179					
V 60	-62	137	-178	-165				
T 61	-127	174	-179	76				
T 62	-69	-8	180	-88				
S 63	-97	-10	177	-43				
D 64	-70	161	-179	-63	136			
V 65	-124	147	177	73				
V 66	-99	132	174	168				
V 67	-112	119	178	-168				
A 68	-116	148	179					
G 69	87	8	-178					
E 70	-83	133	178	9	80	-114		
F 71	-111	-59	-178	178	52			
D 72	-103	141	-179	-163	144			
Q 73	-96	2	-179	-51	-167	-71		
G 74	-112	8	-180					
S 75	-65	134	179	-60				
S 76	-55	-26	179	-62				
S 77	177	135	180	178				
E 78	-136	169	-179	-158	-178	-137		
K 79	-92	98	180	-156	169	-168	174	
I 80	-140	-171	179	75	-67			
Q 81	-129	111	177	-51	-170	45		
K 82	-94	100	-178	-62	129	140	169	
L 83	-98	131	178	-43	159			
K 84	-76	152	179	-52	-135	179	136	
I 85	-97	127	179	-64	168			
A 86	-94	-42	-180					
K 87	-155	149	-179	-163	160	156	-165	
V 88	-107	138	180	158				
F 89	-121	95	-177	-77	83			
K 90	-82	148	178	-38	52	142	-154	
N 91	-67	126	179	176	-89			
S 92	-60	-20	179	-11				
K 93	-78	-14	-178	-73	-28	103	179	
Y 94	-49	132	-176	178	59			
N 95	-119	111	-179	-170	-15			
S 96	-70	-5	178	61				
L 97	-97	-33	-174	-42	139			
T 98	-98	-21	-179	-96				
I 99	60	19	-178	-43	-54			

Residue	Phi	Psi	Omega	χ -1	χ -2	χ -3	χ -4	χ -5
N100	-79	150	179	-160	26			
N101	64	47	-175	-52	-45			
D102	-80	64	179	-167	-157			
I103	-138	155	179	-172	-47			
T104	-149	139	175	-80				
L105	-105	140	-177	-69	180			
L106	-122	118	179	-59	-173			
K107	-100	115	-179	160	100	-152	89	
L108	-75	146	179	-60	-172			
S109	-93	0	-178	105				
T110	-155	108	-178	-52				
A111	-53	152	177					
A112	-74	162	-179					
S113	-116	90	179	-158				
F114	-54	155	-174	-76	18			
S115	-173	-161	-178	72				
Q116	-76	-19	-179	-8	36	-1		
T117	-104	-16	179	61				
V118	-125	120	-179	-179				
S119	-164	-180	-177	70				
A120	-100	149	175					
V121	-87	158	177	66				
C122	-85	154	178	-70	-54	172	-64	
L123	-103	147	175	-130	15			
P124	-82	179	173	35	-32	16		
S125	-99	159	176	-69				
A126	-68	-25	-179					
S127	-92	2	-176	54				
D128	-73	150	180	5	17			
D129	-109	134	178	178	66			
F130	-130	102	-177	-73	109			
A131	-70	141	175					
A132	-49	129	179					
G133	109	-18	-179					
T134	-61	140	179	-39				
T135	-89	109	-180	-63				
C136	-115	-172	-173	-55	-137	99	-89	
V137	-138	146	176	157				
T138	-132	147	-179	-173				
T139	-137	156	176	69				
G140	170	178	179					
W141	-110	21	178	-75	-82			
G142	-71	172	-179					
L143	-52	144	179	-43	174			
T144	-91	-7	179	51				
R145	-158	142	179	-49	135	112	-125	173
Y146	-55	-42		179	60			
A149		162	-179					

Residue	Phi	Psi	Omega	X-1	X-2	X-3	X-4	X-5
N150	-104	14	175	-76	-23			
T151	-99	138	-179	73				
P152	-81	135	178	39	-44	31		
D153	-85	-29	179	-70	152			
R154	-95	140	179	-74	158	-114	-34	3
L155	-64	129	179	-170	74			
Q156	-105	148	-179	-51	-71	-49		
Q157	-139	165	174	65	-165	-28		
A158	-167	144	173					
S159	-93	144	-180	-69				
L160	-158	147	179	75	167			
P161	-89	145	178	43	-37	20		
L162	-77	154	179	-66	-171			
L163	-122	158	172	-78	72			
S164	-81	154	179	-26				
N165	-60	-29	179	-91	-154			
T166	-55	-54	-179	-90				
N167	-66	-42	-179	-51	81			
C168	-65	-31	180	-164	167	-80	-166	
K169	-70	-26	176	-53	-153	171	96	
K170	-51	-39	-180	-84	-174	112	-133	
Y171	-93	-57	-180	-47	-101			
W172	-88	-8	-180	-63	113			
G173	55	-146	-176					
T174	-56	-16	-180	53				
K175	-71	-15	-177	-60	-146	-66	-48	
I176	-91	116	179	-54	-69			
K177	-102	165	-179	-81	150	174	57	
D178	-56	-31	-178	-108	50			
A179	-94	26	179					
M180	-126	151	179	-59	-179	-180		
I181	-139	121	173	139	-135			
C182	-97	150	-180	-51	-166	-80	167	
A183	-148	150	179					
G184	102	-133	-178					
A185	63	18	-174					
S186	-110	8	-178	53				
G187	110	15	-180					
V188	-149	162	-176	-62				
S189	-166	142	177	151				
S190	-72	156	180	-65				
C191	-147	173	177	-155	41	98	-168	
M192	-39	129	180	-80	105	-157		
G193	101	-15	180					
D194	-86	-11	177	-69	140			
S195	-47	138	179	-84				
G196	82	-15	-177					
G197	-75	177	-180					

Residue	Phi	Psi	Omega	χ -1	χ -2	χ -3	χ -4	χ -5
P198	-88	157	170	33	-35	23		
L199	-129	106	180	170	68			
V200	-115	152	-179	-64				
C201	-140	143	-176	-43	-89	99	-137	
K202	-85	122	179	-132	179	160	-159	
K203	-135	128	178	-172	-159	-134	-167	
N204	58	33	178	-65	-16			
G205	73	15	178					
A206	-129	153	177					
W207	-83	130	-180	-64	93			
T208	-119	138	-179	-48				
L209	-76	117	-177	-173	70			
V210	-107	-29	-176	166				
G211	-141	158	171					
I212	-122	124	180	-52	156			
V213	-53	122	-173	-174				
S214	-119	-62	178	-169				
W215	-157	-179	-180	53	-89			
G216	173	-160	-176					
S217	-50	126	-176	162				
S218	-65	-8	-178	36				
T219	-125	4	175	-86				
C220	63	31	180	-60	-168	98	41	
S221	-57	141	-179	168				
T222	-86	-3	178	64				
S223	-94	-8	-171	-92				
T224	-122	140	179	-69				
P225	-80	145	174	19	-13	1		
G226	-80	148	-178					
V227	-114	132	-177	-180				
Y228	-126	152	178	-54	78			
A229	-74	131	-177					
R230	-79	105	-177	-165	179	-76	-90	-176
V231	-59	-40	177	174				
T232	-54	-33	179	-114				
A233	-84	-12	-178					
L234	-113	-6	-173	-78	13			
V235	-68	-40	178	85				
N236	-51	-40	-179	-71	-21			
W237	-65	-45	-180	173	87			
V238	-55	-62	-178	170				
Q239	-51	-41	-179	-60	-55	-9		
Q240	-69	-39	176	-70	-171	-101		
T241	-63	-47	-179	-37				
L242	-57	-46	-180	-84	179			
A243	-58	-36	-178					
A244	-96	-5	-176					
N245	-125	157		-77	-69			

B.) Molecule 2

Residue	Phi	Psi	Omega	X-1	X-2	X-3	X-4	X-5
C 1		157	174	67	72	108	-70	
G 2	71	20	-176					
V 3	-139	95	-179	-179				
P 4	-67	158	178	-5	22	-30		
A 5	-71	-39	177					
I 6	-105	111	-178	-59	150			
Q 7	-66	129	175	62	-59	149		
P 8	-51	141		-22	29	-25		
I 16		132	178	-48	-175			
V 17	-98	125	178	173				
N 18	73	28	179	-159	42			
G 19	-87	-169	180					
E 20	-137	158	177	-49	78	-132		
E 21	-76	132	-179	-171	171	177		
A 22	-80	168	173					
V 23	-80	132	-179	166				
P 24	-51	131	177	-11	17	-17		
G 25	77	1	-177					
S 26	-79	-9	177	55				
W 27	-129	74	-178	-57	110			
P 28	-62	-20	178	34	-48	43		
W 29	-89	-12	-179	43	84			
Q 30	-68	124	176	168	92	41		
V 31	-119	163	178	-62				
S 32	-122	139	175	143				
L 33	-98	131	178	-62	49			
Q 34	-115	150	176	-100	168	34		
D 35	-93	-177	-177	39	91			
K 36	-86	13	178	-112	4	63	16	
T 37	-94	-8	178	81				
G 38	97	11	179					
F 39	-138	171	-173	174	114			
H 40	-55	119	177	-159	33			
F 41	-122	-10	-179	64	88			
C 42	-172	163	176	-96	-149	-91	-91	
G 43	-97	173	176					
G 44	-164	166	179					
S 45	-126	132	177	-67				
L 46	-81	125	-178	-94	169			
I 47	-115	-7	179	51	180			
N 48	-155	179	179	-65	-158			
E 49	-77	-22	-177	-116	178	9		
N 50	-114	-2	-174	-77	-94			
W 51	-139	140	178	-77	91			
V 52	-119	145	179	163				
V 53	-114	132	-179	124				

Residue	Phi	Psi	Omega	X-1	X-2	X-3	X-4	X-5
T 54	-144	-171	177	-167				
A 55	-93	145	180					
A 56	-65	-34	179					
H 57	-63	-24	177	84	-110			
C 58	-60	-5	-179	-62	-91	-91	-149	
G 59	53	36	179					
V 60	-64	143	174	-169				
T 61	-131	158	177	86				
T 62	-64	-4	177	57				
S 63	-98	-6	175	102				
D 64	-89	162	-179	-55	139			
V 65	-124	137	177	142				
V 66	-94	122	-180	167				
V 67	-112	117	-179	176				
A 68	-118	148	-180					
G 69	92	10	-178					
E 70	-90	142	178	-86	-115	111		
F 71	-121	-59	-179	165	74			
D 72	-105	126	-179	-169	164			
Q 73	-82	6	178	-83	166	-8		
G 74	-120	5	-180					
S 75	-46	117	180	59				
S 76	-54	-26	-180	63				
S 77	173	145	-180	-178				
E 78	-152	-179	-180	-70	-169	38		
K 79	-86	56	-178	-130	78	163	-160	
I 80	-67	162	177	59	131			
Q 81	-118	124	177	-72	165	67		
K 82	-103	111	-177	-43	88	-172	-109	
L 83	-113	134	179	-43	152			
K 84	-79	153	-177	-74	59	-143	158	
I 85	-110	118	-180	-59	179			
A 86	-75	-48	-177					
K 87	-152	146	178	-159	148	145	139	
V 88	-106	133	178	176				
F 89	-119	94	-177	-66	88			
K 90	-86	118	-178	42	171	-139	-146	
N 91	-68	120	-179	163	167			
S 92	-56	-10	180	4				
R 93	-98	-4	-178	-69	-154	155	-164	
Y 94	-57	129	-176	171	62			
N 95	-106	105	-177	-166	-63			
S 96	-67	-3	179	80				
L 97	-95	-53	179	-74	177			
T 98	-78	-28	-176	-58				
I 99	75	19	179	-42	-39			
N100	-82	152	179	-171	23			
N101	64	51	-174	-58	-47			

Residue	Phi	Psi	Omega	X-1	X-2	X-3	X-4	X-5
D102	-83	79	178	-168	-167			
I103	-150	145	178	-176	162			
T104	-140	149	177	-85				
L105	-114	139	-180	-68	162			
L106	-109	118	176	-45	-171			
K107	-94	117	-178	153	-178	-83	-63	
L108	-70	122	-180	-49	-176			
S109	-47	-48	179	-141				
T110	-129	125	-179	32				
A111	-59	157	177					
A112	-80	150	177					
S113	-107	111	-180	-36				
F114	-64	146	-172	-74	9			
S115	-161	-171	-178	131				
Q116	-58	-34	-178	-73	80	-4		
T117	-102	-11	179	74				
V118	-128	130	-177	179				
S119	-173	171	-178	-66				
A120	-97	155	179					
V121	-92	158	176	57				
C122	-90	154	177	-51	-70	108	72	
L123	-105	135	177	-75	-168			
P124	-78	180	175	33	-37	28		
S125	-107	156	178	153				
A126	-55	-34	178					
S127	-92	7	179	81				
D128	-74	155	-178	12	157			
D129	-127	110	179	-174	99			
F130	-103	115	-178	-57	89			
A131	-80	152	176					
A132	-59	131	-180					
G133	108	-25	-179					
T134	-49	133	179	-64				
T135	-85	112	-180	-71				
C136	-114	175	-177	-54	-126	107	-95	
V137	-132	138	177	176				
T138	-124	152	179	-173				
T139	-134	163	174	50				
G140	158	176	-180					
W141	-110	23	179	-83	-74			
G142	-79	170	-179					
L143	-45	143	174	-55	-158			
T144	-94	-8	-179	67				
R145	-155	135	-180	-150	142	-132	-45	-37
Y146	-54	131		178	60			
A149		170	-179					
N150	-112	19	178	-82	-23			
T151	-89	117	179	76				

Residue	Phi	Psi	Omega	X-1	X-2	X-3	X-4	X-5
P152	-70	148	178	26	-38	35		
D153	-92	-33	180	-81	94			
R154	-95	141	176	-55	-164	-112	154	-169
L155	-57	136	-175	-178	80			
Q156	-114	154	-179	-54	-80	-48		
Q157	-141	159	176	52	172	55		
A158	-155	140	175					
S159	-93	140	-177	-91				
L160	-166	154	179	78	157			
P161	-85	151	176	12	4	-19		
L162	-82	156	176	-63	171			
L163	-121	157	177	-96	51			
S164	-74	159	-180	23				
N165	-63	-34	-180	-104	-144			
T166	-70	-37	176	90				
N167	-70	-31	-178	-69	114			
C168	-72	-27	-178	-165	175	-84	-172	
K169	-71	-17	178	-89	166	-162	-134	
K170	-59	-54	-179	-92	168	-158	139	
Y171	-72	-45	-175	-65	-78			
W172	-116	-2	-178	-58	102			
G173	51	-127	-180					
T174	-59	-22	-179	-69				
K175	-64	-26	-180	-95	-161	-120	-59	
I176	-84	111	-179	-51	-59			
K177	-104	170	175	-39	-38	-152	157	
D178	-36	-61	180	-43	117			
A179	-85	37	179					
M180	-133	151	177	-54	-179	-178		
I181	-137	141	-179	-158	-168			
C182	-116	157	180	-53	-172	-84	175	
A183	-161	160	176					
G184	105	-138	-178					
A185	63	19	-172					
S186	-120	14	179	59				
G187	113	-2	178					
V188	-133	160	179	-60				
S189	-151	137	173	-179				
S190	-74	152	-179	-80				
C191	-140	174	-179	-155	44	89	-175	
M192	-41	125	-177	-69	101	80		
G193	105	-21	-179					
D194	-78	-26	174	-79	150			
S195	-31	127	177	-104				
G196	87	-14	-179					
G197	-80	174	-177					
P198	-85	152	168	33	-37	25		
L199	-124	111	-177	173	81			

Residue	Phi	Psi	Omega	X-1	X-2	X-3	X-4	X-5
V200	-118	153	176	-78				
C201	-131	147	-179	-43	-95	107	-126	
K202	-86	120	180	-136	-175	169	-177	
K203	-132	117	178	176	176	-167	-65	
N204	61	48	-177	-71	-37			
G205	69	8	-176					
A206	-122	161	-177					
W207	-96	129	179	-73	97			
T208	-121	143	180	-46				
L209	-72	100	-176	-171	65			
V210	-97	-32	179	173				
G211	-143	160	175					
I212	-118	119	-179	-67	155			
V213	-56	123	-175	176				
S214	-115	-62	-176	-169				
W215	-162	178	-178	56	-92			
G216	171	-166	-176					
S217	-41	127	180	175				
S218	-56	-30	-180	59				
T219	-110	10	-177	-95				
C220	56	31	180	-51	-175	89	44	
S221	-59	140	-177	165				
T222	-88	13	179	76				
S223	-122	8	180	-71				
T224	-121	130	-177	-47				
P225	-65	148	174	20	-29	27		
G226	-86	161	-176					
V227	-122	127	-178	-174				
Y228	-124	149	175	-59	78			
A229	-64	133	-179					
R230	-84	106	-173	178	-172	-84	-67	167
V231	-56	-34	-180	167				
T232	-54	-43	180	-105				
A233	-69	-23	-174					
L234	-120	-8	-177	-46	177			
V235	-68	-23	179	72				
N236	-66	-34	-179	-75	-25			
W237	-61	-45	177	164	83			
V238	-55	-65	177	165				
Q239	-52	-49	-180	-57	-58	10		
Q240	-60	-44	176	-51	171	-173		
T241	-61	-49	-180	-50				
L242	-60	-44	-180	-72	-161			
A243	-58	-51	-178					
A244	-67	-20	180					
N245	-132	-175		-46	-78			

APPENDIX D

Appendix D: Hydrogen Bonds in the α -CHT Dimer

A.) Molecule 1

Donor	Acceptor	H-A	D-A	Theta
207 N HN	2 O	2.03	3.02	170.63
16 N HN3	194 OD1	1.71	2.62	148.79
17 N HN	189 O	1.79	2.74	155.49
20 N HN	157 O	1.88	2.82	155.36
157 N HN	20 O	1.68	2.63	155.66
22 N HN	155 O	1.71	2.69	164.95
26 N HN	23 O	2.29	3.28	170.07
30 N HN	27 O	2.20	3.14	155.14
119 N HN	28 O	2.33	3.27	156.23
46 N HN	29 O	1.93	2.86	153.20
30 NE2 HNE2	31 O	2.01	3.00	171.49
31 N HN	44 O	2.04	3.03	168.20
44 N HN	31 O	2.26	3.04	133.95
32 N HN	67 O	2.21	3.16	158.46
67 N HN	32 O	1.78	2.76	163.89
33 N HN	42 O	2.01	3.01	177.13
41 N HN	33 O	1.91	2.85	154.94
34 N HN	65 O	2.00	2.98	166.78
65 N HN	34 O	1.88	2.82	154.31
35 N HN	39 O	2.26	3.21	157.12
39 N HN	35 OD2	2.28	3.21	154.92
38 N HN	35 OD2	1.53	2.52	171.78
40 NE2 HE2	193 O	1.51	2.48	161.63
43 N HN	195 O	1.59	2.56	164.48
45 N HN	53 O	1.88	2.79	150.45
53 N HN	45 O	1.94	2.87	153.59
121 N HN	46 O	2.06	3.02	162.43
47 N HN	51 O	2.03	3.02	171.10
112 N HN	49 O	1.75	2.66	149.21
108 N HN	50 O	1.81	2.81	173.27
52 N HN	106 O	1.83	2.77	155.27
106 N HN	52 O	2.09	2.97	146.40
54 N HN	104 O	1.80	2.77	163.61
55 N HN	54 OG1	1.92	2.61	123.02
104 N HN	54 O	1.96	2.79	139.16
59 N HN	56 O	2.04	3.04	173.76
57 N HN	102 OD2	1.86	2.85	167.78
57 ND1 HD1	102 OD1	1.71	2.68	161.07
61 N HN	64 OD1	1.89	2.83	155.86
63 N HN	61 OG1	2.12	3.08	161.84
64 N HN	61 O	2.16	3.08	150.87
66 N HN	83 O	1.86	2.85	173.74
83 N HN	66 O	1.85	2.75	148.20
68 N HN	81 O	1.86	2.85	173.34

Donor			Acceptor	H-A	D-A	Theta
81	N	HN	68 O	1.87	2.80	153.00
70	N	HN	70 OE1	2.34	3.21	144.55
74	N	HN	72 OD2	1.99	2.95	160.87
73	N	HN	153 O	2.13	2.94	137.28
113	N	HN	81 OE1	2.27	3.15	145.92
81	NE2	HNE1	113 O	1.93	2.80	143.37
109	N	HN	84 O	2.01	3.00	169.32
86	N	HN	107 O	1.86	2.84	164.43
87	N	HN	107 O	2.33	3.29	161.90
107	N	HN	87 O	2.04	3.04	176.71
89	N	HN	105 O	1.63	2.56	152.92
105	N	HN	89 O	1.96	2.96	173.38
91	N	HN	103 O	2.36	3.30	156.69
93	N	HN	91 OD1	2.18	2.86	123.21
94	N	HN	91 O	2.04	3.03	169.48
101	ND2	HND2	93 O	1.98	2.97	169.65
95	N	HN	100 O	2.23	3.11	145.28
99	N	HN	95 O	2.19	2.91	127.04
103	N	HN	101 O	2.12	2.93	136.68
115	N	HN	118 O	1.94	2.81	143.95
117	N	HN	115 OG	2.34	3.21	145.58
118	N	HN	115 OG	2.43	3.26	139.33
116	N	HN	116 OE1	2.02	2.96	154.37
122	N	HN	207 O	2.07	3.06	171.21
209	N	HN	122 O	1.87	2.84	160.30
125	N	HN	128 OD1	1.32	2.20	143.16
128	N	HN	125 O	2.27	3.24	163.85
134	N	HN	131 O	1.90	2.84	154.68
133	N	HN	162 O	1.97	2.92	156.06
162	N	HN	134 O	1.88	2.81	154.49
136	N	HN	160 O	1.62	2.58	160.50
160	N	HN	136 O	2.23	3.12	147.20
137	N	HN	200 O	1.91	2.80	148.01
200	N	HN	137 O	1.87	2.83	158.71
138	N	HN	158 O	1.96	2.92	158.98
158	N	HN	138 O	2.37	3.12	131.53
140	N	HN	156 O	2.20	3.16	159.46
142	N	HN	194 OD2	1.84	2.74	148.21
143	N	HN	192 O	1.92	2.81	145.70
144	N	HN	150 O	1.78	2.73	157.11
145	NH2	HH22	150 OD1	1.96	2.95	172.71
156	NE2	HNE2	154 O	2.01	3.00	169.52
184	N	HN	161 O	1.89	2.85	157.88
163	N	HN	182 O	2.07	3.00	153.69
168	N	HN	164 O	2.03	2.96	155.20
169	N	HN	165 O	2.28	3.14	143.09
167	N	HN	167 OD1	1.98	2.65	121.60
171	N	HN	168 O	2.21	3.05	139.91

Donor		Acceptor	H-A	D-A	Theta
172 N	HN	168 O	1.87	2.85	166.90
173 N	HN	169 O	1.63	2.52	146.41
191 N	HN	194 OD1	1.79	2.76	163.45
194 N	HN	191 O	1.99	2.91	150.56
197 N	HN	194 O	2.21	3.18	165.29
196 N	HN	213 O	1.93	2.91	167.14
213 N	HN	197 O	2.00	2.97	161.62
199 N	HN	211 O	2.06	2.88	138.02
210 N	HN	199 O	1.71	2.69	165.16
211 N	HN	199 O	1.96	2.92	157.88
201 N	HN	208 O	2.04	3.01	160.61
208 N	HN	201 O	1.84	2.80	162.41
203 N	HN	206 O	1.79	2.77	166.06
206 N	HN	203 O	1.92	2.88	161.72
231 N	HN	210 O	2.08	3.06	165.31
212 N	HN	229 O	2.37	3.26	147.58
229 N	HN	212 O	1.88	2.84	159.19
214 N	HN	227 O	2.24	3.08	141.19
215 N	HN	227 O	2.02	3.00	165.17
227 N	HN	215 O	1.88	2.87	169.55
221 N	HN	217 OG	1.97	2.95	167.36
219 N	HN	217 OG	2.12	3.09	161.79
220 N	HN	217 O	1.82	2.75	153.28
233 N	HN	230 O	2.23	3.05	138.27
234 N	HN	231 O	2.05	2.96	150.91
238 N	HN	234 O	2.34	3.25	150.36
239 N	HN	235 O	1.96	2.91	155.87
240 N	HN	236 O	2.11	3.06	157.52
241 N	HN	237 O	2.05	3.01	160.59
242 N	HN	238 O	1.76	2.73	161.11
243 N	HN	239 O	1.82	2.76	154.58
244 N	HN	240 O	2.30	3.16	143.60
H-A Hydrogen to Acceptor Distance					
D-A Donor to Acceptor Distance					
Theta Donor-Hydrogen-Acceptor Angle (Degrees)					

B.) Molecule 2

Donor	Acceptor	H-A	D-A	Theta
2 N HN	120 O	1.93	2.93	173.42
207 N HN	2 O	1.99	2.92	154.49
16 N HN3	194 OD1	1.43	2.33	145.90
17 N HN	189 O	1.87	2.83	160.48
18 ND2 HND2	187 O	2.08	2.97	146.46
20 N HN	157 O	2.06	2.94	144.86
157 NE2 HNE2	20 OE2	2.28	3.24	159.44
157 N HN	20 O	2.00	2.91	150.26
154 NH2 HH22	21 OE1	1.70	2.45	127.79
22 N HN	155 O	1.93	2.92	175.50
26 N HN	23 O	2.19	3.16	164.31
30 N HN	27 O	2.06	2.99	154.52
46 N HN	29 O	2.00	2.92	151.16
30 NE2 HNE2	31 O	1.82	2.78	158.83
31 N HN	44 O	1.98	2.97	170.93
32 N HN	67 O	1.95	2.92	163.07
67 N HN	32 O	1.87	2.85	166.39
33 N HN	42 O	1.77	2.77	173.84
41 N HN	33 O	2.13	3.06	154.92
34 N HN	65 O	1.80	2.80	173.24
65 N HN	34 O	2.03	3.00	164.55
35 N HN	39 O	2.28	3.18	148.91
35 N HN	35 OD2	1.52	2.23	123.42
37 N HN	35 OD1	1.93	2.85	151.43
38 N HN	35 O	2.08	2.98	148.25
43 N HN	195 O	1.61	2.60	168.03
45 N HN	53 O	1.97	2.92	157.69
53 N HN	45 O	1.68	2.65	163.34
121 N HN	46 O	2.18	3.10	152.71
47 N HN	51 O	2.04	3.02	165.53
48 ND2 HND2	47 O	1.60	2.49	145.08
112 N HN	49 O	1.65	2.63	164.08
108 N HN	50 O	1.80	2.78	165.11
52 N HN	106 O	2.04	3.00	158.07
106 N HN	52 O	2.02	2.97	157.40
54 N HN	104 O	1.98	2.95	165.14
55 N HN	54 OG1	2.01	2.69	122.89
104 N HN	54 O	2.08	3.02	156.17
58 N HN	55 O	2.35	3.22	145.56
59 N HN	56 O	2.35	3.30	157.99
57 N HN	102 OD2	1.88	2.79	150.37
57 ND1 HD1	102 OD1	1.59	2.58	169.02
61 N HN	64 CD1	2.23	3.20	161.96
64 N HN	61 O	1.67	2.64	164.04
66 N HN	83 O	1.98	2.96	165.95
83 N HN	66 O	1.77	2.76	168.92

Donor		Acceptor	H-A	D-A	Theta
68	N HN	81 O	1.85	2.83	163.47
81	N HN	68 O	1.66	2.63	161.79
155	N HN	71 O	2.17	3.00	140.24
74	N HN	72 OD2	1.93	2.81	145.08
73	N HN	153 O	1.80	2.76	159.86
80	N HN	78 O	2.15	2.97	138.24
113	N HN	81 OE1	2.05	3.04	174.59
81	NE2 HNE1	113 O	2.41	3.29	147.73
109	N HN	84 O	2.10	3.01	151.68
86	N HN	107 O	1.99	2.89	147.63
107	N HN	87 O	1.86	2.85	170.45
89	N HN	105 O	1.67	2.65	164.35
105	N HN	89 O	1.87	2.81	154.05
94	N HN	91 O	1.95	2.94	169.35
101	ND2 HND2	93 O	2.13	3.03	149.12
95	N HN	100 O	2.22	3.14	152.78
98	N HN	95 OD1	2.04	2.99	158.75
97	N HN	95 OD1	1.88	2.82	154.71
99	N HN	95 O	1.75	2.65	147.70
101	N HN	100 OD1	2.25	2.89	120.32
103	N HN	101 O	2.26	2.99	129.71
115	N HN	118 O	1.87	2.83	158.47
117	N HN	115 OG	2.04	3.00	160.28
118	N HN	115 O	2.41	3.25	141.28
116	N HN	116 OE1	2.31	3.18	145.85
122	N HN	207 O	2.24	3.23	169.02
209	N HN	122 O	1.77	2.74	163.16
125	N HN	128 OD2	1.58	2.40	135.59
125	N HN	128 OD1	2.05	2.99	154.25
128	N HN	125 O	2.03	3.01	166.56
134	N HN	131 O	1.90	2.87	161.49
133	N HN	162 O	1.91	2.82	149.60
162	N HN	134 O	2.21	3.09	146.22
136	N HN	160 O	1.77	2.73	159.63
160	N HN	136 O	1.87	2.79	152.01
137	N HN	200 O	2.05	2.99	154.77
200	N HN	137 O	1.82	2.81	172.76
138	N HN	158 O	1.88	2.80	151.83
158	N HN	138 O	2.12	2.97	141.32
140	N HN	156 O	2.08	2.98	149.83
156	N HN	140 O	2.18	3.14	162.33
142	N HN	194 OD2	1.89	2.77	145.25
143	N HN	192 O	2.16	3.09	155.21
144	N HN	150 O	2.07	3.01	156.25
184	N HN	161 O	2.04	2.94	148.52
163	N HN	182 O	2.10	3.06	159.47
167	N HN	164 OG	2.23	3.14	149.75
168	N HN	164 O	2.10	3.04	154.45

Donor		Acceptor	H-A	D-A	Theta
169 N	HN	165 O	2.13	2.89	131.05
170 N	HN	167 O	2.24	3.05	136.76
172 N	HN	168 O	1.95	2.89	155.53
173 N	HN	169 O	1.84	2.69	140.21
175 N	HN	172 O	2.27	3.22	156.37
180 N	HN	177 O	2.36	3.09	128.93
178 N	HN	178 OD1	2.13	3.12	169.13
180 N	HN	178 O	2.34	3.10	131.39
230 N	HN	179 O	2.04	3.03	167.74
230 NE	HE	180 O	2.11	2.90	134.05
181 N	HN	228 O	2.11	3.05	155.60
228 N	HN	181 O	1.66	2.61	158.89
183 N	HN	226 O	2.29	3.21	151.74
226 N	HN	183 O	2.03	2.94	150.58
187 N	HN	222 O	2.04	2.93	146.31
191 N	HN	194 OD1	1.91	2.88	160.66
194 N	HN	191 O	1.96	2.90	155.46
197 N	HN	194 O	2.25	3.13	146.37
196 N	HN	213 O	1.93	2.92	168.15
213 N	HN	197 O	1.78	2.77	168.26
199 N	HN	211 O	2.12	3.01	146.50
211 N	HN	199 O	2.14	3.12	165.60
210 N	HN	199 O	1.88	2.81	153.26
201 N	HN	208 O	1.81	2.79	166.87
208 N	HN	201 O	1.80	2.77	161.84
203 N	HN	206 O	1.58	2.52	153.35
206 N	HN	203 O	1.94	2.90	161.29
231 N	HN	210 O	1.94	2.92	166.59
212 N	HN	229 O	2.28	3.17	146.71
229 N	HN	212 O	1.98	2.83	140.85
214 N	HN	227 O	2.15	3.01	144.34
215 N	HN	227 O	2.01	2.99	164.42
227 N	HN	215 O	1.88	2.84	160.24
221 N	HN	217 OG	2.03	2.96	153.27
220 N	HN	217 O	2.07	3.06	170.12
223 N	HN	221 OG	2.33	3.30	162.96
233 N	HN	230 O	2.29	3.07	133.54
234 N	HN	231 O	2.04	2.96	153.51
235 N	HN	231 O	1.97	2.68	125.30
238 N	HN	234 O	2.34	3.28	156.51
239 N	HN	235 O	1.75	2.75	170.19
240 N	HN	236 O	1.94	2.93	176.93
241 N	HN	237 O	1.87	2.85	163.00
242 N	HN	238 O	1.84	2.81	162.11
243 N	HN	239 O	1.71	2.68	160.97
244 N	HN	240 O	1.99	2.95	161.47
245 ND2	HND2	241 O	1.94	2.70	131.21

APPENDIX E

Appendix E: Solvent Molecule Positions in the α -CHT Dimer

Exterior

498,499,500,501,502,503,504,505,506,507,508,510,511,514,515
520,521,522,524,526,527,528,531,532,533,534,535,536,537,538
539,542,543,544,546,547,548,551,552,553,554,557,558,559,560
561,562,563,564,565,566,567,569,571,572,573,575,576,577,578
579,580,581,582,583,584,587,588,589,590,591,592,593,594,595
596,597,598,599,600,601,603,605,607,608,609,610,611,612,613
615,616,617,618,620,621,623,624,625,626,627,628,629,631,632
633,634,635,636,637,638,639,640,641,642,643,644,645,646,647
648,649,650,651,652,653,654,656,657,658,659,660,661,663,664
665,666,667,669,670,671,673,674,675,676,677,678,679,680,681
682,683,685,686,687,688,689,691,692,693,694,695,696,697,698
700,701,702,704,705,706,707,708,709,710,711,712,713,714,715
716,717,718,719,720,721,722,723,724,725,726,727,728,729,730
731,732,733,734,735,736,737,738,739,740,741,742

Interior

509,512,513,516,517,518,519,529,530,555,565,603,606,614,703
668,684,690,699,732

Interface

496,497,523,525,540,541,545,549,550,568,570,574,585,586,602
619,622,630,655,662,672,734

APPENDIX F

Appendix F: Protein-Solvent Hydrogen Bonds in the
 α -CHT Dimer

A.) Molecule 1

Donor	Acceptor	H-A	D-A	Theta
3 N HN	564 WAT	1.92	2.90	166.55
6 N HN	544 WAT	1.87	2.86	172.56
16 N HN2	505 WAT	2.10	2.98	146.18
18 N HN	663 WAT	1.92	2.92	171.77
21 N HN	674 WAT	1.93	2.89	159.52
23 N HN	531 WAT	2.00	3.00	174.13
26 OG HO	531 WAT	1.79	2.62	138.12
60 N HN	632 WAT	1.77	2.69	151.44
85 N HN	573 WAT	2.29	3.26	163.62
88 N HN	587 WAT	2.07	3.06	170.91
100 ND2 HND1	593 WAT	2.02	3.01	171.83
101 N HN	514 WAT	2.23	2.88	121.12
102 N HN	514 WAT	2.10	3.09	171.64
115 OG HO	740 WAT	1.97	2.63	121.34
117 OG1 HO	584 WAT	2.35	3.14	135.14
120 N HN	546 WAT	2.30	3.29	170.00
123 N HN	561 WAT	1.77	2.74	161.31
126 N HN	560 WAT	2.26	3.24	165.21
141 N HN	512 WAT	1.86	2.85	174.49
149 N HN2	611 WAT	1.58	2.50	149.36
150 N HN	568 WAT	1.95	2.90	157.06
150 ND2 HND1	523 WAT	2.07	3.05	168.01
157 NE2 HNE1	531 WAT	1.83	2.79	160.86
159 N HN	596 WAT	2.00	2.92	151.81
164 N HN	539 WAT	2.07	3.02	159.28
166 OG1 HO	712 WAT	1.98	2.86	145.77
174 N HN	572 WAT	2.33	3.25	152.30
182 N HN	504 WAT	2.06	3.04	165.77
185 N HN	508 WAT	1.72	2.72	173.11
202 N HN	623 WAT	1.92	2.91	169.37
203 NZ HZ3	543 WAT	2.01	2.97	159.59
204 N HN	581 WAT	1.86	2.84	165.84
205 N HN	705 WAT	2.26	3.23	164.72
214 OG HO	542 WAT	1.82	2.81	170.11
217 N HN	527 WAT	2.26	3.22	161.57
217 OG HO	701 WAT	1.39	2.26	140.17
224 OG1 HO	739 WAT	2.31	3.28	165.74
228 OH HO	498 WAT	2.25	3.10	140.99
232 N HN	520 WAT	1.75	2.72	161.74
236 N HN	510 WAT	2.06	3.05	169.12
237 N HN	652 WAT	2.31	3.21	149.31
239 NE2 HNE2	565 WAT	1.78	2.61	138.34

B.) Molecule 2

Donor		Acceptor	H-A	D-A	Theta
5	N	HN	579 WAT	2.21	2.89
6	N	HN	579 WAT	1.85	2.85
16	N	HN2	515 WAT	2.21	3.19
18	N	HN	563 WAT	1.80	2.78
19	N	HN	661 WAT	2.42	3.25
23	N	HN	511 WAT	2.14	3.13
26	OG	HO	511 WAT	2.07	2.96
27	N	HN	703 WAT	1.91	2.70
27	N	HN	583 WAT	2.07	2.99
34	NE2	HNE2	594 WAT	1.71	2.69
40	NE2	HE2	530 WAT	2.21	3.20
60	N	HN	577 WAT	1.67	2.63
62	N	HN	737 WAT	2.11	3.06
69	N	HN	536 WAT	2.20	3.16
70	N	HN	557 WAT	2.10	3.05
84	N	HN	696 WAT	2.12	3.11
85	N	HN	528 WAT	2.09	3.05
92	N	HN	697 WAT	2.24	3.23
102	N	HN	537 WAT	1.87	2.86
104	OG1	HO	532 WAT	2.19	3.11
114	N	HN	675 WAT	2.40	3.22
120	N	HN	588 WAT	2.14	3.08
123	N	HN	578 WAT	1.96	2.95
135	N	HN	552 WAT	2.09	3.06
141	N	HN	530 WAT	2.02	3.01
150	N	HN	525 WAT	2.20	3.04
159	N	HN	631 WAT	2.20	3.20
159	OG	HO	618 WAT	1.70	2.57
164	N	HN	576 WAT	1.52	2.49
165	ND2	HND2	522 WAT	1.75	2.42
182	N	HN	522 WAT	1.86	2.83
185	N	HN	497 WAT	1.57	2.56
195	N	HN	662 WAT	2.16	2.98
204	N	HN	626 WAT	1.97	2.92
214	OG	HO	636 WAT	2.10	3.06
222	N	HN	660 WAT	1.82	2.79
228	OH	HO	556 WAT	2.05	2.94
230	NH1	HH11	659 WAT	1.99	2.86
232	N	HN	524 WAT	2.02	3.00
236	N	HN	533 WAT	2.14	3.13
239	NE2	HNE2	638 WAT	1.49	2.42

H-A Hydrogen to Acceptor Distance
D-A Donor to Acceptor Distance
Theta Donor-Hydrogen-Acceptor Angle (Degrees)

APPENDIX G

Appendix G: Polar Protein Atoms - Solvent Interactions
in the α -CHT Dimer

Water Number	Protein Atom	Distance(A)
497	VAL (2) 188O	2.53
	THR (2) 224O	2.86
504	LEU (1) 163O	2.62
505	GLY (1) 140O	2.76
	GLY (1) 142O	2.65
506	GLN (2) 34OE1	2.81
508	VAL (1) 188O	2.55
	THR (1) 224O	2.83
509	GLY (1) 196O	2.76
510	THR (1) 232O	2.93
511	GLN (2) 157OE1	2.73
512	GLY (1) 193O	2.70
514	ALA (1) 179O	2.95
515	GLY (2) 140O	2.83
	GLY (2) 142O	2.68
516	THR (1) 139O	2.62
519	GLY (2) 196O	2.94
520	VAL (1) 210O	2.84
522	LEU (2) 163O	2.74
524	VAL (2) 210O	2.77
525	CYS (1) 58O	2.93
528	THR (2) 62O	2.59
	ASP (2) 64O	2.46
529	ASP (2) 194O	2.74
532	ALA (2) 56O	2.60
533	THR (2) 232O	2.79
534	SER (2) 159O	2.99
536	PRO (2) 28O	2.88
	GLN (2) 30O	2.65
537	ALA (2) 179O	2.88
538	PRO (2) 80T	2.67
544	GLY (1) 25O	2.62
	GLN (1) 116O	2.61
547	ASP (1) 128OD2	2.23
549	PHE (1) 41O	2.73
551	PRO (1) 28O	2.90
	GLN (1) 30O	2.54
553	ALA (1) 120O	2.70
554	GLU (1) 20OE1	2.94
557	VAL (2) 67O	2.75
559	VAL (2) 60O	2.39
560	ASN (1) 236OD1	2.48
563	THR (2) 144O	2.81
565	PRO (1) 124O	2.50
566	ASN (1) 204OD1	1.95

Water Number	Protein Atom	Distance(A)
567	ALA (2) 233O	2.87
571	THR (1) 62O	2.57
573	THR (1) 62O	2.21
574	PHE (2) 41O	2.80
576	ASN (2) 167OD1	2.28
579	GLY (2) 25O	2.78
	GLN (2) 116O	2.68
580	LYS (1) 202O	2.69
583	TRP (2) 27O	2.88
584	GLY (1) 69O	2.72
586	MET (1) 192O	3.00
589	LEU (2) 155O	1.92
590	PRO (1) 8OT	2.23
602	SER (2) 96O	2.90
603	TRP (1) 215O	2.95
	VAL (1) 227O	2.90
604	LYS (1) 175O	2.71
606	TRP (2) 215O	2.83
611	ALA (1) 149O	2.73
612	VAL (1) 121O	2.96
613	ALA (2) 55O	2.09
614	GLN (2) 30O	2.90
	GLU (2) 70O	2.87
615	GLY (2) 133O	2.84
619	PHE (1) 41O	2.79
620	ASP (1) 129OD1	2.38
622	LEU (1) 97O	2.93
624	ASP (1) 35O	2.81
628	ASP (1) 129OD1	2.78
630	SER (1) 218O	2.31
631	GLU (2) 20OE1	2.32
634	GLN (1) 156OE1	2.52
638	PRO (2) 124O	2.76
640	LYS (1) 79O	2.61
641	GLY (1) 69O	2.72
647	PRO (2) 4O	2.81
650	THR (1) 117O	2.81
658	ASP (1) 128O	2.48
661	ILE (2) 16O	2.94
	GLN (1) 156OE1	2.54
662	CYS (2) 191O	2.97
663	THR (1) 144O	2.32
665	LYS (2) 175O	2.32
667	SER (1) 77O	2.90
672	GLY (2) 59O	2.94
675	GLU (2) 49OE2	2.21
679	ILE (2) 47O	2.56
680	LYS (1) 169O	2.30

Water Number	Protein Atom	Distance(A)
681	THR (1) 232O	2.87
682	ASN (2) 48OD1	2.93
684	LEU (2) 97O	2.87
685	LYS (2) 90O	2.36
690	ASP (1) 153OD1	1.94
691	ASN (1) 101OD1	2.99
692	ASP (1) 178OD1	2.56
694	SER (1) 159O	2.59
695	GLN (1) 7O	2.86
707	SER (1) 115O	2.44
	GLN (1) 116O	2.03
714	CYS (1) 1O	2.65
715	THR (1) 110O	2.78
716	ALA (2) 22O	2.14
720	PRO (2) 61O	2.08
721	PRO (1) 8O	2.82
722	LYS (1) 170O	2.07
724	ASN (2) 100OD1	2.78
	ASN (2) 101OD1	2.55
730	GLN (2) 50D1	2.11
732	SER (2) 418O	2.31
733	SER (1) 115O	2.80
734	SER (2) 96O	2.93
735	GLN (2) 240O	2.90
738	ALA (2) 126O	2.56
	ASP (2) 128O	2.65
740	SER (1) 113O	2.95

R.M.S. Deviation = 2.67