JOURNAL OF
MEDICINAL
CHEMISTRY



Volume 35, Number 12 June 12, 1992

*Perspective*

# 3D Database Searching in Drug Design

Yvonne C. Martin

*Computer Assisted Molecular Design Project, Pharmaceutical Products Division, Abbott Laboratories, Abbott Park, Illinois 60064.* *Received August 2, 1991*

## Introduction

Molecular modeling[1] and protein structure determination[2] are now often a part of a medicinal chemistry investigation. However, experience over the last decade has shown that, for the computer design of structurally novel molecules, we need tools beyond molecular graphics. Since 1980 3D QSAR methods for the quantitative prediction of potency based on 3D properties have been developed.[3] 3D searching[4] provides other needed capabilities—it designs or recognizes potential bioactive molecules based on their 3D properties. Additionally, several programs use 3D searching to design the molecules to synthesize.[4-8] Of course, the exact molecules to be made also will be governed by ease of synthesis and projected physical properties.

One use of 3D database searching helps one design novel compounds that incorporate conformational constraints. Such compounds might mimic the bioactive conformation of a ligand as established by experiment or molecular modeling. For example, molecular modelers often are asked if the computer would design morphine from enkephalin. A computer design of mimics of the tyrosine in enkephalin suggested a morphine analog 1 as a mimic of one conformation of enkephalin.[9] As discussed below, the designed compounds might also be used to derive or test a pharmacophore model.

3D searching also identifies existing molecules that match a hypothesis of the 3D requirements for bioactivity. It thus can be used to validate such pharmacophores and to suggest other existing compounds for testing to find a

(1) Cohen, N. C.; Blaney, J. M.; Humblet, C.; Gund, P.; Barry, D. C. Molecular Modeling Software and Methods for Medicinal Chemistry. *J. Med. Chem.* **1990**, *33*, 883–984.

(2) (a) Goodford, P. J. Drug Design by the Method of Receptor Fit. *J. Med. Chem.* **1984**, *27*, 557–564. (b) Appelt, K.; Bacquet, R. H.; Bartlett, C. A.; Booth, C. L. J.; Freer, S. T.; Fuhry, M. A. M.; Gehring, M. R.; Herrmann, S. M.; Howland, E. F.; Janson, C. A.; Jones, T. R.; Kan, C-C.; Kathardekar, V.; Lewis, K. K.; Marzoni, G. P.; Matthews, D. A.; Mohr, C.; Moomaw, E. W.; Morse, C. A.; Oatley, S. H.; Ogden, R. C.; Reddy, M. R.; Reich, S. H.; Schoettlin, W. S.; Smith, W. W.; Varney, M. D.; Villafranca, J. E.; Ward, R. W.; Webber, S.; Webber, S. E.; Welsh, K. M.; White, J. Design of Enzyme Inhibitors Using Iterative Protein Crystallographic Analysis. *J. Med. Chem.* **1991**, *34*, 1925–1934. (c) Fesik, S. W.; Gampe, R. T.; Eaton, H. L.; Gemmecker, G.; Olenejniczak, E. T.; Neri, P.; Holzman, T. NMR Studies of [U-C-13]Cyclosporin-A Bound to Cyclophilin-Bound Conformation and Portions of Cyclosoprine Involved in Binding. *Biochemistry* **1991**, *30*, 6574–6583.

(3) (a) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *170*, 5959–5967. (b) Hopfinger, A. J. A QSAR Investigation of Dihydrofolate Reductase Inhibition by Baker Triazines Based upon Molecular Shape Analysis. *J. Am. Chem. Soc.* **1980**, *102*, 7196–7206. (c) Boulu, L. G.; Crippen, G. M.; Barton, H. A.; Kwon, H.; Marletta, M. A. Voronoi Binding Site Model of a Polycyclic Aromatic Hydrocarbon Binding Protein. *J. Med. Chem.* **1990**, *33*, 771–775. (d) Kati, I.; Itai, A.; Iitaka, Y. A Novel Method for Superimposing Molecules and Receptor Mapping. *Tetrahedron* **1987**, *43*, 5229–5236. (e) Doweyko, A. M. The Hypothetical Active Site Lattice. An Approach to Modelling Active Sites from Data on Inhibitor Molecules. *J. Med. Chem.* **1988**, *31*, 1396–1406.

(4) Martin, Y. C.; Bures, M. G.; Willett, P. Searching Databases of Three-Dimensional Structures. In *Rev. in Computational Chemistry*; Lipkowitz, K., Boyd, D., Eds.; VCH Publishers, Inc.: New York, 1990; pp 213–263.

(5) Martin, Y. C. Computer-Aided Design of Potentially Bioactive Molecules by Geometric Searching with ALADDIN. *Tetrahedron Comput. Methodol.* **1990**, *3*, 15–25.

(6) Martin, Y. C.; Van Drie, J. H. Identifying Unique Core Molecules From the Output of a 3D Database Search. In *Proceedings of the 2nd International Conference on Chemical Information, Noordwijkerhout, 1990*, Warr, W., Ed., Springer-Verlag.

(7) Moon, J.; Howe, W. J. Computer Design of Bioactive Molecules: A Method for Receptor-Based de novo Ligand Design. *Proteins: Struct. Funct. Genet.* **1991**, *6*, 314–328.

(8) Böhm, H. J. A New Method for the De Novo Design of Enzyme Inhibitors. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 61–78.

(9) Martin, Y. Unpublished observations.

0022-2623/92/1835-2145$03.00/0 

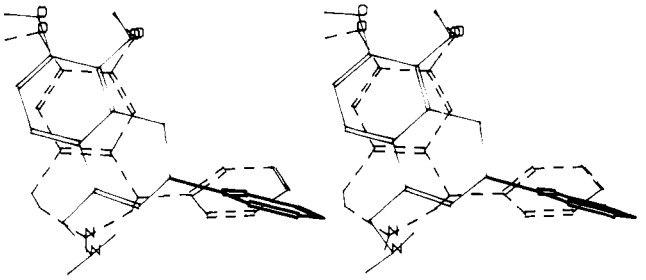


**Figure 1.** The bioactive conformation of the D1 agonist SKF 38393 (dashed lines), the database molecule shown to have dopaminergic activity (solid fine lines), and the location of the added phenyl in the D1 selective compound designed from it (heavy lines).[34]

new lead. Examples of this will also be discussed.

The potential of 3D database searching was recognized years ago,[10-12] yet only recently have several operational systems been implemented.[4,13-28] The current interest in 3D database searching was fueled by the availability of tools for molecular modeling[1] and pharmacophore mapping[29] and by the increasing numbers of 3D protein structures as targets for new drugs.[2] The extensive database of crystallographic structures of small molecules[28] and computer programs that generate 3D structures of small molecules in a few seconds[30-32] typically provide the information to search. 3D searching is complementary to 3D QSAR[3] since it can be used to design series for 3D QSAR analysis and 3D QSAR can be used to rank compounds suggested for synthesis or testing by 3D searching.[33]

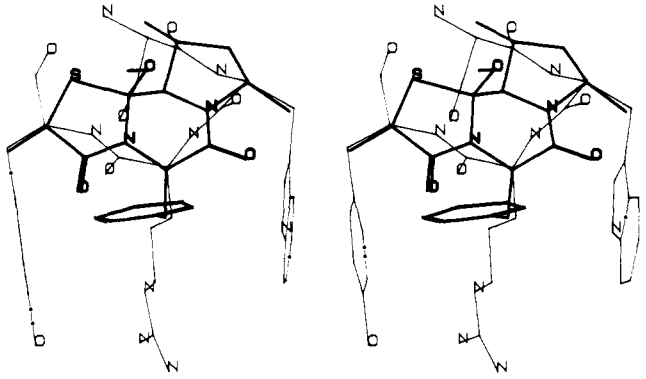


**Figure 2.** The conformation of the proposed bioactive conformation of residues 18–20 (the binding loop) of tendamistat (fine lines) and that of a molecule that has side chains in the geometric relationships (heavy lines). All three side chains could be incorporated as well as the N-terminal portion of the peptide.[21]


(10) (a) Gund, P.; Wipke, W. T.; Langridge, R. Computer Searching of a Molecular Structure File for Pharmacophoric Patterns. *Comput. Chem. Res., Educ., Technol.* **1974**, *3*, 5–21. (b) Gund, P. Three-dimensional Pharmacophoric Pattern Searching. In *Prog. Mol. Subcell. Biol.* **1977**, *11*, 117–143.

(11) Esaki, T. Quantitative Drug Design Studies. V. Approach to Lead Generation by Pharmacophoric Pattern Searching. *Chem. Pharm. Bull.* **1982**, *30*, 3657–3661.

(12) Esaki, T. Approach to Lead Generation by Three-dimensional Pharmacophoric Pattern Searching: Development of a Graphic Module for Prediction of Pharmacological Features of Drugs, unpublished report, 1983.

(13) Borman, S. Software Adds New Dimension to Structure Searching. *Chem. Eng. News* **1989**, *67* (Sept 18), 28–32.

(14) Van Drie, J. H.; Weininger, D.; Martin, Y. C. ALADDIN: An Integrated Tool For Computer-Assisted Molecular Design and Pharmacophore Recognition from Geometric, Steric, and Substructure Searching of Three-Dimensional Molecular Structures. *J. Comput.-Aided Mol. Des.* **1989**, *3*, 225–251.

(15) Jakes, S. E.; Watts, N.; Willett, P.; Bawden, D.; Fisher, J. D. Pharmacophoric Pattern Matching in Files of Three-dimensional Chemical Structures: Evaluation of Search Performance. *J. Mol. Graphics* **1987**, *5*, 41–48.

(16) Jakes, S. E.; Willett, P. Pharmacophoric Pattern Matching in Files of Three-dimensional Chemical Structures. Selection of Inter-atomic Distance Screens. *J. Mol. Graphics* **1986**, *4*, 12–20.

(17) Brint, A. T.; Willett, P. Pharmacophoric Pattern Matching in Files of 3-D Chemical Structures: Comparison of Geometric Searching Algorithms. *J. Mol. Graphics* **1987**, *5*, 50–56.

(18) DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. Using Shape Complementarity as an Initial Screen in Designing Ligands for a Receptor Binding Site of Known Three-dimensional Structure. *J. Med. Chem.* **1988**, *31*, 722–729.

(19) Lewis, R. A.; Dean, P. M. Automated Site-directed Drug Design: The Concept of Spacer Skeletons for Primary Structure Generation. *Proc. R. Soc. London B* **1989**, *236*, 185–140.

(20) Lewis, R. A.; Dean, P. M. Automated Site-directed Drug Design: The Formation of Molecular Templates in Primary Structure Generation. *Proc. R. Soc. London B* **1989**, *236*, 141–162.

(21) Bartlett, P. A.; Shea, G. T.; Telfer, S. J.; Waterman, S. CAVEAT: A Program to Facilitate the Structure-derived Design of Biologically Active Molecules. In *Molecular Recognition: Chemical and Biological Problems*; Roberts, S. M., Ed.; Royal Society of Chemistry: London, 1989; Vol. 78, pp 182–196.

(22) Sheridan, R. P.; Venkataraghavan, R. Designing Novel Nicotinic Agonists by Searching a Database of Molecular Shapes. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 243–256.

(23) Rusinko, A., III; Sheridan, R. P.; Nilakantan, R.; Haraki, K. S.; Bauman, N.; Venkataraghavan, R. Using CONCORD to Construct a Large Database of Three-Dimensional Coordinate from Connection Tables. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 251–255.

(24) Sheridan, R. P.; Rusinko, A., III; Nilakantan, R.; Venkataraghavan, R. Searching for Pharmacophores in Large Coordinate Databases and Its Use in Drug Design. *Proc. Natl. Acad. Sci.* **1989**, *86*, 8125–8169.

(25) Sheridan, R. P.; Nilakantan, R.; Rusinko, A., III; Bauman, N.; Haraki, K. S.; Venkataraghavan, R. 3DSEARCH, A System for Three-dimensional Substructure Searching. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 255–260.

(26) Christie, B. D.; Henry, D. R.; Güner, O. F.; Moock, T. E. MACCS-3D: A Tool for Three-dimensional Drug Design. *Online Inf. 90*, **1990**, 11–13 Dec, 137–161.

(27) Murrall, N. W.; Davies, E. K. Conformational Freedom in 3-D Databases. 1. Techniques. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 312–316.

(28) Allen, F. H.; Davies, J. E.; Galloy, J. J.; Johnson, O.; Kennard, O.; Macrea, C. F.; Mitchell, E. M.; Mitchell, G. F.; Smith, J. M.; Watson, D. G. The Developments of Versions 3 and 4 of the Cambridge Database System. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 187–204.

(29) Martin, Y. C. Overview of Concepts and Methods in Computer-Assisted Rational Drug Design. *Methods Enzymol.* **1991**, *203*, 587–613.

(30) Pearlman, R. S.; Rusinko, A., III; Skell, J. M.; Balducci, R.; McGarity, C. M. CONCORD, Distributed by Tripos Associates, Inc., 1699 S. Hanley Road, Suite 303, St. Louis, MO 63944.

(31) Wipke, W. T.; Hahn, M. A. AIMB: Analogy and Intelligence in Model Building. System Description and Performance Characteristics. *Tetrahedron Comput. Methodol.* **1988**, *1*, 141.

(32) Dolata, D. P.; Leach, A. R.; Prout, K. WIZARD: AI in Conformational Analysis. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 73–85. Leach, A. R.; Dolata, D. P.; Prout, K. Automated Conformational Analysis and Structure Generation: Algorithms for Molecular Perception. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 31–16. Leach, A. R.; Prout, K.; Dolata, D. P. Automated Conformational Analysis: Algorithms for the Efficient Construction of Low-energy Conformations. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 271–282. Leach, A. R.; Prout, K.; Dolata, D. P. The Application of Artificial Intelligence to the Conformational Analysis of Strained Molecules. *J. Comput. Chem.* **1990**, *11*, 680–693. Leach, A. R.; Prout, K. Automated Conformational Analysis: Directed Conformational Search Using the A* Algorithm. *J. Comput. Chem.* **1990**, *11*, 1193–1205.

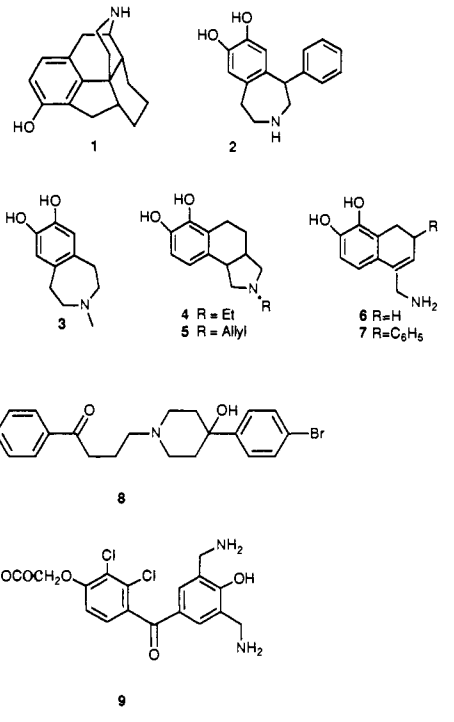
This review will summarize 3D searching from the viewpoint of a medicinal chemist. Emphasis will be methods that search many 3D structures.

## How Has 3D Searching Been Used in Medicinal Chemistry?

Two early 3D searching programs were written to help solve a problem familiar to medicinal chemists,[14-21] that is, how to design a novel structure that matches 3D requirements.

For example, by a synergism of molecular modeling and synthesis of conformationally informative molecules, we established that the bioactive conformation of the D1 dopamine agonist SKF 38393 (2) is that with the phenyl group equatorial (Figure 1).[34] How does one design a mimic that incorporates the N, O, and two phenyls in the proper geometric relationship? Structures imagined with molecular graphics often did not match well enough once built or were a high-energy conformation. Since we were already storing 3D coordinates in a chemical information database,[35] we wrote a computer program for 3D searching of these coordinates.[14]

Our search of the 3D structures of existing compounds identified 3–6 as dopamine D2 agonists. Molecular modeling suggested where to substitute the pendent phenyl (Figure 1). Whereas 6 has a $pK_i$ for the D1 receptor of 4.88, 7 is a D1 selective agonist with a $pK_i$ 6.82.[36] Thus 3D searching identified the lead and molecular modeling identified the correct derivative for synthesis.



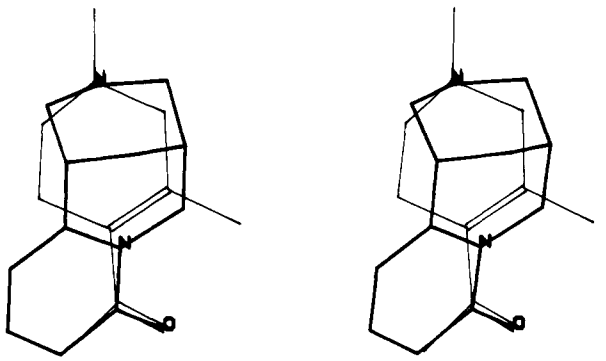



**Figure 3.** A molecule used to derive the nicotinic pharmacophore hypothesis (fine lines) and a molecule designed from steric searching and molecular graphics (heavy lines).[22]

Bartlett et al.[21] had the same frustration when they tried to design conformationally constrained peptide mimics. They wanted to synthesize compounds in which the $C_\alpha$–$C_\beta$ vectors are maintained at the same angles and distances as the corresponding vectors between the side chains in the proposed bioactive conformation of the peptide or protein loop. CAVEAT was written to search the Cambridge Structural Database.[28] It finds structures in which the $C_\alpha$–$C_\beta$ vectors are maintained at the correct angles and distance even though the side chain might be missing or of the incorrect structure. The molecule to be synthesized is then designed by molecular graphics.

Figure 2 shows the side-chain vectors of residues the $\alpha$-amylase binding loop of tendamistat. Superimposed on it is a small molecule from the Cambridge Structural Database. Clearly there is a very close match of the substituent vectors in the two compounds.

The substituent vectors in a cyclic hexapeptide also matched those of the tendamistat binding loop. The compound with the correct side chains is cyclo[Phe-Ala-Trp-Arg-Tyr-D-Pro]. It inhibits $\alpha$-amylase with a $K_i$ of 14 $\mu$M. The corresponding acyclic peptide has a $K_i$ of 2 mM. In another study, a cyclic compound modeled after a CAVEAT hit inhibited thermolysin with a $K_i$ of 7 nM. The corresponding acyclic compound has a $K_i$ of 200 nM. Thus in both cases 3D searching helped design compounds for which the bioactive conformation is heavily populated.

In addition to the design of new synthetic targets, 3D searching can identify new bioactivities in existing molecules. It has identified two new classes of inhibitors of HIV-I protease, 8[37] and 9.[38] Both used criteria derived from the experimental 3D structure of the enzyme. In another case 10 and 11 were correctly recognized to be a new series of plant growth regulators.[39] In this case the search was based on a pharmacophore model. The compounds identified in these three examples are active but not as potent as other known compounds. Thus they will serve as new leads for molecular modification to optimize potency.

(33) Lin, C. T.; Pavlik, P. A.; Martin, Y. C. Use of Molecular Fields to Compare Series of Potentially Bioactive Molecules Designed by Scientists or by Computer. *Tetrahedron Comput. Methodol.*, in press.

(34) Martin, Y. C.; Kebabian, J. W.; MacKenzie, R.; Schoenleber, R. Molecular Modeling-based Design of Novel, Selective, Potent D1 Dopamine Agonists. In *QSAR: Rational Approaches on the Design of Bioactive Compounds 1991*; Silipo, C., Vittoria, A., Eds.; Elsevier: Amsterdam, 1991; pp 469–482.

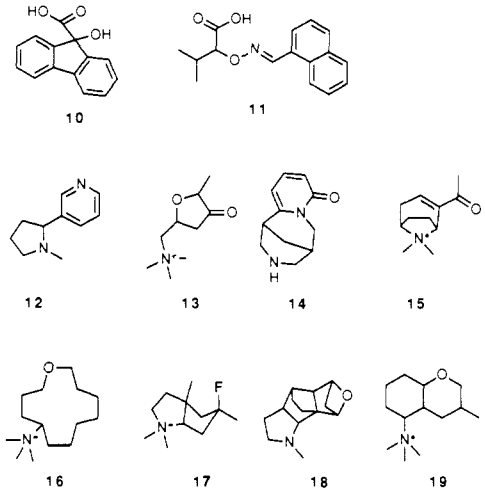
(35) Martin, Y. C.; Danaher, E. B.; May, C. S.; Weininger, D. MENTHOR, A Database System for Three-Dimensional Structures and Associated Data Searchable by Substructure Alone or Combined with Geometric Properties. *J. Comput.-Aided Mol. Des.* **1988**, *2*, 15–29.

(36) Schoenleber, R.; et al. U.S. Patent 4,963,568, 1990.

(37) DesJarlais, R. L.; Seibel, G. L.; Kuntz, I. D.; Furth, P. S.; Alvarez, J. C.; Ortiz de Montellano, P. R.; DeCamp, D. L.; Babé, L. M.; Craik, C. S. Structure-based Design of Nonpeptide Inhibitors Specific for the Human Immunodeficiency Virus 1 Protease. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6644–6648.

(38) Bures, M. G.; Erickson, J. W. Discovery of Novel Inhibitors of HIV-1 Protease by Three-dimensional Substructure Searching. *Tetrahedron Comput. Methodol.*, in press.

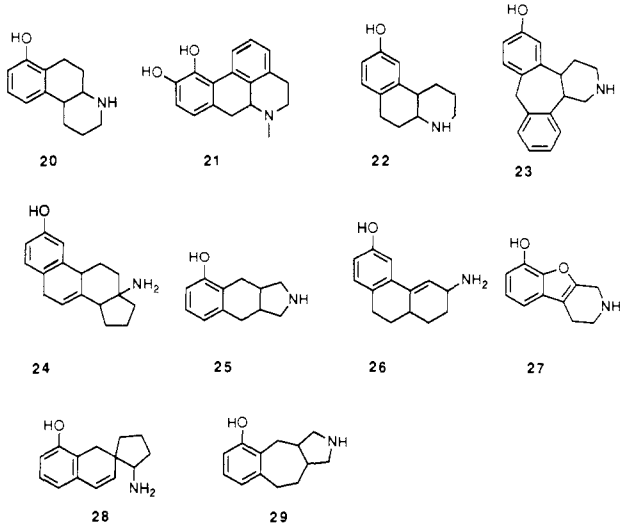
(39) Bures, M. G.; Black-Schaefer, C.; Gardner, G. The Discovery of Novel Auxin Transport Inhibitors by Molecular Modeling and Three-Dimensional Pattern Analysis. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 323–334.



Additionally, 3D searching was the basis of de novo design of molecules with potential bioactivity. For example, Sheridan and Venkataraghavan[22] used a pharmacophore derived from the nicotinic agonists **12–15** (Figure 3) to formulate a search of part of the Cambridge Structural Database.[28] Hundreds of molecules fit into the union surface of these four molecules and have atoms at the location of the pharmacophore basic or quaternary nitrogen atom and the hydrogen-bond acceptor. **16–19** are examples of compounds designed by subsequent molecular graphics analysis of the hits.[22] Figure 3 shows how one of the designed molecules fits onto one of the molecules in the lead series.

In another automated design example, Lewis and Dean searched the structures of a set of ring compounds to find locations on templates at which to add pharmacophore atoms.[19,20] They also identified atoms to be removed in order for the ligand to fit into the binding site. Their work is preliminary because so far they have considered 2D design only.

We used dopamine agonists such as **2** and **20–22** to formulate a pharmacophore for the template-based design of potential dopaminergics.[5] We performed geometric searches on three databases. Our program mutates the database molecules identified into those to be synthesized (see below).[6] Compounds **23–29** are examples of the hundreds of compounds suggested; Figure 4 shows one of them.



The method designed eight of nine classes of known fused-ring phenolic dopaminergic compounds and 62 other classes of fused ring compounds with previously unrecognized potential dopaminergic activity. In addition, at least 200 types of structures with one rotatable carbon–carbon bond were designed. Because we had quantitative CoMFA models[3a] of the dopaminergic receptors, we used the forecast affinities to set priorities for synthesis. The low frequency of finding the same ring type more than once suggests that other searches will design many more potential D2 agonists. Furthermore, compound design based on 3D substructure searching is equally applicable to beginning and mature medicinal chemistry investigations.

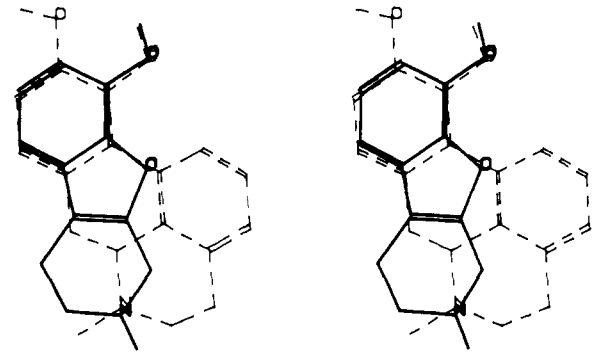


**Figure 4.** The superposition of a compound used to derive our D2 pharmacophore model (dashed line) and a potential dopaminergic designed by the computer (heavy lines).[5]

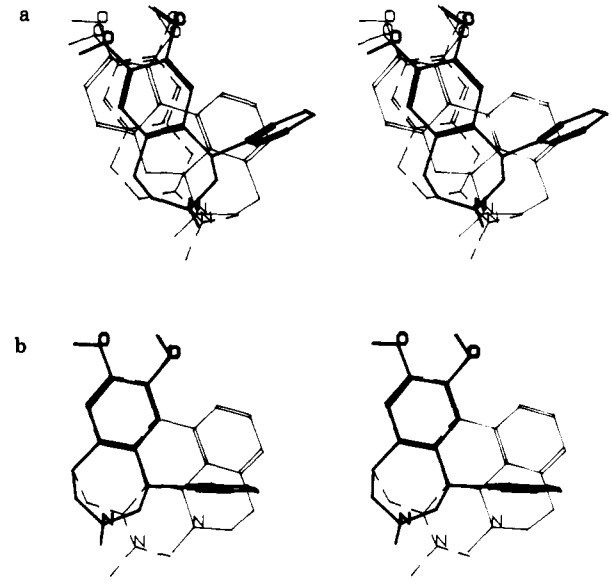


**Figure 5.** Two dopaminergic pharmacophore hypotheses derived from 5-hydroxy-2-aminotetralins (fine lines) and 7-hydroxy-2-aminotetralins (dashed lines).[14] In (a), the molecules are superimposed over their meta OH and amino groups. In (b) they are superimposed over their catechol rings. The solid line shows that the dopaminergic agent SKF 38393 meets the O–N distance requirement of (a), but requires the proposal of a new binding site for the N in model (b).

The fact that many compounds were identified created new problems and opportunities. It was difficult to organize the thousands of designed compounds. Originally this was done manually based on the 2D structures of the compounds. However, later we grouped the molecules by cluster analysis of geometric variables calculated from the pharmacophore and neighboring atoms.[40] Multivariate analyses of the steric fields of the molecules[33] groups them

(40) Martin, Y. C. Opportunities and Problems of 3D Similarity of Molecules for the Computer Design of Bioactive Compounds. *Abstracts of Papers*, 200th National Meeting of the American Chemical Society, Washington, DC, Fall 1990; American Chemical Society: Washington, DC, 1990.

**Table I.** Sources of Requirements for 3D Searching and Design

| source of 3D information for the search question | how the 3D requirements are described | how the 3D requirements are established | results of the search or design | refs |
|---|---|---|---|---|
| pharmacophore model from several active molecules | geometric (see Table II) | molecular graphics | molecules that match the pharmacophore | 14, 15, 24–28, 45 |
| pharmacophore model from several active molecules | geometric (see Table II), superposition rule, and surface description by points, spheres or CoMFA coefficients | molecular graphics and union surface calculation or CoMFA fit[2a] | molecules that match the pharmacophore and have a shape consistent with bioactivity | 14, 25, 45 |
| pharmacophore model from several active molecules | centers of spheres that fill the union surface (plus distance between pharmacophore atoms) | molecular surface analysis of the binding site | molecules that fit into the union surface and fit the pharmacophore | 22 |
| proposed bioactive conformation of a ligand | geometric (see Table II) | molecular graphics | molecules that mimic the ligand | 14, 15, 21, 25–28 |
| a low-energy conformation of a ligand | geometric (see Table II) | molecular graphics | molecules that fix the conformation | 14, 15, 21, 25–28 |
| 3D structure of the protein or DNA binding site | centers of spheres that fill the site | molecular surface analysis of the binding site | molecules that fit into the binding site and occupy some of the sites | 18, 46–49, 51, 55, 56 |
| 3D structure of the protein or DNA binding site | location of H-bond donor and acceptor atoms | potential energy calculations[42,43] or rules derived from experiment[44,47] | molecules that hydrogen bond to some or all of the groups in the binding site | 8–11, 14, 25–27, 19, 20, 47 |
| 3D structure of the protein or DNA binding site | potential energy calculations as the molecule is being designed | | molecules that fill the site | 6 |

by similarity in shape. Thus a chemist can choose a series of varying shape for synthesis, testing, and CoMFA[3a] analysis.

Lastly, by searching the 3D structures of active and inactive compounds, 3D searching methods can validate or refute a pharmacophore hypothesis (Figure 5).[14,24,26] For example, Loew et al.[41] searched 3D databases to confirm their benzodiazepine pharmacophore. All of the known high-affinity classes of ligands for the benzodiazepine receptor matched the pharmacophore, including several that had not been included in the derivation of the model. Only inactive compounds matched constraints that were relaxed from the prefered pharmacophore. 3D searching to derive or validate a pharmacophore will expand when we can quickly search all conformations of molecules of interest.

## Types of 3D Structure Searching

3D searching programs and their applications differ from each other in a number of respects (Table I). Each of these differences affects the characteristics of the molecules found or designed by the search.

**Geometric Searches in Which All 3D Features Are Required To Be Present.** This typically starts with a geometric search. Such searches consider the intramolecular relationships between geometric features such as points, lines, and planes (Table II) calculated from the 3D structure of a ligand.[4] Most pharmacophore hypotheses are based on such geometric relationships.

Because the calculations are intramolecular, distances and angles are independent of the enantiomer used. However, enantiomers have opposite values for torsion angles. If the searching program ignores the sign, then either enantiomer will pass the search; if it includes the sign, then only one enantiomer will pass.

It is handy if the user can specify the substructural environment of the atoms from which the points, lines, and planes will be calculated. For example, if the pharmacophore requires a basic nitrogen atom, neither an amide nor a quaternary ammonium at this site will do. On the other hand, for automated 3D design one might wish to specify atoms of geometric interest such things as "an $sp^3$ aliphatic atom in a ring of any size". A broad atomic specification also can help in searching for hydrophobic regions or hydrogen-bond donors or acceptors.

**Table II.** Typical Geometric Objects Calculated from the 3D Structure of a Database Molecule

- points located at
  - nucleus of an atom or center of a lone pair
  - center of mass of several atoms (such as the center of a ring)
  - projected binding points
    - of Hb acceptors or donors
    - of charged groups
    - at arbitrary locations calculated from other points, lines, and planes of the structure (dummy atoms)
- lines
  - between any two points above
  - least-square line of more than two points
  - through a point (e.g. center of mass) and normal to a plane
- planes
  - calculated from three points
  - least squares calculated from more than three points
  - calculated from two lines

**Table III.** Typical Geometric Constraints

| constraint | geometric objects used |
|---|---|
| distance | two points |
| | any atom and a point (defines a sphere) |
| | any atom and a line (defines a cylinder) |
| | any atom and a plane |
| angle | three points |
| torsion angles | four points, two lines, two planes, or two points and one plane (the latter three may have no solution) |

Table III lists the commonly used geometric constraints for describing the molecules to be identified.

As shown in Table I, the criteria may come from the 3D properties of one ligand, from pharmacophore mapping of a set of active and inactive ligands, or from the 3D structure of the protein or nucleic acid target.

Molecular modeling of an active compound might suggest several possible bioactive conformations. To probe which is correct, one would use 3D searching to design

(41) Loew, G. H.; Villar, H. O.; Jung, W.; Davies, M. F. Computer-Aided Drug Design for the Benzodiazepine Receptor Site. In *Emerging Technologies & New Directions in Drug Abuse Research. NIDA Res. Monograph #112. DHHSHUB Number (80M) 9101812*; Rapaka, R., Makriyanis, A., Kuhar, M. J., Eds.; U.S. Government Printing Office: Washington, DC, 1991; pp 643–661.

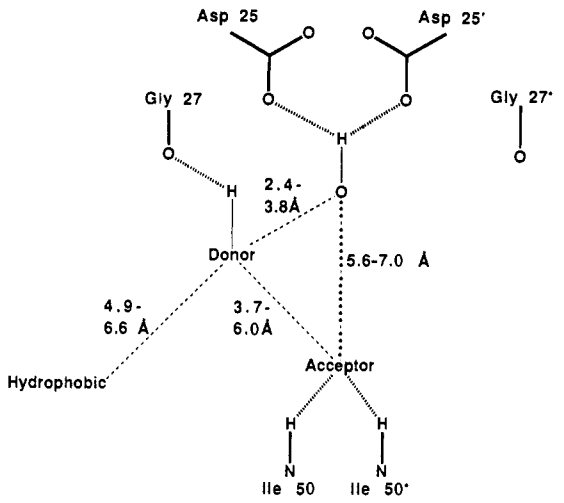



**Figure 6.** The search query for HIV-1 protease inhibitors designed from the 3D structure of the A-74704–protease complex. It required a H-bond donor and acceptor, an OH group, and a hydrophobic region at the indicated distances. The required interactions with the protein are shown by hashed lines.[38]

several mimics of each. The structure–activity relationships of the unconstrained molecule would suggest which groups are necessary to include in the mimics.

Pharmacophore mapping with several active and inactive compounds[29] or a 3D QSAR analysis[3] might suggest regions in space that an active ligand cannot occupy. In geometric searching these forbidden regions are specified with dummy atoms calculated from the structure of the ligand.

An experimental 3D structure of a ligand–macromolecule complex might establish the geometric requirements. For example, geometric criteria for HIV-1 protease inhibition were derived from the structure of a protease-ligand complex.[38] The search shown in Figure 6 identified the inhibitor **9**.[38] Alternatively, one could remove the ligand from the structure of the complex, calculate the location of especially favorable potential interactions,[42,43] and search for or design ligand molecules that match these criteria. One can also describe the principal features of the size and shape of the binding site with dummy atoms. However, in this case steric searching is also especially useful.

Using dummy atoms in geometric searches has limited utility to probe whether a database molecule can fit into a binding site. Sometimes the location of a forbidden region cannot be established precisely from points within the database molecule. For example, notice in Figure 5a that the pharmacophore N and O atoms do not overlap exactly. Thus, the location of a dummy atom described by distances from these atoms will differ from molecule to molecule. Also, for some molecules the supplied construct results in zero or two points. Lastly, unless signed torsion angles are included, geometric searching does not distinguish enantiomers.

**Prescreening To Reduce Search Time for Geometric Searches.** Most geometric searching systems are interactive for simple queries. Searching is made fast by screening out the 95–99% of the compounds that have no chance to meet the 3D constraints.[4,16,21,25-28,44] Screens are established corresponding to values of frequently searched distances or torsion angles. For example, bit 1 of the screen might correspond to an O–N distance of 2.0–3.0 Å, but 2 to 3.01–3.50 Å, bit $n$ to an O–O distance of 2.1–2.8 Å, and bit $n + 1$ to 2.81–3.24 Å, etc. CAVEAT[21] uses screens based on the angles between the substituent vectors. At search time the screens that match the search query are generated and structures that do not match them are eliminated from further consideration. After screening, a geometric search tests if the molecules have the required features.

Screening requires increased computer time during database loading of molecules since one must calculate all the distances and/or torsion angles of the structures and assign the appropriate screens. Screening is not effective if every molecule in the database has the feature.

**Steric Searching in Conjunction with Geometric Searching.** Ligand binding sites on macromolecules are of limited size and definite shape. We might know the shape directly from the 3D structure of the macromolecular target or hypothesize its smallest extent as the surface that encloses all superimposed active molecules. Thus, it makes sense to search for molecules that could fit into the sites of interest. Steric searches do this. The result of a steric search is different for enantiomers of a molecule.

The steric searches in ALADDIN and 3DSEARCH are simply an automation of what one would do interactively in a molecular graphics program. The user supplies the location of the points to use in the superposition. In ALADDIN,[14] one also specifies the location of the dot surface points in the same orientation frame or negative CoMFA[45] coefficients: in 3DSEARCH[25] one specifies the surface as a set of points with specific radii. The programs superimpose the database molecule and check if any atom in it intersects with the supplied surface representation. In ALADDIN[45] the intersecting atoms are labeled for subsequent removal.

Lewis and Dean[19,20] also used a combined geometric–steric searching strategy in their molecular design program OPTIMUS. They first calculate, from the 3D protein structure, the optimal location of hydrogen-bond donors and acceptors in a ligand. They then use geometric searching on generic skeletons to identify spacers that would hold the H-bonding atoms in place. Finally, they orient the spacer in the protein by placing putative pharmacophore atoms of the spacer at optimal positions of hydrogen-bonding groups. Lastly, they identify atoms on the spacer that are inside the protein surface, remove these atoms, and reestablish that the spacer still is one molecule. As noted above, they have published on 2D searches only.

**Searches in Which Not Every Query 3D Feature Need Match.** Kuntz et al.,[46] and later others,[47,48] used clique-detection methods to find possible orientations of

(42) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
(43) Boobbyer, D. N. A.; Goodford, P. J.; McWhinnie, P. M.; Wade, R. C. New Hydrogen-Bond Potentials for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure. *J. Med. Chem.* **1989**, *32*, 1083–1094.
(44) Cringean, J. K.; Pepperrell, C. A.; Poirrette, A. R.; Willett, P. Selection of Screens for Three-Dimensional Substructure Searching. *Tetrahedron Comput. Methodol.* **1990**, *33*, 7–46.
(45) Martin, Y. C.; Danaher, E. B. Unpublished extensions to ALADDIN software.
(46) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. A Geometric Approach to Macromolecule-ligand Interactions. *J. Mol. Biol.* **1982**, *161*, 269–288.
(47) Kuhl, F.; Crippen, G. M.; Friesen, D. A Combinatorial Algorithm for Calculating Ligand Binding. *J. Comput. Chem.* **1984**, *5*, 24–34.
(48) Smellie, A. S.; Crippen, G. M.; Richards, W. G. Fast Drug-Receptor Mapping by Site-Directed Distances: A Novel Method of Predicting New Pharmacological Leads. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 386–392.

ligands in a protein binding site. In DOCK,[46] the binding site is described by the smallest set of spheres that fill it. The potential ligand is described as a set of points located at the atomic nuclei. The algorithm orients the ligands in the binding site by superimposing some of the ligand atoms onto some of the centers of the site-spanning spheres.

A clique is a collection of points with known distances between them. The methods pair a ligand and a site point on the basis of the distance of all other points in the clique to them. A correspondence between the two is found when every distance between the included ligand points is within the tolerance of being equal to the corresponding distance between the included site points. A clique is the collection of these corresponding points.

DOCK keeps only those orientations of the potential ligands for which all atoms are inside the surface of the binding site.[46] The orientations are further ranked by the fraction of the site occupied. As noted above, later workers[22] added the requirement that the oriented database molecule must have atoms within 0.5 Å of the proposed locations of the pharmacophore atoms. These identified atoms will be manually converted into pharmacophore atoms in the designed molecules.

DOCK has been extended to database searching and ligand design.[18,22] In Version 2, chemical points in the binding site serve as docking centers. The scoring function is being changed to a potential-energy calculation including solvation.[49]

The scoring method in the original version of DOCK has been evaluated in a test with a 101 compounds tested as $\alpha$-chymotrypsin inhibitors.[50] Eight of the top ten scoring molecules are active inhibitors. The 3D searching and scoring method produces a statistically significant enhancement of active molecules compared to a random search.

Other steric searching algorithms are being developed. For example, the program SPERM is quite fast with 50 000 compounds searched in 2 h. In this program the 3D surface of a molecule is projected on to an icosahedron. Its shape is described by the lengths of 32 vectors from the center of mass to the vertices. An active molecule or the superposition of several molecules forms the search query and database molecules are tested for shape similarity to it. They rank orientations and molecules by the sum of the squared differences of the lengths of the vectors. A search of 30 000 compounds of the Cambridge Structural Database for those that mimic netropsin identified a eight known binders to DNA in the top 50 scoring molecules. The molecules identified by SPERM and DOCK are not identical.

**Conformational Flexibility.** Clearly a 3D search will not be complete until we can examine all energetically feasible conformations of the molecules. One solution is to store all conformations of every molecule. This clearly suggests huge databases and long searches. Methods that use more chemical knowledge promise to be more efficient.

DesJarlais et al.[52] considered flexible ligands in DOCK. They docked the fragments on either side of the rotatable bond at overlapping adjacent sites on the protein and then checked if the bond length was proper. The user supplied the rotatable bond information. A similar notion can be used in geometric searching by specifying distances not between atoms on a flexible chain, but rather of each to some point that is expected to be conformationally invariant.[53] A trivial example of this would be if one were searching for a phenol in which the orientation of the hydroxy hydrogen atom was important for the pharmacophore. Instead of using the orientation of the hydroxyl in the database, one could calculate the positions of dummy atoms at the two orientations of the H in the plane of the ring from the location of the O, its attached carbon, and the carbons attached to it. Usually, these flexible searches require too much information about the molecules being searched for them to be expected to be thorough.

CHEMX does a rule-based conformational search on each molecule as it is loaded into the database.[27] The low-energy conformations are used to generate the screens. Only the starting or first low-energy conformation is kept. Such conformational keying proceeds at 15–20 000 structures/day on an IBM RS/6000. At 3D search time, if a molecule passes the screen the conformational search is repeated to produce the target conformation.

Haraki et al.[54] compared searching using CHEMX starting structures and conformational flexibility vs the single CONCORD conformation. They made 3D databases of 22 000 compounds with known biological activities. In four of the five searches there were more hits if the molecules were allowed flexibility. However, in the case of $5HT_3$ agonists, the single CONCORD structures produced more hits and identified more of the compounds labeled as serotonergic. This difference suggests that as run, CHEMX[27] did not search the conformations made by CONCORD.[30] Haraki et al.[54] also calculated the fraction of active compounds in the hit list vs in the database. In every case, the searching based on the single conformation produced a higher enhancement. Thus 3D searching on all possible conformations of molecules requires careful attention to accurately generating and evaluating these conformations.

Smellie et al.[48] extended the ideas in DOCK by allowing for conformational flexibility of the ligand. In the example published, the site points are the location of hydrogen bond donors and acceptors on a protein. The 3D structure of the protein fixes the distances between them. Similarly, the ligand atoms are described by their hydrogen-bonding character.

To search all conformations of the ligand they implemented the idea[4] that the 3D structure of the ligand be described by the distance bounds matrix used in distance geometry.[55] Accordingly, for the clique-detecting algor-

(49) Kuntz, I. D.; Shoichet, B.; Bodian, D.; Roe, D.; Lewis, R.; Huang, C.; Ferrin, T.; Langridge, R. Recent Advances in Molecular Docking. *Abstracts of Papers*, 202nd National Meeting of the American Chemical Society: New York, NY, Fall 1991; American Chemical Society: Washington, DC, 1991.

(50) Steward, K.; Bentley, J.; Cory, M. Docking Ligands into Receptors: The Test Case of $\alpha$-Chymotrypsin. *Tetrahedron Comput. Methodol.*, in press.

(51) van Geerestein, V. J.; Grootenhuis, P. D. J.; Haasnoot, C. A. G. 3D Shape Fitting and 3D DB Searching by SPERM. *Tetrahedron Comput. Methodol.*, in press.

(52) DesJarlais, R. L.; Sheridan, R. P.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. Docking Flexible Ligands to Macromolecular Receptors by Molecular Shape. *J. Med. Chem.* **1986**, *29*, 2149–2153.

(53) Güner, O. F.; Henry, D. R.; Moock, T. E.; Pearlman, R. S. Flexible Queries in 3D Searching. Techniques in 3D Formulation. *Abstracts of Papers*, 202nd National Meeting of the American Chemical Society, New York, NY, Fall 1991; American Chemical Society: Washington, DC, 1991.

(54) Haraki, K. S.; Sheridan, R. P.; Ruskino, A., III; Venkataraghavan, R.; Dunn, D. A.; McCulloch, D. Looking for Pharmacophores in 3D Databases: Does Conformational Searching Improve the Yield of Actives? *Tetrahedron Comput. Methodol.*, in press.

(55) Crippen, G. M.; Havel, T. F. Distance Geometry and Molecular Conformation. In *Chemometrics Research Studies Series*; Bawden, D., Ed.; Research Studies Press: Letchworth (Wiley); New York, 1988.

ithm they supply the upper and lower bounds of the distance between the atoms in all possible conformations of the ligand. This distance bounds matrix is quickly calculated from bond lengths, bond angles, and van der Waals radii. In principle, only the 2D structure of the molecule is needed. The matrix is smoothed so that for every three points no one side is longer than the sum of the other two nor shorter than their difference. This raises some lower bounds and lowers some upper bounds.

There is a tentative match between site and ligand points when the distances to the ligand points are within the hydrogen-bonding tolerance of the corresponding atoms in the protein. These matches must be validated by the distance geometry embedding procedure.[55] This tests if the distance ranges of the proposed docking correspond to an actual 3D structure of the ligand. In the embedding they included, but kept rigid, all protein atoms within 6 Å of the ligand. Only 20% of the cliques actually pass this step.

Smellie et al.[48] reported that the distance bounds calculation and clique matching takes approximately 1 s on a low-end workstation. The distance bounds calculation is the more time consuming of the two. The subsequent embedding stage takes much longer than 1 s. Thus, with the described algorithms, searching a 60 000 compound database would take a minimum of 20 h. However, one could calculate and store the distance bounds when the database is built to eliminate the need to calculate them at run time.

Blaney reported a similar solution to the same problem using DOCK.[56] After the matching points are found, several 3D structures of the ligand are generated with the embedding procedure. The maximum distance between site points sets the maximum for all upper distance bounds of the ligand. The final distance geometry embedding uses the radius and location of the spanning spheres to describe the protein. A test with docking methotrexate to dihydrofolate reductase suggests that 10–100 random fits are needed to approximate the experimental binding orientation. Each fit takes 1.7 s on a high-performance workstation. Thus, 100 tries at fitting each ligand will take 3 min. Embedding will always be necessary and with current algorithms it is slow.

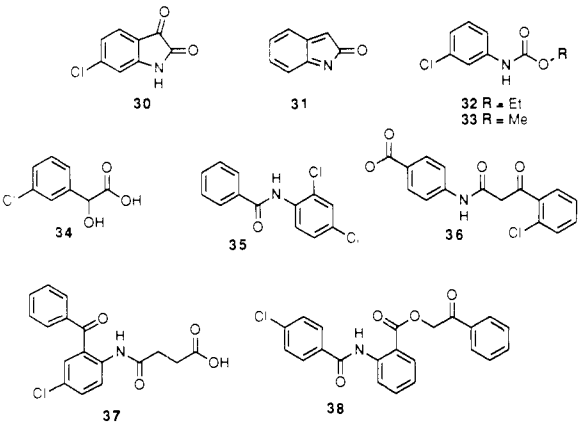
Recently, Clark et al.[57] used the bounded distance matrices to set distance screens analogous to those used for single conformations.[44] They tested a database of 1538 molecules against eight literature pharmacophores. A mean of 16 molecules passed the screening when the CONCORD conformation was tested but 147 passed screening with distance bounds. From the distance bounds search, a mean of 107 3D structures consistent with the pharmacophore could be generated. This is to be compared with 16 for the rigid search.

For each molecule matched in the flexible search there are approximately nine atom matchings possible. Thus although screening eliminates 90% of the molecules in the database, each remaining molecule potentially matches nine ways. Certain applications could stop with the first match, but others would require that all be processed. These workers concluded that, with current algorithms, flexible 3D searching is at least 100× slower than rigid 3D searching.

Conformational flexibility represents a challenge in both computer resources and in further processing the many hits. Advances on both fronts will make this new tool even more useful for medicinal chemists.

**3D Similarity.** By analogy with 2D similarity searching methods,[58] 3D similarity searching ranks database molecules relative to a known active molecule. Similarity searching requires no prior molecular graphics or pharmacophore mapping. Therefore, these methods are useful in identifying compounds for testing to provide the structure–activity information for later pharmacophore mapping.

Pepperrell, Willett, and Taylor explored the ability of different definitions of 3D similarity to detect molecules that have the same biological activity as the input molecule.[59] They found that a method called atom mapping is the most effective. This method identifies atoms in the database molecule that are identical in atomic number and most similar in intra-atomic distance profile to the atoms of the query molecule. The distance profile retains the atomic number of the atoms so that only distances between corresponding pairs are compared between molecules. The atoms of the database molecule are mapped onto the query molecule in order of decreasing atomic similarity. The overall 3D similarity is the sum of the mapped atom similarities. For example, a search based on **30** identified **31–35** as the four most 3D similar molecules. In contrast, **36–39** are most similar in 2D.




Even at this early stage of development, 3D similarity searching is fast, 50–100 compounds per second. It thus would take 10 min to rank 60 000 structures. Yet if the list were shortened to 6000 structures by requiring certain substructures to be present and that there is physical sample to test, even today it would be interactive.

## Programs That Design New Molecules

ALADDIN. This program includes a special language, MODSMI, for the automatic generation of the 2D structures of the compounds suggested from 3D substructure searching for geometric mimics.[6] (The 3D structures of the designed molecules are generated with CONCORD.[30]) MODSMI is used to tell the computer how to transform the database molecules into those that meet the pharmaco-

(56) Blaney, J. A Distance Geometry-based Approach for Docking Conformationally Flexible Molecules from 2D or 3D-Chemical Databases. *Abstracts of Papers*, 202nd National Meeting of the American Chemical Society, New York, NY, Fall 1991; American Chemical Society: Washington, DC, 1991.

(57) Clark, D. E.; Willett, P.; Kenny, P. W. Pharmacophoric Pattern Searching in Files of Three-Dimensional Structures: Use of Smoothed-Distance Matrices for the Representation and Searching of Conformationally Flexible Molecules. *J. Mol. Graphics*, in press.

(58) Willett, P. *Similarity and Clustering Methods in Chemical Information Systems*; Research Studies Press: Letchworth, 1987.

(59) Pepperrell, C. A.; Willett, P.; Taylor, R. Implementation and Use of an Atom-mapping Procedure for Similarity Searching in Databases of 3-D Chemical Structures. *Tetrahedron Comput. Methodol.*, in press.

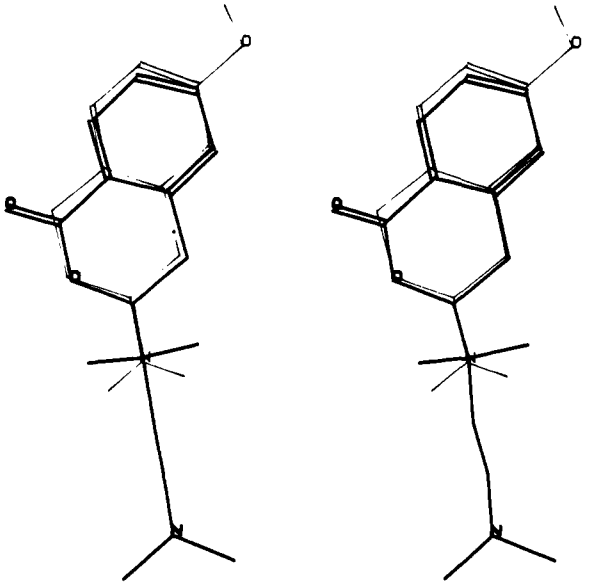


**Figure 7.** The superposition of a database molecule and the known dopaminergic designed from it with ALADDIN.[5]

phore requirements. The verbs "nibble", "replace", "join", and "axe" tell the program to remove the identified atom(s), to replace one set of atoms with another, to join two atoms to form a bond, or to change the bond order by the specified amount. The user also supplies the symbols of the atoms to be added with the "replace" command.

The atoms that are the object of the action are identified by their substructure environment and whether they are labeled as part of a specific geometric object. This provides a way to transform only the appropriate atoms in the database molecules into the pharmacophore atoms.

In our design of dopaminergics and analgetics we identified "any aromatic atom in exactly one six-membered ring". For these designs, we also searched for an $sp^3$ aliphatic atom not attached to an aromatic ring. If the distances between these two types of atoms were correct, we added the OH to the former and converted into a carbon if it were not so. We modified the latter to form the basic nitrogen.

To reduce the number of geometrically similar molecules, we also typically use MODSMI to remove atoms that do not contribute to the geometric relationships of the pharmacophore atoms.[5] For example, side chains and nonpharmacophore substituents on rings are removed. We also usually change aliphatic ring oxygen and nitrogen atoms into carbon atoms and pyridyl nitrogen atoms.

Figure 7 shows the structure of a database molecule superimposed over that designed by ALADDIN.

OPTIMUS. This program automatically removes atoms that penetrate into the protein and checks that this removal does not destroy the integrity of the ligand nor change its geometry substantially.[19,20] It also identifies atoms that should be converted to hydrogen-bond donors or acceptors by some manual procedure. In validation studies, it used the 3D structure of dihydrofolate reductase to design the pteridine ring of methotrexate, a known DHFR inhibitor. It also identified an amidinophenyl grouping for the AAPA binding site on trypsin.[20]

GROW. This computer program builds up the structure of a potential peptide or peptide-like ligand within a binding site of known 3D structure.[7] It uses 3D molecular fragments of the building blocks in different conformations. For example, some fragments are the amides of the natural amino acids in each low-energy conformation of the side chain. The fragments are iteratively pieced together within the binding site. Each stage of ligand growth is evaluated with an energy function that considers van der Waals, coulombic, strain, and desolvation energies. Every fragment is tested for a given position in the chain, and then the $n$ (usually 10) best-fitting are used as starting structures for the next building step. The $n$ best fit side chains are kept and the process repeated on the new growing end. The user specifies a starting point and direction of addition at each stage.

The energy function and search strategy has been validated by demonstrating that GROW reproduces the known binding orientations of inhibitors of rhizopuspepsin and HIV-1 protease. More importantly, GROW designed a substrate for rhizopuspepsin and a peptide inhibitor of renin, K1–30 mM, which is approximately equipotent to molecules patterned after human angiotensinogen.

LUDI. This program is similar to GROW and OPTIMUS in concept.[8] Starting with a binding site of known 3D structure, the position of potential hydrogen bonding and hydrophobic groups in the binding site is estimated from empirical rules or data derived from the Cambridge Structural Database. A database of fragments is searched and placed at favorable positions in the binding site. Then a second database is searched to find spacers that bridge the fragments into a single molecule. The method was validated by showing that the crystal packing of benzoic acid is reproduced and that methotrexate is correctly built into dihydrofolate reductase.

## Sources of 3D Structures for Searching

Unless one is using distance geometry type searches, it is necessary to supply 3D coordinates to search. Experimental 3D data (usually crystallographic) on small molecules is collected in the Cambridge Structural Database of ~85 000 3D structures.[28] Software is supplied for 3D searching of this database. Eyermann used its supplied software for geometric searches to find leads for 3D design.[60] Additionally, DOCK and CAVEAT search subsets of this file.

CONCORD generates a 3D structure from an input 2D structure in a few seconds.[30] Its expert system uses chemical intuition and a novel strain function that optimizes over a single composite variable. CONCORD preserves the stereochemistry from MACCS[61] or stereochemical SMILES files[62] or it will produce all stereoisomers. Small organic molecules are handled well, but peptides are not. CONCORD is the standard program for creating 3D databases of corporate,[5,23] commercially available,[5,63] and biologically active compounds,[5,63] and Chemical Abstracts (4 000 000) molecules.[64] Its most serious limitation is that one conformation does not describe the 3D properties of most molecules.

AIMB also rapidly generates a 3D structure from 2D input.[31] It finds the smallest set of fragments that covers

(60) Eyermann, C. J.; Lam, P. Y-S.; Kerr, J. S.; Ripka, W. C. Using Databases as an Aid in Developing Synthetic Targets. *Abstracts of Papers*, 202nd National Meeting of the American Chemical Society, New York, NY, Fall 1991; American Chemical Society: Washington, DC, 1991.

(61) MACCS-II and MACCS-3D are software products from Molecular Design Limited, San Leandro, CA 94577.

(62) Weininger, D.; Weininger, A. SMILES. A Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1989**, *28*, 31–36.

(63) Henry, D. R.; McHale, P. J.; Cristie, B. D.; Hillman, D. Building 3D Structure Databases. Experiences with MDDR-3D and FCD-3D. *Tetrahedron Comput. Methodol.*, in press.

(64) Fisanick, W. Characteristics of Computer Generated 3D and Related Molecular Property Data for CAS Registry Substances. *Tetrahedron Comput. Methodol.*, in press.

the input molecule with a 2D search on molecules in a 3D database. In the second step it uses these to generate the 3D coordinates. The time to build a structure decreases as the size of the 3D database searched increases. The 3D builder in the CHEMX system[65] uses this concept and has been used to build 3D databases of biologically active compounds.

WIZARD and COBRA generate "all" low-energy conformations of a molecule.[32] They use 3D templates and artificial intelligence criticism of starting structures to generate conformations.

Gasteiger et al.[66] also developed a 3D structure-generating program that applies to the entire range of organic chemistry. It is claimed to handle macrocyclic compounds more satisfactorily than previous programs.

Since these programs are so fast, here may be no advantage to storing the coordinates in databases, which use disk space and need to be regenerated for new versions of the structure generator. For example, ALADDIN can generate with CONCORD the 3D structures it is to search.[45] The user supplies the search query and a list or database of 2D structures.

## Implications of 3D Searching Methods beyond Database Searching

The 3D searching programs need a description of a pharmacophore or experimental binding site that the computer can process. These descriptions provide one the opportunity to make a database of the pharmacophores. Imagine a system in which the 3D structures of molecules proposed for synthesis would be compared with this database of pharmacophore descriptions.[67] Unanticipated potential biological properties might be so identified.

## Directions of the Field

The algorithms for 3D searching are being continually improved. For example, the user frequently supplies a set of distances that implicitly constrain other distances. Clark et al.[68] showed that, if the program generates these other constraints, more compounds are eliminated from consideration.

Cross et al.[69] at Chemical Abstracts are investigating how to integrate 2D and 3D searching to gain the maximum search speed. They use sophisticated screens based on 2D substructure environment and 3D properties to eliminate as many compounds as possible from more detailed checking. Currently, their system can process 2000–6000 structures per minute. Fisanick et al.[64] are investigating flexibility/rigidity indices and minimum–maximum interatomic distances calculated from the 2D structure of the molecule. These numbers can be used to set screens to speed up searching. Additionally, the flexibility/rigidity indices may be useful for property prediction. They are also extending the results of Clark et al.[68] by including more known geometric properties of triangles to speed up the searching procedure.

Bradshaw and Maliski[70] suggested the use of the most restrictive paths between atoms as a means to set the bounds on the distance between them. The most restrictive paths are derived from the 2D structure of the molecule. For example, one would want to avoid full 3D searching on a structure that does not have enough atoms between those of interest to possibly meet the distance constraint. The most restrictive paths would set narrower distance ranges than would distance geometry triangle smoothing and are potentially much faster to calculate. They propose establishing screens using distance bounds generated with this procedure as the molecules are loaded into the database. When a 3D search is done, molecules that cannot meet the criteria are eliminated. Only at this time would one generate the 3D structures for the molecules that pass. This strategy eliminates the storage of large 3D databases and the problem of updating them when a new version of the 3D generator or a totally new 3D generator is available. It also eliminates the computer time used to generate 3D structures that never meet a screen.

The results from 3D searching will be only as good as the data on which the searches were derived. Methods based on the 3D structure of proteins suffer from the lack of accuracy of these structures and from the problems of accurately computing the interaction energies of proposed ligands. We are only beginning to understand how to handle solvation of the unbound molecule and the active site. Bound water molecules and other solutes in the complex also complicate predictions. Other problems arise for searches based on pharmacophore mapping. First, one must derive the model. If the correct set of molecules has not been tested, there may be several rather than one model. Beyond this, the assumption of a common binding mode might be invalid. Continued improvements in these disciplines that supply the 3D information will lead to increased accuracy of the 3D searching results.

Other groups are working on the design of molecules to fit a binding site or to match a pharmacophore.[71,72] The build-up procedures face the problem that the number of designed molecules equals product of the number of parts that fit at each site. This can become a huge number of suggestions. For molecules that are built into a known 3D structure, energy calculations as in GROW can set priorities. Simulated annealing procedures to select the best side chains have also shown promise. However, for molecules that are built to match a pharmacophore, it would seem that some 3DQSAR method would need to be used to set priorities. The integration of the techniques from different laboratories into a useful computer program remains a challenge.

(65) Chem-X molecular modeling software, developed and distributed by Chemical Design Ltd., Oxford, England, 1991.
(66) Gasteiger, J.; Hiller, C.; Rudolph, C.; Sadowski. Automatic Generation of 3D Atomic Coordinates for Organic Molecules. *Tetrahedron Comput. Methodol.*, in press.
(67) Martin, Y. C.; Kim, K. H.; Bures, M. G. Computer-Aided Drug Design in the 21st Century. In *Medicinal Chemistry in the 21st Century*; Wermuth, C., Koga, N., Gujral, P. D., Eds.; Blackwell: Boston, in press.
(68) Clark, D. E.; Willett, P.; Kenny, P. W. Pharmacophoric Pattern Matching in Files of Three-Dimensional Chemical Structures: Use of Smoothed Bounded Distances for Incompletely Specified Query Patterns. *J. Mol. Graphics* **1991**, *9*, 157–163.
(69) Cross, K. Comparison of Atom-by-atom 3D Search Routines for Searching CAS Registry Substances. *Abstracts of Papers*, 202nd National Meeting of the American Chemical Society, New York, NY, Fall 1991; American Chemical Society: Washington, DC, 1991.
(70) Bradshaw, J.; Maliski, E. G. The Use of Most Restrictive Paths in 3D Search Strategy. *Abstracts of Papers*, 202nd National Meeting of the American Chemical Society, New York, NY, Fall 1991; American Chemical Society: Washington, DC, 1991.
(71) Wipke, W. T.; Pitman, M. C.; Koehler, R. T.; Kislin, B. S.; Anderson, G. D. INVENTON: A System for Inventing Molecular Structures. *Abstracts of Papers*, 202nd National Meeting of the American Chemical Society, New York, NY, Fall 1991; American Chemical Society: Washington, DC, 1991.
(72) Gillett, V. J.; Flanagan, K.; Johnson, P. A.; Marshall, C.; Mata, P.; Sike, S.; Stebbings, A. L. Automated Structure Design in 3D. *Tetrahedron Comput. Methodol.*, in press.