Journal of Computer-Aided Molecular Design, 3 (1989) 225–251
ESCOM

J-CAMD 068

ALADDIN: An integrated tool for computer-assisted molecular design and pharmacophore recognition from geometric, steric, and substructure searching of three-dimensional molecular structures

John H. Van Drie[a], David Weininger[b] and Yvonne C. Martin[c,*]

[a] *Meridien Research, Oak Park, IL 60301, U.S.A.*
[b] *Daylight Chemical Information Systems, Irvine, CA 92714, U.S.A.*
[c] *Abbott Laboratories, Computer Assisted Molecular Design Project, D-47E, AP9, Abbott Park, IL 60064, U.S.A.*

Received 9 February 1989
Accepted 7 June 1989

Key words: Dopaminergics; Three-dimensional databases; Peptide mimics; Computer-aided molecular design; Geometric searching; Pharmacophore recognition

SUMMARY

ALADDIN is a computer program for the design or recognition of compounds that meet geometric, steric, and substructural criteria. ALADDIN searches a database of three-dimensional structures, marks atoms that meet substructural criteria, evaluates geometric criteria, and prepares a number of files that are input for molecular modification and coordinate generation as well as for molecular graphics. Properties calculated from the three-dimensional structure are described by either properties calculated from the molecule itself or from the molecule as compared to a reference molecule and associated surfaces. ALADDIN was used to design analogues to probe a bioactive conformation of a small molecule and a peptide, to test alternative superposition rules for receptor mapping of the D2 dopamine receptor, to recognize unexpected D2 dopamine agonist activity of existing compounds, and to design compounds to fit a binding site on a protein of known structure. We have found that series designed by ALADDIN show much more subtle variation in shape than do those designed by traditional methods and that compounds can be designed to be very close matches to the objective.

INTRODUCTION

The new computer tools of molecular graphics and molecular modeling as well as the continuing growth in the solution of three-dimensional protein structures promise increasing accuracy in the design of biologically active small molecules to be used as drugs. Such rational drug design re-

*To whom correspondence should be addressed.

quires a hypothesis on the relationship between three-dimensional structural features of a molecule and its biological activity. This hypothesis may be developed either from the three-dimensional structure of the target macromolecule (usually a protein) [1], or from 'receptor mapping' of the three-dimensional structure-activity relationships of the ligands that bind to the target, sometimes with the aid of NMR or fluorescence measurements [2]. For 'receptor mapping,' one typically needs first to establish the bioactive conformations of the ligands. This is proposed from the biological properties of conformationally defined molecules that hold the critical functional groups (previously identified from structure-activity relationships) in the positions characteristic of one of the several low energy conformations of the least flexible active ligand.

The practice of molecular computer graphics is quite developed [3], the techniques of molecular mechanics and quantum mechanics have been explored for decades [4], and very helpful tools are available for examining structure-activity relationships to derive a hypothesis of the three-dimensional requirements for biological activity [5,6]. However, relatively less attention has been paid to the problem of how to design compounds that meet certain geometric, steric, or conformational constraints.

The problem of designing molecules that bind to a structurally known binding site has been studied. Goodford and others have described computer-based methods that identify the types and relative orientations of atoms that are optimal for binding to a macromolecule [7]. However, the scientist and not the computer proposes new compounds consistent with this information. Kuntz and colleagues have described an algorithm and computer program for the recognition of small molecules that are complementary in shape to a binding site of a macromolecule [8]. However, the chemical properties of the ligand are not considered, so this approach again requires extensive human intervention in compound design. Additionally, if the potential ligand has a small substituent that makes it too large for the binding site, the compound will be rejected. Conversely, molecules that are too small but that could be modified to fit a binding site are not recognized. Thus, neither of these methods meets the needs of a traditional medicinal chemistry program.

Closer to our own work are programs, developed simultaneously with ALADDIN, that search a database of three-dimensional structures based on geometric criteria [9]. That work has concentrated on algorithms to enhance search performance. For our needs, the disadvantages of that work are: (1) the target atoms cannot be constrained to be in a specific chemical environment as precisely as with the use of GENIE SMARTS targets described below nor is the choice of geometric objects so large; (2) the programs are not closely tied to molecular modeling and molecular graphics; (3) the programs do not suggest and evaluate possible new compounds for synthesis; and (4) no provisions have been made to test the fit of a given three-dimensional structure into an experimental or theoretical binding site. As will be discussed later, we have found that the latter point cannot be adequately tested using criteria based on the small molecule alone. Rather, the small molecule must be properly oriented within the binding site and then the fit assessed.

Thus, we needed a computer program that would explicitly suggest for chemical synthesis molecules consistent with the structure of a macromolecular binding site or with the accessible conformations of a flexible ligand. We specifically wanted a program that would propose many molecules of varied molecular backbone to allow synthetic ease and a variety of shapes to play a role in the final choice of molecules to synthesize. The first use of such a program would be to design or to recognize known conformationally constrained analogues to explore the possible bioactive conformation of a flexible ligand. Since any set of newly added conformational constraints may

not be sterically allowed in the ligand-macromolecule complex, a variety of analogues must be suggested and tested biologically. Once the bioactive conformation is established, such a program would be used to suggest or recognize existing compounds that maintain the geometric relationships between the key functional groups but vary the shape of the rest of the molecule. Synthesis and testing of such a set of molecules would provide the data for a quantitative structure-activity relationship (QSAR) analysis with the three-dimensional properties as descriptors as described by Cramer et al. [6]. Of course, if the original 'lead' were a competitor's compound, such an analysis would provide a method to design a novel agent not covered by their patents. A similar use for such a program would be to recognize the potential of an existing compound to show bioactivity in previously unexplored areas. This could refine the hypothesis of the molecular requirements for bioactivity and/or provide a novel chemical structure for chemical modification. Such a use is analogous to those of traditional chemical substructural search techniques. The final use for such a program would be to design molecules that position groups appropriately to bind to a macromolecular binding site of known three-dimensional structure and that do not collide with the site when so positioned. Thus, our proposed computer program would have uses in every facet of computer-aided molecular design.

We have recently developed a new program, ALADDIN, that allows one to describe the subtle three-dimensional requirements for activity as theoretical medicinal chemists describe them. The strength of intermolecular interactions, such as that between a drug and its target biomacromolecule, may be considered to result from the sum of favorable and unfavorable steric and electrostatic interactions. ALADDIN uses a precise geometric description language to define the optimal steric properties of a designed molecule and includes the provision to test the molecule in a standard coordinate system. Currently, ALADDIN uses a chemical topological or substructural language to describe the electrostatic environment of the binding points of the molecule.

Since not all of the methods of computer-assisted molecular design have been perfected, one of our first design objectives was that the program be flexible. For example, we might choose to describe molecular features in terms of points, lines, and planes and to set up the search in these terms. Alternatively, since molecular and union surfaces provide powerful visual descriptors of intermolecular interactions, we expected to need to include such descriptions also. A second example of the requirement for flexibility is the source of small molecules for consideration. Although ALADDIN searches a database of previously modeled structures, we have provided tools for the computer to suggest new molecules. A third type of flexibility required was in the definition of what constitutes an acceptable match between the ideal target and the found molecule. We didn't know if pass/fail criteria would be sufficient or if more quantitative scoring would be more useful. Thus, in contrast to programs designed for more mature scientific research areas, the ultimate functionality of ALADDIN could not be designed but rather an open and flexible framework for research was required.

Our second requirement was the ability to specify the chemical environment of atoms or groups in molecules in either great specificity or great generality. We anticipated that, for example, one should be able to identify 'amide nitrogen atoms in which both the nitrogen and its attached carbonyl carbon atom are part of at least two rings of 5, 6, or 7 aliphatic atoms' as well as 'any aliphatic atom in a 5, 6, or 7-membered ring or attached to such a ring.' We further required that the substructural features not be preselected when the database is built or the program written, but that they be freely selectable at run time.

To ensure flexibility, our third requirement was that the program be based on calculations on the actual coordinates, not on precalculated bitmaps. For example, we needed to access the coordinates in order to calculate a root-mean-square deviation between a candidate molecule and some standard molecule; to calculate the torsion angle between a lone pair on a nitrogen, the nitrogen itself, and two atoms of another functional group in the molecule; or to calculate the distance between every atom in a molecule and an arbitrary surface positioned relative to certain atoms of the molecule. By emphasizing the direct use of atomic coordinates, we retain the flexibility to define new types of geometric and steric criteria to ALADDIN.

Our implementation of ALADDIN is based on the MedChem software [10], particularly GENIE. GENIE is a powerful language for substructure specification, recognition, and enumeration. By building upon it, we incorporate the chemical environment into the search description. We decided to build ALADDIN on this software because it already included facilities to identify specific atoms based on substructural environment, to store and return three-dimensional structures, and to easily add new routines. In addition, a powerful substructure searching program and full database facilities are available. In the current version of ALADDIN, coordinates are retrieved from one of our databases of structural and biological information [11]. ALADDIN also uses files from and produces files for our integrated molecular graphics and chemical/structural database program.

As noted above, the ALADDIN language incorporates both the steric fit to the binding site and specification of the substructural environment of the atoms involved. After GENIE has identified the atoms that match the substructural criteria, the geometric criteria and then steric fit are assessed. The geometric and steric specifications are in subroutines, so it is straightforward to add new types of shape specifications.

Viewed within the overall context of computer-assisted drug design, when designing drugs from the three-dimensional structure of the biomolecular target, ALADDIN is a complement to software such as GRID [7] that identifies the chemical properties for ligand atoms to interact with potential binding sites on the target. When drugs are designed from receptor mapping, ALADDIN is a complement to molecular graphics software and routines which compute union and intersection volumes.

THE ALADDIN LANGUAGE FOR GEOMETRIC, STERIC, SUBSTRUCTURAL SEARCHES

Currently, an ALADDIN search is described using the following elements: required atoms or chemical groups in their topological molecular environment (these are SMARTS targets of GENIE); required distances or angles between these SMARTS targets; and the proposed or known locations of the boundaries of the binding site.

Description of the topological environment of an atom

GENIE is a substructure specification, recognition, and enumeration language [10]. The two key language elements of GENIE are SMILES, a linear notation of the chemical topology of a molecule, and SMARTS, a linear notation of target substructures within a SMILES. Features of GENIE key to the implementation of ALADDIN are described in the following paragraphs.

For any molecule, a unique SMILES can be generated [12]. For example, the unique SMILES of catechol is Oc1ccccc1O. (The lower case atomic symbols refer to an aromatic atom, and the numbers refer to atoms that are connected to form a ring. Hydrogen atoms need not be explicitly indicated.) The unique SMILES forms the key into the database that contains biological activity data, the three-dimensional structure, and other experimental or modeling data about the compound [11].

SMARTS is an extension of SMILES that allows one to symbolically specify substructures. For example, the SMARTS target that describes a phenolic oxygen (an aliphatic oxygen connected to an aromatic carbon atom and to a hydrogen atom) is [O;H]c (the square brackets enclose the definition of one atom, in this case an aliphatic oxygen with one hydrogen attached). Two hits would be recorded in the above catechol for this SMARTS target, since there are two such phenolic oxygens. No hits would be recorded on the SMARTS target [O;H]c for compounds such as formaldehyde, furan, quinone, or alcohols.

The power of SMARTS is apparent when one needs to specify a complex chemical environment. For example, to specify phenolic oxygens connected to six-membered all-carbon aromatic rings, we would specify as a SMARTS target [O;H]c1ccccc1. Further, if one were to require a phenolic oxygen connected to a six-membered all-carbon aromatic ring with a nitrogen para to the oxygen, one would specify as a SMARTS target [O;H]c1ccc(N)cc1. Alternatively, if it didn't matter if the aromatic atoms were carbons or not, the SMARTS specification would be [O;H]a1aaa(N)aa1 for a six-membered ring or [O;H]a:a:aN (the 'a' specifies an aromatic atom and the ':' specifies an aromatic bond) if the ring size didn't matter. An alternative specification for the latter SMARTS target would be [O;H]c:[c,n,o]:[c,n,o]:cN in which the commas designate 'or.' The power of SMARTS is further amplified by the ability to define a particular atom by its environment and to use such definitions in subsequent definitions. The SMARTS target for 'any heavy atom' is *.

Two enhancements have been added to SMARTS for the development of ALADDIN: one is the ability to specify the hydrogen attached to an atom; the other is to specify the lone pair on a nitrogen or those on an oxygen. These are specified simply as:

H-normal SMARTS target: for example, H-Oc, and
LP-normal SMARTS target: for example, LP-N(C)(C)C.

Distance and angles between atoms or chemical groups

It is straightforward to identify the coordinates of interest since GENIE identifies the atoms associated with a particular SMARTS target and the coordinates have been stored in the order of the atom in the unique SMILES (with hydrogens last and in the order of the heavy atom to which they are attached) [11].

There are two ways to describe the required shape properties. The first uses a relative coordinate system and geometric quantities independent of the rotation and translation as well as the absolute stereochemistry of the molecule. The second uses an absolute coordinate system. For the latter, molecules must be rotated and translated to a standard orientation before any tests are applied. ALADDIN uses both coordinate systems. The relative coordinate system is used first since it is computationally simpler. Molecules that pass this first stage may then be tested in an absolute coordinate system.

Some geometric quantities related to the conformation of a candidate molecule that can be calculated in the relative coordinate system are: (1) distances between points (atomic positions, centers of mass of chemical groups, or points calculated from the former); (2) the angles between points as described above; and (3) least-squares planes (i.e., a plane which minimizes the distances from a set of points).

To use ALADDIN one first defines the chemical groups with GENIE. One then describes how the geometric objects (points, planes or lines) are to be derived from these chemical groups. The language for this is as follows:

[object-type] [SMARTS target]

where object-type and the coordinates assigned are one of the following:

POINT [TARGET]: the coordinates of the first atom of the SMARTS target;

POINT LP-[TARGET]: the calculated coordinates of the lone pair;

POINT CM-[TARGET]: the coordinates of the calculated center of mass of the atoms;

POINT H-[TARGET]: the coordinates of the hydrogen atom attached to the target;

POINT (1) (2,a) (3,b): the coordinates of a point in the plane or line defined by geometric object 1 at a distance a from point 2 and b from point 3;

PLANE [TARGET]: the least-squares plane of the atoms target;

PLANE [TARGETa] [TARGETb] [TARGETc]: the plane defined by the coordinates of the first atom of each of the targets;

PLANE (1) (2) (3): the calculated plane defined by the 3 points defined by geometric objects 1, 2, and 3 (numbers to the geometric objects are assigned sequentially as they appear in the input);

PLANE OF MOLECULE: the calculated least-squares plane to all atoms in the entire molecule;

LINE [TARGETa] [TARGETb]: the line between the coordinates of the first atoms of the targets;

LINE (1) (2): the line between geometric objects 1 and 2.

Each geometric object defined this way is sequentially numbered in the ALADDIN search description. Constraints on distances and angles are simply specified as:

DISTANCE(n,m)	low limit	high limit
ANGLE(i,j,k)	low limit	high limit
TORSION(i,j,k,l)	low limit	high limit
PLANE-ANGLE(o,p)	low limit	high limit

where i,j,k,l,m,n,o,p refer to the sequential numbers of the geometric object of interest. For DISTANCE, n and m may be both POINTs; one POINT and one LINE or PLANE; or, if n is a POINT or a PLANE, m may be an * in which case the smallest distance between any atom in the molecule and object n will be used. For ANGLE and TORSION, the geometric objects are POINTS; for PLANE-ANGLE, both geometric objects are planes. Several examples of this language will be described further with the results of the use of ALADDIN.

Location of the boundaries of a binding site

Figure 1 shows three molecules that have D2 agonist activity. They are superimposed to align the hydrogen bond and the basic nitrogen lone pair vectors. The fine dots show the union surface of the most potent agonists reported by Seeman et al. [13], and the heavy dots show the extra regions in space occupied by the three analogues of agonists that do not bind to the receptor.

Thus, such 'receptor mapping' often identifies regions in space that cannot be occupied by the small molecule, presumably because the target biomolecule occupies this space. The location of these forbidden regions using relative coordinates is not entirely satisfactory because one is trying to describe a binding site shape from the geometry of the ligand. However, we have found that sometimes these regions can be described as points in space with a certain radius (i.e., dummy atoms). The location of these points is specified with respect to those points in the molecule that are proposed to form the primary interactions with the target biomolecule. For the D2 example

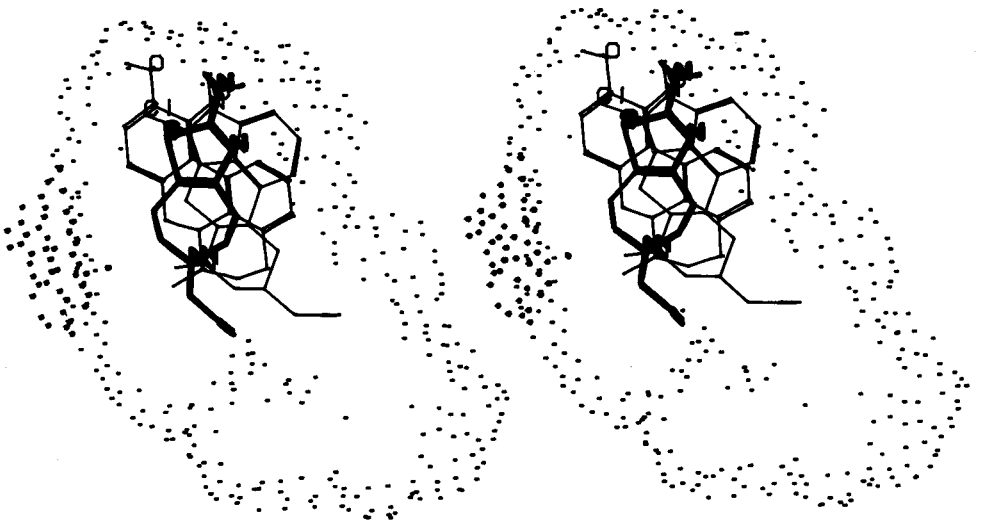

Fig. 1. Representative D2 dopamine agonists in our proposed 'receptor map'.

```
\       First we define the GENIE/SMARTS targets.
DEFINE $HbondON       [O,N,n;H]c
DEFINE $Namide        NC=[O,S]
DEFINE $Naniline      Nc
DEFINE $Nbasic        [N;!$Namide;!$Naniline]
\       The ALADDIN geometric objects are next.
POINT H-[$HbondON]
POINT [$Nbasic]
PLANE OF MOLECULE
POINT (3) (1,10.7) (2,8.6)
\       Then we specify the geometric constraints.
DISTANCE(1,2) 6.8 8.3
DISTANCE(4,*) 4.0  999.0
\       Finally we include the GENIE code.
THORBASE "NEURO"
IF 1- [$HbondON] {
   IF 1- [$Nbasic] {
      CALL "SCORE" "DOPAMINE"  "1 2" }  }
```

Fig. 2. An ALADDIN description of our D2 'receptor map'.

shown in Fig. 1, we have calculated the location of the dummy atom from the location of the basic nitrogen, the electronegative atom, and the hydrogen-bonding hydrogen atom.

If they are not on a straight line, distances to four anchor objects yield a unique solution for the location of a dummy atom. ALADDIN also allows specification of the dummy atom with three anchor objects. This produces two locations of the dummy atom relative to the rest of the molecule. One anchor object may also be a PLANE; for example, specifying a dummy atom by distances to two POINTs and constrained to lie in a specific PLANE yields two relative locations for the dummy atom.

One ALADDIN description of the D2 map is shown in Fig. 2. The atom definitions are, respectively, for an electronegative atom bonded to a hydrogen and an aromatic carbon, an amide nitrogen, an aniline nitrogen, and a nitrogen that is not an amide or an aniline nitrogen. The first geometric object is a point located at the position of the hydrogen-bonding hydrogen; the second is a point located at the position of a basic nitrogen atom; the third is the least-squares plane of the molecule; and the fourth is a point located in the least-squares plane of the molecule, 10.7 Å from POINT 1 and 8.6 Å from POINT 2. The first constraint is a distance of 6.8-8.3 Å between the H-bonding H and the basic N; the second is that no atom can be within 4.0 Å of the dummy atom defined by Geometric Object 4.

Figure 3 shows the selected points, the first required distance, and the location of the dummy atom in apomorphine that defines the forbidden region. This ALADDIN description correctly identified as active all of the compounds included in the composite surface for the receptor map and rejected those compounds that define the 'forbidden region.'

However, we have found that it is not always possible and/or convenient to define the forbidden region from relative coordinates. Sometimes, it is not clear where to place the dummy atoms: it would be tedious to describe the ligand binding site of a protein in this way. At other times, the positions of the forbidden region computed from the candidate XYZ datasets do not superimpose properly; Figure 4 illustrates this. Specifically, the position of the dummy atom varies dramatically even though the corresponding atoms in two structures are rather well superimposable. For the compounds shown in Fig. 4, the dummy atoms shown are 3.03 Å apart. Finally, when we wish

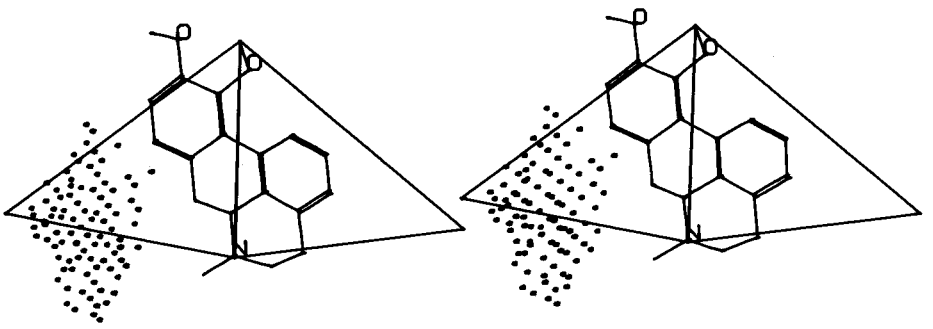

Fig. 3. The geometric objects and constraints of the ALADDIN description in Fig. 2 calculated from the XYZ coordinates of apomorphine.

to evaluate a 'receptor map' hypothesis using known literature data on enantiomerically resolved compounds, the lack of the ability to distinguish enantiomers presents a problem.

Thus, we have added the provision to test in absolute coordinate space those conformations of molecules that meet the first constraints. ALADDIN first brings the candidate XYZ dataset into the proper coordinate space. For this purpose, the user specifies a reference molecule and its atoms for root-mean-square superposition with GENIE/SMARTS specified atoms of the candidate molecule. This reorientation may be skipped if the stored orientation is to be tested. A second input is the file containing the coordinates of points on the surface of the binding site or in the forbidden region. This file is prepared with our molecular graphics program if the fit to a 'receptor map' (such as is shown in Fig. 1) is to be tested or with the location of the surface of the protein if the fit to a binding site on a protein is to be tested. The CLASH routine then tests if any atom of the candidate molecule protrudes into the surface of the binding site. Either only the stored or both enantiomers may be tested.

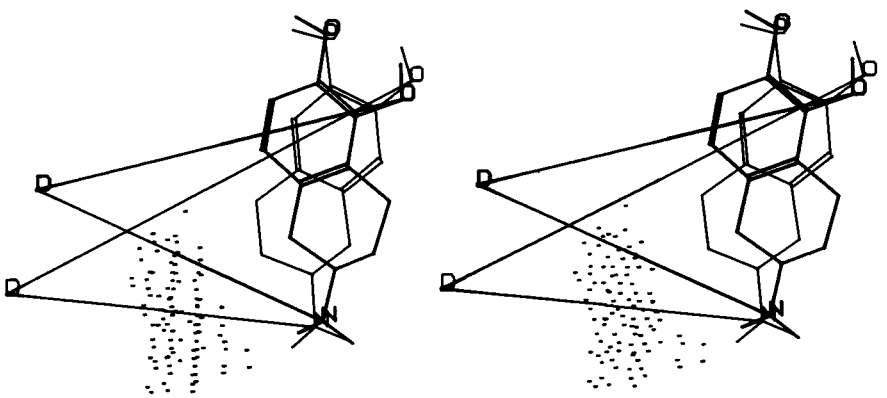

Fig. 4. The location of the dummy atoms from the ALADDIN description in Fig. 2 calculated from the XYZ coordinates of 5,6-dihydroxy-2-aminotetralin and 6,7-dihydroxy-2-aminotetralin.

IMPLEMENTATION

The FORTRAN version of the MedChem software (Version 3.54) running on our VAX 11/785 was used as the basis for implementing ALADDIN. The features of the databases of three-dimensional structures were previously described [11]. In particular, coordinates are stored in a MENTHOR in the order in which the atoms appear in the SMILES string with the hydrogens last and in the order of the heavy atom to which they are attached. Multiple XYZ datasets that correspond to a particular SMILES are often stored; these may represent different conformations or different stereoisomers. The three-dimensional structures were largely determined through CONCORD [14] or molecular mechanics [15], although some of the structures are from molecular orbital calculations or X-ray crystallography.

ALADDIN runs as a FORTRAN subroutine called from GENIE. Thus, all the searching and criteria-selection of GENIE is retained; ALADDIN three-dimensional constraints are additional requirements imposed on the molecules.

When using ALADDIN, one may specify that the search includes only the molecules contained in a specified file of registry numbers, names, name of the set of coordinates (XYZ dataset), and/or SMILES strings. This file may be the result of a substructure or bioactivity search. Alternatively, it might contain the registry numbers of those compounds for which there is enough remaining sample to perform the test or for which the calculated octanol-water partition coefficient lies within a specified range. When one is validating the ALADDIN description of a pharmacophore, this file would contain the SMILES of known active and inactive molecules. One could also use this provision to update a search by indicating the molecules or sets of coordinates added to the database since the previous ALADDIN search.

At the beginning of an ALADDIN run, the user specifies the database to be scanned, the ALADDIN search criteria description file to be used, if all possible hits for a compound are to be recorded or if one hit per compound is sufficient (as in the identification of existing compounds to test for a different activity) and, optionally, which subset of compounds should be examined. ALADDIN then opens the database, parses the search criteria description into an internal form, and retrieves the SMILES structure of each compound to be examined. GENIE tests are performed to determine if the candidate molecule possesses the necessary chemical groups and, optionally, to determine if the user-specified database criteria are met. For example, has it already been tested for D2 activity? Is at least 100 mg present for testing? Only when all of these criteria are satisfied are the coordinates passed to ALADDIN to determine if the geometric and steric constraints are satisfied. If multiple sets of coordinates are present for a particular SMILES, all sets are tested unless one matches and the 'one hit' option is used.

The main ALADDIN output file, after repeating the input search question, optionally logs each compound and conformation (XYZ dataset) tested, each target atom identified as part of a geometric object, and the results of the geometric tests between every tested combination of targets. When a particular XYZ dataset meets all constraints, the pertinent data is also summarized in the results file. If no XYZ dataset for a particular compound meets the constraints, that fact may optionally be logged to the main output file. Figure 5 is an example of part of such an output file.

ALADDIN also produces a file of the SMILES and XYZ dataset name of the compounds that meet the constraints. This file is used as input into further ALADDIN runs or into other Med-

```
NOW EXAMINING: Oc2cccc3C1CNCC1CCc23
 SET #   1 $CC<24C2MOD
  HITS FOR TARGET [$Nbasic]    AT ATOMS:  9
  HITS FOR TARGET H-[$HbondON] AT ATOMS: 15
  >>> MET ALL CRITERIA  1  1  1  1
  ============================HIT 24=====DOPAMINE====
  [3OH]c2cccc3C1C[2NH]CC1CCc23
  $CC<24C2MOD
  WD:[MARTIN.PROGRAMS.ALADDIN.D2.D1.NEWTHOR.RUN]24C2MOD.MSF,
  DISTANCE( 1, 2)    8.02
  DISTANCE( 4, *)    5.70
  DUMMY 4 COORDINATES 4.13   -7.45    3.22
  MATCHING ATOMS:   O1 N9 H15
  ============================HIT 24=====DOPAMINE====


NOW EXAMINING: Oc2cccc3CCC1CNCC1Cc23
 SET #   1 $CC<32C2MOD
  HITS FOR TARGET [$Nbasic]    AT ATOMS: 11
  HITS FOR TARGET H-[$HbondON] AT ATOMS: 16
  SET #  1 FAILED CRITERION   DISTANCE( 1, 2) :    6.61  1  1  1  1
 DID NOT MEET CRITERIA FOR ANY XYZ SET


NOW EXAMINING: C2NC3Cc1cc(O)ccc1C23
 DATABASE DID NOT CONTAIN C2NC3Cc1cc(O)ccc1C23
```

Fig. 5. Part of an ALADDIN output file.

Chem programs to prepare, for example, more detailed printouts of available chemical or biological data, calculate octanol-water partition coefficients, or prepare plots of the structures found.

A third output file is identical to the second except that the atoms in the SMILES that correspond to the ALADDIN/SMARTS targets are identified by the number of the SMARTS target. This file is used to prepare printouts of the structures that pass the search criteria with the target atoms identified. When one uses ALADDIN to find templates, this file is the input to MODSMI, our program that converts the input SMILES into the SMILES of the molecule proposed for synthesis.

The final output file is used as an entry into our molecular graphics system. It contains the SMILES of the molecule, the XYZ dataset name, and the atom numbers of the ALADDIN/SMARTS target atoms. This file can be used as arguments of a molecular graphics macro. For example, one might retrieve the XYZ dataset from a MENTHOR and prepare its enantiomer, superimpose both over a standard molecule matching specified atoms, make a color stereo plot of the display, store the new coordinates in a different MENTHOR, predict the potency from three-dimensional QSAR equations [6], and remove the molecules from the display. This coupling of an ALADDIN output file with the macro capability of our molecular graphics system gives us great flexibility as to how the ALADDIN results may be ultimately analyzed by the scientist. (Indeed, we have found this feature so useful that we sometimes run ALADDIN with very loose search criteria just to prepare this file to use for molecular graphics analysis. This is especially handy when one is in the early phases of a receptor mapping project with many candidate XYZ datasets and several possible superposition rules to examine.)

TABLE 1
SUMMARY OF SEARCH TIMING (VAX 11/785)

Example	Cmpds tested	Distances tested	Plane calcd	Dummy calcd	Torsion calcd	Blocks written	CPU (s)	Distance		s/hit
								Hits	Checks/s	
1a	3378	14082	3478	6956	0	14082	11220	1853	0.8(1.9)[a]	6.1
1b	21565	>100000				1584	5681	328	?	17.3
3a	30[b]	1416	0	0	0	1087	573	234	2.5	2.5
3b	30[b]	1416	0	0	0	501	363	0	3.9	–
3c	158[b]	454	148	296	0	1039	538	136	0.8(1.7)[a]	4.0
4	2338	>100000	0	0	0	1742	37118	18	?	2062
5a	2495	>100000	0	0	0			448	?	
5b	10	6097	0	0	107	2744	1118	180	5.5(5.6)[a]	6.2

[a] The number in parenthesis is the number of geometric calculations per second.

[b] The database contained 3378 compounds, many with more than one conformation, when this run was done.

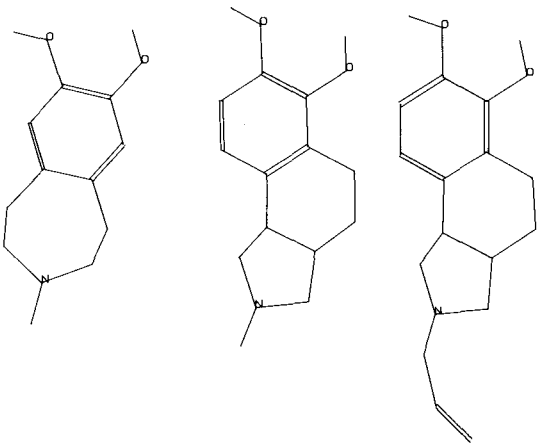

Fig. 6. The structures of three compounds that ALADDIN correctly predicted to have D2 dopamine agonist activity.

We plan to make ALADDIN available at a nominal cost. Coordinates of the molecules included in this report will be available in a MENTHOR database.

RESULTS

Example 1: Search for existing molecules that should have D2 dopaminergic activity

The ALADDIN query shown in Fig. 2 was used to search for compounds that might have D2 dopaminergic activity. In a database of 3378 compounds modeled for various neuroscience projects, 1448 passed the GENIE/SMARTS substructure specifications, and 1853 XYZ datasets met the criteria. The tested distances ranged from 2.2 to 11.52 Å. These 1853 datasets are represented by 912 topologically unique compounds.

The statistics of this run are listed in Table 1. Of the 187 min run time, 33 min were spent on the GENIE/SMARTS recognition of the presence of the appropriate substructures, the rest on the geometric tests and output generation. In comparison, by the use of screens 'a typical search (presumably of 12728 structures) takes about 10 min of elapsed time (on a) VAX 8600' [9d]. However, the searches in that report produced an average of 61 hits, 30 times lower than we found, so some of the decreased time might be due to fewer similar compounds for which detailed geometric tests are necessary. Nonetheless, we expect that the use of distance screens can speed up search time. This is at a cost of the time of creating the distance screens.

Three existing compounds discovered by this search and later shown to have D2 activity are shown in Fig. 6. Although in this case most of the identified compounds that were tested show the predicted activity, the predictions may be expected to be only as robust as is the 'receptor map' on which they were based. ALADDIN provides the tool to find compounds that meet criteria derived from other information.

Example 2: Comparison of two superposition rules for D2 agonists

A subjective but essential part of a 'receptor mapping' investigation is to evaluate different su-

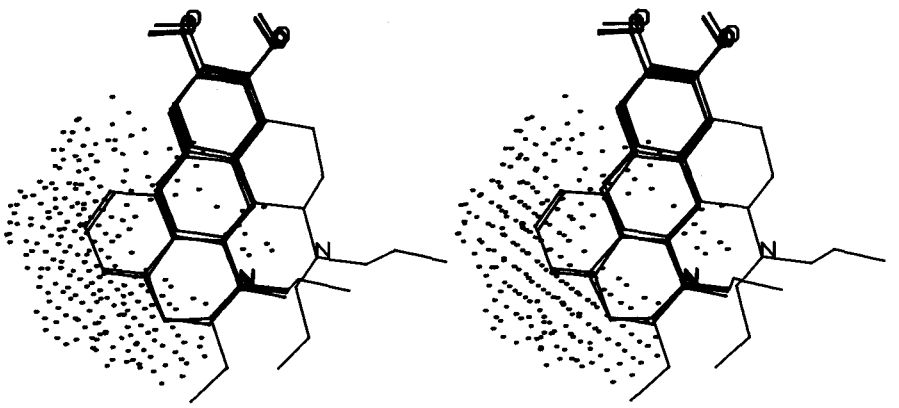

Fig. 7. An alternative superposition of catecholamine dopamine agonists. The forbidden region was calculated from the extra space occupied by the inactive compounds (heavy lines) compared to the active compounds.

perposition rules for a set of compounds. To do this with molecular graphics means that a human is making the evaluation — a task that perhaps a computer could do as well.

For example, although most workers superimpose the various dopamine agonists as we have shown in Fig. 1, there are certain problems with this superposition. The most important problem is the wide range of N-O distances found in active molecules. For example, the N-O distance in 7-hydroxy-2-aminotetralin is 7.4 Å, whereas that in 5-hydroxy-2-aminotetralin is 6.6 Å. Because of this, Grol et al. [16] have suggested that perhaps the molecules should be superimposed over the catechol only, as shown in Fig. 7. Although the nitrogen atoms do not now overlap, they could be imagined to interact with alternate oxygen atoms of a carboxylate in the receptor binding site. Thus, we wondered if this superposition rule is consistent with known structure-activity information.

Since Grol et al. [16] considered only 5,6- and 6,7-dihydroxy-2-aminotetralin, we calculated the

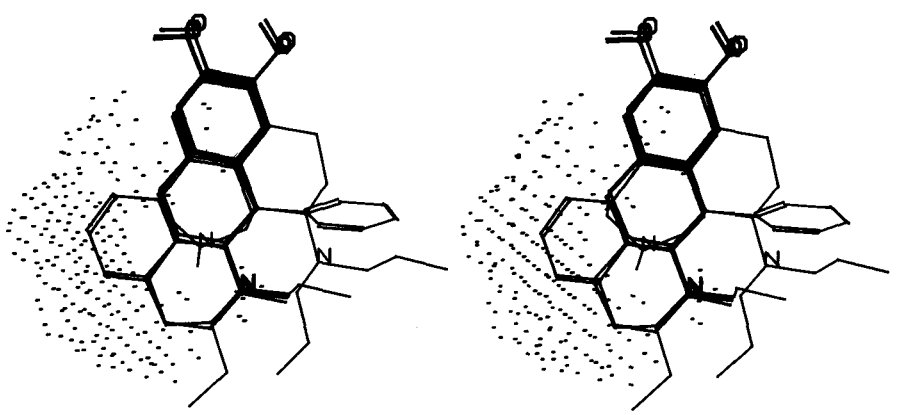

Fig. 8. An ALADDIN-suggested refinement of the Grol proposal. The molecules shown in heavy lines are inactive, the others are active.

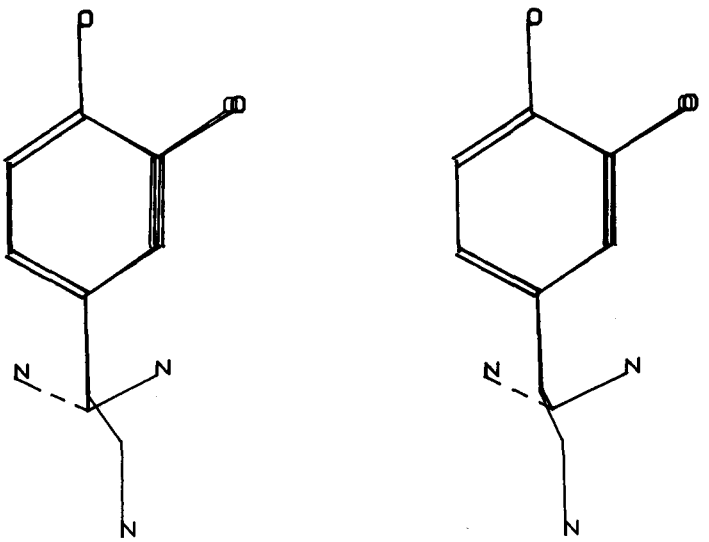

Fig. 9. Three low-energy conformations of dopamine.

location of the forbidden region from these compounds plus the inactive compounds shown in fine lines in Fig. 7. The proposed forbidden region is the unique region of space occupied by inactive compounds compared to the union space occupied by the active compounds. It is shown as dots in the figure.

This data was used in the ALADDIN CLASH test of 27 catecholamines tested for D2 binding by others [13]. The ALADDIN run took 52.3 s VAX 11/785 CPU time. It identified that SKF-38393, a moderately active D2 agonist, protrudes into the forbidden region. However, one may redefine the forbidden region to include this compound in the definition of active compounds, (see Fig. 8), and then there is a clear discrimination in ALADDIN between active and inactive compounds. It should be noted that the basic nitrogen atom of SKF-38393 is not well superimposed with that of the other types of compound, so although this alternative model was not rejected, it too has some deficiencies.

Thus, ALADDIN is a useful tool for evaluating a superposition rule as well as for suggesting a modification to a receptor map hypothesis.

Example 3: Design of compound sets that match each of the three low-energy conformations of dopamine

Since a major intended use of ALADDIN was to design molecules to test conformational requirements for bioactivity, we have made several tests of this capability and have included features to make it easy to perform.

Figure 9 shows the three low-energy conformations of dopamine. Extensive structure-activity analysis has shown that the meta-hydroxyl group is the only one that is essential for bioactivity [17], thus we decided to design phenols rather than catechols. Figure 10 shows the ALADDIN search for templates of the 180° conformation. We searched for an aliphatic atom in a ring or at-

```
        First we define the GENIE/SMARTS targets.
DEFINE $OKA             [A]
DEFINE $aTAR            [a;R1]
\       The ALADDIN geometric objects are next.
POINT [$OKA]
POINT [$aTAR]
\       Then we specify the constraints.
DISTANCE(1,2) 5.5 6.6
\       Finally, we include the GENIE code.
THORBASE "SAMPLE"
IF 1- [$OKA]  {
   IF 1- [$aTAR]  {
      CALL "SCORE" "DOPAMINE" "1 " } }
```

Fig. 10. The ALADDIN search for templates for the 180° conformation of dopamine.

tached to a ring (to become the nitrogen in the designed compound) the correct distance from an aromatic atom bearing a hydrogen (that will be transformed into the OH of the designed compound). A second search was used to identify those hydrogen atoms attached to an aliphatic atom that might be transformed into an amino group. Corresponding searches were performed for the two other conformations.

We examined hydrocarbons that include one aromatic ring fused to a 4-, 5-, 6-, or 7-membered aliphatic ring that is in turn fused to a 4-, 5-, 6-, or 7-membered aliphatic ring. Geometric and spiro isomers were included. If we did not have coordinates for the molecule of interest, they were generated with CONCORD [14] and minimized with MMP2 [15] if CONCORD produced a strained structure.

The specifications in Fig. 10 were used to examine 30 fused ring hydrocarbons. All 30 passed the GENIE/SMARTS specifications, 1416 distance checks were made, 234 hits were recorded,

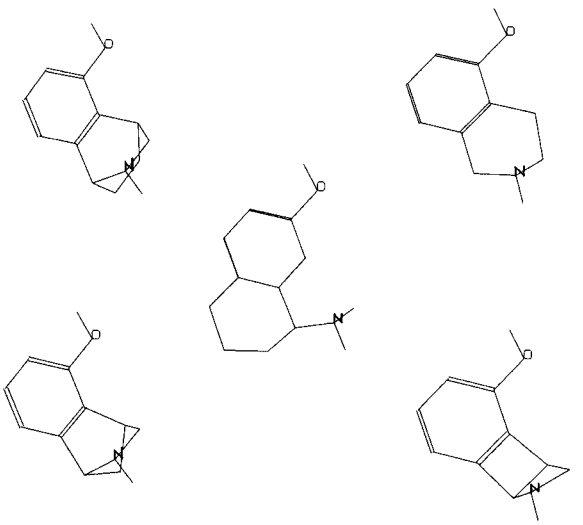

Fig. 11. The structures of compounds suggested by ALADDIN to mimic the 60° conformation of dopamine.

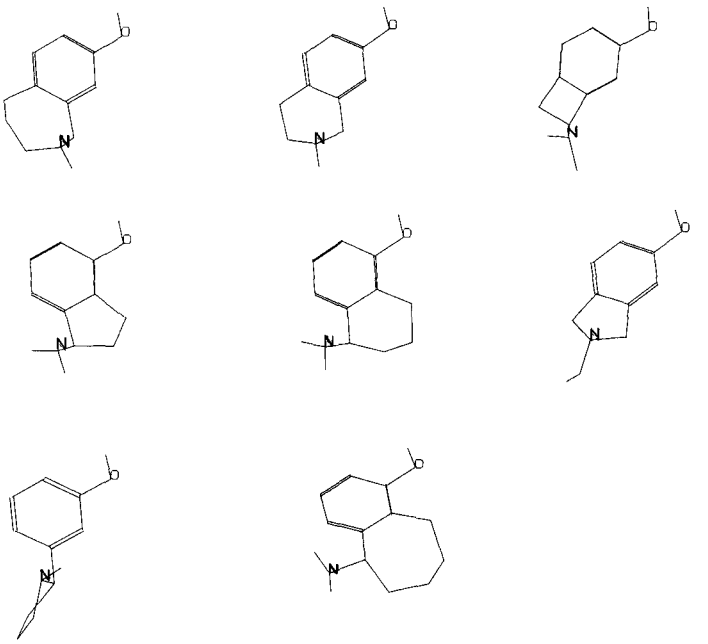

Fig. 12. The structures of compounds suggested by ALADDIN to mimic the 300° conformation of dopamine.

and only one hydrocarbon did not have any aromatic-aliphatic distance that met the criterion. The GENIE part of the search took 12.13 s CPU on the VAX 11/785, whereas the ALADDIN run took 9.55 min. Thus, distance checks and output generation accounted for the majority of the computer time (Example 3a in Table 1). We verified that additional output takes additional computer time by repeating the run with constraints such that no hits were found; this shortened the run by 210 s or 37% (Example 3b in Table 1). Note that distance screens would have made little difference in the timing since 29/30 of the molecules had a distance within the required range.

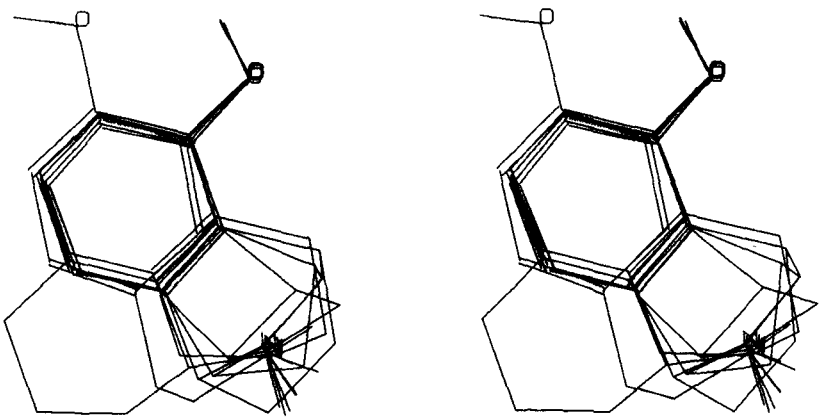

Fig. 13. A stereo plot of the compounds shown in Fig. 11 over the 60° conformation of dopamine.

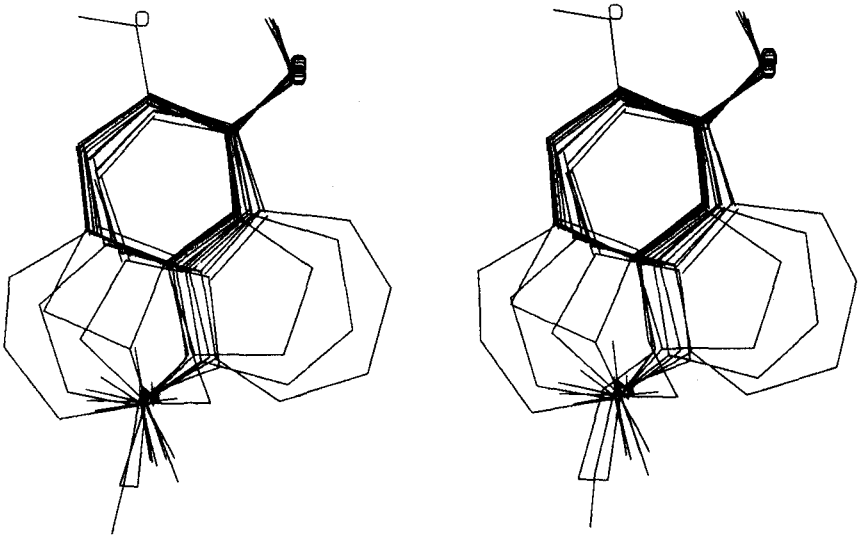

Fig. 14. A stereo plot of the compounds shown in Fig. 12 over the 300° conformation of dopamine.

The labeled SMILES files of hits on the hydrocarbons were then processed to produce the SMILES of the proposed compounds. These SMILES were processed with CONCORD (and MMP2 if necessary) to generate XYZ coordinates of the proposed compounds.

The three final ALADDIN searches tested if the proposed compounds indeed match the original geometric properties of the conformation. This run for the 180° conformation examined 158 compounds and found a total of 134 hits. (Fewer compounds, 158, were examined compared to those suggested by the first ALADDIN run, 241, because this version of SMILES does not retain the stereochemistry of the first 'hits.' Thus, if the original 'hit' was on both cis- and trans-fused

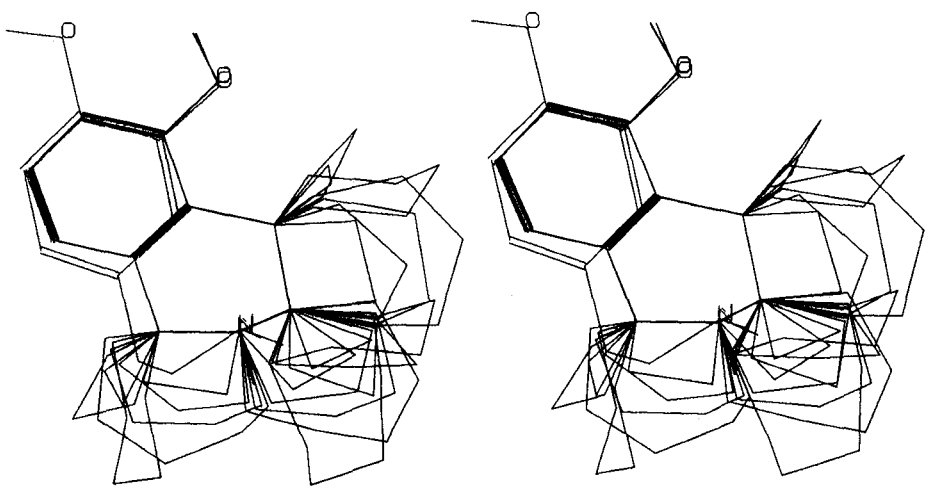

Fig. 15. A stereo plot of compounds with a common backbone but different added rings.

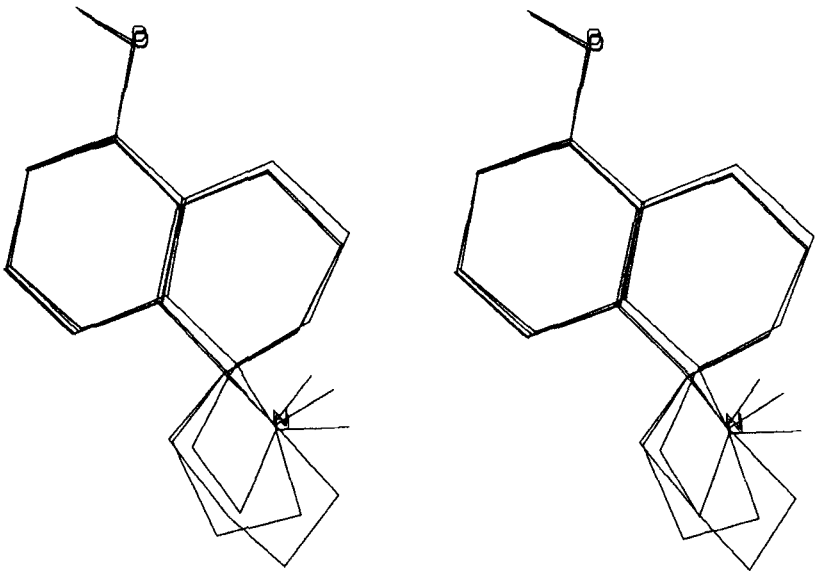

Fig. 16. A stereo plot of compounds that differ only in the direction of the lone pair on nitrogen.

rings, we lost this information and let CONCORD produce the one it preferred. Fewer distance comparisons, 148, were made compared to the number of input compounds, 158, because ten of the compounds did not pass the GENIE/SMARTS specification.) This took 8.96 min CPU time on a VAX 11/785, or approximately 2.5 s/hit (Example 3c in Table 1).

By procedures described above, stereo color plots of the selected compounds were prepared for examination. By comparison, producing these plots took 46 min CPU time.

Figure 11 shows the compounds with unique backbones that were found to match the 60° con-

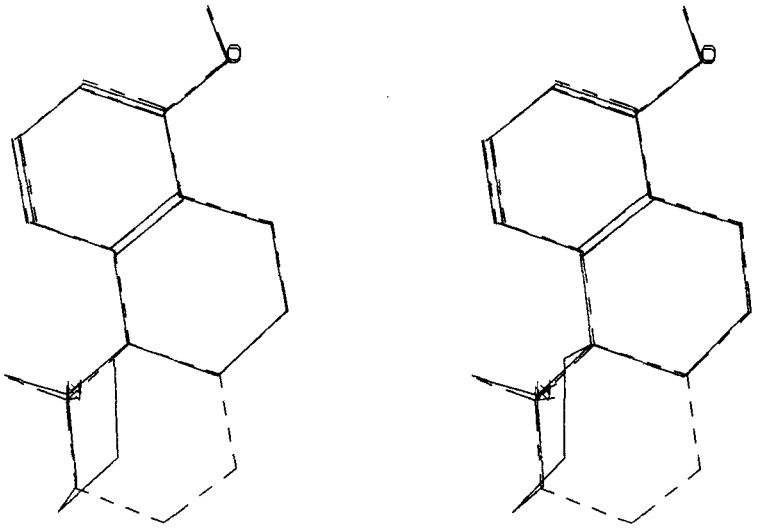

Fig. 17. A stereo plot of two compounds that differ in shape but are the same with respect to the direction of the lone pair on nitrogen.

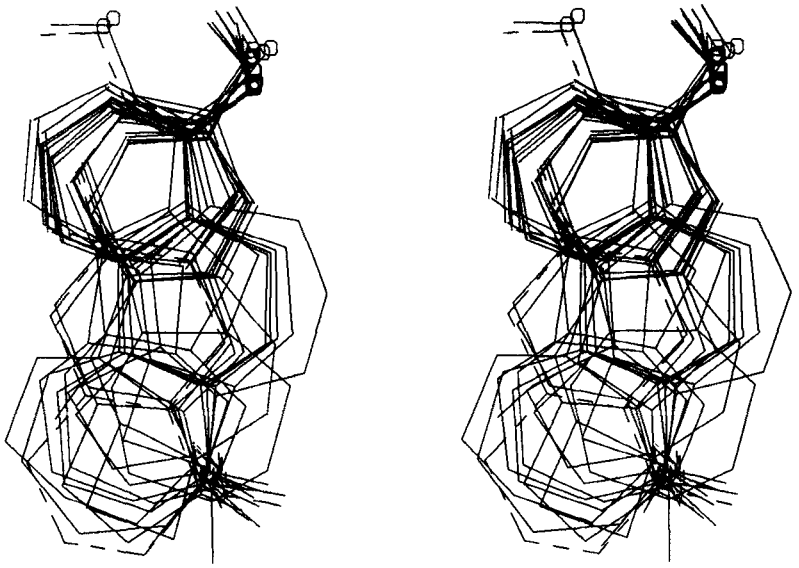

Fig. 18. A stereo plot of the ALADDIN-designed compounds that probe the 'forbidden region' on the D2 dopamine receptor.

formation of dopamine, and Fig. 12 shows those that match the 300° conformation. In Figs. 13 and 14, these are shown superimposed on the corresponding conformation of dopamine. Notice that both sets of compounds vary widely in overall shape. Figure 15 shows this to an extreme; included are compounds that have the constant hydroxy-isoquinoline but varying rings fused to this parent. Figure 16 shows another interesting subset of compounds; they differ only in the size of the spiro-fused ring. Note that the direction of the lone pairs on the nitrogen varies substantially in this set. Thus, ALADDIN suggested a way to probe this shape property of agonists. Figure 17 shows another interesting pair of compounds. In this case, both hold the lone pair in the same direction but use a totally different region of space to form the conformational constraint.

Previous workers have shown that the 180° conformation of dopamine is the bioactive one [17]. Thus, although we were pleased to see that ALADDIN suggested all of the known compounds, this is not surprising since we knew the structures of them when we built the database. However, we were also pleased to see that ALADDIN suggested a great variety of previously unknown compounds. Of the 134 compounds that match the bioactive conformation, Fig. 18 shows a set that probes in a very subtle way the space near the proposed 'forbidden region.' If such a set of molecules were prepared and tested the data would be well-suited for three-dimensional QSAR analysis described by Cramer et al. [6].

Figure 19 shows another interesting molecule, which is a close spatial analogue of apomorphine. This demonstrates that ALADDIN could be used to find novel analogues of known compounds.

Example 4: Design of mimics of the (i + 1) and (1 + 2) side chains in peptide beta turns

The design of peptidomimetics for use as drugs is greatly aided by knowledge of the bioactive conformation of the peptide. Thus, we were interested in demonstrating that ALADDIN can find

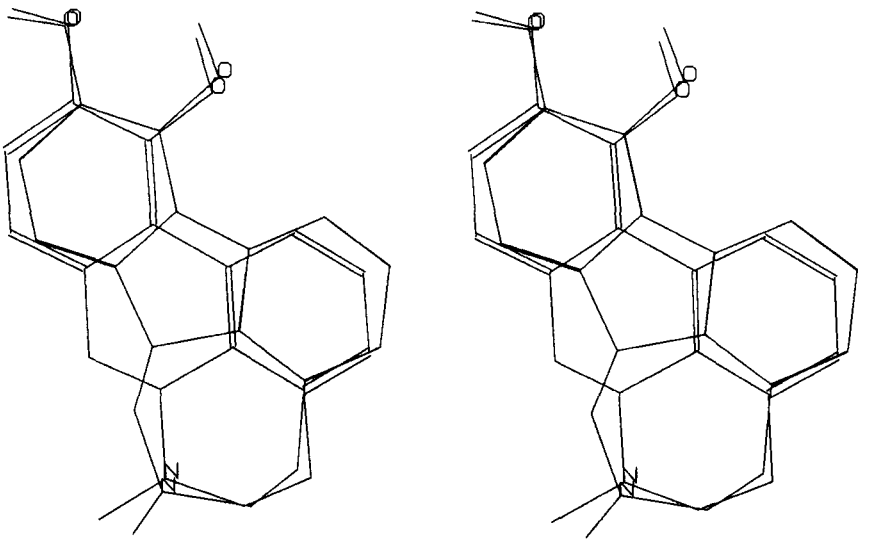

Fig. 19. A stereo plot of a close analogue of apomorphine suggested by the ALADDIN results.

small molecules that match a portion of a peptide chain and/or backbone (i.e., to suggest conformationally constrained mimics of a portion of the peptide). Such a search uses a larger number of geometric constraints than do the ones above.

As an example, we chose to mimic two residues in the beta turn of a theoretical conformation of the peptide LHRH [18]. For this, we searched a database of representative small molecules that had been modeled or retrieved from the Cambridge Crystallographic Database [19] for one or another of our drug design projects. To this database has been added the various conformationally defined molecules (usually cyclic hydrocarbons) suggested by various template-searching exercises such as those described for dopamine above. We searched for compounds that match within 0.1 Å all four nonbonded distances between the following atoms in this conformation: C(beta)6, C(alpha)6, C(beta)7, and C(alpha)7. Figure 20 shows two of the 18 molecules that matched. Indeed, that shown in Fig. 20A matches more than just the specified atoms. It is a better match than the previously suggested mimics of the beta turn [20]. This run took approximately 10 h (Table 1); more than 10000 distance tests were made (we suspend output of individual tests after 9999 because we assume that no one would really use such numbers). An additional 300 matches were found in the search for hydrogen atoms attached to a ring that could be replaced by the respective side chains. In the same database, there were no matches to the C(alpha) and C(beta) of residues 5, 6, and 7.

In a peptide-mimic synthetic chemistry program, one might wish to change the type of atoms in the template to match better the electrostatics of the native peptide or to make the molecule easier to synthesize.

Example 5: Design of compounds to fit into the methotrexate binding site on Lactobacillus casei *dihydrofolate reductase*

To assess the utility of ALADDIN for the design of molecules that fit into a binding site established by protein crystallography, we chose to look for analogues of methotrexate that hold the

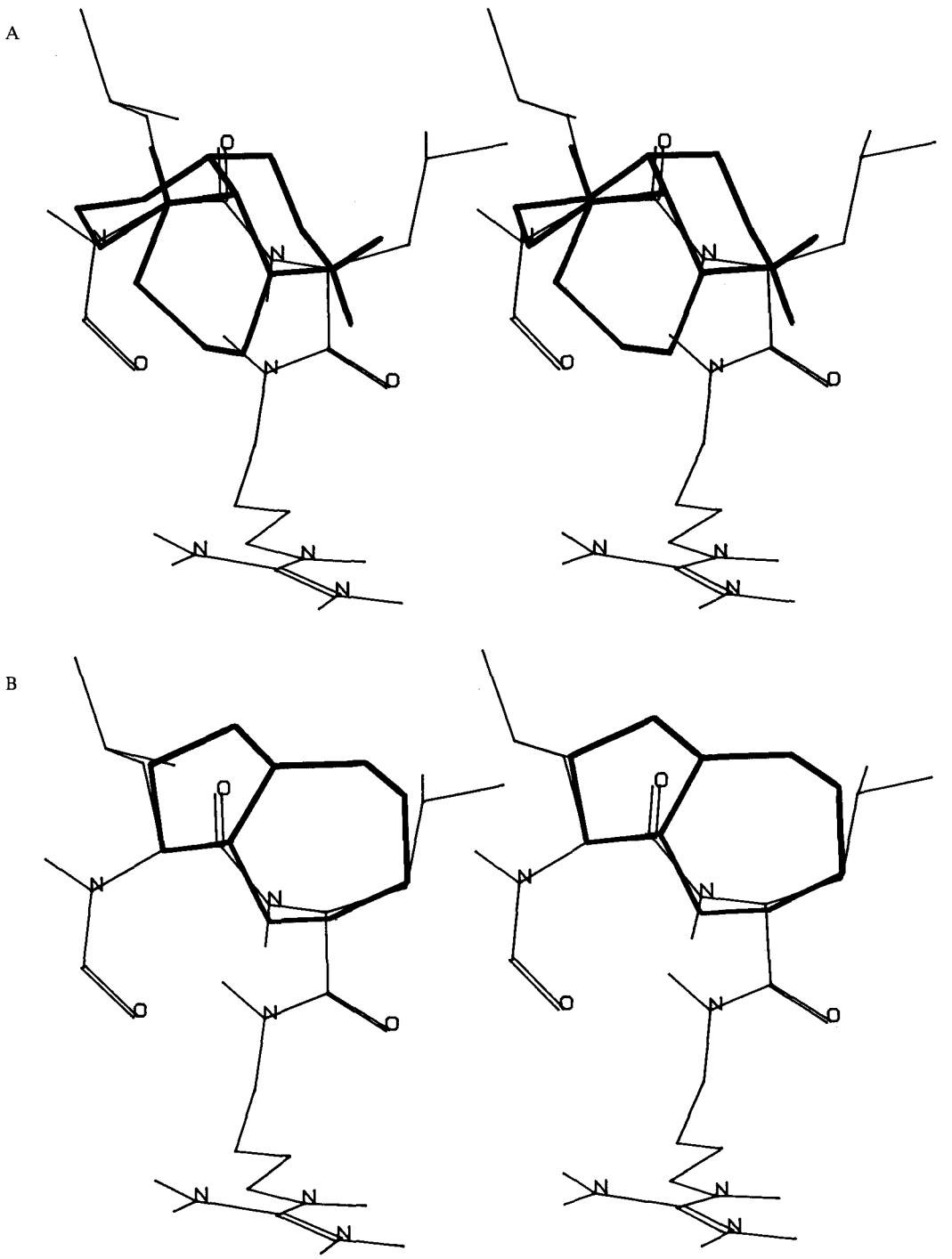

Fig. 20 A and B. Stereo plots of two molecules that mimic the alpha carbon and some of the backbone atoms of two amino acids in a beta turn.

two aromatic rings in the bound conformation [21]. Again, we did a simple distance comparison on a sample of molecules in our small database and identified a total of 448 hits in the first run. However, examination of the two-dimensional structures with the hit atoms labeled suggested that only 25 of these were unique and interesting. Thus, we used molecular graphics rather than the CLASH option of ALADDIN to evaluate them. Figure 21 shows two molecules that fit into the cavity pretty well. Interestingly, several 5-membered ring compounds had the appropriate torsion angles as did some 7-membered ring compounds, but the 6-membered ring compounds that matched best did not fit as well.

The timing for this run, 5a, is listed in Table 1. To get a better estimate of performance we tested ten saturated bicyclic compounds; that run is listed as 5b in the table. Presumably because of the small amount of output, this was the most efficient run logged.

A run on the whole database was terminated after 10602 hits were recorded in 28.78 h VAX 11/785 CPU time. This run produced 40398 blocks of output. Clearly, processing this number of hits requires some additional methods.

DISCUSSION

Only when one can correctly classify the bioactivity of all molecules investigated in a particular biological test does one have a description that is useful for the prediction of the biological activity of untested molecules. The ALADDIN language provides an unambiguous and verifiable description of the geometric, steric, and substructural description of such requirements. Our experience so far is that, when both relative and absolute coordinate systems are available, we have been able to write a search for any type of shape description that we have had.

The next extension of the language will be to add the ability to calculate a quantitative forecast of potency from a three-dimensional-QSAR equation [6]. This addition will allow a more precise definition of both the steric and electrostatic requirements for activity and add the possibility to include explicit quantitation of the hydrogen-bonding properties of the ligands.

Since it is in a computer-recognizable form, the ALADDIN description of the structural requirements for a particular biological activity is especially powerful. One could imagine a catalogue of such descriptions as a very efficient way to design selective compounds or to forecast an unexpected side effect of a proposed compound. Thus, another use of ALADDIN would be to test compounds proposed on any basis against all available valid ALADDIN descriptions.

The examples show that ALADDIN design of new molecules, even though carried out on our rather small database, has provided many suggested structures. (However, we need more experience with designing ligands to fit binding sites on proteins.) Of course, the scientist will change these structures to accommodate synthetic ease and/or to modify physical properties such as octanol-water partition coefficients. Usually, when a group of scientists views the results of an ALADDIN search for templates on which to add the groups that confer bioactivity, many other structures are suggested. MedChem software can be used to quickly see if these compounds were considered by ALADDIN. If not, we then generate three-dimensional coordinates and perform the ALADDIN search. Thus, ALADDIN serves to suggest compounds, and the scientist uses this information to suggest other compounds. Since we are continually adding compounds to the database, the ALADDIN searches for templates become progressively more fruitful.

The ability of ALADDIN plus CONCORD to change a found template into the molecule to be

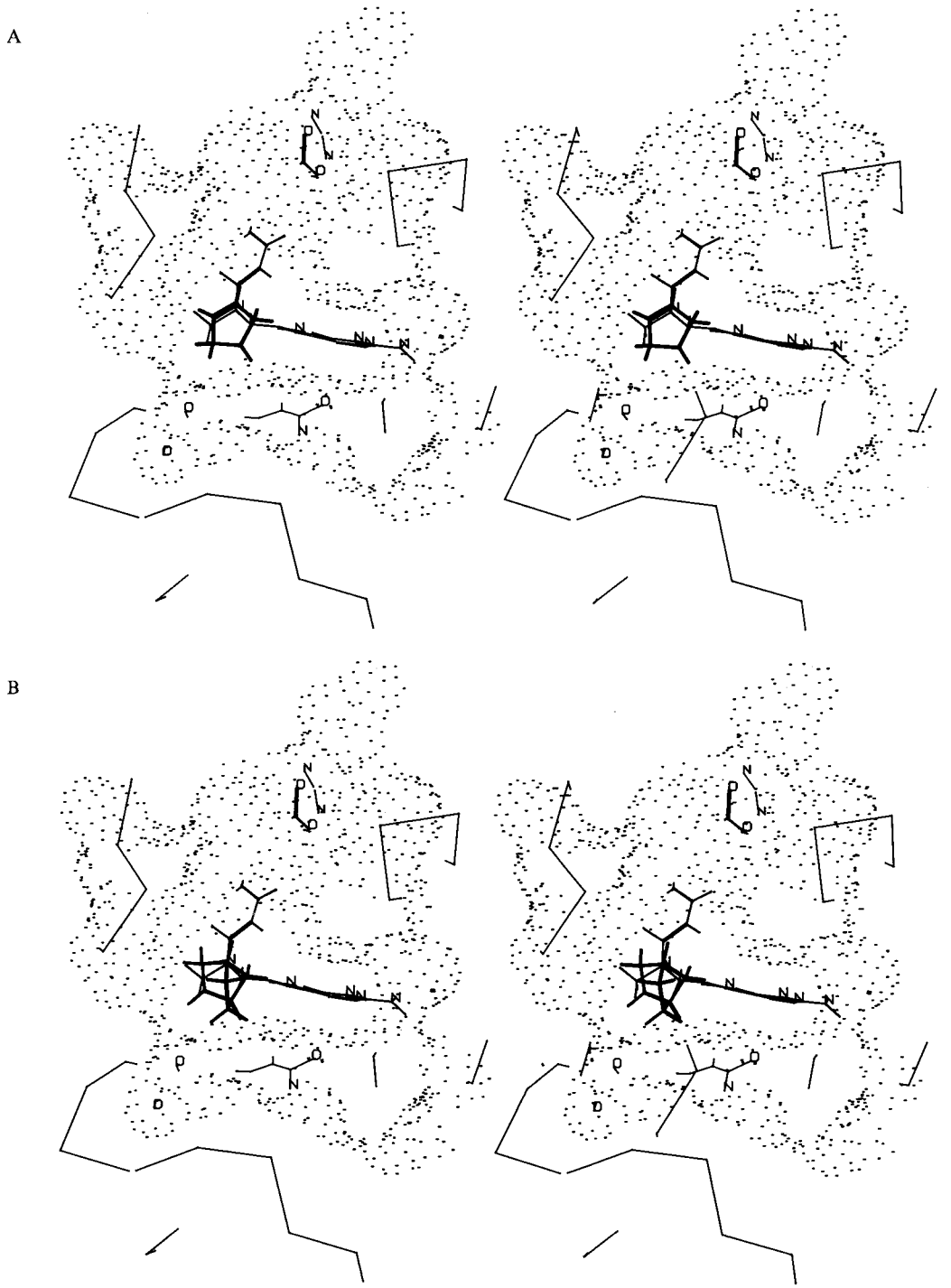

Fig. 21 A and B. A view of methotrexate (fine lines) as bound in the active site of dihydrofolate reductase — established by single-crystal X-ray diffraction. Two different templates that hold methotrexate into its bioactive conformation are shown in heavy black lines.

synthesized and then to evaluate the fit of the designed molecule to the original geometric and steric hypothesis is one key feature of its usefulness. The ties between searching, generation of three-dimensional coordinates, and molecular comparisons make ALADDIN a real computer tool for compound design.

We would like to overcome ALADDIN's use of only the exact data stored in the database. Ideally, one would like to be able to change the conformation retrieved from the database into other potentially favorable conformations, to piece together fragments that separately meet some of the search criteria, and/or to generate new structures de novo. Although the current version of ALADDIN uses coordinates from a database, they could just as well be generated with a program such as CONCORD as needed. Such a capability suggests adding rules for proposing new structures that might transform a near miss into a successful match. (A little additional bookkeeping will be needed, since such suggested structures should be considered only if not previously evaluated.) Similarity searching [22] might also identify compounds that do not meet the original search requirements but are close enough to consider with some structural modifications. Some of these enhancements may require the development of artificial intelligence extensions of ALADDIN.

When, as usual, an ALADDIN design results in more than a few compounds, one must organize the suggestions. Since the usual goal is to design compounds that explore the geometric and steric properties required for bioactivity, the ultimate evaluation of proposed compounds is done in an absolute three-dimensional coordinate system. Currently, this is accomplished by preparing stereo color plots of each suggested molecule superimposed on a reference molecule. These plots are then examined and sorted by a human to recognize related series such as shown in the figures. In principle, additional ALADDIN searches could be used for this pattern recognition; however, for these examples we needed to sort the plots by hand before we could write the GENIE/ALADDIN search. Experience may make such classification easier. However, the ultimate evaluation will still be a human one.

Since the aim of our research was to explore the opportunities opened by such a tool, we have not concentrated on enhancing search speed as have others in a very elegant series of investigations [9]. Nevertheless, the computer time for our studies is of interest. Table 1 shows, as expected, that the time of an ALADDIN run depends on several factors: the number of distances calculated and compared to the search criteria, the number of least-squares planes and dummy atoms calculated, the complexity of the GENIE substructure definition, and the number of hits and thus the number of blocks written to the various files. For simple searches, there is another hit every few seconds — however, for searches with many constraints to be met, the rate of finding hits is much slower.

We have done a few things to shorten the time between submitting an ALADDIN run and viewing some results. First, there are options to examine only certain molecules in the database and to move to the next molecule once a hit is recorded. In addition, the user can view the results while the program is running.

The number of distance checks made during a run will be markedly reduced when ALADDIN is tied to the version of GENIE just developed, since in this new version there will be a FOR construct. This will allow us to define SMARTS targets with respect to only the currently identified SMARTS hit. Thus, rather than doing a distance check to ascertain that two atoms are bonded, the GENIE code will assure this. The amount of time saved will depend on the search question;

the pharmacophore searches in Examples 1, 2, and 3c will not be affected but the template searches would require approximately the square root of the number of distance checks.

Distance screens [9] would undoubtedly make certain pharmacophore searches interactive. As the example on design of dopamine agonists shows, for template searches, screens that indicate whether a molecule contains a particular distance between two atoms will not eliminate enough molecules; perhaps screens based on the presence of a certain triad of distances between three atoms would be adequate. One must still balance the computer time needed to calculate these screens with the benefits of a quick search. For example, if it takes 4 h to prepare the color stereo plots to show the results of an ALADDIN run, it doesn't matter to the scientist if the ALADDIN search takes a few minutes or an hour or two. Furthermore, if it takes a week to understand the ALADDIN results, it doesn't even matter that the color plots are produced so slowly. Rather, it suggests that attention should be paid to developing tools that help the scientist organize and interpret the results.

SUMMARY

In summary, a substantial effort has been devoted over the past few years at Abbott to model and store in databases the three-dimensional structures of compounds with interesting biological properties. We can now effectively exploit this mass of data through ALADDIN, which in effect gives us a technique for computerized screening of existing compounds or designing new compounds for new biological properties. It complements the rest of our tools for computer-assisted molecular design, consisting of molecular mechanics and quantum mechanics programs, molecular graphics, and the MENTHOR databases of experimental and modeling results.

Thus, ALADDIN provides us the ability to:

- objectively describe a receptor map hypothesis;
- scan a database to retrieve previously untested compounds that a receptor map hypothesis predicts to be active;
- quantitatively compare receptor map hypotheses for the same biological activity;
- design compounds that probe the bioactive conformation of a flexible ligand;
- design novel compounds that a receptor map hypothesis predicts to be active; and
- design compounds based on structures from protein crystallography.

REFERENCES

1 Goodford, P.J., J. Med. Chem., 27 (1984) 557–564.

2 a. Marshall, G.R., Barry, C.D., Bosshard, H.E., Dammkoehler, R.A. and Dunn, D.A., In Olson, E.C. and Christoffersen, R.E. (Eds.) Computer-Assisted Drug Design, American Chemical Society, Washington DC, 1979, pp. 205–226.

b. Gund, P., Andose, J.D., Rhodes, J.B. and Smith, G.M., Science, 208 (1980) 1425–1432.

c. Martin, Y.C. and Danaher, E.B., In Williams, M., Glennon, R.A. and Timmermans, P.B.M.W.M. (Eds.) Receptor Pharmacology and Function, Dekker, New York, NY, 1987, pp. 137–171.

3 a. Langridge, R., Ferrin, T.E., Kuntz, I.D. and Connolly, M.L., Science, 211 (1981) 661–666.

b. Feldman, R., In Heller, S.R. and Potenzone Jr., R. (Eds.) Computer Applications in Chemistry, Vol. 9, Elsevier, Amsterdam, 1983, pp. 9–18.

4 Burkert, U. and Allinger, N.L., Molecular Mechanics, American Chemical Society, Washington, DC, 1982.
5 Humblet, C. and Marshall, G.R., Annu. Rep. Med. Chem., 15 (1980) 267–275.
6 Cramer III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959–5967.
7 Goodford, P.J., J. Med. Chem., 28 (1985) 849–857.
8 a. Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T., J. Mol. Biol., 161 (1982) 269–288.
b. DesJarlais, R.L., Sheridan, R.P., Dixon, J.S., Kuntz, I.D. and Venkataraghavan, R., J. Med. Chem., 29 (1986) 2149–2153.
c. DesJarlais, R.L., Sheridan, R.P., Seibel, G.L., Dixon, J.S., Kuntz, I.D. and Venkataraghavan, R., J. Med. Chem., 31 (1988) 722–729.
9 a. Jakes, S.E. and Willett, P., J. Mol. Graph., 4 (1986) 12–20.
b. Brint, A.T. and Willett, P., J. Mol. Graph., 5 (1987) 200–207.
c. Brint, A.T. and Willett, P., J. Chem. Inf. Comput. Sci., 27 (1987) 152–158.
d. Jakes, S.E., Watts, N., Willett, P., Bawden, D. and Fisher, J.D., J. Mol. Graph., 5 (1987) 41–48.
e. Brint, A.T. and Willett, P., J. Mol. Graph., 5 (1987) 49–56.
f. Cambridge Structural Database System, Programs GSTAT, Cambridge Crystallographic Data Centre, Cambridge, 1989.
10 Daylight Software Manual, Release 3.54, Daylight Chemical Information Systems Inc., Claremont, CA, 1988.
11 Martin, Y.C., Danaher, E.B. May, C.S. and Weininger, D., J. Comput.-Aided Mol. Design, 2 (1988) 15–29.
12 Weininger, D. and Weininger, J.L., In Hansch, C., Sammes, P.G. and Taylor, J.B. (Eds.) Comprehensive Medicinal Chemistry, Pergamon, Oxford, in press.
13 Seeman, P., Watanabe, M., Grigoriadis, D., Tedesco, J.L., George, S.R., Svensson, U., Nilsson, J.L.G. and Neumeyer, J.L., Mol. Pharmacol., 28 (1985) 391–399.
14 CONCORD was written by A. Ruskinko III, J.M. Skell, R. Balducci and R.S. Pearlman. It is available from Evans & Sutherland Computer Corporation, Salt Lake City, UT.
15 Allinger, N.L., MM2 Reference Guide, Molecular Design, San Leandro, CA, 1987.
16 Grol, C.J., Jansen, L.H. and Rollema, H., J. Med. Chem., 28 (1985) 679–683.
17 Cannon, J.G., Prog. Drug Res., 29 (1985) 303–414.
18 Momany, F.A., J. Am. Chem. Soc., 98 (1976) 2990–2996.
19 Cambridge Structural Database System User's Manual, Cambridge Crystallographic Data Centre, Cambridge, 1989.
20 a. Krstenansky, J.L., Baranowski, R.L. and Currie, B.L., Biochem. Biophys. Res. Commun., 109 (1982) 1368–1374.
b. Kemp, D.S. and Sun, E.T., Tetrahedron Lett., 23 (1982) 3759–3760.
c. Kemp, D.S. and McNamara, P., Tetrahedron Lett., 22 (1981) 4751–4754.
21 Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C. and Kraut, J., J. Biol. Chem., 257 (1982) 13650–13662.
22 Willett, P., Similarity and Clustering in Chemical Information Systems, Wiley, New York, NY, 1988.