Computational Biology and Chemistry 47 (2013) 89–95

Contents lists available at ScienceDirect

Computational Biology and Chemistry

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

Research article

Mapping the intramolecular signal propagation pathways in protein using Bayesian change point analysis of atomic motions

Mikhail A. Iakhiaev[a], Alexei V. Iakhiaev[b,*]

[a] Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA
[b] Texas College, 2404 N. Grand Avenue, Tyler, TX 75702, USA

ARTICLE INFO

Article history:
Received 25 February 2013
Received in revised form 27 May 2013
Accepted 9 August 2013

Keywords:
Bayesian change point analysis
Protein
Molecular dynamics
Intramolecular signal propagation
Wavelet coherence
Stochastic resonance

ABSTRACT

We propose to use change points of atomic positions in the molecular dynamics trajectory as indicators of the propagating signals in protein. We designate these changes as signals because they can propagate within the molecule in the form of "perturbation wave", transmit energy or information between different parts of protein, and serve as allosteric signals. We found that change points can distinguish between thermal fluctuations of atoms (noise) and signals in a protein despite the differences in the motility of amino acid residues. Clustering of the spatially close residues that were experiencing change points close in time, allowed us to map pathways of signal propagation in a protein at the atomic level of resolution. We propose a potential mechanism for the origin of the signal and its propagation that relies on the autonomic coherence resonance in atomic fluctuations. According to this mechanism, random synchronization of fluctuations of neighboring atoms results in a resonance, which increases amplitude of vibration of these atoms. This increase can be transmitted to the atoms colliding with the resonant atoms, leading to the propagating signal. The wavelet-based coherence analysis of the inter-atomic distances between carbon-alpha atoms and surrounding atoms for the residue pairs that belong to the same communication pathway allowed us to find time periods with temporarily locked phases, confirming the occurrence of conditions for resonance. Analysis of the mapped pathways demonstrated that they form a network that connects different regions of the protein.

1. Introduction

Protein dynamics is essential for its functional activity and includes several types of motions: atomic vibrations, residue motions (including bond stretching and bond angle bending), motions of secondary structure elements, and correlated motions of multiple residues, including motions of large domains (Yonetani and Laberge, 2008; Teilum et al., 2009). These motions carry the signals related to the intramolecular communication, example of which is allosteric regulation of protein function. In allosteric mechanism, ligand binding at an allosteric site is coupled to a structural and/or dynamic change at a distant regulated site of the protein. This implies that there exist links between distant regions of protein which can be defined in terms of long-range intramolecular communication. Historically, two basic models for understanding long-range intra-protein communication where successfully used to explain cooperative nature of the changes in hemoglobin tetramer upon oxygen binding. (a) The sequential or Koshland-Némethy-Filmer (KNF) model (Koshland et al., 1966) and (b) the concerted model or Monod–Wyman–Changeux (MWC) model (Changeux, 2012). Modern versions of theory explaining allosteric mechanisms include several common principles: a protein conformation is not static; rather, it constantly populates a range of distinct conformations according to the Boltzmann distribution, ligands bind to a small subset of the conformations present in the ensemble, stabilizing them, and as a result, the entire ensemble eventually becomes enriched in the ligand-bound, stabilized conformations. All globular proteins could be allosteric and could contain multiple preexisting pathways for long-range intramolecular communication (Tsai et al., 2009). Although allosteric mechanisms are investigated in great detail, mechanistic understanding of the long-range intramolecular communication is still incomplete. Some of the remaining problems include: mapping the communication pathways (identifying atomic chains that transmit the signal), and how propagating signal is sustained and can reach distant sites.

There are several examples of successful mapping of the allosteric pathways in proteins based on their structure, evolutionary conservation, modeling the protein as a network and considering the residue connectivity, and inter-atom force distribution analysis (Chen et al., 2007; Chennubhotla et al., 2008; Stacklies et al., 2009; Lu et al., 2010; Selvaratnam et al., 2011;

* Corresponding author. Tel.: +1 903 593 8311.
E-mail address: aiakhiaev@texascollege.edu (A.V. Iakhiaev).

1476-9271/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiolchem.2013.08.004

Weinkam et al., 2012; Suel et al., 2003). Although insightful, the above mentioned approaches can reveal only a small fraction of existing pathways. For example, the evolutionary approach is based on the analysis of many proteins of the same family and can reveal conserved pathways common for the whole family of proteins, but not unique pathways specific to individual proteins. The approaches that infer links from correlated motions do not even try to map communication pathways: they only take for granted the link between two distant residues, if these residues are involved in coordinated motion. Despite the extensive structural insight, X-ray crystallography provides snapshots of different states of the protein reaction cycle, but not of the transitions between them. There is no established method for mapping the signal transition paths, and also there is no accurate terminology to describe intramolecular communication in proteins. For example, definition of conformation as a specific "three-dimensional arrangement of its constituent atoms" (Price, 2000) can lead to confusion, when applied to protein dynamics, because it allows different interpretation of what to consider a conformation. The object of our study – Activated Protein C (APC) is a serine protease involved in the coagulation (blood clotting) cascade formed by multiple macromolecules. Binding of APC to its receptor or other macromolecules results in dramatic changes in its substrate specificity and its functional activity. Although some of these changes can be attributed to the steric hindrances created by the macromolecules bound to the APC, other changes are clearly allosteric (Yang et al., 2007; Rezaie, 2003). For example, the activity of APC is regulated by the binding of Ca^{2+} and Na^{+} to the sites distant from the active site regions without inducing structural changes detectable by the X-ray crystallography. In this context, computational approaches have the potential to complement the experimental techniques and provide the information that cannot be obtained experimentally. Therefore, in this study we used molecular dynamics (MD) simulations and looked for methods which would allow us to find indicators of the intramolecular communication in the MD trajectory. Because we are dealing with hardly detectable structural changes, protein dynamics, and intramolecular communication which may include energy transmission, here we define our use of terminology, in order to prevent possible confusion. We use a definition of protein conformation that is based on energetic, rather than geometric, considerations. The distinct ensemble of accessible conformational states of proteins corresponds to local minima on the energy surface (with the native conformation corresponding to the global minimum). The native conformation (i.e., protein states corresponding to the global minimum on the energy surface) exhibits a certain degree of heterogeneity, because structural dynamics is an essential feature of any protein system. These energy sublevels are referred in literature as conformations, and sometimes as conformers (Molina et al., 2008). The conformations we are referring to in this study correspond to these energy sublevels and the APC is not expected to undergo strong structural (conformational) changes during our experiments. Other terms that require clarification are collective (or correlated) motions of the protein residues, normal modes, and stochastic resonance as they relate to the change points. The collective motions in the protein are described by its normal modes (NM), which are derived from the elastic network model of the protein and assume the resonance conditions. Protein residues that belong to the same NM move in coordinated fashion, for example, cross the equilibrium position simultaneously (Chennubhotla et al., 2008). These are global motions, as opposed to the local changes in protein dynamics determined with the changepoint analysis. Thus, change points reflect local changes and can be thought of as random (stochastic) phenomena in contrast to collective motions, which reflect the behavior of the whole system. It should be noted that normal modes and related standing waves do not transport energy, although there is a possibility of energy transmission between different modes in certain conditions. Also, low frequency NM correlate with thermal fluctuations, B-factor (residue motility), and direction of the conformational change of the protein. Intramolecular energy transport in proteins is an important phenomenon and there were many studies devoted to modeling of this process. One mechanism that attracted much attention was exciton-soliton mechanism of transmission of the energy released by ATP hydrolysis (Davydov, 1977). This theory explains long-distance energy propagation by formation of solitons characterized by slow dissipation of energy enabling its transmission for long distances. According to the theory of "discrete breathers", energy in a protein can jump from site to site with high yields, covering in many instances remarkably large distances. Such energy transfers mark the spontaneous formation of a localized mode of nonlinear origin at the destination site, which acts as an efficient energy-accumulating center (Piazza and Sanejouand, 2009). The energy accumulation capability of some residues carries some similarity to the properties of the residues undergoing change point in our description of perturbation transmission in proteins.

The first objective of this study is to investigate whether the Bayesian change point analysis of atomic displacements can be used to discriminate between thermal fluctuations and conformational changes (hardly detectable structural changes involved in intramolecular communication). The second objective is to test whether detected change points can allow mapping the signal (perturbation) propagation pathways. Additional goal is to propose a possible explanation of the origin and the mechanism of propagation of these signals.

2. Materials and methods

2.1. The activated protein C (APC) structure used in the study

The three-dimensional structure of the Activated Protein C (APC) was downloaded from Protein Data Bank (PDB ID: 1AUT.pdb). The APC consist of protease domain (heavy chain) and a light chain made of two epidermal growth factor like domains (EGF1 and EGF2) and Gla-domain. In this study, we used the APC fragment containing 293 amino acid residues designated as APCw. The APCw consists of protease and EGF2 domains covalently linked by the disulfide bond. Native APC protease domain contains Na^{+} and Ca^{2+} cations (Rezaie, 2010; Yang et al., 2004) not resolved in the 1AUT.pdb structure. These cations were added to the structure of APCw by homology modeling using as templates the trypsin and factor Xa structures with PDB IDs 1A31.pdb and 2BOK.pdb, respectively.

2.2. Molecular dynamics (MD) experiments

MD simulations of APCw were carried out with periodic boundary conditions and full PME electrostatics using NAMD2 (Phillips et al., 2005) with the all-atom CHARMM27 force field (Foloppe and MacKerell, 2000). The SHAKE algorithm was used to constrain the bonds containing hydrogen atoms to their equilibrium length (Andersen, 1983). The structure was solvated in the TIP3P model water (Mahoney and Jorgensen, 2000) to produce 10 Å thick water shell around the protein. To maintain electrical neutrality sodium ions were added to the system. The Langevin dynamics was utilized to keep a constant temperature of 300 K and the Langevin piston method was used to keep the pressure at 1 atm. in all simulations. The solvated starting structure was minimized using conjugate gradient minimization to an energy gradient tolerance of 0.05. The minimized structure was then heated from 0 K to 300 K in steps of 10 K using velocity reassignment during a 15 ps molecular dynamics run. The equilibrated system was then used for the production runs. The results were saved with frequency 1000, resulting in 2000