Systematic Study of Anharmonic Features in a Principal Component Analysis of Gramicidin A

Martin Kurylowicz, Ching-Hsing Yu, and Régis Pomès*
Molecular Structure and Function Programme, Hospital for Sick Children and Department of Biochemistry, University of Toronto, Toronto, Canada

ABSTRACT We use principal component analysis (PCA) to detect functionally interesting collective motions in molecular-dynamics simulations of membrane-bound gramicidin A. We examine the statistical and structural properties of all PCA eigenvectors and eigenvalues for the backbone and side-chain atoms. All eigenvalue spectra show two distinct power-law scaling regimes, quantitatively separating large from small covariance motions. Time trajectories of the largest PCs converge to Gaussian distributions at long timescales, but groups of small-covariance PCs, which are usually ignored as noise, have subdiffusive distributions. These non-Gaussian distributions imply anharmonic motions on the free-energy surface. We characterize the anharmonic components of motion by analyzing the mean-square displacement for all PCs. The subdiffusive components reveal picosecond-scale oscillations in the mean-square displacement at frequencies consistent with infrared measurements. In this regime, the slowest backbone mode exhibits tilting of the peptide planes, which allows carbonyl oxygen atoms to provide surrogate solvation for water and cation transport in the channel lumen. Higher-frequency modes are also apparent, and we describe their vibrational spectra. Our findings expand the utility of PCA for quantifying the essential features of motion on the anharmonic free-energy surface made accessible by atomistic molecular-dynamics simulations.

INTRODUCTION

Computer simulation has become an essential research tool for understanding how the dynamics of proteins link their structure to their function (1–4). Although molecular-dynamics (MD) simulations have often been successfully used to obtain such insight, the size of the resulting data set makes interpretation difficult. Different parts of a complex molecule may play various functional roles on different length scales and timescales, and it is difficult to identify these motions among the wealth of data resulting from MD trajectories. In this work, we present a number of quantitative measures that identify anharmonic motions for an archetypal ion channel, gramicidin A (gA), and argue that these motions are relevant to function. We obtained these measures not only in the large-covariance regime, but also in the small-covariance regime, which is usually ignored as noise, where we identify short and fast collective motions with known functional consequences in the channel backbone.

Functional dynamics of gramicidin A

Detailed atomistic studies of ion channels present a special opportunity for understanding the structural and dynamic correlates of function. Because of its relative simplicity, gA is by far the most studied and best understood ion channel, and its structural and functional properties have been extensively characterized (5–7). It is a passive transmembrane pore that is selective for small monovalent cations. Its structure has been characterized at high resolution by ^{1}H-NMR in lipid micelles (8) and by solid-state NMR in lamellar-phase lipid bilayers (6,9). The native channel is composed of two monomers that assemble as a head-to-head noncovalently linked dimer, forming a cylindrical pore when solvated in a membrane bilayer (see Fig. S1 in the Supporting Material). Each monomer has 15 alternating L- and D-amino acid residues that form a $\beta^{6.3}$-helix with 2.5 turns per monomer. Four Trp residues stabilize the C-terminals at the water-membrane interface. The gA helix forms a 4-Å-wide cylindrical pore that hosts a single file chain of water molecules traversing the membrane, thereby creating a pathway for cation permeation and a hydrogen-bonded wire for the conduction of protons.

Many computational studies have been performed on gA (7,10). Solvation and hydrogen bonding play an important role in modulating the conduction of protons along water chains. Hydrogen bonding between lumen water molecules and backbone carbonyl groups is thought to play a role in organizing the water wire within the channel by providing surrogate solvation to the hydrated proton, thereby catalyzing the Grotthuss mechanism of proton transport (11–13). It has also been demonstrated that the backbone modes of gA impact its ionic conduction properties (14), especially in comparison with linked analogs of the dimer (13).

The surrogate solvation of cations by carbonyl oxygens in the gramicidin backbone has a long history of study. A peptide-plane libration mechanism was first proposed on the basis of experimental conductance measurements (15). A normal mode analysis (NMA) study concluded that there is a band of short-wavelength (high-frequency) modes between 75 cm^{-1} and 175 cm^{-1} that represent librational motions of the peptide planes (16). Early MD studies

Submitted March 30, 2009, and accepted for publication October 15, 2009.

*Correspondence: pomes@sickkids.ca

Ching-Hsing Yu's present address is SciNet Consortium, University of Toronto, Toronto, Canada.

Editor: Benoit Roux.

0006-3495/10/02/0386/10 $2.00

doi: 10.1016/j.bpj.2009.10.034

concluded that the mobility of the gA channel modulates its conductivity, and suggested that picosecond librations of the carbonyl moieties lining the pore were coupled to the fluctuation of water molecules and of ions in the lumen (14). Tian and Cross (17) reviewed the experimental evidence for carbonyl tilting in gA. NMR studies have provided experimental evidence of peptide plane librations (18,19). Powder-pattern NMR revealed picosecond librations (19), whereas ^{15}N T_1 relaxation measurements indicated a nanosecond timescale (18), although this slower result was interpreted as the effect of damping by slower correlated motions. In recent MD studies, the amplitude of these librations was computed (13), and significant agreement with the amplitudes measured by NMR was found. The frequency of carbonyl librations was also measured by far-infrared (far-IR) spectroscopy (20,21) and found to be in general agreement with the NMA results reported by Roux and Karplus (16).

Principal component analysis and protein dynamics

Principal component analysis (PCA) is a well-established technique (22,23) for extracting collective modes of displacement from an atomistic MD trajectory by diagonalizing the time average of the covariance matrix for any atomic subset of interest. The application of PCA to protein dynamics was pioneered by García (24), who demonstrated that there are multimodal distributions of PCs along a simulated protein trajectory, and hence any harmonic approximation of protein dynamics will fail to capture the essential features of their collective dynamics. This is not to say that the harmonic portions of the landscape are unimportant; indeed, much has been learned from harmonic approximations around an equilibrium structure through the use of NMA (16,25–28) and related elastic network models (29–33). These approximations can discern regions of enhanced flexibility in protein based on the topology of interresidue contacts, which can be useful for identifying ligand-binding sites (29) or understanding anisotropic responses to external forces (31). However, although such work can elucidate the potential energy landscape of an equilibrium protein conformation, it is an approximation based on a single structure, lacking the ''essential''anharmonicity of an atomistic force field that allows for conformational changes on a multimodal free-energy landscape. Furthermore, there is no way to account for temperature or solvent in a static approximation, the influence of which also necessitates the generation of atomistic MD trajectories.

Since it is often the large-amplitude motions that are of interest to biochemists, PCA can afford significant data reduction by concentrating a large fraction of a system's total fluctuations into a small fraction of the collective motions. To this end, many studies have examined the largest few PCs of protein motion (24,26,34–37). Amadei et al. (34) used PCA to extract what they termed the ''essential dynamics'' of proteins, and argued that only the largest non-Gaussian-distributed PCs are sufficient to account for the functional dynamics of a protein. Large collective displacements can be used to study conformational changes, and these are often the best examples of functional motion in proteins. Many proteins bind their ligands through very specific conformational changes around the binding site. Global conformational changes may also exert mechanical forces in the function of molecular motors, such as myosin (38), or facilitate chemical catalysis in the modification of chemical bonds, as in serine proteases (39). Since PCA can be used to compute the root mean-square (RMS) fluctuations along the protein backbone, it has been particularly successful in identifying large collective motions that may be related to functional changes in conformation, such as the hinge-bending motion of thermolysin (40), regions of hydrogen exchange in cytochrome *c* (35), and protein folding (41,42).

Large-scale global conformation changes are not the only interesting feature of protein dynamics. Although the tertiary and quaternary structural changes can span the size of an entire protein, and we would expect the largest PCs to capture these motions, individual residues also have important collective motions at much smaller length scales, and modification of hydrogen bonds within the secondary structure occurs on even shorter scales. Biomolecular processes also span at least 9 orders of magnitude in time, from femtoseconds (bond vibrations) to milliseconds (folding). The covariance eigenvalues of short and fast collective motions are necessarily smaller than those of long and slow motions. Hence, these motions may not be represented in the largest set of PCs, and may even be found among the small-covariance eigenvectors normally ascribed to motions arising from thermal noise. Luckily, there is nothing intrinsic to PCA that gives more meaning to large-covariance motions than to small ones. In this study, we conducted a comprehensive analysis of the entire set of PCs and found non-Gaussian-distributed PCs with small-covariance eigenvalues. We argue that these are also ''essential'' in the same sense as the largest components (34), in that they span the anharmonic portion of the free-energy landscape.

Although most NMA and elastic network studies have focused on the longest wavelengths as in PCA studies, some have studied the shorter wavelengths as well (16,26,28). For example, the normal modes of the binding pocket of wild-type α-lytic protease were found to have a symmetric character, vibrating in phase to maintain the size of the binding pocket, whereas a nonbinding mutant had asymmetric modes that resulted in contraction and expansion of the binding site (28). Regarding the normal modes of gramicidin, it has been argued that the frequency separation of collective modes spanning the whole protein (<50 cm^{-1}) and modes describing amide plane fluctuations (75–175 cm^{-1}) rules out coherent librations of many amide planes (16). The fact that functional features have been found in the short-wavelength regime of NMA justifies the examination of the same regime in PCA of simulated MD trajectories, where the results include the

influence of an anharmonic molecular force field as well as temperature and solvent effects.

METHODS

Molecular dynamics

Simulations were carried out using the CHARMM 31.1 MD package (43) with the TIP3P water model (44) and the CHARMM22 force field (45) for all other atoms. The gA dimer was embedded in a membrane of 122 glycerol-1-monooleate (GMO) molecules and solvated by 3210 water molecules. A periodic bilayer of GMO molecules was constructed in a cubic box with 50 Å sides. After 100 ps of relaxation dynamics, a cylindrical hole was created in the center of the membrane and a gA dimer whose structure was obtained from simulations inside a phospholipid bilayer was inserted (46). The entire system was then equilibrated for 2 ns before the production runs described below were performed.

Two sets of simulations were carried out to probe the long- and short-timescale dynamics of the gramicidin dimer. In one 64 ns production run, longer-timescale dynamics were probed using a 2 fs time step and saving coordinates every 200 fs, and the SHAKE algorithm (47) to constrain stretching of covalent bonds involving hydrogen. In a 10 ns production run, there were no bond constraints within the protein, and a 0.5 fs time step (saving every 10 fs) was used to probe hydrogen dynamics at shorter timescales and yield accurate PCA eigenvalues at the shortest length scales.

The leapfrog Verlet algorithm was used to propagate dynamics with constant surface tension and normal pressure on the membrane based on the Parrinello-Rahman barostat as described previously (48), with a piston mass of 500 amu and a 5 ps coupling constant. The surface tension was set to zero, since the application of external pressure has been shown to be unnecessary for GMO bilayers (49). The area per lipid was stable at 0.25 nm^2, in quantitative agreement with a previous study (49), upon equilibration of the membrane with gA inserted. The Nosé-Hoover algorithm was used to control temperature at 300 K with a thermal piston mass of 1000 kcal ps^2 and 5 ps coupling constant. The simulations were carried out with tetragonal periodic boundary conditions, and with the crystal parameters for box length updated every 200 ps. We used the particle mesh Ewald method of charge summation, with a width $\kappa = 0.3$ $Å^{-1}$, a switching function from 7 to 7.5 Å, a cutoff at 8 Å, and grid point spacing of 1.0 Å. The Lennard-Jones interactions used a force switching function from 10 to 12 Å, with a cutoff at 14 Å.

PCA

Consider a trajectory of N atoms in time $F(X_i, Y_i, Z_i, t)$, where $i = 1,2,\ldots N$ and $t = 1,2,\ldots T$, with T equal to the duration of the trajectory. To study only the internal dynamics of a protein, it is conventional to align each snapshot in the trajectory before computing the average structure, thereby eliminating the translation and rotation of the entire molecule from the trajectory. The mass-weighted covariance matrix is

$$\left\langle M_{ij}\Delta R_i \Delta R_{ij}\right\rangle, \quad (1)$$

where $M_{ij} = M_i^{1/2} M_j^{1/2}$, and $\Delta R = R - \langle R\rangle_t$ is the change of position from the time-averaged structure, for each spatial component $R_i = \{X_i$ or Y_i or $Z_i\}$ of all atoms i and j included in the analysis. Diagonalization of this matrix yields a set of eigenvalues σ_k^2 and eigenvectors $\mathbf{v}_k$, where $k = 1,2,\ldots,3N-6$.

Each eigenvector $\boldsymbol{v}_k$ represents a PC of displacement and may be visualized as a set of N three-dimensional vectors attached to the N atoms analyzed within the molecule. Each of these three-dimensional vectors describes the magnitude and direction of the RMS fluctuations at a given atom, within a given PC. The MD trajectory can be projected onto each eigenvector by forming the dot product of atomic displacements with each eigenvector for all time steps. The resulting distribution of each projection would have a variance σ_k^2 (and standard deviation σ_k); this is the physical meaning of the eigenvalues, which measure the spatial amplitude of each PC across the full trajectory.

RESULTS AND DISCUSSION

Convergence of PCA and conformational sampling

To demonstrate the convergence of conformational sampling for our MD simulations and the PCA results derived from them, we calculate the eigenvalue-normalized overlap introduced by Hess (50), which was adopted as a measure of convergence in a number of studies (51–53). In the Supporting Material we present the overlap of independent PCA results for consecutive trajectory segments of equal duration, as well as the overlap of various subsets of a trajectory with its full duration. These measures quantify the degree to which the phase space of our simulation has ceased to expand. For the NCαC main chain, the average overlap of half the trajectory with its full length is >0.9 for both the 64 ns and 10 ns simulations. The average overlap of two segments of equal duration is >0.8 for all timescales $>$ 1 ns. These values demonstrate that the backbone eigenvectors of gA are well converged.

A similar analysis of the side chains shows that their PCs are not converged, although this measure of overlap is dominated by the convergence of the longest PCs; the shorter PCs, which we discuss below, converge much faster. In addition to the distributions shown in Figs. 2 and 3, in Fig. S4 a plot of PC1 versus PC2 demonstrates multibasin dynamics with limited sampling of each well. Infrequent jumps between basins correspond to a 120° change in χ_1 of a single Trp side chain (Trp 9 or 15 of one monomer, or Trp 11 of the other) or of two side chains simultaneously (Trp 15 of one monomer and Trp 11 of the other). MD studies in vacuo (54) and in dimyristoylphosphatidylcholine (DMPC) (55) described six rotameric states available to each Trp in gA, although only Trp 9 showed a significant number of transitions ($n = 18$) among them in a 100 ns simulation in DMPC (55). Our results are in general agreement with the latter study, indicating that Trp rotameric basins are visited on the 10 ns timescale in the GMO membrane, and therefore the longest side-chain PCs are not be expected to converge within 64 ns.

It is important to note that the spatial shapes of eigenvectors converge much faster than their dynamics: the former requires a quarter oscillation in time, whereas the latter requires many cycles to adequately converge the distribution of states. We address the convergence of dynamics in our treatment of PC distributions below.

Scaling of PCA eigenvalues

The complete PCA eigenvalue spectra for various atomic subsets of gA are shown in Fig. 1. These results are taken from the 10 ns simulation without constraints on hydrogen

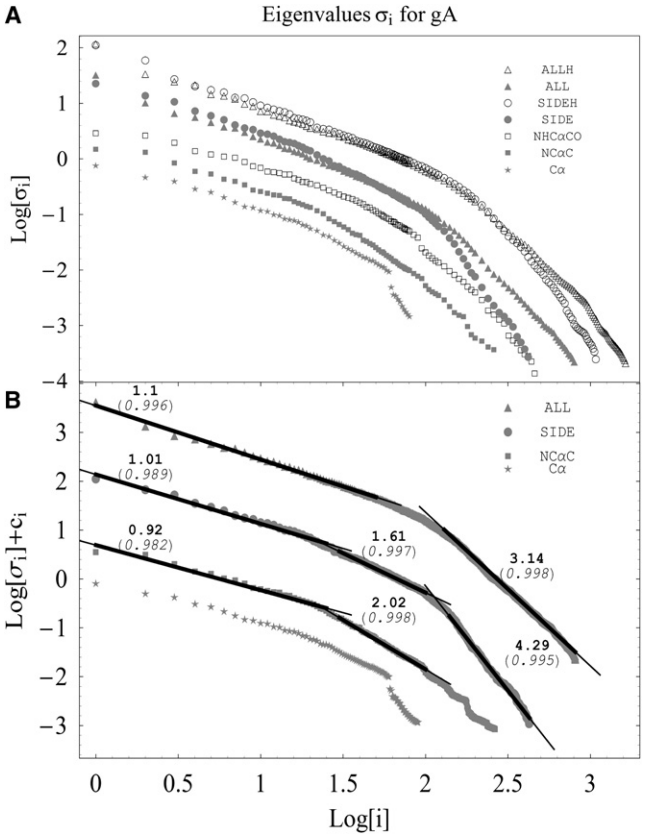

FIGURE 1 Log_{10}-log_{10} plots of the complete PCA eigenvalue spectrum of gA as a function of eigenvalue index *i*. (*A*) Spectra for the backbone, side chains, and whole protein without hydrogen atoms (*gray*: Cα,NCαC, SIDE, ALL) and with them (*black*: NHCαCO, SIDEH, ALLH). The data have been thinned toward the high indices for clarity. (*B*) The complete PC set for heavy atoms, with linear regressions in regions of different power-law scaling. The ALL curve has been translated upward for clarity ($+c_i$). Bold lines indicate the range included in the fit, and the thin lines are guides for the eye. The bold numbers above each line indicate the slope of the fit (i.e., the power α), and the R^2 value for the linear fit is italicized in brackets below the slope.

atoms, and with a 0.5 fs time step. Simulations with SHAKE (47) do not yield the correct eigenvalues at the short-PC end of the spectrum because they freeze the covalent H-bond vibrations. This in turn limits the number of collective degrees of freedom to $<3N$–6 and yields artificially small eigenvalues for degrees of freedom involving hydrogen atoms. On the other hand, the long PCs in the 10 ns simulation (without holonomic constraints) differ very little from those of the 64 ns simulation used for comparison, with an overlap of 0.88.

Each curve in Fig. 1 *A* shows the variance of all PCs for a different atomic subset of the molecular structure: a single atom per residue (C_α), backbone atoms ($NC_\alpha C$ and $NHC_\alpha CO$), side-chain atoms (SIDE and SIDEH), and the combined atom set (ALL and ALLH). Whereas the SIDEH and ALLH sets include all hydrogen atoms, the $NHC_\alpha CO$ curve includes only the amide hydrogen to emphasize the dynamics within the secondary structure involving hydrogen bonds, and to explicitly capture the amide plane motions. Although the long-PC spectra for the whole protein (ALL) and the side chains (SIDE) are almost identical, the scaling of their short PCs is significantly different. This suggests that the small eigenvalues and eigenvectors may encode real physical information about the behavior of our system and are not just noise to be ignored, as was commonly done in previous PCA studies of protein motion.

There are generally two scaling transitions in the spectra: one at ~25 PCs and one at ~100 PCs. Fig. 1 *B* shows three distinct power-law scaling regimes in the heavy-atom PCA eigenvalues $\sigma_k = k^{-\alpha}$, with all linear regressions on the log-log scale scoring $R^2 > 0.98$. Although the largest PCs follow a power of $\alpha \sim 1$, there are significant differences in scaling of the shorter PCs for different parts of the protein. The midsize regime of the backbone scales with $\alpha = 2$, whereas for side chains it is distinctly more shallow, with $\alpha \sim 1.5$. The whole protein (ALL) lacks a clear scaling in this midscale regime, and makes a smooth transition toward steeper scaling at the shortest end of the PC spectrum. In this short-PC regime, the backbone scales with roughly $\alpha \sim 2$, the side chains scale much more steeply with $\alpha = 4$, and the whole protein (ALL) approaches an average between the two, with $\alpha = 3$. Different numbers of PCs span the same scaling features in these spectra for different atomic inclusions (e.g., C_α versus $NC_\alpha C$), suggesting that blocks of PCs span statistically distinct regimes of motion. Hence, the scaling shown in Fig. 1 may be used as a guide to search for components of motion with interesting statistical features and determine the boundaries between distinct regimes of PCs.

Non-Gaussian PC distributions

Gaussian distributions are indicative of motion on a harmonic free-energy landscape, whereas non-Gaussian distributions are the result of anharmonicity on this landscape. This follows from the definition of free energy $F = -k_B T \log(P)$, where a Gaussian probability $P = \exp[-ax^2]$ yields the harmonic function $F = Kx^2$ with spring constant $K = ak_B T$.

To compare the acquired distribution P_k with a Gaussian, we normalized each PC trajectory by its standard deviation σ_k (the square root of the *k*th covariance eigenvalue σ_k^2). We also rebinned the distributions into a common 100 bins to align all distributions with each other for comparison. The resulting normalized distributions P_k^N have $\sigma_k = 1$, and their shape can be compared against a Gaussian distribution of unit variance and height $(2\pi)^{-1/2}$. Fig. 2 shows ΔP_k^N, the difference between the acquired PC distributions and a unit Gaussian, for the gA backbone with and without hydrogen-bonding atoms ($NHC_\alpha CO$, $NC_\alpha C$), and for the heavy side chain atoms (SIDE). Each panel shows results averaged over multiple windows of width 1 ns (64 samples) to 64 ns (one sample), taken from the 64 ns simulation. A single distribution for the entire 10 ns simulation without holonomic restraints is also shown for comparison, since

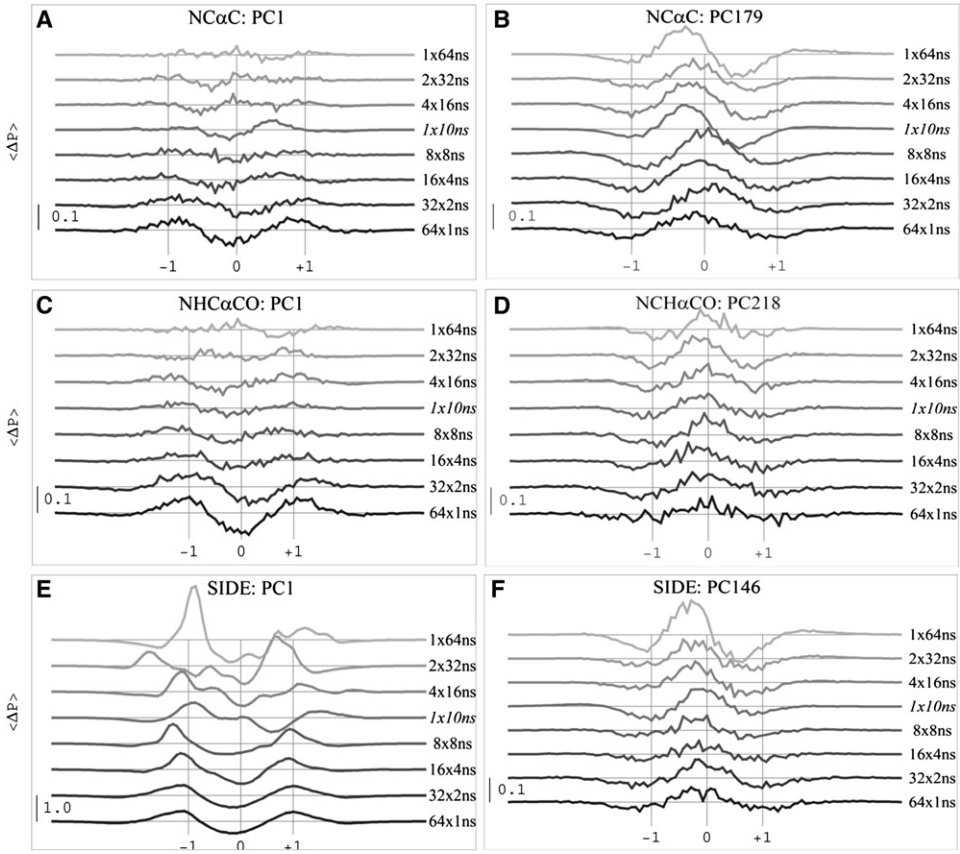

FIGURE 2 Difference between eigenvalue-normalized PC distributions and a unit Gaussian, ΔP, for the longest PC (*left*) and a representative short PC (*right*) of the backbone and side chains of gA. PCA was executed independently for multiple windows at various timescales from the 64 ns simulation (with holonomic restraints), and PC distributions were averaged for a given timescale. Results for the 10 ns simulation (without restraints) are also shown for comparison.

short PCs probe dynamics involving hydrogen atoms. PC1 is shown on the left and a representative short PC is shown on the right.

Although the multimodal distributions of side chains (Fig. 2 *E*) are not surprising, the non-Gaussian features of the longest backbone PCs (Fig. 2, *A* and *C*) exhibit a similar profile at short timescales, though with smaller amplitude. Although these backbone PCs have the appearance of superdiffusive distributions (which undersample near the average and oversample extremes) at short timescales, at long timescales $<\Delta P_k^N>$ approaches zero, indicating that the longest PCs are actually harmonic (Gaussian). This time dependence is an artifact of inadequate sampling, as the distribution of an oscillation sampled over less than the order of a wavelength appears asymmetric. This analysis suggests that the dynamics of the gA backbone converge at ~10 ns. By contrast, there is no dependence on timescale for the non-Gaussian features of the short PCs (Fig. 2, *B*, *D*, and *F*), indicating that these subdiffusive profiles (undersampling extremes and oversampling the average) correspond to persistent anharmonic aspects of backbone and side-chain dynamics. The short PCs shown in Fig. 2 are representative of groups of neighboring PCs that have similar distributions.

Fig. 3 shows ΔP_k^N surfaces for all 270 PCs for $NC_\alpha C$ atoms, 470 PCs for $NHC_\alpha CO$, and 430 PCs for SIDE atoms from the 10 ns simulation. Flat regions indicate PCs with nearly perfect Gaussian distributions ($\Delta P_k^N < 0.01$), whereas peaks and valleys indicate non-Gaussian distributions and suggest anomalous diffusion of those PCs in time. The landscapes show central peaks for short PCs indicating subdiffusion. Although the subdiffusive features at high PC are concentrated at a single length scale for $NC_\alpha C$, this is not the case for the main-chain ($NHC_\alpha CO$) or side-chain atoms. The $NHC_\alpha CO$ results reveal clusters of subdiffusive modes across a number of length scales at a high PC index. This is also true of the heavy side-chain atoms, where it is interesting to note that these features ride on a superdiffusive envelope. Fig. 3 also shows the RMS deviation between the acquired PC distribution and a Gaussian curve for each atomic subset. These plots show the distribution of anharmonic features across the short PC spectrum and reveal distinct bands of subdiffusive components.

Anomalous diffusion in the mean-square displacement autocorrelation function

All of the results presented above describe the spatial characteristics of the system averaged over time. To make proper contact with anomalous diffusivity, we now study the time-ordered behavior of our system by computing the mean-square displacement (MSD) of the cumulative PC projection $x(t)$ (see the Supporting Material for more details):

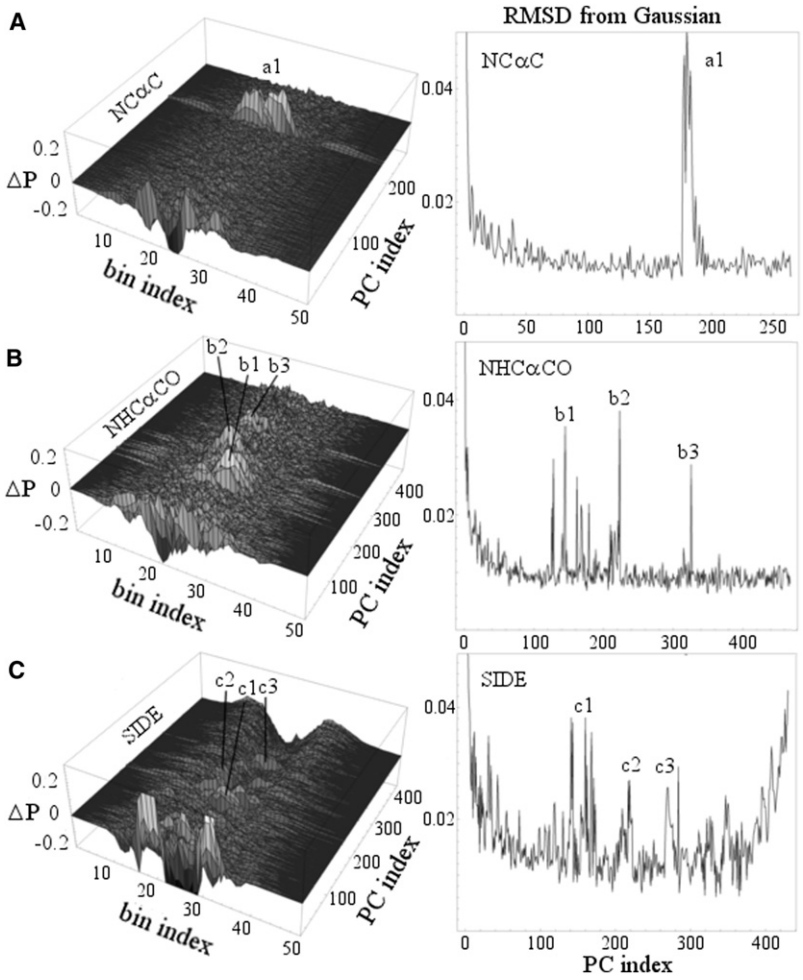

FIGURE 3 Surfaces on the left show the difference between the normalized PC distribution and a unit Gaussian, ΔP, for all components in the PCA of heavy backbone atoms ($NC_{\alpha}C$), main chain ($NHC_{\alpha}CO$), and heavy side-chain atoms (SIDE) in gA. The RMS difference between acquired distributions and a Gaussian distribution is shown on the right.

$$MSD(t) = \left\langle (x(t - t_0) - x(t_0))^2 \right\rangle. \quad (2)$$

The term ''anomalous diffusion'' properly applies to systems whose particles have an MSD that scales nonlinearly in time (56). These are non-Brownian processes that obey a generalized Einstein relation:

$$MSD(t) = 2dD_{\beta}t^{\beta}, \quad (3)$$

where D_{β} is the (anomalous) diffusion coefficient and d is the dimensionality of the system. If $\beta < 1$, a process is subdiffusive in the sense that it moves away from its average more slowly than Brownian diffusion (''sublinear''). A subdiffusive process has antipersistent correlations, where consecutive steps are more likely to move in opposite directions than they would in a random walk. If $\beta > 1$, a process is superdiffusive in that it moves away more quickly than Brownian diffusion (''superlinear''). A superdiffusive process exhibits persistent correlations, where consecutive steps are biased to continue in the same direction.

Fig. 4 shows the MSD for a representative subset spanning all PCs of the $NHC_{\alpha}CO$ and SIDE atomic subsets, across six orders of magnitude in time, from the 10 ns simulation. A careful examination of Fig. 4 reveals a number of interesting features. First, there is a leveling of MSD(t) at long timescales past ~1 ns (with the exception of the first PC). This leveling is a result of the fact that we are analyzing a bounded system of fixed volume: at some timescale, all PCs must cease moving away from their average and return to it. Thus the rollover in

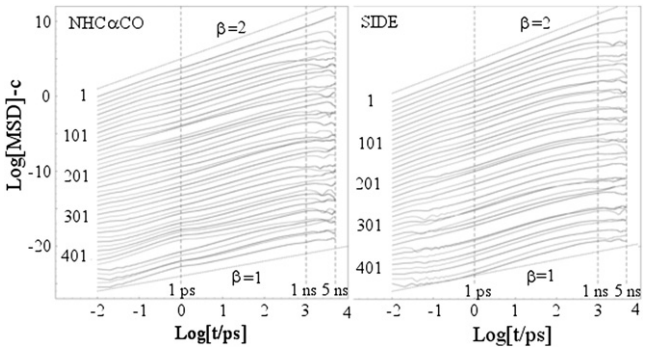

FIGURE 4 MSD of every 11th PC for the $NHC_{\alpha}CO$ and SIDE atomic subsets. The curves are evenly spaced by a constant c at their origin. Linear ($\beta = 1$) and ballistic ($\beta = 2$) values of slope β are shown in dotted gray as a guide for the eye.

the MSD may be considered an ''edge effect'', though it may also contain interesting information about the timescales of the collective motions in our system. For example, we would expect covariant motions at smaller spatial scales in the protein to be bounded at increasingly short timescales, and this is evident in Fig. 4. A comparison of the MSD curves in Fig. 4 with the lines of slopes 1 and 2 (*dotted gray lines*) makes some general trends apparent. The longest PCs scale with $\beta = 2$, indicating ballistic motion unimpeded by thermal perturbation, whereas shorter PCs tend toward $\beta = 1.5$ or even $\beta = 1$. Moreover, there is a nontrivial structure to the groupings of trends (in time) among PCs in the backbone, which is made evident by the changes in spacing between groups of curves in the figure.

The most interesting feature in Fig. 4 is the observation of pronounced oscillations among the shortest PCs at timescales below ~1 ps. These oscillations are most visible in the case of the side chains, although they are also present in the backbone at lower frequencies. This suggests that the subdiffusive features apparent in the non-Gaussian distributions of short PCs are a result of short-timescale oscillations rather than longer-timescale subdiffusive sampling. The superposition of locally subdiffusive PCs on a global superdiffusive envelope in the side chains of Fig. 3 may also be attributable to this interplay of short- and long-timescale behaviors. Note that although oscillations are not ''diffusive'', they meet the definition of subdiffusion in that consecutive steps are anticorrelated at a particular timescale.

To amplify small changes in the scaling of the MSD, in Fig. 5 we plot the instantaneous slope of the MSD as a function of time for long and short groups of PCs. These plots also highlight the fact that consistent power-law scaling is persistent on all timescales up to ~100 ps for all PCs (and up to ~1 ns for the longest PCs). These plots reveal a surprising array of oscillations in the short PC regime, with consistent frequencies across groups of PCs and transitions to higher frequencies for shorter PCs. This figure also makes it clear that there is a general transition at ~1 ps between very short timescale behavior and longer timescale dynamics (100 ps–1 ns). This is the expected ballistic ($\beta = 2$) to diffusive ($\beta < 2$) transition for the longest PCs, indicating the timescale at which atomic motions become restrained by thermal perturbations of their directions and velocities of motion. However, for short PCs the opposite trend is also apparent in the backbone, from slower subdiffusive scaling at short timescales to faster diffusive scaling at long timescales.

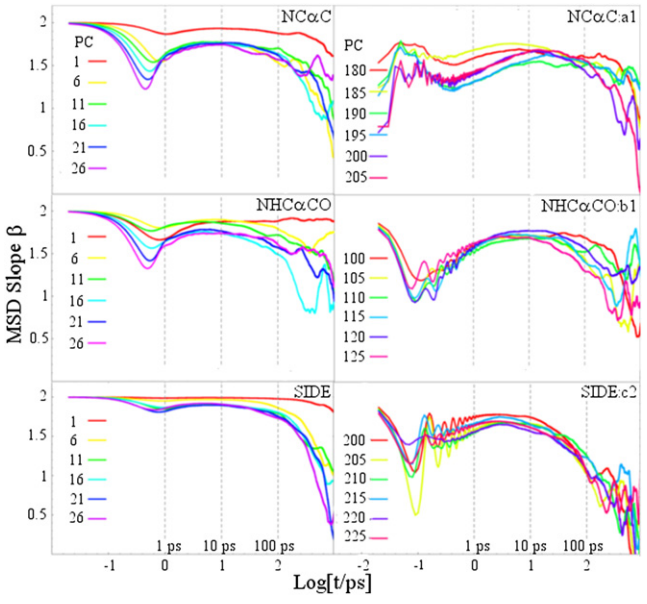

FIGURE 5 Instantaneous slope of the $\log_{10}$(MSD) functions shown in Fig. 4 for the long (*left*) and short (*right*) PCs of the heavy-atom backbone (*top*), main chain (*middle*), and heavy side-chain atoms (*bottom*).

Collective oscillations in the small-covariance regime

To systematically investigate the frequencies of collective motions revealed in the MSD, we computed the Fourier transform of the curves depicted in Fig. 5 for the oscillatory regime below 1 ps. In Fig. 6 we plot the square of the Fourier amplitude for all PCs of our three atomic subsets, representing power in the frequency domain. These results reveal the existence of two dominant collective oscillations in the backbone of gA that can be compared with experimental results from IR and Raman spectroscopy. The first is a broad peak centered at ~5 THz (165 cm^{-1}), spanning PCs 90–120. The second is a sharper peak centered at ~40 THz (1320 cm^{-1}) near PC 250. There is good agreement between the results for the two backbone atomic subsets, with the $NHC_{\alpha}CO$ showing the same dominant features at similar frequencies and PCs as the $NC_{\alpha}C$ set, but with higher-resolution and higher-frequency components in the latter, as expected. The side-chain spectra also show many sharply resolved modes at high PC index, with a pair of dominant modes at 20 THz (660 cm^{-1}) and 40 THz (1330 cm^{-1}), and other distinct modes apparent both above and below these frequencies.

Although it is tempting to attribute these oscillations to covalent bond vibrations, an analysis of the associated eigenvectors reveals that this is not the case in general. In fact, the lowest-frequency oscillations are associated with motions that span many heavy atoms in both the backbone and the side chains, and hence represent collective oscillations across functionally significant portions of our protein. Here we focus on the structure of PC eigenvectors associated with the lowest-frequency backbone oscillations to highlight the possible functional significance of these motions, and the utility of information in the previously ignored small-covariance regime of PCA.

Fig. 7 depicts three sample eigenvectors from the broad 5 THz (165 cm^{-1}) band near PC 100. We illustrate the structure of displacements along each eigenvector by superposition of the NHCαCO backbone projected away

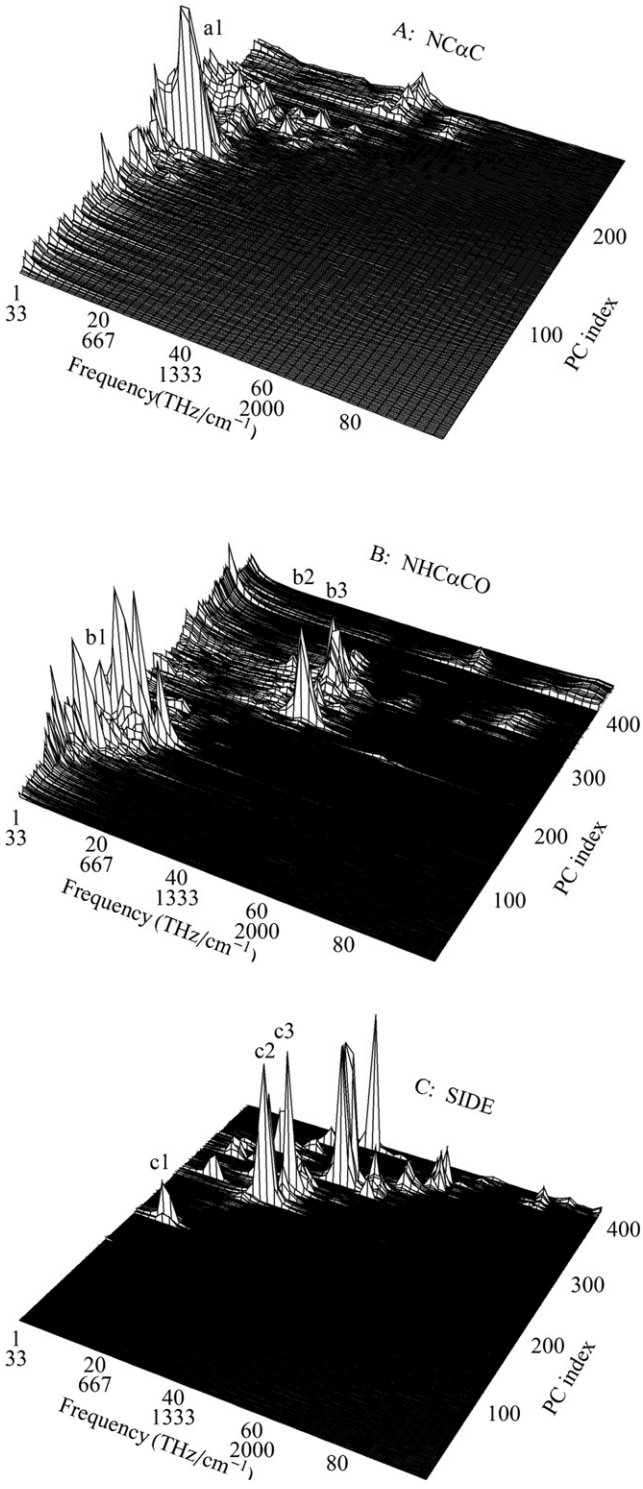

FIGURE 6 Spectral power of the oscillatory regime for β (<1 ps, as shown in Fig. 5).

from the average structure along the positive and negative directions of the PC eigenvector. A careful examination of this figure reveals that in general the displacements are on the scale of a single peptide plane, with tilting of the carbonyl oxygens and amide hydrogens apparent at a number of amino

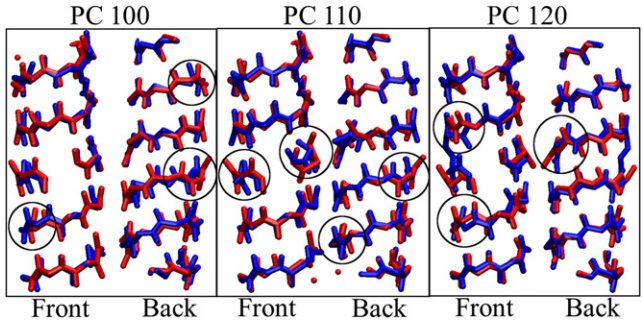

FIGURE 7 Illustration of backbone eigenvectors for subdiffusive PC 100, 110, and 120 of the main-chain $NHC_{\alpha}CO$ atomic subset. The front and back of the helix are shown separately for clarity. The superimposed structures are displaced 5 Å away from the average structure along the appropriate eigenvector, in the positive (*red*) and negative (*blue*) directions. Areas where peptide plane motions result in large displacements of the carbonyl oxygen are highlighted in circles.

acids. This suggests amide plane librations, whose functional significance for cation transport was reviewed in the Introduction. There are ~30 PCs in this group, and an examination of the eigenvectors in time makes it clear that the group as a whole spans tilting motions of each amide plane in the protein (note that there are 30 amide planes in gA). Far-IR and Fourier transform IR spectroscopic measurements of gA without cations have shown that carbonyl librations occupy a band between 75 cm^{-1} and 175 cm^{-1}, and there are other IR-active modes up to 500 cm^{-1} (20,21). This is consistent with the low-frequency features in Fig. 6 *B*, which span the entire far-IR range from ~33 cm^{-1} to 500 cm^{-1}. Moreover, the same experiments measured broad absorption peaks upon addition of Li^+ (20), K^+, Rb^+, and Cs^+ (21) cations to the channel, with the frequencies of cation mobility similar to those of the carbonyl libration band. This shared timescale suggests that the librational modes of the amide planes may be coupled to cation transport through the channel.

We also examined the eigenvectors associated with the higher-frequency backbone mode near 40 THz (1320 cm^{-1}). These are motions within the amide plane associated with stretching of the carbonyl oxygen and amide hydrogen bonds, and are thus clearly visible in the $NHC_{\alpha}CO$ eigenvectors. We conclude that gA has coherent oscillations near 40 THz within the hydrogen bonds that define the secondary structure. Finally, an examination of the side-chain eigenvectors shows that the dominant oscillation modes correspond to bending and torsion of the Trp indole rings (peaks c1 and c2, respectively, in Fig. 6 *C*), which carry a significant dipole moment and form hydrogen bonds with the lipid headgroups in the membrane (46). This suggests that all the MSD oscillations of short PCs are associated with hydrogen bonding, which also explains their subdiffusive distributions as well as their anharmonicity.

The anharmonic character of the Fourier spectra in Fig. 6 is also worthy of comment. The orthogonal decomposition of collective modes in the implementation of atomistic MD

clearly lumps many frequencies of motion together in modes at different length scales. This indicates that collective modes in a protein have complex dynamics with a nonlinear dispersion relation. This finding underlines the need to exercise caution when interpreting the spatial wavelengths from PCA using quasiharmonic approximations that map one wavelength onto one frequency. This is one of the central issues in the interpretation of IR spectra, since this assumption is often used when modes are assigned with the aid of NMA calculations.

CONCLUSION

PCA traditionally has been used in many disciplines to characterize the degrees of freedom that span most of the fluctuations in a system. PCA studies of protein dynamics are no exception, and many have focused on the longest (slowest) PCs, motivated by the need to predict long-timescale dynamics beyond the reach of current simulations (57). In an early and influential study, Amadei et al. (34) defined the ''essential subspace'' as ''a few degrees of freedom in which anharmonic motion occurs that comprises most of the positional fluctuations'' in the system. Here, we have shown that the anharmonic features of the long PCs may be artifacts of insufficient sampling, whereas they are persistent for some shorter PCs. Thus, anharmonicity extends beyond the motions that comprise ''most of the positional fluctuations'', and we suggest that these non-Gaussian-distributed modes are potentially important for the description of function, regardless of their length scale. Although function is difficult to define and quantify, anharmonicity is evidence of coupling among modes, which is likely to be necessary in the complex motions required for function.

A systematic examination of these measures led us to explore features in the short PC regime that have not been examined previously, and we identified collective oscillations with functional implications for gA. A group of backbone oscillations were identified at ~5 THz (165 cm^{-1}) and can be described as peptide plane librations, whose carbonyl oxygens help solvate the lumen and cation in the channel. Our results demonstrate that PCA can be used to isolate interesting covariant motions on a number of different length scales and timescales, in a part of the PCA spectrum that is usually ignored, and highlight the need for an adequate structural and dynamical account of many more PCs than have been conventionally examined in the analysis of protein motion. This analysis is readily applicable to any protein system for which MD simulations are available.

SUPPORTING MATERIAL

Analysis of PCA convergence, description of MSD and anomalous diffusion, and four figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(09)01676-2.

We thank Chris Neale, Nilu Chakrabarti, Boris Steipe, and Ray Kapral for fruitful discussions.

This work was supported in part by operating grant MOP43949 from the Canadian Institute of Health Research, and by a Restracomp scholarship from the Hospital for Sick Children. R.P. is a Canada Research Chair Programme chair holder.

REFERENCES

1. Karplus, M. 1987. Molecular dynamics simulations of proteins. *Phys. Today*. 40:68–70.
2. Karplus, M., and J. Kuriyan. 2005. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. USA*. 102:6679–6685.
3. McCammon, J. A., and S. C. Harvey. 1987. Dynamics of Proteins and Nucleic Acids. Cambridge University Press, New York.
4. Roux, B., and K. Schulten. 2004. Computational studies of membrane channels. *Structure*. 12:1343–1351.
5. Andersen, O. S., and R. E. Koeppe. 1992. Molecular determinants of channel function. *Physiol. Rev*. 72:89S–158S.
6. Ketchem, R. R., W. Hu, and T. A. Cross. 1993. High-resolution conformation of gramicidin A in a lipid bilayer by solid-state NMR. *Science*. 261:1457–1460.
7. Roux, B. 2002. Computational studies of the gramicidin channel. *Acc. Chem. Res*. 35:366–375.
8. Arseniev, A. S., I. L. Barsukov, …, YuA Ovchinnikov. 1985. ^{1}H-NMR study of gramicidin A transmembrane ion channel. Head-to-head right-handed, single-stranded helices. *FEBS Lett*. 186:168–174.
9. Ketchem, R. R., B. Roux, and T. A. Cross. 1997. High-resolution polypeptide structure in a lamellar phase lipid environment from solid state NMR derived orientational constraints. *Structure*. 5:1655–1669.
10. Roux, B., and M. Karplus. 1991. Ion transport in a model gramicidin channel. Structure and thermodynamics. *Biophys. J*. 59:961–981.
11. Pomès, R., and B. Roux. 1996. Structure and dynamics of a proton wire: a theoretical study of H^+ translocation along the single-file water chain in the gramicidin A channel. *Biophys. J*. 71:19–39.
12. Pomès, R., and B. Roux. 2002. Molecular mechanism of H^+ conduction in the single-file water chain of the gramicidin channel. *Biophys. J*. 82:2304–2316.
13. Yu, C. H., and R. Pomès. 2003. Functional dynamics of ion channels: modulation of proton movement by conformational switches. *J. Am. Chem. Soc*. 125:13890–13894.
14. Chiu, S., E. Jakobsson, …, J. A. McCammon. 1991. Time-correlation analysis of simulated water motion in flexible and rigid gramicidin channels. *Biophys. J*. 60:273–285.
15. Urry, D. W., S. Alonso-Romanowski, …, R. D. Harris. 1984. Temperature dependence of single channel currents and the peptide libration mechanism for ion transport through the gramicidin A transmembrane channel. *J. Membr. Biol*. 81:205–217.
16. Roux, B., and M. Karplus. 1988. The normal modes of the gramicidin-A dimer channel. *Biophys. J*. 53:297–309.
17. Tian, F., and T. A. Cross. 1999. Cation transport: an example of structural based selectivity. *J. Mol. Biol*. 285:1993–2003.
18. North, C. L., and T. A. Cross. 1995. Correlations between function and dynamics: time scale coincidence for ion translocation and molecular dynamics in the gramicidin channel backbone. *Biochemistry*. 34:5883–5895.
19. Lazo, N. D., W. Hu, and T. A. Cross. 1995. Low-temperature solid-state ^{15}N NMR characterization of polypeptide backbone librations. *J. Magn. Reson. B*. 107:43–50.
20. Bartl, F., B. Brzezinski, …, G. Zundel. 1998. FT-IR study of the nature of the proton and Li^+ motions in gramicidin A and C. *J. Phys. Chem. B*. 102:5234–5238.
21. Pankiewicz, R., G. Wojciechowski, …, G. Zundel. 2001. FT-IR study of the nature of K^+, Rb^+ and Cs^+ cation motions in gramicidin A. *J. Mol. Struct*. 565:213–217.
22. Berendsen, H. J., and S. Hayward. 2000. Collective protein dynamics in relation to function. *Curr. Opin. Struct. Biol*. 10:165–169.

23. Kitao, A., and N. Go. 1999. Investigating protein dynamics in collective coordinate space. *Curr. Opin. Struct. Biol.* 9:164–169.
24. García, A. 1992. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.* 17:2696–2699.
25. Brooks, B., and M. Karplus. 1983. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. USA.* 80:6571–6575.
26. Hayward, S., A. Kitao, and N. Go. 1994. Harmonic and anharmonic aspects in the dynamics of BPTI: a normal mode analysis and principal component analysis. *Protein Sci.* 3:936–943.
27. Ma, J. 2005. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure.* 13:373–380.
28. Miller, D. W., and D. A. Agard. 1999. Enzyme specificity under dynamic control: a normal mode analysis of α-lytic protease. *J. Mol. Biol.* 286:267–278.
29. Atilgan, A. R., S. R. Durell, …, I. Bahar. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80:505–515.
30. Bahar, I., and A. J. Rader. 2005. Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* 15:586–592.
31. Eyal, E., and I. Bahar. 2008. Toward a molecular understanding of the anisotropic response of proteins to external forces: insights from elastic network models. *Biophys. J.* 94:3424–3435.
32. Tama, F., M. Valle, …, C. L. Brooks, III. 2003. Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. *Proc. Natl. Acad. Sci. USA.* 100:9319–9323.
33. Tirion, M. M. 1996. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 7:1905–1908.
34. Amadei, A., A. B. M. Linssen, and H. J. C. Berendsen. 1993. Essential dynamics of proteins. *Proteins.* 17:412–425.
35. García, A. E., and G. Hummer. 1999. Conformational dynamics of cytochrome c: correlation to hydrogen exchange. *Proteins.* 36:175–191.
36. Hayward, S., and H. J. C. Berendsen. 1998. Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins.* 30:144–154.
37. Maisuradze, G. G., and D. M. Leitner. 2006. Principal component analysis of fast-folding λ-repressor mutants. *Chem. Phys. Lett.* 421:5–10.
38. Mesentean, S., S. Koppole, …, S. Fischer. 2007. The principal motions involved in the coupling mechanism of the recovery stroke of the myosin motor. *J. Mol. Biol.* 367:591–602.
39. Carnevale, V., S. Raugei, …, P. Carloni. 2006. Convergent dynamics in the protease enzymatic superfamily. *J. Am. Chem. Soc.* 128:9766–9772.
40. van Aalten, D. M. F., A. Amadei, …, H. J. C. Berendsen. 1995. The essential dynamics of thermolysin: confirmation of the hinge-bending motion and comparison of simulations in vacuum and water. *Proteins.* 22:45–54.
41. Materese, C. K., C. C. Goldmon, and G. A. Papoian. 2008. Hierarchical organization of eglin c native state dynamics is shaped by competing direct and water-mediated interactions. *Proc. Natl. Acad. Sci. USA.* 105:10659–10664.
42. Maisuradze, G. G., A. Liwo, and H. A. Scheraga. 2009. Principal component analysis for protein folding dynamics. *J. Mol. Biol.* 385:312–329.
43. Brooks, B. R., R. E. Bruccoleri, …, M. Karplus. 1983. CHARMM—a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217.
44. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1983. Comparison of simple potential functions for simulating water. *J. Chem. Phys.* 79:926–935.
45. MacKerell, Jr., A. D., D. Bashford, …, M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.
46. Woolf, T. B., and B. Roux. 1994. Molecular-dynamics of the gramicidin channel in a phospholipid membrane. *Proc. Natl. Acad. Sci. USA.* 91:11631–11635.
47. Ryckaert, J.-P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23:327–341.
48. Zhang, Y., S. Feller, …, R. W. Pastor. 1995. Computer simulation of liquid/liquid interfaces. I. Theory and application to octane/water. *J. Chem. Phys.* 103:10252–10266.
49. Marrink, S. J., and A. E. Mark. 2001. Effect of undulations on surface tension in simulated bilayers. *J. Phys. Chem. B.* 105:6122–6127.
50. Hess, B. 2002. Convergence of sampling in protein simulations. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 65:031910.
51. Faraldo-Gómez, J. D., L. R. Forrest, …, M. S. Sansom. 2004. Conformational sampling and dynamics of membrane proteins from 10-nanosecond computer simulations. *Proteins.* 57:783–791.
52. Grossfield, A., S. E. Feller, and M. C. Pitman. 2007. Convergence of molecular dynamics simulations of membrane proteins. *Proteins.* 67:31–40.
53. Luchko, T., J. T. Huzil, …, J. Tuszynski. 2008. Conformational analysis of the carboxy-terminal tails of human β-tubulin isotypes. *Biophys. J.* 94:1971–1982.
54. Bingham, N. C., N. E. Smith, …, D. D. Busath. 2003. Molecular dynamics simulations of Trp side-chain conformational flexibility in the gramicidin A channel. *Biopolymers.* 71:593–600.
55. Allen, T. W., O. S. Andersen, and B. Roux. 2003. Structure of gramicidin A in a lipid bilayer environment determined using molecular dynamics simulations and solid-state NMR data. *J. Am. Chem. Soc.* 125:9868–9877.
56. Metzler, R., and J. Klafter. 2004. The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics. *J. Phys. A.* 37:R161–R208.
57. Maisuradze, G. G., and D. M. Leitner. 2007. Free energy landscape of a biomolecule in dihedral principal component space: sampling convergence and correspondence between structures and minima. *Proteins.* 67:569–578.