MATCHMOL, an interactive computer graphics procedure for superposition of molecular models

M Cory and J Bentley

Organic Chemistry Department, Wellcome Research Laboratories, Burroughs Wellcome Co, Research Triangle Park, NC, USA

A procedure that facilitates the interactive superposition of computer models of molecules was written. This procedure, a rigid body fit, used a fitting algorithm described by McLachlan and the molecular graphics facilities provided by the NIH — PROPHET system. The flexibility of the PROPHET molecule data type allows easy generation of hypothetical atoms (pseudoatoms) which can serve as positional descriptors of putative pharmacophores. Definition of pseudoatoms which represent the centre of polycyclic aromatic chromophores and the superposition of two DNA binding compounds is used as an example of an application of the MATCHMOL procedure.

Keywords: molecular superposition, PROPHET, MATCHMOL

received 15 March 1984, revised 3 April 1984

The receptor-based approach to drug design uses knowledge of the molecular structure of the putative drug receptor as a basis for generation of likely structural targets. The goal is a compound that will bind more effectively to the receptor and will consequently be a more effective drug[1], assuming equal pharmacological properties. When the drug receptor is a protein with a known 3D structure, elegant molecular modelling approaches can lead to the design of drug analogues which bind to specific ligands on a protein[2]. However, the receptors for most drugs are proteins of which the 3D structure is not known. These protein receptors are usually membrane bound and their sequences are not known. In these cases, studies of molecular models of the ligands can be used to develop a model of the drug receptor site and to identify the pharmacophoric groups on the drugs.

Examples of early efforts to develop pharmacophoric maps of enzyme receptors can be seen in the planar polycyclic models developed by Baker[3] and the pseudo-3D block models[4] which attempt to define an enzyme receptor site.

Molecular models in three dimensions (which can now be rapidly manipulated using the power of computer graphics) and mathematical processing of the shapes of the molecules are the successors to these early 2D paper efforts. They can give insight into the shape of an unknown drug receptor[5,6].

A graphically-based molecular model superposition program would make easier the superposition of molecules and would allow rapid and interactive matching of diverse molecules. Such a program could be designed to take advantage of a graphics tablet and the displayed structure of the molecules as input for the specification of the atom pairs to be matched. The user would not have to be very familiar with the computer representation of molecular data to use the graphics tablet in a program, but could simply pick the matching atoms from the structural diagram displayed on the screen.

IMPLEMENTATION

The PROPHET system

One of the facilities provided by the NIH-PROPHET computer graphics system is a defined single-variable molecule data type (Figure 1)[7]. This provides great flexibility for the medicinal chemist who uses PROPHET molecules, as the data type allows easy modification of the molecular model. Atoms or bonds can be added to, or deleted from, the molecule with interactive commands. New molecules can be made by selecting sets of atoms or bonds in existing molecules and creating a new instance of the data type.

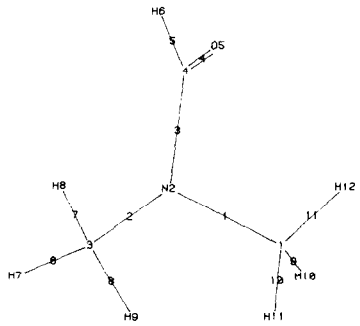
Large sets of molecular models can also be easily manipulated in PROPHET. PROPHET provides a multiple molecule data type as a single variable. A multi-molecule consists of one or more than one instance of the molecule data type along with the coordinate transformation matrices necessary to map the coordinates of the individual molecules into the same display space. Characteristics of the molecules such as atom cartesian coordinates, atom connectivity, atom type, and display selection criteria for the individual molecules that contribute to the multimolecule, are retained within the multi-molecule data type[3]. The easy access to these characteristics provides the user with the ability to query the multi-molecule model for intramolecule or intermolecule geometric information.

ATOMS 12R X 7C

	1 TYPE	2 BOND ARRAY	3 SKETCH X	4 SKETCH Y	5 MODEL X	6 MODEL Y	7 MODEL Z
1	C	[1,9,10,11]	7.62	-3.32	0.00	0.00	0.00
2	N	[1,2,3]	6.70	-4.76	-1.49	0.00	0.00
3	C	[2,6,7,8]	8.18	-6.22	-1.95	-.74	-1.21
4	C	[3,4,5]	4.96	-4.84	-1.95	-.74	1.21
5	O	[4]	3.90	-3.84	-1.65	-1.91	1.45
6	H	[5]			-2.58	-.09	1.82
7	H	[6]			-2.58	-.09	-1.82
8	H	[7]			-2.53	-1.63	-.89
9	H	[8]			-1.08	-1.07	-1.79
10	H	[9]			37	1.04	0.00
11	H	[10]			37	-.52	-.90
12	H	[11]			37	-.52	.90

BONDS 11R X 3C

	1 TYPE	2 ATOM 1	3 ATOM 2
1	S	1	2
2	S	2	3
3	S	2	4
4	D	4	5
5	S	4	6
6	S	3	7
7	S	3	8
8	S	3	9
9	S	1	10
10	S	1	11
11	S	1	12



Figue 1. Connection table portion of the PROPHET molecule data type for dimethylformamide. Above: atoms table with one row for each atom in the molecule. Below left: bonds table, with one row for each bond. Below right: display of the molecule with atom and bond indices shown

The information which is of primary importance to the medicinal chemist when comparing sets of analogous molecules in the series of molecules being studied is the distances between atoms or groups of atoms which might be acting as pharmacophores. These can be represented as the distances between the individual points or as the union or exclusion of volumes[8]. The multi-molecule facility is quite useful in attempts to identify the pharmacophoric groups in sets of molecules. FITMOL[9] was available in the PROPHET system as a PUBLIC procedure for the superposition of sets of molecular models. Once the sets of molecular models were transformed into the same coordinate space a multimolecule could be generated interactively from the individual molecules. The input for FITMOL was a list of atom indices representative of the row of the PROPHET molecule data table (Figure 1) containing the characteristics of each atom. This procedure has been used quite effectively by Rohrer[10] and Fullerton[11] to develop a map of the cardenolide receptor by superimposing the structurally similar portions of sets of molecular models which bind to the receptor. The sources for such molecular models can include atomic coordinates obtained directly from crystallographic studies or conformational energy calculations.

We found the FITMOL program quite useful when very similar molecules with consistent atom indexing or numbering schemes were used. However, when used to superimpose sets of molecules in which the pharmacophores were of relatively diverse type and the molecules were stored in the computer with different numbering schemes, the input to the program required extensive book-keeping for the superposition of series of molecules. This is a problem with diverse molecules, since a molecular modelling superposition program should allow a user to test rapidly various hypotheses when matching pharmacophores. The Cambridge Crystal file[12,13] has recently become the primary source of crystal coordinates for the manipulation of molecular models on the PROPHET system[14]. The Cambridge file stores the atoms of each molecule in the order presented by the original author instead of using a standardized numbering scheme.

Pseudoatoms

Pseudoatoms can be defined in PROPHET by use of a procedure called ADDCENTER. ADDCENTER allows graphic specification of sets of atoms and then adds a row to the atoms table with model coordinates computed as the geometric centre of the specified set of atoms. The atom type of the pseudoatoms is set to *X* so that the specified centres can be identified readily in the molecule display (see Figure 2). Another procedure UNCENTER allows removal of the pseudoatoms.

MATCHMOL

As the first part of the MATCHMOL implementation, an efficient error-free algorithm for superposition of the molecular coordinates was desired. A procedure, ROTATE, was implemented in the PROPHET system by Art Gottleib and Eric Meyer of Bolt Berenek and Newman, the PROPHET system operators. This procedure used the algorithm recently published by

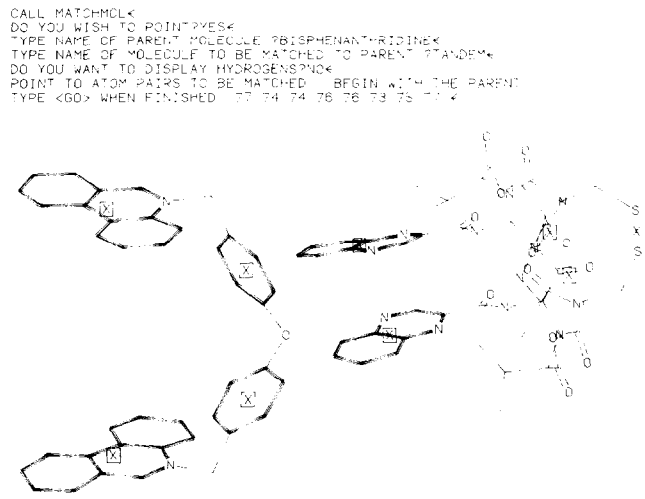


Figure 2. Initiation of MATCHMOL. This shows a call to the procedure, followed by the interactive dialogue that obtains the molecule variable names and sets the display parameters. The left molecule is a bisphenanthridinium (BPA), DNA binding molecule[16]. The right molecule is TANDEM, a synthetic analogue of the DNA double intercalating compound Echinomycin A. It contains two quinazoline ring systems connected by a bicyclic nonapeptide. Coordinate data was obtained from crystal structures[17]. Selected atoms for the match are picked using a graphics tablet and pen. Atom indices are typed on the screen and a box is drawn around the selected atoms. The pseudoatoms, indicated by an X, represent the centres of each of the aromatic ring systems in the BPA molecule and the centres of the quinazoline ring systems, the disulphide bond and the cyclic polypeptide rings in TANDEM

McLachlan[15]. Molecules of various sizes, including proteins, are available in PROPHET. Because of this, the McLachlan algorithm, which was specifically designed to work rapidly on large molecules such as proteins, was seen as particularly attractive. The superposition subroutine was then incorporated into the full interactive graphics program.

The user interaction with the computer begins with the call to the procedure followed by a dialogue (Figure 2) necessary for setting the parameters of the display. The option of nongraphic specification of atom indices was retained for classes of molecules which are numbered similarly or molecules which are too big or complicated for easy graphic display. The molecule variable names are entered in response to the prompts. The first molecule indicated by the user is designated arbitrarily as the parent, and the second molecule is fitted to that parent in a pairwise manner, matching atom for atom for an indicated set. Hydrogen atoms can be suppressed if the user so desires.

Figure 2 also displays the screen after four pairs of atoms have been picked. The atoms are picked in sequence, one from the parent and one from the daughter molecule. A box is drawn around the chosen atoms to confirm the pick visually, and the atom index (row number of the atoms table for the molecule) is typed on the screen.

After at least three pairs of atoms are indicated by the user, an option allows weighting of the atom pairs prior to the fitting (Figure 3). The additional dialogue

```
DO YOU WANT TO ASSIGN WEIGHTING FACTORS TO THE ATOM PAIRS?NO<
DO YOU WANT TO SEE AN OVERLAID DISPLAY OF THE TWO MOLECULES?YES<
DO YOU WANT A TOP-LEVEL COPY OF THE MULTIMOLECULE?YES<
WHAT NAME DO YOU WANT FOR THE MULTIMOLECULE?TANBISPHE<
```

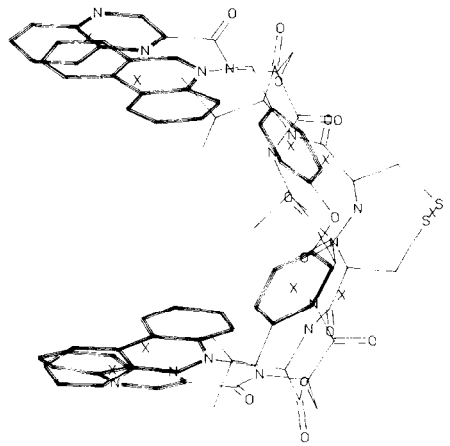


Figure 3. Interaction following the choice of atoms to be matched. The display at this point shows the pair of molecules as a Prophet multi-molecule display. The pseudoatoms are the same as in Figure 2

```
LIST INDIVIDUAL INTERATOMIC SEPARATIONS?YES<
WHAT NAME DO YOU WANT FOR THE TABLE OF INTERATOMIC DISTANCES?TANPHE
```

TANPHE: MATCH OF BISPHENANTHRIDINE AND TANDEM 4R X 6C

	1 PARENT ATOM NAME	2 MATCHED ATOM NAME	3 PARENT ATOM INDEX	4 MATCHED ATOM INDEX	5 INTER-ATOMIC DIST	6 USED IN MATCH?
1	CENTER	CENTER	77	74	2.148	YES
2	CENTER	CENTER	74	76	1.32	YES
3	CENTER	CENTER	76	73	1.138	YES
4	CENTER	CENTER	75	77	1.533	YES

Figure 4. Interaction that generates the PROPHET table display containing information on the matched atoms

```
LIST ANY OTHER INTERATOMIC SEPARATIONS?YES<
DO YOU WANT TO DISPLAY HYDROGENS?NO<
POINT TO THE ATOM PAIRS FOR DISTANCE CALCULATION
BEGIN WITH THE PARENT
TYPE <GO> WHEN FINISHED  3 20 2 26 1 75 <
```

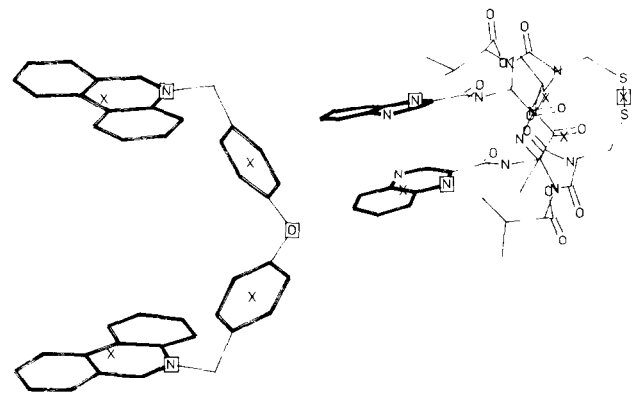


Figure 5. Interaction allowing choice of additional atoms for addition to data table

TANPHE: MATCH OF BISPHENANTHRIDINE AND TANDEM 7R X 6C

	1 PARENT ATOM NAME	2 MATCHED ATOM NAME	3 PARENT ATOM INDEX	4 MATCHED ATOM INDEX	5 INTER-ATOMIC DIST.	6 USED IN MATCH?
1	CENTER	CENTER	77	74	2.148	YES
2	CENTER	CENTER	74	76	1.32	YES
3	CENTER	CENTER	76	73	1.138	YES
4	CENTER	CENTER	75	77	1.533	YES
5	N2	N6	3	20	2.153	NO
6	N1	N'6	2	26	1.523	NO
7	O2	CENTER	1	75	2.999	NO

Figure 6. Final Prophet data table containing both atoms used in generating the matched multi-molecule and other specified interatomic distances

in Figure 3 is used to get a name for the multi-molecule. This name is needed if the user wishes to save the superimposed molecules as a multi-molecule for further work.

Figure 4 displays the initial version of the PROPHET data table. It includes the names and atom indices of the matched atoms and the intermolecule through space distances for the matched pairs of atoms. These distances are computed by the system from the molecule coordinates and presented in Ångström units. An additional column is added which indicates if the particular atom pair was used in determining the match.

Figure 5 shows the interaction which allows selection of additional pairs of atoms for interatomic distance calculation. The atom names, atom indices, and distances are added in rows to the data table. In these cases, column 6 is set to 'no' as these atom pairs were not used in the original match.

Figure 6 shows the final version of the output data table. This data table and the multi-molecule are available to other PROPHET procedures when the MATCHMOL procedure is finished.

While in the PROPHET environment, this graphics procedure uses considerably more CPU time than the old FITMOL procedure. The CPU time is primarily used in the parts of the program that make the procedure interactive, such as the coordinate scaling necessary for the multiple molecular displays, and the generation and display of the data tables used as output.

CONCLUSIONS

This procedure has provided a substantial increase in molecule handling capability for the PROPHET system. The graphic interaction allows the medicinal chemist to use his well-developed capacity to recognize chemical patterns to fit multiple molecules into the same space. Multiple hypotheses about different pharmacophoric overlaps can be tested rapidly. Multiple sets of superpositions can be compared for overall fit by comparison of the data in the output tables. We have used this procedure to generate sets of up to 10 superimposed molecules.

In practice many sets of similar molecules can be overlapped in many ways, so it is important to have an easily-used graphics procedure for testing superposition hypotheses rapidly.

REFERENCES

1 **Humblet, C and Marshall, G R** *Ann. Repts. Med. Chem.* No 15 (1980) p 267

2 **Kuyper, L F, Roth, B, Baccanari, D P, Ferone, R, Beddell, C R, Champness, J N, Stammers, D K, Dann, J G, Norrington, F E A, Baker, D J and Goodford, P J** *J. Med. Chem.* No 25 (1982) p 1120

3 **Baker, B R and Wood, W W** *J. Med. Chem.* No 11 (1968) p 644

4 **Baker, B R, Lee, W W, Skinner, W A, Martinez, A P and Tong, E** *J. Med. Pharm. Chem.* No 2 (1963) p 633

5 **Weaver, D C, Barry, C D, Mcdaniel, M L, Marshall, G R and Lacy, P E** *Mol. Pharmacol.* No 11 (1975) p 833

6 **Marshall, G R, Barry, C D, Bosshard, H E, Dammkoehler, R A and Dunn, D A** in **Olsen, E C and Christoffersen, R E (eds)** *Computer assisted drug design* ACS Symposium Series No 112 American Chemical Society, Washington DC, USA (1979) pp 205–226

7 **Rindone, W P and Kush, T (eds)** *Prophet molecules* Bolt Beranek and Newman, Cambridge, MA, USA (July 1980)

8 **Sufrin, J R and Marshall, G** *Federation Proc.* No 38 (1979) p 562

9 **Rohrer, D C and Smith, G D** *FITMOL in PROPHET molecules* **Rindone, W P and Kush, T (eds)** Bolt, Beranek and Newman, Cambridge, MA, USA, (1980) pp 5–47

10 **Rohrer, D C, Fullerton, D S, Yoshioka, K, From, A H and Ahmed, K** in **Olson, E C and Christofferson, R E** *Computer assisted drug design* ACS Symposium Series, No 112, American Chemical Society, Washington DC, USA (1979) pp 259–279

11 **Fullerton, D S, Yoshioka, K, Rohrer, D C, From, A H and Ahmed, L** *Science* No 205 (1979) pp 917–919

12 **Kennard, O, Allen, F H, Brice, M D, Hummelink, T W A, Motherwell, W D S, Rodgers, J R and Watson, D G** *Pure Appl. Chem.* No 49, (1977) pp 1807–1816

13 **Wilson, S R and Huffman, J C** *J. Org. Chem.* No 45 (1980) pp 540–566

14 **Meyer, E J** *CAMCRYST users manual* Bolt, Beranek and Newman, Cambridge, MA, USA (1983)

15 **McLachlan, A D** *J. Mol. Biol.* No 128 (1979) pp 49–79

16 **Cory, M, McKee, D D, Kagan, J and Henry, D W** Manuscript in preparation

17 **Hossain, M B, van der Helm, D, Olsen, R K, Jones, P G, Sheldrick, G M, Egert, E, Kennard, O, Waring, M J and Viswamitra, M A** *J. Amer. Chem. Soc.* No 104 (1982) p 3401