

RESEARCH

Open Access

Generalized spring tensor models for protein fluctuation dynamics and conformation changes

Tu-Liang Lin^{1*}, Guang Song^{1,2,3*}

From Computational Structural Bioinformatics Workshop 2009
Washington, DC, USA. 1 November 2009

Abstract

Background: In the last decade, various coarse-grained elastic network models have been developed to study the large-scale motions of proteins and protein complexes where computer simulations using detailed all-atom models are not feasible. Among these models, the Gaussian Network Model (GNM) and Anisotropic Network Model (ANM) have been widely used. Both models have strengths and limitations. GNM can predict the relative magnitudes of protein fluctuations well, but due to its isotropy assumption, it can not be applied to predict the directions of the fluctuations. In contrast, ANM adds the ability to do the latter, but loses a significant amount of precision in the prediction of the magnitudes.

Results: In this article, we develop a single model, called generalized spring tensor model (STeM), that is able to predict well both the magnitudes and the directions of the fluctuations. Specifically, STeM performs equally well in B-factor predictions as GNM *and* has the ability to predict the directions of fluctuations as ANM. This is achieved by employing a physically more realistic potential, the $G\ddot{o}$ -like potential. The potential, which is more sophisticated than that of either GNM or ANM, though adds complexity to the derivation process of the Hessian matrix (which fortunately has been done once for all and the MATLAB code is freely available electronically at <http://www.cs.iastate.edu/~gsong/STeM>), causes virtually no performance slowdown.

Conclusions: Derived from a physically more realistic potential, STeM proves to be a natural solution in which advantages that used to exist in two separate models, namely GNM and ANM, are achieved in one single model. It thus lightens the burden to work with two separate models and to relate the modes of GNM with those of ANM at times. By examining the contributions of different interaction terms in the $G\ddot{o}$ potential to the fluctuation dynamics, STeM reveals, (i) a physical explanation for why the distance-dependent, inverse distance square (i.e., $\frac{1}{r^2}$) spring constants perform better than the uniform ones, and (ii), the importance of three-body and four-body interactions to properly modeling protein dynamics.

Introduction

It is now well accepted that the functions of a protein are closely related to not only its structure but also its dynamics. With the advancement of the computational power and increasing availability of computational resources, function-related protein dynamics, such as large-scale conformation transitions, has been probed by various computational methods at multiple scales.

Among these computational methods, coarse-grained models play an important role since many functional processes take place over time scales that are well beyond the capacity of all-atom simulations [1]. One type of coarse-grained models, the elastic network models (ENMs), have been particularly successful and widely used in studying protein dynamics and in relating the intrinsic motions of a protein with its functional-related conformation changes over the last decade [2-5].

The reason why ENMs have been well received as compared to the conventional normal mode analysis

* Correspondence: tlin@cs.iastate.edu; gsong@iastate.edu

¹Computer Science Department, Iowa State University, 226 Atanasoff Hall, Ames, IA 50011, USA

(NMA) lies at its simplicity to use. ENMs do not require energy minimization and therefore can be applied directly to crystal structures to compute the modes of motions. In contrast, minimization is required for carrying out the conventional normal mode analysis (NMA). The problematic aspect of energy minimization is that it usually shifts the protein molecule away from its crystal conformation by about 2 Å. In addition, in ENMs analytical solutions to residue fluctuations and motion correlations can be easily derived. On the other hand, the simplicity of ENMs leaves much room for improvement and many new models have been proposed [6-12].

The two most widely used ENM models are Gaussian Network Model (GNM) and Anisotropic Network Model (ANM). They have been used to predict the magnitudes or directions of the residue fluctuations from a single structure and have been applied in many research areas [4,5], such as domain decomposition [13] and allosteric communication [14-17]. Both models have their own advantages and disadvantages. GNM can predict the relative magnitudes of the fluctuations well, but due to its isotropy assumption, it can not be applied to predict the directions of the fluctuations. In contrast, ANM adds the ability to do the latter, but it loses a significant amount of precision in the prediction of the magnitudes.

Gaussian network model. Gaussian Network Model (GNM) was first introduced in [2] under the assumption that the separation between a pair of residues in the folded protein is Gaussianly distributed. Given its simplicity, the model performs extremely well in predicting the experimental B-factors. The model represents a protein structure using its C_α atoms. The connectivity among the C_α 's is expressed in Kirchhoff matrix Γ (see Eq. (1)). Two C_α 's are considered to be in contact if their distance falls within a certain cutoff distance. The cutoff distance between a pair of residues is the only parameter in the model and is normally set to be 7 Å to 8 Å. Let $\Delta\mathbf{r}_i$ and $\Delta\mathbf{r}_j$ represent the instantaneous fluctuations from equilibrium positions of residues i and j and r_{ij} and $r_{0,ij}$ be the respective instantaneous and equilibrium distances between residues i and j . The Kirchhoff matrix Γ is:

$$\Gamma_{ij} = \begin{cases} -1 & \text{if } i \neq j \cap r_{0,ij} \leq r_c \\ 0 & \text{if } i \neq j \cap r_{0,ij} > r_c \\ \sum_{j,j \neq i}^N \Gamma_{ij} & \text{if } i = j \end{cases} \quad (1)$$

where i and j are the indices of the residues and r_c is the cutoff distance.

The simplicity of the Kirchhoff matrix formulation results from the assumption that the fluctuations of each residue are isotropic and Gaussianly distributed along the X, Y and Z directions. The expected value of residue fluctuations, $\langle \Delta\mathbf{r}_i^2 \rangle$, and correlations, $\langle \Delta\mathbf{r}_i \cdot \Delta\mathbf{r}_j \rangle$, can be easily obtained from the inverse of the Kirchhoff matrix:

$$\langle \Delta\mathbf{r}_i^2 \rangle = \frac{3k_B T}{\gamma} (\Gamma^{-1})_{ii}, \quad (2)$$

$$\langle \Delta\mathbf{r}_i \cdot \Delta\mathbf{r}_j \rangle = \frac{3k_B T}{\gamma} (\Gamma^{-1})_{ij}, \quad (3)$$

where k_B is the Boltzmann constant and T is the temperature. γ is the spring constant. The $\langle \Delta\mathbf{r}_i^2 \rangle$ term is directly proportional to the crystallographic B-factors.

Anisotropic network model. GNM provides only the magnitudes of residue fluctuations. To study the motions of a protein in more details, especially to determine the directions of the fluctuations, normal mode analysis (NMA) is needed. Traditional NMA is all-atom based and requires a structure to be first energy-minimized before the Hessian matrix and normal modes can be computed, which was rather cumbersome. Even after the energy minimization, the derivation of the Hessian matrix is not easy due to the complicated all-atom potential. In Tirion's pioneering work [18], the energy minimization step was removed and a much simpler Hookean potential was used, and yet it was shown that the low frequency normal modes remained mostly accurate. Since then, the Hookean spring potentials have been favored in most coarse-grained C_α models [3,19,20]. One of such models is best known as Anisotropic Network Model (ANM) [3] since it has anisotropic, directional information of the fluctuations. The potential in ANM has the simplest harmonic form. Assuming that a given structure is at equilibrium, the Hessian matrix ($3N \times 3N$) can be derived analytically from such a potential [3]. The $3N \times 3N$ Hessian matrix \mathbf{H}_{ANM} can be repartitioned into $N \times N$ super elements and each super element is a 3×3 tensor.

$$\mathbf{H}_{ANM} = \begin{bmatrix} \mathbf{H}_{1,1} & \mathbf{H}_{1,2} & \mathbf{H}_{1,N} \\ \mathbf{H}_{2,1} & \mathbf{H}_{2,2} & \mathbf{H}_{2,N} \\ \vdots & \vdots & \vdots \\ \mathbf{H}_{N,1} & \mathbf{H}_{N,2} & \mathbf{H}_{N,N} \end{bmatrix} \quad (4)$$

where \mathbf{H}_i is the interaction tensor between residues i and j and can be expressed as:

$$\mathbf{H}_{i,j} = \begin{bmatrix} \frac{\partial^2 V}{\partial X_i \partial X_j} & \frac{\partial^2 V}{\partial X_i \partial Y_j} & \frac{\partial^2 V}{\partial X_i \partial Z_j} \\ \frac{\partial^2 V}{\partial Y_i \partial X_j} & \frac{\partial^2 V}{\partial Y_i \partial Y_j} & \frac{\partial^2 V}{\partial Y_i \partial Z_j} \\ \frac{\partial^2 V}{\partial Z_i \partial X_j} & \frac{\partial^2 V}{\partial Z_i \partial Y_j} & \frac{\partial^2 V}{\partial Z_i \partial Z_j} \end{bmatrix} \quad (5)$$

Let \mathbf{H}^+ be the pseudo inverse of Hessian matrix \mathbf{H}_{ANM} . The mean square fluctuation $\langle \Delta \mathbf{r}_i^2 \rangle$ and correlation $\langle \Delta \mathbf{r}_i \cdot \Delta \mathbf{r}_j \rangle$ can be calculated by summing over the X, Y and Z components:

$$\langle \Delta \mathbf{r}_i^2 \rangle = \frac{3k_B T}{\gamma} \left(H_{3i-2,3i-2}^+ + H_{3i-1,3i-1}^+ + H_{3i,3i}^+ \right) \quad (6)$$

$$\langle \Delta \mathbf{r}_i \cdot \Delta \mathbf{r}_j \rangle = \frac{3k_B T}{\gamma} \left(H_{3i-2,3j-2}^+ + H_{3i-1,3j-1}^+ + H_{3i,3j}^+ \right) \quad (7)$$

Strengths and limitations of GNM and ANM. The advantages of ANM or GNM over the conventional NMA lie in several aspects: (i) it is a coarse-grained model and uses the C_α 's to represent the residues in a structure; (ii) it does not require energy minimization and thus can be applied directly to crystal structures to compute the modes of motions; (iii) it provides analytical solutions to the mean square fluctuations and motion correlations.

The limitations of the GNM model. GNM provides only information on the magnitudes of residue fluctuations but no directional information. Therefore, the modes of GNM should not be interpreted as protein motions or components of the motions, since the potential in GNM is not rotationally invariant [21].

The limitations of the ANM model. In contrast to that in GNM, the potential in ANM is based on simple, harmonic Hookean springs and is rotationally invariant. And thus, the modes of ANM do represent the possible modes of protein motions. In doing this, however, ANM loses a significant amount of precision in predicting the magnitudes of the fluctuations. The reason is that, in GNM, the fluctuations in the separation between a pair of residues are assumed to be Gaussianly distributed and isotropic, while in ANM, because only a Hookean spring is attached between a pair of residues i and j, the fluctuation of residue j is constrained only longitudinally along the axis from i to j. The fluctuation is unconstrained transversely. The interaction spring tensor $\mathbf{H}_{i,j}^{ANM}$ between residues i and j in Eq. (5) becomes the following in the local frame (where the Z axis is along the direction from residues i to j):

$$\mathbf{H}_{i,j}^{ANM} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (8)$$

Because the fluctuation of residue j is unconstrained transversely relative to residue i, the fluctuations given by ANM are less realistic than those by GNM, which are assumed to be isotropic. The isotropy in GNM is equivalent to an interaction spring tensor between residues i and j of the following form:

$$\mathbf{H}_{i,j}^{ANM} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (9)$$

From the two tensors $\mathbf{H}_{i,j}^{ANM}$ and $\mathbf{H}_{i,j}^{GNM}$ given in Eqs. 8 and 9, the causes for the limitations in GNM and ANM are clearly displayed. The unrealistic-ness in ANM is an artifact resulting from its over-simplified potential. The isotropy assumption of GNM, on the other hand, does a better job than ANM in modeling the effect of residue interactions on the magnitudes of the fluctuations, but gives up completely on representing the anisotropic nature that is intrinsic to all physical forces and interactions, since only the magnitudes of the mean-square fluctuations and cross-correlations were of concern when GNM was first proposed. Therefore, to overcome the limitations of GNM and ANM, what is needed is a generalized interaction spring tensor that both is anisotropic and can exert more proper constraints on the fluctuations than the ANM tensor $\mathbf{H}_{i,j}^{ANM}$ does. This calls for a model that has a physically more realistic potential than that of ANM. Since potentials with only two-body interactions can provide only longitudinal constraints, it is necessary to include multi-body interactions in the potential in order to have transversal constraints as well. The multi-body interactions provide additional diagonal and off-diagonal terms to the interaction spring tensor between residues i and j. For example, by properly including three-body interactions, the interaction spring tensor may look like:

$$\mathbf{H}_{i,j}^{STEEM} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & T(i,j) \end{bmatrix} + \sum_k \begin{bmatrix} S_{11}(i,j,k) & S_{12}(i,j,k) & S_{13}(i,j,k) \\ S_{21}(i,j,k) & S_{22}(i,j,k) & S_{23}(i,j,k) \\ S_{31}(i,j,k) & S_{32}(i,j,k) & S_{33}(i,j,k) \end{bmatrix} \quad (10)$$

where k represent the indices of the residues that interact with both residues i and j through three-body interaction S. The first tensor on the right side of the equation represents the two-body interaction, which is similar to $\mathbf{H}_{i,j}^{ANM}$, except that the interaction strength $T(i,j)$ depends on residues i and j, and thus may be distance-dependent as well.

Our contributions. To overcome the limitations of ANM and GNM, we have developed a generalized spring tensor model for studying protein fluctuation dynamics and conformation changes. It is called generalized spring tensor model, or STeM, for the reason that the interaction between a pair of residues i and j is no longer a Hookean spring that has the tensor form of Eq. (8), but takes a generalized tensor form (similar to that in Eq. 10) that can provide both longitudinal and transversal constraints on a residue's fluctuations relative to its neighbours. We obtain the generalized tensor form by deriving the Hessian matrix from a physically more realistic $G\delta$ -like potential (Eq. 11), which has been successfully used in many MD simulations to study protein folding processes and conformation changes [22-24]. In addition to the Hookean spring interactions, the potential includes bond bending and torsional interactions, both of which had been found to be helpful in removing the "tip effect" of the ANM model [9]. The inclusion of the bond bending and torsional interactions is reflected in the generalized tensor spring interaction between residues i and j , in such a way that the tensor now includes not only the two-body interaction between residues i and j , but also three-body and four-body interactions that involve residues i and j (see Eq. 10). In doing this, the STeM model is able to integrate all the aforementioned attractive features of ANM and GNM and overcome their limitations. Specifically, STeM performs equally well in B-factors predictions as GNM and has the ability to predict the directions of the fluctuations as ANM. This is accomplished with virtually no performance slowdown. The only potential drawback of this model is the significantly increased complexity in deriving the Hessian matrix. Fortunately, this has been done once for all and the derivation results are available electronically at <http://www.cs.iastate.edu/~gsong/STeM>. STeM is physically more accurate by explicitly including the bond bending and torsional interactions since they capture the chain behavior of protein molecules, which are neglected in most elastic network models where a protein is treated as an elastic rubber. Therefore, we have reasons to expect this model will further distinguish itself in studying protein dynamics where a correct modeling of bond bending and/or torsional rotations is critical.

Results and discussion

Crystallographic B-factor prediction

Table 1 shows the correlation coefficients between the experimental and calculated B-factors of the 111 proteins in the first dataset. The mean values of the correlation coefficients of ANM, GNM, and STeM are 0.53, 0.59, and 0.60 respectively. STeM provides the directional information of the residue fluctuations as

ANM and has an accuracy even slightly better than GNM in B-factor predictions. Figure 1 shows the distributions of the correlation coefficients between the calculated B-factors and the experimental B-factors. STeM is the only model in which there are instances where the correlation coefficient is above 0.85 and no instances where the correlation coefficient is below 0.25. This implies that the performance of STeM is more steady than either ANM or GNM. The scatter plot of the correlation coefficients between ANM and STeM in Figure 2 shows that STeM performs better than ANM for 80% of the proteins in the dataset. Protein structures of higher resolution have more accurate data on atom coordinates and B-factors. We investigate whether our model's performance can be further improved when the dataset used is limited to structures with higher resolution. We select the 12 structures with resolution better than 1.3 Å from the first dataset. The mean values of the correlation coefficients of these 12 structures are 0.56, 0.62, and 0.63 for ANM, GNM, and STeM, respectively, which gives an improvement of about 5%-6% for all of the three models. Since the improvement is based on a relatively small set of 12 structures, a larger dataset is needed to further examine this potential dependence of B-factor prediction accuracy on structure quality.

The contributions of different interaction terms to the fluctuations

The $G\delta$ -like potential in eq. (11) has four different interaction terms, namely, bond stretching, bond bending, torsional interactions, and the non-bonded interactions. It is of great interest to investigate the relative contributions of these different terms to the agreement with experimental B-factors. Since only the non-bonded interaction term (V_4) is able to provide by itself enough constraints to ensure the Hessian matrix to have no more than six zero eigenvalues, V_4 is used as the base term for the evaluation of different terms' contributions to the mean-square fluctuations. The Hessian matrix of ANM, denoted by \mathbf{H}_{ANM} , is used as another baseline for comparison purposes. Table 2 lists the contributions of these different terms to the improvement of B-factor predictions as they are added to the potential.

First, it is seen that the non-bonded interactions, as are present in \mathbf{H}_{V_4} and \mathbf{H}_{ANM} , play a dominant role in contributing to the B-factors. This is not surprising since the mean-square fluctuations of a residue are mostly constrained by its interactions with its spatial neighbours, most of which are through non-bonded interactions. What is more interesting is that \mathbf{H}_{V_4} term alone performs better than \mathbf{H}_{ANM} . This is in agreement with recent results that the performance of

Table 1 The correlation coefficients between the experimental and calculated B-factors using different models

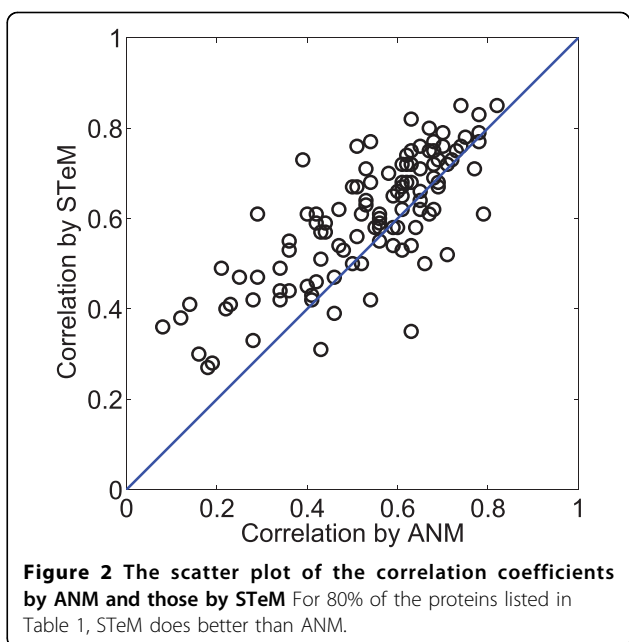
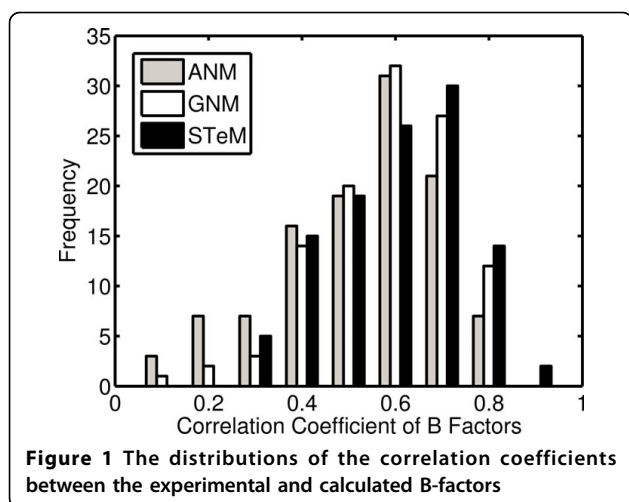
Protein	R(Å)	ANM	GNM	STeM	Protein	R(Å)	ANM	GNM	STeM	Protein	R(Å)	ANM	GNM	STeM
1AAC	1.31	0.7	0.71	0.76	1ADS	1.65	0.77	0.74	0.71	1AHC	2.00	0.79	0.68	0.61
1AKY	1.63	0.56	0.72	0.6	1AMM	1.20	0.56	0.72	0.55	1AMP	1.80	0.62	0.59	0.68
1ARB	1.20	0.78	0.76	0.83	1ARS	1.80	0.14	0.43	0.41	1ARU	1.60	0.7	0.78	0.79
1BKF	1.60	0.52	0.43	0.5	1BPI	1.09	0.43	0.56	0.57	1CDG	2.00	0.65	0.62	0.71
1CEM	1.65	0.51	0.63	0.76	1CNR	1.05	0.34	0.64	0.42	1CNV	1.65	0.69	0.62	0.68
1CPN	1.80	0.51	0.54	0.56	1CSH	1.65	0.44	0.41	0.57	1CTJ	1.10	0.47	0.39	0.62
1CUS	1.25	0.74	0.66	0.76	1DAD	1.60	0.28	0.5	0.42	1DDT	2.00	0.21	-0.01	0.49
1EDE	1.90	0.67	0.63	0.75	1EZM	1.50	0.56	0.6	0.58	1FNC	2.00	0.29	0.59	0.61
1FRD	1.70	0.54	0.83	0.77	1FUS	1.30	0.4	0.63	0.61	1FXD	1.70	0.58	0.56	0.7
1GIA	2.00	0.68	0.67	0.69	1GKY	2.00	0.36	0.55	0.44	1GOF	1.70	0.75	0.76	0.78
1GPR	1.90	0.65	0.62	0.66	1HFC	1.50	0.63	0.38	0.35	1IAB	1.79	0.36	0.42	0.53
1IAG	2.00	0.34	0.52	0.44	1IFC	1.19	0.61	0.67	0.53	1IGD	1.10	0.18	0.44	0.27
1IRO	1.10	0.82	0.51	0.85	1JBC	1.15	0.72	0.7	0.73	1KNB	1.70	0.63	0.66	0.54
1LAM	1.60	0.53	0.63	0.71	1LCT	2.00	0.52	0.57	0.61	1LIS	1.90	0.16	0.43	0.3
1LIT	1.55	0.65	0.62	0.76	1LST	1.80	0.39	0.72	0.73	1MJC	2.00	0.67	0.67	0.61
1MLA	1.50	0.59	0.57	0.54	1MRJ	1.60	0.66	0.49	0.5	1NAR	1.80	0.62	0.76	0.74
1NFP	1.60	0.23	0.48	0.41	1NIF	1.70	0.42	0.58	0.61	1NPK	1.80	0.53	0.55	0.64
1OMP	1.80	0.61	0.63	0.65	1ONC	1.70	0.55	0.7	0.58	1OSA	1.68	0.36	0.42	0.55
1OYC	2.00	0.78	0.73	0.77	1PBE	1.90	0.53	0.61	0.63	1PDA	1.76	0.6	0.76	0.58
1PHB	1.60	0.56	0.52	0.59	1PHP	1.65	0.59	0.63	0.65	1PII	2.00	0.19	0.44	0.28
1PLC	1.33	0.41	0.47	0.42	1POA	1.50	0.54	0.66	0.42	1POC	2.00	0.46	0.52	0.39
1PPN	1.60	0.61	0.64	0.67	1PTF	1.60	0.47	0.6	0.54	1PTX	1.30	0.65	0.51	0.62
1RA9	2.00	0.48	0.61	0.53	1RCF	1.40	0.59	0.63	0.58	1REC	1.90	0.34	0.5	0.49
1RIE	1.50	0.71	0.25	0.52	1RIS	2.00	0.25	0.24	0.47	1RRO	1.30	0.08	0.31	0.36
1SBP	1.70	0.69	0.72	0.67	1SMD	1.60	0.5	0.62	0.67	1SNC	1.65	0.68	0.71	0.72
1THG	1.80	0.5	0.53	0.5	1TML	1.80	0.64	0.64	0.58	1UBI	1.80	0.56	0.69	0.61
1WHI	1.50	0.12	0.33	0.38	1XIC	1.60	0.29	0.4	0.47	2AYH	1.60	0.63	0.73	0.82
2CBA	1.54	0.67	0.75	0.8	2CMD	1.87	0.68	0.6	0.62	2CPL	1.63	0.61	0.6	0.72
2CTC	1.40	0.63	0.67	0.75	2CY3	1.70	0.51	0.5	0.67	2END	1.45	0.63	0.71	0.68
2ERL	1.00	0.74	0.73	0.85	2HFT	1.69	0.63	0.79	0.72	2IHL	1.40	0.62	0.69	0.72
2MCM	1.50	0.78	0.83	0.79	2MHR	1.30	0.65	0.52	0.64	2MNR	1.90	0.46	0.5	0.47
2PHY	1.40	0.54	0.55	0.68	2RAN	1.89	0.43	0.4	0.31	2RHE	1.60	0.28	0.38	0.33
2RN2	1.48	0.68	0.71	0.75	2SIL	1.60	0.43	0.5	0.51	2TGI	1.80	0.69	0.71	0.73
3CHY	1.66	0.61	0.75	0.68	3COX	1.80	0.71	0.71	0.72	3EBX	1.40	0.22	0.58	0.4
3GRS	1.54	0.44	0.57	0.59	3LZM	1.70	0.6	0.52	0.66	3PTE	1.60	0.68	0.83	0.77
4FGF	1.60	0.41	0.27	0.43	4GCR	1.47	0.73	0.81	0.75	4MT2	2.00	0.42	0.37	0.46
5P21	1.35	0.4	0.51	0.45	7RSA	1.26	0.42	0.63	0.59	8ABP	1.49	0.61	0.82	0.62

Column R(Å) gives the resolution of each structure.

B-factor predictions can be improved by using distance-dependent force constants [25,26]. Particularly, the spring constants that take the form of inverse distance square have been shown to be superior in a recent exhaustive study that experimented with different distance-dependent spring constants on a large dataset [10]. The Taylor expansion of the non-bonded interaction term (V_4) shows that it has an equivalent spring constant of the form $\frac{120\epsilon}{r_{0,ij}^2}$ (see Eq. 36), which is exactly proportional to the inverse of the pairwise

distance square. Thus, STeM provides a physics-based explanation for the choice of using inverse square distance spring constants.

The contribution to the improvement in B-factor predictions from each of the bonded interactions, such as that of bond stretching, is small, as had been pointed out by Bahar et al when GNM was first proposed over a decade ago [2]. However, when the contributions of all of these four terms are added up, they together enable the STeM model to gain a significant improvement over ANM to reach the level of accuracy on a par with GNM.



Conformational change evaluation

It is known that the modes derived from the open form of a structure have better overlaps and correlations with the direction of a protein's conformation change than the ones derived from the closed form [20]. Here we apply the STeM model to study the conformation changes between the open and closed forms of 20 proteins and the open forms are used to calculate the normal modes. Table 3 lists the overlaps and correlations of the observed conformation changes and the indices of the modes that are most involved in the conformation changes. GNM is not considered since it cannot provide directional information. The mean values of the overlaps and correlation coefficients of ANM are 0.49 and 0.61 respectively, and 0.52 and 0.64 respectively for STeM. These amount to an improvement of about 5%

for STeM over ANM on both overlap and correlation. Since the results are obtained based on a relatively small set of 20 protein pairs, the significance of the improvement seen here needs to be further tested by conducting a more exhaustive analysis that uses a larger set of proteins and varying parameters, and preferably taking into account the effect of crystal packing as well. We will leave this for future work. It is also worth noting that, in both the overlap and correlation calculations, the modes that are most involved in the conformation change tend to have lower indices in STeM than in ANM (see Table 3), which may imply the modes of STeM be of higher quality than those of ANM.

Conclusions

Protein mean-square fluctuations and conformation changes are two closely related aspects of protein dynamics. However, in the past, two separate groups of models were needed to best explain protein mean-square fluctuations or conformation changes. Specifically, the best models for predicting mean-square fluctuations cannot predict conformation changes, and the models that can predict conformation changes do not have the best performance in predicting mean-square fluctuations. There is thus an obvious gap between the models that work well in predicting one aspect of the dynamics and those in another. Since protein mean-square fluctuations and conformation changes are two closely related dynamic phenomena and share a similar physical origin, we reasoned that models based on a physically more accurate potential should be able to bridge the gap and predict both aspects of the protein dynamics well. Indeed, by using a $G\delta$ -like potential, we have successfully developed a spring tensor model (STeM) that is able to singly predict well both mean-square fluctuations and conformation changes. Specifically, STeM performs equally well in B-factor predictions as GNM *and* has the ability to predict the directions of fluctuations as ANM.

The new STeM model does come with a cost. As is seen, the derivation process of the Hessian matrix in STeM is much more complex than models using only two-body Hookean potentials, such as those used in ANM. However, the introduced complexity in the potential is necessary in resolving the aforementioned gap that is mainly due to over-simplified potentials and in providing a single, unified model for protein dynamics. Moreover, the derivation process, though more complex, needs to be done only once. Examining the different interaction terms in the $G\delta$ potential and their contributions to the agreement with experimental B-factors provides further benefits. Along the way, we have discovered a physical explanation for why the distance-dependent, inverse

Table 2 The contributions of different interaction terms to the agreement with experimental B-factors

Hessian matrices used	Correlation Coefficient with B-factors	Improvement with respect to ANM
H_{ANM}	0.53	0.00
H_{V_4}	0.55	0.02
$H_{V_4} + H_{V_1}$	0.57	0.04
$H_{V_4} + H_{V_2}$	0.57	0.04
$H_{V_4} + H_{V_3}$	0.56	0.03
$H_{V_4} + H_{V_1} + H_{V_2}$	0.59	0.06
$H_{V_4} + H_{V_1} + H_{V_3}$	0.58	0.05
$H_{V_4} + H_{V_2} + H_{V_3}$	0.57	0.04
$H_{V_4} + H_{V_1} + H_{V_2} + H_{V_3}$ (= H_{STeM})	0.60	0.07
$H_{ANM} + H_{V_1}$	0.54	0.01
$H_{ANM} + H_{V_2}$	0.54	0.01
$H_{ANM} + H_{V_3}$	0.54	0.01
$H_{ANM} + H_{V_1} + H_{V_2} + H_{V_3}$	0.56	0.03

H_{ANM} is the Hessian matrix of ANM. H_{V_1} , H_{V_2} , H_{V_3} , and H_{V_4} are the Hessian matrices of the bond stretching (V_1), bond bending (V_2), torsional rotation (V_3), and non-local interaction (V_4) terms, respectively.

distance square (i.e., $\frac{1}{r^2}$) spring constants perform better than the uniform ones. The van der Waals interaction term in the potential naturally renders inverse distance square spring constants! By including the bond bending and torsional interactions and their contributions to the improvement in B-factor predictions, the STeM model

confirms the importance of 3-body and 4-body potentials. The importance of multi-body potentials is made even more evident when their contribution to the interaction spring tensor is examined - the multi-body potentials are shown to be necessary in providing proper constraints on residue fluctuations, even transversely. It is worth noting that the 3-body and 4-body potentials introduced through bond bending and torsional interactions only scratch the surface of the full extensivity of the multi-body potentials since bond bending and torsional interactions are restricted to only consecutive residues along the protein chain. The improvement seen here calls for other generalized spring tensor models that have a thorough treatment of the multi-body potentials. Chain breaking, such as that due to missing residues, has a more felt impact on STeM than on ANM or GNM, since the first, the second, and the third terms of the potential used to derive the model are all related to the continuity of the chain. We have not evaluated such impact in the current work but this could be a future research direction and our STeM model would be a proper tool for evaluating the impact of chain breaking on protein motions. STeM does not always outperform ANM in B-factor predictions - it does better than ANM for 80% of the proteins studied. It would be interesting to find out why this is so. Crystal packing has been known to impact significantly the mean-square fluctuations. Therefore, a proper inclusion of the crystal packing effect may further enhance STeM's performance. Since STeM takes into account bond bending and

Table 3 The overlaps and correlations between the observed conformation changes and the most involved modes using different models and the open conformations

Protein	Overlap in ANM	Correlation in ANM	Overlap in STeM	Correlation in STeM
Adenylate kinase	0.49(1)	0.62(1)	0.55(1)	0.63 (1)
Alcohol dehydrogenase	0.69(3)	0.54(9)	0.73 (2)	0.65 (30)
Annexin V	0.33(1)	0.60(32)	0.33 (1)	0.56 (22)
Aspartate aminotransferase	0.56(9)	0.63(9)	0.68 (6)	0.67 (6)
Calmodulin	0.44(5)	0.62 (77)	0.48 (1)	0.62 (16)
Che Y protein	0.46(1)	0.78(12)	0.40(1)	0.74(1)
Citrate synthase	0.48(7)	0.72(26)	0.49(5)	0.63(5)
Dihydrofolate reductase	0.71(1)	0.65(1)	0.73(1)	0.66(1)
Diphtheria toxin	0.43(1)	0.69(2)	0.50(2)	0.73(2)
Enolase	0.31(1)	0.45(34)	0.32(1)	0.49(53)
HIV-1 protease	0.67(1)	0.78 (10)	0.85 (1)	0.90(1)
Immunoglobulin	0.68(3)	0.57(3)	0.66(3)	0.58(3)
Lactoferrin	0.48(1)	0.64(24)	0.48(1)	0.70(36)
LAO binding protein	0.81(1)	0.74(1)	0.87(1)	0.80(1)
Maltodextrin binding protein	0.77(2)	0.66(2)	0.80(2)	0.70(2)
Seryl-tRNA synthetase	0.21(4)	0.59(10)	0.21(4)	0.60(37)
Thymidylate synthase	0.37(4)	0.69(9)	0.44(3)	0.68(9)
Triglyceride lipase	0.35(15)	0.50(25)	0.30(14)	0.56(24)
Triose phosphate isomerase	0.15(38)	0.28(11)	0.14(7)	0.30(8)
Tyrosine phosphatase	0.41(2)	0.57(27)	0.42(1)	0.59(25)

torsional interactions, it is expected that it should further distinguish itself in studying protein dynamics where a correct modeling of bond bending or torsional rotations is critical, such as in predicting the S^2 order parameters of NMR structures.

Methods

In this section we will show the derivations of the Hessian matrix from a $G\bar{o}$ -like potential proposed by Clementi et al [22].

The $G\bar{o}$ -like potential

The $G\bar{o}$ -like potential in [22] takes the non-native and native (equilibrium) conformations as input and it can be divided into four terms. The first term of this $G\bar{o}$ -like potential (defined as V_1 for later use) preserves the chain connectivity. The second (V_2) and third terms (V_3) define the bond angle and torsional interactions respectively and the last term (V_4) is the nonlocal interactions. The $G\bar{o}$ -like potential has the following expression:

$$\begin{aligned}
 V(X, X_0) &= \sum_{bonds} V_1(r, r_0) + \sum_{angles} V_2(\theta, \theta_0) \\
 &+ \sum_{dihedral} V_3(\phi, \phi_0) + \sum_{i < j - 3} V_4(r_{ij}, r_{0,ij}) \\
 &= \sum_{bonds} K_r (r - r_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 \\
 &+ \sum_{dihedral} \left\{ K_\phi^{(1)} [1 - \cos(\phi - \phi_0)] \right. \\
 &+ \left. K_\phi^{(3)} [1 - \cos 3(\phi - \phi_0)] \right\} \\
 &+ \sum_{i < j - 3} \varepsilon \left[5 \left(\frac{r_{0,ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{r_{0,ij}}{r_{ij}} \right)^{10} \right] \quad (11)
 \end{aligned}$$

In Eq. (11), r and r_0 represent respectively the instantaneous and equilibrium lengths of the virtual bonds between the C_α atoms of consecutive residues. Similarly, the θ (θ_0) and Φ (Φ_0) are respectively the instantaneous (equilibrium) virtual bond angles formed by three consecutive residues and the instantaneous (equilibrium) virtual dihedral angles formed by four consecutive residues. The r_{ij} and $r_{0,ij}$ represent respectively the instantaneous and equilibrium distances between two non-consecutive residues i and j . The $G\bar{o}$ -like potential in Eq. (11) includes several force parameters (K_r , K_θ , $K_\phi^{(1)}$, $K_\phi^{(3)}$ and ε) and the values of these parameters are taken directly from [22] without any tuning. The values of these parameters are: $K_r = 100\varepsilon$, $K_\theta = 20\varepsilon$, $K_\phi^{(1)} = \varepsilon$, $K_\phi^{(3)} = 0.5\varepsilon$ and $\varepsilon = 0.36$.

Anisotropic fluctuations from the second derivative of the $G\bar{o}$ -like potential

Similar to ANM, STeM has a $3N \times 3N$ Hessian matrix that can be decomposed into $N \times N$ super-elements. Each super-element in STeM, $\mathbf{H}_{i,j}$, is a summation of four 3×3 matrices. The first 3×3 matrix is the contribution from bond stretching. The second and third 3×3 matrices are the contributions from bond bending and torsional rotations respectively. The fourth 3×3 matrix is the contribution from nonlocal contacts.

$$\begin{aligned}
 \mathbf{H}_{i,j} &= \begin{bmatrix} \frac{\partial^2 V_1(r, r_0)}{\partial X_i \partial X_j} & \frac{\partial^2 V_1(r, r_0)}{\partial X_i \partial Y_j} & \frac{\partial^2 V_1(r, r_0)}{\partial X_i \partial Z_j} \\ \frac{\partial^2 V_1(r, r_0)}{\partial Y_i \partial X_j} & \frac{\partial^2 V_1(r, r_0)}{\partial Y_i \partial Y_j} & \frac{\partial^2 V_1(r, r_0)}{\partial Y_i \partial Z_j} \\ \frac{\partial^2 V_1(r, r_0)}{\partial Z_i \partial X_j} & \frac{\partial^2 V_1(r, r_0)}{\partial Z_i \partial Y_j} & \frac{\partial^2 V_1(r, r_0)}{\partial Z_i \partial Z_j} \end{bmatrix} + \\
 &\begin{bmatrix} \frac{\partial^2 V_2(\theta, \theta_0)}{\partial X_i \partial X_j} & \frac{\partial^2 V_2(\theta, \theta_0)}{\partial X_i \partial Y_j} & \frac{\partial^2 V_2(\theta, \theta_0)}{\partial X_i \partial Z_j} \\ \frac{\partial^2 V_2(\theta, \theta_0)}{\partial Y_i \partial X_j} & \frac{\partial^2 V_2(\theta, \theta_0)}{\partial Y_i \partial Y_j} & \frac{\partial^2 V_2(\theta, \theta_0)}{\partial Y_i \partial Z_j} \\ \frac{\partial^2 V_2(\theta, \theta_0)}{\partial Z_i \partial X_j} & \frac{\partial^2 V_2(\theta, \theta_0)}{\partial Z_i \partial Y_j} & \frac{\partial^2 V_2(\theta, \theta_0)}{\partial Z_i \partial Z_j} \end{bmatrix} + \\
 &\begin{bmatrix} \frac{\partial^2 V_3(\phi, \phi_0)}{\partial X_i \partial X_j} & \frac{\partial^2 V_3(\phi, \phi_0)}{\partial X_i \partial Y_j} & \frac{\partial^2 V_3(\phi, \phi_0)}{\partial X_i \partial Z_j} \\ \frac{\partial^2 V_3(\phi, \phi_0)}{\partial Y_i \partial X_j} & \frac{\partial^2 V_3(\phi, \phi_0)}{\partial Y_i \partial Y_j} & \frac{\partial^2 V_3(\phi, \phi_0)}{\partial Y_i \partial Z_j} \\ \frac{\partial^2 V_3(\phi, \phi_0)}{\partial Z_i \partial X_j} & \frac{\partial^2 V_3(\phi, \phi_0)}{\partial Z_i \partial Y_j} & \frac{\partial^2 V_3(\phi, \phi_0)}{\partial Z_i \partial Z_j} \end{bmatrix} + \\
 &\begin{bmatrix} \frac{\partial^2 V_4(r_{ij}, r_{0,ij})}{\partial X_i \partial X_j} & \frac{\partial^2 V_4(r_{ij}, r_{0,ij})}{\partial X_i \partial Y_j} & \frac{\partial^2 V_4(r_{ij}, r_{0,ij})}{\partial X_i \partial Z_j} \\ \frac{\partial^2 V_4(r_{ij}, r_{0,ij})}{\partial Y_i \partial X_j} & \frac{\partial^2 V_4(r_{ij}, r_{0,ij})}{\partial Y_i \partial Y_j} & \frac{\partial^2 V_4(r_{ij}, r_{0,ij})}{\partial Y_i \partial Z_j} \\ \frac{\partial^2 V_4(r_{ij}, r_{0,ij})}{\partial Z_i \partial X_j} & \frac{\partial^2 V_4(r_{ij}, r_{0,ij})}{\partial Z_i \partial Y_j} & \frac{\partial^2 V_4(r_{ij}, r_{0,ij})}{\partial Z_i \partial Z_j} \end{bmatrix} \quad (12)
 \end{aligned}$$

The Hessian matrix is the second derivative of the overall potential (equation 11). Let us first consider the first term of the $G\bar{o}$ -like potential and let (X_i, Y_i, Z_i) and (X_j, Y_j, Z_j) be the Cartesian coordinates of two consecutive residues i and j .

$$\begin{aligned}
 V_1(r, r_0) &= K_r (r - r_0)^2 \\
 &= K_r \left\{ \left[(X_j - X_i)^2 + (Y_j - Y_i)^2 + \right. \right. \\
 &\quad \left. \left. (Z_j - Z_i)^2 \right]^{1/2} - r_0 \right\}^2 \quad (13)
 \end{aligned}$$

The first and second partial derivatives of V_1 with respect to the X-direction of residue i are

$$\frac{\partial V_1}{\partial X_i} = -2K_r (X_j - X_i) (1 - r^0 / r) \quad (14)$$

$$\frac{\partial^2 V_1}{\partial X_i^2} = 2K_r (1 + r^0 (X_j - X_i)^2 / r^3 - r^0 / r) \quad (15)$$

We will get similar results for the Y- and Z-directions of residue i. Since we focus only on the equilibrium fluctuations, we can have $r \cong r^0$ at equilibrium and the first and second partial derivatives of V_1 can be further simplified to the following expressions.

$$\frac{\partial V_1}{\partial X_i} = 0 \quad (16)$$

$$\frac{\partial^2 V_1}{\partial X_i^2} = 2K_r (X_j - X_i)^2 / r^2 \quad (17)$$

In a similar way, the second cross-derivatives have the following form:

$$\frac{\partial^2 V_1}{\partial X_i \partial X_j} = -2K_r (X_j - X_i) (Y_j - Y_i) / r^2 \quad (18)$$

Equations 17 and 18 give the elements of the first 3x3 matrix of the super element \mathbf{H}_{ij} in equation 12. For the diagonal super elements \mathbf{H}_{ii} , equations 17 and 18 are substituted by the following:

$$\frac{\partial^2 V_1}{\partial X_i^2} = -\sum_j 2K_r (X_j - X_i)^2 / r^2 \quad (19)$$

$$\frac{\partial^2 V_1}{\partial X_i \partial Y_i} = \sum_j 2K_r (X_j - X_i) (Y_j - Y_i) / r^2 \quad (20)$$

Now let us consider the second term of the potential in Eq. (11) and let (X_i, Y_i, Z_i) , (X_j, Y_j, Z_j) and (X_k, Y_k, Z_k) be the Cartesian coordinates of three consecutive residues i, j and k. Suppose θ is the virtual bond angle formed by these three consecutive residues. Since the second term of the potential is $V_2 = K_\theta (\theta - \theta_0)^2$, the first and second partial derivatives of V_2 are

$$\frac{\partial V_2}{\partial X_i} = 2K_\theta (\theta - \theta_0) \frac{\partial \theta}{\partial X_i} \quad (21)$$

$$\frac{\partial^2 V_2}{\partial X_i^2} = 2K_\theta \left(\frac{\partial \theta}{\partial X_i} \right) + 2K_\theta (\theta - \theta_0) \frac{\partial^2 \theta}{\partial X_i^2} \quad (22)$$

Since θ equals θ_0 at equilibrium, $\frac{\partial^2 V_2}{\partial X_i^2}$ can be further simplified as

$$\frac{\partial^2 V_2}{\partial X_i^2} = 2K_\theta \left(\frac{\partial \theta}{\partial X_i} \right)^2 \quad (23)$$

Likewise, $\frac{\partial^2 V_2}{\partial X_i \partial X_j}$ becomes

$$\frac{\partial^2 V_2}{\partial X_i \partial X_j} = 2K_\theta \left(\frac{\partial \theta}{\partial X_i} \right) \left(\frac{\partial \theta}{\partial X_j} \right) \quad (24)$$

Let $\mathbf{p} = (X_i - X_p, Y_i - Y_p, Z_i - Z_p)$ and $\mathbf{q} = (X_k - X_p, Y_k - Y_p, Z_k - Z_p)$ and define G as the following.

$$G = \frac{(\mathbf{p} \cdot \mathbf{q})}{|\mathbf{p}| |\mathbf{q}|} \quad (25)$$

The θ can be expressed as

$$\theta = \cos^{-1} \left(\frac{(\mathbf{p} \cdot \mathbf{q})}{|\mathbf{p}| |\mathbf{q}|} \right) = \cos^{-1} (G) \quad (26)$$

The partial derivatives of θ are

$$\frac{\partial \theta}{\partial X_i} = \frac{-1}{\sqrt{1-G^2}} \frac{\partial G}{\partial X_i} \quad (27)$$

$$\frac{\partial \theta}{\partial X_j} = \frac{-1}{\sqrt{1-G^2}} \frac{\partial G}{\partial X_j} \quad (28)$$

$$\frac{\partial \theta}{\partial X_k} = \frac{-1}{\sqrt{1-G^2}} \frac{\partial G}{\partial X_k} \quad (29)$$

The derivative of G is

$$\frac{\partial G}{\partial X_j} = \frac{\partial (\mathbf{p} \cdot \mathbf{q})}{\partial X_j |\mathbf{p}| |\mathbf{q}|} = \frac{(X_k - X_j) |\mathbf{p}| |\mathbf{q}| - (\mathbf{p} \cdot \mathbf{q}) \frac{|\mathbf{q}|}{|\mathbf{p}|} (X_i - X_j)}{(|\mathbf{p}| |\mathbf{q}|)^2} \quad (30)$$

We can also get $\frac{\partial G}{\partial X_j}$ and $\frac{\partial G}{\partial X_k}$.

$$\frac{\partial G}{\partial X_j} = \frac{(2X_j - X_i - X_k) |\mathbf{p}| |\mathbf{q}| - (\mathbf{p} \cdot \mathbf{q}) \frac{|\mathbf{q}|}{|\mathbf{p}|} (X_j - X_i)}{(|\mathbf{p}| |\mathbf{q}|)^2} - \frac{(\mathbf{p} \cdot \mathbf{q}) \frac{|\mathbf{p}|}{|\mathbf{q}|} (X_j - X_k)}{(|\mathbf{p}| |\mathbf{q}|)^2} \quad (31)$$

$$\frac{\partial G}{\partial X_k} = \frac{(X_i - X_j) |\mathbf{p} \parallel \mathbf{q}| - (\mathbf{p} \cdot \mathbf{q}) \frac{|\mathbf{p}|}{|\mathbf{q}|} (X_k - X_j)}{(|\mathbf{p} \parallel \mathbf{q}|)^2} \quad (32)$$

Combined eq (23),(27) and (30), we can get the following formula.

$$\frac{\partial^2 V_2}{\partial X_i^2} = \frac{2K\theta}{1-G^2} \left(\frac{(X_k - X_j) |\mathbf{p} \parallel \mathbf{q}| - (\mathbf{p} \cdot \mathbf{q}) \frac{|\mathbf{q}|}{|\mathbf{p}|} (X_i - X_j)}{(|\mathbf{p} \parallel \mathbf{q}|)^2} \right)^2 \quad (33)$$

Similarly, Combined eq (24),(27), (28), (30) and (31), the second cross-derivative $\frac{\partial^2 V_2}{\partial X_i \partial X_j}$ becomes

$$\begin{aligned} \frac{\partial^2 V_2}{\partial X_i \partial X_j} = \frac{2K\theta}{1-G^2} & \left(\frac{(X_k - X_j) |\mathbf{p} \parallel \mathbf{q}| - (\mathbf{p} \cdot \mathbf{q}) \frac{|\mathbf{q}|}{|\mathbf{p}|} (X_i - X_j)}{(|\mathbf{p} \parallel \mathbf{q}|)^2} \right) \\ & \left(\frac{(2X_j - X_i - X_k) |\mathbf{p} \parallel \mathbf{q}| - (\mathbf{p} \cdot \mathbf{q}) \frac{|\mathbf{q}|}{|\mathbf{p}|} (X_j - X_i)}{(|\mathbf{p} \parallel \mathbf{q}|)^2} \right) \\ & - \frac{(\mathbf{p} \cdot \mathbf{q}) \frac{|\mathbf{p}|}{|\mathbf{q}|} (X_j - X_k)}{(|\mathbf{p} \parallel \mathbf{q}|)^2} \end{aligned} \quad (34)$$

Following a similar approach, we can get $\frac{\partial^2 V_2}{\partial X_j \partial X_k}$ and $\frac{\partial^2 V_2}{\partial X_k \partial X_i}$ and these second cross-derivatives form the elements of the second 3×3 matrix of the super element \mathbf{H}_{ij} in equation 12.

Due to the complexity of the derivation process of the Hessian matrix for the third (dihedral angle) term of the potential, we omit the derivation process here. The complete derivation is given in Additional File 1.

Finally, let's consider the final (non-local contact) term.

$$V_4 = \varepsilon \left[5 \left(\frac{r_{0,ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{r_{0,ij}}{r_{ij}} \right)^{10} \right] \quad (35)$$

A Taylor expansion will give us the following form.

$$V_4 = -\varepsilon + \frac{120\varepsilon}{r_{0,ij}^2} (r_{ij} - r_{0,ij})^2 \quad (36)$$

Equation 36 has the same harmonic form as the first term but with a different force constant, so the derivation process is the same as the first term. Therefore, we give only the derivation result here.

$$\frac{\partial^2 V_4}{\partial X_i \partial Y_j} = -\frac{240\varepsilon}{r_{0,ij}^2} (X_j - X_i)(Y_j - Y_i) / r_{ij}^2 \quad (37)$$

After combining the Hessian matrices from all four terms, we can calculate the pseudo inverse of the final Hessian matrix \mathbf{H} . The mean square displacement $\langle \Delta \mathbf{r}_i^2 \rangle$ and inter residue correlation $\langle \Delta \mathbf{r}_i \cdot \Delta \mathbf{r}_j \rangle$ can be calculated by summing the elements over the X, Y and Z directions.

$$\langle \Delta r_i^2 \rangle = \frac{k_B T}{\gamma} (H_{3i-2,3i-2}^+ + H_{3i-1,3i-1}^+ + H_{3i,3i}^+) \quad (38)$$

$$\langle \Delta r_i \cdot \Delta r_j \rangle = \frac{k_B T}{\gamma} (H_{3i-2,3j-2}^+ + H_{3i-1,3j-1}^+ + H_{3i,3j}^+) \quad (39)$$

The protein sets studied

To evaluate the STeM model, we apply it to compute B-factors and to study protein conformation changes and compare the results with those computed from ANM and GNM. For B-factors computations, the protein dataset is from [27] and contains 111 proteins. Two proteins, 1CYO and 5PTP, are removed from the dataset because they no longer exist in the current Protein Data Bank [28]. The proteins in the first dataset all have a resolution that is better than or equal to 2.0 Å. For conformation change studies, the dataset is from [20], which contains 20 pairs of protein structures. Each pair of protein structures have significantly large structure difference from each other.

Evaluation techniques

We used the same evaluation techniques as have been applied before [20,27]. Specifically, the following three numerical measures are used.

The correlation between the experimental and calculated B-factors

The linear correlation coefficient between the experimental and calculated B-factors is calculated using the following formula.

$$\rho = \frac{\sum_i^N (x_i - \bar{x})(y_i - \bar{y})}{\left[\sum_i^N (x_i - \bar{x})^2 \sum_i^N (y_i - \bar{y})^2 \right]^{1/2}} \quad (40)$$

where x_i and y_i are respectively the experimental and calculated B-factors of the C_α atom of residue i and \bar{x} and \bar{y} are the mean values. N is the number of residues.

The overlap between the experimental observed conformation changes and the calculated modes

The overlap measures the directional similarity between a conformation change and a calculated mode. The formula for calculating the overlap is

$$I = \frac{\left| \sum_i^{3N} e_i r_i \right|}{\left[\sum_i^{3N} e_i^2 \sum_i^{3N} r_i^2 \right]^{1/2}} \quad (41)$$

where e_i is the relative displacement of residue i in a selected mode e and r_i is the conformation displacement of residue i .

The correlation between the experimental observed conformation changes and the calculated modes

The correlation measures the magnitude similarity between a conformation change and a calculated mode. The formula used for calculating the correlation is the same as equation (40), with different meaning for x_i and y_i .

$$\rho = \frac{\sum_i^N (x_i - \bar{x})(y_i - \bar{y})}{\left[\sum_i^N (x_i - \bar{x})^2 \sum_i^N (y_i - \bar{y})^2 \right]^{1/2}} \quad (42)$$

where x_i is the magnitude of the displacement of residue i in the conformation change and y_i is the magnitude of the displacement of residue i in the selected mode. \bar{x} and \bar{y} are the corresponding mean values.

Additional file 1: The derivation of the Hessian matrix for the dihedral angle term (V3)

List of abbreviations used

ENM: Elastic Network Model; GNM: Gaussian Network Model; ANM: Anisotropic Network Model; StEM: Spring Tensor Model

Acknowledgements

This article has been published as part of *BMC Structural Biology* Volume 10 Supplement 1, 2010: Selected articles from the Computational Structural Bioinformatics Workshop 2009. The full contents of the supplement are available online at <http://www.biomedcentral.com/1472-6807/10?issue=S1>.

Author details

¹Computer Science Department, Iowa State University, 226 Atanasoff Hall, Ames, IA 50011, USA. ²L H Baker Center for Bioinformatics and Biological Statistics, Iowa State University, Ames, IA 50011, USA. ³Program of Bioinformatics and Computational Biology, Iowa State University, Ames, IA 50011, USA.

Authors' contributions

Tu-Liang Lin collaborated with Guang Song on the ideas of the Spring Tensor Model and applied the model to the predictions of the B-factors and conformation changes. Guang Song conceived the idea of using a more sophisticated potential and suggested the G \ddot{o} -like potential. Most of the implementation was done by Tu-Liang Lin under the supervision of Guang Song. Both authors read and approved the manuscript.

Competing interests

The authors declare that they have no competing interests.

Published: 17 May 2010

References

1. Voth GA: Coarse-graining of condensed phase and biomolecular systems. CRC Press 2009.
2. Bahar I, Atilgan AR, Erman B: Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des* 1997, **2**:173-181.
3. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I: Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 2001, **80**:505-515.
4. Bahar I, Rader A: Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* 2005, **15**:586-592.
5. Ma J: Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure* 2005, **13**:373-380.
6. Ming D, Br \ddot{u} schweiler R: Reorientational contact-weighted elastic network model for the prediction of protein dynamics: Comparison with NMR relaxation. *Biophys. J.* 2006, **90**:3382-3388.
7. Song G, Jernigan RL: An enhanced elastic network model to represent the motions of domain-swapped proteins. *Proteins*. 2006, **63**:197-209.
8. Song G, Jernigan RL: vGNM: a better model for understanding the dynamics of proteins in crystals. *J. Mol. Biol.* 2007, **369**:880-893.
9. Lu M, Poon B, Ma J: A new method for coarse-grained elastic normal-mode analysis. *J. Chem. Theory Comput* 2006, **2**:464-471.
10. Yang L, Song G, Jernigan RL: Protein elastic network models and the ranges of cooperativity. *Proc Natl Acad Sci U S A* 2009, **106**:12347-12352.
11. Zheng W: A unification of the elastic network model and the Gaussian network model for optimal description of protein conformational motions and fluctuations. *Biophys. J.* 2008, **94**:3853-3857.
12. Stember JN, Wriggers W: Bend-twist-stretch model for coarse elastic network simulation of biomolecular motion. *J. Chem. Phys.* 2009, **131**:074112.
13. Kundu S, Sorensen DC, Phillips GN Jr: Automatic domain decomposition of proteins by a Gaussian network model. *Proteins* 2007, **57**:725-733.
14. Lin T, Song G: Predicting allosteric communication pathways using motion correlation network. *Proceedings of the 7th Asia Pacific Bioinformatics Conference (APBC)* Tsinghua University, China 2009, 588-598.
15. Bahar I, Chennubhotla I, Tobi D: Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr. Opin. Struct. Biol.* 2007, **17**:633-640.
16. Zheng W, Brooks B: Identification of dynamical correlations within the myosin motor domain by the normal mode analysis of an elastic network model. *J. Mol. Biol.* 2005, **346**:745-759.
17. Yang Z, M \acute{a} jek P, Bahar I: Allosteric transitions of supramolecular systems explored by network models: Application to chaperonin GroEL. *PLoS Comput. Biol.* 2009, **5**:e1000360.
18. Tirion MM: Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 1996, **77**:1905-1908.
19. Hinsen K: Analysis of domain motions by approximate normal mode calculations. *Proteins*. 1998, **33**:417-429.
20. Tama F, Sanejouand Y: Conformational change of proteins arising from normal mode calculations. *Protein Eng.* 2001, **14**:1-6.
21. Thorpe MF: Comment on elastic network models and proteins. *Phys. Biol.* 2007, **4**:60-63.
22. Clementi C, Nymeyer H, Onuchic JN: Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* 2000, **298**:937-953.
23. Koga N, Takada S: Roles of native topology and chain-length scaling in protein folding: a simulation study with a G \ddot{o} -like model. *J. Mol. Biol.* 2001, **313**:171-180.
24. Koga N, Takada S: Folding-based molecular simulations reveal mechanisms of the rotary motor F1-ATPase. *Proc. Natl. Acad. Sci. USA* 2006, **103**:5367-5372.
25. Riccardi D, Cui Q, Phillips GN: Application of elastic network models to proteins in the crystalline state. *Biophys. J.* 2009, **96**:464-475.

26. Hinsen K, Petrescu AJ, Dellerue S, Bellissent-Funel MC, Kneller GR: **Harmonicity in slow protein dynamics.** *Chem. Phys.* 2000, **261**:25-37.
27. Kundu S, Melton JS, Sorensen DC, Phillips GN Jr: **Dynamics of proteins in crystals: Comparison of experiment with simple models.** *Biophys. J.* 2002, **83**:723-732.
28. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The Protein Data Bank.** *Nucleic Acids Res.* 2000, **28**:235-242.

doi:10.1186/1472-6807-10-S1-S3

Cite this article as: Lin and Song: Generalized spring tensor models for protein fluctuation dynamics and conformation changes. *BMC Structural Biology* 2010 **10**(Suppl 1):S3.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

