Towards an Exhaustive Sampling of the Configurational Spaces of the Two Forms of the Peptide Hormone Guanylin

B.L. de Groot[1], A. Amadei[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

The recently introduced Essential Dynamics sampling method is extended such that an exhaustive sampling of the available (backbone) configurational space can be achieved. From an initial Molecular Dynamics simulation an approximated definition of the essential subspace is obtained. This subspace is used to direct subsequent simulations by means of constraint forces. The method is applied to the peptide hormone guanylin, solvated in water, of which the structure was determined recently. The peptide exists in two forms and for both forms, an extensive sampling was produced. The sampling algorithm fills the available space (of the essential coordinates used in the procedure) at a rate that is approximately six to seven times larger than that for traditional Molecular Dynamics. The procedure does not cause any significant perturbation, which is indicated by the fact that free Molecular Dynamics simulations started at several places in the space defined by the Essential Dynamics sample that complete space. Moreover, analyses of the average free Molecular Dynamics step have shown that nowhere except close to the edge of the available space, there are regions where the system shows a drift in a particular direction. This result also shows that in principle, the essential subspace is a constant free energy surface, with well-defined and steep borders, in which the system moves diffusively. In addition, a comparison between two independent essential dynamics sampling runs, of one form of the peptide, shows that the obtained essential subspaces are virtually identical.

Introduction

Recently, the structure of the peptide hormone guanylin was elucidated by NMR [1]. Two distinct conformations were found, denoted A and B, present in equal amounts, which differ in the way two internal disulphide bridges are arranged with respect to the main chain of the peptide. The hormone as it was studied consists of 13 residues and the two conformations can be classifed as a right handed spiral (A form) and a left handed spiral (B form) [1]. Interchange between the two forms was not observed experimentally, a finding which was supported by computational methods [1].

Guanylin is an endogenous ligand to the heat stable enterotoxin receptor (STaR), an intestinal guanylyl cyclase [2], causing the production of cyclic GMP when activated. For a review, we refer to [3]. Cyclic GMP plays an important role in fluid regulation in the intestines and overproduction leads to severe diarrhea [4]. Guanylin competes with heat stable enterotoxins (STa) in binding to STaR and is homologous to it [2, 5].

*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

Recently, the Essential Dynamics (ED) technique [6] was extended by introduction of a sampling technique that makes use of constraint forces in the essential subspace, where most relevant motions occur [7]. Here, an improved algorithm of this ED sampling technique is presented, which causes less perturbation and performs a rapid filling of the essential subspace. The method is applied to both forms of guanylin, where borders of the allowed region are found in almost every direction, indicating an almost complete sampling. The allowed space found by this method coincides with the space that would be found when a MD simulation would be extended to infinite time. This is shown by starting (free) MD simulations at several places at the border of the essential subspace.

The allowed spaces of the two forms are compared to each other, to the spaces sampled by the initial MD and to the spaces sampled by free MD simulations started at a number of positions in the sampled region. Moreover, we investigated the possible influence of the initial definition of the essential eigenvectors on the results obtained from an extended ED sampling. Careful investigation of the dynamical behaviour of the essential coordinates both during the ED sampling procedure and during MD simulations started at several distinct places in the essential subspace indicate a diffusive behaviour in a constant free energy basin, with well defined borders.

Methods

MD simulations

MD simulations of both the A and the B state were initiated from a corresponding NMR structure (the first structures of the PDB entries 1gna and 1gnb respectively). In both cases, the peptides were surrounded by SPC water molecules [8] (571 and 511 solvent molecules respectively) filling up a truncated octahedron box. The net negative charge was compensated for by a sodium ion which was placed by substitution of the water molecule at lowest potential. Both systems were energy-minimized after which a heatup procedure was used to equilibrate the systems. Subsequently, both systems were simulated for 1 ns, of which the last 750 ps were used for analyses. The temperature was kept constant at 300 K by weak coupling to an external bath [9] (peptide and solvent were coupled separately with a coupling constant of 0.1 ps). The pressure was also kept constant by coupling to a bath with a coupling constant of 0.5 ps. SHAKE [10] was used to constrain bond lengths, allowing a time step of 2 fs. All calculations were performed with the simulation package Gromos [11]. All structure evaluations and visualizations were performed with the program WHAT IF [12].

Initial definition of essential subspace

From the structures of the last 750 ps of both simulations covariance matrices of positional fluctuations (C-α only) were built and diagonalized. Eigenvectors are directions in configurational space and the corresponding eigenvalues indicate the mean square uctuations along these axes [6]. The procedure corresponds to a linear multidimensional leat squares fitting of a trajectory in configurational space [13, 14]. Sorting the eigenvectors by the size of the eigenvalues shows that the configurational space can be divided in a low dimensional (essential) subspace in which most of the positional fluctuations are confined, and a high dimensional (near-constraints) space in which merely small uninteresting vibrations occur.

ED sampling protocol

With ED, all relevant motions, i.e. those with an appreciable amplitude, can be (approximately) described by only a few collective coordinates representing a small fraction of the total number of degrees of freedom. As was shown before [7], this can be used to sample the configurational space more efficiently than by tradi-

tional MD. In a previous paper [7] we introduced the concept of constraint dynamics, applying constraints in the essential subspace in the form of an expanding radius (spanned by e.g. three essential coordinates). Here, a modification of that protocol is introduced. Instead of performing an expansion of a radius with a fixed increment in the radius per step, a choice is now made every step between expanding the radius or keeping the radius fixed at the current value, depending on the direction a normal MD step would have taken.

So, at every (usual MD) step an evaluation is made. If the new position in the chosen essential subspace is further from the starting position than the position in the previous step, no correction is applied. If, however, the new position is closer to the starting position, it is moved back, by means of a constraint force on the alpha-carbon eigenvectors, to a position which has the same distance to the starting position. From the new position the velocities are recalculated. Using the principle of least perturbation, the correction is performed in the direction of the radius vector as described in [7]. In this way, the distance from the origin is not forced to increase if it will not spontaneously do so. Instead, it will move on a sphere with fixed radius, until a direction is found in which the system can expand. If the radius does not increase for a certain time (in this case a criterium of 500 subsequent steps was used), indicating that the system approaches a border, a new expansion cycle is started. The last configuration is used as a center of the new expansion sphere.

To avoid oscillation in a particular direction in subsequent expansion cycles, in every cycle an initial linear expansion (of 1000 steps) along one of the eigenvectors used for the radius expansion is performed. The eigenvector used for this linear expansion and the direction are chosen randomly. The principle for such a linear expansion is the same as for the radius expansion: a step is accepted if the distance from the origin increases in an unperturbed MD step. When the distance decreases, it is put back to the original value.

Thus, there are three major differences with respect to the original ED sampling protocol [7]. First, during expansion cycles, a constraint is only applied to prevent the system from going back, not to push it further from the original position. In this way, the system is not forced to move in unfavourable regions, and expansions will stop automatically if a border is reached. Second, the size of the expansion step is not fixed but is determined by the usual MD step, causing least perturbation at the most efficient expansion speed. Third, the initial linear expansion in an arbitrary direction forces the system to move in a direction other than the reverse of the previous cycle, causing a more rapid filling of the allowed space. The software used is an adaptation of the simulation package Gromos [11].

During the ED sampling of the two states, all MD parameters were kept at the same values as during the free simulations. For both states, a three-dimensional ED sampling (using the first three alpha-carbon eigenvectors from each state respectively) of 100 cycles was performed. Because this calculation showed similar behaviour for the two states, it was decided to concentrate further studies on the A state. An additional 100 cycles of the three-dimensional ED sampling were performed for the A state to investigate the completeness of the ED sampling. Also for the A state, several free MD simulations of 100 ps each were started at the borders to investigate if the allowed space as defined by the ED sampling algorithm is consistent with the behaviour of free MD simulation, i.e. to check the stability of the essential subspace.

Finding borders in the essential subspace

As stated above, in the expansion cycles, the expansion stops when there is no spontaneous increase of the distance from the origin anymore. This will cause the procedure to stop when a border of the allowed region in the essential subspace is
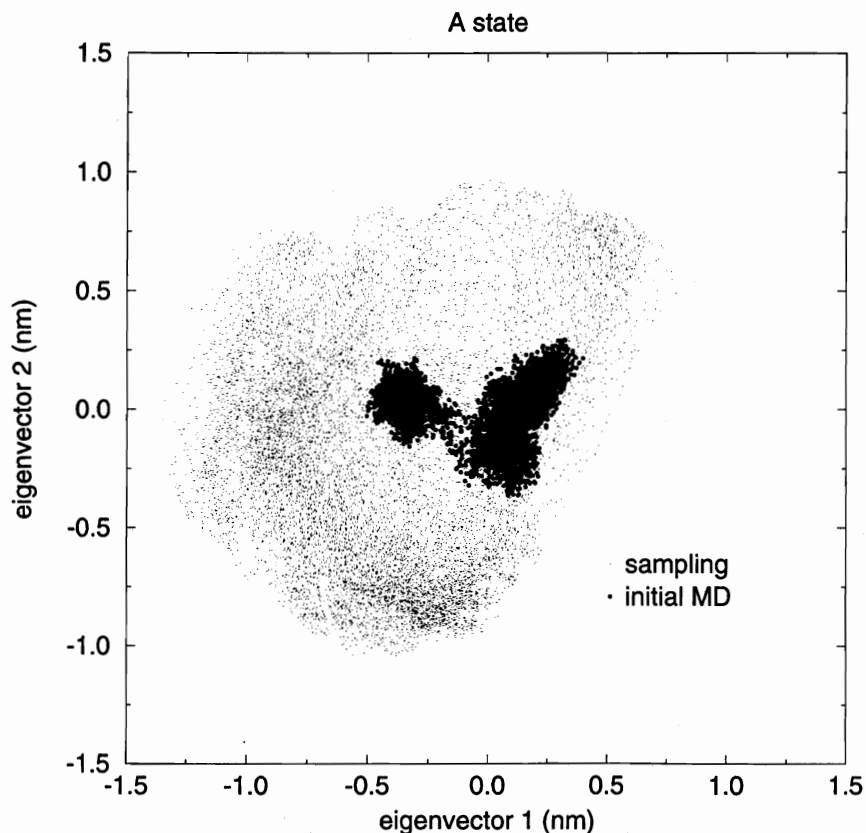
A state

Figure 1: A comparison between the region sampled by the initial free MD simulation of 750 ps and by the ED sampling procedure projected in the plane defined by eigenvectors 1 and 2 from the free MD simulation. A state.

reached. The constraint used to prevent the system from moving back towards the origin of the expansion makes the system move along a sphere with fixed radius, causing additional sampling of the border region.

To define the location of the borders quantitatively, the average free step of the essential coordinates during the free MD steps (so excluding corrections in the essential positions) was evaluated in every position of the essential space, using a grid. In a position not near a border for a specific coordinate, the average free step vector is expected to be zero (indicating an equal probability to move in each direction). So a non-zero average free step indicates the proximity of a border.
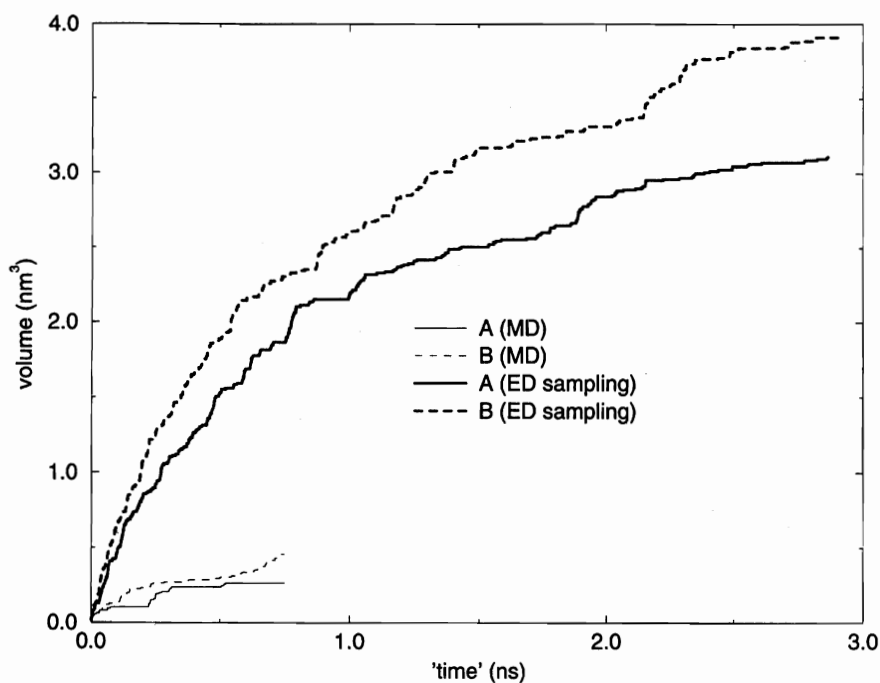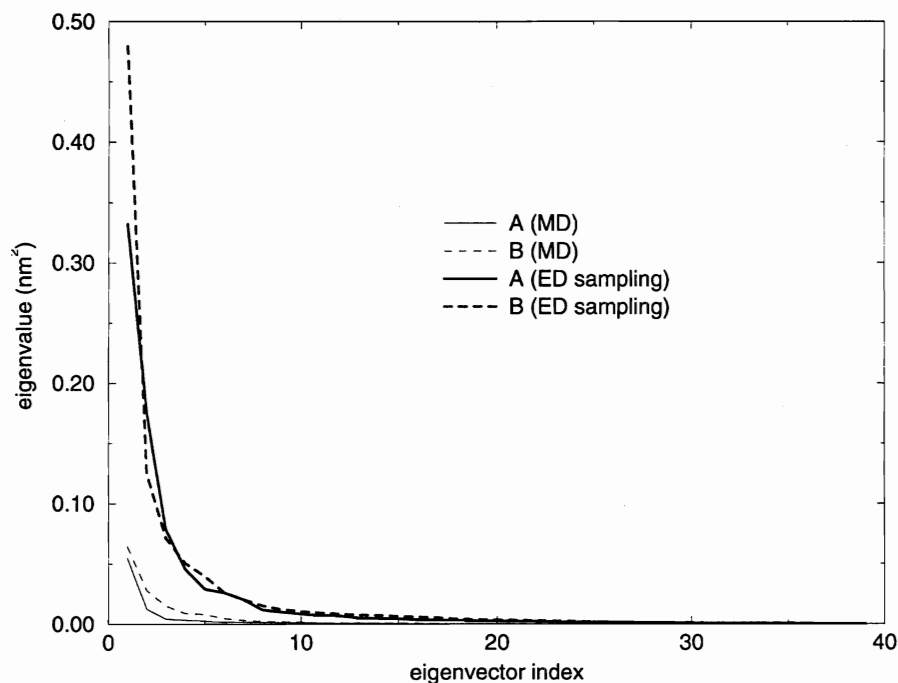

Figure 2: The sampled volume calculated over a grid in the space defined by the first three eigenvectors of each state as a function of the number of integration steps, which corresponds to time in a free MD simulation.

Figure 3: The eigenvalues obtained from an initial MD simulation of 1 ns (of which the last 750 ps were used for analyses) for both the A and B state, as well as those calculated from the set of structures built by an extensive ED sampling.

An ED sampling of the A state using only the first two eigenvectors instead of the first three was performed to investigate the average free step in detail on the grid defined by the first two eigenvectors.

Calculation of the configurational volume

To obtain a quantitative measure of the sampled configurational volume, a cubic grid was put over the space spanned by the first three eigenvectors, which were used in the ED sampling protocol. During the ED sampling procedure the number of non-empty grid elements was multiplied by the volume per grid element to give an estimate of the evolution of the sampled configurational volume in these three dimensions. The grid size must be carefully chosen to represent the sampled volume correctly. A fine grid underestimates the volume and makes it proportional to
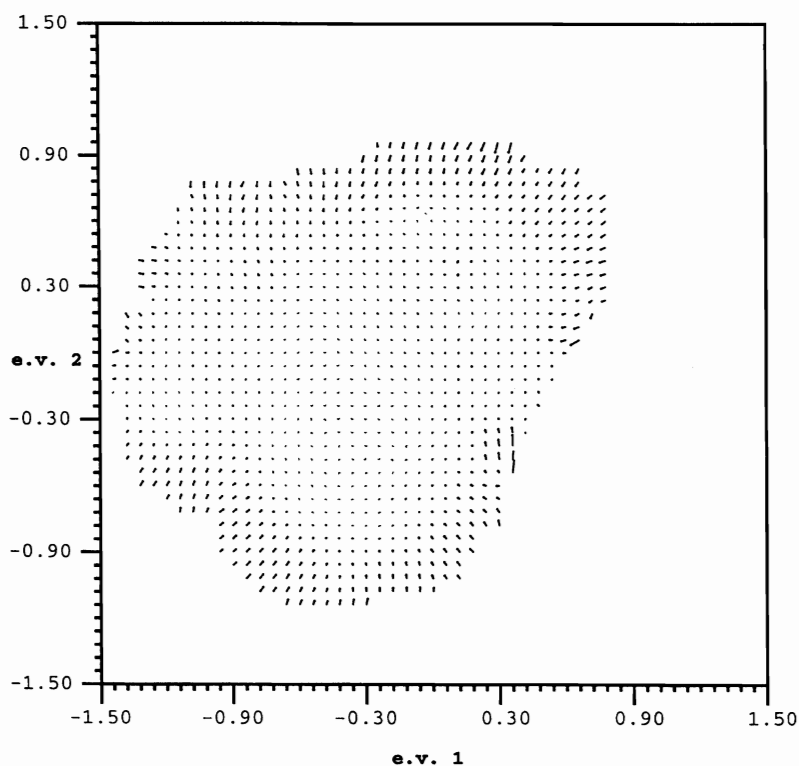

Figure 4: The average free steps calculated over a grid projected in the plane defined by the first two eigenvectors. The arrows indicate the direction and size that a free MD step would take on average in each position. A state.

the number of sampled points, while a coarse grid may introduce incorrect connectivity. A suitable compromise was found when for each of the first three eigenvectors 10 intervals were chosen between -2 nm and 2 nm, dividing the 3D space in 1000 grid elements. For this grid size the volume is practically independent of the density of sampled points.

Results

To estimate the efficiency of the ED sampling protocol, three evaluations were done. First, projections of the trajectories produced by the expansion cycles onto the three planes defined by the first three eigenvectors were compared to the projections of the free MD simulations onto these planes (Fig. 1). For both conformations, the ED sampling run has not only been able to reproduce the complete region that had been sampled by MD, but has significantly enlarged that region in every direction. Second, to obtain a more quantitative measure of the efficiency of the ED sampling protocol, the volume of the space sampled in three dimensions as a function of the number of integration steps was compared for the ED sampling runs and the MD simulations (Fig 2). The slope of the curves is a measure of the efficiency of the sampling protocol, and for both conformations, the ED sampling method produces a significantly steeper plot than the MD, indicating a high efficiency of the ED sampling protocol. The ratio of the slopes of the straight lines fitted to the volume curves of the ED sampling technique and MD was approximately 6 to 7 for both states.

The curves of the volume (Fig. 2) corresponding to the two ED sampling runs both start to level off after approximately 1 million integration steps (corresponding to 2 ns of simulation with a time step of 2 fs), indicating that the allowed space defined by the first three eigenvectors has been completely sampled.

Third, in Figure 3, we compare the eigenvalues of the ED sampling with the eigenvalues of the initial MD runs. This figure shows that the first ten eigenvalues from the ED sampling are much larger than those from the MD simulations. This again
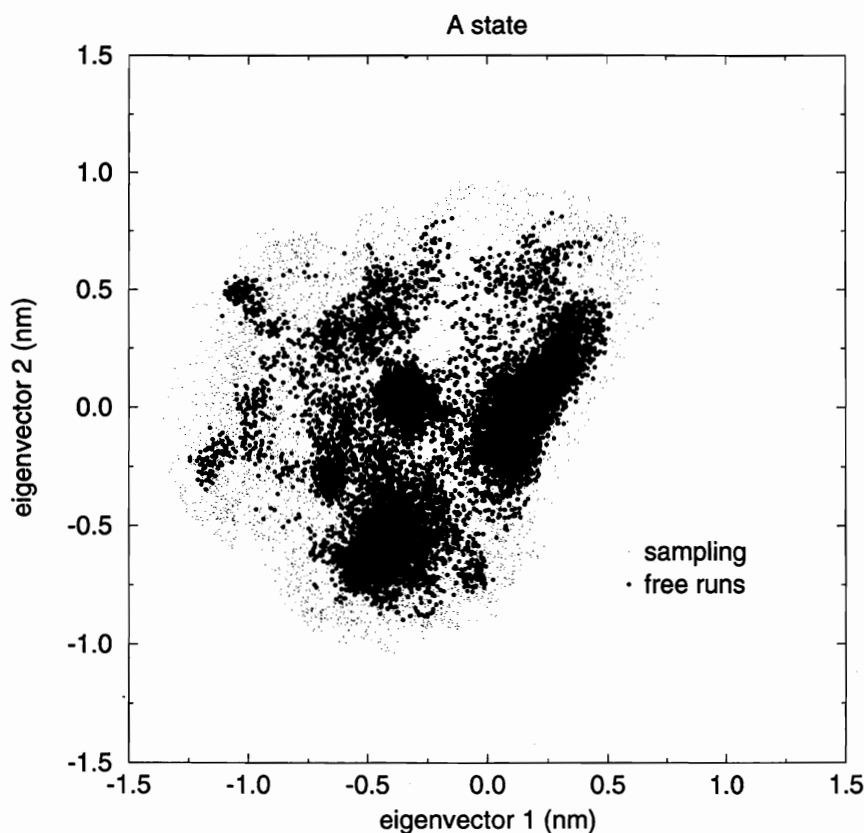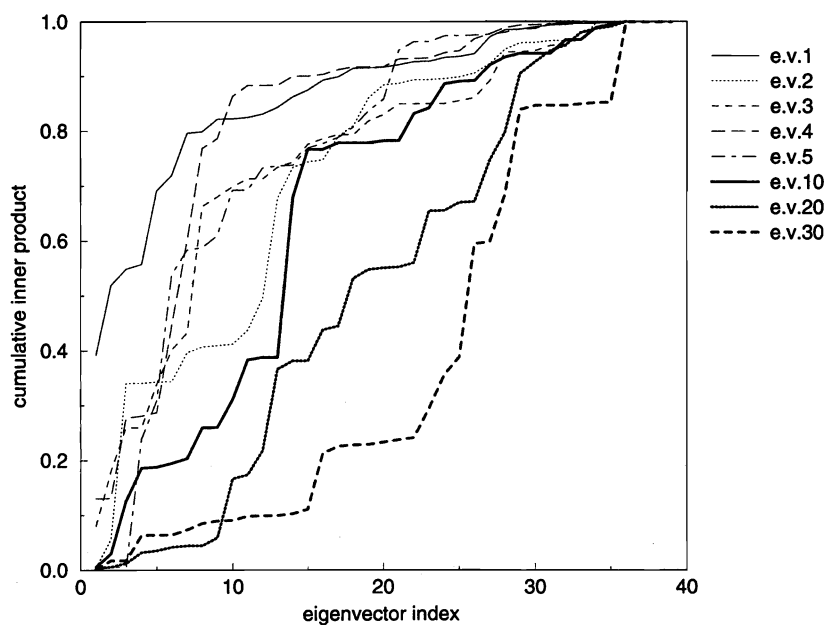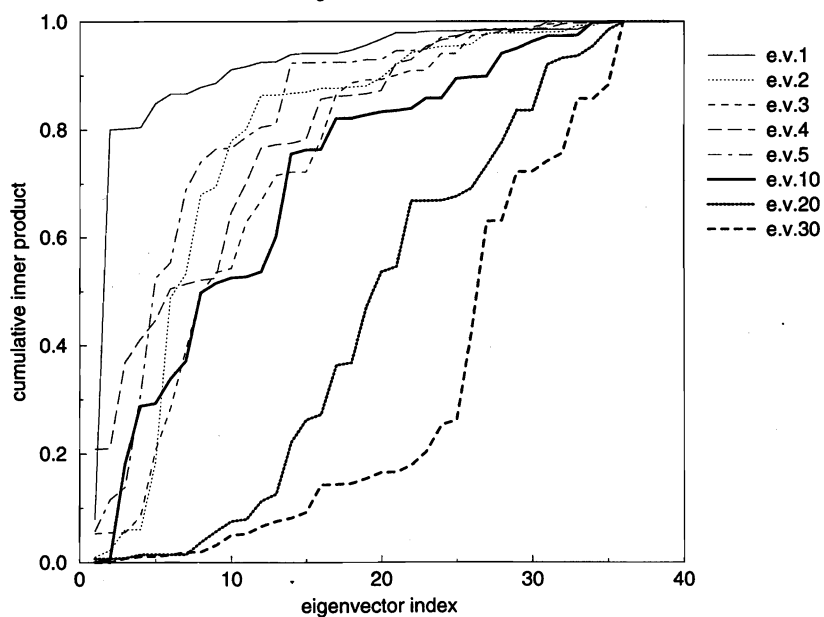

Figure 5: A comparison between the region sampled by the ED sampling procedure and by multiple free simulations of 100 ps started at random places in that space. The structures are projected in the plane defined by eigenvectors 1 and 2 from the initial free MD simulation. A state.
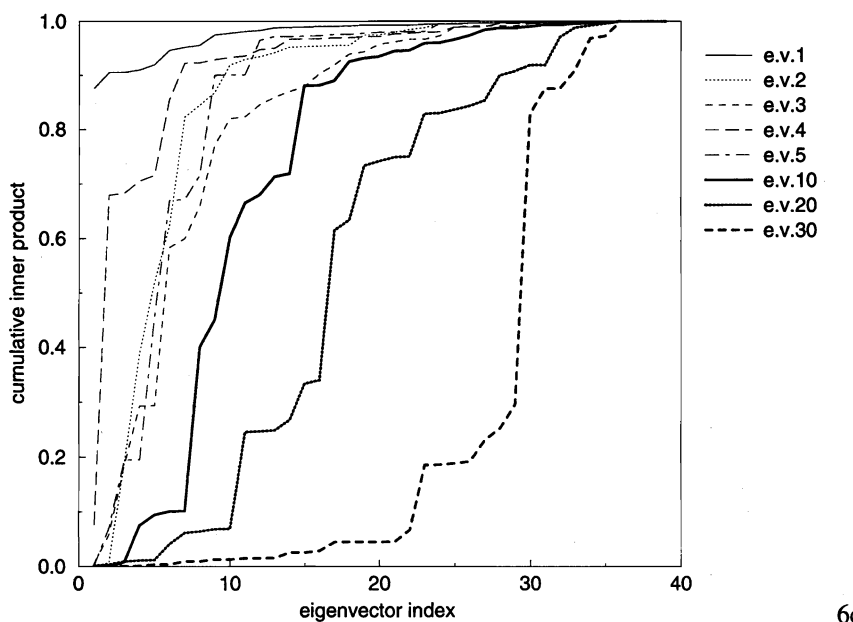
6a

6b

6c

Figure 6a and 6b: Cumulative square inner products between eigenvectors obtained from two 1 ns free MD simulations and eigenvectors built from the complete collection of structures obtained from ED sampling. Along the X axis are the eigenvector indices of the free MD simulation which are used to rebuild single eigenvectors obtained from sampling. A state. Figure 6a shows the results from one MD simulation and figure 6b of another simulation started in a different region of the essential subspace. Figure 6c shows the cumulative inner products between two sets of eigenvectors, obtained from two independent ED sampling runs. A state.

indicates that a much larger essential subspace volume has been covered.

The (2D) ED sampling of the A state in the plane defined by the first two eigenvectors samples the same region in the 1-2 plane as does the 3D ED sampling. The third dimension is somewhat less well sampled because it was not forced to sample the borders, but the sampled 3D volume was very close to that of the three-dimensional ED sampling. Fig. 4 shows the average free step in the plane defined by eigenvectors 1 and 2, calculated from the two-dimensional ED sampling of the A state. The arrows indicate the size and direction of the average free step in every point, spread over a square grid. Almost anywhere close to the border of the sampled essential subspace, the average free step is non-zero and points towards the center of the allowed region. This indicates that the sampled space coincides with almost the complete available space of the A state, in this subspace.

To investigate the effect of the ED sampling algorithm on the definition of the borders of the essential subspace, free MD runs were started at several places at the edges of the sampled space of the A state. If the borders are really located as indicated by the non-zero average free steps, the region where the average free step is almost zero should also be available in a free simulation. Fig. 5 shows the projections of the structures generated by these free runs as well as those from the 2D ED sampling. All free runs move away from the edges for a short time in the direction of the center of the allowed region to fill in the complete allowed region, leaving only the edges unsampled indicating that the whole space produced by the ED sampling is accessible to dynamics.

Energies produced by free runs in regions distinct from the region sampled by the initial free MD simulation of the A state showed no significant differences from the energies produced by the initial free run. The average potential energy of the initial free MD simulation is -25.62 MJ/mol with a standard deviation of 0.17 MJ/mole.
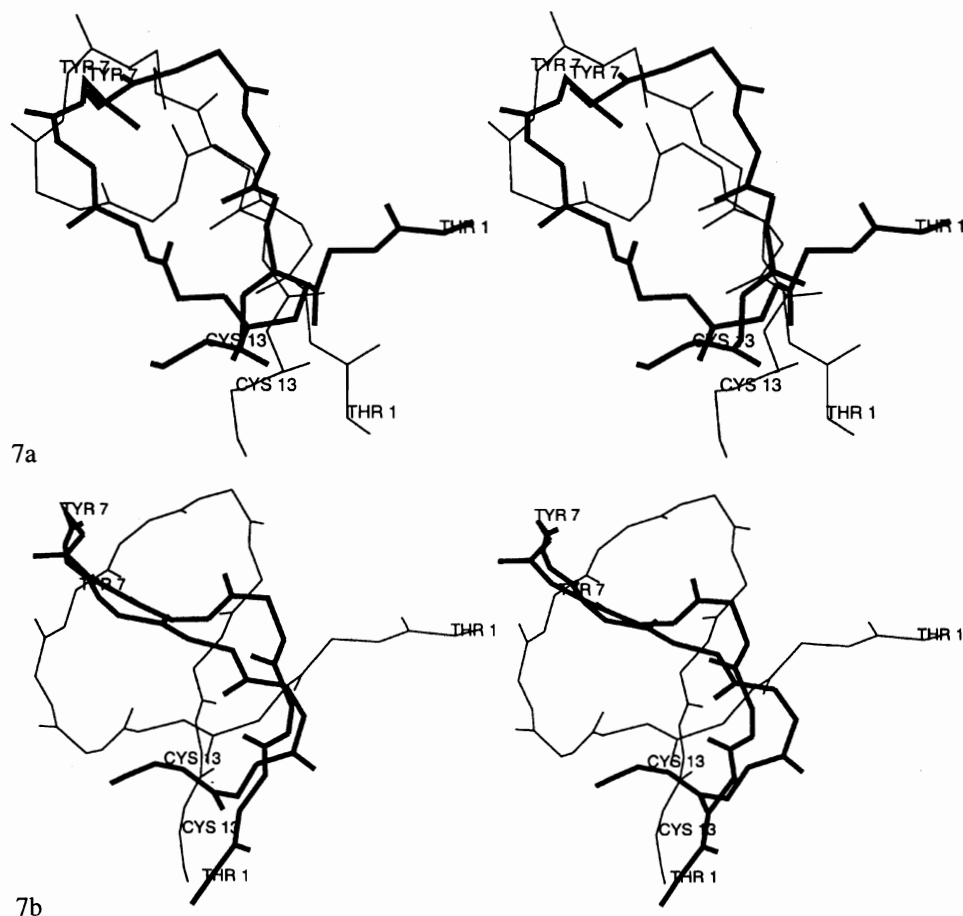

7a

7b

Figure 7: Comparison between structures obtained from free MD simulation and ED sampling. A state. Figure 7a: MD, 7b: ED sampling. In both figures, a stereo picture of the structures corresponding to the minimum and maximum sampled position along eigenvector 1 are shown.

For three free simulations in different parts of the essential subspace the averages were -25.66 MJ/mol (with a standard deviation of 0.16 MJ/mole), -25.64 MJ/mol (0.18 MJ/mol) and -25.69 MJ/mol (0.15 MJ/mole) respectively. Also, energies produced during the ED sampling are similar to energies from free MD simulations, in regions not near the borders (average energy: -25.75 MJ/mol, standard deviation: 0.32 MJ/mole). Thus, the fluctuation of the energy is significantly larger during the ED sampling procedure, whereas the average energy is close to the average obtained in a completely free simulation.

To investigate the accuracy of the definition of the essential subspace, i.e. to check if the essential coordinates and near-constraints are consistent in different MD simulations of about 1 ns, a covariance matrix was built and diagonalized for two (uncorrelated) trajectories of the A state of 1 ns each, started in distinct regions of the initial essential subspace. These analyses are compared to an analysis of the structures produced by the 3D ED sampling of the A state.

A way to compare two sets of eigenvectors is to monitor the cumulative square inner product of one eigenvector from one set with all eigenvectors from the other set. This sum will converge to 1.0 because all eigenvectors of one set will always be able to rebuild the other set. Figs. 6a and 6b show this cumulative square inner product between single eigenvectors of the two 1 ns MD runs and those of the whole ED sampling set. Although the analyses of the simulations contain an appreciable amount of noise, they show a considerable similarity of essential subspaces aproximately defined by, e.g., the first five degrees of freedom. It is also important to note that typical near-constraint eigenvectors of the MD runs (like 20 and 30) do not mix at all with the essential eigenvectors of the ED sampling run. Therefore, 1 ns of free simulation is enough to obtain a basic description of the fully converged essential subspace (as was also found previously [7]) and the definition of the essential subspace is consistent in different parts of the configurational space.

Figure 6 c shows the cumulative square inner product between the eigenvectors from the ED sampling of the A state, and those from another,independent, ED sampling of the A state. The latter sampling was produced by constraining the position along eigenvectors constructed from a different initial MD simulation. Compared to Figs. 6a and 6b, the similarities between the two sets are much higher. This shows that both sets have converged to the same definition of the space. Compared to Figs 6a and 6b, a much better definition of the essential subspace is obtained because the statistics of the covariance matrix is better when it is based on a complete ED sampling rather than a 1 ns MD run.

Fig. 7 shows different structures of the A state produced by the initial MD compared to structures produced by the ED sampling algorithm. The structures produced by the ED sampling deviate much more from the starting structure than the ones produced by the initial MD, illustrating the fact that a much larger part of the configurational volume has been sampled.

Comparison of the structures produced by the ED sampling procedures of the A with those of the B state showed that there is no overlap between the configurational spaces available to the two forms. Also, direct attempts to drive the system from one state to the other (also using the method of least perturbation [7]), constraining the position along eigenvectors that define the differences between the A and B state, failed to accomplish a transition from one state to the other. This suggests that the free energy barrier to move from one state to the other is too high to be passed, and that under usual circumstances the two forms each have their own distinct essential subspaces, with no overlap. This is consistent with experimental data [1], which shows distinct species on a time scale of at least seconds.

Conclusions and Discussion

The results shown before prove that with the ED sampling technique, improved in

this paper, it is possible to approach an almost complete sampling of the essential subspace of a small peptide in water, within a number of integration steps comparable to a simulation time of about 3 ns. The initial slope of the curve of the sampled volume in the subspace defined by the first three eigenvectors which is a measure of the efficiency of a sampling protocol, indicates that the ED sampling algorithm is six to seven times more efficient than usual MD. Therefore, the cluster obtained by ED sampling is much larger than that produced by usual MD of comparable lenghth. This means that for the macroscopical properties that are evaluated by averaging over the ensemble of collected structures, results can be expected to differ between the two clusters. This is especially true for properties that are sensitive to the extreme structures in the ensemble, i.e. properties that depend on the spread in the cluster, rather than on the average.

The fact that the 2D ED sampling reproduces almost all of the 3D space that was obtained by the 3D ED sampling procedure suggests that the use of three essential dimensions is sufficient to obtain a complete sampling of the essential subspace.

The analysis of the average free steps suggests that, apart from the borders, no real free energy gradients are present in the essential subspace. This implies that unconstrained dynamics can be considered as a random walk in the essential subspace, resulting in diffusion like behaviour, as previously observed in the protein HPr [7]. Only those regions where an appreciable average free step is measured (0.0004 nm), remain unsampled by the free MD simulations (Figs. 4 and 5). This suggests that a non zero average free step is a good indication of a border (or a less favourable region) in the essential subspace and that no significant perturbation is induced in the sampling algorithm.

The ED sampling algorithm searches for possible expansion directions until no progress is found anymore, in every single cycle. Numerous of such cycles have been completed for both forms of the peptide. The results of the average free step calculation (Fig. 4) show that, for the A state, in virtually every direction of the essential subspace, a border has been found, indicating that the search in these directions is complete. This suggests that it is improbable that there are paths accessible for the system towards other stable regions in the same essential subspace.

Due to the soft nature of the applied constraints, no unallowed regions have been sampled. This is illustrated by the fact that, for the A state, free MD simulations have filled in the complete allowed space as defined by the ED sampling. Moreover, the free runs span one closed region which indicates that no physical barriers have been passed by the sampling algorithm.

Essential subspaces defined locally from MD simulations in two different parts of the configurational space are similar to the essential subspace defined from the complete ED sampling (Fig 6), which validates the use of a rough initial definition of the essential subspace.

The fact that the two eigenvector sets obtained from two independent sampling runs are so similar (Fig. 6) proves that the definition of the space by this ED sampling protocol is consistent in itself, and not dependent on the initial MD simulation from which eigenvectors are extracted that are used to explore the space. Therefore, the approximated essential subspace defined from an initial MD simulation can indeed be used in an extrapolation protocol to sample the complete allowed space and to refine the description of the essential subspace.

Structures in the NMR cluster are much closer to each other for the A state (mean backbone RMSD with respect to the average structure of 0.47 Å) than for the B state (mean RMSD of 1.07 Å), suggesting more configurational freedom for the B state than for the A state [1]. This is partially reflected by the available configura-

tional volume as obtained from the ED sampling runs of both the A and B states (Fig. 2),which shows a larger accessible volume for the B state than for the A state. These volumes, however, are much larger than those calculated from the NMR clusters. We think that the refinement protocol used to produce the NMR structures might give a too rigid representation of the molecule, because those structures are selected that simultaneously fulfil most NOE restraints. Thus the cloud of structures produced is close to the average structure, and the differences between the structures are not necessarily a good indication of the main modes of motion for the molecule. This is supported by a recent study [15].

The results presented in this paper show that it is possible to obtain a complete sampling of the essential subspace of a small peptide in water, together with an accurate definition of the location of boundaries. This suggests, together with previous results [7], that similar methods can be used to study the configurational space for larger peptides and proteins. We have evidence that this is indeed the case although the computational effort is considerably larger. The borders of the essential space now contain regions that may involve unfolding pathways.

Acknowledgement

It is a pleasure to thank dr. Ruud Scheek for many stimulating discussions.

References and Footnotes

1. N. J. Skelton, K.C. Garcia, D.V. Goeddel, C. Quann, and J.P. Burnier, Determination of the solution structure of the peptide hormone guanylin: Observation of a novel form of topologi cal stereoisomerism, *Biochemistry, 33* , 13581-13592 (1994).
2. M.G. Currie, K.F. Fok, J. Kato, R.J. Moore, F.K. Hamra, K.L. Duffin, and C.E. Smith, Guanylin: an endogenous activator of intestinal guanylate cyclase, *Proc. Natl. Acad. Sci. USA, 89*, 947-951 (1992).
3. L.R. Forte and M.G. Currie, Guanylin: a peptide regulator of epithelal transport, *The FASEB journal, 9* , 643-650 (1995).
4. M. Field, L.H. Graf Jr, W.J. Laird, and P.L. Smith, Heat-stable enterotoxin of Eschericia coli: in vitro effects on guanylate cyclase activity, cyclic GMP concentration and ion transport in small intestine, *Proc. Natl. Acad. Sci. USA, 75*, 2800-2804 (1978).
5. F.J. de Sauvage, S. Keshav, W-J Kuang, N. Gillet, W. Henzel, and D.V. Goeddel, Precursor structure, expression and tissue distributution of human guanylin, *Proc. Natl. Acad. Sci. USA, 89*, 9089-9093 (1992).
6. A. Amadei, A. B.M. Linssen, and H.J.C. Berendsen, Essential Dynamics of Proteins, *Proteins: Str., Funct., and Gen., 17* , 412-425 (1993).
7. A. Amadei, A.B.M. Linssen, B.L. de Groot, D.M.F. van Aalten, and H.J.C. Berendsen. An effi cient method for sampling the essential subspace of proteins. *J. Biom. Str. Dyn. 13*, 615-625 (1996).
8. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, In: *Intermolecular Forces*, B. Pullmann, editor, page 331. Reidel, Dordrecht, 1981.
9. H.J.C. Berendsen, J.P.M. Postma, A. DiNola, and J.R. Haak, Molecular dynamics with cou pling to an external bath, *J. Chem. Phys., 81*, 3684-3690 (1984).
10. 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).
11. W.F. van Gunsteren and H.J.C. Berendsen. *Gromos manual.* BIOMOS, Biomolecular Software, Laboratory of Physical Chemistry, University of Groningen, The Netherlands, 1987.
12. G. Vriend, WHAT IF: a molecular modeling and drug design program, *J. Mol. Graph., 8*, 52-56 (1990).
13. A. E. Garcia, Large-amplitude nonlinear motions in proteins, *Phys. Rev. Lett., 68*, 2696-2699 (1992).
14. A. Amadei, A. B. M. Linssen, B. L. de Groot, and H. J. C. Berendsen. Essential degrees of free dom of proteins. In: *Modelling of Biomolecular Structures and Mechanisms*, A. Pullmann et al., editor, pages 85-93, The Netherlands, 1995. Kluwer.
15. R.M. Scheek, N.A.J. van Nuland, B.L. de Groot, and A. Amadei, Structure from NMR and molecular dynamics: distance restraining inhibits motion in the essential subspace, *J. Biomol. NMR., 6*, 106-111 (1995).