PROTEINS: Structure, Function, and Bioinformatics 65:347–361 (2006)

Pump-Probe Molecular Dynamics as a Tool for Studying Protein Motion and Long Range Coupling

Kim Sharp* and John J. Skinner
Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104

ABSTRACT **A new method for analyzing the dynamics of proteins is developed and tested. The method, pump-probe molecular dynamics, excites selected atoms or residues with a set of oscillating forces, and the transmission of the impulse to other parts of the protein is probed using Fourier transform of the atomic motions. From this analysis, a coupling profile can be determined which quantifies the degree of interaction between pump and probe residues. Various physical properties of the method such as reciprocity and speed of transmission are examined to establish the soundness of the method. The coupling strength can be used to address questions such as the degree of interaction between different residues at the level of dynamics, and identify propagation of influence of one part of the protein on another via "pathways" through the protein. The method is illustrated by analysis of coupling between different secondary structure elements in the allosteric protein calmodulin, and by analysis of pathways of residue–residue interaction in the PDZ domain protein previously elucidated by genomics and mutational studies. Proteins 2006;65: 347–361.** © 2006 Wiley-Liss, Inc.

Key words: molecular dynamics; protein motion; residue coupling; correlation analysis; allostery

INTRODUCTION

Proteins are dynamic objects, and motions of a protein play an important part in their function. Functionally important conformational changes in proteins are typically driven by energies of only a few kT, provided, for example, by the binding of ligands or other proteins. Techniques such as NMR and hydrogen exchange (HX) provide detailed site resolved dynamic information on proteins, revealing the stability, extent, and time scale of motion of individual groups through HX protection factors,[1] the generalized order parameter (S^2), relaxation rates (τ), chemical shift averaging, and other quantities.[2] Molecular dynamics (MD) simulations also provide a detailed description of protein motion and play an important role in the interpretation of experimental probes of protein dynamics. With the routine ability to do all atom simulations on multinanosecond time scales and longer, the amount of information provided by these simulations is enormous. Analyzing the fluctuations in a useful way and relating them to specific experiments is nontrivial.

An important class of methods for studying protein motion is based on frequency analysis. An early example is the now classic method of normal mode (harmonic) analysis,[3] which decomposes the possible motions of a protein around a minimum conformation into harmonic, orthogonal modes. An important insight from this analysis is that the lowest frequency modes represent the softest, most thermodynamically accessible ways a protein could change its conformation (the stiffness being proportional to the square of the frequency). These modes involve long range, concerted motions because they are low frequency, i.e. they have a large effective mass, and hence involve many atoms.[4] A related method is the quasi-harmonic method that uses coordinate fluctuation covariance matrices obtained from MD to model modes of conformational change.[5–8] This allows for a limited amount of anharmonicity. The coordinate covariance matrices may then be analyzed in terms of eigen-vectors and subjected to the same frequency analysis as with normal modes. Principle component and essential dynamics analysis also use effective modes obtained from coordinate fluctuation covariances.[9–11] However, extracting, interpreting, and using the modes obtained by this kind of frequency analysis is not easy.[12–15] Alternatively, coupling between different atoms, residues, or segments may be analyzed directly using the covariance terms.[16] Another approach is to use simplified harmonic models (Elastic network or Tirion type potentials[17,18]) that can be combined with other treatments of large anharmonic motion[19] and sequence/mutation data.[20] Fourier transformation (FT) and filtering of frequencies can be used to simplify and analyze MD trajectories.[21,22] Removal of high frequency motions allows clearer analysis of the putatively more interesting, or at least large scale, low frequency motions.

Grant sponsor: NIH; Grant number: GM48130; Grant sponsor: NSF; Grant number: MCB02-35440.

*Correspondence to: Kim Sharp, Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104. E-mail: sharpk@mail.med.upenn.edu

Received 7 February 2006; Revised 18 April 2006; Accepted 28 June 2006

Published online 24 August 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21146

Other methods take this a step further by actively manipulating selected frequency components of the velocity during MD simulations to probe, or drive conformational changes.[23–26] Dynamics quantities such as amide and methyl NMR order parameters and relaxation rates can be obtained directly from MD simulations, and are most effectively obtained through the frequency domain via fast Fourier transform (FFT).[27–30]

The view of native protein motion as a superposition of oscillatory motions (harmonic or anharmonic) of different frequencies around a minimum energy conformation has provided an attractive model for allosteric (literally "other site") interactions,[7,19,31,32] which gives further impetus to frequency-based methods of analysis. The logic is that allosteric effects require interaction between spatially separated sites. One such mechanism is via collective motion of a large spatial array of atoms, which in turn is characteristic of "low frequency modes" of protein motion. This model (dubbed here the low frequency mode model) naturally invokes analytical methods such as normal mode analysis, essential dynamics, and quasi-harmonic analysis. A different but not necessarily mutually exclusive view of allostery comes from many experiments, the specific residue–residue interaction model. This model does not derive from a low frequency mode view of protein motions, and different ways to analyze MD simulations are required if they are to help interpret these types of experiments.

The specific residue–residue interaction model for allostery emerges from many studies on different proteins with a variety of methods, of which we mention a few pertinent examples. Some classic examples involving oxygen carrying proteins and allosteric enzymes such as glycogen phosphorylase and phosphofructokinase have been reviewed in detail by Perutz.[33] For example, in hemoglobin, oxygen binding to heme iron causes a flattening of the heme-porphyrin plane, which is transmitted to the distal histidine, then via a leucine and isoleucine on the F-helix, to movement of the end of the F-helix, and the CD loop in the cooperative $\alpha 1\beta 2$ and a2b1 interfaces.[34]

Using a statistical mechanical ensemble model of protein fluctuations Friere[35] traced the effect of substrate binding to Lysozyme at helix F through a specific pathway involving residues 24–37 on a neighboring helix, through residues 8–15 on the next helix on to residues in a sheet region on the opposite side of the protein.

In Calmodulin (CaM) residue Y138 has been shown to interact with residues E82, E78, Q79, and D80 of the linker region between helices D and E, which in turn interact with helix A. These specific interactions are required for cooperativity and linkage of Ca^{2+} binding and peptide substrate binding.[36–39]

In the PDZ class of proteins (a peptide binding domain found in signaling proteins) a combination of genomics analysis of sequences, mutations, and binding assays has traced specific reside–residue interactions necessary for allostery, for example from residue H76 through F29 and E57 to A51 on the other side of the protein.[40] This coupling pathway has also recently been detected in dynamic behavior on the ps to ns timescale from changes in NMR-derived backbone and side chain order parameters.[41] A similar sequence/mutation analysis of coupling has been done on the large class of G-coupled protein receptors (GPCR). Significantly, these networks of interactions are quite sparse, i.e. relatively few of the residues mediate allostery, and not all close residues interact in way relevant for allostery.[42] More generally, recent NMR experiments show that mutations cause changes in dynamics that propagate along nonhomogenous, long range, and apparently specific paths.[41,43,44]

Significantly, the specific residue–residue interactions that are mapped out by the experimental and genomic analyses described above are far from obvious by retrospective analysis of these systems in purely structural terms, i.e. in terms of distances between residues. It is hard to explain in purely structural terms why particular residue interactions are important, while others of equal or lesser distance are not. An extra dimension to the interactions must arise from the motions these groups undergo, and the coupling between them, i.e. from protein dynamics. However, in terms of the specific residue–residue interaction model for allostery it is difficult with existing MD techniques to frame and test simple hypotheses about coupling such as the following: Does residue X influence Y, and by how much? Are they more strongly coupled than an arbitrary pair X-Z. How does influence propagate as a function of direction and distance? Can one detect pathways of allosteric action analogous to those found experimentally? Even experiments that probe simpler dynamic properties of proteins than allostery can be difficult to explain. An example is the NMR order parameter, which is a measure of the mobility of a single backbone or side chain group. CaM has anomalous order parameter data such as inverted temperature dependence for some residues (the order parameter increases with T), and an unexpectedly large variation in methionine order parameters.[45,46] These observations imply significant correlation between motions of specific residues, but these correlations are not evident in covariance fluctuation matrices or standard frequency analysis.[16,47] Mayer et al. recently obtained detailed residue–residue correlation matrices in protein G from analysis of the correlation in NMR order parameter changes induced by mutations.[48] A direct comparison with residue–residue correlations obtained by covariance matrix analysis of MD simulations found almost no relationship to the experimental correlations.[49] These and other studies illustrate shortcomings with the available tools for conformational fluctuation analysis in trying to explain experimental data on protein dynamics. This led us to devise a new technique, pump-probe molecular dynamics (PPMD), to address these types of questions and to fill a gap in the simulation analysis toolbox.

METHODS AND MATERIALS

Pump-Probe Molecular Dynamics

The method called here pump-probe molecular dynamics (PPMD) can be applied within any standard MD simu-

lation using existing force field parameters and most simulation conditions. The basic PPMD protocol is as follows:

1. An atom or set of atoms to be pumped is selected prior to the simulation.
2. During the simulation, an oscillating force of a specified magnitude, direction (see below), and frequency ν_0 (period $\tau = 1/\nu_0$) is applied to the pumped atoms(s).
3. Coordinate snapshots are saved throughout the simulation.
4. After the simulation the fluctuation power spectra, or spectral density, $W(\nu)$ of the motions of particular atoms or groups of interest (probe atoms) are obtained via FFT of their Cartesian coordinate trajectories. It should be noted that the power $W(\nu)$ in this context is proportional to the contribution to the mean squared displacement of that atom from motions at frequency ν: summation over the entire power spectrum yields total mean squared displacement of an atom over the simulation period.
5. The pump frequency region of each probe spectrum is examined to see how much the fluctuation power spectrum is increased by the pump, by comparing with the same region of $W(\nu)$ from a control simulation (no pumping force).

The basic PPMD protocol can be applied repeatedly with different pump atoms, force magnitudes, and pump periods as appropriate, to build up a detailed picture of how the pump impulses are transmitted throughout the protein. In the current implementation, pumping of atoms is done in a circular motion around the Z-axis with the same phase and direction of force for all pumped atoms. This application of the force had no particular rationale other than its simplicity, and it was chosen merely for the initial implementation and exploration of the technique. Obviously PPMD is not restricted to this way of applying the force, and other ways of applying it could be selected using some other physical considerations. For example, the axis of the applied circular force could be varied in case the Z-axis direction is atypical for any particular protein/group.

Pumped dynamics is not energy conservative. Elementary considerations show that the power deposited by a particular magnitude of pumping increases as the square of the period τ. Pumping force magnitudes are communicated to the MD program CHARMM[50] in its units (kcal/mole/Å), but since we had no a priori information, suitable magnitudes were determined by experimentation, and typically were 1–3 in these units for the range of frequencies examined here. With the range of pumping forces applied in this work, we have found that the temperature control algorithms in standard MD simulation packages can handle the increase in energy, as judged by a stable value of T during the simulation, and negligible structure distortion.

In practice we found that the increase in power (contribution to rms fluctuation) at the pumped frequency was usually clear enough in the probe atom's power spectrum that comparison of a control power spectrum (no pumping force) of the same atom was not necessary. The increase in fluctuation power spectrum could be detected relative to the baseline power spectrum to either side of ν_0 by a suitable peak finding algorithm, obviating the increase in noise inherent in any spectrum subtraction procedure.

Quantifying the Coupling, or Effectiveness of Transmission of the Pumped Motion

The displacement of an atom in response to a given force will vary depending on how stiff that region of the protein is. Thus, each fluctuation power spectrum is normalized by the sum over $W(\nu)$ (which is just the mean square deviation of the atom) before comparison. After normalization of the power spectra, one can quantify coupling between atoms. If the increase in power of the pumped atom i at frequency ν compared to the control simulation is designated as $\delta W_i(\nu)$, and the corresponding increase in power of a probed atom j is $\delta W_j(\nu)$, then the coupling constant at a particular pump frequency ν_0 can be defined as the ratio of total increase in power (relative contribution to rms fluctuation) of probe to pump atom.

$$C_{ij}(\nu_0) = \int_{\nu_0-\delta}^{\nu_0+\delta} \delta W_j(\nu) d\nu / \delta W_i(\nu_0) \quad (1)$$

C is written in terms of an integration over a frequency range centered on the pump frequency large enough to capture any frequency shifting in transmission to the probe atom due to the nonlinear nature of protein force fields. We allow for the possibility of frequency shifting so as not to miss the coupling, although we have encountered no detectable frequency shifting in applications so far. If a group of atoms are pumped, the denominator contains the sum of $\delta W_i(\nu_0)$ terms over the pumped atoms. Using this coupling metric, one can compute a coupling profile through the sequence of a protein for any pumped frequency. Since C is a dimensionless quantity, and for a given protein/simulation condition only relative values across the sequence convey information, C in plots presented here have been "normalized" and shifted for clarity in graphing. Analysis of residue–residue coupling constants showed that very similar results were obtained if they were computed for the entire backbone of the probe residue (N, C, O, and C_a), its side chain atoms, or just its C_a atom. Results presented here are for coupling constants using just the C_a atom unless otherwise stated.

Comparison of Coupling Profiles

To compare two coupling profiles of a protein, obtained for example before and after a mutation or other perturbation, or with different simulation conditions, two types of analysis were employed. These were developed bearing in mind that only relative values of coupling constant within a single profile have significance, and that most of a typical coupling profile is baseline, with relatively few peaks of interest.

Outlier analysis

One set of coupling constants is treated as the independent variable x_i, the other set as a dependent variable, y_i,

where x_i and y_i are the coupling values of the ith residue say before and after the perturbation. The two sets are subject to linear regression to the equation $y = ax + b$, yielding a minimized root mean squared deviation σ for the fit. The scaled residual $r_i = (y_i - (ax_i + b))/\sigma$ is computed for each residue. A residual of less than -1.5σ or greater than 1.5σ indicates an outlier from the regression, i.e. a significant decrease or increase in coupling, respectively, at that residue due to the perturbation.

Percentile analysis

One set of residue coupling constants (y_i) is plotted against the other (x_i). The resulting scatter plot typically shows the majority of points in a cluster with low coupling constant in both simulations (Fig. 7). This cluster contains the "non-interesting" coupling constants in any profile that are close to baseline, i.e., have coupling that is weak or below the noise level in both simulations. The more interesting coupling constants are the stronger ones corresponding to peaks in the coupling profile. The values that divide the upper 10th percentile from the lower 90th percentile are calculated for both sets of coupling constants, designated $x_{0.1}$ and $y_{0.1}$, respectively. Vertical and horizontal lines are drawn at $x = x_{0.1}$ and $y = y_{0.1}$, respectively. This divides the plot into four regions: lower left, containing residues with low, baseline level coupling in both simulations; upper right, containing residues that have stronger (upper 10th percentile) coupling in both simulations; upper left and lower right regions containing residues that show a coupling peak in one simulation but not the other, i.e. with a significant difference. If for example all the points fell into either the lower left "baseline" region or the upper right "peak" region there is complete agreement between the two coupling profiles at a 10% significance level. Put another way, there is a peak in one simulation if and only if there is a peak in the other. The 10%/90% division was chosen to reflect the typical ratio of residues in peak vs. baseline regions of the coupling profiles analyzed here. The ratio would be adjusted for profiles with less/more peak regions.

Both forms of comparison have the advantage that they are sensitive to the shape of the coupling profile, i.e. the number and location of peaks. They are insensitive to a difference in average coupling between two simulations, or a difference in the range of coupling values between low and high that might occur through differences in scaling, or systematic but nonspecific differences arising from simulation conditions. We note that the linear regression R^2 value alone does not provide a good way either to measure the similarity of profiles or detect differences in peaks: first, the R value is typically always low, dominated as it is by the uncorrelated cluster of baseline points. Second, a low R would occur even if all the peaks occupy in the same positions in the two profiles but they are of different heights.

Frequency and Correlation Analysis

FT, time autocorrelation functions, and time cross-correlation functions for atomic motions were computed from CHARMM format output trajectories using the FORTRAN versions of the routines FFT and CORREL described in Numerical recipes.[51] Time autocorrelation functions for the Brownian harmonic oscillator (BHO) model were computed from standard analytical expressions.[52] The behavior of a BHO is governed by the dimensionless ratio of friction coefficient to harmonic force constant $G = \gamma/2\omega_0$, where γ is the friction coefficient per unit mass, and ω_0 is the harmonic oscillator frequency. Values of $G < 0$, $= 0$, and >0 correspond to under-damped, critically damped, and over-damped conditions, respectively. To match the approximate time scale of relaxation seen in time autocorrelation functions derived from explicit atom simulations of proteins, a constant oscillator frequency of $\omega_0 = 0.05$ radian/ps (oscillation period of 125.7 ps) was chosen with friction coefficients of 0.033, 0.1, or 0.2 ps^{-1} yielding $G = 0.33$ (under-damped), 1 (critically damped), and 2 (over-damped), respectively.

MD Protocol

The PPMD method has been implemented in a Fortran77 subroutine that is called by CHARMM.[50] For the purposes of developing and testing the PPMD methodology, we used the following standard simulation parameters/conditions: CHARMM version 27 force field with all atoms,[53] the Verlet algorithm with a time step of 1 fs, a temperature of 298 K held constant by periodic velocity reassignment, a nonbond cutoff of 14 Å with force shifting, and a total simulation time of at least 10 times the pump frequency period, usually longer than 1 ns. The number of simulation steps and frequency of coordinate saving were adjusted so that exactly 2^n snapshots were generated for FFT. This obviates padding or truncation, and maximizes precision. For most of the simulations we used an approximate but rapid solvent treatment by using the distance dependence dielectric option in CHARMM with a constant of 4. While this model has well-documented shortcomings, it is rapid, and given the large number of simulations required to develop and test the PPMD method, this is an acceptable trade off. The impact of this on our method is assessed with several control simulations as described in results.

For simulations using Langevin dynamics (LD), the Langevin MD command in CHARMM was used. Friction coefficients γ were assigned to all non-hydrogen atoms using

$$\gamma = 6\pi\eta a f / m \qquad (2)$$

where $\eta = 0.89$ cP is the viscosity of water, m is the atomic mass, a is the atomic radius in the CHARMM forcefield, and f is the fraction of exposed surface area of that atom, calculated using the program SURFCV.[54] Thus, buried atoms experience no solvent friction/fluctuation force, while exposed atoms typically have friction coefficients in the range 50–90 ps^{-1}. Hydrogens are assigned a friction coefficient of zero since they are constrained to the heavy atom positions using the SHAKE algorithm.[55]

For simulations using NMR type nuclear Overhauser effect (NOE) type distance restraints, the CHARMM command

NOE was used. A list of restraints was generated for each pair of Cα atoms within 10 Å of each other in the starting structure. For each pair of C_a atoms an NOE type restraint was applied using a harmonic potential with force constant 1 kcal/mole/Å. The force was applied if the distance varied by more than 2.5 Å from the starting structure, other wise no force was applied, i.e., a 5 Å zone of fluctuation with no restraining force was allowed.

Additional PPMD Enhancements

In our initial application of PPMD the pumped frequency does not correspond to any special frequency of natural motion in the protein, (like a prominent normal mode). Thus, it is necessary to scan pumping over as wide a range of frequencies as possible to build up a complete picture of the coupling. To reduce the amount of simulation we investigated pumping at several frequencies simultaneously. Provided the frequencies are well separated, we found that the effects were independent (additive). All results presented here were thus obtained using four simultaneous pumping frequencies: a base frequency with period τ, and three others with periods 3τ, 5τ, and 7τ, respectively.

To explore the sensitivity of the PPMD method simulations of increasing pump magnitude were performed. It was found, not surprisingly, that larger magnitudes eventually resulted in structural distortions in the protein by the end of the simulation. However, by using the NOE-type restraint facility in CHARMM applied to the Cα atoms we could pump with larger magnitudes and so increase sensitivity. NOE-type restraints are convenient for this purpose because they apply no force as long as the distances stay within the upper and lower bounds. These restraints thus can keep a simulation conformation within reasonable bounds while providing minimal bias to the dynamics. PPMD results with and without NOE-type restraints described below show very similar coupling behavior.

Proteins Studied

For the initial development and testing of PPMD, we selected two proteins for which good structures were available, and for which detailed experimental data related to function and allostery was available, CaM and the PDZ domain protein. More specifically, for each protein a variety of experiments and analysis has demonstrated specific residue–residue couplings of functional importance.

CaM (PDB entry 1CDL[56])

CaM is a calcium-regulated protein involved in signal transduction, trafficking, muscle contraction, and many other cellular processes. It is known to recognize over 200 targets,[57] and so it is central in cell regulation and signaling. CaM binds four calcium ions, forming a dumb-bell like structure with two globular domains each containing two ions separated by a long helical domain. Studies of the CaM system include the structural basis for its protein target recognition,[37,58] the thermodynamics of calcium binding,[59] the thermodynamics of smMLCK target peptide peptide binding,[60] and the concerted conformational change upon peptide.[61] Wand et al. have an extensive set of dynamic information for calcium-loaded CaM/smMLCK peptide complex, including NMR order parameters (S^2) and relaxation times (τ) for almost all the amide and methyl groups over the range 15–73°C, and additional order parameter data for the uncomplexed CaM, and for CaM mutants. This work has resulted in an unprecedented amount of site resolved dynamics data on a protein, and interesting dynamic behavior that we have only partly been able to explain with standard MD simulation analysis.[16,47] There is also an extensive series of studies systematically exploring the relationship between fluctuations, conformational changes, binding, and cooperativity.[36–39,62] This experimental work has elucidated specific residue interactions involved in allostery, for example between residue Y138 and helix E cap residues Glu82, and between E78, Q79, and D80 of linker region and residues on helix A. CaM shows two levels of allosteric interaction: between calcium and peptide binding and peptide binding and large binding related changes in conformation.

PDZ domain protein (PDB entry 1BE9[63])

PDZ domain proteins are a family of modular peptide binding domains found in many cytosolic signalling proteins.[64] Extensive sequence data and several high resolution structures are available. Lockless and Ranganathan[40] have used evolutionary sequence analysis and mutation/function analysis to identify specific coupling at the level of residue–residue interactions. These interactions form a network that spans a significant distance in space (i.e. they are allosteric in nature). By examining the spatial sequence of pairs of couplings they can trace pathways of communication. For example residue H76, which is crucial for determining peptide binding specificity, is coupled, through residues F29 and E57 to A51 on the opposite side of the protein. Their combined genomics and experimental analysis also identifies other couplings between some, but by no means all, close residues. The fact that in their analysis not all close residues are coupled forms an important control for the concept of specificity in residue–residue interactions. Recently, Fuentes et al. have shown that very similar coupling pathways are revealed by changes in NMR-derived side chain and backbone order parameters upon peptide binding to PDZ,[41] a direct demonstration of long range propagation of signals.

Starting structures of the single site mutants G33A, G34A, F29K, F44A, S75T, A80V, K84A, and T89S were generated from the wild-type structure (pdb entry 1BE9) using CHARMM by changing the residue template, rebuilding the atoms of the mutated residue and then minimizing the structure using the adaptive basis Newton–Raphson minimizer for 1000 steps. PPMD simulations were then run on the mutants under the same conditions as the wild type.

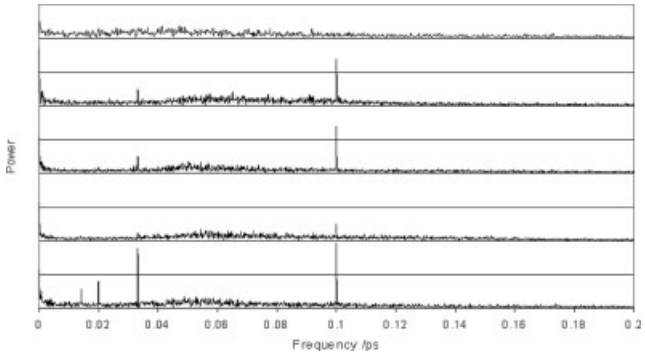

Fig. 1. Fluctuation power spectra of some atomic motions in CaM with 10, 30, 50, and 70 ps period pumping forces applied to helix C, over a total simulation time of 5 ns. Traces are displaced vertically for clarity, and from the bottom up are for the Cα atom in residues Q45 (a pumped residue), T110, L32, and T29. Top trace: Fluctuation power spectrum of Cα atom of Q54 in a control simulation with no pumping force.

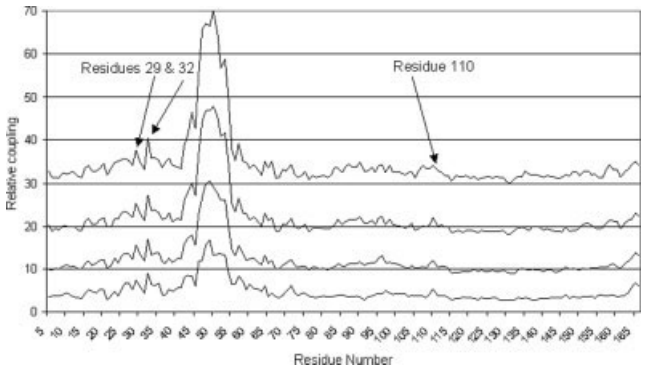

Fig. 2. Coupling profiles for CaM with the C helix pumped. Traces are displaced vertically for clarity. From bottom up, couplings for 10, 30, 50, and 70 ps period pumping. Probe residues whose fluctuation power spectra are shown in Figure 1 are labeled.

RESULTS

PPMD simulations were run on CaM, pumping residues 46–53 in helix C with oscillating forces of 10, 30, 50, and 70 ps period. Figure 1 shows typical fluctuation power spectra obtained from a 5 ns simulation. The upper plot shows a typical power spectrum of an atom from a control simulation with no pumping. The lower plot shows the spectrum of a pumped atom in residue Gln49. Against the rather featureless background four sharp peaks at the pumping frequencies can easily be distinguished as one would expect since this is a pumped atom. The spectra at selected probe atoms in residues T29 and L32 in the neighboring helix and a more distance residue T110, which neighbors residues 29 and 32 but not the pumped helix, show similar features except the spikes have different relative intensities, some being undetectable altogether. This indicates different degrees of transmission to different atoms and at different frequencies.

Figure 2 shows the coupling profile for the simulation in Figure 1, obtained from analyzing the fluctuation power spectra for all the residues, extracting the relative intensities at the pumped frequencies, and applying Eq. (1).

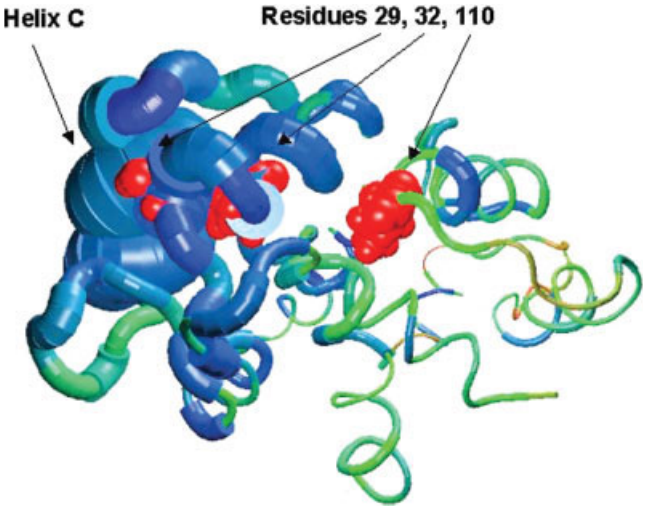

Fig. 3. Coded structural representation of CaM. Worm thickness is proportional to coupling to Helix C at the 10 ps oscillation period. Probe residues whose fluctuation power spectra are shown in Figure 1 are rendered in CPK.

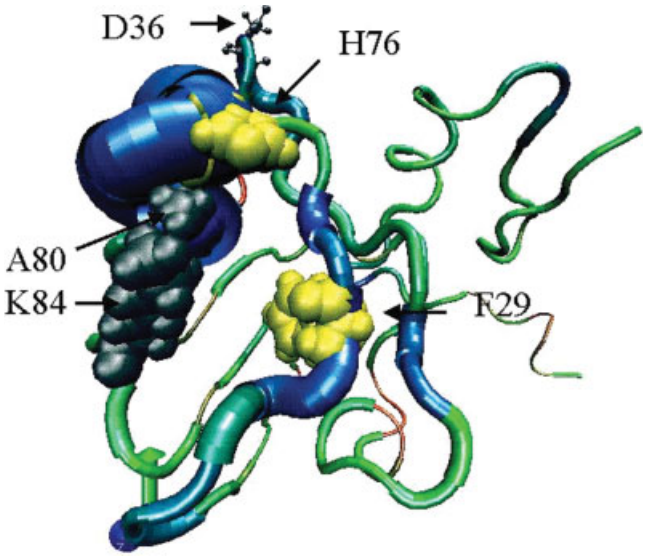

Fig. 4. Coded structural representation of PDZ domain protein. Residue H76 was pumped. Worm thickness is proportional to coupling to H76 at the 10 ps oscillation period. Key coupling pathway residues are shown in CPK.

The coupling profiles exhibit a large peak at the pumped residues, indicating trivial coupling of the pumped residues with themselves. More significantly, there are peaks in coupling at more distant residues in both sequence and space, e.g. at residues 29, 32, and 110, indicating nonhomogeneous spread of the pumped energy and specific couplings.

An alternative way to display the coupling is through a coded structural representation of the protein (Figs. 3 and 4). Here the thickness of the backbone worm indicates the degree of coupling, and enables one to see which parts of the protein are more coupled, and where they are. Figure 3 shows a PPMD simulation on CaM where

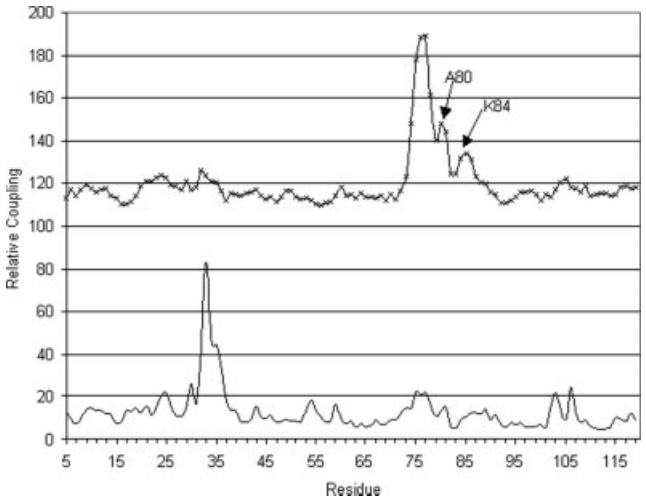

Fig. 5. PDZ domain coupling profiles for pumping at H76 (-x-, trace displaced upward for clarity) and G33 (no symbol). Arrows indicate residues one and two helical turns down from pumped residue H76, as shown in Figure 4. Profiles are profiles averaged over 1, 3, 5, 7, 10, 30, 50, and 70 ps pumping periods.

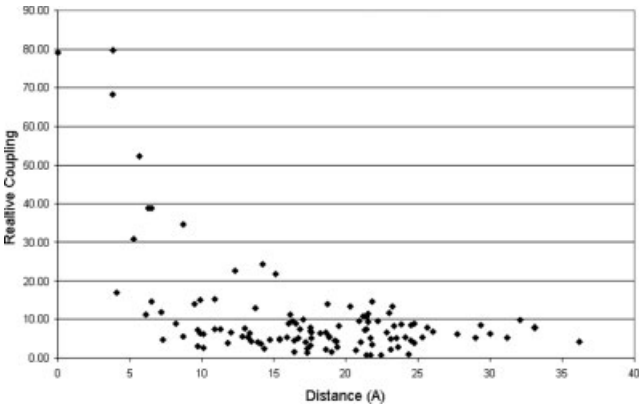

Fig. 6. Plot of coupling strength vs. Cα–Cα distance for PDZ domain pumped at H76. Coupling strengths are taken from the profile averaged over 1, 3, 5, 7, 10, 30, 50, and 70 ps pumping periods (Fig. 5). Distances are measured from the pumped residue H76 Cα.

helix C was pumped at 10, 30, 50, and 70 ps. The coupling profile at 10 ps is coded on the figure. The pumped helix C shows strong coupling to the B helix, which contains residues 29 and 32 and forms one flap closing over the peptide. There is "follow-on" coupling to the helix turn containing residue 110 on the opposite flap. Residues 29, 32, and 110, whose fluctuation power spectra are shown in Figure 1, are rendered in CPK.

Figure 4 shows a similar representation for a PPMD simulation of the PDZ domain protein. In this case residue H76, a key coupled residue identified by Lockless and Raganathan[40] was pumped at 10, 30, 50, and 70 ps. The coupling at 10 ps is shown on the figure. This figure shows that coupling from the pumped residue extends down the helix containing H76 and also across to the β strand containing F29. Interestingly, coupling is much less effective in the other strand direction from H76: round the turn. Heterogeneous coupling to other regions is also apparent. Since the coupling is frequency dependent, average coupling profiles using data from multiple frequencies provide a better picture of the coupling through the protein and increase the sensitivity. To obtain average coupling profiles, profiles from eight different pumping periods, 1, 3, 5, 7, 10, 30, 50, and 70 ps, were averaged. Figure 5 shows two such coupling profiles for the PDZ protein. Each coupling profile represents a set of simulations in which either residue H76 or G33 was pumped. These two residues were shown experimentally to be coupled.[40] Each profile shows a peak in the region of the pumped residue, and other peaks. In the simulations where H76 is pumped, coupling down one side of the helix to the residues one and two helical turns down (A80 and K84) is clearly seen, recapitulating the pathway seen by Lockless and Ranganathan.[40] Significantly, one sees a peak at G33 when H76 is pumped, and a peak at residue H76 when pumping G33, i.e. the coupling is reciprocal. To better illustrate the heterogeneous nature of the coupling, in Figure 6 the residue coupling constants from the H76 pump simulation of PDZ domain illustrated in Figure 5 (upper profile) are plotted against the distance of each residue from the pumped residue H76. The figure illustrates a general decrease in coupling with distance, as one would expect as energy is dissipated across the protein. However, examining the couplings at a given distance in the 5–15 Å region shows a wide range of coupling strengths at each distance, notably even at the very shortest distances that represent neighboring residues.

These coupling profiles illustrate several key features revealed by the PPMD simulations:

i. Heterogenous coupling in space. Not all residues close to the pumped residue(s) are coupled to the same extent (Fig. 6).
ii. Long range coupling. Significant transmission is observed, in some cases from one secondary structure element across another to a third as illustrated in Figures 2 and 3.
iii. Coupling does not necessarily occur via the shortest pathway, as illustrated by the coupling of F29 to H76 via A80 and K84 in PDZ domain protein (Fig. 4).
iv. Reciprocity of coupling between pump and probe residues, as illustrated in Figure 5.

The simulation protocol used for the initial development and testing of the PPMD method was chosen for its speed and to maximize the sensitivity to the pumping signal. However, there are two simulation conditions used here that could potentially affect the coupling profiles. The first is the use of an implicit solvent model. The most serious concern with this model for the PPMD method is the lack of solvent friction and fluctuations. Solvent friction/fluctuation could more rapidly damp the PPMD signal, and so reduce the sensitivity of the method and its ability to detect long range coupling. Also, since solvent does not act uniformly on the protein, but preferentially

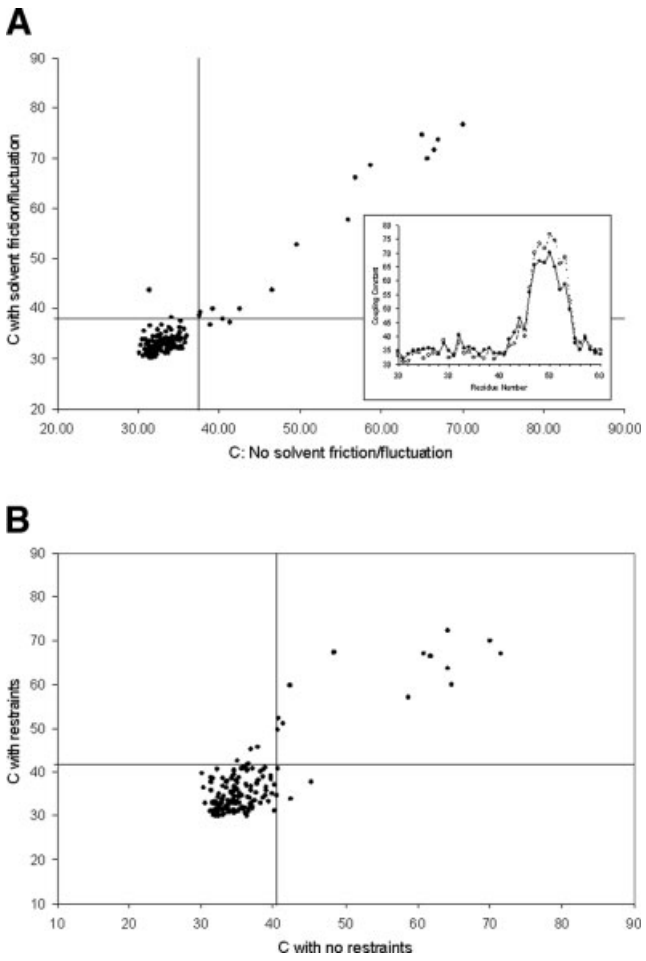

Fig. 7. Percentile analysis of similarity of coupling profiles for CaM pumped at Helix C with a 70 ps period. (**a**) No solvent friction (abscissa) vs. solvent friction (ordinate). Inset shows part of both coupling profiles overlaid to illustrate similarity. (**b**) With NOE restraints vs. no restraints (ordinate).

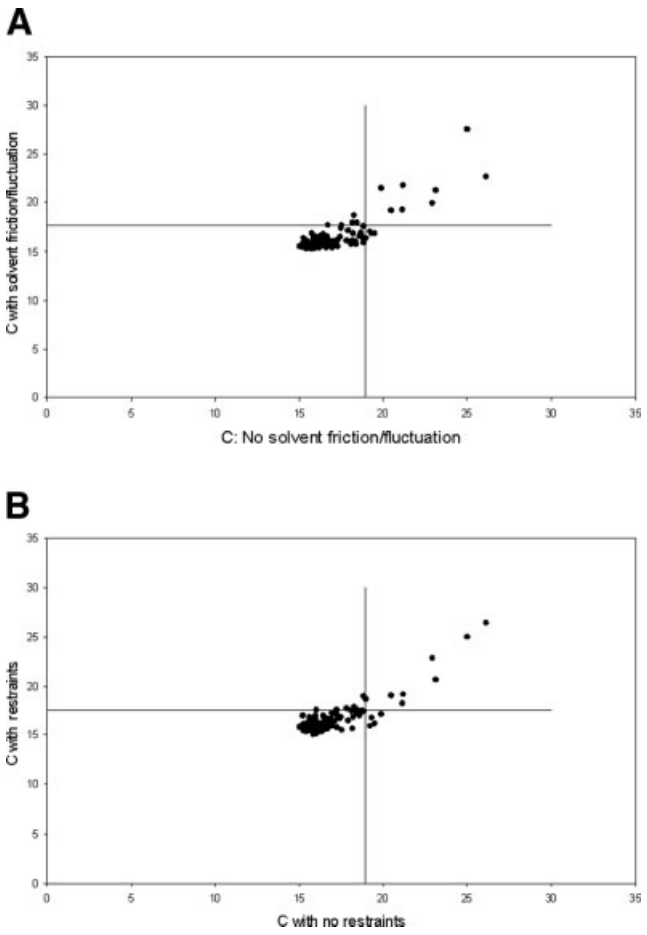

Fig. 8. Percentile analysis of similarity of coupling profiles for PDZ domain protein pumped at residue H76 with a 10 ps period. (**a**) No solvent friction (abscissa) vs. solvent friction (ordinate). (**b**) With NOE restraints vs. no restraints (ordinate).

on surface atoms, it could potentially alter the coupling profiles in a qualitative way. To examine these two possibilities we ran control simulations in which solvent friction/fluctuation on solvent exposed atoms was included via CHARMM's LD facility. Since the solvent friction forces in the LD method would actually be more random than those from an explicit solvent simulation (in which there is a degree of correlation between solvent and protein atom motions due to force reciprocity), the LD control should provide an upper estimate of the effect of neglecting solvent friction/fluctuation. Figure 7(a) shows a comparison of the coupling constants obtained at a pump period of 70 ps applied to the C helix of CaM. The simulation with no solvent friction/fluctuation is the same as that depicted in Figure 2. The simulation with solvent friction was run with a factor of three larger pumping force. The inset to the figure shows an expanded section of the coupling profile between residues 20 and 60, showing that the profiles from the no-friction simulation and the control LD simulation with solvent friction are very similar.

So as to better compare the two coupling profiles we plotted the "with friction" coupling constants against the "no friction" coupling constants for the 70 ps pump period, and used the method of percentile analysis as described in the methods section. The upper 10th percentile lines divide the graph into four regions. Using this analysis we see that Figure 7(a) shows excellent correspondence in coupling profiles with and without solvent friction, with most of the points falling in the lower left baseline regions, or the upper right region indicating a peak found in both simulations. Only one or two residues fall in each of the upper left and lower right regions that indicate a peak in one case but not another. Figure 8(a) shows the same comparison of friction/no friction for a simulation of the PDZ domain, pumping residue H76 at 10 ps. Again, relatively few points fall in either of the upper left or lower right regions that would indicate differences in peaks. Similar levels of agreement are obtained for coupling at the other three pump periods in each simulation. We conclude that solvent friction does not change the coupling profiles qualitatively, although it requires a larger pumping force to

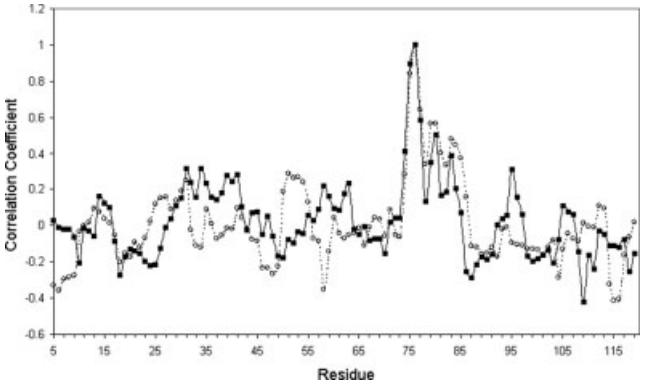

Fig. 9. Fluctuation cross-correlation coefficient $C_{ij} = \langle \Delta x_i \Delta x_j \rangle^2 / \langle \Delta x_i^2 \rangle \langle \Delta x_j^2 \rangle$ between the Cα of H76 and other residue Cα's for PDZ domain protein from a control simulation (no pumping). From batches 1 to 8 (■), and from batches 9 to 16 (○).

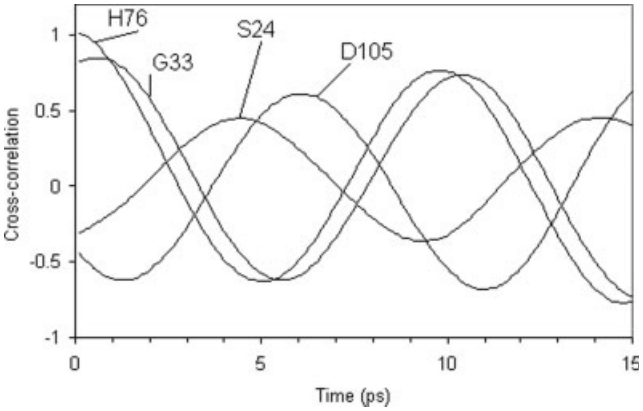

Fig. 10. Time cross-correlation functions for Cα atoms in PDZ domain protein pumped at H76. Distances from the pumped residue are given in Table I.

get the same signal to noise ratio which increases the risk of structure distortion.

The second compromise in the simulation protocol was to use NOE-type restraints on the protein Cα atoms so as to use a larger pumping force to improve signal to noise without distorting too much, or unfolding the protein. While NOE constraints have no force bias when they are satisfied (the majority of the time), they do apply a harmonic restoring force to atoms outside the distance bounds, and this harmonic force may introduce its own coupling behavior and so distort the coupling profile. Figure 7(b) shows a comparison of the coupling profiles obtained at a pump period of 70 ps applied to the C helix of CaM with and without NOE type restraints. Figure 8(b) shows the same comparison for PDZ domain with residue H76 pumped with a 10 ps period. The percentile analysis again shows that there is very good correspondence between the peaks in the two profiles for each protein. The simulations without restraints were done with a factor of 5 lower pumping force to avoid protein shape distortion, i.e. they have a poorer signal to noise ratio which requires a longer simulation. Thus as long as mild NOE type restraints do not qualitatively affect the coupling profile, they have the advantage of better signal to noise and shorter simulation time. There may however be situations in which the type of coupling that is being probed requires rather large excursions of the protein to occur. In this case NOE-type restraints would tend to oppose or mask the coupling, and so they would not be desirable.

In Figure 9 the standard fluctuation cross-correlation coefficients $C_{ij} = \langle \Delta x_i \Delta x_j \rangle^2 / \langle \Delta x_i^2 \rangle \langle \Delta x_j^2 \rangle$ between the Cα of H76 and other residue Cα's are shown for the PDZ domain protein. The plot shows the H76 row extracted from the C_a correlation matrix from a control simulation (no pumping) performed for exactly the same length and conditions as a pumped simulation. The simulation was divided into two halves, and the correlation coefficients calculated for each half to assess the variability. The result is typical of fluctuation correlation analysis, i.e. the residue in question H76, is perfectly correlated with itself, and there is significant coupling to its immediate neighbors. The large majority of correlation coefficients have magnitudes less than 0.5, and moreover most of these vary widely between the first and second halves of the simulation indicating that they are not significant at this length of simulation, at least. The pattern of PPMD couplings, in contrast, persist across different frequencies (Fig. 2) and different simulations (Figs. 7 and 8). Coupling between more distant (in sequence) residues, e.g. H79 and F29, is not apparent from the correlation analysis, nor is any systematic pathway of coupling. Thus, at least for the length of simulation required for PPMD (1–2 ns) this technique has better signal to noise, and is more revealing of coupling pathways than conventional correlation coefficient analysis.

It should be stressed that PPMD is not measuring the mobility of atoms or residues per se i.e. their root mean square (rms) deviation, but the change in mobility due to coupling, or transmission of energy/motions from the pumped residue(s) to the probed residue(s) as measured by the increase in fluctuation power at particular frequencies. Thus, a coupling profile *does not* reflect differences in intrinsic flexibility along the polypeptide chain. The effect of intrinsic differences is removed by the normalization of each atom's fluctuation power spectrum by that atoms rms deviation (which is given by the sum over the power spectrum) prior to the calculation of coupling via Eq. (2). The coupling profile does provide a measure of the interaction between specific atoms or groups at the level of dynamics. To explore the transmission characteristics further, we used time cross-correlation analysis. The cross-correlation function quantifies the correlation in motion between two atoms or groups for different time lags. The same FFT transforms of coordinate trajectories used for the fluctuation power spectra can be used to obtain these time cross-correlation functions in a straightforward manner.[51] From the time axis offset one can obtain the phase delay at different points in the protein. An example is shown in Figure 10, which illustrates the transmission of pumped impulses applied at residue H76 of PDZ domain protein. For the time range shown in the figure, the cross-correlation functions are dominated by the highest frequency pump applied in this simulation,

TABLE I. Phase Delay of a 10 ps Pump Applied to H76 of PDZ

Residue	Lag (ps)	Distance (Å)[a]
H76	0	0
G33	0.7	6
S24	4.5	22
D105	6	16

[a]Distance is from the residue Cα to the H76 Cα.

the 10 ps period pump. The profile for the pumped residue is shown for reference, and has a phase shift of zero. Various phase shifts are observed at other atoms. While the time delay generally increases with distance it is not just dependent on distance, as Table I shows. For example the impulse arrives at residue S24 before residue D105, even though the former residue is farther away. Thus, there is heterogeneous propagation speed through different parts of the protein. These results suggest that variation in distance-normalized transmission speed is another way to quantify coupling strength in the PPMD analysis. This will be examined further in a future study.

The other information obtained from the time cross-correlation analysis is the general rate of transmission of impulses through a protein. The effect propagates at speeds of up to 5 Å/ps, and so it could cross a medium size protein within several picoseconds. It should be noted that this is well within the time scale of the simulations. The upper limit for transmission of any mechanical impulse in a medium is the speed of sound, V_s. While this is not known exactly for proteins, and it also depends on whether the waves are longitudinal or transverse, and how homogenous the mechanical properties are, V_s may be estimated from the relationship:

$$V_s = \sqrt{\frac{Y}{\rho}} \tag{3}$$

where Y is the Young's or bulk modulus and ρ is the density. Since Y is the inverse of the compressibility, it may be estimated from experimental measurements of the latter as $Y \approx 1/(5 \times 10^{-6}/\text{Atm}) = 2 \times 10^{11}$ dynes/cm^3.[65] With a typical protein density of 1.3 g/cm^3, Eq. (3) gives V_s = 14,000 km/h or 39 Å/ps. This estimate lies somewhere between that of water (14 Å/ps) and steel (50 Å/ps), which are accurately estimated by Eq. (3),[66] and it thus seems reasonable. From this we conclude that the transmission rates derived from the data in Table I are similar enough to V_s, but far enough below the estimated upper limit to be reasonable.

It should be noted that the time cross-correlation functions are normalized by the mean squared motions of the two atoms, and so Figure 10 only shows the time lag in transmission and not the decay in intensity with distance. The latter is more accurately depicted by the coupling profiles in Figures 2 and 5. To better understand how transmission of an impulse such as that applied by PPMD might be dissipated in a dense, fluctuating medium such as a protein, we compared the general dynamic behavior

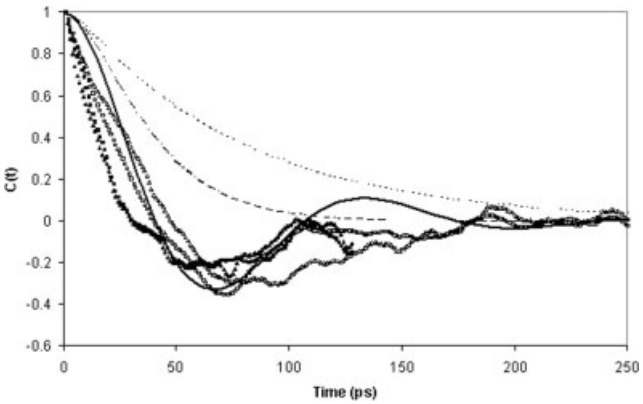

Fig. 11. Time autocorrelation functions for atomic motions of Cα of Asp58 (■, □) and Glu 67 (△, ▲) in CaM simulated with implicit solvent (empty symbols) or explicit solvent (filled symbols). Brownian harmonic oscillator that is under-damped (—), over-damped (. . .), or critically damped (- -).

of the protein to the BHO model. The BHO model is an analytically treatable stochastic model of a system undergoing fluctuation and frictional dissipation around some minimum energy state, with a harmonic restoring force.[52] The BHO model is effectively a one-dimensional analogue of a stable protein undergoing fluctuations around the minimum energy "native state" structure. The time autocorrelation functions for atomic motions in non-PPMD simulations of CaM with implicit and explicit solvent were calculated from the MD trajectories again using FFTs. Characteristic profiles are shown in Figure 11. These are compared to the BHO model using friction and harmonic force constant parameters chosen to match the protein dynamics time-scale. The BHO model has three regimes: where friction forces dominate and there is no oscillatory behavior (over-damped); where the restoring force dominates and the system shows oscillatory behavior (under-damped), and an intermediate regime (critically damped); and where the system has the fastest relaxation and which also shows no oscillatory behavior. As can be seen from Figure 11, the protein behaves as an under-damped system with a characteristic time scale in the 10–100 ps range. Interestingly, the protein motion has strong oscillatory characteristics even with explicit solvent, even though one might expect the frictional effect of the solvent to make the system behave as an over/critically damped one. The under-damped, oscillatory character of protein dynamics is consistent with the detectable transmission of pumped oscillatory impulses over the same length scale as the protein revealed by the PPMD simulations.

A key experimental tool to detect residue–residue coupling is site directed mutagenesis. For example a series of PDZ mutants examined by Lockless and Raganathan[40] revealed very similar coupling to that obtained from their multiple sequence analysis. We examined eight of their mutants using PPMD simulations. Three of these mutants, F44A, S75T, and T89S, involved residues that were determined by sequence analysis and mutation not to be coupled to H76. The other five mutants, G33A, G34A, F29K, A80V,

TABLE II. Summary of Changes in Coupling Due to Mutations in PDZ Domain Protein

	Mutant[b]							
Peak position[a]	*F44A*	*S75T*	*T89S*	**G33A**	**G34A**	**F29K**	**A80V**	**K84A**
24	X[c]	X			X			X
36				X	X	X	X	X
46	X	X					X	
49	X			X	X		X	
54				X				X
75		X	X		X	X	X	X
81	X		X		X		X	
85	X		X				X	X

[a]Mutants in bold involve residues coupled to H76, those in italics are not coupled to H76.
[b]1BE9 residue numbering.
[c]X indicates a peak at that residue that changed due to that mutation (increased or decreased) by $>1.5\sigma$ as determined by the outlier analysis.

and K84A, involved residues that were found experimentally to be coupled to H76. PPMD simulations were performed on each of the mutants by pumping at 1, 3, 5, 7, and 10 ps at either H76 or F29, and the average coupling profiles were computed. Each mutant coupling profile was compared to the corresponding wild-type coupling profile (pumped at either H76 or F29), and significant changes in coupling peaks caused by mutations were determined by the outlier analysis. The results are summarized in Table II. The most notable result is for residue 36, where significant changes in coupling peak are seen in all the mutations involving "on-pathway" residues, but for none of the mutations involving off pathway residues.

DISCUSSION

We present here the development and initial implementation of a method for analyzing the dynamics of proteins using MD simulations, the PPMD method. The method is straightforward to implement and very flexible. It was designed with the goal of helping to interpret experiments that probe residue–residue coupling, particularly in the context of modern NMR measurements of site resolved protein dynamics, and more generally in phenomena that involve long range interactions in proteins, such as allosteric effects. The method is not, however, restricted to these applications. The ability to perturb a protein in a defined way at a defined place and analyze the resulting change in dynamics should be applicable to the understanding of proteins in general as complex dynamical objects. The idea of manipulating MD trajectories in a controlled way is not new, and various methods have been developed.[22–24,67,68] However, the PPMD method described here has significant differences from these earlier methods. Temperature pulse methods[67,69,70] manipulate the total kinetic energy. In effect they operate over the entire frequency regime simultaneously, and usually are applied over the entire protein. In contrast the PPMD method selectively applies a force at a few frequencies at a time (although these can be scanned over the entire spectrum), and to selected atoms of residues. Steered or targeted MD simulations use a force along some "pathway" applied to some part of the system.[71] Here the goal is different from PPMD: it is to drive the system through some specific transition or sequence of events in the MD time scale, events that presumably happen over much longer time scales naturally. Temperature pulse methods are also used with a similar goal, for example to drive protein unfolding on the MD time scale.[69,70]

A method that more closely relates to PPMD is one that manipulates atomic velocity components in a frequency specific way during MD simulation.[23,24] PPMD differs from the methods described in these papers in several key aspects. First, these methods were applied to the atomic velocities. Since the causal sequence in an MD algorithm is Force → Acceleration → Velocity → Position, intervening at the velocity stage will cause atomic velocities and positions to get out of phase. This limits the extent and magnitude of the pulses one can apply.[24] Special precautions such as pulsed applications of the perturbation and quenching can be used to get around this,[24] but this complicates the method, and it can only be applied to a single frequency at a time. In contrast, PPMD avoids this by intervening at the force step. The PPMD method is straightforward, and can be applied at multiple frequencies simultaneously. This is an advantage if a wide range of frequencies is to be explored. Second, the goal of the velocity frequency manipulation method was somewhat different: it was applied principally to drive low frequency motions and torsional motions so as to cause some transition.[23,24] Previously Bowman et al. developed a method called driven molecular dynamics (DMD), which also applies harmonic forces to a protein during an MD simulation.[25,26] In DMD a centro-symmetric force of frequency ν is applied between each pair of atoms, whose strength is proportional to the distance between the atoms. The goal of applying this "non-specific dilating" force uniformly to all the atoms is to excite normal mode vibrations, which are detected by the amount of energy absorbed at each frequency. The goal of DMD is to provide an alternative to the standard normal mode analysis using Hessian matrix diagonalization.[25,26]

The PPMD method thus differs from previous velocity and force driven MD methods in two respects. First, in

the way the frequency-based perturbation is applied—as a force—but only to specific residues. Second, in what the method is designed for: the PPMD method is not designed to drive the protein through some conformational change, low frequency or otherwise, or to excite normal modes, but to examine residue–residue couplings and coupling "pathways."

Another method that was recently applied to the PDZ domain protein so as to detect differential coupling between different parts of the protein is the anisotropic thermal diffusion (ATD) method of Ota and Agard.[68] In this innovative method, the MD simulation is quenched periodically by reducing the simulation temperature to 10 K, then the system is rethermalized with just one residue coupled to a heat bath. The transmission of kinetic energy from this residue to other residues as a function of time is used to extract differences in thermal conductivity. This cycle can be repeated and the results combined to improve the precision. Differences in the rate of thermal diffusion are then interpreted as different couplings between residues. Ota and Agard found that residues H76 and F29 of PDZ domain were coupled, in agreement with the results of Lockless and Raganathan and our analysis. This coupling occurred through the residue I31 that, although not coupled to either H76 or F29 in Lockless and Raganathan's experimental analysis,[40] lies directly between H76 and F29, i.e. the coupling occurs by the shortest pathway. In contrast, in the PPMD analysis presented here, the coupling between H76 and F29 occurs along a helix, through residues A80 and K84, which are coupled to H76 in Lockless and Raganathan's experimental analysis, and then across to F29 on a neighboring strand. This is a more indirect path in space. While the aim of the ATD and PPMD methods applied to the PDZ domain protein is the same, they probe different modes of energy transmission under somewhat different conditions, which may account for these differences in the coupling pathway. In the ATD method it is the transmission of kinetic or heat energy, which comes from all types of motion, whether they can be characterized by a frequency or not. In the PPMD method it is transmission of "acoustic" or mechanical motions at specific frequencies. The other difference is that in the ATD method the protein is quenched and rethermalized, i.e. the method is intermittent and nonequilibrium, while the PPMD method obtains couplings under continuous, equilibrium conditions. The two methods may, therefore, reveal different aspects of coupling, and they could provide complementary information. Further comparison of the two methods on more proteins would be valuable. In this regard, Leitner et al. have done a detailed examination of the thermal conductivity of proteins using normal mode methods,[72] although this study was not undertaken from the perspective of examining specific residue–residue couplings.

In the initial applications of PPMD described here we have shown that nontrivial residue–residue couplings can be detected by the method, and that they reveal long range, nontrivial pathways of interaction. In the PDZ domain for example, these coupling pathways recapitulate those seen experimentally. We emphasize although that the genomics analysis, mutation experiments, and simulations are each examining different properties of the protein, the sequence analysis examines the tolerance of different sites to coupled changes in amino acid residues while preserving biological function. The mutation experiments are probing the contribution of different sidechains to peptide binding energetics. The pump-probe simulations measure the transmission of motional energy between residues. In addition, we are using one wild-type sequence alone, while the statistical analysis extracts couplings from many different sequences. Thus, one would not expect exact correspondence between the three lines of analysis. Indeed it is encouraging that the mutation/binding, sequence analysis, and PPMD agree to the extent they do. However, there are differences in detail. In particular, one expects PPMD to reveal more couplings than those seen in experiment, since the simulations provide atomic level detail on every residue, whereas not all couplings would be important for the correct sequence-to-structure relationship or for binding energetics. An obvious way to more closely mimic the PDZ mutation experiments would be to do pump-probe simulations on PDZ-peptide complexes and analyze binding related changes.

An important consideration for the usefulness of PPMD is distance over which one can detect couplings. The results in Figures 2 and 3 show that one can pick up an interaction between residues on two secondary structure units separated by a third (a helix). Nevertheless, it would be naïve to expect that one simply pump a residue on one side of a protein and always see an effect on the other side. To map out long pathways one might have to "walk" the pump-probe across the protein in shorter steps. This actually parallels how sequences of specific residue–residue interactions that form pathways have been elucidated in experiments on PDZ domain, GPCR proteins, and CaM.[37–40,42] In spite of limitations in the distance over which one can detect couplings, it is important to point out that the PPMD method detects specific residue–residue couplings that we simply cannot detect in standard residue–residue or atom–atom fluctuation covariance matrix analysis. Such couplings, if present in these matrices, are either swamped by the large amount of noise and statistically spurious correlation coefficients likely in the large number ($N \times N$) of terms, or they are not manifest at a simple pair-wise correlation level. In contrast our analysis effectively reduces the $N \times N$ dimensionality in the covariance matrix analysis to a series of N-dimensional analyses, i.e. between each pumped residue and the $N - 1$ remaining residues, and it also increases signal to noise by pumping, rather than trying to use ambient fluctuations. Of course there is an increased computational cost, as a full series of couplings would require additional simulations pumping at the other $N - 1$ residues. However, specific applications are unlikely to require pumping at every residue. Typically only key biologically important residues would need to be pumped to gain useful information.

We examined several of the mutants of PDZ domain protein using PPMD that were studied by Lockless and

Raganathan in elucidating the coupling between H76 and F29, namely three "off-pathway" mutations F44A, S75T, and T89S, and five "on-pathway" mutants, G33A, G34A, F29K, A80V, and K84A. The goal of this analysis was to use changes in coupling profile to identify residues that might be involved in the H76-F29 coupling pathway that are not apparent from the sequence analysis/mutation experiments. The results in Table II show, for example, that at residue 36 significant changes in the coupling are seen in all the mutations involving on-pathway residues, but for none of the mutations involving off-pathway residues. Interestingly, residue D36 sits on the loop directly above residue H76 (as seen at the top of Fig. 4). We tentatively identify this residue as having a secondary role in the pathway. We speculate that it may play a role in "directing" the coupling down the helix from H76 through A80 and K84, rather than in the other strand direction from H76. No other peak shows such a clear distinction between on- and off-pathway mutations, although the peak at residue 49 is prominently perturbed by just one off-pathway mutations, but by the majority of on-pathway mutations. The peak at position 46 is effectively the converse, being perturbed in two of three off-pathway mutations, while being affected by just one on-pathway mutation. This presumably indicates a residue that is quite decoupled from the pathway. It is significant that one should see some such decoupled residues, since if every residue in the protein were involved in a coupling it would make little sense to talk about a coupling pathway at all. Our mutant analysis is far from exhaustive, and clearly one would not expect a clean binary distinction between "on" and "off" pathway cases for every position/mutation, given the complexity of the dynamics of PDZ and the various experimental and simulation uncertainties and limitations. However, this analysis does indicate how one can use PPMD to provide a more mechanistic view of experimentally determined couplings, and how one might use this analysis of MD simulations to examine residue and atomic level details of coupling inaccessible to experiment.

Characterization of CaM dynamics by the time autocorrelation functions revealed that it behaves as an underdamped system, i.e. it has a significant oscillatory characteristic. Other proteins we have examined show the same general behavior. This is significant for the application of the PPMD method. It means that oscillations, whether intrinsic or applied externally as here, will propagate significantly through the protein, as seen here. These oscillations are thus in principle capable of mediating long range coupling, of being detected and analyzed. If a protein acted as an over-damped system, one would expect much less transmission of the PPMD pulse, making the method less sensitive. Indeed, the consequences of a protein being an over-damped dynamic system would be problematic for any frequency-based analysis such as mode analysis or frequency filtering. If protein dynamics had no oscillatory character, it would make little sense to talk about modes or frequencies at all. The oscillatory character of protein dynamics, while clear in the time correlation analysis, is not obvious from the atomic fluctuation power spectra. In a non-PPMD simulation the fluctuation power spectrum is rather featureless with a low, broad peak extending from about 0.02–0.08 ps^{-1} (Fig. 1, top trace). Examination of such profiles for different atoms, or for the same atom from different simulation batches or conditions shows that these profiles are generally similar. No clearly identifiable modes both standout from their neighbors and are persistent across a group of atoms or between different simulation batches. Put another way, it is difficult from examination of the fluctuation power spectrum alone to determine whether protein dynamics is occurring in either the under-damped or over-damped regime. Moreover, the density of vibrational states determined both experimentally, and in harmonic and quasi-harmonic analyses of proteins is also dense enough to approximate a continuum.[73,74] The difficulty in identifying specific and persistent low frequency modes from principal component analysis[12–15] is partly due to how closely spaced in frequency they are, and how variable many modes are. So selection of particular native modes from harmonic analysis so as to analyze coupling or putative allosteric conformational changes is difficult and somewhat subjective. As demonstrated here, the PPMD provides another approach to this problem, where one specifically probes the coupling between residues at a wide range of frequencies. The couplings obtained from PPMD show physically reasonable properties, such as reciprocity, decay with distance, and transmission times consistent with known physical behavior of proteins. The couplings are also robust with respect to the presence of solvent friction and mild restraints on the protein. This is useful since it allows one to use larger pumping forces and get larger signal to noise ratios in the PPMD simulations.

Future Directions

The PPMD method is very flexible in terms of how and where the pumping force is applied. We chose only quite simple pumping schemes at a restricted set of frequencies so as to demonstrate the usefulness of the method. We pumped at periods from 1 to 70 ps. The full range, conservatively, would be from about 5 times the MD time step (5 fs) to about one-fifth of the total simulation length (anywhere from 1 ns upwards with current computers). In addition to the pumping frequency, the selection of pumped atoms, magnitude, type, and phase of force can be varied widely. These features of PPMD can be exploited to maximize control and sensitivity. As PPMD is a new technique, these options need to be explored in a systematic fashion so as to get the most out of it. Some specific considerations include:

How finely to scan the pump frequency: scanning requires multiple simulations, and so one wants to be effective but frugal. Pumping at four frequencies yields a proportional speed up in this regard. The practical upper limit to the number of simultaneous pumping frequencies is not known.

The largest magnitude pumping force that can be applied without causing too much heating or structural distortion so as to maximize sensitivity.

How best to apply the force: options include applying it to a single atom or a collection of atoms, as a planar or circular oscillation, as torsional force around a bond, etc. To increase sensitivity one can also pump more atoms, such as parts of/whole secondary structure units, reflecting the tendency of longer time-scale motions to involve more atoms. The pumping of a portion of Helix C of CaM illustrates this.

PPMD analysis has a close correspondence to NMR experiments that measure propagation of changes in order parameters (S^2) of residues surrounding a site of single residue mutations.[41,43,44] Both provide a measure of residue–residue coupling through dynamics on the same spatial and time scales. The experiments also reveal inhomogeneous transmission of effects. For example in CaM the M119L mutant produces large changes that are confined to the region of the mutation, whereas D58N produces an apparent pathway of changes in methyl order parameters up to 15–20 Å away.[44] While order parameters can be obtained directly from MD simulations and compared to NMR data,[28] both experimental and calculated quantities are single residue values: there is no specific information about correlation between sites. Here the explicit coupling information obtained from PPMD provides an important extra dimension. Analysis of such NMR experiments is a natural future application for PPMD.

More generally, we envision PPMD as a flexible method for analyzing protein dynamics with different, but complementary properties to existing analysis techniques, not as a replacement. It is in fact possible to use PPMD in combination with existing techniques. For example, normal mode and quasi-harmonic analysis provide eigen-values and eigen-vectors that describe the frequency and identity of effective modes. In PPMD it would be possible to pump certain combinations of atoms with the right magnitudes and directions to mimic a mode, thus combining elements of both approaches. Either alone or in combination with other techniques, the PPMD method can help address specific questions about protein dynamics such as how much one residue is coupled to another, how a perturbation at one site affects another site, and whether there is a pathway of coupling between the two sites.

ACKNOWLEDGMENTS

We thank Tobin Sosnick, Walter Englander, and Ryan Coleman for discussions and helpful comments.

REFERENCES

1. Englander SW, Kallenbach NR. Hydrogen exchange and structural dynamics of proteins and nucleic acids. Q Rev Biophys 1983;16:521–655.
2. Palmer AG, Kroenke CD, Loria JP. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. Methods Enzymol 2001;339:204–238.
3. Go N, Scheraga H. Dynamics of a small globular protein in terms of low frequency vibrational modes. J Chem Phys 1969;5:4751–4767.
4. Case DA. Normal mode analysis of protein dynamics. Curr Opin Struct Biol 1994;4:285–290.
5. Levy R, Srinivasen A, Olson W, McCammon JA. Quasi-harmonic method for studying very low frequency modes in proteins. Biopolymers 1984;23:1099–1112.
6. Levy R, Karplus M, Kushnick J, Perahia D. Evaluation of the configurational entropy for proteins: application to molecular dynamics simulations of an α helix. Macromolecules 1984;17: 1370–1374.
7. McCammon JA, Gelin B, Karplus M, Wolynes P. The hinge bending mode in lysozyme. Nature 1976;262:325,326.
8. Karplus M, Kushnick J. Method for estimating the configurational entropy of macromolecules. Macromolecules 1981;14:325–332.
9. Ichiye IT, Karplus M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. Proteins 1991;11:205–217.
10. Hayward S, Kitao A, Go N. Harmonicity and anharmonicity in protein dynamics: a normal mode analysis and principal component analysis. Proteins Struct Funct Genet 1995;23:177–186.
11. Amadei A, Linssen AB, Berendsen HJ. Essential dynamics of proteins. Proteins 1993;17:412–425.
12. Hünenberger PH, Mark AE, van Gunsteren WF. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. J Mol Biol 1995;252:492–503.
13. VanAalten D, DeGroot B, Findlay J, Berendsen H, Amadei A. A comparison of techniques for calculating essential dynamics. J Comp Chem 1997;18:169–181.
14. Abseher R, Nilges M. Are there non-trivial dynamic cross-correlations in proteins? J Mol Biol 1998;279:911–920.
15. Balsera MA, Wriggers W, Oono Y, Schulten K. Principal component analysis and long time protein dynamics. J Phys Chem 1996;100: 2567–2572.
16. Prabhu N, Lee A, Wand JW, Sharp KA. Dynamics and entropy of a calmodulin-peptide complex studied by NMR and molecular dynamics. Biochemistry 2002;42:562–570.
17. Tirion M. Large amplitude elastic motions in proteins from a single parameter, atomic analysis. Phys Rev Lett 1996;77:1905–1908.
18. Kim MK, Chirikjian GS, Jernigan RL. Elastic models of conformational transitions in macromolecules. J Mol Graph Model 2002;21:151–160.
19. Miyashita O, Onuchic JN, Wolynes PG. Nonlinear elasticity, protein quakes, and the energy landscapes of functional transitions in proteins. Proc Natl Acad Sci USA 2003;100:12570–12575.
20. Zheng W, Brooks B, Donaich S, Thirumalai D. Network of dynamically important residues in the open/closed transistion in polymerases is strongly conserved. Structure 2005;13:565–577.
21. Session R, Dauber-Osguthorpe P, Osguthorpe D. Filtering molecular dynamics trajectories to reveal low-frequency collective motions: phospholipase A2. J Mol Biol 1986;209:617–633.
22. Levitt M. Real-time interactive frequency filtering of molecular dynamics simulations. J Mol Biol 1991;220:1–4.
23. Dauber-Osguthorpe P, Maunder C, Osguthorpe D. Molecular dynamics: deciphering the data. J Comput Aided Des 1996;10:177–185.
24. Phillips S, Essex J, Edge C. Digitally filtered molecular dynamics: the frequency specific control of molecular dynamics simulations. J Phys Chem 2000;112:2586–2597.
25. Bowman JM, Zhang X, Brown A. Normal-mode analysis without the Hessian: a driven molecular-dynamics approach. J Chem Phys 2003;119:646–650.
26. Kaledin M, Brown A, Kaledin AL, Bowman JM. Normal mode analysis using the driven molecular dynamics method. II. An application to biological macromolecules. J Chem Phys 2004;121:5646–5653.
27. Akke M, Bruschweiler R, Palmer A. NMR order parameters and free energy. J Am Chem Soc 1993;115:9832,9833.
28. Chatfield D, Szabo A, Brooks BR. Molecular dynamics of *S. nuclease*: comparison of simulation with 15N and 13O relaxation data. J Am Chem Soc 1998;120:5301–5311.
29. Wong KB, Daggett V. Barstar has a highly dynamic hydrophobic core: evidence from molecular dynamics simulations and nuclear magnetic resonance relaxation data. Biochemistry 1998;11:11182–11192.

30. Zabell APR, Post CB. Docking multiple conformations of a flexible ligand into a protein binding site using NMR restraints. Proteins Struct Funct Genet 2002;46:295–307.
31. Ma J, Karplus M. The allosteric mechanism of the chaperonin GroEL: a dynamic analysis. Proc Natl Acad Sci USA 1998;95: 8502–8507.
32. Cooper A, Dryden DTF. Allostery without conformational change. A plausible model. Eur Biophys J 1984;11:103–109.
33. Perutz M. Mechanisms of cooperativity and allosteric regulation in proteins. Q Rev Biophys 1989;22:139–237.
34. Perutz MF, Fermi G, Luisi B, Shaanan B, Liddington RC. Stereochemistry of cooperative mechanisms in hemoglobin. Acc Chem Res 1987;20:309–321.
35. Freire E. The propagation of binding interactions to remote sites in proteins. Proc Natl Acad Sci USA 1999;96:10118–10122.
36. Sun H, Yin D, Coffeen LA, Shea MA, Squier TC. Mutation of Tyr138 disrupts the structural coupling between the opposing domains in vertebrate calmodulin. Biochemistry 2001;40:9605–9617.
37. Jaren OR, Kranz JK, Sorensen BR, Wand AJ, Shea MA. Calcium-induced conformational switching of Paramecium calmodulin provides evidence for domain coupling. Biochemistry 2002;41:14158–14166.
38. Sorensen BR, Faga LA, Hultman R, Shea MA. An interdomain linker increases the thermostability and decreases the calcium affinity of the calmodulin N-domain. Biochemistry 2002;41:15–20.
39. Faga LA, Sorensen BR, VanScyoc WS, Shea MA. Basic interdomain boundary residues in calmodulin decrease calcium affinity of sites I and II by stabilizing helix–helix interactions. Proteins Struct Funct Genet 2003;50:381–391.
40. Lockless S, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. Science 1999;286: 295–299.
41. Fuentes E, Channing J, Lee AL. Ligand dependent dynamics and intramolecular signaling in a PDZ domain. J Mol Biol 2004; 335:1105–1115.
42. Suel G, Lockless S, Wall M, Ranganathan R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat Struct Biol 2003;10:59–69.
43. Clarkson M, Lee AL. Long range dynamic effects of point mutations propagate through side chains in the serine protease inhibitor Eglin c. Biochemistry 2004;43:12448–12458.
44. Igumenova T, Lee AL, Wand AJ. Backbone and side chain dynamics of mutant calmodulin-peptide complexes. Biochemistry 2005;44:12627–12639.
45. Lee AL, Kinnear SA, Wand AJ. Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex. Nat Struct Biol 2000;7:72–77.
46. Lee AL, Wand AJ. Microscopic origins of entropy, heat capacity and the glass transition in proteins. Nature 2001;411:501–504.
47. Lee AL, Sharp KA, Kranz JK, Song X, Wand AJ. Temperature dependence of the internal dynamics of a calmodulin-peptide complex. Biochemistry 2002;41:13814–13825.
48. Mayer K, Earley M, Gupta S, Pichumani K, Regan L, Stone M. Covariation of backbone motion throughout a small protein domain. Nat Struct Biol 2003;10:962–965.
49. Lange O, Grubmuller H, deGroot B. Molecular dynamics simulations protein G challenge NMR-derived correlated backbone motions. Angew Chem Int Ed 2005;44:3394–3399.
50. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization and dynamics calculations. J Comp Chem 1983;4: 187.
51. Press W, Flannery B, Teukolsky S, Vetterling W. Numerical recipes. Cambridge: Cambridge University Press; 1986.
52. van Kampen NG. Stochastic processes in physics and chemistry. Amsterdam: Elsevier; 1992.
53. MacKerell AD, Brooks B, Brooks CL, Nilsson L, Roux B, Won Y, Karplus M. CHARMM: the energy function and its parameterization with an overview of the program. In: Schleyer PVR, editor. The encyclopedia of computational chemistry, Vol. 1. Chichester: Wiley; 1998. pp 271–277.
54. Sridharan S, Nicholls A, Sharp KA. A rapid method for calculating derivatives of solvent accessible surface areas of molecules. J Comp Chem 1995;16:1038–1044.
55. Ryckaert JP, Ciccotti G, Berendsen HJC. Integration of the Cartesian equation of motions of a system with constraints: molecular dynamics of n-alkanes. Numerical J Comp Chem 1977;23: 327–341.
56. Meador W, Means A, Quiocho F. Target enzyme recognition by calmodulin: 2.4 angstroms structure of a calmodulin-peptide. Science 1992;257:1251.
57. Yap KL, Kim J, Truong K, Sherman M, Yuan T, Ikura M. Calmodulin target database. J Struct Funct Genomics 2001;1:8–14.
58. Babu YS, Bugg CE, Cook WJ. Structure of calmodulin refined at 2.2 A resolution. J Mol Biol 1988;204:191–204.
59. Pedigo S, Shea MA. Discontinuous equilibrium titrations of cooperative calcium binding to calmodulin monitored by 1-D 1H-nuclear magnetic resonance spectroscopy. Biochemistry 1995;34:10676–10689.
60. Wintrode P, Privalov P. Energetics of target peptide recognition by calmodulin: a calorimetric study. J Mol Biol 1997;266:1050–1062.
61. Kranz JK, Flynn PF, Fuentes EJ, Wand AJ. Dissection of the pathway of molecular recognition by calmodulin. Biochemistry 2002;41:2599–2608.
62. Vanscyoc WS, Shea MA. Phenylalanine fluorescence studies of calcium binding to N-domain fragments of Paramecium calmodulin mutants show increased calcium affinity correlates with increased disorder. Protein Sci 2001;10:1758–1768.
63. Doyle DA, Lee A, Lewis J, Kim E, Sheng M, MacKinnon R. Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. Cell 1996;85:1067–1076.
64. Songyang Z, Fanning AS, Fu C, Xu J, Marfatia SM, Chishti AH, Crompton A, Chan AC, Anderson JM, Cantley LC. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. Science 1997;275:73–77.
65. Dadarlat VM, Post CB. Adhesive-cohesive model for protein compressibility: an alternative perspective on stability. Proc Natl Acad Sci USA 2003;100:14778–14783.
66. Lide DR, editor. CRC handbook of chemistry and physics, 71st ed. Boca Raton: CRC; 1990.
67. Becker O, Karplus M. Temperature echoes in molecular dynamics simulations of proteins. Phys Rev Lett 1993;70:3514–3517.
68. Ota N, Agard D. Intramolecular signaling pathways revealed by anisotropic thermal diffusion. J Mol Biol 2005;351:345–354.
69. Paci E, Karplus M. Unfolding proteins by external forces and temperature: the importance of topology and energetics. Proc Natl Acad Sci USA 2000;97:6521–6526.
70. Snow CD, Qiu LD, Deguo, Gai F, Hagen SJ, Pande VS. Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy. Proc Natl Acad Sci USA 2004;101:4077–4082.
71. Lu H, Isralewitz B, Krammer A, Vogel V, Schulten K. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. Biophys J 1998;75:662–671.
72. Yu X, Leitner DM. Heat flow in proteins: computation of thermal transport coefficients. J Chem Phys 2005;122:1–11.
73. Cusack S, Smith J, Finney J, Tidor B, Karplus M. Inelastic neutron scattering analysis of picosecond internal protein dynamics. Comparison of harmonic theory with experiment. J Mol Biol 1988;202:903–908.
74. Balog E, Becker T, Oettl M, Lechner R, Daniel R, Finney J, Smith JC. Direct determination of vibrational density of states change on ligand binding to a protein. Phys Rev Lett 2004;93:1–4.