

A Combined Model-Building and Distance Geometry Approach to Automated Conformational Analysis and Search[†]

Andrew R. Leach^{*‡} and Andrew S. Smellie[§]

Department of Chemistry, University of Southampton, Southampton SO9 5NH, England, and BioCAD Corporation, 1091 North Shoreline Boulevard, Mountain View, California 94043

Received March 24, 1992

A combined template-joining/distance geometry approach to automated conformational analysis and search is described. The method is based on a previously reported heuristic, knowledge-based system in which conformations of a molecule are constructed by joining together prestored fragments called conformational units. The program identifies the conformational units to be used in a given molecule from among those contained in a knowledge base. A difficulty with such an approach is that the knowledge base may not contain all the units needed to analyze a particular molecule. Thus in the combined algorithm, distance geometry is used to search the conformational space of those fragments that the system does not 'know' about. By this means the system is able to automatically acquire conformational knowledge. We also describe how the combined approach can be used to construct conformations of highly strained ring systems which can be problematic for template-joining algorithms.

INTRODUCTION

A variety of methods have been reported in recent years for searching the conformational space of small- and medium-sized molecules.^{1,2} Most of these methods are designed for the examination of molecules on an individual basis and are capable of locating many, if not all, minimum-energy structures on the conformational energy hypersurface, provided sufficient computer time is expended. They may also be used to derive low-energy structures which satisfy some form of constraint. In a number of cases a combination of techniques has been used to tackle a conformational search problem. Almost exclusively, these studies have combined distance geometry with molecular dynamics to search the conformational space of a peptide, protein, or nucleic acid.^{3,4} Distance geometry is used to generate a family of initial structures. These then serve as starting points for molecular dynamics simulations which explore the conformational space in the vicinity of the initial structures to find energetically more favorable conformations that still satisfy any experimental constraints. Molecular mechanics minimization is usually employed to refine structures selected at intervals from the dynamics trajectory. In such studies, the two searching algorithms (distance geometry and molecular dynamics) operate independently, each being applied separately to the entire molecule under investigation.

In this paper we describe a combined model-building and distance geometry approach to conformational analysis. The joint algorithm is designed to extend the scope of our previously reported knowledge-based program (COBRA⁵) and also to take advantage of the particular strengths of the two algorithms. Much of the impetus for the current work arose from a desire to improve the ability of the method in performing automatic conformational analysis and search for large numbers of diverse molecules. One application of a program with this capability would then be in the generation of conformations to be stored in a '3D database'. Such databases contain the three-

dimensional structures of molecules, obtained from experiment (e.g., X-ray crystallography), from modeling experiments, or from a structure generation algorithm. Such databases can be searched to find compounds whose three-dimensional structure is complementary to a model of the macromolecular receptor (e.g., a crystal structure or pharmacophore map). Our goals are, however, wider in scope than simply structure generation; rather, we are interested in quantifying the conformational space of a molecule to identify its minimum-energy conformations. A conformational search algorithm (as distinct from an algorithm such as CONCORD,⁶ which provides a single structure for each molecule) can also provide information which is of value when *evaluating* the results of a database search. Searches often yield many 'hits' which must be ranked. Knowledge of the conformational space can be useful in this evaluation. For example, molecules for which the proposed active conformation is much higher in energy than the global energy minimum might be discarded. It may also be desired to eliminate molecules which have many accessible conformations due to the high entropic penalty which would be incurred on binding. An estimate of the free energy associated with this entropic penalty can be calculated from the number of low-energy minima and incorporated into a more quantitative scoring scheme for use in the DOCK program,⁷ as will be reported elsewhere.⁸

COBRA uses many of the techniques which were developed as part of the 'WIZARD-II' project.⁹ This approach has been described previously in a number of publications,⁹⁻¹² and so only an outline will be given here, concentrating on those features most relevant to the current discussion. The molecule is read in (either from a file generated by a molecular modeling program or using the SMILES¹³ notation) and is first analyzed to determine the presence of conformationally relevant features such as rings, bond orders, stereocenters, and aromaticity. The **conformational units** present in the molecule are then determined. A unit is a group of connected atoms about which the system has conformational 'knowledge'. This knowledge includes the atom types, connectivity, bond orders, and stereochemistry of a unit together with information about the conformations it can adopt. To construct a conformation of the molecule, a conformation is first assigned to each unit

[†] A preliminary report of this work was presented at the 202nd meeting of the American Chemical Society (1991).

^{*} To whom correspondence should be addressed.

[‡] University of Southampton.

[§] BioCAD Corp.

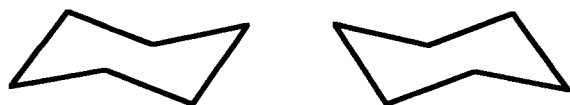


Figure 1. Two unique orientations of the chair conformation of cyclohexane.

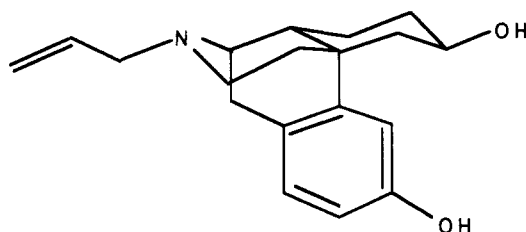


Figure 2. *N*-Allyl-3,6 β -dihydroxymorphinan. This molecule contains a piperidine ring which initially is not matched to a fragment in the knowledge base. However, after generalization this ring is matched to the cyclohexane unit.

present. The unit conformations are then joined together in a stepwise fashion. An implicit assumption in such an approach is that the units behave in a conformationally independent fashion, exhibiting the same behavior in a large molecule as in a small one. Each unit conformation is stored as a separate template file in a library. Typically, the templates of a unit correspond to the minimum-energy conformations the unit can adopt. It may also be necessary to consider more than one orientation of each template; for example, chair cyclohexane has two unique orientations (Figure 1). We define a **subconformation** of a unit as a template in a specific orientation. Different combinations of unit subconformations give rise to different conformations of the molecule, the conformational space searched by the program thus corresponding to all possible combinations of unit subconformations. A number of algorithms can be used to perform this search, including the depth-first and A* algorithms. Not all combinations of unit subconformations give rise to acceptable conformations of the molecule; for example, combinations of subconformations which have high-energy steric interactions or in which the subconformations do not fit together satisfactorily are discarded. A key component of all the algorithms used to perform the search is thus the use of checks which efficiently eliminate unsatisfactory structures.

The approach confers considerable practical advantages by virtue of its speed and the quality of the resulting conformations. In addition, the ability to perform the search automatically makes it possible to examine many molecules without human intervention. However, it is important to recognize some of its limitations, two of which we aim to address in this paper. The first arises when the knowledge base is insufficient to describe the molecule under investigation, i.e., not all the units required are present. When examining individual molecules, the missing unit(s) can be added by the user, but this is impractical when dealing with large numbers of diverse molecules (e.g., in a database) whose identities may not be known, let alone the units which are needed to specify them. A partial solution to this problem with which we have had some success is to use **generalized units**. Should the unit library not contain all the units required to completely define the molecule, successive levels of 'generalization' are employed to try and find a set of units which will describe the molecule. For example, our current system, which does not contain an explicit piperidine unit, will initially be unable to completely describe *N*-allyl-3,6 β -dihydroxymorphinan (Figure 2). At the first level of generalization, however, in which each unit atom is permitted to match molecule atoms of the same

geometrical class (e.g., an sp³ carbon atom is permitted to match any saturated tetrahedral atom in the molecule) the cyclohexane unit would be matched to the piperidine ring. Two subsequent levels of generalization can be employed, if necessary.¹² Prior to the conformational search, the conformations of each generalized unit are **adjusted** so as to give structures closer to the true minimum-energy conformations of the unit as required in the molecule. For example, the bond lengths to the nitrogen in the piperidine ring would be shortened from their initial C–C values to make them closer to the correct C–N bond length. The use of generalized units enables a wide variety of molecules to be analyzed using a relatively small knowledge base. It does, however, require a degree of similarity between the unit and the molecule and is not applicable, for example, to the substitution of a ring for one of different size. In such cases the original algorithm would be unable to further analyze the molecule.

The second problem to be addressed here concerns the construction of highly strained ring systems. An algorithm called **join** is used to construct the molecular conformations from the unit subconformations. This algorithm takes two fragments of the molecule, each containing one or more unit subconformations, and merges them to give a single, larger fragment. To this is added another fragment conformation, the process continuing until a complete structure of the molecule is generated. The join algorithm treats the fragments as rigid bodies, positioning them in space such that common atoms and bonds are optimally superimposed, and then a union structure is derived from the atomic coordinates in the two fragments. An extensive investigation was performed to determine the best algorithms for this task,¹⁴ but there are certain types of molecules (particularly those containing bridged and fused small rings) where the use of rigid templates gives structures that are lower in quality than those obtained for other systems. Indeed, in some cases no conformations are generated because all combinations of unit subconformations do not fit together satisfactorily.

There are a number of ways in which one might address these two issues; the method we report here is to employ an alternative conformational search method for the troublesome pieces of the molecule. An alternative would be to examine a database of experimentally determined conformations (such as the Cambridge Structural Database¹⁵) to identify molecules containing the appropriate fragments. Such an approach is appealing because 'real' experimental data is employed, but suffers from the disadvantage that the appropriate fragments might not be present in sufficient numbers to ensure that the whole conformational space is represented. Indeed, in some cases no examples of the fragment may be present. Currently, methods for searching the conformational space of 'small' molecules other than the model-building approach used in COBRA fall into one of the following classes: systematic search, random search, distance geometry, and molecular dynamics.² In principle, any of these techniques could be used, but we feel that distance geometry has certain advantages over the alternative approaches for the present application. It is relatively rapid in comparison, for example, to random methods^{16,17} or to molecular dynamics. It does not inherently rely on molecular mechanics energy minimization to generate 'reasonable' structures (and so could be used in cases where the required force-field parameters are not available). It is able to deal equally well with both cyclic and acyclic molecules. Distance geometry is especially suited to the generation of strained ring systems due to the large number of distance

constraints which can be derived. In this respect it appears superior to systematic search methods^{18,19} which can suffer from ring closure problems in such cases. The strategy we adopt in the combined algorithm (DG-COBRA) is to use the model-building approach embodied in the original COBRA program until one of the two problems described above is recognized. Appropriate fragments are then formulated, their conformational space is searched using distance geometry, the results are analyzed, and the resulting conformations are then combined with templates from the library in exploring the conformational space of the whole molecule.

The remainder of this paper is structured as follows. First, we describe how the two types of problem are identified and how the fragments to be analyzed using distance geometry are formulated. We then discuss our implementation of the distance geometry algorithm and how the conformations it generates are processed for subsequent use. Finally, we illustrate the operation of the combined approach in a number of examples.

UNDEFINED PARTS OF THE MOLECULE

A substructure searching algorithm is used to determine the units present in the molecule.¹² For all occurrences of each unit in the molecule this algorithm determines a complete set of stereochemically matching atom pairs (i.e., the exact unit-to-molecule correspondence at the atom level, taking full account of the stereochemistry). The molecule is completely defined when all its atoms and bonds have been matched to at least one atom and bond in at least one unit. If, when all the units in the knowledge base have been considered and all levels of generalization have been explored, there remain any atoms or bonds in the molecule which have not been matched to an atom or bond from at least one unit, then the molecule is incompletely defined. It is thus necessary to devise an algorithm which can formulate appropriate fragments for examination using distance geometry. Each fragment so analyzed is treated as a unit (referred to hereafter as a 'DG unit') in the subsequent conformational search, the distance geometry conformations being used in place of prestored templates. Consequently, each DG unit must be defined so that it has atoms in common with units already found in the molecule. The identification and definition of the DG units are performed in two stages. First, undefined rings are identified and analyzed, and then any undefined acyclic parts of the molecule.

The smallest set of smallest rings (SSSR) is used to define the ring system of the molecule. One advantage of using the SSSR is that should there remain one or more rings in the molecule which have not been matched to a ring of the same size from at least one unit, then the molecule's ring system is inadequately determined. Each unmatched ring is defined as a DG unit, formulated to contain the atoms in the ring and all the atoms in the first connectivity shells of the ring atoms. Consider cyclooctanone 4-phenylsemicarbazone (Figure 3), which contains an eight-membered ring, and suppose that no eight-membered ring is present in the knowledge base. The corresponding DG unit contains the eight atoms in the ring together with the atoms immediately adjoining these as indicated in Figure 3. Note that it is not necessary for the fragment to be 'chemically sensible'; simply that the unit has sufficient atoms in common with adjoining units to enable the DG unit to be joined to the remainder of the molecule in the subsequent search of conformational space. Having dealt with any undefined rings, should there remain any atoms or bonds which have not been matched to either a unit from the

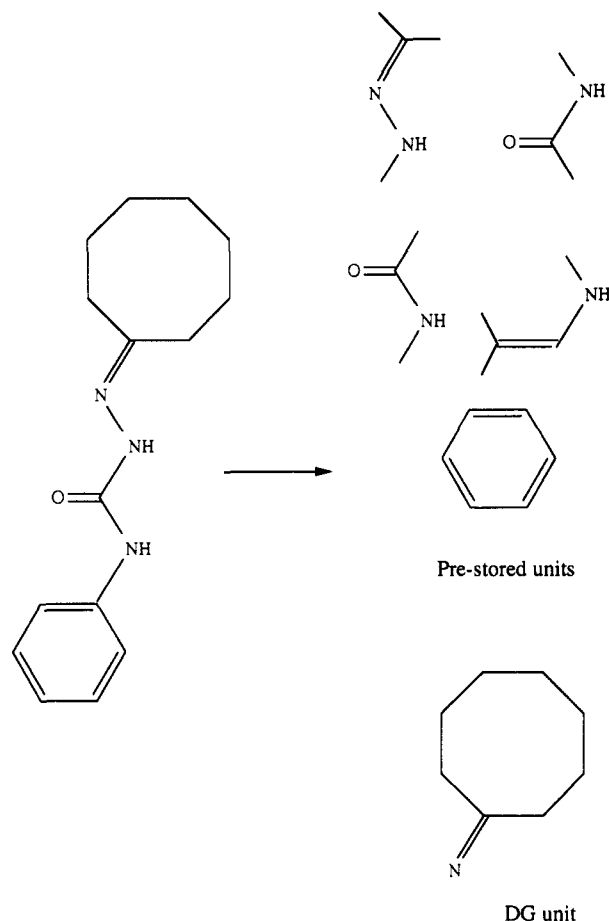


Figure 3. Cyclooctanone 4-phenylsemicarbazone, showing the units recognized within it from the knowledge base together with the fragment selected for distance geometry analysis.

knowledge base or a 'cyclic' DG unit, it is necessary to define appropriate 'acyclic' DG units. An acyclic DG unit is formulated by taking one of the undefined atoms or bonds and then expanding successive shells until the unit overlaps with previously defined units. To illustrate this, consider the cumulene unit present in the knowledge base. The unit would be formulated as indicated in Figure 4. Note that again the unit is defined so as to give appropriate overlap with the other units already found. More acyclic DG units are then specified, if necessary, until the entire molecule is defined.

STRAINED RING SYSTEMS

For many molecules the approach embodied in the WIZARD-II and COBRA programs provides a way to rapidly and effectively search the conformational space. However, in some cases all possible combinations of unit subconformations lead to problems. For some molecules, all combinations may give rise to conformations which contain high-energy steric interactions. The mechanism by which such strains can be resolved has been described.¹⁰ A second type of problem arises when all combinations of subconformations fail to fit together satisfactorily. The join algorithm returns numerical values which indicate the quality of fit for the two fragments joined together. If these values are too large (i.e., the fit is poor), that particular combination is rejected. In our experience such fitting problems only occur when trying to join two fragments which have one or more cyclic bonds in common (i.e., when forming a fused or bridged ring system). However, fitting problems never occur when joining two acyclic units,

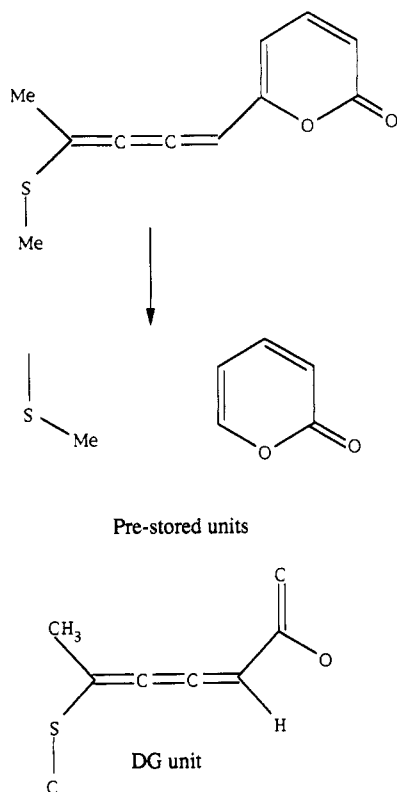


Figure 4. 4-(Methylthio)-1-(2-oxo-2H-pyranyl)penta-1,2,3-triene, found in *compositae*. The cumulene unit which is defined for distance geometry analysis is indicated, together with the units which are used to construct the remainder of the molecule.

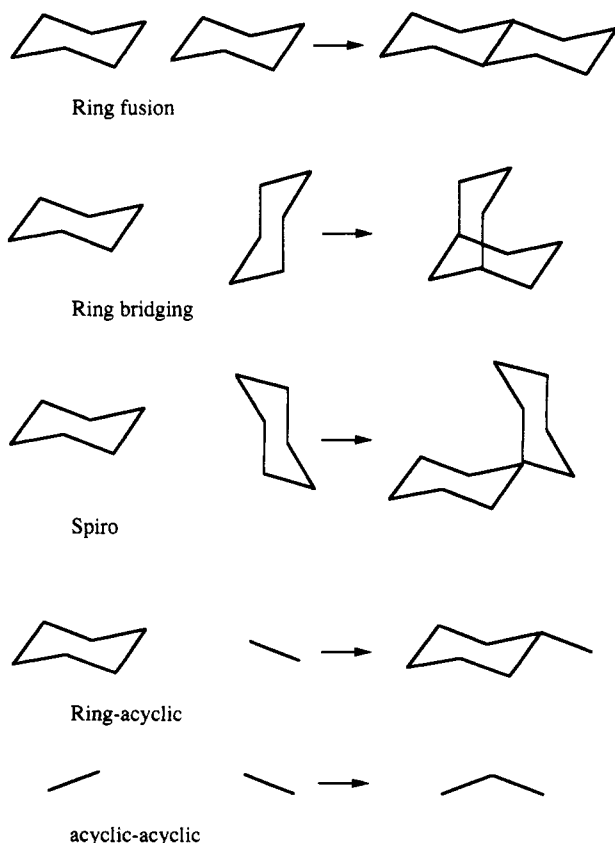


Figure 5. Five kinds of join between units.

when joining a cyclic unit to an acyclic unit, or when forming a spiro join between two rings (Figure 5). Strained ring systems contain a high proportion of ring fusion and bridging joins, and in such cases all combinations of unit subconfor-

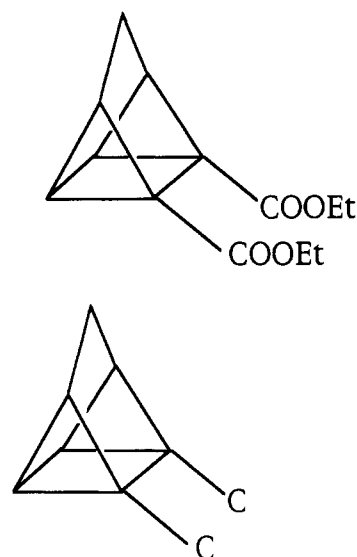


Figure 6. Diethyl tetracyclo[2.2.1.0^{3,5}.0^{2,6}]heptane-2,3-dicarboxylate. The SSSR contains two cyclopropane rings, one cyclobutane ring, and a cyclopentane ring. No conformations of the ring system were constructed by joining together the appropriate unit subconformations using the default fit criteria. The ring system formulated for distance geometry analysis is as indicated.

mations may fail to fit together satisfactorily. Consequently, no conformations would be generated for a molecule containing such a ring system. One solution to this problem would be to relax the criteria which are used to distinguish an acceptable join from an unacceptable one, but this does not address the issue of the quality of the structures which would then be lower than for unstrained systems. For example, the original version of COBRA (using the default fit criteria) failed to construct any conformations for diethyl tetracyclo[2.2.1.0^{3,5}.0^{2,6}]heptane-2,3-dicarboxylate (Figure 6).

Detecting the occurrence of fitting problems is expedited by the use of **conformational entities**. A conformational entity contains one or more units such that no unit is present in more than one entity. Each acyclic unit is defined as a single entity, as is an isolated cyclic unit not joined by ring fusion or bridging joins to another unit. Thus, each of these entities contains a single unit. In contrast, a group of cyclic units which are all interconnected by fused or bridged ring bonds is also defined as a single entity. As an illustration, the entities present in *N*-allyl-3,6 β -dihydroxymorphinan are shown in Figure 7. Conformational entities were originally introduced to deal with the resolution of strain in conformations.¹⁰ They also form the basis of an entity-based searching algorithm¹¹ that is employed in the combined searching method reported here. In this search algorithm, the conformations available to each individual entity are first determined. The conformations permitted to those entities which contain a single unit are identical to the unit subconformations. For entities which contain more than one unit, however, the conformational space of the entity alone is searched by examining all combinations of subconformations of the entity's constituent units using a depth-first search. The resulting set of entity conformations then forms the basis for searching the conformational space of the entire molecule. So for *N*-allyl-3,6 β -dihydroxymorphinan, the conformational space of the ring system would be searched and the resulting conformations then combined with those of the remaining acyclic entities to generate conformations of the entire molecule. The technique is a form of 'divide and conquer' strategy and can give significant reductions in search time. A consequence of the way in which entities are defined is that no two entities will have a cyclic

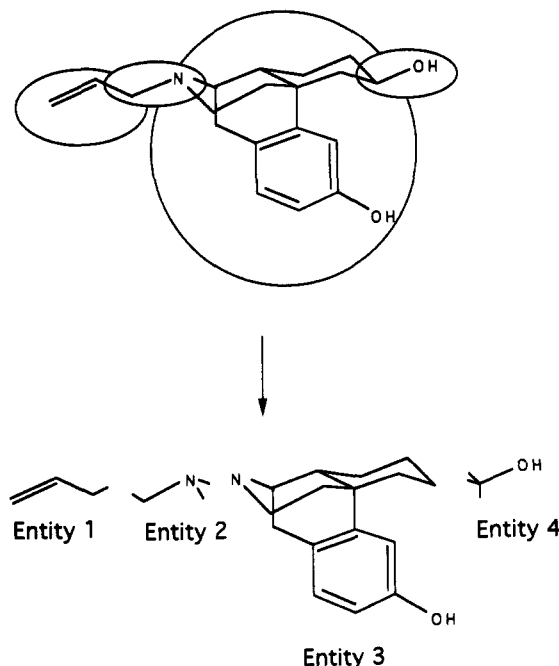


Figure 7. Entities present in *N*-allyl-3,6-dihydroxymorphinan. Rings joined by fusion or bridging joins are defined as a single entity.

bond in common. Thus the joining of entities is never subject to fitting problems; it is only when searching the conformational space of an individual cyclic entity that fitting problems may occur. Should it prove impossible to find any conformations for a cyclic entity because all combinations of its unit subconformations fail to fit satisfactorily, then this is taken as indicative of a strained ring system. In the new combined approach the problematic entity is then analyzed using distance geometry, and the resulting entity conformations are subsequently used in the entity-based conformational search of the entire molecule. In diethyl tetracyclo[2.2.1.0^{3,5}.0^{2,6}]heptane-2,3-dicarboxylate, the tetracyclic ring system comprises a single entity. The conformational space of this ring system is searched using distance geometry, the results are analyzed, and then combined with the other units when examining the entire molecule.

IMPLEMENTATION OF THE DISTANCE GEOMETRY ALGORITHM

Distance geometry is a method for searching conformational space in which a structure is initially formulated in terms of interatomic distances. The mathematical foundations of distance geometry were laid some time ago, but its application to conformational analysis was pioneered more recently, primarily by Crippen, Havel, and Kuntz. Distance geometry is based upon the fact that restrictions can be derived on the values a given interatomic distance can adopt. Some of these restrictions may come from experiment (e.g., 2D NMR experiments); others are derived from chemical principles (e.g., bond lengths). In this paper we are solely concerned with the use of distance geometry to generate structures using chemical information only (i.e., no externally defined distance constraints). Rigorous and detailed descriptions of the different stages involved in the generation of structures by distance geometry have been given elsewhere,²⁰⁻²⁴ so in this paper we only indicate the method we have selected where more than one choice is available. One point to note which distinguishes our use of distance geometry from some other implementations is that while we do employ molecular mechanics minimization to finally refine the distance geometry structures, where this

is not possible (e.g., because the necessary parameters are not available) we desired the 'raw' distance geometry structures to be of sufficient quality that they can be used without further refinement in the subsequent conformational search. Our efforts have thus been directed toward determining the optimal combination of algorithms which gives an adequate coverage of the conformational space and high-quality structures.

In the absence of any user-supplied distance constraints, the initial upper and lower bounds on each interatomic distance in the fragment are assigned in the usual fashion; bounds between atoms in 1-2, 1-3, and 1-4 relationships are allocated using standard bond lengths and angles from an extensive list derived from the COSMIC force-field,²⁵ while the lower and upper bounds on the interatomic distances for atoms in a 1-*n* relationship (*n* > 4) are set to the sum of the *vdw* radii and a large value, respectively. The COSMIC force field is also used within COBRA to calculate a molecular mechanics energy of each conformation as it is constructed. The resulting matrix is then refined using triangle smoothing. We use an all atom (rather than united atom) representation of the fragment. For the typical size of fragment to be analyzed (rarely containing more than 15-20 heavy atoms), the use of an all atom model gives final structures of higher quality without imposing an excessive computational burden.

We use the correlation method of Crippen²³ to assign distances in the initial distance matrix; this also gives a greater probability of a successful embedding. Having derived the metric matrix, we typically embed into four dimensions; in common with others we find that higher quality structures result when embedding and optimizing in a space of higher dimensions than three. A conjugate gradient algorithm is used to refine the initial structure, using the following error function containing terms from both the distance bounds and the chiral terms:

$$E = \sum_{i=0}^{N-1} \sum_{j>i}^N \begin{cases} (l_{ij}^2 - d_{ij}^2)^2 & \text{if } d_{ij} < l_{ij} \\ 0 & \text{if } l_{ij} \leq d_{ij} \leq u_{ij} \\ (d_{ij}^2 - u_{ij}^2)^2 & \text{if } u_{ij} < d_{ij} \end{cases} + \sum_{k=0}^{N_{\text{chiral}}} \begin{cases} (V_{\text{lower},k} - V_k)^2 & \text{if } V_k < V_{\text{lower},k} \\ 0 & \text{if } V_{\text{lower},k} \leq V_k \leq V_{\text{upper},k} \\ (V_k - V_{\text{upper},k})^2 & \text{if } V_{\text{upper},k} < V_k \end{cases}$$

where d_{ij} is the distance between atoms *i* and *j*; u_{ij} and l_{ij} refer to the upper and lower bounds on that distance; V_k is the chiral volume for a given chiral constraint *k* with V_{lower} and V_{upper} defining the lower and upper permitted values. We have also experimented with some of the alternative error functions but find that their behavior is very similar. It is particularly important that the stereochemistry at each atom is correct in the final conformation, as atoms will be overlaid during the joining of units. Prior to the refinement, the total number of chiral constraints which are violated in the initial structure is determined; if more than half are incorrect then the molecule is inverted.²⁶ In addition, for the first few steps of optimization a higher weight is given to the chiral constraint term, enabling the initial conformation to adopt the correct chirality at each such atom. This is facilitated by the use of a higher dimensional space. When convergence is achieved in four dimensions, a few steps of optimization are then performed in three dimensions. This is done because distances calculated in four dimensions will always be larger than the corresponding three-dimensional distance (assuming that the fourth coordinate is non-zero). As a consequence, upper bounds tend to be satisfied more stringently than lower bounds

when the penalty function is computed in more than three dimensions. The final three-dimensional optimization corrects this.

EXAMPLES

In this section we illustrate the application of the combined approach to the three examples described above. In each case, the distance geometry algorithms were used to generate conformations for the appropriate fragments. These were then minimized using the *cosmic* force field. For cyclic systems, a total of 200 structures were generated; for the acyclic fragments, which tend to contain at most two rotatable bonds, a smaller number (25) is used. This smaller value is also employed for the strained ring systems due to the more limited conformational space of such fragments. It is then necessary to decide which of these conformations should be used for the DG unit during the conformational search of the entire molecule. Distance geometry is a random method, though one which preferentially samples the regions of distance space that are more likely to correspond to reasonable three-dimensional structures. The same structure may thus be generated more than once (or, more commonly, two or more structures will minimize to give the same minimum energy conformation). Some means of identifying and removing duplicate structures is therefore required. Typically, the number of distinct minima for acyclic fragments and for strained ring entities is small, and it is desirable to identify and use all energetically accessible conformational minima. However, for cyclic systems the number of minima on the conformational energy hypersurface may be extremely large, especially where there is no or only a limited amount of symmetry in the fragment. Many of these minima correspond to very similar conformations. Under such circumstances it is often preferable to select a set of representative minima which together provide an appropriate indication of the range of accessible conformations. The fragment conformations generated by distance geometry are thus subjected to a clustering procedure; currently we employ the unweighted pair-group method using arithmetic averages (UPGMA).²⁷ The difference between two fragment conformations is calculated as the root mean square (rms) difference of the Cartesian coordinates of all non-hydrogen atoms. In the UPGMA method, the distance between two clusters is equal to the average of the distances of all pairwise combinations of members in the two clusters. The symmetry in the fragment is taken into account during the clustering process (e.g., there are two ways in which the ring in Figure 3 can be superimposed on itself). In the examples described below the conformations were clustered until the minimum distance between any pair of clusters was greater than 0.25 Å. Having analyzed the fragment, it can then be added to the knowledge base, thereby providing the means by which the system can automatically acquire conformational information for use in subsequent analyses.

After generation and clustering, the distance geometry conformations were then combined with the appropriate templates from *COBRA*'s fragment library in searching the conformational space of the entire molecule. A number of parameters determine how this search is performed. In order to enhance search efficiency, a unit is initially not permitted to adopt higher energy subconformations. Thus subconformations whose energy relative to the lowest energy subconformation is greater than some threshold (e.g., 4 kcal/mol) are initially ignored. Only if it proves impossible to construct any acceptable conformations using the lowest energy sub-

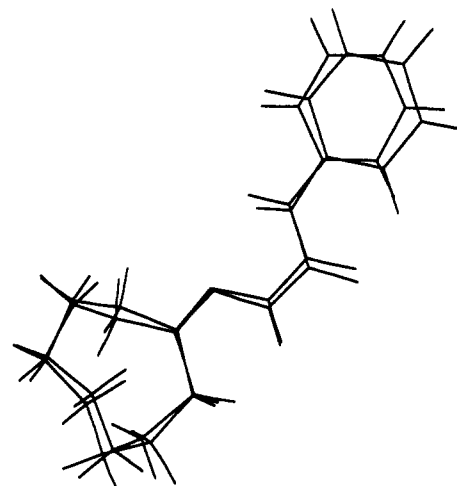


Figure 8. Superposition of the crystal structure of cyclooctanone 4-phenylsemicarbazone and one of the conformations generated by *COBRA*.

conformations are these higher energy structures employed. For example, the boat conformation of the cyclohexane unit would only be considered if it proved impossible to construct acceptable conformations using the chair and twist-boat structures. Conformations must also satisfy a close-contact ratio criterion: if one or more pairs of atoms in a 1,4 relationship or greater approaches to within a specified percentage of the sum of the van der Waals radii of the atoms, then the conformation (or partially constructed conformation) is rejected. The default value of this close-contact ratio (67%) was employed in the examples described below. Note that the entire procedure, including formulation of the distance geometry units, generation of the conformations, clustering, and the subsequent conformational search of the molecule, is performed automatically by the program, without any user intervention.

For cyclooctanone 4-phenylsemicarbazone a total of 24 minimum-energy conformations were obtained for the eight-membered ring, after clustering the initial 200. Five of these were rejected due to their high energy relative to the lowest energy structure. Four conformations are available to the remainder of the molecule, and so a total of 76 conformations for the entire molecule were consequently obtained. One of these structures is shown overlaid on that determined by single crystal X-ray crystallography²⁸ in Figure 8; the rms fit between the two structures (non-hydrogen atoms only) is 0.27 Å.

Five conformations were obtained for the cumulene system in 4-(methylthio)-1-(2-oxo-2*H*-pyran-5-yl)penta-1,2,3-triene. Four of these correspond to the combinations of syn and anti orientations of the *S*-methyl group and of the pyranone ring relative to the cumulene system. The fifth was of very high energy and was rejected by the energy critic. *COBRA* automatically determines the chiral constraints needed to keep odd-numbered cumulenes planar and also ensures that the stereochemical relationships between the substituents in even-numbered cumulenes are maintained. Distance geometry structures in which these relationships are not satisfied are rejected. The use of a higher weight for the chiral term during optimization helps to reduce the number of structures which are rejected for this reason. An improper torsion term was incorporated into the *cosmic* force field to maintain planarity during the energy minimization for odd-numbered cumulenes. There is no conformational variability in the rest of the molecule, so four conformations were obtained in total.

A single conformation was obtained for the ring system in diethyl tetracyclo[2.2.1.0^{3,5}.0^{2,6}]heptane-2,3-dicarboxylate.

When combined with the units for the remainder of the molecule a total of 47 conformations were produced. No crystal structure is available for this molecule. However, each structure was minimized using the COSMIC force field and compared with the initially generated conformation. The rms fits between the initial and final structures ranged from 0.06 to 0.18 Å with an average of 0.10 Å.

CONCLUSIONS

Knowledge-based methods can only deal with problems for which the required information is available. By incorporating an alternative searching algorithm into our approach to conformational analysis and search we have demonstrated one way in which this problem can be addressed. The joint algorithm marries the speed of the fragment-based method with the ability of the distance geometry approach to perform ab initio conformational search. By combining these two methods for performing conformational search with algorithms for perceiving a molecule and a suitable control strategy it is possible to develop a strategy for performing automatic conformational search which uses each searching algorithm for the task to which it is most suited.

ACKNOWLEDGMENT

A.R.L. thanks the SERC and NATO for provision of a postdoctoral fellowship and the SERC for current financial support under the Advanced Fellowship scheme. A significant portion of this work was carried out at the Computer Graphics Laboratory, University of California, San Francisco. A.R.L. would also like to express gratitude to Robert Langridge and the other members of the laboratory for their kind hospitality during his stay. The Computer Graphics Laboratory is supported by the National Institutes of Health through Grant R1081 (R. Langridge PI). The original WIZARD concept was devised by Dr. D. P. Dolata and subsequently developed with Dr. C. K. Prout in the Chemical Crystallography Laboratory, University of Oxford (U.K.).

REFERENCES AND NOTES

- (1) Howard, A. E.; Kollman, P. A. An Analysis of Current Methodologies for Conformational Search of Complex Molecules. *J. Med. Chem.* **1988**, *31*, 1669–1675.
- (2) Leach, A. R. A Survey of Methods for Searching the Conformational Space of Small and Medium-Sized Molecules. In *Reviews in Computational Chemistry, Vol. II*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1991; pp 1–55.
- (3) Lautz, J.; Kessler, J.; Blaney, J. M.; Scheek, R. M.; van Gunsteren, W. F. Calculating Three-Dimensional Molecular Structure from Atom-Atom Distance Information: Cyclosporin A. *Int. J. Pept. Protein Res.* **1989**, *33*, 281–288.
- (4) Koning, T. M.; Davies, R. J.; Kaptein, R. The Solution Structure of the Intramolecular Photoproduct of d(TpA) Derived with the Use of NMR and a Combination of Distance Geometry and Molecular Dynamics. *Nucleic Acids Res.* **1990**, *18* (2), 277–284.
- (5) Leach, A. R.; Prout, K. Automated Conformational Analysis: Directed Conformational Search using the A* Algorithm. *J. Comput. Chem.* **1990**, *11*, 1193–1205.
- (6) Ruskino, A., III; Skell, J. M.; Balducci, R.; McGarity, C. M.; Pearlman, R. S. CONCORD, A Program for the Rapid Generation of High-Quality Approximate 3-Dimensional Molecular Structures. The University of Texas at Austin and Tripos Associates: St. Louis, MO, 1988.
- (7) DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, J. Using Shape Complementarity as an Initial Screen in Designing Ligands for a Receptor Binding Site of Known Three-Dimensional Structure. *J. Med. Chem.* **1988**, *31*, 722–729.
- (8) Shoichet, B. K.; Leach, A. R.; Kuntz, I. D. Desolvation and Conformational Entropy in Inhibitor Design. In preparation.
- (9) Dolata, D. P.; Leach, A. R.; Prout, K. WIZARD: AI in Conformational Analysis. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 73–85.
- (10) Leach, A. R.; Prout, K.; Dolata, D. P. The Application of Artificial Intelligence to the Conformational Analysis of Strained Molecules. *J. Comput. Chem.* **1990**, *11*, 680–693.
- (11) Leach, A. R.; Prout, K.; Dolata, D. P. Automated Conformational Analysis: Algorithms for the Efficient Construction of Low-Energy Conformations. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 271–283.
- (12) Leach, A. R.; Dolata, D. P.; Prout, K. Automated Conformational Analysis and Structure Generation: Algorithms for Molecular Perception. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 316–324.
- (13) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.
- (14) Leach, A. R.; Prout, K.; Dolata, D. P. An Investigation into the Construction of Molecular Models Using the Template Joining Method. *J. Comput.-Aided Mol. Des.* **1988**, *2*, 107–123.
- (15) Allen, F. A.; Bellard, S. A.; Brice, M. D.; Cartwright, B. A.; Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B. G.; Kennard, O.; Motherwell, W. D. S.; Rodgers, J. R.; Watson, D. G. The Cambridge Crystallographic Data Centre: Computer-Based Search, Retrieval, Analysis and Display of Information. *Acta Crystallogr.* **1979**, *B35*, 2331–2339.
- (16) Saunders, M. Stochastic Exploration of Molecular Mechanics Energy Surfaces. Hunting for the Global Minimum. *J. Am. Chem. Soc.* **1987**, *109*, 3150–3152.
- (17) Ferguson, D. M.; Raber, D. J. A New Approach to Probing Conformational Space with Molecular Mechanics: Random Incremental Pulse Search. *J. Am. Chem. Soc.* **1989**, *111*, 4371–4378.
- (18) Lipton, M.; Still, W. C. The Multiple Minimum Problem in Molecular Modeling. Tree Searching Internal Coordinate Conformational Space. *J. Comput. Chem.* **1988**, *9* (4), 343–355.
- (19) 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.
- (20) Crippen, G. M. Rapid Calculation of Coordinates from Distance Matrices. *J. Comput. Phys.* **1978**, *26*, 449–452.
- (21) Crippen, G. M.; Havel, T. F. Stable Calculation of Coordinates from Distance Information. *Acta Crystallogr.* **1978**, *A34*, 282–284.
- (22) Havel, T. F.; Kuntz, I. D.; Crippen, G. M. Theory and Practice of Distance Geometry. *Bull. Math. Biol.* **1983**, *45*, 665–720.
- (23) Crippen, G. M. *Distance Geometry and Conformational Calculations*; Chemometrics Research Studies Series 1; John Wiley: New York, 1981.
- (24) Crippen, G. M.; Havel, T. F. *Distance Geometry and Molecular Conformation*; Chemometrics Research Studies Series 15; John Wiley: New York, 1988.
- (25) Vinter, J. G.; Davis, A.; Saunders, M. R. COSMIC: A General Purpose Molecular Mechanics Force-Field. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 31–51.
- (26) Blaney, J. M.; Crippen, G. M.; Dearing, A.; Dixon, J. S. DGEOM: QCPE program 590.
- (27) Romesburg, H. C. *Cluster Analysis for Researchers*; Lifetime Learning Publications: Belmont, CA, 1984.
- (28) Groth, P. On the Crystal Conformation of Two Eight-Membered Rings: Cyclooctanoxime and Cyclooctanephénylsemicarbazone (at –150 °C). *Acta Chem. Scand., Ser. A* **1981**, 117–121.