Journal of Computer-Aided Molecular Design, 2 (1988) 15–29
ESCOM

J-CAMD 031

MENTHOR, a database system for the storage and retrieval of three-dimensional molecular structures and associated data searchable by substructural, biologic, physical, or geometric properties

Yvonne C. Martin[a], Elizabeth B. Danaher[a], Catherine S. May[a] and David Weininger[b]

[a]*Pharmaceutical Products Division, Abbott Laboratories, Abbott Park, IL 60064, U.S.A. and* [b]*Medicinal Chemistry Project, Pomona College, Claremont, CA 91711, U.S.A.*

Received 10 March 1988
Accepted 22 April 1988

Key words: Molecular graphics; Databases for molecular structures; Binding site models; Substructural searching; Dopaminergic agonists

SUMMARY

MENTHOR is a database system for the storage and retrieval of three-dimensional coordinate and charge information on molecules as well as of traditional biological and physical properties. Our molecular graphics system retrieves from MENTHOR structural information in individual molecules and receptor map/macromolecular binding site hypotheses. Substructural searches of MENTHOR are used to find starting coordinates for molecular modeling and traditional database searches of MENTHOR identify compounds for which modeling is needed. It also forms the data to be searched with ALLADDIN, our substructure/geometric search program. MENTHOR expedites molecular modeling by organizing previous work and facilitating transmission of information between individuals. Examples from modeling of D-2 receptor agonists are shown.

OBJECTIVES

Molecules interact with each other as three-dimensional objects: ideally one would have an understanding of the relationships between three-dimensional structure and properties of the known compounds before an attempt was made to design new molecules with the desired properties. Our goal is the design of new drugs; we study the relationships between chemical structure and biological properties. This report describes one tool that we have developed for such work. It is named MENTHOR, pronounced 'mentor' because it can teach us much, but spelled to remind us that it is based on the THOR software [1].

Figure 1 illustrates one strategy for the design of new drugs. It is based on the three-dimensional structure–activity relationships of the small molecules that interact with a biomacromolecule

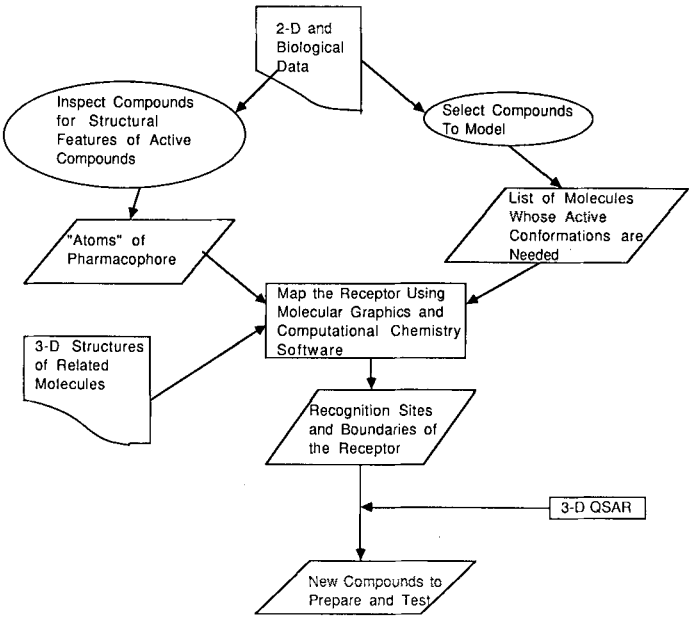


Fig. 1. The information flow in a typical binding site investigation.

[2,3]. Molecular modifications identify the atoms that are necessary for activity. Active compounds that hold these key atoms in somewhat spatially defined positions are selected for detailed molecular modeling. If no such compounds are available, sterically constrained compounds to be synthesized are designed from the low-energy conformations of the natural ligand. The biological properties of these constrained compounds then form the basis for the map of the target binding site.

The main activity in such binding site mapping is the examination and comparison of the many possible conformations and relative superpositions of a number of molecules. In such studies, one quickly realizes the inadequacy of the traditional method of structure storage, namely storing each set of coordinates and charges in a separate computer file. Despite careful naming of directories, sub-directories and files, it is often difficult to locate a desired data set of atomic coordinates and partial atomic charges or indeed to even discover if a particular molecule has been modeled. Sharing modeling results within a group of people further amplifies the difficulty. To solve these problems, we built a database for our modeling data and the biological and physical properties of the compounds. The resulting system provides us and other users of the MedChem software with a convenient tool for molecular modeling and also an exciting resource for three-dimensional compound design.

SYSTEM DESIGN REQUIREMENTS

Our first design requirement was that the database system integrate smoothly with our existing molecular modeling software. It was especially important that we be able to search for and retrieve coordinates and charges while we were using our molecular graphics system INTERACT/

CMD [T.J. O'Donnell, T. Koschmann, Y.C. Martin, F.C. Thomas and H.J.R. Weintraub, Abbott molecular modeling system for small molecules, unpublished]. Thus we required that it be easy to access the data directly from computer programs. This would ultimately also allow retrieval of compounds based on three-dimensional geometric relationships between atoms in particular molecular environments.

Our second design requirement was that compounds be retrievable by substructure. Such searches find starting three-dimensional structures for molecular modeling of analogues. We also required that compounds and specific sets of molecular coordinates be retrievable by name and that there be no limit to the number of names by which a compound could be identified. For example, the drug Eutonyl is also known as pargyline, A-19120, and MO-911; its trade name, generic name, Abbott number, and project number, respectively. However, if someone also wanted to label and retrieve a compound or a set of coordinates by the name 'Best to date', that should also be possible.

However, the capabilities of the proposed system were designed to go beyond mere storage and retrieval of molecular coordinates and charges. The final design requirement was that every coordinate/charge data set for a particular compound be associated with all other data for this compound. This would allow us to retrieve all low-energy conformations of a particular compound and so have these available as starting structures for the modeling of analogues. Both the whole set of conformations and stereoisomers and a specified subset were required to be retrievable. We also wished to be able to identify which is the current best guess as to the active conformation and stereoisomer of the molecule, who made the selection, and for which biological target. A related design requirement was that it be easy to also store biological and physicochemical data on compounds. We were especially eager to conveniently store a wide variety of literature biological data. This close association of all data on one compound is designed to support the creativity of the scientist to formulate and evaluate structure–activity hypotheses.

IMPLEMENTATION

THOR as a basis of our system

The Pomona College Medicinal Chemistry Project chemical information database THOR [1], met most of our objectives. Accordingly, we decided to expand the capabilities of THOR to meet our specific needs.

THOR is a database designed specifically for efficient storage and retrieval of information based on chemical structure. Data may be retrieved by the SMILES [4] or Wisswesser Line Notation linear codes for chemical structures, as well as by CAS numbers, local identifiers such as the Abbott number, and any name. The associated program GENIE [1] provides one a language to specify program control based on the substructures found in a molecule or on the THOR data. Alternatively, the program MERLIN [1] provides true substructure searching of a THOR.

Data in a THOR database are stored as a data tree, TDT. The root of the data tree is the unique SMILES linear structure code of the molecule. All data associated with a particular SMILES, i.e. structure, are stored in the same data tree. This allows one to associate with one SMILES the multiple names, coordinate/charge data sets, and/or chemical and biological data. Adding new data types to an existing THOR is easily achieved, since the data tree is stored as a character string with

data types and data elements separated by delimiters. The description of the data types is kept in a separate file that can be changed with a text editor.

Data security is maintained in a THOR by three levels of password: READ, READ/WRITE, and READ/WRITE/DELETE. Each page has a timestamp of the most recent entry.

New programs and THOR commands written to complete MENTHOR

A number of additional programs and additions to THOR functionality were required to integrate the MedChem software into our existing molecular modeling environment (Fig. 2).

Program to generate the THOR data tree from XYZ coordinates

MSFTOTDT converts a molecular structure file into a THOR data tree that can be read into the MENTHOR. Our molecular structure files are in the format shown for acetic acid in Fig. 3. The atoms may be in any sequence, no bond order is given, and all attached hydrogens are included. MSFTOTDT first determines the SMILES linear notation of the molecule. The initial step in this determination is the identification of the atomic hybridization of each atom from its atomic symbol and connectivity. (Obviously, this can be done only if the molecular structure file has all the atoms of a molecule). Next, all rings in the molecule are identified and aromatic rings noted. Then the bond orders are established, and finally the SMILES is put into a unique linear

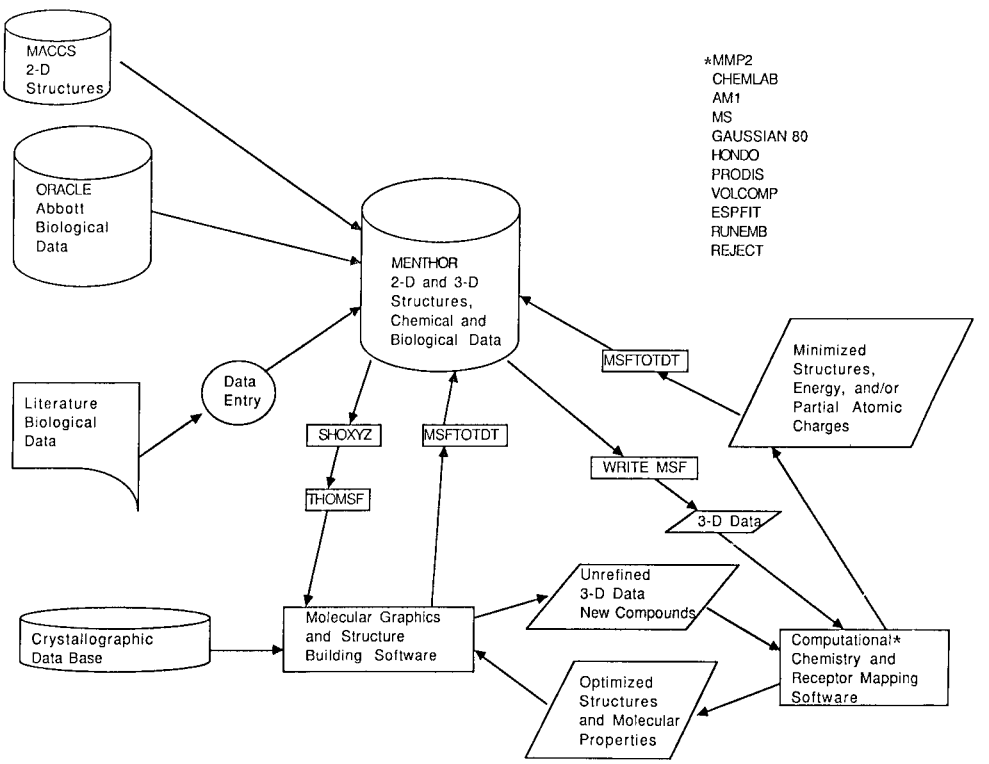


Fig. 2. The relationship between various computer programs and databases described.

```
ACETIC Built from COOH.TEM by REPLACE and ROTATE
Charges from fit to ESP surface from STO-3G
INTERACT/CMD 13-APRIL-1987 YCM                    0     8     7
C1          0.00063   -0.00017   -1.54000   -0.53170 C
C2          0.00000    0.00000    0.00000    0.79750 C
O3          1.17715    0.00000    0.67925   -0.58770 O
O4         -1.05793    0.00000    0.60914   -0.46070 O
H5          0.47196   -0.81627   -1.87381    0.14300 H
H6          0.47183    0.81672   -1.87392    0.14290 H
H7         -0.94224    0.00017   -1.87437    0.13680 H
H8          1.00140    0.00000    1.62264    0.35970 H
C1    C2     1.000
C2    O3     1.000
C2    O4     2.000
C1    H5     1.000
C1    H6     1.000
C1    H7     1.000
O3    H8     1.000
```

Fig. 3. A molecular structure file for acetic acid. The atom labels in columns 1–6 are arbitrary, but the atomic symbol in the sixth field is not. Charges may be zero or sum to a non-integer.

sequence. During this process two checks are made: that the derived bond order is consistent with the input connectivity, and that the number of hydrogen atoms implied by the SMILES is equal to the number of hydrogen atoms in the molecular structure file. In the second stage of MSFTOTDT, the coordinates and charges are associated with the atoms in the SMILES linear notation of the structure. In the data tree, the coordinates of the heavy atoms are stored in the same sequence as they appear in the SMILES. Then follow the coordinates of the hydrogen atoms in the sequence of the heavy atoms to which they are attached. Finally, the charges, if present, are listed in the same sequence as the coordinates. The THOR data tree of acetic acid is shown in Fig. 4. Because it has no embedded blanks, it is roughly 20% the size of the molecular structure file from which it was derived. When the THOR data trees are loaded into MENTHOR all of the data on one compound is associated with its unique SMILES.

THOR command to show stored structures

The THOR command DRAW 'xyz data set' allows the scientist a quick view of the conformation in a particular XYZ data set. It produces four views of the molecule: the standard flat representation and three orthogonal views of the three-dimensional structure. Structures can be drawn with or without hydrogen atoms. On color terminals, the structure is colored according to the partial atomic charges or atom type.

```
$SMI<CC(0)=O>$CC<ACETIC; ACETIC.MSF; MSFIN/INTERACT/CMD 13-APRIL-1987 YCM; ACETIC Built
from
COOH.TEM by REPLACE and ROTATE; Charges from fit to ESP surface from STO-3G;0.00063,-0.00017,
-1.54000,0.00000,0.00000,0.00000,1.17715,0.00000,0.67925,-1.05793,0.00000,0.60914,0.47196,-0.81627,
-1.87381,0.47183,0.81672,-1.87392,-0.94224,0.00017,-1.87437,1.00140,0.00000,1.62264;-0.5317,0.7975,
-0.5877,-0.4607,0.1430,0.1429,0.1368,0.3597>|
```

Fig. 4. The THOR data tree of acetic acid with the coordinates and charges from the molecular structure file shown in Fig. 3. If all partial atomic charges are zero or if the sum of the charges is not an integer, the charges are not stored in the THOR data tree.

NAME Apomorphine								0

XYZ COORDINATE data survey table

Item	Subset	XYZ COOR	MSF FIL	SOURCE	TITLE	SUBTITL	3D-DATA	CHARGE
24	R - ENA	APOMORPI	XRA:[MA	MSFIN/	APOMORP	.MSFFIL	1.91430	-0.5467
25	R - ENA	APO1NINV	WD:[MAR	MSFIN/	INVERSI	1/27/84	1.53786	
26	R - ENA	APO2OMP2	WD:[MAR	MSFIN/	SECOND	10/ 2/8	1.94910	
53	S + ENA	SENANAPO	WD:[MAR	MSFIN/	S ENANT	3/13/86	2.22278	

ACTIVE XYZ data survey table

Item	Subset	ACTIVE XYZ	ACTIVITY	MODELER
22	R - ENANTIOMER	APOMORPHI	D2	Y. C. Martin
23	R - ENANTIOMER	APOMORPHI	D1	Y. C. Martin

Fig. 5. The SHOXYZ screen for apomorphine.

LOCAL NAME Apomorphine	0

Datatype frequencies by page and subset:

Datatype	Total	UNKNOWN	R - ENAN	S + ENAN
SMILES	1	0	0	0
CAS NUMBER	1	0	0	0
ANUMBER	3	1	1	1
NAME	3	0	2	0
SUBSET	3	0	0	0
XYZ COORD	4	0	3	1
ACTIVE XYZ	2	0	2	0
D1 BINDING (lit)	3	0	3	0
D2 BINDING (lit)	7	0	6	1
D1 INVITRO (lit)	1	0	1	0
D2 INVITRO (lit)	1	0	1	0
D1 INVIVO (lit)	1	0	1	0
D1 BINDING DATE	2	2	0	0
D2 BINDING DATE	1	1	0	0
CYCLASE D1 AGON DATE	2	2	0	0
CYCLASE D2 AGON DATE	1	1	0	0
ALPHA1 KI	3	1	1	1
ALPHA2 KI	3	1	1	1
BETA1 KI	3	1	1	1
BETA2 KI	3	1	1	1
TIMESTAMP	1	0	0	0
LOCAL NAME	1	0	0	0
LOGP	1	0	1	0
CLOGP	1	0	0	0
MOLFORM	1	0	0	0
(Total data items)	53	11	25	7

Fig. 6. A summary of the types of data stored for apomorphine.

NAME Apomorphine								0

XYZ COORDINATE data survey table

Item	Subset	XYZ COOR	MSF FIL	SOURCE	TITLE	SUBTITL	3D-DATA	CHARGE
24	R - ENA	APOMORPI	XRA:[MA	MSFIN/	APOMORP	.MSFFIL	1.91430	-0.5467
25	R - ENA	APO1NINV	WD:[MAR	MSFIN/	INVERSI	1/27/84	1.53786	
26	R - ENA	APO2OMP2	WD:[MAR	MSFIN/	SECOND	10/ 2/8	1.94910	
53	S + ENA	SENANAPO	WD:[MAR	MSFIN/	S ENANT	3/13/86	2.22278	

ACTIVE XYZ data survey table

Item	Subset	ACTIVE XYZ	ACTIVITY	MODELER
22	R - ENANTIOMER	APOMORPHI	D2	Y. C. Martin
23	R - ENANTIOMER	APOMORPHI	D1	Y. C. Martin

Fig. 7. A listing of the result of the THOR command TYPE ONLY XYZ for apomorphine.

Program to show the molecular graphics user the XYZ data sets available for a compound

The program SHOXYZ writes to the terminal two tables. The first is of the coordinate/charge data sets available for the compound. The second is the names of the data sets that correspond to the proposed biologically active conformation. For example, Fig. 5 is the SHOXYZ screen for apomorphine. SHOXYZ can be accessed from the operating system command level or from within our molecular graphics program. SHOXYZ finds the MENTHOR page by name, Abbott number, or XYZ data set name.

Programs to transfer coordinates from MENTHOR into our molecular graphics system

The program WRITETSF and the THOR command WRITE TSF write from MENTHOR a specified scratch molecular structure file. This file can be read into our molecular graphics system or processed further for energy minimization, surface or charge calculations, etc. A graphics macro, THOTSF 'filename', creates the file, reads it into INTERACT/CMD, and draws the molecule with the heteroatoms indicated. Both THOTSF and WRITE TSF produce molecular structure files with a different extension, TSF rather than MSF, to identify them as temporary files from a database and thus candidates for deletion when the graphics session is complete. Finally, the command THOR of our molecular graphics system spawns a subprocess with the user executing THOR on a selected database.

MENTHORs currently available

The first MENTHOR we built contains compounds tested for adrenergic or dopaminergic activity. In this MENTHOR are selected literature and proprietary biological data on each compound tested at Abbott plus some structurally related literature compounds, calculated or measured octanol–water logP and pKa values, and atomic coordinate and charges. For example, Fig. 6 shows a summary of the types of data stored for apomorphine. The Abbott data are updated from searches of our MACCS* and biological databases.

*MACCS is the trademark of Molecular Design, Ltd., San Leandro, CA, U.S.A.

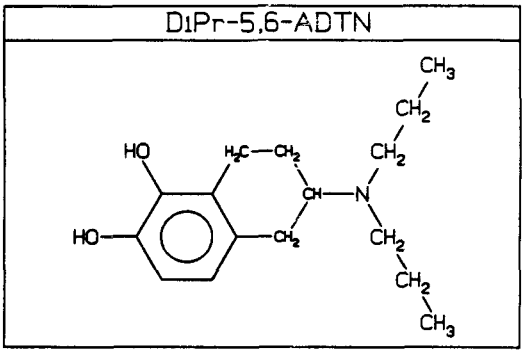
The second MENTHOR contains more diverse structures, each with at least one set of atomic coordinates. It contains the small molecules modeled at Abbott plus many compounds retrieved from the Cambridge Crystallographic Data base [5]. We are using CONCORD [6] to extend it to include even more diverse structures. It is used for MERLIN substructure [1] and substructure/geometric [J.H. Van Drie, D. Weininger and Y.C. Martin, manuscript in preparation] searches to retrieve structures that might be biologically active or that serve as starting points from which to construct coordinates of other analogues.

RESULTS

Typical questions answered by MENTHOR

The simplest question asked of MENTHOR is if any coordinates of a particular molecule are available. Figure 5 (a table) and Fig. 7 (a detailed summary) are examples of such output. Both are produced by standard THOR commands.

A more complicated search would involve questions asked in the early stages of deriving a binding site hypothesis. For example, one might ask 'What compounds have been modeled for D-2 activity?' 'Have I modeled everything that has been tested?' These questions would be answered by



CLOGP	3.201	3
LEVERR	All fragments measured.	
VERSION	3.5	
D2 BINDING	(lit) 1.3UM	5
LIGAND	[3H]spiroperidol	
REFERENCE	SEILER, MARKSTEIN, MOL PHARMACOL, (1982) 22,281	
COMMENT	IC50,IN PRESENCE OF 1UMOL (+)BUTACLAMOL	
SPECIES	CALF	
TISSUE	CAUDATE NUCLEUS	
D2 BINDING	(lit) 2.1	13
LIGAND	[3]spiperone	
REFERENCE	Seeman, et al. Mol. Pharmacol.28:391(1985)	
COMMENT	No Na+, active enantiomer	
D2 BINDING	(lit) 20.4	14
LIGAND	[3]spiperone	
REFERENCE	Seeman, et al. Mol. Pharmacol.28:391(1985)	
COMMENT	No Na+, less active enantiomer	

Fig. 8. The two-dimensional structure of a compound and its biological data drawn with GENIE commands.

```
ANUMBER     5422                                                          0
  TIMESTAMP   1987 Dec 29 13:46:29                                        1
RECCAV NAME  MCDERMED DOPAMINE RECEPTORS                                  14
UNION SURF FILE  NONE
CREATOR      FREEMAN AND MCDERMED,ROYAL SOC OF CHEM SPEC PUB 42,154-66, 1982
             MCDERMED AND FREEMAN,ADV BIOSCI,37,179-87,1982
             MCDERMED,FREEMAN, AND FERRIS, CATECHOLAMINES:BASIC AND CLINICAL
                  FRONTIERS, 568-70,1978
MACRO STATEMENTS TO GET MSF FILES  [DANAHER.DOP.MODELS]MCDERMED.MAC
       THOTSF APOMORPHI
       THOTSF 5OHPRADTM
       THOTSF 271TEHMCD
       THOTSF ISOAPOM
       THOTSF RESTEPRM
       THOTSF TL232MMCD
       THOTSF 8PRADTN1
       THOTSF 5PRADTN2
COMMENTS FOLLOW:
      CONF OF PR ON 8PRADTN1 WAS ARBITRARILY CHOSEN;MMP2 MIN.
      CONF OF PR ON 5PRADTN2 WAS ARBITRARILY CHOSEN;MMP2 MIN.
      5 PR IS NOT ALLOWED STERICALLY, NOR IS EXTRA VOLUME OF ISOAPOMORPHINE.

      HE JUST SAYS DOPAMINE RECEPTORS, BUT THEY
      TESTED SPIPERONE AND APO BINDING ON CALF STRIATAL SO IT DOES APPLY
      TO D2.

      BINDING IS TO THE META OH AND N
      STERIC BULK ON RECEPTOR BLOCKS REQUIRED ORIENTATION OF ISOAPOMORPHINE
      SAYS NOTHING ABOUT DIRECTIONALITY OF LPS ON N OR R GROUPS. THEY ARE AS
      YCM ORIENTED THEM. SAYS NOTHING ABOUT THE DIRECTION OF THE OH'S.
```

Fig. 9. The MENTHOR data on the McDermed dopamine receptor map [7].

running GENIE on the MENTHOR. Output such as Fig. 8 can be produced with MedChem software from output generated by GENIE [1].

When coordinates for a new molecule are to be constructed, the MENTHOR can be queried with MERLIN for coordinates of structurally related molecules. For example, one might construct a starting three-dimensional structure for the compound in Fig. 8 from the coordinates of the corresponding primary amine and those of the *N*-propyl group of some compound for which the biologically active conformation has already been selected.

With the addition of another data type, MENTHOR is also a repository of the lists of compounds and coordinate sets that were included in a particular binding site hypothesis, the criteria for compound selection, the location of related union surface files, etc. These are stored on the MENTHOR page of the natural ligand for the macromolecule. Figure 9 shows this information for the D-2 receptor map proposed by McDermed and Freeman [7].

A final example of a MENTHOR search would be to identify and display different proposed binding site hypotheses for the same receptor. The coordinate sets in each would be identified by a THOR FIND command, and each would be brought into the molecular graphics system for comparison. For example, Figs. 10–12 show three different literature binding site hypotheses for

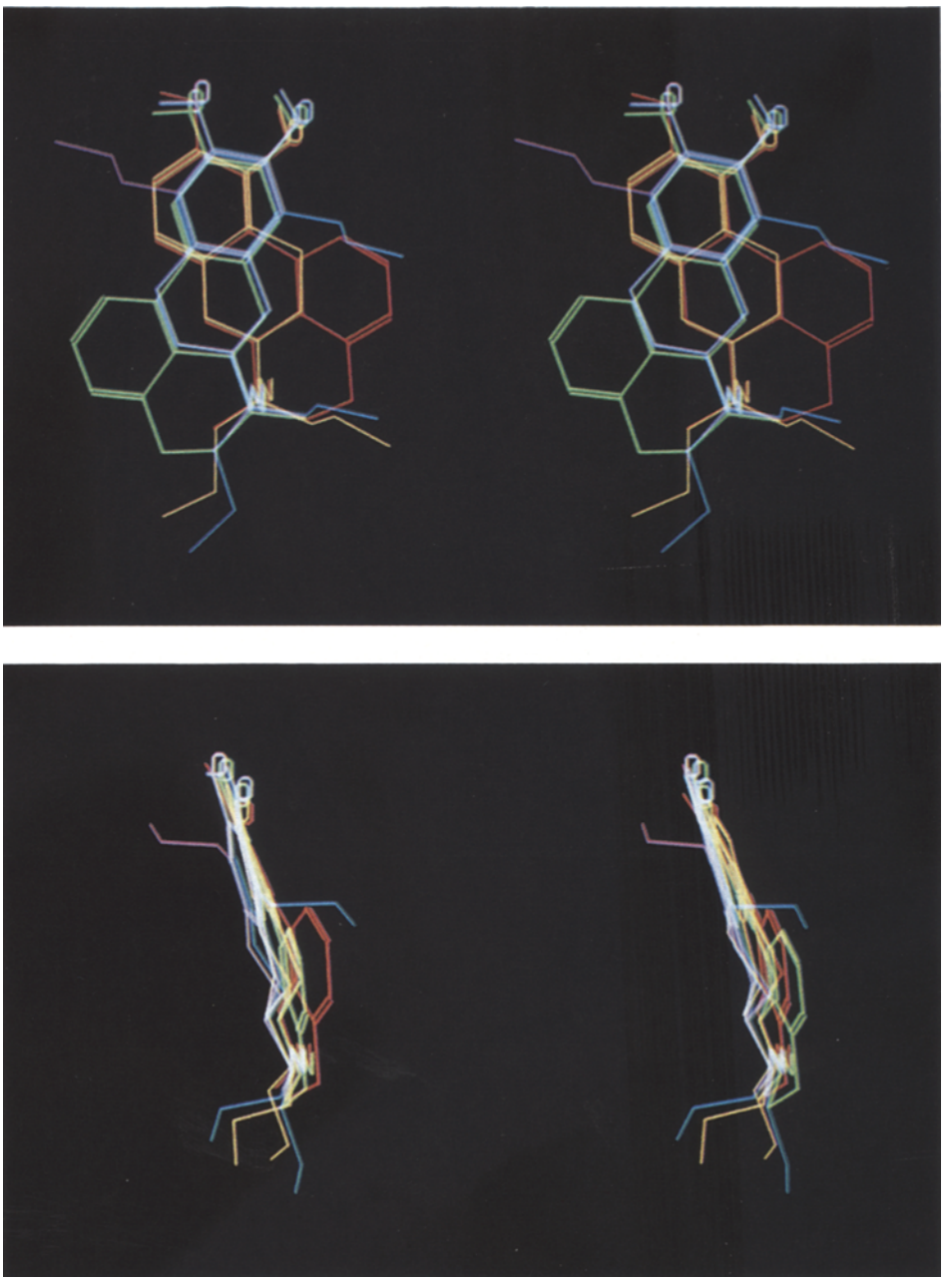


Fig. 10. The D-2 receptor map proposed by McDermed and Freeman [7]. The red compound is inactive, the others active.

D-2 agonists. In each, the coordinates of apomorphine, the viewing angle, and the scale are constant. Of interest first is that the receptor map in Fig. 10 includes only catechol amines whereas the others contain indoles also. Experimental data verified the absolute stereochemistry of the active stereoisomer shown. In contrast, Fig. 11 includes apomorphine and an indole. The superpositions

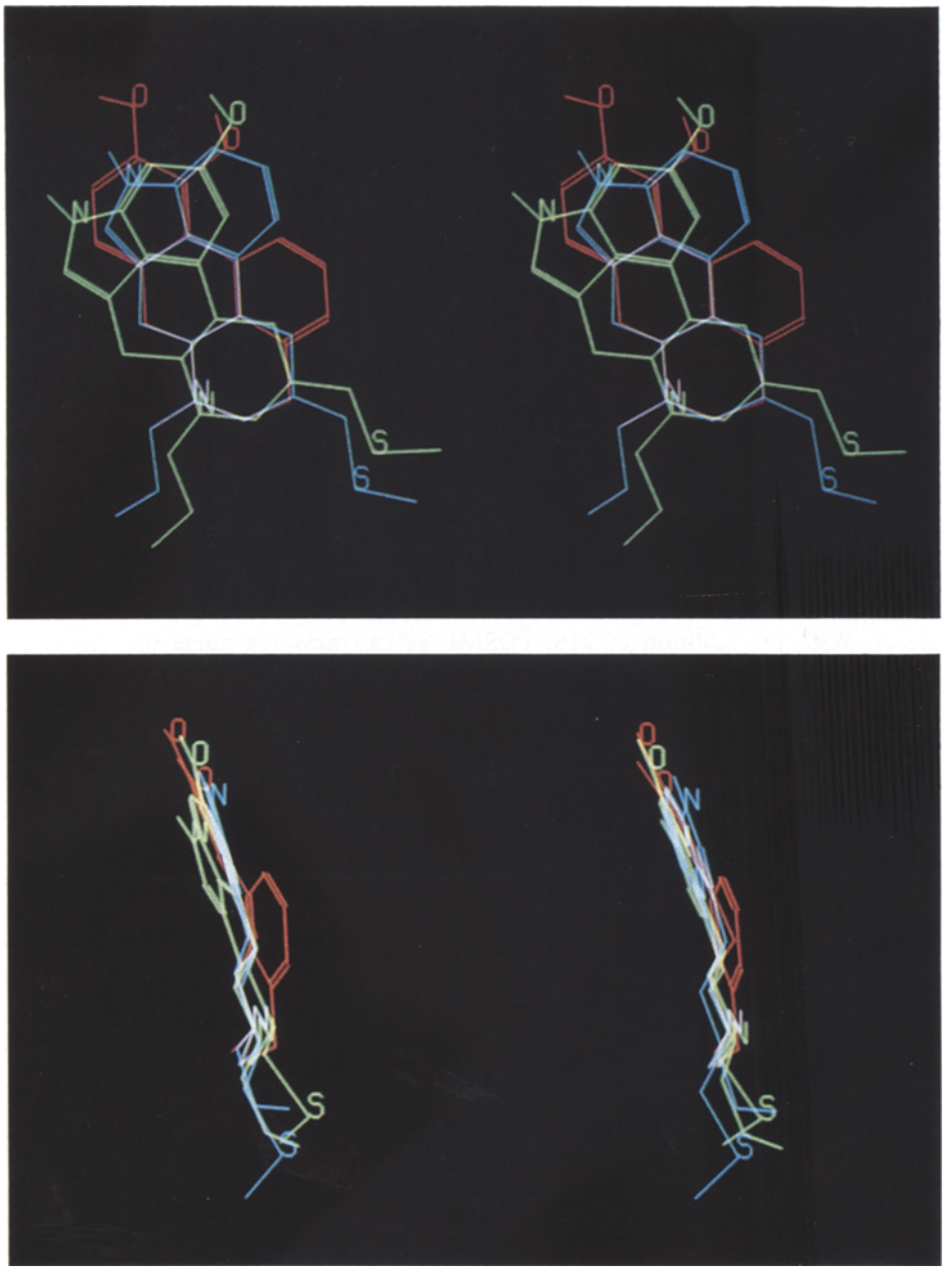


Fig. 11. The superposition of ergolines and apomorphine proposed by Nichols [8].

were chosen to emphasize the requirement for the stereoisomer shown. Note that the hydrogen-bond donors in apomorphine and the indole in the yellow structure do not overlap nor point to the same proposed receptor binding site. In further contrast, Fig. 12 includes only apomorphine and bromocriptine, but two different superpositions are shown. In the superposition shown in

red, the indole ring of bromocriptine is overlapped with the catechol ring of apomorphine as these workers first suggested. Note, however, that the chiral carbon atoms of apomorphine and bromocriptine do not correspond in this superposition. In contrast, the superposition shown in green matches the required stereochemistry but the aromatic rings do not overlap. MENTHOR makes it easy to compare such results since the atomic coordinates for each superposition are retrievable as are the molecular graphics commands to do so. MERLIN substructure searches followed by a mapping of the structures to their data allow one to find similar molecules that have been used by different workers to propose a binding site.

Additional capabilities gained from this work

This project also solved four small but annoying molecular modeling problems. First, we wanted our graphics display to show double bonds. To accomplish this we included MSFTOSMI, similar to MSFTOTDT, into our molecular graphics program to calculate the bond order.

The second problem is the superposition of two molecules of identical two-dimensional structure but different sequence of atoms and different atomic labels. This situation may arise if a given starting structure has been minimized in two different programs or has been constructed from different pieces. With the addition of MSFTOSMI we can now rearrange the atoms into the SMILES sequence. Then a simple root-mean-square superposition, atom-for-atom, accomplishes the required superposition.

Yet another problem in model building or examination of a crystal structure is to know when the object on the screen is a possible molecule or if there is a hydrogen missing from the structure. CMD attempts to calculate the SMILES string of the object any time the connection table is modified: if a SMILES string is printed, the structure is a valid chemical.

Finally, MSFTOSMI is useful for sorting out structures in individual files. For example, if a directory contains molecular structure files for many different compounds, MSFTOSMI may be used to make a summary file of the filename and the SMILES of the structure in that file. A sort of the summary file immediately tells one if there are molecules for which two or more XYZ data sets are present. These may be further examined to eliminate duplicates. Obviously this work is eliminated if MENTHOR is used to store the coordinates.

DISCUSSION

Before we built MENTHOR, we tracked our modeling results in a traditional database. In it were fields for the Abbott number, the file specification of the molecular structure file, and comments. The Abbott number field allowed us to map to databases that contain the Abbott test results, a listing of the parent molecule and indication of the substituents present, and crude two-dimensional pictures of the molecular structure. Thus we could, in principle, accomplish many of the functions of MENTHOR with less ease than with MENTHOR and on a different computer from that on which we did the modeling.

An important issue with such a separate data base is that it is difficult to keep current since it serves no immediate use to the modeler. Moreover, the storage of the actual coordinates and charges (rather than just file specification) in a database helps to prevent corruption or accidental deletion of important information. In contrast, MENTHOR is used as the natural repository for the

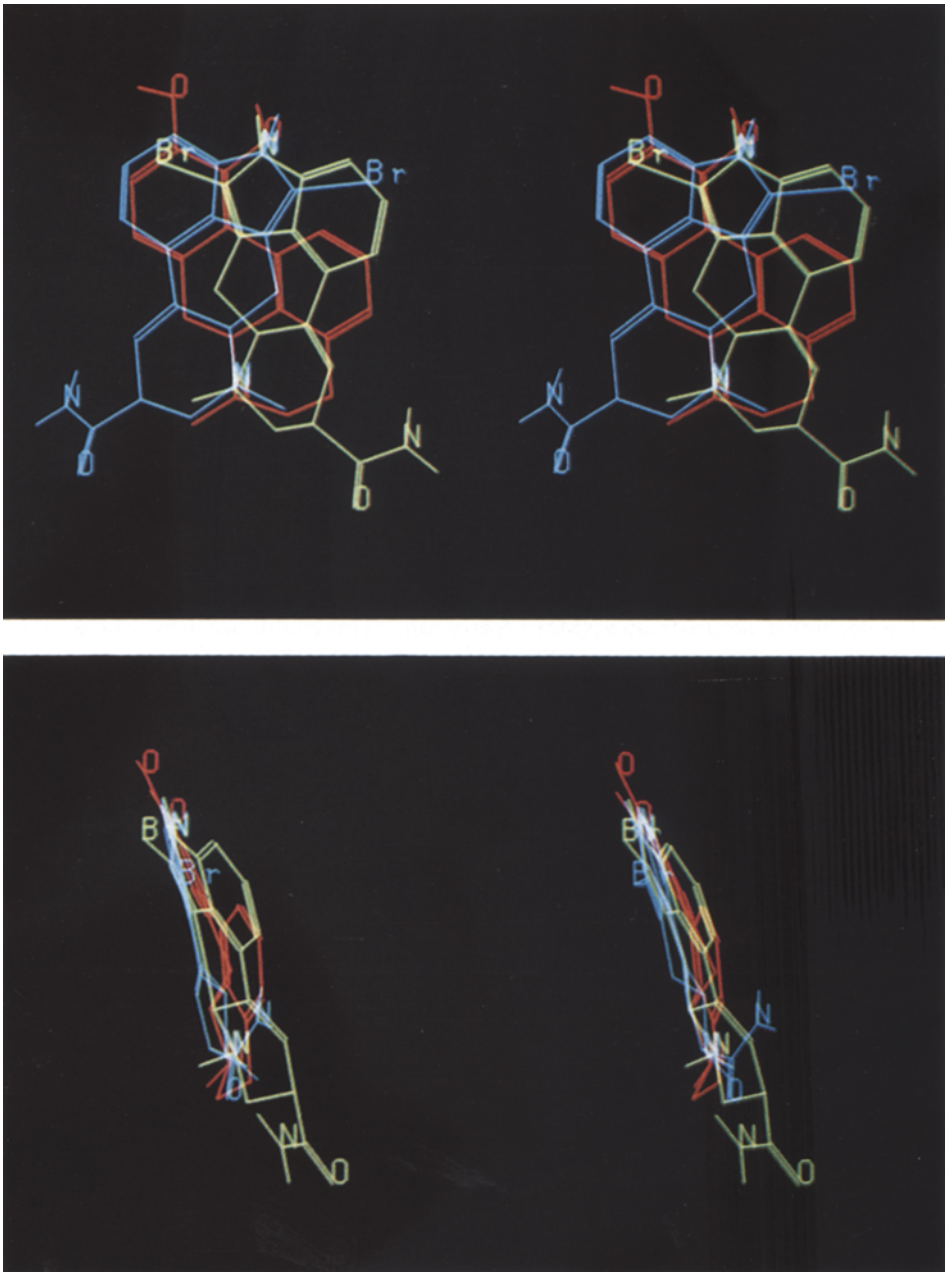


Fig. 12. The two ways to superimpose bromocriptine and apomorphine proposed by Camerman et al. [9]. The bromocriptine superposition shown in red has opposite stereochemistry at the chiral center compared to apomorphine.

coordinates of compounds and the individual structure files are deleted once the coordinates are stored.

A further disadvantage of the original system was that the only way to do sub-structure searching was to key by hand each molecule for specific sub-structures or to do a MACCS search on a computer different from that on which the biological data was stored and that on which our molecular modeling is performed. We did not add literature data to the previous system.

We also experimented with a simple artificial intelligence system to document molecules that had already been modeled and make these results available for new molecules [10]. This system was written in GENIE [1]. Input was a SMILES automatically made from a MACCS structure, and output was a molecular graphics macro. The main problem with this work was that it is time-consuming to write such a program even in GENIE, and one may never need much of the information that has been carefully stored. It is more direct to search a MENTHOR when we need the information. The artificial intelligence system also suffers from the problem of being kept current.

Both MACCS and the Cambridge Crystallographic Data system have sub-structure searching and database facilities. We decided not to build our system on the MACCS software because MACCS is not designed to be accessed by other programs nor are easy interfaces available. We did not base our system on the Cambridge Crystallographic Data because that system has no provision for the addition of non-crystallographic data and there is no correspondence between the topological description of the molecule and the order in which the atomic coordinates are stored.

Recently, we wished to evaluate the program CONCORD [6]. It rapidly generates a three-dimensional molecular structure from a SMILES structure. This evaluation was simple – it involved just five steps, four of which are essentially automatic. First, the SMILES of all compounds in the MENTHOR were written to a file. Second, this file was processed by CONCORD which produced the three-dimensional structures of most of the compounds. Third, the coordinates of each structure were transformed into a THOR data tree by a variant of MSFTOTDT. Fourth, the coordinates were loaded into the MENTHOR. The result is that the CONCORD coordinates are associated in the database with other coordinates for that molecule. The two sets of coordinates may now be compared with a computer program such as ALADDIN [7], or with the usual molecular graphics tools of superposition, etc.

The most exciting aspect of MENTHOR is that it is the first phase of a project to develop software to search the stored three-dimensional structures based on geometric as well as substructural criteria [J.H. Van Drie, D. Weininger and Y.C. Martin, manuscript in preparation]. For example, one might wish to identify existing compounds that satisfy the proposed stereochemical requirements for a particular biological activity. Alternatively, one might wish to identify templates that constrain functional groups to a particular geometric relationship found in one of the low-energy conformations of a known ligand. One key to the success of this project was that the coordinates and charges are stored in the order of the atoms in a unique linear description of the molecule. Thus, once a target substructure in a molecule is identified, the XYZ coordinates of these atoms in each stored conformation may be retrieved. Our work on this project will be described soon [J.H. Van Drie, D. Weininger and Y.C. Martin, manuscript in preparation].

REFERENCES

1 Weininger, D., Weininger, A. and Leo, A.J., MedChem Software Manual, Release 3.52, Med. Chem. Proj., Pomona Coll., Claremont, CA, 1987.

2 Humblet, C. and Marshall, G.R. Three-dimensional computer modeling as an aid to drug design, Drug Dev. Res., 1 (1981) 409–434.

3 Martin, Y.C. and Danaher, E.B., Molecular modeling of receptor–ligand interactions, In Williams, M., Glennon, R. and Timmermans, P. (Eds.) Receptor Pharmacology and Function, Dekker, New York, NY, 1988, pp. 137–171

4 Weininger, D., Chemical structures and computers, In Hansch, C., Sammes, P.G. and Taylor, J.B. (Eds.) Comprehensive Medicinal Chemistry, Pergamon, Oxford, in press.

5 Allen, F.H., Kennard, O. and Taylor R., Systematic analysis of structural data as a research technique in organic chemistry, Acc. Chem. Res., 16 (1983) 146–161.
6 CONCORD was written by A. Ruskinko III, J.M. Skell, R. Balducci and R.S. Pearlman. It is available from Evans & Sutherland Computer Corporation.
7 McDermed, J.D. and Freeman, H.S., Stereochemistry of dopamine receptor agonists, Adv. Biol. Sci., 37 (1982) 179–187.
8 Nichols, D.E., Structural correlation between apomorphine and LSD: Involvement of dopamine as well as serotonin in the actions of hallucinogens, J. Theor. Biol., 59 (1976) 167–177.
9 Camerman, N., Chan, L.Y. and Camerman, A., Stereochemical characteristics of dopamine agonists: Molecular structure of bromocriptine and structural comparisons with apomorphine, Mol. Pharmacol., 16 (1979) 729–736.
10 Martin, Y.C., Danaher, E.B., Weininger, A. and Weininger, D., Abstracts of Papers, Molecular Graphics Society Meeting, Cap d'Agde, April, 1986.