2012

Immunoglobulin motions and antigen binding effects examined by elastic network models

Debkanta Chakraborty

Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/etd

Part of the Biochemistry Commons, Bioinformatics Commons, and the Biophysics Commons

Recommended Citation

Chakraborty, Debkanta, "Immunoglobulin motions and antigen binding effects examined by elastic network models" (2012). Graduate Theses and Dissertations. Paper 12923.

This Thesis is brought to you for free and open access by the Graduate College at Digital Repository @ Iowa State University. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Digital Repository @ Iowa State University. For more information, please contact hinefuku@iastate.edu.
Immunoglobulin motions and antigen binding effects examined by elastic network models

by

Debkanta Chakraborty

A dissertation submitted to the graduate faculty in partial fulfillment of the requirements for the degree of

Master of Science

Major: Bioinformatics and Computational Biology

Program of Study Committee

Robert L. Jernigan, Major Professor
Guang Song
Taner Sen

Iowa State University
Ames, Iowa
2012

Copyright © Debkanta Chakraborty, 2012. All rights reserved.
DEDICATION

To my beloved parents for their unconditional and endless love and support that they have provided to me since my birth.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .................................................................................. vii

ABSTRACT ........................................................................................................ ix

CHAPTER 1: INTRODUCTION ......................................................................... 1

1.1 Overview of Immunology ........................................................................ 1
1.2 Immunoglobulin Structures ..................................................................... 2
1.3 Elastic Network Models .......................................................................... 5
1.4 A guide to the present study .................................................................... 6

CHAPTER 2: IMMUNOGLOBULIN CDR MOTIONS ........................................ 8

2.1 Correlation Analysis of IgG structure ..................................................... 9
  2.1.1 Methods ......................................................................................... 9
    2.1.1.1 Correlated Motions in Normal Modes .................................. 9
    2.1.1.2 Calculation of Combined Effect by Weighting Inverse Eigenvalue . 10
  2.1.2 Results ....................................................................................... 10
2.2 Correlation Analysis for CDR Motions .................................................. 14
2.3 Similar GNM Computed Correlations for the IgG Structure and the CDRs ... 15
2.4 Changes in Internal Distances ................................................................. 17
  2.4.1 Methods ...................................................................................... 18
    2.4.1.1 Computing Changes in Internal Distances for ENMs .......... 18
    2.4.2 Internal Distance Changes of IgG ............................................. 18
2.5 Local and Global Contributions of Correlations ...................................... 21
  2.5.1 Methods ...................................................................................... 21
    2.5.1.1 Local Effect .......................................................................... 21
 CHAPTER 3: ANTIGEN-BINDING EFFECTS ........................................................................... 24

3.1 Similarities and Differences between two homologous Fab structures (bound to antigen and without antigen) ......................................................................................... 25
  3.1.1 Sequence Similarity between Complexed and Uncomplexed Anti-tumor Antibody BR96 Structures ................................................................................................. 26
    3.1.1.1 Methods ........................................................................................................ 26
    3.1.1.2 Results .......................................................................................................... 26
  3.1.2 Structural Similarity between Complexed and Uncomplexed Anti-tumor Antibody BR96 Structures ................................................................................................. 28
    3.1.2.1 Methods ........................................................................................................ 28
    3.1.2.2 Results .......................................................................................................... 29
  3.1.3 Pair-wise Distances (Distances between the coordinates of all the corresponding residues in the two structures) .................................................................................. 29

3.2 Antigen Binding Effects on the Important Motions of the Two Fab Structures (antigen bound and uncomplexed) ................................................................................. 30
  3.2.1 Overlap and Cumulative Overlap in the Normal Modes between Bound and Unbound Structures ................................................................................................. 31
    3.2.1.1 Methods ........................................................................................................ 31
    3.2.1.2 Results .......................................................................................................... 31
  3.2.2 Enumeration of Mean Square Fluctuations ........................................................................ 32

3.3 Replacing the IgG Fab’s in the IgG (each individual Fab in Pdb ID: 1igt) with the antigen bound Fab (Pdb id: 1cly) ......................................................................................... 33
  3.3.1 Structural Similarity between the Fab Structures Used ........................................................................ 34
  3.3.2 Changes to the Important Motions for the Chimeric Structure ........................................................................ 35
CHAPTER 4: PRINCIPAL COMPONENT ANALYSIS OF FAB STRUCTURES .... 40

4.1 Selection of the IgG Fab Structures ................................................................. 41
   4.1.1 Length Distribution of the Fab structures ................................................. 43
   4.1.2 RMSD Distribution of the Fab structures ................................................. 45
   4.1.3 Multiple Structure Alignment of the Fab Structures .................................. 46

4.2 Principal Component Analysis ........................................................................ 49
   4.2.1 Methods ........................................................................................................ 49
   4.2.2 Results .......................................................................................................... 50
       4.2.2.1 Dimensionality Reduction by PCA ......................................................... 50
       4.2.2.2 PCA Scatter Plots ................................................................................ 51

4.3 Correspondence between the Normal Modes obtained from ENM and first few Principal Components .................................................................................. 54
   4.3.1 Methods ........................................................................................................ 55
       4.3.1.1 Overlaps between Normal Modes and PCs ............................................ 55
   4.3.2 Results .......................................................................................................... 56

4.4 Residue Positional Fluctuations of the first three PCs ..................................... 57

CHAPTER 5: SUGAR EFFECTS ..................................................................................... 60

5.1 Various Models to Explain Sugar and Major Hinge Effects .............................. 61
5.2 Sugar Effects on the Major Motions the of IgG ............................................. 62
   5.2.1 Mean Square Fluctuations (MSF) ............................................................... 63
   5.2.2 Overlap across Normal Modes ................................................................. 65
   5.2.3 Correspondence between ANM Temperature Factors and X-ray B Factors ................................................................. 66
   5.2.4 Shape Vectors and Mode Vectors ............................................................. 68
5.3 Exploring the Relocation of Sugar at Different Positions in the Fc .................. 70
CHAPTER 6: CONCLUSIONS AND DISCUSSIONS........................................73

REFERENCES ........................................................................................................75
ACKNOWLEDGEMENTS

I like to take this opportunity to thank my advisor Dr. Robert Jernigan. This dissertation would not have been possible without his constant guidance and inspiration. He has given a lot of technical advice, support and encouragement and more importantly a lot of freedom to explore diverse topics in my research areas. I grew in confidence as a graduate student under his excellent guidance.

I would also like to thank the members of my Program of Study (POS) Committee; Dr. Guang Song and Dr. Taner Sen who gave me a lot of advice during my POS Committee Meeting and also who have supported me to attain my goal. Special thanks go to both former and current chair of our major, Dr. Volker Brendel and Dr. Julie Dickerson, also some other professors of our university including Dr. Edward Yu, Dr. Srinivas Aluru, Dr. Andrzej Kloczkowski, Dr. Drena Dobbs, Dr. Manimaran Govindarasa, Dr. Arun Somani, Dr. Lei Ying and Dr. Ahmed Kamal. Their advice, kind words, excellent lectures helped me a lot and will always be fondly remembered.

I am sincerely thankful to my Laboratory members. Special mention goes to Dr. Michael Zimmermann. He has always been a great source of help. His suggestions, numerous discussions with him, willingness to sort out many issues have helped me all throughout. I sincerely thank Ataur Katebi and Dr. Sumudu Leelananda for their technical and mental support. Ataur and Sumudu have been always available for discussion and they have always inspired me in many ways. Also all past and present members of Dr. Jernigan group deserve worthy mention.

Some of my friends in Ames, mainly Mohammed Saad Bin Shaheed, Mohammed Shaidur Rahman, Sugam Sharma deserve special thanks. They have always stood by me in the days of stress and utter anxiety and have always encouraged me in various ways.
Obviously, I am grateful to my parents, friends and other relatives back in India. My parents have always provided me the strength since when I was a child and they only have given me the inspiration to pursue this goal. Without them, nothing would have been possible in my life. Their sacrifices in my life will never be forgotten. Special thanks go to my grandparents who have not been alive any more, but whom I can hardly forget for all the memorable events in my life and for all their association with me. Also there are my Boro masi (eldest aunt), meshomoshai, my friends Nilanjan, Achintya, Ashik, Atanu, Joydeep, Subhasish, Partha, my undergraduate professors Dr. Asish Kumar De, Dr. Sanjit Kumar Dutta and numerous other people whose contributions in my life are just immeasurable and it is difficult to cite all of them in this small space. Last but not the least, I thank from the core of my heart to Almighty. It is His blessings and kindness which always provides us with immense strength, shows us the light in the utter darkness and guides us like a pole star.
ABSTRACT

Immunoglobulin motions are evaluated using Normal Mode Analysis with Elastic Network Models. By employing this approach, we learn about the important motions of the protein, for the domain motions and other internal motions, and see strong evidence of the dominance of the low frequency normal modes. We particularly investigate the CDR motions. The CDR loops tend to move with their attached domains. By finding internal distance changes, we determine which parts of the structure undergo more rigid body like motions and which parts encounter larger internal distance changes. It turns out that Fab undergoes a large extent of changes in its internal distances and the Fc part moves more rigidly. We also investigate the effects of sugar and antigen binding on the IgG structures. The antigen binding effect is highly significant in the CDR regions, since antigen binding seems to enhance the motions in this region, while the sugar effects are more localized to the Fc region. By performing Principal Component Analysis on the numerous available Fab structures, we gather information about motions apparent in this ensemble of conformations, the correspondence of Principal Components to the Normal Modes of the Elastic Network Models, the residue fluctuations in the first few Principal Components and other important information about Immunoglobulin Dynamics.
CHAPTER 1: INTRODUCTION

1.1 Overview of Immunology

The immune system is designed to combat microbes and chemicals that could damage the host system. It is the principal biological defense against pathogens. The field of Immunology deals with the study of all aspects, especially structures and functions of the components of the immune system in all organisms. The immune system generally can be thought of as possessing two lines of defense (33, 70). The first line of defense against a pathogenic intruder is called the innate immune system, while the second line of defense is called the adaptive immune system. Both the innate and adaptive immune systems have both cellular and humoral components to help them carry out their protective functions. There are a number of ways to distinguish between the innate and the adaptive immune systems. First, while the former represents a non-specific response to intruding organisms, the latter one (adaptive) is more specific, in that, it consists of “memory”. Second, adaptive immune systems take more time to react to an invading organism than the innate ones which are generally constitutively present. Moreover, the adaptive systems are antigen specific while the innate are not.

The specific interactions between antibodies (also known as immunoglobulin) and antigens are generally known as the humoral responses. Humoral immune responses are primarily mediated by antibodies which are secreted by B cells while B cells, being co-stimulated, are transformed into plasma cells. An antigen presenting cell (such as a dendritic cell) can co-stimulate the B cell. CD4+ T-helper cells help this entire process by providing co-stimulation. The antibodies that are thus secreted bind to antigens on the surface of invading microbes (mostly viruses or bacteria) or environmental chemicals and flag them for destruction. Antibody production, which is central to humoral immunity, is also accompanied by other accessory processes, such as, Th2 activation, cytokine production, germinal center
formation, isotype switching, affinity maturation etc. There are effector functions of the antibody, such as pathogen and toxin neutralization, classical complement activation, opsonin promotion of phagocytosis etc.

Cell-mediated immune responses do not involve antibodies, but rather employ the activation of phagocytes, natural killer cells (NK), cytotoxic T-lymphocytes and the release of cytokines in response to antigens. As in this case, the protective function of immunization is involved with various cells, hence this is called cellular immunity. The helper T cells or CD4 cells provide protection against pathogens, whereas cytotoxic T cells bring about apoptosis without using cytokines which leads to death. Other than helping in apoptosis in body cells, cellular immunity activates macrophages and natural killer cells that enable them to kill pathogens and stimulate cells to secrete various cytokines that assist in the functions of other cells involved in adaptive and innate immune responses. The cell-mediated immunity directly attacks microbes that survive in both phagocytes and non-phagocyte cells. Although its main function is to eradicate cells that are infected with viruses, it also helps to fight against fungi, protozoa, cancers etc.

1.2 Immunoglobulin Structures

Antibodies, also known as immunoglobulins, are glycoproteins that belong to the immunoglobulin superfamily. In this manuscript, henceforth, we will use the names antibodies and immunoglobulins interchangeably. Production of immunoglobulins is a major part of the humoral immune system. Immunoglobulin is a large Y-shaped protein produced by B cells and used by the immune system to destroy invading organisms such as bacteria and viruses. The unique molecular part, usually on the surface of the invading microbe that is recognized by an immunoglobulin is called an antigen. By means of specific binding between the antigen and the immunoglobulin an antibody is able to tag a microbe for destruction by other parts of the immune system, or simply by neutralizing its target directly.
Antibodies are large globular plasma proteins of about 150 kDa. The basic functional unit of each antibody is an immunoglobulin monomer (IgG), which consists of a single Ig unit, but these can also occur in multimeric forms - dimeric (IgA), tetrameric or pentameric (IgM). Immunoglobulin structures are well-defined (53 – 57, 33, 23). The Ig monomer consists of four polypeptide chains, two heavy chains (alpha, delta, gamma, epsilon and mu) and two light chains (lambda or kappa). These chains are linked together by disulfide bonds. Each light and heavy chain is again composed of structural domains known as immunoglobulin (Ig) domains. In the CATH database (72), these Ig domains fall under ‘Mainly beta’ class, ‘Sandwich’ architecture, ‘Immunoglobulin-like’ topology and ‘Immunoglobulin’ homologous superfamily. The Ig domains consists of about 70-110 amino acids and are subdivided into various categories according to their corresponding size and function, such as IgV which is variable and IgC which is constant. In each Ig fold, two beta sheets intertwine into a “sandwich” shape.

The different varieties of antibodies are known as isotypes or classes. There are five common antibody isotypes known as IgA, IgD, IgE, IgG and IgM in placental mammals. The type of heavy chains generally determines the immunoglobulin isotype. Alpha and gamma types contain approximately 450 amino acids whereas mu and epsilon have 550 residues. Each heavy chain has two regions – a constant region and a variable region. The constant region is identical in all antibodies of the same isotype, but differs in antibodies of different isotypes. A light chain also has two domains, one constant domain and one variable domain. A light chain has approximately 210-220 residues. Variable regions are included within the amino terminal end of the polypeptide chain (amino acids 1-110, approximately). Across various immunoglobulins, these amino acid sequences exhibit a high level of variability. Constant regions, on the other hand, which contain residues 111-220 approximately, are much more uniform when comparing one immunoglobulin to another within the same isotype.
Various parts of an antibody have unique functions. The arms of the Y of the antibody contain the sites that can generally bind two identical antigens and therefore recognize specific foreign organisms. This part is called the Fab region (Fab is the abbreviation for fragment, antigen binding). The Fab region is constituted from one constant and one variable domain from each heavy and light chain of the antibody. The variable domain is also referred to as the F\textunderscore V region and plays the most vital role for binding to antigens. In fact, variable loops of the β-strands, three each on the light (\(V_L\)) and heavy (\(V_H\)) chains are important for antigen binding and these loops are termed the complementary determining regions (CDRs). CDRs are also known as anti-idiotypes in the framework of immune network theory, according to which, the adaptive immune system is regulated by interactions between idiotypes. The base of the Y is instrumental for immune cell activity. This region is known as the Fc region. This Fc region consists of two heavy chains that contribute generally two or three constant domains. The Fc region of each antibody helps to generate a particular immune response for a given antigen, by binding to a specific class of Fc receptors. In the Fc region, sugar is generally bound to some residues, hence immunoglobulin is also known as a glycoprotein. The major hinge is at the connecting region between the Fab's and the Fc, and it is cysteine-rich and includes disulfide bonds.

In this study, we will investigate the dynamics of the IgG, its components, and the effects of binding of antigens and sugars.
1.3 Elastic Network Models

Elastic Network Models (ENM) have become an important tool and developed into a standard technique for the study of dynamics and structural flexibility of biological macromolecules such as proteins, RNA, etc (15, 16, 19-27, 30, 39). They are primarily used for identifying and characterizing the slowest motions, the most important contributors to the dynamics, by the lowest frequency normal modes (27). Often with Elastic Network Model, the protein is coarse-grained as a set of atoms (alpha carbons or some other representative group of atoms depending upon the level of coarse-graining) that are interconnected by a network of elastic and harmonic springs. It has been found that the coarse-grained representations are as successful in describing the structural dynamics as atomic models.

The Gaussian Network Models (GNM) (5) and Anisotropic Network Model (ANM) (6) are two versions of the common ENMs used to describe, respectively, the amplitude
(GNM) and direction of motions of residues around their static equilibrium positions (ANM). For GNM, a network of N nodes (where N is typically the number of total residues or alpha carbon atoms if coarse-graining is employed) is considered. The ANM will invariably have $3N-6$ internal vibrational modes of motion, and the GNM, $N-1$. The modes are usually obtained by Singular Value Decomposition of the Kirchoff Matrix or contact matrix describing the placement of springs, after the zero eigenvalues are removed ($6$ for ANM and $1$ for GNM).

One of the first Elastic Network Models was proposed by Tirion (1), who showed that an all-atomic Elastic Network Model can reproduce the lowest frequency part of the density of states of a protein as well as the fluctuations of its alpha carbon atom quite well. Over the years, the models (5, 6, 12, 16, 17, 25, 26, 42, 43, 45, 46) have been proven to reproduce known protein functional motions with high fidelity (Atilgan et al., 2001; Bahar et al., 1997; Bahar & Jernigan, 1998; Bahar et al., 1999; Bahar & Jernigan, 1999; Wang & Jernigan, 2005; Bahar et al., 1998; Hinsen 1998; Demirel et al., 1998; Jernigan & Kloczkowski, 2006; Keskin et al., 2002a; Keskin et al., 2002b; Keskin et al., 2000; Kim et al., 2003; Wang et al., 2004). The temperature factors calculated from ENM which are representative of protein dynamics tend to correlate highly with the crystallographic B-Factors (Debye-Waller Factors).

1.4 A guide to the present study

This study mainly focuses on IgG dynamics, the effects of antigen binding, comparisons against the Principal Components of a set of experimentally derived Fab structures and the effects of sugar.

In Chapter 2, we describe the most important IgG motions, CDR motions, and internal distance changes, where we have employed Elastic Network Models to achieve this end.

Chapter 3 is divided into two main sections. In the first part, we compare two homologous structures of Fab’s, one antigen bound and other not bound, to learn whether antigen binding has any significant effect on the motions of the Fab. We
compute the overlaps between the normal modes of these two cases. Their mean square fluctuations serve as a metric to describe the differences. In the second part of this chapter, we also investigate the effect of antigen binding on the entire IgG, where we have replaced one and then the other Fab of IgG (PDB id: 1igt) with the antigen bound Fab (PDB id: 1cly) to investigate impact of the binding on the motions of this protein.

In Chapter 4 we compute the Principal Components (PC) of various sets of Fab experimental structures. We focus on the X-ray Crystal structures of Fab’s and by performing PCA, we gather information about their conformational variability and its relationship to IgG dynamics. In the first part, we discuss how we select the Fab’s, in the second part we derive the Principal Components representing the variations in the structures, and finally we investigate the correspondence between the normal modes of a representative structure and the Principal Components of the ensemble.

In Chapter 5, we explore the effects of sugar on the dynamics. We carry out a sensitivity analysis by relocating sugars to different portions of the structure to see the effects.

Chapter 6 provides a discussion and summary of all results.
CHAPTER 2: IMMUNOGLOBULIN CDR MOTIONS

This chapter discusses in detail the Immunoglobulin Gamma (IgG) structure and its dynamics, with relatively more emphasis on the motions of the Complementary Determining Regions (CDRs). The dynamics of the IgG structure is mainly evaluated using Normal Mode Analysis (NMA) of the Elastic Network Model (ENM). We employ both Anisotropic and Gaussian Network Models (ANM and GNM) to examine the dynamics of the IgG structure.

In Section 2.1, Correlation Analysis of the motions of the parts of the IgG computed with ANM for the first few normal modes are considered. This illustrates the different motions of the Ig fold (in total there are twelve Ig folds in this structure). Also the Fab and Major Hinge motions are investigated with this kind of correlation.

Section 2.2 is an extension of 2.1 for the CDR motions. The CDR regions are the antigen binding domains and some of the CDR loops have very distinct motions in comparison to the global motions of the structure. In general, they show significant correlations with the domains to which they are bound.

Section 2.3 discusses the GNM correlations for the IgG structure and the corresponding CDR motions. As with ANM, GNM also demonstrates the distinct global motions of the two different Fab’s which have two distinct domain motions.

In Section 2.4, the internal distance changes are calculated and shown for several different thresholds of residue distances. This clearly shows that the internal distance changes are not identical for both Fabs’. It is not known whether the differences between the two Fab’s are real or a crystal artifact.

In Section 2.5 we endeavor to find the contributions of nearby residues locally and all the residues globally, by exploring the corresponding local and global contributions of residues on the individual amino acids in the IgG structure. The local effect is found to be more important than global or allosteric effects from this
analysis. But then, we identify the corresponding portions in the structure which are more or less affected by both the nearby and the distant residues.

2.1 Correlation Analysis of IgG structure

IgG plays an important role in the humoral immune system. IgG molecules are composed of two Fab's and one Fc region and each of the Fab’s and Fc’s is comprised of four Ig domains, so altogether there are twelve Ig domains in the IgG structure. This structure mainly consists of two heavy chains and two light chains. While the two heavy chains are distributed through the Fab and Fc regions, the light chain is located only in the Fab. There is generally a distinct hinge region in the connecting region between the Fab’s and the Fc. Also there is a CDR region in each of the two outer portions of the Fab’s where the antigens actually bind.

As it is very well explored elsewhere (5, 6, 15, 19, 21, 25), the dynamics of the protein structures can be explained well by NMA of an ENM. The lowest frequency normal modes are the most dominant universally, and they can characterize the important motions of proteins. For further investigation, we consider the correlations of the motions for each pair of residues. Correlation analysis is able to detect and capture the Fab, Fc, hinge motions and also the different Immunoglobulin (Ig) fold motions.

2.1.1 Methods

2.1.1.1 Correlated Motions in Normal Modes

The standard ANM approach was used, and it provides directly, for each mode, information about the correlated blocks of structure that move together. See Reference (6) for the information about this.
2.1.1.2 Calculation of Combined Effect by Weighting Inverse Eigenvalue

Next we try to see the impact of the combined effect of the first few lowest frequency normal modes for the correlated blocks. This is done to get an approximate look into the main motions of the IgG.

In order to visualize the accumulated effect of first few normal modes, we simply add the motions of the pair of residues for each normal mode and weight these appropriately.

2.1.2 Results

Primarily the domain motions are captured in this analysis. Interestingly, the hinge region is also identified. (Figure 2.1).
Figure 2.1: Correlation of the 1st normal mode for the IgG structure (PDB Id: 1IGT) (Red denotes the highest positive correlation, while blue denotes the highest negative correlation or anti-correlation). Fc region shows the highest correlation, followed by the two domains Fab1 and Fab2. The motions of Fc are mostly anti-correlated with the motions of the two Fab’s.

In Figure 2.1, we show the correlations between parts of the IgG for mode 1. Approximately the first 410 residues constitute the Fab1 region, with 425 to 835 being the Fab2. Around 30 residues (411-425) and (836-850) are the major hinge and then the remaining (850-1316) are the Fc region. Figure 2.1 demonstrates the correlations for the first normal mode for all the residue pairs in the IgG structure. And it is clear that Fc shows the most highly coordinated domain motion in the first normal mode and most of the two Fc’s move in an anti-correlated manner with both Fab1 and Fab2.

Figure 2.2: Correlations for IgG for 4 Normal Modes (2, 3, 4 and 9). The Fab’s exhibit less coordinated domain motions in comparison to Fc. Mode 9 shows distinctly the
coordinated motions of the 12 Ig folds along with the 3 large domain motions (two Fab’s and Fc).

In Figure 2.2, the correlation plots for modes (2-4 and 9) are depicted. In modes 2-4, we again experience very high correlations in the Fc domain, while Fab1 and Fab2 again show less coordinated domain motions compared to Fc. Interestingly, Mode 9 is the one that not only captures the motions of all 12 Ig fold quite distinctly, but also shows the cohesive character of the three main domain motions.

We capture the snapshot for the combined effect of all first 9 normal modes in Figure 2.3. As discussed in the Methods section, it is inverse eigenvalue weighted normalization of the correlations and hence we tend to see the dominance of the lowest modes (mainly modes 1 and 2) in the combined effect diagram. So, we can conclude that the lowest mode, i.e., mode 1 is the most important among the first few modes.

Figure 2.3: Combined Effect of all 9 lowest modes. The lowest frequency mode clearly dominates in the Combined Effect diagram (the correlations for only modes 1 and 9 are shown on the left, shown for comparison).
We further investigate the loop motions for the IgG structure. The loops are the parts that do not fall within categorized secondary structure such as alpha helices and beta sheets. They join the main secondary structures (helices and sheets). They often lie on the surface and are exposed to solvent. Here we consider those loops whose relative solvent exposed area (73) is 85% or higher, and we see that the loops tend to move in an independent fashion which is not dependent on the main structure of IgG.

In Figure 2.4, on the right side we have extracted the loops alone. Here we find mainly that the loops move in a way similar to the entire structure of IgG except for few small cases. (There is more anti-correlation in the loops of the Fc region domain part than Fab1 and Fab2 compared to the entire structure of IgG, even though this difference is not so significant). This kind of approach has been studied before in Jernigan lab (52).
Figure 2.4: Correlations for Loop motions (defined by RSA solvency 85\%) extracted for modes 1 and 9. (The entire structure is composed of 1316 residues, whereas the loops defined in this way constitute around 410 residues). Loops move in a similar coordinated fashion as the entire protein structure.

2.2 Correlation Analysis for CDR Motions

Here we investigate the correlations, as in the previous section but for the CDR regions. The CDR regions are the main antigen-binding parts of the IgG structure, and we look for any pattern in the different CDR loop motions that would inform us whether any of the CDRs has a particular characteristic motion independent of the other parts or if there is any allostery apparent in these CDR loop motions.

Since there are two Fab’s (Fab1 and Fab2) in the IgG structure, we have defined the CDR loops contained in Fab1 as CDR1 and the ones in Fab2 as CDR2.

In Figure 2.5, the correlations for the CDRs (CDR1 and CDR2) are shown. The six CDR1 and six CDR2 loops are shown (color coded in the upper and rightmost bar by blue and red respectively, for the five lower frequency normal modes 2-6 in the order from left to right), while the positive highest correlation and negative highest correlation are color coded by green and black respectively with the white denoting the zero correlation.

In mode 2, CDR1 and CDR2 are highly coordinated within their corresponding domains and also there is a high correlation between these two domains. This high inter-domain correlation between CDR1 and CDR2 gradually decreases as we go from mode 2 to mode 6. In modes 4 and 5, in fact there is a very high anti-correlation between CDR1 and CDR2. In mode 6, we tend to find a minor hinge motion present in the CDR2 with some greater internal motions within CDR2 compared to CDR1. Also in mode 4, CDR2 shows more correlation within its domain than does CDR1.

This type of analysis for CDR regions throws some light on the dynamics of the antigen binding process. From this approach, we can identify which of the corresponding CDR loops are moving in a more coordinated manner to understand their role in the antigen binding function of IgG.
Figure 2.5: CDR motion correlations for CDR1 and CDR2 for 5 Normal Modes (2-6). Blue denotes the 6 CDR1 loops, while red represents the 6 CDR2 loops. Correlations are color-coded from 1 (green) through 0 (white) to -1 (black). The high inter-domain correlations between CDR1 and CDR2 gradually diminish as we move from mode 2 to mode 6. In modes 4 and 5, these two CDRs move in a highly anti-correlated fashion.

2.3 Similar GNM Computed Correlations for the IgG Structure and the CDRs

A similar correlation analysis is performed with GNM for the IgG structure. In GNM, unlike the ANM, the directionality of motions is not obtained. And even though we find the most coordinated motions again within the Fab or Fc domains, unlike ANM, most of the lowest modes of GNM fail to capture the individual Ig fold motions completely distinctly.

In Figure 2.6, the correlation analysis is shown which is calculated from the GNM for the first six normal modes for the IgG structure. These motions are more discrete than from ANM with the magnitudes of motions varying from +1 (highest) to -1.
(lowest) as shown below. We can see bigger blocks (or correlations) for the lowest frequency modes, whereas the motions are for much smaller blocks of structure in modes 5 and 6. We can see the 12 distinct Ig fold motions in modes 5 and 6, but otherwise these are not very much evident since GNM depicts the more coordinated motions within a Fab or Fc domain, unlike ANM.

Figure 2.6: Structural correlations calculated from GNM for first 6 Normal Modes for the IgG structure (White denotes the positive correlation (1), whereas black denotes anti-correlation (-1)). The block sizes for the GNM correlations decrease as the frequency of the normal modes increases. In modes 5 and 6, the 12 Ig fold motions are apparent.
Figure 2.7: Correlations from GNM for the CDRs (see Fig. 2.5 for the key to identifying the loops). Positive correlations are green and anti-correlations are black. Modes 1 and 4 show identical behaviors. Fab2 has internal motions apparent in modes 2, 5, and 6; whereas the CDRs of Fab1 are entirely cohesive in all of these modes.

The CDR plots for the GNM (done in a similar way as was done for the CDRs with ANM), are shown in Figure 2.7, and these show very distinct CDR domain motions. In mode 5, we encounter a minor hinge within CDR2 and also there are distinct internal motions within this Fab2 domain, even in mode 2.

2.4 Changes in Internal Distances

By computing changes in internal distances, we can learn which portion of the protein shows more rigid body motions and which has more relative internal motions within the domain.
2.4.1 Methods

2.4.1.1 Computing Changes in Internal Distances for ENMs

Here we mainly consider the relative displacements of the positions of residue points within the structure with the help of ANM.

The mean square change in internal distances is computed as

\[ \langle \Delta R_i - \Delta R_j \rangle^2 = \langle \Delta R_i^2 \rangle + \langle \Delta R_j^2 \rangle - 2\langle \Delta R_i \cdot \Delta R_j \rangle \] ...........................(1)

These values on the right hand side are obtained directly from the inverse of the contact matrix, from which we derived the normal modes.

Next we compute the normalized change in internal distances. This is the metric which is generally used to compare the magnitude of internal distance changes (23).

\[ \langle (\Delta R_i - \Delta R_j)^2 \rangle' = \frac{\langle (\Delta R_i - \Delta R_j)^2 \rangle}{\sqrt{\Delta R_i^2 \Delta R_j^2}} \] .................................(2)

2.4.2 Internal Distance Changes of IgG

The internal distance changes can shed light on the detailed changes within a protein structure during its dynamics. These changes include understanding how the normal modes stretch and compress the various parts and also alter various pairwise distances between parts of the structure. No internal distance change between a pair of points indicates that the two points move together rigidly, and hence the distance between them remains unchanged.

In figure 2.8, are shown the internal distance changes for different residue thresholds. For this purpose, we use 10, 15 and 20 Å distance thresholds between two points (shown in A, B and C) and also compute the corresponding changes without using such a distance threshold (in D).
We find the changes in internal distances to be higher in Fab1 and Fab2 compared to the Fc region. These findings clearly show that the Fab domains have substantially more internal motion than does the Fc domain.

This is in agreement with the fact that the Fab's (more specifically the CDRs) bind to antigens, and the most functional parts lie in the Fab’s rather than in the Fc. So this might be expected that the Fab’s undergo higher internal motion than Fc. This intrinsic dynamics may even facilitate the binding to different antigens. Figure 2.8 (D) demonstrates that Fab1 furthermore undergoes relatively more internal distance changes compared to Fab2.
Figure 2.8: Changes in Internal Distances for all modes by using various distance thresholds. The first around 410 residues constitute the Fab1, followed by around 250 residues in the first part of Fc, followed by the 410 residues comprising the Fab2 and finally the second part of the Fc is the final 250 residues. The Fab’s undergo more internal motions than Fc. In (D), where no distance threshold is imposed, it is evident that the Fab1 shows significantly more internal distance changes than do the Fab2 or other parts.
2.5 Local and Global Contributions of Correlations

2.5.1 Methods

After we calculate the correlations for each residue pair in the IgG structure, we show the contribution of pair-wise correlations on the structure. For this we employ the approach of calculating the mean pair-wise correlations of each residue and some other residues locally and at a distance and then show these values for each residue on the structure. By this method, we can show which part of the structure is moving with more or less correlation with its nearest residues (domain motion) and with all the residues globally (global motion).

2.5.1.1 Local Effect

Here we calculate the mean pair-wise correlations of each residue and five residues on either side of this residue to yield a matrix of all such mean pair-wise correlation for all the residues. Next we plot the matrix of these mean pair-wise correlations on the structure. We call this as Local Effect or Local Contribution of Correlations since each residue thus shows the relative movement with its neighboring 10 residues (5 on each side) – see Fig. 2.9.

2.5.1.2 Global Effect

We call the contribution of the pair-wise correlations from residues on a single residue as the Global Effect. In order to achieve this, we calculate the mean of all such pair-wise correlations for all residues for each single residue. Here each residue shows the relative motion in comparison to all residues in the structure and hence it does not provide any information about whether a particular residue is moving in correlation with its own domain or not – see Fig. 2.10.

2.5.2 Results

In Figure 2.9, we show the results for the local effects for modes 1 and 2 for the IgG structure. The red to blue spectrum, denotes a gradual decrease in correlations from
highest to lowest (red being highest and blue the lowest). For both of the two lowest frequency normal modes, the contribution of the neighboring 10 residues on each side of each residue demonstrates how most of the IgG structure moves in a highly coordinated fashion within its own domain. There are only a few less correlated parts in the Hinge region and also in Fab1 in comparison with Fab2.

Figure 2.9: The extent of correlation of residues locally, with the most correlated parts colored red, and the less correlated parts in blue. Local Effects, i.e. the contributions to a single residue's correlated motions from five neighboring residues on either side of that particular residue. Most of the IgG structure moves in a highly coordinated fashion with its attached domain. There are relatively few parts showing low correlations at the hinges.

The global contribution for modes 1 and 2 are depicted in Figure 2.10, where we observe which parts of the IgG structure tend to move in a more or less coordinated fashion with all other residues in the structure. As the individual contributions from all residues for a single residue (calculated by means of their pair-wise correlations and then shown in colors on the structure) mostly cancel out with each other, it is somewhat difficult to interpret these results, because it is not clear whether it is moving in more or less coordination with any particular part of the structure. But still
this information of such a global effect is useful because at least we know whether most of the residues are moving in high correlation with the structure globally.

In Figure 2.10, we can see a broad range of correlations (from highest to lowest) in both Fab’s (Fab1 and Fab2) and in the Fc in first and second normal mode, but still it is quite clear from the figure that there is less pairwise correlation in general for each residue with the entire structure, as compared with the local correlations shown in Figure 2.9. This demonstrates clearly that although most of the points in the structure are more or less moving in high correlation with their attached domains, nonetheless all domains generally have substantially independent motions that are not globally the same throughout the entire structure, and therefore we can claim that most parts of the IgG structure have some relative motions with respect to one another even though the domain motions are well shown explicitly with this type of analysis.

Figure 2.10: Global Effects, showing the extent to which a given residue is coordinated in its motions with the remainder of the whole structure. (red denotes highest contribution of correlations, while blue represents the lowest). The various domains of the IgG structure have some distinctive relative motions with respect to one another.
CHAPTER 3: ANTIGEN-BINDING EFFECTS

Antigen binding is the most important biologically related function of the immunoglobulins, and this takes place in the CDR regions of the IgG structure. Antigens generally vary in extreme ways, from some polypeptides to oligosaccharides to various small molecules, having extremely varied sizes. The CDR loop dynamics or motion likely helps them to bind to such a wide variety of different antigens. Understanding the effects and consequences of antigen binding on Fab dynamics helps to characterize the important motions of the immunoglobulin which are instrumental for antigen binding. By learning about the effects of antigen binding, we can begin to classify some of the important IgG motions that are critical for its function of antigen binding. This is important because it can shed some light on the characteristics or the effects of binding of small or large antigens and even regarding the different types of antigens binding to a single structure. This study has the potential to lead to new ways to facilitate drug design for a particular antigen. Here the main intention is to investigate whether the antigen binding CDR loops move in specific coordinated ways with respect to the other parts of the structure or if they have some distinct characteristic motions that are fully independent of the other parts of the IgG structure.

In section 3.1, we examine some of the similarities and differences between homologous Fab structures bound to antigen and without antigen. We investigate both structural and sequence comparisons among the Fab structures (with and without antigen bound to the CDR’s) and find the pair-wise distances (distances between all the residues).

In section 3.2, we examine the antigen binding effects on the lowest frequency modes of these two structures by comparing the important motions in the first few normal modes by employing elastic network models for the bound (antigen-bound)
and unbound homologous structures of Fab’s. We also show the mean square fluctuations for both the bound and unbound structures.

In section 3.3, we first see the sequence and structural similarities between the two Fab’s of IgG and the antigen bound Fab that we studied in section 3.1 and 3.2. Next we replace one or both of the Fab’s of the whole IgG with this corresponding antigen bound Fab and consider its motions while comparing the normal Fab of 1IGT and the antigen bound Fab.

3.1 Similarities and Differences between two homologous Fab structures (bound to antigen and without antigen)

It is difficult to find PDB structures for the entire structure of the IgG (i.e. containing both Fab’s and Fc in the same PDB id structure). One such full-length rare structure is 1IGT which does not have antigen bound to it. In order to determine the effects of antigen binding, we hence take two homologous PDB ids (reported by same group) of separate Fab’s, one with antigen bound (1CLY) and the other without antigen (1UCB). These are structures of only single Fab’s.

1CLY is the X-ray structure of an anti-tumor antibody, in complex with an antigen (58, 60). This is a structure taken at 2.8 Å resolution and is bound to the antigen -nonoate methyl ester derivative of Lewis Y (nLey) by the antibody BR96. This antigen is a carbohydrate (oligosachharide) which is bound in a large pocket formed by residues of all CDR loops except L2.

1UCB, on the other hand, is the X-ray structure of the uncomplexed human chimeric Fab of the anti-tumor antibody BR96 (59). This structure has a stated resolution of 2.6 Å.

Here we explore these structures to learn about their dynamics (or important motions) in the following.
3.1.1 Sequence Similarity between Complexed and Uncomplexed Anti-tumor Antibody BR96 Structures

3.1.1.1 Methods

We employ Clustal 2.1 Multiple sequence Alignment (74) to see the similarities in the sequences between these two homologous structures for both heavy and light chains.

3.1.1.2 Results

As shown below in figures (3.1 and 3.2), we find these two Fab’s have 100% identical sequences for both the light and heavy chains. They have been aligned in Figures 3.1 and 3.2.
Figure 3.1: Clustal Multiple Sequence Alignment for Heavy Chains in pdb structures 1CLY and 1UCB. The sequence identity is 100%.
3.1.2 Structural Similarity between Complexed and Uncomplexed Anti-tumor Antibody BR96 Structures

3.1.2.1 Methods

To learn about the structural similarities between these two Fab structures, we align them by employing the CE align algorithm in the Pymol software package (18). The CE align algorithm provides a fast and accurate structure alignment algorithm for protein structures which was pioneered by Drs. Shindyalov and Bourne (75).
3.1.2.2 Results

In Figure 3.3, these two aligned structures are shown (one bound to antigen and the other without antigen or uncomplexed).

The RMSD between these two structures is 1.19 Å over the 432 residues which is really quite small indicating that these two structures are very much similar. As can be seen in Figure 3.3 there is very substantial overlap between the two structures.

![Figure 3.3: Two Fab's are aligned by CE structural alignment algorithm. The green structure belongs to 1CLY which is the complexed one while the blue one is that of 1UCB which is the uncomplexed structure. The RMSD between these two structures is 1.19.](image)

3.1.3 Pair-wise Distances (Distances between the coordinates of all the corresponding residues in the two structures)

In figure 3.4, we show the pair-wise distances between all individual residues for these two structures with the deviations shown using the 1CLY structure to find which of the portions of the structure show the largest differences. The distances are
color-coded using the rainbow spectrum with red indicating the largest distance difference and at the other end of the spectrum blue indicating the smallest pair-wise deviance.

The CDR side loops show the largest difference with this approach. The cores of the Fab’s are substantially more similar as indicated by the colors in Figure 3.4.

Figure 3.4: Pair-wise Distances between the two Fab structures, with and without antigen bound (Red -> highest Pairwise Distance, Blue -> lowest Distance). The largest pair-wise distances are found in the CDR’s, with the core of the two Fab’s being very similar to one another.

3.2 Antigen Binding Effects on the Important Motions of the Two Fab Structures (antigen bound and uncomplexed)

Next, we characterize several important motions that are affected directly by the antigen binding. We investigate the overlap and the cumulative overlap for first few normal modes between the bound and unbound structures. If there is not much effect of antigen, we will observe high overlaps (approaching values of 1.0) along the
diagonal and zero or smaller overlaps in the other portions of the 2-dimensional overlap matrix. But in cases for significant differences we will see in the other parts of the matrix significant values, and the diagonal values will also be less than 1 revealing that there are some antigen effects. Next we plot the Mean Square Fluctuations on the structures. This will inform us about which parts of the structure have the largest differences between the bound and unbound forms.

3.2.1 Overlap and Cumulative Overlap in the Normal Modes between Bound and Unbound Structures

3.2.1.1 Methods

We compute the normal modes for the elastic network models of the two structures. A measure of the similarity between two given ANM modes $M_i$ and $M_j$ is given by their Overlap, which is defined as

$$O_{ij} = \frac{|M_i . M_j|}{||M_i|| ||M_j||}$$

(1)

The Cumulative Overlap (CO) for the first $k$ modes with a given mode $i$ is defined by

$$CO(k) = \left( \sum_{j=1}^{k} O_{ij}^2 \right)^{1/2}$$

(2)

This is a metric for how well the first $k$ modes together can capture the motion of a single mode $i$.

3.2.1.2 Results

Figure 3.5 shows the Overlap of the first 10 normal modes in the bound and unbound structures and Cumulative Overlap with some number of modes from the unbound structure with first 10 modes of the bound structure. It turns out that in the first and second modes, there is not much effect from the antigen. But the antigen effect is present mainly in the higher modes, and it is most prominent in modes 3 and 5 but also to a significant extent in mode 6. We therefore investigate the movies
for these three particular modes (3, 5 and 6) and we find that the antigen bound portion is the most mobile.

Hence we can conclude that there is a significant effect of antigen on the dynamics of IgG Fab structures and that the effect is more noticeable in the CDR regions to which the antigens bind.

Figure 3.5: Overlap and Cumulative Overlap for First few Normal Modes, comparing Antigen Bound (Model 1) and Free Fab (Model 2). There appear to be some significant antigen binding effects in modes 3, 5 and 6.

3.2.2 Enumeration of Mean Square Fluctuations

The Mean Square Fluctuations calculated from ANM are usually highly correlated with the experimental crystallographic B factors of the protein structures. These capture the extent of motions in 3-dimensional space.

In Figure 3.6, we compare the Mean Square Fluctuations (MSF) for the two Fab structures, one bound and another unbound, and we observe that there is an effect of antigen binding on the structure. It appears that there are some motions in the
vicinity of the antigen binding region in the CDR loops, whereas in the unbound structure, the motions are distributed more uniformly over the other portions of the structure, mainly on the surface parts rather than in the core region.

This result, also similar to our last result, strengthens our claim that the antigen binding has some effect on the FAB dynamics and that effect is most noticeable in the CDR region. Surprisingly the CDR loops show larger fluctuations when antigen is bound.

Figure 3.6: Mean Square Fluctuations of the FAB’s, with and without Antigen. There are larger motions in the antigen binding CDR loops region for the bound structure, whereas the motions are more broadly distributed over other surface portions of the structure.

3.3 Replacing the IgG Fab’s in the IgG (each individual Fab in Pdb ID: 1igt) with the antigen bound Fab (Pdb id: 1cly)

In this section, we first compare the sequence and look at the structural similarities of the two Fab’s of IgG (1igt) and also the antigen bound Fab structure 1cly and then replace each of the Fab’s of 1igt with 1cly in the entire structure. In this way, we try to learn about any differences in important motions within the entire IgG structure when it has original Fab and when it is transplanted with the antigen bound Fab 1cly.
We designate the Fab of 1igt structure containing chains A&B as Fab1 and the one composed of chains C & D as Fab2.

3.3.1 Structural Similarity between the Fab Structures Used

For measuring structural similarities, again we employ the CE algorithm in Pymol between these 3 pairs of Fab’s and notice small RMSD among these pairs.

Figure 3.7 shows the three pairs of CE aligned structures. The RMSD between the two Fab’s of 1igt (Fab1 and Fab2) is 1.99 Å whereas 1igt Fab2 (the one which contains chains C & D) is more structurally similar to 1cly with the RMSD being 1.33 Å, while the RMSD between 1igt Fab1 (which contains chains A & B) and 1cly is found to be 1.84 Å, which is still smaller than the RMSD between the two Fab’s of 1igt. This verifies the high similarity of the Fab that is transplanted.

Since we have strong structural similarities between the antigen bound Fab 1cly and the two original Fab’s of 1igt, we are justified in replacing the 1igt Fab structures with the antigen bound 1cly and then investigating the effects of this replacement on the motions of the entire structure of the modified 1igt. Replacing 1igt Fab2 is more conservative case for replacement than 1igt Fab1 since it has a higher structural similarity to 1cly.

This transplant approach to produce a chimeric structure is quite sensible since it is only a small change and is unlikely to cause any steric clashes or other significant problems in the resulting structure. If there were such problems these could have been resolved by atomic energy minimization or cycles of molecular dynamics equilibration. The fact that we are using coarse-grained structures also makes the
transplantation scheme less likely to be disruptive of structure.

Figure 3.7: Alignment of the Two Fab’s from pdb files 1IGT and 1CLY. Green, blue and purple denote 1IGT Fab1, Fab2 and 1CLY respectively. The three pairs of structures are structurally very similar to one another with the only relatively small differences given in the text.

3.3.2 Changes to the Important Motions for the Chimeric Structure

We replace both the Fab1 and Fab2 structures of 1igt with the 1cly structure and investigate the effects of these changes on the Normal Modes and Mean Square Fluctuations for the elastic network models, and show the results on the chimeric structure. We compute the Overlap, Cumulative Overlap and RMSIP of the normal modes between these two models (the first one is 1igt without its Fab replaced and the second one being 1igt structure with one of the Fab’s replaced by 1cly). We have already defined Overlap, Cumulative Overlap. RMSIP is the Root Mean Square Inner Product and it further captures the Overlap of some number of modes from both the models.
Figure 3.8: Overlap, CO, and RMSIP from ANM for Model 1 (1IGT unmodified) and Model 2 (chimeric 1IGT with Fab1 replaced). In modes 2, 3 and 4, the antigen binding effect is the most significant while causes the least changes to mode 6. The differences between the motions of the original 1IGT and the chimeric IgG with Fab replaced are significant.

In Figure 3.8, we see the effects of replacing 1igt Fab1 with 1cly in the entire structure of 1igt. We observe that for Modes 2, 3 and 4, there is a significant antigen binding effect, since these modes between the two models do not have a high overlap, whereas Mode 6 gives the highest overlap (95%). The cumulative and RMSIP similarly follow these results.

Figure 3.9 shows the similar results for 1IGT when Fab2 is replaced. Here we see high overlaps in some of the modes (Modes 1, 2, 5, 7, 8 and 9) where overlap is above 90%, but in the other modes there are more significant antigen binding effects.
Figure 3.9: Overlap, CO, RMSIP for the ANM comparing Model 1 (1IGT without Fab2 replaced) and Model 2 (chimeric 1IGT with Fab2 replaced). Modes 3, 4 and 6 show the largest antigen binding effects.
Figure 3.10: Highly Similar Mean Square Fluctuations are observed for A) 1IGT with Fab1 replaced with 1CLY, B) 1IGT with Fab2 replaced with 1CLY, C) 1IGT default structure with no Fab’s replaced. (MSF gradually decreases according to the spectrum with red being the highest MSF and blue the lowest). The results are substantially similar with only small differences for the individual cases.
In Figure 3.10, we show the mean square fluctuations on the structures to obtain information about antigen binding effects on the whole structure. Comparison between the MSFs shows that when Fab1 is replaced, the largest changes are to increase the motions in the Fc. Also there is a common behavior that we notice here that is important for our conclusion, but is not easy to observe in this figure. When one of the Fab’s is replaced with the antigen bound Fab, the motions tend to be diminished in the other Fab. Also in general, there is always more MSF in the CDR region of Fab1 than in other parts of the structure. In this structure the CDR region of Fab1 always exhibits somewhat higher fluctuations than does the CDR region of Fab2 for all the cases. It is unclear whether this is a general feature of IgG’s or whether this is an artifact of this individual crystal structure caused by asymmetric intermolecular interactions in the crystal.
CHAPTER 4: PRINCIPAL COMPONENT ANALYSIS OF FAB STRUCTURES

There are many structures of IgG Fab's in the PDB Database, most being crystal structures but with a small number of NMR, Electron Microscopy and other structures too. We explore the crystal Fab structures from the PDB to learn about the similarities and differences among these structures by performing Principal Component Analysis (PCA) on these Fab structures. Also we make comparisons to learn about the extent of correspondence between the Motions from Principal Component Analysis and from the Normal Modes of the Elastic Network Models.

The available numerous IgG Fab structures provide us with a highly significant ensemble of conformations of these structures in different conformational states, and this ensemble of conformations leads directly to information about their dynamics from their structures. This is an approach that has previously been pursued in the Jernigan lab on other structures (2). Also it may aid in informing about the relationship of various structures with respect to their structural details, such as, structures bound to some antigen or unbound, also whether certain loops are closed or open, RMSDs (root mean square distances) among the structures etc. Because there are a large number of such structures in the PDB, this analysis becomes complex at such a high-dimensional space. PCA helps us to reduce the dimensionality of the problem in order to comprehend the key motions for these structures and their dynamics (7, 9, 10, 14).

In section 4.1, we describe the selection of the structures. There are numerous structures of Fab in PDB, which provide the opportunity of being relatively selective, but we still have sufficient numbers of structures to map out the important conformational space. We select on the basis of length, chains, RMSD and some other factors described in this section.
Section 4.2 describes the resulting Principal Components, mainly the distribution of the structures along the first Three Principal Component Directions and also some higher order Principal Components.

Identifying the essential motions of structures by the analysis of multiple structures of the same protein should, in general, reveal key information about the dynamics of the protein. In Section 4.3, we first take a representative Fab structure and then learn about its dynamics by determining the first few normal modes by using an Elastic Network Model and then explore the correspondence between the normal modes obtained from ENMs for the representative structures and first few Principal Components taken from the selection of all Fab structures under consideration.

Section 4.4 describes Residue Positional Fluctuations along the first three PCs in our data set. We learn which residues or portions of the structures undergo the largest fluctuations along several Principal Component axes and from that learn about the significance of the individual Principal Components.

4.1 Selection of the IgG Fab Structures

Antibodies play an extremely important role in our immune system, especially for the adaptive immune system. There is a huge array of antibodies playing highly diverse roles in the immune system. So it is not surprising that there are many antibody Fab structures (more than 2000) reported in the PDB database. These structures vary significantly in terms of the lengths of their light and heavy chains, number of chains, and the structure determination method (X-ray, NMR, Electron Microscopy, etc.) by which their structures have been determined and various other parameters such as the solution conditions for crystallization.

As in the previous chapter, we mainly focus on the antigen bound structure (1cly) which is an anti-tumor antibody (BR96) to see its antigen-binding effects while one of the Fab structures of 1igt is replaced by it. Here, in the current chapter, we sequence blast the PDB database with this structure’s sequence. We screen the Fab structures from the PDB database based on three parameters: Method of structure
determination (X-ray Crystallography taken here), resolution of crystals (3 Å or better) and chain length (at least one chain containing 190-240 residues). The number of Fab structures found from PDB based on these parameters was 1166.

Next we eliminate some structures based on their RMSD (especially removed are the structures with RMSD higher than 8.0 Å over less than 200 residues). We also select only those Fab structures with two chains (one heavy chain and one light chain) and which had total size of the Fab within the range of 400-500 residues (for reasons explained in the next section (4.1.1)) and then the number of Fab structures is reduced to 333 which is the set we have finally settled upon for Principal Component Analysis.

In Figure 4.1, the ensemble of initially unaligned structures and also aligned structures (CE aligned with 1cly) is presented. In (A), 1166 Fab crystal structures (taken at resolution better than 3 Å and before CE alignment) have been shown. (B) are the ones with only two chains and after pair-wise alignment while (C) shows those further screened and aligned structures, based on the RMSD criterion and (D) is the initial representative structure set to which all the other Fab structures have been aligned.

Ideally we would like to obtain the coordinates of all these aligned structures for PCA in a multiple structure alignment, but we have settled on simply using pair-wise alignment since we could not immediately find a simple way to perform the multiple structure alignment on so many Fab structures.
4.1.1 Length Distribution of the Fab structures

The Fab’s in the pdb show huge variation in their sizes. This huge variety in size or chain lengths of Fab’s can be a result of different numbers of chains and individual chain lengths. By examining the distribution of Fab lengths, we find out whether there is a particular range of lengths for a large number of Fab’s or in another word, whether there is a strong distinguishable peak in the Fab length distribution.
In the following figures (Figure 4.2 and Figure 4.3), we show the Fab length distributions, where we identify how many maximum Fab structures fall within a range of Fab lengths. We observe two distinct peaks in the 400-500 and 800-1000 region. In Figure 4.3, this is even clearer. The first or largest peak indicates that the ordinary Fab length lies in the 400-500 range and a second peak occurs because of dimerization of the Fab’s. In Figure 4.3, we see that there are approximately 480 structures whose sizes fall in the range of 400-500 residues. For the purpose of PCA, we therefore choose the Fab structures having lengths within the 400-500 residue range.

Figure 4.2: Fab Length Distribution. Fab lengths are distributed from 200 residues to more than 3000 residues. However, most Fab structures are 400-500 residues in length.
Figure 4.3: Fab Size Distribution. This is a close-up of the data in Figure 4.2, and there are two main peaks, one in the 400-500 range and the other in the 800-1000 range.

4.1.2 RMSD Distribution of the Fab structures

The Root Mean Square Deviations (RMSD) is a significant factor for the selection of Fab’s for Principal Component Analysis, because, if these structures are too diverse in size, it will be impossible to reach any conclusion from the PCs about their conformational space and their dynamics since we would be comparing very different structures that would not be sensible. So, it is important to screen the Fab’s based on their RMSDs and to select those Fab’s with some limited RMSD values within the set.

From the initial 1166 structures, we finally settle on 333 Fab structures based on their length, RMSD, number of chains and RMSD values. In Figure 4.4, we show the RMSD distribution of the 333 Fab’s. As discussed in the last section, we perform
pair-wise alignment with 1cly instead of actually doing a multiple structure alignment, so all these RMSDs are the distances between the other 332 Fab’s and 1cly. The mean of the RMSDs is found to be 4.02 Å and the standard deviation is 1.44 Å. So, in a way, even after eliminating many structures based on their high RMSD values, the RMSDs of the remaining 333 structures are still relatively high. But we do not screen further based on stricter RMSD values since this might eliminate all but a very small number of Fab’s.

Figure 4.4: RMSD Distribution of Fab Structures in the set of 333 structures. Mean of the RMSD values is 4.02 Å and the Standard Deviation is 1.44 Å.

4.1.3 Multiple Structure Alignment of the Fab Structures

Next we perform Multiple Structure Alignment (MSA) on these 333 Fab structures by BioXGEM 3D-BLAST (76) to find out the regions in the structures where the Fab’s are most structurally aligned. This is important because in order to perform PCA on
the Fab structures, we need to cut some of them to all have equal lengths or extract exactly the same number of residues from the Fab’s. So it is worthwhile to do a MSA on the structures to find out the corresponding positions. There are some gaps, insertions, and deletions in various parts of the structures which make it difficult to extract the corresponding residues so they have equal lengths before performing the PCA.

One of the problems of using 3D-BLAST for MSA is that we cannot perform it on the full structure. Rather we have to settle with the corresponding chains. So here we do MSA on both the light and heavy chains.

Figures 4.5 and 4.6 illustrate the MSA results for the light and heavy chains of Fab’s respectively. As we can see in the following figures the structures are mostly aligned in the corresponding positions and we can remove the gaps for the purpose of our analyses.
Figure 4.5: Multiple Structure Alignment Results for Light Chains of Fab's.

Figure 4.6: Multiple Structure Alignment Results for Heavy Chains of Fab's.
4.2 Principal Component Analysis

Next we perform PCA on the structures on the 346 residues in each of the 333 structures, which are the corresponding residues found by 3D-BLAST MSA results. One of the main functions of PCA is for dimensionality reduction. It is shown elsewhere (2, 7-10) that first few PCs (Principal Components) explain a large fraction of the variance.

4.2.1 Methods

Principal Component Analysis (PCA) is performed on the X-ray data set of Fab’s taken from the PDB. The input is an n by p coordinate matrix X where n denotes the number of Fab structures in the data set (in our case 333) and p represents thrice the number of residues, i.e. all the coordinates of the 346 residues (i.e. 1038 for the Cartesian coordinates). The data used are the alpha carbon coordinates from each structure.

From X, the elements of the covariance matrix $C$ are computed as

$$c_{ij} = \langle (X_i - \langle X_i \rangle)(X_j - \langle X_j \rangle) \rangle,$$  \hspace{1cm} \text{------------------------------------------} \hspace{1cm} (1)$$

where the ensemble average includes the n structures and is indicated by the brackets $.\rangle$. The covariance matrix, $C$, is decomposed into its principal components as

$$C = P \Delta P^T,$$ \hspace{1cm} \text{------------------------------------------} \hspace{1cm} (2)$$

where P’s are the eigenvectors, the Principal Components (PCs) and the eigenvalues are the elements of the diagonal matrix, $\Delta$. The eigenvalues are sorted
in a descending order of importance. Each of the eigenvalues is directly proportional to the amount of the variance it captures in the corresponding PC.

4.2.2 Results

4.2.2.1 Dimensionality Reduction by PCA

Figure 4.7 shows the fraction of variances and the cumulative fraction of variances explained by first 10 PCs. It can be seen from the figure that first three PCs are extremely important and account for 42%, 14% and 9% of the variance respectively so that the first 10 PCs collectively explain 87% of the total variance, which is highly significant. So, it can be seen directly that PCA really performs a significant dimensionality reduction in this case so that the first ten PCs are able to capture the most significant fraction of variances in our data set.

Figure 4.7: Percent of Variance Explained by Principal Components for the Motions Apparent in a Set of 333 Fab’s. The percentage of variations explained by first 10 PCs collectively is 87%.
4.2.2.2 PCA Scatter Plots

The PCA scores can provide us with a simple overview of the distribution of all Fab structures in our data set. In general, the Scatter plots of multiple PCA scores demonstrate the distribution of the deviations of the actual structures from the characteristic structure plotted along the directions of the corresponding PCs. Ideally, again, these distributions can provide useful information about the types of motions apparent in the data set of Fab structures. Figure 4.8 shows the three-dimensional scatter plot of the Fab structures along the PC1, PC2 and PC3 directions. There can be found two distinct groups of structures in this plot along PC1 values -100 to 0 and +300 with both PC2 and PC3. Interestingly there are some distinctive clusters apparent in these distributions, and these can be seen most clearly in two of the parts of Figure 4.9.

![Figure 4.8: Distribution of Fab Structures along first three PC Directions. Two distinct clusters are found along PC1 in the range from -100 to 0 and in the region near +300 with each PC2 and PC3. See also Figure 4.9 for the projections of these data.](image)

We take the projections of this three dimensional scatter plot along each pair of first three PCs in Figure 4.9 and it turns out that even though there are two distinct clusters between each PC1 and PC2 and PC1 and PC3, the PC2-PC3 scatter plot
turns out to be completely messy with no distinct group of structures found in this plot.

![Figure 4.9: Distribution of Fab Structures along each pair of Directions in PC1, PC2 and PC3. Two distinct clusters are observable in both PC1-PC2 and PC1-PC3 Scatter Plots, whereas the PC2-PC3 plot shows a much broader scatter of the structures.](image)

The first three PCs explained around 65% of variances, whereas first 10 PCs cumulatively contributed to around 87%, so it seems to be worth-while to check the scatter plots in some higher dimensional PCs.

In Figure 4.10 are shown some scatter plots for some higher PCs. The three-dimensional PCA Scatter Plot along PC4, PC5 and PC6, unlike that for PC1, PC2 and PC3, indicates only one distinct cluster. Interestingly, PC1, as it did with both PC2 and PC3, again identified two clusters when it is considered together with PC4.
Figure 4.10: Distribution of Fab Structures along some Higher PCs. There is only one cluster found in the three dimensional scatter plot involving PC PC4, PC5 and PC6. The PC1 and PC4 plane scatter plot indicates two distinct clusters with relatively few outliers.
4.3 Correspondence between the Normal Modes obtained from ENM and first few Principal Components

The ensemble of structures, most certainly, provides snapshots of the potential motions of proteins. Hence, identifying the important motions by the analysis of multiple structures of the same protein invariably reveals some key information about the dynamics. The large number of available Fab structures in the protein data bank provides a remarkable sampling of conformations of the hugely diverse conformational states, which directly leads us to get certain structural information about the dynamics of the Immunoglobulin Fab structures.

The first few PCs cumulatively capture the majority of the total variance. Thus, the subspace spanned by these PCs can well reflect the most important and dominant motion space of the protein. Again as we have discussed in the earlier chapters, the first few normal modes obtained by Normal Mode Analysis of an Elastic Network Model, also represent the important motions of the protein. So, there must be a good correspondence between the lowest frequency normal modes and first few Principal Components.

Since we have done the pair-wise alignment of all other 332 Fab structures with anti-tumor antibody BR96 structure 1cly, we take that as the central structure since that lies in the center of RMSDs. Next we perform an ENM on this structure, get the first 10 normal modes and then investigate the correspondence between the first 10 normal modes of this representative structure 1cly with the first 3 PCs obtained from all the 333 Fab structures.
4.3.1 Methods

4.3.1.1 Overlaps between Normal Modes and PCs

We can find the alignment between the directions of a given normal mode and a given PC and it is a measure of their overlap. This is similar to the comparisons we have described previously for comparing normal models of two related structures. Tama and Sanejouand (4) first defined this term. Overlap is defined as

\[ O_{ij} = \frac{\lfloor P_i \cdot M_j \rfloor}{\|P_i\|\|M_j\|} \]  

(3)

Where \( O_{ij} \) is the overlap between the \( i^{th} \) PC \( P_i \) and the \( j^{th} \) normal mode \( M_j \).

The Cumulative Overlap (CO) between the first \( k \) normal modes and a given PC \( i \) is defined by

\[ CO(k) = (\sum_{j=1}^{k} O_{ij}^2)^{1/2} \]  

(4)

This is a measure of how well the first \( k \) modes together can capture the motion of a single PC \( i \).

The overlap between the motion spaces of the first \( K \) PCs and the first \( L \) normal modes is defined by the corresponding root mean-square inner product (RMSIP) (29), and it is given by

\[ RMSIP(K, L) = \left( \frac{1}{K} \sum_{k=1}^{K} \sum_{l=1}^{L} (P_k \cdot M_l)^2 \right)^{1/2} \]  

(5)

Where \( RMSIP(K, L) \) is the RMSIP between the \( k^{th} \) PC \( P_k \) and the \( l^{th} \) normal mode \( M_l \).

The RMSIP indicates how well the motion space spanned by first \( k \) PCs is correlated with the first \( l \) lowest frequency modes. Intuitively, the RMSIP between the two spaces (PC and normal modes) measures the percentage of the PC subspace that
is covered by the subspace spanned by the selected lowest frequency normal modes.

4.3.2 Results

In figure 4.11, the overlap, cumulative overlap and RMSIP between the normal modes and the Principal Components are presented. The results show relatively small overlaps, indicating that there is not a high correspondence between these two spaces. This is possibly because of the high RMSD differences among the Fab structures that have been utilized for the PCA.

It appears that PC3 has the highest overlap (21%) with the second normal mode and it also exhibits some other relatively moderate overlaps (like 18%, 17% etc) with some other normal modes. For PC1, the highest overlap happens to be with normal mode 5 (16%) and for PC2 the largest overlap is with mode 4 (again 16%). The cumulative overlap is highest (50%) between PC1 and first 44 normal modes, whereas the RMSIP, which directly follows from the cumulative overlap, is also highest between PC1 and the 44 lowest-frequency normal modes.
Figure 4.11: Overlap, CO and RMSIP between the Normal Modes and PCs. Overlap is highest (21%) between mode 2 and PC3, whereas the cumulative overlap and RMSIP are highest (50%) between PC1 and the 44 lowest-frequency modes.

4.4 Residue Positional Fluctuations of the first three PCs

Since most of the protein displacements can be captured by only a few PCs in terms of the variances of the structures, these PCs can characterize the dominant dynamic behaviors of the proteins. In this section, we are interested to see where the largest conformational displacements or fluctuations are located in the Fab structures for the first three PCs; we call these Residue Positional Fluctuations.
In Figure 4.12, the PC variations or Residue Positional fluctuations have large amplitude in the first half of the Fab’s, i.e. the light chains in case of PC1 and PC2, whereas for PC3, the PC variations are larger for the second half of the Fab’s, i.e. the heavy chains. These are most likely hinge types of motions.
Figure 4.12: Residue Positional Fluctuations for the first three PCs. For PC1 and PC2, there are larger amplitudes of fluctuations within the first half of the Fab's, i.e. in the light chains, while for PC3, the second half of the Fab has larger motions, i.e. in the heavy chain.

In Figure 4.13, we have displayed the pdb names of a few of the Fab's in a PC1-PC2 Scatter Plot. As clearly shown in the figure, there are two distinct clusters across PC1 values, centered near -50 and 300. By identifying the structures as we have displayed in the below figure, we can characterize, identify and visualize the significance of the individual Principal Components. This could be a part of future studies.

Figure 4.13: The Location of Specific Fab PDB Structures in the PC1-PC2 space.
CHAPTER 5: SUGAR EFFECTS

The glycosylation of the IgG structure binds the carbohydrates covalently to the Fc region of the structure. Carbohydrates are considered to be significant for the secretion of Ig molecules, recognition of polymeric antigens and creation of IgG-IgG complexes (61-69). In the literature, researchers have shown that this sugar plays a vital role in complementary activation associated with cell lysis and inflammation. High-mannose glycans in the Fc region of therapeutic IgG antibodies help to increase serum clearance to a certain extent (69). Aggregation happens to be one of the main problems in any protein-based therapeutics and is also related to immunogenicity concerns and decreased efficacy of therapeutic agents. Dynamic fluctuations of interactions between proteins and carbohydrates promote protein aggregation. The significance of intramolecular interactions between aromatic residues and carbohydrates covalently bound in the glycosylated domain has been specifically explored through computer simulations, fluorescence analysis and site-directed mutagenesis. In Reference (68), they reported that chicken IgY binds its receptor at the CH3/CH4 interface in a similar way as in the human IgA – Fc alpha RI interactions. They demonstrated the significance of N-linked carbohydrates by the failure of the CHIR-AB1 interaction after mutation of the glycosylation site.

In this chapter, we will computationally investigate the sugar effects on the IgG structure dynamics. We want to examine this effect of carbohydrates on the protein structure computationally by modeling the sugar. In section 5.1, we discuss various models to explore the sugar effect on the IgG.

Section 5.2 illustrates the effects of sugar on the important motions of the IgG. We explore these effects with various metrics such as Overlap of modes, Mean Square Fluctuations shown on the structure, the Mode Vectors or Shape Vectors, temperature factors, etc.
In Section 5.3, we probe the structure by relocating the sugar at various positions to explore its possible effects on the dynamics of the structure.

### 5.1 Various Models to Explain Sugar and Major Hinge Effects

We investigate the sugar effects computationally by modeling the IgG structure with an attached carbohydrate by utilizing a mixed coarse grain Elastic Network Model where we use an all atomic representation of the carbohydrate or sugar and coarse-grain the other parts of the IgG structure, i.e. represent them by only their alpha carbon atoms. There are two reasons for this. First, the sugar is very small compared to the much larger IgG structure, and the sugar does not have any obvious representative atom in each residue. Second, we are interested in examining the sugar effects and so the atomic details may matter.

In the following Figure (5.1), we show three models. The top left one is the one in which the sugar is represented in the atomic scale, whereas the other parts of the IgG have been modeled in the alpha carbon coarse-grained representation. Generally, we use 12 or 13Å cutoff for the alpha-carbon coarse-grain model and 4 or 5Å cutoff for the atomic model, since atoms are much closer to one another than the alpha carbons in each residue. The top right one represents the model in which the major hinge (the portion which links both the Fab’s to Fc) is represented by the atomic model, whereas the other parts have been coarse-grained by the alpha carbon representation. There is no sugar attached to Fc in this model. In the third or the bottom model, there are different levels of coarse-graining or mixed coarse-graining in the IgG structure. The major hinge and the sugar have been represented by all atomic model while the residues of other parts are represented by alpha carbon atoms. In this model, we often use 13 Å cutoff for the alpha carbon coarse-graining and two different cutoffs which are lower than 7 Å for both sugar and major hinge regions.

These different models are used to compare different parameters like mean square fluctuations, mode vectors etc plotted on the corresponding structures to understand the effect of carbohydrates.
5.2 Sugar Effects on the Major Motions the of IgG

To examine the sugar effects on the IgG (1igt) structure, we compare the different models described in the last section (5.1). The metric or parameters compared between the different models are the Mean Square Fluctuations (MSF), the overlap across modes, Mode Vectors or Shape Vectors displayed on the structures.
5.2.1 Mean Square Fluctuations (MSF)

In Figure 5.2, we show the computed MSF on two different models of IgG structures, the first model having atomic cutoff in the hinge region at 4 Å and all other parts having alpha carbon coarse-graining cutoff at 13 Å, while the second model has exactly these two cutoffs for hinge and other parts, but it has one additional atomic cutoff at 5 Å for the sugar bound to the Fc region.

Because these models employ different numbers of representative atoms, the number of eigenvectors calculated from the Elastic Network Models is also different for the two different models. So, for comparison, for the larger structure we take a smaller subset of vectors and perform Gram-Schmidt re-orthogonalization for the corresponding vectors in order to get the corrected subset of vectors. Gram-Schmidt process (77) is a method for orthogonalizing a set of vectors in an inner product space, most commonly the Euclidean space $R^n$.

The MSF, which is a representation of the motions, are slightly lower (in Figure 5.2) in Fc region for the sugar bound structure, so the sugar effect seems to be more localized and not to show a very extensive or allosteric effect.
Figure 5.2: The sugar effect for the hinge and sugar atomic models. The MSFs have been plotted on the IgG structures, both having the hinge represented by all atoms and one having the sugars represented by atoms, while the other has no sugar. The MSFs are color coded in the rainbow spectrum with the red representing the highest fluctuations and blue indicating the lowest. The motions are smaller in the Fc region for the sugar bound structure.

In Figure 5.3, however, the hinge is not represented by atoms, but rather as for the other parts of the IgG structure, it is represented by alpha carbon atoms (though the hinge cut off is 10 Å here while the other part alpha carbon coarse-grain cutoff is 13 Å). All other parameters remain the same, so the sugar is represented with an atomic cutoff of 5 Å.

Here, no noticeable difference is found in the MSFs in the Fc region between the two models, except for slightly larger motions in the Fab2 (the Fab containing chains C and D) in case of the sugar bound structure.

Figure 5.3: Sugar Effects on the IgG Motions for the Hinge coarse-grained, and Atomic Sugar models; Sugar is present in the right model. There is no noticeable difference in the motions of the Fc region between the two different structures.

In order to learn about the hinge effects, we plot the MSFs (in Figure 5.4) on the two structures, one with the hinge represented as atoms with the atomic cut-off of 5 Å.
and the other with the hinge coarse-grained as alpha carbon atoms with the same coarse-grain cutoff (13 Å) as for the remaining part of the structure. In the right figure of 5.4, i.e. in case of atomic hinge representation, the motions are reduced in the Fab’s and increased considerably within the Fc region. This is because while hinges are represented by atomic cutoff of 5 Å, there are more springs in the ENM model which probably does not allow propagation of the motions from the Fc region to the Fab’s.

![Figure 5.4: Hinge Effects in the IgG. The motions in the Fc are considerably larger and significantly lower in the Fab’s for the model representing the hinge with atoms than for the IgG with coarse-grained hinges.](image)

5.2.2 Overlap across Normal Modes

In order to learn about the sugar effects, we consider two models, one (model 1) bound with sugar and the other (model 2) without sugar, and then consider the overlap between the corresponding normal modes in these two structures, as we have done in one earlier chapter to see the effects of antigen. Here also, since we have to take an equal length eigenvector for evaluating the mode overlaps, we perform the Gram-Schmidt reorthogonalization in order to get the more comparable eigenvectors.
In Figure 5.5, we learn about some sugar effects in modes 2 and 3 since the overlap is somewhat smaller in the cases of these two modes, while the overlap is considerably higher (96-97%) for most other modes.

Figure 5.5: Overlap across Modes (Sugar Effect) for the models shown in Figure 5.2. Model 1 is without sugar, while Model 2 has sugar. The difference between the two models is largest for modes 2 and 3.

5.2.3 Correspondence between ANM Temperature Factors and X-ray B Factors

As we know, the temperature factors calculated from ANM usually show high correlations with those from X-ray crystallization for reliable ENM models. Below in Figure (5.6), Model 1 has no sugar and Model 2 includes sugar. And we immediately find that there is not much difference in the correspondence between
experimental B Factors and ANM temperature factors, and in fact these correspondences are quite similar.

Figure 5.6: Sugar Effect on B factors and ANM Temperature Factors. The correspondence between ANM temperature factors and experimental B factors is similar in case of the two models (one with sugar bound and another without sugar).
5.2.4 Shape Vectors and Mode Vectors

The eigenvectors calculated from NMA with an ENM represent the mode shapes of the system for that motion. The mode shape vectors are not the absolute motions, rather they just describe the relative motions of each point in the structure. Here we use Shape Vectors as a metric to compare between the two models (sugar bound and not bound).

In Figure 5.7, the corresponding Shape vectors across the x, y, and z directions are shown for both the sugar bound model and not bound model. In mode 1, in the sugar bound model, there are smaller fluctuations along the x axis in Fab2 and Fc, with the fluctuations being similar along the y axis between the two models while across the z axis, the fluctuations are larger for the Fab1 in the sugar bound model. In mode 2, differences between these two models are not very significant.
Figure 5.7: Comparisons of Mode Shapes in the x, y, and z directions due to Sugar Effect. In mode 1, in the bound structure there are smaller fluctuations along the x axis in the Fab2 and the Fc than the unbound structure, while along the y axis, the fluctuations are similar between two structures. Along the z direction, there are larger fluctuations in the Fab1 for the bound structure. For mode 2, the differences are smaller.

We also represent the mode vectors by arrows and plot these with the structures. This arrow representation of mode vectors can be used to visualize the direction of motions between two specific states (as for example, targeted Molecular Dynamics, comparing between open and closed structures).

Figure 5.8 demonstrates the mode vectors between the sugar bound and unbound structures for first two normal modes. The top and bottom set of figures represent modes 1 and 2 respectively, while the left and right set of figures show the arrows for mode vectors for bound and unbound structures respectively. For mode 1, there are larger motions in the CDR side of Fab2 (the Fab containing chains C and D) for the sugar bound structure, whereas more motions are noticed in the core side of Fab2 for the unbound structure. For mode 2, however, the differences in the Fc are very significant between the bound and uncomplexed structures, since in the bound structure, there is much less motion in the Fc compared to that in the unbound structure. Thus it confirms our earlier claim of localized sugar effects in mode 2.
Figure 5.8: Mode Vectors, with and without Sugar. (Top: Mode 1, Bottom: Mode 2, Left: With Sugar, Right: Without Sugar). In mode 1, the motions are more localized in the CDR side of the Fab2 of the bound structure and less in the core, while the opposite can be seen in the uncomplexed structure. In mode 2, the motions are significantly smaller in the Fc region for the bound structure compared to the Fc region in the unbound structure.

5.3 Exploring the Relocation of Sugar at Different Positions in the Fc

The sugar is bound covalently in the Fc region of the IgG structure. In order to find the location effect, to learn whether the particular placement of the sugar within the Fc region is important, we considered different models of the sugar bound to the IgG structure where the sugar is bound at the different positions in the Fc of IgG.

In Figure 5.9, we present three such models where the sugar is bound in the lower portion of the Fc (below its original position), at a higher portion in Fc (top right figure of 5.9) and towards right side of the Fc (bottom figure).
Figure 5.9: Different Sugar Positions. Sugar is bound in the lower portion of the Fc (top left), in the upper portion of the Fc (top right), towards the right side of the Fc (bottom figure).

We compare the ANM correlation plots among these three different models for mode 1 in Figure 5.10, the first one with sugar (top left), the second one with sugar at the upper portion of the Fc (top right), in the third, sugar is placed in the lower portion of the Fc (bottom).

It seems that sugar at the top has some apparent effect on the correlation of Fc, which moves as a more correlated domain than in the other three models. The possible reason for this might be that the sugar at the top is near to the hinge and might be restricting its motions to propagate between the Fc and the Fab's, so Fc shows more coordinated motion in case of this structure where sugar is bound at the upper half of Fc, higher than its original position in 1igt structure.
Figure 5.10: Effects of Relocating Sugar in the Fc 1 on Model 1. With Sugar as in Pdb id 1IGT (top left), 2. Sugar at the upper portion of Fc (top right), 3. Sugar at the lower portion of Fc (bottom). Sugar bound at the upper portion of the Fc appears to have some significant effects. It leads to more cohesive motions within the Fc structure for this mode.
CHAPTER 6: CONCLUSIONS AND DISCUSSIONS

Antibodies play an important role in the immune system. In the current study, we have explored the structures and dynamics of antibodies. Our main purpose here has been to comprehend and recognize the important Immunoglobulin domain motions and CDR motions related to its functioning by employing Normal Mode Analysis of Elastic Network Models, which is a proven and well-established method for extracting the most important global motions from the structures. The CDR regions are the parts where antigens bind. By examining the first few normal modes associated with the CDR motions, we investigated and characterized the motions of the CDRs and how they correspond to the global motions of the entire IgG structure. Also we have examined the effects and consequences of antigen binding on the motions. We wanted to verify this by describing the motions apparent among experimental structures, bound and not bound with antigens. There are many Fab structures in the protein data bank. We performed a Principal Component Analysis on the ensemble of experimental Fab structures to extract information about the immunoglobulin dynamics that can be revealed by identifying the apparent motions present in multiple structures of the same protein. Last we examined the sugar effects on the IgG dynamics.

This thesis has primarily considered four topics, namely, Immunoglobulin CDR motions, antigen binding effects on the IgG structure, motions apparent in the Fab structures and the sugar effects on the dynamics.

The individual Ig Fold domain motions can be captured in the first few normal modes of the elastic network models. We also found that CDRs mainly move in high correlation with their attached domains. The inter-domain correlation between CDR1 and CDR2 gradually diminishes as we go from lower modes to higher modes and in some higher modes, the CDRs move in an anti-correlated fashion with respect to one another. Also the well-known major and minor hinges are detected in this kind of
analysis. In the Internal Distance Changes calculation, we have found that the changes in internal distances are larger in Fab1 and Fab2 compared to those in the Fc region. This is in overall agreement with the point of view that the Fab binding to antigens is facilitated by the motions within the Fab’s.

One of the most significant findings of this study is that there is a significant counter-intuitive effect of antigen-binding. We find that the portions of Fab bound with antigen become most mobile and when we try to replace the original Fab in the IgG with an antigen-bound Fab, it appears that the antigen binding suppresses motions in other parts of the structure, primarily in the opposing Fab that does not have antigen bound. In the Principal Component Analysis study, we have seen moderate correlations between the first few PCs and the computed normal modes. Overall the PCs provide a comprehensive representation of the total conformational space.

The effect of the natural sugars on the dynamics appears to be quite localized and serves to make the Fc domain more coherent in its motions, but there is no apparent longer range effect. This localized effect of sugar shows some agreement with reported effects of the sugar impacting the local motions of the Fc, which may be related to effector functions such as fixation of complement, binding to mast cells, macrophages, and natural killer cells.

Overall the IgG is a complex structure and its dynamics are likewise complex, and we have observed effects from binding of sugars and antigen, with some being expected and some being counter-intuitive.
REFERENCES


70. Thomas J. Kindt , Barbara A. Osborne , Richard A. Goldsby ; Kuby Immunology, 6th edition.


74. http://www.ebi.ac.uk/Tools/msa/clustalw2/
