Statistical Molecular Design, Parallel Synthesis, and Biological Evaluation of a Library of Thrombin Inhibitors

Anna Linusson,*,† Johan Gottfries,† Thomas Olsson,† Eivor Örnskov,† Staffan Folestad,† Bo Nordén,‡ and Svante Wold§

AstraZeneca R&D Mölndal, S-431 83 Mölndal, Sweden, AstraZeneca R&D Lund, Box 34, S-221 00 Lund, Sweden, and Research Group for Chemometrics, Umeå University, S-901 87 Umeå, Sweden

Received January 30, 2001

A library of thrombin inhibitors has been designed using statistical molecular design. An aromatic scaffold was used, with three varied positions corresponding to three pockets at the active site of thrombin (the S-, P-, and D-pockets). The selection was performed in the building block space, and previously acquired data were included in the design procedure. The design resulted in six, four, and six building blocks for the first (S), second (P), and third (D) pockets, respectively. A second round of selection applied to the combined selected building blocks resulted in a subset of 18 compounds. The selected library was synthesized in parallel and biologically evaluated. The compounds were analyzed with respect to their inhibition (pIC_{50}) of thrombin; membrane permeability, estimated by migration behavior in micellar media (CE log k') and pK_a; and specificity with respect to inhibition (K_i) of trypsin. Multivariate QSAR studies of the responses yielded valuable results and information that could only be found using statistical molecular design in combination with multivariate analysis.

Introduction

Thrombosis is the result of improper regulation of the hemostasis mechanisms, leading to the formation of intravascular clots, which may cause tissue damage or cell death due to inadequate blood flow. This cardiovascular disorder can lead to deep vein thrombosis, myocardial infarction, and stroke. Much current drug research is focused on finding an antithrombotic drug that is safe and effective and can be administered orally. One prominent target for such a drug is the enzyme thrombin (factor IIa). Thrombin plays a central role in the blood coagulation process as the final key enzyme in the cascade involved. It converts soluble fibrinogen to fibrin, which forms the fibrillar matrix of the blood clot. In addition, the activation of thrombin initiates several other mechanisms including both positive and negative feedback of thrombin generation,[1] together with a variety of cellular effects such as platelet aggregation and tissue remodeling.[2,3] A low molecular weight substance that can inhibit thrombin, and thus its actions, would be a potentially powerful antithrombotic drug. Substances that bind to the active site have been successfully developed, and several inhibitors have been described.[4−13] However, finding an inhibitor that comprise both selectivity and suitable pharmacokinetics has been difficult to identify, prompting the research to continue.[14−17]

In the drug discovery process combinatorial chemistry is routinely used as a complement to sequential synthesis in medicinal chemistry. This automatic approach offers the possibility of synthesizing large arrays of compounds by combining sets of building blocks. However, even though combinatorial chemistry expands synthetic capacity, the number of compounds that can be synthesized is still limited in practice. Along with increases in the number of building blocks and positions varied, the number of possible combinations accelerates rapidly, and selection is often essential.

The features of combinatorial chemistry, allowing many compounds to be synthesized in parallel, together with today's computer capacity provide an excellent basis for statistical selection of library compounds. A smaller subset of molecules can be selected using statistical experimental techniques, i.e., "statistical molecular design" (SMD), without losing information contained in the full library.[18−20] The objectives of a given drug design project, and relevant prior knowledge, should be the focus of the statistical design and thus steer the selection process. It is also important to consider synthetic and financial aspects in the early stages of the selection procedure.

In library design, there is a choice between basing the design either on the building blocks or on the enumerated products.[18,19,21,22] The former has several advantages;[23] e.g., the characterization of physicochemical properties will be made on a smaller set of compounds and the number of reactants will be reduced in a straightforward way. The evaluation process will also be more direct, indicating not only what molecular properties should be varied but also in what part of the molecule appropriate changes should be introduced.

Together with the design of the library of chemical structures, it is equally important to design a battery of biological tests for investigating the structures. The evaluation even at an early stage (preclinical) should include biologically relevant tests with a broad pharmacological profile, so the number of compounds to be discarded in expensive and time-consuming clinical research phases can be minimized.

* Corresponding author. Tel: + 46-(0)31-706 52 07. Fax: + 46-(0)-31-776 37 10. E-mail: anna.linusson@astrazeneca.com.
† AstraZeneca R&D Mölndal.
‡ AstraZeneca R&D Lund.
§ Umeå University.

10.1021/jm010833f CCC: $20.00
Published on Web 09/11/2001

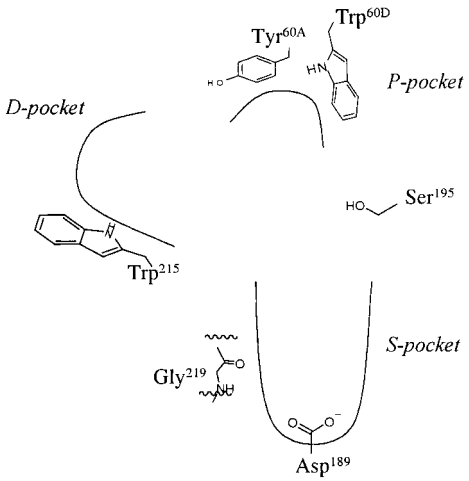


Figure 1. Schematic representation of the active site of thrombin.

The theoretical aspects of library design have been discussed in several papers,[24,25] but little real data have so far been presented and evaluated. In the present work a statistically designed library of thrombin inhibitors has been synthesized and evaluated through biological testing and subsequent structure−activity studies. The design was performed in the building block space, and since thrombin is a well-known enzyme, both previously acquired results ("in-house" data) and the crystal structure were used in the selection procedure. The selected compounds were synthesized in parallel in solution, and the bioassessments included their inhibition of thrombin, their specificity toward thrombin with respect to trypsin, and their tendency to be passively absorbed. The resulting designed library and response values were evaluated using multivariate methods. In addition, we discuss the use of the present approach for any size of medicinal compound series.

Material and Methods

The Active Site of Thrombin. α-Thrombin is a serine protease and a member of the trypsin family. Serine proteases are proteolytic enzymes that all contain the characteristic serine-histidine-aspartate catalytic triad [Ser195, His57, and Asp102 in thrombin (Bode numbering[26])]. The hydroxyl group of the serine acts as a nucleophile, targeting the scissile amide bond of the substrate. There are three principal binding pockets at the cleaving site, the specificity (S) pocket, the proximal (P) pocket, and the distal (D) pocket, modeled from the crystal structure.[26,27] The sides of the S-pocket in thrombin are primarily hydrophobic and at the base of the site there are a carboxylic acid (Asp^{189}) and two backbone carbonyls. In fibrinogen an arginine residue (P1) coordinates into the S-pocket, forming an ion−ion interaction with the Asp^{189}. The P-pocket is hydrophobic, smaller than the S-pocket, and binds the P2 amino acid valine of fibrinogen. The D-pocket is also hydrophobic, larger than the P-pocket, and not so well defined. This pocket is also called the aryl-binding site and it is occupied by a phenylalanine in fibrinogen. A schematic representation of the active site is shown in Figure 1. In addition to the active site, thrombin has several other recognition and binding sites,[28] but these will not be discussed here.

Characterization of Structures. Descriptors from both 2D and 3D structures were used. The 2D-based descriptors were calculated using in-house programmed calculations.[29] These descriptors include properties of size, lipophilicity, polarizability, charge, flexibility, rigidity, and hydrogen-binding capacities, which cover the relevant latent property structures.[29−31] The 3D structures were generated using Concord,[32] and the characterization was made using Spartan software, with the semiempirical AM1.[33]

D-Optimal Design. D-optimal designs maximize det(**X′X**), where **X** is the coded model matrix with n rows (structures) and p terms (model variables) and det denotes the determinant. Here, these designs were created to support a model relating the properties of the compounds and their biological response, $\mathbf{y} = \mathbf{X}^*\mathbf{b} + \mathbf{e}$, where **y** is a column with n responses, **b** is a row vector containing the coefficients, and **e** is the residuals. By maximizing the volume of **X′X** (the determinant) the error variance of the coefficients **b** will be minimized.[34] Discrete variables and scaled scores were used as design variables. The D-optimal designs were augmented by the molecule closest to the calculated centroid. When the final library was selected, an additional design criterion that every building block should be represented at least twice in the selected set was used along with the criteria set by the design program.[35] The D-optimal designs were generated by MODDE software.[35]

Data Analytic Methods. Two data analytical methods were used: principal component analysis (PCA) for compression (summarization) of data sets and partial least squares projection to latent structures (PLS) to relate a structure descriptor matrix, **X**, to an activity matrix, **Y**.

The number of significant principal properties of molecules is often much smaller than the number of structural descriptors. PCA compresses a data set to its main "principal" structure, i.e., reexpresses the information in the structural descriptors using a smaller number of new uncorrelated variables. PCA is expressed in terms of scores (**T**) and loadings (**P**), where $\mathbf{X} = \mathbf{1}^*\bar{\mathbf{x}} + \mathbf{T}^*\mathbf{P}' + \mathbf{E}$, and $\bar{\mathbf{x}}$ is a row vector of variable means and **E** is the residual matrix.[36] The number of principal components was here decided using their eigenvalues and the interpretation of the loadings.

PLS was used for multivariate quantitative structure−activity relationship (MQSAR) analysis of both the historical data and of the library we synthesized and biologically tested. PLS is a latent variable regression method which calculates a model taking into account the variance (in **X**) together with the correlation between **X** and **Y**; i.e., it finds latent structures in **X** that correlate with latent structures in **Y**.[37,38]

The results from PLS can be viewed in score and loading plots. These plots were used to interpret the model, along with the model coefficients for the variables (structural descriptors). The models were validated by investigation of the explained variation in **Y**, R^2Y, and the predictive capability, Q^2 (cross-validation, in five rounds[39,40]). The data set was scaled to zero mean and unit variance prior to the analyses. All calculations were made using SIMCA software.[41]

Biological Evaluation. The synthesized compounds (see Results and Discussion) were subjected to biological

analysis to investigate their capacity to inhibit thrombin (pIC_{50}), their membrane permeability according to their migration behavior in micellar media (CE log k') and pK_a, and their inhibition (K_i) of trypsin.

Due to the lipophilic character of the library compounds and because these model compounds were sparingly soluble in buffer, capillary electrophoreses (CE) was selected as the technique for investigating membrane permeability. In addition, CE is also more straightforward than other in vitro techniques, such as Caco-2 cell permeation, when only minute amounts of substance are available. Further, the use of micelles as separation media in capillary electrophoreses (CE) is a means to experimentally model absorption to a lipid-like pseudostationary phase. Here, the bile acid sodium cholate (SC), a native constituent of intestinal fluids, was used as the micellar electrolyte. The rationale for this so-called micellar electrokinetic chromatography (MEKC) method is that separations are mainly due to differences in molecular size, hydrophobicity, and hydrogen bonding, all of which have a strong influence on in vivo absorption. Notably, correlation has recently been demonstrated between MEKC data and log P,[42] and between MEKC data and in vitro absorption.[43,44] In an MEKC system, micelles are amphiphilic aggregates with an anisotropic microenvironment that provides both hydrophobic and electrostatic sites of interactions. In this way, the micellar phase is structurally more similar to biomembranes than 1-octanol/water partitioning, commonly used for lipophilicity determinations (log P). Moreover, Woodrow and Dorsey[45] have reported that the thermodynamic signature of micelle−water partitioning, as used in MEKC, is similar to biological partitioning.

Because only a minute amount of sample was available in this study and the thrombin inhibitor compounds were degraded at neutral to high pH, common techniques for pK_a determination could not be used. Here, CE enabled determination of pK_a values for the 11 compounds containing a basic moiety, although these were partially degraded. During CE separation, degradation products could be separated from the respective compound.[46]

Results and Discussion

Scaffold. An achiral scaffold was selected for the library (Figure 2), to which three building blocks were connected that coordinate to the S-pocket (S-residue), the P-pocket (P-residue) and the D-pocket (D-residue). A number of compounds containing this scaffold have previously been synthesized and tested for thrombin inhibition,[47] giving the opportunity to include historical data in the design process. Some of these analogues have also been cocrystallized with thrombin, confirming that they bind in the assumed way.[47]

Historical Data. One hundred and five previously synthesized compounds comprising the overall structure showed in Figure 2a[47] were subjected to MQSAR analysis. In this library the S-residue was held constant, while the P-residue (R'_1) was varied with three building blocks, methyl-, chlorine- and methoxy groups, each having 35 different D-residues (R'_2, 35 × 3 compounds). The 35 groups at the D-residue were not varied in a balanced way. The frameworks could be divided into five

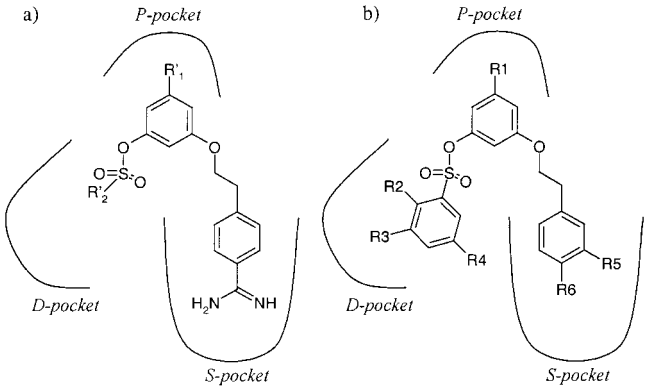


Figure 2. The overall structure of the compounds included in the MQSAR of the historical data (a) and the common substructure of the compounds that were designed and synthesized in the present study (b). The substituents for the previously synthesized compounds (a) were $-CH_3$, $-OCH_3$, $-Cl$ for R′1, with 35 different building blocks for R′2, and the substituents chosen for synthesis (b) were $-H$, $-CH_3$, and $-CH_2CH_3$ for R1; $-H$, $-OCH_3$, and $-Cl$ for R2; $-H$, $-CH_3$, and $-Cl$ for R3; $-H$ and $-Cl$ for R4; $-H$ and $-OH$ for R5; and $-H$, $-CH_3NH_2$, and $-CNHNH_2$ for R6.

groups (without regard to the substituents) consisting of 8 phenyls, 12 hetereoaromatics, 2 aliphatics, 5 bicyclic hetereoaromatics, and 8 bicyclic phenyl−phenyls.

The 105 compounds were described by 53 structural descriptors, including a range of indices, atom counts, and electronic, lipophilicity, polarizability, and size parameters, using an "in-house" program.[23,29,31] In addition, binary descriptors were used both to indicate the three substituents at the P-position (methyl, chloride, and methoxy) and to describe the five groups represented at the D-position (phenyl, hetereoaromatic, aliphatic, bicyclic hetereoaromatic-phenyl, and bicyclic phenyl−phenyl).

MQSAR modeling of these data, using the structural descriptors (as the **X**-block) and the inhibition of thrombin (pIC_{50}) as the response, resulted in a significant model with $R^2Y = 0.68$, $Q^2 = 0.51$ (four components, using linear terms plus square terms of the continuous descriptors). The results showed that the methyl group at the P-residue gave the strongest response. It is important to note that the building blocks for the D-residue were not selected according to a statistical design, making interpretation of these results somewhat uncertain. However, the outcome indicated that a benzene ring or a single hetereoaromatic ring might be preferable at this position.

Building Block Selection. Representative sets of building blocks were selected separately for the three residues shown in Figure 2, i.e., the S-, P-, and *D*-residue. A formal statistical design was used only in the selection of the D building blocks.

The S-Residue. In the previously synthesized library the S-residue was held constant, using a benzamidine (at position 4) to coordinate to the carboxylic acid in Asp[189]. It is well-known that this group effectively inhibits thrombin due to the ion−ion interaction it induces.[7] However, it is also well-known that the amidine group, due to its basicity, will be protonated at pH 7, and hence, the rate of its passive absorption in the body may be low.[15−17] Therefore, one objective of the present study was to investigate if the basicity could be decreased while the inhibitory effect was maintained.

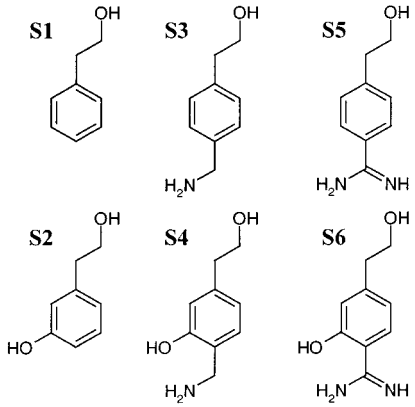


Figure 3. The six building blocks selected for the S-pocket.

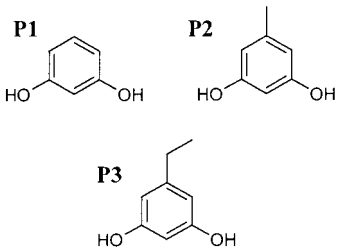


Figure 4. The three building blocks selected for the P-pocket.

A phenyl framework was selected, with two positions to be varied (R5 and R6). At position 4 (R6) the following substituents were selected: −H, $-CH_2NH_2$ (primary amine), and $-CHNHNH_2$ (amidine). When the hydrogen is used at position 4, the building block will lack a basic moiety. The primary amine was selected to introduce a group that was basic, but not as strongly basic as the amidine moiety. The amidine group was included in the design as a reference. In addition, the substituents −H and −OH were investigated at position 3 (R5). A hydroxy group at position 3 may serve as a hydrogen donor to the basic groups at position 4 and thereby decrease pK_a for the protonated forms. The objective was to see if this could enhance the tendency of the compounds to be passively absorbed. Since interaction effects between positions 3 and 4 were expected, all combinations of these five substituents were selected as building blocks for the *S*-residue (Figure 3).

Among these six building blocks, S1 and S2 were commercially accessible, the other four were synthesized.

The P-Residue. The MQSAR of the historical data revealed that a polar group like methoxy or an electronegative group like chlorine was not favorable at this position (Figure 2a, R′1). Consequently, only aliphatic residues were selected in this design. An inspection of the crystal structure indicated that this pocket was relatively small. Therefore, in this investigation we selected substituents with a small size range, −H, $-CH_3$, and $-CH_2CH_3$ (Figure 4).

Two of these three building blocks were commercially available, while P3 was synthesized.

The D-Residue. The MQSAR model of the historical data indicated that a benzene or hetereoaromatic ring was favorable at this position. Even though the model might be somewhat uncertain due to the unbalanced data, these results were used in the design. Investigation of commercially available chemicals and possible synthetic routes showed the phenyl moiety to be advantageous compared to the hetereoaromatics, and hence, a benzene ring was selected as the framework (Figure 2b).

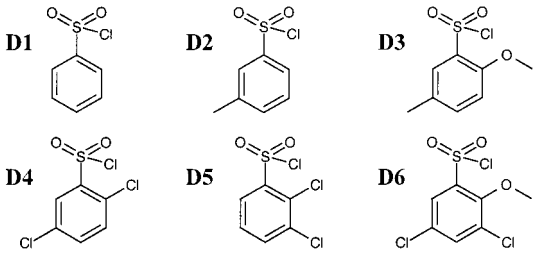


Figure 5. The six building blocks selected for the D-pocket.

The positions 2, 3, and 5 were selected for investigation (R2, R3, and R4), with the following substituents: R2, −H, $-OCH_3$, and −Cl; R3, −H, and −Cl; and R4, −H, $-CH_3$ and −Cl. This choice was partly based on the numerous commercially available chemicals with the listed substituents.

The numbers of selected substituents led to $3 \times 2 \times 3 = 18$ building blocks for the D-residue. Not all combinations were necessary for a screening analysis, so a D-optimal design was applied using discrete design variables (indicator variables) for the substituents. Figure 5 shows the six selected compounds.

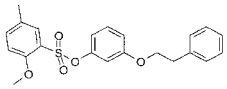
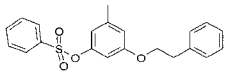
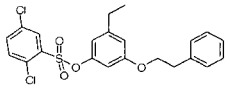
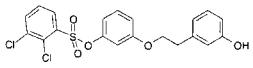
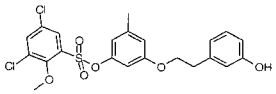
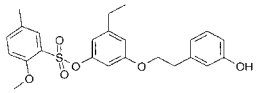
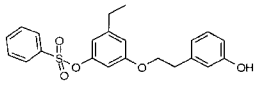
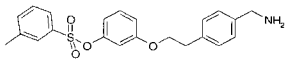
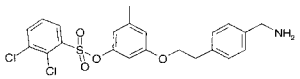
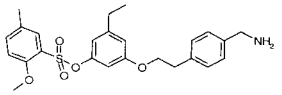
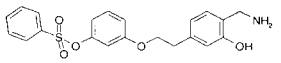
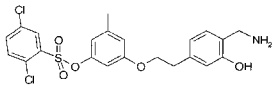
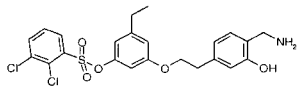
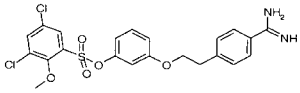
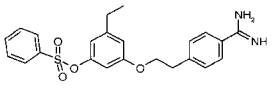
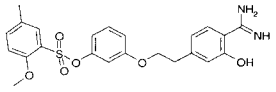
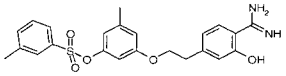
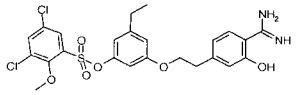
Five out of these six were commercially available; D6 was synthesized.

The Final Library Selection. A full factorial design (all combinations) of the selected building blocks ($6 \times 3 \times 6$) would result in 108 compounds. However, a fractional sampling of these should theoretically be sufficient to investigate their structural space,[19,48] and hence, a smaller subset was selected, based on a multivariate characterization of the compounds.

The characterization was done in the building block space.[23] The selected building blocks ($6 + 3 + 6 = 15$) were described individually using semiempirical calculations (AM1), HOMO, LUMO, polarizability, molecular volume, surface area, log *P*, dipole moment, maximum charge (Mulliken), and minimum charge (Mulliken).[33] In these calculations a methyl group was used to replace the sulfonyl chloride substituent in the reactant set D and the ethanol substituent in the reactant set S. In addition, the molecular weight, the number of hydrogen-bond donors, and acceptors for the S-residue and binary descriptors for the substituent patterns for residues S and D were used. Each of the three building block sets (S, P, and D) was compressed individually by PCA. In total, five principal properties (PCA scores) were used as design variables, two for the residues S and D and one for the P residue. These were combined to represent the properties of the 108 candidate compounds. The D-optimal design (linear and quadratic terms) procedure yielded six different selections of 18 compounds that were approved according to the design criteria (see Material and Methods). From those approved selections the final choice was subjectively made on the basis of medicinal experience, and Table 1 shows the 18 compounds that were finally synthesized.

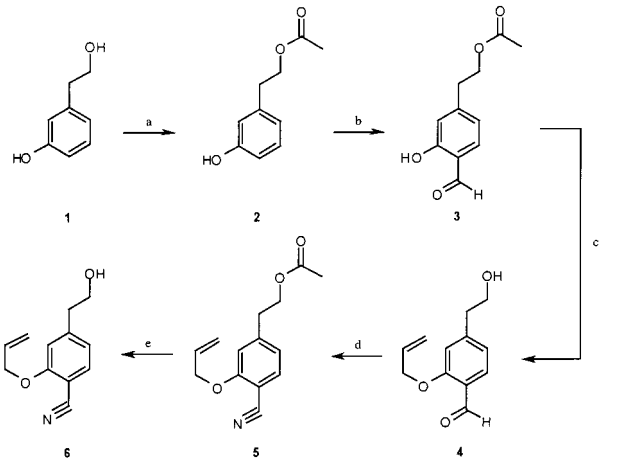
Synthesis of Noncommercial S-Building Blocks. Both the amidine and amine groups of building blocks S3−S6 were synthesized from the corresponding nitriles after synthesizing the library. The nitrile derivative

Table 1. Chemical Structure and Biological Measurement Values for the 18 Selected Compounds

No.	Chemical structure	pIC50 (Thrombin)	pKi Calc.* (Thrombin)	pKi (Trypsin)	CE log k'	pKa
30		-	-	-	1.40	-
36		-	-	-	1.51	-
42		-	-	-	1.85	-
27b		-	-	-	1.41	-
33b		4.25	4.53	-	1.57	-
38b		-	-	-	1.30	-
39b		-	-	-	1.42	-
31b		4.23	4.51	-	0.89	9.4
37b		5.91	6.19	-	0.97	9.0
43b		5.50	5.78	5.21	0.87	9.3
28c		4.97	5.25	-	0.85	8.3
34c		6.03	6.31	-	0.98	8.6
40c		4.95	5.23	-	1.08	8.3
32b		7.25	7.53	6.58	0.96	10.5
44b		7.60	7.88	-	0.88	10.9
29c		6.24	6.52	4.55	1.08	6.2
35c		6.87	7.15	4.49	1.16	6.3
41c		6.54	6.82	-	1.27	6.3

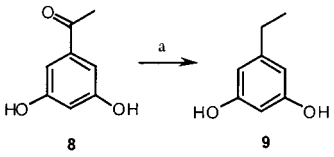
*$K_i = IC_{50}/(1 + [S]/K_m)$; [S] = 0.3 mM, K_m = 0.33 mM.

Scheme 1[a]



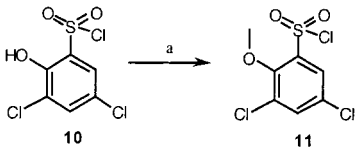
[a] Reagents and conditions: (a) EtOAc, hexane, $NaHSO_4 \cdot SiO_2$, reflux; (b) (1) CH_3CH_2MgBr, Et_2O, 0 °C/rt, (2) paraformaldehyde, Et_3N, benzene, reflux; (c) $CHCl_3$, MeOH, K_2CO_2, CH_2CHCH_2Br, reflux; (d) $CH_3CH_2NO_2$, NaOAc, AcOH, reflux; (e) K_2CO_3, MeOH, reflux.

Scheme 2[a]



[a] Reagents and conditions: (a) Pd/C, H_2, 4% HCl, rt.

Scheme 3[a]



[a] Reagents and conditions: (a) $(CH_3)_3SiCHN_2$, 60% HBF_4, CH_2Cl_2, 0 °C.

yielding building blocks S4 and S6, 2-allyloxy-4-(2-hydroxyethyl)benzonitrile (**6**), was synthesized in five steps (Scheme 1). Selective protection of the primary alcohol of 3-(2-hydroxyethyl)phenol (**1**) gave **2**. Formylation of **2** via the magnesium bromine salt of the phenol resulted exclusively in the *o*-hydroxy aldehyde **3**. The phenol of **3** was protected as the allyl ether **4** and the aldehyde was converted into the nitrile **5** through a one-pot synthesis using nitroethane and sodium acetate in acetic acid.[49] Finally, the primary alcohol was deprotected, yielding the wanted nitrile derivative **6**. In addition, the phenol of **1** was protected as the allyl ether **7** in a manner similar to that used for **3** (to yield building block S2). The nitrile derivative yielding building blocks S3 and S5 was commercially available.

Synthesis of Noncommercial P- and D-Building Blocks. Reduction of 1-(3,5-dihydroxyphenyl)ethanone (**8**) yielded **9** (P3, Scheme 2), and 3,5-dichloro-2-hydroxybenzenesulfonyl chloride (**10**) was converted into **11** through methylation of the hydroxy group using trimethylsilane diazomethane (D6, Scheme 3).

Synthesis of the Library. The library was synthesized in two main steps and five in total (Scheme 4, Table 2). In the first step the P-residue was connected with the D-residue. Mixing the two building blocks in a two-phase system consisting of diethyl ether and saturated sodium hydrogen carbonate yielded the mono sulfonic acid esters **12**−**26**. Increasing the equivalents of 1,3-dihydroxy benzene derivative compared to the sulfonyl chloride resulted in higher yields of the mono-substituted product. In the case of **22**−**26**, this was necessary to obtain any product at all. In the second step, the S-residue was connected to the sulfonic acid esters **12**−**26** through a Mitsunobu reaction, using diethylazodicarboxylate (DEAD) and triphenyl phosphine in THF, to give **27a**−**44a**. The allyl ethers **27a**−**29a**, **33a**−**35a**, and **38a**−**41a** were cleaved reductively with tetrakis($Pd(PPh_3)_4$) and sodium borohydride in THF, yielding **27b**−**29b**, **33b**−**35b**, and **38b**−**41b**. Reduction of the nitriles **31a**, **37a**, **43a**, **28b**, **34b**, and **40b** using borane−dimethyl sulfide complex in THF gave the amines **31b**, **37b**, **43b**, **28c**, **34c**, and **40c**. Finally, the amidines **32b**, **44b**, **29c**, **35c**, and **41c** were synthesized through Pinner reaction of **32a**, **44a**, **29b**, **35b**, and **41b** in ethanol saturated with hydrogen chloride and subsequent treatment with ammonium-saturated ethanol. In the synthesis of the library, purity was favored over yield. The chemical structures and the purity of the final library were verified and determined by high-resolution mass spectroscopy (HRMS) and by proton and carbon nuclear magnetic resonance (^1H NMR, ^{13}C NMR). Two of the final compounds, **36** and **42**, did not ionize during the HRMS measurements, and hence, no masses were recorded for these substances.

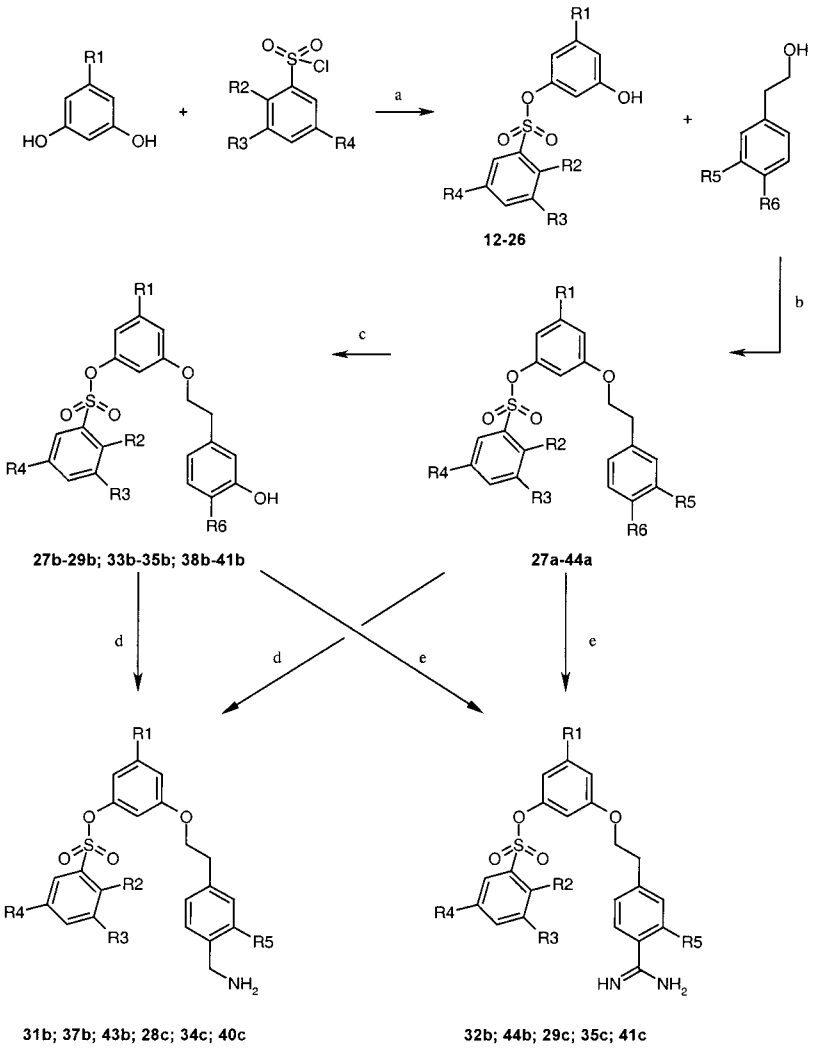
pIC_{50} of Thrombin and CE log *k′*. Twelve of the 18 compounds gave a 50% inhibition of thrombin at concentrations ≤133 μM. Eleven of them contained a basic S-moiety, allowing the possibility of coordination to Asp^{189} in the enzyme. The twelfth, which included a S2 S-residue, was only marginally active ($pIC_{50} = 4.25$).

The 11 compounds containing a basic S-moiety were further used for MQSAR modeling, using two responses, pIC_{50} for the enzyme activity and CE log *k′*, as an estimate of membrane permeability. In this set of 11 compounds the building blocks selected for positions S, P, and D were represented in a balanced way.

A modification of the design descriptors was performed to make the MQSAR easier to interpret (Table 3). Using the same characterization as for the design (semiempirical, AM1[33]), the descriptors were divided into four blocks, according to whether they were based on size, electronic, hydrophobic, or polarizability parameters, for each of the varied positions (S, D, and P). The size and electronic blocks each included more than one descriptor, and local PCAs were used to reduce the dimensions (for each building block set, S, D, and P). The compression resulted in one size descriptor for each building block set and two electronic properties for the S- and D-sets, while the P-set only yielded one.

In addition, to allow the effects of the different substituents at the D-residues to be interpreted, a semiempirical characterization (AM1) was undertaken for each of the three varied positions. These structural descriptors were treated in a manner similar to that described above, yielding four properties (based on size, electronic, hydrophobic, and polarizability characteris-

Scheme 4[a]



[a] Reagents and conditions: (a) Et_2O, aq $NaHCO_3$, rt; (b) PPh_3, DEAD, THF, rt (c) $Pd(PPh_3)_4$, $NaBH_4$, THF, rt; (d) borane−Me_2S, THF, reflux/rt; (e) (1) EtOH, HCl, rt, (2) EtOH, NH_3, rt. The R-groups are displayed in Table 2.

tics) for each of the three positions (2, 3, and 5) at the aromatic ring.

Two end product descriptors were added, log D (pH = 7.4) and log *P*,[50] where the compounds having the *o*-hydroxy amidine were left with missing values, due to the inaccurate log *P* estimates. In addition, the measured pK_a values were included in the model.

To investigate interaction effects, indicator descriptors were added to describe building blocks S and P. The groups in the positions at the S-residue (R_1) were distinguished by two pairs of variables; at the para position −1 was set for the primary amine and +1 for the benzamidine, and at the meta position −1 for the hydrogen and +1 for the hydroxy group. The P-residue was varied with −H, $-CH_3$, and $-CH_2CH_3$ substituents, represented by −1, 0, and, +1, respectively. The substitution pattern for the D-residue was more complex than for the other two and would have generated many binary descriptors. Therefore, the five principal property vectors were used to describe this building block. The structural descriptors used are listed in Table 3 (nos. 1−31).

A PLS using structural descriptors nos. 1−31, square terms of the continuous descriptors (4−31), and two-way interaction terms between nos. 1−3 and 12−16 (Table 3), with pIC_{50} and CE log k' as responses (K_X = 87, M_Y = 2), gave four significant components, explaining 98% of the total variation in **Y**, and a Q^2 of 0.53 (for pIC_{50}, R^2Y = 0.97 and Q^2 = 0.55; for CE log k', R^2Y = 0.98 and Q^2 = 0.30). The relationship between the structural properties and the enzyme activity was nonlinear, as indicated by the large coefficients for the quadratic term for pK_a and for interaction terms between the P-residue and the D-residue. Therefore, a more careful investigation was made of the interaction effects between the properties of the varied positions at the D-residue and the properties of the P-residue. A new MQSAR (K_X = 49, M_Y = 2) was calculated, expanded by the inclusion of two-way interaction terms between no. 1 and nos. 17−28, while the main terms (1−31), the cross terms between no. 1 and nos. 12−16 (Table 3) and the quadratic term of pK_a (no 31) were retained. A refinement of the MQSAR excluding all nonlinear terms with model coefficients of values in the range between −0.05 and 0.05 (K_X = 41, M_Y = 2) gave a model explaining 94% of the total variation and a Q^2 of 0.74 (three PLS components: pIC_{50}, R^2Y = 0.91, and Q^2 = 0.56; CE log k', R^2Y = 0.97 and Q^2 = 0.87). The terms included in the refined model are listed in Table

Table 2. Key to the R-groups in Scheme 4

no.	substituents	no.	substituents
12	R1 = H, R2 = Cl, R3 = Cl, R4 = H	**39a**	R1 = Et, R2 = H, R3 = H, R4 = H, R5 = OCH_2CHCH_2, R6 = H
13	R1 = H, R2 = H, R3 = H, R4 = H	**40a**	R1 = Et, R2 = Cl, R3 = Cl, R4 = H, R5 = OCH_2CHCH_2, R6 = CN
14	R1 = H, R2 = OMe, R3 = H, R4 = Me	**41a**	R1 = Et, R2 = OMe, R3 = Cl, R4 = Cl, R5 = OCH_2CHCH_2, R6 = CN
15	R1 = H, R2 = H, R3 = Me, R4 = H	**42**	R1 = Et, R2 = Cl, R3 = H, R4 = Cl, R5 = H, R6 = H
16	R1 = H, R2 = OMe, R3 = Cl, R4 = Cl	**43a**	R1 = Et, R2 = OMe, R3 = H, R4 = Me, R5 = H, R6 = CN
17	R1 = Me, R2 = OMe, R3 = Cl, R4 = Cl	**44a**	R1 = Et, R2 = H, R3 = H, R5 = H, R6 = CN
18	R1 = Me, R2 = Cl, R3 = H, R4 = Cl	**27b**	R1 = H, R2 = Cl, R3 = Cl, R4 = H, R6 = H
19	R1 = Me, R2 = H, R3 = Me, R4 = H	**28b**	R1 = H, R2 = H, R3 = H, F4 = H, R6 = CN
20	R1 = Me, R2 = H, R3 = H, R4 = H	**29b**	R1 = H, R2 = OMe, R3 = H, R4 = Me, R6 = CN
21	R1 = Me, R2 = Cl, R3 = Cl, R4 = H	**33b**	R1 = Me, R2 = OMe, R3 = Cl, R4 = Cl, R6 = H
22	R1 = Et, R2 = OMe, R3 = H, R4 = Me	**34b**	R1 = Me, R2 = Cl, R3 = H, R4 = Cl, R6 = CN
23	R1 = Et, R2 = H, R3 = H, R4 = H	**35b**	R1 = Me, R2 = H, R3 = Me, R4 = H, R6 = CN
24	R1 = Et, R2 = Cl, R3 = Cl, R4 = H	**38b**	R1 = Et, R2 = OMe, R3 = H, R4 = Me, R6 = H
25	R1 = Et, R2 = OMe, R3 = Cl, R4 = Cl	**39b**	R1 = Et, R2 = H, R3 = H, R4 = H, R6 = H
26	R1 = Et, R2 = Cl, R3 = H, R4 = Cl	**40b**	R1 = Et, R2 = Cl, R3 = Cl, R4 = H, R6 = CN
27a	R1 = H, R2 = Cl, R3−Cl, R4 = H, R5 = OCH_2CHCH_2, R6 = H	**41b**	R1 = Et, R2 = OMe, R3 = Cl, R4 = Cl, R6 = CN
28a	R1 = H, R2 = H, R3 = H, R4 = H, R5 = OCH_2CHCH_2, R6 = CN	**31b**	R1 = H, R2 = H, R3 = Me, R4 = H, R5 = H
29a	R1 = H, R2 = OMe, R3 = H, R4 = Me, R5 = OCH_2CHCH_2, R6 = CN	**37b**	R1 = Me, R2 = Cl, R3 = Cl, R4 = H, R5 = H
30	R1 = H, R2 = OMe, R3 = H, R4 = Me, R5 = G, R6 = H	**43b**	R1 = Et, R2 = OMe, R3 = H, R4 = Me, R5 = H
31a	R1 = H, R2 = H, R3 = Me, R4 = H, R5 = H, R6 = CN	**28c**	R1 = H, R2 = H, R3 = H, R4 = H, R5 = H
32a	R1 = H, R2 = OMe, R3 = Cl, R4 = Cl, R5 = H, R6 = CN	**34c**	R1 = Me, R2 = Cl, R3 = H, R4 = Cl, R5 = OH
33a	R1 = Me, R2 = OMe, R3 = Cl, R4 = Cl, R5 = OCH_2CHCH_2, R6 = H	**40c**	R1 = Et, R2 = Cl, R3 = Cl, R4 = H, R5 = OH
34a	R1 = Me, R2 = Cl, R3 = H, R4 = Cl, R5 = OCH_2CHCH_2, R6 = CN	**32b**	R1 = H, R2 = OMe, R3 = Cl, R4 = Cl, R5 = H
35a	R1 = Me, R2 = H, R3 = Me, R4 = H, R5 = OCH_2CHCH_2, R6 = CN	**44b**	R1 = Et, R2 = H, R3 = Cl, R4 = Cl, R5 = H
36	R1 = Me, R2 = H, R3 = H, R4 = H, R5 = H, R6 = H	**29c**	R1 = H, R2 = OMe, R3 = H, R4 = Me, R5 = OH
37a	R1 = Me, R2 = Cl, R3 = Cl, R4 = H, R5 = H, R6 = CN	**35c**	R1 = Me, R2 = H, R3 = Me, R4 = H, R5 = OH
38a	R1 = Et, R2 = OMe, R3 = H, R4 = Me, R5 = OCH_2CHCH_2, R6 = H	**41c**	R1 = Et, R2 = OMe, R3 = Cl, R4 = Cl, R5 = OH

Table 3. Terms Used in the Refined MQSAR Model

no.	abbrev	explanation	position	model term	no.	abbrev	explanation	position	model term
1	P	discrete variable: −1, 0, +1	P	linear	22	3_size	PCA score, size	D pos 3	linear
2	BA/A	discrete variable: −1, +1	S	linear	23	3_logP	lipophilicity	D pos 3	linear
3	OH/H	discrete variable: −1, +1	S	linear	24	3_pol	polarizability	D pos 3	linear
4	P_elec	PCA score, electronic properties	P	linear	25	5_elec	PCA score, electronic properties	D pos 5	linear
5	P_size	PCA score, size	P	linear	26	5_size	PCA score, size	D pos 5	linear
6	P_pol	polarizability	P	linear	27	5_logP	lipophilicity	D pos 5	linear
7	P_logP	lipophilicity	P	linear	28	5_pol	polarizability	D pos 5	linear
8	S_elec1	PCA score, electronic proeprties 1	S	linear	29	logD	lipophilicity	global	linear
9	S_elec2	PCA score, electronic properties 2	S	linear	30	logP	lipophilicity	global	linear
10	S_size	PCA score, size	S	linear	31	pK_a	basicity	global	linear
11	S_pol	polarizability	S	linear	32	S31×31	pK_a × pK_a	global	quadratic
12	D_elec1	PCA score, electronic properties 1	D	linear	33	C1×12	P × D_elec1	P × D	interaction
13	D_elec2	PCA score, electronic properties 2	D	linear	34	C1×14	P × D_size	P × D	interaction
14	D_size	PCA score, size	D	linear	35	C1×15	P × D_pol	P × D	interaction
15	D_pol	polarizability	D	linear	36	C1×16	P × D_logP	P × D	interaction
16	D_logP	lipophilicity	D	linear	37	C1×17	P × 2_elec	P × D	interaction
17	2_elec	PCA score, electronic properties	D pos 2	linear	38	C1×18	P × 2_size	P × D	interaction
18	2_size	PCA score, size	D pos 2	linear	39	C1×19	P × 2_logP	P × D	interaction
19	2_logP	lipophilicity	D pos 2	linear	40	C1×20	P × 2_pol	P × D	interaction
20	2_pol	polarizability	D pos 2	linear	41	C1×21	P × 3_elec	P × D	interaction
21	3_elec	PCA score, electronic properties	D pos 3	linear					

[a] The abbreviations for nos. 32−41 indicate square (S) and cross (C) terms of the descriptors having the corresponding number.

3, where nos. 1−31 are the main terms and nos. 32−41 are the nonlinear terms.

The calculated responses (pIC_{50} and CE log k'), using the expanded model, versus the experimental values are shown in Figure 6. A visualization of the score and loading vectors from the PLS regression shows the relationship between the compounds, their calculated structural descriptors, and measured responses. A comparison of the score and loading plots (Figure 7a,b) reveals that the first PLS component identified the compounds that have both low CE log k' and low thrombin inhibition (low pIC_{50}), while the second component separated the two responses. The third component (not shown) had similar appearance to the first component in the way that pIC_{50} and the CE log k' were correlated. However, this component had less influence on the model than the two first components. In the score plot (Figure 7a) compounds **44b** and **32b** are located in the lower right corner, which is the corresponding location for pIC_{50} in the loading plot (Figure 7b), indicating that these compounds had good activity but not very high CE log k' values. On the other hand, compounds **40c** and **41c** are located in the upper right corner, at the corresponding location of CE log k', revealing that these compounds had the opposite properties. Compound **29c**'s location between these pairs indicated intermediate pIC_{50} and CE log k' properties. By using several responses in the same model the design of compounds that fulfill more than one criterion is facilitated.

A more general interpretation of the model can be derived using the modeled coefficients (Figure 8). The

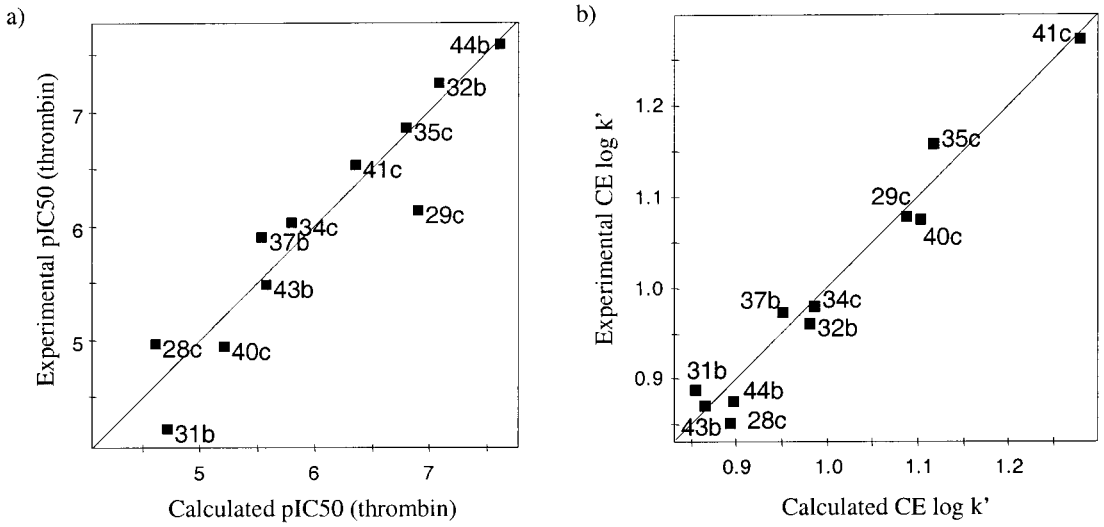


Figure 6. Calculated response values (using the MQSAR) versus the experimental values for (a) the inhibition of thrombin expressed as pIC_{50} ($R^2Y = 0.91$, $Q^2 = 0.56$) and (b) the estimated permeability expressed as CE log k' ($R^2Y = 0.97$, $Q^2 = 0.87$); see Table 1 for chemical structures.

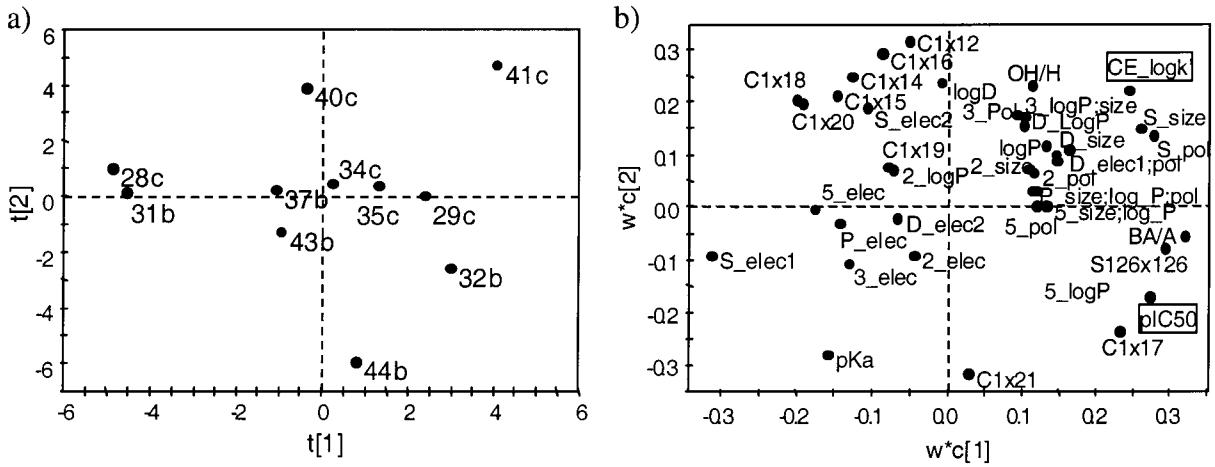


Figure 7. The loading plot (a) and score plot (b) of the MQSAR; for chemical structures see Table 1 and for structural descriptors see Table 3.

properties of the D-residue by themselves were of minor importance when describing the enzyme activity (pIC_{50}), as indicated by their coefficient values, which were close to zero (Figure 8a). However, the interaction terms between the P- and D-residues were large, and it was mainly position 2 at the D-residue that interacted with the P-residue. The coefficients indicated, for example, that if a larger building block, like a methyl or ethyl group, was used at the P-position, a small substituent (C1×18) with low polarizability (C1×20) and large electronic property values (C1×17) should be used at position 2 at the D-residue for increased activity (pIC_{50}). In the synthesized library the second position at the D-residue was varied using hydrogen, methoxy, and chlorine substituents, and thus, hydrogen was shown to be preferable over the other two in combination with a methyl or ethyl group at the P-residue. In addition, the model coefficients for the S-residue showed that stronger basicity at this position led to stronger inhibition of thrombin.

Evaluation of model coefficients for CE log k' revealed that hydroxylation of the S-residue had a positive effect on the possibilities for molecules to be passively absorbed, as indicated by several coefficients (Figure 8b). In addition, the substituents at position 3 on the D-residue, had an effect on CE log k'. At this position, three substituents were varied, hydrogen, methyl, and chlorine, where chlorine was the substituent that best matched the properties that, as indicated by the model, would increase the passive absorption.

pK_a. Consideration of the pK_a values of the investigated compounds with respect to their inhibition of thrombin indicated that hydroxylation next to the amidine (**29c**, **35c**, **41c**) introduced nonlinearity, whereby the pK_a was reduced more than the pIC_{50} (Figure 9, Table 1).

***K_i* of Trypsin.** It was not possible to perform a MQSAR on the selectivity measurements, since low solubility prevented the gathering of sufficient experimental data. However, a comparison of the four derived K_i values (trypsin) with the calculated K_i for thrombin (from the IC_{50} values) indicated that the introduced hydroxy group may decrease the trypsin activity and hence increase specificity toward thrombin (Figure 10, Table 1).

The present study has shown that using a very small series of compounds, SMD can enhance the resulting information content per compound. It is well-known that

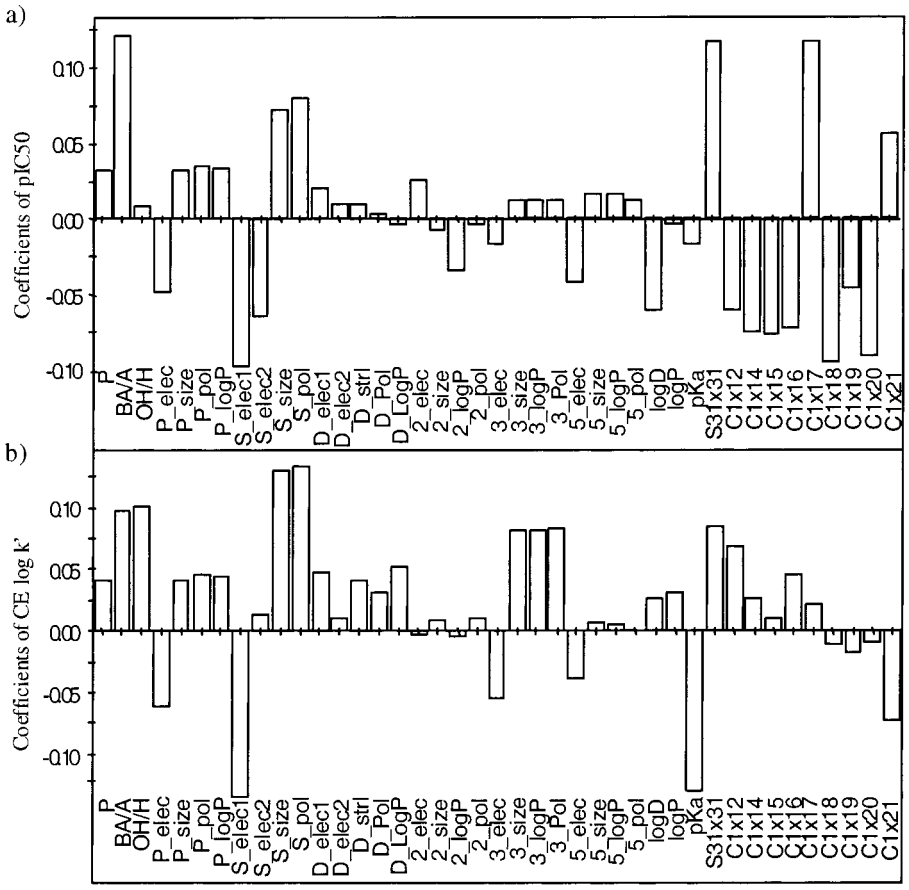


Figure 8. The model coefficients from the MQSAR using two responses, (a) the pIC_{50} and (b) the CE log k' values. A full explanation of the model terms is given in Table 3.

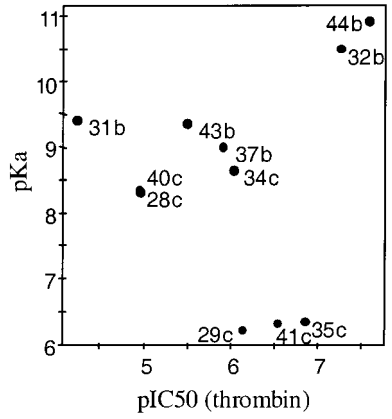


Figure 9. The inhibition of thrombin (pIC_{50}) versus pK_a; for chemical structures see Table 1.

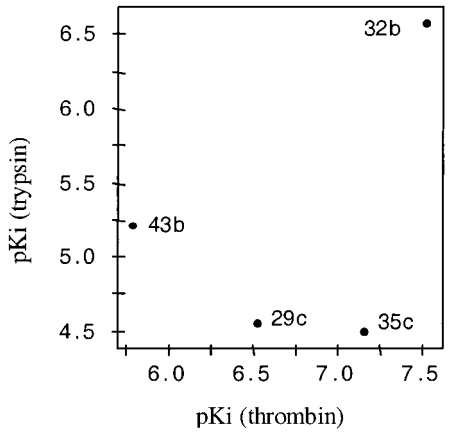


Figure 10. Calculated pK_i for thrombin versus pK_i for trypsin; for chemical structures see Table 1.

scale-up of such series, i.e., the use of parallel or combinatorial approaches, may cause practical problems. However, we cannot foresee any other difficulties and therefore postulate that every compound should have a medicinal consideration, independent of the series size the compound is a member of, before it is made. Thus, we believe that the aforementioned resulting information content enhancement is independent of the size of the compound series and that SMD can be applied to any size of compound series. However, with growing series it may be tempting to allow the depth of medicinal consideration per compound to decrease. Therefore, the use of moderately sized libraries is motivated.

Conclusions

The techniques of statistical experimental design and parallel synthesis were used to make a library of thrombin inhibitors. Prior information and synthetic aspects were included in the design procedure. The biological evaluation showed that a basic moiety was needed in this library to inhibit thrombin. Six of the seven compounds devoid of a basic moiety did not inhibit thrombin.

The MQSAR studies of thrombin activity (pIC_{50}) and the tendency of the compounds to be passively absorbed (CE log k') yielded valuable results. In the structure–activity relationship of thrombin inhibition, the model

showed strong interaction effects between the properties of the D-residue and the properties of the P-residue. This information could only be found using statistical molecular design in combination with multivariate analysis.

The hydroxylation of the benzamidine increased the tendency of the candidate molecules to be passively absorbed without causing large losses of inhibitory effectiveness. In addition, this modification also indicated improved selectivity toward thrombin with respect to trypsin.

To conclude, statistical molecular design is a straightforward method that yields maximal and reliable information from a limited set of molecules, and it can be combined with synthetic considerations and prior information.

Experimental Section

General Chemistry. NMR spectra were recorded on a 400 or 600 MHz Varian Unity plus spectrometer. Chemical shifts (δ) are reported in parts per million (ppm) relative to the deuterated solvent as indicated in the synthetic procedure described below. Spin multiplicities are given as s (singlet), d (doublet), dd (double doublet), ddd (double double doublet), t (triplet), dt (double triplet), q (quartet), dq (double quartet), or m (multiplet). Coupling constants (J) are given in hertz (Hz). Analytical HPLC was performed using a Waters 717 system plus Autosampler, with a Symmetry C_8, 5 μm (4.6 • 50 mm) column, and preparative HPLC was performed using a Prep LC 2000 (Waters) instrument with a Cromasil C_8, 7 μm, (Akzo Nobel) column. In both cases the mobile phases were acetonitrile/water gradients (containing 10 mM ammonium acetate). The UV trace was obtained with a Waters 486 Tunable Absorbance detector (λ = 254 nm). High-resolution mass spectra (HRMS) were produced in a Finnigan MAT900TS (nanospray) spectrometer using an acetonitrile/water matrix. All commercially available chemicals were used without further purification, unless otherwise indicated. Organic extracts were dried using Na_2SO_4 throughout. Column chromatography was performed using Merck silica gel 60 (0.040−0.063 mm), and infrared spectra were recorded using a Mattson FTIR spectrometer.

Acetic Acid 2-(3-Hydroxyphenyl)ethyl Ester (2). The $NaHSO_4$•SiO_2 catalyst was prepared as described by Breton.[51] A stirred mixture of **1** (2.5 g, 18 mmol), ethyl acetate (80 mL), hexane (200 mL), and the catalyst (362 mg) was brought to reflux. After 22 h the solution was cooled to room temperature, filtered, and concentrated at reduced pressure. Purification of the residue by column chromatography (toluene−ethyl acetate 5:1) gave **2** (3.1 g, 95%): ^{1}H NMR ($CDCl_3$) δ 2.03 (s, 3H), 2.86 (t, J = 7.0, 2H), 4.26 (t, J = 7.0, 2H), 6.17 (s, 1H), 6.69−6.71 (m, 3H), 7.14 (t, J = 7.6, 1H).

Acetic Acid 2-(4-Formyl-3-hydroxyphenyl)ethyl Ester (3). Dry THF (40 mL) was added to **2** (2.2 g, 12 mmol). The mixture was cooled to 0 °C under argon and ethylmagnesium bromide (3 M THF solution, 12 mmol) was added dropwise. After removal of the ice bath, the solution was stirred for 30 min and a white salt precipitated. The reaction mixture was concentrated under reduced pressure, benzene was added in two portions (10 mL), and the removal of the solvent was repeated. Benzene (20 mL), paraformaldehyde (2.75 g, 30 mmol), and triethylamine (1.85 g, 18 mmol) were added to the salt.[52] The reaction mixture was refluxed for 3 h and then cooled to room temperature. The solution was made acidic by HCl (2 M) and was extracted three times with diethyl ether. The organic layer was dried and evaporated under reduced pressure to give **3** (2.4 g, 96%), which was used in the next step without further purification: ^{1}H NMR ($CDCl_3$) δ 2.01 (s, 3H), 2.93 (t, J = 6.8, 2H), 4.28 (t, J = 6.8, 2H), 6.83−6.87 (m, 2H), 7.47 (d, J = 7.8, 1H), 9.83 (s, 1H), 11.01 (s, 1H).

2-Allyloxy-4-(2-hydroxyethyl)benzaldehyde (4). Potassium carbonate (5.8 g, 42 mmol), methanol (60 mL), and chloroform (120 mL) were refluxed for 20 min, followed by addition of **3** (2 g, 9.6 mmol). Allyl bromide (1.7 g, 14 mmol) was added after 5 min and the reaction mixture was refluxed overnight. The solution was poured into ice and extracted three times with Et_2O. Evaporation under reduced pressure of the dried organic layer yielded **4** (1.7 g, 85%), which was used without further purification: ^{1}H NMR ($CDCl_3$) δ 2.88 (t, J = 6.3, 2H), 3.87 (t, J = 6.3, 2H), 4.64 (dt, J = 5.2, 1.5, 2H), 5.32 (dq, J = 10.6, 1.5), 5.43 (dq, J = 17.3, 1.5), 6.00−6.10 (m, 1H), 6.80 (s, 1H), 6.88 (d, J = 7.9, 1H), 7.76 (d, J = 7.9, 1H), 10.4 (s, 1H).

Acetic Acid 2-(3-Allyloxy-4-cyanophenyl)ethyl Ester (5). A mixture of freshly distilled nitroethane (1.5 g, 16 mmol), sodium acetate (2.7 g, 33 mmol), acetic acid (4 mL), and **4** (1.7 g, 8 mmol) was refluxed overnight.[49] The solution was poured into ice (40 g) and extracted three times with diethyl ether. The organic layer was washed three times with $NaHCO_3$ (saturated), dried, and concentrated at reduced pressure. Purification of the residue by column chromatography (toluene−ethyl acetate 6:1) gave **5** (1.2 g, 60%): ^{1}H NMR ($CDCl_3$) δ 2.01 (s, 3H), 2.94 (t, J = 6.8, 2H), 4.27 (t, J = 6.8, 2H), 4.64 (dt, J = 5.0, 1.5, 2H), 5.31 (dq, J = 10.6, 1.5, 1H), 5.46 (dq, J = 17.3, 1.5, 1H), 5.98−6.07 (m, 1H), 6.78 (s, 1H), 6.84 (d, J = 7.8, 1H), 7.47 (d, J = 7.8, 1H); IR 1738 (C═O), 2225 (C≡N).

2-Allyloxy-4-(2-hydroxyethyl)benzonitrile (6). A stirred solution of potassium carbonate (0.8 g, 5.9 mmol), dry methanol (30 mL), and **5** (1.2 g, 4.9 mmol) was refluxed for 2 h. The mixture was poured into ice and immediately extracted three times with diethyl ether. The combined organic layers were washed with water and brine. Evaporation under reduced pressure gave **6** (0.77 g, 77%): ^{1}H NMR ($CDCl_3$) δ 2.86 (2H, t, J = 6.1), 3.78 (2H, t, J = 6.1), 4.73 (2H, d, J = 5.1), 5.27−5.31 (1H, m), 5.46−5.52 (1H, m), 6.04−6.14 (1H, m), 6.98 (1H, d, J = 7.8), 7.11 (1H, s), 7.53 (1H, d, J = 7.8).

2-(3-Allyloxyphenyl)ethanol (7). The preparation of **7** (quant. yield) was as for **4** except that **1** was used instead of **3**: ^{1}H NMR ($CDCl_3$) δ 2.72 (2H, t, J = 6.9), 3.74 (2H, t, J = 6.9), 4.53 (2H, dt, J = 5.2, 1.7), 5.22 (1H, dq, J = 10.6, 1.7), 5.40 (1H, dq, J = 17.3, 1.7), 5.99−6.12 (1H, m), 6.74−6.85 (3H, m), 7.16 (1H, t, J = 7.7).

5-Ethylbenzene-1,3-diol (9). A reaction mixture of 2 g (13.1 mmol) of 1-(3,5-Dihydroxyphenyl)ethanone (**8**), 1 g of Pd/C (10%), and 150 mL of aqueous HCl (4%) in a 500 mL round-bottom flask was hydrated overnight (ca 16 h and 900 mL of H_2). The mixture was filtered through Celite and extracted with three portions of Et_2O. The dried solution was evaporated at reduced pressure. Purification of the residue by column chromatography (toluene−ethyl acetate 5:1) gave **9** (1.53 g, 85%) as pale orange crystals: ^{1}H NMR (CD_3COCD_3) δ 1.13 (t, J = 7.6, 3H), 2.45 (q, J = 7.6, 2H), 6.16 (t, J = 2.2, 1H), 6.18 (d, J = 2.2, 2H), 7.98 (s, 1H).

3,5-Dichloro-2-methoxybenzenesulfonyl Chloride (11). To a stirred solution of aqueous HBF_4 (209 mg, 60%) and CH_2Cl_2 (40 mL) was added **10** (10 mmol) at 0 °C. Three portions of $TMSCHN_2$ [2 M hexane solution, 5 mL (10 mmol), 3 mL (6 mmol), 2 mL (4 mmol)] were added dropwise at intervals of 30 min, and the mixture was stirred at 0 °C for an additional 40 min.[53] The reaction mixture was poured into water and extracted with two portions of CH_2Cl_2. The organic layer was washed with water, dried, and evaporated under reduced pressure. The resulting oil was purified by slow distillation (85 °C, 0.1 mmHg) and subsequent filtration through SiO_2 (heptane−ethyl acetate 20:1). Removal of the solvent gave **11** (1.65 g, 60%) as an oil that crystallized in the freezer: ^{1}H NMR ($CDCl_3$) δ 4.10 (s, 3H), 7.72 (d, J = 2.6, 1H), 7.88 (d, J = 2.6, 1H).

General Procedure for Compounds 12−21. The sulfonyl chloride (3.3 mmol) was added to a stirred solution of the benzene diol (3 mmol), diethyl ether (4 mL), and $NaHCO_3$ (saturated, 6 mL). The mixture was kept at room temperature for 2 days. The organic layer was separated and washed with two extractions of K_2CO_3 (saturated). The product was purified and isolated by three extractions with NaOH (1 M), and the water layer was acidified by HCl (2 M), followed by three

extractions using Et_2O. The combined organic layers were evaporated under reduced pressure to give the pure monosubstitued product.

2,3-Dichlorobenzenesulfonic acid 3-hydroxyphenyl ester (12): yield 40%; ^{1}H NMR ($CDCl_3$) δ 5.51 (1H, s, broad), 6.65−6.72 (3H, m), 7.13 (1H, t, J = 8.5), 7.30 (1H, t, J = 8.1), 7.73 (1H, dd, J = 8.1, 1.5), 7.88 (1H, dd, J = 8.1, 1.5).

Benzenesulfonic acid 3-hydroxyphenyl ester (13): yield 60%; ^{1}H NMR ($CDCl_3$) δ 6.24 (1H, s, broad), 6.46 (1H, ddd, J = 8.3, 2.4, 0.7), 6.54 (1H, t, J = 2.4), 6.71 (1H, ddd, J = 8.3, 2.4, 0.7), 7.06 (1H, t, J = 8.3), 7.46−7.51 (2H, m), 7.60−7.64 (1H, m), 7.79−7.82 (2H, m).

2-Methoxy-5-methylbenzenesulfonic acid 3-hydroxyphenyl ester (14): yield 40%; ^{1}H NMR ($CDCl_3$) δ 2.26 (3H, s), 3.93 (3H, s), 5.26 (1H, s, broad), 6.62−6.69 (3H, m), 6.94 (1H, d, J = 8.5), 7.07−7.12 (1H, m), 7.36 (1H, dd, J = 8.5, 2.0), 7.59 (1H, d, J = 2.0).

3-Methylbenzenesulfonic Acid 3-Hydroxyphenyl Ester (15). The equivalents of benzene diol compared to the sulfonyl chloride were increased, from 0.9 equiv in the general procedure to 3 equiv: yield 73%; ^{1}H NMR ($CDCl_3$) δ 2.36 (3H, s), 6.18 (1H, s, broad), 6.48 (1H, ddd, (J = 8.3, 2.4, 1.0), 6.55 (1H, t, J = 2.4), 6.71 (1H, ddd, J = 8.3, 2.4, 1.0), 7.07 (1H, t, J = 8.3), 7.36 (1H, t, J = 7.8), 7.42 (1H, d, J = 7.8), 7.60 (1H, d, 7.8), 7.64 (1H, s).

3,5-Dichloro-2-methoxybenzenesulfonic acid 3-hydroxyphenyl ester (16): yield 23%; ^{1}H NMR ($CDCl_3$) δ 4.07 (3H, s), 4.97 (1H, s, broad), 6.66 (1H, t, J = 2.4), 6.69 (1H, dd, J = 8.3, 2.4), 6.72 (1H, dd, J = 8.3, 2.4), 7.16 (1H, t, J = 8.3), 7.64 (1H, d, J = 2.7), 7.69 (1H, d, J = 2.7).

3,5-Dichloro-2-methoxybenzenesulfonic Acid 3-Hydroxy-5-methylphenyl Ester (17). The equivalents of benzene diol compared to the sulfonyl chloride were increased, from 0.9 equiv in the general procedure to 3 equiv: yield 23%; ^{1}H NMR (CD_3COCD_3) δ 2.80 (3H, s), 4.08 (3H, s), 6.42 (1H, s), 6.47 (1H, s), 6.61 (1H, s), 7.68 (1H, d, J = 2.4), 7.96 (1H, d, J = 2.4), 8.64 (1H, s).

2,5-Dichlorobenzenesulfonic acid 3-hydroxy-5-methylphenyl ester (18): yield 16%; ^{1}H NMR ($CDCl_3$) δ 2.23 (3H, s), 5.16 (1H, s, broad), 6.43−6.44 (1H, m), 6.53−6.54 (2H, m), 7.52−7.53 (2H, m), 7.93−7.94 (1H, m).

3-Methylbenzenesulfonic Acid 3-Hydroxy-5-methylphenyl Ester (19). The equivalents of benzene diol compared to the sulfonyl chloride were increased, from 0.9 equiv in the general procedure to 3 equiv: yield 73%; ^{1}H NMR (CD_3COCD_3) δ 2.43 (3H, s), 3.04 (3H, s), 6.31−6.33 (2H, m), 6.59−6.60 (2H, m), 7.52 (1H, t, J = 7.6), 7.59 (1H, d, J = 7.6), 7.65 (1H, d, J = 7.6), 7.69 (1H, s), 8.60 (1H, s).

Benzenesulfonic acid 3-hydroxy-5-methylphenyl ester (20): yield 73%; ^{1}H NMR ($CDCl_3$) δ 2.19 (3H, s), 5.16 (1H, s), 6.28−6.30 (1H, m), 6.36−6.37 (1H, m), 6.51−6.52 (1H, m), 7.49−7.53 (2H, m), 7.62−7.67 (1H, m), 7.83−7.85 (2H, m).

2,3-Dichlorobenzenesulfonic acid 3-hydroxy-5-methylphenyl ester (21): yield 20%; ^{1}H NMR ($CDCl_3$) δ 2.22 (1H, s), 5.20 (1H, s, broad), 6.41−6.43 (1H, m), 6.52−6.53 (2H, m), 7.31 (1H, t, J = 8.1), 7.73 (1H, dd, J = 8.1, 1.7), 7.9 (1H, dd, J = 8.1, 1.7).

General Procedure for Compounds 22−26. The sulfonyl chloride (3 mmol) was added to a stirred solution of the benzene diol (9 mmol), diethyl ether (15 mL), and $NaHCO_3$ (saturated, 15 mL). The mixture was kept at room temperature for 2 days. The organic layer was separated and concentrated under reduced pressure. Ts-trisamine (1 g, 3.66 mmol) and CH_2Cl_2 (7 mL) were added, and the solution was kept at room temperature for 2.5 h. The mixture was filtered, and the resins were sequentially washed with CH_2Cl_2 and Et_2O three times. The solvent was removed under reduced pressure, and the residue was purified by preparative HPLC to give the monosubstituted product.

2-Methoxy-5-methylbenzenesulfonic acid 3-ethyl-5-hydroxyphenyl ester (22): yield 35%; ^{1}H NMR (CD_3COCD_3) δ 1.08 (3H, t, J = 7.6), 2.27 (3H, s), 2.48 (2H, q, J = 7.6), 3.99 (3H, s), 6.41−6.42 (1H, m), 6.45 (1H, t, J = 2.2), 6.57−6.58 (1H, m), 7.21 (1H, d, J = 8.1), 7.54−7.50 (2H, m).

Benzenesulfonic acid 3-ethyl-5-hydroxyphenyl ester (23): yield 38%; ^{1}H NMR (CD_3COCD_3) δ 1.06 (3H, t, J = 7.6), 2.46 (2H, q, J = 7.6), 6.27−6.28 (1H, m), 6.35 (1H, t, J = 2.2), 7.61−7.62 (1H, m), 7.64−7.68 (2H, m), 7.77−7.81 (1H, m), 7.85−7.88 (2H, m), 8.61 (1H, s, broad).

2,3-Dichloro-benzenesulfonic acid 3-ethyl-5-hydroxyphenyl ester (24): yield 29%; ^{1}H NMR (CD_3COCD_3) δ 1.07 (3H, t, J = 7.6), 2.49 (2H, q, J = 7.6), 6.43−6.45 (2H, m), 6.62−6.63 (1H, m), 7.57 (1H, t, J = 8.3), 7.94 (1H, dd, J = 8.3, 1.5), 8.00 (1H, dd, J = 8.3, 1.5).

3,5-Dichloro-2-methoxybenzenesulfonic acid 3-ethyl-5-hydroxyphenyl ester (25): yield 22%; ^{1}H NMR (CD_3COCD_3) δ 1.09 (3H, t, J = 7.6), 2.51 (2H, q, J = 7.6), 4.08 (3H, s), 7.44−7.46 (2H, m), 7.63−7.64 (1H, m), 7.66 (1H, d, J = 2.4), 7.96 (1H, d, J = 2.4), 8.73 (1H, s, broad).

2,5-Dichlorobenzenesulfonic acid 3-ethyl-5-hydroxyphenyl ester (26): yield 20%; ^{1}H NMR (CD_3COCD_3) δ 1.08 (3H, t, J = 7.6), 2.50 (2H, q, J = 7.6), 6.44−6.45 (1H, m), 6.46 (1H, t, J = 2.2), 7.64−7.65 (1H, m), 7.82−7.83 (2H, m), 7.88−7.89 (1H, m).

General Procedure for Compounds 27a−44a. DEAD (diethylazodicarboxylate, 0.65 mmol) was added dropwise to a stirred solution of the monosubstituted product (**12**−**26**, 0.54 mmol), PPh_3 (0.65 mmol), and 2.8 mL of dry THF. The reaction mixture was kept at room temperature for 16 h, and the solvent was removed under reduced pressure. The residue was purified by preparative HPLC to give the products.

2,3-Dichlorobenzenesulfonic acid 3-[2-(3-allyloxyphenyl)ethoxy]phenyl ester (27a): yield 53%; ^{1}H NMR (CD_3COCD_3) δ 3.00 (2H, t, J = 6.8), 4.15 (2H, t, J = 6.8), 4.55 (2H, dt, J = 5.1, 1.7), 5.23 (1H, dq, J = 10.5, 1.7), 5.40 (1H, dq, J = 17.3, 1.7), 6.01−6.10 (1H, m), 6.70 (1H, ddd, J = 8.3, 2.4, 0.7), 6.73 (1H, t, J = 2.4), 6.81 (1H, ddd, J = 8.1, 2.7, 1.0), 6.86−6.91 (3H, m), 7.21 (1H, t, J = 8.1), 7.25 (1H, t, J = 8.3), 7.54 (1H, t, J = 8.1), 7.92−7.97 (2H, m).

Benzenesulfonic acid 3-[2-(3-allyloxy-4-cyanophenyl)-ethoxy]phenyl ester (28a): yield 35%; ^{1}H NMR (CD_3COCD_3) δ 3.11 (2H, t, J = 6.6), 4.19 (2H, t, J = 6.6), 4.75 (2H, dt, J = 5.2, 1.5), 5.30 (1H, dq, J = 10.6, 1.5), 5.49 (1H, dq, J = 17.5, 1.5), 6.03−6.16 (1H, m), 6.56−6.60 (2H, m), 6.85−6.89 (1H, m), 7.04−7.07 (2H, m), 7.20−7.27 (2H, m), 7.58−7.69 (3H, m), 7.76−7.88 (3H, m).

2-Methoxy-5-methylbenzenesulfonic acid 3-[2-(3-allyloxy-4-cyanophenyl)ethoxy]phenyl ester (29a): yield 49%; ^{1}H NMR (CD_3COCD_3) δ 2.24 (3H, s), 3.11 (2H, t, J = 6.6), 3.97 (3H, s), 4.20 (2H, t, J = 6.6), 4.20 (2H, t, J = 6.6), 4.73 (2H, dt, J = 5.1, 1.6), 5.28 (1H, dq, J = 17.3, 1.6), 6.03−6.13 (1H, m), 6.66−6.69 (2H, m), 6.81−6.84 (1H, m), 7.04 (1H, dd, J = 7.8, 1.5), 7.17−7.23 (3H, m), 7.48−7.53 (2H, m), 7.57 (1H, d, J = 7.8).

2-Methoxy-5-methylbenzenesulfonic acid 3-phenethyloxyphenyl ester (30): yield 84%; ^{1}H NMR (CD_3COCD_3) δ 2.25 (3H, s), 3.03 (2H, t, J = 6.8), 3.97 (3H, s), 4.13 (2H, t, J = 6.8), 6.65−6.69 (2H, m), 6.83 (1H, dd, J = 5.8, 2.4), 7.18−7.23 (3H, m), 7.29−7.31 (4H, m), 7.49−7.53 (2H, m); HRMS calcd for $C_{22}H_{22}O_5S$, 398.1188; observed, 398.1188.

Toluene-3-sulfonic acid 3-[2-(4-cyanophenyl)ethoxy]-phenyl ester (31a): yield 49%; ^{1}H NMR (CD_3COCD_3) δ 2.41 (3H, s), 3.14 (2H, t, J = 6.6), 4.19 (2H, t, 6.6), 6.57−6.61 (2H, m), 7.23 (1H, t, J = 8.3), 7.49−7.72 (8H, m).

3,5-Dichloro-2-methoxybenzenesulfonic acid 3-[2-(4-cyanophenyl)ethoxy]phenyl ester (32a): yield 25%; ^{1}H NMR (CD_3COCD_3) δ 3.17 (2H, t, J = 6.6), 4.07 (3H, s), 4.23 (2H, t, J = 6.6), 6.69−6.73 (2H, m), 6.89 (1H, ddd, J = 8.5, 2.2, 1.0), 7.25−7.29 (1H, m), 7.53−7.56 (2H, m), 7.66 (1H, d, J = 2.7), 7.70−7.73 (2H, m), 7.95 (1H, d, J = 2.7).

3,5-Dichloro-2-methoxybenzenesulfonic acid 3-[2-(3-allyloxyphenyl)ethoxy]-5-methylphenyl ester (33a): yield 25%; ^{1}H NMR (CD_3COCD_3) δ 2.24 (3H, s), 3.30 (2H, t, J = 6.8), 4.07 (3H, s), 4.12 (2H, t, J = 6.8), 4.56 (2H, dt, J = 5.1, 1.7), 5.23 (1H, dq, J = 10.7, 1.7), 5.40 (1H, dq, J = 17.3, 1.7), 6.02−6.11 (1H, m), 6.5 (1H, t, J = 2.2), 6.56−6.57 (1H, m), 6.71−6.72 (1H, m), 6.81 (1H, d, J = 7.8), 6.87 (1H, d, J = 7.8),

6.90−6.91 (1H, m), 7.21 (1H, t, J = 7.8), 7.67 (1H, d, J = 2.7), 7.92 (1H, d, J = 2.7).

2,5-Dichlorobenzenesulfonic acid 3-[2-(3-allyloxy-4-cyanophenyl)ethoxy]-5-methylphenyl ester (34a): yield 48%; ^{1}H NMR (CD_3COCD_3) δ 2.22 (3H, s), 3.10 (2H, t, J = 6.6), 4.19 (2H, t, J = 6.6), 4.73 (2H, dt, J = 5.1, 1.7), 5.29 (1H, dq, J = 10.5, 1.7), 5.49 (1H, dq, J = 17.3, 1.7), 6.04−6.13 (1H, m), 6.56−6.57 (1H, m), 6.71−6.72 (1H, m), 7.03 (1H, dd, J = 7.8, 1.5), 7.18 (1H, d, J = 1.2), 7.57 (1H, d, J = 7.8), 7.79 (2H, d, J = 1.5), 7.89 (1H, t, J = 1.5).

Toluene-3-sulfonic acid 3-[2-(3-allyloxy-4-cyanophenyl)ethoxy]-5-methylphenyl ester (35a): yield 35%; ^{1}H NMR (CD_3COCD_3) δ 2.21 (3H, s), 2.41 (3H, s), 3.08 (2H, t, J = 6.6), 4.15 (2H, t, J = 6.6), 4.74 (2H, dt, J = 5.1, 1.6), 5.29 (1H, dq, J = 10.5, 1.6), 5.49 (1H, dq, J = 17.3, 1.6), 6.04−6.14 (1H, m), 6.38 (1H, t, J = 2.2), 6.43−6.44 (1H, m), 6.68−6.69 (1H, m), 7.03 (1H, dd, J = 7.8, 1.2), 7.18 (1H, d, J = 1.2), 7.51 (1H, t, J = 7.8), 7.57 (2h, d, J = 7.8), 7.63−7.66 (1H, m), 7.68−7.69 (1H, m).

Benzenesulfonic acid 3-methyl-5-phenethyloxyphenyl ester (36): yield 60%; ^{1}H NMR (CD_3COCD_3) δ 2.20 (3H, s), 2.99 (2H, t, J = 6.8), 4.06 (2H, t, J = 6.8), 6.37 (1H, t, J = 2.2), 6.45−6.46 (1H, m), 6.68−6.69 (1H, m), 7.19−7.32 (5H, m), 7.60−7.65 (2H, m), 7.72−7.77 (1H, m), 7.86−7.89 (2H, m); HRMS calcd for $C_{21}H_{20}O_4S$, 368.1082; observed, no ions.

2,3-Dichlorobenzenesulfonic acid 3-[2-(4-cyanophenyl)ethoxy]-5-methylphenyl ester (37a): yield 42%; ^{1}H NMR (CD_3COCD_3) δ 2.21 (3H, s), 3.13 (2H, t, J = 6.6), 4.18 (2H, t, J = 6.6), 6.49−6.51 (1H, m), 6.56−6.57 (2H, m), 6.69−6.70 (1H, m), 7.49−7.52 (2H, m), 7.56 (1H, t, J = 8.1), 7.67−7.70 (2H, m), 7.93−7.98 (2H, m).

2-Methoxy-5-methylbenzenesulfonic acid 3-[2-(3-allyloxyphenyl)ethoxy]-5-ethylphenyl ester (38a): yield 27%; ^{1}H NMR (CD_3COCD_3) δ 1.08 (3H, t, J = 7.6), 2.24 (3H, s), 2.52 (2H, q, J = 7.6), 2.99 (2H, t, J = 6.8), 3.97 (3H, s), 4.11 (2H, t, J = 6.8), 4.55 (2H, dt, J = 5.1, 1.7), 5.23 (1H, dq, J = 10.7, 1.7), 5.41 (1H, dq, J = 17.3, 1.7), 6.02−6.11 (1H, m), 6.49−6.52 (2H, m), 6.67−6.68 (1H, m), 6.81 (1H, ddd, J = 8.3, 2.4, 1.0), 6.85−6.88 (1H, m), 6.90−6.91 (1H, m), 7.17−7.23 (2H, m), 7.47−7.50 (1H, m), 7.53−7.54 (1H, m).

Benzenesulfonic acid 3-[2-(3-allyloxyphenyl)ethoxy]-5-ethylphenyl ester (39a): yield 44%; ^{1}H NMR (CD_3COCD_3) δ 2.98 (2H, t, J = 6.8), 4.10 (2H, t, J = 6.8), 4.56 (2H, dt, (J = 5.1, 1.7), 5.23 (1H, dq, J = 10.5, 1.7), 5.41 (1H, dq, J = 17.3, 1.7), 6.02−6.12 (1H, m), 6.39−6.50 (1H, m), 6.71−6.72 (1H, m), 6.81 (1H, ddd, J = 8.3, 2.7, 1.0), 6.85−6.88 (1H, m), 6.90−6.91 (1H, m), 7.21 (1H, t, J = 7.8), 7.61−7.65 (2H, m), 7.74−7.78 (1H, m), 7.85−7.88 (2H, m).

2,3-Dichlorobenzenesulfonic acid 3-[2-(3-allyloxy-4-cyanophenyl)ethoxy]-5-ethylphenyl ester (40a): yield 32%; ^{1}H NMR (CD_3COCD_3) δ 1.06 (3H, t, J = 7.6), 2.52 (2H, q, J = 7.6), 3.11 (2H, t, J = 6.6), 4.20 (2H, t, J = 6.6), 4.74 (2H, dt, J = 5.1, 1.7), 5.29 (1H, dq, J = 10.5, 1.7), 5.48 (1H, dq, J = 17.3, 1.7), 6.04−6.13 (1H, m), 6.52−6.53 (1H, m), 6.55 (1H, t, J = 2.2), 6.72−6.73 (1H, m), 7.04 (1H, dd, J = 7.8, 1.5), 7.18 (1H, d, J = 1.2), 7.54−7.58 (2H, m), 7.94 (1H, dd, J = 8.1, 1.5), 7.97 (1H, dd, J = 8.1, 1.5).

3,5-Dichloro-2-methoxybenzenesulfonic acid 3-[2-(3-allyloxy-4-cyanophenyl)ethoxy]-5-ethylphenyl ester (41a): yield 17%; ^{1}H NMR (CD_3COCD_3) δ 1.09 (3H, t, J = 7.6), 2.54 (2H, q, J = 7.6), 3.12 (2H, t, J = 6.6), 4.07 (3H, s), 4.21 (2H, t, J = 6.6), 4.75 (2H, dt, J = 5.1, 1.7), 5.29 (1H, dq, J = 10.7, 1.7), 5.49 (1H, dq, J = 17.3, 1.7), 6.03−6.14 (1H, m), 6.53−6.55 (2H, m), 6.73−6.74 (1H, m), 7.06 (1H, dd, J = 7.8, 1.2), 7.20 (1H, d, J = 1.2), 7.59 (1H, d, J = 7.8), 7.65 (1H, d, J = 2.5), 7.95 (1H, d, J = 2.5).

2,5-Dichlorobenzenesulfonic acid 3-ethyl-5-phenethyloxyphenyl ester (42): yield 18%; ^{1}H NMR (CD_3COCD_3) δ 1.09 (3H, t, J = 7.6), 2.54 (2H, q, J = 7.6), 3.04 (2H, t, J = 6.8), 4.15 (2H, t, J = 6.8), 6.53−6.86 (2H, m), 6.75−6.77 (1H, m), 7.20−7.25 (1H, m), 7.30−7.32 (4H, m), 7.80−7.81 (2H, m), 7.89−7.90 (1H, m); HRMS calcd for $C_{22}H_{20}Cl_2O_4S$, 450.0459; observed, no ions.

2-Methoxy-5-methylbenzenesulfonic acid 3-[2-(4-cyanophenyl)ethoxy]-5-ethylphenyl ester (43a): yield 32%; ^{1}H NMR (CD_3COCD_3) δ 1.07 (3H, t, J = 7.6), 2.25 (3H, s), 2.51 (2H, q, J = 7.6), 3.13 (2H, t, J = 6.6), 3.97 (3H, s), 4.18 (2H, t, J = 6.6), 6.48−6.51 (2H, m), 6.66−6.68 (1H, m), 7.20 (1H, d, J = 8.4), 7.68−7.71 (2H, m).

Benzenesulfonic acid 3-[2-(4-cyanophenyl)ethoxy]-5-ethylphenyl ester (44a): yield 44%; ^{1}H NMR (CD_3COCD_3) δ 1.05 (3H, t, J = 7.6), 2.49 (2H, q, J = 7.6), 3.12 (2H, t, J = 6.6), 4.16 (2H, t, J = 6.6), 6.39−6.40 (1H, m), 6.42 (1H, t, J = 2.2), 6.70−6.71 (1H, m), 7.50−7.52 (2H, m), 7.62−7.70 (4H, m), 7.75−7.80 (1H, m), 7.85−7.88 (2H, m).

General Procedure for Compounds 27b−29b, 33b−35b, and 38b−41b. Tetrakis(Pd(PPh_3)$_4$ (8.3 μmol) was added to a solution of **27a** (42 μmol) and 0.44 mL of dry THF. After 5 min, $NaBH_4$ (63 μmol) was added to the stirred solution and the reaction mixture was kept at room temperature for 3 h. The reaction was quenched with 1 M HCl (to neutral pH). THF was removed under reduced pressure, and the aqueous phase was extracted three times with CH_2Cl_2. The solvent was removed under reduced pressure and the residue was purified by preparative HPLC to give the deprotected products.

2,3-Dichlorobenzenesulfonic acid 3-[2-(3-hydroxyphenyl)ethoxy]phenyl ester (27b): yield 80%; ^{1}H NMR (CD_3OD) δ 2.95 (2H, t, J = 6.8), 4.07 (2H, t, J = 6.8), 6.64 (1H, t, J = 2.2), 6.67−6.75 (4H, m), 6.84 (1H, dd, J = 8.3, 1.0), 7.11−7.16 (1H, m), 7.22 (1H, t, J = 8.3), 7.44 (1H, t, J = 8.1), 7.87−7.91 (2H, m); HRMS calcd for $C_{20}H_{16}Cl_2O_5S$, 438.0095; observed, 438.0095.

Benzenesulfonic acid 3-[2-(4-cyano-3-hydroxyphenyl)-ethoxy]phenyl ester (28b): yield 60%; ^{1}H NMR (CD_3OD) δ 2.99 (2H, t, J = 6.6), 4.09 (2H, t, J = 6.6), 6.51 (1H, t, J = 2.2), 6.58 (1H, dd, J = 8.1, 1.7), 6.81−6.89 (3H, m), 7.20 (1H, t, J = 8.1), 7.43 (1H, d, J = 7.8), 7.60 (2H, t, J = 7.8), 7.74 (1H, t, J = 7.6), 7.84 (2H, d, J = 7.8).

2-Methoxy-5-methylbenzenesulfonic acid 3-[2-(4-cyano-3-hydroxyphenyl)ethoxy]phenyl ester (29b): yield 71%; ^{1}H NMR (CD_3 $COCD_3$) δ 2.26 (3H, s), 3.04 (2H, t, J = 6.6), 3.98 (3H, s), 4.16 (2H, t, J = 6.6), 6.65−6.69 (2H, m), 6.82 (1H, ddd, J = 8.3, 2.4, 1.0), 6.95 (1H, dd, J = 7.8, 1.5), 7.04 (1H, d, J = 1.0), 7.19−7.24 (2H, m), 7.49−7.53 (3H, m).

3,5-Dichloro-2-methoxybenzenesulfonic acid 3-[2-(3-hydroxyphenyl)ethoxy]-5-methylphenyl ester (33b): yield 32%; ^{1}H NMR (CD_3 $SOCD_3$) δ 2.21 (3H, s), 2.87 (2H, t, J = 6.8), 3.99 (3H, s), 4.05 (2H, t, J = 6.8), 6.42 (1H, t, J = 2.2), 6.52−6.53 (1H, m), 6.60−6.62 (1H, m), 6.65−6.67 (2H, m), 6.74−6.75 (1H, m), 7.07 (1H, t, J = 7.9), 7.70 (1H, d, J = 2.6), 8.19 (1H, d, J = 2.6), 9.31 (1H, s, broad); HRMS calcd for $C_{22}H_{20}Cl_2O_6S$, 482.0358; observed, 482.0330.

2,5-Dichlorobenzenesulfonic acid 3-[2-(4-cyano-3-hydroxyphenyl)ethoxy]-5-methylphenyl ester (34b): yield 54%; ^{1}H NMR (CD_3 OD) δ 2.27 (3H, s), 3.03 (2H, t, J = 6.3), 4.13 (2H, t, J = 6.3), 6.45−6.47 (1H, m), 6.55−6.56 (1H, m), 6.71−6.72 (1H, m), 6.87 (1H, dd, J = 8.1, 1.2), 6.91 (1H, d, J = 1.2), 7.46 (1H, d, J = 8.1), 7.74−7.75 (2H, m), 7.91 (1H, dd, J = 1.7, 1.0).

Toluene-3-sulfonic acid 3-[2-(4-cyano-3-hydroxyphenyl)ethoxy]-5-methylphenyl ester (35b): yield 78%; ^{1}H NMR ($CDCl_3$) δ 2.17 (3H, s), 2.37 (3H, s), 2.93 (2H, t, J = 6.6), 4.00 (2H, t, J = 6.6), 5.87 (1H, s, broad), 6.32−6.35 (1H, m), 6.54 (1H, s), 6.77 (1H, d, J = 8.1), 6.87 (1H, s), 7.34−7.44 (3H, m), 7.60 (1H, d, J = 7.6), 7.64 (1H, s).

2-Methoxy-5-methylbenzenesolfonic acid 3-ethyl-5-[2-(3-hydroxyphenyl)ethoxy]phenyl ester (38b): yield 55%; ^{1}H NMR (CD_3OD) δ 1.14 (3H, t, J = 7.8), 2.27 (3H, s), 2.55 (2H, q, J = 7.6), 2.95 (2H, t, J = 6.8), 3.98 (3H, s), 4.05 (2H, t, J = 6.8), 6.41 (1H, t, J = 2.2), 6.50−6.51 (1H, m), 6.66−6.76 (4H, m), 7.12−7.18 (2H, m), 7.48−7.54 (2H, m); HRMS calcd for $C_{24}H_{26}O_6S$, 442.1450; observed, 442.1442.

Benzenesulfonic acid 3-ethyl-5-[2-(3-hydroxyphenyl)-ethoxy]phenyl ester (39b): yield 34%; ^{1}H NMR (CD_3COOD) δ 1.07 (3H, t, J = 7.6), 2.49 (2H, q, J = 7.6), 2.93 (2H, t, J = 6.8), 4.02 (2H, t, J = 6.8), 6.35 (1H, t, J = 2.2), 6.37−6.38 (1H, m), 6.64−6.65 (1H, m), 6.70−6.77 (3H, m), 7.12 (1H, t, J =

7.8), 7.54 (2H, t, J = 7.3), 7.67 (1H, t, J = 7.3), 7.84 (2H, d, J = 7.3); HRMS calcd for $C_{22}H_{22}O_5S$, 398.1188; observed, 398.1193.

2,3-Dichlorobenzenesulfonic acid 3-[2-(4-cyano-3-hydroxyphenyl)ethoxy]-5-ethylphenyl ester (40b): yield 67%; ^{1}H NMR (CD_3OD) δ 1.09 (3H, t, J = 7.8), 2.52 (2H, q, J = 7.8), 3.03 (2H, t, J = 6.3), 4.11 (2H, t, J = 6.3), 6.47 (1H, t, J = 2.2), 6.49−6.50 (1H, m), 6.67−6.68 (1H, m), 6.85 (1H, d, J = 8.1), 6.91 (1H, s), 7.42−7.47 (2H, m), 7.87−7.90 (2H, m).

3,5-Dichloro-2-methoxybenzenesulfonic acid 3-[2-(4-cyano-3-hydroxyphenyl)ethoxy]-5-ethylphenyl ester (41b): yield 52%; ^{1}H NMR (CD_3COCD_3) δ 1.09 (3H, t, J = 7.6), 2.54 (2H, q, J = 7.6), 3.04 (2H, t, J = 6.6), 4.07 (3H, s), 4.16 (2H, t, J = 6.6), 6.53 (2H, d, J = 1.7), 6.73 (1H, t, J = 1.7), 6.94 (1H, dd, J = 7.8, 1.2), 7.05 (1H, d, J = 1.2), 7.50 (1H, d, J = 7.8), 7.65 (1H, d, J = 7.8), 7.95 (1H, d, J = 2.7).

General Procedure for Compounds 31b, 37b, and 43b. A solution of **31a** (65 μmol) and dry THF (1 mL) was brought to reflux and BMS (borane−Me_2S, 129 μmol) was added dropwise. Two additional portions of BMS (2 × 129 μmol) were added after 30 min and 1 h, respectively, and the reaction mixture was refluxed for an additional 2.5 h. The condenser was removed and methanol was added and subsequently removed by heating the mixture. This procedure was repeated twice before total removal of the solvent under reduced pressure. The residue was purified by preparative HPLC to yield the product.

Toluene-3-sulfonic acid 3-[2-(4-aminomethylphenyl)-ethoxy]phenyl ester (31b): yield 55%; ^{1}H NMR (CD_3COCD_3) δ 2.4 (3H, s), 2.98 (2H, t, J = 6.8), 4.08 (2H, t, J = 6.8), 4.42 (2H, s), 6.59 (1H, ddd, J = 8.3, 2.4, 1.0), 6.86 (1H, ddd, J = 8.3, 2.4, 1.0), 7.21−7.29 (5H, m), 7.50 (1H, t, J = 7.8), 7.58 (1H, d, J = 7.8), 7.63 (1H, d, J = 7.8), 7.67 (1H, s); HRMS calcd for $C_{22}H_{23}NO_4S$, 397.1348; observed, 397.1345.

2,3-Dichlorobenzenesulfonic acid 3-[2-(4-aminomethylphenyl)ethoxy]-5-methylphenyl ester (37b): yield 50%; ^{1}H NMR (CD_3COOD) δ 2.21 (3H, s), 3.03 (2H, t, J = 6.4), 4.09 (2H, t, J = 6.4), 4.23 (2H, s), 6.48 (1H, s), 6.55 (1H, s), 6.64 (1H, s), 7.29−7.45 (5H, m), 7.84 (1H, d, J = 8.1), 7.91 (1H, d, J = 8.1); HRMS calcd for $C_{22}H_{21}Cl_2NO_4S$, 465.0568; observed, 465.0567.

2-Methoxy-5-methylbenzenesulfonic acid 3-[2-(4-aminomethylphenyl)ethoxy]-5-ethylphenyl ester (43b): yield 25%; ^{1}H NMR (CD_3COOD) δ 1.08 (3H, t, J = 7.6), 2.24 (3H, s), 2.50 (2H, q, J = 7.6), 3.04 (2H, t, J = 6.6), 3.95 (3H, s), 4.11 (2H, t, J = 6.6), 4.20 (2H, s), 6.48−6.49 (1H, m), 6.51 (1H, t, J = 2.2), 6.61−6.62 (1H, m), 7.07 (1H, d, J = 8.5), 7.32 (2H, d, J = 8.3), 7.39−7.44 (3H, m), 7.57 (1H, d, J = 2.2); HRMS calcd for $C_{25}H_{29}NO_5S$, 455.1766; observed, 455.1771.

General Procedure for Compounds 28c, 34c, and 40c. BMS (81 μmol) was added dropwise to a stirred solution of **28b** (63 μmol) and dry THF (1 mL). Two additional portions of BMS (2 × 81 μmol) were added after 30 min and 1 h, respectively, and the reaction mixture was kept at room-temperature overnight. Water was added and the organic solvent was removed under reduced pressure. CH_3CN was added to the aqueous phase and the solution was purified immediately by preparative HPLC to yield the product.

Benzenesulfonic acid 3-[2-(4-aminomethyl-3-hydroxyphenyl)ethoxy]phenyl ester (28c): yield 60%; ^{1}H NMR (CD_3OD) δ 2.99 (2H, t, J = 6.4), 4.07−4.11 (4H, m), 6.53−6.57 (2H, m), 6.81−6.89 (3H, m), 7.18−7.26 (2H, m), 7.60−7.66 (2H, m), 7.74−7.80 (1H, m), 7.84−7.87 (2H, m); HRMS calcd for $C_{21}H_{21}NO_5S$, 399.1140; observed, 399.1138.

2,5-Dichlorobenzenesulfonic acid 3-[2-(4-aminomethyl-3-hydroxyphenyl)ethoxy]-5-methylphenyl ester (34c): yield 57%; ^{1}H NMR (CD_3SOCD_3) δ 2.21 (3H, s), 2.83 (2H, t, J = 7.0), 3.80 (2H, s), 4.03 (2H, t, J = 7.0), 6.42−6.43 (1H, m), 6.51−6.52 (1H, m), 6.58−6.60 (2H, m), 6.75 (1H, s), 6.95 (1H, d, J = 7.7), 7.90−7.91 (3H, m); HRMS calcd for $C_{22}H_{21}Cl_2NO_5S$, 481.0518; observed, 481.0520.

2,3-Dichlorobenzenesulfonic acid 3-[2-(4-aminomethyl-3-hydroxyphenyl)ethoxy]-5-ethylphenyl ester (40c): yield 58%; ^{1}H NMR (CD_3COOD) δ 1.08 (3H, t, J = 7.7), 2.51 (2H, q, J = 7.7), 2.97 (2H, t, J = 6.7), 4.08 (2H, t, J = 6.7), 4.21 (2H, s), 6.52−6.53 (2H, m), 6.66 (1H, s), 6.80 (1H, d, J = 7.7), 6.88 (1H, s), 7.20 (1H, d, J = 7.7), 7.40 (1H, t, J = 8.1), 7.83 (1H, d, J = 8.1), 7.90 (1H, d, J = 8.1); HRMS calcd for $C_{23}H_{23}Cl_2NO_5S$, 495.0674; observed, 495.0667.

General Procedure for Compounds 32b and 44b. A solution of ethanol saturated with HCl (2.5 mL) and **32a** (46 μmol) was kept at room temperature for 2 days. The solvent was removed under reduced pressure and 3 mL of ethanol saturated with NH_3 was added. The reaction mixture was stirred at room temperature for 2 days and the solvent was again removed under reduced pressure. The residue was purified by preparative HPLC to yield the product.

3,5-Dichloro-2-methoxybenzenesulfonic acid 3-[2-(4-carbamimidoylphenyl)ethoxy]phenyl ester (32b): yield 36%; ^{1}H NMR (CD_3COOD) δ 3.18 (2H, t, J = 6.4), 4.07 (3H, s), 6.21 (2H, t, J = 6.4), 6.69 (1H, dd, J = 2.2, 1.0), 6.72−6.73 (2H, m), 6.82−6.86 (1H, m), 7.22 (1H, t, J = 8.1), 7.53 (2H, d, J = 8.6), 7.67 (1H, d, J = 2.6), 7.78 (1H, d, J = 2.6), 7.83 (2H, d, J = 8.6); HRMS calcd for $C_{22}H_{20}Cl_2N_2O_5S$, 494.0470; observed, 494.0471.

Benzenesulfonic acid 3-[2-(4-carbamimidoylphenyl)-ethoxy]-5-ethylphenyl ester (44b): yield 37%; ^{1}H NMR (CD_3COOD) δ 1.06 (3H, t, J = 7.6), 2.42 (2H, q, J = 7.6), 3.15 (2H, t, J = 6.3), 4.16 (2H, t, J = 6.3), 6.33 (1H, s), 6.45 (1H, t, J = 2.2), 6.65 (1H, s), 7.50−7.59 (4H, m), 7.70 (1H, t, J = 7.3), 7.84 (4H, t, J = 8.5); HRMS calcd for $C_{23}H_{24}N_2O_4S$, 424.1457; observed, 424.1458.

General Procedure for Compounds 29c, 35c, and 41c. A solution of ethanol saturated with HCl (3 mL) and **29b** (121 μmol) was kept at room temperature for 5 days. The solvent was removed under reduced pressure and 3 mL of ethanol saturated with NH_3 was added. The reaction mixture was stirred at room temperature for 6 h and the solvent was again removed under reduced pressure. The residue was purified by preparative HPLC to yield the product.

2-Methoxy-5-methylbenzenesulfonic acid 3-[2-(4-carbamimidoyl-3-hydroxyphenyl)ethoxy]phenyl ester (29c): yield 34%; ^{1}H NMR (CD_3COOD) δ 2.23 (3H, s), 3.04 (2H, t, J = 6.1), 3.94 (3H, s), 4.15 (2H, t, J06.1), 6.65 (1H, dd, J = 8.3, 2.2), 6.71 (1H, t, J = 2.2), 6.78 (1H, dd, J = 8.3, 2.2), 6.97 (1H, d, J = 8.3), 7.06−7.18 (3H, m), 7.42 (1H, d, J = 8.3), 7.56 (1H, s), 7.70 (1H, d, J = 8.3); HRMS calcd for $C_{23}H_{24}N_2O_6S$, 456.1355; observed, 456.1362.

Toluene-3-sulfonic acid 3-[2-(4-carbamimidoyl-3-hydroxyphenyl)ethoxy]-5-methylphenyl ester (35c): yield 29%; ^{1}H NMR (CD_3COOD) δ 2.19 (3H, s), 2.38 (3H, s), 3.02 (2H, t, J = 6.4), 4.09 (2H, t, J = 6.4), 6.37−6.39 (2H, m), 6.63 (1H, s), 6.86 (1H, dd, J = 8.4, 1.2), 7.09 (1H, d, J = 1.2), 7.40−7.52 (2H, m), 7.62−7.73 (3H, m); HRMS calcd for $C_{23}H_{24}N_2O_5S$, 440.1406; observed, 440.1406.

3,5-Dichloro-2-methoxybenzenesulfonic acid 3-[2-(4-carbamimidoyl-3-hydroxyphenyl)ethoxy]-5-ethylphenyl ester (41c): yield 28%; ^{1}H NMR (CD_3COOD) δ 1.09 (3H, t, J = 7.7), 2.52 (2H, q, J = 7.7), 3.06 (2H, t, J = 6.4), 4.06 (3H, s), 4.15 (2H, t, J = 6.4), 6.52−6.56 (2H, m), 6.67 (1H, s), 6.99, 1H, d, J = 8. 1), 7.08 (1H, s), 7.65 (1H, d, J = 2.5), 7.71 (1H, d, J = 8.1), 7.78 (1H, d, J = 2.5); HRMS calcd for $C_{24}H_{24}Cl_2N_2O_6S$, 538.0732; observed, 538.0727.

Determination of IC_{50} of Thrombin. A stock solution of 60 NIH U/mL human α-thrombin (cat. no. T6759, Sigma, St Louis, MO) in buffer (0.05 M Tris-HCl, pH 7.4, I = 0.15, 1 g/L bovine serum albumin) was frozen at −70 °C. A working solution was freshly prepared by dilution of the stock solution with the assay buffer to a nominal concentration of 10 NIH U/mL, which corresponds to an assay concentration of 0.02 NIH U/mL.

The 18 compounds were each dissolved to a concentration of 0.1 mM in dimethyl sulfoxide (DMSO). These solutions were then prepared to give assay concentrations in the range between 0.06 nM and 133 μM.

The chromogenic substrate pyroGlu-Pro-pNA (S-2366, Chromogenix AB, Mölndal, Sweden) was dissolved in water to a

concentration of 10 mM and further diluted with the assay buffer to 3.75 mM (assay concentration, 0.3 mM).

The thrombin robot assays were performed in a ROBOSYS Plato 3000 Robotic microplate processor in total volumes of 150 μM. The inhibitor (126 μL) and thrombin (12 μL) were mixed, and the reaction was started by adding 12 μL of the substrate working solution. The absorbance was monitored at 405 nm after 40 min of incubation. The increase in absorbance (405 nm) was linear during the incubation time. The rate dA/dt can therefore be approximated from ΔA 405/40 min, and K_i can be estimated from the relation $K_i = IC_{50}/(1 + [S]/K_m)$, where $[S] = 0.3$ mM and $K_m = 0.33$ mM. The compound concentrations that inhibit 50% of thrombin (IC_{50}) activity were determined from dose−response curves using 10 inhibitor concentrations.

Determination of Dissociation Constant (K_i) of Trypsin. A stock solution of 1 g/L bovine trypsin (cat. no. T8003, Sigma, St Louis, MO) in 1 mM HCl and 1 mM $CaCl_2$ was frozen in aliquots at −70 °C. For the steady state studies, a working solution was freshly prepared by dilution of the stock solution with a buffer (0.05 mol/L Tris-HCl, pH 8.2 at 20 °C, $CaCl_2$ 20 mM) to a final trypsin concentration of 0.25 mg/L.

The inhibitors were dissolved by adding 2% and 0.2% ethanol, respectively, for compounds **29c** and **32b**, and by adding 2.8% and 2.6% 2-propanol, respectively, for compounds **35c** and **43b**. The solutions were diluted using the trypsin buffer to final stock concentrations of 100 μM for compounds **29c**, **35c**, and **43b** and 10 μM for compound **32b**. The stock solutions were prepared to give seven assay concentrations of the inhibitors (0.00, 0.01, 0.02, 0.04, 0.06, 0.08, and 0.10 μM).

The chromogenic substrate Bz-Ile-Glu-(−OR)-Gly-Arg-NH-pNA·HCl; R = 50% H/50% CH_3 (S-2222, Chromogenix AB, Mölndal, Sweden) was dissolved in water to give three concentrations, 0.15, 0.25, and 0.50 mM (assay concentrations, 15, 25, and 50 μM, respectively).

The measurements were carried out using a Cobas Bio centrifugal spectroanalyzer (Roche, Basel, Switzerland) at 37 °C. The inhibitor (90 μL) and trypsin (250 μL) were incubated for 5 min. The reaction was started by adding 40 μL of the substrate and the change in absorbance at 405 nm of the mixture was monitored every 10th second for 110 s. The K_i values were calculated from the best fitting straight line obtained by linear regression analysis.

Determination of CE log k' and pK_a. Two different experimental descriptors were used to estimate the tendency of the test compounds to be passively absorbed: pK_a and migration behavior in micellar media (CE log k'). Both descriptors were determined by means of capillary electrophoresis (CE) separations. The use of micelles as separation media in CE-MEKC mode experimentally models absorption to a lipid-like pseudostationary phase. In this study, all 18 compounds were subjected to MEKC separations using a sodium cholate (SC) bile acid micellar electrolyte. From these measurements retention factors, CE log k' values, were calculated. pK_a values were measured for the compounds that contain basic residues.

Capillary electrophoresis (CE) was carried out on an HP 3D capillary electrophoresis system (Agilent technologies, Waldbronn, Germany) with a diode-array detector. Untreated fused-silica capillaries from Polymicro Technologies of 50 μm i.d., 375 μm o.d., and 26.5 cm effective length were used for separations. The applied voltage was 300 V/cm and the temperature of the capillary was maintained at 37.0 ± 0.1 °C for log k' measurements and 22.2 ± 0.1 °C for the pK_a measurements. Prior to each analytical separation the capillary was rinsed consecutively with 0.1 M NaOH, purified water, and electrolyte. Sample solutions were injected by hydrodynamic injection at 20 mbar (5 s). The MEKC micellar electrolyte was 100 mM sodium cholate (SC) in sodium phosphate buffer (pH 7.4, $I = 0.05$). Standard solutions comprised the respective test compound and a set of marker compounds for the measurements of CE log k'.

Determination of pK_a was carried out using a set of 12 buffers covering a pH range from 3.1 to 12.5. The respective pK_a values were calculated from electrophoretic mobility data, analogously to the procedure described by Gluck and co-workers.[54]

Acknowledgment. We are grateful to Dr. Angela Menschik-Lundin for performing the biological measurements, Dr. Anita Fridström for pK_a measurements, and Drs. Johanna Deinum and Tord Inghardt for valuable biological information and discussions. We are also grateful to Dr. Karl-Erik Karlsson for recording the HRMS spectra. Financial support from the Swedish Natural Science Research Council (NFR) and Astra-Zeneca R&D, Mölndal, is gratefully acknowledged.

References

(1) Fenton, J. W., II Thrombin specificity. *Ann. N. Y. Acad. Sci.* **1981**, *370*, 468−495.
(2) Coughlin, S. R.; Vu, T.-K. H.; Hung, D. T.; Wheaton, V. I. Characterization of a functional thrombin receptor. *J. Clin. Invest.* **1992**, *89*, 351−355.
(3) Van Obberghen-Schilling, E.; Pouysséggur, J. Signaling pathways of the thrombin receptor. *Thromb. Haemostasis* **1993**, *70*, 163−167.
(4) Bajusz, S.; Barabas, E.; Tolnay, P.; Szell, E.; Bagdy, D. Inhibition of thrombin and trypsin by tripeptide aldehydes. *Int. J. Pept. Protein Res.* **1978**, *123*, 217−221.
(5) Kettner, C.; Shaw, E. D-Phe-Pro-Arg CH2Cl−A selective affinity label for thrombin. *Thrombosis Res.* **1979**, *14*, 969−973.
(6) Stürzebecher, J.; Markwardt, F.; Voigt, B.; Wagner, G.; Walsmann, P. Cyclic amides of N-alpha-arylsulfonylaminocylated 4-amidinophenylalanine−Tight binding inhibitors of thrombin. *Thrombosis Res.* **1983**, *29*, 635−642.
(7) Stürzebecher, J.; Walsmann, P.; Voigt, B.; Wagner, G. Inhibition of bovine and human thrombins by derivatives of benzamidine. *Thrombosis Res.* **1984**, *36*, 457−465.
(8) Kam, C.-M.; Fujikawa, K.; Powers, J. C. Mechanism-based isocoumarin Inhibitors for trypsin and blood coagulation serine proteases−New anticoagulants. *Biochemistry* **1988**, *27*, 2547−2557.
(9) Lee, S. L.; Alexander, R. S.; Smallwood, A.; Trievel, R.; Mersinger, L. et al. New inhibitors of thrombin and other trypsin-like proteases: Hydrogen bonding of an aromatic cyano group with a backbone amide of the P1 binding site replaces binding of a basic side chain. *Biochemistry* **1997**, *36*, 13180−13186.
(10) Tucker, T. J.; Brady, S. F.; Lumma, W. C.; Lewis, S. D.; Gardell, S. J. et al. Design and synthesis of a series of potent and orally bioavailable noncovalent thrombin inhibitors that utilize non-basic groups in the P1 position. *J. Med. Chem.* **1998**, *41*, 3210−3219.
(11) Bachand, B.; DiMaio, J.; Siddiqui, M. A. Synthesis and structure−activity relationship of potent bicyclic lactam thrombin inhibitors. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 913−918.
(12) Wiley: M. R.; Weir, L. C.; Briggs, S. L.; Chirgadze, N. Y.; Clawson, D. et al. The design of potent, selective, noncovalent, peptide thrombin inhibitors utilizing imidazole as a S1 binding element. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 2767−2772.
(13) Sall, D. J.; Bailey, D. L.; Bastian, J. A.; Buben, J. A.; Chirgadze, N. Y. et al. Diamino benzo[*b*]thiophene derivatives as a novel class of active site directed thrombin inhibitors. 5. Potency, efficacy, and pharmacokinetic properties of modified C-3 side chain derivatives. *J. Med. Chem.* **2000**, *43*, 649−663.
(14) 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 strucutre of two thrombin-inhibitor complex. *Chem. Biol.* **1997**, *4*, 287−295.
(15) Menear, K. Progress towards the discovery of orally active thrombin inhibitors. *Curr. Med. Chem.* **1998**, *5*, 457−468.
(16) 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.
(17) Sanderson, P. E. J. Small, noncovalent serine protease inhibitors. *Med. Res. Rev.* **1999**, *19*, 179−197.
(18) Martin, E. J.; Blaney, J. M.; Siani, M. A.; Spellmeyer, D. C.; Wong, A. K. et al. Measuring diversity: Experimental design of combinatorial libraries for drug discovery. *J. Med. Chem.* **1995**, *38*, 1431−1436.
(19) Lundstedt, T.; Clementi, S.; Cruciani, G.; Pastor, M.; Kettaneh, N. et al. Intelligent Combinatorial Libraries. *Computer-Assisted Lead Finding and Optimization. Current Tools for Medicinal Chemistry*; Verlag Helvetica Chimica Acta: Basel, 1997; pp 190−208.
(20) Linusson, A.; Wold, S.; Norden, B. Statistical molecular design of peptoid libraries. *Mol. Divers.* **1999**, *4*, 103−114.

(21) Gillet, V. J.; Willet, P.; Bradshaw, J. The effectiveness of reactant pools for generating structurally diverse combinatorial libraries. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 731−740.
(22) Jamois, E. A.; Hassan, M.; Waldman, M. Evaluation of reagent-based and product-based strategies in the design of combinatorial library subsets. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 63−70.
(23) Linusson, A.; Gottfries, J.; Lindgren, F.; Wold, S. Statistical molecular design of building blocks for combinatorial chemistry. *J. Med. Chem.* **2000**, *43*, 1320−1328.
(24) Spellmeyer, D. C.; Blaney, J. M.; Martin, E. Computational approaches to chemical libraries. *Practical application of computer-aided drug design*; Marcel Dekker Inc.: New York, 1997; pp 165−193.
(25) Drewry, D. H.; Young, S. S. Approaches to the design of combinatorial libraries. *Chemometrics Intell. Lab. Syst.* **1999**, *48*, 1−20.
(26) Bode, W.; Mayr, I.; Baumann, U.; Huber, R.; Stone, S. R. et al. The refined 1.9 Å crystal structure of human alpha-thrombin: Interaction with D-Phe-Pro-Arg chloromethyl ketone and significance of Tyr-Pro-Pro-Trp insertion segment. *EMBO J.* **1989**, *8*, 3467−3475.
(27) Bode, W.; Turk, D.; Karshikov, A. The refined 1.9-Å X-ray crystal structure of D-Phe-Pro-Arg chloromethyl ketone-inhibited human alpha-thrombin: Structure analysis, overall structure, electrostatic properties, detailed active site geometry, and structure−function relationship. *Protein Sci.* **1992**, *1*, 426−471.
(28) Fenton, J. W., II; Ofosu, F. A.; Moon, D. G.; Maraganore, J. M. Thrombin structure and function: why thrombin is the primary target for antithrombotics. *Blood Coagulation Fibrinolysis* **1991**, *2*, 69−75.
(29) Olsson, T.; Sherbukhin, V. *Synthesis and Structure Administration (SaSA)*; AstraZeneca R&D Mölndal: S-431 83 Mölndal, Sweden.
(30) Andersson, P. M.; Sjöström, M.; Wold, S.; Lundstedt, T. Comparison between physicochemical and calculated molecular descriptors. *J. Chemom.* **2000**, *14*, 629−642.
(31) Oprea, T.; Gottfries, J. Chemography, the art of navigating in the chemical space. *J. Comb. Chem.* **2001**, *3*, 157−166.
(32) *Concord*; Tripos Inc.: 1699 South Hanley Rd, St. Louis, MO 63144-2913 USA.
(33) *UNIX Spartan 4.0*; Wavefunction, Inc.: 18401 Von Karman Ave., Suite 370, Irvine, CA 92612, USA.
(34) Johnson, M. E.; Nachtsheim, C. J. Some guidlines for constructing exact D-optimal designs on convex design spaces. *Technometrics* **1983**, *25*, 271−277.
(35) *Modde 4.0*; Umetrics: Box 7960, S-907 19 Umeå, Sweden.
(36) Jackson, J. E. *A user's guide to principal components*; Wiley: New York, 1991.
(37) Höskuldsson, A. PLS regression methods. *J. Chemom.* **1988**, *2*, 211−228.
(38) Wold, S. PLS for multivariate linear modeling. *Chemometric methods in molecular design*; VCH: Weinheim, 1995; pp 195−218.
(39) Stone, M. Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc.* **1974**, *36 B*, 111−133.
(40) Wold, S. Cross-validatory estimation of the number of components in factor and principal component models. *Technometrics* **1978**, *20*, 397−405.
(41) *Simca-P 7.01*; Umetrics: Box 7960, S-907 19 Umeå, Sweden.
(42) Trone, M. D.; Leonard, M. S.; Khaledi, M. G. Congeneric behavior in estimations of octanol−water partition coefficients by micellar electrokinetic chromatography. *Anal. Chem.* **2000**, *72*, 1228−1235.
(43) Yang, S. Y.; Bumgarner, J. G.; Kruk, L. F. R.; Khaledi, M. G. Quantitative structure−activity relationships studies with micellar electrokinetic chromatography—Influence of surfactant type and mixed micelles on estimation of hydrophobicity and bioavailability. *J. Chromatogr. A* **1996**, *721*, 323−335.
(44) Örnskov, E.; Gottfries, J.; Folestad, S. Correlation of drug absorption with migration data from capillary electrophoresis using micellar electrolytes. *In preparation*.
(45) Woodrow, B.; Dorsey, J. Thermodynamics of micelle-water partitioning in micellar electrokinetic chromatography—comparisons with 1-octanol−water partitioning and biopartitioning. *Environ. Sci. Technol.* **1997**, *31*, 2812−2820.
(46) Örnskov, E.; Markides, K. E.; Folestad, S. Determination of dissociation constants of labile drug compounds by capillary electrophoresis. Paper 160. *21st International Symposium on Capillary Chromatography and Electrophoresis*: Park City, Utah, June 20−24, 1999.
(47) Unpublished result; AstraZeneca R&D Mölndal: S-431 83 Mölndal, Sweden.
(48) Box, G. E. P.; Hunter, W. G.; Hunter, J. S. *Statistics for experimenters. An introduction to design, data analysis, and model building*; John Wiley & Sons: New York, 1978.
(49) Karmarkar, S. N.; Kelkar, S. L.; Wadia, M. S. A simple unusual one-step conversion of aromatic aldehydes into nitriles. *Synth. Comm.* **1985**, 510−512.
(50) *ACD/Lab*; Advanced Chemistry Development Inc.: 90 Adelaide Street West, Suite 702 Toronto, Ontario M5H 3V9, Canada.
(51) Breton, G. W. Selective monoacetylation of unsymmetrical diols catalyzed by silica gel-supported sodium hydrogen sulfate. *J. Org. Chem.* **1997**, *62*, 8952−8954.
(52) Original procedure: Wang, R. X.; You, X. Z.; Meng, Q. J.; Mintz, E. A.; Bu, X. R. A modified synthesis of O-hydroxyaryl aldehydes. *Synth. Commun.* **1994**, *24*, 1757−1760.
(53) Original procedure: Aoyama, T.; Shioiri, T. Trimethylsilyldiazomethane: A convenient reagent for the O-methylation of alcohols. *Tetrahedron Lett.* **1990**, *31*, 5507−5508.
(54) Cleveland, J. A.; Benko, M. H.; Gluck, S. J.; Walbroehl, Y. M. Automated pK_a determination at low solute concentrations by capillary electrophoresis. *J. Chromatogr. A* **1993**, *652*, 301−308.

JM010833F