Kinematic Manipulation of Molecular Chains Subject to Rigid Constraints

Dinesh Manocha * **Yunshan Zhu** †
Computer Science Department
University of North Carolina
Chapel Hill, NC 27599-3175
{manocha,zhu}@cs.unc.edu

Abstract

We present algorithms for kinematic manipulation of molecular chains subject to fixed bond lengths and bond angles. They are useful for calculating conformations of a molecule subject to geometric constraints, such as those derived from two-dimensional NMR experiments. Other applications include searching out the full range of conformations available to a molecule such as cyclic configurations. We make use of results from *robot kinematics* and recently developed algorithms for *solving polynomial systems*. In particular, we model the molecule as a serial chain using the Denavit-Hartenberg formulation and reduce these problems to inverse kinematics of a serial chain. We also highlight the relationship between *molecular embedding* problems and inverse kinematics. As compared to earlier methods, the main advantages of the kinematic formulation are its *generality* to all molecular chains without any restrictions on the geometry and *efficiency* in terms of performance. The algorithms give us *real time* performance (order of tens of milliseconds) on smaller chains and are applicable to all chains.

Keywords: kinematics, serial chains, loop closure, conformation search, molecular embedding, model building

Introduction

One of the fundamental problem in computational chemistry is the quantitative construction of realistic 3-dimensional structural models of molecules of interest. This includes searching for full range of conformations available to a given molecule and computation of molecular embeddings satisfying some geometric constraints. In this paper we restrict ourselves to generation of molecular conformations and embeddings, given fixed values for bond lengths and vicinal bond angles, but certain dihedral or torsional angles are allowed free rotations. In reality bond lengths and angles are stiff but somewhat flexible, and the physically important conformations are those having low internal energy. The rigid geometry assumption greatly simplifies the geometric complexity. Moreover, the minimal energy conformations found using rigid geometry formulation are a good guess to the actual minima and can be used along with a perturbation treatment for the energy minimization (Go & Scherga 1970).

Many interactive systems for molecular modeling such as *Sybyl* (Tripos 1988) and *Insight* (Biosym 1991) provide capabilities for building and changing molecular models by rotating the torsional angles. The full conformation space of acyclic molecules is the Cartesian product of all the torsional angle with one-dimensional cycles, known as the *n-dimensional toroidal manifold*, where n is the number of rotatable bonds. There is a considerable amount of literature on exhaustive methods. A recent survey of different methods is given in (Howard & Kollman 1988). The complexity of the search methods is an exponential function of the number of rotatable joints and therefore limited to chains with low n. For cyclic molecules having rotatable bonds in the rings, six degrees of freedom are lost due to the closure constraint. Go and Scheraga have proposed solutions to computing the ring conformations (Go & Scherga 1970; 1973). The algorithm in (Go & Scherga 1970) is restricted to molecular chains where the rotation can take place about all backbone bonds and the methods in (Go & Scherga 1973) are for symmetric chains only. They reduce the problem to computing roots of a univariate polynomial. Algorithms based on *distance geometry* to study the ring systems are proposed in (Weiner et. al. 1983; Peishoff & Dixon 1992). They use a distance matrix to generate a set of cartesian coordinates that satisfies the original set of distances as much as possible. A similar approach based on successive infinitesimal rotations is presented by (Sklenar, Lavery, & Pullman 1986). Although these approaches make no assumption about the geometry of molecules, they are relatively slow in practice, may

*Supported in part by Junior Faculty Award, University Research Award, NSF Grant CCR-9319957, ARPA Contract DABT63-93-C-0 048 and NSF/ARPA Science and Technology Center for Computer Graphics and Scientific Visualization, NSF Prime Contract Number 8920219

†Supported in part by NIH National Center for Researc h Resources grant RR-02170

fail to converge and cannot be used for analytic computation of the conformation space for even small chains. More recently, Crippen has introduced the technique of *linear embedding* and applied it to explore the conformation space of cycloalkanes (Crippen 1989; 1992). It is a variant on the usual distance geometry methods and makes use of the metric matrix as opposed to distance matrix. However, the application to ring structures has been limited to molecular chains, where all bond lengths and bond angles are equal. Other approaches for molecular conformations are based on stochastic optimization and Monte Carlo analysis (Go & Scherga 1978; Saunders 1989; Bruccoleri & Karplus 1985). These can be slow in practice and are not guaranteed to find all the solutions.

In this paper we present algorithms for generating the conformation space of molecular chains using kinematic analysis. We model the molecule as a serial kinematic chain using Denavit-Hartenberg formulation and show how problems of chain closure, loop deformations and conformation analysis relate to *inverse kinematics* of serial chains and robot manipulators. The problem of inverse kinematics is well-studied in robotics literature and reduces to solving system of polynomial equations (Spong & Vidyasagar 1989; Raghavan & Roth 1989). We make use of fast solutions to solve these polynomial systems (Manocha 1992; Manocha, Zhu, & Wright 1994) and apply them to study the configurations of cyclooctanes. We also show the application of inverse kinematics to *molecular embedding* problems. That corresponds to finding one or more sets of atomic coordinates such that a given list of geometric constraints is satisfied. This includes the distance constraints from two-dimensional NMR experiments and homology based modeling. The resulting algorithms are also useful for structure prediction, protein folding, model building, 3D molecular matching and drug design (G.R. Marshall et. al. 1979).

The rest of the paper is organized in the following manner. We introduce the inverse kinematics formulation in Section 2 and reduce the problems of molecular conformation to solving system of polynomial equations. In Section 3, we apply the kinematic formulation to compute molecular embeddings. We highlight its application to cyclooctanes and peptide units in Section 4.

Kinematic Formulation of Molecular Chains

Kinematics is the science of motion that treats motion without regard to the forces that cause it. In our context, kinematics for molecular chains refer to its geometrical properties in terms of the position and orientation of the atoms. In order to deal with the complex geometry of the molecules, we affix frames to the various part of a chain and then describe the relationship between these chains. In particular, we use

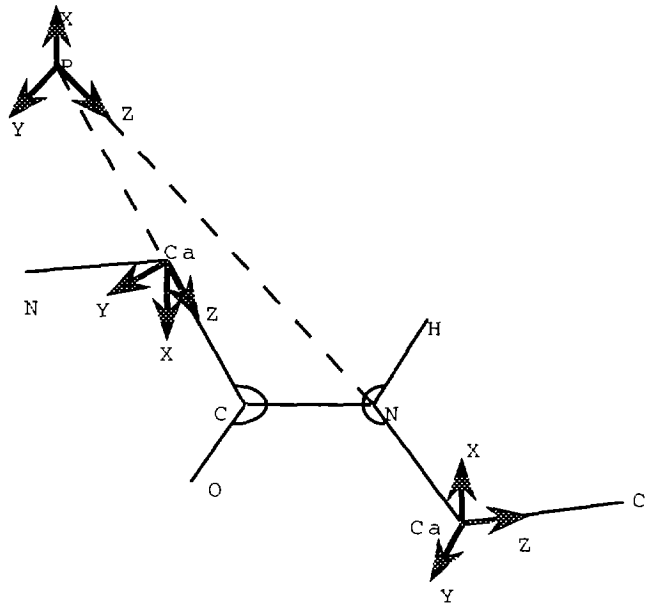


Figure 1: Coordinate systems on a polypeptide units based on DH formalism

the Denavit–Hartenberg (DH) formalism, (Denavit & Hartenberg 1955), to model a molecule with n rotatable bonds. These rotatable bonds need not correspond to a backbone chain. Each bond is represented by the line along its bond axis and the common normal to the last bond axis. In the case of parallel bonds, any of the common normals can be chosen. A coordinate system is attached to each bond for describing the relative arrangements among the various bonds. The z-axis of frame i, called z_i, is coincident with the bond axis i. Initially a base frame is chosen such that the origin is on the z_1 axis and the x_1, y_1 axis are chosen conveniently to form a right hand frame. The origin of frame i, o_i corresponds to the point where the common normal to z_i and z_{i-1} intersection z_i. If z_i and z_{i-1} intersect, the origin z_i is their point of intersection. x_i is along the common normal between z_i and z_{i-1} through o_i, or in the direction normal to the $z_{i-1} - z_i$ plane if z_{i-1} and z_i intersect. Given x_i and z_i, y_i is chosen to complete the right hand coordinate system at o_i. This formulation is shown for a peptide unit in Fig. 1. This is repeated for $i = 2, \ldots, n$. Given a coordinate system, we create the following parameters for the molecular chain:

a_i = distance along x_i from o_i to the intersection of the x_i and z_{i-1} axes.

d_i = distance along z_{i-1} from o_{i-1} to the intersection of the x_i and z_{i-1} axes.

α_i = the angle between z_{i-1} and z_i measured along x_i.

θ_i = the angle between x_{i-1} and x_i measured about z_{i-1}.

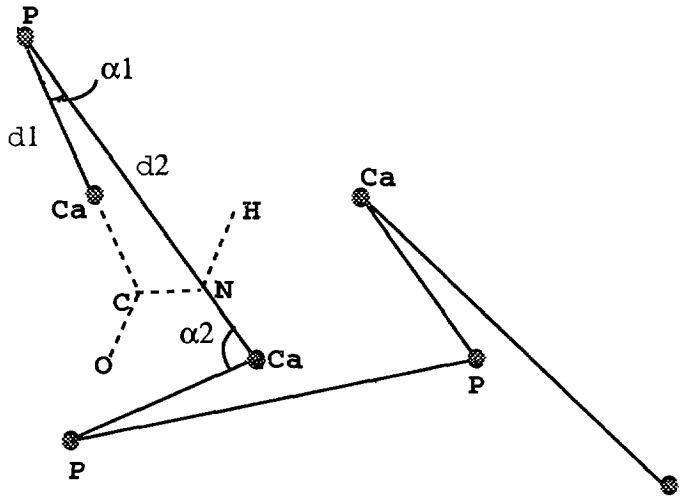


Figure 2: DH parameters for a polypeptide unit

These parameters for the polypeptide unit are shown in Fig. 2. The torsional angles correspond to the θ_is. Given this representation of the coordinate systems for each bond, the 4×4 transformation matrix relating $i-1$ coordinate system to i coordinate system is:

$$\mathbf{A}_i = \begin{pmatrix} c_i & -s_i\lambda_i & s_i\mu_i & a_ic_i \\ s_i & c_i\lambda_i & -c_i\mu_i & a_is_i \\ 0 & \mu_i & \lambda_i & d_i \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad (1)$$

where $s_i = \sin\theta_i$, $c_i = \cos\theta_i$, $\mu_i = \sin\alpha_i$, $\lambda_i = \cos\alpha_i$.

Molecular models also contain attractions and repulsions between non-bonded atoms, charge and hydrophobicity. Attractions hold nearby atoms together, whereas repulsions maintain a minimal separation defined by the atoms' electron shells. In this paper we limit ourselves to building and changing molecules by rotating around the torsional angles for adjusting the interior segments or conformation search. After computing the configuration space of a molecular chain we can apply model attractions, repulsions and charges.

Direct and Inverse Kinematics

We denote the pose of the end of the chain with respect to the base frame as A_E, a 4×4 matrix. Given the n torsional angles, $\theta_1, \ldots, \theta_n$, the pose of the end of the chain corresponds to

$$A_E = A_1 A_2 \ldots A_n. \qquad (2)$$

The computation of A_E from torsional angles is referred to as *direct kinematics*. In most applications, we are given the pose matrix and the problem of *inverse kinematics* corresponds to computing the torsional angles such that (2) is satisfied. The relationship is reduced to a system of algebraic equations by substituting $x_i = \tan\frac{\theta_i}{2}$. Therefore, $\sin\theta_i = \frac{2x_i}{1+x_i^2}$ and $\cos\theta_i = \frac{1-x_i^2}{1+x_i^2}$. Eventually we obtain a system of 6 algebraic equations in n unknowns.

Molecular Conformations and Inverse Kinematics

Many problems related to generation of molecular conformations correspond directly to inverse kinematics. In this section, we show the equivalence for local deformations and chain closures.

Chain Closure The chain closure problem turns out to be a special case of the inverse kinematics problem highlighted above. In particular for chain closure, the right hand side matrix A_E corresponds to the identity matrix. As a result, all the solutions to the chain closure are obtained by substituting for A_E in (2).

Local Deformations The problem of performing conformational variations on a local portion of an acyclic or cyclic chain corresponds to selecting the bonds formulating the chain, choosing the subset of rotatable bonds and the conformation of the end of the local chain. The algorithm proceeds by assigning the frames and computing the DH parameters for the local chain. The rotatable bonds need not be contiguous. The resulting problem can be posed as shown in (2). A_E corresponds to the conformation of the end of the chain and n corresponds to the number of rotatable in the selected chain.

Fast Algorithms for Inverse Kinematics

Inverse kinematics for general serial manipulators has been a fundamental problem for computer controlled robots. Most of the literature in robotics has concentrated on inverse kinematics of chains with six or fewer joints. The complexity of inverse kinematics of a general six jointed chain is a function of the geometry of the chain. While the solution can be expressed in closed form for a variety of special cases, such as when three consecutive joint axes intersect in a common point, no such formulation is known for the general case. It is not clear whether the solutions for such a manipulator can be expressed in closed form. Iterative solutions (based on numerical techniques like optimization or Newton's method) to the inverse kinematics for general $6R$ manipulators have been known for quite some time. However, they suffer from two drawbacks. Firstly they are slow for practical applications and secondly they are unable to compute all the solutions. As a result, most industrial manipulators are designed sufficiently simply so that a closed form exists.

Recently fast algorithms for solving polynomial systems (Manocha 1992) have been applied to the inverse kinematics of serial chains with six or fewer rotatable joints. They are based on the algebraic formulation of the problem and make no assumptions on its geometry (Manocha, Zhu, & Wright 1994). As a result,

they are applicable to all molecular chains. The algorithms make use of the algebraic reduction of the problem to a univariate system presented by Raghavan and Roth (Raghavan & Roth 1989). However the symbolic derivation highlighted by Raghavan and Roth does not take care of all the geometries (e.g. the polypeptide configurations). As opposed to reducing the problem to finding roots of a univariate polynomial, we formulate it as an eigenvalue problem. Depending on the geometry of the molecular chain we obtain matrices of different order, though the total number of solutions is still bounded by 16 (assuming a finite number of configurations). The main advantages of the matrix formulation are in efficiency and accuracy. The reduction to a univariate polynomial involves expanding a symbolic determinant, which is relatively expensive. Moreover the problem of finding roots of polynomial of degree 16 can be numerically ill-conditioned (Wilkinson 1959). As a result, we may not be able to compute accurate results using IEEE double precision arithmetic on current workstations.

Performance

The algorithm for solving inverse kinematics works very well in practice and on the applications highlighted above. It takes Denavit Hartenberg parameters and the pose of end effector as inputs, and outputs all possible solutions for the torsional angles. The algorithm involves symbolic preprocessing, numerical substitution and matrix computation. The symbolic preprocessing is performed using MAPLE computer algebra system, and the algorithm has been implemented in C making use of Fortran implementations of matrix computations from LAPACK (Manocha, Zhu, & Wright 1994).

The Denavit Hartenberg parameters and end poses of a molecular chain can be computed from the atom coordinates if at least one set of these atom coordinates is available. They can also be computed from standard geometry. Once these parameters are computed, they are substituted into the inverse kinematics procedure. For most general cases, the problem is reduced to solving eigenvalues of a 24*24 matrix, which takes no more than 20ms. For the examples of cyclohexane and protein embedding, eigenvalues of 32*32 matrices are computed because of the degeneracy of these structures. The whole inverse kinematics procedure takes 40-50 ms on an SGI/ONYX workstation.

Molecular Embedding and Inverse Kinematics

The molecular embedding problem consists of finding one or more sets of atomic coordinates such that a given list of geometric constraints is satisfied. Such problems arise, when we are given data derived from NMR experiments, where we are given some *experimental constraints* corresponding to the distance between some atoms. In these cases it is assumed that the molecule is under no great strain, so that all bond lengths and bond angles are known from standard values taken from X-ray crystallographic studies on small molecules (Crippen 1989). These constraints are referred to as *holonomic constraints*. The goal of the embedding algorithms is to find all conformations which satisfy these constraints. In this section, we show the equivalence between the embedding problems and inverse kinematics. Furthermore, fast algorithms for inverse kinematics can be used for computing the embeddings.

There are many approaches to computing the molecular embeddings and most general one is based on EMBED (Crippen & Havel ; Crippen 1981). It is frequently used for the determination of conformations of small proteins in solution by NMR. Given the experimental and holonomic constraints, the algorithm finds upper and lower bounds for all distances (which satisfy the constraints) and using some random values in this range it computes the closest corresponding three-dimensional metric matrix. Crippen has extended the algorithm using linearized embedding and taking chirality into account (Crippen 1989). The overall process is iterative and may take considerable running time on some cases. Furthermore, even for small chains it is not guaranteed to compute all the configurations.

We show the equivalence between molecular embedding and inverse kinematics for small chains. It can be combined with model building techniques for application to larger chains as shown in (Leach & Smellie 1992).

Application to Small Chains

In this section we will consider embeddings of molecular chains consisting of up to six rotatable bonds. In the basic version of the problem we are given an initial bond, specified by fixed atom positions, P_1 and P_2, and a final bond specified by fixed atom positions P_6 and P_7. We are interested in finding all the positions of all of the intermediate atoms P_3, P_4 and P_5 such that the bond lengths are all as required (d_2, d_3, d_4 and d_5) and the angles between the bonds are all as required ($\alpha_2, \alpha_3, \alpha_4, \alpha_5$ and α_6). The problem formulation is shown in Fig. 3. We can reduce this problem to inverse kinematics in the following manner:

Assign a frame using the DH formulation at each atom (shown as P_i in Fig. 3), such that the z-axis is along the bonds. We know the position of P_7 in the base coordinate system assigned at P_1. Let the orientation of the frame at P_7 be such that the z-axis is along the bond $P_6 - P_7$ and the x and y axes are chosen to complete a right hand coordinate system. A_E is therefore computed appropriately. The Denavit-Hartenberg parameters are computed using the d_i's and α_i's. The inverse kinematics problem corresponds to computing all the torsional angles, given the pose of the end-effector A_E. Given the torsional angles, the positions of the intermediate atoms can be easily computed.

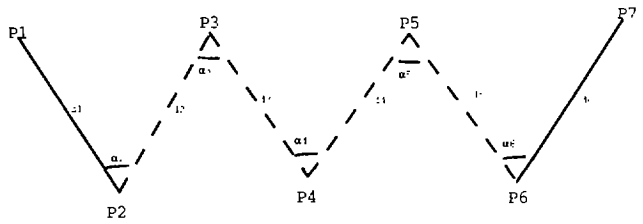


Figure 3: Molecular embedding problem with unknown atom positions

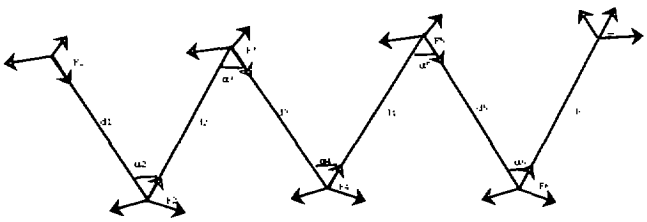


Figure 4: Inverse kinematics formulation of molecular embedding

Theorem 0.1 *Assuming fixed bond lengths and bond angles, the conformation of a molecular chain with up to 7 atoms can be determined its first and last bond positions.*

Proof

We prove the theorem by showing that the molecular chains in Fig. 3 and Fig. 4 have equivalent solution space. Therefore, the conformation of the molecular chain (in Fig. 3 represented in terms of unknown atom positions) can be computed using the algorithm for inverse kinematics highlighted in the previous sections.

We initially show that for each solution to the embedding problem, there exists a corresponding solution to the inverse kinematics problem. Given atom positions P_3,P_4, P_5 that satisfy the bond length and bond angles constraints, the x-axis of local frame at P_i can be computed based on the atom positions. Given the frames and their positions in the global coordinate system, the θ_is are easily computed and they correspond to the solutions of the inverse kinematics problem.

Conversely, if θ_is are the solutions for the inverse kinematics problem, using forward kinematics the positions of points P_2, P_3, ..., P_6 are computed. Given the position and orientation of the frames, we know that P_1 and P_7 correspond to the given positions, and z_0 is along P_1-P_2 bond, and z_7 is along P_6-P_7. Since all rotations are along the bonds, the position of P_2 and P_6 are invariant, and they match the specified positions. The rotations along the bonds do not violate any bond length or bond angle constraints.

Q.E.D.

Although we have considered molecular chains with rotation along all backbone joints, the relationship between embedding problems and kinematics problems extends to all other geometries (like peptide chains) as well.

Applications and Performance

In this section, we describe the applications of inverse kinematics algorithm to larger chains, illustrate the algorithm on loop conformations of cyclooctanes and molecular embedding of protein chains.

Extension to All Molecular Chains

An algorithm for kinematic manipulation of molecular chains with six or fewer dihedral angles is presened in (Manocha, Zhu, & Wright 1994). In this section, we highlight techniques for manipulation of molecular chains with more than six rotatable angles. Given a molecular chain with n torsional angles, the problem of kinematic manipulation reduces to 6 equations in n unknowns. Chains with more than six torsional angles are commonly used in molecular modeling. Examples include cyclo-heptanes, cyclo-nonanes, sugar molecules besides the polypeptide units. Given a chain with $n > 6$ torsional angles, it has $n - 6$ dimensional solution space in the neighborhood of any real solution. A simple strategy to solve for the solutions involves the use of $n - 6$ torsional angles as independent variables and the rest of the six are functions of the independent variables computed using the inverse kinematics algorithm. A simple exhaustive procedure assigns some discrete values to the $n - 6$ independent variables and solves for the rest of the torsional angles based on the algorithm highlighted in the previous section. Values of the independent variables are chosen using exhaustive search methods or randomized techniques. In case there are no real solutions, all the eigenvalues of the matrices formulated in the previous section have imaginary parts. We can use the magnitude of the imaginary part of the eigenvalues in choosing the independent variables and perturbing them to guide to a real solution.

The resulting algorithm combines the analytic approach with exhaustive search or randomized search methods. As compared to purely exhaustive methods, the complexity of the search space goes down by six dimensions, due to the inverse kinematics procedure. For example, to compute all the chain closure configurations of cyclo-nonanes, we only use discrete values of three independent variables as opposed to all the nine independent variables. For torsional angle increments of 60°, we generate 216 configurations of the three independent variables and apply the inverse kinematics procedure. A purely exhaustive method would generate 216^3 configurations. Similarly it can be combined with Monte Carlo methods, we only need to introduce random changes to $n - 6$ torsional angles as opposed to all the n angles.

0.0	0	4	4	6	0	0	0	0	0	6	4	4
30.0	4	4	6	0	0	0	0	0	4	6	3	4
60.0	4	6	0	0	0	0	0	0	6	4	4	4
90.0	6	6	0	0	0	0	0	0	4	8	4	5
120.0	0	0	0	0	0	0	0	0	0	4	6	4
150.0	0	0	0	0	0	0	0	0	0	0	0	0
180.0	0	0	0	0	0	0	0	0	0	0	0	0
210.0	0	0	0	0	0	0	0	0	0	0	0	0
240.0	0	4	6	4	0	0	0	0	0	0	0	0
270.0	6	6	4	8	4	0	0	0	0	0	0	0
300.0	4	4	4	4	6	0	0	0	0	0	0	6
330.0	4	4	3	6	4	0	0	0	0	0	6	4

Table 1: Number of solutions for given values of θ_7 and θ_8. The rows are for different θ_8 values, and the columns are for different θ_7 values.

Cyclooctane Conformations

We have applied inverse kinematics analysis to study the conformations of ring structures of cyclooctanes. The Denavit-Hartenberg parameters of a cyclooctane structure are shown in Fig. 5 ($d = 1.55A^o$ and $\alpha = 115^o$. The pose of end effector in this case is an identity matrix because of the ring closure.

In the case of cyclooctane in Fig. 5. there are eight rotatable bonds. The conformation can not be determined solely based on ring closure property which gives no more than six constraints. In other words, the solution space is a two-dimensional space. In our study, we combined inverse kinematics algorithm along with systematic search. Namely, we assign discrete values to the last two torsional angles and use inverse kinematics to compute the rest of the six torsional angles. We have used different increments of the angles. At 30 degree increments for θ_7 and θ_8, the number of solutions are shown in Table 1. For example, if we fix $\theta_7 = 60^o$ and $\theta_8 = 30^o$, there are 6 solutions. Due to the inherent symmetry in the problem, the resulting two-dimensional table should be symmetric. In general, the algorithm can compute the solutions to a good accuracy. However, near higher multiplicity roots (say double roots), the problem can be ill-conditioned and special processing is needed to compute these high multiplicity solutions, as highlighted in (Manocha 1992).

The performance of the inverse kinematics algorithm for conformation search of a cyclooctane has been highlighted in Table 2. In particular, the running time is shown for different grid size increments. It takes

Grid size increments	30	20	10	5
Time (sec)	7.81	18.12	73.4	295.2

Table 2: Running time of cyclooctane conformation search at various grid sizes

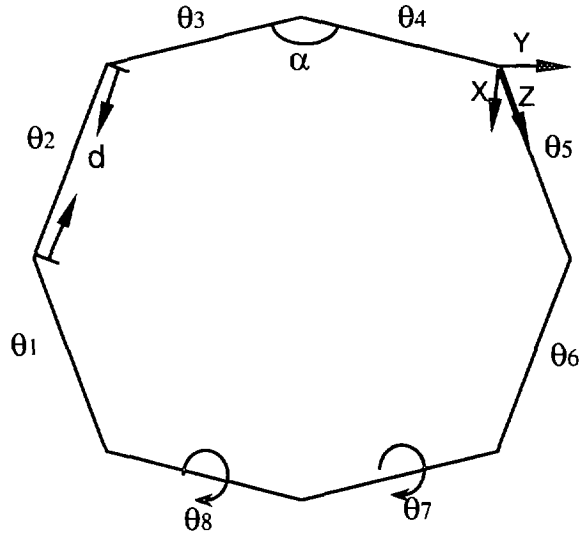


Figure 5: Cyclooctane Conformation : systematic search on θ_7 and θ_8

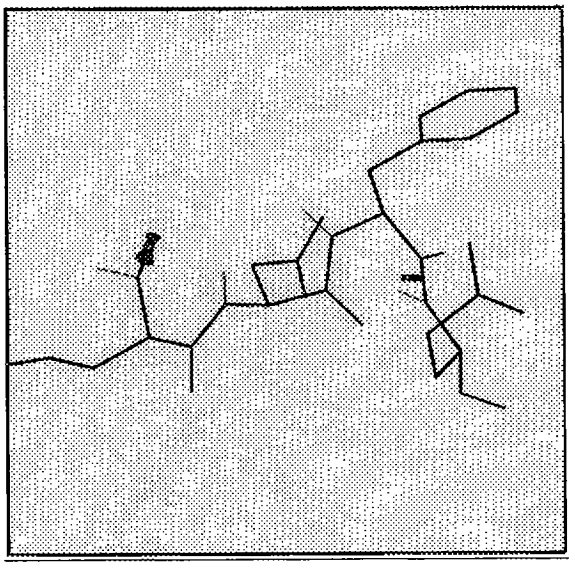
about $40 - 50$ milliseconds for a single execution of the inverse kinematics procedure and about $7 - 8$ seconds to search the two dimensional space of θ_8 and θ_7 at 30^o increments to generate the results in Table 1. Our current implementation is not fully optimized and we believe that the performance and accuracy can be further improved. Our current implementation highlights that 20^o to 30^o degree grid increments generate a good approximation of the configuration space. Furthermore, the performance of the overall algorithm is directly proportional to the grid size. It is straightforward to predict that it will take less than 20 minutes to search the conformation of cyclodecane at 30^o degree increment of the last four dihedral angles.

Our analysis has been based on pure geometry, we are yet to study energy aspects of our results. We believe that these results can be used as starting points for energy minimization procedures. That involves variations in the torsional angles, along with small variations in the bond length and bond angles.

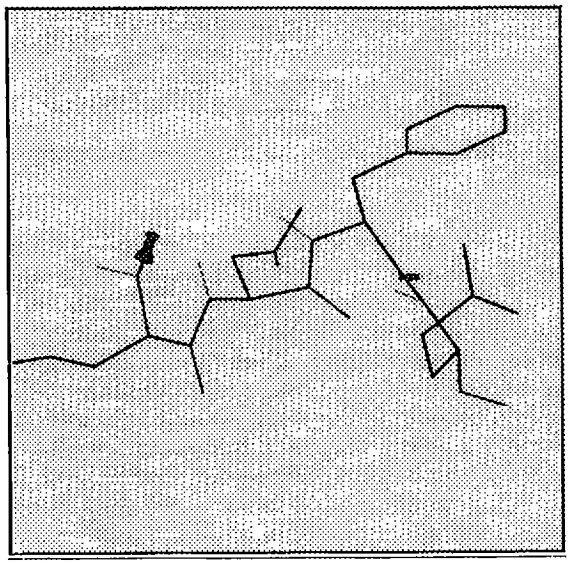
Molecular Embedding of Protein Chains

We applied the inverse kinematics procedure for computing the molecular embeddings of small protein chains. In this section, we demonstrate how to compute backbone atom positions of a protein chain from two end bonds positions based on geometric constraints.

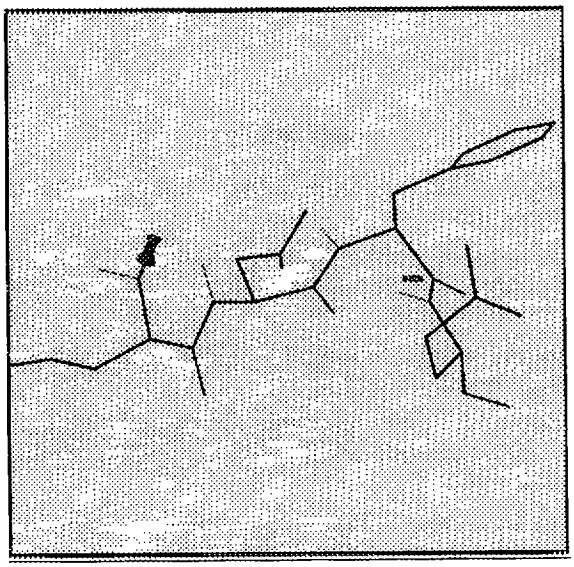
A peptide unit is showed in Fig. 6(a). Due to the the partial double bond character of C-N bond, no rotation along C-N bond is possible and all the atoms of a peptide unit lie in a plane. The only degree of freedom is the rotation along Cα-C and N-Cα bonds. Cα-C and N-Cα intersect at a point P in the same peptide



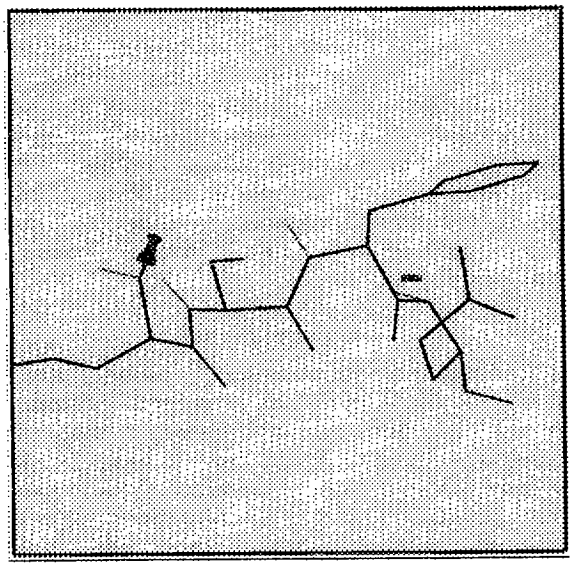
(a)



(b)



(c)



(d)

Figure 7. Four sets of possible atom positions of residues Asn and Phe. The Ca-C bond of Glu and N-Ca bond of Gln are fixed.

	X	Y	Z
Glu Cα	0.128	-2.805	-1.279
Glu C	0.769	-2.365	0.039
Gln N	-1.171	0.389	4.312
Gln Cα	-2.247	-0.494	5.359

Table 3: Atom positions of two end bonds

and N-Cα intersect at a point P in the same peptide plane as showed in Fig. 6(b). (The angle between Cα-C and N-Cα is not exact). Obviously, rotation along Cα-C and N-Cα are the same as rotation along Cα-P and P-Cα, respectively. Therefore, a peptide unit is geometrically equivalent to two consecutive rotatable bonds as showed in Fig. 6(c). Based on standard bond geometry in Fig. 1, all the distances and angles in Fig 6(c) are computed. Following this formulation, Theorem 0.1 is applied to the protein segments, while the planar peptide unit configuration is maintained.

One example of molecular embedding is showed in Fig. 7. In this protein segment, there are *four residues Glu, Asn, Phe and Gln.* The Cα-C bond of Glu is fixed, and the N-Cα bond of Gln is fixed. The four atom positions of these two bonds are listed in Table 3. Four possible conformations of Asn and Phe can be derived based on the bond lengths and bond angles constraints.

We have shown examples of computing backbone atom positions of protein segments. The same technique applies to side chain and other molecular structures. This is based on the fact that we made no assumption about the molecule geometry in the proof of Theorem 0.1 or in the procedure of solving inverse kinematics. This technique is also applicable for inferring backbone atom positions from the side chains. In the electron density mapping data, the Cα-Cβ bonds of side chains are much easier to identify than the backbone atom.

Conclusions

In this paper we have presented efficient techniques for kinematic manipulation of molecular chains. They are used for conformational search or computing the configuration of a molecular chain subject to rigid constraints. We have shown equivalence between the kinematic formulation and molecular embedding problems for short chains. We apply our techniques to compute the conformations of molecular chains with more than six dihedral angles (e.g. cyclooctanes) and compute the embeddings of a peptide units. The main features of our approach lie in its generality to all molecular geometries, efficiency and accuracy.

Acknowledgments.

We are grateful to Fred Brooks, David and Jane Richardson, Alex Tropsha and Bill Wright for many fruitful discussions.

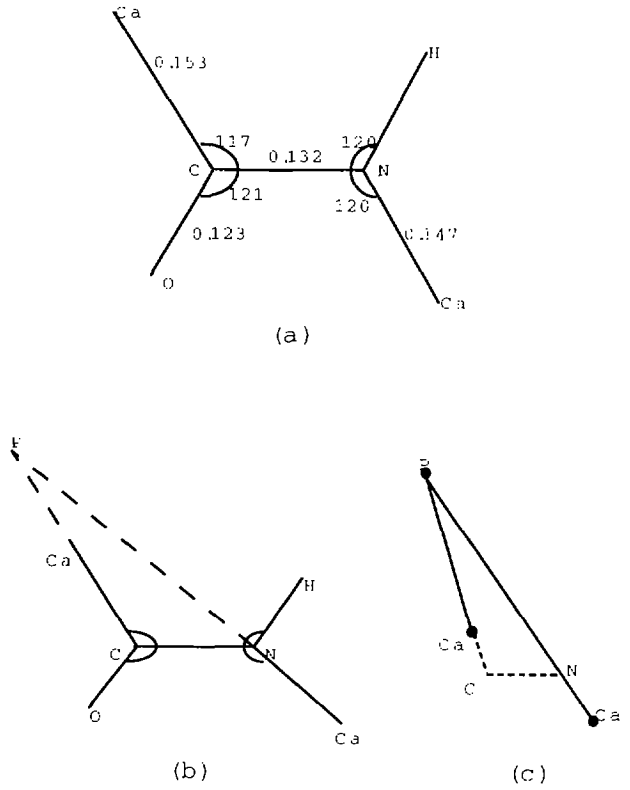


Figure 6: Geometry of a peptide unit

References

Biosym. 1991. *Insight.* San Diego, CA: Biosym Technologies Inc.

Bruccoleri, R., and Karplus, M. 1985. Chain closure with bond angle variations. *Macromolecules* 18(12):2767–2773.

Crippen, G., and Havel, T. Distance geometry and molecular conformation.

Crippen, G. 1981. *Distance Geometry and Conformational Calculations.* Wiley, New York: Research Studies Press.

Crippen, G. 1989. Linearized embedding: A new metric matrix algorithm for calculating molecular conformations subject to geometric constraints. *Journal of Computational Chemistry* 10(7):896–902.

Crippen, G. 1992. Exploring the conformation space of cylcoalkanes by linearized embedding. *Journal of Computational Chemistry* 13(3):351–361.

Denavit, J., and Hartenberg, R. 1955. A kinematic notation for lower-pair mechanisms based upon matrices. *Journal of Applied Mechanics* 77:215–221.

Go, N., and Scherga, H. 1970. Ring closure and local conformational deformations of chain molecules. *Macromolecules* 3(2):178–187.

Go, N., and Scherga, H. 1973. Ring closure in chain molecules with c_n, i or s_{2n} symmetry. *Macromolecules* 6(2):273–281.

Go, N., and Scherga, H. 1978. Calculation of the conformation of cyclo-hexaglycyl. *Macromolecules* 11(3):552–559.

Howard, A. E., and Kollman, P. A. 1988. An analysis of current methodologies for conformational searching of complex molecules. *Journal of Medicinal Chemistry* 31(9):1669–1675.

Leach, A., and Smellie, A. 1992. A combined model-building and distance geometry approach to automated conformational analysis and search. *Journal of Chemical Information and Computer Sciences* 32(4):379–385.

Manocha, D.; Zhu, Y.; and Wright, W. 1994. Conformational analysis of molecular chains using nano-kinematics. In *IEEE Computer Society Workshop on Shape and Pattern Matching in Computational Biology*. To appear. Also available as Technical Report TR94-036, Department of Computer Science, University of N. Carolina, Chapel Hill.

Manocha, D. 1992. *Algebraic and Numeric Techniques for Modeling and Robotics*. Ph.D. Dissertation, Computer Science Division, Department of Electrical Engineering and Computer Science, University of California, Berkeley.

Marshall, G. and et. al. 1979. The conformational parameter in drug design: The active analog approach. In ACS Symposium series. E. C. Olson, A. S., and Christoffersen, R., eds., *Computer-Assisted Drug Design*, volume 112, 205–226.

Peishoff, C., and Dixon, J. 1992. Improvements to the distance geometry algorithm for conformational sampling of cyclic structures. *Journal of Computational Chemistry* 13:565–569.

Raghavan, M., and Roth, B. 1989. Kinematic analysis of the 6R manipulator of general geometry. In *International Symposium on Robotics Research*, 314–320.

Saunders, M. 1989. Stochastic search for the conformations of bicyclic hydrocarbons. *Journal of Computational Chemistry* 10(2):200–208.

Sklenar, H.; Lavery, R.; and Pullman, B. 1986. The flexibility of the nucleic acids: (i) "SIR", a novel approach to the variation of polymer geometry in constrained systems. *Journal of Biomolecular Structure and Dynamics* 3(5):967–987.

Spong, M., and Vidyasagar, M. 1989. *Robot Dynamics and Control*. John Wiley and Sons.

Surles, M. 1992. An algorithm with linear complexity for interactive, physically-based modeling of large proteins. In *ACM SIGGRAPH*, 221–230.

Tripos. 1988. *Sybyl*. St. Louis, MO: Tripos Associates.

Weiner and et. al. 1983. A distance geometry study of ring systems. *Tetrahedron* 39(5):1113–1121.

Wilkinson, J. 1959. The evaluation of the zeros of ill-conditioned polynomials. parts i and ii. *Numer. Math.* 1:150–166 and 167–180.