Published in final edited form as:

Arch Biochem Biophys. 2011 April 1; 508(1): 64–71. doi:10.1016/j.abb.2010.12.031.

Normal Mode Analysis with Molecular Geometry Restraints: Bridging Molecular Mechanics and Elastic Models

Mingyang Lu[†] and Jianpeng Ma^{†,‡,§}

[†]The Verna and Marrs McLean Department of Biochemistry and Molecular Biology Baylor College of Medicine One Baylor Plaza, BCM-125 Houston, TX 77030

[‡]Department of Bioengineering Rice University Houston, TX 77005

Abstract

A new method for normal mode analysis is reported for all-atom structures using molecular geometry restraints (MGR). Similar to common molecular mechanics force fields, the MGR potential contains short- and long-range terms. The short-range terms are defined by molecular geometry, i.e. bond lengths, angles and dihedrals; the long-range term is similar to that in elastic network models. Each interaction term uses a single force constant parameter, and is determined by fitting against a set of known structures. Tests on proteins/non-proteins show that MGR can produce low frequency eigenvectors closer to all-atom force-field-based methods than conventional elastic network models. Moreover, the “tip effect”, found in low frequency eigenvectors in elastic network models, is reduced in MGR to the same level of the modes produced by force-field-based methods. The results suggest that molecular geometry plays an important role, in addition to molecular shape, in determining low frequency deformational motions. MGR does not require initial energy minimization, and is applicable to almost any structure, including the one with missing atoms, bad contacts, or bad geometries, frequently observed in low-resolution structure determination and refinement. The method bridges the two major representations in normal mode analyses, i.e., the molecular mechanics models and elastic network models.

Keywords

Normal mode analysis; energy minimization; all-atom normal modes; coarse-graining; molecular geometry restraints; tip effect

Introduction

Normal mode analysis (NMA) is an important computational tool in studying vibrational motions of biomolecules [1-4]. During the past decades, many mode calculation methods have been developed, including force-field-based methods and network-based methods.

In the force-field-based methods, the molecular potential function is given by a force field, such as CHARMM [5-7] and AMBER [8-11]. Some coarse-grained schemes have also been

[§]To whom the correspondence should be addressed. Mailing address: One Baylor Plaza, BCM-125, Baylor College of Medicine, Houston, TX 77030. jhma@bcm.tmc.edu, Phone: 713-798-8187, Fax: 713-796-9438. .

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

developed to reduce computational costs, such as the Rotations-Translations of Blocks (RTB) [12] and all-atom-derived methods [13, 14]. Recently, Hendrickson group developed a coarse-grained force-field for NMA and molecular dynamics [15]. These methods can keep detailed molecular interactions, but they usually require initial energy minimization step, which in many cases distorts structures [16].

In the network-based methods, such as the coarse-grained elastic network models (ENM) [17-31], the potential functions are solely harmonic terms with equilibrium positions reside on the studied structure. Therefore, they can bypass the initial energy minimization step. However, it was found that the low frequency modes produced by these methods may not be as accurate as force-field-based methods [16, 26, 32]. In particular, some low frequency modes contain abnormally localized motions, known as “tip effect” [33]. Methods have been developed to alleviate this problem by strengthening local stiffness [33-35], or by specific coarse-graining scheme [36]. However, in node-based (or network-based) mode calculation methods, it may become subjective for the selection of nodes, especially for non-protein components, let alone the stiffness between nodes.

Recently, we developed the minimalist network model (MNM) [16, 32] to cope with issues in both conventional force-field-based methods and network-based methods. MNM utilizes a force field to calculate potential, and slightly modifies the Hessian to bypass the initial energy minimization step. Tests show that MNM outperforms both CHARMM normal mode method and all-atom elastic network model in fitting experimental anisotropic displacement parameters of crystal structures. However, in some cases, one may need to deal with low-resolution structures, which commonly contain missing atoms, bad contacts and bad geometries. In those cases, the accurate potentials and Hessian matrices are usually hard to be calculated by a molecular force field. Moreover, many structures contain some components (e.g. organic ligands) that are not conveniently defined in a molecular force field.

To solve this problem, in this study, we developed a new mode calculation method for all-atom structure using molecular geometry restraints (MGR). In addition to non-bonded interactions commonly used in the conventional ENM, the MGR Hamiltonian has harmonic restraining terms on molecular geometry, i.e. bond lengths, bond angles and dihedral angles. Each potential term in MGR uses a single force constant parameter to represent the characteristic stiffness of each interaction type. Unlike force-field-based methods, MGR immediately satisfies the harmonic approximation, allowing one to bypass the lengthy initial energy minimization step. On the other hand, unlike conventional ENM, MGR includes molecular interactions defined by molecular geometry, allowing better description of low frequency modes influenced by the details of molecular structure.

Since the calculation of MGR only requires the information of atomic positions and bond connectivity of the molecules, it is widely applicable to all kinds of macromolecules, such as proteins, DNA, lipids, small molecule ligands, etc. Moreover, MGR can be applied to low quality structures, which commonly contains missing atoms, bad contacts and bad geometries. MGR can also be applied to supramolecular complexes, by combining it with coarse-graining schemes such as RTB method [12].

Our tests show that MGR can systematically produce low frequency eigenvectors closer to the force-field-based methods than ENM. This suggests that, molecular geometry plays an important role, in addition to molecular shape [37], in determining low frequency deformational motions. The MGR method provides a bridge between two major representations in normal mode analyses, i.e., the molecular mechanics models and elastic network models.

In the following, we first describe the methodology of MGR, including the potential functions defined by the molecular geometry. Then, we show the procedure of optimization of parameters and the way to test MGR.

Material and Methods

MGR

The potential function used in MGR method has four terms:

$$V = V_{bond} + V_{angle} + V_{dihedral} + V_{non-bond}. \quad (1)$$

Unlike force-field potentials, each potential term of MGR is a harmonic potential with energy minimum on the studied structure.

The first two terms, V_{bond} and V_{angle} , denote bond length and bond angle potential. They have forms of

$$V_{bond} = \sum k_l (l - l^0)^2, \quad (2)$$

where the summation is over all chemical bonds, l and l^0 are the instantaneous and initial bond lengths; and

$$V_{angle} = \sum k_\theta (\theta - \theta^0)^2, \quad (3)$$

where the summation is over all bond angles, θ and θ^0 are the instantaneous and initial bond angles. The constants k_l and k_θ are the force constants for the bond and bond angle interactions. Here, an assumption is made that, like the non-bonded interactions in conventional ENM, all the bond length (or bond angle) interactions share the same stiffness. It should be noted that although these potential terms in MGR have the same functional forms as those in some force-fields, Equations (2) and (3) are conceptually different from the traditional bonded potentials. Specifically, l^0 and θ^0 in the MGR potential term are not equilibrium values used in molecular mechanics force fields, but values directly taken from the studied structures. It automatically ensures the MGR potential has its energy minimum at any given starting configuration. i.e., MGR does not require initial energy minimization. Such feature also provides an advantage for MGR over regular force-field-based methods when applied to low quality structures, such as the ones reconstructed from low-resolution X-ray crystallographic data, the ones modeled by structure prediction, or the ones modeled in early stages of X-ray crystallographic refinement. Since these structures usually contain defects, such as unphysical bond lengths, missing atoms, or bad contacts, MGR method would have no problem in dealing with those defective structures because all interactions of each type are modeled as uniformly distributed harmonic potentials.

The third term in Equ.1, $V_{dihedral}$ is the dihedral angle potential in a form of

$$V_{dihedral} = \sum k_\phi (\phi - \phi^0)^2, \quad (4)$$

where the summation is over all dihedral angles and improper dihedral angles, ϕ and ϕ^0 are the instantaneous and initial dihedral angles, k_ϕ is the force constant. Although the dihedral angle potential, in some cases, is modeled with a periodic function, we used a simple harmonic term to model dihedral interactions. This treatment is fine in the current study because only one potential minimum is required in the normal mode calculations. In this study, dihedral angles and improper dihedral angles were modeled with the same force

constant k_ϕ , because no apparent improvements were found when modeled with separate parameters.

The last term $V_{\text{non-bonded}}$ is the non-bonded potential in a form of

$$V_{\text{non-bonded}} = \sum_{r^0 \leq r_c} k_r (r - r^0)^2, \quad (5)$$

where r and r^0 are the instantaneous and initial distance of the non-bonded atom pairs, r_c is the cutoff distance and k_r is the force-constant. The non-bonded term is similar to the potential function of the conventional ENM except that MGR excludes the atom pairs that are involved in a chemical bond or a bond angle (1-2, 1-3 interactions). Like many other studies, 1-4 interactions are included in the potential function. Following the literature [17], the cutoff distance r_c for the non-bonded interaction is chosen as 6 Å.

One may wonder whether the non-bonded interactions should be modeled heterogeneously, i.e. to use several parameters to characterize various non-bonded interaction types. To answer this question, we tested MGR on several protein structures with a version in which hydrogen bonds were modeled with a stronger force constant. Hydrogen bond interactions were selected in the test mainly because they are crucial in stabilizing protein structures. However, no apparent improvement was found for the version with stiffened hydrogen bond interactions. Therefore, we believed that it is appropriate to model the non-bonded interactions homogeneously for the purpose of current study.

Once the potential terms are defined in Equ.1, it is then quite straightforward to perform NMA. As already mentioned, MGR does not require an initial energy minimization step. The Hessian matrix of MGR can be obtained by calculating second derivatives of each energy term according to the formulae derived in reference [38]. Although the Hessian matrices are usually highly sparse, diagonalization of the matrices is still computationally expensive for supramolecular structures. To deal with this problem, one can reduce the dimension of Hessian matrix by a coarse-graining scheme, e.g. RTB [12] and sub-structure based method [13, 14, 39]. In this study, we mainly implement and test a version of MGR coarse-grained by the RTB method, in which case each residue is modeled as a rigid-body. By implementing ARPACK package (<http://www.caam.rice.edu/software/ARPACK>) to perform matrix diagonalization, we are able to obtain eigenvectors of reduced dimension, which are then mapped back to full-dimension eigenvectors by the RTB method.

Test Set

In this study, we tested MGR on crystal structures of various biological molecules, including 10 proteins, an RNA/protein complex, a protein with a heme ligand, and a glycoprotein.

The 10 proteins include three all-helical proteins (PDB code 1exr, 1mc2 and 1xmk), three all-sheet proteins (PDB code 1nls, 1od3, 1k5c) and four α - β or α/β proteins (PDB code 3lzt, 1g55, 1n55 and 1ylj) (Table 1). These 10 protein structures are solved in ultra high resolution (higher than 1 Å) and have very few non-protein atoms. When testing on these structures, we only include amino acids in the calculation.

The test set also includes a 2.7Å-resolution RNA/protein complex (PDB code 1vbx). It is composed of a 95-residue polypeptide chain and a 73-residue viral RNA. Since the protein part and the RNA part of the 1vbx structure have similar sizes, it is chosen for testing and comparing the performance of MGR on both structural components. The test set also contains a 1.7Å-resolution deoxy form of myoglobin (PDB code 3h57), which contains a large heme ligand in the interior, and a 4Å-resolution glycoprotein gp120 (PDB code 2bf1),

which contains substantial amount (50 residues) of carbohydrates attached to protein side chains.

Conventional All-atom ENM

MGR is compared to an all-atom version of ENM initially used by Tirion [17]. The cutoff distance is also chosen as 6Å. Similar to MGR, ENM is also coarse-grained with the RTB method. This version of ENM (simply denoted as “ENM” later) is mostly the same as the elNémo method [40], except for the use of a shorter cutoff distance.

Conventional NMA with CHARMM Potential

Normal modes from a force-field-based method are found to be more accurate than those from coarse-grained network models [16, 26, 32]. In this study, the modes calculated by CHARMM force field are served as a reference. To study various types of biomolecules with CHARMM, we use the united atom CHARMM19 force field [6] for proteins and heme ligand, the CHARMM27 force field [7, 41] for nucleotides, and a modified carbohydrate force field [42] for sugars. Our modification on the carbohydrate force field includes the new definition of NAG, NDG and AFL residues, and the patches for covalent links between ASN and NAG/NDG. This hybrid force field is used together with the EEF1 solvation model [43].

Before normal mode calculations, all missing heavy atoms and hydrogen atoms are initially built with CHARMM package. Afterwards, multiple cycles of adapted-basis Newton-Raphson energy minimization are performed with decreasing harmonic constraints until root-mean-square energy gradients reach $1 \times 10^{-8} \text{ kcal} \cdot \text{mol}^{-1} \text{ Å}^{-1}$. Normal modes are then calculated on the minimized structures. Like MGR and ENM, CHARMM normal modes are coarse-grained with RTB scheme. When making comparisons, we also apply MGR and ENM on the minimized structures after truncating hydrogen atoms. Since CHARMM eigenvectors contain hydrogen atoms, each eigenvector is modified by first truncating the hydrogen part and then normalizing to unit vector.

Test Criteria

Two quantities are evaluated during the comparison. The first one is the “overlap index”, representing similarity of the lowest frequency eigenvector subspaces of two normal mode methods. The overlap index is defined as the projection of an eigenvector of one method onto the lowest frequency mode subspace of the other method:

$$P_i = \sum_j (\mathbf{v}_i \cdot \mathbf{u}_j)^2, \quad (6)$$

where $\mathbf{v}_i$, $\mathbf{u}_j$ are i th and j th eigenvectors of the two normal mode methods respectively and the summation is over all $\mathbf{u}$ eigenvectors in the subspace. A P closer to one indicates higher overlap of the mode of one method to the low frequency mode subspace of the other method, therefore higher similarity between the two methods. In this study, the overlap indices P is calculated by using the first 100 non-zero lowest frequency modes of each normal mode method. Because the truncated CHARMM modes are slightly non-orthogonal, to quantify the similarity between CHARMM modes and the modes of the other methods, we calculate P of CHARMM modes by projecting them onto the mode subspace of the other methods. In the test, P usually has large values for the first half number of low-frequency modes (50 modes), but decreases when the frequency increases (see Fig. 1). The magnitude of P dramatically decreases to less than 0.5 for many higher frequency modes. These large decreases are not mainly caused by the difference of the two mode subspaces, but are an artifact caused by the usage of limited number of modes in the calculation.. Thus, to

accurately quantify the difference between two mode subspaces, we define overall overlap index, $\bar{P}$, as the average P of the first 50 lowest frequency modes.

The second test quantity is “local character indicator” [4], used to measure the localized displacement of the eigenvectors,

$$L_i = \sum_{j=1}^n \mathbf{v}_i(j)^4, \quad (7)$$

where $\mathbf{v}_i(j)$ is j th element of i th eigenvector, n is the dimension of eigenvectors and the summation is over all vector elements. Usually, a larger L for an eigenvector indicates larger tip effect. Similarly, the average local character indicator $\bar{L}$ was defined as the average L of the 100 modes.

Results

Parameter Estimation

The potential function of MGR has four harmonic force constants k_l , k_θ , k_ϕ and K_r for the different interaction types. This form of interaction is usually valid for bonded interactions, but may not be suitable for non-bonded interactions. For example, the non-bonded interactions between two positively charged atoms can be constantly repulsive. However, for the whole system, the non-bonded interactions are overall attractive and roughly uniformly distributed [17, 19, 37]. Just like in ENM, the non-bonded interaction is also modeled with the single harmonic force constant k_r .

The four force constants can be initially estimated by

$$k_x = \frac{k_B T}{\sigma_x^2}, \quad (8)$$

where $k_B T$ is the Boltzmann constant, σ_x is the standard deviation of x , $x \in \{l, \theta, \phi, r\}$, K_x is the corresponding force constant. For simplicity, k_r is set to unity. From Equ.8, we get $k_l = \sigma_r^2 / \sigma_l^2$, $k_\theta = \sigma_r^2 / \sigma_\theta^2$ and $k_\phi = \sigma_r^2 / \sigma_\phi^2$. We first estimated σ_x from the statistics based on a high-resolution protein structure database TOP500 (<http://kinemage.biochem.duke.edu/databases/top500.php>). The values of bond lengths and bond angles follow Gaussian distribution well. The standard deviations of the common bonds N-C $^\alpha$, C $^\alpha$ -C, C-O, C-N and C $^\alpha$ -C $^\beta$ were found to be 0.013 Å, 0.013 Å, 0.011 Å, 0.011 Å and 0.016 Å. The standard deviations of the common bond angles N-C $^\alpha$ -C, C $^\alpha$ -C-N, C-N-C $^\alpha$, N-C $^\alpha$ -C $^\beta$ and C $^\beta$ -C $^\alpha$ -O were found to be 3.2°, 1.8°, 2.0°, 1.9° and 2.1°. Accordingly, the average standard deviations can be estimated as $\sigma_l \approx 0.010$ Å, $\sigma_\theta \approx 2.0^\circ / 180^\circ \cdot \pi \approx 0.035$.

On the other hand, the standard deviation of dihedral angle σ_ϕ is hard to be estimated. Compared to bond length and bond angle interactions, dihedral interactions are much weaker. So, they can be easily affected by the other interactions. For example, rotameric angles of protein side-chains can be affected by main-chain atoms. According to the statistics of backbone-dependent rotamer library [44], the deviations of the first rotameric angle χ_1 of small side-chains are roughly about 10°. So, σ_ϕ was roughly estimated as $10^\circ / 180^\circ \cdot \pi \approx 0.175$. Since this σ_ϕ value is only a rough estimation, it needed to be optimized later.

The standard deviation of the non-bonded interactions σ_r is also difficult to be estimated, because the interactions have complicated nature, including Van der Waals, electrostatic and solvation interactions etc. Besides, the potential are not necessarily quadratic. In this study,

we fit harmonic potential to the local minimum of an atomic pairwise knowledge-based potential DFIRE [45],

$$u(r) = u_0(r) + k_B T \log \frac{N_{obs}(r)}{r^{1.61}}, \quad (9)$$

where r is the distance of atom pairs, $u(r)$ is the potential at distance r , $u_0(r)$ is a constant potential shift, $N_{obs}(r)$ is the observed counts of atom pairs whose distance is in the range of $[r, r + \Delta r)$ from the non-redundant structure library TOP500. The constant σ_r is estimated by the force constants calculated by fitting harmonic function to the DFIRE potential. In the study, we found that the DFIRE potentials for the hydrophobic atom pairs have well-characterized minima near the first coordinate shell, for example, the atom pairs $C^\beta(ALA)-C^\beta(ALA)$, $C^\delta(ILE)-C^\delta(LEU)$, $C^\delta(LEU)-C^\delta(LEU)$ and $C^\gamma(VAL)-C^\gamma(VAL)$. Their force constants were estimated to be $5.47 k_B T$, $4.79 k_B T$, $3.89 k_B T$ and $3.89 k_B T$. On average, they correspond to $\sigma_r \approx 0.48 \text{ \AA}$. It should be mentioned that for many other atom pairs, the DFIRE potentials do not have well-defined potential wells for harmonic potentials to fit with. Thus, the actual σ_r should be larger than the estimated value. Similar to σ_ϕ , σ_r also requires further optimization.

According to the above estimation, $k_r = 1$, $k_l = 2.3 \times 10^3$, $k_\theta = 1.9 \times 10^2$ and $k_\phi = 7.5$. The initial parameters are tested on the 10 proteins in the test set. Fig.1 shows the comparison of MGR with ENM and CHARMM on the protein 1nls. Even with the initial estimated parameters, the lowest frequency mode subspace of MGR has consistently higher overlap with the CHARMM modes than the subspace of ENM, indicating MGR modes are always closer to CHARMM modes than ENM (upper panel in Fig.1). MGR vectors also have similar magnitude of the local character indicators as CHARMM vectors (only a few of them are slightly higher), while ENM vectors have much larger local character indicators, indicating ENM has larger tip-effect and MGR has smaller.

Parameter Optimization

The initial estimated σ_ϕ and σ_r are further optimized on the 10 protein test set. Two variable α and β are used to optimize the four parameters of MGR:

$$\begin{aligned} k_r' &= k_r \\ k_l' &= \alpha k_l \\ k_\theta' &= \alpha k_\theta \\ k_\phi' &= \beta k_\phi \end{aligned} \quad (10)$$

where ' denotes the scaled parameters. As shown in Equ.10, α scales the force constants of bond length and bond angles, while β scales that of dihedral angles. From Equ.8 and 10,

$$\sigma_r' = \sqrt{\alpha} \sigma_r, \quad \sigma_\phi' = \sqrt{\alpha/\beta} \sigma_\phi.$$

The values for α and β are optimized against the overall overlap index $\bar{p}$ for MGR and CHARMM eigenvectors on the 10 proteins. All the protein cases have similar trend of $\bar{p}$ with respect to α and β . As shown in Fig.2 for the protein 1nls, larger α usually corresponds to larger $\bar{p}$. But $\bar{p}$ becomes saturated after α reaches about 3 or 4. It suggests that the interactions of bond length and bond angle can affect the lowest frequency modes when the force constants are moderate; while overestimating the force constants only affects the high frequency modes. On the other hand, increasing β largely improves $\bar{p}$ until it reaches maximum. Further increasing β deteriorates $\bar{p}$. It suggests that the interactions of dihedral angles have important roles to the lowest frequency modes. Too weak and too strong dihedral interactions can both worsen the quality of the lowest frequency modes.

$\bar{L}$ as the function of α and β also has a similar trend for these proteins. Increasing of α or β in general decreases $\bar{L}$. However, when α is overestimated, $\bar{L}$ becomes saturated. On the other hand, increasing β always decreases $\bar{L}$. But overestimating β results in less localization in the MGR modes than that in the CHARMM modes, indicating over-restraint of the low-frequency motions. That also explains why $\bar{p}$ is not as good when β is too large.

The optimization results for each protein are shown in Table 1. Best average $\bar{p}$ is obtained when α equals to 4.4 and β equals to 8.3. The corresponding force constants are $k_r = 1$, $k_l = 1.0 \times 10^4$, $k_\theta = 8.2 \times 10^2$ and $k_\phi = 6.2 \times 10^1$. The optimized parameters also help us to reevaluate the standard deviation of non-bonded distances and dihedral angles. Compared to the estimated values, σ_r increases to $1.0 \text{ \AA}$, σ_ϕ reduces to 7.3° . Fig.3 shows the results of MGR with the optimized parameters. For all the 10 proteins, the MGR modes are consistently closer to the CHARMM modes than the ENM modes. Besides, the MGR modes have almost the same average local character indicators as the CHARMM modes, while the ENM modes have consistently higher values. The comparison of the initial parameters and the optimized parameters are shown in Fig.1 for the protein 1nls. The overlap indices of the initial MGR are already substantially higher than that of ENM, while the optimized MGR further improves the overlap indices, but only slightly (upper panel in Fig.1). For the local character indicators, the high values found in some modes of the MGR version with the initial parameters disappear in the version with the optimized parameters (lower panel in Fig.1).

Although the MGR parameters are only optimized against $\bar{p}$, the MGR modes calculated from the version with the optimized parameters have almost the same $\bar{L}$ as the CHARMM modes. This indicates that the geometry restraints in MGR eigenvectors help to eliminate tip effects.

Tests on Nucleotides, Ligands and Carbohydrates

The optimized MGR is further tested on some biomolecules that contain non-protein components.

MGR is first tested on a protein/RNA complex 1vbx. Tests are made on three structures: the protein chain alone, the RNA chain alone, and the whole structure. Each structure is first energy-minimized with CHARMM potential. Then the minimized structures are used to perform various NMAs. As shown in Fig.4, the MGR mode subspace has consistently higher overlaps to the CHARMM modes than the ENM mode subspace on all the three structures. It is noted that the improvements of MGR over ENM on the RNA structure is slightly smaller than that on the protein structure. For the local character indicators, the magnitude of the MGR vectors is similar to that of the CHARMM vectors, and is less than that of the ENM vectors. Hence, the tip effect is successfully reduced in all three cases.

MGR is also tested on the protein/ligand structure 3h57 and the glycoprotein 2bf1. As shown in Fig.5, the results on these structures are similar to the previous protein tests.

Tests on All-atom MGR

Since MGR method uses an all-heavy-atom potential, the performance is also evaluated on the 10 protein structures for the normal modes calculated from all-atom Hessian (instead of using coarse-grained RTB method in previous sections). As shown in Fig.6 (upper panel), the MGR modes are also consistently closer to the CHARMM modes than the ENM modes. Similarly, the local character indicators of the MGR modes are roughly same as those of the CHARMM modes, while those of the ENM modes have much higher values (lower panel in Fig.6). Comparing to the methods using the RTB coarse-graining scheme, all-atom MGR

mode subspace has less overlaps to the CHARMM mode subspace (upper panel in Fig.6). All-atom modes also have higher local character indicators, because the RTB scheme restrains the motions of the atoms in a residue, thus suppressing localized motions in the normal modes. Moreover, it was also found in our previous study that RTB CHARMM modes can fit experimental anisotropic displacement parameters better than the all-atom CHARMM modes [16, 32]. Improvements of MGR over ENM are also compared for the RTB-coarse-grained method and the all-atom method. The improvements of the overlap indices are slightly larger for the all-atom method, except for the protein 1od3. Surprisingly, the improvements of the local character indicators are about 10 times larger for the all-atom method (lower panel in Fig.6). This indicates that all-atom ENM without the RTB scheme contains much larger tip effect, but can be dramatically reduced in the all-atom MGR.

Discussion

In this study, we have developed a new normal mode analysis method for all atom structures with molecular geometry restraints (MGR). In this method, the short-term interactions adopt functional forms similar to those in molecular mechanics force fields, while the long-range term is similar to that of elastic network models (ENM). In order to bypass the initial energy minimization, the minima of all potential terms are set at the studied structures. MGR has four harmonic potential terms for bond lengths, bond angles, dihedral angles and non-bonded interactions. Each of these potential terms has a uniformly distributed force-constant to represent the strength of each interaction type. Those force constants are determined by fitting against a set of training structures.

Unlike common ENMs, in which the network structures only characterize molecular shape and mass distribution, MGR employs more detailed structure information defined by the molecular geometry, such as bond, angle and dihedral. According to the tests, MGR can substantially improve the performance of mode calculation by producing closer low-frequency modes, than conventional network-based methods, to those calculated by a force-field-based method (CHARMM NMA in this study). Our results suggest that molecular geometry plays an important role, in addition to molecular shape [37], in determining low frequency deformational motions. Such an improved representation in molecular interactions also contributes to the reduction of mode localization, i.e., the tip effect, observed in earlier work [16, 33]. Therefore, MGR method bridges the two major representations in normal mode analyses, i.e., the molecular mechanics models and elastic network models.

The four force constant parameters in MGR are directly related to the standard deviations of bond lengths, bond angles, dihedral angles and non-bonded interactions. The corresponding standard deviations σ_b , σ_θ , σ_ϕ and σ_r are found to be 0.01Å, 2.0°, 7.3° and 1.0Å, respectively. These four types of interactions not only have distinct stiffness, but also have different roles to the low frequency deformations. The strength of the bond lengths and bond angles interactions should exceed certain threshold for better calculating the low frequency modes; whereas only moderate range of the dihedral angles interactions is good for the low frequency modes.

MGR can be essentially applied to any macromolecular structure as long as the molecular geometry is pre-defined, e.g., proteins, DNA molecules or small ligands. MGR is a favorable method for structural modeling/refinement at low-resolutions, such as for normal-mode-based X-ray refinement [46-50] as MGR can handle missing atoms in the structures, and can tolerate bad contacts and bad geometries. We hope that MGR will be a useful addition to the tool kit of NMA of biological molecules.

Acknowledgments

The authors acknowledge support of grants from the National Institutes of Health (R01-GM067801), the National Science Foundation (MCB-0818353), and the Welch Foundation (Q-1512). We thank Robert Jernigan for insightful comments on the manuscript.

Abbreviations

NMA	Normal Mode Analysis
MGR	Molecular Geometry Restraints
ENM	Elastic Network Model
MNM	Minimalist Network Model
RTB	Rotations-Translations of Blocks
CHARMM	Chemistry at HARvard Molecular Mechanics
AMBER	Assisted Model Building with Energy Refinement

References

- [1]. Levitt M, Sander C, Stern PS. *J. Mol. Biol.* 1985; 181:423–447. [PubMed: 2580101]
- [2]. McCammon, JA.; Harvey, S. *Dynamics of Proteins and Nucleic Acids*. Cambridge University Press; Cambridge: 1987.
- [3]. Brooks CL III, Karplus M, Pettitt BM. *Adv. Chem. Phys.* 1988; 71:1–249.
- [4]. Brooks BR, Janezic D, Karplus M. *J. Comput. Chem.* 1995; 16:1522–1542.
- [5]. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. *J. Comput. Chem.* 1983; 4:187–217.
- [6]. Neria E, Fischer S, Karplus M. *J. Chem. Phys.* 1996; 105:1902–1921.
- [7]. MacKerell AD, Bashford D Jr, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher I, E. W, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. *J. Phys. Chem.* 1998; B102:3586–3616.
- [8]. Wang JM, Cieplak P, Kollman PA. *Journal of Computational Chemistry*. 2000; 21:1049–1074.
- [9]. Ponder JW, Case DA. *Adv Protein Chem.* 2003; 66:27–85. [PubMed: 14631816]
- [10]. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. *J Comput Chem.* 2004; 25:1157–1174. [PubMed: 15116359]
- [11]. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ. *J Comput Chem.* 2005; 26:1668–1688. [PubMed: 16200636]
- [12]. Tama F, Gadea FX, Marques O, Sanejouand YH. *Proteins*. 2000; 41:1–7. [PubMed: 10944387]
- [13]. Ming D, Wall ME. *Physical Review Letters*. 2005; 95
- [14]. Zhou L, Siegelbaum SA. *Biophys J.* 2008; 94:3461–3474. [PubMed: 18212016]
- [15]. Korkut A, Hendrickson WA. *Proceedings of the National Academy of Sciences*. 2009; 106:15667–15672.
- [16]. Lu M, Ma J. *Proc Natl Acad Sci U S A.* 2008; 105:15358–15363. [PubMed: 18832168]
- [17]. Tirion MM. *Phys. Rev. Lett.* 1996; 77:1905–1908. [PubMed: 10063201]
- [18]. Haliloglu T, Bahar I, Erman B. *Phys. Rev. Lett.* 1997; 79:3090–3093.
- [19]. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. *Biophys J.* 2001; 80:505–515. [PubMed: 11159421]
- [20]. Doruker P, Jernigan RL, Bahar I. *J. Comput. Chem.* 2002; 23:119–127. [PubMed: 11913377]
- [21]. Kurucuoglu O, Jernigan RL, Doruker P. *Qsar & Combinatorial Science*. 2005; 24:443–448.
- [22]. Eom K, Baek SC, Ahn JH, Na S. *J Comput Chem.* 2007; 28:1400–1410. [PubMed: 17330878]

- [23]. Ming D, Kong Y, Lambert M, Huang Z, Ma J. *Proc. Natl. Acad. Sci. USA*. 2002; 99:8620–8625. [PubMed: 12084922]
- [24]. Tama F, Wriggers W, Brooks CL. *J Mol Biol*. 2002; 321:297–305. [PubMed: 12144786]
- [25]. Hinsin K. *Journal of Computational Chemistry*. 2000; 21:79–85.
- [26]. Kondrashov DA, Van Wynsberghe AW, Bannen RM, Cui Q, Phillips GN Jr. *Structure*. 2007; 15:169–177. [PubMed: 17292835]
- [27]. Lyman E, Pfaendtner J, Voth GA. *Biophysical Journal*. 2008; 95:4183–4192. [PubMed: 18658214]
- [28]. Moritsugu K, Smith JC. *Biophysical Journal*. 2007; 93:3460–3469. [PubMed: 17693469]
- [29]. Woodcock HL, Zheng W, Ghysels A, Shao Y, Kong J, Brooks BR. *The Journal of Chemical Physics*. 2008; 129:214109. [PubMed: 19063546]
- [30]. Zheng W. *Biophysical Journal*. 2010; 98:3025–3034. [PubMed: 20550915]
- [31]. Romo TD, Grossfield A. *Proteins*. 2011; 79:23–34. [PubMed: 20872850]
- [32]. Lu, M.; Ma, J. *Proteins: Energy, Heat and Signal Flow*. Leitner, DM.; Straub, JE., editors. Vol. 2009. CRC Press; 2009. p. 229–245.
- [33]. Lu M, Poon B, Ma J. *J. Chem. Theor. Comp*. 2006; 2:464–471.
- [34]. Stember JN, Wriggers W. *Journal of Chemical Physics*. 2009; 131
- [35]. Mendez R, Bastolla U. *Physical Review Letters*. 2010; 104:228103. [PubMed: 20867208]
- [36]. Gohlke H, Thorpe MF. *Biophysical Journal*. 2006; 91:2115–2120. [PubMed: 16815893]
- [37]. Lu M, Ma J. *Biophys. J*. 2005; 89:2395–2401. [PubMed: 16055547]
- [38]. Tuzun RE, Noid DW, Sumpter BG. *Macromolecular Theory and Simulations*. 1996; 5:771–788.
- [39]. Hafner J, Zheng W. *The Journal of Chemical Physics*. 2009; 130:194111–194117. [PubMed: 19466825]
- [40]. Suhre K, Sanjoud YH. *Nucleic Acids Res*. 2004; 32:W610–614. [PubMed: 15215461]
- [41]. Mackerell AD, Feig M, Brooks CL. *Journal of Computational Chemistry*. 2004; 25:1400–1415. [PubMed: 15185334]
- [42]. Guvench O, Greene SN, Kamath G, Brady JW, Venable RM, Pastor RW, A.D.M. *Journal of Computational Chemistry*. 2008; 29:2543–2564. [PubMed: 18470966]
- [43]. Lazaridis T, Karplus M. *Proteins*. 1999; 35:133–152. [PubMed: 10223287]
- [44]. Dunbrack RL Jr. Cohen FE. *Protein Sci*. 1997; 6:1661–1681. [PubMed: 9260279]
- [45]. Zhou H, Zhou Y. *Protein Sci*. 2002; 11:2714–2726. [PubMed: 12381853]
- [46]. Ni FY, Poon BK, Wang QH, Ma JP. *Acta Crystallographica Section D-Biological Crystallography*. 2009; 65:633–643.
- [47]. Chen XR, Poon BK, Dousis A, Wang QH, Ma JP. *Structure*. 2007; 15:955–962. [PubMed: 17698000]
- [48]. Poon BK, Chen XR, Lu MY, Vyas NK, Quirocho FA, Wang QH, Ma JP. *Proc Natl Acad Sci U S A*. 2007; 104:7869–7874. [PubMed: 17470791]
- [49]. Chen X, Lu M, Poon BK, Wang Q, Ma J. *Acta Crystallogr D Biol Crystallogr*. 2009; 65:339–347. [PubMed: 19307715]
- [50]. Chen X, Wang Q, Ni F, Ma J. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107:11352–11357. [PubMed: 20534430]

Research Highlights

1. A new method for normal mode analysis for all-atom structure using molecular geometry restraints (MGR).
2. Producing low frequency eigenvectors closer to all-atom force-field-based methods than conventional elastic network models.
3. Reducing the “tip effect”, found in low frequency eigenvectors in elastic network models, to the same level of the modes produced by force-field-based method.
4. No initial energy minimization is required.
5. Applicable to almost any biomolecular structure, especially that in low resolution, in which case poor geometry (e.g., broken bonds or dihedrals) and bad contacts are common.
6. Showing the important role of molecular geometry, in addition to molecular shape, in determining low-frequency modes.
7. A bridge between two major representations in normal mode analyses, the molecular mechanics models and elastic network models.

Fig.1.

Comparison of the various normal mode methods on the protein 1nls. Upper Panel: the overlap indices of the first 100 lowest frequency CHARMM eigenvectors onto the first 100 lowest frequency subspaces of various normal mode methods. Lower Panel: the local character indicators of the eigenvectors (logarithmic scale). The x-axis is the mode index of the non-zero frequency modes sorted by the value of the frequency, ranging from 7 to 106. In both panels, open squares denote the ENM modes, plus signs denote the MGR modes with the initial parameters, solid circles denote the MGR modes with the optimized parameters, and triangles denote the CHARMM modes.

Fig.2. Performance of MGR as the function of the scaling parameters α and β for the protein 1nls. The two upper panels show the overall overlap indices of the CHARMM modes onto the first 100 lowest frequency subspaces of MGR. The two lower panels show the average local character indicators of the MGR modes.

Fig.3.

Comparison of the various normal mode methods on the 10 proteins in the test set. The upper panel shows the overall overlap indices of the CHARMM modes onto the first 100 lowest frequency subspaces of MGR (solid circles) or ENM (open squares). The lower panel shows the average local character indicators of the MGR (solid circles), the ENM (open squares) and the CHARMM (triangles) modes. The x-axis lists all the 10 proteins.

Fig.4.

Comparison of the various normal mode methods on the protein/RNA complex 1vbx. The three left panels show the overlap indices of the first 100 lowest frequency CHARMM modes onto the first 100 lowest frequency subspaces of MGR (solid circles) or ENM (open squares). The three right panels show the local character indicators (logarithmic scale) of the MGR (solid circles), the ENM (open squares) and the CHARMM (triangles) modes. The two upper panels are for the whole structure 1vbx; the two middle panels are for the protein structure; the two lower panels are for the RNA structure.

Fig.5.

a) Comparison of the various normal mode methods on the hem-bound myoglobin (PDB code 3h57). The notation is the same as Fig.4. b) Comparison of the various normal mode methods on the glycoprotein gp120 (PDB code 2bf1). The notation is the same as Fig.4.

\$watermark-text

\$watermark-text

\$watermark-text

Fig.6. Comparison of the various all-atom normal mode methods on the 10 proteins in the test set. The notation is the same as Fig.3.

Table 1

The optimization results of MGR for each protein. The two parameters α and β were optimized against the average overlap index $\bar{p}$ for each protein. The last line of the table shows the optimized parameters α and β

PDB code	Type	# of residues	Optimal α	Optimal β	Optimal $\bar{P}$
lexr	all- α	146	4.8	15.6	0.965
lmc2	all- α	122	4.0	10.1	0.957
lxmk	all- α	79	2.9	7.3	0.977
lnls	all- β	237	3.3	8.3	0.942
lod3	all- β	132	4.8	9.2	0.944
lk5c	all- β	333	5.1	5.5	0.913
3lzt	α - β	129	3.7	10.1	0.947
lg66	α / β	207	2.9	6.4	0.939
ln55	α / β	249	4.0	14.7	0.929
lylj	α / β	263	4.4	7.3	0.943
Overall			4.4	8.3	0.945