Journal of Biomolecular NMR, 5 (1995) 106–111
ESCOM

J-Bio NMR 294

Structure from NMR and molecular dynamics: Distance restraining inhibits motion in the essential subspace

R.M. Scheek*, N.A.J. van Nuland, B.L. de Groot and A. Amadei

BIOSON Research Institute, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Received 8 March 1995
Accepted 23 May 1995

Keywords: Molecular dynamics; Essential dynamics; Protein structure; Distance restraints

Summary

We address the question how well proteins can be modelled on the basis of NMR data, when these data are incorporated into the protein model using distance restraints in a molecular dynamics simulation. We found, using HPr as a model protein, that distance restraining freezes the essential motion of proteins, as defined by Amadei et al. [Amadei, A., Linssen, A.B.M. and Berendsen, H.J.C. (1993) *Protein Struct. Funct. Genet.*, **17**, 412–425]. We discuss how modelling protocols can be improved in order to solve this problem.

Most protocols for molecular structure determination based on NMR data employ steps in which NOE and J-coupling information is incorporated in the form of distance and dihedral-angle restraints. Such restraints are indispensable for directing the computer search for conformations that agree with the experimental data. A drawback, however, is that restraints will change the dynamic properties of the simulated molecule. We have investigated this problem by examining the effect of distance restraints on the molecular dynamics (MD) of a protein simulated in a computer. While some details of the motion may not be represented accurately in such a computer model, the effects of restraints on the model can probably still be used to address the question how accurate protein models based on restraints can be.

Restrained molecular dynamics techniques were introduced by Van Gunsteren et al. (1984) to find energetically favourable conformations of molecules consistent with distance information from NMR spectra. NOEs between protons are interpreted as bounds on the distances between these protons and the molecular model is adjusted during an MD simulation of the molecule using artificial restraining forces that keep the protons between the upper and lower distance bounds. Torda et al. (1990) have shown that such distance restraining may inhibit motion of the molecule if all restraints are imposed simultaneously. A more gentle, and physically more realistic way to impose distance restraints was proposed, which allows for averaging of distances over a certain time period; the artificial restraining forces are invoked only when distances, properly averaged over this time period, are in conflict with measured NOEs. Obviously, this procedure only solves problems arising from motions that can equilibrate during the averaging period and indeed, heating due to artificial motions, which are driven by the restraining forces, was observed in cases where distances were fluctuating too slowly to equilibrate sufficiently during the averaging period. We looked further into this problem and studied the effects of distance restraining on the different types of motion a protein can undergo.

Amadei et al. (1993) have presented a method to separate a protein's conformational space into two subspaces: a low-dimensional 'essential' subspace, in which most of the protein's motion occurs in a highly correlated, anharmonic fashion, and the remaining subspace, where the motion can be described as rapidly equilibrating 'near-constraints'. We shall demonstrate in this paper, using HPr as an example, that restraining dramatically affects the protein's motion in the essential subspace.

The solution structure of the histidine-containing phosphocarrier protein (HPr) of *Escherichia coli* has been determined recently using a combination of NMR and MD techniques, in which NOEs were modelled as distance restraints (Van Nuland et al., 1994). To assess the effects

*To whom correspondence should be addressed.

TABLE 1
SUMMARY OF DISTANCE RESTRAINT VIOLATIONS IN MD SIMULATIONS OF HPr

MD cluster	Number of violations	Sum of violations (nm)	Largest violation (nm)
S2-r	29	0.722	0.060
S2-f	35	1.995	0.133[a]/0.199[b]
S15-r	25	0.661	0.071
S15-f	71	5.979	0.415[a]/0.145[b]
S21-r	21	0.647	0.077
S21-f	47	2.994	0.244[a]/0.0[b]
Sall-r	21	0.566	0.072
Sall-f	38	2.659	0.289[a]/0.086[b]

The restrained MD simulations in water and the subsequent simulations without distance restraining have been described in detail previously (Van Nuland et al., 1994). In addition to the two simulations described, starting from structures 2 and 21 of the ensemble of 32 in vacuo MD structures, we performed another MD simulation starting with structure 15. These three simulations were performed at 300 K, using the GROMOS 37C4 force-field parameter set (Van Gunsteren and Berendsen, 1987). HPr was surrounded by 2921, 3020 and 2317 water molecules, respectively, in a truncated octahedron. After 30 ps of equilibration, the simulations were continued for 200 ps using time-dependent distance restraints with a memory time constant of 5 ps (Torda et al., 1990) and a distance-restraining force constant of 1000 $kJ\,mol^{-1}\,nm^{-2}$. Restraints involving pseudo-atoms (Wüthrich et al., 1983) were referred to appropriate sites in terms of GROMOS atoms. All three simulations were then continued for another 200 ps without any distance restraining. Snapshots were taken every 10 ps from the three time-averaged distance-restrained MD simulations (sets S2-r, S15-r, S21-r) and from the free MD simulations (sets S2-f, S15-f, S21-f), without energy minimisation. From the combined time-averaged restrained MD simulations and the combined unrestrained MD simulations, 10 snapshots were selected every 20 ps, resulting in two new clusters of 30 snapshots each (sets Sall-r and Sall-f, respectively). For all ensembles the $<r^{-3}>^{-1/3}$-averaged distances are compared with the NOE-derived distance bounds to calculate the violations.
[a] NOE His^{76} $H^{\delta 2}$ - Leu^{80} H^{β}.
[b] NOE Glu^{85} H - Ala^{10} H^{α}.

of the restraints on the final model we also performed three unrestrained MD simulations in water, starting from three different starting conformations. We analysed these MD trajectories using Amadei's essential dynamics technique (Amadei et al., 1993). We investigated the effects of NOE-derived distance restraints on the simulated motion of HPr by projecting restrained MD trajectories onto the 'essential' and 'near-constraints' subspaces.

The MD simulations of HPr used for this analysis have been described in detail by Van Nuland et al. (1994). In short, three different starting conformations of HPr were chosen from the set of conformations derived from the ^1H-^1H NOE data. Each was subjected to 200 ps of MD simulation in water, using time-averaged (Torda et al., 1990) distance restraining (employing a 'memory time constant' of 5 ps), and subsequently to another 200 ps of free MD without any distance restraining. Table 1 shows to what extent the resulting structures agree with the NOE-derived distances. The MD runs that were performed with time-averaged distance restraining result in sets of structures that agree well with the experimental data. However, the free MD simulations result in sets of structures that show some unacceptably large violations of the NOE-derived distance restraints. Usually errors in the force field are held responsible for such a poor agreement, but we shall show that there may be a more fundamental problem in the modelling protocol itself.

The three free MD simulations were combined and used to define the essential and near-constraints subspaces, based on linear covariances between displacements of C^{α} atoms from their mean positions (Amadei et al., 1993). Figure 1 shows a plot of the eigenvalues of the covariance matrix, sorted according to size. As demonstrated before for the dynamics of other proteins, most of the total motion of HPr occurs in a low-dimensional subspace, called the essential subspace. Figure 2 visualises the motion of HPr along the eigenvectors 1, 3 and 4. Note that the regions of HPr involved in the essential motion include the active-centre loop around residue 15, and those parts of the protein that are known to be involved in binding to the A-domains of Enzyme $II^{mannitol}$ and Enzyme I (Van Nuland et al., 1993) of the *E. coli* phosphoenolpyruvate-dependent phosphotransferase system. The motion along eigenvector 2 is not shown; it is dominated by the C-terminal glutamate, which was seen to move significantly in one of the three free MD simulations. Figure 3A shows the projections of the C^{α} displacements during the three free MD simulations onto the eigenvectors 1, 3 and 4, spanning the essential subspace. During the three MD simulations, as noted before by Van Nuland et al. (1994), different, only partially overlap-

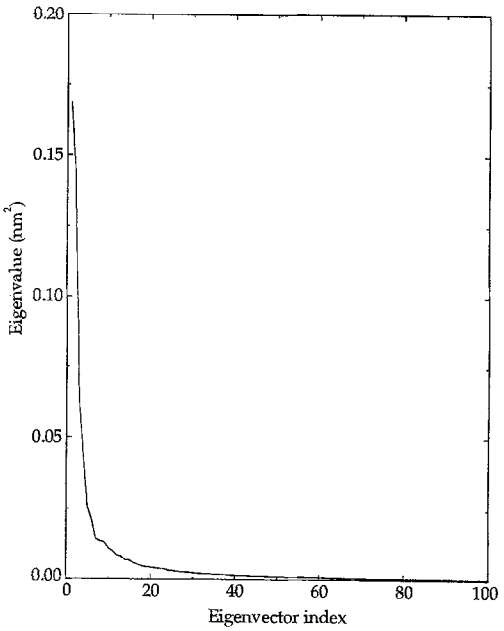

Fig. 1. Plot of the first 100 eigenvalues versus the corresponding eigenvector index of the covariance matrix of the C^{α} atomic displacements from their average positions in three free MD simulations of HPr. The eigenvalues were sorted according to size.

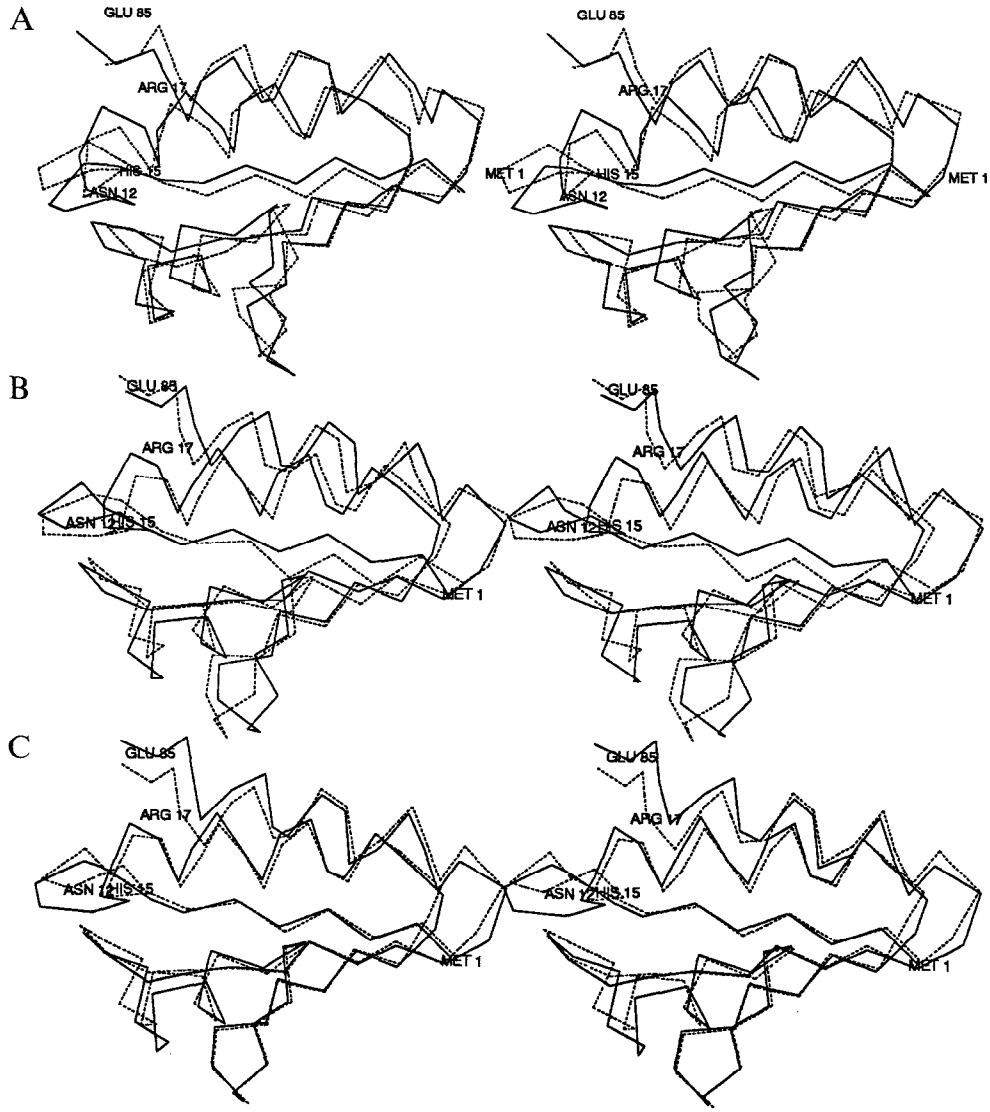

Fig. 2. Stereo pictures showing snapshots of the motion along different eigenvectors of HPr. The continuous and broken lines connect the C^{α} atoms of HPr. The two conformations correspond to the largest displacements from the average conformation along the eigenvectors 1, 3 and 4 in A, B and C, respectively.

ping regions of the conformational space appear to have been sampled. Figure 3B shows the C^{α} displacements during the three restrained MD simulations, projected onto the same eigenvectors. This figure shows that during the restrained MD simulations there has been little motion in the directions defined by these three eigenvectors. Finally, Fig. 4 shows how restraining affects the motion along the first 50 eigenvectors. We have compared the mean displacements along these eigenvectors, as well as the rms fluctuations around these mean displacements, during the three restrained and the three free MD simulations. It is clear from this analysis that restraining inhibits the motion along the essential eigenvectors, whereas no effect is seen on motion in the near-constraints subspace.

An important advantage of NMR as a technique to solve the structure of proteins is that the protein can be studied directly in its native solution state, because the mobility of a protein in solution may be an important aspect of its function. On the other hand, this mobility complicates the modelling procedure, because the measured NMR parameters should be regarded as time and ensemble averages. Ideally, the final protein model resulting from an NMR structure determination would be a set of conformers representative of the actual ensemble that exists under the experimental conditions. However, the structural parameters that can be measured by NMR (NOEs, J-couplings) are not nearly sufficient to define such an ensemble in a unique manner, even if we include our knowledge about bond lengths, bond angles and atomic radii. In practice, therefore, the NMR parameters are usually not modelled as averages. Instead, a static protein model is presented, in which all NMR parameters are accounted for simultaneously. Given a sufficiently large number of NOEs and J-coupling data, this proce-

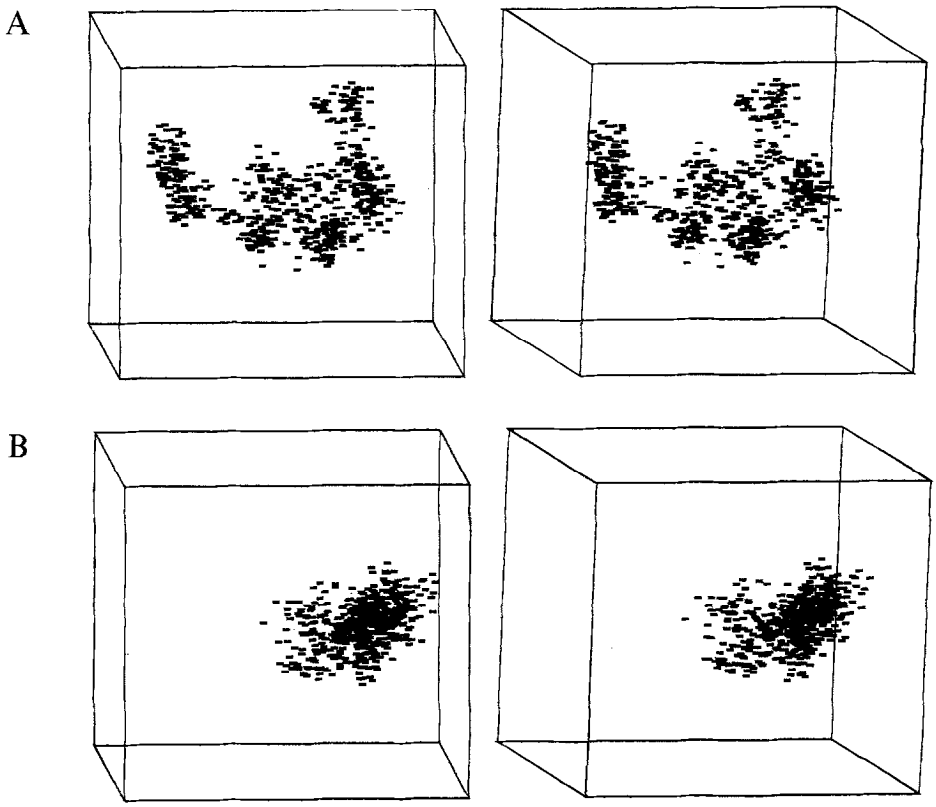

Fig. 3. Stereo representation of the projections of the C^{α} displacements during three MD simulations of HPr onto eigenvectors 1, 3 and 4. (A) Free MD simulations; (B) MD simulations with time-averaged distance restraints. Horizontal, vertical and perpendicular axes correspond to displacements along eigenvectors 1, 3 and 4, respectively, ranging from −1.0 to 1.0 nm for each direction.

dure usually results in a single conformation with high precision, but this precise conformation is not necessarily close to the averaged solution conformation or even to a representative member of the solution ensemble. Moreover, in several cases it has proven impossible to find a single conformation of the molecule under study that can account for all NMR parameters simultaneously (Scarsdale et al., 1986; Kessler et al., 1988,1991; Pepermans et al., 1988; Schmitz et al., 1992; Van der Graaf et al., 1992). A partial solution to this problem was presented

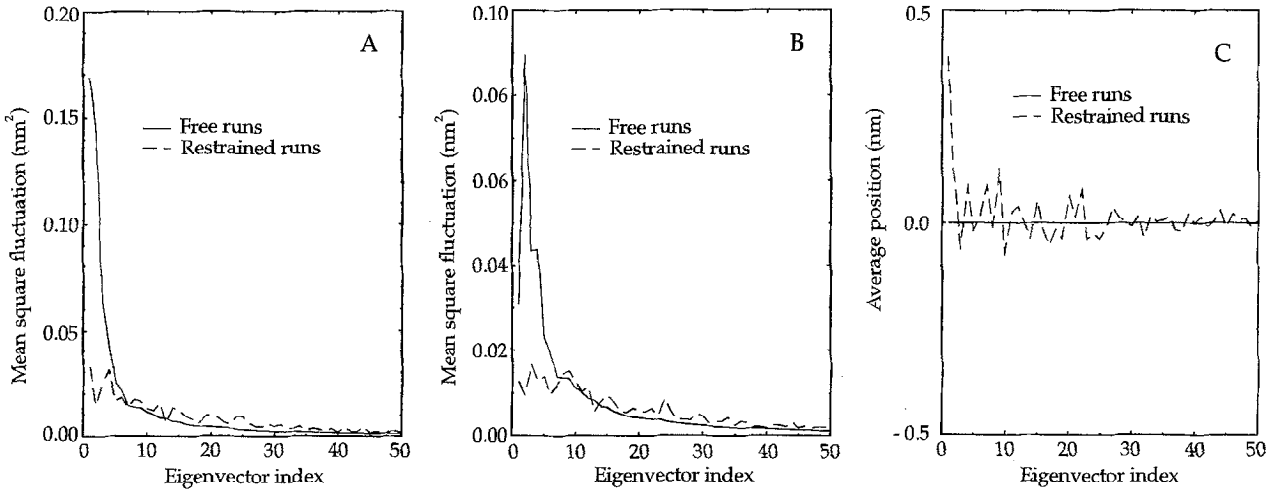

Fig. 4. (A) and (B) Mean square fluctuations of C^{α} atoms from their averaged positions along the first 50 eigenvectors spanning the conformational space of HPr. Continuous lines refer to the free MD simulations, broken lines to the restrained MD simulations. In A, the three MD simulations were combined before calculation of the average C^{α} positions and mean square displacements for each eigenvector; in B, the three simulations were taken separately. From these, three averages and mean square displacements were calculated per eigenvector, which were then averaged for this plot. (C) Shifts in the average positions along the first 50 eigenvectors of HPr during the combined three restrained MD simulations (broken line), relative to the averages calculated from the three free MD simulations (continuous line).

by Torda et al. (1990), who proposed a restrained MD simulation of a protein where a NOE-derived distance restraint is applied only when the corresponding proton–proton distance, properly averaged over a 'memory period', exceeds the experimental distance bounds. Recently, a similar solution was presented for J-coupling restraints (Torda et al., 1993). It was realised that such averaging procedures will affect those motions of the protein that take longer to equilibrate than the memory period (in practice only 10 ps or less) and the results presented here for HPr demonstrate that this can indeed be a serious limitation. We have shown that distance restraining 'freezes' the protein's essential motion, which, in the absence of distance restraints, constitutes the larger part of the protein's motion, as measured by the mean square atomic displacements along the essential degrees of freedom during a free MD simulation.

In the case of HPr, it is clear that not all distance restraints are equally responsible for the observed inhibition of the protein's essential motion. The distance re-

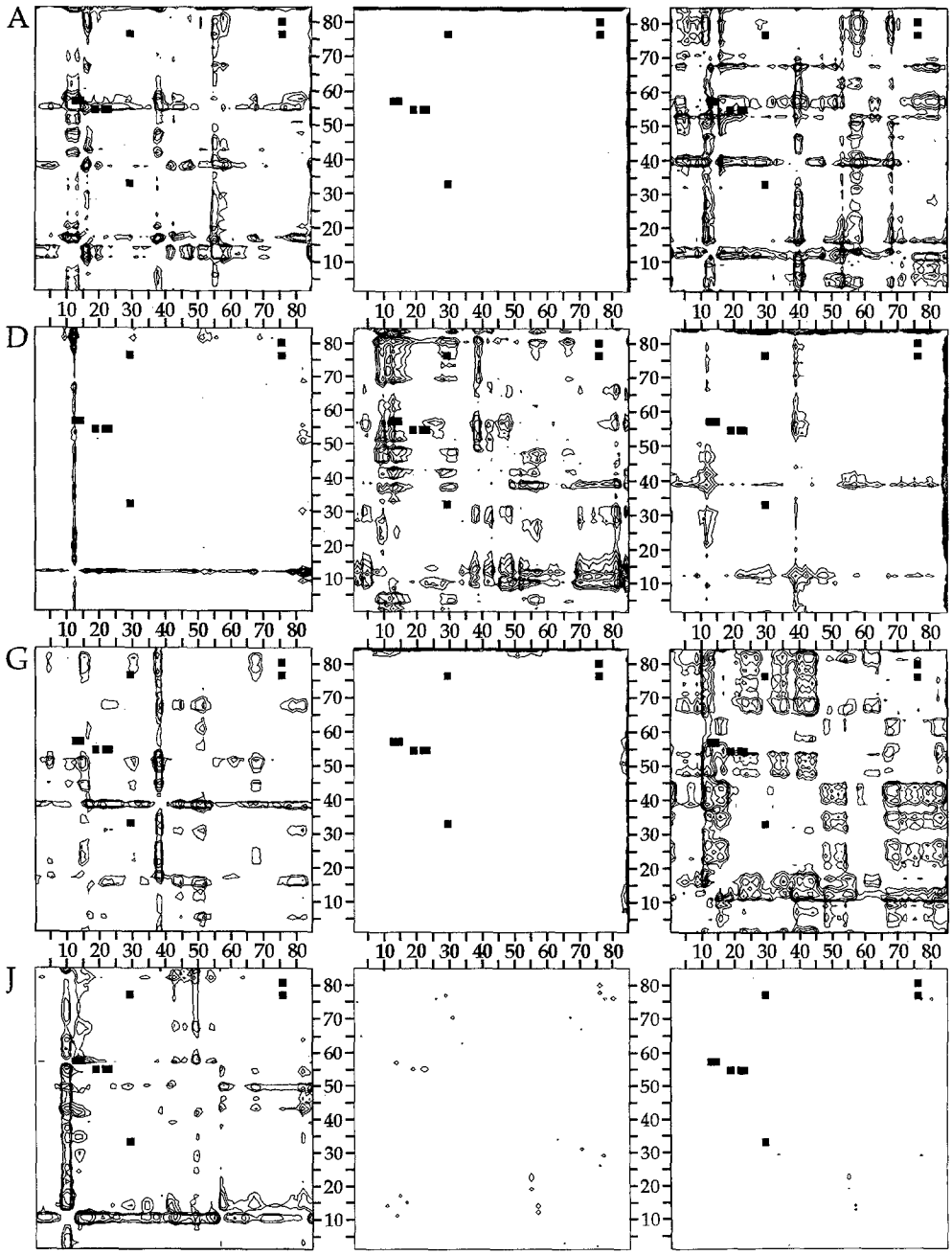

Fig. 5. C^{α}-C^{α} distance fluctuations caused by the simulated motion of HPr along its first 10 eigenvectors (A–J). The square root of the squared fluctuations along eigenvectors 1–10 was contoured at levels corresponding to 0.9, 0.8, 0.7, 0.6 and 0.5 times the maximum C^{α}-C^{α} fluctuation. Also indicated are the pairs of residues that are involved in NOE restraint violations exceeding 0.1 nm (K) and 0.05 nm (L) during the three free MD simulations of HPr, after ensemble-averaging of the corresponding distances over these three MD clusters.

straints that are violated during the free MD simulations are likely to be involved. To test this, we calculated the fluctuations in C^{α}-C^{α} distances caused by the simulated motions along different eigenvectors. Figures 5A–J show the pairs of C^{α} atoms for which the corresponding distance is most affected by the motion along each of the first 10 eigenvectors during the free MD simulations. In the same figure (Figs. 5K and 5L) we have indicated the residue pairs that are involved in the violations of the NMR-derived distance restraints. Indeed, it appears that the motion along several of the essential eigenvectors, observed during the free MD simulations, causes substantial fluctuations in the C^{α}-C^{α} distances of residue pairs that are also involved in some of the most significant NOE violations, most notably those between the active-site loop and residues 55–57. The conclusion is inevitable, then, that these restraints must have been at least partly responsible for the inhibition of the essential motion observed during the restrained MD simulations. Only after release of the restraints was the protein able to move along the essential eigenvectors and to enter regions of conformational space that were inaccessible given the restraints and the way they had been imposed.

Not all NOE violations that showed up during the free MD simulations can be accounted for in this way. For example, the NOE between His^{76} and Leu^{80} (denoted a in Table 1) does not seem to fit into this picture. It was violated during all the restrained MD simulations and, to a larger extent, during the free MD simulations. This particular NOE may have gained intensity due to spin diffusion (the breakdown of the two-spin approximation; Kalk and Berendsen, 1976) in the hydrophobic core of HPr around Phe^{22}, leading to an upper distance bound for this pair of protons that is too tight.

An important conclusion from this analysis is that distance restraining during the final modelling stages of an NMR structure determination freezes the protein's essential motion, as defined by Amadei et al. (1993). Hence, in order to improve the modelling protocol, we need to sample more completely the conformational space accessible to a protein at the temperature of the NMR experiments. From the analysis presented above we learned that distance restraining, even with the use of time-dependent restraints, will limit the accessible conformational space to an unrealistic extent: motions that occur on time scales larger than the averaging period will be missed. Even without distance restraining a serious sampling problem remains, because presently available computers only allow realistic MD simulations of proteins over time spans of nanoseconds. It is encouraging to find that most of the distance violations decrease if averaging is performed over the three free MD calculations taken together, when compared with the three free MD simulations taken separately. Presently we are testing improved sampling schemes that employ the low dimensionality of the essential subspace, in an attempt to improve our NMR structure refinement protocols and, more importantly, to improve our understanding of a protein's behaviour in solution.

Acknowledgements

We thank Dr. Andrew Torda for valuable comments on the manuscript. We thank Prof. Dr. H.J.C. Berendsen, Prof. Dr. G.T. Robillard and Prof. Dr. I.D. Kuntz for stimulating discussions.

References

Amadei, A., Linssen, A.B.M. and Berendsen, H.J.C. (1993) *Protein Struct. Funct. Genet.*, **17**, 412–425.

Kalk, A. and Berendsen, H.J.C. (1976) *J. Magn. Reson.*, **24**, 343–366.

Kessler, H., Griesinger, C., Lautz, J., Müller, A., Van Gunsteren, W.F. and Berendsen, H.J.C. (1988) *J. Am. Chem. Soc.*, **110**, 3393–3396.

Kessler, H., Matter, H., Gemmecker, G., Kling, A. and Kottenbaum, M. (1991) *J. Am. Chem. Soc.*, **113**, 7550–7563.

Pepermans, H., Tourwé, D., Van Binst, G., Boelens, R., Scheek, R.M., Van Gunsteren, W.F. and Kaptein, R. (1988) *Biopolymers*, **27**, 323–338.

Scarsdale, J.N., Yu, R.K. and Prestegard, J.H. (1986) *J. Am. Chem. Soc.*, **108**, 6778–6784.

Schmitz, U., Kumar, A. and James, T.L.J. (1992) *J. Am. Chem. Soc.*, **114**, 10654–10656.

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

Torda, A.E., Brunne, R.M., Huber, T., Kessler, H. and Van Gunsteren, W.F. (1993) *J. Biomol. NMR*, **3**, 55–66.

Van der Graaf, M., Scheek, R.M., Van der Linden, C.C. and Hemminga, M.A. (1992) *Biochemistry*, **31**, 9177–9182.

Van Gunsteren, W.F., Kaptein, R. and Zuiderweg, E.R.P. (1984) In *Proceedings of a NATO/CECAM Workshop on Nucleic Acid Conformation and Dynamics* (Ed., Olsen, W.K.) CECAM, Orsay, pp. 79–82.

Van Gunsteren, W.F. and Berendsen, H.J.C. (1987) GROningen MOlecular Simulation (GROMOS) library manual, Biomos, Groningen.

Van Nuland, N.A.J., Kroon, G.J.A., Dijkstra, K., Wolters, G.K., Scheek, R.M. and Robillard, G.T. (1993) *FEBS Lett.*, **315**, 11–15.

Van Nuland, N.A.J., Hangyi, I.W., Van Schaik, R.C., Berendsen, H.J.C., Van Gunsteren, W.F., Scheek, R.M. and Robillard, G.T. (1994) *J. Mol. Biol.*, **237**, 544–559.

Wüthrich, K., Billeter, M. and Braun, W. (1983) *J. Mol. Biol.*, **169**, 949–961.