Conformational analysis of molecular chains using Nano-Kinematics

Dinesh Manocha * Yunshan Zhu † William Wright ‡
Computer Science Department
University of North Carolina
Chapel Hill, NC 27599-3175
{manocha,zhu,wright}@cs.unc.edu
Fax: (919)962-1799

Abstract: We present algorithms for 3-D manipulation and conformational analysis of molecular chains, when bond length, bond angles and related dihedral angles remain fixed. These algorithms are useful for local deformations of linear molecules, exact ring closure in cyclic molecules and molecular embedding for short chains. Other possible applications include structure prediction, protein folding, conformation energy analysis and 3D molecular matching and docking. The algorithms are applicable to all serial molecular chains and make no asssumptions about their geometry. We make use of results on direct and inverse kinematics from robotics and mechanics literature and show the correspondence between kinematics and conformational analysis of molecules. In particular, we pose these problems algebraically and compute all the solutions making use of the structure of these equations and matrix computations. The algorithms have been implemented and perform well in practice. In particular, they take tens of milliseconds on current workstations for local deformations and chain closures on molecular chains consisting of six or fewer rotatable dihedral angles.

Categories: 3D Protein Modeling, 3D Molecular Matching and Docking, Multiple alignment of genetic sequences

*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

‡Supported in part by NIH National Center for Research Resources grant RR-02170

1 Introduction

Three dimensional manipulation of molecular chains is a fundamental problem in conformational analysis and docking applications. Conformational analysis deals with computation of minimal energy configurations of deformable molecules and docking involves matching one molecular structure to a receptor site of a second molecule and computing the most energetically favorable 3-D conformation. These problems have been extensively studied in biochemistry and molecular modeling literature. Molecular chains are modeled using bond lengths, bond angles and dihedral angles.

In the literature on conformational energy calculations, bond lengths and bond angles have been treated in two different ways, as variables or as fixed parameters[GoS70]. However, modeling the bond-length and bond-angle as variables significantly add to the complexity and even for small chains consisting of six or seven bonds, minimal energy computation becomes a non-trivial problem. In their classic paper, Go and Scheraga highlight that the minimal energy conformations computed by fixing these parameters is a good guess to the actual minima and can be used along with a perturbation treatment for the energy minimization [GoS70]. Therefore, in this paper, we assume that the bond lengths and bond angles are fixed and compute the conformation of the molecular chains by changing the dihedral angles only.

Given the assumptions of fixed bond-lengths and bond-angles, two fundamental problems arise in conformational energy calculations: *ring closure* and *local conformational deformations* [GoS70]. The problem of ring closure arises when we deal with cyclic molecules, i.e. the calculation of dihedral angles which corresponds to exact ring closure. In conformational energy minimization procedures, one selects a starting conformation and alters it step by step such that the corresponding conformational energy decreases monotonically. However, a small change in a dihedral angle located near the middle of the long chain causes a drastic change in the overall conformation of the molecule. It is therefore, desired to have a cooperative variation of angles which confines the conformational changes to a *local* section of the chain. As a result we are interested in calculation of changes in dihedral angles, which cause local deformations of conformations in long polymer chains [GoS70, BA82, CMR89, BK85]. In this paper, We also address the *molecular embedding problem*, that is, to find atom positions that satisfies a set of geometric constraints. It is an important problem in interpreting data from NMR experiment [Cri89].

Many interactive systems for molecular modeling such as *Sybyl* [Tri88] and *Insight* [Bio91] provide capabilities for building and changing molecular models by rotating the dihedral angles. The full conformation space of acyclic molecules is the Cartesian product of all the dihedral 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 [HK88]. 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 [GoS70, GoS73]. The algorithm in [GoS70] is restricted to molecular chains where the rotation can take place about all backbone bonds and the methods in [GoS73] 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 [We83, PD92]. 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 [SLP86]. Although these approaches make no assumption about the geometry of molecules, they are relatively slow in practice, may 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 [Cri89, Cri92]. 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 [GoS78, Sau89, BK85]. These can be slow in practice and are not guaranteed to find all the solutions.

In this paper, we reduce the problem of pose-computation to a system of algebraic equations and show how they relate to the problems of direct and inverse *kinematics* in mechanics and robotics [Cra89, SV89]. In particular, we show the correspondence between molecular conformational analysis and *inverse kinematics* of serial manipulators. We also show how inverse kinematics can be used to solve the molecular embedding problem for short chains. The problem of inverse kinematics has been extensively studied in the robotics literature, but closed form solutions are known for some special geometries only. We present techniques for computing the solutions of the algebraic equations by making use of their sparse structure and reduce the problem to computing the eigenvalues and eigenvectors of a matrix. Based on numerically stable and efficient algorithms for computing the eigendecomposition of a matrix, we present fast and accurate algorithms for ring closure and loop analysis. These algorithms have been implemented and take about of tens

of milliseconds on current workstations for manipulation of molecular chains consisting of up to six rotatable bonds. We demonstrate their performance on chain closures of cyclohexane and loop deformation on a polypeptide chain. These algorithms make no assumptions about the geometry of the molecules and are applicable to all molecular chains. These techniques are based on kinematics and applied to molecular bonds of the order of a few Angstroms. Therefore, we refer to them as *nano-kinematics*

The rest of the paper is organized in the following manner. In Section 2, we show the equivalence between the kinematics problem and molecular modeling problems. In Section 3, we show how the technique can also be applied to molecualr embedding problems. We present the algebraic algorithms in Section 4 and show the reduction to an eigenvalue problem. Depending on the geometry of the molecular chain, we obtain different matrix orders. We discuss extensions of the algorithm to chains consisting of more than six dihedral angles in Section 5. The algorithms are illustrated in Section 6 with several examples of applications. The details of the implementation and performance are presented in Section 7.

2 Molecular Modeling and Inverse Kinematics

The inverse kinematics problem for general serial manipulators is fundamental for computer controlled robots. Given the pose of the end effector (the position and orientation), the problem corresponds to computing the joint displacements for that pose [SV89]. A robot manipulator may consist of prismatic, revolute or spherical joints. A molecular chain with fixed bond lengths and bond angles corresponds to a serial manipulator with revolute joints. Therefore, we only consider the inverse kinematics problem for serial chains with revolute joints. Initially, we consider chains with six rotatable joints (denoted as $6R$) and later on extend the results to arbitrary chains.

We use the Denavit–Hartenberg (DH) formalism, [DH55], to model a serial chain with n revolute joints. Each bond is represented by the line along its joint axis and the common normal to the next joint axis. In the case of parallel joints, any of the common normals can be chosen. A coordinate system is attached to each joint for describing the relative arrangements among the various bonds. The z-axis of frame i, called z_i, is coincident with the joint 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

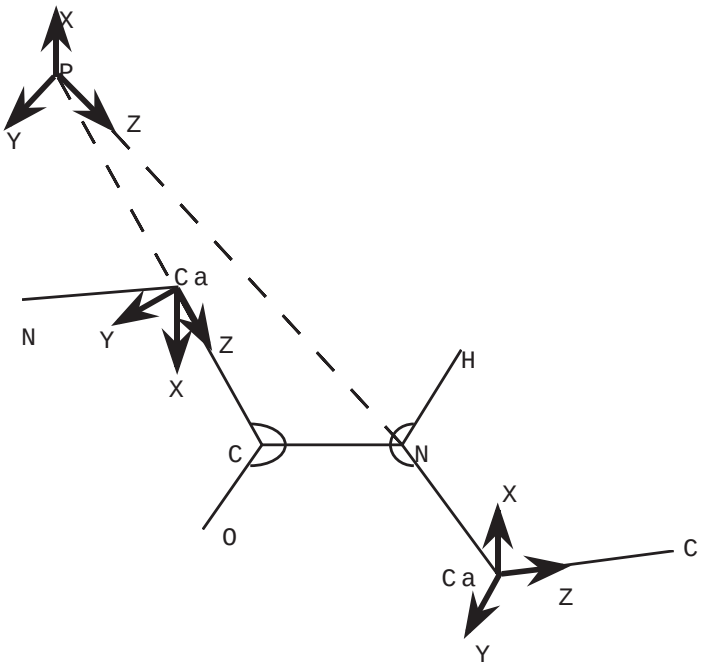

Figure 1: Coordinate systems on a polypeptide units based on DH formalism

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}.

These parameters for the polypeptide unit are shown in Fig. 2. The dihedral angles correspond to the θ_i's. 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:

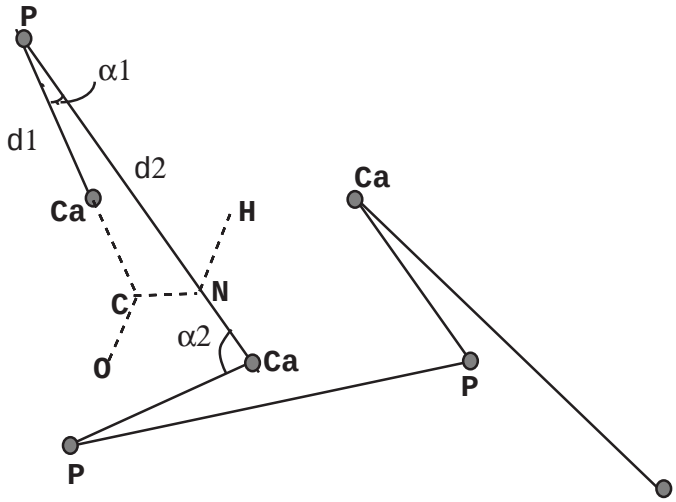

Figure 2: DH parameters for a polypeptide unit

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

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

For a given serial chain or a robot manipulator we are also given the configuration of the end of the chain or the end-effector. This configuration is described with respect to the base link or link 1. We represent this configuration as:

$$\mathbf{A}_{hand} = \begin{pmatrix} l_x & m_x & n_x & q_x \\ l_y & m_y & n_y & q_y \\ l_z & m_z & n_z & q_z \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The problem of inverse kinematics corresponds to computing the joint angles, $\theta_1, \theta_2, \ldots, \theta_n$ such that

$$\mathbf{A}_1\mathbf{A}_2 \ldots \mathbf{A}_n = \mathbf{A}_{hand}. \tag{2}$$

This relationship can be 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.

The problems of ring closure and local deformations turn out to be a special case of inverse kinematics formulation, as shown below.

2.1 Ring Closure

The ring closure problem turns out to be a special case of the inverse kinematics problem highlighted above. In particular for ring closure, the right hand side matrix A_{hand} corresponds to the identity matrix:

$$\mathbf{A}_{hand} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

As a result, all the solutions to the ring closure are obtained by substituting for A_{hand} in (2).

2.2 Local Deformations

The problem of performing conformational variations on a local portion of a 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 coordinate systems and computing the DH parameters for the local chain. The rotatable bonds need not be contiguous (as shown in Fig. 1). The resulting problem can be posed as shown in (2). A_{hand} corresponds to the conformation of the end of the chain and n corresponds to the number of rotatable in the selected chain.

2.3 Cases of Rotation about All Backbone Bonds

Go and Scheraga have considered a restricted version of the inverse kinematics problem, when the rotation can take place about all backbone bonds (as opposed to a few selected dihedral angles) [GoS70]. This turns out to be a special case of the inverse kinematics problem as well. In particular, the Denavit-Hartenberg parameters of such a chain correspond exactly to: $a_i = 0$, $d_i =$ bond length and $\alpha_i =$ bond angle. Such a molecular chain has been shown in Fig. 3.

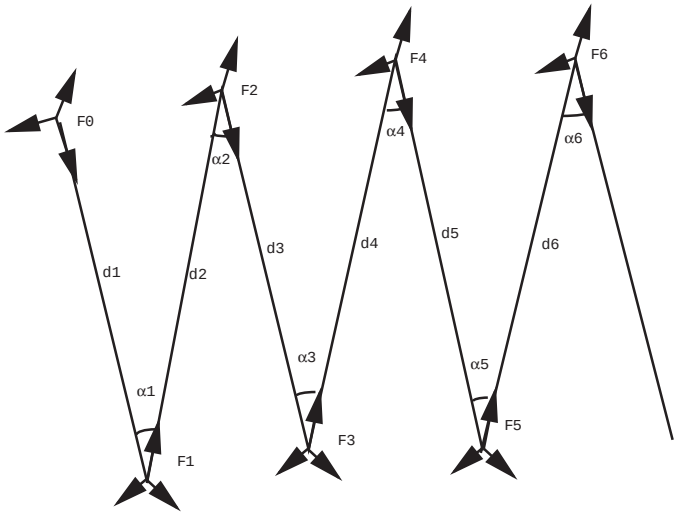

Figure 3: Parameters for Molecular Chain with All Bonds Rotatable

3 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 [Cri89]. 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 [CH, Cri81]. 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 [Cri89]. 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 [LS92].

3.1 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 dihedral angles, given the pose of the end-effector A_E. Given the dihedral angles, the positions of the intermediate atoms can be easily computed.

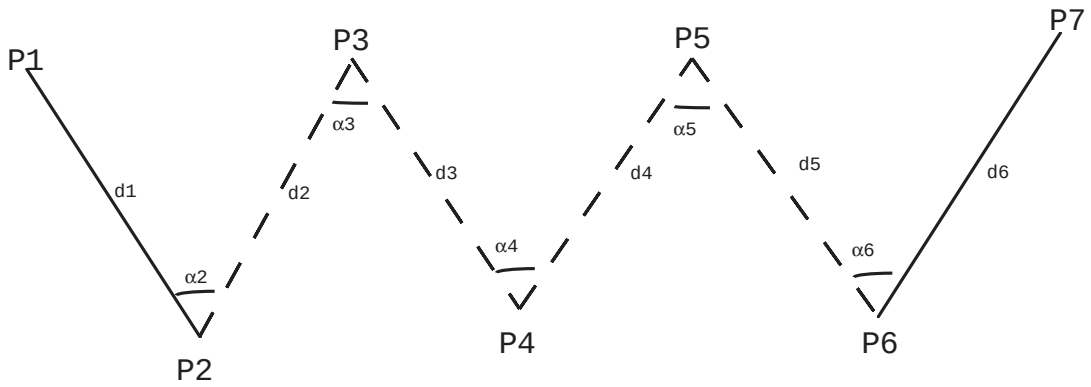

Figure 3: Molecular embedding problem with unknown atom positions

Theorem 3.1 *Assuming fixed bond lengths and bond angles, the conformation of a molecular chain*

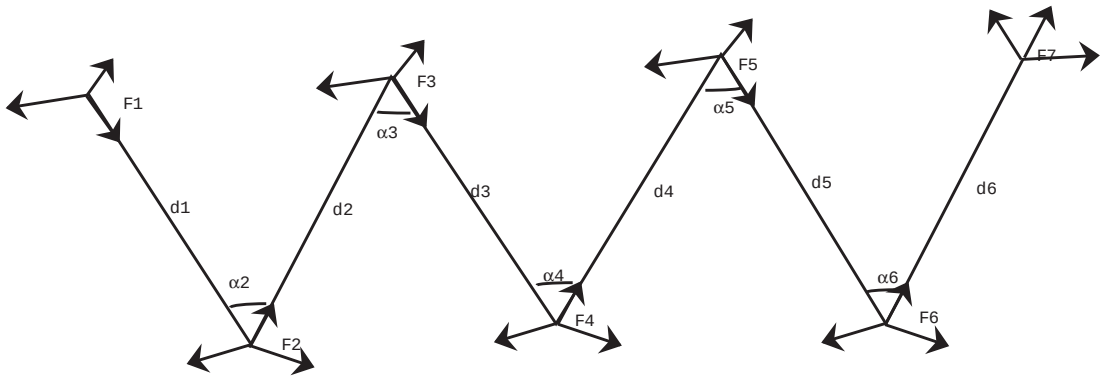

Figure 4: Inverse kinematics formulation of molecular embedding

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.

4 Inverse Kinematics and Matrix Formulation

Most of the literature in robotics has concentrate 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 from exists.

The inverse kinematics problem for six revolute joints has been studied for more than two decades in the robotics literature. Given the multivariate polynomial system, different algorithms to reduce the problem to finding roots of a univariate polynomial are given in [Pie68, RRS73, AA79, DC80]. Given a $6R$ manipulator with a pose of the end-effector, the maximum number of solutions (assuming a finite number of solutions) is 16 as shown in [Pri86, LL88, RR89]. Practical methods to compute all the solutions are presented in [WM91] based on continuation methods and in [MC92] based on matrix computations.

In this section, we consider the problem of inverse kinematics for a molecular chain with six revolute joints. Chains with more than six joints or fewer than six joints are considered in next section. We make use of the algebraic reduction of the problem to a univariate system presented by Raghavan and Roth [RR89]. 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 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 [Wil59]. As a result, we may not be able to compute accurate results using IEEE double precision arithmetic on current workstations.

4.1 Raghavan and Roth Formulation

Given the matrix equation for a six jointed manipulator, (2), Raghavan and Roth rearrange it as:

$$A_3 A_4 A_5 = A_2^{-1} A_1^{-1} A_{hand} A_6^{-1}. \tag{3}$$

As a result the entries of the left hand side matrix are functions of θ_3, θ_4 and θ_5 and the entries of the right hand side matrix are functions of θ_1, θ_2 and θ_6. This lowers their degrees and reduces the symbolic complexity of the resulting expressions. On equating the corresponding entries of the matrix equation, (3), and after simplification, these equations are expressed in a linear formulation given as:

$$(\mathbf{R}) \begin{pmatrix} s_1 s_2 \\ s_1 c_2 \\ c_1 s_2 \\ c_1 c_2 \\ s_1 \\ c_1 \\ s_2 \\ c_2 \end{pmatrix} = (\mathbf{Q}) \begin{pmatrix} s_4 s_5 \\ s_4 c_5 \\ c_4 s_5 \\ c_4 c_5 \\ s_4 \\ c_4 \\ s_5 \\ c_5 \\ 1 \end{pmatrix}, \tag{4}$$

where $\mathbf{R}$ is a 14×8 matrix, whose entries are functions of the geometry of the chain and the configuration of the end of the chain. $\mathbf{Q}$ is a 14×9 matrix, whose entries are linear functions of s_3 and c_3 and their coefficients are functions of the manipulator parameters and the pose. The relationship expressed in (4) helps us in eliminating four of the five variables.

The rest of the algorithm in [RR89] proceeds by using 8 of the 14 equations to solve for the power products on the right-hand-side, and then use these solutions to eliminate these power products from the remaining 6 equations. Eventually, we obtain a linear system of the form:

$$(\mathbf{S}) \, (s_4 s_5 \;\; s_4 c_5 \;\; c_4 s_5 \;\; c_4 c_5 \;\; s_4 \;\; c_4 \;\; s_5 \;\; c_5 \;\; 1)^T = \mathbf{0}, \tag{5}$$

where $\mathbf{S}$ is 6×9 matrix, whose entries are linear combinations of s_3, c_3 and 1.

4.2 Symbolic Representation

In our algorithm, we represent the entries of R and Q as symbolic functions of the geometry of the molecular chain. Given a particular geometry, we substitute the corresponding parameters. The parameters for the polypeptide geometry are highlighted in Section 5.

It is possible that for some special geometries Q is rank deficient. As a result, we perform the rank computation after substitution using the singular-value decompositions (SVD) [GL89]. If the rank, r, is less than 8 we only use r equations to eliminate the power products on the left hand side of (4) and thereby obtain $(14-r)$ equations. Thus, the order of $\mathbf{S}$ in (5) is $(14-r)\times 9$.

Typically, Q has rank 8 and therefore, S consists of 6 rows. In case S consists of more than 6 rows, we are given an overconstrained system. To solve such a system, we take 6 linear combinations of the rows of S and let the matrix S now correspond to this new system of 6 combinations. Once we obtain all the solutions to these new 6 equations, we back substitute them into the original overconstrained system consisting of $14-r$ equations, to find all the solutions of the overconstrained system. In the rest of the paper, we work with the assumption that S consists of 6 rows.

The resulting system is transformed into an algebraic system by replacing the variables s_i and c_i corresponding to the dihedral angle by:

$$s_i = \frac{2x_i}{1+x_i^2},\ c_i = \frac{1-x_i^2}{1+x_i^2}.$$

The resulting system of equations are multiplied by $(1+x_i^2)$ to transform them into polynomial equations. As a result, we get a system of polynomial equations in three unknowns, x_3, x_4, x_5. Furthermore it can be represented as a linear system of the form

$$M(x_3)\ H(x_4,x_5) \quad = \quad \mathbf{0}, \tag{6}$$

where $M(x_3)$ is a 6×9 matrix, whose entries are quadratic polynomial in x_3 and $H(x_4,x_5)$ is a 9×1 vector consisting of entries, $h_1(x_4,x_5),\ldots,h_9(x_4,x_5)$. Each $h_i(x_4,x_5)$ is a monomial function of degree at most four in x_4 and x_5.

4.3 Formulating a Square System

The algebraic equations corresponding to the inverse kinematics have been represented in terms of the matrix polynomial, $M(x_3)$. In our case, the set of equations (6) correspond to a system of non-linear equations (in x_3, x_4 and x_5) and we are interested in computing all their common solutions.

Given the linear system consisting of 6 equations, we multiply it with appropriate power products to convert it into a square system. In particular, we multiply each of the 6 equations by x_4 and obtain 6 new equations represented as.

$$M(x_3)\ \ x_4\ H(x_4,x_5) = \mathbf{0}.$$

We combine these equations into 12 equations represented in the matrix form as:

$$\overline{M}(x_3)\ \overline{H}(x_4, x_5) = \mathbf{0}, \tag{7}$$

where $\overline{M}(x_3)$ is the 12×12 matrix:

$$\overline{M}(x_3) = \begin{pmatrix} M(x_3) & \mathbf{O} \\ \mathbf{O} & M(x_3) \end{pmatrix},$$

and $\mathbf{O}$ is a 6×3 null matrix. Moreover, $\overline{H}(x_4, x_5)$ is a 12×1 vector, whose entries are monomial functions in x_4, x_5 of degree at most 5. $\overline{H}(x_4, x_5)$ is constructed by combining the terms of $H(x_4, x_5)$ and $x_4 H(x_4, x_5)$. As a result, $\overline{M}(x_3)$ represents a square system of order 12 [RR89].

4.4 Matrix Formulation

Given the matrix representation, $\overline{M}(x_3)\ \overline{H}(x_4, x_5)$, the problem of inverse kinematics corresponds to finding all the solutions to the linear system (7).

$\overline{M}(x_3)$ is a 12×12 matrix polynomial of degree 2. It can be represented as:

$$\overline{M}(x_3) = \overline{M}_0 + \overline{M}_1 x_3 + \overline{M}_2 x_3^2, \tag{8}$$

where $\overline{M}_i$ are numeric matrices. The matrix polynomial $\overline{M}(x_3)$ is *regular* if its symbolic determinant is non-zero. Otherwise it is a *singular* matrix polynomial. We are interested in computing the zeros of its determinant. For singular matrix polynomials it corresponds to computing the eigenvalues of the matrix polynomial [VDD83].

The following theorem shows how this problem of zero computation can be linearized, and the problem of inverse kinematics is reduced to computing appropriate zeros of a matrix pencil of the form, $A - xB$ [MC92, VDD83].

Theorem 4.1 *Given the matrix polynomial, $\overline{M}(x_3)$ the zeros of the matrix polynomial correspond to the zeros of the matrix generalized system $A - xB$, where*

$$B = \begin{pmatrix} I & 0 \\ 0 & \overline{M}_2 \end{pmatrix} \quad A = \begin{pmatrix} 0 & I \\ -\overline{M}_0 & -\overline{M}_1 \end{pmatrix}. \tag{9}$$

O and I are null and identity matrices of order 12.

In case the matrix polynomial is regular and $\overline{M}_2$ is well conditioned, this can be further reduced to computing the eigenvalues of [MC92]

$$C = \begin{pmatrix} 0 & I \\ -\overline{M}_2^{-1}\overline{M}_0 & -\overline{M}_2^{-1}\overline{M}_1 \end{pmatrix}. \tag{10}$$

4.5 Solving for All Geometries

A generic molecular chain results in regular matrix polynomials and the solutions are computed using the eigendecomposition of the matrices highlighted in the previous section. However, most of the special geometries (like the polypeptide unit or cyclohexanes) result in singular matrix polynomials and matrix pencils. The current algorithms for eigendecomposition of such pencils can be ill-conditioned and relatively slow. In this section, we make use of the structure of the matrix polynomials corresponding to the kinematics formulation and derive efficient and accurate algorithms.

A simple test to check whether $\overline{M}(x_3)$'s is singular is based upon substituting random values of x_3 into the matrix polynomial and checking the rank of the resulting numeric matrix. Furthermore, the kernel of singular $\overline{M}(x_3)$ may or may not be independent of x_3. The rest of the algorithm involves checking the kernel and performing appropriate matrix computations.

4.5.1 Kernel of the Matrix Polynomial

We check whether the matrix has a fixed kernel based on the following lemma.

Lemma 4.2 *Given the* 12×12 *matrix polynomial* $\overline{M}(x_3)$*, its right kernel consists of* r *constant linearly independent vectors, iff the* 12×36 *matrix* $P = \begin{pmatrix} \overline{M}_0^T & \overline{M}_1^T & \overline{M}_2^T \end{pmatrix}$ *has* r *linearly independent vectors in its left kernel.*

In the above lemma, $\overline{M}_i^T$ refers to the transpose of $\overline{M}_i$. The proof of this lemma is simple. Based on this lemma, it is trivial to check whether $\overline{M}(x_3)$ has a fixed kernel or it is a function of x_3.

4.5.2 Matrix Polynomials with Constant Kernel

In this section, we consider matrix polynomials, $\overline{M}(x_3)$ with a fixed kernel. The algorithm for inverse kinematics follows from the following theorem.

Theorem 4.3 *Given the matrix polynomial,* $\overline{M}(x_3)$*, such that*
(i) The system of equations, (7), has a finite number of solutions.
(ii) $\overline{M}(x_3)$ *has a fixed right kernel.*
For any solution, $(x_3^{'}, x_4^{'}, x_5^{'})$*, to the equations (7),* $x_3^{'}$ *is a solution of* $Det(\overline{Mm}(x_3)) = 0$*, where* $\overline{Mm}(x_3)$ *is a leading minor of* $\overline{M}(x_3)$*.*

We are omitting the proof due to space requirements.

The theorem is constructive and is used to compute all the solutions of inverse kinematics. We take a non-singular minor of $\overline{M}(x_3)$. That corresponds to a regular matrix pencil of order $N - r$ and has degree 2. Using Theorem 4.1, all the roots of its determinant correspond to the eigenvalues of the equivalent matrix, C as shown in (10). The eigenvalues correspond to $x_3^{'}$ of a given solution. Given $x_3^{'}$, we compute $(x_4^{'}, x_5^{'})$ in the following manner.

Take any two linear combinations of the 6 equations in x_4 and x_5:

$$M(x_3^{'})\ H(x_4, x_5) = \mathbf{0}. \tag{11}$$

Let the two equations be $l_1(x_4, x_5) = 0$ and $l_2(x_4, x_5) = 0$. These are two equations of degree 4 each, however the highest degree of x_4 or x_5 in each of these equation is 2. We compute all the solutions to these equations using Sylvester resultant formulation and reducing the problem to an eigenvalue formulation. The resulting matrix is 8×8 and we get at most 8 real solutions of these equations. Given the real solution, we back substitute them into the system (11) and obtain the actual solution.

4.5.3 Matrix Polynomials with Dependent Kernel

In this section we consider matrix polynomials whose kernel is a function of x_3. We make use of the 6 equations represented in (6) and use them to compute all the (x_3, x_4, x_5) corresponding to the solutions of the inverse kinematics problem.

Given the 6 equations being represented as $M(x_3)\ H(x_4, x_5) = 0$. $M(x_3)$ is rank deficient. Depending upon the rank of $M(x_3)$, we use the following approaches.

Rank$(M(x_3)) < 3$ In this case, there are infinite solutions for the configuration of the end of the chain. This follows from the fact that at most 2 of the 6 equations are independent. A system of one or two algebraic equations, in three unknowns corresponds to a two dimensional or one dimensional algebraic set. As a result, there are infinite solutions of the system, $M(x_3)\ H(x_4, x_5) = 0$, which implies that there are infinite solutions of the matrix equation (2).

Rank$(M(x_3)) > 3$ We choose four independent equations and represent them as

$$M^{'}(x_3)\ H(x_4, x_5) = 0,$$

where $M^{'}(x_3)$ is a 4×9 minor corresponding to the four equations. In order to solve this system, we use dialytic elimination and linearize it. This is done by multiplying the four equations by x_4, x_5 and $x_4 x_5$. As a result we obtain additional 12 equations of the form:

$$\begin{aligned} M^{'}(x_3)\ (x_4 H(x_4, x_5)) &= 0 \\ M^{'}(x_3)\ (x_5 H(x_4, x_5)) &= 0 \\ M^{'}(x_3)\ (x_4 x_4 H(x_4, x_5)) &= 0. \end{aligned}$$

These 16 equations are combined and put together in a linear form:

$$\overline{MM}(x_3)\ \overline{HH}(x_4, x_5) = 0, \tag{12}$$

where $\overline{MM}(x_3)$ is a 16×16 matrix polynomial in x_3 of degree 2, and its entries consist of entries of $M^{'}(x_3)$ and null matrices. $\overline{HH}(x_4, x_5)$ is obtained by combining the power products in $H(x_4, x_5),\ x_4 H(x_4, x_5),\ x_5 H(x_4, x_5),\ x_4 x_5 H(x_4, x_5)$. This system is solved by reducing it to a 32×32 eigenvalue problem. From the real solutions of this system, we pick the solutions of the 6 equations, $M(x_3)\ H(x_4, x_5) = 0$ by back substitution.

Rank$(M(x_3)) = 3$ In this case, only 3 of the 6 equations of $M(x_3)\ H(x_4, x_5) = 0$ are independent. The resulting problem corresponds to solving three equations in three unknowns. Each equation has degree 6, however the highest degree term corresponds to $x_3^2 x_4^2 x_5^2$. Such equations are solved using Dixon's formulation of resultant [Dix08] and reducing the problem to eigendecomposition of a 48×48 matrix, followed by back substitution.

5 Extension to All Molecular Chains

In the previous section we had considered molecular chains with six dihedral chains and reduced the problem to an eigendecomposition problem. In this section, we highlight techniques for manipulation of molecular chains with fewer than six dihedral angles or more than six dihedral angles. Given a molecular chain with n dihedral angles, the problem of kinematic manipulation reduces to n equations in 6 unknowns. In case, $n < 6$, there may be no solutions of this system. We solve for all conformations of the chain in the following manner.

Any serial chain with less than 6 dihedral angles is treated as a 6-jointed system. We use additional bonds with rotatable dihedral angles in the configuration. The additional bonds are

placed in such a manner that the dihedral angle corresponding to their coordinate systems are zero. Moreover, this dihedral angle is chosen as θ_3. We treat the chain as a 6-jointed chain and compute all the solutions for the given pose. The original chain has any solutions, iff the 6-jointed chain has a solution with $\theta_3 = 0$.

Chains with more than six dihedral angles are commonly used in molecular modeling. Examples include cyclo-heptanes, cyclo-nonanes, sugar molecules besides the polypeptide units. Given a chain with $n > 6$ dihedral 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 $n-6$ dihedral 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 dihedral 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 opp osed to all the nine independent variables. For dihedral angle increments of 60^o, 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 $n-6$ dihedral angles as opposed to all the n angles.

6 Applications

The algorithm presented in previous sections makes no assumptions about the geometry of molecular chains, and it can be applied to a variety of problems. In this section, we will apply the algorithm to ring closure and local deformation problem. For each case, we will compute its Denavit Hartenberg parameters, and discuss the results.

Number	Link Length	Offset Distance	Twist Angle
i	a_i	d_i	α_i
1	0.0	1.525	69.6
2	0.0	1.526	69.6
3	0.0	1.526	69.6
4	0.0	1.526	69.6
5	0.0	1.526	69.6
6	0.0	1.526	69.6

Table 1: Denavit-Hartenberg Parameters of Cyclohexane Model

6.1 Cyclohexane Conformations

We studied cyclohexane conformation assuming rigid bond geometry. In particular, all C-C bond lengths are fixed at 1.526 Angstrom, and C-C-C bond angles are fixed at 110.4 degree. Since we are only concerned with main chain conformation, the bond lengths and bond angles of C-H are irrelevant to our analysis. It was proved in [GoS73] that cyclohexane has infinite geometrically possible conformation due to its structural symmetry. In the following example, we slightly increase the length of the last bond. Since the symmetry no longer holds, we obtain a finite set of solutions. Using the procedure in Section 2, the Devanit Hartenberg parameters can be computed for this set of bond geometry. They are listed in Table 1. The end effector is an identity matrix because of the ring structure.

Four sets of solutions computed by the algorithm are listed in Table 2. The first two sets of solutions correspond to the dihedral angles of the *chair* conformation in Fig. 5(a). The last two sets correspond to those of the *twisted boat* conformation in Fig. 5(b).

We should point out here that the four solutions are obtained based on the assumption that bond lengths and bond angles are fixed as listed in Table 1. If variation of bond length or bond angle is allowed, other possible conformations exist.

In the data listed in Table 3, we perturb the bond lengths and bond angles by two percent. A different set of solutions are obtained in Table 4. The first two solutions again correspond to the *chair* conformation, while the last two correspond to the *boat* conformation, see Figure 5(c). The *twisted boat* conformation from the standard bond geometry does not show up here.

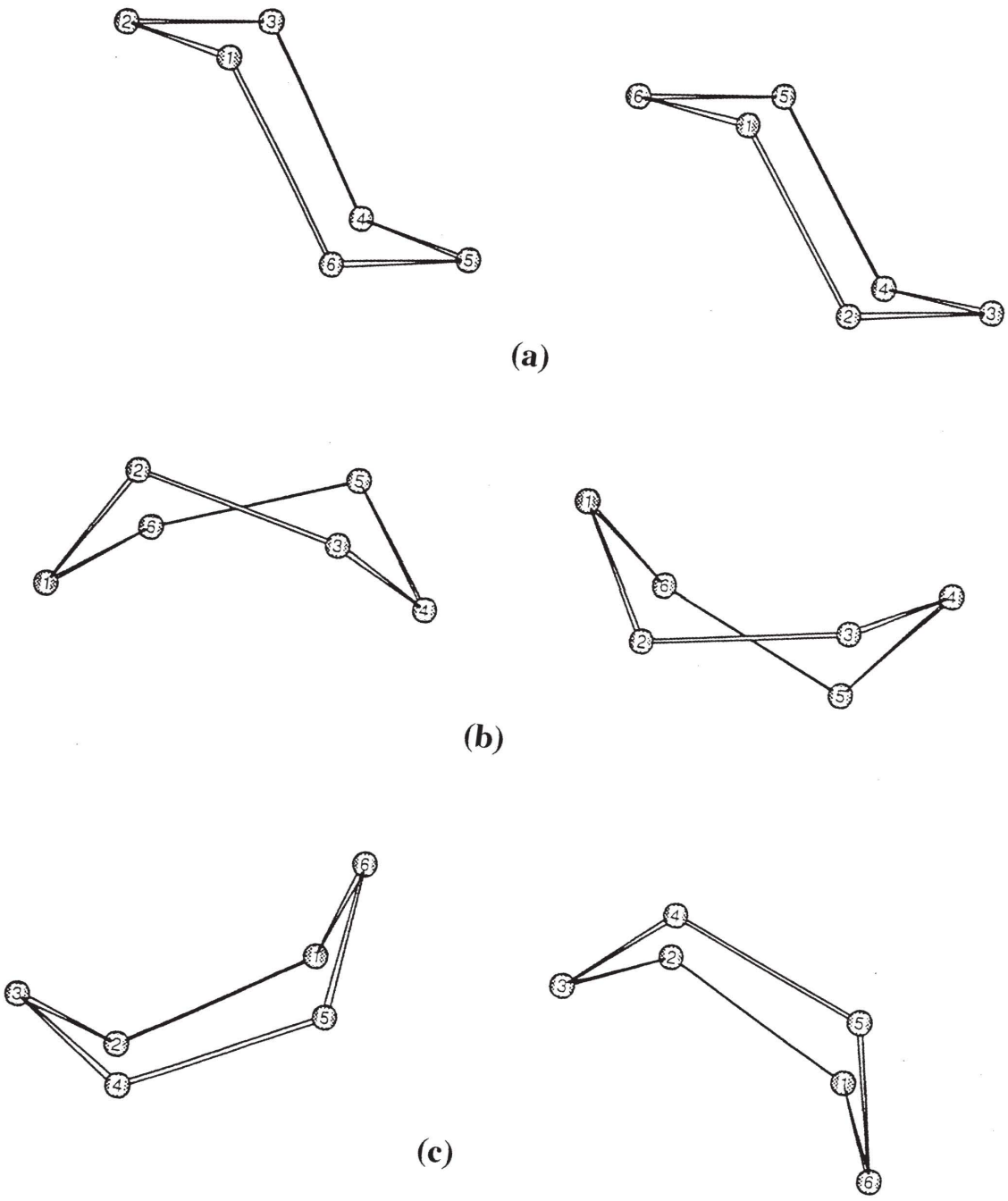

Figure 5: Cyclohexane conformations: (a). (top) *chair* conformation, (b). (middle) *twisted boat* conformation, (c). (bottom) *boat* conformation

Through a series of small perturbations of the bond geometry, we noticed that the *chair* conformation is more stable than the *boat* and *twisted boat* conformation. In all cases of small perturbation we tried, there are no more than four solutions for the ring closure problem. Two of these four solutions constantly correspond to the *chair* conformation, while the other two correspond to either the *boat* or the *twisted boat* conformation. Furthermore, the dihedral angles of the *chair* conformations stay almost constant, while those of either the *boat* or the *twisted boat* vary a lot for small perturbations on bond geometry. Our result is in accord with previous experimental and theoretical result that the *chair* conformation is the favored conformation of cyclohexane.

6.2 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. 6 ($d = 1.55A^o$ 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. 6, 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 dihedral angles and use inverse kinematics to compute the rest of the six dihedral 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 5. 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.

i	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
1	-57.69	57.67	-57.62	57.60	-57.62	57.67
2	57.69	-57.67	57.62	-57.60	57.62	-57.67
3	67.78	-32.04	-31.98	67.70	-31.98	-32.04
4	-67.78	32.04	31.98	-67.70	31.98	32.04

Table 2: The dihedral angles corresponding to the Cyclohexane with standard bond length and bond angles

Number	Link Length	Offset Distance	Twist Angle
i	a_i	d_i	α_i
1	0.0	1.525	69.5
2	0.0	1.525	69.6
3	0.0	1.526	69.8
4	0.0	1.527	69.7
5	0.0	1.526	69.5
6	0.0	1.526	69.6

Table 3: Denavit-Hartenberg Parameters of perturbed Cyclohexane Model

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 [Man92].

The performance of the inverse kinematics algorithm for conformation search of a cyclooctane has been highlighted in Table 6. In particular, the running time is shown for different grid size increments. It takes 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 5. 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

i	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
1	-57.38	57.73	-57.95	57.93	-57.75	57.42
2	57.38	-57.73	57.95	-57.93	57.75	-57.42
3	-60.73	5.87	54.06	-61.25	6.42	53.50
4	60.73	-5.87	-54.06	61.25	-6.42	-53.50

Table 4: The dihedral angles corresponding to the Cylcohexane with perturbed bond length and bond 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	0	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 5: 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.

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 dihedral angles, along with small variations in the bond length and bond angles.

6.3 Local Deformation

The algorithm is also applied to study local deformation of protein segments. In the following example, we compute the Denavit Hartenberg parameters and the end effector from protein datafile instead of using a standard set of bond geometry. In a given datafile from protein databank, the

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

Table 6: Running time of cyclooctane conformation search at various grid sizes

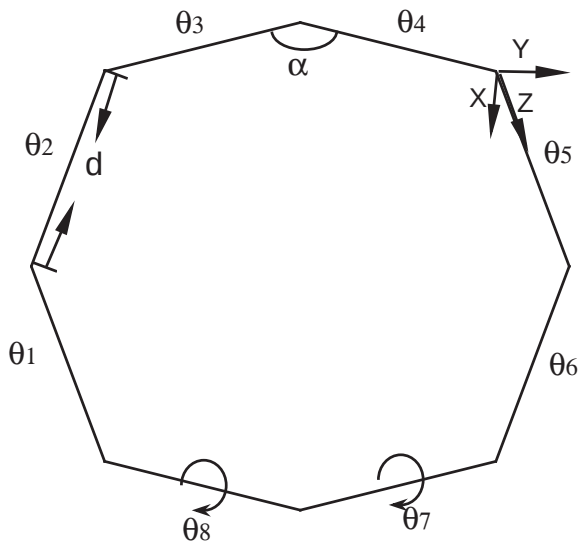

Figure 6: Cyclooctane Conformation : systematic search on θ_7 and θ_8

bond geometry can deviate from the standard geometry for as much as 10 percent. The error from the experimental data is minimized by computing the bond geometry directly from atom coordinates. This approach is useful for analyzing protein models. In protein synthesis applications, standard geometry (for example Pauling-Corey geometry) can be used.

Given atom coordinates, bond lengths and bond angles can be easily computed. We assume all covalent bonds are rigid, and the only freedom in a protein chain is the rotation along C-Ca bond and N-Ca bond. Using the procedure in section 2, Denavit Hartenberg parameters can be computed from the bond geometry. The end effector can be either computed from the original file or specified by the user, dependent on the application.

In our example, we choose a segment of α-helix from protein Felix [Ric93]. Two orthogonal views of this segment are shown in Figure 8(a). The Denavit Hartenberg parameters and the end effector are computed from the atom coordinates.

The D-H parameters are listed in Table 7. The pose of the end effector is:

$$\mathbf{A}_{hand} = \begin{pmatrix} 0.628 & 0.764 & -0.146 & 4.141 \\ -0.552 & 0.569 & 0.609 & -0.152 \\ 0.548 & -0.302 & 0.779 & 3.251 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Two solutions for the given pose are listed in Table 8. The first set of solutions in Table 8

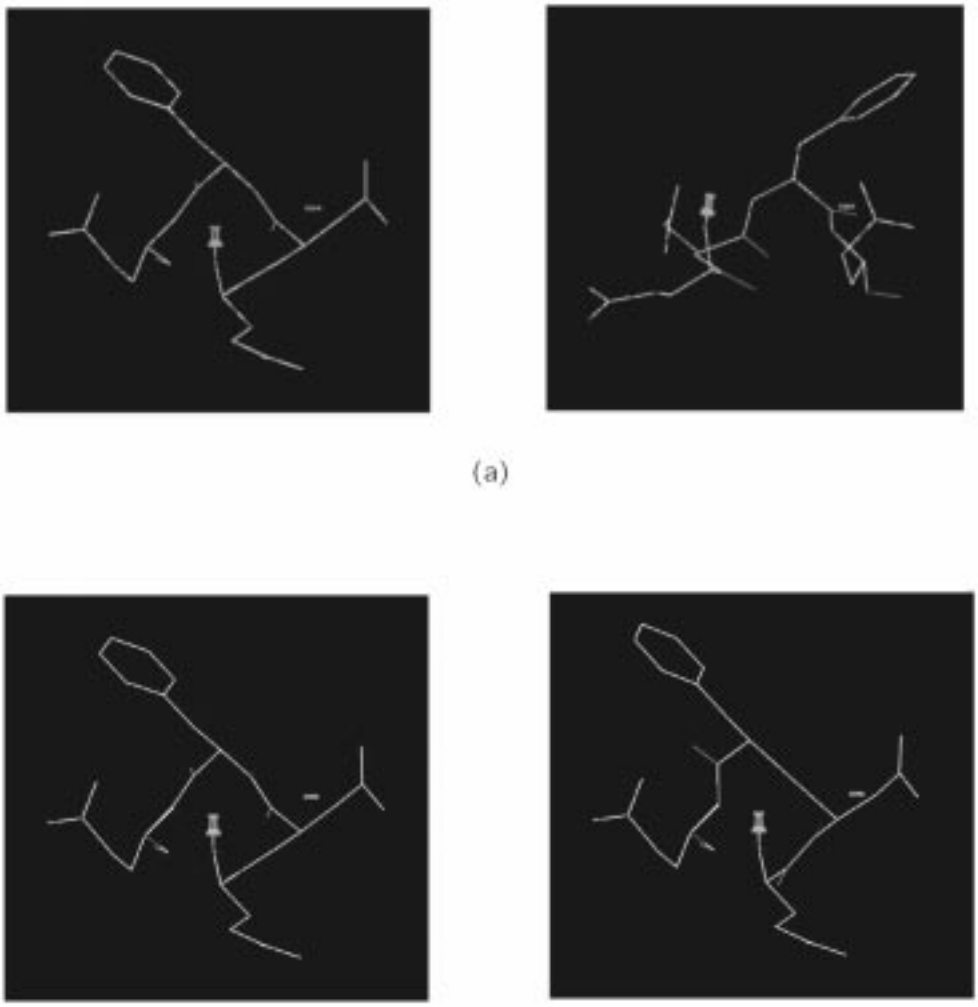

Figure 7: (a). (top) frontview and sideview of an α-helix, (b). (bottom) two possible conformations obtained by inverse kinematics procedure.

Number	Link Length	Offset Distance	Twist Angle
i	a_i	d_i	α_i
1	0.0	-5.81	8.67
2	0.0	9.44	70.05
3	0.0	-5.86	8.61
4	0.0	9.49	70.11
5	0.0	-5.78	8.68
6	0.0	9.42	70.12

Table 7: Denavit-Hartenberg Parameters of a Protein Segment

recovers the structure in the original data, it corresponds to the dihedral angles of a α-helix. The second set corresponds to an alternative conformation. These two solutions are showed in Fig. 7(b).

This technique can be used to exam the *loops* between secondary Structures. Since the loops usually can have irregular structures, they represent a difficult part of the study of protein conformation problem. Our algorithm computes all possible solutions to close the loop. We believe it is a useful tool in analyzing and designing loop structures.

6.4 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. 8(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 planar.

i	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
1	-51.19	124.30	-55.71	136.07	-50.14	124.48
2	-33.93	101.43	-38.57	115.47	-8.90	79.64

Table 8: The dihedral angles of the protein chain

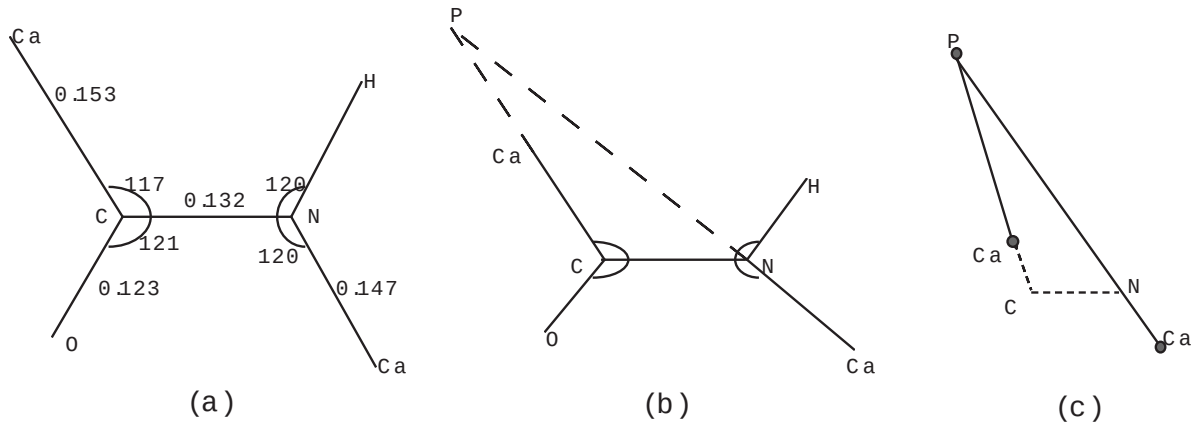

Figure 8: Geometry of a peptide unit

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 plane as showed in Fig. 8(b). (The angle between Cα-C and N-Cα is not shown accurately). 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. 8(c). Based on standard bond geometry in Fig 1, all the distances and angles in Fig. 8(c) can be easily computed. Based on this analysis, Theorem 3.1 can be directly applied for protein segments, while the planar peptide unit is still maintained.

If two adjacent Cα-C bond and C-Cα bond are given, the missing N atom position can be computed as shown in Fig. 8(b). Extend Cα-C bond to P, N is on the line of P-Cα and its distance to Ca is 1.47 A^o. As the two given bonds move further apart, computing intermediate atoms becomes more complicated. Based on Theorem 3.1, at most three intermediate points can be computed if all bonds are rotatable. For the case of protein, it means that there can be no more than four rotatable bonds between the two given end bonds. Otherwise , the intermediate atom positions can not be determined uniquely solely based on bond length and bond angle constraints (there is an entire family of solutions).

One example of molecular embedding is showed in Fig. 9. 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 atoms positions of these two bonds are listed in Table 9. 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

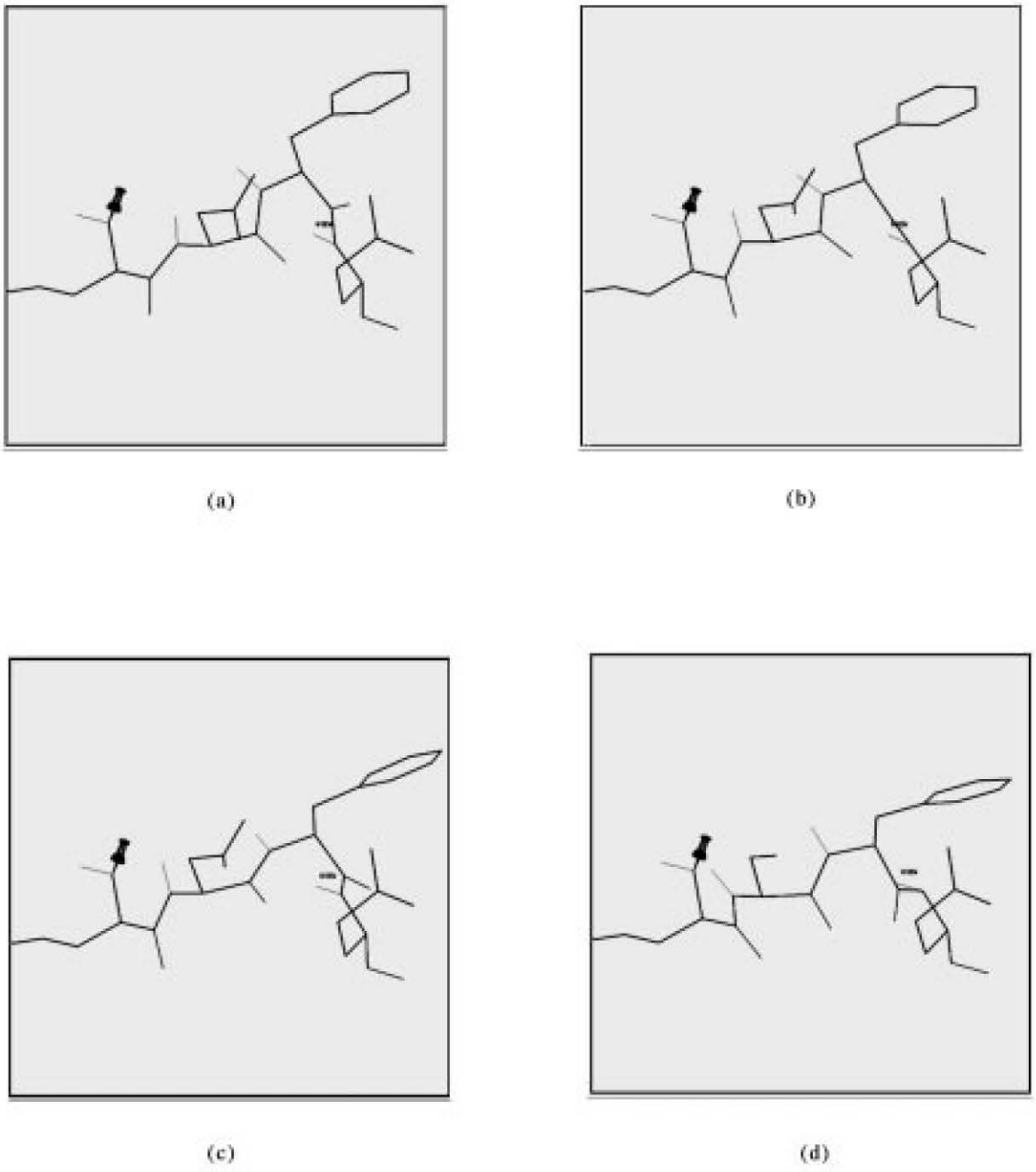

Figure 9: Four sets of possible atom positions of residues Asn and Phe. The Ca-C bond of Glu and N-Ca bond of Gln are fixed.

	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 9: Atom positions of two end bonds

that we made no assumption about the molecule geometry in the proof of Theorem 3.1 or in the procedure of solving inverse kinematics. This technique is also applicable for inferring backbone atom positions from the side chains.

7 Implementation and Performance

A system is implemented to perform real time local deformation. In this section, we will briefly discuss its implementation and performance.

7.1 User Interface

A molecular chain specified by a user is displayed on a projector screen. A robot ARM with 6 degrees of freedom is used as an input device. One terminus of the chain is fixed. The other terminus is manipulated by the user with the robot ARM. When the inverse kinematics procedure finds at least one solution to match the terminus controlled by the user, new atom positions are computed and displayed. Alternative solutions can be displayed by rotating a knob on the ARM.

7.2 System Description

The Denavit Hartenberg parameters and local frames for a given molecular chain are precomputed. Atom positions with respect to the local frames are also precomputed. The computation can be based on either atom coordinates in a protein datafile or standard protein geometry. The parameters are loaded at the system initialization stage. This precomputaion process takes very little computing time.

At runtime, whenever the user moves the ARM, a new pose of the terminus is computed. This

end pose along with the Denavit Hartenberg parameters of the molecular chain is passed to the inverse kinematics procedure. If solutions are found, atom positions for the new set of dihedral angles can be calculated with forward kinematics. These new atom positions are used to updated the display, and the system is ready to read another end pose from the ARM.

7.3 Performance

The system performance is determined mainly by the inverse kinematics procedure. The symbolic computation for the inverse kinematics is done only once, and the result is coded into the system. The bottleneck is the eigendecomposition of a matrix, which takes more than 80% of the total time. Both examples in section 4 (cyclohexane conformation and local deformation) involve computing eigenvalues for 32*32 matrices. The system takes 40-50ms on a SGI/ONYX workstation for each case.

8 Conclusions

In this paper, we have highlighted techniques for 3D manipulation of molecular chains with assumptions of rigid bond length and bond angles. We show the correspondence between the problems of local deformations, ring closure and molecular embedding with inverse kinematics of serial manipulators in robotics. Using results from kinematics, we present an algorithm for kinematic manipulation of molecular chains. It makes no assumption on the geometry of the chain and works very well in practice on chains with six or fewer dihedral angles. It can also be used to reduce the search space on chains with more dihedral angles. It has been implemented and we illustrate its real time performance on computing ring closure of cyclo-hexane and local deformation and embedding of short polypeptide chains.

9 Acknowledgements

We would like to thank Fred Brooks, Jane and David Richardson, Alex Tropsha for useful discussions.

References

[AA79] H. Albala and J. Angeles. Numerical solution to the input output displacement equation of the general 7r spatial mechanism. In *Proceedings of the Fifth World Congress on Theory of Machines and Mechanisms*, pages 1008–1011, 1979.

[BA82] Ulrich Burkert and Norman Allinger. *Molecular Mechanics*. ACS Monographs. Americal Chemical Society, Washington, D.C., 1982.

[Bio91] Biosym. *Insight*. Biosym Technologies Inc, San Diego, CA, 1991.

[BK85] R. Bruccoleri and M. Karplus. Chain closure with bond angle variations. *Macromolecules*, 18(12):2767–2773, 1985.

[CH] G.W. Crippen and T.F. Havel. Distance geometry and molecular conformation. Research Studies Press, New York.

[CMR89] L. Carballeira, R. Mosquera, and M. Rios. Molecular mechanics of peroxides. ii. cyclic compounds. *Journal of Computational Chemistry*, 10(7):911–920, 1989.

[Cra89] J.J. Craig. *Introduction to Robotics: Mechanics and Control*. Addison–Wesley Publishing Company, 1989.

[Cri81] G.M. Crippen. *Distance Geometry and Conformational Calculations*. Research Studies Press, Wiley, New York, 1981.

[Cri89] G.M. Crippen. Linearized embedding: A new metric matrix algorithm for calculating molecular conformations subject to geometric constraints. *Journal of Computational Chemistry*, 10(7):896–902, 1989.

[Cri92] G.W. Crippen. Exploring the conformation space of cylcoalkanes by linearized embedding. *Journal of Computational Chemistry*, 13(3):351–361, 1992.

[DC80] J. Duffy and C. Crane. A displacement analysis of the general spatial 7r mechanism. *Mechanisms and Machine Theory*, 15:153–169, 1980.

[DH55] J. Denavit and R.S. Hartenberg. A kinematic notation for lower–pair mechanisms based upon matrices. *Journal of Applied Mechanics*, 77:215–221, 1955.

[Dix08] A.L. Dixon. The eliminant of three quantics in two independent variables. *Proceedings of London Mathematical Society*, 6:49–69, 209–236, 1908.

[GL89] G.H. Golub and C.F. Van Loan. *Matrix Computations*. John Hopkins Press, Baltimore, 1989.

[GoS70] N. Go and H.A. Scherga. Ring closure and local conformational deformations of chain molecules. *Macromolecules*, 3(2):178–187, 1970.

[GoS73] N. Go and H.A. Scherga. Ring closure in chain molecules with c_n, i or s_{2n} symmetry. *Macromolecules*, 6(2):273–281, 1973.

[GoS78] N. Go and H.A. Scherga. Calculation of the conformation of cyclo-hexaglycyl. *Macromolecules*, 11(3):552–559, 1978.

[HK88] Allison E. Howard and Peter A. Kollman. An analysis of current methodologies for conformational searching of complex molecules. *Journal of Medicinal Chemistry*, 31(9):1669–1675, September 1988.

[LL88] H.Y. Lee and C.G. Liang. A new vector theory for the analysis of spatial mechanisms. *Mechanisms and Machine Theory*, 23(3):209–217, 1988.

[LS92] A.R. Leach and A.S. Smellie. 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, 1992.

[Man92] D. Manocha. *Algebraic and Numeric Techniques for Modeling and Robotics*. PhD thesis, Computer Science Division, Department of Electrical Engineering and Computer Science, University of California, Berkeley, May 1992.

[MC92] D. Manocha and J.F. Canny. Real time inverse kinematics of general 6r manipulators. In *Proceedings of IEEE Conference on Robotics and Automation*, pages 383–389, 1992.

[PD92] C.E. Peishoff and J.S. Dixon. Improvements to the distance geometry algorithm for conformational sampling of cyclic structures. *Journal of Computational Chemistry*, 13:565–569, 1992.

[Pie68] D. Pieper. *The kinematics of manipulators under computer control*. PhD thesis, Stanford University, 1968.

[Pri86] E.J.F. Primrose. On the input–output equation of the general 7r–mechanism. *Mechanisms and Machine Theory*, 21:509–510, 1986.

[Ric93] J. Richardson. Duke University. Personal Communication, 1993.

[RR89] M. Raghavan and B. Roth. Kinematic analysis of the 6r manipulator of general geometry. In *International Symposium on Robotics Research*, pages 314–320, Tokyo, 1989.

[RRS73] B. Roth, J. Rastegar, and V. Scheinman. On the design of computer controlled manipulators. In *On the Theory and Practice of Robots and Manipulators*, pages 93–113. First CISM IFToMM Symposium, 1973.

[Sau89] M. Saunders. Stochastic search for the conformations of bicyclic hydrocarbons. *Journal of Computational Chemistry*, 10(2):200–208, 1989.

[SLP86] H. Sklenar, R. Lavery, and B. Pullman. 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, 1986.

[SV89] M.W. Spong and M. Vidyasagar. *Robot Dynamics and Control*. John Wiley and Sons, 1989.

[Tri88] Tripos. *Sybyl*. Tripos Associates, St. Louis, MO, 1988.

[VDD83] P. Van Dooren and P. Dewilde. The eigenstructure of an arbitrary polynomial matrix: Computational aspects. *Lin. Alg. Appl.*, 50:545–579, 1983.

[We83] P.K. Weiner et. al. A distance geometry study of ring systems. *Tetrahedron*, 39(5):1113–1121, 1983.

[Wil59] J.H. Wilkinson. The evaluation of the zeros of ill–conditioned polynomials. parts i and ii. *Numer. Math.*, 1:150–166 and 167–180, 1959.

[WM91] C. Wampler and A.P. Morgan. Solving the 6r inverse position problem using a generic-case solution methodology. *Mechanisms and Machine Theory*, 26(1):91–106, 1991.