

# Three-Dimensional Quantitative Structure–Activity Relationship Analyses Using Comparative Molecular Field Analysis and Comparative Molecular Similarity Indices Analysis To Elucidate Selectivity Differences of Inhibitors Binding to Trypsin, Thrombin, and Factor Xa

Markus Böhm,<sup>†</sup> Jörg Stürzebecher,<sup>‡</sup> and Gerhard Klebe<sup>\*,†</sup>

Department of Pharmaceutical Chemistry, University of Marburg, Marbacher Weg 6, D-35032 Marburg, Germany, and University of Jena, Center for Vascular Biology and Medicine, Nordhäuser Strasse 78, D-99089 Erfurt, Germany

Received July 27, 1998

Three-dimensional quantitative structure–activity relationship (3D QSAR) methods were applied using a training set of 72 inhibitors of the benzamidine type with respect to their binding affinities ( $K_i$  values) toward thrombin, trypsin, and factor Xa to yield statistically reliable models of good predictive power. Two methods were compared: the widely used comparative molecular field analysis (CoMFA) and the recently reported CoMSIA approach (comparative molecular similarity indices analysis). CoMSIA produced significantly better results for all correlations. Furthermore, in contrast to CoMFA, CoMSIA is not sensitive to changes in orientation of the superimposed molecules in the lattice. The correlation results obtained by CoMSIA were graphically interpreted in terms of field contribution maps allowing physicochemical properties relevant for binding to be easily mapped back onto molecular structures. The advantage of this feature is demonstrated using the maps to design new molecules. Finally, the CoMSIA method was applied to elucidate structural features among ligands which are responsible for affinity differences toward thrombin and trypsin. These selectivity-determining features were interpreted graphically in terms of spatial regions responsible for affinity discrimination. Such indicators are highly informative for the lead optimization process with respect to selectivity enhancement.

## Introduction

The blood clotting cascade is a complex and highly controlled system of several sequential and feedback-regulated processes.<sup>1</sup> A blood clot is formed through the action of a bifurcated cascade of proteolytic reactions involving nearly 20 different glycoproteins or factors. Some of these clotting factors are inactive serine proteinase zymogens that are proteolytically activated by serine proteinases further up in the cascade. The different serine proteinases involved are structurally related,<sup>2</sup> and several of them have been selected as potential targets for drug development, e.g., toward thrombosis. Inhibition at different levels of the cascade, e.g., at thrombin, factor Xa, factor VIIa, factor IXa, or factor XIa, will have a distinct impact on the cascade, resulting in varying therapeutic profiles for specific inhibitors. The structural relationship among these proteinases is also reflected in the shape and properties of inhibitors acting upon them. Accordingly, since high selectivity of these inhibitors is desired, powerful tools to elucidate the features that control selectivity are of utmost importance. Three-dimensional quantitative structure–activity relationship (3D QSAR) methods can be used to extract such criteria from a set of structurally related inhibitors.

In principle, a 3D QSAR study should have two functions: the derivation of a statistically significant

and highly predictive QSAR model that allows to estimate and rank new compounds to be synthesized and, for the design process even more important, the provision of an easily interpretable graphical tool that denotes those areas of known inhibitors that require a particular physicochemical property to increase affinity and selectivity.<sup>3</sup> These properties might be steric bulk, partial charge, local hydrophobicity, or hydrogen-bond donor or acceptor facilities.

In this paper, two 3D QSAR methods were applied: in both approaches, molecular property fields are evaluated between a probe atom and each molecule of a data set at the intersections of a regularly spaced grid. The widely used CoMFA (comparative molecular field analysis) method<sup>4</sup> calculates steric and electrostatic properties according to Lennard–Jones and Coulomb potentials. CoMSIA (comparative molecular similarity indices analysis)<sup>5</sup> is an alternative approach to perform 3D QSAR by comparative molecular field analysis. Molecular similarity is compared in terms of similarity indices. This method allows the consideration of various physicochemical properties, and the resulting contribution maps can be intuitively interpreted. They can also be used to map and pin down those features responsible for selectivity differences among ligands.

A data set of 72 inhibitors derived from 3-amidinophenylalanine is selected for which detailed  $K_i$  values for their inhibition of trypsin, thrombin, and factor Xa have been reported.<sup>6–9</sup> In all cases we develop statistically significant CoMFA and CoMSIA models. The higher statistical significance and greater robustness of

\* Corresponding author.

<sup>†</sup> University of Marburg.

<sup>‡</sup> University of Jena.

**Table 1.** Substituents at Positions R<sup>1</sup> and R<sup>2</sup> and Formal Charges of the 72 Inhibitors Included in the Training Set<sup>a</sup>

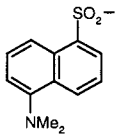
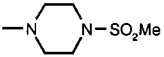
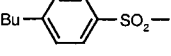
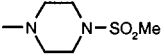
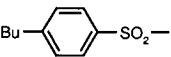
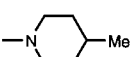
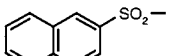
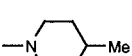
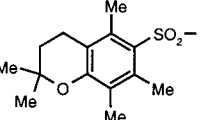
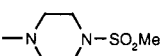
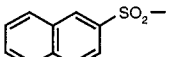
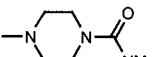
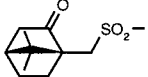
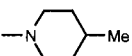
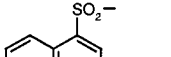
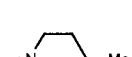
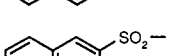
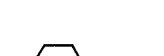
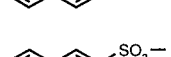
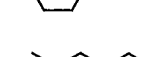
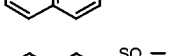
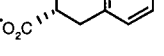
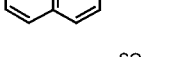
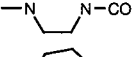
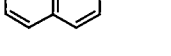
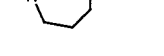
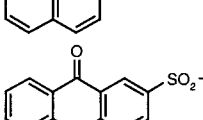
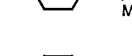
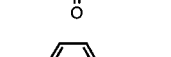
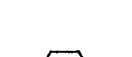
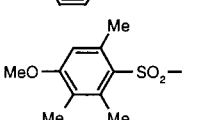
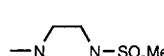
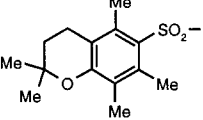
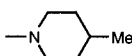
no.	R <sup>1</sup>	R <sup>2</sup>	charge	pK <sub>i</sub>		
				thrombin	trypsin	factor Xa
1			+1	8.377	6.770	5.409
2			+1	8.367	6.796	5.167
3			+1	8.301	6.699	4.921
4			+1	8.208	6.854	4.387
5			+1	8.131	6.119	4.125
6			+1	8.056	6.770	4.620
7			+1	7.854	6.201	4.854
8			+1	7.796	6.201	4.377
9			+1	7.770	7.444	4.367
10			0	7.745	6.886	4.382
11			+1	7.721	7.699	4.114
12			+1	7.678	6.260	4.585
13			+1	7.638	6.854	4.319
14			+1	7.585	7.131	5.638
15			+1	7.585	6.284	4.149
16			+1	7.495	5.745	3.921
17			+1	7.469	6.137	4.097

Table 1 (Continued)

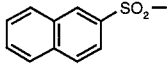
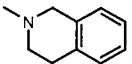
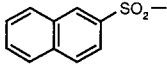
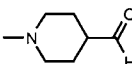
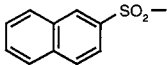
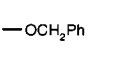
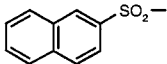
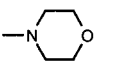
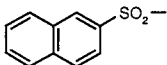
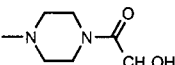
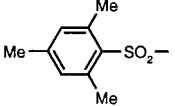
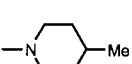
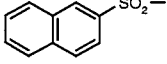
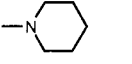
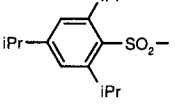
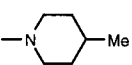
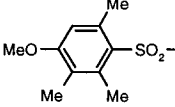
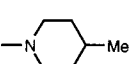
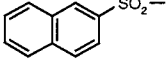
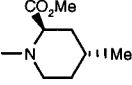
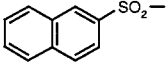
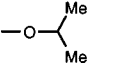
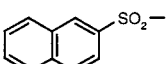
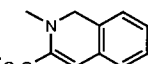
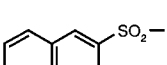
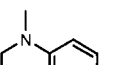
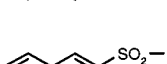
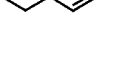
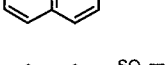
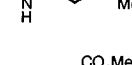
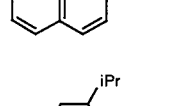
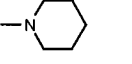
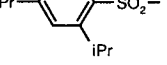
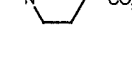
no.	R <sup>1</sup>	R <sup>2</sup>	charge	thrombin	pK <sub>i</sub>	
					trypsin	factor Xa
18			+1	7.432	6.585	4.745
19			+1	7.432	6.658	4.721
20			+1	7.377	6.284	5.658
21			+1	7.377	6.678	4.796
22			+1	7.237	5.959	4.456
23			+1	7.229	5.398	4.022
24			+1	7.187	6.481	4.420
25			+1	7.125	6.161	5.699
26			+1	7.051	6.108	4.000
27			+1	7.018	5.658	4.237
28			+1	6.959	5.854	5.456
29			0	6.921	5.347	4.268
30			+1	6.921	5.824	5.638
31			+1	6.921	5.398	4.328
32			+1	6.824	6.409	4.745
33			+1	6.824	6.569	5.585
34			+1	6.796	6.796	4.119

Table 1 (Continued)

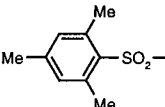
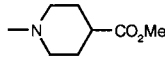
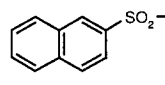
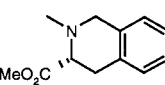
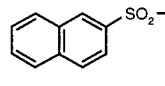
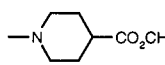
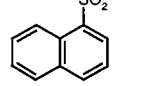
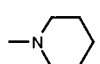
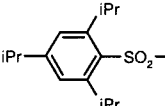
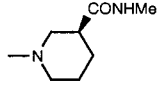
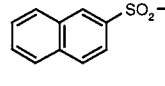
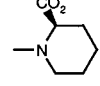
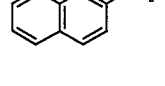
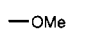
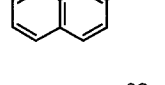
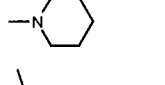
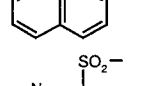
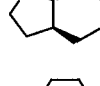
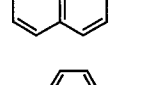
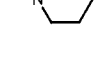
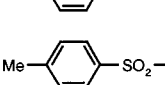
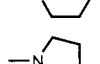
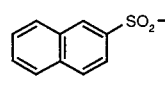
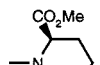
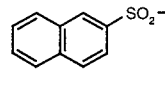
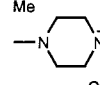
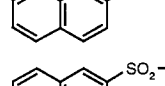
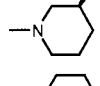
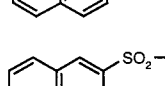
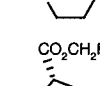
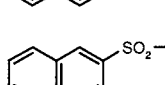
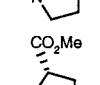




no.	R <sup>1</sup>	R <sup>2</sup>	charge	pK <sub>i</sub>		
				thrombin	trypsin	factor Xa
35			+1	6.745	6.004	4.076
36			+1	6.699	5.921	4.569
37			+1	6.678	7.174	5.092
38			+1	6.638	6.215	4.770
39			+1	6.638	6.102	5.824
40			0	6.585	6.201	4.420
41			+1	6.553	5.602	5.602
42			+1	6.553	6.921	4.770
43			+1	6.495	5.921	4.886
44			+1	6.469	5.444	3.745
45			+1	6.469	5.921	4.824
46			+1	6.456	5.658	4.119
47			+1	6.377	5.678	5.495
48			+2	6.301	6.658	4.678
49			+1	6.292	6.367	4.796
50			0	6.244	6.237	4.363
51			+1	6.201	6.000	4.699
52			+1	6.180	5.092	3.959

Table 1 (Continued)

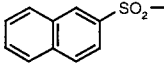
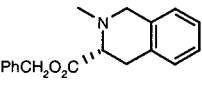
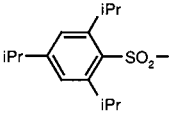
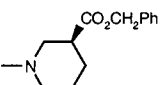
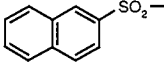

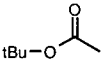
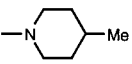
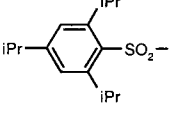
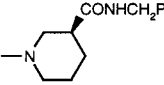
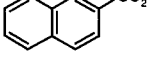
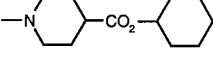
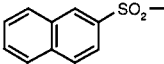

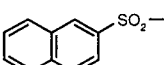

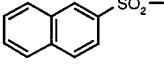
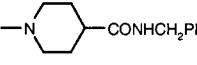
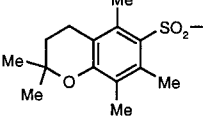

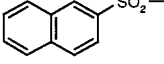

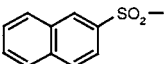

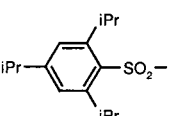

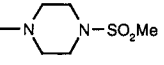
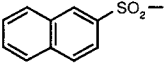
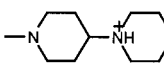
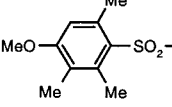

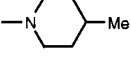
no.	R <sup>1</sup>	R <sup>2</sup>	charge	thrombin	pK <sub>i</sub>	
					trypsin	factor Xa
53			+1	6.161	5.921	5.114
54			+1	6.046	6.071	6.046
55			0	5.959	6.357	4.357
56			+1	5.921	4.854	4.194
57			+1	5.745	6.337	5.602
58			+1	5.678	7.097	5.119
59			+1	5.638	5.102	4.244
60			0	5.538	4.796	3.886
61			+1	5.509	6.569	4.456
62			+1	5.509	4.495	5.000
63			+1	5.244	4.602	4.585
64			+1	5.208	4.796	3.444
65			0	5.137	5.620	4.886
66	H-		+2	4.886	4.538	3.000
67			+2	4.824	6.000	4.658
68			+1	4.770	3.854	3.721
69	H-		+2	4.569	4.538	3.638

Table 1 (Continued)

no.	R <sup>1</sup>	R <sup>2</sup>	charge	pK <sub>i</sub>		
				thrombin	trypsin	factor Xa
70		$-\text{CO}_2^-$	0	4.523	3.928	3.886
71			0	4.456	4.509	4.387
72		$-\text{NHMe}$	+1	4.357	3.000	3.194

<sup>a</sup> The experimental binding affinities toward thrombin, trypsin, and factor Xa are expressed as pK<sub>i</sub> ( $-\log K_i$ ) values; K<sub>i</sub> in mol/L.

the CoMSIA method is demonstrated. The CoMSIA method has been applied by considering five different similarity fields: steric, electrostatic, hydrophobic, hydrogen-bond donor, and hydrogen-bond acceptor properties. The correlation models obtained can be related to the contribution maps, and consequences for a possible design of novel inhibitors are shown. Finally, we used the affinity differences of the ligands in order to derive models to elucidate their selectivity-determining features. Despite the higher relative errors associated with the difference in K<sub>i</sub> values, reliable models are obtained that provide interesting information in the various property maps with respect to selectivity discrimination.

### The CoMSIA Method

Recently, we reported an alternative comparative molecular field analysis, CoMSIA, based on molecular similarity indices.<sup>3,5,10</sup> This method avoids some of the inherent deficiencies arising from the functional form of the Lennard–Jones and Coulomb potentials used in the original version of CoMFA.<sup>4</sup> Both potentials are very steep close to the van der Waals surface and produce singularities at the atomic positions. As a consequence, the potential energy expressed at grid points in the proximity of the surface changes dramatically. However, it is precisely this region that contains the important descriptors for a QSAR analysis.<sup>11</sup> To avoid unacceptably large energy values, the potential evaluations are normally restricted to regions outside the molecules and require the definition of some arbitrarily determined cutoff values. Due to differences in the slope of the Lennard–Jones and Coulomb potentials, these cutoff values are exceeded at different distances from the molecules,<sup>11</sup> requiring further arbitrary scaling of the two fields in a simultaneous evaluation which can involve the loss of information about one of the fields.

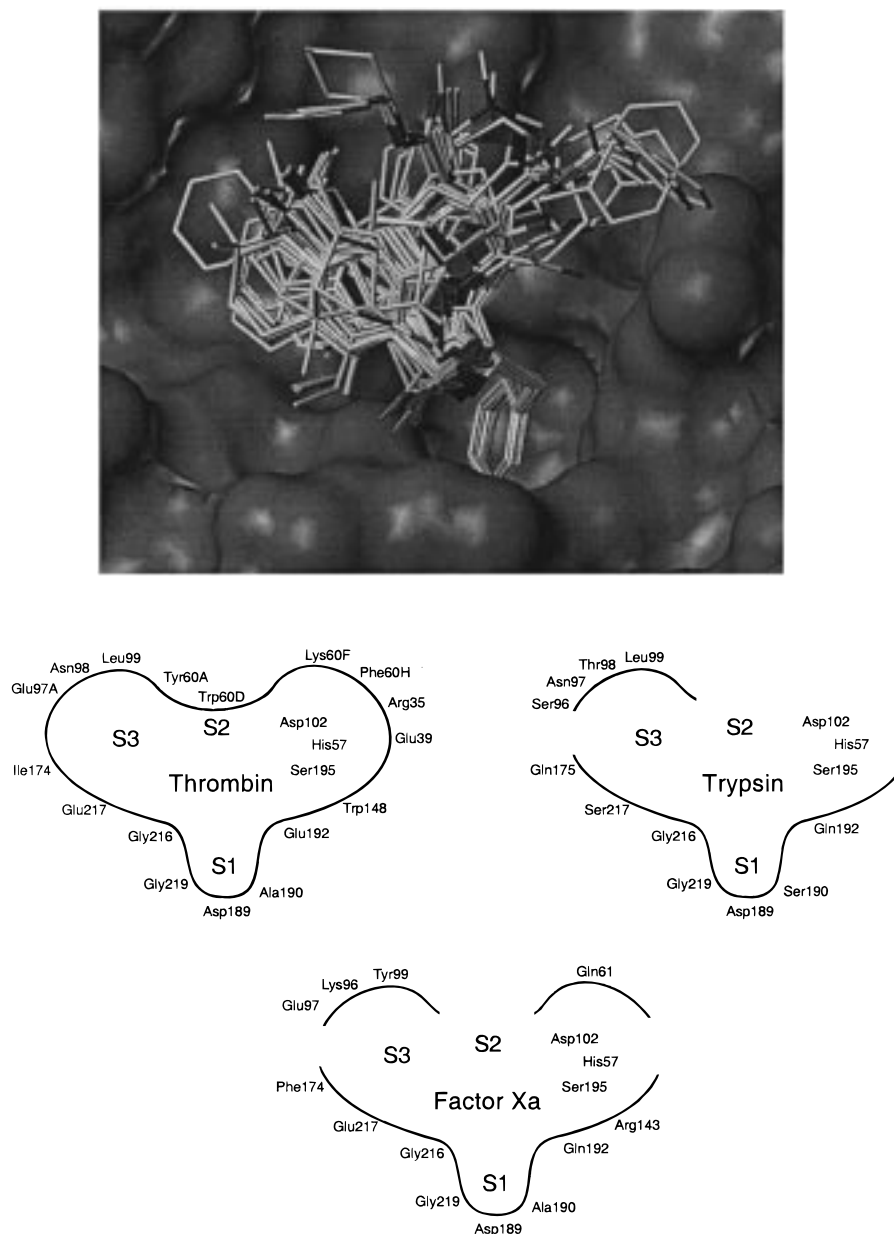
To overcome such problems, CoMSIA evaluates molecular similarity in space. Similarity is expressed in terms of different physicochemical properties: steric occupancy, partial atomic charges, local hydrophobicity, and hydrogen-bond donor and acceptor properties.<sup>3</sup> Using a common probe atom, similarity indices are calculated for a data set of prealigned molecules at regularly spaced grid points. In determining such similarities, the mutual distances between the probe and the atoms of the molecules of the data set are considered. For this distance dependence, a Gaussian-

type functional form is selected that avoids singularities at the atomic positions and requires no arbitrary definition of cutoff limits. The similarity indices can be calculated at all grid points inside as well as outside the molecules and are subsequently evaluated in a PLS analysis<sup>12,13</sup> following the usual CoMFA protocol.

The CoMSIA approach presently considers five different property fields. The purpose of using five (or more) different fields is not to increase the significance and predictive power of the 3D QSAR models. Primarily, the intention is to partition the various properties into spatial locations where they play a decisive role in determining biological activity. This aspect is of utmost importance if a targeted optimization of molecules in a design program is anticipated.

At this point, the major advantage of CoMSIA compared to usual CoMFA becomes apparent: the better ability to visualize and interpret the obtained correlations in terms of field contributions. Contour maps of the relative spatial contributions (isocontours of the coefficients obtained from PLS analysis) of the different fields are commonly used.<sup>4</sup> They indicate those lattice points where a particular property is weighted highest in order to explain trends in affinity data. However, in CoMFA these maps are often neither contiguous nor smoothly connected. Accordingly, they are difficult to interpret, due mainly to the steepness of the potentials close to the molecular surfaces and the previously described cutoff settings. Using molecular similarity indices, substantially improved contour maps are obtained that are easy to interpret in terms of separate property fields and consequently are intuitive as a visualization tool in designing novel compounds. The level-dependent contouring of CoMFA field contributions highlights those regions in space where the aligned molecules would favorably or unfavorably interact with a possible environment. In contrast, the CoMSIA field contributions denote those areas *within* the region occupied by the ligands that “favor” or “dislike” the presence of a group with a particular physicochemical property. This association of such required properties with a possible ligand shape is a more direct guide to check whether all features important for activity are present in the structures under consideration.



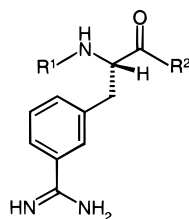


**Figure 1.** Alignment of the 72 inhibitors together with the binding site of trypsin. The solvent-accessible surface is indicated as a solid surface. The three sketches below highlight the relevant amino acids surrounding the binding sites of the related proteins. The catalytic triad is formed by Asp102, His57, Ser195.

## Methods

All molecular modeling and comparative molecular field evaluations were performed using SYBYL<sup>14</sup> versions 6.3 and 6.4 running on a Silicon Graphics O2 (R5000) workstation.

**Data Set and Alignment.** A training set of 72 inhibitors derived from 3-amidinophenylalanine was used for all CoMFA and CoMSIA analyses. Structural variations of the parent structure, present in all molecules, were allowed at the positions R<sup>1</sup> and R<sup>2</sup>.



The considered substituents of the inhibitors are summarized

in Table 1. The protonation state of acidic and basic groups was assumed to be as follows: the benzamidine moiety and basic amino functional groups were protonated, and amides and amino groups adjacent to aromatic portions were treated as uncharged. All carboxylate groups were considered to be deprotonated. To obtain a consistent alignment, the crystal structures of thrombin, trypsin, and factor Xa from the Brookhaven Protein Data Bank (PDB)<sup>15</sup> were used as references (1ETS,<sup>16</sup> 1PPH,<sup>17</sup> and 1HCG<sup>18</sup>). Using the binding geometry of 3-TAPAP (**45**) bound to trypsin (1PPH), the ligands were first placed into the binding site of the latter protein according to the following procedure: the unmodified TAPAP-type parent structure was used as a template for superposition. The R<sup>1</sup> and R<sup>2</sup> substituents were constructed from the SYBYL fragment database and minimized using the TRIPOS force field.<sup>19,20</sup> Conformational flexibility of the substituents was taken into account by a systematic conformational search.<sup>21</sup> By displaying solvent-accessible Connolly surfaces around the binding sites of the three related proteins, sterically unfavored conformations initially fitted to trypsin were detected and subsequently altered to achieve a satisfac-

**Table 2.** Parameters of the Grid Boxes Used for CoMFA and CoMSIA Analyses

	2.0 Å grid			1.5 Å grid			1.0 Å grid		
	<i>x</i>	<i>y</i>	<i>z</i>	<i>x</i>	<i>y</i>	<i>z</i>	<i>x</i>	<i>y</i>	<i>z</i>
lower corner	-6.5	-18.0	-1.0	-6.5	-18.0	-1.0	-6.5	-18.0	-1.0
high corner	19.5	4.0	23.0	19.5	4.0	23.0	19.5	4.0	23.0
step size	2.0	2.0	2.0	1.5	1.5	1.5	1.0	1.0	1.0
number of steps	14	12	13	18	15	17	27	23	25
points	2184			4590			15525		

**Table 3.** Summary of Results from the CoMFA and CoMSIA Analyses

	thrombin		trypsin		factor Xa	
	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA
$q^2$	0.687	0.757	0.629	0.752	0.374	0.594
$S_{press}$	0.594	0.531	0.556	0.469	0.515	0.424
$r^2$	0.881	0.950	0.916	0.972	0.680	0.915
$S$	0.366	0.241	0.264	0.157	0.368	0.194
$F$	124.4	206.6	144.4	240.1	48.2	117.1
components	4	6	5	9	3	6
fraction						
steric	0.624	0.214	0.658	0.167	0.701	0.175
electrostatic	0.376	0.147	0.342	0.158	0.299	0.160
hydrophobic		0.298		0.295		0.345
donor		0.087		0.104		0.090
acceptor		0.253		0.276		0.230
box: stepsize	2 Å					
$q^2$	0.697	0.757	0.609	0.756	0.415	0.589
$S_{press}$	0.584	0.532	0.580	0.467	0.501	0.427
$r^2$	0.873	0.949	0.959	0.972	0.784	0.915
$S$	0.378	0.244	0.189	0.159	0.305	0.194
$F$	115.5	199.6	211.6	235.9	60.6	116.3
components	4	6	7	9	4	6
fraction						
steric	0.608	0.218	0.649	0.172	0.709	0.177
electrostatic	0.392	0.171	0.351	0.176	0.291	0.170
hydrophobic		0.278		0.286		0.334
donor		0.087		0.102		0.090
acceptor		0.246		0.265		0.229
box: stepsize	1.5 Å					
$q^2$	0.691	0.756	0.635	0.754	0.429	0.590
$S_{press}$	0.590	0.532	0.561	0.467	0.495	0.426
$r^2$	0.879	0.949	0.952	0.971	0.798	0.915
$S$	0.370	0.244	0.204	0.159	0.294	0.194
$F$	121.1	200.0	180.2	234.3	66.2	116.5
components	4	6	7	9	4	6
fraction						
steric	0.613	0.217	0.636	0.171	0.701	0.176
electrostatic	0.387	0.172	0.364	0.180	0.299	0.170
hydrophobic		0.279		0.284		0.336
donor		0.088		0.103		0.091
acceptor		0.244		0.262		0.228
box: stepsize	1 Å					

tory alignment. At last, the complete molecules were minimized again using the TRIPOS force field. The structural diversity of the aligned ligands obtained is shown in Figure 1. The finally accepted superposition showed reasonable fit to all three binding pockets. In the following analyses this common alignment has been evaluated. It is likely that the actual binding mode of a particular ligand will deviate slightly among the three related proteins. However, these possible deviations are not considered in our model due to the lack of conclusive information about detailed binding modes. More important and with respect to real-life CoMFA studies where no reference protein structures are available to rationalize such minor alignment differences, we expect our models to be robust and still conclusive enough to cope with these deviations. (The atomic coordinates of all molecules of the data set are available from the authors upon request.)

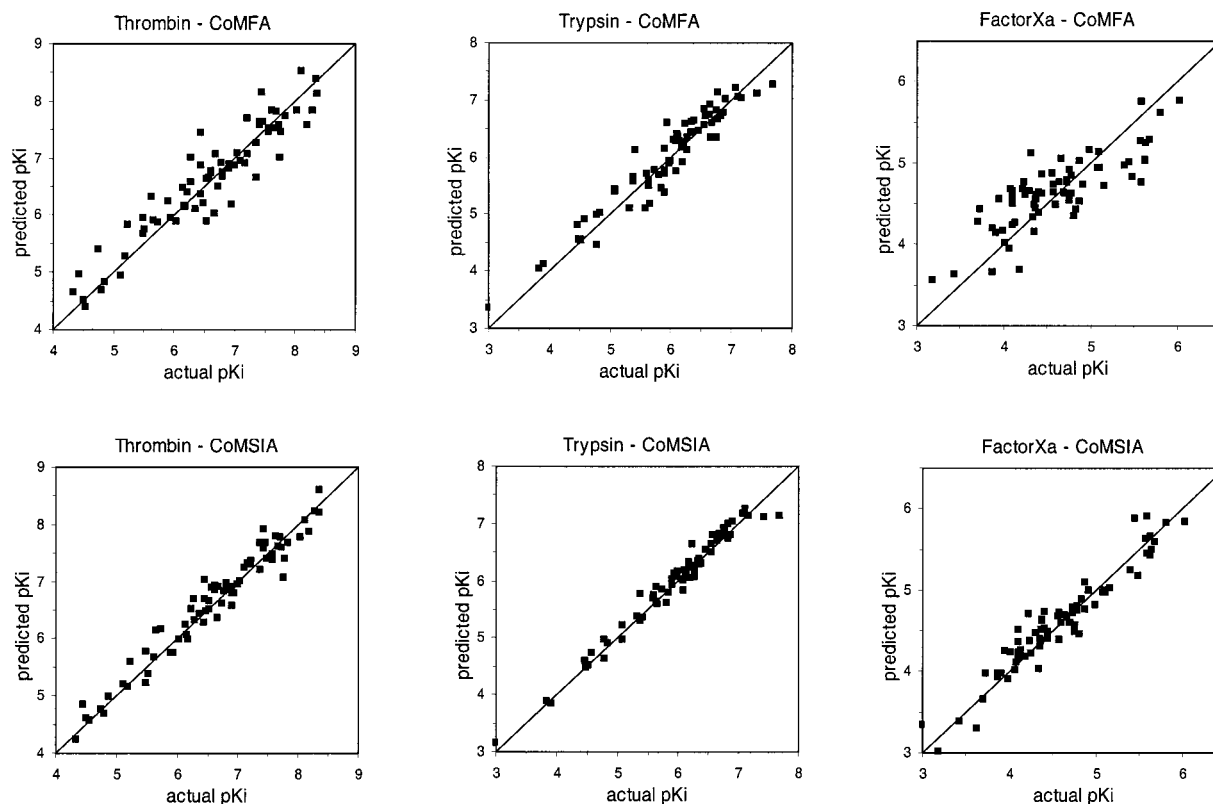
The knowledge of the protein structure as a reference is not a prerequisite to perform 3D QSAR analyses. However, if a crystal structure of the protein is available, a relevant structural alignment of the molecules can be defined with

much higher reliability. Furthermore, it provides the opportunity to interpret features indicated by the contour maps with respect to the protein environment.<sup>3</sup>

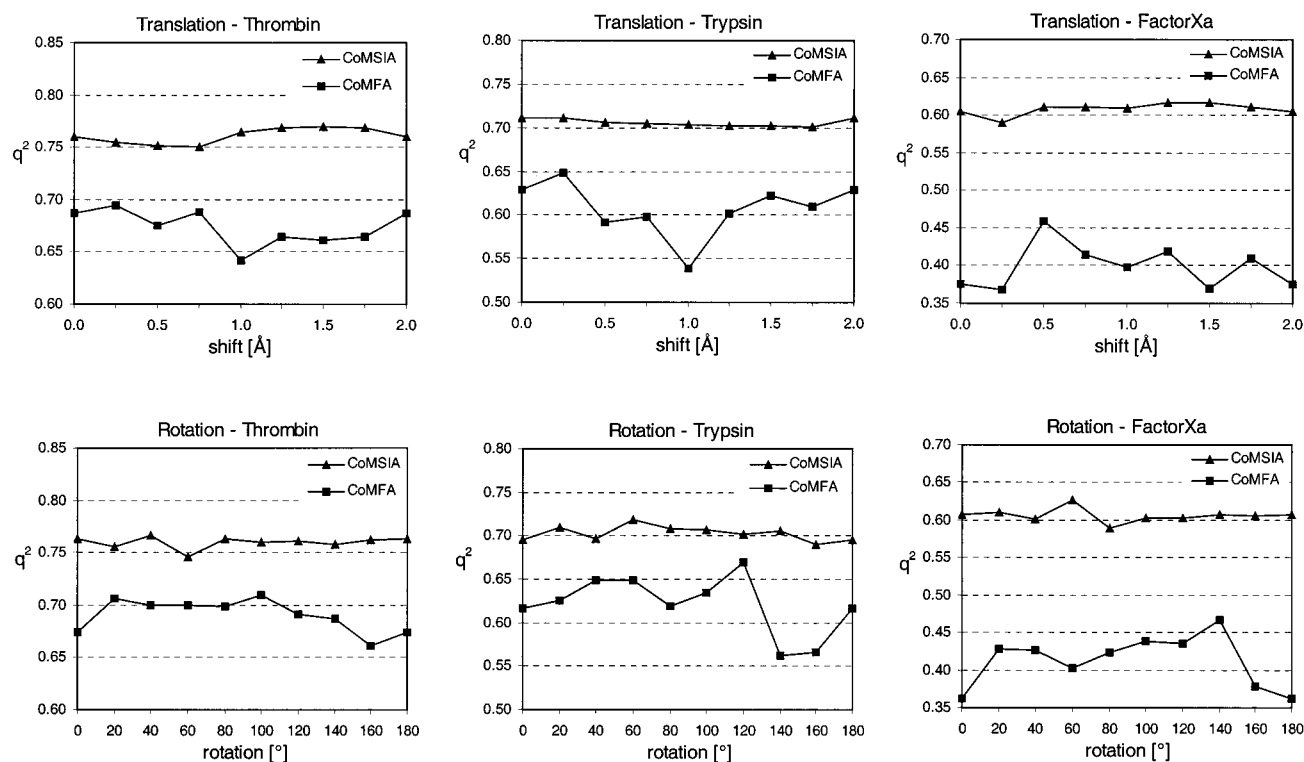
The experimentally determined biological activities of the different inhibitors toward thrombin, trypsin, and factor Xa are given as  $pK_i$  values (Table 1). The affinities toward thrombin and trypsin spread over a satisfactorily large range covering 4.0 and 4.7 logarithmic units, respectively, whereas in the case of factor Xa a variation over 3.0 logarithmic units falls close to the lower limit where the derivation of a statistically significant 3D QSAR model can still be expected. To obtain a better insight into the biological data, we first checked the mutual intercorrelations among the three different data sets. The correlation coefficients of the three combinations are 0.72, 0.28, and 0.46, respectively. The pronounced intercorrelation among the thrombin and trypsin data indicates the closer relationship of the inhibition toward these two serine proteinases.

**CoMFA and CoMSIA Analysis.** Steric and electrostatic CoMFA fields were calculated as implemented in SYBYL<sup>14</sup>





**Figure 2.** Fitted predictions versus actual binding affinities for the 72 inhibitors of the training set. The predicted values were obtained by PLS analyses using the CoMFA and CoMSIA method with 2 Å grid spacing.



**Figure 3.** Dependence of  $q^2$  for CoMFA and CoMSIA on translations and rotations of the entire data set with respect to the surrounding lattice. Translations are in steps of 0.25 Å along the  $xyz$  diagonal; rotations are in steps of 20° around the  $x$ -axis.

using Lennard–Jones and Coulomb potentials, respectively.<sup>4</sup> Partial atomic charges were determined using the AM1 Hamiltonian<sup>22</sup> within the semiempirical package MOPAC 6.0<sup>23,24</sup> taking into account the formal charges as listed in Table 1. All CoMFA calculations were performed with SYBYL standard parameters (TRIPOS standard field, dielectric con-

stant 1/ $r$ , cutoff 30 kcal/mol) using an  $sp^3$  carbon probe atom with a charge of +1.0. Details about the evaluation in CoMSIA are given in our previous papers.<sup>5,10</sup> To shortly summarize, CoMSIA calculates similarity indices at the intersections of a surrounding lattice. The similarity index  $A_r$ <sup>25</sup> for a molecule  $j$  with  $i$  atoms at the grid point  $q$  is determined as follows:

**Table 4.** Five Runs of Cross-Validations Using the "Leave-Five-Out" Procedure<sup>a</sup>

	thrombin			trypsin			factor Xa		
	$q^2$	$s_{press}$	comp	$q^2$	$s_{press}$	comp	$q^2$	$s_{press}$	comp
CoMFA									
run 1	0.666	0.613	4	0.605	0.566	3	0.333	0.535	4
run 2	0.685	0.596	4	0.595	0.573	3	0.381	0.512	3
run 3	0.597	0.679	5	0.648	0.542	5	0.404	0.498	2
run 4	0.680	0.600	4	0.604	0.566	3	0.352	0.524	3
run 5	0.617	0.656	4	0.638	0.549	5	0.360	0.517	2
LOO	0.687	0.594	4	0.629	0.556	5	0.374	0.515	3
CoMSIA									
run 1	0.761	0.530	5	0.760	0.458	8	0.612	0.411	5
run 2	0.772	0.518	5	0.760	0.462	9	0.532	0.452	5
run 3	0.765	0.526	5	0.740	0.481	9	0.594	0.424	6
run 4	0.778	0.507	4	0.757	0.464	9	0.588	0.427	6
run 5	0.785	0.503	5	0.758	0.464	9	0.567	0.438	6
LOO	0.757	0.531	6	0.752	0.469	9	0.594	0.424	6

<sup>a</sup> The "Leave-One-Out" (LOO) results are shown for comparison.

$$A_{F,k}^q(j) = -\sum_i W_{probe,k} W_{ik} e^{-\alpha r_{ij}^2}$$

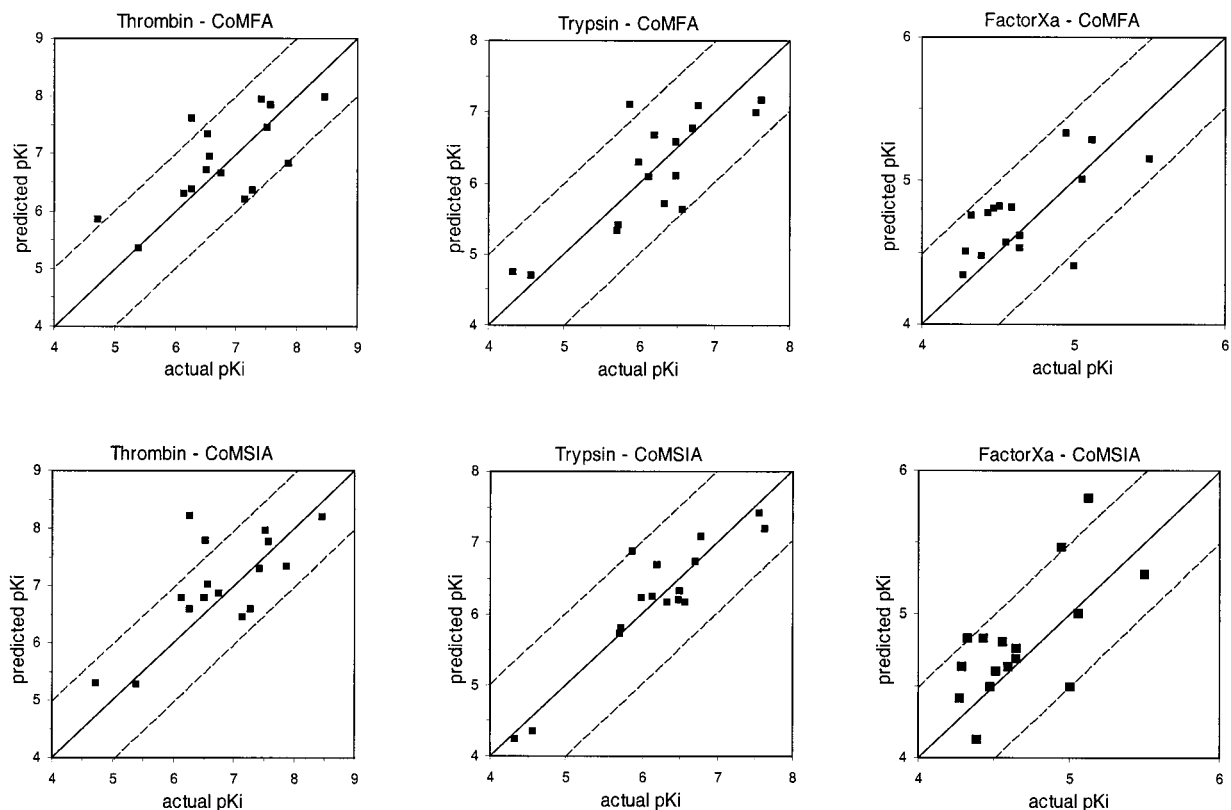
Five physicochemical properties  $k$  (steric, electrostatic, hydrophobic, and hydrogen-bond donor and acceptor) were evaluated, using a common probe atom with 1 Å radius and charge, hydrophobicity, and hydrogen-bond property of +1. A Gaussian-type distance dependence was considered between the grid point  $q$  and each atom  $i$  of the molecule. The value of the so-called attenuation factor  $\alpha$  was initially set to 0.3.

The different CoMSIA fields were calculated with a separate Fortran program.<sup>5,26</sup> Steric property fields were expressed by the third power of the atomic radii. Partial atomic charges attributed to the various atoms in the different molecules were calculated using the AM1 method. Local hydrophobicities were associated using atom-based parameters developed by Viswa-

nadhan et al.<sup>27</sup> Putative positions of possible hydrogen-bond donor and acceptor atoms were assigned by a set of rules derived from experimental values.<sup>28,29</sup> The output from the Fortran program was subsequently transferred to a SYBYL molecular spreadsheet using an SPL macro.

At first, a lattice of 2 Å grid spacing was generated automatically. To allow a comparison of CoMFA and CoMSIA results, the grid box coordinates were changed manually to reveal consistent values. Additional lattices with 1.5 and 1 Å grid spacing with the same orientation were generated (Table 2). CoMFA and CoMSIA fields were subsequently calculated as described above. PLS analyses were performed following the CoMFA standard implementation in SYBYL. The five different descriptor blocks have been scaled to each other using the CoMFA standard scaling option. To check statistical significance of the models, cross-validations were done by means of the "leave-one-out" (LOO) procedure using the enhanced version of PLS, the SAMPLS method.<sup>30</sup> The optimal number of components was determined by selecting the smallest  $s_{PRESS}$  value. Usually this value corresponds to the highest  $q^2$  value. The same number of components was subsequently used to derive the final QSAR models. For all conventional analyses (no cross-validation) the "minimum-sigma" standard deviation threshold was set to 1.0 kcal/mol. The statistical results are summarized in Table 3. The  $q^2$ ,  $s_{PRESS}$ ,  $r^2$ , and  $S$  values were computed as defined in SYBYL. The plots of predicted versus actual binding affinities for the fitted PLS analyses are shown in Figure 2.

Additionally, to perform an even more rigorous statistical test, several runs of a "leave-five-out" procedure were done using the 2 Å lattice spacing (Table 4). In this case, five arbitrarily selected compounds were dropped from the training set and used for prediction. The procedure leaving five molecules out at a time revealed similar  $q^2$  values to the corresponding LOO results. However,  $q^2$  values obtained from the CoMFA method spread over a larger range than those from CoMSIA. In general, the optimal numbers of components for



**Figure 4.** Predicted versus actual binding affinities for the 16 inhibitors not included in the training set. The predicted values were obtained by PLS analyses using the CoMFA and CoMSIA method with 2 Å grid spacing. The dashed lines mark deviations of 1 (factor Xa: 0.5) logarithmic unit from the ideal prediction.

**Table 5.** Substituents at Positions R<sup>1</sup> and R<sup>2</sup> and Formal Charges of the 16 Inhibitors Included in the Test Set<sup>a</sup>

no.	R <sup>1</sup>	R <sup>2</sup>	charge	pK <sub>i</sub>		
				thrombin	trypsin	factor Xa
73			+1	8.481	6.721	4.658
74			+1	7.886	6.585	5.509
75			+1	7.585	6.495	5.013
76			+1	7.523	6.215	4.658
77			+2	7.444	5.886	4.523
78			+1	7.284	6.357	4.301
79			+1	7.155	5.721	4.337
80			+1	6.770	6.149	4.284
81			+1	6.585	6.509	4.481
82			+1	6.553	6.009	5.131
83			+1	6.523	6.796	5.066
84			+1	6.284	7.569	4.569
85			+1	6.284	5.745	4.444
86			+1	6.149	7.638	4.602
87			+1	5.420	4.585	4.959
88			+1	4.745	4.337	4.398

<sup>a</sup> The experimental binding affinities toward thrombin, trypsin, and factor Xa are expressed as pK<sub>i</sub> values.

the final analyses were identical for both the “leave-one-out” and “leave-five-out” procedures.

A common test to check the consistency of the models is to scramble the biological data and repeat the model derivation

process, allowing detection of possible chance correlations. After randomizing our data sets in several distinct ways, in all cases we only observed negative  $q^2$  values in the PLS analyses.

**Table 6.** Predicted Binding Affinities of the 16 Inhibitors Included in the Test Set<sup>a</sup>

no.	thrombin		trypsin		factor Xa	
	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA
73	0.512	0.308	-0.034	-0.001	0.131	-0.023
74	1.081	0.564	0.965	0.442	0.365	0.245
75	-0.234	-0.161	0.408	0.310	0.615	0.531
76	0.084	-0.427	-0.448	-0.464	0.052	-0.093
77	-0.477	0.175	-1.201	-0.977	-0.287	-0.067
78	0.931	0.714	0.668	0.212	-0.200	-0.323
79	0.963	0.718	0.414	0.011	-0.415	-0.484
80	0.132	-0.073	0.070	-0.088	-0.049	-0.120
81	-0.352	-0.411	-0.048	0.197	-0.317	-0.005
82	-0.772	-1.206	-0.265	-0.206	-0.143	-0.668
83	-0.174	-0.233	-0.272	-0.272	0.068	0.070
84	-1.315	-1.905	0.590	0.174	0.009	-0.225
85	-0.086	-0.277	0.355	-0.044	-0.318	-0.380
86	-0.133	-0.610	0.487	0.456	-0.202	-0.022
87	0.099	0.171	-0.089	0.257	-0.365	-0.492
88	-1.092	-0.535	-0.396	0.109	-0.069	0.281
predictive $r^2$	0.470	0.432	0.650	0.842	0.384	0.164

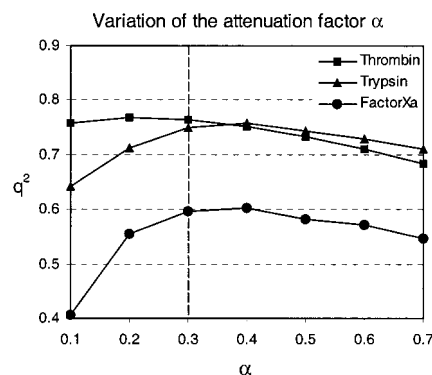
<sup>a</sup> Affinities are given as deviations from the actual values ( $pK_{i\text{ act}} - pK_{i\text{ pred}}$ ).

A well-known problem in CoMFA is that statistical results are dependent on the relative orientation of the molecules of a data set with respect to the lattice.<sup>31,32</sup> To check these inconsistencies and to estimate the robustness of the QSAR models obtained, we translated and rotated the entire data set within the lattice. The translation procedure was performed using the STATIC TRANSLATE command in SYBYL. The original grid box with 2 Å lattice spacing was extended by 2 Å in  $x$ ,  $y$ , and  $z$  directions to guarantee an adequate margin of the lattice surrounding all molecules during the translation. The data set was translated along the  $xyz$  diagonal from 0 to 2 Å in steps of 0.25 Å. Subsequently, the data set was rotated within the lattice using the STATIC ROTATE command. To do so, the data set and the grid box, already used for translation, were moved into the origin of the coordinate system. The molecules were then rotated in steps of 20° around the  $x$ -axis of the grid box. As the grid is orthogonal, a rotation from 0° to 180° was sufficient. CoMFA and CoMSIA fields were calculated, and PLS analyses were subsequently performed after each translation or rotation step. Cross-validated  $q^2$  values were determined for the binding affinities of trypsin, thrombin, and factor Xa (Figure 3).

To check the predictive power of the obtained CoMFA and CoMSIA models, 16 additional inhibitors were selected as a test set (Table 5). The molecules were built and aligned by the same protocol as already described for the inhibitors of the training set. Predictions were performed using the three QSAR models based on the 2 Å lattice spacing. Table 6 summarizes the results obtained from the CoMFA and CoMSIA analyses. The predictive  $r^2$  was calculated according to the definition of Cramer et al.<sup>4</sup> The plots of predicted versus actual binding affinity for the test set molecules are shown in Figure 4.

As mentioned above, similarity indices in CoMSIA are calculated using a Gaussian-type distance dependence. The attenuation factor  $\alpha$  was initially set to 0.3. To decide whether this is an appropriate value,  $\alpha$  was varied in a parameter study within the range from 0.1 to 0.7 in steps of 0.1, and subsequently similarity indices and  $q^2$  values were computed each time (Figure 5).

In CoMSIA five different fields were used for describing the physicochemical properties of the molecules. Since functional group replacement among the considered ligands can affect fields with similar consequences (e.g., removal of a group affects steric and in parallel electrostatic properties or replacement of charged groups influences coherently electrostatic and hydrophobic properties), we checked how much the individual fields contribute to the derived model. All possible combinations of property fields were selected, and for each case a QSAR model was derived. Figure 6 displays the calculated  $q^2$  values of all 31 field combinations.

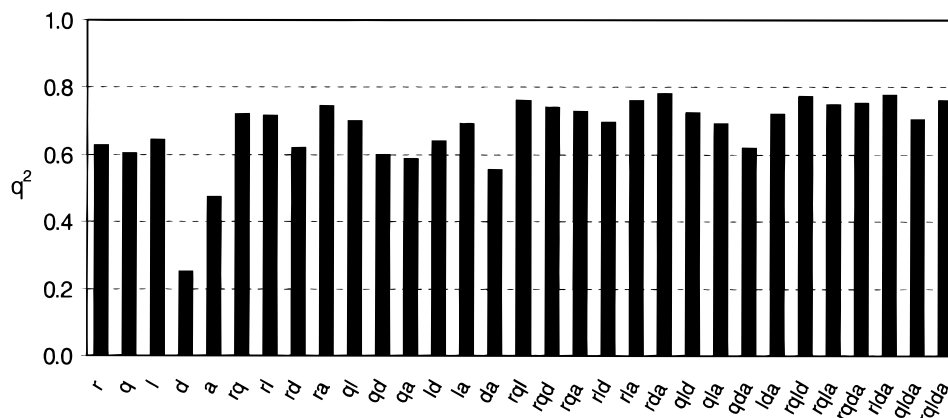


**Figure 5.** Variation of  $q^2$  upon changes of the attenuation factor  $\alpha$  used in the distance dependence between the probe atom and the atoms of the molecules in CoMSIA. For the three considered data sets a value of  $\alpha = 0.3$  is assumed as the best compromise.

The CoMSIA results were graphically interpreted by field contribution maps. We generated coefficient contour maps using the field type "stdev\*coeff". To select appropriate contour levels for each feature, the according histograms of actual field values were analyzed. In an iterative manner a contour level was chosen that produced the best interpretable contour map.<sup>33</sup>

## Discussion

**Predictive Power of the Analyses.** A data set of 72 benzamidine-type inhibitors allows the derivation of three separate QSAR models of statistical significance (Table 3). They provide the successful prediction of binding affinities toward trypsin, thrombin, and factor Xa. A unique alignment has been used for the analyses of the three affinity data sets. Since the affinity data are intercorrelated between 30% and 70%, the derivation of three significant models is not obvious or trivial. In all cases, the CoMSIA analyses reveal significantly better correlations expressed in terms of higher  $q^2$  values (Table 3). Three different grid spacings have been evaluated (1, 1.5, and 2 Å). Whereas for CoMFA some dependence of  $q^2$  on the spacing can be detected, CoMSIA appears to be independent of the lattice grating. Supposedly this can be explained by the greater robustness of QSAR models obtained from the latter method. This assumption is strongly supported by the fact that in the case of CoMSIA  $q^2$  is nearly independent



**Figure 6.** Obtained  $q^2$  values for the 31 possible combinations of the five property fields considered in CoMSIA. Since most of the  $q^2$  values are above 0.6, strong interdependencies among the fields have to be assumed; *r*, *q*, *l*, *d*, and *a* represent the steric, electrostatic, hydrophobic, and hydrogen-bond donor and acceptor property fields, respectively.

of translations and rotations of the superimposed molecules with respect to the lattice, whereas CoMFA produces much stronger and rather unsatisfactory changes in  $q^2$  (Figure 3). These instabilities of CoMFA have been reported by several groups.<sup>31,32</sup> In our opinion, they can be attributed to the shape and steepness of the hyperbolic Lennard–Jones and Coulomb potentials and in consequence to the required arbitrary fixation of cutoff values. For example, it can easily happen for a lattice with 2 Å grid spacing with given spatial orientation that the Lennard–Jones potential at a lattice point  $i+1$  apart from a molecule shows a very small value close to zero. At the lattice point  $i$ , being 2 Å closer to the molecule, the Lennard–Jones potential has increased to a larger value; however, it might still be below the cutoff limit. The contribution at this lattice point could turn out to be significant in the correlation analysis, considering a small shift of the superimposed molecules relative to the surrounding lattice could involve the potential at  $i+1$  to be still rather insignificant. However, at  $i$  the potential can now increase just beyond the cutoff limit. As a consequence, lattice points from this previously important region refrain from any significant variations that can be evaluated in a subsequent correlation analysis. In CoMSIA, at every lattice point steadily increasing or decreasing values are registered and can be correlated. Accordingly a strongly reduced dependency on the lattice orientation and the grid spacing can be expected. Whereas the first aspect gives confidence that the QSAR results will be independent of special modeling conditions, the second has important consequences on the required computational resources: changing from a 2 to 1 Å lattice spacing involves an increase in computing time by a factor of 8.

Table 3 indicates that the numbers of PLS components is higher in CoMSIA compared to CoMFA. Probably this observation results from the significantly higher number of lattice points showing steadily varying field values (e.g., inside the molecules). We selected the optimal numbers of components on the basis of lowest  $SPRESS$ . For all three enzymes a consideration of CoMSIA models with not more than four components does not involve a reduction in  $q^2$  of more than 5%. However, we believe, since the evaluation of the test set molecules underlines convincing predictive power of both methods, that no overfitting of the data is present (see below).

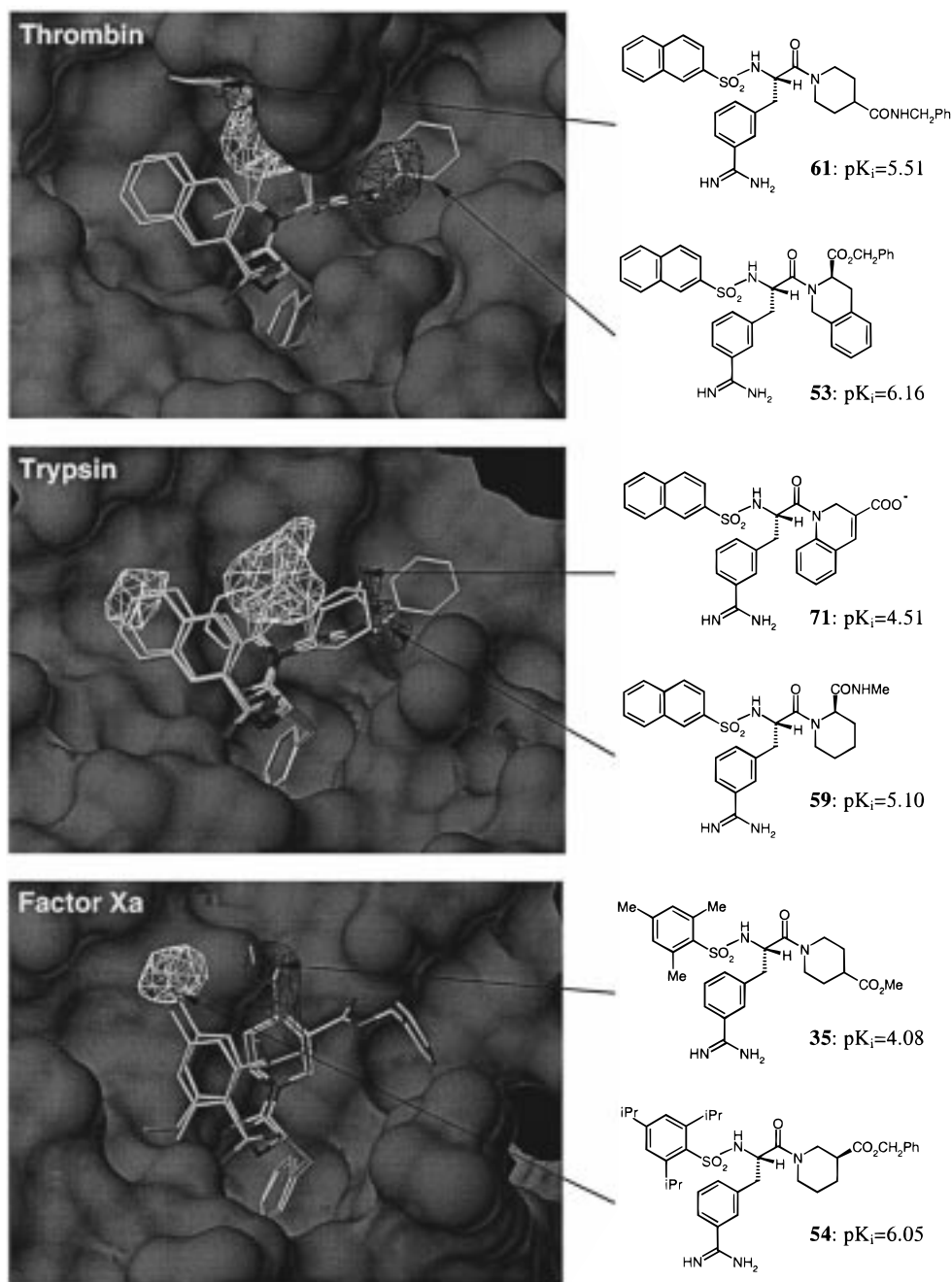
The  $q^2$  values obtained for factor Xa data reveal lower values for both CoMFA and CoMSIA. We explain this observation by the reduced spread of affinity data over only 3 orders of magnitude.

The predictive power of the three models has been checked by predicting the binding constants of 16 additional ligands not included in the training set (Figure 4). In almost all cases, the predicted values fall close to the observed  $pK_i$  values, deviating by not more than 1 (factor Xa: 0.5) logarithmic unit. CoMFA and CoMSIA possess comparable predictive power with respect to these 16 examples. Regarding the affinity data of thrombin, **84** appears to be an outlier. This derivative is substituted at position R<sup>2</sup> by a large 4-phenylpiperazine moiety. None of the molecules in the training set occupies the spatial area accommodated by this substituent in the test molecule in a comparable way. Accordingly the reduced predictability of this derivative can be explained.

In CoMSIA, a Gaussian-type distance dependence is applied. In a preliminary parameter study we calibrated the attenuation factor  $\alpha$  to 0.3.<sup>5</sup> Reducing  $\alpha$  to smaller values means that a probe placed at a particular lattice point detects molecular similarity in its neighborhood more globally. On the other hand, larger values of  $\alpha$  imply a more localized evaluation of similarity. Our systematic parameter study for  $\alpha$  confirms the previous selection of  $\alpha = 0.3$  as optimum (Figure 5).

It is often discussed whether the statistical significance of comparative molecular field analysis increases by considering additional property fields.<sup>34–37</sup> In the present case, we have considered five fields focusing on different physicochemical properties. It is highly unlikely that they are independent from each other; the degree of interdependence is difficult to estimate, however. Accordingly, we used the thrombin data set and computed models based on all 31 possible field combinations (Figure 6). In general,  $q^2$  values between 0.6 and 0.8 are found, except for some combinations based entirely on hydrogen-bonding and electrostatic properties. As mentioned above, this finding indicates strong interdependencies among the individual fields. Focusing solely on the predictive power of a QSAR model cannot justify the consideration of five different fields. However, regarding five (or more) fields opens the opportunity to partition the variance analysis with respect to particular physicochemical properties associated with the mol-





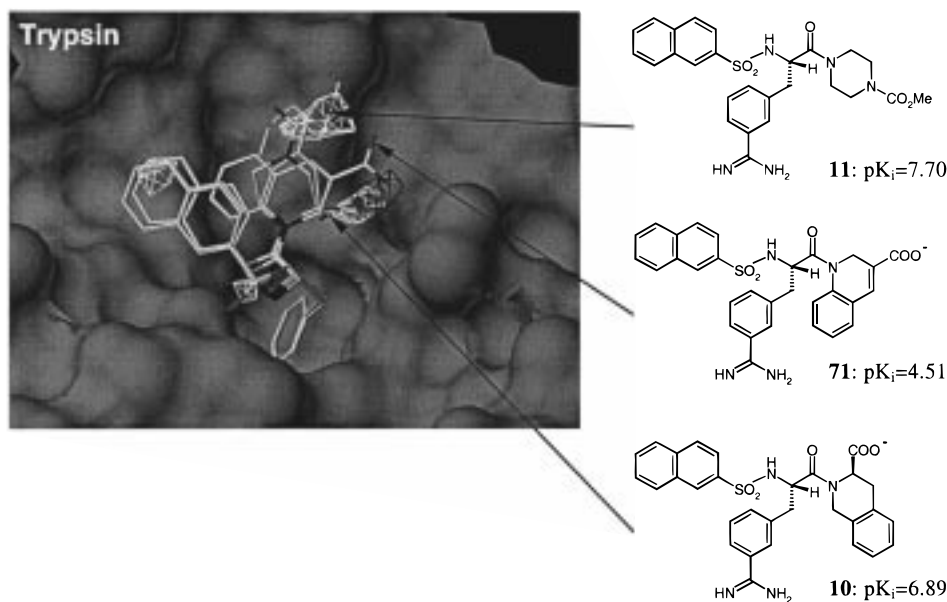
**Figure 7.** CoMSIA stdev\*coeff contour plot elucidating the sterical features with respect to thrombin, trypsin, and factor Xa. White isopleths (contour level  $-0.0028^{33}$ ) enclose areas where steric bulk will enhance affinity. Black contours (contour level  $+0.0022$ ) highlight areas which should be kept unoccupied; otherwise binding affinity will decrease. This is demonstrated by some inhibitors with weak affinities that occupy those black contoured regions (e.g., **61**, **53**, **71**, **59**, and **35**). On the other hand, inhibitor **54** occupying only the white isopleth with its isopropyl moiety in the factor Xa map reveals a higher affinity compared to **35**. The solvent-accessible surfaces are indicated as solid surfaces.

ecules. This aspect gains importance if results of comparative molecular field analysis are used for the design of improved ligands.

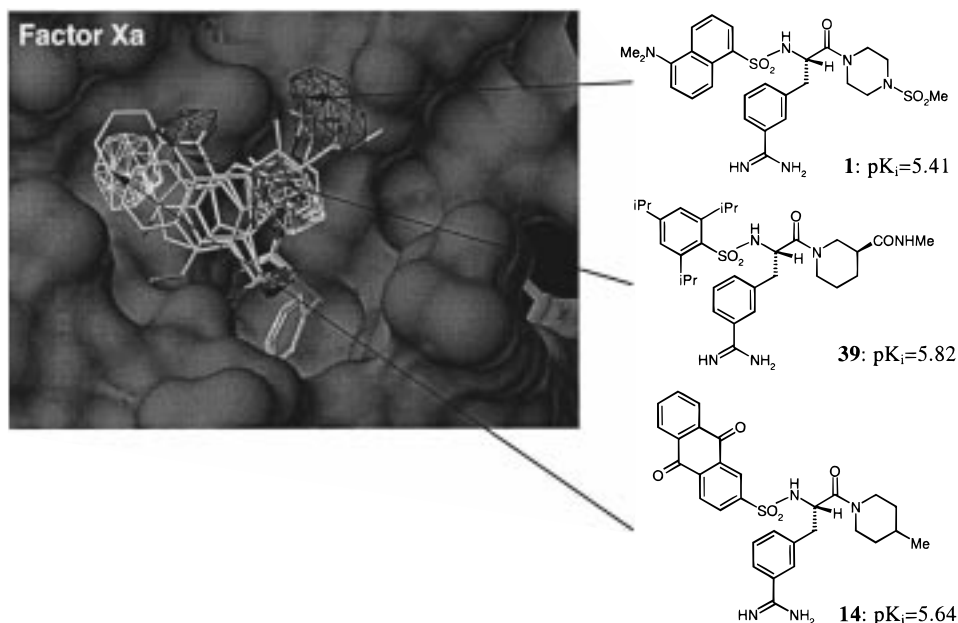
**Graphical Interpretation of the Results.** Besides greater robustness and better predictive power, the CoMSIA method provides significantly improved contour diagrams.<sup>3,5,10</sup> They allow the correlation results to be mapped back onto molecular structures. In the following figures, some selected contour diagrams of field contributions of different properties derived from CoMSIA are given together with some exemplary inhibitors. In addition, the solvent-accessible surface of the corresponding proteins are shown. In Figure 7, the

*steric properties* derived from the thrombin, trypsin, and factor Xa affinity data are displayed. Areas indicated by white contours correspond to regions where steric occupancy with bulky groups will increase affinity. Areas encompassed by black isopleths should be sterically avoided; otherwise reduced affinity can be expected. Different contour diagrams are revealed for the three enzymes. The black contour on the right (next to the catalytic center) is largest for thrombin, reduced in size for trypsin, and completely absent for factor Xa. In the center (proximal pocket, in thrombin below the 60 loop), trypsin shows a sterically favored site. For thrombin, this area splits into a larger favorable and





**Figure 8.** Contour plot of the CoMSIA stdev\*coeff for the electrostatic properties with respect to trypsin binding. White isopleths (contour level  $+0.004^{33}$ ) encompass regions where an increase of negative charge will enhance affinity, whereas in black contoured areas (contour level  $-0.007$ ) more positive charges are favorable for binding properties. The weak binding inhibitor **71** orients its negatively charged carboxylate group into an area which is favorable for more positive charges. The more active molecules **10** and **11** place their carboxylate or methyl ester moieties into regions indicated to be favorable to accommodate more negative charges. The solvent-accessible surface is indicated as a solid surface.

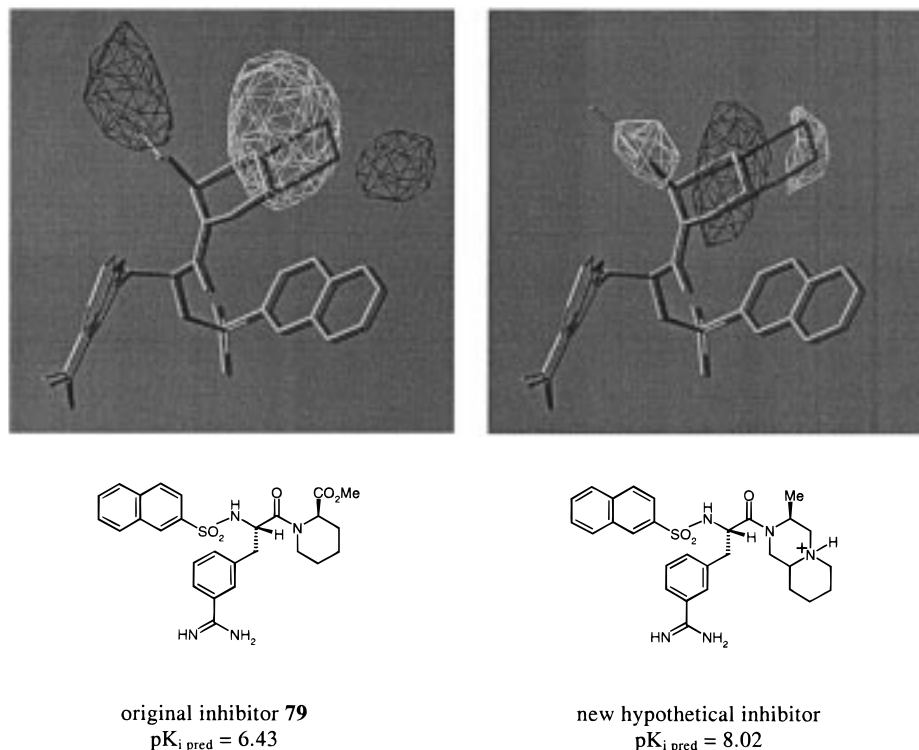


**Figure 9.** Contour plot of the CoMSIA stdev\*coeff for the hydrophobic properties with respect to factor Xa binding. White isopleths (contour level  $-0.005^{33}$ ) encompass regions favorable for hydrophobic groups, whereas in black contoured areas (contour level  $+0.002$ ) more hydrophilic groups are favorable for binding properties. Fairly potent inhibitors such as **14** and **39** orient their bulky aromatic substituents into the S3 pocket of factor Xa indicated to be favorable for binding. The active compound **1** places its piperazylsulfonyle moiety into the black contoured area favorable for hydrophilic groups.

an adjacent unfavorable area. In factor Xa the entire region is disfavored. Trypsin and factor Xa show sterically favorable regions in the distal pocket, however, at slightly different locations. The two molecules, displayed together with the latter map, occupy these regions differently. The less active **35** orients its methyl ester group into the disfavored region whereas the more active **54** fills the white contoured area by its *p*-isopropyl substituent.

The maps of *electrostatic properties* show fewer features in space; however, different areas are highlighted

for the three enzymes. In Figure 8, the contour diagram of electrostatic properties is given for trypsin. Inhibitors orienting groups with increasingly negative charge into areas contoured in white will enhance ligand binding, as will groups with a more positive partial charge placed into areas indicated in black. In the figure the inhibitors **10** and **71** possess distinct affinity. Due to slight stereochemical differences of their skeletons, the more active **10** orients its carboxylate group into an area indicated to be favorable for negatively charged groups (white area behind the black contour). The less active



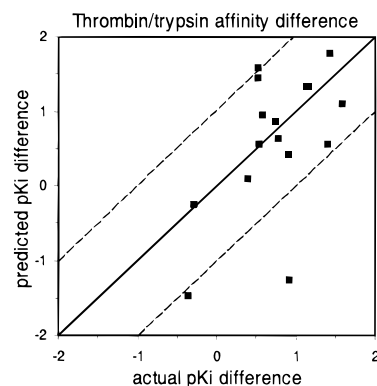
**Figure 10.** CoMSIA stdev\*coeff contour plots for the steric (left, white favorable and black unfavorable for steric bulk) and hydrophobic (right, white favorable for hydrophobic and black favorable for hydrophilic groups) properties with respect to thrombin binding. The maps were used for the design of a new hypothetical inhibitor, displayed in gray, predicted to possess improved affinity. The original molecule is shown in white. Its piperidine moiety was sterically enlarged to a decaline-type ring system in order to better occupy the white contoured area favorable for steric bulk (left) and to accommodate the distal contour favorable for hydrophobic groups (right). The methyl ester substituent of the original molecule occupying the black-colored area (left) indicated as unfavorable for steric bulk was reduced to a methyl group that at the same time correctly occupies the white isopleth (right) favorable for hydrophobic substituents. The required hydrophilicity in the central black contoured area (right) can be enhanced by introducing an additional nitrogen assumed to be protonated in the binding site.

**Table 7.** Statistical Results from the CoMSIA Analysis Using the Thrombin/Trypsin Affinity Differences

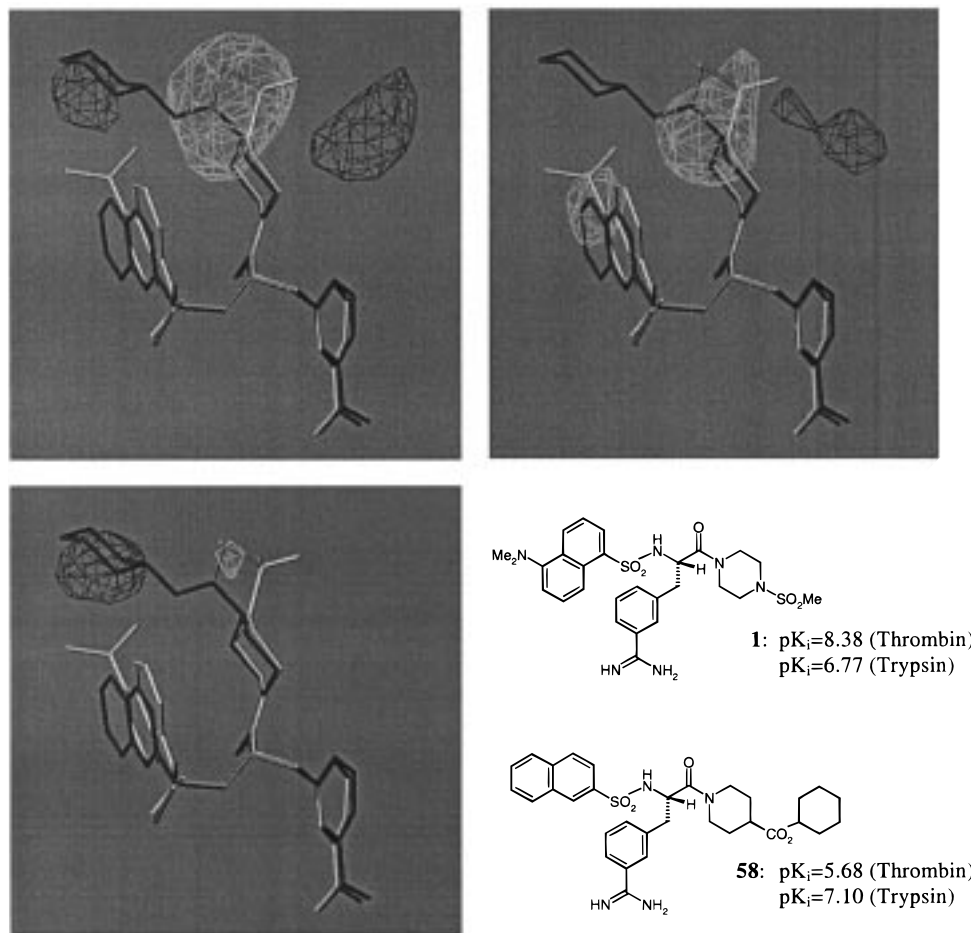
$q^2$	0.574
$S_{\text{press}}$	0.493
$r^2$	0.851
$S$	0.291
$F$	95.9
components	4
fraction	
steric	0.215
electrostatic	0.178
hydrophobic	0.284
donor	0.077
acceptor	0.246
box: stepsize	2 Å

**71** avoids this area. Instead, it partially occupies the black region, unfavorable for negative charge, with a carboxylate group. As an additional example molecule **11** is given, which completely avoids the area next to the catalytic center. It orients an ester group with electronegative oxygens into a site highlighted to be favorable for negatively charged residues. This compound shows the highest affinity among the three cases discussed.

The maps for *hydrophobic properties* show areas where increasing hydrophobicity enhances affinity for trypsin and factor Xa in the distal S3 pocket. Especially in the case of factor Xa (Figure 9) these areas are produced by some potent ligands possessing bulky aromatic moieties, e.g., **14** and **39**. This finding is in agreement with the general observation that factor Xa



**Figure 11.** Predicted versus actual thrombin/trypsin affinity differences for the 16 inhibitors of the test set. The predicted values were obtained by PLS analysis using the CoMSIA method with 2 Å grid spacing. The dashed lines mark deviations of 1 logarithmic unit from the ideal prediction. The high deviation of inhibitor **84** (bottom right) is discussed in the text. favors ligands with large hydrophobic substituents in the S3 pocket.<sup>38</sup> Compared to thrombin and trypsin, the S3 pocket in factor Xa is composed of three adjacent aromatic amino acid residues (Tyr99, Phe174, Trp215), whereas in the two other enzymes only one or two of these neighboring residues are of aromatic nature. For thrombin, hydrophobic groups are favored if they occupy an area below the 60 loop and an area between the 60 loop and the catalytic triad. For all three enzymes, an area coinciding with the spatial location of the saturated heterocyclic moiety (e.g., piperidine or piperazine),



**Figure 12.** Steric contour maps of the CoMSIA stdev\*coeff for the thrombin (left) and trypsin (right) data. White isopleths encompass areas where steric bulk will enhance affinity. Black contours highlight areas where steric bulk will be affinity-reducing. The third map (second row) shows the steric affinity difference map ("selectivity" map) of thrombin/trypsin. The black isopleth indicates the features mainly responsible for selectivity discrimination: if this area is accessed by sterically demanding groups, affinity toward thrombin will drop compared to that for trypsin. The inhibitor **1** (white) with high affinity toward thrombin leaves the selectivity-discriminating region unoccupied, whereas for thrombin less active **58** (black) orients its cyclohexyl substituent into the contoured area.

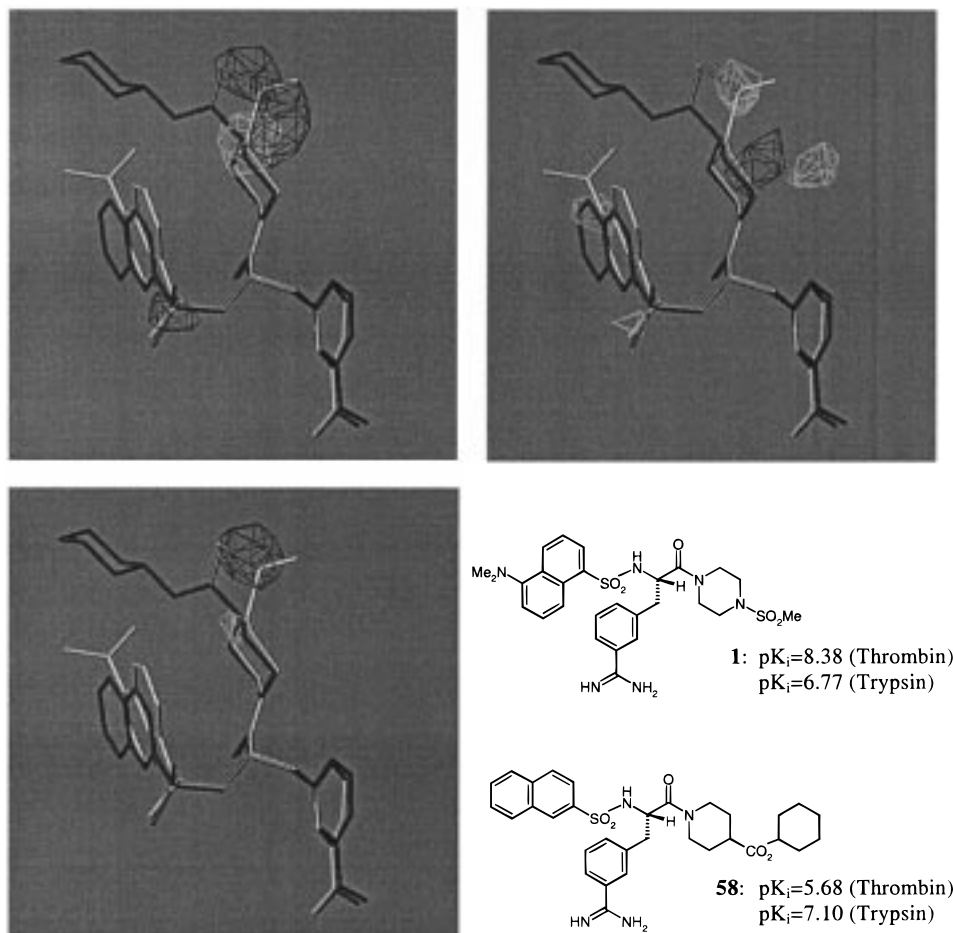
present in most of the inhibitors, is indicated to be favorable for hydrophilic groups. This is in agreement with the fact that at this location piperazyl and piperazylsulfonamide derivatives possess enhanced binding affinity toward all three enzymes compared to the piperidine derivatives.

In summary, the contour diagrams for steric, electrostatic, and hydrophobic properties highlight different areas for the three enzymes. This observation clearly indicates selectivity differences among the inhibitors for the three structurally deviating serine proteinases (see below). In this respect, the contribution maps of the *hydrogen-bonding properties* are less conclusive. One large area is contoured next to the catalytic triad for donor facilities to be favorable; however, it is identically located in all three enzymes. The map for acceptor properties shows no features at all for factor Xa. In the case of thrombin and trypsin an area next to Ser195 is indicated. This region favors the presence of acceptor properties to enhance affinity.

To demonstrate the ability of the CoMSIA contribution maps with regard to the design of improved affinity molecules, the following hypothetical design is suggested (Figure 10). We start with the medium active inhibitor **79**, predicted to possess a  $pK_i$  of 6.43. Con-

sidering the steric contribution map (left), its methyl carboxylate is placed in a sterically unfavorable region (black isopleth). Furthermore, beyond the piperidine moiety a region is highlighted by a white isopleth where steric occupancy should be affinity-enhancing. Consequently, removal of the methyl ester group and replacement of the piperidine ring by a sterically extended decaline-type moiety should enhance affinity. Consulting the maps of hydrophobic properties (right), this proposal can be further refined: a small hydrophobic group next to the piperidine nitrogen should be affinity-enhancing (left white isopleth). A methyl group is chosen to serve this purpose. Furthermore, as indicated by many of the more potent piperazine derivatives compared to the piperidine moieties, an additional hydrophilic center is advantageous for binding (black isopleth). In the hypothetical molecule, this requirement can be realized by an additional nitrogen, presumably protonated under physiological conditions. The suggested compound is predicted by CoMSIA to reveal a  $pK_i$  of 8.02, about 1.6 logarithmic units more than the original structure.

**Elucidating Selectivity Features.** As mentioned above, the contribution maps of steric, electrostatic, and hydrophobic properties show deviating features that



**Figure 13.** Electrostatic contour maps of the CoMSIA stdev\*coeff for the thrombin (left) and trypsin (right) data. White isopleths encompass regions where more negative charge will enhance affinity, whereas in black contoured areas more positive charges are favorable for binding. The third map (second row) shows the electrostatic affinity difference map of thrombin/trypsin. The remaining black isocontour shows that the presence of a more positively charged group in this area will enhance selectivity toward stronger binding to thrombin. The inhibitor **1** places its positively charged sulfur atom at the methylsulfonyl group into this area thus revealing better thrombin binding; see Figure 13.

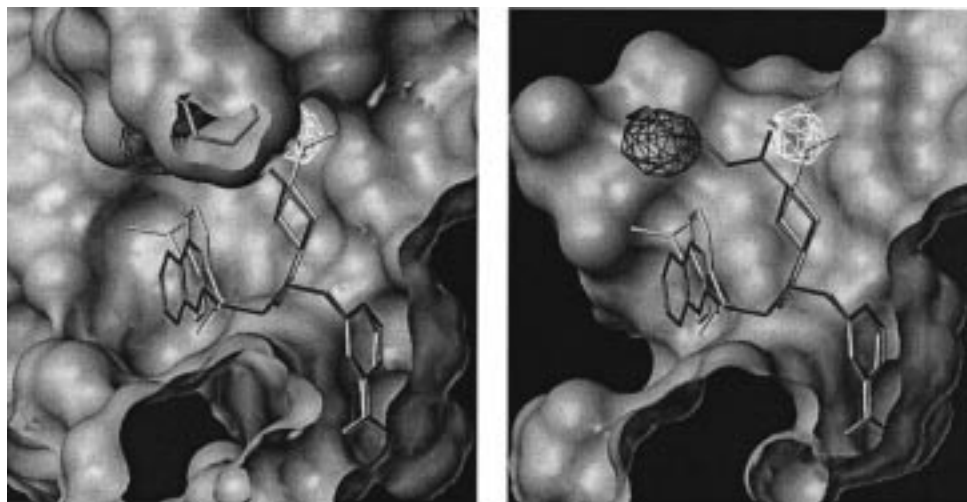
point to structural characteristics responsible for selectivity. The maps for hydrogen-bonding properties do not suggest any discriminating power related to these properties. In any case, it must be remembered that all features derived from a comparative molecular field analysis are only a mirror of the structural variations inherently present in the selected data set. Accordingly, selecting another structurally deviating data set may result in different features leading to alternative conclusions.

To better elucidate the selectivity-discriminating criteria responsible in the data set under consideration, we performed the following data analysis, applied only to the statistically more robust thrombin and trypsin data. In a first step, we determined the affinity differences between thrombin and trypsin for all 72 inhibitors. Even though the differences are associated with a higher experimental uncertainty due to error propagation, we used these values as a dependent property in a CoMSIA analysis. Over all, the difference data spread over more than 3 orders of magnitude. The statistical results of this evaluation are given in Table 7. According to these results, the model obtained appears to have some predictive power. We tried to predict the thrombin/trypsin affinity differences for the 16 compounds not included in the training set. Apart from the above-

discussed outlier **84**, quite convincing results are obtained (Figure 11).

Finally we consulted the contribution maps derived from the affinity differences. In Figure 12, the steric maps obtained for the original thrombin and trypsin data are shown together with the steric map received from the affinity differences. At first glance, the map based on the differences appears as if obtained as a difference between the two maps based on the original affinity data. However, this "selectivity map" is computed subsequently to a PLS analysis based on the thrombin/trypsin  $pK_i$  differences. The steric selectivity map shows one area encompassed by a black isocontour. Leaving this area sterically unoccupied discriminates selectivity toward enhanced thrombin binding; i.e., the binding affinity toward thrombin will increase relative to trypsin. Two inhibitors are shown together with this map. The white-colored inhibitor **1** possesses higher affinity toward thrombin and leaves the indicated area unoccupied. The inhibitor **58** (black-colored) with higher affinity toward trypsin places its terminal cyclohexyl moiety into this affinity-discriminating area. In Figure 13, the corresponding maps displaying the electrostatic properties are shown. Again the selectivity-indicating map presents one strong feature, a region contoured in black requiring a positively charged group to enhance





**Figure 14.** Two selectivity-discriminating areas indicated in the steric (Figure 12) and electrostatic (Figure 13) property maps fall close to the 60 loop in thrombin. In this diagram the contours responsible for steric (left contour in black) and electrostatic (right contour, compared to Figure 11 here in white) discrimination are displayed together with the solvent-accessible surface. Whereas thrombin (left) shows a narrow space-restricted pocket due to the flanking 60 loop, the binding site of trypsin (right) is open and sterically unconstrained in this region. The two isopleths indicating the affinity-discriminating features fall next to the 60 loop in thrombin particularly composed by Tyr60A and Trp60D.

affinity toward thrombin. Considering again the two examples, the more potent thrombin inhibitor **1** (white) orients its methyl sulfonyl group (with sulfur bearing a positive partial charge) into this area, whereas the less potent thrombin ligand **58** (black) leaves this region unoccupied. Comparing the local shape differences of the thrombin versus trypsin binding site, it is interesting to note that both contours highlighted in the steric and electrostatic selectivity-indicating maps fall next to the 60 loop (Figure 14). This loop occurs as a special characteristic in thrombin; accordingly it is reasonable that areas where affinity between both enzymes is discriminated fall close to this 60 loop. Obviously, contour diagrams derived from a CoMSIA analysis based on binding affinity differences highlight plausible spatial characteristics associated with structural differences responsible for selectivity discrimination.

## Conclusions

The present study has shown CoMSIA to possess better predictive power and greater robustness compared to CoMFA. Furthermore, it has been demonstrated that 3D QSAR methods can be successfully applied to derive distinct correlation models for a set of inhibitors binding with deviating affinities to three related enzymes. The contour diagrams obtained for the various CoMSIA field contributions can be mapped back onto structural features accounting for the affinity trends and selectivity discrimination among the inhibitors. Even more conclusive with respect to the latter aspect is a correlation analysis based on affinity differences. It reveals a predictive model and highlights those regions in the ligands that are responsible for the selectivity differences. On the basis of the spatial arrangement of the various field contributions, novel molecules can be designed that are predicted to possess improved binding affinity. 3D QSAR models are a direct mirror of the structural variations inherently present in the selected data set. Accordingly, the selection of a structurally more diverse set of ligands should allow to map features distinct from those highlighted by the

present study. Such a compilation should also allow to analyze affinity variations of the ligands resulting from differences in their hydrogen-bonding properties.

**Acknowledgment.** The authors are grateful to BASF, especially Hugo Kubinyi, for making a copy of the CoMSIA method available to us. Special thanks go to Thomas Mietzner (BASF) for adapting some of the Fortran routines to our data. The effective introduction into the usage of this method to one of us (M. Böhm) by Ute Abraham (BASF) is also gratefully acknowledged. We also thank Milton T. Stubbs (Marburg) for helpful discussions. We are grateful to Pentapharm Ltd. (Basel, Switzerland) for the generous support of the work, especially use of the structures of the whole series of inhibitors synthesized by P. Wikström and H. Vieweg.

## References

- (1) Davie, E. W.; Fujikawa, K.; Kisiel, W. The coagulation cascade: initiation, maintenance and regulation. *Biochemistry* **1991**, *30*, 10363–10370.
- (2) Bode, W.; Brandstetter, H.; Mather, T.; Stubbs, M. T. Comparative analysis of haemostatic proteinases: structural aspects of thrombin, factor Xa, factor IXa and protein C. *Thromb. Haemostasis* **1997**, *78*, 501–511.
- (3) Klebe, G. Comparative molecular similarity indices analysis: CoMSIA. *Perspect Drug Discovery Des.* **1998**, *12*, 87–104.
- (4) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
- (5) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130–4146.
- (6) Stürzebecher, J.; Vieweg, H.; Wikström, P. WO 92/08709, 1991.
- (7) Stürzebecher, J.; Vieweg, H.; Wikström, P. WO 94/18185, 1993.
- (8) Stürzebecher, J.; Prasa, D.; Wikström, P.; Vieweg, H. Structure–activity relationships of inhibitors derived from 3-amidinophenylalanine. *J. Enzyme Inhib.* **1995**, *9*, 87–99.
- (9) Stürzebecher, J.; Prasa, D.; Hauptmann, J.; Vieweg, H.; Wikström, P. Synthesis and structure–activity relationships of potent thrombin inhibitors: piperazines of 3-amidinophenylalanine. *J. Med. Chem.* **1997**, *40*, 3091–3099.
- (10) Klebe, G.; Abraham, U. Comparative molecular similarity indices analysis (CoMSIA) to study hydrogen bonding properties and to score combinatorial libraries. *J. Comput.-Aided Mol. Des.* **1998**, in press.

- (11) Folkers, G.; Merz, A.; Rognan, D. CoMFA: scope and limitations. In *3D QSAR in Drug Design, Theory Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; pp 583–618.
- (12) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J., III. The collinearity problem in linear regression. The partial least squares approach to generalized inverses. *SIAM J. Sci. Stat. Comput.* **1984**, *5*, 735–743.
- (13) Stahle, L.; Wold, S. Multivariate data analysis and experimental design in biomedical research. *Prog. Med. Chem.* **1988**, *25*, 291–338.
- (14) SYBYL molecular modeling software; Tripos Inc., 1699 South Hanley Rd, Suite 303, St. Louis, MO 63144.
- (15) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. E., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535–542.
- (16) Brandstetter, H.; Turk, D.; Höffken, H. W.; Grosse, D.; Stürzebecher, J.; Martin, P. D.; Edwards, B. F.; Bode, W. Refined 2.3 Å X-ray crystal structure of bovine thrombin complexes formed with the benzamidino and arginine-based thrombin inhibitors NAPAP, 4-TAPAP and MQPA. A starting point for improving antithrombotics. *J. Mol. Biol.* **1992**, *226*, 1085–1099.
- (17) Turk, D.; Stürzebecher, J.; Bode, W. Geometry of binding of the N<sup>ε</sup>-tosylated piperidides of m-amidino-, p-amidino- and p-guanidino phenylalanine to thrombin and trypsin. X-ray crystal structures of their trypsin complexes and modeling of their thrombin complexes. *FEBS Lett.* **1991**, *287*, 133–138.
- (18) Padmanabhan, K.; Padmanabhan, K. P.; Tulinsky, A.; Park, C. H.; Bode, W.; Huber, R.; Blankenship, D. T.; Cardin, A. D.; Kisiel, W. Structure of human des(1–45) factor Xa at 2.2 Å resolution. *J. Mol. Biol.* **1993**, *232*, 947–966.
- (19) Vinter, J. G.; Davis, A.; Saunder, M. R. Strategic approaches to drug design. I. An integrated software framework for molecular modelling. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 31–35.
- (20) Clark, M.; Cramer, R. D., III; Van Opdenbosh, N. Validation of the general purpose Tripos 5.2 force field. *J. Comput. Chem.* **1989**, *10*, 982–1012.
- (21) Mayer, D.; Naylor, C. B.; Motoc, I.; Marshall, G. R. A unique geometry of the active site of angiotensin-converting enzyme consistent with structure–activity studies. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 3–16.
- (22) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- (23) Stewart, J. J. P. MOPAC: a general molecular orbital package; QCPE #445; J. Frank Seiler Research Laboratory, United States Air Force Academy, CO 80840.
- (24) Stewart, J. J. P. MOPAC: a semiempirical molecular orbital program. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1–105.
- (25) Kearsley, S. K.; Smith, G. M. An alternative method for the alignment of molecular structures: maximizing electrostatic and steric overlap. *Tetrahedron Comput. Methodol.* **1990**, *3*, 615–633.
- (26) The CoMSIA method described will be made generally available with the SYBYL 6.5.1 release.
- (27) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. Atomic physicochemical parameters for three-dimensional structure directed quantitative structure–activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163–172.
- (28) Klebe, G. The use of composite crystal-field environments in molecular recognition and the de novo design of protein ligands. *J. Mol. Biol.* **1994**, *237*, 212–235.
- (29) Klebe, G.; Mietzner, T.; Weber, F. Methodological developments and strategies for a fast flexible superposition of drug-size molecules. *J. Comput.-Aided Mol. Des.* **1999**, in press.
- (30) Bush, B. L.; Nachbar, R. B., Jr. Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 587–619.
- (31) Cho, S. J.; Tropsha, A. Cross-validated R<sup>2</sup>-guided region selection for comparative molecular field analysis: a simple method to achieve consistent results. *J. Med. Chem.* **1995**, *38*, 1060–1066.
- (32) Kim, K. H.; Brusniak, M. K.; Pearlman, R. S. UniSurCoMFA: for stable and consistent 3D QSAR. In *Rational Molecular Design in Drug Research*; Liljefors, T., Jørgensen, F. S., Krosgaard-Larsen, P., Eds.; Munksgaard: Copenhagen, 1998; pp 67–83.
- (33) The contour levels of CoMFA contour maps are given normally in kcal/mol. As the CoMSIA similarity indices fields are defined in arbitrary units, they do not correspond to any potential describing different partitions of molecular interactions such as CoMFA. Consequently, the CoMSIA contour levels are given without any units.
- (34) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
- (35) Kellogg, G. E.; Semus, S. F.; Abraham, D. J. HINT: a new method of empirical hydrophobic field calculation for CoMFA. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 545–552.
- (36) Gaillard, P.; Carrupt, P. A.; Testa, B.; Boudon, A. Molecular lipophilicity potential, a tool in 3D QSAR: method and applications. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 83–96.
- (37) Kim, K. H.; Greco, G.; Novellino, E. A critical review of recent CoMFA applications. *Perspect Drug Discovery Des.* **1998**, *12*, 257–315.
- (38) Stubbs, M. T. Structural aspects of factor Xa inhibition. *Curr. Pharm. Des.* **1996**, *2*, 543–552.

JM981062R