



This is to certify that the

dissertation entitled

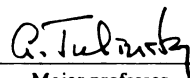
THE STRUCTURE OF A COMPLEX OF RECOMBINANT HIRUDIN
AND HUMAN α -THROMBIN AT 2.3 Å RESOLUTION

presented by

Timothy John Rydel

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Chemistry


Major professor

Date January 28, 1991



PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.

DATE DUE	DATE DUE	DATE DUE
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

MSU Is An Affirmative Action/Equal Opportunity Institution

c:\cir\date\due.pm3-p.1

THE STRUCTURE OF A COMPLEX OF RECOMBINANT HIRUDIN
AND HUMAN α -THROMBIN AT 2.3 Å RESOLUTION

By

Timothy John Rydel

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Chemistry

1991

ABSTRACT

THE STRUCTURE OF A COMPLEX OF RECOMBINANT HIRUDIN AND HUMAN α -THROMBIN AT 2.3 Å RESOLUTION

By

Timothy John Rydel

Crystals of a complex of the recombinant hirudin, rHV2-K47, and human α -thrombin were obtained via vapor diffusion from hanging drops. The hirudin-thrombin complex crystallizes in the tetragonal crystal system, space group $P4_32_12$, with $a = b = 90.54$ Å, $c = 132.04$ Å and one molecule in the asymmetric unit.

The structure was solved by Patterson search techniques using PPACK-human α -thrombin as a model and a 2.8 Å resolution diffractometer intensity data set. The rotational search was conducted in Patterson space using 8.0 to 3.0 Å resolution data and the SEARCH routine in the PROTEIN package of macromolecular crystallographic programs. The translational search also used 8.0 to 3.0 Å resolution data, and was conducted with programs written by Lattman and modified by Deisenhofer and Huber. The R-value of the initial solution was 0.39, but it reduced to 0.34 when the rotational and translational parameters were refined with the rigid body refinement program TRAREF.

The initial structure was improved using 2.3 Å resolution FAST area detector data, the energy restraint-crystallographic refinement program EREF, and graphics interventions. After six iterative stages of model-building and refinement, the hirudin-thrombin complex (includes 207 water molecules) had an R-value of 0.193 for 21,960 reflections between 7.0 and 2.3 Å resolution.

The complex was further refined using PROFFT, the fast Fourier

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

transform version of the stereochemically-restrained least squares refinement program PROLSQ. After five iterative stages of model-building and refinement, the complex (includes 265 water molecules) had excellent overall stereochemistry and an R-value of 0.173 for 21,056 reflections between 7.0 and 2.3 Å resolution.

Hirudin consists of an N-terminal globular domain with a three-cystine core and an extended C-terminal tail. The N-terminal tripeptide of hirudin interacts in a novel manner in the active site region of thrombin. The C-terminal tail of hirudin makes many electrostatic as well as hydrophobic interactions in the anion-binding exosite of thrombin. In all, hirudin engages in 216 contacts of less than 4.0 Å with thrombin. This abundance of interactions is likely the source of the high affinity and specificity of hirudin.

The overall folding of the thrombin structure in the complex is very similar to that of PPACK-thrombin, however many significant main chain and side chain differences exist between the two. Most of the differences can be accounted for in terms of four categories of factors or influences.

To my parents,
Arlene and Joseph Rydel,
and to my wife,
Odette,
for their encouragement, love, and support.

his

aspen

rental

his

rent

exp

exte

stud

stan

ref

Bio

exc

fr

19

20

an

7

8

9

ACKNOWLEDGEMENTS

To Professor Alexander Tulinsky I extend my sincere gratitude for his constant enthusiasm, guidance, interest, and support throughout all aspects of this work. I could not have asked for a better scientific mentor, and it has been a wonderful learning experience to be a part of his laboratory. I am also very grateful to him for allowing me to return to Michigan State to continue work toward a Ph.D. degree in his group.

To Dr. K.G. Ravichandran, a former post-doc of the laboratory, I extend my sincere thanks for his important contributions to the structure solution of the hirudin-thrombin complex. In particular, I am grateful to him for his assistance with the model-building and refinement work on the complex conducted at the Max-Planck-Institut für Biochemie in Martinsried, Germany, and for his friendship during two exciting and exhausting months spent overseas.

I have also benefitted a great deal from being a part of a very friendly, instructive and supportive research group. I would especially like to thank Drs. Vasili Carperos, Anne Mulichak, Pappan Padmanabhan, Ewa Skrzypczak-Jankun, and Manuel Soriano-Garcia for their assistance and helpful discussions.

It is also a pleasure for me acknowledge Dr. Carolyn Roitsch of Transgene, S.A., in Strasbourg, France for supplying us with ample amounts of recombinant hirudin, and for her genuine enthusiasm and interest in the project. The generous financial support of Transgene,

S.A., is also gratefully acknowledged. Thanks is also extended to Dr. John W. Fenton II of the New York State Department of Health in Albany, NY for supplying us with large quantities of human α -thrombin.

It is an honor and a privilege for me to thank Professor Robert Huber of the Max-Planck-Institut für Biochemie in Martinsried, Germany for his role in this project. The invitation to come to his laboratory to collaborate on the structure solution of the complex, and his advice and guidance during my visit are greatly appreciated. I would also like to thank Drs. Wolfram Bode, Albrecht Messerschmidt, and Milton Stubbs, and Ms. Monika Schneider - all members of Professor Huber's laboratory - for their valuable advice and assistance during my stay.

Prior to returning to Michigan State to work on this project, I was very fortunate to have worked in two industrial protein crystallography laboratories where I was given the opportunity to learn and develop new skills in both biochemistry and crystallography. To Drs. Cele Abad-Zapatero and John Erickson of Abbott Laboratories, and Drs. Howard Einspahr and Keith Watenpaugh of The Upjohn Company, I am thankful for these opportunities and for their encouragement of my returning to Michigan State to continue work toward a Ph.D. degree.

I am fortunate to have had so many supportive family members and friends behind me in this effort. I am especially indebted to my wonderful wife Odette, and to my parents, my brother and sisters, and my Grandma for their continual love and encouragement.

Support by the National Institutes of Health is gratefully acknowledged.

TABLE OF CONTENTS

CHAPTER		PAGE
	LIST OF TABLES	ix
	LIST OF FIGURES	xiii
I	INTRODUCTION	1
	A. Thrombin	1
	B. Hirudin	13
II	EXPERIMENTAL	27
	A. Sample Preparation	27
	B. Sample Molecular Weight Estimation	28
	C. Crystallization	30
	D. Crystal Characterization	32
	E. 2.8 Å Resolution Diffractometer Intensity Data Collection	37
	F. 2.3 Å Resolution Area Detector Data Collection . . .	51
III	DATA REDUCTION	62
	A. Converting the 2.8 Å Resolution Diffractometer Intensity Data to Structure Amplitudes	62
	B. Post-Data Collection Processing of the 2.3 Å Resolution Area Detector Structure Factors	68
IV	MOLECULAR REPLACEMENT CALCULATIONS	76
	A. Introduction	76
	B. The Rotation Function Calculations	79
	C. The Translation Function Calculations	81
	D. The Initial Hirudin-Thrombin Electron Density Map Calculation	84
V	MODEL-BUILDING AND EREF REFINEMENT	89
VI	MODEL-BUILDING AND PROLSQ REFINEMENT	95
VII	RESULTS AND DISCUSSION	120
	A. The Hirudin Structure	120
	B. The Hirudin-Thrombin Interaction	139
	C. Crystal Packing	178

CHAPTER		PAGE
VIII	A COMPARISON OF THE THROMBIN STRUCTURE IN THE HIRUDIN-THROMBIN COMPLEX AND IN PPACK-THROMBIN	185
	LIST OF REFERENCES	218

LIST OF TABLES

TABLE		PAGE
1	The Primary Sequence of Human α -Thrombin. The amino acid type is indicated using the standard single letter code. The numbering of residues is based on topological similarities with chymotrypsinogen. A number followed by an alphabetic character indicates an insertion. CP represents a cis-proline. (Numbering taken from [6].)	4
2	Summary of Crystal Data of rHV2-K47-Human α -Thrombin	34
3	The Average Cell Parameters of the HTN2 (3.6-2.8 Å) and of the HTN3 (2.8 Å) Crystals. Cell lengths in Å units, angles in degrees	45
4	Reflection Statistics on the Hirudin-Thrombin FAST Data Set Evaluated by MADNES and Corrected using ABSCOR	74
5	Results of the Translation Search	85
6	Course of Model-Building and EREF Refinement	91
7	Final Model Parameters of the EREF-refined rHV2-K47 Human α -Thrombin Structure	92
8	Summary of Model-Building and PROLSQ Refinement	100
9	Summary of Hirudin-Thrombin PROLSQ Refinement Restraint Parameters	102
10	Final Model Parameters of the PROLSQ-refined Hirudin-Thrombin Complex	103
11	The R-factor vs. Resolution for the Final Hirudin-Thrombin Model Parameters	110
12	Average Individual B-factor and Average Occupancy Factor for the Hirudin-Thrombin Model Parameters	112
13	Secondary Structural Elements of Hirudin	123
14	Sulfur-Sulfur Distances (Å) in Hirudin. A '*' indicates a disulfide bond	125

THE

13

14

15

16

17

18

19

20

21

22

23

TABLE	PAGE
15	Intramolecular Hydrogen Bonds in Hirudin. Hydrogen atoms are assigned geometrically idealized positions. Donor atoms are denoted 'D' and acceptor atoms 'A'. . . . 130
16	Dihedral Angles and Dihedral Energies of the Disulfide Bridges in Hirudin. A '*' indicates the energy value was excluded from the calculation of the mean 137
17	Intermolecular Hydrogen Bonds in the Hirudin-Thrombin Complex. Hydrogen atoms are assigned geometrically idealized positions. Donor atoms are denoted 'D' and acceptor atoms 'A'. A protonated amino terminus is designated 'N+' and an unprotonated amino terminus is designated 'N' for Ile1'. A '*' indicates the mean and standard deviation were calculated using values for an unprotonated Ile1' amino terminus 145
18	Hirudin-Water Molecule Hydrogen Bonds in the Complex. Hydrogen atoms are assigned geometrically idealized positions. Donor atoms are denoted 'D', acceptor atoms 'A' 151
19	Solvent-Bridged Hirudin-Thrombin Hydrogen Bond Interactions. The protein atoms involved in the interaction are represented as 'A' and 'B', and the mediating water 'W' 152
20	Solvent-Bridged Intramolecular Hydrogen Bond Interactions in Hirudin. The protein atoms involved in the interaction are represented as 'A' and 'B', and the mediating water 'W' 152
21	Hirudin α -Thrombin Interactions. Numbers are intermolecular contacts < 4.0 Å; parentheses indicate ion pairs; a dash followed by a number is a hydrogen bond; hydrophobic or neutral contacts are underlined in bold 169
22	Solvent Accessible Surface Area of Thrombin Residues Alone and in Complex with Hirudin. Abbreviations used: TASA= total atom surface accessibility; MASA= main chain atom surface accessibility; SASA= side chain atom surface accessibility 173
23	Solvent Accessible Surface Area of Hirudin Residues Alone and in Complex with Thrombin. Abbreviations used are as defined in Table 22 173
24	Percent Loss of Hirudin Solvent Surface Accessibility Due to Complexation with Thrombin. Results presented with regard to hirudin residue

	ranges. Abbreviations used are as defined in Table 22	174
25	Residue Surface Accessibility of Free Hirudin. Conformation of hirudin as it appears in the complex. Abbreviations used: SASAt= theoretical side chain atom surface accessibility. All other abbreviations used are as defined in Table 22. All surface accessibility results reported in Å ² . SASA/SASAt ratios indicate how the actual side chain accessibility compares to the maximum possible. TASA/SASAt ratios greater than one indicate that there is at least some main chain accessibility. These ratios are not given for glycines as glycine has no non-hydrogen side chain atoms. A '*' preceding a line indicates that the values given are susceptible to error due to disorder in or near these residues. A '+' preceding a line indicates a residue which has at least one direct contact of less than 4.0 Å with thrombin	176
26	Hydrogen Bonds Between Symmetry-Related Molecules. Hydrogen atoms are assigned geometrically idealized positions. Donor atoms are denoted 'D' and acceptor atoms 'A'. The symmetry molecule number refers to the symmetry equivalent position of the symmetry-related molecule in the interaction, as given below. The symmetry-related molecule residue number is preceded by a '*'	182
27	Results of an Optimal Superposition of the Thrombin Structures of the Complex and PPACK-Thrombin [6] using the Program MOLFIT. Fitting based on N, CA, C Atoms. Results expressed in Å units as root mean square deviations (r.m.s.d.). Numbers in the '{}' brackets represent the number of atom pairs contributing to the r.m.s.d	186
28	Structural Differences Between the Thrombin Structures of the Complex and PPACK-Thrombin [6] due to Residues Poorly Defined in Electron Density or Tentatively Placed in Either Structure, or due to Alternate Interpretations of the Density Maps. Optimal superpositioning of thrombin molecules as in Table 27. Abbreviations: HT = hirudin-thrombin complex; PT = PPACK-Thrombin; MCA r.m.s.d. = main chain atom average root mean square deviation; SCA r.m.s.d. = side chain atom average r.m.s.d.; S = surface; AS = active site; conf. = conformation; rot. = rotation. Deviations > 1σ on main chain (0.45 Å) and/or side chain (1.43 Å) atoms are listed	191

29	Structural Differences Between the Thrombin Structures of the Complex and PPACK-Thrombin [6] due to Crystal Packing Interactions in the Complex. Optimal superpositioning of thrombin molecules as in Table 27. Abbreviations: as = aromatic stacking; hb = hydrogen bond; hip = hydrogen-bonding ion pair; hc = hydrophobic contact; pc = polar contact; sympos = symmetry molecule positions (see Table 26), wmi = water mediated interaction. All other abbreviations are defined in Table 28. Residues in brackets [] are not involved in crystal contacts but have significant conformational differences which could be due to packing interactions of nearby residues. A '*' denotes the average MCA and SCA r.m.s.d. for residues Asn143-Gln151. Table 32 gives a complete listing of the MCA and SCA r.m.s.d.'s for the residues. Deviations > 1 σ on main chain (0.45 Å) and/or side chain (1.43 Å) atoms listed	192
30	Structural Differences Between the Thrombin Structures of the Complex and PPACK-Thrombin [6] due to the Interaction with Hirudin. Optimal superpositioning of the thrombin molecules as in Table 27. All abbreviations are defined in Tables 28 and 29. Residues in brackets [] do not interact with hirudin, but have significant conformational differences which could be due to hirudin interactions of nearby residues. Deviations > 1 σ on main chain (0.45 Å) and/or side chain (1.43 Å) atoms listed	193
31	Structural Differences Between the Thrombin Structures of the Complex and PPACK-Thrombin [6] due to Surface Residue Conformational Variability. Optimal superpositioning of the thrombin molecules as in Table 27. Abbreviations: fs = free sidechain; ident. = identical. All other abbreviations defined in Tables 28 and 29. Deviations > 1 σ on main chain (0.45 Å) and/or side chain (1.43 Å) atoms listed	194
32	Structural Differences Between the Thrombin Structures of the Complex and PPACK-Thrombin [6] in the Vicinity of the 149A-149E Insertion Loop. Optimal superpositioning of the thrombin molecules as in Table 27. Abbreviations: r.m.s.d. = root mean square deviation; MCA r.m.s.d. = average main chain atom r.m.s.d.; SCA r.m.s.d. = average side chain atom r.m.s.d. All residues listed have deviations > 1 σ on main chain (0.45 Å) and/or side chain atoms (1.43 Å)	196

FOR

:

2

3

LIST OF FIGURES

FIGURE		PAGE
1	The Prothrombinase Complex. Prothrombin (PT) is composed of fragment 1 (F1), fragment 2 (F2), and prethrombin 2 (Pre-2); Factor Va = Va; factor Xa = Xa; Ca ⁺² ions are represented by filled circles binding to γ -carboxyglutamic acid residues	2
2	The Products of PT Activation. CHO represents a carbohydrate side chain; "a", "b", and "c" represent the sites of cleavage of bovine and human PT; an additional cleavage, "b'" occurs in human PT (taken from [3])	3
3	The Polysaccharide of Human α -Thrombin [8] and its Components: N-Acetylneuraminic Acid (NeuNAc), N-Acetylglucosamine (GlcNAc), Mannose (Man), Galactose (Gal), and Fucose (Fuc)	6
4	Schematic Drawing of the Fibrinogen Molecule [9]. The symmetric molecule is composed of a dimeric central domain containing the amino-termini of all six chains (α , β , γ), two connecting coiled coils, two terminal domains, and two A α polar appendages. Fibrinopeptides A and B (FPA and FPB) are at the amino-terminal ends of the α and β chains. Four carbohydrate clusters (CHO) occur, and are located on each γ chain near the central domain and on the β chains of each terminal domain. Primary cross-linking sites (XL) can be found near the carboxy-termini of the γ chain and in the A α polar appendages	7
5	The Central Bioregulatory Functions of α -Thrombin in Hemostasis. Coagulation factor zymogens are identified by roman numerals; an "a" following a numeral indicates an activated enzyme. Primes and double primes are used to indicate active and inactive forms of factors of V and VIII. Abbreviations: F1.2 = prothrombin fragment 1.2; FpA = fibrinopeptide A; FpB = fibrinopeptide B; PL = phospholipid; and TPN = thromboplastin (taken from [18])	9
6	Amino Acid Sequences of the Three Natural Hirudin Variants: HV1, HV2, and HV3. The boxes	

FIGURE

7

8

9

10

11

12

13

14

15

	delineate regions of homology and the stars (*) refer to variant residues. HV3 contains an additional residue at position 63. Y' indicates a sulfated tyrosine (taken from [66])	14
7	The Disulfide Bond Connectivity Patterns of Hirudin and Epidermal Growth Factor [67]	16
8	The Sequence of Recombinant Hirudin Variant 2-Lysine 47 (rHV2-K47)	18
9	Stereo Drawing of Nine Optimally Superimposed DISMAN Distance Geometry N, CA, C Structures from the 2D NMR Constraint Data of Haruyama and Wuthrich [80]. Disorder exists in the residue ranges 1-3, 31-36 and 49-65 (taken from [80])	19
10	Human α -Thrombin and its Proteolytic Derivatives. The numbering of residues is taken from Table 1. The A-chain and B-chain of thrombin are designated by "A" and "B", respectively. CHO represents the polysaccharide linked to nitrogen ND2 of Asn(N)60G. The location of the catalytic triad residues is noted (H57, D102, and S195). Excised proteolytic fragments are denoted by broken lines	21
11	100Log(Rfx100) vs. %T Gel Concentration. The slope for each line gives the retardation coefficient (Kr) for the corresponding protein oligomer. Abbreviations: L(lysozyme), CA(carbonic anhydrase), CEA(chicken egg albumin), BSA-M(bovine serum albumin-monomer), BSA-D(bovine serum albumin), HT(hirudin-thrombin)	29
12	Log(-Kr) vs. Log(Molecular Weight). Linear regression best fit line to the standard protein retardation coefficient data. Abbreviations used have the same meaning as in Figure 11. The extrapolated molecular weight estimate for the hirudin-thrombin complex is 38.5 kD	31
13	Photograph of an rHV2-K47-Human α -Thrombin Crystal. Crystal size is approximately 1.0 x 0.4 x 0.4 mm. Note the trace of the seed in the center of the crystal	33
14	hk0 Precession Photograph of a Hirudin-Thrombin Crystal. The resolution limit at the edge of the photo is 3.7 Å. (μ angle = 12.0°)	35
15	Morphology of an Idealized Hirudin-Thrombin Crystal	38

FIGURE

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

FIGURE		PAGE
16	A Hirudin-Thrombin Crystal Mounted in a Capillary for X-ray Crystallographic Analysis	39
17	The Goniostat of a Four-Circle Diffractometer [111]	41
18	Axial Intensity Distributions of the Recombinant Hirudin-Human α -Thrombin Crystals	43
19	The ω -Profile of (6,16,17) Taken Before the 3-D HTN2 Data Collection. Peak width at half-height approximately 0.18°	48
20	Intensity Absorption Plot of (0,0,8) Taken Before the HTN3 3.56 Å Resolution Data Collection. $I(\phi)$ in total counts/measurement. The 45° ϕ range employed in the data collection indicated by "D.C."	50
21	Drawing of an Enraf-Nonius Diffractometer with a Conventional Detector [116]	52
22	A Flow Chart of Operations in a MADNES Area Detector Data Collection	55
23	Dependence of X-ray Decay Correction with Scattering Angle for the HTN2 Hirudin-Thrombin (3.6-2.8) Å Resolution Data Collection	65
24	The $\langle F_{obs} ^2 \rangle$ versus $\langle 2\theta \rangle$ Distribution for the 2.8 Å Resolution Hirudin-Thrombin Diffractometer Data Set	69
25	Plot of Sim weighting factor W versus X where $W = I_1(X)/I_0(X)$ and $X = 2 F_{obs}(h) F_{cal}(h) /I_u$ (taken from [137]). $I_1(X)$ and $I_0(X)$ are first- and zero-order modified Bessel functions, respectively. I_u is an estimate of the contribution from the unknown structure, or that which is not accounted for in the model	88
26	Stereoview of the First Unit of Thrombin Carbohydrate Modeled into Map Density. Displayed in bold is N-acetylglucosamine (NAG1) linked to the side chain atom ND2 of Asn60G. Map density is from the final 2Fo-Fc map used for model-building; contours at 0.8σ . Atomic positions are from the final coordinates	98
27	Histogram of ω Angles in the Hirudin-Thrombin Complex	107
28	The Ramachandran Plot of Hirudin-Thrombin. Only non-glycine residue angles are displayed	108

81

82

83

84

85

86

87

88

89

90

91

29	Plot of R-factor vs. Resolution for Hirudin-Thrombin. Large filled circles represent the data points of the complex. The small filled circles represent structure factor amplitude discrepancy data due to specified average coordinate errors (Δr) according to Luzzati [148]	111
30	Plot of the Average Individual B-factor of Main and Side Chain Atoms for the Protein Atoms in the Complex. The $\langle Bi \rangle$ for main and side chain atoms is represented by '+' and 'X' symbols, respectively. Main chain atom $\langle Bi \rangle$ values connected by a line. The break in $\langle Bi \rangle$ in hirudin corresponds to the completely undefined segment Ser32'-Lys35'. Bold arrows point to main and side chain $\langle Bi \rangle$ values of Cys residues engaged in disulfide linkages	114
31	Histogram of the Individual B-factors of the Water Molecules in the Hirudin-Thrombin Structure.	117
32	Histogram of the Occupancy Factors of the Water Molecules in the Hirudin-Thrombin Structure. All water molecules in the last occupancy shell, 1.0 - 1.1, have an occupancy of 1.0	118
33	Stereoview of the Folding of Hirudin in the Complex. N, CA, C atoms only; hirudin disulfides in bold; disordered or poorly defined regions indicated by dashed lines	121
34	The Sequence of Recombinant Hirudin Variant 2-Lysine 47 (rHV2-K47). Isolated loop designated 'A' and loop segments comprising the two interconnected loops designated 'B', 'C', and 'D'.	122
35	Stereoview of the Cystine Disulfide Core of Hirudin. Disulfide linkages in bold	124
36	Stereoview of the Folding of NH ₂ -terminal Domain of Hirudin in the Complex. N, CA, C atoms only; designations 'A', 'B', 'C', and 'D' correspond to those of Figure 34; disorder indicated by dashed lines	127
37	Stereoview of the $\beta 1$ Antiparallel β -Strand in Hirudin. Hydrogen bonds indicated by dashed lines	128
38	Stereoview of the $\beta 2$ Antiparallel β -Strand in Hirudin. Hydrogen bonds indicated by dashed lines	129

NOTE

39

40

41

42

43

44

45

46

47

48

39	Stereoview of the Intramolecular Lys47' Hydrogen Bonds in the Hirudin N-terminal Domain. CA structure of residues 1'-48' including disulfides (disulfide linkages in bold), and hydrogen bonding residues (bold). The disorder in residues 32'-35' as well as the hydrogen bonds are indicated by dashed lines	132
40	Stereoview of the Polyproline II Helix in Hirudin. N, CA, C atoms indicated in bold	133
41	Stereoview of the Disulfides in the Crystal Complex and in the NMR Hirudin Structure [79]. NMR hirudin disulfide side chains indicated by dashed lines; disulfide linkages in bold	136
42	Stereoview of the T4 Type III Helical Reverse Turn in Hirudin. Hydrogen bonds indicated by dashed lines; main chain atoms in bold; Gln65' also included	140
43	Stereoview of the CA Structure of the Hirudin-Thrombin Complex. Hirudin CA and disulfides of hirudin and thrombin in bold. N- and C-terminals of thrombin designated	141
44	Stereoview of PPACK in the Active Site Region of Thrombin. PPACK, from the PPACK-thrombin structure [6], optimally superimposed into thrombin active site based on N, CA, C fitting of PPACK-thrombin and hirudin-thrombin complex thrombin structures; PPACK in bold	143
45	Stereoview of the Interaction of the Hirudin N-terminal Tripeptide in the Active Site of Thrombin. Hirudin N-terminal tripeptide in bold; hydrogen bonds of the N-terminal nitrogen indicated by dashed lines.	144
46	Stereoview of Hydrophobic Contacts of the N-terminal Hirudin Domain with Thrombin. Hirudin residues in bold. Residues in figure: Ile174, Pro60C, Leu13', Val21', and Pro46'.	148
47	Stereoview of Electrostatic Interactions of the N-terminal Hirudin Domain with Thrombin. Hirudin residues in bold. Hydrogen bonds indicated by dashed lines, and ion pairs by +/-	149
48	Stereoview of a Hirudin-Thrombin Water Mediated Interaction. Hirudin residues in bold. Hydrogen bonds indicated by dashed lines, ion pairs by +/-, the water molecule by *	153

2

3

5

52

53

54

55

56

57

58

49	Stereoview of Hirudin Pro46'-Lys47'-Pro48' and the Hirudin N-Terminus-Thrombin Active Site Interaction. Hirudin residues in bold. Hydrogen bonds indicated by dashed lines, ion pairs by +/-, and water molecules by *	154
50	Charged Residues of the Anion-Binding Exosite Region of Thrombin. Thrombin CA structure only with thrombin disulfides and charged anion-binding exosite residues in bold	156
51	Basic Residues of the Postulated Heparin Binding Site of Thrombin. Thrombin CA structure only with thrombin disulfides and basic residues of heparin binding site in bold	157
52	Hirudin(Glu49'-Ser50'-His51')-Thrombin Interactions. Hirudin residues in bold. Hydrogen bonds indicated by dashed lines, ion pairs by +/-, and water molecules by *	159
53	Electrostatic Hirudin(Asp55'-Gln65')-Thrombin Interactions. Hirudin residues in bold. Hydrogen bonds indicated by dashed lines, ion pairs by +/-, and water molecules by *	160
54	Hirudin Phe56' in a Hydrophobic Thrombin Cavity. Phe56' in bold	162
55	Hirudin-Thrombin Interactions near the Type III Helical Reverse Turn of Hirudin. Hirudin residues in bold, with N, CA, C of hirudin bolder. Ion pairs indicated by +/-	164
56	Space-filling Drawing of the rHV2-K47-Human α -Thrombin Complex. Hirudin in sky blue	168
57	The Sequence of Human α -Thrombin. Insertion residues with respect to the sequence of chymotrypsinogen [6] are graphically indicated as protrusions from the linear sequence chain. Disulfide linkage connections are indicated, and insertion residues possessing contacts of $< 4.0 \text{ \AA}$ with hirudin in the hirudin-thrombin complex are circled	171
58	Stereoview of a Hydrophobic Crystal Packing Interaction at a Two-Fold Symmetry Axis. Top: Molecules involved are at symmetry positions 1 and 5 (relation 42, text). Disulfides and hirudin in bold. Two-fold axis indicated. The hydrophobic interactions occur in the circled region and are detailed below. Bottom: Nonpolar/aromatic residues	

FIGURE

59

60

61

62

63

64

- which comprise the hydrophobic cluster. Groups involved: **Leu14G**, **Tyr14J**, **Ile14K**, **Leu129C**, **Tyr134**, **Pro204**, and **Phe204A** (all bold) and the corresponding symmetry-equivalent groups 179
- 59 Stereoview of an Electrostatic Crystal Packing Interaction at a Two-Fold Symmetry Axis. Top: Molecules involved are at symmetry positions 1 and 6 (relation 42, text). Hirudin and all disulfides are in bold. Two-fold axis indicated. The electrostatic crystal packing interactions are within the circled region and are detailed below. Bottom: The double hydrogen-bonding ion pairs involved in the crystal contacts. Dashed lines indicate hydrogen bonds; hirudin glutamate residues are in bold. Two-fold symmetry axis offset approximately 30 degrees from the perpendicular to the plane of the paper 181
- 60 Stereoview of a Crystal Packing Interaction Involving the Thrombin Lys145-Gly150 Undecapeptide Loop. The crystal contact involves molecules at symmetry positions 1 and 7 (relation 42, text). Residues involved: **Lys145**, **Trp148**, **Thr149**, **Asn149B**, **Val149C**, **Gly#1D**, **Ala#1B**, **Leu#123**, **Lys#235**, **Gln#244**, and **Phe#245**. N, CA, C atoms and residues of the Lys145-Gly150 loop engaged in hydrogen bond interactions in bold; hydrogen bonds indicated by dashed lines 184
- 61 Stereoview of the Optimally Superimposed Thrombin CA Structures of the Complex and PPACK-Thrombin. Superpositioning corresponds to the MOLFIT N, CA, C fitting results of Table 27. Residues of the complex in bold 187
- 62 Average R.M.S. Deviation of Main Chain Atoms versus Residue Number for the Superimposed Thrombin Molecules of the Complex and PPACK-Thrombin. Superpositioning as in Figure 61. The horizontal lines represent 1σ and 2σ deviations. Most of residues with deviations $> 1\sigma$ are identified 189
- 63 Average R.M.S. Deviation of Side Chain Atoms versus Residue Number for the Superimposed Thrombin Molecules of the Complex and PPACK-Thrombin. Superpositioning as in Figure 61. The horizontal lines represent 1σ and 2σ deviations. Many of the residues with deviations $> 1\sigma$ are identified 190
- 64 Stereoview of the Superimposed Residues Lys236-Phe245 of the Complex and PPACK-Thrombin. Superpositioning as in Figure 61. Residues of the complex in bold. Symmetry molecule interactions of

	the complex displayed; symmetry molecules indicated by a '#' preceding the residue number. Hydrogen bonds indicated by dashed lines	198
65	Stereoview of the Superimposed Residues Glu146-Gly150 of the Complex and PPACK-Thrombin. Superpositioning as in Figure 61. Residues of the complex in bold	200
66	Stereoview of the Superimposed Residues Glu146-Gly150 of PPACK-Thrombin and the Complex, including the Hirudin N-Terminal Pentapeptide. Superpositioning as in Figure 61. Residues of PPACK-thrombin in bold. Thrombin catalytic triad of the complex displayed	202
67	Stereoview of the Superimposed Residues Lys110-Pro111 of the Complex and PPACK-Thrombin. Superpositioning as in Figure 61. Residues of the complex in bold. Peptide bond preceding Pro111 is cis- in PPACK-thrombin, trans- in the complex	205
68	Stereoview of the Superimposed Residues Lys202-Arg206 of the Complex and PPACK-Thrombin. Superpositioning as in Figure 61. Residues of the complex in bold	207
69	Stereoview of Superimposed Active Site Residues of PPACK-Thrombin and the Complex, including the Hirudin N-Terminal Tripeptide. Residues of the complex in bold. Superpositioning as in Figure 61. Hydrogen bonds indicated by dashed lines	208
70	Stereoview of the Superimposed Anion-Binding Exosite Residues Arg75-Lys81 of the Complex and PPACK-Thrombin. Additional residues of the complex (thrombin Ala113, hirudin Glu57'-Glu58', and the water molecule W467) included. Residues of the complex in bold. Superpositioning as in Figure 61. Hydrogen bonds indicated by dashed lines	211
71	Stereoview of Residues Leu41 and Phe60H-Leu64 of the Complex and PPACK-Thrombin. Residues of the complex in bold. Superpositioning as in Figure 61. Hydrogen bonds indicated by dashed lines	213

A. Th

a

genera

this s

activ

form t

compos

tembra

ions.

and h

activa

and "c

a-thro

sites

a-thro

residu

H

chain

36,600

The A

throug

posse

Thro-

high

(Fac

acti

I INTRODUCTION

A. Thrombin

α -Thrombin is a glycoprotein serine proteinase [1] which is generated in the penultimate step of the blood coagulation cascade. In this step, prothrombin (PT), a zymogen of molecular weight 74,000, is 'activated' by limited proteolysis in the prothrombinase complex [2] to form thrombin. The complex, schematically represented in Figure 1, is composed of PT substrate, the proteolytic enzyme Factor Xa, the membrane-bound cofactor Factor Va, the phospholipid surface, and Ca^{+2} ions. A representation of the prothrombin activation products in bovine and human systems is shown in Figure 2 [3]. During bovine PT activation, Factor Xa cleaves PT at Arg-Thr and Arg-Ile bonds, sites "b" and "c" respectively in Figure 2, to produce PT fragment 1-2 and α -thrombin. In the human system, α -thrombin additionally cleaves at sites "a" and "b'" to produce PT fragment 1, PT fragment 2, and α -thrombin. The result is that the A chain of human thrombin is 13 residues shorter than the A chain of bovine thrombin.

Human α -thrombin, the thrombin form used in this study, is a two chain molecule which possesses a molecular weight of approximately 36,600 [4,5]. The sequence of human α -thrombin is given in Table 1 [6]. The A chain contains 36 residues and is covalently linked to the B chain through a disulfide (1-122). The B chain is 259 residues long and possesses 3 intrachain disulfide linkages (42-58, 168-182, 191-220). Thrombin is a serine proteinase by virtue of the B chain, which is highly homologous to other coagulation/fibrinolytic serine proteinases (Factor IXa, Factor Xa, protein C, urokinase, tissue plasminogen activator, and plasmin) and is approximately 50% homologous with the

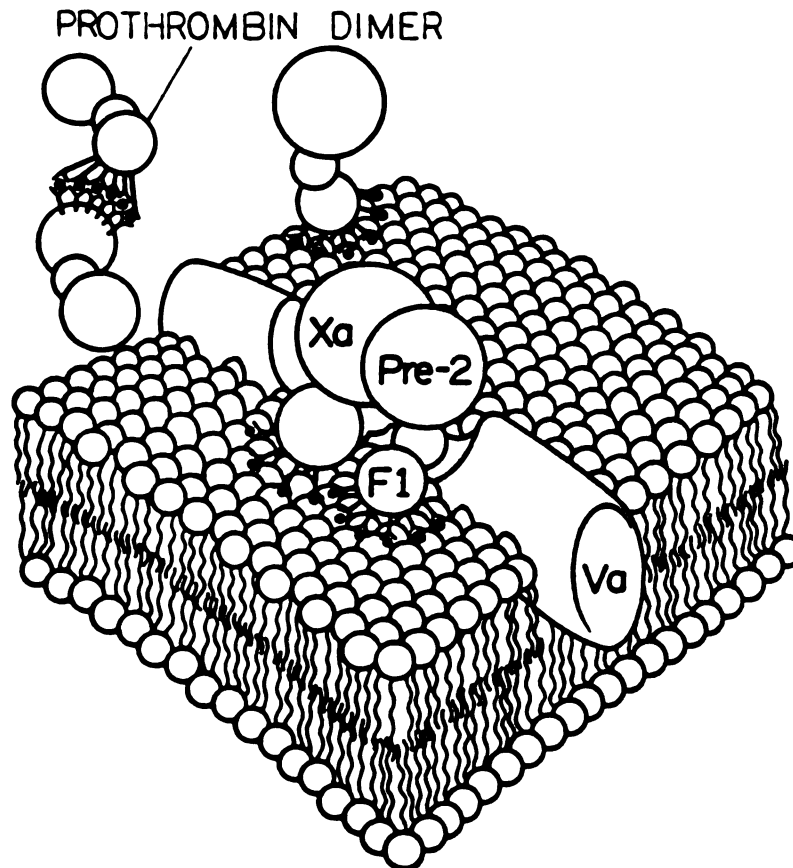


Figure 1. The Prothrombinase Complex. Prothrombin (PT) is composed of fragment 1 (F1), fragment 2 (F2), and prethrombin 2 (Pre-2); Factor Va = Va; factor Xa = Xa; Ca^{+2} ions are represented by filled circles binding to γ -carboxyglutamic acid residues.

Proch

GA

←

←

Figur

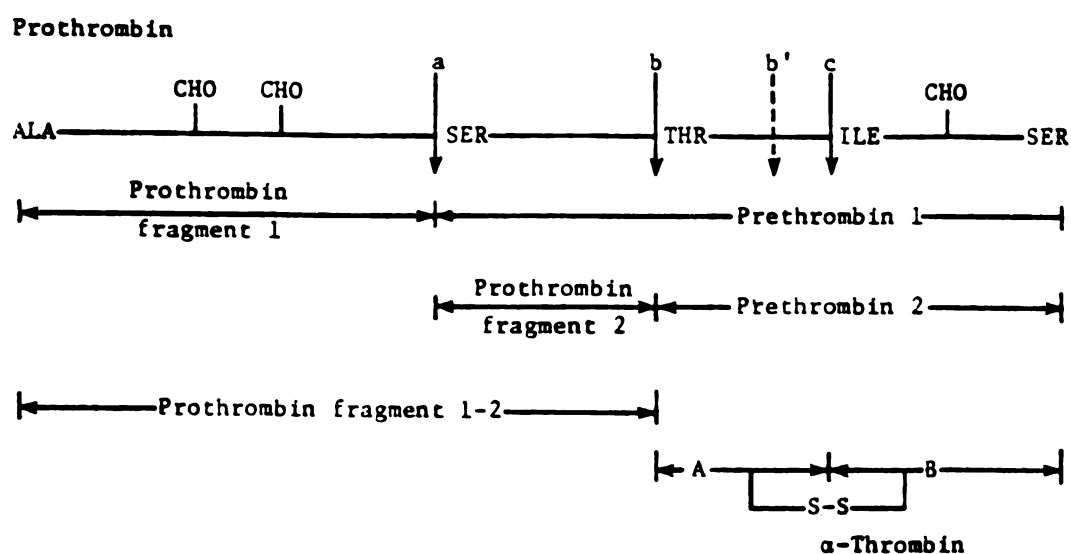


Figure 2. The Products of PT Activation. CHO represents a carbohydrate side chain; "a", "b", and "c" represent the sites of cleavage of bovine and human PT; an additional cleavage, "b'" occurs in human PT (taken from [3]).

Tab 1

H1

T1

L3

E1

S1

Har

I16

M26

K36

S45

A55

E60

V66

Y76

L85

N95

A10

P11

P12

Q13

W14

N14

Q15

P16

I17

K18

C19

M20

Q20

C22

T22

Q23

Table 1. The Primary Sequence of Human α -Thrombin. The amino acid type is indicated using the standard single letter code. The numbering of residues is based on topological similarities with chymotrypsinogen. A number followed by an alphabetic character indicates an insertion. CP represents a cis-proline. (Numbering taken from [6].)

Human α -Thrombin A-chain

T1H	F1G	G1F	S1E	G1D	E1C	A1B	D1A	C1	G2
L3	R4	P5	L6	F7	E8	K9	K10	S11	L12
E13	D14	K14A	T14B	E14C	R14D	E14E	L14F	L14G	E14H
S14I	Y14J	I14K	D14L	G14M	R15				

Human α -Thrombin B-chain

I16	V17	E18	G19	S20	D21	A22	E23	I24	G25
M26	S27	P28	W29	Q30	V31	M32	L33	F34	R35
K36	S36A	CP37	Q38	E39	L40	L41	C42	G43	A44
S45	L46	I47	S48	D49	R50	W51	V52	L53	T54
A55	A56	H57	C58	L59	L60	Y60A	P60B	P60C	W60D
D60E	K60F	N60G	F60H	T60I	E61	N62	D63	L64	L65
V66	R67	I68	G69	K70	H71	S72	R73	T74	R75
Y76	E77	R77A	N78	I79	E80	K81	I82	S83	M84
L85	E86	K87	I88	Y89	I90	H91	P92	R93	Y94
N95	W96	R97	E97A	N98	L99	D100	R101	D102	I103
A104	L105	M106	K107	L108	K109	K110	P111	V112	A113
F114	S115	D116	Y117	I118	H119	P120	V121	C122	L123
P124	D125	R126	E127	T128	A129	A129A	S129B	L129C	L130
Q131	A132	G133	Y134	K135	G136	R137	V138	T139	G140
W141	G142	N143	L144	K145	E146	T147	W148	T149	A149A
N149B	V149C	G149D	K149E	G150	Q151	P152	S153	V154	L155
Q156	V157	V158	N159	L160	P161	I162	V163	E164	R165
P166	V167	C168	K169	D170	S171	T172	R173	I174	R175
I176	T177	D178	N179	M180	F181	C182	A183	G184	Y184A
K185	P186	D186A	E186B	G186C	K186D	R187	G188	D189	A190
C191	E192	G193	D194	S195	G196	G197	P198	F199	V200
M201	K202	S203	P204	F204A	N204B	N205	R206	W207	Y208
Q209	M210	G211	I212	V213	S214	W215	G216	E217	G219
C220	D221	R221A	D222	G223	K224	Y225	G226	F227	Y228
T229	H230	V231	F232	R233	L234	K235	K236	W237	I238
Q239	K240	V241	I242	D243	Q244	F245	G246	E247	

digest

the

residue

cataly

planter

of the

of the

of the

network

round

telecom

-51 A

of the

fibers

Proto

form

clear

subse

relea

NHQ-t

of the

FPA-d

Fiber

Prodi

poly-

over

asso

digestive serine proteinases trypsin, chymotrypsin, and elastase [7]. The B chain of thrombin contains the three invariant catalytic site residues - His57, Asp102, and Ser195 (see Table 1) - commonly called the catalytic triad. Finally, the thrombin B chain also contains a single biantennary chain carbohydrate which is linked to Asn60G; the structure of the carbohydrate moiety is given in Figure 3 [8].

The ultimate step of blood coagulation is also the primary function of thrombin: the conversion of fibrinogen to fibrin. Fibrin forms the network which encapsulates the hemostatic plug or blood clot during wound healing. Fibrinogen is a large, dimeric six chain molecule with a molecular weight of approximately 340,000 and a overall length of about 450 Å [9]. A schematic drawing of fibrinogen is shown in Figure 4 [9].

Fibrin assembly proceeds in two stages [10]. In the first stage, fibrin monomers polymerize end-to-end to form fibrin I polymers or protofibrils. In the second stage, protofibrils associate laterally to form fibers. The first stage of fibrin assembly is initiated by the cleavage of two A α chains of fibrinogen at 16Arg-17Gly with the subsequent release of 2 molecules of fibrinopeptide A or FPA [11]. FPA release leads to an unfolding of a polymerization domain in the NH₂-terminal portion of fibrinogen and exposes a complementary portion of the γ chain, which exists in intact fibrinogen [12]. This FPA-deficient fibrinogen molecule is known as the fibrin monomer. Fibrin monomers polymerize via the available polymerization sites to produce fibrin I polymers or protofibrils. The resulting fibrin I polymers are comprised of fibrin monomers laid end-to-end with an overlap equal to about 225 Å or one-half the monomer length [13].

The second stage of fibrin assembly involves the lateral association of protofibrils to form fibers. A fiber may contain over

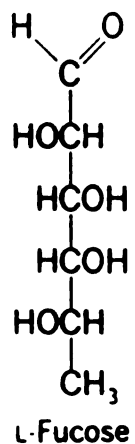
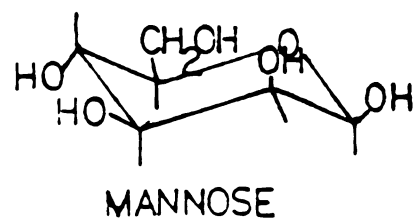
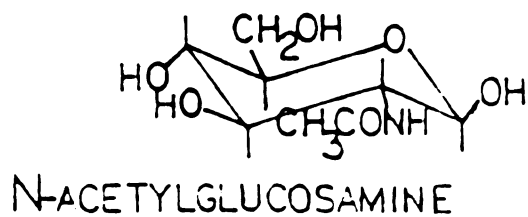
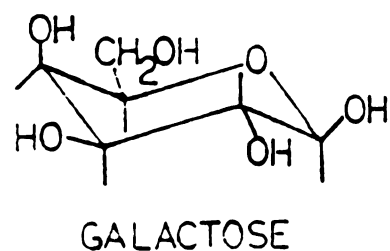
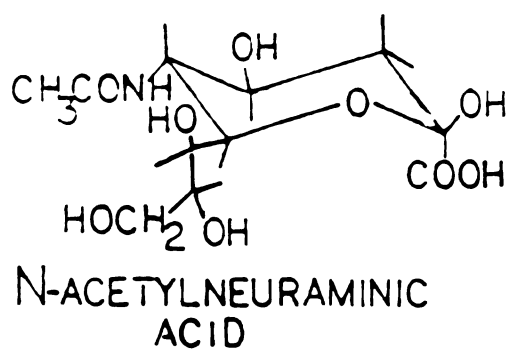


Figure 3. The Polysaccharide of Human α -Thrombin [8] and its Component Hexoses.

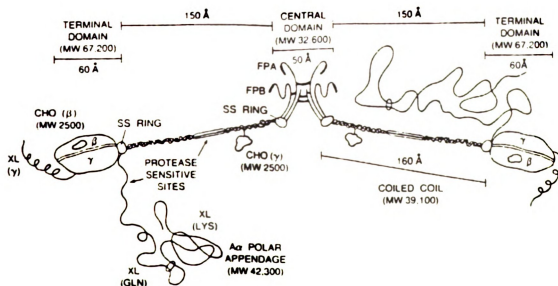


Figure 4. Schematic Drawing of the Fibrinogen Molecule [9].

The symmetric molecule is composed of a dimeric central domain containing the amino-termini of all six chains (α , β , γ), two connecting coiled coils, two terminal domains, and two A α polar appendages. Fibrinopeptides A and B (FPA and FPB) are at the amino-terminal ends of the α and β chains. Four carbohydrate clusters (CHO) occur, and are located on each γ chain near the central domain and on the β chains of each terminal domain. Primary cross-linking sites (XL) can be found near the carboxy-termini of the γ chain and in the A α polar appendages.

11

pr

[1

the

the

for

unf

int

str

rec

of

is

Fac

a-l

acti

the

hemo

bloo

(18)

posi

regi

hem

sev

the

(a

pe

100 protofibrils [14] and it is these fibers which join irregularly to produce the gel-like mass characteristic of in vitro clotted fibrin [15]. The lateral association of fibrin I polymers is accompanied by thrombin cleavage of the B β chains at 14Arg-15Gly in these polymers and the subsequent release of fibrinopeptide B or FPB. The FPB-deficient fibrin I polymers are known as fibrin II polymers. Release of FPB unfolds another unique polymerization site [11] and results in an interaction of polymeric species and in polymers and fibers of increased strand width. It has been demonstrated that although FPB release is not necessary for fiber formation or stage II fibrin assembly, the release of FPB results in more rapid fiber formation [10]. Once fibrin assembly is complete, the network is covalently stabilized by the action of Factor XIIIa (fibrin stabilizing factor), which introduces γ -glutamyl ϵ -lysyl cross-links in fibrin [16].

The role of thrombin in hemostasis is by no means limited to the activation of fibrinogen to clottable fibrin. Thrombin has been called the central bioregulatory enzyme in hemostasis [17]. Thrombin regulates hemostasis on three different levels - the fluid or plasma level, the blood cell or cellular level, and the blood vessel or vascular level [18]. Moreover, thrombin accomplishes these tasks by acting as a positive feedback activator of coagulation [7] as well as a negative regulator [19]. The central bioregulatory functions of thrombin in hemostasis are indicated in Figure 5 [18].

On the fluid or plasma level, thrombin proteolytically cleaves several zymogens to produce active enzymes or stimulate activation [17]. Thrombin activates Factor VIII (antihemophilic factor) [20] and Factor V (accelerator globulin) [21] via limited proteolysis which results in positive feedback to the coagulation pathway. Thrombin also activates

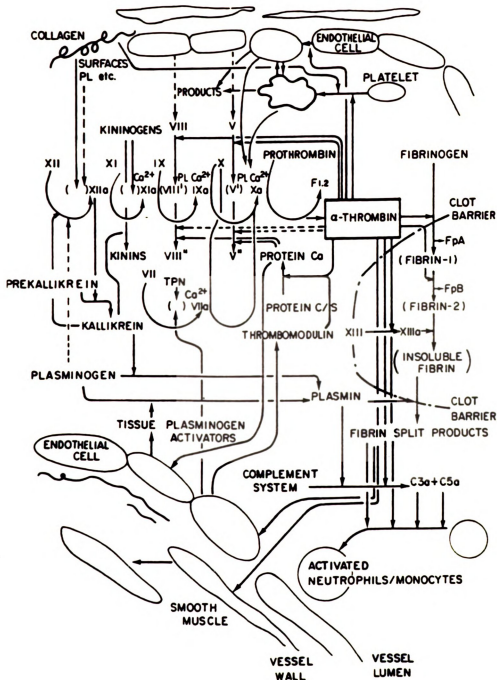


Figure 5. The Central Bioregulatory Functions of α -Thrombin in Hemostasis. Coagulation factor zymogens are identified by roman numerals; an "a" following a numeral indicates an activated enzyme. Primes and double primes are used to indicate active and inactive forms of factors of V and VIII. Abbreviations: Fl.2 = prothrombin fragment 1.2; FpA = fibrinopeptide A; FpB = fibrinopeptide B; PL = phospholipid; and TPN = thromboplastin (taken from [18]).

factor

to be

in the

entire

factor

present

factor

factor

factor

factor

factor

factor

factor

factor

factor

factor

factor

factor

factor

factor

factor

factor

factor

factor

factor

factor

factor

factor

factor

factor

Factor XIV (protein C) by proteolytic cleavage; Factor XIVa is believed to be responsible for the inactivation of Factors Va and VIIIa [22,23]; in this capacity thrombin is no longer a procoagulant but rather an anticoagulant. Thrombin cleavage of Factor XIII (fibrin stabilizing factor) produces Factor XIIIa'b which becomes an active enzyme in the presence of Ca^{+2} ions [24]. The conversion of fibrinogen to clottable fibrin also occurs at this level.

The endogenous proteinase inhibitors antithrombin III, heparin cofactor II, α 1-proteinase inhibitor, and α 2-macroglobulin all inhibit thrombin on the plasma level [25]. Antithrombin III is the primary inhibitor of thrombin in plasma. Antithrombin III and heparin cofactor II are unique in that they display accelerated thrombin inhibition in the presence of heparin and other glycosaminoglycans [26,27]. The most potent naturally occurring plasma-level inhibitor of thrombin is hirudin, a 65-66 amino acid peptide isolated from the salivary glands of the European medicinal leech *Hirudo medicinalis* [28].

On the cellular level, thrombin is an agonist for a variety of cellular activities in many tissues and cell types [29]. Thrombin-cellular interactions stimulate aggregation (platelets), chemotaxis (monocytes), contractility (smooth muscle), proliferation or cell growth (fibroblasts, macrophages, splenocytes), prostaglandin synthesis (platelets, endothelial cells, fibroblasts, neuronal cells), and secretion (platelets, endothelial cells). The induction of platelet aggregation and secretion is a very important procoagulant cellular function of thrombin. On the other hand, a particularly important anticoagulant thrombin-cellular interaction involves the thrombomodulin receptor on endothelial cells. Thrombomodulin is a 74,000 dalton protein which forms a stoichiometric 1:1 noncovalent complex of high

affini

by the

proteol

thromb

thromb

F

thromb

endoth

the

platelet

Willeb

inhibi

T

be de

carbon

[14.3]

prote

indio

acid

'spe

More

the

hydr

afte

try

aspa

Ger

the

affinity ($K_d = 0.48 \text{ nM}$) with thrombin. Moreover, thrombin is inhibited by thrombomodulin and this interaction also stimulates activation of protein C which can in turn inactivate factors Va and VIIIa. Thus, the thrombin-thrombomodulin interaction is yet another means by which thrombin negatively regulates coagulation.

Functions of thrombin on the vascular level also involve thrombin-cellular interactions [17]. Thrombin interacts with the endothelial cells of the blood vessel wall. This interaction results in the synthesis of prostacyclin [30], prostaglandin [30] and platelet-activating factor [31], as well as the release of von Willebrand factor [32], and tissue plasminogen activator (tPA) and its inhibitor [33].

The mechanism by which serine proteinases cleave peptide bonds can be described as a general base-catalyzed nucleophilic attack on the carbonyl carbon of the substrate by the hydroxyl oxygen of Ser195 [34,35]. The x-ray crystal structures of the digestive serine proteinases trypsin [36], α -chymotrypsin [37], and elastase [38] indicate that prior to peptide bond cleavage, the sidechain of an amino acid on the carbonyl-side of the bond to be cleaved fits into a 'specificity pocket' which brings the carbonyl carbon near to Ser195. Moreover, the nature of the binding pocket reflects the specificity of the enzyme [39]. The specificity pocket of α -chymotrypsin is large and hydrophobic; thus, α -chymotrypsin is well-suited to cleave peptide bonds after tryptophan, phenylalanine, tyrosine, and leucine sidechains. In trypsin, the specificity pocket is similar in size but it has an aspartate (Asp189) at the base of the specificity pocket as opposed to a serine. This aspartate deprotonates above about pH 5.0 and thus makes the cavity ideal for the binding of arginyl and lysyl sidechains. The

fact that thrombin is 'trypsin-like' in its ability to cleave Arg-X, Lys-X bonds and that it contains an aspartate at position 189 suggests that thrombin and trypsin share a similar specificity pocket [1].

The diverse functions and regulation thrombin provides in hemostasis, however, clearly indicate that it is a unique trypsin-like serine proteinase which requires more than a specificity pocket to govern its selectivity. While trypsin is omnivorous in its choice of protein substrates and in the number of peptide bonds it will attack [1], thrombin is highly selective in its choice of macromolecular substrates and interactions. Thrombin achieves such high selectivity by using active site region subsites, and exosites which are distinct from the active site [17,40-46]. Moreover, it is these exosites in addition to the catalytic site region which allow thrombin to function in hemostasis by means of enzymatic (proteolytic), nonenzymatic (hormonal), and coupled mechanisms [47]. Coupled mechanisms require the involvement of the thrombin active site and one or more exosites.

The high degree of specificity of α -thrombin interactions with macromolecular substrates is believed to result from binding at three distinct regions on the thrombin surface: the primary binding or specificity pocket [1], an apolar binding site adjacent to the catalytic site [40,41], and an anion-binding exosite which is responsible for fibrinogen recognition [42,44]. The anion-binding exosite appears to contain or be in the vicinity of the β -cleavage loop (residues Arg67-Arg77A [48]) of α -thrombin [44]. The integrity of this loop is important to the interaction of thrombin with fibrinogen, thrombomodulin, and protein C [49-52]. Additional secondary binding sites have been identified for heparin-binding [45] and growth factor activity [46].

relat

focus

climou

report

these

struct

propo

trypsi

a-Intro

referr

*ax-pl

Perman

Eiber

A reso

B. Bir

the

poten

a-Intro

diss

con

app

be

C. H

app

In an effort to truly understand the structure-function relationships underlying thrombin specificity, this enzyme has been the focus of x-ray crystallographic studies for some time [6,52-55]. Although crystals of bovine [52] and human thrombin [53] had been reported for well over ten years, no protein structure has resulted from these crystals. In the absence of definitive crystallographic structures, three-dimensional models of the thrombin B chain were proposed based on the sequence homology with bovine chymotrypsin and trypsin [56,57]. Recently however, nearly identical crystals of human α -thrombin inhibited by D-Phe-Pro-Arg-chloromethylketone (hereafter referred to as PPACK) were obtained by us [55] and independently at the Max-Planck-Institut für Biochemie in Martinsried, Germany [6]. The German research group led by Dr. Wolfram Bode and Professor Robert Huber eventually solved the structure of PPACK-inhibited thrombin at 1.9 Å resolution [6].

B. Hirudin

Hirudin is a small protein which occurs in the salivary glands of the the European medicinal leech, *Hirudo medicinalis* and is the most potent thrombin inhibitor known [28]. The complex of hirudin with α -thrombin is high-affinity, noncovalent, stoichiometric (1:1) and has a dissociation constant in the picomolar to femtomolar range [59,60].

Crude hirudin from the leech is a mixture of polypeptides containing 65 or 66 amino acids which possess a molecular weight of approximately 6900 daltons [61]. The first three hirudin sequences to be determined were classified as hirudin variant 1 [62], hirudin variant 2 [63], and hirudin variant 3 [64] based on their chronological appearance in the literature [65]; these sequences are often referred to

Figure 6. Amino Acid Sequences of the Three Natural Hirudin Variants: HV1, HV2, and HV3. The boxes delineate regions of homology and the stars (*) refer to variant residues. HV3 contains an additional residue at position 63. Y' indicates a sulfated tyrosine (taken from [66]).

as HIV

hirudi

variab

conten

(E). a

trypto

sequer

nomolo

glutar

Varia

(62.6

to

decre

(59.6

have

appro

patte

resid

chara

Cys2

can

homo

inhi

or s

coul

does

whic

as HV1, HV2, and HV3, respectively, and are given in Figure 6 [66]. The hirudin sequences are approximately 80% homologous and contain only 13 variable positions [67]. The amino acid compositions also reveal a high content of asparagine (N), glutamine (Q), aspartate (D) and glutamate (E), as well as the absence of arginine (R), methionine (M), and tryptophan (W) [67]. The COOH-terminal region of these hirudin sequences is rich in acidic residues; nearly half of the last 11 homologous amino acid positions are occupied by an aspartate or glutamate residue. A unique acidic residue found in all three hirudin variants is the sulfated tyrosine (Y') in the antepenultimate position [62,68,69]. This tyrosine acyl sulfate enhances the ability of hirudin to inhibit thrombin; desulfato-hirudins display three-to ten-fold decreased affinity for thrombin with respect to the sulfated forms [59,60,70,71]. Presently, twenty naturally-occurring hirudin sequences have been determined, and they display an overall sequence homology of approximately 66% [61].

Hirudin variants also share a common disulfide connectivity pattern. All hirudin molecules contain six conserved cysteine (C) residues in the NH₂-terminal region which are involved in 3 characteristic cystine linkages: Cys6-Cys14, Cys16-Cys28, and Cys22-Cys39 [72]. Interestingly, all known serine proteinase inhibitors can be grouped into one of ten structural families based on sequence homology, and the topological relationship among disulfide bonds and the inhibitor reactive site [73]. However, hirudin displays no topological or sequence homology to any known serine proteinase inhibitor and thus could represent a heretofore unknown family of inhibitors [72]. Hirudin does share some global features with epidermal growth factor (EGF), which is a structural motif found in numerous proteins [74]. Alignment

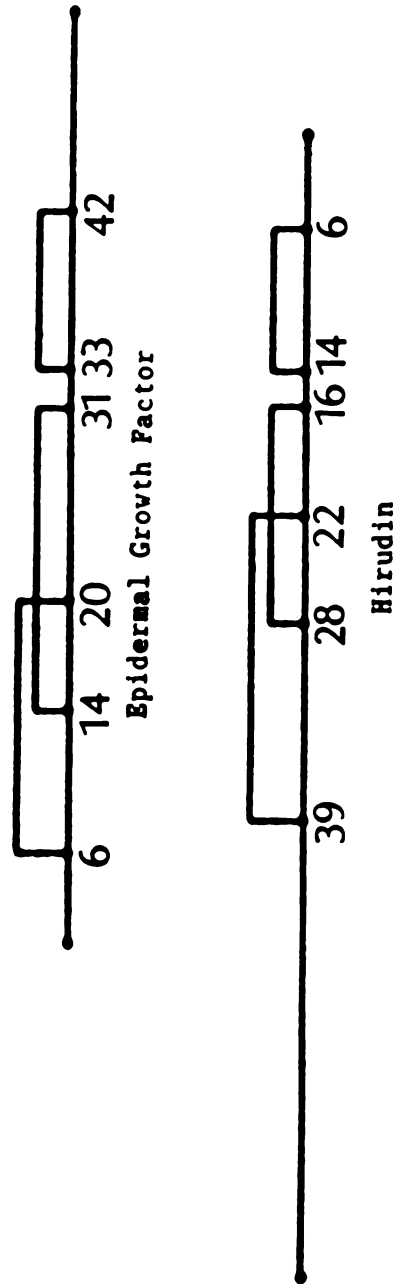


Figure 7. The Disulfide Bond Connectivity Patterns of Hirudin and Epidermal Growth Factor [67].

of the hirudin and EGF disulfide linkage patterns in opposite directions reveals a similar connectivity, as is seen in Figure 7 [67]. It is of note, however, that hirudin and EGF share no functional or sequence homology.

Because leeches contain mixtures of naturally occurring hirudin variants which are difficult to purify and isolate in large quantities, recombinant DNA technology has been employed to obtain homogeneous preparations in ample amounts. To date, five different microbial expression systems have been made to produce recombinant hirudin variant 1 (rHV1) [67,75-77] and two systems have been developed to produce recombinant hirudin variant 2 (rHV2) [63,65]. Recombinant hirudins differ from their natural counterparts in that they lack the sulfated tyrosine. The source of hirudin used in this study is the recombinant hirudin variant 2-Lys47 mutant (rHV2-K47) resulting from a yeast expression system [65]. The sequence and covalent structure of rHV2-K47 are given in Figure 8.

Secondary and tertiary structural information on hirudin in solution has been obtained independently by two research groups employing 2D NMR techniques to study recombinant hirudin variant 1 [78-80]. Qualitatively, the results of both groups are in general agreement. These studies revealed that hirudin is a two domain protein characterized by a compact, well-defined NH₂-terminal core and a flexible, disordered COOH-terminal tail. The core domain (residues 3-30 and 37-48) contains the cystine linkages, several turns and two antiparallel β -strands. Residues 31-36 within the NH₂-terminal domain and residues 49-65 of the COOH-terminal tail exhibited no preferred conformation in solution. A stereo drawing displaying nine optimally superimposed DISMAN distance geometry solutions resulting from the 2D

1

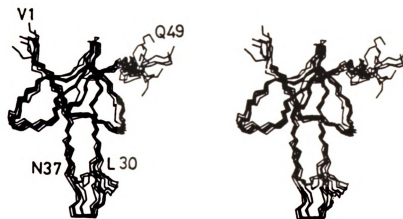


Figure 9. Stereo Drawing of Nine Optimally Superimposed DISMAN Distance Geometry N, CA, C Structures from the 2D NMR Constraint Data of Haruyama and Wuthrich [80]. Disorder exists in the residue ranges 1-3, 31-36 and 49-65 (taken from [80]).

WVP con

Hi

(30) a

dissoci

binding

site of

format

the ap

intrusi

histid

IV in

applic

E

thorpe

involve

B-thro

protec

figur

natio

B-Th

clear

dece

deci

sine

cova

pre

cle

at

NMR constraint data of Haruyama and Wuthrich is shown in Figure 9 [80].

Hirudin reacts rapidly with human α -thrombin ($k_{on} > 1 \times 10^9 \text{ M}^{-1}\text{s}^{-1}$) [59] and forms an amazingly high-affinity, noncovalent complex with a dissociation constant estimated at 80 pM [28] and 20 fM [59]. The apolar binding site and the anion-binding exosite or fibrinogen recognition site of thrombin appear to be important thrombin determinants in complex formation. Hirudin binding to human α -thrombin displaces proflavin from the apolar binding site [41]. In addition, the dissociation constant of hirudin and PPACK-thrombin, in which PPACK alkylates the active site histidine and appears to bind in the apolar binding site [40], is 1×10^6 higher than with α -thrombin [81]. Both of these observations implicate the apolar binding site in the hirudin-thrombin interaction.

Evidence supporting the importance of the anion-binding exosite of thrombin to its interaction with hirudin has come from binding studies involving the proteolytic derivatives of α -thrombin: β -, γ -, and ϵ -thrombin. Schematic drawings of human α -thrombin and these proteolytic derivatives, with the cleavages indicated, are given in Figure 10. β -Thrombin and γ -thrombin are produced by autolysis of native α -thrombin or by limited proteolysis with trypsin [82]. β -Thrombin generated by autolytic degradation is characterized by a cleavage at Arg77A followed by Arg67, and by the concomitant loss of the Ile68-Arg77A undecapeptide [48] (Figure 10 and Table 1). The β -thrombin derivative produced by trypsinolysis, designated β T-thrombin, contains a single cleavage at the Arg77A-Asn78 peptide bond [52]. The exact covalent structure of γ -thrombin appears to depend on the method of preparation. Boissel et al. [48] report γ -thrombin to possess cleavages at Arg67 and Arg77A, as in β -thrombin, and additional breaks at Arg126 and Lys149E, followed by the loss of the resulting 11- and

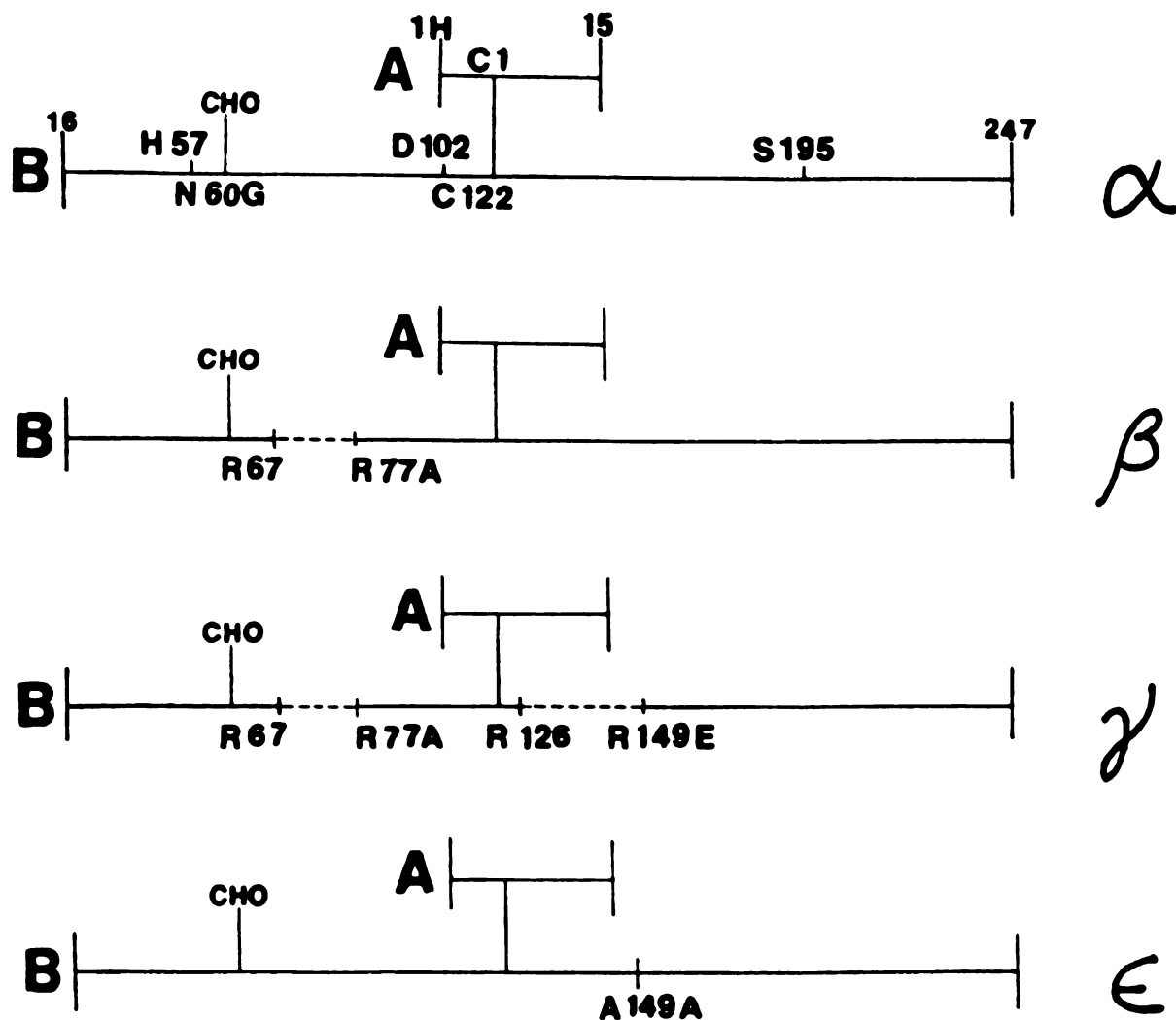


Figure 10. Human α -Thrombin and its Proteolytic Derivatives. The numbering of residues is taken from Table 1. The A-chain and B-chain of thrombin are designated by "A" and "B", respectively. CHO represents the polysaccharide linked to nitrogen ND2 of Asn(N)60G. The location of the catalytic triad residues is noted (H57, D102, and S195). Excised proteolytic fragments are denoted by broken lines.

31-

the

sit

with

(E)

with

of

exos

resp

disp

part

incr

mutat

seri

cont

inter

supra

corre

been

seri

inter

appe

tho

lys

(14)

num

the

31-amino acid peptides (Figure 10 and Table 1). Fenton et al. [82], on the other hand, have reported a γ -thrombin with only two major cleavage sites - Arg77A and Lys149E. ϵ -Thrombin results from limited proteolysis with elastase and contains a single cleavage at the Ala149A-Asn149B bond (Figure 10 and Table 1) [52]. Hirudin complexes with β -thrombin and γ -thrombin, proteolytic derivatives of α -thrombin which possess little or no fibrinogen activity (0.05%) due to a disrupted anion binding exosite [52], have significantly higher dissociation constants with respect to the α -thrombin-hirudin complex [81]. ϵ -thrombin, which displays only about 40% diminished fibrinogen activity with respect to native thrombin [52], in complex with hirudin has only a slightly increased dissociation constant [81]. Hirudin complexes with native human α -thrombin as well as with some of its chemical and proteolytic derivatives are not retained on nonpolymerized fibrin-agarose resin in contrast to free thrombin [44]. The importance of the hirudin interaction with thrombin at the anion-binding exosite is further emphasized by the observation that an antibody raised against a peptide corresponding to the β -cleavage site undecapeptide, Ile68-Arg77A, has been shown to inhibit hirudin binding to α -thrombin [83].

Hirudin is unique among macromolecular substrates for thrombin, and serine proteinase inhibitor proteins in general [84], in that interaction with the primary specificity pocket of thrombin does not appear to be crucial for inhibition. A hirudin lysyl residue was often thought to interact with the thrombin specificity pocket, especially Lys47; residues (39-48) of hirudin display 50% homology with residues (148-157) of PT, which contains an α -thrombin cleavage site utilized in human prothrombin activation [85] (cleavage site "a" in Figure 2). When the basic residues on hirudin were altered by site-directed mutagenesis

and

human

[66.6

minor

conse

inter

import

acidi

of th

flu re

of al

acids

of t

cause

incre

flu a

Desul

cons

ten-s

tail

stud

resid

fibr

as ve

1982-

of th

the

and the binding of the resulting mutants was tested with bovine and human thrombin, only mutation of Lys47 was found to be sensitive [86,87]. However, the increase in the dissociation constant was only minor: four- to ten-fold. Moreover, the fact that a Lys47 is not conserved in natural hirudin sequences further suggests that this interaction is not necessary [61].

The NH₂- and COOH-terminal regions of hirudin both contain residues important to the hirudin-thrombin interaction. The electrostatic and acidic nature of the carboxy-terminal tail appear to be vital. Removal of the last 7 COOH-terminal amino acids of hirudin - which includes two Glu residues and a sulfated Tyr in natural hirudin - results in a loss of about 90% of inhibitory activity, and removal of the last 22 amino acids nearly abolishes inhibition [70]. Single and multiple mutations of the carboxy-terminal glutamates (57,58,60 and 61) to glutamine caused increases in the dissociation constant of the complex which increased with the number of mutations [87], and each negatively charged Glu appears to contribute equally-favorably to the binding energy [88]. Desulfato Tyr63 hirudin-thrombin complexes possess a higher dissociation constant compared to those from sulfated forms, albeit only three- to ten-fold higher [59,60,71]. The hydrophobic residues in the anionic tail also appear to be involved in the interaction. In inhibition studies with a hirudin COOH-terminal peptide mimetic corresponding to residues 55-64, which is the minimum size required to abolish fibrinogen recognition, it was shown that Phe56, Ile59, Pro60, and Leu64 as well as Glu57 were all sensitive to modification [89]. In the NH₂-terminal region of hirudin, it appears that the hydrophobic nature of the first two amino acids, a positively charged amino terminus, and the integrity of the three disulfides are crucial [68,90].

slid

loc

rele

sub

ring

stud

tega

bind

inter

ratio

ring-

clean

inter

(21.20

confir

micro

FFA

report

requir

(56-45

judges

sugges

confon

(25.24

repted

whiten

action

In its ability to interact with thrombin at both the apolar binding site and the anion-binding exosite, hirudin appears to be a bivalent or 'double-headed' inhibitor. Kinetic analysis of complex formation has revealed that the first step is ionic strength-dependent and unaffected by binding at the active site while the second step is influenced by binding at the active site [59]. These results in light of further studies suggest that the first step involves the interaction of the negatively-charged carboxy-terminal tail of hirudin with the anionic binding exosite, while the second step involves amino-terminal domain interaction in the active site region [81,88]. In support of this notion, it has been found in thrombin binding studies that the COOH-terminal hirudin fragments (49-65,53-65) protect against tryptic cleavage, while the NH₂-terminal fragment (1-52) is a competitive inhibitor of the thrombin substrate D-Phe-pipecolyl-Arg-p-nitroanilide [91,92]. Peptide synthesis of the hirudin segment (45-65) first confirmed that the COOH-terminal region of hirudin alone in complex with thrombin was sufficient to inhibit fibrin formation and the release of FPA without affecting amidase activity [93]. Since then it has been reported that amino acid residues (56-64) and (53-64) are minimally required for these results [89,94]. In addition, the finding that a (56-65) hirudin peptide produces a conformational change in thrombin as judged by circular dichroism which does not occur upon PPACK binding suggests that the COOH-terminal tail of hirudin may trigger a conformational change prior to binding of the NH₂-terminal region [95,96]. It is important to note that these carboxy-terminal hirudin peptide mimetics in complex with thrombin have dissociation constants which are about 6 orders of magnitude higher and anticoagulant activities 3-4 orders of magnitude lower than that of intact hirudin

(33).

aboli

of st

agent

conve

phar

near

thes

glyc

inn

Thu

pro

and

oth

14,

Or

cl

di

pl

ac

to

an

ex

Hi

bl

re

Pr

[93].

The ability of hirudin in complex with thrombin to completely abolish the conversion of fibrinogen to fibrin has lead to a great deal of study and interest in the use of hirudin as an antithrombotic medical agent. Many factors make hirudin an attractive alternative to conventional anticoagulant medical therapies. Currently, pharmacological control of thrombin activity in vivo is attained by means of heparins and bezamidine-type thrombin inhibitors, and both of these treatments have drawbacks [97]. Heparins, which are sulfated glycosaminoglycans, inhibit thrombin indirectly by enhancing the inhibitory power of antithrombin III and heparin cofactor II [26,27,97]. Thus, heparin efficacy is dependent on the plasma levels of these proteins. In addition, heparins can be neutralized by natural antiheparins in the blood such as platelet factor 4, can interact with other blood components such as platelets, fibrinolytic constituents, and lipoprotein lipases, and can lead to hemorrhagic side effects [97]. Oral benzamidine-based anticoagulants interfere with the biosynthesis of clotting factor proteins [97]. Hirudin, in contrast, is a selective, direct thrombin inhibitor which is not dependent on or affected by other plasma proteins, and which does not interfere with the biosynthesis or action of these components [28,98]. In addition, in animal studies modeling five pathogenic thrombosis conditions, hirudin exhibited antithrombotic efficacy at the lowest concentrations in all five examples compared to heparin and the best synthetic inhibitor [97]. Hirudin also displays low antigenicity, is rapidly cleared from the bloodstream, and can be produced in ample amounts to accomodate high medical demand through recombinant DNA methods [68,99]. Two main problems with using hirudin as a therapeutic anticoagulant are the

unpara

fento

nessess

fibula

factor

time

the

disrup

fisso

can e

unparalleled affinity of hirudin for thrombin (K_d in the picomolar to femto molar range), which would be a challenge to counteract if necessary, and the fact that hirudin blocks amidolytic as well as fibrinolytic activities in thrombin [59]. Such concerns are not a factor in the promotion and development of hirudin COOH-terminal peptide mimetics for anticoagulant use. These molecules can effectively inhibit the thrombin-catalyzed fibrinogen to fibrin conversion without disrupting thrombin amidolytic activity, have significantly higher dissociation constants in complex with thrombin compared to hirudin, and can easily be synthesized by non-recombinant means [89,93-95].

A. S.

Pentac

York.

(1990)

The r

Dr.

REVD-

purif

lyopon

rolar

1 m

inact

forma

inact

made

again

Dial

chan

pro

usin

78

II EXPERIMENTAL

A. Sample Preparation

Human α -thrombin was generously supplied to us by Dr. John W. Fenton II of the New York State Department of Health in Albany, New York. The thrombin samples were prepared by the method of Fenton et al. [100], and came as frozen vials which were 3.1 mg/mL in 0.75 M NaCl. The recombinant hirudin sample, rHV2-K47, was kindly supplied to us by Dr. Carolyn Roitsch of Transgene, S.A. in Strasbourg, France. The rHV2-K47 was produced by the method of Loison et al. [65], and was purified by the method of Bischoff et al. [101]; the samples were a lyophilized powder of better than 95% purity.

A frozen solution of thrombin was allowed to thaw on ice. A 10-20% molar excess of rHV2-K47 was dissolved in 50 mM sodium phosphate buffer, 1 mM NaN_3 , pH 7.3, and was added to the thrombin solution. The inability of several 5 μL hirudin-thrombin aliquots to produce clot formation in a 0.5 mL human plasma sample verified that thrombin inactivation had occurred.

The hirudin-thrombin sample was then transferred to a dialysis bag made of Spectra/Por 4, 12-14 kD cutoff dialysis membrane, and dialyzed against 25 mM sodium phosphate buffer, 0.375 M NaCl, 1mM NaN_3 , pH 7.3. Dialysis was carried out at 4 °C for several days with several buffer changes to affect a 10^{-9} to 10^{-11} dilution of the excess hirudin. The protein solution was then concentrated in a refrigerated centrifuge using an Amicon Centricon 10 microconcentrator to approximately 6.0 mg/mL.

B.

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

B. Sample Molecular Weight Estimation

Discontinuous polyacrylamide gel electrophoresis [102, 103] was used to estimate the molecular weight of the hirudin-thrombin complex. Slab gels were prepared using the BIORAD Mini-PROTEAN II dual slab cell unit and the power source for electrophoresis was a BIORAD Model 200 constant voltage power supply. The gels were prepared, stained, and destained using Hoefer Scientific Instruments protocols [104]. The procedure used to estimate the molecular weight of electrophoresed proteins is a Sigma Chemical Company protocol [105], and is based on the methods of Bryan [106] and Davis [103]. Four different slab gels of varying separation gel pore size - 7, 8, 9, and 10 %T - were prepared, where %T is defined as :

$$\%T = \{(\text{mass acrylamide monomer} + \text{mass crosslinker}) / \text{total volume}\} \times 100 \quad \dots(1)$$

The pore size and the %T of a gel are inversely correlated; a large value of %T indicates a small pore size. Three microgram samples of the hirudin-thrombin complex, pre- and post-dialysis, as well as of the following calibration standard proteins were applied to each gel : bovine milk α -lactalbumin (MW 14,200), bovine erythrocyte carbonic anhydrase (MW 29,000), chicken egg albumin (MW 45,000), and bovine serum albumin (MW-monomer 66,000; MW-dimer 132,000). Bromophenol blue tracking dye was added to all of the samples. The gels were run at a constant voltage of 200 V. It took approximately one hour for the dye front to move the length of the gel, which concludes the electrophoretic separation. The R_f, or electrophoretic mobility, of each protein oligomer was calculated for each of the four gels, and 100Log(R_f×100) versus percent gel concentration (Figure 11) was plotted for each

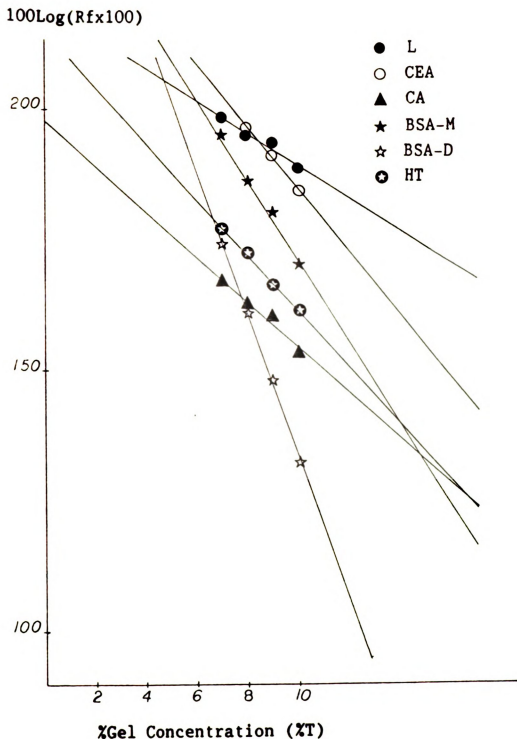


Figure 11. $100\text{Log}(Rfx100)$ vs. %T Gel Concentration. The slope for each line gives the retardation coefficient (K_r) for the corresponding protein oligomer. Abbreviations: L(lysozyme), CA(carbonic anhydrase), CEA(chicken egg albumin), BSA-M(bovine serum albumin-monomer), BSA-D(bovine serum albumin), HT(hirudin-thrombin)

prote

The s

Retar

plac

bro

esper

from

19.3

CP

pro

entr

weig

C.

cta

the

to

the

the

the

the

the

the

the

the

the

protein, where R_f is defined as :

$$R_f = \frac{\text{distance of protein migration}}{\text{distance of tracking dye migration}} \quad \dots(2)$$

The slope for each protein oligomer in such a plot is known as the Retardation Coefficient (K_r). The negative logarithm of K_r was then plotted versus the logarithm of the molecular weight for the standard proteins, producing a linear plot (Figure 12). Using the experimentally-determined K_r value for the complex (-5.4), extrapolation from this plot gave a hirudin-thrombin complex molecular weight of 38,500. Although this molecular weight estimate for the complex is only 1900 daltons larger than that of native human α -thrombin and is about 5000 daltons less than that expected, the discrepancy could be attributed to the difficulty to accurately estimate protein molecular weights from gel electrophoresis.

C. Crystallization

Crystals of the rHV2-K47-human α -thrombin complex were initially obtained by the hanging drop, vapor diffusion method [107] in an incomplete crystallization parameter factorial search [108] using 30 (w/v)% polyethylene glycol (PEG) 4000, 0.2 M MgCl_2 , 0.1 M sodium acetate buffer, pH 4.5, 1 mM NaN_3 , and a protein concentration of about 6 mg/mL. Crystals appeared in 5-7 days. Refinement of these conditions found that the largest crystals resulted from precipitant solutions which contained 27-28% PEG 4000, 0.2 M MgCl_2 , 1 mM NaN_3 , and any of the following buffers at a concentration of 0.1 M: sodium acetate buffer, pH 4.7 or 5.0; sodium citrate buffer, pH 5.5; and ADA (N-[2-Acetamido]-Iminodiacetic Acid) buffer, pH 6.0. Larger crystals (maximum dimensions $1.0 \times 0.4 \times 0.4 \text{ mm}^3$) suitable for x-ray diffraction

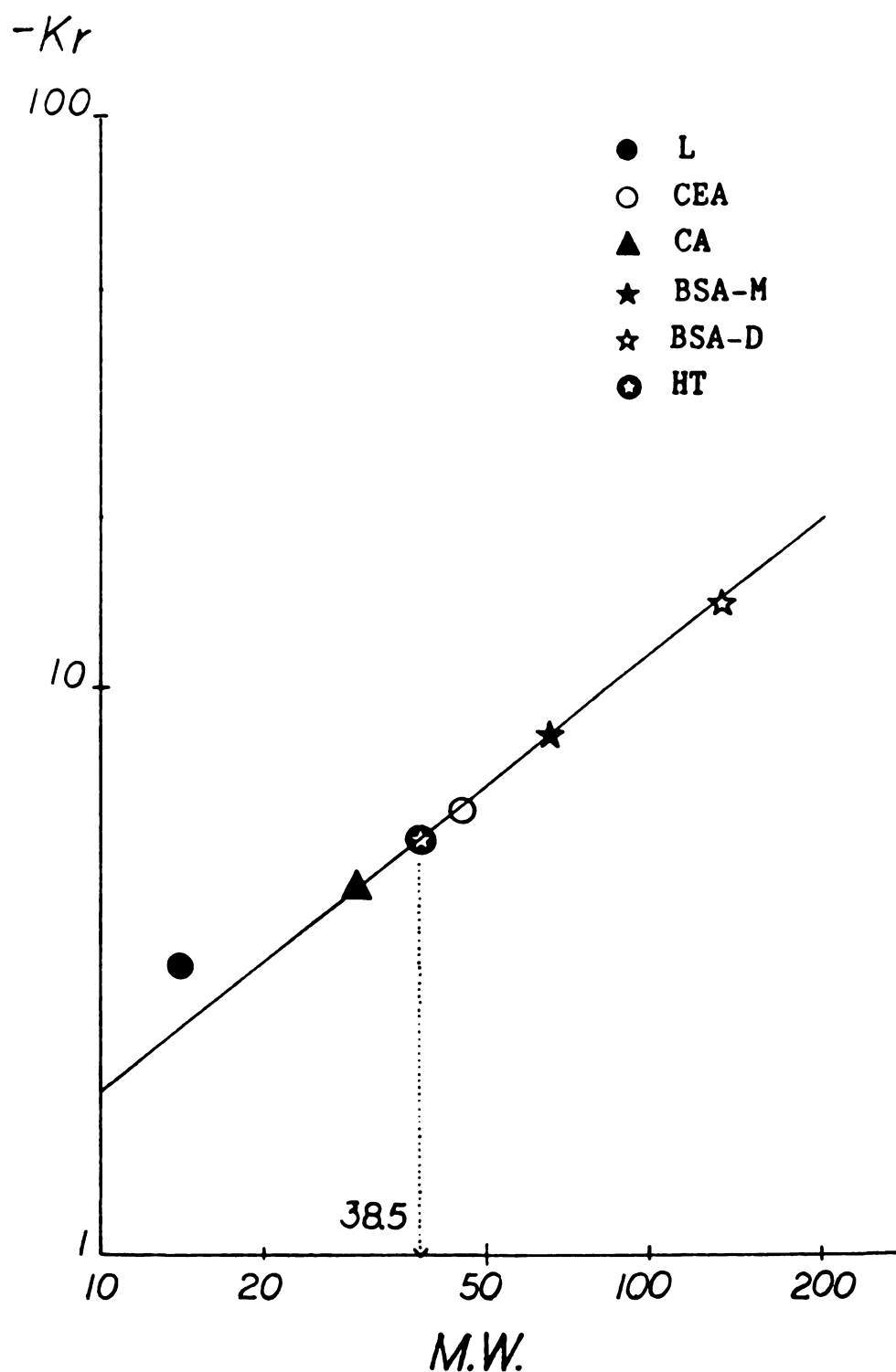


Figure 12. $\text{Log}(-Kr)$ vs. $\text{Log}(\text{Molecular Weight})$. Linear regression best fit line to the standard protein retardation coefficient data. Molecular weight (M.W.) in kilodaltons. Abbreviations used have the same meaning as in Figure 11. The extrapolated molecular weight estimate for the hirudin-thrombin complex is 38.5 kD.

studi

[199]

ident

relece

drop

to

12.5

circu

cryst

D.

W-ra

FD-2

Open

pre

and

for

de

to

to

to

to

to

to

to

to

to

to

to

studies were obtained by macroseeding protein-precipitant solutions [109]. Precipitant solutions used for macroseeding experiments were identical to those mentioned above, but were 26% in PEG 4000. Crystals selected to be used as seeds for macro-seeding in protein-precipitant droplets were rinsed in a solution prior to transfer which was identical to the precipitant solution of the crystallization, but which was 23.5-24.0% in PEG 4000. A photograph of a data collection-quality hirudin-thrombin crystal is shown in Figure 13; the trace of the seed crystal is visible in the center of the large crystal.

D. Crystal Characterization

Crystals of the complex were characterized using Ni-filtered Cu K_{α} x-rays. Still and precession photographs were taken using a Rigaku RU-200B Standard Rotating Anode Source with a 0.5 x 3.0 mm filament operating at 8.8 kW power (55 kV, 160 mA), and a Charles Supper precession camera. We thank Drs. Cele Abad-Zapatero, John Erickson, and Chang Park of the Protein Crystallography Lab of Abbott Laboratories for the use of their facility to carry out these experiments. Frame data were also collected using a Xentronics Area Detector on a Nicolet P3/F diffractometer operating at 1.75 kW (50 kV, 35 mA) with a fine focus x-ray tube. The area detector data were processed using XENGEN data reduction software. We would like to thank Mr. Steven Muchmore of the Protein Crystallography Lab at The Upjohn Company for carrying out the area detector work for us.

The unit cell parameters and relevant crystal data are summarized in Table 2. Still x-ray pictures displayed diffraction to better than 2.5 Å resolution. From (0kl) and (hk0) precession photographs (Figure 14) and/or the reduced area detector data, the rHV2-K47-human α -thrombin

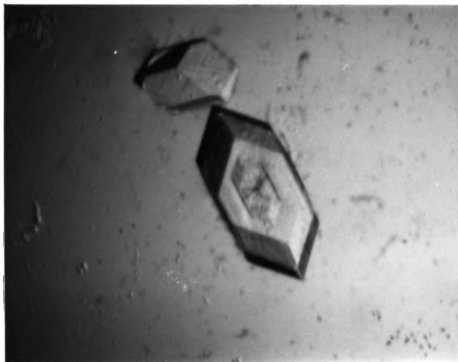


Figure 13. Photograph of an rHV2-K47-Human α -Thrombin Crystal. Crystal size is approximately 1.0 x 0.4 x 0.4 mm. Note the trace of the seed in the center of the crystal.

Table 2. Summary of Crystal Data of rHV2-K47-Human α -Thrombin.

Crystal system	Tetragonal
Space group	$P4_12_12$ or $P4_32_12$
Number of molecules per asymmetric unit	1
a (Å)	90.39(2)
c (Å)	131.97(8)
Solvent fraction	61%
Matthews number (V_m ; Å ³ /dalton)	3.10

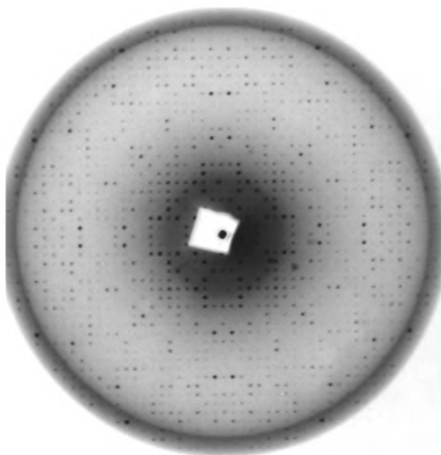


Figure 14. $hk0$ Precession Photograph of a Hirudin-Thrombin Crystal. The resolution limit at the edge of the photo is 3.7 Å. (μ angle = 12.0°.)

crystals were shown to be tetragonal, space group $P4_12_12$ or $P4_32_12$. The unit cell dimensions measured with a Nicolet P3/F diffractometer using 10 high order reflections are: $a = b = 90.39(2)$, $c = 131.97(8)$ Å. Assuming eight hirudin-thrombin complexes ($M_r = 43,500$ D) per unit cell, and a protein specific volume of $0.74 \text{ cm}^3/\text{g}$ gives a solvent fraction (fs) of 61% in the crystals; protein crystals generally have a solvent fraction ranging from about 27 to 65%, with 43% being the most common value [110]. Given the cell parameters and assuming one complex molecule per asymmetric unit, the Matthews number or V_m [110], which is the crystal volume per unit of protein molecular weight, is calculated to be $3.10 \text{ Å}^3/\text{dalton}$; the most commonly observed V_m value is $2.15 \text{ Å}^3/\text{dalton}$ [110]. The high solvent fraction and high Matthews number of the hirudin-thrombin crystals are, however, nearly identical to the values obtained from crystals of the asymmetric molecules human Ig immunoglobulin (fs = 0.61, $V_m = 3.14 \text{ Å}^3/\text{dalton}$) [111] and yeast phenylalanine tRNA (fs = 0.60, $V_m = 3.20 \text{ Å}^3/\text{dalton}$) [112], which suggests that the hirudin-thrombin complex could be asymmetrical in shape.

A storing solution of 30% PEG 4000, 0.375 M NaCl, 0.2 M MgCl_2 , 0.1 M sodium acetate buffer, pH 4.6, 1 mM NaN_3 was found to stabilize the crystals and maintain diffraction quality. The rHV2-K47-thrombin crystals used for x-ray diffraction experiments were gradually equilibrated in this storage solution prior to use by slowly exchanging the crystallizing solution with the storage solution.

E. 2.8 Å Resolution Diffractometer Intensity Data Collection

A 2.8 Å resolution set of diffractometer intensity data was collected on the hirudin-thrombin crystals using monochromatic Cu K α radiation (1.5418 Å) and a Nicolet P3/F four-circle diffractometer controlled by a Data General Nova/4 computer. X-rays were generated using an AEG sealed x-ray tube operating at 2 kW (50 kV, 40 mA), and were monochromated by means of a graphite monochromator crystal. A helium evacuated collimator and beam tunnel were used to minimize the air absorption of the incoming beam and the diffracted x-rays, respectively. The complete data set was collected in two parts, each part using a different crystal. The first data collection contained reflection measurements from 3.63 to 2.80 Å resolution (reflections had 2 θ values ranging from 24.5° to 32.0°), and will be referred to as HTN2. The second set contained intensity data to 3.56 Å resolution (reflection measurements from 2.0° to 25.0° in 2 θ), and will be referred to as HTN3. (The first actual 3.56 Å resolution intensity data set collected, HTN1, was used to generate a file of observed reflections.)

The crystal used for the HTN2 and HTN3 data collections had dimensions of approximately 0.95 x 0.40 x 0.35 mm. These crystals were mounted in siliconized glass capillaries of diameter 1.0 mm. The morphology of an idealized hirudin-thrombin crystal is shown in Figure 15. A crystal of the complex was mounted such that the c* axis (along the longest dimension of the crystal) was coincident with the capillary axis. A drawing of a hirudin-thrombin crystal mounted in a capillary is shown in Figure 16. The crystal was positioned in the capillary bathed in the storing solution and, just prior to sealing the capillary, the crystal was partially 'dried' with a filter paper strip so that only a small amount of liquid remained under the crystal to affix it to the

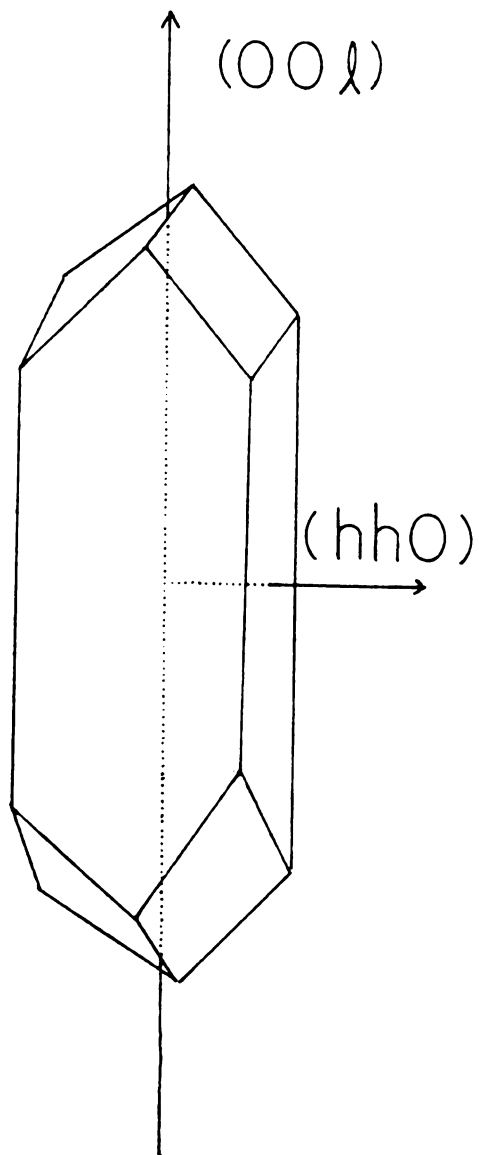


Figure 15. Morphology of an Idealized Hirudin-Thrombin Crystal.

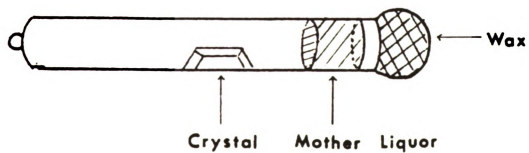


Figure 16. A Hirudin-Thrombin Crystal Mounted in a Capillary for X-ray Crystallographic Analysis.

capillary wall. A small band of storage solution was left at one end of the capillary before sealing it off with wax; as protein crystals contain a significant solvent fraction, the presence of this band of storage solution in the sealed capillary ensured, by vapor equilibration, that the crystal would not dry out. The capillary containing the crystal was then inserted into plasticene mount on a goniometer head.

The protein crystal on a goniometer head was positioned onto the goniostat of a four-circle diffractometer. A schematic drawing of this goniostat is shown in Figure 17 [113]. The goniostat orients the crystal and detector with respect to a stationary x-ray beam, and brings Bragg planes into reflecting position. W.L. Bragg developed the analogy of the the diffraction of x-rays by crystals to ordinary reflection, and he deduced a simple equation - $n\lambda = 2d\sin\theta$ - by treating diffraction as "reflection" from planes of atoms in a lattice [114]. In this equation, d represents the interplanar spacing of the Miller planes and λ is the wavelength of the x-ray radiation. Thus, diffracted x-rays are often referred to as reflections. A reflection is identified by the Miller indices h, k , and l , where the reflection plane (hkl) makes intercepts a/h , b/k , and c/l with the unit-cell axes a , b , and c of the crystal. The four circles of the goniostat refer to the three circles swept out by rotation of the crystal about the ω , ϕ , and χ axes (Figure 17), and the circle swept out by the arm of the detector through rotation about the θ axis (which is coincident with the ω axis). A goniometer head with the mounted protein crystal is placed on a base in the circle which is coincident with the axis. Each diffracted x-ray beam collected with a diffractometer is therefore defined by a 2θ , ω , ϕ , and χ angle as well as by an (hkl) index.

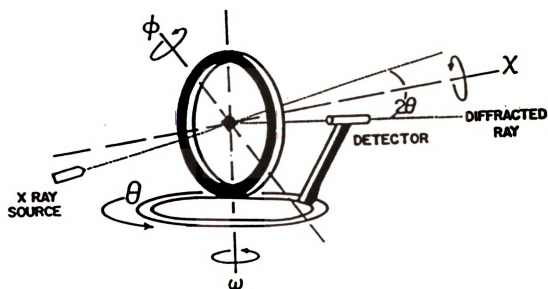


Figure 17. The Goniostat of a Four-Circle Diffractometer [111].

The 2θ , ω , ϕ , and χ angles which the computer-controlled diffractometer drives to in the course of making an intensity measurement result from the unit cell parameters and the orientation of the crystal. The unit cell constants and crystallographic orientation of the crystal are stored in the computer in the form of an orientation matrix A, which relates the reciprocal lattice vectors a^* , b^* , and c^* in terms of their components in the coordinate system of the instrument.

$$A = \begin{pmatrix} a^*x & b^*x & c^*x \\ a^*y & b^*y & c^*y \\ a^*z & b^*z & c^*z \end{pmatrix} \quad \dots(3)$$

In the HTN2 and HTN3 data collections, the orientation matrix was initially determined by manually locating three strong and approximately mutually orthogonal reflections, (0,0,32), (1,16,0), and (16,1,0). These reflections are stored in the reflection array, and were used to calculate the unit cell constants and the initial orientation matrix. As the hirudin-thrombin crystals were mounted with the c^* or (001) direction of the crystal approximately coincident with the capillary axis, it was possible to align the c^* axis to coincide with the ϕ axis of the diffractometer. This alignment facilitated the location of (1,16,0) and (16,1,0), which could be found at $\chi = 0$, since the crystal lattice is an orthogonal system. Axial intensity distributions of the (001) as well as the (h00) and (hh0) directions were then recorded. These measurements, which are displayed in Figure 18, allow one to assess the quality and integrity of the diffraction pattern from crystal to crystal. After this initial orientation matrix was obtained, a more accurate matrix was found by including nine additional high order reflections into the reflection array and by performing a least squares calculation between the calculated and observed angles of the

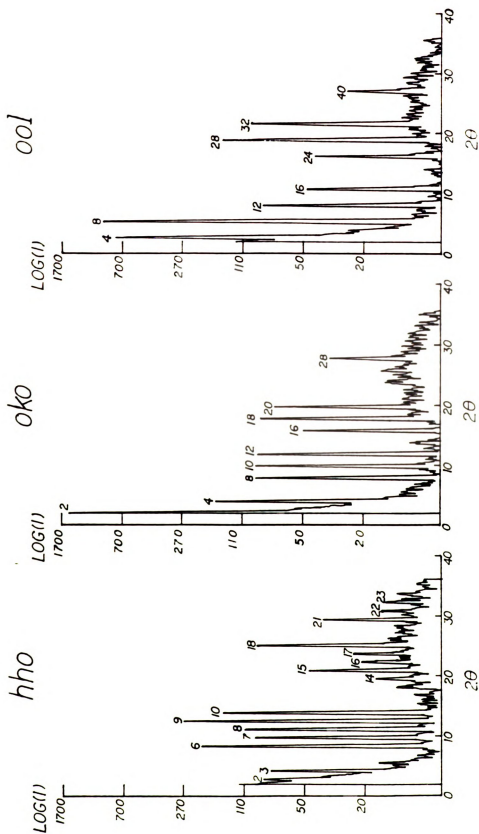


Figure 18. Axial Intensity Distributions of the Recombinant Hirudin-Human α -Thrombin Crystals.

reflections in the array. These additional reflections were all of relatively high 2θ ($15^\circ < 2\theta < 22^\circ$), possessed sufficiently strong intensities so that the reflection centering routine would be able to center them after typical decay (5-15%), and collectively were distributed uniformly in the region of reciprocal space used for intensity measurements - most notably in χ . The cell dimensions of the crystal were updated each time a least squares calculation was performed on the reflections in the reflection array. The average cell parameters obtained by least squares in the HTN2 and in the HTN3 data collections are shown in Table 3.

The method used to measure the intensity data was first proposed by Wyckoff et al. [115], and is known as the Wyckoff ω -step procedure. The Wyckoff procedure is a stationary crystal-stationary detector peak sampling method where the top of the reflection peak is sampled by making a limited number of intensity measurements at small intervals in ω , as opposed to the full profile scan from background to background. It is often used in cases where reflections are closely spaced or where quick, accurate measurements are sought to minimize x-ray-induced crystal decay - situations frequently encountered in macromolecular data collections. Using this method in the HTN2 AND HTN3 data collections, the peak was sampled 0.09° on either side of the calculated value for a given reflection by taking intensity measurements at seven 0.03° step increments, and it took 18 seconds to complete the scan. The four largest intensities measured were then summed, and this sum was taken to be proportional to the integrated intensity. Background measurements were made on either side of the peak by offsetting the ω value 0.40° , and it took approximately 7 seconds to complete the left and right background measurements. Thus, it took approximately 25 seconds to

Table 3. The Average Cell Parameters of the HTN2 (3.6–2.8 Å) and of the HTN3 (2.8 Å) Crystals. Cell lengths in Å units, angles in degrees.

	A	B	C	α	β	γ
HTN2	91.07	90.99	132.60	90.05	89.99	90.03
σ	0.02	0.01	0.04	0.01	0.01	0.01
HTN3	90.95	90.95	132.56	89.99	90.04	90.03
σ	0.02	0.02	0.03	0.04	0.03	0.02

measure each reflection intensity. This procedure also can compensate for minor crystal misalignments by allowing the diffractometer to take additional steps if the peak maximum is too close to an edge of the scan range; six additional steps were allowed in the HTN2 and HTN3 data collections.

Although the Wyckoff ω step scan was used to collect both sets of hirudin-thrombin intensity data, the manner of data collection was different in the two sets. In the HTN2 (3.63-2.80)Å data collection, the intensity data were collected in three 2θ shells of (24.5-27.5)°, (27.5-30.0)°, and (30.0-32.0)°, in the order of lowest to highest resolution. As only 1/16th of the sphere of reflection needs to be measured to collect a unique set of intensity data, only reflections with the k index greater than or equal to the h index were measured. In each shell, data were collected by ascending h,k,l - with l varying fastest, h slowest. A total of 7914 reflections were measured in the HTN2 data collection, and it took approximately 55 hours to obtain these data. In the HTN3 3.56 Å data collection, a reflection file was used based on the observed reflections of the HTN1 3.56 Å data collection; 81% of the possible HTN1 reflections were judged to be observed and were included in this file. The file was created by dividing the unique, observed data into 19 2θ ranges of approximately 305 reflections per range. Each range was sorted in order of descending 2θ , with h varying slowest, l varying fastest, and these individual files were merged in the order of descending average 2θ , so that the data collection file would allow reflections to be measured from high to low resolution. Since high order data are generally weaker, and are more strongly affected by x-ray-induced decay than low order data, this manner of data collection is preferred to that used in the HTN2 data collection. In

the HTN3 data collection, 5983 reflections were measured over the course of about 42 hours.

During the course of each data collection, the alignment of the crystal was monitored by measuring the intensity of three nearly orthogonal reflections every 100 reflections; reflections (1,16,0), (16,1,0), and (0,0,32) were chosen for this purpose. If the intensity of these monitor reflections diminished by more than 15%, the crystal was considered misaligned, and this triggered the re-measurement of the 12 reflections in the reflection array, followed by the calculation of a new orientation matrix and cell parameters by a least squares calculation. Slopes resulting from a least squares fit line of intensity versus exposure hour data for each monitor reflection were used in data processing to obtain decay corrections to the data of $2\theta > 15.0^\circ$.

A number of different measurements preceded and followed the actual intensity data collection; these measurements included the ω profile of the reflection (6,16,17), the intensity absorption of (0,0,8) reflection versus ϕ angle, and the (hk0) zonal reflections to 5.9 Å resolution. The ω profile of a reflection helps determine the parameters in the Wyckoff ω step scan, and provides an indication of the suitability of the crystal for intensity data collection. A split profile, for example, may imply that the crystal is cracked, and that two closely oriented but distinct crystals are contributing to the profile. In the HTN2 and HTN3 data collections, the ω profile of the reflection (6,16,17) was determined by making 10 second intensity measurements at ω values of $\pm 0.50^\circ$ from the calculated value in 0.05° increments of ω ; the ω profile obtained for reflection (6,16,17) prior to beginning the HTN2 data collection is shown in Figure 19. The width of the profile

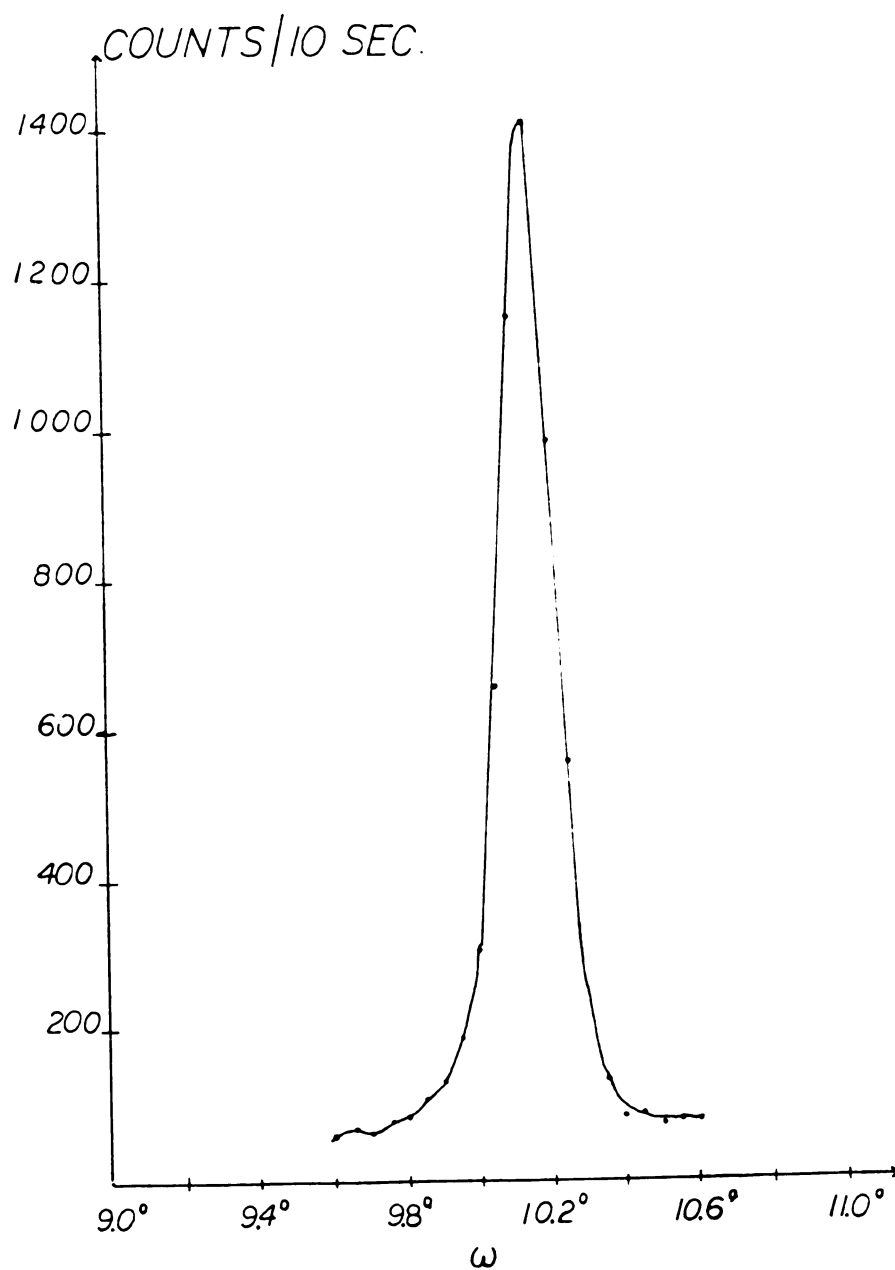


Figure 19. The ω -Profile of (6,16,17) Taken Before the 3-D HTN2 Data Collection. Peak width at half-height approximately 0.18°.

peak at half the intensity maximum defines the region of the peak which should be scanned in the Wyckoff ω step scan; since this width is approximately 0.18° in the profile (Figure 19), a scan range of 0.18° was chosen. From this profile one can also determine a suitable ω offset value at which left and right background measurements can be made; an offset value of 0.40° was chosen from the Figure 19 profile.

The intensity absorption versus ϕ angle for the reflection (0,0,8) was measured to determine the optimum region of ϕ to carry out the data collection. Ideally, one would like to collect data in the region with the lowest overall x-ray absorption or the highest overall intensity. The (0,0,8) reflection was chosen because it has a large intensity (Figure 18) and, more importantly, because each crystal was aligned on the diffractometer so that the c^* axis was coincident with the ϕ axis. This alignment ensured that the (001) Bragg planes were always in reflecting position for all values of ϕ at $\chi = 90^\circ$. Intensity measurements were made every 20° in ϕ over the entire 360° range, and then every 10° over a 150° range containing the 45° data collection ϕ region. A plot of $I(\phi)$ versus ϕ taken before the HTN3 3.56 Å data collection is shown in Figure 20; the 45° range indicated in the plot was the ϕ range employed for the data collection. This region was chosen for the data collection because it was of high intensity or low x-ray absorption. The curve obtained from averaging the absorption curves determined before and after each data collection was used to correct the intensity data for absorption in the data processing.

Intensity data for (hk0) zonal reflections of 2θ (2-15) $^\circ$ were measured before and after the three-dimensional data collections. These data were used in the three-dimensional data processing to assess decay due to x-ray exposure for reflections of the corresponding 2θ range.

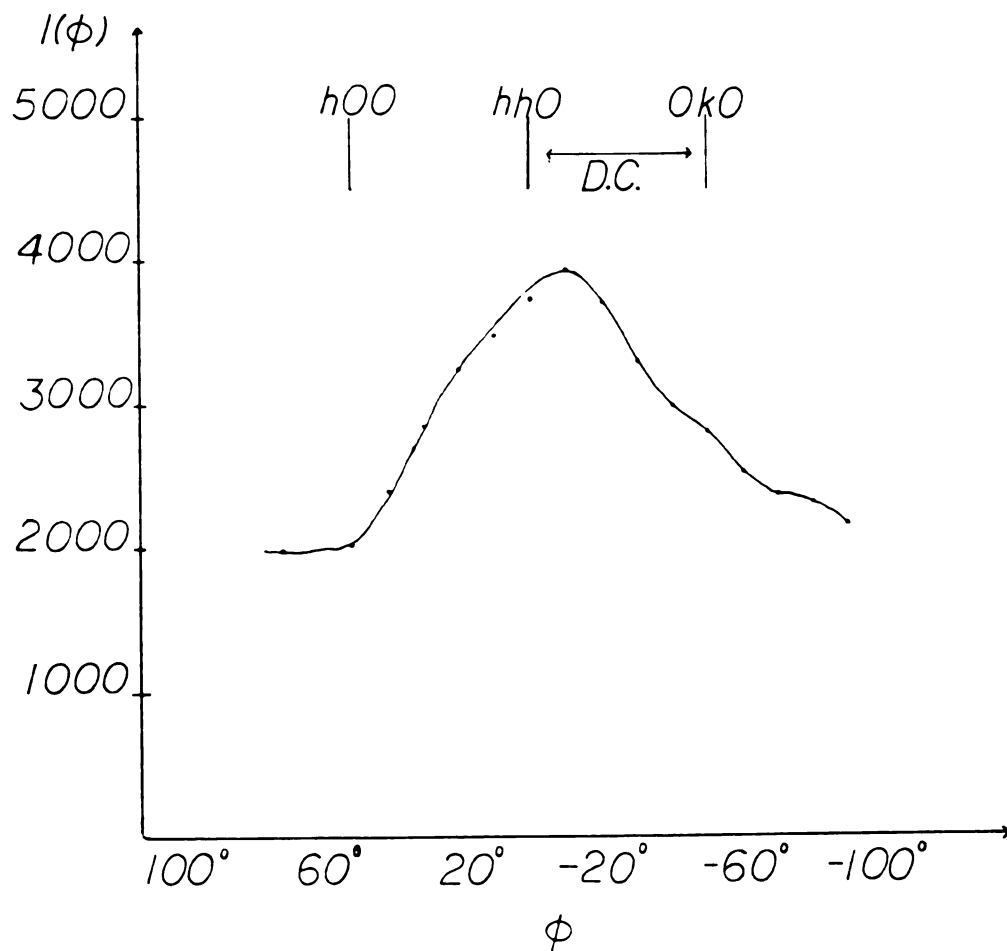


Figure 20. Intensity Absorption Plot of (0,0,8) Taken Before the HTN3 3.56 Å Resolution Data Collection. $I(\phi)$ in total counts/measurement. The 45° ϕ range employed in the data collection indicated by "D.C.".

Word

001

002

F.

00

00

Ma

00

00

00

00

00

00

00

00

00

00

00

00

00

00

00

00

00

00

00

00

00

Moreover, the 5.9 Å resolution zonal intensity data measured before data collection for the HTN2 and HTN3 data sets were used to obtain an initial scale factor for scaling the two data sets.

F. 2.3 Å Resolution Area Detector Data Collection

A 2.3 Å resolution set of area detector intensity data was collected from a recombinant hirudin-thrombin crystal in the laboratory of Professor Robert Huber at the Max-Planck-Institut für Biochemie in Martinsried, Germany. The Enraf-Nonius data collection system consisted of a CAD-4 four-axis goniostat and a FAST television area detector. Cu K_{α} x-rays were produced by a Rigaku RU-200B rotating anode x-ray generator fitted with a 0.3 x 3.0 mm fine focus filament and operating at 5.2 kW (45 kV, 120 mA). The x-rays were monochromated by means of a Ni-filter which removed 99.95% of the Cu K_{α} radiation. A hirudin-thrombin crystal of approximate dimensions 0.85 x 0.50 x 0.50 mm, mounted in the previously described manner with the c^* axis coincident with the capillary axis, was used to collect the intensity data; all measurements were made at 6 °C.

The Enraf-Nonius CAD-4 goniostat is a 4-axis diffractometer which makes use of the same 2θ , ω , and ϕ angles as the Siemens P3/F system, but which uses a κ mechanism, as opposed to the χ circle, to position the goniometer head. A drawing of a ' κ -geometry' goniostat is shown in Figure 21 [116]. The goniometer head is held on an arm which can rotate about an axis 50° from the vertical. The base of the arm rotates about the θ axis, and the motion is identical to that of ω on a conventional goniostat. The upper portion of the arm is held with the ϕ axis inclined 50° to the κ swivel axis. Combined motions of the ω and κ can produce crystal orientations analogous to $-100^\circ < \chi < 100^\circ$ on a conventional

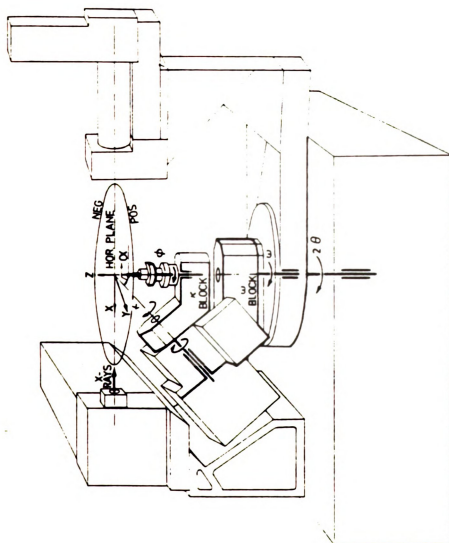


Figure 21. Drawing of an Enraf-Nonius Diffractometer with a Conventional Detector [116].

diffractometer [116].

The Munich Area Detector (NE) System or MADNES software package is an inclusive area detector x-ray diffraction data collection system of programs which was used to carry out all steps involved the FAST experiment. MADNES is an area detector independent system which allows one to perform the following diverse tasks [117]: approximately align crystals; find reflections; autoindex found reflections and use them to determine unit cell parameters and an orientation matrix; refine crystal and detector orientation; predict reflection positions; collect reflections; evaluate reflections as they are collected; monitor and update the crystal orientation matrix and detector alignment throughout data collection; and plot predicted and/or observed reflections.

The data collection geometry used in a MADNES-controlled area detector x-ray diffraction system is similar to that used in rotation or oscillation x-ray photography: the crystal is rotated about an axis perpendicular to the incident x-ray beam. In order to measure a nearly complete set of data from a hirudin-thrombin crystal, two crystal orientations and two data collection ϕ ranges were used. (The ϕ angle in MADNES terminology is identical to the ω angle of the CAD-4 goniostat.) Initially, data were collected by rotating nearly 90° in ϕ (-60° to 24°) about the c^* or capillary axis of the mounted crystal. In addition, data were collected by rotating 100° in ϕ (-10° to 90°) about an axis 30° oblique to the c^* axis.

Unlike conventional scintillation counter diffractometers which measure reflection intensities individually, area detectors have the capability of measuring several hundred reflection intensities at a time. The FAST area detector is a phosphor screen coupled to a television camera, and it is composed of picture elements called pixels.

The pixels are arranged on the screen in a 512 x 512 grid. FAST area detector x-ray intensity data are collected by rotating a crystal through a small angle while x-ray quanta are collected on the phosphor screen. The x-ray quanta collected at each pixel location are integrated, and a two-dimensional array is generated whose elements correspond to the amount of x-rays detected at each detector location; this array is called an image or frame, and it is the basic informational entity used by MADNES to perform its numerous aforementioned tasks. In the hirudin-thrombin FAST experiment, frames of data were generated by rotating 0.1° in ϕ at a rate of 90 seconds per frame. The crystal-to-detector distance was 49 mm, and it took approximately 56 hours to complete the frame data measurements at the two previously mentioned orientations.

The flow which transpires during a MADNES intensity data collection is indicated in Figure 22 [118]. The initial step in the data collection process is a pre-alignment, and it is very similar to the preliminary alignment which must be conducted prior to x-ray precession photography. The purpose of the pre-alignment is to find two major zones of the crystal, and to reposition the crystal so that the center of the zones approximately coincides with the center of the x-ray beam [117,118]. In this experiment, the (0kl) and (h0l) zones were aligned. A least squares fit to the spots which defined the diffraction cones gave the circle center, and the new goniostat angles were calculated to bring the zone and beam centers into alignment. After this step, the missetting angles were approximately zero and rough unit cell parameters could be obtained.

The second step in the FAST experiment utilizes the FIND command and involves determining the centroids of a few hundred reflections at

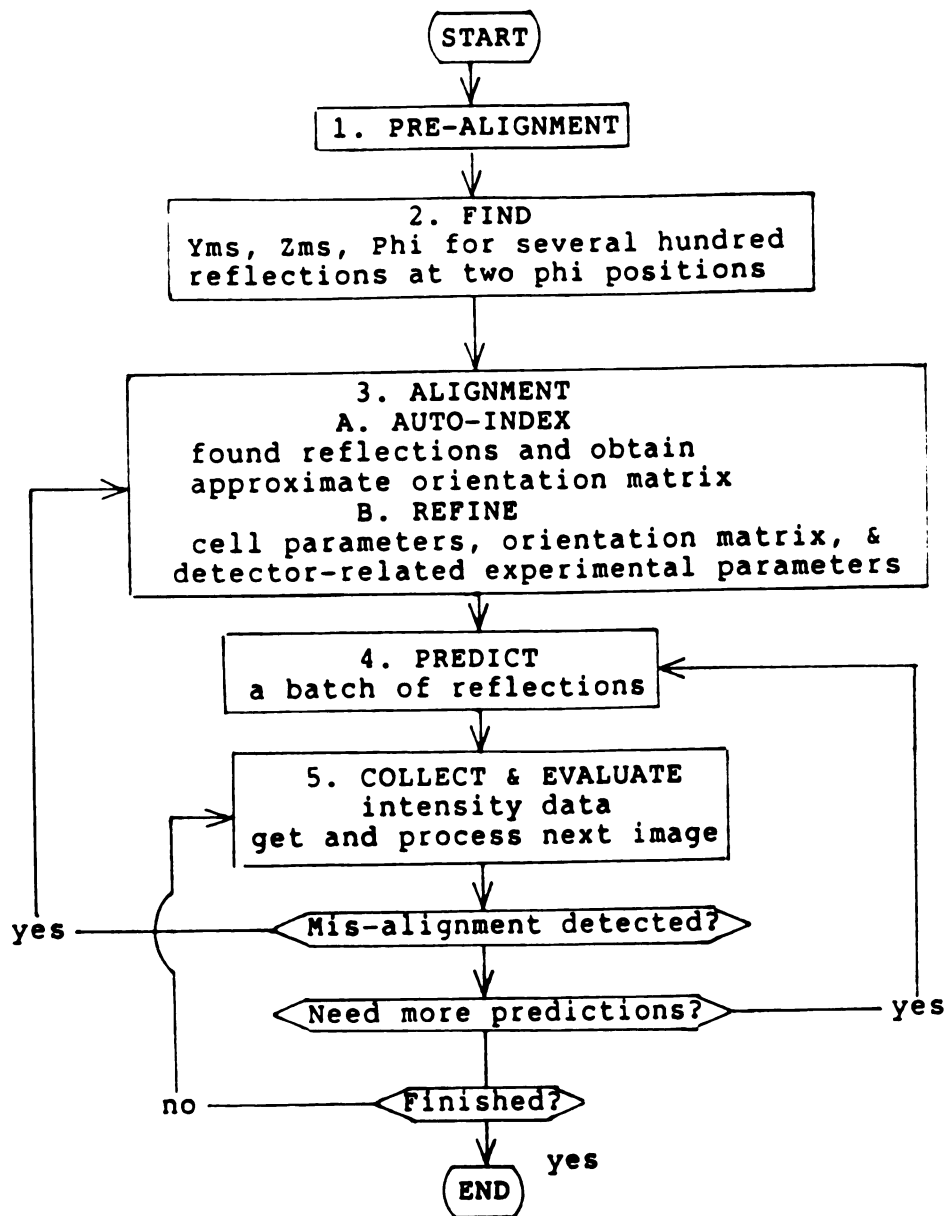


Figure 22. A Flow Chart of Operations in a MADNES Area Detector Data Collection.

the pre-aligned ϕ positions. Three coordinates define a FAST reflection position: Y_{ms} , Z_{ms} , and ϕ . The coordinates Y_{ms} and Z_{ms} ('ms' stands for 'mass store') correspond to the vertical and horizontal directions of the detector, and the ϕ coordinate refers to the crystal rotation angle (the ϕ angle is identical to the ω goniostat angle, as was mentioned earlier). Finding a hundred or so reflections at each of the (h0l) and (0kl) zonal positions was a 4-stage process [118]. First, an initial image was produced by rotating 0.4° in ϕ about each zonal ϕ position. Second, each resulting image was searched for reflections and the Y_{ms} , Z_{ms} values were extracted. Third, ten images spanning 1.0° in ϕ (0.1° rotation per frame), and centered on the zonal ϕ value of step 1, were measured. Finally, a list of Y_{ms} , Z_{ms} , and ϕ coordinates for the reflections contained in these frames was produced. The purpose of the third step was to ensure that any partial reflections measured in step 1 would be completely defined.

The third step of the area detector data collection is an alignment stage which includes an auto-indexing of the found reflections, an orientation matrix calculation, and a refinement of various crystal and experimental parameters. In the rHV2-human α -thrombin data collection, the AUTJ (auto-indexing of reflections by the James Pflugrath method) command was used to auto-index the found reflections and obtain an approximate orientation matrix. The REFIN command was then implemented to refine the crystal orientation matrix, unit cell parameters, the missetting angles, the crystal-to-detector distance, the detector tilt and twist angles, and the center of the primary beam [117]. The unit cell parameters obtained as a result of this alignment step for hirudin-thrombin were: $a = b = 90.53 \text{ \AA}$; $c = 132.05 \text{ \AA}$; $\alpha = \beta = \gamma = 90.00^\circ$ and the error estimates were approximately $0.05 \text{ \AA}$ and 0.05° ,

respectively.

Having completed the pre-alignment, reflection-finding, and alignment steps, one is now ready to actually collect and evaluate the needed frame data. Obtaining a three-dimensional set of reflection data using MADNES basically involves the iterative use of two commands: PREDICT and COLLECT. PREDICT uses the output from the previous alignment step to calculate for a reflection the (hkl) index, central ϕ value, ϕ range, starting ϕ value, Lorentz-polarization correction factor, oblique incidence correction factor, and the mass store coordinates Yms, Zms [117]. Reflections are predicted in small ϕ increments called batches. The major restriction on the batch size is that it contain less than 3000 reflections; in the hirudin-thrombin data collection a batch size of 2.5° was used. Once a batch of reflections has been predicted it is sorted on starting ϕ value and checked for overlaps.

By means of the COLLECT command, frame data collection and data processing are carried out, and controlled by the appropriate parameter input. The data collection proceeds by the exposure of a series of images as the crystal rotates, and involves the extraction of information around each predicted reflection occurring on the images. MADNES stores all the reflection information in a double-linked list [117]. Each time a reflection begins to approach the Ewald sphere, the storage size of data for the reflection is calculated and a contiguous amount of space in this list is allocated. Information for this reflection from a series of adjacent images is transferred to this allocated location as it is obtained. Once the reflection has finished passing through the Ewald sphere, the information for that reflection in the double-linked list is complete and that reflection is processed; the data for the reflection can be written to disk if desired, but most

importantly the data space for this reflection is cleared for use by a new reflection. As an image is being generated by the area detector, MADNES can process the previous image.

Collected (hkl) reflection intensity data were processed to observed structure amplitudes (uncorrected for scaling, absorption, or decay), denoted $|F_{obs}|$, during the course of data collection by application of the following relationship:

$$|F_{obs}|^2 = LORPOL \times I(hkl) \quad \dots(4)$$

where $I(hkl)$ is the background-corrected intensity of a reflection (hkl) and LORPOL is the Lorentz-polarization correction factor, which is a geometric-polarization correction term. The size and shape of reflections was determined by means of the dynamical masking method of Sjolín and Wlodawer [119]. This method generates a mask which is used to evaluate the pixel data. Mask generation proceeds via five stages [118]. In stage 1, the 3-D data are treated with a smoothing algorithm, and the average and standard deviation of the smoothed array are determined. In stage 2, a mask is produced from the smoothed data using the formula:

$$MASK_{y,z,t} = (SMOOTH_{y,z,t} - \text{Average}) / \text{Std. Dev.} \quad \dots(5)$$

In stage 3, a new average background and standard deviation are calculated from the smoothed array by leaving out pixels which deviate too much from the previous average and step 2 is repeated. In step 4, a contiguity criterion is applied to sufficiently positive mask values; the contiguous region is considered to be the peak and non-contiguous regions due to noise or interfering reflections. In step 5, the mask is

used to assign the original data into peak, background, and neither peak nor background pixels so that peak integration and intensity calculation can be achieved. Masks are not produced for low-intensity reflections; these reflections are processed using a mask made from the masks of medium and strong reflections from the same region of the detector. (The detector surface is divided into 16 regions.)

As a result of the COLLECT command, an observed structure factor amplitude ($|F_{obs}|$) file is generated and continually added to during the course of active data collection. The file contains the following information for an individual reflection record: (hkl), $|F_{obs}|$, $\sigma(F_{obs})$, the average background, $\sigma(\text{background})$, error codes, crystal identifier codes, calculated and observed: Y_{ms} , Z_{ms} , ϕ coordinates and the width of the spot center (observed values are only listed if found), and exposure time in hours at the spot center [117]. Each reflection is given a simplified character as well as a detailed numeric error code. Reflections can be given one of four simplified error codes: good, weak, edge, and bad. 'Good' reflections are measurements which were obtained under no obvious error conditions and will be used in further processing (to be described in a later chapter) to generate $|F(hkl)|$ values. Reflections which are labelled with the other three simplified error codes are ignored in later processing, and have detailed numerical error codes which reveal the exact source of the error. 'Weak' reflections were not observed in the data collection. 'Edge' reflections were measured on the edge of a data array in one or more of the Y_{ms} , Z_{ms} , or ϕ coordinates. A 'bad' reflection was labelled as such if one or more of the following problems occurred with regard to a peak: appeared too far from one or more of the predicted detector coordinate values, peak too wide in a particular coordinate value, bad

background or $\sigma(\text{background})$, bad pixel non-uniformity in the data array or a problem with pixel overflow - i.e. too dark [117]. In the hirudin-thrombin data collection, 183,791 reflection measurements were recorded in the output |Fobs| file, of which 164,250 (89.4% of the recorded reflection data) were labelled as 'good' and 19,541 were labelled with the bad, edge, or weak error codes.

MADNES also has a means to continually monitor whether a mis-alignment has occurred and a re-centering is needed [118]. During the course of data collection, the average difference between the predicted and observed Yms, Zms, and ϕ coordinates is calculated for each of the 16 regions of the detector. Moreover, the evaluation status of the 500 most recently processed reflections is stored. If average differences in the predicted and observed coordinates get too high, or if the fraction of 'good' reflections in the 500 most currently evaluated reflection list gets too low, the data collection automatically returns to the REFIN command for a re-alignment. Up to 250 reflections from the current list with the highest |Fobs| and suitable error codes are selected for refinement of missetting angles and detector position. (The unit cell parameters are not refined in a realignment as the ϕ range and the sampling of reciprocal space are too small to yield an accurate orientation matrix calculation.) After the refinement has been conducted, two quantities are calculated to assess if the re-centering has been successful [117]. The root mean square (r.m.s.) difference between the actual and the predicted spot positions on the detector in millimeters (rmscfd), and the r.m.s. difference in angles between the reciprocal lattice vector and the predicted one (rmsdeg) for the reflections selected to perform the refinement are calculated and checked against maximum acceptable values stored in

RmsChk. In the hirudin-thrombin data collection, the acceptable values for for these quantities were 0.2500 and 0.2000, respectively. If the calculated values were below the tolerance limits, the data collection would proceed; otherwise, the data collection would halt and prompt for user intervention.

III DATA REDUCTION

A. Converting the 2.8 Å Resolution Diffractometer Intensity Data to Structure Amplitudes

The conversion of each reflection intensity, $I(hkl)$, to its corresponding structure factor modulus, $|F(hkl)|$, is the initial step in the data processing of x-ray intensity data. The generation of a 2.8 Å resolution set of structure factor amplitudes from the diffractometer x-ray data of the hirudin-thrombin crystals proceeded in two stages. Initially, the 3.56 Å resolution HTN3 and the (3.63–2.80)Å resolution HTN2 data sets were reduced to structure factor amplitudes. Afterward, the data sets were scaled and averaged to produce a unique set of amplitudes to 2.8 Å resolution.

The intensity data were reduced using the program P-DATA, written by C.D. Buck, formerly of this laboratory. The program is able to accomplish this task, using input data from either a Nicolet P3/F or a Picker FACS-I four circle diffractometer, by means of the following equation:

$$|F(hkl)|^2 = \text{CONST} \times \text{ABS} \times \text{DEC} \times \text{LORPOL} \times I(hkl) \quad \dots(6)$$

where CONST is a scaling factor, ABS is an absorption correction factor, DEC is an x-ray induced decay correction factor, LORPOL is the Lorentz-polarization correction factor, and $I(hkl)$ is the background-corrected intensity of the reflection (hkl).

Background-corrected reflection intensities were obtained by averaging the background data associated with the reflections in shells of 2θ . The integrated intensity output for each reflection during the course of data collection is computed from the following equation [120]:

$$I = \left[\frac{\text{total scan count} - \text{sum of background counts}}{\text{background to scan time ratio}} \right] \times \text{scan rate} \quad \dots(7)$$

where measurements of the left and right background as well as of the total scan count for the peak are used in the calculation. Background-averaging of intensity data results in more accurate measurements for two reasons. First, each left and right background was measured only one-fifth as long as the peak in these intensity data collections. Second, since these background count values are generally low, statistical fluctuations can significantly affect the intensities calculated for weak reflections. Averaging the background of the intensity data for the HTN2 and the HTN3 data collections revealed that there was a definite 2θ dependence to the background measurements. The background data were also averaged in shells of ϕ as well as 2θ ; however, since no ϕ dependence could be discerned, the intensity data resulting from the 2θ shell background averaging were used in the P-DATA processing.

The formulation used to correct intensity measurements for absorption in the P-DATA program was based on the method proposed by North, Phillips, and Mathews [121]. The program uses absorption tables of $I_{\text{max}}/I(\phi)$ versus ϕ to apply the absorption correction, and such a table was obtained from the absorption curve depicted in Figure 20.

The decay correction factor, DEC, corrects for the intensity deterioration of a reflection due to x-ray exposure to the crystal, and it is generally a function of both exposure time and 2θ angle. The DEC factor has the form $1/(1-St)$ in the following equation:

$$I'(t) = I(t)[1/(1-St)] \quad \dots(8)$$

In equation (8), $I(t)$ is an intensity measurement recorded at time t in

the data collection, $I'(t)$ is an intensity measurement corrected for decay as if it were recorded at time zero. Furthermore, $S = -s/I'(0)$, where s is either the slope from an intensity versus exposure time plot for a single, periodically-measured monitor reflection, or the slope from an average intensity versus exposure time plot for a periodically-measured group of reflections, and $I'(0)$ is the corresponding extrapolated intensity at zero time.

The decay corrections for the HTN2 (3.6-2.8) Å data set were obtained from the linear plot of S versus average 2θ (Figure 23). Four points were used to generate the line; three points came from intensity decay plots of the three monitor reflections [(1,16,0), (16,1,0), and (0,0,32)], and the fourth point was the slope calculated by comparing peak heights from 5.9 Å resolution (hk0) Patterson projection maps computed using intensity data collected before and after the 3-D data collection. The three decay- and alignment-monitoring reflections decayed approximately (12-16)% over the course of the 57 hour data collection, while the 5.9 Å resolution (hk0) zonal reflections decayed 8% on an average. The HTN2 intensity data were collected from low to high resolution in shells of (24-27.5)°, (27.5-30.0)°, and (30.0-32.0)° in 2θ , respectively. From the S versus average 2θ plot, S values were extrapolated for the 2θ values 27.5° ($S = 0.00373 \text{ hours}^{-1}$) and 30.0° ($S = 0.00404 \text{ hours}^{-1}$). The (24.0-27.5)° intensity data were corrected for decay using $S = 0.00373 \text{ hours}^{-1}$, the (30.0-32.0)° intensity data were corrected for decay using $S = 0.00404 \text{ hours}^{-1}$, and the (27.5-30.0)° data were corrected for decay by interpolating between the two S values.

In the 3.56 Å resolution HTN3 data set, which was collected from high to low resolution using an observed reflection file, the DEC correction factors had a slightly different form than in equation (8),

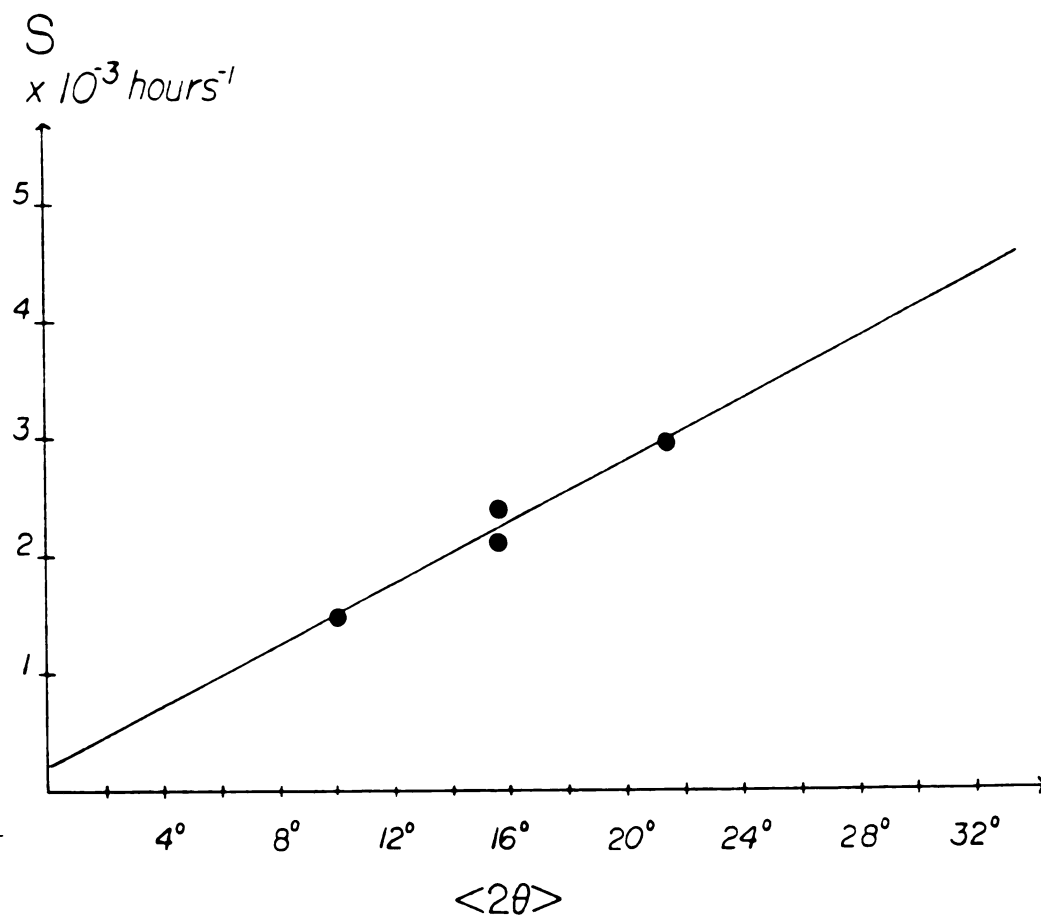


Figure 23. Dependence of X-ray Decay Correction with Scattering Angle for the HTN2 Hirudin-Thrombin (3.6-2.8) Å Resolution Data Collection.

and are contained in the expressions listed below:

$$I'(t) = I(t)[1/(1-2\theta kt)] \quad 2\theta > 12.0^\circ \quad \dots(9)$$

$$I'(t) = I(t)[1/(1-12kt)] \quad 2\theta \leq 12.0^\circ \quad \dots(10)$$

To obtain DEC corrections factors for this data set an S versus average 2θ plot was made, as the one shown in Figure 23, but now the slope of the S versus average 2θ plot, k, as well as the actual 2θ value of the reflection were used to obtain the appropriate decay correction. In the HTN3 data collection the three alignment- and decay-monitoring reflections decayed (5-11)% over the course of the 44 hour data collection, and the 5.9 Å resolution (hk0) zonal reflections decayed on an average of 4%. The k value determined from the S versus average 2θ curve for the HTN3 data collection was $9.04 \times 10^{-5} \text{ hours}^{-1}$.

The factor CONST was used to approximately scale the HTN2 intensity data to the HTN3 intensity data. The CONST value was obtained by finding the mean ratio of corresponding Patterson peak heights from the HTN2 and HTN3 5.9 Å resolution (hk0) Patterson projection maps calculated using intensity data obtained before the start of each 3-D data collection. The value for CONST determined to put the HTN2 intensity data roughly on the same scale as the HTN3 intensity data was 1.04. Moreover, the fact that the CONST value is so close to unity indicates that the x-ray scattering of the two crystals was comparable.

Once the P-DATA reflection processing was completed, it was possible to determine the percentage of collected data which were observed in the HTN2 and HTN3 intensity data collections. The HTN2 (3.6-2.8)Å resolution data were processed to 3483 unique structure factors out of 7597 possible observations, thus this data set had 46% of the unique data observed. In the HTN3 3.56 Å resolution data

collection, 5624 unique measurements were observable out of the 5769 reflections present in the reflection file. This number was unusually high because the data file contained only reflections observed to 3.56 Å resolution from a previous data collection, HTN1. However, if one considers that this observed reflection file represented 81% of the unique data to 3.56 Å resolution, the HTN3 data set had 79% of the unique data observed.

The two sets of reduced data contained 190 common structure factors, and these were used to scale the HTN2 structure amplitudes more accurately to those of HTN3. Before carrying out any further scaling, an R-factor was calculated for the overlapping reflections using the equation:

$$R\text{-factor} = \frac{\sum (|F_{\text{obs}}|_{\text{HTN3}} - |F_{\text{obs}}|_{\text{HTN2}}|)}{\sum (|F_{\text{obs}}|_{\text{HTN3}})} \quad \dots(11)$$

and the value was found to be 0.098. In an effort to reduce the above R-factor and scale the overlapping structure factors (and ultimately the two sets of structure factors more closely), the following ratio was calculated:

$$\sum (|F_{\text{obs}}|_{\text{HTN3}}) / \sum (|F_{\text{obs}}|_{\text{HTN2}}) \quad \dots(12)$$

which was found to be 0.96. Applying a scale factor of 0.96 to the HTN2 $|F(hkl)|$ values and recalculating the R-factor for the overlapping amplitudes yielded a value of 0.091, indicating that the re-scaling brought about improvement. To create a unique set of structure amplitudes to 2.8 Å resolution, all the HTN2 structure factors were

multiplied by the scale factor of 0.96, combined with the HTN3 reduced data, and the overlap amplitudes were averaged; the result was a set of 8917 unique $|F(hkl)|$ values. The $\langle |F(hkl)|^2 \rangle$ versus $\langle 2\theta \rangle$ distribution for these hirudin-thrombin structure factors is shown in Figure 24. The completeness of this 2.8 Å resolution set of $|F(hkl)|$ values was determined by loading these data into the PROTEIN [122] package of macromolecular crystallographic programs, and inserting a CHECK COMPLETENESS subcommand into the LOAD command file; the set of structure factors was found to correspond to 62% (8917/14362) of the possible unique reduced data.

B. Post-Data Collection Processing of the 2.3 Å Resolution

Area Detector Structure Factors

Although the MADNES-controlled FAST area detector diffractometer generated $|F_{obs}(hkl)|$ values for each reflection measured in the hirudin-thrombin data set, these structure factors still had to be corrected for absorption, decay, and scaling before they could be used for x-ray crystal structure analysis. However, the previously described means of carrying out absorption and decay corrections cannot generally be conducted with an area detector system. Azimuthal angle ψ scan data cannot be collected for an empirical absorption correction, nor can specific reflections be measured periodically to carry out a conventional decay correction. Nonetheless, one can make use of the real power of the area detector - its ability to measure large numbers of symmetry-equivalent reflections in a data set - to correct for absorption and decay, as well as for scaling.

By means of the program ABSCOR [123] the hirudin-thrombin FAST area detector diffractometer data were corrected for scaling, absorption, and

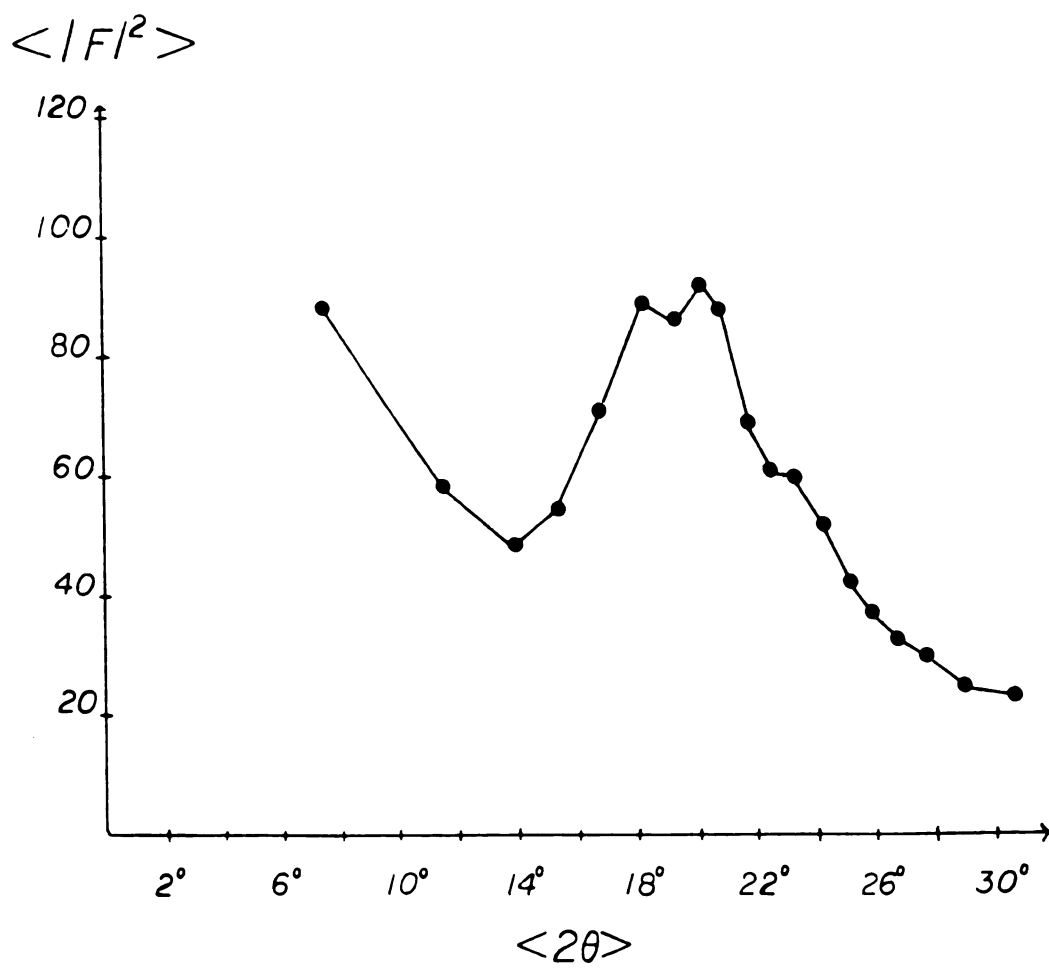


Figure 24. The $\langle |F_{obs}|^2 \rangle$ versus $\langle 2\theta \rangle$ Distribution for the 2.8 Å Resolution Hirudin-Thrombin Diffractometer Data Set.

decay as well as for the non-uniformity of response of the detector. The intensity data output file from the MADNES program (the so-called D 13 file) serves as the input file for ABSCOR.

In a generalized macromolecular crystallographic data collection, it is possible that the crystal will not be fully bathed in the primary x-ray beam, and that variable scattering volumes of the crystal will be irradiated during the course of the data collection. (In the hirudin-thrombin data collections discussed in this dissertation, the crystals should have been bathed uniformly in the x-ray beam throughout the course of the experiments.) In this situation, it is reasonable to assume that reflections measured in a batch or in a block, which consists of one or more contiguous batches, should have similar scale and temperature factors which correct for changes in irradiated volume and for crystal decay due to x-ray exposure. One can represent a structure factor amplitude corrected for absorption, decay, scaling, and non-uniformity of response of the detector by the following formula [123]:

$$\begin{aligned}
 |F|_{\text{corr-h}(i)} = & |F|_{\text{obs-h}(i)} * S_i * \exp[-B_i(\sin(\theta)^2/\lambda^2)] \\
 & \times [A'_{0\text{-h}(i)} * (T_{11}, \dots, T_{33}) * A'_{s\text{-h}(i)} * (T_{11}, \dots, T_{33})]^{-1/2} \\
 & \times (K_{\text{nonuniform}})^{1/2}
 \end{aligned}
 \tag{13}$$

where $|F|_{\text{obs-h}(i)}$ is the observed structure factor amplitude of reflection h of the i th batch; h is the triple index (hkl); i is the index of the reflection batch; S_i is the individual scale factor of the i th reflection batch; B_i is the temperature factor of the i th reflection batch; $A'_{0\text{-h}(i)}$ is the transmission factor of the primary beam of reflection $h(i)$; $A'_{s\text{-h}(i)}$ is the transmission factor of the secondary

beam of reflection $h(i)$; $T_{11}, \dots, T_{33}$ are the components of the symmetric tensor T which describes the transmission properties of the crystal; and, $K_{nonuniform}$ is the post-correction factor for the nonuniformity of response of the FAST area detector.

The transmission factors $A'_{0-h(i)}$ and $A'_{s-h(i)}$ are obtained from the tensor ellipsoids:

$$A'_{0-h(i)}^{-1/2} = (T_{11}u_0^2 + T_{22}v_0^2 + T_{33}w_0^2 + 2T_{12}u_0v_0 + 2T_{13}u_0w_0 + 2T_{23}v_0w_0) \quad \dots(14)$$

$$A'_{s-h(i)}^{-1/2} = (T_{11}u_s^2 + T_{22}v_s^2 + T_{33}w_s^2 + 2T_{12}u_sv_s + 2T_{13}u_sw_s + 2T_{23}v_sw_s) \quad \dots(15)$$

where these factors are generated by the following operations [123]:

$$A'_{0-h(i)}^{-1/2} = [v_{0-h(i)}]^T T v_{0-h(i)} \quad \dots(16)$$

$$A'_{s-h(i)}^{-1/2} = [v_{s-h(i)}]^T T v_{s-h(i)} \quad \dots(17)$$

with

$$v_{0-h(i)} = \begin{matrix} u_0 \\ v_0 \\ w_0 \end{matrix} \quad \text{and} \quad v_{s-h(i)} = \begin{matrix} u_s \\ v_s \\ w_s \end{matrix} \quad \dots(18)$$

being the primary and secondary direction vectors of the normalized beam in the goniometer head coordinate system for reflection $h(i)$.

The scale (S_i) and temperature (B_i) factors for each block of reflections, and the symmetric transmission tensor matrix elements (T_{11} , T_{22} , T_{33} , T_{12} , T_{13} , T_{23}) for the crystal are determined by sorting the reflection structure factors into symmetry-equivalent sets (including the Friedel mates, where $|F(hkl)|$ and $|F(-h-k-l)|$ constitute a Friedel

pair) and minimizing by least squares methods the following sum of squared differences of pairs of symmetry-equivalent reduced data:

$$\sum_{h(i)} \sum_{j=1}^{k[h(i)]} \sum_{l=j+1}^{k[h(i)]} (\ln|F|_{\text{corr}}^{h(i),j} - \ln|F|_{\text{corr}}^{h(i),l})^2 = \min \quad \dots(19)$$

where $k[h(i)]$ is the number of reflections in a set of symmetry equivalent reflections of index h of the i th reflection block. Natural logarithms are used so that the difference is linear with respect to S_i and B_i , and the scale and temperature factors of the first block are fixed at 1.0 and 0.0, respectively.

It is quite common for area detectors to display a non-uniformity in detector response; this problem is generally corrected by means of a pixel-to-pixel correction table obtained by an appropriate calibration method prior to the initiation of the data collection. The aforementioned least squares minimization allows one to introduce a non-uniformity of response post-correction. After the minimization of the scaling and absorption correction parameters has been completed, corrected structure factor amplitudes of the symmetry-equivalent reflections are compared. A deviation from the average value is expressed as a correction factor for particular pixel values. The table which results from this is composed of 64 8x8 pixel boxes and it can be smoothed by various methods. The values from this table are the basis of the correction factor $K_{\text{nonuniform}}$.

In the ABSCOR processing of the hirudin-thrombin FAST output data, 183,791 reflections were input, of which 164,250 were flagged as 'good' and used for further processing and 19,541 with error flags were ignored. The block size used to restrict scale and temperature factors was 2 sequentially measured reflection batches, and since the batch size

used in the data collection was 2.5° in ϕ , the block size was 5° . Previous work has shown that block sizes of 3° to 6° in ϕ give reliable and reproducible results [123]. The end result of this was that the 164,250 FAST |Fobs| were sorted into 31 blocks, and that after a complete least squares minimization of equation (19), the six matrix elements of the symmetric transmission tensor, as well as the 31 scale factors, and 31 temperature factors were determined.

The complete ABSCOR processing of the hirudin-thrombin FAST reflection output data proceeded via 3 stages, and the results are summarized in Table 4. The progress at each stage was assessed by the calculation of the R-factor presented below:

$$R_{\text{merge}} = \frac{\sum_h \sum_j |I(h)_j - \langle I(h) \rangle|}{\sum_h \sum_j I(h)_j} \quad \dots(20)$$

where h represents the triple index (hkl) and j represents the number of equivalent reflections for a given $I(h)$. Prior to any minimization of equation (19), the ABSCOR output revealed that the 164,250 reflection intensities corresponded to 23,521 unique reflections, and that the initial R-merge between individual reflections was 0.168. In the first stage of the ABSCOR processing, scale and temperature factors were refined and applied to the input |Fobs|, and the result was that the Rmerge after scaling was 0.120 between individual reflections. In the second stage of the ABSCOR processing, the scale factors, temperature factors, and the six components of the symmetric transmission tensor T were refined and applied to the |Fobs|, which resulted in an Rmerge of 0.111 between individual reflections. Finally, in the third stage of the ABSCOR processing, the Knonuniform detector non-uniformity of response correction factors were applied and the resultant Rmerge was

Table 4. Reflection Statistics on the Hirudin-Thrombin
FAST Data Set Evaluated by MADNES and Corrected
using ABSCOR

Total number of reflections:	164,250
Number of unique reflections:	23,521
Rmerge*-uncorrected:	0.168
Rmerge*-scaling applied:	0.120
Rmerge*-scaling & absorption corrections applied:	0.111
Rmerge*-scaling, absorption, & post-correction of detector non-uniformity of response applied	0.108
Rmerge**	0.042
Completeness of data to 2.3 Å resolution	0.91

$$R_{\text{merge}} = \frac{\sum (I - \langle I \rangle)}{\sum I}$$

* = between individual reflections

** = between averaged Friedel pairs

0.108.

The resultant output file from the ABSCOR processing contained data for 41,062 reflections, in which there were 23,521 unique data and 17,541 Friedel mate data. The unique and Friedel mate data were averaged by means of the PROTEIN program package to generate a set of 23,521 unique structure factors. The Rmerge calculated in PROTEIN between averaged Friedel pairs was 0.042 and the data set was found to be 91% complete to 2.3 Å resolution.

IV MOLECULAR REPLACEMENT CALCULATIONS

A. Introduction

In any macromolecular crystallographic investigation, once a requisite set of structure factor amplitudes have been obtained, the next major task is to work toward generating an electron density map for structure interpretation and model-building. The electron density at a point (x,y,z) in the crystal can be determined by computing the following Fourier summation:

$$\rho(x,y,z) = 1/V \sum_{\text{all } h, k, l} \sum |F(hkl)| \exp\{i\alpha(hkl)\} \exp\{-2\pi i(hx + ky + lz)\} \dots(21)$$

where $|F(hkl)|$ is the amplitude of the reflection (hkl), $\alpha(hkl)$ is the phase of the reflection, V is the volume of the unit cell, and the summations extend over all of the observable reflections. However, calculating such an electron density map is non-trivial, as it requires a knowledge of the phase angles which cannot be measured experimentally. In this investigation, the molecular replacement method [124] or Patterson search method [125] was used to overcome this "phase problem".

The molecular replacement method has two main types of applications [126]. In crystals which possess more than one molecule per asymmetric unit, the technique can be employed to determine the orientation of non-crystallographic symmetry elements within an asymmetric unit of the unit cell. The method can also be used to generate a set of trial phases for an unknown structure by orienting and positioning a similar molecule of known tertiary structure in the unit cell of the unknown structure. This latter application was employed to solve the structure of the rHV2-K47-human α -thrombin complex using PPACK-human α -thrombin as

a model [127]. The structure solution of the hirudin-thrombin complex was a problem ideally suited for the molecular replacement method. First of all, both the model and the crystal contained the same protein, human α -thrombin. Second, thrombin accounts for roughly 84% of the total molecular weight of the complex. Thus, it seemed reasonable to assume that one could generate a good initial Fourier map based on a Patterson search using the PPACK-thrombin structure as a model.

The molecular replacement method is often referred to as a Patterson search method since calculations are performed using the Patterson functions of the known model and the unknown structure of interest. A Patterson function is a Fourier synthesis which uses the square of the structure factor amplitudes, has the form:

$$P(uvw) = 2/V \sum_{h=0}^{+h} \sum_{k=0}^{+k} \sum_{l=0}^{+l} |F(hkl)|^2 \cos\{2\pi(hu + kv + lw)\}, \quad \dots(22)$$

and produces peaks corresponding to all of the interatomic vectors [128]. A peak at the point (u,v,w) implies that there exist atoms at (x1,y1,z1) and (x2,y2,z2) such that $u = x1 - x2$, $v = y1 - y2$, and $w = z1 - z2$. The height of a peak in the Patterson function is approximately proportional to the product of the atomic numbers of the atoms which define the ends of the vector, (Z1 x Z2). Moreover, if the unit cell of a molecule contains N atoms there will be N^2 peaks in the Patterson map, of which there will be N origin or self-vector peaks and (N^2-N) non-origin peaks (half of which are centrosymmetric). Thus, the Patterson function is a unique Fourier synthesis which contains structural information (relative structural information), but which requires no knowledge of the phases and can be calculated using experimentally-obtainable quantities, the structure factor amplitudes.

The Patterson search conducted with PPACK-human α -thrombin as a model to generate a trial structure of the hirudin-thrombin complex proceeded in three stages common to molecular replacement calculations [124]: (1.) The relative orientation or rotational relationship of the thrombin model with respect to the unknown hirudin-thrombin structure was determined by "rotation function" [126] calculations involving the intramolecular Patterson vectors of the model and the crystal Patterson synthesis. (2.) The correct translation of the properly oriented thrombin model in the unit cell of the hirudin-thrombin crystal was determined by "translation function" [129,130] calculations involving a modified crystal Patterson map and intermolecular model Patterson vectors. Once the first two stages were successfully achieved, the equivalence between a point x' in the unknown hirudin-thrombin structure and the corresponding point x in the model thrombin structure could be expressed by the following relationship:

$$x' = [C]x + d \quad \dots(23)$$

where $[C]$ represents the rotation matrix determined from the first stage of the Patterson search, and d represents the translation vector determined from the second stage. (3.) The properly oriented and translated thrombin model in the hirudin-thrombin unit cell was used to calculate a set of phase angles which were combined with 8.0 to 3.0 Å resolution observed structure factors to produce a trial electron density map for model-building.

The three stages in the Patterson search calculations will now be discussed in more detail.

B. The Rotation Function Calculations

A model Patterson vector set was produced from the coordinates of PPACK-thrombin [6]. The thrombin model (284 residues) contained all but 11 of the protein residues in the PPACK-thrombin crystal structure; the first five residues (1H-1D) and the last four residues (14K-15) of the thrombin A chain, as well as the last 2 residues of the thrombin B chain (246-247) were deleted as they had poor or no density in the PPACK-thrombin electron density maps (Table 1). The search model also contained 24 internal water oxygen atoms, but it lacked the covalently-bound peptidyl inhibitor PPACK. Triclinic P1 structure factors were obtained by Fourier inversion of a model electron density map using the program FFTSF, which contains two main sub-programs, RHOGEN and P1SF. RHOGEN calculated an electron density map of the model positioned in a triclinic P1 orthogonal cell of $120 \times 120 \times 120 \text{ \AA}^3$ using 8.0 to 3.0 $\text{\AA}$ resolution data and an overall temperature factor for the atoms of $20 \text{ \AA}^2$. P1SF then calculated structure factors by fast Fourier inversion of the electron density. Finally, these calculated structure factors were used to generate an 8.0 to 3.0 $\text{\AA}$ resolution Patterson function of the thrombin model in a $120 \times 120 \times 120 \text{ \AA}^3$ orthogonal cell.

However, not all of the peaks in this model Patterson map were used for the rotational search with the hirudin-thrombin crystal Patterson function. In the rotation function calculation one is only interested in intramolecular vectors, thus only Patterson vectors of length greater than 3.0 $\text{\AA}$ and less than 20.0 $\text{\AA}$ were selected. The lower limit of 3.0 $\text{\AA}$ eliminates the huge origin peak from the calculation. The upper limit of 20.0 $\text{\AA}$ is meant to prevent intermolecular vectors from entering the rotation function calculation. Moreover, as PPACK-thrombin is a nearly spherical molecule of dimensions $45 \times 45 \times 50 \text{ \AA}^3$ [6], an upper limit of

20 Å for intramolecular vectors appeared to be a good estimate. 'Weak' or small magnitude Patterson peaks were excluded from the calculation by using an appropriate cutoff limit. Finally, as only half of the complete triclinic model Patterson map is unique, the peak search was restricted to 60 Å along the w axis (although any of the three axes could have been chosen). The end result was that 7565 model Patterson peaks were selected for a rotation function calculation. The hirudin-thrombin crystal Patterson function was calculated using observed diffractometer structure factor amplitudes over the same resolution range.

The rotational search was conducted with the SEARCH routine in the PROTEIN program package. The model thrombin Patterson vector set was rotated and interpolated to the nearest grid points in the hirudin-thrombin crystal Patterson function, and a product function was calculated. This process was repeated until all angles over a specified angular range and increment size were examined. The set of coordinates yielding the highest product function corresponded to the transformation necessary to bring the model and crystal Patterson vectors into the same orientation. The Huber ("ROH") angles ψ , θ , and ϕ [122] were used to conduct the rotational Patterson search. A rotational operation ψ , θ , ϕ corresponds to a rotation of the Cartesian axes by ψ about the z axis, a rotation θ about the new x axis, and a rotation ϕ about the new y axis. The Huber angle rotation range for the calculation was: ψ , 0-100°; θ , 0-180°; and ϕ , 0-180°. All angles were incremented by 5° in the calculations. The highest peak in the search occurred at the Huber angles (5.0°, 175.0°, 80.0°), and it was 7.7 σ above the mean. To refine this solution, a finer rotational search was carried out over the range: ψ , 0-10°; θ , 170-180°; and ϕ , 75-85° using the angular increment of

1.0°. The refinement improved the peak height to 8.1σ above the mean, and the refined coordinates were (3.5°, 176.0°, 81.5°).

C. The Translation Function Calculations

Once the orientation of the model molecule is found in the unit cell of the target crystal, the approximate position is found by conducting a translation search. However, in achieving this translation result one has actually positioned not just one molecule, but the entire set of symmetry-related molecules in the crystal unit cell. This is due to the fact that the relative position between symmetry-related molecules is fixed. For example, the hirudin-thrombin complex crystallizes in space group $P4_32_12$ with one molecule per asymmetric unit and 8 molecules in the unit cell. The relative positions of the 8 symmetry-equivalent molecules in the unit cell are given in the following relations [131]:

Molecule 1:	x	y	z	Molecule 5:	y	x	-z
Molecule 2:	-x	-y	1/2+z	Molecule 6:	-y	-x	1/2-z
Molecule 3:	1/2-y	1/2+x	3/4+z	Molecule 7:	1/2-x	1/2+y	3/4-z
Molecule 4:	1/2+y	1/2-x	1/4+z	Molecule 8:	1/2+x	1/2-y	1/4-z

...(24)

Moreover, the relative positions of the 7 intermolecular vectors are fixed as well; the complete intermolecular vector set is listed below:

1-2:	2x	2y	-1/2	1-6:	(x+y)	(y+x)	-1/2+2z
1-3:	-1/2+(x+y)	-1/2-(x-y)	-3/4	1-7:	-1/2+2x	-1/2	-3/4+2z
1-4:	-1/2+(x-y)	-1/2+(x+y)	-1/4	1-8:	-1/2	-1/2+2y	-1/4+2z
1-5:	(x-y)	(y-x)	2z				

...(25)

The positioning of the thrombin model in the hirudin complex unit cell was achieved by means of a set of programs written by E.E. Lattman (and modified by J. Deisenhofer, R. Huber, and M. Schneider) which calculates a translation function as described by Crowther and Blow [130]. The Crowther and Blow translation function has the form:

$$Ts(t) = \sum_h (|F_{obs}(h)|^2 - \sum_{i=0}^{n-1} |F_m(h \cdot A_i)|^2) \times F_m(h) \cdot F_m'(h \cdot A) \cdot \exp(-2\pi i \cdot h \cdot t) \quad \dots(26)$$

where h represents the triple index (hkl); $F_{obs}(h)$ and $F_m(h)$ are the observed and model structure factors at triple index h and are on an absolute scale; $F_m'(h \cdot A)$ is the complex conjugate of the model structure factor of a symmetry-mate molecule related by the symmetry operator A ; t is the translation vector; and the term $|F_m(h \cdot A_i)|^2$ subtracts out the intramolecular vectors from the observed Patterson function at each of the $n-1$ symmetry related positions using the matrix A_i , A_0 being the identity matrix of the model. The above translation function $Ts(t)$ is specific for the intermolecular vector at the translation t between a model molecule and its symmetry-mate related by the symmetry operator A ; there will be $n-1$ such functions for the $n-1$ unique intermolecular vectors relating a given molecule to its $n-1$ symmetry-mate molecules in the unit cell of a crystal.

Thus, the Crowther and Blow translation function makes use of the intermolecular vector equations defined by the space group of the crystal (relation 25 above) to determine the correct position of the properly-oriented model in the crystal cell. Basically, for each unique intermolecular vector translation function, all the intermolecular vectors were calculated for the model and its appropriate symmetry mate.

Each resulting set of intermolecular vectors is then translated throughout the crystal Patterson synthesis modified to remove the intramolecular vectors. A product function is calculated between the modified crystal Patterson map and the model intermolecular vector set at each point in the translation search. At the correct position of the model in the crystal, the product function will have a maximum value. In many cases, as in the hirudin-thrombin search, if all the translation functions for the set of unique intermolecular vector equations are calculated, the correct position of model in the crystal cell will be overdetermined. In other words, the correct position of the model should account for a high product function (if not the highest) peak in each of the intermolecular translation function searches.

The translational searches were conducted using 8.0 to 3.0 Å resolution crystal and thrombin model structure factors, and a translation function calculation was conducted for each unique intermolecular vector in the $P4_32_12$ space group (relation 25 above). The results of the search are listed in Table 5. The highest peak in all of the calculations occurred in Harker sections at consistent locations which positioned the molecule at $x = 67.42$ Å, $y = 3.00$ Å, and $z = 52.91$ Å, and verified the correct enantiomorphic space group to be $P4_32_12$ as opposed to $P4_12_12$. The correct position of the molecule also gave an outstanding peak in each Harker section; the highest peak in each section was on the average 13.3σ above the mean, and 7.5σ higher than the next highest peak. The R-value for this solution, where the R-value is defined as

$$R\text{-value} = \frac{\sum_h ||F_{obs}(h)| - |F_{cal}(h)||}{\sum_h |F_{obs}(h)|}, \quad \dots(27)$$

was found to 0.39.

The rotational and translational parameters were further refined with the rigid body refinement-Fourier transform fitting program TRAREF [132]. The program TRAREF refined the orientational and translational parameters of a thrombin model by fitting its molecular Fourier transform or search model structure factors to the observed hirudin-thrombin structure factor amplitudes. Model structure factors were calculated based on the position of the model in the crystal cell determined from the translation search (Table 5). The fitting was done by calculating derivatives of the molecular Fourier transform with respect to the three orientational angles, the three positional parameters, a scaling factor, and an overall temperature factor. The system of equations was then reduced to normal equations and improved parameters were determined. The TRAREF refinement shifted the center of gravity position for the thrombin model in the crystal to $x = 67.44 \text{ \AA}$, $y = 2.87 \text{ \AA}$, and $z = 52.87 \text{ \AA}$, and the R-value for this new position was 0.34.

D. The Initial Hirudin-Thrombin Electron Density Map Calculation

An initial hirudin-thrombin Fourier synthesis at 8.0 to 3.0 $\text{\AA}$ resolution was calculated using coefficients of the form

$$W*(2|F_{\text{obs}}(h)| - |F_{\text{cal}}(h)|) * \exp(i*\alpha(h)),$$

...(28)

where h represents the triple index (hkl); $|F_{\text{obs}}(h)|$ is the observed structure factor amplitude at triple index h ; $|F_{\text{cal}}(h)|$ and $\alpha(h)$ represent the model structure factor amplitude and phase at triple index h calculated from the TRAREF-refined position of the thrombin model in

Table 5. Results of the Translation Search

Vector	Harker section	Highest peak Position (Å)			Height (σ)	Second highest peak Height (σ)	Position of the molecule (Å)
		U	V	W			
1-2	W=1/2	43.94	6.00	66.30	13.8	6.7	
1-3	W=3/4	24.87	72.01	33.21	13.9	4.7	
1-4	W=1/4	18.98	24.87	99.38	13.9	4.8	x=67.42
1-5	-	64.52	26.46	105.75	12.5	6.7	y= 3.00
1-6	-	70.28	70.41	39.36	13.0	6.6	z=52.91
1-7	V=1/2	89.26	45.50	6.36	13.6	6.2	
1-8	U=1/2	45.50	51.72	72.81	12.3	5.2	

Peak height is given in standard deviations above the mean value of the function.

The vectors are as follows:

1-2: $2x, 2y, -1/2$

1-6: $(x+y), (y+x), -1/2+2z$

1-3: $-1/2+(x+y), -1/2-(x-y), -3/4$

1-7: $-1/2+2x, -1/2, -3/4+2z$

1-4: $-1/2+(x-y), -1/2+(x+y), -1/4$

1-8: $-1/2, -1/2+2y, -1/4+2z$

1-5: $(x-y), (y-x), 2z$

the crystal; and W represents the Sim weight [133,134] for the model phase angle $\alpha_c(h)$. The map size was 100 x 100 x 140 grids, which corresponded to 0.91 Å/grid along the a and b axes, and 0.94 Å/grid along the c axis.

Such so-called "2Fo-Fc" maps are commonly used in macromolecular crystallography when one is working with an incomplete structure, as in this situation where the phasing model lacked hirudin and solvent structure. The map is basically a point-by-point summation of an F_o and an $(F_o - F_c)$ or ΔF synthesis, and is useful for several reasons. First of all, like an F_o synthesis, it displays the input model. In addition, the map also indicates how additional structure can be accommodated, or whether existing structure can be reoriented or removed from the present model, like a ΔF synthesis [135].

The Sim weighting scheme for the electron density map calculation was designed to weight each $\alpha_c(h)$ by the probability of it being correct, and on the notion that the best weights for structure amplitudes minimize the mean square error in the electron density map due to phase angle error. Based on previous work by Blow and Crick [136], Sim showed that such weights are defined as

$$W = \int_0^{2\pi} \cos E p(E) dE, \quad \dots(29)$$

where

$$E = \alpha_c(h) - \alpha_m(h) \quad \dots(30)$$

and the probability that the phase difference lies between E and $E + dE$ is given by the probability function $p(E)$. This probability function has the form

$$p(E) = \exp(X \cos E) / 2\pi I_0(X), \quad \dots(31)$$

where I_0 is a zero-order modified Bessel function, and X is given by

$$X = 2|F_{obs}(h)| * |F_{cal}(h)| / I_u \quad \dots(32)$$

in which I_u represents the contribution from the "unknown structure" or the atoms in the structure not included in the model. In the Sim weighting scheme used to produce the hirudin-thrombin trial electron density map, I_u was approximated by the quantity

$$I_u = \langle ||F_{obs}|^2 - |F_{cal}|^2 \rangle = \sum_h ||F_{obs}(h)^2 - F_{cal}(h)^2|| / N, \quad \dots(33)$$

where N is the number of observed structure factors included in the calculation, and I_u was calculated in 10 shells of equal thickness in reciprocal space. The Sim weight definition given above in equation (29) can be further simplified to the form

$$W = I_1(X) / I_0(X), \quad \dots(34)$$

where I_1 is a first-order modified Bessel function. The Sim weight is zero when $X=0$ and 1.0 when X is infinity; a plot of the Sim weighting factor W against X is given in Figure 25 [137].

The first Fourier map was displayed on an Evans and Sutherland PS390 interactive graphics display system and it was encouraging, as it clearly had density to accommodate hirudin as well as thrombin. The initial model-building which was conducted using this density map, as well as further graphics interventions and EREF energy refinement which lead to a mildly-refined 2.3 Å resolution hirudin-thrombin structure [127], will be summarized next.

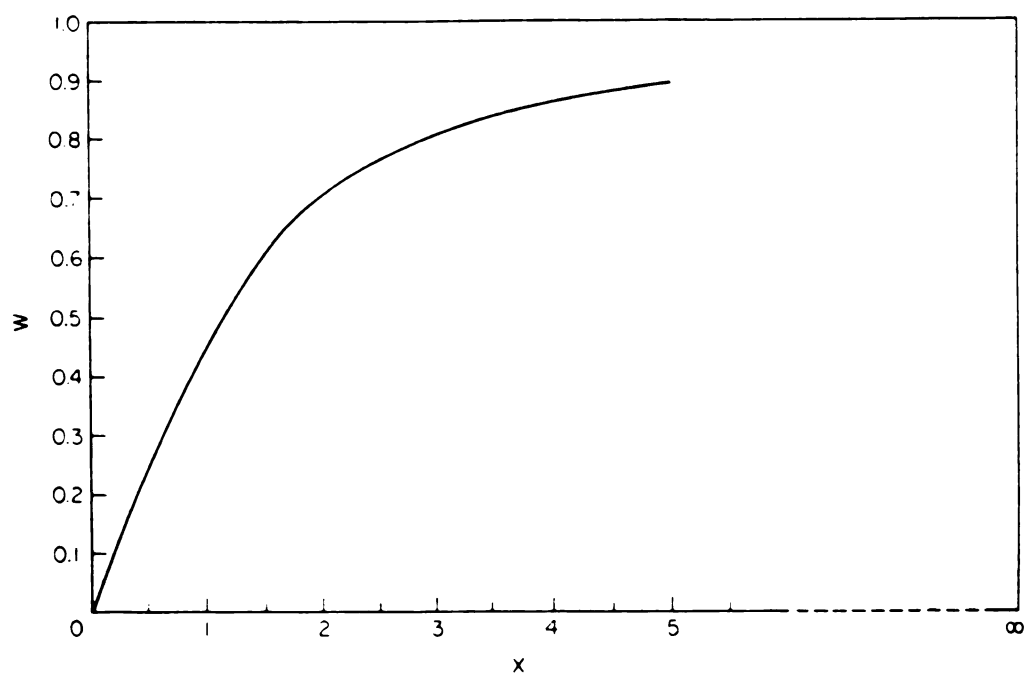


Figure 25. Plot of Sim weighting factor W versus X where $W = I_1(X)/I_0(X)$ and $X = 2|F_{obs}(h)||F_{cal}(h)|/I_u$ (taken from [137]). $I_1(X)$ and $I_0(X)$ are first- and zero-order modified Bessel functions, respectively. I_u is an estimate of the contribution from the unknown structure, or that which is not accounted for in the model.

V MODEL-BUILDING AND EREF REFINEMENT

The hirudin-thrombin complex was initially refined to an R-value of 0.193 with 7.0 to 2.3 Å resolution data by an iterative process of model-building and EREF refinement [127]. The model-building work was conducted interactively using the PSFRODO version [138] of the molecular graphics program FRODO [139] by fitting protein and solvent atoms in Fourier maps viewed on a Evans and Sutherland PS390 graphics display system. The model-building was initiated on the 3.0 Å (2Fo-Fc) Sim-weighted Fourier map resulting from the Patterson search calculations with the thrombin model. The map clearly displayed density for thrombin, which most of the model thrombin structure could easily accommodate, and had a significant fraction of additional density for the recombinant hirudin. At this point, the coordinates of the hirudin structure determined using two-dimensional NMR techniques by Folkers et al. [79] proved to be very useful in tracing residues 3' to 48' (where a prime after a number designates a hirudin residue) of the NH₂-terminal domain of hirudin. The most exciting aspect of these initial model-building studies was to see continuous density for the last ten COOH-terminal residues of the hirudin molecule, especially for the sidechains of Phe 56' and Tyr 63'. This initial hirudin-thrombin model was refined using the energy-restraint crystallographic refinement program EREF [140], and the resulting EREF structure was inspected and modified at the interactive graphics display. Sim-weighted density maps - positive (2Fo-Fc) and negative as well as positive (Fo-Fc) Fourier maps - were used in the course of model-building. The cyclic model-building and refinement procedure was repeated six times with gradual increases in resolution to 2.3 Å; the pertinent details of this

work are summarized in Table 6. In the second and later stages of this process, the FAST area detector structure factors were used solely. Water molecule oxygen atoms were inserted in the third stage, and in later stages. If a strong peak (generally $> 3.5\sigma(\Delta\rho)$) occurred in both the 7 to 2.3 Å resolution (2Fo-Fc) and (Fo-Fc) Fourier maps, and was in a stereochemically reasonable position for a water molecule, a water oxygen atom was inserted. Individual temperature factors were also refined. After the six rounds of model-building and refinement, the structure of the complex had an R-value of 0.193 for 21,960 reflections between 7.0 and 2.3 Å resolution. The structure included 207 water molecules, had a mean individual temperature factor of 35.7 Å², and the r.m.s. deviations for bond lengths and bond angles were 0.013 Å and 2.4°, respectively. The final model parameters of the EREF-refined hirudin-thrombin complex are summarized in Table 7.

The EREF macromolecular refinement program simultaneously minimizes a realistic potential energy function and a crystallographic residual. The complete potential energy function includes terms for bond stretching, bond angle bending, torsion potentials, and non-bonded and electrostatic interactions, and has the following form [141]:

$$\begin{aligned}
 E = & \sum_{\text{bonds}} 1/2K_b(b_i - b_0)^2 + \sum_{\text{bond angles}} 1/2K_t(t_i - t_0)^2 \\
 & \sum_{\text{torsion angles}} K_o\{1 + \cos(m\theta_i + d)\} + \sum_{\text{non-bonded interactions}} (A/r_i^{12} + B/r_i^6) + \sum q_i q_j / r \\
 & \dots(35)
 \end{aligned}$$

where K_b is the bond stretching force constant; b_0 the equilibrium bond length; K_t is the bond angle bending force constant; t_0 is the equilibrium bond angle; K_o is the torsion barrier; m is the periodicity of the torsion barrier; d is the phase of the barrier; and A and B are

Table 6. Course of Model-Building and EREF Refinement

Stage	R-value	Resolution (Å)	No. of atoms	Total Energy (kcal/mole)	Comments
0	0.349	3.0	2328	-	Rotation, translation structure; Nicolet P3/F data set used.
1	0.245	2.8	2495	-1458	Model rebuilt on display (1); 167 protein atoms added; 8 cycles of EREF including resolution extension.
2	0.291	2.5	2495	-1721	Model rebuilt on display (2); 9 cycles of EREF including resolution extension and 2 cycles of B-factor refinement; FAST data set used.
3	0.261	2.3	2677	-1763	Model rebuilt on display (3); 182 protein atoms added; 14 cycles of EREF including resolution extension and 4 cycles of B-factor refinement.
4	0.221	2.3	2787	-2175	Model rebuilt on display (4); 17 protein atoms and 93 water molecules added; 15 cycles of EREF and 4 cycles of B-factor refinement.
5	0.205	2.3	2848	-2128	Model rebuilt on display (5); 10 protein atoms deleted and 71 water molecules added; 19 cycles of EREF refinement and 4 cycles of B-factor refinement.
6	0.193	2.3	2915	-2221	Model rebuilt on display (6); 24 protein atoms and 43 water molecules added; 13 cycles of EREF and 4 cycles of B-factor refinement.

**Table 7. Final Model Parameters of the EREF-refined
rHV2-K47 Human α -Thrombin Structure.**

Number of protein atoms	
a) Thrombin	2296
b) Hirudin	411
Number of solvent atoms	207
r.m.s. deviation from target values	
Bond lengths ($\text{\AA}$)	0.013
Bond angles (deg.)	2.36
Total internal energy (kcal/mole)	-2221
Resolution range ($\text{\AA}$)	7.0 - 2.3
Number of unique reflections used for refinement	21960
R-value	0.193
Mean B value ($\text{\AA}^2$)	35.7

the coefficients of the Lennard-Jones potential for non-bonded interactions. The Lennard-Jones potential parameters used were those specified by Levitt [141], and the values of the force constants were derived from the vibrational spectra of small molecules. Moreover, the bond angle and torsion angle force constants were corrected for the missing hydrogen atoms in the structure [142]. The crystallographic residual has the form

$$X = \sum_h (|F_{\text{obs}}(h)| - |F_{\text{cal}}(h)|)^2. \quad \dots(36)$$

A cycle of EREF refinement proceeds in three steps [142]. In the first step, the system of normal equations for the diagonal-matrix least-squares refinement of atomic parameters is developed from an $(F_o - F_c)$ difference Fourier map. In step two, the potential energy of the model is minimized along with a crystallographic residual to refine atomic positions. If atomic temperature factors are refined, this minimization is bypassed. In the third step, new structure factors are calculated from the model resulting from step two.

The system of normal equations generated in step 1 of a refinement cycle uses the $(F_o - F_c)$ difference Fourier map to minimize the crystallographic residual in equation (36). These equations can be compactly represented by the expression

$$A \cdot dp = b \quad \dots(37)$$

where dp represents the vector of desired parameter shifts (positions, occupancies, B-values); the matrix A is approximated by a diagonal matrix in which the diagonal elements are estimated for each atom type. The components of the vector b , which represent the shifted parameters

of vector dp , are obtained by convoluting the Fourier transform of the atomic form factor (i.e. the density about a given atom) with the gradient of the $(F_o - F_c)$ map around the atomic site in the case of a positional refinement, or with the difference map itself in the case of occupancy or B-factor refinement. The convolution is generally calculated in a box of $3 \times 3 \times 3$ grid points around the atomic positions, and the grid spacing used is approximately one-third of the maximum resolution in angstrom units.

The function which is minimized in step 2 of an EREF cycle is

$$R = E + kX \quad \dots(38)$$

where E represents the conformational energy of equation (35), X represents the crystallographic residual of equation (37), and k is a weighting factor which can be used to emphasize or de-emphasize the diffraction pattern in the minimization process. For example, when $k = 0$ a pure energy refinement is conducted, and when k is large, a pure crystallographic refinement is performed. Generally, during the initial cycles of a refinement run, the k used was 5×10^{-6} , and it was gradually increased in the following cycles from 3×10^{-5} to up to 7×10^{-5} in the last cycles.

The final step of the EREF cycle run involved generating a new set of calculated structure factors based on the energy minimization results or B-factor refinement conducted in the cycle. The FFT (Fast Fourier Transform) program of Ten Eyck [143,144] was used for this purpose. The program initially produces a model electron density map from which it calculates the structure factors by a Fourier inversion.

VI MODEL-BUILDING AND PROLSQ REFINEMENT

Using the final coordinates of the EREF-refined hirudin-thrombin structure as a basis (Table 7), the iterative model-building and refinement procedure was continued using the restrained least squares refinement program of Hendrickson and Konnert [145]. This program, PROLSQ (an acronym for PROtein Least Squares), incorporates stereochemical information into the refinement process by treating this knowledge as additional observations. Basically, PROLSQ minimizes a grand function $\Phi = \sum \phi_i$, where ϕ_i contains ϕ_i functions for: a) structure factors, b) distances, c) planar groups, d) chiral centers, e) non-bonded contacts, f) torsion angles, g) temperature factors, and h) non-crystallographic symmetry. Each piece of information is treated as an observational equation,

$$g\text{-obs} = g\text{-calc}(x) + \epsilon \quad \dots(39)$$

where $g\text{-obs}$ is the observation, $g\text{-calc}(x)$ is the theoretical value computed from the parameters x of the model, and ϵ is the discrepancy between the two [146]. Moreover, as the refinement is based on the principle of least squares, the 'best' set of parameters for equation (39) are those which minimize the weighted sum of squared residuals taken over all the observations,

$$\phi(x) = \sum_h w_h [g_h\text{-obs} - g_h\text{-calc}(x)]^2 \quad \dots(40),$$

in which the appropriate weights w_h are the inverses of the variances of the observations.

The decision was made to further refine the hirudin-thrombin coordinates, and to use the PROLSQ program in particular, for several

reasons. First of all, PROLSQ has been shown over the years to be effective in producing stereochemically-sound structures of biological macromolecules. Moreover, in particular, PROLSQ refinement circumvents two serious shortcomings in the EREF refinement. Temperature factors are not restrained during B-factor refinement in EREF, and in some instances the difference in B-factor between connected atoms in the complex is unreasonably large, as much as 30 Å² or more. Also, the final coordinates of the complex contain 40 ω angles which differ by more than 10° from either side of 180° - one angle as small as 148° - since only moderate energy penalties are placed on such deviations from planarity. As the function minimized by PROLSQ has terms for thermal parameters and planar groups, the aforementioned problems can be averted by making an appropriate choice of target σ values, and therefore weighting factors for equation (40). In addition, upper and lower temperature factor limits were set in the program so that B-factors were not allowed to drop below 15 Å² or rise above 60 Å².

Another important reason for choosing to refine the complex further by means of PROLSQ is the inherent flexibility of this program compared to EREF. By virtue of the different (a-h) functions which comprise the grand function minimized by PROLSQ, as well as the ability to choose the weighting factors for these functions, a refinement protocol can be tailor-made. Through a judicious choice of weighting factors, one can tightly, moderately, or loosely restrain any or all of the (a-h) observational functions; one can also refine atomic occupancies (generally only performed on solvent molecules) in addition to positions and temperature factors, as well as introduce and adjust the extent of shift damping to be applied to parameters during the refinement. The EREF program, on the other hand, affords very little refinement

flexibility; one can simply adjust the crystallographic residual weighting factor k in equation (38), choose positional or B-factor refinement, and control the extent of damping applied to the recommended parameter shifts.

Two new steps were also taken in this second phase of model-building and refinement. The first unit of the carbohydrate moiety in thrombin (Figure 3), N-acetylglucosamine (NAG1), was modeled into a sufficiently large region of 2Fo-Fc electron density near the side chain atom ND2 of Asn60G and included into the refinement (Figure 26). The manner in which water molecules were located and refined in the crystal structure was different. In addition to the criteria used in the EREF refinement and model-building for choosing possible solvent atom positions {good 2Fo-Fc density and significant (generally $> 3.5\sigma$) Fo-Fc density over the resolution range being studied (7 Å to the maximum resolution, 2.5 or 2.3 Å), as well as proximity to a hydrogen-bonding donor or acceptor atom}, it was also required that the position have significant (generally $> 3.5\sigma$) Fo-Fc density over the resolution range 8 Å to the maximum resolution. Refinement of the solvent positions also included alternating block cycles of B-factor and occupancy refinement along with positional refinement.

The interactive model-building, as in the previous work, was carried out on an Evans and Sutherland PS390 graphics display using the PSFR0D0 version of the graphics program FR0D0; however, the 2Fo-Fc and Fo-Fc maps used were not generated using Sim-weighted phases. In addition, the restrained least squares refinement was almost exclusively conducted using a modified version of PROLSQ known as PROFFT, which incorporates fast Fourier transform-based algorithms to speed-up the computation of structure factors and least squares matrix elements, and

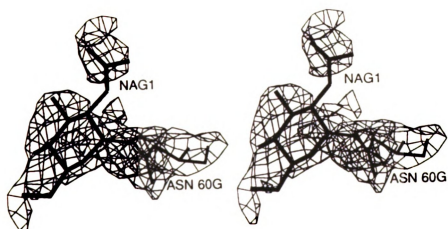


Figure 26. Stereoview of the First Unit of Thrombin Carbohydrate Modeled into Map Density. Displayed in bold is N-acetylglucosamine (NAG1) linked to the side chain atom ND2 of Asn60G. Map density is from the final 2Fo-Fc map used for model-building; contours at 0.8σ . Atomic positions are from the final coordinates.

which can restrain intermolecular contacts [147].

A summary of the model-building and PROLSQ refinement is presented in Table 8. Refinement was initiated using the final coordinates of the EREF-refined hirudin-thrombin structure (including protein atoms and water molecules) in which all the atoms were given an individual B-factor of $35.7 \text{ \AA}^2$, the average individual B-factor of the final EREF structure (Stage -3, Table 8). After two initial rounds of refinement (Stages -2 to -1, Table 8) involving 27 cycles (19 cycles employing tight (T) restraints, 8 employing loose (L) restraints; target σ values used to produce the T and L restraint weights in the refinement are given in Table 9) and a model-building session, however, a new direction had to be taken as the refinement was ill-conditioned. Large coordinate and solvent occupancy shifts were occurring which got progressively larger in size and number, leading to a severe degradation in the stereochemistry and divergence in the refinement. In an effort to avoid these refinement problems, two steps were taken. After difficulties arose in the first round of refinement (Stage -2, Table 8), 857 weak reflections which gave rise to extremely poor R-factors were removed from the reflection file; however, problems persisted in the next stage of refinement (Stage -1, Table 8). As a second attempt to resolve these problems, the refinement was re-started without any of the solvent atoms of the final EREF structure; only the protein and NAG1 carbohydrate atom coordinates from the first graphics rebuilding session (Stage 0, Table 8) were used.

Removal of all the previous EREF-based water molecule oxygen atoms terminated the occurrence of large positional shifts in the protein atomic positions. In the first stage of refinement after this re-start, following only 10 cycles of refinement (8 T and 2 L restraint cycles,

Table 8. Summary of Model-Building and PROLSQ Refinement.

Abbreviations: T, M, L = tight, medium, and loose refinement restraints, respectively (see Table 9); Q = occupancy; NAG1 = N-acetylglucosamine; Bi = individual B (temperature) factor; <> = average; d = d-spacing. Final model parameters given in Table 10.

Stage	R-value	Resolution		Comments
		Maximum (Å)	Total Atoms/Waters	
-3	0.282	2.5	3056/228	EREF structure, all Bi = 35.7 Å ² .
-2	0.237	2.5	3002/174	17 refinement cycles conducted (12 T, 5 L), 4 used (all T). <Bi>= 36.5 Å ² . Refinement ill-conditioned: cannot recover from L cycles; large positional shifts on protein & solvent atoms; low occupancy (Q < 0.4) solvents deleted.
-1	0.224	2.5	3022/228	Model rebuilt on display (1): 14 atom NAG1 sugar unit positioned near Asn60G; 54 new waters added; 48 protein atoms removed. 857 reflections (weak or yielding poor R-factors) removed. 10 refinement cycles conducted (7 T, 3 L). <Bi>= 36.0 Å ² . Refinement ill-conditioned: large positional and occupancy shifts. REMOVE ALL SOLVENTS AND RESTART.
0	0.274	2.5	2795/0	Protein & NAG1 coordinates from model rebuild (1). <Bi>= 35.9 Å ² .
1	0.232	2.5	2795/0	10 cycles of refinement (8 T, 2 L) conducted; <Bi>= 35.5 Å ² .
2	0.202	2.5	2855/76	Model rebuilt on display (2): 76 new waters located; 16 protein atoms removed. 30 refinement cycles (22 T, 8 L) conducted; 16 (14 T, 2 L) used; variable structure factor weighting applied hereafter. <Bi>= 34.5 Å ² .
3	0.199	2.3	2924/132	Model rebuilt on display (3): 56 new waters located; 13 protein atoms added. 35 refinement cycles (29 T, 6 L) conducted; 32 (27 T, 5 L) used. Resolution extension to 2.3 Å. <Bi>= 34.7 Å ² .

Table 8. (cont'd.)

Stage	R-value	Resolution		Total Atoms/Waters	Comments
		Maximum (Å)			
4	0.185	2.3		2987/202	Model rebuilt on display (4): 70 new waters located; 7 protein atoms removed. 47 weak reflections removed, $2.56 \text{ Å} > d > 2.30 \text{ Å}$. 12 refinement cycles (12 T) conducted; 9 used. $\langle Bi \rangle = 35.6 \text{ Å}^2$.
5	0.173	2.3		3048/265	Model rebuilt on display (5): 63 new waters located; 2 protein atoms removed. 32 cycles of refinement (11 T, 21 M) conducted; 17 used (4 T, 13 M). $\langle Bi \rangle = 35.3 \text{ Å}^2$.

Table 9. Summary of Hirudin-Thrombin PROLSQ
Refinement Restraint Parameters.

Target σ values listed for loose, medium, and tight refinement cycles. Weights used in refinement correspond to $1/\sigma^2$. Two values separated by a '/' indicate that the first listed σ value was used in early refinement situations while the second listed value was used later.

	Target σ Values		
	Loose	Medium	Tight
Distances (Å):			
bond lengths	0.030/0.025	0.020	0.015
bond angles	0.050/0.040	0.035	0.030
planar 1-4	0.060/0.050	0.045	0.040
Planes (Å):			
peptides	0.040	0.020	0.015
aromatic groups	0.040	0.020	0.015
Chiral volumes (Å ³):	0.200	0.150	0.150
Non-bonded contacts (Å):			
single torsion	0.55	0.55	0.50
multiple torsion	0.55	0.55	0.50
possible (X..Y)			
hydrogen bond	0.55	0.55	0.50
Isotropic thermal parameters (Å ²):			
main chain bond	0.5	0.5	0.5
main chain angle	1.0	1.0	1.0
side chain bond	0.5/1.0	1.0	0.5/1.0
side chain angle	1.0/1.5	1.5	1.0/1.5

Table 10. Final Model Parameters of the PROLSQ-refined
Hirudin-Thrombin Complex.

The directional arrow and value enclosed in brackets indicate the change with respect to the final hirudin-thrombin EREF structure value.

Protein atoms:

Thrombin:	2322	(↑ 26)
Hirudin:	447	(↑ 36)
Carbohydrate (NAG1):	14	(↑ 14)

Solvent atoms: 265 (↑ 58)

Resolution range (Å): 7.0-2.3

Unique reflections: 21056 (↓ 904)

R-value: 0.173 (↓ 0.020)

Mean B-value (Å²): 35.3 (↓ 0.4)

Geometrical Conformity

r.m.s. deviation in

Bond angles (deg.): 2.8° (↑ 0.4)

r.m.s. deviations from target values:

	target	model
Distances (Å):		
bond lengths	0.020	0.021
bond angles	0.035	0.051
planar 1-4	0.045	0.052

Planes (Å):

peptides	0.020	0.017
aromatic groups	0.020	0.015

Chiral volumes (Å³): 0.150 0.198

Non-bonded contacts (Å):

single torsion	0.55	0.22
multiple torsion	0.55	0.27
possible (X..Y)		
hydrogen bond	0.55	0.27

Isotropic thermal
parameters (Å²):

main chain bond	0.5	0.6
main chain angle	1.0	1.1
side chain bond	1.0	1.2
side chain angle	1.5	1.9

Stage 1, Table 8) the R-factor dropped 0.042 from 0.274 to 0.232, and there was an overall improvement in the stereochemical conformity of the structure. These results clearly imply that at least some of the solvent molecules in the final EREF structure coordinates were the source of the previous ill-conditioned refinement behavior. Moreover, a possible reason why some of the EREF structure water molecules could be causing problems in the PROLSQ refinement is likely linked to the dramatic difference in the functions being minimized by the two programs. The EREF program minimizes a potential energy function along with a crystallographic residual, and likely the energetics of solvation play a strong role in refined water positions. The program PROLSQ, on the other hand, has no direct energy restraints, with refinement being guided primarily by stereochemical restraints. In addition, when one also considers that these crystallographic water molecules have a scattering power of about ten electrons, are generally not fully occupied, and have on an average higher thermal parameters than the protein, it seems reasonable that the EREF and PROLSQ refinement programs would treat such minor contributors to the overall scattering power of the protein crystal differently.

The model-building and refinement progressed very well over the next four stages (Stages 2-5, Table 8). Water molecule oxygen atoms were added and refined in each stage until a total of 265 were included. Also of note is the fact that the resolution of the refinement was extended to 2.3 Å for the last 3 stages of the work, and 47 weak reflections (many yielding poor R-factors) of resolution $2.56 \text{ Å} > d > 2.30 \text{ Å}$ were removed from the reflection file. Finally, a set of refinement restraints intermediate to loose (L) or tight (T), so called 'medium' (M) restraints (Stage 5, Table 8; Table 9), were used in the

last stage of refinement to arrive at the final hirudin-thrombin coordinate set. Overall, in the course of these last four rounds of model-building and refinement (54 T, 13 M, and 7 L restraint cycles used), the R-factor decreased by 0.059 from 0.232 to 0.173, 12 protein atoms were removed, 265 water molecules were added, and the average individual B-factor decreased slightly from 35.5 to 35.3 Å² (Table 8).

An examination of the final model parameters of the PROLSQ-refined hirudin-thrombin complex, along with changes in these values with respect to the final EREF-based complex coordinates (Table 10), reveals that the second phase of model-building and refinement work was clearly worthwhile. The final PROLSQ-based complex coordinates contain 3048 total atoms: 2769 protein atoms (2322 in thrombin, 447 in hirudin), 14 carbohydrate atoms, and 265 water molecules. This atom total represents an increase of 134 over the number of atoms present in the final EREF-based coordinates; moreover, the increase is distributed as follows: 62 more protein atoms (26 in thrombin, 36 in hirudin), 58 more water molecules, and 14 N-acetylglucosamine atoms. Of particular importance is the fact that the R-factor was lowered significantly and the mean B-value decreased as well; the R-factor improved 0.020 from 0.193 to 0.173 and the mean B-factor dropped from 35.7 to 35.3 Å². Also, the final PROLSQ refinement reflection file had 904 fewer reflections than that used in the EREF refinement work; these reflections were removed because they were of low intensity and generally yielded high R-factors.

The geometrical conformity of the PROLSQ-refined hirudin-thrombin structure is also quite good (Table 10). The r.m.s. deviation of bond angles is 2.8°, 0.4° larger than the value for the final EREF complex structure. The r.m.s. deviations on many restrained

parameters are also generally in good agreement with the moderate restraint target values used in the final stages of the refinement (Tables 9, 10). The deviations on bond, angle, and planar 1-4 distances are close to, though generally larger than, the target values, with the deviation on bond angle distances displaying the greatest discrepancy (Table 10). The peptide and aromatic group planarity is excellent, and the r.m.s. deviations on these planar units are below the target values (Table 10). The histogram of ω angles (twist along the C-N peptide bond) in Figure 27 attests to the high degree of peptide bond planarity as better than 91% of the bonds have ω angles of $180^\circ \pm 5^\circ$. In addition, the Ramachandran plot of the ϕ (twist about CA-N axis) and ψ (twist about CA-C axis) angles for non-glycine residues in these coordinates, shown in Figure 28, indicates that the vast majority of the residues have ϕ, ψ angles within conformationally-allowed regions. Only three residues in the complex, which all reside in the thrombin A-chain (Ser1E, Phe7, and Asp14L), deviate significantly from the Ramachandran allowed regions.

The stereoconfiguration at chiral centers is satisfactory; the r.m.s. deviation on chiral volumes is $0.198 \text{ \AA}^3$, which is reasonably close to, but noticeably larger than, the target value of $0.150 \text{ \AA}^3$ (Table 10). On the other hand, the r.m.s. deviations on all non-bonded contacts are outstanding, as all three contact types have deviations which are less than half the target value of $0.55 \text{ \AA}$ (Table 10). The low r.m.s. deviations on non-bonded contacts could in part be due to the fact that an EREF structure was used to initiate the PROLSQ refinement, and that the EREF program, which minimizes a potential energy function containing terms for non-bonded contacts, did well in idealizing these non-bonded interactions. Finally, the deviations on

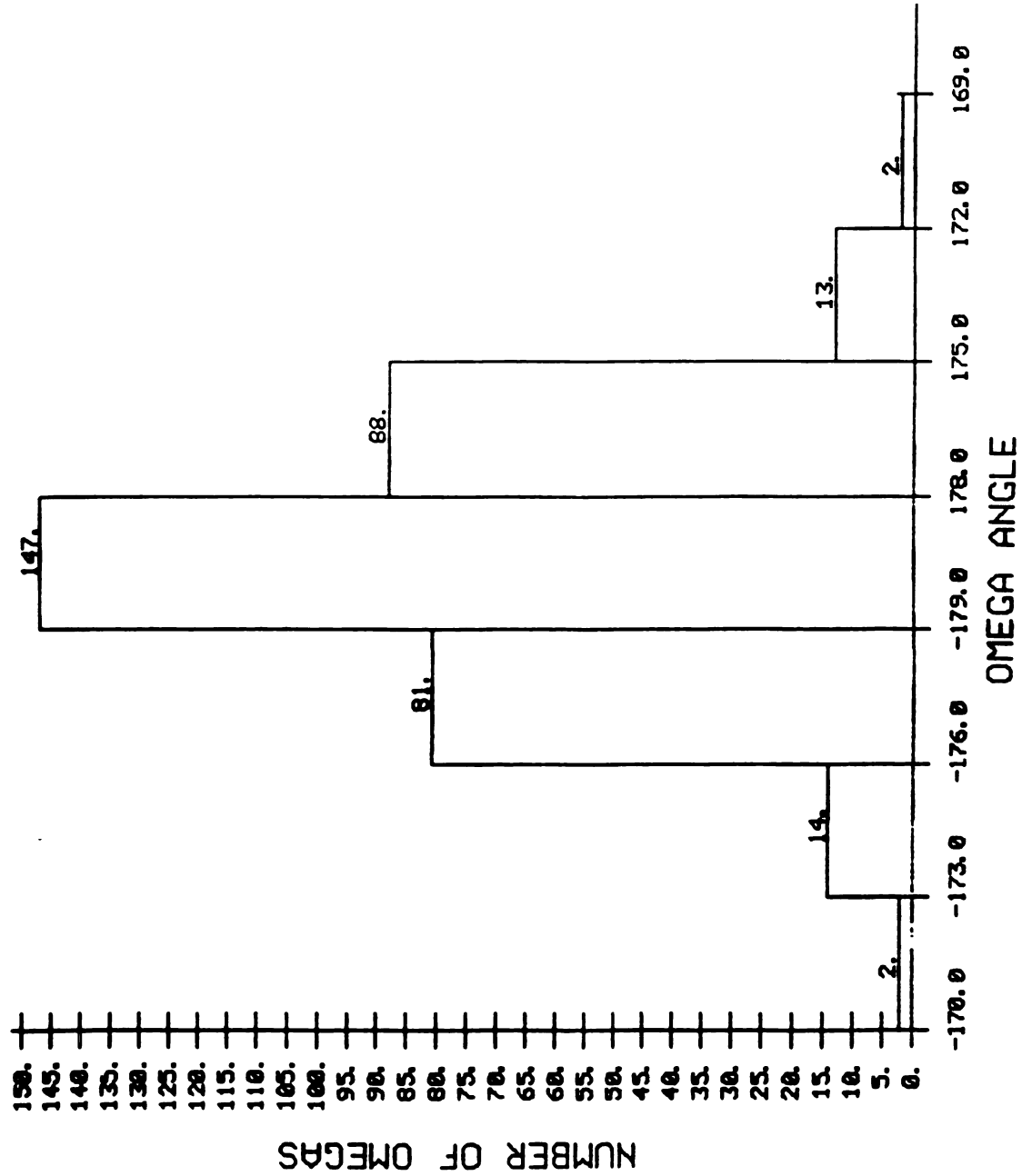


Figure 27. Histogram of ω Angles in the Hirudin-Thrombin Complex.

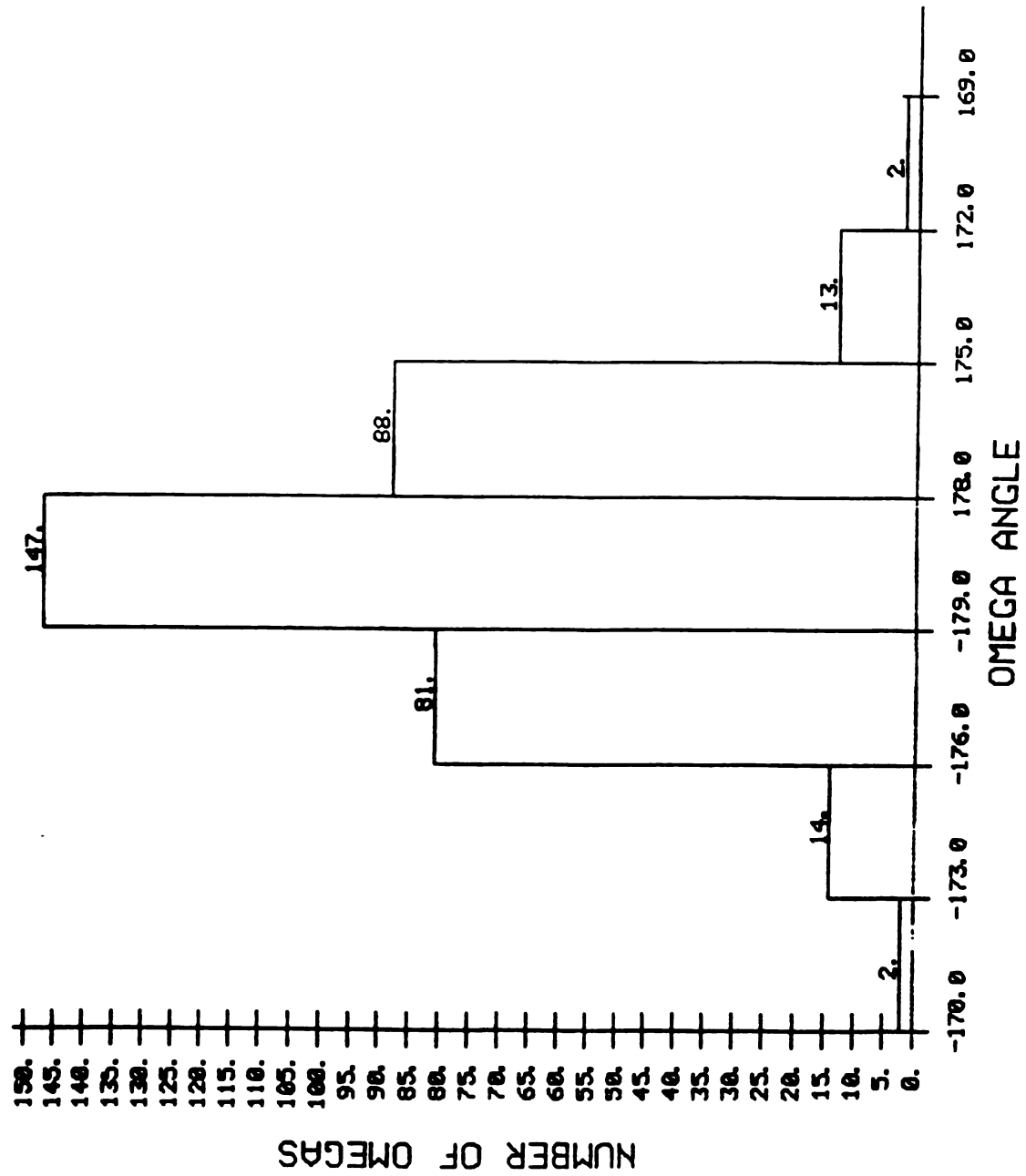


Figure 27. Histogram of ω Angles in the Hirudin-Thrombin Complex.

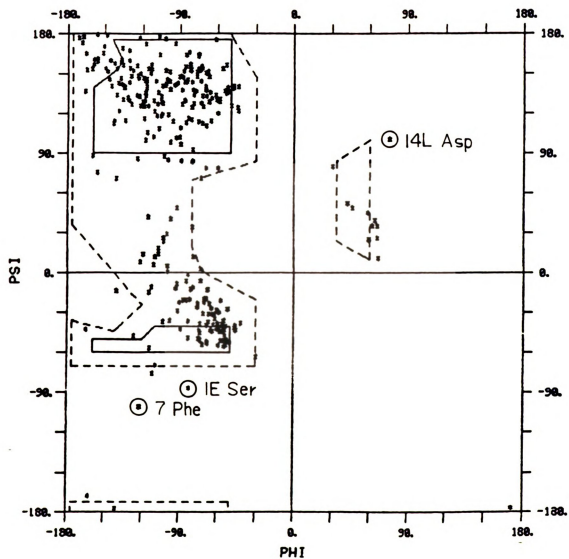


Figure 28. The Ramachandran Plot of Hirudin-Thrombin. Only non-glycine residue angles are displayed.

isotropic thermal parameters among main chain and side chain atoms involved in bonds and angles are all fairly close to the target values, though they are slightly larger (Table 10).

Data pertaining to the R-factor versus resolution shell (seven shells were used, containing approximately 3000 reflections per shell) for the final model parameters of the complex are presented in Table 11 and plotted in Figure 29. The nearly linear, direct correlation between the shell-R and resolution is typical of well-refined structures. Moreover, the fact that the R-factor versus resolution data point curve of the complex falls between the lines corresponding to the Luzzati coordinate error estimates [148] of 0.20 Å and 0.25 Å, suggests that the average error in the atomic positions for the final PROLSQ-refined coordinates is likely close to these estimates (Figure 29). Luzzati plots are based on the assumption that the discrepancies between observed and calculated structure factor amplitudes are due solely to the coordinate error, Δr ; other possible sources of error, such as inaccuracies in the measurement of F_{obs} , are neglected [148]. Thus, the actual error in the coordinates of the complex is probably somewhat larger than 0.25 Å.

While the average individual isotropic temperature factor, $\langle B_i \rangle$, for all the atoms in the hirudin-thrombin complex is 35.3 Å² (Table 10), it is much more informative to consider the $\langle B_i \rangle$ of each component of the crystal and how the $\langle B_i \rangle$ varies within each of these components. Values of $\langle B_i \rangle$ for the various constituents of the hirudin-thrombin crystals are listed in Table 12. Not surprisingly, the $\langle B_i \rangle$ for thrombin (32.1 Å²) is significantly lower than that for hirudin (48.3 Å²), and the thrombin B-chain atoms have a lower $\langle B_i \rangle$ than those of the A-chain (31.5 Å² for the B-chain versus 37.3 Å² for the A-chain). Since

Table 11. The R-factor vs. Resolution for the
Final Hirudin-Thrombin Model Parameters.

The dmin vs. shell R-factor is plotted in Figure 29. The directional arrow and number enclosed in brackets next to shell R-factor values for the the unhydrated structure indicate the change with respect to the corresponding values for the hydrated structure.

dmin (Å)	reflection number	$\langle F_o - F_c \rangle$	sphere R-factor	shell R-factor	shell R-factor no waters	
4.13	3043	89.19	0.150	0.150	0.239	(↑ 0.089)
3.39	3380	63.62	0.143	0.136	0.198	(↑ 0.062)
3.00	3335	55.61	0.151	0.176	0.243	(↑ 0.067)
2.75	3160	46.67	0.157	0.190	0.261	(↑ 0.071)
2.56	3112	42.60	0.163	0.212	0.267	(↑ 0.055)
2.42	2826	39.19	0.168	0.235	0.285	(↑ 0.050)
2.30	2200	38.25	0.173	0.265	0.304	(↑ 0.039)
overall:						
	21056	54.47		0.173	0.242	(↑ 0.069)

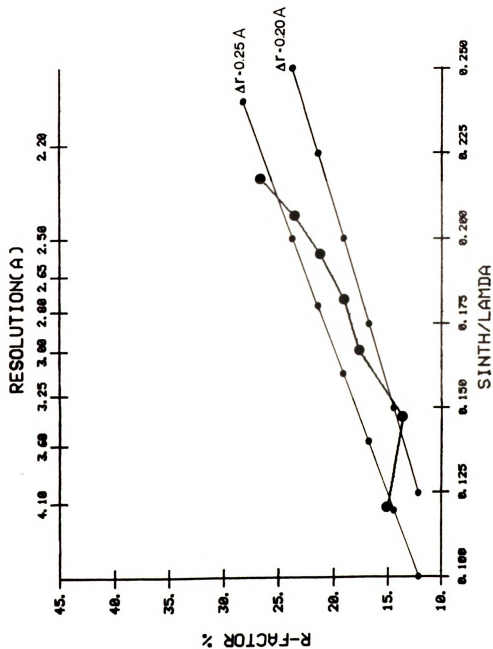


Figure 29. Plot of R-factor vs. Resolution for Hirudin-Thrombin. Large filled circles represent the data points of the complex. The small filled circles represent structure factor amplitude discrepancy data due to specified average coordinate errors (Δr) according to Luzzati [148].

Table 12. Average Individual B-factor and Average Occupancy
Factor Tabulations for the Hirudin-Thrombin Model Parameters.

Stage 'x' waters refer to water molecules located during stage x of the hirudin-thrombin model-building and refinement as outlined in Table 8. (The average occupancy is not given for all the atoms as this value is of no significance.)

Atom Type	Atom Number	$\langle B_i \rangle$ ($\text{\AA}^2$)	$\langle Q \rangle$
All atoms	3048	35.3	----
All protein atoms	2769	34.7	1.00
Thrombin	2322	32.1	1.00
Thrombin A-Chain	253	37.3	1.00
Thrombin B-Chain	2069	31.5	1.00
Hirudin	447	48.3	1.00
All water molecules	265	39.8	0.78
Stage 2 waters	76	31.1	0.91
Stage 3 waters	56	37.8	0.81
Stage 4 waters	70	44.1	0.76
Stage 5 waters	63	47.2	0.63

hirudin is a small inhibitor protein and thrombin is a large serine proteinase with a molecular architecture like chymotrypsin, one would expect atomic vibrations of the atoms in the serine proteinase to be lower than those of the inhibitor. Moreover, since the thrombin A-chain is much smaller than the thrombin B-chain one might expect the B-factors of the covalently linked peptide to be higher, on the average. However, what is startling is the fact that the $\langle Bi \rangle$ for the hirudin atoms is even higher than the average $\langle Bi \rangle$ for all the crystallographically-located water molecules ($39.8 \text{ \AA}^2$) by $8.5 \text{ \AA}^2$.

An examination of the $\langle Bi \rangle$ versus residue number for all of the protein components in the crystal is even more informative. As can be seen in Figure 30, the $\langle Bi \rangle$ varies significantly among the main chain and side chain atoms within each protein unit. The smooth variation in $\langle Bi \rangle$ for main and side chain atoms is likely a result of the isotropic thermal parameter restraints used in the refinement (Table 9). However, the magnitude and variation of $\langle Bi \rangle$ are likely reflective of the conformational flexibility, crystal disorder, and stabilizing interactions in the protein. The $\langle Bi \rangle$ for the side chain atoms is in general larger than for main chain atoms because side chain atoms, by and large, have more conformation flexibility than the main chain atoms. Moreover, in several of the residues where the side chain $\langle Bi \rangle$ is lower than that of the main chain, this occurs because the sidechain atom is a cysteine sulfur involved in a disulfide linkage, which could certainly explain the diminished thermal motion compared to that of the main chain atoms (Figure 30).

Many of the 'high peaks' in the Figure 30 plot correspond to residues (most of which are in loop segments) which are poorly defined in electron density, or are near disordered regions. In the thrombin

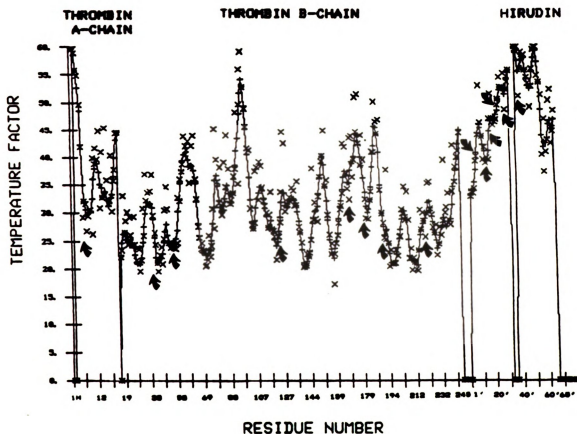


Figure 30. Plot of the Average Individual B-factor of Main and Side Chain Atoms for the Protein Atoms in the Complex. The $\langle Bi \rangle$ for main and side chain atoms is represented by '+' and 'X' symbols, respectively. Main chain atom $\langle Bi \rangle$ values connected by a line. The break in $\langle Bi \rangle$ in hirudin corresponds to the completely undefined segment Ser32'-Lys35'. Bold arrows point to main and side chain $\langle Bi \rangle$ values of Cys residues engaged in disulfide linkages.

molecule, the N- and C-termini of the A-chain and the C-terminus of the B-chain are all missing residues due to disorder in the crystals, and the peak at Glu97A is indicative of the poor density for this residue (weak main chain density and essentially no side chain density). In the hirudin molecule, the residues possessing the largest main and side chain $\langle B_i \rangle$ either adjoin the completely undefined loop residues Ser32'-Lys35' or are near the poorly defined tripeptide segment Asn52'-Asn53'-Gly54'.

Finally, stabilizing interactions in the protein molecules of the complex are probably responsible for some of regions of diminished thermal motion in the Figure 30 plot. The stabilizing or damping effect which the disulfide linkages have on the thermal motion of residues near them is clearly visible, as the cystine linkages generally occur at or near the minima. Also, the low main chain and side chain $\langle B_i \rangle$ values for residues Ile1'-Tyr3' and the trough for residues Asp55'-Gln65' in Figure 30 is probably due to the numerous hirudin-thrombin interactions involving these regions of the hirudin molecule, as will be elaborated on in the following chapter.

All the water molecules in the final PROLSQ-refined coordinates of the complex were located and refined over the course of four stages (Table 8), as was mentioned earlier. Upon examination of the average individual B-factor and average occupancy ($\langle Q \rangle$) for the water molecules added in each of these four steps, an interesting trend was discerned (Table 12). The magnitude of $\langle B_i \rangle$ increases and the magnitude of $\langle Q \rangle$ decreases with each successive round of water molecules added, with values $\langle B_i \rangle = 31.1 \text{ \AA}^2$ and $\langle Q \rangle = 0.91$ for the water molecules added in the first stage of hydration (Stage 2, Table 8) and $\langle B_i \rangle = 47.2 \text{ \AA}^2$ and $\langle Q \rangle = 0.63$ for water molecules added in the last stage (Stage 5, Table

8). This trend seems reasonable if one considers that the difference density peaks selected to be water molecules in the first round of water addition were large in magnitude and significance, thus to satisfactorily account for these peaks a water molecule would generally refine to a low B-factor and a high occupancy factor. However, in the last stage of hydration, the majority of bound solvent atoms have already been located, and frequently many of the significant difference density peaks are likely due to errors. In addition, the largest difference densities are lower in magnitude than they were in the first stage of hydration, and many of the suitable difference density peaks will correspond to second hydration sphere water molecules (more interaction with other water molecules than with protein atoms); thus, the latter water molecules frequently refine to significantly higher B-factors and lower occupancies. The histogram of water individual B-factors in the complex (Figure 31) indicates that although the range in individual temperature factors for water molecules is quite broad, with $15 \text{ \AA}^2 < B_i < 55 \text{ \AA}^2$, nearly half of the crystallographic waters have $40 \text{ \AA}^2 < B_i < 55 \text{ \AA}^2$. The histogram of water occupancies in the complex, Figure 32, shows that majority of the located water molecules are of fairly high occupancy, with nearly 40 water molecules per occupancy shell in each of the five shells between 0.5 and 1.0, and the largest number of waters (55) having the highest possible occupancy, 1.0.

An examination of the R-factor versus resolution data in Table 11 calculated without the water molecules included clearly indicates that the water structure makes a significant contribution to the overall R-factor, and to the the R-factor of each resolution shell. The overall R-factor for the unhydrated complex is 0.242, 0.069 larger than that of the hydrated structure. In addition, while the effect of the water

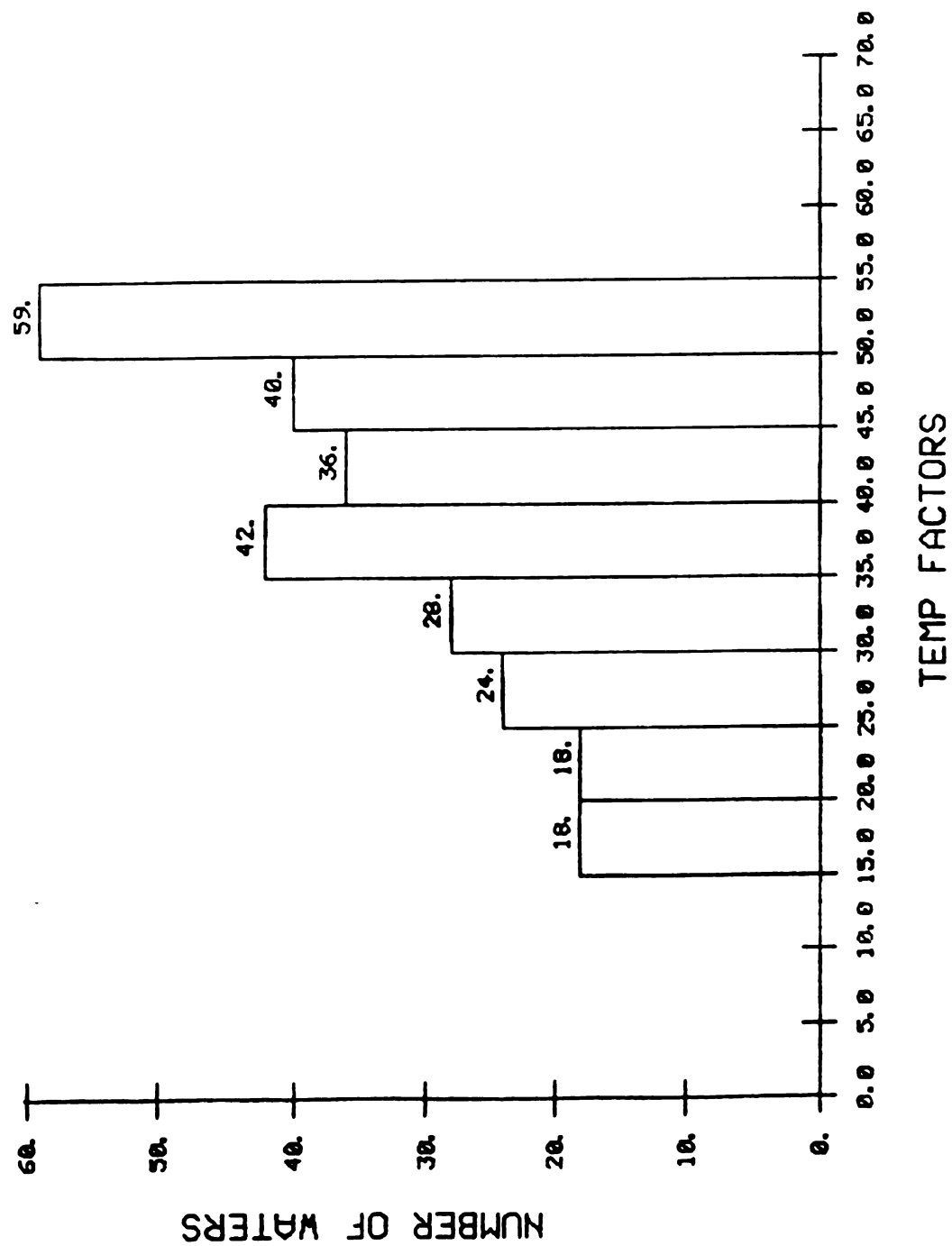


Figure 31. Histogram of the Individual B-factors of the Water Molecules in the Hirudin-Thrombin Structure.

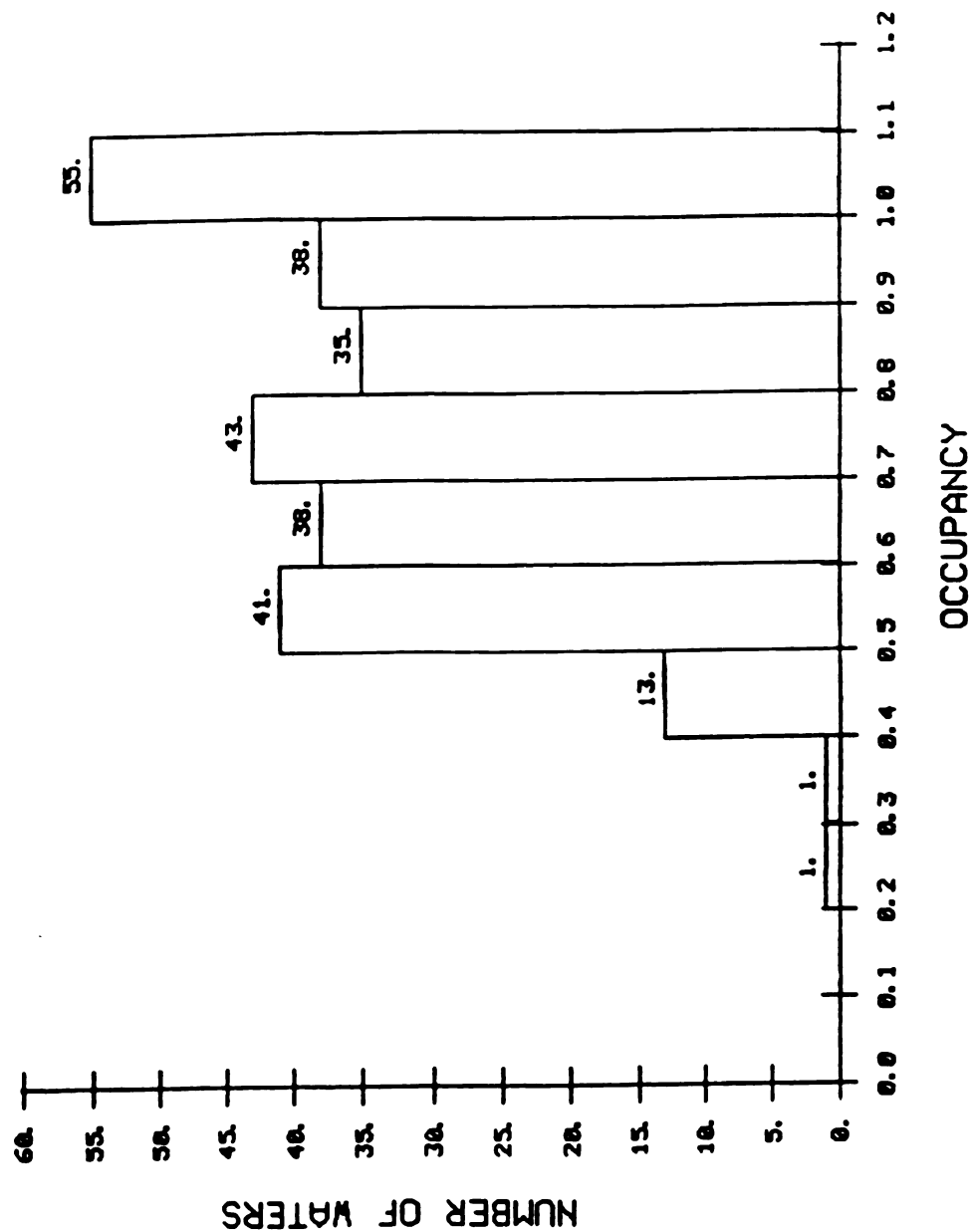


Figure 32. Histogram of the Occupancy Factors of the Water Molecules in the Hirudin-Thrombin Structure. All water molecules in the last occupancy shell, 1.0 - 1.1, have an occupancy of 1.0.

structure is most significant in the lowest resolution shell (0.089 for the 7.00 to 4.13 Å resolution data), where ordered solvent typically makes its largest contribution to x-ray scattering, the water contribution is surprisingly large in all the resolution shells (Table 11). Even in the highest resolution shell (2.42 to 2.30 Å), removal of the water molecules causes the R-factor to increase by 0.039 for the shell. The considerable impact of the water structure at all resolution levels is probably a result of the protein structure re-adjusting in the refinement to accomodate the water molecules optimally.

VII RESULTS AND DISCUSSION

A. The Hirudin Structure

Hirudin appears to possess two distinct domains in the complex (Figure 33). (The hirudin sequence is given in Figure 34.) The NH₂-terminal domain is compact, and contains residues Ile1' to Pro48'. The COOH-terminal domain adopts an extended conformation involving two stretches of polypeptide, residues Glu49'-Gly54' and Asp55'-Pro60', and terminates with a near 3_{10} type III helical reverse turn (Glu61'-Gln65') [149]. The secondary structural elements present in the hirudin molecule are summarized in Table 13.

The globular conformation of the NH₂-terminal domain of hirudin (Figure 33) is intimately related to the presence of the three-disulfide core (Figure 35). The disulfides Cys6'-Cys14' and Cys16'-Cys28' orient nearly perpendicular to one another with the distance between the midpoints of the disulfides being 4.97 Å. In addition, Cys16'-Cys28' and Cys22'-Cys39' are nearly parallel, and the distance between the disulfide bond midpoints is 5.35 Å. Similar close disulfide core interactions are also observed in the kringle structures of prothrombin fragment 1 [150] and plasminogen K4 [151], C3a anaphylatoxin [152], and squash seed trypsin inhibitor [153]. A complete list of the sulfur-sulfur distances in the hirudin molecule is presented in Table 14. It is of particular note that, excluding the actual disulfide bonds themselves, the smallest inter-sulfur distance is 3.74 Å (Cys6' SG-Cys28' SG), the largest distance is only 8.32 Å (Cys14' SG-Cys39' SG), and the mean distance is 5.91 Å $\pm$ 1.30 Å(σ). The result of these close cystine interactions is that the loop segments B, C, and D, which comprise a double loop in hirudin (Figure 34), fold into three unique

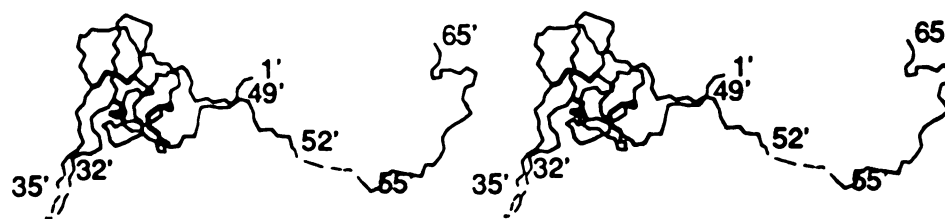


Figure 33. Stereoview of the Folding of Hirudin in the Complex.
N, CA, C atoms only; hirudin disulfides in bold; disordered
or poorly defined regions indicated by dashed lines.

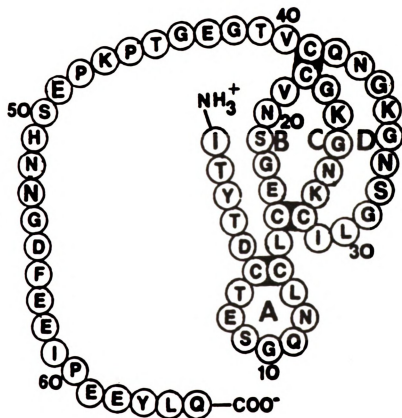


Figure 34. The Sequence of Recombinant Hirudin Variant 2-Lysine 47 (rHV2-K47). Isolated loop designated 'A' and loop segments comprising the two interconnected loops designated 'B', 'C', and 'D'.

Table 13. Secondary Structural Elements of Hirudin.

Element	Type	Residues
β-Structure		
1	Antiparallel	Cys14'-Cys16' Asn20'-Cys22'
2	Antiparallel	Lys27'-Gly31' Gly36'-Val40'
Helix		
H1	Polyproline II	Pro46'-His51'
Reverse Turns		
T1	Type II'	Glu17'-Gly18'-Ser19'-Asn20'
T2	Type II	Gly23'-Lys24'-Gly25'-Asn26'
T3	?	Ser32'-Asn33'-Gly34'-Lys35'
T4	Type III	Glu61'-Glu62'-Tyr63'-Leu64'

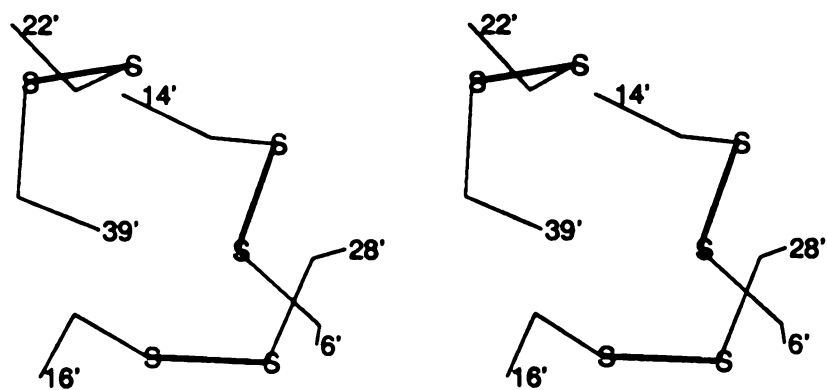


Figure 35. Stereoview of the Cystine Disulfide Core of Hirudin. Disulfide linkages in bold.

Table 14. Sulfur-Sulfur Distances (Å) in Hirudin.
 A '*' indicates a disulfide bond.

Sulfur Atom	6'	14'	16'	22'	28'	39'
6'	0.0	2.01*	4.72	6.11	3.74	7.65
14'		0.0	6.43	6.56	5.43	8.32
16'			0.0	4.70	2.06*	5.22
22'				0.0	5.48	2.05*
28'					0.0	6.58
39'						0.0

loops of protein folding. This, combined with the ordinary Cys6'-Cys14' loop, produces four three-dimensional loops (Figure 36).

The hirudin NH₂-terminal domain appears to be stabilized by antiparallel β -structure. A short segment (β 1, Table 13) results from the interaction of strands Cys14'-Cys16' with Asn20'-Cys22' (Figure 37). The β 1 structure is maintained by the hydrogen bonds Cys22' N-Cys14' O (2.83 Å) and Cys16' N-Asn20' O (3.17 Å). A short hydrogen bond interaction (2.60 Å) between residues Leu13' N and Cys22' O immediately precedes these two strands. The polypeptide strands of this β -structure are connected by a type II' reverse turn (T1, Table 13) [149]. A type II reverse turn (T2, Table 13) [149] is also present in loop segment C involving residues Gly23'-Asn26'. A longer antiparallel β -finger (β 2, Table 13) occurs in loop segment D, and involves residues Lys27'-Gly31' and Gly36'-Val40' (Figure 38). Four hydrogen bonds are apparent in this structural element: Lys27' N-Val40' O (2.39 Å); Val40' N-Lys27' O (2.42 Å); Ile29' N-Gln38' O (2.53 Å); and Gln38' N-Ile29' O (2.72 Å). As with β 1, a reverse turn (T3, Table 13) joins these strands; however, as in solution [78-80], it is completely disordered in the crystal structure of the complex (Figure 33). The structure of the β 2 strand does give some indication that its associated turn will likely be disordered. The strands of peptide appear to diverge just prior to the turn (Figure 38), and this is further evidenced by the fact that what should be a hydrogen bond preceding the turn, Gly31' N-Gly36' O, is a 3.25 Å long contact - too long to be a viable hydrogen bond. The flexibility of this turn could be due to the fact that it is hinged to glycines 31' and 36'.

In addition to the disulfide core, many intramolecular hydrogen bonds also appear to stabilize the N-terminal domain of hirudin in the complex (Table 15). In all, 25 intramolecular hydrogen bonds are found

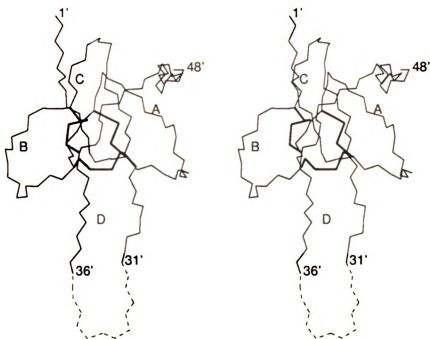


Figure 36. Stereoview of the Folding of NH₂-terminal Domain of Hirudin in the Complex. N, CA, C atoms only; designations 'A', 'B', 'C', and 'D' correspond to those of Figure 34; disorder indicated by dashed lines.

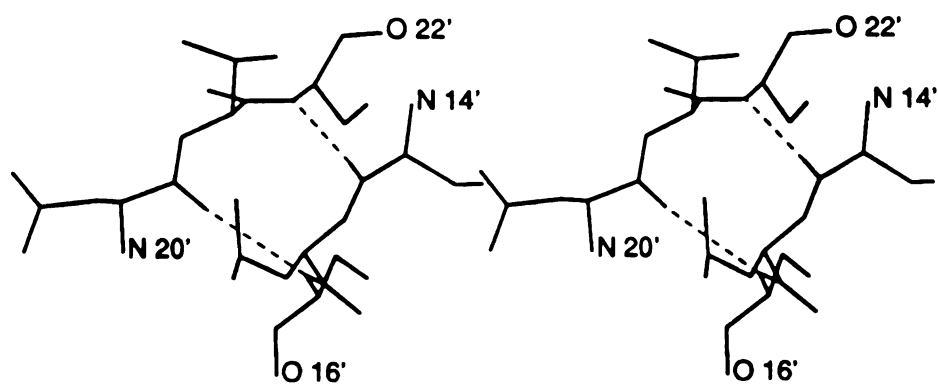


Figure 37. Stereoview of the β_1 Antiparallel β -Strand in Hirudin. Hydrogen bonds indicated by dashed lines.

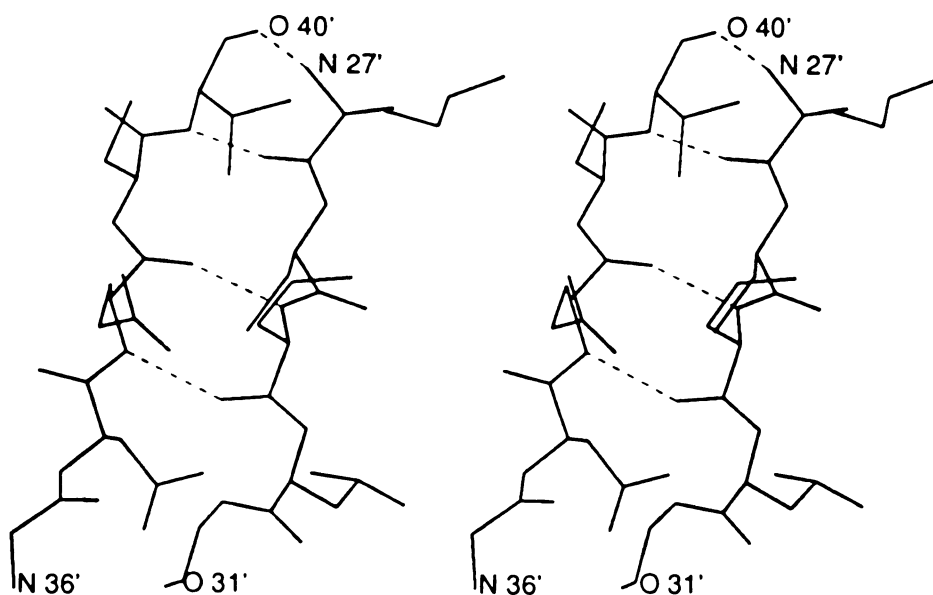


Figure 38. Stereoview of the β_2 Antiparallel β -Strand in Hirudin. Hydrogen bonds indicated by dashed lines.

Table 15. Intramolecular Hydrogen Bonds in Hirudin.
Hydrogen atoms are assigned geometrically idealized positions. Donor atoms are denoted 'D' and acceptor atoms 'A'.

A. Hydrogen Bonds involving Main Chain Atoms.

Element	Donor		Acceptor		Distances (Å)		Angles(deg)		
					H..A	D..A	DHA	CAH	CAD
---	Cys6'	N	Leu15'	O	1.79	2.75	164	145	142
---	Gly10'	N	Cys28'	O	1.44	2.38	159	137	145
---	Asn12'	N	Thr45'	O	1.38	2.24	141	117	130
---	Leu13'	N	Cys22'	O	1.72	2.60	148	166	171
---	Leu15'	N	Thr4'	O	1.85	2.83	161	119	125
β1	Cys16'	N	Asn20'	O	2.17	3.17	162	157	154
T1	Asn20'	N	Glu17'	O	2.63	3.34	130	109	122
β1	Cys22'	N	Cys14'	O	1.89	2.83	149	152	155
T2	Asn26'	N	Gly23'	O	2.46	3.27	137	130	130
β2	Lys27'	N	Val40'	O	1.69	2.39	128	128	146
---	Cys28'	N	Gln11'	O	2.34	3.07	130	125	138
β2	Ile29'	N	Gln38'	O	1.52	2.53	167	156	160
β2	Gln38'	N	Ile29'	O	1.83	2.72	158	140	148
β2	Val40'	N	Lys27'	O	1.53	2.42	147	168	167
---	Gly42'	N	Gly25'	O	2.34	3.07	134	128	120
---	Gly44'	N	Asn26'	O	2.34	2.98	123	138	132
---	Thr45'	N	Gly10'	O	1.91	2.79	145	137	148
---	Lys47'	N	Asn12'	O	1.89	2.70	144	174	170
---	Tyr63'	N	Pro60'	O	1.95	2.90	159	124	128
T4	Leu64'	N	Glu61'	O	2.03	2.82	134	92	106
Mean:					1.93	2.79	146		
Standard Deviation:					0.35	0.31	14		

B. Hydrogen Bonds involving Side Chain Atoms.

Donor		Acceptor		Distances (Å)		Angles(deg)		
				H..A	D..A	DHA	CAH	CAD
Thr7'	N	Gln11'	OE1	1.56	2.48	144	133	137
Glu8'	N	Gln11'	OE1	2.06	3.01	160	83	87
Gln11'	NE2	Glu8'	OE1	1.70	2.63	158	116	114
Leu30'	N	Ser9'	OG	1.71	2.60	150	120	130
Cys39'	N	Glu17'	OE1	1.62	2.57	169	133	129
Lys47'	NZ	Thr4'	OG1	1.69	2.64	149	130	121
Lys47'	NZ	Asp5'	O	1.94	2.90	152	145	150
Mean:				1.75	2.69	155		
Standard Deviation:				0.18	0.19	8		

in the domain, of which 18 involve main chain atoms only (eight reside in turns or β -structure), and seven involve side chain atoms. In particular, while an examination of Figure 36 might lead one to wonder what maintains the conformation of the Val40'-Pro48' polypeptide, impressively nine hydrogen bonds along this nine residue stretch link the chain to the rest of the N-terminal domain.

Hydrogen bonds involving Lys47' also appear to be important in extending the domain to residue Pro48' (Figure 39). The ϵ -amino group of Lys47' is involved in two hydrogen bonds with residues of the amino terminal pentapeptide of hirudin: Thr4'OG1 (2.64 Å) and Asp5'O (2.90 Å) (Figure 39). In addition, Lys47' N hydrogen bonds to Asn12' O (2.70 Å) (Figure 39). Such interactions bring the N- and C-terminals of the domain in relatively close proximity, increasing the size but maintaining the compactness of the domain. The rigidity of the lysine side chain might well be related to the two prolines which flank it, Pro46' and Pro48'. These three residues also initiate the polyproline II helix (H1, Table 13) [154] which extends from Pro46' to His51' (Figure 40); the six residues in this stretch exhibit Ramachandran ϕ , ψ angles similar to those in a collagen helix [155].

A hydrogen bond is a two bond system, D-H...A, consisting of a donor (D-H) and an acceptor (H...A) bond, where D and A represent the electronegative donor and acceptor atoms, respectively. As such, the hydrogen bond contains a large positive (H atom) and a large negative (acceptor) partial charge at close distance. Moreover, since a positive charge assumes its lowest potential energy between two negative charges when all three are aligned, hydrogen bonds are ideally linear. Experimentally, however, it has been shown that the angle decreases from 180° for strong hydrogen bonds to around 130° for weak

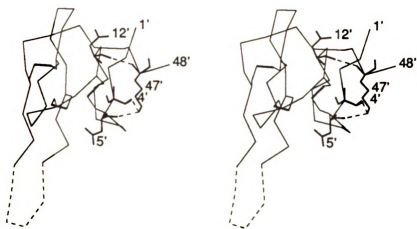


Figure 39. Stereoview of the Intramolecular Lys47' Hydrogen Bonds in the Hirudin N-terminal Domain. CA structure of residues 1'-48' including disulfides (disulfide linkages in bold), and hydrogen bonding residues (bold). The disorder in residues 32'-35' as well as the hydrogen bonds are indicated by dashed lines.

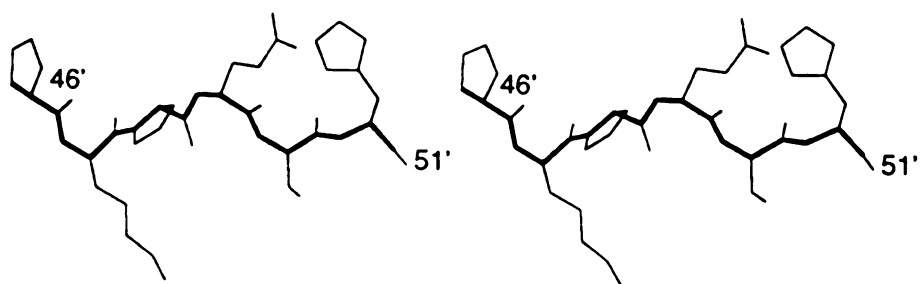


Figure 40. Stereoview of the Polyproline II Helix in Hirudin.
N, CA, C atoms indicated in bold.

ones [156,157]. Hydrogen bond interactions are commonly described in terms of two distances, $r(D..A)$ and $r(H..A)$, and the DHA angle. (Additional angles can also be used to more fully describe the interaction; the CAH and CAD angles can be reported, where 'C' represents the carbon atom to which the acceptor atom A is bonded, but these are not as critical as the DHA angle.)

The mean values of the hydrogen bond descriptors $r(H..A)$, $r(D..A)$, and of the angle DHA were determined for the intramolecular hydrogen bonds in hirudin (25 of the 27 bonds are in the hirudin NH₂-terminal domain), and in general they agree fairly well with mean values determined from organic crystal structures (Table 15). The mean values of the distances H..O and N..O, and of the angle NHO for the 20 intramolecular main chain atom (N, CA, C, O) hydrogen bonds in hirudin were 1.93 Å, 2.79 Å, and 146°, respectively (Table 15). These values are in good agreement with those determined for intramolecular main chain atom hydrogen bonds ($r(H..O) = 1.988$ Å; $r(N..O) = 2.755$ Å; NHO angle = 132.5°) and all (intra- and intermolecular) main chain atom hydrogen bonds in general ($r(H..O) = 1.921$ Å; $r(N..O) = 2.878$ Å; NHO angle = 158.3°) from the literature [158]. Similarly, the corresponding mean values for the intramolecular hirudin hydrogen bonds involving amino acid side chains ($r(H..A) = 1.75$ Å; $r(D..A) = 2.69$ Å; DHA angle = 155°) (Table 15) are in good agreement with those determined from organic crystal structures involving nitrogen donor and oxygen acceptor atoms ($r(H..A) = 1.93$ Å; $r(D..A) = 2.858$ Å; DHA angle = 160.4°) [159]. Finally, while the ideal bond length for general hydrogen bonds involving oxygen and nitrogen-oxygen atoms is considered to be 2.80 Å and 2.85 Å [145], respectively, the 1-4 hydrogen bonds in turns are often accepted to be as long as 3.2 Å [149]. The fact that so

0-
36
r.
st
an
st
:
in
co
st
si
to
st
is
co
Fi
di
en
Cy
ha
th

many of the intramolecular hirudin hydrogen bonds in Table 15 deviate significantly from the ideal values suggests that the diffraction data are not of high enough resolution or quality to resolve them accurately.

The folding of the NH₂-terminal domain of hirudin in the crystalline complex is very similar to that which has been observed for hirudin in solution by 2D NMR studies [78-80]. An optimal superposition of 114 hirudin N, CA, C atoms from the complex (residues 4'-30' and 38'-48') with the average coordinates of 32 NMR structures [79] gave an r.m.s. difference of 0.86 Å, with an r.m.s. difference of 1.95 Å for side chains. The r.m.s. difference of 0.86 Å agrees well with the average atomic r.m.s. differences between the 32 individual NMR structures and with the mean NMR structure, which was approximately 0.7 Å for backbone atoms [79]. The NMR hirudin structures also had disorder in the vicinity of the 32'-35' reverse turn [78-80]. Based on these comparisons, the close correspondence between the NMR and the crystal structures indicates that the hirudin N-terminal domain possesses a similar structure in the free and thrombin-bound states.

Surprisingly, while the main chain atom agreement can be considered to be quite good between the NMR and crystalline complex hirudin structures, the corresponding r.m.s. difference between cystine sulfurs is quite large, 2.43 Å. An optimal superposition of the two disulfide cores based on the N, CA, C fitting mentioned previously is displayed in Figure 41; not surprisingly, the orientation of the disulfides is fairly different between the two structures. The dihedral angles and dihedral energies of these disulfide bridges are presented in Table 16. The Cys6-Cys14 and the Cys22-Cys39 disulfides are displaced approximately a half-bond length and a full bond length in the NMR structure from where they are positioned in the crystal. Moreover, the orientation of the

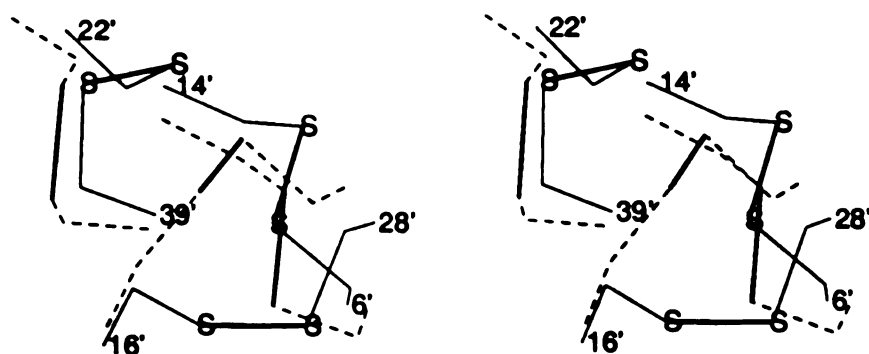


Figure 41. Stereoview of the Disulfides in the Crystal Complex and in the NMR Hirudin Structure [79]. NMR hirudin disulfide side chains indicated by dashed lines; disulfide linkages in bold.

Table 16. Dihedral Angles and Dihedral Energies of the Disulfide Bridges in Hirudin. A '*' indicates the energy value was excluded from the calculation of the mean.

		CA--CB--SG1--SG2--CB--CA						
		X1	X2	X3	X1'	X2'		
Disulfide	Source	X1	X2	X3	X1'	X2'	CA-CA(Å)	E(kcal/mole)
6-14	NMR	-161	-62	87	108	65	4.97	3.32
16-28	NMR	-60	-169	75	-168	-40	6.21	1.92
22-39	NMR	-128	110	165	-83	-121	6.49	15.05*
		Mean Values:					5.89	2.62
							(+/-0.81)	(+/-0.99)
6'-14'	Crystal	-51	-66	97	86	54	4.99	2.07
16'-28'	Crystal	-165	81	85	76	-175	5.94	1.70
22'-39'	Crystal	-175	70	80	65	174	5.55	2.23
		Mean Values:					5.49	2.00
							(+/-0.48)	(+/-0.27)

the Cys16-Cys28 disulfide is different in the two structures; the χ_1 , χ_1' angles and the χ_2 , χ_2' angles differ by approximately 120° and 247° , respectively (Table 16). Available protein structure coordinates indicate that two general disulfide families can be defined based on the chirality of the dihedral χ_3 angle about the sulfur-sulfur bond: right-handed (χ_3 approximately 90°) and left-handed (χ_3 approximately -90°) [160]. Five of the six disulfides presented in Table 16 are right-handed - all of the disulfides from the crystal structure and two of the disulfides from the NMR structure.

The Cys22-Cys39 disulfide of the NMR structure has an unusually large χ_3 value ($\chi_3 = 165^\circ$) which has not been observed in protein structures [160] most likely due to its high dihedral energy. The dihedral energies for the disulfides presented in Table 16 were calculated using a formula from the program AMBER [161]:

$$\begin{aligned}
 E(\text{kcal/mole}) = & 2.0 \cdot (1 + \cos(3 \cdot \chi_1)) + 2.0 \cdot (1 + \cos(3 \cdot \chi_1')) \\
 & + 1.0 \cdot (1 + \cos(3 \cdot \chi_2)) + 1.0 \cdot (1 + \cos(3 \cdot \chi_2')) \\
 & + 3.5 \cdot (1 + \cos(2 \cdot \chi_3)) + 0.6 \cdot (1 + \cos(3 \cdot \chi_3)) \\
 & \dots(41)
 \end{aligned}$$

The dihedral energy for the Cys22-Cys39 disulfide in the NMR structure is 15.05 kcal/mole, drastically larger than the mean value of 3.19 (± 1.43) calculated for right-handed disulfides using available protein crystal structure data [160]. Excluding this outlier value, the mean dihedral energy of 2.62 (± 0.99) kcal/mole for the remaining two disulfides in the NMR structure, and of 2.00 (± 0.27) kcal/mole for the disulfides in the crystal structure agree well with mean value calculated by Katz and Kossiakoff [160]. The mean CA-CA($\text{\AA}$) distance of 5.49 (± 0.48) $\text{\AA}$ (Table 16) observed in the complex also agrees well with the value of 5.07 (± 0.73) $\text{\AA}$ for right-handed disulfides [160]. It is

of note that the mean CA-CA(Å) value of 5.89(+/-0.81) Å for the disulfides in the NMR structure is on the high side due to the fact that two of these distances were long - 6.29 Å (Cys16 CA-Cys28 CA) and 6.49 Å (Cys22 CA-Cys39 CA).

The COOH-terminal domain of hirudin exists as two extended stretches of polypeptide with a bend at Asp55' (Figure 33). The first segment of this domain is approximately 18 Å long and contains residues Glu49' to Gly54'. The first three residues of this segment, Glu49'-Ser50'-His51', complete the polyproline helix displayed in Figure 40. The last three residues of this stretch, Asn52'-Asn53'-Gly54', have very poorly defined electron density in the 2Fo-Fc maps. The second segment of the COOH-terminal domain is approximately 19 Å long and consists of 15 Å in extended conformation (Asp55'-Pro60') followed by a type III 3_{10} helical reverse turn [149]. This reverse turn (T4, Table 13) involves residues Glu61'-Leu64', contains a hydrogen bond between Leu64' N-Glu61' O (2.82 Å) (Table 15), and is shown in Figure 42. A hydrogen bond interaction between Tyr63' N-Pro60' O (2.90 Å) (Table 15) precedes this turn. A general helical structure was also deduced for residues Glu61'-Leu64' by NMR with a synthetic peptide fragment of the hirudin tail [162]. The elongated nature of the C-terminal tail with no stabilizing interactions back to the N-terminal domain probably explains why this portion of the hirudin molecule is disordered in solution.

B. The Hirudin-Thrombin Interaction

The CA structure of the rHV2-K47-thrombin complex is shown in Figure 43, from which it can be seen that the NH2-terminal tail of hirudin binds in the active site region, while the 39 Å long COOH-terminal tail extends across the surface and finishes almost

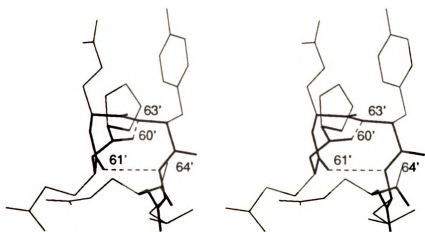


Figure 42. Stereoview of the T4 Type III Helical Reverse Turn in Hirudin. Hydrogen bonds indicated by dashed lines; main chain atoms in bold; Gln65' also included.

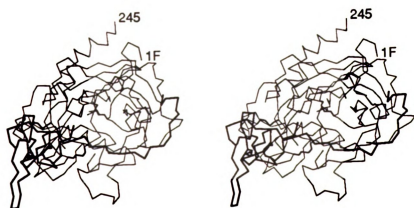


Figure 43. Stereoview of the CA Structure of the Hirudin-Thrombin Complex. Hirudin CA and disulfides of hirudin and thrombin in bold. N- and C-terminals of thrombin designated.

diametrically opposed to the hirudin N-terminus.

A particularly unique and unexpected interaction occurs between the NH₂-terminal tripeptide of hirudin and the the active site region of thrombin from the standpoint of classical serine proteinase inhibition by macromolecular inhibitors. All SERPIN (SERine Proteinase INhibitor)-serine proteinase complexes characterized to date [84,163] display a common mode of interaction in the active site which is also observed in the PPACK-thrombin structure [6] (Figure 44). The polypeptide chain of PPACK aligns antiparallel to the 214-217 strand in thrombin, and the arginyl side chain, which corresponds to the P1 residue of thrombin, sits in the specificity pocket in a substrate-like manner - making an ion pair with residue Asp189 at the base of the pocket. The hirudin NH₂-terminal tripeptide, however, forms a parallel β -strand interaction with the 214-217 thrombin segment and makes no use of the specificity pocket (Figure 45). The hirudin tripeptide forms three hydrogen bonds with the thrombin 214-217 strand (Table 17). Residue Ile 1' engages in two hydrogen bonds (Ile1' N-Ser214 O, 3.02 Å; Ile 1' O-Gly216 N, 3.21 Å) and Tyr3' participates in one (Tyr3' N-Gly216 O, 3.01 Å) (Table 17). In addition, the carbonyl oxygen of Tyr3' makes a hydrogen bond with the amide nitrogen of Gly219 (2.84 Å) (Table 17). Most notable, however, is the fact that the amide nitrogen of Ile1' also interacts through a hydrogen bond with Ser195 OG (3.06 Å) (Table 17).

A consideration of the relative spatial positioning of the hirudin N-terminal tripeptide and PPACK in the thrombin active site reveals both similarities as well as differences. The side chain atoms of Ile1' and Tyr3' occupy basically the same spatial position in the active site region as do Pro and D-Phe of PPACK, respectively. The isoleucyl side chain sits in the S2 subsite of thrombin and makes numerous hydrophobic

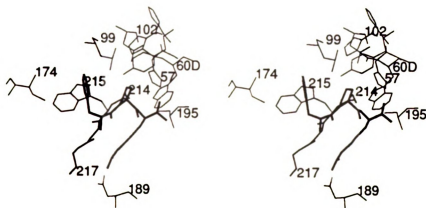


Figure 44. Stereoview of PPACK in the Active Site Region of Thrombin. PPACK, from the PPACK-thrombin structure [6], optimally superimposed into thrombin active site based on N, CA, C fitting of PPACK-thrombin and hirudin-thrombin complex thrombin structures; PPACK in bold.

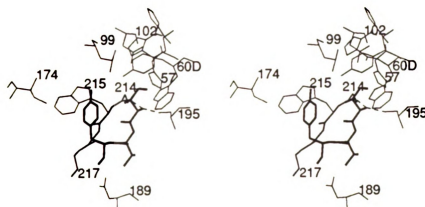


Figure 45. Stereoview of the Interaction of the Hirudin N-terminal Tripeptide in the Active Site of Thrombin. Hirudin N-terminal tripeptide in bold; hydrogen bonds of the N-terminal nitrogen indicated by dashed lines.

Table 17. Intermolecular Hydrogen Bonds in the Hirudin-Thrombin Complex. Hydrogen atoms are assigned geometrically idealized positions. Donor atoms are denoted 'D' and acceptor atoms 'A'. A protonated amino terminus is designated 'N+' and an unprotonated amino terminus is designated 'N' for Ile1'. A '*' indicates the mean and standard deviation were calculated using values for an unprotonated Ile1' amino terminus.

A. Hydrogen Bonds involving Main Chain Atoms.

Donor			Acceptor		Distances (Å)		Angles(deg)		
					H..A	D..A	DHA	CAH	CAD
Ile1'	N+	Ser214	O		2.05	3.01	154	144	152
Ile1'	N	Ser214	O		2.04	3.01	155	149	152
Tyr3'	N	Gly216	O		2.04	3.01	165	161	159
Glu57'	N	Thr74	O		1.84	2.82	173	165	164
Gly216	N	Ile1'	O		2.32	3.21	149	153	158
Gly219	N	Tyr3'	O		2.17	2.84	123	138	155
Lys224	NZ	Ser19'	O		2.17	3.10	149	150	150
Mean*:					2.10	3.00	152		
Standard Deviation*:					0.16	0.15	17		

B. Hydrogen Bonds involving Side Chain Atoms.

Ile1'	N+	Ser195	OG		2.45	3.06	117	126	132
Ile1'	N	Ser195	OG		2.26	3.06	132	121	132
Val21'	N	Glu217	OE1		2.10	3.05	163	133	128
Lys60F	NZ	Glu49'	OE1		2.16	3.20	176	94	94
Arg73	NH2	Asp55'	OD1		1.99	2.92	142	101	110
Arg73	NH1	Asp55'	OD2		2.32	3.22	153	123	116
Tyr76	N	Glu57'	OE2		1.89	2.89	169	130	129
Mean*:					2.12	3.06	156		
Standard Deviation*:					0.16	0.14	17		

contacts with the thrombin residues His57, Tyr60A, Trp60D, Leu99, and Trp215. The aromatic moiety of Tyr3' is in the hydrophobic cleft defined by Leu99, Ile174 and Trp215. This hydrophobic region is likely the apolar binding site for indole reported by Berliner and Shen [40], elaborated upon by others [41,47,58], and delineated by Bode et al., [6] in the PPACK-thrombin structure. The nonpolar contacts of the Tyr3' side chain with thrombin are also complemented by those involving nearby residues in the hirudin molecule; in all, this tyrosyl group is involved in eight van der Waals contacts of less than 4.0 Å with the side chains of Ile1', Leu15', and Val 21'. The side chain of Thr2' only penetrates the edge of the specificity pocket and well below the point of entry of the arginine of PPACK. The fact that hirudin interacts in such a novel fashion in the thrombin active site coupled with the fact that it displays no close sequence or topological similarity to the ten existing classes of serine proteinase inhibitors [73] suggests that hirudin truly represents a heretofore unknown family of inhibitors [72].

It is likely that all naturally occurring hirudins interact in the active site as does rHV2-K47, since they all possess Ile-Thr-Tyr or Val-Val-Tyr NH₂-terminal sequences [61]. Furthermore, the importance of maintaining nonpolar character in the first two positions has been demonstrated experimentally in binding studies with specific 1-2 position site-directed mutants [90]; these experiments showed that replacing N-terminal Val-Val residues with polar amino acids resulted in an increase in the inhibition constant (K_i), while replacement of these residues with aromatic or hydrophobic amino acids had little effect on the value of K_i . This is consistent with the nonpolar character of the S2 subsite.

More difficult to reconcile with the crystal structure is

experimental evidence obtained by Stone, Braun, and Hofsteenge which supports the importance of having a positive charge on the hirudin amino-terminus [81]. In Table 17, appropriate hydrogen bond distances and angles are presented for the two amino-terminal Ile1' N hydrogen bonds to Ser214 O and Ser195 OG assuming both a protonated and an unprotonated hirudin N-terminus. The significantly improved H..A distance (2.26 Å vs. 2.46 Å) and DHA angle (132° vs. 117°) for the Ile1' N-Ser195 OG hydrogen bond with an unprotonated amino terminal nitrogen supports the notion that Ile1' N is uncharged. Since His57 of the catalytic site is also protonated at pH 4.7, a neutral hirudin amino-terminus could alleviate a positive charge build-up in the active site region. Moreover, a protonated Ile1' nitrogen would be only 3.23 Å from a protonated His57 NE2 in the crystal structure.

Although most of the NH₂-terminal domain of hirudin is not in contact with the thrombin surface (Figure 43), many hydrophobic and polar interactions exist at the interface of the two. For instance, three hirudin residues participate in nonpolar interactions at this interface (Figure 46). Residues Leu13' and Pro46' form hydrophobic contacts of 3.34 Å and 3.73 Å, respectively, with Pro60C of the nine residue 60A-60I insertion loop which narrows the active site cleft [6]. In addition, Val21' has two van der Waals contacts of less than 4.0 Å with Ile174. Two ion pair and two hydrogen bond interactions are also found at the interface (Figure 47). The salt bridges are: Asp5'-Arg221A (3.6 Å), and Glu17'-Arg173 (5.0 Å). Hydrogen bonds occur between residues Lys224 NZ-Ser19' OG (3.10 Å) and Val21' N-Glu217 OE1 (3.05 Å) (Table 17). In all, of the 48 amino acids in hirudin N-terminal domain, 15 residues are directly involved in contacts of less

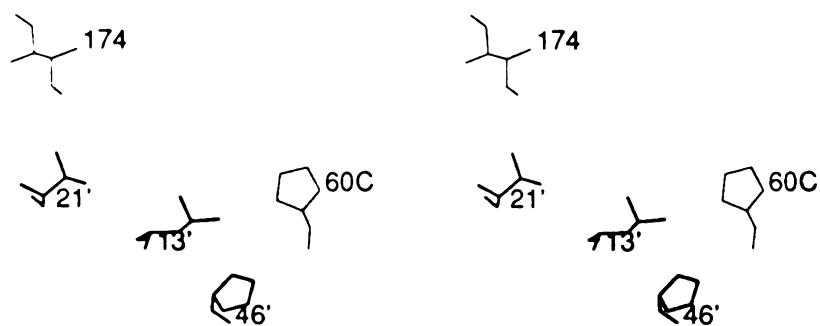


Figure 46. Stereoview of Hydrophobic Contacts of the N-terminal Hirudin Domain with Thrombin. Hirudin residues in bold. Residues in figure: Ile174, Pro60C, Leu13', Val21', and Pro46'.

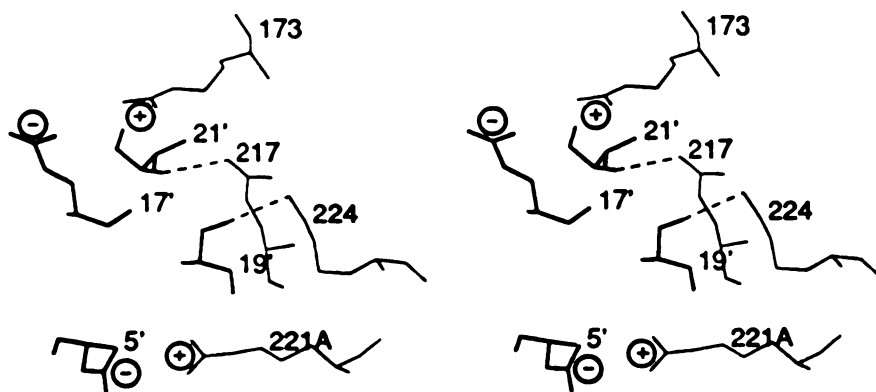


Figure 47. Stereoview of Electrostatic Interactions of the N-terminal Hirudin Domain with Thrombin. Hirudin residues in bold. Hydrogen bonds indicated by dashed lines, and ion pairs by +/-.

than 4.0 Å with thrombin.

A number of water-mediated hirudin-thrombin and hirudin-hirudin hydrogen bond interactions involve residues in the N-terminal hirudin domain (Tables 18, 19, and 20). The carbonyl oxygen of Gly18', side chain oxygens of Asp5' and Asn20', and the ring hydroxyl group of Tyr3' all interact with thrombin residues via water molecules (Tables 18 and 19). In particular, while Asp5' and Arg221A appear to form a strong salt-bridge directly, a nearly pyramidal-coordinate water molecule in the vicinity, W622, appears to bridge this contact as well as bond to the Gly18' O (Figure 48). The result is that the water molecule makes three hydrogen bonds with the three residues involving Asp5' OD2 (2.56 Å), Gly18' O (2.87 Å), and Arg221A NH2 (2.65 Å). The remaining hydrogen-bonding interactions with thrombin through water molecules are Tyr3'-Tyr60A and Asn20'-Arg173. There are also several long range ($d > 3.2$ Å) water-mediated hirudin N-terminal domain-thrombin interactions which involve the following residues: Leu13'-Trp60D; Asn20' with Thr172-Arg173, Glu217, Lys224; and Val21'-Arg173. Lastly, there are four solvent-bridged hydrogen bond interactions between residues within the domain: Thr4'-Leu13'; Asp5'-Gly18'; Leu13'-Lys47'; and Glu17'-Asn20' (Table 20).

The tripeptide segment Pro46'-Lys47'-Pro48' appears to be important in facilitating the binding of the hirudin N-terminus in the active site region of thrombin (Figure 49). However, it is not because this stretch of residues resembles a thrombin substrate-like sequence [164] or because the Cys39'-Pro48' segment bears 50% sequence homology to residues Arg148-Ser157 of the thrombin cleavage site on prothrombin [86]. The hydrogen bonds made by the ϵ -amino group of Lys47' to Thr4' OG and Asp5' O, as well as the solvent-bridged interaction with Arg221A,

Table 18. Hirudin-Water Molecule Hydrogen Bonds in the Complex. Hydrogen atoms are assigned geometrically idealized positions. Donor atoms are denoted 'D', acceptor atoms 'A'.

Donor	Acceptor	Distances (Å)		Angles(deg)	
		H..O	D..A	DHA	COH
Thr2'	N	W498	2.13	3.05	173
Thr2'	OG	W410	1.94	2.85	163
Tyr3'	OH	W606	2.15	3.00	152
Thr4'	N	W469	1.82	2.72	147
Thr4'	OG	W612	2.13	3.10	173
Asp5'	N	W534	2.07	3.05	168
Glu17'	N	W502	2.04	2.99	168
Lys24'	N	W673	2.14	3.08	162
Gly25'	N	W493	1.79	2.70	163
Glu58'	N	W722	1.95	2.99	160
Glu62'	N	W472	2.00	2.86	146
Tyr63'	OH	W657	1.83	2.71	154
W401	Glu57' O		2.36		133
W467	Glu58' OE2		2.75		121
W469	Leu13' O		2.78		133
W472	Glu62' OE2		2.86		101
W483	Thr2' OG1		2.40		121
W502	Asn20' O		2.85		126
W508	Glu49' OE1		3.17		104
W541	Leu15' O		3.08		135
W543	Asp55' O		2.56		150
W573	Asn20' OD1		2.33		130
W622	Asp5' OD2		2.56		121
W622	Gly18' O		2.88		111
W623	His51' O		3.03		89
W672	Leu13' O		3.07		119
W672	Lys47' O		2.56		139
W676	Glu58' OE1		2.30		144
W726	Cys6' O		2.75		157

Table 19. Solvent-Bridged Hirudin-Thrombin Hydrogen Bond Interactions. The protein atoms involved in the interaction are represented as 'A' and 'B', and the mediating water 'W'.

A	Water Molecule	B	Distances (Å)	
			A..W	W..B
Tyr3' OH	W606	Tyr60A OH	3.00	2.38
Asp5' OD	W622	Arg221A NH2	2.56	2.66
Gly18' O	W622	Arg221A NH2	2.88	2.66
Asn20' OD1	W573	Arg173 N	2.33	2.84
Glu49' OE1	W508	Arg35 NH1	3.17	2.93
Glu57' O	W401	Arg67 NH1	2.36	3.00
Glu57' O	W401	Arg67 NH2	2.36	2.94
Glu58' OE2	W467	Arg77A NE	2.75	2.94

Table 20. Solvent-Bridged Intramolecular Hydrogen Bond Interactions in Hirudin. The protein atoms involved in the interaction are represented as 'A' and 'B', and the mediating water 'W'.

A	Water Molecule	B	Distances (Å)	
			A..W	W..B
Thr4' N	W469	Leu13' O	2.72	2.78
Asp5' OD2	W622	Gly18' O	2.56	2.88
Leu13' O	W672	Lys47' O	3.07	2.56
Glu17' N	W502	Asn20' O	2.99	2.85
Glu62' N	W472	Glu62' OE2	2.86	2.86

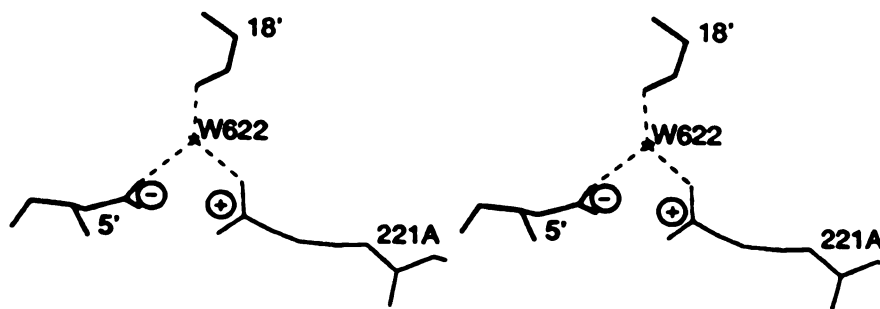


Figure 48. Stereoview of a Hirudin-Thrombin Water Mediated Interaction. Hirudin residues in bold. Hydrogen bonds indicated by dashed lines, ion pairs by +/-, the water molecule by *.

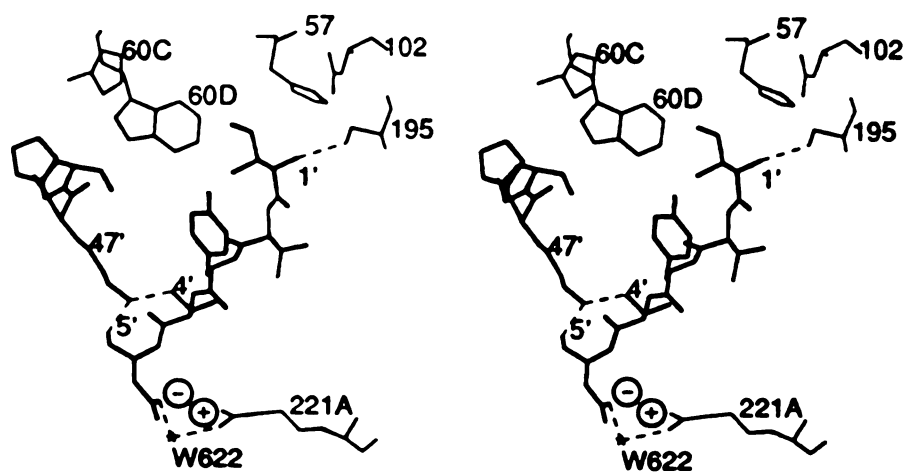


Figure 49. Stereoview of Hirudin Pro46'-Lys47'-Pro48' and the Hirudin N-Terminus-Thrombin Active Site Interaction. Hirudin residues in bold. Hydrogen bonds indicated by dashed lines, ion pairs by +/-, and water molecules by *.

could help stabilize the active site binding by "clamping down" on the 4'-5' dipeptide. Moreover, the van der Waals contact between Pro46' and Pro60C, and the polar contacts between Lys47' and Trp60D could help anchor the N-terminal domain in the proper position for the amino-terminal tripeptide to associate intimately with active site region residues. It is of note that Lys47' has often been viewed as the P1 residue of substrate in the hirudin-thrombin complex [67,70,87]. Though the lysyl side chain of 47' does not occupy the specificity pocket of thrombin and is located approximately 11 Å away (Figure 49), it appears to be a factor in the hirudin-active site region binding.

The COOH-terminal tail in hirudin, in contrast to the NH₂-terminal head, adopts an unusual, extended conformation in the complex (Figure 43). In addition, by means of this extended conformation, residues of the tail can interact intimately with a wealth of residues on the thrombin surface; in all, the side chains of 12 of 17 hirudin C-terminal domain residues are involved in electrostatic or nonpolar contacts with thrombin. The region of thrombin where these residues bind has become known as the anion-binding exosite [47]. In addition, this region, which is defined by the loop segments Phe34-Ser41 and Lys70-Glu80 [6], is particularly rich in charged side chains - especially positively-charged arginyl and lysyl side chains [165,166] (Figure 50). Another extensive and highly electropositive exosite appears to exist on the opposite side of the thrombin molecule and consists of Lys87, Arg93, Arg97, Lys169, Arg175, His230, Arg233, Lys236, and Lys240 (Figure 51); this might be the site of heparin binding since the latter binds at an exosite different from that of fibrinogen [45].

In the first segment of the COOH-terminal hirudin tail, the initial three residues, Glu49'-Ser50'-His51', are involved primarily in

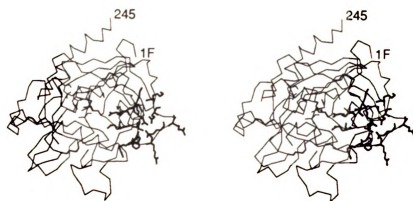


Figure 50. Charged Residues of the Anion-Binding Exosite Region of Thrombin. Thrombin CA structure only with thrombin disulfides and charged anion-binding exosite residues in bold.

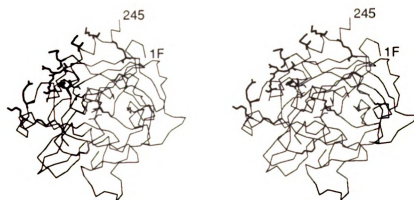


Figure 51. Basic Residues of the Postulated Heparin Binding Site of Thrombin. Thrombin CA structure only with thrombin disulfides and basic residues of heparin binding site in bold.

electrostatic contacts with thrombin (Figure 52). Residue Glu49' forms a hydrogen-bonding ion pair with Lys60F (3.20 Å) (Table 17) and interacts with Arg35 via the water molecule W508 (Tables 18, 19). The OG of Ser50' appears to be involved in close, bifurcated polar contacts of 3.36 Å and 3.68 Å with the side chain oxygen atoms of Glu192. Also, the imidazolium ion of His51' is involved in an ion pair with Glu39 of thrombin (His51' ND1-Glu39 OE2, 3.0 Å). The final three residues of this C-terminal tail segment, Asn52'-Asn53'-Gly54', are poorly defined in the electron density maps.

The second segment of the hirudin COOH-terminal tail, Asp55'-Gln65', can be viewed as being composed of a 15 Å extended stretch of polypeptide from Asp55'-Pro60' followed by a type III 3_{10} helical reverse turn from Glu61'-Leu64'. Impressively, nine of the 11 residues in this hirudin tail region interact with residues on the thrombin surface.

Four of the last eleven residues in the hirudin C-terminal tail participate in electrostatic contacts with thrombin (Figure 53). The carboxylate of Asp55' is involved in a double hydrogen-bonding salt bridge with Arg73 (Arg73 NH2-Asp55' OD1, 2.93 Å; Arg73 NH1-Asp55' OD2, 3.22 Å) (Table 17) and forms an ion pair with Lys149E of 4.9 Å. The Asp55'-Lys149E salt bridge is in agreement with a successful lysine cross-linking experiment carried out using a synthetic dinitrofluorobenzyl-Gly54-Leu64 peptide [167]. In addition, Lys149E is the position of the γ -cleavage site in thrombin [48,82] (Figure 10). The residue Glu57' is engaged in hydrogen bonds with main chain atoms of residues Thr74 (Glu57' N-Thr74 O, 2.93 Å) and Tyr76 (Tyr76 N-Glu57' OE2, 2.89 Å) (Table 17), and has 10 contacts of less than 4.0 Å with main chain and side chain carbon atoms of Arg75. The carbonyl oxygen of

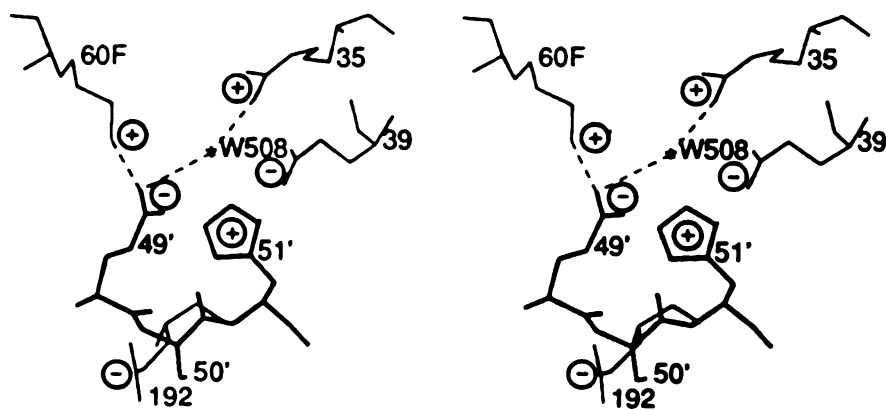


Figure 52. Hirudin(Glu49'-Ser50'-His51')-Thrombin Interactions. Hirudin residues in bold. Hydrogen bonds indicated by dashed lines, ion pairs by +/-, and water molecules by *.

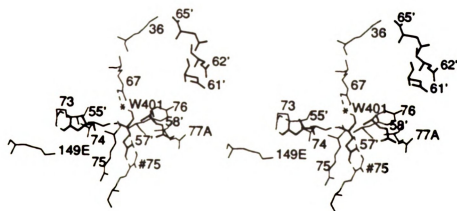


Figure 53. Electrostatic Hirudin(Asp55'-Gln65')-Thrombin Interactions. Hirudin residues in bold. Hydrogen bonds indicated by dashed lines, ion pairs by +/-, and water molecules by *.

Glu57' also interacts with the Arg67 side chain atoms NH1 and NH2 through the water molecule W401 (Table 19). However, the most interesting aspect of this residue is that it participates in a double hydrogen-bonding ion pair with Arg#75 of a symmetry-related molecule in the crystal (Arg#75 NE-Glu57' OE1, 2.81 Å; Arg#75 NH2-Glu57' OE2, 2.93 Å); the proximity of Arg75 to Glu57' suggests that this salt-bridge could likely occur within the complex in solution. The Glu57'-Glu58' residues are also involved in an electrostatic interaction with Arg77A, which is the position of the β -cleavage site in thrombin [48,52]. The side chain of Glu58' forms an ion pair with Arg77A directly while Glu57' can interact with the arginyl group through the mediating water W467, although the distance is long for a hydrogen bond (Glu57'OE2-W467, 3.72 Å). Finally, the C-terminal carboxylate of Gln65' forms an ion pair of 3.33 Å with the ϵ -amino nitrogen of Lys36.

Glutamate residues 61' and 62' have no interactions with thrombin; this was somewhat surprising as site-directed mutagenesis studies have shown that single and multiple mutations of the carboxy-terminal glutamate residues 57', 58', 61', 62' to glutamine caused increases in the hirudin-thrombin dissociation constant [87]. The water molecule W472 appears to bridge an intramolecular and intraresidue hirudin hydrogen bond interaction between Glu62' OE2 and Glu62' N (Table 20).

An unexpected number of van der Waals contacts also occur between the latter half of the extended C-terminal tail and the anion-binding exosite. The residue Phe56' penetrates into a depression on the thrombin surface and makes 11 close contacts with thrombin residues: one with the sulfur of Met32, three with Phe34, one with Leu40, and seven with Thr74 (Figure 54). Most notably, as the planes of Phe56' and Phe34 are nearly perpendicular to one another, edge-on aromatic stacking

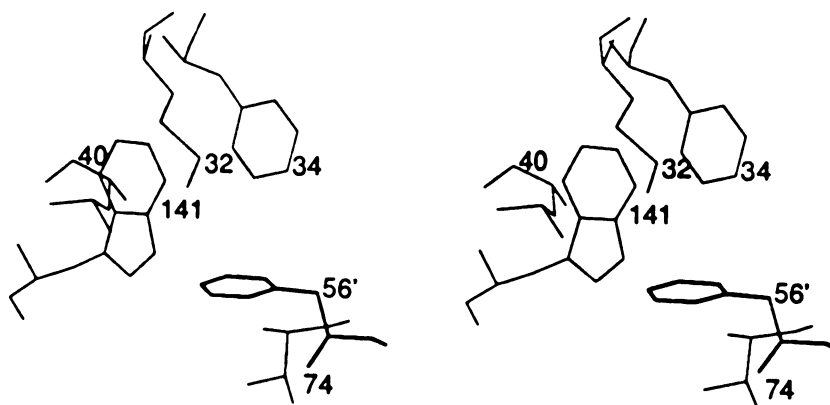


Figure 54. Hirudin Phe56' in a Hydrophobic Thrombin Cavity.
Phe56' in bold.

occurs between the two phenyl groups (Figure 54). There is also a substantial concentration of nonpolar hirudin-thrombin interactions in the vicinity of the type III helical reverse turn (Figure 55). Hirudin residue Ile59' has three close intramolecular contacts with the side chain of Leu64', as well as one contact each with the side chain atoms of Leu65 and Ile82. Residue Pro60' is involved in five intermolecular contacts of less than 4.0 Å with Tyr76 and has one close contact with Ile82. The residue Tyr63', which possesses an acyl-sulfate in natural hirudin, participates in a sum total of five close hydrophobic approaches with Leu65 and Ile82. As a result of the type III 3_{10} helical reverse turn, Leu 64' makes two close contacts each with Ile59' and main chain atoms of Glu61', as well as one close contact with the side chain of Leu65 of thrombin. The interactions in the vicinity of the turn terminate with the Gln65'-Lys36 ion pair.

As the lack of a sulfated tyrosine at position 63' in hirudin has been shown to lead to a six-fold increase in the dissociation constant of the hirudin-thrombin complex with respect to natural Tyr63'-sulfated hirudin [86], it is of some importance to consider residues in the vicinity of a sulfated Tyr63' which could conceivably be involved in a stabilizing interaction. Initially, it seemed reasonable that the nearby thrombin residues Lys81 and Lys109-Lys110 could form ion pairs with a sulfated residue via free bond rotation. In the crystal complex, the ϵ -amino nitrogen of Lys81 forms a hydrogen bond with the carbonyl oxygen of Ala113. The terminal amino groups on the side chains of Lys109 and Lys110, on the other hand, are not engaged in stabilizing interactions - making these two lysyl side chains more likely participants in an ion pair with a sulfated-Tyr63'. However, the

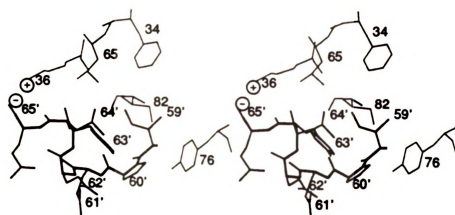


Figure 55. Hirudin-Thrombin Interactions near the Type III Helical Reverse Turn of Hirudin. Hirudin residues in bold, with N, CA, C of hirudin bolder. Ion pairs indicated by +/-.

structure of the hirugen-thrombin complex at 2.2 Å resolution [168] (hirugen is Asn54'-Leu64' with a sulfated Tyr63') reveals that the oxygen atoms of the Tyr63' sulfate make hydrogen bonds to Tyr76 OH and Ile82 N, and that no definitive ion pairs occur. Moreover, the spatial similarity of Tyr63' and Tyr76 in the hirudin-thrombin structure to the corresponding tyrosine rings in the hirugen-thrombin structure suggests that this interaction is likely responsible for the additional binding affinity observed in a sulfated hirudin-thrombin complex.

The aforementioned electrostatic interactions certainly support experimental evidence that acidic residues of the hirudin C-terminal tail [70] and basic residues of thrombin [18,165-166] are important to hirudin binding. Recently, Chang has identified inaccessible lysyl residues in hirudin-blocked thrombin (lysines 36, 60F, 70, 109, 110, 149E) and additionally implied the participation of the arginine residues 67, 73, 75, and 77A in the interaction [166]. The involvement of the lysine residues 36, 60F, 149E, and the arginine residues 67, 73, 75, and 77A in electrostatic interactions with hirudin has already been mentioned. However, the side chains of Lys109 and Lys110 are completely exposed to solvent and have no interactions with hirudin. Moreover, while Lys70 has no contacts with hirudin in the crystal structure of the complex, the Lys 70 ε-amino nitrogen (NZ) is involved in two internal ion pair interactions with Glu77 (3.5 Å) and with Glu80 (3.4 Å). Similar lysyl inaccessibility studies have been carried out by reductive methylation which imply hirudin protection for the lysines 36, 60F, 70, 81, 145, and 224. [169]. Of these newly implied lysine residues, Lys81 forms an internal hydrogen bond with Ala113 (Lys81 NZ-Ala 113 O, 2.73 Å), Lys145 interacts with Glu18 through a solvent-bridged hydrogen bond interaction (Lys145 NZ-W705, 2.87 Å; W705-Glu18 OE2, 2.80 Å), and Lys224

NZ forms a hydrogen bond with Ser19' O (3.10 Å) of hirudin (Table 17), as well as internal hydrogen bonds with Ser171 O (2.76 Å) and with Glu217 OE2 (2.60 Å).

The COOH-terminal tail of hirudin appears to be firmly anchored to thrombin at the end of the anion-binding exosite through hydrophobic interactions of the helical turn and the terminal Gln65'-Lys36 salt bridge. Although removal of Gln65' has little effect on the K_i inhibition or dissociation constant (increases by a factor of 1.3), deletion of Tyr63'-Leu64' raises the K_i by a factor of 40 [170]. The bulky 3_{10} helical turn may very well be important for hirudin-thrombin recognition, but whether this region of thrombin is part of the fibrinogen binding site remains uncertain.

An aspect of the hirudin COOH terminal tail-thrombin interaction in the anion-binding exosite region to emerge in importance as a result of the structure solution of the complex is the extent of the hydrophobic contribution. In the latter half of the C-terminal hirudin tail (Asp55'-Gln65'), nearly half of the residues (five of 11) are hydrophobic or aromatic in nature, and all five are engaged in nonpolar hirudin-thrombin interactions. The importance of these interactions to the anticoagulant activity of hirudin C-terminal tail analogues (like hirugen) has already become evident through thrombin binding studies which indicate that the minimal peptide necessary for detection of anticoagulant activity is Phe56'-Gln65' and, furthermore, that the activity is sensitive to modification of residues Phe56', Ile59', Pro60', Leu64', as well as Glu57' [89,95]. Based on the information in the crystal structure of the rHV2-K47-thrombin complex, it is clear that the hydrophobic contacts are at least as important as the polar interactions in the latter half on the hirudin COOH-terminal tail.

It is likely that hirudin binds with such high affinity to thrombin - forming tight, noncovalent complexes with dissociation constants in the pico-femtomolar range [59,71] - due to the fact that as it meanders across the the thrombin surface it is able to interact with a wealth of residues in a very intimate manner (Figure 56). All in all, 26 of the 65 residues of hirudin and 33 of the 259 residues in the thrombin B-chain are involved in 216 contacts of $< 4.0 \text{ \AA}$, of which eight are ion pairs and 12 are hydrogen bonds.

A summary of the hirudin-thrombin interactions is presented in Table 21. The hirudin N-terminal tripeptide is engaged in a remarkable 41 close contacts which are largely hydrophobic or neutral (30 out of 41 or 73%), as well as five hydrogen bonds. The remainder of the N-terminal hirudin domain, residues Thr4'-Pro48', have 62 close contacts with thrombin in which there are two hydrogen bonds, two ion pairs, and four nonpolar or neutral contacts. Thus, the N-terminal domain is responsible for 103 or 47% of the close contacts with thrombin; it is, however, important to also recognize that 41 or 40% of these 103 contacts are concentrated in the first three residues of hirudin.

As the C-terminal tail begins, there are 32 close contacts within the first three residues, Glu49'-His51', including two ion pair interactions, and one hydrogen bond as well as 19 close contacts along the length of the Glu49' and Trp60D side chains. The last three residues of this first segment of the C-terminal tail, Asn52'-Gly54', have poorly defined density. As the latter half of the C-terminal tail begins, there is again a large number of interactions within a small number of residues; 60 close contacts occur between the Asp55'-Glu58' stretch and thrombin. Among these contacts, the residues Asp55', Glu57', and Glu58' are largely involved in electrostatic interactions -

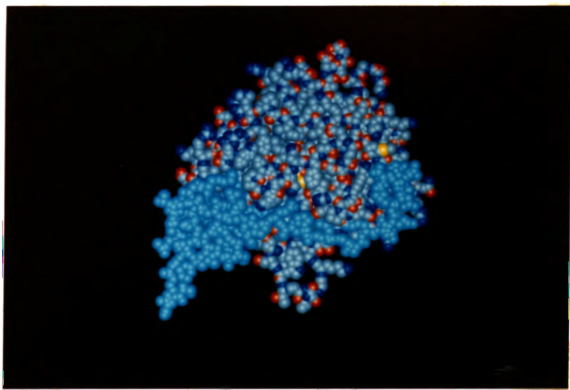


Figure 56. Space-filling Drawing of the rHV2-K47-Human α -Thrombin Complex. Hirudin in sky blue.

Table 21. Hirudin α -Thrombin Interactions. Numbers are intermolecular contacts < 4.0 Å; parentheses indicate ion pairs; a dash followed by a number is a hydrogen bond; hydrophobic or neutral contacts are underlined in bold.

H	I	T	Y	T	D	L	L	E	G	S	N	V	K	P	K	E	S	H	D	F	E	E	I	P	Y	Q	Sum
T	1	2	3	4	5	13	15	17	18	19	20	21	24	46	47	49	50	51	55	56	57	58	59	60	63	65	3
F 34																				<u>2</u>							6
K 36																											6
E 39																											7
H 57	<u>2</u>																										4
Y 60A	<u>1</u>																										5
P 60C						<u>1</u>							<u>3</u>	<u>1</u>	<u>2</u>												23
W 60D	<u>2</u>															19											4-1
K 60F																(4)-1							<u>1</u>				5
L 65																							<u>3</u>				3
R 67																											9-2
R 73																											19-1
T 74																											10
R 75																											15-1
Y 76																											8
R 77A																											2
I 82																											3
W 96																											2
L 99	<u>2</u>																										3
E146																											1
K149E																											1
R173																											15
I174																											2
C191																											2
E192																											6
S195	1-1																										1-1
S214	1-1																										1-1
W215	<u>3</u>																										3
G216	2-1	2	<u>5-1</u>																								9-2
E217			2																								17-1
G219	1	<u>2-1</u>																									5-1
C220	<u>3</u>																										3
R221A																											13
K224																											3-1
SUM	22	10	9	2	4	1	8	1	1	12	16	8	6	1	2	23	3	6	16	12	19	13	4	5	9	3	216-12
H-BONDS	3	2														1			2		2						TOTALS
TOTALS	14	41	3	4	4	1	8	1	1	12	16	8	6	1	2	49	32	51	55	60	58	59	21	65			

accounting for 48 or 80% of the contacts, of which four are hydrogen bonds and three are ion pairs. Moreover, Phe56' interrupts these electrostatic interactions by engaging in 12 close contacts, including 10 neutral or hydrophobic interactions. Lastly, in the region of the type III helical reverse turn, residues Ile59'-Gln65' produce 21 close contacts with thrombin in which 12 are neutral or nonpolar. The last hirudin-thrombin interaction, Lys36 NZ-Gln65 OXT (the terminal carboxylate oxygen), is an ion pair. In all, the hirudin C-terminal residues participate in 113 close contacts with thrombin. Moreover, as in the case of the hirudin N-terminal tripeptide, these contacts are concentrated in small stretches - 32 among Glu49'-His51', 60 among Asp55'-Glu58', and 21 among Ile59'-Pro60' and Tyr63'-Gln65'.

It is also worth noting that 21% (seven of 33) of the thrombin residues involved in contacts with hirudin of less than 4.0 Å are 'insertion residues', and therefore were judged to be inserted amino acids with respect to the sequence of chymotrypsinogen [6]. The major insertions in the thrombin sequence with respect to chymotrypsinogen are indicated in Figure 57, as well as the insertion residues which are directly involved in hirudin-thrombin interactions. Moreover, these residues account for 27% (58 of 216 contacts) of the total contacts. All of the interactions have been reviewed previously, and many of them appear to be important. The fact that insertion residues are responsible for a significant fraction of the hirudin-thrombin contacts, and considering that hirudin does not bind to other serine proteinases or blood proteins, suggests that the insertion residues are likely important to the diverse functionality of thrombin and to its specificity toward macromolecular substrates.

As another means to appreciate the extensive and intimate

hirudin-thrombin interaction, one can examine results obtained from solvent accessible surface calculations; these calculations were performed using a probe radius of 1.4 Å [171] on hirudin, thrombin, and the hirudin-thrombin complex, and the information gathered from this work is summarized in Tables 22-24. In its complex with thrombin, hirudin blocks 11.9% of the thrombin solvent accessible surface - masking 9.9% of the main chain atom accessibility and 12.4% of the side chain atoms (Table 22). This result is not surprising, for, as Figure 56 clearly indicates, hirudin covers only a small fraction of the large, nearly spherical thrombin surface. More impressive are the results presented in Tables 23 and 24. The information in Table 23 indicates that in the complex, thrombin masks 37.7% of the total solvent accessibility of hirudin, in which there is a 23.6% loss in main chain atom accessibility and a 42.4% loss of side chain atom accessibility. In addition, Table 24 reveals that the bulk of this loss is concentrated in the same small stretches of hirudin polypeptide which had a large number of close thrombin contacts in Table 21.

The hirudin N-terminal tripeptide, which displays the highest concentration of close thrombin contacts, also shows the greatest loss in solvent accessibility due to complexation - with a 90.7% total accessibility loss, of which 85.4% and 92.6% are due to a loss of main chain atom and side chain atom accessibility, respectively. As most of the remainder of the hirudin N-terminal domain, Thr4'-Pro48', is located away from the thrombin surface, this region shows the lowest loss of accessibility due to complex formation; among these residues there is a loss of only 20.2% in total atom accessibility, in which 7.2% is due to main chain atoms and 25.0% is due to side chain atoms.

The two short C-terminal hirudin polypeptide stretches which

Table 22. Solvent Accessible Surface Area of Thrombin Residues Alone and in Complex with Hirudin. Abbreviations used: TASA= total atom surface accessibility; MASA= main chain atom surface accessibility; SASA= side chain atom surface accessibility.

	Thrombin Alone (Å ²)	Thrombin in the Hirudin-Thrombin Complex (Å ²)	Loss of Thrombin Residue Accessibility Due to Complexation with Hirudin (Å ²) (%)
TASA	13,669.3	12,044.8	1,624.5 (11.9%)
MASA	2,650.0	2388.9	261.1 (9.9%)
SASA	11,019.3	9655.9	1,363.4 (12.4%)

Table 23. Solvent Accessible Surface Area of Hirudin Residues Alone and in Complex with Thrombin. Abbreviations used are as defined in Table 22.

	Hirudin Alone (Å ²)	Hirudin in the Hirudin-Thrombin Complex (Å ²)	Loss of Hirudin Residue Accessibility Due to Complexation with Thrombin (Å ²) (%)
TASA	4,957.3	3,090.6	1,866.7 (37.7%)
MASA	1,243.0	949.7	293.3 (23.6%)
SASA	3,714.3	2,140.9	1,573.4 (42.4%)

Table 24. Percent Loss of Hirudin Solvent Surface Accessibility Due to Complexation with Thrombin. Results presented with regard to hirudin residue ranges. Abbreviations used are as defined in Table 22.

Hirudin Residue Range:	%Loss TASA	%Loss MASA	%Loss SASA
1-3	90.7	85.4	92.6
4-48	20.2	7.2	25.0
49-51	46.8	30.0	51.0
55-58	56.4	31.9	62.3
59-65	35.7	34.5	36.0

display a high concentration of thrombin interactions (Glu49'-His51' and Asp55'-Glu58') in Table 21 also showed large losses in accessibility due to complexation. In the Glu49'-His51' and Asp55'-Glu58' segments there are 46.8% and 56.4% losses in total surface accessibility, respectively. Most notable, however, is the high loss in side chain atom accessibility in these regions which is 51.0% (Glu49'-His51') and 62.3% (Asp55'-Glu58'), and is consistent with the wealth of interactions involving the side chains of these residues. Finally, in the vicinity of the type III helical reverse turn of hirudin, one sees a modest 35.7% total loss in surface accessibility (34.5% main chain; 36.0% side chain) resulting from complexation. This result is likely due to the fact that intramolecular van der Waals contacts between side chain atoms both in and near the turn restrict access to residues in this region.

A listing of the residue surface accessibility of free hirudin is presented in Table 25. The table includes the total, main chain, and side chain atom surface accessibilities for each residue present in the crystal structure, as well as two ratios: the total atom accessibility (TASA) divided by the theoretical side chain atom accessibility (SASAt) and the side chain atom accessibility (SASA) divided the theoretical side chain accessibility. A TASA/SASAt ratio greater than one implies that there exists at least some main chain atom accessibility, while the SASA/SASAt ratio indicates the fraction of actual side chain atom accessibility with regard to the maximum possible. In general, residues with high TASA/SASAt and SASA/SASAt ratios are solvent accessible, and many of these participate in contacts in the complex. It is of note that while the mean values of the TASA/SASAt and SASA/SASAt ratios for most of the hirudin residues in the complex were 0.57 (+/- 0.33) and 0.47 (+/- 0.27), respectively, (residues poorly defined in electron

Table 25. Residue Surface Accessibility of Free Hirudin.
 Conformation of hirudin as it appears in the complex.
 Abbreviations used: SASAt= theoretical side chain
 atom surface accessibility. All other abbreviations
 used are as defined in Table 22. All surface
 accessibility results reported in Å². SASA/SASAt
 ratios indicate how the actual side chain
 accessibility compares to the maximum possible.
 TASA/SASAt ratios greater than one indicate that
 that there is at least some main chain accessibility.
 These ratios are not given for glycines as glycine has
 no non-hydrogen side chain atoms. A '*' preceding a
 line indicates that the values given are susceptible
 to error due to disorder in or near these residues.
 A '+' preceding a line indicates a residue which has
 at least one direct contact of less than 4.0 Å with
 thrombin.

Res.	MASA	SASA	TASA	$\frac{TASA}{SASAt}$	$\frac{SASA}{SASAt}$
+ Ile 1'	72.7	125.2	197.9	1.14	0.72
+ Thr 2'	29.0	94.9	123.9	0.98	0.75
+ Tyr 3'	8.9	84.2	93.1	0.43	0.39
+ Thr 4'	3.1	61.6	64.7	0.51	0.49
+ Asp 5'	1.0	90.5	91.5	0.73	0.72
Cys 6'	11.3	0.2	11.5	0.10	0.00
Thr 7'	23.7	66.9	90.6	0.72	0.53
Glu 8'	3.8	95.4	99.2	0.63	0.60
Ser 9'	25.2	42.5	67.7	0.71	0.45
Gly10'	19.6	0.0	19.6	***	***
Gln11'	0.0	2.4	2.4	0.01	0.01
Asn12'	0.0	4.6	4.6	0.03	0.03
+ Leu13'	7.6	39.3	46.9	0.28	0.24
Cys14'	0.0	0.0	0.0	0.00	0.00
+ Leu15'	3.8	34.2	38.0	0.23	0.20
Cys16'	10.1	0.0	10.1	0.09	0.00
+ Glu17'	7.6	70.5	78.1	0.49	0.45
+ Gly18'	62.9	0.0	62.9	***	***
+ Ser19'	23.8	77.3	101.1	1.06	0.81
+ Asn20'	2.9	98.8	101.7	0.72	0.70
+ Val21'	14.8	47.5	62.3	0.44	0.33
Cys22'	0.2	0.0	0.2	0.00	0.00
Gly23'	30.2	0.0	30.2	***	***
+ Lys24'	21.5	141.2	162.7	0.83	0.72
Gly25'	44.9	0.0	44.9	***	***
Asn26'	0.0	44.1	44.1	0.31	0.31
Lys27'	0.0	81.1	81.1	0.41	0.41
Cys28'	0.0	0.0	0.0	0.00	0.00
Ile29'	0.8	45.6	46.4	0.27	0.26
Leu30'	19.2	42.4	61.6	0.37	0.25
* Gly31'	77.3	0.0	77.3	***	***
* Gly36'	102.4	0.0	102.4	***	***
Asn37'	15.7	37.0	52.7	0.38	0.26
Gln38'	0.0	104.7	104.7	0.64	0.64

Table 25. (cont'd.)

Cys39'	16.9	20.4	37.3	0.33	0.18
Val40'	4.4	28.8	33.2	0.23	0.20
Thr41'	7.3	116.6	123.9	0.98	0.93
Gly42'	24.7	0.0	24.7	***	***
Glu43'	19.4	70.3	89.7	0.57	0.44
Gly44'	13.1	0.0	13.1	***	***
Thr45'	0.1	94.7	94.8	0.75	0.75
+ Pro46'	2.9	53.4	56.3	0.41	0.39
+ Lys47'	11.1	50.8	61.9	0.32	0.26
Pro48'	17.7	85.0	102.7	0.74	0.62
+ Glu49'	27.8	110.8	138.6	0.88	0.70
+ Ser50'	15.0	72.2	87.2	0.92	0.76
+ His51'	33.5	121.2	154.7	0.90	0.71
* Asn52'	28.8	93.5	122.3	0.87	0.67
* Asn53'	44.0	59.8	103.8	0.74	0.43
* Gly54'	62.2	0.0	62.2	***	***
+ Asp55'	38.0	95.3	133.3	1.07	0.76
+ Phe56'	30.1	127.8	157.9	0.79	0.64
+ Glu57'	18.4	127.1	145.5	0.92	0.80
+ Glu58'	28.3	125.2	153.5	0.97	0.79
+ Ile59'	9.5	82.6	92.1	0.53	0.48
+ Pro60'	12.1	64.6	76.7	0.56	0.47
Glu61'	4.2	122.2	126.4	0.80	0.77
Glu62'	26.8	103.7	130.5	0.83	0.66
+ Tyr63'	25.7	122.0	147.7	0.68	0.56
Leu64'	30.2	80.3	110.5	0.66	0.48
+ Gln65'	80.0	122.4	202.4	1.24	0.75

All residues:

mean:	0.57	0.47
standard deviation:	0.33	0.27

All residues which have
contacts of < 4.0 Å
with thrombin:

mean:	0.72	0.58
standard deviation:	0.29	0.20

density and glycines were excluded), those same ratios for residues directly involved in the hirudin-thrombin interaction were significantly higher - 0.72 (± 0.29) and 0.58 (± 0.20).

C. Crystal Packing

Intermolecular forces and interactions between molecules in the unit cell are an important part of any crystalline molecular structure. Such forces not only stabilize the three-dimensional crystalline array but are also likely important in influencing the symmetry within the unit cell and the structure of the component molecules. The rHV2-K47-thrombin complex crystallizes in the tetragonal system, space group $P4_32_12$, with eight molecules in the unit cell and one molecule in the asymmetric unit. As such, a 4_3 screw axis along c, a 2_1 screw axes along a and b, and a 2-fold symmetry axis along the a/b diagonal relate the crystallographically-independent molecules in the unit cell. The eight symmetry-related equivalent positions, along with that of the reference molecule (x,y,z), are given below [131]:

- | | |
|--------------------------------|--------------------------------|
| 1. x, y, z | 5. y, x, -z |
| 2. -x, -y, $1/2+z$ | 6. -y, -x, $1/2-z$ |
| 3. $1/2-y$, $1/2+x$, $3/4+z$ | 7. $1/2-x$, $1/2+y$, $3/4-z$ |
| 4. $1/2+y$, $1/2-x$, $1/4+z$ | 8. $1/2+x$, $1/2-y$, $1/4-z$ |

...(42)

In the crystals of the complex, both electrostatic and van der Waals forces appear to be important in the crystal packing interactions in the unit cell. A particularly striking hydrophobic cluster of residues occurs at a two-fold symmetry axis between the reference molecule (symmetry position 1, relation 42) and a two-fold symmetry-related molecule (symmetry position 5, relation 42) (Figure 58). Seven hydrophobic/aromatic groups from the two participating

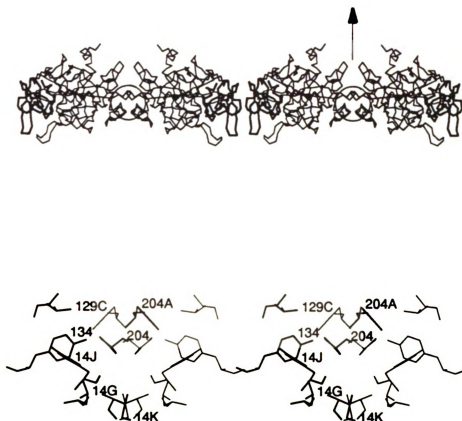


Figure 58. Stereoview of a Hydrophobic Crystal Packing Interaction at a Two-Fold Symmetry Axis. Top: Molecules involved are at symmetry positions 1 and 5 (relation 42, text). Disulfides and hirudin in bold. Two-fold axis indicated. The hydrophobic interactions occur in the circled region and are detailed below. Bottom: Nonpolar/aromatic residues which comprise the hydrophobic cluster. Groups involved: Leu14G, Tyr14J, Ile14K, Leu129C, Tyr134, Pro204, and Phe204A (all bold) and the corresponding symmetry-equivalent groups.

thrombin molecules comprise this cluster - Leu14G, Tyr14J, Ile14K, Leu129C, Tyr134, Pro204, and Phe204A (Figure 58). It is of particular note that three residues at this interface are from the thrombin A chain, a portion of thrombin surface which does not adjoin hirudin in the complex. Residue Leu14G has two nonpolar contacts $< 4.0 \text{ \AA}$ with Ile#14K (#14K indicating residue 14K of a thrombin molecule at symmetry position 5), and Ile14K has two close van der Waals contacts with Ile#14K in addition to the two with Leu#14G. Moreover, residue Tyr14J stacks against Tyr134 (five close contacts), and has two close approaches each with Pro204 and Leu14G. In the thrombin B chain, the side chain atoms of Leu129C have three van der Waals contacts of $< 4.0 \text{ \AA}$ with ring carbon atoms of Phe 204A, and residue Pro204 has three nonpolar contacts with Phe 204A as well as those with Tyr134. Finally, Phe204A has 13 close hydrophobic approaches with residues of the neighboring thrombin molecule - three with Leu#129C, four with Tyr#134, and two each with residues Tyr#14J, Pro#204, and Phe#204A. It is also of interest that this nonpolar cluster occurs in a region on the thrombin surface on the opposite side from the extended site of hirudin binding (Figure 58).

Another packing interaction related by two-fold symmetry is electrostatic, and is also a hirudin-thrombin contact. It involves residues Glu57'-Arg#75 and Glu#57'-Arg75 of the complex (positions 1 and 6, relation 42) (Figure 59). These ion pairs are of special interest for two reasons. First, the fact that they are double hydrogen-bonding ion pairs with good donor-acceptor distances (Arg#75 NH₂-Glu57' OE2, $2.93 \text{ \AA}$; Arg#75 NE-Glu57' OE1, $2.81 \text{ \AA}$) and good donor-proton-acceptor DHA angles (176° and 173°) (Table 26) suggests that the forces are strong and stabilizing packing interactions. Second, the

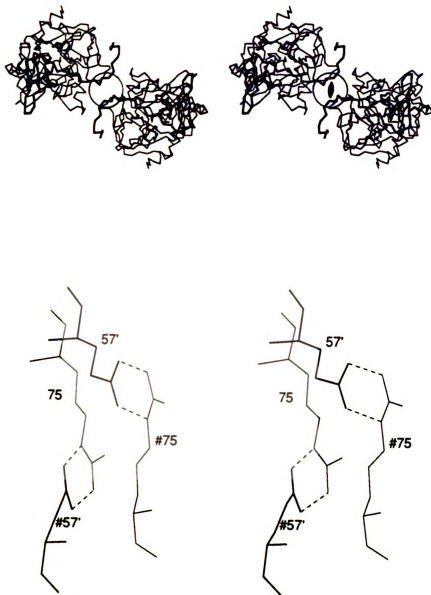


Figure 59. Stereoview of an Electrostatic Crystal Packing Interaction at a Two-Fold Symmetry Axis. Top: Molecules involved are at symmetry positions 1 and 6 (relation 42, text). Hirudin and all disulfides are in bold. Two-fold axis indicated. The electrostatic crystal packing interactions are within the circled region and are detailed below. Bottom: The double hydrogen-bonding ion pairs involved in the crystal contacts. Dashed lines indicate hydrogen bonds; hirudin glutamate residues are in bold. Two-fold symmetry axis offset approximately 30 degrees from the perpendicular to the plane of the paper.

Table 26. Hydrogen Bonds Between Symmetry-Related Molecules. Hydrogen atoms are assigned geometrically idealized positions. Donor atoms are denoted 'D' and acceptor atoms 'A'. The symmetry molecule number refers to the symmetry equivalent position of the symmetry-related molecule in the interaction, as given below. The symmetry-related molecule residue number is preceded by a '#'.
 1. x, y, z
 2. -x, -y, 1/2+z
 3. 1/2-y, 1/2+x, 3/4+z
 4. 1/2+y, 1/2-x, 1/4+z
 5. y, x, -z
 6. -y, -x, 1/2-z
 7. 1/2-x, 1/2+y, 3/4-z
 8. 1/2+x, 1/2-y, 1/4-z

Symmetry-equivalent Positions in P4₃2₁2:

Symmetry Molecule Number	Donor		Acceptor		Distances (Å)		Angles(deg)		
					H..A	D..A	DHA	CAH	CAD
7	Lys145	NZ	Gln#244	O	2.46	3.12	123	134	147
7	Trp148	N	Gly#1D	O	1.69	2.58	147	138	140
7	Thr149	N	Gln#244	OE1	1.99	2.96	161	120	114
7	Asn149B	ND2	Leu#123	O	1.69	2.73	172	145	148
7	Ala#1B	N	Trp148	O	2.04	2.94	155	160	160
6	Arg#75	NH2	Glu57	OE2	1.89	2.93	176	115	114
6	Arg#75	NE	Glu57	OE1	1.80	2.81	173	117	119
7	Lys#235	NZ	Asn149B	O	2.04	2.92	141	148	159
7	Gln#244	NE2	Thr149	O	1.94	2.87	155	122	129
7	Gln#244	NE2	Val149C	O	1.73	2.45	127	160	149
Mean:					1.93	2.83	153		
Standard Deviation:					0.23	0.20	19		

hydrogen-bonding ion pairs form between side chains of adjoining complex molecules, while the spatial proximity of these residues to one another within the complex suggests that in solution the ion pairs could form intramolecularly by free bond rotation.

Finally, there is an extended crystal packing interaction involving residues of the thrombin Lys145-Gly150 undecapeptide loop of one molecule (symmetry position 1, relation 42) and the NH₂- and COOH-terminal chain residues of another molecule (symmetry position 7, relation 42) (Figure 60). At this interface, a remarkable eight hydrogen bonds are formed with residues in the insertion loop (Table 26). The amide nitrogen of Ala#1B and the carbonyl oxygen of Gly#1D form hydrogen bonds with the carbonyl oxygen and amide nitrogen of Trp148, respectively. The C-terminal residue Gln#244 engages in four hydrogen bonds with residues of this thrombin loop: the carbonyl oxygen forms a hydrogen bond with Lys145 NZ, atom OE1 of the glutamyl side chain interacts with N of Thr149, and both protons of the side chain atom NE2 participate in hydrogen bonds with the carbonyl oxygens of Thr149 and Val149C. An aromatic stacking interaction also takes place between the phenyl ring of the C-terminal residue Phe#245 and Trp148. The loop conformation is further stabilized by two hydrogen bonds involving Asn149B and Leu#123, Lys#235. The wealth of interactions involving residues Lys145-Gly150, which is the largest protruding loop structure in the nearly spherical thrombin molecule (Figures 43, 56), could explain in part why it is so well-defined in the crystal structure of the hirudin-thrombin complex. In addition, it is also plausible that in the absence of this extensive network of stabilizing interactions, as in solution, the loop could be disordered.

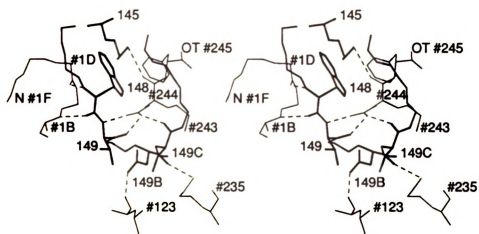


Figure 60. Stereoview of a Crystal Packing Interaction Involving the Thrombin Lys145-Gly150 Undeca peptide Loop. The crystal contact involves molecules at symmetry positions 1 and 7 (relation 42, text). Residues involved: Lys145, Trp148, Thr149, Asn149B, Val149C, Gly1D, Ala1B, Leu123, Lys235, Gln244, and Phe245. N, CA, C atoms and residues of the Lys145-Gly150 loop engaged in hydrogen bond interactions in bold; hydrogen bonds indicated by dashed lines.

VIII A COMPARISON OF THE THROMBIN STRUCTURE IN THE HIRUDIN-THROMBIN COMPLEX AND IN PPACK-THROMBIN

Having examined the structure of hirudin in the complex and the hirudin-thrombin interaction in detail, the thrombin structure of the complex and the manner in which it differs from that in PPACK-thrombin [6] will now be considered. An optimal superposition of the two thrombin structures on the basis of the main chain atoms N, CA, C was obtained by means of the program MOLFIT. This program calculates the rotation matrix and translation vector which 'best' superimpose one orthogonal coordinate set upon another by minimizing the root mean square (r.m.s.) deviation between the basis and transformed coordinate sets. The MOLFIT results for the two thrombin structures are shown in Table 27. The agreement on CA and on the main chain atoms N, CA, C is quite good (Table 27). The r.m.s. difference in N, CA, C reduces to 0.27 Å (609 atom pairs) when deviations larger than 1σ are excluded from the comparison. An examination of Figure 61 verifies that in general the overall folding is very similar between the two molecules. Not surprisingly, the r.m.s. differences of non-main chain atoms are larger (Table 27). Though the r.m.s. difference on sulfur atoms is 0.68 Å, both cysteine and methionine sulfurs are included in this value, and it is more meaningful to consider these different sulfur atoms separately. A calculation of the average distance between corresponding sulfurs reveals that for cysteine sulfur atoms the value is only 0.24 Å, while for methionine sulfurs the value is 0.60 Å. The small deviation with Cys sulfur atoms is likely due to the fact that they are all engaged in disulfide linkages, which are both conformationally restricted and important in maintaining the folding of the thrombin molecule. The

Table 27. Results of an Optimal Superposition of the Thrombin Structures of the Complex and PPACK-Thrombin [6] using the Program MOLFIT. Fitting based on N, CA, C Atoms. Results expressed in Å units as root mean square deviations (r.m.s.d.). Numbers in the '{}' brackets represent the number of atom pairs contributing to the r.m.s.d.

	r.m.s.d. (Å)	
Main Chain (N,CA,C) Atoms	0.45	{818}
Alpha Carbon Atoms	0.46	{272}
Carbonyl Oxygens	0.70	{274}
Sulfurs (Cys, Met)	0.68	{15}
Side Chain Atoms	1.43	{1103}
All Protein Atoms	1.08	{2195}

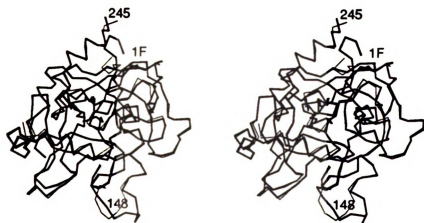


Figure 61. Stereoview of the Optimally Superimposed Thrombin CA Structures of the Complex and PPACK-Thrombin. Superpositioning corresponds to the MOLFIT N, CA, C fitting results of Table 27. Residues of the complex in bold.

larger deviation with Met sulfur atoms is consistent with the fact that these sulfurs are located in side chains which have significantly more flexibility than a cystine linkage. To be more exact, the average deviation in methionine sulfur positions is large mainly because of a single difference between the SD sulfurs of Met84 (2.42 Å), a surface residue.

To determine the location of the significant differences, and in an effort to assess the possible causes for the differences, the mean r.m.s. deviations of main chain and of side chain atoms versus residue number were examined (Figures 62 and 63). All residues with r.m.s. differences greater than 1σ (> 0.45 Å for all main chain atoms, > 1.43 Å for all side chain atoms) were visually examined using computer graphics in the optimally superimposed thrombin structures with the aid of an Evans and Sutherland PS390 stereographics display and the PSFRODO version of the molecular graphics program FRODO. Following this extensive comparison, most of the significant structural differences can be accounted for in terms of the following factors or influences: (a) tentatively placed residues, poorly defined residues in the electron density of either structure, or alternate interpretations of density maps; (b) affect of crystal packing interactions; (c) interaction with hirudin; and (d) variability of surface residue conformation. Listings, with pertinent comments for the residues possessing significant main chain and/or side chain r.m.s. deviations ($> 1\sigma$) believed to result from each of the four categories of factors or influences, are presented in Tables 28 to 31.

Inspection of Figures 62 and 63, as well as a more careful examination of Figure 61, reveals that the most outstanding main chain and side chain differences in the two structures occur at the N- and

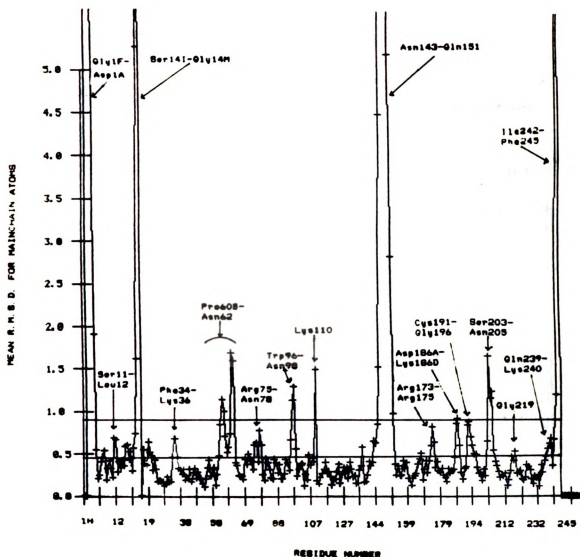


Figure 62. Average R.M.S. Deviation of Main Chain Atoms versus Residue Number for the Superimposed Thrombin Molecules of the Complex and PPACK-Thrombin. Superpositioning as in Figure 61. The horizontal lines represent 1σ and 2σ deviations. Most of residues with deviations > 1σ are identified.

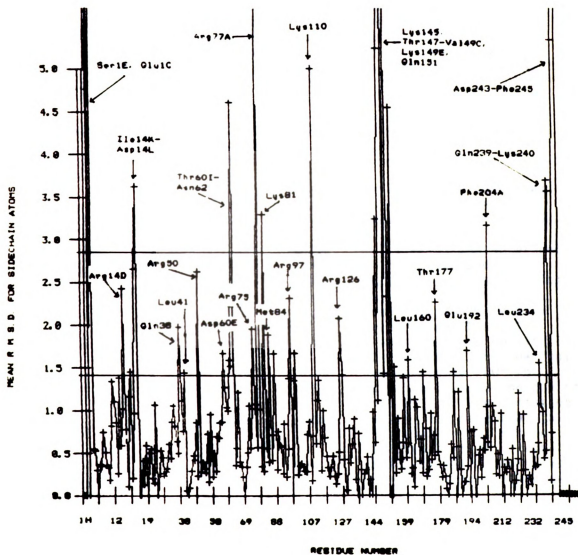


Figure 63. Average R.M.S. Deviation of Side Chain Atoms versus Residue Number for the Superimposed Thrombin Molecules of the Complex and PPACK-Thrombin. Superpositioning as in Figure 61. The horizontal lines represent 1σ and 2σ deviations. Many of the residues with deviations $> 1\sigma$ are identified.

Table 28. Structural Differences Between the Thrombin Structures of the Complex and PPACK-Thrombin [6] due to Residues Poorly Defined in Electron Density or Tentatively Placed in Either Structure, or due to Alternate Interpretations of the Density Maps. Optimal superpositioning of thrombin molecules as in Table 27. Abbreviations: HT = hirudin-thrombin complex; PT = PPACK-Thrombin; MCA r.m.s.d. = main chain atom average root mean square deviation; SCA r.m.s.d. = side chain atom average r.m.s.d.; S = surface; AS = active site; conf. = conformation; rot. = rotation. Deviations $> 1\sigma$ on main chain (0.45 Å) and/or side chain (1.43 Å) atoms are listed.

Residues(s)	Location	MCA r.m.s.d. (Å)	SCA r.m.s.d. (Å)	Comments
Gly1F	S	12.34	----	--Gly1F-Gly1D undefined in PT & tentatively placed along B-chain [6].
Ser1E	S	13.16	19.94	
Gly1D	S	9.71	----	
Ile14K	S	1.62	2.66	--Ile14K-Gly14M undefined in PT & arranged with regard to crystal packing [6].
Asp14L	S	5.28	3.63	
Gly14M	S	11.13	----	
Asp60E	S	1.01	1.67	--no side chain density beyond atom CB in HT
Glu97A	S	1.29	----	--no side chain density in HT new conf. for side chain
Leu99	AS	0.24	1.67	--approximate 180° rot. CB-CG bond; could be due to density map interpretation.
Lys110	S	1.50	5.00	--different main chain & side chain conf. due to cis-Pro111 in PT, trans-Pro111 in HT; density map interpretation difference.

Table 29. Structural Differences Between the Thrombin Structures of the Complex and PPACK-Thrombin [6] due to Crystal Packing Interactions in the Complex. Optimal superpositioning of thrombin molecules as in Table 27. Abbreviations: as = aromatic stacking; hb = hydrogen bond; hip = hydrogen-bonding ion pair; hc = hydrophobic contact; pc = polar contact; sympos = symmetry molecule positions (see Table 26), wmi = water mediated interaction. All other abbreviations are defined in Table 28. Residues in brackets [] are not involved in crystal contacts but have significant conformational differences which could be due to packing interactions of nearby residues. A '*' denotes the average MCA and SCA r.m.s.d. for residues Asn143-Gln151. Table 32 gives a complete listing of the MCA and SCA r.m.s.d.'s for the residues. Deviations $> 1\sigma$ on main chain (0.45 Å) and/or side chain (1.43 Å) atoms listed.

Residues(s)	Location	MCA r.m.s.d. (Å)	SCA r.m.s.d. (Å)	Packing Interactions/ Comments
Gly1F	S	12.34	----	--sympos-1,7; wmi O-W534-Glu#146 OE2.
Ser1E	S	13.16	19.94	--sympos-1,7; OG-Lys#47' NZ (pc) & OG-Thr#4'OG(pc)
Gly1D	S	9.71	----	--sympos-1,7: Trp#148(hb)
[Glu1C]	S	6.79	7.92	
Ala1B	S	1.91	0.55	--sympos-1,7: Trp#148(hb)
[Asp1A]	S	0.55	0.53	
[Ser14I]	S	0.57	0.20	
Tyr14J	S	0.74	0.97	--sympos-1,5: Tyr#134(as), Phe#204A(as).
Ile14K	S	1.62	2.66	--sympos-1,5: Leu#14G(hc), Ile#14K(hc).
Asp14L	S	5.28	3.63	--sympos-1,5: O-Arg#14D NH1 (pc)
[Gly14M]	S	11.13	----	
Arg75	S	0.61	1.96	--sympos-1,6: Glu#57'(hip)
Asn143-Gln151	S	3.99*	3.71*	--sympos-1,7: 8 hydrogen bonds involving these residues (see Table 26).
[Ser203]	S	0.65	1.03	
Pro204	S	1.65	0.62	--sympos-1,5: Ile#14K(hc), Pro#204A(hc)
Phe204A	S	1.15	3.16	--sympos-1,5: Tyr#14J(hc), Leu#129C(hc), Tyr#134(hc), Pro#204(hc), Phe#204A(hc)
[Asn204B]	S	1.23	0.56	
[Asn205]	S	0.54	1.04	
[Ile242]	S	0.67	0.72	
[Asp243]	S	1.19	5.35	
Asn244	S	5.83	11.31	--sympos-1,7: Lys#145(hb), Thr#149(hb), Val#149C(hb).
Phe245	S	6.57	13.52	--sympos-1,7: Trp#148(as)

Table 30. Structural Differences Between the Thrombin Structures of the Complex and PPACK-Thrombin [6] due to the Interaction with Hirudin. Optimal superpositioning of the thrombin molecules as in Table 27. All abbreviations are defined in Tables 28 and 29. Residues in brackets [] do not interact with hirudin, but have significant conformational differences which could be due to hirudin interactions of nearby residues. Deviations $> 1\sigma$ on main chain (0.45 Å) and/or side chain (1.43 Å) atoms listed.

Residues(s)	Location	MCA r.m.s.d. (Å)	SCA r.m.s.d. (Å)	Interacting Hirudin Residues/Comments
Phe34	S	0.46	0.87	Phe56'(hc)
Arg35	S	0.68	1.05	Glu49' via W508
Lys36	S	0.68	0.82	Gln65'(ip)
Pro60C	S	0.85	0.85	Leu13'(hc), Lys24'(pc), Pro46'(hc)
Trp60D	S	1.15	0.87	Ile1'(hc), Lys47'(pc), Glu49'(mostly hc)
Lys60F	S	0.71	1.28	Glu49'(hb)
Arg75	S	0.61	1.96	Glu#57'(hip-sympos=1,6)
[Tyr76]	S	0.64	1.08	
Arg77A	S	0.78	5.79	Glu58'(ip), Glu57' via W467
[Asn78]	S	0.61	1.01	
Asn143-Gln151	S	3.99*	3.71*	Lg. shift possibly explains CD change occurring upon hirudin C-terminal peptide bind- ing but not with PPACK.
Arg173	S	0.60	0.96	Glu17'(ip), Asn20'(pc)
Ile174	S	0.82	0.42	Val21'(pc)
[Arg175]	S	0.64	0.71	
Cys191	AS	0.84	0.34	Thr2'(pc)
Glu192	AS	0.88	1.70	Thr2'(pc), Thr4'(pc), Ser50'(pc)
[Gly193]	AS	0.70	---	
[Asp194]	AS	0.60	0.75	
Ser195	AS	0.50	0.77	Ile1'(hb)
[Gly196]	AS	0.49	---	
Gly219	AS	0.53	---	Thr2'(pc), Tyr3'(hb), Leu15'(pc)

Table 31. Structural Differences Between the Thrombin Structures of the Complex and PPACK-Thrombin [6] due to Surface Residue Conformational Variability. Optimal superpositioning of the thrombin molecules as in Table 27. Abbreviations: fs = free side chain; ident. = identical. All other abbreviations defined in Tables 28 and 29. Deviations $> 1\sigma$ on main chain (0.45 Å) and/or side chain (1.43 Å) atoms listed.

Residues(s)	Location	MCA r.m.s.d. (Å)	SCA r.m.s.d. (Å)	Comments
Pro5	S	0.54	0.75	
Ser11	S	0.69	1.34	--CA,C,O offset; fs, side chain nearly ident.
Leu12	S	0.67	1.11	--peptide atoms offset; fs, rot. approx. 90° CA-CB bond
Lys14A	S	0.34	1.38	-- new side chain conf.; new pc: NZ-Asp 21 O, 3.53 Å.
Arg14D	S	0.59	2.43	--new ip w/ Glu14H, 3.8 Å
Glu14H	S	0.30	1.40	
Glu14E	S	0.61	0.78	--peptide atoms offset; side chain conf. nearly ident. new ip w/ Glu14E, 3.8 Å.
Gly19	S	0.64	----	--main chain atoms of surface
Ser20	S	0.53	0.50	tripeptide Gly19-Asp21
Asp21	S	0.48	0.55	offset
Gln38	S	0.31	1.98	--approx. 90° rot. CA-CB bond
Leu41	S	0.20	1.45	--approx. 90° rot. CA-CB bond; new hc's w/ Phe60H
Arg50	S	0.23	2.63	--new conf. for side chain
Asn60G	S	0.52	1.03	--Thr60I-Asp63 turn shifted;
Phe60H	S	0.58	0.99	new ϕ , ψ angles for Asn62
Thr60I	S	0.74	1.59	classify turn as type I; enhanced hc's between Phe60H,
Glu61	S	1.69	1.53	Leu64, Leu41 in new conf.
Asn62	S	1.61	4.62	
His71	S	0.51	0.51	--main chain atoms offset; approx. 20° twist to histidine ring
Lys81	S	0.45	3.30	--new hb: NZ-Ala 113 O, 2.73 Å
Met84	S	0.23	1.89	--approx. 30° rot CA-CB bond
Lys87	S	0.40	1.67	--fs; approx. 30° rot. CB-CG bond
Trp96	S	0.67	1.39	--main chain atoms offset; side chain conf. nearly ident.

Table 31 (cont'd.)

Arg97	S	1.14	2.32	--main chain atoms offset; new conf. for side chain; new pc: NH1-Asn95 OD1, 3.35 Å
Asn98	S	0.56	1.35	--main chain atoms offset; new conf. for side chain
Arg126	S	0.26	2.08	--new side chain conf.
Glu127	S	0.34	1.42	--new side chain conf.
Leu160	S	0.21	1.59	--approx. 45° rot. CA-CB bond
Val167	S	0.51	0.66	--peptide atoms offset; side conf. nearly ident.
Lys169	S	0.32	1.45	--fs; new side chain conf.
Thr177	S	0.41	2.27	--approx. 180° rot. CA-CB bond
Asp186A	S	0.46	1.46	--fs; new side chain conf.
Glu186B	S	0.86	0.57	--peptide atom conformation significantly different for Glu186B-Lys186D
Gly186C	S	0.92	----	--new ip w/ Glu14E, 3.8 Å
Lys186D	S	0.60	1.22	--approx. rot. 180° CB-CG bond.
Leu234	S	0.37	1.56	--shifting of type III turn Lys235-Ile238
Trp237	S	0.47	0.44	--new side chain conf.
Ile238	S	0.56	0.52	--fs; new side chain conf.
Gln239	S	0.62	3.69	
Lys240	S	0.68	3.56	

Table 32. Structural Differences Between the Thrombin Structures of the Complex and PPACK-Thrombin [6] in the Vicinity of the 149A-149E Insertion Loop. Optimal superpositioning of the thrombin molecules as in Table 27. Abbreviations: r.m.s.d. = root mean square deviation; MCA r.m.s.d. = average main chain atom r.m.s.d.; SCA r.m.s.d. = average side chain atom r.m.s.d. All residues listed have deviations $> 1\sigma$ on main chain (0.45 Å) and/or side chain atoms (1.43 Å).

Residue	MCA r.m.s.d.(Å)	SCA r.m.s.d.(Å)
Asn143	0.65	0.98
Leu144	0.63	0.62
Lys145	0.85	3.24
Glu146	1.52	1.11
Thr147	4.48	5.25
Trp148	7.22	8.07
Thr149	7.54	9.52
Ala149A	7.59	1.43
Asn149B	8.84	5.88
Val149C	6.62	2.33
Gly149D	5.19	----
Lys149E	2.82	4.56
Gly150	0.97	----
Gln151	0.38	1.51

C-termini of the thrombin A-chain, the C-terminus of the thrombin B-chain, and in the loop containing the 149A-149E insertion. Moreover, it appears that the the aforementioned factors (a), (b) and (c) are at the core of these large deviations. The drastic differences in conformation of the tripeptide segments Gly1F-Ser1E-Gly1D and Ile14K-Asp14L-Gly14M at the N- and C-termini of the thrombin A-chain in the two structures (Figure 61), and the large average r.m.s. deviations for main chain (Figure 62, Table 27) and side chain atoms (Figure 63, Table 27) for these polypeptide segments results from the fact that these residues were totally undefined in the PPACK-thrombin structure, and were arranged in reasonable orientations - along the thrombin B-chain for residues 1F-1D, and in a conformation allowed by crystal packing for residues 14K-14M [6]. The probability that the conformation of these positioned residues would be similar to that observed in the hirudin-thrombin crystals is low.

Packing interactions in the crystals of the complex appear to play a role in the conformation of residues at the N- and C- termini of the thrombin A-chain, the C-terminus of the thrombin B-chain and in the Ala149A-Lys149E insertion loop (Tables 26 and 29). As was seen in Figure 60 and mentioned in the previous chapter, at a crystal contact between molecules at symmetry positions 1 and 7 (Table 26), A-chain residues Gly#1D and Ala#1B engage in two hydrogen bonds and B-chain residues Gln#244 and Phe#245 engage in four hydrogen bonds and an aromatic stacking interaction with residues in this insertion loop (Lys145, Trp148, Thr149, Asn149B, and Val149C). Such strong intermolecular forces will clearly have an impact on the conformation of the short polypeptide stretches Gly#1D-Ala#1B and Gln#244-Phe#245, as well as on nearby residues in these regions. In particular, Figure 64

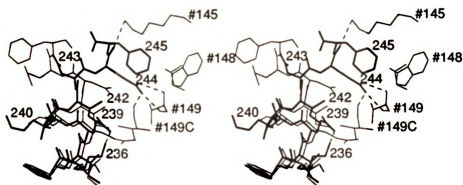


Figure 64. Stereoview of the Superimposed Residues Lys236-Phe245 of the Complex and PPACK-Thrombin. Superpositioning as in Figure 61. Residues of the complex in bold. Symmetry molecule interactions of the complex displayed; symmetry molecules indicated by a '#' preceding the residue number. Hydrogen bonds indicated by dashed lines.

reveals how interactions at an analogous crystal contact, involving Asn244 and Phe245 with residues in the 140-150 thrombin loop of a symmetry-related molecule, are most likely responsible for the conformation of the C-terminal residues Asp243-Phe245 in complex; the C-terminal B-chain folding of residues Lys236-Ile242 agrees quite well in the two structures but at Asp243 the termini diverge. In addition, the conformation of the N-terminus of the thrombin A-chain is likely influenced by the hydrophobic and aromatic stacking interactions involving Tyr14J and Ile14K (Figure 58) and by a close polar contact between the carbonyl of Asp14L and Arg#14D NH1 (3.57 Å) at a crystal contact involving symmetry-related molecules at crystal positions 1 and 5 (Table 26).

The wealth of hydrogen bonds (8) along the protruding thrombin insertion loop from residues Lys145 to Val149C (Table 26, Figure 60) and the aromatic stacking interaction between Trp148 and Phe#245 (Figure 60) at the crystal contact between molecules at symmetry positions 1 and 7 (Table 26) suggests that the conformation of this loop is well-stabilized by, if not dependent on, these interactions. The drastically different conformation this loop adopts in the two structures is shown in Figure 65. The most striking feature of the two loop conformations is the relative position of the indole ring of Trp148. In the complex, the indole is at the center of the loop and makes many contacts, while in PPACK-thrombin it is directed outside of the loop and toward the solvent. In addition, 14 residues in this loop have the r.m.s. differences $> 1\sigma$ on main and/or side chain atoms between the two structures; these values are presented in Table 32.

Crystal contacts alone may not be the only factor involved in influencing or provoking the dramatic shift in the conformation of this

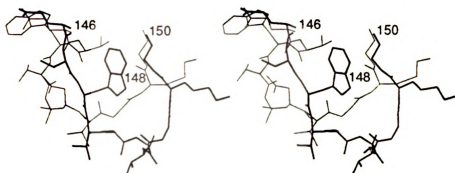


Figure 65. Stereoview of the Superimposed Residues Glu146-Gly150 of the Complex and PPACK-Thrombin. Superpositioning as in Figure 61. Residues of the complex in bold.

thrombin loop in the complex. The orientation of this loop in the complex cannot be the same as in the PPACK-thrombin structure, as it would result in steric problems and would prevent binding of the hirudin N-terminus in the active site region of thrombin; this point is graphically made in Figure 66. If the insertion loop adopted the same conformation in the complex as in PPACK-thrombin, the indole of Trp148 would practically collide with the Thr4'-Asp5' dipeptide of hirudin, and would obstruct the hirudin N-terminal tripeptide from penetrating into the thrombin active site. Thus, a change in the loop conformation between the two structures is necessary. Moreover, since the presence of the hirudin N-terminus in the thrombin active site region is important to the potency with which hirudin binds and inhibits thrombin, it is not surprising that the position of the loop and the indole of Trp148 in the complex are compatible with this important hirudin-thrombin interaction; the thrombin 140-150 loop residues are well away from the five hirudin N-terminal residues and the region where they interact with thrombin (Figure 66).

As hirudin is a 'double-headed' thrombin inhibitor, it is worth considering whether the binding of the N-terminal head or the C-terminal tail is responsible for triggering this necessary loop shift. On the basis of steric restraints and on the fact that the hirudin NH₂-terminus must enter the thrombin active site region for optimal binding, one could argue that the hirudin N-terminal head is responsible for the shift in the 149A-149E insertion loop. However, two different types of experimental evidence point to the likelihood of hirudin C-terminal tail binding to thrombin as the cause of the loop shift. Circular dichroism (CD) studies indicate that the binding of hirudin or a dodecapeptide analogue of the hirudin COOH-terminal tail to thrombin produces a change

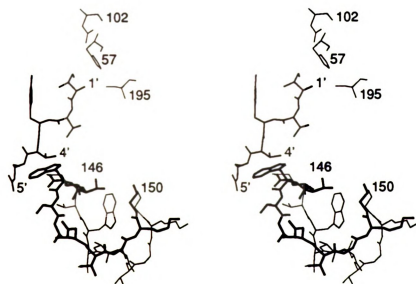


Figure 66. Stereoview of the Superimposed Residues Glu146-Gly150 of PPACK-Thrombin and the Complex, including the Hirudin N-Terminal Pentapeptide. Superpositioning as in Figure 61. Residues of PPACK-thrombin in bold. Thrombin catalytic triad of the complex displayed.

in CD which does not occur upon PPACK inhibition [95,96]. While the exact structural transitions responsible for these CD changes are not known, it is possible that the CD spectral results reflect the altered conformation for the indole of Trp148 and its associated loop. Moreover, the lack of change in CD upon PPACK binding to thrombin and the presence of one upon whole hirudin or C-terminal analogue binding suggest that hirudin binding in the vicinity of the anion-binding exosite, and not the thrombin active site, is responsible for the spectral change. In addition, studies of the kinetics of hirudin inhibition of thrombin indicate that C-terminal tail binding precedes N-terminal head binding, as the rate of the first step in complex formation is decreased by increasing the ionic strength and is unaffected by binding at the active site of thrombin [59]. Taken together, the experimental data suggest that initial C-terminal tail binding of hirudin to the anion-binding exosite region of thrombin is responsible for inducing a CD change, which could be due to a conformational change involving Trp148 and its associated loop. Thus, all in all, it appears as though the hirudin-thrombin interaction as well as crystal contacts are influential in determining the structure of the 149A-149E thrombin insertion loop.

While the largest differences in the thrombin structures have been examined, the majority of the significant differences have yet to be considered in terms of the four categories of influences likely involved in these changes. In category (a), two differences appear to result from poorly defined residues in the complex and two appear to be due to alternate interpretations of electron density maps (Table 28). The residues Asp60E and Glu97A of the complex show large side chain and/or main chain deviations with respect to the PPACK-thrombin structure, most

likely because there is little or no side chain density for these residues in hirudin-thrombin (Table 28). The structural deviations associated with Leu99 and Lys110 appear to stem from map interpretation (Table 28). Residue Leu99, in the active site region, has nearly the same main chain conformation in the two structures but differs by 1.67 Å in its side chain. Though this deviation is large in magnitude, in actuality the side chains in the two structures occupy the same space and simply differ by a near 180° rotation about the CB-CG bond; thus, either could be suitable for modeling in the density map. A much more dramatic difference, resulting from alternate density map interpretations, involves residues Lys110-Pro111 (Figure 67). The peptide bond preceding residue Pro111 was interpreted to be cis-, making this a cis-proline (CP), in PPACK-thrombin while in the complex a more typical trans-peptide bond was modeled. Moreover, while a cis- versus a trans-peptide bond does affect the configuration of Pro111, it places much more severe restraints on the conformation available to the preceding lysine residue, and is why the main and side chain deviations are so large for Lys110 (Figure 67).

Though many of the significant structural differences linked to crystal contact interactions have already been mentioned, there are a few more worth pointing out. The double-hydrogen bonding ion pair between Arg#75 and Glu57' which takes place at the two-fold axis between molecules at symmetry positions 1 and 6 (Table 26, Figure 59) is likely the source of the large main and side chain deviations between the two thrombin structures at Arg75 (Table 29). Also, the hydrophobic cluster created at another two-fold axis by residues from the thrombin A- and B-chains of molecules at symmetry positions 1 and 5 (Table 26, Figure 58) is probably the cause of the significant main and side chain

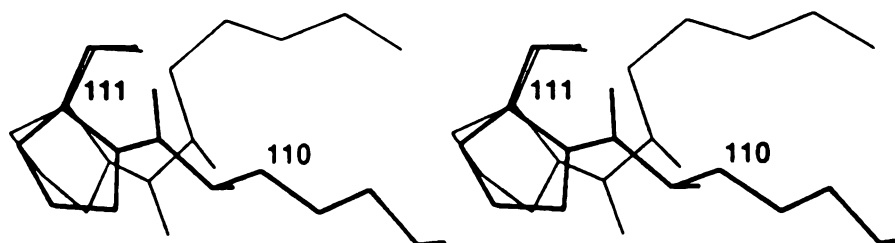


Figure 67. Stereoview of the Superimposed Residues Lys110-Pro111 of the Complex and PPACK-Thrombin. Superpositioning as in Figure 61. Residues of the complex in bold. Peptide bond preceding Pro111 is cis- in PPACK-thrombin, trans- in the complex.

deviations with respect to PPACK-thrombin for residues Pro204 and Phe204A, and those which adjoin them (Ser203, Asn204B, and Asn205). The heptapeptide segments Lys202-Arg206 taken from the main chain fitted thrombin structures are displayed in Figure 68, which reveals the noticeably different positioning of Phe204A in each molecule.

An examination of the entries in Tables 28 to 31 reveals that the majority of the conformational differences in the thrombin structures appear to arise from the hirudin-thrombin interaction and the variability of surface residues. The notion that the hirudin-thrombin interaction is responsible for producing numerous significant structural perturbations in the thrombin molecule with respect to the PPACK-thrombin structure is certainly reasonable when one considers the potency of this interaction, and the fact that only 26 of the 65 hirudin residues are responsible for producing a remarkable 216 contacts $< 4.0 \text{ \AA}$ with thrombin. Also not surprising is the fact that the bulk of the affected thrombin residues are associated with or adjoin important hirudin (and in general macromolecular) binding sites on thrombin - namely, the active site cavity and the anion-binding exosite region (Table 30). Moreover, these hirudin residues are largely found in short peptide stretches displaying high interaction per residue densities - Ile1'-Tyr3' (41 contacts/3 residues or 13.7 contacts/residue), Glu49'-His51' (32 contacts/3 residues or 10.7 contacts/residue), and Asp55'-Glu58' (60 contacts/4 residues or 15.0 contacts/residue) (Table 21).

In an effort to more carefully examine the conformational differences between the thrombin structures in the active site binding cavity, consider Figure 69 which displays the superimposed active sites from the hirudin-thrombin (**bold**) and PPACK-thrombin structures, and

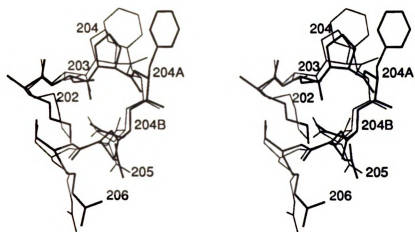


Figure 68. Stereoview of the Superimposed Residues Lys202-Arg206 of the Complex and PPACK-Thrombin. Superpositioning as in Figure 61. Residues of the complex in bold.

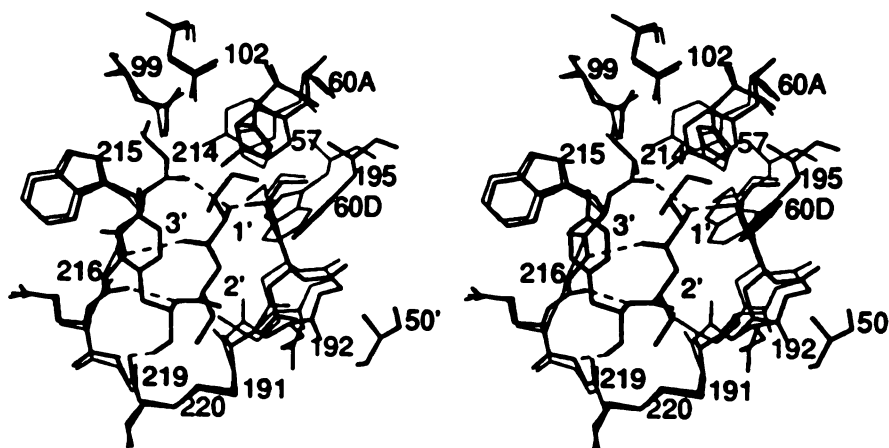


Figure 69. Stereoview of Superimposed Active Site Residues of PPACK-Thrombin and the Complex, including the Hirudin N-Terminal Tripeptide. Residues of the complex in bold. Superpositioning as in Figure 61. Hydrogen bonds indicated by dashed lines.

includes the hirudin N-terminal tripeptide (bold). It is interesting to note that while the residues involved in the β -strand interaction with the hirudin N-terminal tripeptide, Ser214-Gly219, are in nearly the same conformation in the two structures, residues Cys191-Ser195 and in the Tyr60A-Trp60D loop have quite different conformations between the two. The basis for the structural similarity of the Ser214-Cys220 polypeptide segments is likely due to the fact both hirudin and PPACK are involved in strand interactions with this rigid thrombin stretch and, even though the the strand interaction is antiparallel in PPACK and parallel in the complex, most likely the inhibitor tripeptides conform to the thrombin strand rather than vice versa. However, the significant structural perturbations at positions Cys191-Ser195 and Tyr60A-Trp60D most likely arise from hirudin-thrombin interactions which have no analogous match in the interaction of PPACK with thrombin. The large deviation in Cys191-Ser195 strand probably results from the need for Glu192 to adopt a different conformation in the complex, as well as from a unique interaction between Ser195 and the hirudin NH₂-terminus. The Glu192 side chain cannot have the same position in the complex as in PPACK-thrombin for its carboxylate would come too close to the Thr2' side chain of hirudin. This problem coupled with the fact that the side chain can form a stabilizing close polar contact with Ser50' is sufficient to induce the Glu192 main chain and side chain atoms to shift. Similarly, the structural perturbation at Ser195 is a result of the unique hydrogen bond interaction between Ile1' N and Ser195 OG. The residues which adjoin Ser195 and Glu192 - Cys191, Gly193, and Asp194 - undergo significant main chain structural changes due to the Glu192, Ser195 hirudin-thrombin interactions. Moreover, while the proline residue of PPACK and the Ile1' residue of hirudin both have hydrophobic

contacts with Tyr60A and Trp60D of the thrombin 60A-60I insertion loop, the significant structural differences in the conformations of Pro60C, Trp60D, and Lys60F also result from unique hirudin-thrombin interactions which are unmatched in PPACK-thrombin; these include the van der Waals interaction between Pro46' and Pro60C, the close polar contacts between Lys47' and Glu49' with Trp60D, and the ion pair between Glu49' and Lys60F (Table 30).

As PPACK has no interactions with the thrombin anion-binding exosite while hirudin has numerous hydrophobic and electrostatic contacts here, there are expectedly many significant deviations between the two thrombin structures in main and side chain conformation at this binding region (Table 30). Among the Phe34-Ser41 loop of residues which comprise the anion-binding exosite, Phe34, Arg35, and Lys36 display significantly altered main chain and somewhat altered side chain configurations in the complex. These deviations are probably the result of the following contacts: the hydrophobic stacking interaction between Phe34 and Phe56', the water-mediated contact between Arg35 and Glu49', and the ion pair involving the C-terminal carboxylate of Gln65' and Lys36. A number of such structural changes can be found in the Lys70-Lys81 loop of the anion-binding exosite; these differences can be appreciated from Figure 70, which displays Arg75-Lys81 of both thrombin structures superimposed, as well as some pertinent hirudin residues. The difference in the conformation of Arg75 is probably due the double hydrogen-bonding ion pair between this residue and Glu 57' of a symmetry molecule. The major main and side chain deviations at Arg77A stem from the direct ion pair formed between this arginyl side chain and Glu58' of hirudin, as well as from the longer range interaction with Glu57' via the water molecule, W467. Lastly, the differences at Tyr76 and Asn78

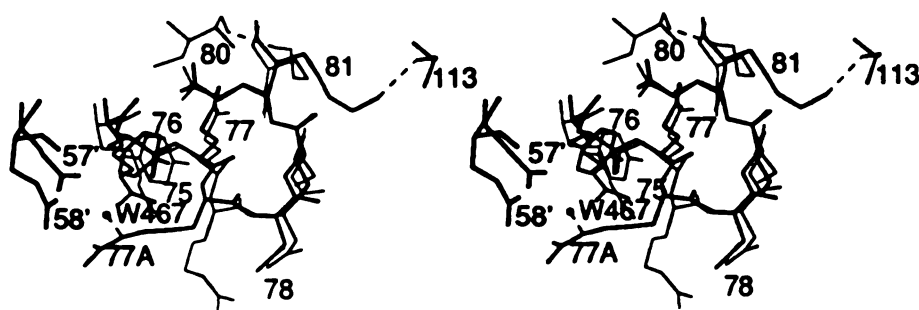


Figure 70. Stereoview of the Superimposed Anion-Binding Exosite Residues Arg75-Lys81 of the Complex and PPACK-Thrombin. Additional residues of the complex (thrombin Ala113, hirudin Glu57'-Glu58', and the water molecule W467) included. Residues of the complex in bold. Superpositioning as in Figure 61. Hydrogen bonds indicated by dashed lines.

are likely a by-product of the hirudin-thrombin interactions of the nearby residues.

Also present in Figure 70 is another type of conformational difference; the side chain NZ of Lys81 participates in a hydrogen bond with the carbonyl oxygen of Ile82 in PPACK-thrombin (2.84 Å), while in the complex it forms a hydrogen bond with the carbonyl oxygen of Ala113 (2.73 Å). This difference is a clear example of the conformational variability possible with surface residues [172-174], and represents the most common type of structural alteration found in the two thrombin molecules; overall 40 thrombin residues have significant conformational differences in the two structures which appear to be due to this (Table 31). Another interesting example this of variability involves the residues Leu41, Phe60G-Leu64 (Table 31), and is displayed in Figure 71. The Thr60I-Asn63 surface turn shifts in the complex with respect to that in PPACK-thrombin. Moreover, while the conformational ϕ , ψ angles of Asn62 in PPACK-thrombin do not conform with any of the three major turn classifications [149], these angles are quite different in the complex, and identify this turn as type I. The shifting of the turn is the main cause of the large side chain deviations for residues Thr60I-Glu61, while the drastic change in the ϕ , ψ conformational angles at Asn62 as well as the turn shift are responsible for the large main and side chain differences for this residue. In addition, several of the new side chain conformations in this loop of the hirudin-thrombin complex appear to stabilize the shift in turn. In particular, the large side chain conformational changes at Leu41 and Leu64 in addition to the significant main chain shift for Phe60H in the thrombin structure of the complex result in there being six close nonpolar contacts of < 4.0 Å between Phe60H and Leu41, Leu64 (three with each residue); in the PPACK-thrombin

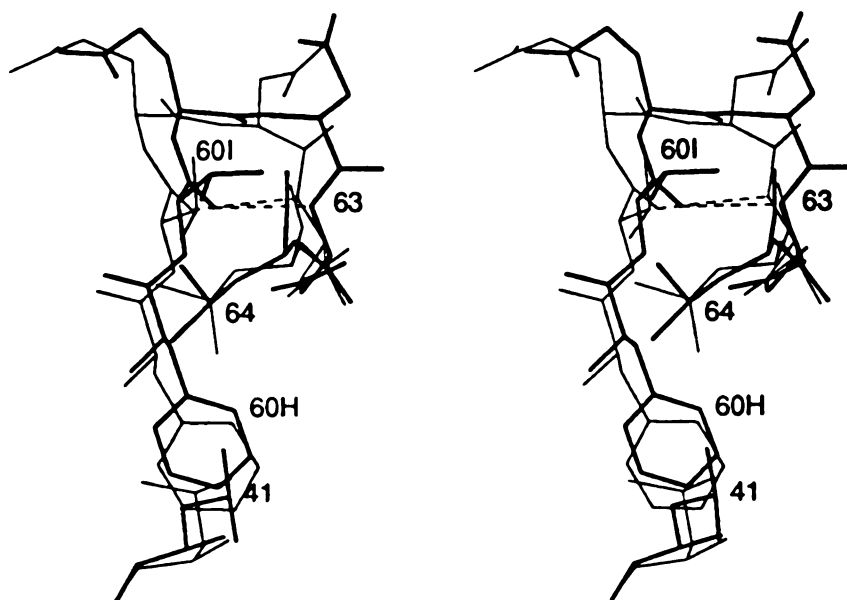


Figure 71. Stereoview of Residues Leu41 and Phe60H-Leu64 of the Complex and PPACK-Thrombin. Residues of the complex in bold. Superpositioning as in Figure 61. Hydrogen bonds indicated by dashed lines.

structure there is only one close contact $< 4.0 \text{ \AA}$ between Phe60H and Leu64, and none involving Leu41. The shift in the side chain orientation at Thr60I in the complex also results in a new close polar contact with Asp63 (Thr60I OG1-Asp63 OD1, $3.27 \text{ \AA}$) which is not present in the PPACK-thrombin structure.

A careful examination of Table 31 reveals that, all in all, not only is the number of residues with main/side chain structural changes likely due to surface residue conformational variability large (40), but so is the extent to which this variability manifests itself. Ten residues are engaged in new contacts: four are involved in ion pairs (Arg14D, Glu14H, Glu14E, and Lys186D), three form close polar contacts (Lys14A, Arg97, and Thr60I), two engage in close van der Waals contacts (Leu41, Phe60H), and one is involved in a hydrogen bond interaction (Lys81). The conformation for eight residues is significantly affected by a shifting in the position of two surface turns: five (Phe60G-Asn62) are affected by the shifting of a type I turn involving residues Thr60I-Asn62, and three (Trp237-Gln239) are affected by the shifting of a type III turn involving residues Lys235-Ile238. In addition, all but 11 residues in Table 31 have significantly altered side chain atomic positions, with six residues displaying dramatically new side chain conformations because they are unstabilized and directed into the solvent (Ser11, Leu12, Lys87, Lys169, Asp189, and Lys240) and six more displaying new side chain conformations simply by reason of free rotation about a single bond (Gln38, Leu41, Lys87, Leu160, Thr177, and Leu234).

It has been possible to associate most of the major main and side chain structural differences between the thrombin molecules in the hirudin-thrombin complex and in PPACK-thrombin with one or another of

the (a)-(d) influences. However, all of these differences have involved solvent-accessible residues located on the thrombin surface - and have been classified in Tables 28 to 31 as belonging to the general thrombin surface (S) or the specialized surface of the active site binding cavity (AS). If one excludes the problematic differences associated with category (a), which probably all appear on the surface only fortuitously, some clear implications can be drawn from this comparison. First of all, only surface residues display significantly altered main/side chain structure in the two thrombin molecules. Second, the deviations associated with these surface residues appear to result not only from binding or crystal contact interactions but also from an inherent conformational flexibility which is likely due to being on a solvent-accessible surface [172-174]. Finally, the main chain and side chain folding of 'internal' or non-surface residues agreed very well - displaying no major ($> 1\sigma$) deviations.

As thrombin is a large, globular, and well-defined serine proteinase, the conclusions drawn cannot be expected to be of general applicability to all structural comparisons involving the same protein or enzyme in free or complexed states. However, it is of particular note that the observations made in this thrombin comparison, in general, agree well with conclusions drawn from a very careful, comprehensive, and detailed structural comparison made on a close relative of thrombin, α -chymotrypsin [174]. In this study, Blevins and Tulinsky examined the structural differences in the two independent molecules of the α -chymotrypsin dimer crystal structure at 1.67 Å resolution, and further compared these independent molecules with monomeric γ -chymotrypsin. Among the conclusions they drew from this study were that the main chain folding was basically identical between the two independent molecules in

the dimer, and that there was marked asymmetry in the side chain structure of the surface residues - some which could be explained by interfacial interactions but much of which had to be attributed to general flexibility in surface side chain structure. Moreover, when the independent molecules of the chymotrypsin dimer were compared to γ -chymotrypsin, although the two molecules of the dimer were more like one another than γ -chymotrypsin, the same conclusions could be drawn from the independent molecules in the dimer comparison. The chymotrypsin comparison, in particular, elegantly displays that even in a situation where the molecules being compared are exposed to the same macroscopic chemical and pH environment, there is still significant surface side chain variability which cannot be explained by interface interactions - emphasizing that protein surface structure is variable as well as adaptable.

One important difference in the conclusions drawn from the thrombin and chymotrypsin structural comparisons is that there were no significant main chain deviations involving surface residues among the chymotrypsin molecules while there were between the two thrombin structures. This discrepancy is probably due to two factors. In the chymotrypsin (CHT) comparison, basically the same free enzyme molecule is examined in all the situations. In the comparison of unique molecules within the dimer of α -CHT, differences simply due dimerization were examined, and in the α -CHT versus γ -CHT comparisons, differences due to pH (α -CHT, pH 3.5; γ -CHT, pH 5.5) and aggregation state (α -CHT, dimer; γ -CHT, monomer) were assessed. In the thrombin comparison, on the other hand, thrombin structures were contrasted which not only differed in pH and chemical environment but, more importantly, were differently inhibited molecules: the thrombin inhibitors differed

significantly in their extent of interaction with the thrombin surface. In particular, it appears that unique hirudin-thrombin interactions are likely the source of many of the significant main (as well as side) chain deviations between the two structures. Also, as the thrombin molecule can be viewed as possessing 12 insertions with respect to chymotrypsinogen (Figure 57) [6], the significant number of main chain differences between the thrombin structures could be due to the fact that the thrombin surface is inherently more adaptable than that of chymotrypsin as a result of these inserted sequences.

LIST OF REFERENCES

LIST OF REFERENCES

1. Magnusson, S. (1971) in **The Enzymes** (Boyer, P.D., Ed.) pp 277-321, Academic Press, New York.
2. Nesheim, M.E., Hibbard, L.S., Tracy, P.B., Bloom, J.W., Myrmel, K.H., & Mann, K.G. (1980) in **The Regulation of Coagulation** (Mann, K.G., & Taylor, F.B., Eds.) pp 145-149, Elsevier-North Holland, New York.
3. Mann, K.G. (1976) **Methods Enzymol.** 45, 123-156.
4. Butkowski, R.J., Elion, J., Downing, M.R., & Mann, K.G. (1977) **J. Biol. Chem.** 252, 4942-4957.
5. Thompson, A.R., Enfield, D.L., Ericsson, L.H., Legaz, M.E., & Fenton II, J.W. (1977) **Arch. Biochem. Biophys.** 178, 356-367.
6. Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S.R., & Hofsteenge, J. (1989) **EMBO J.** 8, 3467-3475.
7. Jackson, C.M. & Nemerson, Y. (1980) **Annu. Rev. Biochem.** 49, 765-811.
8. Nilsson, B., Horne, M.K. III, & Gralnick, H.R. (1983) **Arch. Biochem. Biophys.** 224, 127-133.
9. Doolittle, R.F. (1984) **Annu. Rev. Biochem.** 53, 195-229.
10. Hantgan, R.R., & Hermans, J. (1979) **J. Biol. Chem.** 254, 11272-11281.
11. Blomback, B., Hessel, B., Hogg, D., & Therkildsen, L. (1978) **Nature** 275, 501-505.
12. Blomback, B. (1986) **Ann. N.Y. Acad. Sci.** 485, 120-123.
13. Casassa, E.F. (1955) **J. Chem. Phys.** 23, 596-597.
14. Carr, M.E., Shen, L.L., & Hermans, J. (1977) **Biopolymers** 16, 1-15.
15. Ferry, J.D., & Morrison, P.R. (1947) **J. Am. Chem. Soc.** 69, 388-400.
16. Lorand, L., Downey, J., Gotoh, T., Jacobsen, A., & Tokura, S. (1968) **Biochem. Biophys. Res. Commun.** 31, 222-230.
17. Fenton II, J.W. (1981) **Ann. N.Y. Acad. Sci.** 370, 468-495.

18. Fenton II, J.W. (1986) **Ann. N.Y. Acad. Sci.** 485, 5-15.
19. Esmon, C.T., Esmon, N.L., Kurosawa, S., & Johnson, A.E. (1986) **Ann. N.Y. Acad. Sci.** 485, 215-220.
20. Rick, M.E. & Hoyer, L.W. (1978) **Brit. J. Haematol.** 38, 107-119.
21. Nesheim, M.E. & Mann, K.G. (1979) **J. Biol. Chem.** 254, 1326-1334.
22. Seegers, W.H., Novoa, E., Henry, R.L., & Hassouna, H.I. (1976) **Thromb. Res.** 8, 543-552.
23. Kisiel, W., Canfield, W.M., Ericcson, L.H., & Davie, E.W. (1977) **Biochemistry** 16, 5824-5831.
24. Lorand, L. (1975) in **Proteases and Biological Control** (Reich, E., Rigkin, D.B., & Shaw, E., Eds.) pp 79-84, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
25. Travis, J. & Salvesen, G.S. (1983) **Annu. Rev. Biochem.** 52, 655-709.
26. Bjork, I. & Lindahl, U. (1982) **Mol. Cell. Biochem.** 48, 161-182.
27. Pratt, C.W., Whinna, H.C., Meade, J.B., Treanor, R.E., & Church, F.C. (1989) **Ann. N.Y. Acad. Sci.** 556, 104-115.
28. Markwardt, F. (1970) **Methods. Enzymol.** 19, 924-932.
29. Shuman, M.A. (1986) **Ann. N.Y. Acad. Sci.** 485, 228-239.
30. Weksler, B.B., Ley, C.W., & Jaffe, E.A. (1978) **J. Clin. Invest.** 62, 923-930.
31. Prescott, S.M., Zimmerman, G.A., & McIntyre, T.M. (1984) **Proc. Natl. Acad. Sci. USA** 81, 3534-3538.
32. De Groot, P.G., Gonsalves, M.D., Loesberg, C., van Buul-Wortelboer, M.F., van Aken, W.G., & van Mourik, J.A. (1984) **J. Biol. Chem.** 259, 13329-13333.
33. Gelehrter, T.D., & Sznycer-Laszyk, R. (1986) **J. Clin. Invest.** 77, 165-169.
34. Bender, M.L., & Kezdy, F.J. (1964) **J. Am. Chem. Soc.** 86, 3704-3714.
35. Inward, P.W., & Jencks, W.P. (1965) **J. Biol. Chem.** 240, 1986-1996.
36. Stroud, R.M., Kay, L.M., & Dickerson, R.E. (1974) **J. Mol. Biol.** 83, 185-208.
37. Birktoft, J.J., & Blow, D.M. (1972) **J. Mol. Biol.** 68, 187-240.
38. Shotton, D.M., & Watson, H.C. (1970) **Nature** 225, 811-816.

39. Sawyer, L., Shotton, D.M., Campbell, J.W., Wendell, P.L., Muirhead, H., Watson, H.C., Diamond, R., & Ladner, R.C. (1978) **J. Mol. Biol.** 118, 137-208.
40. Berliner, L.J. & Shen, Y.Y.L. (1977) **Biochemistry** 16, 4622-4626.
41. Sonder, S.A., & Fenton II, J.W. (1984) **Biochemistry** 23, 1818-1823.
42. Berliner, L.J., Sugawara, Y., & Fenton II, J.W. (1985) **Biochemistry** 24, 7005-7009.
43. Olson, T.A., Sonder, S.A., Wilner, G.D., & Fenton II, J.W. (1986) **Ann. N.Y. Acad. Sci.** 485, 96-103.
44. Fenton II, J.W., Olson, T.A., Zabinski, M.P., & Wilner, G.D. (1988) **Biochemistry** 27, 7106-7112.
45. Church, F.C., Pratt, C.W., Noyes, C.M., Kalayanamit, T., Sherrill, G.B., Tobin, R.B., & Meade, J.B. (1989) **J. Biol. Chem.**, 264, 18419-18425.
46. Bar-Shavit, R., Kahn, A.J., Mann, K.G., & Wilner, G.D. (1986) **Proc. Natl. Acad. Sci. USA** 83, 976-980.
47. Fenton II, J.W. (1988) **Semin. Thromb. Hemost.** 14, 234-240.
48. Boissel, J-P., Le Bonniec, B., Rabinet, M-J., Labie, D., & Elion, J. (1984) **J. Biol. Chem.** 259, 5691-5697.
49. Lewis, S.D., Lorand, L., Fenton II, J.W., & Shafer, J.A. (1987) **Biochemistry** 26, 7597-7603.
50. Bezeaud, A. & Guillin, M-C. (1988) **J. Biol. Chem.** 263, 3576-3581.
51. Braun, P.J., Hofsteenge, J., Chang, J-Y. & Stone, S.R. (1988) **Thromb. Res.** 50, 273-283.
52. Hofsteenge, J., Braun, P.J. & Stone, S.R. (1988) **Biochemistry** 27, 2144-2151.
53. Tsernoglou, D., Walz, D.A., McCoy, L.E., & Seegers, W.H. (1974) **J. Biol. Chem.** 249, 999 (1974).
54. McKay, D.B., Kay, L.M., & Stroud, R.M. (1977) in **Chemistry and Biology of Thrombin** (Fenton II, J.W., Lundblad, R.L. & Mann, K.G. Eds.) pp 113-121, Ann Arbor Science, Ann Arbor.
55. Edwards, B.F.P., Kumar, V., Bedford, B.A., Martin, P.D., & Kunjummen, R.D. (1986) **Ann. N.Y. Acad. Sci.** 485, 411-413.
56. Skrzypczak-Jankun, E., Rydel, T.J., Tulinsky, A., Fenton II, J.W., & Mann, K.G. (1989) **J. Mol. Biol.** 206, 755-757.

57. Magnusson, S., Peterson, T.E., Sottrup-Jensen, L., & Claeys, H. (1975) in **Proteases and Biological Control** (Reich, E., Rifkin, D.B., & Shaw, E., Eds.) pp 123-150, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
58. Furie, B., Bing, D.H., Feldmann, R.J., Robison, D.J., Burnier, J.P., & Furie, B.C. (1982) **J. Biol. Chem.** 257, 3875-3882.
59. Stone, S.R. & Hofsteenge, J. (1986) **Biochemistry** 25, 4622-4628.
60. Walsman, P., & Markwardt, F. (1981) **Pharmazie** 36, 653-660.
61. Scharf, M., Engels, J., & Tripier, D. (1989) **FEBS Lett.** 255, 105-110.
62. Dodt, J., Muller, H.P., Seemuller, U., & Chang, J.Y. (1984) **FEBS Lett.** 165, 180-183.
63. Harvey, R.P., Degryse, E., Stefani, L., Schamber, F., Cazenave, J.P., Courtney, M., Tolstoshev, P., & Lecocq, J.P. (1986) **Proc. Natl. Acad. Sci. USA** 83, 1084-1088.
64. Dodt, J., Machleidt, W., Seemuller, U., Maschler, R., & Fritz, H. (1986) **Biol. Chem. Hoppe-Seyler** 367, 803-811.
65. Loison, G., Findeli, A., Bernard, S., Nguyen-Juilleret, M., Marquet, M., Reihl-Bellon, N., Carvallo, D., Guerra-Santos, L., Brown, S.W., Courtney, M., Roitsch, C., & Lemoine, Y. (1988) **Bio/Technology** 6, 72-77.
66. Courtney, M., Loisson, G., Lemoine, Y., Riehl-Bellon, N., DeGryse, E., Brown, S.W., Cazanave, J.P., Defreyn, G., Delebassee, D., Bernat, A., Maffrand, J.P., & Roitsch, C. (1989) **Semin. Thromb. Hemost.** 15, 288-292.
67. Johnson, P.H., Sze, P., Winant, R., P.W. Payne, and J.B. Lazar, **Semin. Thromb. Hemost.** 15, 302 (1989).
68. Bagdy, D., Barabas, E., Graf, L., Peterson, T.E., & Magnusson, S. (1976) **Methods Enzymol.** 45, 669-678.
69. Fink, E. (1989) **Semin. Thromb. Hemost.** 15, 283-287.
70. Chang, J.Y. (1983) **FEBS Lett.** 164, 307-313.
71. Seemuller, U., Dodt, J., Fink, E., & Fritz, H. (1986) in **Proteinase Inhibitors** (Barrett, A., & Salvesen, G., Eds.) pp 337-359, Elsevier, New York.
72. Dodt, J., Seemuller, U., Maschler, R., & Fritz, H. (1985) **Biol. Chem. Hoppe-Seyler** 366, 379-385.
73. Laskowski, M. Jr., & Kato, I. (1980) **Annu. Rev. Biochem.** 49, 593-626.

74. Apella, E., Weber, I.T., Blasi, F. (1988) **FEBS Lett.** 231, 1-4.
75. Dodt, J., Schmitz, T., Schafer, T., & Bergmann, C. (1986) **FEBS Lett.** 202, 373-377.
76. Fortkamp, E., Rieger, M., Heisterberg-Moutsers, G., Schweitzer, S., & Sommer, R. (1986) **DNA** 5, 511-517.
77. Bergman, C., Dodt, J., Kohler, S., Fink, E., & Gassen, H.G. (1986) **Biol. Chem. Hoppe-Seyler** 367, 731-740.
78. Sukumaran, D.K., Clore, G.M., Preuss, A., Zarbock, J., & Gronenborn, A.M. (1987) **Biochemistry** 26, 333-338.
79. Folkers, P.J.M., Clore, G.M., Driscoll, D.C., Dodt, J., Kohler, S., & Gronenborn, A.M. (1989) **Biochemistry** 28, 2601-2617.
80. Haruyama, H., & Wuthrich, K. (1989) **Biochemistry** 28, 4301-4312.
81. Stone, S.R., Braun, P.J., & Hofsteenge, J. (1987) **Biochemistry** 26, 4617-4624.
82. Fenton II, J.W., Landis, B.H., Walz, D.A., & Finlayson, J.S. (1977) in **Chemistry and Biology of Thrombin** (Fenton II, J.W., Lunblad, R.L., & Mann, K.G., Eds.) pp 43-70, Ann Arbor Science, Ann Arbor.
83. Noe, G., Hofsteenge, J., Rovelli, G., & Stone, S.R. (1988) **J. Biol. Chem.** 263, 11729-11735.
84. Read, R.J., & James, M.N.G. (1986) in **Proteinase Inhibitors** (Barrett, A. & Salvesen, G., Eds.) pp 301-333, Elsevier, New York.
85. Peterson, T.E., Roberts, H.R., Sottrup-Jensen, L. & Magnusson, S. (1976) in **Protides of the Biological Fluids** (Peeters, H., Ed.) pp 145-149, Pergamon Press, Oxford.
86. Dodt, J., Kohler, S., & Baici, A. (1988) **FEBS Lett.** 229, 87-90.
87. Braun, P.J., Dennis, S., Hofsteenge, J., & Stone, S.R. (1988) **Biochemistry** 27, 6517-6522.
88. Stone, S.R., Dennis, S., & Hofsteenge, J. (1988) **Biochemistry** 28, 6857-6863.
89. Krstenansky, J.L., Owen, T.J., Yates, M.T. & Mao, S.J.T. (1987) **J. Med. Chem.** 30, 1688-1691.
90. Wallace, A., Dennis, S., Hofsteenge, J., & Stone, S.R. (1989) **Biochemistry** 28, 10079-10084.
91. Dodt, J., Kohler, S., Schmitz, T. & Wilhelm, B. (1990) **J. Biol. Chem.** 265, 713-718.
92. Dennis, S., Wallace, A., Hofsteenge, J. & Stone, S.R. (1990) **Eur. J. Biochem.** 188, 61-66.

93. Krstenansky, J.L. & Mao, S.J.T. (1987) **FEBS Lett.** 211, 10-16.
94. Maraganore, J.M., Chao, B., Joseph, M.L., Jablonski, J. & Ramachandran, K.L. (1989) **J. Biol. Chem.** 264, 8692-8698.
95. Mao, S.J.T., Yates, M.T., Owen, T.J. & Krstenansky, J.L. (1988) **Biochemistry** 27, 8170-8173.
96. Konno, S., Fenton II, J.W., & Villanueva, G.B. (1988) **Arch. Biochem. Biophys.** 267, 158-166.
97. Markwardt, F. (1986) **Ann. N. Y. Acad. Sci.** 485, 204-214.
98. Markwardt, F. (1985) **Biomed. Biochim. Acta** 44, 1007-1013.
99. Markwardt, F., Hauptmann, J., Nowak, G., Klessen, C., & Walsmann, P. (1982) **Thromb. Haemostasis** 47, 226-229.
100. Fenton II, J.W., Fasco, M.J., Stackrow, A.B., Aronson, D.L., Young, A.M., & Finlayson, J.S. (1977) **J. Biol. Chem.** 252, 3587-3598.
101. Bischoff, R., Clesse, D., Whitechurch, O., Lepage, P., & Roitsch, C. (1989) **J. Chrom.** 476, 245-255.
102. Ornstein, L. (1964) **Ann. N.Y. Acad. Sci.** 121, 321-349.
103. Davis, B.J. (1964) **Ann. N.Y. Acad. Sci.** 121, 404-427.
104. Hoefer Scientific Instruments catalogue (1986), pp 110-116, Hoefer Scientific Instruments, San Francisco.
105. Technical Bulletin No. MKR-137 (1986), pp 1-8, Sigma Chemical Company, St. Louis.
106. Bryan, J.K. (1977) **Anal. Biochem.** 78, 513-519.
107. Reid, B.R., Koch, G.L.E., Boulanger, Y., Hartley, B.S., & Blow, D.M. (1973) **J. Mol. Biol.** 80, 199-201.
108. Carter, C.W. Jr. & Carter, C.W. (1979) **J. Biol. Chem.** 254, 12219-12223.
109. Thaller, C., Weaver, L.H., Eichele, G., Wilson, E., Karlsson, R., & Jansonius, J.N. (1981) **J. Mol. Biol.** 147, 465-469.
110. Matthews, B.W. (1968) **J. Mol. Biol.** 33, 491-497.
111. Edmundson, A.B., Wood, M.K., Schiffer, M., Hardman, K.D., Ainsworth, C.F., Ely, K.R., & Deutsch, H.F. (1970) **J. Biol. Chem.** 245, 2763-2764.
112. Ichikawa, T. & Sundaralingam, M. (1972) **Nature New Biol.** 236, 174-176.

113. McPherson, A. (1982) in **Preparation and Analysis of Protein Crystals**, p. 241, John Wiley & Sons, New York.
114. Stout, G.H. & Jensen, L.H. (1989) in **X-ray Structure Determination - A Practical Guide**, pp. 23-25, John Wiley & Sons, New York.
115. Wyckoff, H.W., Doscher, M., Tsernoglou, D., Inagami, T., Johnson, L.N., Hardman, K.D., Allewell, N.M., Kelly, D.M., & Richards, F.M. (1967) **J. Mol. Biol.** 27, 563-578.
116. Wyckoff, H. (1985) **Methods Enzymol.** 114, 330-386.
117. Pflugrath, J.W. & Messerschmidt, A. (1989) **Munich Area Detector NE System User's Guide**, Enraf-Nonius, The Netherlands.
118. Pflugrath, J.W. & Messerschmidt, A. "Structured Software for Data Collection with Area Detectors", from **Crystallography in Molecular Biology Conference**, Bischofsberg, France, September 1985.
119. Sjolín, L. & Wlodawer, A. (1981) **Acta Crystallogr.** A37, 594-604.
120. **Data Collection Operation Manual** (1982), Section D.9 pg 5, Nicolet XRD Corp., Madison.
121. North, A.C.T., Phillips, D.C., & Mathews, F.S. (1968) **Acta Crystallogr.** A24, 351-359.
122. Steigemann, W. (1974) **Ph. D. Thesis**, Technische Universität München.
123. Messerschmidt, A., Schneider, M., & Huber, R. (1990) **J. Appl. Cryst.** 23, 436-439.
124. Rossmann, M.G. (1972) in **The Molecular Replacement Method** (Rossmann, M.G., Ed.) pp 1-42, Gordon and Breach, New York.
125. Huber, R. (1965) **Acta Crystallogr.** 19, 353-356.
126. Rossmann, M.G. & Blow, D.M. (1962) **Acta Crystallogr.** 15, 24-31.
127. Rydel, T.J., Ravichandran, K.G., Tulinsky, A., Bode, W., Huber, R., Roitsch, C., & Fenton II, J.W. (1990) **Science** 249, 277-280.
128. Patterson, A.L. (1935) **Z. Krist.** A90, 517-542.
129. Rossmann, M.G., Blow, D.M., Harding, M.M., & Collier, E. (1964) **Acta Crystallogr.** 17, 338-342.
130. Crowther, R.A. & Blow, D.M. (1967) **Acta Crystallogr.** 23, 544-548.
131. **International Tables for X-ray Crystallography** (1952) (Henry, N.F. & Lonsdale, K., Eds.) Vol. I, p 186, The Kynoch Press, Birmingham.

132. Huber, R. & Schneider, M. (1985) **J. Appl. Cryst.** 18, 165-169.
133. Sim, G.A. (1959) **Acta Crystallogr.** 12, 813-815.
134. Sim, G.A. (1960) **Acta Crystallogr.** 13, 511-512.
135. Stout, G.H. & Jensen, L.H. (1989) in **X-ray Structure Determination - A Practical Guide**, p 337, John Wiley & Sons, New York.
136. Blow, D.M. & Crick, F.H.C. (1959) **Acta Crystallogr.** 12, 794-802.
137. Blundell, T.L. & Johnson, L.N. (1976) in **Protein Crystallography**, p 419, Academic Press, New York.
138. Pflugrath, J.W., Saper, M.A., & Quirocho, F.A. (1984) in **Methods and Applications in Crystallographic Computing** (Hall, S. & Ashida, T., Eds.) pp 404-407, Oxford Univ. Press, Oxford.
139. Jones, T.A. (1978) **J. Appl. Cryst.** 11, 268-272.
140. Jack, A. & Levitt, M. (1978) **Acta Crystallogr.** A34, 931-935.
141. Levitt, M. (1974) **J. Mol. Biol.** 82, 393-420.
142. Deisenhofer, J., Remington, S.J., & Steigemann, W. (1985) **Methods Enzymol.** 115, 303-323.
143. Ten Eyck, L.F. (1973) **Acta Crystallogr.** A29, 183-191.
144. Ten Eyck, L.F. (1977) **Acta Crystallogr.** A33, 486-492.
145. Hendrickson, W.A. & Konnert, J.H. (1980) in **Computing in Crystallography** (Diamond, R., Ramaseshan, S. & Venkatesan, K., Eds.) pp 13.01-13.26, The Indian Academy of Sciences, Bangalore, India.
146. Hendrickson, W.A. (1985) **Methods Enzymol.** 115, 252-270.
147. Finzel, B.C. (1987) **J. Appl. Cryst.** 20, 53-55.
148. Luzzati, P.V. (1952) **Acta Crystallogr.** 5, 802-810.
149. Crawford, J.L., Lipscomb, W.N., & Schellman, C.G. (1973) **Proc. Natl. Acad. Sci. USA** 70, 538-542.
150. Tulinsky, A., Park, C.H., & Skrzypczak-Jankun, E. (1988) **J. Mol. Biol.** 202, 885-901.
151. Mulichak, A.M. & Tulinsky, A. **Blood Coagulation & Fibrinolysis**, manuscript in press.
152. Huber, R., Scholze, H., Paques, E., & Deisenhofer, J. (1980) **Hoppe-Seyler's Z. Physiol. Chem.** 361, 1389-1399.

153. Bode, W., Greyling, H.J., Huber, R., Otlewski, J., & Wilusz, T. (1989) **FEBS Lett.** 242, 285-292.
154. Arnott, S. & Dover, S.D. (1968) **Acta Crystallogr.** B24, 599-601.
155. Yonath, A. & Traub, W. (1969) **J. Mol. Biol.** 43, 461-477.
156. Pedersen, B. (1974) **Acta Crystallogr.** B30, 289-291.
157. Brown, I.D. (1976) **Acta Crystallogr.** A32, 24-31.
158. Taylor, R., Kennard, O., & Versichel, W. (1984) **Acta Crystallogr.** B40, 280-288.
159. Gorbitz, C.H. (1989) **Acta Crystallogr.** B45, 390-395.
160. Katz, B.A. & Kossiakoff, A. (1986) **J. Biol. Chem.** 261, 15480-15485.
161. Weiner, S.J., Kollman, P.A., Case, D., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. Jr., & Weiner, P. (1984) **J. Am. Chem. Soc.** 106, 765-784.
162. Ni, F., Konishi, Y., & Scheraga, H.A. (1990) **Biochemistry** 29, 4479-4489.
163. Huber, R. & Bode, W. (1978) **Acc. Chem. Res.** 11, 114-122.
164. Chang, J.-Y. (1985) **Eur. J. Biochem.** 151, 217-224.
165. Chang, J.-Y. (1986) **Biochem. J.** 240, 797-802.
166. Chang, J.-Y. (1989) **J. Biol. Chem.** 264, 7141-7146.
167. Bourdon, P., Fenton II, J.W., & Maraganore, J.M. (1990) **Biochemistry** 29, 6379-6384.
168. Skrzypczak-Jankun, E. & Tulinsky, A., unpublished results of this laboratory.
169. Zuck, V. & Owen, W., unpublished observations.
170. Degryse, E., unpublished observations.
171. Lee, B. & Richards, F.M. (1971) **J. Mol. Biol.** 55, 379-400.
172. Tulinsky, A., Vandlen, R.L., Morimoto, C.N., Mani, N.K., & Wright, L.H. (1973) **Biochemistry** 12, 4185-4192.
173. Tulinsky, A. (1980) in **Biomolecular Structure, Conformation, Function, and Evolution** (Srinivasan, R., Ed.) Vol. 1, pp. 183-199, Pergamon Press, Oxford.
174. Blevins, R.A. & Tulinsky, A. (1985) **J. Biol. Chem.** 260, 4264-4275.

MICHIGAN STATE UNIV. LIBRARIES



31293009011549