


2016

# Investigating the effects of different force fields on spring-based normal mode analysis

Jaekyun Song  
*Iowa State University*

Follow this and additional works at: <https://lib.dr.iastate.edu/etd>

 Part of the [Bioinformatics Commons](#), [Biophysics Commons](#), and the [Computer Sciences Commons](#)

---

## Recommended Citation

Song, Jaekyun, "Investigating the effects of different force fields on spring-based normal mode analysis" (2016). *Graduate Theses and Dissertations*. 15812.  
<https://lib.dr.iastate.edu/etd/15812>

This Thesis is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

**Investigating the effects of different force fields on spring-based normal mode  
analysis**

by

**Jaekyun Song**

A thesis submitted to the graduate faculty  
in partial fulfillment of the requirements for the degree of  
MASTER OF SCIENCE

Major: Computer Science

Program of Study Committee:

Guang Song, Major Professor

Xiaoqiu Huang

Robert Jernigan

Iowa State University

Ames, Iowa

2016

Copyright © Jaekyun Song, 2016. All rights reserved.

## DEDICATION

To my wife Sunae Kim, my daughter Sophie and my son Kevin.

## TABLE OF CONTENTS

<b>LIST OF TABLES</b> . . . . .	v
<b>LIST OF FIGURES</b> . . . . .	vi
<b>ACKNOWLEDGEMENTS</b> . . . . .	vii
<b>ABSTRACT</b> . . . . .	viii
<b>CHAPTER 1. OVERVIEW</b> . . . . .	1
1.1 Introduction . . . . .	1
1.2 Background . . . . .	3
1.2.1 NMA . . . . .	3
1.2.2 ENM . . . . .	4
1.2.3 GNM . . . . .	5
1.2.4 ANM . . . . .	6
1.2.5 sbNMA and ssNMA . . . . .	7
1.2.6 CHARMM . . . . .	8
1.2.7 AMBER . . . . .	9
<b>CHAPTER 2. METHODS</b> . . . . .	10
2.1 The protein dataset . . . . .	10
2.2 Mean square fluctuations (MSF) and normal mode frequency . . . . .	10
2.3 Single parameters of ssNMA . . . . .	10
2.4 Converting eigenvalues to vibrational frequencies . . . . .	11

<b>CHAPTER 3. RESULTS</b> . . . . .	<b>13</b>
3.1 Evaluation of simplified models by MSF correlations . . . . .	13
3.2 Universality of vibrational spectrum . . . . .	14
3.3 Spectra of three protein groups by different protein folds . . . . .	15
3.4 Evaluation of sbNMA and ssNMA on vibrational spectra . . . . .	18
3.5 Effect of input structure on vibrational spectrum . . . . .	21
3.6 Comparisons between CHARMM and AMBER . . . . .	24
<b>CHAPTER 4. CONCLUSIONS</b> . . . . .	<b>31</b>
<b>BIBLIOGRAPHY</b> . . . . .	<b>34</b>

**LIST OF TABLES**

Table 2.1	Parameters used in ssNMA for CHARMM and AMBER94 . . . . .	11
Table 3.1	MSF correlations of ssNMA and sbNMA with NMA . . . . .	14

## LIST OF FIGURES

Figure 3.1	Distribution of MSF correlation for sbNMA with CHARMM and AMBER	15
Figure 3.2	Scatter plot of the RMSD versus the corresponding MSF correlation of sbNMA . . . . .	16
Figure 3.3	Universality of the density of vibrational modes of NMA with AMBER	17
Figure 3.4	Amide groups in vibrational spectra of NMA for different protein folds	19
Figure 3.5	Vibrational spectra on NMA and two simplified model, sbNMA and ssNMA . . . . .	20
Figure 3.6	Vibrational of the density of vibrational modes of sbNMA and ssNMA with AMBER. . . . .	22
Figure 3.7	Amide groups in vibrational spectra for sbNMA and ssNMA . . . . .	23
Figure 3.8	Vibrational spectra on different input structures . . . . .	25
Figure 3.9	Distribution of the MSF correlation between CHARMM and AMBER on original protein structures. The average coefficient is 0.93. . . . .	26
Figure 3.10	Similarity of minimizations by CHARMM and AMBER . . . . .	27
Figure 3.11	RMSD triangle of PDB structure and minimized structures . . . . .	28
Figure 3.12	Vibrational spectra of sbNMA on original PDB structures with CHARMM and AMBER. . . . .	30

## ACKNOWLEDGEMENTS

I would like to take this opportunity to express my thanks to those who helped me with various aspects of conducting research and the writing of this thesis.

First of all, a sincere thanks to Dr. Guang Song, my major advisor, for advising me on how to do research in computer science. His insights and words of encouragement have often helped me pursuing master degree appropriately. I also would like to thank my committee members, Dr. Xiaoqiu Huang and Dr. Robert Jernigan, for their sincere advice and guidance.

Lastly, I would like to thank my brother Ikkyun, his wife Hwahyun, my nephew Seonu, my father Youngchan, and my mother Seokjeom Kang. Without their lovely supports, I would not have been able to complete this work.



## ABSTRACT

Classical normal mode analysis (CNMA) has been widely acknowledged as one of the most useful simulation tools for studying protein dynamics. CNMA uses a fine-grained all-atom model of proteins and a complex empirical potential. In addition, CNMA requires a structure that must be energetically minimized, which makes the method cumbersome to use, especially for large proteins. In contrast, elastic network models (ENM) use coarse-grained protein models and adopt a simplified potential function. ENM is much faster than CNMA but is less accurate. To take the advantages of both CNMA and ENM, the spring-based normal mode analysis (sbNMA) was developed. It uses a fine-grained all-atom model for proteins and an all-atom empirical force field to maintain accuracy while reducing the computing complexity by eliminating the minimization step. In the previous work on sbNMA, only the CHARMM force field was explored. In this work, we extend the analyses to AMBER, another widely-used force field. We investigate the dependence of sbNMA's performance on force fields. This work provides also insightful understandings of the differences between CHARMM and AMBER.

## CHAPTER 1. OVERVIEW

### 1.1 Introduction

Most proteins have their own unique structures and functions. Many evidences have been provided that show the structure of a protein is directly linked to its function (Henzler-Wildman 2007). Therefore, protein structure and dynamics studies are important for understanding how a protein interacts with itself as well as with other proteins. A protein may have several distinct conformations. The native state of a protein is where it has the global energy minimum.

In the last few decades, simulations for the molecular dynamics (MD) of biomolecules are becoming increasingly popular due to the increasing computing power and many impressive results they have delivered. As a result, more complex systems are being simulated. However, many large proteins or protein complexes still require too much computational resources when it comes to all-atom structure models and all-atom force fields. Normal mode analysis (NMA) is another powerful tool for analyzing protein dynamics (Brook 1983, Go 1983, and Levitt 1983). NMA makes use of all-atom empirical energy potentials such as CHARMM, AMBER and so on. It is powerful but it requires energy minimizations in which many iterations of time-consuming computations are necessary.

Therefore, many efforts have been put into developing simplified models, regarding which a major milestone was Tirion's seminal work (Tirion 1996), which showed that a simple Hookean-like potential was sufficient to reproduce the slow dynamics of proteins. Elastic network model (ENM) further simplified Tirion's model by employing also a simplified structure model. Thus, ENM uses a coarse-grained structure model and a very simplified potential. In ENM, each residue is commonly represented by one node, usually using the alpha carbon. Two alpha carbons are considered interacting if their separation distance is less than a pre-set cutoff

distance. A protein is viewed as a network consisting of residues and Hooke springs that interconnect them (Bahar 1997). One major benefit of this simplification is that ENM does not require the energy minimization step. Consequently, one can obtain vibrational modes quickly and with surprisingly high accuracy (Bahar 2005 and Chennubhotla 2005). Due to its simplicity and fast calculation, it has been widely used (Bahar 1997 and Atilgan 2001).

The first ENM model is Gaussian network model (GNM). It gives fairly accurate predictions of mean square fluctuations, and there is usually a strong correlation between X-ray crystallographic B-factors and GNM predictions (Bahar 1997 and Kundu 2002). However, the problem of GNM is that the directional information of the atomic motions gets lost because GNM modes only have scalar amplitudes rather than vectors. That was one major reason why anisotropic network model (ANM) was developed (Atilgan 2001). With ANM, low frequency modes, which correspond to the motions of structural domains, can be obtained (Atilgan 2001). Because both GNM and ANM have their pros and cons, attempts to unify them were made (Zheng 2008 and Na 2014).

However, according to a recently published paper by Na and Song (2014), ANM still loose significant accuracy when compared to NMA. They evaluated the accuracy of simplified models by measuring correlations of mean squared fluctuations (MSF) between NMA and simplified models such as ANM. ANM was found to correlate poorly with NMA. They then suggested a few simplified models that are derived directly from NMA and correlate well with NMA. Their models use a full-atom empirical force field CHARMM in order to preserve the accuracy. One significant advantage of their model over NMA is that it does not require a computation-intensive minimization step that NMA has to go through. In addition, Na and Song (2016) presented method that can potentially be used to tune parameters in empirical potentials. However, their works were done only with CHARMM force field. Here, I revisit their studies with another widely used force field, AMBER, in order to evaluate their simplified models with a different force field, as well as providing insights in the major differences between two force fields.

## 1.2 Background

### 1.2.1 NMA

The normal mode analysis (NMA) is a technique that provides an analytical description of a system's dynamics around its equilibrium state. The main reason why NMA is a major tool for protein dynamics is that global modes at low frequencies are closely linked to functional motions of proteins. A normal mode refers a motion by which all atoms of proteins are harmonically fluctuating with the same frequency in vicinity of their equilibrium (Goldstein 2001). Normal modes can be obtained from the eigenvalues and eigenvectors by the eigen-decomposition of Hessian matrix that is the second partial derivative of the potential function. The normal mode motions have small amplitudes, which makes them unable to cross energy barriers but rather reside mostly in a potential well. The bottom of the well is referred to as the equilibrium state where the net forces of the system becomes zero. The major disadvantage of NMA is that it assumes a simple harmonic form of the potential, which causes NMA to be valid only in the immediate vicinity to the equilibrium. Therefore, the more an input structure is displaced from the equilibrium, the more uncertain the harmonic approximation holds.

Since NMA was applied to small molecules (Wilson Jr 1955, Shimanouchi 1970 and Itoh 1970), it has been extensively used in chemistry. NMA was introduced for the normal mode analysis of protein independently by Levy (1979) Brooks (1983), Go (1983) and Levitt (1983). The first application of NMA on biological system was on peptide by Levy (1979) followed by small protein bovine pancreatic trypsin inhibitor which has 58 amino acid residues (Brooks 1983, Go 1983 and Levitt 1983). Larger proteins were studied later, such as crambin (Levitt 1985), human lysozyme (Levitt 1985 and Brooks 1985), ribonuclease (Levitt 1985), and myoglobin (Seno 1990).

Theses studies used the same existing empirical potential functions that were used in molecular dynamics simulations. The first three terms corresponds to the internal degree of freedoms: bond stretching, bond angle bending, and torsional interactions, and the last nonbonded inter-

actions for van der Waals and electrostatic interactions.

$$\begin{aligned}
 V = & \frac{1}{2} \sum_{bonds} K_b(b - b_0)^2 + \frac{1}{2} \sum_{angles} K_\theta(\theta - \theta_0)^2 \\
 & + \frac{1}{2} \sum_{torsions} K_\phi[1 + \cos(n\phi - \delta)] + \sum_{nonbonded} K_\phi \left[ \frac{A}{r^{12}} - \frac{B}{r^6} + \frac{q_1 q_2}{Dr} \right]
 \end{aligned}
 \tag{1.1}$$

NMA calculates the global modes of molecules based on the equilibrium state. Therefore, minimization is required for the starting structure. This minimization consumes significant amount of time and computing resources. It is worth noting also that the minimized structure deviates from the original X-ray crystallographic structure (Ma 2005). The global modes of proteins are not sensitive to local interactions or to specific force field energy functions. Therefore, we can get the global scale flexibility of protein structures regarding motion directions and frequencies. To be specific, we can have both the most favorable direction of motion, which requires the least energy to move, and measure its stiffness (frequency).

Once the structure is minimized, a  $3N \times 3N$  Hessian matrix, where  $N$  is the number of atoms, can be constructed by taking the second derivation of the potential energy with respect to the generalized coordinates  $q_i$ :

$$H_{ij} = \frac{\partial^2 V}{\partial q_i \partial q_j}
 \tag{1.2}$$

NMA has been widely used to study proteins since 1980s (Go 1983, Brooks 1983 and Levitt 1983). In the studies of Go (1983) and Levitt (1983), only the torsional angle has been accounted while fixing bond lengths and bond angles for small protein Bovine Pancreatic Trypsin Inhibitor, which reduced the degree of freedom significantly. As a result, Hessian matrix size was shrunked without affecting low frequency vibrational modes too much. In the same year, Brooks and Karplus (Brooks 1983) investigate the same protein but with all degrees of freedom. For all atom force-field potentials, empirical force fields have been developed, such as CHARMM (MacKerell 1998) and AMBER (Wang 2000).

### 1.2.2 ENM

The insensitivity of NMA global modes to fine-grained local interactions and the type of force field being used results in the adoption of elastic network model (ENM) for protein

dynamics. While NMA uses fine-grained all-atom models, ENM models proteins using coarse-grained alpha carbon atoms. In addition, a simplified potential function is used. In other words, ENM takes into account mostly of the topology of protein structure especially the connectivity of atoms. This network representation allows us to implement efficient computational algorithms that can carry for normal mode computations. As a result, we can scale up the system being studied to very large biomolecular complexes. The evaluation of ENM is mainly done by comparisons of RMS fluctuation predictions to X-ray crystallographic B-factors. ENM is mainly divided into gaussian network model (GNM) (Bahar 1997) and anisotropic network model (ANM) (Atilgan 2001)

### 1.2.3 GNM

GNM is inspired by Tirion’s study (1996) and built upon the molecular theory of elasticity (Eichinger 1972 and Flory 1985) for polymer networks. In GNM, the fluctuation  $\Delta r_{ij}$  in the separation  $r_{ij} = |r_j - r_i|$  between two alpha atoms follows a Gaussian distribution around their equilibrium coordinates (Bahar 1997 and Haliloglu 1997). The protein structure is represented as a network of alpha-carbon nodes and elastic springs. The connectivity of alpha-carbon atoms are described by the  $N \times N$  Kirchhoff matrix  $\Gamma_{ij}$ :

$$\Gamma_{ij} = \begin{cases} 0 & \text{if } i \neq j \text{ and } r_{ij} > r_c \\ -\gamma_{ij} & \text{if } i \neq j \text{ and } r_{ij} \leq r_c \\ \sum_{i \neq j} \Gamma_{ij} & \text{if } i = j \end{cases} \quad (1.3)$$

where the cutoff distance  $r_c$  for inter-residue interactions is usually set to be  $7.3 \text{ \AA}$  between  $i$ th and  $j$ th alpha-carbon atoms in the native structure. The diagonal elements  $\Gamma_{ii}$  are the negative sum of the off-diagonal elements of  $i$ th row or  $i$ th column because the Kirchhoff is symmetric. The nodes are connected by a uniform spring of force constant  $\gamma_{ij}$ , which reduce the Kirchhoff to an adjacency matrix. The Kirchhoff matrix controls the fluctuations in alpha carbon positions and their cross-correlations. The potential function for GNM,  $V_{GNM}$ , is define

(Bahar 1997):

$$\begin{aligned} V_{GNM} &= \frac{1}{2} \sum_{ij} \gamma_{ij} (\mathbf{r}_{ij} - \mathbf{r}_{ij}^0)^2 \\ &= \frac{1}{2} \sum_{ij} \gamma_{ij} (\Delta \mathbf{r}_{ij})^2 \end{aligned} \quad (1.4)$$

where  $\mathbf{r}_{ij}^0$  is the position vectors of alpha-carbons in the X-ray Crystallographic structures at equilibrium, and  $\mathbf{r}_{ij}$  is instantaneous position vectors of the residues.

From the inverse of the Kirchhoff matrix, statistically expected value of fluctuations,  $\langle \Delta r_i^2 \rangle$ , and correlations,  $\langle \Delta r_i \cdot \Delta r_j \rangle$ , can be obtained. The eigen-decomposition of the Kirchhoff matrix  $\Gamma$  gives us different normal modes at equilibrium. In GNM, all fluctuations are assumed to be isotropic, so we can only have the magnitudes of the fluctuations without directional information of them.

#### 1.2.4 ANM

To provide directional information GNM does not, Anisotropic Network Model (ANM) is proposed (Doruker 2000, Atilgan 2001, Tama 2001, and Eyal 2006) and it has become one of the most broadly used ENMs. The potential function for ANM,  $V_{ANM}$  is defined by (Hinsen 1998, Hinsen 1999, and Atilgan 2001):

$$V_{ANM} = \frac{1}{2} \sum_{ij} \gamma_{ij} (\Delta r_{ij})^2 \quad (1.5)$$

Then, we can write a closed form expression for the second derivative of potential function for interacting  $i$ th atom and  $j$ th atom by equation 1.5:

$$\frac{\partial^2 V}{\partial x_i \partial y_j} = - \frac{\gamma_{ij} (x_j - x_i)(y_j - y_i)}{r_{ij}^2} \quad (1.6)$$

The dimension of Hessian matrix is  $3N \times 3N$  and it can be expressed as a matrix which have  $N \times N$  super-elements of size  $3 \times 3$ :

$$H^{ANM} = \begin{bmatrix} H_{11} & H_{11} & \dots & H_{1N} \\ H_{21} & H_{22} & \dots & H_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ H_{N1} & H_{N2} & \dots & H_{NN} \end{bmatrix} \quad (1.7)$$

We have the super-element of Hessian matrix,  $H^{ANM}$  for interacting two atoms:

$$H_{ij} = \begin{bmatrix} \frac{\partial^2 V}{\partial x_i \partial x_j} & \frac{\partial^2 V}{\partial x_i \partial y_j} & \frac{\partial^2 V}{\partial x_i \partial z_j} \\ \frac{\partial^2 V}{\partial y_i \partial x_j} & \frac{\partial^2 V}{\partial y_i \partial y_j} & \frac{\partial^2 V}{\partial y_i \partial z_j} \\ \frac{\partial^2 V}{\partial z_i \partial x_j} & \frac{\partial^2 V}{\partial z_i \partial y_j} & \frac{\partial^2 V}{\partial z_i \partial z_j} \end{bmatrix} \quad (1.8)$$

By equation 1.6 and 1.8, we can have closed form expression of off-diagonal super-element:

$$H_{ij} = -\frac{\gamma_{ij}}{r_{ij}^2} \begin{bmatrix} x_j - x_i \\ y_j - y_i \\ z_j - z_i \end{bmatrix} \begin{bmatrix} x_j - x_i & y_j - y_i & z_j - z_i \end{bmatrix} \quad (1.9)$$

where  $\gamma_{ij} = 0$  if  $r_{ij} > r_c$  and the optimal cutoff distance  $r_c$  is about 13-18 Å (Eyal 2006).  $\gamma_{ij}$  can be dependant on the identity pair of the two interaction atoms, but usually it is set to a uniform constant value (Tirion 1996). Lastly, the diagonal super-element is defined by:

$$H_{ii} = -\sum_{i \neq j} H_{ij} \quad (1.10)$$

GNM and ANM have pros and cons respectively. The GNM has been recognized to have better fluctuation predictions than those obtained by ANM (Bahar 2005 and Chennubhotla 2005). The reason of the better results may be explained by its potential function which considers orientational deformations according to the change of distance vectors as well as distance changes (see equation 1.4 and 1.5).

### 1.2.5 sbNMA and ssNMA

With ENM, the normal mode analysis can be performed quickly while trading off accuracy because ENM does not distinguish bonded and non-bonded interactions. NMA is more accurate than ENM but it requires energy minimization which could be the bottleneck for the analysis of large macromolecules. Recently, the model spring-based NMA (sbNMA) was developed, with the hope of performing normal mode analysis faster than NMA while keeping its accuracy (Na 2014). The sbNMA also uses fine-grained all-atom model and all-atom force fields. In order to derive the most optimized model, the original NMA connected to ENM with several intermediate models. The sbNMA does not require an energy minimization step while keeping



most of its accuracy. NMA Hessian matrix can be written as a summation of two terms (Na 2014):

$$H = H_{spr} + H_{frc} \quad (1.11)$$

Na and Song found that the contribution of the spring term  $H_{spr}$  is much greater than the force term  $H_{frc}$ . So they kept only the spring-based term and named their model spring-based NMA (sbNMA). As expected, sbNMA showed superior cross-correlation in MSF with NMA to other simplified models. The next intermediate model, ssNMA, is an even more simplified model than sbNMA. It is almost identical except the fact that the ssNMA uses a much smaller number of parameters (see table 2.1).

### 1.2.6 CHARMM

CHARMM (Chemistry at HARvard Macromolecular Mechanics) is one of widely used force fields set especially for molecular dynamics of biomolecules (Brook 1983). It provides a full atomic empirical potential energy functions (equation 1.12) along with analysis package. CHARMM potential function consists of five terms. They can be categorized into three groups which are two-, three- and four-body interactions. Two-body interactions include bond stretching, van der Waals and electro static interactions, three-body interactions include bond angle and Urey-Bradley interactions, and four-body interactions include dihedral and improper dihedral interactions. For sbNMA and ssNMA, electrostatic interactions are ignored. CHARMM provides several versions of optimized empirical potential function parameter sets, and CHARMM22 is used in this study.

$$\begin{aligned} V = & \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} k_\phi(1 + \cos(n\phi - \delta)) \\ & + \sum_{impropers} k_\omega(\omega - \omega_0)^2 + \sum_{Urey-Bradley} k_u(u - u_0)^2 \\ & + \sum_{nonbonded} \left[ \epsilon \left( \left( \frac{r_{min_{ij}}}{r_{ij}} \right)^{12} - \left( \frac{r_{min_{ij}}}{r_{ij}} \right)^6 \right) + \frac{q_i q_j}{\epsilon r_{ij}} \right] \end{aligned} \quad (1.12)$$

### 1.2.7 AMBER

AMBER (Assisted Model Building with Energy Refinement) is another popular family of force fields for molecular dynamics (Cornell 1995). It also comes along with molecular dynamics software package as well as several versions of parameter sets. Unlike CHARMM, it does not include Urey-Bradley, and the improper dihedral interactions are integrated in dihedral interactions (equation 1.12 and 1.13). While CHARMM has parameters for van der Waals 1-4 interactions, AMBER has vdw-14-scale parameter which can reduce the exaggeration of effect of van der Waals 1-4 interactions by dividing the van der Waals parameters by two.

$$\begin{aligned}
 V = & \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{torsionals} k_\phi(1 + \cos(n\phi - \delta)) \\
 & + \sum_{nonbonded} \left[ \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\epsilon r_{ij}} \right]
 \end{aligned}
 \tag{1.13}$$

## CHAPTER 2. METHODS

### 2.1 The protein dataset

For this study, the same protein dataset as used in a previous study (Na 2016) is adopted. The dataset includes 135 globular proteins which consists of 42 alpha-proteins, 37 beta-proteins, and 56 alpha/beta proteins. The resolution threshold was 2.5 Å. The size of proteins range from 61 amino acid residues to 149 residues, and none of them deviates for more than 6.0 Å during energy minimization. The minimization of X-ray crystallographic structures were done using the program 'minimize' from Tinker package (Ponder 1987) with either CHARMM22 (MacKerell 1998) or AMBER94 (Wang 2000) force field. The dataset was obtained at <http://www.cs.iastate.edu/gsong/CSB/NMAdb/135.html>.

### 2.2 Mean square fluctuations (MSF) and normal mode frequency

The Hessian matrix for NMA is obtained through the program 'testhess' from the Tinker package with either the CHARMM22 or AMBER94 force field. The NMA Hessian matrix is imported, and MSF and normal mode frequencies are calculated using a program written in C# and Matlab (MathWorks). For sbNMA, parameters are imported from Tinker parameter files for CHARMM22 and AMBER94, and Hessian matrix is constructed first followed by calculations of MSF and normal mode frequencies using the same program as that is used for NMA.

### 2.3 Single parameters of ssNMA

The ssNMA use a single parameter for each type of interactions, and they are summarized in table 2.1. The parameters for CHARMM22 are taken from a previous study (Na 2014).

Table 2.1 Parameters used in ssNMA for CHARMM and AMBER94. Unit description: kcal/mol/Å<sup>2</sup> for  $K_b$  and  $K_{UB}$ , kcal/mol/rad<sup>2</sup> for  $K_\theta$ ,  $K_\phi$  and  $K_{imptor}$ , kcal/mol for  $\epsilon$  and Å for all radii. The values in parenthesis are for 1-4 van der Waals interactions.

	$K_b$	$K_\theta$	$K_{UB}$	$K_\phi$	$n_\phi$	$K_{imp}$	$K_{imptor}$	$n_{imptor}$
CHARMM	340	45	10	1	1	70	N/A	N/A
AMBER	384	52	N/A	0.63	3	N/A	4.8	2
	$\epsilon$	$r_H$	$r_N$	$r_O$	$r_C$	$r_S$		
CHARMM	-0.1	1.2	1.85 (1.55)	1.70 (1.40)	1.90	2.0		
AMBER	0.074	1.23	1.83	1.67	1.91	2.0		

For AMBER94, the range and frequency of usage of the parameters of each interaction type is considered. Then, the weighted averages of the parameters are computed. There are some differences between CHARMM and AMBER (see 1.12 and 1.13). First, AMBER potential function does not contain Urey-Bradley term. Second, CHARMM has a separate term for improper, while AMBER does not, so only CHARMM has  $K_{imp}$  parameter. AMBER, however, has separate torsional parameters for improper term  $K_{imptor}$  and  $n_{imptor}$ . Lastly, AMBER sets vdw-14-scale to 2.0 instead of having parameters for 1-4 van der Waals separately to avoid exaggeration on it. Again, the parameters are not just averaged, but are weighted averages over the interactions that exist in the 135 proteins in the dataset.

## 2.4 Converting eigenvalues to vibrational frequencies

From the geometrical information of protein structure and atom types, we can construct Hessian matrix  $\mathbf{H}$  (equation 1.2) and inertia matrix  $\mathbf{M}$  of interacting  $i$ th and  $j$ th atoms with respect to the generalized coordinates  $q_i$ :

$$M_{ij} = \sum_k m_k \frac{\partial \mathbf{r}_k}{\partial q_i} \cdot \frac{\partial \mathbf{r}_k}{\partial q_j} \quad (2.1)$$

where  $m_k$  is the mass of  $k$ th atom and  $\mathbf{r}_k$  is its location vector. From the equation 1.2 and 2.1 we forms a generalized eigenvalue problem:

$$\mathbf{H}\mathbf{v}_i = \lambda_i \mathbf{M}\mathbf{v}_i \quad (2.2)$$

where  $\mathbf{v}_i$  is the  $i$ th eigenmode and  $\lambda_i$  is the corresponding eigenvalue. Now we can get the frequency  $\omega_i$  from  $\lambda_i = \omega_i^2$ . To match the value to experimental result,  $\omega_i$  is divided by  $2\pi c$  where  $c$  is the speed of light, which makes the frequency to have  $cm^{-1}$  unit.

## CHAPTER 3. RESULTS

### 3.1 Evaluation of simplified models by MSF correlations

Mean square fluctuation (MSF) correlations between NMA and simplified models were used to evaluate the quality of simplified models (Na 2014). The Hessian matrix for NMA is obtained from a Tinker program. Then, it is imported and its MSF is calculated using a program written in C# and Matlab. The Hessian matrices for sbNMA and ssNMA are constructed by the program above, followed by a calculation of their MSF. Finally, the correlations between NMA and two simplified models for 135 proteins are performed. The results are summarized in table 3.1 together with those of CHARMM, which is adopted from a previous study (Na 2014).

NMA requires expensive minimization of protein structures while the simplified models sbNMA and ssNMA do not. Therefore, most of analysis time were consumed by the original NMA. The ssNMA is further simplified and uses a single uniform parameter for each potential term because it does not affect low frequency motions which account for the motions of protein domain. The parameters for ssNMA are summarized in table 2.1. Both MSF correlations of CHARMM and AMBER with NMA are comparable regardless of simplified model types used.

Although both ssNMA and sbNMA have high correlation coefficients with two force fields, there exist some differences. To investigate where the difference comes from, their distribution is taken into account (figure 3.1). In AMBER, the overall shape and the peak position remain, but the number of proteins that have low correlation in MSF, is higher than CHARMM. Therefore, it was suspected that the low MSF correlation might be due to the larger RMSD deviation due to the minimization of original structures, so a scatter plot is drawn to see the relationship of MSF and RMSD (figure 3.2). However, the proteins, which shows low MSF with AMBER, do not have high RMSD. Therefore, it is concluded that the relatively higher population of the low-

Table 3.1 MSF correlations of ssNMA and sbNMA with NMA. The dataset of 135 proteins are used to get MSF for ssNMA and sbNMA to correlate with NMA with CHARMM or AMBER. The number in table is averaged from 135 proteins.

	<b>ssNMA</b>	<b>sbNMA</b>	<b>NMA</b>
<b>Minimization</b>	<b>No</b>	<b>No</b>	<b>Yes</b>
<b>Single Parameter</b>	<b>Yes</b>	<b>No</b>	<b>No</b>
<b>Correlation with CHARMM</b>	<b>0.88</b>	<b>0.86</b>	<b>1</b>
<b>Correlation with AMBER</b>	<b>0.83</b>	<b>0.81</b>	<b>1</b>

MSF-correlation proteins with AMBER is not due to different minimization between CHARMM and AMBER, but probably due to the difference in the contribution of the force/torque terms to NMA Hessian matrix (equation 1.11).

### 3.2 Universality of vibrational spectrum

The frequency of vibrational modes are converted to density  $g(\omega)$  (Figure 3.3). The density is properly normalized to see the universal spectrum (ben-Avraham 1993 and Tirion 1993). The figure design is directly adopted from the previous study in order to compare the spectrum of AMBER with that of CHARMM easily (Na 2016). The Hessian matrices for NMA, sbNMA and ssNMA were obtained in section 3.1. Then the vibrational frequencies are calculated by eigen-decomposition of Hessian matrices, followed by conversion from the eigenvalues to frequencies (section 2.4).

The black line indicates the mean of  $g(\omega)$ s on every  $5 \text{ cm}^{-1}$  bins. For each bin, the density distribution is distinguished by different colors. In orange area, 50% densities which are close to mean density are vertically distributed. Similarly, red and gray area contain 90% and 100% densities at a frequency bin respectively. As a result, the universality of density spectrum of vibrational modes is appear to be as universal as it was with CHARMM (Na 2016). Other than the different unique spectrum profile, the distribution at each bin is little bit wider than

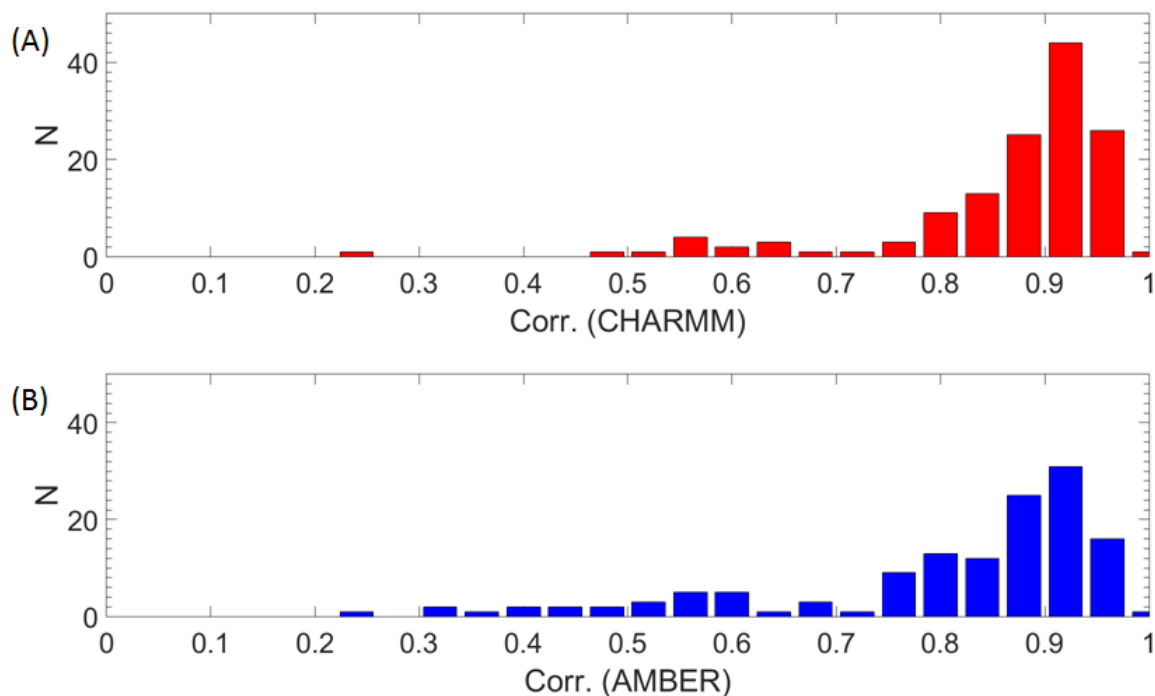


Figure 3.1 Distribution of MSF correlation for sbNMA with CHARMM and AMBER

CHARMM.

The faithful reproduction of spectrum profile throughout the whole frequency range with both CHARMM and AMBER must imply a connection to certain physical characteristics as Na and Song (2016) claimed. They also suggested the possibility that there might be dependency of fluctuations on protein sizes, but due to limit of time, it is left for future work again.

### 3.3 Spectra of three protein groups by different protein folds

The secondary structure difference between  $\alpha$ -helix and  $\beta$ -sheet causes difference in vibrational spectra (Levitt 1985). The protein dataset used in this study was categorized in a previous study (Na 2016) according to the CATH protein structure classification (Sillitoe 2015). The dataset contains 42 alpha-proteins, 37 beta-proteins and 56 alpha/beta proteins. From each one of three groups, the mean of vibrational frequency distributions  $g(\omega)$  is obtained (figure 3.4).



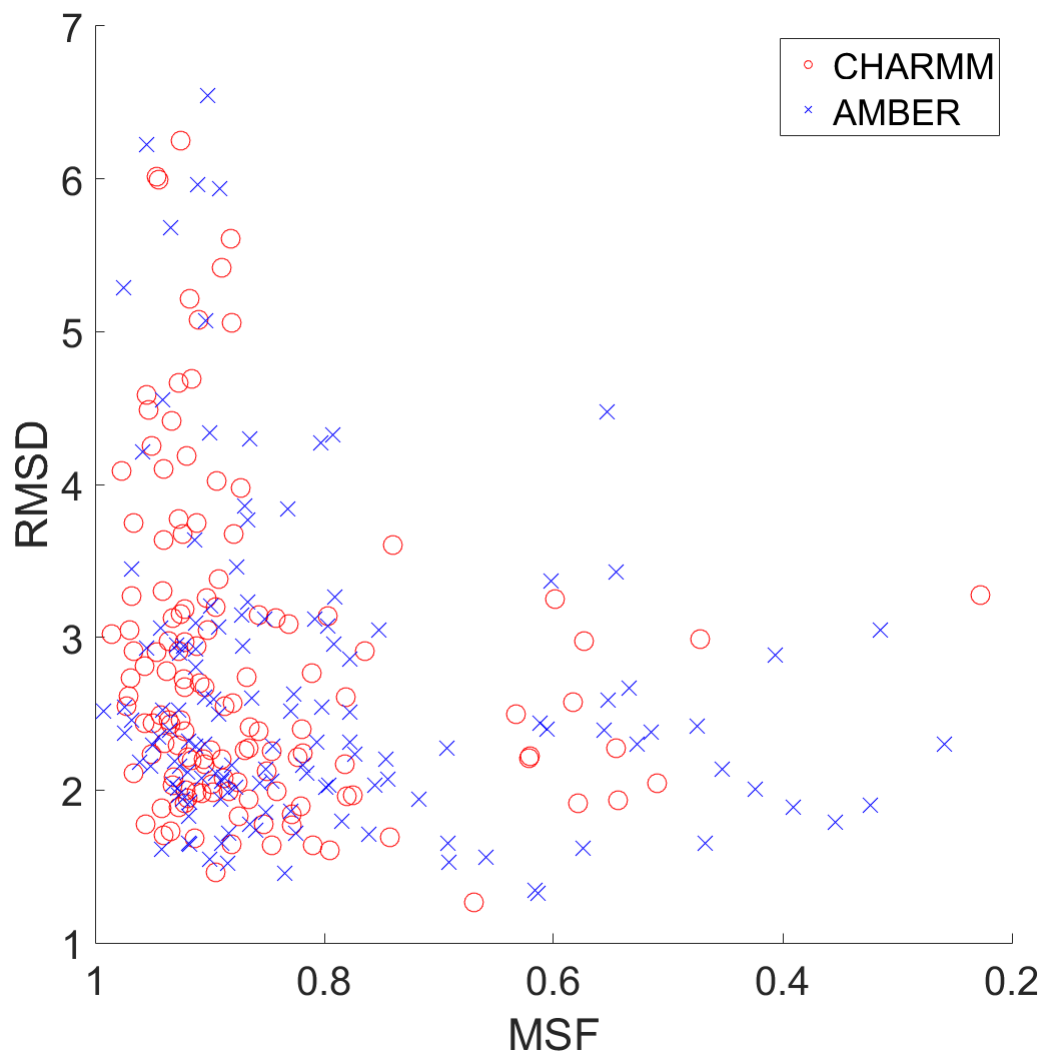


Figure 3.2 Scatter plot of the RMSD versus the corresponding MSF correlation of sbNMA. RMSD between minimized and original protein structures for 135 proteins are calculated and then scatter plot is constructed versus MSF correlation. The number of circles and crosses is 135 respectively and either a circle or cross indicates a protein out of 135.

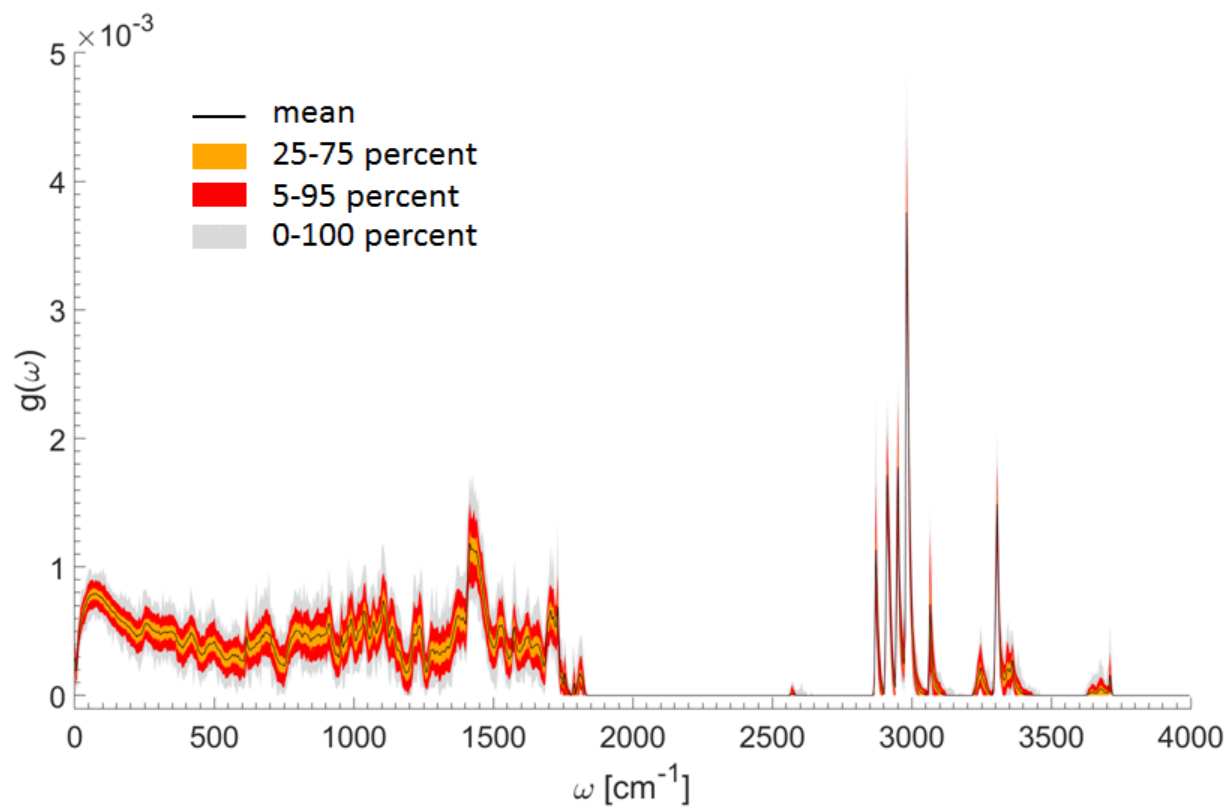


Figure 3.3 Universality of the density of vibrational modes with AMBER. The bin size  $\Delta\omega$  is  $5 \text{ cm}^{-1}$ , the densities at each frequency bin are vertically distributed. The black line connects mean densities at bins. The gray band includes all densities from the 135 proteins of the dataset. The orange and red bands include 50 % and 90 % of densities which are closed to the mean density.

Again, the graph design is adopted from a previous study (Na 2016) for a fair comparison. The overall spectrum profiles for three protein groups are shown in figure 3.4(A) with a full frequency range. The red, blue and gray lines indicate the mean of vibrational spectrum for alpha, beta and alpha/beta proteins respectively. At a glance, the spectra are similar without any significant difference. However, when it is zoomed in for three frequency ranges (figure 3.4(B)), which are known as amide vibration frequencies (Nevskaya 1976, Krimm 1986, Susi 1986, Goormaghtigh 1990, Fu 1994, Cai 1999, Cai 2004, Yang 2015), some differences become visible. Na and Song (2016) observed certain orders of the spectrum peaks of alpha-proteins and beta-proteins in those ranges. The order of amide III region is well matched to the result from CHARMM. Although general locations of three amide groups fit well to the amide ranges reported from experiment results, the orders of peaks in amide I and amide II regions are less clear or appear to be reversed when using AMBER94. The fact that the order of peaks in amide III region is strongly conserved in both CHARMM (Na 2016) and AMBER as well as in experimental study (Cai 2004) may be used to tune parameters of empirical force fields.

### 3.4 Evaluation of sbNMA and ssNMA on vibrational spectra

The vibrational frequency spectra of two simplified models, sbNMA and ssNMA, are compared to the spectrum of NMA to evaluate their quality (figure 3.5). Low frequency ranges are important because they represent the motions of large domains of a protein. The first peak is located at around  $80 \text{ cm}^{-1}$  frequency for NMA (figure 3.5(B)), which is well matched to the previous experimental (Giraud 2003) and simulation result with CHARMM (Na 2016). The second major peak appears at about  $270 \text{ cm}^{-1}$  for CHARMM while for AMBER it is at about  $250 \text{ cm}^{-1}$ .

The spectrum of sbNMA was well overlapped on that of NMA but ssNMA shows large discrepancy as frequency increases (figure 3.5(A)). This is a little surprising since ssNMA showed better MSF correlations with NMA than sbNMA for both CHARMM and AMBER (table 3.1). In the low frequency range, the motions are governed by the relatively large conformational changes that involve many atoms, while the high frequency range is related to the motions of small groups of atoms. Therefore the effects of minor errors are accumulated in

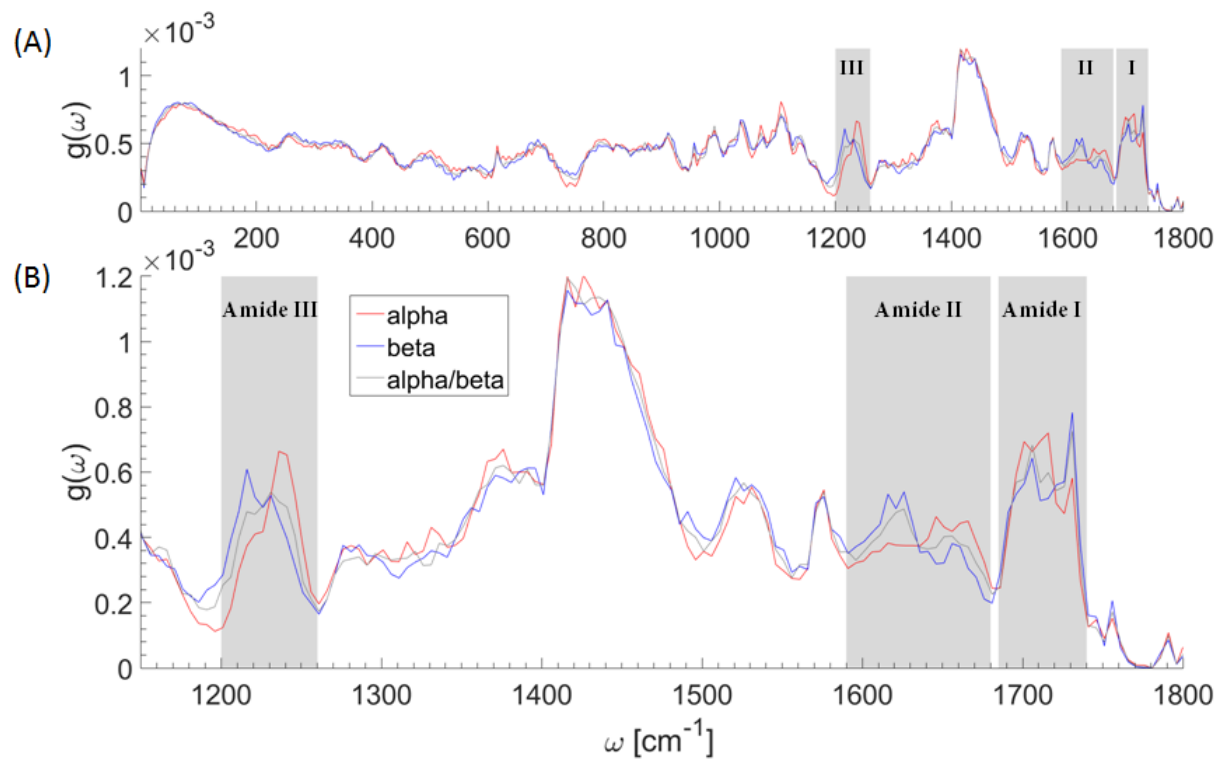


Figure 3.4 Amide groups in vibrational spectra of NMA for different protein folds. (A) The vibrational spectra for 42 alpha-proteins (red), 37 beta-proteins (blue), and 56 proteins with comparable alpha helices and beta sheets (gray). The three gray bands indicate the frequency ranges for amide I, II and III peaks respectively. (B) Same spectra from (A) but zoomed-in to see the spectra for amide regions in detail.

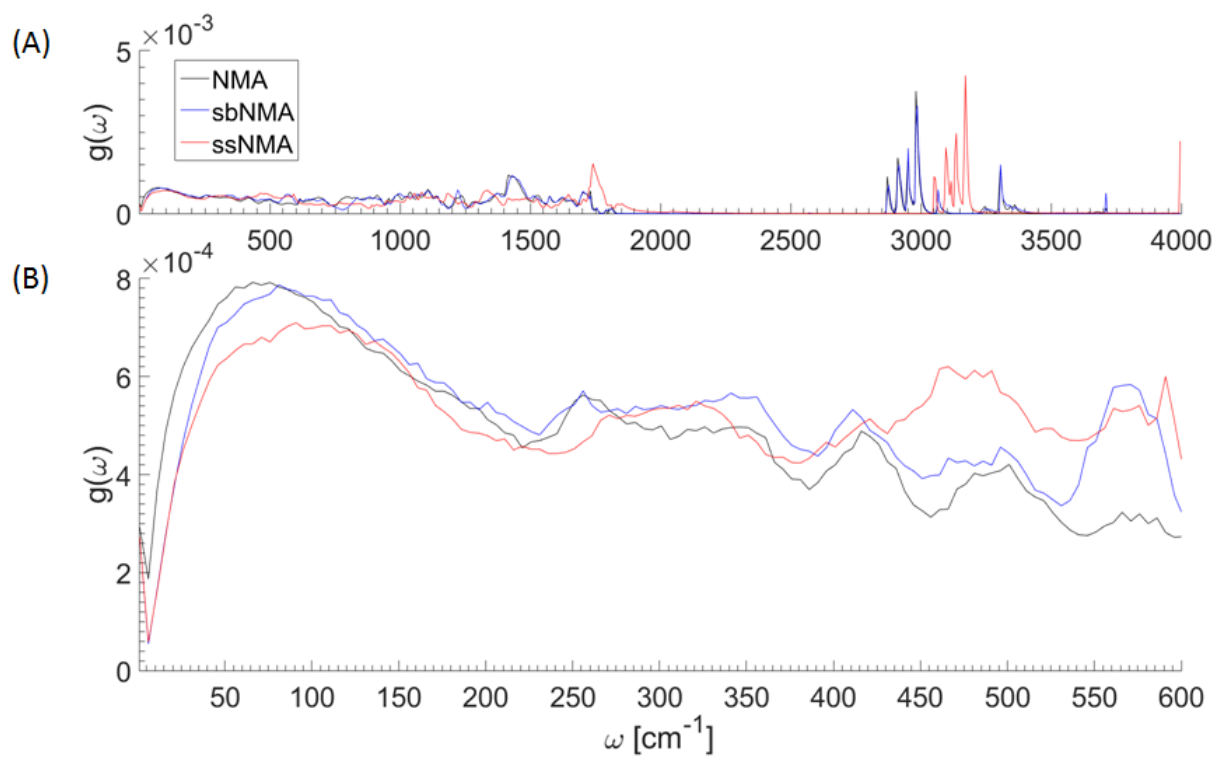


Figure 3.5 Vibrational spectra on NMA and two simplified model, sbNMA and ssNMA

the low frequency motions where many atoms affect each other, which results in more deviation in ssNMA spectrum (figure 3.5 (B)). However, the peak positions at the high frequency range (around  $3000\text{ cm}^{-1}$ ) are just shifted to a higher range (figure 3.5 (A)) without the change of spectrum profile, which is may be because the number of atoms participating in high frequency motions is small at high frequency end is thus more conserved.

The large spectral discrepancy of ssNMA shows that vibrational spectrum should be better for the evaluation of simplified models. To support that, first, the vibrational spectrums for sbNMA and ssNMA are shown (figure 3.6). Both sbNMA and ssNMA show universal spectrum throughout the 135 protein dataset, which again support the universality of vibrational spectrum of normal modes. Next, the order of peaks for the different secondary structure in the amide I range is revisited for sbNMA and NMA with AMBER (figure 3.7). The sbNMA faithfully reproduce the order of peaks in the amide III range as well as amide I and amide II (figure 3.7 (A)). However, with ssNMA, the position of amide I-III frequency ranges are distorted, and furthermore no orders could be found.

The ssNMA showed better MSF correlations with NMA (table 3.1) even if it uses a single uniform parameters for each interactions such as bond, angle, dihedral, improper, Urey-Bradley and van der Waals (table 2.1). Therefore, ssNMA appeared to be a better simplified model than sbNMA since it is well correlated to NMA in terms of fluctuation, and does not require to import force field parameters. However, when the vibrational spectrum is considered, ssNMA was no as good as sbNMA. Therefore, inspecting the spectrum could be a powerful method to verify the soundness of simplified models and for tuning the parameters of empirical force fields.

### 3.5 Effect of input structure on vibrational spectrum

The sbNMA matched well with NMA in both MSF correlations (table 3.1) and vibrational spectrum (figure 3.5). It also preserves the order of peak positions of different secondary structure within the amide III range (figure 3.7), which is also confirmed by experiment result (Cai 2004) as well as NMA (figure 3.4). However, all analysis of sbNMA above are done with minimized structures for fair comparisons.

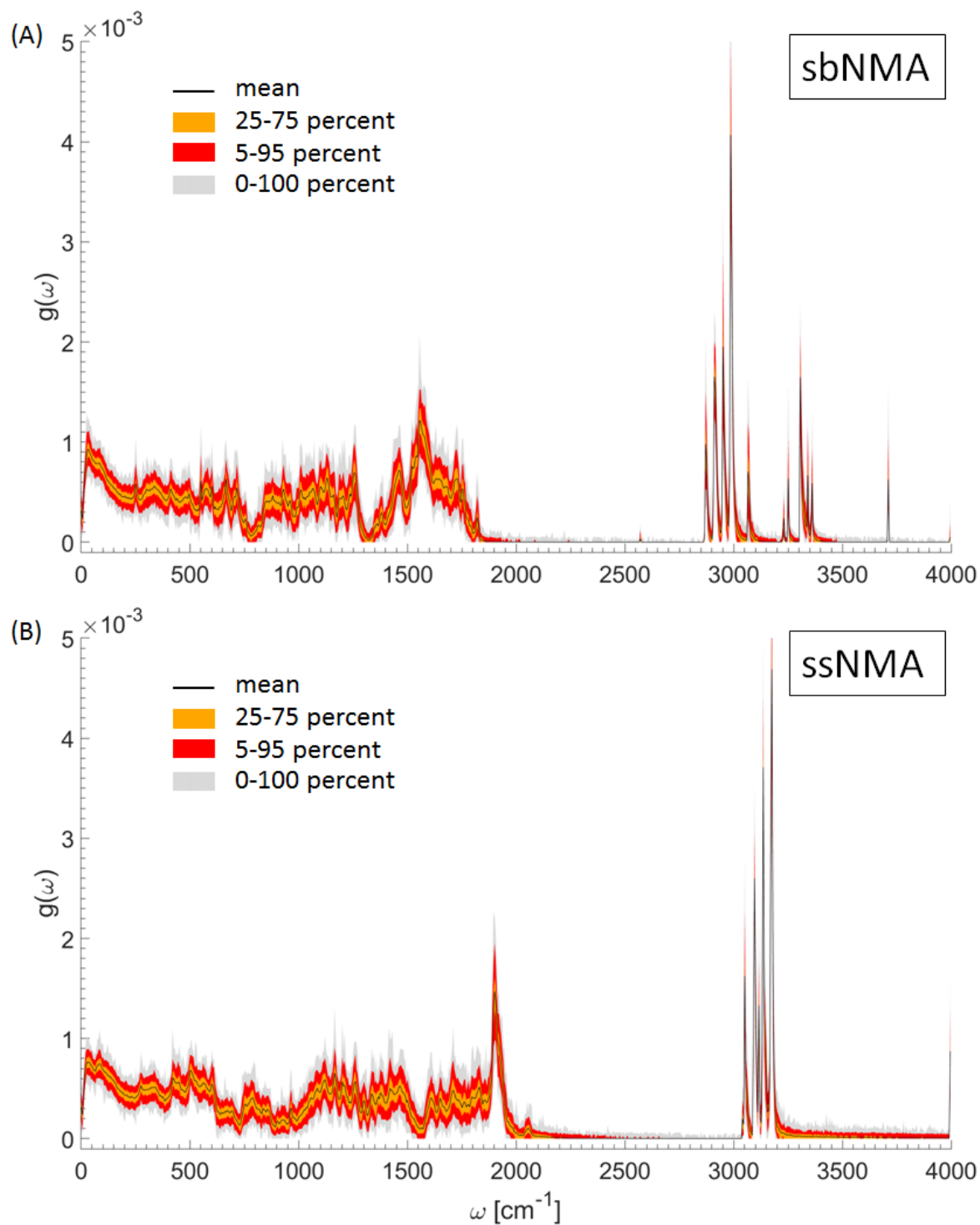


Figure 3.6 Vibrational of the density of vibrational modes of (A) sbNMA and (B) ssNMA with AMBER.

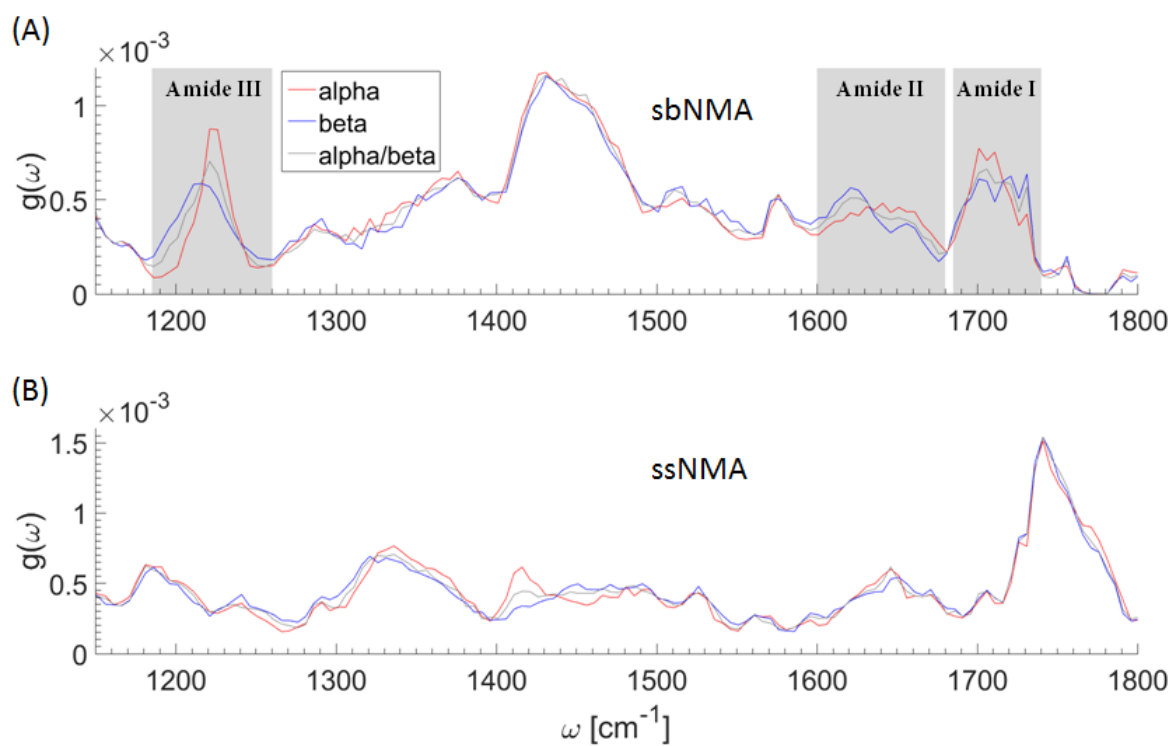


Figure 3.7 Amide groups in vibrational spectra for sbNMA and ssNMA. Two sub-figures for (A) sbNMA and (B) ssNMA are equivalent to the figure 3.4 (B) for NMA.



One of the remarkable advantages of using simplified models including sbNMA is that they do not need the cumbersome minimization step (Tirion 1996). Consequently, sbNMA can be applied to any input structure as long as it is not deviated too much from one to another (Na 2015). To investigate the effect of input structure on vibrational spectrum, the sbNMA is applied to the original X-ray structures from protein data bank (<http://www.rcsb.org/pdb/>) (figure 3.8). Similar to the result of CHARMM in a previous study (Na 2016), the first peak was shifted to a lower frequency. The first peak was positioned around  $30 \text{ cm}^{-1}$  while it was near  $80 \text{ cm}^{-1}$  when minimized structures were used. With CHARMM, its position was also shifted to the left (Na 2016) but a lesser extent. The second peak is placed at about  $250 \text{ cm}^{-1}$  which is well matched with the peak location obtained from minimized structures.

The shift of the first peak position is apparently due to minimization, which is directly linked to the change of protein structure. However, considering the result that different protein structures display the universality of vibrational spectrum for NMA (figure 3.3) as well as simplified models (figure 3.6), it is hard to explain the shift only based on difference in structure. Therefore, the minimization of protein structure may affect the spectra of van der Waals term as Na and Song claimed (Na 2016) because the van der Waals term contributes mainly to the first peak (Na 2016), and elucidating that could be an interesting future work.

### 3.6 Comparisons between CHARMM and AMBER

Since sbNMA can be directly applied to protein structures without minimization, we can use it to investigate and compare the effects of different force fields on normal mode computations without introducing any bias that is due to energy minimization. Here, sbNMA is applied with CHARMM and AMBER. First, the MSF computed from CHARMM correlates extremely well with AMBER (figure 3.9). This indicates that two widely-used force fields, CHARMM and AMBER, have a highly unified spring-based component. The main difference of the two force fields is thus in the force-based term, i.e., in how forces/torques are balanced in a given structure. To examine how differently/similarly force-balancing is done in CHARMM and AMBER, we compare the minimized structures obtained from minimization using CHARMM with those using AMBER. To this end, we draw a scatter plot of the distances between minimized

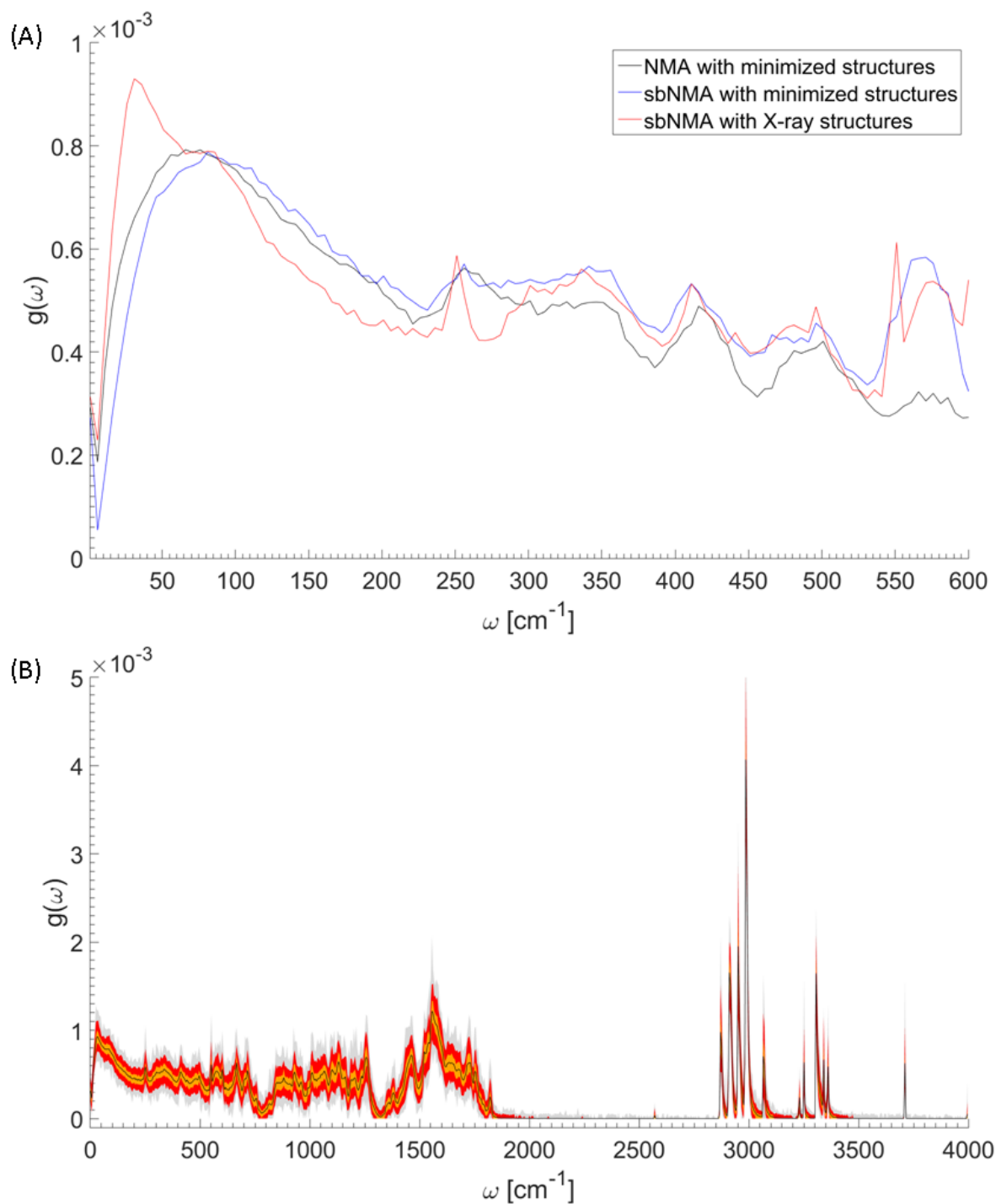


Figure 3.8 Vibrational spectra on different input structures. Normalized structures are obtained by executing Tinker program with AMBER on 135 protein dataset. (A) Vibrational spectra of NMA with minimized structure (black), sbNMA with minimized structure (red), and sbNMA with X-ray structures where the minimization are done. (B) Universality spectra of sbNMA with X-ray structures.

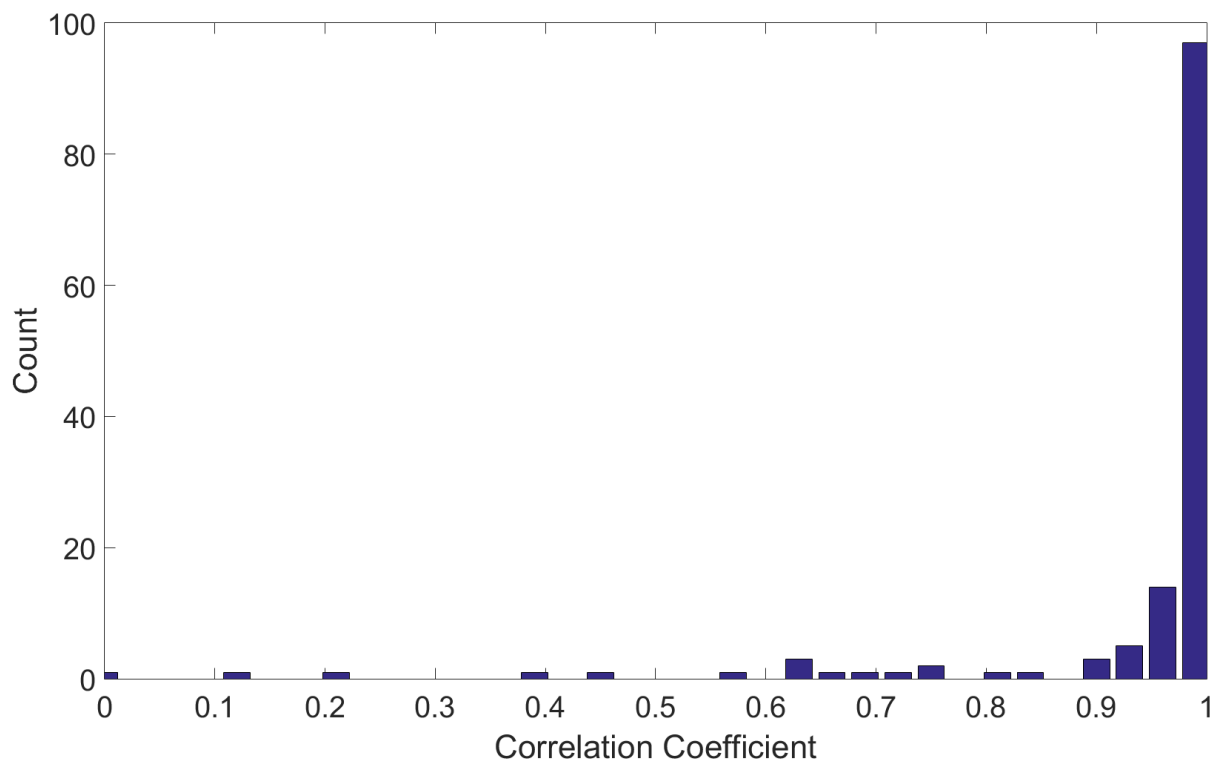


Figure 3.9 Distribution of the MSF correlation between CHARMM and AMBER on original protein structures. The average coefficient is 0.93.

structures and the original protein structures from Protein Data Bank (PDB) and the distances between minimized structures by CHARMM and AMBER (figure 3.10 (A)). Most of the data points fall under the diagonal line, indicating that the distances between minimized structures are in generally less than the extent of structure deviation from the original PDB structures due to minimization by either CHARMM or AMBER. This means the directions of minimization by CHARMM and AMBER are similar. If we draw these three distances as sides of a triangle (see figure 3.11), the angle between the two minimization directions can be computed using the cosine law. Figure 3.10 (B) shows the distribution of this angle  $\theta$  for the 135 proteins in the dataset. It is seen that for most of the proteins, the angle is less than  $60^\circ$ , with a peak at about  $47^\circ$ . Therefore, the minimization directions by CHARMM and AMBER are more similar rather than they are different.

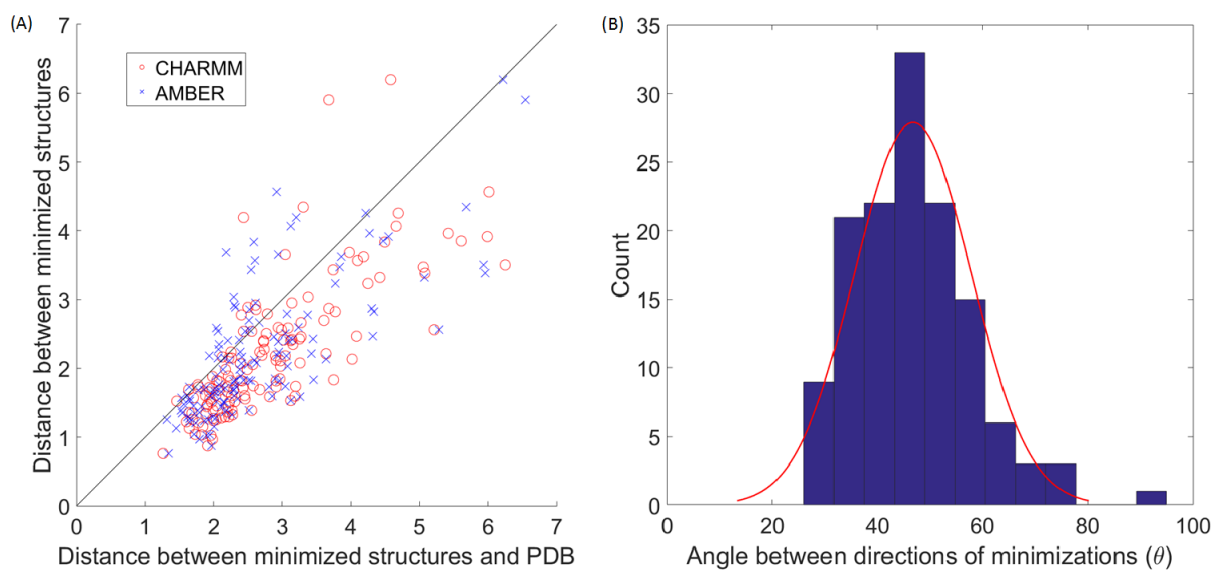


Figure 3.10 Similarity of minimizations by CHARMM and AMBER. (A) Scatter plot of the RMSDs between minimized and original PDB structures versus those between minimized structures by CHARMM and AMBER. All CHARMM/AMBER pairs on the same Y value are separated by X value. (B) Histogram of angles between directions of minimizations. Triangles are constructed by three distances of a pair with the same Y value in (A). The opposite angles of the distance between two minimized structures by CHARMM and AMBER are calculated by the law of cosines.

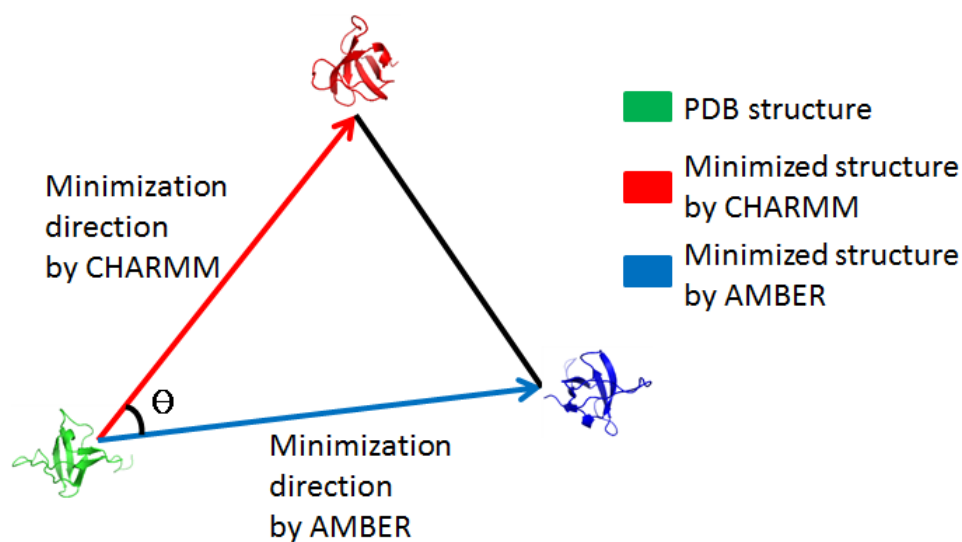


Figure 3.11 RMSD triangle of PDB structure and minimized structures. A PDB structure is shown as green cartoon protein. The red and blue arrow indicate the minimization directions by CHARMM and AMBER respectively. Minimized structures are shown as red (CHARMM) and blue (AMBER) cartoon protein. The lengths of red and blue arrow are RMSDs from the PDB structure by minimizations using CHARMM and AMBER. The length of black line is RMSD between the two minimized structures. The opposite angle ( $\theta$ ) of the black line is calculated by the cosine law.

Next, we plot the vibrational spectra by CHARMM and AMBER. Overall, the two spectra are highly similar (figure 3.12 (A)). Compared with spectra produced from minimized structures, when applied to PDB structures, both force fields cause the first peak to shift to the left, but to a different extent: CHARMM to  $50\text{ cm}^{-1}$  and AMBER to  $30\text{ cm}^{-1}$ . The second peak of AMBER is also shifted toward a lower frequency ( $250\text{ cm}^{-1}$ ) compared to CHARMM (around  $310\text{ cm}^{-1}$ ). Taking all together, sbNMA is stable and has a consistent performance that is independent of the force field being used, whether it is CHARMM or AMBER. CHARMM seems to be more compatible with X-ray crystal structures since its first spectrum peak shifts for a lesser extent than AMBER's.

In summary, the two widely-used force fields, namely CHARMM and AMBER, seem to be highly similar in their spring-based components, i.e., spring-related interactions. Their main differences lie in their force-based components, or the forces and torques at any given conformation, which dictates the direction of minimization. A comparison of their directions of minimization suggests that the force/torques given by the two force fields are similar to a large extent, as their directions of minimization are quite similar.

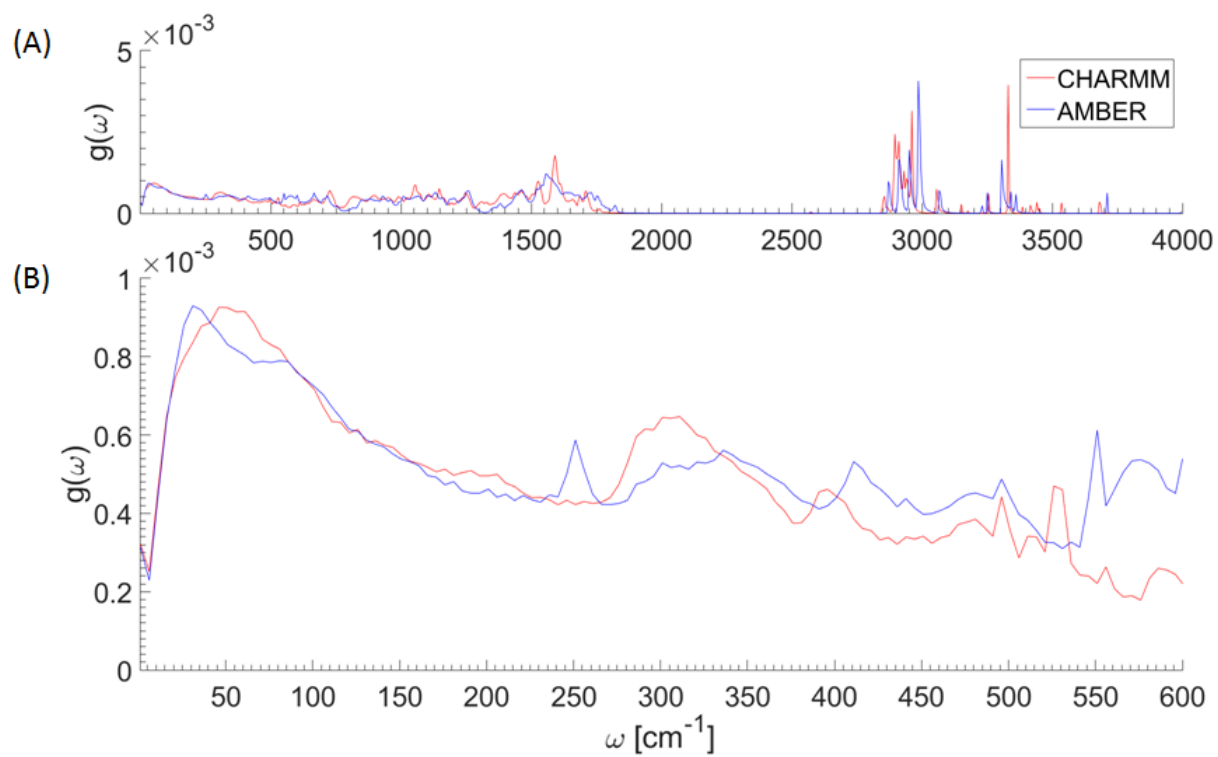


Figure 3.12 Vibrational spectra of sbNMA on original PDB structures with CHARMM and AMBER.

## CHAPTER 4. CONCLUSIONS

In this study, two previous works (Na 2015 and Na 2016), where CHARMM was used, are revisited with the AMBER force field in order to identify the differences between two well-recognized empirical force fields, CHARMM and AMBER as well as to evaluate two simplified models, sbNMA and ssNMA.

First, the MSF correlations of sbNMA and ssNMA with NMA are evaluated using AMBER force field, and the result are compared with that of CHARMM (figure 3.1). The sbNMA, which only takes the spring-based term for Hessian matrix (equation 1.11), was well correlated with NMA with respect to MSF. The more simplified model ssNMA, which is almost identical to sbNMA but uses a single uniform parameter for each interaction (table 2.1), shows slightly better correlations with NMA. Two force fields, CHARMM and AMBER, provide comparable results with sbNMA and ssNMA respectively as expected because their quality has been well acknowledged.

Although two force fields go well together on MSF correlations, there still exist some differences between two force fields. CHARMM provides a slightly better correlation coefficients for both sbNMA and ssNMA (table 3.1). For the AMBER force field, more proteins have low MSF correlation coefficient (figure 3.1). So we hypothesized that those proteins may have higher RMSD by minimization than other proteins with high MSF correlation coefficient. However, there was no clear relationship between MSF correlation and RMSD (figure 3.2). As a conclusion, the force/torque term of AMBER (equation 1.2), which is not considered in sbNMA, may have a more pronounced contribution to Hessian matrix than for the force term of CHARMM.

Using AMBER, the universality of vibrational spectrum was reproduced as before when CHARMM was used (Na 2016) but the universal spectrum profile is different from CHARMM's



(figure 3.3). The order of peaks for alpha proteins and beta proteins was conserved in amide III range (figure 3.4 (B)), which agree with experimental result (Cai 2004) and simulation using CHARMM (Na 2016). Unfortunately, the orders of peaks in amide I and amide II ranges were not clear.

Although the ssNMA, the more simplified model than sbNMA, has slightly better result on the MSF correlations with NMA, it showed some deviation on the vibrational spectra. The universality of spectrum profile with ssNMA is also conserved (figure 3.6 (B)), but the spectrum is distorted and deviate away from that of NMA as the frequency increases (figure 3.5). Furthermore, the order of peaks in amide III frequency ranges totally disappeared (figure 3.7). In contrast, the vibrational spectrum of sbNMA matches well with that of NMA, and the order of the amide III peaks is conserved correctly. Considering the results of ssNMA and sbNMA on the vibrational spectrum, it seems that the vibrational spectrum is a good tool to validate the correctness of simplified models and also possibly to tune up the parameter values of empirical force fields.

The most important advantage of simplified models such as GNM, ANM, ssNMA and sbNMA is that the minimization of original protein structure, which is a bottle neck of simulation for large systems, is not necessary. Therefore, the effect of input structure is investigated with sbNMA. When it comes to MSF correlation with NMA, sbNMA gives good coefficients, 0.86 with CHARMM and 0.81 with AMBER. Regarding vibrational spectrum comparison with NMA, sbNMA does well also (figure 3.5 (A)). A major difference is the shift of the first peak position (figure 3.8 (A)). Because the universality of vibrational spectrum is conserved throughout the different protein structures in the 135 proteins dataset, the difference might be linked to the effect of minimization on van der Waals interactions which mostly contributes to the first peak (Na 2016). Taking all of these together, the original protein structures can be used with confidence to some extent.

Furthermore, we showed that the sbNMA is stable with different force fields. CHARMM and AMBER are used for normal mode analysis by sbNMA on original structures. The MSFs computed from the two force fields highly correlate with one another, with a correlation coefficient of 0.93 (figure 3.9). The vibrational spectra computed from the two force fields also are

similar overall, with slightly shifted peak locations (figure 3.12). These results provide strong evidence that sbNMA works consistently with different force fields.

In this study, AMBER force field was applied to two simplified models, sbNMA and ssNMA, and the results were compared to a previous study that was done with CHARMM (Na 2016). While ssNMA was comparable to sbNMA in MSF correlations, it is significantly worse than sbNMA on reproducing the vibrational spectrum. Therefore, we suggest that spectrum analysis should be performed in validating simplified models. Lastly, by revisiting sbNMA again with AMBER, more supports are added to the soundness of sbNMA.

**BIBLIOGRAPHY**

- Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O. and Bahar, I. (2001). Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophys. J.*, *80*, 505–15.
- Bahar, I., Atilgan, A., Erman, B. (1997). Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.*, *2*, 173–81.
- Bahar, I. and Rader A. J. (2005). Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.*, *15*, 586–92.
- ben-Avraham, D. (1993). Vibrational normal-mode spectrum of globular proteins. *Phys. Rev. B*, *47*, 14559–60.
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem*, *4*, 187–217.
- Brooks, B. and Karplus, M. (1983). Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl Acad. Sci. USA*, *80*, 6571–5.
- Brooks, B. and Karplus, M. (1985). Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. *Proc. Natl Acad. Sci. USA*, *82*, 4995–9.
- Cai, S. and Singh, B. R. (1999). Identification of beta-turn and random coil amide iii infrared bands for secondary structure estimation of proteins. *Biophys. Chem*, *80*, 7–20.

Cai, S. and Singh, B. R. (2004). A distinct utility of the amide iii infrared band for secondary structure estimation of aqueous protein solutions using partial least squares methods. *Biochemistry*, *43*, 2541–9.

Chennubhotla, C., Rader, A. J., Yang, L. W., Bahar, I. (2005). Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. *Phys. Biol.*, *2*, S173–80.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.*, *117*, 5179–97.

Doruker, P., Atilgan A. R., Bahar, I. (2000). Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to alpha-amylase inhibitor. *Proteins*, *40*, 512–24.

Eichinger B. E. (1975). Elasticity Theory. I. Distribution Functions for Perfect Phantom Networks. *Macromolecules*, *5*, 496–505.

Eyal E., Yang L. W. and Bahar, I. (2006). Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics*, *22*, 2619–27.

Flory P. J. (1985). Molecular Theory of Rubber Elasticity. *Polym. J.*, *17*, 1–12.

Fu, F., DeOliveira, D. B., Trumble, W. R., Sarkar, H. K. and Singh, B. R. (1994). Secondary structure estimation of proteins using the amide iii region of fourier transform infrared spectroscopy: Application to analyze calcium-binding-induced structural changes in calsequestrin. *Appl. Spectrosc.*, *48*, 1432–41.

Giraud, G., Karolin, J., and Wynne, K. (2003). Low-Frequency Modes of Peptides and Globular Proteins in Solution Observed by Ultrafast OHD-RIKES Spectroscopy. *Biophys. J.*, *85*, 1903–13.

Go, N., Noguti, T. and Nishikawa, T. (1983). Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl Acad. Sci. USA*, *80*, 3696–700.

Goldstein, H., Poole, C., Safko, J. (2001). *Classical Mechanics Addison Wesley. San Francisco.*

Goormaghtigh, E., Cabiaux, V. and Ruyschaert, J. M. (1990). Secondary structure and dosage of soluble and membrane proteins by attenuated total reflection fourier-transform infrared spectroscopy on hydrated films. *Eur. J. Biochem*, *193*, 409–20.

Haliloglu, T., Bahar, I., and Erman, B. (1997). Gaussian Dynamics of Folded Proteins. *Phys. Rev. Lett*, *79*, 3090.

Henzler-Wildman, K., Kern, D. (2007). Dynamic personalities of proteins. *Nature*, *450*, 964–72.

Hinsen, K. (1998). Analysis of Domain Motions by Approximate Normal Mode Calculations. *Proteins: Struct., Funct., Bioinf.*, *33*, 417–29.

Hinsen, K., Thomas, A., Field, M. J. (1999). Analysis of domain motions in large proteins. *Proteins*, *34*, 369–82.

Itoh, K. and Shimanouchi, T. (1970). Vibrational frequencies and modes of  $\alpha$ -helix. *Biopolymers*, *9*, 383–399.

Krimm, S. and Bandekar, J. (1986). Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. *Adv. Protein Chem*, *38*, 181–364.

Name. (2002). Dynamics of Proteins in Crystals: Comparison of Experiment with Simple Models. *Biophys. J.*, *83*, 723–732.

Levitt, M., Sander, C. and Stern, P. S. (1983). The normal modes of a protein: Native bovine pancreatic trypsin inhibitor. *Int. J. Quant. Chem*, *10*, 181–99.

Levitt, M., Sander, C. and Stern, P. S. (1985). Protein normal-mode dynamics: Trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol*, *181*, 423–47.

- Levy, R. and Karplus, M. (1979). Vibrational approach to the dynamics of an  $\alpha$ -helix. *Biopolymers*, *18*, 2465–95.
- Ma, J. (2005). Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure*, *13*, 373–80.
- MacKerell, A. D., Bashford, D., Bellott, Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B*, *102*, 3586-616.
- Na, H. and Song, G. (2014). Bridging between normal mode analysis and elastic network models. *Proteins: Struct., Funct., Bioinf.*, *82*, 2157–68.
- Na, H. and Song, G. (2014). A natural unification of GNM and ANM and the role of inter-residue forces. *Phys. Biol*, *11*, 036002.
- Na, H and Song, G. (2015). The performance of fine-grained and coarse-grained elastic network models and its dependence on various factors. *Proteins*, *83*, 1273–83.
- Na H, Song G, ben-Avraham D. (2016). Universality of Vibrational Spectra of Globular Proteins. *Phys. Biol*, *13*, 016008.
- Nevskaya, N. A. and Chirgadze, Y. N. (1976). Infrared spectra and resonance interactions of amide-i and ii vibrations of  $\alpha$ -helix. *Biopolymers*, *15*, 637–48.
- Ponder, J. W. and Richards, F. M. (1987). An efficient newton-like method for molecular mechanics energy minimization of large molecules. *J. Comput. Chem.*, *8*, 1016–24.
- Seno, Y. and Go, N. (1990). Deoxymyoglobin studied by the conformational normal mode analysis: I. Dynamics of globin and the heme-globin interaction. *J. Mol. Biol*, *216*, 95–109.

- Shimanouchi, T. (1970). Stable conformations of polymer chains and model compound molecules. *Discuss. Faraday Soc*, *49*, 60–69.
- Sillitoe, I., Lewis, T. E., Cu, A., Das, S., Ashford, P., Dawson, N. L., Furnham, N., Laskowski, R. A., Lee, D., Lees, J. G., Lehtinen, S., Studer, R. A., Thornton, J., and Orengo, C. A. (2015). CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res*, *43*, D376–81.
- Susi, H. and Byler, D. M. (1986). Resolution-enhanced fourier transform infrared spectroscopy of enzymes. *Methods Enzymol*, *130*, 290–311.
- Tama, F., Sanejouand Y. H. (2001). Conformational change of proteins arising from normal mode calculations. *Protein. Eng*, *14*, 1–6.
- Tirion, M. M. and ben-Avraham, D. (1993). Normal mode analysis of g-actin. *J. Mol. Biol*, *230*, 186–95.
- Tirion M. M. (1996). Large amplitude elastic motions in proteins from a single-parameter. *Phys. Rev. Lett.*, *77*, 1905–8.
- Wang, J., Cieplak, P., and Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?. *J. Comput. Chem.*, *21*, 1049–74.
- Yang, H., Yang, S., Kong, J., Dong, A. and Yu, S. (2015). Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy. *Nat. Protoc.*, *10*, 382–96.
- 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*, 3853–7.