A Unification of the Elastic Network Model and the Gaussian Network Model for Optimal Description of Protein Conformational Motions and Fluctuations

Wenjun Zheng
Department of Physics, University at Buffalo, State University of New York, Buffalo, New York 14260-1500

ABSTRACT Coarse-grained elastic models with a Cα-only representation and harmonic interactions have been increasingly used to describe the conformational motions and flexibility of various proteins. In this work, we will unify two complementary elastic models—the elastic network model (ENM) and the Gaussian network model (GNM), in the framework of a generalized anisotropic network model (G-ANM) with a new anisotropy parameter, \( f_{\text{anm}} \). The G-ANM is reduced to GNM at \( f_{\text{anm}} = 1 \) and ENM at \( f_{\text{anm}} = 0 \). By analyzing a list of protein crystal structure pairs using G-ANM, we have attained optimal descriptions of both the isotropic thermal fluctuations and the crystallographically observed conformational changes with a small \( f_{\text{anm}} \) (< 0.1) and a physically realistic cutoff distance, \( R_c \sim 8 \, \text{Å} \). Thus, the G-ANM improves the performance of GNM and ENM while preserving their simplicity. The properly parameterized G-ANM will enable more accurate and realistic modeling of protein conformational motions and flexibility.

INTRODUCTION

Understanding protein conformational dynamics holds the key to decrypting protein functions at a microscopic level. Simplified coarse-grained models (1–4) have been established as valid and efficient means to probe protein conformational motions and flexibility beyond the reach of atomistic molecular simulations (5). Here we focus on two coarse-grained elastic models: the elastic network model (ENM) (6–8) and the Gaussian network model (GNM) (9,10). Both models simplify the atomic interactions in proteins by using elastic interactions between Cα atoms within a cutoff distance \( R_c \). GNM has been shown to perform better than ENM in describing the thermal fluctuations of protein structures measured by the isotropic crystallographic B factors (11–13). Additionally, the GNM-based calculation of B-factors is insensitive to tors (11–13). Additionally, the GNM-based calculation of B-factors is insensitive to tors (11–13). However, the isotropic GNM cannot predict the directions of protein motions. Instead, the normal mode analysis (16) of ENM has been shown to yield a handful of lowest normal modes that quantitatively capture the conformational changes observed between different protein crystal structures (8,17–20). Therefore, GNM and ENM are complementary in describing the thermal fluctuations and conformational motions in proteins, but neither is satisfactory by itself.

In this work, we intend to unify GNM and ENM in the framework of a generalized anisotropic network model (G-ANM) with a new parameter \( f_{\text{anm}} \) that defines the extent of anisotropy between the longitudinal and transverse motions between pairs of neighboring residues (or Cα atoms). At the isotropic limit (\( f_{\text{anm}} = 1 \)), the G-ANM is reduced to a GNM; at the fully anisotropic limit (\( f_{\text{anm}} = 0 \)) the G-ANM is reduced to an ENM. Then we explore the intermediate values of \( f_{\text{anm}} \) to quantitatively assess the performance of a G-ANM in describing both the isotropic thermal fluctuations and the observed conformational changes for a selected list of 18 test cases, each corresponding to a pair of protein structures from the Protein Data Bank (PDB). The systematic evaluation of this list allows us to understand the \( f_{\text{anm}} \)-dependence of the quality of G-ANM. We also consider a range of \( R_c \) values (7 Å ≤ \( R_c \) ≤ 20 Å) to explore the \( R_c \)-dependence of the quality of G-ANM.

Our main findings are as follows: by parameterizing G-ANM at a small \( f_{\text{anm}} \) (\( f_{\text{anm}} \leq 0.1 \)) and a relatively short cutoff distance \( R_c \equiv 8 \, \text{Å} \), we are able to achieve optimal descriptions of both the isotropic thermal fluctuations and the crystallographically observed conformational changes, which are comparable with the best descriptions of the thermal fluctuations attained by GNM (for 8 Å ≤ \( R_c \) ≤ 12 Å) and the best descriptions of the observed conformational changes attained by ENM (for 8 Å ≤ \( R_c \) ≤ 12 Å). Therefore, this study demonstrates an effective way to improve both GNM and ENM without hurting the simplicity of these coarse-grained models.

METHODS

Generalized anisotropic network model

Given the Cα atomic coordinates of a protein crystal structure, we define the G-ANM potential energy as a weighted sum of two harmonic potentials to describe the pairwise interactions between neighboring Cα atoms:

\[
\mathcal{E}_{\text{G-ANM}}(\mathbf{R}) = \sum_{i,j} k_{ij} (R_{ij} - R_{ij}^0)^2 + \sum_{i,j} \frac{k_{ij} \Delta R_{ij}}{2} (R_{ij} - R_{ij}^0) + \sum_{i,j} \frac{k_{ij} \Delta R_{ij}}{2} (R_{ij} - R_{ij}^0)^2
\]

where \( k_{ij} \) is the force constant, \( R_{ij} \) is the distance between atoms i and j, and \( R_{ij}^0 \) is the equilibrium distance. The first term represents the harmonic potential, while the second term represents the anisotropic potential.
\[
E_{\text{G-ANM}} = \frac{1}{2} \sum_{i,j,d^i_0 < R_c} C_{ij} (f_{anm} | \vec{x}_i - \vec{x}_j |^2 + (1 - f_{anm}) |(\vec{x}_i' - \vec{x}_i) \cdot \vec{n}^i_0 |^2),
\]

where \(0 \leq f_{anm} \leq 1\) is the anisotropy weight parameter (see below) and \(i\) or \(j\) is the index for a \(C_a\) atom, \(d^i_0\) is the distance between the equilibrium positions of \(i\) and \(j\), \(\vec{x}_i'\) is the three-dimensional (3D) displacement of \(i\) (\(\vec{n}^i_0\) is the unit vector pointing from the equilibrium position of \(i\) to that of \(j\), \(C_{ij}\) is the force constant of the spring between \(i\) and \(j\); \(C_{ij} = 10\) if \(i\) and \(j\) are bonded, and \(C_{ij} = 0\) otherwise. \(C\) can be determined by fitting the crystalllographic B-factors (see below). The use of two force constants for the bonded and nonbonded residue-residue interactions was shown to improve the performance of ENM (21) and GNM (12).

The physical basis for the weighted combination adopted in Eq. 1 is as follows. For a pair of contacting \(C_a\) atoms (\(i\), \(j\)), \(\vec{x}_i' - \vec{x}_i\) can be partitioned into the longitudinal (parallel to \(\vec{n}^i_0\)) and the transverse (perpendicular to \(\vec{n}^i_0\)) components (see Fig. 1). In ENM, the stiffness for the latter component (the curvature of components (see Fig. 1)) is the isotropic limit, \(E\), and then compute the average (\(CO_{\text{avg}}\) and \(SD_{\text{norm}}\)) and standard deviation (\(CO_{\text{SD}}_{\text{norm}}\)) are computed for \(CO_{\text{norm}}\) among a selected list of 18 test cases (Table 1). A high quality of G-ANM in fitting B-factors is reflected by a high (low) value of \(CO_{\text{avg}}(CO_{\text{SD}}_{\text{norm}})\).

**Evaluation of G-ANM in describing the observed conformational changes**

The quality of G-ANM in describing the observed conformational changes is assessed by the cumulative overlap (CO) for the lowest 15 modes: \(CO = \sqrt{\sum_{m=1}^{15} E_m}\), where \(E_m = \sum_{i=1}^{N} v_{nm} \cdot x_{bm} / \sqrt{\sum_{i=1}^{N} [x_{bm}]^2}\), \(v_{nm}\) is the 3D component of the eigenvector of mode \(m\) at \(C_a\) atom \(i\), and \(x_{bm}\) is the observed structural displacement at \(C_a\) atom \(i\). We perform a similar normalization for CO and then compute the average (\(CO_{\text{avg}}\)) and standard deviation (\(CO_{\text{SD}}\)) of \(CO_{\text{norm}}\) for the lowest 15 modes of the G-ANM and that of the ENM: \(\text{SIM} = (1/15) \sum_{m=1}^{15} \sum_{i=1}^{N} f(m_i, m_j)^2\).

**Comparison between the lowest modes of G-ANM and ENM**

We compute the cumulative similarity score between the lowest 15 modes of the G-ANM and that of the ENM: \(\text{SIM} = (1/15) \sum_{m=1}^{15} \sum_{i=1}^{N} f(m_i, m_j)^2\).
TABLE 1 List of 22 pairs of protein structures from PDB

<table>
<thead>
<tr>
<th>Protein</th>
<th>No. of residues</th>
<th>PDB codes and chains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenylate kinase</td>
<td>218</td>
<td>1aky, 2ak3A</td>
</tr>
<tr>
<td>Alcohol dehydrogenase</td>
<td>373</td>
<td>8ad, 6adA</td>
</tr>
<tr>
<td>Annexin V</td>
<td>317</td>
<td>1av, 1avA</td>
</tr>
<tr>
<td>Calmodulin</td>
<td>144</td>
<td>1ell, 1etr</td>
</tr>
<tr>
<td>Che Y protein</td>
<td>128</td>
<td>3chy, 1chn</td>
</tr>
<tr>
<td>Enolase</td>
<td>436</td>
<td>3enl, 7enl</td>
</tr>
<tr>
<td>HIV-1 protease</td>
<td>99</td>
<td>1hhp, 1ajxA</td>
</tr>
<tr>
<td>Lactoferrin</td>
<td>691</td>
<td>1lfh, 1lfg</td>
</tr>
<tr>
<td>LAO binding protein</td>
<td>238</td>
<td>2hao, 1lst</td>
</tr>
<tr>
<td>Maltodextrin binding protein</td>
<td>370</td>
<td>1omp, 1anf</td>
</tr>
<tr>
<td>Thymidylate synthase</td>
<td>264</td>
<td>3ms, 2tscA</td>
</tr>
<tr>
<td>Triacylglyceride lipase</td>
<td>265</td>
<td>3gl, 4gl</td>
</tr>
<tr>
<td>Tyrosine phosphatase</td>
<td>278</td>
<td>1lytA, 1lyts</td>
</tr>
<tr>
<td>Guanylate kinase</td>
<td>186</td>
<td>1ex7A, 1ex6A</td>
</tr>
<tr>
<td>Serum transferrin</td>
<td>328</td>
<td>1bp5A, 1a8e</td>
</tr>
<tr>
<td>Ras p21 protein catalytic domain</td>
<td>169</td>
<td>4q21, 5p21</td>
</tr>
<tr>
<td>Transducin-α</td>
<td>314</td>
<td>1tag, 1ndA</td>
</tr>
<tr>
<td>5-Enol-pyruvyl-3-phosphate synthase</td>
<td>427</td>
<td>1eps, 1gfsA</td>
</tr>
<tr>
<td>Oligo-peptide binding protein</td>
<td>517</td>
<td>1rk, 2rkA</td>
</tr>
<tr>
<td>RNA helicase</td>
<td>435</td>
<td>8ohm, 1cu1A</td>
</tr>
<tr>
<td>Myosin</td>
<td>730</td>
<td>1vom, 1mma</td>
</tr>
<tr>
<td>Rb69 DNA polymerase</td>
<td>897</td>
<td>1ih7A, 1ig9</td>
</tr>
</tbody>
</table>

Four pairs eliminated from the analysis in bold (the selection removes those low-quality test cases if \( CC_{\text{max}} < 0.5 \) or \( CO_{\text{max}} < 0.5 \). The remaining 18 pairs are used for the evaluation of G-ANM.

\[
I(m_1, m_2) = \sum l_{m_1}^{G-\text{ANM}} v_{m_1}^{G-\text{ENM}} v_{m_1}^{G-\text{SIM}} v_{m_1}^{G-\text{SIM}}
\]

where \( l_{m_1}^{G-\text{ANM}} \), \( v_{m_1}^{G-\text{ENM}} \), and \( v_{m_1}^{G-\text{SIM}} \) is the 3D component of the eigenvector of mode \( m_1 \) of the G-ANM (ENM) at \( C_0 \) atom \( i \). If the two sets of modes span the same subspace, \( \text{SIM} = 1 \); otherwise \( 0 \leq \text{SIM} < 1 \). We note that as \( f_{\text{ann}} \rightarrow 0 \), \( \text{SIM} \rightarrow 0.8 \) instead of 1 because the lowest 3 nonzero modes of the G-ANM converge to the 3 rotational zero modes of the ENM. We compute the average \( \text{SIM}^\text{AVG} \) and standard deviation \( \text{SIM}^\text{SD} \) of \( \text{SIM} \) for the 18 selected test cases as a function of \( f_{\text{ann}} \) and \( R_c \).

RESULTS

We quantitatively assess the performance of G-ANM in describing both the isotropic thermal fluctuations and the observed conformational changes for a selected list of 18 test cases, each consisting of a pair of protein structures from the PDB. The list (Table 1) is compiled from an early work on ENM (17) and our recent work (23,24). We only include the crystal structures that do not have extensive interface between individual structural units.

Evaluation of G-ANM in describing the crystallographic B-factors

The quality of G-ANM in fitting the crystallographic B-factors is assessed by the \( CC \) between theoretical and experimental B-factors (see Methods). We analyze the average \( \text{CC}^{\text{AVG}} \) and standard deviation \( \text{CC}^{\text{SD}} \) of “normalized” \( CC \) over a list of selected test cases (see Methods and Table 1).

\[
\text{CC}^{\text{AVG}} \text{ and } \text{CC}^{\text{SD}} \text{ as a function of } f_{\text{ann}} \text{ and } R_c \text{ are shown in Fig. 2, a and b. The } f_{\text{ann}} \text{-dependence of } CC_{\text{norm}} \text{ at fixed } R_c \text{ is as follows: for } R_c = 7 \text{ Å (8 Å), } CC_{\text{AVG}} \text{ peaks at } f_{\text{ann}} = 0.01(0.1). \text{ For } R_c \geq 10 \text{ Å, the peak shifts to } f_{\text{ann}} = 1 \text{ and its height decreases as } R_c \text{ increases. For } 8 \text{ Å} \leq R_c \leq 20 \text{ Å, } CC_{\text{SD}} \text{ rapidly decreases as } f_{\text{ann}} \text{ increases in } 0 < f_{\text{ann}} < 0.1, \text{ and it becomes flat or slightly increases as } f_{\text{ann}} \text{ increases in } 0.1 \leq f_{\text{ann}} \leq 1. \text{ The observation that } CC_{\text{AVG}} \text{ and } CC_{\text{SD}} \text{ vary substantially in } 0 < f_{\text{ann}} < 0.1 \text{ but change little in } 0.1 \leq f_{\text{ann}} < 1 \text{ suggests that the introduction of small isotropic interactions (the first term of Eq .1) significantly improves the quality of G-ANM in fitting the B-factors to a level comparable with GNM. This improvement is much more pronounced for } R_c = 7 \text{ Å} \text{ or } 8 \text{ Å than for } R_c \geq 10 \text{ Å: for } R_c = 8 \text{ Å, } CC_{\text{AVG}} \text{ increases significantly from 0.64 at } f_{\text{ann}} = 0 \text{ to 0.94 at } f_{\text{ann}} = 0.1, \text{ and } CC_{\text{SD}} \text{ decreases sharply from 0.22 at } f_{\text{ann}} = 0 \text{ to <0.05 at } f_{\text{ann}} = 0.1. \text{ Then we examine the } R_c \text{-dependence of } CC_{\text{norm}} \text{ at fixed } f_{\text{ann}}: \text{ for } f_{\text{ann}} = 0 \text{ (the ENM limit), } CC_{\text{AVG}} (CC_{\text{SD}}) \text{ is maximal (minimal) at } R_c = 20 \text{ Å and minimal (maximal) at } R_c = 7 \text{ Å, suggesting that the optimized fitting of the B-factors by GNM requires high } R_c \text{ (beyond the physical interaction range: } 4.4 \text{ Å to } 12.8 \text{ Å, see Cieplak and Hoang (15)). However, the above } R_c \text{-dependence is changed for } 10^{-4} < f_{\text{ann}} < 0.1: \text{ the maximum (minimum) of } CC_{\text{AVG}} (CC_{\text{SD}}) \text{ is moved to } R_c = 7 \text{ Å or } 8 \text{ Å, which is now within the physical interaction range.} \text{ When } f_{\text{ann}} \text{ and } R_c \text{ are both variable, the optimal fitting of the B-factors by GNM is attained at a physically realistic } R_c \sim 8 \text{ Å and a small } f_{\text{ann}} \sim 0.1, \text{ instead of the ENM limit or the GNM limit.} \text{ Evaluation of G-ANM in describing the observed conformational changes} \text{ The quality of G-ANM in describing the crystallographically observed conformational changes is assessed by the cumulative overlap (CO) between the 15 lowest modes and the observed changes (see Methods). We analyze the average \( CO_{\text{AVG}} \) and standard deviation \( CO_{\text{SD}} \) of “normalized” \( CO \) over a list of selected test cases (see Methods and Table 1).

\[
CO_{\text{AVG}} \text{ and } CO_{\text{SD}} \text{ as a function of } f_{\text{ann}} \text{ and } R_c \text{ are shown in Fig. 2, c and d. The } f_{\text{ann}} \text{-dependence of } CO_{\text{norm}} \text{ at fixed } R_c \text{ is as follows: for } R_c = 7 \text{ Å (8 Å), } CO_{\text{AVG}} \text{ is maximal at } f_{\text{ann}} = 0.001 (0.003); \text{ for } R_c \geq 12 \text{ Å, the maximum shifts toward higher } f_{\text{ann}} \text{ and its height decreases gradually as } R_c \text{ increases. Similarly, for } R_c = 7 \text{ Å (8 Å), } CO_{\text{SD}} \text{ is minimal at } f_{\text{ann}} = 0.001 (0.003); \text{ for } R_c \geq 12 \text{ Å, the minimum shifts toward higher } f_{\text{ann}} \text{ and its value increases gradually as } R_c \text{ increases. Notably, with the exception of } R_c = 7 \text{ Å, } CO_{\text{AVG}} \text{ and } CO_{\text{SD}} \text{ change little in } 0 < f_{\text{ann}} < 0.01, \text{ but vary substantially in } 0.01 < f_{\text{ann}} \leq 1. \text{ Therefore, small isotropic interactions (the first term of Eq .1) do not significantly degrade the quality of G-ANM in describing the observed protein conformational changes when compared with the ENM. Instead, for } R_c = 7 \text{ Å and 8 Å, an improvement in such quality is found.} \text{ Next we study the } R_c \text{-dependence of } CO_{\text{norm}} \text{ at fixed } f_{\text{ann}}: \text{ for } f_{\text{ann}} = 0 \text{ (the ENM limit), } CO_{\text{AVG}} (CO_{\text{SD}}) \text{ is maximal (minimal) at } R_c = 8 \text{ Å and minimal (maximal) at } R_c = 20 \text{ Å, the physical interaction range.
suggesting that the optimized description of the observed conformational changes by ENM requires a relatively small $R_c$ (contrary to the fitting of B-factors). With the exception of $R_c = 7$ Å, the above $R_c$-dependence is essentially maintained in $0 < \text{f}_{\text{anm}} \leq 0.01$, and the maximum (minimum) of $\text{CO}_{\text{AVG}}$ (CO$_{\text{norm}}$) remains at $R_c = 8$ Å.

When both $\text{f}_{\text{anm}}$ and $R_c$ are variable, the optimal description of the observed protein conformational changes is achieved at a physically realistic $R_c = 8$ Å and a small $\text{f}_{\text{anm}}$; $0.003$, which is slightly better than at the ENM limit.

**Comparison between the lowest modes of G-ANM and ENM**

To further understand the $\text{f}_{\text{anm}}$-dependence of the quality of G-ANM in describing the observed conformational changes, we will evaluate how much the lowest modes of the G-ANM differ from that of the ENM as $\text{f}_{\text{anm}}$ varies using a cumulative similarity score $\text{SIM}$ (see Methods).

The $\text{f}_{\text{anm}}$-dependence of $\text{SIM}_{\text{AVG}}$ at fixed $R_c$ resembles that of $\text{CO}_{\text{norm}}$ (Fig. 2c): for $R_c = 7$ Å and 8 Å, $\text{SIM}_{\text{AVG}}$ is peaked at $\text{f}_{\text{anm}} = 0.001$; for $R_c \geq 10$ Å, the peak disappears and the curve’s right-side edge shifts toward higher $\text{f}_{\text{anm}}$ as $R_c$ increases. For fixed $\text{f}_{\text{anm}}$, $\text{SIM}_{\text{AVG}}$ (SIM$^\text{SD}$) is lower (higher) at $R_c = 7$ Å or 8 Å than at $R_c \geq 10$ Å. Thus, for $\text{f}_{\text{anm}} < 0.01$, the lowest modes of the G-ANM differ significantly from that of the ENM only if $R_c$ is relatively small ($R_c \approx 8$ Å). Such difference is mainly due to the occurrence of extra zero modes in ENM for $R_c \leq 8$ Å (in addition to the 6 translational and rotational zero modes), which overestimate the mobility of the sparsely connected regions in ENM (such as a surface loop). The addition of the isotropic interaction energy (the first term of Eq. 1) in the G-ANM eliminates these additional zero modes in all the 22 test cases, thus removing a major source of errors in ENM.

**DISCUSSIONS AND CONCLUSIONS**

This work is, to our knowledge, the first attempt to unify GNM and ENM for the simultaneous modeling of both the thermal fluctuations and conformational motions in protein structures, despite recent efforts for model improvement within the framework of either GNM (12) or ENM (25). Our optimal solution is a generalized anisotropic network model parameterized with a physically realistic cutoff distance $R_c = 8$ Å and a small anisotropy parameter $\text{f}_{\text{anm}} = 0.1$. The optimal values of $\text{f}_{\text{anm}}$ for describing thermal fluctuations and conformational motions are both small, although they are numerically different: the former ($\approx 0.1$) is higher than the latter ($\approx 0.003$). The contradicting parameter optimizations in ENM (the B-factors’ fitting demands high $R_c$ whereas the description of observed conformational changes requires low $R_c$) are resolved in G-ANM: the optimal descriptions of both quantities are achieved at $R_c \approx 8$ Å.

The use of a relatively small $R_c$ is more advantageous because:

1. It agrees with the physical range of residue-residue contact interactions (including van der Waals and screened electrostatic interactions).
2. It enables more realistic modeling of medium-range (8–20 Å) interactions and couplings, which are crucial in allostery but obscured by large R.  
3. Smaller R also leads to lower computational cost in the normal mode analysis of ENM (or GNM), because the Hessian (or Kirchhoff) matrix is more sparse (note that the Hessian matrix of G-ANM is as sparse as that of ENM; thus the computational cost of the G-ANM is similar to that of ENM).

Due to the anisotropic geometry of amino acid side chains, the physical interactions between two contacting residues are intrinsically anisotropic: they depend on both the distance and the orientation between the two residues. The orientation-dependence, which is absent in the ENM potential, is incorporated in the G-ANM by introducing a new parameter $f_{\text{ann}} > 0$ that defines the extent of anisotropy between the longitudinal and transverse motions between pairs of contacting residues. Our result, in favor of a small $f_{\text{ann}}$, suggests that the transverse motions are far less restrained energetically than the longitudinal motions, which may be explained by the high flexibility of side chains that facilitates easy accommodation to transverse motions between residues. This finding also validates the ENM as the zero order approximation to the G-ANM.

In our future studies, through proper parameterization of G-ANM (fitting the thermal fluctuations and/or the observed conformational changes with $f_{\text{ann}}$ and $R$), we will strive to probe several key aspects of conformational dynamics in proteins such as the allosteric couplings (21,26) and the ligand-binding induced conformational motions (27). It will be interesting to assess the performance G-ANM in describing the anisotropic displacement parameters from crystallography (28) or structural fluctuations from NMR data (29). Comparison with other efforts to improve GNM (for example, see Song and Jernigan (30) and Erman (31)) also will be useful.

REFERENCES