Journal of Biomolecular Structure & Dynamics, ISSN 0739-1102
Volume 13, Issue Number 4 (1996),

An Efficient Method for Sampling the Essential Subspace of Proteins

A. Amadei[1]*, A.B.M. Linssen[1], B.L. de Groot[1], D.M.F. van Aalten[2] and H.J.C. Berendsen[1]

[1]Groningen Biomolecular Sciences and Biotechnology Institute (GBB), Department of Biophysical Chemistry, the University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

[2]Department of Biochemistry and Molecular Biology, University of Leeds, Leeds LS2 9JT, UK

Abstract

A method is presented for a more efficient sampling of the configurational space of proteins as compared to conventional sampling techniques such as molecular dynamics. The method is based on the large conformational changes in proteins revealed by the "essential dynamics" analysis. A form of constrained dynamics is performed, forcing the system to move along some of the essential coordinates. This results in a broader sampling of the essential subspace than in a comparable conventional molecular dynamics simulation without constraints. The new sampling method (essential dynamics sampling) was applied to the histidine-containing phosphocarrier protein HPr. The results indicate that the essential dynamics sampling method produces physically allowed structures, as estimated by the evaluation of many geometrical properties. In addition, a study of the motions in the essential subspace reveals a diffusion-like behavior.

Introduction

Folded proteins are stable mechanical constructs able to perform a wide range of functions. These functions are defined by the structure and the dynamical behavior of a folded protein. The 'folded state' itself is a collection of interconvertable structures. Unfortunately it is experimentally not possible to investigate the properties of such an ensemble of structures extensively and show the link between these (dynamical) properties and biological function. Hence, many questions about this link remain unanswered.

Recently we showed [1-4] that the configurational space of a folded protein can be divided in a many-dimensional near-constraints space where simple small fluctuations occur and a low-dimensional 'essential' subspace in which the large concerted motions are confined (this important result was lately also confirmed by others [5]). Previous work has suggested that these 'essential motions' are connected to biological function. [1-4] However, the conformational changes which were revealed by the essential dynamics analysis are limited by the sampling of these motions in the molecular dynamics (MD) trajectories. An exhaustive sampling and investigation of the essential subspace could reveal the extent and characteristics of structural transitions in proteins.

In this paper, a general method to produce an efficient sampling of a protein essential subspace is described. As an example, the investigation into the behavior of the

*Author to whom correspondence should be sent.
Phone: (31) 50-3 634378; Fax: (31) 50 3 634800;
E-mail: H.J.C.Berendsen@chem.rug.nl.

essential coordinates of a protein (histidine containing phosphocarrier protein (HPr)) is presented.

Theory

In the essential dynamics method a covariance matrix based on the fluctuations in atomic (Cartesian) positions is used [1, 6]. No special shape of the potential energy (e.g. harmonic as in the normal modes approach [7]) is required and any set or subset of coordinates can be used. The method is based on the extraction of eigenvectors from the covariance matrix which are then ordered by decreasing corresponding eigenvalue. If approximated linear constraints or even statistically independent directions are present in the configurational space, they will be defined by eigenvectors of the covariance matrix [1, 6]. The eigenvectors of such a matrix correspond also to the 'best fitted' directions through points in configurational space produced by an MD trajectory. The procedure then can be considered as a linear multidimensional fitting (introduced previously in protein MD, [8]), the first eigenvector having the maximum possible average square displacement (eigenvalue). If the overall motion of the protein is mostly confined in a linear n dimensional subspace of the configurational space, this subspace will be defined by the first n eigenvectors of the covariance matrix. We found for several proteins that the essential subspace is confined within approximately the first 10 eigenvectors and all the other eigenvectors are "near-constraints" coordinates [1-4]. A structure can be described as a function of the essential coordinates $\boldsymbol{\xi}$ and of the near-constraints coordinates $\boldsymbol{s}$:

$$\boldsymbol{x}(\boldsymbol{\xi};\boldsymbol{s}) = \boldsymbol{x}^0(\boldsymbol{\xi}) + \Delta\boldsymbol{x}(\boldsymbol{s}) \qquad 1$$

with $\Delta\boldsymbol{x}$ always small and hence $\boldsymbol{x}^0$ determining the relevant structural properties. From the behavior of the near-constraints coordinates (independent Gaussian fluctuations) it is also possible to express the free energy as a potential energy of mean force in the essential coordinates and near constraints ones:

$$A(\boldsymbol{\xi};\boldsymbol{s}) \approx A^0(\boldsymbol{\xi}) + \frac{1}{2}\sum_i k_i s_i^2 \qquad 2$$

the approximation becoming only roughly valid when we use all-atom eigenvectors.

Equations 1 and 2 show that we must be able to sample the essential subspace to obtain and investigate the biologically active states of a protein. Such a sampling is inefficient using conventional MD simulations, because of the time required for a large sampling. The time currently maximally obtainable for a reasonably large protein by MD, in the ns range, does not suffice for large sampling. However, the knowledge of the essential eigenvectors can be used to produce a more efficient sampling in the essential subspace.

Two types of essential dynamics sampling are possible:

1. "Geometrical" sampling. In this mode of sampling we are mostly interested in collecting geometrical properties of the structures as a function of the position in the essential subspace, and use these geometrical properties to define the limits of this subspace. A grid in the essential subspace is defined and at every grid point we construct the structure $\boldsymbol{x}^0$. We equilibrate $\Delta\boldsymbol{x}$ by energy minimization and/or a short MD simulation keeping the position in the essential subspace fixed. Using this method it is possible to construct a large number of structures defined as linear combinations of the essential eigenvectors, with only a local small adjustment for all the other coordinates. Such a collection of structures can be analyzed geometrically and statistically to define 'stable' and 'unstable' regions in the essential subspace.

2. "Physical" sampling. Here the system is moving in the essential subspace in a

dynamical (MD related) way. In every region of this subspace an ensemble of physical structures can be generated, which can be used to investigate the physical properties of the conformational states. In order to efficiently sample the configurational space, constraint forces are used to move the system in the essential subspace while all the other coordinates move freely according to the normal equations of motion. The constraint force is defined using non-stationary holonomic constraints such as a constant step motion along a direction in the essential subspace or a constant step expansion or contraction of the length of the radius between a fixed (reference) position and the actual position, both defined in the essential subspace (we use in general a subset of essential coordinates).

We shall investigate the physical sampling procedure in details here. The displacement in the essential subspace during one time step Δt of MD can be expressed as:

$$\Delta\boldsymbol{\xi} = \Delta\boldsymbol{\xi}_d + \Delta\boldsymbol{\xi}_c \qquad 3$$

$\Delta\boldsymbol{\xi}_d$ is the displacement produced by the dynamics (without constraint) and $\Delta\boldsymbol{\xi}_c$ is the correction for the application of the constraint. The constraint in the essential subspace can be defined as:

$$G(\boldsymbol{\xi}(t); t) = 0 \qquad 4$$

Eq. 4 does not suffice to solve for $\Delta\boldsymbol{\xi}_c; t$ in a unique way. To obtain a unique solution, we add the requirement that the total perturbation $|\Delta\boldsymbol{\xi}_c|$ is minimized. This is achieved using one Lagrangian multiplier.

$$\frac{\partial}{\partial\Delta\xi_c^i}\left[\frac{1}{2}\sum_i(\Delta\xi_c^i)^2 - \lambda G(\boldsymbol{\xi}(t) + \Delta\boldsymbol{\xi}_d + \Delta\boldsymbol{\xi}_c; t + \Delta t)\right] = 0 \qquad 5$$

or

$$\Delta\xi_c^i - \lambda\frac{\partial G}{\partial\Delta\xi_c^i} = 0 \qquad 6$$

or

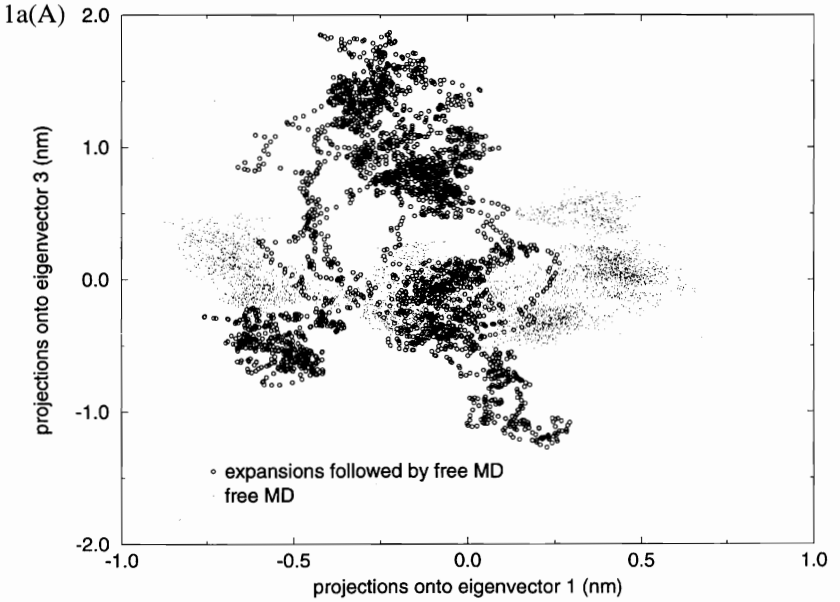
$$\Delta\xi_c^i = \lambda\frac{\partial G}{\partial\Delta\xi_c^i} \qquad 7$$

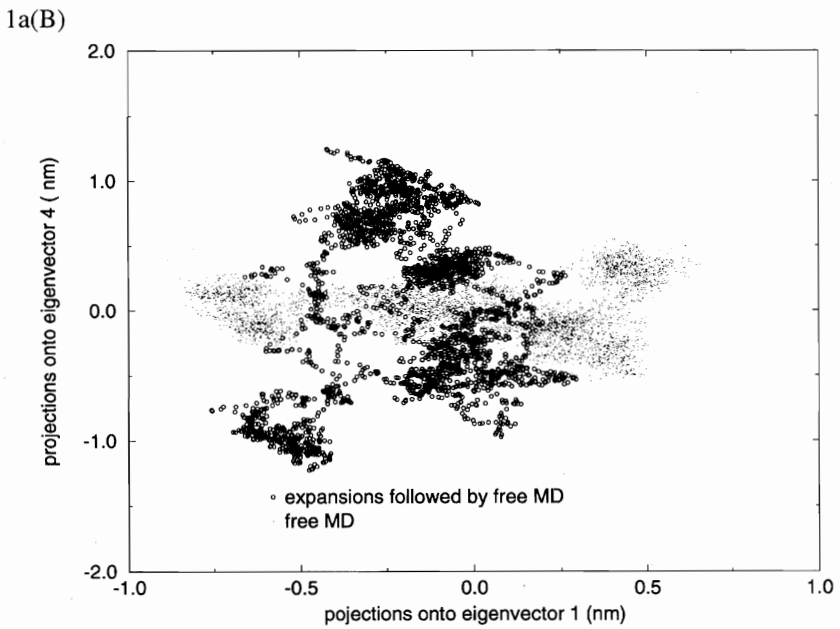
Using equation 7 and equation 4 we can express λ as a function of $\boldsymbol{\xi}(t)$; $\Delta\boldsymbol{\xi}_d$ and t. Such a value of λ can be used to correct $\Delta\boldsymbol{\xi}$ to fulfill our constraint motion with the least perturbation. In appendix A one example is given.

Method

We have produced an extensive essential dynamics (ED) sampling of the 85 residue histidine-containing phosphocarrier protein HPr from E. Coli. Sampling eigenvectors were obtained from MD simulations in water, described elsewhere [9]. Three simulations, initiated from three structures taken from the NMR cluster [9] were concatenated for an essential dynamics analysis. This concatenation of related trajectories has proven to be a suitable method for obtaining eigenvectors representing identical concerted motions in the trajectories [2-4]. Since all three simulations started from structures derived from NMR data, the cluster of structures obtained from the three MD simulations together will form an ensemble of physically allowed configurations (assuming the validity of our force-fields). Taking this ensemble of structures together for an essential dynamics analysis will yield overall directions which describe both motions within the single trajectories, as well as differences between the simulations (which originate in the fact that they were started from different structures).

Only C-α atoms were used to construct a covariance matrix from which the eigenvectors were extracted. Initial experiments showed that the use of all-atom eigenvectors resulted in errors in the sampling caused by apparent correlation's between





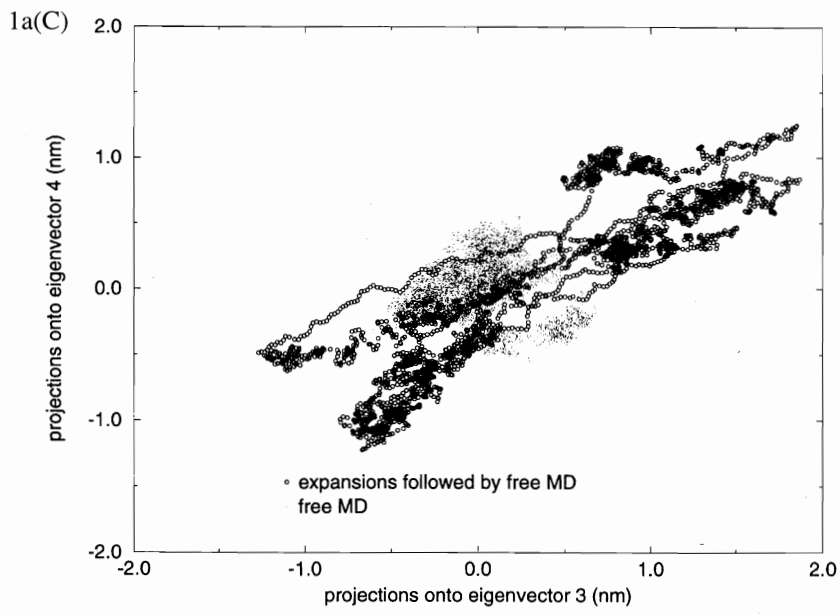


Figure 1a: Projections of the trajectories produced by the conventional MD runs and the ED sampling method onto planes by combinations of the three eigenvectors along which the position was constrained. A. Projections onto the 1-3 plane. B. Projections onto the 1-4 plane. C. Projections onto the 3-4 plane.

large backbone motions and slow side chain rearrangements, because of insufficient statistics. We chose to use eigenvectors 1, 3 and 4 for the ED sampling because these show large concerted motions of residues throughout the protein. The second eigenvector only shows a wobbling of the C-terminus whereas the rest of the protein remains fixed.

The protocol used for the ED sampling consisted of cycles of three parts. The first step is a "radius expansion" (see appendix A) of 5000 steps using the first, third and fourth eigenvector, increasing the radius spanned by these coordinates by 0.0004 nm per step. Initial experiments showed that increasing the stepsize for the constrained coordinates leads to undesired effects (highly strained structures). The expansion is initiated with the final configuration of the previous free MD run (see step 3, below) taken as the center of the expansion sphere. During this expansion, the projections on the selected eigenvectors produced in a normal MD step are corrected in the radial direction of the expansion sphere in such a way that they fulfill the expansion constraint. The second step in the cycle consists of 5000 steps of MD, fixing the positions along the essential coordinates which we used in the expansion. This is meant to equilibrate the system at the new position in the essential space. The third step is a free run of 20000 steps (40 ps) performed to sample the local configurational space.

We started from an equilibrated structure taken as a snapshot from one of the three free MD simulations from which the eigenvectors were extracted. First, a number of expansion steps were performed (no free/fixed MD). Cycle 1 was started from the resulting structure. Cycles 2-5 were started from the final structure of the preceding cycle. A sixth cycle was started directly from the equilibrated conventional MD run. The software used was based on the simulation package GROMOS [10], with modifications to allow constraints on essential eigenvectors. DSSP [11] was used for evaluation of hydrogen bonds, secondary structure and solvent accessibilities. PROCHECK [12] was used to evaluate the number of unfavorable phi/psi combinations. All visualizations and structure evaluations were performed with the modeling package WHAT IF [13].

The system studied consisted of 785 HPr atoms surrounded by 2317 water molecules [14] adding up to a total of 7736 atoms. Periodic boundary conditions were applied using a truncated octahedron box. The temperature was kept constant by coupling to an external bath [15] with a coupling constant of 0.01 ps. SHAKE [16] was used to constrain bond lengths to their equilibrium position, allowing a time step of 2 fs. 4s.

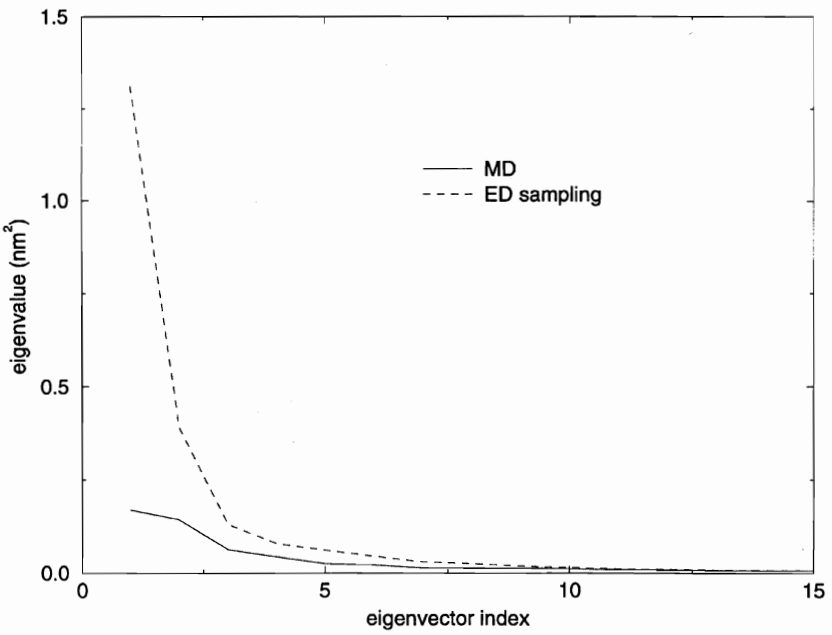


Figure 1b: Comparison of the eigenvalues obtained by the ED sampling method with the eigenvalues obtained by the three initial MD runs.

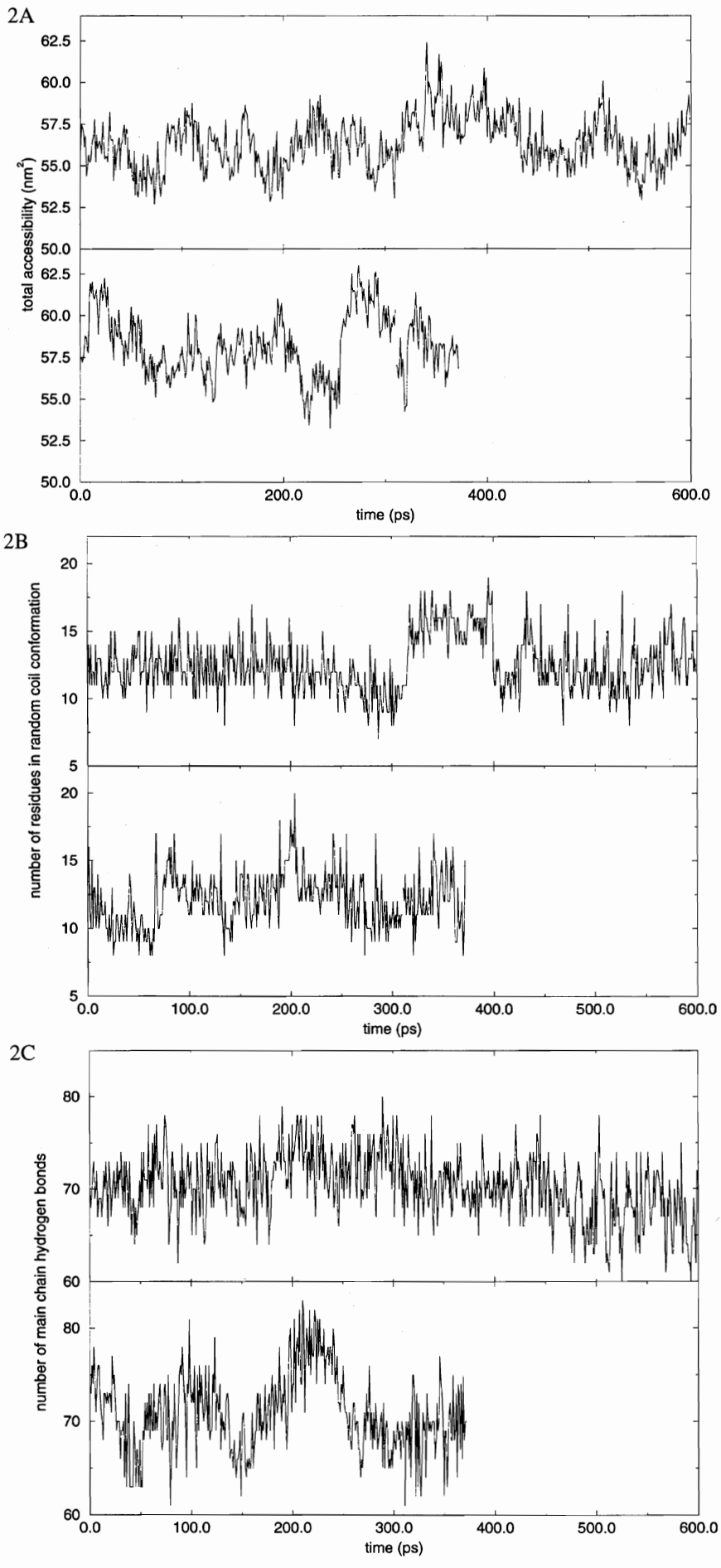


Figure 2: Geometrical properties as a function of time. In each plot the upper graph shows the properties in the three traditional MD simulations of 200 ps each, the lower graph shows the same properties for the six ED sampling cycles as described in the text. A. The total solvent accessible surface. B. The number of residues in random coil conformation. C. The number of backbone-backbone hydrogen bonds. Figure 2D and E continued on the following pages.

To investigate the efficiency of the ED sampling protocol, we compared the initial three MD simulations with the trajectory obtained by the ED sampling. Fig. 1a shows the projections (calculated from the three initial MD runs and the ED sampling run) in the three planes spanned by eigenvectors 1, 3 and 4. The space sampled by the expansion/free runs, in these planes, is larger in almost every direction than the space which was sampled during the three conventional MD simulations. In Fig. 1b the eigenvalues obtained by the three MD runs are compared to the-eigenvalues obtained by the ED sampling. It is clear that the first 10 eigenvalues of the ED sampling are considerably larger then the first 10 eigenvalues of the three MD runs, implying that the volume covered by the ED sampling in the essential subspace is significantly broader. For example the ratio between the square root of the first 5 eigenvalues (root mean square fluctuations) of the ED sampling run and the square root of the first 5 eigenvalues of the 3 MD runs is about 14. The results presented here were produced by six cycles as described previously, with each cycle containing 31000 integration steps, which would correspond to 62 ps of usual MD. The six cycles together would therefore cost the same amount of CPU time as an MD simulation without constraints of 372 ps. The three conventional MD runs were 200 ps each, adding up to a total of 600 ps. Hence, with the ED sampling method a larger volume is sampled at less computational cost.

During each cycle, several geometrical properties were monitored (Fig. 2), like the total solvent accessibility (A), the conservation of secondary structure elements (B), the number of hydrogen bonds (C) and the number of strained phi/psi combinations (D). These geometrical properties proved to give more valuable information than the RMS deviation from a reference structure, which is widely used in MD as a stability measurement. The motions involved are so large that the RMS deviation, see Fig. 3, indicates large structural rearrangements (usually considered to be an indication for an instability of the simulation), while the geometrical properties are not significantly different in the ED sampling cycles compared to the three MD runs. This shows that the ED sampling method produces physically allowed structures. Energies during these calculations showed a similar trend (the energies of the conformations produced by this form of constrained dynamics are fluctuating in the same range as in conventional MD). From the data obtained it seems that almost every direction of the essential subspace is allowed within a certain boundary. These boundaries in the essential subspace could define then a large set of physically allowed structures of a protein. A superposition of structures at the borders of the essential subspace is shown in Fig. 4. Although there are significant

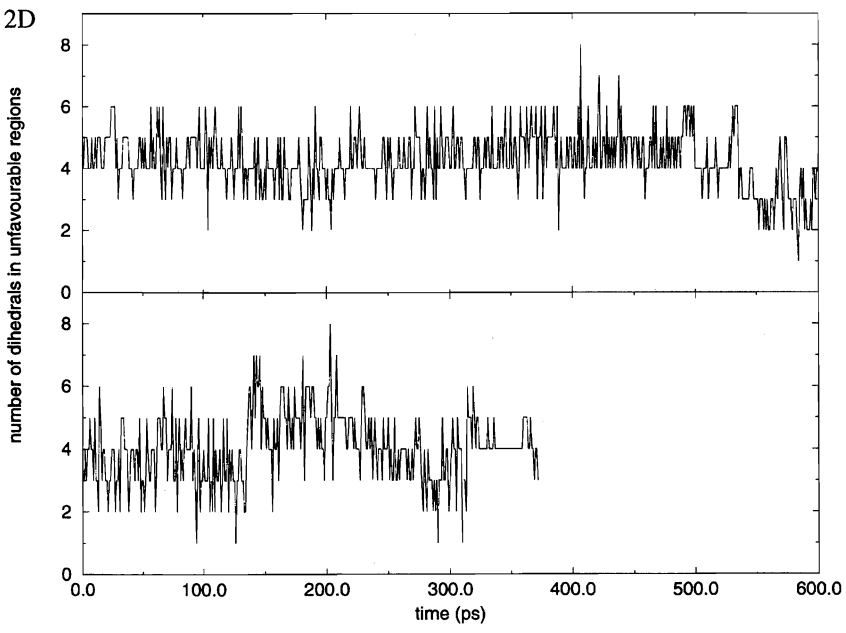


Figure 2 (continued): D. The number of phi/psi torsion angles in unfavourable regions.

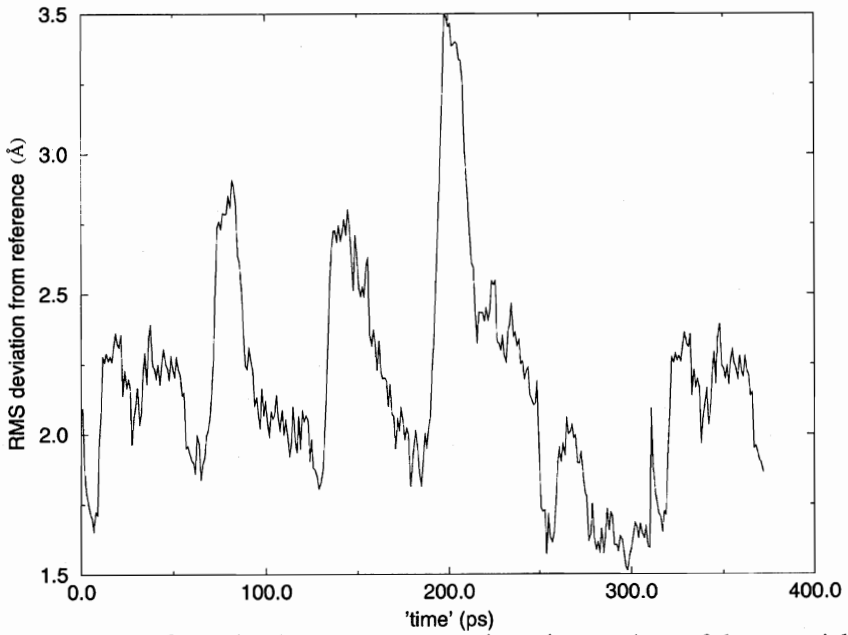


Figure 3: Root mean square deviation (RMS) of the alpha carbon atoms from a reference structure, during the ED sampling cycles.

differences in conformation between structures in various regions of the essential subspace, the protein retains its overall fold.

It should be noted that the density obtained by the ED sampling procedure cannot be used directly for exact thermodynamical evaluations on the system, since its behavior is affected by the presence of the constraints. The ED sampling procedure can be regarded as an efficient way to collect a large ensemble of possible structures that a folded protein can reach. This collection of structures can then be used to evaluate the structural and physical properties in different regions of the essential subspace. In Fig. 5 we also compared the eigenvectors of the initial three MD runs with the eigenvectors obtained by the ED sampling. In the figure are shown the comulative square projections (square inner products) of single eigenvectors of the three MD runs, on the eigenvectors set obtained from the ED sampling. The first 10 eigenvectors (in the figure are shown the first 5 and the tenth) of the three MD runs (defining approximately the essential coordinates of these trajectories) can be almost completely reproduced (at least about 80%) within the first 50 (50 out of 255) eigenvectors evaluated from the ED sampling. In particular eigenvectors 3 and 4 of the three MD runs are almost completely defined by the first 6 or 7 eigenvectors of the ED sampling, and the other essential eigenvectors of the MD runs can be reproduced for 50% or 60% within the first 20 (20 out of 255) eigenvectors obtained from the ED sampling. The partial mixing of the essential subspace obtained from the three initial MD runs, and the first 20 or 30 near constraint eigenvectors of the ED sampling is due to insufficient statistics of both trajectories (MD runs and ED sampling run) to achieve full convergence of the essential subspace in such high dimensional configurational space. From the figure it is also clear that

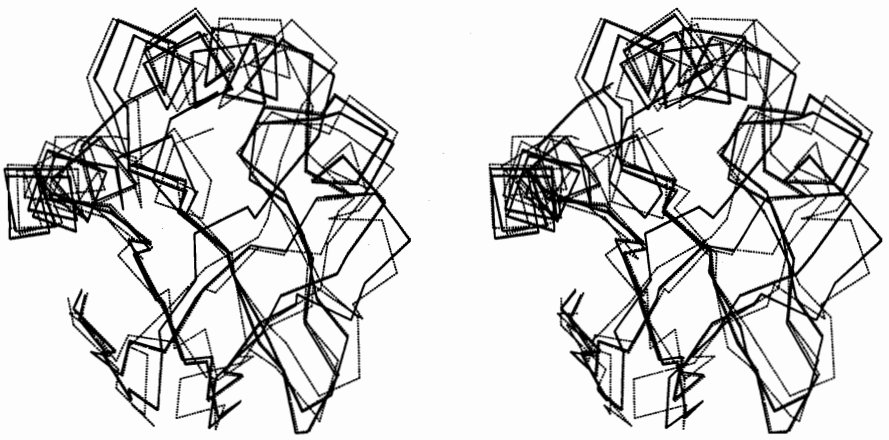


Figure 4: Superposition of the structures having the minimum and maximum accessibility (solid) and number of residues in random coil conformation (dotted) during the ED sampling procedure.

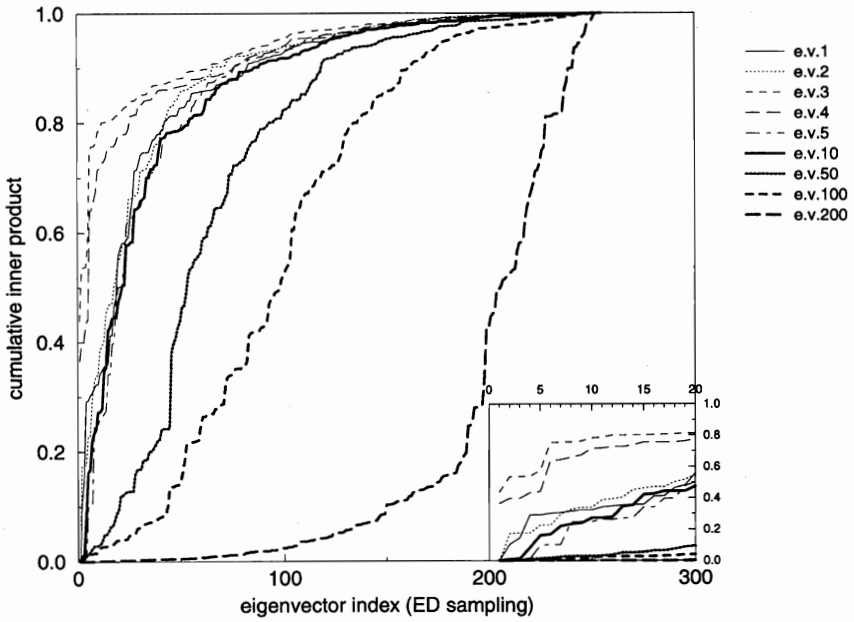


Figure 5: Cumulative square projections (inner products) of eigenvectors 1-5, 10, 50, 100 and 200 obtained from the three MD runs, over the whole eigenvectors set of the ED sampling.

typical near constraint eigenvectors, like eigenvectors 50, 100 and 200, calculated from the MD runs do not have any mixing with the essential subspace of the ED sampling (approximately the first 10 eigenvectors). This significant correspondence between the essential subspace obtained by the ED sampling and the essential eigenvectors obtained from the initial three MD runs, suggests that an initial MD simulation of a few hundreds picoseconds could be enough to obtain a reasonable basic approximation of the fully converged essential subspace of a protein, in order to start the ED sampling procedure.

Finally as a first investigation of the physical properties of the essential subspace we studied the average behavior of motions along the essential coordinates. The average square distance of the projections onto eigenvectors 1, 3 and 4 (with respect to their starting positions) is shown as a function of time in Fig. 6. Thirty stretches of 20 ps were taken from the three free MD simulations to produce these data. The linear dependence of the average square distance on time indicates diffusion-like behavior of the motions in the essential subspace with a diffusion constant (for these essential coordinates) of roughly 4:4 x 10^{-10} m $_2$ s^{-1} . This suggests that there are no large free energy differences between regions inside the boundaries of the essential subspace.

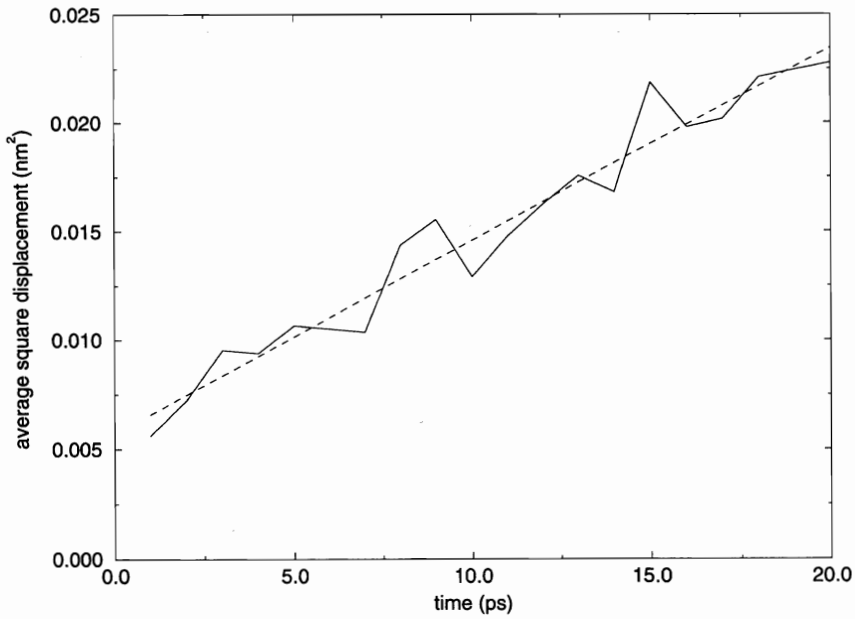


Figure 6: The average square displacement per degree of freedom of 30 parts of 20 ps free MD of the combination of the projections along eigenvectors 1, 3 and 4, from their starting positions. The solid line corresponds to the actual square distances, the dashed line is a linear regression to the experimental data.

Conclusions

The CPU time needed for the calculations described here correspond to 372 ps of usual MD for the system that we studied. As was shown in Fig. 1a and 1b, the volume of the configurational space which was sampled with this new method is significantly larger than the volume sampled during the three conventional MD runs, which added up to a total of 600 ps. Considering that the three free MD simulations were started from three independent structures from the NMR cluster (which makes the volume of the essential subspace sampled by these simulations already significantly larger than it would have been in case one single simulation of 600 ps would have been performed), we can conclude that with this procedure we can sample the configurational space of proteins more efficiently than with usual MD. Since the geometrical properties which were evaluated to check the quality of the protein structures produced by this procedure (Fig. 2) fluctuate in a range comparable to the range in which these properties fluctuate in usual MD simulations, we have no indication that the constraints applied influenced the protein structure in an undesirable fashion. Hence, we have been able to generate a large ensemble of structures quite distant from the starting structure but with comparable geometrical and physical properties (Figs. 2, 4). In addition the results suggest (Fig. 5) that a MD simulation of a few hundred picoseconds could be used to define approximately the exact essential subspace of a protein to start the ED sampling. It should be noted that the ED sampling procedure cannot guarantee the discovery of possible distinct conformational states, of which no information was available. Finally the diffusion-like behavior of the essential coordinates (Fig. 6) suggests that different regions of the essential subspace should have comparable free energies. We are currently improving the ED sampling method and we are performing an extended ED sampling of a 13 residues peptide in order to fully explore the boundaries of the (backbone) configurational space.

Acknowledgments

The authors would like to thank N.A.J. van Nuland for the MD simulations of HPr and R. M. Scheek and G. T. Robillard for many stimulating discussions. A. Amadei was financially supported by The Training and Mobility of Researchers (TMR) Program of the EC.

Appendix A

In this appendix we describe the constraint for a radius expansion or contraction with respect to a reference position in the essential subspace. In this case our constraint G (eq. 4) is:

$$G(\boldsymbol{\xi}(t+\Delta t); t+\Delta t) = |\boldsymbol{\xi}(t) + \Delta\boldsymbol{\xi}_d + \Delta\boldsymbol{\xi}_c - \boldsymbol{\xi}_0|^2 - r^2(t+\Delta t) = 0 \qquad 8$$

$\boldsymbol{\xi}_0$ is the reference position and $r^2(t + \Delta t)$ is a prescribed positive function of time defining the expansion or the contraction. Equation 6 can be written as:

$$\Delta\xi_c^i - 2\lambda(\xi^i(t) + \Delta\xi_d^i + \Delta\xi_c^i - \xi_0^i) = 0 \qquad 9$$

or

$$\Delta\xi_c^i = \frac{2\lambda}{1-2\lambda}(\xi^i(t) + \Delta\xi_d^i - \xi_0^i) \qquad 10$$

Now, by inserting eq. 10 into eq. 8, and by defining $\varepsilon = \frac{2\lambda}{1-2\lambda}$ we obtain:

$$\sum_i [\xi^i(t) + \Delta\xi_d^i + \epsilon(\xi^i(t) + \Delta\xi_d^i - \xi_0^i) - \xi_0^i]^2 - r^2(t+\Delta t) = 0 \qquad 11$$

or

$$\sum [(\epsilon+1)(\xi^i(t) + \Delta\xi_d^i) - (\epsilon+1)\xi_0^i]^2 - r^2(t+\Delta t) = 0 \qquad 12$$

and therefore

$$(\epsilon + 1)^2 = \frac{r^2(t + \Delta t)}{|\xi(t) + \Delta\xi_d - \xi_0|^2} \quad 13$$

or

$$\epsilon = -1 + \frac{r(t + \Delta t)}{|\xi(t) + \Delta\xi_d - \xi_0|} \quad 14$$

combining equs. 10 and 14 we finally obtain:

$$\Delta\xi_c = [\xi(t) + \Delta\xi_d - \xi_0][-1 + \frac{r(t + \Delta t)}{|\xi(t) + \Delta\xi_d - \xi_0|}] \quad 15$$

which gives at every time step Δt the correction that has to be applied. The results presented in this paper were generated by using $r(t)=\gamma t$ with γ being a small positive constant (0.0004 nm per step), and using only three of the essential eigenvectors (1, 3 and 4) to define ξ.

References and Footnotes

1. A. Amadei, A. B.M. Linssen, and H.J.C. Berendsen, Essential Dynamics of Proteins, *Proteins: Str., Funct., and Gen., 17* , 412-425 (1993).
2. D.M.F. van Aalten, A. Amadei, R. Bywater, J.B.C. Findlay, H.J.C. Berendsen, C. Sander, and P.F.W. Stouten. A comparison of structural and dynamic properties of different simulation methods applied to SH3. Submitted to the Biophysical journal.
3. D.M.F. van Aalten, A. Amadei, G. Vriend, A.B.M. Linssen, G. Venema, H.J.C. Berendsen, and V.G.H. Eijsink, The essential dynamics of thermolysin - confirmation of hinge-bending motion and comparison of simulations in vacuum and water, *Proteins: Str. Funct., and Gen. 22* , 45-54 (1995).
4. R.M. Scheek, N.A.J. van Nuland, B.L. de Groot, and A. Amadei, Structure from NMR and mol ecular dynamics: distance restraining inhibits motion in the essential subspace, *J. Biomol. NMR., 6*, 106-111 (1995).
5. T.D. Romo, J.B. Clarage, D.C. Sorensen, and G.N. Phillips Jr., Singular Value Decomposition Analysis of Time-Averaged Crystallographic Refinements, *Proteins: Str., Funct., and Gen. 22*, 311-321 (1995).
6. A. Amadei, A. B. M. Linssen, B. L. de Groot, and H. J. C. Berendsen. Essential degrees of free dom of proteins, *Modelling of Biomolecular Structures and Mechanisms*, A. Pullmann et al., editor, pages 85-93, The Netherlands, 1995. Kluwer.
7. M. Levitt, C. Sander, and P. S. Stern, Protein normal-mode dynamics-trypsin-inhibitor, crambin, ribonuclease and lysozyme, *J. Mol. Biol., 181* , 423-447 (1985).
8. A. E. Garcia, Large-amplitude nonlinear motions in proteins, *Phys. Rev. Lett., 68*, 2696-2699 (1992).
9. Nico A.J. van Nuland, Ilona W. Hangyi, Ren e C. van Schaik, Herman J.C. Berendsen, Wilfred F. van Gunsteren, Ruud M. Scheek, and George T. Robillard, The high-resolution structure of the histidine-containing phosphocarrier protein HPr from Escherichia coli determined by restrained molecular dynamics from NMR-NOE data, *J. Mol. Biol., 237* , 544-559 (1994).
10. W.F. van Gunsteren and H.J.C. Berendsen. Gromos manual. BIOMOS, Biomolecular Software, Laboratory of Physical Chemistry, University of Groningen, The Netherlands, 1987.
11. Wolfgang Kabsch and Christian Sander, Dictionary of Protein Secondary Structure: Pattern recognition of Hydrogen-Bonded and Geometrical features, *Biopolymers, 22*, 2577-2637 (1983).
12. R.A. Laskowski, M.W. MacArthurand D.S. Moss, and J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl Cryst., 26* , 283-291 (1993).
13. G. Vriend, WHAT IF: a molecular modeling and drug design program, *J. Mol. Graph., 8*, 52-56 (1990).
14. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, *Intermolecular Forces*, B. Pullmann, editor, page 331. Reidel, Dordrecht, 1981.
15. H.J.C. Berendsen, J.P.M. Postma, A. DiNola, and J.R. Haak, Molecular Dynamics With Coupling to an External Bath, *J. Chem. Phys., 81*, 3684-3690 (1984).
16. J.P. Ryckaert, G. Ciccotti, and H.J.C. Berendsen, Numerical Integration of the Cartesian Equations of Motion of a System with Constraints; Molecular Dynamics of n-Alkanes, *J. Comp. Phys., 23* , 327-341 (1977).

Date received: July 18, 1995

Communicated by the Editor C. W. Hilbers