HINGE PREDICTION USING ELASTIC NETWORK MODELS

by

Uğur Emekli

B.S.,Chemical Engineering, Boğaziçi University, 2004

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Chemical Engineering

Boğaziçi University

2006

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my thesis advisor Prof. Dr. Türkan Haliloğlu for her continous guidance, invaluable contributions and confidence in me. She has always been an inspiration for me along my undergraduate and graduate education.

I would also like to present my special thanks to Dina Schneidman and Prof. Dr. Ruth Nussinov for their support in my thesis work and my appreciation to Prof. Dr. Pemra Doruker for her contribution to enrichment of my knowledge and perspective.

Thanks to my colleagues in Polymer Research Center, Aslı Ertekin, Özge Kürkçüoğlu, Nevra Özer, Nigar Kantarcı and Sinem Özel, for all their helps and sharing their knowledge.

Finally I want to thank to my parents for their endless support...

ABSTRACT

HINGE PREDICTION USING ELASTIC NETWORK MODELS

The motion and interaction of rigid parts in a protein connected by hinges are important for the function of the protein. This makes identification of the hinge region particularly important. In this work a method and a web server, HingeProt, which makes use of this method are developed and presented. The method employs elastic network (EN) models for predicting the rigid parts and the flexible hinge regions connecting them in the native topology of protein chain. The elastic network models used in the method are Gaussian and Anisotropic network models (GNM and ANM, respectively). GNM calculates the mean-square fluctuations and the correlation between the fluctuations of residues in the most dominant (slowest two) modes. These correlations suggest hinge regions and the cooperation between them, whereas ANM provides the direction of the fluctuations of residues in the corresponding modes. The method requires a 3D structure of a single protein chain or multiple chains. Using this static conformation, a list of the rigid parts, a list of connecting hinge residues, short flexible loops and the predicted fluctuation of the residues, in the slowest two modes are calculated. The efficiency and validity of the method are tested by a method devised on two different data sets. The method is highly efficient and successful in providing the results within seconds for small size proteins and within several minutes for proteins consisting of a few thousands of residues. It can be used toward applications including protein function, flexible protein-protein and protein-ligand docking, flexible docking of protein structures. The server can be reached through http://www.prc.boun.edu.tr/.

ÖZET

ELASTİK AĞ YAPI MODELLERİ KULLANILARAK PROTEİNLERİN BAĞLANMA NOKTALARININ TAHMİNİ

Protein hareketleri ve proteinlerin esnek olmayan kısımlarının destek noktalarından kırılarak birbirleriyle ve diğer proteinlerle etkileşmesi bir proteinin işlevini yerine getirebilmesi açısından önemlidir. Bu durum, proteinlerin esnek olmayan kısımlarının ve destek noktalarının belirlenmesini daha önemli kılar. Bu çalışmada, bu amacı gerçekleştirmek üzere bir yöntem ve bu yöntemi kullanan bir internet ara birimi hazırlandı. Bu yöntemde elastik ağ yapı modelleri olan Gaussian Network Model (GNM) ve Anisotropic Network Model (ANM) kullanılarak esnek olmayan kısımlar ve bunları birleştiren destek noktaları tahmin edilmeye çalışıldı. Bu yaklaşımda GNM metodu protein yapı taşlarının ortalama hareketlerinin boyunu ve birbirleriyle olan korelasyonunu hesaplamak için, ANM ise hareketlerin yönünü belirlemek için kullanıldı. Yöntem herhangi bir proteinin statik 3D yapısından yola çıkarak, bir dizi esnek olmayan bölge ve destek noktaları listesi ve proteinin tahmin edilen hareketini vermektedir. Geliştirilen yöntemin geçerliliği ve doğruluğu iki farklı veri kümesi kullanılarak test edildi. Bu test sonucunda yöntem küçük proteinler için bir kaç saniyede büyük proteinler için bir kaç dakika içinde başarılı ve doğru sonuçlar verdi. Bu yöntemin proteinlerin işlevlerinin belirlenmesinde, protein-protein ve protein-ligand etkileşimlerinin tahmininde ve protein kenetlenme çalışmalarında faydalı olması amaçlanmaktadır. Bu yöntemin web ara birimine http://www.prc.boun.edu.tr/ sayfasındaki kısa yol yardımıyla ulaşılabilir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

k_b	Boltzman constant
R_i	Position vector of i^{th} residue
H	Hessian matrix
H_{ij}	$(i,j)_{th}$ superelement of Hessian Matrix
ΔR_i	Displacement vector for R_i
ΔR_{ij}	The magnitude of the vector $\Delta R_i - \Delta R_j$
T	Temperature
u_i	i^{th} eigenvector
V_{tot}	Total potential
α	Alpha helix
β	Beta strand
γ	Force Constant
Γ	Kirchhoff - connectivity matrix
λ_i	i^{th} eigenvector
3D	Three dimensional
ANM	Anisotropic Network Model
EM	Electron Microscopy
GNM	Gaussian Network Model
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank/ Protein structure deposited in PDB
RMSD	Root mean square deviation

1. INTRODUCTION

Proteins are complex organic compounds that are formed by linkage of smaller units, called aminoacids, via peptide bonds. Each one has a specific aminoacid sequence and distinct function. Hence all proteins are unique in the sense that they are adapted for specific task. This task is determined by the motion of the protein dictated by its structure. The structure of the protein is determined by its aminoacid sequence.

The study of macromolecular motions is important for understanding and controlling the mechanisms of biomolecular interactions and functions. Once the motion is identified in all aspects, many ways to favor or inhibit this motion can be explored. Hence there are numerous applications for the identification of protein motion.

Computational prediction of 3D associations of the proteins without any prior knowledge of the interaction is a very complicated task[1]. Even the simplest cases are known to have failed to converge to a correct solution in simulations. Predicted protein motions can be used to generate a set of protein conformations for such studies to enhance convergence[2, 3]. Alternatively, regions identified to have cooperative motions can be identified as rigid parts to be used in computational methods [4, 5]. Another application is refinement of the low-resolution EM structures by extraction of functionally important structural motions [6, 7, 8] from the structural fluctuations in the principal directions.

Several approaches are available for the identification of molecular flexibility. The approaches can be divided into two major groups [9]. The first group identifies structural flexibility through analysis of two or more different conformations of the same or homologous proteins [10, 11]. On the other hand, methods belonging to the second group start from a single protein structure. These methods are not limited by the availability of different conformations. Molecular Dynamics (MD) and Monte-Carlo (MC) are classical representatives of the methods in this group. However, these methods are computationally expensive and therefore not capable of capturing large-scale

protein rearrangements. Other very efficient approaches attempt to estimate protein rigidity and flexibility using graph [12] or constraint theory [13]. Methods based on the Normal Mode Analysis (NMA) fall in between the accurate computationally expensive MD/MC methods and the less accurate, however very efficient graph theoretic approaches.

In this work a method, based on Normal Mode Analysis, is devised for identifying the location and motion of the rigid parts, pivoted by the hinge locations, and also the motion exhibited by the rigid parts using the the native topology of protein chain. Two elastic network models are used in the method: Gaussian and Anisotropic network models (GNM and ANM, respectively). The efficiency and validity of the method are tested by a procedure (explained in detail) on two different data sets. Also a web server which employs the method, is built. The web server is expected to useful in a range of applications including protein function determination, flexible protein-protein and protein-ligand docking and flexible docking of protein structures.

In comparison to this work, there are a number of available NMA and hinge prediction methods and servers: Among those, iGNM [14], having the same theoretical background with HingeProt, provides an extensive dynamic mode analysis for macromolecular monomeric and complex structures using the GNM. Other available servers such as elNemo [15] and WEBnm [16] carry out the normal mode analysis of structures to predict the mobility or to generate possible conformations. There are also available online databases and web servers to predict hinges such as Database of Macromolecular Movements [17], FlexProt [18] or DynDom [19] , which require two conformations to predict hinge sites. Among these, the present method particularly focuses on the identification of hinge sites, rigid segments, flexible loops, and the direction of the fluctuations of residues in a structure using both the GNM, which is similar to iGNM, and ANM, which is similar to normal mode servers by providing the direction of the motions.

2. PROTEIN STRUCTURE : GENERAL INFORMATION

Proteins are polymers formed by linkage of 20 different kinds of monomers, i.e. aminoacids. Size of a protein may vary from 50 aminoacids to thousands depending on the function and location of the protein. Each aminoacid has a central atom (C^{α}) to which are attached a hydrogen atom, an amino group (NH_2), a carboxyl group (COOH) and a side chain. According to the complexity of this side chain each aminoacid acquires certain chemical properties such as hydrophobicity, polarity, electrical charge etc. The size of a side chain also brings in some stereochemical properties to the aminoacid.

Table 2.1: The aminoacid types

NAME	Symbol 1	Symbol 2	NAME	Symbol 1	Symbol 2
Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Aspargine	Asn	N	Methionine	Met	M
Aspartic Acid	Asp	D	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamine	Gln	Q	Serine	Ser	S
Glutamic Acid	Glu	E	Threonine	Thr	T
Glycine	Gly	G	Tryptophan	Trp	W
Histidine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V

The primary structure of the protein is the aminoacid sequence of 20 different kinds of aminoacids. The aminoacid sequence causes the protein to shape according to chemical and physical properties of each aminoacid in the sequence. This local geometry of the protein is called the secondary structure of the protein. The major determinant of the protein function and the protein folding is the secondary structure. Two well known secondary structure types which arise from local hydrogen bondings are α-helices and β-strands. α-helix is a righthanded coiled conformation and looks like

a spring. Each aminoacid in the chain corresponds to 100^o turn. What gives the α-helix its shape is the H-bonding between the amino (NH_2) group of an aminoacid and the carboxyl group of the four earlier aminoacid. The length of an α-helix structure may vary from four to 50 but typically it is about 10 aminoacid length. β-strand on the other hand, is fully extended conformation of aminoacids, forming a regular stretch. A β-strand is about five to 10 aminoacids in length. Beta strands at different locations of the proteins may form hydrogen bonds if they come close in a right orientation. Then the resulting structure is β-sheet which consists of a number of parallel beta strands. Below are given the representations for α-helix and β-strand. Further reorganization and folding of the protein structure is represented by tertiary structure.

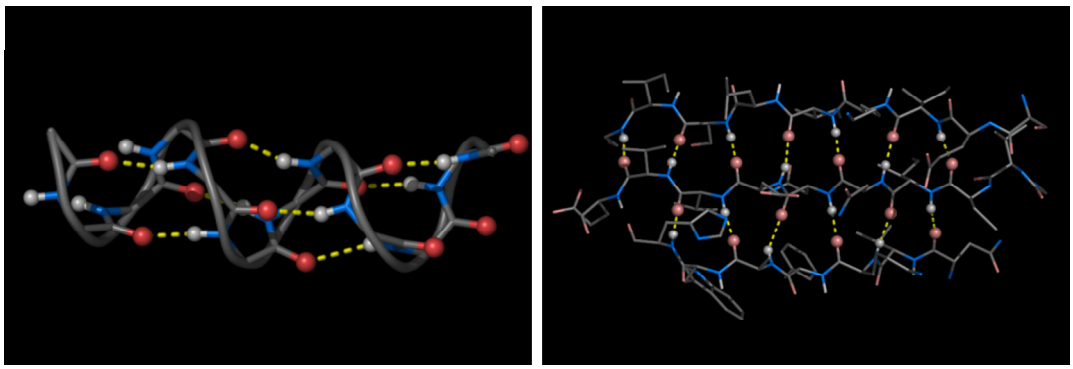

Figure 2.1: Secondary structure motifs (a)α-helix (b)β-strand.

The secondary and tertiary structure of the protein are the consequences of the interactions of the aminoacid side chains. They represent the thermodynamically stable static state of the protein which is called conformation. The proteins in solution are in continuous motion, colliding and interacting with other structures or complexes in the solution. Those collisions, if they occur at the correct locations with enough energy to exceed the barrier, result in new bond formation and/or bond breakage. The function of the protein is determined by such events. Hence the function is directly related to collisions and indirectly related to secondary and tertiary structure of the protein which enhance those collisions.

The conformational changes may take place before or after a protein structure accomplishes its function. Those changes may be local or may be encountered by

extensive regions of the protein structure, i.e. domains. Key-Lock model is an example where no conformational change is observed. Alternatively, domain motions, motion of a domain in relation to another, may occur in the form of "hinge motion" or "shear motion". In hinge motion domains exhibit a large motion pivoted by the connecting regions of different domains. On the other hand in shear motion the movement of domains are relatively smaller and no pivot is available. Among those, this study mainly focuses on the Hinge Motions of the protein.

Hinge motion is characterized by large changes in main-chain torsional angles occurring at a localized region, which is called a hinge. Hinge motions may involve a small number of residues, since even one bond can provide the required rotational freedom. This kind of protein motion is free of packing constraints. When a chain exhibits hinge motion at the region connecting two structural domains, each domain behaves as a rigid body and packing interactions can appear/disappear between the interfaces of those rigid bodies. Hinge motions usually occur upon binding to another molecule, or upon activation/deactivation of the protein.

One of the most interesting examples is calmodulin. Calmodulin is a calcium-binding protein that can bind to and regulate different protein targets. Therefore its function is crucial for the cellular functions. Upon binding to its ligands, there is large-scale movement of calmodulin involving splitting of one long helix. The total rotation of one domain relative to the other is upwards of 150^{O}.

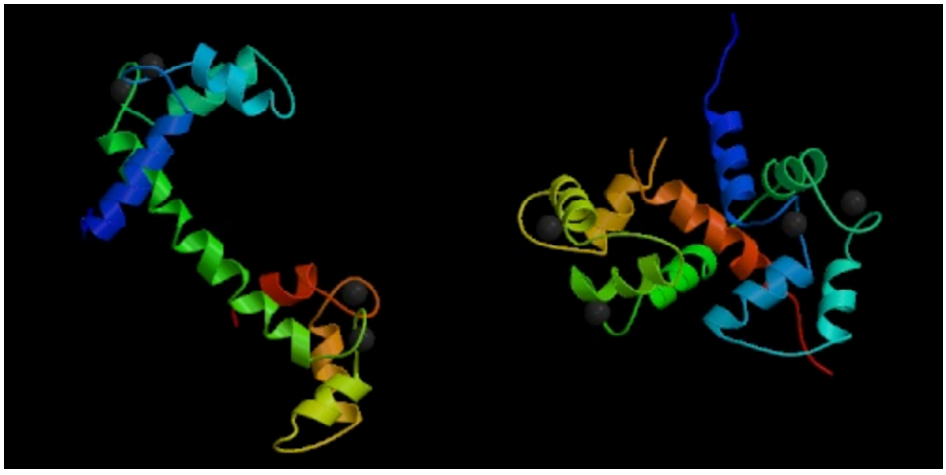

Figure 2.2: Calmodulin protein before binding (4cln) and after binding (2bbm)

Note that four letter codes in the parentheses in the figure caption stand for the protein code. Each protein has a unique code which makes it easier to identify and search in the large databases. Once the 3D structure of a protein is determined by the means of experimental methods such as X-Ray crystallography or NMR, the related information is deposited to a database called Protein Databank [20]. It is given a four letter code starting with a number for identification which is called the PDB code.

When Figure 2.2 is inspected, it can be observed that the change in protein conformation upon binding is dramatic for this protein. The protein structure encounters extensive domain motions and rigid parts, labeled with different color codes, change in the orientation in 3D space. If bending locations of this motion can be predicted with coarse-grained methods before it is already determined with time consuming and exhaustive methods, the results can be used to save time and effort.

Therefore the hinge regions are the mechanistically informative regions of the structure and are of importance in mediating cooperative motions that have functional importance. The method devised in this work focuses on the prediction of the rigid parts, the hinge regions and also motion suggested by those hinges using a single static conformation of a protein structure.

3. THE MODEL AND THE METHODS

3.1. General Outlook: Normal Mode Analysis

In protein dynamics, the conformational mechanics defined by the equilibrium native structure plays a significant role. NMA [21, 22] is a unique tool for providing an analytical solution for the modes of motion accessible to given structure in its global minimum. With the computationally simpler coarse grained elastic network (EN) models, NMA has recently shown its success for predicting the intrinsic collective motions. The dynamics have importance for identifying residues that are critical for function, folding and refining experimental data from macromolecular structures [23, 24, 25]. It has been shown that known protein motions can be predicted by perturbing the original structures along the direction of the two slowest frequency normal modes [26]. This suggests that associated motions overlap with those smoothest ascent directions (lowest-energy modes) [26]. It may further suggest that the structures may have evolved in a way that facilitates their biological functions through the use of their intrinsic elastic modes. Hence in this work the rigid parts of the protein structure and the flexible joints (the hinge regions) are identified from the equilibrium fluctuations of the structure in a given topology by the elastic network models, the Gaussian network model (GNM) [27] and the Anisotropic network model (ANM) [28].

The hinge regions are the mechanistically informative regions of the structure and are of importance in mediating cooperative motions that have functional importance. We use the GNM to predict the mean-square fluctuations of the residues and the correlation between them and the ANM to describe the direction of the fluctuations of the residues. With these findings, this method is expected to be useful in a range of potential applications such as functional mechanisms of macromolecular structures, and assemblies, flexible docking in resolving protein-protein association and in refining the complex structures, and fitting flexible hinge-bent protein structures into EM density maps and refining the EM structures.

3.2. Models - GNM and ANM

GNM [25, 27] and its extension ANM [28] are coarse-grained residue level elastic models. GNM predicts the relative magnitudes of the fluctuations, whereas ANM predicts the directionality of the collective motions in addition to their magnitudes. Earlier studies have shown that the GNM results are more robust, and thus are preferentially used for evaluating mean-square fluctuations of residues in low frequency modes [28].

In the Gaussian Network Model (GNM) [25, 27], each aminoacid in the protein is assumed to exhibit fluctuations from its mean positions which we observe in its static conformation (its mean positions are defined by their conformation deposited in the PDB file). Those fluctuations can be modeled by Gaussian Distribution. To simplify the model each residue -aminoacid-, is represented by a single mass center, a node, at its C^{α} and assumed to be connected to its neighboring nodes, which are defined as "at a distance smaller than a cut-off distance", by elastic springs. Each spring has a uniform force constant so a perfectly elastic network model is formed. The total potential energy is the sum of harmonic potential energies in the springs of the neighboring centers.

$$V_{tot} = \sum_i \sum_j V(R_i, R_j) = \sum_i \sum_j \frac{1}{2}\gamma(\Delta R_{ij}.\Delta R_{ij}) \tag{3.1}$$

where γ is the Hookean force constant between interacting centers, R_i is the position vector of i^{th} residue, R_{ij} is the fluctuation in the distance vector $R_{ij} = R_j - R_i$ from the equilibrium position. By rearranging the equation and defining Kirchhoff connectivity matrix, the equation becomes:

$$V_{tot} = \frac{\gamma}{2} tr[\Delta R^T \Gamma \Delta R] \tag{3.2}$$

The Kirchhoff connectivity matrix, Γ, is defined as n×n matrix for a protein consisting of n residues, in which the diagonal holds the number of neighbors, which

is called the coordination number, and ij^{th} element which equals -1 if i^{th} residue is in contact with j^{th} residue, 0 otherwise. Note that the sum of the elements for all rows and columns add up to zero in this matrix. ΔR is the matrix consisting of distances.

When this potential function is diagonalized and solved, a number of eigenvalues and for each eigenvalue a fluctuation vector is obtained. Each element of each vector represents the extent of fluctuation for the corresponding aminoacid. Eigenvalues signify the frequency of the motion. So since the number of eigenvalues is equal to the number of residues, a number of fluctuation vectors (n) of different frequency will be reached. Although n eigenvectors are available only (n-1) eigenvectors will be valid since the smallest eigenvalue will be trivial, equal to zero, due to the degrees of freedom of the system, which is one. The vectors with highest eigenvalue represents the local high frequency motions and the vectors with lowest non-trivial eigenvalue represents the global cooperative motions of the protein. Then it can be deduced that the domain motions of proteins, if available, are represented best by the slowest modes of motion.

Once the decomposition is carried out, the degree of cooperation between fluctuations of two residues, in specifically one mode of motion or a number of modes of motion, i.e. cross-correlation, can be calculated using the equation given below. Note that if this equation is used for only one mode of motion then the correlation coefficient will equal to 1.0 or -1.0 , whereas if a number of modes are considered, the values vary between -1.0 and 1.0. This equation is especially useful in the slowest modes of motion because the most cooperative motion of the protein are observed in those modes.

The correlation between fluctuations of i^th and j^th residue, ΔR_i and ΔR_j respectively, are given as:

$$< \Delta R_i \Delta R_j >= (\frac{3k_BT}{\gamma})[\Gamma^{-1}] = \sum_{k=1}^{n} (\frac{3k_BT}{\gamma}) k [\lambda_k^{-1} u_k u_k^T]_{ij} \tag{3.3}$$

In this equation, λ_k is the k^{th} eigenvalue, u_k is the eigenvector associated with this eigenvalue, Γ is the Kirchhoff matrix and n is the total number of modes subject

to consideration.

Experiences have shown that GNM is a very robust model in calculating the extent of the fluctuations however it does not provide any information on the direction of the movement. The direction of motion is quite important in predicting the protein motion and the function resulted by this motion. Another aim of this work is to provide predicted conformations for detailed dynamic studies and simulations of proteins. So GNM is not adequate in that sense despite its robustness.

To predict the direction of movement another elastic network model, ANM [28] which is a simplified version of NMA can be used. This method also uses the static conformation of protein structures based on the same assumptions of GNM such as all residues in contact, residues that are closer to each other than a cut-off distance, are connected with springs of a force constant, γ, and no other interaction between the residues exist. Although this assumption may seem too strong, since it does not account bonds, studies have shown that it can produce satisfactory results in protein dynamics [28]. In this method the potential function of Gaussian Network Model is replaced by an anisotropic model where the direction of the movement is taken into account by treating x,y and z cartesian coordinates separately. With this method the following equation for the potential function is obtained.

$$V_{tot} = \frac{\gamma}{2} \sum_i \sum_j h(r_{cut} - R_{ij})(\Delta R_j - \Delta R_i)^2 \tag{3.4}$$

where ΔR_i is the fluctuation vector of the position vector R_i, R_{ij} is the distance between the i^{th} and j^{th} centers, h the heaviside step function - 1 if i^{th} and j^{th} residues are in contact, that is $(r_{cut} - R_{ij}) > 0$, and 0 otherwise- and γ is the force constant of the imaginary spring.

For the solution of this equation, the second derivative of the potential function must be calculated and this introduces the Hessian matrix in the solution. The Hessian matrix may be thought as the ANM replacement of Kirchhoff matrix in GNM. The

form of the Hessian matrix is given below.

$$\mathbf{H} = \begin{bmatrix} H_{11} & H_{12} & \dots & H_{1N} \\ H21 & \dots & & H_{2N} \\ \vdots & & & \\ H_{N1} & \dots & & H_{NN} \end{bmatrix}$$

Each element of the Hessian matrix represents a 3×3 matrix such that

$$H_{ij} = \begin{bmatrix} \frac{\partial^2 V}{\partial X_i \partial X_j} & \frac{\partial^2 V}{\partial X_i \partial Y_j} & \frac{\partial^2 V}{\partial X_i \partial Z_j} \\ \frac{\partial^2 V}{\partial Y_i \partial X_j} & \frac{\partial^2 V}{\partial Y_i \partial Y_j} & \frac{\partial^2 V}{\partial Y_i \partial Z_j} \\ \frac{\partial^2 V}{\partial Z_i \partial X_j} & \frac{\partial^2 V}{\partial Z_i \partial Y_j} & \frac{\partial^2 V}{\partial Z_i \partial Z_j} \end{bmatrix}$$

When the derivatives are taken, then the resulting matrix is diagonalized and the inverse is taken to solve for ΔR. The solution obtained is quite similar to the GNM solution except instead of inverse of Kirchhoff matrix, (Γ^{-1}), the inverse of the Hessian matrix, (H^{-1}) occurs in the solution. The inverse of the Hessian matrix contains $(3N-6)$ eigenvectors for $(3N-6)$ non trivial eigenvalues. Notice that six eigenvalues are zero since the degrees of freedom for the set of the equations are six. Each element of each vector represents the extent of fluctuation for each axis. Therefore when the extent of fluctuations for three coordinate axes are gathered up, the direction of the movement is obtained. Again eigenvalues signify the frequency of the motion and will be sorted with their corresponding eigenvectors to obtain the slowest modes of motion (low frequency high amplitude motion) and the fastest modes of motion (high frequency and low amplitude motion). The calculation of cross correlation for the ANM is also similar.

$$< \Delta R_i \Delta R_j >= (\frac{3k_B T}{\gamma}) tr[H^{-1}] = (\frac{3k_B T}{\gamma}) \sum_{k=1}^{n} [\lambda_k^{-1} u_k u_k^T]_{ij} \tag{3.5}$$

where $tr[H^{-1}]$ is the sum of the diagonal elements of 3 x 3 small matrices in H^{-1}.

Consequently with the use of GNM and ANM together, theoretically it is possible to determine the rigid parts and the corresponding direction of movement for those rigid parts. The method is able to suggest both hinge locations and possible conformations. Here it should be noted that cutoff distance is an important parameter for the method. The Gaussian Network Model studies have shown that the best results are obtained when the cutoff distance is 10Å so this value is taken by default. However on the web server this value may be altered. With N-1 modes of motion, Γ yields N dimensional eigenvectors u_k, the components of which give the motions of N residues along the k^{th} mode with the corresponding eigenvalue γ_k. In ANM [28], Kirchhoff matrix, Γ, is replaced by Hessian matrix, H, with a larger cutoff distance ($< 18\mathring{A}$) again adjustable over the server. H gives 3N-6 eigenvectors u_k with each in 3N dimensional, the components of which describe the motions of N residues along x, y and z directions in the k^{th} mode. The knowledge of fluctuation vectors allows us to construct and explicitly view pairs of alternative conformations sampled by the individual modes, simply by adding the fluctuation vectors ΔR_i in the respective modes to the equilibrium position vectors. Here, specifically the fluctuations in the slowest two modes are considered. The slowest modes are the most cooperative global modes of motions that are related to functional motions in a structure.

3.3. Prediction of sequential rigid segments and hinges using GNM

In this work, GNM is used to calculate the mean-square fluctuations and the correlation between the fluctuations of residues, and ANM to generate the conformations that describe the fluctuations of residues from the average in the principal directions of motion.

First the GNM decomposes the fluctuations of N residues of a structure into a series of N modes, given the Cartesian coordinates of its α-carbon coordinates. N-1 GNM nonzero modes are organized in ascending order such as $\lambda_1 < \lambda_2 < \ldots < \lambda_{N-1}$. The eigenvectors corresponding to the slowest first and seconds modes, $u_k = 1$ and 2, are extracted. The square of these vectors (Equation 3.3 with i=j for k=1 and 2) describes the mean-square fluctuations (the autocorrelations) of residues from

equilibrium positions along the k^{th}, principal coordinate (first and second modes here). u_k gives the shape of the k^{th} mode as a function of residue index. The local minima of those mode shapes point to the flexible joints of the structure, i.e. the hinge regions, which connect the rigid units and mobile loops. These hinge regions mediate the correlation between these structural units and can thus be easily identified by the values of the correlation between the fluctuations of residues in a given mode or a set of modes (Equation 3.3). The cross-correlations in the slowest first and second mode (Equation 3.3 with i,j for k=1 and 2) are considered in conformity with the mean-square fluctuations. Since the GNM is a one dimensional model in its description of the fluctuations, the change in the sign of the correlation values between - and + points to a flexible joint that connects the rigid structural units. We use the change of sign for identifying the rigid segments and hinges.

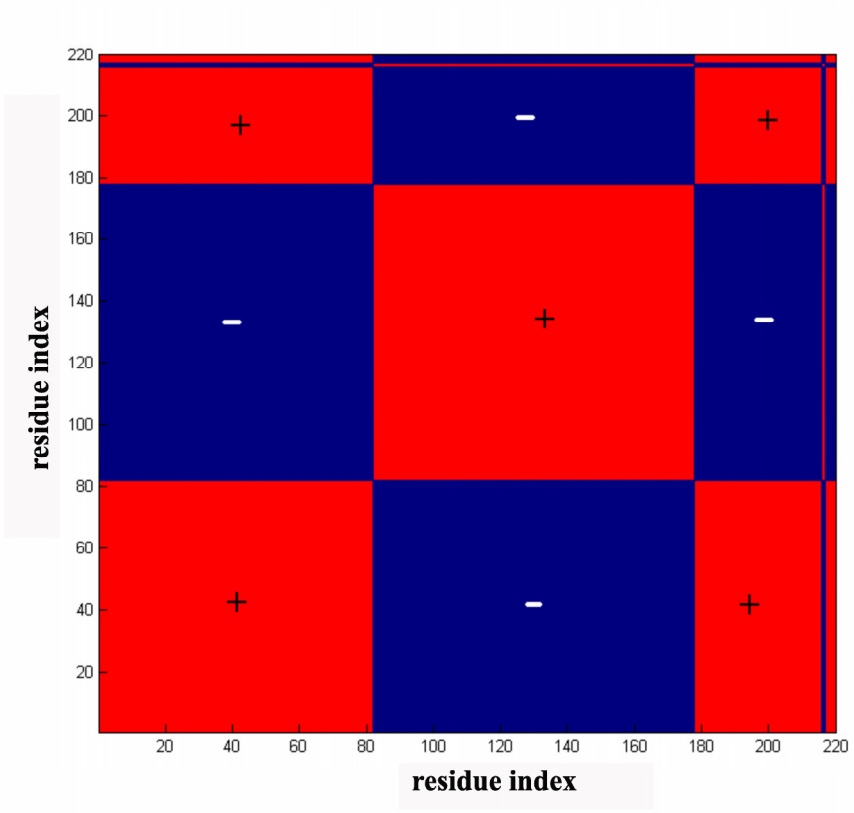

Figure 3.1: Cross Correlation graph for the slowest mode of the Glutamin binding protein

Figure 3.1 is the graphical representation of the cross correlation data. Each

axis corresponds to residue numbers. The graph is always symmetric by definition, if there is a vertical line along cutting the x axis at the i^{th} aminoacid, then there is another horizontal line cutting the y axis at this aminoacid. This aminoacid by definition equals where the change of sign occurs. The change on the sign of a residue represented in this figure, as mentioned, refers to the cooperation of the residues. Each set of residues restricted by two consecutive lines in the figure are said to move cooperatively. Once this information is known, a computational procedure to extract the initial rigid segments may be formed. Those are named as initial segments because they need to be treated afterwards and coupled if required. The need for this will be explained in the following sections. So the initial segmentation was the first step of this study.

3.4. Mode Mapping and Extracting the direction of the movement

The GNM fluctuations are isotropic by definition and thus the directions of the fluctuations are characterized by ANM. ANM with 3N-6 eigenvectors of u_k describes the fluctuations of N residues in the x, y and z directions from the average structure (X-ray or NMR) according to k^{th} mode. So once the rigid parts are determined, the ANM mode corresponding to the slowest two GNM modes which are already used for rigid part and hinge determination, can be used to calculate the motion of the protein.

GNM and ANM use different models and matrix solutions to provide answer despite the fact that they are based on the same theory. The difference arises because rearrangement of the vector indexes may differ slightly in the sorted version in two methods. So the first method of the GNM may not correspond to the first mode of ANM. Fortunately the difference is not dramatic or frequent, only small shifts occur. So it is possible to determine the ANM modes which correspond to the slowest two modes of GNM, in a couple of slowest modes most of the time.

Mapping of the modes are carried out based on the physical definitions of the GNM and ANM. GNM calculates the mean square fluctuations for each residue and ANM calculates the direction of the movement for each residue. GNM outputs a vector

length and ANM outputs the vector itself. Although they are in different extents the behavior of ANM vector lengths along the protein structure must be similar for the GNM output. However for direct comparison, the extent of the average length must be brought to the same level so a normalization is required. All fluctuation vectors of the modes which are considered for mapping then need to be normalized as the case in the method. The slowest mode fluctuation vector of the GNM is taken, it is normalized so that the sum of all fluctuation extents add up to one, in order to find its corresponding ANM mode. Then the fluctuation vectors of the slowest 10 modes of motion for ANM are taken. For each displacement vector of one mode, fluctuation extents are calculated from the displacement vectors and all fluctuation extent vectors for ANM are also normalized so that the sum of the elements of the vector add up to one. By this way 10 ANM fluctuation vectors and two GNM fluctuation vectors of the same order are obtained. Then the mapping of each GNM mode can be carried out by comparing the differences in the fluctuation extents. Below is given an example.

In Figure 3.2 it can be observed that the location of peaks and hinges overlap for the second graph, which includes the third mode of ANM whereas for the first figure shows that the graphical behavior of the ANM's first mode is different than the GNM's first mode. This correspondence then signifies that the GNM mode is represented by the motion that is obtained from the third ANM mode. In short, for this case,the rigid parts and hinges are calculated with the slowest mode of GNM and direction of movement for each residue are obtained from the third slowest mode of ANM.

After mapping the ANM modes to GNM modes by comparing the square fluctuations, between the resulting modes in the two models, the directions of the fluctuations of residues in the slowest first and second modes of GNM are obtained by the ANM analysis. ANM provides an extent of motion for each cartesian coordinate axis for each residue while GNM was able to provide only an extent for the vector consisting of x,y and z cartesian coordinates. With this, it has been possible to give a predicted orientation for the fluctuations of residues in the most dominant two modes. As the fluctuations are symmetric with respect to the equilibrium positions, ANM-predicted deformed structures could readily be obtained by adding and subtracting the fluctu-

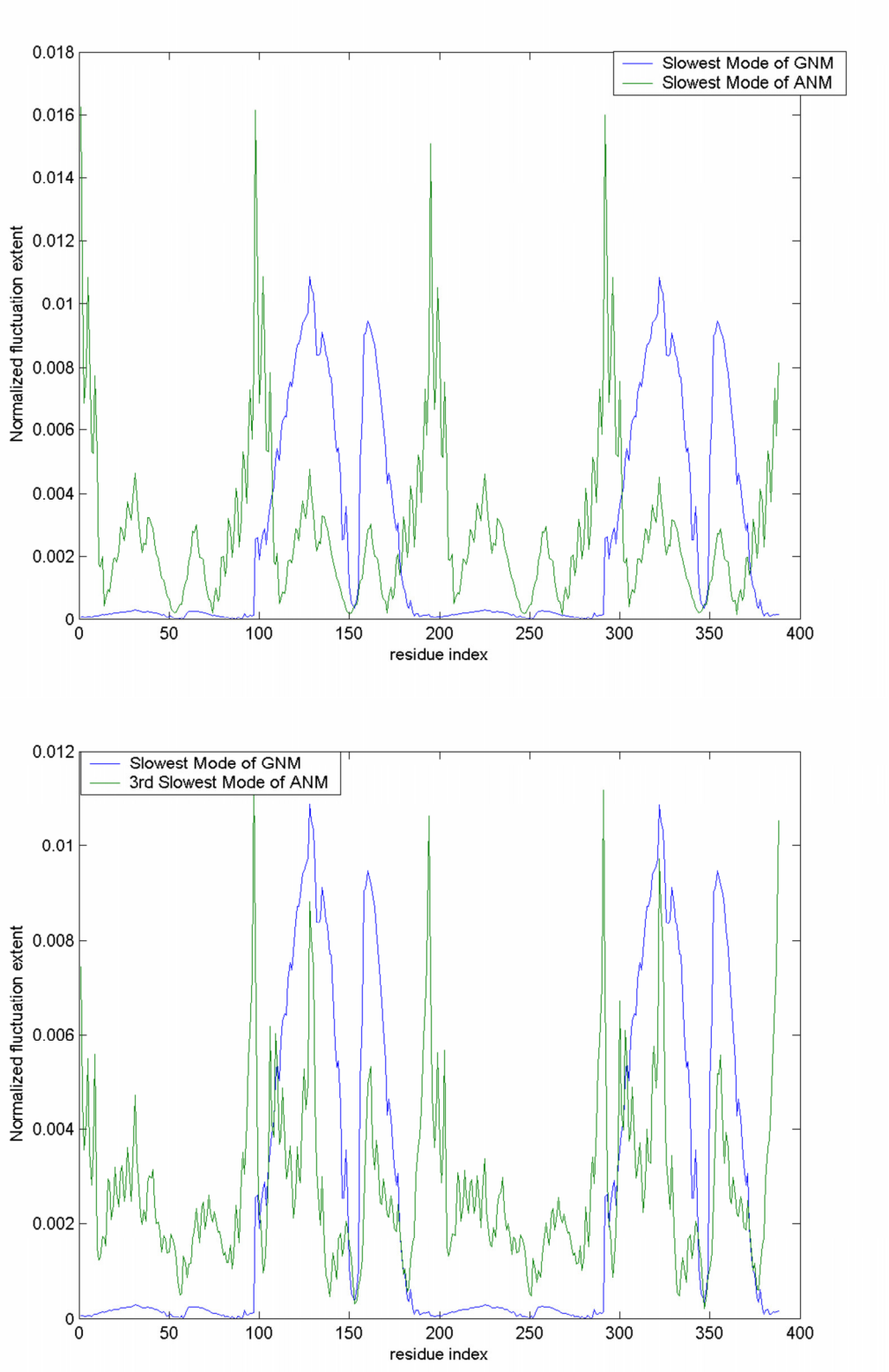

Figure 3.2: The extent of fluctuations for 1bl8 protein calculated by GNM and ANM (a) The slowest GNM mode is compared with the slowest ANM mode (b) The slowest GNM mode is compared with the third slowest ANM mode

ations of each residue to/from its equilibrium position. It should be noted that the absolute magnitudes of the fluctuation vectors can not be predicted by GNM or ANM. The extent of fluctuation vectors depends on the force constant, γ, in the potential equation. Thus, the extents of deformations from the average structure are meaningful and relative if they are compared between the residues/structural units of the same protein in a given mode.

3.5. Spatial segment clustering

When the protein chain is divided into initial sequential rigid segments where the sign of the correlation changes from + to -, there may be some very small segments which can not be considered as rigid bodies. Besides the change of sign represents only the change in cooperative behavior of neighboring atom but does not say anything about the cooperation of the first and the last residue. That is to say, among the nonconsecutive initial segments there might be some cooperating ones or they all might be irrelevant. This bears the need of an algorithm to determine the cross correlation between the rigid parts this time.

An algorithm is also devised to solve this problem. This algorithm may be explained as follows: For example, the (i,j) segment starts with residue i and ends with residue j. Rigid segments consisting of less than 15 residues are considered as short flexible fragments instead of rigid part, and merged with their neighboring rigid parts. Since a domain or domain motion can not be defined by small segments, this procedure was required to classify domains. The short flexible segments are also recorded as a separate list.

Then CAST clustering method [29] is applied to the initial fragments. The following graph is constructed: each rigid segment is a node. Two non-sequential rigid segments (i, j) and (i', j') are connected by an edge, if their endpoints are within 12Åfrom one another, i.e. the distance between the C^{α} atoms of residue i and residue i' and the distance between the C^{α} atoms of residue j and residue j' is less than 12Å. The segments in the opposite direction are also attempted to be joined, by checking the

distance between the C^{α} atoms of residue i and residue j' and the distance between the C^{α} atoms of residue j and residue i'. If one of the endpoints is the beginning or the end of the protein chain, this endpoint is not checked. Each cluster produced by the CAST algorithm represents one rigid part and may include several initial rigid segments. These rigid parts are the output of the method devised here.

4. PRELIMINARY STUDIES ON THE METHOD

Before building up an automated procedure to extract all the information mentioned in the theory and a web server, some preliminary studies are carried out to determine roughly whether the method is promising. For this purpose 21 cases are collected from Database of Macromolecular Movements [17] and FlexProt [18] and their results are compared with the results of the method devised here at each step of the procedure.

The first step was to calculate the fluctuations of residues (α-carbons) by GNM for the set of structure obtained. The potential hinges were identified from the distribution of the fluctuations in the slowest modes and the results are compared with those predicted by FlexProt [18] and given in Database of Macromolecular Movements [17]. Then a more elaborated analysis on a few structures chosen from the above set by both GNM and ANM were done as case studies. In that several points were tried to be addressed regarding the potential hinges: the most dominant hinge, the hinges that operate cooperatively -the hinges that appear in the same mode-, the hinges that has dependence on the others nearby -a hinge may have contribution from the residues that are sequentially distant but close in space- and the direction of the fluctuations. Additionally, more examples were provided so as to display the predicted conformations that describe fluctuations in the most dominant modes of motion by ANM.

4.1. Prediction of hinges by GNM

The minimum fluctuating residues, the hinge residues, in the slowest first mode (Slow1), in the second slowest mode (Slow2) and in the average of the slowest two modes (Slow12) are identified by GNM cross correlation analysis [25, 27] for 21 proteins. The correlations between the fluctuations in the respective modes also pointed to the same regions of the structure that mediate the hinge behavior. The comparative analysis of the hinge regions in the two slowest modes and in the average of these two modes implies possible ranks of the identified hinges. If for example one hinge location is also

Table 4.1: Detailed results of the preliminary study on the data set

PDB CODE	size	SLOW1	SLOW2	SLOW12	DB1	Match	DB2	Match
1K9K	89	48	**(27, 68)**	**(27, 68)**	59-74	**S2, S12**	50 63	**S1** **N/A**
1JKN	165	**(17, 47,106)**, 78	**(21,72,110)**,47, 78, 100, 135	**(17,47,106)**,78, 135	94-117	**S1, S2, S12**	87 106	**N/A** **S1, S12**
1BJZ	194	164	61, 102, 164, 185	61, 102, 164	94-110 123-140	**S2, S12** **N/A**	164	**S1, S2, S12**
1HFV	164	**(22, 41, 153)**, 87, 119	**(22, 153)**, 87	**(22, 41, 153)**, 87, 119	73-95	**S1, S2, S12**	41 153	**S1, S12** **N/A**
2A8V	118	50	9, 30, **(53,96)** , **(65, 78)**	**(50, 96)**	40-59	**S1,S2,S12**	73	**N/A**
1CLL	144	77	23, 101	77	59-82	**S1,S12**	77	**S1,S12**
1ADJ	420	321	**(58, 261)** ,161, 329	201, 321	**Not Available**		296-320	**S1,S12**
1GGG	220	**(82, 180)**	20, 37, 88, 151, 190	**(82, 180)**			87 180	**N/A** **S1,S12**
1DAN	132	80	38, 93	47, 86			78-84	**S1**
1TCR	236	113	**(20, 76)**, **(153,202)**, **(142,202)**, 88 103, 122, 169, 182	**(204, 224)**, 113, 149			116-120	**N/A**
1TOP	162	36, 88	72, 120	88			36 70 88	**S1** **S2** **S12**
1UKE	193	**(11, 116)**, 30, 125, 168	**(11, 116)**, 50, 63, 168	**(11, 116)**, 168			110-120 135	**S1, S2, S12** **N/A**
3FRU	269	169	8, 25, **(35, 51)** , **(93, 114, 119)** , **(183, 202, 210, 226, 239)**, 257, 263	169			87 165	**N/A** **S1, S12**
2AK3	226	20, **(118, 196)**, 175	20, 38, 116, 191	38, 116, 191			120 160	**S1** **N/A**
1A67	108	21, **(39, 56, 87)**	5, 34	**(39, 56, 87)**	48-70	**S1, S12**	**Not Available**	
1A03	89	27	**(19,69)**	27,69	21-44	**S1, S12**		
2RAN	320	87, 242	71, 118, 237	89, 242	227-249	**S1, S2, S12**		
3KTQ	536	156, 268, 286, 308	176, 249, 319, **(430, 464)**	156, 268, 308	170-185	**S2**		
6Q21	171	**(8, 57, 79)**, 21, 26	5, 21, 31, 51, 79, 108, 118	21, **(57, 79)**, 108, 148	66-77	**S1, S2, S12**		
4ICB	73	**(23, 28)**	11, 58	23, 58	40-58	**S2, S12**		
2PRG	271	29, 83, 22, 173	50, 88, 122, 151, 229	88, 122, 173	76-104	**S1, S2, S12**		

observable in the average of two modes, this hinge location is said to be more dominant with respect to a hinge location which can be observed in one of the modes but not in the second.

Table 4.1 includes the results of the analysis of these 21 structures, 14 and 13 of them being compared with the predictions by FlexProt [18] and those given in Database of Molecular Movements [17], respectively. The results for the six of the cases are compared with the predictions of both in the latter. The first column gives the PDB identity for each structure, the second column gives the number of residues in the corresponding structure and the following three columns identify the index of the hinge locations determined according to GNM cross correlation study for the first

slowest mode, the second slowest mode and the weighted average of the slowest two modes respectively. The bold entries in parentheses identify the hinge locations which are in close vicinity (10 neighbourhood in space) to each other. Such groups of hinge locations are assumed to behave cooperatively since they are very close in space and treated as one hinge location.

Table 4.2: Overall results for the overlap of the hinges

	GNM / GERSTEIN	GNM / FLEXPROT
OVERALL	93% (13/14)	61% (14/23)
SLOW1	64% (9/14)	52% (12/23)
SLOW2	71% (10/14)	13% (3/23)
SLOW12	86% (12/14)	39% (9/23)

In table 4.2, DB1 refers to the Database of Macromolecular Movements [17] and DB2 refers to the FlexProt [18] hinge predictions for the same structure. If those predictions of DB1 and DB2 match with that of GNM results for the first and the second slowest modes, then the corresponding matching mode is written on the match column for each database. So it has been able to determine the success of the such prediction based on the results obtained from the available databases. However it should be noted that the validity of this success measure is limited with the validity of the results obtained from the databases and no method is available up to now providing predictions with 100 per cent success. The results of the comparison may be summarized as follows.

Table 4.2 gives an idea about the extent of the agreement between the predictions by GNM [25, 27] and those by FlexProt [18] and given in Database of Macromolecular Movements [17]. 13 of 14 hinge regions defined by the data in Database of Macromolecular Movements [17] covers at least one of the hinges determined by GNM. This hinge is mostly assessed as the main hinge observed in Slow12. On the other hand, the comparison of the predictions by GNM [25, 27] with those of FlexProt [18] seems to be relatively less correlated with consensus in 14 of 23 cases. This might be possibly due to the specific assignment of one or two residues as a hinge point in most of the cases

by FlexProt [18] rather than a region.

4.2. Case Studies by GNM/ANM

Protein structures from Table 4.1 are presented in detail in this section. For each structure, the mean-square fluctuations in the slowest modes (Slow1, Slow2 and Slow12), the correlation between the fluctuations (Cross1, Cross2, Cross12) in the respective modes by GNM [25, 27], and the conformations that describe the fluctuations from the average structure in the slowest first two modes by ANM [28] are presented. The hinges that were observed in the predicted conformations in the slowest modes by ANM [28] overlap the hinges that are described by GNM [25, 27], although mode-to-mode correspondence may not be applicable and mapping may be required as mentioned before. In the following subsections case studies will be presented.

4.2.1. Calcium Sensor

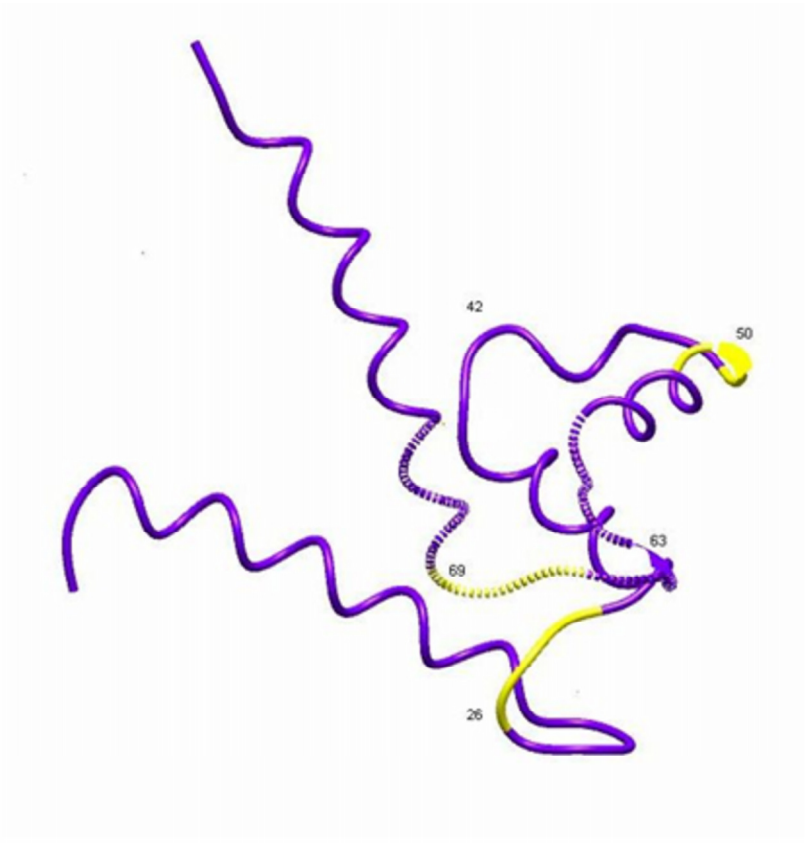

Figure 4.1: Calcium Sensor protein cartoon representation

Figure 4.1 displays Calcium Sensor which is a crucial protein for Calcium intake. Yellow displays the hinges identified in the slowest modes by GNM, spring region points to the "hinge part" given in Database of Macromolecular Movements [17] and the two strand points at index location of 50 and 63 show the hinge residues given in the Flexprot [18]. Hinge region given by Database of Macromolecular Movements [17] covers one of three hinges described by GNM. Database of Macromolecular Movements [17] agrees with one of two hinges by Flexprot [18]. 1 of 3 hinges by GNM overlaps with one of two hinges by Flexprot [18].

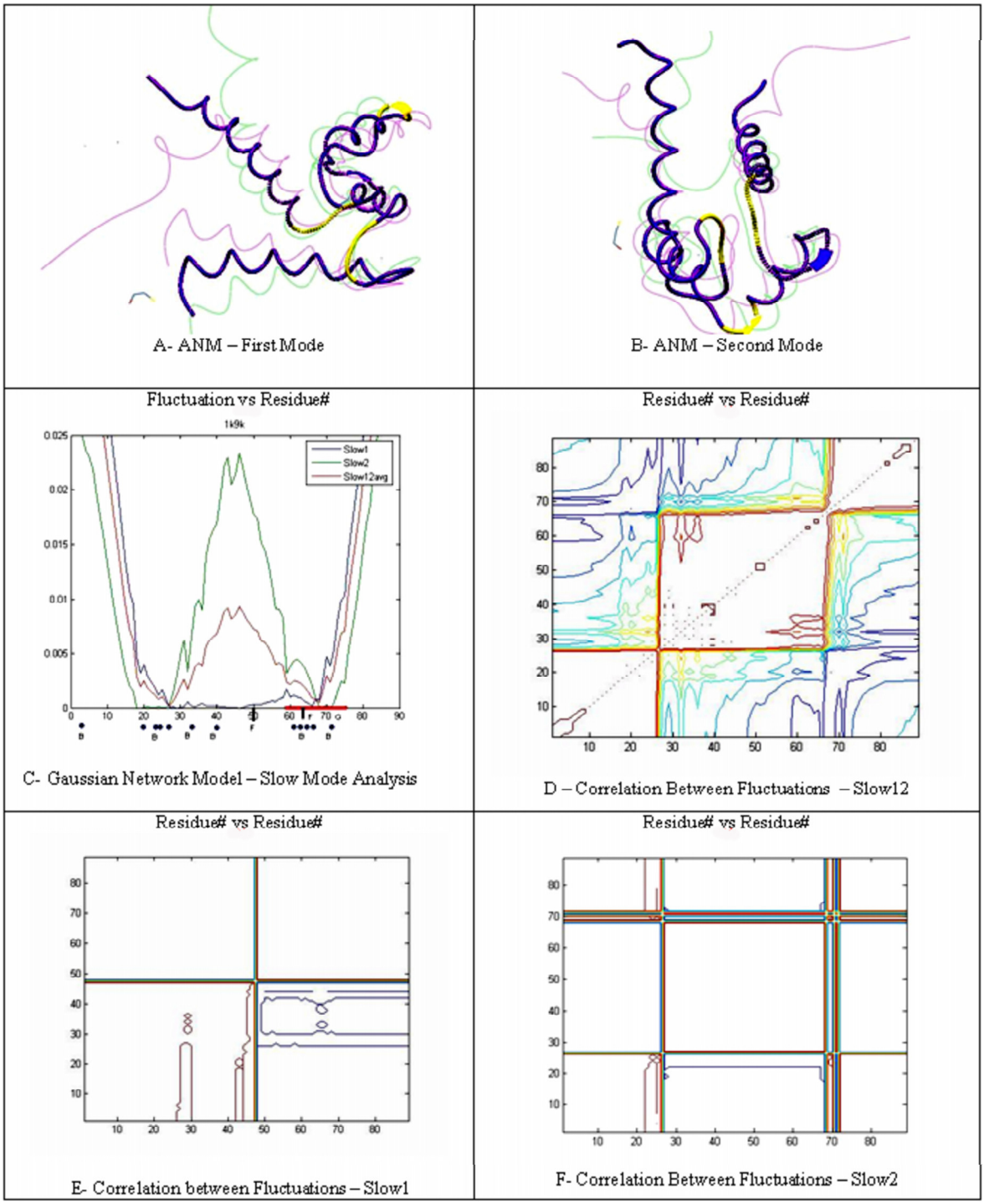

Figure 4.2: Calcium Sensor protein detailed study

Figure 4.2 elaborates the results by both GNM [25, 27] and ANM [28] (Panels A-F). The conformations that describe the fluctuations (green and magenta) from the average structure (blue) in the principle modes of motion by ANM [28] are presented in Panels A (slowest first mode) and B (slowest second mode), respectively. The analysis of the distribution of the fluctuations and the correlations between the fluctuations by GNM [25, 27] eases identification of the hinges in the latter panels. Panel C displays the fluctuations in the slowest two modes and in the average of the slowest two modes; whereas Panel D, E and F display the correlation between the fluctuations in the respective modes. Although, the hinge at residues around 50 appears in the most cooperative mode and divides the structure into two dynamic domains, it doesn't survive in the average of the slowest two modes. However, it should be contributing to a breathing motion in the structure (Panels A and B). The two more hinges at residues around 27 and 68 in the average mode appear with an effective contribution from the second slowest mode of motion. One can conclude here that the primary hinges are these latter two hinges and they operate cooperatively (Panel F), one being agreed by the data in Database of Macromolecular Movements [17]. These two hinges divide the structure into three dynamic domains.

4.2.2. Bound Calcyclin

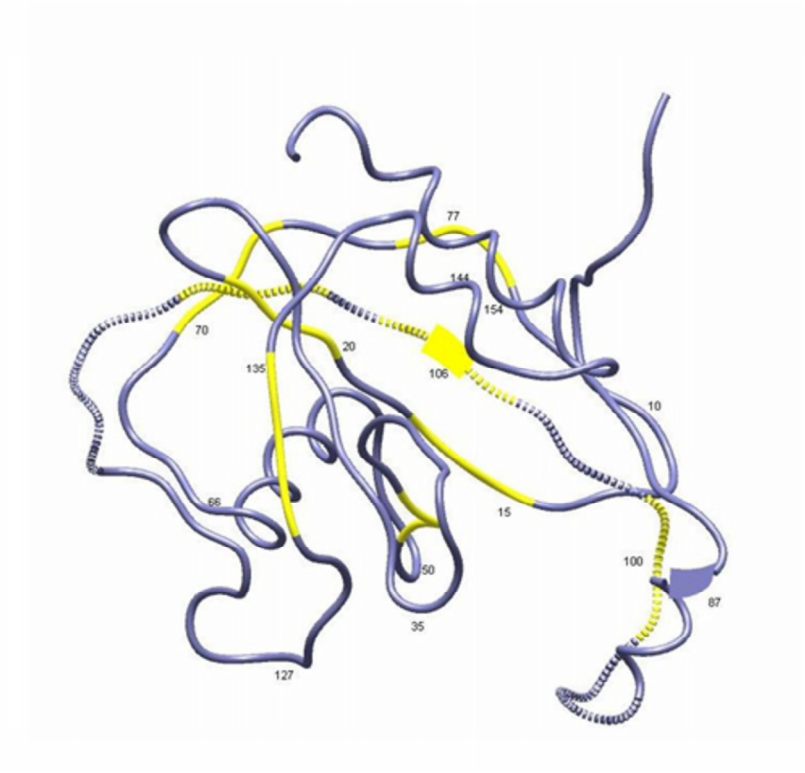

Figure 4.3: Bound Calcyclin protein cartoon representation

Figure 4.3 displays ribbon diagram of Bound Calcyclin (PDB code: 1JKN). Hinge regions by GNM [25, 27] (using the fluctuations in the first two slowest modes), Database of Macromolecular Movements [17] and Flexprot [18] are presented on the structure. Yellow Region implies main slow mode hinges determined with GNM. Spring region implies the "hinge part" given in Database of Macromolecular movements. Two strand point shows the hinge residues given in Flexprot. Some important residues are labeled with residue number. ANM figures show that the loop between residue 87 and 100 seems to display enhanced amplitude of motion (as will be seen below) mediated by the hinge residues around 100, 106, 77, and 15. A relatively small movement, vibration, can be observed at around residue 127 controlled by residue 135. A very long hinge segment defined by Database of Macromolecular Movements [17] cover three hinges by GNM [25, 27] and one hinge point by Flexprot [18]. One of the two hinges by Flexprot [18] overlap one of the hinge points by GNM [25, 27].

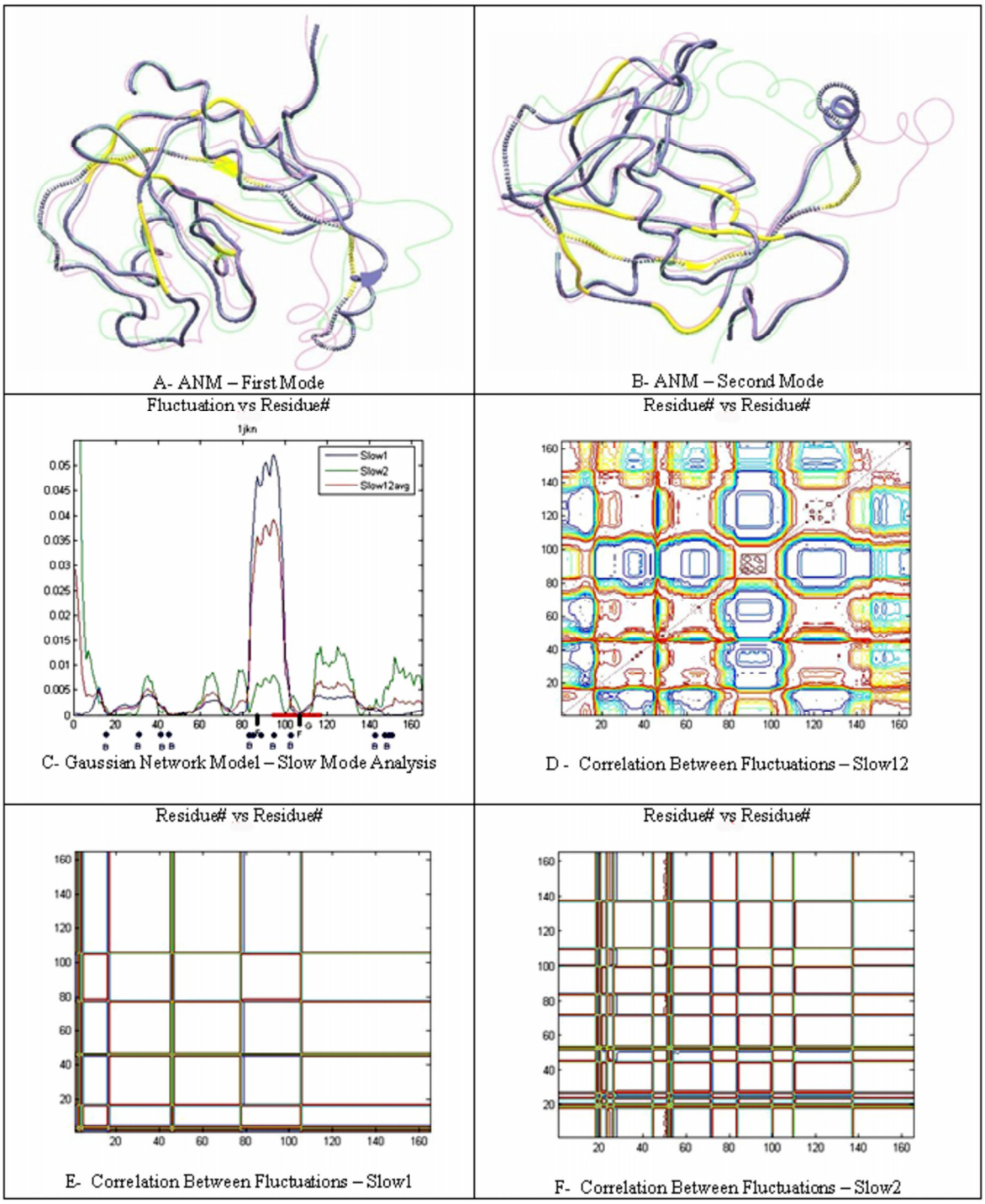

Figure 4.4: Bound Calcyclin protein detailed study

Figure 4.4 elaborates the results by both GNM [25, 27] and ANM [28]. Panels A and B display the conformations that describe the fluctuations from the average structure in the two principle modes of motion, the slowest first and second modes, respectively. When the distribution of the fluctuations in the slowest modes is analyzed, it can be observed that the loop region between 80 and 100 displays a very enhanced motion. The hinges at residues around 105 and 80 mainly control this motion. Additionally, there are some other hinge points in the vicinity as indicated in the figure. They are all correlated as it is observed in Panel D, where those hinges operate in the same mode of motion and survive as well in the average of the slowest first two modes. In the slowest second mode, a new hinge appears at residues around 135. This case is a good example, where correlated hinges spatially both close and far can be observed.

4.2.3. Lupin Ap4A hydrolase

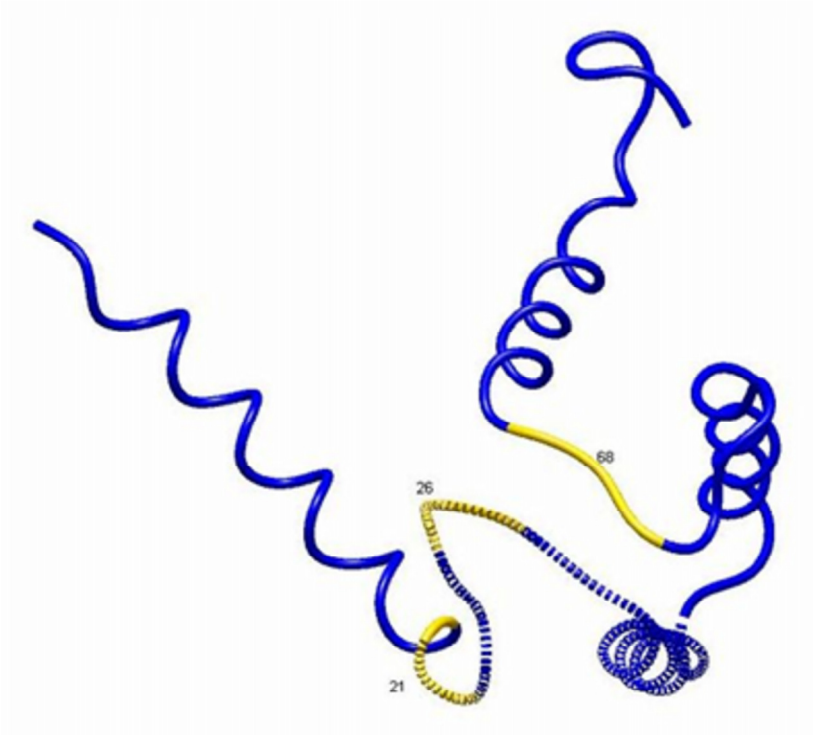

Figure 4.5: Lupin Ap4A hydrolase protein cartoon representation

Figure 4.5 displays ribbon diagram of Lupin Ap4A hydrolase (PDB code: 1A03). Hinge regions by GNM [25, 27] (using the fluctuations in the slowest first two modes), Database of Macromolecular Movements [17] and Flexprot [18] are presented on the structure. Yellow Region implies main slow mode hinges determined with GNM. Spring region implies the "hinge part" given in Database of Macromolecular Movements [17]. When ANM figures are inspected the main two movements of the molecule are observed at two ends as opening-closing action. The main hinges causing this movement can be seen in GNM output (yellow region). The hinge region given in Database of Macromolecular Movements [17] covers two (might be considered as one hinge -an extended hinge as well) of three hinges determined by GNM [25, 27]. There is another hinge in the vicinity, which might possibly correlate with the latter hinges. With these hinges, one would expect and opening and closing behavior as will be discussed below.

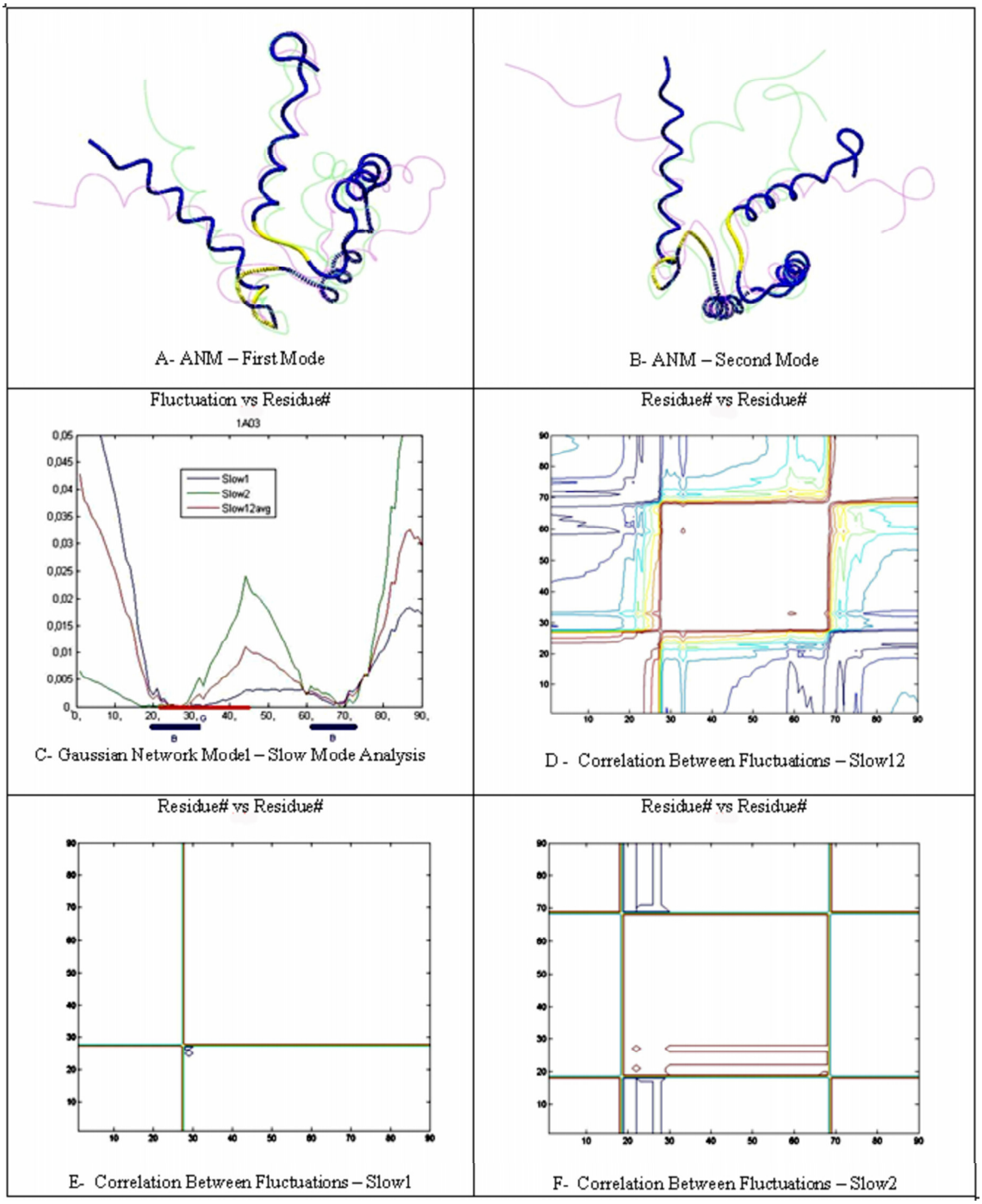

Figure 4.6: Lupin Ap4A hydrolase protein detailed study

Figure 4.6 elaborates the results by both GNM [25, 27] and ANM [28]. Panels A and B display the conformations that describe the fluctuations from the average structure in the two principle modes, in the slowest first and second modes, respectively. In the slowest first mode, the hinges at 21 and 26 (the two might be considered as one hinge as well) seem to dominate, whereas, the hinge at residues around 70 incorporates as well in the second mode. The analysis of the correlations between the fluctuations (Panels D-F) reveal that the hinge at residues around 28 is indeed the primary hinge as it appears in the slowest first and average of the slowest two modes. On the other hand, the one at residues around 20 appears in the next slowest mode and seems to be a part of the primary hinge. The hinge at residues around 70 is another hinge and correlates with the primary hinge.

4.2.4. Biotin Carboxylase

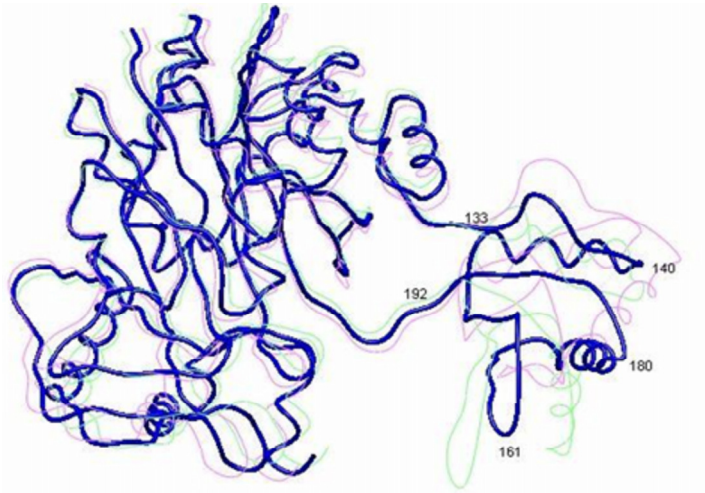

Figure 4.7: Biotin Carboxylase ANM Motion - Mode 1

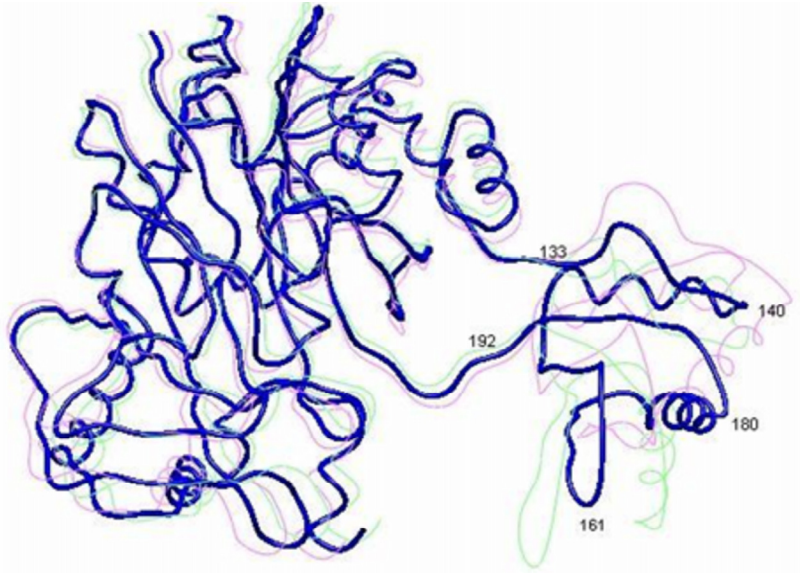

Figure 4.8: Biotin Carboxylase ANM Motion - Mode 2

Figures 4.7 and 4.8 display the conformations (green and magenta) that describe the fluctuations from the average structure (blue) of Biotin Carboxylase (PDB code 1BNC) in the slowest first and second modes of motion by ANM [28], respectively.

Another figure, Figure 4.9 is available to display the distribution of the mean-square fluctuations of the residues by GNM [25, 27]. The dominant motion observed here belongs to the loop regions between residues 133 and 192. Database of Macromolecular Movements [17] gives a hinge region between residues 218 to 233, which is in the left crowded region of the molecule and is not displayed here. The motion observed

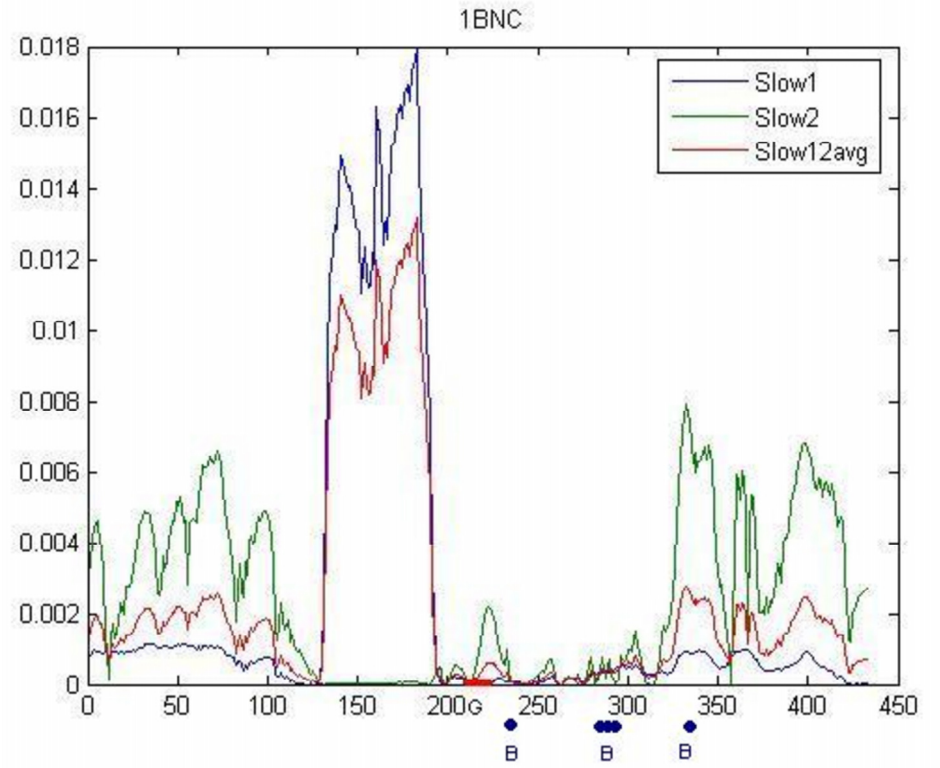

Figure 4.9: Biotin Carboxylase GNM results

between residues 133 and 192 is a necking motion. A similar motion is described in Database of Macromolecular Movements [17]. In the figure below the blue region on the x-axis spot the hinges suggested by the Database of Macromolecular Movements [17] and red region on the x-axis labels the hinge predicted by Flexprot [18].

5. RESULTS AND DISCUSSION

Preliminary test results and case studies had shown that the method is quite successful in many cases so a package which is capable of joining the mentioned methods, extracting the required information and making use of it, was worth the effort. Such a package would be useful for not only it allows the method to be tested on large data sets, but also it enables for the other user to make use of the method. So a package which can employ the method with the input and gives a summary of the results was constructed for this purpose. The package is named as HingeProt, which is inspired by its starting point, devising a method to predict the "Hinges".

5.1. Case Studies

Many case studies have been carried out to test the success of the method. Most of them successfully identified the hinges, and predicted the motion of the protein. In this section a few of those case studies will be presented with the results.

Figure 5.1: HingeProt output for Calmodulin Molecule (PDB 4cln)

Figure 5.1 represents the HingeProt output for Calmodulin molecule (4cln). Note that this molecule was presented before in the Theory section and the conformational

changes into the molecule upon binding was explained. There is a very large conformational change in the calmodulin molecule upon ligand binding. This involves splitting of one large helix. When the HingeProt output is analyzed it can be observed that the first slowest mode divides the molecule into two rigid parts in the middle of the helix that was proposed to be split.

Figure 5.2: HingeProt output for Glutamin Binding Protein - 1ggg

Figure 5.2 represents the HingeProt output for Glutamine binding protein. The protein undergoes conformational change from the open to the closed form upon binding of glutamine. Again the first slowest mode correctly recognizes the two domains that move one relative to other, as two rigid part. Note that the domains are intertwined, i.e. one of the rigid parts consists of two rigid segments.

As a multichain example Hemoglobin molecule which accomplishes a vital function for the body by carrying oxygen molecules in the body, can be given. Figures 5.3 (a) and 5.3 (b) provides outputs for a multichain example, Hemoglobin. For oxygen binding the Fe atom at the center is exposed by the expansion of the interface between the chains. The motion can be observed by comparing Figures 5.3 (a) and 5.3 (b). Rigid parts subject to this motion can be discerned from the color coding. HingeProt method predicts the hinge location correctly.

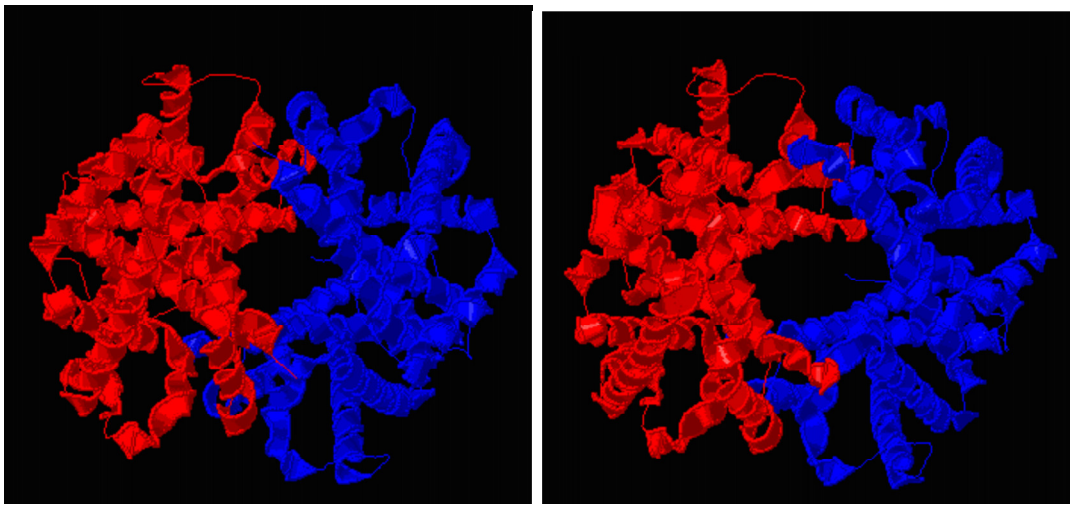

Figure 5.3: HingeProt output for Hemoglobin - 1bz0. (a) Closed conformation (b) Open conformation

5.2. Statistical Validation

After the method was developed and the package is constructed, a procedure to test the applicability and the success of the method was needed. Preliminary studies and case studies following the development of the package had shown that the success of the method was promising. However after the method is improved and assembled the further statistical tests to prove the validity of the method was needed. Since the method is unique in the way it predicts the hinges, a one to one comparison with other available methods was not possible.

To test the accuracy and validity of the method we devised the following procedure. Given two different conformations of the same protein, if there is a hinge movement between the two conformations, than the rigid alignment of their structures will not be able to obtain alignment that covers the full length of the protein structure. Only part of the structure will be aligned. For example, when aligning calmodulin protein in two different conformation (PDB codes: 1rfj:A and 1up5:A) only half of the length can be aligned. Calmodulin consists of two domains with the hinge between them and only one of the domains is aligned in rigid alignment. However if we know where the hinge is, we can split the structure into rigid fragments and align the full length of the protein. This observation was used for HingeProt validation. If

the hinges recognized by HingeProt can describe the conformational change between the two structures, than the size of the alignment by rigid fragments will be higher compared to the rigid alignment of the whole structure. So using this idea a procedure is devised. The procedure is summarized in Figure 5.4.

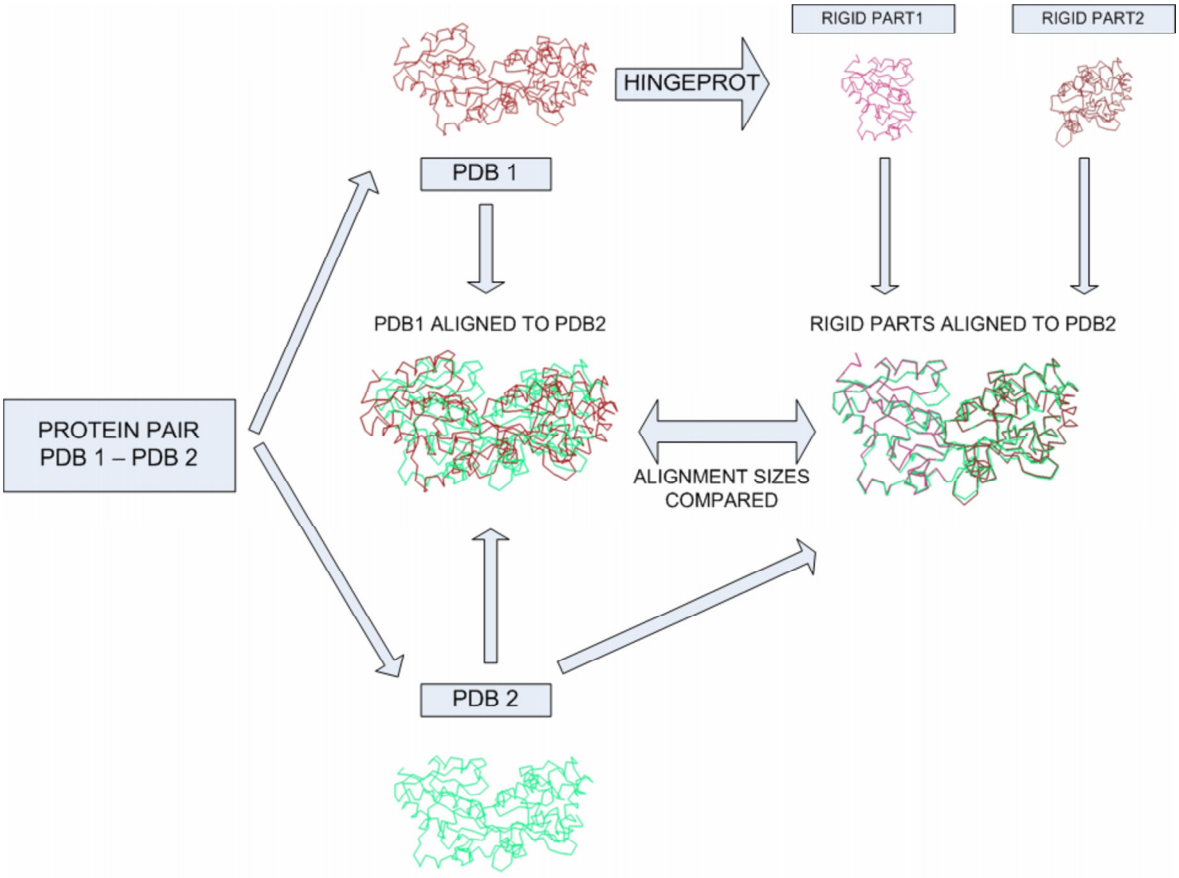

Figure 5.4: Statistical Validation Procedure for HingeProt

For this purpose two different data sets are used. The first data set consisted of 39 protein chain pairs collected from Database of Molecular Motions Morph Server [17] and the second data set consisted of 1226 protein chain pairs collected from DynDom database [19]. Among the two data sets the first one is composed of more distinctive and explored pairs which are already known to exhibit the hinge motion. So this is considered as the primary data set. In the second data set, the type of motion between two conformations is not taken into account but since it is large in amount, it is especially good for statistical data extraction purpose and to verify the results of the first data set. Each pair in both data sets represents two conformations of a structure

which transition from one to another is possible by the motion pivoted by the hinges. Two proteins in each pair are aligned using MultiProt [30] -multiple sequence alignment software which is capable of aligning many protein structures at the same time- and alignment size and RMSD are recorded for each pair. Then for each pair, rigid segments and hinges of the protein using the method explained are determined. PDB file is split into fragment files according to determined rigid segments. The fragments of the first conformation are aligned as separate structures again using Multiprot, with the second structure of the pair. The total alignment size - i.e. total number of residues that are aligned with the second structure in all segments - and the average root mean square distance between all aligned residues of the first protein chain and the second protein chain are recorded. The aim was to determine to increase in alignment percentages before and after the split. Alignment percentage, here, refers to the total number of aligned residues divided by number of residues in the smaller structure (the highest possible alignment size). Results are also compared with FlexProt [18] which uses a different methodology and requires both of the conformations as input.

Table 5.1: Results for the statistical analysis

Set 1 - Database of MacroMolecular Movements		
	Alignment Size	RMSD
Original PDB Alignment	74%	0.61
HingeProt 1^{st} Slowest Mode	88.2%	0.54
HingeProt 2^{nd} Slowest Mode	85.5%	0.53
Set 2 - DynDom Database		
	Alignment Size	RMSD
Original PDB Alignment	70.7%	0.67
HingeProt 1^{st} Slowest Mode	89.7%	0.54
HingeProt 2^{nd} Slowest Mode	86.5%	0.54

The results for the statistical analysis are summarized on Table 5.1. Those are the average values and obtained by an extensive analysis. Details of those numbers and the full data set are given at the Appendix.

The results of the analysis point out the quite considerable success of the method. For the first data set which consisted of 39 protein pairs the improvement in alignment percentage was 14.2 per cent in average (increase from 74 per cent to 88.2 per cent) for the slowest mode and 11.5 per cent (increase from 74 per cent to 85.5 per cent) for the second slowest mode. For the second data set which contained 1226 cases improvement in alignment percentage was even better, 19 per cent for the slowest mode (from 70.7 per cent to 89.7 per cent) and 15.8 per cent (from 70.7 per cent to 86.5 per cent).

For the first data set, 15 of 39 cases increase from 80 per cent average to over 95 per cent, giving an average of 97.8 per cent, in the first slowest mode. That increase in alignment percentage show that the rigid body fragmentation of the first protein chain is carried out successfully at the correct hinge locations so that the protein motion is favored from the first conformation to the second. The results are similar for the second data set. Alignment percentage increases up to 90 per cent with nearly 20 per cent improvement. For 112 cases of the second data set the increase in alignment percentage is from 80 per cent in average to 99 per cent and higher for the first slowest mode. Nearly 100 per cent alignment means nearly all residues of two conformations are aligned for those 112 cases. Those results show that the methodology is quite promising, supporting the success of the method especially in the first slowest mode.

6. THE WEB SERVER

This study aimed not only to devise the method but also present this method to the science community by making it available over the web. To serve this purpose a web server is constructed. When an individual uses such a server, he expects is to be simple and comprehensive at the same time. Extensive calculations can be made based on very complex theories but the need to present the results as compact and as comprehensive as possible is crucial in today's fast developing science environment. User-friendly interface is also very important to meet the needs of a user. All these concerns are considered when building the server. Another important point to improve such servers is user feedbacks. This also has been accounted. Since the server has been running for quite a while, many feedbacks are received and every feedback is taken into account. In fact, the web server is currently running in its third version and has improved a lot since the first version. The address for the third version of the server is http://www.prc.boun.edu.tr/appserv/prc/HingeProt3. The latest version can be reached by the link at the address http://www.prc.boun.edu.tr/

As mentioned, the web server is quite easy to use and fast. The input to the server is the coordinate file of a protein chain in PDB format. The user may enter four letter PDB code or upload his own PDB file (Figure 6.1). The user will choose whether he/she wants to carry out the analysis on single chain or the whole structure. If the input file includes more than one chain and the user selects the "Select chain" option the user will be able to select one or more of the chains on the next page. The user may optionally enter an e-mail address, to which the results will be sent once the job has finished. In any case, the results will be displayed in the browser window. Figure 6.1 is the initial page of HingeProt. The following figure, Figure 6.2 represents the chain selection page where the user able to select one or more chains of a protein structure.

The web server calculates the hinge residues and predicts the rigid parts between the hinge sites using the slowest first and second modes of motion. The rigid parts and the hinge residues are listed for the two slowest modes. The rigid parts and the

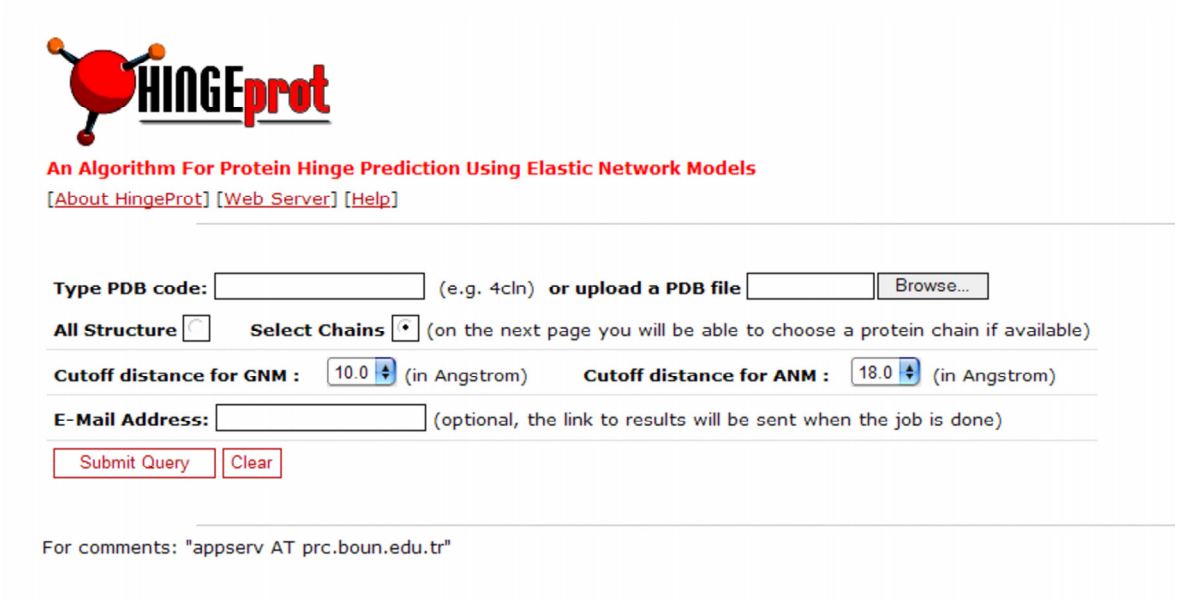

Figure 6.1: The Welcome page of HingeProt

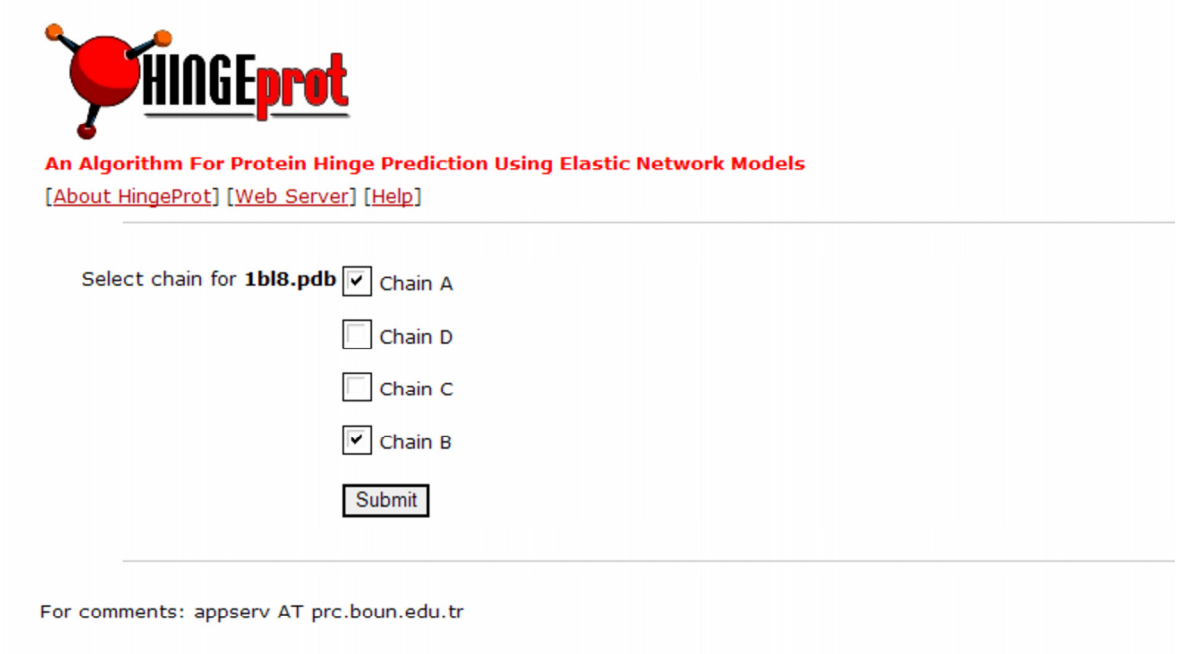

Figure 6.2: Chain Selection page of HingeProt

fluctuations of the structure at each mode can be viewed by Jmol script on the results page. The Jmol script provides an animation for the visualization of the protein motion calculated by the method. The server also provides the PDB file for each of the two modes. This PDB file allows the user to visualize the rigid parts and fluctuation using the temperature factor column and the model, respectively, for color coding in his/her own computer. Below, we provide an example with its behavior in the slowest first mode, as which might possibly be the mode mainly contributing to the function of the protein.

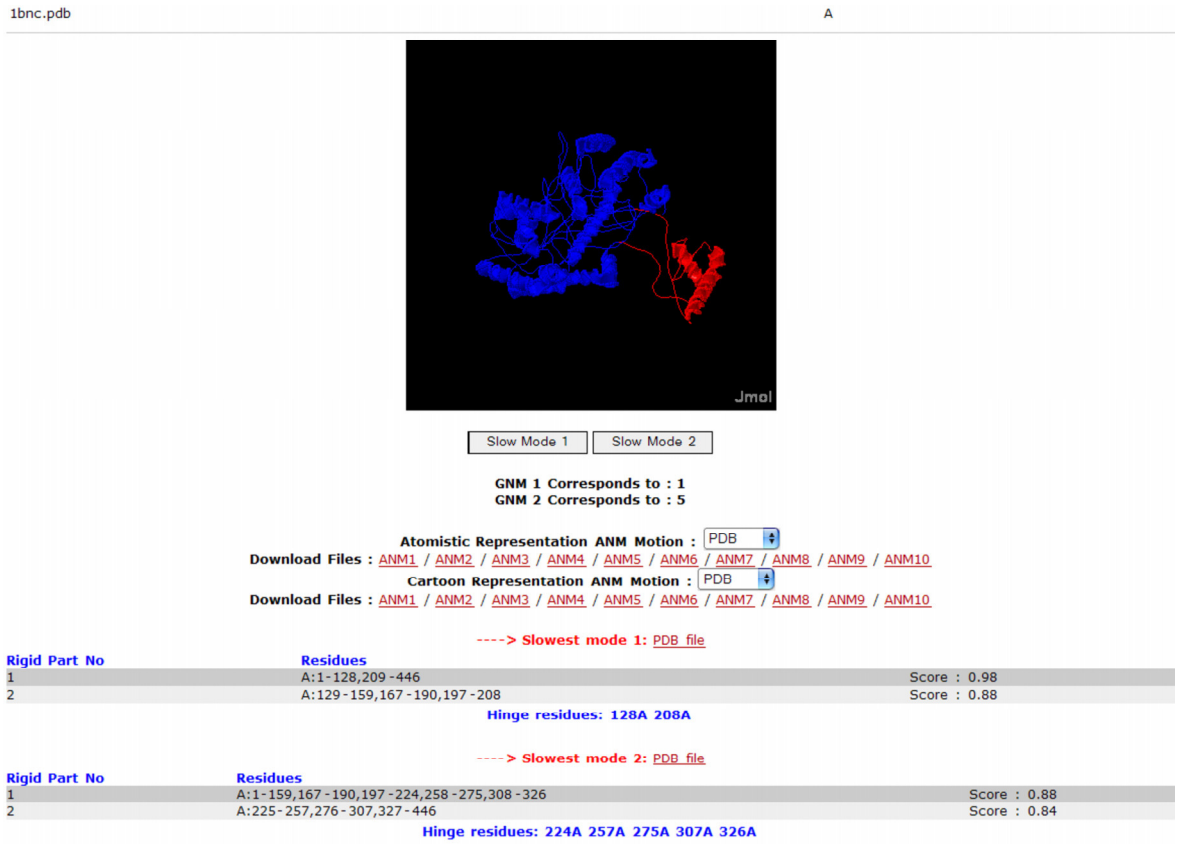

Figure 6.3: The Results page of HingeProt for Biotin Carboxylase - 1BNC

Figure 6.3 provides an example for the results page of HingeProt. As seen in the figure, the results page summarizes both rigid parts and the hinges, whereas the figure displayed in the results page visualizes both the motion and the partitioning of the structure. The displayed results belong to Biotin Carboxylase which is discussed before as a case study and shown in Figure 4.7 and Figure 4.8.

7. CONCLUSIONS

The method described here and the HingeProt server is based on the predictions of the elastic network models, GNM and ANM. With this method we were able to predict the rigid parts and the hinge regions using a single conformation of the protein structure. Further, it provides the direction of the fluctuations of the residues in the structure, which shows how the flexible joint mediates the structural motions.

The validity of the results obtained by the methodology described here was tested both specifically and statistically. Both the case studies and the statistical analysis carried on an extensive data set pointed out the considerable success of the method.

A web server employing the method was also built with a user-friendly interface. It provides the results and the visualization PDB files within seconds to several minutes for proteins consisting of up to a few thousands of residues.

The method is expected to be useful in a range of potential applications such as flexible protein-protein and protein-ligand docking, and fitting flexible hinge-bent protein structures into EM density maps and refining the EM structures, and also comprehension of functional mechanisms of macromolecular structures and assemblies.

8. RECOMMENDATIONS

Although the HingeProt methodology is shown to provide quite satisfactory results, further tests and development on the method are possible. Many available databases give information on the residues which accomplish the function of the protein, i.e. the interacting residues, after detailed protein studies. A more elaborated case specific study on a large data set using this information may give further idea about the validity of the methodology in predicting the function.

The assumptions of the methodology can be further examined. HingeProt is currently working on a coarse grained level, that is, all residues are considered as indistinguishable particles. So if an atomistic version of the method, which identifies the differences in different residues, can be devised, results may be compared to see if the improvement is worth the increase in the running time.

Another study can be carried out to explore the effect of the parameters on the protein size. Many trials have shown that the parameters may have considerable effects on the results depending on the size of the protein. Those parameters are taken constant in the statistical analysis. A more detailed study could show whether a procedure to adjust the parameters dynamically according to the protein input is possible or not.

Currently the size of the protein input is limited by the physical capabilities of the computers the method is running on. So an upgrade of the computers would also increase the limits of the method. The web server has been running since a while and each feedback is evaluated so as to meet the user needs in a better fashion. This policy should be preserved in order to keep the utilization of the method by the third parties.

APPENDIX A: DETAILED STATISTICAL RESULTS FOR DATABASE OF MACROMOLECULAR MOVEMENTS

Table A.1: The Detailed Results of HingeProt for each protein chain

PDB1		# PDB1	PDB2		# PDB2	ALG_PDB	%COVERAGE	RMS_PDB	#M1	ALG_MODE1	%COVERAGE	RMS_M1	#M2	ALG_MODE2	%COVERAGE	RMS_M2
1gtm	A	417	1hrd	A	449	258	***61.9%***	0.93	2	303	***72.7%***	0.89	3	273	***65.5%***	0.83
1cnp	A	90	1a03	B	90	30	***33.3%***	0.95	2	38	***42.2%***	0.86	2	35	***38.9%***	0.78
1f3y	A	165	1jkn	A	165	109	***66.1%***	0.85	2	115	***69.7%***	0.80	2	111	***67.3%***	0.84
1gtr	A	529	1nyl	A	523	409	***78.2%***	0.75	2	430	***82.2%***	0.72	3	397	***75.9%***	0.65
1e8b	A	160	1e88	A	160	91	***56.9%***	0.78	2	138	***86.3%***	0.72	2	100	***62.5%***	0.66
1g0x	A	192	1p7q	D	183	108	***59.0%***	0.75	2	172	***94.0%***	0.70	3	145	***79.2%***	0.69
6tim	A	249	1tre	A	255	198	***79.5%***	0.71	2	204	***81.9%***	0.70	3	202	***81.1%***	0.69
1k20	A	310	1k23	A	302	176	***58.3%***	0.66	2	284	***94.0%***	0.69	4	247	***81.8%***	0.62
1ser	-	793	1ses	-	842	717	***90.4%***	0.74	2	734	***92.6%***	0.68	3	773	***97.5%***	0.62
1byu	A	202	1rrp	A	204	133	***65.8%***	0.66	2	136	***67.3%***	0.66	2	133	***65.8%***	0.64
2efg	A	582	1fnm	A	655	376	***64.6%***	0.59	2	563	***96.7%***	0.65	3	461	***79.2%***	0.50
1aa7	A	158	1ea3	A	157	132	***84.1%***	0.78	2	141	***89.8%***	0.63	3	144	***91.7%***	0.61
1qln	A	862	1msw	D	863	511	***59.3%***	0.66	2	533	***61.8%***	0.63	6	660	***76.6%***	0.52
1ex6	A	186	1ex7	A	186	123	***66.1%***	0.67	2	149	***80.1%***	0.62	3	176	***94.6%***	0.66
1oxs	C	352	1oxu	C	353	247	***70.2%***	0.92	2	342	***97.2%***	0.62	3	337	***95.7%***	0.55
1n0v	C	825	1n0u	A	819	425	***51.9%***	0.69	2	640	***78.1%***	0.61	4	554	***67.6%***	0.50
1fgu	B	238	1jmc	A	238	110	***46.2%***	0.67	2	204	***85.7%***	0.60	4	205	***86.1%***	0.61
2cbl	A	305	1b47	B	304	226	***74.3%***	0.48	2	263	***86.5%***	0.55	3	295	***97.0%***	0.44
1l96	-	162	1l97	-	328	104	***64.2%***	0.77	2	151	***93.2%***	0.53	3	143	***88.3%***	0.66
4crx	B	322	1crx	A	322	236	***73.3%***	0.81	2	293	***91.0%***	0.52	3	284	***88.2%***	0.51
1joj	A	351	1jg6	A	351	233	***66.4%***	0.61	2	328	***93.4%***	0.52	4	314	***89.5%***	0.65
1ecb	A	475	1ecc	A	492	431	***90.7%***	0.66	2	436	***91.8%***	0.51	2	435	***91.6%***	0.52
1dkx	A	219	1dky	A	211	183	***86.7%***	0.66	2	208	***98.6%***	0.48	2	191	***90.5%***	0.64
1ckm	A	317	1ckm	B	317	242	***76.3%***	0.47	2	288	***90.9%***	0.47	4	303	***95.6%***	0.46
2lao	-	238	1laf	-	239	143	***60.1%***	0.41	2	236	***99.2%***	0.47	3	175	***73.5%***	0.36
3dap	A	320	1dap	B	320	235	***73.4%***	0.44	2	262	***81.9%***	0.44	4	309	***96.6%***	0.40
1l5b	A	101	1l5e	A	101	51	***50.5%***	0.40	2	99	***98.0%***	0.42	4	98	***97.0%***	0.43
1bnc	A	433	1dv2	A	450	368	***85.0%***	0.38	2	416	***96.1%***	0.41	2	369	***85.2%***	0.37
1fto	B	257	1ftm	B	258	172	***66.9%***	0.53	2	246	***95.7%***	0.39	5	203	***79.0%***	0.41
2nac	A	374	2nad	A	391	294	***78.6%***	0.82	2	370	***98.9%***	0.37	3	371	***99.2%***	0.42
1tgl	-	265	4tgl	-	265	251	***94.7%***	0.39	3	252	***95.1%***	0.37	3	252	***95.1%***	0.36
1ttp	A	256	1ttq	A	256	254	***99.2%***	0.38	2	255	***99.6%***	0.37	2	255	***99.6%***	0.38
1ake	A	214	1ank	A	214	213	***99.5%***	0.44	3	213	***99.5%***	0.37	5	213	***99.5%***	0.37
2ktq	A	528	3ktq	A	539	477	***90.3%***	0.37	2	477	***90.3%***	0.35	3	493	***93.4%***	0.43
3enl	-	436	7enl	-	436	419	***96.1%***	0.40	2	424	***97.2%***	0.35	3	423	***97.0%***	0.34
1lua	A	287	1lu9	A	287	287	***100.0%***	0.55	2	287	***100.0%***	0.32	3	287	***100.0%***	0.52
1dqz	B	280	1dqy	A	283	253	***90.4%***	0.34	4	255	***91.1%***	0.31	3	253	***90.4%***	0.32
1gu0	A	149	1gu1	A	149	143	***96.0%***	0.26	2	144	***96.6%***	0.26	2	143	***96.0%***	0.25
5cro	A	61	6cro	A	60	50	***83.3%***		2	49	***81.7%***		2	52	***86.7%***	0.00

Table A.1 gives the detailed results of statistical analysis done to test the validity of HingeProt. PDB codes and chain identifiers for each pair are given in the table #PDB1 column is the size (number of aminoacids) of the first protein, #PDB2 column is the size of the second protein, ALG_PDB is the alignment size calculated with MultiProt before partitioning, %COV is the percent coverage of alignment with respect to maximum possible alignment size (the size of smaller PDB,) RMS_PDB is the RMSD before partitioning, #M1 is the number of rigid parts determined with HingeProt, ALG_MODE1 is the alignment size determined according to the slowest mode, %COV is the percent coverage of the alignment, RMS_M1 is the RMSD after partitioning based on the first slowest mode and the following four columns are the results for the second slowest mode likewise.

Table A.2: The Detailed Results of FlexProt for each protein chain

PDB1		# PDB1	PDB2		# PDB2	flex0	flex0 rmsd	flex1	flex1 rmsd	flex2	flex2 rmsd	flex3	flex3 rmsd	flex4	flex4 rmsd
1gtm	A	417	1hrd	A	449	146	1.45	208	1.46	248	1.43	276	1.43	302	1.41
1cnp	A	90	1a03	B	90	20	1.31	39	1.41	53	1.36	64	1.33	73	1.35
1f3y	A	165	1jkn	A	165	81	1.49	119	1.48	139	1.48	150	1.47	149	1.44
1gtr	A	529	1nyl	A	523	201	1.50	290	1.49	374	1.48	440	1.48	481	1.43
1e8b	A	160	1e88	A	160	67	1.43	122	1.29	160	1.01	141	1.31	146	1.17
1g0x	A	192	1p7q	D	183	56	1.44	111	1.47	157	1.41	171	1.21	178	1.22
6tim	A	249	1tre	A	255	132	1.49	170	1.26	200	1.28	224	1.25	236	1.26
1k20	A	310	1k23	A	302	121	1.39	189	1.43	235	1.44	267	1.39	291	1.27
1ser	-	793	1ses	-	842	367	1.50	718	1.48	760	1.21	793	0.98	783	1.01
1byu	A	202	1rrp	A	204	105	1.46	132	1.46	157	1.45	174	1.45	181	1.43
2efg	A	582	1fnm	A	655	349	1.49	562	1.29	580	1.15	408	1.28	424	1.29
1aa7	A	158	1ea3	A	157	157	1.38	49	1.39	65	1.24	75	1.26	91	1.29
1qln	A	862	1msw	D	863	220	1.50	360	1.49	479	1.44	577	1.45	663	1.43
1ex6	A	186	1ex7	A	186	111	1.48	163	1.08	186	0.84	184	1.03	173	1.37
1oxs	C	352	1oxu	C	353	329	1.50	337	1.48	344	1.50	306	1.45	316	1.45
1n0v	C	825	1n0u	A	819	430	1.50	651	1.49	730	1.42	788	1.27	819	1.15
1fgu	B	238	1jmc	A	238	107	1.44	189	1.37	225	1.37	230	1.31	217	1.38
2cbl	A	305	1b47	B	304	241	1.50	253	1.47	264	1.46	274	1.46	279	1.46
1l96	-	162	1l97	-	328	110	1.48	162	1.05	156	1.20	160	1.14	160	1.10
4crx	B	322	1crx	A	322	181	1.44	308	1.44	316	1.36	319	1.27	293	1.43
1jej	A	351	1jg6	A	351	194	1.49	351	1.09	284	0.85	298	0.84	308	0.79
1ecb	A	475	1ecc	A	492	332	1.41	462	1.41	470	1.30	375	1.38	385	1.36
1dkx	A	219	1dky	A	211	211	0.99	54	1.22	71	1.29	84	1.06	97	1.12
1ckm	A	317	1ckm	B	317	237	1.38	305	0.78	317	0.71	284	1.38	286	1.39
2lao	-	238	1laf	-	239	108	1.46	194	0.69	238	0.50	222	0.77	186	1.22
3dap	A	320	1dap	B	320	153	1.39	247	0.90	320	0.74	283	1.01	292	1.00
1l5b	A	101	1l5e	A	101	52	1.20	101	0.70	71	1.19	77	1.28	84	1.33
1bnc	A	433	1dv2	A	450	249	1.23	386	1.27	408	1.28	420	1.27	426	1.27
1fto	B	257	1ftm	B	258	138	1.49	220	1.48	257	0.67	249	0.92	252	1.46
2nac	A	374	2nad	A	391	374	1.18	41	1.04	56	1.11	71	1.11	83	1.17
1tgl	-	265	4tgl	-	265	182	1.37	263	1.30	265	0.99	226	1.38	231	1.24
1ttp	A	256	1ttq	A	256	256	0.40	45	1.43	66	1.25	86	1.31	102	1.33
1ake	A	214	1ank	A	214	214	0.45	39	1.46	52	1.30	65	1.24	78	1.24
2ktq	A	528	3ktq	A	539	356	1.50	526	1.48	417	1.49	436	1.49	454	1.49
3enl	-	436	7enl	-	436	436	0.91	64	1.31	85	1.31	105	1.32	124	1.34
1lua	A	287	1lu9	A	287	287	0.55	48	1.46	70	1.46	84	1.44	100	1.39
1dqz	B	280	1dqy	A	283	211	1.46	268	1.45	275	1.35	265	1.45	268	1.28
1gu0	A	149	1gu1	A	149	149	0.98	42	1.31	59	1.38	74	1.34	86	1.33
5cro	A	61	6cro	A	60	57	1.43	28	1.39	37	1.16	43	1.09	48	1.01

Table A.2 gives the detailed results of FlexProt which is another method to determine the hinge locations. It requires 2 protein conformations. The results are alignment size and RMSD according to the number of hinge locations tried (from 0 Hinge Location to 4 Hinge Locations). The aim of this method is to find the hinge locations so as to maximize the alignment size for a given number of Hinges.

APPENDIX B: THE DYNDOM DATABASE

Table B.1: The DynDom data set - protein pairs PDB codes

PDB1	CH	SIZE	PDB2	CH	SIZE	PDB1	CH	SIZE	PDB2	CH	SIZE	PDB1	CH	SIZE	PDB2	CH	SIZE	PDB1	CH	SIZE	PDB2	CH	SIZE
1oc0	A	364	1lj5	A	379	1iz4	A	241	1ge8	A	238	1cyy	B	251	1cy9	A	245	1gae	O	330	1dc6	B	330
1nb7	A	566	1gx5	A	518	1oj7	C	390	1oj7	A	390	1qpg	-	415	1fw8	A	415	1t3f	B	219	1t04	B	218
1tub	B	427	1tvk	B	426	1wdw	L	385	1v8z	B	387	1dpj	A	329	1fmu	A	322	1yj7	B	154	1yj7	A	154
1j3b	A	513	1j3b	B	518	1owr	Q	284	1a02	N	280	1m2v	A	705	1m2o	C	718	1sva	3	342	1sva	1	348
1c7j	A	485	1qe3	A	467	1owr	Q	284	1pzu	L	276	1jej	A	351	1m5r	A	351	1s7o	B	105	1s7o	C	108
1lvk	-	743	1vom	-	730	1n9g	C	364	1n9g	B	364	1wdn	A	223	1ggg	B	220	1rii	A	243	1rii	C	237
1lvk	-	743	1jwy	A	753	1rpj	A	288	1gud	A	288	1vhl	A	208	1n3b	B	203	1qr4	B	175	1qr4	A	175
1kcm	A	256	1t27	A	269	1owr	M	284	1pzu	M	276	1skq	A	416	1skq	B	416	1v58	A	229	1v58	B	229
1o89	A	320	1o8c	B	324	1tjy	A	316	1tm2	A	314	1ggk	C	727	1gg9	A	727	1mau	A	328	1i6m	A	326
1j1w	A	738	1itw	D	740	1f6m	F	320	1trb	-	316	1cza	N	898	1bg3	B	902	1w72	M	210	1w72	L	210
1xr7	B	460	1tp7	A	450	1tlb	Q	326	1tkl	B	325	1b02	A	279	1bkp	B	278	1a3q	B	285	1a3q	A	285
1ihm	B	511	1ihm	C	492	1jr3	E	334	1xxh	F	328	1m46	A	148	1m45	A	142	1euz	A	416	1euz	F	416
1yjq	A	293	1ks9	A	291	1uae	-	418	1ejd	B	418	1i6j	A	256	1d0e	A	259	1u2z	C	385	1u2z	B	379
1ecc	A	492	1ecf	B	500	2bhv	F	188	2bhv	A	191	1yy9	A	613	1mox	B	501	1hf2	C	202	1hf2	A	196
1rkm	-	517	1jet	A	517	1nkt	A	836	1nkt	B	836	1rz8	B	228	1yyl	B	229	1nr0	A	610	1pev	A	610
1ihm	B	511	1ihm	A	492	1jr2	B	260	1jr2	A	260	1e2w	B	251	1cfm	C	251	1w72	H	223	1w72	I	223
1mrp	-	309	1nnf	A	308	1h54	B	754	1h54	A	752	1s9a	A	256	1s9a	B	256	1f06	A	320	1f06	B	320
1s3r	B	473	1s3r	A	473	1yyq	A	353	1yyq	B	351	1ocj	A	360	2bvw	B	360	1e02	P	221	1e02	H	221
1jqo	A	904	1jqo	B	904	1iho	A	282	1iho	B	282	1xpy	D	370	1xs2	B	360	1yv9	B	257	1yv9	A	257
1jv1	A	490	1jvd	B	486	1n9g	B	364	1h0k	B	364	1q1b	B	367	1q1b	D	367	1t5o	D	340	1t5o	C	340
1q14	A	289	1szd	A	291	1ba2	A	271	1urp	C	271	1qun	J	279	1qun	B	279	1e3a	A	258	1kec	A	206
1aro	P	774	1h38	A	857	1dqy	A	283	1dqz	A	280	1hf0	B	128	1hf0	C	131	1rlm	B	269	1rlt	B	268
1aro	P	774	1cez	A	862	1n7g	D	313	1n7h	B	334	2bhn	D	211	2bhn	A	211	8fab	C	206	8fab	A	206
1ny6	N	247	1ny6	L	246	1a81	G	220	1a81	K	220	1xw5	B	217	1hnb	A	217	1v1h	A	103	1v1i	A	102
1m3i	C	465	1pfo	-	471	1cdo	B	374	1cdo	A	374	1tkk	A	359	1jpm	A	359	1ooa	A	313	1ooa	B	313
1pfo	-	471	1m3i	B	465	1i36	B	258	1i36	A	258	1viy	C	208	1vhl	A	208	1fmc	B	255	1ahh	A	253
1ek5	A	346	1ek6	B	345	1rf5	C	427	1rf5	D	427	1f06	B	320	2dap	-	320	1b4a	C	146	1b4a	B	146
1tpz	A	395	1tqd	B	379	1yzy	B	412	1yzy	A	412	1xmm	B	288	1xml	B	282	1l4u	A	165	1u8a	A	163
1i00	B	278	1ju6	D	286	1xdv	B	759	1xdv	A	758	1or3	A	136	1gs9	A	144	1h47	B	158	1h48	E	157
1uiv	A	497	1uiu	A	496	1owr	M	284	1owr	Q	284	1z4o	B	215	1o08	A	221	1v1h	A	103	1v1i	C	98
1sw4	B	270	1sw5	D	270	1pzu	L	276	1pzu	B	276	1f0k	B	351	1nlm	B	350	1za3	H	224	1za3	B	224
1m3i	C	465	1m3i	A	465	1kam	B	175	1kaq	C	186	1rp5	A	687	1rp5	B	680	1ux2	F	207	1uv6	F	205
1e6e	C	456	1e1l	A	455	1pxa	-	394	1k0i	A	394	1rzm	B	338	1vr6	A	343	1qb3	B	119	1qb3	A	113
1bkd	S	439	1nvv	S	469	1uf2	A	967	1uf2	B	1019	1on7	B	172	1jbo	B	171	1axk	B	394	1axk	A	393
1vjy	A	299	1b6c	H	326	1h4s	A	473	1hc7	A	464	1fcj	B	304	1fcj	A	302	1x6v	B	586	1x6v	A	564
2nmt	A	422	1iic	A	422	1i7d	A	620	1d6m	A	603	1aov	-	686	1dot	-	686	1np3	A	327	1np3	B	327
1q3s	E	517	1q3q	D	518	1ig9	A	901	1q9x	D	903	1ge8	A	238	1iz5	A	240	1fe8	I	210	1fe8	H	210
1xhx	B	571	1xi1	B	571	1uke	-	193	1qf9	A	194	1f7v	A	606	1f7u	A	606	1ihg	A	364	1iip	A	297
1si1	A	316	1q35	A	317	1u4n	A	308	1qz3	A	309	1n0t	B	257	1nnk	A	258	1tjh	L	213	1u8k	A	214
1k1y	B	636	1k1y	A	636	1ypw	B	692	1yqi	C	690	1gp9	C	171	1gmo	C	172	1gji	B	275	1gji	A	275
1ft9	A	210	1ft9	B	206	1kqm	A	777	1dfk	A	728	1wvc	A	254	1tzf	A	251	1b4a	C	146	1b4a	A	146
2udp	A	338	1lrk	A	338	1ryx	A	686	1ovt	-	682	1eft	-	405	1mj1	A	405	3fru	A	269	3fru	C	269
1dpe	-	507	1dpp	A	507	1ldj	A	725	1u6g	A	715	1ee8	B	266	1ee8	A	266	2gbp	-	309	1gcg	-	309
1uiv	A	497	1uiv	B	496	1gru	N	524	1pcq	G	524	1qo0	E	194	1qo0	D	189	1lu9	A	287	1lua	C	287
1gxb	C	341	1o17	A	339	1kqm	A	777	1dfl	A	772	1obh	A	762	1h3n	A	814	1q08	B	89	1q0a	B	89
1lls	A	370	1a7l	A	380	1lh0	B	206	1lh0	A	213	1obz	B	165	1v1t	A	164	1pn0	A	652	1pn0	C	656
1i8t	A	367	1i8t	B	367	1xko	A	150	1squ	B	154	1n0v	D	825	1n0v	C	825	1u9a	A	160	1u9b	-	159
8acn	-	753	5acn	-	754	1b7u	A	689	1b1x	A	689	1fwl	B	296	1fwl	D	296	1kke	B	205	1kke	A	206
1q8y	A	351	1q8y	B	357	1gru	N	524	1gr5	J	517	1akz	-	223	1emh	A	223	1qiu	D	264	1qiu	C	264
1ojl	C	252	1ojl	A	292	1hwx	A	501	1nr1	C	496	1akm	C	322	1duv	H	333	1fg9	E	193	1fg9	D	205
1mc5	A	373	1teh	A	373	1d0n	A	729	1h1v	G	327	1bg3	B	902	1hkb	B	899	1jq5	A	366	1jpu	A	361
1pix	A	586	1pix	B	586	1f7v	A	606	1bs2	A	603	1aby	A	283	1c7d	A	284	1bvu	E	416	1bvu	D	416
1n8k	A	374	1ye3	A	374	1tfw	C	437	1r8c	A	437	1wpm	B	308	1k23	C	303	1opl	B	365	1opk	A	449
1suo	A	465	1po5	A	465	1sw6	B	254	1sw6	A	254	1ses	A	421	1set	B	421	1f3d	H	217	1f3d	K	215
1ix3	A	222	1irm	C	196	1yy9	A	613	1nql	A	612	1kzh	A	550	1kzh	B	530	1k46	A	124	1huf	A	123
1e7p	A	655	1qla	A	655	1rke	A	262	1rkc	A	258	1b6r	A	349	1b6s	B	355	1q57	E	483	1q57	A	483

Table B.1: The DynDom data set - protein pairs PDB codes

PDB1	CH	SIZE	PDB2	CH	SIZE	PDB1	CH	SIZE	PDB2	CH	SIZE	PDB1	CH	SIZE	PDB2	CH	SIZE	PDB1	CH	SIZE	PDB2	CH	SIZE
1e4v	A	214	4ake	B	214	1hf0	B	128	1cqt	A	134	1cko	-	317	1ckm	B	317	1kke	B	205	1kke	C	206
1xp5	A	994	1iwo	A	994	1ze2	A	300	1ze1	C	308	1xdv	A	758	1xd4	A	824	1g5c	A	169	1g5c	F	169
1noc	A	372	1vaf	A	419	1x86	A	362	1txd	A	352	1lbq	B	354	1lbq	A	356	1zka	A	110	1zk9	A	110
1k2o	A	406	1yrc	A	405	1sx4	H	524	1kp8	A	525	1cyy	B	251	1cyy	A	250	1ad9	H	219	1ad9	B	219
1e4v	A	214	1e4y	B	214	1eer	B	213	1ern	B	209	2bgz	A	293	2bgy	A	293	2tmg	C	408	2tmg	F	408
1ka2	A	497	1k9x	B	497	1hyr	C	275	1b3j	A	264	1jlh	C	556	1g98	A	555	1gr5	J	517	1gr5	C	517
1jr3	E	334	1a5t	-	324	1kdt	B	223	2cmk	A	221	1kf0	A	416	1vjc	A	416	1ra0	A	423	1r9z	A	423
1xdo	A	687	1xdp	A	687	1oen	-	524	1os1	A	537	1xx5	C	212	1xx5	B	216	1q57	E	483	1q57	C	483
1w8j	A	722	1w8j	D	717	1mu2	B	402	1mu2	A	538	1dpf	A	178	1s1c	B	179	1q57	D	483	1q57	F	483
1gq3	A	294	2atc	A	305	1v4s	A	448	1v4t	A	424	1cko	-	317	1ckm	A	317	1ct8	B	220	1ct8	D	220
1aky	-	218	1dvr	A	220	1x9q	A	231	2mpa	L	219	1fgu	B	238	1fgu	A	246	1q57	D	483	1q57	B	483
1ais	A	181	1pcz	A	183	1k9a	F	436	1k9a	A	439	1n0t	B	257	1p1u	B	258	1q57	B	483	1q57	E	483
1br2	E	673	1br1	C	787	1iy2	A	245	1ixz	A	238	1evl	D	401	1evk	B	401	1za6	E	220	1za6	G	220
1f20	A	435	1tll	A	630	1ng1	-	294	1jpn	A	296	1xcg	E	358	1xcg	A	355	1wiq	B	363	1wiq	A	363
1h38	A	857	1s77	D	828	1qvi	Y	141	1s5g	Y	142	1ykd	A	383	1ykd	B	380	1i1c	A	205	1i1c	B	205
1x91	A	149	1x90	A	147	1xhx	D	571	1xhx	B	571	2bgz	A	293	1wlg	A	293	1von	W	173	1vow	W	173
1ydi	A	256	1rke	A	262	1kix	A	446	1jb7	A	460	1iq8	B	577	1iq8	A	577	1von	W	173	1voy	W	173
1f6a	B	217	1o0v	B	320	1vic	A	255	1vh3	B	242	1tuu	A	399	1g99	B	398	1r9s	I	119	1r5u	I	118
1xdv	B	759	1xd4	B	824	1i3q	B	1083	1twf	B	1094	1rzu	B	478	1rzu	A	477	1wiq	A	363	1wip	A	363
1ovt	-	682	1n04	A	683	1fp5	A	208	1f6a	B	217	1ezf	B	324	1ezf	A	323	1wiq	A	363	1wio	A	363
1jdn	A	407	1jdp	A	396	1ser	B	421	1ses	A	421	1tpl	B	426	1tpl	A	426	1m0w	A	481	1m0t	B	455
1fhu	A	298	1r6w	A	321	1md7	A	304	1md8	A	314	1a4i	B	295	1a4i	A	285	2tys	B	395	1qoq	B	394
1wy5	A	311	1wy5	B	311	1qun	B	279	1klf	P	279	1vbk	A	307	1vbk	B	296	1m5y	B	377	1m5y	C	376
1zt4	C	278	1zt4	A	272	1sl0	A	658	1t8e	A	698	1k3p	A	426	1owc	B	426	1coz	A	126	1n1d	A	126
2bhs	C	291	2bht	C	291	1mbt	-	340	1uxy	-	340	1k20	A	310	1wpp	A	310	1ufq	C	212	1uei	B	209
1ung	A	288	1unl	B	292	1jt0	D	185	1jus	D	184	1z69	D	327	1z69	A	327	1jpz	B	458	1bvy	A	439
1gll	O	494	1bu6	Y	499	1i7s	C	511	1i7q	A	517	1gxm	A	324	1gxo	A	320	1ctq	A	166	1nvv	R	166
1qmn	A	365	2ach	A	337	1ebu	A	358	1ebf	B	358	1ybi	A	284	1ybi	B	284	1kph	D	285	1kp9	A	270
1j1u	A	299	1u7d	B	284	1i3q	B	1083	1rtf	B	106	1we3	L	526	1we3	M	525	1unk	A	87	1ayi	-	86
1k30	A	363	1iuq	A	350	1gh6	B	326	1n4m	A	345	1f0k	B	351	1f0k	A	351	1r5u	A	1380	1i3q	A	1414
1s2o	A	244	1tj5	A	244	1v9d	B	321	1v9d	A	308	1q3x	A	321	1q3x	B	315	1bih	B	391	1bih	A	391
2tpt	-	440	1otp	-	440	1mwk	A	320	1mwm	A	316	1ups	A	402	1ups	B	397	1mok	D	522	1mo9	A	522
1khd	D	329	1khd	B	328	1jrr	A	357	1by7	A	354	1o58	B	285	1o58	A	293	1s8f	A	168	1wms	A	170
1t93	A	403	1se6	B	402	2bpa	1	426	1cd3	F	426	1ots	C	221	1ots	E	221	1typ	B	486	1fec	B	485
1gq3	A	294	1q95	B	310	1okk	A	290	1ng1	-	294	1r4l	A	597	1r42	A	597	1kku	A	216	1kqo	E	233
1wgz	C	510	1wgz	A	510	1si7	A	340	1szw	A	330	1dkf	A	217	1g5y	A	228	1uaa	A	636	1uaa	B	633
1gqc	A	242	1h7h	B	241	1hx1	A	377	1kay	-	378	1bcc	C	379	3bcc	C	379	1n7t	B	255	1n7t	A	255
1dxh	A	335	1ort	B	335	1x86	A	362	1x86	E	336	1ta8	A	313	1tae	C	321	1qib	A	161	1gen	-	200
1ba2	A	271	2dri	-	271	1dgm	A	346	1lii	A	331	1bgx	T	828	1taq	-	807	1npp	A	244	1m1g	A	240
1qn9	B	186	1vok	A	192	1pea	-	368	1qo0	B	374	1mmk	A	309	1j8u	A	307	1hzv	A	514	1hzu	A	521
1rf6	D	427	1rf5	C	427	16pk	-	415	13pk	C	415	1q4x	A	239	1n46	A	248	1fig	L	215	1dqd	L	214
1p7h	N	286	1owr	N	284	1m43	B	331	1lee	A	331	2pjr	A	542	1pjr	-	623	1v7b	A	175	1v7b	B	172
1p7h	N	286	1pzu	H	276	1n0v	D	825	1zm9	C	822	2pjr	A	542	3pjr	A	646	1y8r	B	512	1y8q	B	510
1ovn	A	229	1ovn	B	226	1vdw	B	248	1vdw	A	248	1jib	A	585	1jl8	B	585	1efu	D	282	1efu	B	282
1gl3	A	367	1t4b	A	367	1nam	H	275	1s7t	D	276	1lvo	E	299	1p9u	B	300	1ors	A	214	1ob1	D	215
1k5h	B	391	1q0q	B	398	1ifg	A	140	1azz	D	138	1ed3	A	275	1kjm	A	277	1jih	A	509	1jih	B	509
1ni0	B	157	3pvi	A	156	1fvf	B	543	1epu	A	527	1fwy	B	324	1fxj	A	327	3bjl	B	216	2fb4	L	216
1go4	G	100	1go4	E	87	1ohh	B	479	1e79	A	492	1s5u	E	136	1s5u	C	130	1sh8	A	153	1sh8	B	149
1eov	A	487	1asy	A	490	1e9i	D	431	1e9i	A	430	1q3u	F	322	1f44	A	316	1k9a	B	441	1k9a	C	438
1igt	D	444	1igt	B	444	1tqb	C	219	1cr9	L	219	1t33	A	220	1t33	B	214	1n48	A	342	1s9f	C	341
1gdt	B	183	1gdt	A	183	2trt	-	198	1bjz	-	194	1ozv	A	429	1p0y	B	441	1t09	A	414	1t09	B	414
1l6k	E	73	1l6k	D	72	1yz6	A	261	1yz7	A	176	1yp4	D	426	1yp2	B	426	1k1q	A	333	1k1s	A	341
1rtj	A	543	1fko	A	542	1v9p	A	584	1v9p	B	584	1ktl	C	274	1mhe	A	273	1mjg	N	728	1mjg	P	728
1w3b	A	388	1w3b	B	368	1ulk	B	126	1ulk	A	126	1kem	H	218	1q9l	B	222	1ibv	F	228	1hq6	B	228
1c0a	A	585	1eqr	A	590	1rka	A	305	1rkd	-	306	1o9l	A	468	1m3e	A	459	1a3x	B	487	1a3w	A	492
1e6c	B	170	1shk	B	159	1ass	-	152	1asx	-	152	1psd	A	404	1yba	D	398	1yvl	A	652	1yvl	B	653
1gk4	A	79	1gk4	F	74	1h4v	B	404	1ady	C	420	1xag	A	353	1xah	A	323	1vzy	A	290	1vzy	B	286
1mmi	A	366	1jql	A	366	1pvp	A	323	1q3u	F	322	1u4g	A	298	1ezm	-	298	1z5b	B	460	1mu5	A	460
1go4	G	100	1go4	H	93	1wp9	C	475	1wp9	F	475	1sqf	A	425	1sqg	A	424	1vsc	B	196	1vca	A	199
1r9s	I	119	1y1v	I	119	1dn1	B	222	1ez3	B	124	1urz	B	387	1urz	F	388	1qxx	A	76	1ihr	B	74

Table B.1: The DynDom data set - protein pairs PDB codes

PDB1	CH	SIZE	PDB2	CH	SIZE	PDB1	CH	SIZE	PDB2	CH	SIZE	PDB1	CH	SIZE	PDB2	CH	SIZE	PDB1	CH	SIZE	PDB2	CH	SIZE
1e7d	B	157	1e7l	B	157	1sdc	A	192	1vrr	A	203	1p2c	E	211	1mlc	B	218	1ub6	L	213	1a3l	L	217
3pro	D	152	2pro	C	129	1ex6	B	186	1ex6	A	186	1ues	C	191	1ues	A	191	1dg1	G	385	1efc	A	386
1y10	A	360	1y10	C	363	1vlk	-	142	1j7v	L	150	1f0y	A	291	1f17	B	291	1kb5	L	214	1qbm	L	214
1ini	A	225	1vgt	A	207	1c0m	B	216	1c0m	C	222	1r45	D	200	1r45	B	201	1bx2	B	191	1jwm	B	187
1w2d	B	250	1w2f	A	270	1yi8	B	331	1yi8	A	331	1jvk	B	214	1jvk	A	215	1xcq	F	218	1mhh	D	214
1d9c	A	121	1d9g	A	121	1fx7	D	220	1u8r	A	221	1yt5	A	256	1yt5	D	256	1k5h	A	398	1k5h	B	391
2bw0	A	309	1s3i	A	307	1bbu	A	486	1bbw	A	469	1jn6	B	214	1jnh	H	214	1w39	B	189	1auy	A	163
2bm1	A	660	1ktv	B	632	1i1a	D	205	1i1c	A	205	1a4j	H	217	1i7z	D	220	1u56	A	188	1u55	B	187
1hjb	E	68	1hbj	D	68	1jqb	A	351	1kev	B	351	1gmo	C	172	1gmo	H	171	1t09	A	414	1t0l	B	414
1sn4	A	994	1vfp	A	994	1h03	P	125	1h2q	P	119	1b8g	B	425	1ynu	A	417	1h99	A	220	1tlv	A	204
1ytz	C	159	1a2x	A	158	1kp3	A	439	1k92	A	444	1e0t	D	446	1e0u	C	461	1i4j	B	110	1i4j	A	110
1i5d	A	190	1i5a	B	177	1npr	A	242	1npp	A	244	1n73	E	315	1lwu	B	315	1ae6	L	219	1mju	L	219
1yuh	L	211	1q0x	L	212	1upm	C	123	1aus	S	123	1sz9	C	140	1sza	B	140	1e0c	D	196	1djs	A	202
1be3	E	196	3bcc	E	196	1oqx	B	209	1e4k	B	216	1wxd	A	263	1wxd	B	263	1u94	A	306	1xmv	A	293
1sjp	A	447	1sjp	B	445	1w72	M	210	1adq	L	213	1e4w	H	213	1e4x	I	217	1ic1	A	190	1ic1	B	190
2tn4	-	155	1tcf	-	156	1g3j	A	522	2bct	-	502	1rzf	L	213	1q1j	M	215	1d5x	A	179	1fv1	D	178
1ukl	A	876	1ukl	B	876	1g38	D	393	2adm	B	385	1u42	A	101	1u36	A	100	1r37	A	347	1jvb	A	339
1xxh	H	366	1xxh	G	364	1pkx	D	589	1pl0	A	589	1zq9	B	278	1zq9	A	278	1a79	A	171	1a79	B	171
1b6d	B	212	1bey	L	214	1y0z	B	322	1y0z	A	321	1n71	C	179	1b87	A	181	1ewk	B	449	1ewv	B	448
1juo	B	172	1juo	A	172	1e5l	A	449	1e5q	A	449	1qya	B	307	1qy9	D	291	1au7	A	130	1au7	B	128
1vrt	A	525	1ep4	A	522	1d5i	H	221	1d5b	B	221	8fab	D	222	8fab	B	214	1bjm	A	216	3bjl	B	216
1mi7	R	103	3wrp	-	101	1k2y	X	459	1p5d	X	454	2gfb	J	219	1kno	D	220	1c5d	B	215	1c5d	H	214
1hvu	D	554	1vrt	A	525	1y1x	A	182	1y1x	B	174	1g83	B	161	1g83	A	161	1v7m	I	217	1v7n	K	217
1nkr	-	195	2dl2	A	197	1iaw	B	304	1ev7	B	293	1g9k	A	455	1o0q	A	453	1es7	C	104	1reu	A	103
1joc	A	123	1joc	B	123	1q6u	A	213	1q6h	B	208	1zcz	B	454	1zcz	A	452	1s9f	C	341	1ryr	A	341
1w0j	E	466	1h8e	D	467	1j4g	C	241	1j4g	D	241	1lqs	M	142	1lqs	L	142	1rih	L	212	1ct8	A	214
2091	-	165	172l	-	164	1ckc	A	212	1kdt	B	223	1l5b	B	101	3ezm	A	101	1eeh	A	431	1e0d	A	429
1nak	L	217	1cgs	L	219	1hkc	A	899	1cza	N	898	1sva	4	331	1sva	2	348	1wcw	A	254	1wd7	B	255
1vkx	A	273	1lei	A	273	1tw2	B	350	1tw3	A	340	1zhh	A	344	1jx6	A	338	1l5h	A	405	1m1n	F	477
1gxj	A	161	1gxl	A	205	1uax	B	211	1uax	A	211	1nd0	F	222	1rum	H	222	1s7g	E	248	1s7g	A	252
2ram	B	273	1vkx	A	273	1lgp	A	113	1lgq	A	112	1sva	4	331	1sva	3	342	1z5h	B	780	1z5h	A	780
1ni0	B	157	1h56	A	156	1kn1	B	161	1b33	M	160	1xf3	H	216	1i8m	B	213	1v8c	D	165	1v8c	B	160
1i5d	A	190	1i58	A	189	1okr	A	120	1sd6	B	118	1igt	C	214	1igt	A	214	1bjm	A	216	3bjl	A	216
1e0s	A	173	1hfv	B	164	1wvg	A	352	1wvg	B	350	1m5y	D	389	1m5y	B	377	1orq	A	215	1ors	A	214
1c0w	B	219	1p92	A	218	1swv	A	257	1rdf	F	263	1r49	A	548	1k4t	A	564	1ffo	B	408	1p4e	D	408
1hvu	D	554	1dlo	A	556	1knv	B	291	1knv	A	290	1fnt	Z	212	1g65	Y	211	1c2a	A	120	1tx6	J	114
1p1g	L	215	1clz	L	219	1e22	A	485	1lyl	B	483	1p1f	B	496	1p1j	A	525	1bp5	D	334	1ryo	A	324
1vf7	C	246	1vf7	M	232	1am9	D	76	1am9	C	82	1qu3	A	880	1ffy	A	917	1t9a	A	597	1jsc	A	541
1p9m	B	163	1alu	-	157	1cdc	B	96	1a64	B	94	1acc	-	665	1t6b	X	676	1q5v	C	115	1q5v	A	115
1xdt	T	518	1f0l	B	522	1e7d	B	157	1e7d	A	157	1d5f	B	350	1c4z	B	350	2buj	A	291	2buj	B	286
1f59	B	440	1ibr	B	458	1gmj	C	59	1gmj	A	65	2pjr	A	542	2pjr	F	544	1eak	B	419	1eak	A	421
2hap	D	75	2hap	C	76	1aup	-	427	1hrd	C	449	1lk3	A	136	1vlk	-	142	1hzv	A	514	1nir	A	538
1gvm	F	134	1h8g	A	92	1q06	B	126	1q06	A	122	1stz	C	311	1stz	B	311	1tc6	A	222	1qye	A	225
2mpa	L	219	1pz5	A	215	1k0b	D	259	1jzr	C	255	1e7v	A	850	1e7u	A	872	1txv	B	440	1tye	B	440
1iqp	C	319	1iqp	D	326	1sjs	-	415	1hj6	A	414	1lkx	A	650	1lkx	C	679	1ewk	B	449	1ewt	B	456
1ytz	C	159	1top	-	162	1jbg	A	106	1r8d	A	109	1ajs	A	412	1ajs	B	411	1e6z	B	498	1e6n	A	496
1y11	A	364	1y10	B	357	1r4c	G	110	1g96	A	111	1s59	F	153	1s59	C	149	1a0q	L	211	1fj1	C	213
1ev7	A	295	1iaw	B	304	1jr1	A	436	1nf7	A	454	1c1g	C	284	1c1g	A	284	1ff6	H	219	1ff5	B	219
1i4m	A	108	1uw3	A	106	1deb	A	54	1deb	B	53	1l6l	8	74	1l6k	E	73	1dqm	L	214	1gpo	M	218
1ko5	B	172	1knq	A	171	1tj7	A	455	1tj7	B	451	1av1	D	201	1av1	C	201	1d6v	L	211	1axs	L	211
172l	-	164	214l	-	163	1u28	A	378	1l7d	B	358	1l6l	S	74	1l6l	8	74	1moe	B	240	1moe	A	240
1bbd	L	219	1hin	L	217	1q6u	A	213	1q6h	A	208	1mco	L	216	3mcg	2	216	1meo	A	202	1men	C	201
1zme	C	70	1zme	D	70	1st0	A	300	1xmm	B	288	1qzx	A	425	1qzw	G	432	1s9f	C	341	1s97	A	341
1kw2	B	453	1kxp	D	438	1igy	D	434	1igy	B	434	1c1g	C	284	1c1g	B	284	1k0r	B	326	1k0r	A	326
1j1d	A	159	1j1e	D	159	1vdd	A	199	1vdd	D	198	1rid	A	244	1g40	A	243	1f58	L	216	1iqw	L	215
1o5t	A	378	1r6t	A	428	1rfx	A	89	1rfx	C	89	1l6l	8	74	1l6l	K	74	1gig	H	221	2vir	B	221
1gxd	A	624	1ck7	A	619	1qlf	A	276	1s7v	A	276	1u4q	B	317	1u5p	A	211	1ktw	A	457	1h80	B	430
1e7d	B	157	1en7	B	157	1dee	F	223	1hez	D	213	1yl3	H	164	1rl6	A	164	1oa2	B	217	1olq	A	217
1m8w	B	341	1m8y	A	341	1f9n	C	148	1f9n	A	147	1hqc	A	314	1ixr	C	308	1jn6	A	206	1jnh	C	208
1nwq	C	60	1nwq	A	60	1o9l	D	464	1ooy	A	463	1j1d	A	159	1dtl	A	149	1nkd	-	59	1gmg	A	56

Table B.1: The DynDom data set - protein pairs PDB codes

PDB1	CH	SIZE	PDB2	CH	SIZE	PDB1	CH	SIZE	PDB2	CH	SIZE	PDB1	CH	SIZE	PDB2	CH	SIZE	PDB1	CH	SIZE	PDB2	CH	SIZE
1g8l	A	403	1fc5	B	396	1q4k	B	224	1q4o	B	208	1l6l	S	74	1l6l	X	75	1w0d	C	337	1w0d	D	337
1yuh	L	211	1f4w	L	210	1loo	A	431	1iwe	A	430	1gk4	B	79	1gk4	A	79	1a21	B	197	1a21	A	194
1nc2	C	215	1yuh	L	211	1qfy	B	295	1sm4	B	296	1tlf	C	296	1efa	C	286	1rih	L	212	1fh5	L	213
1ehi	A	360	1ehi	B	347	1qlf	A	276	1s7w	J	276	1req	C	727	2req	A	725	1egg	B	144	1egg	A	132
1jth	C	69	1sfc	K	73	1b3u	A	588	1b3u	B	588	3mcg	2	216	1mcb	A	216	1af7	-	274	1bc5	A	269
1za3	R	91	1du3	A	90	1eap	A	213	1a0q	L	211	1ep4	B	395	1n5y	B	429	1drw	-	272	1arz	C	271
2bku	D	854	2bku	B	857	1hvv	D	61	1jth	B	67	1l6l	S	74	1l6l	J	74	1yy9	C	211	1yy8	C	213
1w7j	A	752	1w8j	A	722	1upn	E	125	1h03	P	125	1i6v	C	1113	1smy	M	1119	1qbm	L	214	1fdl	L	214
1b9n	A	258	1b9m	B	244	1f9n	A	147	1f9n	B	149	1jpp	B	502	1g3j	A	522	1dky	A	211	1dkx	A	219
1vkx	B	312	1nfk	B	312	1g8l	A	403	1g8r	A	403	1do2	B	407	1do0	C	406	1jqd	A	286	1jqe	A	280
1y10	B	357	1y10	D	356	2bex	B	457	2bex	A	456	1e5w	A	346	1ef1	B	289	1xwl	-	580	1ua0	A	580
1iqp	C	319	1iqp	F	326	1yf2	A	425	1yf2	B	425	1iw7	F	345	1ku2	A	240	1miz	B	201	1y19	H	192
1g51	B	580	1efw	A	580	1vbg	A	874	1vbh	A	862	1mco	L	216	2mcg	1	216	1ngy	B	216	1n7m	L	216
1kfi	A	570	1kfq	A	571	1mre	L	218	1t66	L	219	1c1g	B	284	1c1g	D	284	1ex2	A	185	1exc	B	185
1q12	B	367	1q1e	A	367	1liw	A	517	1lix	C	515	1ytz	C	159	1yv0	C	148	1z5h	B	780	1z1w	A	780
1hvv	A	67	1hvv	D	61	1rku	B	206	1rku	A	206	1u4q	B	317	1cun	C	213	1pg7	M	213	1jpt	L	213
1vio	A	230	1vio	B	230	1h4v	B	404	1adj	D	420	1ygy	B	527	1ygy	A	527	1aj7	H	217	2rcs	H	217
1nbv	H	219	1cbv	H	219	1w0j	E	466	1h8e	E	454	1nx5	A	411	1y64	B	411	1d3u	B	201	1ais	B	193
1rgi	G	346	1d0n	A	729	1t3e	A	412	1t3e	B	405	1pfi	A	46	1ql1	A	46	2gfb	K	214	1kno	C	214
1hlo	B	73	1hlo	A	80	1721	-	164	1501	B	162	1vf5	Q	168	1vf5	D	168	1p13	B	102	1o4g	A	105
2rsl	A	115	2rsl	B	120	1zag	B	274	1t7y	A	273	1sv5	A	552	1n6q	A	558	1miz	B	201	1mk9	D	192
1c4p	B	137	1qqr	B	133	1pn3	A	391	1pn3	B	384	1jqj	D	325	1xxh	F	333	1s78	F	222	1l7i	H	222
2bct	-	502	1t08	A	510	1p9o	B	280	1p9o	A	269	1wtd	A	264	1wte	A	272	1n7m	H	213	1ngy	A	213
1cgs	H	214	2cgr	H	214	1w72	H	223	1dfb	H	229	1bgy	Q	196	1sqb	E	196	1qye	A	225	1u2o	A	226
1p53	A	250	1p53	B	250	2gd1	R	334	1nq5	A	334	1p9m	A	298	1bqu	A	208	1tyq	A	399	1k8k	A	401
1jzd	A	219	1eej	B	216	1sm3	L	210	2vit	A	210	1ffx	D	410	1sa0	B	419	1i9j	H	220	1l7t	H	220
1irx	B	508	1irx	A	507	1evk	A	401	1evl	D	401	2bpt	A	860	2bku	D	854	1dn0	A	215	1rhh	C	215
1za6	B	324	1za6	H	324	1gd2	G	64	1gd2	F	64	1s7b	B	107	1s7b	G	106	1rz8	A	214	1yym	Q	214
1jyu	A	96	1fyr	D	95	2fmt	A	314	1fmt	A	308	2bsu	D	275	1i4f	A	275	1qpo	F	284	1qpr	F	284
1vkx	B	312	1svc	P	311	1a4j	A	217	1mre	L	218	1bpy	A	326	1bpe	-	284	2rcs	L	214	1gaf	L	214
1iqp	E	326	1iqp	B	326	1yke	C	98	1ykh	A	95	1orq	C	223	1ors	C	132	1aip	A	373	1aip	B	373
1efx	A	278	1qqd	A	273	1eia	-	207	2eia	A	206	1clq	A	903	1rv2	D	825	1j49	B	332	1j4a	D	332
1tjl	C	145	1tjl	F	145	1fzc	B	308	1re3	F	301	1ep4	B	395	1hvu	D	554	1opg	H	227	1bm3	H	216
1y10	B	357	1y10	A	360	1z7d	D	375	1z7d	E	374	1t11	A	376	1t11	B	374	1e6n	A	496	1e6z	A	497
1ucu	A	494	1io1	A	395	1bbd	H	213	1txv	H	219	1ii8	B	174	1f2u	B	145	1vdx	A	184	1vgj	A	181
1nks	D	194	1nks	A	194	1psd	A	404	1sc6	B	377	2bsu	D	275	1lp9	H	275	2nad	A	391	2nac	B	374
1bpy	A	326	7ico	A	327	1b47	B	304	1yvh	A	304	1h3e	A	427	1h3f	B	406	1nvt	A	287	1nvt	B	287
2bnh	-	456	1dfj	I	456	1svm	C	363	1svo	B	362	3bcc	E	196	1bgy	Q	196	1ko7	A	285	1ko7	B	285
3wrp	-	101	1tro	E	97	1m	B	247	1mkm	A	246	1qgk	A	876	1ukl	A	876	1oao	D	728	1mjg	N	728
1pxy	B	477	1pxy	A	465	1ozv	A	429	1p0y	A	430	1ojw	B	251	1ok3	A	252	1rfy	A	89	1rfy	B	88
1j1d	A	159	1j1d	D	160	1m1j	B	402	1m1j	E	401	1sk6	E	143	1ahr	-	146	2rcs	L	214	1aj7	L	214
1mox	B	501	1mox	A	499	1u9o	A	194	1t56	A	193	1tui	B	397	1eft	-	405	1n8s	A	449	1lpb	B	449
1r22	B	97	1rt2	B	99	15	H	217	1fgn	H	214	1nh8	A	276	1nh7	A	274	1d5x	B	183	1d6e	B	179
1f34	B	138	1f32	A	127	1qfh	B	212	1wlh	B	306	1efw	A	580	1l0w	A	580	1ewk	B	449	1ewk	A	448
1vf7	M	232	1t5e	H	231	15c8	H	217	25c8	H	217	1n8z	C	581	1s78	B	568	1d2e	C	397	1d2e	D	397
1rl3	B	259	1rl3	A	268	1jk9	B	243	1qup	A	219	1vfp	A	994	1xp5	A	994	1gto	A	62	1qx8	B	51
1m8y	A	341	1m8z	A	339	1v2d	A	365	1v2e	B	367	1iqp	D	326	1iqp	A	326	2dld	A	337	2dld	B	337
1y4u	A	472	1y4u	B	454	2a1t	R	313	1efv	A	312	1ft4	A	140	1ext	A	160	1f3j	B	187	1es0	B	190
1do0	A	406	1g4b	F	393	1aup	-	427	1bgv	A	449	1x8z	A	151	1x8z	C	150	1jn6	A	206	1jnh	A	208
1oqx	B	209	1adq	A	206	1wqg	A	184	1wqf	A	183	1nx5	A	411	1nx4	A	410	1wej	H	223	1qbl	H	219
1d7m	B	101	1d7m	A	101	1rlm	B	269	1rlo	C	267	1iqp	A	326	1iqp	E	326	1f58	L	216	1ggb	L	215
1tij	A	114	1r4c	G	110	1jr3	D	338	1xxh	A	333	1ynj	D	1238	1hqm	D	1175	1ydk	A	207	1k3o	A	198
1ic2	A	79	1ic2	C	79	1gqq	A	428	1p3d	A	449	1ce2	A	689	1jw1	A	689	1k33	A	62	1df4	A	57
1u16	A	449	1tju	D	449	1igc	H	222	2hrp	H	226	1dkg	A	158	1dkg	B	151	1bqs	A	209	1gsm	A	206
1nfi	C	301	2ram	B	273	1sst	C	236	1ssq	D	257	1rj2	J	318	1kz7	A	327	1dpg	B	485	2dpg	-	485
1jal	A	348	1jal	B	335	1ck0	L	216	1q9w	A	219	1k8t	A	498	1xfx	D	735	1dn0	A	215	1qlr	C	215
1sd6	B	118	1sax	B	120	1xwr	A	79	1xwr	B	77	1n6m	B	335	2ncd	A	358	1ult	A	533	1v25	B	507
1e5l	A	449	1ff9	A	447	1f3u	E	118	1f3u	G	118	1go4	H	93	1go4	F	87	1fe8	L	211	1fe8	M	211
1nay	C	56	1nay	A	56	1p78	A	92	1p51	D	92	1jnm	A	56	1s9k	F	52	1wdl	B	706	1wdl	A	715
1blb	D	183	2bb2	-	181	1tqb	C	219	1yeg	L	219	1s1c	Y	70	1s1c	X	69	1osj	A	345	1dr0	B	346

Table B.1: The DynDom data set - protein pairs PDB codes

PDB1	CH	SIZE	PDB2	CH	SIZE	PDB1	CH	SIZE	PDB2	CH	SIZE	PDB1	CH	SIZE	PDB2	CH	SIZE	PDB1	CH	SIZE	PDB2	CH	SIZE
1emt	L	214	1fbi	L	214	1gig	L	210	1sm3	L	210	1ktv	B	632	1elo	-	632	1n78	B	468	1n78	A	468
1do0	C	406	1do0	A	406	1iz1	B	294	1ixc	A	284	1i6v	C	1113	1ynj	C	1114	1hpu	C	525	1oi8	A	525
1pwe	D	315	1pwh	D	327	1b3r	D	428	1li4	A	430	1iz1	B	294	1iz1	Q	292	1hzh	L	215	1n0x	L	215
1gvm	F	134	2bml	B	126	1v7c	C	351	1uin	A	350	2dl2	A	197	1m4k	A	193	1f58	H	228	2f58	H	228
1ztm	A	416	1ztm	B	418	2lao	-	238	1lst	-	239	1nam	H	275	1mwa	I	274	1e0o	D	196	1e0o	B	197
1ooa	A	313	1vkx	B	312	1jyu	A	96	1jyr	A	96	1rzi	F	217	1rzi	H	216	1yqv	L	211	1e6j	L	210
1nui	B	239	1nui	A	242	1a81	C	254	1a81	A	254	1tfy	D	437	1tfw	C	437	1dqq	B	210	1dqm	H	209
1fos	E	60	1a02	F	53	2bq0	B	231	2bq0	A	230	1z47	A	345	1z47	B	342	1osj	A	345	1idm	-	343
1h1h	A	134	1dyt	A	133	1cb6	A	691	1lcf	-	691	1tno	K	314	1ni1	A	312	2a1w	I	225	2a77	H	219
1cts	-	437	1csh	-	435	1o75	B	402	1o75	A	399	1jl5	A	353	1g9u	A	353	2mcg	1	216	3mcg	1	216
1ykh	B	114	1yke	B	119	1eih	A	540	1rd6	A	540	1do0	C	406	1do0	B	406	1f7o	A	116	1f7d	A	118
1tcf	-	156	1tn4	-	157	1vf6	B	60	1vf6	A	58	1eqr	B	590	1c0a	A	585	1dqd	L	214	1bgx	L	210
1pc6	A	141	1pc6	B	133	1o9l	D	464	1o9l	A	468	1nsn	H	210	1ors	B	221	1c04	D	122	1whi	-	122
1fpo	C	157	1fpo	A	171	1opl	B	365	1opj	B	288	1ex6	B	186	1ex7	A	186	1mt0	A	241	1xef	C	241
1m45	A	142	1n2d	B	145	1eia	-	207	2eia	B	204	1b3q	E	370	1b3q	A	368	1cu1	A	645	1hei	A	443
1i2d	B	572	1m8p	A	573	1ya0	B	464	1ya0	A	458	1wsj	E	153	1wsg	C	151	1cd3	4	146	1al0	3	140
1dva	M	101	1fak	L	108	1pbq	B	266	1pb7	A	281	1rve	A	244	1rve	B	244	1g7t	A	569	1g7s	A	576
1qvr	A	803	1qvr	C	803	1nyt	B	271	1nyt	D	270	1clz	L	219	1tet	L	216	1z8u	D	135	1yzi	A	141
1f59	B	440	1gcj	A	447	1ku1	B	211	1ku1	A	211	3ktq	A	539	5ktq	A	534	1ek8	A	185	1ise	A	184
2pcp	B	215	2pcp	D	215	1f9n	C	148	1f9n	F	149	1qxx	A	76	1u07	A	90	1rki	A	101	1rki	B	97
1hk7	B	247	1usv	E	251	1nc2	D	215	1nc2	B	218	1hv9	B	450	1fwy	B	324	1tim	B	247	1ssg	A	247
1jmc	A	238	1fgu	B	238	1br4	H	148	1br1	F	148	1txv	B	440	1u8c	B	599	1m7h	C	204	1m7g	C	209
1kit	-	757	1w0p	A	753	1cnz	A	363	1cm7	A	363	1p7q	D	183	1ugn	A	186	1flz	A	228	1eui	A	221
1d2h	D	252	1d2h	A	252	1sg6	B	378	1dqs	B	380	1wwj	A	99	1r5p	B	93	1ae6	L	219	1uz8	L	218
1sg2	A	141	1u2m	C	137	1kht	A	190	1kht	C	191	1v9m	A	312	1r5z	B	320	1o9x	A	579	1e7e	A	582
1qs8	B	328	1miq	B	375	1vq3	B	84	1vq3	C	87	1s9h	B	267	1u0j	A	261	1p5v	B	136	1p5u	C	130
1c0a	A	585	1il2	A	585	1qlf	A	276	1ffc	A	273	1i5s	A	330	1vfv	A	324	1n78	B	468	1gln	-	468
1kyw	A	350	1kyz	E	361	1pmz	D	82	1pmz	C	84	1knx	E	310	1knx	A	303	1k9a	B	441	1k9a	F	436
1w24	A	180	1z2w	A	182	1zav	A	178	1zax	A	174	1rve	A	244	4rve	B	240	1br9	-	182	1bqq	T	184
1ira	Y	311	1itb	B	310	1opo	C	268	1opo	A	267	1e3o	C	132	1hf0	B	128	1nyr	B	637	1nyr	A	642
1qyr	B	252	1qyr	A	252	1pg7	I	213	1jps	H	213	1lil	A	212	1lil	B	212	1si4	D	146	1uiw	D	146
1jft	A	340	1dbq	A	276	1a49	D	519	1aqf	H	519	1iai	L	214	12e8	M	214	1f4q	A	161	1k94	B	165
1g8g	A	510	1r6x	A	386	1dn0	D	217	1qlr	B	209	1iq5	A	144	1xa5	A	144	1ii6	A	340	1yrs	A	330
1k90	B	465	1k90	A	485	2rcs	H	217	1gaf	H	217	1u12	B	127	1vp6	C	133	1i3q	A	1414	1twf	A	1419
1gh7	A	408	1gh7	B	408	1wkq	A	158	1tiy	A	155	1s7k	A	158	1z9u	A	172	1s1e	A	181	1s6c	A	165
1fmt	B	307	2fmt	A	314	2vit	A	210	1mfe	L	211	1op3	K	211	1op5	K	212	1wdl	B	706	1wdm	B	710
1qvr	A	803	1qvr	B	803	1tzh	B	222	1fvd	B	223	1t04	C	214	1s78	C	214	1bey	L	214	1t04	C	214
1vf7	A	237	1vf7	C	246	1e0d	A	429	4uag	A	428	1g0y	R	310	1ira	Y	311	1ck7	A	619	1gxd	B	623
1ggi	H	215	1ggc	H	215	1h9g	A	223	1hw1	B	226	1pvd	A	537	1pvd	B	537	4prg	D	270	3prg	A	267
1dva	M	101	1qfk	L	96	1tij	A	114	1tij	B	112	1m4u	L	112	1lxi	A	104	1ovz	A	193	1ow0	D	190
1v5b	A	356	2ayq	B	357	1hi6	A	214	1cfs	A	214	1xwv	B	129	1xwv	A	129	1hi3	A	135	2bex	C	135
1ym7	C	607	1omw	A	614	1ltl	C	239	1ltl	F	239	4at1	D	146	1q95	K	153	1fcf	A	387	1fc9	A	386
1ebh	B	439	1one	A	436	1fzr	C	129	1m0d	A	129	1k40	A	126	1k04	A	142	1eix	C	232	1l2u	A	224
1ha3	A	391	1aip	A	373	1nvm	F	309	1nvm	B	312	1j19	A	316	1gc7	A	297	1hmp	B	209	1bzy	D	214
1qo3	A	274	1ddh	A	274	1g8l	A	403	1g8l	B	403	1yvr	A	520	1yvp	B	532	1mp4	B	289	1h5r	C	290
1ny5	B	385	1ny5	A	384	1qgt	B	143	1qgt	C	142	1cr9	L	219	1yed	L	219	3hfm	H	215	1nak	H	214
1f3u	F	154	1f3u	D	153	1kcr	H	218	1kcu	H	217	1hq4	A	215	1ken	L	213	1na6	B	395	1na6	A	365
1jnl	L	211	1xf3	A	213	1m1j	F	389	1m1j	C	390	1bjm	A	216	4bjl	A	216	1rym	A	185	1ryb	A	186
2bgw	A	219	2bhn	D	211	1r8j	A	272	1r8j	B	264	1tbp	A	180	1nh2	A	180	1yqy	A	514	1pwq	A	732
1eej	B	216	1tjd	A	216	1qdn	C	190	1qcs	A	190	1fch	A	302	1fch	B	297	1jt6	A	186	1jt0	D	185
1y50	A	87	1y51	C	87	1xr9	A	276	1a9e	A	277	1eym	A	107	1j4r	A	107	1iq5	A	144	1g4y	R	147
1jqj	D	325	1jqj	C	328	1n8z	B	220	1tzh	B	222	1j09	A	468	1n78	B	468	8ohm	-	435	1cu1	A	645
1q12	B	367	1q1b	B	367	1tjl	A	145	1tjl	C	145	1for	L	210	1ots	F	211	1xwl	-	580	1lv5	B	580
1t08	A	510	1jdh	A	508	1q2w	B	297	1uk2	B	302	1n1e	A	349	1evy	A	346	1fc9	A	386	1fc6	A	382
1d6s	A	322	1fcj	B	304	1yy9	D	220	1yy8	D	221	1iai	M	215	1yqv	L	211	1tfe	-	142	1aip	C	195
1h0d	A	216	1afv	L	217	1rq0	C	314	1rq0	A	319	1fbi	L	214	1fai	L	214	1w80	A	248	1kyu	A	247

REFERENCES

1. Janin, J., "Assessing predictions of protein-protein interaction: the CAPRI experiment", *Protein Science*, Vol. 14, No. 2,pp. 278-283, 2005.

2. Delarue, M. and Y. H. Sanejouand, "Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the Elastic Network Model", *Journal of Molecular Biology*, Vol. 320, 2002.

3. Tobi, D. and I. Bahar, "Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state", *The Proceedings of the National Academy of Sciences*, Vol. 102, No. 52, pp. 18908-18913, 2005.

4. Sandak, B., H. J. Wolfson and R. Nussinov, "Flexible docking allowing induced fit in proteins: insights from an open to closed conformational isomers", *Proteins*, Vol. 32, No. 2, pp. 159-74, 1997.

5. Schneidman-Duhovny D., Y. Inbar, R. Nussinov and H. J. Wolfson, "Geometry-based flexible and symmetric protein docking", *Proteins*, Vol. 60, No. 2, pp. 224-231, 2005.

6. Delarue, M., and Y. H. Sanejouand, "On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models", *The Proceedings of the National Academy of Sciences*, Vol. 101, 2004.

7. Ming, D ., Y. F. Kong, A. M. Lambert, Z. Huang and J. P. Ma, "How to describe protein motion without amino acid sequence and atomic coordinates", *The Proceedings of the National Academy of Sciences*, Vol. 99, 2002.

8. Tama, F., W. Wriggers and C. L. Brooks, "Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory", *Journal of Molecular Biology*, Vol. 321, 2002.

9. Kovacs, J. A., P. Chacon and R. Abagyan, "Predictions of protein flexibility: first-order measures", *Proteins*, Vol. 56, No. 4, pp. 661-668, 2004.

10. Boutonnet, N., M. Rooman, S. Wodak, "Automatic analysis of protein conformational changes by multiple linkage clustering", *Journal of Molecular Biology*, Vol. 253, pp. 633-647, 1995.

11. Shatsky, M., R. Nussinov and H. J. Wolfson, "Flexible protein alignment and hinge detection", *Proteins*, Vol. 48, pp. 242-256 , 2002.

12. Jacobs, D. J., A. J. Radler, L. A. Kuhn and M. F. Thorpe, "Protein Flexibility predictions using graph theory", *Proteins*, Vol. 44, pp. 150-165, 2001.

13. Thorpe, M. F., M. Lei, A. J. Rader, D. J. Jacobs and L. A. Kuhn, "Protein flexibility and dynamics using constraint theory", *Journal of Molecular Graphics and Modelling*, Vol. 19, pp. 60-69, 2001.

14. Lee-Wei, Y., X. Liu, C. J. Jursa, M. Holliman, A.J. Rader, H. Karimi and I. Bahar, "iGNM: A Database of Protein Functional Motions Based on Gaussian Network Model", *Bioinformatics*, Vol. 21, pp. 2978-2987, 2005.

15. Suhre, K. and Y. H. Sanejouand, "ElNemo: a normal mode web-server for protein movement analysis and the generation of templates for molecular replacement", *Nucleic Acids Research*, Vol. 32, pp. 610-614, 2004.

16. Hollup, S. M., G. Salensminde and N. Reuter, "WEBnm@: a web application for normal mode analysis of proteins", *BMC Bioinformatics*, Vol. 11, 2005.

17. Flores, Echols, Milburn, Hespenheide, Keating, Lu, Wells, Yu, Thorpe and Gerstein, "Molecular Movements Database", *Nucleic Acids Research*, Vol. 34, pp. 296-301, 2006.

18. Shatsky M., H. J. Wolfson and R. Nussinov, "Flexible protein alignment and hinge detection", *Proteins: Structure, Function, and Genetics*, Vol. 48, pp. 242-256, 2002.

19. Qi, G., R. Lee and S. Hayward, "A comprehensive and non-redundant database of protein domain movements", *Bioinformatics*, Vol. 21, No. 12, pp. 2832-2838, 2005.

20. Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, "The Protein Data Bank", *Nucleic Acids Research*, Vol. 28, pp. 235-242, 2000.

21. Go, N., T. Noguti and T. Nishikawa, "Dynamics of small globular proteins in terms of low-frequency vibration modes", *The Proceedings of the National Academy of Sciences*, Vol. 80, pp. 3696-3700, 1983.

22. Kitao, A. and N. Go, "Investigating protein dynamics in collective coordinate space", *Current Opinion in Structural Biology*, Vol. 9, pp. 164-169, 1999.

23. Ma, J., "Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes", *Current Opinion in Structural Biology*, Vol. 13, pp. 373-380, 2005.

24. Bahar, I. and A. J. Rader, "Coarse-grained normal mode analysis in structural biology", *Current Opinion in Structural Biology*, Vol. 15, pp. 586-592, 2005.

25. Bahar, I., A. R. Atilgan and B. Erman, "Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential", *Folding and Design*, Vol. 2, pp. 173-181, 1997.

26. Krebs, W. G., V. Alexandrov, C. A. Wilson, N. Echols, H. Yu and M. Gerstein, "Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistics", *Proteins*, Vol. 48, pp. 682-695, 2002.

27. Haliloglu, T., I. Bahar and B. Erman, "Gaussian Dynamics of Proteins", *Physical Review Letters*, Vol. 79, pp. 3090-3093, 1997.

28. Atilgan, A. R., A. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin and I.

Bahar, "Anisotropy of fluctuation dynamics of proteins with an elastic network model", *Biophysical Journal*, Vol. 80, pp. 505-515, 2001.

29. Ben-Dor, A., R. Shamir and Z. Yakhini, "Clustering gene expression patterns", *Journal of Computational Biology*, Vol. 6, pp. 281-297, 1999.

30. Shatsky, M., R. Nussinov and H. J. Wolfson, "A method for simultaneous alignment of multiple protein structures, *Proteins:Structure, Function, and Bioinformatics*, Vol. 56, pp. 143-156, 2004.