Chapter 5
Generalized Spring Tensor Models for Protein Fluctuation Dynamics and Conformation Changes

Hyuntae Na, Tu-Liang Lin, and Guang Song

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 cannot 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 book chapter, we present 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

H. Na (✉)
Department of Computer Science, Iowa State University, 226 Atanasoff Hall, Ames, IA 50011, USA
e-mail: htna@iastate.edu

T.-L. Lin (✉)
Department of Management Information Systems, National Chiayi University, 580 Sinmin Rd., Chiayi City 600, Taiwan
e-mail: tuliang@mail.ncyu.edu.tw

G. Song (✉)
Department of Computer Science, Iowa State University, 226 Atanasoff Hall, Ames, IA 50011, USA

L.H. Baker Center for Bioinformatics and Biological Statistics, Iowa State University, 226 Atanasoff Hall, Ames, IA 50011, USA

Program of Bioinformatics and Computational Biology, Iowa State University, 226 Atanasoff Hall, Ames, IA 50011, USA
e-mail: gsong@iastate.edu

K. Han et al. (eds.), *Protein Conformational Dynamics*, Advances in Experimental Medicine and Biology 805, DOI 10.1007/978-3-319-02970-2_5,

ANM. This is achieved by employing a physically more realistic potential, the Gō-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. In addition, we show that STeM can be further extended to an all-atom model and protein fluctuation dynamics computed by all-atom STeM matches closely with that by Normal Mode Analysis (NMA).

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 Gō potential to the fluctuation dynamics, STeM reveals, (i) a physical explanation for why the distance-dependent, inverse distance square (i.e., $1/r^2$) spring constants perform better than the uniform ones, and (ii), the importance of three-body and four-body interactions to properly modeling protein dynamics.

STeM is not limited to coarse-grained protein models that use a single bead, usually the alpha carbon, to represent each residue. The core idea of STeM, deriving the Hessian matrix directly from a physically realistic potential, can be extended to all-atom models as well. We did this and discovered that all-atom STeM model represents a highly close approximation of NMA, yet without the need for energy minimization.

Keywords Normal mode analysis • Hessian matrix • Spring tensor model • Protein dynamics • Mean-square fluctuations

Abbreviations

ENM	Elastic Network Model
GNM	Gaussian Network Model
ANM	Anisotropic Network Model
STeM	Spring Tensor Model
NMA	Normal Model Analysis
ANMr2	ANM using $1/r^2$ as spring constant

5.1 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 function-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 (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 [2, 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 cannot 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.

5.1.1 Gaussian Network Model

Gaussian Network Model (GNM) was first introduced in [3] 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 $\mathbf{\Gamma}$ (see Eq. (5.1)). 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 7–8 Å. Let $\Delta \boldsymbol{r}_i$ and $\Delta \boldsymbol{r}_j$ 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 $\mathbf{\Gamma}$ is:

$$\mathbf{\Gamma}_{ij} = \begin{cases} -1 & \text{if } i \neq j \cap r_{0,ij} \leq r_c \\ 0 & \text{if } i \neq j \cap r_{0,ij} > r_c \\ \sum_{j,j\neq i}^{N} \mathbf{\Gamma}_{ij} & \text{if } i = j \end{cases} \tag{5.1}$$

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, $\langle \Delta \boldsymbol{r}_i \cdot \Delta \boldsymbol{r}_j \rangle$, can be easily obtained from the inverse of the Kirchhoff matrix:

$$\left\langle \Delta \boldsymbol{r}_i^2 \right\rangle = \frac{3k_B T}{\gamma} \left(\mathbf{\Gamma}^{-1}\right)_{ii}, \tag{5.2}$$

$$\left\langle \Delta \boldsymbol{r}_i \cdot \Delta \boldsymbol{r}_j \right\rangle = \frac{3k_B T}{\gamma} \left(\mathbf{\Gamma}^{-1}\right)_{ii}, \tag{5.3}$$

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.

5.1.2 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 [4, 19, 20]. One of such models is best known as Anisotropic Network Model (ANM) [4] 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 [4]. The $3N \times 3N$ Hessian matrix H_{ANM} can be repartitioned into $N \times N$ super elements and each super element is a 3×3 tensor.

$$\boldsymbol{H}_{\mathrm{ANM}} = \begin{bmatrix} \boldsymbol{H}_{1,1} & \boldsymbol{H}_{1,2} & \dots & \boldsymbol{H}_{1,N} \\ \boldsymbol{H}_{2,1} & \boldsymbol{H}_{2,2} & \dots & \boldsymbol{H}_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{H}_{N,1} & \boldsymbol{H}_{N,2} & \dots & \boldsymbol{H}_{N,N} \end{bmatrix} \tag{5.4}$$

where $\boldsymbol{H}_{i,j}$ is the interaction tensor between residues i and j and can be expressed as:

$$\boldsymbol{H}_{i,j} = \begin{bmatrix} \frac{\partial^2 V}{\partial X_i \partial X_j} & \frac{\partial^2 V}{\partial X_i \partial Y_j} & \frac{\partial^2 V}{\partial X_i \partial Z_j} \\ \frac{\partial^2 V}{\partial Y_i \partial X_j} & \frac{\partial^2 V}{\partial Y_i \partial Y_j} & \frac{\partial^2 V}{\partial Y_i \partial Z_j} \\ \frac{\partial^2 V}{\partial Z_i \partial X_j} & \frac{\partial^2 V}{\partial Z_i \partial Y_j} & \frac{\partial^2 V}{\partial Z_i \partial Z_j} \end{bmatrix} \tag{5.5}$$

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

$$\left\langle \Delta \mathbf{r}_i^2 \right\rangle = \frac{3k_B T}{\gamma} \left(\boldsymbol{H}^{+}_{3i-2,3i-2} + \boldsymbol{H}^{+}_{3i-1,3i-1} + \boldsymbol{H}^{+}_{3i,3i} \right) \tag{5.6}$$

$$\left\langle \Delta \mathbf{r}_i \cdot \Delta \mathbf{r}_j \right\rangle = \frac{3k_B T}{\gamma} \left(\boldsymbol{H}^{+}_{3i-2,3j-2} + \boldsymbol{H}^{+}_{3i-1,3j-1} + \boldsymbol{H}^{+}_{3i,3j} \right) \tag{5.7}$$

5.1.3 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_a'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 $\boldsymbol{H}_{i,j}^{\mathrm{ANM}}$ between residues i and j in Eq. (5.5) becomes the following in the local frame (where the Z axis is along the direction from residues i to j):

$$\boldsymbol{H}_{i,j}^{\mathrm{ANM}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5.8}$$

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:

$$\boldsymbol{H}_{i,j}^{\mathrm{GNM}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5.9}$$

From the two tensors $\boldsymbol{H}_{i,j}^{\mathrm{ANM}}$ and $\boldsymbol{H}_{i,j}^{\mathrm{GNM}}$ given in Eqs. (5.8) and (5.9), 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 $\boldsymbol{H}_{i,j}^{\mathrm{ANM}}$ 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:

$$\boldsymbol{H}_{i,j}^{\mathrm{STeM}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \mathrm{T}(i,j) \end{bmatrix} + \sum_{k} \begin{bmatrix} \mathrm{s}_{11}(i,j,k) & \mathrm{s}_{12}(i,j,k) & \mathrm{s}_{13}(i,j,k) \\ \mathrm{s}_{21}(i,j,k) & \mathrm{s}_{22}(i,j,k) & \mathrm{s}_{23}(i,j,k) \\ \mathrm{s}_{31}(i,j,k) & \mathrm{s}_{32}(i,j,k) & \mathrm{s}_{33}(i,j,k) \end{bmatrix} \tag{5.10}$$

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 Eq. (5.10)

represents the two-body interaction, which is similar to $\boldsymbol{H}_{i,j}^{\mathrm{ANM}}$, except that the interaction strength T(i,j) depends on residues i and j, and thus may be distance-dependent as well.

5.1.4 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. (5.8), but takes a generalized tensor form (similar to that in Eq. (5.10)) that can provide both longitudinal and transversal constraints on a residue's fluctuations relative to its neighbors. We obtain the generalized tensor form by deriving the Hessian matrix from a physically more realistic Gō-like potential (Eq. (5.11)), which has been successfully used in many MD simulations to study protein folding processes and conformation changes [22–24]. In additional 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. (5.10)).

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.

STeM is not limited to coarse-grained protein models that use a single bead, usually the alpha carbon, to represent each residue. The core idea of STeM, deriving the Hessian matrix directly from a physically realistic potential, can be extended to all-atom models as well. We did this and discovered that all-atom STeM model represents a closer approximation of NMA than most other models.

5.2 Results and Discussion

5.2.1 Crystallographic B-Factor Prediction

Table 5.1 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 5.1 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 Fig. 5.2 shows that STeM performs better than ANM for 80 % of the proteins in the dataset.

Table 5.1 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
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

(continued)

Table 5.1 (continued)

Protein	R(Å)	ANM	GNM	STeM	Protein	R(Å)	ANM	GNM	STeM
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

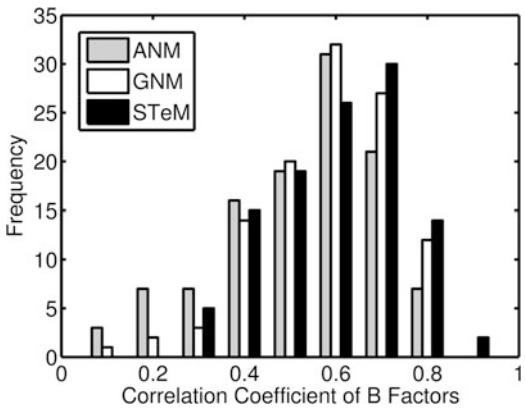

Fig. 5.1 The distributions of the correlation coefficients between the experimental and calculated B-factors

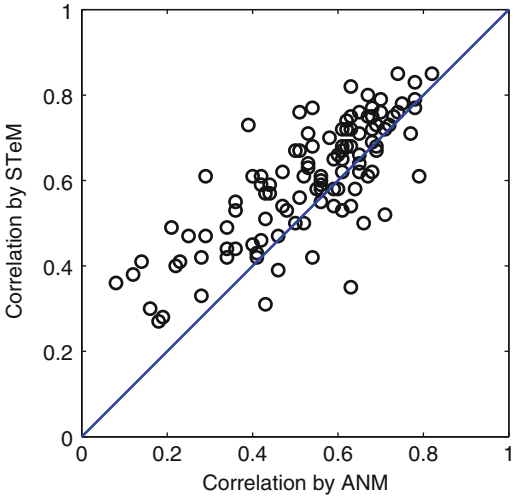

Fig. 5.2 The scatter plot of the correlation coefficients by ANM and those by STeM. For 80 % of the proteins listed in Table 5.1, STeM does better than ANM

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.

5.2.2 *The Contributions of Different Interaction Terms to the Fluctuations*

The Gō-like potential in Eq. (5.11) 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 $\boldsymbol{H}_{\mathrm{ANM}}$, is used as another baseline for comparison purposes. Table 5.2 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 $\boldsymbol{H}_{V_4}$ and $\boldsymbol{H}_{\mathrm{ANM}}$, 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

Table 5.2 The contributions of different interaction terms to the agreement with experimental B-factors H_{ANM}

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

$\boldsymbol{H}_{\mathrm{ANM}}$ is the Hessian matrix of ANM. $\boldsymbol{H}_{V_1}$, $\boldsymbol{H}_{V_2}$, $\boldsymbol{H}_{V_3}$, and $\boldsymbol{H}_{V_4}$ 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

with its spatial neighbors, most of which are through non-bonded interactions. What is more interesting is that $\boldsymbol{H}_{V_4}$ term alone performs better than $\boldsymbol{H}_{\mathrm{ANM}}$. This is in agreement with recent results that the performance of 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 [16]. The Taylor expansion of the non-bonded interaction term (V_4) shows that it has an equivalent spring constant of the form $\frac{120\varepsilon}{r_{0,ij}^2}$ (see Eq. (5.36)), 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 [3]. 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.

5.2.3 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

Table 5.3 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)

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. The open forms are used to calculate the normal modes. Table 5.3 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 5.3), which may imply the modes of STeM be of higher quality than those of ANM.

5.2.4 *Protein Fluctuation Dynamics Predicted by All-Atom STeM Matches Closely with That of NMA*

In this section, we apply all-atom STeM model to a large number of proteins and show that the fluctuation dynamics produced by STeM matches closely with that of NMA. To avoid the uncertainties existing in experimental B-factors due to crystal packing and disorder, the atomic fluctuations computed from NMA and STeM are compared with each other and not with the experimental B-factors.

To compute the fluctuations, all the structures are first energetically minimized using the Tinker program [27] with the Charmm22 force field. The minimized structures are then used by NMA, STeM, and later on, by ANM and ANMr2 models, to compute the mean-square fluctuations. Some of the force field parameters from Charmm22 are used in computing the STeM Hessian matrix. Let $\boldsymbol{M}$ be the $N \times N$ diagonal mass matrix, I be the 3×3 identity matrix, and $\otimes$ be the operator of the Kronecker product. Let $\boldsymbol{b}^{\mathrm{NMA}}$ and $\boldsymbol{b}^{\mathrm{STeM}}$ denote the mean-square fluctuations from NMA and STeM, respectively. The following procedure details how they are determined:

1. Use Tinker [27] to determine the minimized conformation C whose potential energy as defined by Charmm22 is fully minimized;
2. Obtain $\boldsymbol{b}^{\mathrm{NMA}}$ using Tinker;
3. Compute the Hessian matrix $\boldsymbol{H}^{\mathrm{STeM}}$ of C using STeM;
4. Determine frequencies f_i and modes $\boldsymbol{m}_i$ of $\boldsymbol{H}^{\mathrm{STeM}}$ in the mass-weighted Cartesian coordinate as follows, where $i = 7, 8, \ldots, 3N$:
   (a) $\tilde{\boldsymbol{H}}^{\mathrm{STeM}} \leftarrow \left(\boldsymbol{M}^{1/2} \otimes I\right)^{-1} \boldsymbol{H}^{\mathrm{STeM}} \left(\boldsymbol{M}^{1/2} \otimes I\right)^{-1}$;
   (b) $\langle f_i, \tilde{\boldsymbol{m}}_i \rangle \leftarrow i$th eigenvalue and eigenvector of $\tilde{\boldsymbol{H}}^{\mathrm{STeM}}$;
   (c) $\boldsymbol{m}_i \leftarrow \left(\boldsymbol{M}^{1/2} \otimes I\right)^{-1} \tilde{\boldsymbol{m}}_i$;
5. Compute the B-factor $\boldsymbol{b}^{\mathrm{STeM}}$ using f_i and $\boldsymbol{m}_i$;
6. Compute the correlation between $\boldsymbol{b}^{\mathrm{NMA}}$ and $\boldsymbol{b}^{\mathrm{STeM}}$.

The procedure is repeated for a dataset of 306 proteins.

5.2.4.1 STeM Outperforms ANM in Matching with NMA

Figure 5.3 compares the correlations between computed B-factors: $\boldsymbol{b}^{\mathrm{STeM}}$, $\boldsymbol{b}^{\mathrm{ANM}}$, or $\boldsymbol{b}^{\mathrm{ANMr2}}$, with $\boldsymbol{b}^{\mathrm{NMA}}$. 306 proteins, listed in Table 5.4, are used to compute these correlations. Denote corr($\boldsymbol{a}$,$\boldsymbol{b}$) by the correlation between two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. Figure 5.3a shows the scatter plot of corr($\boldsymbol{b}^{\mathrm{NMA}}$,$\boldsymbol{b}^{\mathrm{ANM}}$) and corr($\boldsymbol{b}^{\mathrm{NMA}}$,$\boldsymbol{b}^{\mathrm{STeM}}$),

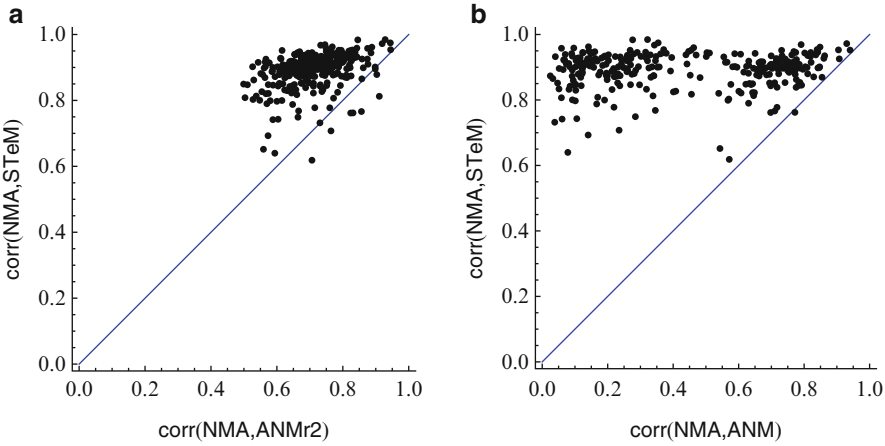

Fig. 5.3 The scatter plots of the correlation coefficients with NMA by different all-atom models. (**a**) the scatter plot of corr($\boldsymbol{b}^{\mathrm{NMA}}$,$\boldsymbol{b}^{\mathrm{ANM}}$) and corr($\boldsymbol{b}^{\mathrm{NMA}}$,$\boldsymbol{b}^{\mathrm{STeM}}$), and (**b**) the scatter plot of corr($\boldsymbol{b}^{\mathrm{NMA}}$,$\boldsymbol{b}^{\mathrm{ANMr2}}$) and corr($\boldsymbol{b}^{\mathrm{NMA}}$,$\boldsymbol{b}^{\mathrm{STeM}}$), where $\boldsymbol{b}$ denotes the mean-square fluctuations computed by the different models. 306 proteins as listed in Table 5.4 are used for the computation

Table 5.4 List of proteins used in the all-atom STeM study

1BKR	2A28	2R8U	3B0F	3JQU	3NAR	3R87	3TME	4BBD
1C5E	2AAJ	2RK5	3B7H	3JTN	3NBC	3RDM	3TOE	4D8D
1DBF	2BT9	2RKL	3BD1	3K6F	3NGP	3RDO	3TP5	4DCZ
1G2B	2CKK	2VE8	3BRI	3K6T	3NJK	3RDS	3TPX	4DRO
1G2R	2F5K	2VKL	3BRL	3KBL	3NJM	3RE6	3TSI	4DRP
1GK7	2FE5	2VZC	3BZY	3KIK	3NS6	3RGR	3TWE	4E34
1GU1	2FL4	2WJ5	3CCD	3KJL	3NTW	3RHB	3TXQ	4E35
1HG7	2GBJ	2WPU	3CNK	3KNG	3NXA	3RJS	3TXS	4E6I
1I2T	2GBN	2WQ0	3CPO	3KOV	3O48	3RKV	3U1C	4E6S
1IHR	2HIN	2X3D	3CX2	3KXY	3O5Z	3RL8	3U80	4E8O
1J8Q	2HO2	2X48	3D4W	3L1F	3OBL	3RNJ	3UD8	4EDL
1J9E	2I5C	2X5H	3DS4	3L1X	3OJB	3RNV	3UJ3	4EDM
1JCD	2IC6	2X5T	3E2B	3L7H	3OMT	3RQ9	3VA9	4ERR
1JO0	2IGD	2XDH	3E56	3LKY	3OV4	3RSW	3VBG	4ES3
1MFG	2IZX	2XEM	3FG7	3LLB	3P38	3RY2	3VEJ	4EWI
1MG4	2J5Y	2XF6	3FX7	3LLO	3P6J	3RZW	3VGN	4EZA
1MK5	2JDC	2XF7	3FZW	3LNQ	3P7J	3S02	3VI6	4F26
1MM9	2JDD	2XG3	3G21	3LNW	3PA7	3SEI	3VMX	4F55
1N9M	2JKU	2XRH	3G9R	3LRD	3PE9	3SFM	3VMY	4F8A
1NWW	2LIS	2XUS	3GZ2	3LRG	3PO8	3SGP	3ZR8	4FQN
1OOT	2O31	2XW6	3H00	3M0R	3PYJ	3SGR	3ZSK	4FYH

(continued)

Table 5.4 (continued)

1R29	2O37	2XX6	3HA4	3M0U	3Q47	3SHU	3ZSL	4GCN
1R6J	2O9V	2Y2T	3HFO	3M8J	3Q9Q	3SK4	3ZW2	4GCO
1T6F	2OEI	2Y3E	3HGM	3M9H	3QF3	3SK6	3ZZL	4GMQ
1TG0	2ON8	2Y3F	3HSH	3M9J	3QGL	3SK8	3ZZQ	4GS3
1U07	2ONQ	2Y4X	3HTU	3MAB	3QMQ	3SQF	4A1H	4GSW
1URR	2OVG	2Y9F	3I4O	3MBT	3QMX	3SSQ	4A6S	4HBX
1W53	2PMR	2Y9G	3ID1	3MCB	3QWG	3SSU	4A75	4HE6
1WM3	2PWO	2Y9R	3ID2	3MCE	3QWS	3SWY	4A9F	4HK2
1WYX	2PZV	2YEL	3ID3	3MHE	3R27	3T6F	4ABM	4HTH
1Y0M	2QCB	2YH5	3ID4	3MP9	3R3M	3T6L	4AEQ	4HTI
1YO3	2QCP	2YIZ	3IG9	3MSH	3R45	3T8N	4AGH	4HTJ
1Z96	2QJL	2ZWM	3IGE	3N27	3R69	3T8U	4B27	4HX8
2A26	2R6Q	3AXC	3IPT	3N4W	3R85	3TDM	4B6X	4IOG

while (b) the scatter plot of corr($\boldsymbol{b}^{\mathrm{NMA}}$,$\boldsymbol{b}^{\mathrm{ANMr2}}$) and corr($\mathbf{b}^{\mathrm{NMA}}$,$\mathbf{b}^{\mathrm{STeM}}$), respectively. The average correlation over all proteins is 0.44 for corr($\boldsymbol{b}^{\mathrm{NMA}}$,$\boldsymbol{b}^{\mathrm{ANM}}$), 0.71 for corr($\boldsymbol{b}^{\mathrm{NMA}}$,$\boldsymbol{b}^{\mathrm{ANMr2}}$), and 0.89 for corr($\boldsymbol{b}^{\mathrm{NMA}}$,$\boldsymbol{b}^{\mathrm{STeM}}$). STeM clearly outperforms both ANMr2 and ANM in matching with NMA, having a high average correlation in mean-square fluctuations with those of NMA. The results thus underscore the importance of including multi-body interactions for a finer portrait of protein fluctuation dynamics.

5.3 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 Gō-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 Gō 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 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. In [28] we noted that the 3-body and 4-body potentials introduced through bond bending and torsional interactions in the coarse-grained STeM model only scratched the surface of the full extensity of the multi-body potentials. Indeed, results from all-atom STeM where the multi-body interactions are most accurately represented demonstrate that all-atom STeM has reached an even higher correlation with NMA in predicting mean-square fluctuations, yet without the need for energy minimization.

Finally, since STeM takes into account bond bending and 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 S^2 order parameters of NMR structures.

5.4 Methods

In this section we present the derivations of the Hessian matrix for a coarse-grained model from a Gō-like potential [23]. The derivations are mostly the same as what appeared in [28]. In addition, we show how the core idea of STeM can be extended to derive the STeM Hessian matrix for an all-atom model using an all-atom potential.

5.4.1 The Gō-Like Potential

The Gō-like potential in [23] takes the non-native and native (equilibrium) conformations as input and it can be divided into four terms. The first term of this Gō-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 Gō-like potential has the following expression:

$$\begin{aligned} V(X, X_0) &= \sum_{\text{bonds}} V_1(r, r_0) + \sum_{\text{angle}} V_2(\theta, \theta_0) + \sum_{\text{dihedral}} V_3(\phi, \phi_0) + \sum_{i<j-3} V_4\left(r_{ij}, r_{0,ij}\right) \\ &= \sum_{\text{bonds}} K_r (r - r_0)^2 + \sum_{\text{angle}} K_\theta (\theta - \theta_0)^2 \\ &\quad + \sum_{\text{dihedral}} \left\{ K_\phi^{(1)} \left[1 - \cos(\phi - \phi_0)\right] + K_\phi^{(3)} \left[1 - \cos 3(\phi - \phi_0)\right] \right\} \\ &\quad + \sum_{i<j-3} \varepsilon \left[5 \left(\frac{r_{0,ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{r_{0,ij}}{r_{ij}} \right)^{10} \right] \end{aligned} \tag{5.11}$$

In Eq. (5.11), 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 Gō-like potential in Eq. (5.11) includes several force parameters ($K_r, K_\theta, K_\phi^{(1)}, K_\phi^{(3)}$ *and* ε) and the values of these parameters are taken directly from [23] 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* $\varepsilon = 0.36$.

5.4.2 Anisotropic Fluctuations from the Second Derivative of the Gō-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, $\boldsymbol{H}_{i,j}$, 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.

$$\boldsymbol{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 Z_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} \tag{5.12}$$

The Hessian matrix is the second derivative of the overall potential (Eq. (5.11)). Let us first consider the first term of the Gō-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.

$$\begin{aligned} V_1(r, r_0) &= K_r (r - r_0)^2 \\ &= K_r \left\{ \left[\left(X_j - X_i\right)^2 + \left(Y_j - Y_i\right)^2 + \left(Z_j - Z_i\right)^2 \right]^{1/2} - r_0 \right\}^2 \end{aligned} \tag{5.13}$$

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

$$\frac{\partial V_1}{\partial X_i} = -2K_r \left(X_j - X_i\right)\left(1 - r^0/r\right) \tag{5.14}$$

$$\frac{\partial^2 V_1}{\partial X_i^2} = 2K_r \left(1 + r^0 \left(X_j - X_i\right)^2 / r^3 - r^0/r\right) \tag{5.15}$$

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.

$$\frac{\partial V_1}{\partial X_i} = 0 \tag{5.16}$$

$$\frac{\partial^2 V_1}{\partial X_i^2} = 2K_r\left(X_j - X_i\right)^2/r^2 \tag{5.17}$$

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

$$\frac{\partial^2 V_1}{\partial X_i \partial Y_j} = -2K_r\left(X_j - X_i\right)\left(Y_j - Y_i\right)/r^2 \tag{5.18}$$

Equations (5.17) and (5.18) give the elements of the first 3×3 matrix of the super element $\boldsymbol{H}_{ij}$ in Eq. (5.6). For the diagonal super elements $\boldsymbol{H}_{ii}$, Eqs. (5.17) and (5.18) are substituted by the following:

$$\frac{\partial^2 V_1}{\partial X_i^2} = -\sum_j 2K_r\left(X_j - X_i\right)^2/r^2 \tag{5.19}$$

$$\frac{\partial^2 V_1}{\partial X_i \partial Y_i} = \sum_j 2K_r\left(X_j - X_i\right)\left(Y_j - Y_i\right)/r^2 \tag{5.20}$$

Now let us consider the second term of the potential in Eq. (5.11) 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

$$\frac{\partial V_2}{\partial X_i} = 2K_\theta\left(\theta - \theta_0\right)\frac{\partial \theta}{\partial X_i} \tag{5.21}$$

$$\frac{\partial^2 V_2}{\partial X_i^2} = 2K_\theta\left(\frac{\partial \theta}{\partial X_i}\right)^2 + 2K_\theta\left(\theta - \theta_0\right)\frac{\partial^2 \theta}{\partial X_i^2} \tag{5.22}$$

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

$$\frac{\partial^2 V_2}{\partial X_i^2} = 2K_\theta\left(\frac{\partial \theta}{\partial X_i}\right)^2 \tag{5.23}$$

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

$$\frac{\partial^2 V_2}{\partial X_i \partial X_j} = 2K_\theta \left(\frac{\partial \theta}{\partial X_i}\right)\left(\frac{\partial \theta}{\partial X_j}\right) \tag{5.24}$$

Let $\boldsymbol{p} = (X_i - X_j, Y_i - Y_j, Z_i - Z_j)$ and $\boldsymbol{q} = (X_k - X_j, Y_k - Y_j, Z_k - Z_j)$ and define G as the following.

$$G = \frac{(\boldsymbol{p} \cdot \boldsymbol{q})}{|\boldsymbol{p}|\,|\boldsymbol{q}|} \tag{5.25}$$

The θ can be expressed as

$$\theta = \cos^{-1}\left(\frac{(\boldsymbol{p} \cdot \boldsymbol{q})}{|\boldsymbol{p}|\,|\boldsymbol{q}|}\right) = \cos^{-1}(G) \tag{5.26}$$

The partial derivatives of θ are

$$\frac{\partial \theta}{\partial X_i} = \frac{-1}{\sqrt{1 - G^2}} \frac{\partial G}{\partial X_i} \tag{5.27}$$

$$\frac{\partial \theta}{\partial X_j} = \frac{-1}{\sqrt{1 - G^2}} \frac{\partial G}{\partial X_j} \tag{5.28}$$

$$\frac{\partial \theta}{\partial X_k} = \frac{-1}{\sqrt{1 - G^2}} \frac{\partial G}{\partial X_k} \tag{5.29}$$

The derivative of G is

$$\frac{\partial G}{\partial X_i} = \frac{\partial}{\partial X_i}\left(\frac{(\boldsymbol{p} \cdot \boldsymbol{q})}{|\boldsymbol{p}|\,|\boldsymbol{q}|}\right) = \frac{\left(X_k - X_j\right)|p|\,|q| - (\boldsymbol{p} \cdot \boldsymbol{q})\,\frac{|q|}{|p|}\left(X_i - X_j\right)}{(|\boldsymbol{p}|\,|\boldsymbol{q}|)^2} \tag{5.30}$$

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

$$\frac{\partial G}{\partial X_j} = \frac{\left(2X_j - X_i - X_k\right)|\boldsymbol{p}|\,|\boldsymbol{q}| - (\boldsymbol{p} \cdot \boldsymbol{q})\,\frac{|q|}{|p|}\left(X_j - X_i\right)}{(|\boldsymbol{p}|\,|\boldsymbol{q}|)^2} - \frac{(\boldsymbol{p} \cdot \boldsymbol{q})\,\frac{|p|}{|q|}\left(X_j - X_k\right)}{(|\boldsymbol{p}|\,|\boldsymbol{q}|)^2} \tag{5.31}$$

$$\frac{\partial G}{\partial X_k} = \frac{\left(X_i - X_j\right)|\boldsymbol{p}|\,|\boldsymbol{q}| - (\boldsymbol{p} \cdot \boldsymbol{q})\,\frac{|p|}{|q|}\left(X_k - X_j\right)}{(|\boldsymbol{p}|\,|\boldsymbol{q}|)^2} \tag{5.32}$$

Combined Eqs. (5.23), (5.27) and (5.30), we can get the following formula.

$$\frac{\partial^2 V_2}{\partial X_i^2} = \frac{2K_\theta}{1-G^2}\left(\frac{\left(X_k - X_j\right)|\boldsymbol{p}|\,|\boldsymbol{q}| - (\boldsymbol{p}\cdot\boldsymbol{q})\,\frac{|\boldsymbol{q}|}{|\boldsymbol{p}|}\left(X_i - X_j\right)}{\left(|\boldsymbol{p}|\,|\boldsymbol{q}|\right)^2}\right)^2 \tag{5.33}$$

Similarly, Combined Eqs. (5.24), (5.27), (5.28), (5.30), and (5.31), the second cross-derivative $\frac{\partial^2 V_2}{\partial X_i X_j}$ becomes

$$\frac{\partial^2 V_2}{\partial X_i X_j} = \frac{2K_\theta}{1-G^2}\frac{\left(X_k - X_j\right)|\boldsymbol{p}|\,|\boldsymbol{q}| - (\boldsymbol{p}\cdot\boldsymbol{q})\,\frac{|\boldsymbol{q}|}{|\boldsymbol{p}|}\left(X_i - X_j\right)}{\left(|\boldsymbol{p}|\,|\boldsymbol{q}|\right)^2}$$
$$\left(\frac{\left(2X_j - X_i - X_k\right)|\boldsymbol{p}|\,|\boldsymbol{q}| - (\boldsymbol{p}\cdot\boldsymbol{q})\,\frac{|\boldsymbol{q}|}{|\boldsymbol{p}|}\left(X_j - X_i\right)}{\left(|\boldsymbol{p}|\,|\boldsymbol{q}|\right)^2} - \frac{(\boldsymbol{p}\cdot\boldsymbol{q})\,\frac{|\boldsymbol{p}|}{|\boldsymbol{q}|}\left(X_j - X_k\right)}{\left(|\boldsymbol{p}|\,|\boldsymbol{q}|\right)^2}\right) \tag{5.34}$$

Following a similar approach, we can get $\frac{\partial^2 V_2}{\partial X_j X_k}$ and $\frac{\partial^2 V_2}{\partial X_k X_i}$ and these second cross-derivatives form the elements of the second 3×3 matrix of the super element $\boldsymbol{H}_{ij}$ in Eq. (5.6).

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 can be found in [28].

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

$$V_4 = \varepsilon\left[5\left(\frac{r_{0,ij}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{0,ij}}{r_{ij}}\right)^{10}\right] \tag{5.35}$$

A Taylor expansion will give us the following form.

$$V_4 = -\varepsilon + \frac{120\varepsilon}{r_{0,ij}^2}\left(r_{ij} - r_{0,ij}\right)^2 \tag{5.36}$$

Equation (5.36) 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.

$$\frac{\partial^2 V_4}{\partial X_i \partial Y_j} = -\frac{240\varepsilon}{r_{0,ij}^2}\left(X_j - X_i\right)\left(Y_j - Y_i\right)/r_{ij}^2 \tag{5.37}$$

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

$$\langle \Delta \boldsymbol{r}_i^2 \rangle = \frac{k_B T}{\gamma} \left(\boldsymbol{H}^+_{3i-2,3i-2} + \boldsymbol{H}^+_{3i-1,3i-1} + \boldsymbol{H}^+_{3i,3i} \right) \tag{5.38}$$

$$\langle \Delta \boldsymbol{r}_i \cdot \Delta \boldsymbol{r}_j \rangle = \frac{k_B T}{\gamma} \left(\boldsymbol{H}^+_{3i-2,3j-2} + \boldsymbol{H}^+_{3i-1,3j-1} + \boldsymbol{H}^+_{3i,3j} \right) \tag{5.39}$$

5.4.3 Extending STeM to an All-Atom Model

STeM [28] was originally based on the Gō potential [29, 30] and was applied to coarse-grained systems. Here we show how STeM can be extended to all-atom models. Consequently, the force field parameters used in STeM for the interactions among the atoms are adopted from an all-atom force field, for example, the Charmm22 force field.

All-atom STeM is different from NMA. Though all-atom STeM share some similarity with NMA, such as both are all-atom models and can be applied to an equilibrated structure to compute normal modes, STeM is different from NMA in the sense that it is fully spring-based models and does not consider the effect of inter-atomic forces. Indeed, as in Gō model, STeM assumes the input structure is at equilibrium, and in addition, the inter-atomic forces are all zero. NMA, however, does not make the second assumption. NMA has been often applied to locally energetically-minimized structures, where the systems are at equilibrium, but the inter-atomic forces are clearly not zero. Thus, the difference between NMA and STeM mostly represents the effect of inter-atomic forces on a system.

All-atom STeM is also different from all-atom ANM (anisotropic network model [4]). In ANM, atoms interact through two-body Hookean springs only. In STeM, atoms interact via generalized spring tensors (thus the name STeM – spring tensor model) and include three-body and four-body interactions. STeM and ANM do share some similarity. Both models are purely spring-based models and do not take into account the effect of inter-atomic forces when studying protein fluctuations and conformation changes. STeM is especially similar to a particular variant of ANM, the ANMr2, or ANM using $\frac{1}{r^2}$ as spring constants, as was thoroughly investigated in [10]. This is because, the effect of non-bonded terms in STeM, especially the van der Waal interactions, is similar to $\frac{1}{r^2}$ springs [28].

In the following, we will show how STeM is a close approximation of NMA, and how ANM is a further approximation of STeM.

5.4.3.1 The Close Relationship Between NMA to All-Atom STeM

The close relationship between STeM and NMA is illuminated in the following derivation of STeM Hessian matrix.

First, let us consider the three-body interactions, specifically the bond angle interactions. Let $\theta = \angle ijk$ be the instantaneous angle formed by three sequential atoms i, j, and k. The bond angle potential of atoms i, j, and k is defined as $V_\theta = \frac{1}{2}k_\theta(\theta - \theta_0)^2$, where k_θ is the bond angle spring constant, θ_0 is the equilibrium angle. The block Hessian matrix $\boldsymbol{H}_\theta$ for the angle interaction is a 9×9 second derivative matrix of V_θ with respect to x, y and z coordinates of atoms i, j, and k. Write out one component $\frac{\partial V_\theta}{\partial X_i \partial Y_k}$ of $\boldsymbol{H}_\theta$ as follows:

$$\begin{aligned}\frac{\partial^2 V_\theta}{\partial X_i \partial Y_k} &= \frac{\partial}{\partial X_i}\left(\frac{\partial V_\theta}{\partial \theta}\frac{\partial \theta}{\partial Y_k}\right) \\ &= \frac{\partial^2 V_\theta}{\partial \theta^2}\frac{\partial \theta}{\partial X_i}\frac{\partial \theta}{\partial Y_k} + \frac{\partial V_\theta}{\partial \theta}\frac{\partial^2 \theta}{\partial X_i \partial Y_k} \\ &= k_\theta \cdot \frac{\partial \theta}{\partial X_i}\frac{\partial \theta}{\partial Y_k} + f_\theta \cdot \frac{\partial^2 \theta}{\partial X_i \partial Y_k}\end{aligned} \tag{5.40}$$

where $f_\theta = \frac{\partial V_\theta}{\partial \theta}$ is the bending force (which actually is a torque). Notice that Eq. (5.40) is a combination of the physical terms (k_θ and f_θ) and geometric terms (the rest of the terms), which represent the projection of physical interactions into a particular coordinate system. In a similar fashion, the rest of the elements of the block hessian matrix $\boldsymbol{H}_\theta$ can be written out using k_θ and f_θ. Finally, the block Hessian matrix H_θ can be rewritten as a summation of two terms:

$$\boldsymbol{H}_\theta = \boldsymbol{H}_\theta^{\mathrm{NMA}} = k_\theta \cdot \boldsymbol{H}_{\theta|K_\theta} + f_\theta \cdot \boldsymbol{H}_{\theta|f_\theta} \tag{5.41}$$

where $\boldsymbol{H}_{\theta\big|k_\theta}$ and $\boldsymbol{H}_{\theta\big|f_\theta}$ are 9×9 matrice that are fully determined by protein geometry and atom coordinates, where k_θ is a force field parameter and $f_\theta = k_\theta(\theta - \theta_0)$ is the torque acting on the bond angle. In STeM, the bending torque f_θ is assumed to be 0, i.e., $f_\theta \to 0$. This simplifies the $\boldsymbol{H}_\theta$ in Eq. (5.41) and it becomes:

$$\boldsymbol{H}_\theta^{\mathrm{STeM}} = k_\theta \cdot \boldsymbol{H}_{\theta\big|k_\theta}. \tag{5.42}$$

Now for the four-body interactions. Let $\boldsymbol{H}_\phi$ be the 12×12 block Hessian matrix for the dihedral interaction among four atoms i, j, k, and l. Let $k_\phi = \frac{\partial^2 V}{\partial \phi^2}$ and $f_\phi = \frac{\partial V}{\partial \phi}$ be the dihedral spring constant and bending force (torque), respectively. Similar to Eq. (5.41), the Hessian matrix $\boldsymbol{H}_\phi$ can be written as a function of k_ϕ and f_ϕ:

$$\boldsymbol{H}_\phi = \boldsymbol{H}_\phi^{\mathrm{NMA}} = k_\phi \cdot \boldsymbol{H}_{\phi\big|K_\phi} + f_\phi \cdot \boldsymbol{H}_{\phi\big|f_\phi}. \tag{5.43}$$

Since $V(\phi) = K_\phi(1 - \cos(n(\phi - \phi_0)))$ in most force fields, where K_ϕ and ϕ_0 are force field parameters and n is the multiplicity, $k_\phi = \frac{\partial^2 V}{\partial \phi^2} = n^2 K_\phi \cos(n(\phi - \phi_0))$. In STeM, the torque f_ϕ is assumed to be zero. In addition, STeM assumes that the

input structure has the equilibrium values for all the dihedral angles, i.e., $\phi = \phi_0$. Therefore,

$$\boldsymbol{H}_{\phi}^{\mathrm{STeM}} = k_{\phi} \cdot \boldsymbol{H}_{\phi\big|_{K_\phi}} = n^2 K_{\phi} \cdot \boldsymbol{H}_{\phi\big|_{K_\phi}}. \tag{5.44}$$

Improper is a special type of dihedral interactions. Improper potential takes the form of $V(\psi) = K_{\psi}(\psi - \psi_0)^2$, where K_{ψ} and ψ_0 are force field parameters. To simplify notations for improper interaction, we define $\boldsymbol{H}_{\psi\big|_{H_\psi}}$ in the same way as $\boldsymbol{H}_{\phi\big|_{H_\phi}}$, and its spring constant $k_{\psi} = \frac{\partial^2 V(\psi)}{\partial \psi^2} = 2K_{\psi}$. Therefore,

$$\boldsymbol{H}_{\psi}^{\mathrm{STeM}} = k_{\psi} \cdot \boldsymbol{H}_{\psi\big|_{K_\psi}} = 2K_{\psi} \cdot \boldsymbol{H}_{\psi\big|_{K_\psi}}. \tag{5.45}$$

Likewise, the Hessian matrix $\boldsymbol{H}_l$ for two-body interaction between a pair of atoms i and j can be determined: $\boldsymbol{H}_l = k_l \cdot \boldsymbol{H}_{l\big|_{k_l}} + f_l \cdot \boldsymbol{H}_{l\big|_{f_l}}$. In STeM, the force term is again assumed to be zero. As for the first term, there are three types of two-body interactions in an all-atom potential, i.e., bond stretching, van der Waals interactions, and electrostatic interactions, and thus different k_l. For the bond stretching potential, or V_{bond}, which is usually expressed as $V_{\mathrm{bond}} = K_{\mathrm{bond}}(r - r_0)^2$, we have

$$k_{\mathrm{bond}} = \frac{\partial^2 V_{\mathrm{bond}}}{\partial r^2} = 2K_{\mathrm{bond}}. \tag{5.46}$$

For van der Waal term, whose potential is $V_{\mathrm{vdW}} = \epsilon\left(\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^{6}\right)$, where ϵ and r_0 are force field constants. We have

$$k_{\mathrm{vdW}} = \frac{\partial^2 V_{\mathrm{vdW}}}{\partial r^2} = \frac{12\epsilon}{r^2}\left(13\left(\frac{r_0}{r}\right)^{12} - 7\left(\frac{r_0}{r}\right)^{6}\right). \tag{5.47}$$

Lastly, for the electrostatic term, since $V_{\mathrm{elec}} = \frac{332 q_i \cdot q_j}{rD}$, where q_i is the partial charge of atom i, and D is the dielectric constant and is set to be 80, k_l is thus:

$$k_{\mathrm{elec}} = \frac{\partial^2 V_{\mathrm{elec}}}{\partial r^2} = \frac{2 \cdot 332 q_i \cdot q_j}{80 r^3} = \frac{8.3 q_i \cdot q_j}{r^3}. \tag{5.48}$$

Finally, the spring constant k_l for two-body interaction is

$$k_l = k_{\mathrm{bond}} + k_{\mathrm{vdW}} + k_{\mathrm{elec}}. \tag{5.49}$$

This spring constant may become negative. In that case, we set k_l to be zero to avoid producing negative eigenvalues from the STeM Hessian matrix.

Finally, let N be the number of atoms, the $3N \times 3N$ full Hessian matrix $\boldsymbol{H}^{\mathrm{NMA}}$ for the whole system can be written as a summation of a spring constant related term $\boldsymbol{H}_{spr}^{\mathrm{NMA}}$ and a force/torque related term $\boldsymbol{H}_{\mathrm{frc}}^{\mathrm{NMA}}$:

$$\boldsymbol{H}^{\mathrm{NMA}} = \boldsymbol{H}^{\mathrm{NMA}}_{\mathrm{spr}} + \boldsymbol{H}^{\mathrm{NMA}}_{\mathrm{frc}}, \tag{5.50}$$

where $\boldsymbol{H}^{\mathrm{NMA}}_{\mathrm{spr}}$ and $\boldsymbol{H}^{\mathrm{NMA}}_{\mathrm{frc}}$ are

$$\boldsymbol{H}^{\mathrm{NMA}}_{\mathrm{spr}} = \sum_{\theta\in\Theta} k_\theta \boldsymbol{H}_{\theta}\big|_{k_\theta} + \sum_{\phi\in\Phi} k_\phi \boldsymbol{H}_{\phi}\big|_{k_\phi} + \sum_{\psi\in\Psi} k_\psi \boldsymbol{H}_{\psi}\big|_{k_\psi} + \sum_{l\in L} k_l \boldsymbol{H}_{l}\big|_{k_l}, \tag{5.51}$$

$$\boldsymbol{H}^{\mathrm{NMA}}_{\mathrm{frc}} = \sum_{\theta\in\Theta} f_\theta \boldsymbol{H}_{\theta}\big|_{k_\theta} + \sum_{\phi\in\Phi} f_\phi \boldsymbol{H}_{\phi}\big|_{k_\phi} + \sum_{\psi\in\Psi} f_\psi \boldsymbol{H}_{\psi}\big|_{k_\psi} + \sum_{l\in L} f_l \boldsymbol{H}_{l}\big|_{k_l}, \tag{5.52}$$

where Θ, Φ, Ψ, and L are the sets of angular, dihedral, improper and pairwise interactions.

STeM assumes that all forces and torques are zero. Therefore,

$$\boldsymbol{H}^{\mathrm{STeM}} \approx \boldsymbol{H}^{\mathrm{NMA}}_{\mathrm{spr}}. \tag{5.53}$$

It is approximately equal since STeM additionally assumes that the input structure has the equilibrium values for the dihedral potentials, while NMA does not. Specifically, $\boldsymbol{H}^{\mathrm{STeM}}$ is,

$$\boldsymbol{H}^{\mathrm{STeM}} = \sum_{\theta\in\Theta} k_\theta \boldsymbol{H}_{\theta|K_\theta} + \sum_{\phi\in\Phi} k_\phi \boldsymbol{H}_{\phi|K_\phi} + \sum_{\psi\in\Psi} k_\psi \boldsymbol{H}_{\psi|K_\psi} + \sum_{l\in L} k_l \boldsymbol{H}_{l|K_l} \tag{5.54}$$

Our original work on STeM [28] details how $\boldsymbol{H}^{\mathrm{STeM}}$ can be computed. To compute $\boldsymbol{H}^{\mathrm{NMA}}$, one may use software packages such as gromacs [31] or tinker [27].

5.4.3.2 The Relationship Between STeM and ANM, the Role of Multi-body Interactions

ANM [4] is a widely-used coarse-grained model for proteins. A particular variant of ANM, ANMr2, which uses $\frac{1}{r^2}$ as spring constants, was thoroughly investigated in [10] and was shown to have better performance than the regular ANM.

ANM, particularly ANMr2, is closely related to STeM in that the former is a simplification of the latter [28]. STeM is a close approximation of the NMA, and ANM/ANMr2 is a further approximation of STeM. STeM ignores the contributions of inter-atomic forces that are considered in NMA (Eq. (5.50)), while ANMr2/ANM takes into account only the two-body interactions and ignores the contributions of multi-body interactions (bond angle and torsional angle interactions) that are considered in STeM.

5.4.4 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 [32] 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 [33]. 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 has significantly large structure difference from each other.

5.4.5 Evaluation Techniques

We used the same evaluation techniques as have been applied before [20, 32]. Specifically, the following three numerical measures are used.

5.4.6 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.

$$\rho = \frac{\sum_i^N (x_i - \overline{x})\,(y_i - \overline{y})}{\left[\sum_i^N (x_i - \overline{x})^2 \sum_i^N (y_i - \overline{y})^2\right]^{1/2}} \tag{5.55}$$

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

5.4.7 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

$$I = \frac{\left|\sum_i^{3N} e_i r_i\right|}{\left[\sum_i^{3N} e_i^2 \sum_i^{3N} r_i^2\right]^{1/2}} \tag{5.56}$$

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.

5.4.8 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 Eq. (5.55), with different meaning for x_i and y_i.

$$\rho = \frac{\sum_i^N (x_i - \overline{x})(y_i - \overline{y})}{\left[\sum_i^N (x_i - \overline{x})^2 \sum_i^N (y_i - \overline{y})^2\right]^{1/2}} \tag{5.57}$$

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. $\overline{x}$ and $\overline{y}$ are the corresponding mean values.

Acknowledgments Funding from National Science Foundation (CAREER award, CCF-0953517) is gratefully acknowledged.

Competing Interests The authors declare that they have no competing interests.

References

1. Voth GA (2009) Coarse-graining of condensed phase and biomolecular systems. CRC Press, Boca Raton. xviii, 455 p., 16 p. of plates
2. Ma J (2005) Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. Structure 13(3):373–380
3. Bahar I, Atilgan AR, Erman B (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des 2(3):173–181
4. Atilgan AR et al (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 80(1):505–515
5. Bahar I, Rader AJ (2005) Coarse-grained normal mode analysis in structural biology. Curr Opin Struct Biol 15(5):586–592

6. Ming D, Bruschweiler R (2006) Reorientational contact-weighted elastic network model for the prediction of protein dynamics: comparison with NMR relaxation. Biophys J 90(10): 3382–3388
7. Song G, Jernigan RL (2006) An enhanced elastic network model to represent the motions of domain-swapped proteins. Proteins 63(1):197–209
8. Song G, Jernigan RL (2007) VGNM: a better model for understanding the dynamics of proteins in crystals. J Mol Biol 369(3):880–893
9. Lu M, Poon B, Ma J (2006) A new method for coarse-grained elastic normal-mode analysis. J Chem Theory Comput 2(3):464–471
10. Yang L, Song G, Jernigan RL (2009) Protein elastic network models and the ranges of cooperativity. Proc Natl Acad Sci USA 106(30):12347–12352
11. Zheng W (2008) A unification of the elastic network model and the Gaussian network model for optimal description of protein conformational motions and fluctuations. Biophys J 94(10):3853–3857
12. Stember JN, Wriggers W (2009) Bend-twist-stretch model for coarse elastic network simulation of biomolecular motion. J Chem Phys 131(7):074112
13. Kundu S, Sorensen DC, Phillips GN Jr (2004) Automatic domain decomposition of proteins by a Gaussian Network Model. Proteins 57(4):725–733
14. Bahar I, Chennubhotla C, Tobi D (2007) Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. Curr Opin Struct Biol 17(6):633–640
15. Zheng WJ, Brooks B (2005) Identification of dynamical correlations within the myosin motor domain by the normal mode analysis of an elastic network model. J Mol Biol 346(3):745–759
16. Yang Z, Majek P, Bahar I (2009) Allosteric transitions of supramolecular systems explored by network models: application to chaperonin GroEL. PLoS Comput Biol 5(4):e1000360
17. Lin TL, Song G (2010) Generalized spring tensor models for protein fluctuation dynamics and conformation changes. BMC Struct Biol 10(Suppl 1):S3
18. Tirion MM (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett 77(9):1905–1908
19. Hinsen K (1998) Analysis of domain motions by approximate normal mode calculations. Proteins 33(3):417–429
20. Tama F, Sanejouand YH (2001) Conformational change of proteins arising from normal mode calculations. Protein Eng 14(1):1–6
21. Thorpe MF (2007) Comment on elastic network models and proteins. Phys Biol 4(1):60–63; discussion 64–55
22. Koga N, Takada S (2006) Folding-based molecular simulations reveal mechanisms of the rotary motor F1-ATPase. Proc Natl Acad Sci USA 103(14):5367–5372
23. Clementi C, Nymeyer H, Onuchic JN (2000) 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 298(5):937–953
24. Koga N, Takada S (2001) Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model. J Mol Biol 313(1):171–180
25. Riccardi D, Cui Q, Phillips GN Jr (2009) Application of elastic network models to proteins in the crystalline state. Biophys J 96(2):464–475
26. Hinsen K et al (2000) Harmonicity in slow protein dynamics. Chem Phys 261(1–2):25–37
27. Ponder JW, Richards FM (1987) An efficient Newton-like method for molecular mechanics energy minimization of large molecules. J Comput Chem 8(7):1016–1024
28. Lin TL, Song G (2010) Generalized spring tensor models for protein fluctuation dynamics and conformation changes. BMC Struct Biol 10(Suppl 1):S3
29. Go N (1983) Protein folding as a stochastic-process. J Stat Phys 30(2):413–423
30. Taketomi H, Ueda Y, Go N (1975) Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. Int J Pept Protein Res 7(6):445–459

31. Pronk S, Páll S et al (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29(7):845–854
32. Kundu S et al (2002) Dynamics of proteins in crystals: comparison of experiment with simple models. Biophys J 83(2):723–732
33. Berman HM et al (2000) The protein data bank. Nucleic Acids Res 28(1):235–242