

Convergence of Molecular Dynamics Simulations of Membrane Proteins

Alan Grossfield,^{1*} Scott E. Feller,² and Michael C. Pitman¹

¹IBM T. J. Watson Research Center, Yorktown Heights, New York 10598

²Department of Chemistry, Wabash College, Crawfordsville, Indiana 47933

ABSTRACT The central question in evaluating almost any result from a molecular dynamics simulation is whether the calculation has converged. Unfortunately, assessing the ergodicity of a single trajectory is very difficult to do. In this work, we assess the sampling of molecular dynamics simulations of the membrane protein rhodopsin by comparing the results from 26 independent trajectories, each run for 100 ns. By examining principal components and cluster populations, we show that rhodopsin's fluctuations are not well described by 100 ns of dynamics, and that the sampling is not fully converged even for individual loops. The results serve as a reminder of the caution required when interpreting molecular dynamics simulations of macromolecules. *Proteins* 2007;67:31–40. © 2007 Wiley-Liss, Inc.

Key words: molecular dynamics; rhodopsin; ergodicity; principal component; cluster analysis; convergence

INTRODUCTION

Molecular dynamics simulation has long been considered a powerful tool to help develop our understanding of biomolecular systems, allowing us to connect atomic-scale interactions and fluctuations to the kinetics and thermodynamics measured experimentally. However, making this connection is complicated by issues of scale: experiments typically track macroscopic numbers of protein molecules for time periods ranging from microseconds to minutes, while molecular dynamics is typically restricted to a single macromolecule, simulated for tens or hundreds of nanoseconds. The situation is further complicated when membrane proteins are involved, because of the long range structural correlations and slow relaxation times of lipid membranes. As a result, one of the primary problems associated with molecular simulation is convergence. That is, do we know if the simulation has run long enough that the results are well determined? This is an extremely difficult question, because it involves attempting to use what has been measured to deduce whether there is anything of importance which remains unmeasured. Experimentalists often attempt to answer the analogous question by replicating the experiment; repeatability is a necessary but not sufficient condition for a good measurement. This sort of test, while conceptually straightforward, has not been broadly adopted by the simulation

community for the simple reason that it is too computationally expensive; available computer time is consumed simply by running each simulation as long as possible, despite evidence that multiple short simulations sample more broadly than a single long trajectory.¹

It has long been known that standard molecular dynamics is not a particularly efficient technique for thermodynamic sampling, and as a result many specialized techniques have been developed to improve its convergence. For example, umbrella sampling² and steered dynamics³ have been used to enhance sampling to allow the calculation of potentials of mean force along a particular reaction coordinate.^{4–6} The methods are extremely useful in a broad range of calculations, such as ion permeation,⁷ peptide conformation,⁸ and ligand unbinding.⁹ However, these calculations are not without their difficulties. For one thing, one must determine in advance the relevant reaction coordinate, which in a complex system may not be obvious. Second, the addition of artificial restraint potentials to the system precludes the analysis of any kinetic phenomenon. Finally, both methods rely on sampling of all degrees of freedom orthogonal to the reaction coordinate, some of which may relax slowly and inhibit simulation convergence.

Replica exchange or parallel tempering molecular dynamics has become very popular in recent years.¹⁰ The technique combines a large number of simulations at different temperatures, with periodic exchange of temperatures to enhance the sampling at low temperatures. Despite significant successes,¹¹ this method too has a number of drawbacks. The number of replicas needed for efficient sampling increases rapidly with system size;¹² this is especially significant when considering membrane proteins, which typically require large numbers of lipids and waters to form a realistic environment. Moreover, much of the computational effort is spent sampling at artificially high temperatures (>400 K). Although several recent methods have been introduced to reduce the number replicas required,^{13,14} they have yet to be broadly adopted.

Grant sponsors: National Science Foundation, Dreyfus Foundation.

*Correspondence to: Alan Grossfield, IBM T. J. Watson Research Center, Yorktown Heights, NY 10598. E-mail: agrossf@us.ibm.com

Received 31 May 2006; Revised 22 September 2006; Accepted 23 October 2006

Published online 22 January 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21308

Finally, there are a number of sampling techniques that work by smoothing the potential energy surface, either analytically as in algorithms based on Tsallis statistics¹⁵ or multicanonical dynamics,¹⁶ numerically as in adaptive umbrella sampling,^{17,18} or algorithmically by conducting a random walk in energy space, as in density of states Monte Carlo.^{19,20} These methods can also be combined with replica exchange²¹ or other external variables²² to further improve sampling. However, they do suffer some of the same deficiencies as the methods mentioned above: they are most efficient for small systems, require a significant amount of preliminary knowledge of the system, and distort the kinetic information that is necessary to compare to many experiments.

Further, standard molecular dynamics has a number of important advantages. First, the kinetics of the system are as realistic as can be managed classically, especially if the microcanonical ensemble is used. Second, one does not have to impose a particular reaction coordinate, biasing function, etc. at the beginning of the simulation, and thus do not have to worry about artifacts due to those perturbations. Third, unperturbed molecular dynamics can make contact with experiments that have both thermodynamic and kinetic components; for example, magnetization transfer experiments in magic angle spinning NMR are related to the time correlation function for proton–proton distances.²³ For these reasons, the performance of large scale molecular dynamics is still quite important.

Recent advances in supercomputer technology^{24,25} have significantly increased our ability to run long time scale molecular dynamics simulations. Recently, we presented the results of 26 independently constructed molecular dynamics simulations of the G-protein coupled receptor rhodopsin embedded in an explicit lipid membrane.^{26,27} Where those works focused on the interactions between the protein and lipid, we examine here the behavior of the protein, focusing in particular on its fluctuations, and on the degree to which the individual simulations produce consistent results. We do so using two techniques, principal component analysis (PCA) and cluster population analysis. The former technique seeks to isolate large scale cooperative motions, while the latter focuses on the relative populations of substates on the energy landscape. Despite their differences, both methods give us the same basic message: the protein fluctuations are not quantitatively converged by a single simulation on the 100 ns time scale.

METHODS

Molecular Dynamics

We performed 26 independent 100 ns molecular dynamics simulations of rhodopsin in a membrane containing 50 1–stearoyl–2–docosahexaenoyl–phosphatidylethanolamine molecules, 49 1–stearoyl–2–docosahexaenoyl–phosphatidylcholine molecules, and 24 cholesterol molecules. The details of system construction and simulation methodology have been discussed previously,²⁶ and so we will only

briefly summarize them here. The all-atom CHARMM27 force field was used to model the protein,²⁸ while the most recent CHARMM parameters were used for the saturated²⁹ and polyunsaturated³⁰ chains, and cholesterol.³¹ The initial protein coordinates were taken from a 2.2 Å crystal structure of rhodopsin (PDB code 1U19^{32,33}). In each case, the initial coordinates for the lipid and cholesterol were regenerated, to ensure that the membrane environments are truly independent. Construction and equilibration were performed using CHARMM version 27,³⁴ while production calculations were performed using Blue Matter,²⁵ a molecular dynamics package specially written to take advantage of the Blue Gene/L hardware.²⁴ Analysis was performed using Tinker version 4.2,³⁵ with local additions and modifications. To simplify the analysis, we used only the backbone α -carbons. The first 20 ns of each simulation were excluded from the analysis as equilibration. Coordinate snapshots were saved every 100 ps, yielding roughly 800 snapshots per trajectory.

Average Structures

The average structure for a flexible molecule in a molecular dynamics simulation is usually computed by first aligning the coordinates from each snapshot onto a reference structure, then averaging the resulting coordinates. However, the choice of reference structure can clearly affect the result, and in the case of a complex system like a protein it is not clear that any of the obvious choices (e.g. the crystal coordinates, final snapshot, etc.) are optimal. To avoid this difficulty, we computed the average using an iterative procedure. First, all snapshot coordinates were aligned against the first frame, and the average structure computed. Next, the first pass average structure was used as the new reference structure, and a new average computed. The procedure was repeated until the RMS difference between successive average structures was less than 0.001 Å, typically 3 or 4 iterations. These average structures computed in this manner always had significantly lower average RMS deviations from the trajectory coordinates than average structures computed in the conventional manner. The procedure was implemented as part of the Tinker simulation package.³⁵

Principal Component Analysis

PCA is one of the most commonly used tools for analyzing protein fluctuations.^{10,36–39} This technique seeks to describe protein motions using a new basis set that directly reflects the collective motions undergone by the system. The principal components are computed by finding the eigenvectors (\vec{v}) and eigenvalues (λ) of the fluctuation covariance matrix

$$C_{ij} = \left\langle (x_i - x_i^o)(x_j - x_j^o) \right\rangle \quad (1)$$

where x_i is a coordinate of the protein, x_i^o is the equivalent coordinate for the reference structure, and the brackets denote an ensemble average.

The simplest way to assess the similarity of eigenvectors from different simulations is to compute dot products, which in effect measure the cosine of the angle between two eigenvectors. Because \vec{v} and $-\vec{v}$ are effectively the same eigenvector, we use the absolute value of the dot product

$$d_{ij} = \left| \sum_{k=1}^{3N} v_k^i \cdot v_k^j \right| \quad (2)$$

instead, where N is the number of atoms. To compare two sets of eigenvectors (e.g. from separate trajectories), we use the covariance overlap^{40,41}

$$\Omega_{A,B} = 1 - \left[\frac{\sum_{i=1}^{3N} (\lambda_i^A + \lambda_i^B) - 2 \sum_{i=1}^{3N} \sum_{j=1}^{3N} \sqrt{\lambda_i^A \lambda_j^B} (v_i^A \cdot v_j^B)^2}{\sum_{i=1}^{3N} (\lambda_i^A + \lambda_i^B)} \right]^{1/2} \quad (3)$$

The covariance overlap ranges from 0, if all pairs of eigenvectors are orthogonal, to 1, if they are identical, and is in effect an average of the squared pairwise dot products, weighted by the average displacement along the vector.

Cluster Populations

We examined the convergence of structural fluctuations of the protein using the technique recently developed by Lyman et al.⁴² First, we combined the data from all 26 of our simulations, and performed a simple clustering procedure. After choosing the first frame as a reference structure, we collected all other frames with an RMS deviation less than a given threshold value (Table I) and defined them to be part of a cluster. We then chose the first unallocated frame to be the new reference structure, and continued the procedure until all structures had been allocated to a cluster. We then compared the individual simulations to the cluster reference structures, assigning each frame to the cluster it is most similar to, allowing us to compute the populations for the clusters. This procedure was implemented as part of the Tinker simulation package.³⁵

Examining the deviations of the cluster population distributions for the various trajectories provides a measure of their convergence. We computed the population variance for a pair of trajectories as the sum of the squares of the population differences for each cluster

$$\sigma^2(P) = \sum_{i=1}^N (P(i) - P_0(i))^2 \quad (4)$$

where $P(i)$ is the population of cluster i in a particular simulation, $P_0(i)$ the population of cluster i for all simulations, and N is the number of clusters. The population variance in itself is not easily interpretable, since the probability of a given value of σ^2 depends on P_0 . Rather, we compared the mean population variance to that com-

TABLE I. Details of the Cluster Population Analysis

Structure	Length	Threshold	Clusters	Sample Size
Whole	348	3.0	79	2
Helices	232	1.5	70	2–3
N-terminus	33	1.0	426	11
		1.5	50	2
		2.0	12	*
		3.0	3	*
C-I	8	1.0	9	1–2
E-I	8	0.6	37	1–2
		0.8	10	*
		1.0	4	*
		1.0	42	3
C-II	7	1.0	42	3
E-II	28	1.0	52	3
C-III	18	2.0	143	5
E-III	9	1.0	39	2

“Length” is the number of residues in a particular portion of the molecule, “Threshold” is the value used to determine which structures should be clustered with a given reference structure, in Å, “Clusters” is how many clusters were found, and “Sample Size” is the estimated number of independent samples in a trajectory. A “*” in the “Sample Size” column meant the population was too skewed to allow estimation of the effective sample size via bootstrap.

puted via a bootstrap procedure,⁴³ where we generated artificial distributions by sampling randomly from P_0 . Since each datum in the artificial distributions is truly independent, we can compare the bootstrapped variance as a function of sample size to the mean variance seen in the simulations to estimate the effective number of independent samples in the real data.

We began each calculation using an RMS threshold of 1.0 Å; reviewing Table I shows that in general this led to a reasonable number of clusters. The primary exceptions are the C-terminus, where even a 4.0 Å threshold generated hundreds of poorly populated clusters, and the C-III loop (connecting helices 5 and 6), where a threshold of 2.0 Å still leads to 143 clusters, as opposed to thousands for a 1.0 Å threshold. In the former case, we decided there was no reasonable threshold that would produce a manageable number of clusters, while in the latter case we only pursued the higher threshold results.

RESULTS

Average Structures

The simplest way to check for convergence of a set of protein simulations is to examine the average structures. It is important to note that the average structure from a simulation may not be physically meaningful, or even energetically accessible, if the molecule has multiple distinct conformations. This does not diminish its value as a tool for characterizing the sampling, in that large variations in the average structures from the individual simulations indicate a lack of convergence. In Table II, we list the RMS deviations of the average structures for the various pieces of rhodopsin, while Figure 2 shows the probability distributions. Both show that the RMS deviation of the protein

TABLE II. Average Root Mean Square Deviations of Rhodopsin Structures

Molecule	Pairwise RMS	RMS to average	Traj RMS
Whole	2.80	1.94	1.93
Helices	1.36	0.94	0.94
N-terminus	1.02	0.72	0.90
C-I	0.39	0.28	0.32
E-I	0.25	0.17	0.31
C-II	1.05	0.74	0.64
E-II	0.65	0.45	0.64
C-III	2.12	1.47	1.62
E-III	0.74	0.53	0.61
C-terminus	5.89	4.20	4.32

Superposition was performed using only the backbone α -carbons. The pairwise RMS was computed by computing the average of the RMS deviations between the average structures from the of the 26 trajectories. The RMS to average was computed by comparison to the average of the 26 average structures. The “Traj RMS” is the average of the RMS of each trajectory to its own average structure. “Whole” indicates the entire rhodopsin molecule, while “Helices” indicates the transmembrane portion. The loops are labelled “C” and “E”, indicating the cytoplasmic or extracellular faces. Thus, loop C-I connects helices 1 and 2, E-I connects helices 2 and 3, and so forth (Fig. 1). All RMS values are in Å.

as a whole is larger than that of the transmembrane portion. This is unsurprising, since the helices are expected to be very well structured. Perhaps more surprisingly, the average loop structures for the most of the loops are very similar; the mean pairwise RMS for the average structures from the individual simulations are generally ≤ 1 Å. The only exception is C-III, which connects helices 5 and 6 (see upper right hand corner of Fig. 1): it is a long loop without significant secondary structure. By contrast, the longer E-II loop is well converged, most likely because it is stabilized by a short beta sheet.

It is important to distinguish between the RMS deviation of the average structures (columns 1 and 2 from Table II, the pairwise average and the average deviation from the global average), which should tend to zero in the limit of infinite sampling, and the RMS fluctuations within a given simulation (column 3), which reflect the flexibility of the molecule and must remain finite. This last number is particularly sensitive to the choice of a template structure; the iterative procedure we used to define the average structure consistently reduced the trajectory RMS, frequently by as much as 50%. Unsurprisingly, the trajectory RMS and average structure RMS are strongly correlated; the average is most easily sampled in the cases where the structural fluctuations are small.

The C-terminus is a clear outlier, with much higher RMS deviations than any other portion of the protein. There are a number of possible reasons for this: it is a long peptide (25 residues), without secondary structure, and unlike the other loops it is tethered on only one end. Further, residues 330–333 are missing in several of the published crystal structures (e.g. 1F88,⁴⁵ 1HZX,⁴⁶ 1GZM,⁴⁷ 1L9H,⁴⁸) which suggests that this region is highly flexible. In any event, it is clear that the structure of this peptide is not well sampled by 100 ns of dynamics.

Principal Component Analysis

PCA is a standard tool for extracting the large scale motions from molecular dynamics simulations. The principal components of protein motion are computed as the eigenvectors of the C- α fluctuation covariance matrix. The eigenvalue associated with each vector represents the variance of the molecular along that vector, which in the quasi-harmonic approximation has an associated frequency and timescale. It is thus ironic that the most important eigenvectors have the slowest timescales, and thus are most difficult to determine accurately in a finite molecule dynamics simulation.^{40,49} As shown by Balsera et al.,⁴⁹ the orthogonality of the eigenvectors means that inaccuracy in the low frequency modes “pollutes” the high-frequency modes as well.

Figure 3 shows an attempt to measure the similarity of the lowest frequency modes computed from the 26 independent simulations. Specifically, we compute the absolute value of the dot product for the first eigenvalue for each pair of simulations; in the limit of good sampling, the eigenvectors would be very similar, and thus the dot products would approach 1. Figure 3 shows the probability distribution for the dot products, using the eigenvectors computed using the whole protein and various subsets of it.

Figure 3(A) shows the probability distribution computed for the whole protein and for just the transmembrane helix region. The probability distribution for the transmembrane helices is roughly Gaussian in shape, and appears to be indicative of statistical fluctuations about a single well-defined average eigenvector. By contrast, the first eigenvector dot products for the whole protein are much less similar; some simulations produce similar eigenvectors (dot products ≈ 0.7), while others are totally different. This is not surprising; the transmembrane portion of the protein is expected to undergo limited fluctuations about a single average structure, in contrast to loops which may have a number of distinct, well-populated conformations. In particular, the inclusion of the long unstructured C-terminal tail, which we have already concluded is poorly sampled, is expected to reduce the overall similarity of the modes.

Parts B and C of Figure 3 show the same probability distributions, computed for the first eigenvectors of the loops and terminal tails. The results show that the simulations do not as a rule produce consistent eigenvectors; that is, 100 ns of sampling is not sufficient for the lowest frequency mode to converge. The exception is the C-I loop, which connects helices 1 and 2. This is a relatively short loop (8 residues), which appears to be quite rigid; Table II shows that the RMS C α deviation within a simulation is 0.32 Å. Even here, there is evidence of at least 2 distinct populations of eigenvectors; the vast majority of them are quite similar, with dot products ≥ 0.9 , but there is a secondary peak around 0.5; these eigenvectors are associated with the simulations which produced the small secondary peak in the C-I curve in Figure 2(C).

In a sense, selecting out only the eigenvector with the largest eigenvalue is an arbitrary if reasonable approach to measure the similarity of fluctuations. However, be-



Fig. 1. Ribbon diagram of rhodopsin, built from the 1U19³³ crystal structure. The helices are colored from blue (N-terminus and Helix 1) to red (Helix 8 and C-terminus). The extracellular (or intradiscal) side of protein is on the bottom, while the cytoplasmic face is on top. The figure was generated using Pymol.⁴⁴

cause it concentrates on only the largest mode, it does not measure the overall similarity of the fluctuation spaces. For example, if the second largest mode from one simulation were very similar to the largest-eigenvalue mode from another, that would not show up in Figure 3. A better and more rigorous technique is the covariance overlap [Eq. (3)], which compares all pairs of eigenvectors, weighted by their respective eigenvalues. The probability distribution for the covariance overlap between different simulations is shown in Figure 4.

Figure 4(A) shows the probability distributions for the whole protein and the transmembrane region. As with Figure 3, these results indicate that the fluctuations of the transmembrane portion are more similar than those for the protein as a whole. In both cases, the overlap values are similar to but smaller than the ones reported by Faraldo-Gómez et al.⁴¹ The authors of that paper found that the overlap values for rhodopsin dropped as they considered longer trajectories. The trajectories considered here are significantly longer (26 independent 100 ns trajectories, as opposed to a single 40 ns trajectory divided into pieces). Thus, it seems likely that their conclusion—that large-scale protein motions look like random diffusion

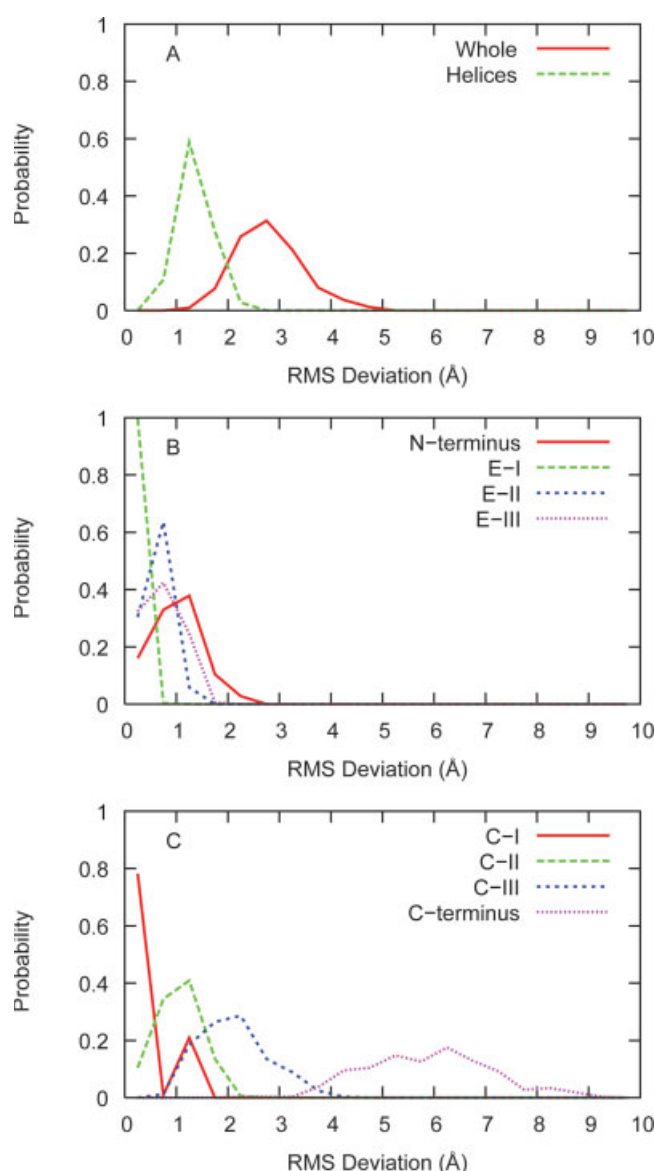


Fig. 2. Probability distribution of pairwise RMS deviation of average structures. Part A shows the distributions for the whole protein and the transmembrane helices. Parts B and C show the distributions for loops on the cytoplasmic and extracellular faces of the protein.

on the 10 ns time scale—continues to hold for the time scales considered here.

Parts B and C of Figure 4 show the probability distributions for the covariance overlaps of the loops and termini. On the whole, the extracellular face of the protein (Part B) appears better converged than the cytoplasmic face. This could imply that the cytoplasmic loops are more flexible, a notion that could have functional significance, since these loops are expected to undergo significant rearrangement upon rhodopsin activation. However, in the present context, the results mostly indicate that 100 ns is not sufficient to sample the fluctuations of even a single loop with any degree of statistical certainty. Once again, the C-I loop is an exception; most of the trajectories

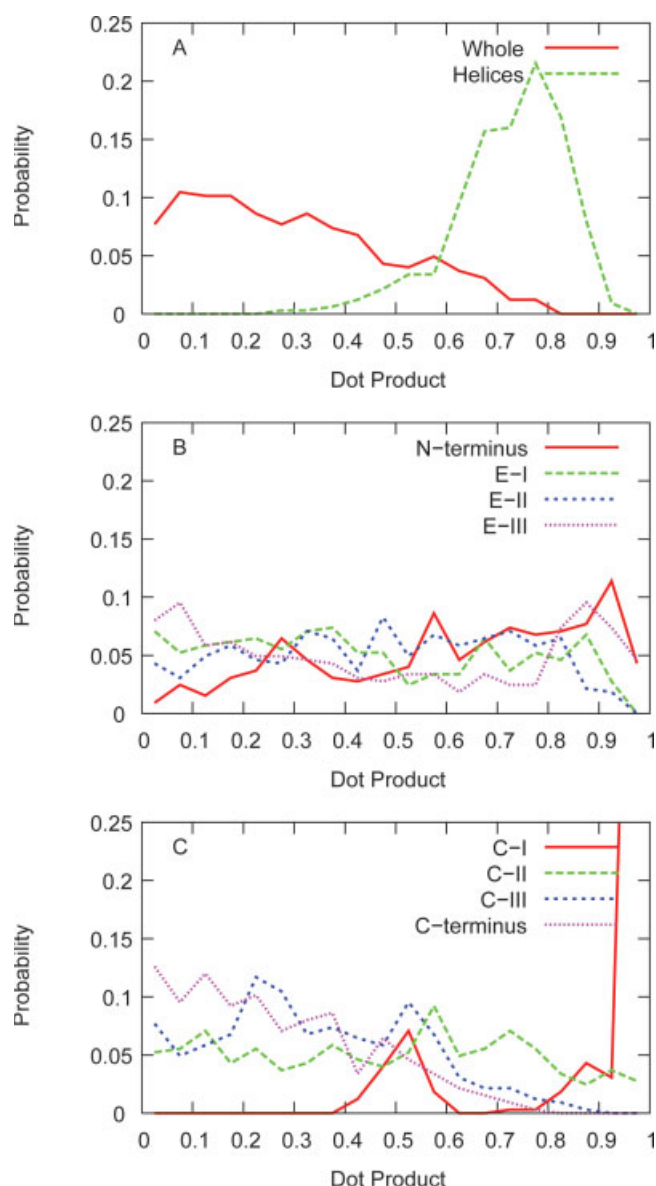


Fig. 3. Probability distribution for pairwise dot products of the first eigenvector for different portions of rhodopsin. Part A contains the whole protein and the transmembrane helices, while Parts B and C contain the extracellular and cytoplasmic loops.

sample very similar conformations spaces, although as before there is evidence of a small secondary population of simulations where the loop behaves differently. On the extracellular face, E-I and (surprisingly) the N-terminus appear to be reasonably well sampled, while E-III is more poorly converged than any portion of the protein except the C-terminus.

Cluster Population Analysis

PCA, while a valuable tool for characterizing the fluctuations captured by a simulation, does not directly describe the protein energy landscape or how it was sampled, infor-

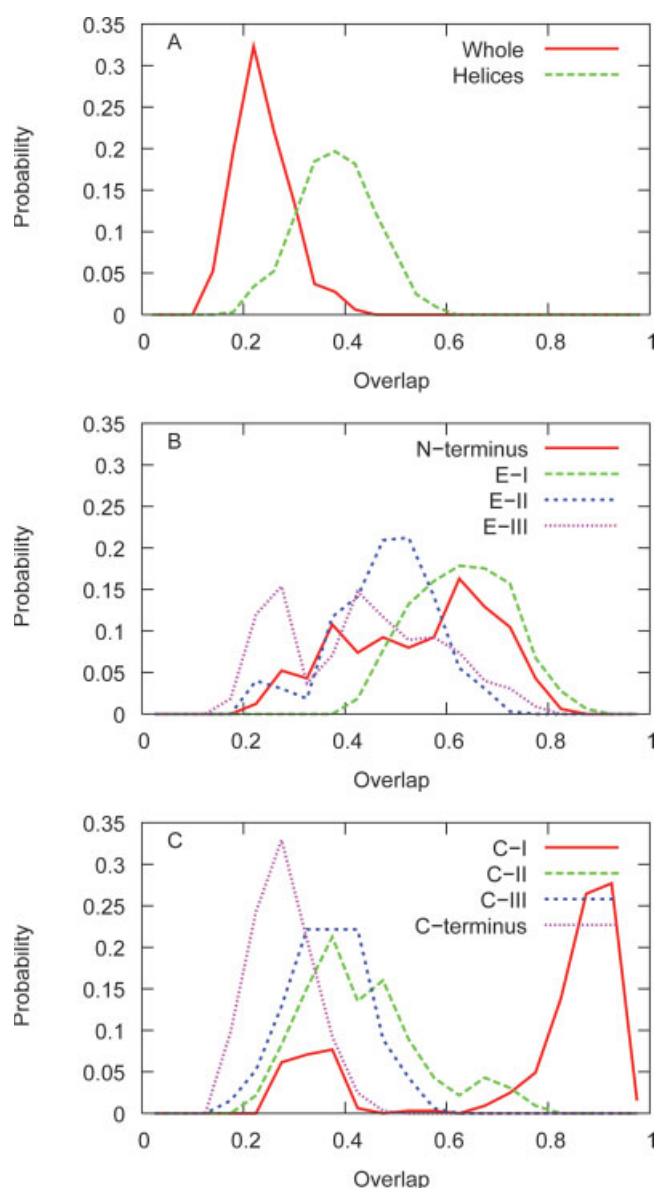


Fig. 4. Covariance overlap [Eq. (3)] for different portions of rhodopsin. Part A contains the whole protein and the transmembrane helices, while Parts B and C contain the extracellular and cytoplasmic loops.

mation that is often critical to make contact with experimental thermodynamics. In this context, it would be valuable to have a convergence test which is more directly correlated with the probability distributions in the system. Accordingly, we apply the cluster population test recently published by Lyman and Zuckerman⁴² to our rhodopsin trajectories. This method proceeds in two steps: First, the structures from all trajectories are combined and clustered (see earlier section for details). Second, each structure is assigned to a cluster, and the cluster probability distribution for each individual simulation is computed.

Each trajectory would produce the same probability distribution in the limit of infinite sampling time, as predicted by the ergodic hypothesis. In practice, sampling is

limited, and there will be statistical variations; the magnitude of these variations is a measure of the degree to which the simulations have converged. Unfortunately, there is no simple theoretical tool for interpreting the population variance directly. Instead, we applied a bootstrap Monte Carlo approach,⁴³ attempting to deduce the expected variations for a given sample size and thus quantify convergence. Specifically, we generate a series of artificial probability distributions by sampling randomly from the “true” distribution—in this case, we combine the 26 simulations and use the aggregate probability distribution as the best estimate—and compare their variance to that seen in the individual simulations. As the number of samples in each bootstrapped data set is increased, the population variance will drop. We varied the sample number such that the bootstrap variance matched that of the trajectories, which tells us the effective number of independent samples per trajectory. Given a known amount of sampling time (≈ 80 ns per simulation), this in turn gives us a measure of the correlation time. Of course, there is not a single correlation time for the entire simulation. Rather, the correlation time depends on the details of the measurement, including what part of the protein is being examined, and the resolution with which it is examined.

The results are summarized in Table I. The effective sample sizes for the various components of the protein typically range from 2 to 5. Given that the sampling portions of our simulations are roughly 80 ns, this is equivalent to saying that the correlation times for protein or loop structural fluctuations are roughly 16–40 ns. These correlation times are comparable to the simulation times of current molecular dynamics simulations of proteins, suggesting that in many cases these simulations are not long enough for their results to be well converged. This is not surprising, given the poor PCA convergence described above.

While we believe these numbers to be qualitatively correct, there is a question as to how sensitive they are to the details of the procedure, in particular to the size of the clusters. For example, the results for the N-terminus in Table I show there is clearly some dependence. The estimate of the effective sample size changes from 2 to 11 when the cluster threshold is reduced to 1 Å from 1.5 Å.

In some cases, it is possible to use other structural analyses to independently estimate the timescales for protein motion. For example, fluctuations in the axis for individual helices ought to be at least roughly related to the timescales measured by cluster analysis for the transmembrane region. Accordingly, we computed the fluctuation correlation function for each helical axis, by computing the principle moment time series for each helix, subtracting the average orientation vector, and constructing the renormalized time correlation. The correlation functions from the individual simulations are extremely noisy, but averaging over the 26 simulations produces smooth monotonic decay. However, the behavior is clearly nonexponential, and at long time intervals (20–40 ns) all 7 transmembrane helices show statistically significant negative correlation values. As such, it is difficult to quantify

the long-time dynamics of helical orientation, other than to say they are slow enough to be poorly sampled.

Further, Table I shows we were not always able to estimate the sample size, because the mean simulation variance was larger than the bootstrap variance for a sample size of 1; this occurred only when the probability distribution was dominated by a single state (population $>95\%$), with the other states poorly populated. In this case, we may violate one of the cardinal assumptions underlying the use of bootstrap Monte Carlo methods, that the source probability distribution function is reasonably accurate. If this is not the case, the variance in the actual data may be larger than that seen in the artificial data sets, even when only 1 data point is used to generate each set. This can occur if the clustering is either exceedingly coarse or exceedingly fine. For example, if the probability distribution is dominated by a single state, it follows that statistics for the other states are necessarily less precise, which in turn propagates to the artificial data sets, where it can cause them to significantly underestimate the variance. This is because the true substructure of the potential surface is usurped into a single state, with the other “states” being the edges of the distribution. On the opposite extreme, if the binning is too fine, none of the state probabilities are accurately determined, and the entire procedure devolves into noise. Further effort must be invested in order to better understand this phenomenon.⁴²

However, even this significant uncertainty does not change the basic message, that flexible loops and termini have relatively long correlation times, and are thus difficult to sample on the molecular dynamics time scale. While further research is clearly necessary to determine the optimal threshold size, etc., for performing these analyses, it is clear that cluster population analysis can be a valuable tool for assessing simulation convergence, especially when used to complement to other techniques, such as PCA.

DISCUSSION

From the beginning of the protein molecular dynamics field,^{50,51} the issues of sampling and convergence have been of major concern. A number of techniques have been developed to test the degree of ergodicity in a simulation,^{52–55} but the simple fact remains that, outside of certain trivial cases, there is no way to prove that simulation has converged. Rather, in direct analogy to scientific proof in general, we can only demonstrate either that a simulation is not converged or that it may be converged. In some sense, we know a priori that any protein simulation has not fully sampled phase space, because extremely rare events (such as spontaneous unfolding of a soluble protein under native conditions) are not seen. However, the absence of such rare events does not necessarily undermine the general utility of the simulation, if the rare events do not make a significant contribution to the partition function. Rather, the relevant question is, has the simulation run long enough that we can have confidence in its pre-

dictions? Even this question must be refined, because the question depends on which predictions are evaluated, since different phenomena have different relaxation times. For example, the conformations of individual lipids are well-sampled on the molecular dynamics time scale, but overall lipid reorganization is not.²⁶

The present work is an attempt to measure the ability of 100 ns of molecular dynamics to sample the structure and fluctuations of a well-studied membrane protein, rhodopsin. However, we believe that its implications extend well beyond the context of membrane proteins. Our analysis of rhodopsin's loops has direct implications for the loops of other proteins, including soluble proteins. Moreover, a protein loop could be considered the best-case scenario for sampling a soluble peptide; the effective tethering of the ends of the loop greatly reduces the conformation space that needs to be explored, without strong loop-protein interactions to slow the kinetics. Certainly, the poor convergence seen for the C-terminal tail should act as a caution in any analysis of an unstructured peptide, especially if only a single trajectory is examined. Indeed, the cluster population method used here was developed to analyze the fluctuations of Met-enkephalin, where it showed that many tens of nanoseconds were required to produce ergodic behavior for even a 5 residue peptide.⁴² However, the rest of the loops are more ordered, and as a rule their average structures are fairly consistent from simulation to simulation.

PCA has been applied to macromolecular simulations under several names, including quasiharmonic analysis^{56,57} and essential dynamics.^{36,37} In many cases, the goal was to extract the large-scale (and presumably low-frequency) motions of the system, with the goal of reducing the system to a manageable number of modes. These modes were then used for extrapolation to longer time scales,³⁶ visualizing system motions,⁵⁸ or as effective reaction coordinates^{10,38,39} for potentials of mean force.

Unfortunately, these applications presuppose that the large-scale principal moments derived from molecular dynamics are sufficiently well-converged as to be meaningful. However, there is significant evidence that this may not be the case. Balsera et al.⁴⁹ published an examination of PCA suggesting that until the sampling time approaches the longest time scales relevant to the system (likely to be of order a microsecond or longer), none of the computed modes will be reliable. They followed this analysis with a numerical test, comparing the modes computed in two halves of a 490 ps simulation of G-actin, demonstrating qualitatively that there was little agreement between them. More recently, Faraldo-Gómez et al.⁴¹ applied the same basic test to 10 ns trajectories for several membrane proteins, drawing similar conclusions. Hess^{40,59} performed a rigorous quantitative analysis, concluding that in a molecular dynamics simulation, the "long" time scale motions of folded proteins resemble random diffusion rather than systematic motion.

The simulations involved in the present work are an order of magnitude longer than the ones examined previously,^{41,59} and we compare a large number of independent

trajectories instead of different portions of a single trajectory. However, we still find that, despite our much larger computational investment, the principal modes from the individual simulations still are not well converged. Further, the added computational effort does not appear to have improved the convergence; the covariance overlaps for two halves of a 10 ns rhodopsin simulation⁴¹ and the individual 100 ns trajectories from the present work are comparable. Furthermore, the situation does not improve significantly when smaller subsets of the protein are considered. With the exception of the C-I loop, even the fluctuations of individual loops do not appear to be well sampled, as measured by either principal component analysis or cluster population analysis.

The analysis presented here is specific to a particular membrane protein, rhodopsin. However, the results are consistent with previous results from other membrane proteins⁴¹ and soluble proteins,^{40,59} and it is reasonable to assume that the dynamics of rhodopsin's loops are comparable to those of soluble proteins. It is more difficult to generalize the results for the transmembrane helices; the lipid membrane is ordered on far larger length- and time-scales than water, and this clearly will have implications for protein dynamics; lipid lifetimes at the surface of the protein are in the tens of nanoseconds, as opposed to tens of picoseconds for water. Moreover, the basic GPCR architecture is quite different from anything found outside a membrane, and it is these gross geometric features which largely determine the large-scale fluctuations.⁶⁰

CONCLUSIONS

Molecular dynamics simulation has long been used to develop our understanding of biomolecules, including proteins, nucleic acids, and membranes. As computer power has increased, so too have the lengths of molecular dynamics simulations. However, assessing the degree to which the simulations have converged has remained problematic, largely due to difficulties in analyzing a single trajectory. The present work seeks to resolve this difficulty by comparing the results from 26 independent trajectories, using the variations between the simulations as a measure of ergodicity. We find that 100 ns of simulation, while long by today's standards, is not sufficient to describe the fluctuations of rhodopsin in any quantitative way. Further, we find that even the individual loops are not well sampled, due to their high flexibility and long correlation times. In particular, the results of principal component analysis, a very popular technique for characterizing large protein motions, are not consistent among the individual simulations.

This work has major implications for the design of future simulations involving proteins and peptides. Rhodopsin is similar in size to many membrane proteins, especially other G-protein coupled receptors, and its loops can serve as exemplars for the sampling requirements of structured peptides and the loops of soluble proteins. These results demonstrate clearly even relatively small

systems require extensive sampling before their results converge, and serve to caution us in our attempts to interpret simulations of ever larger systems. Moreover, we believe that it is important to consider the advantages of running several relatively long trajectories, as this approach not only improves convergence but also provides the means to measure it.

ACKNOWLEDGMENTS

We acknowledge the contributions of the Blue Matter team (B. Fitch, R. Germain, A. Rayshubskiy, T.J.C. Ward, M. Eleftheriou, F. Suits, Y. Zhestkov, R. Zhou, J. Pitera, and W. Swope), and thank Dan Zuckerman, Ed Lyman, and Jay Ponder for helpful discussions.

REFERENCES

1. Caves LSD, Evanseck JD, Karplus M. Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin. *Protein Science* 1998;7:649–666.
2. Patey GN, Valeau JP. A Monte Carlo method for obtaining the interionic potential of mean force in ionic solution. *J Chem Phys* 1975;63:2334–2339.
3. Schulten K. Manipulating proteins by steered molecular dynamics. *J Mol Graph Mod* 1998;16:289.
4. Roux B. The calculation of the potential of mean force using computer simulation. *Comp Phys Comm* 1995;91:275–282.
5. Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J Chem Phys* 2003;119:3559–3566.
6. Park S, Schulten K. Calculating potentials of mean force from steered molecular dynamics simulations. *J Chem Phys* 2004;120:5946–5961.
7. Bernèche S, Roux B. Energetics of ion conduction through the K⁺ channel. *Nature* 2001;414:73–77.
8. Drozdov AN, Grossfield A, Pappu RV. Role of solvent in determining conformational preferences of alanine dipeptide in water. *J Am Chem Soc* 2004;126:2574–2581.
9. Saam J, Tajkhorshid E, Hayashi S, Schulten K. Molecular dynamics investigation of primary photoinduced events in the activation of rhodopsin. *Biophys J* 2002;83:3097–3112.
10. Sanbonmatsu KY, Garcia AE. Structure of met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Proteins: Struct Func Gen* 2002;46:225–234.
11. Zhou R. Trp-cage: folding free energy landscape in explicit water. *Proc Nat Acad Sci USA* 2003;100:13280–13285.
12. Kofke DA. On the acceptance probability of replica-exchange Monte Carlo trials. *J Chem Phys* 2002;117:6911–6914.
13. Cheng X, Cui G, Hornak V, Simmerling C. Modified replica exchange simulation methods for local structure refinement. *J Phys Chem B* 2005;109:8220–8230.
14. Liu P, Kim B, Friesner RA, Berne BJ. Replica exchange with solute tempering: a method for sampling biological systems in explicit water. *Proc Nat Acad Sci USA* 2005;102:13749–13754.
15. Andricioaei I, Straub JE. On monte carlo and molecular dynamics methods inspired by Tsallis statistics: methodology, optimization, and application to atomic clusters. *J Chem Phys* 1997;107:9117–9124.
16. Okamoto Y. Generalize-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. *J Mol Graph Model* 2004;22:425–439.
17. Bartels C, Karplus M. Probability distributions for complex systems: adaptive umbrella sampling of the potential energy. *J Phys Chem B* 1998;102:865–880.
18. Bartels C, Karplus M. Multidimensional adaptive umbrella sampling: applications to main chain and side chain peptide conformations. *J Comput Chem* 1997;18:1450–1462.
19. Rathore N, Knotts IVTA, de Pablo JJ. Configurational temperature density of states simulations of proteins. *Biophys J* 2003;85:3963–3968.
20. Mastny EA, de Pablo JJ. Direct calculations of solid-liquid equilibrium density-of-states Monte Carlo simulations. *J Chem Phys* 2005;122:124109–1–124109–6.
21. Mitsutake A, Okamoto Y. Replica-exchange extensions of simulated tempering method. *J Chem Phys* 2004;121:2491–2504.
22. Chopra M, Muller M, de Pablo JJ. Order-parameter-based Monte Carlo simulations of crystallization. *J Chem Phys* 2006;124:134102–1–134102–8.
23. Soubias O, Gawrisch K. Probing specific lipid-protein interaction by saturation transfer difference NMR spectroscopy. *J Am Chem Soc* 2005;127:13110–13111.
24. Allen F, Almasi G, Andreoni W, Beece D, Berne BJ, Bright A, Brunheroto J, Cascaval C, Castanos J, Coteus P, Crumley P, Curioni A, Denneau M, Donath W, Eleftheriou M, Fitch B, Fleisher B, Georgiou CJ, Germain R, Giampapa M, Gresh D, Gupta M, Haring R, Ho H, Hochschild P, Hummel S, Jonas T, Lieber D, Martyna G, Maturu K, Moreira J, Newns D, Newton M, Philhower R, Picunko T, Pitera J, Pitman M, Rand R, Royyuru A, Salapura V, Sanomiya A, Shah R, Sham Y, Singh S, Snir M, Suits F, Swetz R, Swope WC, Vishnumurthy N, Ward TJC, Warren H, Zhou R. Blue Gene: a vision for protein science using a petaflop supercomputer. *IBM Syst J* 2001;40:310.
25. Fitch BG, Germain RS, Mendell M, Pitera J, Pitman M, Rayshubskiy A, Sham Y, Suits F, Swope WC, Ward TJC, Zhestkov Y, Zhou R. Blue Matter, an application framework for molecular simulation on blue gene. *J Para Distrib Comp* 2003;63:759–773.
26. Grossfield A, Feller SE, Pitman MC. A role for direct interactions in the modulation of rhodopsin by omega-3 polyunsaturated lipids. *Proc Nat Acad Sci USA* 2006;103:4888–4893.
27. Grossfield A, Feller SE, Pitman MC. Contribution of omega-3 fatty acids to the thermodynamics of membrane protein solvation. *J Phys Chem B* 2006;110:8907–8909.
28. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack JrR, Evanseck J, Field M, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kucera K, Lau F, Mattos C, Michnick S, Ngo T, Nguyen D, Prodhom B, Reiher W III, Roux B, Schlenkerich M, Smith J, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling dynamics studies of proteins. *J Phys Chem B* 1998;102:3586–3616.
29. Klauda JB, Brooks BR, MacKerell Jr AD, Venable RM, Pastor RW. An ab initio study on the torsional surface of alkanes and its effect on molecular simulations of alkanes and a DPPC bilayer. *J Phys Chem B* 2005;109:5300–5311.
30. Feller SE, Gawrisch K, MacKerell AD, Jr. Polyunsaturated fatty acids in lipid bilayers: Intrinsic and environmental contributions to their unique physical properties. *J Amer Chem Soc* 2002;124:318–326.
31. Pitman M, Suits F, MacKerell J AD, Feller SE. Molecular-level organization saturated and polyunsaturated fatty acids in a phosphatidylcholine bilayer containing cholesterol. *Biochem* 2004;43:15318–15328.
32. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. *Nucleic Acids Res* 2000;28:235–242. Available at <http://www.rcsb.org/pdb>.
33. Okada T, Sugihara M, Bondar AN, Elstner M, Entel P, Buss V. The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *J Mol Biol* 2004;342:571–583.
34. Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy minimization and dynamics calculations. *J Comput Chem* 1983;4:187–217.
35. Ponder JW. TINKER: Software Tools for Molecular Design 2004; 4.2: <http://dasher.wustl.edu/tinker/>.
36. Amadei A, Linsem ABM, Berendsen HJC. Essential dynamics of proteins. *Proteins: Struct Func Gen* 1993;17:412–425.
37. deGroot BL, Amadei A, Scheek RM, van Nuland NAJ, Berendsen HJC. An extended sampling of the configurational space of HPr from *E. coli*. *Proteins: Struct Func Gen* 1996;26:314–322.
38. Nymeyer H, Garcia AE. Simulation of the folding equilibrium of α -helical peptides: a comparison of the generalized Born approximation with explicit solvent. *Proc Nat Acad Sci USA* 2003;100:13934–13939.
39. Zhang W, Wu C, Duan Y. Convergence of replica exchange molecular dynamics. *J Chem Phys* 2005;123:154105–1–154105–9.

40. Hess B. Similarities between principal components of protein dynamics and random diffusion. *Phys Rev E* 2000;62:8438–8448.
41. Faraldo-Gomez JD, Forrest LR, Baaden M, Bond PJ, Domene C, Patargias G, Cuthbertson J, Sansom MSP. Conformational sampling and dynamics of membrane proteins from 10-nanosecond computer simulations. *Proteins* 2004;57:783–791.
42. Lyman E, Zuckerman DM. Ensemble based convergence analysis of biomolecular trajectories. *Biophys J* 2006;91:164–172.
43. Efron B, Tibshirani RJ. An introduction to the bootstrap. Boca Raton, FL: Chapman and Hall/CRC; 1998.
44. DeLano WL. The PyMol user's manual. DeLano Scientific; 2002.
45. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BJ, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. Crystal structure of rhodopsin: a g protein-coupled receptor. *Science* 2000;289:739–745.
46. Teller DC, Okada T, Behnke CA, Palczewski K, Stenkamp RE. Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of g-protein-coupled receptors (GPCRs). *Biochem* 2001;40:7761–7772.
47. Li J, Edwards PC, Burghammer M, Villa C, Schertler GFX. Structure of bovine rhodopsin in a trigonal crystal form. *J Mol Biol* 2004;343:1409–1438.
48. Okada T, Fujiyoshi Y, Silow M, Navarro J, Landau EM, Shichida Y. Functional role of internal water molecules in rhodopsin revealed by X-ray crystallography. *Proc Natl Acad Sci U S A* 2002;99:5982–5987.
49. Balsera MA, Wriggers W, Oono Y, Schulten K. Principal component analysis and long time protein dynamics. *J Phys Chem* 1996;100:2567–2572.
50. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. *Nature* 1977;267:585–590.
51. Karplus M, McCammon JA. Protein structural fluctuations during a period of 100 ps. *Nature* 1979;277:578–578.
52. Allen MP, Tildesley DJ. Computer simulations of liquids. Oxford: Clarendon Press; 1987.
53. Mountain RD, Thirumalai D. Measures of effective ergodic convergence in liquids. *J Phys Chem* 1989;93:6975–6979.
54. Flyvbjerg H, Petersen HG. Error estimates on averages of correlated data. *J Chem Phys* 1989;91:461–466.
55. Cao J, Berne BJ. On energy estimators in path integral Monte Carlo simulations: dependence of accuracy on algorithm. *J Chem Phys* 1989;91:6359–6366.
56. Karplus M, Kushick J. Method for estimating the configurational entropy of macromolecules. *Macromolecules* 1981;14:325–332.
57. Andricioaei I, Karplus M. On the calculation of entropy from covariance matrices of atomic fluctuations. *J Chem Phys* 2001;115:6289–6292.
58. Garcia AE. Large-amplitude nonlinear motions in proteins. *Phys Rev Lett* 1992;68:2696–2699.
59. Hess B. Convergence of sampling in protein simulations. *Phys Rev E* 2002;65:031910–031910–10.
60. Rader AJ, Anderson G, Isin B, Khorana HG, Bahar I, Klein-Seetharam J. Identification of core amino acids stabilizing rhodopsin. *Proc Natl Acad Sci USA* 2004;101:7246–7251.