



3 1293 00914 1064

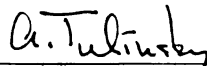
This is to certify that the
dissertation entitled

Probing the Nature of Thrombin Binding Sites
Using X-ray Crystallography
presented by

Xiayang Qiu

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Chemistry


Major professor

Date 1 / 11 / 93



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:\circ\date\due.pm3-p.1

**PROBING THE NATURE OF THROMBIN BINDING SITES
USING X-RAY CRYSTALLOGRAPHY**

By

Xiayang Qiu

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Chemistry

1992

ABSTRACT

PROBING THE NATURE OF THROMBIN BINDING SITES USING X-RAY CRYSTALLOGRAPHY

By

Xiayang Qiu

Thrombin is a serine proteinase that plays many critical roles in blood coagulation and wound healing processes, and has been known for a long time as a unique enzyme because of its extraordinary specificity and versatility. The study of the nature of thrombin binding sites will not only contribute to the understanding of its physiological behavior, but also should benefit the search for antithrombin drugs that could promise cures for thousands.

X-ray crystallographic methods were applied to obtain the three-dimensional structures of human α -thrombin-inhibitor complexes in order to reveal their detail interactions on the molecular level. Among the inhibitors that have been studied, three bind to the active site (cyclotheonamide A and PPACK, Figure 13; CV863-1, Table 3), one binds to the fibrinogen recognition exosite (MDL-28050, Table 3) and one binds to both the active site and exosite (hirulog 3, Table 3). Cyclotheonamide A is a macrocyclic peptide from a Japanese sponge that has an α -keto amide group that is capable of binding thrombin as a transition state analog. Synthetic peptide CV863-1 is one of the most potent small molecule antithrombotics with binding features of

Xiayang Qiu

cyclotheonamide A, while PPACK provides an additional model for the active site binding. The synthetic peptide MDL-28050-thrombin complex suggests that 55'-59' is the recognition sequence and shows that the binding of the C-terminal 3_{10} helical turn is flexible. Hirulog 3, a promising drug mimicking the bivalent nature of hirudin, reveals the nature of the S' subsites of thrombin for the first time, which are two cavities formed by the back-side of the S2 apolar site (His57, Tyr60A and Trp60D) and residues Glu39, Lys60F, Phe60H, Asn143, Gln151 and Glu192. Except for the S1 specificity pocket and a few positions within the exosite recognition sequence, tolerance for the imprecision of binding was observed as typical at subsites of thrombin, which is most likely the source of its unique specificity and versatility.

All five single-crystal structures were solved by isomorphous replacement methods, and were refined by restrained least-squares methods to excellent accuracy (typically ± 0.2 Å in position at 2.3 Å resolution with R-factors less than 0.14). The structures of thrombin in the various complexes showed very good agreement.

Acknowledgement

It has been a very special experience for me coming to the United States of America and pursuing my dream of becoming a good scientist. Four years in this Chemistry Department of Michigan State University, I am glad that I have had a very good opportunity to learn from Professor Alexander Tulinsky and to develop myself towards maturity. Therefore, I would like to express my deep gratitude here to Professor Tulinsky for his guidance and help, which were crucial for every success I might have achieved.

I would like to acknowledge here the people I have been pleasantly working with, especially Dr. Pushpa Padmanabhan, who provided me with precious crystals for my work in this thesis, and Dr. Pappan Padmanabhan, who gave me many advises and suggestions. I also owe a lot to other people in this laboratory whose work was cited in this thesis, especially Tim, Ewa, Vasili, and Mali. I thank Drs. Mathews and Ganesh for comments on this manuscript. In addition, I thank Drs. Maraganore, Krstenansky, Fenton, Brunck and Maryanoff for providing various samples.

In the end, I would like to thank my wonderful family, my parents and my sister, for their constant support and encouragement. It is always important for me to know that they are there to give me their best wishes. Without them, any achievement of mine would have been meaningless. Therefore, it is my honor and pleasure to dedicate this thesis to them.

Table of Contents

CHAPTER 1. INTRODUCTION	1
A. Blood Coagulation and Thrombin	1
B. Previous Structural Studies	14
C. Present Work	23
(1) Active site with cyclotheonamide A, CV863-1 and PPACK	23
(2) Fibrinogen recognition exosite with MDL-28050	27
(3) S' subsites with hirulog 3	28
CHAPTER 2. EXPERIMENTAL	31
A. Crystallization	31
(1) Sample information	31
(2) Preparation of the complexes	32
(3) Crystallization methods and conditions	34
B. Data Collection and Reduction	42
(1) Preparation	44
(2) Area Detector	46
(3) Frame acquisition	48
(4) Data reduction	52
C. Isomorphous Replacement	56
D. Model Building and PROLSQ Refinement	58
CHAPTER 3. RESULTS	72
A. The Cyclotheonamide A-Hirugen-Thrombin Complex	72
(1) General	72
(2) Overall structure	75
(3) Cyclotheonamide A-thrombin interaction	87
B. The CV863-1-Hirugen-Thrombin Complex	94

(1) General	94
(2) Overall Structure	96
(3) CV863-1-thrombin inter action	101
C. The PPACK-Thrombin Com plex	105
(1) General	105
(2) The structure	108
D. The MDL-28050-Thrombin Complex	115
(1) General	115
(2) Overall Structure	117
(3) MDL-28050-thrombin in teraction	126
E. The Hirulog 3-Thrombin Complex	128
(1) General	128
(2) Overall Structure	129
(3) Hirulog 3-thrombin inter action	135
CHAPTER 4. DISCUSSION AND ACCOMPLISHMENT	146
A. Active Site	146
B. Fibrinogen Binding Exosite	153
C. The S1'-S4' Subsites of Thrombin	159
D. Other Inhibitor Utilizing S' Subsites	163
E. Model of Thrombin Autolysis	167
F. Conclusion	170
LIST OF REFERENCES	173

CHAPTER 1. INTRODUCTION

A. Blood Coagulation and Thrombin

There are two principal mechanisms to stop the loss of blood in higher organisms following vascular injury. Platelets are activated and adhere to the site of injury to form a platelet plug that reduces or temporarily stops the loss of blood and the coagulation cascade is triggered that leads to fibrin formation and the generation of an insoluble fibrin clot that strengthens the platelet plug. After many years of in depth studies, the mechanism of blood coagulation is understood more and more [1]. The extrinsic pathway initiates the blood coagulation cascade while the intrinsic pathway is responsible for the growth and maintenance of fibrin formation (Figure 1). There are many proteins in human plasma that are involved in this process, most of them initially identified in patients with bleeding complications. The proteins circulate in blood in a precursor or inactive form and are activated by minor proteolysis in a series of stepwise reactions, many being converted to serine proteinases (Figure 1). In the final step, fibrinogen is cleaved by thrombin to form clottable fibrin.

Thrombin is a serine proteinase that plays a critical role in blood coagulation. It is generated by activating prothrombin in the "prothrombinase complex" [2] that consists of the substrate prothrombin, the proteolytic enzyme factor Xa, the membrane-bound cofactor, factor Va, calcium ions and a phospholipid surface (Figure 2). The products of this prothrombin activation reaction, shown in Figure 3, are prothrombin fragment 1, prothrombin fragment 2 and

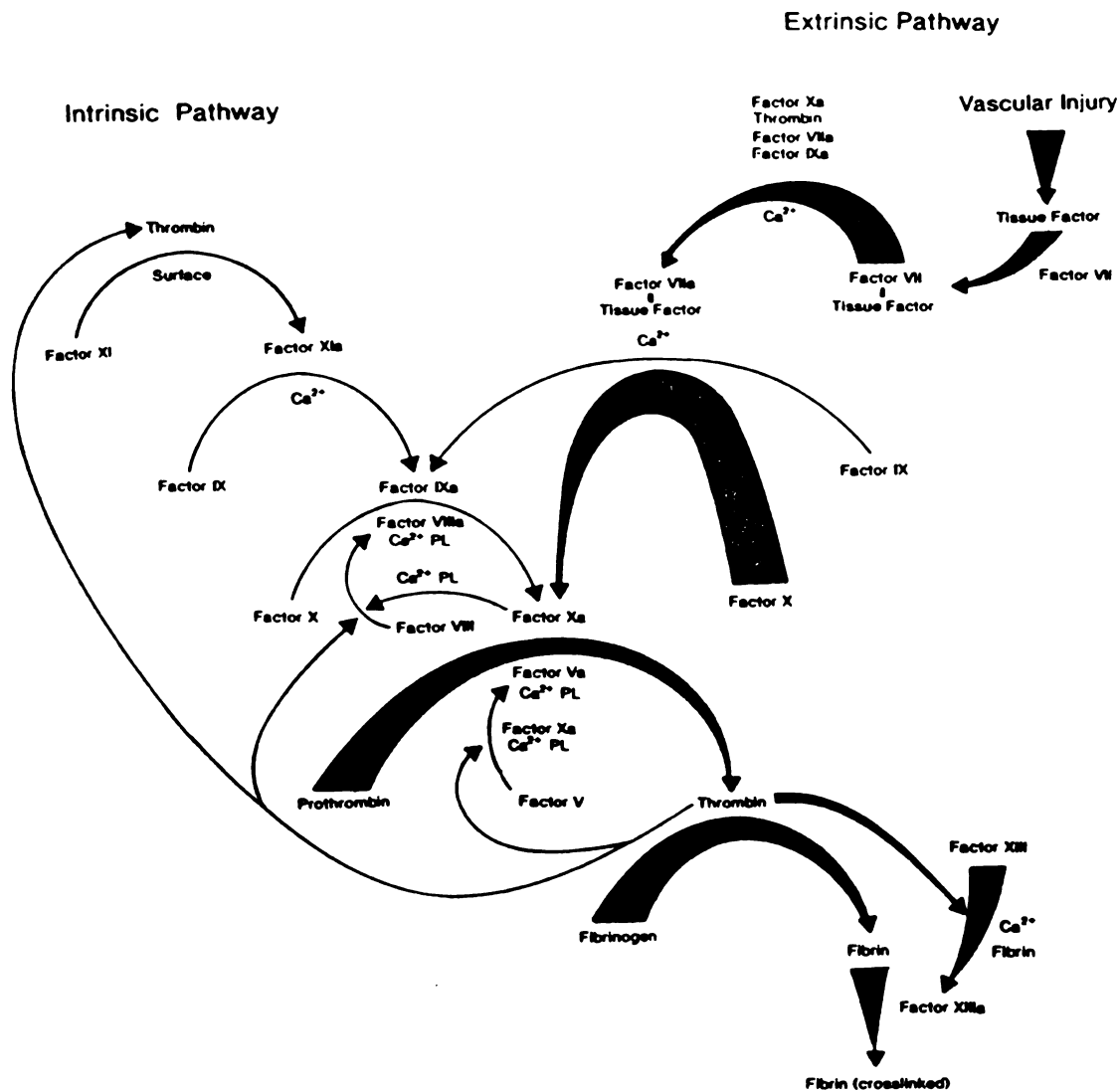


Figure 1. Coagulation cascade and fibrin formation by the intrinsic and extrinsic pathways. Vascular injury causes tissue factor to be exposed to blood, which triggers the extrinsic pathway (right side) shown in heavy arrows. The intrinsic pathway (left side) can be triggered when thrombin is generated. The activated clotting factors (except for thrombin) are designated by lowercase a, i.e., IXa, Xa, XIa, etc. PL refers to phospholipid. Taken from Davie et al. (1991) [1].

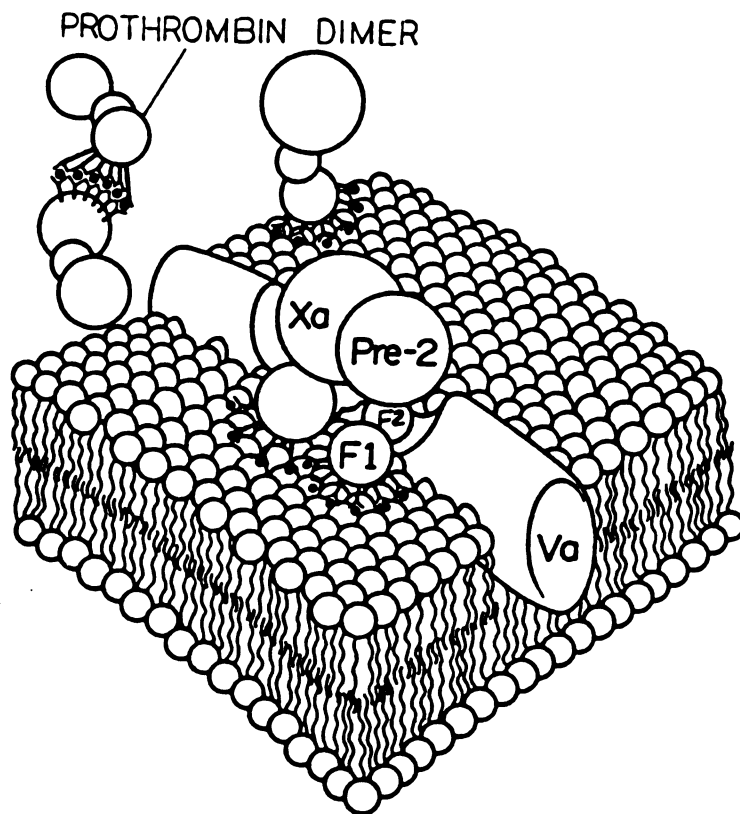


Figure 2. The "prothrombinase complex". Prothrombin is composed of fragment 1 (F1), fragment 2 (F2) and prethrombin 2 (Pre-2); Factor Va = Va; Factor Xa = Xa; calcium ions are represented by filled circles binding to γ -carboxyglutamic acid residues and the phospholipid surface. Taken from Rosenberg (1987) [2].

Prothrombin:

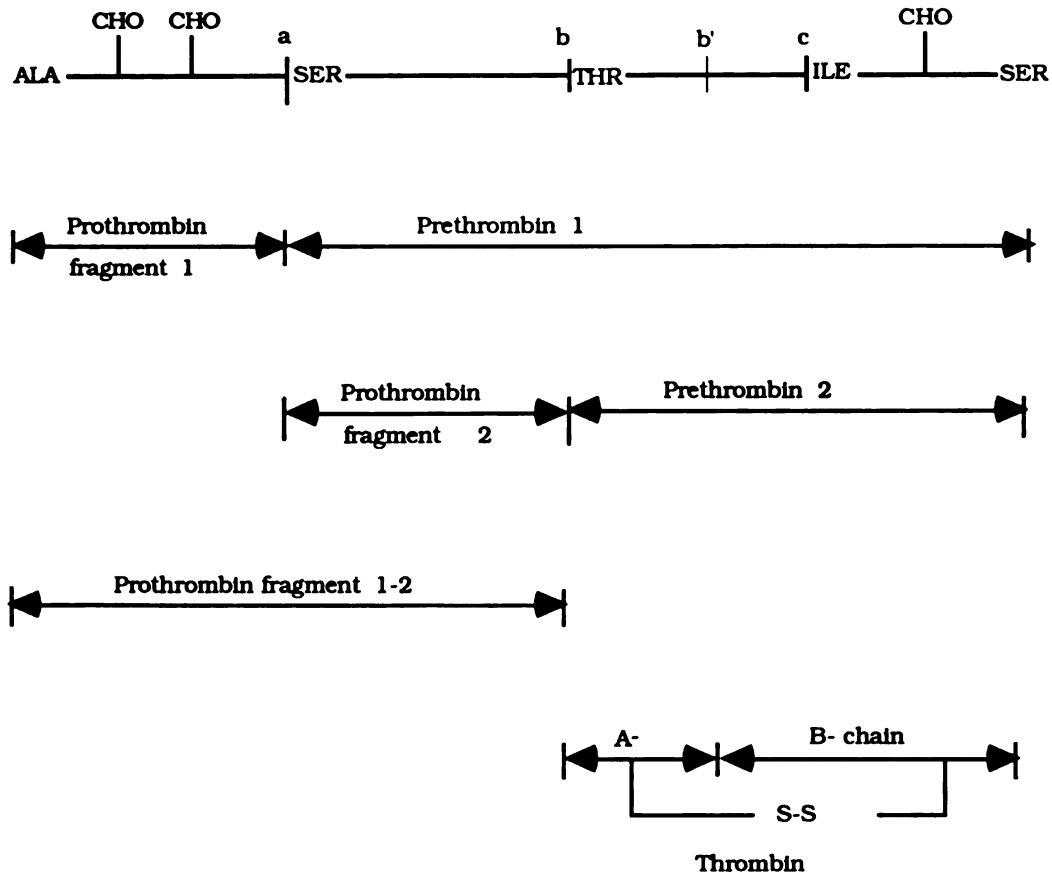


Figure 3. The products of prothrombin activation. CHO represents a carbohydrate side chain; a, b, b' and c are the cleavage sites of human prothrombin.

α -thrombin on cleavages at sites a, b, b' and c [3]. The cleavage at the c site produces the A chain and B chain of thrombin that are linked together by a disulfide bond.

Human α -thrombin, used in all the studies of this thesis, is a two chain enzyme that has a molecular weight of about 36,600 [4,5]. Its sequence, in chymotrypsinogen numbering, is listed in Table 1 [6]. There are 36 residues in the A chain and 259 residues in the B chain. Four disulfide bonds link the residues 1 to 122, 42 to 58, 168 to 182 and 191 to 220. Like all the other serine proteinases, human α -thrombin also has a catalytic triad consisting of Ser195, His57 and Asp102. Moreover, the overall sequence of the thrombin B chain, which encompasses the catalytic site, is also highly homologous to other serine proteinases, like trypsin, chymotrypsin, elastase, factor IXa, factor Xa, urokinase, tissue plasminogen activator, etc. [7]. A single chain carbohydrate is attached to the Asn60G residue of thrombin [8].

One of the major roles of thrombin is to cleave fibrinogen. Fibrinogen is a large dimeric molecule with three non-identical chains in each monomer (Figure 4). The α - (M.W. 66,000), β - (M.W. 52,000) and γ - (M.W. 46,500) polypeptide chains are held together by 29 inter- and intrachain disulfide bonds to form a trinodular structure of approximately 450 X 90 Å [1]. Thrombin catalyzes the removal of fibrinopeptide A (FPA) from the two α chains and fibrinopeptide B from the two β chains (Figure 4) to produce the fibrin monomer (Figure 5) and expose charged polymerization sites on the center domain of fibrin (referred to as the A and B sites). These are able to bind to oppositely charged regions on the

Table 1. The 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 chymotrypsin. A number followed by an alphabetic character represents an insertion. CP is a cis-proline.

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				

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	H70	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
T299	H230	V231	F232	R233	L234	K235	K236	W237	I238
Q239	K240	V241	I242	D243	Q244	F245	G246	E247	

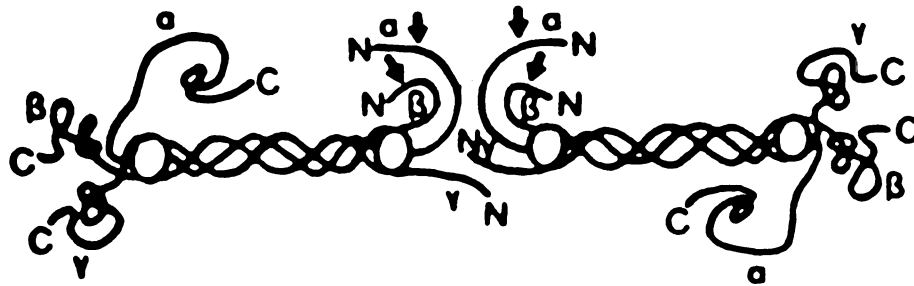


Figure 4. Schematic drawing of the fibrinogen molecule [1]. The molecule is a dimer with three chains (α , β , γ). The N-terminal of each chain is designated by N and the C-terminal of each chain is designated by C. Thrombin cleaves the N-terminal parts of the fibrinogen α and β chains (shown in arrows) to release fibrinopeptide A (FPA) and fibrinopeptide B (FPB).

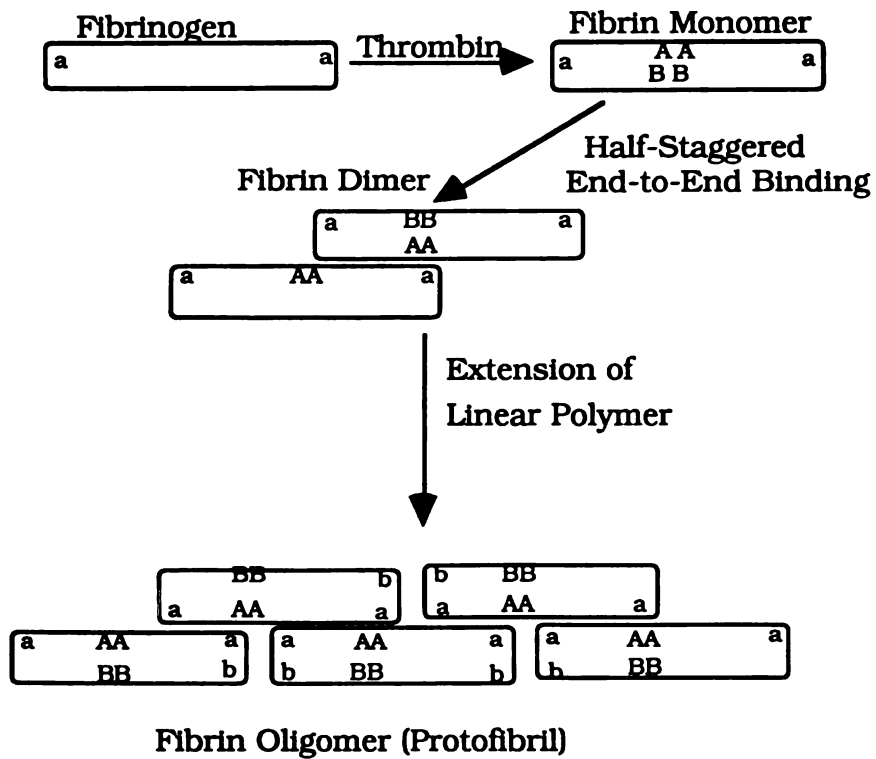


Figure 5. The formation of the fibrin oligomer (Protofibril). This is the earliest polymerization event leading to clot formation.

C-terminal domains (termed the a and b sites) of adjacent fibrinogen molecules, eventually producing an elongated, staggered fibrin polymer called the protofibril [9]. A more detailed analysis of the polymerization event indicates that FPB is not cleaved until the A site has made contact with the corresponding a-site. The B-b interaction may serve to stabilize the protofibril once the primary interaction has occurred [10]. The protofibrils will then be loosely linked to each other through a lateral attachment site in the a chain [11]. Finally, the entire clot is stabilized by factor XIIIa (fibrinoligase) to form covalent, end-to-end, antiparallel bonds between the carboxy-termini of adjacent γ chains and lateral attachments between adjacent chains so that the entire clot becomes a covalently cross-linked meshwork of fibrin [12].

Interestingly, thrombin also participates in the natural anticoagulant mechanism (Figure 6) [2]. In human plasma, antithrombin III is the major thrombin inhibitor even though there are a number of others (heparin cofactor II, α_1 -proteinase inhibitor and α_2 -macroglobulin) [13]. Thrombin can also bind to thrombomodulin. Protein C is converted to activated protein C by this thrombin-thrombomodulin complex. Activated protein C then complexes with protein S that is capable of inactivating factor Va and VIIIa that will eventually shut down the "prothrombinase complex" thrombin production and fibrin formation [14].

From Figure 1, thrombin is also responsible for the activation of other plasma proteins. From Figure 7, it should be clear why thrombin is called the central bioregulatory enzyme in hemostasis [15]. Thrombin can activate factor V and factor VIII through limited

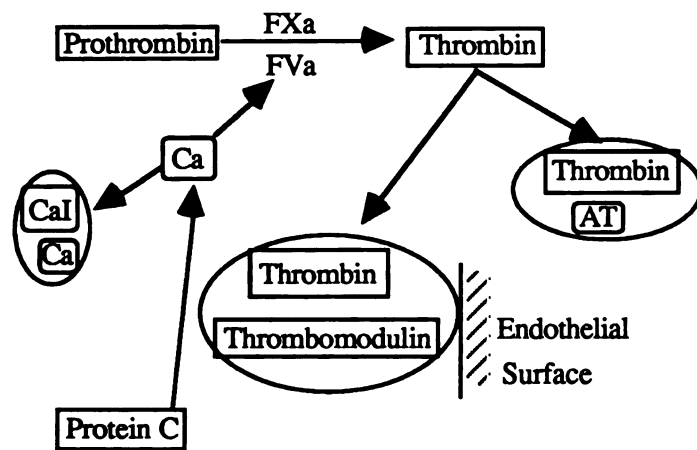


Figure 6. The natural anticoagulant mechanisms of the endothelium. AT represents antithrombin III; Ca indicates activated protein C; CaI is activated protein C inhibitor; complexes are circled.

Figure 7. The central bioregulatory functions of α -thrombin in hemostasis. Coagulation factor zymogens are indicated by roman numerals and their activated enzymes are followed by an a. Primes and double primes are used to indicate active and inactive forms of factor V and VIII. F1.2 is prothrombin fragment 1.2; FpA and FpB are fibrinopeptide A and B; PL is phospholipid and TPN is thromboplastin. Taken from Fenton (1986) [15].

proteolysis that results in its procoagulant function and thrombin activates factor XIII that leads to a stable cross-linked fibrin; thrombin is also involved in the activation of protein C that results in anticoagulant function. Moreover, thrombin is an agonist for a variety of cellular activities in many tissues and cell types [16]. Thrombin-cellular interactions have also been said to stimulate the aggregation of platelets, the chemotaxis of monocytes, the contractility of smooth muscle, the proliferation or cell growth of fibroblasts, macrophages and splenocytes, the prostaglandin synthesis of platelets, endothelial cells, fibroblasts and neuronal cells, the secretion of platelets and endothelial cells, the synthesis of prostacyclin, prostaglandin and platelet-activating factor on the blood vessel wall [17, 18] and the release of von Willebrand factor, tissue plasminogen activator (tPA), which is involved in fibrinolysis, and tPA inhibitor [19, 20]. Among these, thrombin induced platelet aggregation and secretion have an important direct effect upon coagulation and the role of thrombin as a growth factor in stimulating repair to tissue damage of the wound itself has been attracting much attention recently. The multiple functions of thrombin are summarized in Table 2 [21].

Unlike trypsin and other serine proteinases, which cleave almost all the arginyl and lysyl bonds of substrates, thrombin cleaves only four peptide bonds per fibrinogen molecule among the 376 available arginyl and lysyl bonds. Thus, thrombin has long been thought of as a highly specific enzyme. Nonetheless, thrombin is also highly versatile, as discussed earlier, in interacting with a large variety of different substrates. Therefore, the structural details of thrombin on the molecular level are very important, not only because

Table 2. The multiple functions of thrombin.

ENZYMATIC	Cleaves:	Influenced by	Function
	fibrinogen -> fibrin	---	coagulation
	factor V -> Va	---	cascade amplification
	factor VIII -> VIIIa	---	cascade amplification
	factor XI -> XIa	negative surface	cascade initiation
	Protein C -> PCa	thrombomodulin	cascade stoppage
	factor XIII -> XIIIa	Ca ²⁺ , fibrin	clot cross-linking
	platelet receptor	---	platelet activation
	Modulated by		
	thrombomodulin	---	cascade stoppage
	heparin	---	inhibition acceleration
	Inhibited by		
	fibrin	---	thrombin sequestration
	thrombomodulin	---	cascade stoppage
	α_2 -macroglobulin	---	inhibition
	antithrombin III	heparin	inhibition (serpin)
	heparin cofactor II	heparin	inhibition (serpin)
	protease nexin I	heparin	inhibition (serpin)
	hirudin	---	inhibition (from leech)
CELLULAR	induces platelet aggregation		
	stimulates platelet secretion		
	causes mitogenesis in lymphocytes		
	exhibits chemotactic properties		
	binds to endothelial cells and extracellularmatrix		
	plays a role in fibrin internalization		

of their physiological and medical implications, but also in their own right. X-ray crystallography is the most reliable technique for revealing these details, which is the subject matter of this thesis.

B. Previous Structural Studies

The first crystal structure of thrombin to be determined, human α -thrombin complexed with D-Phe-Pro-Arg-chloromethylketone (PPACK), was solved in 1989 at 1.9 Å resolution by Bode, et al. [6,22]; the crystallization conditions were almost identical to those reported earlier by our laboratory [23]. Soon after, Rydel et al. reported the structure of thrombin complexed with hirudin, a leech protein that is the most potent thrombin inhibitor [24,25]. These two structures provided many details of thrombin-inhibitor interaction and provided foundation for future structural studies of thrombin interactions. Other structures of thrombin complexed with different synthetic peptide inhibitors were reported later on [26-34], a few of them were completed by the author of this thesis [32-34] and some of which will be described here with detail.

The crystal structure of the PPACK-thrombin complex [22] showed that the overall folding of thrombin B chain is similar to that of chymotrypsin, except for the three insertion loops 60A-I, 149A-E and 186A-D. The Tyr60A-Trp60I insertion is not only bulky but also restricts accessibility to the active site while Ala149A-Lys149E forms a loose loop that accounts for autolytic and proteolytic susceptibility at Trp148, Ala149A and Lys149E by cathepsin, elastase and thrombin respectively to produce ζ -, ϵ - and γ -thrombin [35-37] (Figure 8). The structure also confirmed the position of the thrombin catalytic site and additionally revealed the manner in which a

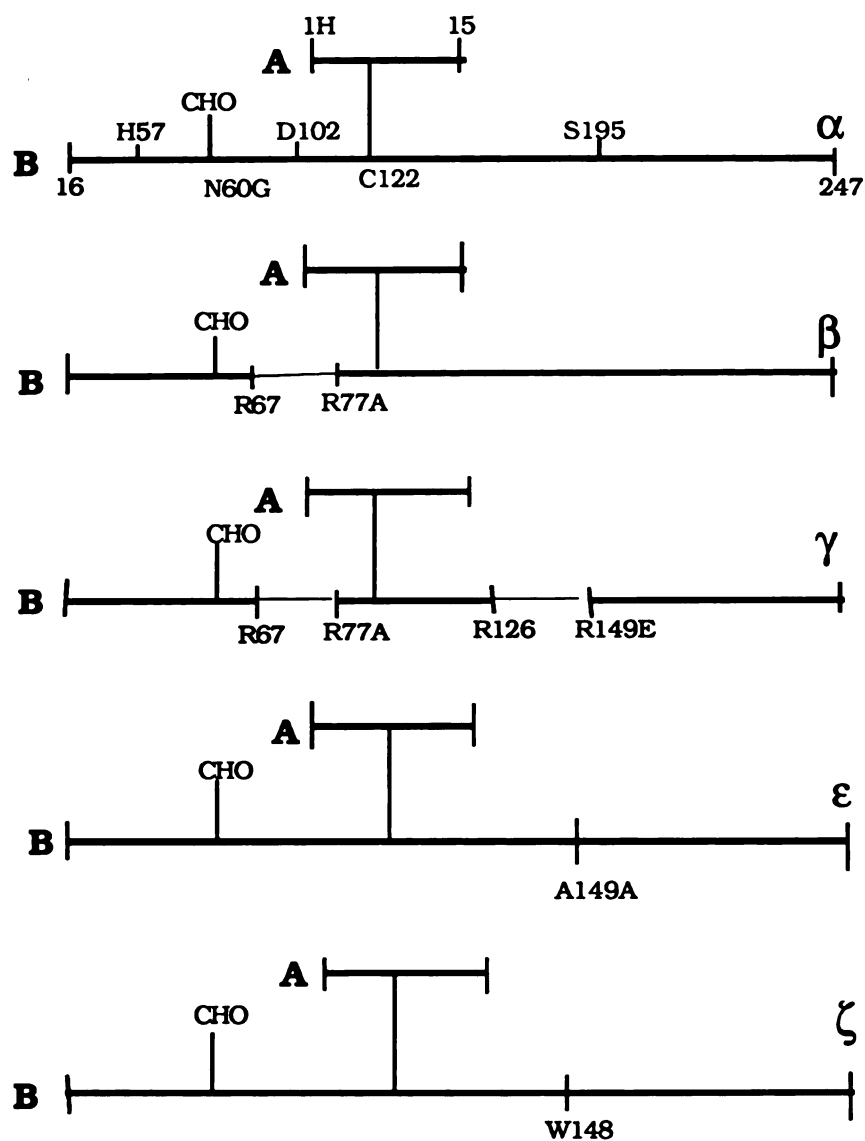


Figure 8. Human α -thrombin and its proteolytic derivatives. Capital A and B designate the A and B chain of thrombin. CHO represents the carbohydrate attached to Asn(N)60G. The catalytic residues are H57, D102 and S195. Excised cleavage sites are denoted.

substrate may interact with the active site. The inhibitor PPACK binds to thrombin covalently, similar to the base-catalyzed nucleophilic attack on the carbonyl carbon of substrate by the OG atom of Ser195 that has been observed in other serine proteinase crystal structures [38,39]. Other important details are (see Figure 9 also): (1) the chloromethylketone alkylates His57 and forms a hemiketal with Ser195 OG, both of the catalytic triad, (2) arginine binds in the S1 specificity site with its guanidinium group making a hydrogen bonding ion pair with Asp189 of thrombin, (3) the proline and D-Phe of PPACK occupy an extended S2 apolar binding site and (4) the peptidyl chain forms an antiparallel two strand β -sheet with Ser214-Gly216 of thrombin. The latter interaction is also similar to the primary interaction observed in all natural serine proteinase-protein inhibitor complexes.

Hirudin, produced in the salivary gland of the blood-sucking leech *Hirudo medicinalis*, prevents the clotting of ingested blood and is the most potent natural thrombin inhibitor with a K_d reported as low as 20 fM [40,41]. It is a 65 residue protein that forms a 1:1 noncovalent complex with thrombin. The crystal structure of human α -thrombin complexed with recombinant hirudin variant 2-lysine 47 (rHV2-K47) (Figure 10) shows that the 48 residue globular N-terminal, which is stabilized by three disulfide bridges, binds at the active site region. However, the binding is completely different than that of PPACK: (1) the amino terminus of hirudin forms a hydrogen bond with Ser195 OG of the catalytic site, (2) the S1 site of thrombin is unoccupied and (3) although Ile1' and Tyr3' (prime notation designating inhibitor throughout this thesis) are encapsulated in the

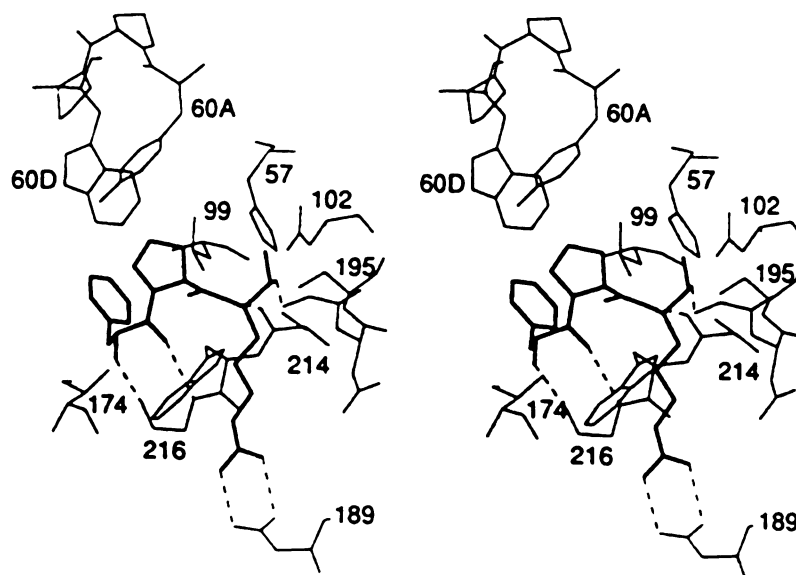


Figure 9. Stereoview of the interaction of PPACK with the active site of thrombin. PPACK shown in bold; hydrogen bonds broken; coordinates used were those of the present work (see Chapter 3, section C)

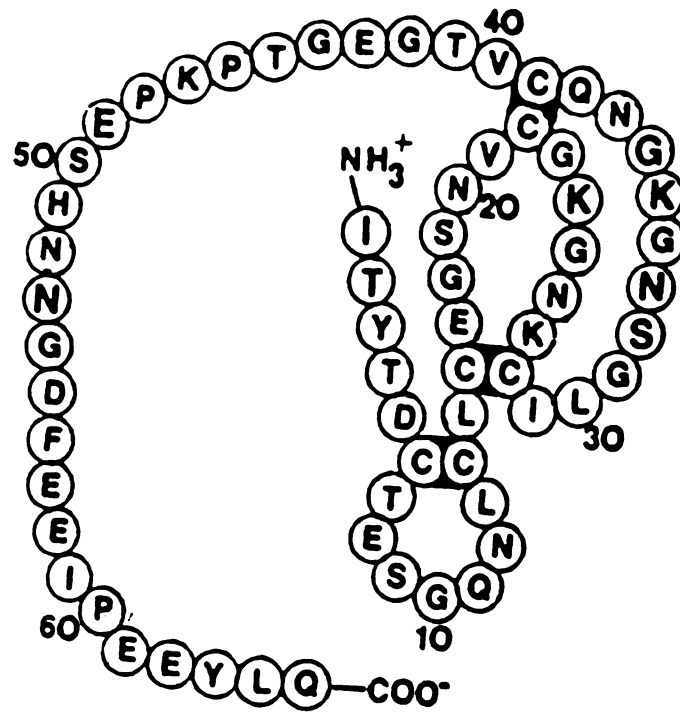


Figure 10. The sequence of recombinant hirudin variant 2-lysine 47 (rHV2-K47).

S2 apolar subsite and bind in the corresponding positions of D-Phe-Pro of PPACK, the Ile1'-Tyr3' of hirudin forms a parallel rather than antiparallel β -strand with Ser214-Gly219. The extended heptadecapeptide C-terminal of hirudin, which contains a high content of negatively charged amino acids, binds at the fibrinogen recognition "anion binding" exosite that is mainly composed of the Arg67-Glu80 loop of thrombin [24,25]. The thrombin-hirudin interactions are summarized schematically in Figure 11, where a putative heparin binding site is also indicated [42].

Hirudin and analogs of its functional domains have been growing in importance in the development of new antithrombotic drugs [43]. Hirugen, a peptide consisting of the 53'-64' sequence of hirudin, and hirulogs, bivalent thrombin inhibitors, are examples of peptide inhibitors designed to mimic thrombin-hirudin interactions [44-48]. The crystal structures of thrombin complexed with hirugen and hirulog 1, D-Phe-Pro-Arg-Pro-(Gly)₄-desulfato-Tyr63'-hirugen, have been reported earlier from this laboratory [27].

The role of the fibrinogen recognition exosite of thrombin has been studied extensively by examining the interaction of thrombin with synthetic hirudin C-terminal analogs [49-53]. Thrombin cleavage of fibrinogen and several other macromolecules can be inhibited effectively by blocking only the fibrinogen recognition exosite without inhibiting the amidolytic function of the thrombin active site for small chromogenic substrates [49,54]. Hirugen, the hirudin C-terminal analog, is an exosite inhibitor. Its K_d of 150 nM is decreased by about one order of magnitude upon sulfation of Tyr63' [41,51,52]. The three dimensional structure of the human

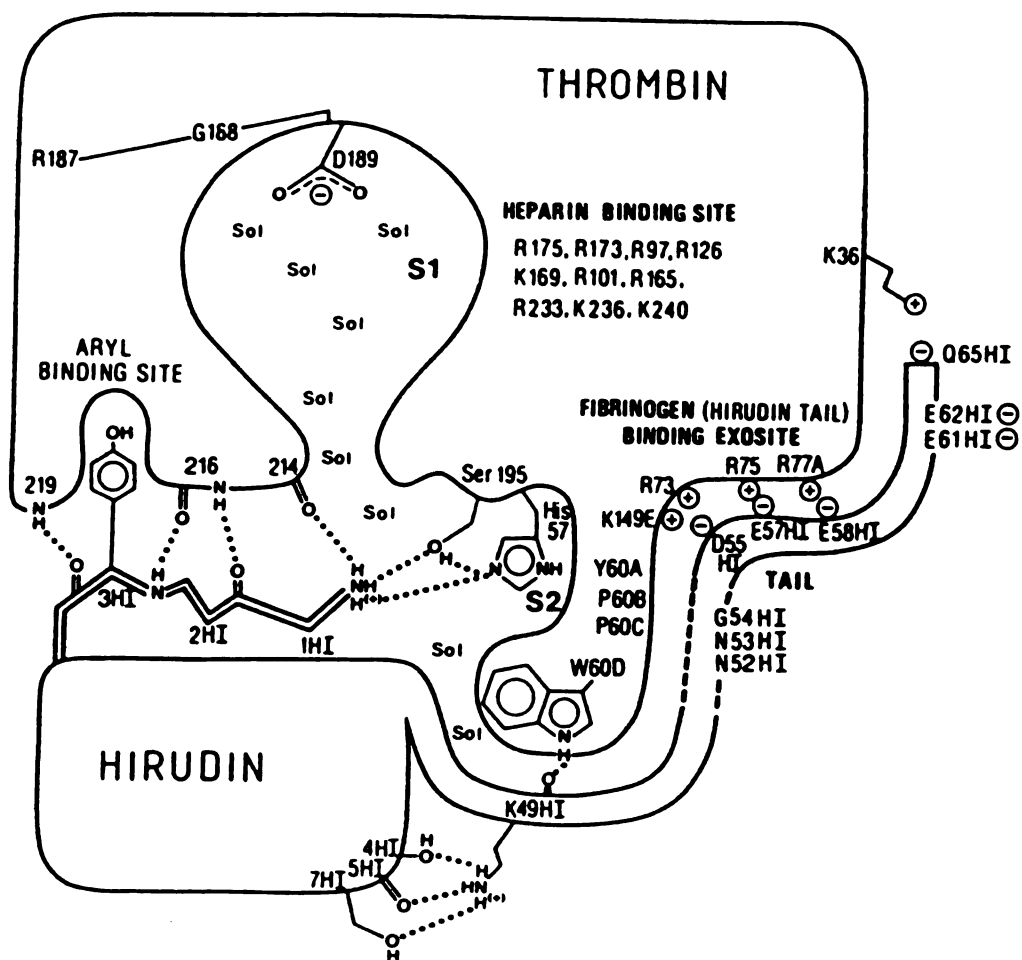


Figure 11. The canonical substrate-like inhibition mode of hirudin. Sol represents solvent molecules. Hirudin residues are designated by HI following the corresponding residue number. Taken from Bode and Huber (1992) [42].

α -thrombin-hirugen complex [27] has shown binding associations very similar to that of the C-terminal undecapeptide of hirudin, with sulfato Tyr63' of hirugen involved in an extended hydrogen bonding network utilizing the three sulfato oxygen atoms (Figure 12). The active site of thrombin in this structure is unoccupied and the positions of catalytic residues and the patterns of hydrogen bonding are essentially identical to those of other serine proteinases. Binding of hirudin or PPACK in the active site produces only a minor effect on the positions of the catalytic triad [27].

The thrombin inhibition by hirugen has been improved by mimicking the hirudin-thrombin interaction with the design of a class of bivalent peptides that can interact with both the active site and the anion binding exosite. Thus, hirulogs link a hirudin C-terminal peptide and an active site inhibitor together using a short oligo-glycine spacer and they show effective inhibition of thrombin at nanomolar concentrations, and they are by far the most efficient thrombin inhibitors outside more extensive natural peptides like hirudin [45,48]. The structure of the hirulog 1-thrombin complex, with an N-terminal D-Phe-Pro-Arg targeted to bind in the active site, has practically identical interactions with that of the PPACK-thrombin complex in the active site [22], and with that of the hirudin or hirugen-thrombin complexes at the exosite [27]. However, hirulog 1 is also a substrate so that the Arg3'-Pro4' bond is cleaved [54] although at a slow rate ($k_{\text{cat}} = 0.3 \text{ min}^{-1}$); moreover, the Pro4'-tetra-glycine spacer is disordered in the crystal structure of the complex so that the S1', S2', ... sites of thrombin (for substrates of serine proteinase, centering at the scissile bond, residues toward the

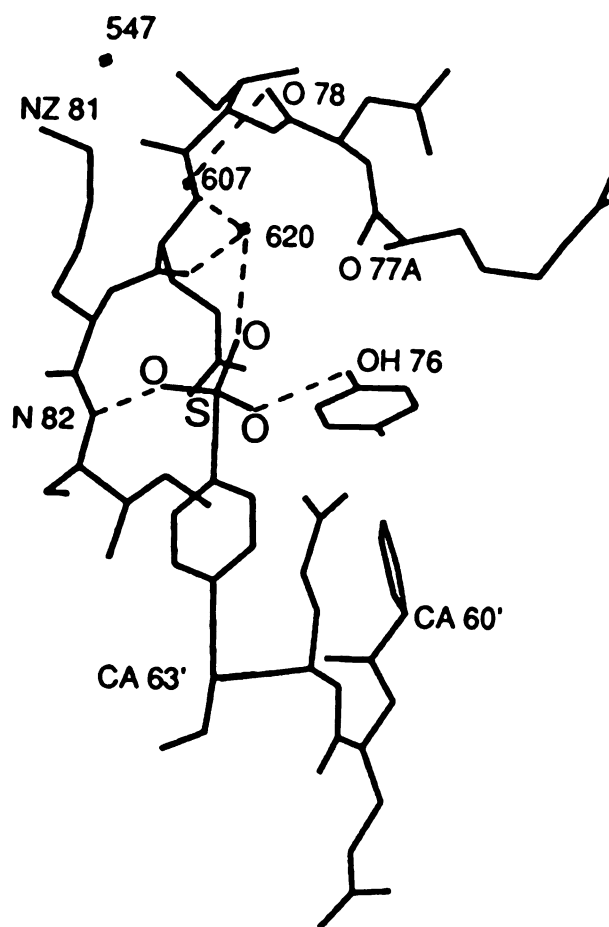


Figure 12. Hydrogen bond interaction of Tyr63' SO₃ of hirugen. Hydrogen bonds are drawn as broken lines. Water molecules 547, 607 and 620 are also shown. Taken from Skrzypczak-Jankun et al., (1991) [27].

N-terminal are named P1, P2, P3... and residues toward the C-terminal are named P1', P2', P3'... while the complementary subsites of the enzyme are named S1, S2', S3... and S1', S2' S3'...respectively) [58,39] could not be described. The electron density of the C-terminal of hirulog 1 (Pro60'-Leu64') was also not resolved, which precluded further elaboration of the role of Tyr63' desulfation on binding in this region.

C. Present Work

(1) Active site with cyclotheonamide A, CV863-1 and PPACK

Enzyme transition-state analogues have played a crucial role in the study of enzyme mechanisms and in the practice of drug design [59-61]. Recently, the macrocyclic immunosuppressants FK-506 and rapamycin (Figure 13a) have been identified as transition-state inhibitors of immunophilins, binding proteins with rotamase (peptidyl-prolyl cis-trans isomerase) activity [62-68]. The immunosuppressant ligand (Figure 13a) binds to its target immunophilin at the active site, with the α -keto amide group in a twisted conformation (dihedral angle of 95-100°) that mimics the transition state for interconversion between peptidyl-prolyl cis and trans amide rotamers [59,63]. Given the rarity of such macrocyclic α -keto amides in nature, the disclosure by Fusetani and co-workers of macrocyclic peptides that contain this α -keto amide group, namely cyclotheonamide A and B (Figure 13b), was particularly exciting to people in various fields [69-71].

Cyclotheonamide A (Figure 13b), a natural product from the Japanese marine sponge *Theonella* sp. (collected off Hachijo-jima Island, 300 km south of Tokyo, and the sponge was characterized by

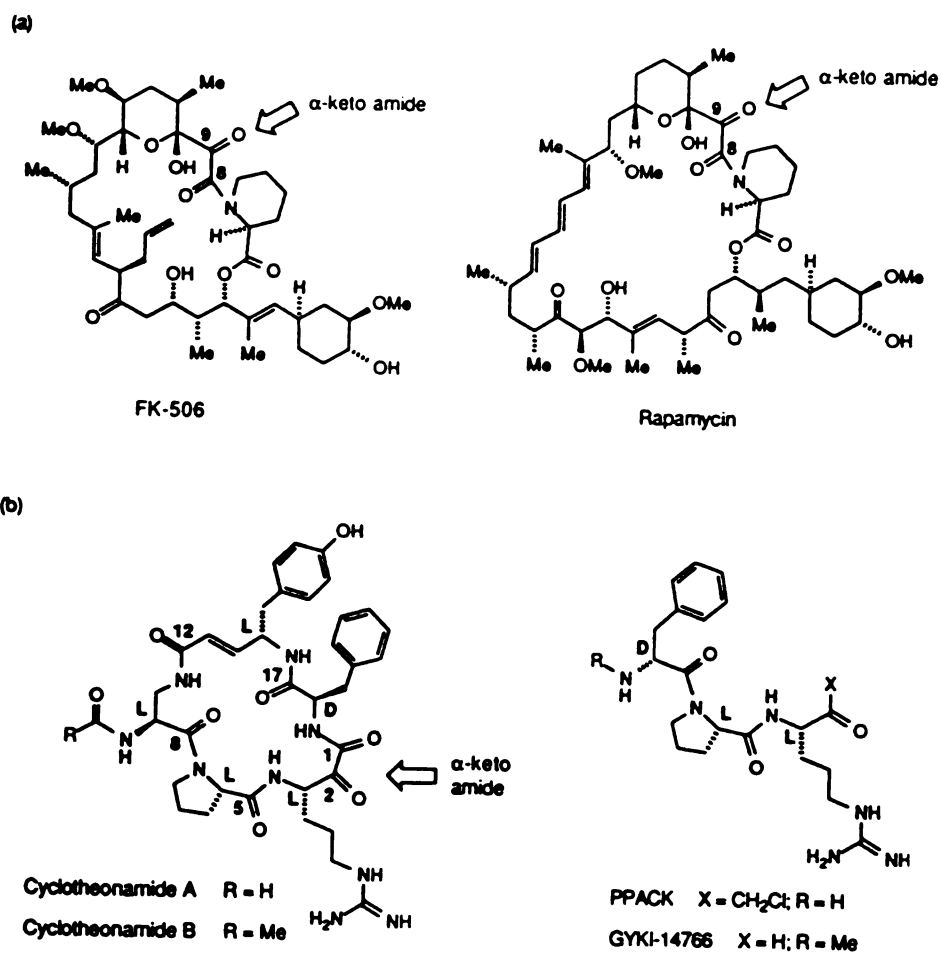


Figure 13. Cyclic peptide inhibitors. (a) FK-506 and rapamycin with the α -keto amide [C8-C9] denoted. (b) Cyclotheonamide A and B with ring numbering shown; and two reference thrombin inhibitors PPACK and GYKI-14766 (stereocenters labeled with descriptors).

a brilliant yellow inner body that differs from those of others), is reported to be a moderate potent ($K_d = 0.1$ mM) active site inhibitor of thrombin [69]. Interestingly, the Pro-Arg motif in cyclotheonamide A correlates to the P2-P1 positions in a class of thrombin inhibitors represented by PPACK and GYKI-14766 (Figure 13b); moreover the α -keto amide group of cyclotheonamide could be poised to act as a transition state analogue. For serine proteinase inhibition, as opposed to PPlase inhibition, this would entail interaction of the electrophilic ketone carbonyl with the hydroxy group of Ser195 to generate a tetrahedral intermediate [72], which may be comparable but different to previous transition state analog studies of other serine proteinases [73-75]. Therefore, the structure of the cyclotheonamide-thrombin complex would define the molecular basis for inhibition of human α -thrombin by cyclotheonamide A. Furthermore, cyclotheonamide A, a cyclic ligand for thrombin, might reveal new modes of molecular recognition in the thrombin active site cleft, particularly regarding the unique Tyr60A-Thr60I insertion loop. Although the Pro-Arg motif of cyclotheonamide A was expected to interact in a manner similar to PPACK [22], the cyclic array and the α -keto group offered intriguing possibilities for new interactions and novel design of antithrombotic drugs. The vinyl tyrosine residue in cyclotheonamide A (Figure 13b), which is not natural, shows that nature is creating kind of an "unnatural" product, lending confidence to organic chemists trying to design peptidyl mimics for different kinds of purposes.

A synthetic peptide, CV863-1, with a K_d of 0.5 nM, has the sequence of Boc-Asp-Pro-ketoArg-phenethyl amide, where Boc is a

(CH₃)₃-O-(CO)- group, ketoArg is an Arg*(C=O)-(C=O)- group (an normal arginine minus the C=O group, that is, decaboxy-arginine, is designated by Arg* throughout this thesis) and phenethyl amide is actually a decarboxy phenylalanine. It is by far the most potent small molecule inhibitor of thrombin. The CV863-1 peptide is designed to mimic cyclotheonamide A that also has a sequence of Pro-ketoArg-D-Phe. Furthermore, the sequence of Boc-Asp-Pro-ketoArg in CV863-1 is similar to the thrombin platelet receptor active site binding sequence of Leu-Asp-Pro-Arg [76]. Therefore, the determination of the structure of the CV863-1-thrombin complex would firmly establish a new class of thrombin inhibitory drugs derived from the marine sponge cyclotheonamide and in addition, would provide information about the thrombin-receptor interaction in the active site. Interestingly, test results showed that the thrombin inhibition by CV863-1 is more directed against receptor binding than against fibrinogen binding.

In addition, because a number of aspects of the original refined structure of the PPACK-thrombin complex [22] differ significantly from those of the other thrombin complex structures determined in this laboratory (resolution, mode of intensity data collection, and structure refinement, total number of solvent molecules, etc.) [25,27,32-34], the structure of PPACK-thrombin complex was re-determined and refined using the protocols of this laboratory to allow for more meaningful comparisons between comparably determined structures.

(2) Fibrinogen recognition exosite with MDL-28050

The C-terminal of hirudin binds to the non-catalytic fibrinogen recognition "anion binding" exosite of thrombin and crystal structures of thrombin with sulfato and desulfato forms of this peptide and its analogs have been extensively studied [24-33]. Leeches do not produce only a single form of hirudin, and thrombin does not only bind to hirudin but also to a large number of variants of hirudin; moreover, those that have been examined all appear to be highly potent much like hirudin [77]. In addition, there are a number of other natural macromolecular substrates that bind to thrombin at the exosite, such as fibrinogen [78], thrombomodulin [14] and heparin cofactor II [13]. While a hirudin C-terminal peptide with sulfato-Tyr63' (hirugen) exhibits a well-defined helical turn structure with the sulfato oxygens hydrogen bonded with thrombin residues (Figure 12) [27], non-sulfated analogs have shown only poor electron density for the helical turn (hirudin, hirulog 1) [24,27], implying a significant hydrophilic contribution to the binding of this region of hirudin. The focus of the present study of the thrombin exosite was to examine hirudin-based peptides of significantly different sequence that bind to the exosite and the manner in which they can be accommodated by the same site. A non-native peptide sequence derived from extensive structure activity studies on the hirudin54'-65' region is MDL-28050 that was designed to have good potency without the presence of the sulfato-Tyr63' residue [79]. It is selected using single amino acid replacement that starts with the hirudin sequence and preceeds by changing each amino acid, one at a time, to test the binding constants and decide which binds best. The

structure of MDL-28050, together with the that of the hirullin P18 peptide (also called hirudin P18, derived from a different species of leech *Hirudinaria manillensis*, which has the C-terminal sequence most divergent from that of hirudin) [80], which has also been determined in our laboratory, provides results and compelling experimental evidence that elaborate on the dogma of the nature of the anion binding exosite.

(3) S' subsite with hirulog 3

Hirulog 3 is a synthetic peptide inhibitor in the hirulog class (hirulog 1 was discussed above, hirulog 2 has fibrinopeptide A 7-16 attached to hirugen through a prolyl-tetra-glycine spacer) [47]. Its sequence, designed as D-Phe-Pro- β -homoarginine-(Gly)₅-desulfato-Tyr₆₃'-hirugen, has a β -homoArg at position 3'. Although an amide, the β -homoArg-Gly bond is not a true peptide bond because of the methylene group insertion between the C α atom and carbonyl carbon (equivalent to β position); however, this makes it proteolytically stable. The high resolution refined structure of hirulog 3 complexed with human α -thrombin reveals the nature of the S1', S2' and S3' subsites of thrombin that have heretofore been unreported but have been shown to be important for binding to serine proteinases in general [39,58]. Therefore, the structure of hirulog 3 represents an important thrust in the direction of better design of hirulog analog drugs. In addition, the C-terminal penta-peptide is ordered in a helical turn, and Asn₅₃'-Gly₅₄' are disordered as in the hirudin, hirugen and hirulog 1 structures; the latter is most likely related to the susceptibility of asparagine to deamidate in an Asn-Gly link [81-

83]. A number of the thrombin inhibitor peptides that have been mentioned above are listed in Table 3.

Table 3. Nature and synthetic peptide inhibitors of thrombin

Hirudin ^{53'-65'}	N G D F E E I P E E Y L Q
Hirugen	N G D F E E I P E E Y L
Hirullin ^{50'-62'}	S D F E E F S L D D I E Q
MDL-28050	Suc Y E P I P E E A Cha (D-E)
FPA(6'-16')	G D F L A E G G G V R
Hirulog 1	(D-F) P R P G G G G -Hirugen*
Hirulog 2	(FPA) P G G G G -Hirugen*
Hirulog 3	(D-F) P (h-R) P G G G G -Hirugen*
CV863-1	Boc D P (k-R) F*
Receptor ^{37'-46'}	T L D P R S F L L R

Y' = Sulfated tyrosine

Suc = Desamino aspartic acid, or succyl-

Cha = β -Cyclohexylalanine

(D-) = d-Enantimorph

FPA = Fibrinopeptide A

h-R = β -Homoarginine, or Arg^{*}-(CH₂)-(CO)-

Boc = (CH₃)₃-O-(CO)-

k-R = α -Keto-arginine, or Arg^{*}-(CO)-(CO)-

F* = Phe*, phenethyl amide, or decarboxyl phenylalanine

Hirugen* = desulfato-Tyr^{63'}-Hirugen.

CHAPTER 2. EXPERIMENTAL

A. Crystallization

(1) Sample information

All the human α -thrombin samples used in this laboratory, including those used in the projects of this thesis, were provided as generous gifts by Dr. John W. Fenton II of New York State Department of Health, Albany, New York as part of a continuing collaboration with our laboratory. The preparation method was described earlier by Lundblad et al. [84] and Fenton et al. [85]. As mentioned before, thrombin samples are very unstable at room temperature because they are vulnerable to autolysis and proteolysis. Therefore, native α -thrombin samples were treated with great care and kept deeply frozen (-80°C). A typical human α -thrombin sample comes in dry ice as 1 mL frozen vials containing 1.48 mg/mL thrombin in 0.75 M NaCl aqueous solution.

The cyclotheonamide A was provided by Dr. Bruce E. Maryanoff of the R. W. Johnson Pharmaceutical Research Institute, Spring House, Pennsylvania. Although the total chemical synthesis of cyclotheonamide A had been achieved [34], the material used in the present crystal structure study was the natural extract from the Japanese marine sponge originally obtained by Professor Nobuhiro Fusetani and co-workers of University of Tokyo, Tokyo, Japan [69]. The sample came as a dry powder and was reported as free of impurities such as cyclotheonamide B [69].

The CV863-1 sample was provided by Dr. Terence K. Brunck of Corvas International, Inc., San Diego, California. It was synthesized at

Corvas and was sent to us as a powder.

PPACK was purchased from Calbiochem and used as received. The properties and preparation methods of PPACK were published originally by Kettner and Shaw [86].

The MDL-28050 compound was supplied by Dr. John L. Krstenansky of Merrell Dow Research Institute, Cincinnati, Ohio. It was synthesized [79] by solid-phase techniques using an Applied Biosystems (Applied Biosystem, Inc., Foster City, California) Model 430-A peptide synthesizer [87] and purified by preparative HPLC (High Performance Liquid Chromatography).

Hirulog 3 and hirugen samples were provided Dr. John M. Maraganore of Biogen, Cambridge, Massachusetts. Hirugen was synthesized by standard solid-phase techniques while hirulog 3 required a few more chemical reaction steps [47]. The samples came as aqueous solutions.

(2) Preparation of the complexes

There are a few precautions that must be applied when complex preparation is undertaken. First, thrombin is very unstable so that it should be inhibited as soon as the frozen sample melts. Second, the original concentration of thrombin samples is 1.48 mg/mL, compared to 4 to 10 mg/mL as the usual protein concentration used in most of the crystallization experiments; thus the thrombin solution has to be concentrated. Moreover, there is a large amount of NaCl (0.75 M) in the original thrombin sample that is considered to be helpful for stabilizing thrombin [85], but is a deterrent in our search for crystallization conditions because NaCl itself can behave as

a precipitant. Therefore, NaCl must be diluted to a suitable level for crystallization of thrombin complexes.

In a typical experiment, a thrombin inhibitor (or two inhibitors in cases of ternary complexes) was added over a frozen 1 mL thrombin sample while it was thawing on ice or in a 4 °C cold room. Then, the solution was diluted to 2 mL using 0.1 M sodium phosphate buffer (pH 7.3) and allowed to stand approximately two hours to allow for the completion of the reaction. A Centricon-10 miniconcentrator, which has a membrane to allow molecules of 10 kDa or smaller to pass through, concentrates the protein complex to an ideal concentration and, in the mean time, keeps the diluted concentration of NaCl unchanged.

In the case of cyclotheonamide A, a ten-fold molar excess of the inhibitor and hirugen (by experience, hirugen was added to ensure a better quality of the crystal, since it does not interfere with the cyclotheonamide A-thrombin interaction nor induces any notable conformational change [27]) were added over a frozen 1 mL sample of human α -thrombin solution (1.48 mg/mL) at 4 °C. This was then diluted to 2 mL with 0.1 M sodium phosphate buffer (pH 7.3). The solution of the complex was concentrated to approximately 6 mg/mL using a Centricon-10 miniconcentrator in a refrigerated centrifuge.

For CV863-1, a ten-fold molar excess of the inhibitor and hirugen were added over a frozen 1 mL sample of human α -thrombin solution (1.48 mg/mL) at 4 °C. This was then diluted to 2 mL with 0.1 M sodium phosphate buffer (pH 7.3). The solution of the complex was concentrated to approximately 6 mg/mL using a Centricon-10 miniconcentrator in a refrigerated centrifuge.

In preparing the PPACK complex, a ten-fold molar excess of the inhibitor was added over a frozen 1 mL sample of human α -thrombin solution (1.48 mg/mL) at 4 °C. This was then diluted to 2 mL with 0.1 M sodium phosphate buffer (pH 7.3). The solution of the complex was concentrated to approximately 8 mg/mL using a Centricon-10 miniconcentrator in a refrigerated centrifuge.

For the MDL-28050 complex, a ten-fold molar excess of the inhibitor was added over a frozen 1 mL sample of human α -thrombin solution (1.48 mg/mL) at 4 °C. This was then diluted to 2 mL with 0.1 M sodium phosphate buffer (pH 7.3). The solution of the complex was concentrated to about 6 mg/mL using a Centricon-10 miniconcentrator in a refrigerated centrifuge.

In the case of hirulog 3, a ten-fold molar excess of the inhibitor was added over a frozen 1 mL sample of human α -thrombin solution (1.48 mg/mL) at 4 °C. This was then diluted to 2 mL with 0.1 M sodium phosphate buffer (pH 7.3). The solution of the complex was concentrated to approximately 5 mg/mL using a Centricon-10 miniconcentrator in a refrigerated centrifuge.

(3) Crystallization methods and conditions

Good quality X-ray diffracting single crystals are essential for structure determination. We are fortunate that Dr. Pushpa Padmanabhan was able to grow most of the crystals in our laboratory, including all the crystals used in the studies of this thesis. The methods, used in the crystallization experiments for this thesis, are hanging drop method, sitting drop method, repetitive seeding and incomplete factorial search.

The hanging drop method, illustrated in Figure 14, is the most commonly used method in protein crystallization experiments [88]. Implementing the traditional vapor diffusion technique on a microscopic scale, a hanging drop of protein-containing buffer solution (usually 10-20 μL) is placed over a reservoir of solution (usually 1 mL in volume) that contains a higher concentration (usually slightly smaller than the one that precipitates the protein) of precipitants (usually inorganic salts, organic solvents or polyethylene glycol PEG) in a sealed chamber (a cover glass sealed over a well of a tissue culture linbro plate is used in our laboratory). The solvent in the drop is gradually transferred through the vapor phase to the more concentrated reservoir solution, the drop is slowly concentrated and the protein sample can crystallize over a period of days, weeks, months or even years. The method is very inexpensive to set up and requires very small amount of protein.

An alternative way of utilizing vapor diffusion is the sitting drop method. It uses a similar setup, only the drop is sitting in a small container (much smaller than the reservoir) that is placed inside the sealed chamber. This way, a larger drop can be used to grow bigger crystals; sometimes proteins that crystallize in sitting drop do not crystallize in a hanging drop, which might be due to the different gravitational effect on the drops.

Sometimes, seeds or small crystals will not grow bigger through refining the crystallization conditions while the repetitive seeding method could be very effective for the enlargement of small crystals. These crystalline seeds can be sucked out of the drop with a thin glass capillary and moved to a fresh drop of mother liquor,

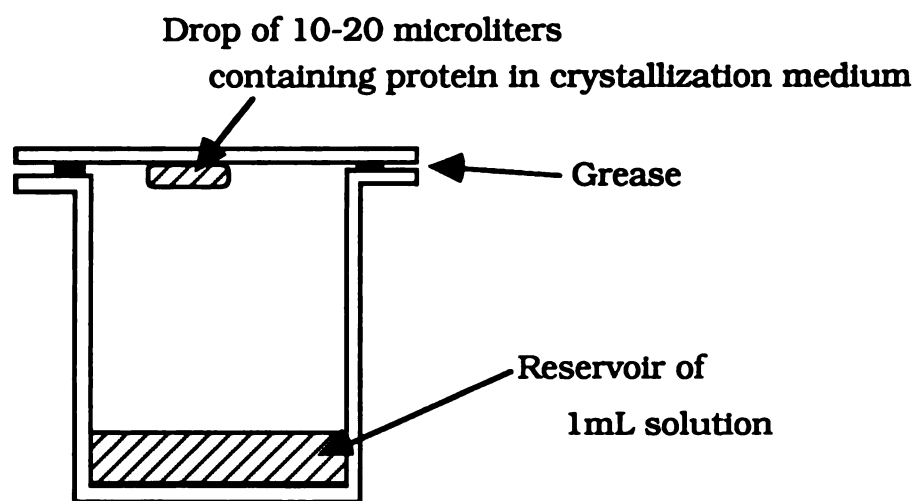


Figure 14. The hanging-drop method for protein crystallization.

which washes the crystal until it is clean and free of other seeds. When the seed is slowly dissolving and providing a fresh crystalline surface, it can then be moved to another protein-containing drop and set up as hanging or sitting drop (under same conditions, or with a slightly lower concentration of precipitants). Sometimes, the seed crystal will grow bigger but still be small for X-ray diffraction experiments, so that the procedure will have to be applied repetitively until the crystal grows big enough.

Protein crystallization is not an easy task because the crystallization condition is very sensitive to the ionic strength of the solution, the chemical composition and pH of the buffer, temperature, precipitants, etc. The search for a good condition usually might take months, years or longer. The incomplete factorial search [89-91] has proven to be a very helpful way of screening possible conditions. It is a set of conditions for initial crystallization experiments; the conditions are chosen by systematic assignment of variables in accordance with principles of randomization and balance [89]. The method allows us to screen a wide range of conditions using the least number of statistically effective trials. Refining the search will then be carried out near the conditions that produce crystals, microcrystals or precipitates to obtain bigger and better single crystals. The set of factorial solutions used in this laboratory is listed in Table 4.

For the human α -thrombin-cyclotheonamide A-hirugen complex, seed crystals were obtained by using 30% PEG 4K, 0.2 M sodium acetate in 0.1 M Tris buffer (pH 8.5) in a reservoir, over which was suspended a hanging-drop of 2 μ L of the ternary complex diluted by

Table 4. The factorial conditions for initial crystallization search

No.	Salt	Buffer	Precipitant
1	0.2M CaCl ₂	0.1M Acetate	30% MPD
2	1.0M Na,K Tart.	0.1M ADA	---
3	---	---	0.4M (NH ₄) ₃ PO ₄
4	---	0.1M Tris	3.0M (NH ₄) ₂ SO ₄
5	0.2M Na Citrate	0.1M Hepes	30% MPD
6	0.2M MgCl ₂	0.1M Acetate	30% PEG 4K
7	1.2M Na Citrate	0.1M Hepes	---
8	0.2M Na Citrate	Adj. to pH5.5	2.0M (NH ₄) ₂ SO ₄
9	0.2M NH ₄ Ac	0.1M Citrate	30% PEG 400
10	---	0.1M Acetate	1.5M (NH ₄) ₃ PO ₄
11	0.2M (NH ₄) ₂ SO ₄	0.1M Hepes	2.0M Na,K Phst.
12	0.2M Na Citrate	0.1M Tris	20% PEG 400
13	0.2M CaCl ₂	0.1M Hepes	25% PEG 4K
14	0.1M MgCl ₂	0.1M ADA	30% PEG 8K
15	0.2M Li ₂ SO ₄	0.1M Citrate	30% PEG 4K
16	1.0M Li ₂ SO ₄	0.2M Acetate	---
17	0.2M (NH ₄) ₃ PO ₄	0.1M Tris	30% MPD
18	0.2M NH ₄ Ac	0.1M Tris	2.0M Na,K Phst.
19	0.1M (NH ₄) ₂ SO ₄	0.1M Citrate	30% PEG 8K
20	---	0.1M ADA	30% MPD
21	0.2M MgCl ₂	0.1M Hepes	30% PEG 400
22	0.2M NaAc	0.1M Tris	30% PEG 4K
23	0.2M NaAc	0.1M Acetate	1.0M Na,K Phst.
24	0.2M CaCl ₂	0.1M Tris	---
25	0.5M NH ₄ Ac	0.1M Citrate	30% MPD
26	2.0M NaAc	0.1M ADA	---
27	0.2M Na,K Tart.	0.1M ADA	30% PEG 8K
28	1.0M Na,K Tart.	0.1M Hepes	---
29	0.2M (NH ₄) ₂ SO ₄	0.1M Acetate	30% PEG 400
30	0.1M (NH ₄) ₂ SO ₄	0.1M Hepes	20% PEG 4K
31	2.0M (NH ₄) ₂ SO ₄	0.1M ADA	---
32	0.2M Na Citrate	0.1M ADA	30% EtOH
33	0.2M MgCl ₂	0.1M Hepes	30% EtOH
34	0.2M NH ₄ Ac	0.1M Tris	30% EtOH
35	0.2M CaCl ₂	0.1M Acetate	30% EtOH
36	0.2M NaAc	0.1M Hepes	30% EtOH

*Tart. = Tartarate, Ac = CH₃COO⁻, Phst. = Phosphate, Et = C₂H₅, PEG = polyethylene glycol, MPD= 2-methyl-2,4-pentanediol, ADA = N-[2-Aceamidol]-2-iminodiacetic acid, Hepes = N-[2-Hydroxyethyl] piperazine-N'-[2-ethane]sulfonic acid.

2 μ L of the reservoir solution. The crystal used for intensity data collection, 0.6 mm X 0.3 mm X 0.2 mm in size, was obtained by enlarging a small crystal using repetitive seeding, employing the vapor-diffusion sitting drop method (other conditions were the same). A photograph of the crystals is shown in Figure 15.

Crystals of the human α -thrombin-CV863-1-hirugen complex were grown using the hanging-drop method with 3 mg/mL of protein complex, 0.13 M sodium phosphate buffer (pH7.3), 0.24 M NaCl, 13% PEG 4K and 1mM NaN_3 in 10 μ L drops equilibrating against 1mL well solution of 0.2 M sodium phosphate buffer (pH7.3), 26% PEG 4K and 1 mM NaN_3 . The crystal used for intensity data collection, 0.6 mm X 0.5 mm X 0.3 mm in size, grew in about two weeks and has the same appearance as those in Figure 15.

In the case of the human α -thrombin-PPACK complex, crystals were grown using the hanging-drop method with 4 mg/mL of protein complex, 0.075M sodium phosphate buffer (pH7.3), 0.24 M NaCl, 12% PEG 4K and 1mM NaN_3 in 10 μ L drops equilibrating against 1mL well solution of 0.1M sodium phosphate buffer (pH7.3), 24% PEG 4K and 1 mM NaN_3 . These conditions are practically identical to what was reported earlier by this laboratory [23], except for a difference in the crystallization methods (dialysis method was used in the report). The crystal used for intensity data collection, 0.8 mm X 0.4 mm X 0.4 mm in size, grew in about two weeks. The morphology of the crystals is also identical to those reported, which is shown in Figure 16 [23].

For growing the crystals of the human α -thrombin-MDL-28050 complex, hanging-drop methods were also used with 3 mg/mL



Figure 15. Crystals of human α -thrombin-cyclotheonamide A-hirugen complex. The size of the biggest crystal is about 0.6 mm X 0.3 mm X 0.2 mm.

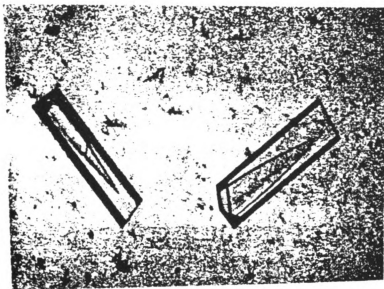


Figure 16. Crystals of human α -thrombin-PPACK complex. Lengths of the crystals are about 1 mm. Taken from [23].

protein complex, 0.075M of sodium phosphate buffer (pH7.3), 0.24 M NaCl, 12% PEG 4K and 1mM NaN₃ in 10 μ L drops equilibrating against 1mL well solution of 0.1M sodium phosphate buffer (pH7.3), 24% PEG 4K and 1 mM NaN₃. The crystal used for intensity data collection, 0.6 mm X 0.4 mm X 0.3 mm in size, grew in about two weeks and has the same appearance as those in Figure 15.

For the human α -thrombin-hirulog 3 complex, crystals were grown using the hanging-drop method with 2.5 mg/mL of protein complex, 0.075M of sodium phosphate buffer (pH7.3), 0.19 M NaCl, 12% PEG 8K and 1mM NaN₃ in 10 mL drops equilibrating against 1mL well solution of 0.1M sodium phosphate buffer (pH7.3), 24% PEG 8K and 1 mM NaN₃. The crystal used for intensity data collection, an approximate cube of 0.4 mm on edge, grew in about two weeks and has the same appearance as those in Figure 15.

The appearances of the crystals (Figure 15, Figure 16) would seem to suggest that only the PPACK crystals are distinct, while the others might possibly to be in the same crystal system.

B. Data Collection and Reduction

For all the crystals mentioned above, X-ray diffraction intensity data sets were collected using a Siemens multi-wire area detector. Graphite monochromated CuK α radiation (1.5418 Å) was generated by a Rigaku RU200 rotating anode source (typically operated at a power of 50 kV x 150 mA). Crystals were mounted on a Nicolet P3/F four-circle diffractometer gonistat (Figure 17). Only one crystal was necessary for collecting each high resolution (2.2-2.4) data set, and each was exposed to X-rays for about 40 hours during a data collection. The program XENGEN [92] was used to process the raw

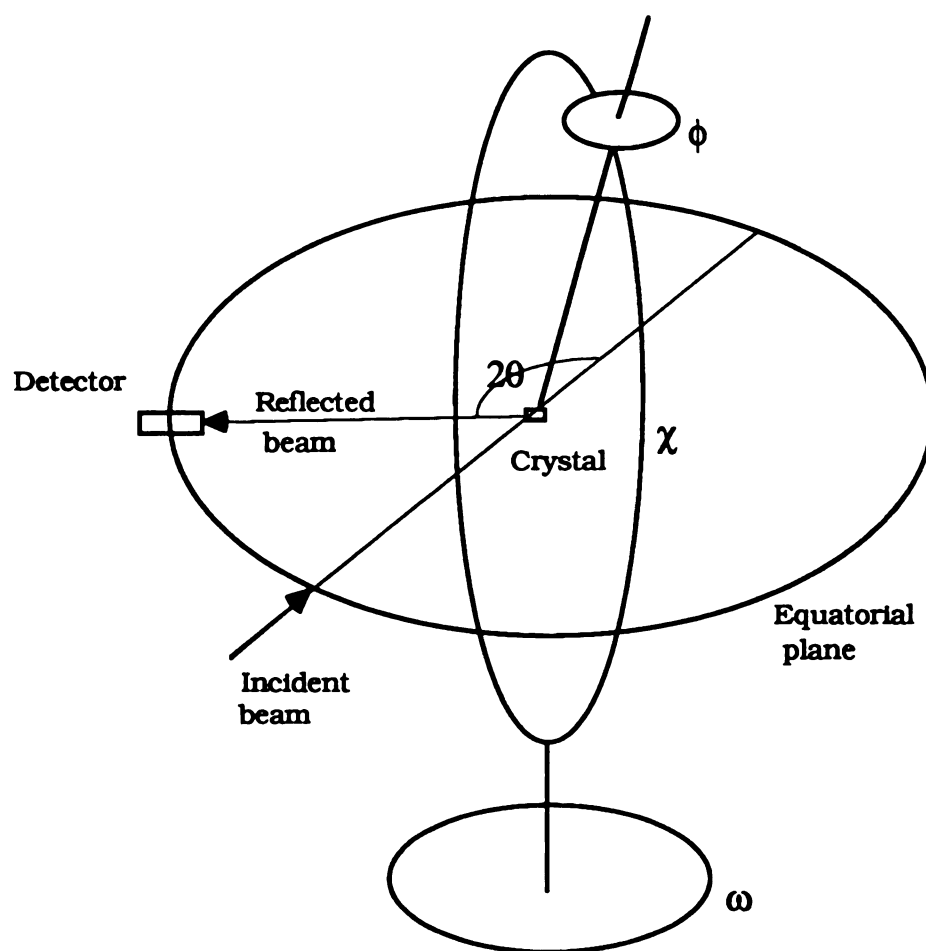


Figure 17. The geometry of A four-cycle diffractometer. The angles 2θ , ω , χ and ϕ are moved with computer controlling commands.

data sets to obtain integrated intensities.

(1) Preparation

Usually, a protein crystal will contain 27 to 65 % of solvent molecules [88], and will dry out and lose its X-ray diffracting ability quickly if the crystal is left in the open air. Therefore, all protein crystals must be mounted in a glass capillary sealed with mother liquor. The capillaries used for the above crystals (Lindeman) were 1.0 mm in diameter and siliconized. The capillary mounted crystal (Figure 18) will then be inserted into plastic clay on a goniometer head for X-ray diffraction experiments. The crystals were mounted in a random orientation in the capillary because the orientation matrix can be obtained later in the data reduction using XENGEN. The amount of mother liquor that bathes the crystal is very important: too much of it will increase the background significantly and cause the crystal to move slowly in the solution, while too little mother liquor will not provide enough support for the crystal and result in the incidental sliding of the crystal.

Once a crystal on a goniometer head is placed onto the goniostat of a four-circle diffractometer, it can be oriented using computer controlling commands by driving its ω , χ and ϕ angles (Figure 17), as well as the detector arm (2θ angle). While keeping the X-ray beam stationary, this setup allows us to detect any desired Bragg diffraction planes.

The actual explanation of the physical phenomenon involved in X-ray diffraction is very complicate in mathematical forms. Fortunately, a theory of visualizing the scattering of X-rays by a single crystal in terms of reflections from planes of atoms was introduced

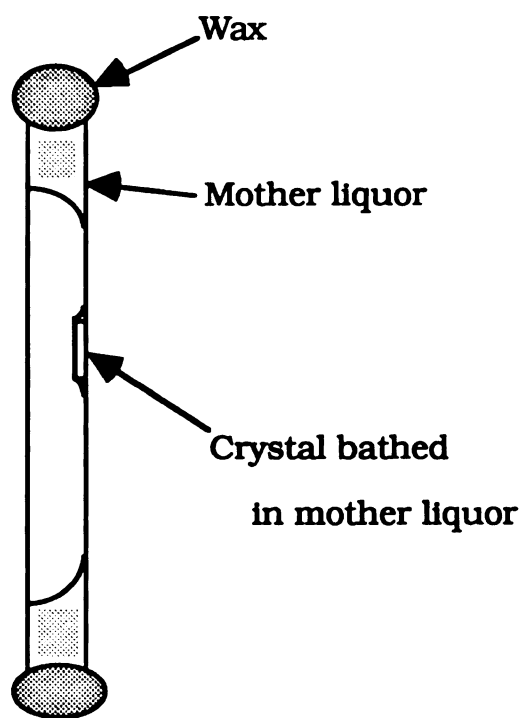


Figure 18. A protein crystal mounted in a sealed capillary for X-ray diffraction experiments. Mother liquor and the crystal are held by surface tension of the siliconized wall of the glass capillary.

by Bragg, which became one of the fundamental equations for X-ray diffraction [93]:

$$n\lambda = 2d \sin\theta,$$

where the integer n refers to the order of the spectrum, the wavelength of X-ray is λ , the distance between successive parallel planes of atoms is d (sometimes cited as the resolution) and the reflection angle is θ . This is why diffracted X-rays are often referred as "reflections". Therefore, a reflection can be identified by a set of $(2\theta, \omega, \chi, \phi)$ angles on the goniostat, as well as by a set of Miller indices (h, k, l) where reflection planes (h, k, l) intercept the crystal axes a, b and c by $(a/h, b/k, c/l)$. After the determination of the unit cell of the crystal, an orientation matrix becomes available to provide conversions between the $(2\theta, \omega, \chi, \phi)$ setting and the (h, k, l) setting. In the case of area detector data collection, this orientation matrix will normally be determined during the course of data collection, not before the data collection as in diffractometry.

(2) Area detector

An area detector is a very powerful instrument especially when applied to crystallographic studies of biological macromolecules. It enables us to collect a large number of reflections simultaneously not only to save time in data collection but also to reduce problems caused by the rapid decay of protein or DNA crystals. The detector itself (imaging proportional counter) is shown in Figure 19. Once a beam of X-rays strikes the detector inner chamber, the signal is transformed into an electronic one to be recorded. The detector screen is divided into 512×512 grids referred to as pixels. The X-ray intensity collected at each pixel is integrated, and all the pixels

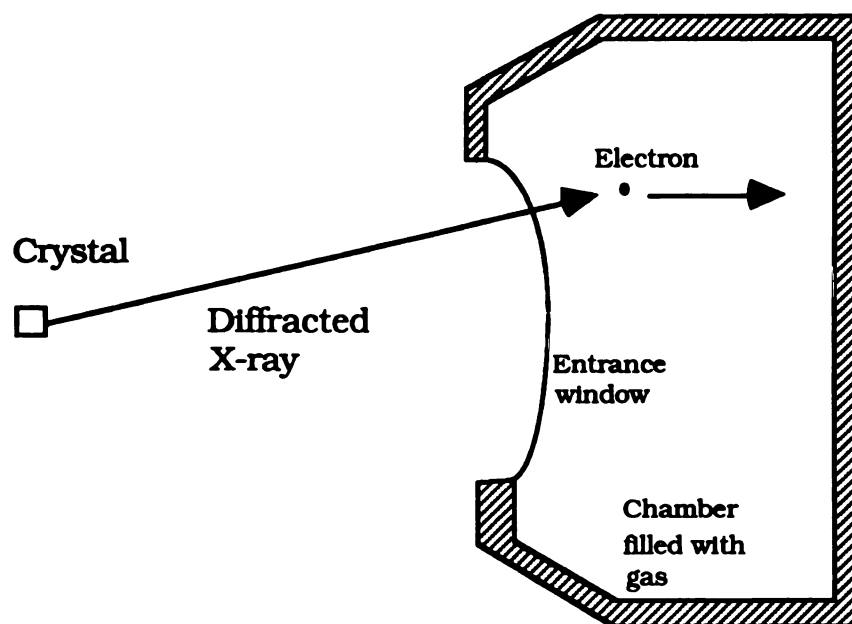


Figure 19. The area detector (imaging proportional counter).

add up to produce a two dimensional image of the diffracted X-rays (called a frame of measurement).

The Siemens X-1000 multiwire area detector system is composed of an area detector with a position decoding circuit (PDC), frame buffer, real time color display and network communications, and requires a MicroVAX computer workstation as a host for full functionality (Figure 20). The area detector is mounted on a sliding track of the 2θ arm of a four-circle goniometer and can be controlled from either the frame buffer or the VAX station using a menu user interface with easy-to-use pop-up menus and panels. For each X-ray event detected in a designated period of time, the output signals are decoded by PDC and the resulting digital positions are then sent to the frame buffer to form frame information. All these frames can be transferred via network (ethernet in our laboratory) to the VAX computer for storage and processing.

The main software for data collection consists of FRAMBO, SADIE and SANTA. The program FRAMBO (FRAME Buffer Operation) is used for frame acquisition and analysis and has to be set up in the very beginning of the data collection, the program SADIE (Siemens Area Detector Integrated Environment) is used when the host VAX terminal controls the frame buffer and goniometer, and the program SANTA (Siemens Network Transfer Agent) must be operating on the VAX to receive the frames when the detector is being controlled through the frame buffer.

(3) Frame acquisition

Major steps involved in area detector data collection are described as follows:

i) Flood field is performed at the desired crystal-detector distance (about one eighth of the longest crystal axis in angstroms in centimeters). The closest allowed distance for the detector in our laboratory is 11.65 cm and the value was used in all the data collections in this thesis), the detector collects radiation from an iron radiation source (bias set to the center value of the bias plot that will be checked every few weeks) over a certain period of time (usually 45 minutes FLOOD and 5 minutes ADD). This step allows the correction of background for pixel sensitivities so that every pixel on the detector apparently responds to signals uniformly.

ii) Brass plate is measured using the same iron source (same bias) and a special brass plate with a square matrix of holes. The detector collects the radiation through these holes for setting up the conversion from pixel to cm (usually collecting for an hour).

iii) Main beam position is determined. Taking off the iron source and lowering the value of bias by 0.14, a rotation frame is collected on an aligned Ni-Ni Packman crystal for 60 seconds, where the only movement is the rotation of the crystal about one axis (ϕ here) so that the picture will show an even distribution of reflections in layers [88]. A few Friedel pairs that are related by an inversion center are measured. The average center values correspond to the center position of the incident beam (subtracting these values from 511 gives the center position for XENGEN data processing).

iv) The crystal is optically aligned the at the center of the four-circle goniometer so that the incident X-ray aims at the center of the crystal no matter how the crystal is oriented during the data collection.

v) Collect one frame. This normally tells us how well the crystal diffracts X-rays and helps us to decide a swing angle (2θ), a collecting time for each frame and a power setting for the X-ray source. For a typical crystal of the thrombin complexes previously mentioned, the swing angle was about 15° , each frame was collected for 90 seconds and usually at a power of 50 kV x 150 mA (7.5 kW).

vi) Rocking curve of selected reflection is plotted after collecting 11 frames and displaying the sixth frame. It provides information of the width of the reflections to help to choose a scan range (usually 0.2° per step, that is, collect 5 frames for every degree). These profiles also determine the quality of the crystal (that is, the wider the profile, the poorer the quality of the crystal).

vii) Setup for data collection. We usually set up one $190^\circ \phi$ scan run and five $50^\circ \omega$ scan runs (50° apart at a different χ angle) for a 2.3 Å resolution data set of a monoclinic crystal. Triclinic protein crystals are uncommon, while crystals of higher symmetry need fewer frames (for example, an orthorhombic crystal requires only 90° of ϕ scan and three sets of $50^\circ \omega$ scan). We normally start to process the data after acquiring 100 frames, which will likely give the unit cell of the crystal, and we can interrupt the data collection and set up a new collecting scheme if necessary. The program R-SPACE, which can imitate the sweeping area of the detector against a three dimensional Ewald construction of the crystal, was used to determine the original sweeping scheme to ensure a complete set of reflection data. The Ewald construction [88] is a sphere drawn with center at the crystal (C) and radius $1/\lambda$ (Figure 21). The origin of the reciprocal lattice (**a***, **b***, **c***) (bold letters represent vectors) is

placed at point O where the unreflected X-ray traveling along **AC** meets the sphere. The condition for a diffraction along CB is that B is a reciprocal lattice point (h,k,l) or $\mathbf{S} = \mathbf{OB} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$. Therefore, in order to collect the (h,k,l) reflection, the crystal has to be oriented in such a way that the reciprocal lattice point cuts the reflection sphere.

(4) Data Reduction

Raw data can be refined and reduced by the programs XENGEN [92] on the VAX station to obtain integrated intensities. The major steps are listed below.

i) Set up. It sets up sub-directories and necessary files for the following data processing.

ii) Calibrate. By inputting the position of the origin (deduced from the main beam position mentioned previously), it provides conversions between pixel and cm.

iii) Border. It defines the active region of the detector.

iv) Spots. Usually using 100 frames, it produces a list of bright spots for the crystal/detector parameter refinement.

v) Refine. It performs the auto-indexing of the reflections and least-squares refinements. This provides information about the orientation matrix, the unit cell of the crystal, and refines the crystal/detector parameter.

vi) Integration. Integrated data (background is also determined here) is obtained with updated parameters. This is the most time-consuming step of the data reduction.

vii) Reduce. It reformates the data set and sorts the reflections according to the updated orientation matrix and unit cells. A small

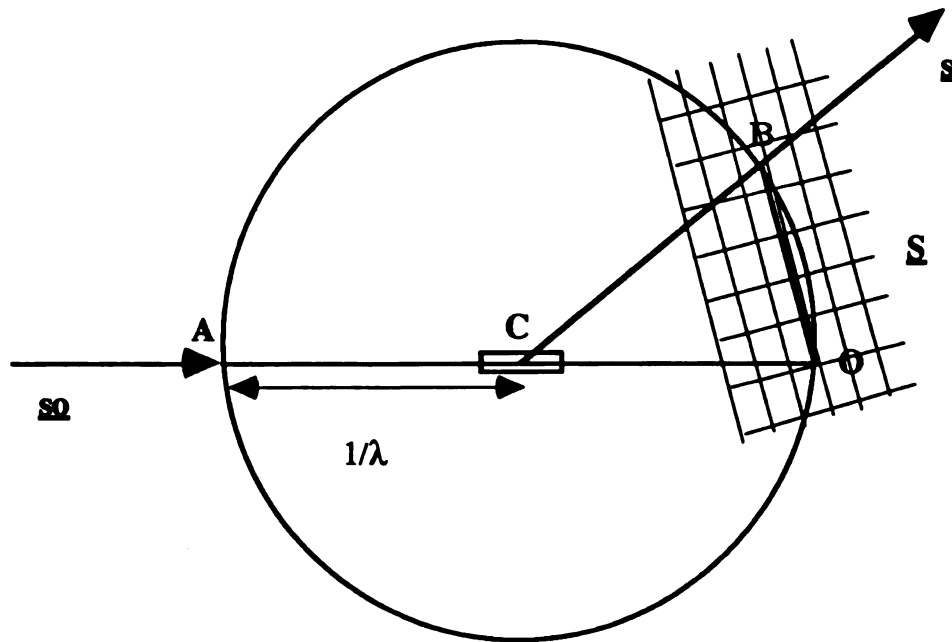


Figure 21. The Ewald construction. Crystal is at C, the wavelength of X-ray is λ , the origin of the reciprocal lattice (shown by crossing lines) is O, the incident beam is $\underline{s_0}$, the reflected beam is $\underline{s}$ and $\underline{S} = \underline{OB}$.

program XPREP can be used before the final REDUCE to analyze the systematic absences of the data set and ensure a correct space group and unit cell.

viii) Mrmerge. It merges the list of sorted reflections.

ix) Scale1. It computes scales and R factors.

x) Reject. It deletes outliers. A g.o.f. (goodness of fit) factor is calculated between the available redundant reflections; a g.o.f. of 5 means the reflection deviates by 5 times of the standard deviation. Usually we will first delete reflections with g.o.f. greater than 15, perform scale1 again, then reject reflections of g.o.f. greater than 10, perform scale1, reject reflections of g.o.f. greater than 5 and perform scale1 to obtain the final scale and R factors.

xii) Stats. It calculates final statistics of the data set, which includes the averaged unit cell, the number and averaged intensity of reflections by resolution shells, and R factors of several kinds.

xiii) Makemu. It makes an output file of a set of merged average intensities (I) or structure factors (F), which contains h, k, l, I (or F), σ and flags. The σ value is the standard deviation among the redundancies of the reflection; higher flags are set for bad reflections such as the ones with negative intensities.

Structure factors (or amplitudes), proportional to the square root of intensities, are usually obtained using the following formula:

$$|F(hkl)|^2 = k \times A \times D \times L \times P \times I(hkl),$$

where the background-corrected intensity $I(hkl)$ of the reflection (hkl) needs to be corrected for a scaling factor k, absorption factor A, a decay factor D, Lorentz factor L and polarization factor. When a crystal changes its orientation, it exposes different thicknesses to

the X-ray beam and thus has different absorptions. The absorption factor takes into account these differences in the absorption of X-rays. The decay factor takes into account the decay of the protein crystal with respect to its exposure (time and power) to X-rays. The Lorentz factor is a geometrical factor that takes into account the relative time each reflection spends in the reflection position. It equals to $1/\sin^2\theta$ when the rotation axis is normal to the scattering plane, but is different for other geometries. The polarization factor at a Bragg angle θ , equal to $(1 + \cos^2 2\theta)/2$, takes into account the partial polarization of the diffracted beam if the incident beam is unpolarized.

In processing the area detector data using XENGEN, the corrections were applied automatically by internal scaling among the redundancies of a reflection and among different sets of frames collected upon different exposure time. When collecting the data sets for the thrombin complex crystals of this thesis, the treatment by XENGEN was quite reliable because: (1) the thicknesses of the crystals were fairly close in each dimension so that the absorption correction is small; (2) the crystals were not decaying very rapidly over the relatively short exposures (40-50 hours), so that the decay of these crystals is not tremendous; (3) there were very many redundant reflections collected: therefore, the internal scaling scheme should work well in a macroscopic or statistical sense. Nonetheless, the method is not flawless. For example, the intensity of a specific reflection could increase with exposure so that the scaling correction for the reflection is actually increasing the error.

It is always a good idea to use another fresh crystal to continue the data collection if significant decay is observed.

After data processing, the quality of a data set is usually judged by the value of R_{merge} , which measures the average deviation between redundant reflections, and

$$R_{\text{merge}} = \sum |I_i - \langle I \rangle| / \sum I_i.$$

The typical value for data sets mentioned previously is about 0.05.

C. Isomorphous Replacement

In order to solve a crystal structure using structure factors observed in X-ray diffraction experiments, we need to calculate an electron density map $\rho(\text{xyz})$:

$$\rho(\text{xyz}) = V^{-1} \sum_h \sum_k \sum_l \mathbf{F}(\text{hkl}) \exp[-2\pi i(\text{hx} + \text{ky} + \text{lz})],$$

$$\mathbf{F}(\text{hkl}) = |\mathbf{F}(\text{hkl})| \exp[i\alpha(\text{hkl})].$$

The structure amplitude $|\mathbf{F}(\text{hkl})|$ is proportional to the square root of diffraction intensity, the unit cell volume is V , but the phases $\alpha(\text{hkl})$ have to be acquired before the calculation can be performed. During a diffraction data collection, only intensities of the reflections are recorded, phase information is lost: this is the so-called "phase problem" in crystallography. Usually, phases have to be determined by special methods, for example, molecular replacement, multiple isomorphous replacement (MIR), multi-wavelength anomalous dispersion (MAD) or their combinations. Since the crystals mentioned above were found to be isomorphous to those of known thrombin complexes, that is, they all have almost the same chemical content and crystal packing (same space group and unit cell), we have been able to solve all the structures in using isomorphous replacement methods thrombin coordinates.

Theoretically, if we know the crystal structure of a thrombin-inhibitor complex TH (T stands for thrombin; H stands for inhibitor), that is, we know both the structure amplitudes and the phases of TH, we can state that $\mathbf{F}_{TH}$ is known and:

$$\mathbf{F}_{TH} = \mathbf{F}_T + \mathbf{F}_H,$$

where the contribution to structure factors from thrombin is $\mathbf{F}_T$ and the contribution to structure factors from the inhibitor is $\mathbf{F}_H$ (capital letters in bold designate vectors). Similarly, for an unknown thrombin complex TX, we have:

$$\mathbf{F}_{TX} = \mathbf{F}_T + \mathbf{F}_X.$$

Now, if these two complexes are isomorphous, the contribution to structure factors from crystal packing should be the same in both crystals, and, since inhibitors are only a very small portion of the complex (usually one to two kDa of molecular weight, compared to 36.6 kDa of thrombin), the protein complexes should also be approximately the same, so that we have:

$$\mathbf{F}_{TX} @ \mathbf{F}_{TH} \equiv \mathbf{F}_T.$$

or more precisely:

$$\alpha_{TX} \equiv \alpha_{TH} \equiv \alpha_T.$$

Therefore, we can calculate an approximate electron density map for the unknown complex structure TX by using the phases of the known structure TH (or T) using TH (or T) as a model structure for TX, which can be presented as:

$$\rho(xyz) = V^{-1} \sum_h \sum_k \sum_l |\mathbf{F}_{TX}(hkl)| \exp[i\alpha_{TX}(hkl)] \exp[-2\pi i(hx+ky+lz)]$$

$$\equiv V^{-1} \sum_h \sum_k \sum_l |\mathbf{F}_{TX}(hkl)| \exp[i\alpha_{TH}(hkl)] \exp[-2\pi i(hx+ky+lz)].$$

The electron density map (called $|F_o|$ map because $|F_{TX}(hkl)|$ is the observed structure amplitude) can then be improved by adjusting the model structure and refining. This is the isomorphous replacement method.

Many forms of electron density maps can be calculated besides the $|F_o|$ map. The electron density maps used here are the $(2|F_o| - |F_c|)$ maps and $(|F_o - F_c|)$ difference maps, where the calculated structure factor (F_c) is obtained from the model structure. The difference map $(|F_o - F_c|)$ shows the lack of agreement between the "real" and modeled structures; it can be used to bring out the unknown part of the structure or measure the accuracy of the current model. The advantage of using $(2|F_o| - |F_c|)$ (that is, $|F_o| + (|F_o| - |F_c|)$) maps instead of (F_o) maps is that the "signal for change" is amplified, which is more helpful for the improvement of the model structure. In this work, we started with part of the thrombin molecule as a model. In all the cases, the initial $(2|F_o| - |F_c|)$ and $(|F_o - F_c|)$ maps showed good density for the inhibitors, and the starting crystallographic R factor is about 0.32, where:

$$R = \Sigma ||F_o| - |F_c|| / \Sigma |F_o|.$$

D. Model Building and PROLSQ Refinement

Model structures and electron density maps were refined using the restrained least squares program PROLSQ, a package with many options and flexibilities [94]. Protein structures refined by PROLSQ usually have better geometry than those by several other refinement packages (XPLOR, EREF). Assuming the best model would have the smallest "sum of weighted squares of differences between F_o and F_c " (represented by Q), PROLSQ minimizes the function Q :

$$Q = \sum W_i [F_{o_i} - F_{c_i}([X])]^2$$

where the observed structure factor (F_o) and the calculated structure factor (F_c) of a reflection i are designated as F_{o_i} and F_{c_i} , the weight applied to the reflection i is represented by W_i , and $[X]$ is a set of restraints on distances (bond distances, angle distances and planar 1,4 distances), planar groups, chiral groups, van der Waals contacts, torsion angles, thermal parameters and non-crystallographic symmetry.

There are two subroutines, SCATT and PROTIN, that precede the main routine PROLSQ. The subroutine SCATT combines information of observed structure factors (h,k,l,F) and atomic scattering factors into a binary file SCATT.BIN. The subroutine PROTIN deals with the coordinates (x,y,z,B) of the model structure and a dictionary file, which contains ideal geometrical information of each of the amino acids and special groups that exists in the protein complex, and generates a binary file PROTIN.BIN as another input for PROLSQ. With a control file containing information such as the unit cell and symmetry of the crystal, refinement resolution, geometrical and thermal restraints, etc., the main routine can then be carried out, which produces shift information at the end of each cycle and updates the SCATT.BIN file with new phases. A subroutine MERGE applies the shifts to the old coordinates and produces a set of new and refined coordinates. A flow chart of the whole procedure is illustrated in Figure 22.

A standard dictionary file contains ideal models and types of restraints for the 20 normal amino acids, several types of N- and C-termini, cis- peptide links and a few more special groups that are

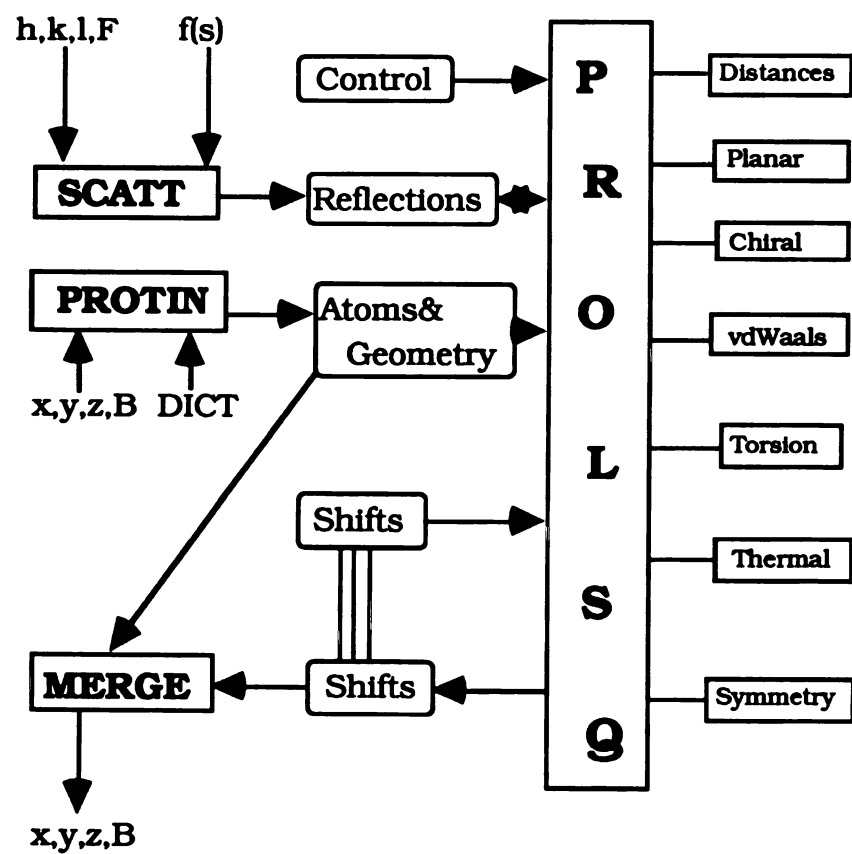


Figure 22. The flow chart of PROLSQ programs.

often seen in protein crystals (for example, H_2O , SO_4^{2-} and the heme group). Because many of the synthetic peptides that we were studying possess unusual groups (see Table 3 on page 31) in the sequences, we had to modify this dictionary file in order to refine the structures containing these special groups. This would only be an ordinary problem if the "abnormal peptide" differs from a natural peptide only at side chains. Because of the fact that all proteins are amino acids joined together through peptide links (Figure 23), the dictionary file is written in a way that assumes all the main chain structures are repeating units of N, CA, C and O atoms. Therefore, if the main chain connection of the peptide is different from natural (for example, in cases of β -homoarginine and α -keto-arginine), it had to be dealt with separately. This problem was solved, which was applied to other similar work of this laboratory.

The numbering systems for the peptides of present work are presented in Table 5 and will be used consistently throughout this thesis. Some of the special groups are shown in Figure 24 while others are clear enough in Table 3.

In the case of refining the cyclotheonamide A structure, the standard dictionary file should be able to recognize Ala*1' (the abbreviation of N-form-Ala or N-formyl-alanine), Pro2' and Arg3'. As with D-Phe5' or any other residue that is a d-enantiomorph, the only revision of dictionary file is to change the IHAND option of methionine in the chiral center specification card to 0 (instead of 1) so that all the chiral centers are restrained to have ideal chirality with the hand of current model. Therefore, only the CO and TNT groups need to be specified (Figure 24). The ideal model of TNT was

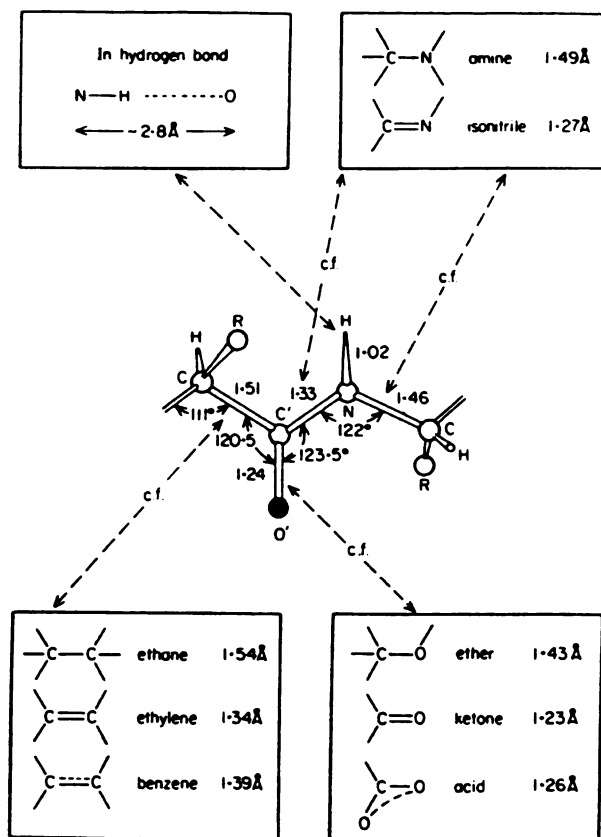


Figure 23. Standard bond lengths and angles of the peptide link. Lengths of similar bonds are also shown for comparison. Because of the delocalization of the π -electrons, atoms C, C', O', N, and C are in a planar structure (the C's are usually designated by CA). Taken from Blundell and Johnson (1976) [88].

Table 5. The numbering system used in this thesis.

											
Cyclotheonamide A:	N-form-Ala-Pro-Arg-(CO)-D-Phe-TNT										
Numbering:	1'	2'	3'	4'	5'	6'					
CV863-1:	Boc-Asp-Pro-Arg -(CO)-Phe*										
Numbering:	0'	1'	2'	3'	4'	5'					
PPACK:	D-Phe-Pro-Arg- chloromethylketone										
Numbering:	1'	2'	3'								
MDL-28050:	Suc-Tyr-Glu-Pro-Ile-Pro-Glu-Glu-Ala-Cha-D-Glu										
Numbering:	55'	56'	57'	58'	59'	60'	61'	62'	63'	64'	65'
Hirulog 3:	D-Phe-Pro-β-homoArg-Gly-Gly-Gly-Gly-Gly-Hirugen*										
Numbering:	1'	2'	3'	4'	5'	6'	7'	8'	53'-64'		

*The abbreviations used here (e.g. Boc, Suc, Cha) are the same with those of Table 3 (page 31).

*N-form-Ala1' of cyclotheonamide A is also abbreviated to Ala*1'.

*CO is a carbonyl group (carbon and oxygen).

*TNT is shown in Figure 24 (page 65).

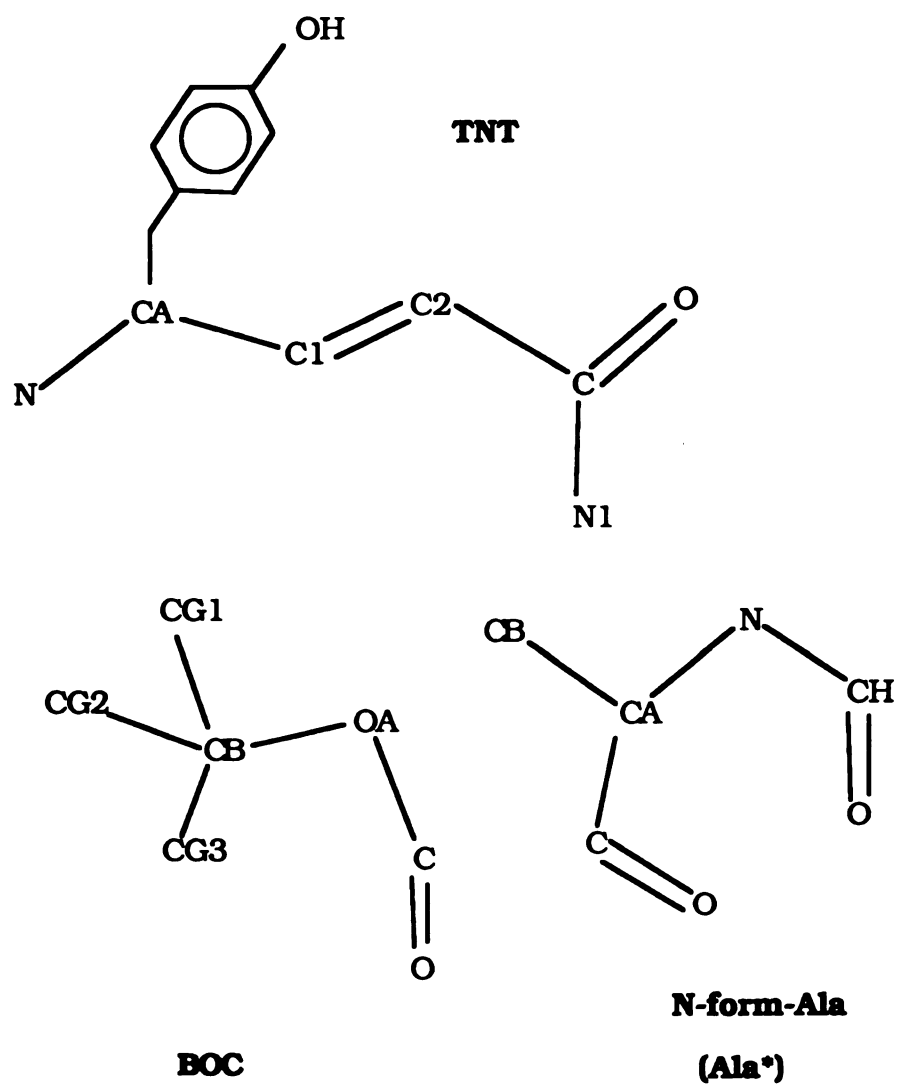


Figure 24. Several unusual groups referred to in the thesis.

built based on the model of a tyrosine and bond lengths and angles of standard values (C1-C2 1.41 Å, C2-C 1.50 Å, C-O 1.21 Å, C-N1 1.33 Å; CA-C1-C2 125°, other angles around double bonds are about 120°). A series of distances and planar restraints were added externally in the control file of the subroutine PROTIN, which also contains the definition of restraints for the connection of TNT6' to D-Phe5' and Ala*1', and the connection of CO4' to Arg3' and D-Phe5' (Table 6). In the actual refinement, the residue number of D-Phe5' had to be named intentionally far apart from Arg3' so that there will be no default restraints between them. The residue number of CO and TNT had to be named further apart so that they would not fall into the range that was defined as normal peptide chains in the control file. Moreover, the coordinates of CO4' and TNT6' must be placed at the end of the coordinates file, just before those of the solvent molecules. The Arg3' C-Ser195 OG distance was not restrained originally, which was found to be 1.9 Å during the refinement, and which was then restrained very loosely at 1.9 Å (Table 6).

In the refinement of CV-863-1, Phe*5' is treated as a regular phenylalanine missing the two carboxyl atoms C and O. The CO4' group is treated exactly the same way as that of CO4' in cyclotheonamide A. The model for BocO' was built using the standard geometry of similar molecules (Figure 24, the tetrahedral centered at CB has bond angles of 109.5° and bond lengths of 1.55 Å except for CB-OA 1.43 Å, C-OA 1.39 Å, C-O 1.21 Å, CB-OA-C 110°, OA-C-O 121°). The coordinates of BocO', as well as those of CO4', were placed at the end of the coordinates file with residue numbers far apart from those defined in peptide chains. The restraints and connection

Table 6. The external restraints of cyclotheonamide A in the PROTIN control file.

Bond distances:

CO4' C - Arg3' C	1.49 A
CO4' C - D-Phe5' N	1.32 A
TNT6' N - D-Phe5' C	1.32 A
TNT6' N1 - Ala*1' CB	1.47 A
Arg3' C - Ser195 OG	1.90 A **

Angle distances:

CO4' C - Arg3' CA	2.62 A
CO4' C - Arg3' O	2.37 A
CO4' O - Arg3' C	2.37 A
CO4' C - D-Phe5' CA	2.47 A
CO4' O - D-Phe5' N	2.29 A
Arg3' C - D-Phe5' N	2.44 A
TNT6' N - D-Phe5' CA	2.47 A
TNT6' CA - D-Phe5' C	2.47 A
TNT6' N - D-Phe5' O	2.29 A
TNT6' N1 - Ala*1' CA	2.47 A
TNT6' C - Ala*1' CB	2.47 A

Planar_{1,4} distances:

CO4' O - D-Phe5' CA	2.71 A
TNT6' CA - D-Phe5' O	2.71 A

Planar Groups:

(Arg3' CA- C- O) & (CO4 C)
 (Arg3' C), (CO4' C- O) & (D-Phe5' N- CA)
 (D-Phe5' CA-C-O)& (TNT6' N-CA)
 (TNT6' CA-C1-C2-C-O-N) & (Ala*1' CB)
 (TNT6' CB-CG-CD1-CE1-CZ-CE2-CD2-OH)

* Abbreviation for N-form-Ala1'.

** Very loosely restrained.

of Boc0' (Table 7) were defined externally in the PROTIN control file. The restraints for Arg3' C to Ser195 OG distances are also restrained loosely at 1.9 Å.

For refining PPACK, restraints of D-Phe1' were treated the same way as those of D-Phe5' in cyclotheonamide A. The chloromethylketone group had only shown the methyl carbon atom in the crystal structure, therefore, the whole group was not included in PROLSQ refinement. Because Arg3' was treated as a C-terminal of a peptide chain (that is, Arg-OXT where an extra oxygen was included) and oxygen differs carbon by only two electrons, the OXT atom was changed to a carbon atom at the end of the refinement without introducing observable errors.

For MDL-28050, Suc55' was treated as an aspartic acid excluding the coordinates for the amino nitrogen atom. The β -cyclohexylalanine group differs from a phenylalanine residue by only the side chain conformation. Therefore, a model structure of Cha64' was constructed based on a phenylalanine, and the cyclohexane ring (1.52 Å on every edges and 109.5° for every bond angles) was in a chair conformation. Thus, restraints of three planar groups (CG-CD1-CE2-CZ, CG-CD2-CE1-CZ and CD1-CE1-CD2-CE2) need to be specified in the PROTIN control file, in addition to the usual revisions of the PROLSQ dictionary file.

For hirulog 3, a model of the β -homoArg ($\text{Arg}^*-\text{CH}_2-\text{C}=\text{O}$) was built based on an arginine (CA-CH₂ 1.54 Å, CH₂-C 1.51 Å and CA-CH₂-C 109.5°). The coordinates of the homoArg3' also had to be placed at the end of the coordinates file with a residue number far apart from those defined in peptide chains, and its restraints and

Table 7. The external restraints of Boc0' and homoArg3' in the PROTIN control file.

Boc0'

Bond distances:

Boc0' C - Asp1' N	1.32 A
-------------------	--------

Angle distances:

Boc0' OA - Asp1' N	2.31 A
--------------------	--------

Boc0' O - Asp1' N	2.29 A
-------------------	--------

Boc0' C - Asp1' CA	2.47 A
--------------------	--------

Planar Groups:

(Boc0' OA- C- O) & (Asp1' N- CA)

homoArg3'

Bond distances:

homoArg3' N - Pro2' C	1.32 A
-----------------------	--------

homoArg3' C - Gly4' N	1.32 A
-----------------------	--------

Angle distances:

homoArg3' N- Pro2' CA	2.47 A
-----------------------	--------

homoArg3' N - Pro2' O	2.29 A
-----------------------	--------

homoArg3' CA - Pro2' C	2.47 A
------------------------	--------

homoArg3' CH ₂ - Gly4' N	2.47 A
-------------------------------------	--------

homoArg3' C - Gly4' CA	2.47 A
------------------------	--------

homoArg3' O - Gly4' N	2.29 A
-----------------------	--------

Planar 1,4 distances:

homoArg3' CA - Pro2' O	2.71 A
------------------------	--------

homoArg3' O - Gly4' CA	2.71 A
------------------------	--------

Planar groups:

(homoArg3' N-CA)& (Pro2' C- O- CA)

(homoArg3' CH₂- C- O) & (Gly4' N- CA)

(homoArg3' NE- CZ- NH- NH₂).

* Loosely restrained.

connection were also added externally in the PROTIN control file (Table 7).

The PROLSQ least squares refinements were carried out by a series of geometrically "tight-loose-tight" restrained cycles, starting with an overall thermal parameter (typically $30 \text{ \AA}^2$) for all the atoms at 2.8 $\text{\AA}$ resolution and no solvent, then proceeding to individual thermal parameters for each atom and gradually increased resolution (typically 2.8, 2.5, 2.3 $\text{\AA}$). Water molecules were introduced at 2.5 $\text{\AA}$ resolution, where two different ($|F_o - F_c|$) difference density maps were calculated at (7.0-2.5 $\text{\AA}$) and (8.0-2.5 $\text{\AA}$) respectively, and where a water was required to have prominent density in both maps and in the ($2|F_o| - |F_c|$) map, and to also have reasonable close contacts from other oxygen or nitrogen atoms. Both the thermal parameters and occupancies of water molecules were refined, and only reasonably good waters were retained (typically required to have $2|F_o| - |F_c|$ electron density with $B < 35 \text{ \AA}^2$ and occupancy > 0.3). Between each set of refinement cycles, new electron density maps were calculated and the model was updated to the new maps with computer graphics programs of FRODO [95] on a Evans & Sutherland PS390 stereographics system, adjusting torsion angles and adding newly resolved residues. Typically, the final structure contains about 240 water molecules and is refined under a variable weighting scheme on structure amplitudes.

The quality and accuracy of final structures are usually determined by their R factors (a structure with R factor below 0.20 is usually a good one), which measure the relative discrepancies of the observed and calculated structure factors. The r.m.s. (root mean

squares) discrepancies of the final refinement with respect to their target geometries also measure the quality of the final structure, while the torsion angles are usually required to be close to planar for ω -angles and in the allowed regions of the Ramachandran plot for ϕ and ψ angles, where ω = CA-N-C-CA, ϕ = C-CA-N-C and ψ = N-CA-C-N. Each of the allowed regions in the Ramachandran plot are usually close to the conformational angles of a certain type of secondary structure (α helix or β sheet); if a torsion angle is not in the allowed regions of the plot, it usually means bad close contacts exist among the atoms near the residue. A way of estimating the errors in atomic positions was proposed by Luzzati [96] and shown in Figure 25, assuming that the errors (Δr) are the sole cause of the F_o and F_c differences, which are dependent on R factors and resolution range. Furthermore, the thermal parameter B, which causes a decrease in diffraction intensity by a factor $\exp[-B(\sin\theta/\lambda)^2]$ because of the thermal motion of atoms, also shows the quality of the structure. A B-factor of 15-20 $\text{\AA}^2$ corresponds to a mean displacement of 0.15 $\text{\AA}$ to 0.5 $\text{\AA}$, but this estimate is very rough because the B-factors in the actual refinement are normally determined empirically and take into account of a variety of other factors such as static disorder, incorrect scaling, and absorption. As a rule, a higher thermal parameter usually means poorer accuracy of the structure.

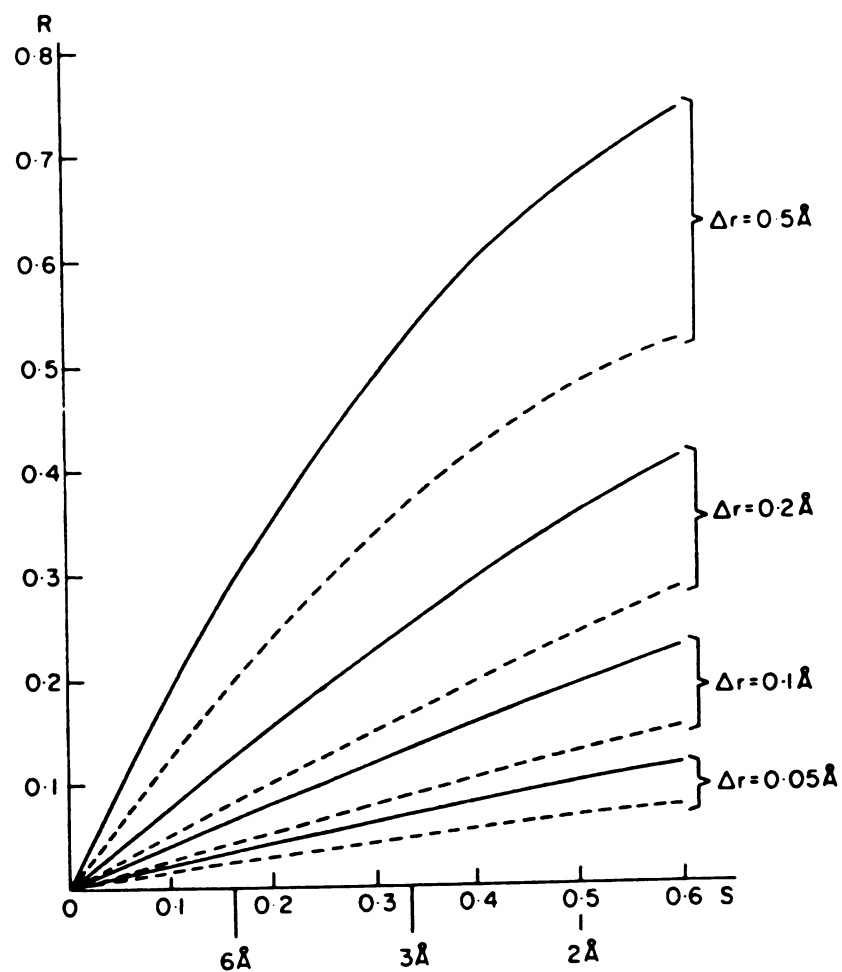


Figure 25. The estimation of errors (Δr) in atomic position. Taken from Luzzati, 1952 [96]. Crystallographic R factor is plotted versus $S = 2\sin\theta/\lambda$.

CHAPTER 3. RESULTS

A. The Cyclotheonamide A-Hirugen-Thrombin Complex

(1) General

Using methods described earlier, a set of diffraction data on the crystal of the cyclotheonamide A-hirugen- α -thrombin complex was collected. The X-ray generator was operated at 50 kV x 150 mA (7.5 kW power) and a 0.3 mm collimator was used for the incident beam. The crystal-to-detector distance was set at 11.65 cm. The detector swing angle was 15°. The scan range was 0.2° per frame, and each frame was collected for 90 seconds. The crystal diffracted well to 2.3 Å resolution, with the average peak width of a reflection being about 0.6°. After XENGEN intensity data reduction, a total of 14,017 independent reflections was obtained from 46,395 observations. After eliminating weak reflections with $I/\sigma(I) < 2$, a data set of 12,650 unique reflections was used for solving the structure (79% observed, $R_{\text{merge}} = 0.043$). This data set is complete to 2.5 Å resolution, and contains half of the possible reflections between 2.3 to 2.5 Å resolution.

The crystal was found to be isomorphous to those of the hirugen and hirulog 1 complexes [27]: monoclinic, space group C2, four molecules per unit cell with $a = 70.61$ Å, $b = 72.38$ Å, $c = 73.35$ Å and $\beta = 101.10^\circ$. The average crystal volume per unit molecular weight, V_m , is equal to 2.4 Å³/dalton. The protein fraction of the crystal was calculated to be 51% (empirically assuming a protein specific volume of 0.74 cm³/g for protein crystals, the protein fraction in this crystal is equal to 0.74 multiplied by the total weight

of protein in a unit cell and then divided by the volume of the unit cell).

The structure was solved by the isomorphous replacement method described earlier in Chapter 2. The phases of this structure were approximated using the thrombin coordinates of hirulog 3-thrombin complex [32]. The starting model includes residues Ser1E-Ile14K of the A chain, Ile16-Thr147 and Gln151-Glu247 of the B chain (see Table 1 on page 6). Both hirugen and cyclotheonamide A were not included in the initial phase calculation and crystallographic R factor started at 0.32. Beginning with an overall thermal parameter of $30 \text{ \AA}^2$ at 2.8 Å resolution, we directly proceeded with 25 cycles of PROLSQ refinement and improved the model to an R factor of 0.23 with individual thermal parameters. The first $(2|F_O| - |F_C|)$ electron density map at 7.0-2.8 Å resolution showed good density for thrombin, hirugen and most of the cyclotheonamide A residues; the latter two were also prominent in the $(|F_O - F_C|)$ difference electron density map. There was no density for the thrombin autolysis loop residues Trp148-Lys149E, similar to that of the hirugen and hirulog 1 complexes [27]. Models of hirugen and the clearly resolved part of cyclotheonamide A were then fitted into the electron density using computer graphics. Further refinements of the structure were carried out with a modified dictionary file that allowed proper restraints to the cyclotheonamide A model, and were continued with gradually increased resolution (2.8, 2.5, 2.3 Å) and addition of water molecules. The details of the course of the refinement are listed in Table 8.

The final structure of the complex converged at $R = 0.138$ with

Table 8. The details of the PROLSQ refinement of the cyclotheonamide A-hirugen-thrombin complex structure.

# of cycles	Resolution (Å)	Reflections/ Variables	R factor	Remarks
25	7.0-2.8	7986/9109	0.230	2276 atoms, $\langle B \rangle = 26 \text{ Å}^2$
11	7.0-2.8	7986/9342	0.214	2335 atoms, $\langle B \rangle = 28 \text{ Å}^2$
21	7.0-2.5	11114/9625	0.217	hirugen and CtA* included
11	7.0-2.5	11114/10070	0.181	add 89 waters $\langle B \rangle = 28 \text{ Å}^2$
11	7.0-2.3	12056/10290	0.172	add 44 more waters
6	7.0-2.3	11777/10254	0.164	deleted 279 weak reflections
11	7.0-2.3	11777/10556	0.149	add 54 waters , loose CtA*
12	7.0-2.3	11777/10821	0.142	correct scale, variable weight
14	7.0-2.3	11777/10816	0.138	239 waters, restrain B (15-60)

*CtA = Cyclotheonamide A

an average thermal parameter of 30 Å² using 239 water molecules and 11,777 reflections (7.0-2.3 Å resolution). The r.m.s. deviations from ideal restraints and the R-factors by resolution ranges are listed in Table 9 for the final structure. The ω -angles of 97% of the peptide bonds are with $\pm 6^\circ$ of planarity, and only the same few residues reported before [25] are slightly out of the conformationally allowed regions (Figure 26). The mean error in coordinates is estimated to be approximately 0.2 Å, as described in Chapter 2 [96].

(2) Overall structure

In the cyclotheonamide A-hirugen-thrombin complex, the structure of thrombin is well defined in electron density except for the few terminal and autolysis loop residues, similar to those in the hirugen and hirulogs complexes (Thr1H-Ser1E, Asp14L-Arg15, Glu247 and Trp148-Lys149E) [27,32]. A few more residues on the surface of the protein, namely Asn62, Lys81 and Gln244, have side chain electron density only to their CB atoms. Excluding these residues that are loosely arranged on the surface of the protein, the calculation of the r.m.s. differences between the thrombin structure in different crystals show good consistency (Table 10). The structure of hirugen is also well defined, although the side chains of Glu58' and Glu61' are directed into the solvent region and have no electron density. The details of this part of the structure are practically the same as those of the hirugen structure reported earlier [27]. The distribution of thermal parameters of the structure is shown in Figure 27, which is also very similar to that of the hirudin-thrombin complex [97]. The thermal parameters of hirugen, the terminal regions of thrombin (N- and C- terminals of A chain and C-terminal if

Table 9. Final least squares restraints/deviations and R factor statistics of cyclotheonamide A-hirugen-thrombin.

	Target Sigma	RMS Deviation
distances (A)		
bond lengths	0.020	0.019
angle lengths	0.035	0.049
planar 1,4	0.050	0.052
planes (A)		
peptides	0.020	0.020
aromatic groups	0.020	0.020
chiral volumes (A ³)	0.15	0.20
non-bonded contacts (A)		
single torsion	0.55	0.21
multiple torsion	0.55	0.29
possible H-bond	0.55	0.28
torsion angles (deg)		
planar	3	3
starggered	15	23
orthonomal	20	26
thermal parameters (A ²)		
main chain bond	1.5	1.3
main chain angle	2.0	2.2
side chain bond	2.0	2.0
side chain angle	2.5	3.1

d _{min}	Number of reflections	$\sigma(F_o)$	$\langle F_o - F_c \rangle$	R factor shell	R factor sphere
4.40	1645	33.1	81.3	0.149	0.149
3.64	1701	28.5	58.3	0.109	0.129
3.23	1665	25.5	51.5	0.124	0.127
2.95	1601	23.1	44.5	0.146	0.131
2.73	1697	21.0	40.0	0.154	0.134
2.55	1771	19.0	36.0	0.164	0.137
2.30	1697	17.1	32.0	0.158	0.138

diffraction pattern : $\sigma(|F_o|) = (22.5) + (-150.0)[\sin(\theta/\lambda) - (1/6)]$
 $\langle ||F_o| - |F_c|| \rangle = 49.0$

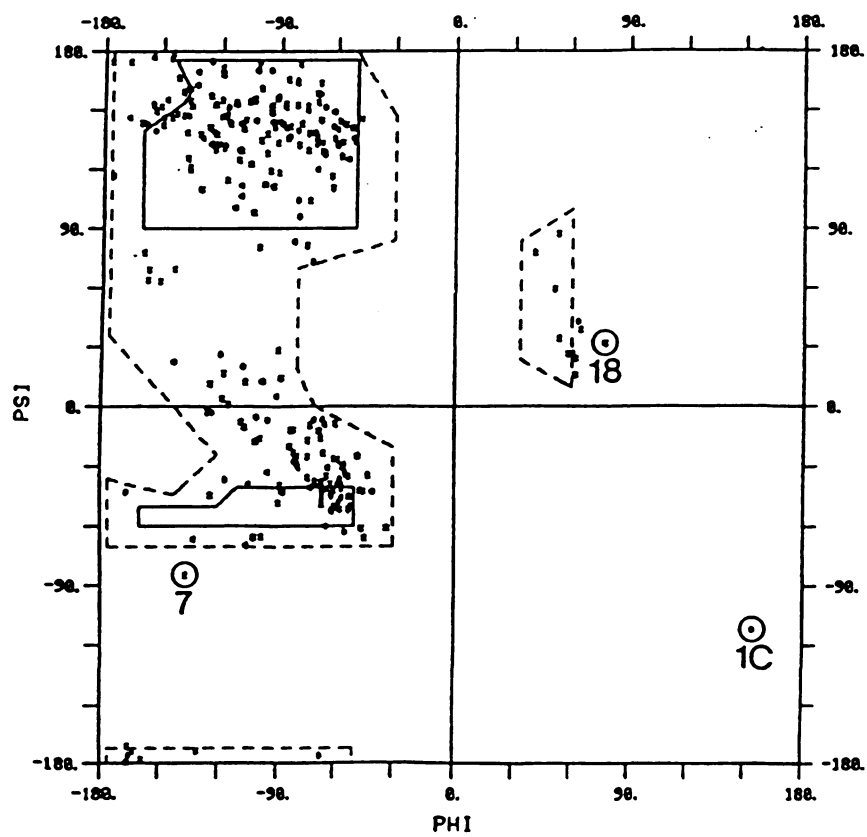


Figure 26. The Ramachandran plot of the cyclotheonamide A-hirugen-thrombin complex structure.

Table 10. The RMS deviations between similar thrombin structures.

	A - E	B - E	C - E	A - B	G - E
main chain (A)	0.3	0.3	0.4	0.2	0.5
side chain (A)	0.7	0.5	1.1	0.6	1.4
sulfurs (A)	0.3	0.3	0.3	0.4	0.3
55'- 64' (A)	1.5	1.4	---	0.5	1.4
1'- 3' (A)	1.1	1.3	0.8	0.3	---
P1'- P4' (A)	---	---	---	---	5.8*

	D - F	D - E	E - F	H - C
main chain (A)	0.3	0.3	0.3	0.3
side chain (A)	1.1	0.6	1.1	1.2
sulfurs (A)	0.4	0.4	0.2	0.4
55'- 59' (A)	0.8	---	0.7	---
60'- 64' (A)	6.8(0.7**)	---	0.9	---
1'- 3' (A)	---	---	---	0.3

A = Cyclotheonamide A-hirugen-thrombin

B = CV863-1-hirugen-thrombin

C = PPACK-thrombin

D = MDL-28050-thrombin

E = Hirulog 3-thrombin

F = Hirugen-thrombin

G = Hirudin-thrombin

H = PPACK-thrombin of Bode et al.[22]

*Average estimated value between the main chain of Gly4'-Gly7' and hirudin Glu49'-His51'.

**Average estimated value between main chain atoms when the 60'-64' segments were fitted to superpose each other.

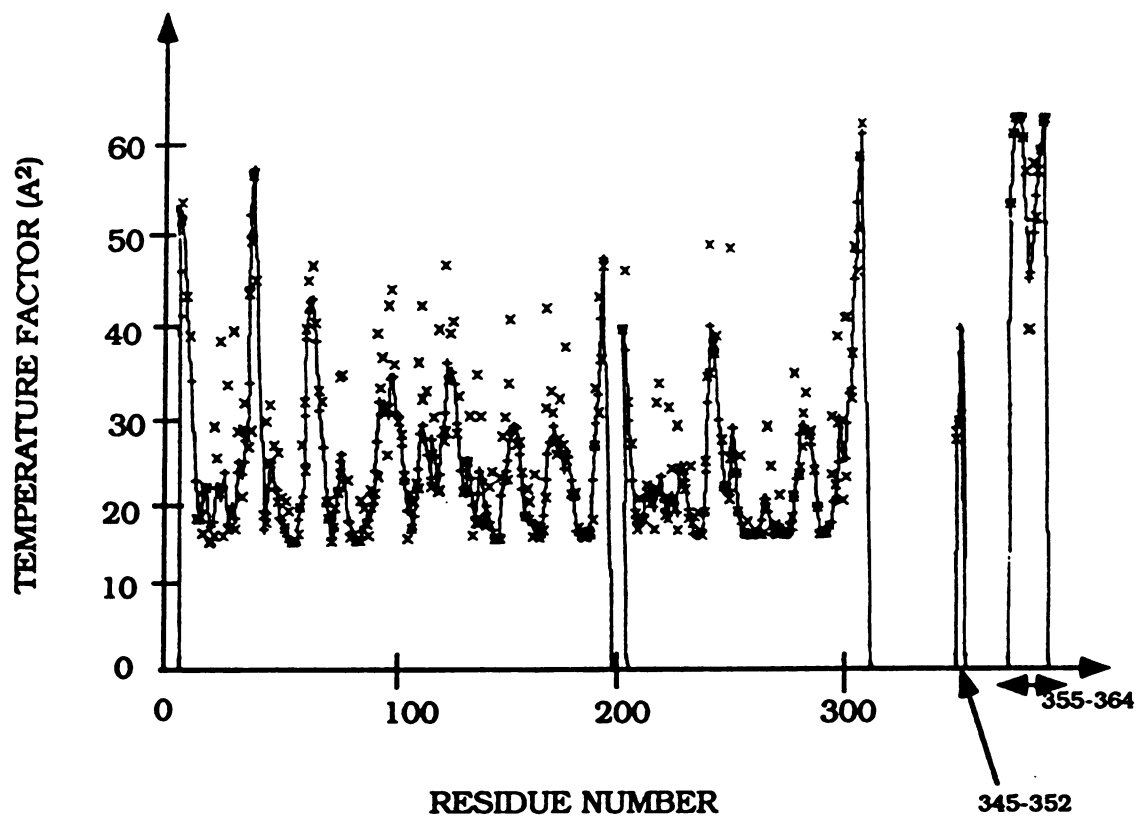


Figure 27. The distribution of thermal parameters of the cyclotheoamide A-hirugen-thrombin complex. In this figure, thrombin is from residues 4 to 295; cyclotheoamide A is from residues 345 to 352; hirugen is from residues 355 to 365.

B chain) and the insertion loop residues are higher than the rest of the structure. The overall solvent pattern, including the channel of about 20 water molecules that connect the thrombin active site to the 186A-D insertion loop [25], is also very consistent in all the structures. The distributions of thermal parameters and occupancies of water molecules are plotted in Figure 28. Cyclotheonamide A is well defined in the electron density except for a small break at D-Phe5' CB (Figure 29). The average thermal parameter of cyclotheonamide A is about $40 \text{ \AA}^2$, which is higher than that of thrombin (less than $30 \text{ \AA}^2$) but lower than that of the hirugen ($55 \text{ \AA}^2$).

Macrocyclic peptides are fairly common in nature, and many of them have been classified as toxins or antibiotics, especially those from primitive plants like fungus or mushrooms (see [98] for a review of this field). The spatial structures of natural and synthetic cyclic peptides have also been studied extensively in crystals and solutions [99-101]. Generally, backbone conformations of cyclic peptides are less flexible than those of linear peptides. They are generally favored in two antiparallel pleated β -sheets with transannular hydrogen bonds stabilizing the folding (Figure 30a). Nonetheless, the actual ring conformation of a cyclic peptide is affected largely by the length of ring, the chirality of each amino acid, the side groups and the polarity of its environment. For example, the favored conformation is different in nonpolar solvent than in a polar one (Figure 30b).

Cyclotheonamide A is not a normal cyclic peptide and therefore should have a conformation that is somewhat different from a normal one. There are a total of five amide bonds, five chiral centers

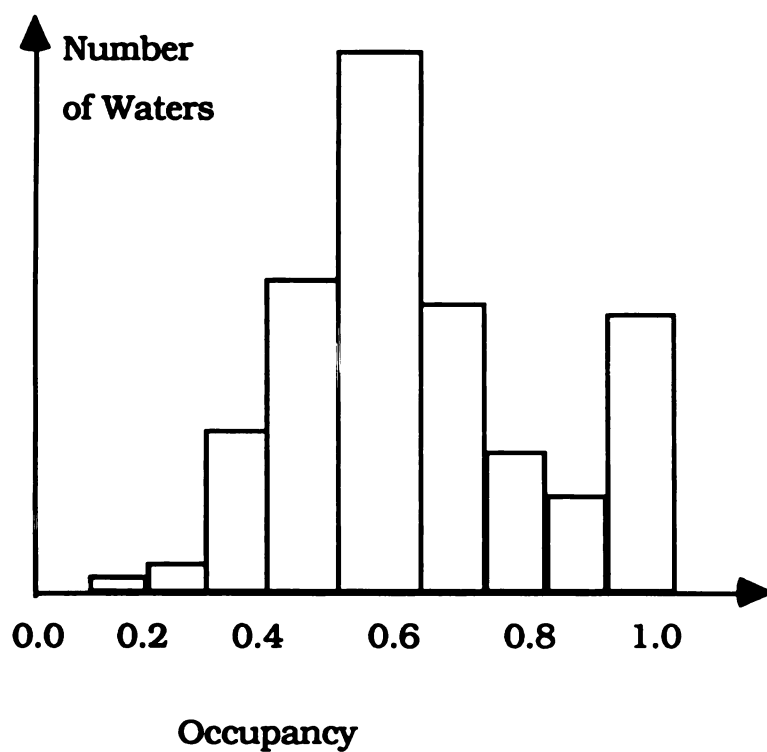
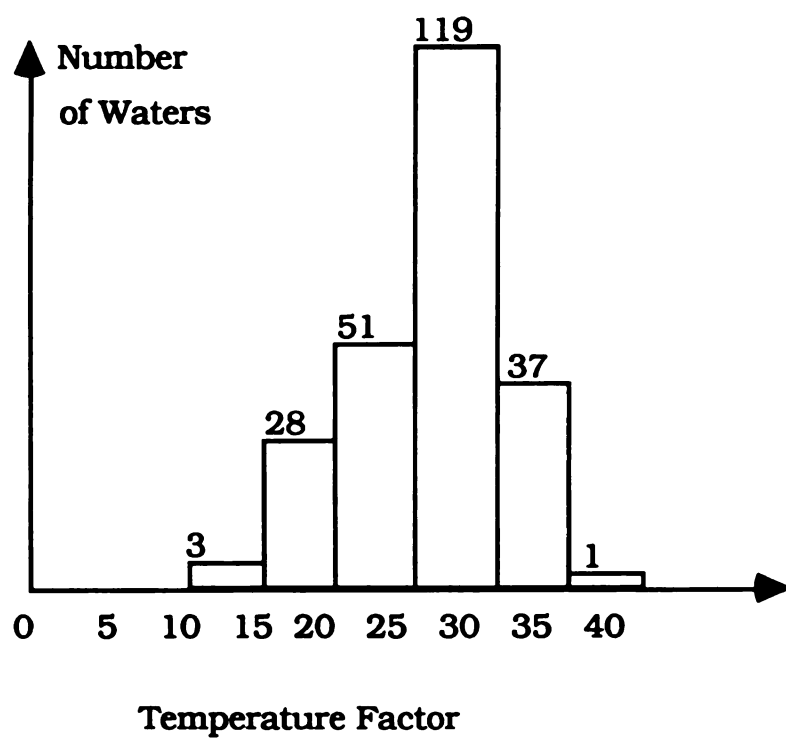


Figure 28. Distributions of thermal parameter and occupancy of water molecules.

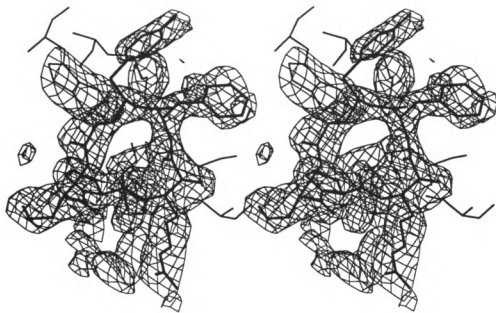
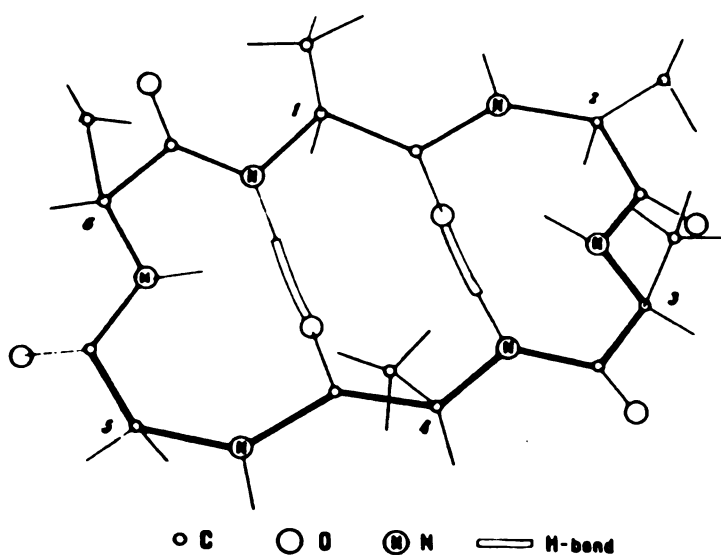


Figure 29. Stereoview of the electron density for cyclotheonamide A.

(a)



(b)

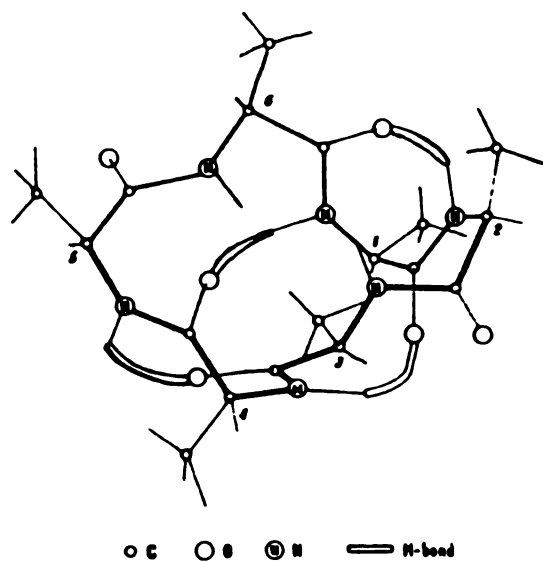
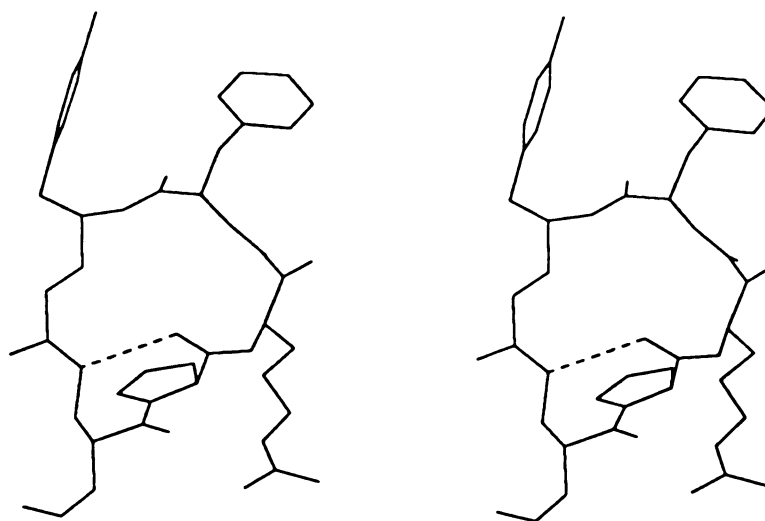


Figure 30. Favored conformations of cyclohexopeptides. (a) The "pleated sheet" conformation in polar media. (b) The conformation of cyclohexa-L-alanine in nonpolar media. Taken from [99].

(L L L D D) and five "side groups" so that it could be called a cyclopenta-peptide [100]; however, there are 19 atoms in the ring so that it is likely to be closer to the conformation of a cyclohexa-peptide [101], which has a 18-membered ring. A closer view at the backbone structure of cyclotheonamide A (Figure 31) shows that the ring is twisted, probably due to its interactions with thrombin. Among the five amide carbonyls of the 19-membered ring (Arg3' CO is not counted), only that of Pro2' is pointed inwards, while the rest are all pointing to the outside of the ring. As a result, only one transannular hydrogen bond is found to stabilize the ring (Pro2' O-TNT6' N1, 3.1 Å), although two other hydrogen-bond like van der Waals contacts are also present (Ala*1' O-TNT6' N1 3.3 Å, CO4' O-Arg3' N 3.2 Å).

Interestingly, the conformation of cyclotheonamide A is closer to the favored conformation of the cyclohexa-peptide in nonpolar media (Figure 30b) than that in a polar one (Figure 30a). This is probably because of the predominant nonpolar environment of the thrombin-cyclotheonamide A binding region: the side groups of Pro2', D-Phe5' and TNT6' all form hydrophobic interactions with thrombin, only that of Arg3' forms hydrophilic interactions. It is therefore reasonable to suggest that the conformation of a unbound cyclotheonamide A molecule in aqueous solutions would be closer to that of the pleated sheets (Figure 30a) than that of the bent ring (Figure 30b). The torsion angles of the cyclotheonamide A ring (Table 11) show that the Ala*1'-Arg3' stretch is still in a pleated β -sheet conformation, while the 5'-6' stretch can be regarded as a distorted pleated sheet. This probably implies that the D-Phe2'-Arg3'

(a)



(b)

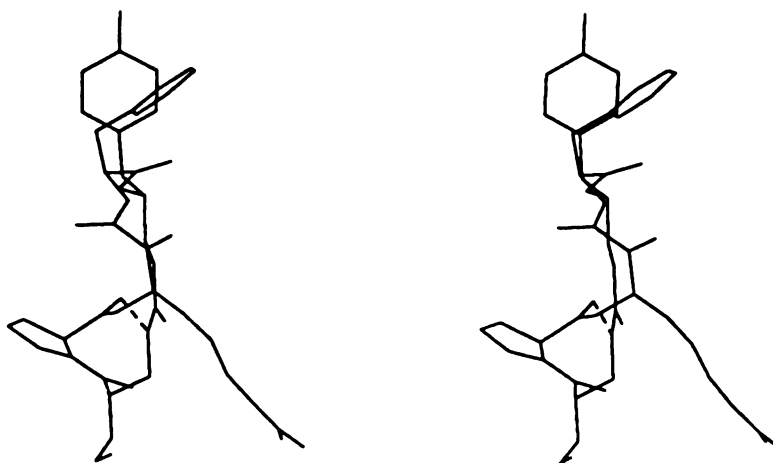


Figure 31. The backbone ring conformation of cyclotheonamide A in the crystal of its complex with thrombin. (a) Top-view. (b) Side-view. The transannular hydrogen bond (Pro2' O-TNT6' N1) is represented by a broken line.

Table 11. Torsion angles of the nineteen-membered ring in cyclotheonamide A.

Definition	Value (degree)
1' CB-1' CA-1' C -2' N	-177
1' CA-1' C -2' N -2' CA	179
1' C -2' N -2' CA-2' C	-56
2' N -2' CA-2' C -3' N	144
2' CA-2' C -3' N -3' CA	-177
2' C -3' N -3' CA-3' C	-97
3' N -3' CA-3' C -4' C	38
3' CA-3' C -4' C -5' N	95
3' C -4' C -5' N -5' CA	180
4' C -5' N -5' CA-5' C	156
5' N -5' CA-5' C -6' N	-144
5' CA-5' C -6' N -6' CA	-172
5' C -6' N -6' CA-6' C1	-112
6' N -6' CA-6' C1-6' C2	-161
6' CA-6' C1-6' C2-6' C	-179
6' C1-6' C2-6' C -6' N1	0
6' C2-6' C -6' N1-1' CB	-178
6' C -6' N1-1' CB-1' CA	128
6' N1-1' CB-1' CA-1' C	54

motif binds to thrombin tighter and at a faster rate, while the D-Phe5'-TNT6' segment binds to thrombin only after a conformational change (and probably an adopted unfavored energy state for the ring). Experimental observations also show that cyclotheonamide A reaches its full thrombin-binding potency very slowly, and the D-Phe5'-TNT6' addition does not contribute significantly to the binding [33].

The structure also shows that the C1=C2 double bond in TNT6' is in an ordinarily unfavored cis-conformation, bringing the TNT6' N1 atom close to the Pro2' O atom to form the transannular hydrogen bond. The α -keto amide group and the Ala*1' CB-TNT6' N1 connection correspond to the β -turn regions of a normal cyclohexapeptide [100-101]. The presence of the CO4' group is important, not only because it forms the transition-state interaction and hydrogen bonds with thrombin, but also because it allows the correct spacing (one extra bond length and a reversal of the peptide chain) for the novel D-Phe5' interactions with thrombin. Although α -keto amide transition state complexes for immunosuppressant macrocycles bound to immunophilins have been visualized crystallographically [63,67], this is the first visualization of an α -keto amide transition state complex for a serine proteinase.

(3) Cyclotheonamide A-thrombin interaction

As predicted from its sequence, cyclotheonamide A binds to the active site of thrombin, with the Pro2'-Arg3' sequence binding in the same way as that of PPACK in PPACK-thrombin complexes [22, 32]. As listed in Table 12, the S1 specificity site of thrombin is occupied by Arg3', with its guanidinium group forming a hydrogen-bonded ion pair with Asp189 (Figure 32). The S2 apolar site is occupied by

Table 12. Cyclotheonamide A-thrombin interactions (< 4.0 Å).

covalent:

CO4' C-Ser195 OG	1.8 Å
------------------	-------

ion pair:

Arg3' NH1-Asp189 OD1	2.7 Å
Arg3' NH2-Asp189 OD2	2.9 Å

hydrogen bond:

Ala*1' N-Gly216 O	2.9 Å
Ala*1' O-Gly216 N	3.1 Å
Arg3' N-Ser214 O	3.1 Å
Arg3' N-Ser195 OG	2.9 Å
Arg3' O-Ser195 N	2.6 Å
Arg3' O-Gly193 N	2.8 Å
CO4' O-His57 NE2	2.7 Å

van der Waals:

Ala*1' C-Trp215 CB	3.7 Å
Pro2' CB-His57 CD	3.4 Å
Pro2' CB-Trp60D CH2	3.9 Å
Pro2' CB-Leu99 CD	3.7 Å
Pro2' CG-Tyr60A CZ	3.7 Å
D-Phe5' CZ-Leu40 CD2	3.6 Å
D-Phe5' CE2-Leu41 C	3.5 Å
D-Phe5' CE2-Leu41 O	2.5 Å
D-Phe5' CA-Trp60D CH2	3.5 Å
D-Phe5' CE2-Gly193 CA	2.9 Å
TNT6' C1-Trp60D CZ2	3.5 Å
TNT6' N-Trp60D CH2	2.8 Å
TNT6' C1-Glu192 CG	3.8 Å

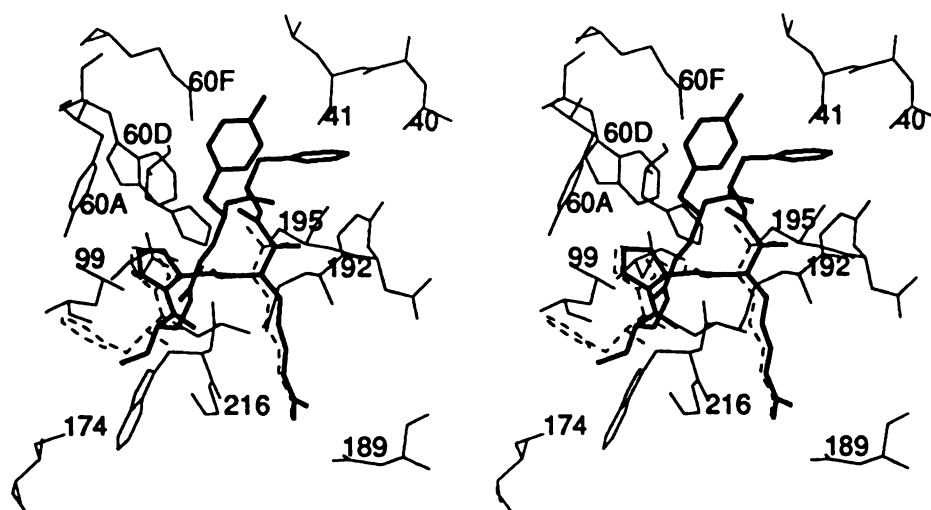


Figure 32. Stereoview of cyclotheonamide A bound in the active-site cleft of thrombin. Cyclotheonamide A, dark; thrombin, light; PPACK, broken.

Pro2' involving hydrophobic contacts with His57, Tyr60A, Trp60D and Leu99 (Table 12). Moreover, the Ala*1'-Arg3' segment forms a hydrogen-bonded, double-stranded antiparallel β -sheet with the Ser214-Gly216 sequence of thrombin, which includes a van der Waals contact between this two β -strands (Table 12). Other than its main chain interactions mentioned above, the side chain and the formyl group of Ala*1' do not contribute to the thrombin binding; the later is in the vicinity of the D-Phe ring of PPACK, but too polar and too small to interact with thrombin the same way as that of the D-Phe ring.

The most impressive feature of the cyclotheonamide A-thrombin binding is the formation of the complicated hydrogen-bonded tetrahedral intermediate (hemiketal) that resembles a transition state for peptide hydrolysis using the α -keto amide group (Figure 33). The two carbonyl groups of the α -keto amide group are oriented at a dihedral angle of 109° , akin to the carbonyls of FK-506 in its complex with the immunophilin FKBP (Figure 13 on page 24) [63]. The transition state "tetrahedral intermediate" bond, CO4' C-Ser195 OG, is at a distance of 1.8 Å, longer than the same bond in the PPACK-thrombin complex reported by Bode et al. [22] (about 1.6 Å) and the standard C-O single bond length (1.45 Å). The OG oxygen impinges nearly orthogonally on the α -keto group (Ser195 OG-Arg3' C-Arg3' O = 86° , CO4' C-Arg3' C-Ser195 OG = 93° , Arg3' CA-Arg3' C-Ser195 OG = 97°). Moreover, the keto oxygen (Arg3' O) makes a bifurcated hydrogen bond with the nitrogen atoms of Gly193 and Ser195 of thrombin (Table 12); the former was observed in the transition state intermediate of the phenylethane boronic acid

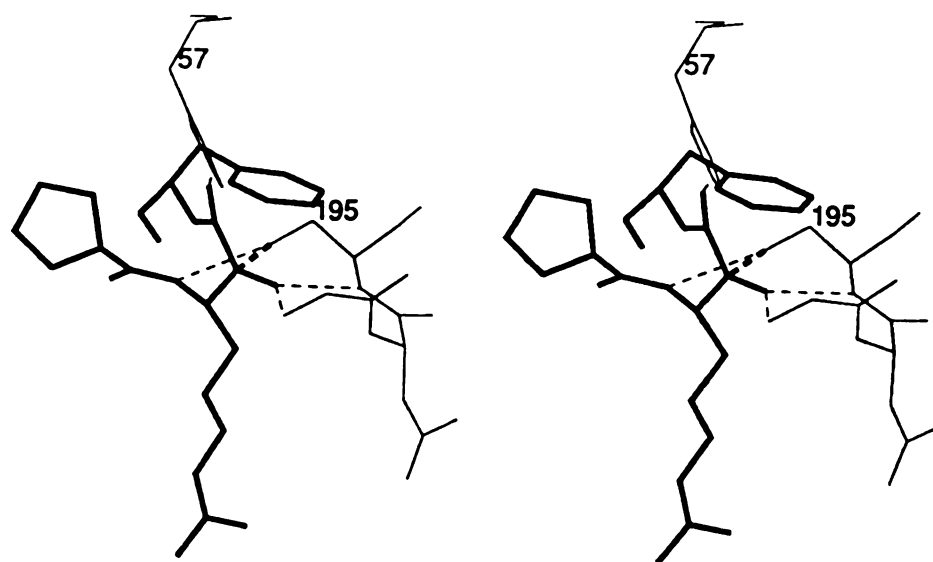


Figure 33. Stereoview of the α -keto amide group interacting with thrombin. Cyclotheonamide A, dark; thrombin, light; the transition state "tetrahedral intermediate" bond, dark broken; hydrogen bonds, light broken.

(PEBA)-chymotrypsin complex (2.7 Å) [73], and the latter was observed in the PPACK-thrombin intermediate (weaker at 3.1 Å). The main chain nitrogen atom of Arg3' is also hydrogen-bonded with Ser195 OG, similar to that in the PPACK-thrombin complex [22]. These similarities notwithstanding the oxygen atom of CO4' of the α -keto amide group is involved in a very strong hydrogen bond with His57, which disrupts the geometrical arrangement between Ser195 and His57. This agrees with our earlier observations that the region is very sensitive to the binding of different types of inhibitor or substrate, and plays an active role in the catalytic hydrolysis of serine proteinase (the distance Ser195 OG-His57 NE2 is about 2.4 Å in both the hirugen and MDL-28050 complexes, about 3.0 Å in both the PPACK and cyclotheonamide A complexes; it is even further separated in the hirulog 3 complex with a distance of about 3.5 Å). The geometrical arrangement between Asp102 and His57 are practically the same in both active site unoccupied (hirugen-thrombin and MDL-28050-thrombin) and occupied (PPACK and cyclotheonamide A) states, even though it is also essential to the catalytic function. In this structure, the carbon atom of CO4' is 2.4 Å from Ser195 OG and 3.6 Å from His57 NE2, compared to 2.1 Å and 1.9 Å for that of PPACK [22] that represents the state of binding after hydrolysis. Therefore, the transition state of thrombin hydrolysis scissile cleavage of a substrate, in which the substrate is under the influence of catalysis but has not yet been cleaved, can be approximated by the binding mode of the α -keto group in this structure. It is an amide bond in a normal substrate, but it will be cleaved by the enzyme so that the structure is not observable.

Many new thrombin-inhibitor interactions are revealed by the novel chain reversal (different from that of FPA, which will be discussed later), although the cyclic main chain has to be distorted for the better use of the D-Phe5' and TNT6' side groups. The D-Phe5'-TNT6' stretch appears to be "sandwiched" between the Tyr60A-Thr60I insertion loop and the Glu192-Ser195 segment, making "numerous" van der Waals contacts (or hydrophobic interactions) (Table 12, Figure 32). The D-Phe5' residue interacts with Leu40, Leu41, Trp60D and Gly193 of thrombin, while the TNT6' residue (also has a D- chiral center) interacts with Trp60D and Glu192. Ironically, aromatic rings of D-Phe5' and TNT6' do not make good aromatic stacking interactions (within 4.0 Å distances) with the aromatic thrombin residues Tyr60A and Trp60D, although they do make many hydrophobic contacts with Tyr60A and Trp60D and the TNT6' ring is perpendicular to that of Trp60D (the usually favored conformation for aromatic stacking at 4.2 Å). Therefore, the limited flexibility of the cyclic peptide does not seem to be optimized for thrombin binding, which agrees with the experimental results from testing its binding properties [33].

Cyclotheonamide A was found to be a much better inhibitor for trypsin (K_d 23 nM, compared to 180 nM with thrombin) [33]. Its availability in the Japanese sponge might be to protect against trypsin digestion of the sponge by trypsin-containing sponge-eating animals and thus represents a way of survival for the sponge. Nonetheless, the novel features of the cyclotheonamide A-thrombin binding did provide many very important new ideas for the future

design of thrombin inhibitory drugs, which is illustrated by the design of the CV863-1 peptide.

B. The CV863-1-Hirugen-Thrombin Complex

(1) General

The diffraction data set of the CV863-1-hirugen-thrombin complex crystal was also collected as described previously. Parameters of interest are listed in Table 13 and are compared with those of the cyclotheonamide A complex. The data set is complete up to 2.5 Å resolution, and contains half of the possible reflections between 2.3 to 2.5 Å resolution. The crystal was also found to be isomorphous to the hirugen, hirulog 1 [27] and cyclotheonamide A complexes of thrombin [33] (Table 13). Therefore, the structure was also solved by the isomorphous replacement method that has been described in Chapter 2. The phases of this structure were approximated using the thrombin coordinates of hirulog 3-thrombin complex [32]. The initial model included residues Ser1E-Ile14K of the A chain, Ile16-Thr147 and Gln151-Glu247 of the B chain. Both hirugen and CV863-1 were not included in the first phase calculation and the initial crystallographic R factor was 0.30. Starting with an overall thermal parameter of 30 Å² at 2.8 Å resolution, we proceeded with 22 cycles of PROLSQ refinement and improved the model to an R factor of 0.23 with individual thermal parameters. The first ($2|F_O| - |F_C|$) electron density map at 7.0-2.8 Å resolution showed good density for thrombin, hirugen and most of the CV863-1 residues; the latter two were also prominent in the ($|F_O - F_C|$) difference map. There was no density for Trp148-Lys149E of the autolysis loop of thrombin, similar to the hirugen and hirulog 1 complexes [27].

Table 13. Data collection, processing and crystal characterization of the crystals of the hirugen-thrombin complexes of CV863-1 and cyclotheonamide A.

	CV863-1	Cyclotheonamide A
power of X-ray	50 kV x 150 mA	50 kV x 150 mA
collimator size	0.3 mm	0.3 mm
crystal-detector	11.65 cm	11.65 cm
swing angle	15°	15°
scan range	0.2°	0.2°
collecting time	90 sec.	90 sec.
average peak width	0.6°	0.6°
resolution	2.3 Å	2.3 Å
# collected	47,784	46,395
# independent	14,312	14,017
# with $I/\sigma(I) > 2$	13,409	12,650
data completion	86%	79%
R_{merge}	0.032	0.043
crystal system	monoclinic	monoclinic
space group	C2	C2
# of molecules	4 per unit cell	4 per unit cell
a	71.20 Å	70.61 Å
b	72.36 Å	72.38 Å
c	73.52 Å	73.35 Å
β	101.17°	101.10°
V_m	2.4 Å³/dalton	2.4 Å³/dalton
protein fraction	51 %	51 %

Models of hirugen and the clearly resolved parts of CV863-1 were then fitted into the electron density using computer graphics. Further refinements of the structure were carried out with a modified dictionary file that allowed proper restraints to the CV863-1 model, and were then continued to gradually increased resolution (2.8, 2.5, 2.3 Å) and inclusion of water molecules. The course of refinement was similar to that of the cyclotheonamide A complex structure (Table 8).

The final structure of the complex, which includes 234 water molecules, converged at $R = 0.142$ with an average thermal parameter of 28 Å^2 using 12,293 reflections (7.0-2.3 Å resolution). The r.m.s. values and R factors by resolution ranges are listed in Table 14. The ω -angles of 97% of the peptide bonds are with $\pm 6^\circ$ of planarity and only the same few residues reported before [25] are slightly out of the conformationally allowed regions (the Ramachandran plot is omitted because it is practically the same as Figure 26). The mean error in coordinates is estimated, as described in Chapter 2 [96], to be about 0.2 Å.

(2) Overall structure

In the CV863-1-hirugen-thrombin complex, thrombin residues are mostly well defined in the electron density. The same few residues mentioned previously in the cyclotheonamide complex, Thr1H-Ser1E, Asp14L-Arg15, Glu247 and Trp148-Lys149E, have no density [27,32]. Residue Asn62 also has electron density only to its CB atom. Excluding the few disordered residues, the r.m.s. differences between thrombin structures in different crystals, listed in Table 10, indicated a very good agreement. The structure of

Table 14. Final least squares restraints/deviations and R factor statistics of CV863-1-hirugen-thrombin.

	Target Sigma	RMS Deviation
distances (A)		
bond lengths	0.020	0.019
angle lengths	0.035	0.049
planar 1,4	0.050	0.054
planes (A)		
peptides	0.022	0.018
aromatic groups	0.022	0.019
chiral volumes (A ³)	0.15	0.20
non-bonded contacts (A)		
single torsion	0.55	0.21
multiple torsion	0.55	0.26
possible H-bond	0.55	0.24
torsion angles (deg)		
planar	3	3
starggered	15	22
orthonomal	20	32
thermal parameters (A ²)		
main chain bond	1.5	1.3
main chain angle	2.0	2.2
side chain bond	2.0	2.1
side chain angle	2.5	3.1

d _{min}	Number of reflections	$\sigma(F_o)$	$\langle F_o - F_c \rangle$	R factor shell	R factor sphere
4.41	1647	34.6	84.4	0.153	0.153
3.63	1745	30.0	61.5	0.112	0.132
3.21	1777	26.9	53.5	0.128	0.131
2.92	1754	24.4	47.0	0.151	0.135
2.70	1763	22.2	41.5	0.154	0.137
2.54	1738	20.3	37.2	0.166	0.140
2.30	1869	18.6	34.5	0.167	0.142

diffraction pattern : $\sigma(|F_o|) = (24.5) + (-150.0)[\sin(\theta/\lambda) - (1/6)]$
 $\langle ||F_o| - |F_c|| \rangle = 51.0$

hirugen is also well defined, except for the side chains of Glu58' and Glu61', which are directed into the solvent region and do not have electron density. The details of the hirugen part are really the same as the originally reported hirugen structure [27] and the hirugen in the cyclotheonamide A complex. The distribution of thermal parameters of the structure is also similar to Figure 27, with hirugen and terminals of thrombin having the highest B factors. The overall solvent pattern is quite consistent with other thrombin complex structures. The distribution of thermal parameters and occupancies of water molecules are similar to that shown in Figure 28. All the residues of the CV-863-1 peptide are well defined in the electron density map. The average thermal parameter of the peptide is only about 35 Å².

Since the CV863-1 peptide has the same Pro2'-Phe*5' sequence as that of cyclotheonamide A, the structure of this segment was anticipated to be closely related to that of cyclotheonamide A (Figure 34). The phenyl ring of Phe*5' is rotated approximately 30° with respect to that of the D-Phe5' of cyclotheonamide A. CV863-1 is not a cyclic peptide, thus it is not required to possess chain reversals. Nonetheless, the α-keto groups in both structures are really identical, rendering the Phe*5' interactions with thrombin possible. The Phe*5'-thrombin interaction in the CV863-1 complex indicates a better binding mode for the phenyl ring since there is no main-chain restraint like that observed in cyclotheonamide A.

The "main chain" of BocO' is fitted in a cis-conformation in this structure, but the observed electron density is also large enough to fitting a trans-isomer (Figure 35). We decided to model the

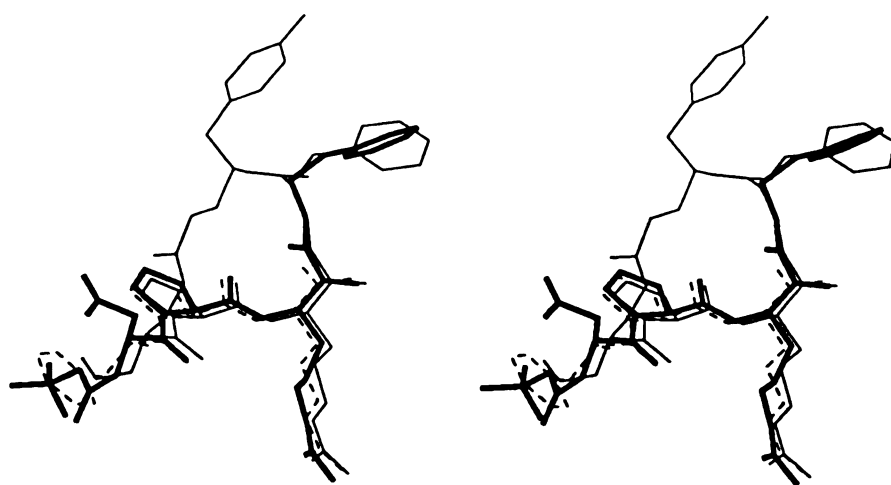


Figure 34. Stereoview of the CV863-1 peptide compared with cyclotheonamide A and PPACK. Cyclotheonamide A, light; CV863-1, dark; PPACK, broken.

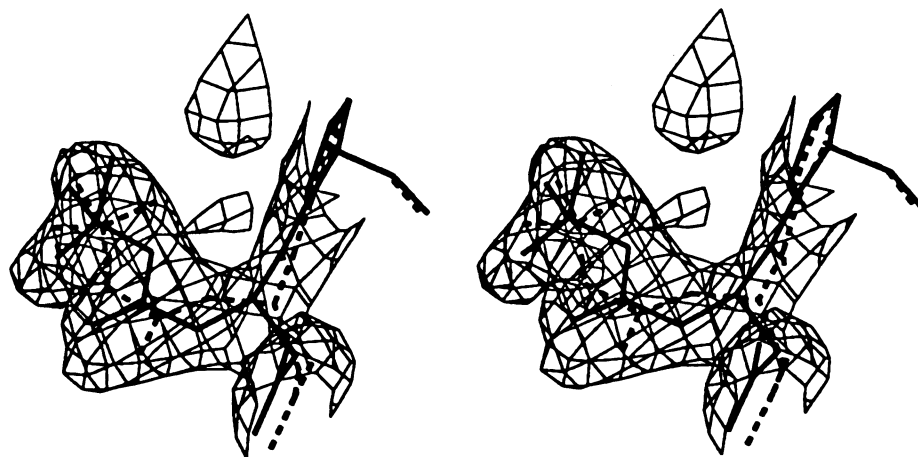


Figure 35. Stereoview of the electron density for BocO'-Pro2' of CV863-1. In bold is cis-conformation; in broken is trans-conformation.

structure as the *cis*-conformation since it appeared to fit the electron density. However, the energy barrier between the two conformations should not be as high as that in normal peptides. In fact, the crystal structure of thrombin complexed with a similar inhibitor (Unpublished result of this laboratory), which also contained a Boc group but was crystallized under a different pH value, exhibited electron density that are more suitable to be fitted as a *trans*-conformation. Therefore, the conformation in the crystal could be a mixture of both *cis*- and *trans*-isomers.

The main chain of Asp1' is also similar to that of the Ala*1' residue of cyclotheonamide A, although their side groups are different. As mentioned in Chapter 1, the Boc0'-Arg3' sequence is similar to the active site binding sequence (LDPR) of thrombin platelet receptor. There is a hydrogen-bond like contact (possibly a weak hydrogen bond) between CO4' O and Arg3' N (3.2 Å), as in the cyclotheonamide A structure.

(3) CV863-1-thrombin interaction

The Pro2'-Phe*5' sequence was designed to mimic the binding in cyclotheonamide A. The crystal structure of the CV863-1 complex shows that this was achieved. The above segment binds to thrombin in a very similar way as that of cyclotheonamide A and corresponds to the Pro2'-Arg3' part of PPACK (Figure 34, Figure 36 and Table 15). The Asp1'-Arg3' segment of the CV863-1 peptide is also antiparallel to the Ser214-Gly216 sequence of thrombin, with three hydrogen bonds linking the two segments (Table 15). The α -keto amide group of Arg3'-CO4' forms the same hydrogen-bonded transition state "tetrahedral intermediate" that binds to thrombin (Table 15)

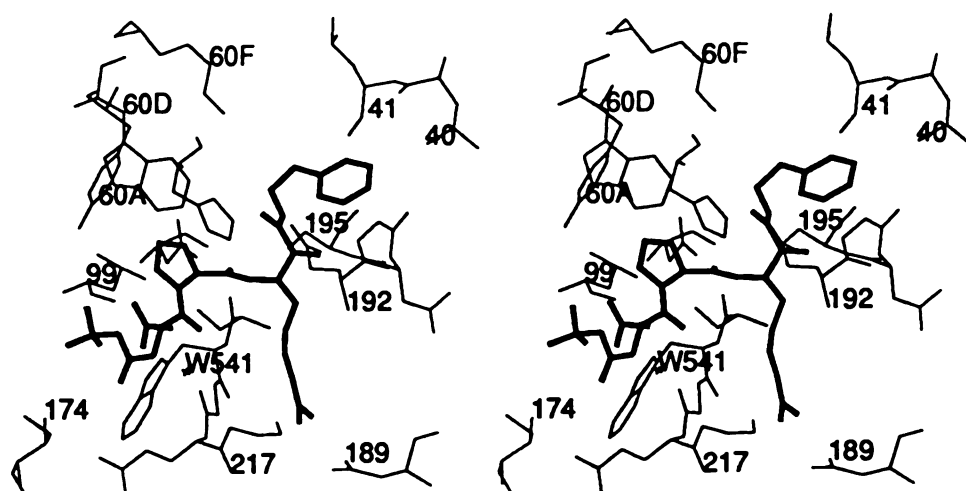


Figure 36. Stereoview of the thrombin-CV863-1 interaction. Thrombin, light; CV863-1, dark.

Table 15. Thrombin-CV863-1 interactions.

covalent:

CO4' C-Ser195 OG	2.0 A
------------------	-------

ion pair:

Arg3' NH1-Asp189 OD1	2.7 A
----------------------	-------

Arg3' NH2-Asp189 OD2	2.6 A
----------------------	-------

hydrogen bond:

Asp1' N-Gly216 O	2.8 A
------------------	-------

Asp1' O-Gly216 N	3.0 A
------------------	-------

Arg3' N-Ser214 O	3.1 A
------------------	-------

Arg3' O-Ser195 N	2.8 A
------------------	-------

Arg3' O-Gly193 N	2.8 A
------------------	-------

Arg3' N-Ser195 OG	2.9 A
-------------------	-------

CO4' O-His57 NE2	2.5 A
------------------	-------

Asp1' OD1-W541-Gly219 N	2.3 A, 3.4 A
-------------------------	--------------

van der Waals:

Boc0' CG1-Leu99 CD1	3.9 A
---------------------	-------

Boc0' CG2-Ile174 CD1	4.0 A
----------------------	-------

Boc0' CG3-Trp215 CD2	3.7 A
----------------------	-------

Asp1' C-Trp215 CB	3.7 A
-------------------	-------

Pro2' CB-His57 CD	3.5 A
-------------------	-------

Pro2' CB-Trp60D CH2	3.8 A
---------------------	-------

Pro2' CB-Leu99 CD	4.0 A
-------------------	-------

Pro2' CG-Tyr60A CE2	3.6 A
---------------------	-------

Phe*5' CE2-Leu40 O	3.9 A
--------------------	-------

Phe*5' CD2-Leu41 O	3.1 A
--------------------	-------

Phe*5' CE1-Glu192 CD	3.2 A
----------------------	-------

Phe*5' CE2-Gly193 CA	3.4 A
----------------------	-------

(dihedral angle $O=C-C=O$ 113°; angles Ser195 OG-Arg3' C-Arg3' O 91°, Arg3' CA-Arg3' C-Ser195 OG 95°, CO4' C-Arg3' C-Ser195 OG 91°).

The similarities notwithstanding several new features have been revealed by the CV863-1-thrombin interaction. First, since the phenyl ring of Phe*5' is able to rotate freely, it adopts an optimized position to interact with thrombin. Surprisingly, the aromatic ring interacts with only Leu40, Leu41, Glu192 and Gly193 (Table 15). The interaction with Trp60D, as observed in cyclotheonamide A, does not occur; even the hydrophobic interactions with Leu40 and Leu41 side chains have diminished. Moreover, hydrophobic interactions involving carbon atoms of the Glu192 side chain, which were available only to the tyrosine ring of TNT6' in the cyclotheonamide complex, are new to the phenyl ring of Phe*5' of the CV863-1 complex. The side chain of Glu192 is completely different in these two structures, which is not only due to the differences in their thrombin binding modes, but also due to the fact that the Glu192 side chain is very flexible and appears to be an active player in thrombin substrate binding [102,103].

The additional sequence of Boc0'-Asp1' was found to contribute to the CV863-1-thrombin binding. The side chain of Boc0' interacts with Leu99, Ile174 and Trp215 (Table 15), similar to the D-Phe residue in PPACK (Figure 34) but may not be as efficient because Boc0' makes less contacts at longer distances. The function of the Asp1' residue in the peptide sequence is likely to be for the placement of Boc0' (or a leucine in the platelet receptor sequence) side chain close to the D-Phe1' (of PPACK) binding region of the S2

apolar site, although there could be a water mediated weak hydrogen bond between Asp1' side chain and Gly219 (Asp1' OD1-W541 2.4 Å, W541-Gly219 N 3.4 Å). The cis-conformation of BocO' is not likely to be the conformation of the leucine in platelet receptor-thrombin binding. The CV863-1 is a much more effective thrombin inhibitor (K_d 0.5 nM) than cyclotheonamide A (K_d 180 nM), and is one of the most potent small molecule thrombin inhibitors. Nonetheless, the binding potency of the CV863-1 peptide will likely be improved with future designs based on structural results, e.g., replacing BocO' with a residue of a bigger hydrophobic side chain.

C. The PPACK-Thrombin Complex

(1) General

The PPACK-thrombin structure reported by Bode et al. [22] included 409 water molecules, which is a significantly different from that found usual in our laboratory. Therefore, one of the main reasons for our re-determination of the structure was to include fewer water molecules in the final refinement. The 409 water molecules exhibit thermal parameters ranging from 3 Å² to 108 Å² with the average value being 50 Å². Thus, we were inclined to believe that many of the 409 waters were unreliably determined, which would lead to problems if the structure was used for other related structures.

The intensity data set for the crystal of the PPACK-thrombin complex was collected using the above mentioned methods. Some critical parameters are listed in Table 16, in comparison with those reported by Bode et al. [22]. Although the PPACK-thrombin crystals do not scatter X-rays as well as the other crystals in this thesis, it still diffracts well beyond 2.4 Å resolution. The data set is

Table 16. Data collection, processing and crystal characterization of PPACK-thrombin complex compared to that of Bode et al.[22].

		of Bode et al.
power of X-ray	50 kV x 150 mA	---
collimator size	0.3 mm	---
crystal-detector	11.65 cm	40 mm
swing angle	14°	10°
scan range	0.2°	---
collecting time	120 sec.	---
average peak width	0.5°	---
resolution	2.4 Å	1.9 Å
# collected	51,362	66,592
# independent	13,334	16,910
# with $I/\sigma(I) > 2$	10,427	---
data completion	69%	65%
R_{merge}	0.064	0.101
crystal system	orthorhombic	orthorhombic
space group	$P2_12_12_1$	$P2_12_12_1$
# of molecules	4 per unit cell	4 per unit cell
a	88.26 Å	87.74 Å
b	68.02 Å	67.81 Å
c	61.22 Å	61.07 Å
V_m	2.4 Å ³ /dalton	2.4 Å ³ /dalton
protein fraction	51 %	51 %

--- Not available

complete up to 2.5 Å resolution, and contains half of the possible reflections between 2.4 to 2.5 Å resolution. The crystal was found to be identical to that of the originally reported PPACK-thrombin [22] (Table 16). Thus, the structure was solved by the isomorphous replacement method described in Chapter 2. The initial phases of this structure were approximated using the thrombin coordinates of hirulog 3-thrombin complex [32], which were then transformed to fit the PPACK unit cell. A transformation matrix was found by superposing the hirulog 3-thrombin structure on that of the PPACK-thrombin complex coordinates reported by Bode et al. [22]. The initial model includes residues Ser1E-Ile14K of the A chain, Ile16-Thr147 and Gln151-Glu247 of the B chain. The autolysis loop of thrombin and the PPACK peptide were not included in the first phase calculation and the starting crystallographic R factor was 0.37. Beginning with an overall thermal parameter of 30 Å² at 2.8 Å resolution, we carried out 30 cycles of PROLSQ refinement and improved the model to an R factor of 0.26 with individual thermal parameters. The initial ($2|F_O| - |F_C|$) electron density map showed good density for thrombin, its autolysis loop and PPACK, and the latter two also appeared in the ($|F_O - F_C|$) difference map. Models of PPACK and the autolysis loop were then fitted into the electron density using computer graphics. The refinement was continued gradually, increasing resolution (2.8, 2.4 Å) and water molecules, taking a similar course as that of the refinement of the cyclotheonamide complex structure (Table 8).

The final structure of the complex converged at $R = 0.144$ with an average thermal parameter of 28 Å² using only 208 water molecules

and 9,536 reflections from 7.0 Å to 2.4 Å resolution. The r.m.s. deviations from target values and R factor by resolution ranges are listed in Table 17. The ω -angles of 97% of the peptide bonds are with $\pm 6^\circ$ of planarity and only the same few residues reported before [25] are slightly out of the conformationally allowed regions (the Ramachandran plot is omitted because it is practically the same with Figure 26). The mean error in coordinates is estimated, as described in Chapter 2 [96], to be about 0.2 Å.

(2) The structure

In the PPACK-thrombin complex, the structure of thrombin is well defined in the electron density except for the few terminal residues as those in the earlier reported PPACK-thrombin complex [6] and hirugen or hirulogs structures (Thr1H-Glu1C, Asp14L-Arg15 and Glu247) [27,32]. In addition, residue Ser11 does not have side chain electron density. Omitting these few residues, which are loosely arranged on the surface of the protein, the r.m.s. differences between thrombin in the present complex and that of Bode et al. [22] are listed in Table 10. The values of 0.3 Å for main chain positions and 1.2 Å for side chain positions show a very good consistency of the overall thrombin structures in the two different crystals (including residues of the flexible autolysis loop). The differences of side chain positions are mainly due to larger acidic and basic side groups that are on the surface of thrombin, for example, Asp60E, Arg75, Arg77A, Lys81, Arg97, Arg126, Arg165, Lys169, Asp170, Arg173, Glu127 and Lys236. Interestingly, the few residues near S' subsites (to be discussed in Chapter 4), namely Arg35', Glu39, Gln151 and Glu192, also show shifts greater than average, which may

Table 17. Final least squares restraints/deviations and R factor statistics of PPACK-thrombin.

	Target Sigma	RMS Deviation
distances (A)		
bond lengths	0.020	0.017
angle lengths	0.035	0.045
planar 1,4	0.050	0.050
planes (A)		
peptides	0.023	0.016
aromatic groups	0.023	0.017
chiral volumes (A ³)	0.15	0.19
non-bonded contacts (A)		
single torsion	0.55	0.23
multiple torsion	0.55	0.27
possible H-bond	0.55	0.27
torsion angles (deg)		
planar	3	2
starggered	20	23
orthonomal	25	31
thermal parameters (A ²)		
main chain bond	1.5	1.1
main chain angle	2.0	1.7
side chain bond	2.0	1.6
side chain angle	2.5	2.5

d _{min}	Number of reflections	$\sigma(F_o)$	$\langle F_o - F_c \rangle$	R factor shell	R factor sphere
4.59	1358	28.2	64.2	0.156	0.156
3.84	1381	24.5	48.8	0.118	0.137
3.42	1355	21.9	43.5	0.127	0.134
3.11	1338	19.8	39.6	0.149	0.137
2.87	1360	17.8	32.5	0.155	0.139
2.67	1353	16.0	31.3	0.167	0.142
2.40	1391	14.0	27.1	0.160	0.144

diffraction pattern : $\sigma(|F_o|) = (18.0) + (-140.0)[\sin(\theta/\lambda) - (1/6)]$
 $\langle ||F_o| - |F_c|| \rangle = 41.0$

be a result of the higher-than-average solvent accessibility of these residues. The distribution of thermal parameters of the structure is also similar to that in Figure 27, with the terminals of thrombin having the highest thermal parameters.

We included only 208 waters in our refinement of the structure and reduced the R factor to 0.144 (0.156 from 8.0-1.92 Å in that of Bode et al.). The overall solvent pattern, including the channel of about 20 waters that connects the thrombin active site and the 186A-D insertion loop [25], is consistent with other structures solved in this laboratory, despite the fact that many are in different crystal systems. The distributions of thermal parameters and occupancies of water molecules are plotted in Figure 37, which clearly indicate generally good quality for water molecules.

A stereoview comparison of the PPACK binding region between this structure and that reported by Bode et al. [22] is shown in Figure 38. The two structures are almost identical in this region of the active site, and the PPACK-thrombin interactions are also similar in the two structures (Table 18). The values of the "tetrahedral intermediate" bonds are found to be almost identical to those of cyclothienoamide A and CV863-1 (Table 19); compared to about 1.6 Å in the Bode structure [22]. This difference is probably caused by the difference in two refinement methods, or simply is about the range of experimental errors (0.2 Å). The geometry of the tetrahedral intermediates of PPACK is comparable but not very consistent with the two structures (Table 19). The PPACK-thrombin complex structure is the only reported thrombin structure here that does not have an exosite inhibitor. It could have been a very useful structure

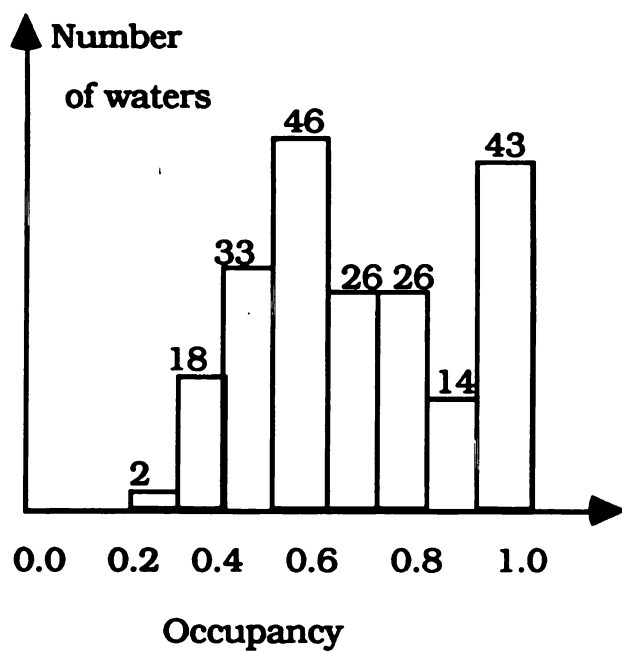
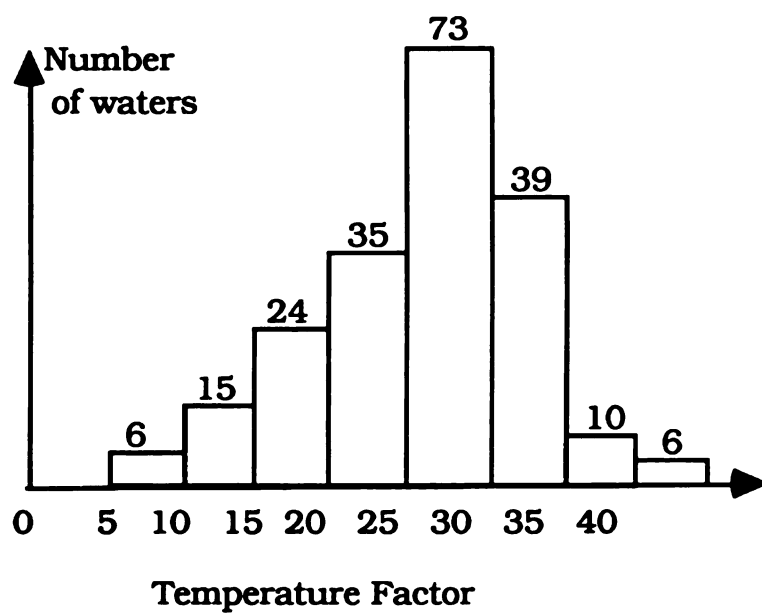


Figure 37. The distribution of thermal parameters and occupancies of water molecules in present PPACK-thrombin.

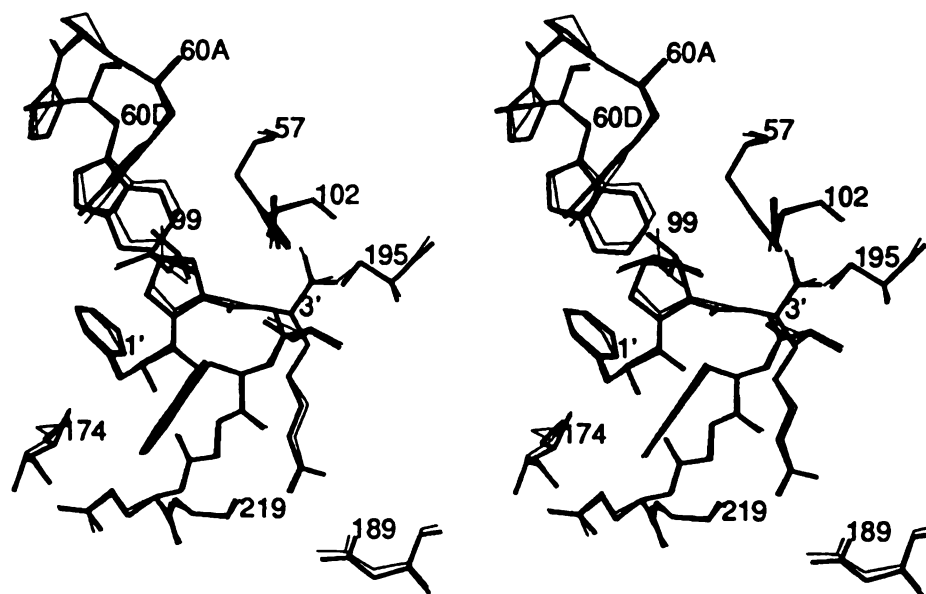


Figure 38. Stereoview comparing of the active site-PPACK region. Present PPACK-thrombin structure in dark; PPACK-thrombin structure of Bode et al. [22] in light.

Table 18. PPACK-thrombin interactions compared to those reported by Bode et al. (in brackets) [22].

covalent:

Arg3' C-Ser195 OG	1.8 (1.6) A
Arg3' CM*-His57 NE2	1.9 (1.4) A

ion pair:

Arg3' NH1-Asp189 OD2	2.9 (2.8) A
Arg3' NH2-Asp189 OD1	3.0 (2.6) A

hydrogen bond:

D-Phe1' N-Gly216 O	2.4 (2.7) A
D-Phe1' O-Gly216 N	2.9 (3.1) A
Arg3' N-Ser214 O	3.1 (2.9) A
D-Phe1' N-W460-Glu192 OE1	2.6,2.3 (3.0,2.7) A
Arg3' NE-W408-Gly 219 O	3.0,2.8 (3.0,2.9) A

van der Waals:

D-Phe1' CE2-Leu99 CD2	3.6 (3.5**) A
D-Phe1' CD2-Trp215 CB	3.5 (3.8) A
D-Phe1' CD2-Ile174 CD1	3.7 (4.0) A
Pro2' CB-Leu99 CD2	3.8 (3.8) A
Pro2' CB-His57 CD2	3.6 (3.5) A
Pro2' CG-Tyr60A CE2	3.8 (3.5) A
Pro2' CG-Trp60D CH2	3.9 (3.8) A
Arg3' CD-Trp215 C	3.4 (3.8) A
Arg3' CZ-Gly216 CA	3.8 (3.8) A
Arg3' CZ-Ala190 C	3.9 (3.6) A

***CM is the carbon atom of the original chloromethylketone group.**

****distances of van der Waals contacts in the Bode structure only designates the closest ones between the two residues, not necessarily between the same two atoms as indicated.**

Table 19. The tetrahedral intermediates in various structures.

	cyclotheonamide	CV863-1	PPACK	PPACK of Bode
D1 (Å)	1.8	2.0	1.8	1.6
D2 (Å)	2.7	2.5	1.9	1.4
$\alpha 1$ (°)	86	91	93	116
$\alpha 2$ (°)	97	95	107	114
$\alpha 3$ (°)	93	91	92	95

*D1 is the distance between Ser195 OG and Arg3' C.

*D2 is the distance between His57 NE2 and CO4' O (or Arg3' CM).

* $\alpha 1$ is the angle of Ser195 OG-Arg3' C-Arg3' O.

* $\alpha 2$ is the angle of Ser195 OG-Arg3' C-Arg3' CA.

* $\alpha 3$ is the angle of Ser195 OG-Arg3' C-CO4' C (or Arg3' CM) .

because certain conformational changes could have been observed by comparing the structure to others that are bound with an exosite inhibitor. Unfortunately, the structure can not reveal an unbound thrombin exosite, because the exosite interacts intermolecularly with a symmetry related neighboring molecule in the crystal structure [25].

D. The MDL-28050-Thrombin Complex

(1) General

The X-ray diffraction experiments performed on the crystal of the MDL-28050-thrombin complex were carried out using the methods as described previously. Parameters of the data collection are listed in Table 20, with those of the hirulog 3 complex referred for comparison. The data set is complete to 2.5 Å resolution, and contains half of the possible reflections between 2.2 to 2.5 Å resolution. The crystal was found to be isomorphous to those of the hirugen, hirulog 1 [27], cyclotheonamide A and CV863-1 complexes [33] (Table 20). Therefore, the isomorphous replacement method, described in Chapter 2, was applied to solve the structure. The initial model was built using the thrombin coordinates of the hirulog 3-thrombin complex [32], including residues Ser1E-Arg15 of the A chain, Ile16-Leu144 and Gln151-Glu247 of the B chain. The MDL-28050 peptide was not included in the first phase calculation, and the starting crystallographic R factor was 0.31 with an overall thermal parameter of 34 Å² at 2.8 Å resolution. We then proceeded with 22 cycles of PROLSQ refinement and improved the model to an R factor of 0.22 with individual thermal parameters. The first ($2|F_o| - |F_c|$) electron density map at 7.0-2.8 Å resolution was in good

Table 20. Data collection, processing and crystal characterization of the MDL-28050-thrombin complex compared to hirulog 3.

	MDL-28050	Hirulog 3
power of X-ray	50 kV x 100 mA	50 kV x 100 mA
collimator size	0.3 mm	0.3 mm
crystal-detector	11.70 cm	11.70 cm
swing angle	15° or 17°	12° or 15°
scan range	0.2°	0.2°
collecting time	75 sec.	75 sec.
average peak width	0.6°	0.6°
resolution	2.2 Å	2.3 Å
# collected	46,652	59,257
# independent	15,771	13,756
# with $I/\sigma(I) > 2$	13,885	12,167
data completion	75%	77%
R_{merge}	0.037	0.045
crystal system	monoclinic	monoclinic
space group	C2	C2
# of molecules	4 per unit cell	4 per unit cell
a	72.28 Å	71.44 Å
b	72.73 Å	72.10 Å
c	73.56 Å	73.07 Å
β	100.55°	101.02°
V_m	2.4 Å³/dalton	2.4 Å³/dalton
protein fraction	51 %	51 %

agreement for thrombin and the MDL-28050 residues; the latter was also prominent in the ($|F_o - F_c|$) difference map. There was no density for Trp148-Lys149E of the autolysis loop of thrombin, similar to the hirugen and hirulog 1 complexes [27]. A model of the clearly resolved part of MDL-28050 was then fitted into the electron density using computer graphics. Further refinements of the structure were carried out with a modified dictionary file that allowed proper assignment of restraints to β -cyclohexylalanine. The refinement was continued to gradually increased resolution (2.8, 2.5, 2.2 Å) and water molecules, following a similar course as that of the refinement of the cyclotheonamide complex structure (Table 8).

The PROLSQ refinement converged at $R = 0.145$ with an average thermal parameter of 29 Å^2 . The final structure includes 244 water molecules and 12,908 reflections from (7.0-2.2 Å) resolution. The r.m.s. deviations from target values and R factor by resolution ranges are listed in Table 21. The ω -angles of 97% of the peptide bonds are with $\pm 6^\circ$ of planarity and only the same few residues reported before [25] are slightly out of the conformationally allowed regions (Ramachandran plot is similar to Figure 26). The mean error in coordinates is estimated to be about 0.2 Å (see Chapter 2).

(2) Overall structure

Most of the thrombin residues are well defined in the electron density for the structure of the MDL-28050 complex. A few terminal and autolysis loop residues were found to be disordered, similar to those of the hirugen and hirulogs complex (Ser1E-Glu1C, Ile14K-Arg15, Glu247; Trp148-Lys149E) [27,32]. Excluding these few residues, the r.m.s. differences between thrombin in the

Table 21. Final least squares restraints/deviations and R factor statistics of MDL-28050-thrombin.

	Target Sigma	RMS Deviation
distances (A)		
bond lengths	0.020	0.019
angle lengths	0.035	0.049
planar 1,4	0.050	0.052
planes (A)		
peptides	0.023	0.019
aromatic groups	0.023	0.021
chiral volumes (A ³)	0.15	0.20
non-bonded contacts (A)		
single torsion	0.55	0.21
multiple torsion	0.55	0.30
possible H-bond	0.55	0.29
torsion angles (deg)		
planar	3	3
starggered	23	22
orthonomal	25	28
thermal parameters (A ²)		
main chain bond	1.5	1.3
main chain angle	2.0	2.1
side chain bond	2.0	2.1
side chain angle	2.5	3.1

d _{min}	Number of reflections	$\sigma(F_o)$	$\langle F_o - F_c \rangle$	R factor shell	R factor sphere
4.32	1808	34.9	87.2	0.154	0.154
3.57	1886	30.2	60.4	0.113	0.134
3.15	1853	27.0	52.8	0.136	0.134
2.86	1844	24.4	46.1	0.158	0.138
2.66	1790	22.2	40.9	0.164	0.141
2.49	1858	20.3	36.2	0.168	0.144
2.20	1869	18.2	33.9	0.163	0.145

diffraction pattern : $\sigma(|F_o|) = (24.5) + (-150.0)[\sin(\theta/\lambda) - (1/6)]$
 $\langle ||F_o| - |F_c|| \rangle = 51.0$

MDL-28050 and other thrombin complexes shows a very good consistency of the thrombin structure in different crystals (Table 10). The distribution of thermal parameters of the structure is also similar to that of Figure 27, with residues of MDL-28050 and terminals of thrombin having the highest thermal parameters. The overall solvent pattern is also consistent in all the structures. The distributions of thermal parameters and occupancies of solvent molecules are similar to that of Figure 28. The carbohydrate chain attached to Asn60G was also not defined in the electron density.

Most of the residues of MDL-28050 are well defined in the electron density. The side chains of Glu61' and Glu62', as well as D-Glu65', are directed into the solvent region and do not have electron density. There is a water molecule in the vicinity of Glu62', which could be part of the unresolved side chain electron density for Glu62'. The MDL-28050 peptide is in an extended conformation from Suc55' to Ile59', finishing with a 3_{10} helical turn from Pro60' to Cha64' (Figure 39). All the ω -angles of the peptide are close to planar and all the ϕ , ψ angles are in the conformationally allowed regions. The main chain of the first stretch of residues is similar to that of the corresponding residues at the hirudin C-terminal; while the helical turn structure is very different from that of hirudin (Table 10, Figure 40). Two hydrogen bonds (Pro60' O-Glu62' N, 2.5 Å; Pro60' O-Ala63' N, 2.9 Å) stabilize the helical turn folding. Moreover, since the side chain of Ile59' is less than 4 Å away from the side groups of Pro60', Ala63' and Cha64', the four residues form a small hydrophobic cluster that might help to strengthen the 3_{10} folding of the peptide.

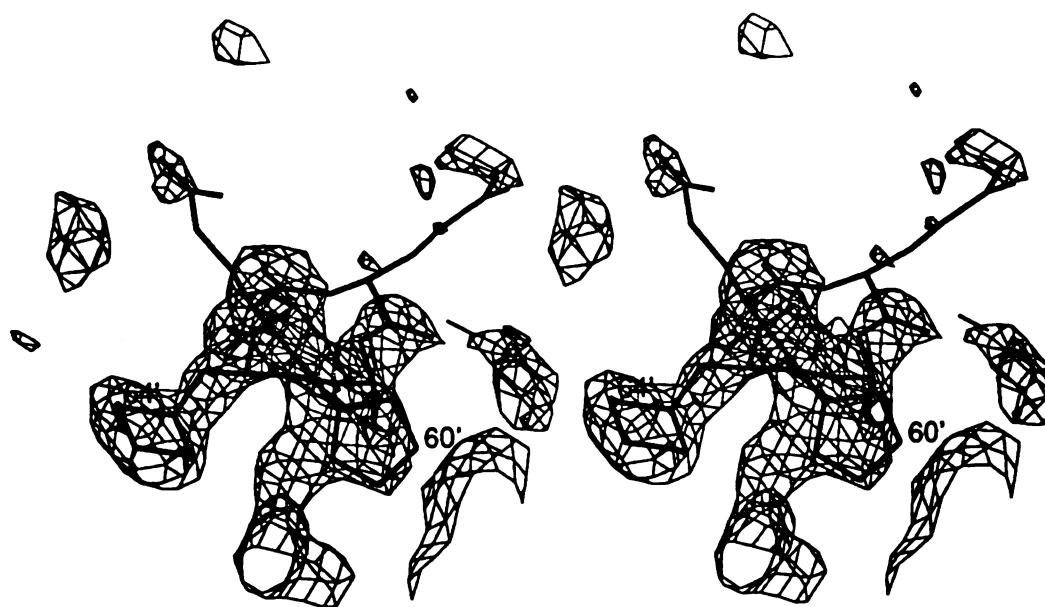


Figure 39. Steroview of the electron density of MDL-28050 near the helical turn region (Pro60'-Cha64'). Residues in bold are those of MDL-28050.

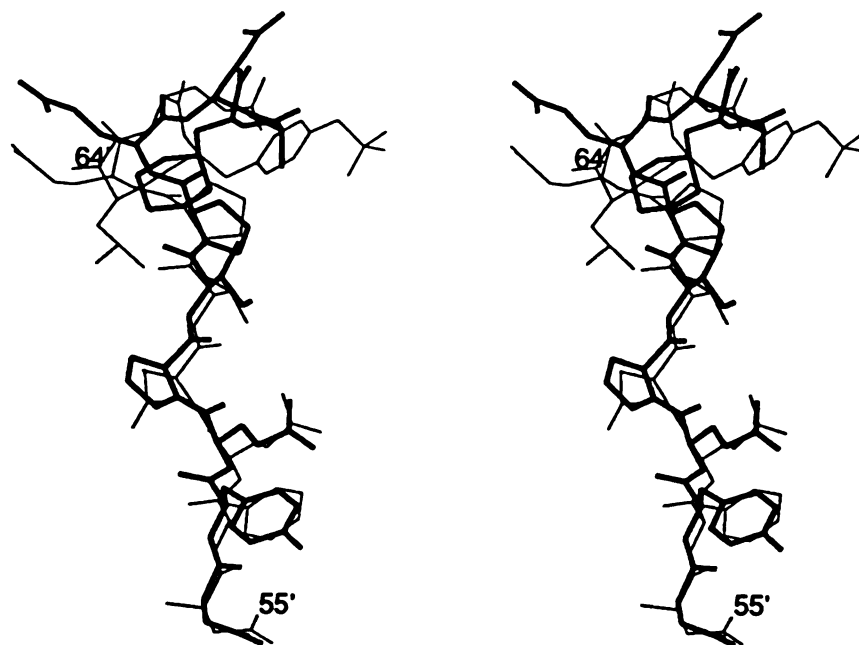


Figure 40. Stereoview comparing of MDL-28050 to hirugen in the structures of their respective complexes with thrombin. Hirugen, light; MDL-28050, dark.

(3) MDL-28050-thrombin interaction

The MDL-28050 peptide, like hirugen or hirudin55'-65', binds to an exosite of thrombin that is proposed to be the fibrinogen recognition "anion binding " exosite [25,27]. The exosite is a cliff-like wall formed by the Phe34-Leu40 and Arg73-Met84 loops of thrombin. The interaction of the MDL-28050 peptide with thrombin was expected to be identical to that of the hirudin C-terminal, with Suc55'-Cha64' of the MDL-28050 peptide surrounded by residues of these two loops.

The first stretch of MDL-28050 residues, Suc55'-Ile59', was almost as expected (Figure 41). The interactions include the two ion pairs observed in the thrombin exosite binding: (1) Suc55' makes a hydrogen-bonded salt bridge with Arg73 (Arg73 NH₂-Suc55' OD1, 3.5 Å; Arg73 NH₁-Suc55' OD2, 2.9 Å; 3.9 and 2.5 Å respectively in the hirugen complex); (2) Glu57' makes an ion pair with Arg75* of a crystallographically twofold related neighboring molecule (Arg75* NE-Glu57' OE1, 2.4 Å; Arg75* NH₁-Glu57' OE1, 3.2 Å), although the side chain of Arg75 shifts about 1 Å compared to that in the hirugen complex. In addition, several hydrogen bonding interactions are also conserved in this region (Glu57' N-Thr74 O, 3.1 Å; Glu57' O-W634-Arg67 NH₁, 2.6 and 3.0 Å). The presence of these highly conserved ion-pair and hydrogen bonding interactions explains the critical role of a glutamic acid at position 57' to the binding potency of the peptide [79]. Moreover, the hydrophobic interactions observed in this region are similar to those of the hirudin complex. There is a tyrosine at position 56' in MDL-28050, replacing the phenylalanine in hirudin, which interacts similarly with the aromatic side

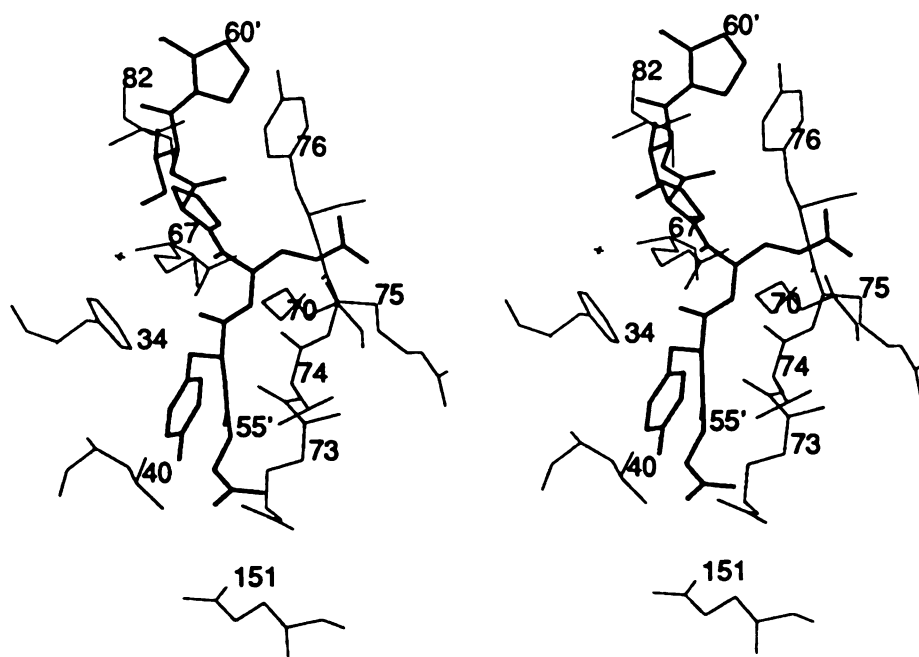


Figure 41. Stereoview of the Suc55'-Pro60' stretch of MDL-28050 interacting with thrombin. Thrombin, light; MDL-28050, dark. The small cross is a water molecule (W634).

group of Phe34 (Tyr56' CD2-Phe34 CE1, 3.5 Å). The phenolic oxygen atom sets up a hydrogen-bond like contact with Arg73 (Tyr56' OH-Arg73 NH₂, 3.4 Å). It is noteworthy that the thrombin platelet receptor also has a tyrosine in this position. The Ile59' residue interacts with the same thrombin residues as that of hirudin, having hydrophobic contacts with Ile82 (Ile59' CD1-Ile82 CD1, 3.5 Å), although the Ile59' side chain shifted notably (Figure 40). All this agrees with the observation that residues with hydrophobic side chains are most favored at positions 56' and 59' [79].

Interestingly, the structures of the hirudin, hirugen and hirulogs complexes [24,27,32] indicate that the the exosite interaction is predominately hydrophobic. Most of the anionic positions of the inhibitor point only towards the solvent and do not interact with thrombin (like Glu58', Glu61' and Glu62'). This observation contradicts the conventional terminology, "anion binding exosite", and also the fact that anionic character was used to predict the exosite binding of thrombin substrates by counting the number of aspartic and glutamic acids in the local sequence. Although there remain uncertainties regarding the precise role of the Tyr63' sulfation, it is clear that the side chain of Glu58' in hirudin or its derivatives is not involved in interaction with thrombin. Thus, a proline was designed at position 58' in MDL-28050 to replace the glutamic acid. As expected, the change did not alter the binding at this position and the main chain conformation remained the same (Figure 40). Considering the strong binding potency of this peptide to thrombin, residues with negatively charged side chains are not an essential component of the exosite binding. This explains why

position 58' tolerates residues with either charged or hydrophobic side chain without any significant change in binding constants [79]. Glu61' and Glu62' also have side chains disordered in the solvent in the crystal structure, substituting them by alanine only causes a slight decrease in binding potency [79]. This is to be expected from the nature of the positions 61' and 62', which are similar to that of position 58'.

Although the structures of hirudin and analogs complexed with thrombin are similar, there are new aspects to the MDL-28050 complex that were not anticipated in the original design of the peptide inhibitor. For instance, both hirudin and hirugen contain Asp55'. Since the MDL-28050 sequence starts at position 55', placing Suc55' at the position will not introduce a positively charged N-terminal as for a normal amino acid will. The alteration eliminates the possibility of a positive charge interfering with the binding between the negatively charged aspartic side chain and Arg73. Evidently, Suc55', along with Gln151 and Arg73, shifted by an average of 1.0 Å, which results in different binding details (Figure 41). The carboxy oxygen of Suc55' moves about 3.0 Å, pointing in the opposite direction to that of Asp55' of hirugen and forms a new hydrogen bond with Thr74 (Suc55' O-Thr74 OG1, 2.7 Å; 3.7 Å in the hirugen complex) and a hydrogen-bond like contact with Gln151 (Suc55' OD2-Gln151 NE2, 3.4 Å; more than 4.0 Å in hirugen). With the hydrogen-bonded salt bridge (Suc55' to Arg73) unchanged, the change of the sequence disrupted a few interactions, most notably, Suc55' OD1-Thr74 OG1, 3.7 Å, which was 3.2 Å in hirugen. Overall, Suc55' of MDL-28050 seems to provide a stronger binding with

thrombin than that of an aspartic acid. The binding potency was improved significantly when this change of sequence was effected [79].

The second stretch of MDL-28050 residues (Pro60'-Cha64') starts with a proline, which is often found prior to a helical secondary structure. Although notably shifted, Pro60' nevertheless makes the same hydrophobic contacts with Tyr76 of thrombin as in hirudin (Pro60' CG-Tyr76 CE2, 3.1 Å) (Figure 42). Therefore, position 60' seems to be playing two roles, one as an initiator for the helical turn structure, and the other as a contributor to the lipophilic binding of thrombin.

The most striking feature of the MDL-28050-thrombin structure is the 3_{10} helical turn at the C-terminal. The β -cyclohexylalanine was designed to mimic the lipophilic component of the helical turn sequence (Tyr63'-Leu64') of hirudin. As the observed in the fibrin clotting attests, as well as the crystal structure, the modification provides a good binding potency to thrombin, *but in a completely different binding manner*. The cyclohexane ring occupies the hydrophobic pocket of thrombin very differently than the tyrosine 63' ring of the hirudin sequence (Figure 40). The side chain of Cha64' is stacked in a large hydrophobic pocket, with the side chains of Leu65, Ile82, Met84, Ile59', Pro60' and Ala63' situated within 4.0 Å (Figure 42). There are virtually no polar interactions in the region, unlike the case of the sulfated tyrosine of hirugen, yet the overall binding potency remains high [79]. Considering the fact that D-Glu65', as well as side chains of Glu61' and Glu62', are disordered in solvent regions and do not contribute to thrombin binding, we

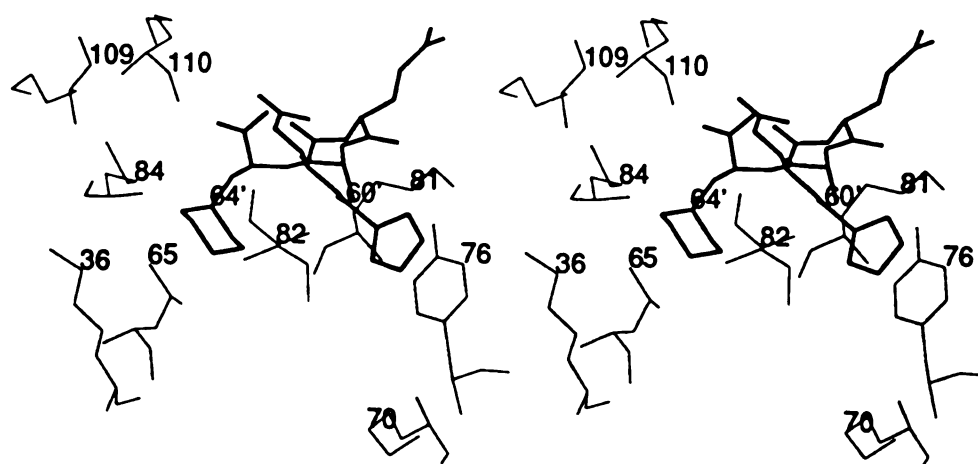


Figure 42. Stereoview of the Pro60'-Cha64' stretch of MDL-28050 interacting with thrombin. Thrombin, light; MDL-28050, dark.

conclude that the binding with thrombin in the exosite is mainly hydrophobic in nature. Hence, residues Ala63'-Cha64' are important to the binding potency, while Glu61', Glu62' and D-Glu65' are not [79]. These negatively charged residues may contribute to the long range recognition of positively charged thrombin residues near the exosite (Lys36, Arg67, Arg73, Lys81, Lys109 and Lys110), which has been predicted by electrostatic potential energy analyses of thrombin [104,105].

E. The Hirulog 3-Thrombin Complex

The diffraction data set of the hirulog 3-thrombin complex crystal was obtained as described previously (Table 20). The data set is complete to 2.5 Å resolution, and contains half of the possible reflections between 2.3 to 2.5 Å resolution. The crystal was found to be isomorphous to those of the hirugen, hirulog 1 [27], cyclotheonamide A [33] and MDL-28050 complexes [34] (Table 20). Thus, it was possible to solve the structure by the isomorphous replacement method. The phases of this structure were approximated using the thrombin coordinates of hirugen-thrombin complex [27]. The initial model includes residues Ser1E-Arg15 of the A chain, Ile16-Leu144 and Gln151-Glu247 of the B chain and the Lys145-Gly150 autolysis loop of the hirudin-thrombin complex [25]. The hirulog 3 peptide was not included in the first phase calculation and the initial crystallographic R factor was 0.31. The first ($2|F_O| - |F_C|$) electron density map at 7.0-2.8 Å resolution showed good electron density for thrombin and most of the hirulog 3 residues; the latter was also prominent in the ($|F_O - F_C|$) difference map. There was no density for Trp148-Lys149E of the autolysis loop of thrombin,

similar to the hirugen and hirulog 1 complexes [27]. A model of the clearly resolved part of hirulog 3 was then fitted into the electron density using computer graphics, and the unresolved autolysis loop of thrombin was excluded from the model. Starting with an overall thermal parameter of $34 \text{ \AA}^2$ at $2.8 \text{ \AA}$ resolution, refinement was continued to gradually increased resolution ($2.8, 2.5, 2.3 \text{ \AA}$) and water molecules (Table 8). The refinement of the structure was carried out with a modified dictionary file that allowed proper restraints to the β -homoarginine model.

The final structure of the complex converged at $R = 0.132$ with an average thermal parameter of $26 \text{ \AA}^2$ using 246 water molecules and 11,408 reflections from ($7.0\text{-}2.3 \text{ \AA}$) resolution. The r.m.s. deviations from the target values and R factor by resolution ranges are listed in Table 22. The ω -angles of 97% of the peptide bonds are with $\pm 6^\circ$ of planarity and only the same few residues [25] are slightly out of the conformationally allowed regions. The mean error in coordinates is estimated [96] to be $0.2 \text{ \AA}$.

(2) Overall structure

In the hirulog 3 complex structure, thrombin is well defined in the electron density except for the same few terminal and autolysis loop residues as those in the hirugen and hirulogs complex structures (Thr1H-Glu1C, Ile14K-Arg15, Glu247; Trp148-Lys149E) [27,32]. There is also some discontinuity of the electron density for the side chains of Asn62, Lys87, Gln151, Glu192 and Gln244. Excluding these few residues that are loosely arranged on the surface of the protein, the r.m.s. differences between thrombin in the hirulog 3 and several other complexes (Table 10) showed a very

Table 22. Final least squares restraints/deviations and R factor statistics of hirulog 3-thrombin.

	Target Sigma	RMS Deviation
distances (A)		
bond lengths	0.020	0.018
angle lengths	0.035	0.048
planar 1,4	0.050	0.053
planes (A)		
peptides	0.023	0.018
aromatic groups	0.023	0.019
chiral volumes (A ³)	0.15	0.20
non-bonded contacts (A)		
single torsion	0.55	0.21
multiple torsion	0.55	0.28
possible H-bond	0.55	0.29
torsion angles (deg)		
planar	3	3.1
starggered	20	25.2
orthonomal	25	32.2
thermal parameters (A ²)		
main chain bond	1.5	1.3
main chain angle	2.0	2.1
side chain bond	2.0	1.9
side chain angle	2.5	3.0

d _{min}	Number of reflections	$\sigma(F_o)$	$\langle F_o - F_c \rangle$	R factor shell	R factor sphere
4.40	1661	35.1	82.2	0.140	0.140
3.64	1701	30.5	63.0	0.109	0.124
3.22	1713	27.5	55.1	0.120	0.123
2.95	1612	25.1	46.9	0.140	0.126
2.73	1729	23.0	42.9	0.146	0.129
2.56	1537	21.1	38.6	0.149	0.131
2.30	1455	19.1	36.2	0.146	0.132

diffraction pattern : $\sigma(|F_o|) = (24.5) + (-150.0)[\sin(\theta/\lambda) - (1/6)]$
 $\langle ||F_o| - |F_c|| \rangle = 52.6$

good consistency of the overall thrombin structure in different crystals. The distribution of thermal parameters of the structure is also similar to Figure 27, with residues of hirulog 3 and terminals of thrombin exhibiting the highest thermal parameters. The distributions of thermal parameters and occupancies of water molecules are similar to that shown Figure 28. The carbohydrate chain attached to Asn60G is not defined in the electron density, although there are two peaks in the electron density, which correspond to two water molecules in size, but which are not large enough for the sugar hexose.

Most of the residues of the hirulog 3 structure are well defined in the electron density including those of the penta-glycine bridge (Figure 43) and most of those of the C-terminal helical turn that was disordered in the hirulog 1 complex [27]. The side chains of Glu61' and Glu62' are directed towards the solvent region and did not have electron density, while Tyr63' and Leu64' show some discontinuity. There are two water molecules near the Glu62' side chain and another one in the vicinity of Glu61', which could be part of their disordered side chain electron density. In the complex, hirulog 3 is in an extended conformation bending at Gly4' and Asp55', then finishing with a 3_{10} helical turn formed by the last four residues. It embraces a full diameter of the thrombin surface (Figure 44) displaying many interactions at both the catalytic site and the fibrinogen recognition anion binding exosite of thrombin that resemble those formed by an enzyme-substrate complex. The D-Phe1'-Pro2'- β -homoArg3' moiety and the C-terminal of hirulog 3 have similar conformational angles to PPACK in PPACK-thrombin and the

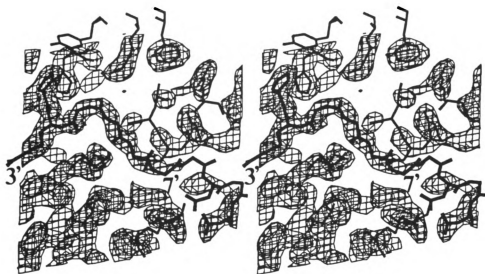


Figure 43. Stereoview of the final electron density near the pentaglycine region. Hirulog 3 in bold in electron density; β -homoArg3' and Gly7' designated; no density for Gly8'-Asn53'-Gly54', which is only modeled.

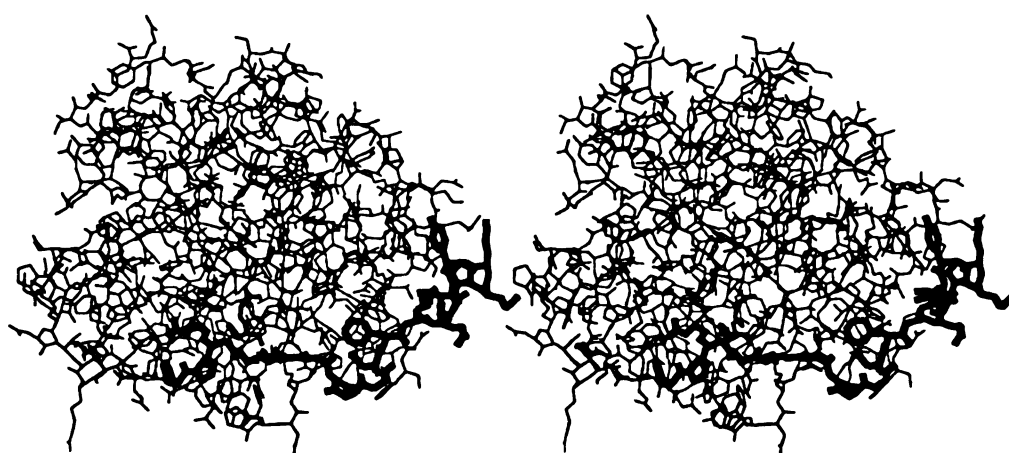


Figure 44. Stereoview of the hirulog 3-thrombin complex. Hirulog 3 and catalytic triad, bold; S1', S2'...subsites to immediate right of the triad; no density for Gly8'-Asn53'-Gly54', which structure is only modeled.

C-terminal of hirudin in the hirudin-thrombin complex, respectively (Table 10) [22,25]. All the ω -angles of hirulog 3 are close to planar with only three residues out of conformationally allowed ϕ , ψ regions; Glu62' and Tyr63' have only limited side chain electron density while interestingly, the residue Gly4', at $\phi = 100^\circ$, $\psi = -100^\circ$ is out of conformationally allowed regions. The conformational angles of Gly4' are acceptable for glycine, but this conformation will not be favored for amino acids with side chains. This may be related to the fact that it is adjacent to β -homoArg3' that displaces it by a bond length from a natural peptide sequence. Not surprisingly, other β -homoArg derivatives with non-glycyl 4' residues show reduced antithrombin activity [47].

Considering the resolution of the structure (2.3 Å), for which carbonyl oxygen atoms of the main chain are only marginally defined, there remains a certain degree of ambiguity in interpreting the detailed placement of the glycine residues in the electron density (Figure 43). Thus, although the glycine spacer can be traced confidently, the detailed orientation of each peptide bridge is not as certain as other parts of the structure with well-defined side chains. This has introduced some ambiguity in defining the S3'-S4' subsites extending from the active site.

As in hirudin, hirugen and hirulog 1 complexes [25,27], residue Gly8' and Asn53'-Gly54' have no electron density in the hirulog 3 structure (Figure 43). This was originally thought to be simple disorder but now appears to be the result of deamidation of Asn53'. Spontaneous aspartyl and asparaginyll deamidation, isomerization and racemization has been studied [81-83]. The reaction occurs in a

matter of days at physiological temperatures and pH and is sequence-dependent; Asn-Gly gives a maximum reaction rate. The reaction goes through a succinimide intermediate and produces structural isomers (Figure 45). The nonenzymatic deamidation of recombinant hirudin at Asn53'-Gly54' has also been recently characterized [83] showing the formation of both α and γ isomers through the differential cleavage of an unstable succinimide intermediate between Asn53' and Gly54'. The final $\alpha : \gamma$ (Asn/iso-Asn) ratio was found to be about 1:3. Since the sequence of Asn53'-Gly54' also occurs in hirulog 3, and the period and conditions of crystallization are favorable for deamidation, multiple isomeric deamidation products are probably involved in obscuring the resolution of the Gly8'-Asn53'-Gly54' sequence of the structure.

(3) Hirulog 3-thrombin interaction

The interaction of D-Phe1'-Pro2'- β -homoArg3' of hirulog 3 at the thrombin active site is similar to that of D-Phe1'-Pro2'-Arg3' in the PPACK and hirulog 1-thrombin structures with only minor positional shifts (Table 10, Figure 46) [22,27]. The D-Phe1' and Pro2' side chains are in the extensive hydrophobic region formed by Tyr60A, Trp60D, Leu99, Ile174 and Trp215 that is the apolar site of thrombin. Even with a β -homoarginine instead of an arginine at the S1 specificity site, the guanidinium group of β -homoArg3' nevertheless participates in a strong hydrogen bonding ion pair with Asp189, similar to a normal arginine (β -homoArg3' NH2-Asp189 OD1 2.7 and β -homoArg3' NH1-Asp189 OD2 2.9 Å). The side chain conformation of β -homoArg3' appears to be in a less favored rotamer state and the CA atom is noticeably shifted (Figure 46). Electron

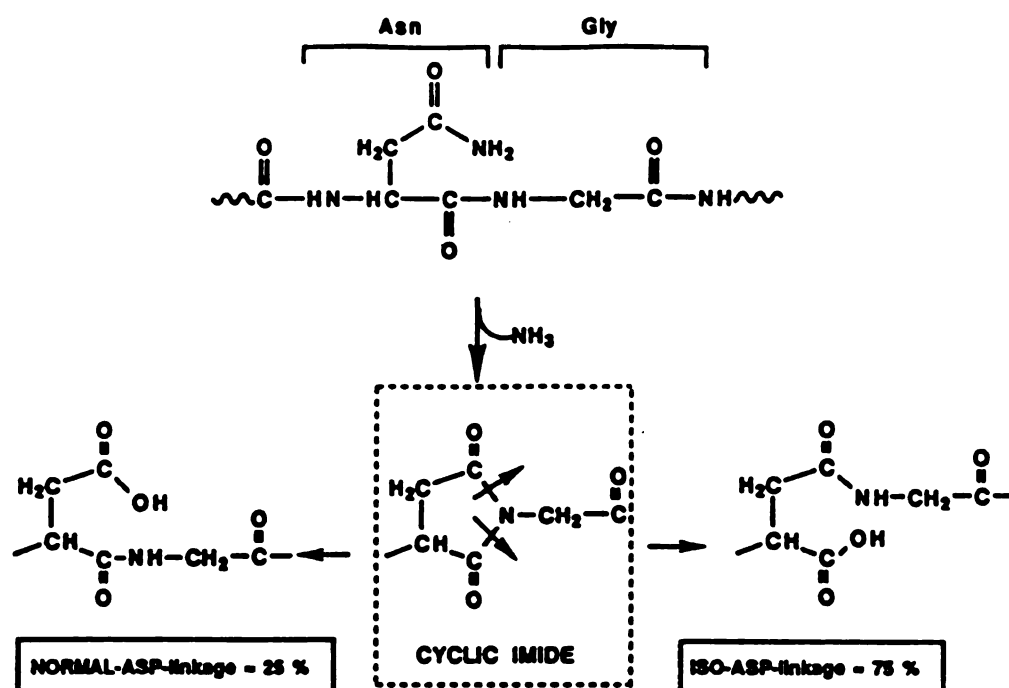


Figure 45. The nonenzymic deamidation of Asn-Gly sequence through the succinimide intermediate. Taken from [83].

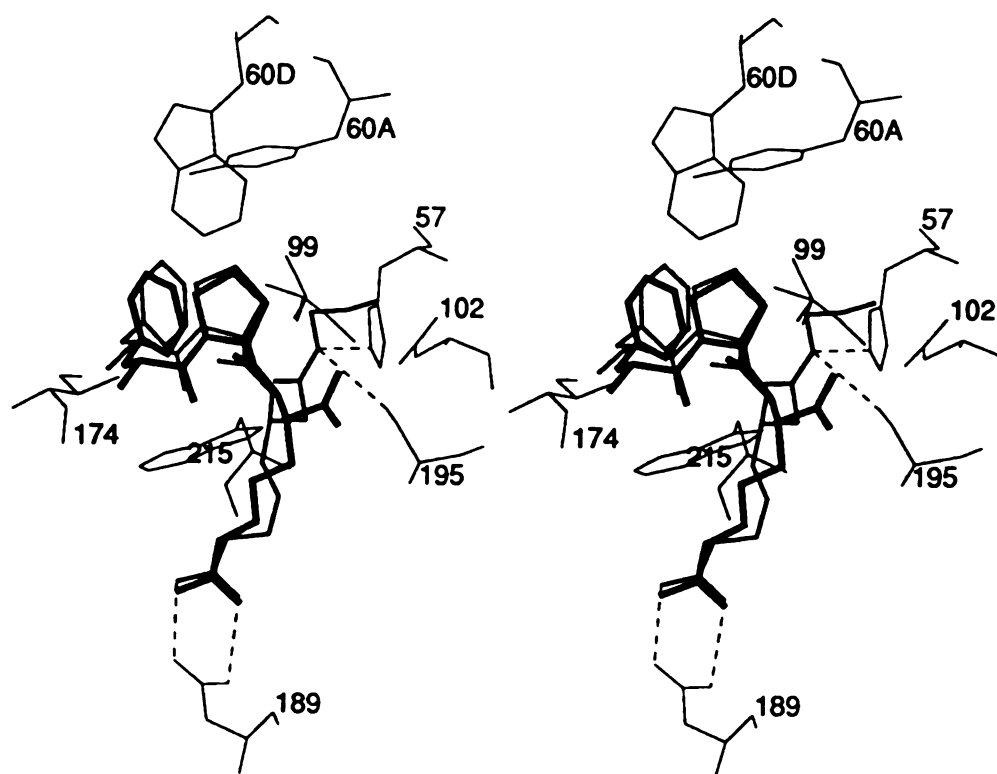


Figure 46. Stereoview comparing interactions in active site between **D**-Phe-Pro- β -homoArg-Gly in the hirulog 3 complex and PPACK in the hirulog 1 complex. Former in medium, latter in bold lines; hydrogen bonds in broken lines.

density maps calculated excluding this residue did not alter the shape of the density in the site suggesting that the different side chain conformation is probably an affect of the methylene insertion.

The most significant change in the active site of thrombin is associated with Ser195 OG. The χ_1 angle changes from -70° in hirugen-thrombin and -92° in hirulog 1-thrombin [27] to $+28^\circ$ in hirulog 3-thrombin. The χ_1 values in hirugen and hirulog 1 orient Ser195 OG toward homoArg3' CA, which is in the place of the scissile carbonyl group of a substrate (Figure 46). The change of more than 90° in χ_1 in the hirulog 3 complex results in the formation of a hydrogen bond between Ser195 OG of thrombin and Gly4' N of hirulog 3 (Figure 46, Table 16). With the concomitant disruption of the hydrogen bond between Ser195 OG and His57 NE2 of the catalytic site is the formation of one between Gly4' N and His57 NE2 in the hirulog 3-thrombin complex. Thus, in addition to being unable to effect catalysis, general base function is disrupted and considerable hydrogen bond reorganization takes place in the catalytic site as a result of the introduction of the methylene group at the scissile bond position.

The interaction of Asp55'-Leu64' of hirulog 3 at the fibrinogen recognition anion binding exosite is also similar to that of the hirudin C-terminal, hirugen and the resolved portion of the hirulog 1 structure (Figure 47) [25,27]. The most important interactions include: (1) Asp55' makes a hydrogen bonded salt bridge with Arg73 (Arg73 NH2-Asp55' OD1 2.6 Å, Arg73 NH1-Asp55' OD2 3.3 Å), (2) Glu57' makes an ion pair with Arg75 of a crystallographically two-fold related neighboring molecule, (3) Phe56', Ile59', Pro60',

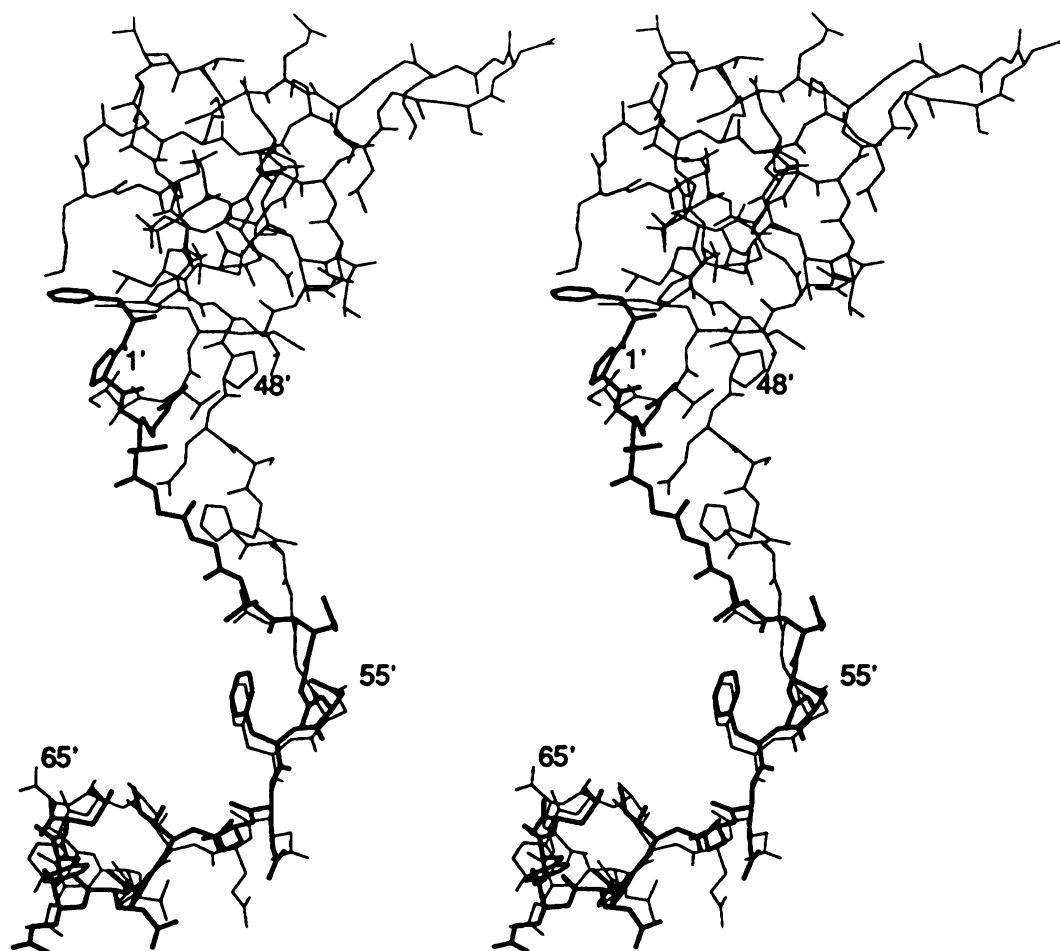


Figure 47. Stereoview comparing the structure of hirulog 3 and the C-terminal of hirudin in their respective thrombin complexes. Hirulog 3, bold; Gly8'-Asn53'-Gly54' and Asn52'-Asn53'-Gly54' of respective inhibitors only modeled; pertinent residues numbered.

Tyr63' and Leu64' are stacked in an extensive nonpolar cavity formed by Phe34, Gln38, Leu65, Ile82 and Tyr76. In addition, the distances of Lys81 NZ to Tyr63' OH and Lys36 NZ to the carboxylate group of Leu64' are less than 4.0 Å, so that Lys81 and Lys36 may be involved in aiding the docking of the C-terminal helical turn in a polar way even though the main interaction of the region is generally undirectional and hydrophobic.

In the hirudin structure, Glu58' makes an ion pair with Arg77A, the autolytic β -cleavage site of thrombin; in the hirugen structure, the Glu58' side chain is disordered and possibly interacts with the disordered autolysis loop. The hirulog 3 complex is also ambiguous in the area of the Glu58' side chain, which showed a lack of electron density for CB and CG atoms of its side chain but with density that may be solvent near a position that may be the carboxylate group. Thus, the protection against β -cleavage at Arg77A by hirugen and hirulog 3 appears to be more by physical obstruction rather than by specific interaction. However, the absence of Gln65' and lack of sulfation of Tyr63' in hirulogs does not significantly affect the binding interaction of hirulog 3 even though sulfation produces a ten-fold increase in binding avidity [106,107]. The crystal structure of the hirugen-thrombin complex reveals that although the sulfato oxygen atoms of Tyr63' hydrogen bond with thrombin directly or through mediating water molecules, these oxygen atoms do not ion-pair with oppositely charged residues of thrombin. An unexpected result from the hirulog 1-thrombin crystal structure was that the Pro60'-Leu64' residues of the peptide are disordered. Consequently, detailed structural evaluation of the lack of Tyr63' sulfate and lack of Gln65'

could not be made. In the hirulog 3 structure, Pro60'-Leu64' are fairly well-resolved although the quality of the electron density is not as good as that in hirugen. Examining the helical turn in hirulog 3 shows that the binding region is mainly hydrophobic and that the sulfation of Tyr63' augments the binding. Thus, the disorder of the region in the hirulog 1-thrombin complex is not the result of lack of sulfation or lack of a Gln65' C-terminus as was previously speculated [27].

The methylene insertion in β -homoArg3' of hirulog 3 prevents proteolysis. Moreover, the first four glycine residues of the pentaglycine spacer (Gly4'-Gly8') are well resolved in the electron density (Figure 44). This enabled the first structural documentation of the P1' to P4' positions of a thrombin substrate bound at the S1' to S4' subsites. Hirugen [27], PPACK [22] and several other active site inhibitors [28, 108] do not interact with these subsites, while hirudin takes a different path out of the active site where it binds through Ile1'-Tyr3' [25]. In the case of hirulog 1, the Arg3'-Pro4' bond is cleaved, like that of a substrate, with the glycine spacer bridge to Asp55' being totally disordered in the structure [27]. The importance of the S1' to S4' sites in the binding of fibrinopeptide A (FPA) fragments is indicated by the decrease of K_d by an order of magnitude with the presence of the P3'-P4' residues [109]. The addition of two or three more amino acids of the fibrinogen peptide essentially does not alter K_d further; however, this might be dependent upon the nature of the P5' to P7' side chains at these positions.

The structure of the penta-glycine spacer of hirulog 3 and nearby thrombin residues is shown in Figure 48. After an initial turn at Gly4', the remaining spacer residues are in an extended conformation, and since they have no side chains, their interaction with the host only involves main chain atoms. The position of Gly4' is firmly anchored by three hydrogen bonds with His57, Lys60F and Ser195 (Table 23) and the S1' site inferred from the vicinity of the CA position of Gly4' is lined, in part, by the "back-side" of the apolar S2 subsite (His57, Tyr60A, Trp60D). The lysyl side chain of Lys60F, some 5.0 to 6.0 Å from the aromatic residues, completes closing off the region to produce a small cavity. The carboxylate oxygen atom of the next glycine of the spacer hydrogen bonds weakly to Gly193 N (Table 16). Although the side group of Glu192 is also nearby, it is not clear whether the latter contributes to the Gly5' interaction since there is no electron density for its carboxylate group. In fact, reorienting the side chain of Glu192 may place the carboxylate group at an alternate position that was interpreted as two water molecules. However, in this case there would be no density corresponding to the CB and CG atoms of Glu192. Thus, Glu192 may be distributed between two alternate positions. A critical role for the position of Glu192 is suggested in protein C activation by the thrombin-thrombomodulin complex, which displays sensitivity to the substrate sequence binding to thrombin [102]. This is also in agreement with our own observations of the crystal structures of several different substrate sequences [103]. In either position, Glu192 appears to be too distant to interact with the CA vicinity of P2' (the present position of Gly5' CA). Since the position of the latter residue cannot

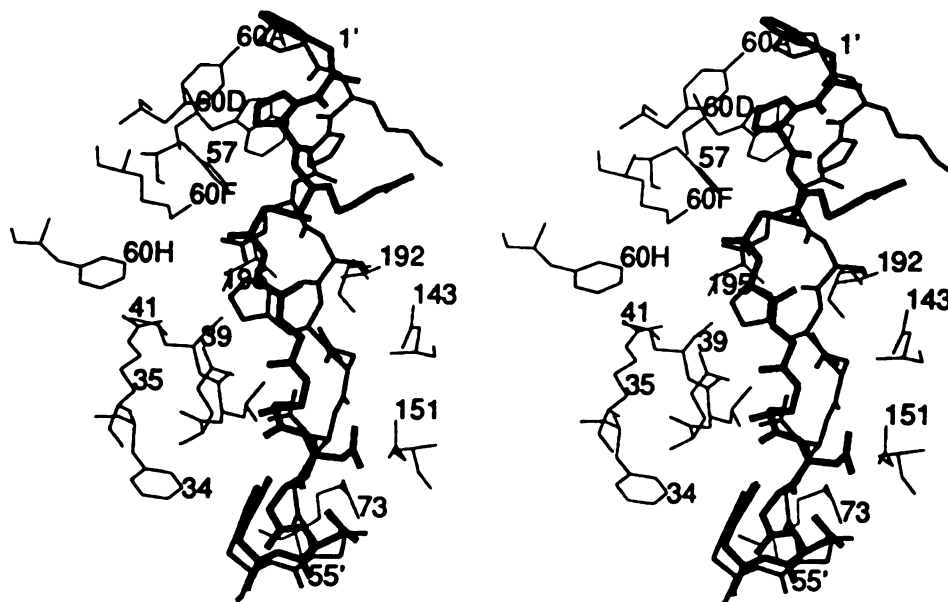


Figure 48. Stereoview of S1' to S4' subsites of thrombin. Hirulog 3 (D-Phe1'-Phe56'), dark; hirudin (Pro46'-Phe56'), medium; thrombin, light. Residues Gly8'-Asn53'-Gly54' and Asn52'-Asn53'-Gly54' of respective inhibitors only modeled.

Table 23. Interactions between hirulog 3 pentaglycine spacer and thrombin residues.

Donor	Acceptor	Distances (Å)	Type
Lys60F NZ	Gly4' O	3.0	hydrogen bond
Gly4' N	Ser195 OG	3.1	hydrogen bond
His57 NE2	Gly4' N	3.0	hydrogen bond
Gly193 N	Gly5' O	3.2	hydrogen bond
Gly6' N	Leu41 O	3.5	van der Waals
Leu40 N	Gly7' O	2.6	hydrogen bond
Gly7' O	Glu39 OE2	3.6	van der Waals

be significantly altered without affecting Gly4', Glu192 most likely is not involved in the S2' subsite.

The residues of Lys60F on one side, Leu41, Phe60H in back, and Glu39 from another side form a much larger cavity than that of the S1' site (Figure 48). This region, which appears capable of harboring two moderately sized side chains, appears to constitute the S2'-S3' sites of thrombin. In the hirudin complex, Glu49' and His51' side chains occupy these positions from a distance and participate in a complicated multi-charged interaction involving three positive charges from Arg35 (mediated by a water), Lys60F, His51' and negatively charged ones from Glu39, Glu49' and possibly Glu192, while Ser50' extends out into solvent and is near Glu192 [25]. Thus, the larger cavity might actually correspond to the S2' and S4' sites with the S3' position of substrate simply involving a surface

interaction with thrombin in the vicinity of Asn143 and Glu192, which is in agreement with the effects of substituting glutamine for Glu192 in thrombin. However, Glu39 also appears to be a component of the S3' subsite, at least when P3' is an aspartate [102].

The spacer interaction continues from Gly5' with an extended hydrogen bond-like contact between Gly6' N and Leu41 O (Table 16), while Gly7' is hydrogen bonded to Leu40. If Gly7' had a side chain, it would be directed towards solvent in the present conformation, similarly to Ser50' of the hirudin-thrombin complex, but with a twist of the peptide in one direction or another, it could interact with Glu39 on one side or Asn143 and Gln151 on the other. Since there is precedence for the former interaction in the hirudin-thrombin complex, Asn143 seems more suitable for assisting S3' site interactions, while Gln151 is closer to Asn53' at the S6' site. The electron density of the last glycine spacer residue deteriorates in the vicinity of the Asn53'-Gly54' disorder (Figure 44) rendering the S5'-S7' sites unclear.

CHAPTER 4. DISCUSSION AND ACCOMPLISHMENT

A. Active Site

In studying the detail interactions of thrombin with fibrinogen or other thrombin substrates, as well as in the search for more effective antithrombotics, numerous crystal structures of thrombin-peptide complexes were undertaken [22-34]; the main target has been the active site of thrombin. The thrombin complex of PPACK is the first available model for the thrombin active site binding [6,22,27,32], followed by those of hirudin [24-26], benzamidine [28,31], NAPAP [28,31], MPQA (MD-805) [28,31], fibrinopeptide A [29,30], 4-TAPAP [31], DAPA [103,108] and FPAM (an FPA peptidomimic) [110] (Figure 49). Among them, the coordinates of the PPACK, hirudin, DAPA, FPA and FPAM complex structures are available in this laboratory for quantitative comparisons. As discussed in the previous chapters, complexes of cyclotheonamide A and CV863-1 with thrombin provide a novel type of binding because of the distinct α -keto amide group and the new chain reversal possible with it [33].

Thus far, the fibrinopeptide A-bovine α -thrombin complex [29,30] most thoroughly reveals the interaction in the active site of thrombin. A schematic drawing of the FPA-thrombin interaction is shown in Figure 50, where Ser195 and His57 belong to the catalytic triad of serine proteinases. The trypsin-like S1 specificity pocket ends at Asp189 that favors an oppositely charged arginyl or lysyl side group, which has been observed in all the crystal structures of thrombin complexed with active site binding sequences [22,27-33]

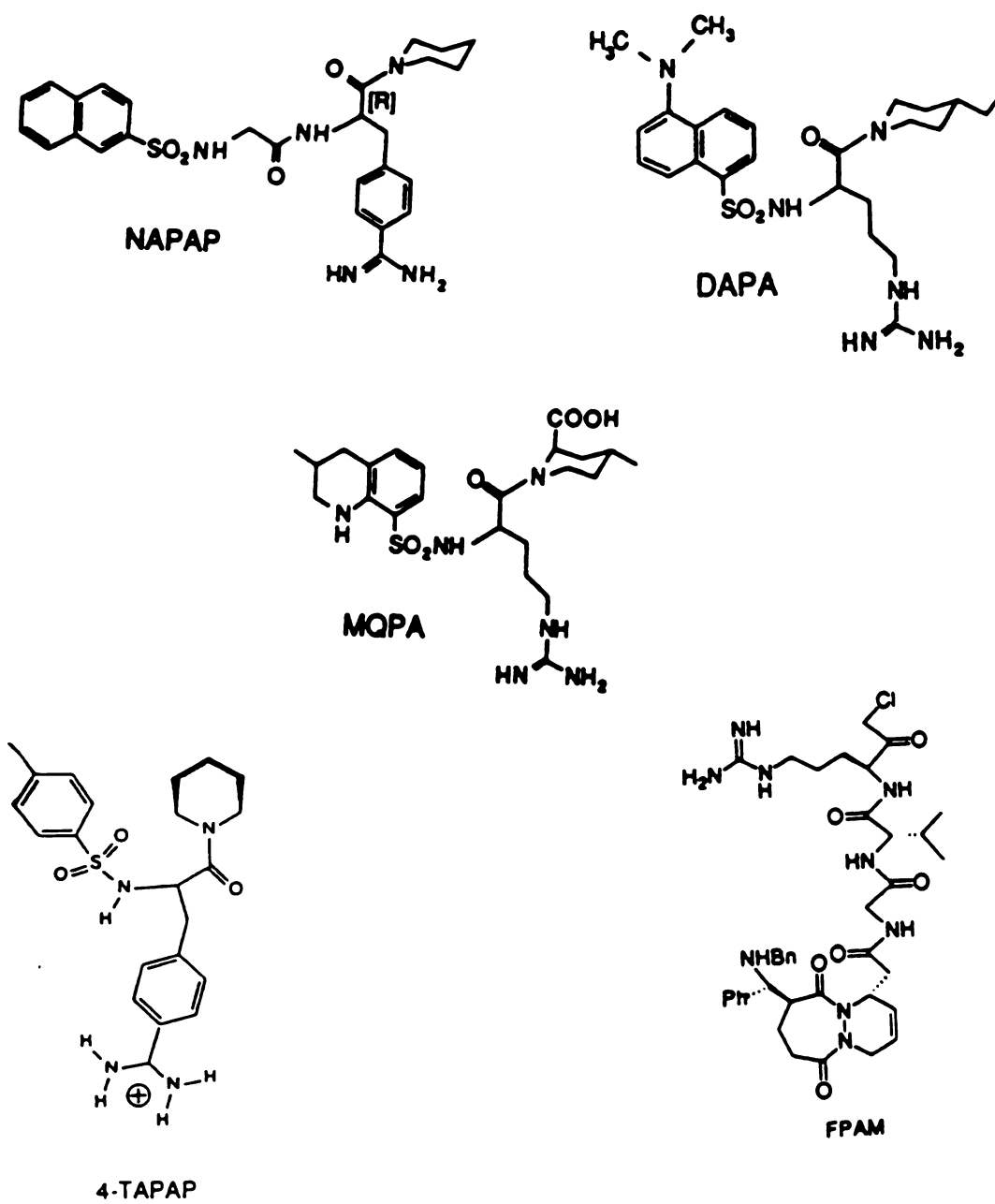


Figure 49. Chemical formula of thrombin active site inhibitors.

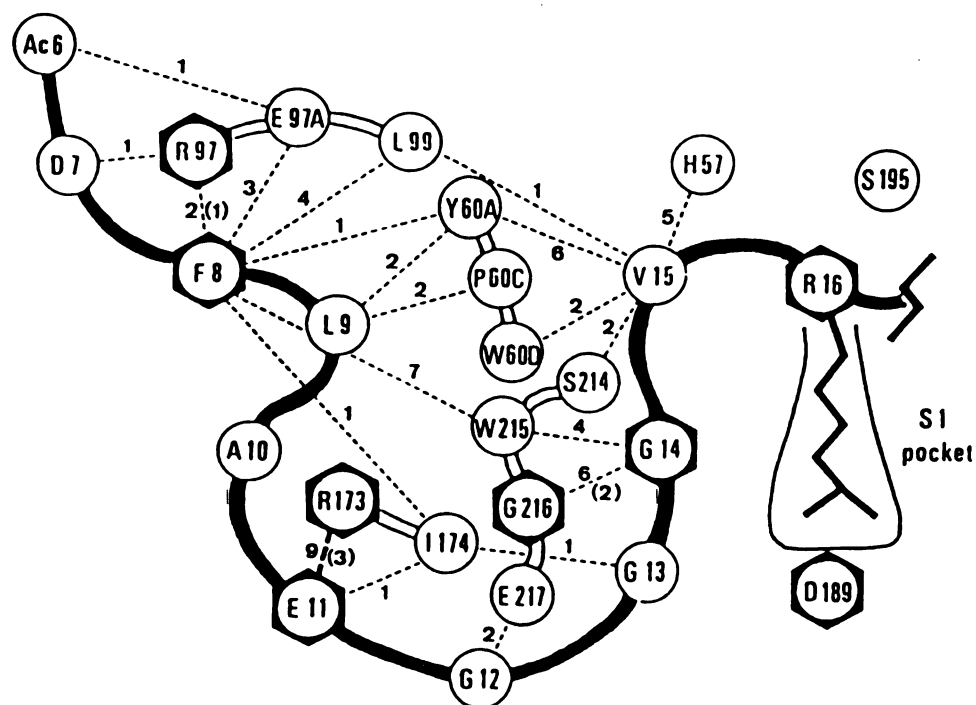


Figure 50. Schematic drawing of the thrombin-fibrinopeptide A interaction. Residues (polar in hexagons) of FPA are connected by the heavy line. Interactions are designated in broken lines and the number of close contact (less than 4.0 Å) are given (polar interactions in brackets). Taken from Stubbs et al.[30].

except for hirudin, in which the S1 specificity pocket is unoccupied while Ile1'-Tyr3' binds only to an apolar site [25] of thrombin. The apolar site has been proposed long before any crystal structures became available [111], and is most likely composed of the apolar side groups of His57, Tyr60A, Trp60D, Leu99, Ile174 and Trp215 [22]. Unlike the ion-pair interactions at the end of the thrombin S1 pocket that require a precise positioning of the arginine with respect to Asp189, the inhibitor binding at the apolar site is more flexible because of the nature of hydrophobic interactions [108]. Since hydrophobic residues (valine in FPA or proline in PPACK) appear frequently at the P2 position of thrombin inhibitors or substrates, the S2 subsite of thrombin is defined as one side of the apolar cavity. The apolar site is used differently in the PPACK complex because of the D-Phe1' group at the P3 position, which, by virtue of its enantiomorphous conformation, reaches into the other side of the apolar site, thus defining this half of the apolar cavity as the S3 subsite of thrombin. Other examples, like NAPAP, MPQA, 4-TAPAP and DAPA, are also distinct in their P2 and P3 interactions [28,31,103] with thrombin subsites. Therefore, the definition of thrombin subsites is sequence-dependent. The S2-S11 subsites of thrombin are defined, in this chapter, according to that of the FPA-thrombin complex (Figure 50); thus the half of the apolar site, which is occupied by the D-Phe1' group of PPACK in PPACK-thrombin, is defined as S8 and S9 subsites (Figure 50).

Crystal structures of the cyclotheonamide A and CV863-1 complexes reveal a number of novel features of thrombin active site binding, mostly because of the intriguing α -keto amide group in both

inhibitors. The α -keto amide group forms a complicated hydrogen-bonded transition state "tetrahedral intermediate" in its interaction with thrombin, which most likely accounts for a large portion of the total binding potency of the inhibitor (Figure 33). Although "tetrahedral intermediates" in serine proteinase inhibitor-serine proteinase interaction are not new [73-75], those of the α -keto groups proved to be distinct from those of other serine proteinase inhibitors (Table 19). The α -keto group also forms similar but different interactions from those of the immunosuppressant ligand-immunophilin complex that mimics the transition state for the interconversion between peptidyl-prolyl cis- and trans- amide rotamers [59,63].

The unnatural D-Phe1' residue in PPACK binds to one side of the apolar site (left side in Figure 50). Fibrinopeptide A is known to have a β -turn that brings a reversal of the amino acid main chain and enables its P8-P9 residues bind to the same site [112]. The extra carbonyl group of the α -keto amide group (CO4'), as well as the TNT6'-Ala*1' connection, also produce a reversed chain segment (D-Phe5'-TNT6') in cyclotheonamide A, but in a clearly different manner (Figure 51). Since the D-Phe5'-TNT6' segment is closer to the S1'-S3' subsites that were defined earlier in the hirulog 3-thrombin structure, its interaction with thrombin will be discussed later. The excellent binding potency of CV863-1, a synthetic antithrombotic drug taking advantage of the novel binding features revealed by the cyclotheonamide A-thrombin interaction, is the direct evidence of the importance of structural results in drug design.

Because uninhibited α -thrombin will autolysis to γ -thrombin,

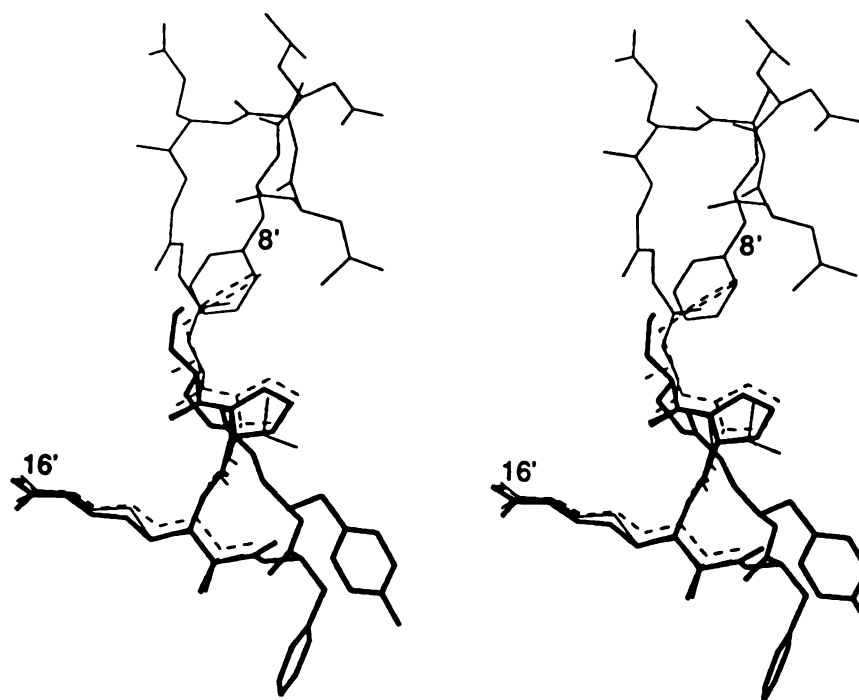


Figure 51. Stereoview comparing the structure of cyclotheonamide A to that of fibrinopeptide A and PPACK. Residues Phe8' and Arg16' of FPA (light) labeled. Cyclotheonamide A, dark; PPACK, broken.

the crystal structure of native α -thrombin has yet to be determined. The only structures with an unoccupied active site are the hirugen and hirullin (sequence is in Table 3) complexes [27,34], and the MDL-28050 structure. The structure of the hirulog 3 complex showed a PPACK sequence in the active site, but the crystal was found to be in the same packing field. More information about the behavior of solvent molecules in the active site can be obtained by comparing these three structures and the hirulog 3 structure.

The positions of the thrombin active site residues are consistent in all the structures, except for the OG atom of Ser195, which displays small differences in each. The side chain of Glu192 also seems to be fluctuating in all the structures (unpublished results of this laboratory), but the residues in the 186A-D insertion loop are all in the same position. This seems to imply that any conformational change in the thrombin active site will be small upon exosite binding. The channel of some 20 water molecules that link the active site to the 186A-D loop is also conserved in these structures [25]. The specificity pocket is filled with some of these solvent molecules when not occupied by an inhibitor. The binding of PPACK to the active site seems to effect a displacement of these solvent molecules. There appear to be three conserved water molecules (not four as reported previously [27]), originally hydrogen bonded in the vicinity of Asp189 side chain, that are displaced by the arginyl side chain upon inhibitor binding. In addition, at least two conserved water molecules are displaced by the phenyl ring of the D-Phe residue. It is possible that additional water molecules are also be displaced, although the result is not consistent in all the complexes. The

inhibitor binding in the active site of thrombin may only be displacing these solvent molecules with only small adjustments of the catalytic residues Ser195, His57 and Asp102 and without drastic conformational changes, although an intact thrombin structure is needed for a conclusion to be reached.

B. Fibrinogen Recognition Exosite

Since the determination of the crystal structure of the hirudin-thrombin complex [24], extensive studies have been carried out to explore the details of the binding interactions at both the active site and fibrinogen recognition "anion binding" exosite of thrombin [108]. Nearly all of the studies reported thus far targeting the exosite have been limited to close analogs of hirudin [79,113-117]. It has been suggested that the exosite is not only acting as a recognition site for thrombin enzymatic activities but also acting as a regulatory site for the non-enzymatic functions [115,116]. Examination of some of the various exosite sequences of natural thrombin macromolecular substrates (Table 24) clearly shows that there is still much to understand about binding in the thrombin exosite. The structure of the MDL-28050-thrombin complex, along with the structure of thrombin with hirullin50'-62' (Table 3), which has the most distinct exosite-binding sequence among hirudin variants [80], provide alternative sequences that bind well in the exosite. Thus, the structures may reveal new insights concerning the nature of the exosite. Moreover, the structures could also provide valuable information for future design of thrombin inhibitory drugs.

Comparing the peptide structure of MDL-28050 to that of hirugen (Table 10, Figure 40) shows that they seem to have two

Table 24. Definite exosite binding sequences of thrombin macromolecular substrate from human plasma.

Protein*	Sequence
Fibrinogen Aα 27-50	AAKDSDWPFCSDWDWNYKCPSGCR
Fibrinogen Bβ 66-80	LHADPDLGVLCPTGC
Fibrinogen γ 8-23	CCILDERFGSYCPTTC
Platelet Receptor 49-61	NDKYEPFWEDEEK
Thrombomodulin 408-426	ECPEGYILDDGFICTDIDE
Heparin Cofactor II 57-75	DDDYLDLEKIFAEDDDDYID

* See [76,78,118,119] for references.

parts. The first part, corresponding to hirudin55'-59', has peptide conformations that are similar in both structures. The remainder, corresponding to hirudin60'-64', is quite different in the two. Except for position 58', which is not utilized for thrombin binding and tolerates substitutions, the peptides of the first part are engaged in a concerted and efficient binding pattern with thrombin (Figure 41). As a result, the nature of the residues in these positions has been very specific and highly conserved. Acidic residues were frequently found at positions 55' and 57', while hydrophobic (or aromatic) residues often appear at positions 56' and 59' [79, 80]. The sequence combination appears to be unique and is most likely used for the specific recognition of the exosite binding.

The hirudin54'-60' peptide was reported to have reasonable anticoagulant and thrombin binding activities. It is capable of protecting residues Lys36, Lys70, Lys81, Lys109 and Lys110 of thrombin from chemical modification, whereas protection by the hirudin54'-58' peptide is virtually undetectable [114]. It has been shown by structural studies that the lysine residues are located on the surface of thrombin; prior to the binding of exosite inhibitors, these may play an important role in initial long-range recognition between the inhibitor and the exosite of thrombin. They are buried underneath the inhibitor after the inhibitor binding, thus protecting these residues from chemical modification (Lys70 is rather far from the inhibitor and buried under Arg67 of thrombin; therefore, the exact correspondence for Lys70 is not as obvious). From the hirullin50'-62' structure (by Mali in this laboratory) [34], we believe that Ser50' (corresponding to position 54' of hirudin) maybe act as a protecting group to eliminate an N-terminal positive charge, while the hydrophobic Pro60' of hirudin is a polar serine in hirullin, which stabilizes the 3_{10} helical turn that follows (Table 3) [34]. Therefore, it appears that the hirudin 55'-59' segment is sufficient for exosite recognition and inhibitory function (that is, large enough to compete with fibrinogen binding). Interestingly, the 55'-59' segment was the only resolved portion observed in the crystal structure of the hirulog 1-thrombin complex that bound to the exosite, even though hirulog 1 has a hirudin sequence to Leu64' but Tyr63' is not sulfated [27].

The sequence of the second part of the MDL-28050 peptide, as well as that of hirullin50'-62', is not homologous to that of hirudin60'-65' (Table 3), and as a result, the crystal structure for this

part of the peptide varies (Figure 40, Figure 52). A closer examination of the structures proved to be very interesting and informative. First, the main chain conformations of MDL-28050 and hirullin are not very different (Figure 52), and the interactions with thrombin residues are also similar [34]. The conformation of the last five residues in MDL-28050 is also reasonably close to the 3_{10} helical turn structure of hirugen (Table 10), while the difference between the two is the result of a free bond rotation (about 40°) of the helical turn at the CA-C of Ile59' (Figure 52). As a result, the side group of Cha64' of MDL-28050 points in the opposite direction compared to that of Tyr63' in hirugen, although they both fit approximately into the same lipophilic pocket, with Cha64' penetrating deeper into thrombin (Figure 52, Figure42).

Recent ^1H NMR studies also revealed the behavior of a hirudin55'-65' peptide in aqueous solution. Ordered conformations were found to be highly populous for the hirudin peptides: residues Phe56', Ile59', Pro60', Tyr63' and Leu64' are somewhat ordered; Pro60'-Gln65' segment is identified as a 3_{10} helical turn, and there is a rotation at the $\alpha\text{CH}-\beta\text{CH}$ bond of Ile59' upon binding to thrombin [117]. Molecular dynamic simulation and Monte Carlo conformational searches suggest that the structure of the Ile59'-Gln65' portion is rather stable, while that of the Phe56'-Glu58' part fluctuates in solution [117]. These results agree well with our observations in crystal structures. Therefore, the process of binding these peptides to thrombin can be thought of as following: Pro60'-Gln65' is in a 3_{10} helical turn in solution prior to its binding to the exosite; Phe56' and Ile59 are also somewhat ordered because their hydrophobic

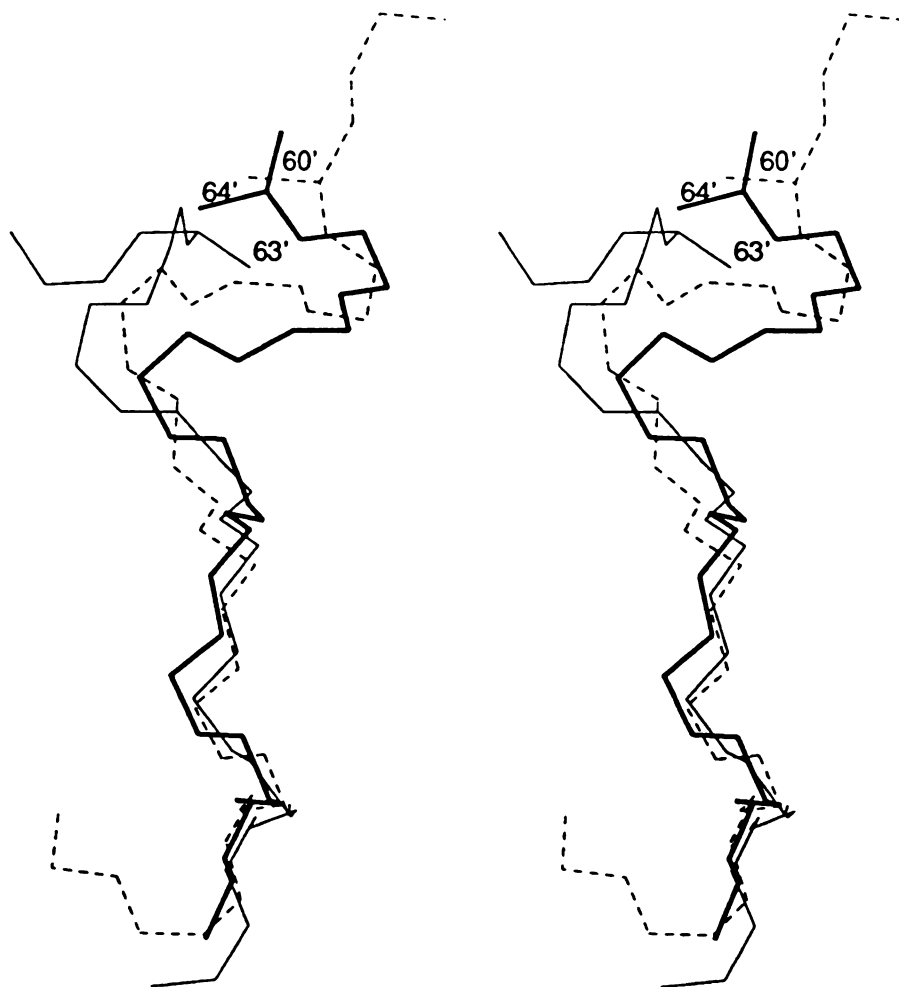


Figure 52. Stereoview comparing the N-CA-C Structures of MDL-28050, hirullin and hirugen as they are bounded with thrombin. Hirugen, light; MDL-280 50, dark; hirullin, broken. Major hydrophobic residues in the helical turn were labeled at their CB positions.

hydrophobic side groups can cluster with those in the helical turn. Since the exosite is highly positively charged [104,105], the negatively charged glutamic or aspartic acids of the peptides are used to direct the peptides toward the exosite of thrombin. When the peptide gets close enough to thrombin, hydrophobic contacts come into effect and predominate. Because the exosite is not the only positively charged site of thrombin [104,105], the special 55'-59' recognition sequence is used to confirm the correct binding. Finally, the helical turn rotates to adopt a conformation that is the most efficient for binding to thrombin, that is, to utilize the hydrophobic pocket formed by Leu65, Ile82, Met84 and Tyr76 of thrombin. Unlike the tyrosine at position 63' in hirudin, both MDL-28050 and hirullin have bulky hydrophobic side groups at position 64'. The differences in these sequential positions are probably the reason for their similar helical turn conformations, which differ from those of hirudin and hirugen by a free bond rotation near 59'-60'. It is clear that thrombin is not as specific or selective as previously thought, as it can tolerate different degrees of imprecision in the binding sequences [108]. These findings will probably provide useful guidelines for the molecular modeling of exosite binding sequences of other natural macromolecular thrombin substrates, which are listed (Table 24).

The only available α -thrombin structure with an "unoccupied" exosite is that of the PPACK-thrombin complex [22,32]. The crystal structure reveals, however, that the thrombin exosite makes contacts with the 186 insertion loop [116] of a symmetry related thrombin molecule (designated with *), including hydrophobic

contacts (Pro186* CG-Tyr76 CD1, 3.4 Å) and hydrogen bonds (Asp186A* OD2- Gln 38 OE1, 3.0 Å; Asp186A* OD1 is 3.4 Å from a water molecule that is 3.0 Å from Thr74 O and 3.1 Å from the guanadinium group of Arg67). From this, the side chain conformations of the exosite in PPACK-thrombin structure as an unoccupied exosite are probably still unreliable. The main chain conformation, on the other hand, should not have been affected significantly by the contacts of the 186 loop, and is found similar before and after the inhibitor binding (an r.m.s. value of about 0.5 Å between the exosite residues in the PPACK-thrombin and the MDL-28050-thrombin structures).

C. The S1'-S4' Subsites of Thrombin

Since the P1'-P4' residues of hirulog 3 are all glycine, their interaction with thrombin is diminished because of the lack of side groups. The antithrombotic potency of hirulog 3 or analogs thereof could be improved by rationally optimizing the interaction of the spacer residues with thrombin from interactive computer graphics molecular modeling. Side chains at S1' will be limited to small groups because of space limitations. This would suggest glycine or alanine as suitable P1' residues. However, with the location of the ϵ -amino group of Lys60F in the same vicinity, small polar groups, such as serine, might be better. This is also borne out by examining the P1' site of natural substrates of thrombin (Table 25). Threonine is a possibility, if it is small enough to fit into the cavity. This would be especially so if it were oriented such that its methyl group abuts the backside of the apolar S2 wall made up of His57, Tyr60A, Trp60D and its hydroxyl group hydrogen bonds or otherwise interacts with

Table 25. Peptide sequence around the cleavage site of thrombin macromolecular substrate from human plasma.

Protein*	Cleavage Site
	I
Fibrinogen A 12-24	G G G V R G P R V V E R H
A 15-27	V R G P R V V E R H Q S A
B 10-22	F F S A R G H R P L D K K
Prothrombin 151-163	M V T P R S E G S S V N L
282-294	F F N P R T F G S G E A N
Factor V 706-718	A L G I R S F R N S S L N
1014-1026	P L S P R T F H P L R S E
1541-1553	A W Y L R S N N G N R R N
Factor VIII 368-380	F I Q I R S V A K K H P K
736-748	A I E P R S F S Q D S R H
1309-1321	T Q S K R A L K Q F R L P
1685-1697	N Q S P R S F Q K K T R H
Factor XIII 33-42	G V V P R G V N L Q E F
509-518	E G V M K S R S N V D M
Platelet Receptor 37-50	T L D P R S F L L R N P N
Antithrombin III 389-399**	V I A G R S L N P N R
Factor XI 365-377**	K I K P R I V G G T A S R
Protein C 8-17**	Q V D P R L I D G K V

* See [76,78,120 -126] for references.

** Cofactor needed.

Lys60F NZ, as serine must do. Binding in the S2' subsite is modeled well by a small bulky hydrophobic residue making numerous van der Waals contacts with the methylene carbon atoms of Lys60F and the ring of Phe60H; the latter lines the interior wall along with Leu41. Valine, leucine and isoleucine are observed frequently at the P2' position of natural thrombin substrates (Table 25); phenylalanine is also found at the P2' position, most likely because it is small enough to be accommodated and can enter into an aromatic stacking interaction with Phe60H, or will behave like Phe*5' in the CV863-1 peptide that interacts with Leu40, Leu41, Glu192 and Gly193. Side chain conformations of Glu192 are completely different in the cyclotheonamide A and CV863-1 complexes, which agrees with the flexibility that has been observed other thrombin complex structures (unpublished results of this laboratory). It is clear that Glu192 plays a critical role in thrombin-substrate binding, although its contribution to the S' subsite binding is probably not very significant.

The nature of the S3'-S4' subsites is much less definitive than the S1'-S2' sites, primarily because of the uncertainty of the conformation of CA of P3', which may be directed away from the complex and therefore not capable of interacting at all, as with certain hirudin subsites [108], or possibly produce surface interaction with Glu192. The S3' site would seem to be modeled well by His51' of hirudin, which forms an ion pair with Glu39 [25]. Fibrinogen and Factor V have an arginine at the P3' position while other substrates either have a carboxyamide, which could hydrogen bond to Glu39, or a small side group (Table 25), which could just point out in the solvent (but require a twist of the main chain near

P3'). The P4' position appears to be also capable of interacting with S3' through a long, positively charged side chain or carboxyamide. Otherwise, the most likely interactions of P4' residues with thrombin are with Asn143 and/or Gln151.

Two other considerations are pertinent to S' subsite binding. The first is that the β -homoArg3' substitution introduces a displacement of the peptide chain of about one bond length. This might be of some consequence in the immediate vicinity of the scissile bond but its perturbing effect should dwindle and disappear as the distance from the point of catalysis increases. A more puzzling aspect of the exosite binding is the potentially strained conformation of Gly4'. Natural substrates have either glycine or serine at this position (Table 25). Although the $\phi = 100^\circ$ and, $\psi = -100^\circ$ are unacceptable conformational angles for serine and must be closer to allowed regions, they could be in a strained conformation to enhance catalysis. Conversely, the Gly4' conformation might be an artifact of β -homoArg3' and its additional carbon-carbon bond. This agrees with the tolerance of thrombin of a proline at the P1' position in hirulogl1 and hirulog 2, both of which are substrates for thrombin, although poor ones [57].

In inferring suitable residues for S1'-S4' subsite interactions, the consideration of the different rotamers that are possible from free bond rotation was kept to a minimum. Thus, other side chain groups or orientations may also be fairly suitable in addition to those indicated. It is of note that bulky Leu/Ile groups appear at the S1' subsite, but only with substrates requiring a cofactor (Table 25). The present work would seem to suggest that this would only be possible

with a conformational change, in this case, accompanying cofactor binding.

The thrombin-substrate interactions of the hirulog 3-thrombin complex provide revealing conclusions concerning details of the S' subsites at the molecular level immediate to the active site. The most evident factor emerging from all the foregoing analysis is that the corresponding P' amino acid residues of substrates and inhibitors must confer substantial additional specificity and stability to thrombin action. In this context, the correlation between the complementarity of the S' sites and the corresponding P' positions of natural substrates (Table 25) is particularly striking and impressive.

D. Other Inhibitors Utilizing S' Subsites

The structure of hirulog 3 revealed the S1'-S4' subsites of thrombin, even though substrate side chain interactions with thrombin have not been directly observed because of the use of the penta-glycine spacer [32]. The sites will be very useful for modeling various inhibitors or substrates C-terminal to the scissile bond. These sites have been shown to be important for binding to serine proteinases in general [39,58]. Not surprisingly, then, the D-Phe5'-TNT6' segment of cyclotheonamide A and Phe*5' of CV863-1 are both found in the S' subsite regions (Figure 48 and 53).

Although main chain atoms of TNT6' form hydrophobic interactions with Trp60D and Glu192 side groups (Table 12), the tyrosine ring of TNT6' appears to be too far away from thrombin to make any significant interactions (Figure 32). The ring points into the solvent region, close to where the main chain of Glu49' of hirudin is situated in the thrombin complex (Figure 48). We have

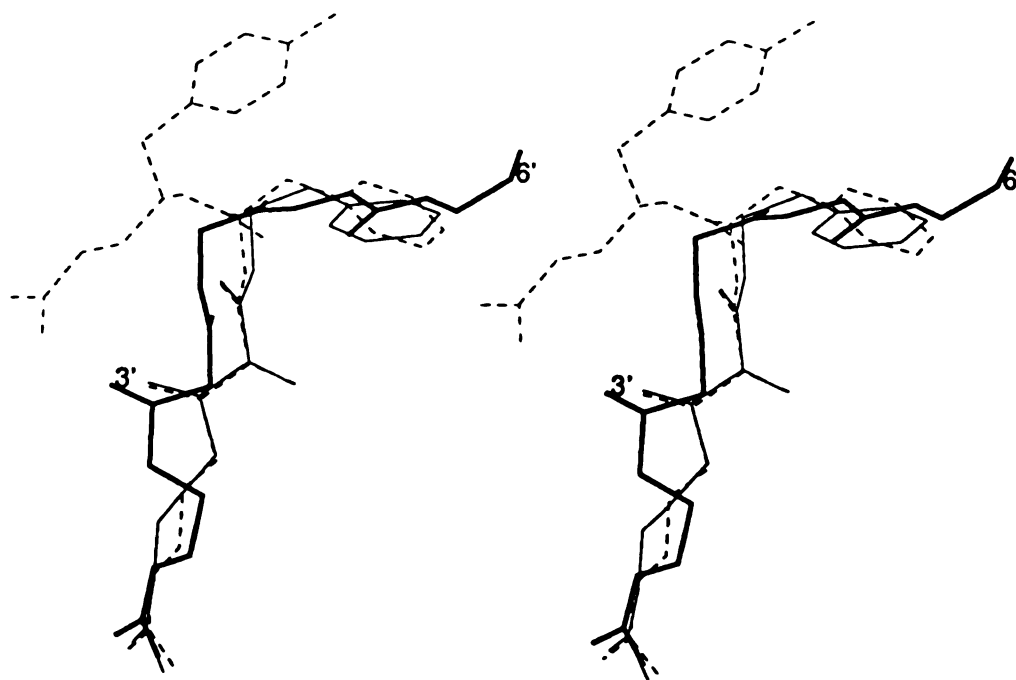


Figure 53. Stereoview comparing S' subsite binding sequences. Cyclotheonamide A, broken; hirulog 3, dark; CV863-1, light. Hirulog 3 positions labeled.

already discussed the small hydrophobic cavity at the back side of the thrombin S2 apolar site (His57, Tyr60 A and Trp60D), corresponding to the S1' subsite; the TNT6' ring is located somewhat differently in the cavity (Figure 48 and 53). Considering the fact that the position of TNT6' is relatively distant to thrombin, the TNT6' residue most likely does not contribute much to the S1'-S4' binding.

Unlike the inefficient positioning of TNT6', the binding site of D-Phe5' in cyclotheonamide A (Figure 32) or Phe*5' of CV863-1 (Figure 36) is in the middle of the S1'-S3' binding region (Figure 53). The position of the Phe*5' phenyl ring of CV863-1 is a model for the position of the site because it is not restrained as is D-Phe5' in cyclotheonamide A. This phenyl ring superimposes well upon the main chain of the penta-glycine bridge of hirulog 3 (Figure 53), while making many van der Waals contacts with residues Leu40, Leu41, Glu192 and Gly193 of thrombin (Table 15, Figure 36). The region formed by Leu40-Leu41 and Glu192-Gly193 appears to be a critical site in prothrombin to produce the folding of the catalytic machinery of the active enzyme. When prothrombin is activated in the "prothrombinase complex" (Figure 2), the zymogen is cleaved at the Arg15-Ile16 bond (Figure 3). Residue Ile16 is then folded inwards and buried within the protein, forming a salt bridge with Asp194 through its N-terminal positive charge. The salt bridge is right adjacent to the region of Leu40-Leu41 and Glu192-Gly193, which has also been suspected to promote destabilization of the zymogen [22].

There is a bond length displacement introduced by CO4' of the α -keto amide group, which causes a special chain reversal of the

inhibitors (Figure 50). Thus, the Leu40-41 and Glu192-Gly193 sites are not likely to be the S1' subsite for natural thrombin substrates, but rather, possibly the region of S2' and S3' subsites, as indicated by the hirulog 3 structure. For natural thrombin substrates, the S1' site is still likely on the "back side" of the S2 apolar site (His57, Tyr60A and Trp60D), or Lys60F if there is a polar (or charged) side group at P1'.

The structure of hirudin-thrombin complex [25] also provides evidence implicating the S1'-S4' subsites of thrombin. The superposition of the hirulog 3 and hirudin structures (Table 10, Figure 47 and 48) shows that the two are very similar from Asp55'-Leu64'. However, although the peptide from Lys47'-Gly54' (or Gly4'-Gly54') is in an extended conformation in both structures, the path of the chain is far from identical and is displaced with respect to thrombin by about 6.0 Å on the average. The main chain of the pentaglycine spacer of hirulog 3 makes much closer contacts with thrombin (Figure 48). The position of Lys47'-Pro48' of hirudin does not correspond to the D-Phe1'-Pro2'- β -homoArg3' positions of hirulog 3 because they do not bind in the active site like the latter. Instead, it is the N-terminal Ile1'-Thr2'-Tyr3' of hirudin that binds there, in a "non-substrate" orientation. Thus, the hirudin C-terminal in the vicinity of the S' subsites is positioned very differently from hirulog 3. Nonetheless, the side chains of Glu49'-His51' of hirudin interact impressively in an intricate and concerted way with thrombin. The tolerance of thrombin for imprecision in binding at the active site and exosite has already been pointed out by us [108]; the accommodation of Glu49'-His51' by the S2'-S4' subsites extends

this flexibility to yet an additional region of the fibrinogen recognition site. The only question that remains is whether natural substrates also display the close contact, main chain hydrogen bonding interactions of the penta-glycine spacer (Table 23). Judging from the general binding behavior of the exosite substrates [108], it would seem that the extent of utilization of main chain interactions would be variable (as with hirudin where there are none) and sequence dependent.

E. Model of Thrombin Autolysis

Since α -thrombin is degraded to β - and γ -thrombin through autolysis when the Arg77A-Asn78 (β -cleavage site) and Lys149E-Gly150 (γ -cleavage site) bonds are cleaved, the structure of α -thrombin can be regarded also as a structure of a macromolecular substrate of thrombin, so that it should be possible to use it to model a substrate-enzyme interaction. The Lys149E-Gly150 bond belongs to the loosely attached autolysis loop that has been shown to be flexible. The structure of the loop is ordered in PPACK-thrombin and is different from that in hirudin-thrombin only as a result of different crystal packing, but is completely disordered in crystal structures like that with hirugen and hirulogs (in C2 space group). Consequently, any modeling using the autolysis loop would be unreliable.

The overall folding near Arg77A in the exosite of thrombin is consistent in all the different thrombin structures where the exosite is in an inhibited state (Table 10). Although the structure of a thrombin complex with a macromolecular substrate is not yet available, there is precedence for the folding of the main chain near

the β -cleavage site being similar to that observed in the inhibited states. Because Arg77A is at the center of a sharp bend on the surface of thrombin, it is very easy to position Tyr76 and Arg77A in the S1 specificity pocket and the S2 apolar site without changing the main chain folding of either (Figure 54). Moreover, this model shows that the highly positively charged exosite of the substrate can make numerous interactions with the highly negatively charged active site region of the enzyme, which was revealed by electrostatic potential calculations [104,105]. The detailed interactions will not be accurate, but may still be useful in providing general information about the global thrombin-substrate contact.

In this model, the exosite of the substrate can make many contacts with the active site region of the enzyme (including the apolar site and the 60A-I insertion loop). The rim of the putative heparin binding site (Arg173) [104,105] also seems to be reachable. Residues Lys109'-Ile118' of thrombin substrate are in the vicinity of the autolysis loop of the enzyme and could also produce some contacts. More detailed predictions of the model are illustrated in Figure 54 (side chain conformations of the substrate are not changed). Possible contacts in the direction of the S' sites include: (1) Asn78' of the substrate in between the side groups Asn143 and Glu192 so that the S1' interactions would be polar or hydrogen-bonding interactions; (2) Ile79' in between the side group of Tyr117' and the "back-side" of Trp60D; (3) Lys81' with Glu146 or Glu192 in the opposite direction; (4) Asp116' between Gln151 and Arg73 that could hydrogen bond or for ion pair interactions; (5) Tyr117' close to Asn143 and Glu39 that could form hydrogen bonds or van der Waals

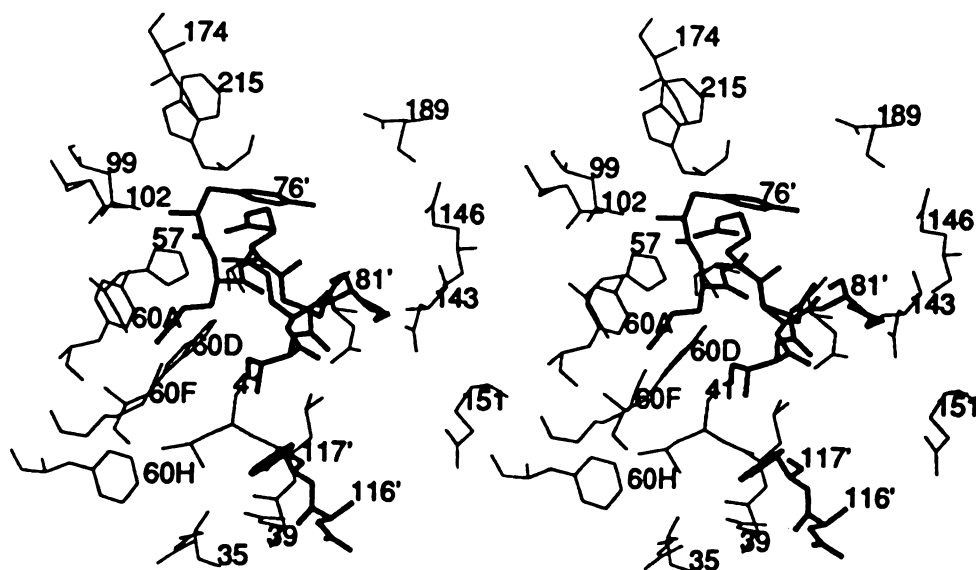


Figure 54. Stereoview of a Model for β -Cleavage Autolysis. Thrombin as substrate, dark with prime designations; thrombin as enzyme, light.

contacts. Therefore, the S1'-S4' subsites of the model are likely to be Asn143, Trp60D, Glu192 Gln151 and Arg73, which agree well with that deducted from the hirulog 3-thrombin.

D. Conclusion

Crystal structures of thrombin in the presence of different types of inhibitors revealed the details of several binding sites of thrombin, which are endowed with considerable flexibility [108]. The hirudin C-terminal binding at the proposed fibrinogen recognition site [25] provides a credible model for fibrinogen binding at the exosite, in the absence of a structural determination of a fibrinogen-exosite complex. Fibrinogen binding in the active site has been described with the structure of a fibrinopeptide A-thrombin complex [29,30]. The PPACK-thrombin structure and the fibrinopeptide A-thrombin structures show the positioning of an arginine residue in the primary specificity site and the nature of the apolar subsite, and are excellent models for substrate binding at the catalytic site [22,29,30]. The hirulog 3 structure does much to define the S' binding sites of substrates on the other side of the catalytic site.

In addition to the enzymatic active site, fibrinogen recognition exosite and the S' subsites that have been defined with these structures, a putative heparin binding site (Arg126, Lys236, Lys240, Arg93, Arg101, Arg233, Arg165, Lys169, Lys235, Arg175, Arg173 and Arg97) has also been inferred for thrombin based on electrostatic potential calculations [104,105]. This region was confirmed to be a binding site in the crystal structure of the prothrombin fragment 2-thrombin complex [127]. The site is a 10-12 Å wide and 20-25 Å long groove harboring the positively charged

arginines or lysines. The groove is large enough to accommodate a sulfated tetra- or penta-saccharide. The heparin site is also an exosite of thrombin, and like the exosite for fibrinogen recognition, it operates in concert with the active site of thrombin to attain certain specific responses.

One of the reasons for the versatility of thrombin (Table 2, Figure 7) is the existence of the independent binding sites. Unlike trypsin, chymotrypsin or other enzymes, thrombin is more selective in its enzymatic action, which is the result of the multiple binding features. While the active site pocket is similar to other trypsin-like serine proteinases, the exosites of thrombin are unique among the known enzymes. The fibrinogen recognition exosite is only some 10 Å away from the catalytic site and apparently extends for a length of about 40 Å; the heparin binding site is about 15 Å away from the active site but is on the opposite side of the fibrinogen recognition exosite of the molecule. The S' subsites comprise the region that connects the active site and the fibrinogen recognition exosite and is sometimes considered as part of the fibrinogen binding exosite. Indications are, however, that only a penta-peptide sequence is required for fibrinogen recognition (unpublished results of this laboratory).

Another reason for the versatility of thrombin is the possibility for multiple binding modes within these sites, which includes different combinations of the available subsites or simply promiscuous binding among these subsites. The fidelity of active site binding is only very precise in the S1 specificity pocket. The S2-S11 subsites can be used in many different ways and are forgiving of

imprecision. For example, the apolar site, where FPA uses it as S2 and S8-S9 subsites while PPACK uses the site as its S2 and S3 subsites. The multiple binding modes also reveal themselves the complexes of hirudin, MPQA, NAPAP, 4-TAPAP, DAPA, FPAM, cyclotheonamide A and CV863-1 with thrombin. The binding in the fibrinogen recognition exosite is also forgiving of different sequences and modes at binding even in the conserved recognition segment (55'-59'). Examples are the substitutions of Glu58' to Pro58' and Phe56' to Tyr56' in the MDL-28050 structure and Ile59' to Phe59' in hirullin. Moreover a even greater flexibility of binding is associated with the 3_{10} helical turn region, displayed in the structures of hirudin, MDL-28050 and hirullin50'-62'. Binding in the S' subsites also seems to be sequence-dependent and could also show different binding modes.

Because of the multiple choices among the major binding sites and among the different binding modes of the subsites inside a major site, the number of the total combination of thrombin binding possibilities are much greater than the known number of functions for thrombin. It can be concluded, therefore, that the functional diversity (specificity and versatility) of thrombin is inherently related to the availability of the few major binding sites and the explicit binding modes that result from a tolerance of imprecision by specific interactions among subsites.

LIST OF REFERENCES

1. Davie , E. W., Fujikawa, K. and Kisiel, W. (1991) *Biochemistry* 30, 10363-10370.
2. Rosenberg, R. D. (1987) in *The Molecular Basis of Blood Diseases* (Stamatoyannopoulos, G., Nienhuis, A.W., Leder, P. and Majerus, P. W., eds.) pp. 534-574, W. B. Saunders Company, Philadelphia.
3. Mann, K. G. (1976) *Methods Enzymol.* 45, 123-156.
4. Butkowski, R. J., Elion, J., Downing, M. R. and Mann, K. G. (1977) *J. Biol. Chem.* 252, 4942-4957.
5. Thompson, A. R., Enfield, D. L., Ericsson, L. H., Legaz, M. E. and Fenton, J. W. II (1977) *Arch. Biochem. Biophys.* 178, 356-367.
6. Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S. R., and Hofsteenge, J. (1989) *EMBO J.* 8, 3467-3475.
7. Jackson, C. M. and Nemerson, Y. (1980) *Annu. Rev. Biochem.* 49, 756-811.
8. Nilson, B., Horne, M. K. III and Gralnick, H. R. (1983) *Arch. Biochem. Biophys.* 244, 127-133.
9. Crabtree, G. R. (1987) in *The Molecular Basis of Blood Diseases* (Stamatoyannopoulos, G., Nienhuis, A. W., Leder, P. and Majerus, P. W., eds.) pp. 631-661, W. B. Saunders Company, Philadelphia.
10. Higgins, D. L., Lewis, S. D. and Schafer, J. A. (1983) *J. Biol. Chem.* 258, 9276-9281.
11. Hantgan, R. and Hermans, J. (1979) *J. Biol. Chem.* 254, 11272-11279.
12. Laki, K. (1972) *Ann. N.Y. Acad. Sci.* 202, 1-21.
13. Travis, J. and Salvesen, G. S. (1983) *Annu. Rev. Biochem.* 52, 655-709.
14. Esmon, C. T., Taylor, F. B. and Snow, R. (1991) *Thromb. Haemostas.* 66, 160-165.
15. Fenton, J. W. II (1986) *Ann. N. Y. Acad. Sci.* 485, 5-15.

16. Shuman, M. A. (1986) *Ann. N. Y. Acad. Sci.* 556, 104-115.
17. Weksler, B. B., Ley, C. W. and Jaffe, E. A. (1978) *J. Clin. Invest.* 62, 923-930.
18. Prescott, S. M., Zimmerman, G. A. and McIntyre, T. M. (1984) *Proc. Natl. Acad. Sci. USA* 81, 3534-3538.
19. DeGroot, P. G., Gonsalves, M. D., Loesberg, C., von Buul-Wortelboer, M. F., van Aken, W. G. and van Mourik, J. A. (1984) *J. Biol. Chem.* 259, 13329-13333.
20. Gelehrter, T. D., and Sznycer-Laszyk, R. (1986) *J. Clin. Invest.* 77, 165-169.
21. Bode, W. and Stubbs, M. T. (1992) *Semin. Thromb. Hemostas.*, in press.
22. Bode, W., Turk, D. and Karshikov, A. (1992) *Protein Sci.* 1, 426-471.
23. Skrzypczak-Jankun, E., Rydel, T. J., Tulinsky, A., Fenton, J. W. and Mann, K. G. (1989) *J. Mol. Biol.* 206, 755-757.
24. Rydel, T. J., Ravichandran, K. G., Tulinsky, A., Bode, W., Huber, R., Roitsch, C., and Fenton, J. W. II (1990) *Science* 249, 277-280.
25. Rydel, T. J., Tulinsky, A., Bode, W. and Huber, R. (1991) *J. Mol. Biol.* 221, 583-601.
26. Grutter, M. G., Priestle, J. P., Rahuel, J., Grossenbacher, H., Bode, W., Hofsteenge, J. and Stone, S. R. (1990) *EMBO J.* 9, 2361-2365.
27. Skrzypczak-Jankun, E., Carperos, V., Ravichandran, K. G., Tulinsky, A., Westbrook, M. and Maraganore, J. M. (1991) *J. Mol. Biol.* 221, 1379-1393.
28. Banner, D. W. and Hadvary P. (1991) *J. Biol. Chem.* 266, 20085-20093.
29. Martin, P. D., Robertson, W., Turk, D., Huber, R., Bode, W. and Edwards, B. F. P. (1992) *J. Biol. Chem.* 267, 7911-7920.
30. Stubbs, M. T., Oschkinat, H., Mayr, I., Huber, R., Anglikar, H., Stone, S. R. and Bode, W. (1992) *Eur. J. Biochem.* 1542, 187-195.

31. Brandstetter, H., Turk, D., Hoeffken, H. W., Grosse, D., Sturzebecher, J., Martin, P. D., Edwards, B. F. P. and Bode, W. (1992) *J. Mol. Biol.* 226, 1085-1099.
32. Qiu, X., Padmanabhan, K. P., Carperos, V. E., Tulinsky A., Kline, T., Maraganore, J. M. and Fenton, J. W. II (1992) *Biochemistry*, 31, 11689-11697.
33. Maryanoff, B., Qiu, X., Padmanabhan, K. P., Tulinsky, A., Almond, H. R., Andrade-Gordon, P., Greco, M. N., Kauffman, J. A., Nicolaou, K. C., Liu, A., Brungs, P. H. and Fusetani, N. (1992) *Proc. Natl. Acad. Sci. U. S. A.*, submitted.
34. Qiu, X., Yin, M., Padmanabhan, K., Krstenansky, J. L. and Tulinsky A. (1992), *J. Biol. Chem.*, submitted.
35. Boissel, J. P., LeBonniec, B., Rabiet, M. J., Lable, D. and Elion, J. (1984) *J. Biol. Chem.* 259, 5691-5697.
36. Kawabata, S., Morita, R., Iwanaga, S. and Igarashi, H. (1985) *J. Biochem. (Tokyo)* 97, 325-331.
37. Brezniak, D. V., Brower, M. S., Witting, J. I., Walz, D. A. and Fenton, J. W. II (1990) *Biochemistry* 29, 3536-3540.
38. Huber, R. and Bode, W. (1978) *Acc. Chem. Res.* 11, 114-122.
39. Read, R. J. and James, M. N. G. (1986) in *Protein Inhibitors* (Barret, A. J. and Salveen, G. eds.), pp. 301-336, Elsevier, Amsterdam.
40. Markwardt, F. (1970) *Method Enzymol.* 19, 924-932.
41. Stone, S. R. and Hofsteenge, J. (1986) *Biochemistry* 25, 4622-2628.
42. Bode, W. and Huber, R. (1992) *Eur. J. Biochem.* 1275, 433-451.
43. Sawyer, R. T. (1991) *Bio/technology* 9, 513-519.
44. Maraganore, J. M., Chao, B., Joseph, M. L. and Ramachandran, K. L. (1989) *J. Biol. Chem.* 264, 8692-8698.
45. Maraganore, J. M. (1990) in *"Protein C and Related Anticoagulants"* (Bruley, D. F. and Drohan, W. N. eds.) pp. 103-118, Gulf Publishing Co., Houston.
46. Maraganore, J. M., Bourdon, P., Jablonski, J., Ramachandran, K. L. and Fenton, J. W. II (1990) *Biochemistry* 29, 7095-7101.

47. Kline, T., Hammond, L., Bourdon, P. and Maraganore, J. M. (1991) *Biochem. Biophys. Res. Commun.* 177, 1049-1055.
48. DiMaio, J., Gibbs, B., Munn, D., Lefebvre, J., Ni, F. and Konishi, Y. (1990) *J. Biol. Chem.* 265, 21689-21703.
49. Krstenansky, J. L. and Mao, S. J. T. (1987) *FEBS Lett.* 221, 10-16.
50. Bourdon, P., Fenton, J. W. II and Maraganore, J. M. (1990) *Biochemistry* 29, 6379-6384.
51. Chang, J.-Y., Ngai, P. K., Rink, H., Dennis, S. and Schlaeppli, J.-M. (1990) *FEBS Lett.* 261, 287-290.
52. Jakubowski, J. A. and Maraganore, J. M. (1990) *Blood* 75, 399-406.
53. Prescott, S. M., Seeger, A. R., Zimmerman, G. A., McIntyre, T. M. and Maraganore, J. M. (1990) *J. Biol. Chem.* 265, 9613-9616.
54. Naski, M. C., Fenton, J. W. II, Maraganore, J. M., Olsen, S. T. and Shafer, J. A. (1990) *J. Biol. Chem.* 265, 13484-13489.
55. Hofsteenge, J., Stone, S. R., Donella Deana, A. and Pinna, L. A. (1990) *Eur. J. Biochem.* 188, 55-59.
56. Niehrs, C., Hutter, W. B., Carvallo, D. and Degryse, E. (1990) *J. Biol. Chem.* 265, 13484-13489.
57. Witting, J. I., Bourdon, P., Brezniak, D. V., Maraganore, J. M. and Fenton, J. W. II (1992) *Biochem. J.* 283, 737-743.
58. Laskowski, M. and Kato, I. (1980) *Ann. Rev. Biochem.* 49, 593-626.
59. Silverman, R. B. (1988) in *Mechanism-Based Enzyme Inactivation: Chemistry and Enzymology*, pp. 5-6, CRC Press, Boca Raton.
60. Rich, D. N., (1985) *J. Med. Chem.* 28, 263-271.
61. Huff, J. R., (1991) *J. Med. Chem.* 34, 2305-2311.
62. Michnick, S. W., Rosen, M. K., Wandless, T. J., Karplus, M. and Schreiber, S. L. (1991) *Science* 252, 836-839.
63. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. and Clardy, J. (1991) *Science* 252, 839-842.

64. Moore, J. M., Peattie, D. A., Fitzgibbon, M. J. and Thompson, J. A. (1991) *Nature* 351, 248-250.
65. Rosen, M. K., Standaert, R. F., Galat, A., Nakatsuka, M. and Schreiber, S. L. (1990) *Science* 248, 863-866.
66. Albers, M. W., Walsh, C. T. and Schreiber, S. L. (1990) *J. Org. Chem.* 55, 4984-4987.
67. Van Duyne, G. D., Standaert, R. F., Schreiber, S. L. and Clardy, J. (1991) *J. Am. Chem. Soc.* 113, 7433-7438.
68. Harrison, R. K. and Stein, R. L. (1992) *J. Am. Chem. Soc.* 114, 3464-3469.
69. Fusetani, N., Matsunaga, S., Matsumoto, H. and Takebayashi, Y. (1990) *J. Am. Chem. Soc.* 112, 7053-7054.
70. Fusetani, N., Sugawara, T. and Matsumoto, H. (1991) *J. Am. Chem. Soc.* 113, 7811-7814.
71. Hagihara, M. and Schreiber, S. L. (1992) *J. Am. Chem. Soc.* 114, 6570-6573.
72. Trainor, D. A. (1987) *Trends Pharmacol. Sci.* 8, 303-327.
73. Tulinsky, A. and Belvins, R. A. (1987) *J. Biol. Chem.* 262, 7737-7743.
74. Vijayalakshimi, J., Meyer, E. F., Kam, C. and Powers, J. C. (1991) *Biochemistry* 30, 2175-2183.
75. Chow, M. M., Meyer, E. F., Bode, W., Kam, C.-M., Radhakrishnan, R., Vijayalakshimi, J. and Powers, J. C. (1990) *J. Am. Chem. Soc.* 112, 7783-7789.
76. Vu, T. H., Wheaton, V. I., Hung, D. T., Charo, I. and Coughlin, S. R. (1991) *Nature* 355, 674-677.
77. Scharf, M., Engels, J. and Tripler, D. (1989) *FEBS Lett.* 255, 105-110.
78. Blomback, B., Blomback, M., Hessel, B. and Iwanaga, S. (1967) *Nature* 215, 1445-1448.
79. Krstenansky, J. L., Broersma, R. J., Owen, T. J., Payne, M. H., Yates, M. T. and Mao, S. J. T. (1990) *Thrombosis and Haemostasis* 63, 208-214.

80. Krstenansky, J. L., Owen, T. J., Yates, M. T. and Mao, S. J. T. (1990) *FEBS Lett.* 269, 425-429.
81. Geiger, T. and Clarke, S. (1987) *J. Biol. Chem.* 262, 785-794.
82. Wright, H. T. (1991) *Protein Engineering* 4, 283-294.
83. Bischoff, R., Lepage, P., Jaquinod, M., Cauet, G., Acker-Klein, M., Clesse, D., Laporte, M., Bayol, A., Van Dorsseelaer, A. and Roitsch, C. (1992) *J. Biol. Chem.*, submitted.
84. Lundblad, R. L., Kingdom, H. and Mann, K. G. (1976) *Methods Enzymol.* 45, 1626-1634.
85. Fenton II, J. W., Fasco, M. J., Stackrow, A. B., Aronson, D. L., Young, A. M. and Finlayson, J. S. (1977) *J. Biol. Chem.* 252, 3587-3598.
86. Kettner, C. and Shaw, E. (1981) *Methods Enzymol.* 80, 826-842.
87. Krstenansky, J. L. and Mao, S. J. T. (1987) *FEBBS Lett* 211, 10-16.
88. Blundell, T. L. and Johnson L. N. (1976) in *Protein Crystallography*, Academic Press, New York.
89. Carter Jr., C. W. and Carter, C. W. (1979) *J. Biol. Chem.* 254, 12219-12223.
90. Abergel, C., Moulard, M., Moreau, H., Loret, E., Cambillau, C. and Fontecilla-Camps, J. C. (1991) *J. Biol. Chem.* 266, 20131-20138.
91. Jancarik, J. and Kim, S.-H. (1991) *J. Appl. Cryst.* 24, 409-411.
92. Howard, A. J., Gilliland, G. L., Finzel, B. C., Poulous, T. L., Olhendorf, D. H. and Salemme, F. R. (1987) *J. Applied Cryst.* 20, 383-387.
93. Bragg, W. L. (1913) *Proc. Camb. Phil. Soc.* 17, 43-61.
94. Hendrickson, W. A. and Konnert, J. H. (1980) in "*Biomolecular Structure, Function, Conformation and Evolution*" (R. Srinivasan, Ed.) pp.43-57, Pergamon, Oxford.
95. Jones, T. A. (1982) in "*Computational Crystallography*" (D. Sayre, Ed.) pp.303-317, Clarendon Press, Oxford.

96. Luzzati, V. (1952) *Acta Cryst.* 5, 802-810.
97. Rydel, T. J. (1990) *PhD Thesis*, Department of Chemistry, Michigan State University.
98. Wieland, T. and Bodanszky, M. (1991) "*The World of Peptides*", Springer-Verlag, Berlin.
99. Ovchinnikov, Y. A. (1973) in "*The Chemistry of Polypeptides*" (P. G. Katsoyannis, Eds.) pp.169-204, Plenum Press, New York-London.
100. Karle, I. L. (1986) *Int. J. Peptide Protein Res.* 28, 420-427.
101. Karle, I. L., Kishore, R. Raghothama, S. and Balaram, P. (1988) *J. Am. Chem. Soc.* 110, 1968-1963.
102. Le Bonniec, B. F. and Esmon, C. T. (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88, 7371-7475.
103. Mathews, I.I., unpublished results of this laboratory.
104. Bode, W., Huber, R., Rydel, T. J. and Tulinsky A. (1991) in "*Thrombin: Structure and Function*" (Berliner, L. J. Ed.), pp. 3-55, Plenum Press, New York.
105. Karshikov, A., Bode, W., Tulinsky, A. and Stone, S. R. (1992) *Protein Sci.* 1, 738-755.
106. Braun, P. J., Dennis, S., Hofsteenge, J. and Stone, S. R. (1988) *Biochemistry* 27, 6517-6522.
107. Dodt, J., Kohler, J. and Balci, A. (1988) *FEBS. Letters* 229, 87-90.
108. Tulinsky, A., and Qiu, X. (1992) *Blood Coagulation and Fibrinolysis*, in press.
109. Marsh, H. C., Meinwald, Y. C., Thannhauser, T. W. and Scheraga, H. A. (1983) *Biochemistry* 22, 4170-4174.
110. Wu, T.-P., Vivien, C.-Y., Tulinsky, A., Chruscial, R. A., Nakanishi, H., Shen, R., Priebe, C. and Kahn, M.(1992) *Prot. Eng.* submitted.
111. Berliner, L. J. and Shen, Y. Y. L. (1977) *Biochemistry* 16, 4622-4626.

112. Ni, F., Meinwald, Y. C., Vasquez, M. and Scheraga, H. A. (1989) *Biochemistry* 28, 3094-3105.
113. Betz, A., Hofsteenge, J. and Stone, S. R. (1991) *Biochemistry* 30, 9848-9853.
114. Chang, J.-Y. (1991) *Biochemistry* 30, 6656-6661.
115. Fenton, J. W. II (1987) in "Control of Animal Cell Proliferation" (Boynton, A. L. and Lefferf, H. L., Eds.) pp.133-152, Academic Press, New York.
116. Fenton, J. W. II (1988) *Advan. Clin. Enzymol.* 6, 186-193.
117. Ni, F., Ripoll, D.R. and Purisima, E.O. (1992) *Biochemistry* 31, 2545-2554.
118. Hayashi, T., Zushi, M., Yamamoto, S. and Suzuki, K. (1990) *J. Biol. Chem.* 265, 20156-20159.
119. Hortin, G. L., Tollefsen, D. M. and Benutto, B. M. (1989) *J. Biol. Chem.* 264, 13979-13982.
120. Downing, M. R., Butkowski, R. J., Clark, M. M. and Mann, K. G. (1975) *J. Biol. Chem.* 250, 8897-8906.
121. Kane, W. H. and Davie, E. W. (1988) *Blood* 71, 539-555.
122. Tahakashi, N., Tahakashi, Y. and Putnum, F. W. (1986) *Proc. Natl. Acad. Sci. U.S.A.* 83, 8019-8023.
123. Pratt, C. W. and Church, F. C. (1991) *Semin. Hematol.* 28, 3-9.
124. Kisiel, W. (1979) *J. Clin. Invest.* 64, 761-769.
125. Fujikawa, K., Chung, D.W., Hendrickson, L. and Davie, E. W. (1986) *Biochemistry* 25, 2417-2424.
126. Hagen, F. S., Gary, C. L., O'Hara, P., Grant, F. J., Saari, G. C., Woodbury, R. G., Hart, C. E., Insley, M., Kisiel, W., Kurachi, K. and Davie, E. W. (1986) *Proc. Natl. Acad. Sci. U.S.A.* 83, 2412-2416.
127. Arni, R. K., Padmanabhan, K., Padmanabhan, K. P., Wu, T.-P. and Tulinsky A. (1992) *J. Biol. Chem.*, submitted.

MICHIGAN STATE UNIV. LIBRARIES



31293009141064