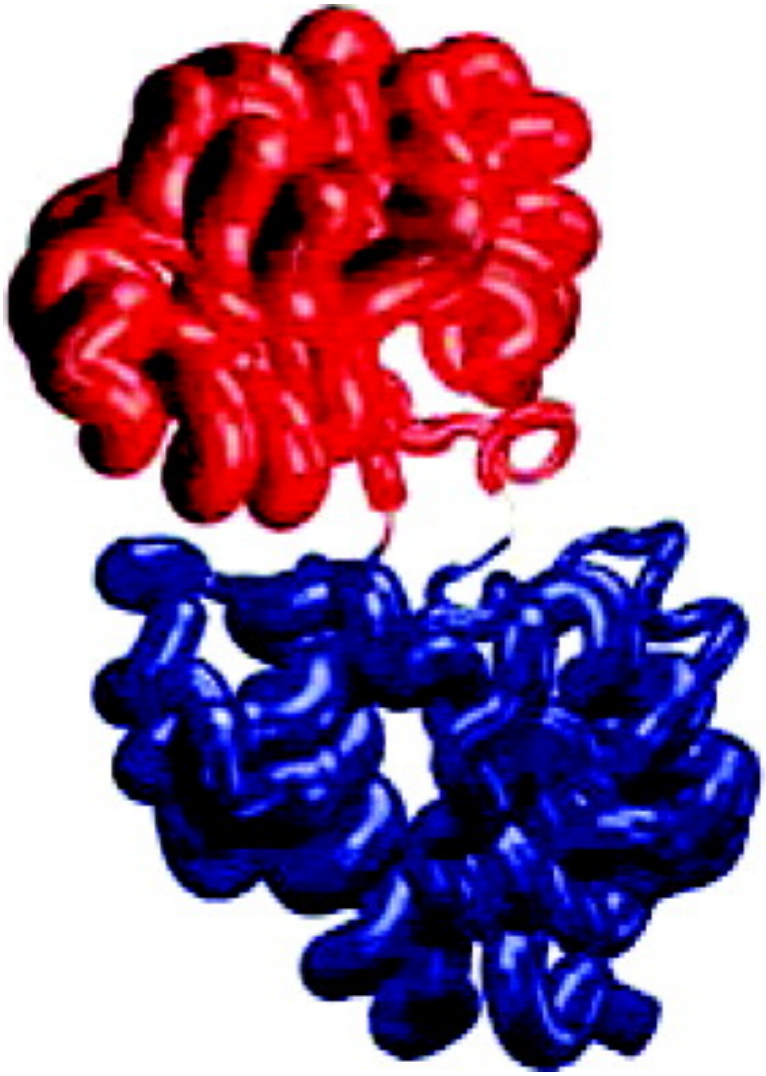
Article

Characterization of Protein Conformational States by Normal-Mode Frequencies

Benjamin A. Hall, Samantha L. Kaye, Andy Pang, Rafael Perera, and Philip C. Biggin

J. Am. Chem. Soc., **2007**, 129 (37), 11394-11401• DOI: 10.1021/ja071797y • Publication Date (Web): 23 August 2007



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 08/23/2007

Characterization of Protein Conformational States by Normal-Mode Frequencies

Benjamin A. Hall,[†] Samantha L. Kaye,[†] Andy Pang,[†] Rafael Perera,[‡] and Philip C. Biggin*[,†]

Contribution from the Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, U.K., and Department of Primary Care, University of Oxford, Rosemary Rue Building, Old Road Campus, Headington, Oxford OX3 7LF, U.K.

Received March 14, 2007; E-mail: Philip.biggin@bioch.ox.ac.uk

Abstract: Conformational change in polymers including proteins is central to many molecular processes. Defining conformational states, however, remains a difficult and increasingly common problem, with many existing methods based on arbitrary or potentially unrepresentative measures. Furthermore, the expanding length of molecular dynamics simulations and direct observation of transitions between different energy basins suggest that this issue will only become evermore important. Methods commonly used to characterize conformational states include principal component analysis, root-mean-square deviation-based clustering, and geometric measurements such as hinge angles and distances. Here we present a method where the eigenvector frequencies derived from a Gaussian network model (Bahar, I.; Atilgan, A. R.; Erman, B. *Folding Des.* **1997**, *2*, 173−181) of a trajectory of structures from a molecular dynamics simulation are used to describe the state of the protein at each time point. We apply the method to three proteins that share the same fold as the type II periplasmic binding proteins: The lysine-arginine-ornithine-binding protein, the glutamine-binding protein, and the ligand-binding domain from the NR1 *N*-methyl-D-aspartate receptor. We find that the method can distinguish different states in good agreement with a variety of previous analyses and additionally provides information on the dynamic properties of that system at a given time point.

Introduction

Conformational change in polymer systems, including proteins, is quite often central to their function. For example, channel gating in response to voltage in voltage-dependent potassium channels[2] or allostery in the conformational spread response to chemical stimuli in the chemotaxis system.[3] Despite this importance, the details of many conformational changes remain unknown. This is in part due to experimental difficulty of obtaining high-resolution crystal structures in multiple states. In addition, simply capturing two states does not necessarily tell us about the path taken in a conformational change. Such details of the change may play a key role in applications of models of conformational change; for example, in drug design, where molecular dynamics (MD) has already defined roles for induced fit[4] and water.[5] The energy landscape view of proteins[6,7] describes these changes as movement from one energy basin (or state) to another.[8] Indeed, structural information obtained from X-ray crystallography or NMR experiments usually reflects a time-averaged view of an ensemble of structures within one energy basin. Different basins equate to different conformational states, and these states can be explored by careful manipulation of the conditions, such as changing the solvent, as exemplified by the behavior of melittin,[9,10] or mutating a residue in the protein.

Tracking changes between states is still very challenging for experiments, although time-dependent NMR is able to provide some insight.[11] Simulations can, however, provide fully atomistic time-dependent information, and although we are still a long way from obtaining reasonable statistics for large conformational changes, the timescales (routinely tens of nanoseconds) are now such that some infrequent events that correspond to basin-hopping can be observed in an MD simulation.[12] More recently, structure-based Hamiltonians were used to move between conformational states.[8,13] This raises an interesting problem if we consider a transition between two states A and

[†] Department of Biochemistry.
[‡] Department of Primary Care.

(1) Bahar, I.; Atilgan, A. R.; Erman, B. *Folding Des.* **1997**, *2*, 173−181.
(2) Sands, Z.; Grottesi, A.; Sansom, M. S. P. *Curr. Biol.* **2005**, *15*, R44−R47.
(3) Bray, D.; Duke, T. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 53−73.
(4) Henchman, R. H.; Wang, H. L.; Sine, S. M.; Taylor, P.; McCammon, J. A. *Biophys. J.* **2005**, *88*, 2564−2576.
(5) Arinaminpathy, Y.; Sansom, M. S. P.; Biggin, P. C. *Mol. Pharmacol.* **2006**, *69*, 11−18.
(6) Leopold, P. E.; Montal, M.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 8721−8725.
(7) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 167−195.
(8) Okazaki, K.-I.; Koga, N.; Takada, S.; Onuchic, J. N.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11844−11849.
(9) Kemple, M. D.; Buckley, P.; Yuan, P.; Prendergast, F. G. *Biochemistry* **1997**, *36*, 1678−1688.
(10) Roccatano, D.; Columbo, G.; Fiorini, M.; Mark, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12179−12184.
(11) Harwood, N. E.; Price, N. C.; McDonnell, J. M. *FEBS Lett.* **2006**, *580*, 2129−2134.
(12) Pang, A.; Arinaminpathy, Y.; Sansom, M. S. P.; Biggin, P. C. *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 809−822.
(13) Hyeon, C.; Lorimer, G. H.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18939−18944.

10.1021/ja071797y CCC: $37.00

B. How does one know, or classify, when one is in state A (as defined by some experimentally determined structure, which in itself may not be the best representative of that state, but for these purposes will suffice) rather than in state B? Metrics that have been used in the past include angles between domains,[14] torsion angles of individual residues, interdomain distances,[15,16] and number of contacts as a fraction of contacts unique to each state.[17] Although such metrics are useful and intuitive, they often suffer from large amounts of noise in the simulation data and sometimes can lead to difficulties in assigning intermediate states to one state or the other (which may of course not always be desirable anyway). Another problem is how best to characterize the motion within each state. At the simplest level, this is often described by the rmsd of key structural components of the protein. A more detailed methodology utilizes principal component analysis (PCA) or essential dynamics to describe the low-frequency motions.[18,19] These methods are perfectly suitable to describe the motion but also suffer from the problem of (relatively) short trajectories[20] and large loop motions that can sometimes dominate the motion in the trajectory.

To address this, we present a novel method that makes use of a Gaussian network model (GNM)[1] in combination with MD simulations. GNMs have been heavily exploited to provide information on protein flexibility usually starting from just one structure.[21] Motions from a Hessian matrix, generated using an abstract model of structure, are decomposed into eigenvectors and eigenvalues, providing information on the concerted motions of the protein and their dominance, respectively. These have been found to correlate well with global motions seen from MD simulations in bimolecular systems spanning a range of different folds, functions, and sizes, including potassium channels,[22−24] SecY,[25] viral capsids,[26] and ribosome.[27] Its role in structural classification has been explored,[28] demonstrating its sensitivity to fold and conformation.

By applying GNM calculations to a series of snapshots from an MD simulation, we obtain a probability distribution of the frequency for the first eigenvector (see Supporting Information for a discussion on the validity of this approach). We show here by application to proteins from a similar fold how the variation in the probability distribution can help discriminate trajectories that are in different states. Our results demonstrate that the frequency (and the overall motion) corresponds to a different conformational state. We also demonstrate that the method is surprisingly sensitive and can, for example, discriminate subtle changes due to the presence of ligand. As we are considering ensembles of coordinates rather than single entities, the method may represent an improved way of characterizing snapshots and motions from a simulation. We demonstrate the method by application to a series of proteins (Figure 1) that are known to undergo similar conformational changes: the lysine-arginine-ornithine-binding protein (LAOBP), the glutamine-binding protein (GBP), and the ligand-binding domain (LBD) from the NR1 *N*-methyl-D-aspartate (NMDA) receptor.

Figure 1. Cartoon representation of the three different proteins used in this study. (A) LAOBP in the open-cleft form (PDB code: 2LAO). (B) GBP in the open-cleft form (PDB code: 1GGG). (C) The ligand-binding domain from the NR1 NMDA receptor in an open-cleft form (PDB code: 1PBQ. See ref 16 for more details). Proteins are colored in a blue to red color ramp from the N- to C-termini.

The motions in these proteins are central to their function. Furthermore, they belong to the same fold and represent an ideal test case.[29] LAOBP and GBP are periplasmic binding proteins

(14) Hayward, S.; Berendsen, H. J. C. *Proteins: Struct., Funct., Genet.* **1998**, *30*, 144−154.
(15) Arinaminpathy, Y.; Sansom, M. S. P.; Biggin, P. C. *Biophys. J.* **2002**, *82*, 676−683.
(16) Kaye, L. S.; Sansom, M. S. P.; Biggin, P. C. *J. Biol. Chem.* **2006**, *281*, 12736−12742.
(17) Caflisch, A. *Curr. Opin. Struct. Biol.* **2006**, *16*, 71−78.
(18) Garcia, A. E. *Phys. Rev. Lett.* **1992**, *68*, 2696−2699.
(19) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412−425.
(20) Balsera, M. A.; Wriggers, W.; Oono, Y.; Schulten, K. *J. Phys. Chem.* **1996**, *100*, 2567−2572.
(21) Jernigan, R. L.; Bahar, I. *Curr. Opin. Struct. Biol.* **1996**, *6*, 195−209.
(22) Grottesi, A.; Domene, C.; Hall, B. A.; Sansom, M. S. P. *Biochemistry* **2005**, *44*, 14586−14594.
(23) Haider, S.; Grottesi, A.; Hall, B. A.; Ashcroft, F. M.; Sansom, M. S. P. *Biophys. J.* **2005**, *85*, 3310−3320.
(24) Shrivastava, I. H.; Bahar, I. *Biophys. J.* **2006**, *90*, 3929−3940.
(25) Haider, S.; Hall, B. A.; Sansom, M. S. P. *Biochemistry* **2006**, *45*, 13018−13024.
(26) Rader, A. J.; Vlad, D. H.; Bahar, I. *Structure* **2005**, *13*, 413−421.
(27) Wang, Y.; Rader, A. J.; Bahar, I.; Jernigan, R. L. *J. Struct. Biol.* **2004**, *147*, 302−314.
(28) Kundu, S.; Sorensen, D. C.; Phillips, G. N., Jr. *Proteins: Struct., Funct., Bioinf.* **2004**, *57*, 725−733.

(PBPs) that are involved in the uptake of a variety of substrates into the cell of Gram-negative bacteria. They have a bilobed structure that acts via a Venus fly trap mechanism to bind the ligand in the cleft between the two lobes.[30,31] The bound complex is recognized by a receptor located in the inner membrane that transports the substrate into the cell by mechanisms that remain to be clarified. Detailed knowledge of the dynamics of both open-cleft and substrate-bound closed-cleft conformations of the PBPs will improve our understanding of these processes. PBPs have also been heavily exploited in the field of biosensors,[32] but problems still arise with respect to predicting which combinations of modifications will be successful, perhaps reflecting interactions with solvent and the overall dynamics.

The NMDA receptor is a subclass of the ionotropic glutamate receptors. Glutamate is the major excitatory neurotransmitter in the brain and is the physiologically relevant ligand. The ligand-binding domain shows structural similarity to the PBPs, but the closing of the cleft (due to glutamate binding) in this instance induces a mechanical strain on the transmembrane region causing it to open.[33] This opening allows the passage of cations into and out of the cell. The resulting depolarization across the membrane forms the start of neuronal signal propagation that allows neurons to communicate with each other. The receptor is thought to be arranged as a dimer of dimers,[34] and although no structure exists for the tetramer, presumably the subunits sit adjacent to each other. Thus, the dynamics of one ligand-binding domain will exert an effect on its neighbors. These interactions will play a central role in defining their response to neurotransmitters.

Although the principal motion in all three of these proteins is the cleft closure, the remaining less dominant motions will contribute to their function. For LAOBP and GBP, these less dominant motions may play a role in the transfer of the substrate to the transmembrane receptor. In the case of the NMDA receptor, these secondary motions may play a role in "tuning" the response of the receptor to the neurotransmitter. These motions may thus be important in the context of dynamic allostery.[35] The method of analysis we present suffers from less bias and system dependence compared to geometric-based analysis approaches such as a distance or angle and may therefore represent a general approach to characterizing conformational change.

Methods

Simulations. Molecular dynamics trajectories were obtained from previous reported simulations in the literature. LAOBP was reported by Pang et al.,[12] GBP was reported by Pang et al.,[36] and the ligand-binding domain from the NR1 NMDA receptor was reported by Kaye et al.[16] The LAOBP series of simulations considered here were apo simulations where ligands (arginine, ornithine, histidine, or lysine) if present were removed before the start of the simulation. In the case of GBP, the authors had extended the simulations to 10 ns post publication. Simulations corresponding to the open-apo, glutamine-bound, and closed-apo (glutamine bound state but with glutamine removed) states were considered. For LAOBP and GBP, data from 10 ns were used. For the NR1-LBD simulations, all trajectories were 20-ns long. NR1 simulations with and without the presence of the glycine ligand for both wild-type and a C744A-C798A mutant (which removes a disulfide bond linking the two domains; Figure 1) were considered. Snapshots were taken every 5 ps from each trajectory during the GNM processing. All simulations were performed with GROMACS.

Gaussian Network Model. Each structure from the simulations was analyzed with the Gaussian network model as described in ref 37. Briefly, a Kirchoff matrix was constructed based on the relative positions of carbon α atoms in the structure; if two residues were within a cutoff distance of 7 Å they were considered bound by a spring. This value was chosen because it is a typical and widely used cutoff for GNM models[37−39] and as such would appear to be the most applicable for analyzing a range of related structures. However, we tested the sensitivity of the method to the choice of cutoff from 5 to 9 Å for the LAOBP-Arg-apo simulation (Supporting Information). The results are robust to the choice of cutoff, but at lower cutoffs the singular value decomposition does not always converge. For the higher cutoffs, the interpretation of the data does not change. All bonds have equal, arbitrary spring constants. Singular value decomposition is then performed on the matrix to give eigenvalues and eigenvectors of motion. The square root of the eigenvalue is equal to the frequency of the motion, which is inversely proportional to the dominance of a given motion. In contrast with normal-mode analysis (NMA), which has six zero eigenvalues to represent rotation and translation, the GNM has a single zero eigenvalue representing rotation and translation. This is as NMA explicitly considers the Cartesian components of each spring, whereas the GNM considers the inner products of the Cartesian vectors[40] and as such is isotropic. The motion with the lowest positive nonzero eigenvalue therefore represents the dominant motion undergone by a given structure.[29]

Statistical Analysis. All statistical analysis was performed using *R*.[41] One-way analysis of variance (ANOVA[42]) was used to detect a difference between means. If statistically significant differences in the means were detected with ANOVA, Tukey's post-hoc tests were carried out for paired comparisons. Tukey's method was chosen as it adjusts for multiple comparisons.[43]

Results

Abstract Model. To demonstrate the principle of the analysis, we considered the smallest possible system that demonstrates changes in GNM frequencies. This is described in the Supporting Information.

Example Proteins. By way of application, we apply the methodology to proteins where we already have some existing knowledge about the conformational dynamics: LAOBP, GBP, and the ligand binding domain from the NR1 NMDA receptor. These proteins are all members of the same fold, and our

(29) Keskin, O.; Jernigan, R. L.; Bahar, I. *Biophys. J.* **2000**, *78*, 2093−2106.
(30) Ames, G. F.-L. *Annu. Rev. Biochem.* **1986**, *55*, 397−425.
(31) Felder, C. B.; Graul, R. C.; Lee, A. Y.; Merkle, H.-P.; Sadee, W. *AAPS PharmSci.* **1999**, *1*, article 2.
(32) Dwyer, M. A.; Hellinga, H. W. *Curr. Opin. Struct. Biol.* **2004**, *14*, 495−504.
(33) Dingledine, R.; Borges, K.; Bowie, D.; Traynelis, S. F. *Pharmacol. Rev.* **1999**, *51*, 7−61.
(34) Schorge, S.; Colquhoun, D. *J. Neurosci.* **2003**, *23*, 1151−1158.
(35) Popovych, N.; Sun, S.; Ebright, R. H.; Kalodimos, C. G. *Nat. Struct. Biol.* **2006**, *13*, 831−838.
(36) Pang, A.; Arinaminpathy, Y.; Sansom, M. S. P.; Biggin, P. C. *FEBS Lett.* **2003**, *550*, 168−174.
(37) Bahar, I.; Atilgan, A. R.; Erman, B. *Folding Des.* **1997**, *2*, 173−181.
(38) Doruker, P.; Atilgan, A. R.; Bahar, I. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 512−524.
(39) Su, J. G.; Jiao, X.; Sun, T. G.; Li, C. H.; Chen, W. Z.; Wang, C. X. *Biophys. J.* **2007**, *92*, 1326−1335.
(40) Rader, A. J.; Chennubhotla, C.; Yang, L.-W.; Bahar, I. The Gaussian Network Model: Theory and Applications. In *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems;* Cui, Q., Bahar, I., Eds.; Chapman & Hall/CRC: Boca Raton, FL, 2006.
(41) Team, R. D. C. **2006**.
(42) Miller, J. C.; Miller, J. N. *Statistics for Analytical Chemistry;* Ellis Horwood: New York, 1989.
(43) Winer, B. J.; Brown, D. R.; Michels, K. M. *Statistical Principles in Experimental Design;* McGraw-Hill: New York, 1991.
(44) Kaye, S. L.; Sansom, M. S. P.; Biggin, P. C. *Biochemistry* **2007**, *46*, 2136−2145.

Table 1. Summary of Simulation Data Used

simulation	description	time (ns)	ref
LAOBP-open-apo	open-cleft	10	12
LAOBP-Lys-apo	ligand removed before start of MD simulation	10	12
LAOBP-Arg-apo		10	12
LAOBP-Orn-apo		10	12
LAOBP-His-apo		10	12
GBP-open-apo	open-cleft form	10	36
GBP-Gln-bound	closed-cleft, ligand bound	10	36
GBP-closed-apo	ligand removed before simulation	10	36
NR1-WT-apo	open-cleft WT simulation	20	44
NR1-C744A-C798A-apo	mutant based on open-cleft WT	20	44
NR1-WT-closed-Gly	WT with glycine bound	20	44
NR1-C744A-C798A-Gly	mutant with glycine bound	20	44

simulation data set (Table 1) includes trajectories of mutants and different ligands. Thus, the influence of these factors on the conformational states the proteins explore can be assessed. Frames from each trajectory were used as input for the GNM analysis.

Lysine-Arginine-Ornithine-Binding Protein. Applying this method to related structures in the periplasmic binding protein family such as LAOBP reveals similar patterns and can be used to clearly distinguish between different states. Figure 2 shows results from LAOBP simulations that started from a closed-cleft conformation but with the ligand in each case removed. Figure 2A−C suggests a distribution that is in accord with there being only one state. This agrees with the earlier analysis of these trajectories that demonstrated that they remained in a closed state as defined by the separation of the two lobes.[12]

Figure 2D (LAOBP-Arg-apo) reveals a different pattern, a bimodal distribution, which reflects the observation that this simulation undergoes a distinct transition from closed- to open-cleft conformation. This hypothesis is supported by analysis of the LAOBP-open-apo trajectory that reveals one distinct peak centered around an eigenvalue of 0.07 (Figure 2E). That the LAOBP-His-apo, LAOBP-Lys-apo, and LAOBP-Orn-apo are one state and differ from the LAOBP-open-apo state is highlighted further by comparing box plots of their distributions (Figure 3A). These distributions can be further rationalized by a clustering analysis of the trajectory frames, based on the Cα rmsd. The frames from the LAOBP-Arg-apo can be clustered into two distinct groups using a threshold of 2.65 Å using the method described by Daura et al.[45] (Figure 3B). The same threshold gives just one cluster for all of the frames extracted from the LAOBP-His-apo, LAOBP-Lys-apo, and LAOBP-Orn-apo simulations and one cluster from the LAOBP-open-apo simulation. In addition, this change in structure can also be observed with other methods, and the eigenvalues can be seen to closely follow both the domain distance (here defined as the distance between the centers of mass of the two domains) and the projection down the dominant eigenvector defined by PCA (Figure 3C). This is discussed further later in the article.

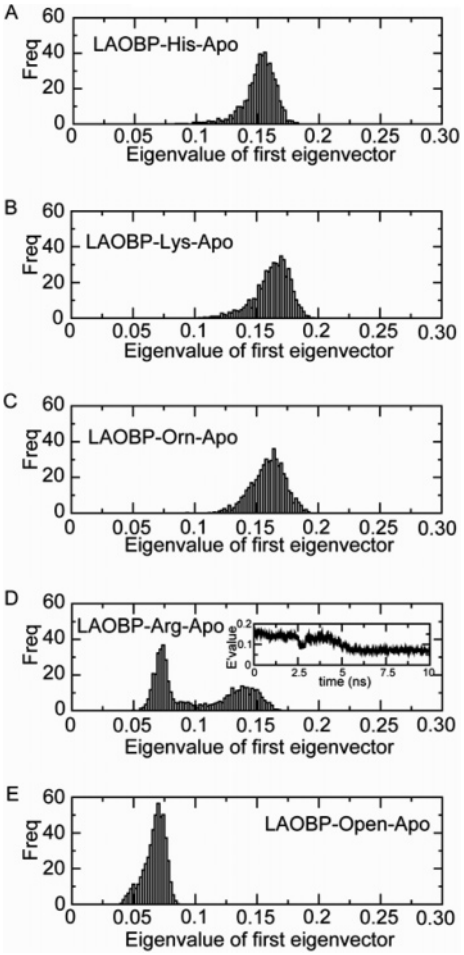


Figure 2. Analysis of the LAOBP trajectories. (A−C) Results for the closed-apo simulations from LAOBP-His, LAOBP-Lys, and LAOBP-Orn, respectively. Peaks are in approximately the same position, but they are statistically different. (D) Result for the LAOBP-Arg simulation. This appears to have two peaks: one centered around 0.14 and another peak at 0.07. This simulation has been reported to undergo a transition from a closed-cleft conformation to an open-cleft conformation. The inset shows how the eigenvalue varies with time and demonstrates that the simulation is moving in one direction rather than fluctuating between these states. (E) Analysis of the LAOBP-open-apo simulation where there is a single peak centered around 0.07.

We wanted to know, however, to what extent the eigenvector distributions were similar. Further statistical analysis (ANOVA with post comparison tests) reveals that the distributions for the LAOBP-open-apo, LAOBP-His-apo, LAOBP-Lys-apo, and LAOBP-Orn-apo are in fact all significantly different from one another at the 95% confidence level (Tables 2 and 3). That the three "closed-apo" simulations that remained in a single state (LAOBP-His-apo, LAOBP-Lys-apo, and LAOBP-Orn-apo) are all significantly different implies two things: (i) the method is surprisingly sensitive to very small changes in conformation and (ii) these three simulations have not converged (they only differ in their initial conformation). One can, however, see substantial overlap in the closed-cleft simulations (Figures 2A−C and 3A) and that these distributions have little overlap with the LAOBP-open-apo simulations (Figures 2E and 3A). This is supported by the ANOVA analysis (Tables 2 and 3) that suggests that the differences in the means is small between closed-cleft simulations compared to the difference in the mean

(45) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; Van Gunsteren, W. F.; Mark, A. E. *Angew. Chem., Int. Ed.* **1999**, *38*, 236−240.
(46) Daura, X.; van Gunsteren, W. F.; Mark, A. E. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 269−280.
(47) Preusser, A. *ACM Trans. Math. Software* **1989**, *15*, 79−89.

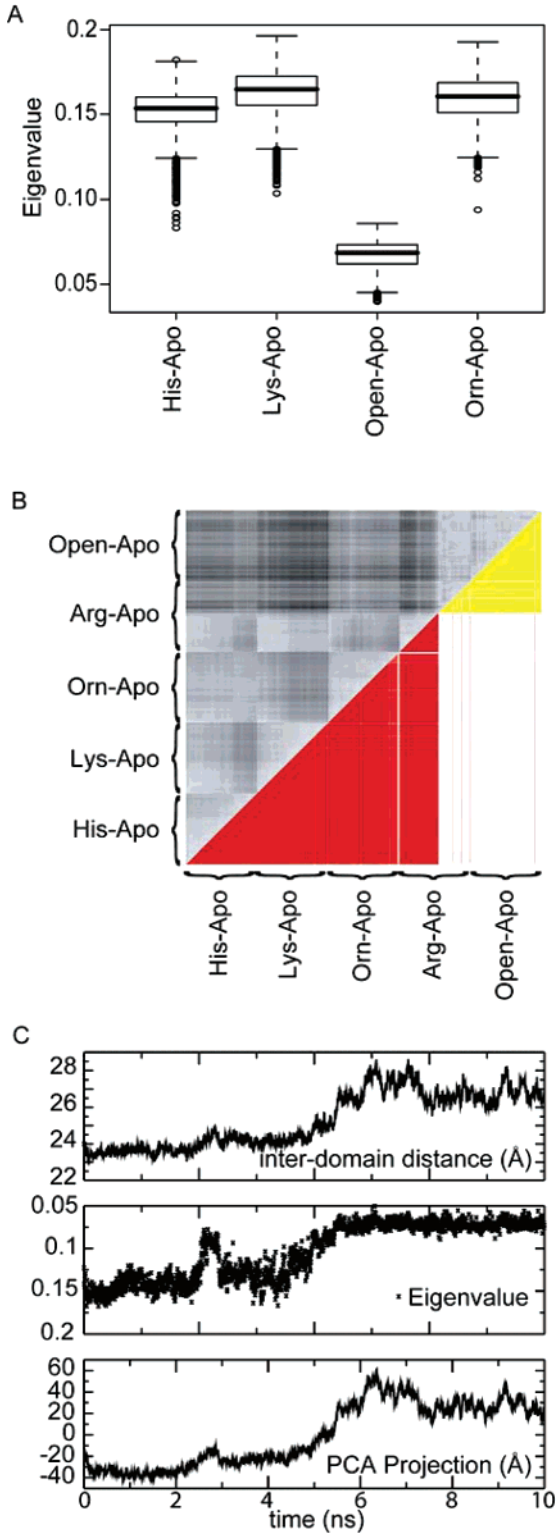


Figure 3. (A) Box plots for the LAOBP simulations that remained in a single conformational state. Whisker lines are drawn at 1.5 times the interquartile range. Outliers are indicated explicitly. The overlap between the LAOBP-His-apo, LAOBP-Lys-apo, and the LAOBP-Orn-apo simulations is substantial. Conversely, there is very little overlap between those simulations and the LAOBP-open-apo simulation. (B) The rmsd matrix of all five LAOBP simulations and the results of a cluster analysis using the method as described by Daura et al.[46] Two different states are illustrated by the yellow and red blocks in the lower half according to the clustering. It can clearly be seen that the LAOBP-Arg-apo simulation has two clusters of conformations. (C) Changes in interdomain distance (top), eigenvalue (middle), and the projection onto the first principle component (bottom) from the LAOBP-Arg-apo simulation.

Table 2. ANOVA Analysis of the LAOBP Apo Simulations (LAOBP-Open-Apo, LAOBP-His-Apo, LAOBP-Lys-Apo, and LAOBP-Orn-Apo)

	df	sum sq	mean sq	*F* value	P[a]
apo trajectories	3	12.5488	4.1829	26680	$<2.2e^{-16}$
residuals	8000	1.2543	0.0002		

[a] The low P value here implies that there is a significant difference between the mean eigenvalue for the simulations (at least between two).

Table 3. Tukey Multiple Comparisons of the Means for the LAOBP Apo Simulations (Excluding LAOBP-Arg-Apo); All Are Significant at the 95% Confidence Limit

simulation	diff[a]
LAOBP-open-apo: LAOBP-His-apo	0.084747408
LAOBP-open-apo: LAOBP-Orn-apo	0.092447208
LAOBP-open-apo: LAOBP-Lys-apo	0.095736632
LAOBP-His-apo: LAOBP-Orn-apo	0.007699800
LAOBP-His-apo: LAOBP-Lys-apo	0.010989224
LAOBP-Orn-apo: LAOBP-Lys-apo	0.003289424

[a] Note that the difference between the LAOBP-open-apo distribution and the other distributions is much larger (approximately 1 order of magnitude) than that between the LAOBP-His-apo, LAOBP-Orn-apo, or LAOBP-Lys-apo distributions.

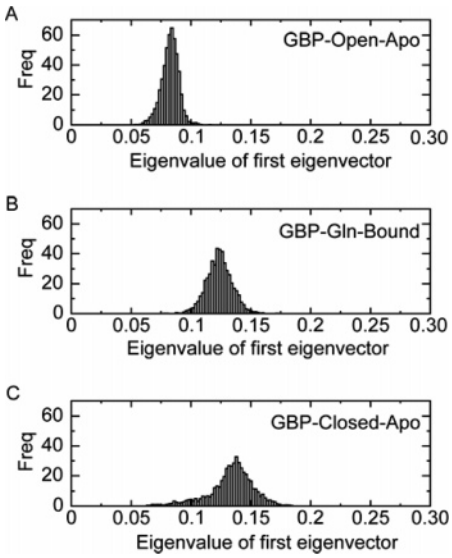


Figure 4. Results from GBP simulations for the open-apo (A), Gln-bound (B), and closed-apo (C) simulations.

between LAOBP-open-apo and the closed-cleft simulations (an order of magnitude).

Similar results are obtained for the GBP where three simulations were analyzed: an open-apo state, a glutamine-bound state, and a closed-cleft state (Figure 4). There is clear distinction between the GBP-open-apo and the two closed-cleft distributions (GBP-Gln-bound and GBP-closed-apo). Comparison of GBP-Gln-bound and GBP-closed-apo, however, raises an interesting issue. Are the differences in the distributions due to the presence of ligand or do they simply reflect a sampling issue as we suggested earlier for the LAOBP simulations? We cannot answer that comprehensively at this stage. However, given that the protein is in the same starting configuration we tentatively suggest that this difference may be attributable to the presence of the ligand.

NR1: Can Effects of Mutations Be Observed? We were interested to know whether the method could distinguish

between mutations in proteins that might affect the dynamics. To this end, we used a mutation in the NR1-LBD protein (C744A-C798A) that removed a disulfide bond that tethered the two lobes together.[44] Our previous simulation data suggested that the mutated protein more readily underwent transitions between open- and closed-cleft conformations.[44] The distributions of the principal eigenvector for WT and C744A-C798A NR1-LBD mutant in open-apo and glycine-bound states are shown in Figure 5. Comparing the distributions for NR1-WT-apo-1 and NR1-C744A-C798A-apo-1, Figure 5A,B shows two peaks common to both simulations (at 0.065 and 0.11) that appear to correspond to open and closed states, respectively. In addition, there is a third peak for the NR1-WT-apo-1 centered around 0.15. This could reflect the formation of a conformation that is more closed than the ligand-bound closed conformations.[44] In comparison, the simulations starting from closed structures with glycine bound for the wild type (Figure 5C) and the mutant (Figure 5D) show a similar set of peaks, although in the wild-type simulation the eigenvalues are dominated by a peak in the region of the closed-state peak (0.11), with a small shoulder (at 0.8). The mutant protein has a single, broad peak, residing between the two closed peaks from the NR1-WT-apo-1 simulation (Figure 5A). Note that there is no peak in either glycine-bound simulation, reflecting the maintenance of a closed-cleft conformation throughout. Overall, it appears that the mutation does not affect the frequency of the first eigenvector in the open-cleft state, but may have a stronger influence on the closed-cleft conformational state.

Comparisons with Existing Methods. By comparing the eigenvalues over time against domain distance and the projection of a trajectory down a dominant eigenvector defined by principle component analysis, we can demonstrate both the key similarities and differences between these approaches. Figures 3C and 5E show the changes in eigenvalue, projection of the trajectory along the dominant eigenvector defined by PCA, and interdomain distance. The simulation of closed apo LAOBP (based on the structure with arginine bound) demonstrates a movement over time between two distinct states: a closed conformation (from 0 to 5 ns) and an open conformation (from 5.5 to 10 ns). Both domain distance and the PCA projection continue to vary after opening, whereas the eigenvalues from the GNM remain relatively stable, reflecting what is concluded from visual inspection.

In comparison, while three states were observed in the simulation of NR1 wild type where closure was observed, the GNM eigenvalues most clearly reflected each of the three states, whereas the discrimination was less clear for one state for the other metrics. The steps observed in domain distance reflect the eigenvalues most closely, but like the LAOBP simulation the open state showed greater variation that could not be identified in the changes in eigenvalues over time. The projection onto the first principle component demonstrates the first step clearly, but the discrimination between the final two states is less clear, possibly masked by the domination of the first two states in the PCA calculation.

GNM Dynamics over Time. The elastic network model is most frequently used as a predictor of dynamics, and the change in frequency of the dominant GNM mode reflects a change in dominance of the mode over the course of the trajectory. It is thus also interesting to examine the changes in the predicted dynamics over time, to understand if the changes in conformation alter the precise shape of the dynamics. Figure 6A shows

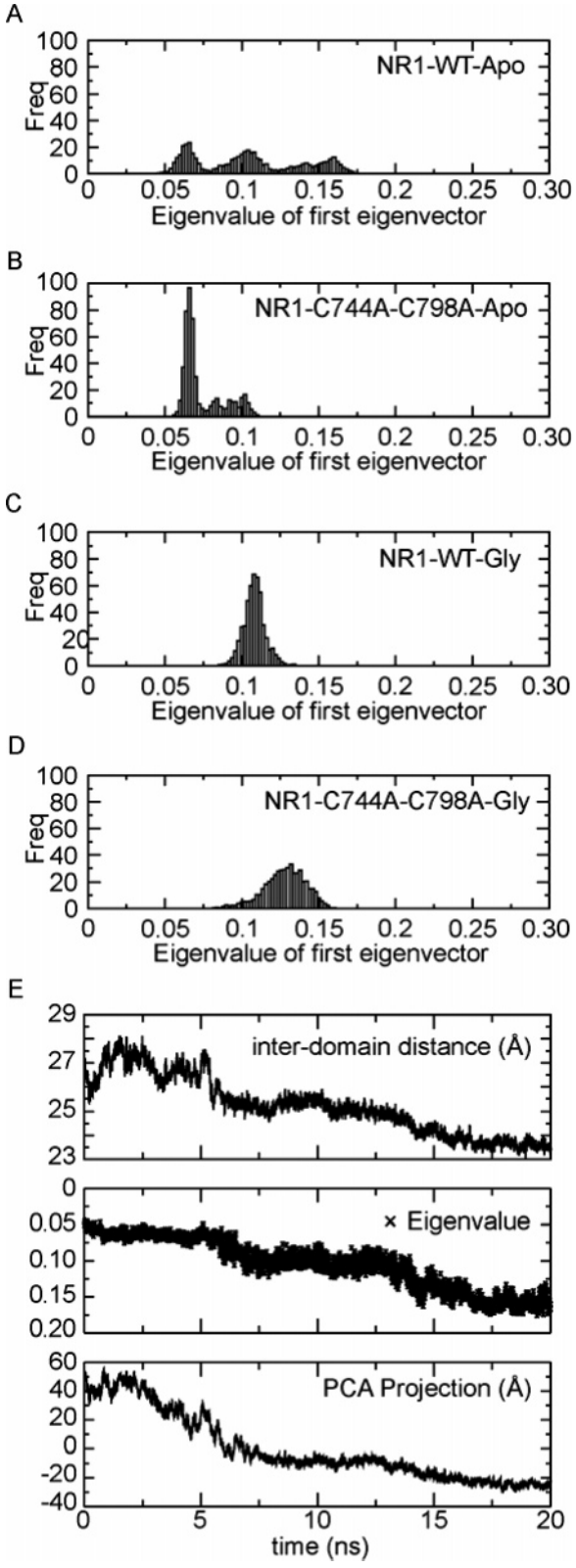


Figure 5. Results for the ligand-binding domain from the NR1 NMDA receptor: eigenvalue distributions for the first eigenvector are shown for the NR1-WT-apo simulation (A), the NR1-C744A-C798A-apo simulation (B), the NR1-WT-closed-apo simulation (C), and the NR1-C744A-C798A-Gly simulation (D). (E) Change in interdomain distance (top), eigenvalue (middle), and PCA projection (bottom) with time for the NR1-WT-apo simulation.

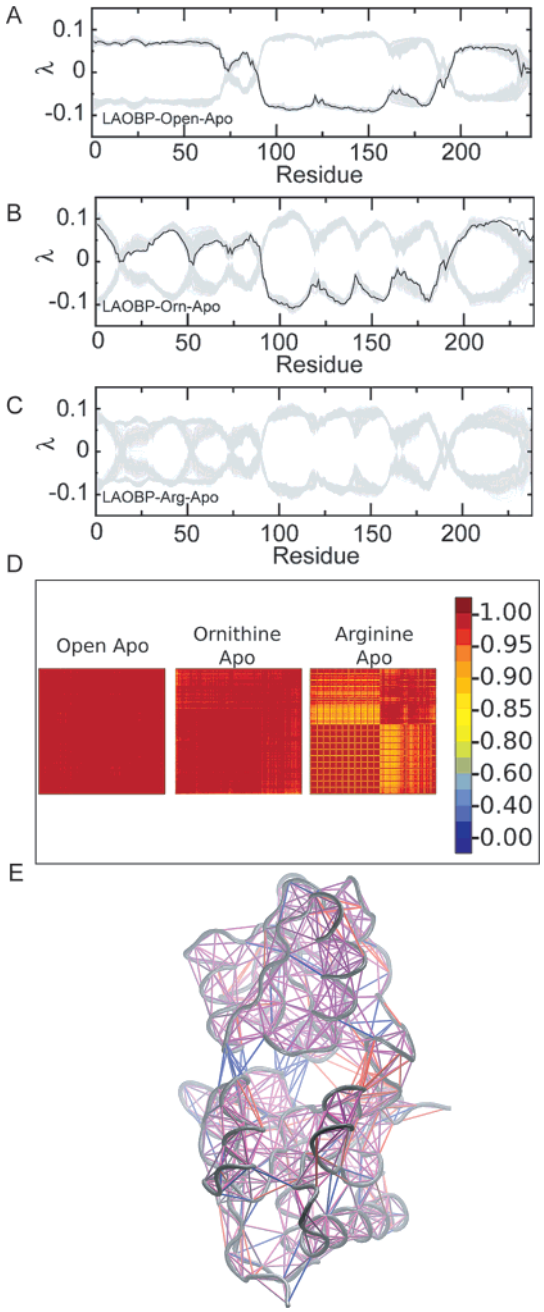


Figure 6. Comparisons of the dynamics over the course of multiple trajectories. Plot of the individual eigenvectors per residue for the LAOBP-open-apo (A), LAOBP-Orn-apo (B), and LAOBP-Arg-apo (C) simulations. Each eigenvector contains a single value for each Cα atom, due to the $n \times n$ matrix used to generate modes. The gray lines are composed of 2001 lines (the number of frames extracted from each trajectory) and are symmetrical about the zero line. This is because the GNM method is not able to distinguish phase. The LAOBP-Arg-apo simulation (C) displays a pattern that has elements of both the LAOBP-open-apo simulation (A) and the LAOBP-closed-apo (B) simulations. That they are two different populations is illustrated further by taking a frame at 1.5 and 7.5 ns from the LAOBP-Arg-apo simulation and overlaying them on top of the data for the LAOBP-open-apo (A) and LAOBP-Orn-apo (B) simulations shown as a black line. (D) Inner products of every tenth eigenvector over time for LAOBP simulations without conformational changes (LAOBP-open-apo and LAOBP-Orn-apo) and with conformational changes (LAOBP-Arg-apo). Orthogonal vectors will have a value of zero; parallel vectors will have a value of 1. Diagrams are generated with Xfarbe.[47] (E) Plot of contact matrixes from different points of the LAOBP-Arg-apo simulation onto the structure. Common contacts between Cα are plotted in purple, contacts unique to the closed state (at 2 ns) are blue, and contacts unique to the open state (at 8 ns) are red.

all eigenvectors over the trajectory overlaid. As the absolute sign of any eigenvector does not influence the dynamics (that is, an arbitrary eigenvector is interpreted as the same as the eigenvector multiplied by −1), it is possible to see that the eigenvectors all show the same patterns in the simulations where conformational change is not undergone. Furthermore, when considering the simulation where conformational change takes place, visual inspection indicates that it resembles a mix of the eigenvectors from the open and closed states. To further examine this, we calculated the normalized inner products of all vectors from a trajectory with one another, giving a value between −1 and 1 for each eigenvector. To take account of vectors that are parallel but in opposite directions, we then make all values positive, giving a value of zero for orthogonal vectors and 1 for parallel vectors. We can see from the results in Figure 6B that within a trajectory where conformational change is not undergone, the eigenvectors are very similar (>0.95). Where conformational change occurs, we can again see that the eigenvectors within a conformation are almost identical, but between conformations the inner products remain similar (>0.8). While these differences could be used in combination with the eigenvalue results to distinguish the different conformations, they are too slight and variable to be useful for changes on this scale. This, however, might not be true for all complexes undergoing a conformational change; the difference may be greater where larger changes are undergone.

By analyzing the changes in eigenvalue over all simulations, it can be seen that more closed conformations have lower eigenvalues than open conformations (Figures 3A and 5A). An immediate conclusion that can be made is that the more open conformations have fewer interdomain contacts, and thus the domains are more free to move than when closed. Hence, the eigenvalue is lower and the motion more dominant. Looking in more detail at the difference between the contact matrixes for a structure in its open and closed states (LAOBP apo simulation, based on the arginine complex at 2 and 8 ns), we can see that the sum of total contacts is only different by 6 (Figure 6E). However, when the contacts are overlaid onto the structures it can be seen that in the closed structure most unique contacts are between domains, particularly distal to the hinge region. In contrast, the open structure has multiple unique contacts in the hinge region. On the basis of these data, we suggest that the additional contacts between the domains are sufficient to cause the large change in eigenvalue by locking the cleft shut at a point opposite the hinge.

Conclusions

As atomistic MD simulations are increasingly able to sample longer timescales, structures are more able to move between two states. However, the definition of states using common metrics such as clustering, PCA, interdomain distances, or hinge angles can be inappropriate or insufficiently sensitive for many problems. We present here a novel approach that utilizes GNM calculations to characterize the (isotropic) dynamic properties of each snapshot. This approach is able to readily distinguish between open- and closed-cleft conformations in the PBP class of proteins, for example. Although the data are limited, the method may also be able to pick out quite discrete changes in dynamics attributable to mutations in the protein structure. The method has the advantage that it encompasses the contribution from all residues, but unlike distance metrics it is not subject

to undue bias from one region of the protein (e.g., a large loop motion). The approach also offers an advantage over methods such as native state contacts analysis as no prior knowledge of the states is required to characterize them.

Acknowledgment. We thank Mark Sansom for helpful discussions. B.A.H. and S.L.K. thank the MRC for a studentship. This work was supported by the Wellcome Trust.

Supporting Information Available: Details of the influence of the choice of cutoff and of the method applied to a simple three-particle system (water molecules). This material is available free of charge via the Internet at http://pubs.acs.org.

JA071797Y