Application of Fragment Screening and Fragment Linking to the Discovery of Novel Thrombin Inhibitors[†]

Nigel Howard,[§] Chris Abell,[§] Wendy Blakemore,[‡] Gianni Chessari,[‡] Miles Congreve,[‡] Steven Howard,*[,‡] Harren Jhoti,[‡] Christopher W. Murray,[‡] Lisa C. A. Seavers,[‡] and Rob L. M. van Montfort[‡]

Astex Therapeutics, 436 Cambridge Science Park, Milton Road, Cambridge CB4 0QA, United Kingdom, and University Chemical Laboratory, Lensfield Road, Cambridge, CB2 1EW, United Kingdom

Received August 26, 2005

The screening of fragments is an alternative approach to high-throughput screening for the identification of leads for therapeutic targets. Fragment hits have been discovered using X-ray crystallographic screening of protein crystals of the serine protease enzyme thrombin. The fragment library was designed to avoid any well-precedented, strongly basic functionality. Screening hits included a novel ligand (**3**), which binds exclusively to the S2−S4 pocket, in addition to smaller fragments which bind to the S1 pocket. The structure of these protein−ligand complexes are presented. A chemistry strategy to link two such fragments together and to synthesize larger drug-sized compounds resulted in the efficient identification of hybrid inhibitors with nanomolar potency (e.g., **7**, IC$_{50}$ = 3.7 nM). These potent ligands occupy the same area of the active site as previously described peptidic inhibitors, while having very different chemical architecture.

Introduction

Thromboembolic disorders are a major cause of mortality in Western societies. At present, clinical treatment of thrombosis involves the administration of heparin and its low molecular weight derivatives or oral anticoagulants of the dicumarol type, which all indirectly inhibit the trypsin-like serine protease thrombin. These drugs have limitations with regard to their efficacy and also their therapeutic index which, being low, leads to the need for extensive monitoring. Thrombin has been under intense investigation for over a decade with the aim of identifying potent inhibitors that might be useful as anticoagulation agents.[1−3] An orally active thrombin inhibitor should be useful for treatment of pathological states characterized by thrombosis, such as deep vein thrombosis and stroke.[4] Thrombin acts in the blood coagulation cascade by catalyzing the conversion of fibrinogen to fibrin and also converting factor XIII to factor XIIIa, which then cross-links the fibrin clot. Thrombin also activates upstream zymogens factors V, VIII, and XI, which then further accelerate the clotting cascade by thrombin synthesis.[5]

An injectable form of a direct thrombin inhibitor, argatroban, (Chart 1) has been approved by the FDA, but only for the relatively rare condition of heparin-induced thrombocytopenia.[6] Very recently, the first orally active thrombin inhibitor, ximelagatran (Exanta), was licensed for the treatment of venous thromboembolic events in Europe. Ximelagatran is a double-prodrug molecule (Chart 1) and is converted in vivo to the active parent (melagatran) that contains a highly basic amidine functionality which is key to inhibitor potency through salt bridge formation with aspartate 189 in the S1 specificity pocket.[7,8] This amidine group has been incorporated in many

Chart 1. Licensed Drugs that Directly Inhibit Thrombin

argatroban

ximelagatran, R^1 = OH, R^2 = Et
melagatran,   R^1 = R^2 = H
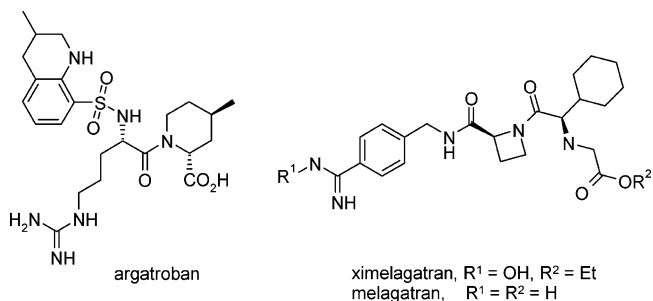
classes of thrombin inhibitors, but while such compounds demonstrate good in vivo efficacy, their use is generally limited by poor oral bioavailability. Although a prodrug approach offers one solution to this problem, the challenge over recent years has been to develop second-generation compounds that lack a highly basic P1 motif and therefore have superior druglike properties.[9]

Over the last 20 years there has been considerable interest in new approaches to drug discovery that offer improvements in the process of identifying new therapeutic agents. Large investments in platform technologies such as high-throughput screening and combinatorial chemistry have been made across the pharmaceutical industry. One area that has received significant attention is the drug-likeness of hits and leads, because the quality of lead compounds is thought to have a major impact on the attrition rates in drug development.[10−12] More recently, interest has grown in a different approach to lead generation that involves screening libraries of molecules that are much smaller and functionally simpler than drugs themselves, often referred to as "fragment-based" drug discovery.[13−21] Fragments are low molecular weight compounds (typically 100−250 Da) and as a consequence will usually exhibit low binding affinities (100 μM−mM). A fragment-based approach is argued to have advantages over conventional screening methods for several reasons.[12,18] First, the number of compounds that need to be screened to find useful hits is in the order of only a few hundreds, because the lower complexity compounds (fragments) have a higher probability of matching a target protein binding

site. Second, a higher proportion of the atoms in a low molecular weight fragment hit are directly involved in the desired protein–ligand binding interaction, so fragments can be seen as more efficient binders when compared with conventional (HTS) screening hits. Last, when guided by protein–ligand crystal structures, fewer molecules will need to be synthesized in order to optimize the hits to potent leads, and the chemical synthesis approach itself will often be simple and straightforward due to the low level of complexity of the initial fragment hits.

Recently, Congreve et al. analyzed a diverse set of fragment hits that had been identified against a range of targets.[22] This study indicated that such hits seemed to obey, on average, a "rule of three" in which molecular weight is <300, the number of hydrogen bond donors and acceptors is, in each case, ≤3, and clogP ≤ 3. The authors suggest that this "rule of three" provides a useful guideline for physicochemical properties when designing fragment-based screening sets.

Due to the low binding affinity of fragments they are usually difficult to detect using bioassay-based screening methods. A variety of alternative biophysical methods have therefore been used to detect the binding of such fragments, and the area has recently been reviewed.[18–20,23] Recently, Nienaber et al. identified novel inhibitors for urokinase using protein X-ray crystallographic screening.[21] In our laboratories we have developed methodologies for screening fragments, that we call "Pyramid", which include the high-throughput X-ray protein crystallographic screening of fragment libraries.[24,25] In our previous reports the subsequent optimization of hits identified from fragment screening has been based on a "fragment-growing" strategy. This involves using structure-based drug design to form additional interactions by growing out from the starting fragment. In this current work, we describe an example of an alternative, "fragment-linking", approach. This general approach has been described by others and is associated with its own particular set of difficulties.[13,14,26]

We have already reported our preliminary findings of fragment-based crystallographic screening against thrombin in a previous publication.[24] We now describe the fragment hits identified from Pyramid screening in more detail and how the subsequent structure-based fragment-linking strategy led to the identification of novel, potent, thrombin inhibitors. In the screen we soaked libraries of fragments into thrombin crystals with the aim of identifying new leads for this important drug target. In particular, our rationale was to exclude fragments that contained a strongly basic motif, such as an amidine or guanidine, and instead to focus on identifying hits with better druglike properties. We also wished to establish our approach as a useful method for identifying potent ligands for a representative protease target. As a class of enzymes, proteases have often been challenging as drug discovery targets since leads against these targets are often large and peptidic in nature, making the development of druglike inhibitors difficult.[27]

Results and Discussion

Overview of Pyramid Screening. The Pyramid screening approach and its application to a number or targets, including thrombin, has been described previously.[24] Further details on specific steps in the application to thrombin are described in other sections. In brief, thrombin screening consisted of the following steps: First, thrombin crystals suitable for soaking experiments were generated, and the robustness of soaking techniques against a variety of fragments and fragment mixtures was demonstrated. Second, a library of 80 fragments was defined

using virtual screening as described below. Third, thrombin crystals were soaked in compound mixtures, followed by elucidation of the protein–ligand structures. Fragment hits were identified in Fo–Fc electron density maps calculated after initial automatic refinement against an unliganded thrombin structure and fitted and scored using AutoSolve.[17] Protein–ligand structures of bound fragments were then subjected to further refinement steps. The output from the Pyramid process was a set of experimentally determined binding modes for thrombin-bound fragments which were then manually examined using AstexViewer.[28] Ligands chosen for hit-to-lead chemistry were selected based on a number of criteria which are described below.

Construction of a Focused Fragment Library for Thrombin. Virtual screening was used to build a focused set of fragments to screen against thrombin. Filtering on the basis of 1D and 2D properties was employed to create databases of compounds from the Astex Therapeutics library of available substances (ATLAS).[29] 1D and 2D properties included heavy atom count (HAC), number of hydrogen bond donors (NDON), and clogP. Compounds in these databases were subsequently docked against different protein conformations of thrombin using a proprietary version of the program GOLD.[30] This method has been previously reported by Hartshorn et al.[24] Pharmacophores were designed to reward the placement of hydrophobic groups in the S1 pocket and subsequently applied during selected dockings. Goldscore[31] and Chemscore[32] scoring functions were used to score and to rank the different fragments.

This approach generated a library of 80 fragments, which could be categorized as follows:

(1) Charged S1 binders (HAC ≤ 13, NDON ≤ 4, clogP ≤ 3): fragments carrying a weakly basic group able to interact with Asp189 (e.g., aminopyridine and aminoquinoline fragments). Strong bases and obvious arginine mimetics were avoided.

(2) Neutral S1 binders (HAC ≤ 13, NDON ≤ 1, clogP ≤ 3.5): uncharged fragments able to bind deeply in the S1 pocket through hydrophobic interaction (e.g., substituted chlorophenyl and indole fragments). Although there were no neutral fragments reported that bind to thrombin, drug-sized inhibitors with neutral S1 binding groups have been described, suggesting that such fragments might be identified.[33–35]

(3) Neutral S1 binders with a solubilizing group (HAC ≤ 18, NDON ≤ 3, clogP ≤ 2.5): in order to test some uncharged scaffolds it was necessary to introduce a solubilizing group to make screening at high concentrations possible.

(4) S1–S1β and larger S1–S2 binders (HAC ≥ 16, NDON < 5, clogP < 5): a set of slightly larger compounds selected with the aim of simultaneously targeting the S1 pocket and either the S1β pocket or the S2 pocket of the active site.

Thrombin Crystal Structure Determinations. Crystallographic screening of the focused thrombin set was carried out on crystals of α-thrombin complexed with the peptidic inhibitor hirugen which renders the protein inactive by binding to the thrombin anion-binding exosite. However, hirugen binding does not obstruct the thrombin active site, thus providing a crystal system suitable for fragment soaking.[36] To maximize throughput and minimize subsequent deconvolution experiments, Pyramid screening of the targeted library was carried out using cocktails (mixtures) of two fragments. Even after overnight soaking in high concentrations of fragments, the thrombin–hirugen crystals generally diffracted to about 2.0 Å resolution, which allowed high-throughput X-ray data collection on in-house equipment using an automatic sample-changing robot ACTOR.[37]

Overview of Screening Results. Although fragments often have low affinity, they usually exhibit a high ligand efficiency, i.e., a high value for the ratio of free energy of binding to the number of heavy atoms.[38,39] This concept of ligand efficiency was one of the factors used to help prioritize the fragments selected for hits-to-leads chemistry. Overall, hits from the fragment screening were selected based upon the following criteria: moderate to good ligand efficiency, good druglike properties (such as neutral or weakly basic, nonpeptidic compounds), and potential for rapid optimization to potent leads based on a consideration of their protein–ligand binding modes.

The Pyramid screen identified several fragments bound to the S1 pocket of the enzyme. In each case, the crystal structure of the thrombin-bound fragment was solved and demonstrated a clearly defined binding mode by fitting to electron density in the active site. The hits included weakly basic fragments, such as aminopyridines and aminoquinolines, and also neutral fragments, such as chlorophenyl and indole-based compounds. For the purpose of this work, two chlorophenyl fragments, **1** (Figure 1a) and **2** (Figure 2a) were chosen for chemistry optimization. Crystal structures of thrombin containing uncharged chlorophenyl P1 motifs have (as noted earlier) been previously reported in the literature, but only in the context of larger, more potent, inhibitors.[33–35] The detection of small uncharged fragments, such as compound **1**, binding to the S1 pocket, without the requirement for additional hydrogen bonding in adjacent subsites, is a key finding. This provides validation of high-throughput X-ray crystallography's ability to detect fragments driven primarily by hydrophobic binding to active-site pockets.

Figure 1a shows the crystal structure of the chlorophenyl–tetrazole fragment **1**. The chlorophenyl group binds deeply in the S1 pocket with the chlorine atom located above the aromatic ring of the Tyr228 side chain. The tetrazole ring makes close contact (3.6 Å) with the sulfur atom of Cys220 in the S1β pocket, possibly forming a donor-atom–π interaction.[35] These type of interactions have been identified in other systems such as flavoenzymes.[40,41] The higher in vitro potency of compound **1** (IC$_{50}$ = 330 μM) made this compound the most potent, and also ligand efficient, of the uncharged S1 binders identified from the fragment screen.

Figure 2a shows the binding mode of compound **2**. The chlorophenyl group binds in the S1 pocket in an analogous fashion to **1** (Figure 1a). The amino group forms hydrogen bonds with the hydroxyl group of the catalytic Ser195, with the backbone carbonyl of Ser214, and also interacts via water mediation with Gly216. Although compound **2** was insufficiently soluble under the assay conditions for an IC$_{50}$ to be measured accurately (IC$_{50}$ > 1 mM), its very simple structure makes it a synthetically attractive starting point for hit-to-lead chemistry.

The Pyramid screening approach not only identified S1 binders but also identified a larger compound, **3**, which unexpectedly bound to the S2–S4 region, while leaving the S1 pocket vacant (Figure 1b). To our knowledge the crystal structure of compound **3** bound to thrombin provides the first reported structural characterization of a small molecule binding exclusively in the enzyme's S2 and S4 pockets. Compound **3** had been included in the screening set following the initial GOLD-based virtual screening which had placed the benzyl motif of this compound in the S1 pocket. The structure of compound **3** complexed with thrombin showed that the compound bound in a folded conformation where the naphthyl and the benzyl groups formed an intramolecular aromatic interaction while binding to the S4 and the S2 pockets, respectively.
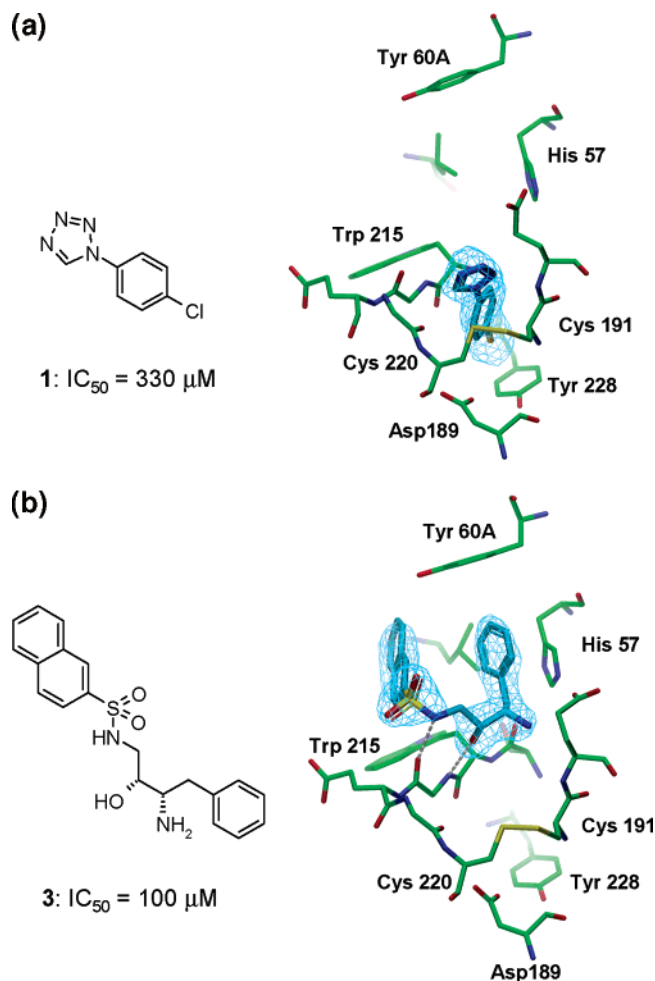
Figure 1. Fragment hits from thrombin Pyramid screening. (a) Crystal structure showing the binding of **1** in the thrombin S1 pocket at 2.3 Å resolution. **1** is shown in blue with the initial Fo–Fc electron density identifying the presence of the ligand superimposed in light blue and contoured at 3σ. Key protein residues are shown with carbon atoms in green, nitrogen in blue, oxygen in red, and sulfur in yellow. (b) Crystal structure showing the binding of **3** in the thrombin S2–S4 pockets at 2.2 Å resolution. Ligand and protein residues are shown using the same color scheme as in panel a, and the electron density is contoured at 3σ. Chemical structures of the fragment hits are shown on the left-hand side of the panels. All figures of protein–ligand structures were made with AstexViewer. IC$_{50}$ values are given as the average of two or more determinations.

Moreover, the hydroxyl group and the sulfonamide NH formed hydrogen bonds with Gly216. Similar hydrogen-bonding interactions to Gly216 are commonly observed with many peptidomimetics thrombin inhibitors, including melagatran and argatroban. The folded conformation adopted by **3** was analyzed using molecular dynamics calculations in explicit water at room temperature.[42] It was observed that after 55 ps of dynamics simulation, the structure of compound **3** changed from a high-energy open conformation (rmsd between the X-ray structure and the open conformation was 3.4 Å) to a low-energy folded conformation, similar to the one observed in the crystal structure (rmsd equal to 1.3 Å). These results suggested that S2–S4 binder **3** may adopt a prearranged, folded, conformation in solution which is similar to the observed bioactive form. This phenomenon, often termed "hydrophobic collapse", has been described for a number of important ligands, including thrombin inhibitors.[43,44] Due to the novel binding mode and the nonpeptidic nature of this molecule, compound **3** offered an attractive start-point for chemistry optimization.
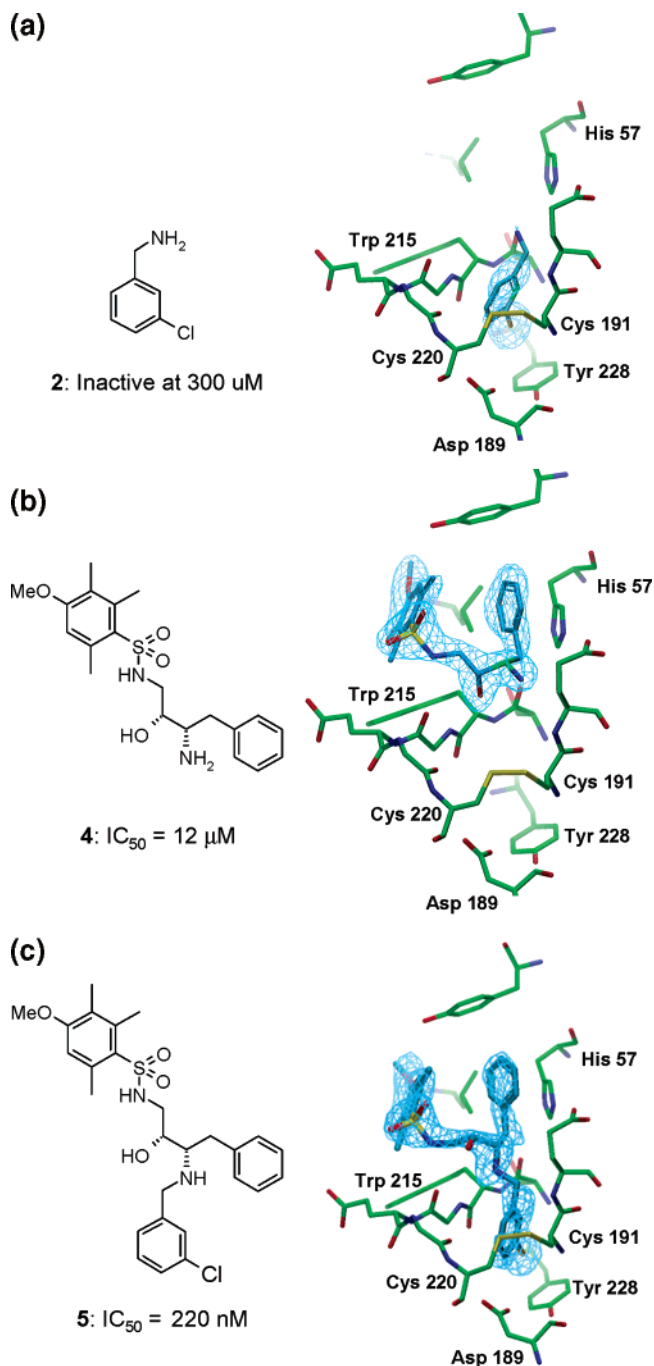
Figure 2. Fragment linking of thrombin S1 and S2−S4 binders. (a) Binding of a chlorophenyl fragment in the thrombin S1 pocket revealed at 2.2 Å resolution. (b) Crystal structure of the S2−S4 binder **4** at 2.2 Å resolution. (c) Crystal structure of **5**, designed by linking of **2** and **4**. In all panels the color scheme and electron density contour levels are the same as in Figure 1. IC$_{50}$ values are given as the average of two or more determinations.

Hit Optimization. The results from fragment screening against thrombin demonstrate the ability of this approach to probe different pockets within the active site. Furthermore, the high-quality structural information generated by the Pyramid screen can then be used to facilitate a rapid hit-to-lead optimization phase. We first looked at optimizing the novel S2−S4 binder, **3**. Analysis of the crystal structure of compound **3** suggested the naphthyl motif was suboptimal with regard to its steric complementarity with the S4 pocket. A small number of analogues aimed at modifying the sulfonamide motif were designed and synthesized. Of these, the methoxy-trimethyl-

benzenesulfonamide analogue **4** was found to be the most potent compound (IC$_{50}$ = 12 μM, Figure 2b). This substituent had previously been identified for another class of thrombin inhibitors.[45] The crystal structure of compound **4** bound to thrombin was solved and revealed a similar binding mode to S2−S4 binder **3** with the methoxy-trimethylbenzene group showing exceptional shape complementarity with the S4 pocket. In addition, compound **4** demonstrated good selectivity over other serine proteases such as trypsin and FXa, being essentially inactive at 1 mM.

Having identified different ligands which bind in adjacent pockets, a "fragment-linking" approach seemed to present an attractive strategy, particularly since the global protein structure did not change upon ligand binding. It has been well established that there should be a "superadditivity" of potency when two fragments are linked in an optimal fashion.[46,47] Achieving this in practice is often very difficult, but there are a number of cases reported in which two fragments have been successfully linked together to give highly potent ligands. For example, NMR-directed fragment linking has been used to identify potent ligands for a number of targets, including the FK506-binding protein, stromelysin, and protein tyrosine phosphatase 1B (PTP1B).[13,14,26]

To define the best fragment-linking strategy, thrombin structures which contained a fragment in the S1 pocket were overlaid with the structure of the optimized S2−S4 binder **4** in the same frame of reference. The overlay allowed direct comparison between the different binding modes and the possible attachment points between fragments. Compound **4** contains a hydroxyl group and a primary amine which are directed toward the S1 pocket and therefore provided attractive points through which S2−S4 and S1 fragments could be linked. Guided by structure-based design and consideration of synthetic tractability, the S2−S4 binder **4** was linked through the basic nitrogen to the meta position of the chlorophenyl ring of the S1 fragment **2** via an aminomethyl linker. The hybrid compound **5** (IC$_{50}$ = 220 nM) demonstrated a 50-fold improvement in affinity over compound **4** and was at least 1000-fold selective over trypsin (**5**: IC$_{50}$ trypsin > 100 μM). The fragment-linking strategy is summarized in Figure 2a−c which illustrates the crystal structures of compounds **2**, **4**, and **5** respectively.

The crystal structure of compound **5** bound to thrombin shows the benzyl group and the phenyl sulfonamide occupying the S2 and S4 pockets, respectively, as observed in compound **4**, with the sulfonamide NH hydrogen-bonded to the carbonyl of Gly216. The secondary amine points toward Gly216 in order to allow the 3-chlorophenyl to bind to the S1 pocket, as seen previously for S1 fragment **2**. Finally, the hydroxyl group is directed away from Gly216, and the hydrogen bond to the protein, as observed in compound **4**, is absent.

Although a 50-fold increase in potency of the hybrid compound **5** over the S2−S4 binder **4** was achieved, the improvement was not of the magnitude expected for optimally linked fragments.[46] The conformational changes observed for compound **5** with the concomitant loss of one hydrogen bond may go some way to explain this. To investigate further optimization, we turned our attention to fragment linking with the more potent S1 fragment **1** (IC$_{50}$ = 330 μM). Two hybrid compounds were designed and synthesized, one analogue **6** with the fully substituted P4 motif (methoxy-trimethylbenzene), and a second analogue **7** with a less substituted P4 motif (*p*-methoxybenzene) aimed at minimizing molecular weight, lipophilicity, and molecular complexity (Chart 2). Compound **6** (IC$_{50}$ = 1.4 nM) was the most potent analogue identified, demonstrat-

Chart 2[a]

[a]**6**: IC$_{50}$ = 1.4 nM

[a]**7**: IC$_{50}$ = 3.7 nM
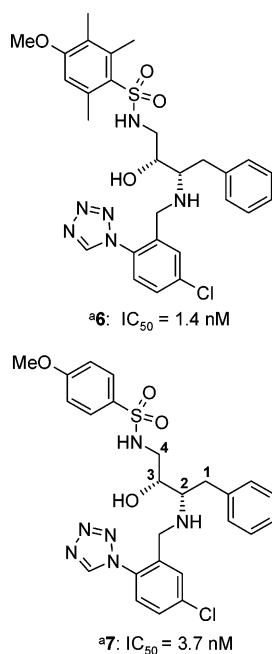
[a] Average of two or more determinations.
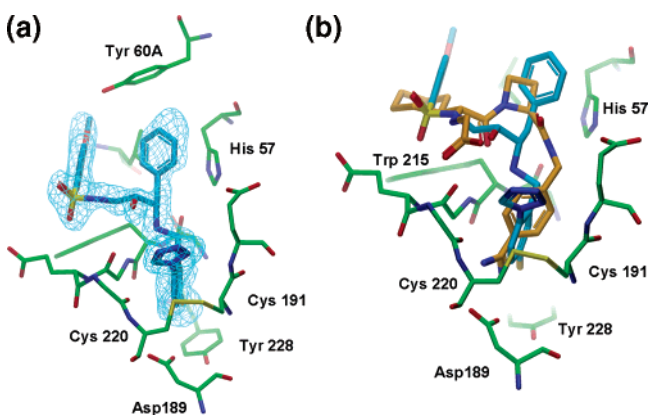
Figure 3. (a) Binding of **7**. Crystal structure showing the binding of **7** at 2.0 Å resolution. In all panels the color scheme and electron density contour levels are the same as in Figure 1. (b) Binding mode of **7** overlaid with melagatran.

ing a 200-fold improvement in affinity over compound **5**, and the simplified analogue **7** (IC$_{50}$ = 3.7 nM) was also only marginally less potent. The crystal structure of the complex between thrombin and analogue **7** (Figure 3a) showed that the relative binding modes of the P1, P2, and P4 functionalities were analogous to those observed in both the initial fragments and in compound **5**. However, some notable differences were again observed, particularly regarding the conformation of the ethanolamine linker. Rotation around the C1–C2 and C2–C3 bonds of the ethanolamine chain in compound **7** causes both the hydroxyl group and the sulfonamide NH to point away from the Gly216 backbone residue, thereby preventing hydrogen bonding to Gly216. The binding of compound **7** to thrombin appears to be driven purely by hydrophobic interactions with no obvious contribution from hydrogen bonding.

The binding mode of compound **7** is compared with that of melagatran in Figure 3b. An overlay of the two structures reveals some similarities with respect to the key P1, P2, and P4 motifs of the two compounds. Both compounds present hydrophobic groups which occupy the S2 and S4 pockets, although, in the case of melagatran, the P2 motif is relatively small. Furthermore,

both compounds present aromatic motifs which bind to the S1 pocket, although these groups are chemically quite different. In the case of compound **7**, P1 is a nonbasic chlorophenyl group, whereas melagatran presents a basic benzamidine motif. Although they occupy the same active-site region, the two compounds are based on very different scaffolds. Melagatran is a pseudo-peptide, whereas compound **7** is nonpeptidic.

In this example, fragment linking has produced the potent inhibitor **6** (IC$_{50}$ = 1.4 nM) which can be attributed to the binding affinity of the initial fragments being combined in an additive manner. However, the "superadditivity" which could theoretically be expected from optimal linked fragments was not observed. This lack of optimal ligand efficiency can be ascribed to the fact that the interactions of the starting hits were not completely retained in the fragment-linked compounds. This result fits with the assertion that linking two fragments in an optimal fashion is often very difficult in practice and, furthermore, that large molecular weight inhibitors are often a consequence of such an approach.

An alternative strategy to fragment linking is to start from a fragment with high ligand efficiency and to use structure-based drug design to form additional interactions by growing out from the fragment. It has been demonstrated elsewhere that the fragment-growing approach can maintain ligand efficiency and thus identify attractive lead molecules.[39] Although both approaches can deliver potent inhibitors, a fragment-growing approach is likely to deliver compounds which are better starting points for lead optimization.

Conclusions

The fragment-screening and linking strategy has facilitated the rapid identification of novel nanomolar inhibitors of thrombin which are nonpeptidic and lack a highly basic guanidine/amidine mimetic. Only 80 targeted fragments were screened, and less than 40 compounds were synthesized. Fragment hits with novel binding modes, e.g., compound **3**, were discovered which demonstrates the potential of fragment-based screening to probe beyond just the key S1 specificity pocket in thrombin.

It is recognized that compound **7** is not ideal with respect to Lipinski's molecular weight guidelines for an oral drug. Nonetheless, this compound presents a useful starting point for further optimization strategies. Optimization of this series and a comparison with potent lead compounds identified from a fragment-growing approach will be the topic of a future publication.

Experimental Section

Crystallography. Freeze-dried human prothrombin and the hirugen peptide were purchased from Enzyme Research Laboratories, South Bend, IN and Bachem, Bubendorf Switzerland, respectively. Human α-thrombin was activated and purified using a modified method described by Ngai and Chang.[48] Crystals of the α-thrombin–hirugen complex were obtained by a procedure based on the method of Skrzypczak-Jankun et al.[36] In brief, human α-thrombin and the hirugen peptide were complexed in a 1:3 molar ratio in 20 mM Tris/HCl pH 7.5 and 250 mM NaCl, and the thrombin–hirugen complex was concentrated to a final concentration of 10 mg/mL. Initial thrombin–hirugen crystals were grown by vapor diffusion using drops composed of a 1:1 mixture of protein and reservoir solution, equilibrated over reservoirs containing 28–33% PEG4000, 0.1 M phosphate buffer pH 7.3 and 0–0.3 M NaCl. Diffraction quality crystals could be obtained by microseeding, whereby drops composed of a 1:1 mixture of protein solution and thrombin–hirugen seed crystals in 30% PEG4000 and 0.1 M phosphate buffer pH 7.3 were equilibrated against reservoirs of 23%

Table 1. Crystallographic Data Collection and Refinement Statistics

compound	1	2	3	4	5	7
			Data Collection			
X-ray source			in house, $\lambda = 1.54$ Å			
resolution (Å)	2.3	2.1	2.2	2.2	2.0	2.0
no. unique reflections	15571	17595	18049	17066	17227	23682
completeness (%)^a	97.3 (96.3)	95.4(86.3)	99.7 (99.1)	97.4(92.7)	93.6(87.3)	97.9 (90.0)
ave multiplicity	1.96	1.96	2.91	1.97	2.07	3.3
Rmerge^{a,b}	0.136(0.38)	0.081(0.26)	0.133(0.44)	0.061(0.16)	0.044(0.11)	0.060(0.15)
			Refinement			
Rcryst	0.220	0.176	0.185	0.164	0.153	0.164
Rfree	0.306	0.246	0.259	0.242	0.229	0.219
rmsd bond lengths (Å)	0.011	0.011	0.012	0.011	0.012	0.011
rmsd bond angles (deg)	1.4	1.3	1.4	1.4	1.4	1.3
ave B-factor protein (Å^2)	39.2	34.7	33.5	21.4	23.9	20.8
ave B-factor ligand (Å^2)	35.3	36.4	32.2	15.3	25.8	34.1
ave B-factor solvent (Å^2)	40.2	43.8	40.0	29.5	33.7	19.6

^{a} Numbers in parentheses indicate the highest shell values. ^{b} Rmerge $= \sum_h\sum_i|I(h,i) - \langle I\rangle(h)|/\sum_h\sum_i\langle I\rangle(h)$; $I(h,i)$ is the scaled intensity of the *i*th observation of reflection *h*, and (*h*) is the mean value. Summation is over all measurements. Rcryst $= \sum_{hkl,\text{work}}||F_{obs}| - k|F_{calc}||/\sum_{hkl}|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, *k* is a weighting factor, and work denotes the working set of 95% of the reflections used in the refinement. Rfree $= \sum_{hkl,\text{test}}||F_{obs}| - k|F_{calc}||/\sum_{hkl}|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, *k* is a weighting factor, and test denotes the test set of 5% of the reflections used in cross validation of the refinement. λ refers to wavelength, rmsd to root-mean-square deviations. All ligands have been refined with occupancies of 1.0.

PEG4000, 0.1 M phosphate buffer pH 7.3 and 0−0.3 M NaCl at room temperature. Data collection and refinement statistics for selected crystal structures are presented in Table 1.

All soaking experiments were carried out overnight in a standard mother liquor consisting of 30% PEG4000 and 0.1 M phosphate buffer pH 7.3. Data were collected using our in-house equipment on either a Rigaku-MSC Jupiter CCD or an R-AXIS IV image plate detector mounted on an RU-H3R rotating anode generator equipped with Osmic confocal multilayer optics. All data sets were processed with d*trek[49] and scaled using SCALA.[50] Structure solution and initial refinement of the protein−ligand complexes were carried out with our automated scripts using a modified thrombin−hirugen structure (PDB id code 1QJ1) as a starting model. Bound ligands were identified and fitted in the Fo−Fc electron density using AutoSolve and further refined using our automated scripts followed by rounds of rebuilding and refinement using Refmac.[50]

Virtual Screening. In thrombin, the side chain conformation of Glu192 can vary significantly changing the size and the shape of the pockets S1 and S1β.[51] To take these conformational changes into account, we decided to virtually screen compounds against two structures: PDB entry 1ETS and a modified version of 1ETS where the Glu192 side chain was moved away from the S1β pocket. For each protein structure two binding sites were constructed: a large binding site constituted by all protein atoms within 6 Å of an atom in the 1ETS ligand and a small binding site constituted by all protein atoms within 6 Å of an atom in the benzamidine substructure contained in the 1ETS ligand. All the virtual screens were run on a Linux cluster using the Astex web-based virtual screening platform[29] using methods and settings previously described by Verdonk et al.[52] A number of filters (heavy atoms count, MW, number of donors, number of acceptors, ClogP, and PSA) combined with the scoring functions Goldscore and Chemscore were applied to score and rank the docked libraries. Finally, virtual screens were visually analyzed, and fragments were selected based on the binding mode, the chemical tractability, and the score.

The S1 hydrophobic pharmacophore was designed using the Factor Xa structure 1MQ5, where the ligand contains a chlorophenyl ring buried deeply in the S1 pocket. The S1 pocket residues of the Factor Xa structure 1MQ5 were overlaid with the respective S1 pocket residues in the thrombin structure 1ETS. The coordinates of chlorine atom in 1MQ5 were chosen to define the center of the pharmacophore. Subsequently, dockings experiments of chlorobenzene fragments against the thrombin structures 1ETS were used to optimize the pharmacophore radius to 2.5 Å and the pharmacophore weight to 10.
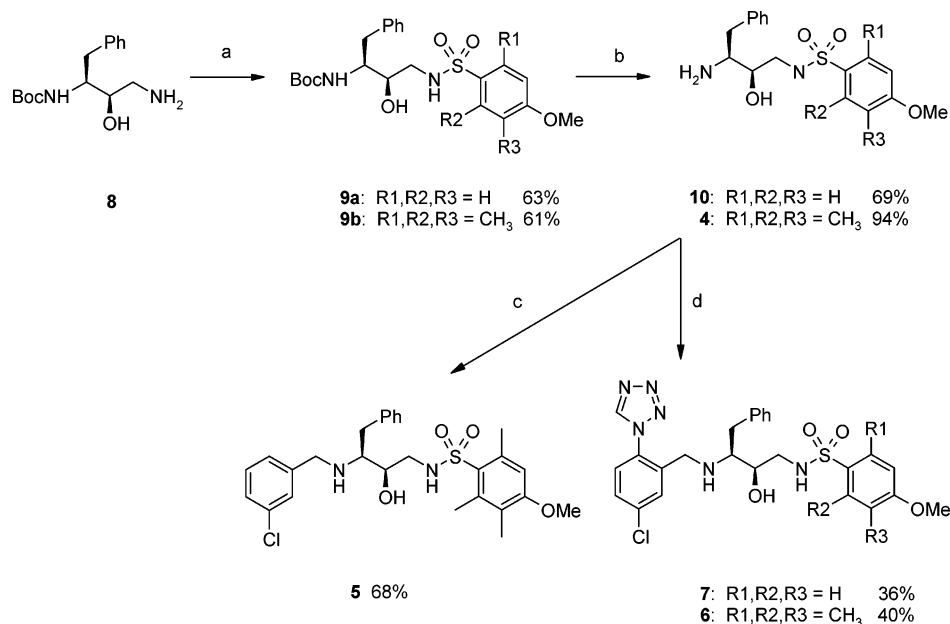
Molecular Dynamics. The molecular dynamics simulation on **3** was performed with cubic periodic boundary condition (NVT ensemble) using the program Discover. The system consisted of **3** solvated with 236 water molecules placed in a cubic box (with dimensions 20 Å × 20 Å × 20 Å). All atoms of the system were considered explicitly, and their interaction were computed using CFF91 force field[53] using a dielectric constant equal to 1. First a 500 iteration potential energy minimization was carried out with harmonic restraint on **3**, followed by 3ps molecular dynamics at 298 K. A final dynamics production run of 500 ps (2 fs step) was carried out at 298 K with no restraint on the ligand. Snapshots from the trajectories were saved every 0.5 ps.

Chemistry. The compounds **5**, **6**, and **7** were synthesized from the amino-alcohol **8** (Scheme 1), which can be prepared in five steps from L-phenylalanine.[54] Sulfonamides **9a** and **9b** were prepared by treatment of the amino-alcohol **8** with the corresponding sulfonyl chloride (Scheme 1). Removal of the Boc-protecting group, followed by reductive amination with the relevant benzaldehyde, gave the final products **5**, **6**, and **7**. 5-Chloro-2-tetrazol-1-yl-benzaldehyde **14** was synthesized from 2-amino-5-chlorobenzoic acid **11** in three steps (Scheme 2). Tetrazole **12** was prepared from amine **11** using a modification of the procedure described by Young et al.[35] Subsequent PyBOP-mediated^a coupling with *N,O*-dimethylhydroxylamine hydrochloride gave the Weinreb amide **13** which, upon treatment with lithium aluminum hydride, gave the aldehyde **14**.

All reactions were carried out under a N$_2$ atmosphere with commercial grade reagents from Sigma-Aldrich-Fluka, Avocado, Breckland, or Novabiochem. Solvents were distilled prior to use or commercially available anhydrous solvents used (Aldrich). The ^{1}H and ^{13}C NMR spectra were recorded on a Bruker DPX400N spectrometer. Chemical shifts are reported in ppm relative to residual solvent peak. Flash silica chromatography was performed using Breckland silica gel 60 (230−400 mesh). Analytical LC−MS was performed using either method A (Waters 2795 separation module/Waters micromass ZQ detector with a Waters Atlantis dC$_{18}$ 4.6 mm × 30 mm, 3 μm column and a mobile phase of acetonitrile/10 mM aq ammonium acetate with 0.1% formic acid) or method B (Waters 2795 separation module/Waters 2996 PDA detector with a Phenomenex Synergi 4 μm MAX-RP 80A, 2.0 mm × 50 mm column and a mobile phase of acetonitrile/water with 0.1% formic acid). Purification by HPLC was carried out using a Waters Fractionlynx/Waters ZQ LC/MS platform with a Phenomenex Synergy MAX-RP, 10 μm, 150 mm × 15 mm column and a mobile phase of acetonitrile/water with 0.1% formic acid.

---

^{a} Abbreviations: DMF, *N,N*-dimethylformamide; EtOAc, ethyl acetate; PyBOP, (benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate.

Scheme 1. Synthesis of Inhibitors **5**, **6**, and **7**[a]

8

9a: R1,R2,R3 = H 63%
9b: R1,R2,R3 = CH$_3$ 61%

10: R1,R2,R3 = H 69%
4: R1,R2,R3 = CH$_3$ 94%

5 68%

7: R1,R2,R3 = H 36%
6: R1,R2,R3 = CH$_3$ 40%

^{a} Reagents: (a) Mtr-Cl or 4-MeOC$_6$H$_6$SO$_2$Cl, ^{i}Pr$_2$EtN, DMF; (b) TFA, CH$_2$Cl$_2$; (c) 3-chlorobenzaldehyde, NaCNBH$_3$, MeOH, AcOH; (d) **14**, NaCNBH$_3$, MeOH, AcOH.

Scheme 2. Synthesis of Aldehyde **14**[a]

11   **12**   **13**   **14**
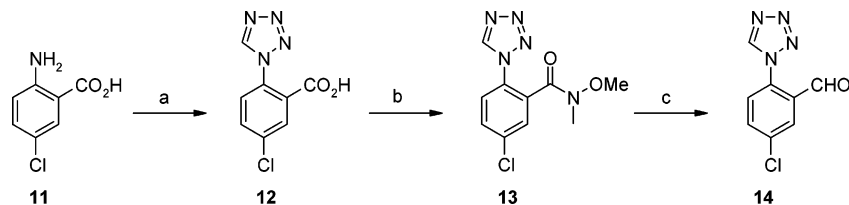
^{a} Reagents: (a) CH(OMe)$_3$, NaN$_3$, AcOH, 88%; (b) MeNHOMe·HCl, PyBOP, ^{i}Pr$_2$EtN, CH$_2$Cl$_2$, 83%; (c) LiAlH$_4$, THF, −78 °C, 94%.

(1*S*,2*R*)-[1-Benzyl-2-hydroxy-3-(4-methoxy-benzenesulfon-ylamino)-propyl]-carbamic Acid *tert*-Butyl Ester (9a). A solution of 4-methoxy-benzenesulfonyl chloride (427 mg, 2.07 mmol) in DMF (5 mL) was added dropwise to a solution of (1*S*,2*R*)-(3-amino-1-benzyl-2-hydroxy-propyl)-carbamic acid *tert*-butyl ester **8** (527 mg, 1.88 mmol) and *N,N*-diisopropylethylamine (655 μL, 3.76 mmol) in DMF (15 mL) over the period of 10 min. The reaction was stirred for 18 h. The mixture was concentrated in vacuo, the residue was dissolved in EtOAc (25 mL) and washed with water (25 mL), 10% aq KHSO$_4$ (25 mL), saturated aq NaHCO$_3$ (25 mL), and saturated brine (25 mL). The organic phase was dried (Na$_2$-SO$_4$), filtered, and concentrated in vacuo. The residue was purified by silica chromatography with hexane/EtOAc (1.25:1) to give the title compound **9a**, a white solid, 531 mg, 63%. ^{1}H NMR (CDCl$_3$, 400 MHz): δ 7.78 (m, 2H), 7.16−7.34 (m, 5H), 6.94 (m, 2H), 5.74 (br, 1H), 4.42 (d, *J* = 8.2 Hz, 1H), 3.84 (s, 3H), 3.77 (dt, *J* = 4.6, 8.2 Hz, 1H), 3.45 (br, 1H), 3.15 (ddd, *J* = 3.1, 9.1, 12.7 Hz, 1H), 3.01 (dd, *J* = 4.6, 14.2 Hz, 1H), 2.85 (m, 2H), 1.29 (s, 9H) ppm. ^{13}C NMR (CDCl$_3$, 100 MHz): δ 163.3, 157.2, 137.4, 131.7, 129.8, 129.6, 129.2, 127.1, 114.7, 80.7, 71.8, 56.0, 53.8, 45.8, 36.7, 28.5 ppm. LC−MS (method A) *m/z* (ES$^+$) 351 [M − Boc + H], R_t = 4.12 min.

(1*S*,2*R*)-[1-Benzyl-2-hydroxy-3-(4-methoxy-2,3,6-trimethyl-benzenesulfonylamino)-propyl]-carbamic Acid *tert*-Butyl Ester (9b). Prepared in an analogous fashion to **9a** using **8** (89 mg, 303 μmol) and 4-methoxy-2,3,6-trimethyl-benzenesulfonyl chloride to give the title compound **9b**, a white solid, 95 mg, 61%. ^{1}H NMR (CDCl$_3$, 400 MHz): δ 7.18−7.34 (m, 5H), 6.58 (s, 1H), 5.96 (dd, *J* = 3.9, 7.8 Hz, 1H), 4.44 (d, *J* = 8.4 Hz, 1H), 3.85 (s, 3H), 3.82 (m, 1H), 3.42 (br, 2H), 3.05 (m, 2H), 2.87 (m, 2H), 2.68 (s, 3H), 2.60 (s, 3H), 2.14 (s, 3H), 1.33 (s, 9H) ppm. ^{13}C NMR (CDCl$_3$, 100 MHz): δ 158.1, 155.5, 137.4, 135.9, 133.0, 128.2, 127.8, 127.4, 125.5, 124.1, 110.9, 79.0, 70.3, 54.3, 52.0, 43.3, 35.1, 27.0, 23.0,

16.5, 10.8 ppm. LC−MS (method B) *m/z* (ES$^+$) 493 [M + H], (ES$^−$) 491 [M − H], R_t = 3.45 min.

(2*R*,3*S*)-*N*-(3-Amino-2-hydroxy-4-phenyl-butyl)-4-methoxy-benzenesulfonamide (10). (1*S*,2*R*)-[Benzyl-2-hydroxy-3-(4-meth-oxy-benzenesulfonylamino)-propyl]-carbamic acid *tert*-butyl ester **9a** (347 mg, 770 μmol) was dissolved in (1:1) trifluoroacetic acid/CH$_2$Cl$_2$ (10 mL), and stirred for 2 h, then concentrated in vacuo. The residue was partitioned between EtOAc (20 mL) and (1:1) saturated aq NaCO$_3$/water (20 mL). The organic phase was dried (Na$_2$SO$_4$), filtered, and concentrated in vacuo. The residue was purified by silica chromatography with methanol/CH$_2$Cl$_2$ (1:9) to give the title compound **10**, a white foam solid, 187 mg, 69%. ^{1}H NMR (CD$_3$OD, 400 MHz): δ 7.80 (m, 2H), 7.29 (m, 2H), 7.22 (m, 3H), 7.06 (m, 2H), 3.86 (s, 3H), 3.60 (td, *J* = 4.7, 6.8 Hz, 1H), 3.15 (td, *J* = 4.7, 9.5 Hz, 1H), 3.06 (dd, *J* = 4.7, 13.3 Hz, 1H), 2.98 (dd, *J* = 4.7, 13.8 Hz, 1H), 2.92 (dd, *J* = 6.8, 13.3 Hz, 1H), 2.52 (dd, *J* = 9.5, 13.8 Hz, 1H) ppm. ^{13}C NMR (CD$_3$OD, 100 MHz): δ 164.4, 139.7, 133.0, 130.4, 130.2, 129.7, 127.6, 115.4, 73.3, 56.3, 56.2, 46.5, 38.4 ppm. LC−MS (method A) *m/z* (ES$^+$) 351 [M + H], (ES$^−$) 349 [M − H], R_t = 2.96 min.

(2*R*,3*S*)-*N*-(3-Amino-2-hydroxy-4-phenyl-butyl)-4-methoxy-2,3,6-trimethyl-benzenesulfonamide (4). Prepared in an analogous fashion to **10** but using **9b** (840 mg, 1.71 mmol) as starting material to give the title compound **4**, an off-white solid, 630 mg, 94%. ^{1}H NMR (CDCl$_3$, 400 MHz): δ 7.25 (m, 2H), 7.15 (m, 3H), 6.80 (s, 1H), 4.82 (br, 1H), 3.83 (s, 3H), 3.29 (dd, *J* = 5.4, 10.7 Hz, 1H), 2.96 (dd, *J* = 4.7, 12.8 Hz, 1H), 2.76 (m, 3H), 2.60 (s, 3H), 2.50 (s, 3H), 2.30 (m, 1H), 2.09 (s, 3H) ppm. ^{13}C NMR (CDCl$_3$, 100 MHz): δ 158.6, 139.9, 138.3, 138.0, 129.6, 129.4, 128.2, 125.9, 124.2, 112.4, 72.4, 55.7 (×2), 45.4, 39.0, 23.9, 17.7, 12.0 ppm. LC−MS (method A) *m/z* (ES$^+$) 393 [M + H], (ES$^−$) 391 [M − H], R_t = 3.06 min.

(2R,3S)-N-[3-(3-Chloro-benzylamino)-2-hydroxy-4-phenyl-butyl)-4-methoxy-2,3,6-trimethyl-benzenesulfonamide (5). (2R,3S)-N-(3-Amino-2-hydroxy-4-phenyl-butyl)-4-methoxy-2,3,6-trimethyl-benzenesulfonamide **4** (158 mg, 402 μmol) and 3-chlorobenzaldehyde (46 μL, 406 μmol) were dissolved in methanol (10 mL) and stirred for 30 min. Acetic acid (75 μL, 1.31 mmol) and sodium cyanoborohydride (51 mg, 812 μmol) were added, and the reaction mixture was stirred for 18 h. The mixture was concentrated in vacuo, and the residue was partitioned between EtOAc (25 mL) and 10% aq K_2CO_3 (25 mL). The organic phase was dried (Na_2-SO_4), filtered, and concentrated in vacuo. The residue was purified by silica chromatography with hexane/EtOAc (1:1) to give the title compound **5**, a white foam solid, 141 mg, 68%. ^{1}H NMR (CDCl$_3$, 400 MHz): δ 7.11−7.29 (m, 5H), 7.06 (m, 2H), 6.98 (t, J = 1.6 Hz, 1H), 6.89 (td, J = 1.6, 6.8 Hz, 1H), 6.58 (s, 1H), 3.83 (s, 3H), 3.66 (td, J = 4.4, 7.7 Hz, 1H), 3.58 (d, J = 13.5 Hz, 1H), 3.52 (d, J = 13.5 Hz, 1H), 3.11 (dd, J = 4.4, 12.9 Hz, 1H), 2.97 (dd, J = 7.7, 12.9 Hz, 1H), 2.86 (td, J = 4.4, 9.3 Hz, 1H), 2.73 (dd, J = 4.4, 13.9 Hz, 1H), 2.67 (s, 3H), 2.62 (m, 1H), 2.58 (s, 3H), 2.14 (s, 3H) ppm. ^{13}C NMR (CDCl$_3$, 100 MHz): δ 159.8, 142.1, 139.2, 138.2, 134.6, 130.1, 129.5, 129.1, 129.0, 128.5, 127.7, 127.1, 126.6, 125.7, 112.5, 69.9, 60.8, 55.9, 51.4, 45.5, 35.9, 24.7, 18.3, 12.4 ppm. LC−MS (method A) m/z (ES$^+$) 517, 519 [M + H], (ES$^-$) 515, 517 [M − H], R_t = 3.53 min.

(2R,3S)-N-[3-(5-Chloro-2-tetrazol-1-yl-benzylamino)-2-hydroxy-4-phenyl-butyl)-4-methoxy-2,3,6-trimethyl-benzene-sulfonamide (6). Prepared in an analogous fashion to **5** but using **4** (451 mg, 1.15 mmol) and 5-chloro-2-tetrazol-1-yl-benzaldehyde **14** (240 mg, 1.15 mmol) in methanol (35 mL) in the presence of activated 3 Å molecular sieves. The residue was applied to a silica chromatography column with methanol/CH$_2$Cl$_2$ (3:97) to give the crude product (355 mg). A 100 mg sample was further purified by preparative HPLC to give the title compound **6**, a white solid, 73 mg, 40%. ^{1}H NMR (CD$_3$OD, 400 MHz): δ 9.26 (s, 1H), 7.51 (m, 2H), 7.44 (m, 1H), 7.10−7.22 (m, 3H), 7.05 (m, 2H), 6.69 (s, 1H), 3.81 (s, 3H), 3.73 (m, 1H), 3.54 (d, J = 13.9 Hz, 1H), 3.44 (d, J = 13.9 Hz, 1H), 2.97 (m, 2H), 2.82 (m, 2H), 2.60 (s, 3H), 2.57 (m, 1H), 2.52 (s, 3H), 2.09 (s. 3H) ppm. ^{13}C NMR (CD$_3$OD, 100 MHz): δ 159.7, 144.5, 139.0, 138.8, 138.6, 136.5, 135.4, 131.9, 131.5, 129.2, 129.0, 128.6, 127.4, 126.6, 125.2, 112.2, 69.7, 61.3, 55.0, 46.5, 44.8, 34.7, 23.5, 17.2, 11.1 ppm. LC−MS (method A) m/z (ES$^+$) 585, 587 [M + H], (ES$^-$) 583, 585 [M − H], R_t = 4.51 min.

(2R,3S)-N-{3-[5-Chloro-2-tetrazol-1-yl-benzylamino]-2-hydroxy-4-phenyl-butyl}-4-methoxy-benzenesulfonamide (7). Prepared in an analogous fashion to **5** but using (2R,3S)-N-(3-amino-2-hydroxy-4-phenyl-butyl)-4-methoxy-benzenesulfonamide **10** (264 mg, 753 μmol) and 5-chloro-2-tetrazol-1-yl-benzaldehyde **14** (157 mg, 753 μmol). The residue was applied to a silica chromatography column with methanol/CH$_2$Cl$_2$ (3:97) to give the crude product **7** (294 mg). A 100 mg sample was further purified by preparative HPLC to give a white solid, 50 mg, 36%. ^{1}H NMR (CD$_3$OD, 400 MHz): δ 9.22 (s, 1H), 7.77 (m, 2H), 7.47 (m, 2H), 7.40 (m, 2H), 7.20 (m, 3H), 7.15 (m, 1H), 7.08 (m, 2H), 7.03 (m, 2H), 3.85 (s, 3H), 3.62 (td, J = 4.8, 7.4 Hz, 1H), 3.44 (d, J = 13.9 Hz, 1H), 3.37 d, J = 13.9 Hz, 1H), 3.03 (dd, J = 4.8, 13.2 Hz, 1H), 2.80 (m, 3H), 2.50 (dd, J = 8.6, 13.3 Hz, 1H) ppm. ^{13}C NMR (CD$_3$OD, 100 MHz): δ 164.4, 145.7, 140.5, 138.8, 137.5, 133.0, 132.8, 130.4, 130.2, 129.6, 129.5, 128.7, 127.4, 115.4, 71.8, 62.1, 56.2, 47.8, 46.8, 36.9 ppm. LC−MS (method B) m/z (ES$^+$) 543, 545 [M + H], (ES$^-$) 541, 543 [M − H], R_t = 2.33 min.

5-Chloro-2-tetrazol-1-yl-benzoic Acid (12). 2-Amino-5-chlorobenzoic acid **11** (10.90 g, 63.53 mmol) and sodium azide (12.38 g, 190.43 mmol) were suspended in trimethyl orthoformate (21.0 mL, 191.95 mmol) and cooled to 0 °C. Glacial acetic acid (220 mL) was added, and the mixture was stirred at 0 °C for 3 h then at ambient temperature for 16 h. The slurry was concentrated in vacuo, and the residue was partitioned between EtOAc (750 mL) and 3 N HCl (500 mL). The organic phase was dried (Na$_2$SO$_4$), filtered, and concentrated in vacuo to give the title compound **12**, a light yellow solid, 12.57 g, 88%. ^{1}H NMR (DMSO-d_6, 400 MHz): δ

13.75 (s, 1H), 9.81 (s, 1H), 8.05 (d, J = 2.5 Hz, 1H), 7.92 (dd, J = 2.5, 8.5 Hz, 1H), 7.77 (d, J = 8.5 Hz, 1H) ppm. ^{13}C NMR (DMSO-d_6, 100 MHz): δ 164.6, 145.2, 135.9, 133.2, 131.6, 131.0, 130.1, 129.9 ppm. LC−MS (method A) m/z (ES$^+$) 225, 227 [M + H], R_t = 3.70 min.

5-Chloro-N-methoxy-N-methyl-2-tetrazol-1-yl-benzamide (13). 5-Chloro-2-tetrazol-1-yl-benzoic acid **12** (1.80 g, 8.01 mmol), N,O-dimethylhydroxylamine hydrochloride (781 mg, 8.01 mmol), and PyBOP (4.17 g, 8.01 mmol) were suspended in CH$_2$Cl$_2$ (150 mL). N,N-Diisopropylethylamine (4.2 mL, 24.03 mmol) was added, and the resultant solution was stirred for 36 h. The solution was concentrated in vacuo to ~15−20 mL, diluted with EtOAc (200 mL), and washed with water (200 mL), 10% aq KHSO$_4$ (200 mL), saturated aq NaHCO$_3$ (200 mL), and saturated brine (200 mL). The organic phase was dried (Na$_2$SO$_4$), filtered, and concentrated in vacuo. The residue was purified by silica chromatography (EtOAc) to give the title compound **13**, a white solid, 1.78 g, 83%. ^{1}H NMR (CD$_3$OD, 400 MHz): δ 9.53 (s, 1H), 7.76 (d, J = 1.9 Hz, 1H), 7.73 (m, 2H), 3.45 (s, 3H), 3.18 (s, 3H) ppm. ^{13}C NMR (CD$_3$OD, 100 MHz): δ 145.1, 133.9, 133.0, 132.7, 131.4, 130.6, 127.6, 62.6, 33.1 ppm. LC−MS (method A) m/z (ES$^+$) 268, 270, R_t = 3.15 min.

5-Chloro-2-tetrazol-1-yl-benzaldehyde (14). A solution of 5-chloro-N-methoxy-N-methyl-2-tetrazol-1-yl-benzamide **13** (1.72 g, 6.34 mmol) in THF (20 mL) was added dropwise to a solution of lithium aluminum hydride (1 M in THF) (12.9 mL, 12.9 mmol) at −78 °C over the period of 30 min. The mixture was stirred for 1 h, then cold water (5 mL) added dropwise to quench the reaction. The resultant mixture was diluted with EtOAc (250 mL) and washed with 1 N HCl (240 mL) and saturated brine (240 mL). The organic phase was dried (Na$_2$SO$_4$), filtered, and concentrated in vacuo to give the title compound **14**, a yellow solid, 1.26 g, 94%. ^{1}H NMR (CDCl$_3$, 400 MHz): δ 9.85 (s, 1H), 8.98 (s, 1H), 8.07 (d, J = 2.4 Hz, 1H), 7.79 (dd, J = 2.4, 8.5 Hz, 1H), 7.52 (d, J = 8.5 Hz, 1H) ppm. ^{13}C NMR (CDCl$_3$, 100 MHz): δ 186.9, 144.0, 138.4, 135.3, 132.4, 132.0, 131.6, 128.3 ppm. LC−MS (method A) m/z (ES$^+$) 209 [M + H], R_t = 2.96 min.

Thrombin Assay. Compounds were incubated with activated thrombin and 14 μM fluorogenic substrate, BOC-Val-Pro-Arg-MCA (Bachem), in 50 mM Tris pH 8, 5 mM CaCl$_2$, 100 mM NaCl, 5% DMSO in 96-well plates. The cleavage of the substrate was followed by monitoring the change in fluorescence at 460 nm (excitation at 365 nm) for 25 min at room temperature on a Packard Fusion plate reader. Initial reaction rates were measured, and IC$_{50}$s were calculated from replicate curves using GraphPad Prizm software, and standard errors and P values were within statistically acceptable limits. Assay standards were provided by benzamidine and proflavin, both of which are well described in the thrombin literature.[55,56] Benzamidine and proflavin were found to have IC$_{50}$ values of 440 and 12 μM, respectively.

References

(1) Bode, W. The structure of thrombin, a chameleon-like proteinase. *J. Thromb. Haemostasis* [Online] **2005**, Epub ahead of print, 09/05/2005.

(2) Stubbs, M. T.; Bode, W. The clot thickens: clues provided by thrombin structure. *Trends Biochem. Sci.* **1995**, *20*, 23−28.

(3) Riester, D.; Wirsching, F.; Salinas, G.; Keller, M.; Gebinoga, M.; Kamphausen, S.; Merkwirth, C.; Goetz, R.; Wiesenfeldt, M.; Sturzebecher, J.; Bode, W.; Friedrich, R.; Thurk, M.; Schwienhorst, A. Thrombin inhibitors identified by computer-assisted multiparameter design. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 8597−8602.

(4) Steinmetzer, T.; Sturzebecher, J. Progress in the development of synthetic thrombin inhibitors as new orally active anticoagulants. *Curr. Med. Chem.* **2004**, *11*, 2297−2321.

(5) Stassen, J. M.; Arnout, J.; Deckmyn, H. The hemostatic system. *Curr. Med. Chem.* **2004**, *11*, 2245−2260.

(6) Matsuo, T.; Koide, M.; Kario, K. Development of argatroban, a direct thrombin inhibitor, and its clinical application. *Semin. Thromb. Hemostasis* **1997**, *23*, 517−522.

(7) Sorbera, L. A.; Bayes, M. C. J.; Silvestre, J. Melagatran and Ximelagatran. *Drugs Future* **2001**, *26*, 1155−1170.

(8) Dullweber, F.; Stubbs, M. T.; Musil, D.; Sturzebecher, J.; Klebe, G. Factorising ligand affinity: a combined thermodynamic and crystallographic study of trypsin and thrombin inhibition. *J. Mol. Biol.* **2001**, *313*, 593−614.

(9) Sturzebecher, J.; Hauptmann, J.; Steinmetzer, T. Thrombin. In *Proteinase and Peptidase Inhibition*; Smith, H. J., Simons, C., Eds.; Taylor and Francis: London, 2002; pp 178−201.

(10) Oprea, T. I.; Davis, A. M.; Teague, S. J.; Leeson, P. D. Is there a difference between leads and drugs? A historical perspective. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1308−1315.

(11) Teague, S. J.; Davis, A. M.; Leeson, P. D.; Oprea, T. The design of leadlike combinatorial libraries. *Angew. Chem., Int. Ed.* **1999**, *38*, 3743−3748.

(12) Hann, M. M.; Leach, A. R.; Harper, G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 856−864.

(13) Hajduk, P. J.; Sheppard, G.; Nettesheim, D. G.; Olejniczak, E. T.; Shuker, S. B.; Meadows, R. P.; Steinman, D. H.; Carrera, G. M., Jr.; Marcotte, P. A.; Severin, J.; Walter, K.; Smith, H.; Gubbins, E.; Simmer, R.; Holzman, T. F.; Morgan, D. W.; Davidsen, S. K.; Summers, J. B.; Fesik, S. W. Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. *J. Am. Chem. Soc.* **1997**, *119*, 5818−5827.

(14) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996**, *274*, 1531−1534.

(15) Carr, R.; Jhoti, H. Structure-based screening of low-affinity compounds. *Drug Discovery Today* **2002**, *7*, 522−527.

(16) Lesuisse, D.; Lange, G.; Deprez, P.; Benard, D.; Schoot, B.; Delettre, G.; Marquette, J. P.; Broto, P.; Jean-Baptiste, V.; Bichet, P.; Sarubbi, E.; Mandine, E. SAR and X-ray. A new approach combining fragment-based screening and rational drug design: application to the discovery of nanomolar inhibitors of Src SH2. *J. Med. Chem.* **2002**, *45*, 2379−2387.

(17) Blundell, T. L.; Abell, C.; Cleasby, A.; Hartshorn, M. J.; Tickle, I. J.; Parasini, E.; Jhoti, H. High-throughput X-ray crystallography for drug discovery. In *Drug Design: Special Publication*; Flower, D. R., Ed.; Royal Society of Chemistry: Cambridge, U.K., 2002; pp 53−59.

(18) Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R. Fragment-based lead discovery. *Nat. Rev. Drug Discovery* **2004**, *3*, 660−672.

(19) Erlanson, D. A.; McDowell, R. S.; O'Brien, T. Fragment-based drug discovery. *J. Med. Chem.* **2004**, *47*, 3463−3482.

(20) Zartler, E. R.; Shapiro, M. R. Fragnomics: fragment-based drug discovery. *Curr. Opin. Chem. Biol.* **2005**, *9*, 366−370.

(21) Nienaber, V. L.; Richardson, P. L.; Klighofer, V.; Bouska, J. J.; Giranda, V. L.; Greer, J. Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nat. Biotechnol.* **2000**, *18*, 1105−1108.

(22) Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A "rule of three" for fragment-based lead discovery? *Drug Discovery Today* **2003**, *8*, 876−877.

(23) Muchmore, S. W.; Olson, J.; Jones, R.; Pan, J.; Blum, M.; Greer, J.; Merrick, S. M.; Magdalinos, P.; Nienaber, V. L. Automated crystal mounting and data collection for protein crystallography. *Struct. Folding Des.* **2000**, *8*, R243−R246.

(24) Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Frederickson, M.; Tickle, I. J.; Jhoti, H. Fragment-based lead discovery using X-ray crystallography. *J. Med. Chem.* **2005**, *48*, 403−413.

(25) Gill, A. L.; Frederickson, M.; Cleasby, A.; Woodhead, S. J.; Carr, M. G.; Woodhead, A. J.; Walker, M. T.; Congreve, M. S.; Devine, L. A.; Tisi, D.; O'Reilly, M.; Seavers, L. C. A.; Davis, D. J.; Curry, J.; Anthony, R.; Padova, A.; Murray, C. W.; Carr, R. A. E.; Jhoti, H. Identification of Novel p38α MAP Kinase Inhibitors using fragment-based lead generation. *J. Med. Chem.* **2004**, *48*, 414−426.

(26) Szczepankiewicz, B. G.; Liu, G.; Hajduk, P. J.; Abad-Zapatero, C.; Pei, Z.; Xin, Z.; Lubben, T. H.; Trevillyan, J. M.; Stashko, M. A.; Ballaron, S. J.; et al. Discovery of a potent, selective protein tyrosine phosphatase 1B inhibitor using a linked-fragment strategy. *J. Am. Chem. Soc.* **2003**, *125*, 4087−4096.

(27) Leung, D.; Abbenante, G.; Fairlie, D. P. Protease inhibitors: current status and future prospects. *J. Med. Chem.* **2000**, *43*, 305−341.

(28) Hartshorn, M. J. AstexViewer: a visualisation aid for structure-based drug design. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 871−881.

(29) Watson, P.; Verdonk, M. L.; Hartshorn, M. J. A web-based platform for virtual screening. *J. Mol. Graphics Modell.* **2003**, *22*, 71−82.

(30) Verdonk, M. L.; Cole, J. C.; Hartshorn, M.; Murray, C. W.; Taylor, R. D. Improved protein−ligand docking using GOLD. *Proteins* **2003**, *52*, 609−623.

(31) Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245*, 43−53.

(32) Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. Flexible docking using Tabu search and an empirical estimate of binding affinity. *Proteins* **1998**, *33*, 367−382.

(33) Sanderson, P. E.; Cutrona, K. J.; Dyer, D. L.; Krueger, J. A.; Kuo, L. C.; Lewis, S. D.; Lucas, B. J.; Yan, Y. Small, low nanomolar, noncovalent thrombin inhibitors lacking a group to fill the "distal binding pocket". *Bioorg. Med. Chem. Lett.* **2003**, *13*, 161−164.

(34) Tucker, T. J.; Brady, S. F.; Lumma, W. C.; Lewis, S. D.; Gardell, S. J.; Naylor-Olsen, A. M.; Yan, Y.; Sisko, J. T.; Stauffer, K. J.; Lucas, B. J.; Lynch, J. J.; Cook, J. J.; Stranieri, M. T.; Holahan, M. A.; Lyle, E. A.; Baskin, E. P.; Chen, I. W.; Dancheck, K. B.; Krueger, J. A.; Cooper, C. M.; Vacca, J. P. Design and synthesis of a series of potent and orally bioavailable noncovalent thrombin inhibitors that utilize nonbasic groups in the P1 position. *J. Med. Chem.* **1998**, *41*, 3210−3219.

(35) Young, M. B.; Barrow, J. C.; Glass, K. L.; Lundell, G. F.; Newton, C. L.; Pellicore, J. M.; Rittle, K. E.; Selnick, H. G.; Stauffer, K. J.; Vacca, J. P.; Williams, P. D.; Bohn, D.; Clayton, F. C.; Cook, J. J.; Krueger, J. A.; Kuo, L. C.; Lewis, S. D.; Lucas, B. J.; McMasters, D. R.; Miller-Stein, C.; Pietrak, B. L.; Wallace, A. A.; White, R. B.; Wong, B.; Yan, Y.; Nantermet, P. G. Discovery and evaluation of potent P1 aryl heterocycle-based thrombin inhibitors. *J. Med. Chem.* **2004**, *47*, 2995−3008.

(36) Skrzypczak-Jankun, E.; Carperos, V. E.; Ravichandran, K. G.; Tulinsky, A.; Westbrook, M.; Maraganore, J. M. Structure of the hirugen and hirulog 1 complexes of α-thrombin. *J. Mol. Biol.* **1991**, *221*, 1379−1393.

(37) Sharff, A. J. High throughput crystallography on an in-house source, using ACTOR. *Rigaku J.* **2004**, *20*, 10.

(38) Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today* **2004**, *9*, 430−431.

(39) Carr, R. A.; Congreve, M.; Murray, C. W.; Rees, D. C. Fragment-based lead discovery: leads by design. *Drug Discovery Today* **2005**, *10*, 987−992.

(40) Breinlinger, E. C.; Keenan, C. J.; Rotello, V. M. Modulation of flavin recognition and redox properties through donor atom−π interactions. *J. Am. Chem. Soc.* **1998**, *120*, 8606−8609.

(41) Rotello, V. M. The donor atom−π interaction of sulfur with flavin. A density functional investigation. *Heteroat. Chem.* **1998**, *9*, 605−606.

(42) Rossky, P. J.; Karplus, M. Solvation. A molecular dynamics study of a dipeptide in water. *J. Am. Chem. Soc.* **1979**, *101*, 1913−1937.

(43) Lim, M. S.; Johnston, E. R.; Kettner, C. A. The solution conformation of (D)Phe-Pro-containing peptides: implications on the activity of Ac-(D)Phe-Pro-boroArg-OH, a potent thrombin inhibitor. *J. Med. Chem.* **1993**, *36*, 1831−1838.

(44) Obst, U.; Banner, D. W.; Weber, L.; Diederich, F. Molecular recognition at the thrombin active site: structure-based design and synthesis of potent and selective thrombin inhibitors and the X-ray crystal structures of two thrombin inhibitor complexes. *Chem. Biol.* **1997**, *4*, 287−295.

(45) Reers, M.; Koschinsky, R.; Dickneite, G.; Hoffmann, D.; Czech, J.; Stuber, W. Synthesis and characterisation of novel thrombin inhibitors based on 4-amidinophenylalanine. *J. Enzyme Inhib.* **1995**, *9*, 61−72.

(46) Murray, C. W.; Verdonk, M. L. The consequences of translational and rotational entropy lost by small molecules on binding to proteins. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 741−753.

(47) Page, M. I.; Jencks, W. P. Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 1678−1683.

(48) Ngai, P. K.; Chang, J. Y. A novel one-step purification of human α-thrombin after direct activation of crude prothrombin enriched from plasma. *Biochem. J.* **1991**, *280* (Part 3), 805−808.

(49) Pflugrath, J. W. The finer things in X-ray diffraction data collection. *Acta Crystallogr., Sect. D* **1999**, *55* (Part 10), 1718−1725.

(50) Collaborative Computational Project, N. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D* **1994**, *50*, 760−763.

(51) Engh, R. A.; Brandstetter, H.; Sucher, G.; Eichinger, A.; Baumann, U.; Bode, W.; Huber, R.; Poll, T.; Rainer, R.; von der Saal, W. Enzyme flexibility, solvent and "weak" interactions characterize thrombin−ligand interactions: implications for drug design. *Structure* **1996**, *4*, 1353−1362.

(52) Verdonk, M. L.; Hartshorn, M.; Mooij, W. T. M.; Murray, C. W.; Taylor, R. D.; Watson, P. Virtual screening using protein−ligand docking: avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*(3), 793−806.

(53) Maple, J. R.; Hwang, M.-J.; Jalkanen, K. J.; Stockfish, T. P. H. A. T. Derivation of class II force fields: V. Quantum force field for amides, peptides and related compounds. *J. Comput. Chem.* **1998**, *19*, 430−458.

(54) Barrish, J. C.; Gordon, E.; Alam, M.; Lin, P. F.; Bisacchi, G. S.; Chen, P.; Cheng, P. T.; Fritz, A. W.; Greytok, J. A.; Hermsmeier, M. A. Aminodiol HIV protease inhibitors. 1. Design, synthesis, and preliminary SAR. *J. Med. Chem.* **1994**, *37*, 1758−1768.

(55) Ascenzi, P.; Fruttero, R.; Amiconi, G.; Pugluse, L.; Bolognesi, M.; Massimo, C.; Onesti, S.; Guarneri, M.; Menegatti, E. Inhibition of α-, and β- and γ- thrombin by mono-, bis-, tris- and tetra-benzamidine structures: thermodynamic study. *J. Enzyme Inhib.* **1992**, *6*, 131−139. 1992.

(56) Magnusson, K. K. A. The binding of proflavin to thrombin. *Arch. Biochem. Biophys*. **1974**, *160*, 175−184.