Biochimica et Biophysica Acta 1834 (2013) 836–846

Contents lists available at SciVerse ScienceDirect

Biochimica et Biophysica Acta

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

Characterizing conformation changes in proteins through the torsional elastic response ☆

Helena G. Dos Santos [1], Javier Klett [1], Raúl Méndez, Ugo Bastolla *

Centro de Biologia Molecular Severo Ochoa, CSIC-UAM, 28049 Madrid, Spain

ARTICLE INFO

Article history:
Received 29 November 2012
Received in revised form 22 January 2013
Accepted 6 February 2013
Available online 19 February 2013

Keywords:
Normal mode analysis
Elastic network model
Torsion angle
Conformation change
Allostery
Energy barrier

ABSTRACT

The relationship between functional conformation changes and thermal dynamics of proteins is investigated with the help of the torsional network model (TNM), an elastic network model in torsion angle space that we recently introduced. We propose and test a null-model of "random" conformation changes that assumes that the contributions of normal modes to conformation changes are proportional to their contributions to thermal fluctuations. Deviations from this null model are generally small. When they are large and significant, they consist in conformation changes that are represented by very few low frequency normal modes and overcome small energy barriers. We interpret these features as the result of natural selection favoring the intrinsic protein dynamics consistent with functional conformation changes. These "selected" conformation changes are more frequently associated to ligand binding, and in particular phosphorylation, than to pairs of conformations with the same ligands. This deep relationship between the thermal dynamics of a protein, represented by its normal modes, and its functional dynamics can reconcile in a unique framework the two models of conformation changes, conformational selection and induced fit. The program TNM that computes torsional normal modes and analyzes conformation changes is available upon request. This article is part of a Special Issue entitled: The emerging dynamic view of proteins: Protein plasticity in allostery, evolution and self-assembly.

1. Introduction

Proteins are molecular machines that perform their biological function dynamically [1,2]. Stability, i.e., the existence of a well-defined average three-dimensional structure, and at the same time flexibility, i.e., the existence of intrinsic collective movements of large amplitude, characterize the native state of ordered proteins and are key for their catalytic activity [3], ligand binding ability [4], and allosteric regulation [5].

For ordered proteins, the topology of the native state determines to a large extent both the stability of the protein and its intrinsic collective dynamics, which can be predicted by elastic network models (ENM) [6–8,12]. They are Go-like models [9,10] that represent the energetics of the native state based only on its topology. Go models and ENMs fulfill the principle of minimal frustration [11], which assumes that all native interactions are at their energy minimum. Several flavors of ENM have been described in the literature [13]. We have recently introduced the torsional network model (TNM) [14], which adopts the torsion angles of the protein backbone as degrees of freedom, similar to other methods of normal mode analysis and protein dynamics in torsion angle space [15–19]. The TNM has the advantage that it represents all protein atoms with a computationally affordable cost, concentrating on physically allowed motions that do not modify bond lengths and bond angles.

The intrinsic dynamics of the native state of a protein modeled through the ENM can be analytically studied using normal mode analysis (NMA) [20]. NMA approximates the native energy landscape as the harmonic well in the neighborhood of the equilibrium position, and decomposes the native ensemble into a set of independent motions, the normal modes. Low frequency normal modes tend to represent collective motions that produce the largest displacements from the average position. It has been observed that the low frequency normal modes of the ENMs correlate with the intrinsic motions of the protein measured by crystallographic B-factors [15,21,22], despite the fact that B-factors are largely influenced by rigid body motions not represented by NMA [23], they correlate with the essential motions produced by very long molecular dynamics simulations [24,25], and with functional conformation changes such as those upon binding of a physiological ligand, in the sense that often a few low frequency normal modes almost perfectly reproduce the functional motion [26–29].

In this work we investigate the relationship between intrinsic protein motions predicted through ENMs and protein motions observed

☆ This article is part of a Special Issue entitled: The emerging dynamic view of proteins: Protein plasticity in allostery, evolution and self-assembly.

* Corresponding author. Tel.: +34 911964633.
E-mail address: ubastolla@cbm.uam.es (U. Bastolla).

[1] These authors contributed equally to this work.

1570-9639/$ – see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.bbapap.2013.02.010

as two different conformations of the same protein determined in X-ray crystallography experiments or NMR spectroscopy. Linear response theory predicts that the response of the protein to a generic perturbation, for instance ligand binding, is mostly influenced by low frequency normal modes [30,31]. Based on linear response theory, we recently proposed a null model of the response to a "random" perturbation that is independent of the intrinsic dynamics of the proteins, and introduced the parameter ρ that quantifies significant deviations from this null model [14]. When this parameter is large, low frequency normal modes contribute to the conformation change significantly more than expected based on the null model, and the energy barrier opposing to the conformation change is reduced. We observe that in this case only a small number of low frequency normal modes are sufficient to reproduce the conformation change. We interpret these observations as hints of a co-evolution between the functional motion and the intrinsic dynamics of the protein. Here we perform a large scale study of the relationship between conformation changes and intrinsic protein dynamics, analyzing all pairs of crystallized protein structures having the same amino acid sequence and at least 1 Å of root mean square deviation (RMSD).

The paper is organized as follows. In the first section we review the torsional network model used in the computations, reporting computational details omitted in the original publication. In the second section we present our null model of conformation change and the parameter ρ that measures deviations from the null model. In the third section we present the results of a massive analysis of the protein data bank.

2. Torsional network model (TNM)

2.1. Degrees of freedom and kinematics

A molecule composed of n atoms with masses m_i can be represented either through their Cartesian coordinates $\left\{\vec{r}_i\right\}$ or, equivalently, through a set of $n-1$ bonds connecting pairs of atoms, each characterized by the bond length l_a, the bond angle θ_a that it forms with the previous bond and the torsion angle φ_a with respect to the plane of the two previous bonds. In the TNM, only the backbone torsion angles phi (rotation around to the N–C_α bond) and psi (rotation around the C_α–C bond) are allowed to vary, and the other degrees of freedom are kept fixed. Our computer program allows us to additionally select the backbone angle omega and side-chain torsion angles as degrees of freedom, but usually this introduces noise and considerably increases the computation time. We then have to select reference atoms for computing kinetic energy. These can be only α carbons, only β carbons, all backbone atoms, all backbone atoms plus β carbons, or all heavy atoms, which are the best choice, on which the results presented in this paper are based. Bond lengths and bond angles are treated as degrees of freedom if the bond is not a covalent bond, such as the virtual bond connecting two residues separated by a disordered loop whose coordinates cannot be determined in the X-ray experiment. When we analyze a conformation change, we consider for the computation of kinetic energy only residues that are aligned in the two structures and treat gaps in the alignment as disordered loops, nevertheless all atoms are used for computing native interactions (see below). We denote with d the number of degrees of freedom in the torsional space.

The Jacobian matrix that relates infinitesimal torsional and Cartesian displacements is

$$\vec{J}_{ia} \equiv \frac{\partial \vec{r}_i}{\partial \varphi_a} = \chi_{ia}\left(\vec{\tau}_a + \vec{v}_a \times \vec{r}_i\right) \tag{1}$$

where $\chi_{ia} \in \{0,1\}$ is one if atom i is upstream of axes a and zero otherwise (by convention, torsional perturbations are propagated from the N-terminus to the C-terminus of the protein), $\times$ denotes the vector product and $\vec{\tau}_a$ and $\vec{v}_a$ are the translation and the rotation associated with the degree of freedom a, respectively. If the degree of freedom represents the torsion around the axis with unit vector $\vec{e}_a$ and origin $\vec{s}_a$, it is easy to see that $\vec{v}_a = \vec{e}_a$ and $\vec{\tau}_a = -\vec{e}_a \times \vec{s}_a$. If the degree of freedom represents the bond length, it holds $\vec{v}_a = \vec{0}$ and $\vec{\tau}_a = \vec{e}_a$ and if it represents the bond angle, it holds $\vec{v}_a = \left(\vec{e}_{a-1} \times \vec{e}_a\right) / \left|\vec{e}_{a-1} \times \vec{e}_a\right|$ and $\vec{\tau}_a = -\vec{v}_a \times \vec{s}_a$.

The degree of freedom a modifies both the internal degrees of freedom and the rigid body degrees of freedom. To get rid of the latter, we have to impose the Eckart conditions [32]

$$\sum_i m_i \vec{J}'_{ia} = 0, \sum_i m_i \vec{r}_i \times \vec{J}'_{ia} = 0 \tag{2}$$

resulting in $\vec{J}'_{ia} = \chi_{ia}\left(\vec{\tau}_a + \vec{v}_a \times \vec{r}_i\right) + \left(\vec{\tau}'_a + \vec{v}'_a \times \vec{r}_i\right)$, i.e., we have to apply the rigid body transformations $\vec{\tau}'_a$ and $\vec{v}'_a$ that compensate the rigid body motion of the molecule. In the reference frame in which the center of mass is at the origin, it holds $\vec{\tau}'_a = -\frac{Ma}{M}\left(\vec{\tau}_a + \vec{v}_a \times \vec{R}_a\right), I\vec{v}'_a = -I_a\vec{v}'_a - M_a\vec{R}_a \times \vec{\tau}_a$, with $M = \sum_i mi$, $M_a = \sum_i \chi_{ia} m_i$, $M_a\vec{R}_a = \sum_i X_{ia} m_i \vec{r}_i$, I is the inertia tensor and I_α is its restriction to the set having $\chi_{ia} = 1$ [15].

The kinetic energy matrix T in torsion angle space is

$$T_{ab} = \sum_i m_i \frac{\partial \vec{r}_i}{\partial \varphi_a}\frac{\partial \vec{r}_i}{\partial \varphi_b} = \sum_i m_i \vec{J}'_{ia}\vec{J}'_{ib} \times \tag{3}$$

In matrix notation, $T = J'^t M J' = K^t K$, where J' is represented as a $3n \times d$ matrix, the superscript t indicates matrix transposition, M is the diagonal mass matrix (not to be confused with the total mass $M = \sum_i m_i$), and we introduce the notation $K_{ia} = \sqrt{m_i} J'_{ia}$. Taking advantage of the Eckart conditions, we can simplify the formula as

$$T_{ab} = M_{ab}\vec{\tau}_a \cdot \vec{\tau}_b + \vec{v}_a \cdot I^{ab}\vec{v}_b + M_{ab}\vec{R}_{ab} \cdot \left(\vec{\tau}_a \times \vec{v}_b + \vec{\tau}_b \times \vec{v}_a\right) - M\vec{\tau}'_a \cdot \vec{\tau}'_b - \vec{v}'_a \cdot I\vec{v}'_b. \tag{4}$$

Here M_{ab}, $\vec{R}_{ab}$ and I_{ab} are the mass, center of mass and inertia tensor of the set of atoms that are moved by both degrees of freedom a and b, i.e., $\chi_{ia} = \chi_{ib} = 1$. We now exploit the fact that the degrees of freedom are nested, i.e., if axis b is downstream of axis a (which we denote as $b > a$), then $\chi_{ib} = 1$ implies $\chi_{ia} = 1$, so that $M_{ab} = M_b$, unless a and b represent degrees of freedom of different side-chains, in which case $M_{ab} = 0$.

2.2. Potential energy

In ENMs, the effective potential energy of the protein is modeled as a sum of pairwise terms that only runs over native interactions, $V = \sum_{ij} C_{ij} v(r_{ij})$. $r_{ij} = \left|\vec{r}_i - \vec{r}_j\right|$ is the distance between interacting atoms. $C_{ij} = 1$ if the atom i and j are in contact in the native state, 0 otherwise. We use a definition of contacts in which for each pair of residues the two heavy atoms at shorter distance interact provided that their distance is smaller than 4.5 Å. The Go model (or, equivalently, the principle of minimum frustration) requires that each interaction term has a minimum corresponding to the native interaction distance r^0_{ij}. For small displacements from the equilibrium position $\left\{\vec{r}^{\,0}_i\right\}$, the potential energy can be expanded in Taylor series up to second order. Since the constant term can be ignored, and the force