Getting that hit: 3D database searching in drug discovery

David T. Manallack

At each stage in the drug discovery process, 3D database searching can play a significant role – from finding a lead compound to searching for alternative backup structures. Medicinal chemists have many tools in their armoury for assisting in the early stages of a new project, and 3D database searching is now an essential part of the drug discovery process. To meet the challenges of the next millennium in the pharmaceutical industry, the time taken to find leads to begin research into new therapeutic areas must be as short as possible. 3D database searching offers one way of *getting that hit*.

The pharmaceutical industry has grown leaner and meaner in the past decade. Mergers, takeovers and rationalization have created enormous companies that under increased pressure to succeed. Alongside these changes have emerged small biotechnology companies who can rapidly initiate research into new (or established) therapeutic areas without the burden of an elephantine administration. At an appropriate juncture, these smaller companies seek to liaise with one of the pharmaceutical giants to develop their drug candidates. This is an attractive proposition for the larger company as they are buying into established research, typically at the beginning of clinical trials.

To make this prospect more appealing and to stay ahead, small companies often utilize novel research methods. Examples include structure-based drug design *via* X-ray crystallography, combinatorial chemistry, chiral expertise or combinations of these. Such companies also need to move fast. This means that access to novel drug molecules for testing and lead generation must be readily available. One way of doing this is *via* combinatorial chemistry or efficient searching methods to select compounds from 3D databases; or preferably a combination of the two. In order to use combinatorial chemistry effectively, one needs to be able to test the large number of molecules produced. Unfortunately, not all biological testing methods lend themselves to this approach and other complementary strategies need to be considered.

3D database searching allows the chemist to select either a diverse set of compounds or a biased set based either on known active compounds or from a knowledge of the binding site of known leads. This brief overview will cover some aspects of database searching and associated topics, as well as giving some practical information regarding software and databases. Detailed listings of databases, associated software, suppliers and Internet resources are provided in boxed text throughout the review, and addresses for suppliers are provided at the end of the article in Box 5 (superscript letters used in the text and boxes denote entries in the address listing).

2D – 3D

Chemical structures have traditionally been stored as 2D molecular depictions; also, when stereochemical information is included, the term '2.5D' is used[1]. In order to make effective use of collections of compounds for lead discovery, their electronic storage should include 3D structures. Research into methods of converting 2D databases into 3D has led to the development of a number of software packages[1–4] (Box 1). Much of this work has been devoted to obtaining reasonable 3D structures in as short a time as possible.

David T. Manallack, Chiroscience, Cambridge Science Park, Milton Road, Cambridge, UK CB4 4WE. tel: +44 1223 420430, fax: +44 1223 428678, e-mail: davidm@chirosci.demon.co.uk

  1359-6446/96/$15.00. PII: S1359-6446(96)10021-0

Box 1. 2D to 3D structure conversion programs*

Catalyst	Molecular Simulations Inc.[b]
Cobra	Oxford Molecular Ltd[f]
CONCORD	Tripos[d]
Converter	Molecular Simulations Inc.[b]
CORINA	Sadowski and Gasteiger[2,27] and Oxford Molecular Ltd[f]
Chem-X Builder	Chemical Design Ltd
Galaxy	AM Technologies Inc.[g]
Interchem	Interprobe Chemical Services[h]
RUBICON	Daylight CIS[i]

*For superscript numbers, see reference list; for superscript letters, see Box 5

Before conversion of a database from 2D to 3D, each molecule must be in a suitable format. Typically, this will take the form of SMILES strings[5], although other formats have been catered for (e.g. MACCS Mol files, MDL Information Systems[a]). The literature suggests that these packages will convert well over 90% of structures. If resources are available, the failed compounds can be put through a second conversion package, and so on. The remainder can be examined to determine reasons for their failure and, if the number of compounds is small, they may be modelled by hand using a suitable molecular modelling package. Some 2D to 3D conversion packages fail if the molecule is particularly 'exotic' or contains many ring systems. It has been argued that such compounds can be safely left out of the database because they may never make a good drug or inspire a chemist to begin lead optimization if the chemistry appears to be too difficult. The alternative view is that potentially useful information is being discarded. Continued improvements in this area should bring closer the desired 100% successful conversion rate.

Conformational flexibility

The need to be able to perform flexible searches stems from the fact that most molecules in compound collections possess rotatable bonds. The first studies to search 3D databases for compounds matching a pharmacophore model used databases and/or methods that examined single conformations of stored molecules[4,6]. Software solutions dealing with the problem of flexibility only began to emerge at the beginning of this decade[4] (Box 2). Two such methods, discussed below, involve either storage of multiple conformations for each compound or conformer generation at search time.

Multiple conformer storage has several drawbacks, including storage requirements and uncertainties about adequate coverage of conformational space. The 'Poling' technique devised by Smellie and coworkers[7] is better able to explore conformational space and has been implemented in the Catalyst (Molecular Simulations Inc.[b]) software[8]. Until recently, disk storage space has been a resource issue for molecular modelling, and methods that avoided storing large numbers of conformations were investigated. Chemical Design Ltd[c]

Box 2. Database and associated software*†

Name	Capability/Function	Supplier
Catalyst/Info/Hypo/DB Server	2D, 3D, Flexible	Molecular Simulations Inc.[b]
CAVEAT	Vector	Molecular Simulations Inc.[b,j]
ChemDBS-2D/3D	2D, 3D, Flexible	Chemical Design Ltd[c]
Iditis	Proteins	Oxford Molecular Ltd[f]
Ingres	Relational	Computer Associates International Inc.[k]
ISIS/Base, ISIS/Host, ISIS/3D	2D, 3D, Flexible	MDL Information Systems Inc.[a]
MACCS II	Relational	MDL Information Systems Inc.[a]
Oracle	Relational	Oracle[l]
REACCS	Reaction	MDL Information Systems Inc.[a]
SYBYL/UNITY	2D, 3D, Flexible	Tripos[d]
Thor/Merlin	2D, 3D, Flexible	Daylight CIS Inc.[i]

*For superscript letters, see Box 5

†This information has been extracted from a list of software for Unix workstations compiled by Silicon Graphics in August 1994

Box 3. Databases of compounds available for searching*†‡

Database	Size/Dimension	Supplier
Available Chemicals Directory (ACD) – 3D	160K/3D	MDL Information Systems Inc.[a]
Beilstein Crossfire	6,000K/2D	Beilstein[m]
BioByte Masterfile Database	24K/3D	Molecular Simulations Inc.[b]
Brookhaven Protein Databank	4.2K/3D	Brookhaven National Laboratory[n]
Cambridge Crystallographic Database	107K/3D	Cambridge Crystallographic Data Centre[o]
CAST-3D	up to 370K/3D	American Chemical Society[p]
Chapman and Hall Chemicals	105K/3D	Chemical Design Ltd[c]
Chapman and Hall Drugs	16K/3D	Chemical Design Ltd[c]
Chapman and Hall Natural Products	69K/3D	Chemical Design Ltd[c]
Cipsline	24K/2D	Prous[q]
Comprehensive Medicinal Chemistry – 3D	6.5K/3D	MDL Information Systems Inc.[a]
Derwent World Drug Index	41K/3D	Chemical Design Ltd[c], MSI[b]
Drugs of the Future	1.3K/2D	Prous[q]
Flavour and Fragrance Materials	2.5K/3D	Chemical Design Ltd[c]
HTS chemicals on CD ROM	210K/3D	Chemical Design Ltd[c]
ILIAD, TRIAD databases	11K and 40.5K, resp./Vector	Molecular Simulations Inc.[b,j]
Index Chemicus	290K/2D	ISI Inc.[r]
Maybridge	47K/3D	Daylight CIS[i], Tripos[d] and others
Metabolite	16.3K/2D	MDL Information Systems Inc.[a]
MDL Drug Data Report – 3D	57K/3D	MDL[a], Prous[q]
MedChem 95	28.5K/2D	Daylight CIS[i]
Pomona Masterfile	24K/3D	Daylight CIS[i]
NCI database	126K/3D	MSI[b], Tripos[d], MDL[a] and others
New Research Intermediates	1.6K/3D	Chemical Design Ltd[c]
NME Express	Variable/2D	Prous[q]
SPRESI 95	3,200K/3D	Daylight CIS[i]
TSCA 93	100K/2D	Daylight CIS[i]
World Drug Index 95	41K/3D	Daylight CIS[i]

*For superscript letters, see Box 5

†K = 1,000

‡Based on a list of databases assembled by Silicon Graphics in August 1994 and updated to include recent releases. In some cases, third-party resellers of databases are listed who may not have compiled the original information

introduced the use of a set of distance keys[9], which described the possible distances between defined atom (or group) centres derived from rule-based conformational searches. The 3D keys act as a filter to remove structures that cannot fit the 3D query. Conformations of the structures that pass this filter are regenerated to determine which conformer fits the query; these are then stored for further examination[9]. When this software was used to search the *Derwent Standard Drug File* for thromboxane synthase inhibitors, 53 of the 63 compounds were recognized as inhibitors of the enzyme[4]. Of the ten inhibitors not identified by the search, nine did not have the correct nitrogen atom type and the remaining compound did not physically fit the pharmacophore.

The alternative searching method developed recently uses a torsional minimization of candidate molecules to adjust rotatable bonds to fit a 3D search query[10,11]. Some researchers have argued that the method of storing 3D distance keys described above may miss some important conformations, because these were created using a rule-based technique (e.g. the default increment for single bonds is 120°, although this parameter is adjustable). Most software companies now offer the torsional fitting method [e.g. ChemDBS-3D (flexifit)[c], SYBYL/UNITY[d], ISIS/3D[a], MACCS-II/3D[a]; see Box 2].

Mason conducted a comparative study of the different software packages on a database of 58,000 3D structures[3]. His investigations showed that many factors influence the

number of hits obtained from a series of searches based on simple 3D pharmacophores. Both the rule-based conformation technique and the torsional minimization technique have advantages and disadvantages and, in general, these methods were found to be comparable. The torsional minimization method has been shown to be fast and better able to handle extremely flexible molecules (> 14 rotatable bonds) using small query distance constraints[3,4]. The limitation of the method is that it is not systematic and can only produce a few (or, in some cases, one) possible conformational solution(s) to a 3D query. The rule-based method, on the other hand, will provide all possible mappings to the 3D search. As each of these techniques has its own useful features, it is very difficult to state that one method is better than another, because this will depend to a large extent on the nature of the 3D search query[3,4].

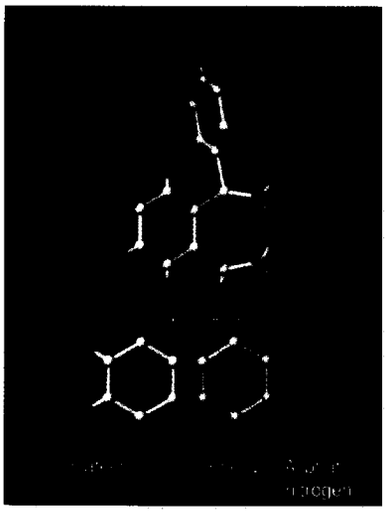

Figure 1. *Example of a 2D search based on a lead compound. The top structure depicts SKF 38393, a potent dopamine receptor (D_1) agonist. The 2D search, below, is aimed at finding compounds containing all of the elements listed: catechol, phenyl ring and aliphatic nitrogen.*

Search strategies

In the early stages of a drug discovery programme, researchers may be faced with little or no SAR information regarding suitable drug candidates. At this point, random screening may be used to find leads. Typically, the medicinal chemistry team will choose a set of compounds from their in-house collection, as well as screening commercially available compounds. Some companies have recently emerged who specialize in providing sets of compounds suitable for screening (e.g. SPECS and BioSPECS[c]; see Boxes 3 and 4 for detailed listings of databases and Internet resources). Choice may be random or involve the calculation of a variety of physicochemical properties for each compound to assist in assessing the molecular diversity of the set. Statistical techniques can be applied to choose a subset of compounds to span the parameters used to describe the molecules. The aim of these analyses is to select and test fewer compounds, whilst gaining as much information as possible about the dataset. Each molecule chosen should therefore represent a family of compounds from the entire set. Neither random selection nor choices based on molecular diversity can be considered to be right or wrong, and the topic of molecular diversity is deserving of a review in its own right. If, however, a lead is known, then a more directed approach can be adopted by searching (2D or 3D) for compounds with similar structures to the lead candidate (Figure 1). As more bioactive compounds are found, or information becomes available from the literature, pharmacophore models can be constructed to facilitate 3D database searching.

Pharmacophore development can either be conducted by classical molecular modelling techniques or through the use of automated pharmacophore searching software. The automated approach has the advantage of removing some of the bias that a molecular modeller may impart when seeking to create a pharmacophore model (for reviews, see Refs 12 and 13). Automated approaches can suggest a number of models, many of which can be discarded as nonsensical, leaving reasonable models requiring validation. Test compounds with known biological activities can be fitted to each model in turn to determine if it can rationalize the activities of these compounds. Alternative validation strategies may involve database searching to ensure that the model picks out known bioactive compounds; or validation may employ a quantitative approach, such as the application of a 3D QSAR technique.

A recent example of the generation and use of a pharmacophore model to search a 3D database involved the examination of a series of dopamine (D_1) receptor agonists. The original impetus for the work was to compare a pharmacophore derived using classical molecular modelling techniques[14] (Figure 2a) with an automated pharmacophore generation package (Catalyst, Molecular Simulations Inc.[b]). Remarkable similarity was found between the original pharmacophore model and the best hypothesis derived by Catalyst[15] (Figure 2b). This hypothesis was used to search the BioByte Masterfile database (Molecular Simulations Inc.[b]) for compounds fitting this model, and a number of known and selective D_1 agonists were identified in this way.

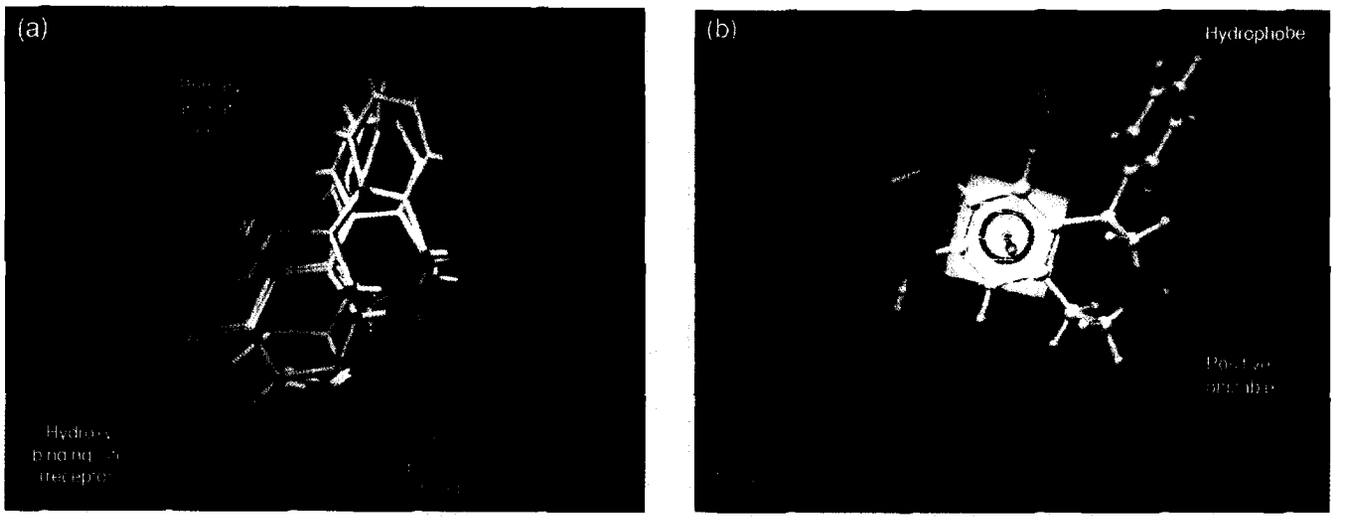

Figure 2. *(a) Key features of the dopamine receptor (D_1) agonist pharmacophore model derived using classical molecular modelling methods[14]. (b) Pharmacophore hypothesis generated with the Catalyst program using a similar set of D_1 agonists[15]. SKF 38393 is shown in the model as a representative agonist demonstrating the fit to the key pharmacophoric elements.*

Two other successful examples of 3D database searching have been described by Ashton and coworkers[16]. The lead dibenzamide compound in their low-density lipoprotein receptor upregulator programme had weak activity (2 μM) and could not be improved upon. 3D and 3D-flexible searches identified a number of compounds from their corporate database, one of which had an activity of 6 nM (Ref. 16). In their endothelin receptor antagonist programme, 3D searches were based on two potent antagonists, and a series of hit compounds was found giving the chemists suitable leads. Evidence demonstrating that the 3D relationship of the pharmacophoric elements used in the search were important came following the screening of 400 additional compounds. These compounds were related to the 3D hits by their 2D structure and failed to show any activity[16].

One factor that will influence a pharmaceutical company that is considering initiating a drug discovery programme in a therapeutic area is the availability of an X-ray structure of the target biosystem, preferably one with a bound ligand. This so-called structure-based drug design gives a direct insight into the nature of the target binding site. Design of molecules can be based on this site or use the geometry of bound ligands. Following synthesis and testing, the ligands themselves can be crystallized into the protein and the structure determined. The resultant complex can be compared to predicted modes of binding and, using this iterative approach, potent and selective ligands can be developed[17].

In the absence of X-ray crystallographic facilities, the binding site of a published target structure can be used as a constraint for database searching. This form of search has been implemented in the DOCK program[18]. A negative image of the binding site (Figure 3) is calculated by filling the site with overlapping spheres. Ligands are 'docked' by matching their structure to the sphere centres and a score is produced on the resultant fit. This score can be based on steric and/or electrostatic interactions and recent improvements have been made to perform rigid body minimizations[19]. An example of the use of DOCK to identify novel ligands for a binding site came from a study of the HIV enzyme showing that haloperidol could fit the binding site[20]. Subsequent testing of haloperidol demonstrated modest affinity for the enzyme. The crystal structure of a related haloperidol analogue bound to the HIV-1 enzyme has been determined and shows that a conformational change occurred in the binding site and that the ligand did not bind as DOCK predicted[21]. Most searching software is now able to exploit binding-site information for database mining and considerable efforts are being invested in this area (see Ref. 4).

Several other methods have been developed for database searching, but for the sake of brevity only two of these are discussed here. The first, SPERM (Ref. 22), employs a search that finds molecules matching the shape of known compounds.

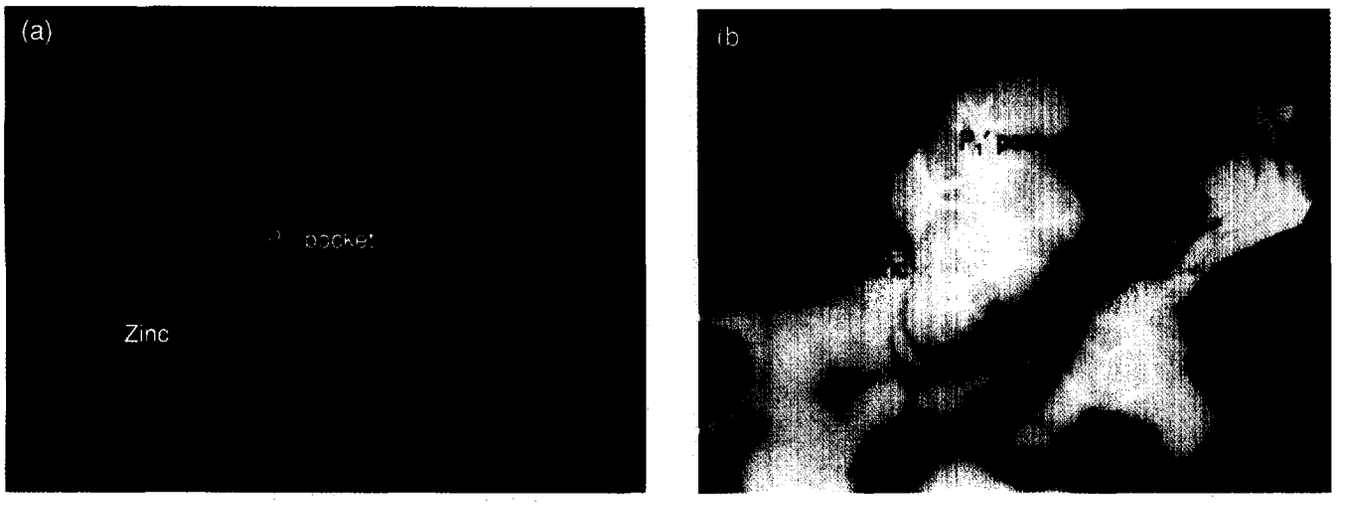

Figure 3. *(a) Representation of the surface of the catalytic site of human fibroblast collagenase (1HFC). (b) Negative image of the catalytic site of human fibroblast collagenase clearly showing the protruding shape of the P_1' pocket.*

This method, which has been compared to various molecular similarity techniques, employs a gnomonic projection of molecular shape onto a sphere. This allows for rapid searching to compare a lead candidate with a database of compounds. CAVEAT[23] was developed to provide database searching for vector-based 3D queries, the vectors in this case being the bonds between selected atoms. This has proved to be of use in finding rigid nonpeptidic backbone structures to mimic peptides.

Experience at Chiroscience

The databases at Chiroscience have been successfully used to assist each of the drug discovery programmes. The approach, as with any database search, has been to sift through many thousands of compounds to find a small subset as potential leads that fit our requirements. The simplest example described here employed 2D searching for compounds containing combinations of particular functional groups based on lead candidates. Using the lists of compounds emerging from these searches, it was possible to select a number of compounds for biological testing.

The medicinal chemistry and molecular modelling groups work closely together at Chiroscience to increase the efficiency of the lead generation process. Through this approach, low micromolar hits were found for a phosphodiesterase (PDE-IV) programme[24]. In summary, 20 compounds were purchased for testing, four of which were shown to have activity of less than 10 μM. This enabled the chemists to make a choice of leads for optimization on the basis of various criteria, including patentability and synthetic considerations. Also, at this point, bias may be directed at molecules that lend themselves to lead optimization *via* combinatorial chemistry; this approach has been successfully achieved within Chiroscience. Information from the literature, together with the results of in-house testing, allowed generation of a pharmacophore model for the PDE-IV binding site. This model has been used to perform 3D and 3D-flexible searches to identify further lead compounds.

In addition to 2D and 3D searching, Chiroscience have used the binding site(s) of matrix metalloproteases[25] to search for potential inhibitors. The UNITY system[d] allows points in space to be defined, which are used as exclusion areas for the database search. While this does not allow a perfect representation of the binding site, the approximation proved to be useful and hit structures may be viewed 'docked' into the binding site. Such lists of molecules can be quickly reduced in size by the molecular modeller in conjunction with a medicinal chemist.

From this approach, using 3D-flexible searching, we have found low micromolar, nonpeptidic hits that have initiated new areas of chemistry. In this study, 11 compounds were selected for testing, resulting in two hits with activities in the range 1–10 μM. Furthermore, each hit was selected from families of structures and these related compounds were also purchased for testing. For the first compound, six related molecules were tested, resulting in three hits with similar activity.

Box 4. Internet resources

This list, although not exhaustive, provides examples of Internet sites with information regarding databases and associated software. The four World Wide Web (WWW) sites worthy of investigation listed below provide excellent summaries of available databases and molecular modelling software.

•National Institutes of Health (NIH): http://molbio.info.nih.gov/modeling/
•Computational Chemistry List (CCL): http://ccl.osc.edu/chemistry.html
•NetSci: http://www.awod.com/netsci. Also associated with this site is a monthly series of articles concerning different topics relating to R&D in the pharmaceutical industry. The September issue of NetSci was dedicated to 3D database searching.
•CAOS/CAMM: http://www.coas.kun.nl/b_pack/camm.html
Information regarding X-ray crystal or NMR structures of proteins can be found at the Brookhaven WWW site: http://www.pdb.bnl.gov/. This site provides an excellent service for searching and obtaining 3D protein structures.

Small molecule X-ray crystal data can be found at the Cambridge Crystallographic Data Centre: http://csdvx2.ccdc.cam.ac.uk/. Information concerning the database and CSD software is given, and details are provided about database searching (this is free to academic groups).

Each of the major suppliers of software or databases also have WWW sites. One interesting site is maintained by the Daylight corporation[i] listing their services: http://www.daylight.com/. An associated site – http://fox.pomona.claremont.edu/chem/Daylight/index.html – can be used to access the Mjollnir e-mail system which can be used for searching Daylight's chemical databases. This service is free but there are restrictions on the frequency of use.

Box 5. Supplier addresses

a MDL Information Systems Inc., 14600 Catalina St, San Leandro, CA 94577, USA
b Molecular Simulations Inc., 16 New England Executive Park, Burlington, MA 01803, USA
c Chemical Design Ltd, Roundway House, Cromwell Park, Chipping Norton, Oxfordshire UK OX7 5SR
d Tripos, 1699 South Hanley Rd, St Louis, MO 63144, USA
e SPECS and BioSPECS, B.V., Fleminglaan 16, 2289 Rijswijk, The Netherlands
f Oxford Molecular Ltd, The Medawar Centre, Oxford Science Park, Oxford, UK OX4 4GA
g AM Technologies Inc., 5 Great Valley Parkway, GV Corporate Center, Suite 269, Malvern, PA 19355, USA
h Interprobe Chemical Services, Lenzie, Glasgow, UK G66 4HX
i Daylight CIS Inc., 18500 Von Karman Ave., #450 Irvine, CA 92715, USA
j Dr Paul Bartlett, Department of Chemistry, University of California, Berkeley, CA 94720, USA
k Computer Associates International Inc., One Computer Associates Plaza, Islandia, NY 11788-7000, USA
l Oracle, Oracle Centre, The Ring, Bracknell, Berkshire, UK RG12 1BW
m Beilstein Informationssysteme GmbH, Varrentrapp Strasse 40–42, D-60846, Frankfurt, Germany
n Brookhaven National Laboratory, Chemistry Department, Bld 555, PO Box 5000, Upton, NY 11973-5000, USA
o Cambridge Crystallographic Data Centre, 12 Union Rd, Cambridge, UK CB2 1EZ
p American Chemical Society, PO Box 3337, Columbus, OH 43210, USA
q J.R. Prous S.A., PO Box 540, 08080 Barcelona, Spain
r ISI Inc., Brunel Science Park, Brunel University, Uxbridge, UK UB8 3PQ,
s Maybridge Chemical Company, Trevillett, Tintagel, Cornwall, UK PL34 0HW
t Bionet Research Ltd, 3A Highfield Industrial Estate, Camelford, Cornwall, UK PL32 9QZ
u Menai Organics Ltd, Unit 5, Menai Technology Centre, Deiniol Rd, Bangor, Gwynedd, UK LL57 2UP
v Receptor Research, Unit 5, Cory Close, Wainhouse Corner, Bude, Cornwall, UK EX23 0AR
w Peakdale Fine Chemicals Ltd, 7 Brookfield Industrial Estate, Glossop, Derbyshire, UK SK13 9LQ

Six of the 15 compounds relating to the second hit had activities below 10 μM, the best being 500 nM. This approach allowed rapid build up of SAR to direct future medicinal chemistry efforts. Each compound may also be docked into the enzyme active site to determine whether the SAR information is consistent with the hypothetical mode of binding. If the data make sense, then molecular modelling can suggest appropriate substitution to be made to the parent molecule or allow *de novo* drug design packages[26] to help suggest modifications to improve activity and/or selectivity.

Conclusions

Database searching has been demonstrated by a number of groups to suggest not only active lead compounds, but also molecules appropriate for exploitation using classical lead optimization strategies or combinatorial chemistry. If improvements are to be made to database searching software, those researchers involved in drug discovery need to outline their requirements carefully. 3D database searching will improve as search paradigms become refined in terms of speed and adequate coverage of conformational space.

Hardware advances should also speed up search and access times. Further advances may come when alternative representations are used that deviate from describing molecules as a collection of geometric points in space. Binding-site models that include shape and molecular fields may allow us to get closer to reality, so that potential lead molecules are not missed in the quest for *getting that hit.*

Other useful information

The NCI database is available *via* anonymous FTP from *helix.nih.gov* in the directory *ncidata/3D*. This directory contains a compressed version of the database produced by Drs Jens Sadowski and Johann Gasteiger (Universität Erlangen-Nuernberg) using the CORINA software[2,28]. The NCI also publishes the results of biological screening; two databases of compounds are available with results for anticancer and anti-AIDS activity (20,000 and 25,000 compounds, respectively). These are available from: http://epnws1.ncifcrf.gov:2345/dis3d/dtp.html as compressed MDL SDF format files. The ACD database[a] contains listings from numerous suppliers. The addresses of three companies[s,t,u] listed in the ACD who release their database free of charge are given in Box 5. Three companies not listed in the ACD who stock compounds suitable for screening by the pharmaceutical industry also provide their databases free of charge[e,v,w].

In addition to the databases listed in Boxes 1–5, two World Wide Web sites on the Internet provide small databases of compounds: http://ibc.wustl.edu/klotho/ and http://ws05.pc.chemie.th-darmstadt.de/vrml/pdbvis.html.

ACKNOWLEDGEMENTS

I thank Drs A. Good, J. Mason, J. Montana, S. Beasley, J. Davies, R. Scott, A. Richards and J. Padfield for helpful discussions.

REFERENCES

1 Pearlman, R.S. (1993) in *3D QSAR in Drug Design: Theory, Methods and Applications* (Kubinyi, H., ed.), pp. 41–70, ESCOM
2 Sadowski, J. and Gasteiger, J. (1993) *Chem. Rev.* 93, 2567–2581
3 Mason, J.S. (1995) in *Molecular Similarity in Drug Design* (Dean, P.M., ed.), pp. 138–162, Blackie Academic and Professional
4 Good, A.C. and Mason, J.S. (1995) in *Reviews in Computational Chemistry*, Vol. 7 (Boyd, D.B. and Lipkowitz, K.B., eds), pp. 67–117, VCH
5 Weininger, D. (1988) *J. Chem. Inf. Comput. Sci.* 28, 31–36
6 Gund, P. (1977) in *Progress in Molecular and Subcellular Biology*, Vol. 5 (Hahn, F.E., ed.), pp 117–143, Springer-Verlag
7 Smellie, A., Teig, S.L. and Towbin, P. (1995) *J. Comput. Chem.* 16, 171–187
8 Sprague, P.W. (1991) in *Recent Advances in Chemical Information – Proceedings of the 1991 Chemical Information Conference* (Collier, H., ed.), pp. 107–111, Royal Society of Chemistry, Cambridge
9 Murral, N.W. and Davies, E.K. (1990) *J. Chem. Inf. Comput. Sci.* 30, 312–316
10 Hurst, T. (1994) *J. Chem. Inf. Comput. Sci.* 34, 190–196
11 Moock, T.E. *et al.* (1994) *J. Chem. Inf. Comput. Sci.* 34, 184–189
12 Wermuth, C-G. and Langer, T. (1993) in *Drug Design: Theory, Methods and Applications* (Kubinyi, H., ed.), pp. 117–136, ESCOM
13 Golender, V.E. and Vorpagel, E.R. (1993) in *3D QSAR in Drug Design: Theory, Methods and Applications* (Kubinyi, H., ed.), pp. 137–149, ESCOM
14 Manallack, D.T. (1994) in *Molecular Modelling and Drug Design* (Vinter, J.G. and Gardner, M., eds), pp. 333–376, Macmillan Press
15 Scott, R.K., Hoffmann, R.D. and Manallack, D.T. (1995) paper presented at the SCI conference on the Design of Bioactive Compounds, September 1995, Potsdam (in preparation)
16 Ashton, M.J., Jaye, M.C. and Mason, J.S. (1996) *Drug Discovery Today* 1, 71–78
17 Erion, M.D. *et al.* (1993) *J. Med. Chem.* 36, 3771–3783
18 Kuntz, I.D. *et al.* (1982) *J. Mol. Biol.* 161, 269–288
19 Bodian, D.L. *et al.* (1993) *Proteins* 17, 266–276
20 Dejarlais, R.L. *et al.* (1990) *Proc. Natl Acad. Sci. USA* 87, 6644–6648
21 Rutenber, E. *et al.* (1993) *J. Biol. Chem.* 268, 15343–15346
22 Perry, N.C. and van Geerestein, V.J. (1992) *J. Chem. Inf. Comput. Sci.* 32, 607–616
23 Lauri, G. and Bartlett, P. (1994) *J. Comput.-Aided Mol. Design* 8, 51–66
24 Lowe, J.A. and Cheng, J.B. (1992) *Drugs Future* 17, 799–807
25 Stocker, W. *et al.* (1995) *Protein Sci.* 4, 823–840
26 Muller, K. (ed.) (1995) *Perspect. Drug Design* 3, issue dedicated to *de novo* drug design
27 Sadowski, J., Gasteiger, J. and Klebe, G. (1994) *J. Comp. Inf. Comput. Sci.* 34, 1000–1008