Journal of Computer-Aided Molecular Design, 10 (1996) 397–416
ESCOM

J-CAMD 354

Active-site-directed 3D database searching: Pharmacophore extraction and validation of hits

David E. Clark*, David R. Westhead**, Richard A. Sykes and Christopher W. Murray***

Proteus Molecular Design Ltd, Proteus House, Lyme Green Business Park, Macclesfield, Cheshire SK11 0JL, U.K.

Received 30 March 1996
Accepted 18 May 1996

Keywords: Drug design; Lead generation; Thrombin inhibitors; Scoring function

Summary

Two new computational tools, PRO_PHARMEX and PRO_SCOPE, for use in active-site-directed searching of 3D databases are described. PRO_PHARMEX is a flexible, graphics-based program facilitating the extraction of pharmacophores from the active site of a target macromolecule. These pharmacophores can then be used to search a variety of databases for novel lead compounds. Such searches can often generate many 'hits' of varying quality. To aid the user in setting priorities for purchase, synthesis or testing, PRO_SCOPE can be used to dock molecules rapidly back into the active site and to assign them a score using an empirical scoring function correlated to the free energy of binding. To illustrate how these tools can add value to existing 3D database software, their use in the design of novel thrombin inhibitors is described.

Introduction

Over the last decade, the searching of databases of three-dimensional chemical structures (3D databases) has become an accepted and valuable tool in drug discovery [1]. There are at least two reasons why this has been able to occur. Firstly, the advent of rapid computer programs capable of generating 3D molecular coordinates from 2D connection tables has enabled the conversion of many large corporate and commercial chemical structure databases into 3D format. Several such computer programs are now available, e.g., Refs. 2–6, and these have been evaluated and compared in a number of articles [7–11]. Secondly, software algorithms for searching these databases have been developed and refined [12]. Broadly speaking, these algorithms fall into three classes: methods which seek complementarity to a specified active site such as DOCK [13–18], CLIX [19], LUDI [20] and FLOG [21]; geometric (or pharmacophore; a pharmacophore is an arrangement of atoms or functional groups whose spatial orientation with respect to one another may confer a particular biological activity upon the molecule in which it is present) searching methods [22–32] and similarity searching methods [33]. Both complementarity-based and pharmacophore searching have proven effective in the discovery of novel lead molecules in a variety of drug discovery programs [34–40]. Three-dimensional similarity searching is still at too early a stage of development to be applied routinely to drug discovery projects, although significant progress is being made in this area [41–44]. For detailed reviews of 3D database searching methods and applications, the reader is referred to Refs. 12, 33 and 45.

In general, the complementarity-based programs like DOCK are used in 'direct' drug design scenarios, i.e. where a high-resolution 3D structure of the target macromolecule is available. Conversely, pharmacophore searching has traditionally been applied in 'indirect' drug design situations once a pharmacophore has been elucidated from a series of active (and inactive) molecules using pharmacophore mapping techniques, e.g. Refs. 46–56. However, with the growing number of pharmacologically relevant protein structures being solved, pharmacophores can now be derived directly from an active site, perhaps with a cocrystallised inhibitor in place. The work of Lam

*Present address: Dagenham Research Centre, Rhône-Poulenc Rorer Ltd., Rainham Road South, Dagenham, Essex RM10 7XS, U.K.
**Present address: EMBL Outstation, European Bioinformatics Institute, Hinxton Hall, Hinxton, Cambridge CB10 1RQ, U.K.
***To whom correspondence should be addressed.

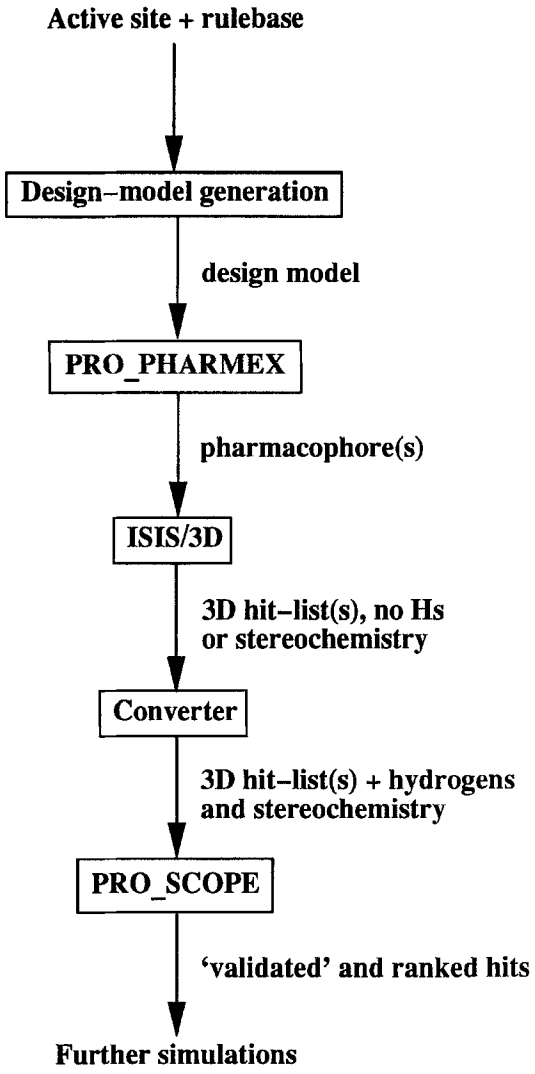

Fig. 1. Overview of the methodology employed in this work showing the relationship between PRO_PHARMEX, PRO_SCOPE and commercial software.

et al. [38] is a good example of the successful application of this type of methodology.

This paper is concerned with the use of 3D database searching methods in structure-based drug design. In particular, two computational tools are described which support, and add value to, existing commercial software for database searching. The first of these programs, PRO_PHARMEX, allows rapid and flexible derivation of pharmacophores from a given active site in the macromolecule of interest. These pharmacophores can then be used as the basis for conducting searches of 3D databases which may be commercial, proprietary or generated de novo [57–59]. The resulting hit lists are then submitted to the second program, PRO_SCOPE, which performs a rapid docking and scoring of the hit molecules in the active site.

An overview of the methodology employed in this work is given in Fig. 1. The following sections of this paper describe in detail the pharmacophore generation software, the techniques used for 3D database searching and the program for docking and evaluating database hits. The use of these software tools in the structure-based design of potential thrombin inhibitors will then be presented and the results discussed before the Conclusions.

Pharmacophore generation

As mentioned earlier, most of the existing pharmacophore generation programs have been developed for the case of 'indirect' drug design. In general, the aim of such programs is to extract from a series of molecules (usually together with their biological activities against the target of interest) a set of common structural features and interfeature distance ranges believed to be responsible for the observed activity. Over the years, a variety of techniques have been developed for this purpose including the active analogue approach [47–50] and ensemble distance geometry [51]. More recently, methods based on clique detection [46,52] and genetic algorithms [54] have emerged together with other novel techniques [53,55,56].

In the case of direct drug design, there are several difficulties that need to be addressed, or at least borne in mind, when trying to construct pharmacophores from an active site [60]. The first of these is how to select the important features within the site to which the ligands are intended to bind. Then there is the issue of defining distance (or other) constraints between those selected points in the site. Another problem is how to include some representation of the receptor-excluded volume in the pharmacophore, if this is deemed necessary. One approach to the automation of this pharmacophore extraction process has been recently presented by Upton and Davies [61]. In what follows, we describe our methodology for the extraction of pharmacophores from active sites and the ways in which we have attempted to address some of the problems mentioned above.

PRO_PHARMEX methodology

Our approach to pharmacophore generation begins by locating regions of space within the active site where functional groups with particular properties are able to bind to the enzyme. Currently, these are hydrogen bond donor regions (where, for instance, NH or OH groups could bind), hydrogen bond acceptor regions (where atoms able to accept H bonds could bind) and lipophilic regions (where lipophilic groups like phenyl rings or alkyl chains can bind). As an option, lipophilic regions can be split into two types, aliphatic and aromatic. Each region located in this manner corresponds to a potential pharmacophore *feature*, and information about the position and spatial extent of these features can be used to derive upper and lower limits on the interfeature distances. Con-

version of this information into a pharmacophore query for a 3D database search is then simply a matter of choosing a set of features to comprise the query, and associating with each chosen feature a suitable chemical substructure for which to search. We choose to ignore the addition of receptor-excluded volume to the pharmacophore at this stage. This is because we find that it is not a particularly effective constraint for a 3D database search, and that it tends to lead to large increases in the CPU time required. Ignoring receptor-excluded volumes may lead to hits in the database search which will not bind because the required conformation forms steric clashes with the receptor. However, all the hits from the database search will be postprocessed by the PRO_SCOPE software, which checks that the hits are able to fit in the active site, meeting the pharmacophore constraints and not forming clashes. We find that the strategy of dealing with receptor clashes at this stage, rather than as excluded volumes in the database search, is much more effective.

Interaction site generation

The location of regions within the site with specific properties makes use of PRO_LIGAND, a computational tool which has been described previously [62]. The design-model generation module of PRO_LIGAND can be used to derive interaction sites from atoms on the enzyme surface and within a user-defined active site. These interaction sites represent spatial positions of functional groups on the ligand which will permit favourable interaction with the enzyme.

Enzyme atoms able to accept hydrogen bonds give rise to **D** sites where donor hydrogen atoms from the ligand can be placed. Each of these sites is in fact linked with another site **X**, defining the position of the heteroatom to which the donor hydrogen is bonded. Sites are generated at a user-specified density within limits defined by hydrogen bond geometries observed in crystal structures [63–65]. Similarly, each enzyme donor hydrogen gives rise to interaction sites where acceptor atoms on the ligand may be placed. These sites are labelled **A** and are linked to sites of type **Y**, which give the position of the ligand atom bonded to the acceptor.

Lipophilic interaction sites are generated from all enzyme atoms defined by the user to be lipophilic. A typical choice would be all carbon atoms. These sites are labelled **L** and, unlike the hydrogen bond type sites, are not paired with any other type of site. The sites are generated at user-specified uniform density on the surface of a sphere of radius 4 Å around the generating atom. The user is free to differentiate between aromatic and aliphatic atoms within the enzyme, by instructing that lipophilic sites generated from aromatic atoms be labelled as type **R** rather than **L**.

The design model comprises all the sites generated as described above, with those sites clashing with enzyme atoms removed. An illustration of a hypothetical design model is shown in Fig. 2. The figure shows **A-Y** vector sites emanating from an amine in the receptor and **D-X** sites from a receptor carbonyl group. Also shown are lipophilic sites mapping the surface of the active site. Note particularly the cluster of lipophilic sites in the 'pocket' in the top right-hand corner of the active site.

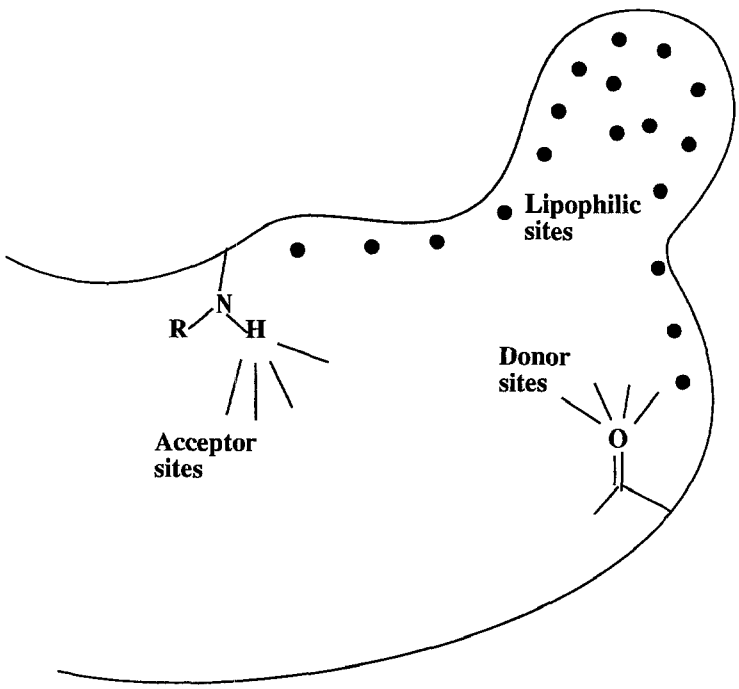

Fig. 2. Schematic of an active site showing donor and acceptor vector sites and lipophilic point sites.

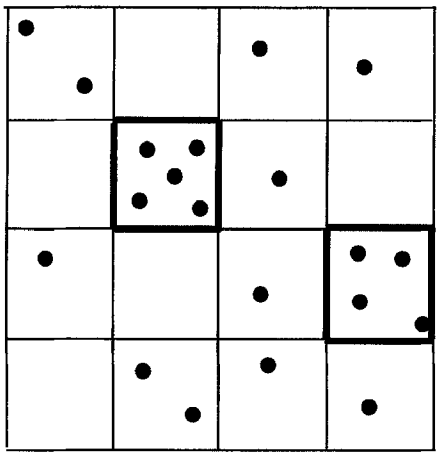

Fig. 3. Detection of significant groups of lipophilic sites using the grid method.

It is worth mentioning at this point that the design-model generation module is a very flexible tool, allowing the user much control over the generation of sites [62]. In particular, allowed hydrogen bond geometries can be varied, as can the density with which sites are generated. Furthermore, the user can choose to generate sites only from specified active-site atoms.

Pharmacophore feature generation

Pharmacophore features are derived by collapsing the information held within the design model into a smaller set of features. Each feature consists of a type (either **D**, **A**, **L** or **R**), and a sphere representing its position and extent. The sphere encloses a region of space in which a functional group of the same type as the feature should reside in order to interact with the enzyme. Features of types **D** and **A** are generated from design model **D-X** and **A-Y** sites, respectively. Each enzyme atom giving rise to these types of sites generates a spatially localised cluster of sites; the sphere is generated as a minimal bounding sphere for the cluster using an algorithm by Ritter [66].

The use of the spherical approximation to features has the advantage that it leads in a very natural way to simple pharmacophores that consist of features and upper and lower bounds on their separation. The drawback is that for the vector hydrogen bond sites any directional information is lost, and sometimes the sphere encloses some space containing no sites. It is possible therefore that the resulting pharmacophores yield hits from the database search that are unable to fit back onto the interaction sites. However, this is not a serious problem, since the PRO_SCOPE software described below is designed to act as a filter on the hits from the database search which, amongst other things, rejects molecules unable to be fitted onto the interaction sites. Our philosophy is thus to generate rather unsophisticated pharmacophores for database searching, and to use further software to prioritise the hits.

It is more difficult to locate parts of the design model where specific lipophilic interactions are available to the ligand (**L**- and **R**-type features). This is because the **L**- and **R**-type design-model sites are not localised near the atom from which they are generated. Indeed, a lipophilic interaction of sufficient significance to include in a pharmacophore would probably involve contacts with many enzyme atoms, as would occur for instance in a lipophilic 'pocket' of an enzyme. In order to locate such regions of significant lipophilic interactions, we first set up a Cartesian grid over the active site. The cells within the grid are cubic and typically of side length 2–4 Å. The number of **L** sites (and **R** sites if used) within each grid cell is counted, and those cells with a significantly higher than average number of sites are used to mark the regions of significant lipophilic interactions. Here, significance is decided by requiring the number of sites to be higher than a user-specified number of standard deviations (SDs) from the mean number (typically 2–2.5 SDs is used). This criterion was preferred to the more straightforward possibility of simply counting the number of sites in each square and comparing with a threshold, because it is possible for the user to generate design models in which the lipophilic sites are created at differing spatial densities around lipophilic protein atoms. The quantity reflecting a genuine lipophilic feature is therefore not the number of sites in a grid cell but rather whether a grid cell is significantly more lipophilic than the rest of the design model.

L- and **R**-type features are then created as minimal bounding spheres for the sites within the chosen cells of the grid, which is illustrated in Fig. 3. This process of spatial discretisation is subject to errors if significant clusters of sites span a number of grid cells. However, this problem can be assessed by perturbing the position of the grid and comparing the two feature sets. In general, for pharmacophore generation, a highly accurate definition of the position of lipophilic regions is not required, since uncertainty can be reflected in larger tolerances on inter-feature distances.

An example of the collapse of the hypothetical design model into a set of spheres representing pharmacophore features is shown in Fig. 4. The figure shows how the cluster of **D-X** sites has been reduced to a single donor feature and, likewise, the cluster of **A-Y** sites to an acceptor feature. The grid analysis of the active site has located a single significant lipophilic feature in the pocket mentioned earlier. Although in this simple example only three features have been located, it is not uncommon in real design situations to find that 10–20 features may be identified from a design model. The user may exert some control over this number by choosing to form a more 'focused' design model (where only a few active-site atoms give rise to interaction sites), in which case fewer features would be formed, and this approach has been used in the work in this paper (vide infra).

Database query creation

Conversion of the set of derived features into a 3D database search query requires input from the user. In particular, the user must specify which of the features should form part of the query, and what chemical substructures should be associated with each feature. The decision about which features to choose is not always straightforward, and it depends to a large extent on the design criteria. Some features may be known to be crucial, for instance those emanating from important catalytic residues, and evidence from the known binding modes of other ligands may point to other important features. The user may wish to study feature distributions from several related proteins in order to assess specificity issues. Of course, in order to define an effective search query some simple rules for feature choice must be followed. A pharmacophore with too large a number of features will produce too few hits in the database; we find that 3–6-feature pharmacophores usually produce reasonable results. Similarly, the interfeature distances must not be so large that only very large molecules could fit the constraints, or so small as to be chemically unreasonable. Even with these considerations it may be necessary for the user to experiment with feature choice in order to produce interesting results from the database search.

The decision about what chemical substructure to associate with each feature depends on the user's requirements and on the chemical nature of the 3D database entries. For instance, the user may require that a given acceptor be satisfied by a carbonyl substructure or a more general acceptor substructure may be desired. Equally, the user may wish to tailor his query to suit the chemical nature of the entries within a particular 3D database.

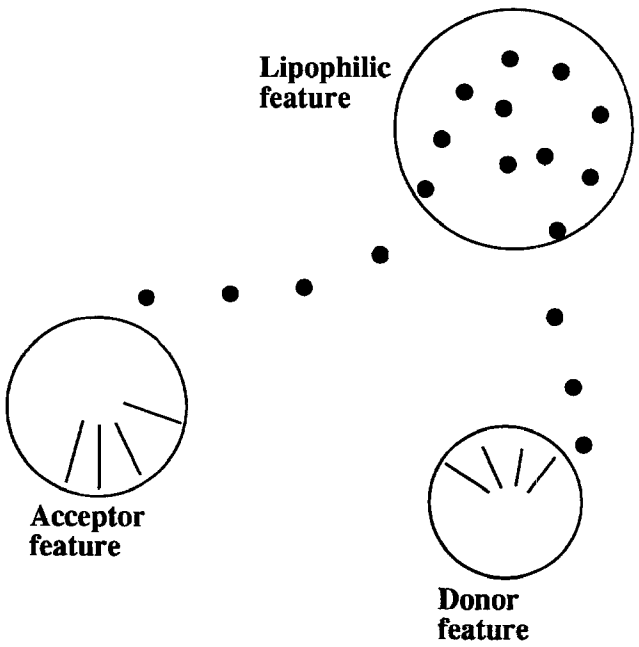

Fig. 4. Collapse of interaction sites into pharmacophore features.

It is clear from the previous paragraphs that pharmacophore generation software should support a great deal of user interaction, and therefore requires a graphical user interface. The PRO_PHARMEX software is thus designed with 'point-and-click' facilities to generate pharmacophore features and to delete existing features. Having decided on a feature set, it is possible to enumerate and display candidate pharmacophores. Enumeration is subject to user-defined constraints on the total number of features included, the number of each type of feature included, the interfeature distances, and which features must be included (or not included) in every pharmacophore. Finally, it is possible to associate, with each feature, a chemical substructure for which to search. This is chosen by the user from a library of possibilities and it is straightforward for the user to add new substructures to the library.

A possible query resulting from our hypothetical example is shown in Fig. 5. Here it can be seen that the lipophilic feature has been represented by an aromatic ring, the acceptor feature by a carbonyl group and the donor feature by a hydroxyl group or amine. The tolerance on the distance ranges between the features is easily controlled by the user through alteration of the radii of the spheres. When a suitable query has been created, it can be written out in MDL MOL file format [67]. In many instances, several such queries may be generated from a given set of features.

Three-dimensional database searching

Once pharmacophores have been extracted from the active-site region and saved as MOL files, they can be imported into the available 3D database system, in our case ISIS/3D [68]. Database searches for a given pharmacophore may be carried out over just the stored conformers ('rigid' search) or conformational exploration may be allowed ('flexible' search). The latter option will invariably generate up to an order-of-magnitude more hits than the former and is considerably more time-consuming. The molecules retrieved from the database can be exported as 3D structures in an SD file [67]. A variety of databases are available, but the studies in this paper use only the Available Chemicals Directory (ACD) [69].

Validation and ranking of hits

Introduction

The further processing of the output from a 3D database search is desirable for a number of reasons. As mentioned earlier, we have chosen not to use any representation of receptor-excluded volume in the pharmacophores generated by PRO_PHARMEX. Thus, it is necessary to check that the molecules retrieved by the database search

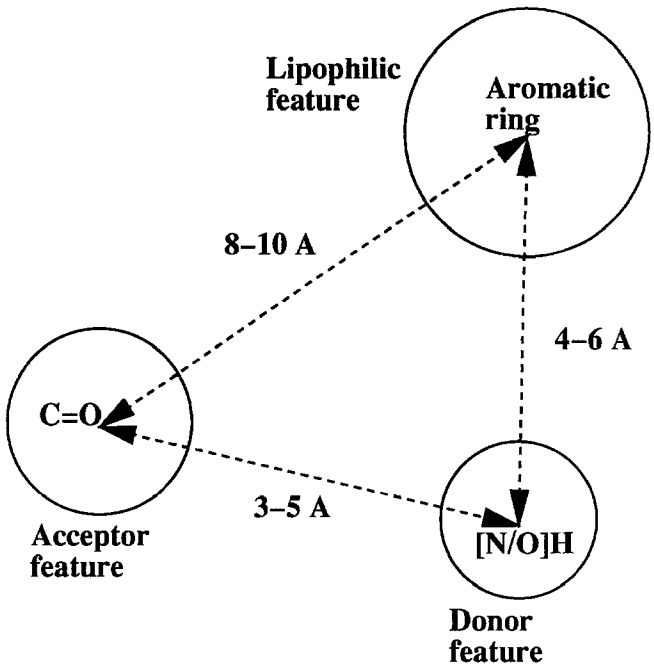

Fig. 5. Conversion of pharmacophore features into a specific pharmacophore query for database searching.

can indeed fit into the active site without making steric clashes with the receptor. A further constraint is that the molecule should be able to adopt a low-energy conformation on binding. It would also be interesting to be able to explore alternative binding modes for a given molecule at the receptor. Finally, when faced with a large number of hits from a database search, some fast method of ranking them would be very helpful in deciding which hits to model further, synthesise or purchase for testing.

The PRO_SCOPE methodology is outlined in Fig. 6 and each of the steps will be described in detail in what follows.

Preprocessing of the SD file

In our work, we have encountered a variety of difficulties with the commercial software in use, and in this section we describe the work-arounds employed to create molecules in a suitable format for processing by PRO_SCOPE.

Three particular problems were encountered:

(i) When structures were saved in three dimensions from ISIS/3D databases, it was noticed that hydrogens attached to sp^2 nitrogens had poor geometries due to a deficiency in the hydrogen sprouting functionality.

(ii) A further problem with saving 3D structures was that the stereochemical parity flags in the connection tables were all set to the D/L indicator value of 3, even if the 2D structure's connection table indicated that the structure was a particular stereoisomer.

(iii) Saving 2D structures was also beset with a problem. Specifically, some amides are sketched in a cis-conformation and this is carried through to the CONVERTER software [5] which is used to generate 3D coordinates from the connection table.

A complete solution to this problem would be to save both 2D and 3D SD files and write a utility to copy the correct stereochemical parity flags from the 2D to the 3D files. For this work, however, it was decided to save the structures as 3D files and allow CONVERTER to select randomly the chirality of the stereocentres during the 3D conversion process. To 'reconvert' 3D structures, it is first necessary to set the dimensionality flags in the input SD file to '2D'; CONVERTER will not process the molecules otherwise.

Molecular property screens

Before subjecting the potential ligand to the more computationally demanding subgraph isomorphism and directed tweak checks, some rapid molecular property screens are used to eliminate unsuitable structures. Thus, the user may set acceptable ranges for a number of properties: molecular weight, number of atoms, number of rotatable bonds and log P (calculated using the method of Viswanadhan et al. [70]). Any structure which falls outside the acceptable ranges is automatically rejected.

Molecule labelling

Before the graph-matching process can take place, the molecule must be labelled with the appropriate interaction sites. Thus, for instance, if the 3D database search is to

be performed with a pharmacophore consisting of two donor groups and an acceptor group, the molecule will need to be labelled with interaction sites representing these groups. This labelling is accomplished by means of a rule-based procedure where each rule denotes a substructure in a SMILES-like notation [71]. For each feature in the pharmacophore, the user must specify what substructures can correspond to that feature by means of one such rule. Each rule further indicates if and how each of the atoms in that substructure should be labelled. Thus, for example, the first rule in Fig. 7 instructs the program to search the molecule for any matches to the specified substructure (C(=NH)N(H)H) and to label the second and fourth atoms of the match as **X** sites and the third and fifth atoms of the match as **D** sites. A powerful regular expression-based syntax is available within the SMILES-like notation which permits very flexible definitions of the rules; for instance, the second rule in Fig. 7 indicates that any OH or NH group attached to a carbon atom should be labelled as a donor group.

In general, there will be more than one mapping of the pharmacophore onto the molecule, and so each of these mappings is stored as a separate group of labels. Each of these groups is explored in sequence during the matching process.

Three-dimensional graph matching

The 3D matching process loops over the number of specified receptor conformations together with their associated design models. These conformations are generally obtained from a molecular dynamics history. This allows some treatment of receptor flexibility in cases where the induced fit is known to be important. This is not believed to be the case for thrombin [72]. Being able to test molecules against multiple receptor models could also be used as a simple test of specificity if more than one receptor type is used (e.g. trypsin versus thrombin). This option has not been tried in this work, however. The user may specify a minimum number of these receptor conformations which must be hit by any molecule which is to be considered as a valid hit.

Thus, once a molecule has been correctly labelled, the program proceeds to seek a 3D match between each group of interaction sites belonging to the molecule and the interaction sites of the design model in question. This is accomplished using the subgraph isomorphism algorithm of Ullmann [73] which has been used successfully in many chemical structure applications. Up to a user-specified maximum number of isomorphisms between each group of labels and the design model are found and stored.

```
COMMENT Loop over hit list
Do i=1, no_of_molecules_in_hit list
   Read molecule from SD file
   Perform molecular property checks
   If FAIL goto next molecule
   Label molecule according to specified rules
COMMENT Loop over number of receptor conformations
Do j=1, no_of_receptor_conformations
COMMENT Begin 3D checks
      Do k=1, no_of_groups_of_labels
         Find Nmatch isomorphisms with design model(j)
         Do l=1, Nmatch
            Access an isomorphism at random
            Seek Directed Tweak fit to design model
            If FAIL goto next isomorphism
            Minimise ligand and ligand/receptor complex (optional)
            Score ligand
         Enddo
      Enddo
   Enddo
Enddo
```

Fig. 6. Pseudocode description of the PRO_SCOPE algorithm.

In order to account for the conformational flexibility of the molecules during matching, distance bounds matrices are calculated using the directed tweak routines which seek to establish the maximum and minimum distances that can be attained between all pairs of atoms [74]. The subgraph isomorphism algorithm then uses these distance ranges in establishing a match in the manner described by Clark et al. [26].

If no matches are found for the molecule's groups of labels, it is rejected and the algorithm returns to consider the next molecule from the hit list.

Directed tweak fitting

The finding of a match for a molecule in the subgraph isomorphism check is a necessary, but not sufficient condition for a molecule to be accepted. This is because the distance bounds matrix does not include correlation effects, i.e. the effect that one interatomic distance having one value might have on the possible values attainable by the other interatomic distances. Thus, the next step in validating a hit molecule is to seek to generate a specific conformation which matches the interaction sites using some form of conformational exploration procedure [26].

The procedure adopted in this work is known as 'directed tweak' [31] and was originally developed for 3D database searching applications, where it has been shown to be both efficient and effective [27,31]. We have recently demonstrated its utility in the field of de novo design [74] and thus it was a natural choice for use in this work.

The directed tweak algorithm selects at random one of the matches established by the subgraph isomorphism algorithm and then seeks to verify it by performing a torsional optimisation, adjusting the molecule's torsion angles so as to minimise the distances between corresponding interaction sites on the molecule and in the design model [74]. If, after a user-specified number of attempts, no satisfactory conformation can be found, the match is rejected and another isomorphism will be considered. Alternatively, if a conformation can be attained which allows a close match between the specified interaction sites and which is also free from van der Waals clashes either internally or with the active site, then the molecule is accepted. It is then passed on for force-field optimisation (optional) and empirical scoring.

Force-field optimisation

As an optional step, the accepted conformation can be minimised in the presence and absence of the active site to calculate an approximation to the strain energy incurred by the molecule on binding. During this minimisation process, the active site is held rigid allowing just the molecule to relax. All molecular mechanics calculations employ the fast and approximate 'Clean' force field developed by Hahn [75]. Partial charges are calculated using the method of Gasteiger and Marsili [76]. The Clean force field bears many similarities to the 'generalised atom' force field incorporated in the CHEM-X software [3,77] in that it does not rely on extended force-field atom types. Only element type, hybridisation and bond type are used in calculating the energy of a system [75]. This ensures that the force field can cope with a diverse range of chemistries without the need for additional parameter development.

```
("C(=NH)N(H)H", "I", 2, "X", 3, "D", 4, "X", 5, "D")

("C^[ON]$H", "I", 2, "X", 3, "D")
```

Fig. 7. Rules for labelling molecules.

Scoring

The (minimised) conformation of the molecule is then assigned a score using a scoring function developed by Böhm [78] for use in the de novo design program LUDI. Böhm's scoring function permits an approximate calculation of binding free energy (kJ/mol) of the molecule in terms of readily calculable quantities such as lipophilic contact surface area, the number and quality of hydrogen bonds formed, and the number of rotatable bonds. Following Böhm [78], the form of the equation used is:

$$\Delta G_{binding} = \Delta G_0 + \Delta G_{hb} \sum_{hbonds} f(\Delta R, \Delta\alpha) + \Delta G_{ionic} \sum_{ionicint} f(\Delta R, \Delta\alpha) + \Delta G_{lipo} |A_{lipo}| + \Delta G_{rot} NRot$$

The values used for the various coefficients are those adopted by Böhm [78] for the LUDI program: $\Delta G_0 = 5.4$, $\Delta G_{hb} = -4.7$, $G_{ionic} = -8.3$, $\Delta G_{lipo} = -0.17$ and $\Delta G_{rot} = 1.4$. Full details of the derivation and calculation of this scoring function may be found in Ref. 78.

Using this scoring function, it is possible to rank the accepted molecules according to the strength of interaction they are likely to make with the receptor. Since the first accepted conformation is not necessarily the highest scoring one available to the molecule, a user-specified number of acceptable conformations will be sought by the 3D matching process and scored. After the specified number of conformations has been found, or all possible attempts exhausted, the highest scoring conformer against the receptor conformation under study is saved for future reference.

Once the loop over the number of receptors is complete, the best-scoring conformation of the molecule taken over all receptor conformations is saved in a separate file.

Example: Searching the ACD for potential thrombin inhibitors

Introduction

The final step in the process of blood clot formation is the hydrolysis of fibrinogen to fibrin by the serine protease thrombin [79]. This enzyme thus constitutes a good target for the development of antithrombotic agents. Human α-thrombin consists of an A chain of 36 amino acids and a B chain of 259 amino acids which are covalently linked by a disulphide bridge. In our studies, we have used the structure of thrombin contained in the Brookhaven Protein Databank as entry 1DWD [80].

The active site of thrombin contains three principal binding pockets [81] which, following the notation of Banner and Hadvary [80], are denoted as S1, D and P. The S1 recognition pocket contains Asp^{189}, which interacts with a guanidinium group from arginine residues or an ammonium group from lysine residues. The D-pocket (signifying its distal relation to the catalytic site) is a hydrophobic pocket which is a favourable binding site for aromatic rings which can interact with the indole ring of Trp^{215}. The last of the binding pockets, the P-pocket (denoting its proximal relation to the catalytic site), is also hydrophobic in nature and is important for thrombin specificity. A further key interaction feature in the active site is the amide NH of Gly^{216}; many low molecular weight thrombin inhibitors form a hydrogen bond with this group. These key active-site features are illustrated in Fig. 8, which also shows the binding of PPACK, a well-known thrombin inhibitor. For a recent computational study of the active site of thrombin using the MCSS method, see Ref. 82.

The aim of the work described in the following sections was: (i) to validate the PRO_PHARMEX and PRO_SCOPE software; and (ii) to locate novel, small organic molecules that bind to thrombin. These might either constitute new lead compounds themselves or, more likely, suggest new templates or scaffolds upon which further structure-based design could take place.

Design-model generation

It was decided to focus upon a subset of these features of the thrombin active site, in particular to target the region containing the S1 recognition pocket and the important Gly^{216} residue. Using the PRO_LIGAND design-model generation module [62], a design model was generated to characterise the complementary interactions for

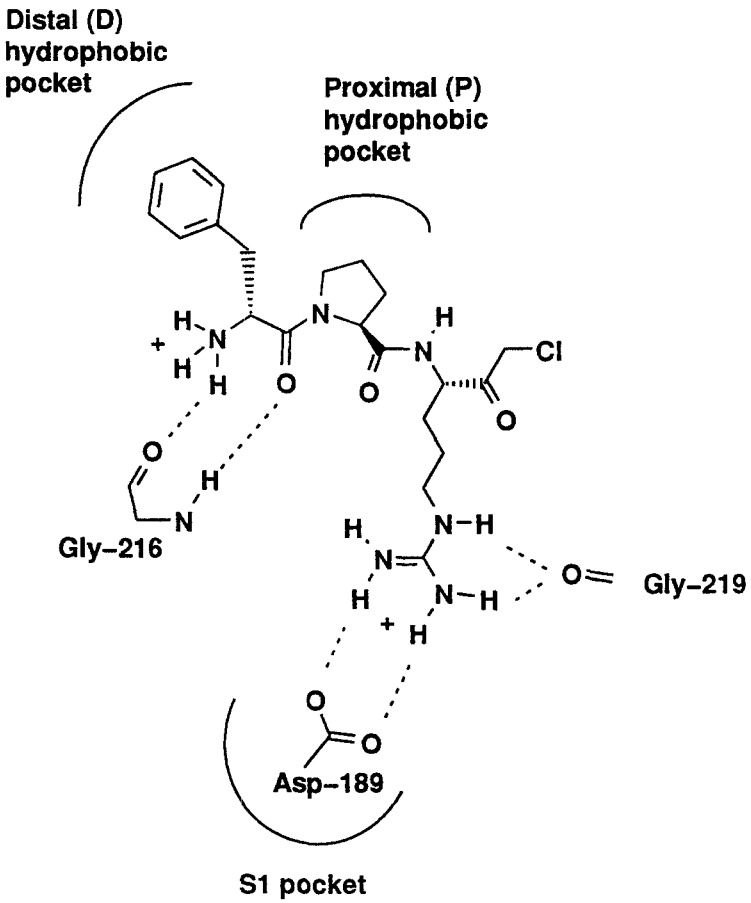

Fig. 8. The thrombin inhibitor, PPACK, showing key interactions with the thrombin active site.

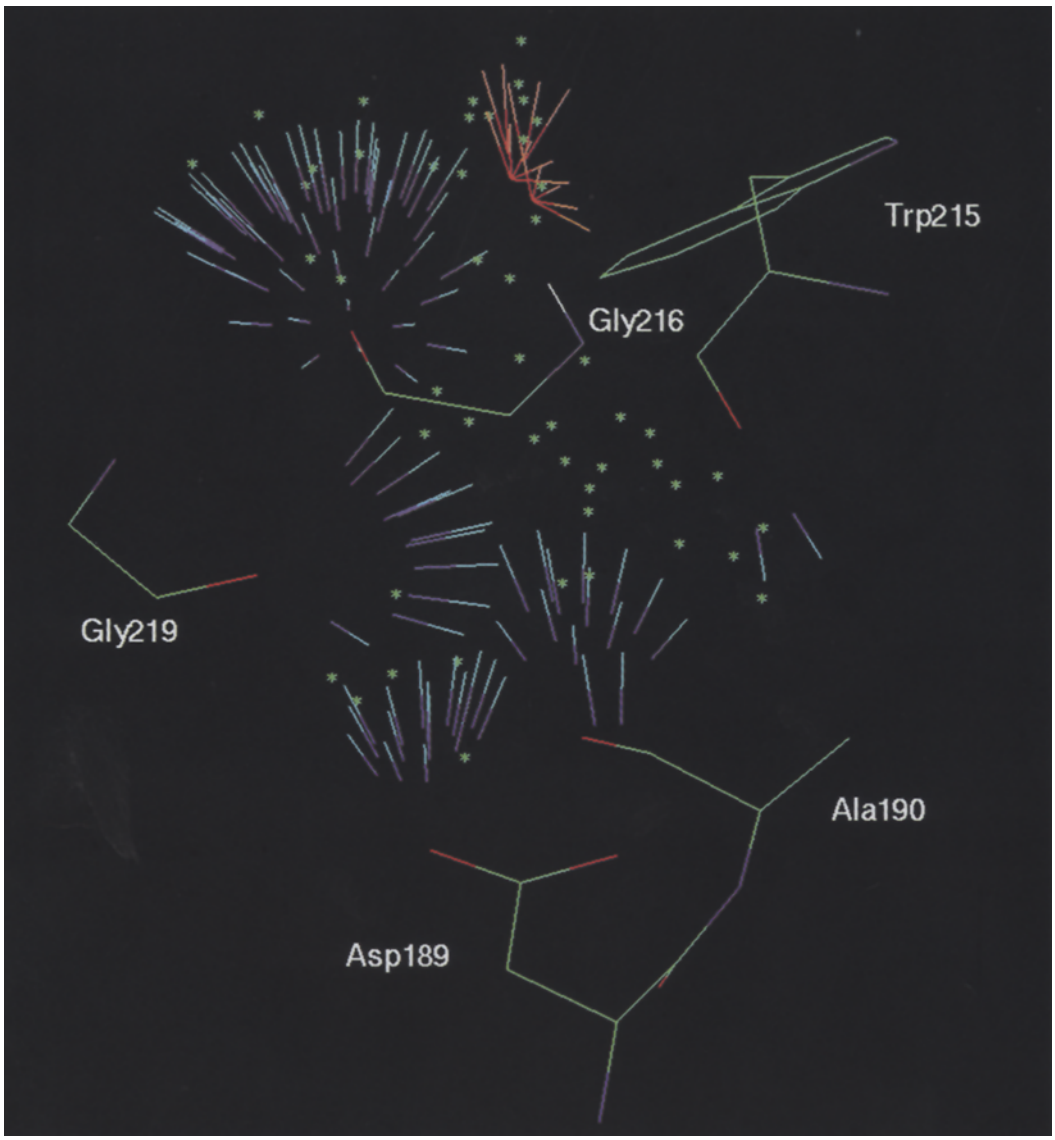

Fig. 9. 'Focused' design model for thrombin. Blue vectors represent hydrogen bond donor sites, orange-red vectors represent hydrogen bond acceptor sites and green asterisks denote lipophilic sites. Also shown are the active-site residues giving rise to the design-model features.

this region of the thrombin active site. The resulting design model, which we shall term the 'focused' design model, is shown in Fig. 9. In addition, a design model for the full active site was prepared; this is shown in Fig. 10 and will be termed the 'full' design model.

Pharmacophore generation

PRO_PHARMEX was run using the 'focused' design model as input. The program identified a total of seven donor and acceptor features (no lipophilic features were requested because the primary interest was in targeting the hydrogen bonding groups in and around the S1 pocket). These seven features are shown superimposed upon the design model in Fig. 11. Of these, features 3, 5 and 7 were selected to constitute a pharmacophore. This pharmacophore is shown in Fig. 12 as it appears in PRO_PHARMEX, and schematically in Fig. 13. As can be seen, the spheres have been resized to a radius of 0.5 Å to give an interfeature distance tolerance of ±1.0 Å in the pharmacophore. This magnitude of tolerance, while somewhat arbitrary, is rather more practical (in terms of the number of hits retrieved from a database search) than the very large distance ranges that would result from using the radii of the minimal bounding spheres. The features have also been assigned an appropriate substructure. Feature 7 has been assigned a carbonyl group to interact with the amide NH of Gly^{216}. Feature 5 has been assigned an NH group to interact with the C=O of Gly^{216}. Finally, feature 3 has been assigned an NH_2 moiety in the hope that it will bind to one of the acceptor features in the S1 pocket.

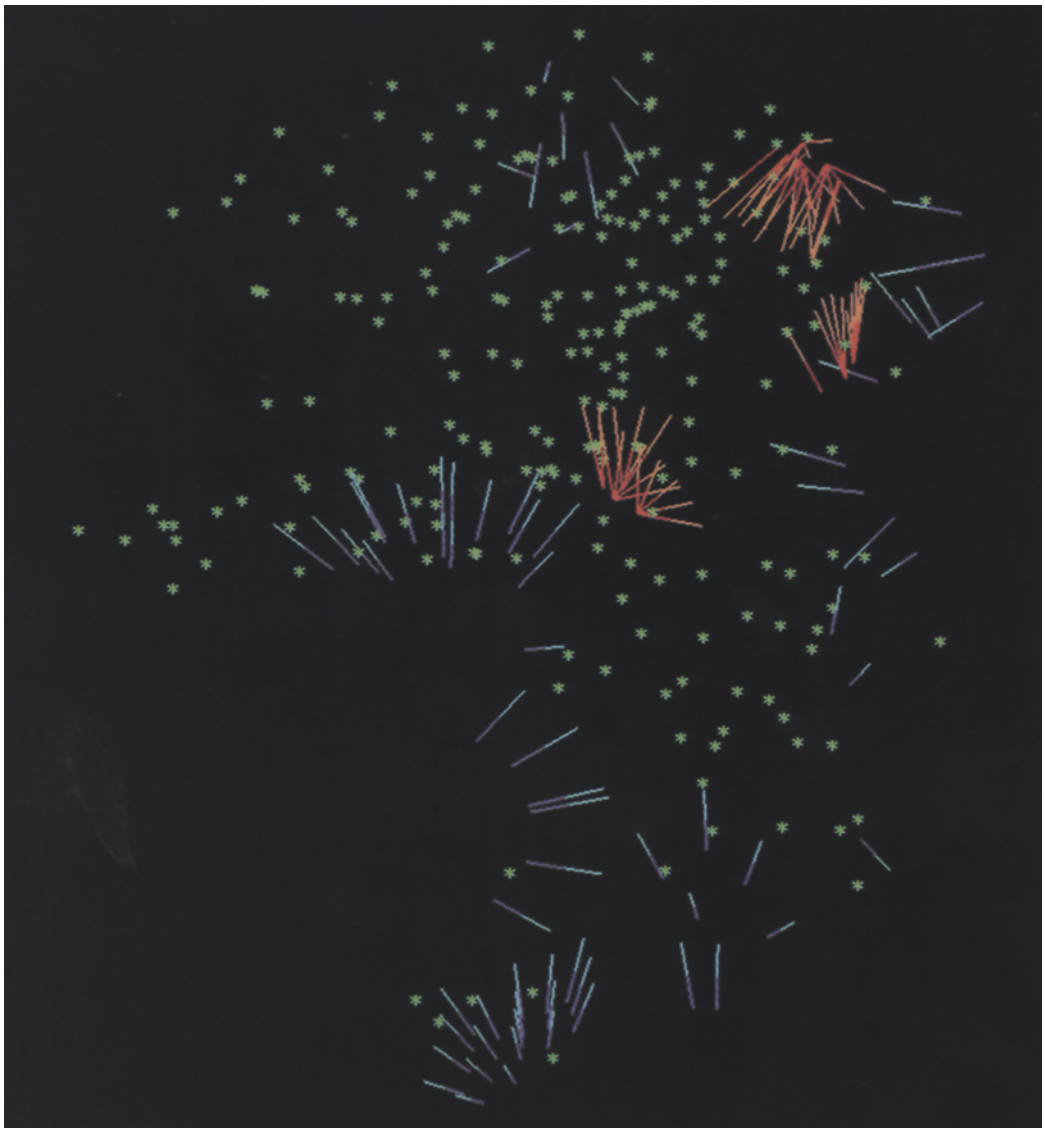

Fig. 10. 'Full' design model for thrombin. Blue vectors represent hydrogen bond donor sites, orange-red vectors represent hydrogen bond acceptor sites and green asterisks denote lipophilic sites.

This pharmacophore was saved from PRO_PHARMEX in MDL's MOL file format [67] ready for importing into ISIS/3D for database searching.

Database searches

The pharmacophore query derived above was used in two separate database searches, one a 'rigid' search and the other a 'flexible' search.

Rigid search

The rigid search was carried out using the SSS search option within ISIS/3D [68] with the 95.1 release of the Available Chemicals Directory [69]. The database was first filtered using a search for compounds with a molecular weight of less than 450. A search of the resulting subset using the pharmacophore of Fig. 13 yielded 779 hits. The hit list was saved as a 3D SD file [67] and, having set the dimensionality flags to 2D, the structures were processed by CONVERTER to add the necessary hydrogen atoms and generate a low energy conformer. The final number of structures after the CONVERTER run was 768. This set of compounds will be termed the 'rigid' set hereafter.

Flexible search

Subsequently, the pharmacophore was used in a conformationally flexible search of a later release of the ACD (95.2). Once again, a subset of the database was created using molecular weight filters, this time consisting of those molecules having a molecular weight of 170–500. This subset was then searched using the CFS (Conforma-

TABLE 1
PRO_SCOPE RESULTS USING THE 'RIGID' SET OF MOLECULES

MaxGraphs	No. of valid hits	Böhm scores (kJ/mol)			Total CPU time (h)
		Worst	Mean	Best	
1	21	9.9	3.2	−6.6	2.9
5	57	11.5	−0.2	−14.3	5.0
10	56	7.8	−1.3	−11.9	7.4
25	70	6.5	−1.8	−16.6	10.5
50	78	6.5	−2.9	−15.5	16.3

tionally Flexible Searching) search option within ISIS/3D [32,68]. The default settings were employed except that van der Waals bump checking was employed at search time as well as at view time, and a cutoff of 7 was imposed on the torsional degrees of freedom permitted between the groups in the molecule which match the pharmacophore [32]. This search yielded 1745 hits, all of which were processed successfully by CONVERTER. Hereafter, this second set of compounds will be termed the 'flexible' set.

PRO_SCOPE runs

Having generated these two sets of compounds satisfying the pharmacophore, PRO_SCOPE was then used to see if these molecules could indeed fit into the active site of thrombin and make favourable interactions with some of the important binding features. For this purpose, it was decided to use the full design model. Although this would allow the compounds to bind to active-site features other than those used to derive the search pharmacophore, it was felt that it would be interesting to see if other binding modes were possible for the compounds. Of course, the focused design model could equally well be used if it were wished to restrict the possible interactions made by the molecules. In order to label the molecules with interaction sites appropriate to the search pharmacophore, three labelling rules were employed. These rules ensured that each group of labels for each molecule consisted of two donor features (corresponding to an NH_2 group and an NH group) and one acceptor feature (corresponding to a C=O group).

Rigid set

The key PRO_SCOPE parameters used with the rigid set of compounds are shown in Fig. 14. The first six parameters are simple molecular property filters used to reject molecules that fall outside the specified ranges. The 'Breadth' parameter indicates that PRO_SCOPE was permitted to find and score up to 10 possible fitting conformations in the active site for each molecule. The 'Tolerance' parameter controls the tolerance permitted in the graph matching of the molecule to the design model. The final parameter indicates the termination criterion used with the Clean force-field minimisation.

A series of five runs were carried out allowing the 'MaxGraphs' parameter to vary. MaxGraphs controls the number of mappings explored for each of the possible labellings of the molecules. The higher the value of MaxGraphs, the greater the number of fittings of a molecule to the design model that will be explored. The results of these experiments are shown in Table 1.

It is interesting to note that over half of the 768 molecules were rejected by the molecular property filters: five failed the number of atoms check, 86 failed the molecular weight check (note that no lower bound was specified in the original database search) and 304 failed the rotatable bond check. These filters left a total of 373 molecules to be investigated by the detailed 3D checks. Looking at Table 1, the results indicate what might be an expected trend: as MaxGraphs is increased, and thus the amount of searching, the mean Böhm scores improve with a concomitant rise in computational expense (CPU times refer to an SGI R4000 Indy machine). The best of the Böhm scores overall, −16.6 kJ/mol, corresponds to a predicted activity of about 1.2 mM.

Flexible set

For the flexible set of compounds, the same parameters were employed as for the rigid set, except that the upper molecular weight limit was increased to 500 and the upper limit on the number of atoms to 100. Again, a number of runs were carried out in which MaxGraphs was allowed to vary. The results of these are shown in Table 2.

Once again, a large proportion of the molecules were eliminated by the molecular property filtering. It is of interest that in spite of the filters employed in the ISIS/3D searches, 65 molecules failed the molecular weight checks and 1033 failed the rotatable bond checks. The reasons for these failures are that, firstly, PRO_SCOPE strips away solvent molecules and counterions that may be included in the molecular weight of an MDL database entry. This can result in a molecule falling below the lower molecular weight limit in PRO_SCOPE, even though it passed the check in MDL. Secondly, the torsional degree of freedom cutoff in the CFS search meas-

TABLE 2
PRO_SCOPE RESULTS USING THE 'FLEXIBLE' SET OF MOLECULES

MaxGraphs	No. of valid hits	Böhm scores (kJ/mol)			Total CPU time (h)
		Worst	Mean	Best	
1	62	10.5	3.5	−6.7	5.8
5	127	14.0	0.7	−11.1	7.7
10	150	10.9	−0.8	−12.8	9.8
25	180	9.7	−2.0	−18.4	13.4
50	205	8.5	−2.3	−15.3	18.3
100	218	6.9	−3.1	−19.2	24.9

TABLE 3
RESULTS OF SIMULATIONS ON PREFERRED HITS

Molecule	PRO_SCOPE (kJ/mol)	Böhm scores (kJ/mol)			Total CPU time (h)
		Worst	Mean	Best	
1 (Maybridge S 15004)	−14.9	−20.2	−23.4	−26.5	12.9
2 (SALOR S91,324-3)	−13.4	−16.7	−18.6	−19.5	11.3
3 (Maybridge CD 03535)	−15.7	−18.1	−19.5	−24.8	11.8
4 (Maybridge S 14081)	−12.9	−8.8	−17.1	−23.2	13.8
5 (Maybridge BTB 03490)	−14.2	−10.8	−17.8	−23.1	13.4
6 (SALOR S77,680-7)	−13.9	1.4	−5.4	−10.2	14.0
7 (SALOR S98,534-1)	−12.7	−5.9	−14.2	−18.2	12.2

ures only the number of rotatable bonds on the paths between the points in the molecule which 'hit' the pharmacophore. Thus, it is often the case that this is only a percentage of the total number of rotatable bonds in the molecule, which is what is measured by PRO_SCOPE. Adding six failures on the number of atoms check, the filters left a total of 641 molecules for 3D fitting in the active site.

Looking at Table 2, the trends are identical to the rigid set. The best overall Böhm score is a little higher: −19.2 kJ/mol corresponding to a predicted activity of about 440 μM. Note that in this table, the CPU times refer to one node of a Convex Exemplar machine.

Simulation of preferred hits

From the two sets of experiments above, 15 molecules scoring −12.0 kJ/mol or less were scrutinised in the active site to examine the binding modes suggested by PRO_SCOPE. From these molecules, a set of seven 'preferred hits' were selected, two from the rigid search and five from the flexible search. These molecules were judged to be attractive because: (i) they are small organic molecules possessing novel chemistry (in the context of thrombin inhibition); and (ii) they form good interactions with the active-site region. The seven molecules are shown in Fig. 15.

These seven molecules were subjected to detailed molecular simulation in the thrombin active site using the DISCOVER program [83] with the CFF95 force field [84] and a distance-dependent dielectric. The model of the thrombin active site was derived from the PDB entry 1DWD and included no water molecules. For each of the molecules, the following simulation protocol was followed:

(i) Minimise the ligand molecule in the active site using a conjugate gradient minimiser for a maximum of 5000 iterations or until the maximum derivative is less than 0.1 kcal/Å. Hold the enzyme fixed.

(ii) Minimise the ligand molecule in the active site using a conjugate gradient minimiser for a maximum of 5000 iterations or until the maximum derivative is less than 0.1 kcal/Å. Allow selected enzyme residue side chains to relax. Save the final geometry as the first snapshot of the MD trajectory.

(iii) Perform 25 ps of molecular dynamics at 150 K, allowing the selected enzyme residue side chains to move. Take a snapshot every 5 ps.

(iv) Perform 75 ps of molecular dynamics at 300 K, allowing the selected enzyme residue side chains to move. Take a snapshot every 5 ps.

(v) Minimise all snapshots using a conjugate gradient minimiser for a maximum of 5000 iterations or until the maximum derivative is less than 0.1 kcal/Å.

The result of this protocol for each of the molecules is a set of 21 minimised snapshots. Each of these snapshots was then scored using the Böhm scoring function. The best, worst and mean of these scores are collated in Table 3 along with the CPU time (on the Convex Exemplar) for the simulation protocol for each of the molecules. Also shown for comparison is the Böhm score for the PRO_SCOPE-fitted conformation of each of the molecules prior to simulation.

TABLE 4
ASSAY RESULTS FOR THE SIX AVAILABLE PREFERRED HITS

Molecule	K_i (thrombin) (mM)	K_i (trypsin) (mM)	K_i (factor Xa) (mM)	Comment
1	1.06 ± 0.14	Inactive	1.10 ± 0.09	
2	1.41 ± 0.18	Inactive	Inactive	Not fully soluble
3	Inactive	Inactive	Inactive	Not fully soluble
5	1.46 ± 0.22	Inactive	1.05 ± 0.08	
6	Inactive	Inactive	Inactive	Not fully soluble
7	1.97 ± 0.34	Inactive	2.75 ± 0.45	

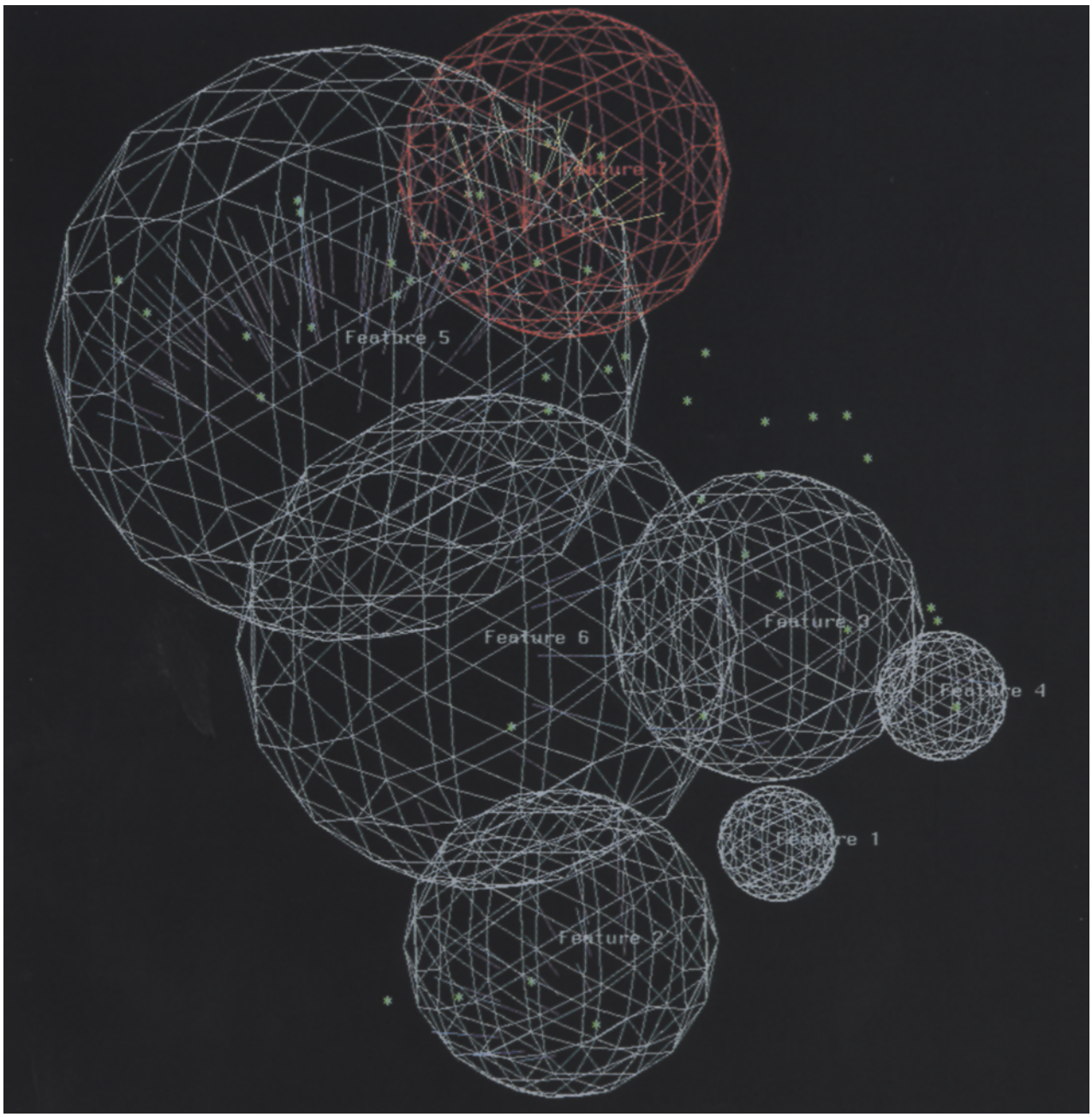

Fig. 11. Pharmacophore features extracted by PRO_PHARMEX from the focused design model. The red sphere encloses the hydrogen bond acceptor sites and the white spheres enclose hydrogen bond donor sites.

Examining the results in Table 3, it can be seen that, in all cases except molecule 6, the simulation results in the molecule achieving a mean Böhm score over the 21 snapshots that is better than that corresponding to the PRO_SCOPE-generated conformation. This is generally attributable to the fact that the CFF95 force field seems to encourage the formation of hydrogen bonds more strongly than does the Clean force field. This results in the conformations generated by the simulation often having a much larger hydrogen bond contribution to the Böhm score than their PRO_SCOPE-generated counterparts. In the case of molecule 6, the simulation reveals that the binding mode suggested by PRO_SCOPE is in fact poor. The proposed hydrogen bonds are transient and there is insufficient lipophilic contact area to compensate for this in terms of predicted binding affinity. Molecule 7, although attractive from the point of view of its chemistry, proved to be very mobile during simulation and even its highest score was only average from the point of view of molecules 1–5. Some of these show very good predicted binding affinities at points during the simulation, although the mean value is probably a safer

indication of any probable activity. As a rough guide, Böhm scores of −17.1 and −22.8 kJ/mol correspond to predicted binding affinities of 1 mM and 100 μM, respectively. Thus, on the basis of these simulations and the Böhm score predictions, we might expect molecules 1–5 to exhibit some activity against thrombin. In the following subsections, we examine the possible binding modes for molecules 1–5 predicted by the simulations.

Molecule 1

Molecule 1 is relatively stable throughout the molecular dynamics trajectory and forms several strong interactions with the thrombin active site. In particular, the benzamide carbonyl group forms a hydrogen bond with the amide NH of Gly^{216} while the NH of the benzamide interacts strongly with the hydroxyl oxygen of Ser^{195}. The NH of the hydrazide group forms a strong hydrogen bond with Ala^{190} while the terminal NH_2 forms a weaker interaction with Trp^{215}. It is apparent too that the molecule can make favourable lipophilic contact with the active site; in particular the CF_3 group is positioned to fit in the P pocket well. These interactions are shown schematically in Fig. 16.

Molecule 2

The interactions made by molecule 2 for much of its trajectory are shown in Fig. 17. As can be seen, there is again the key interaction with Gly^{216}, and the thiocarbonyl group apparently forms a hydrogen bond with Gly^{219}. The hydrazide moiety of this molecule interacts slightly differently with the active site compared to molecule 1. In this case, the NH of the hydrazide forms a

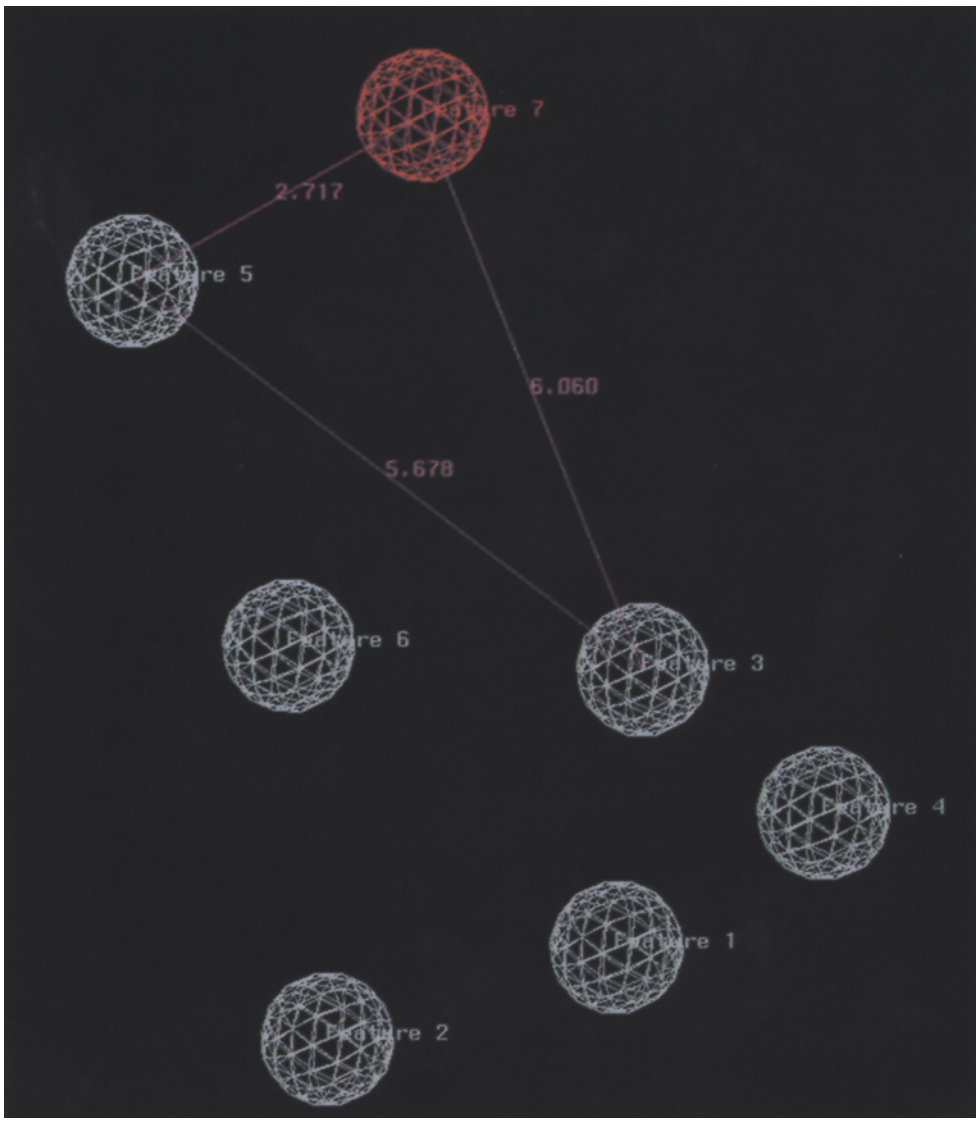

Fig. 12. Pharmacophore generated by PRO_PHARMEX. The red sphere encloses the hydrogen bond acceptor sites and the white spheres enclose hydrogen bond donor sites. Note the resizing of the spheres to give acceptable tolerances on the interfeature distances in the pharmacophore.

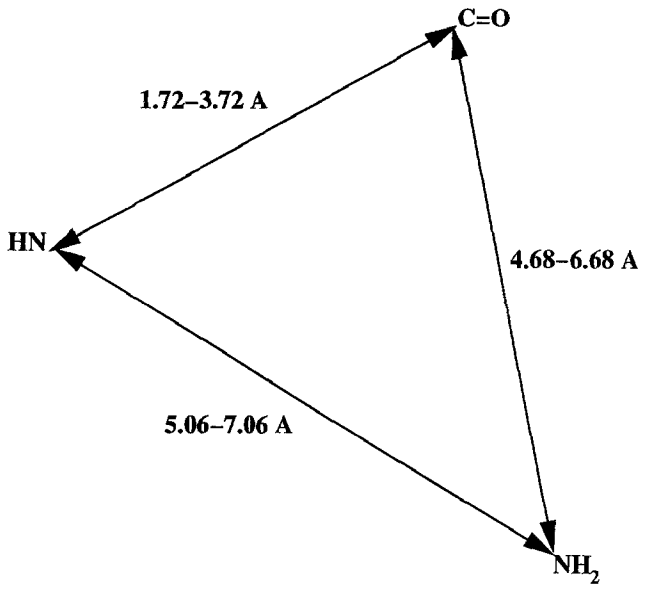

Fig. 13. Schematic representation of a generated pharmacophore.

strong interaction with Gly^{219} while the terminal NH_2 hydrogen bonds to Ala^{190}.

Molecule 3

Molecule 3 exhibits a very different binding mode from those of molecules 1 and 2. It does not enter the S1 recognition pocket at all, but makes several strong interactions in the region around the key Gly^{216} residue. The ring amide forms a kind of β-sheet motif with the amide group of Gly^{216} and the phenylamine moiety curves round to form a further hydrogen bond with the amide carbonyl of that same residue. These three interactions (illustrated in Fig. 18) persist throughout the simulation in what appears to be a very stable binding mode.

Molecule 4

Molecule 4 is distinctly more mobile during the molecular dynamics trajectory than the three preceding molecules, particularly in the flexing of the amide-containing ring. This mobility is reflected in the variation in its Böhm scores during the simulation. Nonetheless, as the simulation progresses, the molecule adopts a geometry forming good interactions with the Asp^{189} of the S1 recognition pocket. Additional hydrogen bonds are formed to the amide carbonyls of Gly^{216} and Gly^{219}. This binding mode is illustrated in Fig. 19.

Molecule 5

Once again, considerable mobility is exhibited by molecule 5 and this is again reflected in the variability of its Böhm scores during the simulation. For part of the simulation, however, an interesting binding mode is observed which is illustrated in Fig. 20. Here it can be observed that the molecule forms interactions with Gly^{216} and Ser^{195} in a manner similar to molecule 1. However, its slightly different position in the S1 pocket means that, whereas molecule 1 formed no interactions with its terminal NH_2 group, molecule 5 can form a moderate hydrogen bond with Asp^{189}. The hydrazide moiety also hydrogen bonds to Gly^{219} and Ala^{190}.

Assay results

Six of the seven 'preferred hits' were purchased from the relevant suppliers (molecule 4 was unavailable for purchase). Molecules 1, 3 and 5 were purchased from Maybridge Chemical Company Ltd. (Tintagel, U.K.). Compounds 2, 6 and 7 were purchased from Aldrich Chemical Company (Gillingham, U.K.). These six molecules were tested for inhibition of thrombin, trypsin and factor Xa using a colorimetric microplate assay with synthetic peptide substrates as described by Tapparelli et al. [85]. The results obtained are presented in Table 4.

As can be seen, four of the six compounds exhibit low millimolar activity against thrombin and three show similar activity against factor Xa. The fact that none of these compounds is active against trypsin may be of interest, in that it may suggest some selectivity for the classes of compound concerned. However, at such low activity levels, it is difficult to be certain of this hypothesis. It was difficult to determine the activity or otherwise of some of the compounds because of their poor solubility. It is thus possible that these compounds are also weakly active against thrombin and factor Xa.

Discussion

The 'mining' of databases of 3D chemical structures has become a valuable means of lead generation over the last decade or so. The growing amount of available, detailed structural information about target macromolecules provides exciting opportunities for the structure-based derivation of search queries and the structure-based validation of the resulting hits. PRO_PHARMEX and PRO_SCOPE are designed to take advantage of these

Minimum molecular weight = 170.0
Maximum molecular weight = 450.0
Minimum number of rotatable bonds = 0
Maximum number of rotatable bonds = 7
Minimum number of atoms = 15
Maximum number of atoms = 75
Breadth = 10
Tolerance = 0.5
Convergence criterion = 0.1

Fig. 14. Input parameters for PRO_SCOPE runs.

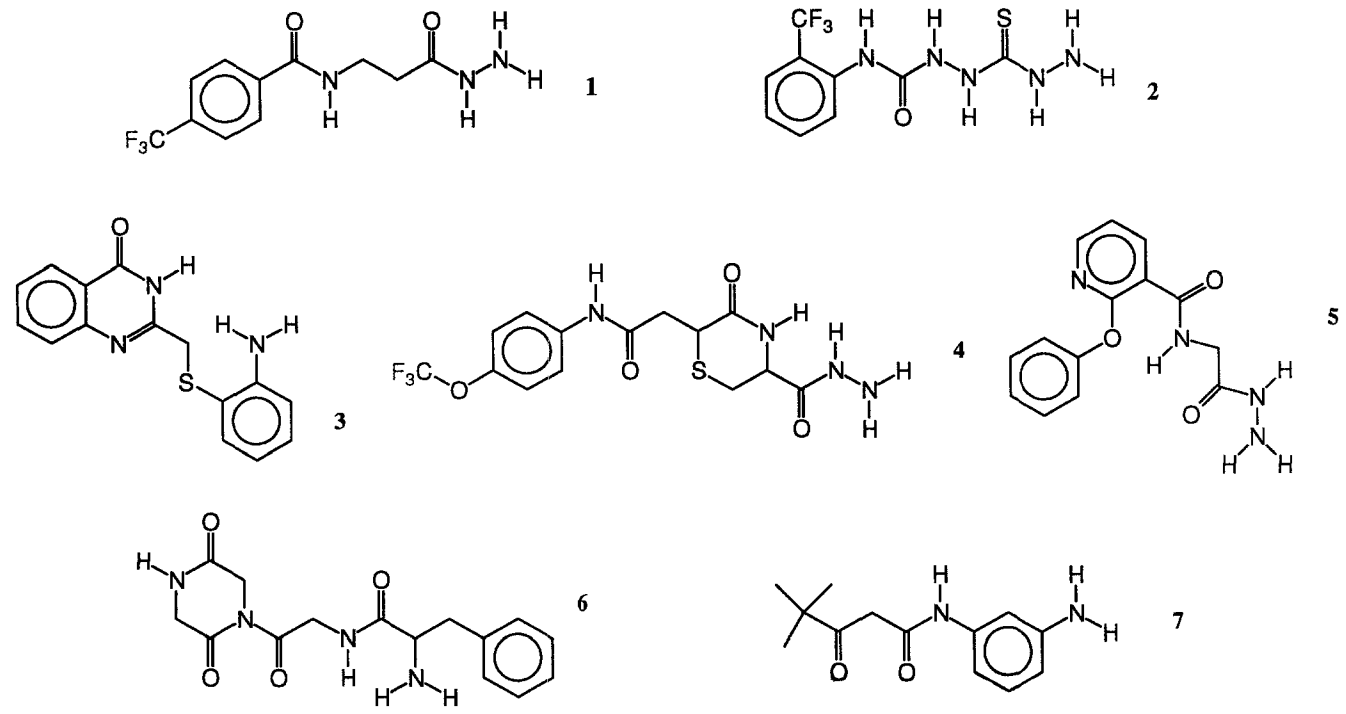

Fig. 15. 'Preferred hits' from PRO_SCOPE runs.

opportunities and to facilitate the work of the molecular designer using 3D databases in structure-based drug design projects. Work of a similar nature has been described briefly by Upton and Davies [61] but a complete description and validation of their method has not been presented in the literature to date.

PRO_PHARMEX represents a novel approach to receptor-based pharmacophore generation. It links naturally with our in-house de novo design package, PRO_LIGAND [62], through its use of the *design model* to characterise the salient physicochemical characteristics of the binding site under study. Being graphically based, PRO_PHARMEX is extremely amenable to user interaction and has many features enabling the user to control the number and nature of the pharmacophores generated. The output of pharmacophores in MDL MOLfile format enables the production of search queries for commercial 3D database systems.

While the underlying concept of PRO_SCOPE, that of rapid molecular docking, is not new, the use of conformational exploration in conjunction with interaction-site guided docking is novel and builds upon the rigid docking studies using LUDI [20]. Of significant interest at the present time is the application of the empirical scoring function developed by Böhm [78] to the ranking of the docked database hits. A computational method for fast

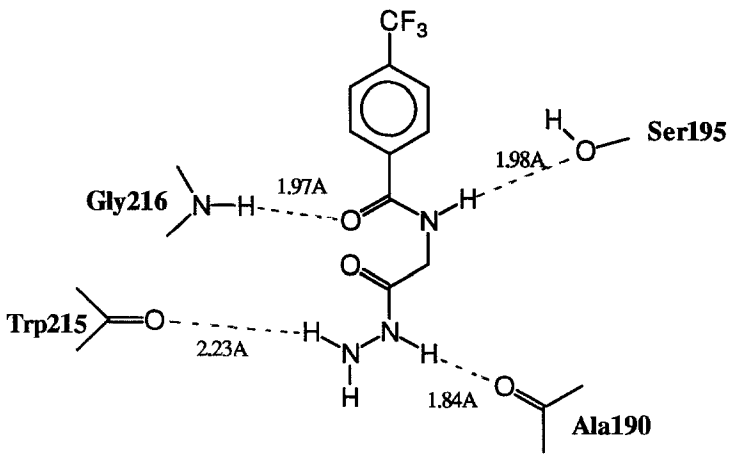

Fig. 16. Interactions of preferred hit 1 during MD simulation.

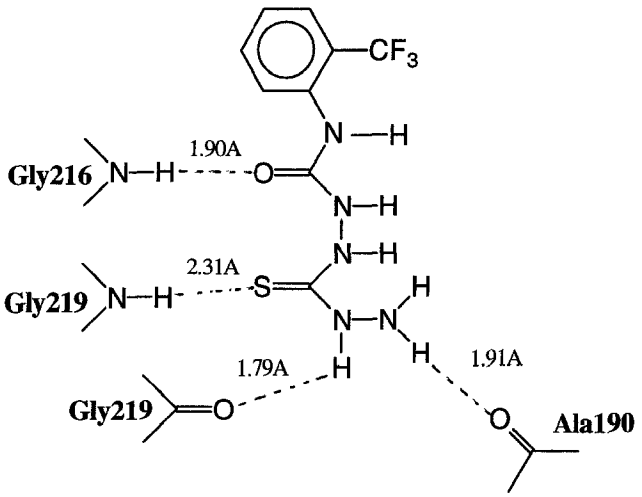

Fig. 17. Interactions of preferred hit 2 during MD simulation.

and reliable estimations of binding affinity is one of the most pressing needs in computer-aided molecular design at the moment [86]. It is thus of interest to see how Böhm's function performs in this situation.

In Table 5, we tabulate the Böhm score, the predicted binding affinity and the actual binding affinity for the four molecules found to be active against thrombin. The Böhm scores and predicted activities are based on the mean values obtained during the MD simulations described earlier. It is clear from Table 5 that the scoring function is capable of predicting the binding affinity of these compounds to within the 1.4 orders of magnitude quoted by Böhm [78]. The rank ordering of the active molecules is also correct, although this may be just fortuitous. While these results are encouraging, other work in-house has shown that such reliability cannot be taken for granted. Clearly, more work is required in the development and testing of such empirical scoring functions.

In terms of the active molecules themselves, molecule 1 is of particular interest because of its relatively simple chemistry and its potential as a template for further design. In particular, it might be possible to attach further substituents to the benzene ring in an attempt to bind to the D pocket. The resulting increase in hydrophobic interactions would be likely to increase the activity of the molecule. The positioning of the CF_3 group in the P pocket is also encouraging for the design of selective thrombin inhibitors. Thus, this molecule in particular satisfies the design aims we set out earlier: it is a small organic molecule, shows activity against thrombin and is a potentially useful starting point for further studies. This result also indicates the value of PRO_PHARMEX and PRO_SCOPE in the active-site-directed searching of 3D chemical structure databases.

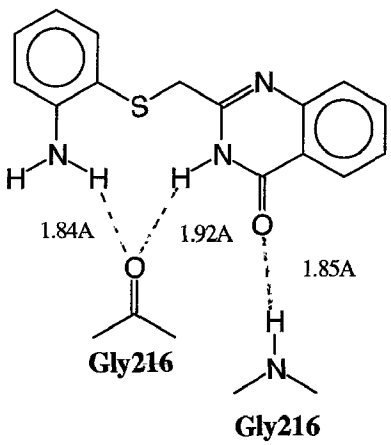

Fig. 18. Interactions of preferred hit 3 during MD simulation.

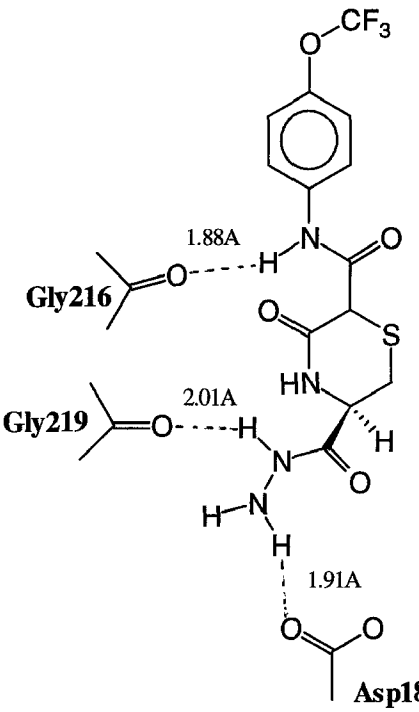

Fig. 19. Interactions of preferred hit 4 during MD simulation.

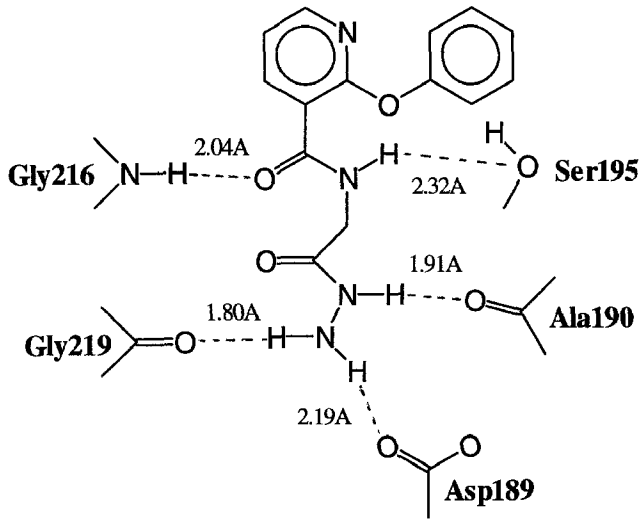

Fig. 20. Interactions of preferred hit 5 during MD simulation.

TABLE 5
COMPARISON OF PREDICTED AND ACTUAL BINDING AFFINITIES FOR ACTIVE MOLECULES

Molecule	Böhm score (kJ/mol)	Predicted binding affinity (mM)	Actual binding affinity (mM)
1	−23.4	0.08	1.06
2	−18.6	0.55	1.41
5	−17.8	0.76	1.46
7	−14.2	3.26	1.97

Future developments in PRO_SCOPE will centre around improvements to the empirical scoring function and also to the sampling technique used to search the conformational space of the molecules in the active site. Another issue that is worthy of attention is the prediction of the aqueous solubility of the database compounds prior to purchase and assay. All of the six molecules in this study presented some difficulty from a solubility point of view and this was only resolvable in four cases. Clearly then, some means of calculating whether or not a particular molecule is likely to be problematic would be an additional and useful screen to apply to the database compounds. The approach of Klopman and co-workers [87] might be helpful in this regard.

Conclusions

We have described two new computational tools, PRO_PHARMEX and PRO_SCOPE, for use in the active-site-directed searching of 3D databases. PRO_PHARMEX is a flexible, interactive, graphically based program enabling the easy formulation of database search queries derived from the active site of a target macromolecule. PRO_SCOPE is designed to aid in the validation of the resulting hits from a database search by checking that the molecules can fit comfortably in the active site and by assigning each molecule a score based on an empirical energy function correlated to binding affinity. The two programs have been used to help discover novel millimolar inhibitors of the enzyme thrombin.

Acknowledgements

The authors would like to acknowledge the contributions of a number of their colleagues at Proteus to this piece of work: Dr. Mike Firth provided programming support, Dr. Harry Martin and Jacqui Mahler assayed the compounds mentioned in the paper, Dr. Steve Young gave valuable assistance in the solubilisation of the active compounds and Dr. Bohdan Waszkowycz provided the molecular simulation protocols described in the paper as well as helping to select the 'preferred hits'. Finally, Dr. Jin Li gave support and encouragement during the course of the research.

References

1 Martin, Y.C., J. Med. Chem., 35 (1992) 2145.
2 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 3-dimensional molecular structures, The University of Texas at Austin, TX, and Tripos Associates, St. Louis, MO, U.S.A., 1988.
3 CHEM-X, Chemical Design Ltd., Chipping Norton, Oxfordshire, U.K., 1995.
4 Sadowski, J. and Gasteiger, J., Chem. Rev., 93 (1993) 2567.
5 CONVERTER, v. 2.3, Molecular Simulations Inc., San Diego, CA, U.S.A., 1995.
6 COBRA, v. 3.0, Oxford Molecular Group, Oxford, U.K., 1993.
7 Hendrickson, M.A., Nicklaus, M.C. and Milne, G.W.A., J. Chem. Inf. Comput. Sci., 33 (1993) 155.
8 Nicklaus, M.C. and Milne, G.W.A., J. Chem. Inf. Comput. Sci., 33 (1993) 639.
9 Ricketts, E.M., Bradshaw, J., Hann, M., Hayes, F. and Tanna, N., J. Chem. Inf. Comput. Sci., 33 (1993) 905.
10 Pearlman, R.S., In Kubinyi, H. (Ed.) 3D QSAR in Drug Design: Theory, Methods and Applications, ESCOM, Leiden, The Netherlands, 1993, pp. 41–79.
11 Sadowski, J., Gasteiger, J. and Klebe, G., J. Chem. Inf. Comput. Sci., 34 (1994) 1000.
12 Bures, M.G., Martin, Y.C. and Willett, P., Top. Stereochem., 21 (1994) 467.
13 Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E., J. Mol. Biol., 161 (1982) 269.
14 DesJarlais, R.L., Sheridan, R.P., Seibel, G.L., Dixon, J.S., Kuntz, I.D. and Venkataraghavan, R., J. Med. Chem., 31 (1988) 722.
15 Meng, E.C., Shoichet, B.K. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 505.
16 Meng, E.C., Gschwend, D.A., Blaney, J.M. and Kuntz, I.D., Proteins Struct. Funct. Genet., 17 (1993) 266.
17 Shoichet, B.K. and Kuntz, I.D., Protein Eng., 6 (1993) 723.
18 Meng, E.C., Kuntz, I.D., Abraham, D.J. and Kellogg, G.E., J. Comput.-Aided Mol. Design, 8 (1994) 299.
19 Lawrence, M.C. and Davis, P.C., Proteins Struct. Funct. Genet., 12 (1992) 31.
20 Böhm, H.-J., J. Comput.-Aided Mol. Design, 8 (1994) 623.
21 Miller, M.D., Kearsley, S.K., Underwood, D.J. and Sheridan, R.P., J. Comput.-Aided Mol. Design, 8 (1994) 153.
22 Gund, P., Prog. Mol. Subcell. Biol., 5 (1977) 117.
23 Jakes, S.E. and Willett, P., J. Mol. Graph., 4 (1986) 12.
24 Jakes, S.E., Watts, N.J., Willett, P., Bawden, D. and Fisher, J.D., J. Mol. Graph., 5 (1987) 41.
25 Brint, A.T. and Willett, P., J. Mol. Graph., 5 (1987) 49.
26 Clark, D.E., Willett, P. and Kenny, P.W., J. Mol. Graph., 10 (1992) 194.
27 Clark, D.E., Jones, G., Willett, P., Kenny, P.W. and Glen, R.C., J. Chem. Inf. Comput. Sci., 34 (1994) 197.
28 Sheridan, R.P., Nilakantan, R., Rusinko III, A., Bauman, N., Haraki, K.S. and Venkataraghavan, R., J. Chem. Inf. Comput. Sci., 29 (1989) 255.
29 Van Drie, J.H., Weininger, D. and Martin, Y.C., J. Comput.-Aided Mol. Design, 3 (1989) 225.
30 Murrell, N.W. and Davies, E.K., J. Chem. Inf. Comput. Sci., 30 (1990) 312.
31 Hurst, T., J. Chem. Inf. Comput. Sci., 34 (1994) 190.

32 Moock, T.E., Henry, D.R., Ozkabak, A.G. and Alamgir, M., J. Chem. Inf. Comput. Sci., 34 (1994) 184.
33 Downs, G.M. and Willett, P., In Boyd, D.B. and Lipkowitz, K.B. (Eds.) Reviews in Computational Chemistry, Vol. 7, VCH, New York, NY, U.S.A., 1996, pp. 1–66.
34 Kuntz, I.D., Science, 257 (1992) 1078.
35 Shoichet, B.K., Stroud, R.M., Santi, D.V., Kuntz, I.D. and Perry, K.M., Science, 259 (1993) 1445.
36 Ring, C.S., Sun, E., McKerrow, J.H., Lee, G.K., Rosenthal, P.J., Kuntz, I.D. and Cohen, F.E., Proc. Natl. Acad. Sci. USA, 90 (1993) 3583.
37 Watts, C.R., Kerwin, S.M., Kenyon, G.L., Kuntz, I.D. and Kallick, D.A., J. Am. Chem. Soc., 117 (1995) 9941.
38 Lam, P.Y.S., Jadhav, P.K., Eyermann, C.J., Hodge, C.N., Ru, Y., Bacheler, L.T., Meek, J.L., Otto, M.J., Rayner, M.M., Wong, Y.N., Chang, C.-H., Weber, P.C., Jackson, D.A., Sharpe, T.R. and Erickson-Viitanen, S.E., Science, 263 (1994) 380.
39 Wang, S., Zaharevitz, D.W., Sharma, R., Marquez, V.E., Lewin, N.E., Du, L., Blumberg, P.M. and Milne, G.W.A., J. Med. Chem., 37 (1994) 4479.
40 Kiyama, R., Homma, T., Hayashi, K., Ogawa, M., Hara, M., Fujimoto, M. and Fujishita, T., J. Med. Chem., 38 (1995) 2728.
41 Pepperrell, C.A. and Willett, P., J. Comput.-Aided Mol. Design, 5 (1991) 455.
42 Turner, D.B., Willett, P., Ferguson, A.M. and Heritage, T.W., SAR QSAR Environ. Res., 3 (1995) 101.
43 Wild, D.J. and Willett, P., J. Chem. Inf. Comput. Sci., 36 (1996) 159.
44 Thorner, D.A., Wild, D.J., Willett, P. and Wright, P.M., J. Chem. Inf. Comput. Sci., 36 (1996) 900.
45 Good, A.C. and Mason, J.S., In Boyd, D.B. and Lipkowitz, K.B. (Eds.) Reviews in Computational Chemistry, Vol. 7, VCH, New York, NY, U.S.A., 1996, pp. 67–117.
46 Golender, V.E. and Vorpagel, E.R., In Kubinyi, H. (Ed.) 3D QSAR in Drug Design: Theory, Methods and Applications, ESCOM, Leiden, The Netherlands, 1993, pp. 137–149.
47 Marshall, G.R., Barry, C.D., Bosshard, H.E., Dammkoehler, R.A. and Dunn, D.A., In Olson, E.C. and Christoffersen, R.E. (Eds) Computer-Assisted Drug Design, ACS Symposium Series, Vol. 112, American Chemical Society, Washington, DC, U.S.A., 1979, pp. 205–226.
48 Mayer, D., Naylor, C.B., Motoc, I. and Marshall, G.R., J. Comput.-Aided Mol. Design, 1 (1987) 3.
49 Dammkoehler, R.A., Karasek, S.F., Shands, E.F.B. and Marshall, G.R., J. Comput.-Aided Mol. Design, 3 (1989) 3.
50 Dammkoehler, R.A., Karasek, S.F., Shands, E.F.B. and Marshall, G.R., J. Comput.-Aided Mol. Design, 9 (1995) 491.
51 Sheridan, R.P., Nilakantan, R., Dixon, J.S. and Venkataraghavan, R., J. Med. Chem., 29 (1986) 899.
52 Martin, Y.C., Bures, M.G., Danaher, E.A., DeLazzar, J., Lico, I. and Pavlik, P.A., J. Comput.-Aided Mol. Design, 7 (1993) 83.
53 Ghose, A.K., Logan, M.E., Treasurywala, A.M., Wang, H., Wahl, R.C., Tomczuk, B.E., Gowravaram, M.R., Jaeger, E.P. and Wendoloski, J.J., J. Am. Chem. Soc., 117 (1995) 4671.
54 Jones, G., Willett, P. and Glen, R.C., J. Comput.-Aided Mol. Design, 9 (1995) 532.
55 Hodgkin, E.E., Miller, A. and Whittaker, M., J. Comput.-Aided Mol. Design, 7 (1993) 515.
56 Barnum, D., Greene, J., Smellie, A.S. and Sprague, P., J. Chem. Inf. Comput. Sci., 36 (1996) 563.
57 Ho, C.M.W. and Marshall, G.R., J. Comput.-Aided Mol. Design, 9 (1995) 69.
58 Glen, R.C. and Payne, A.W.R., J. Comput.-Aided Mol. Design, 9 (1995) 181.
59 Clark, D.E., Firth, M.A. and Murray, C.W., J. Chem. Inf. Comput. Sci., 36 (1996) 137.
60 Van Drie, J.H., Network Science (http://www.awod.com/netsci), 1 (1995).
61 Upton, R. and Davies, E.K., Poster presented at the 14th International Conference of the Molecular Graphics and Modelling Society, Cairns, Australia, August 27–September 1, 1995.
62 Clark, D.E., Frenkel, D., Levy, S.A., Li, J., Murray, C.W., Robson, B., Waszkowycz, B. and Westhead, D.R., J. Comput.-Aided Mol. Design, 9 (1995) 13.
63 Böhm, H.-J., J. Comput.-Aided Mol. Design, 6 (1992) 61.
64 Böhm, H.-J., J. Comput.-Aided Mol. Design, 6 (1992) 593.
65 Klebe, G., J. Mol. Biol., 237 (1994) 212.
66 Ritter, J., In Glassner, A.S. (Ed.) Graphics Gems, Academic Press, London, U.K., 1990, pp. 301–303.
67 Dalby, A., Nourse, J.G., Hounshell, W.D., Gushurst, A.K.I., Grier, D., Leland, B.A. and Laufer, J., J. Chem. Inf. Comput. Sci., 32 (1992) 244.
68 ISIS/3D, MDL Information Systems Inc., San Leandro, CA, U.S.A., 1995.
69 Available Chemicals Directory, MDL Information Systems Inc., San Leandro, CA, U.S.A., 1995.
70 Viswanadhan, V.N., Ghose, A.K., Revankar, G.R. and Robins, R.K., J. Chem. Inf. Comput. Sci., 29 (1989) 163.
71 Weininger, D., J. Chem. Inf. Comput. Sci., 28 (1988) 31.
72 Grootenhuis, P.D.J. and Van Galen, P.J.M., Acta Crystallogr., D51 (1995) 560.
73 Ullmann, J.R., J. Assoc. Comput. Machin., 23 (1976) 31.
74 Murray, C.W., Clark, D.E. and Byrne, D.G., J. Comput.-Aided Mol. Design, 9 (1995) 381.
75 Hahn, M., J. Med. Chem., 38 (1995) 2080.
76 Gasteiger, J. and Marsili, M., Tetrahedron, 36 (1980) 3219.
77 Davies, E.K. and Murrall, N.W., Comput. Chem., 13 (1989) 149.
78 Böhm, H -J., J. Comput.-Aided Mol. Design, 8 (1994) 243.
79 Walkinshaw, M.D., Med. Res. Rev., 12 (1992) 317.
80 Banner, D.W. and Hadvary, P., J. Biol. Chem., 266 (1991) 20085.
81 Obst, U., Gramlich, V., Diederich, F., Weber, L. and Banner, D.W., Angew. Chem. Int. Ed. Engl., 34 (1995) 1739.
82 Grootenhuis, P.D.J. and Karplus, M., J. Comput.-Aided Mol. Design, 10 (1996) 1.
83 DISCOVER, v. 2.9.5, Molecular Simulations Inc., San Diego, CA, U.S.A., 1995.
84 CFF95 force field, implemented in DISCOVER 2.9.5., Molecular Simulations Inc., San Diego, CA, U.S.A., 1995.
85 Tapparelli, C., Metternich, R., Ehrhardt, C., Zurini, M., Claeson, G., Scully, M.F. and Stone, S.R., J. Biol. Chem., 268 (1993) 4734.
86 Ajay and Murcko, M.A., J. Med. Chem., 38 (1995) 4953.
87 Klopman, G., Wang, S. and Balthasar, D.M., J. Chem. Inf. Comput. Sci., 32 (1992) 474.