Collective protein dynamics in relation to function

Herman JC Berendsen* and Steven Hayward†

Several techniques for the analysis of the internal motions of proteins are available – separating large collective motions from small, presumably uninteresting motions. Such descriptions are helpful in the characterization of internal motions and provide insight into the energy landscape of proteins. The real challenge, however, is to relate large collective motions to functional properties, such as binding and regulation, or to folding. These issues have been recently addressed in several papers.

Addresses
*Groningen Biomolecular Sciences and Biotechnology Institute (GBB), Department of Chemistry, The University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands; e-mail: berendsen@chem.rug.nl
†School of Information Systems, University of East Anglia, Norwich NR4 7TJ, UK; e-mail: sjh@sys.uea.ac.uk

Current Opinion in Structural Biology 2000, **10**:165–169

0959-440X/00/$ – see front matter

Abbreviations

EDA	essential dynamics analysis
MD	molecular dynamics
MSF	mean-square fluctuation
NMA	normal mode analysis
PCA	principal component analysis
rmsd	root mean square deviation

Introduction

Collective motions within proteins can be derived from atomic interactions given a single experimental structure, from knowledge of at least two molecular conformations or from an analysis of many conformations, as can be generated by the computer simulation technique of molecular dynamics (MD). The analysis of collective motions can be used to investigate the conformational energy landscape, to improve sampling efficiency and in the refinement of X-ray and NMR data. These aspects have been covered recently in an excellent review by Kitao and Go [1••], and earlier by Hayward and Go [2]. In this review, we concentrate on the application of techniques to derive collective motions in relation to function and folding. Aspects of protein function related to motion include substrate binding and product release, regulation and allosteric behavior, and contractile and motor functions. Collective motions of protein fragments must also be involved in the folding process, but their systematic investigation is still in an early stage.

Computational techniques to determine collective motions

Analysis of collective behavior must be based on knowledge of the structural fluctuations that occur as a result of thermal motion in the protein. Such fluctuations can be obtained in various ways.

Normal mode analysis (NMA) is based on the assumption that, over the range of thermal fluctuations, the conformational energy surface can be characterized by the parabolic approximation at a single energy minimum. This assumption is false at physiological temperatures. It ignores both the effect of solvent, although some bound water molecules may be included in the calculation, and the multiple minima nature of the conformational energy surface. NMA determines the independent harmonic modes of the molecule, each single mode comprising the concerted motions of many atoms. The standard application to a large biological molecule is computationally expensive; however, several sophisticated numerical techniques can be applied to extract the lowest frequency modes of large molecules [3]. Despite its approximations, NMA has been found to be useful in determining functionally relevant motions [1••,3]. NMA uses a harmonic approximation of the full atomic force-field, whereas the CONCOORD method of de Groot *et al.* [4] uses a very crude approximation of atomic interactions to generate a large number of conformations that satisfy a set of atomic distance constraints. For both methods, a single experimental structure suffices.

MD simulations generate an overwhelming amount of information contained in the trajectory of atomic coordinates. The viewing of such a trajectory on a graphics screen reveals the tremendous complexity of protein motion, but little else. In order to reveal the concerted fluctuations with large amplitudes, a principal component analysis (PCA) can be carried out on a large number of configurations chosen from the MD trajectory. As the dynamic consistency of the chosen configurations is not relevant, Monte Carlo methods [5,6] can be used as well. Alternatively, PCA can be applied to a sufficiently large set of experimental structures or to a set of configurations that is compatible with the freedom allowed for atomic contacts of various types, as can be generated by CONCOORD. PCA involves diagonalization of the covariance matrix of atomic fluctuations (after a translational/rotational fit to remove overall translation and rotation) to yield collective variables that are sorted according to their contribution to the total mean-square fluctuation (MSF). One application of PCA, quasi-harmonic analysis, uses mass-weighted coordinates of all atoms, thus relating it to NMA. This is useful in studies that concern the nature of the conformational energy surface [1••]. For studies that focus on relating large-scale motions to function, however, there exists the possibility of reducing the computational task by selecting only backbone or Cα atoms for the PCA. This analysis is often termed essential dynamics analysis (EDA) [7] or molecular optimal dynamic coordinate analysis [8]. In projecting the trajectory onto the collective variables with the largest contributions to the total MSF, one must always be aware that the resulting motion may not be independent of that

of other variables, which may also make a significant contribution to the total MSF [9]. Although MD simulations are, in principle, to be preferred to NMAs, the problem of convergence during the MD simulation has prompted a number of studies [10,11]. It is prudent, therefore, to make a comparison (by taking inner products) between the eigenvectors of the first half of the trajectory and those of the second half, or between two independent simulations, in order to distinguish their intrinsic nature from noise.

The motion in the 'essential' subspace of the first few eigenvectors seems, in most cases, to be best described as diffusion in a rather flat potential of mean force. MD simulations of nanoseconds length sample only a fraction of the accessible space and, hence, the rmsd from the initial configuration will continue to increase with time; such increase is not necessarily a sign of insufficient equilibration or conformational instability. Amadei *et al.* [12,13] have shown that combining the motions in a limited set of harmonic wells is sufficient to characterize the motion as determined by MD and EDA. Diffusional hopping between shallow wells causes a negative tail in the velocity correlation function, with an inhibitive effect on the rate of diffusion. Kitao *et al.* [14] have developed a combined PCA and NMA analysis that is based on the jumping among minima (or JAM) model. Applying this analysis to a MD simulation of lysozyme in water, it was possible to demonstrate the hierarchical nature of the conformational energy surface and determine the distribution of barrier heights.

Although the graphical presentation of the motion along a single EDA eigenvector is easier to comprehend than the presentation of the unprocessed trajectory, in many cases it appears to be possible to simplify the protein motion even further. This is the case when there are collective motions for which parts of the protein behave as quasi-rigid bodies. The motion along any individual normal mode eigenvector or essential dynamics eigenvector is defined by two conformations. Several programs now exist that can take two conformations of a protein and express the conformational change as the relative movement of the rigid parts using screw axes. There are two main approaches to this problem. The first searches for rigid domains within the protein using a multiple least-squares fitting procedure [15], for which the program HingeFind is available. The second determines the rigid bodies or dynamic domains by locating clusters of rigid-body parameters. The DynDom program [16] determines clusters of rotation vectors before determining interdomain screw axes and interdomain bending regions. The program DomainFinder [17] is based on the same idea, but includes relative translation in the analysis. If appropriate, the motion implied by each normal mode eigenvector or essential mode eigenvector can be graphically represented by a single structure, displayed together with arrows that represent hinge axes indicating the relative motion of the domains. This depiction is usually only sensible for the modes that correspond to fluctuations with the largest MSFs and that are associated with global, rather than local, motions. This depiction allows a direct comparison of the same analysis of two experimentally determined structures that represent a functional movement [18•]. Such a comparison is both easier and more meaningful, from a biochemist's point of view, than a mathematical similarity measure, such as inner products of eigenvectors. Figure 1 illustrates such a comparison for bacteriophage T4 lysozyme.

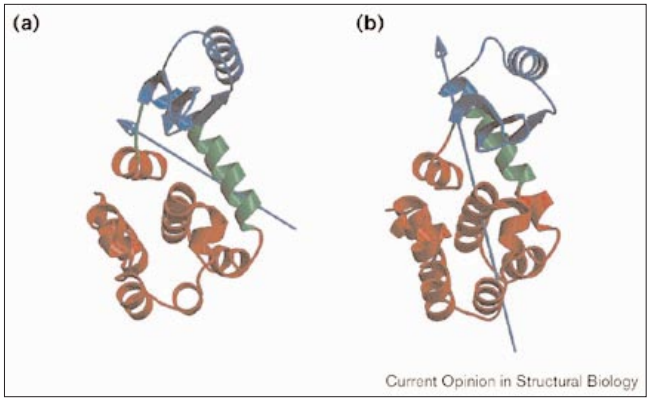
Applications

Functional motions

In this section, we review recent applications of NMA and EDA that concentrate on understanding functional motions.

GroEL is a chaperonin of considerable interest because of its role in *in vivo* protein folding. It is a double-ringed cylinder, each ring comprising seven identical subunits. Interest in its allosteric mechanism and, in particular, in the nature of inter-ring communication has stimulated two recent studies. de Groot *et al.* [19•] have applied the distance fluctuation method CONCOORD to GroEL alone and GroEL bound to GroES. An EDA was then applied to the resulting 'trajectories'. Fluctuations in the direction of the conformational change between the GroES-free and GroES-bound structures of GroEL were correlated with fluctuations in the direction associated with nucleotide binding. It was also found that this coupling was caused by constraints on residues at the interface between the two rings. NMA analyses of parts of GroEL have also been reported [20]. An NMA of three equatorial domains, two from the *trans* ring and one from the *cis*, have indicated how

Figure 1



Comparing simulation with experiment for bacteriophage T4 lysozyme. **(a)** The results of the domain decomposition (using the program DynDom [16]) of the first essential mode of motion derived from an EDA of 38 crystal structures of T4 lysozyme. The N-terminal domain is colored blue, the C-terminal domain red and the interdomain bending region green. The axis indicates how the red domain moves relative to the blue domain as a quasi-rigid body. The motion is predominantly one of closure. **(b)** A DynDom analysis of the first essential mode of a MD simulation. Although the domain decomposition is basically the same as in the X-ray case, some differences are discernable, in particular the direction of the hinge axis, which indicates a twisting mode of motion, rather than a closing mode. Adapted from [27].

displacements in the *cis* ring associated with ATP binding are coupled to complementary motions in the *trans* ring.

Another large allosteric protein is aspartate transcarbamylase. An NMA of this multimeric protein with some 2770 residues has been performed using a technique called diagonalization in a mixed basis [21]. Using this technique, it has been possible to calculate 53 of the lowest frequency modes for both the Tensed (T) and the Relaxed (R) structures of this enzyme. The modes determined for both structures have been compared with the T to R transition itself. A recent study has focused on the tertiary deformations that accompany quaternary changes [22]. In particular, the domain motions within the regulatory and catalytic chains were analyzed. At least in the regulatory chain, the domain demarcation implied by the lowest frequency modes was in agreement with that implied by the two X-ray structures. The NMA showed that the coupled motions of the regulatory and catalytic chains are correlated with intrachain domain motions. Regions that couple tertiary deformations to quaternary changes were identified. Hinsen *et al.* [17] have performed a domain decomposition on an NMA of aspartate transcarbamylase.

NMA has also been recently applied to α-lytic protease in an attempt to understand substrate specificity [23••]. The wild-type is specific to substrates that have a residue with a small hydrophobic sidechain preceding the scissile bond. On the other hand, the M190A mutant shows much lower specificity and is able to bind larger sidechains. NMAs were performed on both the wild-type and mutant structures in order to understand the mechanism of the difference in specificity. It was shown that, in the wild-type structure, the walls of the binding pocket moved together in a symmetric fashion, thus preserving the size of the pocket, whereas in the mutant, the walls moved in an antisymmetric way, causing the pocket to expand and contract. The difference in the substrate specificity was attributed to this difference in the character of the motion at the binding pocket. An antisymmetric motion of two DNA-binding wings was also found from an EDA of a MD simulation on a single stranded DNA binding protein [24]. The movement of the wings opens and closes the DNA-binding site. Unfortunately, in both these studies, the mechanisms underlying these movements were not elucidated.

Chau *et al.* [25] have studied the dynamics of serum retinol-binding protein using MD and EDA. They apply a technique that may prove useful in future applications, that is, the projection of the MD trajectory onto the space defined by the essential dynamics eigenvectors derived from an analysis of different X-ray structures of the protein [26,27]. Such a projection, usually made onto the first two eigenvectors for ease of presentation, can reveal whether the MD trajectory visits the regions of the configuration space defined by the X-ray structures, as indeed it did in this case. Their analysis showed that bound retinol did not appear to greatly influence the dynamics of the protein or *vice versa*.

Lins *et al.* [28] have performed two MD simulations on the catalytic domain of HIV-1 integrase, one with no magnesium ion in the active site, the other with. EDA clearly reveals how the flexibility of this site is quenched in the presence of the ion.

Kazmierkiewicz *et al.* [29] have performed MD and EDA on neurophysins with and without ligands. The main modes of motion indicate allosteric communication between the site of ligand binding and the intermonomeric interface.

Folding motions

Only a handful of investigations that involve EDA applied to protein folding or stability has been reported. Several unfolding simulations that reveal the relative stability of domains (e.g. [30]) are not covered in this review, as they do not exploit PCA or EDA.

Creveld *et al.* [31] have performed three MD simulations on cutinase, a protein of considerable interest to the washing powder industry because of its ability to remove fat stains. Unfortunately, it is unstable in the presence of surfactants. A simulation at room temperature was performed to reveal functional sites and two unfolding simulations were carried out, one at 393K and one with a modified protein–solvent interaction allowing water to penetrate the interior of the protein more easily, to reveal sites associated with unfolding. EDA was used to identify these sites. The consistency of unfolding from the two unfolding simulations suggested sites at which mutations may stabilize the enzyme without affecting its function.

Roccatano *et al.* [32] have looked at the possibility that a β hairpin in protein G forms a nucleation site by performing MD simulations at various temperatures and analyzing the trajectories using EDA. They found that the peptide retained a β-hairpin-like structure over a wide range of temperatures, in agreement with experimental evidence.

Kazmirski *et al.* [33] have performed multiple unfolding simulations on three different proteins. Although they do not use EDA in the conventional sense, they do perform PCA in a space of properties that are found to vary during the simulations. Collective property variables are determined that characterize the unfolding.

García and Hummer [34•] analyzed collective motions of cytochrome *c*, in simulations at four temperatures, in terms of the involvement of formation and breaking of intramolecular hydrogen bonds. Large fluctuations are not necessarily correlated with experimentally observed amide hydrogen exchange rates, but may involve rigid-body displacements with stable hydrogen bonds.

Conclusions and outlook

A lot of effort has gone into describing and characterizing the internal motions of proteins. From a biochemical

point of view, the connection to function is a crucial aspect, which is just beginning to be uncovered; the days have passed that a mere description of protein internal motions from MD simulations was considered to be of interest. How functional motions relate to detailed molecular characteristics and amino acid sequences is still to be discovered. A practical bottleneck is the time limitation of present day MD simulations, which may cover times up to tens of nanoseconds. Such simulations may not grasp the essential motions related to function at much longer time-scales. Examples are slow conformational changes triggered by effector binding, macromolecular interactions or external fields. Improvements in computational power and the implementation of more sophisticated techniques will undoubtedly close the present gap between simulation and reality.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. •• Kitao A, Go N: **Investigating protein dynamics in collective coordinate space.** *Curr Opin Struct Biol* 1999, **9**:164-169.

A broader review than this – it describes applications of collective coordinate space analyses to the nature of the conformational energy surface and refinement of experimental data. It gives a clear idea of the importance of this field and its future directions.

2. Hayward S, Go N: **Collective variable description of native protein dynamics.** *Annu Rev Phys Chem* 1995, **46**:223-250.

3. Hayward S: **Normal mode analysis of biological molecules.** In *Computational Biochemistry and Biophysics*. Edited by Becker OM, MacKerell AD Jr, Roux B, Watanabe M. New York: Marcel Dekker; 2000:in press.

4. de Groot BL, van Aalten DMF, Scheek RM, Amadei A, Vriend G, Berensdsen HJC: **Prediction of protein conformational freedom from distance constraints.** *Proteins* 1997, **29**:240-251.

5. Bahar I, Erman B, Haliloğlu T, Jernighan RL: **Efficient characterization of collective motions and interresidue correlations in proteins by low-resolution simulations.** *Biochemistry* 1997, **36**:13512-13523.

6. Haliloğlu T, Bahar I: **Coarse-grained simulations of conformational dynamics of proteins: applications to apomyoglobin.** *Proteins* 1998, **31**:271-281.

7. Amadei A, Linssen ABM, Berendsen HJC: **Essential dynamics of proteins.** *Proteins* 1993, **17**:412-425.

8. García AE: **Large amplitude nonlinear motions in proteins.** *Phys Rev Lett* 1992, **68**:2696-2699.

9. Basu G, Kitao A, Hirata F, Go N: **A collective motion description of the 3_{10}-/α-helix transition: implications for a natural reaction coordinate.** *J Am Chem Soc* 1994, **116**:6307-6315.

10. Balsera MA, Wriggers W, Oono Y, Schulten K: **Principal component analysis and long time protein dynamics.** *J Phys Chem* 1996, **100**:2567-2572.

11. de Groot BL, van Aalten DMF, Amadei A, Berendsen HJC: **The consistency of large concerted motions in proteins in molecular dynamics simulations.** *Biophys J* 1996, **71**:1707-1713.

12. Amadei A, de Groot BL, Cerusa M-A, Paci M, Di Nola A, Berendsen HJC: **A kinetic model for the internal motions of proteins: diffusion between multiple harmonic wells.** *Proteins* 1999, **35**:283-292.

13. Amadei A, Cerusa MA, Di Nola A: **On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations.** *Proteins* 1999, **36**:419-424.

14. Kitao A, Hayward S, Go N: **Energy landscape of a native protein: jumping-among-minima model.** *Proteins* 1998, **33**:496-517.

15. Wriggers W, Schulten K: **Protein domain movements: detection of rigid domains and visualization of hinges in comparison of atomic coordinates.** *Proteins* 1997, **29**:1-14.

16. Hayward S, Berendsen HJC: **Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme.** *Proteins* 1998, **30**:144-154.

17. Hinsen K, Thomas A, Field MJ: **Analysis of domain motions in large proteins.** *Proteins* 1999, **34**:369-382.

18. • Hayward S: **Structural principles governing domain motions in proteins.** *Proteins* 1999, **36**:425-435.

Proteins for which at least two experimentally determined conformations are available were analyzed in terms of dynamic domains, hinge axes and hinge-bending regions. The purpose was to determine whether any structural motifs involved in controlling interdomain motions could be identified. Individual cases often represent an experimentally determined functional motion and provide an opportunity to assess simulations by comparison.

19. • de Groot BL, Vriend G, Berendsen HJC: **Conformational changes in the chaperonin GroEL: new insights into the allosteric mechanism.** *J Mol Biol* 1999, **286**:1241-1250.

CONCOORD simulations of single-ring and double-ring structures of GroEL were made. Molecular dynamics simulations of structures of this size would not be currently feasible. The results indicate coupling between the conformational change associated with nucleotide binding and that associated with GroES binding, but only in the double-ring simulations, not in simulations of single rings.

20. Ma J, Karplus M: **The allosteric mechanism of the chaperonin GroEL: a dynamic analysis.** *Proc Natl Acad Sci USA* 1998, **95**:8502-8507.

21. Mouawad L, Perahia D: **Diagonalization in a mixed basis: a method to compute low-frequency normal modes for large macromolecules.** *Biopolymers* 1993, **33**:599-611.

22. Thomas A, Hinsen K, Field MJ, Perahia D: **Tertiary and quaternary conformational changes in aspartate transcarbamylase: a normal mode study.** *Proteins* 1999, **34**:96-112.

23. •• Miller DW, Agard DA: **Enzyme specificity under dynamic control: a normal mode analysis of alpha-lytic protease.** *J Mol Biol* 1999, **286**:267-278.

Given the level of approximation in normal mode analysis, it is surprising that the results of this study can be linked so well to experimental results. It is shown how a mutant can affect the dynamics and, thus, the function of an enzyme by influencing fluctuations in the size of the substrate-binding pocket.

24. Horstink LM, Abseher R, Nilges M, Hilbers CW: **Functionally important correlated motions in the single-stranded DNA-binding protein encoded by filamentous phage Pf3.** *J Mol Biol* 1999, **287**:569-578.

25. Chau P-L, van Aalten DMF, Bywater RP, Findlay JBC: **Functional concerted motions in the bovine serum retinol-binding protein.** *J Comput Aided Mol Des* 1999, **13**:11-20.

26. van Aalten DMF, Conn DA, de Groot BL, Findlay JBC, Berendsen HJC, Amadei A: **Protein dynamics derived from clusters of crystal structures.** *Biophys J* 1997, **73**:2891-2896.

27. de Groot BL, Hayward S, van Aalten DMF, Amadei A, Berendsen HJC: **Domain motions in bacteriophage T4 lysozyme; a comparison between molecular dynamics and crystallographic data.** *Proteins* 1998, **31**:116-127.

28. Lins RD, Briggs JM, Straatsma TP, Carlson HA, Greenwald J, Choe S, McCammon JA: **Biophysical theory and modeling – molecular dynamics studies on the HIV-1 integrase catalytic domain.** *Biophys J* 1999, **76**:2999-3011.

29. Kazmierkiewicz R, Czaplewski C, Lammek B, Ciarkowski J: **Essential dynamics/factor analysis for the interpretation of molecular dynamics trajectories.** *J Comput Aided Mol Des* 1999, **13**:21-33.

30. Lu H, Schulten K: **Steered molecular dynamics simulations of force-induced protein domain unfolding.** *Proteins* 1999, **35**:453-463.

31. Creveld LD, Amadei A, van Schaik RC, Pepermans HAM, de Vlieg J, Berendsen HJC: **Identification of functional and unfolding motions of cutinase as obtained from molecular dynamics computer simulations.** *Proteins* 1998, **33**:253-264.

32. Roccatano D, Amadei A, Di Nola A, Berendsen HJC: **A molecular dynamics study of the 41-56 β-hairpin from B1 domain of protein G.** *Protein Sci* 1999, **8**:2130-2143.

33. Kazmirski SL, Li A, Daggett V: **Analysis methods for comparison of multiple molecular dynamics trajectories: applications to protein unfolding pathways and denatured ensembles.** *J Mol Biol* 1999, **290**:283-304.

34. García AE, Hummer G: **Conformational dynamics of cytochrome c:**
• **correlation to hydrogen exchange.** *Proteins* 1999, **36**:175-191.
Molecular dynamics simulations of horse heart cytochrome *c* at different temperatures were analyzed with a view to understanding the nature of the motion within the rugged energy landscape and hydrogen exchange. Large fluctuations were found to be associated with regions that move concertedly and were not necessarily correlated with hydrogen exchange.