Journal of Computer-Aided Molecular Design, **14:** 411–425, 2000.
KLUWER/ESCOM

Computer based screening of compound databases: 1. Preselection of benzamidine-based thrombin inhibitors

Thomas Fox* & Eric E.J. Haaksma
Boehringer Ingelheim Pharma KG, Department of Chemical Research/Structural Research, D-88397 Biberach an der Riß, Germany

Received 9 September 1999; Accepted 22 December 1999

Key words: active site mapping, flexible query, GRID, structure based screening, 3D database, thrombin

Summary

We present a computational protocol which uses the known three-dimensional structure of a target enzyme to identify possible ligands from databases of compounds with low molecular weight. This is accomplished by first mapping the essential interactions in the binding site with the program GRID. The resulting regions of favorable interaction between target and ligand are translated into a database query, and with UNITY a flexible 3D database search is performed. The feasibility of this approach is calibrated with thrombin as the target. Our results show that the resulting hit lists are enriched with thrombin inhibitors compared to the total database.

Introduction

The first step in the drug discovery process is the identification of a biological target whose function should be modified in order to correct a pathophysiological state. Once this has been accomplished, so-called lead molecules have to be identified. These are the starting points for large-scale synthesis programs during which activity, bioavailability, pharmacokinetics etc. are optimized. As at the beginning of a drug discovery process it is not clear whether a compound can be successfully optimized, usually several lead compounds are worked on in parallel. The discovery of structurally different lead structures is therefore a critical success factor in the first phase of a drug discovery project.

The search for lead structures is traditionally performed by biological screening of in-house compound collections, a procedure which can be extremely fast if a so-called high throughput system can be set up. The speeds of these screening programs make a rational compound selection superfluous, as in a short period of time all compounds can be tested. However, these screens are strictly limited to physically available compounds. Therefore, computational methods have been developed that can search for any compound which is electronically available, e.g. compounds which were once in the compound pool but now have decomposed or have been used up, or compounds which are only available in external compound collections [1–5]. Recently, even huge 'virtual' databases have been introduced, which contain up to billions of possible molecules which do not even exist in the real world and only are synthesized after they have been identified as promising compounds [6, 7].

Moreover, not every assay can be automated for an HTS system. Therefore, another possible application of computational screening approaches are low-throughput screenings. Here, only a few hundreds of compounds can be tested per week, and obviously a procedure which helps to select compounds that should preferentially be tested would be of great value.

In order to perform such a computational screening, either knowledge about compounds that interact with a given pharmacological target, or the three-dimensional (3D) structure of this target has to be available. Such a structure or that of a target-ligand complex provides information on the possible and/or essential points of interaction between target and lig-

*To whom correspondence should be addressed. E-mail: thomas.fox@bc.boehringer-ingelheim.com

and. Therefore, close inspection of the binding site and the possible interactions in it should significantly improve the chances to find new or better ligands. One possibility to use the information provided by the 3D structure is to identify interaction points in the binding site and use this knowledge as a starting point of a 3D search in proprietary and non-proprietary databases [8–10]. An alternative could involve an automated docking procedure which tries to fit potential ligands into the active site of the enzyme. Then a scoring function which has been designed to give a reliable estimate of the free energy of binding is evaluated, thus identifying interesting compounds for further analysis [11–14]. However, despite recent impressive advances, these methods still need considerable computational resources and certainly would benefit from a preselection based on pharmacophore features.

Our approach is to use the program GRID [15, 16] to map the active site of our target and the spacial distribution of the possible interactions. After identifying the energetically most important interactions, a pharmacophore model is defined in which the types of interactions and their geometrical relationships are directly read out from the GRID results. This pharmacophore model is used to start a conformationally flexible, 3D database query on corporate and/or external databases with the program UNITY [17].

This strategy was tested with the serine protease thrombin as the pharmacological target. For this enzyme, a large amount of experimental data is available which should make it possible to calibrate the method and to assess its viability and success. The goal of our procedure is therefore twofold. First, we want to extract known thrombin inhibitors from a database of compounds that were tested for thrombin activity. This would correspond to the situation where compounds have to be selected for screening. Second, we want to find new lead structures different from our in-house database, which is accomplished by searching an external database for molecules satisfying the derived pharmacophore model.

The test system: thrombin

Thrombin is a trypsin-like serine protease which – besides other important physiological tasks [18] – plays a central role in the blood coagulation cascade [19]. Therefore, inhibition of thrombin is a promising approach towards the prevention and treatment of thrombosis. Accordingly, significant effort has been put into the development of orally active thrombin inhibitors, and recently the first such inhibitors have been reported to be in clinical trials [20–23].

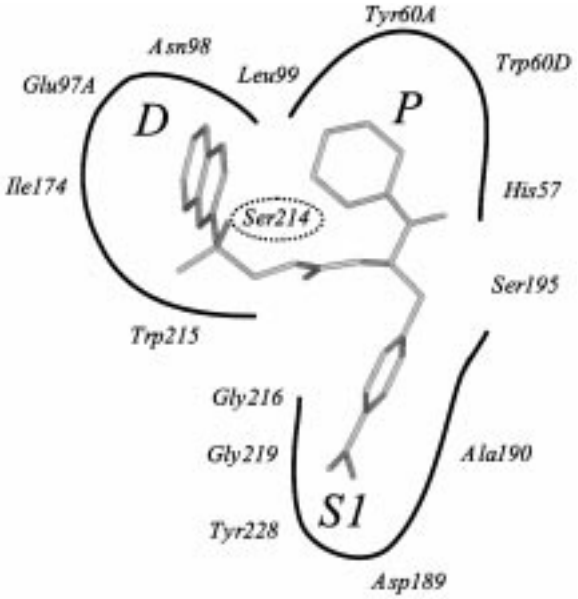

Figure 1. Schematic drawing of the unprimed side of the active site of thrombin with its S1, P and D pockets. Several amino acids that define the active site and/or are important for binding of ligands are highlighted. Also shown is NAPAP as an example inhibitor. Ser214 is located at the back portion of the active site.

The X-ray crystallographic structures of thrombin with various specific active-site directed inhibitors have been solved [19, 24]. Figure 1 shows a schematic drawing of the active site of thrombin, together with a number of amino acids implicated in the binding of ligands. On the unprimed side of the catalytic triad (Ser195, His57, Asp102; the numbering of the thrombin residues corresponds to that in chymotrypsinogen, see [19]), the deep S1 pocket is located. With Asp189 at its bottom, which is positioned to make a salt bridge with basic P1 residues of substrates, it is an important determinant for the specificity of thrombin. A characteristic feature of thrombin is the '60-insertion loop', which is a primary cause for the narrower specificity of thrombin compared to other serine proteases. Together with His57, Ser214, and L99, this loop forms the S2 site, which is ideally suited to bind medium sized nonpolar residues. A second hydrophobic pocket, the so-called 'aryl-binding site', is lined by the residues Ile174, Trp215, and the segment 97–99. Although these two subsites form one single hydrophobic surface, they usually are distinguished by the nomenclature P ('proximal') and D

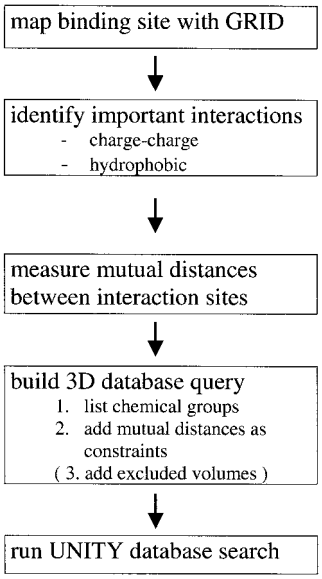

Figure 2. Workflow chart of the important steps of our procedure.

('distal') pocket. As a consequence of this structure, most peptidic or non-peptidic active site inhibitors of thrombin share a similar binding mode: A basic group reaches into the S1 pocket, additionally two mostly hydrophobic groups bind into the adjacent P and D sites.

From the point of structure based drug discovery one fact is especially fortunate: comparison of various thrombin crystal structures shows that thrombin is not much affected by the binding of the different ligands; its overall and detailed structure is essentially maintained. Therefore, one can assume that flexibility of the receptor sidechains will not play a dominant role and does not have to be taken into account during the drug design efforts.

Another reason for choosing thrombin as a system to develop and test our virtual screening approach is the large amount of data accumulated at Boehringer Ingelheim on this target: we have available the three-dimensional structures of 30 thrombin inhibitor complexes, solved crystallographically in our X-ray laboratory [24]; additionally a large number of compounds have been tested in vitro for inhibition of thrombin. Thus, we have a large pool of experimental data at our hands which should allow a rigorous test of our approach.

Procedure

Figure 2 shows a workflow chart of the individual steps of our procedure. Starting from an analysis of the binding site with the GRID method, we use these results to derive a three-dimensional database query in UNITY. A detailed description of these tasks is given in the following sections.

Mapping of the active site

The first step of our proposed approach is the identification of possible interaction sites in the receptor. For this task, we used the program GRID, which is a computational procedure to determine energetically favorable interaction sites on small molecules or proteins of known three-dimensional structure [15, 25–28]. This is achieved by calculating the energies of interaction between a small chemical group (the 'probe') and a second molecule (the 'target') on the points of a regularly spaced grid in the region of interest. The interaction energy includes contributions from van der Waals and electrostatic interactions, as well as hydrogen bond interactions. Regions of negative energy then indicate energetically favorable binding sites for the particular probe; they may be displayed as three-dimensional contour surfaces.

The different GRID probes are characterized by size, charge, polarizability, and hydrogen-bond donor/acceptor capabilities. The values for these descriptors are chosen such as to reproduce the experimentally available data on binding strengths and directionality [15].

In our experience, the exact nature of the probe type has only a minor influence on the shape and position of the resulting contour values. For example, the different oxygen types O (carbonyl oxygen), OS (sulfone/sulfoxide), ON (nitro group), while differing in details and in their absolute energy values, usually yield very similar areas of favorable interaction. Therefore, from the multitude of probe types supplied with GRID, we chose a few representatives to sample the different interaction types, typically these include the hydrophobic probe DRY, the alkyl hydroxy group OH, an alcoholate probe (O−), sp^2 oxygen (e.g., O::, O=), an ammonium probe (N2+, NM3), and an organic halogen (BR, CL). In addition to these single atom probes, probes corresponding to multi-atom entities such as water, amides or amidine are employed.

The resulting GRID energy maps were displayed as contour surfaces with INSIGHT97 [29]. The energy thresholds were chosen in such a way that the individual contours formed 'blobs' with diameters of at least 1–1.5 Å. These values guarantee a high degree of specificity (i.e., not the whole active site is mapped as 'favorable'), and on the other hand the problem of too small regions of very favorable interaction is avoided: These are only rarely occupied by a ligand atom and, therefore, are not relevant for our purpose.

Derivation of the database query

Searching in databases of chemical compounds, different classes of search criteria or descriptors can be used. For the purpose of this article, we want to define 2D queries as those which only depend on a two-dimensional representation of the molecule, i.e., we then essentially perform a substructure search and only look for the number of bonds between certain atom types. A 3D query additionally takes into account the relationship between these patterns in the three-dimensional space, e.g., how many Å is pattern 1 away from pattern 2.

Therefore, in the derivation of a 2D/3D query to search a compound database, the following questions have to be addressed:

1. Which are the key molecular features that are responsible for the favorable interaction between ligand and receptor (relevant for 2D/3D)?
2. How should these key features be described in the database system one uses (relevant for 2D/3D)?
3. Which are the appropriate geometric relationships between these features and the corresponding geometric tolerances (relevant for 3D only)?

It is generally assumed [10, 30] and in accordance with our experience that for a selective database query, the pharmacophore model should consist of at least three interaction points. Consistently, it seems reasonable to assume that a database query should include this amount of information. If only two interaction points are employed, the database query is not specific enough and will return many (false-positive) hits. Conversely, with more than three features one risks that the query is too focused and the resulting hit list is too small.

To choose the type of interaction, one has to consider the three major forces that determine molecular recognition and/or selective binding: electrostatics, hydrophobic interactions, and hydrogen bonding. To assess which of them is important for a given problem, different routes are possible: One can compare different known ligands and extract common features present in all or many of them. This approach has been used extensively in the past [31] and is the only possibility if structural information about the receptor is missing. However, at least a couple of ligands have to be known, and one also risks that the novelty is limited. With the 3D structure of the target at hand, a direct determination of regions for a certain type of interaction is possible. This can be done either by hand, or guided by interaction databases (e.g., IsoStar [32]), or by LUDI-like procedures [33–36], which for each interaction type determine favorable interaction points within the binding region. One problem with these statistical approaches – especially if one wishes to derive an exact and focused database query – is to decide where exactly the most favorable interaction region is.

To overcome this problem, we decided to use the results of GRID calculations to identify the pertinent interactions at the receptor site. This offers the distinct advantage that not only a qualitative, but also a quantitative measure of favorable interaction is available, i.e., the question where is the best region for an interaction of a certain type can be easily addressed. A potential difficulty, however, is that one might have to choose between different competing interactions in the same spatial region.

A special note deserves the case of hydrogen bonding. Here one has to keep in mind that the net energy gain is always the result of a delicate balance between the energetics of a hydrogen bond in solution and an alternative hydrogen bond formed between ligand and receptor. While in aqueous solution there is a high probability that an ideal hydrogen bond can be formed, this is less likely in the active site of the enzyme. Therefore, the introduction of a hydrogen bond partner in a ligand may or may not improve its binding affinity [37–39]. As a consequence, as a first approximation we assume that hydrogen bonds can be ignored if other interactions are available.

The next point, after derivation of the query model, is the question of how these features can be described in the context of the database system that we use. For the searches in the databases we employed the program UNITY [17]. It allows both 2D and 3D queries by means of the Sybyl Line Notation (SLN) [40], a SMILES-like system to describe chemical structures and database queries. 2D queries are specified by chemical substructures that have to be matched in the resulting 'hits'. 3D queries are defined using 'fea-

tures' and 'constraints'. Features are points, vectors or planes derived from 2D patterns that have to be present in the desired hits; with constraints the spatial relationships between features are defined.

For charged centers, a simple way would be searching for atoms with a non-zero formal charge. However, this would assume that the formal charges are encoded in the database entries. Moreover, compounds which are stored in the database in their neutral forms could be ionized at physiological pH.

To circumvent these problems in our queries, we decided to simply list the possible groups that bear a non-zero charge and thus may favorably interact with a charge center at the receptor. For a positive charge, these include primary, secondary, and tertiary aliphatic amines, as well as substituted or unsubstituted amidines and guanidines. For the amines, the nitrogen atom is considered to be the center of interaction, for the amidines and guanidines the carbon atom connecting the nitrogens is used. While for the latter one could imagine more elaborate schemes for where to place the charge center (e.g., depending on the substitution pattern), we believe that our simple model is sufficient, especially in view of the geometric tolerances we use (see below). Correspondingly, for a negatively charged center, we would use the central carbon, sulfur, or phosphorus atoms of terminal CO_2, SO_3, or PO_3 groups, or the centroid of a tetrazole ring. We note that the exact location of a charged center is only important for a 3D query.

An immediate advantage of such a scheme is that it can easily be expanded to special proprietary groups, or groups that are only partly ionized, such as, e.g., imidazoles or sulfonamides with electron-withdrawing substituents.

The hydrophobic feature is much more difficult to describe, as there is a continuum ranging from slightly to highly hydrophobic, depending on the chemical environment (substitution pattern), and, moreover, hydrophobicity is much less an inherent atomic or group characteristic as, e.g., the charge state.

Many queries in the past identified hydrophobic regions by the centroid of a molecular fragment, which often is a phenyl or another aromatic ring [41, 42]. In UNITY, the predefined 'hydrophobic feature' for example is defined as the centroid of any 5- or 6-membered ring. However, this definition encompasses also non-hydrophobic rings such as tetrazole; moreover, aliphatic chains, which also may show considerable hydrophobic character, are completely neglected. The other extreme, assuming that every carbon atom is hydrophobic, is certainly not correct either, as a carbon next to, e.g., a charged group probably would not interact well with a hydrophobic pocket in the target enzyme.

A possible solution of this problem would be a case-to-case decision whether an atom or group is hydrophobic or not. Various schemes have been suggested how to quantify the hydrophobicity of atoms or groups and their contribution to binding affinity [43, 44], and their implementation in the context of database searching has been investigated [43]. However, these schemes tend to be complicated (and thus relatively slow), and usually they require specific software.

Therefore, in the spirit of the approach used for charged groups, and compatible with the UNITY software, we define hydrophobic features in putative ligands by a list. A hydrophobic center is either a halogen atom (F, Br, Cl, I), or a CH group which

1. is not bound to a polar or charged (or chargeable) group such as –OH, $–NH_2$, –COOH, $–PO_3H$, etc.,
2. is not bound to such a carbon atom,
3. is not doubly bonded to an O, S, N.

Using a CH group instead of just a single carbon atom alleviates the problem of buried carbon atoms as in, e.g., t-butyl groups, which do not contribute to the hydrophobic interaction energy. This is, in some sense, a straightforward way to increase the probability that the hydrophobic feature is at the surface of the ligand.

Once the 2D patterns (i.e., atoms or functional groups) have been determined, appropriate geometric constraints and corresponding tolerances have to be added. Within UNITY, several types of geometric constraints are available. Either the geometric relationships between features (with, e.g., distances or angles), or the desired location of features in 3D space, or the location of the whole molecule with respect to the target can be specified. The type we found the most useful and straightforward to apply are mutual distances between the pharmacophoric features derived from the GRID mappings.

To this end, we identify the local minima of the interesting GRID maps (or the geometric average if several minima are close, i.e., within one contour surface) and simply measure the corresponding distances between these minima. In the past, benchmark queries in the literature employed tolerances in interatomic distances as low as $\pm$ 0.4 Å or even less [8, 9, 45, 46]. This is in contrast to other studies that show that pharmacophore identification with tolerances of 1–2 Å on distances leads to reasonable results [47]. Similar

values have been suggested by Greene et al. [43] based on an analysis of the variation of interaction energies with changes in the geometry.

While it is tempting to view a query with low tolerances as accurate and capable of selectively retrieving molecules with high affinity, in our opinion this is inappropriate for various reasons:

1. Using crystallographic structures of proteins to derive the database query, the accuracy of the atomic positions in the crystal is a lower bound on the proper tolerances.
2. The receptor usually can and will respond to the interactions with the ligand: side chains may move to allow for a better interaction.
3. The favorable interaction is not limited to one single point.
4. The calculation of the interactions using GRID is approximate.
5. The way we define our interaction groups (see above), the minimum of the interaction as derived by GRID need not coincide with the atomic positions (as in the case of hydrophobic groups) or the center of interaction may not be the center on the ligand (e.g. amidines).
6. A practical aspect is that in flexible searches the computational effort increases with lower tolerances.

With the 3D structure of a target to derive the pharmacophore model, additionally the size and shape of the receptor site are known. This can be used to further refine the database query by imposing additional constraints on the extent and shape of possible ligands, thus reducing the number of false positives [48]. In UNITY, this can be done with so-called 'excluded volumes' – spheres which define regions in space where no ligand atoms are allowed. The locations of excluded volumes can be obtained from the X-ray structure by manually selecting atoms of the protein target which line the active site.

The flexible query

Early 3D searching systems represented structures by single conformations that were generated using programs such as CONCORD [49] or CORINA [50], which provide a rapid means of creating low-energy conformations for a large portion of typical compound databases. However, the majority of pharmaceutically important chemicals is flexible (e.g., the average number of rotatable bonds in the Boehringer Ingelheim compound pool is 9). If only one conformation for each structure is stored in the database, rigid 3D queries will not find the vast majority of possible hits: The use of such a rigid search is likely to fail to identify flexible structures for which a possible conformation that can adopt the query pharmacophore is different from the stored one [51].

Generally, there are two approaches to solve this problem. The first is to generate not just one, but a large collection of conformers for each compound and store them in the database for subsequent searching [52]. However, the number of conformations required to span the available conformational space may become extremely large and still may not include the conformation that fulfills the given pharmacophore model.

An alternative approach, exploited by UNITY, is to investigate the flexibility of the molecule at search time, using the 'directed tweak' algorithm [9] to adjust the rotatable bonds of the molecules at search time to produce a conformation which matches the given 3D query as closely as possible. A check for van der Waals interactions then selects chemically reasonable conformations. Using analytical derivatives, this method is reasonably fast; however, it is only applicable when the query contains only distance constraints. If additionally excluded volumes are used in the database query, UNITY first generates a conformation satisfying the distance constraints via directed tweak, then this conformation is tested for violation of the volume constraints.

Analysis of the UNITY hit lists

The ideal query would retrieve all active compounds from a database but not any inactive compound. This suggests the use of two characteristic numbers that cover these two quality aspects of a resulting hit list. The first of these numbers is the 'yield' which measures the fraction of active compounds in the database which are found by the query. To further quantify the quality of a hit list, we calculate the 'purity' as the ratio between the active compounds in a hit list and the total number of hits. The two values vary, independently, from 0 to 100%. E.g., if all the hits were active compounds, the purity would equal 100, if no active compounds were in the hit list, the purity would equal 0. No attempt was made to assess the quality of the resulting hit geometries.

Table 1. Results of the GRID calculations: Minimum interaction energy for a grid encompassing the whole protein, E(all), and for a grid around the active site only, E(box), calculated for a collection of 19 different probe types. Also given are the energies to yield the 'ideal' contour blobs as described in the text, E(cont), and the locations of the resulting contour surfaces

Probe	Description	Charge	E(all)	E(box)	E(cont)	Location[a]
DRY	hydrophobic		−2.6	−2.6	−1.5	Active site pockets
N-:	anionic tetrazole N	−	−11.2	−8.0	−6.0	h2o sites
N1#	sp NH with one hydrogen	+	−19.5	−19.5	−15.0	Channel, h2o sites
N1	neutral flat NH, e.g. amide		−10.6	−9.3	−7.5	Channel, h2o sites, amidine
N2:	sp^3 NH_2 with lone pair		−18.2	−15.9	−12.0	Channel, h2o sites, amidine
N2+	sp^3 amine NH_2 cation	+	−26.4	−23.1	−17.0	Channel, h2o sites, amidine
NM3	trimethyl-ammonium cation	+	−15.7	−15.7	−9.0	Amidine, bottom channel
O-	sp^2 phenolate oxygen	−	−15.6	−12.1	−8.0	h2o sites
O1	alkyl hydroxy OH group		−14.2	−12.9	−9.0	h2o sites
O::	sp^2 carboxy oxygen atom		−15.5	−12.2	−8.0	h2o sites
O=	O of sulfate/sulfonamide		−13.0	−9.8	−6.0	h2o sites
PO4	PO_4 phosphate dianion	2−	−23.9	−19.1	−15.0	1 specific h2o site
PO4H	PO_4H phosphate anion	−	−20.9	−20.9	−15.0	Same h2o site as PO4
F	organic fluorine atom		−6.2	−3.8	−2.0	h2o site
CL	organic chlorine atom		−10.2	−7.8	−6.0	Channel, h2o sites, D pocket, amidine
OH2	water		−17.7	−15.1	−12.0	h2o sites
amide			−15.3	−12.6	−9.0	h2o sites, near P pocket, amidine
amidine		+	−38.6	−38.6	−22.0	Channel, amidine
carboxy		−	−21.3	−18.5	−10.0	h2o sites

[a]Channel: water channel extending from the bottom of the S1 pocket; h2o site: various sites outside or at the periphery of the active site cleft, in the crystal structures often occupied by water molecules; amidine: bottom of S1 pocket, in the crystal structures occupied by amidine group.

Computational details

GRID calculations

The starting point of our GRID calculations was the crystal structure of thrombin co-crystallized with inhibitor. Of the various in-house structures we chose the one with the highest resolution (1.6 Å); the backbone RMS deviation to the thrombin X-ray structure from the PDB (entry 1dwd) amounts to 0.53 Å. With the program GRIN (part of the GRID program package), hydrogen atoms were added where appropriate, the crystallographic water molecules were not taken into account. For the GRID calculations, the protein was considered rigid.

For the calculations considering the active site only, the size of the GRID box was chosen to encompass all thrombin amino acid residues that have heavy atoms closer than 5 Å to the co-crystallized inhibitor. As Figure 3 shows, the whole active site with the specificity pocket and the D and P binding sites is covered by this $21.5 \times 28.5 \times 15.5$ Å^3 box. For the calculations with the whole protein, the box had a size of $61 \times 58 \times 58$ Å^3. The distance between grid points was set to 0.5 Å, all other GRID input parameters retained their default values. The local minima of the GRID energy maps were obtained with the program MINIM [16].

The test databases

To test the feasibility of our approach, four databases were used in this study. The first two contain 30 in-house thrombin inhibitors for which X-ray structures have been determined. In the database XRAY_EXP, the experimental conformations of the ligands are stored, whereas the conformations in the database XRAY_CORINA were generated by CORINA [53].

The database THR_ACT contains 3107 thrombin ligands that have been tested for thrombin inhibition, against either human or bovine thrombin. In the few cases where both activities have been measured, the resulting IC_{50} values are identical within the error margins. Therefore, we are confident that values from these different sources are comparable and can be used together. The compounds showed IC_{50} values from

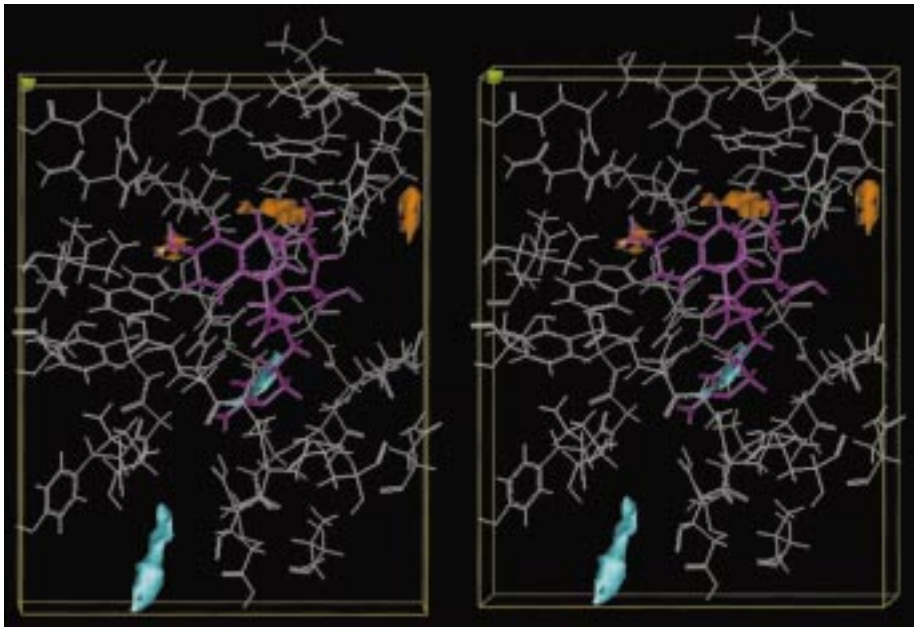

Figure 3. Stereoview of the X-ray structure of the active site of thrombin (gray), with the inhibitor argatroban (MQPA) bound (magenta). The yellow lines show the extent of the GRID array. Orange: GRID contours of the DRY probe at an energy level of -1.5 kcal mol^{-1}; blue: GRID contours of the NM3 probe at -9 kcal mol^{-1}; green: GRID contours of the PO4H probe at -15 kcal mol^{-1}.

100 μmol l^{-1} to the low nanomolar range. Compounds that were not active in the assay were assigned an activity of 100 μmol l^{-1}.

All molecules in these three databases contain a benzamidine moiety for the interaction with the S1 pocket of thrombin.

In contrast to the first three databases, which either contain known thrombin inhibitors or compounds that had been synthesized with the known features of thrombin in mind, the Available Chemicals Directory (ACD) is an unbiased collection of mainly relatively small molecules (94% of the entries have a molecular weight < 500) which usually are classified as 'non-druglike' [54]. The ACD was chosen to see if our approach is able to select reasonable compounds from this collection. From a 2D version of the ACD, which contained 200167 entries, we removed all inorganic compounds and compounds with more than 20 rotatable bonds or a molecular weight above 750. Then the 3D coordinates were generated with CORINA [53]. The resulting UNITY database contains 174421 UNITY registration entries, corresponding to 167659 unique structures.

In all cases, the databases were generated with the UNITY utility *dbimport*, followed by standard 2D and macro fingerprint generation with *dbmkscreen*.

UNITY queries

The queries were performed with the command *dbsearch* from the command line on an SGI PowerChallenge (R8000 CPU) or SGI Origin (R10000 CPU). The maximal search time per database entry was set to 60 or 600 CPU s. In Table 4 the different queries employing various levels of complexity which are used in this study are summarized.

Results

GRID receptor mapping

In Table 1 we have compiled, for a few selected GRID probes, the maximum (negative) interaction energy between the probe and the target, both for the calculation with the whole thrombin as target, and for a calculation that only encompasses the box around the active site.

One recognizes immediately that the minimum energies span a broad range, from -2.6 kcal mol^{-1} for the DRY probe to -39 kcal mol^{-1} for amidine. Generally, charged groups tend to have higher absolute interaction energies than neutral probes (with the exception of the rather delocalized N-: and NM3 probes, this value is below -20 kcal mol^{-1}), and multi-atom

Table 2. Distances R1–R5 (in Å) used in the database query (see Figure 4)

distance (Å)	R1	R2	R3	R4	R5
	9.8	9.3	6.5	7.5	8.6

probes give stronger interactions than single-atom probes. Uncharged probes without hydrogen bonding capabilities, e.g., DRY and C3, populate the opposite side of this spectrum. The obvious result of these findings is that the interaction energy scales of different probes are not directly comparable, i.e., for the derivation of a pharmacophore model one cannot simply pick those probes and interaction areas with the most negative energy values.

As a working hypothesis, we assume that for an interaction which is important enough to be included in a pharmacophore model, its interaction energy is close or equal to the maximum value for the whole protein. Therefore, in a next step we compare the maximum interaction energies within the active site box with the corresponding values for the whole protein. Indeed, only for a few probes the maximum is within the active site box: these are the hydrophobic DRY probe, the cationic N1#, NM3, amidine and the anionic PO4H probes. For all other probes, the maximum energy in the vicinity of the active site is – sometimes significantly – smaller. This indicates that these probes, despite a high absolute energy, are not really important for selective binding of a ligand within the active site.

To investigate these findings further, we generated contour plots at the energy thresholds given in Table 1. This revealed that most probes have good interactions not inside the active site cleft, but at various sites in the box where crystallographic water molecules are found. A very important example is a channel extending from the bottom of the S1 pocket: here all probes with hydrogen bonding abilities show a more or less pronounced interaction. Indeed, this behaviour can be used to determine crystal water sites in X-ray structures [55].

In Figure 3 we illustrate these results by plotting the contour surfaces for the DRY, NM3, and PO4H probes. The different probes yield contours at different regions in the active site, making it rather straightforward to derive a pharmacophore model.

Of the probes that have their global minimum within the active site box, PO4H has a single interaction area away from the active site at a position which in many of our X-ray structures is occupied by a crystallographic water molecule (see Figure 3). The other probes – all positively charged nitrogen probes – show significant attraction for the water channel mentioned above, but also for the bottom of the S1 pocket, about where the amidinium groups of the known inhibitors are bound. For NM3, this area is actually the global minimum.

With the DRY probe, we see three interaction sites (see Figure 3), two of which correspond to the well-known P and D pockets in the thrombin active site. The third is found on the solvent-facing side of Trp60D, pointing towards the far less intensively explored S1′ pocket of thrombin.

These results suggest a three-point model which consists of one positively charged center (at the bottom of the S1 pocket) and two hydrophobic sites hp1 and hp2 (in the P and D pockets of thrombin). We note that this model also very nicely corresponds to the well-known features of the thrombin active site [19].

The exact locations of the hydrophobic sites hp1 and hp2 were obtained as follows: At first the local minima of the DRY contours were calculated and those minima were selected which could be associated with one of the 2 hydrophobic blobs inside the active site of thrombin. Then for each of the two blobs the coordinates of the corresponding local minima were averaged to yield one interaction point for each of the two hydrophobic sites. (Slightly different locations are obtained if just the lowest minimum in each hydrophobic area is considered.)

Owing to the fact that our test databases only contain benzamidine compounds, as the positive center the amidine carbon atom of benzamidine (C0) was used. To take into account the form of the S1 pocket and, consequently, the ‘upright’ orientation of the benzamidine, we additionally introduced the para carbon atom (C4) in the phenyl ring as a fourth point in the pharmacophore and modelled the benzamidine into the active site to minimize possible sterical clashes. The coordinates for C0 were taken from the global minimum of the NM3 probe, this point is almost the same as the local minima of probes with contours at the bottom of the S1 pocket, and also very close to the crystallographic position of the amidine carbon atom. The C4 point was derived from the modelled structure. Then the distances R1 to R5 between these four sites were calculated and used as constraints for the 3D query. The final pharmacophore model is shown in Figure 4. Additionally, the torsion angle C0-C4-hp1-hp2 was determined.

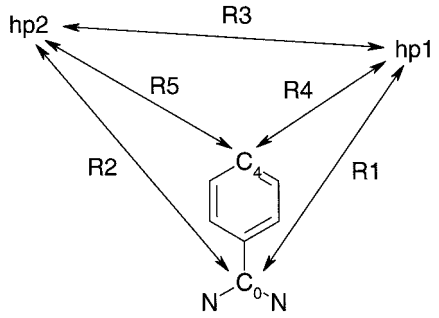

Figure 4. Pharmacophore model as derived from the GRID results and used in the UNITY database search; hp = hydrophobic feature.

Table 3. Minimum and maximum value, average and standard deviation σ of distances R1–R5 (in Å) deduced from the X-ray structures of the inhibitors in thrombin. All: all 30 X-ray structures; All-2: longest and shortest distance excluded

		R1	R2	R3	R4	R5
All	Min	9.0	9.3	5.1	4.8	8.1
	Max	11.6	11.2	7.5	8.0	9.8
	Avg	10.0	10.2	6.5	6.5	8.8
	σ	0.65	0.37	0.66	0.82	0.53
All-2	Min	9.3	9.8	5.8	5.6	8.2
	Max	10.7	10.6	7.3	7.6	9.6
	Avg	10.0	10.1	6.6	6.5	8.8
	σ	0.41	0.18	0.46	0.56	0.42

Validation of the pharmacophore model

To test the plausibility of the deduced database query, we first analyzed the 30 in-house X-ray structures of thrombin with co-crystallized ligand. We identified the moieties in the ligand which interact with the hydrophobic pockets of thrombin, and based on these results we measured the distances R1 to R5 as defined in Figure 4.

However, it is not a priori clear which part of a hydrophobic group is actually responsible for the interaction: is it e.g. the whole phenyl ring, or just one or two atoms in it?

To answer this question, we used the GRID energy values obtained with the DRY probe in the P and D pockets of thrombin and identified the parts of the ligands that lie within the -1.5 kcal mol^{-1} contour surfaces. In most cases, one or two carbon atoms, CH_2 or CH_3 groups were found inside one of these contour surfaces, in a few cases the bond between two carbon atoms was covered.

Following the identification of interaction atoms, the distances R1 to R5 were measured in the experimental conformations. In Table 3, the measured minimum, maximum and average distances R1 to R5 between these carbon atoms (or the centroid of two carbon atoms, if applicable) and the benzamidine anchor group are listed. Some of the ligands do not have any atoms within these contour blobs and were omitted from this analysis.

Generally, the experimental distances correspond reasonably well to the values deduced from the GRID calculation. For all five distances, the standard deviation σ is significantly below 1 Å, and using the average value $\pm 2\sigma$ as the tolerance, with few exceptions all the X-ray conformations would be covered. Omitting the longest and the shortest distance in each of the sets to exclude possible outliers hardly influences the results. The largest change is in R2 where the range of distances decreases from 1.9 to 1.0 Å, whereas all average distances remain unaltered. This also holds when the two longest and shortest distances are removed.

In addition to the assurance that the GRID-deduced 3D search corresponds well with the experimental inhibitor binding modes, these numbers also justify our choice of a rather high tolerance of ± 1–1.5 Å in the UNITY queries: With the exception of R4, where a value of 1.5 Å is used, the tolerance is set to 1.0 Å. For the torsion constraint, a value of $0 \pm 20°$ was used.

Validation of the database query

From these results we derived a hierarchy of database queries, from a rather simple 2D query which only searches for the benzamidine moiety (bza) over 3D queries employing distance constraints to those who additionally use excluded volume constraints (see Table 4).

These queries were then applied to the test database XRAY_EXP. As this database contains all 30 ligands in their X-ray conformation, an 'ideal' query should retrieve all compounds whose conformations match our pharmacophore model. Indeed, as Table 5 shows, with simple queries without excluded volume constraints all molecules are returned. Analyzing the hits, one finds that for the hydrophobic features not necessarily the atoms are selected which cover the hydrophobic spot in the crystal structure, but that *any* two hydrophobic atoms matching the distance criteria are returned. This is a consequence of our simplistic way of defining hydrophobicity as an on/off characteristic – UNITY does not allow a ranking into more or

Table 4. Overview over the various UNITY database queries used in this study

Query name	Type	Description[a]
bza	2D	benzamidine
bza_3pt	3D	benzamidine (C0), 2 hydrophobic points, R1, R2, R3
bza_4pt	3D	benzamidine (C0, C4), 2 hydrophobic points, R1–R5
bza_tor	3D	bza_4pt plus additional torsion constraint (C0-C4-hp1-hp2)
ex30	3D	bza_tor + 4 excluded volumes (3 Å shell)
ex35	3D	bza_tor + 13 excluded volumes (3.5 Å shell)
ex40	3D	bza_tor + 42 excluded volumes (4 Å shell)
ex40_08	3D	query ex40, radius of excluded volumes set to 0.8 times the default
ex40_05	3D	query ex40, radius of excluded volumes set to 0.5 times the default
ex40_025	3D	query ex40, radius of excluded volumes set to 0.25 times the default
amd	2D	amidine
amd_3pt	3D	amidine (C0), 2 hydrophobic points, R1, R2, R3
amd_ex35	3D	amd_3pt + 13 excluded volumes (3.5 Å shell)
amd_ex40	3D	amd_3pt + 42 excluded volumes (4 Å shell)

[a]See also text.

Table 5. Results of searching the database XRAY_EXP (30 ligands in their experimental conformation, see text) with UNITY. For various queries as described in Table 4, and two values for the timeout parameter, the number of entries in the hit lists (#Hits), the number of compounds that could not exhaustively be searched within the given time-span (#Timeouts), the average RMSD of the returned conformation with respect to the X-ray conformation, and the CPU time on an SGI R10000 processor are shown

Query	Timeout (s)	# Hits	# Timeouts	Average RMSD ± σ	Time (min)
bza_3pt	60	30	0	0.9 ± 0.51	0.1
bza_4pt	60	30	0	0.9 ± 0.78	0.1
bza_tor	60	30	0	1.2 ± 0.81	0.8
ex30	60	29	1		5.3
ex30	600	29	1	1.2 ± 0.78	6.3
ex35	60	23	7		15
ex35	600	29	0	1.1 ± 0.63	24
ex40	60	9	21		25
ex40	600	27	3	1.0 ± 0.62	95
ex40_08	600	27	3	1.1 ± 0.62	85
ex40_05	600	28	2	1.0 ± 0.63	65
ex40_025	600	28	2	1.2 ± 0.67	59

less hydrophobic patterns. As a consequence, we expect the occurrence of false positive hits – compounds with groups classified as hydrophobic and in the right spatial orientation, but not able to bind into the hydrophobic pockets of the target enzyme. However, we note that this is not in contrast with our general philosophy of using UNITY as a preselection tool.

Despite this problem, the conformations returned by UNITY from searching THR_EXP stay reasonably close to the X-ray conformations: The average RMSD of all heavy atoms is 1.0 Å, less than 10% of the compounds have an RMSD > 2 Å. Incorporation of excluded volume constraints increases the computer time needed for the search by orders of magnitude. Consequently, the conformational search for an increasing number of compounds cannot be completed within the default value for the maximum search time per compound (60 CPU s). Only with a timeout value of 600 s, most compounds are found even for the most complex search considered in this study. However, no significant improvement of the conformations is ob-

Table 6. Searching the database XRAY_CORINA (30 ligands, conformation generated by CORINA) with UNITY. For the meaning of the column headings see Table 5

Query	Timeout (s)	# Hits	# Timeouts	Average RMSD $\pm \sigma$	Time (min)
bza_3pt	60	29	0	2.1 ± 0.80	0.2
bza_4pt	60	29	0	2.1 ± 0.74	0.5
bza_tor	60	29	0	2.1 ± 0.83	0.8
ex30	60	29	0	2.1 ± 0.78	6.3
ex35	60	21	8		16
ex35	600	29	0	2.0 ± 0.77	22
ex40	60	5	24		26
ex40	600	21	7	1.6 ± 0.67	140
ex40_08	600	22	5	1.8 ± 0.78	119
ex40_05	600	24	3	1.9 ± 0.81	96
ex40_025	600	27	2	2.0 ± 0.82	70

Table 7. Searching the database THR_ACT (3107 thrombin ligands) with UNITY: For three activity thresholds (100, 10 and 1 μM) and various queries, the number of entries in the hit lists (Hits), the number of compounds which timed out (To), the number of active compounds in the hit list (Actives) as well as the yield Y and the purity P (for definitions see text), and the CPU time on an SGI R10000 processor (Time) are shown. For the definitions of the database queries see Table 4

Query	Hits	To	100 μM			10 μM			1 μM			Time
			Actives	Y	P	Actives	Y	P	Actives	Y	P	
bza	3107	0	2160	100.0	69.5	1332	100.0	42.9	794	100.0	25.6	25.2 s
bza_2pt	2606	0	1890	87.5	72.5	1150	86.3	44.1	697	87.8	26.7	1.4 h
bza_4pt	2568	0	1886	87.3	73.4	1148	86.2	44.7	697	87.8	27.1	2.5 h
bza_tor	2465	2	1854	85.8	75.2	1139	85.5	46.2	693	87.3	28.1	5.8 h
ex30	2414	3	1828	84.6	75.7	1136	85.3	47.1	695	87.5	28.8	18.9 h
ex35	2277	27	1766	81.8	77.6	1117	83.9	49.1	688	86.6	30.2	58.6 h
ex40	1110	1042	878	40.6	79.1	554	41.6	50.0	329	41.4	29.6	269 h
ex40_08	1563	683	1213	56.2	77.6	767	57.6	49.1	470	59.2	30.1	217 h
ex40_05	1964	379	1513	70.0	77.0	937	70.3	47.7	563	70.9	28.7	161 h
ex40_025	2174	210	1659	76.8	76.3	1033	77.6	47.5	627	79.0	28.8	122 h

served either with the additional torsion constraint or with excluded volume constraints.

As a next test, we applied our database queries to the database XRAY_CORINA, which contains the 30 thrombin ligands in a CORINA generated conformation. Compared to the experimental conformations, the average RMSD of these structures is 2.2 Å, with 7 compounds $>$ 3.0 Å, and 9 with an RMSD $<$ 1.5 Å. If we use only internal constraints in the database query (bza_2pt, bza_4pt, bza_tor, see Table 4), 29 of the 30 compounds are found. The fact that we miss one of the molecules compared to the search of THR_EXP, where all were found, points to an inefficiency of UNITY in scanning the conformational space of the ligands. The 29 hits have an average RMS deviation to the X-ray conformations of 2.1 Å, four compounds have an RMSD of more than 3 Å. Eight compounds have an RMSD of 1.5 Å or less, and, therefore, resemble the experimental binding conformations reasonably well. Moreover, compared to the CORINA generated conformations stored in the database, the UNITY conformations are only little closer to the experimentally determined structures.

Using excluded volumes improves the conformations if at least a 4 Å shell is used, the average RMSD to the crystal conformations then drops to 1.6 Å. This is mainly due to the fact that the four compounds with an RMSD $>$ 4 Å are not found with this sterically rather demanding query. Indeed, even with a long

timeout-value of 600 s, only 21 of the 30 compounds are found.

We, therefore, attempted to relieve this steric demand by reducing the size of the excluded volumes by a factor of 0.8, 0.5 and 0.25. As can be seen from Tables 5 and 6, with decreasing size of the spheres, the number of hits increases, and the computational effort is reduced. However, with decreasing size of the exclusion spheres, the average RMSD of the resulting hits increases; for the smallest radii investigated, no improvement over the query without excluded volumes is obtained.

One reason for the poor agreement of the UNITY generated conformations with the experimental ones is certainly the rather general, qualitative definition of a hydrophobic center that we use in the UNITY database search, which matches a rather high number of atoms in the ligands. This problem is even more pronounced in the present queries, where not just one, but two of these fuzzy hydrophobic centers have to be identified in the ligands. In queries with features that are more rigorously defined (such as hydrogen bond donors/acceptors, or charged functional groups), one would expect a better agreement between the conformations obtained by UNITY and the actual binding modes of the ligands.

Therefore, the use of excluded volumes with the goal of better describing the active site of the target does not – at least with the given version of the database software – seem to be a feasible option for larger, realistic data sets, given the rather significant increase in computer time without a concomitant improvement of the resulting ligand conformations.

Thrombin database

In Table 7, the results for the thrombin test database THR_ACT, which contains 3107 compounds, are summarized. For each query, the total number of hits returned (hits), as well as the number of compounds that could not exhaustively be searched within the given timeout (to) are given. As expected from the results discussed above, with the queries employing only internal constraints almost all compounds can be processed within the timeout value of 600 s. In contrast to this, the queries with excluded volume constraints yield a significant number of 'timed-out' compounds.

For each of the hit lists obtained, we then determined the number of active compounds, i.e. those whose IC_{50} is smaller than a given threshold level (actives), as well as the yield and purity (in %).

Table 8. Number of hits (#Hits), number of time-outs (#Timeout), and CPU times on an SGI R10000 (Time) for different UNITY queries of the ACD. For the definitions of the database queries see Table 4

Query	# Hits	# Timeouts	Time (min)
whole database	167459	–	–
bza	29	0	2.3
bza_3pt	2	0	2.3
amd	1057	0	2.6
amd_3pt	199	0	3.8
amd_ex35	195	0	44
amd_ex40	141	16	540

We start the discussion with the results for a threshold of 100 μM. It is obvious from Table 7 that we deal with a highly biased data set: of the 3107 compounds, 70% bind with an IC_{50} of 100 μM or better to thrombin. This of course is a consequence of the fact that for the purpose of this study we only concentrate on molecules that contain a benzamidine group which by itself binds with an IC_{50} of 220 μmol l^{-1} [21]. This also mirrors the fact that many of the compounds in this database were designed and synthesized with the known characteristics of the thrombin active site in mind. Nevertheless, it is very gratifying that the inclusion of the additional constraints deduced from the GRID calculations still further improves on these results. Using benzamidine plus the two lipophilic sites in the D and P pockets pushes the purity to about 75%; with additional volume constraints, up to 80% of the compounds in the hit list are classified as active.

Parallel to the increase of the purity of the resulting hit lists, not only the total number of hits decreases, but also the number of active compounds that are returned. This is reflected in the decreasing values for the yield; for the ex35 query (see Table 6), only little more than 80% of the possible active compounds are contained in the hit list. This behavior is even worse in the queries employing the 4 Å shell of excluded volumes – depending on the size of the volume spheres only 41–77% of the possible active compounds are in the resulting hit lists.

Similar results are obtained with activity thresholds of 10 or 1 μmol l^{-1}. Again, starting from a rather high level of active compounds, the purity improves with incorporation of more and more complex

internal and external (volume) constraints in the query, whereas the yield decreases correspondingly.

ACD

As discussed above, the thrombin test database THR_ACT is highly biased towards thrombin inhibitors. Therefore, we decided to apply our approach to the Available Chemicals Directory (ACD), consisting of about 167 000 molecules. It mainly contains small, non-drug-like molecules, and, therefore, should be an interesting test case for the query.

Indeed, as Table 8 shows, a 2D query for benzamidines immediately reduces the large number of compounds to a hit list containing only 29 compounds. These are mostly small benzamidine derivatives, and if two additional hydrophobic sites are required, the number of compounds reduces to 16. With the appropriate values for R1, R2 and R3, only two compounds remain. One of these, MFCD00079213, is a bis-benzamidine which is not expected to bind well to thrombin; the other, MFCD00079526, is NAPAP, a well-known thrombin inhibitor with an IC_{50} of about 0.2 μM.

If we search for amidine or guanidine groups instead of the more restricting benzamidine anchor fragment, we obtain 1057 hits. With the additional requirement for two hydrophobic centers at the correct mutual distances R1–R3, 199 compounds match the query. Employing a 3.5 Å shell of excluded volumes does not change the hit list significantly, with a shell of 4 Å the number of hits drops to 141, albeit accompanied by a substantial increase of computer time, and 16 compounds timed out.

Nevertheless, this example demonstrates that our approach can be successfully employed to reduce the large number of compounds available in the ACD to a manageable size. 150 to 200 compounds are an order of magnitude which easily allows a manual inspection of the hit list, with the goal of either identifying likely thrombin binders, or excluding undesirable compounds, e.g. those with long hydrocarbon chains or with reactive functional groups.

The methodology can be easily expanded to other possible P1 moieties. Both 2-amino-pyrimidine and 2-amino-imidazole have been used as basic anchor fragments [22]. Searching the ACD for these groups yields 61 and 2 hits, respectively. The additional requirement for two hydrophobic centers in the appropriate distances reduces the hit lists to zero compounds. This is due to the above-mentioned nature of the compounds compiled in the ACD, which contains only a small number of drug-like molecules.

Conclusions

The present paper shows that the identification of new lead structures can be successfully accomplished by computer-aided screening techniques. The 3D structure of a biological target molecule presents detailed structural requirements for its ligands. The nature of these requirements as well as their geometrical orientation can be extracted computationally with the program GRID. We feel that this procedure cannot be fully automatized as the success will critically depend on the identification of an anchor fragment [56]. An anchor fragment is defined as a minimal recognition motif which is essential for binding. For the identification of these fragments we presently use the de novo design program LUDI. The top scoring LUDI fragments are inspected visually and a fragment is subsequently selected manually. This fragment is then included in the GRID results and a functional 3D query is obtained. As the above study shows, this can be successfully applied to 3D database searches for the identification of new lead structures.

Our method can serve two purposes. Firstly, select compounds which have a high probability of binding and thereby reduce the number of compounds which have to be tested. This is worthwhile only if no high-throughput screening, where up to 100 000 compounds per week can be tested, is available. However, in cases where not all available in-house compounds can be tested, a careful subselection has to be made. A major problem with the presented method is the priorization of the selected compounds. Especially when the most likely candidates have to be selected, a ranking is essential.

The second role of the presented selection method is the identification of new and/or additional lead structures, which can be useful for HTS as well as for non-HTS projects. In both cases the screening of electronically available compounds for new lead structures can be extremely helpful. However, now a different kind of ranking is required: one could use one of the many available diversity programs for the selection of a small, diverse subset. Optimally, such compounds can be ordered and then tested. If this is not possible, they still are extremely helpful to serve as ideas for additional templates which can be used to correctly position the pharmacophoric groups.

Acknowledgements

We are grateful to our colleague Dr. Hans Briem for stimulating discussions. We also thank TRIPOS Inc. for temporary additional UNITY licenses.

References

1. Martin, Y.C., Bures, M.G. and Willett, P., Rev. Comput. Chem., 1 (1990) 213.
2. Martin, Y.C., J. Med. Chem., 35 (1992) 2145.
3. Güner, O.F., Henry, D.R. and Pearlman, R.S., J. Chem. Inf. Comput. Sci., 32 (1992) 101.
4. Downs, G.M. and Willett, P., Rev. Comput. Chem., 7 (1996) 1.
5. Van Drie, J.H., J. Comput.-Aided Mol. Design, 11 (1997) 39.
6. Walters, W.P., Stahl, M.T. and Murcko, M.A., Drug Discovery Today, 3 (1998) 160.
7. Van Drie, J.H. and Lijiness, M.S., Drug Discovery Today, 3 (1998) 274.
8. Clark, D.E., Jones, G., Willett, P., Kenny, P.W. and Glen, R.C., J. Chem. Inf. Comput. Sci., 34 (1994) 197.
9. Hurst, T.J, J. Chem. Inf. Comput. Sci., 34 (1994) 190.
10. Clark, D.E., Westhead, D.R., Sykes, R.A. and Murray, C.W., J. Comput.-Aided Mol. Design, 10 (1996) 397.
11. Makino, S. and Kuntz, I.D., J. Comput. Chem., 18 (1997) 1812.
12. Rarey, M., Kramer, B., Lengauer, T. and Klebe, G., J. Mol. Biol., 261 (1996) 470.
13. Jones, G., Willett, P., Glen, R.C., Leach, A.R. and Taylor, R., J. Mol. Biol., 267 (1997) 727.
14. Baxter, C.A., Murray, C.W., Clark, D.E., Westhead, D.R. and Eldridge, M.D., Proteins Struct. Funct. Genet., 33 (1998) 367.
15. Goodford, P.J., J. Med. Chem., 28 (1985) 849.
16. GRID, version 1.5, Molecular Discovery Ltd., Oxford.
17. UNITY, version 3.x, Tripos Inc., St. Louis, MO.
18. Stubbs, M.T. and Bode, W., Curr. Biol., 4 (1994) 823.
19. Stubbs, M.T and Bode, W., Thromb. Res., 69 (1993) 1.
20. Tapparelli, C., Metternich, R., Ehrhardt, C. and Cook, N.S., Trends Pharmacol. Sci., 14 (1993) 366.
21. Stürzebecher, J., In Machovich, R. (Ed.) Thrombin, Vol. 1, CRC Press, Boca Raton, FL, 1984, pp. 131–160.
22. Wiley, M.R. and Fisher, M.J., Exp. Opin. Ther. Patents, 7 (1997) 1265.
23. Kaiser, B., Drugs Future, 23 (1998) 423.
24. Nar, H. and Tong, L., unpublished results.
25. Boobbyer, D.N., Goodford, P.J., McWhinnie, P.M. and Wade, R.C., J. Med. Chem., 32 (1989) 1083.
26. Wade, R.C., Clark, K.J. and Goodford, P.J., J. Med. Chem., 36 (1993) 140.
27. Wade, R.C. and Goodford, P.J., J. Med. Chem., 36 (1993) 148.
28. Goodford, P.J., J. Chemom., 10 (1996) 107.
29. InsightII (Release 98.0), Molecular Simulations Inc., San Diego, CA.
30. Martin, Y.C., Bures, M.G., Danaher, E.A., DeLazzer, J., Lico, I. and Pavlik, P.A., J. Comput.-Aided Mol. Design, 7 (1993) 83.
31. Marshall, G.R., Barry, C.D., Bosshard, H., Dammkoehler, R.A. and Dunn, D.A., In Olson, E.C. and Christoffersen, R.E. (Eds) Computer Assisted Drug Design, ACS Symp. Ser. 112, ACS, Washington, DC, 1979, pp. 205–226.
32. Bruno, I.J., Cole, J.C., Lommerse, J.P.M., Rowland, R.S., Taylor, R. and Verdonk, M.L., J. Comput.-Aided Mol. Design, 11 (1997) 525.
33. Böhm, H.J., J. Comput.-Aided Mol. Design, 6 (1992) 61.
34. Böhm, H.J., J. Comput.-Aided Mol. Design, 6 (1992) 593.
35. Böhm, H.J., J. Comput.-Aided Mol. Design, 8 (1994) 243.
36. Böhm, H.J., J. Comput.-Aided Mol. Design, 8 (1994) 623.
37. Korolkovas, A., Essentials of Molecular Pharmacology, Wiley-Interscience, New York, NY, 1970, pp. 137–186.
38. Morgan, P.B., Scholtz, J.M., Ballinger, M.D., Zipkin, I.D. and Bartlett, P., J. Am. Chem. Soc., 113 (1991) 297.
39. Weber, A.E., Steiner, M.G., Krieter, P.A., Colletti, A.E., Tata, J.R., Halgren, T.A., Ball, R.G., Doyle, J.J., Schorn, T.W., Stearns, R.A., Miller, R.R., Siegl, P.K.S., Greenlee, W.J. and Patchett, A.A., J. Med. Chem., 35 (1992) 3755.
40. Ash, S., Cline, M.A., Homer, R.W., Hurst, T. and Smith, G.B., J. Chem. Inf. Comput. Sci., 37 (1997) 71.
41. Hibert, M., Hoffman, R., Miller, R. and Carr, A., J. Med. Chem., 33 (1990) 1594.
42. Lloyd, E. and Andrews, P., J. Med. Chem., 29 (1986) 453.
43. Greene, J., Kahn, S., Savoj, H., Sprague, P. and Teig, S., J. Chem. Inf. Comput. Sci., 34 (1994) 1297.
44. Ghose, A. and Crippen, G., J. Comput. Chem., 7 (1986) 565.
45. Clark, D.E., Willett, P. and Kenny, P.W., J. Mol. Graph., 11 (1993) 146.
46. Nicklaus, M.C., Neamati, N., Hong, H.X., Mazumder, A., Sunder, S., Chen, J., Milne, G.W.A. and Pommier, Y., J. Med. Chem., 40 (1997) 920.
47. Martin, Y.C., Bures, M.G., Danaher, E.A. and DeLazzer, J., In Wermuth, C. (Ed.) Trends in QSAR and Molecular Modelling 92, ESCOM, Leiden, The Netherlands, 1993, p. 20.
48. Greenidge, P.A., Carlsson, B., Bladh, L.G. and Gillner, M., J. Med. Chem., 41 (1998) 2503.
49. Rusinko III, A., Skell, J.M., Balducci, R., McGarity, C.M. and Pearlman, R.S., CONCORD, A Program for the Rapid Generation of High Quality Approximate 3D Molecular Structures. University of Texas, Austin, TX. Distributed by Tripos Associates, St. Louis, MO.
50. Sadowski, J. and Gasteiger, J., Chem. Rev., 93 (1993) 2567.
51. Haraki, K.S., Sheridan, R.P., Venkataraghavan, R., Dunn, D.A. and McCulloch, R., Tetrahedron Comput. Method., 3 (1990) 565.
52. Smellie, A., Teig, S.L. and Towbin, P., J. Comput. Chem., 16 (1995) 171.
53. Sadowski, J., Schwab, C.H. and Gasteiger, J., CORINA, 3D Structure Generator, Version 2.0, Universität Erlangen-Nürnberg, 1997.
54. Sadowski, J. and Kubinyi, H., J. Med. Chem., 41 (1998) 3325.
55. Fox, T. and Haaksma, E.E.J., in preparation.
56. Rejto, P.A. and Verkhivker, G.M., Proc. Natl. Acad. Sci. USA, 93 (1996) 8945.