Virtual Screening and Scaffold Hopping Based on GRID Molecular Interaction Fields

Marie M. Ahlström,*[,†,‡] Marianne Ridderström,[†] Kristina Luthman,[‡] and Ismael Zamora[§,||]

DMPK & BAC Department, AstraZeneca R&D Mölndal, SE-431 81 Mölndal, Sweden, Department of Chemistry, Medicinal Chemistry, Göteborg University, SE-412 96 Gothenburg, Sweden, Lead Molecular Design, S.L., Vallés 96-102 (27) E-08190, Sant Cugat del Vallés, Spain, and Institut Municipal d'Investigació Medica (IMIM), Universitat Pompeu Fabra, Doctor Aiguader 80, 08003 Barcelona, Spain

Received December 10, 2004

In this study, a set of strategies for structure-based design using GRID molecular interaction fields (MIFs) to derive a pharmacophoric representation of a protein is reported. Thrombin, one of the key enzymes involved in the blood coagulation cascade, was chosen as the model system since abundant published experimental data are available related to both crystal structures and structurally diverse sets of inhibitors. First, a virtual screening methodology was developed either using a pharmacophore representation of the protein based on GRID MIFs or using GRID MIFs from the 3D structure of a set of chosen thrombin inhibitors. The search was done in a 3D multiconformation version of the Available Chemical Directory (ACD) database, which had been spiked with 262 known thrombin inhibitors (multiple conformers available per compound). The model managed to find 80% of the known thrombin inhibitors among the 74 291 conformers in the ACD by only searching 5% of the database; hence, a 15-fold enrichment of the library was achieved. Second, a scaffold hopping methodology was developed using GRID MIFs, giving the scaffold interaction pattern and the shape of the scaffold, together with the distance between the anchor points. The scaffolds reported by Dolle in the Journal of Combinatorial Chemistry summaries (2000 and 2001) and scaffolds built or derived from ligands cocomplexed with the thrombin enzyme were parameterized using a new set of descriptors and saved into a searchable database. The scaffold representation from the database was then compared to a template scaffold (from a thrombin crystal structure), and the thrombin-derived scaffolds included in the database were found among the top solutions. To validate the usefulness of the methodology to replace the template scaffold, the entire molecule was built (scaffold and side chains) and the resulting compounds were docked into the active site of thrombin. The docking solutions showed the same binding pattern as the cocomplexed compound, hence, showing that this method can be a valuable tool for medicinal chemists to select interchangeable core structures (scaffolds) in an easy manner and retaining the binding properties from the original ligand.

INTRODUCTION

Recent developments in combinatorial chemistry and high-throughput screening (HTS) facilitate the synthesis and pharmacological profiling of large numbers of interesting compounds. The results generate large databases of molecular structures, which play an increasingly important role in drug discovery.

To enable efficient use of stored data, molecular similarity searches in large chemical databases need representations of molecules that can correctly describe the relevant chemical information and, at the same time, be fast enough to allow an evaluation of thousands of compounds in a short period of time. The most accurate molecular representations can be defined by quantum chemical descriptors, but such calculations usually tend to be slow, but at the other extreme, simple atom and bond counts are generally too trivial to discriminate between many molecules; hence, a suitable compromise is needed. Relevant descriptors are those based on 2D and 3D substructural fragments or properties.[1] Several techniques for performing a similarity search based on such descriptors with the aim of selecting compounds with a high probability to be active in a certain biological test system have been published.[2−11] An important strategy in the development of novel lead compounds is virtual screening. Virtual screening can either be performed by the docking of potential ligands into a protein binding site or by similarity-based virtual screening, for example, by using a pharmacophore searching procedure. In this study, distances between GRID molecular interaction fields (MIFs) obtained from the protein structure were used as a "pharmacophore". The distance-based description of the protein is compared to the distances between structural properties in the possible ligand. The MIF calculation and similarity process is computationally inexpensive and enables the analysis of large data sets, of which interesting hits could be further analyzed by more accurate methods such as docking procedures. One important element of virtual screening is the ability to perform similarity searches to find biologically similar but chemically different compounds. This is possible if the substructure

* Corresponding author tel.:+46 31 776 1274; fax: +46 31 776 3786; e-mail: marie.m.ahlstrom@astrazeneca.com.
[†] DMPK & BAC Department, AstraZeneca R&D Mölndal.
[‡] Göteborg University.
[§] Lead Molecular Design, S.L., Sant Cugat del Vallés, Spain.
[||] IMIM, Universitat Pompeu Fabra.

10.1021/ci049626p CCC: $30.25
Published on Web 07/12/2005

analysis is based on 3D descriptors but not if it is based on 2D descriptors. Such a similarity search is critical if scaffold replacement is to be performed.

A problem that has been addressed in combinatorial library design is the selection of suitable scaffolds useful in the development of libraries and to find better hits. Different approaches to identify new chemical scaffolds that share or improve the biological activity of known drug target ligands or techniques to select suitable scaffolds useful in the development of libraries targeted for a specific activity can be found in the literature.[12−19]

The aim of the present study has been to develop a unique and efficient procedure to perform virtual screening. A methodology using GRID MIFs as a description of a molecule to derive pharmacophoric points possible for use in the discrimination between ligands and nonligands for receptors or enzymes has been developed. The method is highly useful in similarity searches of large databases and will hopefully lead to more focused libraries comprised of compounds fulfilling specific desired properties. The method has been validated using a set of 262 known thrombin inhibitors that have been added to a 3D version of the Available Chemical Directory (ACD) database to test whether it was possible to retrieve the compounds from the database or not.[20−32]

A second aim of this work was to explore the usefulness of GRID MIF-derived descriptors in performing a scaffold or substructure replacement, so-called *scaffold hopping*.[3] The scaffolds could be used to construct new molecular structures, which could be evaluated from docking experiments. In contrast to most other techniques that only consider geometrical fits of core templates, the scaffold-hopping methodology proposed in this study takes into account the shape, geometry, and protein−ligand interaction similarities as well as synthetability.

We have used the thrombin enzyme as the model system as there are several crystal structures of the thrombin apoprotein and thrombin cocomplexed with inhibitors available in the Protein Data Bank (PDB). Moreover, abundant experimental data are published, and there are structurally diverse sets of inhibitors known for this enzyme. The basic understanding of the molecular mechanism of thrombin is extensive. Its role in the blood coagulation cascade is to convert the soluble plasma protein fibrinogen into the insoluble fibrin that forms part of the blood clot;[33] thus, the development of inhibitors of thrombin is expected to lead to drugs useful in the treatment of several cardiovascular diseases.[34]

The thrombin enzyme active site consists of three main pockets (Figure 1), the S1 pocket with the acidic side chain of Asp189 positioned at the bottom of the specificity site; the S2 pocket where the residues His57, Tyr60A, Trp60D, and Leu99 define the shape of a small lipophilic pocket; and the S3 pocket consisting of Leu99, Ile174, and the aromatic side chain of Trp215.[35,36] Lots of efforts have been devoted to the discovery of novel classes of inhibitors. Most inhibitors known today form a covalent bond to Ser195, which is part of the catalytic triad (Ser195, His57, and Asp102) in the active site of thrombin.[37] Noncovalent thrombin inhibitors are compounds that rather form strong hydrophobic and electrostatic interactions with the enzyme. In the present study, we have focused on noncovalent thrombin inhibitors.

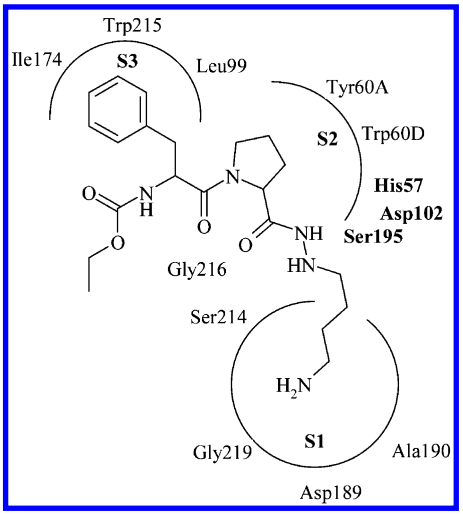


Figure 1. Schematic picture of the active site of thrombin, highlighting the relevant amino acids surrounding the inhibitor Eoc-D-Phe-Pro-Abh (PDB: 1ae8). The thrombin enzyme active site consists of three main pockets, the S1, S2, and S3 pockets. The catalytic triad of thrombin consists of the residues Ser195, His57, and Asp102.

MATERIALS

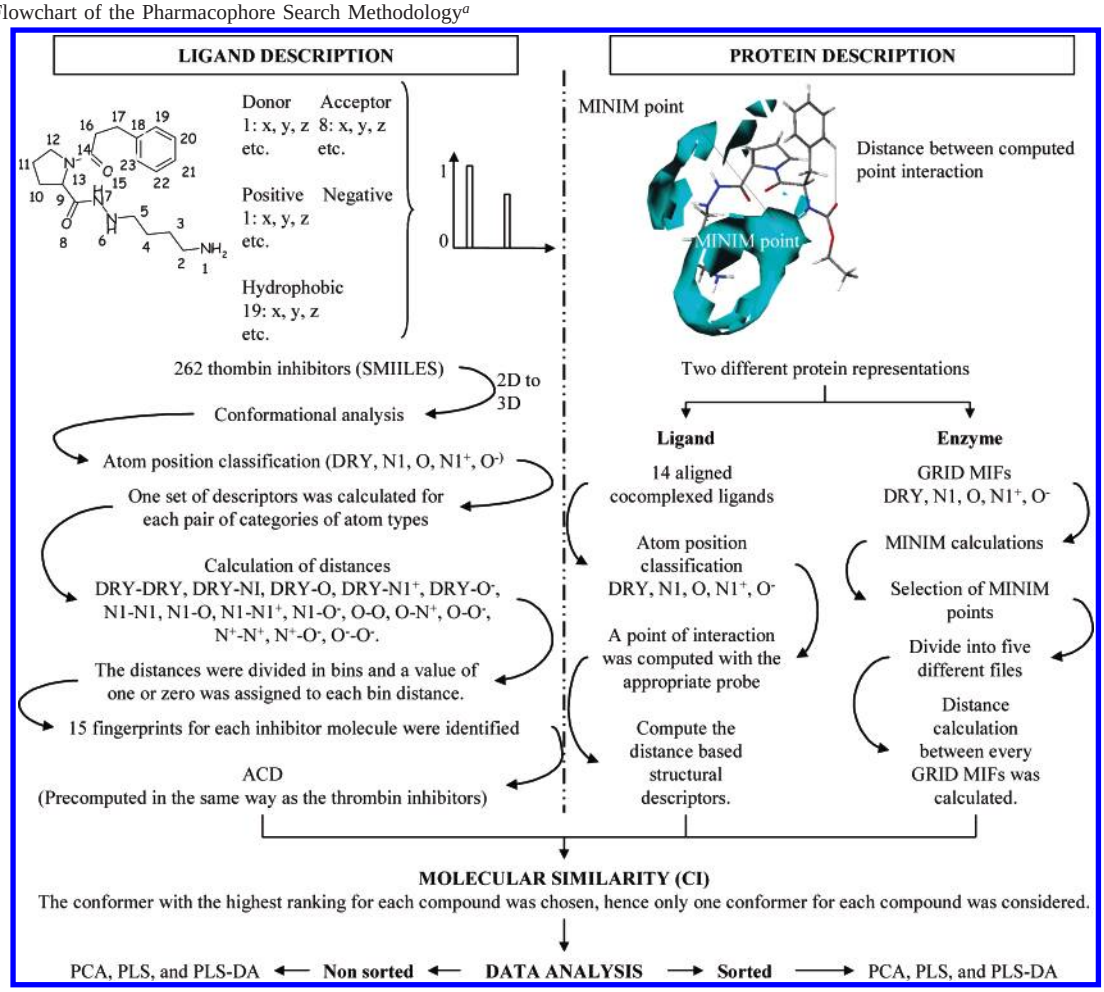
Computer Hardware and Software. All calculations were performed on a Pentium IV 2.3 GB PC in a Linux environment or on a Silicon Graphics Octane workstation (Silicon Graphics Inc., Mountain View, CA). The software utilized in the computational analyses were GRID version 21 (Molecular Discovery, http://moldiscovery.com), SYBYL 6.9 (Tripos Associates Inc., St. Louis, MO), FlexX,[38] and GOLD.[39] Chemical structures were typed in SMILES,[40] and Xconfort 6.0[41,42] was used for fast conformational analyses. The pharmacophore search and the scaffold hopping were performed using ad-hoc developed software, which was programmed using the Perl language following the method explained later in this section.

The GRID Force Field and GRID MIFs. In the program GRID,[43,44] a regular grid is built over the molecular region of interest or inside a protein cavity. Probes, i.e., functional groups that represent different physicochemical properties, are moved stepwise from grid point to grid point, and for each probe/point, the interaction energy is computed using the GRID force field. The force field is based on a Lennard-Jones potential, a distance-dependent electrostatic function, and a hydrogen-bond term. The force field has been parameterized from crystal structure data. Thus, for each grid point, the energy of interaction is obtained, which together form a MIF that represents the potential interaction of the compound/protein with a certain chemical group (probe).

The probes selected for this study were the DRY probe representing steric and hydrophobic interactions, the N1 (amide nitrogen) and $N1^+$ (sp^3 amine NH cation) probes to represent hydrogen-bond-donor groups, and the O (sp^2 carbonyl oxygen) and O^- (sp^2 phenolate oxygen) probes to represent hydrogen-bond-acceptor groups with different formal charges.

Virtual Screening. Data Set. A set of 262 structurally diverse compounds was selected from among published thrombin inhibitors.[20−32] The compounds had K_i values ranging from 45 pM to 5.37 μM and IC_{50} values ranging from 5 nM to 130 μM. The molecular weight of the

Scheme 1. Flowchart of the Pharmacophore Search Methodology[a]



[a] On the left-hand side, it is shown how the spiked ACD database is built, and on the right-hand side, the computations of the protein and the ligand are described. The two procedures are combined to perform a similarity analysis followed by a multivariate data analysis.

compounds ranged between 166 g/mol and 596 g/mol. Since the bioactive conformation was unknown, several conformations per compound were calculated to obtain a representation of the conformational diversity. The program Xconfort 6.0 was used to perform 2D-to-3D conversions and the conformational analyses using the diverse set of conformer options and selecting a maximum of 20 conformations per compound. In total, 3705 conformers were produced (within an energy range of 10 kJ/mol from the global minimum of each compound) and added to a 3D version of the Available Chemical Directory (ACD) database, which contains 74 291 different conformers of 33 127 compounds.

Thrombin Crystal Structures. The X-ray crystal structures of 14 thrombin cocomplexes with ligands were retrieved from the PDB.[45] The criteria for the selection of these structures were that the resolution had to be ≤ 2 Å and the ligand should not be covalently bound to the enzyme. The following crystal structures from the PDB database were chosen: 1a4w,[46] 1ae8,[47] 1afe,[47] 1bmm,[48] 1bmn,[48] 1c5n,[49] 1eb1,[50] 1ghy,[51] 1lhc,[52] 1lhd,[52] 1lhe,[52] 1lhf,[52] 1lhg,[52] and 7kme.[53]

METHODS

Virtual Screening. For the virtual screening methodology, a pharmacophore representation of the thrombin protein based on GRID MIFs was used. The search was performed in a 3D multiconformation version of the ACD database, which had been spiked with 262 known thrombin inhibitors (multiple conformers available per compound). Two different protein representations were used, one based on the GRID minima points from the enzyme crystal structure (points in the 3D MIFs that were surrounded by points with higher energy) and one derived from energy calculations using GRID based on the positions of the ligands in the thrombin crystal structures.

The crystal structure of thrombin complexed with hirugen and an active site inhibitor (PDB: 1a4w)[46] was used as a template for the ligand alignment. A protein backbone alignment to the template was performed for each of the other 13 crystal structures. Once all protein−ligand complexes were aligned to 1a4w, the ligands were extracted and their atom coordinates concatenated and divided into five different files depending on their atom type properties, that is, hydrophobic, hydrogen donor, hydrogen acceptor, or potential charge (positive or negative) at physiological pH. These atomic positions were used to compute the energies of interaction to represent the protein cavity.

The virtual screening methodology developed in this study is a four-step procedure (Scheme 1); first, descriptors of the protein and the inhibitors are generated; thereafter, a similarity analysis followed by a multivariate data analysis is performed. The virtual screening methodology is a completely new method using a different matrix compared to those of other known virtual screening methods. The validation steps that were used are as follows.

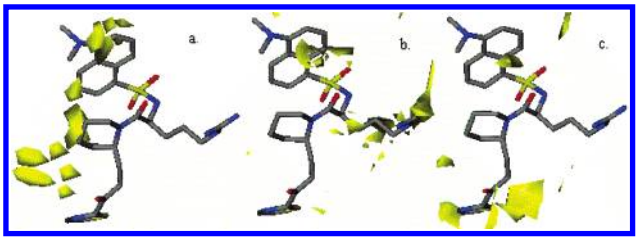


Figure 2. MIFs calculated in the thrombin crystal structure (PDB: 1a4w) using three different probes. (a) DRY (−1.1 kcal/mol), (b) N1 (−6.1 kcal/mol), and (c) O (−4.7 kcal/mol) are viewed together with the cocomplexed compound, RWJ-50215, to visualize that the interaction fields correspond to actual atomic positions.

1. Protein Description. MIFs can be computed on biomolecules such as receptors or enzymes, to suggest important interaction points between the protein and a ligand, or directly from the ligands, to suggest positions where a possible receptor would interact with the ligand.[54] In this study, MIFs were computed on the thrombin enzyme (PDB: 1a4w) using the program GRID version 21. The probes selected for this study were DRY, N1, $N1^+$, O, and O^-. The energies of interaction for the five selected probes were superimposed inside the protein cavity to indicate positions where molecular fragments or functional groups had to be placed in the receptor pocket to obtain optimal interactions with the protein. To visualize that the interaction fields corresponded to actual atomic positions, the MIFs calculated for three probes, DRY, N1, and O, were viewed together with the cocomplexed compound *N*-[4-(aminoiminomethyl)amino-1-[2-(thiazol-2-ylcarbonylethyl)-piperidin-1-ylcarbonyl[butyl]-5-dimethylaminonaphthalenesulfonamide trifluoroacetate hydrate (RWJ-50215; Figure 2). In default mode, a grid spacing of 1.0 Å was used with the grid extending 5 Å beyond the template ligand structure. Although the MIFs covered the region of interaction, the atomic position of the ligand did not exactly match the minima position found by GRID. Because of this, the energies involved in the interaction between the protein and the ligand were computed in two different ways: one based on the GRID minima points inside the thrombin enzyme active site (1a4w) and the other based on the GRID point calculation on the atomic positions from 14 different ligand crystal structures.

1a. GRID Minima Points from the Enzyme Crystal Structure. To obtain some information about the influence of water molecules on the interaction between the protein and the ligand, an analysis of the competition of the probes with water molecules was performed. Thus, the GRID program was used to calculate the interactions between the GRID polar probes (N1, $N1^+$, O, and O^-) and the 1a4w crystal structure. The GRID option LEAU was kept equal to 3, which means that GRID would compute the interaction energy taking into account the competition between probe and water. When GRID was used in this mode, the program calculated in each grid point if water or the probe itself had the strongest interaction with the protein. In the water interaction calculation, an entropy factor was introduced by the program to account for the possible formation of an interaction between water and the protein. This water interaction energy was used in the decision whether water made a stronger interaction than the probe in a specific grid point. A positive value of the energy of interaction in the GRID computation represented an unfavorable interaction compared to water, and GRID would detect regions where this occurred if the directive LEAU is set to LEAU = 3. The water molecules found in the crystal structure, which correspond to the position of the grid points where the water molecules had stronger interactions than the probes, were kept in the next GRID analysis (Figure 3). Prebound water molecules may influence the ligand such that an altered orientation of the ligand or a changed selectivity between ligands could be observed. Therefore, it was important to identify unfavorable regions for certain probes on the enzyme.

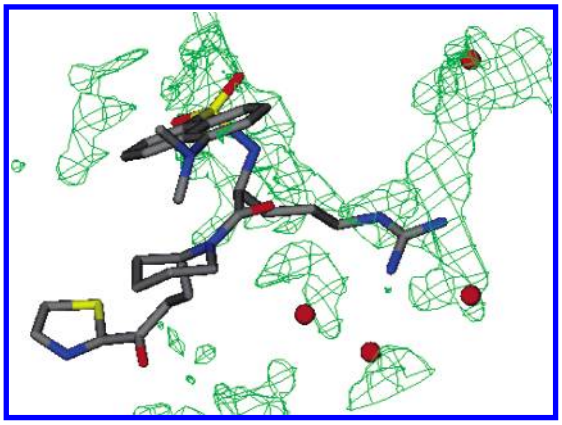


Figure 3. RWJ-50215 and the interaction energy of the O probe (−2.1 kcal/mol) and the water molecules, which had stronger interactions than the probes and, therefore, were kept in the next GRID calculation. To get some information about the influence of the water molecules on the interaction between the protein and the ligand, an analysis of the competition of the probes with the water molecules was performed. The GRID program was used to calculate the interactions between the GRID polar probes (N1, $N1^+$, O, and O^-) and the 1a4w crystal structure (LEAU = 3).

The GRID fields with the probes DRY, N1, $N1^+$, O^-, and O (LEAU = 0) were computed for 1a4w, and the preselected water molecules were included in the calculation (W2298, W2301, W2315, and W2316). MINIM calculations were performed for each probe field in order to identify the minima in the GRID interaction field. The GRID box is rectangular; therefore, several minima points were outside the region of interest, that is, distant from the ligand. A few points, which were at a maximum distance of 3 Å from the 14 ligands and had low energy, were chosen. These energy points were divided into five different files according to which probe they originated from. Then, the distance between every selected GRID point was calculated, and a set of descriptors was obtained. This procedure is similar to the GRIND methodology developed by Pastor et al.[54] where a matrix that describes the energy of interaction (every pairwise interaction type) versus the distance was computed.

1b. GRID Points from the Ligands in the Crystal Structures. The atomic positions for all 14 aligned cocomplexed ligands classified by their interaction type were submitted to GRID using the appropriate probe in order to compute point interactions (POSI option in GRID) instead of a GRID field; for example, if hydrophobic properties were analyzed, the DRY probe was used. The energies and the positions were used to compute the distance-based structural descriptors, as described above.

2. Ligand Description. The atoms constituting the thrombin inhibitors were sorted into five GRID classes depending on their hydrophobic (DRY), hydrogen-bond-donor (N1), hydrogen-bond-acceptor (O), positive ($N1^+$), or negative (O^-)

charge properties. Tripos force field atom types were used to do this straightforward classification of substrate atoms. The distances between the different atom positions, classified using the previous criteria, were then transformed into binned distances. The distance between the classified atoms were calculated, and all pairs of interactions were taken into account: DRY−DRY, DRY−N1, DRY−O, DRY−$N1^+$, DRY−O^-, N1−N1, N1−O, N1−$N1^+$, N1−O^-, O−O, O−N^+, O−O^-, N^+−N^+, N^+−O^-, O^-−O^-. Then, the distances were divided in bins and a value of one or zero was assigned to each bin distance, indicating the presence or absence of such a distance in the compound. One set of descriptors was calculated for each pair of categories of atom types as mentioned before, thus, generating 15 structural fingerprints for each inhibitor.[55]

3. Molecular Similarity. Similarities between the different ligand structures were calculated using the Carbo Index (CI; eq 1).[56]

$$R_{AB} = \frac{\int P_A P_B dv}{(\int P_A{}^2 dv)^{1/2} (\int P_B{}^2 dv)^{1/2}} \quad (1)$$

R_{AB} is determined from the structural properties P_A and P_B of the two molecules being compared. The numerator measures property overlap, whereas the denominator normalizes the similarity results. Hence, the value of the CI runs from −1 (perfect dissimilarity) passing via 0 (no similarity) to +1 (perfect similarity). Electron density, electrostatic potential, and shape have all been used as structural properties for CI computation. In this study, the distances between different atom types were used, which are defined according to the five probes used in the GRID calculation (DRY, N1, O, $N1^+$, O^-).

4. Data Analysis. To improve the discrimination between the thrombin and non-thrombin inhibitors, a statistical analysis was performed. Two training sets were chosen; the first was based on 20 thrombin inhibitors randomly chosen ("nonsorted") from the ranking list, together with 20 non-inhibitors. The compounds chosen represented a structurally diverse dataset, which was considered suitable for virtual screening. To estimate how the method could perform in the best case, another set of known thrombin inhibitors was selected; this time, the top 20 ranked thrombin inhibitors ("sorted") from the average similarity search were chosen, together with the same 20 non-thrombin inhibitors used in the first training set.

Two multivariate data analysis methods were used, principal component analysis (PCA) and partial least squares (PLS).[57,58] PCA represents the multidimensional data structure using the minimum number of variables (principal components). The direction that best describes the relative distance between the objects is the direction with maximum variance. This direction is called the first principal component (PC1). The second principal component (PC2) is orthogonal to PC1 and has, again, the maximum variance. PCA has two main applications: (a) visualization of multivariate data by scatter plots and (b) data reduction and transformation, especially if features are highly correlating. PCA uses the criterion of maximum variance of the data in a matrix **X** (independent variable), whereas PLS uses the criterion of maximum covariance between data in matrix **X** and matrix **Y** (dependent variable). PLS is a useful tool when the number of descriptors is greater than the number of compounds. The result from PLS can be viewed in score and loading plots. These plots were used to interpret the model, along with a determination of the model coefficients for the variables. Cross-validated (in five rounds) PLS analyses were used to check the predictive capability (Q^2) of the models and to determine the optimal number of components to be used in the final model. To sharpen the separation between groups of observations (type of interactions), a PLS discriminant analysis (PLS-DA) was performed. In this case, the *y* variable was an indicator for whether the compounds were a known thrombin inhibitor or not.

Scaffold Hopping. In this study, a new procedure to perform a bioisosteric substitution of one scaffold to another, so-called scaffold hopping, has also been established.[3] The starting point of the procedure was the selection of a template scaffold, which could be used during the search. When the scaffold had been chosen, the number and positions of anchor points (atomic positions where the scaffold can be chemically modified) were decided. Three parameters based on the distance between anchor points, the scaffold interaction pattern, and the shape of the scaffold were calculated and compared on the basis of a similarity search using the CI (eq 1). Thereafter, the scaffolds were sorted in classes from high to low similarity compared to the template scaffold. A scaffold with high similarity to the template scaffold was then chosen and aligned to the template scaffold in 3D space. The search template and the new scaffold were aligned by volume overlap and by comparing the root-mean-square distance (RMSD) values related to the anchor points. The new scaffold was rotated by 10° at a time, and at each angle, the RMSD of the anchor points in the template scaffold was calculated. Equal weights were given to these two parameters. The side chains connected to the anchor points in the template scaffold were then attached to the scaffold selected in the search. This new compound was energy minimized and docked into the active site of the thrombin enzyme to verify that it could adopt a similar binding pattern as that of the template.

Template Scaffolds. To find suitable scaffolds, a search in the literature for compounds cocomplexed with the thrombin enzyme was performed. Two template scaffolds (Figure 4a,b) were retrieved from a protein structure (1ae8) in the PDB, in which thrombin was cocomplexed with 1(*N*-ethoxycarbonyl-D-phenylalanyl-L-prolyl)-2(4-aminobutyl) hydrazine (Eoc-D-Phe-Pro-Abh). When choosing the scaffold, it is important to take into consideration whether it is possible to synthesize new ligands based on its structure or not, otherwise there is a risk that only new scaffolds not possible for use in the preparation of new chemical libraries will be identified. For the definition of scaffold and anchor points, see Figure 4c.

Database. Since a scaffold database with appropriate molecular structures was not available, a new database was created, composed of compounds from the summaries in The Journal of Combinatorial Chemistry from 2000 and 2001.[59,60] The scaffolds in the database were written in SMILES; thereafter, each scaffold was converted from 2D to 3D and a conformational analysis was performed using the Xconfort 6.0 program. The published scaffolds were divided into categories depending on the number of possible attachment

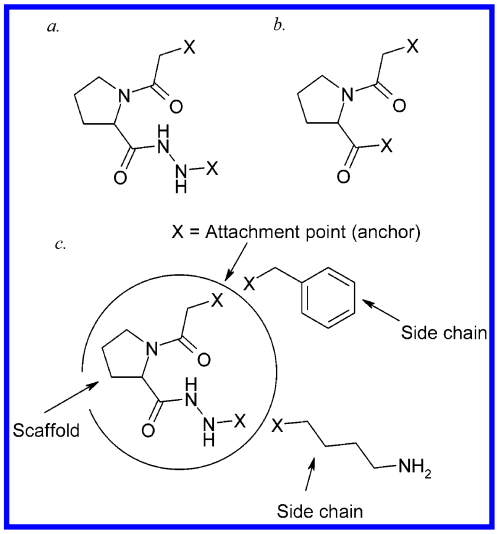


Figure 4. Template scaffolds retrieved from a protein structure (1ae8), which is thrombin cocomplexed with Eoc-D-Phe-Pro-Abh. (a) 1ae8_1. (b) 1ae8_2. (c) Explanation of what is considered a scaffold, a side chain, and an attachment point (anchor).

points (anchors) for which some chemical derivatization had been reported. The anchor points were chosen such that they were useful in the synthesis of new compounds (Figure 4). This database consisted of 89 different scaffolds, with two anchor points, and the outcome from the conformational analysis gave, in total, 4055 different structures.

The scaffolds in the database were sorted into different "baskets" according to the number of anchor points, anchors 1−5. For each scaffold, three different types of information were computed (Figure 5):

1. The distance between anchor points was computed.

2. GRIND-like parameters[54] were calculated to describe the scaffold interaction pattern. The main difference between the method used in our study and the GRIND description previously described in the literature, where the distances were computed from grid points, was that, in our analysis, the distance bins had the maximum energy of interaction for the distance between the anchor points and the GRID MIF energies. In this study, each scaffold was analyzed using all five probes (DRY, N1, O, O^-, and $N1^+$), and consequently, descriptor vectors were obtained in quantities that were 5 times the number of anchor points; that is, two anchor points yielded 10 vectors to describe the scaffold interaction pattern.

3. A new function was developed on the basis of the frequency distances between the attachment point in the scaffold and the positive interaction GRID points computed for the N1 probe in order to consider the shape of the scaffolds. As in the GRIND-like descriptors, one shape function was obtained for each anchor point.

Molecular Similarity. Similarities between the compounds were calculated using the CI (eq 1).[56] As stated above, the following distances were calculated and used as a measurement for similarity: the distance between the anchor points (point at the scaffold where the side chains from the template should be added), the distance between the anchor points and GRID fields, and the distance between the anchor points and the surface measured as described above. These three parameters were weighted equally, giving an overall similarity index that could be compared between the template scaffold and the scaffolds in the database.

Docking. Once the scaffolds from the database were selected and the new ligand structures had been built, they were docked into the active site of thrombin. Three different docking programs were used to validate the method: AutoGroup, which is a GRID-based docking program, and two other docking programs, FlexX and GOLD, which are not GRID-based.

One thrombin crystal structure (PDB: 1ae8) was selected as the target for the docking simulations. The modifications

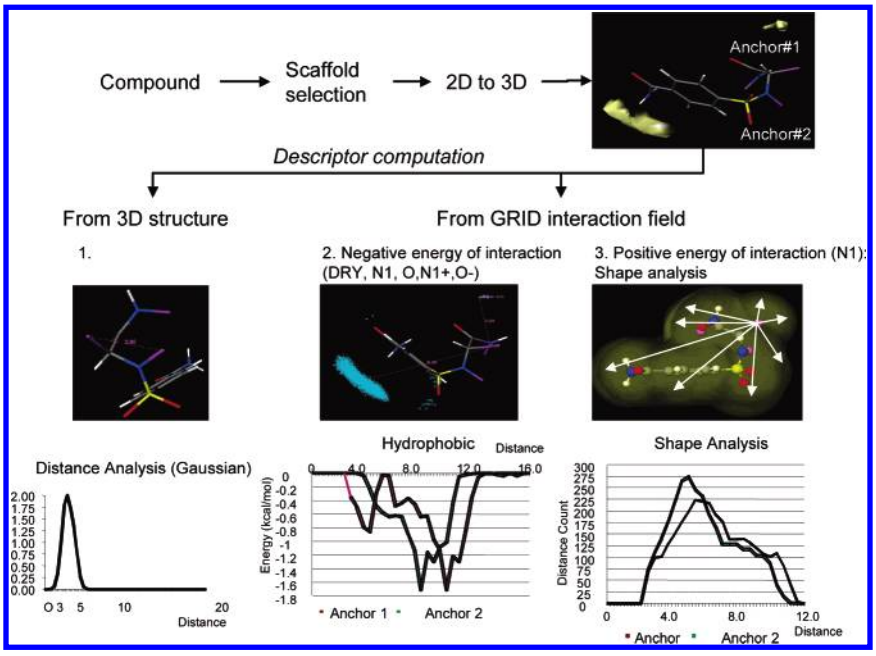


Figure 5. Diagram illustrating that the scaffolds were selected and, for each scaffold, three different types of information were computed: (1) The distance between anchor points was computed from the 3D structure. (2) The negative energy of interaction for five probes (DRY, N1, O, O^-, and $N1^+$), describing the scaffold interaction pattern, and (3) the positive energy of interaction for the N1 probe, describing the shape of the scaffold, were computed from GRID interaction fields.

that were done to the protein were to remove the ligand from the binding site and, in a first step, the water molecules as well. It has been established that thrombin inhibitors interact via water molecules with the protein mainly by hydrogen-bonding interactions (e.g., with Glu192).[30] To obtain some information about the influence of water molecules on the interaction between the protein and the ligand, an analysis on the competition of the probes with water molecules was performed as described above. To do this, the GRID option LEAU was kept equal to 3, as described in the virtual screening section. The GRID program was used to calculate the interactions between the GRID polar probes (N1, $N1^+$, O, and O^-) and the 1ae8 crystal structure. The water molecules found in the crystal structure, which correspond to the position of the grid points where the water molecules had stronger interactions than those of the probes, were kept in the next GRID analysis. As a result, the docking experiments were performed having the most relevant water molecules present in the active site (W2487, W2505, W2509, W2510, W2526, W2539, W2559, W2610, W2620, and W2622). These two different approaches, with and without water molecules, gave different results depending on the docking program used.

AutoGroup. In AutoGroup, the PDB file had to be modified before being imported in Greater,[61] which converted the PDB format to the format of the input files (kout files) required to run the docking procedure. The box size was defined to extend 5 Å from the crystallographic ligand position. The compounds were built and imported into Xconfort 6.0 for energy minimization and conformational analysis. The diverse set of conformers was used, and a maximum of 40 conformations per compound were saved (within an energy range of 10 kJ/mol from the global minimum of each compound). All these conformers were used in the docking, and the conformers with binding patterns similar to that of the crystal ligand and with energies of binding, as computed by GRID, close to the energy of the crystal ligand were kept for visual inspection.

GOLD, Version 1.1. The compounds were built and energy minimized using the MMFF94s force field and MMFF94 charges.[62] The active site cavity was defined with a radius of 20 Å from the β-carbon atom in His57 in the protein. The genetic algorithm implemented in GOLD was used to optimize the orientation of the ligand in the active site. During this optimization process, the ligand was considered flexible whereas the active site of the enzyme was considered rigid. For each ligand, 30 dockings were allowed, with an early termination if the RMSDs were within 1.5 Å for the top three solutions. Default values were used for all other parameters.

FlexX, Version 1.11.1. The same structure files obtained from the conformational analysis for the GOLD dockings were also used in the FlexX dockings. However, formal charges were assigned within FlexX. The active site cavity was defined with a radius of 10.1 Å from residue Trp215. Again, the ligand was considered flexible whereas the active site of the enzyme was considered rigid during this optimization process. For each ligand, 30 dockings were allowed.

RESULTS AND DISCUSSION

Virtual Screening. In this study, two different approaches to compute the mode of interaction between thrombin and the ligand have been investigated: one based on the GRID minima points inside the thrombin enzyme active site (1a4w) and the other one based on the GRID point calculation on the atomic positions from 14 different ligand crystal structures. Depending on the pairwise comparison between the different interaction types (DRY−DRY, DRY−N1, DRY−O, DRY−$N1^+$, DRY−O^-, N1−N1, N1−O, N1−$N1^+$, N1−O^-, O−O, O−N^+, O−O^-, N^+−N^+, N^+−O^-, and O^-−O^-), 15 different similarity indexes were computed. The structures in the 3D version of the ACD database were ranked from high to low similarity, taking into account all the different interactions. From this list, the conformer with the highest ranking for each compound was chosen; hence, only one conformer for each compound was considered. To find the impact of the different properties in the final discrimination between the thrombin inhibitors and the noninhibitors, a multivariate data analysis was performed either considering (PLS) or not considering (PCA) the classification of the compounds in the training set (Figure 6). To sharpen the separation between groups of observations (type of interactions), a PLS-DA was performed. The PLS analysis shows that there is a good separation of the inhibitors and noninhibitors already, before taking the classification into account. This procedure was repeated twice using two methods to extract the training set, either by random selection of 20 thrombin inhibitors and 20 non-thrombin inhibitors, the so-called "nonsorted" training set, or by selection of the 20 highest ranked thrombin inhibitors and 20 non-thrombin inhibitors, the so-called "sorted" training set.

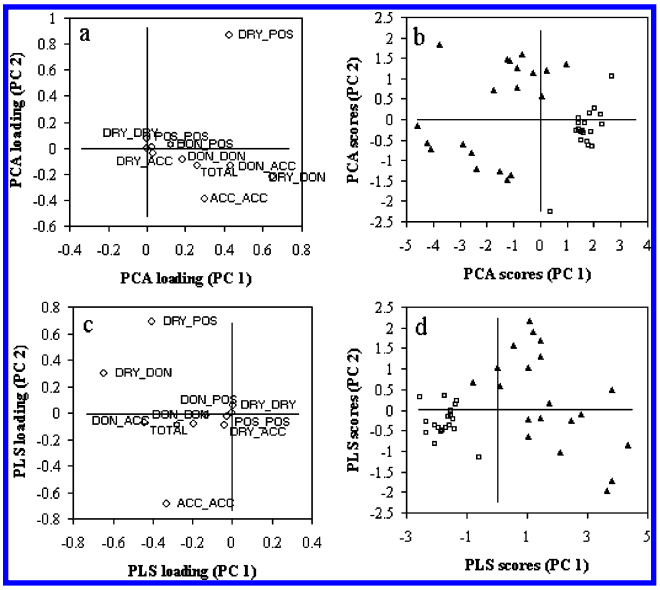


Figure 6. PCA and PLS models for the method derived from the 14 aligned ligands, "nonsorted". (a) PCA loading plot, showing that DRY_POS, DRY_DON, and ACC_ACC are the interactions that best distinguish between thrombin inhibitors and non-thrombin inhibitors. (b) PCA score plot: square, thrombin inhibitors; triangle, non-thrombin inhibitors. (c) PLS loading plot, showing that DRY_POS, DRY_DON, and ACC_ACC are the interactions that best distinguish between thrombin inhibitors and non-thrombin inhibitors. (d) PLS score plot: square, thrombin inhibitors; triangle, non-thrombin inhibitors.

Predictions based on GRID points from the ligands in the crystal structure gave better results than the calculations based on GRID minima points from the enzyme crystal structure (Figure 7). One explanation could be that the atomic positions from the 14 crystal ligands are well-defined. The calculation

Table 1. PCA/PLS Coefficients for the PCA/PLS Analysis and the Enrichment of the Database for the Four Different Methods

		PCA		PLS	known thrombin inhibitors found by searching 5% of the database
		component 1	component 2	(R^2)	
ligand	nonsorted	46%	68%	0.52	80%
	sorted	65%	79%	0.85	27%
enzyme	nonsorted	49%	67%	0.23	55%
	sorted	55%	74%	0.85	30%

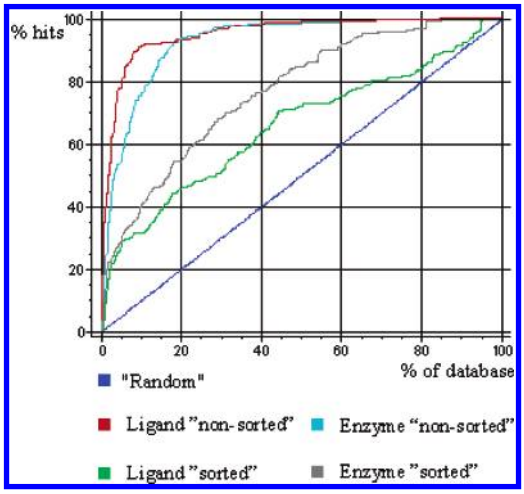


Figure 7. Enrichment factor curve showing how the different methods focus the ACD database. Ligand "nonsorted" (red), enzyme "nonsorted" (cyan), enzyme sorted (grey), ligand sorted (yellow), and random (blue) curves are shown. The random curve shows how it would look if we found one compound for every 0.4% of the database.

from the enzyme generates more minima points, thus introducing more noise. The difference in the ability to focus the library using the "sorted" or "nonsorted" training set was smaller for the method based on the enzyme than that based on the ligands, which can be clearly seen in Figure 7.

This can be due to the fact that, during the search, the enzyme method covered a larger volume in space because it was based on GRID MIFs instead of the exact atomic position as in the ligands case. As a result, a set of more structurally diverse ligands was found when the enzyme-based method was used. So, although the method was based on sorted compounds, they were not as structurally similar as those for the method based on ligand structures. This means that when using the top 20 candidates obtained from the enzyme method in the PCA/PLS analysis, they were not as similar to each other as in the case for the top 20 from the ligand method and will, therefore, give a better representation of the ACD database. This indicates that the enzyme method was not as sensitive for compound selection in the PCA/PLS analysis as the ligand method.

The R^2 value for the ligand "sorted" method ($R^2 = 0.85$) was higher than the R^2 value for the "nonsorted" method ($R^2 = 0.52$), but the method that focused the ACD to a higher extent was the method based on the "nonsorted" ligands (Table 1). This indicated that when the PLS/PCA analysis was performed on the "sorted" compounds, the method became too restricted; the compounds were not structurally diverse enough to represent the compounds in the ACD.

The external prediction of the ligand "nonsorted" method was good; the overall correct prediction was 85% (Table 2). The percentage of false positive and false negative was 5

Table 2. External Prediction for the Ligand "Nonsorted" Method

	predicted	
	thrombin inhibitor	non-thrombin inhibitor
experimental		
thrombin inhibitor	40%	10%
non-thrombin inhibitor	5%	45%

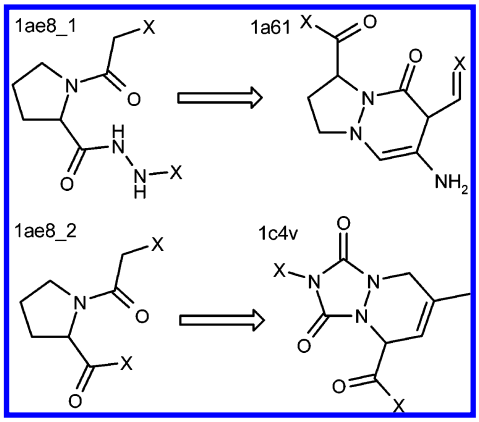


Figure 8. Two new scaffolds chosen from the database search.

and 10%, respectively. The ligand method based on the "nonsorted" thrombin inhibitors managed to find 80% of the known thrombin inhibitors in the ACD by only searching 5% of the database; hence, a 15-fold concentration of the library was achieved (Table 1).

Such enrichment of a database of that size is of great importance in pharmaceutical lead discovery. This makes it possible, very early in the lead discovery process, to search through a virtual library for compounds that could be used in initial screening campaigns to develop new therapeutic agents.

Scaffold Hopping. Two scaffolds originating from the same compound (Eoc-D-Phe-Pro-Abh), 1ae8_1 and 1ae8_2 (Figure 4), were used as templates for the search in the scaffold hopping. From the database search, two new scaffolds were chosen (Figure 8).

These two were chosen partly because they could be used for the design of new compounds in 3D space using the side chains from the template scaffold and partly because structural information was available. Both scaffolds are building blocks from compounds cocrystallized with the thrombin enzyme (PDB: 1a61 and PDB: 1c4v). These two scaffolds (1a61, 1c4v) were used for validation of the method. The search was done using two anchor points, which gave two possibilities to add the side chains to the new scaffold. The first was the alignment of the arbitrarily named side chain 1 of the template with anchor 1 of the scaffold in the database and side chain 2 of the template with the arbitrarily named anchor 2 of the scaffold (align0) and vice versa, template side chain 1 with scaffold anchor 2 and

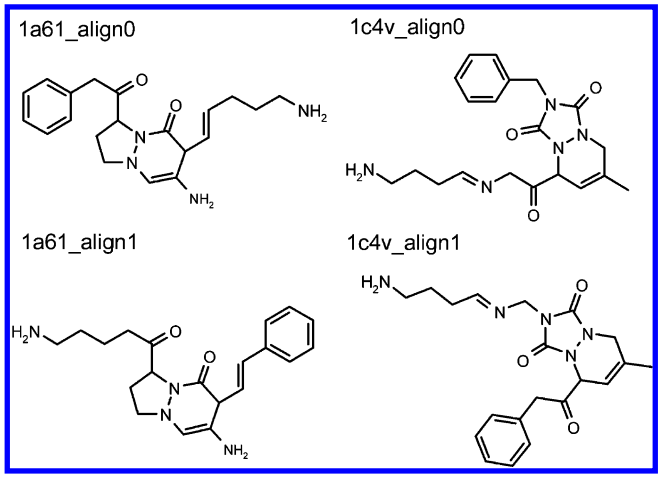


Figure 9. Four compounds that were docked. The solutions that had similar binding patterns to the cocomplexed compound, i.e., within 2.5 Å, were all ranked among the top 10 solutions in the docking programs FlexX and GOLD of, in total, 30 solutions.

Table 3. Docking Result for the Four New Compounds and the Cocrystallized Compound

	GOLD (water)		FlexX (no water)		AutoGroup (water)	
	ranking	RMSD[a]	ranking	RMSD[a]	kcal/mol	RMSD[a]
1a61_0	7	1.31	2	1.01	−7.60	1.96
1a61_1	2	0.962	2	1.24	−5.00	1.59
1c4v_0	7	1.36	*b*	*b*	−9.09	2.57
1c4v_1	8	1.66	2	1.71	−9.28	0.822
1ae8_key	1	0.684	1	1.10	−13.20	0.724

[a] RMSD value for the side chains compared to the ligand in the crystal (PDB: 1ae8). [b] No solution was docked in the same way as the cocrystallized compound.

template side chain 2 with scaffold anchor 1 (align1). It is of outmost importance to consider the chemistry when doing the alignment; otherwise, there is a risk of ending up with compounds difficult or even impossible to synthesize. To determine that the docking programs worked properly, the cocomplexed compound (Eoc-D-Phe-Pro-Abh) was docked in addition to the new compounds. These docking results showed that the three docking programs AutoGroup, FlexX, and GOLD docked the compound in the same way that it was positioned in the crystal structure with a RMSD < 1.0 Å.[63]

The performance of the three docking programs was overall equal. One difference that could be seen was that GOLD performed best if the water molecules were included, whereas the opposite was true for FlexX. The program AutoGroup docked all four compounds in a way similar to that of the cocrystallized compound in both cases, that is, water included and water excluded.

The four compounds (Figure 9) were docked, and all docking solutions that showed a binding pattern similar to that of the cocomplexed compound, that is, within 2.5 Å[63] (Table 3), were found among the top 10 ranked solutions in the docking programs FlexX and GOLD. This indicates that the scoring functions for the programs were consistent with the docking performance. GOLD managed to dock all four compounds, whereas FlexX only managed to dock three of the four compounds in the same way as the (bioactive conformation) cocomplexed ligand.

In AutoGroup, the solutions were not ranked; instead, the interaction energies were given for each solution. The

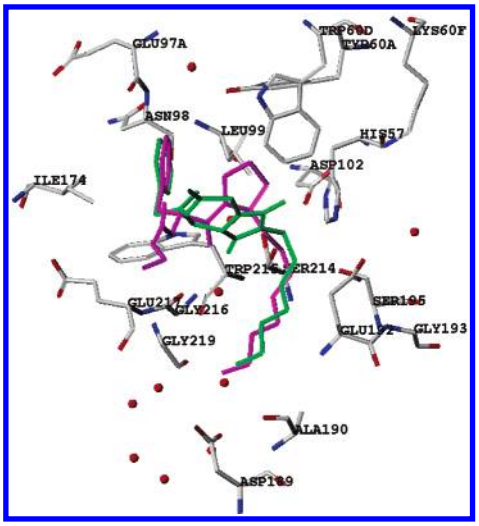


Figure 10. Thrombin crystal (PDB: 1ae8). Crystal Eoc-D-Phe-Pro-Abh: magenta. Docked 1c4v_align1: green. Docking performed with AutoGroup (with water).

interaction energies for the compounds with binding patterns similar to that of the cocomplexed compound were between −5.00 kcal/mol and −13.20 kcal/mol, which indicated that there was a high probability that the compounds could interact with the enzyme. The docking result for the new compounds evidently showed that they have the same binding pattern as the cocomplexed compound. One example is the new compound 1c4v_align1, with a RMSD value of 0.82 Å (Figure 10).

The other new compounds had RMSD values between 0.96 and 2.57 Å; the main reason for the high RMSD value was that the phenyl ring in the new compound was somewhat twisted compared to that in the cocrystallized compound. That twist does not affect the hydrophobic interaction with the side chains Trp60A, Trp215, Leu99, and Ile174. The interaction with Asp189 is the same for all compounds irrespective of the RMSD value. The interactions that can be seen between the template scaffold and the thrombin enzyme can also be found between the new scaffold and the thrombin enzyme, that is, interaction with Gly216.

From these results, it is clear that this method could be a helpful tool for medicinal chemists when designing new compounds in the lead identification phase to receive information about new potential scaffolds. Changing the core structure of a compound can be one important approach to overcome, for example, selectivity problems. The potential utility of this technique resides in the scaffold hopping itself, which can be an effective idea generator when medicinal chemists are interested in changing the scaffold in compounds because of unwanted side effects or absorption, distribution, metabolism, and excretion problems.

CONCLUSIONS

This study describes a method that is both practical and effective for library enrichment using distance descriptors between atoms in compounds with known binding affinities to thrombin and GRID-computed energies in combination with PLS/PCA analysis. In an early stage of the drug development process, fast methods for identifying interesting lead structures with reasonable accuracy may be more efficient than methods with high accuracy, which are often

extremely time-consuming. This technique should be used as a quick screening method, decreasing the number of compounds for more extensive studies.

This study also shows that it is possible to use GRID MIFs as descriptors to substitute scaffolds and still retain a binding pattern similar to that of the thrombin enzyme active site. This method can be a helpful tool for medicinal chemists when designing new compounds in the lead identification phase.

Supporting Information Available: Figures of the docking of a known thrombin inhibitor (Eoc-D-Phe-Pro-Abh) and two new possible thrombin inhibitors, 1a61 and 1c4v, using three different programs, AutoGroup, FlexX, and Gold. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Willett, P. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983−996.

(2) Frimurer, T. M.; Bywater, R.; Naerum, L.; Lauritsen, L. N.; Brunak, S. Improving the odds in discriminating "drug-like" from "non drug-like" compounds. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1315−1324.

(3) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. "Scaffold-Hopping" by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angew. Chem., Int. Ed.* **1999**, *38*, 2894−2896.

(4) Sadowski, J.; Kubinyi, H. A scoring scheme for discriminating between drugs and nondrugs. *J. Med. Chem.* **1998**, *41*, 3325−3329.

(5) Forino, M.; Jung, D.; Easton, J. B.; Houghton, P. J.; Pellecchia, M. Virtual docking approaches to protein kinase B inhibition. *J. Med. Chem.* **2005**, *48*, 2278−2281.

(6) Evers, A.; Klabunde, T. Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha1A adrenergic receptor. *J. Med. Chem.* **2005**, *48*, 1088−1097.

(7) Mozziconacci, J. C.; Arnoult, E.; Bernard, P.; Do, Q. T.; Marot, C. Optimization and validation of a docking-scoring protocol; application to virtual screening for COX-2 inhibitors. *J. Med. Chem.* **2005**, *48*, 1055−1068.

(8) Christmann-Franck, S.; Bertrand, H. O.; Goupil-Lamy, A.; der Garabedian, P. A.; Mauffret, O. Structure-based virtual screening: an application to human topoisomerase II alpha. *J. Med. Chem.* **2004**, *47*, 6840−6853.

(9) Krovat, E. M.; Fruhwirth, K. H.; Langer, T. Pharmacophore identification, in silico screening, and virtual library design for inhibitors of the human factor Xa. *J. Chem. Inf. Model.* **2005**, *45*, 146−159.

(10) Martin, Y. C. 3D database searching in drug design. *J. Med. Chem.* **1992**, *35*, 2145−2154.

(11) Bender, A.; Mussa, H. Y.; Gill, G. S.; Glen, R. C. Molecular surface point environments for virtual screening and the elucidation of binding patterns (MOLPRINT 3D). *J. Med. Chem.* **2004**, *47*, 6569−6583.

(12) Honma, T.; Hayashi, K.; Aoyama, T.; Hashimoto, N.; Machida, T. Structure-based generation of a new class of potent Cdk4 inhibitors: new de novo design strategy and library design. *J. Med. Chem.* **2001**, *44*, 4615−4627.

(13) Naerum, L.; Norskov-Lauritsen, L.; Olesen, P. H. Scaffold hopping and optimization towards libraries of glycogen synthase kinase-3 inhibitors. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1525−1528.

(14) Schmidt, J. M.; Mercure, J.; Tremblay, G. B.; Page, M.; Kalbakji, A. De novo design, synthesis, and evaluation of novel nonsteroidal phenanthrene ligands for the estrogen receptor. *J. Med. Chem.* **2003**, *46*, 1408−1418.

(15) Lloyd, D. G.; Buenemann, C. L.; Todorov, N. P.; Manallack, D. T.; Dean, P. M. Scaffold hopping in de novo design. Ligand generation in the absence of receptor information. *J. Med. Chem.* **2004**, *47*, 493−496.

(16) Lewell, X. Q.; Jones, A. C.; Bruce, C. L.; Harper, G.; Jones, M. M. Drug rings database with web interface. A tool for identifying alternative chemical rings in lead discovery programs. *J. Med. Chem.* **2003**, *46*, 3257−3274.

(17) Jenkins, J. L.; Glick, M.; Davies, J. W. A 3D Similarity Method for Scaffold Hopping from Known Drugs or Natural Ligands to New Chemotypes. *J. Med. Chem.* **2004**, *47*, 6144−6159.

(18) Cramer, R. D.; Jilek, R. J.; Guessregen, S.; Clark, S. J.; Wendt, B. "Lead hopping". Validation of topomer similarity as a superior predictor of similar biological activities. *J. Med. Chem.* **2004**, *47*, 6777−6791.

(19) Rush, I. T.; Grant, J. A.; Mosyak, L.; Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein−protein interaction. *J. Med. Chem.* **2005**, *48*, 1489−1495.

(20) Bohm, M.; St. Rzebecher, 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.

(21) Bone, R.; Lu, T.; 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.

(22) Burgey, C. S.; Robinson, K. A.; Lyle, T. A.; Sanderson, P. E.; Lewis, S. D. Metabolism-directed optimization of 3-aminopyrazinone acetamide thrombin inhibitors. Development of an orally bioavailable series containing P1 and P3 pyridines. *J. Med. Chem.* **2003**, *46*, 461−473.

(23) Carroll, A. R.; Pierens, G. K.; Fechner, G.; De Almeida Leone, P.; Ngo, A. Dysinosin A: a novel inhibitor of Factor VIIa and thrombin from a new genus and species of Australian sponge of the family Dysideidae. *J. Am. Chem. Soc.* **2002**, *124*, 13340−13341.

(24) Hauel, N. H.; Nar, H.; Priepke, H.; Ries, U.; Stassen, J. M. Structure-based design of novel potent nonpeptide thrombin inhibitors. *J. Med. Chem.* **2002**, *45*, 1757−1766.

(25) Jhoti, H.; Cleasby, A.; Reid, S.; Thomas, P. J.; Weir, M. Crystal structures of thrombin complexed to a novel series of synthetic inhibitors containing a 5,5-trans-lactone template. *Biochemistry* **1999**, *38*, 7969−7977.

(26) Linusson, A.; Gottfries, J.; Olsson, T.; Ornskov, E.; Folestad, S. Statistical molecular design, parallel synthesis, and biological evaluation of a library of thrombin inhibitors. *J. Med. Chem.* **2001**, *44*, 3424−3439.

(27) Narasimhan, L. S.; Rubin, J. R.; Holland, D. R.; Plummer, J. S.; Rapundalo, S. T. Structural basis of the thrombin selectivity of a ligand that contains the constrained arginine mimic (2S)-2-amino-(3S)-3-(1-carbamimidoyl- piperidin-3-yl)-propanoic acid at P1. *J. Med. Chem.* **2000**, *43*, 361−368.

(28) Noteberg, D.; Branalt, J.; Kvarnstrom, I.; Linschoten, M.; Musil, D. New proline mimetics: synthesis of thrombin inhibitors incorporating cyclopentane- and cyclopentenedicarboxylic acid templates in the P2 position. Binding conformation investigated by X-ray crystallography. *J. Med. Chem.* **2000**, *43*, 1705−1713.

(29) Pierce, A. C.; Jorgensen, W. L. Estimation of binding affinities for selective thrombin inhibitors via Monte Carlo simulations. *J. Med. Chem.* **2001**, *44*, 1043−1050.

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

(31) Supuran, C. T.; Scozzafava, A.; Briganti, F.; Clare, B. W. Protease inhibitors: synthesis and QSAR study of novel classes of nonbasic thrombin inhibitors incorporating sulfonylguanidine and O-methylsulfonylisourea moieties at P1. *J. Med. Chem.* **2000**, *43*, 1793−1806.

(32) Zhang, H. C.; Derian, C. K.; Andrade-Gordon, P.; Hoekstra, W. J.; McComsey, D. F. Discovery and optimization of a novel series of thrombin receptor (par-1) antagonists: potent, selective peptide mimetics based on indole and indazole templates. *J. Med. Chem.* **2001**, *44*, 1021−1024.

(33) Walker, C. P.; Royston, D. Thrombin generation and its inhibition: a review of the scientific basis and mechanism of action of anticoagulant therapies. *Br. J. Anaesth.* **2002**, *88*, 848−863.

(34) Fenton, J. W., II; Ofosu, F. A.; Moon, D. G.; Maraganore, J. M. Thrombin structure and function: why thrombin is the primary target for antithrombotics. *Blood Coagulation Fibrinolysis* **1991**, *2*, 69−75.

(35) Bode, W.; Mayr, I.; Baumann, U.; Huber, R.; Stone, S. R. The refined 1.9 A crystal structure of human alpha-thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J.* **1989**, *8*, 3467−3475.

(36) Chirgadze, N. Y.; Sall, D. J.; Briggs, S. L.; Clawson, D. K.; Zhang, M. The crystal structures of human alpha-thrombin complexed with active site- directed diamino benzo[*b*]thiophene derivatives: A binding mode for a structurally novel class of inhibitors. *Protein Sci.* **2000**, *9*, 29−36.

(37) Martin, P. D.; Malkowski, M. G.; DiMaio, J.; Konishi, Y.; Ni, F. Bovine thrombin complexed with an uncleavable analog of residues 7−19 of fibrinogen A alpha: geometry of the catalytic triad and interactions of the P1′, P2′, and P3′ substrate residues. *Biochemistry* **1996**, *35*, 13030−13039.

(38) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.

(39) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.

(40) Weininger, D. SMILES, a chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31−36.

(41) Pearlman, R. S.; Balducci, R. Confort: A Novel Algorithm For Conformational Analysis. *National Meeting of the American Chemical Society*, New Orleans, 1998.
(42) *Confort*; Tripos Associates Inc.: St. Louis, MO. http://www.tripos.com.
(43) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.
(44) 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.
(45) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(46) Matthews, J. H.; Krishnan, R.; Costanzo, M. J.; Maryanoff, B. E.; Tulinsky, A. Crystal structures of thrombin with thiazole-containing inhibitors: probes of the S1′ binding site. *Biophys. J.* **1996**, *71*, 2830−2839.
(47) De Simone, G.; Balliano, G.; Milla, P.; Gallina, C.; Giordano, C. Human alpha-thrombin inhibition by the highly selective compounds *N*-ethoxycarbonyl-D-Phe-Pro-alpha-azaLys *p*-nitrophenyl ester and N-carbobenzoxy-Pro-alpha-azaLys *p*-nitrophenyl ester: A kinetic, thermodynamic and X-ray crystallographic study. *J. Mol. Biol.* **1997**, *269*, 558−569.
(48) Malley, M. F.; Tabernero, L.; Chang, C. Y.; Ohringer, S. L.; Roberts, D. G. M. Crystallographic determination of the structures of human alpha-thrombin complexed with BMS-186282 and BMS-189090. *Protein Sci.* **1996**, *5*, 221−228.
(49) Katz, B. A.; Mackman, R.; Luong, C.; Radika, K.; Martelli, A. Structural basis for selectivity of a small molecule, S1-binding, submicromolar inhibitor of urokinase-type plasminogen activator. *Chem. Biol.* **2000**, *7*, 299−312.
(50) Friedrich, R.; Steinmetzer, T.; Bode, W. Complex structure of human thrombin with *N*-methyl-arginine inhibitor. To be published.
(51) Katz, B. A.; Elrod, K.; Luong, C.; Rice, M. J.; Mackman, R. L. A novel serine protease inhibition motif involving a multi-centered short hydrogen bonding network at the active site. *J. Mol. Biol.* **2001**, *307*, 1451−1486.
(52) Weber, P. C.; Lee, S. L.; Lewandowski, F. A.; Schadt, M. C.; Chang, C. H. Kinetic and crystallographic studies of thrombin with Ac-(D)-Phe-Pro- boroArg-OH and its lysine, amidine, homolysine, and ornithine analogs. *Biochemistry* **1995**, *34*, 3750−3757.
(53) Mochalkin, I.; Tulinsky, A. Structures of thrombin retro-inhibited with SEL2711 and SEL2770 as they relate to factor Xa binding. *Acta Crystallogr., Sect. D* **1999**, *55*, 785−793.
(54) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. GRid-INdependent Descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. *J. Med. Chem.* **2000**, *43*, 3233−3243.
(55) Zamora, I.; Afzelius, L.; Cruciani, G. Predicting drug metabolism: a site of metabolism prediction tool applied to the cytochrome P450 2C9. *J. Med. Chem.* **2003**, *46*, 2313−2324.
(56) Carbo, R.; Leyda, L.; Arnau, M. How Similar is a Molecule to Another? An electron density measure of the similarity between two compounds. *Int. J. Quantum Chem.* **1980**, *17*, 1185−1189.
(57) Stahle, L.; Wold, S. Multivariate data analysis and experimental design in biomedical research. *Prog. Med. Chem.* **1988**, *25*, 291−338.
(58) Leach, A. R. *Molecular Modelling Principles and Applications*, 2nd ed.; Longman: New Jersey, 2001; pp 497−499, 706−708.
(59) Dolle, R. E. Comprehensive survey of combinatorial library synthesis: 2001. *J. Comb. Chem.* **2002**, *4*, 369−418.
(60) Dolle, R. E. Comprehensive survey of combinatorial library synthesis: 2000. *J. Comb. Chem.* **2001**, *3*, 477−517.
(61) *GRID*, version 21; Molecular Discovery Ltd.: Pinner, Middlesex, U. K. http://moldiscovery.com.
(62) Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parametrization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.
(63) Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein−ligand docking. *Proteins* **1999**, *37*, 228−241.

CI049626P