PROTEINS: Structure, Function, and Genetics 3:209–218 (1988)

Combined Procedure of Distance Geometry and Restrained Molecular Dynamics Techniques for Protein Structure Determination From Nuclear Magnetic Resonance Data: Application to the DNA Binding Domain of Lac Repressor From *Escherichia coli*

J. de Vlieg,[1] R.M. Scheek,[1] W.F. van Gunsteren,[1] H.J.C. Berendsen,[1] R. Kaptein,[2] and J. Thomason[3]
[1]*Laboratory of Physical Chemistry, University of Groningen, 9747 AG Groningen and* [2]*Department of Organic Chemistry, University of Utrecht, 3584 CH Utrecht, The Netherlands;* [3]*Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, California 94143*

ABSTRACT **The technique of two-dimensional nuclear magnetic resonance (2D-NMR) has recently assumed an active role in obtaining information on structures of polypeptides, small proteins, sugars, and DNA fragments in solution. In order to generate spatial structures from the atom–atom distance information obtained by the NMR method, different procedures have been developed. Here we introduce a combined procedure of distance geometry (DG) and molecular dynamics (MD) calculations for generating 3D structures that are consistent with the NMR data set and have reasonable internal energies. We report the application of the combined procedure on the lac repressor DNA binding domain (headpiece) using a set of 169 NOE and 17 "hydrogen bond" distance constraints. Eight of ten structures generated by the distance geometry algorithm were refined within 10 ps MD simulation time to structures with low internal energies that satisfied the distance constraints.**

Although the combination of DG and MD was designed to combine the good sampling properties of the DG algorithm with an efficient method of lowering the internal energy of the molecule, we found that the MD algorithm contributes significantly to the sampling as well.

Key words: distance-restrained molecular dynamics, 2D NOE-spectroscopy, tertiary structure, solution conformations

INTRODUCTION

In recent years methods have been developed to determine the 3-dimensional (3D) structure of proteins by nuclear magnetic resonance (NMR).[1–4] The NMR method has the advantage of yielding a 3D structure in solution and avoiding the crystallization process. The latter is still the bottleneck in protein structure determination by X-ray crystallography. In many cases it is very difficult, if not impossible, to grow protein crystals. An example is the *lac* repressor protein of *Escherichia coli* for which no crystals of sufficient quality for an X-ray diffraction study have been obtained so far.

For a number of polypeptides, for example glucagon (29 residues),[5] proteinase inhibitor IIA from bull seminal plasma (57 residues),[6] phoratoxin (46 residues),[7] hirudin (65 residues)[8] and the globular domain of chicken histone H5 (79 residues),[9] the 3D structure has been derived by 2D NMR methods. Once the proton resonances are assigned, the nuclear Overhauser enhancements (NOE) between proton pairs can be identified and translated into proton–proton distances, which can be used to construct a 3D model structure. This can be done by a combination of different techniques, such as model building[10] on a graphics display system, distance geometry calculations,[11] and distance-restrained molecular dynamics.[12]

For *lac* repressor headpiece a 3D solution structure has been obtained by a combination of model building and restrained molecular dynamics.[10] Starting from a crude initial hand-made model, a solution structure was derived by an iterative process of manual adjustments on a graphics display system and restrained molecular dynamics calculations. In this way it was possible to obtain a solution structure that satisfied all NMR data and had a low internal energy.

However, the process of graphics-display sessions become cumbersome when large molecules with many distance constraints are considered, and a certain amount of subjectivity will probably be introduced during the model-building step. Also the configurational space accessible to the molecule under the distance constraints will certainly not be searched completely in this way of structure determination. Therefore, we introduce here an alternative procedure to generate 3D-structures in which the model building step is replaced by a distance geometry cal-

Received February 12, 1988; accepted March 18, 1988.

Address reprint requests to J. de Vlieg, Laboratory of Physical Chemistry, University of Groningen, Nijenborgh 16, 9747 AG Gronigen, The Netherlands.

culation (DG).[11] Since the DG methods start from randomly selected distances without requiring a predefined model structure, the configurational space can be searched by generating a cluster of initial structures, all satisfying the distance constraints, by repeated distance geometry calculations.

The DG structures are generally highly strained and need further energy refinement. Therefore we used a refinement procedure consisting of restrained energy minimization (EM) and molecular dynamics (MD) techniques to search for low-energy structures within the allowed range of violations. In order to reduce the computing effort during the energy refinement procedure, the MD simulation time was limited to 10 ps. In this way the good sampling properties of the DG algorithm are combined with an efficient method of energy refinement.

The combined DG and MD procedure was applied to the *lac* repressor headpiece using a set of 169 NOE distance and 17 "hydrogen bond" distance constraints. In order to investigate the merits of the combined procedure we generated in total ten initial headpiece structures by repeated distance geometry calculations. These initial structures were subsequently refined. In this paper the following points will be addressed:

1. A detailed comparison of the refined and unrefined structures is made to test the sampling properties and the ability of the EM/MD refinement procedure to transform highly strained DG structures into energetically acceptable structures.
2. The structures are compared to our previous solution structure of *lac* repressor headpiece, which was obtained by a combination of model-building techniques and restrained molecular dynamics calculations[10] in order to estimate the degree of convergence during the molecular dynamics simulations.
3. The configurational space searched by DG and the combined use of DG and MD is compared to the configurational space searched during molecular dynamics alone.

COMPUTATIONAL METHODS

Translation of Experimental NOEs to Distance Constraints

For the headpiece, resonances of all backbone amide and Cα protons (except those of Ile 48) and of the majority of the side chain protons have been assigned.[13] Analysis of the 50 ms mixing time 2D NOE spectra of the headpiece resulted in a list of 169 NOEs.[14]

The 2D NOE spectra contain information on the secondary structure of the protein because different patterns of short interproton distances prevail for different secondary structural elements. For the headpiece three α- helices were found; residues 6–13, 17–25, and 34–45.[15] In the α-helical regions, cross peaks between Cα protons of residue i and amide protons of both residue i + 1 and i + 3 were just visible. In a regular α-helix, the distances between these protons are approximately 0.35 nm. Therefore we have taken this value as a basic upper limit of the NOE distance constraints for all 169 proton-proton pairs that give rise to cross peaks in the 2D NOE spectrum with a mixing time of 50 ms. In order to facilitate the preservation of the α-helices, 17 attractive constraints (r_0 = 0.25 nm) between the amide proton of residue i and the carbonyl oxygen of residue i-4, with i = 10–13, 21–25, 38–45, were added to the 169 attractive NOE constraints.

For some proton-proton pairs, corrections have to be applied to the basic upper limit distance constraint of 0.35 nm. This may arise in cases where protons cannot be distinguished in the NMR spectra, either because of the lack of stereospecific assignments or because of rapid dynamic processes. In these cases the distance constraint must be applied to the average position (pseudoatom) and the upper limit bound is corrected for the difference between the proton position and the actually used pseudoposition.[16] A full description of the concept of the use of pseudoatoms in GROMOS[17] is given in reference 18.

The Distance Geometry Algorithm

The distance geometry algorithm used was developed from the EMBED program[11] and will be described in more detail elsewhere. It was implemented on a CRAY 2 computer at the Boeing supercomputer facility in Seattle. In the first phase of the program, upper and lower bounds on interatomic distances are calculated for all pairs of atoms separated by less than two variable dihedral angles. Omega dihedral angles in proteins are fixed to their trans values to ensure planar peptide bonds. For pairs of atoms separated by one variable dihedral angle, upper and lower bonds are calculated corresponding to the trans and cis conformations, respectively, of the atom pair with regard to the rotatable bond. If no dihedral angles separate a pair of atoms upper and lower bounds are set equal to the interatomic distance.

The upper and lower bounds thus generated are augmented by the distance constraints that follow from the 169 observed NOE and 17 "hydrogen bond" distance constraints. For all remaining atom pairs, lower bounds are taken to be equal to the sum of the atomic radii. To allow for hydrogen bond formation between atoms of the proper type, the lower bound for such pairs are lowered by 0.1 nm. The DG algorithm then proceeds with an iterative bound smoothing procedure. For all possible triplets containing a certain atom pair, the upper bound on the distance between the atoms of that pair is lowered and the lower bound is raised to the triangulation limits imposed on it by the bounds on the two other distances in that triplet. Distance matrices are then randomly selected between the upper and lower bound matrices and subsequently embedded in three-dimensional

space according to the procedure described elsewhere.[11,19] Since in the embedded structures many distances usually fall outside their bounds, a conjugate gradient optimization is needed to minimize the remaining atom–atom distance violations and to impose the proper chirality to the various asymmetric centers in the molecule.

Structure Refinement by EM and MD Using Distance Constraints

In the method of restrained EM and MD refinement, an arbitrary initial structure is refined by minimizing the violations of the distance constraints and the potential energy function of the molecule. By application of EM, one searches for a minimum energy conformation by moving along the gradient:

$$\Delta \mathbf{r}_i \sim -\partial/\partial \mathbf{r}_i V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \qquad (1)$$

through configurational space[20] where $\mathbf{r}_i$ are the cartesian position vectors of N atoms with masses m_i (i = 1, 2, . . ., N). Since in this way one moves basically only downhill over the energy hypersurface, EM yields only a local minimum energy conformation, which is generally not far away from the initial one.

In an MD simulation a trajectory (conformations as a function of time) of the system is generated by simultaneous integration of Newton's equations of motion:

$$\begin{aligned} d^2\mathbf{r}_i/dt^2 &= m_i^{-1}\,\mathbf{F}_i \\ \mathbf{F}_i &= -\partial/\partial \mathbf{r}_i\, V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \end{aligned} \qquad (2)$$

for all the atoms in the system. Here, the force on atom i is denoted by $\mathbf{F}_i$. In MD the kinetic energy makes it possible to cross over energy barriers of the order of kT. This feature makes MD a more efficient technique to generate low-energy conformations than EM techniques.[10,21]

Molecular Model and Force Field

The potential energy function consists of the usual terms representing bond angle bending, harmonic (out-of-plane, out-of-tetrahedral configuration) dihedral bending, sinusoidal dihedral torsion, van der Waals, and electrostatic (Coulomb) interactions.[22,23] All atoms are treated explicitly except for the hydrogens bound to carbon atoms, for which united (CH_1, CH_2, CH_3) atoms are used. The covalent bond lengths are kept fixed by applying the SHAKE method.[24,25] Since no solvent molecules are considered, we have neutralized the charged groups in order to compensate for the missing shielding effect of the solvent.[18] The values of the parameters of the potential energy function can be found elsewhere.[17]

In order to obtain 3D structures that satisfy the set of distance constraints, an extra term has been added to the potential function:

$$\begin{aligned} &V_{dc}(r_{nm}, r^0{}_l) \\ &= 0 && \text{if } (0 \leqslant r_{nm} \leqslant r^0{}_l) \\ &= \tfrac{1}{2}\, K_{dc}(r_{nm} - r^0{}_l)^2 && \text{if } (r^0{}_l \leqslant r_{nm} \leqslant r^1{}_l) \\ &= \tfrac{1}{2}\, K_{dc}(2r_{nm} - r^0{}_l - r^1{}_l)(r^1{}_l - r^0{}_l) && \text{if } (r_{nm} \geqslant r^1{}_l) \end{aligned} \qquad (3)$$

where $r^0{}_l$ is the value for the basic upper limit constraint plus the correction terms for pseudo atoms,[18] and r_{nm} is the actual distance between atoms n and m involved in the specific constraint. The potential energy terms V_{dc} (r) is taken linear beyond $r^1{}_l$ in order to avoid the occurrence of very large constraining forces. In this way a conformational change from one local minimum to another one, which are separated by a conformational path with large violations of the distance constraints is allowed. Here a value of $r^1{}_l = r^0{}_l + 0.1$ nm is used. The force constant k_{dc} can be gradually changed during the refinement so that the flexibility of the molecule can be adapted in the different stages of refinement.

Computational Procedure

The distance geometry technique was repeatedly applied to the *lac* repressor headpiece using the set of 169 NOE and 17 "hydrogen bond" distance constraints and resulted in ten structures.

Before applying MD, the strain that is present in structures generated using distance geometry must be relaxed by EM. This was done in two consecutive steps. First, the structures were subjected to 100 steepest descent EM steps in which the bond vibrations were explicitly treated. Second, 200 bond-length constrained conjugate gradient steps were performed.[20] The first step was necessary to prevent convergence problems in the SHAKE algorithm[24,25] caused by the high strain occurring in the DG structures. In both steps a distance constraint force constant of $K_{dc} = 500\ kJ \cdot mol^{-1} \cdot nm^{-2}$ is used.

When starting an MD simulation, the initial velocities are taken from a Maxwellian distribution at the desired temperature (usually 300 K), and the translational and rotational motion of the molecule is halted. The MD time step was $\Delta t = 0.002$ ps, and the neighbor list of nonbonded atom pairs was updated every 10 MD steps, using a cut-off radius of 0.8 nm.

Furthermore, the molecule was weakly coupled to a thermal bath with reference temperature T_0 using a coupling constant of $\tau = 0.1$ ps.[17] In order to reduce computing effort we limited the MD refinement time to 10 ps. During the first 5 ps a distance constraint force constant of $K_{dc} = 500\ kJ \cdot mol^{-1} \cdot nm^{-2}$ was used. Subsequently the MD refinement was continued for 5 ps during which the force constant was increased gradually at a rate of $750\ kJ \cdot mol^{-1} \cdot nm^{-2} \cdot ps^{-1}$ to a final value of $4{,}000\ kJ \cdot mol^{-1} \cdot nm^{-2}$. Finally the kinetic energy was removed from the system by another cycle of conjugate gradient energy minimization using the same high value for the force constant.

TABLE I. Average Results of the Refinement Procedure*

	Distance geometry	After EM refinement	After EM+MD refinement
E_{pot}	12,468 (810)	−1,351 (145)	−2,806 (89)
E_{bond}	7,880 (110)	—	—
E_{angle}	355 (79)	470 (29)	346 (24)
$E_{im.dih}$	336 (48)	165 (12)	99 (11)
E_{dih}	646 (111)	492 (40)	286 (29)
E_{el}	−1,007 (142)	−1,502 (89)	−1,995 (51)
E_{lj}	4,269 (774)	−976 (45)	−1,550 (62)
Σ_{dviol}	0.84 (0.13)	1.74 (0.30)	0.77 (0.24)
RMSD	0.11 (0.01)	0.12 (0.01)	0.19 (0.06)

*The mean potential energy contributions, total distance constraints violations and rms differences averaged over ten headpiece structures obtained by distance geometry, refinement of these structures using EM, and refinement using EM+MD calculations. The energy terms for bond-length, bond-angle bending, harmonic dihedral bending, sinusoidal dihedral torsion, electrostatic and van der Waals interactions are denoted by E_{bond}, E_{angle}, $E_{im.dih}$, E_{dih}, E_{el}, and E_{lj}, respectively, and are given in kJ.mol^{-1}. The sum over these terms is denoted by the symbol E_{pot}. During EM and MD calculations the E_{bond} term has no meaning because bond-lengths are kept rigid by applying the SHAKE method.[20,21] The total distance constraint violations, i.e., the total sum of the violations of the 169 NOE and 17 "hydrogen bond" distance constraints is denoted by the symbol Σ_{dviol} and given in nm. The symbol RMSD denotes the average root mean square differences between the ten structures and is given in nm. For fitting and comparison of the structures only the C_α atom positions of residues 4–47 are used. In parentheses the standard deviation of the average values are given.

TABLE II. Individual Results of Refinements of Ten Distance Geometry Structures of the *lac* Repressor Headpiece*

	Initial		After EM		After EM+MD	
Structure	Epot	Σ_{dviol}	Epot	Σ_{dviol}	Epot	Σ_{dviol}
S1	13,937	0.72	−1,393	1.46	−2,686	1.13
S2	12,457	0.85	−1,269	1.88	−2,798	0.86
S3	11,955	0.95	−1,406	2.32	−2,776	0.81
S4	13,230	0.76	−1,199	2.07	−2,716	0.63
S5	11,404	0.88	−1,577	1.49	−2,927	0.37
S6	11,418	0.78	−1,545	1.28	−2,926	0.55
S7	12,442	0.59	−1,196	1.57	−2,671	1.07
S8	12,560	1.03	−1,098	1.95	−2,847	0.67
S9	11,551	1.05	−1,301	1.78	−2,819	1.07
S10	13,707	0.82	−1,379	1.61	−2,891	0.58

*The total potential energy and total distance constraint violations for ten distance geometry-generated *lac* repressor headpiece structures (denoted by the symbols S1, S2, . . . , S10) before and after EM and EM+MD refinement. The various quantities are defined in the footnote to Table I.

RESULTS AND DISCUSSION

In view of testing the quality of our simple refinement procedure and the sampling behavior of distance geometry and molecular dynamics calculations, the structural quantities of ten *lac* repressor headpiece structures generated by repeated distance geometry calculations (DG) before and after EM/MD refinement were compared.

We also compared the refined and unrefined structures to our previous solution structure, which was obtained by a combination of model building and restrained EM/MD refinement.[10]

Average Results of the Refinement Procedure

The mean values of the potential energy contributions, total violation of the distance constraints and total root mean square positional differences averaged over ten headpiece structures obtained by 1) distance geometry calculations and 2) combined use of DG and EM/MD refinement are shown in Table I.

The DG structures have on the average a very low total violation of the distance constraints (0.84 nm) but an extremely high potential energy (12,468 kJ·mol^{-1}) because of a large bond energy term (7,880

$kJ \cdot mol^{-1}$) and Lennard-Jones interaction (4,269 $kJ \cdot mol^{-1}$). The high value of the potential energy is above all caused by the use of slightly smaller standard van der Waals radii during the DG calculations than those used by the GROMOS[17] force field. By restrained energy minimization (EM) the potential energy is brought down considerably. However, the mean value of the total violation is increased during the energy minimization by a factor 2 from 0.84 to 1.74 nm. This is apparently due to the fact that the distance constraint energy term and the physical force field have opposite effects. This is not very surprising because the DG structures are only determined by intermolecular distances but not by other internal energy terms.

During the MD refinement the mean potential energy drops strongly from $-1{,}351$ $kJ \cdot mol^{-1}$ to $-2{,}806$ $kJ \cdot mol^{-1}$, and only minimal violations on the distance constraints remain in the final structures. The drop in potential energy is caused by a decrease of all energy terms, of which the Lennard-Jones and electrostatic interaction terms are most important. The mean value of the total distance constraint violation of the MD refined structures is even lower than the value found for the DG structures. This observation reflects the power of dynamic modeling. Generally, MD searches a larger part of configurational space than EM, since the available kinetic energy allows the molecule to cross over barriers of the order of kT.

Differences Between Individual Refinements

In Table II, the constraint violations and potential energies of the structures during the ten individual refinements are shown. It is observed, as expected, that for all ten EM refined structures the total violation of the distance constraints is larger than those found for the initial DG structures. For six of the MD refined structures, however, the values of the total violation are lower; for two they are approximately equal (S2, S9), and for two (S1, S7) they are higher than for the DG structures.

In all cases the potential energies of MD refined structures are much lower than for those refined only with EM. In two cases (structure S1 and S7) the MD algorithm was not able to generate structures with a lower or approximately equal value for the total violation as found for the initial DG structure. Also the potential energies of structures S1 and S7 are among the highest of the refined structures ($-2{,}686$ and $-2{,}671$ $kJ \cdot mol^{-1}$). Apparently, for these two the initial distance geometry structures are too strained to be refined to reasonable structures with low energies while satisfying the set of distance constraints. However, by using a more elaborate refinement procedure, large improvements can be obtained as will be shown below. We note here that structures S1 and S7 have the lowest total distance constraint violations among the distance geometry structures. This shows that the best DG structure is not necessarily the best one when the internal energy is taken into account.

In Figure 1 the refinement is followed by displaying the number of NOE distance constraint violations as a function of the size of the violations. We chose the refinement of structure 4 as a typical example. The number and the size of the violations occurring in structures generated by the distance geometry algorithm is normally small (Fig. 1a). However, the number and the size of the violations increase dramatically by performing EM refinement (Fig. 1b). During the EM refinement of structure 3 the maximum violation became even as large as 3.2 nm! In spite of the large number and size of violations occuring in EM structures, only small violations remain after MD refinement (Fig. 1c).

RMS Positional Differences

Before comparing two structures, they were fitted to each other by superimposing the centers of mass and by subsequently performing a rotational least-squares fit around the common centers of mass. Because the positions of the N- and C-termini are not or only weakly determined by NMR data, only the C_α atom positions of residues 4 to 47 are used for fitting and comparison. In Table III, the RMS positional differences among refined and unrefined distance geometry structures are compared. The spread between the RMS positional differences among DG structures is low (Table III, upper righthand triangle). The standard deviation of the average RMS differences of 0.11 nm is only 0.01 nm. After the MD refinement the spread is increased strongly as can be observed in the lower lefthand triangle of Table III. In most cases the difference between two structures increased after refinement, but, in particular, the increases for structures 1 and 7 are extremely large. The RMS difference between structures 6 and 8 before and after refinement decreased significantly from 0.12 nm to 0.08 nm.

In Table IV the RMS positional shifts during EM and MD refinements are compared. Energy minimization does not change the conformation substantially (shifts range from 0.02 to 0.04 nm) and apparently relaxes the strain in the molecule by small local positional adjustments. When MD is applied, the structures change considerably, which leads to a large decrease of the internal energy.

Comparisons With a Model-Built Structure

In Table V we compared the refined and unrefined structures with our previous solution structure.[10] This structure was obtained by a combination of model building and restrained EM/MD calculations. The potential energy of this structure is $-3{,}091$ $kJ \cdot mol^{-1}$ and the total violation is 0.40 nm for the 169 NOE and 17 "hydrogen bond" constraints used for the generation and refinement of the DG structures. Apart from these distance constraints, non-NOE data and information on J-coupling were also included during the refinement of the model- built structure. We note that the internal energy of the refined model-built

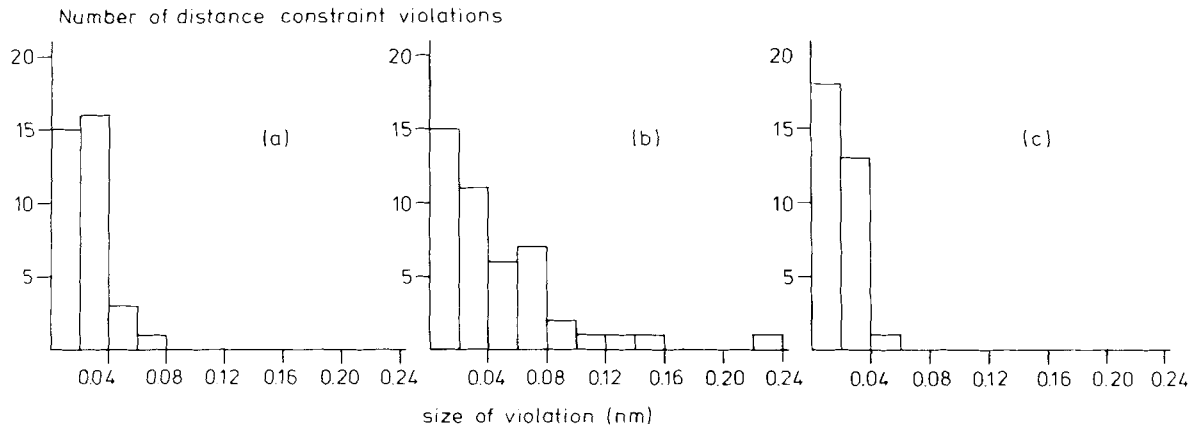

Fig. 1. Number of distance constraint violations as a function of the size (in nm) of the violations in the course of a typical refinement (S4). **a:** Initial distance geometry structure. **b:** Structure after EM. **c:** Structure after 10 ps MD and EM.

TABLE III. RMS Differences Between Various *Lac* Repressor Headpiece Structures*

Structure	1	2	3	4	5	6	7	8	9	10
1		0.12	0.10	0.10	0.10	0.12	0.10	0.12	0.13	0.12
2	0.26		0.11	0.13	0.13	0.13	0.12	0.12	0.11	0.14
3	0.26	0.20		0.09	0.11	0.10	0.09	0.11	0.11	0.12
4	0.24	0.15	0.18		0.10	0.11	0.11	0.12	0.13	0.12
5	0.24	0.15	0.16	0.09		0.11	0.11	0.11	0.14	0.10
6	0.27	0.15	0.22	0.11	0.11		0.11	0.12	0.13	0.11
7	0.21	0.28	0.23	0.29	0.26	0.32		0.10	0.12	0.12
8	0.25	0.16	0.20	0.11	0.12	0.08	0.30		0.11	0.11
9	0.22	0.13	0.23	0.17	0.17	0.17	0.26	0.16		0.13
10	0.21	0.16	0.19	0.10	0.10	0.13	0.29	0.13	0.16	

*Root mean square differences in nm between various headpiece structures for C_α positions of residues 4–47. Upper righthand triangle: RMS differences between DG structures. Lower lefthand triangle: RMS differences between the EM + MD refined structures.

TABLE IV. RMS Positional Shift During the EM and EM + MD Refinement Procedure for Structures S1–S10*

Structure	EM refinement	MD refinement
S1	0.02	0.26
S2	0.03	0.18
S3	0.03	0.15
S4	0.03	0.16
S5	0.03	0.09
S6	0.04	0.15
S7	0.03	0.23
S8	0.03	0.20
S9	0.03	0.13
S10	0.02	0.17

*For the comparison of the initial distance geometry structures to the EM and EM + MD refined structures only the C_α positions of residues 4–47 are used.

structure is lower than the energies found among the refined DG structures, while the total violation approaches the best of the refined DG structures. From Table V it is observed that the rms differences of the refined model-built structure and the ten DG structures are approximately equal and range between 0.14 to 0.17 nm. However, after MD refinement structures S4, S5, S8, and S10 correspond more closely to the refined model-built structure than before. It is striking that the two structures that are the closest to the model-built structure (S5 and S10) are characterized by low values of their potential energies and total violation of the distance constraints (Table II). Structure S5 can even be called the best-refined DG structure according to these two criteria.

On the other hand, for structures S1 and S7 the correspondence to the model-built structure decreased greatly during MD refinement. As noted before, these two refined structures are characterized by (relatively) high internal energies and large violation distances.

A More Elaborate Refinement Procedure

For two structures (S1 and S7) the refinement procedure of 10 ps simulation time was not sufficient, as shown above. In order to investigate if the refinement can be improved by using a more elaborate procedure, the worse structure (S7) was simulated over a period

TABLE V. Root Mean Square Difference (in nm) Between the MD-Refined Model-Built Structure and 1) DG Structures and 2) EM+MD Refined Structures*

Structure	Distance geometry	EM+MD refined
S1	0.15	0.20
S2	0.17	0.17
S3	0.15	0.17
S4	0.16	0.13
S5	0.17	0.10
S6	0.14	0.15
S7	0.17	0.25
S8	0.17	0.14
S9	0.15	0.17
S10	0.16	0.09

*Only C_α atom positions of residues 4–47 are compared.

TABLE VI. Results of a More Elaborate Refinement Procedure*

Structure	E_{pot}	Σ_{dviol}
Initial DG structure	12,442	0.59
MD 10 ps	−2,671	1.07
20 ps	−2,782	1.01
40 ps	−2,872	0.73
50 ps	−2,938	0.55

*The potential energy and distance constraint violation during a more elaborate refinement protocol used for structure S7 (see for definitions of the quantities the footnote to Table I). During this protocol the value of the force constant K_{dc} and temperature T_0 are changed as follows. From 0 to 5 ps: K_{dc} = 500 kJ.mol^{-1}.nm^{-2}; from 5 to 15 ps: 4,000 kJ.mol^{-1}.nm^{-2}, from 15 to 35 ps: 500 kJ.mol^{-1}.nm^{-2}; and from 35 to 50 ps: K_{dc} = 4,000 kJ.mol^{-1}.nm^{-2}. The change of K_{dc} was done gradually at a rate of 750 kJ.mol^{-1}.nm^{-2}.ps^{-1}. The temperature during the run was changed as follows: from 0 to 40 ps: T = 300 K; from 40 to 45 ps: T = 600 K; from 45 to 50 ps: T = 300 K.

of 50 ps in which the temperature was increased to 600 K and decreased again to 300 K over the last 10 ps. As can be seen from Table VI the refinement of structure 7 is improved greatly by using this elaborate procedure. After 40 ps simulation time at normal temperature (300 K) an energetically acceptable structure is generated that satisfies reasonably the experimental distance constraints. By raising the temperature to 600 K (from 40–45 ps) and subsequently cooling down to 300 K (from 45–50 ps), even lower values for the potential energy and distance violations are attained, which are comparable to the values found for the structure S5 (best-refined DG structure). Also, the correspondence to the refined model-built structure is improved from 0.25 at t = 10 ps to 0.19 nm at t = 50 ps.

RMS Positional Fluctuations and NOE-Determination Value

In order to judge roughly the extent to which the relative position of each residue is determined by the NOE data, we defined a NOE distance constraint determination value per residue (rdv). This was done as follows:

$$\mathrm{rdv} = \frac{\sum_{i=1}^{\text{number of (pseudo) atoms}} \sum_{j=1}^{\text{max. of 3 NOEs}} \frac{r_{i,j}^{o,u}}{r_{i,j}^{o}}}{N_{dih}} \quad (4)$$

where

- rdv = NOE distance constraint determination value per residue
- $r_{i,j}^{o,u}$ = basic upper limit distance constraint for the j'th NOE involving (pseudo) atom i
- $r_{i,j}^{o}$ = actual distance constraint corrected for example, dynamic effects or lack of stereospecific assignments
- N_{dih} = number of dihedral degrees of freedom per residue

In Figure 2 the reciprocal value of the NOE-determination value of the residues of *lac* repressor headpiece is shown (bottom). In this figure it is observed that the constraints are not uniformly distributed over the molecule. Especially the N- and C-terminal parts and the loop region between helix I and II are weakly determined by NOE distance constraints. The reciprocal of the NOE determination value is compared to the rms fluctuations in C_α positions for refined and unrefined DG structures (Fig. 2, top). The rms positional fluctuations are calculated by only averaging over the reasonably refined structures. Therefore, structures S1 and S7 are excluded from the calculation. From Figure 2 it is observed that, especially for those parts of the molecule that are poorly determined by NOE distance constraints, the fluctuations are larger in refined than in unrefined structures. Thus, the structures adopt a wider range of conformations after the MD refinement than was generated by the repeated distance geometry calculations, most notably in the poorly determined parts of the molecule.

RMS Positional Differences Between MD-Generated Structures

Although the combination of DG and MD techniques was designed to combine the good sampling properties of the DG algorithm with an efficient method of energy minimization, we found that the MD algorithm contributes significantly to the sampling as well. However, the occurrence of an enlarged spread in the MD-refined DG structures is partly caused by differences between the DG structures used as starting structures of the MD runs. It is not to be expected that during the course of an MD simulation at moderate temperatures the same large differences between generated structures will be observed. In order to test this assumption, a separate 40-ps-long

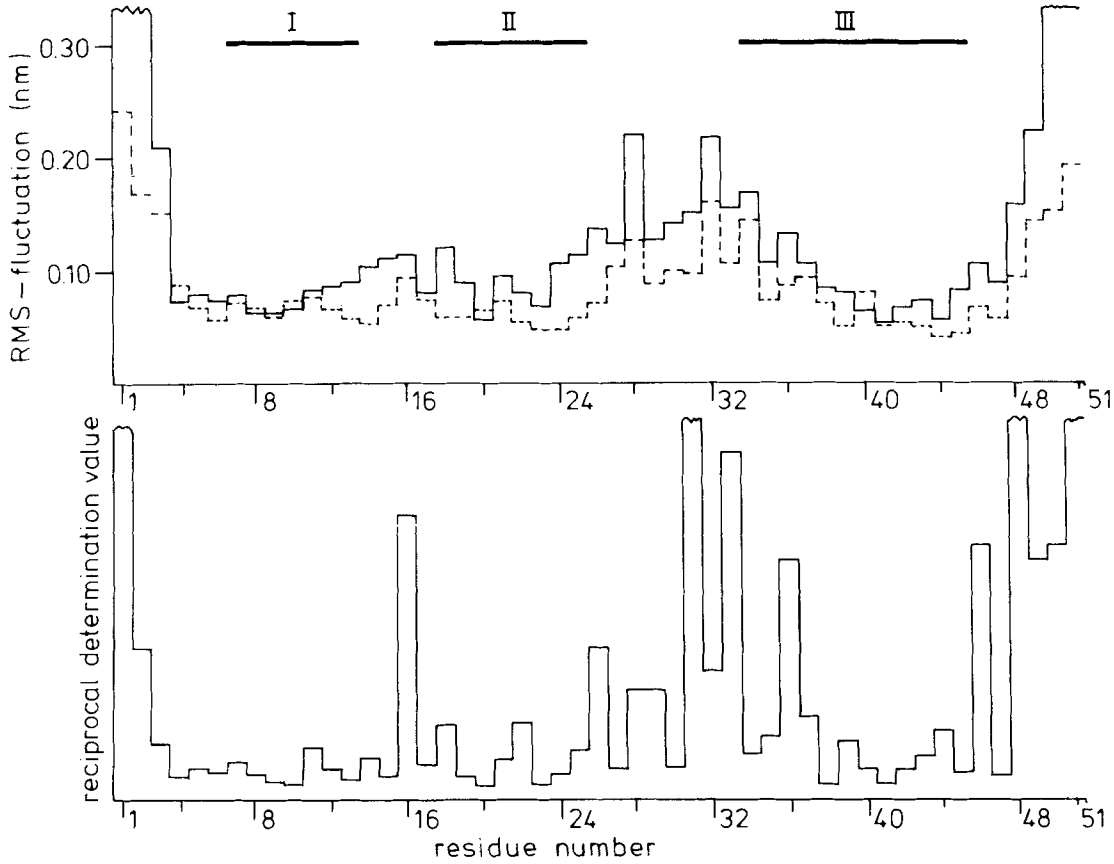

Fig. 2. **Bottom:** Reciprocal value of the NOE-determination value per residue (see text). **Top:** rms positional fluctuations in nm for headpiece C_α atoms averaged over eight structures (S2–S6, S8–S10) obtained by distance geometry calculations (dotted lines) and after EM + MD refinement (solid lines). Ony Ca positions of residues 4–77 are used for superimposing. The thick lines represent the positions of the α-helices I, II and III.

MD simulation was performed (at 300 K), starting from the refined model-built structure and using a force constant of 500 $kJ \cdot mol^{-1} \cdot nm^{-2}$. Every 5 ps a structure was selected and used as an initial structure for another 5 ps MD run in which the force constant was gradually increased to 4,000 $kJ \cdot mol^{-1} \cdot nm^{-2}$ at a rate of 750 $kJ \cdot mol^{-1} \cdot nm^{-2} \cdot ps^{-1}$. Subsequently the structures of the eight side-line MD simulations were energy minimized using again a force constant of 4,000 $kJ \cdot mol^{-1} \cdot nm^{-2}$. In this way the refinement procedure of the distance geometry structures was mimicked as well as possible.

In Table VII the average potential energies and total violation, averaged over eight headpiece structures, and the mean rms positional differences among these structures obtained by DG, DG+ EM+MD, and selection from a MD run, are compared. (Again, structures S1 and S7 are excluded.)

From Table VII it is observed that the spread between structures selected from an MD run is clearly lower than the spread found among unrefined and refined DG structures. At the same time the internal energies and violation distances of the snapshots are lower because of the better starting structure. Pictures of the C_α atoms of the eight unrefined, refined, and MD snapshot structures of the *lac* repressor headpiece are given in Figure 3.

TABLE VII. The Average Potential Energies and Total Violation Distances Averaged Over Eight Headpiece Structures and the Mean rms Potential Differences Among These Structures Obtained by DG, DG+MD, and Selection from an MD Run Eight Structures; See Text*

Structures obtained by	E_{pot}	Σ_{dviol}	RMSD
DG	12,285	0.89	0.12
DG+MD refinement	−2,838	0.69	0.15
Selected from MD run	−3,044	0.52	0.08

*The various quantities are defined in the footnote to Table I.

CONCLUSIONS

From the application of combined DG and restrained MD techniques to the *lac* repressor headpiece using an experimentally determined set of NOE

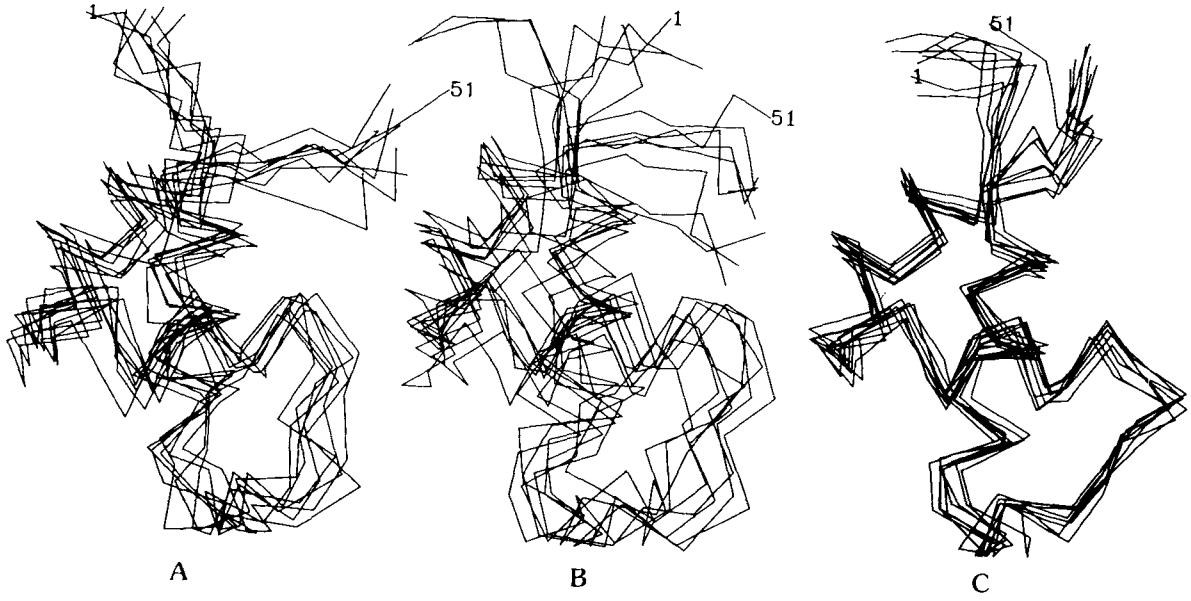

Fig. 3. Pictures of the C_α atoms of eight *lac* repressor headpiece structures generated by the distance geometry algorithm before **(A)** and after EM + MD refinement **(B)**. In **C:** snapshot structures from an MD run are shown. The structures (S2–S6, S8–S10) were superimposed such as to minimize the RMS differences in the C_α positions of residues 4–47.

distance constraints, a number of conclusions can be drawn. Using DG techniques, we generated ten structures with minimal violations of the distance constraints. During NOE-restrained EM refinement the highly strained structures could only be relaxed at the expense of a dramatic increase in the NOE violations. By applying MD refinement even lower values for the internal energies are reached than with EM, and for eight out of ten structures the total violation distance is again brought down to values equal to or even lower than those found among the DG structures. This result is expected, since MD searches a larger part of the conformational space and generally results in a lower energy minimum than EM. For two of the structures it was concluded that the simulation time of 10 ps was insufficient, although these were the best structures generated by the DG algorithm as judged by the violation distance. This emphasizes the necessity of refining DG structures with the use of an energy function that describes the intramolecular potential. A more elaborate procedure totaling 50 ps run time applied to one of these two structures was able to transform it into an acceptable one.

Although the combination of DG and restrained MD techniques was designed to combine the good sampling properties of the DG algorithm with an efficient method of energy minimization, we found that the MD algorithm contributes significantly to the sampling as well, most notably in the poorly defined parts of the molecule. However, the enlarged spread appearing in the MD refined DG structures is mainly caused by the already existing differences in the initial structures. This is confirmed by the fact that during the course of one MD run the differences between MD snapshot structures is lower than that found between different DG structures.

In four cases (S4, S5, S8, and S10), the correspondence to a solution structure, which was found earlier by applying restrained MD to a model-built structure, increased significantly during the MD refinement of the DG structures. In only two cases did the correspondence to the model-built structure decrease greatly (S1, S7) during MD refinement.

We note that all of the distance constraints used can be satisfied simultaneously in all *lac* repressor headpiece structures generated by the combined procedure of DG and MD calculations. This implies that, at the present accuracy of NMR data, there is no evidence for an equilibrium of multiple conformations that could lead to mutally exclusive distance constraints. This is in contrast to the case of the cyclic decapeptide antamanide, where a dynamic exchange between multiple conformations of the molecule occurs, leading to a NMR data set that cannot be satisfied by one conformation.[26]

Finally we conclude that the combination of DG algorithms[11,27,28] to sample the allowed part of conformational space and restrained MD to search for low-energy conformations within the allowed range is a very powerful technique to obtain 3D molecular structures from atom–atom distance information. The same conclusions have been reached by Clore et al.,[29] and Lautz et al.,[30] in their recent comparisons of DG and MD.

ACKNOWLEDGMENTS

This work was supported by the Netherlands Foundation for Chemical Research (S.O.N.) with financial aid from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

REFERENCES

1. Ernst, R.R, Bodenhausen, G., Wokaun, A. "Principles of Nuclear Magnetic Resonance in One or Two Dimensions." Oxford: Clarendon Press, 1987.
2. Wüthrich, K. "NMR of Proteins and Nucleic Acids." New York: J. Wiley and sons, 1986.
3. Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R., van Gunsteren, W.F. A protein structure from nuclear magnetic resonance data: *lac* repressor headpiece. J. Mol. Biol. 182:179–182, 1985.
4. Wüthrich, K., Wider, G., Wagner, G., Braun, W. Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance. J. Mol. Biol. 155:311–319, 1982.
5. Braun, W., Wider, G., Lee, K.H., Wüthrich, K. Conformation of glucagon in a lipid-water interphase by ^{1}H nuclear magnetic resonance. J. Mol. Biol. 169:921–948, 1983.
6. Williamson, M.P., Havel, T.F., Wüthrich, K. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ^{1}H nuclear magnetic resonance and distance geometry. J. Mol. Biol. 182:295–315, 1985.
7. Clore, G.M., Sukumaran, D.K., Nilges, M., Gronenborn, A.M. Three-dimensional structure of phoratoxin in solution: Combined use of nuclear magnetic resonance, distance geometry and restrained molecular dynamics. Biochemistry 26:1732–1745, 1987.
8. Clore, G.M., Sukumaran, D.K., Nilges, M., Zarbock, J., Gronenborn, A.M. The conformations of hirudin in solution: A study using nuclear magnetic resonance, distance geometry and restrained molecular dynamics. EMBO J. 6:529–537, 1987.
9. Clore, G.M., Gronenborn, A.M., Nilges, M., Sukumaran, D.K., Zarbock, J. The polypeptide fold of the globular domain of histone H5 in solution. A study using nuclear magnetic resonance, distance geometry and restrained molecular dynamics. EMBO J. 6:1833–1842, 1987.
10. de Vlieg, J., Boelens, R., Scheek, R.M., Kaptein, R., van Gunsteren, W.F. Restrained molecular dynamics procedure for protein tertiary structure determination from NMR data: A *lac* repressor headpiece structure based on information on J-coupling and from presence and absence of NOE's. Isr. J. Chem. 27:181–188, 1986.
11. Havel, T., Kuntz, I.D., Crippen, G.M. The theory and practice of distance geometry. Bull. Math. Biol. 45:655–720, 1983.
12. Clore, G.M., Gronenborn, A.M., Brünger, A.T., Karplus, M. Solution conformation of a heptadecapeptide comprising the DNA binding helix F of the cyclic AMP receptor protein of *Escherichia Coli*: Combined use of ^{1}H-nuclear magnetic resonance and restrained molecular dynamics. J. Mol. Biol. 186:435–455 1985.
13. Zuiderweg, E.R.P., Kaptein, R. and Wüthrich, K. Sequence-specific resonance assignments in the ^{1}H nuclear-magnetic-resonance spectrum of the *lac* repressor DNA-binding domain 1–51 from *Escherichia coli* by two-dimensional spectroscopy. Eur. J. Biochem. 137:279–292, 1983.
14. Zuiderweg, E.R.P., Scheek, R.M. and Kaptein, R. Two-dimensional ^{1}H-nmr studies on the *lac* repressor DNA binding domain: Further resonance assignments and identification of nuclear overhauser enhancements. Biopolymers 24:2257–2277, 1985
15. Zuiderweg, E.R.P., Kaptein, R., Wüthrich, K. Secondary structure of the *lac* repressor DNA-binding domain by two-dimensional ^{1}H nuclear magnetic resonance in solution. Proc. Natl. Acad. Sci. U.S.A. 80:5837–5841, 1983.
16. Wüthrich K., Billeter M., Braun W. Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. J. Mol. Biol. 169:949–961, 1983.
17. van Gunsteren, W.F., Berendsen, H.J.C. "Groningen Molecular Simulation (GROMOS) Library Manual." Nijenborgh 16, Groningen, The Netherlands: Biomos B.V., 1987.
18. van Gunsteren, W.F., Boelens, R., Kaptein, R., Scheek, R.M., Zuiderweg, E.R.P. An improved restrained molecular dynamics technique to obtain protein tertiary structure from nuclear magnetic resonance data. In: "Molecular Dynamics and Protein Structure." Hermans, J., ed. Western Springs, IL: Polycrystal Bookservice 1985:92–99.
19. Crippen, G.M. Distance geometry and conformational calculations. In: "Chemometrics Research Studies Series." Bawden, D., ed. New York: J. Wiley and sons. 1983:1–58.
20. van Gunsteren, W.F., Karplus, M. A method for constrained energy minimization of macromolecules. J. Comput. Chem. 1:266–274, 1980.
21. van Gunsteren, W.F., Kaptein, R., Zuiderweg, E.R.P. Use of molecular dynamics computer simulation when determining protein structure by 2D-NMR. "Report of a NATO/CECAM Workshop on Nucleic Acid Conformation and Dynamics." Olson, W.K. ed. CECAM, Université de Paris Sud, Orsay, France, 1983:18–29.
22. Äqvist, J., van Gunsteren, W.F., Leijonmarck, M., Tapia, O. A molecular dynamics study of the C-terminal fragment of the L7/L12 ribosomal protein. J. Mol. Biol. 183:461–477, 1985.
23. van Gunsteren, W.F., Berendsen, H.J.C. Molecular dynamics simulations: techniques and applications to proteins. In: "Molecular Dynamics and Protein Structure." Hermans, J. ed. Western Springs, IL: Polycrystal Bookservice. 1985:5–14.
24. Ryckaert, J.P., Ciccotti, G., Berendsen, H.J.C. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. J. Comput. Phys. 23:327–341,1977.
25. van Gunsteren, W.F., Berendsen, H.J.C. Algorithms for macromolecular dynamics and constraint dynamics. Mol. Phys. 34:1311–1327, 1977.
26. Kessler, H., Griesinger, C., Lautz, J., Müller, A., van Gunsteren, W.F., Berendsen H.J.C. Conformational dynamics detected by nuclear magnetic resonance NOE-values and J-coupling constants. J.A.C.S., in press, 1988
27. Havel, T., Wüthrich, K. A distance geometry program for determining the structures of small proteins and other macromolecules from NMR measurements of intramolecular proton-proton proximities in solution. Bull. Math. Biol. 46:673–698, 1984.
28. Braun, W., and Gō, N. Calculation of protein conformations by proton–proton distance constraints: A new efficient algorithm. J. Mol. Biol. 186:611–626, 1985.
29. Clore, G.M., Nilges, M., Brünger, A.T., Karplus, M., Gronenborn, A.M. A comparison of the restrained molecular dynamics and distance geometry methods for determining three-dimensional structures of proteins on the basis of interproton distances. FEBS Lett. 213:269–277, 1987.
30. Lautz, J., Kessler, H., Blaney, J.M., Scheek, R.M., van Gunsteren, W.F. On calculation three-dimensional molecular structure from atom–atom distance information. Submitted to Int. J. Peptide Protein Res.