RAMBLE: A conformational search program

T. D. J. Perkins and D. J. Barlow

Department of Pharmacy, King's College London, London, UK

A program, RAMBLE, is described which searches the conformational space of looped or cyclic molecules by random assignment of internal coordinates. Conformations produced are screened for energetically unfavorable intramolecular contacts and are subject to user-supplied constraints in terms of both bonded and nonbonded distances, bond angles, and torsion angles. The methodology employed is discussed in relation to alternative search strategies, and a description is given of its successful application in modeling the structures of cyclohexane, reverse-turn tetrapeptides, and a bacterial siderophore complex.

Keywords: conformational search, molecular modeling, peptide structure, structure prediction

INTRODUCTION

Conformational search techniques are essential to many aspects of molecular modeling. They are widely used in the study of structure–activity relationships of drug molecules,[1] in the correlation of molecular structure and spectroscopic data,[2] and in the prediction of loop conformations in proteins modeled by homology.[3] In most cases the procedures are designed to sample the conformational space available to a molecule systematically, by performing iterative rotations about all single bonds. This type of approach is satisfactory when dealing with small and/or conformationally restricted molecules, but rapidly becomes intractable as the size and flexibility of the structure increases. Moreover, in the analysis of molecules in which there are significant variations in bond lengths or bond angles (e.g., metal–ion complexes), or those involving cyclic systems (e.g., disulfide-linked peptides) this method is entirely unsuitable. In these situations, a simple solution is to undertake random sampling of the possible conformations, and then to eliminate those structures in which the conformational constraints are not satisfied.

In this report, we describe a suite of programs (RAMBLE, MAKRAM, NFIT, and GLUE) that are designed specifically for this purpose. The methods are discussed in relation to other conformational search strategies[4–7] and their reliability is tested by modeling structures that have well-characterized conformational families: cyclohexane, a tetrapeptide hydrogen-bonded reverse turn, and a bacterial siderophore complex.

Address reprint requests to Dr. Barlow at the Department of Pharmacy, King's College London, Manresa Road, London SW3 6LX, UK.
Received 6 February 1990; accepted 13 March 1990

METHODS

For any noncyclic or monocyclic system the conformational search can be broken down into four steps:

(1) The definition of the molecular structure and setting of geometrical constraints.
(2) The generation of a set of conformers that satisfy these constraints.
(3) The removal of redundant multiple structures, and identification of a representative subset of conformers.
(4) The calculation of the potential energies of the representative conformers, as a measure of their relative stabilities.

The program RAMBLE (Random Automated Method of Building Loops Etc.) carries out stages 2 and 3 above, after reading a text file that defines the molecular structure and constraints. For polycyclic systems, the above procedures are repeated to produce two or more overlapping fragments of the molecule, and these are then bolted together to achieve the complete structure.

1. Definition of geometrical structure and constraints

To illustrate the style of input to RAMBLE, a tetraglycyl peptide reverse turn is used as an example structure. Figure 1 shows a schematic illustration of the structure, while the corresponding input file, together with some explanatory comments, is shown in Figure 2. The input file contains a number of keywords, followed by data relating to that keyword (e.g., ATOM specifies the number of atoms). Some keywords that define optional parameters may be omitted from the file, in which case sensible default values are assumed. The relevant keywords are noted as each topic is discussed, allowing reference to be made to the sample file in Figure 2.

The input file can be created directly using any text editor, or, more easily for a large or complex structure, by using the program MAKRAM, which takes the required confor-

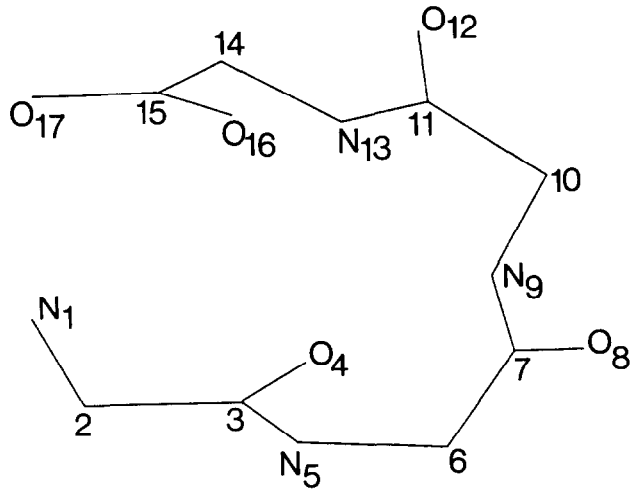
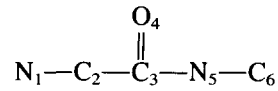


Figure 1. Schematic stick drawing of a tetraglycyl peptide, showing the atom numbering system used by RAMBLE. Atom 1 is the N-terminal amino group N atom; atom 17 is the C-terminal O atom. A reverse turn is forced between peptide carbonyl oxygen atom (4) and the peptide amide nitrogen (13)

mational data from a coordinate file created using the CHEMMOD molecular modeling system.[8] MAKRAM then gives user-friendly prompts for the additional geometrical parameters required, to create a complete input file for RAMBLE. In the case of cyclic systems, it is not necessary for the starting structure to satisfy the constraints for ring closure, since these are specified explicitly in the input to MAKRAM.

For simplicity and speed, only the basic framework of the molecule is ever considered, that is, the backbone atoms plus one-atom branches. In the example structure, therefore, this includes the peptide main chain N, C_α, C′, and O atoms. The connected atoms are numbered sequentially, except that at branch points, the branch atoms receive priority, for example, for a peptide fragment,



The structure is defined by a number of constraints, in terms of both bonded and nonbonded geometry. For a molecule of n atoms the bonded geometry is described by $n - 3$ torsion angles (TORR), $n - 2$ bond angles (ANGR), and $n - 1$ bond lengths (BNDL). In the input to RAMBLE the value of each of these parameters is specified by two numbers, representing the lower and upper bounds of the allowed geometry.

Additional distance (LOOP, DSTR) and optional bond (FANG) and torsion angle (FTOR) constraints may be specified, relating either to nonbonded interactions or to bonds formed by cyclization. In the modeling of the reverse turn,

```
ATOM      17 atoms in structure
17
CONF      100 conformations satisfying constraints to be produced
100
FITT      only use atoms 3-13 for the superpositions of structures required in cluster analysis
3 13
TORR      list of 14 torsion angle ranges
-180 180        atoms 1-2-3-4   ψ − 180
180 180         atoms 1-2-3-5   (ψ − (ψ − 180)) = 180
180 180         atoms 2-3-5-6   ω
-180 180        atoms 3-5-6-7   φ
-180 180
180 180
180 180
-180 180
-180 180
180 180
180 180
-180 180
-180 180
180 180
ANGR      list of 15 bond angle ranges
109.47 109.47   atoms 1-2-3
121 121         atoms 2-3-4
114 114         atoms 2-3-5
123 123         atoms 3-5-6
109.47 109.47
121 121
114 114
123 123
109.47 109.47
121 121
114 114
123 123
109.47 109.47
121 121
121 121
BNDL      list of 16 bond length ranges
1.47 1.47       atoms 1-2
1.53 1.53       atoms 2-3
1.24 1.24       atoms 3-4
1.32 1.32       atoms 3-5
1.47 1.47
1.53 1.53
1.24 1.24
1.32 1.32
1.47 1.47
1.53 1.53
1.24 1.24
1.32 1.32
1.47 1.47
1.53 1.53
1.24 1.24
1.24 1.24
LOOP      monitor distance between atoms 4 and 13
4 13
DSTR      allow a range of 2.7-3.5Å
2.7 3.5
FANG      allow any final angles
0 180 0 180
BRCH      4 branch atoms, attached to atoms 3, 7, 11, and 15
4
3 7 11 15
TYPE      list of 17 CHEMMOD atom types
6
1
2
11
6
1
2
11
6
1
2
11
6
1
2
11
10
QLEV      cutoff for adding new members to a cluster
0.05
QBMP      use 1.0 × Van der Waals radius as cutoff for non-bonded atomic contacts
1.0
```

Figure 2. Input file for tetraglycyl reverse-turn peptide. The text that is part of the input required for the program is printed in typewriter type, while explanations are in slanted type. Note that torsion angles which define the positions of main-chain atoms following branch points are specified as an increment to the previous torsion angle. For example, the torsion angle defined by atoms 1-2-3-5 is specified as an increment to the angle defined by atoms 1-2-3-4. Increments generally take values of 180° (trigonal atom geometry) or ±120° (tetrahedral atom geometry)

for example, a hydrogen bond is forced by constraining the distance between atoms 4 (the peptide carbonyl oxygen of residue 1) and 13 (the peptide nitrogen of residue 4) to be in the range 2.7–3.5 Å (see Figure 2).

In a cyclic system, the choice of the endpoints of the structure (i.e., those points to complete the loop) is often significant, since an additional two bond angles and three torsion angles are defined on loop closure. Although the program does allow constraints to be imposed on each of these bond and torsion angles, it is more sensible if the join is positioned at a flexible part of the molecule. For example, cyclic peptides containing glycine should have this residue (rather than a C_β-containing residue) at the join, since it has a wider range of allowable ϕ, ψ torsion angles.

2. Generation of conformers and selection of those satisfying the constraints

Each conformer is constructed atom-by-atom as specified by the bond length, bond angle, and torsion angle lists. If any particular parameter is fixed the defined value is used; otherwise, a random value from a uniform distribution between the lower and upper limits is chosen.

As each atom is added the distances between it and all other atoms are determined, excluding those separated by less than four bonds and those involved in monitored distances. If any one of the distances measured is less than the sum of the van der Waals radii of the atoms, the structure is rejected immediately, thereby providing early screening against the generation of high-energy conformers. Note that it is also possible to carry out the hard-sphere calculations using a particular fraction of the van der Waals radii, or a universal minimum separation figure, depending on the value entered for QBMP.

When a conformer that satisfies the above van der Waals criteria is completed, the distance (and optional bond angle and torsion angle) constraints are checked. If the structure does not conform to any of the above parameters, it is rejected; otherwise, its Cartesian coordinates are recorded and the next conformer is started.

3. Selection of representative conformations by cluster analysis

Once the specified number of successful conformations have been completed, it is necessary to remove any multiply occurring structures. This is done on the basis of the minimum root mean square (RMS) difference between the atomic coordinates of each pair of conformers. The RMS difference is a one-dimensional projection of the distance in conformational space between two structures, and thus represents a simple quantity to use in cluster analysis. RMS differences

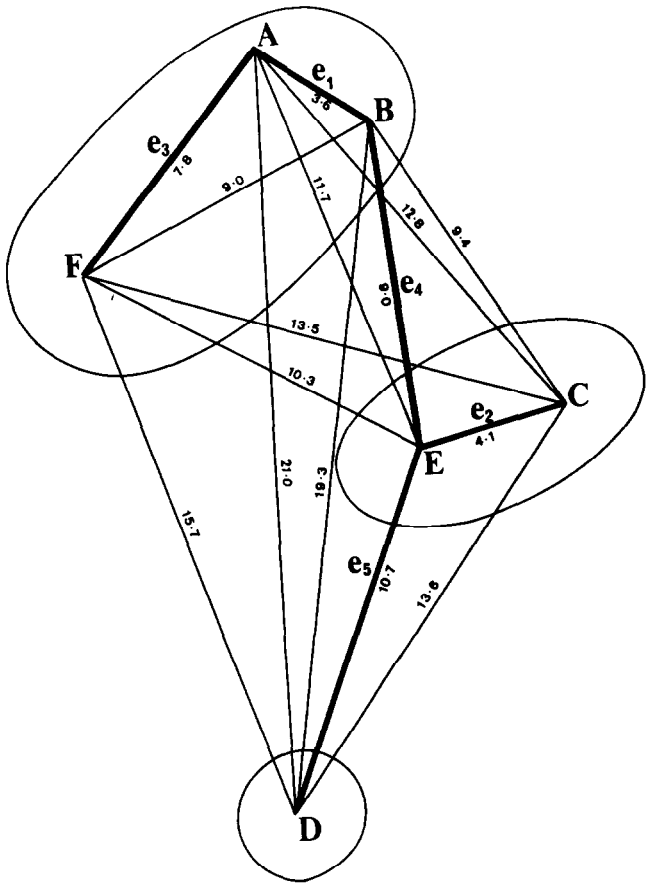


Figure 3. Construction of single-linkage map and clustering of data. The figure shows a two-dimensional representation of the RMS distances between n = 6 *conformers* A–F. *The single-linkage map for the conformers is constructed as an ordered list of 5 (i.e.,* n − 1*) pairs of conformers covering minimum total distance, such that all conformers are included and there are no cycles. The edges included in the single-linkage map are shown with heavy lines and are marked as* $e_1, \ldots, e_5$. *These connections are used to assign conformers to clusters, by repeatedly adding the next closest element connected to any existing member of the current cluster, until some condition is violated. A new member will be added to a cluster only if the average RMS distance between it and the existing members is less than some cutoff value (in the example here, 10 units).*

*The method of clustering is as follows: The conformers (*A, B*) at the vertices of the shortest edge* e_1 *become the nucleus of cluster 1. Next, the closest connected conformer (*F, *along* e_3*) is tested as a potential member of cluster 1. Since (*$|AF| + |BF|)/2 < 10$*, then* F *is added. The process is then repeated for the next closest connected conformer (*E, *along* e_4*), but since (*$|AE| + |BE| + |FE|)/3 > 10$*, then* E *is not added, and cluster 1 is completed with members* A, B, *and* F. *The next cluster (2) is started in the same way as the first, with the conformers (*E, C*) at the vertices of the shortest remaining edge (*e_2*) forming its nucleus. Next, the closest connected conformer (*D, *along* e_5*) is tested as a potential member of cluster 2, but because (*$|DE| + |CE|)/2 > 10$ *it is excluded. The clustering process is thus completed, with* E *and* C *in cluster 2, and* D *as a singlet.*

Note that this transformation of single-linkage map to clusters represents a modification of the more usual procedure (e.g., Ref. 15), where the final number of clusters is fixed by the user

are calculated by the method of McLachlan[9] using the routine MATFIT,[10] and are then used by the routine LINKER[11] to produce a single-linkage map. The single-linkage map is then partitioned to produce clusters of related conformations by the method detailed in Figure 3.

For each cluster, one conformation is chosen as a representative and its coordinates are output in the format of the CHEMMOD molecular modeling system.[8] The representative conformation is identified as the element that has the smallest total of RMS differences with the remaining members of the cluster.

This strategy for identifying the representative conformations from a set of structures may not always be successful, because of the loss of dimensionality incurred in using RMS differences as a measure of conformational similarity. In the examples considered below, however, it is encouraging to see that there is generally greater similarity of structures within clusters than between clusters. In any event, the performance of the clustering procedure can easily be monitored (by visual inspection) and any anomalies can be rectified by changing the value of QLEV.

To deal with polycyclic systems, the simplest procedure is to build the constituent monocyclic fragments separately using RAMBLE, and to then combine them. This is carried out by the program NFIT, which bolts together two fragments that have a number of atoms in common. NFIT takes the coordinates of each of the representative conformations of both fragments, and carries out matches of the common atoms between all possible pairs of conformers. If the RMS error in superimposing the matched atoms is less than a specified cutoff, the coordinates and connectivities of the two fragments are merged to produce one complete structure. This method has been used successfully to model the siderophore iron complex ferrichrome.

An additional program, GLUE, is designed specifically for peptides, and bolts the amino acid side chains onto a RAMBLE-built main chain (so long as C_β atoms are present for all nonglycine residues). The side-chain conformations used in the program are the 2–4 most abundant conformations (comprising at least 10% of the total) identified from a survey of known protein crystal structures in the Brookhaven Protein Databank,[12] (Perkins and Barlow, unpublished data). For each main chain all the possible side-chain rotamers are built, producing a very large number of conformers. Once built, the potential energy is calculated for each conformation and energy minimization may then be undertaken.

All programs are written in standard FORTRAN 77, and have been implemented on a Vax 8800 or 8700 operating under VMS at King's College London Computer Centre. RAMBLE has also been implemented on an Amdahl 5890/300 operating under Phoenix/MVS at the University of London computer center, and a U-Man 1000 microcomputer operating under MIRAGE. Performance times, where given, are for the VAX/VMS system.

RESULTS

1. Cyclohexane: a simple small cyclic system

Cyclohexane was constructed by RAMBLE such that the given bond lengths were 1.54 Å, bond angles were tetrahedral (109.47°), and torsion angles were variable over the range −180° to +180°. Ring closure was brought about by constraining the final bond length to be 1.54 ± 0.04 Å, and final bond angles to be 109.47 ± 2.0°. In a typical run, the program required of the order of 10^6 iterations to produce 100 conformations to these specifications. On a Vax 8700 this took around 900 s of CPU time.

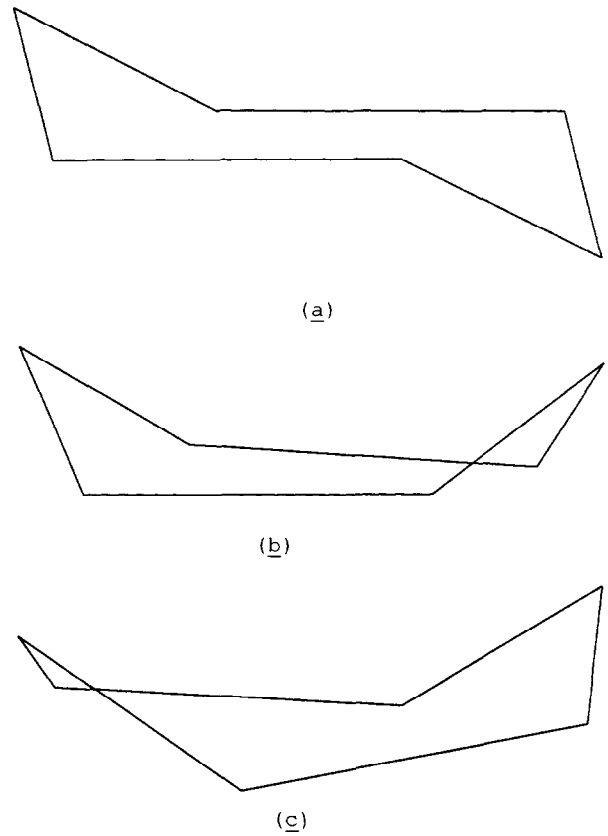


Figure 4. Selected representative conformations of cyclohexane generated by RAMBLE: (a) chair conformation, (b) boat conformation, (c) twist-boat conformation

The clustering algorithm separated the conformers into the recognized forms of cyclohexane: boat, chair, and twist-boat (see Figure 4). One complete cluster with all members superimposed is shown in Color Plate 1. A total of 6 clusters were produced, which, because of the inclusion of singlet structures, led to 8 representatives. The discrepancy between the number of representatives and the number of conformational forms is due in part to the degeneracy of conformations arising from the rotational symmetry of cyclohexane. To produce representatives with the half-chair conformation, the constraints for ring closure were relaxed to allow bond angles of 109.47 ± 20°.

2. Reverse-turn tetrapeptide: a hydrogen-bonded loop system

The reverse-turn tetraglycyl peptide, which is used as the example input file in this paper, has a number of well-defined conformations in protein crystal structures.[13] These conformations are known as types I, I′, II, and II′, and are defined with reference to the torsion angles for the central

two residues (i.e., ϕ_2, ψ_2 and ϕ_3, ψ_3). Each of these types was reproduced by RAMBLE, although a number fell outside these boundaries (see Table 1 for details). The clustering again generally placed conformations of different types into different clusters, giving a final set of 11 clusters and 17 representatives. Figure 5 shows representatives of each class of reverse turn, and Color Plate 2 shows one complete cluster of type II′ structures. The CPU required for the task was about 350 s and around 10^5 iterations were required to produce 100 successful structures.

Table 1. Definition of reverse-turn criteria and details of those produced by RAMBLE[a]

	Type I	Type II	Type I′	Type II′
Richardson (1981)	$-140 < \phi_2 < -10$ $-80 < \psi_2 < 50$ $-110 < \phi_3 < -10$ $-80 < \psi_3 < 20$	$-110 < \phi_2 < -10$ $70 < \psi_2 < 170$ $30 < \phi_3 < 130$ $-50 < \psi_3 < 50$	$10 < \phi_2 < 140$ $-50 < \psi_2 < 80$ $10 < \phi_3 < 110$ $-20 < \psi_3 < 80$	$10 < \phi_2 < 110$ $-170 < \psi_2 < -70$ $-130 < \phi_3 < -30$ $-50 < \psi_3 < 50$
Number found	1	18	22	1
Generalization of above	$\phi_2 < 0$ $\phi_3 < 0$	$\phi_2 < 0$ $\psi_2 > 0$ $\phi_3 > 0$	$\phi_2 > 0$ $\phi_3 > 0$	$\phi_2 > 0$ $\psi_2 < 0$ $\phi_3 < 0$
Number found	6	45	38	11

[a]The total number of conformations generated by RAMBLE was 100. The table shows the classification and reclassification of the same set of data. In the Richardson classification, note that 54 of the generated structures fell outside the limits. The generalized version of the Richardson limits[13] was obtained by considering only whether the ranges allow positive or negative values of each torsion angle

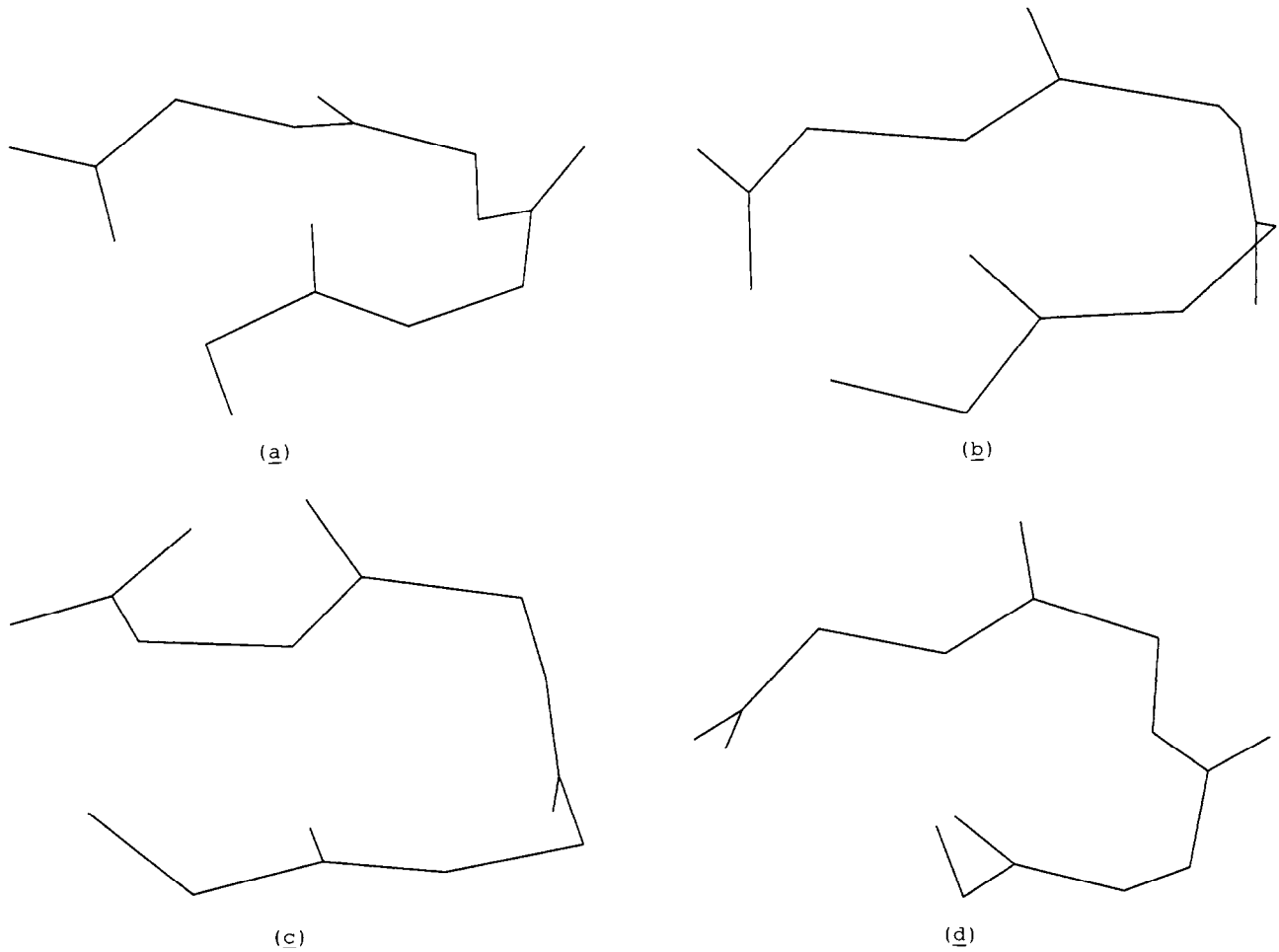


Figure 5. Selected representative conformations of reverse turn tetrapeptides generated by RAMBLE: (a) type I, (b) type II, (c) type I′, (d) type II′

3. Siderophore complex: a polycyclic system coordinating an Fe^{3+} ion

The bacterial siderophore complex ferrichrome consists of a cyclic hexapeptide with three ornithine side chains providing hydroxamate ligands to an Fe^{3+} ion (Figure 6). Since the molecule is polycyclic, it was built using RAMBLE in a number of stages:

(1) Modeling of the hexapeptide unit (including C_β atoms) (Figure 7*a*).
(2) Modeling of a loop of two ornithine side chains and the metal (Figure 7*b*).
(3) Combination of the first two stages using the program NFIT which overlaps and merges common atoms.
(4) Modeling of third ornithine link (Figure 7*c*), so that the distance between its endpoints was constrained to be within the range found for the common atoms in the fragments already generated.
(5) Combination of the third fragment with the composites already obtained from stage 3 above, again using the program NFIT.

From the 100 conformers generated for each of the three fragments, produced in 300, 600, and 6 min of CPU time, respectively, clustering produced a list of 19 representative hexapeptides and 13 second-stage structures. These structures were merged to give 95 molecules, which were combined with the 43 representative third-stage fragments to produce a total of 108 complete ferrichrome conformations. Several of these conformations had the same overall shape as the crystal structure.[14] The least similar structure had an RMS difference of 4.08 Å, while the closest conformer had an RMS difference of only 1.16 Å (see Color Plate 3 for overlaid structures). These results illustrate the ability of the method to model closely the actual conformations observed with this type of molecule.

DISCUSSION

A method of random assignment of torsion angles using the program RAMBLE has been shown to be successful at producing the known conformations of a variety of cyclic structures. It also allows for variation of bond angles and bond lengths, which is a vital feature for modeling the structures of metal–ion complexes. Cluster analysis has been used to simplify the data by grouping structures into conformationally similar families. In addition to the structures described above, successful conformational analyses of a number of

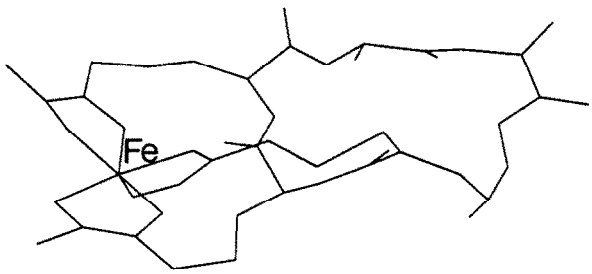


Figure 6. Molecular structure of the ferrichrome–Fe^{3+} complex

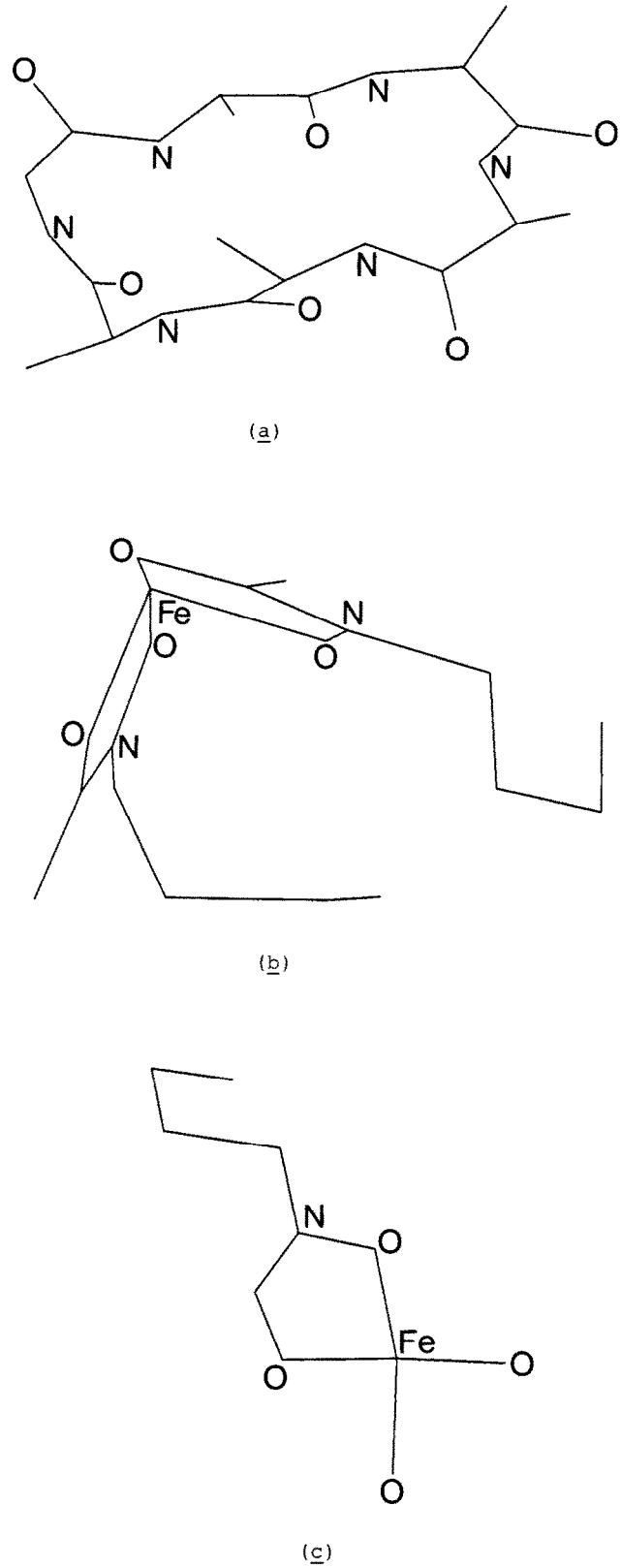


Figure 7. Fragments of the bacterial siderophore complex ferrichrome, used in its construction by RAMBLE and NFIT: (a) cyclohexapeptide fragment, (b) the stage 2 fragment (consisting of two ornithine side-chains and the Fe^{3+} ion), (c) the third arm (consisting of an ornithine side-chain and the Fe^{3+} ion)

other molecules have also been undertaken. These include novel hexadentate metal–ion chelating agents, and a disulfide-linked fragment of the peptide hormone urotensin II.

A number of other methods have also been used to explore the conformational space of cyclic molecules (see Ref. 4 for a review). The simplest is the grid search in which each rotatable bond is systematically rotated by a fixed increment, and the resulting structures are tested for correct ring closure. This is a very straightforward approach, but the time required for the search increases exponentially with increasing number of atoms. A number of techniques have been used to improve on this method, including the ring closure algorithm of Gō and Scheraga.[5] Here, $n-6$ out of n rotatable torsion angles in a loop are fixed, and the values of the six remaining torsion angles are solved mathematically to give exact ring closure. This is thus equivalent to exploring a subspace of the allowed conformations, and has the advantage that the time required rises exponentially with $n-6$ rather than n, reducing the scale of the problem. In later studies made by Bruccoleri and coworkers[6,7] this method was extended to allow for variation in bond angles in addition to torsion angles.

A rather different approach has been taken by Shenkin *et al.*[3] in modeling antibody hypervariable loops. In this method, linear fragments of the loops are generated by random assignment of torsion angles. A series of "directed tweaks" are then applied to the structures, using an iterated Lagrange multiplier method, to close the loop with minimum perturbation. Since nearly all the generated fragments are forced into acceptable conformations, the efficiency of the method appears quite impressive.

RAMBLE is not necessarily the fastest of the available programs, therefore, but it does have a number of unique advantages. It is very flexible since all the data are input by the user to the directives file, and hence is not limited to any one type of molecule. The user specifies exactly the range of values any torsion angle may occupy, and this can be used, for example, to exclude any peptide (ϕ, ψ) values not allowed by the Ramachandran map. Similarly, the ability to use flexible bond angle and bond length values is included in the program, and is particularly useful when modeling metal–ion complexes. In this case the lack of reliable force fields for energy minimization makes the generation of sensible starting structures particularly important. The incorporation of van der Waals checking is another way in which the quality of the structures that are output is verified. The built-in cluster analysis routine provides a simple way of classifying the structures produced, and ignoring any that are too similar to others. At the same time, it can give an indication of the possible conformations allowed for a structure, provided that there is thorough sampling of the available conformational space. While this may not be possible for systems with a large number of degrees of freedom, it does not appear to present problems for molecules of the size and complexity considered here.

In future versions of RAMBLE the use of some technique such as the Gō and Scheraga ring-closure algorithm to complete more of the generated loops is envisaged. However, this may be difficult to implement if the ability to include flexible bond angles and lengths is retained. In addition, more specific versions may be developed for peptides to allow "intelligent" sequence-specific building of the backbone and automatic addition of side chains.

ACKNOWLEDGMENTS

TDJP acknowledges receipt of a SERC quota award, and DJB a University of London Central Research Fund grant.

REFERENCES

1 Dean, P. M. *Molecular Foundations of Drug-Receptor Interaction*. Cambridge University Press, Cambridge, 1987, pp. 122–128
2 Wüthrich, K. Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science* 1989, **243**, 45–50
3 Shenkin, P. S., Yarmush, D. L., Fine, R. M., Wang, H. and Levinthall, C. Predicting antibody hypervariable loop conformation. I. Ensembles of random conformations for ringlike structures. *Biopolymers* 1987, **26**, 2053–2085
4 Burt, S. K. and Greer, J. Search strategies for determining bioactive conformers of peptides and small molecules. *Annu. Rep. Med. Chem.* 1988, **23**, 285–294
5 Gō, N. and Scheraga, H. A. Ring closure and local conformational deformations of chain molecules. *Macromolecules* 1970, **3**, 178–187
6 Bruccoleri, R. E. and Karplus, M. Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers* 1987, **26**, 137–168
7 Bruccoleri, R. E., Haber, E. and Novotný, J. Structure of antibody hypervariable loops reproduced by a conformational search algorithm. *Nature* 1988, **335**, 564–568
8 White, D. N. J., Tyler, J. K. and Lindley, M. R. High-performance microcomputer molecular modelling. *Comput. Chem.* 1986, **10**, 193–199
9 McLachlan, A. D. Gene duplications in the structural evolution of chymotrypsin. *J. Mol. Biol.* 1979, **128**, 49–79
10 Remington, S. J. and Matthews, B. W. A systematic approach to the comparison of protein structures. *J. Mol. Biol.* 1980, **140**, 77–99
11 Späth, H. *Cluster-Analysis Algorithms for Data Reduction and Classification of Objects*. Ellis Horwood, Chichester, 1980, pp. 172–174
12 Bernstein F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. and Tasumi, M. The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 1977, **112**, 535–542
13 Richardson, J. S. Anatomy and taxonomy of protein structure. *Adv. Protein Chem.* 1981, **34**, 167–339
14 Zalkin, A., Forrester, J. D., and Templeton, D. H. Ferrichrome-A tetrahydrate. Determination of crystal and molecular structure. *J. Am. Chem. Soc.* 1966, **88**, 1810–1814
15 Dean, P. M. and Callow, P. Molecular recognition: identification of local minima for matching in rotational 3-space by cluster analysis. *J. Mol. Graphics* 1987, **5**, 159–164