

---

# Resultants and Loop Closure

---

EVANGELOS A. COUTSIAS,<sup>1</sup> CHAOK SEOK,<sup>2</sup> MICHAEL J. WESTER,<sup>3</sup>  
KEN A. DILL<sup>4</sup>

<sup>1</sup>*Department of Mathematics and Statistics, University of New Mexico,  
Albuquerque, New Mexico 87131, USA*

<sup>2</sup>*Department of Chemistry, Seoul National University, Seoul, 151-742, Republic of Korea*

<sup>3</sup>*Office of Biocomputing, University of New Mexico, Albuquerque, New Mexico 87131, USA*

<sup>4</sup>*Department of Pharmaceutical Chemistry, University of California in San Francisco,  
San Francisco, California 94107, USA*

*Received 21 May 2005; accepted 27 May 2005*

*Published online 29 August 2005 in Wiley InterScience (www.interscience.wiley.com).*

*DOI 10.1002/qua.20751*

---

**ABSTRACT:** The problem of tripeptide loop closure is formulated in terms of the angles  $\{\tau_i\}_{i=1}^3$  describing the orientation of each peptide unit about the virtual axis joining the  $C_\alpha$  atoms. Imposing the constraint that at the junction of two such units the bond angle between the bonds  $C_\alpha-N$  and  $C_\alpha-C$  is fixed at some prescribed value  $\theta$  results in a system of three bivariate polynomials in  $u_i := \tan \tau_i/2$  of degree 2 in each variable. The system is analyzed for the existence of common solutions by making use of resultants, determinants of matrices composed of the coefficients of two (or more) polynomials, whose vanishing is a necessary and sufficient condition for the polynomials to have a common root. Two resultants are compared: the classical Sylvester resultant and the Dixon resultant. It is shown that when two of the variables are eliminated in favor of the third, a polynomial of degree 16 results. To each one of its real roots, there is a corresponding common zero of the system. To each such zero, there corresponds a consistent conformation of the chain. The Sylvester method can find these zeros among the eigenvalues of a  $24 \times 24$  matrix. For the Dixon approach, after removing extraneous factors, an optimally sized eigenvalue problem of size  $16 \times 16$  results. Finally, the easy extension to the more general problem of triaxial loop closure is presented and an algorithm for implementing the method on arbitrary chains is given. © 2005 Wiley Periodicals, Inc. *Int J Quantum Chem* 106: 176–189, 2006

**Key words:** loop closure; loop modeling; protein kinematics; polynomial systems; resultants

*Correspondence to:* E. A. Coutsias; e-mail: vageli@math.unm.edu

## 1. Introduction

**D**etermining the structure of a molecule from its chemical composition is the central problem of stereochemistry. Especially for large macromolecules with complex topologies and unique compositions, such as proteins and nucleic acids, the extreme complexity of the configuration space makes this problem one of the grand challenges of our time. The recent advances in the field of genomics have resulted in ever-increasing numbers of proteins whose sequence can be deduced from the genome, but whose structure and function are not understood. Computer prediction has thus become an increasingly alluring alternative to costly and time consuming experimental structure determination, such as by crystallographic or nuclear magnetic resonance (NMR) techniques.

Proteins are fascinating objects of study from a geometric standpoint. They are one-dimensional polypeptide chains that fold in specific ways to form three-dimensional structures, which are the building and functional blocks of all living things. These structures are characterized by certain motifs, such as  $\alpha$ -helices and  $\beta$ -sheets, but various other constructions are also present, e.g., a variety of types of turns, hinges, and bridges, underpinning and supporting the stability of larger structural units.

The basic links of the protein chain are the amino acids. Each amino acid is composed of a central carbon atom,  $C_\alpha$ , bonded to a hydrogen atom, an amino group, and a carboxyl group. This central carbon is also attached to a side chain, a variable group of atoms that differentiates one amino acid from another. The amino acids come in 80 different types, 20 of which together with some variants form all known proteins. Two special amino acids, glycine, which has a second hydrogen as its side chain, and proline, whose side chain of three carbon atoms loops around to connect to the amino nitrogen, are special structural elements, the first acting as a flexible hinge and the second as a provider of structural stiffness. Amino acids join by forming a peptide bond between the carboxyl and amino groups of successive residues. The resulting polypeptide chain inherits certain structural properties that must be understood before a systematic exploration of the conformation space of such chains can be attempted.

In studying the structure of a protein by computer, one typically considers ensembles of candi-

date structures. Although most proteins are believed to exist at a unique (native) conformation representing a global minimum of the Gibbs free energy, many proteins require flexibility in order to perform their function and exist within a range of alternative conformations. In general, deciding which are "native" or native-like conformations is often done on the basis of comparisons with the known structures of other proteins with strong sequence similarity if available (homology modeling), or by searching for a conformation minimizing some energy functional. Thus stated, the problem is one of optimization: given a certain scoring functional or energy, physics-based or information-theoretic, find the conformation or ensemble of conformations that minimize(s) that functional. Given the overwhelmingly large number of possible conformations even for a small-sized protein (<70 residues) and the extreme granularity of the energy landscape that is being searched for minimal points, it is important to refine and extend techniques further for the efficient exploration of the conformation space of proteins as an indispensable component of any methodology for determining their structure *in silico*.

From the modeling point of view, it is important to recognize that the various degrees of freedom (DoF) of a protein possess distinctly different flexibility properties. The stiffness constants associated with the deformation of dihedral angles from their optimal values are an order of magnitude smaller than those associated with deforming bond angles (with the exception of the  $\omega$  torsion angle associated with the peptide bond, whose constant is comparable to that of typical bond angles). In contrast, the constants associated with deforming bond angles are an order of magnitude smaller than those associated with deforming bond lengths [1]. Thus, bond-length distortion can be practically ignored, while bond angles may vary just a little: variations of  $\pm 5$ –10% account for almost all the bond-angle variability observed in the Top500 database of high-resolution, nonredundant protein structures [2]. As a result, the shape variability of a polypeptide chain is mostly due to the flexible torsion angles associated with the bonds at the  $C_\alpha$  carbon in the backbone ( $\phi$  and  $\psi$  torsions) and to various side-chain ( $\chi$ ) torsions not associated with rings [3].

Thus, the exploration of the low-energy conformation space of a protein involves mainly examination of all the different arrangements associated with deforming the  $\phi$ ,  $\psi$ , and  $\chi$  dihedrals, with small perturbations in the bond angles and the  $\omega$

torsions associated with peptide bonds also playing an important role, while even very small variations in bond lengths incur large energy penalties. Topological obstructions (steric clashes between distant atoms) pose obstacles to this exploration, and to achieve a native-like conformation often requires an arrangement of polar or aliphatic residues in three dimensions relative to each other and to the solvent in ways that minimize the electrostatic and other solvent-mediated interactions.

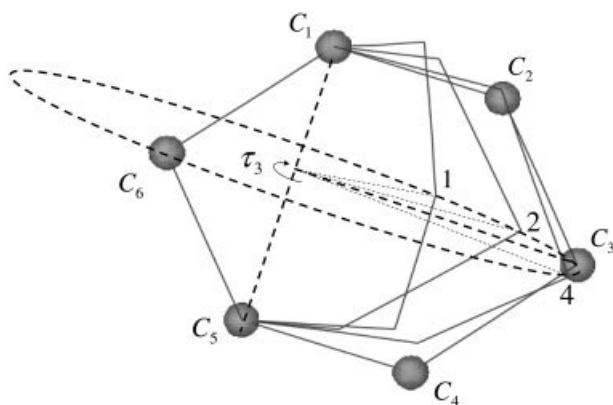
In the context of exploring the conformation space of proteins (a central task in protein folding) and other macromolecules, it is often useful to proceed by considering alternative structures differing only by the transposition of a few localized segments [4–7]. Such localized transpositions have been employed as localized moves in Monte Carlo methods [8] for efficient searches for the free energy minimum of a protein [5, 7, 9–12], and they offer the advantage that they can help achieve local refinement in the structure of a molecule without altering more distant elements.

It is often the case that the structure of a protein is known (or can be predicted with reasonable accuracy due to sequence similarity), except for certain regions: this is the missing loop problem. In other instances, a certain segment contains errors or needs to be determined with higher precision than afforded by a given model: this is the loop refinement problem. Problems of this type require the systematic exploration of the conformation space of a subsegment (loop) in a protein within the constraints imposed by the attachment to the rest of the molecule, whose structure is known. For this exploration, we can deform the chain locally by only changing the  $\phi$  and  $\psi$  torsions while sampling a certain range of values of other DoF. The problem of determining the structure of ring molecules is of a similar nature, as changes in various torsional and other DoF must respect the closure of the ring. Mathematically these problems are manifestations of the simple idea of deforming a polyhedral line by only changing its dihedrals such that certain distance (and/or angle) constraints between remote atoms or bonds are conserved, exactly or to within a certain tolerance. In molecular structure studies, the problem has a long history, especially since the pioneering work of Gō and Scheraga [4], and several algorithmic implementations are available. A relatively recent review can be found in Ref. [7]. The relationship of this problem to the problem of inverse kinematics in modern robotics has been investigated recently by several investigators [7,

13–18]. Because 6 DoF must be specified for placing an object at a given point with a given orientation, at least six torsions need to be changed to achieve a local deformation of a polypeptide chain so that it remains attached to fixed ends, while all other DoF (bond lengths, bond angles,  $\omega$  torsions) retain prescribed values. Concerted variation in seven (or more) torsions allows continuous motions, while when only six torsions are allowed to vary, discrete sets of at most 16 alternative arrangements will result. In robotics, these problems are called the **7R** and **6R** problems of inverse kinematics, respectively [19]. The most general solution known in the robotics literature [20] allows arbitrary arrangements of the torsional axes. In the molecular context, it is often the case that pairs of freely rotatable bonds share a vertex (atom), making the formulation of the problem considerably simpler.

Recently we proposed a general method of solving this problem when the six torsions are associated with three pairs of coterminal axes (meeting at points  $\mathbf{R}_i$ ,  $i = 1, 2, 3$ ), with arbitrary structure between the pairs [7]. This formulation leads to a system of three biquadratic polynomials in terms of the variables  $u_i = \tan(\tau_i/2)$ ,  $i = 1, 2, 3$ , where the  $\tau_i$  describe the orientation of the rigid units about virtual axes joining the points  $\mathbf{R}_i$  (see Fig. 2), which is a generalization of a system derived originally for the study of the conformation of octahedra [21]. We present the reduction of this polynomial system to a univariate polynomial of degree 16 by means of the Sylvester [22] and Dixon [23] resultants. In both cases, the implementation of this reduction with Maple is accomplished efficiently by first effecting certain algebraic simplifications as direct reductions proved to be intractable, leading either to numerically unstable results in the case of the Sylvester resultant or, for the Dixon resultant, to extraneous factors [24] whose presence led to a polynomial of degree 32. Moreover, in the Dixon case, the direct implementation masked the symmetry of the final form, which proved useful for deducing interesting geometric properties associated with certain special cases such as the conformational problem of a six-membered ring [13]. Further applications to structural problems in biological systems become possible with this formulation, and we are currently pursuing a connection with tensegrity studies of the stability of the cytoskeleton [25].

The smallest molecule for which alternative conformations may exist that satisfy all hard constraints is a hexagonal ring or loop with all its bonds rotatable (Fig. 1). Since each residue in a



**FIGURE 1.** Alternative conformations preserving bond lengths and bond angles of a slightly asymmetric hexagon. In the frame of the triangle  $C_5, C_6, C_1$ , atom  $C_3$  (shown at position 3) traces a circle about the axis  $C_1-C_5$ , parametrized by the torsion  $\tau_3$ . Conformations 1 and 4 are “boats,” while 2 and 3 are “chairs.”

protein backbone contributes two rotatable bonds, the smallest protein segment that can be considered possessing alternative conformations with respect to a given protein backdrop is a tripeptide unit whose first and last rotatable bonds are fixed in space. If other chains, e.g., a chain including a disulfide bridge, are considered, the rotatable bonds may also be found on segments of the side chains. The extension to these cases is straightforward as we discuss in the last section.

This article is organized as follows: in Section 2 we define the problem and derive the polynomial equations describing loop closure. In Section 3, we discuss their solution by Sylvester and Dixon resultants. Section 4 presents an algorithm for using loop closure to introduce local perturbations in a chain.

## 2. Loop Closure: Basic Definitions

When bond lengths are fixed, the conformational problem of a tripeptide unit in the protein backbone can be described in terms of 15 internal coordinates, i.e., the seven bond and eight dihedral angles. Since the closure conditions introduce six constraints, the conformation of the tripeptide unit can be completely characterized by specifying nine of these and determining values for the remaining six such that closure is satisfied. Conveniently, the three pairs of  $\phi, \psi$  torsions about each  $C_\alpha$  can be chosen for this task. Of course, other combinations of six

DoF are possible and have been used by various authors [14, 20].

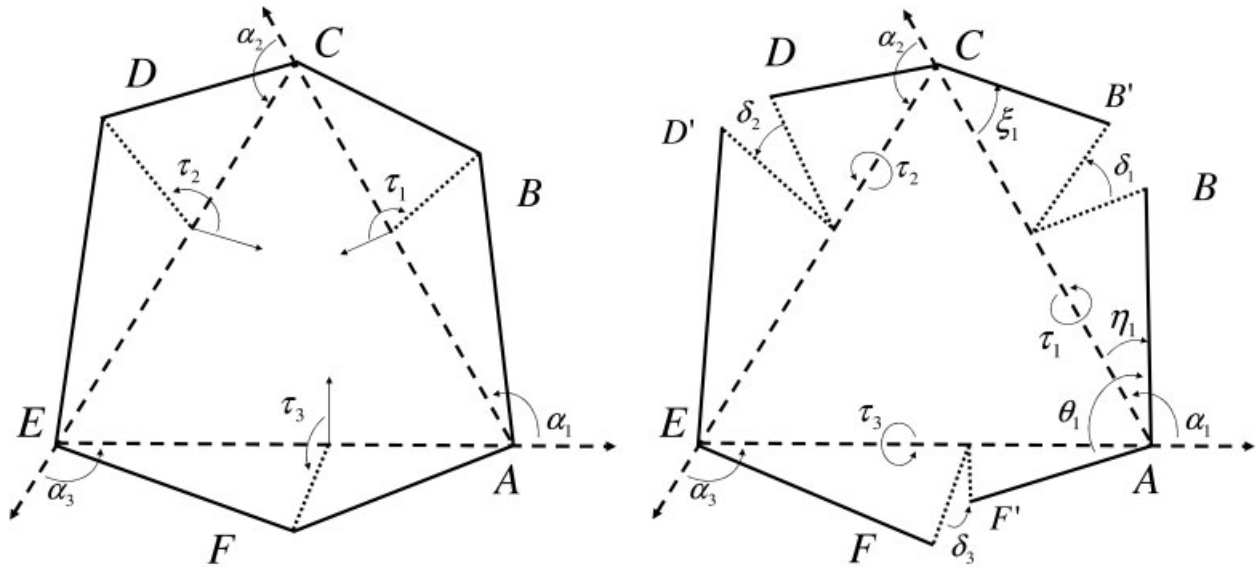
In this paper, we consider an important special case in which the **6R** problem has an intuitively simple description: consider all the motions of a chain molecule that involve changes in only six backbone torsions. If these are arranged so that they form three coterminal pairs, the segments between successive pairs will effectively form a coarser chain of three (closed case) or four (open case) rigid bodies, joined at the locations of the paired torsion axes. Illustrations are given in Figures 2 and 3 for a six-membered ring and a tripeptide example, respectively. For the former, the three rigid bodies are the triangles  $(ABC)$ ,  $(CDE)$ , and  $(EGA)$  (Fig. 2, left). For the latter, the four rigid bodies are  $(N_1 C_{\alpha 1})$ ,  $(C_{\alpha 1} C_1 N_2 C_{\alpha 2})$ ,  $(C_{\alpha 2} C_2 N_3 C_{\alpha 3})$ , and  $(C_{\alpha 3} C_3)$ .

We focus on the more general tripeptide case, which is similar to introducing nonplanarity to the three triangles while keeping each one rigid, (Fig. 2, right). If we require the two end segments of the chain  $(N_1 C_{\alpha 1})$  and  $(C_{\alpha 3} C_3)$  to remain at a fixed position relative to each other,  $(C_{\alpha 3} C_3 N_1 C_{\alpha 1})$  forms a third, virtual segment. Now, each of the three rigid units  $(C_{\alpha 1} C_1 N_2 C_{\alpha 2})$ ,  $(C_{\alpha 2} C_2 N_3 C_{\alpha 3})$ , and  $(C_{\alpha 3} C_3 N_1 C_{\alpha 1})$  has two junctions on it, attaching to the other two units. Define the line connecting the two junctions on a unit as the virtual axis of the unit  $(C_{\alpha 1}-C_{\alpha 2}, C_{\alpha 2}-C_{\alpha 3},$  and  $C_{\alpha 3}-C_{\alpha 1})$ . The motions of the first two segments relative to the rest of the chain can only be composed of individual rotations of each about their respective virtual axes  $(C_{\alpha 1}-C_{\alpha 2}$  and  $C_{\alpha 2}-C_{\alpha 3})$  or joint rotations of the two as a unit about the (fixed) virtual axis  $C_{\alpha 1}-C_{\alpha 3}$ . The angles involved in the statement of the problem are shown in Figure 3 in the context of the  $C_\alpha$  triangle, while Figure 2 (right) shows them for the most general chain since arbitrary but rigid structure can be introduced between the points  $B-B', D-D',$  and  $F-F'$ .

The angles  $\tau_i$  and  $\sigma_i$  are related to each other because  $C_{\alpha_i}-C_i$  and  $N_{i+1}-C_{\alpha_{i+1}}$  are rotated together as a rigid body. Figure 4(a) shows that  $\tau_i$  and  $\sigma_i$  are related by the simple relation

$$\sigma_i = \tau_i + \delta_i, \quad (1)$$

where  $\delta_i$  is the dihedral angle defined by the three vectors  $(C_i C_{\alpha i}, C_{\alpha i} C_{\alpha_{i+1}}, C_{\alpha_{i+1}} N_{i+1})$ . The generalization to arbitrary chains containing atoms with paired rotatable bonds (2) is straightforward and is discussed in detail in Ref. [7].

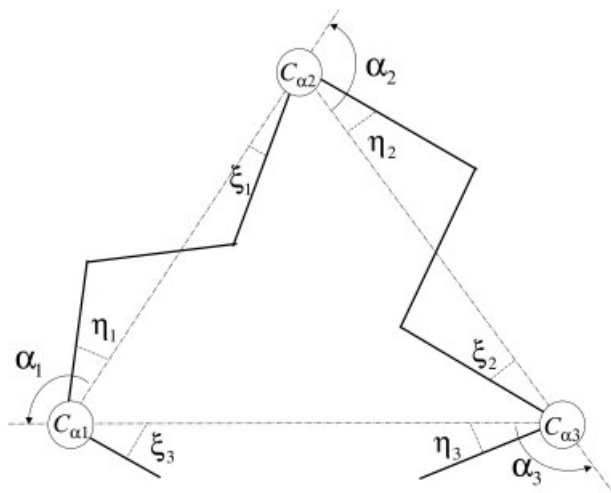


**FIGURE 2.** Definition of the three variables  $\tau_1, \tau_2,$  and  $\tau_3,$  and the three constraints  $\theta_1, \theta_2,$  and  $\theta_3$  in the simple hexagonal loop closure problem.

**2.1. COEFFICIENTS OF THE POLYNOMIALS**

Requiring that the bonds  $N_i-C_{\alpha_i}$  and  $C_{\alpha_i}-C_i$  maintain a given angle  $\theta_i$  leads to the relation [7]:

$$\begin{aligned} \cos \theta_i + \cos \eta_i \cos \xi_{i-1} \cos \alpha_i &= \sin \alpha_i (\sin \xi_{i-1} \cos \eta_i \cos \sigma_{i-1} \\ &+ \cos \xi_{i-1} \sin \eta_i \cos \tau_i) + \sin \xi_{i-1} \sin \eta_i (\sin \tau_i \sin \sigma_{i-1} \\ &+ \cos \alpha_i \cos \tau_i \cos \sigma_{i-1}). \end{aligned} \quad (2)$$



**FIGURE 3.** Definition of angle parameters  $\alpha_i, \eta_i,$  and  $\xi_i.$

Equation (2) can then be written as a double Fourier series

$$0 = a_i + b_i \cos \sigma_{i-1} + c_i \cos \tau_i + d_i \cos \sigma_{i-1} \cos \tau_i + e_i \sin \sigma_{i-1} \sin \tau_i \quad (3)$$

where the coefficients are

$$\begin{aligned} a_i &= -\cos \theta_i - \cos \eta_i \cos \xi_{i-1} \cos \alpha_i \\ b_i &= \sin \alpha_i \sin \xi_{i-1} \cos \eta_i \\ c_i &= \sin \alpha_i \cos \xi_{i-1} \sin \eta_i \\ d_i &= \cos \alpha_i \sin \xi_{i-1} \sin \eta_i \\ e_i &= \sin \xi_{i-1} \sin \eta_i. \end{aligned}$$

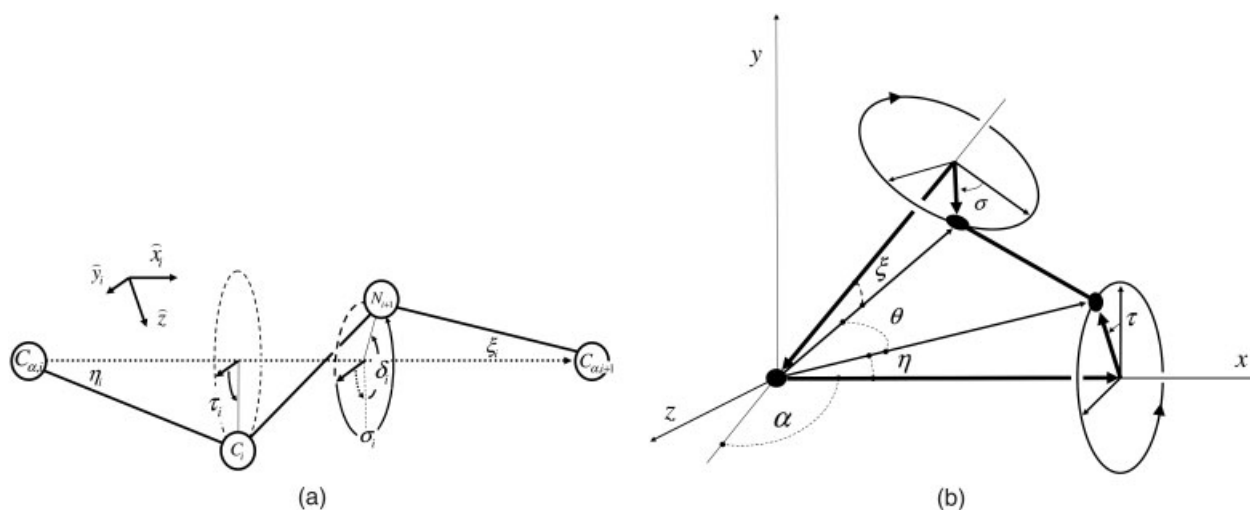
Equation (3) is now converted into polynomial form in the variables  $w_i, u_i,$  where

$$w_i := \tan \frac{\sigma_i}{2}, \quad u_i := \tan \frac{\tau_i}{2}.$$

To do this, introduce the half-angle formulas

$$\cos \tau = \frac{1 - u^2}{1 + u^2}, \quad \sin \tau = \frac{2u}{1 + u^2}, \quad u = \tan \frac{\tau}{2},$$





**FIGURE 4.** (a) A peptide unit along the  $C_{\alpha,i}-C_{\alpha,i+1}$  virtual bond. In the local coordinate system,  $\tau_i$  and  $\sigma_i$  are related by  $\sigma_i = \tau_i + \delta_i$ . (b) Geometric definitions at the  $C_{\alpha,i}$  junction. The black dot at the origin denotes the  $C_{\alpha,i}$  atom, while the circle centers correspond to  $N_i$  ( $\sigma$ -circle) and  $C_i$  ( $\tau$ -circle).

into (3) to arrive at a system of three biquadratics in  $w_i, u_i$  ( $i = 1, 2, 3$ ):

$$0 = a_i + b_i \left( \frac{1 - w_{i-1}^2}{1 + w_{i-1}^2} \right) + c_i \left( \frac{1 - u_i^2}{1 + u_i^2} \right) + d_i \left( \frac{1 - w_{i-1}^2}{1 + w_{i-1}^2} \right) \times \left( \frac{1 - u_i^2}{1 + u_i^2} \right) + e_i \left( \frac{2w_{i-1}}{1 + w_{i-1}^2} \right) \left( \frac{2u_i}{1 + u_i^2} \right),$$

or, equivalently,

$$0 = a_i(1 + w_{i-1}^2)(1 + u_i^2) + b_i(1 - w_{i-1}^2)(1 + u_i^2) + c_i(1 + w_{i-1}^2)(1 - u_i^2) + d_i(1 - w_{i-1}^2)(1 - u_i^2) + e_i 4w_{i-1}u_i. \quad (4)$$

Expanding and regrouping the results in Eq. (4),

$$A_i w_{i-1}^2 u_i^2 + B_i w_{i-1}^2 + C_i w_{i-1} u_i + D_i u_i^2 + E_i = 0, \quad (5)$$

where

$$A_i = a_i - b_i - c_i + d_i = -\cos \theta_i - \cos(\alpha_i - \xi_{i-1} - \eta_i)$$

$$B_i = a_i - b_i + c_i - d_i = -\cos \theta_i - \cos(\alpha_i - \xi_{i-1} + \eta_i)$$

$$C_i = 4e_i = 4 \sin \xi_{i-1} \sin \eta_i$$

$$D_i = a_i + b_i - c_i - d_i = -\cos \theta_i - \cos(\alpha_i + \xi_{i-1} - \eta_i)$$

$$E_i = a_i + b_i + c_i + d_i = -\cos \theta_i - \cos(\alpha_i + \xi_{i-1} + \eta_i).$$

We now eliminate the variables  $w_i$ . Using the twist transformation

$$w_i = \frac{u_i + \Delta_i}{1 - \Delta_i u_i}, \quad \Delta_i = \tan \delta_i / 2,$$

in Eq. (5), we find

$$A_i \left( \frac{u_{i-1} + \Delta_{i-1}}{1 - \Delta_{i-1} u_{i-1}} \right)^2 u_i^2 + B_i \left( \frac{u_{i-1} + \Delta_{i-1}}{1 - \Delta_{i-1} u_{i-1}} \right)^2 + C_i \frac{u_{i-1} + \Delta_{i-1}}{1 - \Delta_{i-1} u_{i-1}} u_i + D_i u_i^2 + E_i = 0.$$

Finally, the derivation of the coupled biquadratic polynomials is carried out by multiplying through by  $(1 - \Delta_{i-1} u_{i-1})^2$  and regrouping. Since

$$\Delta = \frac{\sin \delta}{1 + \cos \delta}, \quad \Delta^2 = \frac{1 - \cos \delta}{1 + \cos \delta},$$

we multiply the resulting expressions through by  $(1 + \cos \delta_{i-1})/2$  to arrive at the expression for the coefficients:

$$p_{22}^{(i)} = -\cos \theta_i - \cos \xi_{i-1} \cos(\alpha_i - \eta_i) - \cos \delta_{i-1} \sin \xi_{i-1} \sin(\alpha_i - \eta_i)$$

$$p_{21}^{(i)} = -2 \sin \delta_{i-1} \sin \xi_{i-1} \sin \eta_i$$

$$p_{20}^{(i)} = -\cos \theta_i - \cos \xi_{i-1} \cos(\alpha_i + \eta_i) - \cos \delta_{i-1} \sin \xi_{i-1} \sin(\alpha_i + \eta_i)$$

$$p_{12}^{(i)} = -2 \sin \delta_{i-1} \sin \xi_{i-1} \sin(\alpha_i - \eta_i)$$

$$p_{11}^{(i)} = 4 \cos \delta_{i-1} \sin \xi_{i-1} \sin \eta_i$$

$$p_{10}^{(i)} = -2 \sin \delta_{i-1} \sin \xi_{i-1} \sin(\alpha_i + \eta_i)$$

$$p_{02}^{(i)} = -\cos \theta_i - \cos \xi_{i-1} \cos(\alpha_i - \eta_i) + \cos \delta_{i-1} \sin \xi_{i-1} \sin(\alpha_i - \eta_i)$$

$$p_{01}^{(i)} = 2 \sin \delta_{i-1} \sin \xi_{i-1} \sin \eta_i$$

$$p_{00}^{(i)} = -\cos \theta_i - \cos \xi_{i-1} \cos(\alpha_i + \eta_i) + \cos \delta_{i-1} \sin \xi_{i-1} \sin(\alpha_i + \eta_i).$$

The polynomials themselves will be given by

$$P_1(u_3, u_1) = \sum_{k=0}^2 \left[ \sum_{j=0}^2 p_{jk}^{(1)} u_3^j \right] u_1^k = \sum_{k=0}^2 L_k u_1^k, \quad (6)$$

$$P_2(u_1, u_2) = \sum_{j=0}^2 \left[ \sum_{k=0}^2 p_{jk}^{(2)} u_2^k \right] u_1^j = \sum_{j=0}^2 M_j u_1^j, \quad (7)$$

and

$$P_3(u_2, u_3) = \sum_{j=0}^2 \left[ \sum_{k=0}^2 p_{jk}^{(3)} u_3^k \right] u_2^j = \sum_{j=0}^2 N_j u_2^j, \quad (8)$$

where

$$L_k := L_k(u_3) := \sum_{j=0}^2 p_{jk}^{(1)} u_3^j,$$

$$M_j := M_j(u_2) := \sum_{k=0}^2 p_{jk}^{(2)} u_2^k,$$

and

$$N_j := N_j(u_3) := \sum_{k=0}^2 p_{jk}^{(3)} u_3^k. \quad (9)$$

All three constraints must be satisfied at once. Therefore, we are looking for common real roots of these polynomials: each common root is a triplet  $(u_1, u_2, u_3)$  that satisfies all three polynomials. The corresponding torsions  $\{\tau_i\}_1^3$  give a placement of the three chains so that the bond angles are correct at the three “pivot” junctions.

In the next section, we describe the method of the resultant for finding these common zeros. Although deriving the resultant naturally leads to a 16th-degree polynomial in one of the  $u_i$ , it is possible to prove that a system of three biquadratics of the above form must generically have 16 zeros, real or complex. This follows from a straightforward application of the **BKK** theorem [26, 27].

### 3. Systems of Polynomials and Resultants

The vanishing of the resultant of a system of multivariate polynomials is a necessary and sufficient condition for the existence of a common root. For two polynomials,  $F_m(u)$  and  $F_n(u)$  of degrees  $m$  and  $n$ , to have a common root  $u$ , they must have a factor in common; i.e., there must exist polynomials  $g(u)$  and  $h(u)$  of degrees  $\leq n - 1$  and  $\leq m - 1$ , respectively, such that

$$gF_m + hF_n = 0.$$

This leads to a system of  $m + n$  linear homogeneous equations for determining the coefficients of  $g$  and  $h$ , and the resultant is the determinant of the matrix associated with that system. We demonstrate how this works for two second-order equations in a single variable. Let

$$f_1(u) = a_2 u^2 + a_1 u + a_0 = 0$$

$$f_2(u) = b_2 u^2 + b_1 u + b_0 = 0.$$

If these have a common root, say  $u^*$ , they must be of the form

$$f_1(u) = a_2(u - u^*)(u - u_1)$$

$$f_2(u) = b_2(u - u^*)(u - u_2)$$

and so there are two polynomials of degree 1,  $g(u) = b_2(u - u_2)$  and  $h(u) = -a_2(u - u_1)$ , such that

$$g(u)f_1(u) + h(u)f_2(u) = 0. \quad (10)$$

Since the roots are generally unknown, we simply write

$$g(u) = g_1u + g_0, \quad h(u) = h_1u + h_0,$$

and Eq. (10) becomes

$$(g_1u + g_0)(a_2u^2 + a_1u + a_0) + (h_1u + h_0)(b_2u^2 + b_1u + b_0) = 0$$

or, grouping like powers of  $u$  together,

$$(g_1a_2 + h_1b_2)u^3 + (g_1a_1 + g_0a_2 + h_1b_1 + h_0b_2)u^2 + (g_1a_0 + g_0a_1 + h_1b_0 + h_0b_1)u + (g_0a_0 + h_0b_0) = 0,$$

which can be written in the equivalent form

$$(g_1 \ g_0 \ h_1 \ h_0) \begin{pmatrix} a_2 & a_1 & a_0 & 0 \\ 0 & a_2 & a_1 & a_0 \\ b_2 & b_1 & b_0 & 0 \\ 0 & b_2 & b_1 & b_0 \end{pmatrix} \begin{pmatrix} u^3 \\ u^2 \\ u \\ 1 \end{pmatrix} = 0.$$

The left and right null vectors give, respectively, the coefficients of the two factor polynomials and the (common) zero of the original pair. The rank deficiency of the coefficient matrix (and the vanishing of its determinant, i.e., the resultant) is the necessary and sufficient condition for the existence of these null vectors.

Once the vanishing of the determinant above has been established, finding  $u$  is straightforward; discarding the third equation implied above for the right null vector (since it is dependent on the others), and moving the column associated with the component 1 to the right-hand side, we solve the resulting system for  $u$  using Cramer's rule:

$$u = \frac{\begin{vmatrix} a_2 & a_1 & 0 \\ 0 & a_2 & -a_0 \\ 0 & b_2 & -b_0 \end{vmatrix}}{\begin{vmatrix} a_2 & a_1 & a_0 \\ 0 & a_2 & a_1 \\ 0 & b_2 & b_1 \end{vmatrix}}.$$

### 3.1. SUCCESSIVE ELIMINATIONS AND THE SYLVESTER RESULTANT

The resultant of  $P_1$  and  $P_2$ , whose vanishing guarantees a common root in  $u_1$ , is given by the determinant

$$R_8(u_2, u_3) = \begin{vmatrix} L_2 & L_1 & L_0 & 0 \\ 0 & L_2 & L_1 & L_0 \\ M_2 & M_1 & M_0 & 0 \\ 0 & M_2 & M_1 & M_0 \end{vmatrix} = \begin{vmatrix} L_2 & L_0 \\ M_2 & M_0 \end{vmatrix}^2 - \begin{vmatrix} L_2 & L_1 \\ M_2 & M_1 \end{vmatrix} \begin{vmatrix} L_1 & L_0 \\ M_1 & M_0 \end{vmatrix}.$$

Since all the nonvanishing elements are products of two quadratics in  $u_2$  and two quadratics in  $u_3$ , the resultant is a biquartic in these variables, and has the form

$$R_8(u_2, u_3) = \sum_{j,k=0}^4 q_{jk} u_2^j u_3^k.$$

Here, the  $5 \times 5 = 25$  quantities  $q_{jk}$  are found in terms of products of the  $a_{jk} := p_{jk}^{(1)}$  and  $b_{jk} := p_{jk}^{(2)}$  by expressing  $R_8$  as a sum of six tensor products.

We write  $R_8$  as a quartic in  $u_2$  introducing the functions  $Q_j$ , quartics in  $u_3$ :

$$R_8 = \sum_{j=0}^4 \left( \sum_{k=0}^4 q_{jk} u_3^k \right) u_2^j =: \sum_{j=0}^4 Q_j u_2^j. \quad (11)$$

The final resultant, which eliminates  $u_2$  to arrive at a degree 16 polynomial in  $u_3$  is given by

$$R_{16} = \det(S),$$



where the matrix  $S$  is given as

$$S(u_3) := \sum_{k=0}^4 S_k u_3^k = \begin{pmatrix} N_2 & N_1 & N_0 & 0 & 0 & 0 \\ 0 & N_2 & N_1 & N_0 & 0 & 0 \\ 0 & 0 & N_2 & N_1 & N_0 & 0 \\ 0 & 0 & 0 & N_2 & N_1 & N_0 \\ Q_4 & Q_3 & Q_2 & Q_1 & Q_0 & 0 \\ 0 & Q_4 & Q_3 & Q_2 & Q_1 & Q_0 \end{pmatrix}, \quad (12)$$

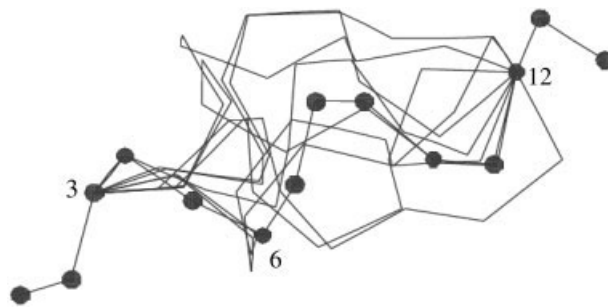
so that

$$S_k := \begin{pmatrix} c_{2k} & c_{1k} & c_{0k} & 0 & 0 & 0 \\ 0 & c_{2k} & c_{1k} & c_{0k} & 0 & 0 \\ 0 & 0 & c_{2k} & c_{1k} & c_{0k} & 0 \\ 0 & 0 & 0 & c_{2k} & c_{1k} & c_{0k} \\ q_{4k} & q_{3k} & q_{2k} & q_{1k} & q_{0k} & 0 \\ 0 & q_{4k} & q_{3k} & q_{2k} & q_{1k} & q_{0k} \end{pmatrix},$$

where we defined  $c_{ij} := p_{ij}^{(3)}$ , with  $c_{i3} = c_{i4} = 0$ ,  $i = 0, 1, 2$ . These matrices can be used directly in the matrix polynomial approach, which finds the solutions as eigenvalues of a ‘‘companion’’ matrix pencil. The computation of the polynomial coefficients for the direct approach requires some additional computations given in Ref. [7].

From these expressions, whose computation involves only four distinct  $2 \times 2$  determinants, we can compute the final polynomial. This computation can be done analytically, by deriving the lengthy expressions for the coefficients of the final polynomial in terms of the coefficients of the original polynomials. These analytical expressions can be useful, especially if one wants to study the effect of varying parameters on the behavior of the solution of the tripeptide loop closure. For the calculation shown in Figure 5, the computation of the coefficients was done numerically, using the numerical algorithm `tripep_closure.f90` [29], based on the Sylvester resultant.

Once  $u_3$  is obtained,  $u_2$  and  $u_1$  can be found via the equations [7]:



**FIGURE 5.** Example of the general chain loop closure: a chain of 14 atoms corresponding to two residues in an RNA molecule. Here, atoms 3, 6, and 12 act as pivots for the move. Atoms shown for original,  $45^\circ$  helix together with nine alternative conformations.

$$u_2 = \frac{\begin{vmatrix} N_2 & N_1 & N_0 & 0 & 0 \\ 0 & N_2 & N_1 & N_0 & 0 \\ 0 & 0 & N_2 & N_1 & 0 \\ 0 & 0 & 0 & N_2 & -N_0 \\ 0 & Q_4 & Q_3 & Q_2 & -Q_0 \end{vmatrix}}{\begin{vmatrix} N_2 & N_1 & N_0 & 0 & 0 \\ 0 & N_2 & N_1 & N_0 & 0 \\ 0 & 0 & N_2 & N_1 & N_0 \\ 0 & 0 & 0 & N_2 & N_1 \\ 0 & Q_4 & Q_3 & Q_2 & Q_1 \end{vmatrix}},$$

where  $N_j$  and  $Q_j$  are functions of  $u_3$  defined by Eqs. (9) and (11), respectively, and

$$u_1 = \frac{\begin{vmatrix} L_2 & L_1 & 0 \\ 0 & L_2 & -L_0 \\ 0 & M_2 & -M_0 \end{vmatrix}}{\begin{vmatrix} L_2 & L_1 & L_0 \\ 0 & L_2 & L_1 \\ 0 & M_2 & M_1 \end{vmatrix}},$$

where  $L_j$  and  $M_j$  are functions of  $u_3$  and  $u_2$ , respectively.

### 3.2. SIMULTANEOUS ELIMINATION AND THE DIXON RESULTANT

The Dixon resultant provides a powerful alternative to the previous discussion. The idea is to find the condition on the coefficients of  $n$  polynomials  $p_i(x_1, \dots, x_{n-1})$  in  $n - 1$  variables for the existence of a common root,  $(x_1^*, \dots, x_{n-1}^*)$ . In practice, to use this method for the solvability of a system of  $n$  polynomials in  $n$  variables, the variable  $x_n$  is treated as a parameter and the resulting condition is a

polynomial in  $x_n$ . Depending on the method employed, this polynomial may exhibit various kinds of singular or redundant behavior and may only constitute a necessary (but not sufficient) condition for the existence of a common root. We present a brief outline of Dixon's method as developed in the paper by Kapur et al. [30] and give a Dixon resultant for the triaxial loop closure problem of minimal size.

Consider a system of polynomials  $p_i(x_1, \dots, x_{n-1})$  with  $i = 1, \dots, n$ . Form the cancellation matrix

$$A(x_1, \dots, x_{n-1}, y_1, \dots, y_{n-1}) = \begin{bmatrix} p_1(x_1, x_2, \dots, x_{n-1}) & \cdots & p_n(x_1, x_2, \dots, x_{n-1}) \\ p_1(y_1, x_2, \dots, x_{n-1}) & \cdots & p_n(y_1, x_2, \dots, x_{n-1}) \\ p_1(y_1, y_2, \dots, x_{n-1}) & \cdots & p_n(y_1, y_2, \dots, x_{n-1}) \\ \cdots & \cdots & \cdots \\ p_1(y_1, y_2, \dots, y_{n-1}) & \cdots & p_n(y_1, y_2, \dots, y_{n-1}) \end{bmatrix}$$

and define the Dixon polynomial

$$\delta(x_1, \dots, x_{n-1}, y_1, \dots, y_{n-1}) = \frac{\det A}{\prod_{i=1}^{n-1} (x_i - y_i)} := \sum_{l \in \mathcal{L}, m \in \mathcal{M}} D_{lm} x^l y^m,$$

where  $l, m$  are ordered  $(n-1)$ -tuples,  $l = (l_1, l_2, \dots, l_{n-1})$ ,  $m = (m_1, m_2, \dots, m_{n-1})$  so that  $x^l := x_1^{l_1} x_2^{l_2} \cdots x_{n-1}^{l_{n-1}}$  and, similarly for  $y^m$ , with  $\mathcal{L}$  the set of all exponents corresponding to  $x$ -monomials present in the Dixon polynomial and similarly for  $\mathcal{M}$  and the  $y$ -monomials. In general, there is no guarantee that the matrix of coefficients

$$\mathbf{D} := [D_{lm}]$$

is square. If it turns out to be so, its determinant,

$$\mathcal{D} := \det \mathbf{D}$$

or, more precisely, the irreducible part of  $\mathcal{D}$  as a function of the polynomial coefficients, is called the Dixon resultant.

In general, the dimensions associated with  $l$  and  $m$  could be different, leading to a nonsquare Dixon matrix. However, for the triaxial loop closure problem this turns out to not be the case; as we shall see, the Dixon matrix is  $8 \times 8$  and composed of quadratic terms in  $u_3$ . We begin with the three poly-

nomials (6), (7), and (8), written with their dependence on  $u_3$  not shown explicitly:

$$P_1(u_3, u_1) = \sum_{k=0}^2 \left[ \sum_{j=0}^2 p_{jk}^{(1)} u_3^j \right] u_1^k = \sum_{k=0}^2 L_k u_1^k =: p_3(u_1),$$

$$P_2(u_1, u_2) = \sum_{j=0}^2 \left[ \sum_{k=0}^2 p_{jk}^{(2)} u_2^k \right] u_1^j = \sum_{j=0}^2 \sum_{k=0}^2 M_{jk} u_1^j u_2^k =: p_1(u_1, u_2),$$

and

$$P_3(u_2, u_3) = \sum_{j=0}^2 \left[ \sum_{k=0}^2 p_{jk}^{(3)} u_3^k \right] u_2^j = \sum_{j=0}^2 N_j u_2^j =: p_2(u_2),$$

where

$$L_k(u_3) := \sum_{j=0}^2 p_{jk}^{(1)} u_3^j,$$

$$M_{jk} := p_{jk}^{(2)},$$

and

$$N_j(u_3) := \sum_{k=0}^2 p_{jk}^{(3)} u_3^k.$$

We now have the cancellation matrix

$$A(u_1, u_2, v_1, v_2) = \begin{bmatrix} p_1(u_1, u_2) & p_2(u_2) & p_3(u_1) \\ p_1(v_1, u_2) & p_2(u_2) & p_3(v_1) \\ p_1(v_1, v_2) & p_2(v_2) & p_3(v_1) \end{bmatrix}$$

and the Dixon polynomial

$$\delta(u_1, u_2, v_1, v_2) = \frac{\det A}{(u_1 - v_1)(u_2 - v_2)} := \sum_{l \in \mathcal{L}, m \in \mathcal{M}} D_{lm} u^l v^m$$

It turns out that this matrix results in a system that is too large. Indeed, a symbolic computation of the resultant produces a polynomial of degree 32. Careful hand optimization of the formulas with the left and right null vectors defined as

$$V_i := [1 \quad v_1 \quad v_2 \quad v_1v_2 \quad v_2^2 \quad v_1v_2^2 \quad v_2^3 \quad v_1v_2^3]^T$$

and

$$V_r := [1 \quad u_1 \quad u_1^2 \quad u_1^3 \quad u_2 \quad u_1u_2 \quad u_1^2u_2 \quad u_1^3u_2]^T$$

helps isolate an extraneous factor of degree 16, resulting in a much more elegant and compact expression for the irreducible part  $\mathbf{R}_D$ , that is, the Dixon resultant:

$$\mathcal{D} := (M_{22}N_2L_2)^4 \det \mathbf{R}_D$$

with

$$\mathbf{R}_D := [D_{lm}] = \begin{bmatrix} 0 & A_0 & A_1 & A_2 & 0 & B_0 & B_1 & B_2 \\ A_0 & A_1 & A_2 & 0 & B_0 & B_1 & B_2 & 0 \\ 0 & B_0 & B_1 & B_2 & 0 & C_0 & C_1 & C_2 \\ B_0 & B_1 & B_2 & 0 & C_0 & C_1 & C_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & D_0 & D_1 & D_2 \\ 0 & 0 & 0 & 0 & D_0 & D_1 & D_2 & 0 \\ 0 & D_0 & D_1 & D_2 & 0 & 0 & 0 & 0 \\ D_0 & D_1 & D_2 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

where, for  $i = 0, 1, 2$ :

$$A_i := M_{i1}N_0 - M_{i0}N_1,$$

$$B_i := M_{i2}N_0 - M_{i0}N_2,$$

$$C_i := M_{i2}N_1 - M_{i1}N_2,$$

$$D_i := L_i.$$

These coefficients are all quadratic in the third variable,  $u_3$ . In this way, the reduced Dixon matrix  $R_D$  can be written as a matrix polynomial. We have

$$A_i = A_{i2}u_3^2 + A_{i1}u_3 + A_{i0},$$

and similarly for  $B, C, D$ . The coefficients are defined as

$$A_{ij} := p_{i1}^{(2)}p_{0j}^{(3)} - p_{i0}^{(2)}p_{1j}^{(3)},$$

$$B_{ij} := p_{i2}^{(2)}p_{0j}^{(3)} - p_{i0}^{(2)}p_{2j}^{(3)},$$

$$C_{ij} := p_{i2}^{(2)}p_{1j}^{(3)} - p_{i1}^{(2)}p_{2j}^{(3)},$$

$$D_{ij} := p_{ji}^{(1)}.$$

Then

$$\mathbf{R}_D := \mathbf{R}_2u_3^2 + \mathbf{R}_1u_3 + \mathbf{R}_0$$

$$= \sum_{i=0}^2 u_3^i \begin{bmatrix} 0 & A_{0i} & A_{1i} & A_{2i} & 0 & B_{0i} & B_{1i} & B_{2i} \\ A_{0i} & A_{1i} & A_{2i} & 0 & B_{0i} & B_{1i} & B_{2i} & 0 \\ 0 & B_{0i} & B_{1i} & B_{2i} & 0 & C_{0i} & C_{1i} & C_{2i} \\ B_{0i} & B_{1i} & B_{2i} & 0 & C_{0i} & C_{1i} & C_{2i} & 0 \\ 0 & 0 & 0 & 0 & 0 & D_{0i} & D_{1i} & D_{2i} \\ 0 & 0 & 0 & 0 & D_{0i} & D_{1i} & D_{2i} & 0 \\ 0 & D_{0i} & D_{1i} & D_{2i} & 0 & 0 & 0 & 0 \\ D_{0i} & D_{1i} & D_{2i} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The above form of the Dixon resultant has been verified by symbolic calculation, using Maple [31, 32]. We note that the presence of the extraneous factor was not precluded by the general theory because the triaxial loop closure system is not generic  $n$ -degree [24].

### 3.2.1. Generalized Eigenproblem Formulation

To apply this formulation to the triaxial loop closure problem, we first note that the Sylvester method discussed previously amounts to essentially applying the Dixon method of the preceding section in turns: first  $u_1$  is eliminated between  $P_1(u_3, u_1)$  and  $P_2(u_1, u_2)$ , resulting in a new polynomial in  $u_2$  and  $u_3$ ,  $R_8(u_2, u_3)$ . Then  $u_2$  is eliminated between  $R_8(u_2, u_3)$  and  $P_3(u_2, u_3)$ , producing a polynomial in  $u_3$  alone. The computational characteristics of this process depend clearly on the order in which variables are eliminated, and in some cases [23] this has been reported to cause numerical difficulties. However, there is another, algebraically subtler effect resulting from the “unbalancing” of the process by this preferential order of elimination: the resultant matrix can be written as a matrix polynomial. In the successive elimination method, the final matrix (whose determinant is the resultant) arises from elimination between a biquadratic and a biquartic (by which we mean quartic in both variables). It therefore has size  $6 \times 6$  and can be expanded as a matrix polynomial of degree 4. However, in the determinantal expansion every term is the product of two quartic and four quadratic polynomials, so the resulting polynomial is of degree 16, as expected from the **BKK** root count for a polynomial system with elements in the  $2 \times 2 \times 2$  box.

As is well known, the determinant of a matrix polynomial can be written as the characteristic polynomial of a companion matrix, in this case of size  $(4 \times 6)^2$ . The root system of this polynomial contains the  $u_3$ -components of the common roots. However, these must be separated from extraneous

zeros, whose computation adds unnecessary overhead to the method. As a result, the alternative offered by the Dixon resultant becomes attractive for two reasons: it is both simpler to state and compute, and all its terms are quadratic in  $u_3$ . This latter property implies that the companion matrix will be block  $2 \times 2$  and, since the blocks are of size  $8 \times 8$ , the resulting generalized eigenproblem has dimension 16 and is a constant multiple of the resultant. However, here one needs to identify the extraneous factor  $(M_{22}N_2L_2)^4$  resulting from the computation of the Dixon resultant, whose presence made the problem practically uncomputable by automatic means.

We now give the formulation of the above polynomial equations as generalized eigenproblems. Following Manocha [13], we write  $R_{16}(u_3)$  as a determinant of a matrix polynomial with matrix coefficients  $S_k$ :

$$\det\left(\sum_{k=0}^4 S_k u_3^k\right) = 0,$$

which for a generic set of matrices  $S_k$  is equivalent to

$$\det(\mathcal{B}u_3 - \mathcal{A}) = 0$$

with

$$\mathcal{B} := \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & S_4 \end{pmatrix},$$

$$\mathcal{A} := \begin{pmatrix} 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \\ -S_0 & -S_1 & -S_2 & -S_3 \end{pmatrix},$$

where all blocks are of size  $6 \times 6$ .

For the Dixon version,

$$\det\left(\sum_{k=0}^2 \mathbf{R}_k u_3^k\right) = 0,$$

which for a generic set of matrices  $\mathbf{R}_k$  is equivalent to

$$\det(\mathcal{B}_D u_3 - \mathcal{A}_D) = 0$$

with

$$\mathcal{B}_D := \begin{pmatrix} I & 0 \\ 0 & \mathbf{R}_2 \end{pmatrix}, \quad \mathcal{A}_D := \begin{pmatrix} 0 & I \\ -\mathbf{R}_0 & -\mathbf{R}_1 \end{pmatrix},$$

where all blocks are of size  $8 \times 8$ .

The resulting generalized eigenproblems, of form  $u_3 \mathcal{B}Z = \mathcal{A}Z$ , can be solved numerically with the LAPACK routine `dggev.f`, for example. Having found the roots, the reconstruction of the chain with the altered torsion angles is straightforward. However, in practical applications, one must guard against solutions made unfeasible by the clashing of distant parts of the chain due to the rearrangement. Especially when additional driver angles are introduced, solution branches can disappear or new solutions can bifurcate from old branches. Topological considerations become important and powerful continuation and other topological methods can be brought to bear to identify feasible solution sets and speed up the solution process [33].

## 4. Reconstructing the Chain

In the previous sections, we saw how to reduce the problem to either a 16th-degree polynomial or a generalized eigenproblem for  $u_3$ , and how to then determine for each solution the corresponding values of  $u_2$  and  $u_1$ .

Reconstructing the configuration corresponding to each triplet is a straightforward geometrical exercise, which we describe here for completeness:

### 4.1. ALGORITHM: TRIAXIAL LOOP CLOSURE

1. Define Cartesian coordinates for a chain of  $N$  atoms,

$$\mathcal{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N],$$

where it is possible that  $\mathbf{x}_1$  is connected to  $\mathbf{x}_N$ ; i.e., the chain may be a closed ring.

2. Identify three atoms in the chain that will serve as pivots for the loop closure. These atoms must be connected to their nearest neighbors via rotatable bonds. Let these be defined as  $\mathbf{a}_1 \equiv \mathbf{x}_a$ ,  $\mathbf{a}_2 \equiv \mathbf{x}_b$  and  $\mathbf{a}_3 \equiv \mathbf{x}_c$ , with  $1 \leq a < b < c \leq N$ , where we require that no two of these pivotal atoms are neighbors.
3. Subdivide the chain into three subchains:

$$\mathcal{X}_1 = [\mathbf{x}_a, \dots, \mathbf{x}_b], \quad \mathcal{X}_2 = [\mathbf{x}_b, \dots, \mathbf{x}_c],$$

$$\mathcal{X}_3 = [\mathbf{x}_c, \dots, \mathbf{x}_a].$$

Note that the third chain is looped back, and a bond is introduced between  $\mathbf{x}_N$  and  $\mathbf{x}_1$ . This may be a real bond or a virtual bond, i.e., the third piece may be a contiguous chain (part of a closed ring), or a virtual chain formed by connecting the beginning and end pieces. We also introduce the notation  $\mathbf{n}_i$  (respectively,  $\mathbf{c}_i$ ) for the previous (respectively, next) neighbors of the atoms  $\mathbf{a}_i$ . It is possible that  $\mathbf{c}_i = \mathbf{n}_{i+1}$  for some  $i = 1, 2$ , and/or 3.

4. Chains may now be perturbed in an arbitrary fashion: for example, some dihedrals may be changed, or some bond angles or even bond lengths can be altered in a prescribed fashion. The new configuration will be known in Cartesian form, but the perturbed chains will, in general, no longer be cohesive with each other.
5. If subchain  $\mathcal{X}_3$  is perturbed, it must be anchored to an absolute coordinate system in some fashion. In the case of an open chain, this means that the actual ends are anchored to some fixed positions as is typical for missing-loop closure problems. For a ring molecule, this serves to fix the arbitrary affine transformation that may be applied to the entire molecule. As a result of this procedure, the atoms  $\mathbf{a}_3$  and  $\mathbf{a}_1$  as well as  $\mathbf{c}_3 \equiv \mathbf{x}_{c+1}$  and  $\mathbf{n}_1 \equiv \mathbf{x}_{a-1}$  become fixed to absolute locations that will subsequently serve as the anchors of the loop closure algorithm. From this point on, subchain  $\mathcal{X}_3$  remains fixed.
6. Calculate the triangle scaffold for loop closure in its own body frame. That is, form the vectors

$$\mathbf{d}_i = \mathbf{a}_{i+1} - \mathbf{a}_i, \quad i = 1, 2, 3,$$

where  $\mathbf{a}_4 \equiv \mathbf{a}_1$ , and also calculate their norms,  $d_i = \|\mathbf{d}_i\|$ . Calculate the exterior angles

$$\alpha_i = \cos^{-1} \frac{\mathbf{d}_{i-1} \cdot \mathbf{d}_i}{d_{i-1}d_i}$$

(with  $\mathbf{d}_0 \equiv \mathbf{d}_3$ ) at vertices  $\mathbf{A}_i$ ,  $i = 1, 2, 3$  corresponding to the atoms  $\mathbf{a}_i$ , where  $\mathbf{A}_3$  is placed at the origin [coordinates (0, 0, 0)],  $\mathbf{A}_1$  along the positive  $x$ -axis [coordinates ( $d_3$ , 0, 0)] and

$\mathbf{A}_2$  on the positive  $xy$ -half-plane [coordinates ( $-d_2 \cos \alpha_3$ ,  $d_2 \sin \alpha_3$ , 0)].

7. Reduce subchains  $\mathcal{X}_1$  and  $\mathcal{X}_2$  each to its own body frame. These are defined in terms of a local  $\hat{x}_i, \hat{y}_i, \hat{z}_i$  orthogonal coordinate system, where

$$\hat{x}_i = \frac{\mathbf{d}_i}{d_i}, \quad \hat{z}_i = \frac{\hat{x}_i \times (\mathbf{c}_i - \mathbf{a}_i)}{\|\hat{x}_i \times (\mathbf{c}_i - \mathbf{a}_i)\|},$$

$$\hat{y}_i = \hat{z}_i \times \hat{x}_i \quad i = 1, 2.$$

Then the rotation-to-body matrices are defined by

$$\mathcal{Q}_i = [\hat{x}_i, \hat{y}_i, \hat{z}_i],$$

and the transformation to body coordinates for each subchain is given by

$$\mathcal{X}_i^b \equiv \mathcal{Q}_i^T(\mathcal{X}_i - \mathbf{a}_i).$$

8. Compute the remaining loop closure polynomial parameters ( $i = 1, 2, 3$ ):

$$\delta_{i-1} = \text{dihedral } \angle(\mathbf{c}_{i-1}, \mathbf{a}_{i-1}, \mathbf{a}_i, \mathbf{n}_i)$$

$$\eta_i = \text{bond } \angle(\mathbf{c}_i, \mathbf{a}_i, \mathbf{a}_{i+1})$$

$$\xi_{i-1} = \text{bond } \angle(\mathbf{a}_{i-1}, \mathbf{a}_i, \mathbf{n}_i)$$

using the normal definitions for dihedral and bond angles, while the constraint parameters,  $\theta_i$ , must be computed from the initial chain as the bond angles at the pivotal atoms (and their values may be retained or perturbed as desired).

9. Now solve the loop closure problem, and determine the number of real solution triplets,  $\tau_1, \tau_2, \tau_3$ . For each triplet:

a. Rotate  $\mathbf{d}_1$  and  $\mathbf{d}_2$  by angle  $\pi - \tau_3$  and compute the frame vector  $\hat{\mathbf{z}}$ .

b. Place atom  $\mathbf{a}_2$  using frame  $\mathcal{Q}_3$ :

$$\mathbf{a}_2 \leftarrow \mathbf{a}_3 + \mathcal{Q}_3 \mathcal{R}_x(\pi - \tau_3) \mathbf{A}_2$$

c. Rotate chains  $\mathcal{X}_i^b$ ,  $i = 1, 2$  about their respective body  $x$ -axes by the angles  $\tau_i$  and place them in the rotated triangle frame using corresponding edges and  $\hat{\mathbf{z}}$ .

We are in the process of implementing a version of this algorithm that will also include side chain placement.

### ACKNOWLEDGMENTS

The authors thank Michael Barnett for organizing a very stimulating conference. One of the authors (E. A. C.) would like to acknowledge the hospitality of the Dill group during several visits to UCSF. E. A. C. also acknowledges helpful suggestions by Manfred Minimair, who introduced him to the Dixon resultant and shared a Maple script for its computation.

---

### References

1. Leach, A. R. *Molecular Modeling: Principles and Applications*; 2nd Ed.; Pearson Education EMA: Harlow, UK, 2001.
2. <http://kinemage.biochem.duke.edu/databases/top500.php>.
3. Branden, C.; Tooze, J. *Introduction to Protein Structure*; Garland: New York, 1999.
4. Gō, N.; Scheraga, H. A. *Macromolecules* 1970, 3, 178.
5. Dodd, L. R.; Boone, T. D.; Theodorou, D. N. *Mol Phys* 1993, 78, 961.
6. Wedemeyer, W. J.; Scheraga, H. A. *J Comput Chem* 1999, 20, 819.
7. Coutsias, E. A.; Seok, C.; Jacobson, M. P.; Dill, K. A. *J Comput Chem* 2004, 25, 510.
8. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J Chem Phys* 1953, 21, 1087.
9. Li, Z.; Scheraga, H. A. *Proc Natl Acad Sci USA* 1987, 84, 6611.
10. Dinner, A. R. *J Comput Chem* 2000, 21, 1132.
11. Ulmschneider, J. P.; Jorgensen, W. L. *J Chem Phys* 2003, 118, 4261.
12. Wu, M. G.; Deem, M. W. *Mol Phys* 1999, 97, 559.
13. Manocha, D. In *Proceedings of Symposia in Applied Mathematics*; Cox, D. A.; Sturmfels, B., Eds.; American Mathematics Society: Boston, MA, 1997, p 41.
14. Wu, M. G.; Deem, M. W. *J Chem Phys* 1999, 111, 6625.
15. Canutescu, A. A.; Dunbrack, R. L. *Protein Sci* 2003, 12, 963.
16. Cortés, J.; Siméon, T.; Remaud-Siméon, M.; Tran, V. *J Comput Chem* 2004, 25, 956.
17. Lotan, I.; van den Bedem, H.; Deacon, A. M.; Latombe, J. C. *Computing Protein Structure from Electron Density Maps: The Missing Loop Problem*; preprint, 2005.
18. Noonan, K.; O'Brien, D.; Snoeyink, J. *Int J Robotics* 2005 (in press).
19. Duffy, J. *Analysis of Mechanisms and Robot Manipulators*; Arnold: London, 1980.
20. Lee, H. Y.; Liang, C. G. *Mech Mac Theory* 1988, 23, 219.
21. Bricard, R. *J Math Pure Appl* 1897, 3, 113.
22. van der Waerden, B. L. *Algebra*; Springer: Berlin, 2003.
23. Saxena, T. *Efficient Variable Elimination Using Resultants*; Ph.D. thesis; SUNY: Albany, 1997.
24. Kapur, D.; Saxena, T. In *Proceedings of the 1997 International Symposium on Symbolic and Algebraic Computation*; Küchlin, W. W., Ed.; Maui, Hawaii, July 21–23, 1997; p 141.
25. Lazopoulos, K. *Int J Solids Struct* 2005, 42, 3459.
26. Bernshtein, D. N. *Funct Anal Appl* 1975, 9, 183.
27. Gelfand, I. M.; Kapranov, M. M.; Zelevinsky, A. V. *Discriminants, Resultants and Multidimensional Determinants*; Birkhäuser: Boston, 1994.
28. Iyanaga, S.; Kawada, Y., editors, *Encyclopedic Dictionary of Mathematics*; MIT Press: Cambridge, MA, 1980; art 107D, 349.
29. <http://www.math.unm.edu/~vageli/codes/loop-closure.tar.gz>.
30. Kapur, D.; Saxena, T.; Yang, L. In *Proceedings of the ACM International Symposium on Symbolic and Algebraic Computation*; Oxford, England, July 1994.
31. Monagan, M. B.; Geddes, K. O.; Heal, K. M.; Labahn, G.; Vorkoetter, S. M.; McCarron, J.; DeMarco, P. *Maple 9 Introductory Programming Guide*; Maplesoft: Waterloo, Canada, 2003.
32. Monagan, M. B.; Geddes, K. O.; Heal, K. M.; Labahn, G.; Vorkoetter, S. M.; McCarron, J.; DeMarco, P. *Maple 9 Advanced Programming Guide*; Maplesoft: Waterloo, Canada, 2003.
33. Sommese, A. J.; Wampler, C. W. II *The Numerical Solution of Systems of Polynomial Arising in Engineering and Science*; World Scientific: Hackensack, NJ, 2005.