

COMMUNICATION

An alternative derivation of the equations of motion in torsion space for a branched linear chain

Christopher Bystroff

Department of Biology, RPI, Troy, NY 12180, USA. E-mail: bystrc@rpi.edu

Torsion space molecular dynamics may be more efficiently encoded if the global motions are separated from the internal motions. The equations of motion for single, non-cyclic chains are shown to be first order in the backbone angle parameters when the global frame of reference is ignored and second order otherwise. Adding a simple heuristic substitute for the global motions enables the encoding of dynamics for mixed constrained/unconstrained model systems.

Keywords: angle space/internal coordinates/molecular dynamics/peptides/proteins

Introduction

The natural coordinate system for simulations of conformational changes in single-molecule, branched-linear systems is the space of all dihedral angles or torsion space, where bond lengths and bond angles may be fixed to their equilibrium values. Nonetheless, the majority of molecular simulations today are still done in Cartesian space and geometrical constraints are handled using harmonic distance restraints and/or numerical coordinate resetting (Ryckaert *et al.*, 1977). This paper introduces a simpler, exact solution to the problem of encoding the torsion space equations of motion. The new approach is more computationally efficient than existing torsion space solutions.

Previous formulations of the equations of motion for internal coordinates were rigorously derived using LaGrange multipliers (Bae and Haug, 1987; Mazur and Abagyan, 1989). The resulting equations describe the gradient of atomic shifts relative to a fixed frame of reference, with respect to internal coordinates, including bond lengths, bond angles and dihedral angles. The solution is exact and completely general and may be applied to any system that can be described as a branched linear chain (or 'BKS tree'), even when such chains connect non-bonded atoms or have no geometric constraints. However, the power of the method exists in its ability to freeze out geometrical constraints, thus potentially speeding the computation. The method is not exact for covalent cycles, such as the cyclic peptides oxytocin and bacitracin, but there is a numerical solution to this problem (Abagyan and Mazur, 1989).

The shortcoming of the general formulation is its second-order dependence on the parameters, which requires non-linear least-squares or numerical integration for the solution of the torsion space gradient at each step of the simulation. Here it is noted that by separating the global motions from the internal motions, one can partition the second-order problem into two first-order problems. The first part is the solution of the torsion space gradient for pairwise internal distances, which is sensible because the energy calculations commonly used in molecular simulations are functions of pairwise distance, not absolute

position. The second part is the recovery of the global reference frame by least-squares superposition. The new approach amounts to solving two sets of linear equations at each step of the simulation.

Internal coordinate molecular dynamics has been implemented in several molecular simulation programs, especially as a tool in X-ray and NMR structure refinement of proteins (Guntert *et al.*, 1997; Rice and Brünger, 1994). It has been shown to be robust and fast relative to the Cartesian space equivalent, providing the twin advantages of a longer time step and freedom from geometric distortion at high temperatures.

Methods

The following is the derivation of the differential equations for the torsion angle energy gradient, as a function of the energy gradient in Cartesian space.

Newton's second law may be expressed as

$$d\vec{x}_i = \vec{v}_i dt + \frac{\vec{F}_i}{2m_i} dt^2 \quad (1)$$

where x , v , F and m are the coordinates, velocity, force and mass for atom i and t is time. The change in internal distances, D , is

$$dD_{ij} = (d\vec{x}_j - d\vec{x}_i) \cdot \hat{u}_{ij} \quad (2)$$

where u_{ij} is a unit vector from atom i to atom j . The same quantity may also be expressed as the sum of the contributions of infinitesimally small torsion angular shifts over the set of all intervening rotatable bonds K_{ij} :

$$dD_{ij} = \sum_{k \in K_{ij}} \frac{dD_{ij}}{d\theta_k} d\theta_k \quad (3)$$

where θ_k is the k th rotatable bond. In Equation (3), the molecular topology is taken into account, so that only rotatable bonds K_{ij} that occur along the covalent connection from atom i to atom j are used in the sum. The derivative in Equation (3) is calculated as (see Figure 1)

$$\frac{dD_{ij}}{d\theta_k} = \hat{u}_{ij} \cdot (\hat{u}_k \otimes \vec{r}_{kj}) \quad (4)$$

where \hat{u}_{ij} is a unit vector from x_i to x_j , \hat{u}_k is a unit vector along the axis of right-handed rotation (i.e. along the chemical bond between the atom closest to i along the chain to the atom closest to j along the chain, see Figure 2) and r_{kj} is a vector from any point on the axis of rotation to x_j . By substitution, we obtain

$$(d\vec{x}_j - d\vec{x}_i) \cdot \hat{u}_{ij} = \sum_{k \in K_{ij}} \hat{u}_{ij} \cdot (\hat{u}_k \otimes \vec{r}_{kj}) d\theta_k \quad (5)$$

All terms except $d\theta$ can be computed from the coordinates. For example, the Cartesian space energy gradient F may be computed using the non-covalent terms in an AMBER (Cheatham *et al.*, 1999) or CHARMM force field (Lazaridis and Karplus, 1999). There are as many such Equations (5) as

there are internal distances D_{ij} , which in principle number $N(N - 1)/2$ where N is the number of atoms. In practice, bonded atoms and atoms separated by only one atom along the chain do not have intervening rotatable bonds and thus have zero terms on the right-hand side of Equation (5). The same is true for atoms that are both part of the same rigid group, such as a phenyl ring.

Least squares

The torsional shifts $d\theta$ are solved using linear least squares. To do so, the set of Equations (5) is multiplied by the transpose of the matrix on its right-hand side:

$$\underline{m}_{k(ij)} = \hat{u}_{ij} \cdot (\hat{u}_k \otimes \vec{r}_{kj}) \quad (6)$$

Note that the indices of \underline{m} are k and ij , where the latter index represents an atom pair; thus the double sum in the equation below. Next we define the squared matrix \underline{M} , whose elements are

$$\underline{M}_{ki} = \sum_j \sum_l \underline{m}_{k(ij)}^T \underline{m}_{k(ij)} = \sum_i \sum_j [\hat{u}_{ij} \cdot (\hat{u}_k \otimes \vec{r}_{kj})] [I \hat{u}_{ij} \cdot (\hat{u}_i \otimes \vec{r}_{ij})] \quad (7)$$

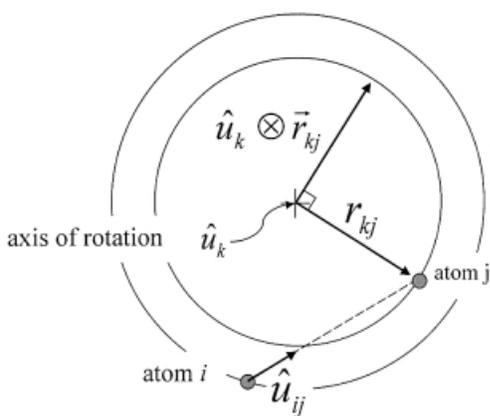


Fig. 1. The geometry of torsional rotation, projected along the rotation axis (center). The derivative of the atom pair distance is the height of the triangle ikj projected on a plane perpendicular to the rotation axis.

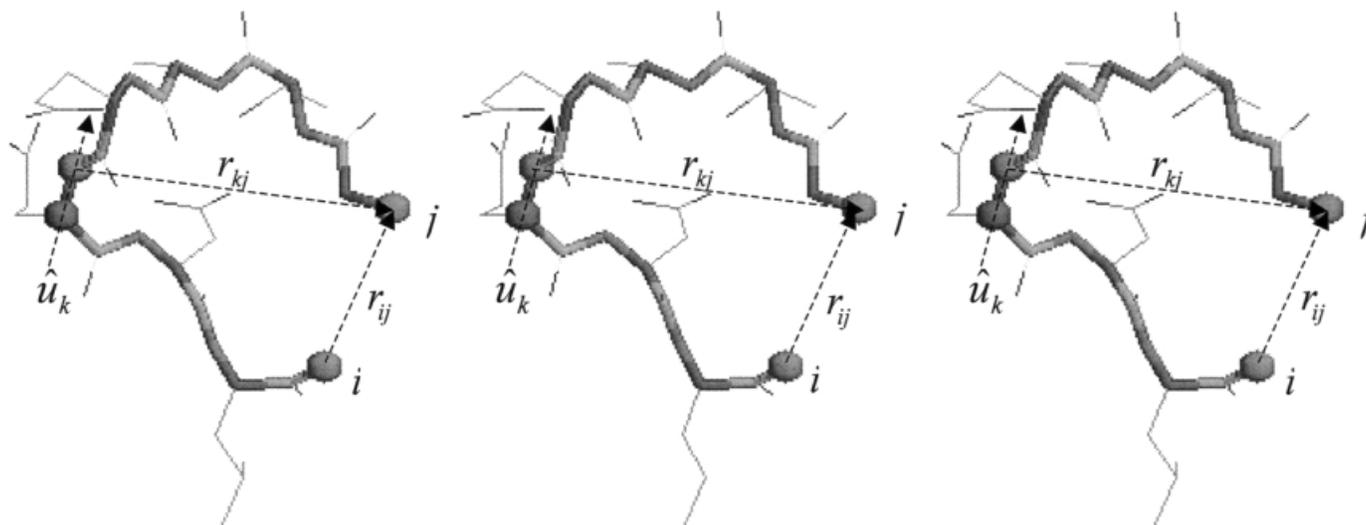


Fig. 2. The set K_{ij} of torsion angles to be included in the summation for the internal distance $i \rightarrow j$ in Equation (5) are the rotatable bonds along the intervening covalent path (thick lines).

and we define the vector A :

$$\vec{A}_k = \sum_i \sum_j \underline{m}_{k(ij)}^T (d\vec{x}_j - d\vec{x}_i) \cdot \hat{u}_{ij} = \sum_i \sum_j [\hat{u}_{ij} \cdot \hat{u}_k \otimes \vec{r}_{kj}] (d\vec{x}_j - d\vec{x}_i) \cdot \hat{u}_{ij} \quad (8)$$

Using Equations (7) and (8), the products of Equations (5) and the transpose of (6), summed over all atom pairs ij , becomes

$$\vec{A} = \underline{M} d\vec{\theta} \quad (9)$$

Inverting \underline{M} and multiplying, we obtain

$$\underline{M}^{-1} \vec{A} = d\vec{\theta} \quad (10)$$

A unique solution exists if the number of equations exceeds the number of variables θ . However, care must be taken to remove equations for torsion angles which have no dependent internal distances. The only step left out in this derivation is the matrix inversion in Equation (10), which is done using the Cholesky decomposition (Press *et al.*, 1992).

Note that the least-squares solution is an exact (within numerical limits) correction for the angular decoupling effect, also known as the Coriolis force, whereas previous, fast heuristic approaches do not account for coupling between all angular shifts simultaneously (Jain *et al.*, 1993). In the absence of an exact solution, the approximation errors ($\sin x \approx x$) accumulate along the chain, manifesting in unstable behavior at the ends of the chain and forcing a much smaller time step.

Generating Cartesian coordinates from torsion angles

The method has been applied to proteins. Using idealized templates for the amino acids, the torsion-shifted protein coordinates (\vec{x}^T) may be quickly regenerated from the torsion angles one residue at a time. The frame of reference for the first residue is defined by the first three atoms in the chain. The template coordinates are rotated and translated to the reference frame and torsional rotations are applied to the appropriate atoms. A set of three 'dummy' atoms, attached to the carbonyl carbon, define the frame of reference for the next residue and the cycle is repeated. At the end of the chain, the first dummy atom becomes the terminal oxygen.

Cartesian space velocities $[v_i]$, see Equation (1)] may be calculated in the usual way,

$$\vec{v}_i(t) = \vec{v}_i(t-1) + [\vec{F}_i(t-1)/m_i]\Delta t$$

or numerically,

$$\vec{v}_i(t) = [\vec{x}_i(t) - \vec{x}_i(t-1)]/\Delta t$$

from the coordinates after least-squares superposition (see below). The numerical method has the advantage of including the effects of the constraints.

Global frame of reference

The previous derivations of these equations considered a global frame of reference by linking the first atom in the chain to the origin by its polar coordinates and two angles of global orientation (Mazur *et al.*, 1989; Jain *et al.*, 1993). Since there are no forces between the atoms and the origin, the least-squares approach described above breaks down. In a single molecule system, there is no potential energy difference between globally rotated or translated molecules that are otherwise identical. However, a global frame of reference must be considered, even in a single molecule system, to account for velocity-dependent forces.

The local frame of reference for each molecule may be defined, arbitrarily, as the reference frame of the first residue in the first cycle of the simulation. For subsequent cycles, the coordinates, $\vec{x}^f(t)$, are least-squares superimposed (Kabsch, 1976) on the Cartesian-shifted coordinates, $\vec{x}^f(t-1) + d\vec{x}$. This gives the model the same global translational and angular momenta as an equal time step in Cartesian space, while enforcing torsional constraints, since least-squares superposition effectively cancels overall translational and angular momenta (Zhou *et al.*, 2000).

Using this two-step approach, a torsion space molecular dynamics simulation should conserve momentum to the degree that a Cartesian space simulation conserves momentum, since the torsion space shifts have been fit to the Cartesian space shifts. Conservation of energy was demonstrated for an earlier implementation of constrained torsion space dynamics (Mazur *et al.*, 1992), even for time steps that are an order of magnitude longer than the longest Cartesian space time step.

This treatment of the global motions in Cartesian space is preferable to the alternative of treating the molecule as a rigid body and using Euler's equations to apply the overall motions. A rigid body model of the intermolecular collisions is likely to be a poor approximation, since such flexible-chain collisions are likely to be more inelastic than elastic.

Results and discussion

By simply converting the problem to internal distances, solving the equations of motion in torsion space becomes a linear least-squares problem that is easily and rapidly computed at each step of a constrained molecular dynamics simulation.

Although bond stretching and bond bending were omitted for simplicity, a close analogy of Equation (3) can be drawn for these types of motion. Since both stretching and bending are linear sums over the set of intervening bonds in the chain, the derivation proceeds as for torsion angles and only the book-keeping of angle/distance dependences is changed. (Bending is identical with torsional rotation, but with a different axis of rotation.) However, the best justification for the extra computational expense of using internal coordinate equations of motion is that bond bending and stretching, the fastest

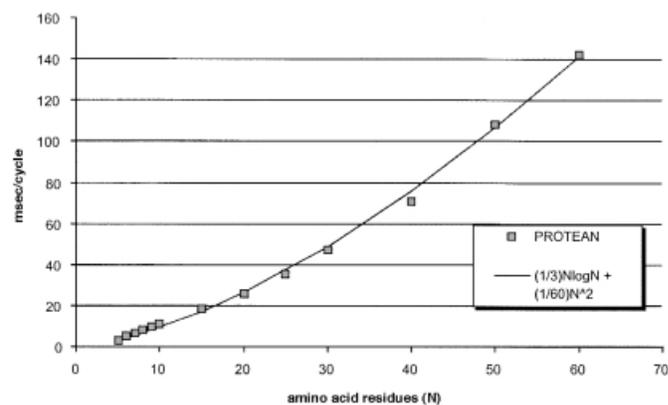


Fig. 3. Runtime complexity of program PROTEAN. CPU time per cycle (ms) was averaged over 1000 cycles using a typical protein sequence, using a 733 MHz Pentium III computer running Linux 7.0. Runtime is $O(N\log N)$ for short chains and $O(N^2)$ for longer chains.

oscillating and perhaps least interesting of the internal motions, are easily eliminated from the equations. Eliminating the fastest motions from the simulation allows a longer time step to be used and may speed up the simulation by as much as 10-fold (Mazur *et al.*, 1992).

Implementation

Torsion space dynamics using the new formulation are implemented in a new Fortran program (PROTEAN) that simulates polypeptide chains in implicit solvent. Testing on well-studied peptide systems and a comparison with AMBER 6.0 using will be published separately. The torsion angle to atom-pair dependences (the lists K_{ij} of intervening torsion angles) are 'hard-wired' for non-cyclic polypeptides, enforcing the requirement that the sum in Equation (5) is carried out over only the torsion angles that fall between the two atoms in the chain. Cyclic structures (i.e. the proline pyrrolidine ring) are treated as rigid bodies and disulfide bonds, which would create covalent cycles, are not constrained but instead restrained to their ideal bond lengths. Flexible bond lengths and angles are not allowed in the current implementation.

Runtime complexity

The runtime complexity for this algorithm is quadratic in the number of atoms [note the double sum in Equations (7) and (8)] and of order $O(N\log N)$ in the number of variable torsion angles (the runtime order of the Cholesky decomposition). These runtimes represent sequential program units and are therefore additive rather than multiplicative. Since the summation step is generally fast relative to the matrix inversion step, the effective runtime order is $N\log N$ for chains up to 40 amino acid residues in length and thereafter quadratic. Figure 3 shows actual runtimes in milliseconds/cycle for variable-length fragments of a compact 60-residue protein, performed using a 733 MHz Pentium machine running Linux 7.0.

References

- Abagyan, R.A. and Mazur, A.K. (1989) *J. Biomol. Struct. Dyn.*, **6**, 833–845.
- Bae, D.S. and Haug, E.J. (1987) *Mech. Struct. Mach.*, **15**, 359–382.
- Cheatham, T.E., Cieplak, P. and Kollman, P.A. (1999) *J. Biomol. Struct. Dyn.*, **16**, 845–862.
- Guntert, P., Mumenthaler, C. and Wuthrich, K. (1997) *J. Mol. Biol.*, **273**, 283–298.
- Jain, A., Vaidehi, N. and Rodriguez, G. (1993). *J. Comput. Phys.*, **106**, 258–268.
- Kabsch, W. (1976) *Acta Crystallogr.*, **A32**, 922–923.
- Lazaridis, T. and Karplus, M. (1999) *Proteins*, **35**, 133–152.

C.Bystroff

- Mazur,A.K. and Abagyan,R.A. (1989) *J. Biomol. Struct. Dyn.*, **6**, 815–832.
Mazur,A.K., Dorofeev,V.E. and Abagyan,R.A. (1992) *J. Comput. Phys.*, **92**, 261–272.
Press,W.H., Teukolsky,S.A., Vetterling,W.T. and Flannery,B.P. (1992) *Numerical Recipes in C*. Cambridge University Press, New York.
Rice,L.M. and Brunger,A.T. (1994) *Proteins*, **19**, 277–290.
Ryckaert,J.P., Cicotti,G. and Berendsen,H.J.C. (1977) *J. Comput. Phys.*, **23**, 327–341.
Zhou,Y., Cook,M. and Karplus,M. (2000) *Biophys. J.*, **79**, 2902–2908.

Received March 3, 2001; revised July 26, 2001; accepted August 1, 2001