Computer-based screening of compound databases for the identification of novel leads

Paul W. Finn

In the increasingly competitive pharmaceutical industry, novel targets and mechanisms of action are being investigated in the search for the high-quality drugs required by patients and regulatory authorities. As a result of this move towards novel targets, the initial stage of drug discovery, finding a lead structure, becomes increasingly difficult. Computer-based screening is gaining recognition as a key tool for this part of the drug discovery process. A symbiosis of computational analysis and high-speed synthesis offers exciting potential for the future.

One approach to lead finding is empirical screening: i.e. testing compounds in an assay until an active molecule is found. The compounds screened could be commercially available, natural products or, in the case of the pharmaceutical industry, the historical collection of compounds synthesized in-house. Screening technologies are highly advanced, and often many hundreds of thousands of compounds can be screened in a reasonable time (months), hence the term high-throughput screening (HTS).

High-throughput screening

Although a powerful and proven method of lead discovery, HTS suffers from some disadvantages. One is cost: each individual screening event is cheap, but with many events the total cost can be high, and the problem increases as compound collections grow. Another problem is sensitivity: the ability to find weakly active leads. For example, technical considerations may limit the maximum concentration that can be used in a HTS. A third disadvantage of HTS is that the development of suitable assays can be slow and, for some targets, even impossible.

Computer-based screening offers a complementary approach to lead finding. Using a specific criterion, a computer database of structures is searched (screened) for compounds that match. Compounds identified in this way can be tested, or analysed further. Different strategies will be adopted, depending on the circumstances.

The raw materials: 2D and 3D databases

Databases of chemical structures are central to the activities of pharmaceutical companies, defining the history of the synthesis and acquisition of compounds, and representing the body of chemicals available for screening. These databases can contain hundreds of thousands of structures. 'Two-dimensional' databases, containing the 2D-chemical-structure diagram, were the first to be developed and are the most widespread. In these databases, the structure is represented as a *connection table* (a list of the atoms and their bonded connectivity). Common applications include substructure and similarity searching (see below). 2D databases do not contain information about the 3D structures of the compounds.

The development of fast and effective programs (e.g. Concord, available from Tripos Associates, 1699 South

Paul W. Finn, Pfizer Central Research, Ramsgate Road, Sandwich, Kent, UK CT13 9NJ. tel: +44 1304 618469, fax: +44 1304 618422, e-mail: finnpw@pfizer.com

 Copyright © Elsevier Science Ltd. All rights reserved. 1359-6446/96/$15.00. PII: S1359-6446(96)10033-7

Hanley Road, St Louis, MO 63144, USA, and Cobra, available from Oxford Molecular PLC, Magdalen Centre, Oxford Science Park, Sandford-on-Thames, Oxford, UK) that can take as input a 2D representation of a molecule and generate as output a 3D conformation of that molecule[1], has led to the development of 3D databases. In general, a single conformation is generated, even though many of the molecules represented will be conformationally flexible. As an indication of the performance that can be achieved, the Corina and Concord programs have been used to convert a database of 300,000 in-house compounds from 2D to 3D, with only 107 compounds failing to be converted – a success rate of 99.96%. The failures were mostly structures containing elements rare in our database (for example boron). The conversion required approximately 15 h of CPU time on a Silicon Graphics R4400 processor. A compendium of available 2D–3D conversion programs and 3D databases has appeared recently[2]. In addition to the computer-generated databases, the Cambridge Structural Database System[3] provides an invaluable source of experimentally determined structures. These databases form the raw materials for the searching algorithms used in the identification of novel leads.

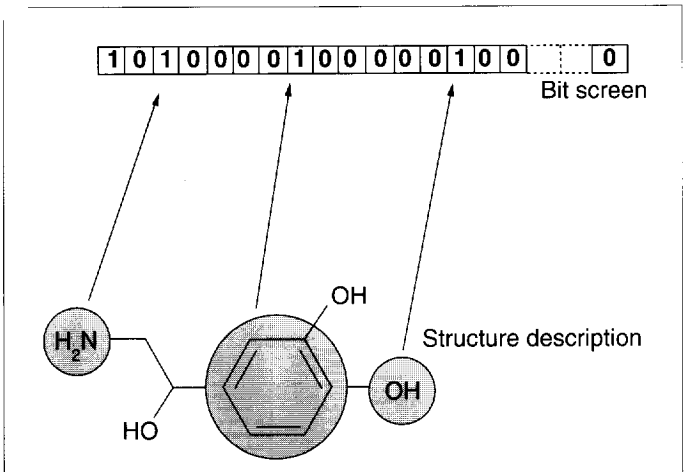


Figure 1. *Example of the definition of a fragment bit screen. The occurrence of a particular structure sets a specific bit in the screen to 1 (on). These bit screens can be checked very rapidly, enabling only a small number of structures to be forwarded to atom matching.*

Following up a known lead

A number of computational methods exist for following up a known active compound. Which method is chosen will depend on the information available.

Substructure searching

Substructure searching will typically be undertaken after a lead has been identified. The query will retrieve those structures from the database that contain groups present in the active molecule (or molecules) are believed to be important for activity. These molecules can then be screened in a biological assay. With databases of more than 10^5 molecules, substructure searches must be efficient. The atom–atom matching (which must be performed to accurately identify a hit) is slow, so methods are required to minimize its use. Fragment-based bit screens (not to be confused with biological screens) have been developed for this purpose. The bit screens encode the presence or absence of small, atom- or bond-centred substructures, and can effectively eliminate a very high percentage of the database before atom–atom matching (Figure 1). Screens of this sort are widespread in database software and also form the basis for many implementations of similarity searching. This type of search is very widespread. As an example, substructure searching has been used to identify a potent tyrosine kinase inhibitor[4] (IC_{50} = 40 nM), starting from an initial lead that was, itself, identified via 3D searching (see below).

Similarity searching

Substructure searching may not find all the structures of interest in a database. For example, if a substructure search specifies a phenyl ring, no other aromatic ring systems will be retrieved, even though they might be quite acceptable. Of course, multiple substructure searches could be performed, but this is tedious and it is difficult to cover all possibilities. Similarity searching provides a way forward by retrieving structures that are similar, but not identical, to the query. This sort of approach is particularly apt in the early stages of a project, when there may be insufficient SAR data to define the important functionality accurately.

In order to proceed in this way, similarity must be defined. Graph theory techniques can be used to identify the maximal common substructure, which gives an accurate indication of the degree of similarity between two molecules. The computational demands of this process make it impractical for large databases, although an application for identifying similarities between protein tertiary structures has been described[5]. The most frequently used similarity measure employs the fragment bit-screen definitions already present in many databases for substructure searching

(Figure 1). It should be borne in mind that these fragment screens were designed to increase the efficiency of substructure searching rather than to be optimal for similarity searching. The similarity of two molecules can then be assessed quantitatively on the basis of how many screen bits they share, calculated via the Tanimoto coefficient. The Tanimoto coefficient has been shown to be appropriate for database applications[6,7], and is defined according to Eqn 1:

$$N(AB) / [N(A) + N(B) - N(AB)] \quad (1)$$

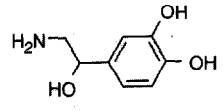
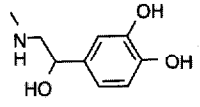
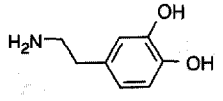
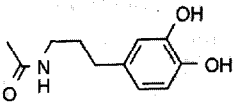
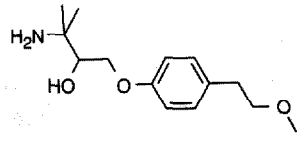
where $N(AB)$ is the number of bits set in common by A and B, $N(A)$ is the total number set by A, and $N(B)$ is the total number set by B. This coefficient ranges from 1 for identical molecules to 0 for molecules that share no common bits. An example of how this coefficient translates into structural terms is shown in Box 1. Functionality of this type exists in commercially available chemical database systems, such as those from Molecular Information Systems (MDL Information Systems Inc., 14600 Catalina St, San Leandro, CA 94577, USA), Daylight (Daylight Chemical Information Inc., #450, 18500 Von Karman Ave, Irvine, CA 92715, USA) and Tripos (Tripos Associates, 1699 South Hanley Rd, St Louis, MO 63144, USA).

While these bit-string descriptions are often used alone for similarity searching, they can be supplemented with other descriptors that can be calculated readily from a 2D chemical diagram. Descriptors of this type include hydrophobicity, topological indices and molecular weight. Recently, researchers at Merck have published findings on the use of physicochemical property descriptors in similarity searching[8]. They extend earlier work on atom-pair[9] and topological-torsion[10] descriptors by devising alternative definitions based on, for example, hydrogen bonding properties. Interestingly, they found that overall these more broadly based descriptors performed less well than the original definitions (as measured by their ability to find molecules in the Derwent Standard Drug File that shared an activity keyword), but that performance was very case-specific. This highlights the advantage of using several different types of descriptors.

Analogously, 3D-similarity searching can be performed, where the similarity between two structures can be defined in terms of, for example, the distribution of interatomic distances, atom mappings, or the maximal common subgraph[11]. For 3D measures, the conformation used will affect the similarity score obtained, and thus potentially similar molecules could be missed because of the arbitrary conformations stored in the database. Nonetheless, the 3D-similarity searches found molecules with significant 3D overlap that were very different at the 2D level[12]. Sheridan and coworkers[13] have compared the performance of a 3D-similarity measure, based on interatomic distances, with 2D-similarity measures. They found that the descriptors are about equally effective, and that there is a broad correlation between the similarity rankings produced by the methods. There are exceptions, however; for example, where compounds are similar because they adopt a folded conformation. The conformational flexibility of the molecules was addressed by using multiple conformations for each structure[14].

Box 1. Examples of similarity measured by the Tanimoto coefficient

Noradrenaline (**1**) is used as the basis for comparison. The related molecules, adrenaline (**2**) and dopamine (**3**), have high Tanimoto coefficients. More distantly related molecules (**4** and **5**) score lower. Below a coefficient of 0.5, little structural similarity remains.

Compound	Structure	Tanimoto coefficient
1		1.0
2		0.91
3		0.78
4		0.6
5		0.5

3D (pharmacophore) searching

Protein–ligand binding is a complex event. This is demonstrated by our limited ability to predict binding affinities, even when we have a detailed structural knowledge of the complex. Some of the factors involved, such as the displacement of ordered water from the protein surface, are hard to estimate. However, the complementarity of shape and electronic properties of the two molecules is undoubtedly a necessary (although not sufficient) condition for binding, and it is this aspect that 3D-search systems attempt to encode and use.

The pharmacophore is the central concept of these systems. The activity of a molecule is deduced to reside in certain key groups (features such as hydrogen bond donors or acceptors or hydrophobic, acidic or basic groups) that interact with the protein. In this model, the rest of the molecule acts either as a skeleton to hold the groups in the right place, or is 'excess baggage'. Many systems have been developed for the deduction of the pharmacophore from series of active (and sometimes inactive) molecules. Typically, the derived pharmacophores consist of 3–5 features, and the distances between them (angles and other geometric measures) are also sometimes used. 3D-searching systems allow specification of pharmacophoric points such as chemical groups (e.g. carbonyl), and most also allow specification of functional features such as hydrogen bond donor or acceptor sites. As might be expected, these latter queries, which encode more generally the physicochemical nature of the interactions at the binding site, find more hits[15], and can yield enhanced SAR information.

The systems initially developed for 3D searching used a single conformation for each structure[16–18]. Just as in 2D, bit screens are used to reduce the number of molecules reaching the CPU-intensive geometric-matching stage. More recent systems take conformational flexibility into account. At present three main approaches are used:

- Storing multiple conformations: a number of studies have been performed aimed at developing methods to produce a set of conformations suitable for database searching[19–22].
- Performing conformational analysis during screen generation and at search time[23].
- Performing torsional minimization at search time[24,25].

Torsional minimization and a genetic algorithm have been found to be efficient for flexible searching[26], whereas distance geometry and systematic search approaches were found to be too slow for database applications.

Applications of 3D searching

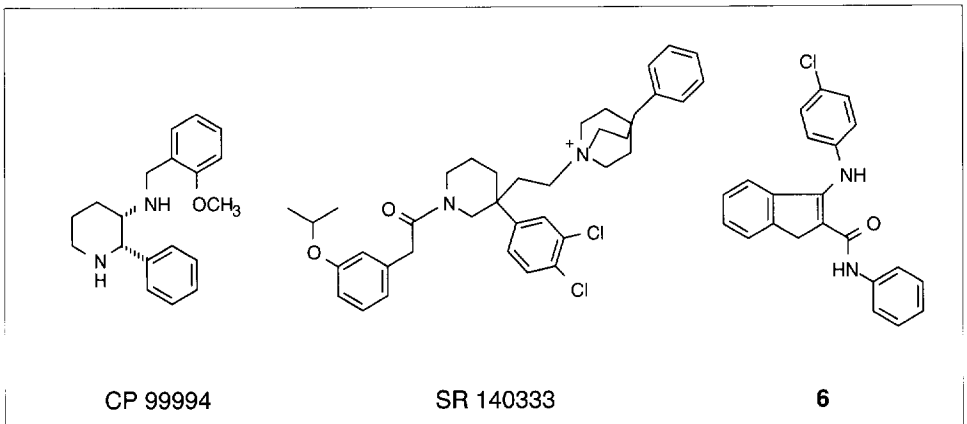
While there is much anecdotal evidence of the efficacy of 3D-database searching for the discovery of novel leads, the number of examples in the literature is still rather small. Using a simple three-point pharmacophore for protein kinase C binding[27], Wang and coworkers[28] searched through 206,876 structures in the National Cancer Institute 3D database[29]. Of the 535 compounds retrieved, 286 were available for testing; of these, 161 were deemed to possess the necessary hydrophobic substituents (which could not be included in the pharmacophore with the software available at the time), and were submitted for testing. Eleven compounds, five of which possessed a K_i <40 μM, were found. Also, inactive molecules that contained the pharmacophore could be converted into active ones by addition of hydrophobic functionality.

Angiotensin II (Ang II) receptor antagonists have also been discovered using a combination of 3D searching and molecular modelling[30]. In this study, the aim was not to identify compounds for screening, but rather to identify replacements for the biphenyltetrazole moiety present in many Ang II antagonists. The MACCS-3D program was used to search through the MACCS-II Drug Data Report 3D (MDDR-3D) database for structures matching a three-point query derived from an overlay of potent, nonpeptidic Ang II antagonists. There were 139 hits, clustered into families based on structural similarity, and these were examined to assess their suitability in terms of novelty, rigidity and synthetic feasibility. The final targets incorporated a dibenzo[*a,d*]cycloheptene tricyclic system, fitted the pharmacophore model, and possessed binding activity in the subnanomolar range.

Two examples from our own work illustrating the use of flexible 3D-database searching are given below.

Example: NK_1 antagonists. This first application led to the discovery of a novel, potent, neurokinin 1 (NK_1) antagonist. Conformational analysis and distance mapping of a number of structurally diverse NK_1 antagonists (e.g. CP 99994 and SR 140333; Figure 2) indicated that hydrophobic interaction was possible for active compounds but not for inactive ones, and that the preferred distance was approximately 5.6 Å. This formed the basis for 3D-database searches using UNITY. Because the query was very general, initial problems were experienced with retrieving large numbers of hits. By using a small distance tolerance of 0.01 Å, we were able to limit the number of hits retrieved to 64. However,

such a small value for the tolerance is physically unreasonable, and subsequent studies have indicated that other filters for limiting hit list size, such as molecular weight or rotatable bond counts, are much better. These hits fell into a number of structural classes because, as is true for most compound collections, there are strong biases for certain structural types. The hits were reviewed manually and representative examples were selected. As we had a relatively small number of compounds left after this selection phase, it was possible to perform some further modelling work to assess the energetic reasonableness of the conformation returned by UNITY, that is, to ascertain that it was close to a local minimum, and not too high in energy (within 10 kcal/mol of the global minimum). Compounds that overcame this last hurdle were submitted for screening without any modification. This screening process identified compound **6** (Figure 2). Its binding affinity for the NK_1 receptor is 300 nM, and it is structurally unrelated to known antagonists.



CP 99994 SR 140333 **6**

Figure 2. *CP 99994 and SR 140333 are representative examples of structures used in the development of the NK_1 antagonist pharmacophore. The hit structure from the UNITY search, compound* ***6****, has a K_i of 300 nM for the NK_1 receptor.*

Example: muscarinic antagonists. The second example illustrates the use of UNITY in template design, in work that we performed for a muscarinic antagonist project. Using a series of known antagonists, we developed a simple three-point pharmacophore model consisting of a basic nitrogen, phenyl ring and hydrogen bond acceptor, illustrated here as a carbonyl oxygen (Figure 3). This pharmacophore was used to search our corporate databases. An overlay of some of the hits (Figure 4) illustrates the diversity of structural types that can be found with this type of search. As expected, many known muscarinic antagonists were found to contain the well-known benzhydryl headgroup (Figure 5a), representing the phenyl and carbonyl part of the pharmacophore, as part of the hit. However, structural fragments that were not contained in any antagonist known to us were also found, as illustrated by the α-substituted cyclohexyl ring (Figure 5b). Subsequent work by D'Agostino and coworkers[31] identified a closely related structure, LG 50643 (Figure 5c) as a potent muscarinic antagonist.

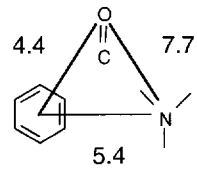
No known or suitable leads

When there are no leads, or those that are available are undesirable, database-searching techniques can still be applied, especially if there is structural information about the target. In this case, protein-derived pharmacophore queries or docking searches can be performed. If there is no protein information, then clustering or dissimilarity ordering of the database can be performed to optimize the screening strategy.

Protein-derived pharmacophore queries

In order to use this approach, the structural information of the active site must be transformed into a 3D-search query. This can be done manually, via molecular modelling, or by the use of a program such as GRID, which identifies energetically favourable binding sites[32]. Using a 3D query derived in this way, database searching will find molecules that can orientate these features correctly for binding. At present, most programs only take limited account of the steric environment of the active site: for example, by the use of excluded-volume spheres (regions of space that must not be occupied by the hit) in the query. In the CLIX method[33],

Figure 3. *Muscarinic pharmacophore definition. Distances are specified in Ångstroms. A tolerance of 0.7 Å was allowed for each distance.*



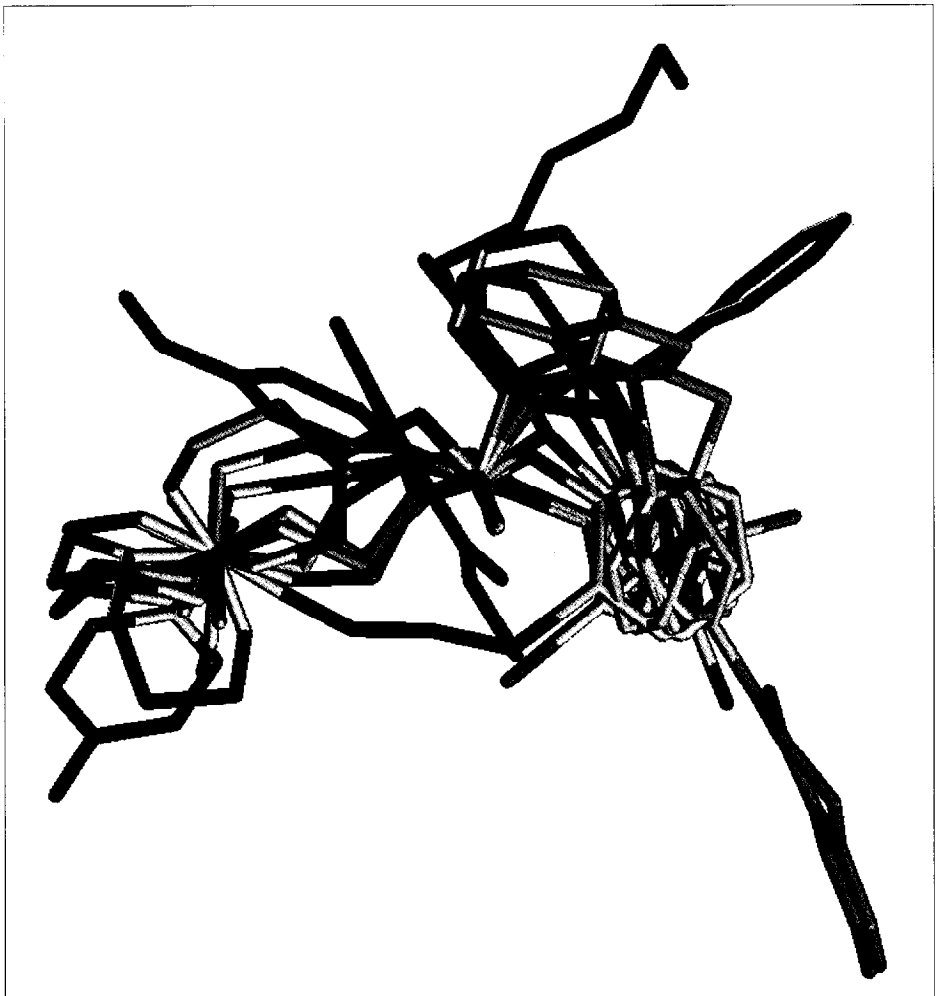


Figure 4. *An example of the results of a 3D pharmacophore search. The pharmacophoric groups are shown in yellow. The common spatial arrangement is achieved with many different chemical frameworks.*

the GRID-derived pharmacophore hits are optimized in the site with molecular mechanics minimization to reduce poor steric interactions.

There are two examples of the discovery of HIV-1 protease inhibitors via database searching. In the first example[34], key hydrogen bonding interactions with a peptidic ligand were identified, and queries were specified to mimic these interactions. A benzophenone derivative of K_i 11 μM was found. In the second example, a database hit was used as the starting point for modelling studies and protein-structure-based drug design, leading to potent nonpeptide inhibitors[35].

Docking methods

These methods differ from the pharmacophore type in concentrating on shape complementarity. The DOCK[36] program is the best known of this family. Originally, the ligands were ranked simply on how badly ligand atoms clashed with the protein; later versions incorporated favourable steric and also electrostatic terms[37,38]. There are several reports of the use of DOCK to discover lead compounds[39–41], including the use of protein structures built by homology modelling[42]. An excellent review of DOCK and related automated docking methods has been published[43].

Clustering methods

When there are no leads or protein structural information on which to base searches for substructure, similarity or pharmacophore, a subset of compounds may be selected by clustering. The objective here is to optimize the use of screening resources by selecting a subset of the total number of compounds. Similar structures will fall into clusters, and similar structures are likely to have comparable biological activities. There are two main decisions to be made when clustering: which description of the structures to use, and which method of clustering. Most work has been performed with the fragment bit-strings used in substructure and similarity searches and the Tanimoto coefficient of similarity, using the Jarvis–Patrick, non-hierarchical clustering method[44], which is fast enough to cluster files of hundreds of thousands of compounds. However, Jarvis–Patrick clustering did not perform as well as the Ward and the group-average hierarchical methods when molecular property, rather than structural data, was used as input[45]. A method has also been described for assessing the geometric diversity of functional groups in chemical databases[46].

That chemical structure is correlated with biological activity sounds intuitively correct, but how strong is the correlation? In a recent, very thorough, analysis of descriptors and clustering methods, Brown and Martin[47] have demonstrated that, in the cases examined, 2D descriptors performed better than 3D ones, and that the Ward clustering method

was consistently the best. Active molecules did tend to fall into the same clusters, but, interestingly, this was less true for high-throughput datasets than for datasets derived from medicinal chemistry programmes. For example: for the medicinal chemistry set, there was a 57% chance that the most similar molecule to an active was itself active; for the HTS set, this was reduced to 8%. To achieve good separation of actives from inactives requires a small cluster. As Brown and Martin state, 'it is probably unreasonable to hope to characterize a large diverse dataset with a small set of representatives but rather that more than 20% of the dataset may have to be tested'. It will be interesting to see how applicable these findings are to datasets generated by combinatorial chemistry.

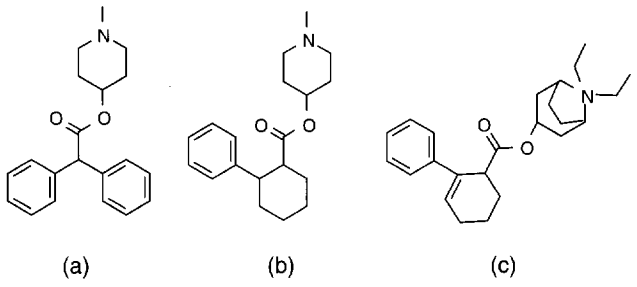


Figure 5. *(a) An example hit containing the benzhydryl headgroup. (b) An alternative hit from the UNITY work. (c) The potent muscarinic antagonist LG 50643, which contains a closely related headgroup.*

Combinatorial chemistry and virtual libraries

Whilst database searching clearly offers great potential for the identification of active molecules, there is always the limitation that only molecules present in the database can be found. Even the large proprietary collections of pharmaceutical companies represent only a tiny fraction of the total potential diversity of drug-like molecules. It is, therefore, an attractive idea to supplement these collections with additional databases of compounds that are not immediately available, but which would be synthesized if they were to be found as hits from a database search. Such databases have become known as virtual libraries.

A method for the generation of 3D databases based on the manipulation of SMILES-string representations of molecules has been described[48], as has a method based on graph-theory methods[49]. Both methods allow significant user specification to control the types of structures built: for example, the maximum number of atoms and rings, or the functionality included or excluded from the molecules. Hits from databases constructed in this manner would then be synthesized without further changes or modified to produce a molecule better fitting the search criteria, or more readily synthesized. A program (MODSMI)[50] has been developed that modifies the scaffold molecules retrieved from a database, so that they match the pharmacophore functionality. An 'ideas' database has been used in the design of Ang II antagonists[51]. The construction of virtual databases is not a new idea: the databases of cyclic systems used by the CAVEAT program[52], which searches on the basis of matching bond vectors rather than atom positions, can be considered as virtual databases because hits would normally act as scaffolds for holding the pharmacophoric groups required in the correct spatial positions.

A major practical consideration for users of these methods is the synthetic accessibility of the target molecules. Given the uncertainties present in most pharmacophore models, and the difficulty of predicting binding affinity even when we have a high-resolution structure of the target protein, to embark on synthesis of a target molecule must be considered a risky business, with a high probability of failure (i.e. the target shows no activity). If the synthesis is long, or unprecedented, the risk is even higher. This difficulty is shared by the programs for *de novo* drug design, which, rather than searching for molecules in a database, construct them piece by piece to match the pharmacophore or active-site requirements. In both these areas, methods such as CAESA (G.J. Myatt, PhD Thesis, University of Leeds, 1994), which estimate the synthetic accessibility of molecules, can be used as a guide to target selection.

Perhaps the most exciting area of current development is the synergy that is being established between combinatorial chemistry and database searching. The techniques of combinatorial chemistry make accessible millions of potential compounds within a particular synthetic scheme, and yet, it may be desirable to make only a selection, perhaps even a small selection, of these compounds. There are obviously a number of ways of making the selection, including a random choice. However, the techniques of clustering, diversity selection and 3D-database searching, discussed in the context of selecting molecules for testing from a large collection of existing compounds, can also be applied to the potential products of combinatorial synthesis by construction of, and selection from, a virtual library of products. Using this approach, only the molecules that best fit the

criteria would be synthesized. Furthermore, active molecules could easily be followed up, via construction of further libraries based on the actives. The guaranteed accessibility of the members of a combinatorial chemistry virtual library offers a powerful advantage.

Despite increasing hardware capabilities, no systems are available to store in a database the astronomical numbers of compounds that may potentially be present in a combinatorial virtual library. Either some filtering must be performed (for example based on diversity) to reduce the library size to manageable numbers, or other methods that do not rely on exhaustive enumeration of the library must be employed. In an interesting example of this latter category, a genetic algorithm has been used to suggest peptoid combinatorial libraries based on 2D measures of similarity to known active molecules[53]. The algorithm succeeded in finding high-scoring molecules, even though the virtual library contained approximately 20 billion compounds.

The way forward

Computer-based screening will continue to be a frontier research area for the foreseeable future, with developments in database technology, pharmacophore generation, pharmacophoric descriptions, flexible ligand docking and scoring functions. Further improvements in high-throughput screening and high-speed synthesis can also be anticipated. The most efficient and effective lead-finding and optimization strategies for the future will use increasingly tight integration between the computational and the experimental methods.

ACKNOWLEDGEMENTS

I thank Drs M. Snarey, C. Burt and A. Alex for their helpful discussions, and Dr Y.C. Martin for the provision of reference 47 prior to publication.

REFERENCES

1 Sadowski, J., Gasteiger, J. and Klebe, G. (1994) *J. Chem. Inf. Comput. Sci.* 34, 1000–1008
2 Manallack, D.T. (1996) *Drug Discovery Today* 1, 231–238
3 Allen, F.H. *et al.* (1991) *J. Chem. Inf. Comput. Sci.* 31, 187–204
4 Ward, W.H.J. *et al.* (1994) *Biochem. Pharmacol.* 48, 659–666
5 Grindley, H.M. *et al.* (1993) *J. Mol. Biol.* 229, 707–721
6 Willett, P., Winterman, V. and Bawden, D. (1986) *J. Chem. Inf. Comput. Sci.* 26, 36–41
7 Moock, T.E. *et al.* (1988) *Tetrahedron Comput. Methodol.* 1, 117–128
8 Kearsley, S.K. *et al.* (1996) *J. Chem. Inf. Comput. Sci.* 36, 118–127
9 Carhart, R.E., Smith, D.H. and Venkataraghavan, R. (1985) *J. Chem. Inf. Comput. Sci.* 25, 64–73
10 Nilakantan, R. *et al.* (1987) *J. Chem. Inf. Comput. Sci.* 27, 82–85
11 Pepperrell, C.A. and Willett, P. (1991) *J. Comput.-Aided Mol. Design* 5, 455–474
12 Willett, P. (1995) *J. Mol. Recog.* 8, 290–303
13 Sheridan, R.P. *et al.* (1996) *J. Chem. Inf. Comput. Sci.* 36, 128–136
14 Kearsley, S.K. *et al.* (1994) *J. Comput.-Aided Mol. Design* 8, 565–582
15 Greene, J. *et al.* (1994) *J. Chem. Inf. Comput. Sci.* 34, 1297–1308
16 Jakes, S.E. *et al.* (1987) *J. Mol. Graph.* 5, 41–48
17 Van Drie, J.H., Weininger, D. and Martin, Y.C. (1989) *J. Comput.-Aided Mol. Design* 3, 225–251
18 Sheridan, R.P. *et al.* (1989) *J. Chem. Inf. Comput. Sci.* 29, 255–260
19 Leach, A.R. (1994) *J. Chem. Inf. Comput. Sci.* 34, 661–670
20 Smellie, A., Teig, S.L. and Towbin, P. (1995) *J. Comput. Chem.* 16, 171–187
21 Smellie, A., Kahn, S.D. and Teig, S.L. (1995) *J. Chem. Inf. Comput. Sci.* 35, 285–294
22 Smellie, A., Kahn, S.D. and Teig, S.L. (1995) *J. Chem. Inf. Comput. Sci.* 35, 295–304
23 Murrall, N.W. and Davies, E.K. (1990) *J. Chem. Inf. Comput. Sci.* 30, 312–316
24 Moock, T.E. *et al.* (1994) *J. Chem. Inf. Comput. Sci.* 34, 184–189
25 Hurst, T. (1994) *J. Chem. Inf. Comput. Sci.* 34, 190–196
26 Clark, D.E. *et al.* (1994) *J. Chem. Inf. Comput. Sci.* 34, 197–206
27 Rando, R.R. and Kishi, Y. (1992) *Biochemistry* 31, 2211–2218
28 Wang, S. *et al.* (1994) *J. Med. Chem.* 37, 4479–4489
29 Milne, G.W.A. *et al.* (1994) *J. Chem. Inf. Comput. Sci.* 34, 1219–1224
30 Kiyama, R. *et al.* (1995) *J. Med. Chem.* 38, 2728–2741
31 D'Agostino, G. *et al.* (1994) *J. Pharm. Pharmacol.* 46, 332–336
32 Goodford, P.J. (1985) *J. Med. Chem.* 28, 849–857
33 Lawrence, M.C. and Davis, P.C. (1992) *Proteins* 12, 31–41
34 Bures, M.G. *et al.* (1990) *Tetrahedron Comput. Methodol.* 3, 673–680
35 Lam, P.Y.S. *et al.* (1994) *Science* 263, 380–384
36 Kuntz, I.D. *et al.* (1982) *J. Mol. Biol.* 161, 269–288
37 DesJarlais, R.L. *et al.* (1988) *J. Med. Chem.* 31, 722–729
38 Meng, E.C., Shoichet, B.K. and Kuntz, I.D. (1992) *J. Comput. Chem.* 13, 505–524
39 Grootenhuis, P.D.J. *et al.* (1990) *Anti-Cancer Drug Design* 5, 237–242
40 Shoichet, B.K. *et al.* (1993) *Science* 259, 1445–1450
41 DesJarlais, R.L. *et al.* (1990) *Proc. Natl. Acad. Sci. U. S. A.* 87, 6644–6648
42 Ring, C.S. *et al.* (1993) *Proc. Natl. Acad. Sci. U. S. A.* 90, 3583–3587
43 Blaney, J.M. and Dixon, J.S. (1993) *Perspect. Drug Des. Discovery* 1, 301–319
44 Willett. P., Winterman, V. and Bawden, D. (1986) *J. Chem. Inf. Comput. Sci.* 26, 109–118
45 Downs, G.M., Willett, P. and Fisanick, W. (1994) *J. Chem. Inf. Comput. Sci.* 34, 1094–1102
46 Boyd, S.M. *et al.* (1995) *J. Comput.-Aided Mol. Design* 9, 417–424
47 Brown, R.D. and Martin, Y.C. (1996) *J. Chem. Inf. Comput. Sci.* 36, 572–584
48 Ho, C.M.W. and Marshall, G.R. (1995) *J. Comput.-Aided Mol. Design* 9, 65–86
49 Clark, D.E., Firth, M.A. and Murray, C.W. (1996) *J. Chem. Inf. Comput. Sci.* 36, 137–145
50 Martin, Y.C. (1992) *J. Med. Chem.* 35, 2145–2154
51 Mason, J.S. (1995) in *Molecular Similarity in Drug Design* (Dean, P.M., ed.), pp. 138–162, Blackie Academic and Professional
52 Lauri, G. and Bartlett, P.A. (1994) *J. Comput.-Aided Mol. Design* 8, 51–66
53 Sheridan, R.P. and Kearsley, S.K. (1995) *J. Chem. Inf. Comput. Sci.* 25, 310–320