Journal of Molecular Structure 972 (2010) 41–50

Contents lists available at ScienceDirect

Journal of Molecular Structure

journal homepage: www.elsevier.com/locate/molstruc

Shape-dependent global deformation modes of large protein structures

Gennady V. Miloshevsky [a,*], Ahmed Hassanein [a], Peter C. Jordan [b,*]

[a] School of Nuclear Engineering, Purdue University, West Lafayette, IN 47907, USA
[b] Department of Chemistry, MS-015 Brandeis University, P.O. Box 549110, Waltham, MA 02454-9110, USA

ARTICLE INFO

Article history:
Received 23 November 2009
Received in revised form 3 January 2010
Accepted 5 January 2010
Available online 11 January 2010

Keywords:
Minimization
Normal modes
Conformational changes
Gating transitions
Protein shape
Mode following

ABSTRACT

Conformational changes are central to the functioning of pore-forming proteins that open and close their molecular gates in response to external stimuli such as pH, ionic strength, membrane voltage or ligand binding. Normal mode analysis (NMA) is used to identify and characterize the slowest motions in the gA, KcsA, ClC-ec1, LacY and $LeuT_{Aa}$ proteins at the onset of gating. Global deformation modes of the essentially cylindrical gA, KcsA, LacY and $LeuT_{Aa}$ biomolecules are reminiscent of global twisting, transverse and longitudinal motions in a homogeneous elastic rod. The ClC-ec1 protein executes a splaying motion in the plane perpendicular to the lipid bilayer. These global collective deformations are determined by protein shape. New methods, all-atom Monte Carlo Normal Mode Following and its simplification using a rotation–translation of protein blocks (RTB), are described and applied to gain insight into the nature of gating transitions in gA and KcsA. These studies demonstrate the severe limitations of standard NMA in characterizing the structural rearrangements associated with gating transitions. Comparison of all-atom and RTB transition pathways in gA clearly illustrates the impact of the rigid protein block approximation and the need to include all degrees of freedom and their relaxation in computational studies of protein gating. The effects of atomic level structure, pH, hydrogen bonding and charged residues on the large-scale conformational changes associated with gating transitions are discussed.

1. Introduction

X-ray crystallography [1] yields atomically resolved protein structures in near-native conformations; these can be explored to illuminate protein structure–function relationships. Although such structures, captured in a variety of conformations, provide useful architectural insights, they remain only static images of dynamic systems. To understand protein function requires characterizing dynamic behavior during conformational changes (gating) that permit signaling (ion channels) or substrate transport (transporters). Experimentally derived B-factors [2,3] identify highly mobile regions, but these are often flexible polypeptide loops and protein blocks not necessarily involved in gating. In addition, B-factors are not vectorial, and cannot provide insight into gating movements. Fluorescent resonance energy transfer and nuclear magnetic resonance methods [4,5] can provide partial information with respect to the directionality of conformational changes, but full dynamical detail is unresolved. Molecular modeling approaches such as molecular dynamics (MD) simulations [6] and normal mode analysis (NMA) [7] are also used to study protein conformational dynamics.

Reliable computational sampling of gating transitions is a long-standing challenge [8]. There are two distinct difficulties. First, transitions are many orders of magnitude slower than atomic vibrations. Second, nature tends to be perverse, and crystallography usually provides conformational data for only one of many functionally important states. Thus, not only is standard MD, limited to nsec simulations, an inadequate analytical tool but it is also generally necessary to "bootstrap" migration on a potential energy surface for which the reliable experimental data only characterizes a single functionally important configuration. Numerous approaches have been devised to address these issues. If two conformational endpoints are known, a new technique [9,10], carrying out repeated trajectory analysis at successive points along a transition pathway is a promising MD extension of line search methods [7] for migrating on complex potential surfaces. However, the more typical situation is a single endpoint structure established crystallographically. In steered MD, a probable direction for gating is arbitrarily designated; however, the imposed high-speed perturbations have recently been shown to introduce intrinsic bias in the simulations [11]. All-atom NMA [12,13], coarse-grained Gaussian Elastic Network models [13,14] and Principal Component Analysis [15,16] are useful techniques for identifying large-scale protein motions. However, they only identify the initial steps in exiting a

* Corresponding authors. Tel.: +1 765 494 8618; fax: +1 765 496 2233 (G.V. Miloshevsky); tel.: +1 781 736 2540; fax: +1 781 736 2516 (P.C. Jordan).
E-mail addresses: gennady@purdue.edu (G.V. Miloshevsky), jordan@brandeis.edu (P.C. Jordan).

0022-2860/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.molstruc.2010.01.009

near-native protein conformation, and thus only probe the onset of a gating transition. In these applications a target structure is usually generated from an initial structure via a single-step atomic displacement along one (or several) low-frequency modes until a preset root-mean-square displacement (RMSD), typically 2.0–3.5 Å, between the two structures is attained. However, local dynamics, representative of a single near-native configuration, is generally inadequate for describing large-scale conformational rearrangements. In such processes the low frequency, highly collective protein normal modes (NMs) may well undergo dramatic changes in character, precisely features that cannot be captured in analyzing local dynamics. To account for such changes and establish gating pathways and transition-state structures, requires following or tracking the low-frequency NMs using an appropriate procedure [17,18].

All NMA approaches are vacuum simulations that ignore the influence of the surroundings. The underlying hypothesis, well established by long experience, is that large-scale conformational rearrangements are governed by the shape and backbone connectivity of the biomolecule itself, and the intrinsic directions of the cooperative displacements are effectively encoded in the native fold [19,20]. Placing a biomolecule in a realistic water/lipid environment results in overdamping the intrinsically allowed cooperative displacements encoded in the protein architecture [21]. Structural motion becomes slow and diffusive, not vibrational as in vacuum, meaning that the NM eigenvalues have no physical significance in terms of both the timescales and the amplitudes of harmonic oscillations as derived from eigenvalues. However, the NM eigenvectors are physically meaningful both in vacuum and in biophysically relevant milieus indicating the tendency of the protein structure to undergo a conformational change in a particular direction upon internal or external perturbations [20,21]. A biomolecular skeleton, alpha carbons connected by springs, placed in arbitrary surroundings is a purely harmonic system that follows a harmonic trajectory in the space of collective NM coordinates. However, this harmonic trajectory is altered by anharmonic contributions due to the presence of the side chains in a real protein and their diffusive, overdamped motion in a solvent [21]. Small conformational changes of the side chains and the surrounding's frictional effects are actually imposed on the large-scale motions of the backbone. Thus, the net effect of environmental influences on the structural dynamics of proteins evolving along gating pathways is to enforce an effectively harmonic trajectory [20,21]. The surroundings also may be significant if they can perturb the protein structure in ways that affect either shape or backbone connectivity.

Here we review our all-atom NMA results on gating initiation in five proteins (gA [22], KcsA [23], ClC-ec1 [24,25], LacY [26] and $LeuT_{Aa}$ [27]) and contrast these studies with analyses of gA and KcsA using our new methods, all-atom Monte Carlo Normal Mode Following (MC-NMF) [17] and its simplification, RTB MC-NMF [18], based on a rotation-translation of blocks (RTB) approximation [28]. Normal coordinates provide the collective degrees of freedom. Perturbing these proteins along the low-frequency NMs yields the largest collective structural changes at the smallest energetic cost. We show that cylindrically-shaped gA, KcsA, LacY and $LeuT_{Aa}$ biomolecules exhibit deformational modes reminiscent of global twisting, transverse and longitudinal motions in a homogeneous elastic rod. ClC-ec1's gross shape is totally different. Viewed perpendicular to the membrane plane it is rhomboid [29], but from within the membrane it looks somewhat like a butterfly. Its global deformations are also different. The subunits swing relative to one another, leading to a symmetric splay; its cytosolic and periplasmic regions alternate in electrolyte accessibility. We confirm anew the established fact that protein shape governs the character of the low-frequency NMs [19]. What conventional NMA-based methods cannot do is establish possible gating pathways because they do not describe changes in the collective coordinates during gating deformation. Our new methods are designed for just this purpose, and were applied to elucidate the large-scale conformational changes that take place in gA [17] and KcsA [18] along their gating pathways. In MC-NMF along their transition pathways, the continuous change of protein shape dramatically alters the global character of the low-frequency NMs. We discuss the effects of atomic level structure, hydrogen bonding and charged residues on large-scale conformational changes associated with gating.

2. Computational methods

2.1. Energy minimization and NMA

Minimization and harmonic analysis [7] are integral to our strategy for conformational searching of the relevant crystallographic structures. We treat these, incorporating explicit hydrogens, in full atomic detail. Structurally significant waters, ions and substrate molecules are retained, and the all-hydrogen CHARMM22 topology and parameter set [30] describes molecular systems. To remove atomic overlaps and relax the protein, we minimize using steepest descent with a random step length [7], which usually takes 500–2000 steps. The initially assembled molecular system is subject to a number of separate minimizations in Cartesian coordinates, each generating a different final configuration reflecting the random step length used in steepest descent [7]. All minimized structures conserve the native protein's fold, but each exhibits different side chain conformations. The lowest energy structure is selected and further minimized using a new nonlinear conjugate gradient method with guaranteed descent [31], one which is globally convergent. This approach is required to locate a geometry with unprecedentedly low strain energy ($<5\times10^{-10}$ kcal mol^{-1} Å^{-1}) for a large system (14,809 atoms in ClC-ec1). Other minimization methods such as standard Conjugate Gradient, Quasi-Newton, TNPACK and L-BFGS-B [7] do not converge. In order to carry out NMA on large proteins, the maximum derivative must be less than 10^{-8} kcal mol^{-1} Å^{-1} because low-frequency eigenvectors crucially reflect the quality of the minimized structure [7]. Properly minimized, the first six NMs, corresponding to translations and rotations of the whole molecule, must exhibit vibrational frequencies of zero. Imperfectly minimized, first-order terms in the Taylor expansion of the potential energy couple global molecular rotation with internal motions; these perturb the internal eigenvalues and eigenvectors, leading to apparent non-zero eigenvalues for the first six NMs. For maximum derivatives $>10^{-5}$ kcal mol^{-1} Å^{-1}, imaginary vibrational frequencies arise, indicating the structure is not properly minimized. Only in the fashion outlined are physically viable minima generated, with six zero and $3N-6$ positive eigenvalues [7].

NMA [7,12] approximates the full molecular potential by a harmonic function around a minimum energy structure. We approximate the CHARMM potential energy [30] E_c as a harmonic function E_h of the atomic mass-weighted Cartesian coordinates $\mathbf{x}$ around a stable conformation $\mathbf{x_0}$ using a Taylor expansion

$$E_h(\mathbf{x}) \approx E_c(\mathbf{x_0}) + \mathbf{g}^T(\mathbf{x}-\mathbf{x_0}) + \frac{1}{2}(\mathbf{x}-\mathbf{x_0})^T\mathbf{H}(\mathbf{x}-\mathbf{x_0}) \quad (1)$$

where $\mathbf{g}$ is the mass-weighted gradient vector, $\mathbf{x}-\mathbf{x_0}$ is the displacement vector and $\mathbf{H}$ is the mass-weighted Hessian matrix. The superscript T denotes the transpose of vectors $\mathbf{g}$ and $\mathbf{x}-\mathbf{x_0}$. For the bond and non-bonded energy terms the first and second derivatives of the potential energy E_c with respect to the mass-weighted Cartesian coordinates are calculated analytically. For other energy terms (angle, dihedral, improper and Urey–Bradley terms) a