


2013

# Application of Finite Element Method in Protein Normal Mode Analysis

Chiung-fang Hsu

*University of Massachusetts Amherst, [chiungfa@engin.umass.edu](mailto:chiungfa@engin.umass.edu)*

Follow this and additional works at: <http://scholarworks.umass.edu/theses>

 Part of the [Applied Mechanics Commons](#), [Biomechanical Engineering Commons](#), and the [Computer-Aided Engineering and Design Commons](#)

---

Hsu, Chiung-fang, "Application of Finite Element Method in Protein Normal Mode Analysis" (2013). *Masters Theses 1911 - February 2014*. 1014.

<http://scholarworks.umass.edu/theses/1014>

This thesis is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Masters Theses 1911 - February 2014 by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact [scholarworks@library.umass.edu](mailto:scholarworks@library.umass.edu).

**APPLICATION OF FINITE ELEMENT METHOD  
IN PROTEIN NORMAL MODE ANALYSIS**

A Thesis Presented

by

CHIUNG-FANG HSU

Submitted to the Graduate School of the  
University of Massachusetts Amherst in partial fulfillment  
of the requirements for the degree of

MASTER OF SCIENCE IN MECHANICAL ENGINEERING

February 2013

Mechanical and Industrial Engineering

© Copyright by CHIUNG-FANG HSU 2013

All Rights Reserved

**APPLICATION OF FINITE ELEMENT METHOD  
IN PROTEIN NORMAL MODE ANALYSIS**

A Thesis Presented

by

CHIUNG-FANG HSU

Approved as to style and content by:

---

Byung H. Kim, Chair

---

Ian Royce Grosse, Member

---

Ching-Shung Chang, Member

---

Donald L. Fisher, Department Chair  
Mechanical and Industrial Engineering

*To Mom, Dad, Jennifer and Alvin.*

## ACKNOWLEDGMENTS

I would like to express my most sincere gratitude to the following persons who have made the completion of this thesis possible:

Professor Byung Kim, my advisor, for providing me with all the freedom, the necessary tools, and the guidance that was crucial in exploring.

Professor Ian Grosse and Professor C.S. Chang, for accepting to stay on my thesis defense committee and providing me with critical insights in mathematical aspects of my research.

Ming-Wen Hu, my lab mate, for his unsparing and consistent help and great company in the lab.

Most especially to my family and friends, near and far, for their good wishes, blessings and constant encouragement.

And to God, who made all things possible.

## ABSTRACT

# APPLICATION OF FINITE ELEMENT METHOD IN PROTEIN NORMAL MODE ANALYSIS

FEBRUARY 2013

CHIUNG-FANG HSU

B.S.M.E., NATIONAL TAIWAN UNIVERSITY, TAIWAN

M.S.M.E., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Byung H. Kim

This study proposed a finite element procedure for protein normal mode analysis (NMA). The finite element model adopted the protein solvent-excluded surface to generate a homogeneous and isotropic volume. A simplified triangular approximation of coarse molecular surface was generated from the original surface model by using the Gaussian-based blurring technique. Similar to the widely adopted elastic network model, the finite element model holds a major advantage over standard all-atom normal mode analysis: the computationally expensive process of energy minimization that may distort the initial protein structure has been eliminated. This modification significantly increases the efficiency of normal mode analysis. In addition, the finite element model successfully brings out the capability of normal mode analysis in low-frequency/high collectivity molecular motion by capturing protein shape properties. Fair results from six protein models in this study have fortified the capability of the finite element model in protein normal mode analysis.

**Keywords:** protein normal mode analysis, finite element analysis, protein elastic network model.



# TABLE OF CONTENTS

	Page
<b>ACKNOWLEDGMENTS</b> .....	<b>v</b>
<b>ABSTRACT</b> .....	<b>vi</b>
<b>LIST OF TABLES</b> .....	<b>xi</b>
<b>LIST OF FIGURES</b> .....	<b>xii</b>
<b>CHAPTER</b>	
<b>1. INTRODUCTION</b> .....	<b>1</b>
1.1 Background .....	1
1.1.1 Definition of Proteins .....	1
1.1.2 Protein Folding and Conformational Change .....	3
1.2 Current Studies of Protein Collective Motions .....	6
1.3 Research Objective .....	8
<b>2. LITERATURE REVIEW</b> .....	<b>9</b>
2.1 Normal Mode Analysis (NMA) .....	9
2.1.1 Introduction .....	9
2.1.2 Theory .....	11
2.2 Elastic Network Model in Protein NMA .....	16
2.2.1 Introduction .....	16
2.2.2 Theory .....	17
2.3 Finite Element Model in Protein NMA .....	19
2.3.1 Introduction .....	19
2.3.2 Theory .....	21

<b>3. METHODOLOGY</b> .....	<b>23</b>
3.1 Protein Data Acquiring .....	24
3.2 Protein Solvent-excluded Surface Computation .....	26
3.2.1 MSMS/Fine Molecular Surface .....	28
3.2.2 Coarse Molecular Surface .....	29
3.2.3 Surface Data Processing .....	30
3.3 Model Generation and Meshing .....	30
3.3.1 Mesh elements .....	30
3.3.2 Density and Constitutive Behavior .....	33
3.3.2.1 Density .....	33
3.3.2.2 Young's Modulus .....	34
3.3.2.3 Poisson's Ratio .....	34
3.4 Normal Mode Analysis .....	35
3.5 Post-Processing .....	35
3.5.1 Relative Displacement .....	36
3.5.2 Overlap and Cumulative Overlap .....	36
3.5.3 Degree of Collectivity .....	36
3.5.4 B-factor .....	37
<b>4. RESULTS AND DISCUSSION</b> .....	<b>38</b>
4.1 MSMS/Fine Molecular Surface .....	38
4.1.1 HIV-1 Protease .....	38
4.1.2 Che Y Protein .....	46
4.1.3 LAO Binding Protein .....	49
4.2 Coarse Molecular Surface .....	53
4.2.1 HIV-1 Protease, Che Y Protein, LAO Binding Protein .....	54
4.2.1.1 HIV-1 Protease .....	54
4.2.1.2 Che Y Protein .....	54
4.2.1.3 LAO Binding Protein .....	56
4.2.2 Maltodextrin Binding Protein, Enolase, Lactoferrin .....	56
4.2.2.1 Maltodextrin Binding Protein .....	56
4.2.2.2 Enolase .....	60
4.2.2.3 Lactoferrin .....	60

4.3 Overall Discussion ..... 62

**5. CONCLUSIONS ..... 67**

**BIBLIOGRAPHY ..... 69**

## LIST OF TABLES

Table	Page
3.1 List of targeted proteins; the two pdb codes of each protein represent its “open” and “closed” conformations respectively. ....	24
3.2 List of protein weight and mass density. ....	34
4.1 List of targeted proteins with different surface models; figures in parentheses are the numbers of residues. ....	38
4.2 An overview of protein collectivities and cumulative overlaps. The variation pattern of cumulative overlaps implies that the predictability of conformational changes might be highly correlated to degrees of collectivity. ....	66

## LIST OF FIGURES

Figure	Page
1.1 (a) Amino acids, which usually refer to <i>alpha-amino acids</i> in biochemistry, are molecules containing an amine group ( $\text{H}_2\text{N}-$ ), a carboxylic acid group ( $-\text{COOH}$ ), and a side-chain (R) that is specific to each amino acid. The first carbon that attaches to a functional group refers to the alpha-carbon ( $\text{C}_\alpha$ ) in organic chemistry. (b) Every peptide has a N-terminus residue and a C-terminus residue on the ends of the peptide. ....	2
1.2 Gregory A. Petsko, Dagmar Ringe, 2004 [54]; Levels of protein structure illustrated by the catabolite activator protein: (a) The amino-acid sequence (primary structure) contains all the information needed to specify (b) the regular repeating patterns of hydrogen-bonded backbone conformations (secondary structure) such as alpha helices (red) and beta sheets (blue), as well as (c) the way these elements pack together to form the overall fold of the protein (tertiary structure) (protein PDB: 2CGP); (d) The relative arrangement of two or more individual polypeptide chains is called quaternary structure (protein PDB: 1CGP). ....	4
1.3 Ivett Bahar et al., 2010 [7]; Equilibrium motions of proteins: An overview of the broad range of equilibrium motions accessible under native state conditions, ranging from bond length vibrations, of the order of femtoseconds, to coupled movements of multimeric substructures, of the order of milliseconds or seconds. Collective motions (i.e. global or essential modes) usually occur at the low frequency (i.e. long time scale) end of the mode spectrum and engage large-scale structural rearrangements. ....	7
2.1 In the harmonic approximation of a normal mode analysis, a protein is presumed as an assemblage of (a) harmonic oscillators, and (b) the conformational energy surface at an energy minimum can be approximated by a parabola over the range of thermal fluctuations. ....	10

2.2	Three well known normal mode motions of the water molecule ( $H_2O$ ): (b) symmetric stretching, (c) bending and (d) asymmetric stretching. These three motion patterns are also often found in molecules of various sizes. ....	15
2.3	Moon K Kim et al., 2002 [41]; A representation of protein structure as an elastic network. The backbone trace (i.e. trace of $\alpha$ -carbons) is shown dark lines. The grey lines represent the spring connections between $\alpha$ -carbons within a specific <i>cutoff distance</i> $R_C$ . ....	18
3.1	A schematic illustration of the finite element analysis procedure. The aforementioned three steps are rather conceptual and are presented by blue arrow text boxes. The rectangle boxes represent the actual course for accomplishing the analysis procedure. A detailed explanation of each step is presented in the following sections. ....	25
3.2	Michel F. Sanner et al., 1996 [61]; 2D illustration of the concept of solvent-excluded surface (SES); the SES, shown by black solid line, is defined by rolling a solvent-sized probe-sphere over the van der Waals surface of the protein. ....	27
3.3	Michel F. Sanner et al., 1996 [61]; 3D illustration of the solvent-excluded surface probe-sphere; an adequate probe-sphere radius is usually set as 1.5 Å, representing the size of a water molecule. ....	28
3.4	Too large of a probe-sphere radius will cause separate parts to connect. The demonstrated protein is HIV-1 protease. (Image rendered by PMV 1.5.4) ....	29
3.5	(a) 3D structure of the LAO binding protein; (b) a triangular approximation of the molecular surface; (c) a coarser molecular surface is calculated by using the Gaussian-based blurring technique. ....	31
3.6	Illustration of finding a closed surface; the main concept is to extend the surface in a hierarchical manner. On a closed surface, every edge/side is shared by two areas. Therefore, given randomly a known side (i.e. the red line), there are two adjacent triangles found (i.e. triangles with blue lines). To each of these known triangles, there are also two other triangles adjacent to it (i.e. triangles with gray solid lines). Accordingly, the search process will continue until no more new triangles are found (e.g. yellow lines indicate a halt when two known triangles meet). ....	32

3.7	Illustration of the mesh element SOLID 185.[4] This eight-node element allows for prism and tetrahedral degenerations when used in irregular regions which is suitable for protein models regarding their irregular shapes. ....	33
4.1	HIV-1 protease: Comparisons of the <i>relative displacements</i> for the lowest eight normal modes. (FE model - thick red line; EN model - blue line) The residue index represents the number of alpha-carbons. Correspondence between the FE model and the EN model is quite satisfying regarding two end sections (residue index 1 to 15 and 85 to 99) and the middle sections (residue index 45 to 60) of the charts. Furthermore, the forementioned three sections all have high relative displacements except in the first mode. Indeed, these three sections of the protein chain all undergo relatively high positional changes according to the experimental data. ....	40
4.2	HIV-1 protease: $\alpha$ -carbon/residue displacements between two end conformations, i.e. "open" conformation to "closed" conformation. The middle section and two end sections of the protein chain undergo relatively high positional changes. This corresponds well to the results shown in Figure 4.1. These three sections will later be proven significant regarding the property of the protein shape. (Figure 4.4) ....	41
4.3	HIV-1 protease: B-factors from FE results and experimental data. (FE model - thick red line; experiment - blue line) Data from the FE model differs greatly from the experiment since those B-factors are derived from only twenty modes. However, the B-factors successfully reveal flexibility of residues who play predominant roles in conformational change; both the B-factors from the FE model and $\alpha$ -carbon displacements (Figure 4.2) have peaks around three crucial sections. ....	41
4.4	HIV-1 protease 3D surface model: Residue 1 to 10 (blue) , 45 to 60(red) and 85 to 99(green). These regions located exactly on the protruded parts of the protein which imply that FE model performs well by capturing the property of the molecular shape.(Image rendered with PMV 1.5.4) ....	42

4.5	HIV-1 protease: Values of <i>overlap</i> and <i>cumulative overlap</i> from (a) the FE model and (b) the EN mode. Although two bar charts of overlaps seem quite different at first glance, they both acquired high overlaps within the first few modes. Cumulative overlaps in both charts reached a plateau at approximately the fifth mode. The maximum overlaps of the FE model and the EN model occurred at the second and the third mode, respectively. ....	44
4.6	Comparison of normal modes that hold the maximum overlaps (i.e. dominant modes) : (a) The FE model matches well with the EN model. Two sets of data are similar even though they are retrieved from different modes. (b) Data from the FE model and experiments also yield good correspondency. According Tama et al., [69] the second mode from the FE model might not only be one of the dominant modes but also be that <i>very single mode</i> bearing most of the conformational change. ....	45
4.7	Che Y protein: Comparisons of <i>relative displacement</i> of the first four normal modes. (FE model - thick red line; EN model - blue line) In general, most portion of the data yield great correspondency. ....	47
4.8	Che Y protein: Values of <i>overlap</i> and <i>cumulative overlap</i> from (a) the FE model and (b) the EN mode. Both models have their first two maximum overlaps at the same modes – mode 1 and mode 6. They have comparable ability in predicting dominant modes. However, modal overlaps between the FE results and experimental data were not well predicted. The maximum overlap occurred at the first mode only yielding a value of 0.412. This also happened to the EN model which yielded a value of 0.415. Cumulative overlaps from both models share a poor ascending rate. ....	48
4.9	Che Y protein: Comparison of B-factors between the FE results and experimental data has similarities in the first 20 modes. The FE results have made good predictions of the overall flexibility of the protein. ....	49
4.10	Che Y protein: Both of the two end conformations (a) and (b) (PDB code: 3CHY & 1CHN), have rather globular shape, and thus result in a more complex conformational change. The movement of the protein might be rather localized, and the moving direction of residues might change constantly during conformational change. (Image rendered with Jmol from <a href="http://www.rcsb.org">http://www.rcsb.org</a> ) ....	50



4.11 LAO binding protein: Comparison of <i>relative displacement</i> of the first four normal modes. The two models yield great correspondency especially in the first three modes. (FE model - thick red line; EN model - blue line) . . . . .	51
4.12 LAO binding protein: Values of <i>overlap</i> and <i>cumulative overlap</i> from (a) the FE model and (b) the EN mode. Both models have very high overlaps at the first normal modes and the cumulative overlaps remain almost static after the first mode. These indicate that the first mode is the dominant mode which bearing most of the conformational change. . . . .	52
4.13 LAO binding protein: The necking shape in the middle of the protein might be the cause of a hinge-like motion during conformational change. (Image rendered with PMV 1.5.4) . . . . .	53
4.14 HIV-1 protease: Cumulative overlap from (a) fine surface model and (b) coarse surface model; (c) comparison of relative displacement between fine and coarse surface model at their most dominant modes. Both models capture dominant modes (i.e. modes with highest overlaps) within the first few normal modes. The maximum overlap of coarse models occurs at the fourth mode while for fine model the maximum occurs at the second mode. . . . .	55
4.15 Che Y protein: Cumulative overlap from (a) the fine model and (b) the coarse model. Since the Che Y protein has a pattern of motion in which directions change constantly, the results have poor motion prediction (i.e. low overlap values) for not only the fine model but also the coarse model. In terms of overall performance in over 20 normal modes, the coarse model closely parallels the fine model. Both models yield the maximum overlaps at the first normal mode. (c) Comparison of relative displacement between the fine and the coarse models at their most dominant modes. The tight correspondence between two lines confirms again the capability of the coarse model. . . . .	57
4.16 LAO binding protein: Cumulative overlap from (a) the fine model and (b) the coarse model. (c) Comparison of relative displacement between the fine and the coarse models at their most dominant modes. The coarse model yields not only high overlap at the first mode but almost identical predictions to the fine model. This capability was shown in the previous case (Figure 4.12). . . . .	58

4.17 Maltodextrin binding protein: (a) Values of overlap and cumulative overlap. The pattern of results of maltodextrin binding protein has a great resemblance to those of the LAO binding protein (Figure 4.17(a)(b)). Likewise, the maximum overlap happens at the first normal mode with a remarkably high value. (b) The comparison of relative displacement between the most dominant mode and experimental data provides a prominent demonstration to the correspondency in (a); the majority of the data at the first mode coinciding with the experimental data. ....	59
4.18 Enolase: (a) Relative displacement and (b) values of <i>overlap</i> and <i>cumulative overlap</i> . The bar chart shows quite a dispersive distribution of overlap of enolase protein. While the first three maximum overlaps all happen within the first four normal modes, the cumulative rate of overlaps over twenty normal modes is relatively low. ....	61
4.19 Enolase: The movement of the second normal mode. This normal mode holds the maximum overlap and is also the only collective motion predicted in the dominant modes. (Image rendered with ANSYS 13.0) ....	62
4.20 Lactoferrin: (a) Relative displacement and (b) values of <i>overlap</i> and <i>cumulative overlap</i> . Lactoferrin has a relatively concentrated distribution of overlaps. All dominant modes (i.e. high overlap modes) occurred within the lowest few modes. Cumulative overlaps reach a fair plateau around the eighth normal mode. The maximum overlap is found at the third normal mode which corresponds to the inference by Tama et al.– the most dominant modes are very often found within the first four modes.[69] ....	63
4.21 Lactoferrin: (a)(b) – Conformational change from experimental data. (c)(d) – Motion pattern of the 3 <sup>rd</sup> normal mode. (e)(f) – Motion pattern of the 4 <sup>th</sup> normal mode. The similarity among these three sets of motion pattern indicates that these two dominant normal modes perform well on predicting the collective motion. (Image rendered with Rasmol and ANSYS) ....	64

# CHAPTER 1

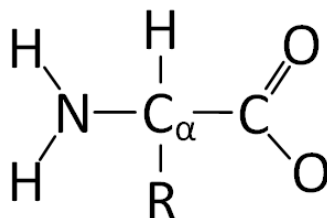
## INTRODUCTION

### 1.1 Background

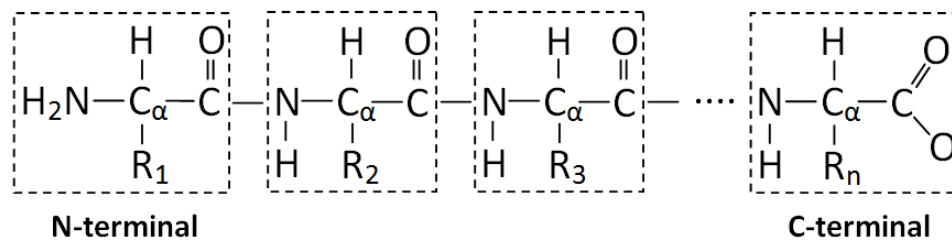
The pursuit of understanding protein structures and functional activities has long been a critical concern in life sciences. It is well known that proteins are not static objects with fixed configurations but are instead dynamic actors. In addition, protein dynamics play a fundamental role in many diseases, and spans a wide timescale, from picoseconds to milliseconds or even longer. To understand why and how the proteins hold their functionally significant behavior, it is essential to review some basic knowledge about proteins. Hereafter, we will present a review of the biological relevance of protein structural informations.

#### 1.1.1 Definition of Proteins

Proteins are fundamental biochemical compounds and are building blocks of most living organisms. They are linear chains of *amino acids* (Figure 1.1) linked by peptide bonds. The genetic code specifies 22 standard amino acids, which underlie countless linear combinations of proteins. In general, amino acid chains consist of more than two and less than fifty units are called *peptides*. Amino acids which have been incorporated into a peptide are termed *residues*, every peptide has a N-terminus and C-terminus residue on the ends of the peptide. Amino acid chains in much greater sizes (e.g. 500 residues) are known as *polypeptides*. In other words, a protein is one or more polypeptides with specific dimensional structure and assigned functions. Biochemists often refer to four distinct aspects of a protein's structure (Figure 1.2):



(a)



(b)

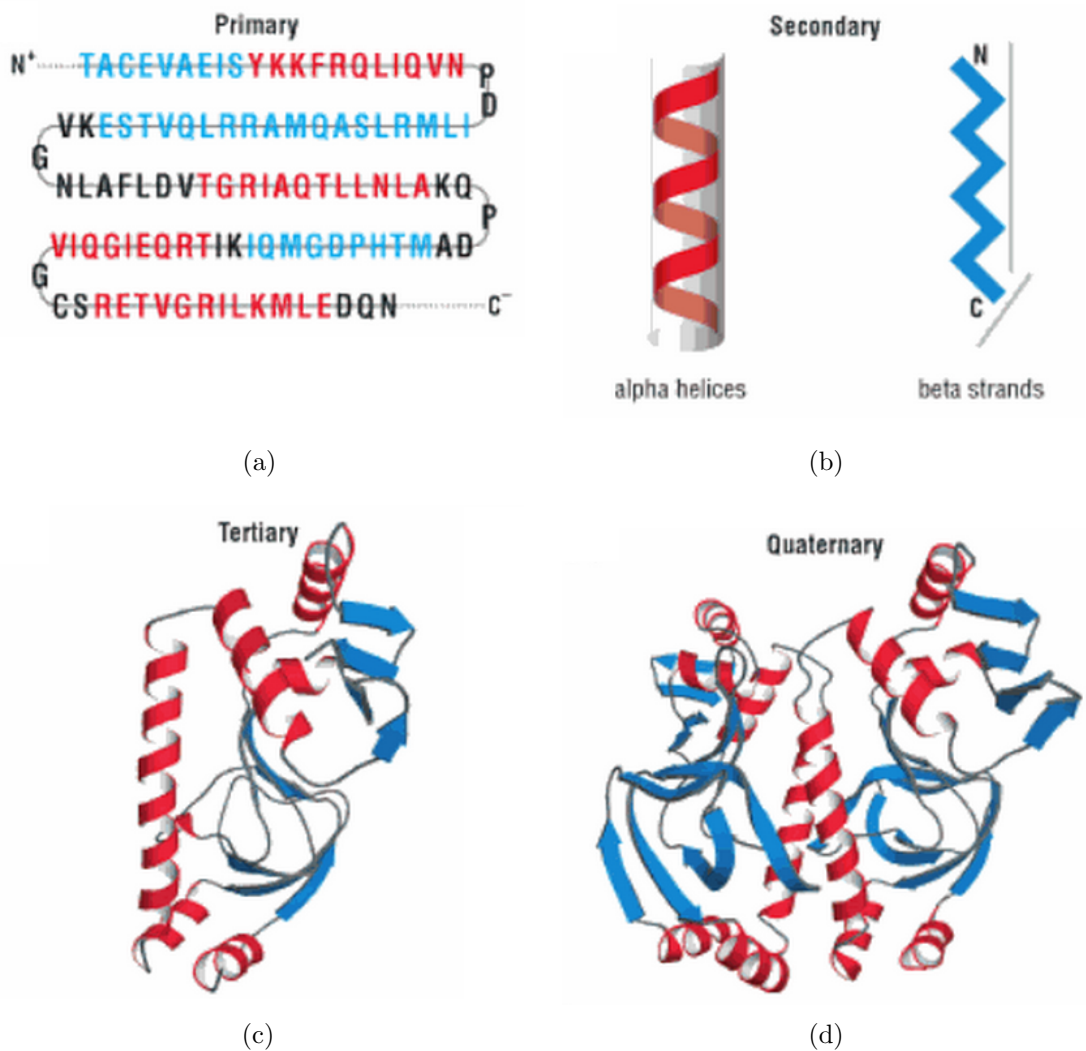
**Figure 1.1.** (a) Amino acids, which usually refer to *alpha-amino acids* in biochemistry, are molecules containing an amine group ( $\text{H}_2\text{N}-$ ), a carboxylic acid group ( $-\text{COOH}$ ), and a side-chain ( $\text{R}$ ) that is specific to each amino acid. The first carbon that attaches to a functional group refers to the alpha-carbon ( $\text{C}_\alpha$ ) in organic chemistry. (b) Every peptide has a N-terminus residue and a C-terminus residue on the ends of the peptide.

- *Primary structure*: the amino acid sequence. (Figure 1.2(a))
- *Secondary structure*: regularly repeating local structures stabilized by hydrogen bonds. The most common examples are the *alpha helix*, *beta sheet* and *turns*. (Figure 1.2(b))
- *Tertiary structure*: the overall shape of a single protein molecule; the spatial relationship of the secondary structures to one another. The term “tertiary structure” is often used as synonymous with the term *fold*. The tertiary structure is what controls the basic function of the protein. (Figure 1.2(c))
- *Quaternary structure*: the structure formed by several protein molecules (i.e. polypeptide chains) , which function as a single protein complex. (Figure 1.2(d))

In terms of the tertiary structure, each protein has its unique three-dimensional structure which usually refers to “*conformation*”. A *native* conformation of a protein is acquired in the first place by folding its lengthy peptide chains and forming into a certain structure. This protein forming process is known as “*folding*”. In some cases, a protein can have more than one native conformation.

### 1.1.2 Protein Folding and Conformational Change

Folding is a natural but delicate process that a protein undergo in order to achieve its stable and unique conformation. In this regard, protein structures are confined to a global energy minimum (or the aforementioned *native state*).<sup>[7]</sup> This process takes place in a highly crowded, complex, molecular environment within the cell. Many proteins can fold unassisted, simply through the chemical properties of their amino acids, while others require the aid of molecular chaperones to fold into their native states. Notably, folding accounts for the main structural difference between proteins and other chain molecules; by folding, proteins achieve narrower distribution of conformations comparing to disordered polymers.<sup>[7]</sup>



**Figure 1.2.** Gregory A. Petsko, Dagmar Ringe, 2004 [54]; Levels of protein structure illustrated by the catabolite activator protein: (a) The amino-acid sequence (primary structure) contains all the information needed to specify (b) the regular repeating patterns of hydrogen-bonded backbone conformations (secondary structure) such as alpha helices (red) and beta sheets (blue), as well as (c) the way these elements pack together to form the overall fold of the protein (tertiary structure) (protein PDB: 2CGP); (d) The relative arrangement of two or more individual polypeptide chains is called quaternary structure (protein PDB: 1CGP).

When a protein acquired its native conformations, it does not remain static; there is a *dynamic equilibrium* which is induced by thermal fluctuation. The thermal fluctuation of atoms that were originally viewed as random and stochastic events, actually account for a local relaxation phenomena in nanoseconds regime.[7] They may facilitate, for instance, diffusion of oxygen into myoglobin[73] or permeation of ions across ion channels.[63, 59, 58] Furthermore, these thermal fluctuations may also facilitate concerted domain movements or allosteric interactions.[26] In other words, these atomic scale movements may intrinsically influence protein behavior in broad length (i.e. angstroms to millimeters) and time (i.e. femtoseconds to milliseconds) scales.(Figure 1.3) Moreover, numerous studies have indicated that those equilibrium dynamics underlie the “*collective motions*”, which play important roles in protein biological functions.[38, 55, 19] As a result, the conformation of a protein is usually flexible and dynamic. A protein can change its conformations in response to changes in its environment or other factors; the transition is called a *conformational change*.

Notably, the mechanisms behind folding and conformational change, both biologically and mechanically, are considerably distinct. While protein folding results solely from the arrangement of amino acid sequence, a conformational change may be induced by many factors such as a change in temperature, pH, voltage, ion concentration, phosphorylation, or the binding of a ligand. However, to some degree, there is an analogy between folding and conformational changes; amino acid sequence encodes structure (i.e. folding) while structure encodes equilibrium dynamics (i.e. conformational change).[7]

In protein dynamics, conformational changes hold functional significance which is the cause for this study. In general, functional movements involve collections of either microstates or substates in a dynamic equilibrium. However, collective motions which engage large substructures or even the entire structure, are dominant in overall conformational changes. These collective motions are designated as global or essential

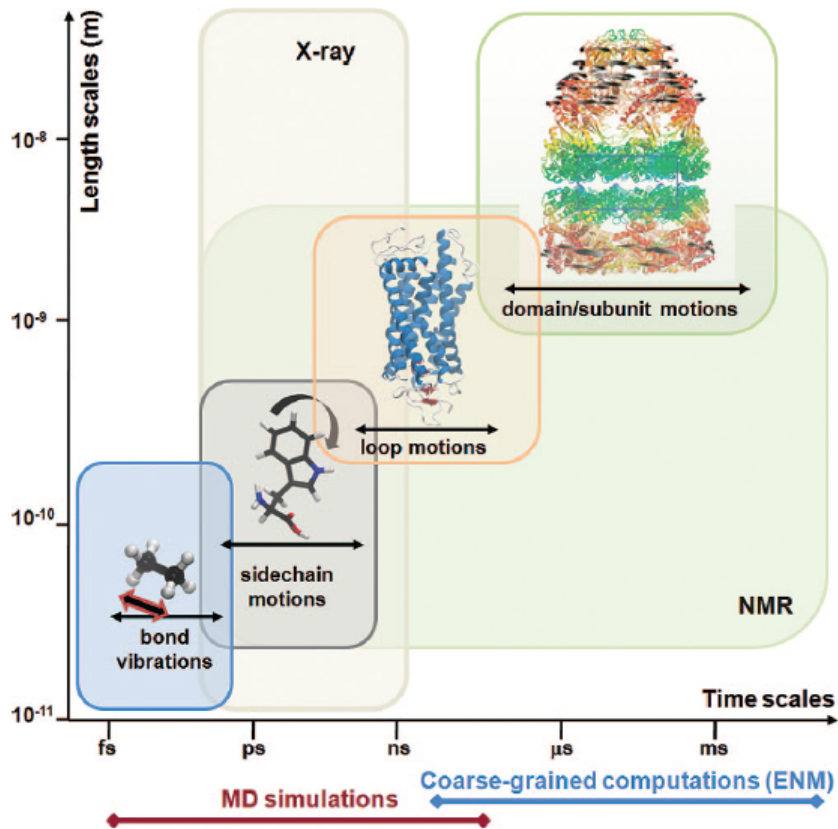
modes.[7] They usually occur at the low frequency (i.e. long time scale) end of the mode spectrum and engage large-scale structural rearrangements.[49] (see Figure 1.3)

## 1.2 Current Studies of Protein Collective Motions

Function-related collective motions have always been a major concern in protein dynamics studies. In addition to many biophysical techniques, such as X-ray crystallography, nuclear magnetic resonance (NMR), and electron paramagnetic resonance (EPR) that are widely used to study macromolecular motions. Mathematical methods based on principal components analysis (PCA) have also been introduced to this area. In the past two decades, a large number of studies based on PCA have provided great contributions for understanding protein low-frequency and collective motions. Normal mode analysis (NMA) of equilibrium structure [55, 19], essential dynamics analysis (EDA) of the covariance matrices retrieved from MD runs [2], and singular value decomposition (SVD) of Molecular Dynamic (MD) or Monte Carlo (MC) trajectories [37, 27, 57] are all in the category of PCA-based methods; these methods can provide better insights into protein motions by bridging with the PCA.

An efficient and predominant approach among the aforementioned methods, in terms of increasing computational efficiency and decreasing modeling resolution, is the normal mode analysis (NMA). Normal mode analysis assumes the system is stabilized by harmonic potentials at its native state and provides information of equilibrium modes by solving the system as a free vibration problem. Its application to proteins can date back to the early 1980s.[53, 12, 67, 50] For the past decade, it has been widely used for exploring protein functional motions. The major reason to its broader use is the global modes unraveled by normal mode analysis bear functional significance. This feature became even more evident with the use of simplified models in coarse-grained normal mode analysis (CG NMA). A widely adopted CG NMA technique is





**Figure 1.3.** Ivet Bahar et al., 2010 [7]; Equilibrium motions of proteins: An overview of the broad range of equilibrium motions accessible under native state conditions, ranging from bond length vibrations, of the order of femtoseconds, to coupled movements of multimeric substructures, of the order of milliseconds or seconds. Collective motions (i.e. global or essential modes) usually occur at the low frequency (i.e. long time scale) end of the mode spectrum and engage large-scale structural rearrangements.

the elastic network (EN) model. It enables NMA computation for very large biological molecules in a very short amount of time.[49]

### **1.3 Research Objective**

In consequence of the success of normal mode analysis in protein functional motions, this study performs an analysis based on normal mode analysis for understanding protein conformational changes. However, the protein models are solved by the finite element (FE) method, which is a less common approach in protein dynamics. The FE method is a relatively new technique in protein normal mode analysis. It was first applied by Bathe et al. for solving protein and macromolecular assemblies based on normal mode analysis.[9] This study aims to establish an analytic procedure of FE-based normal mode analysis to provide an efficient and competent analysis option for understanding protein conformational changes.

## CHAPTER 2

### LITERATURE REVIEW

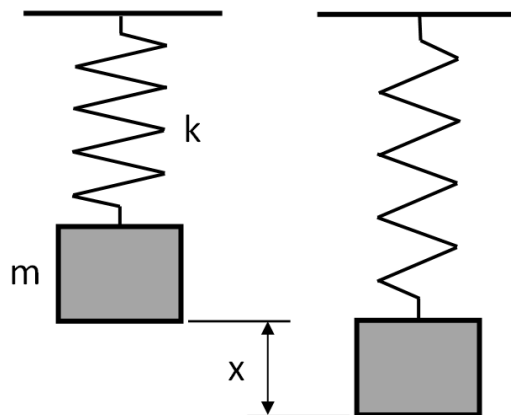
#### 2.1 Normal Mode Analysis (NMA)

##### 2.1.1 Introduction

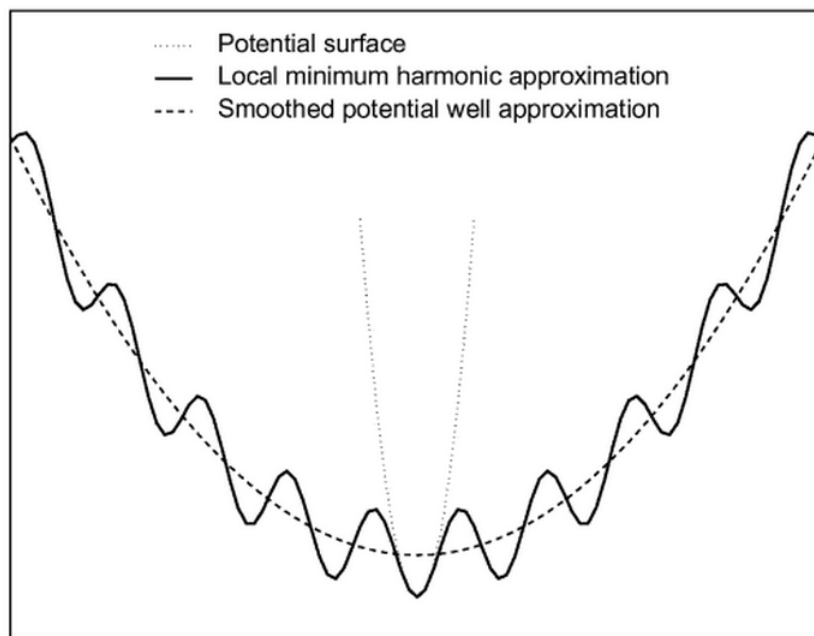
Normal mode analysis (NMA) is a technique based on classical mechanics. It is the study of motion patterns and dynamic response of a system (e.g. structures or fluids) when excited by an input. A normal mode of a system is a pattern of concerted motion that all parts of the system move sinusoidally with the same frequency. Under the motion of a certain normal mode, the center of mass of the system does not move and all parts pass through their equilibrium positions at the same time. Normal modes are independent; they do not interact with each other. The frequencies of the normal modes are known as *natural frequencies* or *resonant frequencies*. Any system has a set of normal modes that depend on its structure and mechanical properties (e.g. Young's modulus, Poisson's ratio).

In normal mode analysis, the modes of greatest fluctuation are those with the lowest frequencies. In this regard, normal mode analysis is one of the major simulation techniques used to probe low-frequency, large-scale, shape-changing motions in biological molecules.[7, 69] In the harmonic approximation of a normal mode analysis, the conformational energy surface at an energy minimum is approximated by a parabola over the range of thermal fluctuations.(Figure 2.1(b))

Originally, the normal mode analysis requires three main steps of performance in the Cartesian coordinate space: (1) minimization of the conformational potential energy as a function of the atomic Cartesian coordinates; (2) the calculation of the so-



(a) A simple harmonic oscillator. The behavior of a harmonic oscillator is described as  $m \frac{d^2x}{dt^2} = -kx$ , where  $m$  is the mass,  $k$  is an elastic constant and  $x$  is the displacement.



(b) Sen, T. Z. and Jernigan, R. L., 2006 [65]; A schematic one-dimensional view of the potential energy surface of a protein showing two kinds of harmonic approximations: an approximation to a local minimum, and an approximation to the smoothed-out potential well.

**Figure 2.1.** In the harmonic approximation of a normal mode analysis, a protein is presumed as an assemblage of (a) harmonic oscillators, and (b) the conformational energy surface at an energy minimum can be approximated by a parabola over the range of thermal fluctuations.

called Hessian matrix, which is the matrix of second derivatives of the potential energy with respect to the atomic coordinates; and (3) the diagonalization of the Hessian matrix. The final step yields eigenvalues and eigenvectors of the *normal modes*. Usually, the first and final steps are the bottlenecks which led to the late revival of the application of normal mode analysis. The energy minimization and diagonalization are computationally demanding, both of CPU time and memory, especially while the number of atoms increases. Normal mode analysis (NMA) that requires the aforementioned manners is called *standard* NMA, to distinguish it from other coarse-grained NMAs such as elastic network modeled NMA.

### 2.1.2 Theory

A standard NMA is usually performed in a vacuum, and the dynamics of system is represented as a set of harmonic oscillators. Consider a system containing  $N$  interaction sites (e.g. a molecule with  $N$  atoms). In Cartesian coordinates, the potential energy function  $V$ , near the equilibrium conformation, can be expressed in a Taylor series [7]:

$$\begin{aligned}
 V(\mathbf{q}) = & V(\mathbf{q}^0) + \sum_i \left( \frac{\delta V}{\delta q_i} \right)^0 (q_i - q_i^0) + \\
 & \frac{1}{2} \sum_{i,j} \left( \frac{\delta^2 V}{\delta q_i \delta q_j} \right)^0 (q_i - q_i^0)(q_j - q_j^0) + \dots
 \end{aligned}
 \tag{2.1}$$

, where  $\mathbf{q}^0$  is the equilibrium conformation. The first term is the minimum value of the potential, which may be set to zero. The second term is identically zero at any local minimum of the potential. If the expansion is terminated at the quadratic level, the potential energy can be expressed as:

$$\begin{aligned}
V(\mathbf{q}) &= \frac{1}{2} \sum_{i,j} \left( \frac{\delta^2 V}{\delta q_i \delta q_j} \right)^0 (q_i - q_i^0)(q_j - q_j^0) \\
&= \frac{1}{2} \sum_{i,j} (q_i - q_i^0) H_{ij} (q_j - q_j^0) = \frac{1}{2} \Delta \mathbf{q}^T \mathbf{H} \Delta \mathbf{q}
\end{aligned} \tag{2.2}$$

, where  $\mathbf{H}$  is the Hessian matrix obtained from the second derivatives of the potential with respect to the components of  $\mathbf{q}$  [7]:

$$H_{ij} = \left( \frac{\delta^2 V}{\delta q_i \delta q_j} \right)^0 \tag{2.3}$$

The Hessian matrix  $\mathbf{H}$  is an  $N \times N$  matrix of  $3 \times 3$  submatrices, each describes the energetic contribution from the interaction of two sites. Two important properties of the Hessian are: (1) The  $\mathbf{H}$  is real and symmetric and is therefore diagonalized by an orthogonal transformation. If  $\mathbf{H}$  is not symmetric, its eigenvectors would not form an orthonormal basis over the full space of molecular motions and normal mode analysis could not be performed. (2) None of the eigenvalues of  $\mathbf{H}$  can be negative if  $\mathbf{H}$  is constructed at a local potential energy minimum. The sign of a given eigenvalue indicates the local curvature of the potential along the corresponding mode directional vector or eigenvector: Positive eigenvalues indicate local minima, and negative eigenvalues indicate local maxima. The local potential energy landscape for a system in a potential energy minimum will have only positive or zero curvature in all directions. Eigenvalues that are identically zero indicate conformational changes that have no effect on the systems (internal) potential energy. Typically,  $\mathbf{H}$  has six zero eigenvalues, corresponding to the *rigid-body rotations* and *translations* of the whole molecule, which yield to  $3N - 6$  internal degrees of freedom (i.e.  $3N - 6$  sets of valid solutions).

Given the potential of the system, the total energy of the system can be described by the *Hamiltonian* as [49]:

$$\begin{aligned}
K(\mathbf{q}) + V(\mathbf{q}) &= \frac{1}{2} \sum_i m_i \frac{dq_i^2}{dt} + \frac{1}{2} \sum_{i,j} \left( \frac{\delta^2 V}{\delta q_i \delta q_j} \right)^0 (q_i - q_i^0)(q_j - q_j^0) \\
&= \frac{1}{2} \sum_i m_i \frac{dq_i^2}{dt} + \frac{1}{2} \Delta \mathbf{q}^T \mathbf{H} \Delta \mathbf{q}
\end{aligned} \tag{2.4}$$

, where  $K(\mathbf{q})$  represents the *kinetic energy*, and  $m_i$  represents the mass of the atom  $i$ . Another commonly used equation for describing the mechanical behavior of the system is the *equation of motion* which can be written as [7]:

$$\mathbf{M} \frac{d^2 \Delta \mathbf{q}}{dt^2} + \mathbf{H} \Delta \mathbf{q} = 0 \tag{2.5}$$

, where  $M$  represents a diagonal matrix containing the masses of the atoms. Both equations lead to same solutions of eigenvalues and eigenvectors. Usually, the Hessian matrix is solved by transforming the system coordinates into *mass-weighted* coordinates,  $X_i = \sqrt{m_i}(q_i - q_i^0) = \sqrt{m_i} \Delta q_i$ . As a result, Eq.2.4 can be rewritten as:

$$\begin{aligned}
K(\mathbf{X}) + V(\mathbf{X}) &= \frac{1}{2} \sum_i \frac{dX_i^2}{dt} + \frac{1}{2} \sum_{i,j} \left( \frac{\delta^2 V}{\delta X_i \delta X_j} \right)^0 X_i X_j \\
&= \frac{1}{2} \sum_i \frac{dX_i^2}{dt} + \frac{1}{2} \mathbf{X}^T \tilde{\mathbf{H}} \mathbf{X}
\end{aligned} \tag{2.6}$$

, where the *mass-weighted* Hessian can be acquired from the above equation [49]:

$$\tilde{H}_{ij} = \frac{\delta^2 V}{\delta X_i \delta X_j} \tag{2.7}$$

The diagonalization of mass-weighted Hessian yields to a generalized eigenvalue equation:

$$\begin{aligned}
\tilde{\mathbf{H}} \mathbf{u}_k &= \lambda_k \mathbf{u}_k \\
&= \omega_k^2 \mathbf{u}_k
\end{aligned} \tag{2.8}$$

, where  $\mathbf{u}_k$  and  $\lambda_k$  are the  $k$ th eigenvector and eigenvalue respectively, and  $\omega_k$  represents the frequency of the mode of motion. The solutions are normally sorted in ascending order of the eigenvalue, providing the eigenvector matrix  $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{3N-6})$  and eigenvalue matrix  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{3N-6})$ . Again, it is noted that among the total of  $N$  normal modes, only  $3N - 6$  of them are meaningful; the first six normal modes have eigenvalues equal to 0 and correspond to rigid-body translations and rotations of the whole system. Consequently, the dynamics of the system can be described as a linear combination of “independent” normal mode oscillators; the atomic displacements can be expressed as the sum of normal mode contributions [20]:

$$X_i = \sqrt{m_i} \Delta q_i = \sum_k \mathbf{u}_{ki} Q_k \quad (2.9)$$

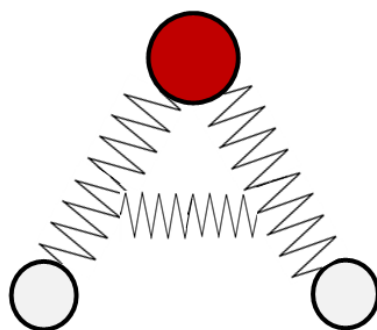
, where  $Q_k$  is the  $k$ th normal mode coordinates. Note that the sum is over all normal modes at site  $i$ , and  $\sum_k \mathbf{u}_{ki} = 1$ . In other words, in normal mode coordinates, the  $k$ th normal mode variable,  $Q_k$ , oscillate with the frequency  $\omega_k$  in a set of directions given by the eigenvector,  $\mathbf{u}_k$ . The original Cartesian coordinates is therefore derived as:

$$\Delta q_i = \frac{1}{\sqrt{m_i}} \sum_k \mathbf{u}_{ki} Q_k \quad (2.10)$$

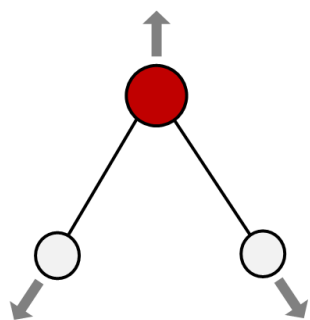
Figure 2.2 shows an example of the resulting normal modes obtained for a small model such as a water molecule. Normal mode analysis reveals three well known motions of the water molecule, i.e. *bending* mode, *symmetric stretching* mode and *asymmetric stretching* mode.

Furthermore, according to *equipartition theorem* of a thermal equilibrium system, the vibrational energy is equally partitioned among all the modes [6]. That is, the average potential of each mode is equal to  $k_B T/2$ , where  $T$  is the absolute temperature

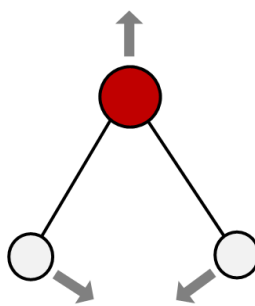




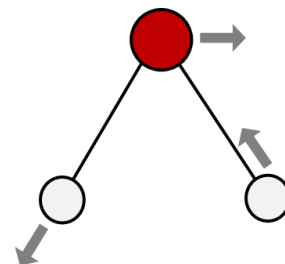
(a) A molecular model of  $H_2O$



(b) Symmetric stretching



(c) Bending



(d) Asymmetric stretching

**Figure 2.2.** Three well known normal mode motions of the water molecule ( $H_2O$ ): (b) symmetric stretching, (c) bending and (d) asymmetric stretching. These three motion patterns are also often found in molecules of various sizes.

and  $k_B$  is the *Boltzmann constant*. To this end, the potential energy of the  $k$ th mode is  $\frac{1}{2}\omega_k^2 \langle Q^2 \rangle = \frac{1}{2}k_B T$ , which yields the following relation:

$$\langle Q^2 \rangle = \frac{k_B T}{\omega_k^2} \quad (2.11)$$

, where  $\langle Q^2 \rangle$  denotes the average value of  $Q^2$ . This implies that the average amplitude of oscillation along mode  $k$  scales with  $1/\omega_k^2$ . Thus, the molecule experiences the greatest displacement along the lowest frequency modes (i.e. the “slowest” modes).

## 2.2 Elastic Network Model in Protein NMA

### 2.2.1 Introduction

One of the major differences between the elastic network normal mode analysis (EN NMA) and the standard NMA is the energy minimization. As elucidated previously, a standard NMA requires energy minimization prior to performing NMA, to ensure that the first derivative of the total potential is zero with respect to all degrees of freedom and to evaluate the second derivatives (i.e. Hessian matrix). However, energy minimization is computationally expensive and generally distorts the initial conformation. As a result, normal mode analysis is usually performed on a structure altered from the original. These drawbacks aroused a surge of studies in coarse-grained NMAs, including the EN NMA.[55, 19] The term coarse-grained refers to the simplification of models in which numbers of atoms are greatly cut down. This significantly decreases the size of the Hessian matrix while evaluating solutions. Besides, no energy minimization is required prior to the EN NMA since the initial conformation, taken directly from the crystallography structure, is assumed to be an energy minimum. These features enable the elastic network model to analyze large biomolecular complexes and assemblies with competent efficiency.

Another motivation of the application of elastic network normal mode analysis (EN NMA) is the robustness of protein global motions. As mentioned in the previ-

ous chapter, low-frequency global motions usually hold functional significance. These global motions are found to be favored by the overall architecture despite of detailed structural and energetic changes.[21, 52, 74] More specifically, the property that dominates the global modes is the network of inter-residue contacts, which is a purely geometric quality defined by the overall shape or native contact topology. To this end, the EN NMA, which simplifies the model without distorting its overall shape, can greatly enhance computational efficiency and still, conserve the essence of global movements. Indeed, in recent years, the EN-based NMAs have contributed greatly to improve the understanding of collective/global dynamics in membrane proteins.

Two commonly used elastic network (EN) models are the Gaussian network model (GNM) [22, 23] and the anisotropic network model (ANM) [5, 24, 15]. In both models, the structure is represented as a network of nodes, i.e.  $\alpha$ -carbons ( $C_\alpha$ ), and elastic springs (Figure 2.3). The springs connect the  $\alpha$ -carbon pairs that lie within a pre-specified cutoff distance,  $R_C$ , in the native structure. The EN model then approximates the potential energy as that of a classical network of masses coupled by springs. The main difference between GNM and ANM is that the former uses an  $N \times N$  Kirchhoff matrix, as opposed to the  $3N \times 3N$  Hessian in ANM. As a consequence, the ANM became the most broadly used EN model since no information on the three-dimensional direction of motions can be obtained by the GNM. Herein, we only consider the ANM, and any latter discussions of EN model will only refer to the ANM.

### 2.2.2 Theory

In the case of elastic network model (i.e. ANM), the Hessian is derived from the following potential energy function [7]:

$$V = \frac{1}{2} \sum_{|R_{ij}^0| < R_C} \gamma_{ij} (R_{ij} - R_{ij}^0)^2 \tag{2.12}$$



**Figure 2.3.** Moon K Kim et al., 2002 [41]; A representation of protein structure as an elastic network. The backbone trace (i.e. trace of  $\alpha$ -carbons) is shown dark lines. The grey lines represent the spring connections between  $\alpha$ -carbons within a specific *cutoff distance*  $R_C$ .

, where  $R_{ij}$  is the distance between atoms  $i$  and  $j$  and  $R_{ij}^0$  is the distance between the atoms in the original crystallographic structure. The summation here is only performed over atoms within the cut-off distance  $R_c$ . The force constant,  $\gamma_{ij}$ , of the spring represents chemical bond or atomic force between atoms  $i$  and  $j$ . In most applications,  $\gamma_{ij}$  is taken as a constant,  $\gamma$ , for all pairs of residues connected in the network.[71] However, some studies adopted varying force constants: Hinsen proposed using a force constant that decays rapidly with distance [34]; Sen and Jernigan demonstrated how the force constants vary with the residue coordination numbers [65]; the adoption of stiffer springs for sequentially neighboring residues [44] or amino acid-specific force constants [31, 45] has been shown to improve the agreement with experiments.

Nonetheless, the choice of the specific spring constants has little effect on the global modes. Since the global modes are mostly dominated by intrinsic properties of

the shape of a protein, they have been confirmed in several studies to be insensitive to model parameters.[21, 52, 74] Additionally, the absolute value of spring constant for a given level representation does not affect the mode shapes (i.e. the eigenvectors of  $\mathbf{H}$ ) but their frequencies, since the eigenvalues,  $\lambda_k$ , are proportional to  $\gamma$ . Likewise, the global modes are insensitive to the adoption of residue-specific force constants. Therefore, In this study, the force constant is set as *unity* for all interactions. Once Eq.2.12 has been calculated, the procedure is the same as for the standard NMA; the Hessian is calculated and its eigenvalues and eigenvectors are determined. As mentioned, the elastic network model is carried out on a subset of atoms, which are usually the  $\alpha$ -carbons,  $C_\alpha$ , instead of all atoms in the protein. This would result in a Hessian approximately tenfold lower in order compared with the standard NMA. The computational cost for calculating the eigenvalues and eigenvectors is, therefore, reduced considerably.

## 2.3 Finite Element Model in Protein NMA

### 2.3.1 Introduction

The finite element (FE) method is a powerful numerical technique which was developed for solving complex problems in structural mechanics. Its origin can be traced back to the matrix analysis of structures where the concept of displacement or stiffness matrix approach was introduced.[72] It is common to apply the FE method for modal analysis (i.e. normal mode analysis), since the object being analyzed can have arbitrary shape and the results of the calculations are acceptable. In the FE method, the structural system is modeled by a set of appropriate finite elements (whether one-, two-, or three-dimensional) interconnected to each other. These elements are generated by “*meshing*”, which is a technique that discretize a continuous domain(i.e. whole system) into a set of discrete sub-domains. Elements may have physical properties such as thickness, coefficient of thermal expansion, density, Young’s modulus,

shear modulus and Poisson's ratio. Three-dimensional elements are used for modeling 3D solids such as machine components, dams, soil masses or ,most specially in this study, proteins. Common three-dimensional element shapes include tetrahedrals and hexahedrals. (Nodes to be concerned in solution are placed at the vertexes of elements and possibly in the element faces or within the element.) In the early applications of the FE method, molecular shapes of proteins and their assemblies were assumed a priori, i.e. molecules are typically modeled as regular geometric objects such as cylinders, spheres, and sheets.[47, 70] However, while the meshing technique has been improved, molecules are modeled according to their intrinsically irregular shape.[9, 40]

Similar to elastic network models, the proposed finite element based procedure offers several distinct advantages over standard NMA. Firstly, the costly energy minimization that may distort the initial protein structure is eliminated. Secondly, it provides direct applicability to X-ray data of proteins with unknown atomic structure.[48, 68] Thirdly, the FE method is suitable to calculate the mechanical response of proteins and their supramolecular assemblies to applied bending, buckling, and other generalized loading scenarios.[47, 36, 42, 51, 16, 66, 35, 70] Nonetheless, a great feature that facilitates the applicability of FE method is the robustness of protein global motion. As emphasized previously, global modes are widely recognized to be intrinsic properties of the shape of the protein and are insensitive to minor perturbation of parameters. Since the protein structures in FE model is isotropic and shape/structure sensitive, it favors the intrinsic properties of global modes.

Indeed, the FE method has achieved considerable success in the computational modeling of tissue and cell mechanics.[39] Moreover, in the cases of further application, the FE-based protein model may be coupled directly to field calculations including the Poisson-Boltzmann equation to model aqueous electrolyte-mediated elec-

trostatic interactions [33, 29] and the Stokes equations to model solvent damping [14, 18, 60].

### 2.3.2 Theory

In the finite element (FE) model, given the protein volume, constitutive behavior, and boundary conditions, a set of equations of motion can be derived as follows [28]:

$$\mathbf{M} \frac{d^2 \Delta \mathbf{q}}{dt^2} + \mathbf{K} \Delta \mathbf{q} = \mathbf{r} \quad (2.13)$$

, where  $\Delta \mathbf{q}$  is the finite element nodal displacement degrees of freedom,  $\mathbf{M}$  is the diagonal mass-matrix,  $\mathbf{K}$  is the elastic stiffness matrix, and  $\mathbf{r}$  is a force vector that results from boundary conditions. In the case of the free vibration problem, i.e.  $\mathbf{r} = 0$ , Eq.2.13 is basically the same as Eq.2.5. Substitution of the oscillatory solution:  $\Delta \mathbf{q}_k = \mathbf{u}_k \cos(\omega_k t + \phi)$ , into the free vibration form of Eq.2.13 results in the generalized eigenvalue problem:

$$\mathbf{K} \mathbf{u}_k = \lambda_k \mathbf{M} \mathbf{u}_k \quad (2.14)$$

, where  $\lambda_k$  is the eigenvalue and  $\lambda_k = \omega_k^2$ . The solutions to the problem, similar to the EN model, will yield pairs of eigenvalues and eigenvectors, i.e.  $(\lambda_k, \mathbf{u}_k)$ . The corresponding solutions to the  $\alpha$ -carbons can be derived by mapping the finite element nodal degrees of freedom onto the original atomic coordinates.

While the FE model is conceptually similar to elastic network based models, there is a great distinction: The EN models typically connect  $C_\alpha$  atoms by springs of equal stiffness, which results in a locally *anisotropic* and *inhomogeneous* elastic material with length-scale dependent mechanical properties. In contrast, the FE model treats the protein as a *homogeneous continuum* solid with an *isotropic* elastic material response.

Hereafter, we will focus on the application of FE method to the computation of protein normal modes and readers are referred to the references [8, 75] for comprehensive details on its theoretical foundations.



## CHAPTER 3

### METHODOLOGY

To generate the finite element model, three steps are required: (1) definition and discretization of the protein volume; (2) definition of constitutive behavior and the mass density of the protein; and (3) application of boundary conditions such as loading.

Regarding the first step, the protein volume is defined by its bounding *solvent-excluded surface*, which is also called the Richards molecular surface or simply the molecular surface. This surface is defined by the closest point of contact of a solvent-sized probe-sphere that is rolled over the van der Waals surface of the protein. The molecular volume defined by the solvent-excluded surface is never penetrated by any part of the solvent probe-sphere.[30, 17] Herein, the solvent-excluded surface is computed by using MSMS 2.6.1 [61], a computational software package which generates triangulated approximation to the solvent-excluded surface. The protein volume bound by the closed solvent-excluded surface is then generated and discretized for further analysis by using proper finite element programs or software. The finite element software used here is ANSYS<sup>®</sup> 13.0.[1]

Secondly, the protein constitutive response is modeled using the standard Hooke's Law, which treats the protein as a homogeneous, isotropic, elastic continuum with Young's modulus  $E$  and Poisson ratio  $\nu$ . [28] As pointed out in previous chapter, while the finite element (FE) model is conceptually similar to the elastic network (EN) model, a great distinction still exists. The EN model connects  $\alpha$ -carbon atoms by springs of equal stiffness, which results in a locally anisotropic and inhomogeneous

elastic material with length-scale dependent mechanical properties. In contrast, the FE model defined here treats the protein as homogeneous material, with an isotropic elastic response that is length-scale invariant.

Finally, in the third step, arbitrary boundary conditions that consist of displacement or force based loading may be applied to the molecule, modeling after the effects of the protein environment. In this application, all proteins are solved as free vibration problems in the absence of any boundary condition.

The following diagram (Figure 3.1) shows the comprehensive procedure of this finite element based normal mode analysis. While the three steps mentioned before are rather conceptual, this flow chart demonstrates the actual course for accomplishing the application. A detailed explanation of each step is presented in the following sections.

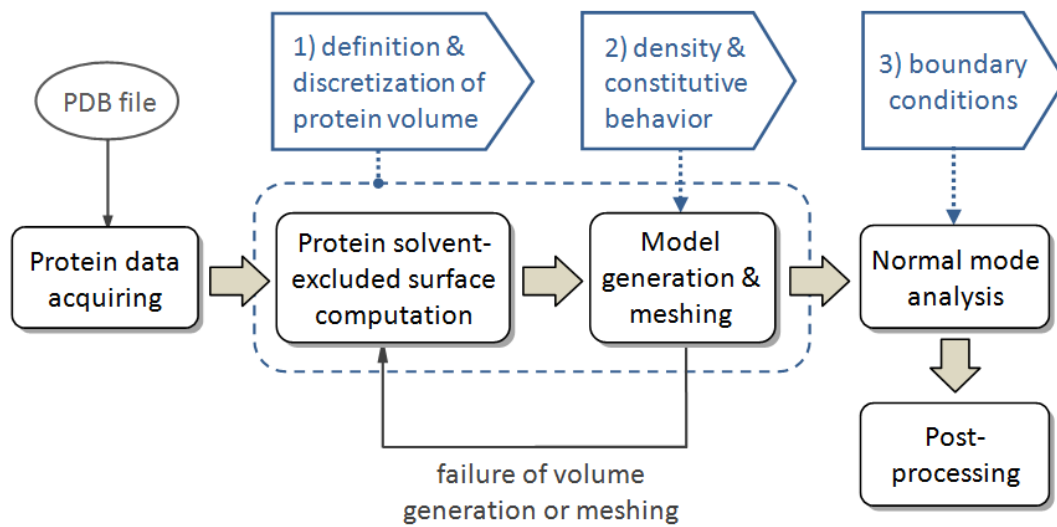
### 3.1 Protein Data Acquiring

This study targets six proteins, as shown in Table 3.1 in ascending order of size, with their two end conformational PDB codes.

**Table 3.1.** List of targeted proteins; the two pdb codes of each protein represent its “open” and “closed” conformations respectively.

Protein Name	Residue Numbers	PDB Codes
HIV-1 protease	99	1HHP, 1AJX
Che Y protein	128	3CHY, 1CHN
LAO binding protein	238	2LAO, 1LST
Maltodextrin binding protein	370	1OMP, 1ANF
Enolase	436	3ENL, 7ENL
Lactoferrin	691	1LFH, 1LFG

Each PDB code stands for a specific conformation of a protein, which is held in a Protein Data Bank (.pdb) file. The pdb file format is a textual file format describing the three dimensional structures of molecules held in Protein Data Bank



**Figure 3.1.** A schematic illustration of the finite element analysis procedure. The aforementioned three steps are rather conceptual and are presented by blue arrow text boxes. The rectangle boxes represent the actual course for accomplishing the analysis procedure. A detailed explanation of each step is presented in the following sections.

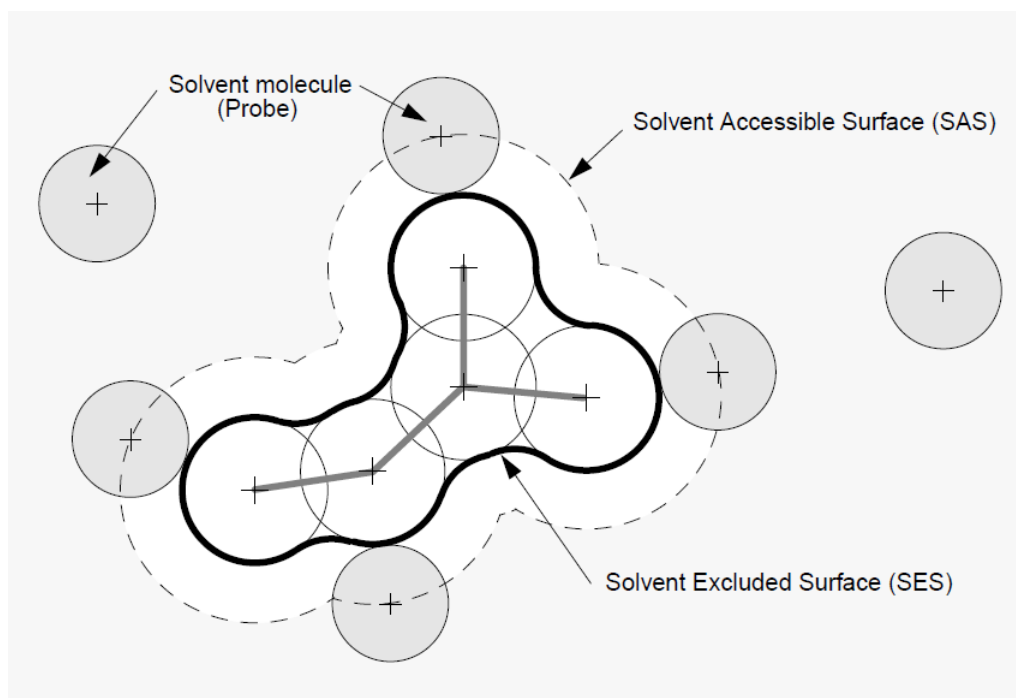
[11]. The pdb format contains description and annotation of protein and nucleic acid structures including atomic coordinates, observed sidechain rotamers, secondary structure assignments, as well as atomic connectivity.

The two PDB codes of each protein corresponds to its “open” (left) and “closed” (right) conformation structures respectively . As pointed out by F. Tama et al. [69], normal mode analysis performs better with open conformations which usually have a less compact shape with dispersed domains. Therefore, only “open” conformers are implemented in the normal mode analysis process within this study. The “closed” conformers are studied only in post-processing, along with “open” conformers, as experimental data.

A number of Molecular viewers are now available online for users to visualize protein structures and perform further computations with pdb files. The Python Molecular Viewer (PMV) 1.5.4 is preferably adopted in this work regarding its extensive computational package, including the MSMS mentioned before. Application details of this software are provided in the following section.

### **3.2 Protein Solvent-excluded Surface Computation**

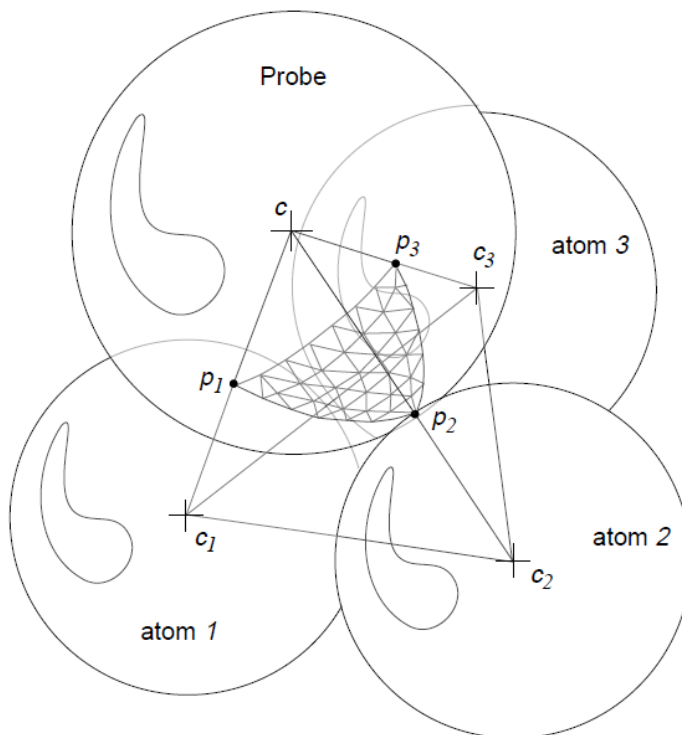
Given structural data from Protein Data Bank, the protein model is ready for generation. As mentioned, a protein’s volume is defined by its bounding solvent-excluded surface. The solvent-excluded surface, also known as the molecular surface, is defined by the closest point of contact of a solvent-sized probe-sphere that is rolled over the van der Waals surface of the protein, which defines the molecular volume that is never penetrated by any part of the solvent probe-sphere. This solvent probe-sphere represents the solvent molecule, which is usually set as a water molecule with a radius of 1.5 Å. (Figure 3.2)



**Figure 3.2.** Michel F. Sanner et al., 1996 [61]; 2D illustration of the concept of solvent-excluded surface (SES); the SES, shown by black solid line, is defined by rolling a solvent-sized probe-sphere over the van der Waals surface of the protein.

### 3.2.1 MSMS/Fine Molecular Surface

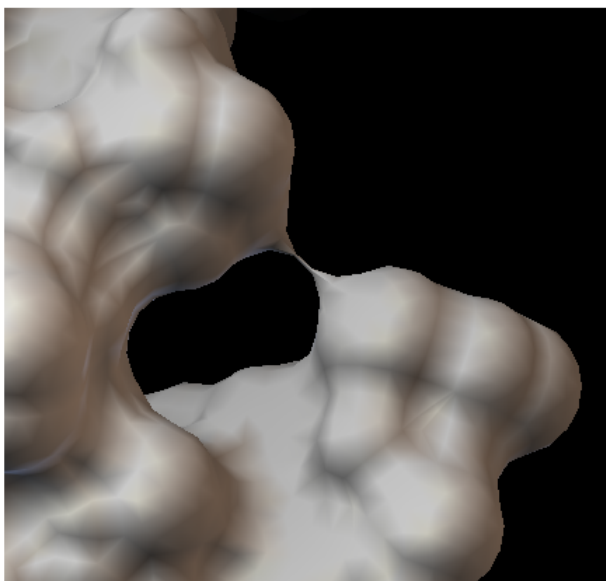
As mentioned, the MSMS is adopted here to compute protein solvent-excluded surface. Redundant molecules such as water, ions etc., are removed prior to protein surface calculation. The MSMS generates a high density triangulated approximation (one triangular vertex per  $\text{\AA}^2$ ) to the exact solvent-excluded surface. (Figure 3.3)



**Figure 3.3.** Michel F. Sanner et al., 1996 [61]; 3D illustration of the solvent-excluded surface probe-sphere; an adequate probe-sphere radius is usually set as  $1.5 \text{\AA}$ , representing the size of a water molecule.

It is noted that the radius of the probe-sphere should not be too small nor too large. Too small of a radius will result in surface points that are too dense, while a radius that is too large will cause separate parts to connect (Figure 3.4). An adequate

probe-sphere radius is usually set as 1.5 Å. Therefore, a default radius (i.e. 1.5 Å) is used here for all proteins except HIV-1 protease, and a smaller radius (i.e. 1.0 Å) is used for HIV-1 protease to avoid the connection in Figure 3.4.



**Figure 3.4.** Too large of a probe-sphere radius will cause separate parts to connect. The demonstrated protein is HIV-1 protease. (Image rendered by PMV 1.5.4)

### 3.2.2 Coarse Molecular Surface

In comparison to the MSMS computation, a more efficient way to generate the protein volume is through computing a coarser molecular surface which can also be done in PMV. This calculation is done by using the Gaussian-based blurring technique.[62] In this computation process, selected atoms are first blurred as gaussians into a grid. The grid is then isocontoured at a user specific value and an indexed polygon geometry is added to the viewer. Similar to MSMS computation, the Gaussian blurring approach yields surface approximation comprising of triangulated areas. The result-

ing surface is a decimated version of the original surface and the number of faces are greatly reduced by user definition.

### 3.2.3 Surface Data Processing

The molecular surfaces derived from the above methods are saved as files comprising of information on vertices and triangles, and used for volume generation. However, these surfaces require improvement because they often contain duplicated vertices or residual areas due to triangle degeneration resulting from computation. This will lead to failure of volume generation in the following process. In some cases, the problem is solved easily by removing all duplicated points prior to generating the volume. On the other hand, some cases need a more involved process to acquire clean surface models. Figure 3.6 shows the concept of finding a closed molecular surface.

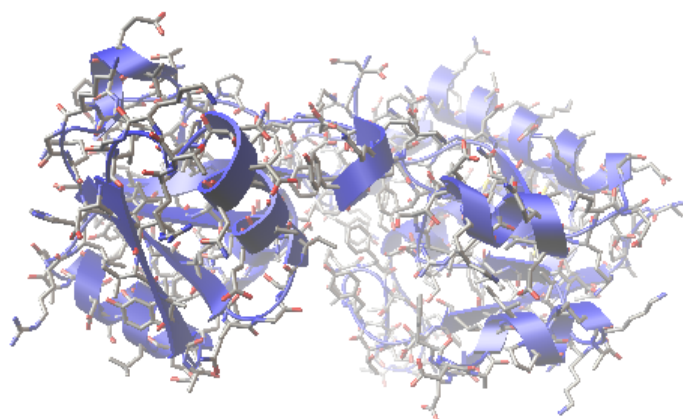
## 3.3 Model Generation and Meshing

### 3.3.1 Mesh elements

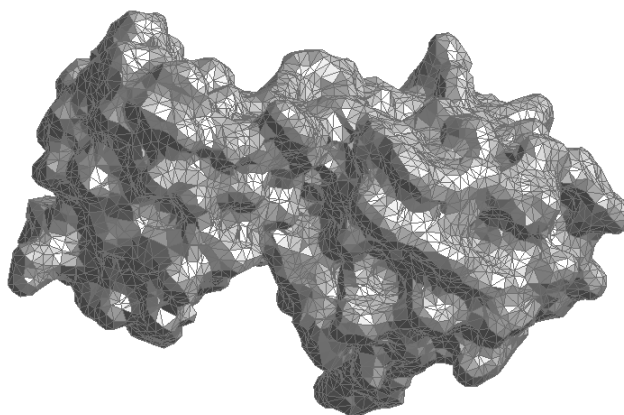
Once a surface model is generated, it can be imported into the finite element software ANSYS to define a protein volume. The protein volume that is bounded by the closed solvent-excluded surface is subsequently discretized with three-dimensional tetrahedral elements via automatic mesh generation using built-in mesher in ANSYS.

The mesh element used here for finite element analysis is SOLID 185 (Figure 3.7) [4], which is a 8-node structural solid in ANSYS element library. This three-dimensional element has three degrees of freedom at each node: translations in the nodal x, y, and z directions. It allows for prism and tetrahedral degenerations when used in irregular regions which is suitable for protein models regarding their irregular shapes. Additionally, in the case of a protein model, the adoption of this element results in better efficiency in comparison with the more commonly used tetrahedral element SOLID 187. In contradiction to SOLID 187, the SOLID 185 element does not

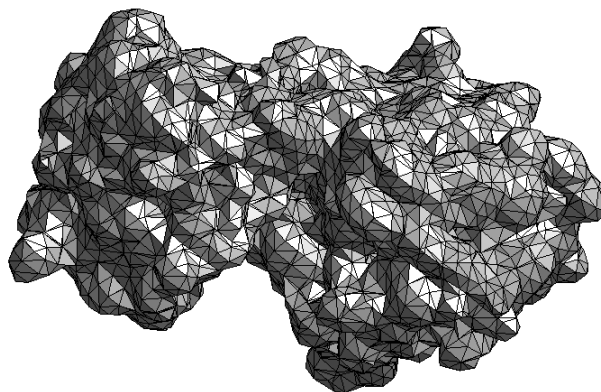




(a) LAO binding protein 3D structure.

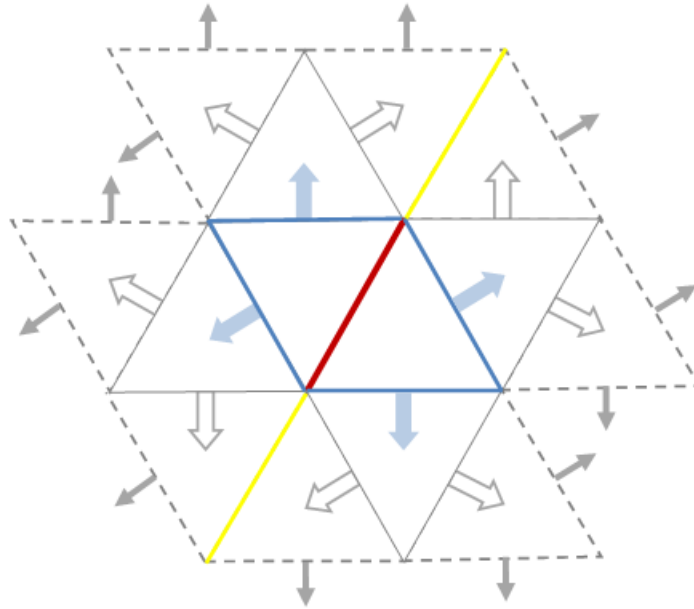


(b) MSMS molecular surface



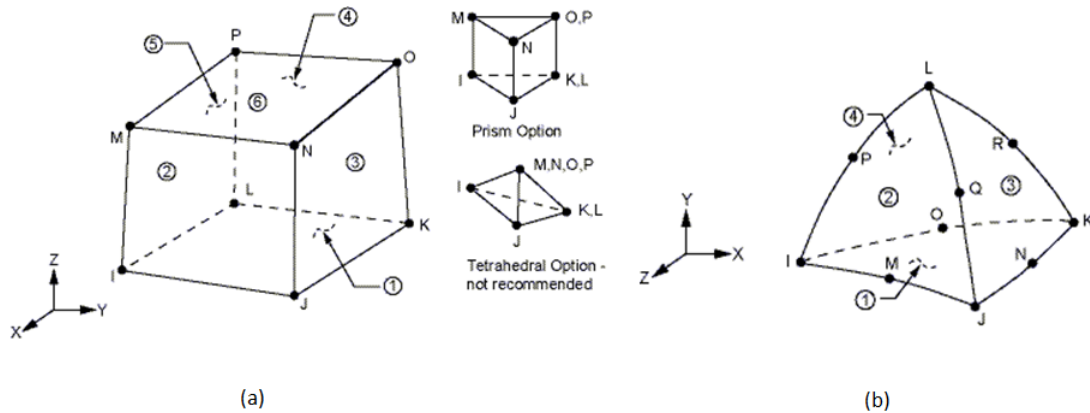
(c) Coarsen molecular surface

**Figure 3.5.** (a) 3D structure of the LAO binding protein; (b) a triangular approximation of the molecular surface; (c) a coarser molecular surface is calculated by using the Gaussian-based blurring technique.



**Figure 3.6.** Illustration of finding a closed surface; the main concept is to extend the surface in a hierarchical manner. On a closed surface, every edge/side is shared by two areas. Therefore, given randomly a known side (i.e. the red line), there are two adjacent triangles found (i.e. triangles with blue lines). To each of these known triangles, there are also two other triangles adjacent to it (i.e. triangles with gray solid lines). Accordingly, the search process will continue until no more new triangles are found (e.g. yellow lines indicate a halt when two known triangles meet).

contain mid-side nodes which significantly reduce the time for meshing and solving when the model is complex or large.



**Figure 3.7.** Illustration of the mesh element SOLID 185.[4] This eight-node element allows for prism and tetrahedral degenerations when used in irregular regions which is suitable for protein models regarding their irregular shapes.

### 3.3.2 Density and Constitutive Behavior

#### 3.3.2.1 Density

The mass density of the protein is taken to be homogeneous. A simple molecular-weight-depending function proposed by Fischer et al.[25] is provided below for accurate estimation of the average protein density:

$$\rho(M) = \left[ 1.410 + 0.145 \cdot \exp \left( -\frac{M(kDa)}{13} \right) \right] g/cm^3$$

, where  $\rho$  is the mass density of a protein and  $M$  is the weight of a protein measured in  $kDa$ . Given protein weight information from Protein Data Bank, density of each protein is derived and listed in Table 3.2.

**Table 3.2.** List of protein weight and mass density.

Protein	Weight (kDa)	density ( $g/cm^3$ )
HIV-1 protease	10.80	1.4739
Che Y protein	13.97	1.4602
LAO binding protein	26.03	1.4302
Maltodextrin binding protein	40.71	1.4169
Enolase	46.63	1.4146
Lactoferrin	76.14	1.4110

### 3.3.2.2 Young's Modulus

The effective Young's modulus is generally unknown for proteins, however, it is determined sometimes by matching protein stretching stiffness to an experiment, as has been performed by Tirion et al.[10, 43] In other cases, Young's modulus can be obtained by fitting thermal fluctuations of  $\alpha$ -carbon atoms in the FE model to those obtained using either the all-atom NMA or the RTB procedure, which generally ranges from 2 to 5 GPa.[9, 64] Herein, all cases are computed using a Young's modulus of 2 GPa, representing an approximate lower bound on protein stiffness.[9] Nonetheless, since the precise value of the Young's modulus affects the magnitude of thermal fluctuations linearly, the precision of Young's modulus will not affect much in the FE model. Furthermore, the results from the EN model are actually "scaled" values due to assumption of unity stiffness. After all, it is the "relative" fluctuation that is concerning therefore all results from both models will be scaled for comparison in post-processing.

### 3.3.2.3 Poisson's Ratio

Studies have indicated that it is most appropriate to model the protein interior as crystalline.[3] Therefore, the Poisson's ratio is said to be 0.3, which is typical for crystalline solids, for all cases in this study. In addition, Bathe et al.[9] have found that the precise value of Poisson's ratio does not affect the computation results within

the range of 0.3~0.5. The material compressibility does not play an important role in normal mode analysis.

### 3.4 Normal Mode Analysis

Similar to the coarse-grained elastic network model, the initial structure of the finite element (FE) model extracted from the pdb file is assumed to be the ground-state structure. In other words, the FE model is presumed to be at its minimal energy state, and no energy minimization is performed prior to the FE-based NMA.

As pointed out previously, all cases here are solved as free vibration problems in the absence of any boundary condition. The solution method used here is the PCG Lanczos method.[32] The PCG Lanczos method internally uses the Lanczos algorithm [56], combined with the PCG iterative solver. This method will be significantly faster when working with large models that are dominated by three-dimensional solid elements. Moreover, it works well when only a few of the lowest modes are requested. These features fit the method to protein model since only low-frequency/dominant modes are interested in protein normal mode analysis.

### 3.5 Post-Processing

In post-processing, normal mode solutions of  $\alpha$ -carbons are obtained by mapping the original atomic coordinates onto the FE model; modal displacements (i.e. eigenvectors) and frequencies (i.e. eigenvalues) of every  $\alpha$ -carbon are retrieved at its corresponding nodal degrees of freedom. Given normal mode solutions, the following quantities of interest are commonly adopted for analysing the results and providing better understanding. In order to assess the applicability of the FE model, the aforementioned quantities will be discussed along with those from the EN model.

### 3.5.1 Relative Displacement

For the sake of comparison, modal displacements of  $\alpha$ -carbons, from both FE model and EN model, will be scaled to proper magnitude. Usually, each  $\alpha$ -carbon is represented by its corresponding residue index/number. This will be seen in figures from the latter chapter.

### 3.5.2 Overlap and Cumulative Overlap

*Overlap*,  $I_k$ , is a measure of the similarity between the direction of the conformational change and the one given by normal mode  $k$  [46]:

$$I_k = \frac{\Delta \mathbf{q}_{AB} \cdot \mathbf{u}_k}{\Delta \mathbf{q}_{AB}} \quad (3.1)$$

, where  $\mathbf{u}_k$  is the normalized directional vector, and  $\Delta \mathbf{q}_{AB} = \mathbf{q}_B - \mathbf{q}_A$  represents the actual displacements between two conformations. A value of unity for the overlap means that the direction given by the normal mode  $k$  is identical to  $\Delta \mathbf{q}_{AB}$ . Moreover, the potential contribution of every normal mode to the transition may be deduced from the *cumulative overlap*,  $Q_k$  [46]:

$$Q_k = \left[ \sum_k I_k^2 \right]^{1/2} \quad (3.2)$$

, where the summation is performed over the subset of modes of interest. Note that the summation is identically equal to unity if it is performed over all  $3N - 6$  modes/eigenvectors, which form a complete orthonormal basis set for the  $3N - 6$  dimensional space of conformational changes.

### 3.5.3 Degree of Collectivity

The *degree of collectivity*,  $\kappa_k$ , is a measure which implies how collective a concerted motion (e.g. conformational change) is [13]:

$$\kappa_k = \frac{1}{N} \exp \left[ - \sum_{i=1}^N \alpha (\Delta \mathbf{R}_i)^2 \Big|_k \log(\alpha \Delta \mathbf{R}_i)^2 \Big|_k \right] \quad (3.3)$$

,  $\Delta \mathbf{R}_i$  is the displacement/fluctuation of residue  $i$ , and  $\alpha$  is a normalized constant chosen so that  $\sum_i (\Delta \mathbf{R}_i)^2 \Big|_k = 1$ . This quantity reflects the number of atoms that are affected during the conformational change. In addition, it has entropic significance that the mode with the highest degree of collectivity has the highest entropy. In other words, the mode with higher  $\kappa$  value is distributed over a larger number of residues rather than being orderly confined to a few residues. It is of interest to identify the most collective modes since they are intrinsically favored by the functional movements.

### 3.5.4 B-factor

The *B-factor* or *Boltzmann factor* (also known as *temperature-factor* or *Debye-Waller factor*) of protein crystal structures reflects the fluctuation of atoms above their average positions. The B-factor is given by [49]:

$$\mathbf{B}_i = \frac{8\pi^2}{3} \langle (q_i - q_i^0)^2 \rangle \quad (3.4)$$

, where  $q_i^0$  is the coordinates of residue  $i$  at its native/equilibrium conformation. Comparisons of B-factors derived from the normal mode results and experimentally measured B-factors give an indication on differences in protein flexibility between the free protein and the protein in a crystallographic environment.

## CHAPTER 4

### RESULTS AND DISCUSSION

Herein, the finite element based normal mode analysis results are discussed mainly in three aspects: (1) *relative displacement* of low-frequency modes; (2) *overlaps* of the first twenty modes with respect to experimental data; and (3) quality of the *cumulative overlaps*. As shown in Table 4.1, the first three proteins, sorted in ascending order, are analyzed with both fine (MSMS) and coarse molecular surface for comparison. The remaining three proteins are analyzed with coarse surface solely for the sake of efficiency. Results from the two surface models are presented separately followed by an overall discussion.

**Table 4.1.** List of targeted proteins with different surface models; figures in parentheses are the numbers of residues.

	Targeted Protein
Fine Molecular Surface (MSMS)	HIV-1 protease (99), Che Y protein (128), LAO binding protein (238)
Coarse Molecular Surface	HIV-1 protease (99), Che Y protein (128), LAO binding protein (238), Maltodextrin binding protein (370), Enolase (436), Lactoferrin (691)

#### 4.1 MSMS/Fine Molecular Surface

##### 4.1.1 HIV-1 Protease

As indicated in the previous chapter, the finite element (FE) model is defined directly from the atomic structure without initial energy minimization. That is,

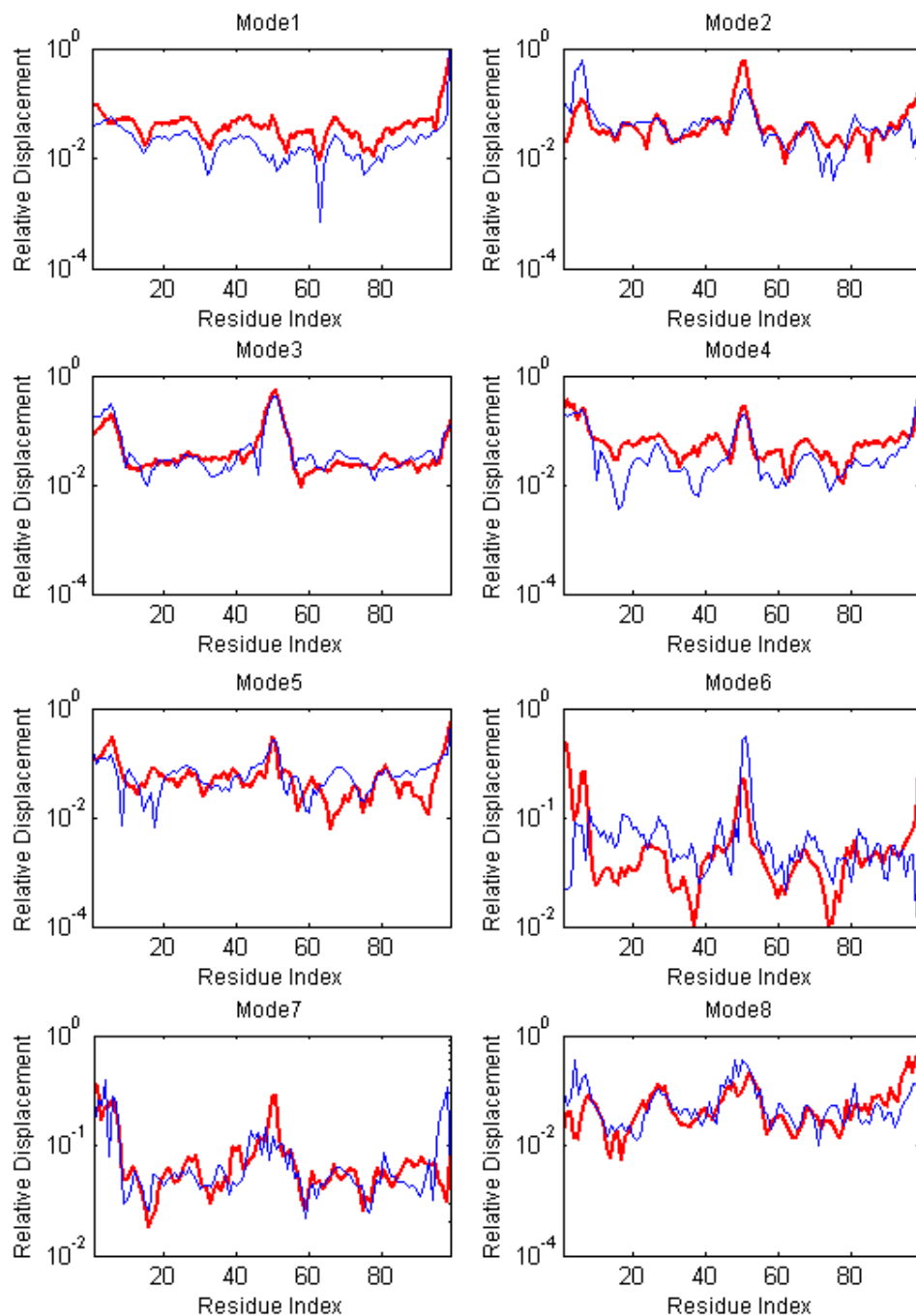


the atomic structure extracted from HIV-1 protease PDB file is presumed energy-minimized, which is consistent with the elastic network (EN) model.

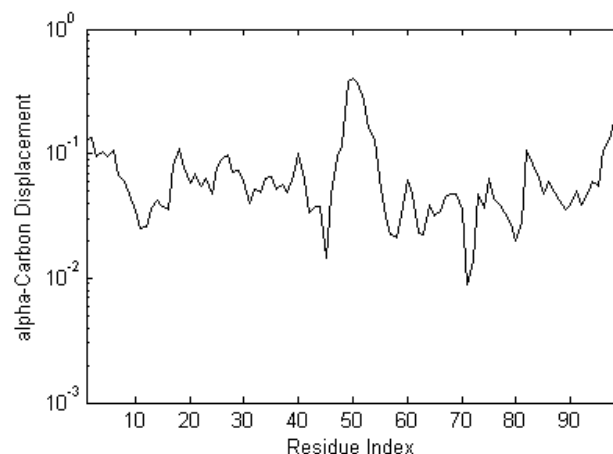
In post-processing,  $\alpha$ -carbon coordinates are mapped onto the FE model to retrieve corresponding relative displacements of each  $\alpha$ -carbon/residue; displacements of residue are usually represented by displacements of  $\alpha$ -carbons in convention. Figure 4.1 is the comparison of relative displacements of the first eight modes between the FE and the EN models. Correspondence between the FE model and the EN model is quite satisfying regarding two end sections (residue 1 to 15 and 85 to 99) and the middle section (residue 45 to 60) of the protein chain. Furthermore, the forementioned three sections all have high relative displacements except in the first mode. This feature can be self-explaining while we look at the residue displacements extracted from two end conformations (Figure 4.2). Indeed, the middle section and two end sections of the protein chain undergo relatively high positional changes. This similarity of relative displacement corresponds well to the theory that low frequency modes are dominant modes, which hold significant character of the conformational change. In this case, the FE model works as well as the EN model for capturing the dominant modes movements.

Figure 4.3 shows scaled B-factors from the FE results and experimental data. Data from the FE model differs greatly from the experiment since those B-factors are derived from only twenty modes. However, the B-factors successfully reveal flexibility of residues who play predominant roles in conformational change; both the B-factors from the FE model (Figure 4.3) and  $\alpha$ -carbon displacements (Figure 4.2) have peaks around three crucial sections.

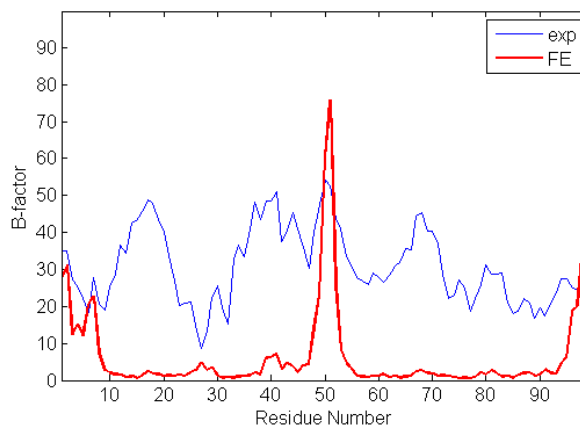
Furthermore, Figure 4.4 provides a rather vivid demonstration of this feature. Residues in the three forementioned regions, which perform large displacement during conformational change, are colored in blue, red, and green respectively. These regions are located exactly on the protruded parts of the protein, which implies that the FE



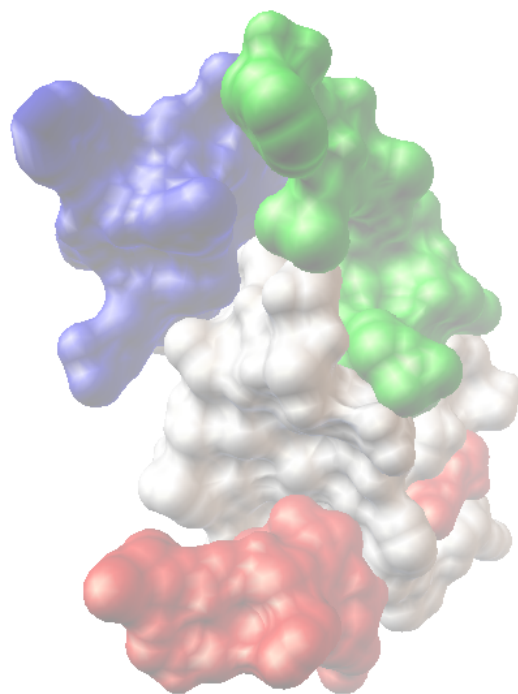
**Figure 4.1.** HIV-1 protease: Comparisons of the *relative displacements* for the lowest eight normal modes. (FE model - thick red line; EN model - blue line) The residue index represents the number of alpha-carbons. Correspondence between the FE model and the EN model is quite satisfying regarding two end sections (residue index 1 to 15 and 85 to 99) and the middle sections (residue index 45 to 60) of the charts. Furthermore, the forementioned three sections all have high relative displacements except in the first mode. Indeed, these three sections of the protein chain all undergo relatively high positional changes according to the experimental data.



**Figure 4.2.** HIV-1 protease:  $\alpha$ -carbon/residue displacements between two end conformations, i.e. "open" conformation to "closed" conformation. The middle section and two end sections of the protein chain undergo relatively high positional changes. This corresponds well to the results shown in Figure 4.1. These three sections will later be proven significant regarding the property of the protein shape. (Figure 4.4)



**Figure 4.3.** HIV-1 protease: B-factors from FE results and experimental data. (FE model - thick red line; experiment - blue line) Data from the FE model differs greatly from the experiment since those B-factors are derived from only twenty modes. However, the B-factors successfully reveal flexibility of residues who play predominant roles in conformational change; both the B-factors from the FE model and  $\alpha$ -carbon displacements (Figure 4.2) have peaks around three crucial sections.



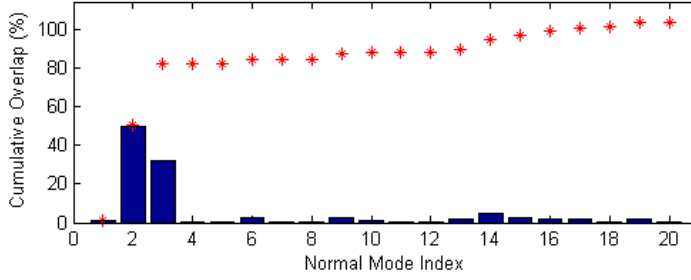
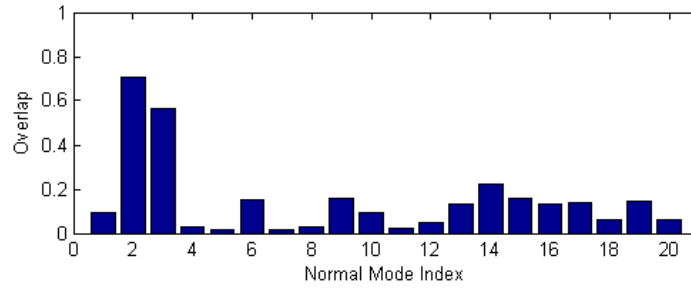
**Figure 4.4.** HIV-1 protease 3D surface model: Residue 1 to 10 (blue) , 45 to 60(red) and 85 to 99(green). These regions located exactly on the protruded parts of the protein which imply that FE model performs well by capturing the property of the molecular shape.(Image rendered with PMV 1.5.4)

model performs well by capturing the property of the molecular shape. Again, this confirms the importance of protein shape to its conformational change.

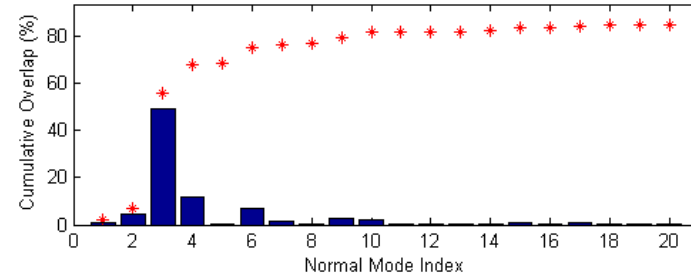
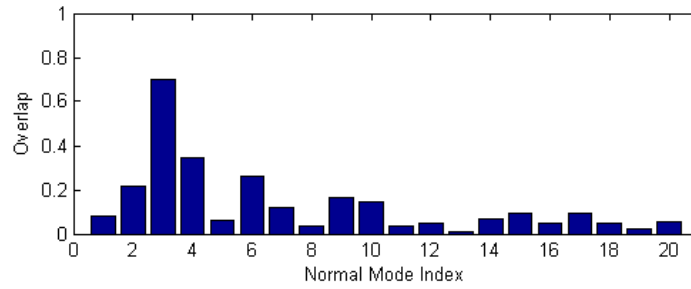
Overlaps and cumulative overlaps of the first twenty modes, with respect to experimental data, are shown in the following bar charts (Figure 4.5)<sup>1</sup>. Although two bar charts of overlaps seem quite different at first glance, they both acquired high overlaps within the first few modes. Furthermore, cumulative overlaps in both charts reached a plateau at approximately the fifth mode. The maximum overlaps of the FE model and the EN model occurred at the second and the third mode, respectively. Figure 4.6 shows the comparison of residual displacements between the FE model and the EN model, regarding normal modes that hold the maximum overlaps. Predictably, two sets of data are similar even though they are retrieved from different modes. In addition, Figure 4.6(b) shows the comparison of residual displacements between the FE model and experimental data which also yields good correspondency. According Tama et al., it is often found that a single normal mode can carry a lot of information.[69] Moreover, this single mode is very often found within the first four modes. At this point, the second mode from the FE model is not only one of the dominant modes but also that *very single mode* bearing most of the conformational change.

---

<sup>1</sup>Before reviewing the charts, it is noted that the value of cumulative overlap should never exceed 100% (Figure 4.5(a) bottom). The summation of overlaps over  $3N - 6$  modes should be unity. The reason that the cumulative overlap of HIV-protease went beyond reasonable value can be attributed to the underlying distinction between FE model and EN model. In EN model, the number of concerned components is determined by the number of residues, which means a molecule with  $N$  residues (i.e.  $N$  components) yields in  $3N - 6$  degrees of freedom and  $3N - 6$  normal modes. These numbers are predetermined and fixed for a molecule. On the contrary, the number of components/nodes in a FE model is determined by mesh since the molecule is deemed as a homogenous solid. For instance, the FE model of HIV-1 Protease has 11,951 nodes, thus, there should be  $3 * 11,951 - 6 = 35,847$  normal modes. However, herein, only the 99 residues are considered in the post-processing step. In other words, a 11,951-node model is now replaced by a 99-node model, although these 99 nodes are crucial and representative. Therefore, a large portion of the overlap value from all nodes is missed and replaced with that from only 99-nodes due to simplification, and result in "seemingly unreasonable" values of cumulative overlap. However, considering all nodal results in the FE model is undoubtedly inefficient and unnecessary. As long as one bear in mind this very distinction of FE

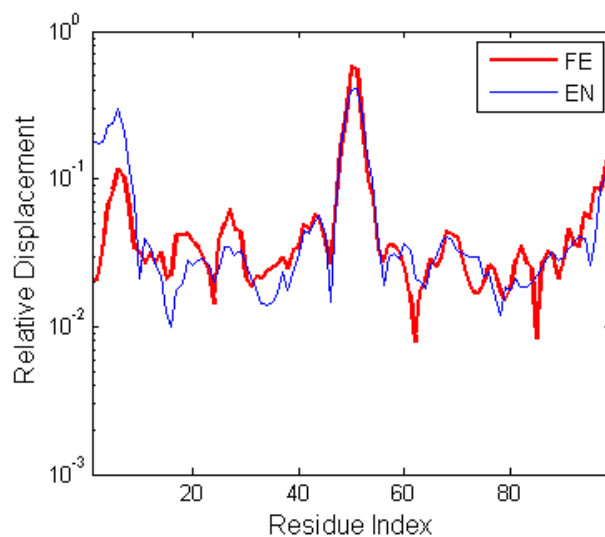


(a) FE model : *overlap* and *cumulative overlap*

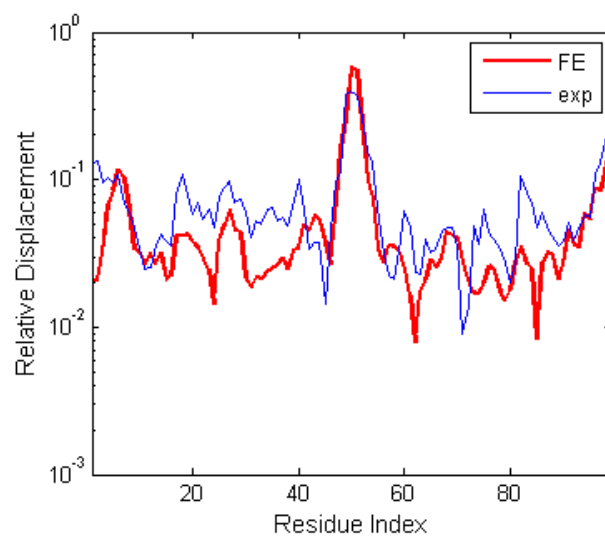


(b) EN model : *overlap* and *cumulative overlap*

**Figure 4.5.** HIV-1 protease: Values of *overlap* and *cumulative overlap* from (a) the FE model and (b) the EN mode. Although two bar charts of overlaps seem quite different at first glance, they both acquired high overlaps within the first few modes. Cumulative overlaps in both charts reached a plateau at approximately the fifth mode. The maximum overlaps of the FE model and the EN model occurred at the second and the third mode, respectively.



(a) FE mode 2 and EN mode 3



(b) FE mode 2 and Experimental Data

**Figure 4.6.** Comparison of normal modes that hold the maximum overlaps (i.e. dominant modes) : (a) The FE model matches well with the EN model. Two sets of data are similar even though they are retrieved from different modes. (b) Data from the FE model and experiments also yield good correspondency. According Tama et al., [69] the second mode from the FE model might not only be one of the dominant modes but also be that *very single mode* bearing most of the conformational change.

### 4.1.2 Che Y Protein

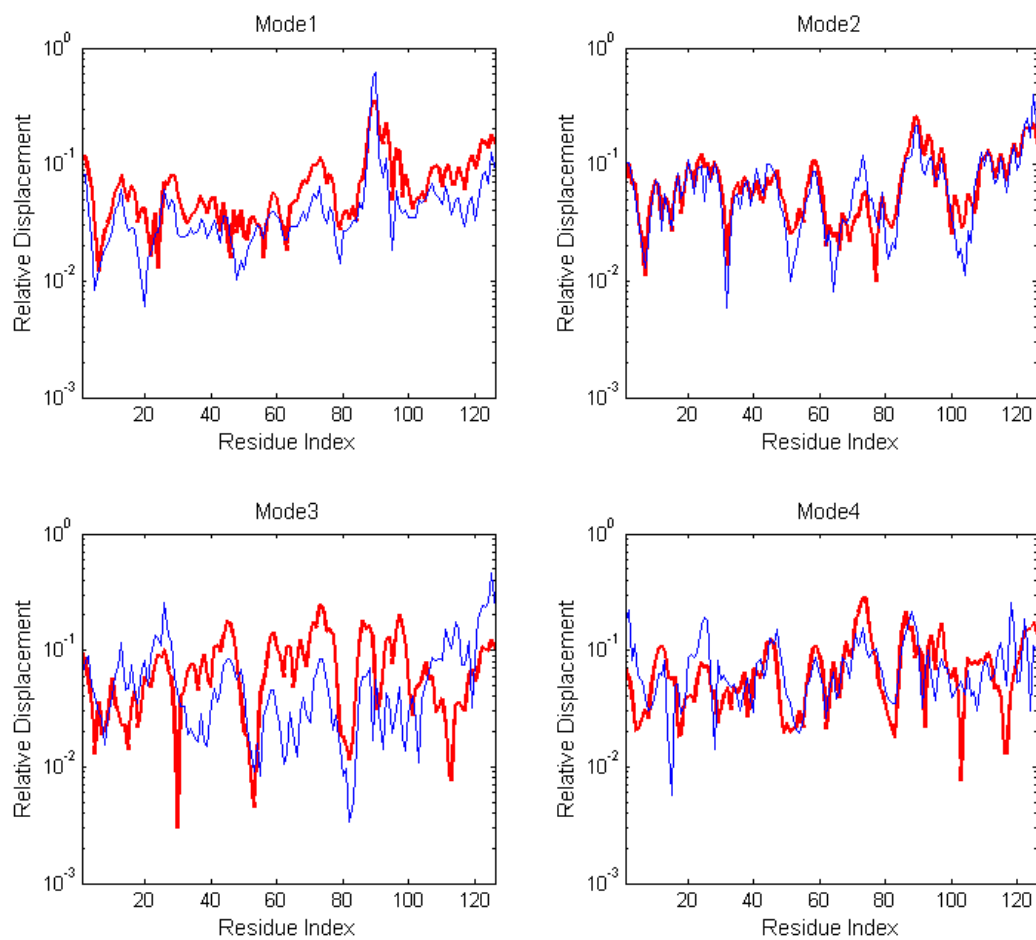
Likewise, the FE model of Che Y protein is defined directly without initial energy minimization. Relative displacement comparisons of the first four modes between the finite element (FE) and the elastic network (EN) models yield great correspondence (Figure 4.7). Moreover, both models have their first two maximum overlaps at the same modes, i.e. mode 1 and mode 6. (Figure 4.8) In other words, they have comparable ability in predicting dominant modes. However, modal overlaps between the FE results and experimental data were not correctly predicted. The maximum overlap occurred at the first mode only yielding a value of 0.412. This also happened to the EN model which yielded a value of 0.415. Cumulative overlaps from both models share a poor ascending rate.

Notwithstanding, comparison of B-factors between the FE results and experimental data has similarities in the first 20 modes. (Figure 4.9) To this end, the FE results have indeed made good predictions of the overall flexibility of the protein. In fact, the failure of poor overlaps has a reasonable explanation stemming from the shape of the protein. In comparing to other cases in this study, Che Y protein has a fairly globular shape which results in a more complex conformational change. (Figure 4.10) The movement of the protein might be rather localized, and the moving direction of residues might change constantly during conformational change. [69] While the FE model predicts the instantaneous movement at each normal mode, experimental data (i.e. displacements between two end conformations) reveal only movement concerning the two end states. Therefore, even if the FE model predicts normal modes with proper movements, it fails to obtain overlaps with expected values.

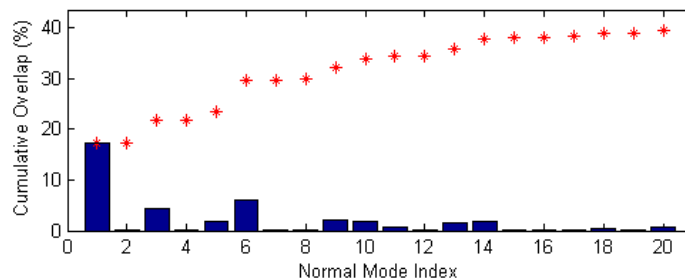
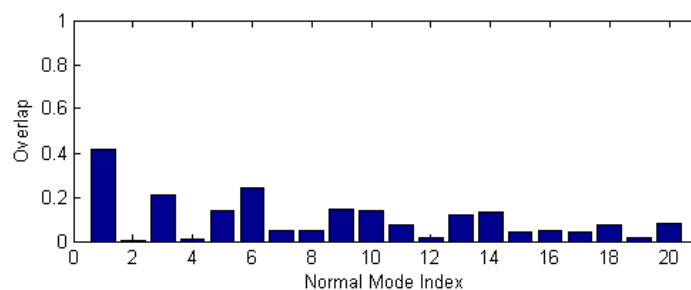
---

model, the cumulative overlap analysis can still stand its referential importance. In this study, it is the rate of cumulation that is concerned rather than the cumulation value itself.

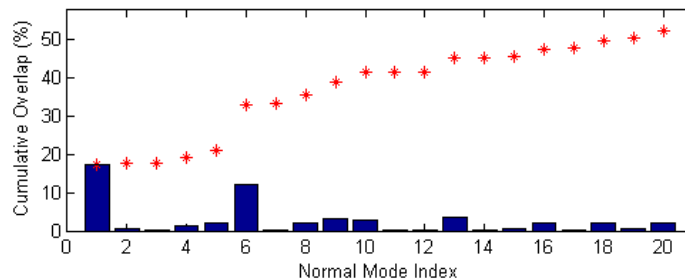
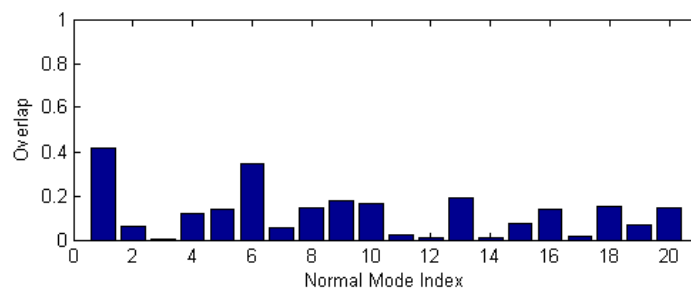




**Figure 4.7.** Che Y protein: Comparisons of *relative displacement* of the first four normal modes. (FE model - thick red line; EN model - blue line) In general, most portion of the data yield great correspondency.

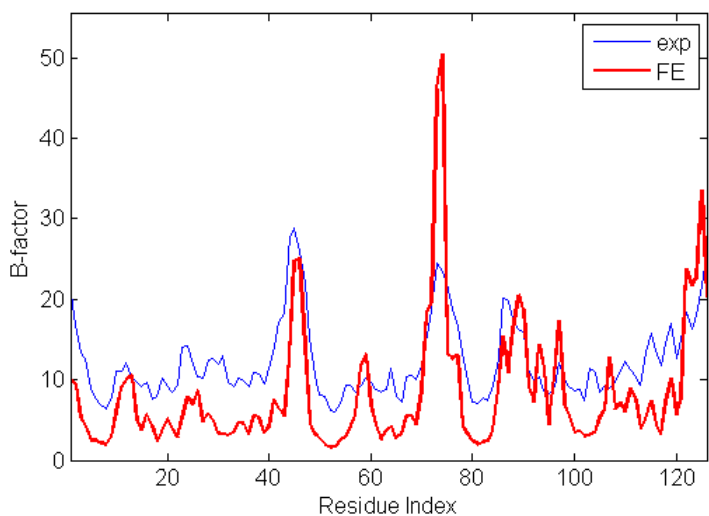


(a) FE model : *overlap* and *cumulative overlap*



(b) EN model : *overlap* and *cumulative overlap*

**Figure 4.8.** Che Y protein: Values of *overlap* and *cumulative overlap* from (a) the FE model and (b) the EN mode. Both models have their first two maximum overlaps at the same modes – mode 1 and mode 6. They have comparable ability in predicting dominant modes. However, modal overlaps between the FE results and experimental data were not well predicted. The maximum overlap occurred at the first mode only yielding a value of 0.412. This also happened to the EN model which yielded a value of 0.415. Cumulative overlaps from both models share a poor ascending rate.

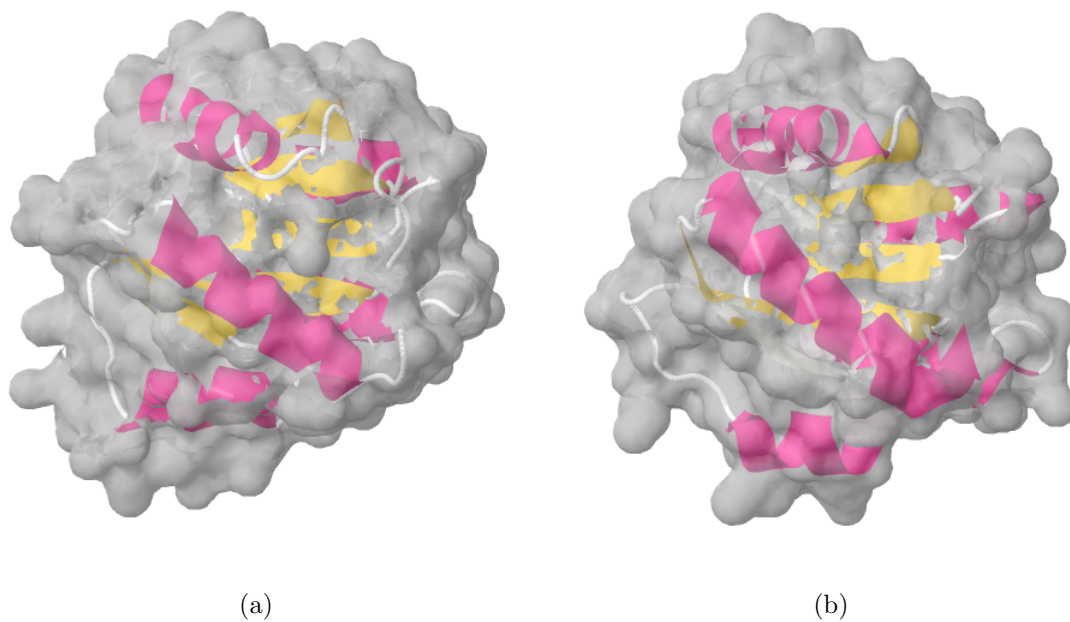


**Figure 4.9.** Che Y protein: Comparison of B-factors between the FE results and experimental data has similarities in the first 20 modes. The FE results have made good predictions of the overall flexibility of the protein.

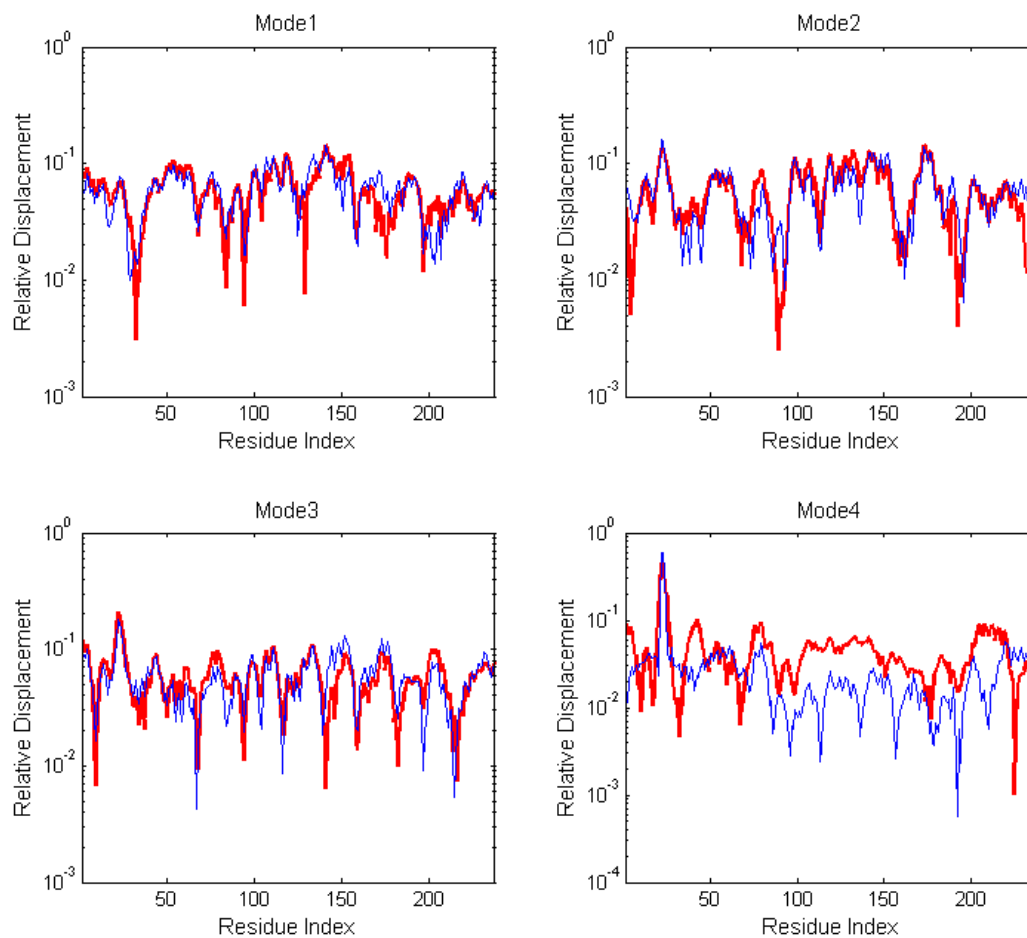
### 4.1.3 LAO Binding Protein

Similar to previous cases, initial energy minimization is omitted for defining the finite element (FE) model of the LAO binding protein. Figure 4.11 shows relative displacement comparisons of the first four modes between the FE and the elastic network (EN) models. The two models yield great correspondency especially in the first three modes. Encouragingly, both models have very high overlaps at the first normal modes (Figure 4.12), and cumulative overlaps remain almost static after the first mode.

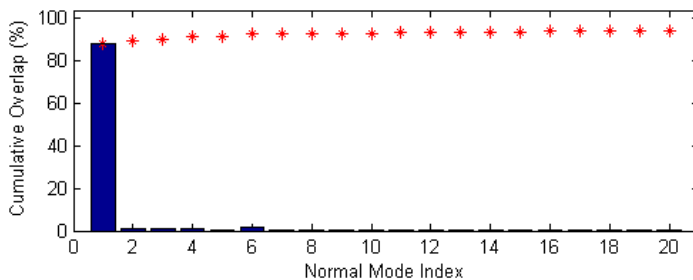
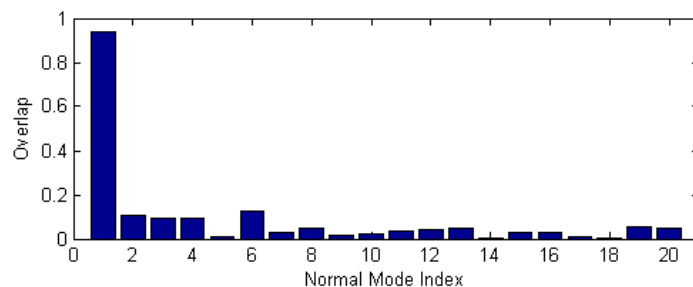
In this case, the first mode is the best fit to the very single mode which bearing most of the conformational change. Again, the shape of the protein, as shown in Figure 4.13, provides a decent explanation. The necking shape in the middle of the LAO binding protein might be the cause of a hinge-like motion during a conformational change. This collective motion involves a large amount of residues moving in almost



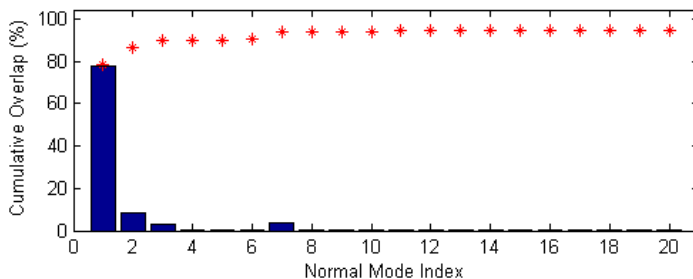
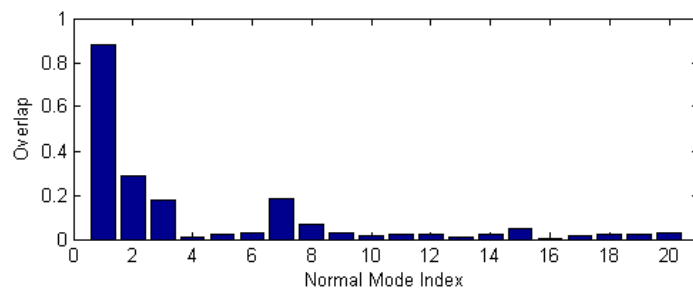
**Figure 4.10.** Che Y protein: Both of the two end conformations (a) and (b) (PDB code: 3CHY & 1CHN), have rather globular shape, and thus result in a more complex conformational change. The movement of the protein might be rather localized, and the moving direction of residues might change constantly during conformational change. (Image rendered with Jmol from <http://www.rcsb.org>)



**Figure 4.11.** LAO binding protein: Comparison of *relative displacement* of the first four normal modes. The two models yield great correspondency especially in the first three modes. (FE model - thick red line; EN model - blue line)



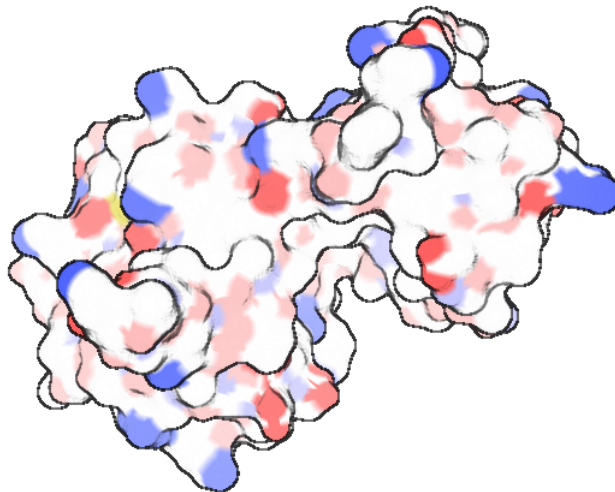
(a) FE model : *overlap* and *cumulative overlap*



(b) EN model : *overlap* and *cumulative overlap*

**Figure 4.12.** LAO binding protein: Values of *overlap* and *cumulative overlap* from (a) the FE model and (b) the EN mode. Both models have very high overlaps at the first normal modes and the cumulative overlaps remain almost static after the first mode. These indicate that the first mode is the dominant mode which bearing most of the conformational change.

the same direction. Consequently, the movement is quite straightforward and any change of direction is limited.



**Figure 4.13.** LAO binding protein: The necking shape in the middle of the protein might be the cause of a hinge-like motion during conformational change. (Image rendered with PMV 1.5.4)

Again, the degree of collectivity has its contribution to the normal mode results. The LAO binding protein has a high degree of collectivity, i.e. 0.68. This implies that the conformational change involves fairly collective movement. Likewise, detailed discussion about degree of collectivity will be mentioned latter.

Regarding the previous three cases based on MSMS/fine surface approximation, the FE model unfolds good potential in predicting protein normal modes. Although, results in the second case seem questionable at first glance, queries can be answered by scrutinizing the pattern of motion throughout conformational change.

## 4.2 Coarse Molecular Surface

In previous sections, data from the finite element (FE) model have been compared, in many aspects, to those from the elastic network (EN) model as well as experimental

data. Normal mode analysis based on the FE model shows satisfactory results. In this section, discussions will focus on contrasts and analogies between MSMS/fine molecular surface model and coarse molecular surface model. The same three proteins above are analyzed again by coarse molecular surface model.

Hereafter, for the sake of convenience, the MSMS/fine molecular surface model and the coarse molecular surface model are simply called the *fine model* and the *coarse model*, respectively.

## **4.2.1 HIV-1 Protease, Che Y Protein, LAO Binding Protein**

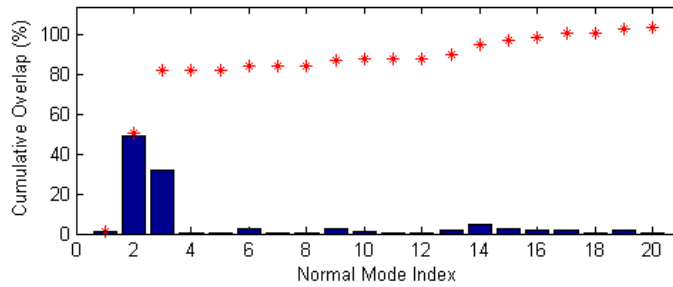
### **4.2.1.1 HIV-1 Protease**

Figure 4.14(a)&(b) show cumulative overlaps for both fine and coarse surface model. Both models capture dominant modes, i.e. modes with highest overlaps, within the first few normal modes. The maximum overlap of coarse models occurs at the fourth mode while for fine model the maximum occurs at the second mode. Relative displacement comparison between these two maximum overlaps is shown in Figure 4.14(c). In terms of collective motion, there are good analogies between these two lines; both lines reveal high relative displacements at two end sections (residue 1 to 15 and 85 to 99) and the middle section (residue 45 to 60) of the protein chain. Overall, the results prove good competence of the coarse model in comparison with the fine model.

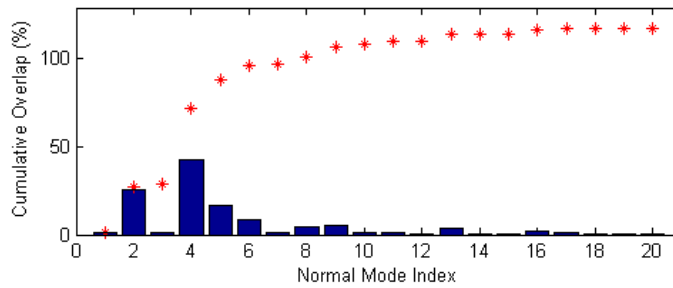
### **4.2.1.2 Che Y Protein**

As mentioned previously, Che Y protein has a pattern of motion in which directions change constantly. This results in poor motion prediction (i.e. low overlap values) for not only the fine model but also the coarse model.(Figure 4.15) Notwithstanding this condition, it is still worth observing the behavior of this two models for understanding the distinction of their capabilities. Cumulative overlaps of the first few modes from the coarse model fall slightly behind those from the fine model.

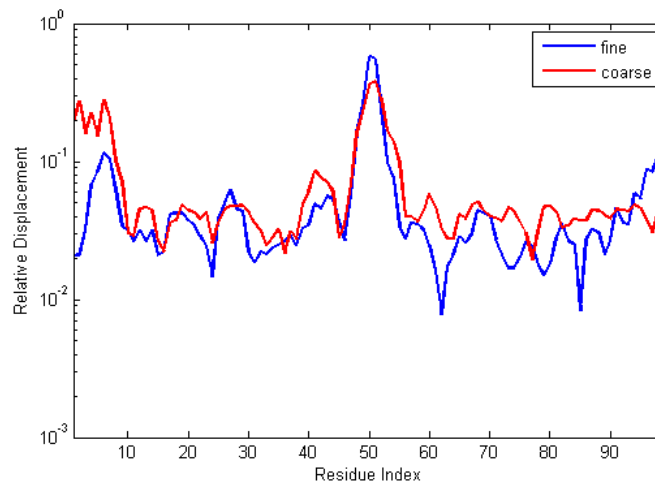




(a) Fine surface model : *cumulative overlap*



(b) Coarse surface model : *cumulative overlap*



(c) Relative displacement : FE – mode 2 and EN – mode 4

**Figure 4.14.** HIV-1 protease: Cumulative overlap from (a) fine surface model and (b) coarse surface model; (c) comparison of relative displacement between fine and coarse surface model at their most dominant modes. Both models capture dominant modes (i.e. modes with highest overlaps) within the first few normal modes. The maximum overlap of coarse models occurs at the fourth mode while for fine model the maximum occurs at the second mode.

However, in terms of overall performance in over 20 normal modes, the coarse model closely parallels the fine model. Both models yield the maximum overlaps at the first normal mode. Figure 4.15(c) is the comparison of relative displacement between the two maximum overlaps. The tight correspondence between two lines confirms again the capability of the coarse model.

#### **4.2.1.3 LAO Binding Protein**

Recalling from the previous section, the consistency of the motion of LAO binding protein has yielded excellent predictability (i.e. considerably high overlaps). Since the coarse model has shown its capability in previous cases, it is conceivable that in this case, the coarse model provides a satisfactory prediction as well. Indeed, the coarse model yields not only high overlap at the first mode but almost identical predictions to the fine model. (Figure 4.16)

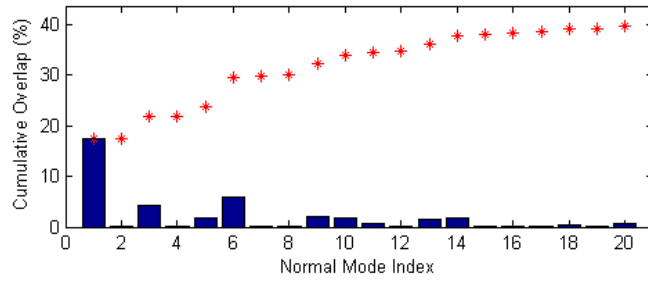
#### **4.2.2 Maltodextrin Binding Protein, Enolase, Lactoferrin**

In previous sections, results have confirmed the capability of the coarse model for conformational change prediction. For the sake of efficiency, hereafter, the rest of the sizable proteins will be analyzed solely by the coarse model.

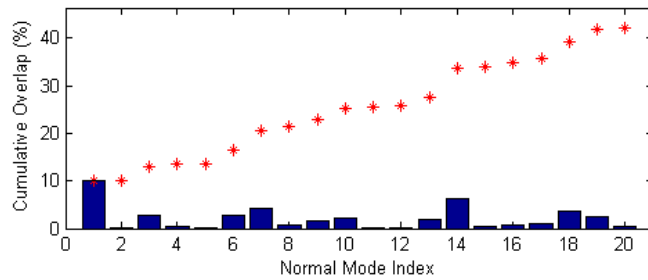
##### **4.2.2.1 Maltodextrin Binding Protein**

The pattern of results of maltodextrin binding protein has a great resemblance to those of the LAO binding protein. (Figure 4.16(a)(b) and Figure 4.17(a)(b)) Likewise, the maximum overlap happens at the first normal mode with a remarkably high value. Furthermore, relative displacement results in Figure 4.17(b) provides a prominent demonstration to this correspondency; the majority of the data at the first mode coinciding with the experimental data.

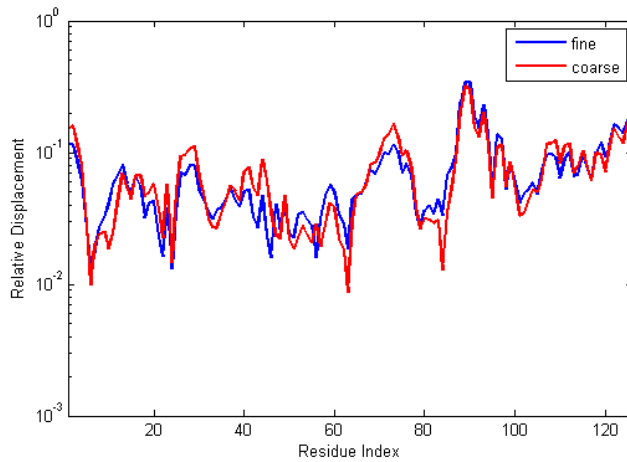
The first normal mode in this case can be deemed as that *very single mode* featuring the most in conformational change. Presumably, the shape of the protein plays



(a) Fine model : *cumulative overlap*

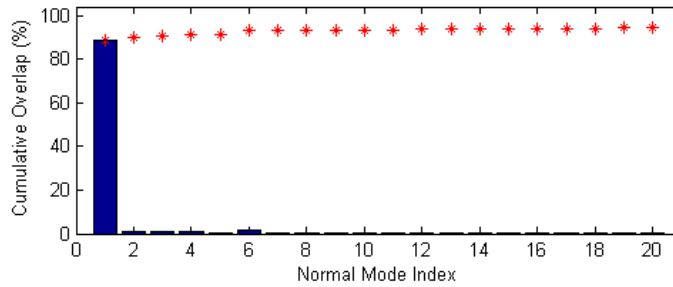


(b) Coarse model : *cumulative overlap*

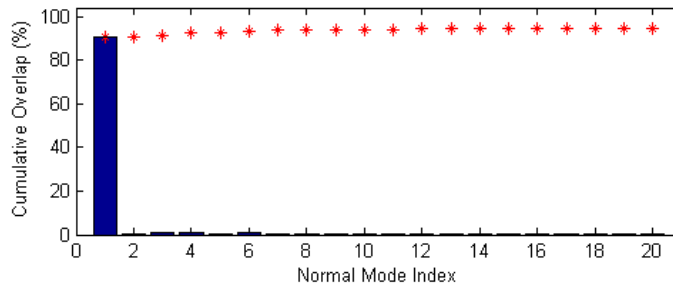


(c) Relative displacement : FE – mode 2 and EN – mode 4

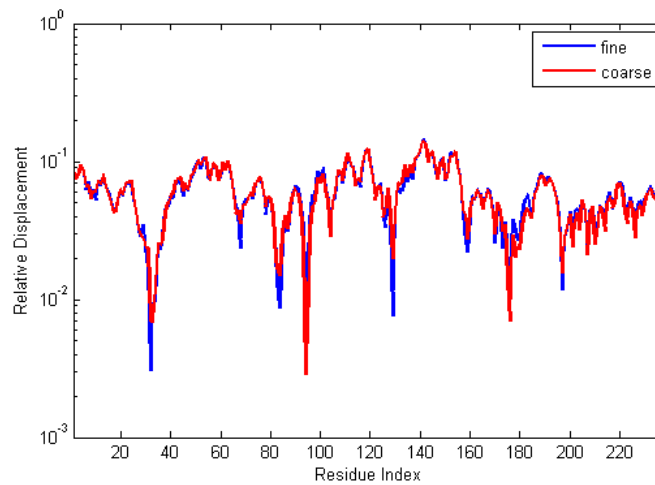
**Figure 4.15.** Che Y protein: Cumulative overlap from (a) the fine model and (b) the coarse model. Since the Che Y protein has a pattern of motion in which directions change constantly, the results have poor motion prediction (i.e. low overlap values) for not only the fine model but also the coarse model. In terms of overall performance in over 20 normal modes, the coarse model closely parallels the fine model. Both models yield the maximum overlaps at the first normal mode. (c) Comparison of relative displacement between the fine and the coarse models at their most dominant modes. The tight correspondence between two lines confirms again the capability of the coarse model.



(a) Fine model : *cumulative overlap*

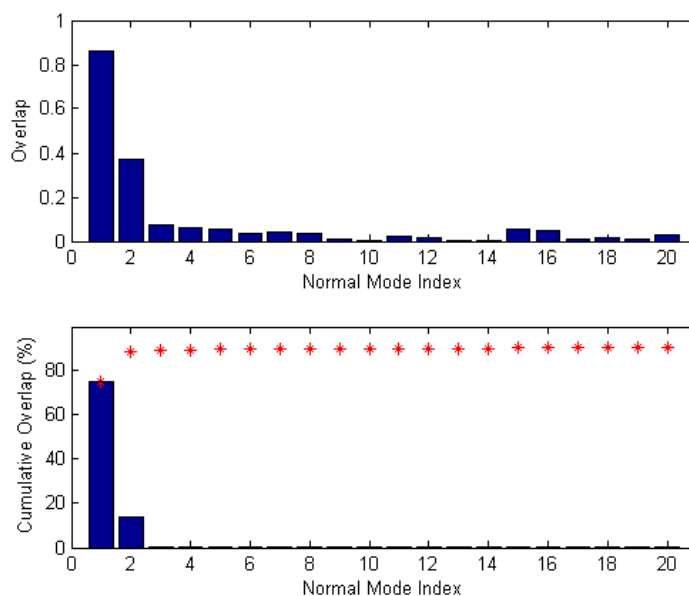


(b) Coarse model : *cumulative overlap*

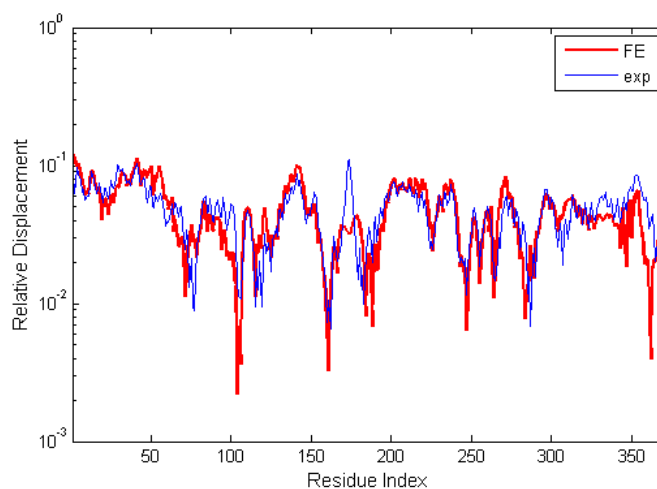


(c) Relative displacement : FE – mode 2 and EN – mode 4

**Figure 4.16.** LAO binding protein: Cumulative overlap from (a) the fine model and (b) the coarse model. (c) Comparison of relative displacement between the fine and the coarse models at their most dominant modes. The coarse model yields not only high overlap at the first mode but almost identical predictions to the fine model. This capability was shown in the previous case (Figure 4.12).



(a) *Overlap and cumulative overlap.*



(b) Relative displacement : FE – mode 1 vs. experimental data

**Figure 4.17.** Maltodextrin binding protein: (a) Values of overlap and cumulative overlap. The pattern of results of maltodextrin binding protein has a great resemblance to those of the LAO binding protein (Figure 4.17(a)(b)). Likewise, the maximum overlap happens at the first normal mode with a remarkably high value. (b) The comparison of relative displacement between the most dominant mode and experimental data provides a prominent demonstration to the correspondency in (a); the majority of the data at the first mode coinciding with the experimental data.

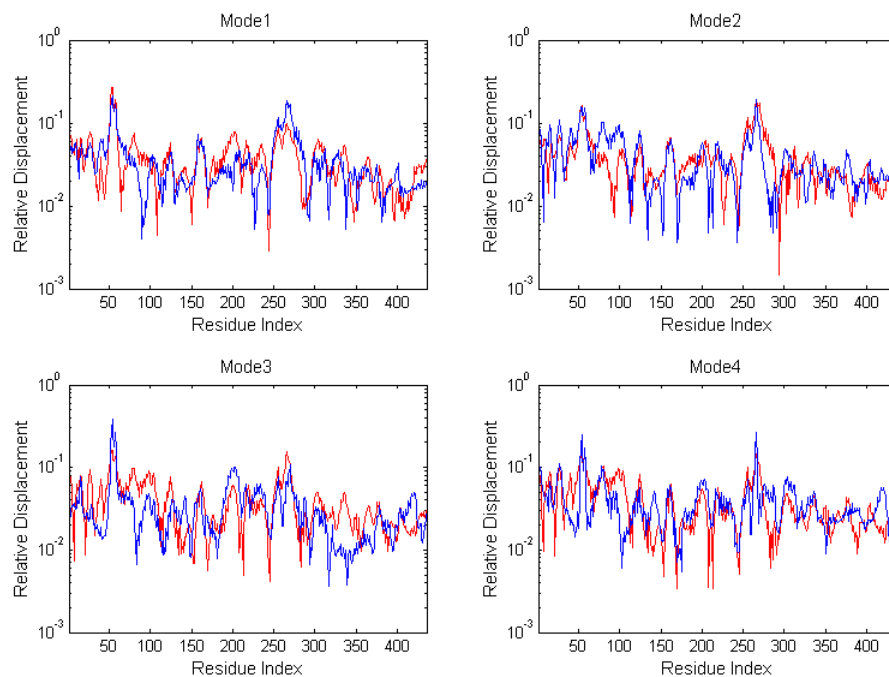
an important role in this upshot. The maltodextrin binding protein has a smaller cross-section in the middle and performs a hinge-like motion, with a bit of it twisting, during conformational change. The conformational change is fairly collective which involves a large amount of residue.

#### 4.2.2.2 Enolase

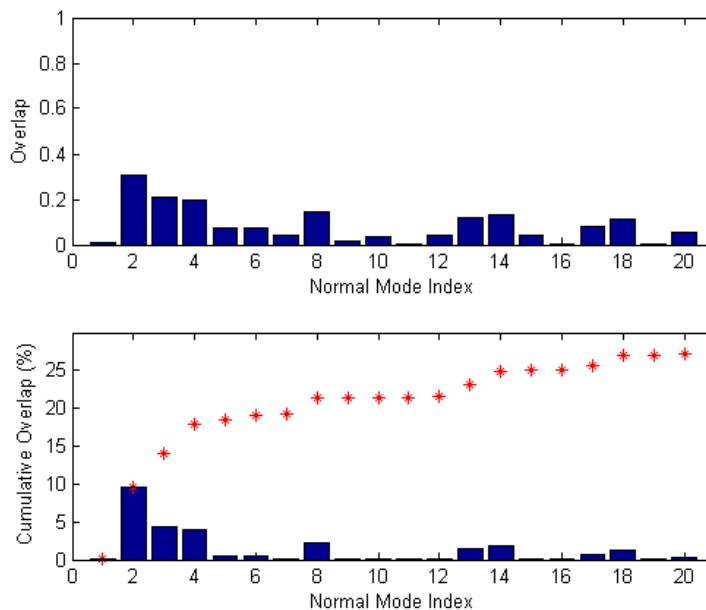
In this case, the bar chart (Figure 4.18(b)) shows quite a dispersive distribution of overlap of enolase protein. While the first three maximum overlaps all happen within the first four normal modes, the cumulative rate of overlaps over twenty normal modes is relatively low. However, this consequence is predictable in terms of collectivity. Enolase has a fairly low collectivity, i.e. 0.10, which implies a more localized motion during conformational change. Figure 4.19 is a visualization of the second normal mode. The second mode holds the maximum overlap and is also the only collective motion predicted in the dominant modes. Other dominant modes with high overlaps have motions confined to a small amount of residue and are not as collective as the second mode.

#### 4.2.2.3 Lactoferrin

In comparison with the previous case, lactoferrin has a relatively concentrated distribution of overlaps. All dominant modes (i.e. high overlap modes) occurred within the lowest few modes (Figure 4.20(b)). Cumulative overlaps reach a fair plateau around the eighth normal mode. Furthermore, the maximum overlap is found at the third normal mode which also corresponds to the inference by Tama et al.– the most dominant modes are very often found within the first four modes.[69] Figure 4.21 shows the open and closed conformations of lactoferrin (Figure 4.21(a)&(b)), accompanying the motion pattern of normal modes with the first two maximum overlaps (Figure 4.21(c)–(f)). It is evident that these two dominant normal modes (i.e. mode 3 and mode 4) yield good predictions of the collective motion.

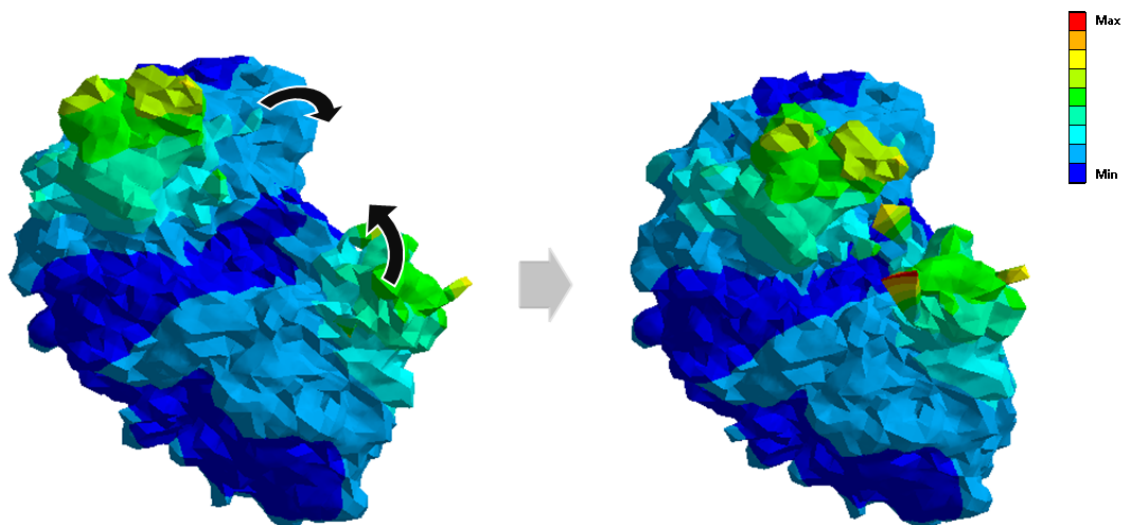


(a) Enolase: Comparison of *relative displacements* of the first four normal modes. (FE model - red line; EN model - blue line)



(b) EN model : *overlap* and *cumulative overlap*

**Figure 4.18.** Enolase: (a) Relative displacement and (b) values of *overlap* and *cumulative overlap*. The bar chart shows quite a dispersive distribution of overlap of enolase protein. While the first three maximum overlaps all happen within the first four normal modes, the cumulative rate of overlaps over twenty normal modes is relatively low.



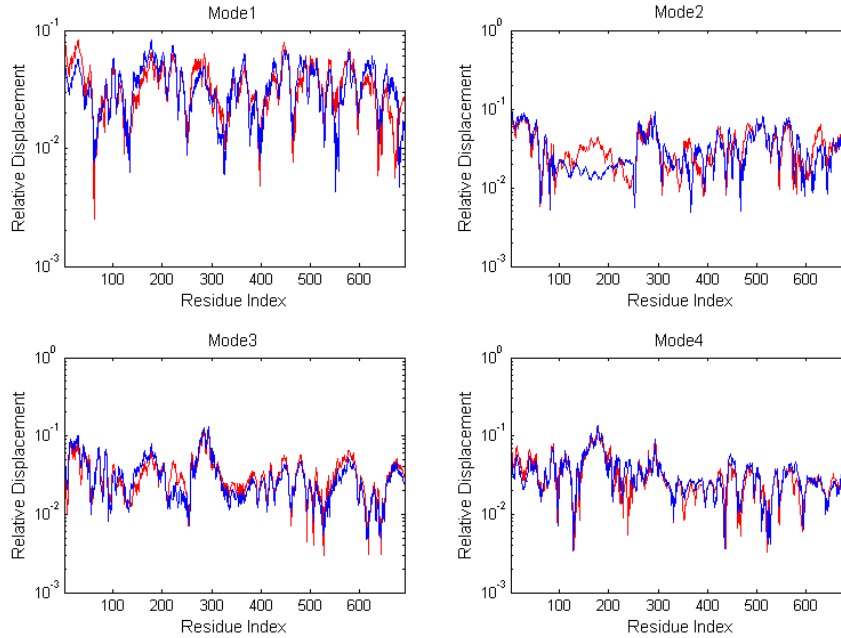
**Figure 4.19.** Enolase: The movement of the second normal mode. This normal mode holds the maximum overlap and is also the only collective motion predicted in the dominant modes. (Image rendered with ANSYS 13.0)

### 4.3 Overall Discussion

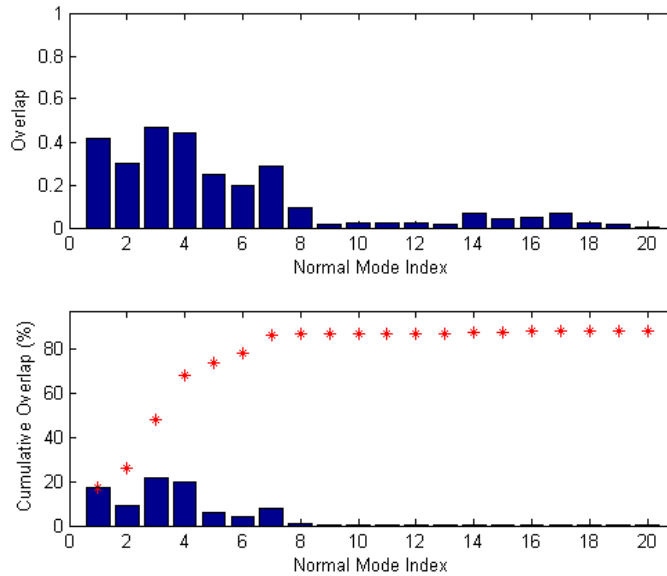
Generally speaking, finite element based normal mode analysis is competent in comparison with elastic network based analysis. In the second section of this chapter, the finite element (FE) model performs as well as the elastic network (EN) model dose for all three fine models. Moreover, the coarse model from the third section also shows its great potential. In terms of all six proteins, four of them yield pleasing results; low frequency modes yield high overlap values in general. The unfavorable results from the other two proteins might refer to an important factor: degree of collectivity.

As shown in Table 4.2, both proteins (i.e. Che Y protein and enolase) have fairly low degrees of collectivity. In other words, their conformational changes are rather localized. Since normal mode analysis is favorable for predicting collective motions, it is conceivable that low degrees of collectivity might lead to poor overlap results. Fur-



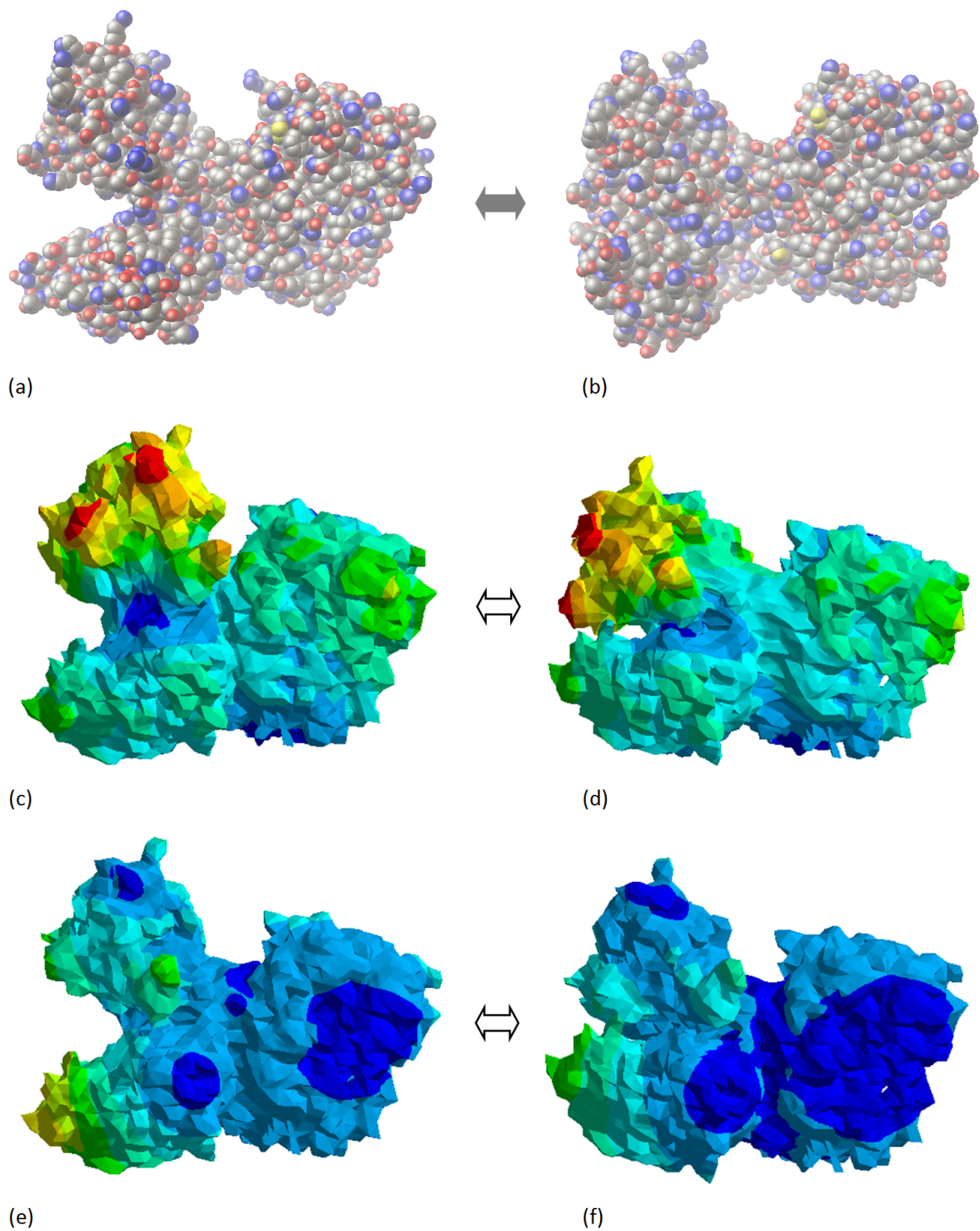


(a) Lactoferrin: Comparison of *relative displacements* of the first four normal modes. (FE model - red line; EN model - blue line)



(b) EN model : *overlap* and *cumulative overlap*

**Figure 4.20.** Lactoferrin: (a) Relative displacement and (b) values of *overlap* and *cumulative overlap*. Lactoferrin has a relatively concentrated distribution of overlaps. All dominant modes (i.e. high overlap modes) occurred within the lowest few modes. Cumulative overlaps reach a fair plateau around the eighth normal mode. The maximum overlap is found at the third normal mode which corresponds to the inference by Tama et al.– the most dominant modes are very often found within the first four modes.[69]



**Figure 4.21.** Lactoferrin: (a)(b) – Conformational change from experimental data. (c)(d) – Motion pattern of the 3<sup>rd</sup> normal mode. (e)(f) – Motion pattern of the 4<sup>th</sup> normal mode. The similarity among these three sets of motion pattern indicates that these two dominant normal modes perform well on predicting the collective motion. (Image rendered with Rasmol and ANSYS)

thermore, Table 4.2 implies that the predictability of conformational changes might be highly correlated to degrees of collectivity. Both LAO and maltodextrin binding proteins, which have the highest collectivities, yield admirably high overlaps values at the lowest mode. HIV-1 protease and lactoferrin, which have degrees of collectivity higher than the lowest two, also yield satisfactory results for the first few normal modes. However, overlaps are quite dispersed for the Che Y protein and enolase over the first twenty modes, and dominant modes are less obvious.

In addition, the single dominant mode proposed by Tama et al. may be greatly affected by degrees of collectivity. As mentioned in previous sections, the very single dominant mode was found at the first normal mode for both LAO and maltodextrin binding proteins, which have the highest degrees of collectivity. Moreover, for HIV-1 protease and lactoferrin, the single dominant mode was also found within the first four modes (i.e. second and third mode, respectively). However, for Che Y protein and enolase which have the low degrees of collectivity, overlap values are low and the single dominant mode is unrecognizable. As a result, it is likely that higher degrees of collectivity will result in a higher degree of recognition in the single dominant mode. This corresponds to the spirit of normal mode analysis that movement with higher collectivity is more likely to be well predicted with few dominant modes.

Nonetheless, there are some mechanical responses resulting from molecular shape that cannot be learned solely from analytical data. At this point, it is crucial to take a close look at the three-dimensional protein structure. For instance, sometimes the moving direction of residues changes notably during conformational change. Since normal mode analysis predicts the instantaneous movement at the ground state, while experimental data (i.e. displacements between two end conformations) reveals movement concerning the two end states, so comparison of two sets of data might result in poor correspondency. Therefore, it is critical to observe and compare all the information in different aspect for acquiring a comprehensive understanding of the results.

**Table 4.2.** An overview of protein collectivities and cumulative overlaps. The variation pattern of cumulative overlaps implies that the predictability of conformational changes might be highly correlated to degrees of collectivity.

Protein	Collectivity	Cumulative Overlap
HIV-1 protease (99)	0.23	
Che Y protein (128)	0.17	
LAO binding protein (238)	0.68	
Maltodextrin binding protein (370)	0.67	
Enolase (436)	0.10	
Lactoferrin (691)	0.48	

## CHAPTER 5

### CONCLUSIONS

This study proposed a finite element (FE) procedure for protein normal mode analysis (NMA). The FE model adopted the protein solvent-excluded surface to generate a homogeneous and isotropic volume. A simplified triangular approximation of the coarse molecular surface was generated from the original surface model by using the Gaussian-based blurring technique. Similar to the elastic network (EN) model, the FE model holds a major advantage over all-atom NMA; the computationally expensive process of energy minimization that may distort the initial protein structure has been eliminated. This modification significantly increases the efficiency of NMA. Fair results from the FE model also fortify the capability of this simplification manner.

In comparison with the EN model, the FE model successfully brings out the capability of NMA in low-frequency/high collectivity molecular motion. In spite of the hypothesis of homogeneity, which contradicts the protein intrinsic property and other NMA methods, the FE model performs encouragingly by capturing protein shape properties. This also underlines the predominance of shape in protein dynamic behavior. In addition, the coarsen molecular surface model yields competent results in comparison with the original solvent-excluded surface model.

An exclusive advantage of FE method proposed here is the forthright visualization feature. Protein behavior can be complicated to analyze and elucidate. Observations on both protein structure and visualized motion pattern provide perspectives that may clarify potential blind spots and deepen the understanding.

However, a great drawback of this FE model is the slow meshing process which can take hours to days. Moreover, the irregular shapes of proteins can lead to problematic and sometimes unsuccessful meshing. To enhance the practicality of the FE model, it is imperative to improve the quality and speed of meshing. On the other hand, except the built-in mesher in the FE software used here, numerous developing mesh programs providing better performance are available from other resources. By coupling a superior mesh program to the FE model, the issue might be overcome.

Nonetheless, a couple more advantages, which have not been shown in this study, should also be addressed. Firstly, the FE method is able to calculate the mechanical response of proteins and apply to bending, buckling, and other generalized loading scenarios. This feature is crucial for probing the structure-function relation of protein supramolecular assemblies, as has been studied by Bathe et al.[9]. Secondly, the Poisson-Boltzmann equation used to model aqueous electrolyte-mediated electrostatic interactions in proteins may be coupled directly to the FE model. Since protein dynamics can be influenced by various factors such as pH, temperature, ionic strength, etc., coupling analysis can provide more realistic and comprehensive perspectives. Finally, the atomic Hessian can be projected onto the FE space in order to incorporate atomic-level interactions into the model, thereby eliminating the a priori assumption of homogeneous isotropic elastic response.

The application of the FE method in protein NMA is still in its infancy. To the authors' knowledge, research with this technique up until now has only been studied by Bathe et al.[9, 40]. There is plenty of room to explore in this area. It is worth expecting that future studies of FE-based NMA will provide distinct contributions in protein dynamics prediction.

## BIBLIOGRAPHY

- [1] ANSYS ®. *Academic Research, Release 13.0.*
- [2] Amadei, A.; Linssen, A. B.; Berendsen H. J. *Proteins* 17, 412 (1993).
- [3] Anfinsen, C.B. Thermodynamics of structural stability and cooperative folding behavior in proteins. In *Advances in Protein Chemistry* (Orlando, Florida, 1984), vol. 36, Academic Press Inc, pp. 316–334.
- [4] ANSYS, Inc. ANSYS mechanical APDL element reference. *Academic Research, Release 13.0.*
- [5] Atilgan, A. R.; Durell, S. R.; Jernigan R. L.; Demirel M. C.; Keskin O.; Bahar I. *Biophys. J.* 80, 505 (2001).
- [6] Attard, P. Equipartition theorem. In *Thermodynamics and Statistical Mechanics: Equilibrium by Entropy Maximisation* (525 B Street, Suite 1900, San Diego, CA 92101-4495, 2002), Academic Press, pp. 165–166.
- [7] Bahar, I.; Lezon, T. R.; Bakan A.; Shrivastava I. H. Normal mode analysis of biomolecular structures: Functional mechanisms of membrane proteins. *ACS Chem. Rev.* 110, 3 (2010), 1463–1497.
- [8] Bathe, K. J. In *Finite element procedures* (Upper Saddle River, NJ, 1996), Prentice-Hall.
- [9] Bathe, M. A finite element framework for computation of protein normal modes and mechanical response. *Proteins* 70 (2008), 1595–1609.
- [10] ben Avraham, D.; Tirion, M.M. Dynamic and elastic properties of factin: a normal modes analysis. *Biophys J* 68 (1995), 1231–1245.
- [11] Berman, H.M.; Westbrook, J; Feng Z; Gilliland G; Bhat T.N.; Weissig H.; Shindyalov I.N.; Bourne P.E. The protein data bank. *Nucleic Acids Research* 28 (2000), 235–242 (<http://www.pdb.org/>).
- [12] Brooks, B.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* 80, 6571 (1983).
- [13] Bruschiweiler, R. *J. Chem. Phys.* 102 (1995).
- [14] Case, D. A. Normal mode analysis of protein dynamics. *Curr Opin Struct Biol* 4 (1994).

- [15] Chennubhotla, C.; Rader, A. J.; Yang L. W.; Bahar I. Elastic network models for understanding biomolecular machinery: From enzymes to supramolecular assemblies. *Physical Biology* 2 (2005).
- [16] Claessens, M.M.A.E.; Bathe, M.; Frey E.; Bausch A.R. Actin-binding proteins sensitively mediate f-actin bundle stiffness. *Nat Mater* 5 (2006).
- [17] Connolly, M.L. *J. Appl. Cryst.* 16 (1983), 548–558.
- [18] Cortis, C.M.; Friesner, R.A. Numerical solution of the poissonboltzmann equation using tetrahedral finite-element meshes. *J Comput Chem* 18 (1997).
- [19] Cui, Q.; Bahar, I. E. In *Normal Mode Analysis. Theory and Applications to Biological and Chemical Systems* (Boca Raton, FL, 2006), Taylor & Francis Group.
- [20] de Groot, Steven Hayward; Bert L. Normal modes and essential dynamics. In *Molecular modeling of proteins* (Clifton, N.J., 2008), Andreas Kukol, Ed., Totowa, NJ : Humana Press, pp. 89–106.
- [21] Doruker, P.; Jernigan, R. L.; Bahar I. J. *Comput. Chem.* 23, 119 (2002).
- [22] Erman, I. Bahar; A. R. Atilgan; B. *Fold. Des.* 2, 173 (1997).
- [23] Erman, T. Haliloglu; I. Bahar; B. *Phys. ReV. Lett.* 79, 3090 (1997).
- [24] Eyal, E.; Yang, L. W.; Bahar I. *Bioinformatics* 22, 2619 (2006).
- [25] Fischer, H; Polikarpov, I; Craievich A.F. Average protein density is a molecular-weight-dependent function. *Protein Sci.* 13, 10 (October 2004), 2825–2828.
- [26] Frauenfelder, H.; McMahon, B. H.; Austin R. H.; Chu K.; Groves J. T. *Proc. Natl. Acad. Sci. U.S.A.* 98, 2370 (2001).
- [27] Garcia, A. E.; Harman, J. G. *Protein Sci.* 5, 62 (1996).
- [28] Gere, J.M.; Timoshenko, S. Linear elasticity, hooke’s law, and poisson’s ratio. In *Mechanics of Materials* (1997), Boston : PWS-Kent Pub. Co., pp. 27–31.
- [29] Green, D.F.; Tidor, B. Escherichia coli glutaminyl-trna synthetase is electrostatically optimization for binding of its cognate substrates. *J Mol Biol* 342 (2004).
- [30] Greer, J.; Bush, B.L. *Proc. Natl. Acad. Sci. USA.* 75 (1978), 303–307.
- [31] Hamacher, K.; McCammon, J. A. *J. Chem. Theory Comput.* 2, 873 (2006).
- [32] Help System, Structural Guide, ANSYS Inc. ANSYS ®. *Academic Research, Release 13.0.*



- [33] Hendsch, Z.S.; Tidor, B. Do salt bridges stabilize proteins? a continuum electrostatic analysis. *Prot Sci* 3 (1994).
- [34] Hinsen, K. *Proteins* 33, 417 (1998).
- [35] Howard, J. In *Mechanics of Motor Proteins and the Cytoskeleton* (Sunderland, MA, 2001), Sinauer Associates, Inc.
- [36] Ivanovska, I.; Wuite, G.; Jonsson B.; Evilevitch A. Internal dna pressure modified stability of wt phage. *Proc Natl Acad Sci USA* 104 (2007).
- [37] Jernigan, I. Bahar; B. Erman; T. Haliloglu; R. L. *Biochemistry* 36 (1997).
- [38] Jolliffe, I. T. In *Principal Component Analysis* (New York, 2002), Springer.
- [39] Kaazempur-Mofrad, M.R.; Bathe, M.; Karcher H.; Younis H.F.; Seong H.C.; Shim E.B.; Chan R.C.; Hinton D.P.; Isasi A.G.; Upadhyaya A.; Powers M.J.; Griffith L.G.; Kamm R.D. Role of simulation in understanding biological systems. *Comput Struct* 81 (2003).
- [40] Kim, D.-N.; Nguyen, C.-T.; Bathe M. Conformational dynamics of supramolecular protein assemblies. *Journal of Structural Biology* 173 (2011), 261–270.
- [41] Kim, M.K.; Chirikjian, G.S.; Jernigan-R.L. Elastic models of conformational transitions in macromolecules. *Journal of Molecular Graphics and Modelling* 21 (2002), 151–160.
- [42] Kis, A.; Kasas, S.; Babic-B.; Kulik A.J.; Benoit W.; Briggs GAD; Schonenberger C.; Catsicas S.; Forro L. Nanomechanics of microtubules. *Phys Rev Lett* 89, 248101 (2002).
- [43] Kojima, H.; Ishijima, A.; Yanagida-T. Direct measurement of stiffness of single actin-filaments with and without tropomyosin by in-vitro nanomanipulation. *Proc. Natl. Acad. Sci. USA* 91 (1994), 12962–12966.
- [44] Kondrashov, D. A.; Cui, Q.; Phillips G. N. Jr. *Biophys. J.* 91, 2760 (2006).
- [45] Lyman, E.; Pfaendtner, J.; Voth G. A. *Biophys. J.* 95, 4183 (2008).
- [46] Marques, O.; Sanejouand, Y. H. *Proteins* 23 (1995).
- [47] Michel, J.P.; Ivanovska, I.L.; Gibbons M.M.; Klug W.S.; Knobler C.M.; Wuite G.J.L.; Schmidt C.F. Nanoindentation studies of full and empty viral capsids and the effects of capsid protein mutations on elasticity and strength. *Proc Natl Acad Sci USA* 103 (2006).
- [48] Ming, D.; Kong, Y.F.; Lambert M.A.; Huang Z.; Ma J.P. How to describe protein motion without amino acid sequence and atomic coordinates. *Proc Natl Acad Sci USA* 99 (2002).

- [49] Miyashita, O.; Tama, F. Normal mode analysis techniques in structural biology. *Encyclopedia of life sciences* (2007).
- [50] Naik, V. M.; Krimm, S.; Denton J. B.; Nemethy G.; Scheraga H. A. *Int. J. Pept. Protein Res.* *24*, 613 (1984).
- [51] Needleman, D.J.; Ojeda-Lopez, M.A.; Raviv U.; Ewert K.; Miller H.P.; Wilson L.; Safinya CR. Radial compression of microtubules and the mechanism of action of taxol and associated proteins. *Biophys J* *89* (2005).
- [52] Nicolay, S.; Sanejouand, Y. H. *Phys. Rev. Lett.* *96*, 078104 (2006).
- [53] Nishikawa, N. Go; T. Noguti; T. *Proc. Natl. Acad. Sci. U.S.A.* *80*, 3696 (1983).
- [54] Petsko, G.A.; Ringe, D. From sequence to structure. In *Protein Structure and Function* (23 Plumtree Road, Sunderland, MA 01375, 2004), Miranda Robertson Eleanor Lawrence, Ed., New Science Press., Sinauer Associates Inc.
- [55] Rader, I. Bahar; A. J. *Curr. Opin. Struct. Biol.* *15*, 586 (2005).
- [56] Rajakumar, C.; Rogers, C.R. The lanczos algorithm applied to unsymmetric generalized eigenvalue problem. *International Journal for Numerical Methods in Engineering* *32*, 5 (1991), 1009–1026.
- [57] Romo, T. D.; Clarage, J. B.; Sorensen D. C.; Phillips G. N. Jr. *Proteins* *22*, 311 (1995).
- [58] Roux, B. *Acc. Chem. Res* *35*, 366 (2002).
- [59] Roux, S. Berneche; B. *Nature* *414*, 73 (2001).
- [60] Rugonyi, S.; Bathe, K.J. On finite element analysis of fluid flows fully coupled with structural interactions. *CMES: Comput Model Eng Sci* *2* (2001).
- [61] Sanner, M.F.; Olson, A.J. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* *38*, 3 (1996), 305–320.
- [62] Sanner, M.F. A component-based software environment for visualizing large macromolecular assemblies. *Structure* *13* (March 2005), 447–462.
- [63] Sansom, I. H. Shrivastava; M. S. *Biophys. J.* *78*, 557 (2000).
- [64] Sedeh, R.S.; Bathe, M.; Bathe K.J. The subspace iteration method in protein normal mode analysis. *J. Comput. Chem.* *31* (2010), 66–74.
- [65] Sen, T. Z.; Jernigan, R. L. In *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems* (Boca Raton, FL, 2006), Chapman & Hall CRC.
- [66] Shemesh, T.; Geiger, B.; Bershadsky A.D.; Kozlov M.M. Focal adhesions as mechanosensors: A physical mechanism. *Proc Natl Acad Sci USA* *102* (2005).

- [67] Stern, M. Levitt; C. Sander; P. S. *J. Mol. Biol.* 181, 423 (1985).
- [68] Tama, F.; Wrighers, W.; Brooks C.L. Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory. *J Mol Biol* 321 (2002).
- [69] Tama, F.; Sanejouand, Y.-H. Conformational change of proteins arising from normal mode calculations. *Protein Eng.* 14, 1 (2001), 1–6.
- [70] Tang, Y.Y.; Cao, G.X.; Chen X.; Yoo J.; Yethiraj A.; Cui Q. A finite element framework for studying the mechanical response of macromolecules: application to the gating of the mechanosensitive channel mscl. *Biophys J* 91 (2006).
- [71] Tirion, M. M. *Phys. Rev. Lett.* 77 (1996).
- [72] Weaver, Jr. W.; Gere, J.M. In *Matrix Analysis Of Framed Structures, 3rd Edition* (1966), Springer-Verlag New York, LLC.
- [73] Wittenberg, J. B. Wittenberg; B. A. *Annu. Rev. Biophys. Biophys. Chem.* 19, 217 (1990).
- [74] Zheng, W.; Brooks, B. R.; Thirumalai D. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7664 (2006).
- [75] Zienkiewicz, O.C.; Taylor, R.L. In *The finite element method* (Boston, 2000), Butterworth-Heinemann.