

# Influence of Internal Dynamics on Accuracy of Protein NMR Structures: Derivation of Realistic Model Distance Data from a Long Molecular Dynamics Trajectory

Thomas R. Schneider<sup>1</sup>, Axel T. Brünger<sup>2</sup> and Michael Nilges<sup>1\*</sup>

<sup>1</sup>Universität Göttingen, Institut für Anorganische Chemie  
Tammannstr. 4  
D-37077 Göttingen, Germany

<sup>2</sup>Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University  
New Haven, CT 06520, USA

<sup>3</sup>Structural Biology Programme, EMBL  
Meyerhofstr. 1, D-69117 Heidelberg, Germany

In order to study the effect of internal dynamics on the accuracy of NMR structures in detail, we generated NOE distance data from a long molecular dynamics trajectory of BPTI. Cross-relaxation rates were calculated from the trajectory by analysis of the appropriate proton-proton vector autocorrelation functions. A criterion for the convergence of correlation functions was developed, and the analysis was restricted to those correlation functions that had converged within the simulation time. Effective distances were determined from the calculated cross-relaxation rates. Internal dynamics affected the derived distances in a realistic way, since they were subject both to radial averaging (which increases the cross-relaxation rate) and angular averaging (which decreases the cross-relaxation rate). The comparison of the effective distances with average distance between the protons during the trajectory showed that for most the effects of angular and distance averaging essentially cancel out. For these distances, the effective distance derived from an NOE is therefore a very good estimate of the average distance, or the distance in the average structure. However, for about 10% of the distances, the effective distance was more than 10% larger than the average distance, while for about 5%, it was more than 10% smaller, in some cases by more than 2 Å. Little correlation is observed between the effects on cross-relaxation rates to different protons of the same residue. The results of this analysis have implications for the way structures are calculated from NOE distance data. For many distances, the assumption of a rigid structure is valid, and large error bounds would result in the loss of too much information content. On the other hand, the error bounds very often employed are not wide enough for some of the effects seen in our study.

© 1999 Academic Press

**Keywords:** NMR; solution structure; NOE; molecular dynamics; structure accuracy

\*Corresponding author

## Introduction

NOE cross-relaxation rates are sensitive to both molecular structure and dynamics. Models fitted to NOE data should therefore ideally represent the structure and the dynamics of the molecule. Internal dynamics of proteins has proven one of the more difficult problems in the interpretation of

inter-proton NOEs in terms of distances, and has been the focus of several studies using molecular dynamics (MD) (Olejniczak *et al.*, 1984; LeMaster *et al.*, 1988; Post, 1992; Brüschweiler *et al.*, 1992; Palmer & Case, 1992; Abseher *et al.*, 1995). Calculations of NMR relaxation parameters from MD simulations and comparison with NMR experiment have a long history (e.g. see Lipari *et al.*, 1982; Chandrasekhar *et al.*, 1992; Palmer, 1997; Philippopoulos *et al.*, 1997). In short, internal dynamics may lead to an under- or over-estimation of the inter-proton distance, depending on the predominance of radial or angular averaging, respectively. In deriving three-dimensional structures

Abbreviations used: NOE, nuclear Overhauser effect; MD, molecular dynamics; BPTI, bovine pancreatic trypsin inhibitor; RMS, root-mean-square.

E-mail address of the corresponding author: nilges@embl-heidelberg.de

from NOE distance data, an underestimation of distances is more problematic, since it can lead to mutually inconsistent distance restraints. Hence, internal dynamics can be viewed as a source of noise in the data. In most cases, the measured cross-relaxation rates are therefore translated into estimates for distances between atoms, and all sources of imprecision in the data are treated by generously set error bounds (see Nilges & O'Donoghue, 1998, for a recent review). Approximately, structure ensembles calculated with generous error bounds may represent the dynamic behaviour of the molecule well (Brünger, 1997; Abseher *et al.*, 1998). Several model studies investigated the effect of noise on the derived NMR structures (e.g. Clore *et al.*, 1993; Zhao & Jardetzky, 1994).

Various methods have been suggested for directly treating internal dynamics in the structure refinement (Kim & Prestegard, 1989; Koning *et al.*, 1990; Torda *et al.*, 1989; Pearlman & Kollman, 1991; Withka *et al.*, 1992; van Gunsteren *et al.*, 1994; Kemmink & Scheek, 1995; Bonvin & Brünger, 1995a). While these studies established the potential power of fitting averages over structure ensembles or trajectories to the data, they have either not been rigorously tested in a realistic model system, or the model studies contained tacit assumptions about the system or the data quality that make it difficult to judge the general implications. It is noteworthy that a decade after the introduction of the first of these methods, the large majority of structures is still determined with standard distance restraints. It is one aim of the present study to construct a realistic and consistent model system that we will use to develop and rigorously check refinement algorithms, similar in spirit to a study performed more than ten years ago in X-ray crystallography (Kuriyan *et al.*, 1986). MD is the method of choice to generate model data, since NMR parameters can be directly calculated with few approximations from the time course of atomic positions generated by an MD simulation. MD has been used in some studies to derive model data in a more qualitative way (Pearlman & Kollman, 1991; Bonvin & Brünger, 1995a). The MD trajectory provides the ultimate reference "structure" to which the result of a structure determination, employing the simulated NOE data, can be compared.

We studied a small stable protein (Bovine Pancreatic Trypsin Inhibitor, BPTI) for which an MD trajectory *in vacuo* with an implicit solvent model can produce satisfactory results (Loncharich & Brooks, 1989; Steinbach & Brooks, 1994; van Aalten *et al.*, 1995; Abseher *et al.*, 1998). We chose a simulation *in vacuo*, since for our study the length of the trajectory was more important than to use a truly state-of-the-art simulation method. NOEs were extracted from the trajectory by calculating spectral densities from vector autocorrelation functions. This introduces only minimal approximations; in particular, no type of averaging needs to be explicitly

assumed. For many inter-proton vectors, the dynamics is too slow to be sampled properly even in a 6.6 ns trajectory. We developed a heuristic criterion to determine a convergence length for each correlation function. The criterion limited the maximum internal correlation time to roughly 10% of the simulation time. This criterion also identified almost half of the correlation functions as non-converged. One can estimate a cross-relaxation rate for the non-converged correlation functions by assuming slow dynamics with  $\langle r^{-6} \rangle^{-1/6}$  averages, to obtain an estimate for every single cross-relaxation rate in the molecule (Chaloux *et al.*, 1998). However, here, we restricted the analysis to the converged correlation functions.

Here we describe the MD trajectory and give a detailed analysis of the distances extracted from the correlation function analysis. We show that the calculated cross-relaxation rates are self-consistent in the framework of the model-free approach put forward by Lipari & Szabo (1982). For many proton pairs we found that the distance derived from the cross-relaxation rate is close to the average distance; that is, distance and angular averaging effects approximately cancel out (e.g. LeMaster *et al.*, 1988). Around 14% of the distances, however, deviate by more than 10%, some by more than 2 Å. We discuss simple approaches to correct the extracted distances for internal dynamics.

## Theory

The theory relating relaxation rates accessible by NOE experiments to correlation functions describing molecular motions has been reviewed in several publications (LeMaster *et al.*, 1988; Brüschweiler *et al.*, 1992; Post, 1992). We follow closely Brüschweiler *et al.* (1992); we briefly present some equations in the following, mostly to explain where our calculation method deviated from previous papers, and to introduce the notation.

The main quantity derived in 2D NOE experiments is the cross-relaxation rate  $\sigma_{ij}$  describing the rate at which magnetization is transferred between spins  $i$  and  $j$  via dipolar coupling:

$$\sigma_{ij} = \frac{\pi}{5} \gamma^4 \hbar^2 [6J_{ij}(2\omega) - J_{ij}(0)] \quad (1)$$

with the Larmor frequency  $\omega$  and the gyromagnetic ratio  $\gamma$  of protons. The spectral densities  $J_{ij}(\omega)$  characterize the modulation of dipolar coupling between nuclei with time, caused by the fluctuations of the internuclear vector  $\vec{r}_{ij}$  relative to the external magnetic field. The spectral densities  $J_{ij}(\omega)$  are given as Fourier cosine transforms of correlation functions  $C_{ij}(t)$ :

$$J_{ij}(\omega) = 2 \int_0^{\infty} C_{ij}(t) \cos(\omega t) dt \quad (2)$$

Assuming that internal motions are uncorrelated with overall tumbling, and that overall tumbling is isotropic, the correlation function can be comple-

tely factorized into contributions arising either from overall tumbling ( $C_O(t)$ ) or from internal motions ( $C_I(t)$ ) (Hubbard, 1970; Tropp, 1980).  $C_O(t)$  is a simple exponential decay, and  $C_I(t)$  can be simplified with help of the addition theorem for spherical harmonics (Wallach, 1967):

$$C_{ij}(t) = \underbrace{\frac{1}{4\pi} e^{-t/\tau_R}}_{C_O(t)} \underbrace{\left\langle \frac{P_2(\hat{\mu}_D(0) \cdot \hat{\mu}_D(t))}{r_{ij}^3(t)r_{ij}^3(0)} \right\rangle}_{C_I(t)} \quad (3)$$

$\hat{\mu}_D(t)$  is the unit vector in the direction of the inter-proton vector, measured in the diffusion frame rigidly attached to the protein.  $P_2$  is the second order Legendre polynomial  $P_2(x) = (3x^2 - 1)/2$ .

In the framework of the “model free approach” (Lipari & Szabo, 1982) the internal correlation function is approximated by a three-parameter model (two parameters in the case of constant distances) involving a single exponential that decays to a constant value:

$$C_I^{LS}(t) = \langle r^{-6} \rangle (S^2 + (1 - S^2)e^{-t/\tau_e}) \quad (4)$$

where the order parameter  $S^2$  ( $0 \leq S^2 \leq 1$ ) describes the degree of order in the motion of the inter-proton vector, and  $\tau_e$  is a time constant describing the effective decay of the correlation function. The normalization constant can be chosen arbitrarily and is here set to  $\langle r^{-6} \rangle$  for convenience. This model together with the assumption of isotropic rotational motion gives a four-parameter model for the overall correlation function:

$$C^{LS}(t) = C_O(t)C_I^{LS}(t) = \frac{1}{4\pi} e^{-t/\tau_R} \langle r^{-6} \rangle (S^2 + (1 - S^2)e^{-t/\tau_e}) \quad (5)$$

The spectral density for a correlation function of this type is:

$$J^{LS}(\omega) = \frac{1}{2\pi} \langle r^{-6} \rangle \left( S^2 \frac{\tau_R}{1 + \omega^2 \tau_R^2} + (1 - S^2) \frac{\tau_{tot}}{1 + \omega^2 \tau_{tot}^2} \right) \quad (6)$$

where

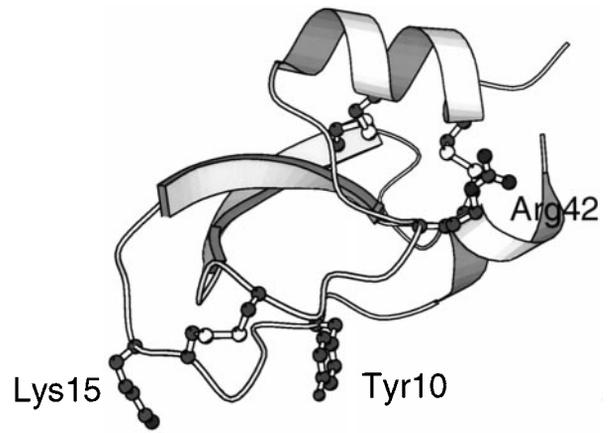
$$\frac{1}{\tau_{tot}} = \frac{1}{\tau_O} + \frac{1}{\tau_e} \quad (7)$$

The angled brackets in the equations denote ensemble averages. Here, we mark quantities calculated from time-averages over the MD trajectory by  $\sim$  to distinguish them from the theoretical quantities.

## Results and Discussion

### Molecular dynamics simulation

The overall fold of BPTI as found in all three crystal forms (I, Marquart *et al.*, 1983; II, Wlodawer *et al.*, 1984; III, Wlodawer *et al.*, 1987) and in the solution structure (Berndt *et al.*, 1993) is shown in



**Figure 1.** Molscript plot (Kraulis, 1991) of BPTI. The disulphide bridges and a few residues discussed in the text are indicated.

Figure 1. The N-terminal  $3_{10}$  helix extending from residue 2 to 7 is followed by a stretch of random coil comprising residues 8 to 17 containing the “active site” residue Lys15 (Figure 2). Residues 18 to 24 and 29 to 35 form the two strands of a  $\beta$ -hairpin. The part of polypeptide backbone connecting the  $\beta$ -hairpin with the C-terminal  $\alpha$ -helix (residues 48 to 55) is attached to the the first strand of the  $\beta$ -hairpin by an isolated  $\beta$ -bridge between Phe45N and Tyr21O. The whole structure is reinforced by three disulphide bridges (5-55, 14-38 and 30-51). The three C-terminal residues Gly-Gly-Ala have been found in a variety of conformations in different structural studies (Berndt *et al.*, 1993; Wlodawer *et al.*, 1987).

The root mean square (RMS) deviation from the starting structure for all non-hydrogen atoms reaches a constant value of ca. 2.9 Å after 2 ns, a value only a little larger than in a previous much shorter *in vacuo* simulation of BPTI (Levitt & Sharon, 1988). The mean deviation for N-C $\alpha$ -C atoms plateaus at ca. 1.6 Å (see Figure 3).

To assess whether the protein remains intact during the simulation the secondary structure was checked at regular time intervals using the program PROCHECK (Laskowski *et al.*, 1993; Figure 4). The secondary structure present in the starting structure is mostly preserved during the entire simulation. Apart from breakage and reformation of hydrogen bonds involving residues on the edges of secondary structure elements and fluctuations in the C-terminal helix three main events were observed. The N-terminal  $\alpha$ -helix vanishes during the equilibration phase and is immediately restored within the first nanosecond of the simulation. At  $\approx 3$  ns, one turn of  $3_{10}$  helix is formed in the turn region of the  $\beta$ -hairpin for a short period of time. This region shows substantial differences in different crystal forms (Wlodawer *et al.*, 1987) and has been recognized as flexible in normal mode calculations (Brüschweiler, 1992). At  $\approx 4$  ns,

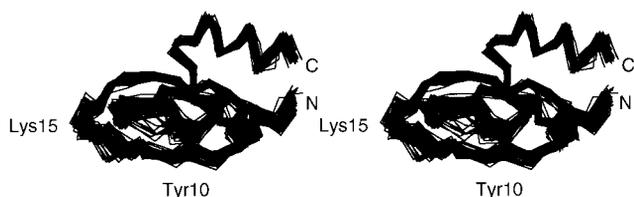


Figure 2. Snapshots of the trajectory every 100 ps.

the isolated  $\beta$ -bridge at residue Phe45 expands by two more hydrogen bonds forming a short piece of  $\beta$ -strand adjacent to the first strand of the  $\beta$ -hairpin. Around 5 ns residues Phe33 and Val34 bulge out of the surface of the protein, interrupting both strands of the  $\beta$ -hairpin for a short period of time.

The RMS fluctuations for  $C^\alpha$  atoms around the average structure of the simulation are consistently larger than the average RMSD around the average structure found for the 20 conformers used to represent the NMR solution structure (Berndt *et al.*, 1993), with the exception of N terminus and C terminus, which are better defined in the simulation than in the NMR-structure (Figure 5). The locations of increased motion show qualitative agreement. Pro9 and Tyr10 are solvent-exposed and located in a region where the mean NMR structure and the crystal structures exhibit substantial differences (Berndt *et al.*, 1993). The large RMS fluctuation of Phe33, Val34, Tyr35 is due to a bulging of the region away from the protein surface and probably is an artifact due to the missing solvent.

### Rotational diffusion

Three separate correlation functions for overall tumbling were calculated for  $x$ ,  $y$ , and  $z$  directions. Inspection of the three correlation functions for rotational diffusion on the interval  $t \in [0, 1000 \text{ ps}]$ ,

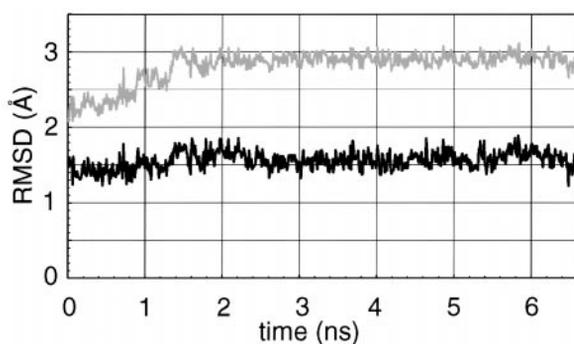


Figure 3. Time course of the RMS deviations relative to the crystal structure 4PTI. The lower and upper curves show the RMS deviations for the backbone (N- $C^\alpha$ -C) atoms and all non-hydrogen atoms of the protein.

and the characteristic times for rotation (Figure 6) showed an approximate isotropy of motion. A mean correlation function was calculated by averaging.

The mean value of the characteristic times corresponds directly to the correlation time for rotational diffusion determined experimentally by NMR methods. Based on the experimentally determined value of  $2.0(\pm 0.5)$  ns obtained at  $36^\circ\text{C}$  (Szyperski *et al.*, 1993) the value of the rotational diffusion time at 300 K can be calculated to be  $2.5(\pm 0.6)$  ns, if one takes into account the different viscosity of water at different temperatures (Cantor & Schimmel, 1980; Weast & Astle, 1983). While the mean correlation time of  $\approx 1.4$  ns observed in the trajectory is smaller than the experimental value, similar behaviour has been observed in other studies using MD in solvent (Ahlström *et al.*, 1989; Brüschweiler *et al.*, 1992; Smith & van Gunsteren, 1994), and in our case may be partly explained by the missing hydration shell. Solvation effects were included in our simulation *via* the friction and sto-

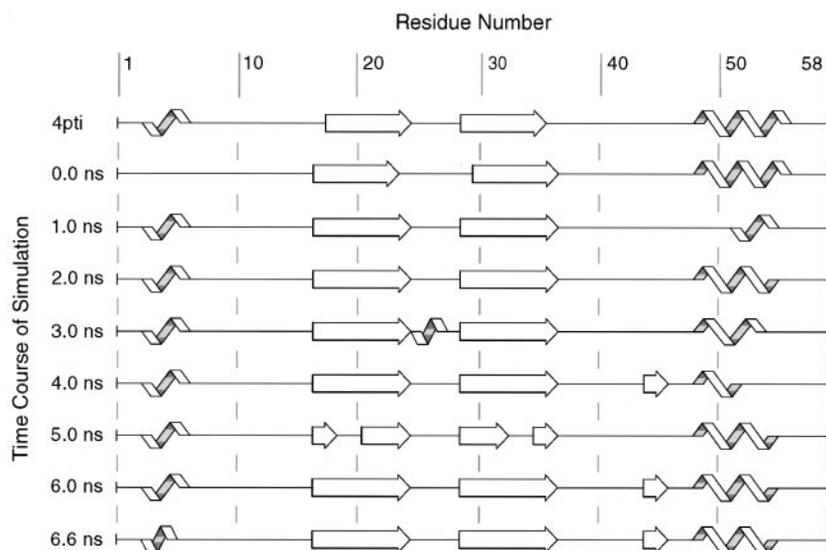
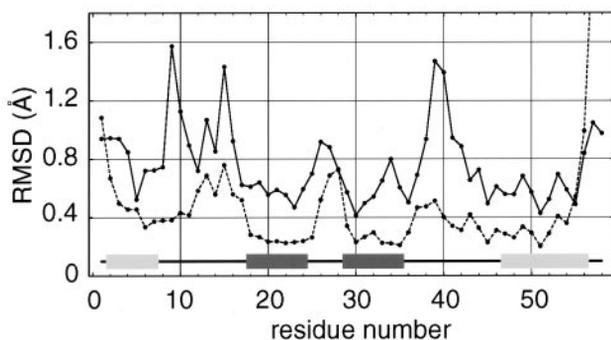


Figure 4. Snapshots of secondary structure (Kabsch & Sander, 1983) of the crystal structure 4PTI and of the structure at different stages of the molecular dynamics simulation. Plots were obtained with PROCHECK (Laskowski *et al.*, 1993).



**Figure 5.** RMS fluctuation of  $C^\alpha$  atoms of BPTI relative to the mean structure of the simulation as a function of the amino acid sequence (continuous line), and average of the global backbone displacements of the 20 energy minimized conformers relative to the mean NMR structure (broken line).  $\beta$ -Strands are marked in light grey,  $\alpha$ -helices are marked in dark grey.

chastic force terms in the Langevin equation, which mimic mechanic effects of an explicit solvent (see Methods). Experimentally determined correlation times for rotational diffusion vary strongly with concentration (Wüthrich & Baumann, 1976; Richarz *et al.*, 1980). The assumption of infinite dilution in the MD trajectories will therefore also contribute to the observed discrepancy.

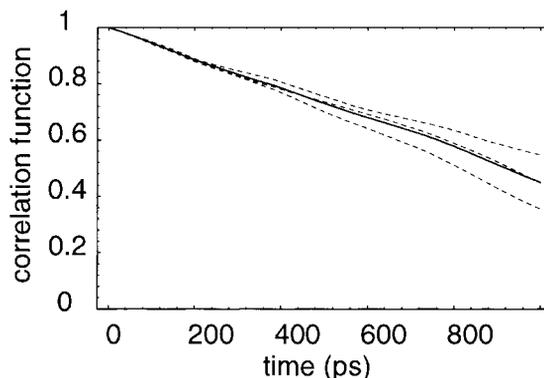
The calculated correlation function for rotational diffusion was only employed to extract the internal part of the total correlation function. For calculation of cross-relaxation rates the experimental value of  $\tau_R = 2.5$  ns was used.

### Selection of proton pairs

To guarantee that all proton pairs that give an NOE-derived distance of less than 4.5 Å were included, the analysis comprised 2779 non-methyl proton pairs satisfying the condition  $\langle r^{-6} \rangle^{-1/6} \leq 4.5$  Å, a cutoff close to the maximum distance commonly observed in NOE experiments. Interactions involving methyl groups were taken into account if at least one of the methyl hydrogen atoms fulfilled the criterion given above. Each methyl group was then represented by a pseudo-atom located at the centre of mass of its hydrogen atoms. This gave an additional number of 367 NOE pairs, increasing the overall number of considered interactions to 3146.

### Convergence of internal correlation functions

Despite the length of the trajectory, there are types of motions which were not sufficiently explored to lead to a reliable estimate of a correlation function. A convergence time  $t_{\text{conv}}$  (the maximum time for which a correlation function was reliable) was estimated by comparison of two correlation functions calculated from the entire trajec-



**Figure 6.** Correlation functions for rotational diffusion. Broken lines, three vectors rigidly attached to the protein. Continuous line, mean correlation function. The characteristic times (equation (13)) are  $\tau_x = 1127$ ,  $\tau_y = 1393$ ,  $\tau_z = 1711$ ,  $\langle \tau \rangle = 1355$ .

tory and a sub-trajectory (see Methods). All correlation functions with  $t_{\text{conv}} \leq 10$  ps were discarded. This restricted further analysis to 1686 correlation functions.

For analysis of the convergence behaviour, the complete set of inter-proton vectors was divided into nine groups. Proton pairs with fixed internuclear distance were collected into one group, pairs with varying distance were categorized into intra- and inter-residue. The latter groups were further subdivided depending on the local environment of the protons involved: “bb-bb”, both protons bound to the polypeptide backbone; “bb-sc”, one proton bound to the backbone, the other belonging to a side-chain; “sc-sc”, both protons belonging to a side-chain; “met” stands for pairs where at least one partner belongs to a methyl group.

All correlation functions calculated for proton pairs with a fixed internuclear distance (geminal protons and protons on aromatic rings) fulfilled the convergence criterion. For all but two (see Figure 7) intra-residue backbone-backbone correlation functions, sufficient sampling is achieved. A crankshaft motion around the peptide plane between Cys14 and Lys15 at around 4.7 ns changes torsion angles all the way to Ala16, giving rise to a drastic change in the dynamics of the inter-proton vectors within this region. This change is reflected in the correlation functions and in the case of 15-h-15-ha leads to insufficient sampling. The second proton pair, 42-h-42-ha, is also situated in a region of the protein not stabilized by secondary structure. As for the region around Lys15, two distinct conformations are occupied during the trajectory, with a transition at about 1 ns. Although the mean behaviour of the inter-proton vector is clearly dominated by the situation corresponding to the second conformation, the correlation function is still to a significant extent influenced by the first. This gives rise to a large estimated error. In principle such situations can be handled by treating

**Table 1.** Statistics for convergence time  $t_{\text{conv}}$  for different classes of inter-proton vectors

	$n_{\text{all}}$	$n_{\text{conv}}$	%conv	Mean	$\sigma$	Min	Max
Fixed dist.	116	116	100.0	550	247	14	999
bb-bb-intra	59	57	96.6	678	138	295	999
bb-sc-intra	337	224	66.5	586	226	34	999
sc-sc-intra	481	337	70.1	520	261	12	999
met-intra	125	91	72.8	576	260	14	999
bb-bb-inter	284	228	80.3	653	163	167	993
bb-sc-inter	684	325	47.5	568	232	29	999
sc-sc-inter	818	219	26.8	541	288	15	999
met-inter	242	89	36.8	597	246	52	999
All	3146	1686	53.6	573	242	12	999

$n_{\text{all}}$  is the number of inter-proton vectors in the respective group.  $n_{\text{conv}}$  is the number of correlation functions with  $t_{\text{conv}} \geq 10$  ps. %conv is the percentage of correlation functions considered as converged. The Mean, the standard deviation  $\sigma$ , minimum, and maximum of  $t_{\text{conv}}$  are based only on converged correlation functions. The interactions were grouped as follows: fixed, the distance is fixed by the covalent geometry; bb-bb, both protons bound to the polypeptide backbone; bb-sc, one proton bound to the backbone, the other belonging to a side-chain; sc-sc, both protons belonging to a side-chain; met pairs where at least one partner belongs to a methyl group.

both conformers separately (Fadel *et al.*, 1995; Fushman *et al.*, 1994). However, for the number of correlation functions in the present study such a manual intervention was not practical.

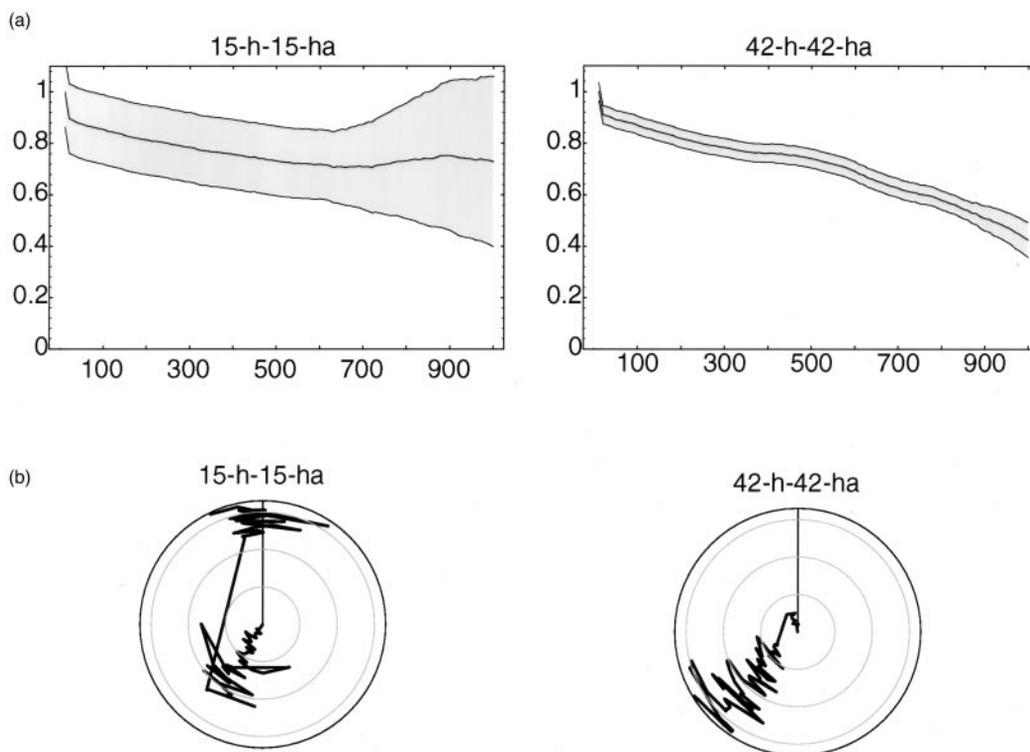
The situation is substantially worse for backbone-side-chain and side-chain-side-chain interactions. Out of the correlation functions describing interactions within the same residue still about 60% are considered as converged reflecting the relatively small number of configurations possible for one side-chain. For interactions between protons on different residues the motion of the inter-proton vector in many cases is too complicated to

be covered by the trajectory in a statistically meaningful way. Only 77.1, 40.2, 15.6 and 28.9% are reliably estimated beyond  $\tau = 10$  ps, thus leaving a large fraction of the calculated correlation functions not meeting the convergence criterion.

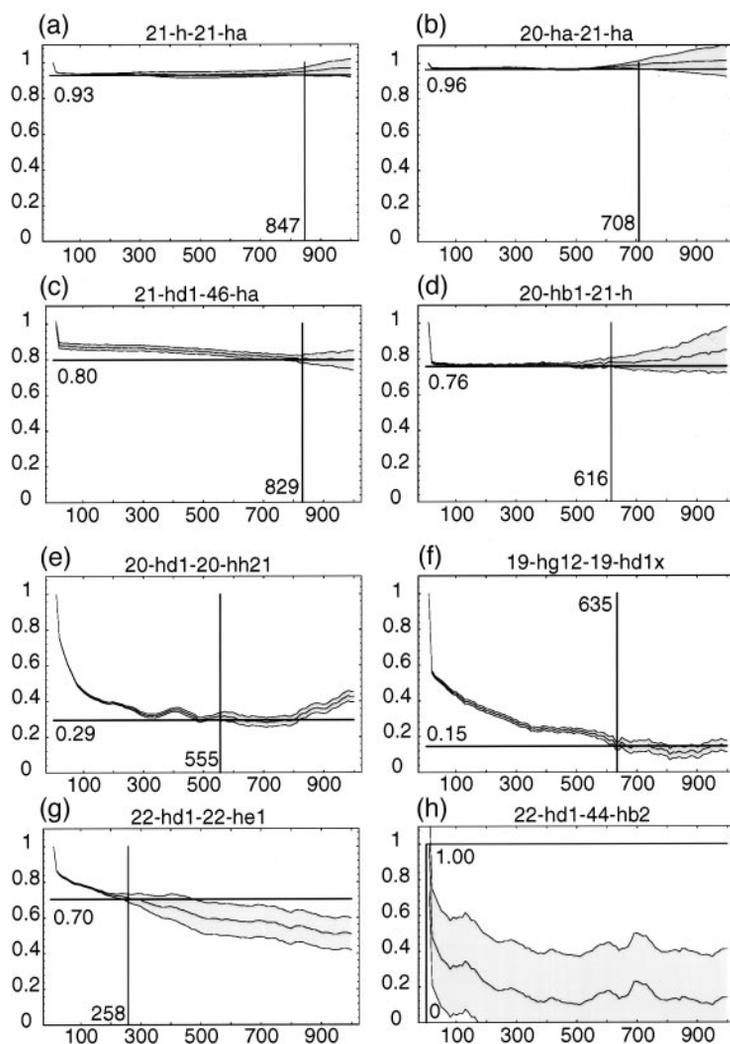
The overall mean for  $t_{\text{conv}}$  for converged correlation functions was 435 ps. Hence, on average the calculated correlation functions were reliable for less than one tenth of the trajectory length.

### Parameters for internal motion

In general the correlation functions exhibit, as found in previous studies (Post, 1992; Brüschweiler



**Figure 7.** (a) The two intra-residue backbone-backbone correlation functions that did not converge. The shaded area corresponds to the estimated error of the correlation function (equation 15). (b) Time-series of the corresponding backbone torsion angle  $\phi$ .



**Figure 8.** Some examples for correlation functions. The horizontal line indicates the extracted order parameter, the number next to it gives its value; the vertical line indicates the convergence time, the number next to it its value. (a) An intra-residue converged correlation function with high order parameter; (b) an inter-residue converged correlation function with high order parameter; (c), (d) inter-residue converged correlation functions with intermediate order parameters; (e), (f) intra-residue correlation functions with slow decay and lower order parameter; (g) short convergence time; (h), non-converged.

*et al.*, 1992), a rapid initial decay on a time-scale of  $\approx 10$ -20 ps immediately reaching the final plateau value or followed by a multi-exponential decay (Figure 8).

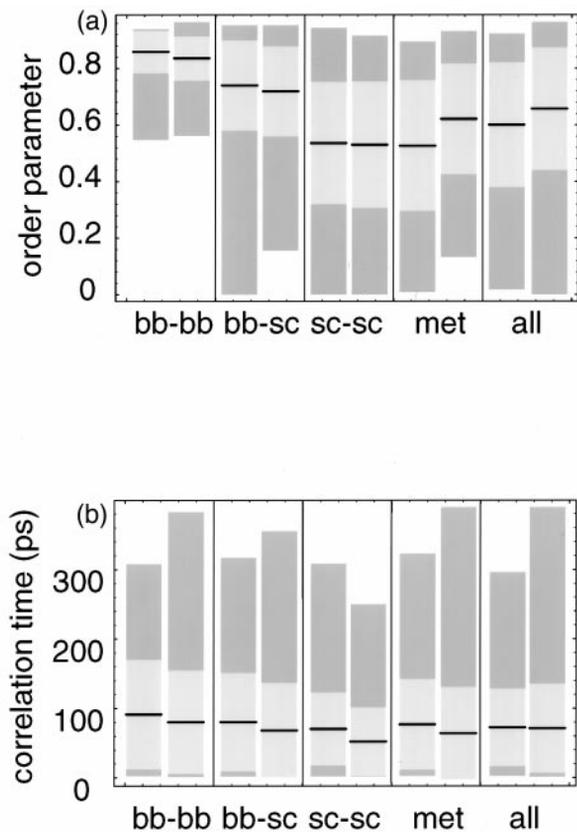
Overall, the order parameters  $\bar{S}^2$  derived from the trajectory vary substantially ( $0.66(\pm 0.22)$ ). Broken down into groups in the same way as for the convergence behaviour, the highest order parameters are found when both protons are rigidly attached to the backbone ( $0.87(\pm 0.06)$  and  $0.85(\pm 0.07)$ ), reflecting the limited flexibility of the backbone (Figure 9). If both protons are located on side-chains, the order parameters are on average substantially lower ( $0.54(\pm 0.22)$ ,  $0.53(\pm 0.22)$ ). Methyl groups show somewhat higher order parameters than other side-chain protons, reflecting their shorter average distance to the backbone (in particular for Ala residues), and that they tend to be buried in the interior of the protein.

We also analysed the order parameter in terms of residue number and restraint class (Figure 10). Somewhat surprisingly, the order parameter does not follow the same pattern as the RMS fluctuation plot throughout the sequence. Thus, while the order parameter tends to be low for the flexible

residue Arg39, it is not significantly reduced for residue Lys15. On the other hand, residues 18 and 19, which are part of the  $\beta$ -sheet, show low order parameters. Interestingly, however, there are values of  $\bar{S}^2$  close to one even for residues with low average order parameter.

The analysis of effective correlation times (Figure 9(b)) shows that the dependency on local environment is weaker than for the order parameters. For all groups the correlation times cover a wide range. While the effective correlation times are not real correlation times and do not strictly have a physical interpretation (it is determined by the area under the correlation function), it is worth remembering that order parameters and correlation times describe fundamentally different physical phenomena. The time scale of the motion is determined by the energetic barriers between conformational substates (e.g. side-chain rotamers), while the order parameters are related to the spatial extension of the motion, i.e. the shape of the energy surface.

We use only two exponential decays to approximate the correlation function. A model employing more basis functions (Lipari & Szabo, 1982; Clore



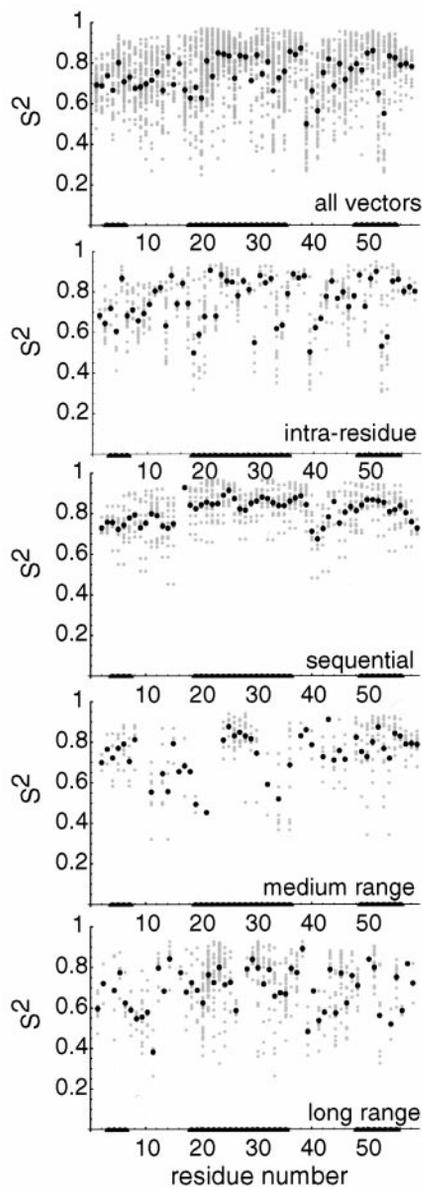
**Figure 9.** Statistics for order parameters  $\tilde{S}^2$  (a) and effective correlation times (b) for different classes of inter-proton vectors. Each bar indicates the mean value, standard deviation, minimum and maximum. The classes are: bb-bb-intra/inter; bb-sc-intra/inter; sc-sc-intra/inter; met-intra/inter; fixed, all; where bb stands for backbone, sc for side-chain, met for methyl group, and fixed for distances fixed by the covalent geometry.

*et al.*, 1990) might allow a more detailed description of the dynamics. However, we show below that the simplest model is entirely sufficient for the purpose of NOE calculation.

### Distance calculation from spectral densities

The MD trajectory allows us to assess the influence of different approximations in deriving distances from the cross-relaxation rates. From the trajectory itself, we can calculate the  $\langle r^{-6} \rangle^{-1/6}$  average and the arithmetic average  $\langle r \rangle$  of the inter-proton distance, and the cross-relaxation rates  $\tilde{\sigma}$ . From the cross-relaxation rates we can determine different distances  $r_{\text{eff},i}$ . Our calculation of the cross-relaxation rates  $\tilde{\sigma}$  made use of numerical integration of the simulated correlation functions up to  $t_{\text{conv}}$  (see Methods). Hence, the  $\tilde{\sigma}$  contain contributions from multi-exponential decay of the correlations.

We first examined how much replacing the correlation functions by the two-parameter Lipari-Szabo model influences the results. In the Lipari-Szabo model given in equation (5), a distance can

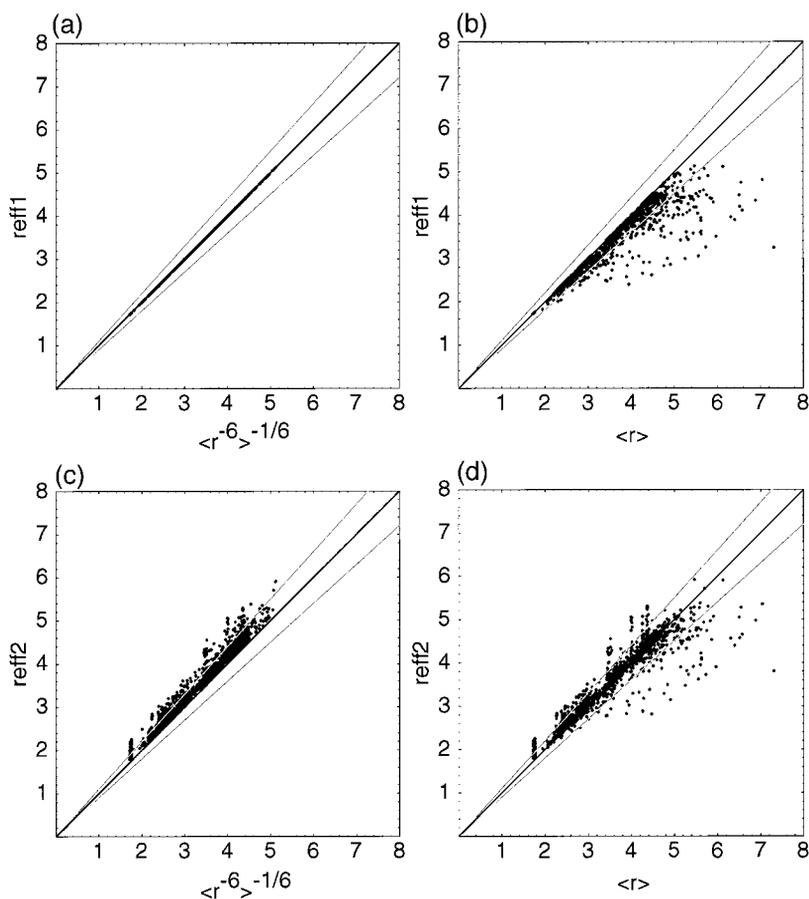


**Figure 10.** All order parameters  $\tilde{S}^2$  along the sequence (grey dots) and average order parameter for each residue (black dots), for different classes of inter-proton vectors. The secondary structure is indicated by black lines in the bottom of the Figure.

be calculated from the cross-relaxation rate  $\tilde{\sigma}_{ij}$ , the order parameter  $\tilde{S}_{ij}^2$  and the correlation times  $\tilde{\tau}_{\text{tot},ij}$  and  $\tau_R$ . This distance,  $r_{\text{eff},1}$  should be close to the  $\langle r^{-6} \rangle^{-1/6}$  distance between the protons  $i$  and  $j$ :

$$r_{\text{eff},1} = \left[ \tilde{\sigma}_{ij}^{-1} \frac{4\pi}{25} \gamma^4 \hbar^2 \left( \tilde{S}_{ij}^2 \tau_R \left( \frac{6}{1 + 4\omega^2 \tau_R^2} - 1 \right) + (1 - \tilde{S}_{ij}^2) \tilde{\tau}_{\text{tot},ij} \left( \frac{6}{1 + 4\omega^2 \tilde{\tau}_{\text{tot},ij}^2} - 1 \right) \right) \right]^{1/6} \quad (8)$$

The consistency between  $r_{\text{eff},1}$  and  $\langle r^{-6} \rangle^{-1/6}$  is excellent (Figure 11(a), Table 2). The multi-exponential behaviour of the correlation functions, which is



**Figure 11.** (a)  $r_{\text{eff},1}$  plotted against  $\langle r^{-6} \rangle^{-1/6}$  and (b)  $\langle r \rangle$  (c)  $r_{\text{eff},2}$  plotted against  $\langle r^{-6} \rangle^{-1/6}$  and (d)  $\langle r \rangle$ .

neglected in the simple approximation with effective correlation time and order parameter, has very little influence on the derived distance. In order to determine structures, it is however more important to obtain reliable estimates for  $\langle r \rangle$ . The deviations between  $r_{\text{eff},1}$  and  $\langle r \rangle$  are much larger, with a tendency to underestimate the distance. (206 out of 1686 (ca. 14%) distances are underestimated by more than 10%.) In an experimental structure determination, it is also unrealistic to assume knowledge of an order parameter and effective correlation time for every NOE. We studied therefore

in detail the simplest and most realistic case where we assume a rigid molecule and set  $S = 1$  for all NOEs. In this case, equation (8) reduces to:

$$r_{\text{eff},2} = \left[ \tilde{\sigma}_{ij}^{-1} \frac{4\pi}{25} \gamma^4 \hbar^2 \left( \frac{6\tau_R}{1 + 4\omega^2\tau_R^2} - 1 \right) \right]^{1/6} \quad (9)$$

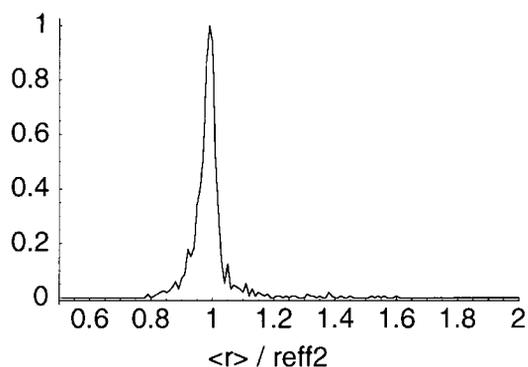
The scatter-plot (Figure 11(d)) shows that there are now under- and overestimations of the distance. The fact that  $S^2$  is always smaller than 1.0 leads to a partial cancellation of the underestimation of the distance by  $\langle r^{-6} \rangle^{-1/6}$  fluctuation, since  $\langle r^{-6} \rangle^{-1/6}$  is always smaller than  $\langle r \rangle$ . The number of under-estimated estimates for  $\langle r \rangle$  is reduced from 206 to 81. While there is no cancellation of effects for every single distance, the cancellation averaged over all NOEs is very good (Figure 12). This is also seen on a residue-by-residue level (Figure 13) if one considers all NOEs, less so for medium range or long range NOEs alone. For the long range NOEs, the systematic under-estimations are found around residue Lys15 and close to Arg39 and Cys55. Large under-estimations for single distances are found all over the sequence.

Our results are in essence in agreement with previous studies (Post, 1992; Brüschweiler *et al.*, 1992). Perceived differences between these studies and an earlier study (LeMaster *et al.*, 1988) are partially

**Table 2.** Number of distances that are under- or over-estimated by more than 10%, when compared to  $\langle r^{-6} \rangle^{-1/6}$  and  $\langle r \rangle$  averages over the trajectory

	$\langle r^{-6} \rangle^{-1/6}$		$\langle r \rangle$	
	Over	Under	Over	Under
$r_{\text{eff},1}$	0	0	0	206
$r_{\text{eff},2}$	284	0	152	81
$r_{\text{eff},3}$	90	0	61	178
$r_{\text{eff},4}$	70	0	47	187

$r_{\text{eff},1}$  was calculated from cross-relaxation rates with exact values for  $\tilde{S}^2$  and  $\tau_e$  for each individual inter-proton vector;  $r_{\text{eff},2}$  with  $\tilde{S}^2 = 1$  for all vectors;  $r_{\text{eff},3}$  with  $\tilde{S}^2$  and  $\tau_e$  set to their respective averages over all inter-proton vectors; and  $r_{\text{eff},4}$  with  $\tilde{S}^2$  and  $\tau_e$  set to separate averages for backbone-backbone, backbone-side-chain and side-chain-side-chain vectors.



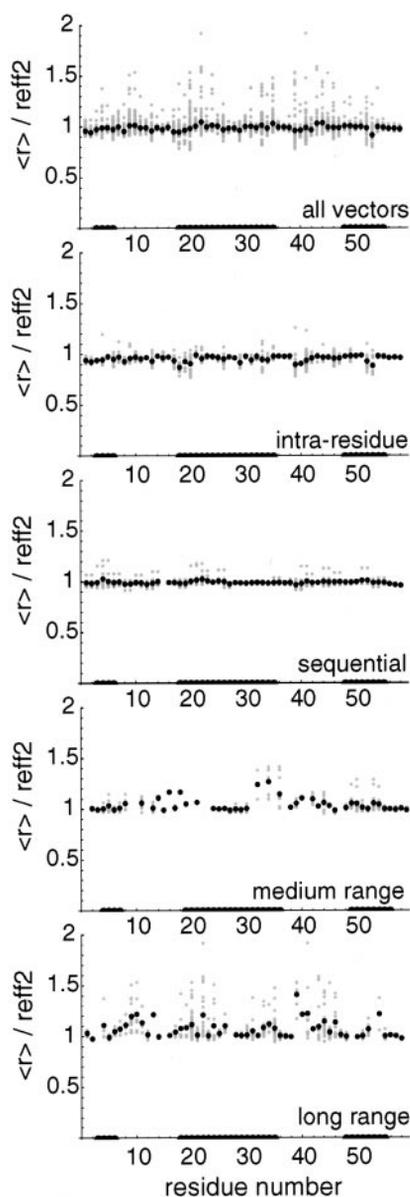
**Figure 12.** Histogram of  $\langle r \rangle / r_{\text{eff},2}$  for all cross-relaxation rates in the analysis.

due to different emphasis given to cancellation effects and underestimation of distances by averaging. With much longer simulation times in our study, and convergence analysis for the correlation functions, we do observe cancellation of effects for the large majority of NOEs. LeMaster *et al.* (1988) had restricted their analysis to conformationally constrained inter-proton vectors, i.e. for vectors for which convergence could be expected.

The convergence analysis is a central point in our study. We cannot say much about NOEs from non-converged correlation functions. Slower dynamics can lead to two possible effects: broadening of peaks and therefore smaller NOEs, or dominance of  $\langle r^{-6} \rangle^{-1/6}$  averaging and therefore larger NOEs. Since absence of NOEs is usually not used directly in structure calculations, the latter pose a more severe problem. In the first use of the data derived here we have therefore assumed the latter (Chaloux *et al.*, 1998).

### Uniform corrections for internal motions

The use of information on dynamics extracted from MD trajectories has been suggested to complement experimental information (Koning *et al.*, 1990; Abseher *et al.*, 1995). To account for the effects of fluctuations in the inter-proton-vectors in a crude approximations, we replaced  $\tilde{S}_{ij}^2$  and  $\tilde{\tau}_{\text{tot},ij}$  by their means over all inter-proton vectors. There is a dramatic improvement for the estimate of  $\langle r^{-6} \rangle^{-1/6}$ . Instead of 284 outliers for the rigid molecule, there are now only 90. However, the estimate of  $\langle r \rangle$  is not improved. While the number of overestimated distances is reduced, the number of underestimated distances increases from 81 to 178. If we apply the averages for order parameters and effective correlation times separately for backbone-backbone, backbone-side-chain and side-chain-side-chain vectors, the further improvement is small for the estimate of  $\langle r^{-6} \rangle^{-1/6}$ , while the number of under-estimations of  $\langle r \rangle$  increases further (Table 2).



**Figure 13.** Ratio of  $\langle r \rangle / r_{\text{eff},2}$  for different classes of NOEs (grey dots) and the average ratio for each residue (black dots). The secondary structure is indicated by black lines in the bottom of the Figure.

## Conclusions

We have presented a detailed NMR analysis of a long MD trajectory and applied a heuristic criterion to check for convergence of the correlation functions. In spite of the length of the trajectory, only around 50% of the correlation functions met the convergence criterion. The maximum time-scale that we could reliably extract from the correlation function was limited to less than 10% of the trajectory length (500 ps for 5.5 ns trajectory). In spite of the application of the convergence criterion, the order parameters are on average lower than in previous studies (Brüschweiler *et al.*, 1992; Post, 1992,

1992; Abseher *et al.*, 1995), probably due to the significantly longer simulation time.

For many NOEs, the effects of radial and angular averaging cancel. However, we find that in 14% of the NOEs the internal dynamics lead to an underestimation of the average distance by more than 10%. This fraction might increase if, with longer simulation time, more correlation functions converge. Our results indicate that the length of the simulation may be more important than the exact treatment of solvent environment, if an implicit solvent model is used and a suitable system is chosen (i.e. a globular protein without long loops).

Our results have implications for the development of NMR structure refinement protocols. For once, we obtained good  $\langle r^{-6} \rangle^{-1/6}$  estimates with a simple overall correction with the average order parameter and effective correlation time. These corrected distances could be used in an ensemble refinement using  $\langle r^{-6} \rangle^{-1/6}$  averaging.

More importantly, good estimates of the (arithmetic) average distance were obtained from most cross-relaxation rates. This suggests that for the majority of distances, tighter error bounds could be employed in a structure calculation than are usually employed in structure calculations. On the other hand, some of the distances would lie outside even these generous bounds. One possible solution to this problem could be an ensemble refinement technique which partitions the data into a "static" set, for which standard distance restraints are applied, and a "dynamic" set, which is used with  $\langle r^{-6} \rangle^{-1/6}$  averaging. This could solve the most difficult problem with ensemble averaging techniques, over-fitting of the data (Brünger *et al.*, 1993; Bonvin & Brünger, 1995a,b).

Our results underline again that, in contrast to spin diffusion, it is difficult to imagine a correction for the effects of internal dynamics in closed analytical form. One of the primary aims of this paper was to derive and characterize a model data-set from an MD trajectory. We will use this data-set to develop and test refinement protocols, and the data-set will be made available (<http://www.nmr.embl-heidelberg.de>).

## Methods

### Molecular dynamics simulation

A molecular dynamics simulation of 6.6 ns was performed using X-PLOR (Brünger, 1992) employing the CHARMM extended atom force field PARAM19 (Brooks *et al.*, 1983), which treats only polar hydrogen atoms explicitly. The initial set of atomic coordinates was obtained from the crystal form I structure of BPTI at 1.5 Å resolution (Marquart *et al.*, 1983) as deposited in the protein data bank (Bernstein *et al.*, 1977). Polar hydrogen atoms were built on the crystal structure coordinates (Brünger & Karplus, 1988). The structure was then minimized for 100 steps. This was followed by 50 ps of equilibration during which the C $\alpha$  coordinates were restrained to their initial positions (Bruccoleri &

Karplus, 1986). The velocities were reassigned to a Maxwell distribution at 300 K every 10 ps. Only side-chains that were more than 50% solvent-accessible were treated with Langevin dynamics with a friction coefficient of 20 ps $^{-1}$ . The interior of the protein was simulated by pure Newton dynamics without any temperature coupling. A distance-dependent dielectric constant  $\epsilon = R$  was employed throughout. Charges on Lys, Arg, Glu, and Asp side-chains were scaled by a factor of 0.3. A switching function for the non-bonded energy terms was applied between 5 and 9 Å (Loncharich & Brooks, 1989). The non-bonded cutoff was set to 9.5 Å. The non-bonded list was updated whenever an atom had moved by more than 0.25 Å. Bond lengths were kept rigid during the simulation by use of the SHAKE-method (Ryckaert *et al.*, 1977), and the integration step was 2.0 fs. Complete coordinate sets were written every 0.1 ps. After completion of the dynamics run non-polar hydrogen atoms were generated onto the respective heavy atoms for each coordinate set using the HBUILD routine in X-PLOR (Brünger & Karplus, 1988).

### Calculation of correlation functions and rotational diffusion

By means of the addition theorem for spherical harmonics, the expression for the correlation function for the vector between protons  $i$  and  $j$  (see. e.g. Wallach, 1967) can be written as:

$$C(t) = \frac{1}{5} \left\langle \frac{P_2(\hat{\mu}_{L,ij}(t) \cdot \hat{\mu}_{L,ij}(0))}{r_{ij}^3(t)r_{ij}^3(0)} \right\rangle \quad (10)$$

where  $\hat{\mu}_{L,ij}$  is the Cartesian unit vector in the direction of the inter-proton vector in the laboratory frame, and  $r_{ij}$  is the distance of the two protons. The ensemble average denoted by the angled brackets in the above equation is then evaluated as a time average to obtain the correlation function  $\tilde{C}(t_n)$  evaluated from the trajectory:

$$\tilde{C}(t_n) = \frac{1}{5} \frac{1}{M-n} \sum_{m=1}^{M-n} \frac{P_2(\hat{\mu}_L(t_m) \cdot \hat{\mu}_L(t_{m+n}))}{r_{ij}^3(t_m)r_{ij}^3(t_{m+n})} \quad (11)$$

Due to the length of the simulation, overall tumbling plays a significant role in the calculation of correlation functions. Rather than following standard practice and fitting the entire trajectory to the initial frame of the trajectory, we decided to calculate a correlation function for rotational diffusion and to apply the factorization of equation (3) explicitly. The two methods are entirely equivalent; however, the calculation of a rotational correlation function simplifies the discussion of rotational correlation times (cf. the discussion about the rotational diffusion time by Brüschweiler *et al.*, 1992). In complete analogy with the correlation function for an inter-proton vector, the correlation function for rotational diffusion is written as:

$$C_O(t) = \frac{1}{5} \langle P_2(\hat{\mu}_{LD}(0) \cdot \hat{\mu}_{LD}(t)) \rangle \quad (12)$$

where  $\hat{\mu}_{LD}$  is now a unit vector rigidly attached to the protein, and  $M$  is the total number of trajectory frames. A set of vectors rigidly attached to the protein can be defined by the  $x$ ,  $y$ ,  $z$  unit vectors at the end of the equilibration period ( $t = t_{\text{ref}}$ ). The time series for this set of vectors are accumulated by mass-weighted best fits between the backbone atoms of each coordinate set

$\vec{r}(t)$  and the reference coordinate set  $\vec{r}(t_{\text{ref}})$ . Characteristic times for rotational diffusion were calculated *via* (Ahlström *et al.*, 1989):

$$\tau_c = \frac{1}{M-10} \sum_{m=1}^{M-10} \frac{10\text{ps}}{(\ln(\tilde{C}_O(t_m))) - (\ln(\tilde{C}_O(t_m+10)))} \quad (13)$$

We can then obtain the internal correlation function  $C_I(\tau)$  directly by exploiting the factorization of the total correlation function (equation (3)) into contributions from internal motions and overall tumbling:

$$C_I(\tau) = \frac{C(\tau)}{C_O(\tau)} \approx \frac{\tilde{C}(\tau_n)}{\tilde{C}_O(\tau_n)} = \tilde{C}_I(\tau_n) \quad (14)$$

### Ergodic convergence of correlation functions

Throughout this study ensemble averages are approximated by time averages over discrete phase points. For the case of correlation functions the error in the estimate obtained by finite time averaging has been shown to be proportional to  $\sqrt{\tau_c/T}$  (Zwanzig & Ailawadi, 1969), where  $\tau_c$  is the correlation time characterizing the correlation function, and  $T$  is the length of the averaging interval. The correlation times for the motion of different inter-proton vectors vary substantially and cannot be determined prior to the calculation of the correlation function and its error. An estimate of the error of the correlation function should give a small error for correlation times  $\tau_c$  small compared to  $T$ , and a large error for  $\tau_c$  large comparable to  $T$ . To estimate the statistical error, two internal correlation functions were calculated for each inter-proton vector. The first evaluation was based on the full 6.6 ns simulation ( $\tilde{C}_{I,lg}(t)$ ), the second used phase points limited to a 10% shorter time interval ( $\tilde{C}_{I,sh}(t)$ ). This allows estimation of the error for the internal correlation at each time point  $t$  *via*:

$$\tilde{\epsilon}(t) = \text{Max}(|\tilde{C}_{I,lg}(t') - \tilde{C}_{I,sh}(t')|)t' \epsilon[0, t] \quad (15)$$

The time point at which the estimated error  $\tilde{\epsilon}$  exceeded 2.5% was defined as the convergence time  $t_{\text{conv}}$  of the correlation function, and values for  $t > t_{\text{conv}}$  were discarded. The value for the maximally tolerated error was determined empirically, and was found to reliably exclude parts of the correlation functions that showed a marked increase after the initial decay. An increase in the correlation function (often to values larger than one) is unphysical, and entirely due to insufficient sampling. Correlation functions with  $t_{\text{conv}}$  of less than 10 ps were marked as “not converged” and excluded from further analysis.

A better estimate of the error of the correlation function would be the standard deviation of each point in the correlation function. This can be readily calculated with the summation method. We compared the standard deviation and our criterion for some correlation functions and found them to be qualitatively similar. The summation method was however too slow for a calculation of all correlation functions and their standard deviations with a trajectory of several ns.

### Calculation of order parameters and correlation times

Due to its limited length, the trajectory cannot cover all possible motions of the inter-proton vectors. We expect that a plateau value reached by a calculated corre-

lation function represents an upper limit for the real value of the generalized order parameter. In general, the estimate from a trajectory of limited length could also be too low, in case there are rare motions present which are not sufficiently sampled in the trajectory. However, correlation functions with contributions from rare motions were excluded by our convergence criterion.  $\tilde{S}^2$  was therefore set to the minimum of the correlation function on the converged interval  $[0, t_{\text{conv}}]$ . We preferred this approach to averaging over the whole correlation function (e.g. Chandrasekhar *et al.*, 1992), since the value of  $\tilde{S}^2$  is not influenced by parts of the correlation function with large statistical uncertainties. An effective correlation time  $\tau_c$  was calculated *via* (Lipari & Szabo, 1982):

$$\tilde{\tau}_{c,ij} = (1 - \tilde{S}_{ij}^2)^{-1} \sum_{i=1}^{t_{\text{conv}}} \tilde{C}_{ij}(t_i) - \tilde{S}_{ij}^2 \quad (16)$$

### Calculation of spectral densities

For evaluation of spectral densities internal correlation functions were prolonged to infinity using the plateau value  $\tilde{S}^2$ :

$$\tilde{C}_I^*(t_n) = \begin{cases} \tilde{C}_I(t_n) & \text{for } t_n < lt_{\text{conv}} \\ \tilde{S}^2 & \text{for } t_n > gt_{\text{conv}} \end{cases} \quad (17)$$

Assuming a simple exponential decay with a correlation time  $\tau_R$  for the rotational correlation function  $C_O(t)$  a total correlation function  $\tilde{C}^*(t_n)$  which is defined from  $t = 0$  to  $t = \infty$  can then be constructed:

$$\begin{aligned} \tilde{C}^*(t_n) &= \tilde{C}_I^*(t_n)C_O(t_n) \\ &= \frac{1}{5} \tilde{C}_I^*(t_n)e^{-t/\tau_R} \end{aligned} \quad (18)$$

The spectral densities  $\tilde{J}(\omega)$  at given frequency can be calculated from this combined function *via* Fourier cosine transformation:

$$\begin{aligned} \tilde{J}(\omega) &= 2 \int_0^\infty C^*(t') \cos(\omega t') dt' \\ &= \frac{2}{5} \int_0^{t_{\text{conv}}} \tilde{C}_I(t') e^{-t'/\tau_R} \cos(\omega t') dt' \\ &\quad + \frac{2}{5} \int_{t_{\text{conv}}}^\infty \tilde{S}^2 e^{-t'/\tau_R}(\tau_R) \cos(\omega t') dt' \end{aligned} \quad (19)$$

The first integral can be evaluated by a summation employing the calculated values for the internal correlation function. The second integral can be evaluated analytically:

$$\begin{aligned} \tilde{J}(\omega) &= \frac{2}{5} \Delta t \sum_{n=1}^{n_{\text{conv}}} \tilde{C}_I(t_n) e^{-t_n/\tau_R} \cos(\omega t_n) \\ &\quad + \frac{2}{5} \frac{\tilde{S}^2 e^{-t_{\text{conv}}/\tau_R}}{t_{\text{conv}}^{-2} + \omega^2} \{t_{\text{conv}}^{-1} \cos(\omega t_{\text{conv}}) - \omega \sin(\omega t_{\text{conv}})\}. \end{aligned} \quad (21)$$

Note that the numerical evaluation of the first integral does not make use of the above defined effective correlation time and therefore includes effects corresponding to a non-mono-exponential decay of the correlation function in the spectral density.

Cross-relaxation rates are now accessible *via* equation (1).

## Acknowledgements

This work was in part supported by an EMBL predoctoral fellowship to T.R.S.

## References

- Abseher, R., Lüdemann, S., Schreiber, H. & Steinhauser, O. (1995). NMR cross relaxation investigated by molecular dynamics simulation: a case study of ubiquitin in solution. *J. Mol. Biol.* **249**, 604-624.
- Abseher, R., Horstink, L., Hilbers, C. W. & Nilges, M. (1998). Essential spaces defined by NMR structure ensembles and molecular dynamics simulation show significant overlap. *Proteins: Struct. Funct. Genet.* **31**, 370-382.
- Ahlström, P., Teleman, O., Kördel, J., Forsen, S. & Jönsson, B. (1989). A molecular dynamics simulation of bovine Calbindin  $D_{9k}$ . Molecular structure and dynamics. *Biochemistry*, **28**, 3205-3211.
- Berndt, K., Güntert, P., Orbons, L. & Wüthrich, K. (1993). Determination of a high-quality nuclear magnetic resonance solution structure of the bovine pancreatic trypsin inhibitor and comparison with three crystal structures. *J. Mol. Biol.* **227**, 757-775.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. J., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.
- Bonvin, A. M. J. J. & Brünger, A. T. (1995a). Conformational variability of solution nuclear magnetic resonance structures. *J. Mol. Biol.* **250**, 80-93.
- Bonvin, A. M. J. J. & Brünger, A. T. (1995b). Do NOE distances contain enough distance information to access the relative populations of multi-conformer structures? *J. Biomol. NMR*, **5**, 72-76.
- Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **4**, 187-217.
- Brucoleri, R. E. & Karplus, M. (1986). Spatially constrained minimization of macromolecules. *J. Comput. Chem.* **7**, 165-175.
- Brünger, A. & Karplus, M. (1988). Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison. *Proteins: Struct. Funct. Genet.* **4**, 148-156.
- Brünger, A. T. (1992). *X-PLOR. A System for X-ray Crystallography and NMR*, Yale University Press, New Haven.
- Brünger, A. (1997). X-ray crystallography and NMR reveal complementary views of structure and dynamics. *Nature Struct. Biol.* **4**, S862-S865.
- Brünger, A. T., Clore, G. M., Gronenborn, A. M., Saffrich, R. & Nilges, M. (1993). Assessing the quality of solution nuclear magnetic resonance structures by complete cross-validation. *Science*, **261**, 328-331.
- Brüschweiler, R. (1992). Normal modes and NMR order parameters in proteins. *J. Am. Chem. Soc.* **114**, 5341-5344.
- Brüschweiler, R., Roux, B., Blackledge, M., Griesinger, C., Karplus, M. & Ernst, R. (1992). Influence of rapid intramolecular motion on NMR cross-relaxation rates. A molecular dynamics study of antamanide in solution. *J. Am. Chem. Soc.* **114**, 2289-2302.
- Cantor, C. R. & Schimmel, P. R. (1980). *Biophysical Chemistry*, Freeman, San Francisco.
- Chaloux, F.-R., O'Donoghue, S. I. & Nilges, M. (1998). Molecular dynamics and accuracy of NMR structures: effects of error bounds and data removal. *Proteins: Struct. Funct. Genet.* in the press.
- Chandrasekhar, I., Clore, G., Szabo, A., Gronenborn, A. & Brooks, B. (1992). A 500 ps molecular dynamics simulation study of interleukin 1 $\beta$  in water. *J. Mol. Biol.* **226**, 239-250.
- Clore, G., Szabo, A., Bax, A., Kay, L., Driscoll, P. & Gronenborn, A. (1990). Deviation from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* **112**, 4989-4991.
- Clore, G. M., Robien, M. A. & Gronenborn, A. M. (1993). Exploring the limits of precision and accuracy of protein structures determined by nuclear magnetic resonance spectroscopy [published erratum appears in *J. Mol. Biol.* (1994) Mar 25;237(2):243]. *J. Mol. Biol.* **231**, 82-102.
- Fadel, A. R., Jin, D. Q., Montelione, G. T. & Levy, R. M. (1995). Crankshaft motions of the polypeptide backbone in molecular dynamics simulations of human type-alpha transforming growth factor. *J. Biomol. NMR*, **6**, 221-225.
- Fushman, D., Ohlenschläger, O. & Rüterjans, H. (1994). Determination of the backbone mobility of ribonuclease T1 and its 2'GMP complex using molecular dynamics simulations and NMR relaxation data. *J. Biol. Struct. Dynam.* **11**, 1377-1402.
- Hubbard, P. S. (1970). Nonexponential relaxation of rotating three-spin systems in molecules of a liquid. *J. Chem. Phys.* **52**, 563-568.
- Kabsch, W. & Sander, C. (1983). A solution for the best rotation to relate two sets of vectors. *Biopolymers*, **22**, 2577-2637.
- Kemmink, J. & Scheek, R. M. (1995). Dynamic modelling of a helical peptide in solution using NMR data: multiple conformations and multi-spin effects. *J. Biomol. NMR*, **5**, 33-40.
- Kim, Y. & Prestegard, J. H. (1989). A dynamic model for the structure of acyl carrier protein in solution. *Biochemistry*, **28**, 8792-8797.
- Koning, T. M. G., Boelens, R. & Kaptein, R. (1990). Calculation of the nuclear Overhauser effect and the determination of proton-proton distances in the presence of internal motion. *J. Magn. Reson.* **90**, 111-123.
- Kraulis, P. (1991). Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.
- Kuriyan, J., Petsko, G. A., Levi, R. M. & Karplus, M. (1986). Effect of anisotropy and anharmonicity on protein crystallographic refinement. An evaluation by molecular dynamics. *J. Mol. Biol.* **190**, 227-254.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.
- LeMaster, D. M., Kay, L. E., Brünger, A. T. & Prestegard, J. H. (1988). Protein dynamics and distance determinations by NOE measurement. *FEBS Letters*, **236**, 71-76.
- Levitt, M. & Sharon, R. (1988). Accurate simulation of protein dynamics in solution. *Proc. Natl Acad. Sci. USA*, **85**, 7557-7561.
- Lipari, G. & Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance

- relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546-4558.
- Lipari, G., Szabo, A. & Levy, R. (1982). Protein dynamics and NMR relaxation: comparison of simulations with experiment. *Nature*, **300**, 197-198.
- Loncharich, R. J. & Brooks, B. R. (1989). The effects of truncating long-range forces on protein dynamics. *Proteins: Struct. Funct. Genet.* **6**, 32-45.
- Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983). The geometry of the reactive site of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallog. sect. B*, **39**, 480-487.
- Nilges, M. & O'Donoghue, S. I. (1998). Ambiguous NOEs and automated NOESY assignment. *Prog. NMR Spectrosc.* **32**, 107-139.
- Olejniczak, E. T., Dobson, C. M., Karplus, M. & Levy, R. M. (1984). Motional averaging of proton nuclear Overhauser effects in proteins. Predictions from a molecular dynamics simulation of Lysozyme. *J. Am. Chem. Soc.* **106**, 1923-1930.
- Palmer, A. G., III (1997). Probing molecular motion by NMR. *Curr. Opin. Struct. Biol.* **7**, 732-737.
- Palmer, A. G., III & Case, D. A. (1992). Molecular dynamics analysis of NMR-relaxation in a zinc-finger peptide. *J. Am. Chem. Soc.* **114**, 9067-9075.
- Pearlman, D. A. & Kollman, P. A. (1991). Are time-averaged restraints necessary for nuclear magnetic resonance refinement? A model study for DNA. *J. Mol. Biol.* **220**, 457-479.
- Philippopoulos, M., Mandel, A. M., Palmer, A. G., III & Lim, C. (1997). Accuracy and precision of NMR relaxation experiments and md simulations for characterizing protein dynamics. *Proteins: Struct. Funct. Genet.* **28**, 481-493.
- Post, C. B. (1992). Internal motional averaging and three-dimensional structure determination by nuclear magnetic resonance. *J. Mol. Biol.* **224**, 1087-1101.
- Richarz, R., Nagayama, K. & Wüthrich, K. (1980). Carbon-13 nuclear magnetic resonance relaxation studies of internal mobility of the polypeptide chain in basic pancreatic trypsin inhibitor and a selectively reduced analogue. *Biochemistry*, **19**, 5189-5196.
- Ryckaert, J., Ciocotti, G. & Berendsen, H. (1977). Numerical-integration of Cartesian equations of motion of a system with constraints - molecular dynamics of N-alkanes. *J. Comput. Phys.* **23**, 327-341.
- Smith, P. E. & van Gunsteren, W. F. (1994). Translational and rotational diffusion of proteins. *J. Mol. Biol.* **236**, 629-636.
- Steinbach, P. J. & Brooks, B. B. (1994). New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comp. Chem.* **15**, 667-683.
- Szyperski, T., Luginbühl, P., Otting, G. & Wüthrich, K. (1993). Protein dynamics studied by rotating frame <sup>15</sup>N spin relaxation times. *J. Biomol. NMR*, **3**, 151-164.
- Torda, A. E., Scheek, R. M. & van Gunsteren, W. F. (1989). Time-dependent distance restraints in molecular dynamics simulations. *Chem. Phys. Letters*, **157**, 289-294.
- Tropp, J. (1980). Dipolar relaxation and nuclear Overhauser effects in nonrigid molecules: the effect of fluctuating internuclear distances. *J. Chem. Phys.* **72**, 6035-6043.
- van Aalten, D. M., Amadei, A., Linssen, A. B., Eijssink, V. G., Vriend, G. & Berendsen, H. J. (1995). The essential dynamics of thermolysin: confirmation of the hinge-bending motion and comparison of simulations in vacuum and water. *Proteins: Struct. Funct. Genet.* **22**, 45-54.
- van Gunsteren, W. F., Brunne, R. M., Gros, P., van Schaik, R. C., Schiffer, C. A. & Torda, A. E. (1994). Accounting for molecular mobility in structure determination based on nuclear magnetic resonance spectroscopic and X-ray diffraction data. *Methods Enzymol.* **261**, 619-654.
- Wallach, D. (1967). Effect of internal rotation on angular correlation functions. *J. Chem. Phys.* **47**, 5258-5268.
- Weast, R. & Astle, M. (eds) (1983). *CRC Handbook of Chemistry and Physics*, CRC Press, Inc., Boca Raton, FL, USA.
- Withka, J. M., Swaminathan, S., Srinivasan, J., Beveridge, D. L. & Bolton, P. H. (1992). Toward a dynamical structure of DNA: comparison of theoretical and experimental noe intensities. *Science*, **255**, 597-599.
- Wlodawer, A., Walter, J., Huber, R. & Sjolín, L. (1984). Structure of bovine pancreatic trypsin inhibitor. results of joint neutron and X-ray refinement of crystal form ii. *J. Mol. Biol.* **180**, 301-329.
- Wlodawer, A., Nachman, J., Gilliland, G., Gallager, W. & Woodward, C. (1987). Structure of form III crystals of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **198**, 469-480.
- Wüthrich, K. & Baumann, R. (1976). <sup>13</sup>C spin relaxation studies of the basic pancreatic trypsin inhibitor. *Org. Magn. Res.* **8**, 532-535.
- Zhao, D. & Jardetzky, O. (1994). An assessment of the precision and accuracy of protein structures determined by NMR: dependence on distance errors. *J. Mol. Biol.* **239**, 601-607.
- Zwanzig, R. & Ailawadi, N. K. (1969). Statistical error due to finite time averaging in computer experiments. *Phys. Rev.* **182**, 280-283.

Edited by G. Von Heijne

(Received 14 April 1998; received in revised form 23 September 1998; accepted 1 October 1998)