Journal of Structural Biology 190 (2015) 338–347

Contents lists available at ScienceDirect

Journal of Structural Biology

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

Robust elastic network model: A general modeling for precise understanding of protein dynamics

Min Hyeok Kim [a], Byung Ho Lee [b], Moon Ki Kim [a,b,*]

[a] SKKU Advanced Institute of Nanotechnology (SAINT), Sungkyunkwan University, Suwon 440-746, South Korea
[b] School of Mechanical Engineering, Sungkyunkwan University, Suwon 440-746, South Korea

ARTICLE INFO

Article history:
Received 28 January 2015
Received in revised form 7 April 2015
Accepted 9 April 2015
Available online 16 April 2015

Keywords:
Elastic network model
Normal mode analysis
Protein dynamics
B-factor
Vibrational density of states

ABSTRACT

In the study of protein dynamics relevant to functions, normal mode analysis based on elastic network models (ENMs) has become popular. These models are usually validated by comparing the calculated atomic fluctuation for a single protein in a vacuum to experimental temperature factors in the crystal packing state. Without reflecting the crystal packing effect, in addition, their arbitrary assignment of spring constants leads to inaccurate simulation results, yielding a low correlation of the *B*-factor. To overcome this limitation, we propose a robust elastic network model (RENM) that not only considers the crystalline effect by using symmetric constraint information but also uses lumped masses and specific spring constants based on the type of amino acids and chemical interactions, respectively. Simulation results with more than 500 protein structures verify qualitatively and quantitatively that one can obtain the better correlation of the *B*-factor by RENM without additional computational burden. Moreover, an optimal spring constant in physical units (dyne/cm) is quantitatively determined as a function of the temperature at 100 and 290 K, which enables us to predict the atomic fluctuations and vibrational density of states (VDOS) without a fitting process. The additional investigation of 80 high-resolution crystal structures with anisotropic displacement parameters (ADPs) indicates that RENM could give a full description of vibrational characteristics of individual residues in proteins.

1. Introduction

Proteins undergo conformational changes, which are closely related to specific biological functions, including catalysis, regulation, transport, ligand binding, and allosteric regulation (Henzler-Wildman and Kern, 2007). However, it is a great challenge for experimental studies to resolve the protein dynamics, owing to the difficulty in direct observation (Kondrashov et al., 2006). Simulation methods, typically molecular dynamics (MD) simulation, can alternatively explore the fine details of protein dynamics, but the computational cost of these all-atom force field-based calculations is very expensive, which limits the timescale (nanoseconds to microseconds) and data size despite advanced computing technology, including supercomputers. In order to reduce such computational burden, various coarse-grained (CG) methods have been proposed by using the simplified potential and structure. One of the simple but robust CG methods is the elastic network model (ENM). In this model, a target protein is represented as a system composing identical masses, typically a Cα representation, connected by a harmonic linear spring with a uniform (Atilgan et al., 2001; Tama and Sanejouand, 2001; Tirion, 1996) or distance-dependent force constant (Hinsen et al., 2000; Yang et al., 2009). Such dramatic simplification leads to efficient calculation of the CG normal mode without the energy equilibrium. Moreover, the collective motions of a protein, which are often correlated to its intrinsic biological functions, are observed from a few of the lowest frequency normal modes. They are not sensitive to the CG level but only sensitive to their topological features (Tirion, 1996). Several numerical studies have shown that ENM efficiently captures the functionally relevant protein dynamics nearly without limitation in both size and timescale (Bahar and Rader, 2005; Tama and Brooks, 2006).

Abbreviations: ENM, elastic network model; RENM, robust elastic network model; VDOS, vibrational density of states; ADPs, anisotropic displacement parameters; MD, molecular dynamics; CG, coarse-grained; NM, normal mode; ANM, anisotropic network model; GNM, Gaussian network model; CC, correlation coefficient; TLS, translation–liberation–screw; MWCENM, mass weighed chemical elastic network model; NMA, normal mode analysis; EOM, equation of motion.

* Corresponding author at: SKKU Advanced Institute of Nanotechnology (SAINT), Sungkyunkwan University, Suwon 440-746, South Korea.
E-mail address: mkkim@me.skku.ac.kr (M.K. Kim).

http://dx.doi.org/10.1016/j.jsb.2015.04.007
1047-8477/

With a great interest in the application of ENM to protein conformational change, it is a natural issue to improve the validity of ENMs. Experimental *B*-factors, which physically represent the atomic fluctuations, have normally been used to validate the ENMs by comparing with fluctuation values calculated from the normal mode (NM). It is expected that a more accurate ENM can reproduce better *B*-factors. Both the anisotropic network model (ANM) and its isotropic variant, the Gaussian network model (GNM), have yielded reasonable results: namely, the average correlation coefficient (CC) for 113 proteins is 0.55 (Eyal et al., 2006) and 0.59 (Yang et al., 2005), respectively.

After that, more elaborate ENMs, usually considering the experimental crystallization environment, such as the crystal packing state (Hafner and Zheng, 2010; Kundu et al., 2002; Riccardi et al., 2009; Soheilifard et al., 2008; Song and Jernigan, 2007; Zheng, 2010), have been proposed to improve the accuracy of *B*-factor prediction. The computed atomic fluctuations from NM are targeted to an isolated protein, whereas the experimental *B*-factors are determined under the crystalline state. This discrepancy would cause inaccurate simulation results. Indeed, an early study by Yang et al. (2007) has underlined the effect of crystal contact, showing the better correlation (0.75) between GNM and NMR data under the soluble state compared to that (0.49) between GNM and *B*-factors measured by X-ray crystallography. Therefore, to accurately describe protein dynamics in the crystalline state, it is important to properly consider the crystal packing effect. To this end, by taking into account the crystal packing effect, Kundu et al. (2002) have improved the average CC of GNM from 0.59 to 0.66. They also found that the GNM model achieved better correlation with experiments than a simplified translation–liberation–screw (TLS) model without any fitting parameter (Schomaker and Trueblood, 1968; Soheilifard et al., 2008). The TLS model, which regards a protein as an assembly of rigid subunits, emphasizes the rigid body motion as the significant contribution for the crystalline *B*-factor. It usually fits the *B*-factors by optimizing the translation, rotation, and screwing motions. Although several studies considering the rigid body motion as another crystallization environment have improved the *B*-factor correlation (Soheilifard et al., 2008; Song and Jernigan, 2007), its extent of contribution to the *B*-factor is under debate (Meinhold and Smith, 2005; Song and Jernigan, 2007; Zheng, 2010). As the criterion used for optimizing the model is not always the same for every protein and the possibility of overfitting with many parameters still exists, the induced models may not have a physical basis. Moreover, the arbitrary contributions of parameters from the fitting process could benefit the understanding of the dynamics of specific target proteins, but this makes it difficult to anticipate the experimental *B*-factors. There is no doubt that, as the number of fitting parameters increases, much higher CC values are naturally expected, but the prediction of a high variance of parameters becomes much harder.

This paper presents a new systematic extension of ENM by not only reflecting the crystal packing effect but also incorporating the chemical and inertia information into the coarse-grained model. Unlike the previous methods to consider the crystal packing effect, we have adopted a more efficient strategy, which focuses on a single protein molecule as our main structure model while the packing effect of surrounding protein structures is mathematically applied by symmetric constraints. This moderate model could keep the computational cost at the level of a single protein structure by efficiently reflecting the crystal effect without limit. Moreover, consideration of the real chemical and physical properties, such as chemical interactions and residue inertia, not completely but somewhat improves the average CC values without further computational burden. The assignment of the chemical information according to the types of chemical bonds also leads to the reliable vibrational density of states (VDOS) distribution, showing the bimodal state with low and high frequency spectrum ranges, which is in good agreement with that using the all-atom CHARMM force-field model. Unlike the previous *B*-factor fitting methods, the proposed model only needs a single parameter, such as the spring constant; thus, it can be empirically determined based on the experimental *B*-factor. Finally, the universal stiffness value obtained in this study not only achieved the reliable atomic fluctuation in terms of low root mean square deviation (RMSD) value with experimental *B*-factors but also described the accurate peaks in low vibrational frequency ranges, which are very important fingerprints in the dynamics of individual proteins but laborious to indicate with existing experimental methods (Balog et al., 2004; He et al., 2011).

2. Materials and methods

As the details of ENM have already been presented elsewhere (Kim et al., 2002), here is given only a brief description. ENM for protein dynamics is based on the underlying assumption that the total potential energy, *V*, of a given protein is expressed as the sum of harmonic interactions such that:

$$V = \frac{1}{2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} k_{ij}\{\|\vec{r}_i(t) - \vec{r}_j(t)\| - \|\vec{r}_i(0) - \vec{r}_j(0)\|\}^2, \quad (1)$$

where $k_{i,j}$ is a harmonic force constant between the *i*th and *j*th atom, *N* is the total number of particles, $\vec{r}_i(0)$ is the initial equilibrium position of the *i*th atom and $\vec{r}_i(t) = \vec{r}_i(0) + \vec{\delta}_i$. Here, $\vec{\delta}_i$ is assumed to be a small displacement of the initial position $\vec{r}_i(0)$. In general ENM, the representative atoms are usually Cα atoms. In the pioneering study (Tirion, 1996), this quadratic potential was be sufficient to describe low-frequency collective motions of proteins.

In this study, we propose a more robust ENM, called RENM, with the following two strategies. First, we define the force constants assigning the various stiffness values based on the types of chemical interactions. Moreover, the total mass of each residue is also assumed to be a lumped mass on the Cα atom. This mass weighted chemical ENM, called MWCENM (Kim et al., 2013), enables us to generate more practical and accurate simulation results, such as the frequency spectrum in the form of the vibrational density of states (refer to the following subchapter). Second, we develop the symmetry-constrained elastic network model, called SCENM (Kim et al., 2003; Lee et al., 2014), to take the crystal packing effect into account. With the space group information in PDB, the corresponding constraints on crystal contact are properly assigned to the main protein structure. In short, these two strategies are intended to effectively define the intra-connection and inter-connection, respectively. The following subchapters introduce them in more detail.

2.1. Intra-connection model: mass weighed chemical elastic network model (MWCENM)

MWCENM was originally introduced by Kim et al. (2013) to describe a feasible conformational change with realistic spring connections and lumped masses. It is especially useful for closed proteins, which have often been difficult to properly show the dynamics of in the standard ENM due to their excessive spring connections. Another benefit of MWCENM is the capability to generate vibration characteristics of proteins quantitatively. That is, real physical values of stiffness and mass used in MWCENM enable us to obtain not only the frequency spectra but also the corresponding vibration modes of target proteins.