PRO_SELECT: Combining Structure-Based Drug Design and Array-Based Chemistry for Rapid Lead Discovery. 2. The Development of a Series of Highly Potent and Selective Factor Xa Inhibitors

John W. Liebeschuetz,*,† Stuart D. Jones,† Phillip J. Morgan,† Chris W. Murray,† Andrew D. Rimmer,† Jonathan M. E. Roscoe,† Bohdan Waszkowycz,† Pauline M. Welsh,† William A. Wylie,† Stephen C. Young,† Harry Martin,† Jacqui Mahler,† Leo Brady,‡ and Kay Wilkinson‡

Protherics Molecular Design, Beechfield House, Lyme Green Business Park, Macclesfield SK11 0JL, U.K., and Department of Biochemistry, University of Bristol, Bristol BS8 1TD, U.K.

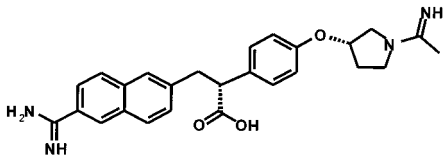
Received June 6, 2001

In silico screening of combinatorial libraries prior to synthesis promises to be a valuable aid to lead discovery. PRO_SELECT, a tool for the virtual screening of libraries for fit to a protein active site, has been used to find novel leads against the serine protease factor Xa. A small seed template was built upon using three iterations of library design, virtual screening, synthesis, and biological testing. Highly potent molecules with selectivity for factor Xa over other serine proteases were rapidly obtained.

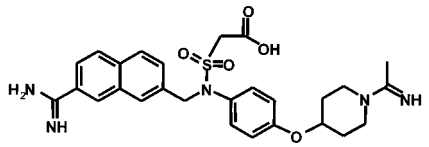
Introduction

Serine proteases represent a class of enzymes of great therapeutic importance. Members of the class which have been targeted for drug design include tryptase and urokinase and, in the blood coagulation cascade, thrombin, factor VIIa, and factor Xa. Factor Xa lies at the junction of the intrinsic and extrinsic pathways of the coagulation cascade. It is the active enzyme present in the prothrombinase complex, which converts prothrombin into thrombin. Thrombin is the final enzymatic product of the blood coagulation cascade and is responsible for the conversion of fibrinogen into fibrin. Much effort has been spent targeting thrombin, in particular, and, more recently, factor Xa, with the aim of designing antithrombotic drugs which are orally available and which show a reduced potential for bleeding as a side effect.[1] Current therapies include the heparins, which are not orally available, and the coumarins, which have a narrow therapeutic window with regard to bleeding. Factor Xa has been claimed to be a better antithrombotic target than thrombin because there are indications that factor Xa inhibitors may have less propensity to show bleeding side effects.[2,3] Additionally, a rebound effect has been observed following cessation of therapy with direct thrombin inhibitors.[4] Potential indications are for deep vein and arterial thrombosis, post operative prophylactic use, myocardial infarction, and stroke.

Crystal derived structural models are available for quite a number of serine proteases. Their mode of catalytic action is well understood, and the structural features that give rise to substrate selectivity have in many cases been elucidated. Recently, for both thrombin and factor Xa, structures have been published which have a variety of competitive inhibitors bound in the active sites. Despite this wealth of structural information, the role of structure-based design, to date, has generally been to help suggest analogues of an existing lead and to post-rationalize the activity data, rather than as a tool for the de novo design of inhibitors.[5,6,7]



DX-9065a



YM-60828

Figure 1. Bis-amidine factor Xa inhibitors DX-9065a and YM-60828.

The first crystal structure of the factor Xa enzyme was published by Bode's group in 1993 (Brookhaven code 1HCG).[8] This crystal structure is missing the Gla (γ-carboxyglutamic acid) domain (N terminal residues 1−45) and also residues Glu 146−Gln 151, close to the active site, which are apparently autocleaved during crystallization. The S1 pocket of the active site is occupied by the A-chain terminal Arg 439 of a neighboring factor Xa molecule, which hydrogen bonds to Asp 189 in the standard bidentate fashion**.** The active site is similar to trypsin, but differs from many other serine proteases in having a large S4 pocket. The S1 pocket differs from trypsin in that Ser 190 is replaced with hydrophobic Ala 190. These features suggest that selective small molecule inhibitors for factor Xa can be obtained and, indeed, many are already known. For example, DX-9065a is a dicationic inhibitor with a K_i of 41 nM against factor Xa and a K_i of 630 nM against trypsin and >2000 nM against thrombin (Figure 1.).[9] YM-60828 is another dicationic inhibitor with a K_i of 2.3 nM against factor Xa, 159 nM against trypsin, and

* Corresponding author: John W. Liebeschuetz, Tularik Ltd, Beechfield House, Lyme Green Business Park, Macclesfield SK11 0JL, U.K. Tel: (44) 1625 427369. Fax: (44) 1625 612311. E-mail: jliebeschuetz@tularik.com.
† Protherics Molecular Design.
‡ University of Bristol.

10.1021/jm010944e CCC: $22.00
Published on Web 02/16/2002

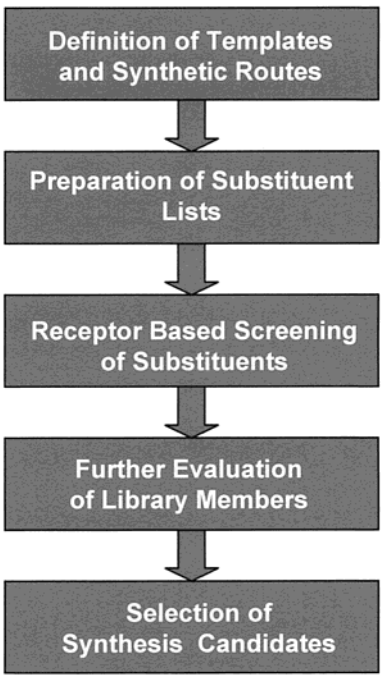


Figure 2. Flow diagram demonstrating the PRO_SELECT approach.

>10 000 nM against thrombin.[10] Both these compounds are currently in preclinical and clinical trials and show some promise as oral antithrombotics without bleeding side effects.[11,12] A crystal derived structure for DX-9065a bound to factor Xa (Brookhaven code 1FAX) exists. This indicates that the naphthamidine portion sits in the S1 pocket, making a single hydrogen bond to Asp 189, while the acetamidinopyrrolidine portion sits in the electron rich S4 pocket and makes a hydrogen bond to the pendant Glu 97 side chain.[13]

Virtual Screening Methodology

We have recently published a methodology for the computational or 'virtual' screening of combinatorial libraries against the active site of an enzyme.[14] The program central to this methodology is called PRO_SELECT. A flow diagram illustrating the methodology is given in Figure 2. The idea is to generate relatively small candidate libraries (10−1000 compounds) which have a high ratio of hits to inactives. These libraries are generally based on a template chemistry and are designed to be synthetically accessible in a timely and cost-effective manner. The catholic nature of the screening procedure allows a wide diversity of structures to be explored within the constraints imposed by the template.

We start with a template structure. This is designed to fit into part of the active site, normally in a central position, and make favorable interactions with the enzyme. The template has attachment points to which substituents can be affixed using simple chemical reactions. Each attachment point is ideally directed toward a different pocket in the binding site, in which a suitable substituent may find favorable interactions. The substituents must be readily accessible to allow rapid chemical synthesis. Therefore the substituents are directly derived from lists of appropriate reagents, each selected from a directory of commercially available chemicals. The selection is normally carried out by searching according to a simple pharmacophore appropriate to the target pocket. Each list may contain several thousand different chemicals.

Each substituent from a list is computationally screened for fit to the target pocket. This was done in two stages in the work presented here. First each substituent is attached to the template and then a user defined number of conformations are roughly assessed for goodness of fit and favorable interactions. This was done by assessing the substituent for complementarity to an 'interaction site model' in the manner of Klebe.[15] The number of attempts at finding graph matches is of the order of a thousand. A match may not represent a viable conformation for a substituent, and this is checked by carrying out a local conformationally flexible fitting of the substituent onto the interaction sites using a directed tweak algorithm. The number of attempts at finding a viable conformation from a given match is of the order 30 to 120. Many substituents may be rejected at this stage. Accepted substituents can be further refined in a second stage, using molecular mechanics to optimize internal geometries and substituent:receptor interactions, via an implementation of the 'CLEAN' force field.[16] Each substituent is scored using an empirical scoring function to find and preserve the best match for each substituent. More recent versions of PRO_SELECT use a docking protocol on both template and substituent to obtain a good binding conformation in which the template position can be adjusted.[17] The empirical scoring function represents binding energy and is derived by regression analysis of measured binding affinity to terms known to be important in determining affinity and calculable from existing crystal structures. A number of different such functions have been published. The empirical scoring function used in this work is that of Böhm, although we have subsequently derived our own scoring function, ChemScore.[18−20]

The score is thus used to drive placement of the substituent. It is also used by the molecular designer to differentiate between substituents in the design of the final sublibrary. It is generally used only as a cutoff filter to pick out those substituents which have the best chance of making good binding interactions. It is not expected that the Böhm score will correlate well with actual binding affinities of the library members, as the Böhm score is derived from complexes in which only favorable ligand:protein interactions are made. Unfavorable interactions which are not penalised by the Böhm score, for instance polar:lipophilic contacts, are liable to frequently arise in the PRO_SELECT placements. For this reason, other criteria are also used in the selection process. These can include strain energy estimation, a diversity metric or calculated physical chemical properties. Manual inspection of the predicted binding modes also plays an important role. The process may be repeated for separate substituent lists attached to different points on the template. Thus the final library will consist of an array of substituents for a single attachment point or a combinatorial array constructed from two or more lists each corresponding to a separate attachment point.

PRO_SELECT was first validated using thrombin as a target.[21] We now report on the use of this methodology

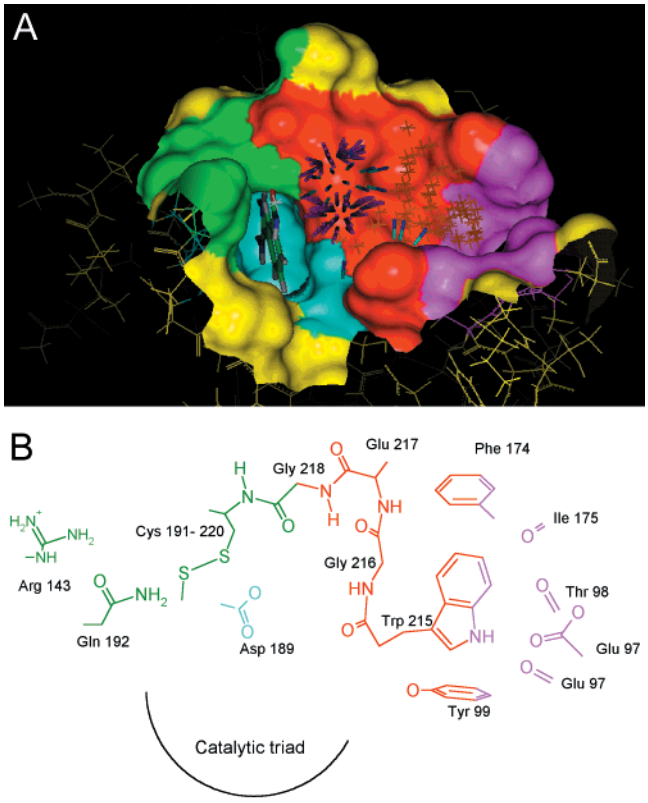


Figure 3. Factor Xa active site VdW surface (A) and schematic (B) illustrating the strategy used in the iterative design process. The initial template, 3-benzamidinecarbonyl, and the interaction site model for the first library are shown (A).

to design factor Xa inhibitors as the primary stage in an ongoing program to discover novel antithrombotic drugs. This has led to a new class of chemistry that is highly active and specific for factor Xa. Several groups have described similar 'Virtual Screening' approaches that have lead to active inhibitors against other targets.[22,23] To our knowledge this is one of the first examples where such an approach has led to molecules sufficiently active to show therapeutic effect in relevant animal disease models. It is also the first published example of the de novo design of inhibitors for factor Xa.

Design Strategy

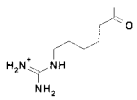
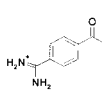
It is usual, when considering the design of a combinatorial library, to build the library around a central template to which two or more substituents are attached or incorporated via facile chemistry. The placement of such a template in an active site would necessarily be central. Examination of the factor Xa 1HCG structure reveals that the central region of the active site is very broad, and the disposition of polar 'anchoring' sites is sparse (Figure 3A). Therefore we felt no confidence that a central template could be designed which would be guaranteed to bind in a single predictable manner. An alternative strategy was chosen in which a template would be placed in one of the specificity pockets and PRO_SELECT would be used to find potential substituents to fit the central portion of the active site. Synthesis and screening of this initial library would then allow a lead molecule to be selected. This could be further elaborated, again by use of PRO_SELECT, to exploit other specificity pockets within the active site. The S1 pocket was chosen as the most suitable pocket in which to place a template. Inhibitors of the trypsin class of serine proteases generally have a cationic moiety in this pocket which can make a single or bidentate H-bond with Asp 189. This provides a good anchor and a predictable binding orientation for a variety of ligands. The strategy of iterative library design to grow into different regions of the factor Xa active site is demonstrated in Figure 3, with reference to the 1HCG structure. This was the only structure openly available at the start of this work and is therefore the one that was used. Initial placement of the template in S1 (blue) was to be followed by development of the first library to probe the central (red) region. This was to be followed up by second and third libraries to occupy the green and purple pockets. The red region incorporates the Ser 214 to Gly 218 backbone, the position of which is well conserved in many serine proteases and which is known to be capable of providing H-bonding recognition interactions with natural substrates. It also partially contains the characteristic 'aromatic box' of factor Xa, constructed from the Tyr 99, Trp 215, and Phe 174 side chains. This 'box' constitutes the majority of the S4 pocket. The purple region represents the back of the S4 pocket and is characterized by three backbone carbonyls from Thr 98, Glu 97, and Lys 96 and the anionic Glu 97 side chain which can overhang the pocket. In theory these groups are available to hydrogen bond strongly to electropositive and cationic groups. The green region represents a hydrophobic pocket that has as its base the Cys 191−Cys 220 disulfide bridge, Gln 192 and Arg 143 side chains as the left-hand wall, and the Gly 218 backbone as the right-hand wall. This pocket is well conserved in many serine proteases but has not been exploited frequently in the design of inhibitors, perhaps because the pocket can often be occupied by the mobile side chain of residue 192. The structure of the potent anti-Xa protein, tick anticoagulant peptide, bound to factor Xa, shows that this ligand can make use of this region, albeit with considerable reorganization of the active site.[24] We were of the opinion, after consideration of the 1HCG structure, that this pocket could be a prime target area for substituent placement in factor Xa.

Results and Discussion

Template Selection and First Library Design. It was decided to employ an amidino group to anchor the S1 template via a bidentate hydrogen bond to Asp 189. We were aware that such a group has in the past lead to problems in oral availability and rapid clearance. Nevertheless it was felt important to use a template that had reasonable base activity of its own so that structure/activity trends would become immediately apparent.

It was envisaged that the synthetically facile linkage of the template to the substituent would be via an amide bond. It was also envisaged that this amide bond might, itself, be able to make hydrogen bonding interactions with the active site. Several possible template candidates were considered. It was decided to use PRO_SELECT to design a library with each template and to compare the quality of the libraries in order to select the best template. Three of the templates examined are shown in Table 1.

Table 1. Hit Rate and Hit Quality of Libraries Designed around Three Different S1 Templates and a Single Substituent List, Using PRO−SELECT

Template			
No. of Substituents satisfying the pharmacophore	2331	2331	2331
No. of hits after PRO_SELECT	672	585	615
No. of hits with Böhm Score < -3.0 kJ mol^{-1}	178	32	163

The initial stage in using PRO_SELECT was to create a 'Design Model'. This is a simple 'interaction site model" of the active cleft.[14] The Design Model for the region targeted for the first library is included in Figure 3A. The dark/light blue vectors represent hydrogen bond donor sites, the blue/purple vectors represent hydrogen bond acceptor sites, and lipophilic point sites are represented as orange crosses. The first step in substituent evaluation was to find matches between interaction sites in the substituent, attached to template, and complementary sites in the design model. Also illustrated in Figure 3A is one of the templates bound in the S1 pocket in the binding orientation employed in the PRO_SELECT job.

The choice of substituent lists and the virtual screening protocol used are given in the Experimental Section. A summary of the output performance is given in Table 1 for each of the three templates. The number of substituents that passed the matching stage was roughly the same for each template. However, the number of substituents binding reasonably well (Böhm score < −3) was much less for the 4-benzamidinocarbonyl template than for the other two templates. The arginine template had marginally more high quality hits than the 3-benzamidinocarbonyl template. However, the Böhm score for the former was 10 kJ mol $^{-1}$ higher than that for the latter (−5.3 versus −16.4), corresponding to a shortfall of roughly 2 orders of magnitude in terms of binding affinity. It was concluded therefore that the 3-benzamidinocarbonyl template was the most appropriate template to use.

A disappointing feature, even in the case of the 3-benzamidinocarbonyl template, was the paucity of high scoring substituents (score < −10). Substituents could be found which made either the desired polar interactions or hydrophobic interactions but rarely both. The number of highly promising substituents therefore appeared limited. The reason for this is likely to be the restriction on substituent diversity, placed upon us by using solely the Available Chemicals Directory as a source. Six targets from this first PRO_SELECT run were selected for synthesis.

One substituent that did score exceptionally well was the glycine-2-naphthylamide, **1** (Figure 4). The naphthyl group was found to sit well in the hydrophobic S4 pocket, and the glycine C═O was able to H-bond with Gly 218. Both of these effects led to a good Böhm score. When modeled against the active site, it was found that the NH from the benzamide amide group also could make a hydrogen bond to Gly 216. This naphthylglycinamide substituent was found to be unobtainable, and the presence of the highly carcinogenic β-naphthylamine substructure mitigated against its synthesis. Nevertheless the motif looked of sufficient interest to be investigated further. Accordingly it was decided to append the glycine to the 3-amidinobenzoyl group and use this molecule as a larger template. PRO_SELECT was used to search for substituents which, when attached via an amide bond to the glycine carbonyl, would probe deeper into the S4 pocket. It was envisaged that it might be possible to use the carbonyl groups at the back of the S4 pocket for hydrogen bonding. For this reason, the library of substituents chosen for virtual screening was selected to be bisamines separated by a hydrophobic group. PRO_SELECT offers the ability to interconvert functional groups in silico prior to virtual screening. This 'deprotection' facility mimics a synthetically facile functional group transforming reaction.[13] This was exploited here to broaden the diversity of possible substituents by inclusion of a list of bis-nitriles, converted on the fly to bis-amines.

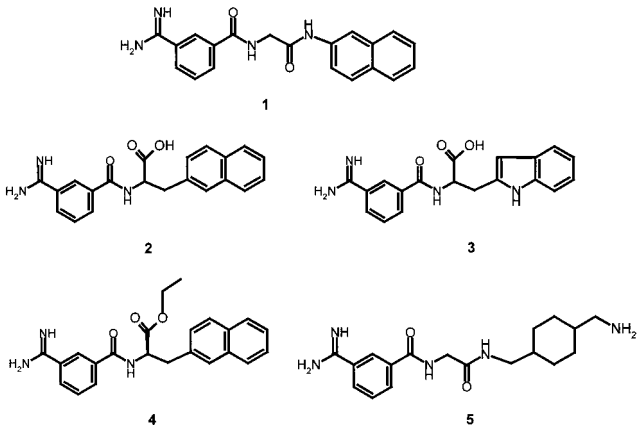


Figure 4. Lead compounds arising out of library 1.

Further details of the virtual screening protocol are given in the Experimental Section. Eight substituents were chosen which had Böhm contributions of better than −10 and strain energies of lower than 25 kJ mol $^{-1}$.[25] Good hits from this second search were incorporated with those of the first, and 14 compounds in total were selected for synthesis (library 1a).

Addition of bulky and partially hydrophobic substituents to a small template could arguably be expected to lead to increased efficacy through nonspecific lipophilic interactions. We wanted to be confident that any increase in activity in the initial library was not generated this way. Therefore it was decided to prepare a second library using some of the same substituents selected for the 3-amidinobenzoyl template but attached to the 'wrong' 4-amidinobenzoyl template. Eight substituents were chosen (library 1b).

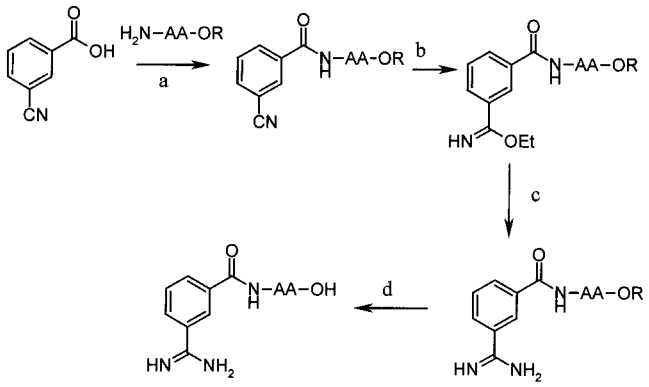
The synthetic routes used for the preparation of these compounds are shown in Schemes 1 and 2. Simple amine substituents, as exemplified by amino acids and their esters, were prepared by coupling with 3-cyanobenzoic acid (using TBTU in DMF), convertion to the imidate (HCl in ethanol), and then to the amidine (using ammonia in ethanol). Hydrolysis of the ester was accomplished using aqueous sodium hydroxide in ethanol, Scheme 1.

Table 2. Comparison of Activities against Factor Xa, Trypsin, and Thrombin for Benzamidine and Libraries 1a−c, 2a−e, and 3a[a]

		factor Xa				trypsin				thrombin			
library	*n*	mean pK_i	±SD	mean K_i[b] (μM)	best K_i (μM)	mean pK_i	±SD	mean K_i[b] (μM)	best K_i (μM)	mean pK_i	±SD	mean K_i[b] (μM)	best K_i (μM)
benzamidine		**3.7**		**200**	**200**	**4.8**		**17**	**17**	**4.6**		**25**	**25**
1a	**14**	**4.3**	**±0.6**	**50**	**8.5**	**4.4**	**±0.8**	**38**	**6**	**4.5**	**±0.7**	**31**	**5**
1a (subset)	7	4.2	±0.8	58	8.5	4.2	±0.9	61	6	4.6	±0.8	23	5
1b	7	3.1	±0.3	780	162	3.7	±0.9	180	2.3	3.6	±0.5	230	89
1a + 1c	**36**	**4.4**	**±0.8**	**36**	**2.1**	**4.2**	**±0.8**	**61**	**6.5**	**4.3**	**±0.9**	**51**	**3.0**
2a	**6**	**5.5**	**±0.8**	**3.5**	**0.22**	**4.6**	**±0.6**	**26**	**9**	*c*		*c*	*c*
2b	6	4.7	±0.7	21	1.0	4.3	±0.4	44	21	*c*		*c*	*c*
2c	9	4.7	±0.4	21	2.9	4.3	±0.4	43	8	*c*		*c*	*c*
2d	11	4.5	±0.5	29	3.0	4.3	±0.4	47	15	*c*		*c*	*c*
2a + 2e	**34**	**5.7**	**±0.8**	**1.9**	**0.063**	**4.9**	**±0.7**	**11**	**0.79**	**4.7**[d]	**±0.5**	**18**	**1.4**
3a	**106**	**6.5**	**±0.8**	**0.34**	**0.016**	**5.4**	**±0.6**	**4.1**	**0.040**	**5.2**[e]	**±0.6**	**7**	**0.023**

[a] Figures not in bold represent benchmark libraries. [b] K_i figures are geometric means calculated as the reciprocal antilog of the pK_i mean. [c] Insufficient compounds tested against thrombin. [d] Only 23 compounds tested against thrombin. [e] Only 99 compounds tested against thrombin.

Scheme 1. Solution-Phase Synthesis of Inhibitors[a]



[a] Conditions: (a) TBTU in DMF; (b) HCl in ethanol; (c) ammonia in ethanol; (d) NaOH in ethanol.

Where the S4 unit was a bis-amine a solid-phase route was preferred, Scheme 2. The bis-amine was attached to 2-chlorotrityl polystyrene resin and coupled to an Fmoc protected amino acid using TBTU in DMF. The Fmoc protection was removed with piperidine in DMF and free amine coupled to 3-amidinobenzoic acid TFA salt using DIPCI and HOBt. The product was then cleaved from the resin using TFA/triethylsilane.

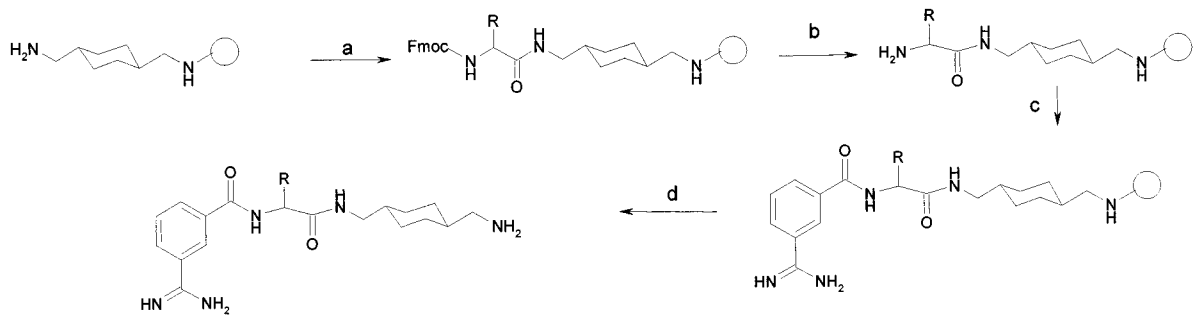
Compounds were tested in chromogenic assays and K_i's calculated against a range of serine proteases. These included factor Xa, trypsin, and thrombin.

The mean, standard deviation, and best activities of libraries 1a and 1b against factor Xa, trypsin, and thrombin are given in Table 2. The corresponding activities for the subset of library 1a with common substituents to library 1b are also given. The designed library 1a with the 3-amidinobenzoyl template demonstrates on average markedly improved activity over benzamidine (K_i of 200 μM, pK_i of 3.7). Moreover, the 'control' library with the 4-amidinobenzoyl template, 1b, shows no such improvement and is, on average, more than an order of magnitude less active than the corresponding subset of compounds from library 1a. The elevation of activity in library 1a is therefore not simply a function of increasing the size of the molecule by addition of a lipophilic fragment.

It was decided to broaden the library 1a by making some simple structure-predicated modifications of some of the hits. Thus a variety of more lipophilic esters of the naphthylalanine and tryptophan analogues **2** and **3** were prepared, with the idea they might better fill the S4 pocket (library 1c). This led to the low micromolar lead **4**. This compound is selective for factor Xa over trypsin (K_i of 2.0 μM vs K_i of 12.5 μM) despite the fact that benzamidine itself is 10 fold better for trypsin (K_i of 20 μM vs trypsin).

Another one of the most active compounds in library 1a was **5** (K_i, 14 μM), and this was selected as a second possible lead compound. The more active of the two leads, **4**, did not readily lend itself to modification via easily accessible chemistry. Compound **5**, on the other hand, looked ideally set up for quick variation, both by replacement of the glycine with other amino acids and by replacement of the 1,4-bis-aminomethylcyclohexane fragments. This was therefore chosen as the lead molecule from which to develop the second library.

Scheme 2. Solid-Phase Synthesis of Inhibitors[a]



[a] Conditions: (a) TBTU/DIPEA and Fmoc-amino acid in DMF; (b) 20% piperidine in DMF; (c) 3-amidinobenzoic acid TFA salt and DIPCI/HOBt in DMF; (d) 10% triethylsilane in TFA.

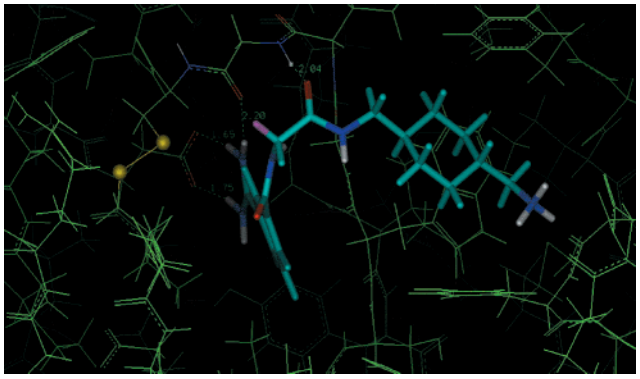


Figure 5. Compound **5** modeled in the 1HCG factor Xa structure.

Second Library Design. Figure 5 illustrates **5** docked into the 1HCG structure. The terminal nitrogen of the S4 binding portion has been modeled as protonated. The glycine in **5** can be elaborated from either of the prochiral hydrogens. Elaboration involving a D-amino acid or analogue thereof (i.e., coming off the hydrogen marked purple in Figure 5) appeared to have a good chance of exploiting the lipophilic disulfide pocket (green area, Figure 3, disulfide in yellow in Figure 5) according to the model. The enantiomeric L-amino acids appear not to be able to access good binding sites in the vicinity. PRO_SELECT jobs were carried out using the list of available α-amino acids. Both L and D configurations were examined, and both polar and hydrophobic interactions were sought. The list of amino acids that resulted consisted mainly of D-amino acids, which docked into the disulfide pocket. A library of seven compounds from this list, all D-enantiomers, was synthesized (sublibrary 2a), all with the 1,4-bis-aminomethylcyclohexane S4 fragment. The corresponding L-enantiomers were also made (sublibrary 2b). A variety of other D-amino acids (9 in all, sublibrary 2c) and L-amino acids (11 in all, sublibrary 2d) was also utilized. Thus libraries 2b, 2c, and 2d represent benchmark libraries to compare with 2a. The solid-phase synthetic route in Scheme 2, was applicable to all the library 2 compounds.

Table 2 gives the mean, standard deviation, and best activities for all four sublibraries against factor Xa, trypsin, and thrombin.

The figures in Table 2 indicate that designed sublibrary 2a is, on average, almost an order of magnitude more active against factor Xa than any of the three comparison libraries, 2b, 2c, and 2d. However, this is not mirrored in the average trypsin activities. The compound within the library that showed the highest activity was **6** (Figure 6), with a K_i of 220 nM against factor Xa and 7.4 μM against trypsin. Analysis of the PRO_SELECT run and further modeling revealed that the phenyl group appeared able to sit well into the disulfide pocket. It also appeared able to make an edge-on interaction with the disulfide bridge. This compound was selected as the lead molecule for the third cycle of PRO_SELECT driven optimization.

Analysis of the 1HCG structure suggested that there was plenty of room left in the disulfide pocket for further hydrophobic elaboration, especially in the vicinity of the 3 and 4 positions of the phenyl ring in **6**. Accordingly, to exploit this extra binding possibility, further analogues were designed, using both medicinal chemistry principles and modeling (library 2e). Only modest increases in activity were achieved. The reasons for this remained unclear until the crystal structure of DX-9065a bound to factor Xa was published (1FAX).[12] This structure retains the autolysis loop, missing in the 1HCG structure. This loop sits at the back of the disulfide pocket area, severely curtailing its size and depth.

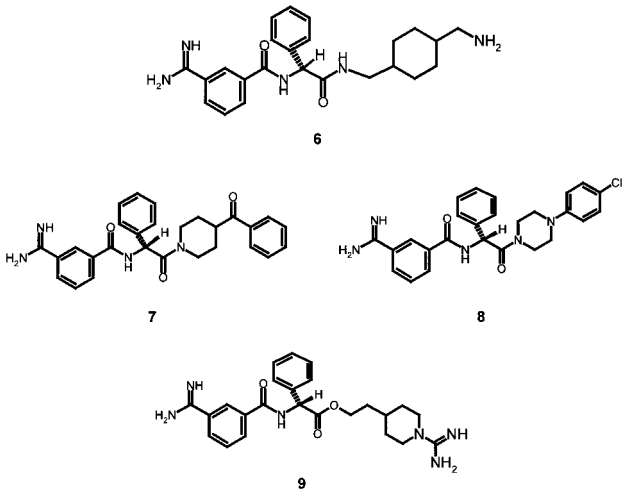
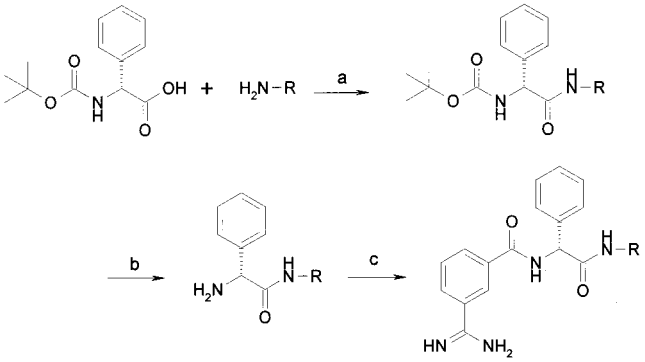


Figure 6. Lead molecule for library 3 (**6**) and examples of library 3 compounds.

Third Library Design and Activity. Only a limited number of substituents designed to access the S4 pocket was tried in libraries 1 and 2. Many of these were diamines. However, it was accepted that there was considerable scope to find better S4 pocket binders and that these could either contain cations or hydrogen bond donors or, alternatively, they could be hydrophobic in nature and interact strongly with the aromatic 'box'. Therefore it was decided to carry out PRO_SELECT jobs using a docked conformation of **6** as the template, with the aim of replacing the terminal diamine with a hydrophobic primary or secondary monoamine. Several jobs were run. The list of available monoamines was large, and the pocket to be filled is also sizable. Therefore, a much bigger list of quality substituents was found from these jobs than from those run previously. Approximately 100 targets were selected (library 3a). The vast majority of these compounds contained a lipophilic S4 binder. Where the S4 binder was lipophilic, a solution-phase synthetic route was used (Scheme 3). Boc-D-phenylglycine was coupled to an S4 component in DMF using EDC/HOBt or HOAt. The products thus obtained were deprotected, using TFA in DCM, and then coupled to 3-amidinobenzoic acid TFA salt. Where the S4 binder was a diamine, the solid-phase route (Scheme 2) could be used.

Summary activity data for library 3a is given in Table 2. Activity, in relation to the lead compound **6**, was generally retained and, in many cases, improved upon in this library, despite the fact that there was usually no possibility of obtaining a hydrogen bond to the back of the S4 pocket in the new series. The most active targets out of this set of analogues, **7**, K_i 16 nM, and **8**, K_i 16 nM (Figure 6), showed a greater than 10-fold increase in activity over **6** and retained selectivity over other serine proteases (K_i's of 980 and 1700 nM against

Scheme 3. Solution-Phase Synthesis of Inhibitors[a]



[a] Conditions: (a) coupling agent—see text; (b) 25% TFA in DCM; (c) 3-amidinobenzoic acid TFA salt and DIPCI/HOBt in DMF.

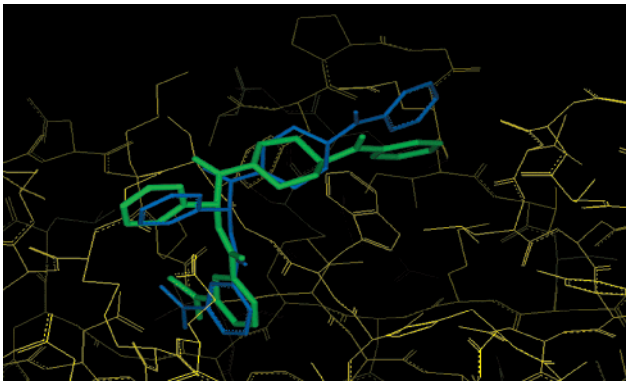


Figure 7. Compound **7** bound in trypsin (green) and superimposed with the predicted binding mode in Factor Xa (blue).

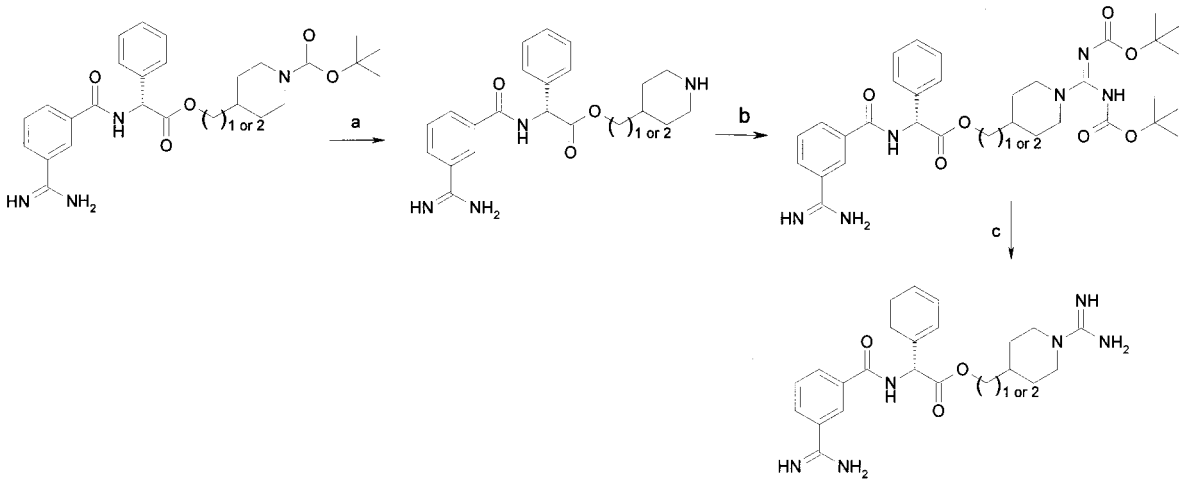
trypsin, 1100 and 1600 nM against thrombin). These molecules are both amenable to further optimization at the S4 end of the molecule and thus represent good third cycle leads. Selectivity for the library as a whole against thrombin and trypsin was slightly improved, despite the fact that both these enzymes have sizable lipophilic pockets that correspond to the Xa S4 pocket.

A cocrystal of compound **7** bound in trypsin, was successfully obtained. Figure 7 compares the predicted binding mode in factor Xa (blue) with that found in trypsin (green). The benzamidine is found in S1, hydrogen bonding in a bidentate fashion to Asp189; the phenyl portion of the phenylglycine linker is found in the disulfide pocket, which can also be accessed in trypsin, and the benzoyl piperidine group sits centrally in the S4-pocket with the carbonyl pointing upward. Both amide bonds in the ligand make the predicted hydrogen bonds. The first H-bonds Gly 216 through NH, the second, Gly 218, through C═O. The bound ligands appear least well superimposed in the S4 region. However, trypsin and factor Xa differ in the S4 pocket, most noticeably at residues 99 and 174 (Leu and Gln in trypsin). Leu 99 allows more room for the terminal phenyl group to sit on the right-hand side of the pocket, than does Tyr 99 in factor Xa.

The third library contained a wide diversity of active chemistries. This gave rise to a number of structurally different lead molecules that could be exploited further using either classical medicinal chemistry or a structure informed approach. Compound **9** represents one example where structure-based modification of a PRO_SELECT lead gave rise to other chemistries with potent activity and selectivity. The synthesis of this type of compound is outlined in Scheme 4. The Boc-N-protected S4 intermediates were first prepared as shown in Scheme 3. The Boc protection was removed using TFA/DCM, and the resulting amine reacted with 1,3-bis-*tert*-butoxycarbonyl methyl pseudothiourea in the presence of mercury II chloride to give the bis-butoxycarbonyl guanidines. Final treatment with TFA/DCM and purification by preparative HPLC gave the final products isolated as TFA salts. Compound **9** has a K_i of 26 nM against factor Xa but only 1.6 μM against trypsin and 8 μM against thrombin. Other related compounds had similar activity and selectivity factors of 250.

Overview. Figure 8 illustrates how the activity of the series evolved. The benzamidine template and lead molecules from each library are indicated. Each library contains, as well as the original PRO_SELECT members, additional structure-based and medicinal-chemistry-based analogues, some of which are mentioned above. To obtain predicted binding affinities for all compounds under a consistent set of docking conditions, all molecules were subsequently redocked, keeping the appropriate template portion of each molecule rigid as in the original PRO_SELECT run. Predicted binding affinity was calculated from the docking score for each molecule. Predicted binding affinity is plotted against

Scheme 4. Preparation of S4 Guanidino Compounds[a]



[a] Conditions: (a) TFA/DCM; (b) 1,3-bis-*tert*-butoxycarbonyl methyl pseudothiourea/$HgCl_2$; (c) TFA/DCM.

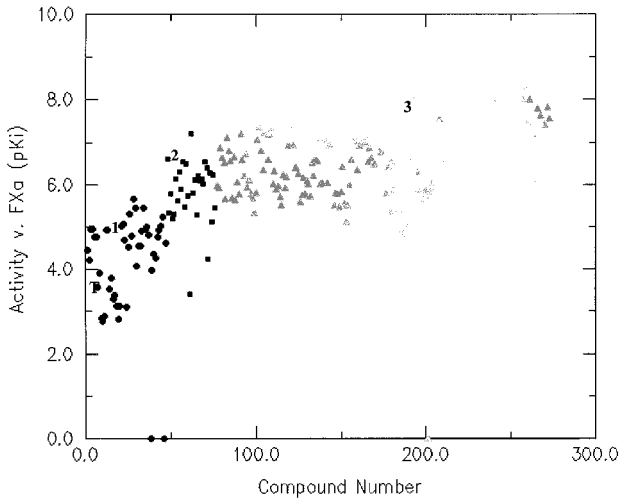


Figure 8. Activity progression in the benzamidine series from the first library (circles) to the second (squares) and third (triangles) libraries. Original template and lead molecules for subsequent library development are marked.

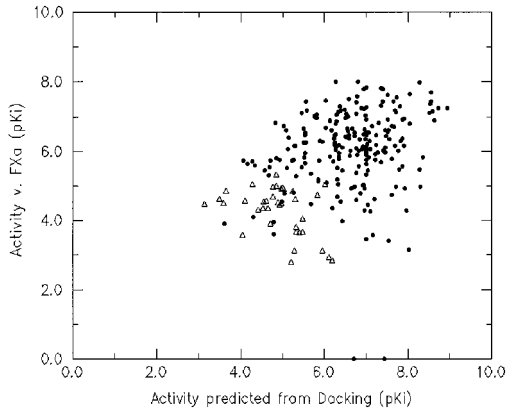


Figure 9. Activity against factor Xa in the benzamidine series versus predicted activity calculated from docking score. Compounds are from both the PRO_SELECT (●) and benchmark (△) libraries.

factor Xa activity in Figure 9. Both those compounds selected through virtual screening and those not so selected (libraries 1b and 2b,c,d) are plotted. There is not a tight correlation. Nevertheless, out of the selected compounds, there is only a small proportion which score well but which show poor activity. In addition, those compounds not selected through virtual screening generally show both poor predicted and actual activity. These things are what we would hope to see as the primary aim of the virtual screening approach is to concentrate synthetic resource in the areas where reasonable activity is most likely to reside. One of the reasons the correlation of predicted and measured activity is not better is because the empirical scoring function is only able to describe positive features of the binding mode and not negative ones other than rotational entropy. In addition, current scoring functions ignore subtle electrostatic effects such as π stacking and so on, which can greatly influence activity. So quite a spread of activities is obtained in the final data set. This spread of activity can be very useful, however, as it often envelops a rudimentary structure−activity relationship among subgroups of similar chemistry within the library. This provides a springboard for a classical lead optimization approach. For this reason, we found it important to have good compound integrity in each library. Impure compounds gave rise to data that confounded early stage SAR development in some cases.

Several other points are worth making. First, the lead molecule chosen at the beginning of each cycle was not necessarily the most active molecule in the previous library. It is more important that the lead be reasonably easy to chemically modify and also be likely to have the least pharmacokinetic problems.

Second, this approach is synergistic with modern combinatorial chemical techniques and it could be used with benefit alongside them, to design large focused libraries. However, in such an approach only one synthetic route is generally followed, and it is accepted that a fraction of library members will not be successfully made. All compounds in a PRO_SELECT designed library are to some extent 'cherished', and therefore there is more incentive to make all members of the library than is usual in combinatorial chemical exercises. If medium throughput array chemistry is employed and the libraries are relatively small, then it is practical and useful to develop and employ more than one synthetic route, as was done in this study.

Conclusions

We have described the successful application of 'virtual' screening in the rational design of potent and selective factor Xa inhibitors. The starting point for the program was a simple template, benzamidine, placed in the S1 pocket. Iterative library design built upon the template in order to access other pockets in the active site. The chemistry involved in the synthesis of these libraries was designed to be straightforward, allowing rapid access of targets. Libraries using the same chemistry were also synthesized which were not designed through use of PRO_SELECT but which were reasonable from a medicinal chemistry standpoint. These consistently showed poorer average activity against (by roughly a factor of 10) and selectivity for factor Xa than those designed by 'virtual' screening.

Several lead molecules with diverse structure and K_i 's in the range of 10 to 50 nM were obtained, representing 4 orders of magnitude increase with regard to the binding affinity of the starting template. Further testing established that some of these compounds showed antithrombotic activity when given intraperitoneally in the Wessler stasis model of venous thrombosis in rat and therefore have potential therapeutic use as an injectable treatment. None of the compounds showed an effect when given orally, however. It was felt this was likely to be because of the highly basic benzamidine moiety. Further work has since been carried out to replace this group with a group of moderate basicity and to optimize potency by modification elsewhere in the molecule. Highly potent and selective compounds with strong oral antithrombotic activity were found. These will be described in later publications.

The design of libraries through structure-based 'virtual screening' is a methodology of drug design that is currently of high interest. We have demonstrated that it can be an efficient method for the generation of potent and selective lead molecules, in cases where a target protein structure exists.

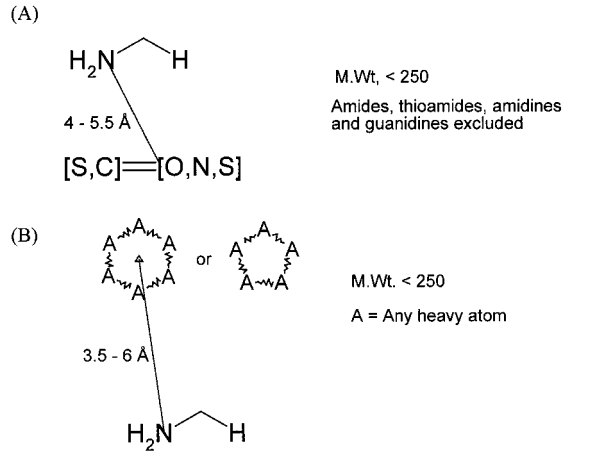


Figure 10. Pharmacophores used in finding substituent lists for library 1a: (A) pharmacophore for H-bond acceptor substituents, (B) pharmacophore for lipophilic substituents.

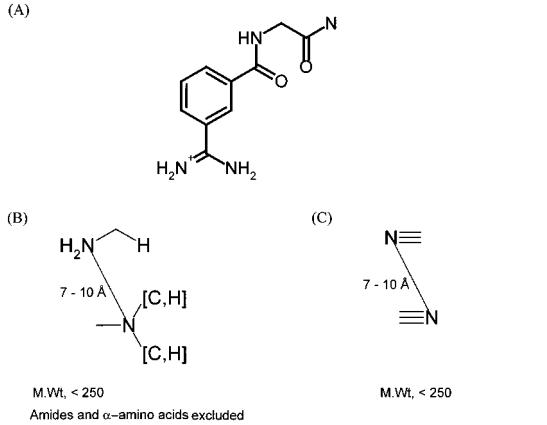


Figure 11. Template and pharmacophore for library 1b: (A) template, (B) pharmacophore for bis-amine substituents, (C) pharmacophore for bis-nitrile list.

Experimental Section

Computational Details. Manipulation and inspection of receptor, template, and ligands before and after minimization or simulation was carried out using InsightII 95.0.[25] All molecular mechanics minimizations and molecular dynamics simulations were carried out using the Discover 2.97 program[26] with the CFF95 force field. The Discover calculations were carried out on a Convex Exemplar (16 × HP7100s) running SPP-UX 3.1. The pharmacophore searches of the Available Chemical Directory (ACD)[27] and substituent list generation were performed using ISIS/Base 1.2[28] and ISIS/Draw 1.[28] Searches were carried out allowing full conformational flexibility. Inspection of the structures and associated numerical data generated by PRO_SELECT was carried out using in-house graphics software (XMOLBROWSE). Substituent list generation and graphical inspection was carried out on SGI Indigo R3000 workstations running IRIX 4.0.5.

Protocol for the Design of Library 1a Members Based on Three S1 Templates. The template positioning for the arginine template (Table 1) was taken from a minimized structure of PPACK in FXa. The docked conformation of PPACK was derived from the 1PPB (PDB designation) PPACK/thrombin structure. The template positioning of the benzamidines was derived from a docking of DX-9056a into the 1HCG (PDB designation) factor Xa structure. The carbonyl group was placed at the ring position proximal to the lip of S4 in the case of the 3-amidinobenzoyl template (Table 1). Both possible orientations of the carbonyl group planar to the ring were treated as valid, and PRO_SELECT runs were carried out on each. Only one orientation was deemed useful for the 4-amidinobenzoyl template.

Two lists of substituents were prepared using the ACD as a source. Pharmacophores for the substituent selection were calculated from the crystal structure assuming reasonable placement of the template and are given in Figure 10. One pharmacophore was targeted toward the polar functionality at the lip of S4, with the primary aim of picking up interactions with the N−H of Gly 218 (Figure 10A). The other was targeted at the hydrophobic 'aromatic box' (Figure 10B). The amino group common to both is the linking group to the template. The initial aim was to only probe the central region of the active site. Therefore a molecular weight limit of 250 was set. Conformationally flexible searching of the ACD with these pharmacophores using ISIS/Base software generated 2D structure lists which were than converted to 3D using Converter[26] (Molecular Simulations Inc.).

The final 'polar' list numbered 1534 substituents, the 'nonpolar' one numbered 797. Each list of substituents was evaluated separately by PRO_SELECT. The polar substituents were assigned hydrogen bonding acceptor sites on double bonded oxygen, nitrogen, and sulfur (this was found as effective at finding hits as assigning both acceptor and donor sites). Hydrophobic sites were assigned to carbons in five- and six-membered carbocyclic rings. Passes from the interaction site matching stage were minimized using the Clean force field, keeping the template rigid, and scored for binding affinity using the empirical scoring function of Böhm.[18] The best scoring conformation per substituent was retained. The final list of substituents was filtered to pick out only those which had Böhm contributions of −3 kJ mol^{-1} or better.

Protocol for the Further Design of Library 1a Members Based around the 3-Amidinobenzoylglycine Template. Molecular dynamics simulations were carried out on ligands **1** and **4**, among others, docked into the 1HCG factor Xa structure. Snapshots were taken from the simulations at 1 ps intervals and each snapshot minimized. Analysis of these simulations allowed the selection of two significantly different template geometries, both of which gave reasonably good Böhm scores. PRO_SELECT runs were carried out using each geometry. The template structure is given in Figure 11A.

The substituent list was prepared by searching the ACD using the pharmacophore in Figure 11B. The list contained 422 substituents. A second bis-nitrile list was prepared according to the pharmacophore in Figure 11C. PRO_SELECT runs were carried out separately on this list which contained 656 substituents. Substituents for library 1a were selected out of that set of substituents with Böhm contributions of −9 kJ mol^{-1} or better, and strain energies of 27 kJ mol^{-1} or lower.[25] The hits were clustered using the Jarvis-Patrick method within XMOLBROWSE, and substituents were selected from those that passed the criteria via manual inspection of each ligand docking. It was found that the two different template positions gave rise to quite different sets of high scoring substituents.

Protocol for the Design of Library 2. The origin of the templates for this library was an "annealed" geometry of **5** manually docked into the 1HCG FXa X-ray structure. The receptor geometry was held rigid while the ligand was subjected to molecular dynamics at temperatures up to 300 K with subsequent slow cooling followed by minimization. This final geometry then provided an appropriate ligand template geometry. The disconnection point for this template was taken to be either of the prochiral hydrogens of the ligand glycine methylene. Each of these hydrogen atoms was substituted during separate PRO_SELECT runs as it was desired to look at both L- and D-amino acids.

The list of potential "amino acid" substituents was obtained by searching the ACD for all free α-amino acids. There were 1230 possible amino acid substituents after removal of high molecular weight compounds (MW > 250) filtering of undesirable chemistries and conversion into 3D. The substituent disconnection point employed here was the amino acid side chain → C-α bond.

Several SELECT jobs were then carried out to focus on lipophilic or polar interactions arising from the amino acid

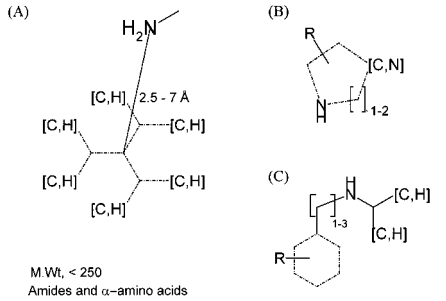


Figure 12. Pharmacophores used in finding substituent lists for library 3: (A) pharmacophore for hydrophobic primary amines, (B) pharmacophore for hydrophobic cyclic secondary amines, (C) pharmacophore for hydrophobic acyclic secondary amines. Aromatic and fused bicyclic systems were allowed in the lists generated by pharmacophores A, B, and C.

substituent with the receptor interaction sites. A list of substituents suitable for synthesis was generated after removal of duplicates, inappropriate substituents, e.g., side chains from substituents available only in L form where the D was preferred, and filtering of poor scoring substituents using the criterion that the Böhm score be less than -3.0 kJ mol^{-1}.

Protocol for the Design of Library 3. Compound **6** and also an analogue of compound **6** which had the bis(aminomethyl)cyclohexane replaced by a 1-adamantylamine, docked into the 1HCG factor Xa structure, were simulated at 300 K. The receptor was kept rigid, and snapshots were taken at 5 ps intervals and minimized. Low energy snapshots were selected to provide two different template positionings. Three pharmacophores were used to search the ACD for appropriate hydrophobic amines.

These pharmacophores, given in Figure 12, represent respectively primary amines, secondary cyclic amines, and secondary acyclic amines. The hydrophobic parts of the pharmacophores were designed so as to avoid linear hydrocarbons. Hits that had several polar groups were excluded, as were certain classes of reactive chemistry. A molecular weight limit of 250 for free base was used. The primary amine list numbered, after conversion to 3D and enumeration of enantiomers and diastereomers, 1053 molecules, and the cyclic amine list numbered 250 compounds. The secondary acyclic amine list, which was restricted to substituents containing a six-membered carbocycle, numbered 366 compounds. The link site was chosen to be the N−H trans to the carbonyl of the phenyl glycine, in the case of the primary amine list, but the C(=O)−N bond for both the secondary and the cyclic lists. Hydrophobic interaction sites were placed at carbons adjacent to a hydrophobic branch point. Passes from the interaction site matching stage were treated as described above. Priority substituent lists were selected on the basis of having favorable Böhm contributions (generally -12 kJ mol^{-1} or better) and low strain energies (generally less than -21 kJ mol^{-1}). These lists were clustered according to chemistry and processed by manual inspection of the binding mode to generate synthetic candidates.

Chemistry. Abbreviations used follow IUPAC−IUB nomenclature. Additional abbreviations are HPLC, high performance liquid chromatography; DMF, dimethylformamide; DCM, dichloromethane; HATU, *O*-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; HOBt, 1-hydroxybenzotriazole; TBTU, 2-(1*H*-(benzotriazol-1-yl)-1,1,3,3-tetramethyluroniumtetrafluoroborate; DIPEA, diisopropylethylamine; TEA, triethylamine; HOAt, 1-hydroxy-7-azabenzotriazole; Fmoc, 1-(9*H*-fluoren-9-yl)methoxycarbonyl; TFA, trifluoroacetic acid; MALDI-TOF, matrix assisted laser desorption ionization−time-of-flight mass spectrometry. Unless otherwise indicated, amino acid derivatives, resins, and coupling reagents were obtained from Novabiochem (Nottingham, U.K.) and other solvents and reagents from Rathburn (Walkerburn, U.K.) or Aldrich (Gillingham, U.K.) and were used without further purification.

Purification was by gradient reverse-phase HPLC on a Waters Deltaprep 4000 at a flow rate of 50 mL/min using a Deltapak C18 radial compression column (40 mm × 210 mm, 10−15 mm particle size) and solvent mixtures consisting of eluant A (0.1% aq TFA) and eluant B (90% MeCN in 0.1% aq TFA) with gradient elution.

Analytical HPLC was on a Shimadzu LC6 gradient system equipped with an autosampler, a variable wavelength detector at flow rates of 0.4 mL/min. Eluents A and B as for preparative HPLC used the following columns: Luna2 C18 2 × 150 mm 5 μm, Symmetry C8 4.6 × 30 mm 3.5 μm (Phenomenex). Purified products were further analyzed by Maldi TOF and/or LCMS and ^{1}H NMR.

Compound libraries were prepared using both solid-phase and solution-phase parallel synthetic methods as described below.

Simple Amine S4 Units (Scheme 1). Amino acid ester hydrochlorides were either (i) coupled to 3-amidinobenzoic acid TFA salt using DIPCI/HOBt in DMF containing 1 equiv of DIPEA, purified by reverse-phase preparative HPLC and isolated as the TFA salt, or (ii) coupled to 3-cyanobenzoic acid using TBTU/HOBt in DMF containing 1 equiv of DIPEA and the nitrile converted to the amidine by sequential treatment with HCl gas in ethanol and ammonia gas in ethanol. The products were purified by reverse-phase preparative HPLC and isolated as the TFA salt. Compounds **3** and **4** were prepared by these routes.

3-Amidinobenzoyl-D-tryptophan TFA Salt, 3. To a solution of 3-cyanobenzoic acid (500 mg, 3.4 mmol), D-tryptophan methyl ester hydrochloride (866 mg, 3.40 mmol), and HOBt (459 mg, 3.4 mmol) in DMF (10 mL) were added TBTU (1.09 g, 3.40 mmol) and DIPEA (592 μL, 3.4 mmol). The reaction was stirred until complete by TLC and then partitioned between ethyl acetate and water. The organic solution was evaporated in vacuo to give 3-cyanobenzoyl-D-tryptophan methyl ester (954 mg, 81%).

HCl gas was bubbled into a solution of 3-cyanobenzoyl-D-tryptophan methyl ester (925 mg, 2.66 mmol) in ethanol, and the mixture was left overnight before evaporating to dryness in vacuo. The solid was taken up in ethanol, and the solution was saturated with ammonia gas. After being stirred overnight, the mixture was evaporated to dryness in vacuo and the resulting solid purified by preparative HPLC to give a mixture of 3-amidinobenzoyl-D-tryptophan methyl and ethyl ester TFA salts.

To a solution of 3-amidinobenzoyl-D-tryptophan methyl and ethyl ester TFA salts (50 mg) in ethanol (5 mL) was added 1 M aqueous sodium hydroxide (1 mL), and the mixture was stirred overnight. The mixture was evaporated to dryness in vacuo and purified by preparative HPLC to give 3-amidinobenzoyl-D-tryptophan TFA salt (11 mg). ^{1}H NMR (CD_3CN) δ 8.14 (1H, s, Ar); 8.08 (1H, d, Ar); 7.82 (1H, d, Ar); 7.68 (2H, m, Ar); 7.45 (1H, d, Ar); 7.05−7.30 (3H, m, Ar); 4.96 (1H, m, α-proton); 3.3−3.5 (d-ABq, β-proton). Homogeneous by HPLC Luna C18, Symmetry C8. High resolution MS $(M+1)^+$ found 351.14555 ($C_{19}H_{18}N_4O_3$ requires 351.14568).

3-Amidino-D-2-naphthylalanine Ethyl Ester TFA Salt, 4. 3-Amidinobenzoic acid TFA salt (100 mg) was added to a mixture of HOBT (48.6 mg) and DIPCI (57 μL) in DMF that had been stirring for 10 min. To this mixture was added a solution of 2-naphthylalanine ethyl ester hydrochloride (100.5 mg) and triethylamine (50 μL). After being stirred overnight, the crude reaction mixture was purified by preparative HPLC to give 3-amidinobenzoyl-D-2-naphthylalanine ethyl ester TFA salt. ^{1}H NMR (CD_3CN) δ 7.98 (1H, s, Ar); 7.88 (1H, d, Ar); 7.67 (5H, m, Ar); 7.50 (1H, t, Ar); 7.3−7.4 (3H, m, Ar); 4.80 (1H, dd, α-proton); 4.0 (2H,q, Et); 3.35−3.1 (d-ABq, β-proton); 1.1 (3H, t, Et). Homogeneous by HPLC Luna C18, Symmetry C8. LCMS 390 $(M+1)^+$, high resolution MS $(M+1)^+$ found 390.18077 ($C_{23}H_{23}N_3O_3$ requires 390.181740).

Bis-amine S4 Units by Solid-Phase Methodology (Scheme 2). The S4 component (bis-1,4-aminomethylcyclohexane) was supported on 2-chlorotrityl resin (1.2 mmol/g) and coupled with an Fmoc protected amino acid using TBTU/

DIPEA in DMF. The washed resin was treated with 20% piperidine in DMF to remove the Fmoc protection and then reacted with 3-amidinobenzoic acid TFA salt using DIPCI/HOBt in DMF. The product was then cleaved off the washed resin using 10% triethylsilane in TFA ,and the crude product obtained was purified by preparative reverse-phase HPLC and isolated as the TFA salt. Compounds **5** and **6** were made by this route.

3-Amidinobenzoyl-glycinyl-(4-aminomethylcyclohexyl)methylamine, 5. ^{1}H NMR (CD_3CN/D_2O) mixture of cis/trans isomers, major isomer only: δ 8.15 (s, 1H, Ar); 8.1 (d,1H, Ar); 7.9 (d, 1H, Ar); 7.66 (t, 1H, Ar); 4.97 (s, 2H, "Gly CH_2"); 3.0 (d, 2H, amide CH_2); 2.70 (d, 2H, amine CH_2); 1.70 (m, 4H, cyclohexyl); 1.40 (m, 3H, cyclohexyl; 0.90 (m, 3H, cyclohexyl). Homogeneous by HPLC Luna C18, Symmetry C8. LCMS 346 $(M+1)^+$, high resolution MS $(M+1)^+$ found 346.22378 ($C_{18}H_{27}N_5O_2$ requires 346.22426).

3-Amidinobenzoyl-D-phenylglycinyl-(4-aminomethylcyclohexyl)methylamine, 6. ^{1}H NMR (D_2O) mixture of cyclohexyl cis and trans isomers 8.09 (1H, s); δ 8.05 (1H, d, J = 7.5 Hz); 7.90 (1H, d, J = 7.5 Hz); 7.66 (1H, t, J = 7.5 Hz); 7.43 (5H, m); 5.47 (1H, s); 3.05 (2H, m); 2.78 (2H, m); 1.48 (7H, m); 0.86 (3H, m), Homogeneous by HPLC Luna C18, Symmetry C8. LCMS 422 $(M+1)^+$, high resolution MS $(M+1)^+$ found 422.25548 ($C_{24}H_{31}N_5O_2$ requires 422.25556).

Solution Phase II (Scheme 3). Boc-D-Phenylglycine was coupled to an S4 component in DMF using HATU or TBTU/DIPEA or, alternatively, EDCI or DIPCI with HOBt or HOAt as an additive. When the S4 component was an alcohol, catalytic DMAP was added. The products thus obtained were deprotected using TFA in DCM and then coupled to 3-amidinobenzoic acid TFA salt using DIPCI/HOBt in DMF. The products were purified by reverse-phase preparative HPLC and isolated as the TFA salts. Compounds **7** and **8** were made by this route.

1-(3-Amidinobenzoyl-D-phenylglycinyl)-4-benzoylpiperidine, 7. To a solution of Boc-D-phenyl glycine (251 mg, 1 mmol) and a mixture of DMF (1 mL) and DCM were added 4-benzoylpiperidine (339 mg 1.5 mmol), DIPEA (348 μL, 2 mmol), and TBTU (353 mg 1.1 mmol). After being stirred at room temperature overnight, the mixture was partitioned between ethyl acetate (6 mL) and 10% hydrochloric acid (2 mL). The organic layer was washed with 10% hydrochloric acid (2 mL), saturated aqueous sodium bicarbonate, and then brine. Evaporation of solvent gave the crude product which was taken up in dichloromethane (2 mL) and treated with trifluoroacetic acid (2 mL) until removal of the Boc group was complete. Solvent was evaporated in vacuo, and the residue was taken up in ethyl acetate and washed with saturated aqueous sodium bicarbonate and then brine before evaporating to dryness. The residue was dissolved in DMF (5 mL), and to this was added a mixture of HOAt (150 mg, 1.1 mmol), 3-amidinobenzoic acid TFA salt (300 mg, 1.08 mmol), and DIPCI (180 μL, 1.15 mmol), and the mixture was stirred overnight. Any solids were removed by filtration, and solvent was removed in vacuo. The residue was taken up in ethyl acetate, washed with saturated aqueous sodium bicarbonate, dried ($MgSO_4$), and evaporated in vacuo. The residue was converted to the TFA salt by addition and evaporation of 25% TFA in acetonitrile and then dissolved in a minimum of aqueous acetonitrile for purification by preparative RPHPLC to give 1-(3-amidinobenzoyl-D-phenylglycinyl)-4-benzoylpiperidine TFA salt (159 mg, 27% over three steps). ^{1}H NMR (DMSO-d_6) δ 8.40 (2H, m); 8.10 (1H, d); 7.70 (1H,t); 7.50 (10H, m); 5.55 (1H, s); 3.60 (1H, m); 2.5 (2H, m); 1.00 (6H,m). Homogeneous by HPLC Luna C18, Symmetry C8. LCMS 469 $(M+1)^+$, high resolution MS $(M+1)^+$ = 469.22282 ($C_{28}H_{28}N_4O_3$ requires 469.223935).

1-(3-Amidinobenzoyl-D-phenylglycinyl)-4-chlorophenylpiperazine TFA Salt, 8. 1-(3-Amidinobenzoyl-D-phenylglycinyl)-4-chlorophenylpiperazine TFA salt was prepared from 4-chlorophenylpiperazine in a manner similar to that described above. ^{1}H NMR (CD_3CN) δ 8.05 (1H, s); 8.00 (1H, d); 7.87 (1H, d); 7.55(1H, t); 7.31 (5H, m); 7.08,(2H,d); 6.75,(2H,d); 5.95 (1H, s); 3.70 (1H,m); 3.55 (2H,m); 3.45 (1H,m); 3.12 (1H,m); 3.00 (1H,m); 2.85 (1H,m); 2.35 (1H,m), Homogeneous by HPLC Luna C18, Symmetry C8. LCMS 476 $(M+1)^+$, high resolution MS $(M+1)^+$ = 476.18340 ($C_{26}H_{26}N_5O_2$ requires 476.18529).

Amidino compounds such as **9** were prepared initially using the solution-phase method II described above to give Boc-N-protected S4 intermediates. The Boc protection was removed using TFA/DCM, and the resulting amine reacted with 1,3-bis-*tert*-butoxycarbonyl methyl pseudothiourea in the presence of mercury II chloride to give the bis-butoxycarbonyl guanidines. Final treatment with TFA/DCM and purification by preparative HPLC gave the final products isolated as TFA salts (Scheme 4).

3-Amidinobenzoyl-D-phenylglycine 1-Amidinopiperidin-4-ylethyl Ester, 9. 3-Amidinobenzoyl-D-phenylglycine 1-Boc-piperidin-4-ylethyl ester was prepared using the general solution-phase method described above for compounds **7** and **8** and then treated with 25% TFA in DCM to remove the Boc protection. Treatment with 1,3-bis-*tert*-butyloxycarbonylmethylthiopseudourea (1 equiv), TEA (3 equiv), and mercury(II) chloride (1 equiv) in DMF overnight followed by extraction into ethyl acetate, washing with 2 N aqueous sodium hydroxide and water, and drying ($MgSO_4$) gave 3-amidinobenzoylphenylglycine 1-(1,3-bis-*tert*-butyloxycarbonyl amidine)-4-piperidin-4-ylethanol ester.

The above compound was treated with 25% TFA in DCM until the Boc protection was removed and then evaporated in vacuo. The residue was purified by preparative RPHPLC to give 3-amidinobenzoyl-D-phenylglycine 1-amidinopiperidin-4-ylethyl ester. ^{1}H NMR (D_2O) 8.17 (1H, m); δ 8.07 (1H, d); 7.93 (1H, d); 7.70 (1H, t); 7.45 (5H, m); 5.60 (1H,s); 4.25 (2H,m); 3.55 (2H,m); 2.75 (2H, m); 1.60 (4H, m); 1.25 (1H, m); 1.00 (2H, m). Homogeneous by HPLC Luna C18, Symmetry C8. LCMS451 $(M+1)^+$, high resolution MS $(M+1)^+$ = 451.244439 ($C_{24}H_{30}N_6O_3$ requires 451.245725).

Biology. Inhibition of factor Xa was assessed at room temperature in 0.1 M phosphate buffer, pH 7.4, according to the method of Tapparelli et al.[29] Purified human factor Xa was purchased from Alexis corporation, Nottingham, U.K. Chromogenic substrate pefachrome-FXA was purchased from Pentapharm AG, Basel, Switzerland. Product (4-nitroaniline) was quantified by absorption at 405 nm in 96-well plates using a Dynatech MR5000 reader (Dynex Ltd, Billingshurst, U.K.). K_m and K_i were calculated using SAS PROC NLIN (SAS Institute, Cary, NC, Release 6.11). K_m values were determined as 100.9 μM for factor Xa/pefachrome-FXA. Inhibitor stock solutions were prepared at 40 mM in dimethylsulfoxide and tested at 500 μM, 50 μM, and 5 μM. Accuracy of K_i measurements was confirmed by comparison with K_i values of known inhibitors of factor Xa.

Crystallization, Data Collection, and Refinement. Bovine trypsin (Sigma, Type III) was further purified by ion-exchange chromatography (Mono S, 0.1 M sodium phosphate pH 6.0, eluted with a 0−1 M sodium chloride gradient). A complex of the purified trypsin with compound **7** was prepared by incubating a 3-fold molar excess of the inhibitor with the enzyme, which was then concentrated to 15 mg/mL in 0.05 M Tris pH 8, 3 mM calcium chloride, 18% acetonitrile, and 5% DMF. Crystals were grown by vapor diffusion against a well containing 2.1 M ammonium sulfate and 0.05 M Tris pH 8.15. Nucleation of crystal growth required streak seeding using low-density crystals grown in the presence of benzamidine according to the procedure of Batunik (1989).[30] Crystals of the bovine trypsin-compound **7** complex belonged to space group $P2_12_12_1$, with a = 60.08 Å, b = 63.83 Å, and c = 70.04 Å, and diffracted to beyond 2.0 Å. A complete native data set, comprising 17 272 unique reflections in the range 30−2.0 Å and with an average redundancy of 4.0, was collected at 100 K using station PX7.2 of the Daresbury SRS synchrotron (wavelength 1.488 Å). These data were processed using DENZO and SCALEPACK,[31] and the structure solved by molecular replacement using AMORE[32] with the coordinates from PDB entry 3PTN as search model. The structure was refined using iterative cycles of simulated annealing refine-

ment with X-PLOR[33] and manual rebuilding using O.[34] The final model has good geometry and an R_{cryst} of 17.8% and R_{free} of 24.0% (calculated with data in the range 15−2.0 Å). The coordinates have been deposited in the PDB (reference code 1eb2).

Acknowledgment. The authors thank Allen Miller for encouragement and for helpful suggestions during the preparation of this manuscript.

References

(1) (a) Wiley, M. R.; Fisher, M. J. Small-molecule direct thrombin inhibitors. *Expert Opin. Ther. Pat.* **1997**, *7* (11), 1265−1282. (b) Menear, K. Progress towards the discovery of orally active thrombin inhibitors. *Curr. Med. Chem.* **1998**, *5*, 457−468. (c) Rewinkel, J. B. M.; Adang, A. E. P. Strategies and progress towards the ideal orally active thrombin inhibitor. *Curr. Pharm. Des.* **1999**, *5*, 1043−1075. (d) Zhu, B.-Y.; Scarborough, R. M. Recent advances in inhibitors of factor Xa in the prothrombinase complex. *Curr. Opin. Cardiovasc., Pulm. Renal Invest. Drugs* **1999**, *1* (1), 63−88. (e) Walenga, J. M.; Jeske, W. P.; Hoppensteadt, D.; Kaiser, B. Factor Xa inhibitors: Today and beyond. *Curr. Opin. Cardiovas., Pulm. Renal Invest. Drugs* **1999**, *1* (1), 13−27. (f) Al-Obeidi, F.; Ostrem, J. A. Factor Xa inhibitors. *Expert Opin. Ther. Pat.* **1999**, *9* (7), 931−953.

(2) Chi, L.; Rogers, K. L.; Uprichard, A. C. G.; Gallagher, K. P. The therapeutic potential of novel anticoagulants. *Expert Opin. Invest. Drugs* **1997**, *6* (11), 1591−1622.

(3) Morishima, Y.; Tanabe, K.; Terada, Y.; Hara, T.; Kunitada, S. Antithrombotic and Hemorrhagic Effects of DX-9065a, a Direct and Selective Factor Xa: Comparison of a Direct Thrombin Inhibitor and Antithrombin III-Dependent Anticoagulants. *Thromb. Haemostasis* **1997**, *78*, 1366−1371.

(4) Gold, H. K.; Torres, F. W.; Garabedian, H. D.; Werner, W.; Jang, I.; Khan, A.; Hagstrom, J. N.; Yasuda, T.; Leinbach, R. C.; Newell, J. B.; Bovill, E. G.; Stump, D. C.; Collen, D. Evidence for a Rebound Coagulation Phenomenon after Cessation of a 4-hour Infusion of a Specific Thrombin Inhibitor in Patients with Unstable Angina Pectoris. *J. Am. Coll. Cardiol.* **1993**, *21*, 1039−1047.

(5) Galemmo, R. A.; Maduskuie, T. P.; Dominguez, C.; Rossi, K. A.; Knabb, R. M.; Wexler, R. R.; Stouten, P. F. W. The *de novo* design and synthesis of cyclic urea inhibitors of Factor Xa: Initial SAR studies. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2705−2710.

(6) Klein, S. I.; Czekaj, M.; Gardner, C. J.; Guertin, K. R.; Cheney, D. L.; Spada, A. P.; Bolton, S. A.; Brown, K.; Colussi, D.; Heran, C. L.; Morgan, S. R.; Leadley, R. J.; Dunwiddie, C. T.; Perrone, M. H.; Chu, V. Identification and Initial Structure−Activity Relationships of a Novel Class of Nonpeptide Inhibitors of Blood Coagulation. *J. Med. Chem.* **1998**, *41*, 437−450.

(7) Dominguez, C.; Duffy, D. E.; Han, Q.; Alexander, R. S.; Galemmo, R. A.; Park, J. M.; Wong, P. C.; Amparo, E. C.; Knabb, R. M.; Luettgen, J.; Wexler, R. R. Design and Synthesis of Potent and Selective 5,6-fused Heterocyclic Thrombin Inhibitors. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 925−930.

(8) Padmanabhan, K. P.; Tulinsky, A.; Park, C. H.; Bode, W.; Huber, R.; Blankenship, D. T.; Cardin, A. D.; Kiesel, W. Structure of Human Des(1−45) Factor Xa at 2.2 Å Resolution. *J. Mol. Biol.* **1993**, *232*, 947−966.

(9) Hara, T.; Yokoyama, A.; Ishihara, H.; Yokoyama, Y.; Nagahara, T.; Iwamoto, M. DX-9065a, a New Synthetic, Potent Anticoagulant and Selective Inhibitor for Factor Xa. *Thromb. Haemostasis* **1994**, *71* (3), 314−319.

(10) Hirayama, F.; Koshio, H.; Taniuchi, Y.; Sato, K.; Hisamichi, N.; Sakai, Y.; Katayama, N.; Kawasaki, T.; Matsumoto, Y.; Yanagisawa, I. *Abstracts of Papers*, 214th National Meeting of the American Chemical Society, Las Vegas, NV, 1997; American Chemical Society: Washington, DC, 1997; MEDI049.

(11) Yamazaki, M.; Asakura, H.; Aoshima, K.; Saito, M.; Jokaji, H.; Uotani, C.; Kumbashiri, I.; Morishita, E.; Ikeda, T.; Matsuda, T. Protective Effects of DX-9065a, an Orally Active Novel Synthesized and Selective Inhibitor of Factor Xa, Against Thromboplastin-Induced Experimental Disseminated Intravascular Coagulation in Rats. *Sem. Thromb. Hemostasis* **1996**, *22* (3), 255−259.

(12) Sato, K.; Taniuchi, Y.; Hirayama, T.; Koshio, H.; Matsumoto, Y.; Iizumi, Y. Comparison of the Anticoagulant and Antithrombotic Effects of YM-75466, a novel orally-Active Factor Xa Inhibitor and warfarin in Mice. *Jpn. J. Pharmacol.* **1998**, *78*, 191−197.

(13) Brandstetter, H.; Kühne, A.; Bode. W.; Huber, R.; von der Saal, W.; Wirtensohn, K.; Engh, R. A. X-ray Structure of Active Site-inhibited Clotting Factor Xa. *J. Biol. Chem.* **1996**, *271* (47), 29988−29992.

(14) Murray, C. W.; Clark, D. E.; Auton, T. R.; Firth, M. A.; Li, J.; Sykes, R. A.; Waszkowycz, B.; Westhead, D. R.; Young, S. C. PRO_SELECT: Combining structure-based drug design and combinatorial chemistry for rapid lead discovery. 1. Technology. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 193.

(15) Klebe, G. J. The Use of Composite Crystal-field Environments in Molecular recognition and the *de Novo* Design of Protein Ligands. *J. Mol. Biol.* **1994**, *237*, 212.

(16) Hahn, M. Receptor Surface Models. 1. Definition and Construction. *J. Med. Chem.* **1995**, *38*, 2080.

(17) 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: Struct., Funct., Genet.* **1998**, *33*, 367.

(18) Böhm, H.-J. The development of a simple empirical scoring function to estimate the binding constant for a protein−ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243.

(19) Eldridge, D. E.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425−445.

(20) Murray, C. W.; Auton, T. R.; Eldridge, M. D. Empirical Scoring Functions. II. The testing of an empirical scoring function for the prediction of ligand−receptor binding affinities and the use of Bayesian Regression to improve the quality of the model. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 503.

(21) Li, J.; Murray, C. W.; Waszkowycz, B.; Young, S. C. Targeted Molecular Diversity in Drug Discovery − Integration of Structure-Based design and Combinatorial Chemistry. *Drug Discovery Today* **1998**, *3* (3), 105−112.

(22) Kick, E. K.; Roe, D. C.; Skillman, A. G.; Liu, G.; Ewing, T. J. A.; Sun, Y.; Kuntz, I. D.; Ellman, J. A. Structure-based design and combinatorial chemistry yield low nanomolar inhibitors of cathepsin D. *Chem. Biol.* **1997**, *4*, 297−307.

(23) Böhm, H.-J.; Banner, D. W.; Weber, L. Combinatorial docking and combinatorial chemistry: Design of potent non-peptide thrombin inhibitors. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 51−56.

(24) Wei, A.; Alexander, R. S.; Duke, J.; Ross, H.; Rosenfeld, S. A.; Chang, C.-H. Unexpected Binding Mode of Tick Anticoagulant Peptide Complexed to Bovine Factor Xa. *J. Mol. Biol.* **1998,** *283*, 147−154.

(25) The strain energies quoted here are generally much higher than the associated calculated binding energy. This is because they are calculated by different methods, the strain energy arising out of estimates, derived using the 'Clean' force field. Therefore they cannot be compared and are used independently from one another in the process of ranking the substituents.

(26) Copyright 1995, BIOSYM/Molecular Simulations, San Diego.

(27) Copyright 1990−1994, MDL Information Systems, Inc. San Leandro, CA. All Rights Reserved.

(28) MDL Information Systems, Inc. San Leandro, CA. All Rights Reserved.

(29) Tapparelli, C.; Metternich, R.; Ehrardt, C.; Zurini, M.; Claeson, G.; Scully, M. F.; Stone, S. R. In Vitro and In Vivo Characterization of a Neutral Boron-containing Thrombin Inhibitor. *J. Biol. Chem.* **1993**, 268, 4734−4741.

(30) Bartunik, H. D.; Summers, L. J.; Bartsch, H. H. Crystal structure of bovine b-trypsin at 1.5 Å resolution in a crystal form with low molecular packing density. *J. Mol. Biol.* **1989**, *210*, 813−828.

(31) Otwinoski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1996**, *276*, 307−326.

(32) Collaborative Computational Project, Number 4. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr.* **1994**, *D50*, 760−763.

(33) Brunger, A. T. 1992 X-PLOR Manual Version 3.1.

(34) Jones, T. A.; Zou, J.-Y.; Cowan, S. W.; Kjeldgaard, M. Improved methods for building protein structures in electron-density maps and the location of errors in these models. *Acta Crystallogr.* **1991**, *A47*, 110−119.

JM010944E