Journal of Biomolecular NMR, **11:** 241–263, 1998.
KLUWER/ESCOM

Distributed torsion angle grid search in high dimensions: A systematic approach to NMR structure determination

Garry P. Gippert*, Peter E. Wright & David A. Case**
Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, U.S.A.

Received 5 May 1997; Accepted 30 September 1997

Key words: conformation filter, distance geometry, parallel computing, systematic search

Abstract

Two complementary approaches for systematic search in torsion angle space are described for the generation of all conformations of polypeptides which satisfy experimental NMR restraints, hard-sphere van der Waals radii, and rigid covalent geometry. The first procedure is based on a recursive, tree search algorithm for the examination of linear chains of torsion angles, and uses a novel treatment to propagate the search results to neighboring regions so that the structural consequences of the restraints are fully realized. The second procedure is based on a binary combination of torsion vector spaces for connected submolecules, and produces intermediate results in Cartesian space for a more robust restraint analysis. Restraints for NMR applications include bounds on torsion angles and internuclear distances, including relational and degenerate restraints involving equivalent and nonstereoassigned protons. To illustrate these methods, conformation search results are given for the tetrapeptide APGA restrained to an idealized β-turn conformation, an alanine octapeptide restrained to a right-handed helical conformation, and the structured region of the peptide SYPFDV.

Introduction

The calculation of three-dimensional structures of peptides and proteins from NMR-derived restraints normally involves a pseudorandom sampling of conformation space with a distance geometry algorithm, followed by restrained molecular dynamics and simulated annealing (Braun, 1987; Gippert et al., 1990; Case and Wright, 1993). The distance geometry algorithms fall into two categories depending on whether the initial sampling is performed in distance space for metric-matrix distance geometry (Crippen and Havel, 1988; Havel, 1991) or in torsion angle space for variable target function distance geometry (Braun and Gō, 1985; Güntert and Wüthrich, 1991). In contrast to more-or-less random sampling of conformation space offered by these methods, systematic search methods offer the possibility to identify all unique, satisfactory conformations for a given set of restraints. It has long been recognized, however, that the large number of degrees of freedom in flexible biomolecules precludes an exhaustive treatment. Conformation space search can be made more efficient if dihedral angles are limited to values within the low-energy regions of conformation space for polypeptides (Ramachandran and Sasisekharan, 1968), a result which can be mimicked by calibration of hard-sphere van der Waals radii (Iijima et al., 1987) with freely rotatable ϕ and ψ torsion angles. Systematic search algorithms for the determination of individual amino acid side-chain conformations together with stereospecific assignments have been developed (Güntert et al., 1989; Nilges et al., 1990) with documented improvements in the accuracy of NMR structures. Similar treatments have been used to simultaneously determine the conformations and stereospecific assignments of heme propionates (Morikis et al., 1993). Improved computer speed has allowed the systematic determination of ensembles of satisfactory structures for NMR-restrained,

*Current address: Physical Chemistry 2, Lund University, Chemical Center, Box 124, S-22100 Lund, Sweden; http://www.fkem2.lth.se/~garry.
**To whom correspondence should be addressed.

cyclic peptide systems whose unique elements number in the millions (Beusen et al., 1990, 1996).

Here we describe a procedure for the systematic search in torsion angle space for the efficient production of conformations that satisfy all restraints. The search procedure is composed of two distinct algorithms. The first, Distance-resolved Torsion Angle Grid Search (DTAGS), operates in distance space and is mostly concerned with the efficient pruning of conformation space and the production of distance bounds extrema resulting from the search. Conformation space is represented and stored as high-dimensional grids in computer memory, and the search is limited to linear chains of torsion angles, although suitable algorithms are utilized to redistribute search results in the vicinity of a branch in the chain (i.e., side chain). The second algorithm, NEWMOL, operates in Cartesian space, and is used to extend the search results from DTAGS to high dihedral dimensions. Conformations are stored as lists of torsion angle vectors and are combined to restrict new trial conformations based on a binary tree approach. NEWMOL produces fully three-dimensional models including side-chain branching and flexible five-membered rings for proline residues. In addition to providing a detailed description of these algorithms, we describe several applications of systematic search for the identification of accessible conformation space for small peptides restrained by simulated and experimental NMR distance bounds. We also describe several ways in which large conformational ensembles may be analyzed and compared to NMR data, including the back-calculation of distances for the accumulation of the distance extrema matrix, and the superposition of a large number of three-dimensional models to a mean-distance coordinate frame for the visualization of conformational fluctuations allowed by the restraints.

Methods

The general subject of conformational search has been addressed elsewhere (Braun, 1987; Leach, 1991). We use a rigid-geometry approximation (Gō and Scheraga, 1970), which restricts the search to flexible torsion angles, and uses fixed values for bond lengths and bond angles. Steric interactions and experimental restraints are represented by hard spheres, so that conformations are either allowed or disallowed. A conceptually simple conformation search algorithm can be encoded by a series of nested loops, where each loop corresponds to the assignment of a torsion angle to its possible values. Within the innermost loop all degrees of freedom have been assigned and transformed appropriately to produce a three-dimensional conformer. The conformer is evaluated by comparison to restraints, and the results are stored in a conformational array, or grid, having as many dimensions as there are degrees of freedom in the system. After the search, the ensemble containing only allowed conformers can be obtained from examination of the contents of the conformation grid. Although this simple algorithm is guaranteed to find all satisfactory conformations that can be represented by the grid, it is not generally the fastest method. We impose several conditions on the conformational model that together reduce the number of trial conformations that must be examined, thus permitting the search to proceed more rapidly, at the expense of increased complexity for the search algorithm.

The key idea to improve efficiency is to break the conformation search up into smaller pieces, so that the allowed conformation space for small fragments is used to restrict trial conformations for larger fragments. Conformers are evaluated in an ‘all-or-nothing’ manner, and the corresponding point in a conformation grid is set to ‘1’ if there are no restraint violations, and is set to ‘0’ if there are restraint violations. The allowed conformation space is then read directly by locating the ‘1’s remaining in the grid. This is similar to the ‘build-up’ procedure (Simon et al., 1991) used for energy-based calculations, except that all allowed conformers are used in later stages of the calculation, not only those having low energy. A complication that arises in the hierarchical search described here is that a given submolecule conformation may satisfy its internal restraints, but may produce consistent restraint violations when considered as part of a larger fragment. Thus, local torsional correlations may arise as a result of non-local restraints. To take advantage of this information, we introduce a ‘build-down’ procedure, where, in the grid search algorithm, the results of higher dimensional searches are used to restrict the allowed conformation space of lower dimensional submolecules. This approach distributes the consequences of the restraints to neighboring degrees of freedom in the chain, thus producing an efficient pruning of conformation space.

Distance-space grid search

In this section we describe a search in torsion angle space for conformers that satisfy a matrix of interatomic distance bounds. Torsion angles are independently sampled on a torsion grid which is stored as a high-dimensional array in computer memory. Three-dimensional conformations are not produced directly. This method will be referred to as the 'distance-space' or 'grid' search.

Consider a linear molecular chain composed of n atoms. The effects of closed rings, branches and additional atoms attached to the chain are neglected for the moment. The distance between atoms 1 and n is defined in terms of intervening degrees of freedom, and may be computed by an iterative process whereby the coordinates of atom 1 are transformed to a coordinate system at the i-th atom position (Hendrickson, 1961):

$$\mathbf{x}_i = p_{i-1} + \mathbf{T}_{i-2}\,\mathbf{R}_{i-3}\,\mathbf{x}_{i-1} \qquad (1)$$

where

$$\mathbf{p}_i = \begin{bmatrix} d_i \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{T}_i = \begin{bmatrix} -\cos\theta_i & -\sin\theta_i & 0 \\ \sin\theta_i & -\cos\theta_i & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{R}_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\omega_i & \sin\omega_i \\ 0 & -\sin\omega_i & \cos\omega_i \end{bmatrix}$$

Here, d_i is the i-th bond length (joining atoms i and i+1), θ_i is the i-th bond angle (formed by atoms i, i+1, and i+2), and ω_i is the i-th torsion angle (formed by atoms i, i+1, i+2, and i+3, with increasing ω_i corresponding to a clockwise rotation of atom i+3 with respect to atom i looking along the i+1 $\rightarrow$ i+2 bond).

A systematic variation of torsion angles has been encoded into a computer program called DTAGS, which produces all combinations of torsion angles sampled independently at chosen values. A conformation array is allocated in computer memory to represent all combinations of torsion angle values that the chain may adopt. Each grid element is initially 1 (allowed) and is set to 0 (disallowed) if bounds violations are detected. Subsequent search procedures utilize the stored conformational information in order to avoid reexamination of disallowed regions in conformation space.

Each trial conformer is examined for violations of restraints defined between groups of atoms at the two ends of the chain, called aggregates (Dammkoehler et al., 1989). This permits relational restraints to be applied to equivalent and prochiral groups (methylene protons and isopropyl pseudoatoms). For instance, distances to equivalent protons may be subject to r^{-6} averaging (Gippert et al., 1990; Fletcher et al., 1996) or combined to form a 'degenerate distance restraint' that is satisfied if at least one of the equivalent distances satisfy the bound. Relational restraints may also be imposed as bounds on the ratio between two distances, for example a restraint that requires one distance to be shorter than another, independent of the precise details of the calibration of NOE intensity to distance bound. A similar approach permits relational restraints to be applied for ^{3}J couplings to both protons in a methylene group.

Examination of all linear subchains in a complex system occurs as part of a hierarchical search for chains of increasing size. The results from lower dimensional searches are used to restrict conformation space for higher dimensional searches. First examine all single torsion angle chains (1-chains) and update the corresponding grids. Next, examine pairs of adjacent torsion angles (2-chains) and update the corresponding grids. This process is continued (3-chains, etc.) until conformation space for the longest linear subchain(s) in the system has been systematically explored. At each search depth, the disallowed points identified for small chains are 'pruned' from the search tree for larger chains. An algorithm that performs a hierarchical, recursive, systematic search in torsion angles for a linear chain is presented in Algorithm A (see Appendix). The notation (a, n) designates a subchain of length n beginning with torsion angle number a. When the *i*-th torsion angle along the chain from a is encountered the decision to continue recursion is made based on preexisting conditions in the conformation grid.

A recurrence relation (line A16 calls A10) lends itself to the evaluation of Equation 1 for the calculation of end-to-end distances for atoms in a chain. For efficiency, the rotation matrices corresponding to the product $\mathbf{T}_i\mathbf{R}_{i-1}$ are precomputed based on the geometry of the subchain and the chosen torsional grid points. Within the innermost loop of Algorithm A (lines A18–A24) the chain end-to-end distance is computed and compared to distance bounds. When a violation is detected, the corresponding point in the conformation matrix C_j is set to zero (line A21). The

subscript j indicates the depth of recursion, from 1 to n. Note that recursion may be 'pruned' according to the elements of the conformation matrices $C_{j<n}$ (line A13) which correspond to smaller subchains (a, j) which were examined earlier in the search hierarchy. Algorithm A also illustrates that torsion angles do not necessarily need to be sampled uniformly; line A12 shows that torsion angle i is indexed through a list of values containing M_i elements which may be chosen arbitrarily before the search begins.

To give an example of a hierarchical linear-chain grid search, consider a torsion chain consisting of the backbone torsions in a single amino acid residue, including flanking peptide bonds (Figure 1). As Algorithm A implies, the search order for subchains is top-to-bottom (increasing size of submolecule) and left-to-right (torsion position along the chain). The submolecules in Figure 1A are labelled with the overall index order of the search. Figure 1B gives the identity of aggregate atoms for the corresponding chains given in Figure 1A which are involved in bump-checks and restraint evaluation. A subset of chains with a proton in both aggregates (underlined) is sensitive to geometric restraints obtained from homonuclear NMR measurements, for example, the ϕ subchain (2) is used to compute the magnitude of the scalar coupling between HN and Hα protons (Pardi et al., 1984), and the $\psi\omega_+$ (7) and $\phi\psi\omega_+$ (9) subchains are used to compute the sequential Hα-HN and HN-HN NOE distances, respectively (Billeter et al., 1982; Wüthrich, 1986). In the DTAGS program, peptide bonds are treated as explicit bond rotations, although in some applications they are permitted to adopt only a single angle corresponding to a cis (0°) or trans (180°) conformation. All torsion angles are otherwise permitted free rotation through 360° except in the case of torsional restraints or pseudorotation.

The line segments in Figure 1A connect a chain (below) with its two immediate subchains (above). Search results from *left* subchains (thick lines) are directly available within the recursion algorithm and may be used to reduce the number of trial conformations during execution of the search procedure in Algorithm A. For example, it is not necessary to examine torsion combinations ABC for any value of C if it was previously determined that AB is disallowed. The asymmetry of this graph reflects an inherent property of recursion algorithms. Indexing of conformational array elements for left subchains, as illustrated here, is performed during recursion without additional cost.

An additional reduction in the number of trial conformations may arise from consideration of conformational array elements for *right* subchains, indicated by thin lines in Figure 1A. This is accomplished by preparing a (trial) conformational array to contain only allowed combinations of elements in both *left* and *right* subarrays. This additional information is then used during recursion (line A13) to avoid redundant steps in the calculation. The *initialize* subroutine called in Algorithm A (line A6) performs the described operation and is shown in Algorithm B (see Appendix). Recursion is used (line B11 calls B7) to sample all elements of the relevant conformational grids. The instructions that relate a conformation array to its left and right subarrays are shown in lines B13 and B14. For example, a trial conformation ABC is disallowed if either AB or BC substate is disallowed. Many elements of Algorithm B are identical to those found in Algorithm A.

Algorithm B is itself recursive in that initialization is applied to all subchains in a 'build-up' direction, i.e., from smaller to larger subchains beginning with chains of dimension 1. Again, the notation (a, n) describes a linear subchain of length n starting at position a in a larger chain. The left sub-chain is designated (a, n–1) and the right sub-chain is designated (a+1, n–1). Before the conformational array for chain (a, n) is initialized (line B5), the initialization procedure is applied to both left and right subchains (lines B3 and B4, respectively). As a consequence, trial conformations within a conformational space are restricted to combinations of torsion angles in allowed regions of its lower dimensional, constituent subspaces.

After a search, conformational information may also pass from larger to smaller chains by using a 'build-down' procedure which is essentially the reverse of that described above. For example, if conformation ABC is disallowed for all values of C, then AB is also disallowed even when, considered in isolation, it produces no restraint violations. This occurs often for peptide systems in that a search of $\omega\phi\Psi\omega_+$ is more restrictive than a search of $\phi\Psi$ alone due to steric restrictions imposed by atoms in adjacent peptide groups. A procedure which updates the conformation array for left (a, n–1) and right (a+1, n–1) subchains based on a 'skyline' projection of the conformation array for chain (a, n) is given in Algorithm C (see Appendix). For skyline projection within recursion, it is necessary to treat left and right subchains using separate routines. The projection of (a, n) onto (a, n–1) along the last ((a+n–1)-th) dimen-

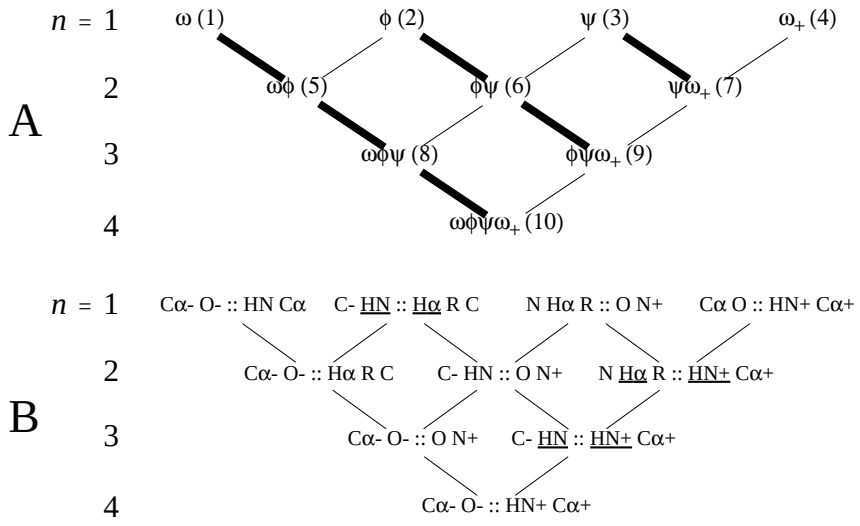


Figure 1. Search tree for a dipeptide containing ω, ϕ, Ψ, and ω_+ torsion angles (the '+' indicates the following residue) showing (A) subchains and search order, and (B) corresponding atom aggregates used to evaluate restraints.

sion is treated by lines C8–C20, and the projection of (a, n) onto (a+1, n–1) along the first (a-th) dimension is treated by lines C21–C33. The '∨' symbol used in lines C16 and C29 is the Boolean 'OR' operator. Algorithm C is itself recursive in that projection is applied to all constituent subchains in a 'build-down' direction, i.e., from larger to smaller subchains ending with chains of dimension 1. Initialization (Algorithm B) and projection (Algorithm C) procedures traverse the entire subtree.

During the initialization and projection operations described above, consideration must be taken for the relative orientation of the subchains in the search tree. For a completely linear molecule with no side-chain branches, all subchains have the same orientation, i.e., are parallel to one another. In this circumstance, the recursion indices for a chain and its immediate subchains will have the same order. However, an unavoidable consequence of side-chain branching is that some of the subchains will have an antiparallel orientation with respect to one another, illustrated in Figure 2 for a single amino acid residue composed of backbone and side-chain degrees of freedom. The purely recursive form for the initialization and projection algorithms (Algorithm B and Algorithm C, respectively) cannot be used for comparing subchains with an antiparallel relationship. Modified forms for these algorithms (not shown) are used for the three possible subchain orientations: head-to-tail (parallel) ($\rightarrow\rightarrow$), head-to-head (antiparallel) ($\rightarrow\leftarrow$), and tail-to-tail (antiparallel) ($\leftarrow\rightarrow$). The DTAGS program automatically takes care of the appropriate bookkeeping functions and program calls for all subchain orientations.

Cartesian-space binary tree search

In this section we describe a search in torsion angle space for conformers that satisfy a matrix of interatomic distance bounds and a variety of other restraints. Conformers are represented by vectors of torsion angles, which may be generated independently from the actual search procedure, for example, by a systematic, biased or random sampling. Torsion vectors for connected (adjacent) submolecules are combined to form trial conformers for larger submolecules using a binary tree approach. Trial conformations are computed as three-dimensional models, which allows for side-chain branching and closed five-membered rings, in addition to a significant increase in the types of restraints available for NMR structure determination or more general modelling applications. This method will be referred to as the 'Cartesian-space' or 'binary tree' search.

Consider chain elongation generated by specifying a value for a single torsion angle ω_i. For chain elongation to occur, positions for the first three atoms in the chain must be predefined, represented by the vectors $\mathbf{x}_1 = \{0, 0, 0\}$, $\mathbf{x}_2 = \{d_1, 0, 0\}$, and $\mathbf{x}_3 = \{d_2 - d_1 \cos\theta_1, d_1 \sin\theta_1, 0\}$. The position of atom i+3 is determined in terms of local chain parameters by

$$\mathbf{x}_{i+3} = \mathbf{x}_{i+2} + \mathbf{R}_{i+2}\,\mathbf{r}_{i+3}$$

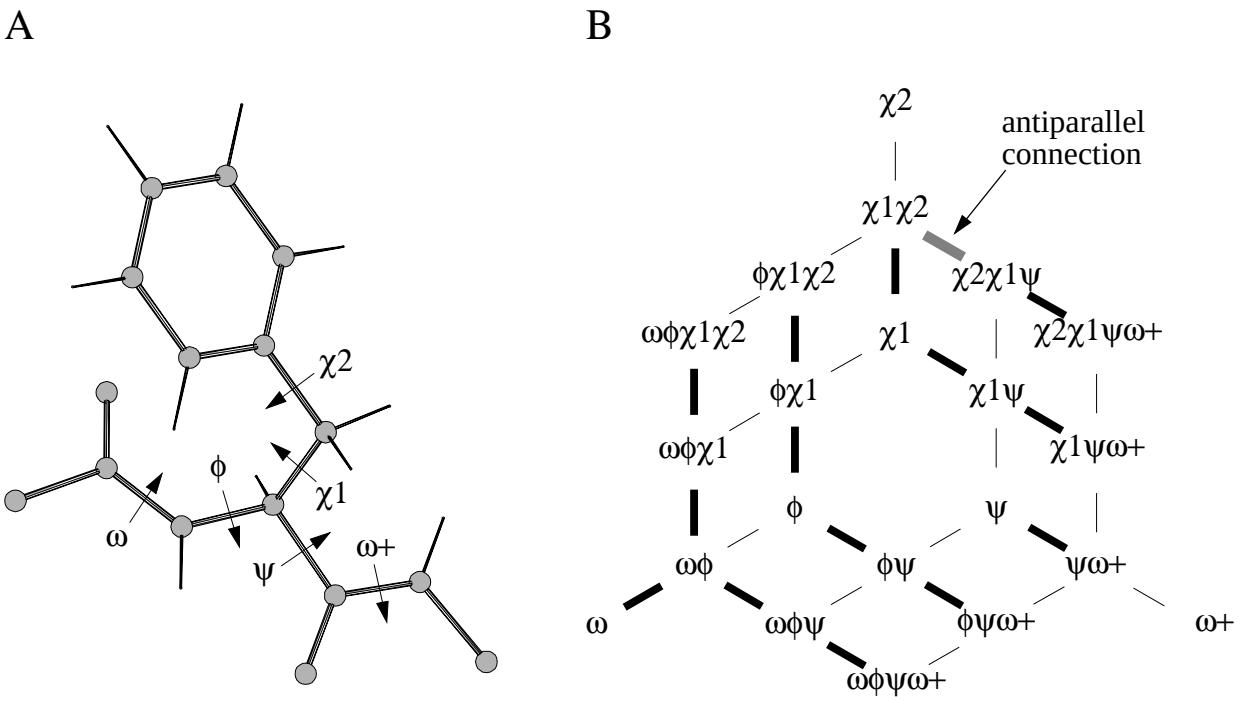


Figure 2. (A) Ball-and-stick model for phenylalanine residue with backbone and side-chain torsion angles, and (B) corresponding search tree showing an antiparallel orientation for $\chi^1\chi^2$ and $\chi^2\chi^1\Psi$ subchains.

where

$$\mathbf{R}_{i+2} = \begin{bmatrix} e_{1x} & e_{2x} & e_{3x} \\ e_{1y} & e_{2y} & e_{3y} \\ e_{1z} & e_{2z} & e_{3z} \end{bmatrix} \qquad (2)$$

and

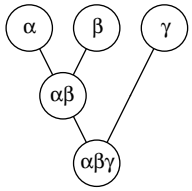
$$\mathbf{r}_{i+3} = d_{i+2} \begin{bmatrix} -\cos\theta_{i+1} \\ \sin\theta_{i+1}\cos\omega_i \\ \sin\theta_{i+1}\sin\omega_i \end{bmatrix}$$

The displacement vector $\mathbf{r}_{i+3}$ defines the position of atom $i + 3$ in a coordinate system defined with atom $i + 2$ at the (local) origin, and with orientation defined by the components of the matrix $\mathbf{R}_{i+2}$. The unit-length, principal axis vectors $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ form a right-handed, rectangular coordinate system defined with $\mathbf{e}_1$ parallel to the rotation axis in the direction atom $i+1$ to atom $i+2$, and $\mathbf{e}_3$ along $\mathbf{e}_1 \times (\mathbf{x}_i - \mathbf{x}_{i+1})$.

A convenient encoding of Equation 2 in a computer program NEWMOL is used to produce three-dimensional coordinates for a molecule as a function of its torsion angles and internal geometry. The rigid-geometry approximation allows for aggregates of atoms with internally rigid geometry to appear at the chain ends (Dammkoehler et al., 1989), so that the orientation matrix **R** needs to be calculated only once per chain elongation. Conformational information supplied to NEWMOL includes a list of torsion angle names in chain-elongation order followed by a list of torsion vectors, each containing one value for each torsion angle in the submolecule. By application of Equation 2, a three-dimensional model is produced in computer memory, and the conformer is evaluated with respect to bounds on distances and other restrained geometric quantities. Successful (allowed) conformers are printed out to an accumulating list of torsion vectors, whereas unsuccessful (disallowed) conformers produce no output. This can be contrasted to the scheme used to store the allowed/disallowed status for conformers generated using recursion, where a binary 1 or 0 is stored in a high-dimensional grid in computer memory. The use of torsion vectors both as input and output to NEWMOL suggests the terminology of conformation filter rather than conformation search.

A hierarchical scheme is used to generate a systematic search using torsion vectors. The set of allowed conformers for submolecule $\alpha\beta$ is a subset of the product of all conformers for its connected fragments α and β. Here, α and β represent adjacent submolecules

A $\{\alpha\} \bullet \{\beta\} \bullet \{\gamma\} \rightarrow \{\alpha\beta\} \bullet \{\gamma\} \rightarrow \{\alpha\beta\gamma\}$



B $\{\alpha\} \bullet \{\beta\} \bullet \{\gamma\} \rightarrow \{\alpha\} \bullet \{\beta\gamma\} \rightarrow \{\alpha\beta\gamma\}$

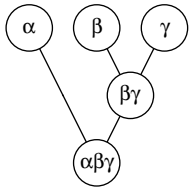


Figure 3. Binary composition trees for three adjacent submolecules α, β, and γ with intermediate compositions (A) α with β, and (B) β with γ.

with one or more internal degrees of freedom such that their combination yields a contiguous submolecule. It is convenient to write this combination (pruning out conformations that fail to satisfy the restraints) as

$$\{\alpha\} \cdot \{\beta\} \rightarrow \{\alpha\beta\} \tag{3}$$

The set of allowed conformations of submolecule α is denoted {α}, etc. If conformation sets {α} and {β} are consistent with restraints that depend entirely on internal degrees of freedom for α and β submolecules, respectively, then the evaluation of trial conformations in the set {αβ} need only consider restraints that depend on combined degrees of freedom for the αβ submolecule.

The choice of binary search strategies is crucial for completion of the search. Consider the example of three connected submolecules α, β, and γ that form a linear molecule αβγ. Both αβ and βγ are possible intermediate steps in a search calculation (Figure 3), and both search orders will produce identical results for {αβγ}. It is easy to construct a problem where the intermediate conformation set {βγ} contains many fewer elements than set {αβ}. In this case, the second composition order (Figure 3B) is likely to be more efficient (other factors being equal), written as an expression with parentheses to indicate nesting of operations:

$$\{\alpha\} \cdot (\{\beta \cdot \gamma\}) \rightarrow \{\alpha\beta\gamma\} \tag{4}$$

In general, there are $O_n = (2n)!/n!(n+1)!$ possible orders for n compositions. Sometimes an optimal search strategy may be inferred by inspection of the constraint topology.

In practical terms, a conformation set is stored in an *enumeration* file for the submolecule. This file contains the identity of the torsion angles in the submolecule and a list of position vectors in torsion angle space. The input to NEWMOL is a list of commands that tell it to produce and evaluate three-dimensional structures. The program SPACECAT generates NEWMOL commands for evaluating conformers listed in an enumeration file:

$$\text{SPACECAT a} \mid \text{NEWMOL} > \text{a}' \tag{5}$$

where a and a′ are names of enumeration files corresponding to trial and filtered conformations for the submolecule, respectively. Similarly, the program SPACEJOIN generates NEWMOL commands for combining elements from two enumeration files. Hence, $\{\alpha\} \cdot \{\beta\} \rightarrow \{\alpha\beta\}$ can be written as

$$\text{cat a} \mid \text{SPACEJOIN b} \mid \text{NEWMOL} > \text{ab} \tag{6}$$

where a, b and ab are names of enumeration files corresponding to α, β and αβ submolecules, respectively. The meaning of expression 6 is 'Combine each conformer in file a with each conformer in file b, examine the combinations and write out the allowed conformers to a file ab.' Using this approach, several program executions can be written as a single command line:

$$\text{cat a} \mid \text{SPACEJOIN b} \mid \text{NEWMOL} \mid \text{SPACEJOIN c} \mid \text{NEWMOL} > \text{abc} \tag{7}$$

where intermediate results (for example, file ab) are not written to disk, but exist transiently within the pipeline. The order of operations in expression 7 is represented by the sequential binary tree (Figure 3A). The non-sequential order (Figure 3B) can also be written on a single command line:

$$\text{cat b} \mid \text{SPACEJOIN c} \mid \text{NEWMOL} \mid \text{SPACEJOIN -exch a} \mid \text{NEWMOL} > \text{abc} \tag{8}$$

where the '-exch' option to SPACEJOIN exchanges the positions of the input submolecules. The extension to additional processes is straightforward. The pipeline permits data transfer over a network in a multiprocessor environment, which means that elements of the conformation search, i.e., independent NEWMOL executions, are performed in parallel.

Combined grid/tree search

In many situations it is advantageous to perform a combined grid/binary tree search. Fewer computer instructions are needed to evaluate restraint violations for a grid search calculation compared to a binary tree search. The grid search also determines low-dimensional (near-neighbor) torsional correlations relatively early on, and propagates the correlations by projection and initialization operations. Generally, a grid search is faster than a binary tree search for fairly small systems and for a preliminary search of short subchains in larger systems. However, the grid search strategy becomes less attractive when the conformation arrays become sparsely populated, as can happen particularly for well-restrained systems. Under these circumstances, it becomes more efficient to store and manipulate points in conformation space as position vectors in torsion space, rather than as elements of a high-dimensional grid.

To illustrate the combination of grid and binary tree strategies, consider the example of a blocked octapeptide restrained to be helical by $HN_i - O_{i-4}$ distance restraints, the details of which are unimportant here. A torsion search is applied to identify all allowed ϕ and Ψ backbone angles. Two of many possible combined grid/binary trees are shown in Figure 4. Submolecules consisting of allowed ϕ and Ψ combinations are indicated by circles at the nodes of the tree. Submolecules that directly engage a distance restraint (spanning brackets) are indicated by filled symbols (see also Figure 9).

The strategy depicted in Figure 4A performs an independent grid search on a submolecule consisting of the first four residues 1–4 and a submolecule consisting of the last four residues 5–8. A binary composition is then used to generate the conformation space for the intact molecule. The strategy depicted in Figure 4B performs grid search on overlapping submolecules up to four residues in length, i.e., residues 1–4, 2–5,..., 5–8. For B the maximum number of torsions examined is the same as for A. For both A and B the final composition operation is between submolecules 1–4 and 5–8 only.

Conformation search calculations were carried out for both strategies depicted in Figure 4, with a uniform 10° increment for ϕ and Ψ torsion angles. For A the grid search produced 3434 and 4426 conformations for submolecules 1–4 and 5–8, respectively. The (final) composition of these submolecules produced $3434 \times 4426 \approx 15$ million trial conformations, of

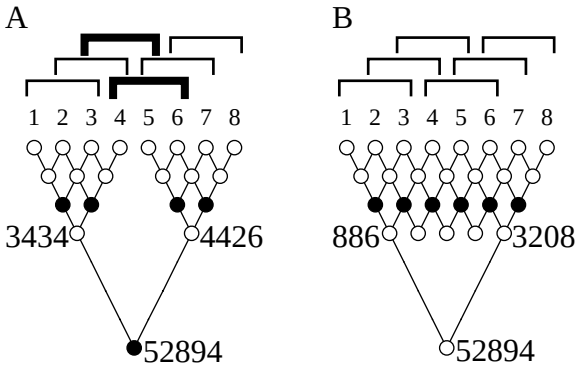


Figure 4. Combined grid/binary trees for an octapeptide restrained by HN_i–O_{i-4} distance bounds (spanning brackets). Submolecule searches that directly engage a restraint are indicated by filled symbols. (A) Separate grid search for submolecules 1–4 and 5–8 can not encounter two of the restraints (bold lines) until the final binary composition. (B) Grid search for submolecules 1–8 to the same depth as that in (A), followed by binary composition. The number of conformations after grid search and binary composition are shown for selected submolecules.

which 52 894 had no bounds violations. Highlighted by bold lines in Figure 4A are the fragment-spanning distance restraints which were not addressed during the grid search in A, but which come into play in producing the final conformation set during binary composition. For B the grid search produced 886 and 3208 conformations for submolecules 1–4 and 5–8, respectively, which in the final composition produced $886 \times 3208 \approx 2.8$ million trial conformations, but yielded the identical final set of 52 894 conformations. Strategy B, facilitated by an initial grid search using initialization and projection, is about 5 times faster than A.

Construction of bounds

The distance bounds matrix produced for conformation search is essentially identical to that used for metric-matrix distance geometry calculations (Crippen and Havel, 1988). It is desirable to allow torsional flexibility for five-membered rings such as proline side chains and nucleic acid sugars in a conformation search. A basic problem which must be solved is that variable 1–4 distances governed by endocyclic torsion angles must also satisfy 1–3 distances which are fixed by covalent bond angles on the opposite side of the ring. To accommodate a limited flexibility for five-membered rings during torsion search, we apply a uniform bounds widening to selected elements of the distance bounds matrix corresponding to 1–2 (bond length) and 1–3 (bond angle) distances in the ring

Table 1. Proline intraring distance bounds

Atom pair		Distance bounds (Å)	
Atom i	Atom j	Lower	Upper
Bond length distance ±0.050 Å			
N	Cα	1.403	1.503
Cα	Cβ	1.470	1.570
Cβ	Cγ	1.470	1.570
Cγ	Cδ	1.470	1.570
Cδ	N	1.410	1.510
Bond angle distance[a] ±0.100 Å			
Cδ	Cα	2.329	2.529
N	Cβ	2.259	2.459
Cα	Cγ	2.326	2.537
Cβ	Cδ	2.333	2.558
Cγ	N	2.236	2.460
C–H distance across 'broken' Cβ-Cγ bond ±0.100 Å			
Cβ	Hγ{2,3}	2.059	2.259
Cβ	Qγ[b]	1.873	2.073
Cγ	Hβ{2,3}	2.065	2.265
Cγ	Qβ[b]	1.880	2.080

[a] Extrema of bond angle (1–3) distances for *up* and *down* pucker states in ECEPP/2 parameters (Momany et al., 1975; Némethy et al., 1983) (deviations in parentheses): Cα-Cγ (0.011 Å), Cβ-Cδ (0.025 Å) and Cγ-N (0.024 Å).

[b] Q represents the pseudoatom at the center of the respective protons (Wüthrich et al., 1983).

(Table 1). By widening 1–3 bounds, single ring torsions may vary from planarity, and by widening 1–2 bounds, conformation space for pairs of adjacent ring torsions may be searched without solving for exact ring closure. Bounds are widened uniformly, i.e., by equal amounts for upper and lower bounds, so that the average distance (needed to describe the exact covalent geometry in the ring) is unchanged. For proline, the starting set of distances used for bounds widening are the extrema of distances in the *up* and *down* conformers in the ECEPP/2 parameter set (Momany et al., 1975; Némethy et al., 1983), shown in Table 1.

Ring flexibility is treated differently depending on whether the conformation search is performed in distance space or Cartesian space. For the distance-space search, only linear chains of torsion angles are examined. No more than two endocyclic torsions are permitted in a given chain, so that we avoid the calculation of a distance from an atom to 'itself' around the closed ring, while allowing all atoms to remain available for use in distance restraints and van der Waals bump-check. With intraring distance bounds given in Table 1, a distance space grid search produces maps of allowed $\chi^n\chi^{n+1}$ space (Figure 5), with a maximum deviation of single ring torsions $\approx 25°$ from planarity. (The endocyclic ring torsion $\chi^0(C\delta - N - C\alpha - C\beta)$ is related to backbone torsion angle ϕ by a constant projection angle, which in ECEPP/2 parameters is $\phi = \chi^0 - 57.58°$.)

For the Cartesian-space search we use a symmetric pseudorotation scheme whereby the three-dimensional conformation of the entire ring is governed by a single degree of freedom, the pseudorotation phase angle P (Saenger, 1984). Torsion vectors $X = (\chi^0, \chi^1, \chi^2, \chi^3, \chi^4)$ are produced by evaluating the expression

$$\chi^n = \chi_{max}\cos(P - n6\pi/5) \qquad (9)$$

for $n = 0 \ldots 4$ at discrete values of P. Two-dimensional projections of X are shown as continuous curves in Figure 5 for values of the pseudorotation amplitude $\chi_{max} = \chi^0/\cos(P)$ corresponding to the up and down pucker states in the ECEPP/2 parameter set, which have $P = 113.5°$, $\chi_{max} = 25.2°$ (up), and $P = -152.5°$, $\chi_{max} = 19.6°$ (down). For conformation search, an average value of $\chi_{max} = 22.4°$ is used. With bounds widening implemented as described above, this scheme permits the unbiased reconstruction of both pucker states to less than 0.04 Å all-atom rms coordinate deviation.

A subset of four out of five ring torsions is needed to produce coordinates for all atoms in the proline side chain and adjacent backbone. We have chosen the subset $X = (\chi^0, \chi^1, \chi^4, \chi^3)$ as a basis set. Residual deviations in χ^2 are smaller than 0.6° for all values of P. Due to small (less than 0.1 Å) imperfections in ring closure distances, additional C–H (1–3) distance bounds spanning the 'broken' Cβ-Cγ bond are widened to permit a full 360° rotation in P without intraring distance bounds violations (Table 1).

Contact radii

Van der Waals radii are used to describe the minimum contact distance between atoms separated by more than one rotatable bond. An initial set of atomic radii was chosen to be consistent with distance geometry programs (Braun and Gō, 1985; Güntert and Wüthrich, 1991; Havel, 1991) that use ECEPP/2 parameters (Momany et al., 1975; Némethy et al., 1983) to define covalent geometry. For each atom type, the minimum value from these sources was used to minimize the excluded conformation space. Radii were

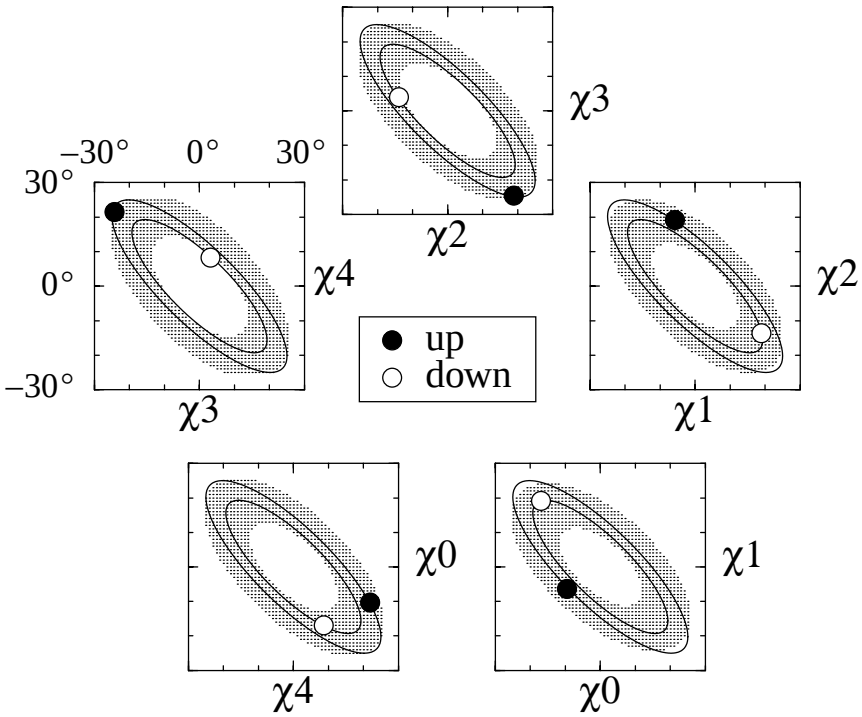


Figure 5. Allowed $\chi^n\chi^{n+1}$ regions for a proline residue after bounds widening (modifications to selected intraring 1–2 and 1–3 distance bounds) as described in the text and given in Table 1. χ^0 is the endocyclic torsion (Cδ–N–Cα–Cβ). The gray regions were produced by a 1° uniform sampling in a distance-space grid search. Superimposed are continuous pseudorotation trajectories computed as described in the text for χ_{max} values back-calculated from *up* and *down* pucker states in the ECEPP/2 parameter set (indicated by filled and open circles, respectively).

then calibrated (Iijima et al., 1987) so that allowed $\phi\Psi$ regions for a terminally blocked alanine encompass all low-energy conformation regions (Ramachandran and Sasisekharan, 1968). In some distance geometry calculations the four atoms of a methyl group are replaced by a single pseudoatom at the center of the proton positions (Wüthrich et al., 1983). In such cases the methyl pseudoatom is assigned a radius of from 1.5 to 2.0 Å (Havel, 1991) to represent the steric bulk of the atoms which it replaces. It was noted in trial torsion search calculations that for a methyl-containing side chain such as isoleucine, the use of a methyl pseudoatom radius greater than 1.5 Å excluded a significant number of sidechain conformations. We have chosen to retain pseudoatoms as a convenient reference point for NOE distance restraints, but have assigned all pseudoatoms a radius of 0.0 Å independent of rotation of the methyl group. The final set of radii used are (units of Å): C(1.40), N(1.25), O(1.20), H(0.95), S(1.60), with a smaller aromatic carbon radius of 1.35. In addition, the contact distance between (donor) hydrogen atoms and (acceptor) heavy atoms for all potential hydrogen bonds is reduced to 1.80 Å.

Figure 6 shows the allowed $\phi\psi$ regions consistent with these radii for a uniform, 10° search in ϕ and Ψ for a blocked alanine residue; the allowed regions occupy $428/(36^2) \approx 1/3$ of the total. In a separate calculation, $5 \times 10^5 \phi\Psi$ points were chosen from a pseudorandom distribution, producing 173 847 allowed points for a similar ratio. An average of 3244 allowed points (9332 random trials) were required so that each $10° \times 10°$ box was occupied at least once. Nearly 80% of the allowed random points were 'duplicates', and about 7% fell more than 10° from an allowed grid point (shown as dark, speckled regions lining the allowed regions in Figure 6).

Input files used by DTAGS and NEWMOL programs are generated from a residue library and primary sequence by the RESLIB program included in the source code distribution (see later). Files describing covalent modifications and experimental restraints for the system may also be given. Restraints include bounds on inter-atomic distances, distance rations, scalar couplings, scalar coupling ratios, virtual and covalent bond angles, virtual and covalent torsion angles, absolute value torsion angles, signed tetrahedral

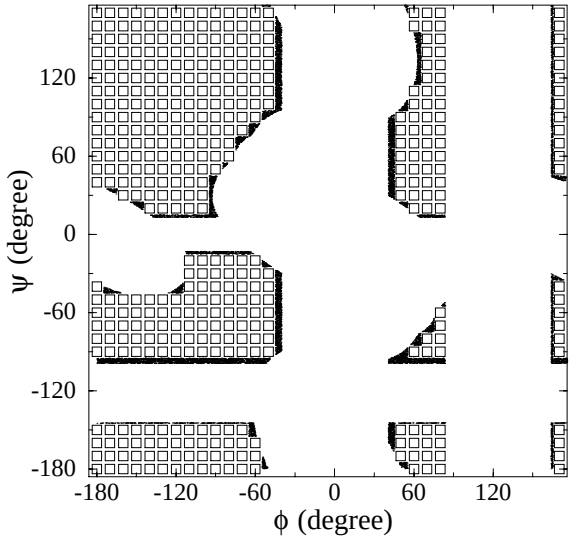


Figure 6. Allowed $\phi\Psi$ regions for a blocked alanine, with van der Waals radii as given in the text, and torsion sampling on a $10^\circ \times 10^\circ$ grid (open squares, reduced in size for clarity) augmented by randomly chosen points (speckled 'edges' of allowed regions).

volumes, oriented hydrogen bonds, and normalized dot products between internuclear or plane-normal vectors.

Results

β-*Turn peptide*

Reverse turns are common structural motifs in proteins which bring about a nearly 180° chain reversal within the space of two residues. β-Turns are classified by the values of dihedral angles at positions 2 and 3 within a four-residue turn sequence, and are further categorized by the presence or absence of a stabilizing hydrogen bond between NH at position 4 and CO at position 1 (Venkatachalam, 1968; Richardson, 1981). Type I turns have ϕ and Ψ dihedral values near $\phi_2 = -60^\circ$, $\Psi_2 = -30^\circ$, $\phi_3 = -90^\circ$, and $\Psi_3 = 0^\circ$; and type II turns have dihedral values near -60°, 120°, 90°, and 0°, respectively. As the smallest class of regular secondary structure, we undertook grid search calculations for a peptide with sequence APGA restrained by a 4–1 hydrogen bond that closes a 10-membered ring. The 'hydrogen bond' consists of distance bounds $N_4 - O_1 \leq 3.0$ Å and $HN_4 - O_1 \leq 2.0$ Å.

The distance bounds matrix was subjected to triangle inequality bound smoothing either before (B) or after (A) the hydrogen bond restraints were added,

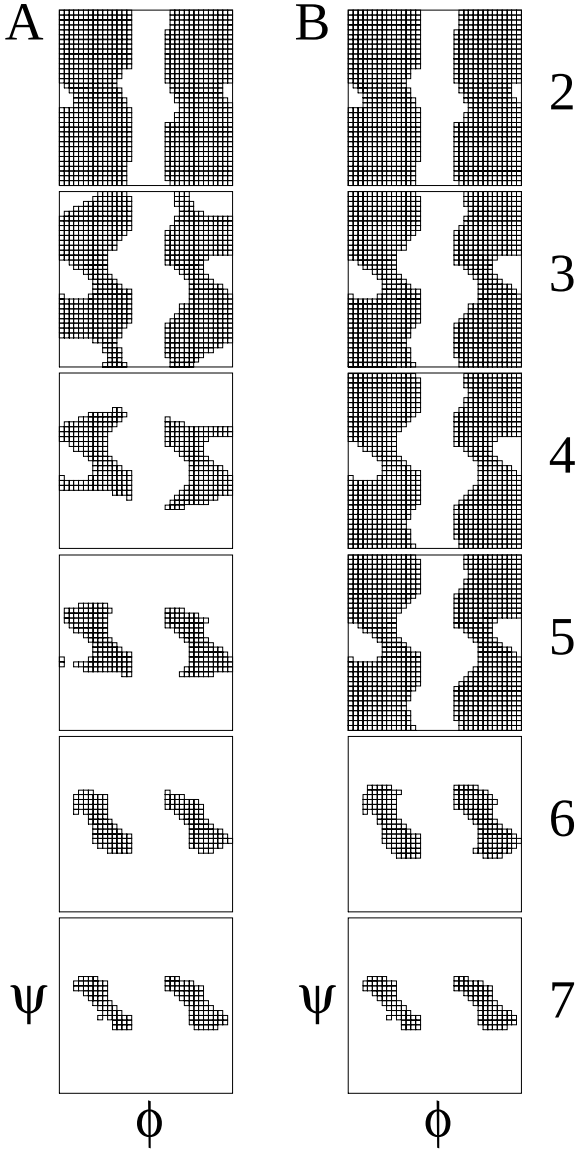


Figure 7. Allowed $\phi\Psi$ regions for residue Gly^3 after depth 2–7 in a distance-space grid search of APGA restrained by HN_4–O_1 distances where triangle inequality smoothing was applied (A) after and (B) before restraints were inserted to the distance bounds matrix.

and was used for grid search calculations performed in distance space. The progress of the grid search was followed by plotting the allowed $\phi\psi$ points for Gly^3 after completion of the conformation search for chains of length 2–7 (Figure 7). Although the starting and ending points of searches B and A are identical, the intermediate results are different. The allowed conformation space for search B remains un-

changed for chain lengths 3–5, i.e., until the N_4–O_1 bound becomes directly engaged during the search of a length-6 chain (spanning $\omega_2\phi_2\Psi_2\omega_3\phi_3\Psi_3$). An additional reduction in allowed conformation space occurs at chain length 7 ($\omega_2\phi_2\Psi_2\omega_3\phi_3\Psi_3\omega_4$) when the $HN_4 - O_1$ bound becomes directly engaged. In contrast, there is a contraction of the allowed conformation space for search A for chain lengths 3–7 caused by the distribution of bounds throughout the matrix by bounds smoothing. The result of an early reduction in the number of allowed torsion combinations is that fewer trial conformations are required for subsequent search procedures for longer chains. It has been noted elsewhere (Clark et al., 1991, 1992) that the use of smoothed distance bounds may considerably improve search efficiency during directed search of small molecules.

Allowed $\phi\Psi$ maps for residues Pro^2 and Gly^3 (Figure 8) illustrate that restrained APGA conformers fall into two distinct families for each residue. The $\Psi_2\phi_3$ projection of the four-dimensional space $\phi_2\Psi_2\phi_3\Psi_3$ connects regions in $\phi_2\Psi_2$ with regions in $\phi_3\Psi_3$ with no 'cross-over'; these regions correspond to type I and type II β-turns. A number of search calculations were carried out with different combinations of distance bounds and torsional conditions (Table 2). When both hydrogen bond and van der Waals restraints are present, there are about twice as many conformations in type II compared to type I for the APGA peptide. The ratio of type I to type II remains essentially unchanged for different torsional conditions; by contrast, a 1:1 ratio was found in unrestrained searches. Type I turns dominate a library of β-turn conformations extracted from high-resolution protein structures (Wilmot and Thornton, 1988). However, the trend is reversed and type II predominates, as observed here, when positions 2 and 3 of the turn have the sequence Pro-Gly.

Table 2. APGA grid search results

Torsion conditions[a]	Restraints[b]		Number of conformations		
	VDW	H-BOND	Total	Type I[c]	Type II[c]
A	no	no	326592	163296	163296
A	yes	no	113855	58933	54922
A	no	yes	1930	980	950
A	yes	yes	312	108	204
B	yes	yes	108	33	75
C	yes	yes	7802	2520	5282

[a] All torsion angles were searched at uniform 10° increments with the following restrictions: (A) fixed peptide torsions ω_2, ω_3, $\omega_4 = 180°$, flexible proline $\phi_2 \in [-90°, -30°]$, flexible backbone $\Psi_2, \phi_3, \Psi_3 \in [-180°, +170°]$; (B) same as (A) but with fixed proline $\phi_2 = -60°$]; (C) same as (A) with flexible peptide torsions ω_2,ω_3, $\omega_4 \in [+170°, +190°]$.

[b] Grid search calculations were performed in the presence (yes) and absence (no) of van der Waals (VDW) and hydrogen bond (H-BOND) distance restraints defined in the text.

[c] Subsets corresponding to type I and type II turns are defined by $\phi_3 < 0°$ and $\phi_3 \geq 0°$, respectively.

Table 3. Helical peptide distance restraints

Restraint category[a]	Atom pair		Distance bounds (Å)	
	Atom i	Atom j	Lower	Upper
Idealized NOE[b]	HN_i	HN_{i+1}	2.3	3.3
	$H\alpha_i$	HN_{i+1}	3.0	4.0
	$H\alpha_i$	HN_{i+3}	2.9	3.9
	$H\alpha_i$	$M\beta_{i+3}$[c]	2.5	4.4
α-Helical hydrogen bond	O_i	HN_{i+4}	1.8	2.0[d]
3_{10}-Helical hydrogen bond	O_i	HN_{i+3}	1.8	2.0[d]

[a] Restraints in addition to van der Waals radii and hydrogen bond allowances as described in the text.

[b] Idealized NOE distance bounds to unique protons (i.e. not the methyl pseudoatom) are obtained by adding and subtracting 0.5 Å from the values listed in Table 7.1 of Wüthrich (1986).

[c] Distance bounds refer to the methyl β pseudoatom positioned at the center of the Hβ atoms, and are subject to no additional correction.

[d] Upper distance bounds for simulated hydrogen bonds were varied within a range of 1.9–2.4 Å in separate conformation search calculations.

Helical peptides

Helices are elements of secondary structure with successive $\phi\Psi$ values near $(-50°,-60°)$, often characterized by HN_i–O_{i-4} or HN_i–O_{i-3} hydrogen bonds for α and 3_{10} conformations, respectively. Conformation search calculations were carried out for a terminally blocked alanine octapeptide model for several combinations of idealized NOE and hydrogen bond restraints, and torsion grid spacing for backbone torsion angles ϕ and Ψ (Table 3). The purpose of these calculations is to gain practical experience in applying conformation search to larger peptide systems and to explore methods for the analysis of large conformational ensembles.

The distance bounds matrix prepared for conformation search contained, in addition to van der Waals and covalent distances, bounds between all possible HN_i–O_{i-n} pairs spanning the central octapeptide including atoms in blocking groups (Figure 9). For an

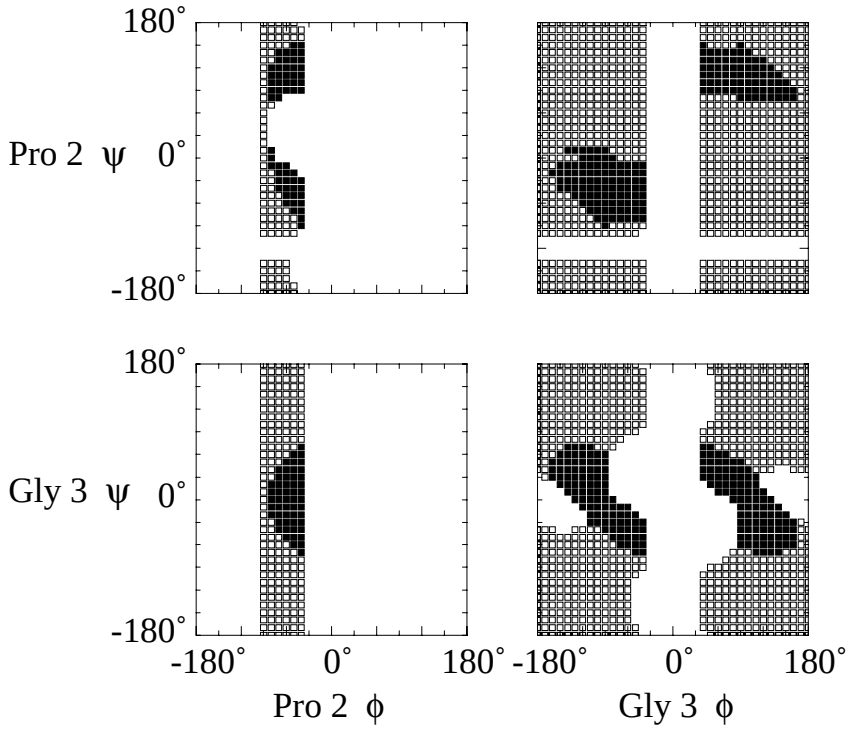


Figure 8. Two-dimensional projections of allowed ϕ and Ψ regions for residues Pro^2 and Gly^3 in the peptide APGA with van der Waals radii only (open symbols) and for HN_4–O_1 distance restraints (filled symbols) determined by distance-space grid search as described in the text. For the restrained search, the $\phi\psi$ 'Ramachandran' projections (upper left for Pro^2 and lower right for Gly^3) show two allowed conformation regions consistent with type I and type II reverse turns. The $\Psi_2\phi_3$ projection (upper right panel) demonstrates that the allowed conformational regions are distinct.

α-helix (n = 4) there are six HN–O pairs (Figure 9B), while for a 3_{10}-helix (n = 3) seven restraints are produced (Figure 9C). The upper distance bound for hydrogen bond restraints was set to 2.0 Å, and was varied from 1.9 to 2.4 Å in separate calculations. The calculations used an additional torsional restraint $\phi_4 < 0°$ and $\phi_5 < 0°$ to restrict the search to right-handed helical conformers. All remaining ϕ and Ψ torsion angles were sampled over 360°.

Conformation search calculations were carried out for the helical peptide using several different strategies. It was the goal to be able to search more than a few residues at a time that drove the development of the search algorithms from distance space to real space. Here is a brief description of the development.

Initially, conformation search calculations were attempted entirely in distance space using a grid search strategy. The distance-space algorithm stores a separate conformation grid in computer memory for every linear subchain in the system. Therefore, the number of simultaneous bond rotations, i.e., the maximum chain length, is limited by the amount of available computer memory. Counting peptide bonds (even when given a single fixed value) it was possible to allocate sufficient memory for up to 13 bond rotations out of a total of 25 ($(\omega\phi\Psi)_{1-8}\omega_9$) for the helical peptide with tight restraints, but only 9–10 bond rotations for loose restraints. The grid search was limited by the requirement to allocate storage for the product conformation space in which, for some subchains, allowed conformation regions occupied as little as 10^{-4} to 10^{-6} of the total. Although conformational grids are relatively straight forward to manipulate using projection and initialization algorithms described above, a large fraction of computer memory may be used to store disallowed points in conformation space.

A second method was devised to perform a hierarchical search similar to that described above, but using a list of conformations stored in a computer file instead of addressing memory locations. This second approach also propagated search results to higher dimensions by performing a 'build-up' of trial conformations based on overlapping submolecules similar to Algorithms A and B, and a 'build-down' of

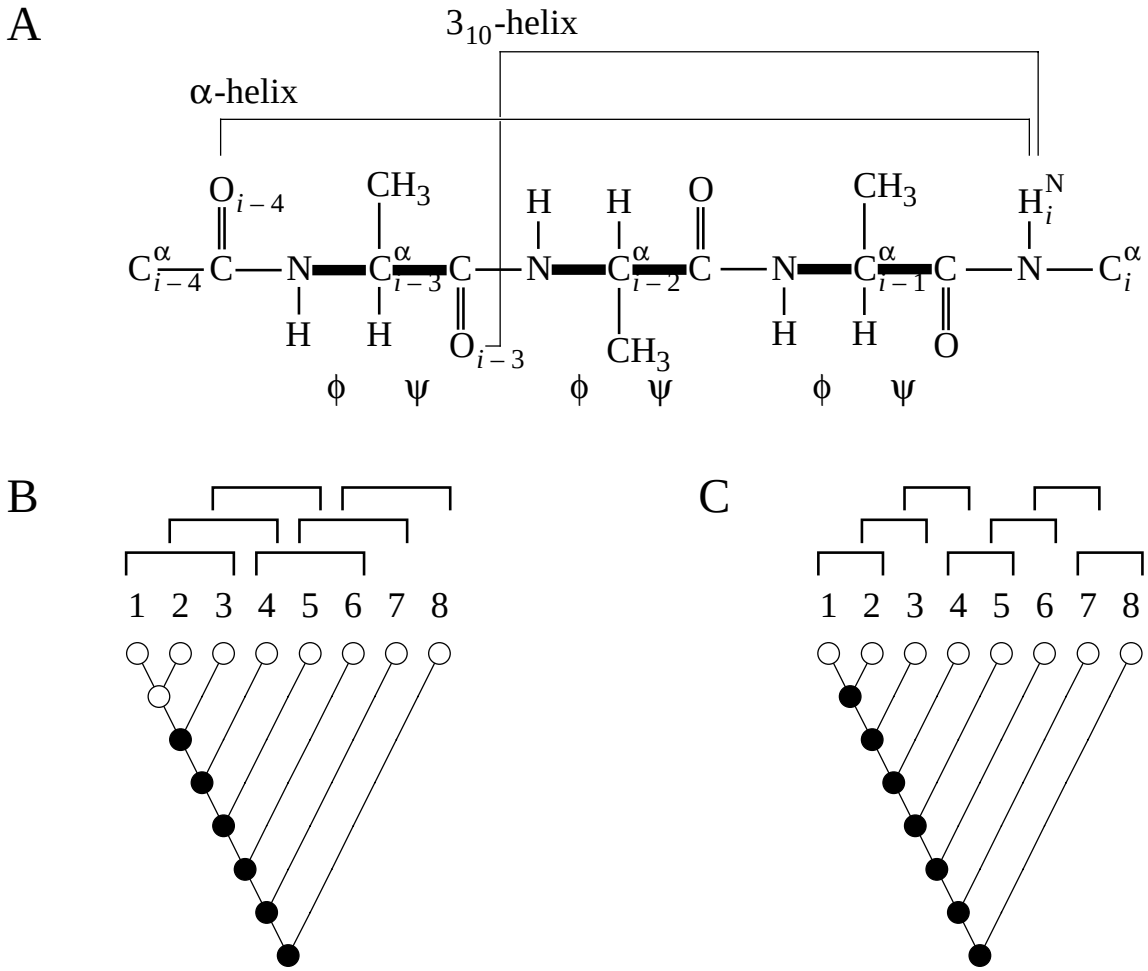


Figure 9. Schematic diagram showing restraint topology (A) and (B, C) binary tree search strategy for a blocked octapeptide restrained by 'helical' hydrogen bonds. Shown are (B) an α-helix, for which restraints $d(HN_i–O_{i-4})$ span three residues, and (C) a 3_{10}-helix, for which restraints $d(HN_i–O_{i-3})$ span two residues. Filled symbols represent submolecules for which a distance restraint is directly engaged (see also Figure 4).

searched conformations similar to Algorithm C. Using this approach it was possible to examine the conformation space for the 25-bond blocked octapeptide for a total of 13 161 unique conformations. A diagram showing the progression of the search is shown in Figure 10. For this particular demonstration, the distance restraints included idealized NOE and α-helical hydrogen bond distances (Table 3) and no torsion restraints other than fixing peptide bonds to 180°.

It was evident that the recursive initialization and projection procedures did not operate efficiently due to a high demand for file storage of intermediate results, and the relatively high cost of file read/write operations. The build-up of trial conformations based on overlapping subchains was abandoned for the file-based search, and a new approach was tried where trial conformations were produced by combining allowed conformers for adjacent (rather than overlapping) subchains using a binary tree. The initialization and projection procedures were not used in this approach, which, for the helical model, significantly increased the speed of the search calculation. An added benefit of the binary tree approach is that the search calculation can be distributed over a network in a multiprocessor environment.

For the helix, a composition order that engages distance restraints for relatively short chains could be determined by inspection of the restraint topology (Figure 9B) and an efficient search was carried out nearly entirely by composition. As described above,

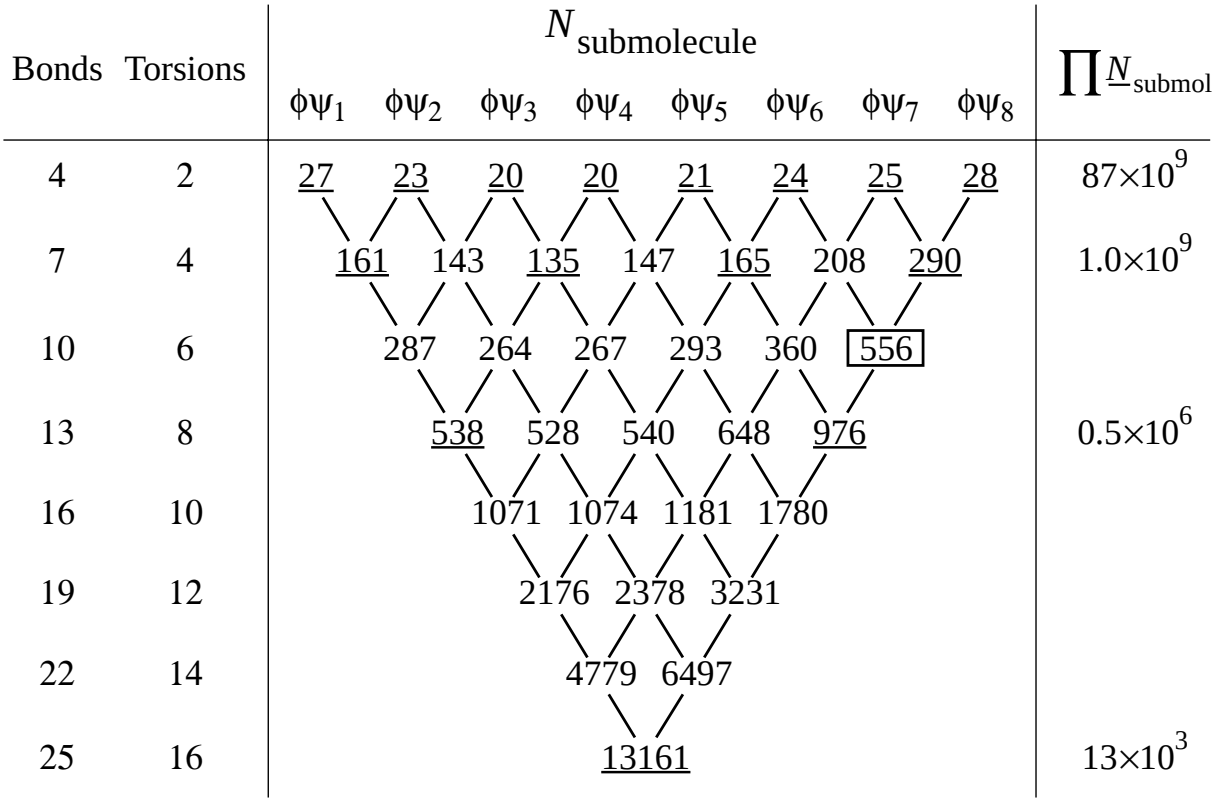


Figure 10. Grid search tree for a helical octapeptide subject to simulated hydrogen bond and NOE distance restraints. Each node in the tree represents the number of allowed conformations for the submolecule with the given number of bonds and subset of rotatable torsions (left columns) in the spanning subtree. For example (boxed) there were 556 allowed combinations after search for the six torsions $(\phi\Psi)_{6-8}$ (10 bonds including surrounding peptide bond torsions). The product of the number of conformations (right column) in adjacent submolecules (underlined) is an estimate of the number of trial conformations needed to perform a purely combinatorial search.

for a more general solution, the binary tree approach is best used in combination with a prescreening of conformation space by grid search so that local correlations in torsion space implied by the restraints are discovered early on. Table 4 summarizes the results for a series of search calculations carried out using either grid/binary tree or binary tree (alone) search strategy for different hydrogen bond arrangements and torsional increments. Simulated NOE restraints were not used for this series of calculations. All search methods (grid search, overlap build-up, and binary composition) produced identical ensembles of allowed conformations for all systems for which a direct comparison was made.

NMR-restrained conformation search of SYPFDV

The peptide SYPFDV is one of a class of turn-forming peptides that has been shown to contain significant populations of a type VI β-turn conformation

Table 4. Search results for helical peptide

Distance restraints[a]		Uniform	Number of
Topology	Upper bound HN_i to O_j	torsion increment (°)	conformations
α, i − j = 4	1.860[b]	10	7
	1.9	10	176
	2.0	10	16739
	2.1	10	189109
	2.2	10	753438
3_{10}, i − j = 3	2.0	10	4644
	2.2	10	411852
	2.4	10	7818296
α, i − j = 4	1.9	9	3057
	2.0	9	185600

[a] Hydrogen bond distance restraints as given in Table 3 were augmented by van der Waals radii and 'right-handed' ϕ restraints described in the text.

[b] Search with uniform 1.855 Å hydrogen bond restraints found no allowed conformations.

in aqueous solution in the *cis*-Pro state (Dyson et al., 1988). The peptide conformation has been characterized structurally by NMR using NOE distance restraints and scalar J-coupling constants together with distance geometry and molecular dynamics calculations (Yao et al., 1994b). In the NMR structures, the central Tyr-*cis*-Pro-Phe region adopts a well-defined conformation with the aromatic rings of Tyr^2 and Phe^4 stacked against the proline side chain. This stacking arrangement buries a large hydrophobic surface area which is thought to stabilize the turn conformation in solution (Yao et al., 1994a,b; Demchuk et al., 1997). To further characterize the solution conformation of the SYPFDV peptide, we have performed conformation search calculations on the central tripeptide, which encompasses all reported NMR-derived restraints.

A distance bounds matrix was prepared for the amino acid sequence SY*cis*PFDV with covalent distances and van der Waals radii as described. The covalent geometry of the *cis*-Pro peptide bond differs slightly from that of *trans*-Pro in the ECEPP/2 parameter set (Momany et al., 1975; Némethy et al., 1983), and was specified using an appropriate entry in the residue library. Limited flexibility for proline ring torsions and peptide bond torsions was introduced by bounds widening. The experimental distance bounds used previously to determine NMR structures (Yao et al., 1994b) were inserted into the bounds matrix. The bounds matrix was subjected to triangle inequality distance bounds smoothing (Crippen and Havel, 1988) prior to conformation search. During the search, degenerate distance restraints were used for NOE bounds to equivalent aromatic ring protons. These are satisfied if at least one of a set of equivalent distances satisfy the original bound, without pseudoatom correction, but with multiplicity correction (Fletcher et al., 1996). The resulting allowed conformation space (Figure 11) is similar to that allowed by r^{-6} average distance bounds (Gippert et al., 1990) and is significantly smaller than pseudoatom correction while still avoiding the 'assignment' of equivalent protons.

Initially all torsion angles were allowed full rotation over the complete range: 360° for ϕ, Ψ and χ^1 for Tyr^2 and Phe^4, and ψ of Pro^3; 180° for χ^2 for Tyr^2 and Phe^4 due to two-fold symmetry; and $-60 \pm 30°$ for ϕ of Pro^3; $0 \pm 30°$ for ring torsions of Pro^2; and $\pm 10°$ deviations from a planar conformation for peptide bond torsions. Distance-space grid search and real-space binary tree search calculations were carried out under different conditions described in detail below, and summarized in Table 5.

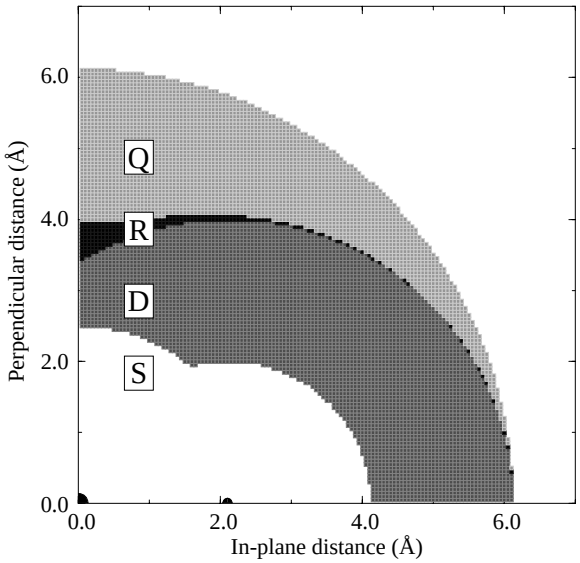


Figure 11. Allowed positions of an external proton due to an NOE distance bound involving equivalent aromatic protons: (S) sterically disallowed region; (D) degenerate distance restraint; (R) r^{-6} average distance; (Q) pseudoatom distance correction. Depicted is a 4.0 Å NOE bound. The figure is taken as a slice perpendicular to the plane of the aromatic ring through the proton positions at ($x = \pm 2.1$, $y = 0$ Å), and has mirror symmetry with respect to both x (ring plane) and y (ring perpendicular) axes.

Two distance-space searches were carried out at uniform sampling increments of 10° and 5°, search A and B, respectively. There are 18 rotatable bonds in the blocked YPF peptide system but at most 10 flexible torsion angles in the longest linear (unbranched) subchain in the system. Search A found no satisfactory conformations, with violations occurring for the five-dimensional subchain $\chi_2^2\chi_2^1\Psi_2\omega_3\chi_3^4$ due to restraints between atoms of Tyr^2 and Pro^3. An identical search was performed with a smaller torsion increment (B) and found satisfactory conformations for all subchains. Under these conditions, a single allowed point in conformation space was found for the side chain for Tyr^2, with increased flexibility progressing through the backbone of Pro^3 to the side chain of Phe^4, depicted by a projection of internal coordinates in Figure 12. Variations in flexibility between residues 2–3 and 3–4 are emphasized by counting the number of substates allowed for the subchain $\chi_2^2\chi_2^1\Psi_2\omega_3\phi_3$, for which only a single conformation was found, and the adjacent subchain $\Psi_3\omega_4\phi_4\chi_4^1\chi_4^2$, for which 27 439 conformers were found. Furthermore, the *cis*-Pro peptide bond has an ω angle of

Table 5. Search results for blocked Y*cis*PF peptide fragments

Search	Torsion increment (°)	Search protocol	Result
A	10	Distance space[a]	Failed in five-dimensional subsearch involving Tyr^2-Pro^3 restraints
B	5	Distance space[a]	Completed 10-dimensional search found single Tyr^2 conformer
C	2	Real space[b] (blocked Y*cis*P only)	Completed search found two Tyr^2 conformers

[a] Peptide bond torsions allowed to vary by ±10° from planar. Torsion angles in the proline ring were varied independently over a range ±30° from planar.
[b] Peptide bond torsions were held at a planar conformation. Proline ring torsions were correlated by pseudorotation.

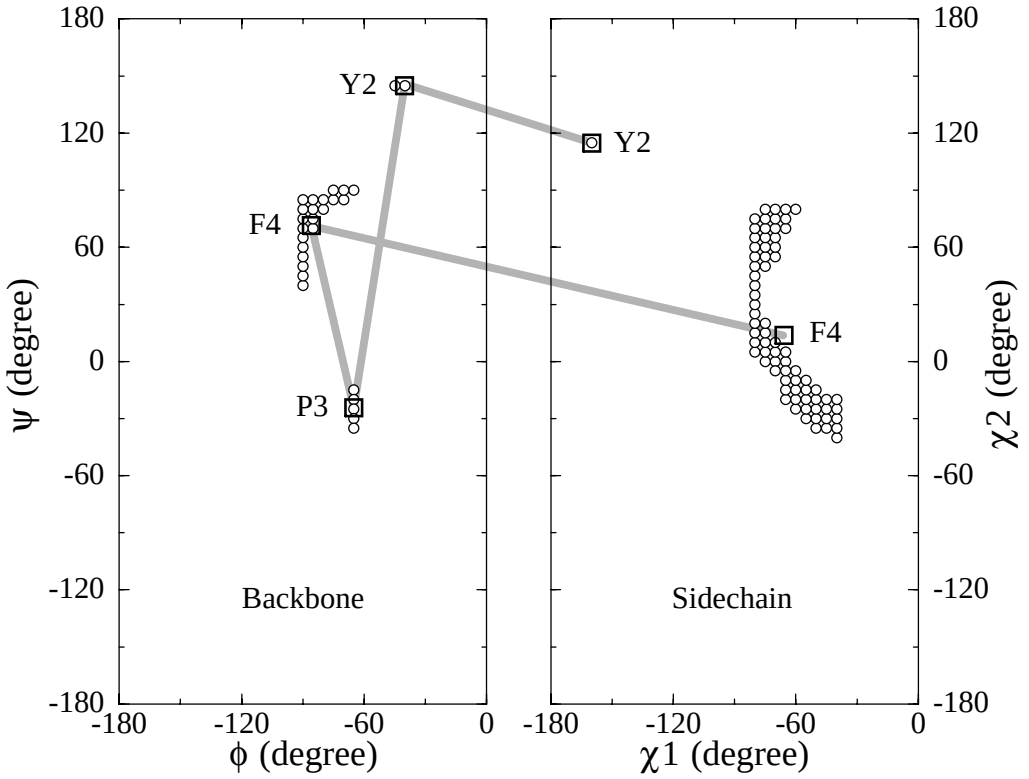


Figure 12. Two-dimensional projection of allowed ϕΨ (backbone) and $\chi^1\chi^2$ (side-chain) regions for a Y*cis*PF peptide subject to a 5° distance-space search and experimental NOE distance bounds. Average torsion angles for the ensemble (computed from the embedded average distance structure) are indicated by open boxes, and are joined by a line progressing from the side chain of Tyr^2 to the backbone of Tyr^2, Pro^3 and Phe^4, and to the side chain of Phe^4.

−10°, which is a significant distortion away from the expected planar conformation. Table 6 compares distances back-calculated from the ensemble of allowed conformations to the original NMR-derived distance restraints (Yao et al., 1994b). The Tyr^2-Pro^3 conformation is determined by van der Waals contact of the two side chains, as well as NOE-based evidence for close proximity between Tyr^2 Hϵ and Pro^3 Hα (3.5 Å, only just satisfied), Tyr^2 Hϵ and Pro^3 Hδ3 (4.0 Å, satisfied by 0.1 Å), and Tyr^2 Hβ3 and Pro^3 Hα (4.0 Å, satisfied by 0.1 Å). In contrast, two of the most easily satisfied restraints also join protons of Tyr^2 and Pro^3: between Tyr^2 Hδ and Pro^3 Hβ3 (5.0 Å, satisfied on

Table 6. NOE distances for Y*cis*PF conformation search

Distance restraints from Yao et al. (1994b)			Search results[a]		
Atom i	Atom j	Upper bound (Å)	Average distance (Å)	Number of substates	Restraint satisfied (Å)
Tyr^2Hα	Pro^3Hα	2.5	2.1	1	0.4
	Phe^4HN	3.5	3.2	29	0.3
	Asp^5HN	4.0	3.1	23106	0.9
Tyr^2Hβ3	Pro^3Hα	4.0	3.9	1	0.1
Tyr^2Hδ	Pro^3Hα	3.0	2.5,4.1[b]	1	0.5
	Pro^3Hβ3	5.0	3.2,5.9[b]	3	1.8
	Pro^3Hδ3	4.0	2.4,6.6[b]	2	1.6
Tyr^2Hϵ	Pro^3Hα	3.5	3.5,4.7[b]	1	0.0
	Pro^3Hβ3	4.0	2.9,5.7[b]	3	1.1
	Pro^3Hγ3	5.0	3.9,7.6[b]	12	1.1
	Pro^3Hδ3	4.0	3.9,7.3[b]	2	0.1
Pro^3Hα	Phe^4Hδ	5.0	4.6,7.7[b]	27439	0.4
Pro^3Hβ2	Phe^4HN	5.0	3.6	64	1.4
	Phe^4Hα	5.0	4.2	738	0.8
Pro^3Hγ2	Phe^4HN	5.0	3.3	268	1.7
	Phe^4Hα	5.0	4.7	2631	0.3
	Phe^4Hδ	4.5	2.9,6.0[b]	148902	1.6
	Phe^4Hϵ	5.0	3.5,6.3[b]	148902	1.5

[a] Results for 5° distance-space search with uniform torsion sampling.
[b] Averages of maximum and minimum distances for pairs of equivalent aromatic protons.

average by 1.8 Å) and between Tyr^2 Hδ and Pro^3 Hδ3 (4.0 Å, satisfied by 1.6 Å).

A representative subset of the allowed conformers from search B was reconstructed in three dimensions (Figure 13). Shown are 100 conformers (chosen every 44th out of 4383) superpositioned for minimum coordinate rms (Diamond, 1992) to the average distance structure (Figure 13C), obtained by embedding (Crippen and Havel, 1988) the average distance matrix collected for the entire ensemble of allowed conformers. Two different superpositions were chosen. The superposition depicted in Figure 13A was performed using all atoms defined in the linear peptide Y*cis*PF (excluding the Pro^3 side chain). This superposition masks, to some degree, the lack of disorder in the side-chain position of residue Tyr^2 by 'mixing in' conformational heterogeneity of the Phe^4 side chain. A similar effect is seen for loop regions of protein NMR structures, where, for example, β-turns appear to be disordered when the entire molecule is used for superposition, but appear well ordered when the superposition is made for atoms only within the local structure (Moore et al., 1991). Figure 13B shows the same conformers as in Figure 13A, but with a superposition determined using Cα, Cβ, Cγ, Cδ1 atoms in the side chain of Tyr^2. Variations in flexibility apparent in the plot of internal coordinates (Figure 12) are better seen in the biased overlay.

The real-space conformation search (C) with 2° uniform torsional increment succeeded in finding the same Tyr^2 side chain conformation as found for search B, as well as found a second conformation with a more parallel stacking arrangement between side chains of Tyr^2 and Pro^3 (Yao et al., 1994b).

Conclusions

We have described two complementary algorithms for producing a high-dimensional, systematic search in torsion angles for conformations of peptides and protein fragments that satisfy a distance geometry model. Satisfactory conformations obey all constraints implied by fixed bond lengths and angles, hard-sphere atomic radii, and restraints implied by bounds on interatomic distances and other functions of three-dimensional coordinates. All satisfactory conforma-

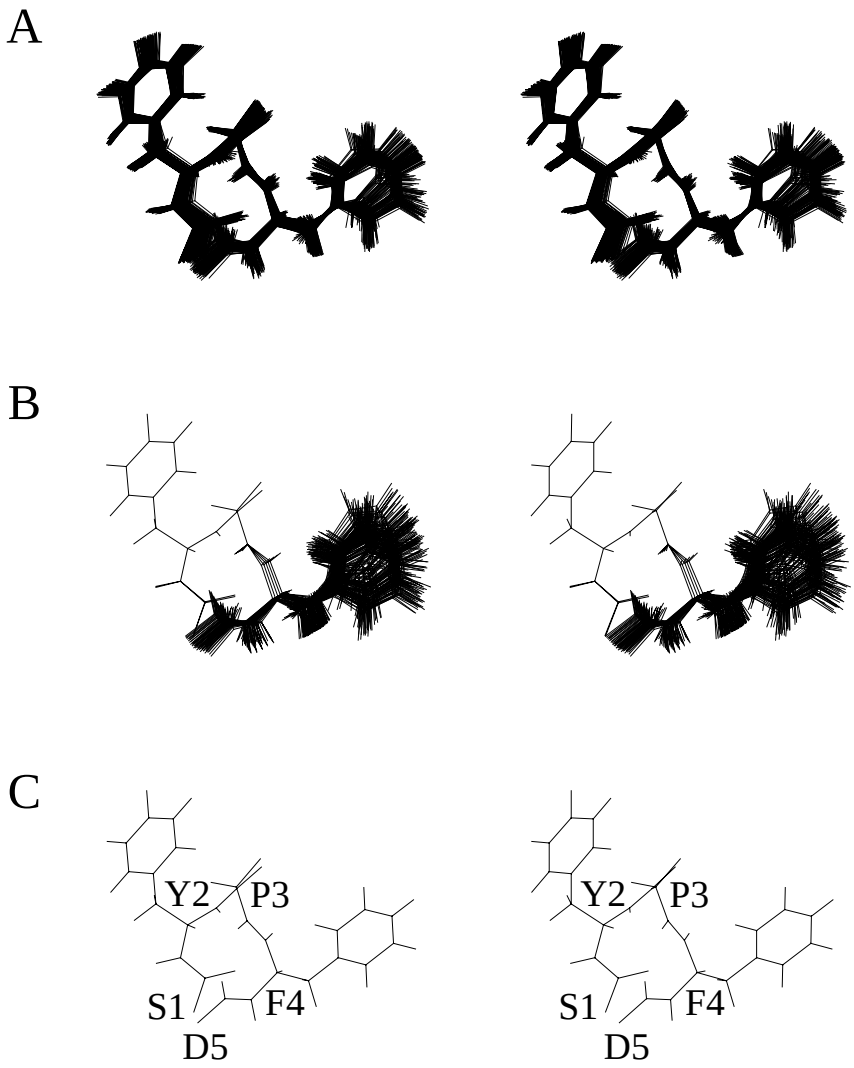


Figure 13. Stereo image of representative Y*cis*PF conformers corresponding to the allowed torsion regions shown in Figure 12: (A) all-atom superposition, and (B) Tyr^2 side-chain superposition. Shown in (C) is the average distance structure used as the target for superposition. The proline side-chain Cγ and hydrogen atoms are omitted from this figure. The view is along the Pro N to C^{α} bond. This figure was prepared using a modified version of MOLSCRIPT v.1.1 (Kraulis, 1991).

tions are found that can be represented by an initial choice of discrete torsion points for independent degrees of freedom. Alternatively, the real-space algorithm described is suitable for performing conformation search using high-dimensional, correlated points in torsion space, for example from a conformation library, for a systematic build-up approach.

We have applied these algorithms to the problem of systematic conformation search for a hydrogen-bonded β-turn peptide with sequence APGA; a right-handed, fully hydrogen-bonded α-helix; and an NMR-restrained fragment of the peptide SYPFDV whose conformation was previously determined in solution. These examples serve to illustrate the principles and limitations of systematic search. Conformational populations produced by systematic search are consistent with expected regions of conformation space determined by other methods. As discussed elsewhere (Beusen, et al., 1996), systematic search provides an alternative to traditional distance geometry approaches for the determination of biomolecular structure.

The size of conformational searches that can be carried out depends upon the number of rotatable bonds, the scanning increment, the effectiveness of

restraints in reducing conformational space, and the available computer resources. Our algorithms do not evade the 'combinatorial explosion' inherent in systematic search, but do extend the size of practical problems. We have performed calculations for isolated peptide sequences up to 13 residues, and on overlapping 3–4 residue fragments of medium-sized proteins. Most of these searches used torsion increments of 5–20°. The key size limitation generally involves the number of *allowed* conformers, since it becomes unwieldy to collect and compare more than about 10^9 conformations. (The largest set used to date contained about 235 million unique allowed conformations for an unrestrained system consisting of eight independent degrees of freedom (Bashford et al., 1997).) For macromolecules, the most common application of ideas presented here will probably involve the systematic search for allowed conformations on chain fragments such as loops and for the validation of restraints and stereospecific NMR assignments.

Acknowledgements

This work was supported by GM45813 (D.A.C.) and GM36643 (P.E.W.). We thank Mike Christensen, Jane Dyson, Thomas Macke, Jin Yao and Ping Yip for helpful discussions. G.P.G. thanks the Carlsberg Laboratory for support during the completion of this manuscript. Instructions for obtaining the DTAGS and NEWMOL programs from the World Wide Web are given at http://www.scripps.edu/~garry/programs.html.

References

Bashford, D., Case, D.A., Choi, C. and Gippert, G.P. (1997) *J. Am. Chem. Soc.*, **119**, 4964–4971.

Beusen, D.D., Iijima, H. and Marshall, G.R. (1990) *Biochem. Pharmacol.*, **40**, 173–175.

Beusen, D.D., Shands, E.F.B., Karasek, S.F., Marshall, G.R. and Dammkoehler, R.A. (1996) *J. Mol. Struct. (THEOCHEM)*, **370**, 157–171.

Billeter, M., Braun, W. and Wüthrich, K. (1982) *J. Mol. Biol.*, **155**, 321–346.

Braun, W. and Gō, N. (1985) *J. Mol. Biol.*, **186**, 611–626.

Braun, W. (1987) *Q. Rev. Biophys.*, **19**, 115–157.

Case, D.A. and Wright, P.E. (1993) In *NMR in Proteins* (Eds., Clore, G.M. and Gronenborn, A.M.), MacMillan, New York, NY, pp. 53–91.

Clark, D.E., Willett, P. and Kenny, P.W. (1991) *J. Mol. Graph.*, **9**, 157–160.

Clark, D.E., Willett, P. and Kenny, P.W. (1992) *J. Mol. Graph.*, **10**, 194–204.

Crippen, G.M. and Havel, T.F. (1988) *Distance Geometry and Molecular Conformation*, Wiley, New York, NY.

Dammkoehler, R.A., Karasek, S.F., Shands, E.F.B. and Marshall, G.R. (1989) *J. Comput.-Aided Mol. Design*, **3**, 3–21.

Demchuk, E., Bashford, D., Gippert, G. and Case, D.A. (1997) *J. Mol. Biol.*, **270**, 305–317.

Diamond, R. (1992) *Protein Sci.*, **1**, 1279–1287.

Dyson, H.J., Rance, M., Houghten, R.A., Lerner, R.A. and Wright, P.E. (1988) *J. Mol. Biol.*, **201**, 161–200.

Fletcher, C.M., Jones, D.N.M., Diamond, R. and Neuhaus, D. (1996) *J. Biomol. NMR*, **8**, 292–310.

Gippert, G.P., Yip, P.F., Wright, P.E. and Case, D.A. (1990) *Biochem. Pharmacol.*, **40**, 15–22.

Gō, N. and Scheraga, H.A. (1970) *Macromolecules*, **3**, 178–187.

Güntert, P., Braun, W., Billeter, M. and Wüthrich, K. (1989) *J. Am. Chem. Soc.*, **111**, 3997–4004.

Güntert, P. and Wüthrich, K. (1991) *J. Biomol. NMR*, **1**, 447–456.

Havel, T.F. (1991) *Prog. Biophys. Mol. Biol.*, **56**, 43–78.

Hendrickson, J.B. (1961) *J. Am. Chem. Soc.*, **83**, 4537–4547.

Iijima, H., Dunbar Jr., J.B. and Marshall, G.R. (1987) *Proteins Struct. Funct. Genet.*, **2**, 330–339.

Kraulis, P.J. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.

Leach, A.R. (1991) In *Reviews in Computational Chemistry II* (Eds., Lipkowitz, K.B. and Boyd, D.B.), VCH, New York, NY, pp. 1–55.

Momany, F.A., McGuire, R.F., Burgess, A.W. and Scheraga, H.A. (1975) *J. Phys. Chem.*, **79**, 2361–2381.

Moore, J.M., Lepre, C., Gippert, G.P., Chazin, W.J., Case, D.A. and Wright, P.E. (1991) *J. Mol. Biol.*, **221**, 533–555.

Morikis, D., Brüschweiler, R. and Wright, P.E. (1993) *J. Am. Chem. Soc.*, **115**, 6238–6246.

Némethy, G., Pottle, M.S. and Scheraga, H.A. (1983) *J. Phys. Chem.*, **87**, 1883–1887.

Nilges, M., Clore, G.M. and Gronenborn, A.M. (1990) *Biopolymers*, **29**, 813–822.

Pardi, A., Billeter, M. and Wüthrich, K. (1984) *J. Mol. Biol.*, **180**, 741–751.

Ramachandran, G.N. and Sasisekharan, V. (1968) *Adv. Protein Chem.*, **23**, 283–437.

Richardson, J.S. (1981) *Adv. Protein Chem.*, **34**, 167–339.

Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer, New York, NY.

Simon, I., Glasser, L. and Scheraga, H.A. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 3661–3665.

Venkatachalam, C.M. (1968) *Biopolymers*, **6**, 1425–1436.

Wilmot, C.M. and Thornton, J.M. (1988) *J. Mol. Biol.*, **203**, 221–232.

Wüthrich, K., Billeter, M. and Braun, W. (1983) *J. Mol. Biol.*, **169**, 949–961.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.

Yao, J., Brüschweiler, R., Dyson, H.J. and Wright, P.E. (1994a) *J. Am. Chem. Soc.*, **116**, 12051–12052.

Yao, J., Dyson, H.J. and Wright, P.E. (1994b) *J. Mol. Biol.*, **243**, 754–766.

Appendix: Algorithms

Algorithm A. (hierarchical, recursive, systematic search)

Search conformation space for linear chain (a, N). Search is first applied to constituent, lower dimensional subchains, and the results are used to prune the search tree.

```
A1    for (n ← 1 . . . N)
A2        for (a ← 1 . . . N − n + 1){
A3            x⃗_a ← {0, 0, 0}
A4            x⃗_{a+1} ← {d_a, 0, 0}
A5            x⃗_{a+2} ← {d_{a+1} − d_a cosθ_a, d_a sinθ_a, 0}
A6            call initialize(a, n)
A7            call search(a, 1, n)
A8            call projection(a, n)
A9        }

A10   subroutine search(a, j, n){
A11       i ← a + j − 1
A12       for (ω_i ← {ω_{i1}, . . . , ω_{iM_i}})
A13       if (C_j^{ω_a,...,ω_i} = 1){
A14           x⃗_{i+3} ← p⃗_{i+2} + T̃_{i+1} R̃_i x⃗_{i+2}
A15           if (j < n)
A16               call search(a, j + 1, n)
A17           else {
A18               r ← sqrt(x²_{i+3} + y²_{i+3} + z²_{i+3})
A19               if ((r > upper_bound) or (r < lower_bound)){
A20                   print 'Conformer ω_a, . . . , ω_i is disallowed.'
A21                   C_j^{(ω_a,...,ω_i)} ← 0
A22               }
A23               else
A24                   print 'Conformer ω_a, . . . , ω_i is allowed.'
A25           }
A26       }
A27   }
```

Algorithm B. (recursive initialization)

Initialize high-dimensional conformational array from constituent, lower dimensional arrays. Initialisation is first applied to constituents.

```
B1    subroutine initialize(a, n){
B2        if (n = 1) return
B3        call initialize(a, n − 1)
B4        call initialize(a + 1, n − 1)
```

Algorithm B. (Continued)

```
B5          call recurs_init(a, 1, n)
B6      }
Initialize C_j^(ω_a,...,ω_i) from left subspace C_{j−1}^(ω_a,...,ω_{i−1}) and right subspace C_{j−1}^(ω_{a+1},...,ω_i).
B7      subroutine recurs_init(a, j, n){
B8          i ← a + j − 1
B9          for (ω_i ← {ω_i1, . . . , ω_iM_i}){
B10             if (j < n)
B11                 call recurs_init(a, j + 1, n)
B12             else
B13             if ((C_{j−1}^(ω_a,...,ω_{i−1}) = 0) or C_{j−1}^(ω_{a+1},...,ω_i) = 0))
B14                 C_j^(ω_a,...,ω_i) ← 0
B15         }
B16     }
```

Algorithm C. (recursive projection)

Project high-dimensional conformation array onto constituent, lower dimensional arrays. Projection is subsequently applied to constituents.

```
C1      subroutine projection(a, n){
C2          if (n = 1) return
C3          call recurs_left_proj(a, 1, n − 1)
C4          call recurs_right_proj(a + 1, 1, n − 1)
C5          call projection(a, n − 1)
C6          call projection(a + 1, n − 1)
C7      }
Project C_{j+1}^(ω_a,...,ω_{i+1}) onto left subspace C_j^(ω_a,...,ω_i).
C8      subroutine recurs_left_proj(a, j, n){
C9          i ← a + j − 1
C10         for (ω_i ← {ω_i1, . . . , ω_iM_i}){
C11             if (j < n)
C12                 call recurs_left_proj(a, j + 1, n)
C13             else if (C_j^(ω_a,...,ω_i) = 1){
C14                 ξ ← 0
C15                 for (ω_{i+1} ← {ω_(i+1)1, . . . , ω_(i+1)M_{i+1}})
C16                     ξ ← ξ ∨ C_{j+1}^(ω_a,...,ω_{i+1})
C17                 C_j^(ω_a,...,ω_i) ← ξ
C18             }
C19         }
C20     }
```

Algorithm C. (Continued)

Project $C_{j+1}^{(\omega_{a-1},\ldots,\omega_i)}$ onto right subspace $C_j^{(\omega_a,\ldots,\omega_i)}$.

```
C21     subroutine recurs_right_proj(a, j, n){
C22             i ← a + j − 1
C23             for (ω_i ← {ω_i1, . . . , ω_iM_i}){
C24                     if (j < n)
C25                             call recurs_right_proj(a, j + 1, n)
C26                     else if (C_j^(ω_a,...,ω_i) = 1){
C27                             ξ ← 0
C28                             for (ω_a−1 ← {ω_(a−1)1, . . . , ω_(a−1)M_a−1})
C29                                     ξ ← ξ ∨ C_j+1^(ω_a−1,...,ω_i)
C30                             C_j^(ω_a,...,ω_i) ← ξ
C31                     }
C32             }
C33     }
```