(12) Arteca, G. A.; Jammal, V. B.; Mezey, P. G. Shape Group Studies of Molecular Similarity and Regioselectivity in Chemical Reactions. *J. Comput. Chem.* **1988**, *9*, 608-619.
(13) Walker, P. D.; Arteca, G. A.; Mezey, P. G. A Complete Shape Group Characterization for Molecular Charge Densities Represented by Gaussian-Type Functions. *J. Comput. Chem.* **1990**, *12*, 220-230.
(14) Cioslowski, J.; Fleischmann, E. D. Assessing Molecular Similarity from Results of ab Initio Electronic Structure Calculations. *J. Am. Chem. Soc.* **1991**, *113*, 64-67.
(15) Graham, M. S. Merck Sharpe & Dohme, Sea Program, *Q.C.P.E. 567*.
(16) Meyer, A. M.; Richards, W. G. Similarity of Molecular Shape. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 426-439.
(17) Szabo, A.; Ostland, N. S. *Modern Quantum Chemistry*; Macmillan: Basingstoke, 1982; pp 410-412.
(18) Gill, P. E.; Murray, W. Algorithms for the Solution of Non-linear Least Squares Problems. *J. Numer. Anal.* **1978**, *15*, 977-992.
(19) Nelder, J. A.; Mead, R. Simplex Method for Function Minimization *Comput. J.* **1965**, *7*, 308-313.
(20) Takashi, S.; Katsutoshi, M.; Hiroyuki, T.; Yasuo, S.; Takeshi, F.; Yutaka, K. Study of Antidiabetic Agents. Synthesis of AL321 and Related Compounds. *Chem. Pharm. Bull.* **1982**, *30*, 3563-3573.
(21) Chem-X, Chemical Design Ltd., Unit 12, 7 West Way, Oxford OX2 0JB, United Kingdom.
(22) Stewart, J. J. P. MOPAC 5, *Q.P.C.E. 455*.
(23) Ferenzcy, G.; Reynolds, C. A.; Richards, W. G. Semi-Empirical AM1 Electrostatic Potential and AM1 Electrostatic Potential Derived Charges, a Comparison with ab Initio Values. *J. Comput. Chem.* **1990**, *11*, 159-159.

Compare_Conformer: A Program for the Rapid Comparison of Molecular Conformers Based on Interatomic Distances and Torsion Angles

ISTVÁN KOLOSSVÁRY† and WAYNE C. GUIDA*

Research Department, Pharmaceuticals Division, CIBA-GEIGY Corporation, Summit, New Jersey 07901

Received September 11, 1991

A computer program for comparison of the conformations of a number of related molecular structures is described. The comparisons are performed on either interatomic distances or torsion angles. The comparisons are accomplished on ordered pairs of distances or torsion angles, and the distance comparisons can be performed in a manner that allows permutation of the distance pairs being compared. The algorithm utilizes bit-string Boolean operations that allow the comparisons to be performed rapidly. The program should be useful for computer-assisted molecular modeling studies in which the viable conformers of bioactive analogues are compared in order to locate those conformers that place key substituents in the same spatial orientation.

INTRODUCTION

A central theme of computer-assisted molecular modeling has been the structural comparison of compounds that bind to the same molecular receptor. Such comparisons, particularly those among related structural analogues, have been employed to derive templates for drug design and to assist in the elucidation of receptor-ligand associations. A number of procedures have emerged that are of utility in molecular modeling studies. For example, a method termed the "Active Analog Approach" (AAA) pioneered by Marshall and coworkers[1] relies on the concept of a pharmacophoric pattern. This so-called pharmacophore is defined as the three-dimensional arrangement of atoms and/or functional groups essential for favorable receptor-ligand interactions. For compounds that lack structural rigidity, the potential pharmacophoric pattern for a particular molecule is not fixed but can vary substantially depending upon the conformational properties of the molecule under consideration. Thus, identification of the pharmacophore for flexible molecules is often an arduous task. The goal of a molecular modeling study involving AAA is to locate the pharmacophoric pattern common to conformers in a series of active molecules, and this goal can be achieved by coupling a systematic grid search of conformational space (without energy minimization) to orientation map calculations (identification of common intergroup distances).[2,3] An advantage of this method is that distance constraints derived from relatively rigid analogues can be used to limit the conformational search for more flexible analogues. The utility of AAA has been convincingly demonstrated with the suggestion of a unique bioactive conformation for a series of angiotensin-converting enzyme inhibitors.[4]

Numerous other methods have been employed for the identification of a common pharmacophore within a series of bioactive analogues[5-11] and for the location of molecules within three-dimensional databases that possess a particular pharmacophoric pattern.[12-16] All of these methods rely upon definition of the pharmacophore as a set of interatomic or intergroup distances and can be categorized (perhaps superficially) as distance geometry and/or pattern recognition approaches.

Of course, the identification of conformers sharing the same pharmacophoric pattern may be insufficient for complete rationalization of three-dimensional structure versus activity. Subsequent analysis may reveal whether steric and electrostatic properties of these conformers are optimal for binding. Volume mapping techniques can be used to provide additional steric information.[17,18] Coulombic and hydrophobic matching, which can be formulated in terms of electrostatic potentials and fields (potential gradient), may provide additional useful information concerning whether a particular conformer is likely to be active.[19-22] Furthermore, the approaches mentioned above rely on simplifying assumptions such as a single-binding mode. Nevertheless, these techniques have proven to be useful in the rationalization of three-dimensional structure-activity relationships.

Approaches such as AAA require collection of multiple conformational states of the analogues under consideration and comparison of interatomic or intergroup distances among the allowed conformations. In this paper, we describe an algorithm for the rapid comparison of a series of structural analogues, each of which may be associated with multiple conformers that may (or may not) have been previously subjected to energy minimization. We have developed a computer program that we call Compare_Conformer and abbreviate CP (for Com-

* Author to whom correspondence should be addressed.
† On leave from the Department of General and Analytical Chemistry, Technical University of Budapest, Szt. Gellért tér 4, H-1111 Budapest, Hungary.

0095-2338/92/1632-0191$03.00/0

Pare). The comparison performed by CP is based on distance or torsion angle variations that occur in the individual conformers of each molecular structure. The strategy utilized in CP is similar to the one described by Murral and Davies[16] in that a rapid screening step, which relies upon the use of bit maps, is followed by a more time-consuming but comprehensive comparison of individual distances or torsion angles.

METHOD

(A) Pairwise Distance Comparison. One method for comparison of molecular conformers involves pairwise matching of distances. CP allows the user to specify distances by designating atom pairs within the conformers of a reference molecule that are matched against corresponding distances (also designated as atom pairs) in the conformers in all the other molecular structures of interest. The pairwise correspondence between the designated distances in the reference structure and the other structures is also explicitly specified by the user. A match occurs when the distances between corresponding atom pairs deviate from one another by no more than a user-specified tolerance, and the tolerance may be specified separately for each designated distance (atom pair). For a match to occur, the distance between a specific atom pair of a conformer in the reference molecular structure must match the distance between the corresponding atom pair in at least one conformer in the second molecular structure; it must also match the distance between the corresponding atom pair in at least one conformer in the third molecular structure, etc. Furthermore, this must be true for all the distances specified relative to the reference structure before CP considers that it has found a match. Note that designation of a reference structure is necessary since the tolerances are assigned relative to the distances computed for the reference structure.

To illustrate this process with a concrete example, let us assume that for the conformers in a reference structure three atom pairs (distances) are to be compared with three corresponding atom pairs (distances) for the conformers in a second structure. Furthermore, let us assume that the three atom pairs will be selected from only three atoms (a common situation for investigation of the spatial relationship between key atoms in the substituents that comprise a postulated pharmacophore but not one that is required by CP). If we designate these three atoms in the reference structure as a, b, and c and the three atoms in the other structure as a′, b′, and c′, the program might be instructed to perform the comparison so that distance ab corresponds to a′b′, bc to b′c′, and ca to c′a′. CP will then locate matches such that distance a′b′ is identical to distance ab within the user-specified tolerance, and likewise distance b′c′ must match bc and distance c′a′ must match ca to within the specified tolerance for at least one conformer of each of the two different structures. Of course, additional matches might exist for other conformers, and CP locates all the matches (see Algorithmic Details below). Implicit in the interatomic distance comparison we have just described is correspondence between the atoms themselves, and it follows that for conformers where these distances match, the molecular structures can be oriented in such a way that atom a can be superimposed with atom a′, b with b′, and c with c′.

(B) Permuted Distance Comparison. CP also has an option for permuted distance matching in which the user does not explicitly designate atom pairs nor is the correspondence between such atom pairs designated. Instead the user simply indicates a set of atoms for comparison in each molecular structure. CP automatically permutes the interatomic distance relationships of the specified atoms in the various molecular structures relative to those in the reference structure. This operation corresponds to the hypothetical example described above where the user would merely specify that atoms a, b, and c in the reference structure must match atoms a′, b′, and c′ in the other structure. The program would then test for matches in which the three possible interatomic distances in the reference structure are compared to every permutation of the three possible interatomic distances in the second structure. The comparison, as for the pairwise distance matching, is done on ordered distance pairs, and thus the comparison performed on the example described above for pairwise distance matching would be only one of the possible permutations; another would be the one where ab corresponds to b′c′, bc corresponds to c′a′, and ca corresponds to a′b′. The function of this capability of the program is to reveal different relative spatial orientations of the selected atoms within the conformers of the molecular structures that match corresponding atoms in conformers of the reference file in ways that may not have been previously considered. It also allows the distance comparison to be performed in an unbiased manner.

In the case of permuted distance comparison, only one tolerance is specified. Since the distances to be compared are permuted, it would not make sense to compare the distances with different tolerances.

The program also allows for the specification of user-defined atom types as part of the permuted distance comparison option. The atom type descriptor is an integer and can be assigned any arbitrary value, e.g., −1 for negatively charged atoms, 1 for positively charged atoms, and 0 for atoms that bear no charge. The specification of atom types allows for the suppression of matches in which the matched interatomic distances lead to situations where atoms of different types match one another, destroying, for example, a desired electrostatic pattern.

(C) Torsion Angle Comparison. CP has as well a facility for comparison of torsion angles rather than distances. The torsion angle comparison is virtually identical to the pairwise distance comparison except that torsion angles (specified as atom quartets) are specified instead of atom pairs.

(D) Algorithmic Details. An exhaustive (or "brute force") approach for comparison of two structures, both of which contain several conformers, requires ij comparisons, where i and j are the number of conformers belonging to structures 1 and 2. For three or more structures, ijk ... comparisons, where i, j, and k ... are the number of conformers belonging to structures 1, 2, and 3 ..., are required. In other words, the brute force approach requires the comparison of every conformer of the reference structure with every conformer of all the other structures.

CP uses a different strategy that is much more rapid, if the number of matches is far fewer than the total number possible. This situation is the one that is frequently observed in practice since the user is generally not interested in data where every conformer of the reference structure matches every conformer of all the other structures. In CP, a "two-pass" procedure is employed. The first pass involves the rapid location of conformers whose specified distances or torsion angles fall into ranges where matches can possibly occur. Perhaps, more importantly, conformers are excluded from further comparisons that possess one or more distances (or torsion angles) that fall outside of the allowed ranges. First, the distance (or torsion angle) ranges are coded as bits in a bit string (composed of an array of normal 32-bit computer words). This coding allows the extremely rapid extraction of distances (or torsion angles) that are within the allowed ranges by utilization of bitwise Boolean algebra available within many computer languages. Next, the conformers that possess distances falling into these ranges are extracted (also using bitwise Boolean operations). The second pass is performed on these conformers and is simply the brute force procedure applied to this (usually small) subset of match-candidate conformers provided by

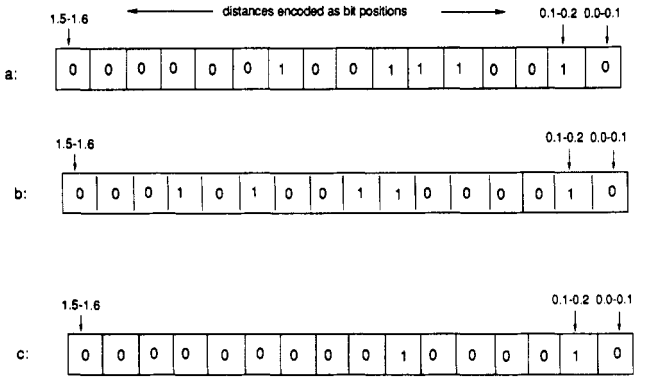


Figure 1. (a) Bit string encoding distances found in each conformer in a reference structure for the first atom pair. (b) Bit string encoding distances found in each conformer in a second structure for the corresponding first atom pair. (c) Boolean .AND. of the bit strings depicted in a and b, which indicates those distances that conformers of the reference structure have in common with the conformers in the second structure for the first atom pair.

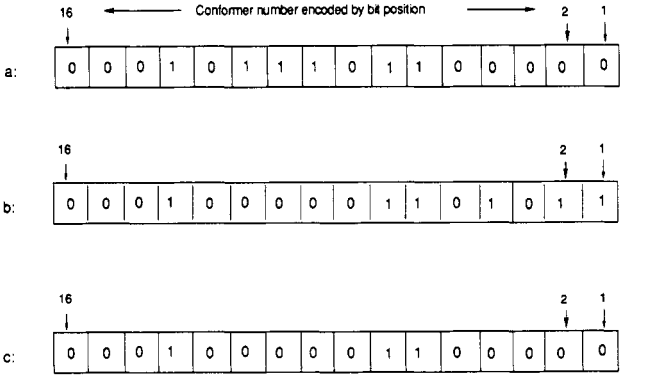


Figure 2. (a) Bit string encoding conformer numbers of conformers in the reference structure that have distances in common with conformers of the second structure for the first atom pair specified. (b) Bit string encoding conformer numbers of conformers in the reference structure that have common distances for the second atom pair specified. (c) Boolean .AND. of the bit strings depicted in a and b, which indicates those conformers in the reference structure that match conformers in the second structure with respect to the two atom pairs specified.

application of the Boolean techniques described above.

Figure 1 depicts how the bitwise encoding of distances (or torsion angles) followed by Boolean operations is accomplished by CP. FIgure 1a illustrates a bit string, which is shown as one composed of 16 bits (i.e., a short integer computer word; CP actually uses an array of 32-bit long integers for its bit encoding). Each bit position corresponds to a particular distance (within a resolution of 0.1 Å), and CP would encode into the same bit string all the distances calculated for the first atom pair of each conformer in the reference file. Similarly, encoding of all the distances calculated for the corresponding atom pair of all the conformers in a second structure would be encoded in a second bit string as illustrated in Figure 1b. The number of such bit strings, of course, would depend upon the number of structures being matched. The Boolean .AND. operation applied to the two bit strings shown in Figure 1, parts a and b, would then yield the set of *distances* for the first atom pair that conformers of the two structures have in common as illustrated in Figure 1c. This procedure is then applied for each of the other atom pairs being matched. The tolerance is handled by setting additional bits adjacent to the one appropriate for the calculated distance, although for the sake of the present illustration tolerance bits are not shown. A second bit string as depicted in Figure 2a is used to store the conformer number of those conformers in the reference structure that possess distances in common with distances in the second structure for the first atom pair being compared. If more than two structures are being compared, the conformer numbers refer to those conformers of the reference structure that possess distances in common with all other structures. Figure 2b illustrates a second bit string used to store conformer numbers for the next atom pair being compared. The number of such bit strings would depend on the number of atom pairs selected by the user. The Boolean .AND. operation is then employed to yield a bit string that indicates those conformers in the reference structure that match conformers in the second structure for every distance specified. This procedure is then applied for the second structure (and any consecutive structures).

At this point the first pass is complete, and the second pass begins. The information provided by pass one is a list of conformations of the reference structure that can conceivably match at least one conformer in each and every other structure, and a similar list is generated for all other structures. These lists are stored as bit strings as described above. Thus, each conformer in the list for the reference structure matches at least one conformer in the list for the second structure, at least one conformer in the list for the third structure, etc. Furthermore, each conformer in the list for the second structure matches at least one conformer in the list for the reference structure, at least one conformer in the list for the third structure, etc. Nonetheless, the individual conformer-to-conformer correspondence is unknown at this point. This correspondence is determined, in the second pass, by application of the brute force matching procedure on the match candidates found in the lists described above. The output of pass two is a set of so-called match groups. A match group is a set of conformers from every structure that match the very same conformer, which we refer to as the *template* conformer, of the reference structure.

Simple benchmarks can be used to determine the utility of this approach. The time required to do a complete brute force comparison on two structures, each with 100 conformers, can be estimated by allowing CP to operate on two identical test sets of structures. We chose to employ a rigid structure that had been "cloned" 100 times, but of course, any single conformer of any structure could be employed. Thus, both data sets contained the same structure, a tetrasubstituted adamantane. The six possible interatomic distances between the substituents were compared such that each distance (atom pair) in the first data set was matched against its corresponding distance (atom pair) in the second data set. Therefore, in this "worst case scenario," every "conformer" in the reference structure matches every "conformer" in the second structure and hence there are 100 × 100 or 10 000 legitimate matches. The Boolean, first pass, when applied to the test data described above required 0.15 CPU s on a VAX 8820, and the brute force second pass required 7.46 CPU s. A second benchmark in which two identical test sets of structures were set up as before but in which there were no matches (since the program had been erroneously instructed to compare distances that could not possibly match each other) required 0.12 CPU s for the Boolean first pass and 0.02 CPU s for the second pass.[23] Thus, the second pass is more than 350 times faster in the second test case than in the first with an investment of only roughly 0.1 CPU s. Of course, the actual gain is determined by the total number of potential matches relative to the actual number of matches. In real cases the expected gain in performance by doing the Boolean first pass should be between 1 and 2 orders of magnitude!

The permuted distance comparison is also carried out in two passes. Bitwise Boolean operations are also applied; however, the permuted distance comparison is more closely related to a brute force procedure than the pairwise distance matching. This is because the distances to be compared are permuted

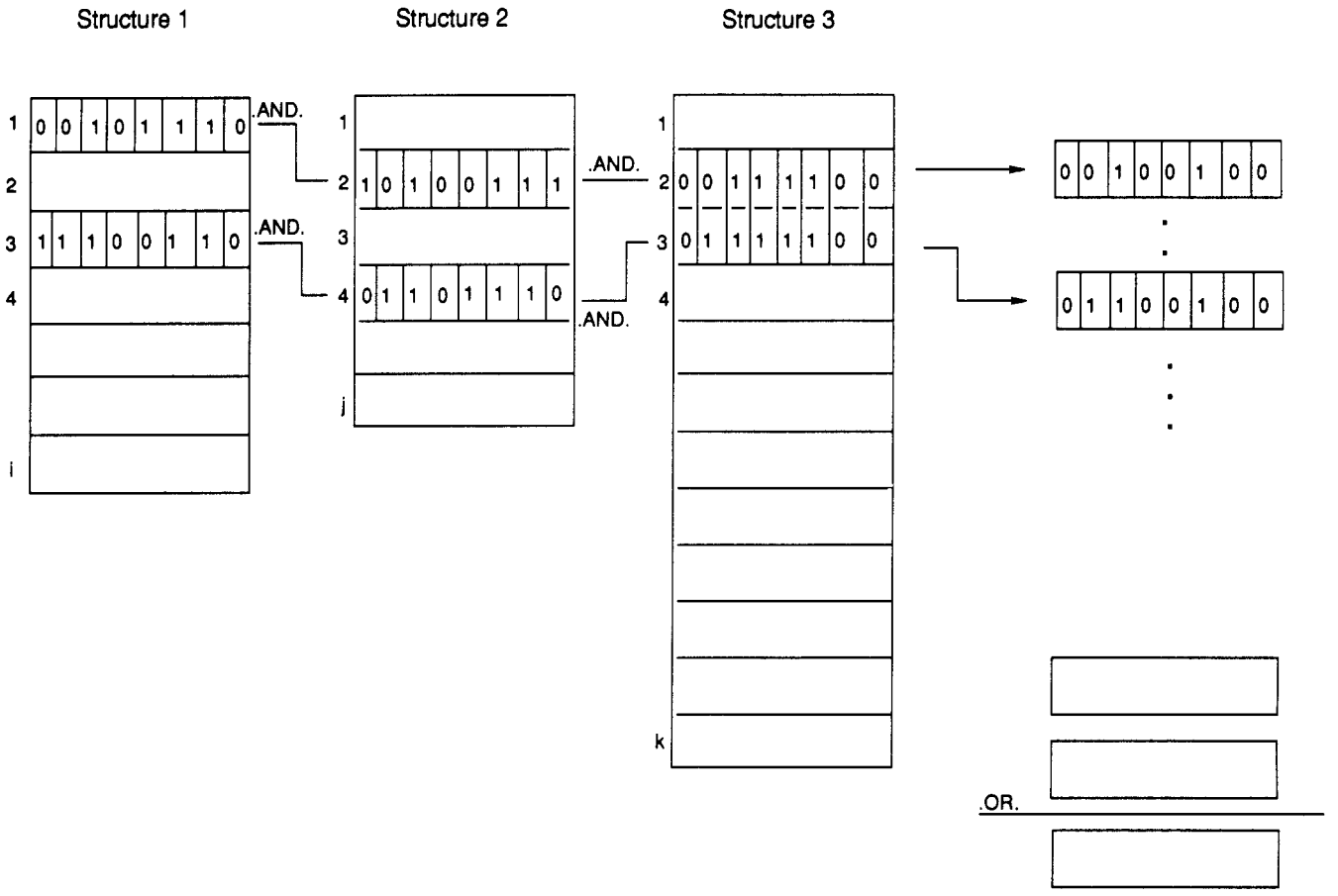


Figure 3. Bit string representation of the $N(N-1)/2$ interatomic distances associated with N atoms selected from each of three sets of conformers of three different structures. (Only an eight-bit segment of the bit string is shown.) All possible bit string triplet combinations are separately .AND.-ed (horizontally), affording the bit strings shown on the right. These bit strings are then .OR.-ed (vertically) yielding a single bit string as shown on the bottom right. The bits that are set in this bit string correspond to distances that are common to conformers contained in every horizontal combination.

and therefore cannot be handled separately. In other words, the distance information present in the conformer geometries is compressed into a single bit string rather than in several bit strings, with a concomitant loss of information. Consequently, the algorithm for permuted distance comparison is different from the algorithm for the pairwise distance comparison described above.

Before presenting the algorithm, one more item must be considered. If we select N atoms (some of which may be dummy atoms, e.g., the centroid of a benzene ring) to define the pharmacophore, there are N factorial ways to superimpose the N atoms selected from each of two different structures. We describe this operation as permuting the N atoms. If, on the other hand, we consider working in distance space and permute the corresponding $N(N-1)/2$ interatomic distances, a significant amount of redundancy is introduced. In fact, $3N-6$ distances are enough to define the relative spatial orientation of the atoms, which is a smaller number of distances than $N(N-1)/2$ for $N > 4$. Thus for $N > 3$ there will be redundancy relative to permuting individual atoms if all interatomic distances are permuted. The algorithm applied in CP uses all the $N(N-1)/2$ distances, nonetheless, a little more intelligently than exploring $N(N-1)/2$ factorial permutations. If, in fact, there are far less ways to match the N atoms than the total number of permutations, which is usually expected to be the real situation, the algorithm employed by CP becomes quite efficient.

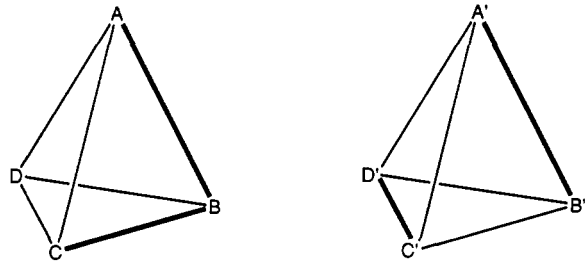
In pass one, a nested loop system is generated where the number of loops is equal to the number of molecular structures. The outermost loop represents the reference structure, the second loop the second structure, etc. The loop indices span from one to the number of conformers of each corresponding structure. This loop system generates an exhaustive set of index n-uples (where n is the number of structures). One such n-uple corresponds to one particular combination of conformers taking one conformer from each structure. This situation is illustrated in Figure 3 in which a schematic diagram for three structures is shown. The number of blocks illustrated for each structure corresponds to the number of conformers of each structure (i, j, k). Each block shows eight bits of the distance bit string representation of the corresponding conformer with all of the $N(N-1)/2$ distance bits set. For sake of simplicity, tolerance bits are not shown.

Now, for each horizontal block combination (two combinations or n-uples are shown in Figure 3) the Boolean .AND. operator is used to extract those distances that are common to all the conformations of that combination. All these .AND.-ed bit strings (shown in the fourth column in Figure 3) are then combined using the Boolean .OR. operator. This operation gives a single bit string with bits set, which correspond to the common distances arising from all possible conformer combinations.

The next step is similar to what we have described with the pairwise matching. All conformations of all structures are tested against these common distances, and the conformations with $N(N-1)/2$ hits are kept as match candidates. In other words, all those conformations that have at least one non-common distance, and could not possibly match any other conformation, are discarded. Certainly, in all comparisons, the equivalence of two distances means equivalent within the bounds of the user-defined tolerance. Thus, the outcome of pass one is a list of match candidates.

In pass two, a permuted brute force matching takes place with the set of match candidates. Now instead of matching designated distances, two distance matrices are compared, both with $N(N-1)/2$ distances. If a unique pairwise correspondence can be found between the elements of two matrices within the tolerance, then this situation represents a hit or a pseudohit. The term pseudohit refers to situations in which the numerical distance values match but the distances are in an incorrect geometrical orientation to one another. Since the distances are not just numerical values but also represent dummy bonds between the atom pairs associated with them, the pairwise correspondence between distances implicitly determines a connectivity pattern. This means that a hit not only requires the unique pairwise correspondence of the elements of two distance matrices but also requires the equivalence of the associated connectivity pattern.

The occurrence of pseudomatches can be illustrated by the following example. Consider, for instance, an atom quartet A-B-C-D, which spans a regular tetrahedron. Also consider an atom quartet A′-B′-C′-D′ from a second structure, which also spans a regular tetrahedron. Assume that A, B, C, and D represent atomic positions in at least one conformer in a reference structure and that A′, B′, C′, and D′ represent atomic positions in at least one conformer belonging to a second structure. Also assume that a permuted distance match is to be performed. In this case two identical 6 × 6 distance matrices will be compared (representing the single common length of the edges of the regular tetrahedron), and based solely upon distance criteria, the six edges of tetrahedron A-B-C-D could correspond to any permutation of the six edges of A′-B′-C′-D′. Nonetheless, a situation where edge AB corresponds to A′B′ and BC corresponds to C′D′, for example, is a pseudomatch since AB and BC share a common vertex, whereas A′B′ and C′D′ do not. Generally, such pseudomatches occur rarely unless a loose tolerance is specified.



The equivalence of the connectivity pattern can be tested in different ways. For example, Jakes et al.[13] have described the so-called set reduction technique for this purpose. Our approach is different and is based on the fact that any number of simultaneous swappings of both the rows and columns i and j of a connectivity matrix do not change the connectivity pattern, it simply corresponds to a different atom numbering system. A simple function, the determinant of a matrix, is invariant to such simultaneous row and column swappings, and therefore the determinant is a good descriptor of a connectivity pattern. Thus, the connectivity test is applied as follows. After the pairwise correspondence between the elements of the two distance matrices has been established, every element of the distance matrix for the conformer being compared to the template conformer is replaced with the corresponding element of the distance matrix for the template. In this way we obtain two matrices, the original one for the template and the one just described with the very same elements in a different order. Now calculating the determinant of both matrices will reveal if they represent the same connectivity pattern. The two matrices are considered to be equivalent, and therefore a hit is found if the determinants are identical, otherwise a pseudohit is found, which is not of interest. It is instructive to note that in some cases the difference between two determinants is only displayed in the 14th or 15th decimal place. Conceivably, depending upon the algorithm used to compute the determinant, the 14th–15th digit is not significant due to numerical inaccuracies. Nonetheless, this does not affect our connectivity test, because if the two matrices are indeed equivalent as far as the connectivity pattern is concerned, the two determinants will be *exactly* the same regardless of the actual numerical value (if calculated directly with a noniterative algorithm).

At this point a hit is found. Nonetheless, it could well be that more than one unique correspondence can be found between the elements of the two distance matrices. This leads to potential matches where the template conformer in the reference structure can match the very same conformer in one of the other structures in different ways. The worst case scenario occurs when all of the elements of the two distance matrices are identical within the tolerance. In this situation there are $N(N-1)/2$ factorial permutations that should be considered. This would be very time consuming even with $N = 5$ atoms. Fortunately, the real situation is usually not even close to being this formidable, unless a very loose tolerance is specified.

The pursuit of such permuted matches is established as follows. For every matrix pair, a nested loop system is generated. The number of nested loops is $N(N-1)/2$. The outermost loop corresponds to the first element of the reference matrix, and the innermost loop corresponds to the last element of the reference matrix. The loop indices span from 1 to the number of distances found in the other matrix that are identical within the tolerance with the distance in the reference matrix correspondng to the loop number. If all the upper limits of the loop indices are 1 then there is only one potential way of matching, and the connectivity test described above is then applied for this match. If the upper limit of one or more loop indices is greater than 1, there are more permutations that will then all undergo the connectivity test. The total number of these permutations is equal to the product of the $N(N-1)/2$ upper loop index limits. It is obvious that this second pass is likely to be the bottleneck of the algorithm, and it would be a difficult task to predict how much time this process would require for any given situation.

The output of pass two is similar to the one for pairwise matching: a set of match groups is produced with additional information specifying exactly how the distances (and implicitly how the N atoms of the pharmacophore) in the member conformations of a match group align with the pharmacophore of the template.

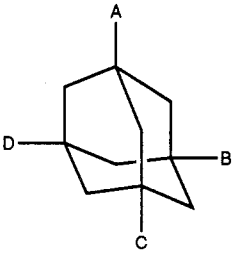
Having established the relative spatial orientation of key pharmacophoric substituents in a hit, an atom type match can be employed as well. A user-defined atom type specifier may be assigned for each atom (real or dummy) of the pharmacophore, e.g., −1 for negatively charged atoms, 1 for positively charged atoms, and 0 for nonpolar atoms. The program will then check whether or not the atoms selected for permuted matching match each other with respect to their assigned atom type. A permutation passes the atom type test if for all the $N(N-1)/2$ distances, the associated atom types (two per distance according to the atom pair associated with the distance) match each other (head-to-head or head-to-tail) according to the pairwise correspondence of the distances in that permutation. Since such a permutation determines a unique connectivity pattern (see the connectivity test above) and therefore an atom-to-atom correspondence, the atom type based upon the comparison of the atom type pairs of all the $N(N-1)/2$ interatomic distances is equivalent with the atom type test based upon the comparison of the single atom type of the N atoms.

```
1           !switch 1=pairwise (distance),2=permuted (distance),3=torsion angle
3           !Number of Files
4           !Number of distance pairs to be considered per molecule
0.7         !Tolerance in Angstroms, >= 0.0 (+/-, resolution = 0.1)
0.7         !Tolerance for second distance pair
0.1         !Tolerance for third distance pair
0.6         !Tolerance for fourth distance pair
ACCHO.OUT   !Name of the 1st file [REFERENCE] (acetylcholine conformers)
1 5         !1st atom pair
1 7         !2nd atom pair
5 7         !3rd atom pair
1 6         !4th atom pair
NIC.OUT     !Name of the 2nd file (nicotine conformers)
7 13        !1st atom pair
7 1         !2nd atom pair
13 1        !3rd atom pair
7 5         !4th atom pair
MUSC.OUT    !Name of 3rd file (muscarone conformers)
7 3         !1st atom pair
7 12        !2nd atom pair
3 12        !3rd atom pair
7 4         !4th atom pair
```

Figure 4. Instruction file for pairwise distance matching of acetylcholine, (−)-nicotine, and (−)-muscarone conformers with respect to the atoms shown.

Finally, the issue of molecular chirality must be considered. CP compares structures for permuted matching based solely upon distances, and thus molecular chirality information (if it exists in a particular structure) is not considered during the comparison. It is possible, for example, that a match will occur between the template and a conformer in one of the other structure files and that a *permuted* match will be found that corresponds to the enantiomer of this same conformer. There is no explicit option for chirality checking in the current version of CP. Thus, molecular superimposition (which is the graphics output option in CP) cannot be performed between the template conformer and the enantiomeric conformer *since it does not exist in the structure file.* CP will attempt to superimpose the template conformer with the conformer that actually exists within the structure file, and in fact, currently, this is the only way to detect enantiomeric matches with CP. If the user finds that some of the superimposition RMSs within a match group are considerably greater than the others in the same match group, then this may be a consequence of chirality incompatibilities. Note that this "feature" in CP might be of some utility since enantiomeric matches not previously considered may be identified.

A simple test case can be used to illustrate the chirality aspect of the permuted distance comparison. For this test case we have selected an asymmetrically tetrasubstituted adamantane, although any other chiral molecule could serve as an example. The four different substituents, located at the tertiary carbon atoms of the adamantane skeleton, span an almost perfect tetrahedron.



Permuted distance matching of the four substituents can be performed on two *identical* files containing one of the two enantiomeric forms of the tetrasubstituted adamantane. In this case (N = 4) there are four factorial permutations. However, the 24 permutations represent an enantiomeric pair. The reference structure (one of the enantiomers) can only be superimposed with itself in 12 different ways, the other 12 permutations correspond to the 12 different ways in which the enantiomeric structure can be superimposed with the template. Indeed, CP is able to locate all 24 permutations when an appropriate tolerance is employed. Twelve permutations have a superimposition RMS less than 0.2 relative to the template,

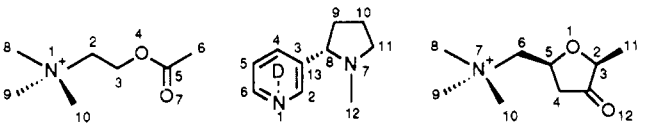
```
(A)nalysis or create MMOD (O)utput files of superimposed matching structures
from a previous analysis? [A]

Provide the name of the matchlist file! [MATCHLIST]

Do you want to create a new set of MMOD multiconformer input files
with the non-matching conformers removed? [N]

The total number of conformers read from file     ACCHO.OUT is   491.
The total number of conformers read from file     NIC.OUT is     6.
The total number of conformers read from file     MUSC.OUT is    4.

SUMMARY: 4 match-groups, 14 hits
File 1   file 2      file 3
Conf #   # of confs  # of confs  # of hits     <RMS>  SD
  67        2           2           4        0.501  0.023
 124        2           2           4        0.501  0.023
 320        1           2           3        0.486  0.024
 398        1           2           3        0.486  0.024

Dist #      1      2      3      4
Toler     0.7    0.7    0.1    0.6
<Diff>   0.53   0.66   0.00   0.46
SD       0.11   0.05   0.00   0.19
Maxdiff  0.68   0.75   0.01   0.66

Do you want to (S)ave the matchlist or (E)xit? [E] s
Matchlist file created.

You are advised to study the matchlist file before using this option!
There are 4 groups of matching structures.

Do you want to create a MMOD file of the superimposed structures? [N] y

Select a group and provide a filename! [1 GRAPHICS]

Conformer   67 in file   ACCHO.OUT has been superimposed with
conformer    1 in file   NIC.OUT,    RMS = 0.2135
conformer    2 in file   NIC.OUT,    RMS = 0.1801
conformer    1 in file   MUSC.OUT,   RMS = 0.3177
conformer    2 in file   MUSC.OUT,   RMS = 0.3440

Specify the RMS limit above which you want matches to be discarded! [None]
Do you want to create another MMOD file? [N]

CPU Time: 7.9 seconds*

*For a VAX 6510
```

Figure 5. Dialogue between the user and the CP program for pairwise distance matching of nicotinic acetylcholine agonists. The default responses are shown in brackets.

and the other 12 permutations have a superimposition RMS between 1.9 and 2 since, as discussed above, these matches refer to the enantiomer not actually present in the second structure file.

RESULTS

In order to demonstrate the utility of the CP program for a problem of pharmacological interest, a series of compounds that are purported to bind to the nicotinic acetylcholine receptor as agonists[24] was chosen for pairwise distance comparison. The molecules chosen were (−)-nicotine, (−)-muscarone, and acetylcholine itself. In order to obtain a viable set of conformers for input into the CP program for comparison, each molecule was subjected to a systematic grid search procedure using the MULTIConformer submode of the MacroModel program.[25] A dummy atom was added to the centroid of the pyridine ring to facilitate comparison of the purported pharmacophoric substituents.



acetylcholine (−)-nicotine with dummy atom placed at centroid (−)-muscarone

The search was performed at 30° resolution, and conformers were rejected that possessed an occurrence of at least one nonbonded contact less than 1.5 Å. No energy minimization was performed, but the conformers passing the nonbonded rejection criterion were "filtered" based upon their MM2 energies. Only conformers within 30 kJ of the lowest energy

```
File 1 is ACCHO.OUT with 491 conformers.
File 2 is  NIC.OUT with   6 conformers.
File 3 is MUSC.OUT with   4 conformers.

Selected atom pairs:

File 1:  1- 5  1- 6  5- 6  1- 7
File 2:  5- 13  5- 12  13- 12  5- 9
File 3:  7- 11  7- 12  11- 12  7- 1

Distance tolerance for pair  1 = 0.7 Angs.
            for pair  2 = 0.7 Angs.
            for pair  3 = 0.1 Angs.
            for pair  4 = 0.6 Angs.

 Match Group  Distances between selected atom pairs
```

# 1	
File 1 Cnf 67	4.4925 5.4296 1.2089 4.5222
File 2 Cnf 1	3.8150 4.6837 1.2016 4.3441
File 2 Cnf 2	3.8359 4.7984 1.2016 4.2148
File 3 Cnf 1	4.8941 6.0825 1.2074 3.9333
File 3 Cnf 2	4.9509 6.1204 1.2074 3.8668

Figure 6. Text output for pairwise distance matching of nicotinic acetylcholine agonists. Only distances for the first match group are shown.

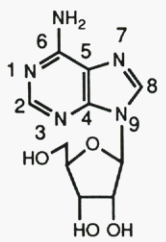
conformer were kept for further analysis. Since CP uses MacroModel multiconformer structure files for its input (and writes MacroModel multistructure files for its graphics output), the files produced by MacroModel as a result of the MULTIC search served as input files to CP.

An instruction file is used to set up the pairwise distance matching, and an example of such a file for this particular example is shown in Figure 4. In the instruction file the user specified the file names for the structures being compared. The first file specified is considered to be the reference file. In addition, the atom pairs defining the distances to be compared and the tolerances for the comparison are specified. The order in which the atom pairs are listed in the instruction file indicates the pairwise correspondence to the atom pairs listed for the other structures.

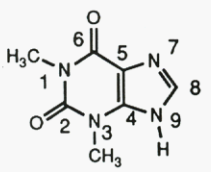
Figure 5 depicts the CP user dialogues and Figure 6 shows the text output for this example. Figure 7 shows the graphics output displayed using MacroModel for one of the match groups located by CP. In Figure 7, conformers of nicotine and muscarone are superimposed with a conformation of acetylcholine that is identical (within the 30° resolution employed for the systematic search) to the one derived by NMR for acetylcholine bound to its nicotinic receptor.[26] Although the relevance of this conformation of acetylcholine has been called into question,[27] the analysis described above with the CP program demonstrates that both nicotine and muscarone are able to adopt a conformation of reasonable energy that favorably matches the NMR-derived conformer of acetylcholine.

To demonstrate the utility of the permuted distance comparison option of the CP program, we have chosen an example that on the surface appears trivial but, upon closer inspection, reveals a situation that arises in the molecular modeling of active analogues that is of considerable practical importance. Peet et al. have recently reported a new binding mode for xanthines which bind to the adenosine receptor.[28] In particular, two different receptor binding modes for theophylline relative to adenosine have been proposed. One binding mode, referred to as the "standard overlay mode", in which all four ring nitrogen atoms are superimposed and another, in which only three ring nitrogens superimpose, have been described. In the nonstandard overlay mode, the theophylline molecule has been flipped and rotated relative to the standard orientation.

The CP program was able to locate both binding orientations when the three distances between ring nitrogen atoms N1, N3, and N9 of adenosine and theophylline were selected for permuted distance comparison. Since rigid structures were being



Adenosine



Theophyline

compared, obviously only one conformer for each molecule was employed. In the standard overlay mode N1, N3, N7, and N9 of theophylline are overlaid with N1, N3, N7, and N9 of adenosine, respectively. In the nonstandard overlay mode, N9, N3, and N1 of theophylline are overlaid with N1, N3, and N9 of adenosine. Thus not only was the, perhaps obvious, standard overlay mode "discovered" by CP but also the nonobvious alternative overlay mode was also obtained "automatically". Of course, some insight by the user is necessary as it is required that the user select the appropriate atoms for matching. Figure 8 shows the nonstandard superimposition of theophylline onto the adenosine template.

To demonstrate the utility of CP for torsion angle matching, the file containing the acetylcholine conformers as described above was compared to a file containing the X-ray structure for acetylcholine.[29] Two conformers were located when a tolerance of 15° was employed. Quite reasonable agreement (within the 30° resolution of the systematic search) with the X-ray structure was observed. Figure 9 shows a superimpo-

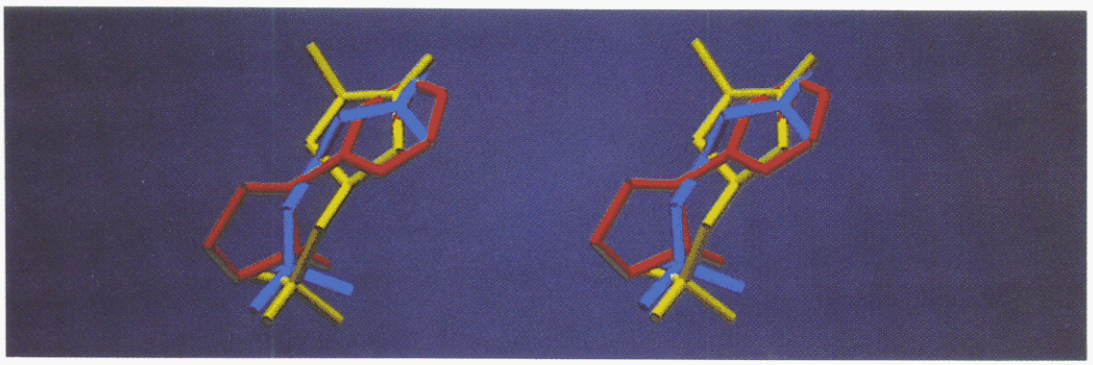


Figure 7. Superimposition (stereoview) of a template conformer of acetylcholine (blue) with conformers of (−)-nicotine (red) and (−)-muscarone (yellow) from one of the match groups produced by CP as a result of a pairwise distance matching procedure.

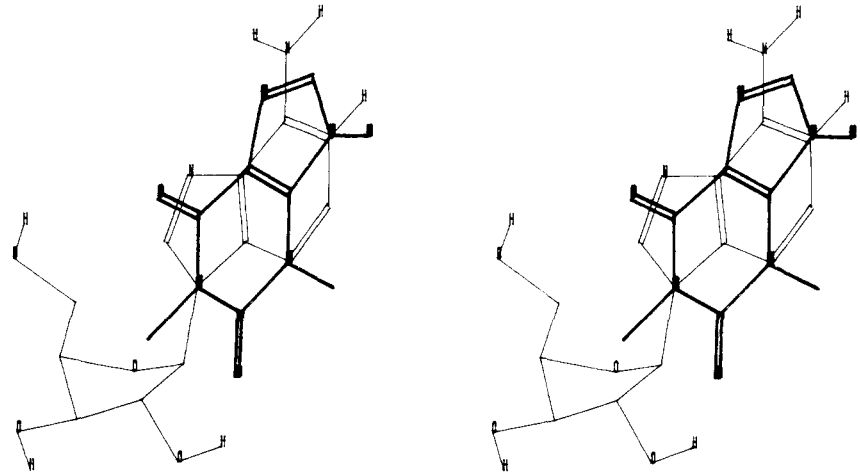


Figure 8. Superimposition (stereoview) of the template (adenosine) with theophylline (in bold) in the nonstandard overlay mode produced by CP as the result of a permuted distance matching procedure.

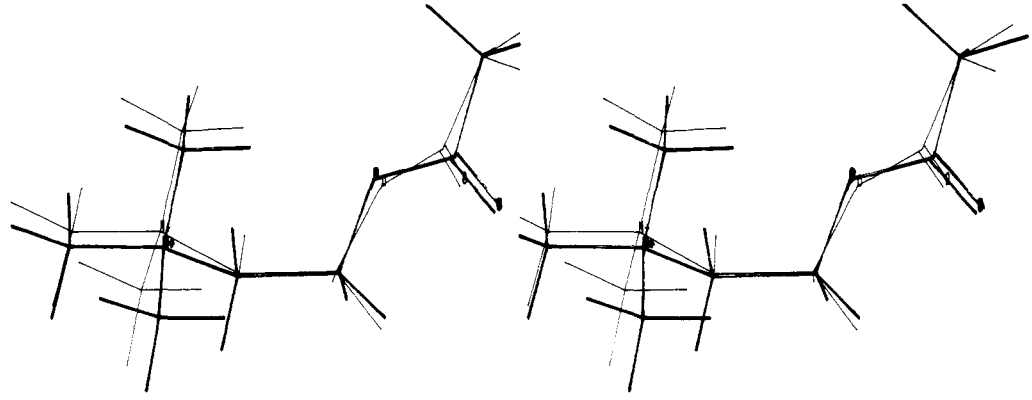


Figure 9. Superimposition (stereoview) of the template (X-ray structure of acetylcholine, shown in bold) with one of the conformers located by CP from a set of acetylcholine conformers as the result of a torsion angle matching procedure.

sition of the X-ray structure and one of the matching conformers for acetylcholine.

SUMMARY

In conclusion, CP allows the rapid comparison of conformers of different molecular structures based on either interatomic distances or torsion angles. Unlike methods such as ensemble distance geometry,[7] CP requires prior collection of the conformations of the molecules being compared. In our experience, this requirement is not a liability and is frequently an advantage. Since distance geometry methods do not explicitly take molecular energetics into account, structures produced using distance geometry methods are likely to be unfavorable energetically, particularly when flexible molecules are involved. CP can be applied to low energy structures that have been subjected to energy minimization, or at least have been filtered with respect to their internal energies. Furthermore, as force fields are improved so that solvation effects are taken into account,[30] CP will be useful for determining whether reasonable pharmacophores can be derived from solution conformations.

Like other methods for locating conformers of active analogues that allow superimposition of key atoms or functional groups, CP obligates the user to supply some information concerning the pharmacophoric substituents. However, CP allows considerable flexibility in this specification. The permuted match option of CP only requires identification of the substituents and, in most cases, the atom type to which they belong. No presumption concerning the correspondence of the substituents is necessary, and hence this option of CP should be highly useful for identification of the pharmacophore without bias.

We believe that CP will prove to be quite useful for drug design purposes in locating conformers of the relevant structures having key pharmacophoric substituents in the same relative spatial orientation.[31]

ACKNOWLEDGMENT

We thank Dr. Regine Bohacek for discussions that eventually led to the Compare_Conformer program. We also thank Dr. Greg Paris for assisting in the CP user interface.

REFERENCES AND NOTES

(1) Marshall, G. R.; Barry, C. D.; Bosshard, H. E.; Dammkoehler, R. A.; Dunn, D. A. The Conformational Parameter in Drug Design: The Active Analog Approach. In *Computer-Assisted Drug Design*; *ACS Symposium Series 112*; Olson, E. C., Christoffersen, R. E., Eds.; American Chemical Society: Washington, DC, 1979; pp 205–226.

(2) Motoc, I.; Dammkoehler, R. A.; Mayer, D.; Labanowski, J. Three-Dimensional Quantitative Structure–Activity Relationships. I. General Approach to the Pharmacophore Model Validation. *Quant. Struct.–Act. Relat.* **1986**, *5*, 99–105.

(3) Motoc, I.; Dammkoehler, R. A.; Marshall, G. R. Three-Dimensional Structure–Activity Relationships and Biological Receptor Mapping. In *Mathematical and Computational Concepts in Chemistry*; Trinajstic, N., Ed.; Ellis Horwood Ltd.: Chichester, 1986; pp 222–251.

(4) Mayer, D.; Naylor, C. B.; Motoc, I.; Marshall, G. R. A Unique Geometry of the Active Site of Angiotensin-Converting Enzyme Consistent with Structure–Activity Studies. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 3–16.

(5) Crandell, C. W.; Smith, D. H. Computer-Assisted Examination of Compounds for Common Three-Dimensional Substructures. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 186–197.

(6) Ghose, A. K.; Crippen, G. M. Use of Physicochemical Parameters in Distance Geometry and Related Three-Dimensional Quantitative Structure–Activity Relationships: A Demonstration Using *Escherichia coli* Dihydrofolate Reductase Inhibitors. *J. Med. Chem.* **1985**, *28*, 333–346.

(7) Sheridan, R. P.; Nilakantan, R.; Dixon, J. S.; Venkataraghavan, R. The Ensemble Approach to Distance Geometry: Application to the Nico-

tinic Pharmacophore. *J. Med. Chem.* **1986**, *29*, 899-906.
(8) Takahashi, Y.; Maeda, T. S.; Sasaki, S. Automated Recognition of Common Geometrical Patterns Among a Variety of Three-Dimensional Molecular Structures. *Anal. Chim. Acta* **1987**, *200* (1), 363-377.
(9) Ghose, A. K.; Crippen, G. M.; Revankar, G. R.; McKernan, P. A.; Smee, D. F.; Robins, R. K. Analysis of the in Vitro Antiviral Activity of Certain Ribonucleosides against Parainfluenza Virus Using a Novel Computer-Aided Receptor Modeling Procedure. *J. Med. Chem.* **1989**, *32*, 746-756.
(10) Takahashi, Y.; Akagi, T.; Sasaki, S. Three-dimensional pharmacophoric pattern search using COMPASS: Nootropic Agents. *Tetrahedron Comput. Methodol.* **1990**, *3*, 27-35.
(11) Austel, V.; Muller, P.; Reiffen, M. Systematic Consideration of Conformational Properties in Series Design. In *QSAR in Drug Design and Toxicology*; Hadzi, D., Jerman-Blazic, B., Eds.; Elsevier Science Publishers B.V.: Amsterdam, 1987; pp 85-89.
(12) Jakes, S. E.; Willett, P. Pharmacophoric Pattern Matching in Files of 3-D Chemical Structures: Selection of Interatomic Distance Screens. *J. Mol. Graphics* **1986**, *4*, 12-20.
(13) Jakes, S. E.; Watts, N.; Willett, P.; Bawden, D.; Fisher, J. D. Pharmacophoric Pattern Matching in Files of 3-D Chemical Structures: Evaluation of Search Performance. *J. Mol. Graphics* **1987**, *5*, 41-48.
(14) Brint, A. T.; Willett, P. Pharmacophore Pattern Matching in Files of 3D Chemical Structures: Comparison of Geometric Searching Algorithms. *J. Mol. Graphics* **1987**, *5*, 49-56.
(15) Sheridan, R. P.; Rusinko, A.; Nilakantan, R.; Venkataraghavan, R. Searching for Pharmacophores in Large Coordinate Data Bases and Its Use in Drug Design. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8165-8169.
(16) Murral, N. W.; Davies, E. K. Conformational Freedom in 3-D Databases. I. Techniques. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 312-316.
(17) Motoc, I.; Marshall, G. R.; Dammkoehler, R. A.; Labanoski, J. Molecular Shape Descriptors. 1. Three-Dimensional Molecular Shape Descriptor. *Z. Naturforsch.* **1985**, *40A*, 1108-1113.
(18) Bohacek, R. S.; Guida, W. C. A Rapid Method for the Computation, Comparison and Display of Molecular Volumes. *J. Mol. Graphics* **1989**, *7*, 113-117.
(19) Weiner, P. K.; Langridge, R.; Blaney, J. M.; Schaeffer, F.; Kollman, P. K. Electrostatic Potential Molecular Surfaces. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 3754-3758.
(20) Náray-Szabó, G. Electrostatic Complementarity in Molecular Associations. *J. Mol. Graphics* **1989**, *7*, 76-81.
(21) Náray-Szabó, G. Electrostatics in Computer-Aided Drug Design. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1989**, *16*, 87-99.
(22) For a nonelectrostatic-based approach to lipophilicity see: Furet, P.; Sele, A.; Cohen, N. C. 3-D Molecular Lipophilicity Potential Profiles: A New Tool in Molecular Modeling. *J. Mol. Graphics* **1988**, *6*, 182-189.
(23) Since there are no matches for the "brute force" procedure to operate on, the CPU time required for the second pass in this case is due to "overhead" in setting up the appropriate loops.
(24) Spivak, C. E.; Waters, J.; Witkop, B.; Albuquerque, E. X. Potencies and Channel Properties Induced by Semirigid Agonists at Frog Nicotinic Acetylcholine Receptors. *Mol. Pharmacol.* **1983**, *23*, 337-343.
(25) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. MacroModel—An Integrated Software System for Modeling Organic and Bioorganic Molecules Using Molecular Mechanics. *J. Comput. Chem.* **1990**, *11*, 440-467.
(26) Behling, R. W.; Yamane, T.; Navon, G.; Jelinski, L. W. Conformation of Acetylcholine Bound to the Nicotinic Acetylcholine Receptor. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6721-6725.
(27) Spivak, C. E.; Waters, J. A.; Aronstam, R. S. Binding of Semirigid Nicotinic Agonists to Nicotinic and Muscarinic Receptors. *Mol. Pharmacol.* **1989**, *36*, 177-184.
(28) Peet, N. P.; Leutz, N. L.; Meng, E. C.; Dudley, M. W.; Ogden, H. M. L.; Demeter, D. A.; Weintraub, H. J. R.; Bey, P. A Novel Synthesis of Xanthines: Support for a New Binding Mode for Xanthines with Respect to Adenosine at Adenosine Receptors. *J. Med. Chem.* **1990**, *33*, 3127-3130.
(29) Herdlklotz, J. K.; Sass, R. L. The Crystal Structure of Acetylcholine Chloride: A New Conformation for Acetylcholine. *Biochem. Biophys. Res. Commun.* **1970**, *40*, 583-588.
(30) See, for example: Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127-6129. Gilson, M. K.; Honig, B. The Inclusion of Electrostatic Hydration Energies in Molecular Mechanics Calculations. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 5-20.
(31) CP is available from QCPE.

Chemsits' Use of Libraries

BARBARA CHARTON

Hunter College Library, 695 Park Avenue, New York, New York 10021

Received September 17, 1991

Interviews have been conducted with a number of chemists, predominantly in the United States, in industry, government, and academia. Chemists working in some areas of the science depend heavily upon the literature, and such chemists show a uniformity in their work patterns which transcends the differences in their organizations. There are striking similarities in their need for literature and in the practical arrangements that they have made for access to it.

INTRODUCTION

The relationship of scientists to their literature is well-documented. As a professional group, scientists are strongly dependent on the work of their colleagues and predecessors. Unlike law, a field which also is heavily dependent on its literature, science depends on conferences, workshops, and informal networks as well. All these secondary means of generating literature are within the public domain, and all are capable of being accessed by workers in the various scientific fields.

It has been noted that in general, scientists are very aware of their literature. The chemists interviewed are frequent and proficient users of their own literature. They view this as an accomplishment. Examination of some university undergraduate catalogues shows that generally instruction in the use of the chemical literature is not taught by the chemistry department to its majors in a separate course. (The undergraduate catalogues of 12 colleges were examined. Each was the home institution of an interviewed chemist. Only one mentioned library involvement in the catalogue. That one specified the use of library materials in the senior thesis course.) However, when asked, practically every chemist who teaches will make a point of saying that he/she incorporates the literature in the courses taught and brings classes to the library consistently. This is vital since chemistry and its allied fields generate a considerable portion of the scientific literature. It is a reflection of the prominence of chemical and related industries in society.

One constant complaint of the working scientist in the pressure of the need to keep up with an ever-growing literature. Chemists, having the largest literature to cope with, will list their library time as a major component of their professional lives.[1] However, within the large discipline of chemistry in general, a particular group of chemists in the special fields that might be termed "theoretical chemistry" tend to stress their dependence on literature and libraries. Computational chemists who work in physical organic chemistry or molecular design are such specialists.

Such chemists can be found in pharmaceutical companies, government, universities, and chemical companies manufac-

0095-2338/92/1632-0199$03.00/0