Journal of Biomolecular NMR, 5 (1995) 33–40
ESCOM

J-Bio NMR 270

Dynamic modelling of a helical peptide in solution using NMR data: Multiple conformations and multi-spin effects

Johan Kemmink[a] and Ruud M. Scheek[b,*]

[a]*European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69012 Heidelberg, Germany*
[b]*The Groningen Biomolecular Science and Biotechnology Institute, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands*

Received 11 November 1994
Accepted 9 January 1995

Keywords: Molecular dynamics; Ensemble dynamics; Peptide conformation; NOE; Spin diffusion; Bovine pancreatic trypsin inhibitor (BPTI)

Summary

Nuclear Overhauser effect (NOE) measurements on molecules in solution provide information about only the ensemble-averaged properties of these molecules. An algorithm is presented that uses a list of NOEs to produce an ensemble of molecules that on average agrees with these NOEs, taking into account the effect of surrounding spins on the buildup of each NOE ('spin diffusion'). A simplified molecular dynamics simulation on several copies of the molecule in parallel is restrained by forces that are derived directly from differences between calculated and measured NOEs. The algorithm is tested on experimental NOE data of a helical peptide derived from bovine pancreatic trypsin inhibitor.

Introduction

High-resolution NMR has become a well-established technique for determining three-dimensional structures of small proteins and nucleic acid fragments in solution (Wüthrich, 1986). The richest source of information is the nuclear Overhauser effect (NOE), which results in cross peaks in a multidimensional NOE experiment at the resonance frequencies of spins that are in close proximity to each other. In the classical modelling strategy, the intensities of the cross peaks are first translated into spin–spin distances. This is often accomplished by a qualitative classification of the observed NOE cross peaks into strong, medium and weak intensities, corresponding to certain distance bounds. This information is then used to derive a set of static structures using methods like distance geometry, usually followed by a simulated annealing or a restrained molecular dynamics (MD) protocol.

A complicating phenomenon, often observed in multidimensional NOE spectra recorded with relatively long mixing times, is the presence of cross peaks resulting from indirect magnetization transfer ('spin-diffusion effects'). Such a cross peak should not be converted directly into a distance, because its intensity depends not only on the distance between the spins for which the cross peak is observed, but also on the relative positions of nearby other spins. The relative contributions of the 'direct' and 'indirect' magnetization transfer can be calculated correctly only if, among other things, the 3D structure of the biomolecule is known. As it is the structure that is being determined, this is a circular problem, which can be tackled by using relaxation matrix approaches like IRMA, MARDIGRAS and NO2DI (Boelens et al., 1989; Borgias and James, 1990; Van de Ven et al., 1991). In this paper a different method is described, in which differences between measured and calculated NOE intensities are directly converted into forces (Yip and Case, 1989; Nilges et al., 1991; Stawarz et al., 1992; Bonvin et al., 1994) acting during the dynamic annealing of an ensemble of molecules.

In the classical approach of structure determination by

*To whom correspondence should be addressed.

NMR, a structure is accepted as a member of the solution set only if it is consistent with all or most of the measured NMR distances simultaneously. Such an approach denies, however, the fact that biomolecules like proteins and nucleic acid fragments are flexible and can adopt a variety of conformations. Recently, Torda et al. (1989,1990) presented a method based on restrained MD, which generates a molecular dynamics trajectory in which distance restraints are satisfied as averages over time. Here we present an alternative algorithm, which generates multiple copies of a single molecule that satisfy NOE-derived distance restraints as ensemble averages.

The modelling strategy will be illustrated on a peptide representing the 14 C-terminal residues 45–58 of bovine pancreatic trypsin inhibitor (BPTI) dissolved in a 1:1 mixture of H_2O and TFE (trifluoroethanol). In native BPTI, this segment of the protein comprises the α-helix ranging between residues 48–55. In aqueous solution, the corresponding peptide ($P_{45\ 58}$) shows only a small preference for an α-helical conformation, as evidenced both by CD and ^1H NMR spectroscopy (Goodman and Kim, 1989; Kemmink and Creighton, 1993). Dissolved in an H_2O/TFE mixture, however, the peptide P_{45-58} shows a clear preference for helical conformations.

Theory

Multi-spin effects and chemical exchange

Relaxation and exchange in a system of N spins can be described by the generalized Bloch equations (Ernst et al., 1987). The time evolution of the longitudinal magnetization, relaxing to thermal equilibrium, is described by a set of coupled first-order differential equations. In a matrix representation, this is written as:

$$\frac{d\Delta\mathbf{M}_z(t)}{dt} = \mathbf{L}\Delta\mathbf{M}_z(t) \tag{1}$$

The vector $\Delta\mathbf{M}_z(t)$ of dimension N represents the deviations of the longitudinal magnetization from thermal equilibrium for the N spins:

$$\Delta\mathbf{M}_z(t) = \mathbf{M}_z(t) - \mathbf{M}_0 \tag{2}$$

while the rate matrix **L** incorporates relaxation rates (relaxation matrix **R**) and chemical exchange rate constants (kinetic matrix **K**):

$$\mathbf{L} = -\mathbf{R} + \mathbf{K} \tag{3}$$

The diagonal elements of the relaxation matrix **R** describe the effect of spin-lattice relaxation, while the off-diagonal elements of **R** express the effects of cross-relaxation. The diagonal matrix elements R_{ii} consist of the usual dipolar term (see for example Goldman (1988)) and R_{leak}, a phenomenological term that accounts for all other sources of relaxation:

$$R_{ii} = 0.1\gamma^4\hbar^2\left(\frac{\mu_0}{4\pi}\right)^2 (J(0) + 3J(\omega_0) + 6J(2\omega_0)) \times \sum_{j\neq i}\langle r_{ij}^{-6}\rangle + R_{leak} \tag{4}$$

where ω_0 is the Larmor frequency. The off-diagonal elements R_{ij} contain only a dipolar term:

$$R_{ij} = 0.1\gamma^4\hbar^2\left(\frac{\mu_0}{4\pi}\right)^2 (6J(2\omega_0) - J(0))\langle r_{ij}^{-6}\rangle \tag{5}$$

In this approximation, only fluctuations in the magnitude of the vector connecting spins i and j are taken into account, whereas angular fluctuations are neglected. Brackets $\langle\ \rangle$ denote averaging over the molecules of the ensemble. Whether the r_{ij} dependence is of the type $\langle r_{ij}^{-6}\rangle$ or $\langle r_{ij}^{-3}\rangle^2$ depends on the approach chosen. Since we consider the initial ensemble as consisting of identical conformers, we chose $\langle r_{ij}^{-3}\rangle^2$ averaging in Eq. 5, as discussed later. In the case of uniform isotropic tumbling, the spectral density J(ω) is a simple Lorentzian function. When τ_c is the time constant of the exponential auto-correlation function, J(ω) is described as follows:

$$J(\omega) = \frac{\tau_c}{1 + (\omega\tau_c)^2} \tag{6}$$

The formal solution of Eq. 1 is:

$$\Delta\mathbf{M}_z(t) = [e^{+\mathbf{L}t}]\ \Delta\mathbf{M}_z(t = 0) \tag{7}$$

The exponential matrix can be written as a series expansion:

$$\exp[+\mathbf{L}t] = \mathbf{1} + \mathbf{L}t + 1/2\mathbf{L}^2t^2 + \ldots\ldots \tag{8}$$

For a sufficiently short time t, the exponential matrix can be approximated by the first two terms only, treating spins i and j as an isolated pair (two-spin approximation). In this case, the buildup of an NOE varies linearly with the mixing time. Multi-spin effects (relay of magnetization to a third spin and further), which are often apparent in 2D NOE spectra, are not taken into account in this approximation.

Equation 1 was solved by numerical integration. For each NOE that was measured, the redistribution of longitudinal magnetization was calculated in a network of n spins that affect the development of this NOE. This network consists of the two spins i and j involved in the NOE_{ij}, all the spins k that are close to spin i or j (significant R_{ik} or R_{jk} elements in matrix **L**), and spins k that are coupled via a kinetic process to spin i or j (significant K_{ik}

or K_{jk} elements in **L**). To calculate the NOE between spins i and j, the network is first initialized by giving all n spins zero excess magnetization, except spin i ($\Delta M_l(0) = \delta_{li}$, where $1 \leq l \leq n$; $\delta = 1$ if $l = i$ and $\delta = 0$ if $l \neq i$).

The distances between the spins in the network are calculated from a set of structures and averaged ($\langle r_{ij}^{-6} \rangle$ or $\langle r_{ij}^{-3} \rangle^2$, see Discussion). From these averaged distances the relaxation rates can be calculated using Eqs. 4 and 5, assuming that values for τ_c and R_{leak} are known from other measurements or can be estimated. These relaxation rates are combined with rate constants for relevant kinetic processes (e.g., ring flips of tyrosine and phenylalanine side chains) in the matrix **L**, as described by Eq. 3. Next, the redistribution of longitudinal magnetization over the network is calculated using time steps Δt that are much smaller than the reciprocal of the largest matrix element of matrix **L** (Eq. 3). The time course of the n-dimensional vector $\mathbf{\Delta M_z}$ can then be approximated by using only the first two terms of Eq. 8:

$$\mathbf{\Delta M_z}(t + \Delta t) \approx \mathbf{\Delta M_z}(t) + \mathbf{L}\Delta\mathbf{M_z}(t)\,\Delta t \tag{9}$$

After integration from $t = 0$ to $t = \tau_m$ (the mixing time used in the NOE experiment), the NOE intensity between spins i and j can be derived from the magnetization transferred from spin i to spin j:

$$NOE_{ij}^{calc} = \Delta M_j(\tau_m) \tag{10}$$

This procedure is repeated for all spin pairs for which experimental NOE intensities are available, and a list of apparent distances for a given ensemble of molecules results:

$$\langle d_{ij}^{app} \rangle = \left(\frac{NOE_{ij}^{calc}}{0.1\gamma^4 \hbar^2 \left(\frac{\mu_0}{4\pi}\right)^2 (J(0) - 6J(2\omega_0))\tau_m} \right)^{-1/6} \tag{11}$$

These calculated apparent distances can be compared with upper (u_{ij}) and lower bounds (l_{ij}) derived from the NOE experiment. Violations of the experimental distance restraints (u_{ij} or l_{ij}) are directly translated into forces which act on the spins i and j in a simplified molecular dynamics protocol, as will be described below. Defining the apparent distances in this way implies that they equal the real Cartesian distances if τ_m is very small compared to the inverse of the cross-relaxation rates. In this sense, classical distance restraining is a limiting case of the present algorithm.

Restrained molecular dynamics on an ensemble of molecules

To adjust a starting set of structures so that it is collectively consistent with experimental NOE data, we applied the technique of molecular dynamics, using a simplified force field containing the following two terms (Scheek et al., 1991):

$$V_{tot} = V_{holo} + V_{NOE} \tag{12}$$

The holonomic term serves to preserve the chemical integrity of the molecule and can be considered as a simplistic version of the physical force field used in molecular dynamics simulation programs. The NOE term serves to impose the experimental restraints.

The holonomic potential is composed of a distance and a chiral contribution:

$$V_{holo} = K_{DC}\left(\sum_{d_{ij} > u_{ij}} \left(d_{ij}^2 - u_{ij}^2\right)^2 + \sum_{d_{ij} < l_{ij}} \left(l_{ij}^2 - d_{ij}^2\right)^2 \right) + K_{CH} \sum_{\text{quadruples i}} \left(V_i - V_i^t\right)^2 \tag{13}$$

where u_{ij} and l_{ij} stand for the upper and lower bounds on the holonomic distance d_{ij} between atoms i and j and the summation is over all the atom pairs in the molecule that violate an upper or lower bound. V_i is the signed volume of the tetrahedron spanned by the four ligands of a (pro)chiral center i, with target value V_i^t. This term serves to impose the proper chirality on such quadruples of atoms and the sum runs over all these quadruples in the molecule. K_{DC} and K_{CH} are weighting factors which have to be balanced with the weighting factor of the experimental (NOE) restraints (K_{NOE}; see below). 'Holonomic forces' acting on each atom pair i and j are derived from this potential by taking the gradient with respect to the vector connecting i and j ($\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$; Scheek et al., 1991).

The NOE potential has the following form, Eq. 14:

$$V_{NOE} = K_{NOE} \sum_{\langle d_{ij}^{app} \rangle > u_{ij}} \left(1 - \left(\frac{\langle d_{ij}^{app} \rangle}{u_{ij}} \right)^2 \right)^2 + K_{NOE} \sum_{\langle d_{ij}^{app} \rangle < l_{ij}} \left(1 - \left(\frac{\langle d_{ij}^{app} \rangle}{l_{ij}} \right)^2 \right)^2 \tag{14}$$

in which u_{ij} and l_{ij} are the experimental restraints and $\langle d_{ij}^{app} \rangle$ is given by Eq. 11. This form of the NOE potential was chosen because it emphasizes violations of the short upper and lower bounds. This was done for two reasons: (i) these distances are in general the most reliable, and (ii) they are primarily the result of direct magnetization transfer. In this way the algorithm treats the spin-diffusion effects as corrections.

An expression for the restraining forces can be obtained as follows:

$$\mathbf{F}_i = -\frac{\partial V_{NOE}}{\partial \langle d_{ij}^{app} \rangle} \frac{\partial \langle d_{ij}^{app} \rangle}{\partial d_{ij}^{app}} \frac{\partial d_{ij}^{app}}{\partial d_{ij}} \frac{\partial d_{ij}}{\partial \mathbf{r}_i} \tag{15}$$

Application of Eq. 15 in the limiting case where d_{ij} equals $\langle d_{ij}^{app} \rangle$ yields an expression for these forces. The results

presented in this paper were obtained using the following slightly modified expression for the restraining forces:

$$\mathbf{F}_i = \begin{cases} -4K_{NOE}\left(1-\left(\dfrac{\langle d_{ij}^{app}\rangle}{u_{ij}}\right)^2\right)\dfrac{(\mathbf{r}_i-\mathbf{r}_j)}{u_{ij}d_{ij}} & \langle d_{ij}^{app}\rangle > u_{ij} \\ 0 & l_{ij} \le \langle d_{ij}^{app}\rangle \le u_{ij} \\ -4K_{NOE}\left(1-\left(\dfrac{\langle d_{ij}^{app}\rangle}{l_{ij}}\right)^2\right)\dfrac{(\mathbf{r}_i-\mathbf{r}_j)}{l_{ij}d_{ij}} & \langle d_{ij}^{app}\rangle < l_{ij} \end{cases}$$

$$\mathbf{F}_j = -\mathbf{F}_i \qquad (16)$$

Note that every atom experiences the same restraining force in each molecule of the ensemble.

Summary of the algorithm

The complete procedure can now be summarized. A starting structure is generated using a distance geometry program and the experimental NOE data (Havel et al., 1983). An ensemble of identical copies of this starting structure is created. With this ensemble, a parallel MD simulation is started by giving all atoms randomly distributed initial velocities, corresponding to the desired temperature. Next, the holonomic forces (F_{holo}) are calculated from V_{holo} (Eq. 13) for each member of the ensemble separately. Proton–proton distances are collected for all relevant proton pairs, averaged over the ensemble and converted into restraining forces (using Eq. 16). Finally, holonomic and NOE forces are summed and used to calculate new atomic positions after a molecular dynamics time step. Usually, simulations are performed in two stages: the first is at elevated temperature with rather tight coupling to a bath of that temperature (typically 1000 time steps at elevated temperature with a coupling time of 10 steps (Berendsen et al., 1984)); in the second stage, the temperature is slowly lowered (usually in 1000 steps with a temperature coupling of 50 steps).

Materials and Methods

General

Peptide P_{45-58} (with the sequence (in one-letter code) FKSAEDSMRTSGGA) was synthesized using 9-fluorenyl-methoxycarbonyl chemistry. The sequence is identical to residues 45 to 58 in BPTI, except that the original cysteine residues at positions 51 and 55 were replaced by serines to avoid problems with oxidation. The N-terminus was blocked with an acetyl group in order to mimic the corresponding fragment in BPTI.

^{1}H NMR experiments were performed on a sample containing 10 mg of peptide dissolved in 50% H_2O/50% TFE-d_3 v/v (Cambridge Isotopes). The pH was adjusted to 4.6 by adding small amounts of KOH or HCl solutions.

^{1}H assignments, the observed backbone NOE patterns and CD spectra of P_{45-58} will be presented elsewhere (Kemmink, J. and Creighton, T.E., manuscript in preparation).

NOE buildup series

An NOE buildup series was recorded on a Bruker AMX-600 spectrometer operating at 600.141 MHz at 271 K. The NOE mixing times were 25, 50, 100, 200, 300 and 500 ms. A spectral width of 6024 Hz was used. Solvent suppression was accomplished by presaturation of the H_2O signal. FIDs were acquired with 2048 data points,

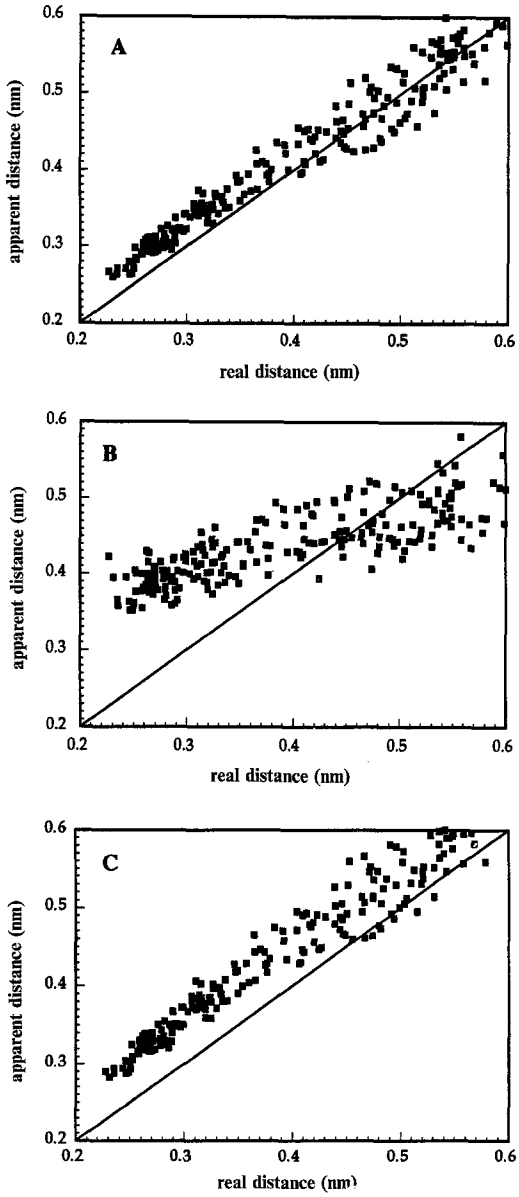

Fig. 1. Comparison of a set of actual ^{1}H-^{1}H distances (real distance) versus the distances calculated using a mixing time of 500 ms (apparent distance). The actual distances were taken from an ensemble of helical structures of P_{45-58}, generated using NOE data extrapolated to zero mixing time. The apparent distances were calculated from Eqs. 10 and 11 using the following parameters: (A) $\tau_c = 1$ ns, $R_{leak} = 0$ s^{-1}; (B) $\tau_c = 10$ ns, $R_{leak} = 0$ s^{-1}; (C) $\tau_c = 1$ ns, $R_{leak} = 1$ s^{-1}.

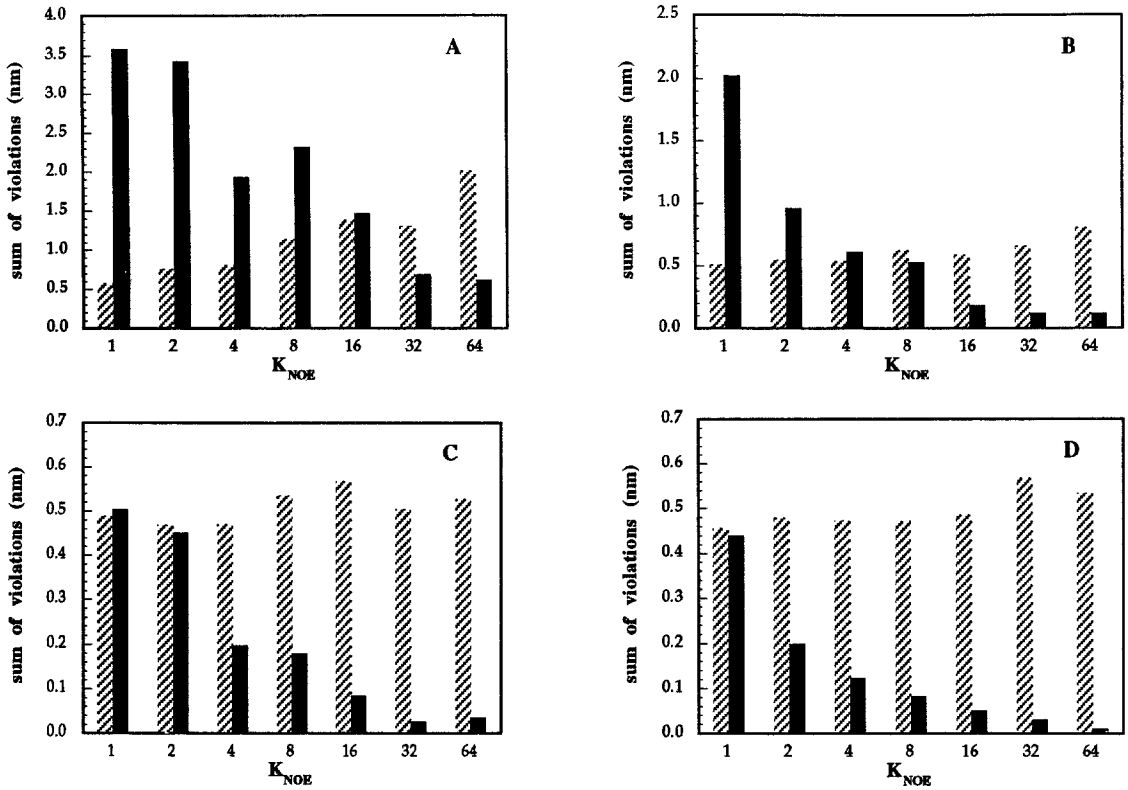

Fig. 2. The sum of the violations of the holonomic constraints (dashed bars) and the NOE restraints (solid bars) as a function of the NOE restraining force constant, K_{NOE} (Eq. 14). Four different ensemble sizes were used for the calculations: (A) one, (B) two, (C) four and (D) eight structures per ensemble. Experimental NOE data were used extrapolated to 0 ms mixing time. Note that the vertical scale is expanded in B, C and D compared to A.

accumulating 32 scans. For each spectrum, 512 t_1 values were recorded using the TPPI method for quadrature detection in this dimension (Marion and Wüthrich, 1983).

Data processing

Data were processed on a Silicon Graphics Indigo workstation using the program SNARF (F.H.J. van Hoesel, University of Groningen, Groningen). Data points were weighted by a Lorentzian–Gaussian transformation in the t_2 dimension and a shifted sine bell in the t_1 dimension. After zero-filling and Fourier transformation, the final matrix contained 1024×1024 real points in both dimensions. Polynomial baseline corrections were routinely applied in both dimensions.

Distance data collection

Analysis of the collected data was performed using the program SNARF. A total of 253 NOE cross peaks were selected from the 500 ms NOESY spectrum, and their cross-peak volumes were determined in each of the six NOESY spectra with different mixing times. Subsequently, all cross-peak volumes were converted into distances such that the distances between the pairs of protons of four methylene groups (Phe^{45} ($C^{\beta}H_2$), Ser^{47} ($C^{\beta}H_2$), Asp^{50} ($C^{\beta}H_2$) and Met^{52} ($C^{\gamma}H_2$)), extrapolated to 0 ms mixing time, were equal to 0.18 ± 0.01 nm. The apparent distances belonging to each cross peak were subjected to a linear fit with respect to the mixing time. In case of a very poor fit, which was sometimes caused by measurements of low-intensity NOEs at relatively short mixing times, such points were discarded until a reasonable fit was obtained. This fit was used to calculate upper and lower bounds on the apparent distances for the different mixing times. Note that only the uncertainties in the measurement of NOE intensities determine the bounds on these apparent distances.

Following this procedure, a total of 372 entries were stored into each restraint file, corresponding to mixing times of 0, 25, 50, 100, 200, 300 and 500 ms, respectively. Double entries for undistinguishable methylene and aromatic protons account for the fact that the number of entries is greater than the 253 cross peaks mentioned above.

Results

Properties of peptide P_{45-58}

Fully folded BPTI consists of two major elements of secondary structure: an antiparallel β-sheet consisting of residues Ile^{18}–Asn^{24} and Leu^{29}–Tyr^{35} and an α-helix consisting of residues Ala^{48}–Cys^{55}. These are packed together to form a very stable core of hydrophobic side chains

(Deisenhofer and Steigemann, 1975). Although comprising the helical residues in folded BPTI, peptide P_{45-58} in aqueous solution shows only very limited preference for a helical conformation, as concluded from both CD and NMR measurements (Goodman and Kim, 1989; Kemmink and Creighton, 1993). This is not atypical; in general, small peptide fragments in aqueous solution show only very limited or no preference for a specific conformation (see, for example, Dyson and Wright (1991)).

Addition of TFE greatly enhances the helix propensity of this sequence, resulting in CD and ^1H NMR spectra that are typical of helical conformations (Kemmink, J. and Creighton, T.E., unpublished observation). However, 2D NOE spectra of P_{45-58} dissolved in 50% H_2O/50% TFE recorded at low temperature (271 K) show not only the short- and medium-range NOEs typical of a pure α-helix (see Wüthrich (1986)), but also other cross peaks, indicative of the existence of multiple conformations. Another phenomenon observed in NOE spectra of this sample recorded with relatively long mixing times (300–500 ms) is the vast number of NOE cross peaks present, compared to the same spectra recorded in 100% H_2O. This difference between the H_2O and H_2O/TFE spectra can only be explained in part by the presence of helical structure in the latter solvent, giving rise to a number of medium-range cross peaks. Most of the extra NOE cross peaks are spin-diffusion effects caused by the difference in physical properties of the two solvents, especially the viscosity.

Spin-diffusion effects

To illustrate the implications of spin diffusion, Fig. 1 presents comparisons of the actual distances for a set of spin pairs in the model helical peptide P_{45-58} with the corresponding apparent distances, calculated according to Eqs. 10 and 11. Long mixing times τ_m and large values of the effective correlation time τ_c emphasize the effects of the onset of spin diffusion (Figs. 1A and B): relay of magnetization from a spin pair ij to other spins in the network tends to increase the apparent distance between i and j, while indirect magnetization transfer between i and j via pathways involving other spins tends to decrease the apparent distance between them. In the limiting case, all apparent distances tend to the same value (all NOEs become equal). Kinetic processes like aromatic ring flips may function as very efficient pathways for magnetization transfer: spins that are as far apart as 0.8 nm in a model structure may appear to be at a distance of only 0.4 nm (not shown). Note that in this approach the use of pseudoatoms and corresponding pseudocorrections for aromatic ring protons is not necessary. Increasing the value of R_{leak} results in an increase of all apparent distances (Fig. 1C).

Ensemble dynamics

Extrapolated initial buildup rates of the measured NOEs of $P_{45\ 58}$ in H_2O/TFE were used to generate a list of upper and lower distance bounds without the compli-

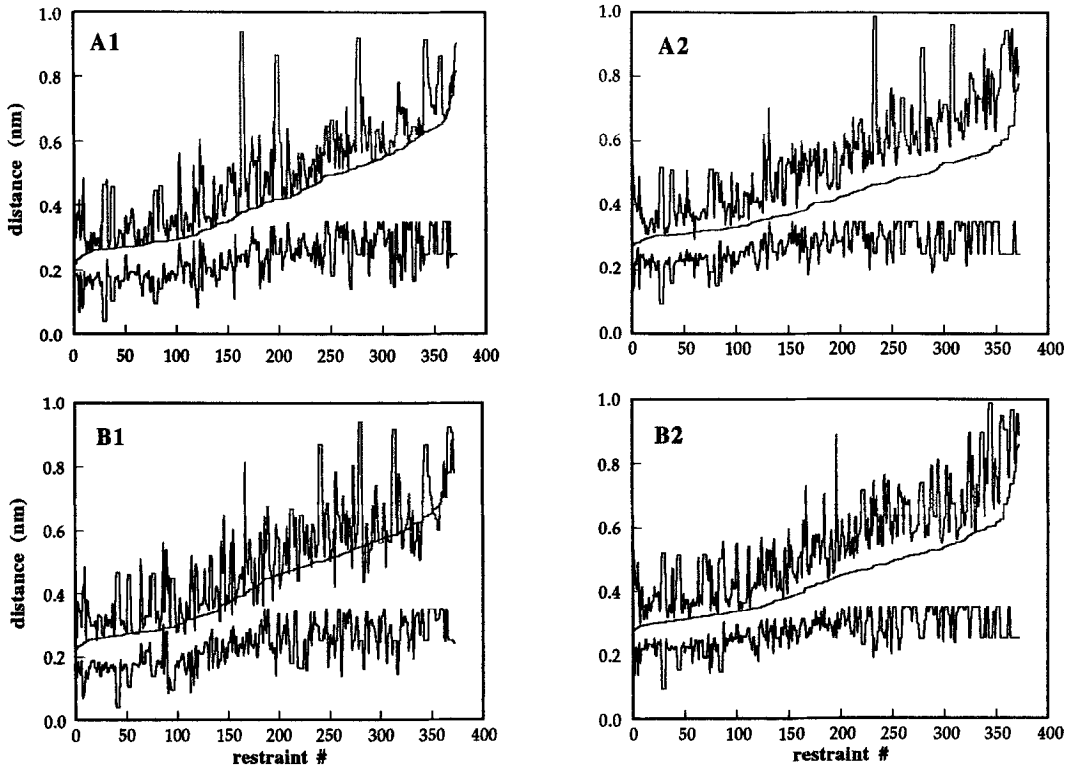

Fig. 3. Comparison of averaged distances (middle trace) calculated from ensembles of P_{45-58} in H_2O/TFE, generated using (A) 0 ms and (B) 300 ms mixing time NOE data with the upper bounds (top trace) and lower bounds (bottom trace) derived from NOE data with 0 ms (**1**; extrapolated data) and 300 ms (**2**) mixing times. The restraints are sorted with respect to the calculated distance.

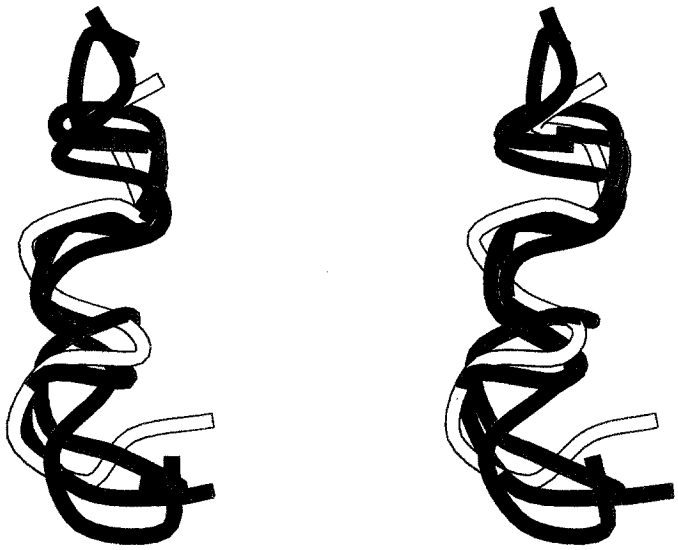

Fig. 4. Stereoscopic schematic representation of the backbone of the final four ensemble members of P_{45-58}, generated using the 100 ms NOE data ($R_{leak} = 5\ s^{-1}$ and $\tau_c = 4$ ns). The backbone N, C^{α}, C and O atoms of the structures were superimposed. Four identical extended structures were used as starting structures. The final ensemble did not show any violations of the imposed distance bounds. The drawing was created using the program MOLSCRIPT (Kraulis, 1991).

cation of spin diffusion, as described in the Methods section. These distances were used as restraints in structure calculations employing our dynamic annealing protocol, in which we investigated the effect of the magnitude of the restraining force constant in combination with the size of the modelled ensemble. With only one structure per 'ensemble' (Fig. 2A), it was not possible to obtain good agreement with the experimental data (low sum of violations of NOE restraints) without distorting bond distances and angles (holonomic constraints). Good agreement with experimental data could be obtained only at the expense of the holonomic part of the potential used, by applying a high force constant for the restraining forces. This situation can be improved by introducing more structures per ensemble: a significantly better agreement was obtained using two structures per ensemble (Fig. 2B). With an ensemble of at least four conformers, the conflict between experimental and holonomic parts of the 'potential' disappeared (Fig. 2C). A further increase in the size of the ensemble had little effect (Fig. 2D).

Modelling of peptide P_{45-58}

Obviously, a good model ensemble should be consistent with the experimental data obtained at all mixing times. Thus, the experimental restraints obtained for P_{45-58} at a mixing time of 300 ms, and from the extrapolated initial buildup rates, were used to generate two ensembles of four conformers each, using the dynamic annealing protocol. These two ensembles were used to calculate the apparent distances corresponding to the NOEs measured at mixing times of 0 (extrapolated data) and 300 ms. Figure 3 shows a comparison of these calculated apparent distances (middle trace) with the corresponding experimental upper and lower distance bounds (top and bottom traces, respectively). The success of the annealing protocol when using the mixing time at which the ensemble was generated in the first place is clear from the comparison of the apparent distances that were calculated from the two ensembles (Figs. 3A1 and B2). Individual conformers of the peptide from the two ensembles were in agreement with the holonomic constraints in all cases, but not necessarily with the experimental restraints, as these were imposed only as ensemble averages. When apparent distances were calculated from an ensemble using a longer mixing time than that used to generate the ensemble, they agreed with the corresponding experimental bounds as well (Fig. 3A2). The reverse is not the case: ensembles obtained from longer mixing time NOEs do not agree as well with the NOEs obtained at shorter mixing times. Thus, the ensemble generated from the extrapolated 0 ms NOE data is also consistent with the 300 ms NOE data (Fig. 3A2), but the ensemble generated using the 300 ms NOE data is not in agreement with the 0 ms data set (Fig. 3B1). Note that the apparent distances have a distribution that is biased towards the experimental upper bounds. This probably reflects the absence of nonbonded attractive interactions in the 'force field' used, which is expected to lead to somewhat extended structures.

Figure 4 shows the ensemble obtained from the 100 ms data set. In spite of the limitations of the simplified force field used, it is clear that the peptide adopts a partly helical conformation under the experimental conditions employed. This demonstrates the intrinsic tendency of this peptide to adopt a native-like conformation in solution.

Discussion and Conclusions

A new approach for the dynamic modelling of mol-

ecular structures using NMR data is described, in which inconsistencies between an ensemble of model structures and experimental data are directly translated into restraining forces that can be used in a dynamic annealing protocol to adjust the ensemble as a whole. Any inconsistency between the magnitudes of a measured NOE between two spins i and j and an NOE calculated from the ensemble of model structures is corrected by applying restraining forces according to Eq. 16. In this equation, the forces only approximate the proper derivatives of the corresponding potential (Eqs. 14 and 15) with respect to the atomic position vectors (Yip and Case, 1989). Thus, the direction of these forces is taken to be precisely along the vector $\mathbf{r}_i - \mathbf{r}_j$, and multi-spin effects are taken into account only in the sign and magnitude of the restraining forces. The direction of the restraining forces is close to optimal for the shorter distances, but may be far from optimal for the longer distances, where deviations from the two-spin approximation are largest. For this reason, the forces restraining the shorter distances are weighted more strongly by choosing the form given in Eq. 16. The rationale behind these choices is that, by improving the shorter distances preferentially, improved pathways for indirect magnetization transfer are created, so that the longer apparent distances (corresponding to weaker NOEs) will also be calculated more reliably. Moreover, due to thermal fluctuations, atoms will hardly ever move precisely along the actual gradients of the force field during a molecular dynamics simulation, so exact calculation of both the magnitude and the direction of the restraining forces is probably unnecessary.

The procedure can be implemented directly in more sophisticated MD simulation programs, because forces are delivered. The present version is valid only for molecules that tumble isotropically. More complicated situations can be handled, assuming some model of the dynamic behaviour (e.g. Lipari and Szabo, 1982a,b). Another possibility, which we prefer, is to measure parameters describing molecular motions directly (Peng and Wagner, 1992) and to use these to improve the calculation of the apparent distances (NOEs).

The parallel MD simulations presented here start with an ensemble of molecules having the same initial conformation. The size of the ensemble is limited to the minimum needed to account for all experimental data. Ideally, with a correct force field and a sufficiently long simulation time, larger ensembles can be used: the ensemble will approach a proper Boltzmann distribution over conformations that are accessible at the temperature used in the simulation. As in the approach advocated by Torda et al. (1989,1990), the effects of fast local motions are dealt with by using $\langle r^{-3} \rangle^2$ averaging over the ensemble. Our calculations demonstrate that this approach is able to adjust the ensemble as a whole, such that the calculated apparent ensemble-averaged spin–spin distances agree with the measured experimental values. Apparently, the conformers needed to account for all NOEs are interconvertible during the simulation period. In other cases, it may be necessary to start with an ensemble of conformers that are not interconvertible on the MD time scale, which is presently on the order of nanoseconds for small molecules. In this case, $\langle r^{-6} \rangle$ averaging should be considered, together with weighting factors reflecting the concentrations of the different conformers.

Acknowledgements

We thank Dominique Nalis and Richard Jacob for peptide synthesis and purification. The advice given by Herman Berendsen and Thomas Creighton is gratefully acknowledged.

References

Berendsen, H.J.C., Postma, J.P.M., Van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) *J. Chem. Phys.*, **81**, 3684–3690.

Boelens, R., Koning, T.M.G., Van der Marel, G.A., Van Boom, J.H. and Kaptein, R. (1989) *J. Magn. Reson.*, **82**, 290–308.

Bonvin, A.M.J.J., Boelens, R. and Kaptein, R. (1994) *J. Magn. Reson.*, **4**, 143–149.

Borgias, B.A. and James, T.L. (1990) *J. Magn. Reson.*, **87**, 475–487.

Deisenhofer, J. and Steigemann, W. (1975) *Acta Crystallogr.*, **B31**, 238–250.

Dyson, H.J. and Wright, P.E. (1991) *Annu. Rev. Biophys. Biophys. Chem.*, **20**, 519–538.

Ernst, R.R., Bodenhausen, G. and Wokaun, A. (1987) *Principles of NMR in One and Two Dimensions*, Clarendon Press, Oxford.

Goldman, M. (1988) *Quantum Description of High-Resolution NMR in Liquids*, Clarendon Press, Oxford, pp. 226–264.

Goodman, E.M. and Kim, P.S. (1989) *Biochemistry*, **28**, 4343–4347.

Havel, T.F., Kuntz, I.D. and Crippen, G.M. (1983) *Bull. Math. Biol.*, **45**, 665–720.

Kemmink, J. and Creighton, T.E. (1993) *J. Mol. Biol.*, **234**, 861–878.

Kraulis, P.J. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.

Lipari, G. and Szabo, A. (1982a) *J. Am. Chem. Soc.*, **104**, 4546–4559.

Lipari, G. and Szabo, A. (1982b) *J. Am. Chem. Soc.*, **104**, 4559–4570.

Marion, D. and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.*, **113**, 967–974.

Nilges, M., Habazettl, J., Brünger, A.T. and Holak, T.A. (1991) *J. Mol. Biol.*, **219**, 499–510.

Peng, J.W. and Wagner, G. (1992) *Biochemistry*, **31**, 8571–8586.

Scheek, R.M., Torda, A.E., Kemmink, J. and Van Gunsteren, W.F. (1991) In *Computational Aspects of the Study of Biological Macromolecules by NMR* (Eds, Hoch, J.C., Poulsen, F.M. and Redfield, C.) Plenum Press, New York, NY, pp. 209–271.

Stawarz, B., Genest, M. and Genest, D. (1992) *Biopolymers*, **32**, 633–642.

Torda, A.E., Scheek, R.M. and Van Gunsteren, W.F. (1989) *Chem. Phys. Lett.*, **157**, 289–294.

Torda, A.E., Scheek, R.M. and Van Gunsteren, W.F. (1990) *J. Mol. Biol.*, **214**, 223–235.

Van de Ven, F.J.M., Blommers, M.J.J., Schouten, R.E. and Hilbers, C.W. (1991) *J. Magn. Reson.*, **94**, 140–151.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.

Yip, P. and Case, D.A. (1989) *J. Magn. Reson.*, **83**, 643–648.