

---

## RESEARCH ARTICLES

---

# Torsion Angle Dynamics: Reduced Variable Conformational Sampling Enhances Crystallographic Structure Refinement

Luke M. Rice and Axel T. Brünger

*The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520*

**ABSTRACT** A reduced variable conformational sampling strategy for macromolecules based on molecular dynamics in torsion angle space is evaluated using crystallographic refinement as a prototypical search problem. Bae and Haug's algorithm for constrained dynamics [Bae, D.S., Haug, E.J. A recursive formulation for constrained mechanical system dynamics. *Mech. Struct. Mach.* 15:359–382, 1987], originally developed for robotics, was used. Their formulation solves the equations of motion exactly for arbitrary holonomic constraints, and hence differs from commonly used approximation algorithms. It uses gradients calculated in Cartesian coordinates, and thus also differs from internal coordinate formulations. Molecular dynamics can be carried out at significantly higher temperatures due to the elimination of the high frequency bond and angle vibrations. The sampling strategy presented here combines high temperature torsion angle dynamics with repeated trajectories using different initial velocities. The best solutions can be identified by the free  $R$  value, or the  $R$  value if experimental phase information is appropriately included in the refinement. Applications to crystallographic refinement show a significantly increased radius of convergence over conventional techniques. For a test system with diffraction data to 2 Å resolution, slow-cooling protocols fail to converge if the backbone atom root mean square (rms) coordinate deviation from the crystal structure is greater than 1.25 Å, but torsion angle refinement can correct backbone atom rms coordinate deviations up to approximately 1.7 Å. © 1994 Wiley-Liss, Inc.

**Key words:** simulated annealing, molecular dynamics, torsion angle dynamics, conformational sampling, X-ray crystallography, refinement

## INTRODUCTION

Many applications of computer-aided modeling rely strongly on the ability to efficiently search the conformational space of macromolecules. Free energy perturbation calculations,<sup>1,2</sup> structure prediction,<sup>3,4</sup> and structure determination by X-ray crystallography and solution NMR spectroscopy<sup>5,6</sup> all depend critically on conformational sampling. While an exhaustive search of conformational space is usually impossible,<sup>7</sup> approximate searches for the global minimum of an objective function can be accomplished using simulated annealing.<sup>8,5</sup> Many different conformations are generated using molecular dynamics or Monte Carlo procedures. Both methods generally endow a system consisting of  $N$  atoms with the full  $3 \times N$  degrees of freedom obtained by considering each atom as independent of all others. A considerable simplification can be obtained by recognizing that the torsion angles, which represent approximately one tenth of the total degrees of freedom, actually encode most of the conformational variability. In fact, Diamond showed the effectiveness of this simplification of the search space as early as 1971,<sup>9</sup> but in the context of least squares refinement, it is a less powerful optimization technique. A sampling strategy using molecular dynamics restricted to the space of torsion angles should profit from this 10-fold reduction in the number of variables, and search conformations more effectively.

In the context of conformational searching, X-ray

---

Abbreviations: MIR, multiple isomorphous replacement; rms, root mean square.

---

Received February 14, 1994; accepted April 18, 1994.  
Address reprint requests to Axel T. Brünger, The Howard Hughes Medical Institute and Department of Biophysics and Biochemistry, Yale University, New Haven, CT 06520.

structure determination and refinement provide a good testing ground for sampling algorithms. Unlike search dependent applications such as de novo drug design that are error prone due to inaccuracies in empirical energy functions, structure determination has a clearly defined target, or objective function: maximal agreement with diffraction data. This unambiguous target makes it considerably easier to judge the merits of different sampling strategies. The target is usually a function of the difference between observed diffraction data and data calculated from a model set of atomic coordinates. The process of solving an X-ray crystal structure is aimed at deducing an atomic arrangement within the unit cell of the crystal that agrees with the observed diffraction intensities, i.e., at minimizing the objective function. Macromolecular diffraction data, however, are usually insufficient to precisely define atomic positions, so empirical knowledge about covalent bonding geometries and nonbonded interactions is included to better define the model.<sup>10</sup> The solution of a crystal structure can thus be viewed as a global optimization problem of the target function in a coordinate space where the atomic positions are restrained to ideal covalent geometry.

Molecular dynamics in the presence of holonomic constraints is not a new idea. One method, the SHAKE algorithm,<sup>11</sup> uses iterative corrections at each integration step to correct constraint violations. This is an approximation, however, and is impractical at high temperatures. Another alternative derives the equations of motion explicitly in internal (torsion angle) coordinates.<sup>12</sup> This eliminates the need for approximation algorithms, but introduces new complications: the force calculation in internal coordinates becomes significantly more complex. A more recent formulation (a "spatial operator" formulation<sup>13</sup>) of constrained molecular dynamics uses small ( $6 \times 6$ ) matrices to efficiently convert Cartesian forces into arbitrary internal coordinates. The formalism is very similar to that of Kalman filtering and smoothing.<sup>14</sup> The approach we chose<sup>15, 16</sup> is equivalent to the spatial operator one in that it makes use of small matrices to map Cartesian forces into internal coordinates. The resulting equations of motion are more physically transparent, however, because they are derived using D'Alembert's Principle instead of complicated matrix inversion techniques.

Here we employ high temperature torsion angle molecular dynamics as a sampling strategy, and use applications to crystallographic structure refinement as a test of its merits. Sampling is increased by repeating refinements with different initial conditions, and the best models are identified by the free  $R$  value,<sup>17</sup> or by the  $R$  value if experimental phase information is included in the refinement. The algorithms for crystallographic refinement and torsion angle molecular dynamics are briefly described in

the Theory section. The Methods section describes the test case and computational protocols. In Results and Discussion, a conformational sampling strategy is outlined, and its performance evaluated using results from several test refinements at both high (5–2 Å) and medium (5–3 Å) resolution. Finally, the Appendix contains a more thorough derivation of the equations of motion for dynamics with holonomic constraints.

## THEORY

### Crystallographic Refinement

X-Ray crystallography for macromolecules proceeds in a series of stages: crystallization, data collection, phasing, model building, and refinement. The third step, phasing, arises because of what is known as the *phase problem*: a monochromatic diffraction experiment on a single crystal provides the amplitudes of the reflections, but not the phases.<sup>18</sup> The phases are necessary for computing an electron density map by Fourier transformation of the structure factor, described by *both* amplitudes and phases. This map is usually used to construct a set of atomic coordinates, a model. Once a reliable model has been obtained, refinement begins.

Refinement can be formulated as the global minimization of the objective function

$$E = E_{\text{chem}} + w_{\text{X-ray}} E_{\text{X-ray}} \quad (1)$$

where  $E_{\text{chem}}$  comprises empirical information about chemical interactions,  $E_{\text{X-ray}}$  describes the difference between observed and calculated data, and  $w_{\text{X-ray}}$  is a weight chosen to balance the forces arising from each term.  $E_{\text{chem}}$  is a function of all atomic positions, describing covalent (bond lengths, bond angles, torsion angles, chiral centers, and planarity of aromatic rings) as well as nonbonded (van der Waals, hydrogen bonding, and electrostatic) interactions.

$$\begin{aligned} E_{\text{chem}} = & \sum_{\text{bonds}} k_b(r - r_0)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 \\ & + \sum_{\text{dihedrals}} k_\phi \cos(n\phi + d) \\ & + \sum_{\text{chiral, planar}} k_\omega(\omega - \omega_0)^2 \\ & + \sum_{\text{atom-pairs}} (ar^{-12} + br^{-6} + cr^{-1}) \quad (2) \end{aligned}$$

In practice, since crystallographic refinement is not very sensitive to the accuracy of the empirical energy function, a simpler "geometric" energy function is often substituted for  $E_{\text{chem}}$  in which all the nonbonded terms are replaced by a purely repulsive quartic one.<sup>19,10</sup> Conjugate gradient minimization of this function is roughly equivalent to least squares optimization provided the cosine dihedral term is replaced by squared differences. The various

forms of  $E_{\text{chem}}$  produce nearly identical structures after refinement.<sup>20</sup>

The most common form of  $E_{\text{X-ray}}$  is the crystallographic residual, defined as the sum over the squared differences between observed  $[F_{\text{obs}}(\mathbf{h})]$  and calculated  $[F_{\text{calc}}(\mathbf{h})]$  structure factor amplitudes

$$E_{\text{X-ray}} = \sum_{\mathbf{h}} [|F_{\text{obs}}(\mathbf{h})| - k|F_{\text{calc}}(\mathbf{h})|]^2. \quad (3)$$

Other forms of  $E_{\text{X-ray}}$  are adjusted to include phase information. For example, a penalty term ("phase restraints")<sup>21</sup> based on the difference between observed phases and those calculated from the model can be added to the residual

$$E_{\text{X-ray}} = \sum_{\mathbf{h}} [|F_{\text{obs}}(\mathbf{h})| - k|F_{\text{calc}}(\mathbf{h})|]^2 + w_p \sum_{\mathbf{h}} f[\phi_{\text{obs}}(\mathbf{h}) - \phi_{\text{calc}}(\mathbf{h})] \quad (4)$$

$w_p$  is the weight given to the phase restraint, and  $f$  is a square well function with a width equal to the arc cosine of the figure of merit  $[fom(\mathbf{h})]$  for each reflection. Another possible form of  $E_{\text{X-ray}}$  does not use the amplitude residual at all but instead simultaneously restrains the real and imaginary parts of the structure factor<sup>22</sup>; we refer to it as the vector residual. It has the form

$$E_{\text{X-ray}} = \sum_{\mathbf{h}} fom(\mathbf{h}) \{ [A_{\text{obs}}(\mathbf{h}) - kA_{\text{calc}}(\mathbf{h})]^2 + [B_{\text{obs}}(\mathbf{h}) - kB_{\text{calc}}(\mathbf{h})]^2 \} \quad (5)$$

where  $A, B$  denote the real and imaginary parts of the structure factor.

Electron density maps computed from native crystal amplitudes and experimentally observed phases are in most cases insufficient to allow a complete,

unambiguous tracing of the macromolecule. Furthermore, electron density maps for macromolecules are usually obtained at lower than atomic resolution and are therefore prone to errors during interpretation. Initial atomic models are likely to contain (partially) incorrect regions, and thus require refinement techniques with a large radius of convergence.

Of the commonly used algorithms for refining macromolecular crystal structures,<sup>5,10,23,24</sup> simulated annealing has the largest radius of convergence. Annealing denotes a physical process wherein a solid is heated until all particles randomly arrange themselves in the liquid phase, and then slowly cooled so that all particles arrange themselves in the lowest energy state. By formally identifying the target  $E$  [Eq. (1)] with the potential energy of the system, "simulated" annealing can be carried out.<sup>8</sup> Simulated annealing is an approximation algorithm: there is no guarantee that it will find the global minimum (except asymptotically).<sup>25</sup>

Simulated annealing requires the definition of a target function, a generation mechanism to create a Boltzmann distribution of conformations at a given temperature  $T$ , and an annealing schedule, that is, a sequence of temperatures  $T_1 > T_2 > \dots > T_l$  at which the Boltzmann distribution is computed. One common implementation uses molecular dynamics to generate changes in atomic coordinates consistent with the Boltzmann distribution at a given temperature. Another implementation generates a Boltzmann distribution using a Metropolis Monte Carlo procedure to sample conformations. The molecular dynamics approach has the advantage of being driven by the gradient, and thus restricts changes of the system to physically reasonable ones. The iner-

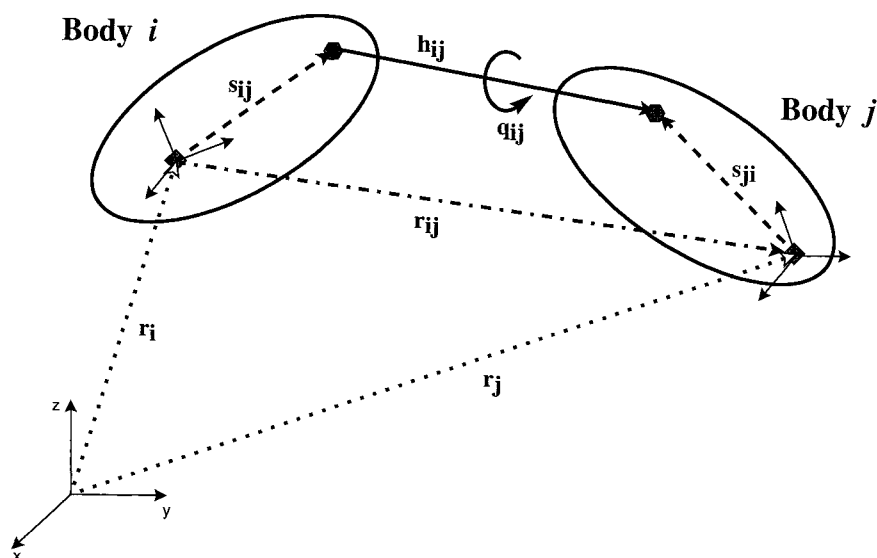


Fig. 1. Schematic diagram defining vectors for the simplest example of torsion angle dynamics.

tia of the system also seems to facilitate barrier crossing.

Molecular dynamics consists of the numerical integration of Newton's equations of motion:

$$m_i \frac{\partial^2 \mathbf{r}_{i,u}}{\partial t^2} = - \frac{\partial E}{\partial \mathbf{r}_{i,u}}. \quad (6)$$

Here  $\mathbf{r}_{i,u}$  and  $m_i$  are, respectively, the coordinates and mass of atom  $i$ , and  $E$  is the potential energy. In the context of simulated annealing,  $E$  denotes the target function comprised of  $E_{\text{chem}}$  and  $E_{\text{X-ray}}$ . Simulated annealing also requires temperature control. Temperature coupling<sup>26</sup> accomplishes this goal by adding pseudo-friction forces which are proportional to the velocities of each atom  $\mathbf{v}_i$ :

$$\mathbf{f}_i^{\text{frict}} = \beta_i \left( \frac{T_0}{T} - 1 \right) \mathbf{v}_i. \quad (7)$$

Here  $T_0$  is the bath temperature,  $\beta_i$  is a force constant, and  $T$  is the actual temperature computed from the kinetic energy of the system. This pseudofriction adds or removes "heat" from the system by forcing or damping atomic velocities.

### Torsion Angle Dynamics

Unconstrained molecular dynamics for a system of  $N$  particles solves the well known Newton equations of motion  $\delta \rightarrow \partial \mathbf{x}_i / \partial t^2 = \mathbf{a}_i(\mathbf{r}) = \mathbf{F}_i(\mathbf{r})/m_i$ . For systems with holonomic constraints, there are essentially two possible approaches. The first is to switch from Cartesian coordinates  $\mathbf{r}_i$  to generalized internal ones  $\mathbf{q}_i$ . Having thus redefined the system, one solves equations of motion for the generalized coordinates analogous to those above. Formulating the dynamics in terms of generalized coordinates has the disadvantage that it is difficult (but not impossible) to calculate the gradients, but has the advantage that they are functions of the coordinates only, and do not involve the velocities. Thus once one has expressed the gradients, conventional Verlet type integration schemes<sup>27</sup> can be used. We preferred a formulation in Cartesian coordinates, where the gradient calculations remain relatively straightforward and topology independent. The acceleration, however, becomes a function of positions and velocities:

$$\mathbf{a}(\mathbf{r}, \dot{\mathbf{r}}) = \mathbf{M}^{-1}(\mathbf{r}) \mathbf{Q}(\mathbf{r}, \dot{\mathbf{r}}) \quad (8)$$

where  $\mathbf{a}$  represents the system acceleration vector, and  $\mathbf{M}$  and  $\mathbf{Q}$  denote the  $(6 \times 6)$  system inertia matrix and  $(6 \times 1)$  generalized force vector. This does not present insurmountable difficulties, but instead requires different integration schemes. Our implementation uses a fourth-order Runge Kutta integration scheme.<sup>28</sup>

The equations of motion for constrained dynamics in Cartesian coordinates are derived in complete

generality by Bae and Haug.<sup>15,16</sup> Here we derive the equations in a simpler form specific to torsion angle dynamics. The following derivation is more suggestive than complete; a more complete form appears in the Appendix.

Consider two bodies,  $i$  and  $j$ , connected by a bond of fixed length  $|\mathbf{h}_{ij}|$ . We make the assumption that the only allowable relative motion between the two bodies is a rotation about  $\mathbf{h}_{ij}$ . Several vectors are defined in Figure 1:  $\mathbf{r}_i$  and  $\mathbf{r}_j$  locate (with respect to an inertial frame) the center of mass of body  $i$  and  $j$ , respectively.  $\mathbf{s}_{ij}$  and  $\mathbf{s}_{ji}$  define the points of attachment for each body with respect to its center of mass. The position of the center of mass of body  $j$  with respect to that of body  $i$  is simply  $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ . Finally, the scalar  $q_{ij}$  measures the relative angle of rotation about the bond  $\mathbf{h}_{ij}$ .

The assumption that the only allowable relative motion between the two bodies is a rotation about the bond connecting them implies a relationship between the angular velocity  $\omega$  of their respective centers of mass measured in an inertial ("lab") frame:

$$\omega_j = \omega_i + \dot{\mathbf{h}}_{ij} \dot{q}_{ij} \quad (9)$$

where  $q_{ij}$  denotes the time derivative of the relative angle between the two bodies and  $\mathbf{h}_{ij} = \mathbf{h}_{ij}/|\mathbf{h}_{ij}|$  is the unit vector along the bond connecting them. The expression for  $\mathbf{r}_j$  can be rewritten:

$$\begin{aligned} \mathbf{r}_j &= \mathbf{r}_i + \mathbf{r}_{ij} \\ &= \mathbf{r}_i + \mathbf{s}_{ij} + |\mathbf{h}_{ij}| \mathbf{h}_{ij} - \mathbf{s}_{ji}. \end{aligned} \quad (10)$$

This expression can be differentiated and then rearranged, resulting in an expression for the center of mass velocity of body  $j$  in terms of that of body  $i$ :

$$\begin{aligned} \dot{\mathbf{r}}_j &= \dot{\mathbf{r}}_i + \dot{\mathbf{s}}_{ij} + |\dot{\mathbf{h}}_{ij}| \mathbf{h}_{ij} - \dot{\mathbf{s}}_{ji} \\ &= \dot{\mathbf{r}}_i + \mathbf{w}_i \times \mathbf{s}_{ij} + |\mathbf{h}_{ij}| \mathbf{w}_i \times \mathbf{h}_{ij} - \mathbf{w}_j \times \mathbf{s}_{ji} \\ &= \dot{\mathbf{r}}_i - \mathbf{s}_{ij} \times \mathbf{w}_i - |\mathbf{h}_{ij}| \dot{\mathbf{h}}_{ij} \times \mathbf{w}_i + \mathbf{s}_{ji} \times \mathbf{w}_i \\ &\quad - \dot{q}_{ij} \mathbf{h}_{ij} \times \mathbf{s}_{ji} \\ &= \dot{\mathbf{r}}_i - \mathbf{r}_{ij} \times \mathbf{w}_i - (\mathbf{h}_{ij} \times \mathbf{s}_{ji}) \dot{q}_{ij}. \end{aligned} \quad (11)$$

The following notation will simplify the form of the equations:

$$\mathbf{Y}_i = \begin{bmatrix} \dot{\mathbf{r}}_i \\ \mathbf{w}_i \end{bmatrix} \quad (12)$$

$$\mathbf{q}_{ij} = [q_{ij}] \quad (13)$$

$$\delta \mathbf{Z}_i = \begin{bmatrix} \delta \mathbf{r}_i \\ \delta \boldsymbol{\pi}_i \end{bmatrix} \quad (14)$$

where  $\mathbf{Y}$  denotes the  $6 \times 1$  vector of translational and angular velocities,  $\mathbf{q}$  denotes the  $1 \times 1$  vector of relative angular velocity, and  $\delta \mathbf{Z}$  denotes the  $6 \times 1$  vector of translational ( $\delta \mathbf{r}_i$ ) and angular ( $\delta \boldsymbol{\pi}_i$ ) virtual displacements. ( $\mathbf{q}$  is defined as a vector for consistency with the original derivation and for convenience in later matrix operations.) Using this vector

notation, the velocity relationships can be written in a more compact form:

$$\mathbf{Y}_j = \begin{bmatrix} \mathbf{I} - \hat{\mathbf{r}}_{ij} \\ \mathbf{O} & \mathbf{I} \end{bmatrix} \mathbf{Y}_i + \begin{bmatrix} -\hat{\mathbf{h}}_{ij} \times \mathbf{s}_{ji} \\ \mathbf{h}_{ij} \end{bmatrix} \dot{\mathbf{q}}_{ij} \quad (15)$$

$$= \mathbf{B}_{ij}^{(1)} \mathbf{Y}_i + \mathbf{B}_{ij}^{(2)} \dot{\mathbf{q}}_{ij}$$

where  $\hat{\mathbf{r}}_{ij}$  denotes the skew-symmetric matrix associated with the vector cross product of  $\mathbf{r}_{ij}$ . One can also use this notation to relate a virtual displacement of body  $j$  ( $\delta \mathbf{Z}_j$ ) to a virtual displacement of body  $i$  ( $\delta \mathbf{Z}_i$ ):

$$\delta \mathbf{Z}_j = \mathbf{B}_{ij}^{(1)} \delta \mathbf{Z}_i + \mathbf{B}_{ij}^{(2)} \delta \mathbf{q}_{ij}. \quad (16)$$

Finally, the velocity relationship can be differentiated, yielding an acceleration one:

$$\dot{\mathbf{Y}}_j = \mathbf{B}_{ij}^{(1)} \dot{\mathbf{Y}}_i + \mathbf{B}_{ij}^{(2)} \ddot{\mathbf{q}}_{ij} + \mathbf{D}_{ij} \quad (17)$$

$$\ddot{\mathbf{q}}_{ij} = \mathbf{B}_{ij}^{(1)} \dot{\mathbf{Y}}_i + \dot{\mathbf{B}}_{ij}^{(2)} \dot{\mathbf{q}}_{ij}.$$

All that remains is to solve for the dynamics of the system. This is done using D'Alembert's principle,<sup>29</sup> which states that the virtual work done by the applied forces must vanish (see the Appendix for a more complete discussion):

$$\sum_{\text{bodies } k} \delta \mathbf{Z}_k^T (\mathbf{M}_k \dot{\mathbf{Y}}_k - \mathbf{Q}_k) = 0 \quad (18)$$

where "T" denotes the transpose, and the equation must hold for all  $\delta \mathbf{Z}_k$  which do not violate the constraints and where  $\mathbf{M}_k$  and  $\mathbf{Q}_k$  are the inertia matrix and generalized force vector, respectively. One can then solve for the acceleration of one body and the relative acceleration between them to obtain  $\mathbf{a}(\mathbf{r}, \mathbf{r})$ . Finally, a Runge Kutta method is used to integrate the partial differential equation  $\delta \rightarrow \partial x \rightarrow \mathbf{r} / \delta t^2 = \mathbf{a}(\mathbf{r}, \mathbf{r})$  [Eq. (8)].

This is a recursive algorithm, so the equations of motion for two bodies easily extend to many. Our implementation groups atoms into rigid bodies, allowing only torsion-angle motions between bodies. The connectivity of these bodies defines a tree-like topology for the macromolecule, with one arbitrarily chosen body identified as the base (or root). As any molecular dynamics algorithm, the torsion angle one begins with positions, velocities, and forces for all atoms. Then center of mass positions, velocities, and forces are computed for each rigid group. Starting at the outer ends of the tree topology, each chain is "reduced" one body at a time by solving for the relative acceleration between the tip and the direct inner body. Then the tip's inertial properties are mapped, or aggregated, into the inner body's, resulting in a chain one link shorter. This process continues until an expression for the acceleration of the base body is obtained. After solving for the base's acceleration (which only requires inversion of a  $6 \times 6$  matrix), the aggregation of bodies is reversed. The acceleration of the body "outboard" of the base is

determined by the base's acceleration and the relative acceleration between the bodies, and so on for the next outboard body. This outward expansion is continued until the tree has been completely covered (see the Appendix for more details). Then a Runge Kutta integration step updates positions and velocities. Finally, new forces are calculated, and the whole process begins anew. The formalism is a general one: several tree-like topologies can be handled, as can closed topological loops such as those formed by disulfide bonds. Currently, however, our implementation treats these closed loops approximately by harmonic distance restraints.

## METHODS

Test refinements were carried out on the  $\alpha$ -amylase inhibitor 1HOE-467A, a small protein of 74 amino acids. All calculations were done using an extended version of the program X-PLOR,<sup>30</sup> which will be made available in a future version. Force field parameters were taken from the parameter set PARHCSOX developed by Engh and Huber.<sup>31</sup> The diffraction data was 70% complete for reflections with  $|F_{\text{obs}}| > 2\sigma$ . The mean figure of merit of the MIR phases was 0.69 for data from 5 to 3 Å resolution, with a maximum of around 0.82 (near the low resolution end) and a minimum of around 0.57 (near the high resolution end). The resolution of the data used in test refinements was from 5 to 2 Å ("high resolution refinements") and 5 to 3 Å ("medium resolution refinements"). Refinements began with successively worse models, the generation of which is described in the next section. Each model was subjected to 100 steps of conjugate gradient minimization against the target  $E_{\text{chem}}$  in order to obtain close to ideal geometry for bond lengths and bond angles which would remain fixed during torsion angle molecular dynamics.  $C^\alpha$  coordinates were harmonically restrained during this phase so that the local geometry could be corrected without affecting the errors introduced intentionally. The refinement stage consisted of 4 ps of constant temperature (5000 K or 10,000 K) molecular dynamics—conventional or torsion angle—against the target  $E^* = E_{\text{chem}}^* + w_{\text{X-ray}} E_{\text{xray}}$ , where  $E_{\text{chem}}^*$  denotes the conventional empirical energy with the usual nonbonded terms substituted by a purely repulsive one.<sup>19,10</sup>  $E_{\text{x-ray}}$  is simply the residual [Eq.(3)] for high resolution refinements or the vector residual [Eq. (5)] for medium resolution refinements. For torsion angle refinements a 0.002 ps timestep was used for integrating the equations of motion, and for conventional dynamics a 0.0005 ps timestep. The resulting model then underwent a short (0.1 ps) quenching by conventional molecular dynamics at 300 K, against the target,  $E = E_{\text{chem}} w_{\text{X-ray}}, E_{\text{x-ray}}, +$  where  $E_{\text{x-ray}}$  is the residual for both high and medium resolution refinements; 120 steps of minimization were applied as the last step in the refinement.

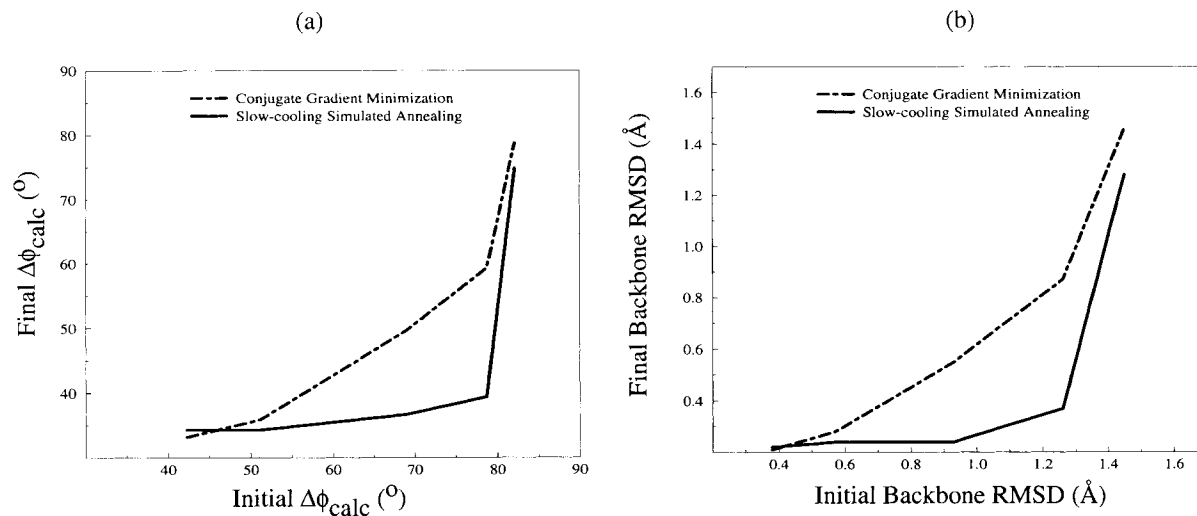


Fig. 2. Two measures of the radius of convergence of two common refinement techniques. Dashed lines describe 100 steps of conjugate gradient minimization, solid lines slow-cooling simulated annealing. The initial temperature for the simulated annealing protocol was 3000 K, and was decreased by 25 K every 0.025 ps until 300 K, at which point 120 steps of conjugate gradient minimization were applied. (a) The final calculated mean phase

error  $\Delta\phi_{\text{calc}}$  as a function of the initial calculated phase error, where  $\Delta\phi_{\text{calc}}$  is defined as the average (figure of merit weighted) difference between phases calculated from the crystal structure and phases calculated from the model. (b) The final backbone atom rms coordinate deviation from the crystal structure versus initial backbone atom rms coordinate deviations.

## RESULTS AND DISCUSSION

Conventional refinement methods such as slow-cooling simulated annealing are powerful, but nevertheless show some limitations.<sup>5,21,32–39</sup> These limitations may be quantified by starting with atomic coordinates close to those of the crystal structure and arbitrarily introducing errors. A series of these increasingly bad models could then be refined—the point at which a refinement technique fails to return the answer will give a good indication of that methods' radius of convergence.

In the present study, arbitrarily "scrambled" models were generated from an initial model built by Pflugrath et al.<sup>35</sup> using experimental phase information from multiple isomorphous replacement (MIR) diffraction data. This initial model is quite close to the refined crystal structure, with backbone and nonhydrogen atom root mean square (rms) coordinate deviations of 0.4 and 0.8 Å, respectively. Scrambling of this initial model was introduced by increasingly long molecular dynamics simulations at 600 K computed *without* reference to the X-ray data. Errors are thereby distributed throughout the structure, and are typical for molecular replacement models or poorly built initial models. A series of these models was refined using two standard techniques: conjugate gradient minimization and slow-cooling simulated annealing.

Results are presented in Figure 2, which depicts two different measures of model accuracy, calculated mean phase error  $\Delta\phi_{\text{calc}}$  (between 5 and 2 Å resolution) and backbone atom rms coordinate deviations.  $\Delta\phi_{\text{calc}}$  is defined as the average figure of

merit weighted phase difference between phases computed from the model and those computed from the refined crystal structure. A similar graph for a perfect refinement technique would simply be a straight line along the x-axis: no matter how great the initial errors, the result would be in good agreement with the answer. Clearly this is not the case for conjugate gradient minimization, or even for slow-cooling simulated annealing. For refinements carried out between 5 and 2 Å, slow-cooling simulated annealing can correct mean phase errors up to about 80°. This corresponds to backbone atom rms coordinate deviations of around 1.3 Å. Figure 2 also illustrates the highly nonlinear relationship between phase error and rms coordinate difference: the average phase error can never exceed 90° while the coordinate rms difference is considerably less restricted.

In order to evaluate the torsion angle refinement strategy, three models were taken from the scrambled series. The first model, the "least scrambled model," sits just inside the radius of convergence for a standard slow-cooling protocol for this system. The next two, the "medium scrambled model" and "most scrambled model" lie increasingly outside the radius of convergence of slow-cooling simulated annealing. These models are characterized by calculated mean phase differences  $\Delta\phi_{\text{calc}}$  (at 5 to 2 Å resolution) of 79, 82, and 84°, respectively. They have respective backbone atom rms coordinate deviations from the crystal structure of 1.25, 1.45, and 1.63 Å.

The refinement strategy incorporates three notable features. The most significant one is the use of

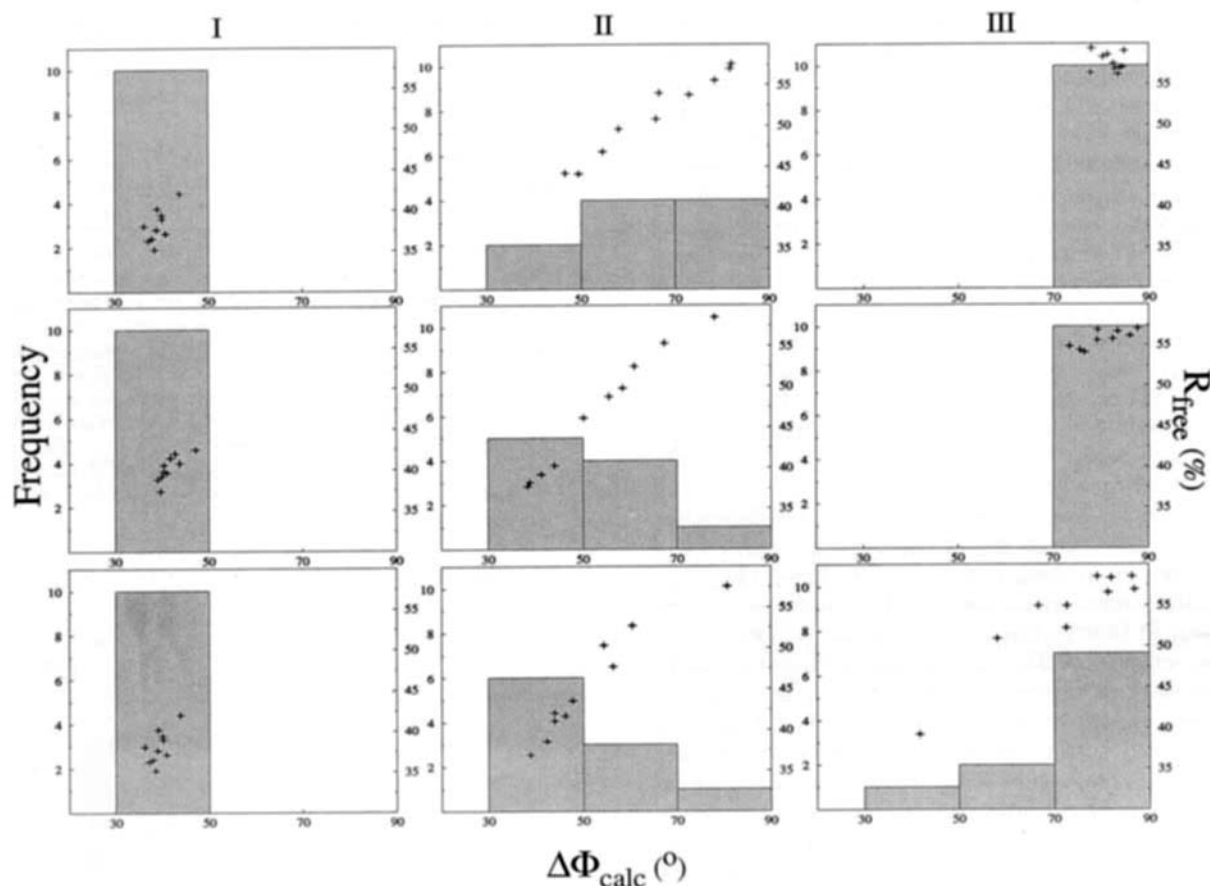


Fig. 3. Results from three series of refinements at 5 to 2 Å resolution (repeated 10 times with different initial velocities) of three increasingly incorrect initial models. Each bin counts the number of times a refinement yields a model with a calculated phase error between 30 and 50, 50 and 70, and 70 and 90°. The crosses above each bin show the value of  $R_{\text{free}}$  for each refinement. **Column I** describes refinements of the "least scrambled

model" (see text), **column II** refinements of the "medium scrambled model," and **column III** refinements of the "most scrambled model." The top row presents results from conventional molecular dynamics constant temperature (5000 K) refinements, the middle row torsion angle dynamics refinements at the same temperature, and the bottom row torsion angle dynamics refinements at 10,000 K.

torsion angle molecular dynamics in order to simplify the conformational space to be searched. Second, the slow-cooling protocol is replaced by constant high temperature molecular dynamics (to increase sampling), and is followed by a relatively fast cooling stage. Finally, as in solution NMR structure determination,<sup>6</sup> refinements are repeated 10 times from the same initial model but with different initial velocities, again to increase sampling.

Results for refinements at 5 to 2 Å resolution are presented in Figure 3. From left to right, the three columns depict test refinements for progressively worse models, from the least scrambled to the most scrambled. From top to bottom, the refinement protocol goes from constant temperature (5000 K) conventional dynamics to constant temperature (5000 K) torsion angle dynamics to constant higher temperature (10,000 K) torsion angle dynamics. Refinements are binned in three categories, according to the degree of convergence as measured by the phase

difference  $\Delta\phi_{\text{calc}}$ . This measure is only available if the answer is known to advance, so another measure, the free  $R$  value ( $R_{\text{free}}$ ),<sup>17</sup> is shown for each model in each bin.  $R_{\text{free}}$  is the measure that one would use in practice to judge the quality of a refined model. The figure shows that this use of  $R_{\text{free}}$  is justified since there is a strong correlation between  $R_{\text{free}}$  and the phase error.

Several trends are apparent from Figure 3. For the least scrambled model (column I), every refinement converges on the correct answer. As model quality degrades, however, it becomes clear that the convergence of standard refinement techniques is significantly improved upon by the modifications described above. Column II (representing refinements of the medium scrambled model) of Figure 3 is the most suggestive. The top graph demonstrates the value of repeated refinements, even for conventional dynamics, with an almost equal distribution between the converged, partially converged, and un-

converged bins. The middle graph indicates that at the same temperature, torsion angle molecular dynamics refinement succeeds more frequently, implying more efficient conformational searching. The lower graph shows that raising the temperature of the heat bath further improves the success rate of these refinements.

To place this series in the context of slow-cooling simulated annealing, 10 conventional slow-cooling refinements starting from the medium scrambled model do not converge (not shown). Thus, the slow-cooling sampling protocols do not search conformational space as effectively. The temperatures in the constant temperature refinements are high enough to provide good searching, but low enough to ensure that the model remains in the vicinity of any significant minimum it might encounter.

Column III of Figure 3, representing refinements of the most scrambled structure, demonstrates the power of high temperature torsion-angle sampling: neither conventional or torsion angle dynamics converge at 5000 K, but some torsion angle refinements converge at 10,000 K. The initial model for these refinements lies well outside the radius of convergence of conventional techniques (cf. Fig. 2). The extremely high temperature of 10,000 K is unattainable for unconstrained molecular dynamics without incurring significant and prohibitive computational cost due to numerical instabilities. In simplifying the search space by constraining bond lengths and bond angles to equilibrium values the high frequency vibrations which effectively place limits on the maximum attainable temperature are eliminated. Only through a combination of repeated refinements and high temperature torsion-angle molecular dynamics can one achieve convergence for models of this poor quality.

All of the models from converged refinements have calculated phase differences  $\Delta\phi_{\text{calc}}$  greater than  $30^\circ$ . This value might seem rather high, but in fact one cannot expect much better for this particular system. A conventional slow-cooling refinement beginning with the crystal structure coordinates only converges to a phase difference of  $26^\circ$ , with backbone and nonhydrogen rms coordinate deviations of 0.17 and 0.40 Å, respectively. A conjugate gradient minimization of 200 steps applied to the crystal structure coordinates results in a phase difference of  $15.5^\circ$ , corresponding to backbone and nonhydrogen rms coordinate deviations of 0.14 and 0.18 Å, respectively. These phase errors between  $15$  and  $30^\circ$  indicate the limited extent to which the structure factor amplitudes determine the structure at 2 Å resolution.

Besides increasing convergence, constraining molecular dynamics refinement to torsion angles has another practical benefit: it requires less computer time. For identical refinements consisting of a constant temperature 5000 K stage of 4 ps, torsion an-

TABLE I. Representative Timings for One Refinement

Refinement type*	Average time (s)
Conventional slowcool	4627
5,000 K conventional dynamics	5186
5,000 K torsion angle dynamics	3546
10,000 K torsion angle dynamics	4058

\*Conventional slowcool describes a simulated annealing refinement which begins after 100 steps of conjugate gradient minimization against  $E_{\text{chem}}$ . The initial temperature is 3000 K, and is decreased by 25 K every 0.025 ps until 300 K, resulting in a dynamics trajectory of 2.7 ps. 120 steps of conjugate gradient minimization are applied to end the refinement. The other refinement protocols begin with the same 100 steps of conjugate gradient minimization, but then undergo 4 ps of constant temperature molecular dynamics against  $E_{\text{chem}} + E_{\text{X-ray}}$ , followed by 0.1 ps of conventional dynamics at 300 K against the same target. These refinements also end with 120 steps of conjugate gradient minimization.

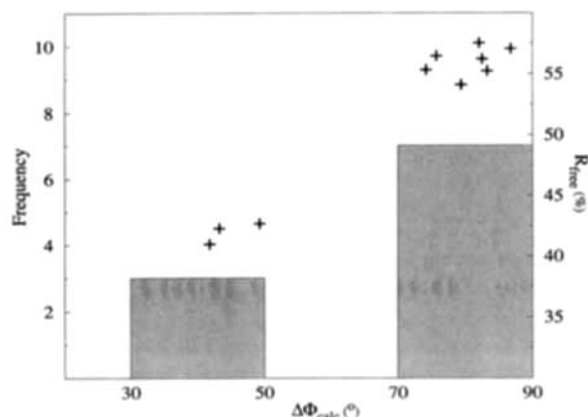


Fig. 4. Similar to Figure 3, but representing the result of doubling the length of the high temperature stage for a 10,000 K torsion angle refinement of the most scrambled structure. It should be compared with the lower right corner of Figure 3.

gle refinement is about 20% faster than conventional refinement. Increasing the temperature to 10,000 K incurs only a relatively modest increase in computational cost (representative timings for a Hewlett Packard 735 are shown in Table I).

An immediate advantage of this time saving is the possibility to run longer simulations, increasing the ability to search conformational space. Figure 4 shows the result of doubling the length of the constant high temperature stage from 4 to 8 ps for a 10,000 K torsion angle refinement of the most scrambled model. In this case longer sampling increases the success rate of the refinements. This series of refinements is not surprisingly more computationally intensive than one standard slow-cooling refinement, but the computer time required remains small in comparison to the human time needed to manually rebuild the model or repeat an experiment to obtain better initial phases.



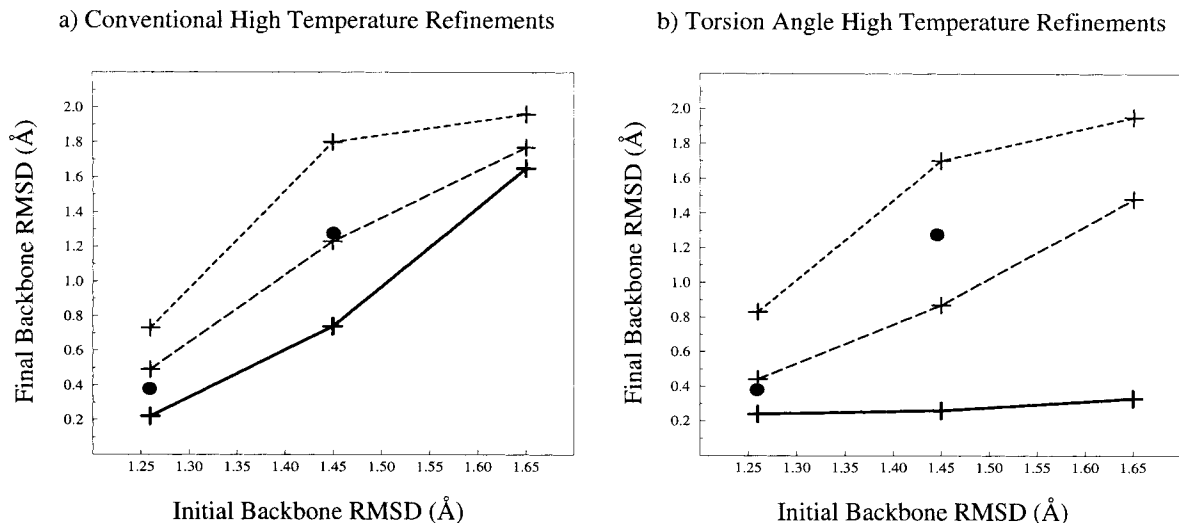


Fig. 5. Convergence of (a) 5000 K conventional refinements and (b) 10,000 K torsion angle refinements of the least, medium, and most scrambled models. Solid lines show the lowest backbone atom rms coordinate deviation obtained from each series of refinements, and long-dashed lines show the average. The best models can be identified by a low value of  $R_{\text{free}}$  (cf. Fig. 6). Short-

dashed lines show the highest backbone atom rms coordinate deviation from each series. Ten refinements were averaged for each model, except for the torsion angle refinements of the most scrambled model, where 20 (10 times 4 ps and 10 times 8 ps) were averaged.

The convergence of the different high resolution refinement protocols is summarized in Figure 5, where convergence is measured by final backbone atom rms coordinate deviations from the crystal structure. Solid lines show the lowest values obtained from the series of refinements shown in Figures 3 and 4. Long-dashed lines show the average backbone atoms rms coordinate deviation from the crystal structure from all refinements of a given model, and short-dashed lines show the highest. Filled circles indicate the relevant points taken from the slow-cooling refinements shown in Figure 2. Torsion angle refinements (Fig. 5b) outperform conventional ones (Fig. 5a) on average, dramatically so if only one considers the best model from each series. Also, the worst torsion refinement is no worse than the worst conventional one, so there is nothing to lose by using torsion angle refinement. Clearly, the backbone atom rms coordinate deviation is only available if one knows the answer in advance. Figure 6, however, shows a strong correlation between  $R_{\text{free}}$  and backbone rms coordinate deviations. Thus in practice  $R_{\text{free}}$  can be used to consistently identify the best models from a series of refinements. Averaging the structure factors over the four best refinements of the most scrambled model yields a phase difference  $\Delta\phi_{\text{calc}}$  of  $41.8^\circ$ . This is a modest improvement over the average individual  $\Delta\phi_{\text{calc}}$  of  $44^\circ$ , and may reduce model bias. (The four best refinements have mean phase errors of  $41.7^\circ$ ,  $41.8^\circ$ ,  $43.2^\circ$ , and  $49.2^\circ$ .)

To give a more qualitative sense of the errors that torsion angle dynamics refinement can correct, Fig-

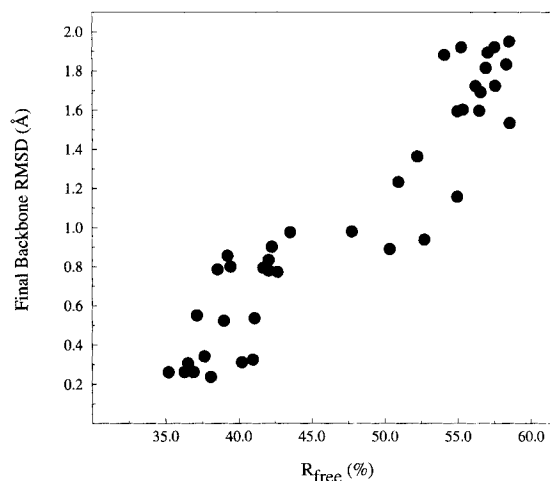


Fig. 6. Final backbone atom rms coordinate deviations plotted against  $R_{\text{free}}$  for all 10,000 K torsion angle refinements.

ure 7 shows  $(2|F|_o - |F|_c)e^{i\phi_c}$  electron density maps (a) before and (b) after refinement of the most scrambled model. Both maps are contoured at  $1.4\sigma$ . Also shown on both sides are the model coordinates (in red) and the crystal structure coordinates (in blue). The electron density calculated from the initial model is uninterpretable, having very little density for the correct chain trace. The right-hand figure shows the same region for the model with the lowest value of  $R_{\text{free}}$ . The model's chain trace lies very close to the correct one, and the density is far more easily interpreted. Backbone and side chain atom rms coordinate differences from the crystal structure be-

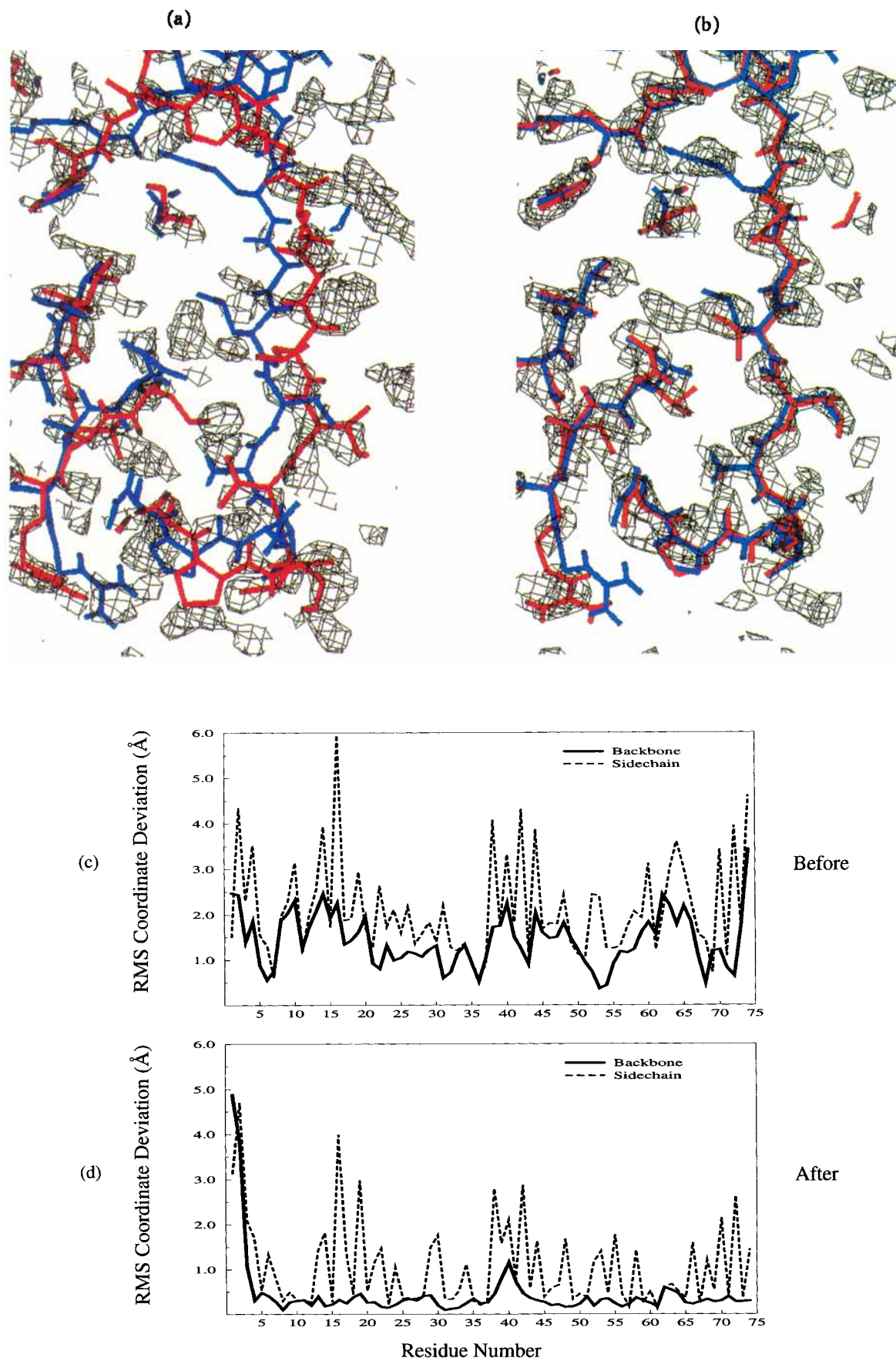


Fig. 7.

fore and after refinement are shown in Figure 7c,d. These improvements in both backbone and side chain deviations are remarkable, though backbone improvements are more pronounced. The large deviations for the first two residues correlate with large thermal factors ( $B > 45 \text{ \AA}^2$ ) observed in the crystal structure—this region is simply not very well determined by the diffraction data. The models resulting from these refinements are of course not fully refined, but are close enough to the crystal structure that more conventional techniques—which were previously inadequate—should now be applicable.

The simplification of the conformational search space provided by torsion angle dynamics significantly extends the radius of convergence of refinement *at relatively high resolution*. One might expect that it would similarly extend the radius of convergence of refinement at medium resolution. Refinement at 5 to 3 Å resolution with the same sampling strategy, however, does not significantly extend the radius of convergence (not shown). In fact, refinements of the least scrambled model show very sparse convergence: reducing the resolution limits to 5 to 3 Å reduces the number of reflections (observables) by a factor of 5, and results in a severely underestimated problem.

This adverse observable to parameter ratio can be improved using phase information from multiple isomorphous replacement (MIR) diffraction data. Using phase restraints [Eq. (4)] improves the radius of convergence so that all ten 5000 K refinements (for both conventional dynamics and torsion angle dynamics) of the least scrambled model converge, where convergence is now judged by a final model with a backbone atom rms coordinate deviation of less than 1 Å. Only two 5000 K conventional refinements and no 5000 K torsion angle refinements of the medium scrambled model converge. Increasing the temperature of the torsion angle refinements to 10,000 K increases the success rate: four of the refinements of the medium scrambled model converge.

Medium resolution refinements against the vector residual [Eq. (5)] show a significantly increased radius of convergence. Ten refinements starting with the medium scrambled model converge 6 times out of 10 for torsion angle dynamics and 5 times out of 10 for conventional dynamics. This shows a clear improvement over phase restraints [Eq. (4)], however, in this case this probably reflects more on the

target than on the dynamics since relatively little is gained by high temperature reduced variable techniques. Ten conventional 5000 K refinements of the most scrambled model only converge once whereas ten 10,000 K torsion-angle refinements converge 5 times, indicating the benefits of high temperature torsion angle sampling. Refinements of an even worse model (with initial backbone atom rms coordinate deviation of 1.83 Å) confirm the value of high temperature torsion angle sampling: four 10,000 K torsion-angle refinements converge while no 5000 K conventional ones do.

Figure 8 summarizes convergence for the 5 to 3 Å refinement protocols, again using backbone atom rms coordinate deviations from the crystal structure as a measure of convergence. As for high resolution refinements, torsion angle refinements (Fig. 8b) consistently outperform their conventional counterparts (Fig. 8a). These refinements were performed without cross-validation because of the adverse parameter to observable ratio which makes  $R_{\text{free}}$  prone to larger fluctuations. Figure 9, however, shows a strong correlation between the  $R$  value and backbone atom rms coordinate deviations. This is in general not the case, as  $R$  values for medium resolution refinements performed without phase information do not distinguish as well between good and bad models.

## CONCLUSIONS

Molecular dynamics constrained to torsion angles provides a powerful tool for conformational searching. Applications to crystallographic refinement show significantly increased convergence over conventional techniques both at high and medium resolution (Figs. 5 and 8). The success of the method derives from the combination of simplifying the conformational space by constraining bond lengths and bond angles, sampling at high temperature, and repeating refinements with different initial conditions. The best solutions can be identified by the free  $R$  value, or by the  $R$  value if experimental phase information is appropriately included in the refinement.

Torsion angle refinement is not presented as a replacement for more standard techniques such as slow-cooling simulated annealing, but as a way to bring bad models within their reach without human intervention. Improved convergence for relatively high resolution refinements with no phase information should be particularly useful for molecular replacement techniques: more distant initial models will be tolerated. Medium resolution refinements also show improved convergence under torsion angle refinement, provided that experimental phase information is included appropriately. This should eliminate some intermediate model building, and may be useful for phase combination, again because more distant models are tolerated. In combination with

Fig. 7.  $(2|F_o| - |F_c|)e^{i\phi_c}$  electron density maps (a) before and (b) after refinement of the most scrambled structure. The final model shown is the one that has the lowest  $R_{\text{free}}$ . The crystal structure coordinates are shown in blue, and the model coordinates in red. Backbone atom (solid) and side chain atom (dashed) rms coordinate deviations from the crystal structure (c) before and (d) after refinement.

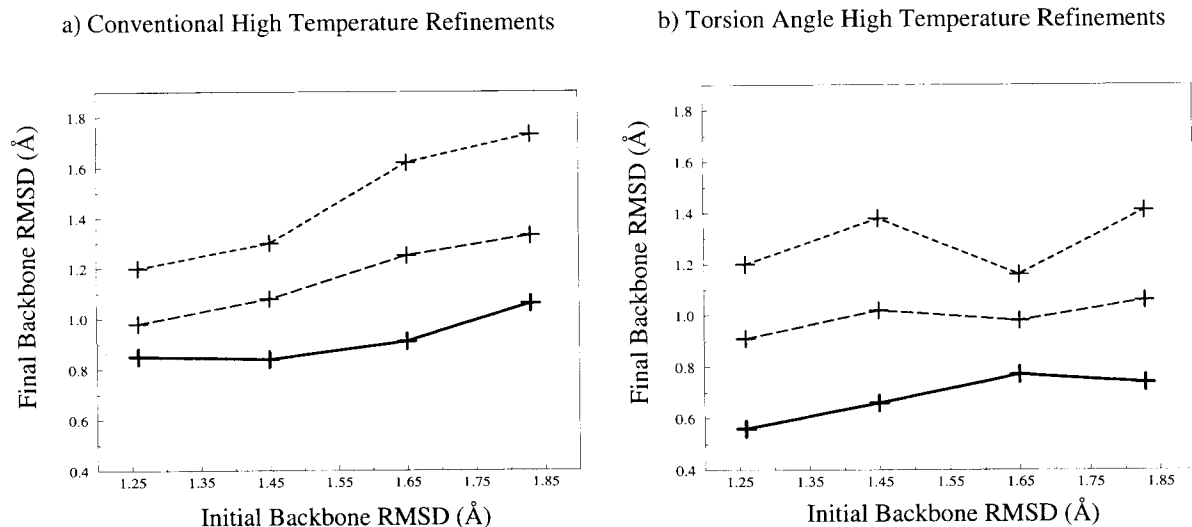


Fig. 8. Convergence of medium resolution (a) 5000 K conventional and (b) 10,000 K torsion angle refinements against the vector residual [Eq. (5)] of increasingly worse models. Convergence is measured by backbone atom rms coordinate deviations from the crystal structure, with lowest values shown in solid lines,

average values in long-dashed lines, and highest values in short-dashed lines. The best models can be identified by a low  $R$  value (cf. Fig. 9). Averages were calculated over 10 refinements for every model.

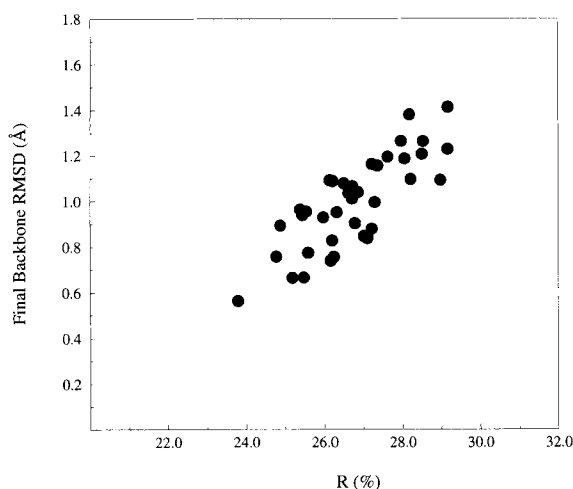


Fig. 9. Final backbone atom rms coordinate deviations plotted vs.  $R$  for all 10,000 K torsion angle refinements against the vector residual [Eq. (5)].

automatic side chain placement,<sup>36</sup> high temperature torsion angle sampling could conceivably make the refinement process fully automatic.

Our combined reduced variable high temperature sampling strategy could be generalized by restricting even more degrees of freedom, for example, by constraining secondary structure elements or tertiary domains. This could potentially extend the scope of refinement to significantly lower resolution, perhaps allowing refinement of very large macromolecular complexes where only low resolution data are available.

Crystallographic structure determination is but one of many applications of computer modeling of macromolecules that depend on the ability to search conformational space effectively. The use of torsion angle molecular dynamics as a sampling strategy should not be limited to crystallographic structure determination: other applications that depend critically on efficient conformational searching such as solution NMR structure determination or de novo structure prediction should also benefit from a reduced variable sampling technique.

#### ACKNOWLEDGMENTS

We thank P. Adams, T. Burling, P. Gros, and W. DeLano for valuable discussions and suggestions, and H. Chun, J. Myer, and C. Padilla at Moldyn, Inc. for exposure to reduced variable molecular dynamics algorithms. J. Pflugrath kindly provided diffraction data and coordinates for the  $\alpha$ -amylase inhibitor. This work was supported in part by a grant from the National Science Foundation to ATB (DIR 9021975).

#### REFERENCES

1. Beveridge, D.L., DiCapua, F.M. Free energy via molecular simulation: Applications to biomolecular systems. *Annu. Rev. Biophys. Chem.* 18:431–492, 1989.
2. Hodel, A., Simson, T., Fox, R.O., Brünger, A.T. Conformational substates and uncertainty in macromolecular free energy calculations. *J. Phys. Chem.* 97:3409–3417, 1993.
3. Di Nola, A., Roccatano, D., Berendsen, H.J.C. Molecular dynamics simulation of the docking of substrates to proteins. *Proteins* 19:174–182, 1994.
4. Kostrowicki, J., Scheraga, H.A. Application of the diffusion equation method for global optimization to oligopeptides. *J. Phys. Chem.* 96:7442–7449, 1992.
5. Brünger, A.T., Kuriyan, J., Karplus, M. Crystallographic

- R factor refinement by molecular dynamics. *Science* 235: 458–460, 1987.
6. Brünger, A.T., Nilges, M. Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR-spectroscopy. *Quat. Rev. Biophys.* 26:49–125, 1993.
  7. Levinthal, C. Are there pathways of protein folding? *J. Chem. Phys.* 65:44–45, 1968.
  8. Kirkpatrick, S., Gelatt, C.D., Jr., Vecchi, M.P. Optimization by simulated annealing. *Science* 220:671–680, 1983.
  9. Diamond, R. A Real-space refinement procedure for proteins. *Acta Cryst.* A27:436–452, 1971.
  10. Hendrickson, W.A. Stereochemically restrained refinement of macromolecular structures. *Meth. Enzymol.* 11: 252–270, 1985.
  11. Ryckaert, J.P., Ciccotti, G., Berendsen, H.J.C. Numerical integration of cartesian equations of motion of a system with constraints—molecular dynamics of N-alkanes. *J. Comp. Phys.* 23:327–341, 1977.
  12. Mazur, A.K., Abagyan, R.A. New methodology for computer-aided modelling of biomolecular structure and dynamics. (I) Non-cyclic structures. *J. Biomol. Struct. Dynam.* 4:815–832, 1989.
  13. Jain, A., Vaidehi, N., Rodriguez, G. A fast recursive algorithm for molecular dynamics simulations. *J. Comp. Phys.* 106:258–268, 1993.
  14. Rodriguez, G. Kalman filtering, smoothing, and recursive robot arm forward and inverse dynamics. *IEEE J. Robotics Autom.* 6:624–639, 1987.
  15. Bae, D.-S., Haug, E.J. A recursive formulation for constrained mechanical system dynamics: Part I. Open loop systems. *Mech. Struct. Mach.* 15:359–382, 1987.
  16. Bae, D.-S., Haug, E.J. A recursive formulation for constrained mechanical system dynamics: Part II. Closed loop systems. *Mech. Struct. Mach.* 15:481–506, 1988.
  17. Brünger, A.T. The free  $R$  value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature (London)* 355:472–474, 1992.
  18. Hauptman, H.A. The phase problem of X-ray crystallography. *Phys. Today* 42(11):24–29, 1989.
  19. Nilges, M., Clore, G.M., Gronenborn, A.M. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett.* 229:317–324, 1988.
  20. Hahn, M., Heinemann, U. DNA helix structure and refinement algorithm: Comparison of models for d(CCAGGCm<sup>5</sup>CTGG) derived from NUCLSQ, TNT, and X-PLOR. *Acta Cryst.* D5:468–477, 1993.
  21. Brünger, A.T. Crystallographic refinement by simulated annealing: Application to a 2.8 Å resolution structure of aspartate aminotransferase. *J. Mol. Biol.* 203:803–816, 1988.
  22. Arnold, E., Rossmann, M.G., The use of molecular-replacement phases for the refinement of the human rhinovirus 14 structure. *Acta Cryst.* A44:270–282, 1988.
  23. Jack, A., Levitt, M. Refinement of large structures by simultaneous minimization of energy and  $R$  factor. *Acta Cryst.* A34:931–935, 1978.
  24. Tronrud, D.E., Ten Eyck, L.F., Matthews, B.W. An efficient general purpose least squares refinement program for macromolecular structures. *Acta Cryst.* A43:489–500, 1987.
  25. Laarhoven, P.J.M., Aarts, E.H.L., eds. "Simulated Annealing: Theory and Applications." Dordrecht: D. Reidel, 1987: 187.
  26. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690, 1984.
  27. Verlet, L. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* 159:98–105, 1967.
  28. Abramowitz, M., Stegun, I. "Handbook of Mathematical Functions, Applied Mathematics Series," Vol. 55. New York: Dover Publications, 1968: 897.
  29. Goldstein, H. "Classical Mechanics." Reading, MA: Addison-Wesley, 1980: 16–21.
  30. Brünger, A.T. X-PLOR, Version 3.1. A System for X-ray Crystallography and NMR. New Haven: Yale University Press, 1992.
  31. Engh, R.A., Huber, R. Accurate bond and angle parameters for X-ray protein-structure refinement. *Acta Cryst.* A47:392–400, 1991.
  32. Gros, P., Betzel, Ch., Dauter, Z., Wilson, K.S., Hol, W.G.J. Molecular dynamics refinement of a thermitase-eglin-c Complex at 1.98 Å resolution and comparison of two crystal forms that differ in calcium content. *J. Mol. Biol.* 210: 347–367, 1989.
  33. Kuriyan, J., Brünger, A.T., Karplus, M., Hendrickson, W.A. X-ray refinement of protein structures by simulated annealing: Test of the method on myohemerythrin. *Acta Cryst.* A45:396–409, 1989.
  34. Brünger, A.T., Krukowski, A., Erickson, J.W. Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Cryst.* A46:585–593, 1990.
  35. Pflugrath, J.W., Wiegand, G., Huber, R., Vertesey, L. Crystal structure determination, refinement, and the molecular model of the  $\alpha$ -amylase inhibitor Hoe-467A. *J. Mol. Biol.* 189:383–386, 1989.
  36. Jones, T.A., Zou, J.Y., Cowan, S.W., Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.* A47:110–119, 1991.
  37. Bae, D.S. A Recursive formulation for constrained mechanical system dynamics. Ph.D. Dissertation, University of Iowa, Iowa City, Iowa, 1986.

## APPENDIX

Here the equations of motion for molecular dynamics constrained to torsion angles are derived, again considering the simplest example of two connected bodies. Referring to Figure 1, the vectors  $\mathbf{s}_{ij}$  and  $\mathbf{s}_{ji}$  define the points of attachment in each body with respect to its center of mass. These vectors are fixed in the center of mass reference frame of each body, and the "bond" vector  $\mathbf{h}_{ij}$  is fixed in the center of mass frame of body  $i$ . The time derivatives of vectors  $\mathbf{s}_{ij}$  and  $\mathbf{h}_{ij}$  are therefore given by  $\omega_i \times \mathbf{s}_{ij}$  and  $\omega_i \times \mathbf{h}_{ij}$ . Similarly, the time derivative of the vector  $\mathbf{s}_{ji}$  is given by  $\omega_j \times \mathbf{s}_{ji}$ . These relationships give rise to the simpler ones presented before:

$$\begin{aligned} \mathbf{Y}_j &= \begin{bmatrix} \mathbf{I} & -\mathbf{r}_{ij} \\ \mathbf{O} & \mathbf{I} \end{bmatrix} \mathbf{Y}_i + \begin{bmatrix} -\dot{\mathbf{h}}_{ij} \times \mathbf{s}_{ji} \\ \dot{\mathbf{h}}_{ij} \end{bmatrix} \dot{\mathbf{q}}_{ij} \quad (19) \\ &= \mathbf{B}_{ij}^{(1)} \mathbf{Y}_i + \mathbf{B}_{ij}^{(2)} \dot{\mathbf{q}}_{ij} \end{aligned}$$

where the notation  $\mathbf{r}_{ij}$  denotes the  $3 \times 3$  antisymmetric matrix representing the vector cross product.

Bae and Haug<sup>15,16</sup> used D'Alembert's principle to derive the equations of motion. If a system of particles is in equilibrium, then the net force vanishes:  $\sum_i \mathbf{F}_i \cdot \delta \mathbf{r}_i = 0$ . For a system of particles not in equilibrium, one may still write  $\sum_i (\mathbf{F}_i - \dot{\mathbf{p}}_i) \cdot \delta \mathbf{r}_i = 0$ , where  $\delta \mathbf{r}_i$  represents an arbitrary displacement. To obtain equations of motion for dynamics under holonomic constraints, one considers only those displacements  $\delta \mathbf{Z}_i$  which do not violate the constraints. Unlike the completely arbitrary set of  $\delta \mathbf{r}_i$ , the  $\delta \mathbf{Z}_i$  are related through the constraints. This dependency can be manipulated to give the equations of motion for a constrained system.

Under these assumptions, one obtains the equations of motion presented earlier:

$$\sum_{\text{bodies } k} \delta \mathbf{Z}_k^T (\mathbf{M}_k \dot{\mathbf{Y}}_k - \mathbf{Q}_k) = 0 \quad (20)$$

where “ $T$ ” denote the transpose, and the equation must hold for all  $\delta \mathbf{Z}_k$  which do not violate the constraints. Here the matrix  $\mathbf{M}_k$  is given by

$$\mathbf{M}_i = \begin{bmatrix} m_i & 0 & 0 & 0 & 0 & 0 \\ 0 & m_i & 0 & 0 & 0 & 0 \\ 0 & 0 & m_i & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{xx} & I_{xy} & I_{xz} \\ 0 & 0 & 0 & I_{xy} & I_{yy} & I_{yz} \\ 0 & 0 & 0 & I_{xz} & I_{yz} & I_{zz} \end{bmatrix} \quad (21)$$

and the generalized force vector  $\mathbf{Q}_k$  by

$$\mathbf{Q}_i = \begin{bmatrix} F_i^x \\ F_i^y \\ F_i^z \\ N_i^x - (\omega_i \times \mathbf{I} \omega_i)_x \\ N_i^y - (\omega_i \times \mathbf{I} \omega_i)_y \\ N_i^z - (\omega_i \times \mathbf{I} \omega_i)_z \end{bmatrix}$$

where  $\mathbf{I}$  is the  $3 \times 3$  inertia tensor, and  $N_i^x, N_i^y, N_i^z$  are the components of the net torque (measured in the lab frame) about the center of mass. The components of  $\omega_i \times \mathbf{I} \omega_i$  also appeared in  $\mathbf{Q}_k$  to account for the fact that the net torque  $\mathbf{N}_i$  is not given by  $\mathbf{N}_i = \mathbf{I} \omega_i$  but rather by  $\mathbf{N}_i = \mathbf{I} \dot{\omega}_i + \omega_i \times \mathbf{I} \omega_i$ , and only  $\mathbf{I} \omega_i$  appears in the term  $\mathbf{M}_i \dot{\mathbf{Y}}_i$ . All that remains is to work through the algebra to show that this does indeed provide the dynamic information in a format useful for a molecular dynamics algorithm. Expanding the sum for the simple two-body case:

$$\delta \mathbf{Z}_i^T (\mathbf{M}_i \dot{\mathbf{Y}}_i - \mathbf{Q}_i) + \delta \mathbf{Z}_j^T (\mathbf{M}_j \dot{\mathbf{Y}}_j - \mathbf{Q}_j) = 0. \quad (23)$$

Substituting for

$$\delta \mathbf{Z}_j = \mathbf{B}_{ij}^{(1)} \delta \mathbf{Z}_i + \mathbf{B}_{ij}^{(2)} \delta \mathbf{q}_{ij} \quad (24)$$

and

$$\mathbf{Y}_j = \mathbf{B}_{ij}^{(1)} \mathbf{Y}_i + \mathbf{B}_{ij}^{(2)} \dot{\mathbf{q}}_{ij}$$

and

$$\dot{\mathbf{Y}}_j = \mathbf{B}_{ij}^{(1)} \dot{\mathbf{Y}}_i + \mathbf{B}_{ij}^{(2)} \ddot{\mathbf{q}}_{ij} + \mathbf{D}_{ij}$$

yields

$$0 = \delta \mathbf{Z}_i^T (\mathbf{M}_i \dot{\mathbf{Y}}_i - \mathbf{Q}_i) + \delta \mathbf{Z}_j^T (\mathbf{M}_j \dot{\mathbf{Y}}_j - \mathbf{Q}_j) \quad (25)$$

$$= \delta \mathbf{Z}_i^T (\mathbf{M}_i \dot{\mathbf{Y}}_i - \mathbf{Q}_i) + (\delta \mathbf{q}_{ij}^T \mathbf{B}_{ij}^{(2)T} + \delta \mathbf{Z}_i^T \mathbf{B}_{ij}^{(1)T}) \{ \mathbf{M}_j [\mathbf{B}_{ij}^{(1)} \dot{\mathbf{Y}}_i + \mathbf{B}_{ij}^{(2)} \ddot{\mathbf{q}}_{ij} + \mathbf{D}_{ij}] - \mathbf{Q}_j \}.$$

Factoring terms in  $\delta \mathbf{Z}_i$  and  $\delta \mathbf{q}_{ij}$  gives

$$\begin{aligned} & \delta \mathbf{Z}_i^T \{ \mathbf{M}_i \dot{\mathbf{Y}}_i + \mathbf{B}_{ij}^{(1)T} \mathbf{M}_j \mathbf{B}_{ij}^{(1)} \dot{\mathbf{Y}}_i + \mathbf{B}_{ij}^{(1)T} \mathbf{M}_j \mathbf{B}_{ij}^{(2)} \ddot{\mathbf{q}}_{ij} + (26) \\ & \quad \mathbf{B}_{ij}^{(1)T} \mathbf{M}_j \mathbf{D}_{ij} - \mathbf{Q}_i - \mathbf{B}_{ij}^{(1)T} \mathbf{Q}_j + \\ & \quad \delta \mathbf{q}_{ij}^T \{ \mathbf{B}_{ij}^{(2)T} \mathbf{M}_j \mathbf{B}_{ij}^{(1)} \dot{\mathbf{Y}}_i + \mathbf{B}_{ij}^{(2)T} \mathbf{M}_j \mathbf{B}_{ij}^{(2)} \ddot{\mathbf{q}}_{ij} + \\ & \quad \mathbf{B}_{ij}^{(2)T} \mathbf{M}_j \mathbf{D}_{ij} - \mathbf{B}_{ij}^{(2)T} \mathbf{Q}_j \} = 0. \end{aligned}$$

But in this equation the quantities  $\delta \mathbf{Z}_i$  and  $\delta \mathbf{q}_{ij}$  are, by construction, linearly independent. Thus both coefficients must vanish for the sum to vanish. There are two unknowns,  $\mathbf{Y}_i$  and  $\mathbf{q}_{ij}$ , the acceleration of body  $i$  and the relative acceleration about the bond  $\mathbf{h}_{ij}$ , respectively. First one sets the coefficient of  $\delta \mathbf{q}_{ij}$  to zero and solves for  $\mathbf{q}_{ij}$ .

$$\begin{aligned} \ddot{\mathbf{q}}_{ij} = & - (\mathbf{B}_{ij}^{(2)T} \mathbf{M}_j \mathbf{B}_{ij}^{(2)})^{-1} \{ \mathbf{B}_{ij}^{(2)T} \mathbf{M}_j \mathbf{B}_{ij}^{(1)} \dot{\mathbf{Y}}_i \quad (27) \\ & + \mathbf{B}_{ij}^{(2)T} [\mathbf{M}_j \mathbf{D}_{ij} - \mathbf{Q}_j] \}. \end{aligned}$$

This expression for  $\mathbf{q}_{ij}$  can now be substituted into the coefficient of  $\delta \mathbf{Z}_i$  and providing an expression for  $\mathbf{Y}_i$ , the acceleration of body  $i$ . This value of  $\mathbf{Y}_i$  determines the value of  $\mathbf{q}_{ij}$ , and thus gives the accelerations for the system. (Proof the existence of inverses of necessary matrices appears in Bae.<sup>37</sup>)

A more complicated topology is treated analogously. For a long branch free chain of  $n$  bodies, the first step is to write  $\delta \mathbf{Z}_n$  as a function of  $\delta \mathbf{Z}_{n-1}$  and  $\delta \mathbf{q}_{(n-1)n}$ . But  $\delta \mathbf{q}_{(n-1)n}$  is linearly independent of all other virtual displacements, so one can use Eq. (27) to solve for the relative acceleration  $\delta \mathbf{q}_{(n-1)n}$  between the end body and one body inboard. An expression is thereby obtained for a chain of  $(n-1)$  bodies, and the coefficient of  $\delta \mathbf{Z}_{n-1}$  contains additional terms to account for contributions from the  $n$ th body (cf. Eq. (26)). Then one writes  $\delta \mathbf{Z}_{n-1}$  as a function of  $\delta \mathbf{Z}_{n-2}$  and  $\delta \mathbf{q}_{(n-2)(n-1)}$ , and the reduction continues until an expression for the first body's acceleration is obtained. Proceeding back out the chain, relative accelerations can be computed. Branched chains and closed loops can be dealt with as well; details appear in Bae and Haug.<sup>16</sup>