THE DOMAIN DECOMPOSITION OF A SINGLE-DOMAIN PROTEIN

K. ANTON FEENSTRA & HERMAN J. C. BERENDSEN
Bioson Research Institute and Laboratory of Biophysical Chemistry
University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands
e-mail: anton@chem.rug.nl

Recently a novel method was developed to determine Dynamical Domains in proteins from molecular dynamics simulations [1] and from X-ray conformers [2]. This method is primarily intended to be used in the analysis of motions in multi-domain proteins. We will show that, using this method, also in a typical single-domain protein of 85 residues characteristic motions can be filtered out, which can be translated into regions of increased and decreased rigidity.

Abbreviations

ED	Essential Dynamics
HPr	Histidine containing phosphocarrier Protein
LD	positional Langevin Dynamics
MD	Molecular Dynamics
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank

1 Introduction

Understanding the complexity of proteins is one of the major challenges in science today [3,4]. Models that can tell us something about proteins will have to incorporate at least some of this complexity, but dealing with it head-on, without simplification, is currently far beyond reach of our cognitive or computational capabilities. This applies above all to aspects of folding. Experimental techniques are not yet able to measure directly parameters that give detailed information [5], although much progress is being made in some directions [6,7]. For this purpose, methods have been developed that are able to reduce the overwhelming complexity of proteins to a level that is either directly comprehensible to humans, or computationally accessible. An example of the former can be a simple lattice model with P and H beads (e.g. Shakhnovitch, 1997 [8]), where the unaided human mind can follow events and derive certain properties. An example of the latter can be an MD simulation with full atomic description (e.g. Daura *et al.*, 1997 [9]), which with increasing computer power will yield information on larger and larger time-scales and systems.

In general reducing the complexity of the system means that a choice will have to be made for the level of detail and for the number of degrees of freedom (which will also be dependent on the level of detail). The detail level is mostly determined by the number of explicit entities that will be included (ranging from full Quantum description with all electrons included, via full atom description to description with simplified side-chains and no explicit solvent). The number of degrees of freedom is of course limited by the number of independent entities in the model, but may be further restricted to some subset of allowable degrees of freedom (e.g. on- and off-lattice descriptions).

The common notion that proteins are built up from structural elements such as helices, sheets, specific turns, hydrophobic cores, etc., suggests a simplification of protein structure and dynamics in terms of such elements [10]. But a scientifically valid subdivision must be based on a proper analysis; such structural elements must be objectively distinguishable on the basis of their properties. If structural elements exist which have more internal rigidity than the protein as a whole, they can be considered as units with a significant role in function and in folding, thus reducing the complexity to hopefully manageable proportions.

Information on the motional freedom of proteins can be obtained from from x-ray data on different conformations, from NMR measurements or from molecular simulation. Simulation in full atomic detail is now possible over several nanoseconds. The trajectories can then be analyzed in terms of collective fluctuations of large amplitude, which are likely to describe the most important aspects of the overall dynamics, by the method of *Essential Dynamics* [11]. The method of essential dynamics analysis is quite straightforward to use, and is being used routinely by an increasing number of people. The results have proven to be useful in a number of studies [11,12,13,14].

Recently a 'model-free' method has been described to analyze motions occurring in proteins [1,2] in terms of rigid-body motions of internal domains, and has been used on x-ray data [2] and molecular simulations [1]. The selection of domains is based on the relative magnitude of intra-domain to inter-domain motion, and is thus not necessarily related to standard secondary structure elements. It is along these lines that we have investigated the internal domain structure of a typical single-domain protein. The purpose of this study is to establish the applicability of the method. A more detailed paper is in preparation, with the purpose of identifying internal elements which could subsequently be used to simplify the complexity of the protein for folding and functional studies.

2 Methods

2.1 Langevin Dynamics Trajectories

The results of 6 Langevin Dynamics (LD) simulations started from structure numbers #1, #2, #12, #17, #19 and #26 from the NMR PDB entry [15] 1HDN [16] (numbering corresponds to the consecutive structures in the PDB file), were provided by B. Hess [a,17], see Table 1. Each trajectory is 1 μs long, they are denoted by $l1$, $l2$, $l12$, $l17$, $l19$ and $l26$ (for LD).

2.2 Trajectory of MD with extended sampling

The results of 2 MD simulations with 'extended sampling' were provided by B.L. de Groot [18]. In this simulation, the eigenvectors from an ED analysis were used to

[a] Currently under development at our lab is a simplified description of proteins for use with positional LD simulations [17]. In this simplified description all backbone atoms are present, but side-chains are represented by at most three atoms. This makes large time-steps of approximately 20 fs possible. No explicit solvent is included in the simulation. These conditions make the extremely long simulation time of 1 μs (1000 ns) possible. A paper is in preparation.

Table 1: Overview of simulations used from other sources. Type MD* stands for MD with extended sampling. *nmr* denotes structures from the NMR PDB entry 1HDN [16]

Name	Started from	Type	Source
*l*1	nmr1	LD	[17]
*l*2	nmr2	LD	[17]
*l*12	nmr12	LD	[17]
*l*17	nmr17	LD	[17]
*l*19	nmr19	LD	[17]
*l*26	nmr26	LD	[17]
s1	nmr21	MD*	[18]
s2	nmr21	MD*	[18]

increase the sampling efficiency of an MD simulation [13]. These trajectories extend to 1057 and 685 ps respectively [18] and are denoted by s1 and s2.

2.3 MD Simulations

Starting Structures

2 MD simulations of HPr were performed from 2 different starting structures taken from the NMR PDB entry [15] 1HDN [16] (#1 and #12, numbering corresponds to the consecutive structures in the PDB file), the resulting trajectories will be referred to as n1 and n12 (for *n*ative). 2 more MD simulations were performed starting from structure #1 at elevated temperatures (400 K and 500 K), which will be referred to as h and 2h (for *h*ot and *too h*ot). 10 MD simulations were started from various frames (summarized in Table 2) of the Langevin Dynamics trajectories (see Section 2.1), they are denoted by f1 .. f26 and f26.5 .. f26.500 (for *f*olding, because they start from partially unfolded structures). In Table 2 a comprehensive summary of all performed simulations is given.

The starting structures taken from the *l*1 .. *l*26 trajectories were first unsimplified (i.e. all missing atoms in the simplified side-chain representation were added back). Note that this results in a random side-chain placement where the orientation of the full-atom parts cannot be deduced from the simplified description. This structure was energy minimized using steepest descents method, and 10 ps MD with position restraints with a force constant of 1000 kJ $mol^{-1}nm^{-2}$ was performed. For f1 .. f26 the position restraints were on all heavy atoms, for f26.5 .. f26.500 on the α-atoms only.

The protein was solvated in a cubic box of simple point charge (SPC) water [19]. For the n1 and n12 simulation this was done with a minimum distance of 0.9 nm between the protein and the box, resulting in a 5.7 nm cubic box. In the other simulations smaller box sizes are chosen. This will introduce some artifacts, but large deviations are already expected from the distorted protein structure (for the f1 .. f26 and f26.5 .. f26.500 simulations), or from the elevated temperature (for the h and 2h simulations), so the artifacts introduced by smaller box sizes will be negligible in comparison.

Table 2: Overview of MD simulations that were performed, a total of 51.2 ns. *nmr* denotes structures from the NMR PDB entry 1HDN [16], numbered according to the position of each conformation in the PDB file. Other references are to LD trajectories at specific times (see Table 1). Simulation temperature, total length of the simulation and the size of the cubic box is shown.

Name	**Started from**	**T (K)**	**Length (ns)**	**Box (nm)**
n1	nmr1	300	5.0	5.7
n12	nmr12	300	5.0	5.7
f1	$l1@1\mu$s	300	3.3	5.0
f2	$l2@1\mu$s	300	1.9	5.0
f17	$l17@1\mu$s	300	3.3	5.0
f19	$l19@1\mu$s	300	3.5	5.0
f26.5	$l26@0.005\mu$s	300	3.0	4.9
f26.10	$l26@0.01\mu$s	300	3.0	4.9
f26.50	$l26@0.05\mu$s	300	2.9	4.9
f26.100	$l26@0.1\mu$s	300	3.0	4.9
f26.500	$l26@0.5\mu$s	300	3.0	4.9
f26	$l26@1\mu$s	300	3.4	5.0
h	nmr1	400	5.3	5.1
2h	nmr1	500	5.6	5.4

For the n1, n12, h and 2h simulations, the resulting solvated structures were directly used as starting conformations for the MD simulations. For the f1 .. f26 and f26.5 .. f26.500 simulations, the solvated structures were again subjected to energy minimization and position restrained MD (as described above).

MD Parameters and Details

All MD simulations were performed using the following parameters and methods: The GROMOS-87 forcefield [20] was used, with increased repulsion between water oxygen and carbon atoms, as suggested by Van Buuren *et al.* [21]. Explicit hydrogens were defined for the aromatic rings [22], the resulting parameter set is the one referred to as SW by Daura *et al.* [23]. The MD time step was 2 fs. In simulation n1 SHAKE [24] was used for all covalent bonds up to 2290 ps, after that LINCS [25] was used. For all other simulations LINCS was used throughout the whole simulation. SHAKE and LINCS have been shown to give the same solution up to machine precision [25].

A twin-range cut-off for non-bonded interactions was employed with short-range cut-off for Van der Waals and Coulomb interactions of 1.0 nm which were calculated every simulation step, and a long-range cut-off of 1.8 nm for Coulomb interactions which were calculated during neighbor-list generation, every 20 fs. The temperature was controlled using weak coupling to a bath [26] of 300 K (400 K for h and 500 K for 2h; see Table 2) with a time constant of 0.1 ps. Protein and solvent were independently coupled to the heat bath. In the h and 2h simulations the system was heated from 300 K to 400 K resp. 500 K at a rate of 5 $\mathrm{K\,ps^{-1}}$ during the first part (20 ps for h and 40 ps for 2h) of the simulation. The pressure was also controlled using weak coupling with a time constant of 1.0 ps. A relative dielectric

constant ϵ_r of 1 was used.

The MD simulations were carried out using the GROMACS molecular dynamics package [27,28] on a Silicon Graphics Power Challenge with MIPS R10 000 processors and on SGI O2 Workstations with R5000 processors. On the Power Challenge on a single processor the n1 and n12 simulations (a total of 10 ns) with parameters as stated above took 88 days of computer time.

2.4 Essential Dynamics analysis

Essential dynamics analysis [11] was applied to all MD trajectories, with the first 10% of each discarded to exclude any possible equilibrational effects. The analysis was performed using the WEDTRA, WEDEIG and WEDPRJ tools from the WHATIF program [29]. First WEDTRA is used to fit all frames in the trajectory to one single reference frame (the first frame of the trajectory). Second WEDEIG is used to build the covariance matrix and diagonalize it, this yields the eigenvectors and eigenvalues, i.e. the essential modes and their corresponding root-mean-square fluctuations. Last, WEDPRJ is used to project the trajectory on each of the first 6 eigenvectors, the extreme structures in these projections are used as input for the Domain Decomposition analysis.

2.5 Domain Decomposition

The new method of Hayward *et al.* [1,2] of automatic domain decomposition based on rotational analysis was applied to the results of all ED analyses. To distinguish the domains determined by this method from domains determined by other methods (e.g. methods based on structural, genetical or functional information) they should be called *dynamical domains*, in this paper we will however mostly refer to them simply as *domains*.

This method is implemented in a program called DYNDOM, which was used for analysis. As input the extreme structures of a projection of a trajectory on one of its eigenvectors are used, these describe the motion that is contained in the eigenvector to the extent that is present in the trajectory. Parameters were set as follows: maximum number of clusters 50, number of iterations for clustering routine 100, fitting segment window length 5 residues, minimum domain size 8 residues, minimum ratio of external to internal motion 1.0.

3 Results

3.1 MD Simulations

Both native simulations (n1 and n12) show excellent stability in all aspects. The mean RMS deviation over both trajectories is 0.15 nm. Also the simulation at 400 K (h) is very stable, although more fluctuations are present (e.g. in the secondary structure elements) with a mean RMS deviation of 0.24 nm, which rises to around 0.3 at the end. All secondary structure elements, as determined by DSSP [30], are fully intact at the end of n1, n12 and h. The simulation at 500 K (2h) shows a severe, but gradual, degradation of structure. Most secondary structure elements

are lost at the end of this simulation, only parts of the N-terminal β-strand, and the C-terminal α-helix and β-strand remain. The RMS deviation rises to a mean value of 0.42 after about 1 ns.

3.2 Other Trajectories

In the LD trajectories ($l1 .. l26$) HPr is partially denatured, gradually most secondary structure disappears until only a fraction (some 20%) is left at the end of the 1 μs trajectories. The RMS deviation of all backbone heavy atoms goes up to 0.46 nm.

The trajectories from the MD with extended sampling (s1 and s2) are very stable, comparable to n1 and n12. The mean RMS deviation is slightly higher, around 0.24 nm, which is to be expected because of the extended sampling.

3.3 Essential Dynamics analysis

On all trajectories mentioned in Tables 1 and 2, Essential Dynamics analysis was performed, yielding for all trajectories a cumulative fluctuation (*cfl*) in the first 6 eigenvectors (denoted as *cfl(6)*) between 0.58 and 0.85, i.e. between 58% and 85% (depending on the trajectory) of the motion present in these trajectories can be described by only 6 degrees of freedom. These values are comparable to those reported by others [11,12,13,14]. Both the largest (0.85) and smallest (0.58) cfl(6) are from a LD trajectory. The sampling trajectories (s1 and s2) also have a large cfl(6), 0.85 and 0.78 respectively. The native trajectories (n1 and n12) have a cfl(6) of 0.66 resp. 0.63, the heated trajectories (h and 2h) of 0.64 resp. 0.74. The f1 .. f26 and f26.5 .. f26.500 have cfl(6)'s ranging from 0.63 and 0.79.

3.4 Domain Decomposition

Domain assignments

For all trajectories, the first 6 essential modes (from now on: *modes*), which comprises for the 22 trajectories a total of 132 modes, were analyzed for their domain composition using DYNDOM [1,2]. The resulting domain assignments are summarized in Fig. 1. For each of the trajectories, 6 horizontal bars are present, corresponding to the 6 essential modes. The color coding corresponds to the domain assignment by DYNDOM, however, the coloring is not consistent between different essential modes. In itself this summarized representation is not very informative, but it does give a comprehensive global view of the results of the analysis.

The first thing that should be noted is that no domain assignments are made for the first 2 and the last 2 residues in the protein, as can be seen in Fig. 1. This is due to the 5 residue long fitting window that is used by DYNDOM while determining the curl (i.e. rotational motion) of the residues. This means that for the first 2 and last 2 residues, no complete window is present, so no fitting and no curl determination is done for these residues.

From Fig. 1 it is immediately clear that not for all modes a consistent domain assignment could be made, these are all the rows that are completely white. For

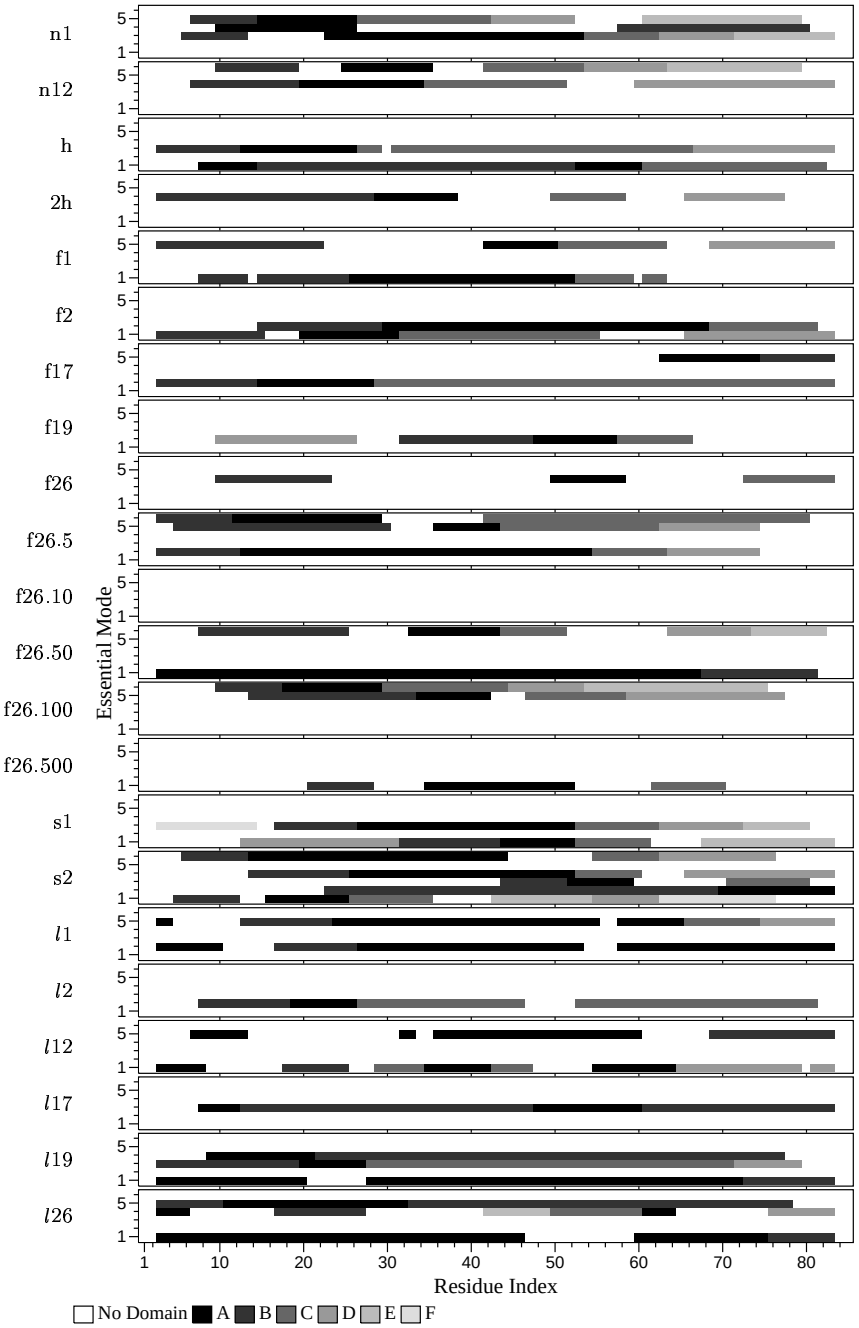

Figure 1: Domain decompositions per residue of HPr (along the x-axis) for the first 6 essential modes (along the y-axis) for each of the trajectories. From top to bottom n1, n12, h, 2h, f1 .. f26, f26.5 .. f26.500, s1, s2, and $l1$ .. $l26$. Color coding indicates different domains as assigned by the DYNDOM program [1,2]. Domain coloring does not necessarily match between different essential modes.

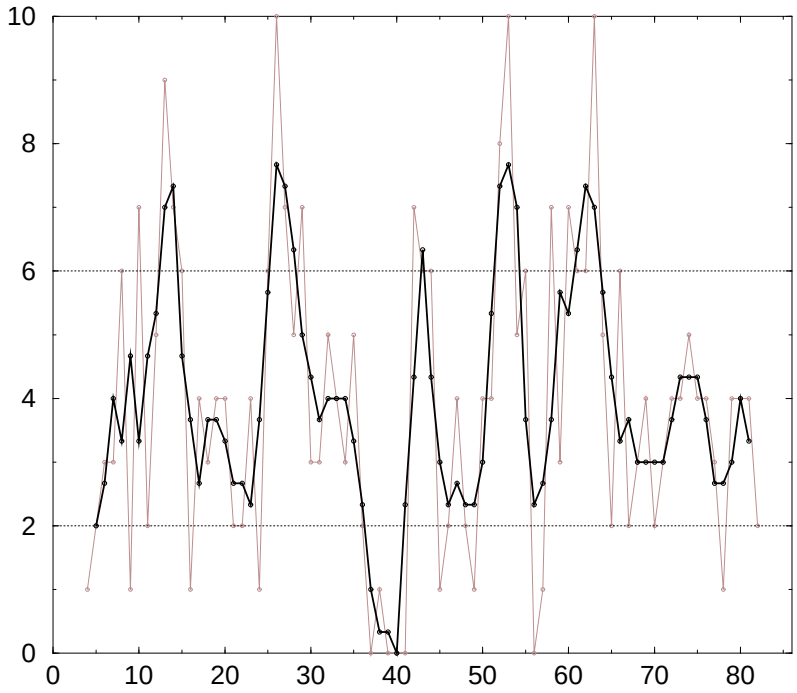

Figure 2: Total count of domain demarcations as assigned by the DYNDOM program [1,2] for all trajectories for the first 6 essential modes of HPr. The thin grey line is the actual count of which the thick black line is a running average over 3 points. The theoretical maximum count is 43; the number of modes for which a domain assignment has been made. The dotted lines correspond to the color coding used in Fig. 3.

43 modes out of the total 132 an assignment was made. Also, not all residues need to be assigned to a domain, which can be seen as white stretches in a number of rows. Counting over all modes for which a domain assignment was made and ignoring the first 2 and last 2 residues, a total of 72% of all residues were assigned to domains. The residues that were not assigned to domains, will typically be involved in inter-domain motion (i.e. they serve as hinges or connecting regions).

Domain demarcations

A more compact view of the results is given in Fig. 2 which shows the count of domain boundaries (or demarcations) at each residue position. The boundaries at the third residue from the termini are not taken into account, as they arise directly from the 5 residue fitting window (as mentioned in Section 3.4). The actual count is displayed as a thin grey line, a running average is taken over three points which is shown as a thick black line. Clearly visible are 5 'hot spots' (at or around positions 14, 26, 43, 53 and 62), where domain boundaries often occur and one 'cold spot' (around position 40) where hardly any domain boundaries occur. A number of regions of intermediate values is present.

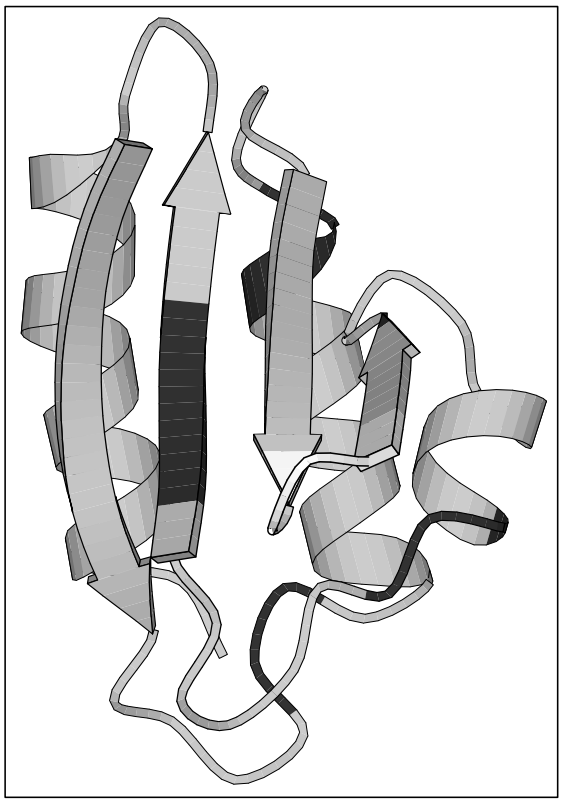

Figure 3: 3 residue running average of the count of domain demarcations as assigned by the DYNDOM program [1,2] (see Fig. 2) color coded onto a cartoon representation of HPr. White, grey and black correspond to a count of 0 .. 2, 2 .. 6 and 6 .. 8 resp (see also Fig. 2). Note that shading makes the white darker and the black lighter (Plot made by MolScript [31] with recent extensive modifications [32]).

The 3-residue running averaged demarcation counts (see Fig. 2) are color coded onto a graphical representation of HPr in Fig. 3, with white, grey and black corresponding to a count of 0 .. 2, 2 .. 6 and 6 .. 8 resp. So 'cold spots' will show up as white regions, the 'hot spots' as black, and intermediate regions will be grey. The first 3 and last 3 residues for which no domain demarcations are possible, as was mentioned in the Section 3.4 are colored according to the fourth residue from the respective termini (which is grey in both cases).

The cold spot is easily visible as the white (with greyish shadings) β-turn (type I$'$) between the two right-most β-strands. The hot spots are also readily identified as the black regions (with dark-grey shading). Starting from the C-terminus, the first is found in the loop region between the left β-strand and the middle α-helix (at the bottom), the second at the top of the middle α-helix, the third (which is only a small spot) at the end of the rightmost, short, β-strand, the fourth at the end (bottom) of the rightmost α-helix and the fifth in the middle of the second (from the left) β-strand.

4 Discussion

The large number of MD simulations we performed, represent the state of the art. Of these, n1 and n12 are the best in terms of methodology, i.e. they have a large long-range cutoff of 1.8 nm and a correspondingly large computational box with water. Of the others (h, h2, f1 .. f26 and f26.5 .. f26.500), some artifacts may be expected from the somewhat smaller box sizes. This does not present a problem, because of the different conditions of the systems in these MD simulations, i.e. elevated temperature and very non-native like conformations. This means that some detailed aspects might not be totally correct as a result of the smaller box-size, but the general behavior of the system and the qualitative results contained in the trajectories will be correct, because they are dominated primarily by these deviating conditions.

The new method of determining a domain composition based on rotational correspondence of residues appears to be applicable also to analyze properties of small, single domain proteins. The results will of course depend heavily on the quality of the MD simulations on which the essential dynamics analyses were performed. In the present study, a large variety of starting structures and simulation conditions is used to ensure that results do not depend on the choice of a specific starting conformation, or on particular conditions.

The transitions between the hot spots, the cold spots and the intermediate regions (see Fig. 2) is fairly sharp, indicating that these results are meaningful at least in the sense that the distribution is not random. More detailed analyses should be performed, linking the hot- and cold spots to other local structural properties of the protein. Correlations with the actual motions found in the essential spaces of the separate trajectories should also exist, so correspondences should be present between details of these motions and the hot- and cold spots, linking them to dynamical properties of the protein.

The location of most hot spots do not bring any surprise: in a loop or at the ends of β-strands or α-helices. So generally, secondary structure elements (especially α-helices) tend to move as more or less rigid bodies. One hot spot, however, is located in a strange place: in the middle of a β-strand, which is in turn in the middle of the β-sheet. This means that, very often, the β-sheet does not move as a whole, but it breaks up in at least two separate parts.

The location of the cold spot seems very significant, i.c. right at a β-turn. Apparently this is a very stable part of the protein since in all the trajectories (including those at 400 K and 500 K, the LD trajectories and the f1 .. f26 and f26.5 .. f26.500 trajectories) it is not divided into separate domains. This might even indicate that this region is crucial to the folding behavior of HPr.

Because all the analyses are based on a 'grand synthesis' of the results of a number of very different simulations, comprising 132 different dynamical modes, it will be interesting to see whether there are significant differences in the results if the analysis were repeated on particular subsets of the simulations. It might be that certain features (e.g. the existence and locations of the hot and cold spots) arise mainly from a certain type of simulation, or that combining two different types of simulation, obscures the presence of other features.

5 Conclusions and future aspects

In this preliminary study we have analyzed dynamic trajectories of the single-domain protein HPr by essential dynamics and by the new method of automatic domain decomposition based on rotational analysis, as implemented in Hayward's program DYNDOM. The results depend strongly on the quality of the MD simulations, for which a variety of methods and initial structures have been used. It was found to be possible to identify stable regions (cold spots) in the protein, as well as transitional regions (hot spots), which can hopefully be used in further studies on the nature of the folding process. This warrants further detailed investigation.

Acknowledgments

We thank Dr. Steven Hayward for stimulating discussions and help with using DYNDOM and Drs. Bert L. De Groot for providing the trajectories of the MD simulation with extended sampling and Ir. Berk Hess for providing the trajectory of the positional LD simulation. K.A.F. acknowledges support from the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for Scientific Research (NWO).

References

1. Hayward, S., Kitao, A., and Berendsen, H. J. C. Model-free methods of analyzing domain motions in proteins from simulation: A comparison of a normal mode analysis and a molecular dynamics simulation of lysozyme. *PROTEINS: Struct. Funct. Gen.*, 27:425–437, 1997.
2. Hayward, S. and Berendsen, H. J. C. Systematic analysis of domain motions in proteins conformational change; new results on citrate synthase and t4 lysozyme. *PROTEINS: Struct. Funct. Gen.* (in press).
3. Dill, K. A. and Fersht, A. R. Folding and binding. *Curr. Opin. Struct. Biol.*, 6:1–2, 1996.
4. Dobson, C. M. and Ptitsyn, O. B. Folding and binding. from theory to therapy. *Curr. Opin. Struct. Biol.*, 7:1–2, 1997.
5. Dill, K. and Chan, H. From levinthal to pathways to funnels. *Nature Struct. Biol.*, 4:10–19, 1997.
6. Diehl, M., Doster, W., Petry, W., and Schober, H. Water-coupled low-frequency mode of myoglobin and lysozyme observed by inelastic neutron scattering. *Biophys. J.*, 73:2726–2732, 1997.
7. Eaton, W. E., Muñoz, V., Thompson, P. A., Chan, C.-K., and Hofrichter, J. Submillisecond kinetics of protein folding. *Curr. Opin. Struct. Biol.*, 7:10–14, 1997.
8. Shakhnovich, E. I. Theoretical studies of protein-folding thermodynamics and kinetics. *Curr. Opin. Struct. Biol.*, 7:29–40, 1997.
9. Daura, X., van Gunsteren, W., Rigo, D., Jaun, B., and Seebach, D. Studying the stability of a helical β-heptapeptide by molecular dynamics simulations. *Chem. Eur. J.*, 3:1410–1417, 1997.

10. Fersht, A. R. Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.*, 7:3–9, 1997.
11. Amadei, A., Linssen, A. B. M., and Berendsen, H. J. C. Essential dynamics of proteins. *PROTEINS: Struct. Funct. Gen.*, 17:412–425, 1993.
12. van Aalten, D. M. F., Amadei, A., Linssen, A. B. M., Eijsink, V. G. H., and Berendsen, H. J. C. The essential dynamics of thermolysin: Confirmation of the hinge-bending motion and comparison of simulations in vacuum and water. *PROTEINS: Struct. Funct. Gen.*, 22:45–54, 1995.
13. Amadei, A., Linssen, A. B. M., de Groot, B. L., Aalten, D. M. F., and Berendsen, H. J. C. An efficient method for sampling the essential subspace of proteins. *J. Biomol. Struct. Dyn.*, 13:615–626, 1996.
14. van Aalten, D. M. F., de Groot, B. L., Berendsen, H. J. C., Findlay, J. B. C., and Amadei, A. A comparison of techniques for calculating protein essential dynamics. *J. Comp. Chem.*, 18(2):169–181, 1997.
15. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Jr., E. F. M., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.*, 112:535–542, 1977.
16. 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. The high-resolution structure of the histidine-containing phosphocarrier protein HPr from Escherichia coli determined by restrained molecular dynamics from NMR nuclear overhauser effect data. *J. Mol. Biol.*, 237:544–559, 1994.
17. Hess, B. Preliminary results from positional Langevin Dynamics simulations on HPr. Private Communication.
18. de Groot, B. L., Amadei, A., Scheek, R. M., van Nuland, N. A. J., and Berendsen, H. An extended sampling of the configurational space of HPr from E. coli. *PROTEINS: Struct. Funct. Gen.*, 26(3):314–322, 1996.
19. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., and Hermans, J. Interaction models for water in relation to protein hydration. In Pullman, B., editor, *Intermolecular Forces*, pages 331–342. D. Reidel Publishing Company, Dordrecht, 1981.
20. van Gunsteren, W. F. and Berendsen, H. J. C. *Gromos-87 manual.* Biomos BV, Nijenborgh 4, 9747 AG Groningen, The Netherlands, 1987.
21. van Buuren, A. R., Marrink, S. J., and Berendsen, H. J. C. A molecular dynamics study of the decane/water interface. *J. Phys. Chem.*, 97:9206–9212, 1993.
22. van der Spoel, D., van Buuren, A. R., Tieleman, D. P., and Berendsen, H. J. C. Molecular dynamics simulations of peptides from BPTI: A closer look at amide-aromatic interactions. *J. Biomol. NMR*, 8:229–238, 1996.
23. Daura, X., Oliva, B., Querol, E., Avilés, F. X., and Tapia, O. On the sensitivity of MD trajectories to changes in water-protein interaction parameters: The potato carboxypeptidase inhibitor in water as a test case for the GROMOS force field. *PROTEINS: Struct. Funct. Gen.*, 25:89–103, 1996.
24. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. 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.
25. Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.*, 18:1463–1472, 1997.
26. Berendsen, H. J. C., Postma, J. P. M., DiNola, A., and Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.
27. Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.*, 91:43–56, 1995.
28. van der Spoel, D., van Buuren, A. R., Apol, E., Meulenhoff, P. J., Tieleman, D. P., Sijbers, A. L. T. M., van Drunen, R., and Berendsen, H. J. C. GROMACS *User Manual version 1.5.* Nijenborgh 4, 9747 AG Groningen, The Netherlands. Internet: `http://rugmd0.chem.rug.nl/~gmx`, 1997.
29. Vriend, G. WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.*, 8:52–56, 1990.
30. Kabsch, W. and Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22:2577–2637, 1983.
31. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.*, 24:946–950, 1991.
32. Esnouf, R. M. An extensively modified version of MOLSCRIPT that includes greatly enhanced coloring capabilities. *J. Mol. Graphics*, 15:132–134, 1997.