

Torsion-Angle Molecular Dynamics as a New Efficient Tool for NMR Structure Calculation

EVAN G. STEIN, LUKE M. RICE, AND AXEL T. BRÜNGER*

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

Received July 18, 1996; revised October 2, 1996

Molecular dynamics in torsion-angle space was applied to nuclear magnetic resonance structure calculation using nuclear Overhauser effect-derived distances and J -coupling-constant-derived dihedral angle restraints. Compared to two other commonly used algorithms, molecular dynamics in Cartesian space and metric-matrix distance geometry combined with Cartesian molecular dynamics, the method shows increased computational efficiency and success rate for large proteins, and it shows a dramatically increased radius of convergence for DNA. The torsion-angle molecular dynamics algorithm starts from an extended strand conformation and proceeds in four stages: high-temperature torsion-angle molecular dynamics, slow-cooling torsion-angle molecular dynamics, Cartesian molecular dynamics, and minimization. Tests were carried out using experimental NMR data for protein G, interleukin-8, villin 14T, and a 12 base-pair duplex of DNA, and simulated NMR data for bovine pancreatic trypsin inhibitor. For villin 14T, a monomer consisting of 126 residues, structure determination by torsion-angle molecular dynamics has a success rate of 85%, a more than twofold improvement over other methods. In the case of the 12 base-pair DNA duplex, torsion-angle molecular dynamics had a success rate of 52% while Cartesian molecular dynamics and metric-matrix distance geometry always failed. © 1997 Academic Press

INTRODUCTION

The goal of nuclear magnetic resonance structure calculation is to simultaneously satisfy experimentally observed NMR data (nuclear Overhauser effects, J coupling constants, and chemical shifts) and chemical information (stereochemistry and nonbonded interactions). Structure-calculation methods based on metric-matrix distance geometry or on molecular dynamics in Cartesian space sometimes show a low success rate for large molecules or poorly determined systems (1). To increase the success rate, the two methods are often used in combination (2). However, even this combined method is sometimes unsuccessful; thus, there is a need for improvement in algorithms for NMR structure calculation,

especially as advances in NMR technology allow larger molecules to be studied.

Molecular dynamics-based structure calculation has been shown to be capable of folding a protein starting from an extended strand (3). Since the original implementation was prone to numerical instabilities, a “soft” repulsive nonbonded potential and gradient-limited distance restraints were introduced (5, 6). Although very successful at refining small proteins and small nucleic acids, even this improved implementation sometimes fails to converge when applied to large structures.

Fixed-length and fixed-angle constraints can be imposed in order to reduce approximately 10-fold the number of adjustable parameters that characterize a model, thus creating a more favorable observable-to-parameter ratio for an NMR structure-calculation method. These constraints are most useful at high simulated annealing temperatures where conventional molecular dynamics allows significant deviations from ideal geometry. Indeed, fixed-length constraints have been applied to structure calculation by gradient descent minimization (7, 9), and by Monte Carlo minimization (8, 10, 11). It is only recently, however, that efficient and robust algorithms have become available for molecular dynamics in torsion-angle space (12–16).

In this paper, a new NMR structure-calculation method is described that uses molecular dynamics constrained to torsion-angle space. Convergence to a correct final model is achieved, starting from an extended strand conformation. The first stage of the protocol consists of an initial search of torsion-angle space at a high temperature with a decreased weight on the repulsive energy term. A second stage of torsion-angle dynamics follows, in which the temperature of the system is gradually reduced while the weight on the repulsive energy term is linearly increased to unity. Two stages follow where the bond lengths and bond angles are allowed to relax.

Several test cases are used to compare the torsion-angle dynamics method to Cartesian molecular dynamics and to combined metric-matrix distance geometry and Cartesian

* To whom correspondence should be addressed.

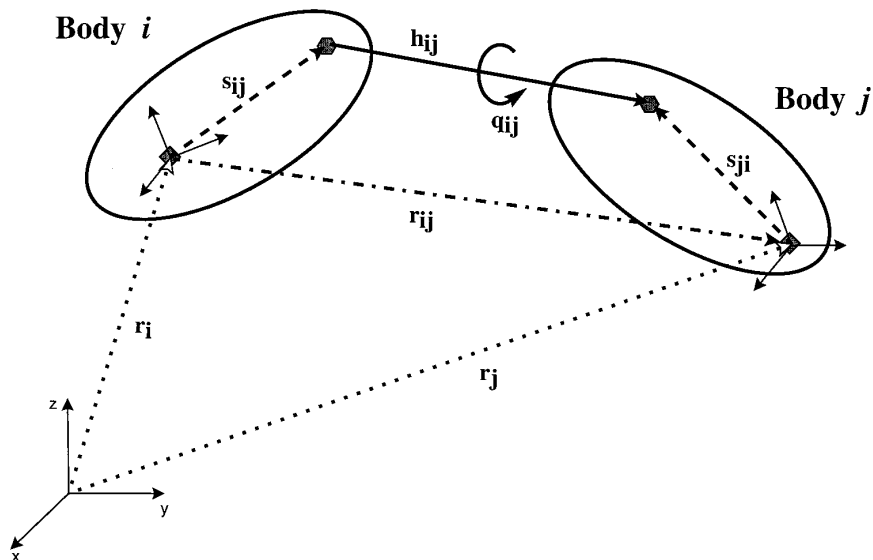


FIG. 1. Diagram showing variables and definitions for torsion angle molecular dynamics involving two connected bodies. Vectors \mathbf{r}_i and \mathbf{r}_j locate the centers of mass (in an arbitrary inertial frame) of bodies i and j , respectively. See (15) for more details.

molecular dynamics. The methods are compared by their computational efficiency and their success rate.

tion-based optimization problem (18). The latter approach uses the energy function

METHODS

Energy Function

NMR structure calculation can be formulated as a distance-geometry problem (17) or as an hybrid-energy-func-

$$E = E_{\text{chem}} + E_{\text{nmr}} \quad [1]$$

$$E_{\text{nmr}} = w_{\text{NOE}}E_{\text{NOE}} + w_{\text{dihedral}}E_{\text{dihedral}} \quad [2]$$

$$E_{\text{chem}} = E_{\text{geom}} + w_{\text{vdw}}E_{\text{vdw}}, \quad [3]$$

TABLE 1
NMR Data Statistics

	BPTI	Protein G	Interleukin-8	Villin 14T	DNA
Number of residues	58	56	144(dimer)	126	24
NOE-derived distance restraints					
Intraresidue ^a	0	307	532	73	102
Interresidue short range ^b	472	223	768	622	107
Interresidue long range ^c	240	259	420	625	19
Total	712	789	1720	1320	228
Hydrogen bond restraints					
Intraresidue ^a	0	0	0	0	0
Interresidue short range ^b	0	30	72	36	6
Interresidue long range ^c	0	38	52	50	24
Total	0	68	124	86	30
Dihedral angle restraints					
ϕ	0	54	136	68	N/A
ψ	0	0	122	0	N/A
χ_1	0	39	104	52	N/A
χ_2	0	12	0	0	N/A
Total	0	105	362	120	136

^a Distance restraints within the same residue.

^b Distance restraints between residues whose sequence separation is less than or equal to four residues.

^c Distance restraints between residues whose sequence separation is greater than four residues.

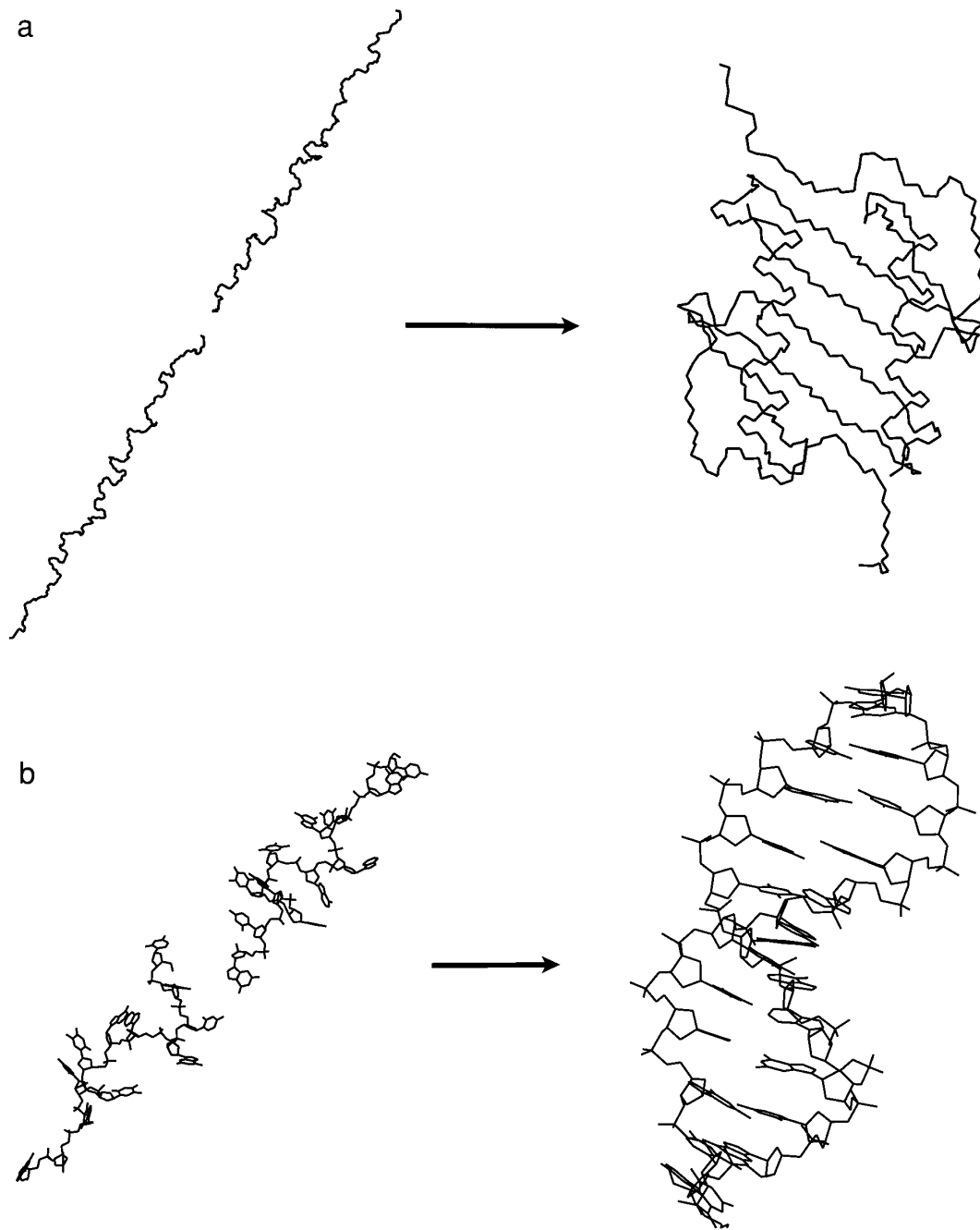


FIG. 2. (a) The initial extended strands after regularization and average structure of interleukin-8, a 144-residue dimer, obtained by torsion-angle molecular dynamics starting from the extended strands. (b) The initial extended strands after regularization and average structure of the DNA dodecamer (CGCGPATTCGCG), obtained by torsion-angle molecular dynamics starting from the extended strands.

where E_{chem} describes agreement with expected values for bond lengths, bond angles, planarity, chirality, and non-bonded interactions, consisting of van der Waals, hydrogen bonding, and electrostatic contributions (18). However, because solvent is neglected from the structure calculations, electrostatic interactions are excluded and hydrogen bonds

are modeled as pseudo-NOEs. Rather than use the Lennard-Jones potential,

$$E_{\text{vdw}} = \left\{ 4\epsilon \left[\left(\frac{\sigma}{R} \right)^{12} - \left(\frac{\sigma}{R} \right)^6 \right] \right\}, \quad [4]$$

TABLE 2
Torsion-Angle Molecular Dynamics Protocol

	Stage 1	Stage 2	Stage 3	Stage 4
	High-temperature torsion-angle molecular dynamics	Slow-cooling torsion-angle molecular dynamics	Slow cooling Cartesian molecular dynamics	1000 steps conjugate gradient minimization
Temperature	50,000 K (20,000 K)	50,000 K (20,000 K) → 1000 K	1000 K → 300 K	—
Time step	0.015 ps	0.015 ps	0.003 ps	—
Δt	15 ps (60 ps)	15 ps (60 ps)	6 ps	—
w_{NOE}	150	150	150	50
w_{dihedral}	100 (5)	100 (5)	100	300
w_{vdw}	0.1	0.1 → 1.0	1.0	1.0

Note. Listed are the temperature for each stage of the protocol, duration of the stage, and weights for the energy terms (Eqs. [1], [2], [3]) for protein structure calculations. An arrow (→) indicates that the value of that parameter linearly changes over the course of the particular stage. Parameters for nucleic acid refinement are shown in parentheses.

the van der Waals interactions were described by a purely repulsive quartic potential,

$$E_{\text{vdw}} = [(0.8\sigma\sqrt{2})^2 - R^2]^2, \quad [5]$$

where R is the distance between two atoms and ϵ and σ are the Lennard-Jones parameters for a particular atom pair. However, for the final analysis of the refined NMR structures, Eq. [4] was used. Force field parameters were taken from a parameter set designed for NMR refinement of proteins (parallhdg.pro) and of nucleic acids (parallhdg.dna) (2, 19).

NOE-derived distance restraints were described by a flat-bottomed parabolic (square-well) function with a soft asymptote (6, 18, 22)

$$E_{\text{NOE}} = \min \begin{cases} \Delta^2 & d_{\text{upper}} + 0.5 > R \\ a + \frac{b}{\Delta_{\text{softexp}}} + \Delta & d_{\text{upper}} + 0.5 < R \end{cases} \quad [6]$$

where Δ is defined as

$$\Delta = \begin{cases} (R - d_{\text{upper}}) & d_{\text{upper}} < R \\ 0 & d_{\text{lower}} < R < d_{\text{upper}} \\ (d_{\text{lower}} - R) & R < d_{\text{lower}}. \end{cases} \quad [7]$$

Here R is the distance between a particular pair of spins in the model, d_{lower} is the lower bound for the distance, d_{upper} is the upper bound, and a and b are determined such that E_{NOE} is a differentiable function at the point $R = d_{\text{upper}} + 0.5$. The sum is carried out over all the NOEs.

Dihedral angle restraints derived from J coupling constant measurements were described by

$$E_{\text{dihedral}} = \sum_{\text{dihedrals}} \begin{cases} (\phi - \phi_{\text{upper}})^2 & \phi_{\text{upper}} < \phi \\ 0 & \phi_{\text{lower}} < \phi < \phi_{\text{upper}} \\ (\phi_{\text{lower}} - \phi)^2 & \phi < \phi_{\text{lower}}. \end{cases} \quad [8]$$

The torsion-angle molecular dynamics method in principle allows for the inclusion of other functional forms such as direct refinement against NOEs, J -coupling values based on the Karplus equation (20), and restraints derived from chemical shifts (21).

Molecular Dynamics

Structure calculation based on molecular dynamics (3, 4) consists of the numerical integration of Newton's equations of motion

$$m_i \frac{\partial^2 r_{i,u}}{\partial t^2} = - \frac{\partial E}{\partial r_{i,u}}, \quad [9]$$

where $r_{i,u}$ and m_i are the coordinates and mass, respectively, of atom i , and E is the hybrid energy function (Eq. [1]). Temperature control, required for simulated annealing (23), was performed by temperature coupling (24),

$$m_i \frac{\partial^2 r_{i,u}}{\partial t^2} = - \frac{\partial E}{\partial r_{i,u}} + \beta_i \left(\frac{T_0}{T} - 1 \right) \dot{v}_i, \quad [10]$$

where T_0 is the temperature of the bath to which the system is coupled, β_i is a force constant, T is the temperature of the system, and \dot{v}_i is the velocity of each atom i . Temperature coupling will cause ‘‘heat’’ to either be added or removed

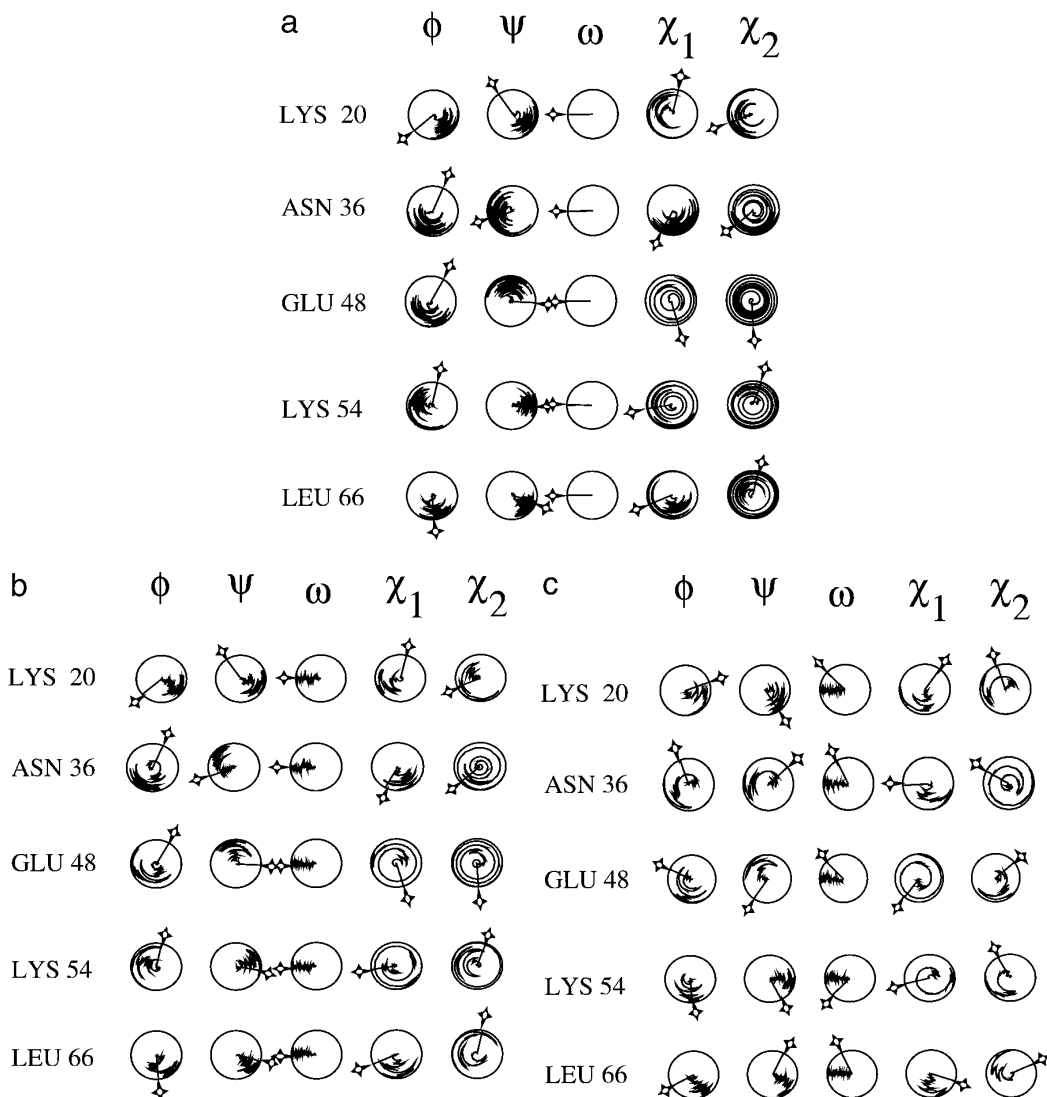


FIG. 3. Dial plots (30) showing the change in specified torsion angles versus molecular dynamics time (Eq. [9]). The initial torsion angle is indicated by an arrow. The radius of each sphere represents time while the rotation represents the torsion angle. A representative set of residues was chosen. Results are shown for (a) torsion-angle molecular dynamics, (b) SA, and (c) DGSA.

from the system (in the form of kinetic energy) as it is needed to maintain the temperature.

Torsion-Angle Molecular Dynamics

What follows is a simplified sketch of one implementation of torsion-angle constrained molecular dynamics (15), following the algorithm of Bae and Haug (12, 13). Consider two bodies (Fig. 1), i and j , connected by a bond of fixed length $|\mathbf{h}_{ij}|$. Assuming that the only allowable relative motion between the two bodies is a rotation about \mathbf{h}_{ij} , let \mathbf{r}_i and \mathbf{r}_j locate (with respect to an arbitrary inertial frame) the center of mass of body i and j , respectively. Let \mathbf{s}_{ij} (\mathbf{s}_{ji}) locate the endpoint of \mathbf{h}_{ij} on body i (j) with respect to its center of

mass. Thus, \mathbf{s}_{ij} is a vector from the center of mass of body i to the end of \mathbf{h}_{ij} . 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} (cf. (15) for more details).

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 \mathbf{w} of their respective centers of mass measured in an inertial (“lab”) frame:

$$\mathbf{w}_j = \mathbf{w}_i + \hat{\mathbf{h}}_{ij} \dot{q}_{ij}. \quad [11]$$

TABLE 3
RMS Difference from the Average Structure

	BPTI	Protein G	Interleukin-8	Villin 14T
Torsion-angle mol. dyn.	$0.34 \pm 0.07 \text{ \AA}$	$0.26 \pm 0.05 \text{ \AA}$	$1.68 \pm 0.40 \text{ \AA}$	$1.38 \pm 0.27 \text{ \AA}$
SA	$0.43 \pm 0.15 \text{ \AA}$	$0.35 \pm 0.05 \text{ \AA}$	$1.81 \pm 0.24 \text{ \AA}$	$1.42 \pm 0.36 \text{ \AA}$
DGSA	$0.43 \pm 0.14 \text{ \AA}$	$0.30 \pm 0.04 \text{ \AA}$	$1.72 \pm 0.28 \text{ \AA}$	$1.54 \pm 0.17 \text{ \AA}$

Note. The mean and standard deviation of the root-mean-square difference (\AA) from the average structure for backbone atoms (O, C, C $^\alpha$, N).

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

$$\begin{aligned} \mathbf{r}_j &= \mathbf{r}_i + \mathbf{r}_{ij} \\ &= \mathbf{r}_i + \mathbf{s}_{ij} + |\mathbf{h}_{ij}|\hat{\mathbf{h}}_{ij} - \mathbf{s}_{ji}. \end{aligned} \quad [12]$$

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 :

$$\begin{aligned} \dot{\mathbf{r}}_j &= \dot{\mathbf{r}}_i + \dot{\mathbf{s}}_{ij} + |\dot{\mathbf{h}}_{ij}|\hat{\mathbf{h}}_{ij} - \dot{\mathbf{s}}_{ji} \\ &= \dot{\mathbf{r}}_i + \mathbf{w}_i \times \mathbf{s}_{ij} + |\mathbf{h}_{ij}|\mathbf{w}_i \times \hat{\mathbf{h}}_{ij} - \mathbf{w}_j \times \mathbf{s}_{ji} \\ &= \dot{\mathbf{r}}_i - \mathbf{s}_{ij} \times \mathbf{w}_i - |\mathbf{h}_{ij}|\hat{\mathbf{h}}_{ij} \times \mathbf{w}_i + \mathbf{s}_{ji} \times \mathbf{w}_i \\ &\quad - \dot{q}_{ij}\hat{\mathbf{h}}_{ij} \times \mathbf{s}_{ji} \\ &= \dot{\mathbf{r}}_i - \mathbf{r}_{ij} \times \mathbf{w}_i - (\hat{\mathbf{h}}_{ij} \times \mathbf{s}_{ji})\dot{q}_{ij}. \end{aligned} \quad [13]$$

Thus, assuming certain constraints act between atoms or groups of atoms, one can obtain an expression for the velocity of one group in terms of the velocity of another. This relationship can be differentiated to give a relationship between accelerations, and integrated to give a relationship between positions (15).

The current implementation of the torsion-angle molecular dynamics algorithm cannot treat nonrigid closed bonding

networks exactly (15). The algorithm introduces an approximation whereby one bond in the closed network is allowed to vibrate. This could potentially cause numerical instabilities at high simulation temperatures for nucleotide ribose rings, and therefore, lower simulation temperatures are required (see below).

Test Cases

Structure calculations were carried out on protein G (25), BPTI (26), interleukin-8 (IL8) (27), villin 14T (28), and a short (12 base-pair) duplex strand of DNA (CGCGPATTC-GCG) (29) (Table 1). Nearly all ϕ , ψ , and χ_1 dihedrals are restrained for both the monomers of IL8, and most ϕ , χ_1 , and χ_2 dihedrals are restrained for protein G. There are fewer dihedral angle restraints for villin 14T: one dihedral restraint per residue on average, with regions of secondary structure more highly restrained. There are no dihedral angle restraints for BPTI. The dodecamer nucleic acid test case includes 136 dihedral angle restraints and 30 hydrogen bond restraints in addition to the NOE-derived distance restraints.

Torsion-Angle Molecular Dynamics Protocol

Initial structures consisted of extended strand conformations which were generated by sequentially placing all atoms along the x axis at tenth of an angstrom intervals, with y and z coordinates set to random numbers between zero and one. The initial coordinates were regularized using simulated annealing and conjugate-gradient minimization against E_{chem} (Eq. [3]) in order to obtain good local geometry (Fig. 2).

Details of the structure-calculation protocol are described in Table 2. In the first stage, the regularized extended strands are subjected to 15 ps of torsion-angle molecular dynamics at 50,000 K using the hybrid energy function E (Eq. [1]). To facilitate rotational barrier crossings, w_{vdw} (Eq. [3]) is set to 0.1. The structures are then subjected to a slow-cooling torsion-angle molecular dynamics stage in which the temperature is reduced from 50,000 to 1,000 K over a period of 15 ps while w_{vdw} is linearly increased from 0.1 to 1.0. The third stage of the protocol consists of a slow-cooling stage from 1,000 to 300 K for 6 ps using Cartesian molecular dynamics. Finally, the structure is subjected to 1000 steps of conjugate-gradient minimization.

TABLE 4
RMS Difference between Structure Calculation Methods

	BPTI	Protein G	Interleukin-8	Villin 14T
SA vs DGSA	0.06 \AA	0.08 \AA	0.71 \AA	0.93 \AA
SA vs torsion-angle mol. dyn.	0.21 \AA	0.13 \AA	0.60 \AA	0.85 \AA
DGSA vs torsion-angle mol. dyn.	0.21 \AA	0.14 \AA	0.90 \AA	0.49 \AA

Note. The root-mean-square difference (\AA) between the specified average structures for backbone atoms (O, C, C $^\alpha$, N).

TABLE 5
Geometry and Energy Statistics

	Torsion-angle molecular dynamics	SA	DGSA
BPTI			
$\langle \Delta_{\text{NOE}} \rangle$	0.047 \pm 0.0025 Å	0.049 \pm 0.0029 Å	0.049 \pm 0.0022 Å
$\langle \Delta_{\text{dihedral}} \rangle$	0	0	0
$\langle \Delta_{\text{bonds}} \rangle$	0.0034 \pm 0.00015 Å	0.0038 \pm 0.00025 Å	0.0038 \pm 0.00023 Å
$\langle \Delta_{\text{angles}} \rangle$	0.67° \pm 0.013°	0.67° \pm 0.029°	0.67° \pm 0.023°
$\langle E_{\text{vdw}} \rangle$	-50.3 kcal mol ⁻¹	47.1 kcal mol ⁻¹	43.5 kcal mol ⁻¹
Protein G			
$\langle \Delta_{\text{NOE}} \rangle$	0.017 \pm 0.00075 Å	0.016 \pm 0.00080 Å	0.016 \pm 0.00057 Å
$\langle \Delta_{\text{dihedral}} \rangle$	0.24° \pm 0.017°	0.33° \pm 0.033°	0.31° \pm 0.047°
$\langle \Delta_{\text{bonds}} \rangle$	0.0018 \pm 0.00003 Å	0.0017 \pm 0.00007 Å	0.0016 \pm 0.00005 Å
$\langle \Delta_{\text{angles}} \rangle$	0.48° \pm 0.036°	0.47° \pm 0.043°	0.46° \pm 0.026°
$\langle E_{\text{vdw}} \rangle$	-103.5 kcal mol ⁻¹	-12.6 kcal mol ⁻¹	-13.1 kcal mol ⁻¹
Interleukin-8			
$\langle \Delta_{\text{NOE}} \rangle$	0.025 \pm 0.0026 Å	0.028 \pm 0.0029 Å	0.026 \pm 0.0033 Å
$\langle \Delta_{\text{dihedral}} \rangle$	0.17° \pm 0.035°	0.23° \pm 0.061°	0.29° \pm 0.050°
$\langle \Delta_{\text{bonds}} \rangle$	0.0025 \pm 0.00036 Å	0.0029 \pm 0.00036 Å	0.0026 \pm 0.00041 Å
$\langle \Delta_{\text{angles}} \rangle$	0.54° \pm 0.012°	0.53° \pm 0.014°	0.54° \pm 0.014°
$\langle E_{\text{vdw}} \rangle$	-112.9 kcal mol ⁻¹	32.3 kcal mol ⁻¹	17.2 kcal mol ⁻¹
Villin 14T			
$\langle \Delta_{\text{NOE}} \rangle$	0.025 \pm 0.0047 Å	0.029 \pm 0.0041 Å	0.025 \pm 0.0023 Å
$\langle \Delta_{\text{dihedral}} \rangle$	0.50° \pm 0.070°	0.52° \pm 0.080°	0.61° \pm 0.093°
$\langle \Delta_{\text{bonds}} \rangle$	0.0027 \pm 0.00044 Å	0.0024 \pm 0.00055 Å	0.0025 \pm 0.00017 Å
$\langle \Delta_{\text{angles}} \rangle$	0.48° \pm 0.033°	0.56° \pm 0.039°	0.45° \pm 0.029°
$\langle E_{\text{vdw}} \rangle$	-75.19 kcal mol ⁻¹	191.7 kcal mol ⁻¹	175.6 kcal mol ⁻¹

Note. $\langle \Delta_{\text{NOE}} \rangle$, average deviation of NOE-derived distances from target values; $\langle \Delta_{\text{dihedral}} \rangle$, average deviation of restrained dihedral angles from target values; $\langle \Delta_{\text{bonds}} \rangle$, average deviation of bond lengths from ideal values; $\langle \Delta_{\text{angles}} \rangle$, average deviation of bond angles from ideal values; $\langle E_{\text{vdw}} \rangle$, average van der Waals energy using Eq. [4]. Means and standard deviations are computed for the specified ensembles.

TABLE 6
Success Rate and Computational Efficiency

	BPTI	Protein G	Interleukin 8 ^a	Villin 14T
Torsion-angle molecular dynamics				
Success rate	98.0%	87.7%	89.3%	84.7%
Structure calculation	921 s	959 s	2992 s	1942 s
Computational efficiency	940 s	1093 s	3352 s	2892 s
SA				
Success rate	78.1%	78.1%	69.4%	36.8%
Structure calculation	523 s	563 s	5597 s	4612 s
Computational efficiency	670 s	720 s	8060 s	12546 s
DGSA				
Success rate	64.0%	100.0%	24.3%	32.5%
Structure calculation	557 s	558 s	1631 s	1351 s
Computational efficiency	870 s	558 s	6690 s	4164 s

Note. The success rate, the computer time required to generate a single structure ("struc. calc"), and computational efficiency are shown for each protocol and test case. All computations were carried out on a Hewlett-Packard 735 computer.

^a In order to get any accepted structures for interleukin-8 or villin 14T from either the DGSA or the SA protocol, it was necessary to quadruple the molecular dynamics period.

TABLE 7
Results for the DNA Dodecamer Test Case

	Torsion-angle molecular dynamics	Cartesian molecular dynamics (SA)	Distance geometry with Cartesian molecular dynamics (DGSA)	Original structure (29)	Rerefined original structure
Success rate	52.0%	0.0%	0.0%		
$\langle N_{\text{NOE}} \rangle$	0.0	0.9	19.5	0.0	0.0
$\langle \Delta_{\text{NOE}} \rangle$	0.050 Å	0.089 Å	0.26 Å	0.054 Å	0.051 Å
$\langle \Delta_{\text{dihedral}} \rangle$	0.45°	11.89°	14.52°	1.68°	0.54°
$\langle \Delta_{\text{bonds}} \rangle$	0.012 Å	0.019 Å	0.021 Å	0.0086 Å	0.012 Å
$\langle \Delta_{\text{angles}} \rangle$	0.99°	1.80°	1.82°	5.29°	1.00°
$\langle E_{\text{vdw}} \rangle$	-269.2 kcal mol ⁻¹	7820 kcal mol ⁻¹	458.5 kcal mol ⁻¹	-352.3 kcal mol ⁻¹	-253.4 kcal mol ⁻¹
RMS deviation from original structure (29)	2.66 ± 0.27 Å	5.23 ± 0.61 Å	6.23 ± 0.56 Å	0	2.26 ± 0.04 Å

Note. No acceptable structures were generated by the SA or DGSA algorithms so the best ten of the first fifty generated structures are shown. $\langle N_{\text{NOE}} \rangle$, average number of NOE distances further than 0.5 Å outside of the distance bounds; $\langle \Delta_{\text{NOE}} \rangle$, average deviation of NOE-derived distance from target values; $\langle \Delta_{\text{dihedral}} \rangle$, average deviation of dihedral angle from target values; $\langle \Delta_{\text{bonds}} \rangle$, average deviation of bond lengths from ideal values; $\langle \Delta_{\text{angles}} \rangle$, average deviation of bond angles from ideal values; $\langle E_{\text{vdw}} \rangle$, average van der Waals energy using Eq. [4].

The structure-calculation protocol was repeated with different initial velocities drawn from a random Maxwellian distribution in order to obtain an ensemble of structures. Acceptance of the resulting structure was checked using the criterion described below.

The parameters that most affected the protocol are the temperature and the duration of the torsion-angle dynamics stages. Temperatures greater than 50,000 K accelerated the convergence (thereby increasing computational efficiency) for large molecules but lowered the success rate for smaller molecules (data not shown). A temperature of 50,000 K appeared to be a good compromise for protein structures with 10 to 15 NOE restraints per residue, as shown by trial calculations with BPTI and protein G. In our experience, the protocol is sufficient to refine protein molecules ranging from 1000 to 3000 atoms with approximately 10–15 NOE restraints per residue. For structures comprising more than 3000 atoms, it may be necessary to increase the length of the torsion-angle molecular dynamics stages.

Minor changes in the protocol were necessary for refining nucleic acid structures due to vibrations in nonrigid ribose rings. The simulation temperature had to be reduced to 20,000 K for the torsion-angle molecular dynamics stages, the coefficient for the dihedral-restraints energy term (E_{cdih}) had to be reduced from 100 to 5 during both stages of torsion-angle dynamics, and the length of both torsion-angle molecular dynamics stages had to be tripled in order to obtain acceptable structures.

A major advantage of the torsion-angle molecular dynamics method is its simplicity. It consists of only four stages with two parameters changing: w_{vdw} and w_{dihedral} (Eqs. [2] and [3], and Table 2). In contrast, during the five stages of the SA method, w_{vdw} , w_{dihedral} , the weights for bond angle

and improper energy terms, the van der Waals radii, and the asymptote for the NOE-derived distance function (Eq. [6]) are frequently changed. The DGSA algorithm, which consists of eight stages of molecular dynamics and conjugate-gradient minimization, has an even larger number of changing parameters.

Comparisons

The torsion-angle molecular dynamics algorithm was compared to a Cartesian-molecular-dynamics-based simulated annealing method starting from an extended strand (referred to as SA) (6) and a protocol which uses metric-matrix distance geometry combined with Cartesian molecular dynamics (referred to as DGSA) (2). The two algorithms are implemented in X-PLOR, version 3.1 (5) (files SA.INP and DGSA.INP), and were not modified except to extend the Cartesian molecular dynamics stage for interleukin-8 and villin 14T by a factor of 4 in order to obtain a reasonable acceptance rate (not shown).

Acceptance Criterion

The three algorithms (torsion-angle molecular dynamics, SA, and DGSA) were repeated with different initial velocities until they each produced 50 acceptable structures, where an acceptable structure is defined as one that contains no violations of NOE restraints greater than 0.5 Å and no dihedral angle violations greater than 5°. Structures were also rejected if the root-mean-square (RMS) deviation of bonds from ideal values was greater than 0.02 Å, or the RMS deviation of angles was greater than 2.0°. Success rate is defined as the ratio of the number of accepted structures to the total number of trials. Computational efficiency is de-



fined as the average computing time in order to obtain one acceptable structure.

Computer Program

All calculations were carried out with X-PLOR (on-line) (5), which is available over the Internet (URL <http://xplor.csb.yale.edu>).

RESULTS AND DISCUSSION

Dials plots (30) were used to compare the ability to sample conformational space for structure calculation using torsion-angle molecular dynamics, Cartesian-molecular-dynamics-based simulated annealing, and its combination with distance geometry. Torsion angles are visualized by rotation around a circle in which radial displacement is proportional to the molecular dynamics simulation time t (Eq. [9]). Figure 3 shows that the dial plots corresponding to the torsion-angle molecular dynamics structure calculations are significantly better sampled than those corresponding to the other protocols.

Table 3 shows that the RMS difference from the average structure is approximately (within a standard deviation) the same for all three structure-calculation methods. The RMS differences between the average structures generated by each method are similar for all pairwise combinations, and they are within the RMS differences from the corresponding average structures (Table 4). The ensembles generated by the three structure-calculation methods satisfy both the experimental data and chemical restraints to the same degree except that the van der Waals energies are lower for the torsion-angle molecular dynamics structures (Table 5).

The success rate and computational efficiency of torsion-angle molecular dynamics is higher than that of the other two methods for larger proteins (Table 6). For both interleukin-8 and villin 14T, SA takes longer than torsion-angle molecular dynamics to generate a single structure (which may or may not be acceptable) and an average of two to four times longer to generate an acceptable one. DGSA is still faster than torsion-angle dynamics at generating a single structure but takes about twice as long to generate an acceptable one (Table 6).

The results of the structure calculation of the DNA dodecamer are presented in Table 7. Since no acceptable structures were generated by either the DGSA or SA algorithm, statistics were generated by choosing the 10 best structures (according to NOE and dihedral angle violations) of the first 50 generated. In contrast, torsion-angle molecular dynamics

produced acceptable structures in 52% of the trials. It should be noted that the original structure (29) was obtained by restrained molecular dynamics refinement, starting from canonical A- and B-form DNA.

Average structures are shown in Fig. 4. The ensemble generated by torsion-angle molecular dynamics agrees most closely with the original structure (29) (RMS deviation = 2.67 Å) while the helices generated by SA and DGSA deviate significantly from the original structure and have large NOE violations (Table 7). The SA and DGSA protocols have paired the bases correctly, but the proper helicity was not achieved. The original structure apparently does not satisfy the statistical criteria to the same degree as the structures generated by torsion-angle molecular dynamics (Table 7). These differences are artificial: they are a consequence of using a different energy function in the original structure calculation. When the original structure is subjected to a brief (6 ps) Cartesian molecular dynamics refinement and conjugate gradient minimization (1000 steps) against the same energy function used in this paper, the ensemble of structures moves away from the original structure (atomic RMS difference = 2.26 Å) and moves toward the torsion-angle molecular dynamics generated structure (atomic RMS difference = 1.25 Å), and the statistical quantities become very similar to the torsion-molecular dynamics method (Table 7).

Torsion-angle molecular dynamics was able to successfully fold an extended strand of DNA into B-DNA formation with the correct helicity while other methods failed. It is remarkable that B-form DNA is achieved without the imposition of additional restraints and without starting from either A- or B-form DNA.

CONCLUSIONS

Molecular dynamics constrained to torsion angles provides a powerful tool for structure calculation with NMR data. The method has a higher success rate and efficiency than conventional simulated annealing algorithms which use Cartesian molecular dynamics or distance geometry combined with Cartesian molecular dynamics. A significant difference between the computing time required for torsion-angle molecular dynamics and other methods can be seen with proteins larger than 100 residues (Table 6). Furthermore, torsion-angle molecular dynamics is capable of folding extended DNA strands into B-form DNA. As structures analyzed by NMR increase in size, we expect that the advantage of torsion-angle molecular dynamics will become increasingly more important.

ACKNOWLEDGMENTS

We thank G. Warren, A. Bonvin, P. Adams, T. Burling, and W. DeLano for valuable discussions, and G. M. Clore for providing the NMR data of interleukin 8 and the DNA dodecamer, and M. Nilges for the BPTI test case. We also thank M. A. Markus and G. Wagner for providing the data on villin 14T. This work was supported by a grant of the National Science Foundation to A.T.B. (ASC93181159). L.M.R. is an HHMI predoctoral fellow.

REFERENCES

1. W. J. Metzler, D. R. Hare, and A. Pardi, *Biochemistry* **28**, 7045–7052 (1989).
2. M. Nilges, G. M. Clore, and A. M. Gronenborn, *FEBS Lett.* **229**, 317–324 (1988).
3. A. T. Brünger, G. M. Clore, A. M. Gronenborn, and M. Karplus, *Proc. Nat. Acad. Sci.* **83**, 3801–3805 (1986).
4. R. Kaptein, E. R. P. Zuiderweg, R. M. Scheek, R. Boelens, and W. F. van Gunsteren, *J. Molec. Biol.* **182**, 179–182 (1985).
5. A. T. Brünger, "X-PLOR, Version 3.1. A System for X-Ray Crystallography and NMR," Yale Univ. Press, New Haven, 1992.
6. M. Nilges, J. Kuszewski, and A. T. Brünger, in "Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy" (J. C. Hoch, F. M. Poulsen, and C. Redfield, Eds.), pp. 451–455, Plenum Press, New York, 1991.
7. R. Diamond, *Acta Crystallogr. Sect. A* **27**, 436–452 (1971).
8. J. Curri, *J. Chem. Phys.* **61**, 1204–1207 (1974).
9. W. Braun and N. Gö, *J. Molec. Bio.* **186**, 611–626 (1985).
10. N. B. Ulyanov, U. Schmitz, and T. L. James, *J. Bio. NMR* **3**, 547–568 (1993).
11. Y. Xu and N. R. Krishna, *J. Magn. Reson. B* **108**, 192–196 (1995).
12. D. S. Bae and E. J. Haug, *Mech. Struct. Mach.* **15**, 359–382 (1987).
13. D. S. Bae and E. J. Haug, *Mech. Struct. Mach.* **15**, 481–506 (1988).
14. A. Jain, N. Vaidehi, and G. Rodriguez, *J. Comput. Phys.* **106**, 258–268 (1993).
15. L. M. Rice and A. T. Brünger, *Proteins* **19**, 277–290 (1994).
16. A. M. Mathiowetz, A. Jain, N. Karasawa, and W. A. Goddard, *Proteins* **20**, 227–247 (1994).
17. G. M. Crippen and T. F. Havel, "Distance Geometry and Molecular Conformation," Wiley, New York, 1988.
18. A. T. Brünger and M. Nilges, *Q. Rev. Biophys.* **26**, 49–125 (1993).
19. J. Kuszewski, M. Nilges, and A. T. Brünger, *J. Biomol. NMR* **2**, 33–56 (1992).
20. D. S. Garrett, J. Kuszewski, T. J. Hancock, P. J. Lodi, G. W. Vuister, A. M. Gronenborn, and G. M. Clore, *J. Magn. Reson. B* **104**, 99–103 (1994).
21. J. Kuszewski, A. M. Gronenborn, and G. M. Clore, *J. Magn. Reson. B* **107**, 293–297 (1995).
22. M. Nilges, A. M. Gronenborn, A. T. Brünger, and G. M. Clore, *Protein Eng.* **2**, 27–38 (1988).
23. A. T. Brünger and L. M. Rice, in "Data Handling in Science and Technology" (J. H. Kalivas, Ed.), Vol. 15, p. 259–280, Elsevier, New York, 1995.
24. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, *J. Chem. Phys.* **81**, 3684–3690 (1984).
25. A. M. Gronenborn, D. R. Filpula, N. Z. Essig, A. Achari, M. Whitlow, P. T. Wingfield, and G. M. Clore, *Science* **253**, 657–661 (1988).
26. T. Schneider, and M. Nilges, personal communication.
27. G. M. Clore, E. Appella, E. Yamada, K. Matsushima, and A. M. Gronenborn, *Biochemistry* **29**, 1689–1696 (1990).
28. M. A. Markus, T. Nakayama, P. Matsudira, and G. Wagner, *Protein Sci.* **3**, 70–81 (1994).
29. G. M. Clore, H. Oschkinat, L. W. McLaughlin, F. Bensele, C. Scalfi-Happ, E. Happ, and A. M. Gronenborn, *Biochemistry* **27**, 4185–4197 (1988).
30. S. Swaminathan, G. Ravishanker, D. Beveridge, R. Lavery, C. Etchebest, and H. Sklenar, *Proteins* **8**, 179–193 (1990).