GRID/CPCA: A New Computational Tool To Design Selective Ligands

Mika A. Kastenholz,[†] Manuel Pastor,[‡] Gabriele Cruciani,[‡] Eric E. J. Haaksma,[†] and Thomas Fox*[,†]

Department of Chemical Research/Structural Research, Boehringer Ingelheim Pharma KG, 88397 Biberach/Riss, Germany, and Laboratory for Chemometrics, University of Perugia, Via Elce di Sotto 10, 06123 Perugia, Italy

Received March 6, 2000

We present a computational procedure aimed at understanding enzyme selectivity and guiding the design of drugs with respect to selectivity. It starts from a set of 3D structures of the target proteins characterized by the program GRID. In the multivariate description proposed, the variables are organized and scaled in a different way than previously published methodologies. Then, consensus principal component analysis (CPCA) is used to analyze the GRID descriptors, allowing the straightforward identification of possible modifications in the ligand to improve its selectivity toward a chosen target. As an important new feature the computational method is able to work with more than two target proteins and with several 3D structures for each protein. Additionally, the use of a 'cutout tool' allows to focus on the important regions around the active site. The method is validated for a total number of nine structures of the three homologous serine proteases thrombin, trypsin, and factor Xa. The regions identified by the method as being important for selectivity are in excellent agreement with available experimental data and inhibitor structure−activity relationships.

Introduction

Selectivity toward a single biological target is an essential requirement for potential drugs to minimize side effects. Therefore, it is desirable to involve selectivity considerations as early as possible in the drug design efforts. In the past, the design of selective drugs relied on trial and error or on the close inspection of structural data. This task was especially difficult when targeting a single protein within a highly homologous family.

With the same purpose, QSAR or CoMFA approaches were often used in the past. Here, one tries to correlate the biological test results for the studied systems with structural features present in the tested ligands and to identify functional groups responsible for activity and selectivity.[1,2] However, this requires a series of compounds that have already been synthesized and tested and relies critically on the alignment and superposition of the compounds. Moreover, this approach is difficult if the structural variation within the test compounds is too high. QSAR approaches also suffer from the principal problem that they only can interpolate within the given set of protein ligands: properties that are not present in the training set (e.g. selectivity in a set of unselective inhibitors) cannot be found. This is solved by the computational procedure presented in this article.

The availability of 3D structures of the target proteins considerably simplifies the search for selective ligands. Then it is possible to compare the binding sites of different targets, looking for differences in sequence or structure that can be exploited for selective target−ligand interactions. However, this is by no means trivial, as competing contributions may have to be balanced, and there is always the danger of overlooking structural differences that might have a strong impact on selectivity. To avoid these problems, the GRID/PCA approach was developed and successfully applied in the past to investigate the selectivity between pairs of biomolecular targets.[3−6] The original approach was based on the analysis of GRID[7−9] generated molecular interaction fields (MIFs) using principal component analysis (PCA[10]) to highlight the most relevant differences between the given target enzymes. Visual interpretation of the results guided the understanding of the contributions of various interactions in the binding process. However, the way the problem was formulated made it very difficult if not impossible to understand the relative importance of the GRID probes for selectivity. As a consequence, regions that infer selectivity through hydrophobic interactions could never be identified as important with this method. Moreover, the original GRID/PCA method allows only the investigation of pairs of targets. If more than two homologous enzymes are being studied and/or a certain selectivity profile of the ligand is desired, the interpretation becomes extremely complex.

In this paper, we present an improved method based on the GRID/PCA approach that offers a solution to these problems. Starting from one or more 3D structures of several target proteins, a multivariate description of the binding sites is performed using the program GRID. Then the MIFs are analyzed with the consensus principal component analysis (CPCA). As a result we obtain contour plots highlighting both the regions and the type of interactions in these regions that can be used to introduce selectivity into a potential ligand of these protein targets.

Test System

This novel chemometric approach will be illustrated for three serine proteases of the chymotrypsin family: thrombin, factor Xa, and trypsin. Thrombin and factor

* To whom correspondence should be addressed. Tel: ++49+7351-54 7585. Fax: ++49+7351−54 5137. E-mail: thomas.fox@bc.boehringer-ingelheim.com.
[†] Boehringer Ingelheim Pharma KG.
[‡] University of Perugia.

10.1021/jm000934y CCC: $19.00
Published on Web 07/25/2000

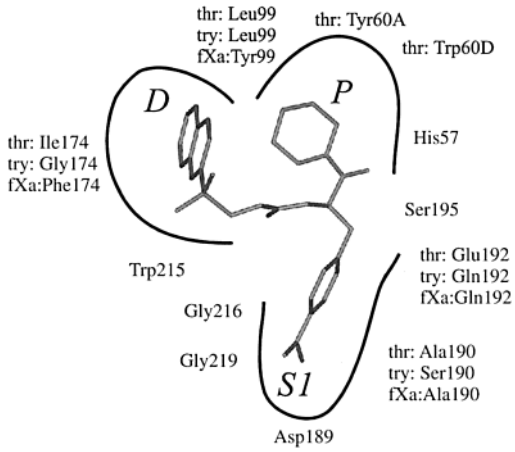


Figure 1. Schematic drawing of the unprimed side of the active site of the three enzymes thrombin, trypsin, and factor Xa. The crystallographic binding mode of the inhibitor NAPAP is shown to indicate the S1, P, and D pockets. Amino acids important for inhibitor binding are highlighted. Residues which differ among the enzymes are preceded by a 3-letter code to indicate the enzyme: thr = thrombin, try = trypsin, fXa = factor Xa.

Xa are prominent players in the blood clotting cascade.[11] They are, therefore, important targets for the development of new anticoagulant/antithrombotic drugs.[12−17] Trypsin is an enzyme excreted by the pancreas for helping in the digestion of aliments and has classically been used as a model enzyme for the whole serine protease family. Thus, to minimize side effects of thrombin/factor Xa inhibitors, and to enhance their bioavailability, potential drugs should exhibit selectivity toward thrombin/factor Xa with respect to trypsin.[18]

Crystallographic structures for all three enzymes show a similar structure of their active sites.[19] Figure 1 shows a schematic drawing of the active sites with the amino acids highlighted which differ between the three enzymes. The main determinant for the specificity of ligands toward proteins of the chymotrypsin family is the deep hydrophobic S1 pocket. Asp189 (the numbering of the amino acid residues follows the chymotrypsin scheme[19]) is located on the bottom of the S1 pocket, where it can form salt bridges with basic residues of the substrate peptides. This is the most conserved region of the three studied serine proteases: all residues are conserved except for a A190S mutation in trypsin which makes its pocked slightly more hydrophilic. Two other pockets on the unprimed side of the catalytic triad are important for substrate and inhibitor binding. Both the P and the D pockets are mostly hydrophobic, and the differences in their amino acid residues have been used in the past to enhance selectivity in potential protein ligands.

The most striking difference between thrombin and the other two proteases is its insertion loop Tyr60A-Pro60B-Pro60C-Trp60D that rests as a lid on the active site and forms the P pocket. The hydrophobic P pocket thus imposes steric constraints on potential inhibitors and also offers the selective exploitation of hydrophobic interactions.[13] Therefore, designing ligands which point small hydrophobic functional groups into the P pocket results in enhanced selectivity for thrombin.[20,21] Another significant structural difference is the D pocket of factor Xa which is formed by residues Phe174, Tyr99, and Trp215. They produce an 'aromatic box' able to accommodate hydrophobic and positively charged functional groups,[22,23] and this has been exploited in the design of selective factor Xa inhibitors.[24−26] Another difference that was used in the past is the Q192E mutation in thrombin at the entrance of the S1 pocket. Compounds pointing a carboxylate group in this direction interact more favorably with the glutamine environment of trypsin or factor Xa than with the glutamate in thrombin.[25,27−29]

In the past, only a few QSAR studies have investigated the selectivity in the thrombin/trypsin/factor Xa system.[21,30] They support the experimentally derived structure−activity relationships and highlight the importance of the interactions of ligands in the three subpockets (S1, P, D) for the selectivity.

Chemometric Analysis

GRID/PCA[4,5] was one of the first methods proposed to rationally design selective ligands. In this method, only two targets were studied, and both were characterized using GRID with n different probes representing different chemical groups. The problem was formulated, from a chemometric point of view, as a collection of $2n$ objects, each one representing a different target−probe interaction. To extract relevant information from this **X** matrix (which contains all target−probe interaction energies collected at the grid points), the PCA decomposes the matrix into a product of two smaller matrices, **T** (score matrix) and **P′** (loading matrix), that explain at best the overall variance of the original **X** matrix. The score matrix contains a few variables (so-called principal components, PCs) that are used to describe the objects, while the loading matrix relates the original variables with the PCs. From a practical point of view, the PCA of the **X** matrix allows on the one hand a simplified view of the system: in the scores plot two clusters, corresponding to the two target proteins, can be recognized. On the other hand, the loading plot can be used to identify the variables with highest participation in the PCA that discriminate these two clusters. When these variables are represented in the space around the targets using appropriate isocontour plots, they identify selectivity-relevant regions.

However, this procedure has several problems that make a selectivity analysis difficult. First, since the PCA is quite sensitive to the scale of interaction energies, the information given by probes representing weak nonbonded interactions (van der Waals and hydrophobic) is masked by the effect of probes representing stronger interactions (Coulombic and hydrogen bonds). This can lead to the erroneous appreciation that hydrophobic or van der Waals interactions cannot be exploited for selectivity, while the importance of hydrophobic interactions for drug binding has become increasingly evident over the years. Therefore, the measure of the relevance assigned by the method to the different probes can be unreliable and misleading, if only based on the numerical scale.

Another important problem is that the PCA model is general, and thus the spatial regions highlighted represent generally important regions, not linked to any

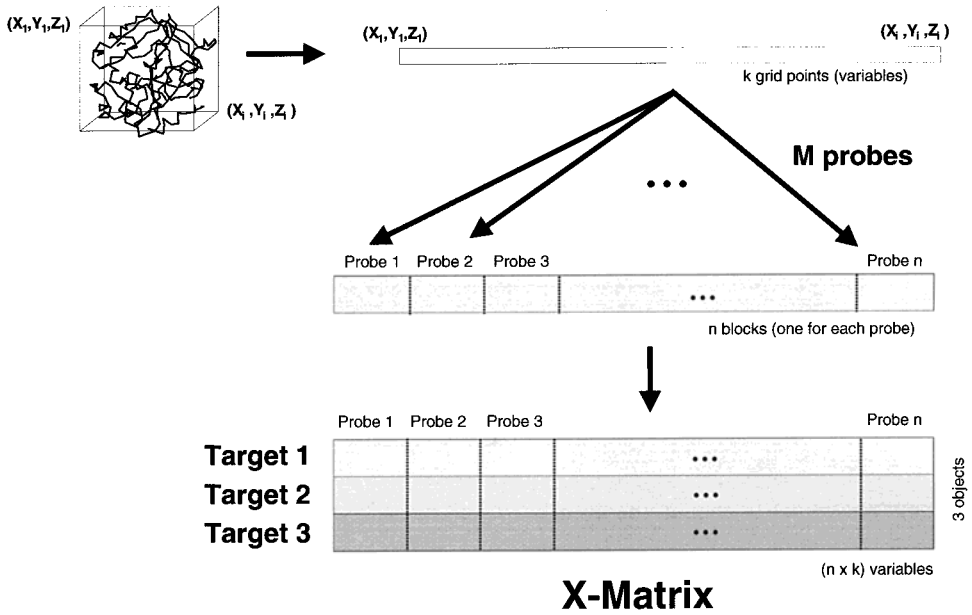


Figure 2. Procedure used to build the **X** matrix for the CPCA. Starting from the GRID calculations for one probe, a vector containing all interaction energies at the *k* grid points is constructed. Then the vectors for *n* probes are compiled into one long vector containing $n \times k$ data points. Stacking the long vectors for every target protein results in the final **X** matrix.

particular probe. From the point of view of designing selective ligands, it would be much more useful to highlight such selectivity-relevant regions for different probes representing potential chemical groups.

Moreover, the original method relies on the visual inspection of the scores plot to identify the PC which participates most in the discrimination of the clusters. This can be difficult when there are more than two clusters, since two or more PCs might be relevant for the discrimination. To overcome the mentioned difficulties, the new GRID/CPCA approach has been developed. It is based on the GRID/PCA method but incorporates several differences and enhancements, which are detailed in the following paragraphs.

New Problem Formulation. In the GRID/CPCA approach, the 3D structures representing the targets are analyzed using GRID, but the MIFs obtained for the different probes are added, side-by-side, adding new variables to the same object (see Figure 2). At the end, the matrix describing the systems has a row for each 3D structure studied and $n \times k$ columns, corresponding to *n* probes $\times$ *k* variables present in a single MIF.

The PCA of such a matrix produces a scores plot where each 3D structure is represented by a single point (see Figure 3). When several different structures are used to represent each target protein, the scores plots should show them clustered, thus showing that the differences between the 3D structures of the same target are less important than the differences between the proteins.

This problem formulation has a number of additional advantages. First, since the method is not limited to the description of two objects, several protein targets can be included in the analysis and selectivity profiles between groups of targets can be generated. In addition, many 3D structures representing the same target protein can be used. This allows minimizing spurious results resulting from minute differences in the crystal

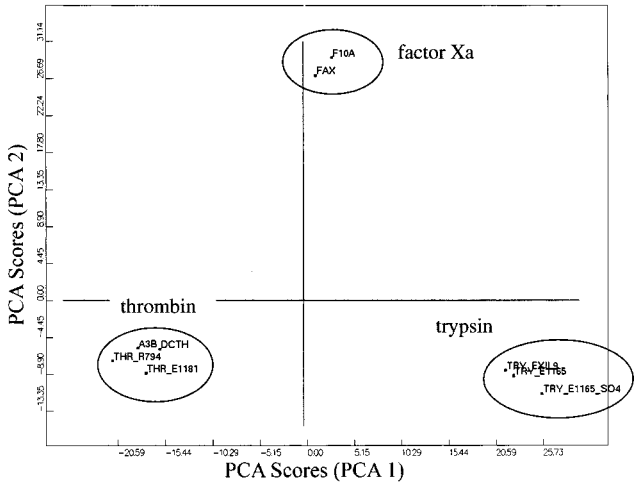


Figure 3. CPCA scores plot for the carboxyl oxygen probe. The clustering of the various structures for each protein used in the analysis is evident. PC 1 can be used to discriminate between thrombin and trypsin, whereas PC 2 discriminates between factor Xa on the one hand and trypsin and thrombin on the other.

structures of the targets. Indeed, the method highlights the differences between the common features of the targets.

Block Unscaled Weights. One of the main drawbacks of the original method originates from the different overall range of the interaction energies obtained for different probes. In the new problem formulation, the different probes are organized in different blocks of variables. Therefore, it is possible to apply a scaling procedure which normalizes their importance in the model. The weighting procedure is called block unscaled weights (BUW) and serves to fulfill the outlined objective. The operation scales each variable block separately, whereas the relative scales of single variables within each block remain unchanged. The weighting coefficients are obtained through an equalization of the

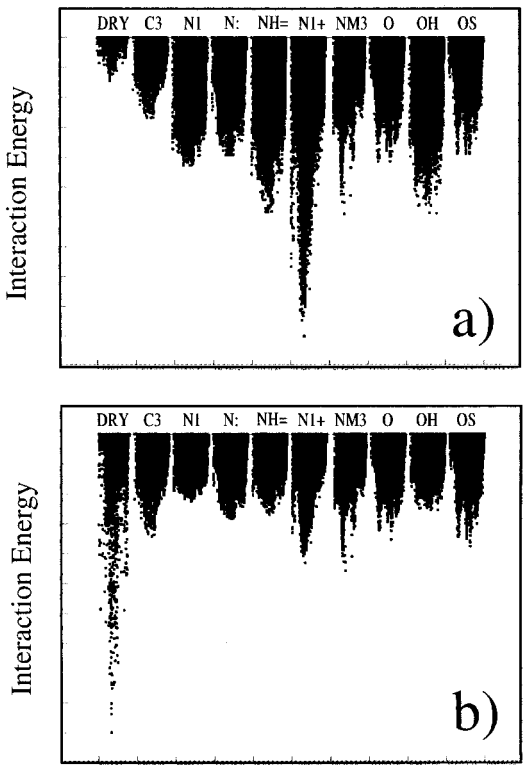


Figure 4. Effect of BUW on the distribution of the *x* variables (i.e. the interaction energies obtained with GRID) for each probe: (a) distribution before and (b) distribution after BUW to unit variance.

block variances, which results in giving each probe the same importance within the model. Figure 4 illustrates the BUW procedure, showing the initial energy distribution of the *x* variables for each probe and the normalized distribution after the variable block weighting.

CPCA. The matrix produced by this new problem formulation can be analyzed using regular PCA. However, since it is structured in meaningful blocks, hierarchical PCA methods such as hierarchical principal component analysis[31,32] and consensus principal component analysis[33] provide interesting information regarding the relative importance of the different blocks (i.e. probes) in the analysis. Among the available methods, we decided to apply CPCA, as it is implemented in the GOLPE[34] program. The algorithm used is an adaptation of NIPALS[10] and is described in the Appendix.

Basically, CPCA uses exactly the same objective function as PCA and tries to best explain the overall variance of the **X** matrix, but the analysis is made on two levels: the block level, which expresses the "opinion" of each of the probes, and the superlevel, which expresses the "consensus" of all blocks. As implemented in GOLPE, CPCA provides a solution on the superlevel that is identical to a solution found in regular PCA, i.e., the same **T** and **P** matrices are obtained. Additionally, the method produces block scores $\mathbf{T_b}$ and block loadings $\mathbf{P_b}$ for each of the probes used and a weight matrix which expresses the participation of each block in the overall scores. The block loadings are essentially identical to the "piece" of the loading corresponding to this block but use a different normalization. On the contrary, block scores represent a particular point of view of the model given by a certain probe and provide unique information not present in regular PCA. Object distances in the block scores are used in GOLPE to assess the relative importance of the different probes in their discrimination.

Region Cutouts. Due to the use of a large rectangular box to enclose all relevant regions within the binding site, data will also be collected for regions that do not lie within the immediate active site binding pockets. Since the homology tends to be higher in the binding site than further away, these regions are often only badly superimposed and the PCA finds significant structural differences that mask more relevant but more subtle differences in the region of primary interest. To overcome this problem, GOLPE includes the 'cutout tool', which allows to select irregularly shaped regions within the original GRID box. This not only leads to a significant reduction of the *x* variables but also facilitates the overall analysis. The resulting contours gain much transparency due to focusing on relevant regions in the binding sites.

Identification of Important Variable Blocks for Selectivity. Plots showing the distribution of the *x* values combined with a distinctive use of the cutout tool allow the identification of probes that are able to distinguish between the different objects within the selected region. Therefore, maxima within the interaction energy distribution of a variable block allow a direct translation into selectivity for the corresponding protein.

Contour Plots. The PCA or CPCA loadings can be translated into contour plots describing the interaction fields between a GRID probe and the target protein structure. For a selectivity study one is interested in the loadings discriminating different target proteins. Unfortunately, often more than one PC contributes to separate these objects in the scores plot, and therefore, any single loadings plot can only partially explain the structural features which were found as important by the model. In this GOLPE offers the possibility to use active plots.[35] Here, one draws a vector linking pairs of objects in a 2D scores plot which is then translated into isocontour plots that identify those variables which contribute most to differentiate the selected objects. To obtain such isocontours, GOLPE calculates the difference between the two points for the first and second principal component and projects these differences back into the original space (a pseudofield) using the PCA loading. The result is a grid plot of the differences in the pseudofields that highlight the object differences for the corresponding probe. Using these plots, one is able to answer directly the questions of (a) where are the regions that can produce selective interaction with respect to the start and end points of the drawn vector (which e.g. could connect a pair of protein targets) and (b) which interaction (i.e. probe) is responsible for this difference. It should be emphasized that these plots are different for each probe, while in the original GRID/PCA method, only a single loadings plot was obtained.

Computational Details

Protein Structures. The 3D structures of thrombin, trypsin, and factor Xa used in this study were taken from in house X-ray data[36] or from the Research Collaboratory for Structural Bioinformatics (RCSB) protein database[37] and are compiled in Table 1. Crystallographic water molecules, bound

Table 1. List of Protein Structures Used in This Study

enzyme (ligand)	organism[a]	PDB entry code[b]	resolution (Å)	rms[c]
thrombin (E1181)	H	in house	2.2	
thrombin (R794)	H	in house	*d*	0.379
thrombin (apo-form)	H	1A3B	1.8	0.249
thrombin (DCTH)	B	in house	*d*	0.599
trypsin (E1165)	H	in house	1.8	0.669
trypsin (E1165, SO_4)	H	in house	1.8	0.669
trypsin (EXIL9)	H	in house	1.75	0.665
factor Xa (DX-9065A)	H	1FAX	3.0	0.940
factor Xa (apo-form)	H	in house	3.0	0.972

[a] Human (H) or bovine (B) origin. [b] Entry code for the Protein Data Bank; 'in house' refers to structures described in ref 36. [c] Root-mean-square deviation in Å of the Cα atoms used for alignment on the thrombin (E1181) structure. [d] Not available, but <3.0 Å.

Table 2. List of the GRID Probe Types Used in This Study

name	chemical group	additional properties
DRY	hydrophobic probe	empirical term for entropy
C3	methyl group	
N1	neutral flat NH, e.g. amide	
N:	sp^3 N with lone pair bonded to 3 heavy atoms	
NH=	sp^2 NH with lone pair	
N1+	sp^3 amine NH cation	charge +1
NM3	trimethylammonium cation	charge +1
O	sp^2 carbonyl oxygen	
OH	phenol or carboxy OH	
OS	oxygen of sulfone/sulfoxide	
COO−	carboxylic acid anion	multiatom probe, charge −1
O::	sp^2 carboxy oxygen atom	

ligands, and counterions were removed. For the protein structure alignment, we first determined structurally conserved regions (SCR) of the different protein targets by visual inspection. Then the protein structures were aligned according to the Cα traces of these SCRs using the InsightII[38] modeling software. As we use more than one structure per target protein, the exact details of the alignment protocol have only a minor influence on the results.

GRID Calculations. The calculations were performed with version 17 of the GRID software;[39] the proteins were considered rigid. Hydrogens were added with the program GRIN (part of the GRID package). The GRID box dimensions were chosen to encompass all relevant residues within the respective actives sites of the proteins. This resulted in a box size of 33 Å × 33 Å × 36 Å. The grid spacing was set to 1 Å; all other GRID input parameters retained their default values. Changing the distance between grid points to 0.75 Å had no significant effect on the results. The GRID probes were chosen to represent all relevant interactions (hydrophobic, charge−charge, and hydrogen bond donor/acceptor) and to cover the most common chemical groups used in known thrombin and factor Xa inhibitors. Table 2 lists the GRID probes used, including a short description which functional groups they represent.

Data Pretreatment. The calculated GRID fields were imported into GOLPE. Then all data points with an absolute value smaller than 0.01 or with a standard deviation below 0.03 were set to zero. A major complication in the interpretation of the CPCA pseudofield contours arises when the sign of the interaction energy changes between different regions highlighted as important for selectivity. Moreover, a positive interaction energy is in most cases due to a unfavorable steric interaction. Therefore, the data generally was pretreated by setting the maximum positive cutoff to zero. By doing so, one can focus on favorable interaction energies between the objects and the probes, which significantly facilitates the interpretation of the contour plots. We note, however, that in certain cases just the consideration of the positive interaction energy (i.e. setting a minimum cutoff at zero) may be instructive.

Region Cutouts. The region cutout tool was used to focus the data analysis on the region of the active site. The relatively large inhibitor NAPAP was used as a template to define the relevant region for binding, and all data points within 3 Å of the inhibitor atoms were retained. This reduces the number of field variables to 3% of the original amount. To focus on each of the three binding main pockets (S1, P, and D) separately, NAPAP was cut into three fragments according to the functional groups residing in the corresponding binding pockets and the cutout tool was used with these NAPAP fragments.

Results

In the following we will discuss the results of the CPCA approach which point toward the main differences between the three serine proteases that can be used to design selective ligands. For each of the three main binding pockets, we first try to establish which interaction type could infer selectivity, then the exact region of this interaction is identified. The first task is accomplished by examining the distribution of the *x* variable values within the binding pocket; the exact region is then displayed as a CPCA differential pseudofield contour.

S1 Pocket. With the exception of the A190S mutation in trypsin, the S1 pockets are very similar in all three proteins. The amino acid change at position 190 makes the S1 pocket of trypsin smaller and less hydrophobic than the corresponding pockets in thrombin and factor Xa. Figure 5a shows the distribution of the calculated interaction energies (after BUW) for 10 different GRID probes, color-coded by enzyme. As expected for similar pockets, for all 10 probes the interaction energies span the same ranges for the different enzymes; none of the interactions represented by the GRID probes are particularly favorable in one of the proteins. Although Ser190 in trypsin should favor hydrophilic interactions and thus higher energies for the corresponding probe types, no significant difference to the other two enzymes is found. This could be a first indication that this difference probably cannot successfully used to drive selectivity in a desired direction.

Asp189 at the bottom of the S1 pocket is known to be important for overall affinity of ligands with positively charged groups at the P1 position. However, since Asp189 is conserved within the three studied serine proteases, none of the enzymes show especially high interaction energies of the positively charged probes (N1+ or NM3). The interaction values of the other probes are of the same order of magnitude as the charged probes − this is a consequence of the BUW procedure, which scales all interactions to unit variance.

Figure 6 shows the pseudofield differences for the GRID carbonyl oxygen probe between thrombin and trypsin. The large cyan contour indicates a region where interaction between the O probe and the enzyme is more favorable in trypsin than in thrombin. Therefore, making a potential ligand more hydrophobic should shift selectivity toward thrombin. Indeed, there is experimental evidence for this behavior.[40,41] Below the cyan region a yellow spot close to the hydrogen atom of the serine residue indicates an interaction that is less favorable in trypsin than in thrombin. Since the mutation from alanine to serine decreases the size of the S1 pocket, the introduction of electrostatically favorable groups does not necessarily lead to a selectivity increase toward trypsin, since the sterical demand of the ligand

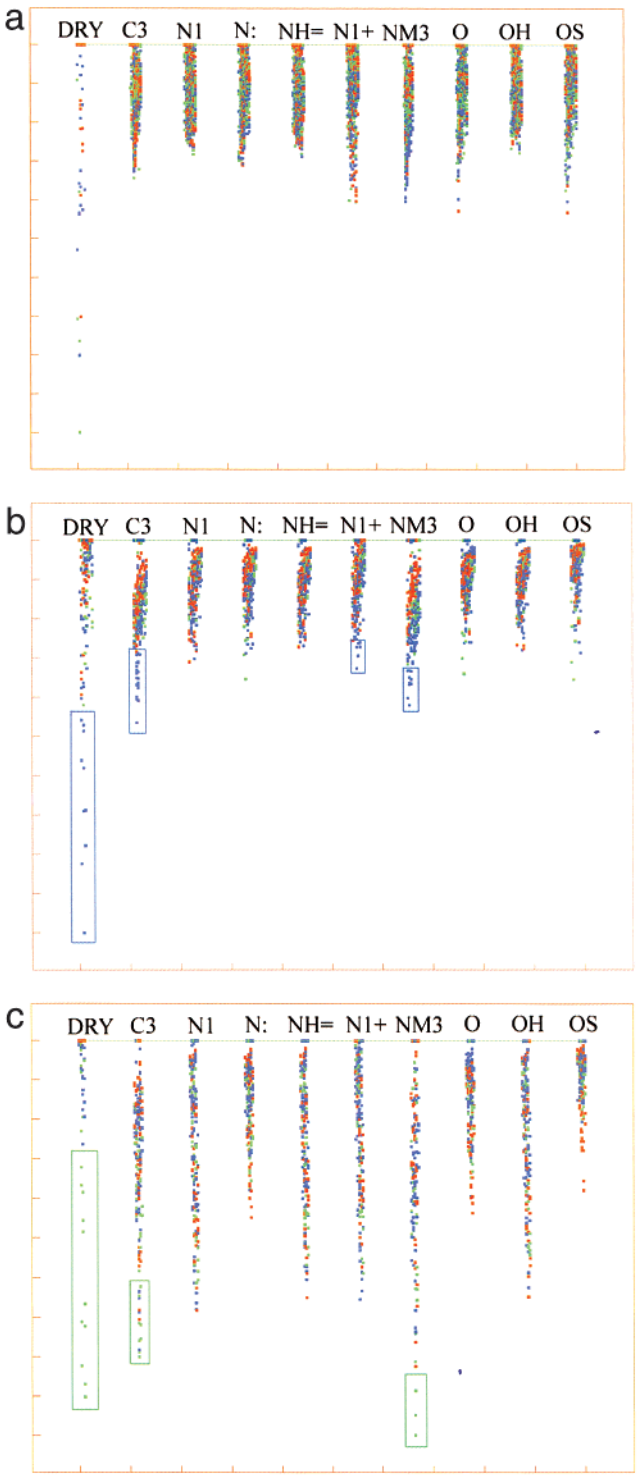


Figure 5. *x* Variable distribution for the 10 grid probes within the main pockets of the active sites (after BUW). Blue dots indicate energies in thrombin, red dots in trypsin, and green dots in factor Xa: (a) interactions in the S1 pocket, (b) interactions in the P pocket, and (c) interactions in the D pocket. Interaction energies that show a prevalence for one of the enzymes are highlighted with colored boxes. Although the overall analysis was performed with several crystal structures for each protease, for the plots only one representative structure for each enzyme was used to improve the readability of the figures.

may collide with the reduced space in the S1 pocket of trypsin. This becomes even more obvious with the sterically even more demanding GRID OH probe (data not shown). In this case the favorable electrostatic interaction is completely offset by the unfavorable steric demands. Therefore, even though a hydrophilic group will be electrostatically favored by trypsin, the steric restraints will eventually shift selectivity away from trypsin toward thrombin or factor Xa. This observation has also been reported in the literature: methyl-substituted aminopyridyl groups appear to be more selective for thrombin vs trypsin than the corresponding nonsubstituted ones,[40,41] and replacing the prototypical arginine side chain in a peptidic inhibitor by tryptophane leads to a more than 1000-fold selectivity increase.[42] Similarly, going from a benzamidine P1 group to the more hydrophobic and bulkier 1-aminoisoquinoline increases selectivity for thrombin by a factor of about 20.[43]

P Pocket. The energy distributions for the GRID probes in the P pocket are shown in Figure 5b. In contrast to the results for the S1 pocket, several probes have particulary high interaction energies in one of the proteins. For both the hydrophobic DRY and C3 probes, the highest interaction energies are found in the thrombin P pocket. The cationic probes N1+ and NM3 are also particularly favorable for thrombin. Therefore, a ligand with either hydrophobic or positively charged functional groups in the region of the P pocket should improve selectivity toward thrombin.

Figure 7 shows the CPCA differential plot between thrombin and trypsin for the DRY probe. Right below the thrombin insertion loop residues Tyr60A and Trp60D a large yellow contour indicates that the introduction of hydrophobic groups in a potential ligand at that position would increase its selectivity for thrombin. The steric repulsion of the Trp60D residue cannot be seen in this plot, as the postive cutoff was set to zero, thus excluding all unfavorable interactions. This hydrophobic region has been identified before as important for selective inhibitors of thrombin, and most thrombin inhibitors point either proline side chain atoms[19,40] or another lipophilic group in this direction.[20,28,41,44−47] As can be seen in Figure 7, the hydrophobic piperidine group of NAPAP, for example, occupies the same region as the yellow contour, thus contributing to NAPAP's selectivity for thrombin.

Very similar contour plots are obtained if the P pockets of thrombin and factor Xa are compared. These results underline the known importance of the insertion loop of thrombin in determining its specificity toward substrates and inhibitors.

The results of the interaction energy distributions suggest that a positive interaction (N1+ and NM3 probes) also could be exploited to increase selectivity toward thrombin. However, inspection of the CPCA differential plots for both the N1+ and the NM3 probe shows that besides a favorable area for thrombin in the P pocket, nearby regions favor interactions with both factor Xa and trypsin. Therefore, it would probably be difficult to alter the specifity of ligands with this type of interaction. This is in agreement with experimental findings that a positive charge does not improve selectivity toward thrombin.[20]

As expected from the *x* variable distribution, with the other probes no significant selectivity regions can be identified.

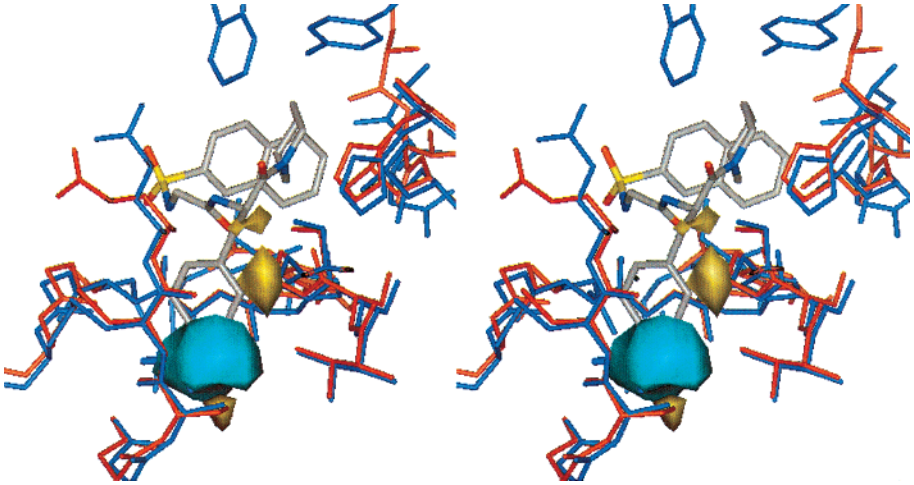


Figure 6. CPCA pseudofield plot for the GRID carbonyl oxygen probe within the S1 pocket. Shown are field differences between thrombin (blue) and trypsin (red). Cyan contours indicate regions where interaction energies are more favorable for trypsin and yellow contours regions where interaction energies are more favorable for thrombin (or less favorable for trypsin). To aid the orientation, the inhibitor NAPAP is shown in gray.

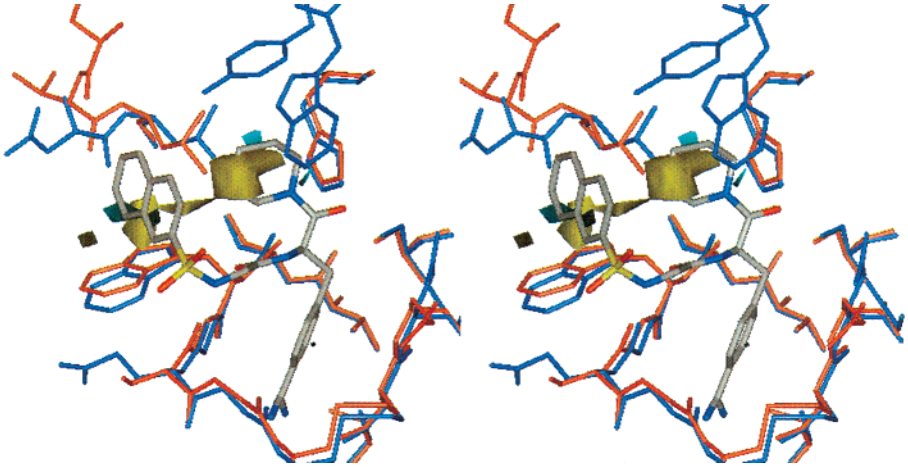


Figure 7. CPCA pseudofield plot showing field differences between thrombin (blue) and trypsin (red) for the GRID DRY probe within the P and D pockets. Yellow contours indicate regions where a hydrophobic interaction increases selectivity toward thrombin. NAPAP is drawn in gray.

D Pocket. Figure 5c shows the energy distribution for the probes in the D pocket. The DRY, C3, and the NM3 probes show the highest interaction energies for factor Xa, and therefore, these interactions are expected to be the most important ones for selectivity toward this enzyme. In addition, for trypsin the O and OS probes are particularly favorable.

In factor Xa, the D pocket is lined by aromatic residues in addition to the negatively charged Glu97; therefore, the finding that hydrophobic or cation−π interactions are important again underlines the strength of our *x* variable weighting procedure.

Figure 8 shows the CPCA pseudodifference field plot for the cationic NM3 probe. One large favorable cyan contour for the NM3 probe is located in the hydrophobic box, pointing to the possible cation−π interaction in factor Xa. Toward the carboxyl group of Glu97, a second favorable spot can be seen. This region has been called the factor Xa cation hole.[29,48] Therefore, the introduction of positively charged or polarized groups at either position will increase selectivity for factor Xa over thrombin or trypsin. Indeed, a number of highly active and specific factor Xa inhibitors have positively charged groups directed toward the contour blobs in the D pocket.[22,24,49−51] A similar picture is obtained for the hydrophobic DRY probe when going from thrombin to factor Xa. Various favorable cyan blobs are situated within the hydrophobic box of factor Xa. Therefore, the introduction of hydrophobic groups in a potential ligand that reach into the D pocket is especially favorable for factor Xa and will increase selectivity over thrombin and trypsin.[52,53] In a series of 1,2-dibenzamidobenzene-derived inhibitors of factor Xa, optimization of the hydrophobic interaction of the inhibitor with the D pocket significantly enhanced selectivity for the inhibition of factor Xa relative to thrombin and trypsin.[54]

The fact that in trypsin the interaction of the O and OS probes is particularly favorable is probably due to the E217S mutation, which removes the negatively charged acid and provides a hydrogen bond donor for interaction with the oxygens. Although this residue is at the edge of the D pocket and is pointing its side chain toward the solvent, this again highlights the importance of hydrophobic interactions for selective inhibitors of thrombin and factor Xa.

The Glu/Gln192 Position. Another amino acid difference between the three proteases involves position 192, located at the entrance of the S1 pocket. The

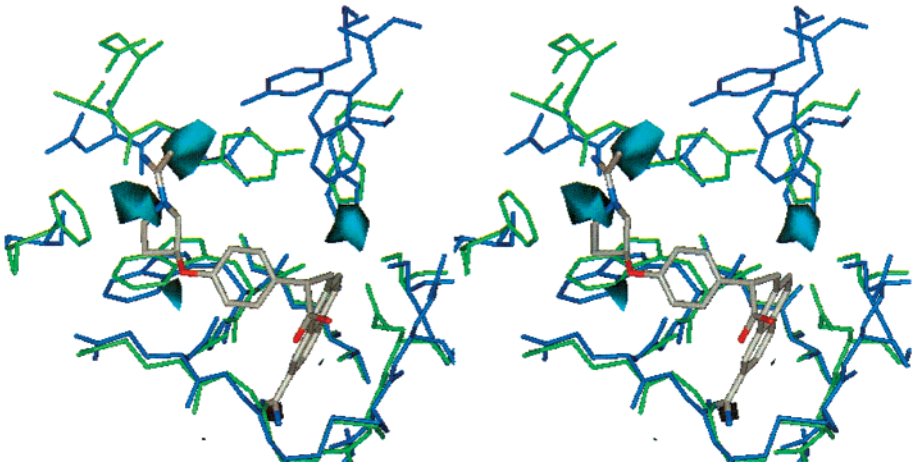


Figure 8. CPCA pseudofield plot of the field differences between thrombin (blue) and factor Xa (green) for the GRID NM3 probe. The cyan contours indicate regions where interactions of the NM3 type enhance selectivity toward factor Xa. For comparison, the selective factor Xa inhibitor DX-9065A is shown in gray.

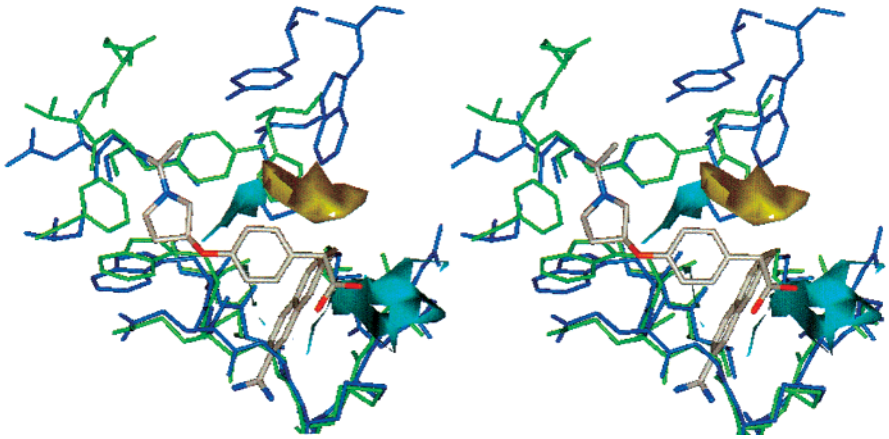


Figure 9. CPCA pseudofield plot of the field differences between thrombin (blue) and factor Xa (green) for the carboxylate probe within the P pocket and the entrance of the S1 pocket. Here, the pretreatment retained the positive interaction energies (see text). Blue contours indicate regions were a negatively charged group in the ligand increases selectivity toward factor Xa. The yellow contour shows the steric interaction with Trp60D in thrombin which is unfavorable due to the choice of positive interaction energies. The inhibitor DX-9065A is shown in gray.

glutamic acid in thrombin is replaced by glutamine both in trypsin and factor Xa. Therefore, we decided to use the GRID carboxylate probe to investigate the differences between the enzymes. In Figure 9 the pseudodifference plot for the carboxylate probe is shown for the difference from thrombin to factor Xa. Two cyan contour surfaces show regions where such a functional group in a ligand would improve selectivity for factor Xa over thrombin. The contour on the right is close to position 192 and is due to the fact that a negatively charged group in a ligand would interact more unfavorably with Glu192 in thrombin than with Gln in factor Xa where even a hydrogen bond interaction may be possible. An example where this interaction has been used is the selective factor Xa inhibitor DX-9065A. If its acid group which is close to this favorable area is removed, the factor Xa selectivity of DX-9065A over thrombin is reduced by a factor of more than 100.[55] Similarly, in the stereoisomer with the acid pointing in the other direction, the selectivity against thrombin is reduced by a factor of 15.[27] In a series of factor Xa inhibitors disclosed by Berlex, the addition of an appropriately placed carboxy group leads to a dramatic improvement of the selectivity versus thrombin; MD simulations suggested that this is a result of an electrostatic repulsion between the carboxy group of the inhibitor and Glu192;[56] the selectivity toward trypsin remains unchanged. Incorporation of an acidic functional group in 1,2-dibenzamidobenzene inhibitors of factor Xa also leads to an increased selectivity for factor Xa relative to thrombin, but not to trypsin.[54] These findings are consistent with the fact that only in thrombin is residue 192 a negatively charged glutamate.

A second cyan contour is located inside the P pocket close to the hydroxyl group of Tyr99. Thus, substituting a ligand with an acid group which points into this region could be another way to increase selectivity for factor Xa. An interesting example for this design criterion is the selective factor Xa inhibitor FX-2212A.[25] However, its acid group cannot directly interact with the tyrosine OH, and is not close to the cyan contour surface. In the crystal structure, a water occupies the area of the cyan blob, thus mediating the interaction between the acid group and the Tyr99 hydroxyl group. We note that for this example the positive part of the GRID interaction energies was used (i.e. the *minimum* cutoff was set to zero; see also Computational Details) to obtain contour plots that are easy to interpret.

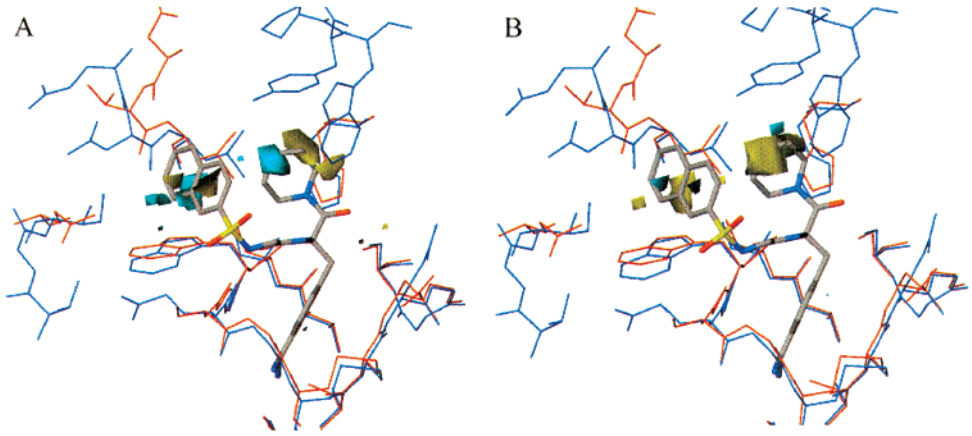
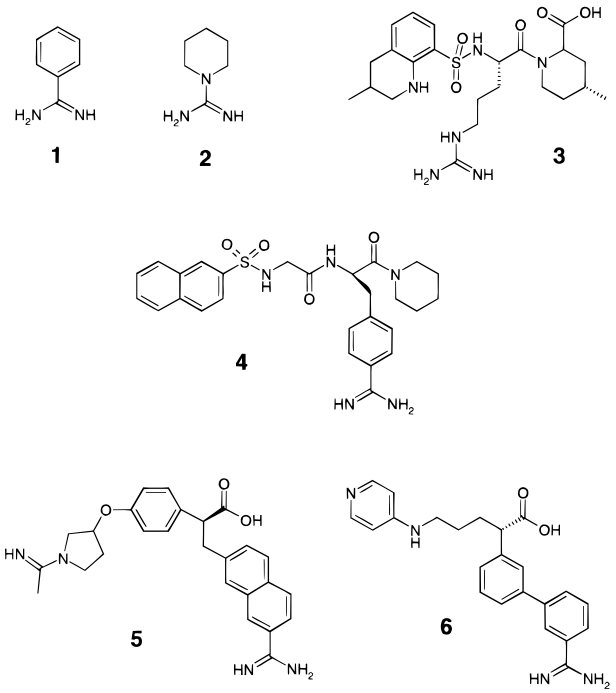


Figure 10. Comparison of the contour plots obtained by a simple difference of the GRID fields and the results of a CPCA for thrombin (blue) and trypsin (red): (A) difference of the DRY fields and (B) CPCA pseudofield plot of the field differences for the DRY probe. For comparison NAPAP is drawn in gray.

Table 3. Experimental Biological Activities and Structures of the Inhibitors Mentioned in This Study



compd	no.	measure	thrombin	trypsin	factor Xa	ref
benzamidine	**1**	K_i (μM)	0.300	0.031		28
amidinopiperidine	**2**	K_i (μM)	0.150	0.360		28
(2*R*,4*R*)-MQPA	**3**	K_i (μM)	0.019	5	210	44
NAPAP	**4**	K_i (μM)	0.0060	0.69	7.9	45
DX 9065A	**5**	IC_{50} (μM)	>2000		0.07	26, 55
FX-2212A	**6**	IC_{50} (μM)	>100		0.00272	25

Discussion

Since the meaning of the pseudofield differences is essentially the same as the field differences between two objects, i.e., structures, a naive approach would be to simply calculate those field differences from the original object fields. As an example, we compare such a simple difference of the hydrophobic fields (the DRY probe) of thrombin and trypsin with the pseudofield differences obtained from a CPCA. From Figure 10 it is obvious that the contour plot of the simple field difference between two objects (Figure 10A) is far more difficult to interpret than the corresponding pseudofield plot (Figure 10B). Especially in the P pocket, the difference method does not yield a clear picture: is the addition of hydrophobic groups in the ligand beneficial for selectivity toward thrombin (yellow contour) or not (blue contour)? In contrast to this, the pseudofield difference plot unequivocally predicts that the placement of hydrophobic residues in the P pocket will be favorable for trypsin compared to thrombin. Similarly, in the D pocket, the slightly different orientation of the tryptophan rings in thrombin and trypsin dominates the sign and magnitude of the field differences in this area.

Compared to the previous GRID/PCA method, the GRID/CPCA approach offers several important advantages. The first improvement is the possibility to better balance the different interaction energy scales as given in the GRID parametrization: due to the fact that charged probes span a much higher energy range than uncharged or hydrophobic probes, they tend to dominate a straightforward PCA. With the BUW pretreatment, the GRID energy scales do not longer bias the results against hydrophobic interactions. As a consequence, for this type of analysis, we could show for the first time that hydrophobic interactions are important determinants of selectivity between proteins. Moreover, as we can include many different enzymes in the analysis, the chance of artificially increasing the variance of a probe type is greatly reduced. The second improvement is based on the fact that every target structure is treated as one object of the PCA. This allows to compare more than two targets simultaneously and makes it possible to derive complicated selectivity profiles (i.e. how to modify the ligand to inhibit targets A and B, but not C). One can also include several coordinate sets for the same target protein for the analysis, either from independent crystal structures or from snapshots of a MD simulation. This allows some 'averaging' over the individual structures and thus a reduction of the noise that is due to subtle experimental differences that are present but do not really contribute to specificity differences of the targets.

The results for the GRID/CPCA selectivity are in excellent agreement with the experimental data on selectivity in the thrombin/trypsin/factor Xa system. The S1 pocket, with its mainly conserved residues, can be used on its own to drive selectivity of a compound toward one of the three enzymes, if both electrostatic

and steric requirements are adequately satisfied. In the P pocket, the CPCA predicts hydrophobic moieties to enhance selectivity for thrombin. This has been an important design criterion for a large number of selective thrombin inhibitors. Less explored has been the chance of enhancing selectivity for thrombin with a positively charged functional group in this pocket.

The D pocket can also be used to drive selectivity of potential protease inhibitors. Especially the design of selective factor Xa inhibitors has taken advantage of the hydrophobic box in factor Xa which can accommodate cationic functional groups.[49] While the potential role of cation$-\pi$ interactions has been suggested in the past,[24,57] we were able to show for the first time that, indeed, this is also a direct selectivity criterion that can be used in the design of new factor Xa inhibitors. Therefore, the combination of a hydrophobic group carrying positively polarized atoms that might additionally interact with Glu97 should result in a potent and selective factor Xa inhibitor. Furthermore, we show that also the E192Q mutation can be used to increase the selectivity toward factor Xa.

Recently, the primed side of the thrombin active site has attracted some interest as an additional possibility for the design of potent and selective thrombin inhibitors.[58−60] However, much less known is experimentally about these binding pockets and the factors that determine specificity. Therefore, in this validation study of the CPCA method, the primed pockets were not further investigated.

Conclusions

The GRID/CPCA allows a detailed analysis of structural differences important for selectivity within a given family of target proteins. Based on the 3D structures of the proteins, this method analyzes selectivity differences from the view of the receptor and is therefore independent of the availability of appropriate ligands for a ligand-based QSAR analysis. The new procedure overcomes several of the disadvantages identified in selectivity analyses by the GRID/PCA approach which has been used in the past. Graphical representation of the pertinent differences responsible for selectivtiy can be directly translated into suggestions on how an existing ligand can be modified (or which functional groups a novel ligands should have) to enhance selectivity toward a given target.

Appendix

The GOLPE CPCA Algorithm. CPCA uses the same objective function as PCA. It explains at best the overall variance of the matrix and yields an overall model nearly identical to the PCA model. The original algorithm is an adaptation of the NIPALS algorithm.[10]

The algorithm starts with an initial guess of superscores t_T, which are used to compute the different block loadings p_b. Then, the block loadings are used to compute the block scores t_b. All the block scores are collected in a matrix T. The relationship between the block scores T and the superscores is defined by a weight matrix (w_T) which indicates the participation of each block to the overall scores. The whole procedure is then iterated until convergence of the superscores:

guess initial t_T

loop until convergence of t_T

$$p_b = X^T{}_b t_T / t^T{}_T t_T$$

normalize p_b

$$t_b = X_b p_b$$

$$T = [t_1 ... t_B]$$

$$w_T = T^T t_T / t_T^T t_T$$

normalize w_T

$$t_T = T w_T$$

end loop

As is usual in NIPALS, the algorithm works dimension after dimension, deflating the original matrix using the block scores and loadings:

$$p_b = X_b^T t_T / t_T^T t_T$$

$$X_b = X_b - t_T p_b$$

As was first shown by Westerhuis et al.,[33] the block scores and loading can be obtained from a regular PCA, because at convergence the PCA scores and the CPCA superscores are identical $(t = t_T)$. If it is so, no iterative algorithm is needed, and we obtain the block loading as

$$p_b = X_b^T t / t^T t$$

and the block scores and weights as

$$t_b = X_b p_b$$

$$T = [t_1 ... t_B]$$

$$w_T = T^T t / t^T t$$

In GOLPE, the original CPCA algorithm was modified to best suit our needs, and the matrix is deflated using the PCA scores and loadings instead of the block scores and block loadings.

References

(1) Kubinyi, H. *3D QSAR in Drug Design: Theory Methods and Applications;* ESCOM: Leiden, 1993.
(2) Kim, K. H.; Greco, G.; Novellino, E. A critical review of recent CoMFA applications. *Perspect. Drug Discovery Des.* **1998**, *12*, 257−315.
(3) Cruciani, G.; Goodford, P. J. A search for specificity in DNA-drug interactions. *J. Mol. Graph.* **1994**, *12*, 116−129.
(4) Pastor, M.; Cruciani, G. A novel strategy for improving ligand selectivity in receptor-based drug design. *J. Med. Chem.* **1995**, *38*, 4637−4647.
(5) Matter, H.; Schwab, W. Affinity and Selectivity of Matrix Metalloproteinase Inhibitors: A Chemometrical Study from the Perspective of Ligands and Proteins. *J. Med. Chem.* **1999**, *42*, 4506−4523.
(6) Filipponi, E.; Cecchetti, V.; Tabarrini, O.; Bonelli, D.; Fravolini, A. Chemometric rationalization of the structural and physicochemical basis for selective cyclooxygenase-2 inhibition: toward more specific ligands. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 277−292.
(7) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.

(8) Boobbyer, D. N.; Goodford, P. J.; McWhinnie, P. M.; Wade, R. C. New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. *J. Med. Chem.* **1989**, *32*, 1083−1094.
(9) Wade, R. C.; Clark, K. J.; Goodford, P. J. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. Ligand probe groups with the ability to form two hydrogen bonds. *J. Med. Chem.* **1993**, *36*, 140−147.
(10) Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37−52.
(11) Davie, E. W.; Fujikawa, K.; Kisiel, W. The coagulation cascade: initiation, maintenance, and regulation. *Biochemistry* **1991**, *30*, 10363−10370.
(12) Kaiser, B. Thrombin and factor Xa inhibitors. *Drugs Future* **1998,** *23*, 423−436.
(13) Wiley, M. R.; Fisher, M. J. Small-molecule direct thrombin inhibitors. *Exp. Opin. Ther. Patents* **1997**, *7*, 1265−1282.
(14) Tapparelli, C.; Metternich, R.; Ehrhardt, C.; Cook, N. S. Synthetic low-molecular weight thrombin inhibitors: molecular design and pharmacological profile. *Trends Pharmacol. Sci.* **1993**, *14*, 366−376.
(15) Hauptmann, J.; Stiirzebecher, J. Synthetic inhibitors of thrombin and factor Xa: from bench to bedside. *Thromb. Res.* **1999**, *93*, 203−241.
(16) Babine, R. E.; Bender, S. L. Molecular Recognition of Protein−Ligand Complexes: Applications to Drug Design. *Chem. Rev. (Washington, D.C.)* **1997**, *97*, 1359−1472.
(17) Claeson, G.; Scully, M. F.; Kakkar, V. V.; Deadman, J. *The Design of Synthetic Inhibitors of Thrombin;* Plenum Press: New York and London, 1993.
(18) Krishnan, R.; Zhang, E.; Hakansson, K.; Arni, R. K.; Tulinsky, A.; Lim-Wilby, M. S. L.; Levy, O. E.; Semple, J. E.; Brunck, T. K. Highly selective mechanism-based thrombin inhibitors: Structures of thrombin and trypsin inhibited with rigid peptidyl aldehydes. *Biochemistry* **1998**, *37*, 12094−12103.
(19) Stubbs, M. T.; Bode, W. A player of many parts: The spotlight falls on thrombin's structure. *Thromb. Res.* **1993**, *69*, 1−58.
(20) Stiirzebecher, J.; Prasa, D.; Hauptmann, J.; Vieweg, H.; Wikstroem, P. Synthesis and structure-activity relationships of potent thrombin inhibitors: Piperazides of 3-amidinophenylalanine. *J. Med. Chem.* **1997**, *40*, 3091−3099.
(21) Bursi, R.; Grootenhuis, P. D. J. Comparative molecular field analysis and energy interaction studies of thrombin-inhibitor complexes. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 221−232.
(22) Stubbs, M. T.; Huber, R.; Bode, W. Crystal structures of factor Xa specific inhibitors in complex with trypsin: structural grounds for inhibition of factor Xa and selectivity against thrombin. *FEBS Lett.* **1995**, *375*, 103−107.
(23) Ewing, W. R.; Becker, M. R.; Manetta, V. E.; Davis, R. S.; Pauls, H. W.; Mason, H.; Choi-Sledeski, Y. M.; Green, D.; Cha, D.; Spada, A. P.; Cheney, D. L.; Mason, J. S.; Maignan, S.; Guilloteau, J.-P.; Brown, K.; Colussi, D.; Bentley, R.; Bostwick, J.; Kasiewski, C. J.; Morgan, S. R.; Leadley, R. J.; Dunwiddie, C. T.; Perrone, M. H.; Chu, V. Design and Structure−Activity Relationships of Potent and Selective Inhibitors of Blood Coagulation Factor Xa. *J. Med. Chem.* **1999**, *42*, 3557−3571.
(24) Maduskuie, T. P., Jr.; McNamara, K. J.; Ru, Y.; Knabb, R. M.; Stouten, P. F. W. Rational Design and Synthesis of Novel, Potent Bis-phenylamidine Carboxylate Factor Xa Inhibitors. *J. Med. Chem.* **1998**, *41*, 53−62.
(25) Kamata, K.; Kawamoto, H.; Honma, T.; Iwama, T.; Kim, S. H. Structural basis for chemical inhibition of human blood coagulation factor Xa. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6630−6635.
(26) Herbert, J. M.; Bernat, A.; Dol, F.; Herault, J. P.; Crepon, B.; Lormeau, J. C. DX 9065A, a novel, synthetic, selective and orally active inhibitor of factor Xa: in vitro and in vivo studies. *J. Pharmacol. Exp. Ther.* **1996**, *276*, 1030−1038.
(27) Nagahara, T.; Yokoyama, Y.; Inamura, K.; Katakura, S.-I.; Komoriya, S.; Yamaguchi, H.; Hara, T.; Iwamoto, M. Dibasic (Amidinoaryl)propanoic Acid Derivatives as Novel Blood Coagulation Factor Xa Inhibitors. *J. Med. Chem.* **1994**, *37*, 1200−1207.
(28) Hilpert, K.; Ackermann, J.; Banner, D. W.; Gast, A.; Gubernator, K.; Hadvary, P.; Labler, L.; Müller, K.; Schmid, G.; et al. Design and Synthesis of Potent and Highly Selective Thrombin Inhibitors. *J. Med. Chem.* **1994**, *37*, 3889−3901.
(29) Brandstetter, H.; Kühne, A.; Bode, W.; Huber, R.; von der Saal, W.; Wirthensohn, K.; Engh, R. A. X-ray structure of active site-inhibited clotting factor Xa. Implications for drug design and substrate recognition. *J. Biol. Chem.* **1996,** *271*, 29988−29992.
(30) Böhm, M.; Stürzebecher, J.; Klebe, G. 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. *J. Med. Chem.* **1999**, *42*, 458−477.
(31) Wold, S.; Hellberg, S.; Lundstedt, T.; Sjostrom, M.; Wold, H. Symposium on PLS Model Building: Theory and Application, Frankfurt am Main, Germany, 1987.
(32) Wold, S.; Kettaneh, N.; Tjessem, K. Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection. *J. Chemom.* **1996**, *10*, 463−482.
(33) Westerhuis, J. A.; Kourti, T.; Macgregor, J. F. Analysis of multiblock and hierarchical PCA and PLS models. *J. Chemom.* **1998,** *12*, 301−321.
(34) GOLPE4.5; Multivariate Infometric Analysis, S.r.l., Viale dei Castagni, 16, Perugia, Italy; 1999.
(35) Cruciani, G. Presentation at the 12th European Symposium on QSAR, Copenhagen, Denmark 1998.
(36) Nar, H.; Tong, L. To be published.
(37) RCSB Protein Data Bank, operated by the Research Collaboratory for Structural Bioinformatics; http://www.rcsb.org/pdb/index.html.
(38) InsightII(98.0); Molecular Simulations Inc., San Diego, CA, 1998.
(39) GRID version 17; Molecular Discovery Ltd., West Way House, Elms Parade, Oxford, U.K.; 1998.
(40) Feng, D. M.; Gardell, S. J.; Lewis, S. D.; Bock, M. G.; Chen, Z. G.; Freidinger, R. M.; Naylor-Olsen, A. M.; Ramjit, H. G.; Woltmann, R.; Baskin, E. P.; Lynch, J. J.; Lucas, R.; Shafer, J. A.; Dancheck, K. B.; Chen, I. W.; Mao, S. S.; Krueger, J. A.; Hare, T. R.; Mulichak, A. M.; Vacca, J. P. Discovery of a novel, selective, and orally bioavailable class of thrombin inhibitors incorporating aminopyridyl moieties at the P1 position. *J. Med. Chem.* **1997**, *40*, 3726−3733.
(41) Sanderson, P. E. J.; Cutrona, K. J.; Dorsey, B. D.; Dyer, D. L.; McDonough, C. M.; Naylor-Olsen, A. M.; Chen, I. W.; Chen, Z.; Cook, J. J.; Gardell, S. J.; Krueger, J. A.; Lewis, S. D.; Lin, J. H.; Lucas, B. J., Jr.; Lyle, E. A.; Lynch, J. J., Jr.; Stranieri, M. T.; Vastag, K.; Shafer, J. A.; Vacca, J. P. L-374,087, an efficacious, orally bioavailable, pyridinone acetamide thrombin inhibitor. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 817−822.
(42) Malikayil, J. A.; Burkhart, J. P.; Schreuder, H. A.; Broersma R. J., J.; Tardif, C.; Kutcher L. W., I.; Mehdi, S.; Schatzman, G. L.; Neises, B.; Peet, N. P. Molecular design and characterization of an a-thrombin inhibitor containing a novel P1 moiety. *Biochemistry* **1997**, *36*, 1034−1040.
(43) Rewinkel, J. B. M.; Lucas, H.; Van Galen, P. J. M.; Noach, A. B. J.; Van Dinther, T. G.; Rood, A. M. M.; Jenneboer, A. J. S. M.; Van Boeckel, C. A. A. 1-Aminoisoquinoline as benzamidine isostere in the design and synthesis of orally active thrombin inhibitors. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 685−690.
(44) Kikumoto, R.; Tamao, Y.; Tezuka, T.; Tonomura, S.; Hara, H.; Ninomiya, K.; Hijikata, A.; Okamoto, S. Selective inhibition of thrombin by (2R,4R)-4-methyl-l-[N2-[1,2,3,4-tetrahydro-8-quinolinyl)sulfonyl]-L-arginyl]-2-piperidinecarboxylic acid. *Biochemistry* **1984**, *23*, 85−90.
(45) Renatus, M.; Bode, W.; Huber, R.; Stürzebecher, J.; Stubbs, M. T. Structural and functional analyses of benzamidine-based inhibitors in complex with trypsin: Implications for the inhibition of factor Xa, tPA, and urokinase. *J. Med. Chem.* **1998**, *41*, 5445−5456.
(46) Bone, R.; Lu, T. B.; Illig, C. R.; Soll, R. M.; Spurlino, J. C. Structural analysis of thrombin complexed with potent inhibitors incorporating a phenyl group as a peptide mimetic and aminopyridines as guanidine substitutes. *J. Med. Chem.* **1998**, *41*, 2068−2075.
(47) Wagner, J.; Kallen, J.; Ehrhardt, C.; Evenou, J. P.; Wagner, D. Rational design, synthesis, and X-ray structure of selective noncovalent thrombin inhibitors. *J. Med. Chem.* **1998**, *41*, 3664−3674.
(48) 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.
(49) Al-Obeidi, F.; Ostrem, J. A. Factor Xa inhibitors by classical and combinatorial chemistry. *Drug Discovery Today* **1998**, *3*, 223−231.
(50) Phillips, G.; Davey, D. D.; Eagen, K. A.; Koovakkat, S. E.; Liang, A.; Ng, H. P.; Pinkerton, M.; Trinh, L.; Whitlow, M.; Beatty, A. M.; Morrissey, M. M. Design, synthesis, and activity of 2,6-diphenoxypyridine-derived factor Xa inhibitors. *J. Med. Chem.* **1999**, *42*, 1749−1756.
(51) Gabriel, B.; Stubbs, M. T.; Bergner, A.; Hauptmann, J.; Bode, W.; Stürzebecher, J.; Moroder, L. Design of benzamidine-type inhibitors of factor Xa. *J. Med. Chem.* **1998**, *41*, 4240−4250.
(52) Becker, M. R.; Ewing, W. R.; Davis, R. S.; Pauls, H. W.; Ly, C.; Li, A.; Mason, H. J.; Choi-Sledeski, Y. M.; Spada, A. P.; Chu, V.; Brown, K. D.; Colussi, D. J.; Leadley, R. J.; Bentley, R.; Bostwick, J.; Kasiewski, C.; Morgan, S. Synthesis, SAR and in vivo activity of novel thienopyridine sulfonamide pyrrolidinones as factor Xa inhibitors. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 2753−2758.

(53) Choi-Sledeski, Y. M.; McGarry, D. G.; Green, D. M.; Mason, H. J.; Becker, M. R.; Davis, R. S.; Ewing, W. R.; Dankulich, W. P.; Manetta, V. E.; Morris, R. L.; Spada, A. P.; Cheney, D. L.; Brown, K. D.; Colussi, D. J.; Chu, V.; Heran, C. L.; Morgan, S. R.; Bentley, R. G.; Leadley, R. J.; Maignan, S.; Guilloteau, J.-P.; Dunwiddie, C. T.; Pauls, H. W. Sulfonamidopyrrolidinone Factor Xa Inhibitors: Potency and Selectivity Enhancements via P-1 and P-4 Optimization. *J. Med. Chem.* **1999**, *42*, 3572−3587.

(54) Wiley, M. R.; Weir, L. C.; Briggs, S.; Bryan, N. A.; Buben, J.; Campbell, C.; Chirgadze, N. Y.; Conrad, R. C.; Craft, T. J.; Ficorilli, J. V.; Franciskovich, J. B.; Froelich, L. L.; Gifford-Moore, D. S.; Goodson, T., Jr.; Herron, D. K.; Klimkowski, V. J.; Kurz, K. D.; Kyle, J. A.; Masters, J. J.; Ratz, A. M.; Milot, G.; Shuman, R. T.; Smith, T.; Smith, G. F.; Tebbe, A. L.; Tinsley, J. M.; Towner, R. D.; Wilson, A.; Yee, Y. K. Structure-Based Design of Potent, Amidine-Derived Inhibitors of Factor Xa: Evaluation of Selectivity, Anticoagulant Activity, and Antithrombotic Activity. *J. Med. Chem.* **2000**, *43*, 883−899.

(55) Katakura, S.; Nagahara, T.; Hara, T.; Iwamoto, M. A novel factor Xa inhibitor: structure−activity relationships and selectivity between factor Xa and thrombin. *Biochem. Biophys. Res. Commun.* **1993**, *197*, 965−972.

(56) Guilford, W. J.; Shaw, K. J.; Dallas, J. L.; Koovakkat, S.; Lee, W.; Liang, A.; Light, D. R.; McCarrick, M. A.; Whitlow, M.; Ye, B.; Morrissey, M. M. Synthesis, characterization, and structure-activity relationships of amidine-substituted (bis)benzylidene-cycloketone olefin isomers as potent and selective factor Xa inhibitors. *J. Med. Chem.* **1999**, *42*, 5415−5425.

(57) Lin, Z.; Johnson, M. E. Proposed cation-π mediated binding by factor Xa: a novel enzymatic mechanism for molecular recognition. *FEBS Lett.* **1995**, *370*, 1−5.

(58) Rezaie, A. R.; Olson, S. T. Contribution of lysine 60f to S1′ specificity of thrombin. *Biochemistry* **1997**, *36*, 1026−1033.

(59) St. Charles, R.; Matthews, J. H.; Zhang, E. L.; Tulinsky, A. Bound structures of novel P3−P1′ β-strand mimetic inhibitors of thrombin. *J. Med. Chem.* **1999**, *42*, 1376−1383.

(60) Martin, P. D.; Malkowski, M. G.; DiMaio, J.; Konishi, Y.; Ni, F.; Edwards, B. F. P. Bovine Thrombin Complexed with an Uncleavable Analogue of Residues 7−19 of Fibrinogen Aα: Geometry of the Catalytic Triad and Interactions of the P1′, P2′, and P3′ Substrate Residues. *Biochemistry* **1996**, *35*, 13030−13039.

JM000934Y