Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of thrombin complexed with selective non-electrophilic inhibitors having cyclohexyl moieties at P1

Raman Krishnan,† Igor Mochalkin,‡ Raghuvir Arni§ and A. Tulinsky*

Department of Chemistry, Michigan State University, East Lansing, Michigan 48824-1322, USA

† Present address: BioCryst Pharmaceuticals Inc., Birmingham, AL 35244, USA.
‡ Present address: Center for Molecular Genetics, University of California at San Diego, La Jolla, CA 92093-0634, USA.
§ Present address: Department of Physics, IBILCE/UNESP, CP136, São Jose do Rio Preto, SP, CEP 15054-000, Brazil.

Correspondence e-mail: tulinsky@cem.msu.edu

Received 14 September 1999
Accepted 5 January 2000

PDB References: thrombin–MOL354 complex, 1c4v; thrombin–MOL356 complex, 1c4y; thrombin–MOL376 complex, 1d9i; thrombin–MOL592 complex, 1d6w; thrombin–MOL1245 complex, 1c4u.

The crystal structures of five new non-electrophilic β-strand-templated thrombin active-site inhibitors have been determined bound to the enzyme. Four co-crystallize with hirugen and inhibitor isomorphously to produce thrombin–hirugen crystals (monoclinic, space group $C2$), while one co-crystallizes in the hexagonal system, space group $P6_5$. A 1,4-substituted cyclohexyl moiety is conserved at the P1 position of all the inhibitors, along with a fused hetero-bicyclic five- and six-membered ring that occupies the P2 site. Amino, amidino and aminoimidazole groups are attached to the cyclohexyl ring for recognition at the S1 specificity site, while benzylsulfonyl and diphenyl groups enhance the binding at the S3 subsite. The cyclohexyl groups at the P1 positions of three of the inhibitors appear to be in the energetically favored chair conformation, while the imidazole-substituted cyclohexyl rings are in a boat conformation. Somewhat unexpectedly, the two cyclohexyl-aminoimidazole groups bind differently in the specificity site; the unique binding of one is heretofore unreported. The other inhibitors generally mimic arginyl binding at S1. This group of inhibitors combines the non-electrophilicity and selectivity of DAPA-like compounds and the more optimal binding features of the S1–S3 sites of thrombin for peptidic molecules, which results in highly potent (binding constants 12 nM–16 pM, one being 1.1 μM) and selective (ranging from 140 to 20 000 times more selective compared with trypsin) inhibitors of thrombin. The binding modes of these novel inhibitors are correlated with their binding constants, as is their selectivity, in order to provide further insight for the design of therapeutic antithrombotic agents that inhibit thrombin directly at the active site.

1. Introduction

The blood-coagulation cascade is a series of activation reactions that utilizes precursor protein molecules and regulating factors circulating in the blood to seal an area of endothelial damage in order to stop blood loss (Davie *et al.*, 1991). Initially triggered by tissue factor, a membrane glycoprotein released at the site of vascular injury that complexes factor VIIa, the coagulation cascade ultimately results in the formation of a cross-linked fibrin matrix that stabilizes a platelet plug. The multifunctional serine protease thrombin plays a prominent role in the process by participating in both pro- and anticoagulant activities. In the former, it cleaves fibrinogen, which leads to fibrin, and activates platelets and coagulation factors V, VII, VIII and XIII (Colman *et al.*, 1994). Bound to thrombomodulin, thrombin promotes anticoagulation by activating protein C, which converts factors Va and VIIIa to inactive forms in the presence of the cofactor protein S (Esmon, 1993).

Thrombin possesses a catalytic triad of His57–Asp102–Ser195 (chymotrypsinogen numbering; Bode *et al.*, 1992), like other members of the serine protease family, and electropositive fibrinogen- and heparin-binding exosites. The catalytic triad of thrombin is surrounded by Ser214–Cys220 and a lengthy insertion loop at position 60, forming the active site of the enzyme. The Ser214–Cys220 structural segment consists of two almost orthogonal arrays. The equatorial one (Ser214–Gly216) is critical for the association of the P1–P3 segment of the main chain of thrombin substrates and active-site peptidic inhibitors through an antiparallel (Bode *et al.*, 1992; Mathews *et al.*, 1994) or parallel (Tabernero *et al.*, 1995; Mochalkin & Tulinsky, 1999) β-like strand. The axial array, consisting of Glu217–Cys220, forms part of the arginyl specificity pocket that contains Asp189 at the bottom, both key recognition residues of the binding event. The 60A–D insertion loop partially occludes the active site at the top and is responsible for some of the selectivity of the active site compared with trypsin and other serine proteases (Locht *et al.*, 1997).

Since the determination of the crystal structure of thrombin inhibited with PPACK[1] (Bode *et al.*, 1992), a large array of different thrombin inhibitors have been developed and studied (Balasubramanian, 1995). Selective inhibition of thrombin has been the principal target of these antithrombotic drug-design ventures, which are aimed at essentially replacing the clinical use of heparin, which is neither highly optimal, convenient nor controllable. There are numerous serine proteases, unrelated to coagulation, that play other important physiological roles in life processes. Thus, inhibitor design should be selective since such enzymes could be accidental targets of less discriminating thrombin inhibitors and could lead to unwanted side effects.

Small-molecule peptidic thrombin inhibitors binding at the active site of thrombin can be broadly classified as electrophilic or not by virtue of their ability to form a covalent tetrahedral intermediate through an activated carbonyl group of the P1 residue and Ser195 OG of the catalytic triad. Since all serine proteases have a structurally similar catalytic center, the electrophilic inhibitors interact with many different serine proteases without much selectivity. Such was the case of a series of boronate analogues of PPACK, which had picomolar binding constants for both thrombin and trypsin (Elgendy *et al.*, 1992; Tapparelli *et al.*, 1993; Claeson *et al.*, 1993). A class of non-electrophilic inhibitors designed using a DAPA template is very selective for thrombin, but individual members do not usually bind optimally like substrate to the S1–S3 sites (Banner & Hadvary, 1991; Brandstetter *et al.*, 1992; Mathews & Tulinsky, 1995). Molecumetics Ltd (Bellevue, WA) have used a combinatorial approach to generate a diverse variety of conformationally constrained mimetics of the principal structural elements of proteins (α-helicies, β-strands, reverse turns; Kahn *et al.*, 1988; Nakanishi *et al.*, 1992; Wu *et al.*, 1993; Ogbu *et al.*, 1998). The syntheses (Boatman *et al.*, 1999) and the crystal structures (St Charles *et al.*, 1999) of four Molecumetics first generation (P3–P1′) β-strand-templated thrombin active-site inhibitors bound to thrombin have recently been reported. The template was the bicycle at the P2 position (MOL106; Fig. 1), with substituents chosen to match those of PPACK at other positions to produce a number of potent and selective inhibitors of thrombin. The present group of second-generation Molecumetics inhibitors (Fig. 1) combines the non-electrophilicity and selectivity of DAPA-like compounds and the more optimal binding features at the S1–S3 sites of peptidic molecules, which results in highly potent and selective inhibitors of thrombin. The new P2 templates mimic the extended β-strand secondary structure of peptide substrates bound to their cognate enzymes within the protease superfamily (Ogbu *et al.*, 1998). These templates were designed with the goal of developing a versatile scaffold on which a wide variety of pharmacophore units could be rapidly introduced. The initial work produced a series of thrombin inhibitors of the general structure shown in Fig. 1 (except MOL106; Ogbu *et al.*, 1998).

Here, we report five crystal structures of new non-electrophilic β-strand-templated thrombin active-site inhibitors as they bind to thrombin. A 1,4-substituted cyclohexyl moiety is conserved at the P1 position of all the inhibitors, as is a fused hetero-bicyclic five- and six-membered ring occupying the P2 site. Amino, amidino and amino-imidazole groups are attached to the cyclohexyl ring for recognition, whereas benzylsulfonyl and diphenyl groups enhance the binding at P3. Somewhat unexpectedly, the two cyclohexyl-imidazole groups bind differently in the specificity site (the unique binding of one unreported previously), while the bromine of the bromobenzylsulfonyl of MOL1245 makes 11 impressive van der Waals contacts with thrombin. The binding modes of the inhibitors are correlated with binding constants and provide further insight for the design of therapeutic antithrombotic agents that inhibit thrombin directly.

2. Experimental

2.1. Crystallization

An approximately sevenfold molar excess of hirugen was added at 277 K to a frozen human α-thrombin solution (1.0 mg ml^{-1} in 0.75 *M* NaCl) to form a 1:1 thrombin–hirugen complex and prevent autocleavage of the enzyme. After approximately 24 h, the solution of the binary complex was concentrated in a refrigerated centrifuge (277 K) to about 3.5 mg ml^{-1} using a Centricon 10 concentrator with a molecular-weight cutoff of 30 kDa. Since the MOL356 inhibitor was not very soluble, probably because of the hydrophobic biphenyl–cyclohexyl–aminoimidazole combination (Fig. 1), it was added to the concentrated thrombin–hirugen complex at about a maximum of only one molar excess, while the other inhibitors were at approximately 50 molar excess; both were then left to equilibrate overnight.

The complexes were crystallized using the hanging-drop crystallization technique. The well solution contained 24% PEG 8000 in 0.1 *M* sodium phosphate buffer pH 7.3 and was

[1] Abbreviations used: PPACK, D-phenyl-prolyl-arginyl chloromethylketone; DAPA, dansyl arginine (*N*-ethyl-1,5-pentanediyl) amide; hirugen, sulfated Tyr63-*N*-acetyl hirudin 53–64.

mixed in a 1:1 ratio with the thrombin–hirugen–MOL inhibitor ternary complex solution and equilibrated over 1 ml of the well solution to give monoclinic crystals isomorphous with binary thrombin–hirugen crystals (Skrzypczak-Jankun *et al.*, 1991). Unlike the other complexes, the ternary complex of MOL356 crystallized differently in a hexagonal space group.

2.2. Intensity data collection

The X-ray diffraction data of the ternary complexes of MOL354, MOL376 and MOL592 were collected at 123 K with an R-AXIS II imaging-plate detector, using 15% glycerol as cryo-solvent in the well solution; those of MOL356 and MOL1245 were measured at room temperature. The radiation generated from a Rigaku RU-200 rotating-anode generator operating at 5 kW power with a fine-focus filament (0.3 × 3.0 mm) was monochromated (Cu $K\alpha$) and intensified by focusing with Molecular Structure Corporation/Yale Mirrors. The crystal-to-detector distance was 10.0 cm, with a swing angle of 0°; coverage completeness was in the range 85–90%. Except for the hexagonal MOL356 complex, the complexes scattered X-rays to about 2.3–2.0 Å resolution. Autoindexing and processing of the measured intensity data were carried out with the Rigaku *R-AXIS* software package (Higashi, 1990). Characteristics of the crystals and intensity data-collection statistics are given in Table 1.

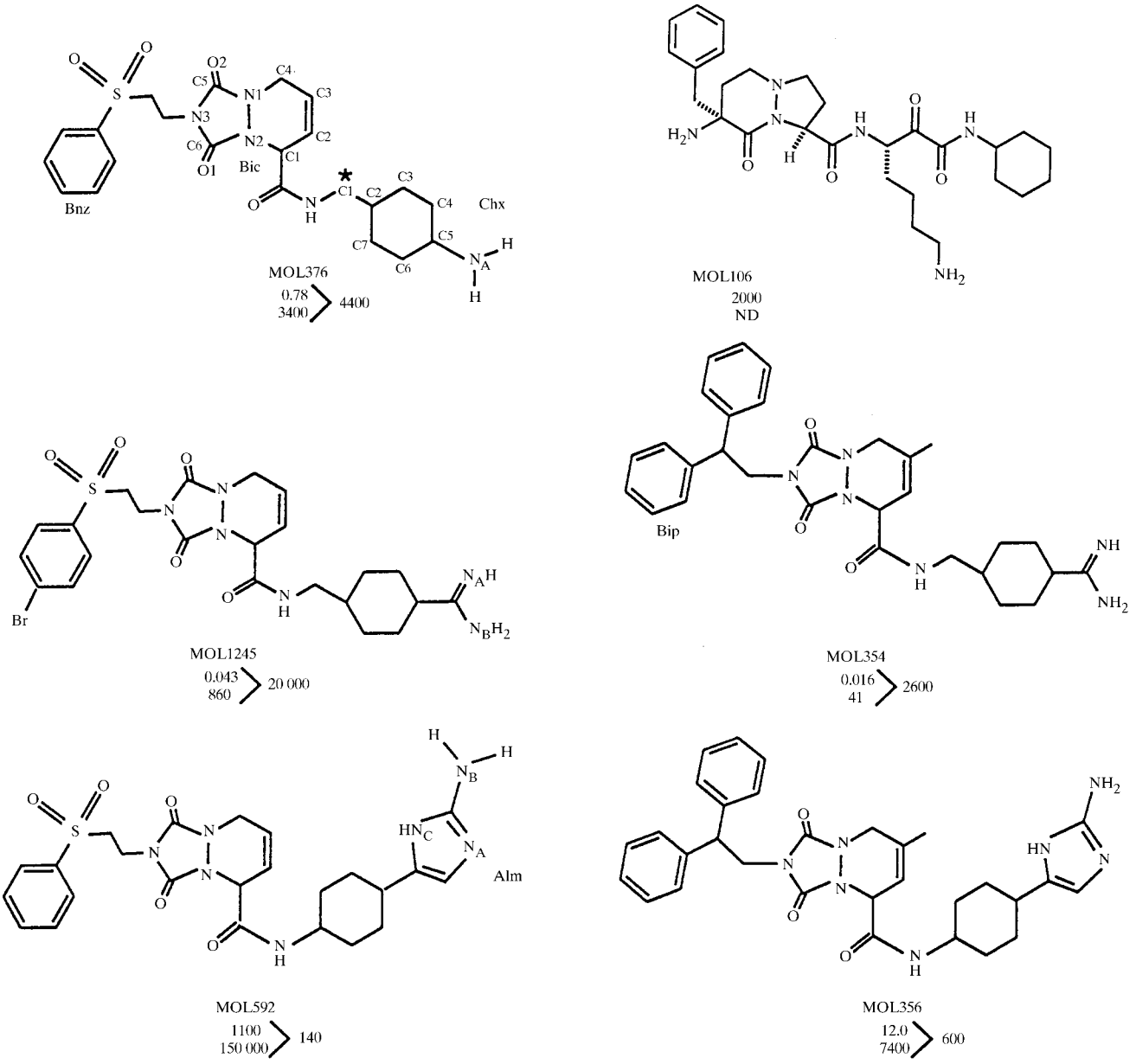

Figure 1
Selective non-electrophilic Molecumetics active-site inhibitors of thrombin. Numbering defined on MOL376 and other appropriate sites. Below the MOL number are shown the K_i values (nM) for thrombin and for trypsin; the ratio of the K_is is given on the right, showing the improved degree of thrombin inhibition. ND, not determined (Boatman *et al.*, 1999).

Table 1
Intensity data collection of Molecumetic inhibitor–thrombin complexes.

	MOL354	MOL356 (hexagonal)	MOL376	MOL592	MOL1245
Crystal dimensions (mm)	0.2 × 0.15 × 0.15	0.3 × 0.2 × 0.2	0.2 × 0.2 × 0.1	0.3 × 0.2 × 0.2	0.15 × 0.15 × 0.2
Space group	*C*2	$P6_5$	*C*2	*C*2	*C*2
Unit-cell parameters					
a (Å)	70.10	93.10	71.46	71.43	71.51
b (Å)	71.27	93.10	72.19	71.67	72.02
c (Å)	71.84	95.29	73.12	72.45	72.91
β (°)	100.34	120†	100.7	101.0	100.8
Molecules per asymmetric unit	1	1	1	1	1
Resolution (Å)	2.3	2.7	2.3	2.0	2.1
Observations	18059	23354	14987	33066	29913
Independent reflections	10994‡	6509	9294	17371	14337
Redundancy	1.6	3.6	1.6	1.9	2.1
Outermost shell (Å)	2.5–2.3	3.0–2.7	2.5–2.3	2.2–2.0	2.3–2.1
$R_{\rm merge}$ (%)	8.1	8.4	5.8	5.7	6.0
Outermost shell (%)	14.2	13.8	13.1	15.8	15.7
$I/\sigma(I)$ (outermost shell)	3.8	2.3	2.6	2.8	2.2

† γ angle. ‡ $I/\sigma(I) > 1.5$; for others, $I/\sigma(I) > 1.0$.

2.3. Structure determination

The MOL354, MOL376, MOL592 and MOL1245 complex crystals belong to the monoclinic space group *C*2, with four ternary complexes per unit cell. Their crystal structures were determined directly using the isomorphous thrombin coordinates of the binary thrombin–hirugen complex (Vijayalakshmi *et al.*, 1994; PDB code 1hah). The initial coordinates were optimized by rigid-body rotation–translation refinement in the 9.0–3.0 Å resolution range using the *X-PLOR* program package (Brünger, 1992). The structures were refined with the programs *PROLSQ* (Hendrickson, 1985) and *PROFFT* (Finzel, 1987), generally following procedures described previously (Mathews & Tulinsky, 1995; Matthews *et al.*, 1996).

The V_m value of crystals of the MOL356 complex is 3.25 Å Da^{-1} (38% protein fraction); systematically absent reflections along the *c* axis of the hexagonal crystals indicate that $l = 6n$ and that the space group is therefore either $P6_1$ or $P6_5$. The hexagonal crystal structure was determined by molecular replacement in the 20.0–4.0 Å resolution range using the program *AMoRe* (Navaza, 1994). Cross-rotation/translation functions were calculated for the two possible space groups using PDB code 1hah as the thrombin model. This gave correlation coefficients after rigid-body refinement of 64% ($P6_5$) and 35% ($P6_1$). Refinement proceeded smoothly with *PROLSQ* to $R = 14.4\%$ (with 20 water molecules).

The final stereochemistry of all the ternary complex structures was examined using the program *PROCHECK* (Laskowski *et al.*, 1993), which indicated that all had good molecular geometry. The root-mean-square deviations (r.m.s.Δs) of the bond distances, angle distances and planar 1,4 distances were in the ranges 0.016–0.023, 0.028–0.057 and 0.028–0.068 Å, respectively (target values were 0.02, 0.03 and 0.05 Å, respectively). The r.m.s.Δs of the isotropic thermal factors of the complexes were in the ranges 0.9–1.0, 1.6–1.9, 2.0–2.5 and 2.8–3.8 $Å^2$ (target values were 1.0, 1.5, 2.0 and 3.0 $Å^2$, respectively) for main-chain bonds, main-chain angles, side-chain bonds and side-chain angles, respectively. The omit maps of the two weakest diffracting complexes (MOL356 and MOL376) are shown in Fig. 2. Final refinement statistics are summarized in Table 2.

3. Results and discussion

3.1. Thrombin

Electron density is well defined for most of the residues of thrombin in all the complexes, except for the terminal regions of the *A* chain (Thr1H–Ala1B, Ile14K–Arg15) and the autolysis loop in the vicinity of Trp148–Lys149E, both of which are also typically disordered in most other isomorphous thrombin structures. The final thrombin structures of the ternary complexes are all similar, along with their accompanying hirugen peptides, and show no significant conformational differences from each other as well as from other thrombin structures (r.m.s.Δ < 0.4 Å between 240 and 250 CA positions). Two sodium ion sites have been found in the monoclinic crystal structures, similar to those described previously (DiCera *et al.*, 1995; Zhang & Tulinsky, 1997). One is intramolecular, the other is intermolecular; the former is responsible for a slow to fast kinetic transition in thrombin (Wells & Di Cera, 1992). Both sodium ions adopt slightly distorted octahedral arrangements coordinated by solvent water molecules and carbonyl O atoms: those of Arg221A and Lys224 for the intramolecular site, while those of Lys169 and Thr173 as well as Phe204A from a symmetry-related molecule coordinate the intermolecular ion (Zhang & Tulinsky, 1997). The intramolecular sodium site was also found in the hexagonal structure of the MOL356 complex.

3.2. MOL inhibitors

All the thrombin–MOL inhibitor complexes (except MOL356) have been refined at a resolution of 2.3 Å or better, whereas diffraction data of MOL356 only extended to 2.7 Å resolution (Tables 1 and 2). Therefore, the conclusions drawn from the structures of the complexes are essentially of comparable reliability. The inhibitors also have good electron density, except for one of the biphenyl rings of MOL354, which is only partially defined. The electron density of the cyclohexyl-aminoimidazole group of MOL356 could be fitted equally well with a boat or chair conformation for the cyclohexyl group. However, a *cis* attachment (axial/equatorial) would be required for the chair conformation, which was not synthetically possible, thus limiting the conformation to a boat form. All the inhibitors bind to thrombin in an extended conformation like substrate and the archetype inhibitor PPACK (Bode *et al.*, 1992).

Table 2
Individual refinement indicator parameters.

Parameter	MOL354	MOL356 (hexagonal)	MOL376	MOL592	MOL1245
Resolution range (Å)	7.0–2.3	7.0–2.7	9.0–2.3	9.0–2.0	7.0–2.1
No. of reflections	10518	5872	9092	17152	11633
$\|F_o\|/\sigma$ greater than	1.5	1.5	1.5	1.5	2.0
Observational completeness (%)	69	49	56	72	64
Outermost range (%)	50	15	38	42	47
R value (%)	17.6	14.4	15.3	17.4	16.5
No. of water molecules	98	20	98	134	100
$\langle B \rangle$ (Å^2)	25.5	20.1	28.4	28.3	26.6

3.3. P1 moieties

The inhibitors do not interact directly with Ser195 of the catalytic triad of thrombin (distances to Ser195 OG are all >3.5 Å) because they lack a terminal carboxylate or peptide-like group (Fig. 1). This is the only similarity between generic non-electrophilic inhibitors and the five MOL inhibitors. The P2–P1′ groups of DAPA-like inhibitors bind to thrombin with P1′ in the S2 site and P2 in the D-enantiomorphic S3 site, whereas the present MOL inhibitors bind like PPACK. Moreover, the P1 groups of the MOL inhibitors approach the carboxylate of Asp189 in the specificity pocket directly and not at an angle as in the DAPA-like counterparts (Banner & Hadvary, 1991; Brandstetter *et al.*, 1992; Mathews & Tulinsky, 1995). This lack of a direct interaction with Ser195 OG leaves the oxyanion hole of thrombin unoccupied as the P1 groups have no O atom to hydrogen bond with Gly193 N and Ser195 N of the hole, indicating that the dominant interactions of these potent inhibitors (Fig. 1) must reside among the S1–S3 binding subsites and not in transition-state intermediate formation.

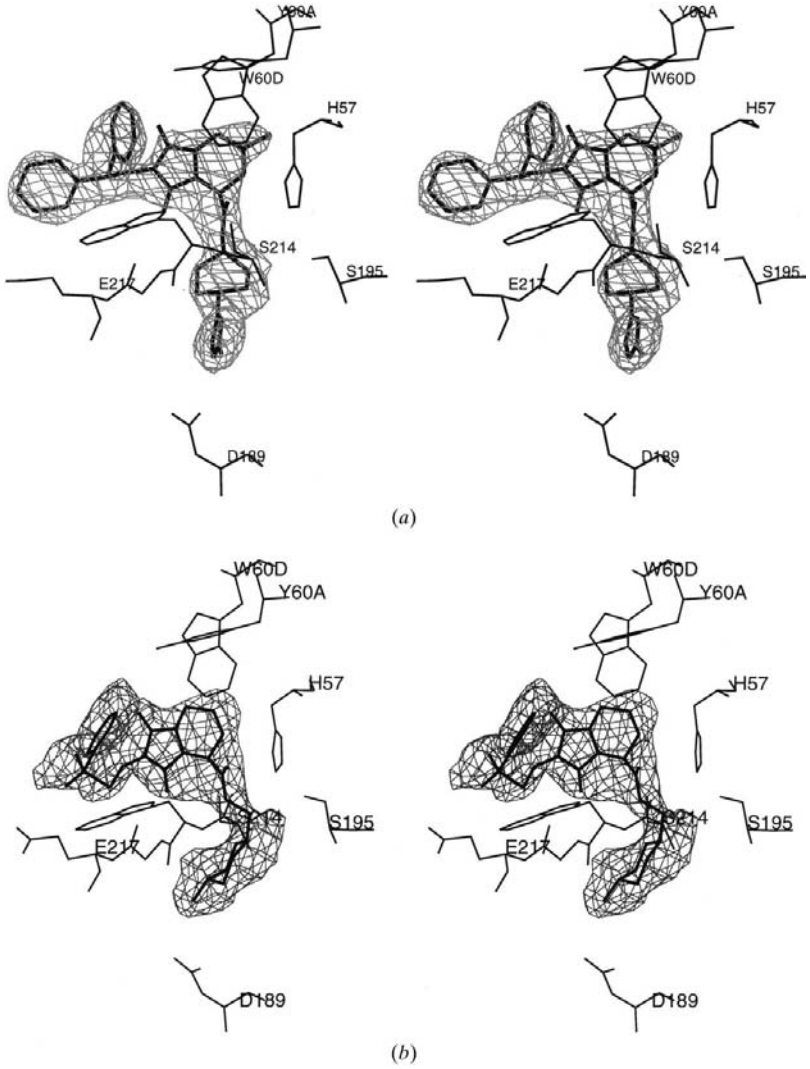

Figure 2
Stereoview of omit maps of (*a*) MOL356 and (*b*) MOL376. Contoured at 1σ level; inhibitors in bold lines; thrombin labeled and in thin lines.

The long fairly narrow S1 specificity pocket of thrombin, bottomed out with the negative charge of the carboxylate of Asp189, is ideally designed to accommodate an arginine group (Krishnan *et al.*, 1998). However, thrombin is known to bind a variety of moieties in the specificity pocket, such as cyclohexyl- and benzylamines, aminopyridyls and thiazoles (Tucker *et al.*, 1997; Feng *et al.*, 1997), guanidinopiperidyls (Krishnan *et al.*, 1998) and even a bulky hydrophobic group like indole (Malikayil *et al.*, 1997); the variety is even greater when arginine/guanidine bioisosteres of DAPA-related inhibitors are included (St Laurent *et al.*, 1995), which generally fit the S1 pocket more loosely (Mathews & Tulinsky, 1995). In most cases, the binding of all these alternate P1 side groups still tends to mimic the binding of arginine (Krishnan *et al.*, 1998).

The MOL inhibitors, like some of the foregoing, have an amino, amidino or amino-imidazole charged group attached to a cyclohexyl ring at the P1 position (Fig. 1). Not surprisingly, these charged groups form hydrogen-bonded salt bridges with the negatively charged carboxylate of the Asp189 side chain of the S1 pocket, with the P1 group of each inhibitor extended in the specificity pocket (Figs. 3 and 4). Differences in the chemical nature of the charged basic groups, however, and the overall size and exact conformation of the P1 groups produce variations in binding interactions and the role of solvent in the stabilization of the thrombin–MOL inhibitor complexes (Fig. 5). The cyclohexyl groups at the P1 position of three MOL inhibitors appear to be in the energetically favored chair conformation (Fig. 3), while the imidazole-substituted cyclohexyl rings (MOL356 and MOL592) are in a boat conformation (Fig. 4).

The cyclohexylamino group of MOL376 interacts with Asp189 of thrombin more like the lysine residue of the first-generation MOL106 inhibitor (Fig. 1; St Charles *et al.*, 1999) than the lysine analogue of DUP714 (boronic acid counterpart of PPACK; Weber *et al.*, 1995). This is most likely to be because of the similarity of the large hydrophobic bicyclic P2 groups of MOL106 and MOL376. The P1 group of MOL376 only resembles lysine in having a terminal amino group, so the interactions of the group differ from lysine DUP714. In the thrombin–MOL376 complex, the amino group makes a hydrogen-bonded salt bridge with Asp189 OD1 (Figs. 3 and 5), differing from the thrombin complex of lysine DUP714, where a hydrated doubly hydrogen-bonded water molecule mediates the amino group and the Asp189 carboxylate, with no other contacts between the two side chains. A direct contact, however, is observed between the amine and the carboxylate of Asp189 with a DUP714 homolysine inhibitor containing a methylene insertion in the P1 group (Weber *et al.*, 1995). In the MOL376 complex, the bond to the cyclohexyl ring is equatorial and an equatorial cyclohexylamino group occupies a position equivalent to one of the terminal N atoms of the guanidinium group of arginine- or amidino-based inhibitors (Figs. 3 and 5). Both the bonds are to the cyclohexyl apical atoms of a chair conformation. Other conserved interactions of such inhibitors are also observed in the S1 site (Zhang & Tulinsky, 1997); these are a hydrogen bond to Gly219 O and a doubly hydrogen-bonded water molecule mediating Asp189 OD2 and Tyr228 OH (Fig. 5). Another water molecule, O_w603, positionally mimics an NH1 or NH2 atom of thrombin-bound arginine and completes the interactions of the amino group bridging it and Asp189 OD2 (Fig. 5). The second solvent shell of the MOL376 complex includes the conserved water site O_w426 hydrogen bonded to Phe227 O and O_w603. Although Asp189 OD1 of thrombin and the cyclohexylamine of MOL376 hydrogen bond to a water molecule (O_w605) as in thrombin complexes with arginine-based inhibitors, unlike the latter no further interactions ensue from this water because it is positionally a little different. Some of the foregoing interactions have also been observed with other thrombin-bound inhibitors containing a cyclohexylamine or an aminopyridine group at the P1 position (Tucker *et al.*, 1997; Feng *et al.*, 1997).

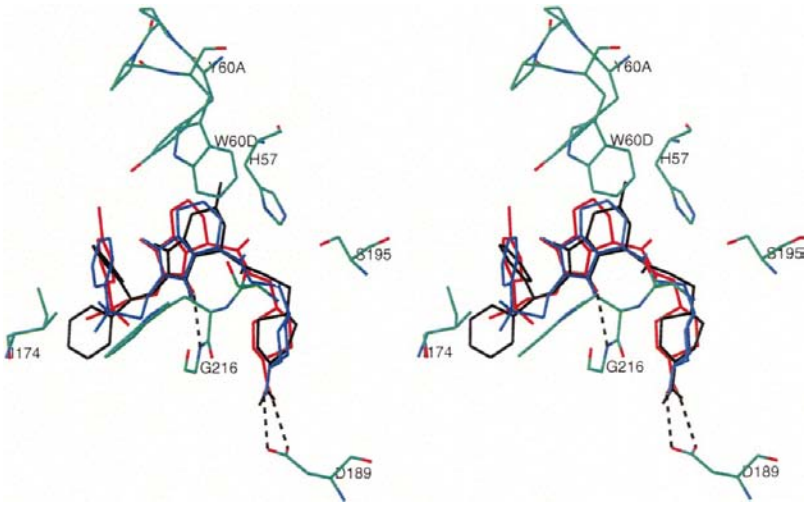

Figure 3
Stereoview of superposed MOL354, MOL376 and MOL1245 bound to thrombin. MOL354 (black), MOL376 (blue), MOL1245 (red); thrombin of MOL1245 complex numbered in atom colors: green, carbon; blue, nitrogen; red, oxygen; broken lines, principal hydrogen bonds. Note the chair conformation of MOL376.

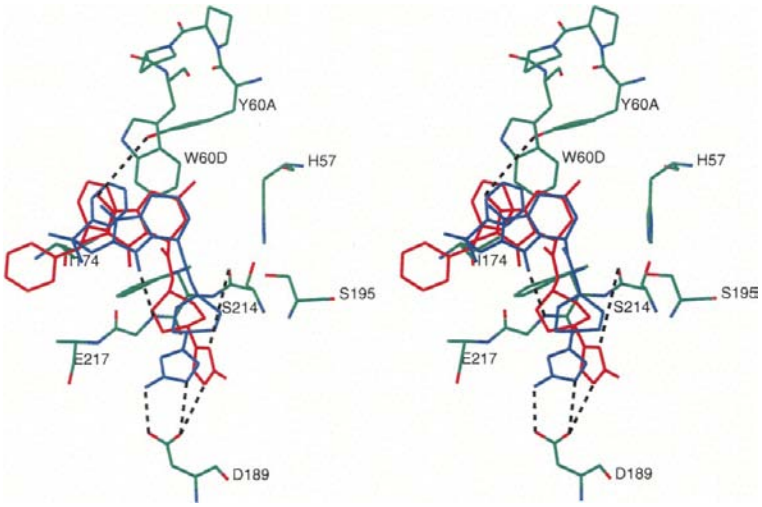

Figure 4
Stereoview of superposed MOL356 and MOL592 bound to thrombin. MOL356 (red), MOL592 (blue); thrombin of MOL356 complex numbered in atom colors; broken lines, principal hydrogen bonds.

Not surprisingly, the cyclohexylamidino groups of the MOL354 and MOL1245 inhibitors form doubly hydrogen-bonded salt bridges with the carboxylate group of Asp189, similar to arginine (Figs. 3 and 5). As in MOL376, the bonds to the cyclohexyl ring and to the amidino group are both equatorial and emanate from apical positions of the chair, but the cyclohexylamidino rings differ by being rotated about 80° with respect to the ring of MOL376 (Fig. 3): this is caused by O_w637 of the MOL376 structure, which hydrogen bonds with Chx N (2.4 Å), Bic O1 (2.8 Å) and Gly216 O (3.2 Å) but is not present in the S1 vicinity of MOL354 and MOL1245. Conserved hydrogen bonds of guanidinium groups in the specificity pocket (Gly219 O, water bridge to Phe227 O) are also found, along with water-mediated interactions between the carboxylate O atoms of Asp189, Glu217 O and Tyr228 OH (Fig. 5). However, in the MOL354 complex, the water bridging Phe227 O and the amidino group appears to be a little closer to Phe227 O (2.7 Å) and further (3.4 Å) from the amidino group.

The absence of an N atom corresponding to NE of arginine-based inhibitors in the amino and amidino inhibitors (Fig. 1) leads to the loss of the neighboring water molecule that also hydrogen

bonds to a hydrogen-bonding network involving Glu192, Gly216 and Gly219 (Zhang & Tulinsky, 1997). The Glu192 side chain, which is part of the network, is sometimes disordered in structures lacking this water molecule (Krishnan *et al.*, 1998). The Glu192 side chain is in the 'up' conformation in all the thrombin–MOL structures as in the thrombin–hirugen complex (Vijayalakshmi *et al.*, 1994), with anywhere from one to three water molecules hydrating and apparently stabilizing its conformation. Another factor contributing to the network disruption in the complexes is that a portion of the bulky hydrophobic cyclohexyl ring protrudes out along the surface of the S1 pocket, similar to that of guanidino piperidyl groups (Krishnan *et al.*, 1998), thereby interfering with the otherwise stabilizing protection of the Glu192 side chain in the 'bent' conformation which shields the specificity P1 side group from bulk solvent. The MOL354 and MOL1245 structural environments differ near Asp189 OD1, where an additional water molecule O_w425 is found in the MOL354 structure (Fig. 4). This changes some of the contact distances because O_w461 of MOL1245 is located between O_w408 and O_w425 of MOL354. The water structure in this region tends to vary somewhat because space is limited and can be easily encroached upon by the functionality of the P1 group. Sometimes one of the water molecules in this vicinity is even linked to Arg221A N of the sodium ion binding site (Fig. 5).

Both thrombin–MOL356 and thrombin–MOL592 have cyclohexyl-aminoimidazole groups occupying the S1 site (Fig. 1), but since the amide N atom lacks a methylene C atom, it is linked directly to the cyclohexyl ring; the latter therefore does not penetrate as deeply into the S1 site and only overlaps about half of the cyclohexyl ring of MOL354, MOL376 and MOL1245 (Fig. 6). This is evidenced further by the P2–P3 groups of MOL356 being additionally displaced away from the pocket toward the 60-insertion loop (Figs. 4 and 6). Surprisingly, the interactions of the aminoimidazole groups are very different in the specificity pocket of the MOL356 and MOL592 complexes (Figs. 4 and 5). The cyclohexyl ring is in a boat conformation in both MOL356 and MOL592. Furthermore, the imidazole ring is parallel to the plane of the boat in MOL592, whereas it is rotated about 75° in MOL356, with the stereochemistry of the bonds to the cyclohexyl and imidazole groups (both apical positions) being axial and equatorial, respectively. The MOL592 inhibitor forms a doubly hydrogen-bonded salt bridge with Asp189 through an N atom of the imidazole ring and its amino group, unlike MOL356, which interacts with Asp189 uniquely (Fig. 5). Thus, the terminal imidazole of the P1 group of MOL592 is more like a guanidinium of arginine, utilizing all three N atoms (Figs. 4 and 5). This aminoimidazole mimics arginine further by having other conserved hydrogen-bonded interactions found in thrombin–arginine complexes (Fig. 5). The long axis through the cyclohexylimidazole group of the MOL356 inhibitor makes an angle of about 30° to that of MOL592 (the two axes intersecting near the apical cyclohexyl atom carrying the imidazole; Fig. 4). The imidazole ring of MOL356 is flipped 180° with respect to MOL592 about its bond to the cyclohexyl ring and only one imidazole N atom forms a hydrogen bond with Asp189 OD2, *while the amino N atom makes a direct hydrogen-bonding contact with Phe227 O* occupying the position of a conserved water molecule of other complexes (Fig. 5). This water molecule is linked to the charged terminal guanidinium or amidine group by a hydrogen bond in P1 arginyl or amidino inhibitors. The lack of the methylene C atom in MOL356 and MOL592 notwithstanding, the boat conformation of the cyclohexyl ring and the equatorial bond to the imidazole ring of MOL356 lead to a deeper lateral penetration into the interior of the S1 pocket, thereby shifting the P2–P3 groups 1.0–1.5 Å toward the 60-insertion loop (Figs. 4 and 6). The same also generally applies to MOL592, but to a lesser extent. This penetration is far enough in MOL356 for the other ring N atom of the imidazole (N_C) to make a unique contact and *possibly*

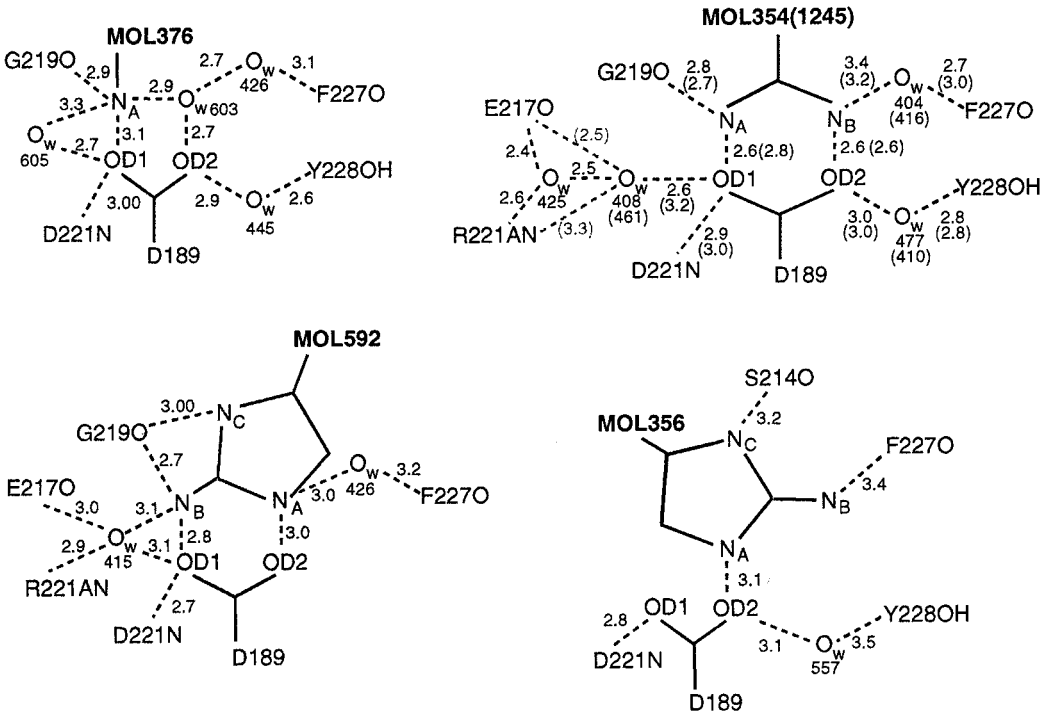

Figure 5
Schematic of hydrogen-bonded salt bridges of Molecumetics inhibitors with Asp189 of the S1 specificity site of thrombin. MOL354 and MOL1245 shown together in one schematic; values for MOL1245 are in parentheses.

Table 3
MOL1245 bromine close contacts with thrombin (Å).

Tyr60A OH	3.8
Trp96 O	3.8
Glu97A C	3.6
Glu97A O	3.2
Asn98 N	3.6
Asn98 C	3.8
Asn98 CA	3.7
Leu99 N	3.6
Leu99 CD1	3.9
O_w406	3.4
O_w433	3.5

a longer hydrogen bond with Ser214 O (Fig. 5). In the other MOL complexes, Ser214 O contacts the amide N atom of the P1 moiety.

A point about which something should be said is the hydrogen bond between Asp189 OD1 and Asp221 N of thrombin, which appears to operate in concert with a water molecule mediating Asp189 OD2 and Tyr228 OH (Fig. 5). The two produce a steering effect on the orientation of the carboxylate group of Asp189. The net effect is to position and hold the carboxylate O atoms ideally for a doubly hydrogen-bonded interaction. This geometry occurs frequently for inhibitors of thrombin (Lau *et al.*, 1995), but is rare in other protein structures (less than 2%; Singh *et al.*, 1987). Since the steering interactions also occur in the unoccupied S1 site of the thrombin–hirugen complex (Vijayalakshmi *et al.*, 1994), they are most likely to be responsible for the high incidence of doubly hydrogen-bonded salt bridges found in the S1 specificity site of inhibitor-bound thrombin structures.

3.4. P2 residue

All the P2 hetero-bicyclic (Bic) groups are practically planar, with some small deviations from planarity. The Bic group of each inhibitor occupies the apolar S2 and part of the S3 region of the active site, making hydrophobic contacts with the side chains from His57, Tyr60A, Trp60D and Trp215 (Figs. 3 and 4) like other similar bicyclic groups (St Charles *et al.*, 1999). The six-membered ring occupies approximately the same position as the proline of PPACK in the thrombin–PPACK complex (Bode *et al.*, 1992).

The O1 atoms of the Bic groups of the inhibitors form a hydrogen bond with Gly216N like the bicyclic P2 group of first generation Molecumetics inhibitors (St Charles *et al.*, 1999). This interaction is in all the thrombin–MOL structures (<3.0 Å) and corresponds to one of the three usually found in the antiparallel β-strand between the P2–P3 residues of the inhibitor and the Ser214–Gly216 stretch of thrombin. The (usually) accompanying interaction from a P3 N atom to Gly216 O is missing with the present MOL inhibitors because they lack the N atom. A slightly longer hydrogen-bond contact, which varies between 3.1 and 3.3 Å, in the antiparallel β-stretch is found between the amide N atom of the MOL354, MOL376 and MOL1245 inhibitors and Ser214 O. It is longest in the MOL1245 complex, along with a correspondingly long Bic O1–Gly216 N distance (3.4 Å). The longer distance is a consequence of a subtle displacement of the Bic ring in the direction of the 60A–D segment of the insertion loop of thrombin (Fig. 3), which is most likely to be the result of the numerous bromine interactions of its P3 group with thrombin (Table 3). Since the P2–P3 groups of MOL356 are displaced 1.0–1.5 Å toward the 60-insertion loop because the imidazole of the P1 group is laterally buried deeper in the S1 pocket, its Bic O2 atom makes a unique hydrogen bond directly with Tyr60A OH (Fig. 4). The hydrogen bond has previously not been seen in any thrombin–inhibitor complexes and increases the number of direct hydrogen-bond contacts of MOL356 with thrombin to four (Fig. 5).

3.5. P2–P3 sulfonyl bridge

The sulfonyl groups of MOL376, MOL592 and MOL1245 are located on the surface and have no interactions with thrombin, with only an O atom of MOL1245 hydrogen bonding to a water molecule. This is in agreement with other P3 benzylsulfonyl inhibitors bound to thrombin (Krishnan *et al.*, 1998). In comparable complexes with trypsin, the benzyl group is disordered and the sulfonyl group is positionally displaced and interacts with Gly218N through a hydrogen bond (Krishnan *et al.*, 1998). A similar hydrogen bond has been reported with thrombin and two benzylsulfonamide inhibitors (Tucker *et al.*, 1997; Feng *et al.*, 1997). However, these particular inhibitors have a D-Phe at P3 which occupies P3 with the benzylsulfonated N-terminal of the inhibitor. The benzyl group thus occupies a uniquely different position close to a P1 cyclohexyl or pyridylamine.

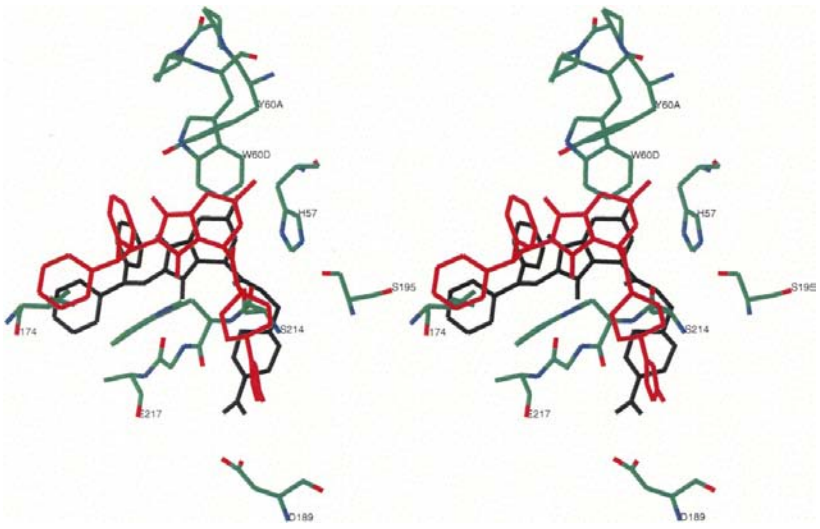

Figure 6
Stereoview of superposed MOL354 and MOL356 bound to thrombin. MOL354 (black), MOL356 (red); thrombin of MOL354 complex numbered in atom colors. Note 1.0–1.5 Å shift of MOL356.

3.6. S3 residue

The benzyl and bromobenzyl groups of MOL376, MOL592 and MOL1245 curl up toward the 60-insertion loop of thrombin and make numerous hydrophobic interactions with thrombin residues Asn98–Leu99, Ile174 and

Trp215 of the D-enantiomorphic S3 binding site. The edge of the benzyl ring approaches the face of the indole of Trp215 in a fashion typical of aromatic–aromatic interactions in proteins (Figs. 3 and 4). The P3 biphenyl rings in MOL354 and MOL356 are positioned similarly, with one phenyl ring occupying the D-S3 aromatic site while the other is positioned slightly above the L-S3 site (Fig. 4). The latter surface site accommodates the aspartate of the Leu-Asp-Pro-Arg sequence of thrombin platelet receptor (Mathews *et al.*, 1994). This mode of biphenyl binding has also been observed with other peptidic thrombin inhibitors (Tucker *et al.*, 1997).

The 4-Br atom of the P3 benzylsulfonyl group of MOL1245 makes 11 impressive close van der Waals contacts with thrombin (Table 3). The expected covalent radius of bromine is about 1.14 Å, with a C—Br distance of 1.8–1.9 Å, so a van der Waals contact with carbon, nitrogen or oxygen is in the neighborhood of 3.6–3.8 Å. Since a Br atom is fairly polarizable, it is unsurprising that most of the contacts are polar. Moreover, several of the interactions (Table 3) involve residues of the S4 binding site of thrombin (Mochalkin & Tulinsky, 1999) and factor Xa (Brandstetter *et al.*, 1996). It is of importance to note that the Arg93–Arg101 stretch has been implicated in thrombin–heparin binding (Karshikov *et al.*, 1992; Gan *et al.*, 1994; Ye *et al.*, 1994; Sheehan *et al.*, 1994).

3.7. Thrombin–trypsin binding constants

A final point about which something should be said is the trends of the respective binding constants of the MOL inhibitors for thrombin and trypsin. The two best inhibitors for thrombin are MOL354 and MOL1245, both having a cyclohexylamidino P1 group and sub-nanomolar binding constants (Fig. 1). Although the P3 groups are different, from the binding constants their contribution to binding must be comparable. The source of the tenacious sub-nanomolar binding appears to reside in the amidino groups mimicking the guanidinium of arginine and the binding of the benzylsulfonyl and one of the biphenyl P3 groups in the D-S3 site as in the D-Phe of the thrombin–PPACK complex. The majority of the loss of a factor of about 20–30 in thrombin binding of MOL376 is clearly the result of the replacement of the amidino with a simple amine group at P1. Although the amine of MOL376 copies amidino binding to some extent (Fig. 5), from the binding constants the overall energetics of binding appears to be weakened. The cyclohexyl-aminoimidazole inhibitors are the poorest inhibitors of thrombin (Fig. 1), which must be because of their bulky and unwieldy P1 group. It is difficult, however, to account for the markedly poorer binding constants of MOL356 and MOL592 from their specificity site hydrogen-bonding interactions (Fig. 5), which suggest comparable binding to the other MOL inhibitors, especially for MOL592. Thus, entropic effects arising from the conformation of the inhibitor required to make the hydrogen- bond interactions may make significant contributions to the binding. That MOL356 is a better inhibitor of thrombin than MOL592 by two orders of magnitude could arise from its having four spatially well distributed and direct binding interactions with thrombin rather than a larger number of less definitive water-mediated ones (Fig. 5).

The three best thrombin inhibitors (MOL354, MOL376 and MOL1245) also are the most selective (two to three orders of magnitude) against trypsin (Fig. 1). This is primarily because of the inability of trypsin to bind a P3 benzylsulfonyl (Krishnan *et al.*, 1998), which probably also applies to the rings of the P3 diphenyl group of MOL354. The difference of a factor of 20 between MOL356 and MOL592 for trypsin is in the same direction and of comparable magnitude as the thrombin-binding constants, suggesting that the details of the binding in the S1 site of trypsin may be similar to those found in thrombin (Fig. 5).

This work was supported by NIH Grant HL43229 and in part by Molecumetics Ltd. We would also like to thank Molecumetics Ltd for providing the inhibitors and their binding constants for thrombin and trypsin, and Hiroshi Nakanishi for numerous discussions about the work.

References

Balasubramanian, B. N. (1995). *Bioorg. Med. Chem.* **3**, 999–1156.
Banner, D. & Hadvary, P. (1991). *J. Biol. Chem.* **266**, 20085–20093.
Boatman, P. D., Ogbu, C. O., Eguchi, M., Kim, H.-O., Nakanishi, H., Cao, B., Shea, J. P. & Kahn, M. (1999). *J. Med. Chem.* **42**, 1367–1375.
Bode, W., Turk, D. & Karshikov, A. (1992). *Protein. Sci.* **1**, 426–471.
Brandstetter, H., Kuhne, A., Bode, W., Huber, R., van der Saal, W., Wirthensohn, K. & Engh, R. A. (1996). *J. Biol. Chem.* **271**, 29988–29992.
Brandstetter, H., Turk, D., Hoeffken, H. W., Gross, D., Sturzebecher, J., Martin, P. D., Edwards, B. F. P. & Bode, W. (1992). *J. Mol. Biol.* pp. 1085–1099.
Brünger, A. T. (1992) *X-PLOR Version* 3.1. *A System for Crystallography and NMR*. New Haven: Yale University Press.
Claeson, G., Phillip, M., Agner, E., Scully, M. F., Metternich, R., Kakkas, V. V., Desoyza, T. & Niu, L. (1993). *Biochem. J.* **290**, 309–315.
Colman, R. W., Marder, V. J., Salzman, E. W. & Hirsch, J. (1994). *Hemostasis and Thrombosis*, 3rd ed., edited by R. W. Colman, J. Hirsch, V. Marder & E. W. Salzman, pp. 3–18. Philadelphia: J. B. Lippincott Co.
Davie, E. W., Fujikawa, K. & Kisiel, W. (1991). *Biochemistry*, **30**, 10363–10370.
DiCera, E., Guinto, E. R., Vindigni, A., Dang, Q. O. Alaya, Y. M., Wuyi, M. & Tulinsky, A. (1995). *J. Biol. Chem.* **270**, 22089–22092.
Elgendy, S., Deadman, J., Patel, G., Green, D., Chino, N., Goodwin, C. A., Scully, M. F., Kakkar, V. V. & Claeson, G. (1992). *Tetrahedron Lett.* **33**, 4209–4212.
Esmon, C. T. (1993). *Thromb. Haemost.* **70**, 29–35.
Feng, D. M., Gardell, S. J., Lewis, S. D., Bock, M. G., Chen, Z., Freidinger, R. M., Naylor-Olsen, A. M., Ramjit, H. G., Woltmann, R., Baskin, E. P., Lynch, J. J., Lucas, R., Schafer, J. A., Dancheck, K. B., Chen, I.-W., Mao, S. S., Krueger, J. A., Hare, T. R., Mulichak, A. M. & Vacca, J. P. (1997). *J. Med. Chem.* **40**, 3726–3733.
Finzel, B. C. (1987). *J. Appl. Cryst.* **20**, 53–55.
Gan, Z.-R., Li, Y., Chen, Z., Lewis, S. D. & Shafer, J. A. (1994). *J. Biol. Chem.* **269**, 1301–1305.
Hendrickson, W. A. (1985). *Methods Enzymol.* **115**, 252–270.
Higashi, T. (1990). *J. Appl. Cryst.* **23**, 253–257.
Kahn, M., Wilke, S., Chen, B. & Fujita, K. (1988). *J. Am. Chem. Soc.* **110**, 1638–1639.

Karshikov, A., Bode, W., Tulinsky, A. & Stone, S. R. (1992). *Protein Sci.* **1**, 727–735.
Krishnan, R., Zhang, E., Hakansson, K., Arni, R. K., Tulinsky, A., Lim-Wilby, M. S. L., Levy, O. E., Semple, J. E. & Brunck, T. K. (1998). *Biochemistry*, **37**, 12094–12103.
Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. (1993) *J. Appl. Cryst.* **26**, 283–286.
Lau, W. F., Tabernero, L., Sack, J. S. & Iwanowicz, E. J. (1995). *Bioorg. Med. Chem.* **3**, 1039–1048.
Locht, A., Bode, W., Huber, R., Le Bonniec, B. F., Stone, S. R., Esmon, C. & Stubbs, M. T. (1997). *EMBO J.* **16**, 2977–2984.
Malikayil, J. A., Burkhart, J. P., Schreuder, H. A., Brocrsma, R. J. Jr, Tardif, C., Kutcher, L. W. III, Mehdi, S., Schatzman, G. L., Neizes, B. & Peet, N. P. (1997). *Biochemistry*, **36**, 1034–1040.
Mathews, I. I., Padmanabhan, K. P., Ganesh, V., Tulinsky, A., Ischii, M., Chen, J., Turck, C. W., Coughlin, S. R. & Fenton, J. W. (1994). *Biochemistry*, **33**, 3266–3276.
Mathews, I. I. & Tulinsky, A. (1995). *Acta Cryst.* D**51**, 550–559.
Matthews, J. H., Krishnan, R., Costanzo, M. J., Maryanoff, B. E. & Tulinsky, A. (1996). *Biophys. J.* **71**, 2830–2839.
Mochalkin, I. & Tulinsky, A. (1999). *Acta Cryst.* D**55**, 785–793.
Nakanishi, H., Chrusciel, R. A., Shen, R., Bertenshaw, S., Johnson, M. E., Rydel, T. J., Tulinsky, A. & Kahn, M. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 1705–1709.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Ogbu, C. O., Qabar, M. N., Boatman, P. D., Urban, J., Meara, J. P., Ferguson, M. D., Tulinsky, J., Lum, C., Babu, S., Blaskovich, M. A., Nakanishi, H., Ruan, F., Cao, B., Minarik, R., Little, T., Nelson, S., Nguyen, M., Gall, A. & Kahn, M. (1998). *Bioorg. Med. Chem. Lett.* **8**, 2321–2326.
Sheehan, J. P., Tollefsen, D. M. & Sadler, J. E. (1994). *J. Biol. Chem.* **269**, 32747–32751.
Singh, J., Thornton, J. M., Sharey, M. & Campbell, S. F. (1987). *FEBS Lett.* **224**, 161–171.
Skrzypczak-Jankun, E., Carperos, V., Ravichandran, K. G., Tulinsky, A., Westbrook, M. & Maraganore, J. (1991). *J. Mol. Biol.* **221**, 1379–1393.
St Charles, R., Matthews, J. H., Zhang, E., Tulinsky, A. & Kahn, M. (1999). *J. Med. Chem.* **42**, 1376–1383.
St Laurent, D. R., Balasubramanian, N., Han, W. T., Trehan, A., Federici, M. E., Meanwell, N. A., Wright, J. J. & Seiler, S. M. (1995). *Bioorg. Med. Chem.* **3**, 1145–1156.
Tabernero, L., Chang, C. Y., Ohringer, S. L., Lau, W. F., Iwanowicz, E. J., Han, W. C., Wang, T. C., Seiler, S. M., Roberts, D. G. M. & Sack, J. S. (1995). *J. Mol. Biol.* **246**, 14–20.
Tapparelli, C., Metternich, R., Ehrhardt, C., Zurini, M., Claeson, G., Scully, M. F. & Stone, S. R. (1993). *J. Biol. Chem.* **268**, 4734–4741.
Tucker, T. J., Lumma, W. C., Mulichak, A. M., Chen, Z., Naylor-Olsen, A. M., Lewis, S. D., Lucas, R., Freidinger, R. M. & Kuo, L. C. (1997). *J. Med. Chem.* **40**, 830–832.
Vijayalakshmi, J., Padmanabhan, K. P., Mann, K. G. & Tulinsky, A. (1994). *Protein Sci.* **3**, 2254–2271.
Weber, P. C., Lee, S. L., Lewandowski, F. A., Schadt, M. C., Chang, C. H. & Kettner, C. A. (1995). *Biochemistry*, **34**, 3750–3757.
Wells, C. M. & Di Cera, E. (1992). *Biochemistry*, **31**, 11721–11730.
Wu, T.-P., Yee, V., Tulinsky, A., Chrusciel, R. A., Nakanishi, H., Shen, R., Priebe, C. & Kahn, M. (1993). *Protein Eng.* **6**, 471–478.
Ye, J., Rezaie, A. R. & Esmon, C. T. (1994). *J. Mol. Biol.* **269**, 17965–17970.
Zhang, E. & Tulinsky, A. (1997). *Biophys. Chem.* **63**, 185–200.