Identification of Ligand Binding Sites of Proteins Using The Gaussian Network Model

by

Ceren Tüzmen

A Thesis Submitted to the

Graduate School of Engineering

in Partial Fulfillment of the Requirements for

the Degree of

Master of Science in

Computational Sciences and Engineering

Koc University

September 2010

Koc University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Ceren Tüzmen

and have found that it is complete and satisfactory in all respects,

and that any and all revisions required by the final

examining committee have been made.

Committee Members:

_____

Prof. Burak Erman (Advisor)

_____

Prof. Türkan Haliloğlu

_____

Assoc. Prof. Özlem Keskin

Date: _____

To my family,

ABSTRACT

Biomolecular interactions play key roles in biological activity. Investigation of those interactions, including protein-ligand interactions, is crucial for understanding the way that nature designed its biological machinery. Ligand binding particularly requires, recognition of the ligand by the protein, which in turn arranges the three dimensional structure of the protein, mostly directed by the energetic interactions involved. Based on these requirements, ligand-binding has been considered as a local process. Yet, it has been recently shown that ligand binding depends not on the local structure, but rather on an interaction pathway, that takes part in rearrangement of the protein into the most favorable conformation upon binding.

The nonlocal nature of the protein-ligand binding problem is investigated via the Gaussian Network Model with which the residues lying along interaction pathways in a protein and the residues at the binding site are predicted. The predictions of the binding site residues are verified by using several benchmark systems where the topology of the unbound protein and the bound protein-ligand complex are known. Predictions are made on the unbound protein. Agreement of results with the bound complexes indicates that the information for binding resides in the unbound protein. Cliques that consist of three or more residues that are far apart along the primary structure but are in contact in the folded structure are shown to be important determinants of the binding problem.

Comparison with known structures shows that the predictive capability of the method is significant.

ÖZET

Biyomoleküler etkileşimler biyolojik aktivelerde önemli görevler üstlenmektedir. Protein-ligand etkileşimleri de dahil olmak üzere, bu etkileşimlerin araştırılması, biyolojik sistemlerin çalışma mekanizmasını anlayabilmek ve çözümleyebilmek açısından büyük önem taşır. Ligand protein etkileşmeleri sonucunda, ligand proteine bağlanırsa, proteinin konformasyonunda, bu süreçte görev alan enerjetik etkileşimler tarafından yönetilen bir takım değişiklikler meydana gelir. Bu değişimler göz önünde bulundurulduğunda, bu bağlanma süreci lokal bir süreç olarak düşünülmüştür. Ancak, yakın zamanda yapılan araştırmalar göstermiştir ki; ligandın proteine bağlanması esnasında, lokal etkileşimler değil, protein üzerinde varolan ve proteinin kalıntılarını, protein-ligand kompleksini en uygun hale getirecek şekilde düzenleyen bir etkileşim yolağı rol oynamaktadır.

Bu çalışmada, lokal olmayan etkileşim yolağını belirlemek ve bağlanma yerindeki kalıntıları öngörmek amacıyla Ağ Yapı Modeli adı verilen bir model kullanılmıştır. Ek olarak, birincil yapıda birbiriden uzak olup, katlanmış yapıda birbiriyle etkileşen üç veya daha fazla kalıntıdan oluşan kliklerin de bağlanmayı önemli ölçüde tayin ettikleri gösterilmiştir. Proteinin liganda bağlı olmadığı duruma ait kristal yapılar kullanılarak öngörülen sonuçları doğrulamak için, aynı proteinin liganda bağlı olduğu haldeki kristal yapıdan edinilen bilgiler kullanılmıştır.

Bilinen sistemlerin karşılaştırılmasıyla elde edilen sonuçların uyuşması göstermektedir ki, bağlanmayla ilgili bilgi, serbest halde bulunan proteinin yapısı kullanılarak elde edilebilmektedir ve kullanılan metot bağlanma yeri kalıntılarını tahmin etmekte önemli ölçüde başarılı olmuştur.

ACKNOWLEDGEMENTS

I owe my deepest gratitude to my advisor Prof. Burak Erman for his guidance, encouragement and incredible patience throughout the development of this thesis. It was a great pleasure for me to have the opportunity of meeting him and working with him.

I am extremely grateful to my thesis committee members Prof. Türkan Haliloğlu and Assistant Prof. Özlem Keskin for sparing their precious time for the critical reading of my thesis and for their valuable comments.

Special thanks go to Ayşe Küçükyılmaz, Gözde Özbek and Şeyda İpek, for their warm, true and lasting friendships. I feel lucky to find you and I know that I will never lose you. I should also thank to my oldest friend, Ekin Ertemiz for being by my side whenever I need, accompanying me in good and bad. I am grateful to Mert Gür, for his endless support and his scientific assistance during the past two years.

I thank to my officemates, Besray Ünal, Özge Engin, Beytullah Özgür, Çiğdem Sevim and Bora Karasulu for making the office such a nice environment. I would like thank to my old housemates Gözde and Senay, to my old officemates Gözde, Mali and Hakan and to all my grad friends that made Koç such a joyful place.

Finally, I would like to thank to my parents Yalçın & Emine Tüzmen and my sister Cansu Tüzmen, for their love, guidance and continuous support all through my life. I dedicate this thesis to them.

Table of contents

1 Introduction 1

2 Overview 4

2.1. Literature Review For Selected Systems 4

2.1.1. Oxireductases: Human Heme-Oxygenase-1 4

2.1.2. Transferases: Human Glutathione Transferase A1-1 5

2.1.3. Hydrolases: Catalytic Domain Of Protein Tyrosine Phosphatase 1b 7

2.1.4. Ligases: Biotin Carboxylase Domain Of Acetyl Co-A Carboxylase 2 9

2.1.5. Lyases: Human Carbonic Anhydrase Ii 10

2.1.6. Ca^{+2} Binding S100a6 12

3 Methods 17

3.1. Gaussian Network Model 17

3.2. Graph Theory And Cliques 19

3.3. Formulation Of The Problem 20

4 Results And Discussion 27

4.1. Human Heme-Oxygenase-1 27

4.2. Human Glutathione Transferase A1-1 30

4.3. Human Protein Tyrosine Phosphatase 1b 33

4.4. Biotin Carboxylase Domain Of Acetyl-Coa Carboxylase 2 36

4.5. Human Carbonic Anhydrase Ii 39

4.6. S100A6 42

5 Conclusion 45

Bibliography 47

Supplemantary 1 57

Supplemantary 2 59

Vita 71

List of Figures

2.1 Ribbon representation of heme bound human heme oxygenase-1 (pdb ID: 1N3U). 4

2.2 Ribbon representation of human glutathione S-transferase A1-1 in complex with two S-benzyl-glutathiones (pdb ID: 1GUH). 6

2.3 Ligplot of interactions of catalytic domain of protein tyrosine phosphatase 1b with ligand (pdb ID: 1bzc) 8

2.4 Ribbon representation of the catalytic domain of protein tyrosine phosphatase 1B with ligand (pdb ID: 1bzc) 8

2.5. Ribbon representation of biotin carboxylase domain of acetyl co-A carboxylase 2 complexed with Soraphen A (pdb ID: 3gid) 9

2.6. Ligplot showing the interactions of carbonic anhydrase complexed with TPD (pdb ID: 1bnw) 11

2.7. Ribbon representation of carbonic anhydrase complexed with TPD and Zn^{+2} (pdb ID: 1bnw) 11

2.8. Ribbon representation of s100A6 with Ca^{+2} ions (pdb ID: 1k9p) 12

4.1 a) Total correlation C_T of residues as a function of residue indices b) Contour plot of distance fluctuations $\left\langle \left(\Delta R_{ij} \right)^2 \right\rangle$ of 1NI6.pdb. 27

4.2 Three dimensional structure of human HO-1 B chain with Heme 28

4.3 Identified interaction path residues and cliques of HO-1 29

4.4 a) Total correlation C_T of residues as a function of residue indices. b) Contour plot of distance fluctuations $\left\langle \left(\Delta R_{ij} \right)^2 \right\rangle$ of 1K3O.pdb.. 30

4.5. Three dimensional structure of human GST A1-1 with S-benzyl-glutathione..... 31

4.6 Interaction path residues and cliques of GST A1-1. 32

4.8 Three dimensional structure of human PTP 1B with TPI..... 34

4.9 Interaction path residues and the cliques of PTP 1B 35

4.10 a) Total correlation C_T of residues as a function of residue indices. b) Contour plot of distance fluctuations $\left\langle \left(\Delta R_{ij} \right)^2 \right\rangle$ of 3GLK.pdb. 36

4.11 Three dimensional structure of BC domain of ACC2 with Sarophen A. 37

4.12 Interaction path residues and the cliques of ACC2 37

4.13 a) Total correlation C_T of residues as a function of residue indices. b) Contour plot of distance fluctuations $\left\langle \left(\Delta R_{ij} \right)^2 \right\rangle$ of 2CBE.pdb. 39

4.14 Three dimensional structure of Carbonic anhydrase II with Brinzolamide and Zn^{+2}. .. 40

4.15 Interaction path residues and cliques of Carbonic anhydrase II. 40

4.16 a) Total correlation C_T of residues as a function of residue indices. b) Contour plot of distance fluctuations $\left\langle \left(\Delta R_{ij} \right)^2 \right\rangle$ of 1K9P.pdb. 42

4.17 Three dimensional structure of S100A6 with Ca^{+2} ions. 43

4.15 Interaction path residues and cliques of S100A6. 44

S1 Alcohol Dehydrogenase 59
S2 Ispc 59
S3 Adenosine Kinase 600
S4 Map Kinase P38-α 60
S5 Kinase Domain TRP- Ca^{+2} Channel 61
S6 Cyclin-Dependent Kinase 61
S7 M-phase inducer phosphatase 2 (Cdc25b) 62
S8 Angiogenin 62
S9 Carboxypeptidase A 63
S10 Gamma Chymotyrpsin 63
S11 Glyoxalase I 64
S12 Lysozyme 64
S13 Tyrosyl-DNA phosphodiesterase 65
S14 Beta-lactam synthetase 65
S15 Vacuolar protein sorting Protein29 (VPS29) 66
S16 Phospholipase C 66
S17 Pancreatic α-amylase 67
S18 Ferrochelatase 67
S19 Hydroxynitrile Lyase 68
S20 Adipocyte Lipid binding Protein 688
S21 Copper Resistance Protein 69
S22 L-leucine Binding Protein 69
S23 Fibrillin-1 70
S24 TNF receptor associated factor (Traf 6) 700

List of Tables

2.1 Test Set Proteins with PDB codes for both ligand-free and ligand-bound structures, and chain IDs used in calculations.. 16

S1 Summary of results for the whole data set.. 57

Chapter 1

INTRODUCTION

A ligand can be defined as a substrate or an inhibitor of an enzyme, a hormone or a growth factor of a receptor, or an antibody of an antigen. When the ligand finds a suitable location with the right shape on the protein, the formation of a ligand-protein complex is directed particularly by the energetic interactions involved. Irrespective of whether the interaction of interest is specific or cooperated, ligand binding requires both structural recognition and energetic interplay between the two molecules. Based on these requirements, ligand binding has been considered as a local process. However, observation of both short and long range conformational changes upon binding led to the suggestion that the full topology of the protein should be taking part in the ligand binding process [1].

According to this hypothesis, binding should depend not on the local structure, but rather on an interaction pathway on the protein that takes part in the collective reorganization of the residues to accommodate for the best and most favorable conformation of the protein-ligand complex. Numerous experimental observations are in support of this hypothesis. The changes in conformation in calcium binding proteins is cited in the first comprehensive review of this phenomenon [2]. All experimental evidence points out to the fact that the full topology of the protein should take part in such rearrangements. Thus, the information needed for determining the interaction pathway should somehow be hidden in the topology.

In the simplest case, a coarse grained picture of the protein is satisfactory. The topology of the protein in this case is represented by the connectivity matrix, or the contact map, of the three dimensional structure, where the ij^{th} element of the matrix is unity if the i^{th} and j^{th} residues are in contact, and zero otherwise. Several successful models of proteins exist at this level of the topology, i.e. the residue based, coarse grained topology. One of them is the Gaussian Network Model (GNM) [3] which uses the connectivity matrix as its force constants matrix. In several recent papers [4-7], using the GNM, a statistical thermodynamics argument was proposed, that leads to the determination of structurally and functionally important residues in relation to ligand-protein interactions, as well as the "interaction path" that the protein uses in transferring information form one point to the other. The method, which we term as the 'maximum eigenvalue method' [8] is based on determining the residues that exchange energy with their neighbors and the surrounding medium. In this thesis, we present several examples where we show that these residues which are closely associated with binding are located on well defined paths.

The concept of interaction pathways or networks, in relation to ligand binding, have been addressed from different perspectives. Lockless and Ranganathan [9] suggested that correlations between two residues resulting in energy transfer among them lead to interaction paths and are evolutionarily conserved. Nelson et al proposed a relation between long range perturbations and the interaction path [10]. Pan et al [11] and Freire and collaborators [12, 13] associated conformational changes in allostery with interaction pathways. Amitai et al introduced the topological closeness measure as a determinant of interaction paths [14]. Our approach is an addition to this series of papers that emphasize the significance of topology in binding. The prediction of binding sites based on GNM is simple and easy to apply as demonstrated in our examples.

A new additional concept that we used is the 'clique', defined as a subset of three or more pairs of vertices, with each pair being connected by an edge, i.e. contacting (or interacting with) each other[15]. Cliques are expected to have great significance in protein-protein or protein-ligand interactions, as they are stiff regions, therefore likely to be conserved throughout evolution. In our test set, cliques made up of residue triads are identified since triads are frequently observed as spatial forms in the active sites of the proteins. In this thesis, we show the significance of cliques in relation to ligand binding.

Our test set is a diverse set, composed of 30 proteins most of which are enzymes catalyzing different chemical reactions. The set also holds proteins which are not engaged with an enzymatic reaction, but instead acts as a receptor or a mediator which is involved in cell-signalling pathways. For each protein, we collected both ligand-free and ligand-bound crystal structures from Protein Data Bank[16]. We identified structurally and functionally important residues by using the ligand free structures of those proteins. We combined the results obtained by the two methods and compared them with the experimental results which can be obtained using ligand bound structures.

In Chapter 2, we provide necessary background on six selected protein-ligand systems out of thirty. At the end of this chapter, there is a table (Table 2.1), which gives the names, pdb IDs and specific functions of each protein found in our test set. Chapter 3 gives information about the models that are used in this study and formulates the problem. We present the selected results for six test proteins in Chapter 5, with the discussion of the results.

The thesis is concluded with a short summary of the performed study and a short outlook of the possible future research directions that can be followed for a more in-depth study of energy transfer and interaction pathways in proteins.

Chapter 2

OVERVIEW

2.1. Literature Review for selected systems

In this section, we perform a literature survey for selected proteins included in the data set, five of which are enzymes catalyzing different chemical reactions while the other protein is involved in a cellular pathway.

2.1.1. OXIREDUCTASES: Human Heme-Oxygenase-1

Heme oxygenase (HO) is responsible from catalysis of NADPH, O_2 and degradation of heme to biliverdin with the release of iron and carbon monoxide (CO)[17]. Enzymatic reduction of biliverdin is biologically important, since the potent antioxidant bilirubin is yielded[18].

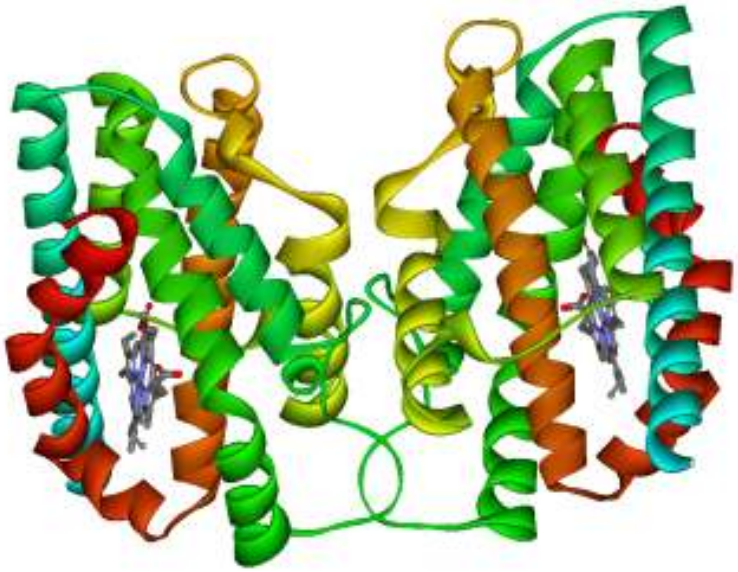


Figure 2.1 Ribbon representation of heme bound human heme oxygenase-1 (pdb ID: 1N3U).

Humans and other mammals have two isozymes; HO-1and HO-2, which are the products of separate genes. The highest level of human HO-1 is found in the spleen, where recycling of erythrocytes takes place. It is also found in liver and other tissues. Regulation of HO-1 is at the transcriptional level by porphyrins, metals, progesterone and a variety of other molecules. It functions in response to oxidative stress, ischemia, hypoxia and other disease states [19]. In addition, the absence of functional HO-1 has been found to be related with severe growth retardation, anemia and enhanced endothelial cell injury[20, 21].

In the heme bound state, human HO-1 arranges its helical shape with the help of highly conserved, distal helix residues, so that it supplies flexibility to accommodate substrate binding and product release (Figure 2.1). Human HO-1 has a dynamic active-site pocket, which is enlarged in the apo state as distal and proximal helices surrounding the heme plane move farther apart. [22] In the halo form, active site residues Thr21, Val24, Thr23, Thr26, Ala28 and Glu29, which reside on the proximal helix, and Tyr-134, Thr-135, Leu-138, Gly-139, Ser-142, and Gly-143, which reside on the distal helix, are important as they interact with heme.[23, 24]

2.1.2. TRANSFERASES: **Human Glutathione Transferase A1-1**

Glutathione S-transferases are involved in the catalysis of xenobiotics, carcinogens and conjugations with endogenous ligands. In addition, they can perform a variety of functions in metabolic pathways which are not related with detoxification, such as the intracellular storage or transport of a variety of other hydrophobic, non-substrate compounds including hormones, metabolites and drugs. Besides, due to the elevation of GST levels in tumor cells, they have been the focus of significant interest with regard to drug resistance [25-29].

In mammals, gene classes of GSTs, namely alpha, mu, pi, kappa, omega and theta are distinguished by their primary amino acid sequence homologies, tissue distribution, and substrate specificities.[29] All of the six principal structures have the same basic protein fold consisting of two domains; domains I and II which are often loosely mentioned as

the GSH and xenobiotic substrate binding domains, respectively[27, 28]. (Figure 2.2) The most observed mammalian GST are the class alpha, mu, and pi enzymes. Their regulation has been studied in detail, which is complex as they display sex-, age-, tissue-, species-, and tumorspecific patterns of expression. The expression of GSTs is also controlled by a structurally diverse range of xenobiotics and, so far, at least 100 chemicals have been identified [26].

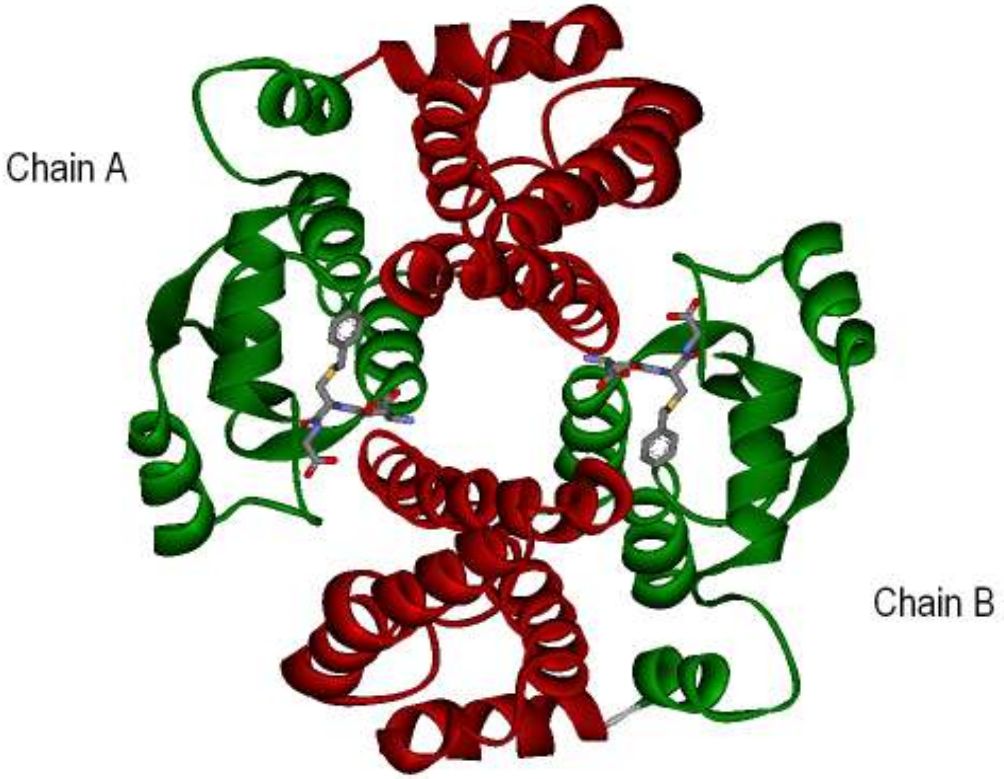


Figure 2.2 Ribbon representation of human glutathione S-transferase A1-1 in complex with two S-benzyl-glutathione**s** (pdb ID: 1GUH).

Although the catalytic mechanism of GSTs still remains unclear, it probably requires the activation of the thiol group in GSH, for nucleophilic addition to a variety of the hydrophobic substrates. In three-dimensional structures, a conserved amino acid residue, within hydrogen-bonding distance of the sulfur of glutathione with a hydroxy group, which is either a tyrosine residue, as in GST A1-1, class alpha or, less frequently, a serine residue, has been shown to be central in catalysis, likely by leading

stabilizing GSH thiol [27, 30]. In addition, in the instance of the class alpha enzymes the side chain of Arg15 is thought to be involved in the inner coordination sphere of the sulfur. The thiolate anion accepts a hydrogen bond from the seryl or tyrosyl hydroxyl group and gathers additional stabilization from positive charge of Arg15 [29].

2.1.3. HYDROLASES: Catalytic Domain of Protein Tyrosine Phosphatase 1B

The protein tyrosine phosphatases work complementarily with protein tyrosine kinases in regulating signal transduction pathways which control many physiological processes, such as cell growth or cell differentiation [31, 32].

Unlike protein tyrosine kinases which are originated from a common ancestor, protein tyrosine phosphatases display a great diversity both in structure and mechanism. Yet, they are recognized by the motif HCX_5R at their active sites, with an essential cysteine residue (Cys 215 in PTP1B) which acts as a nucleophile during catalysis [33, 34]. The active-site cleft recognizes the phosphate of the substrate and a cysteinyl-phosphate intermediate is formed. Then the phophoenzyme intermediate is hydrolyzed mediated by Gln262 and Asp181, and the phosphate ion is released. If the cysteine residue at the active-site is oxidized, it abolishes its function, so does the PTP activity. This oxidation is reversible and depends on the surroundings of the catalytic site. In PTP1B, the residues His214, Cys215 and Ser 216 have central roles in the activation of the active-site cysteine residue[35]. In Figure 2.3, there is a ligplot of PTP 1B, which shows schematically the interactions between the ligand and the protein itself. Figure 2.4 is a ribbon representation of the catalytic domain of PTP 1B in complex with an inhibitor.

Cellular pathways regulated by tyrosine phosphorylation led to the idea of the development of PTP-based therapeutics. Still, the PTPs are challenging targets for developing active-site-directed inhibitors because of the highly conserved and positively charged active-site pocket. Nevertheless, tremendous progress has been made to address potency, selectivity and bioavailability problems related to in drug discovery. [36-38]. Mainly, PTP 1B control over regulation of insulin makes it an effective target for the treatment of type II diabetes and obesity[39]. Moreover, a secondary allosteric

site has recently been discovered for PTP1B which provide a promising approach to overcome the potential challenges of targeting the active-site pocket. Numerous small-molecule inhibitors are known which bind to this site and stabilize the inactive conformation of PTP1B [40]. Since this allosteric site is not very well conserved and do not have negative charges compared to the phospho-Tyr binding active-site, it presents an alternative strategy for therapeutic development [41].

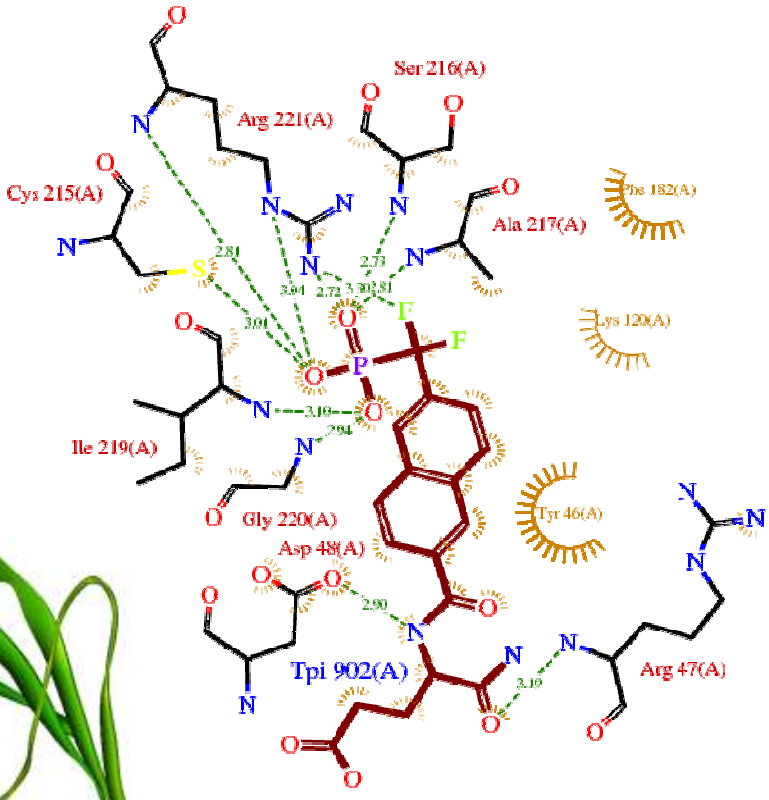


Figure 2.3 Ligplot of interactions of catalytic domain of protein tyrosine phosphatase 1b with ligand (pdb ID: 1bzc)

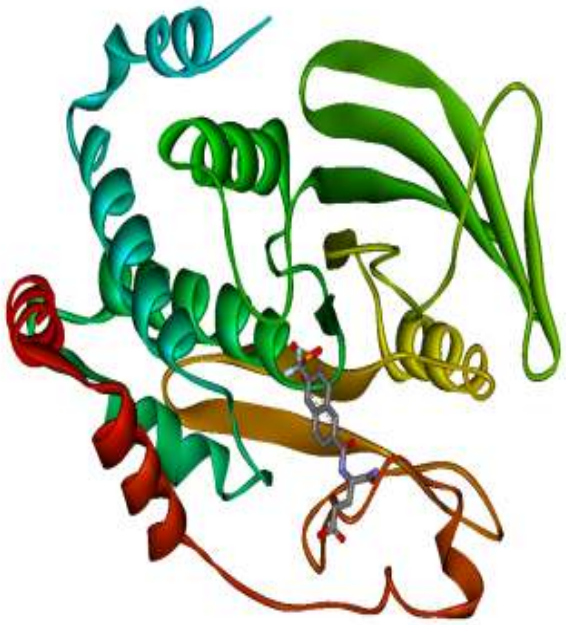


Figure 2.4 Ribbon representation of the catalytic domain of protein tyrosine phosphatase 1B with ligand (pdb ID: 1bzc)

2.1.4. LIGASES: Biotin Carboxylase Domain of Acetyl Co-A Carboxylase 2

Acetyl Co-A Carboxylase (ACC) is responsible from the biotin-dependent synthesis of malonyl-CoA, through its catalytic domains, biotin carboxylase (BC) (residues Val259-761Ala) and carboxyltransferase (CT) (residues Leu1809-Gly2305).

Since ACC has a crucial role in fatty acid metabolism, ACC has become a target for therapeutic intervention against the treatment of diseases such as type II diabetes, cardiovascular diseases and atherosclerosis, metabolic syndrome in general, and in the control of obesity [42-45].

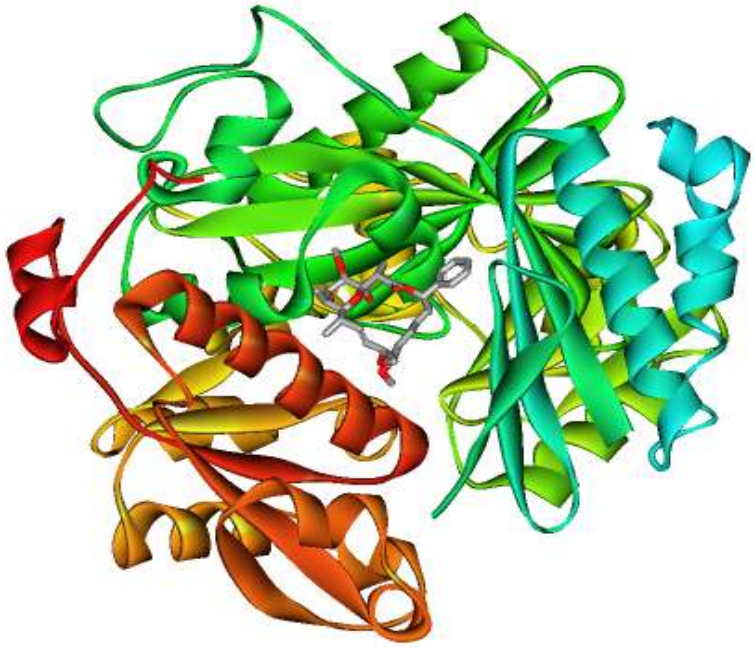


Figure 2.5. Ribbon representation of biotin carboxylase domain of acetyl co-A carboxylase 2 complexed with Soraphen A (pdb ID: 3gid)

In mammals, ACC is present in two forms; ACC1 and ACC2. ACC1 is mostly found in lipogenic tissues, such as liver or adipose, involved in biosynthesis of long-chain fatty-acids. Its product malonyl-CoA is used as a building block for extending the fatty acid chains, which will in turn converted into triacylglycerides and phospholipids.

ACC2, on the other hand, is expressed in the heart and skeletal muscle cells where it regulates the fatty acid oxidation via its malonyl-CoA product which is used as a potent inhibitor [45-49].

In ACC2-deficient mice, due to the depletion of malonyl-CoA and a continuous fatty acid oxidation, a reduced body fat mass and body weight is observed, regardless of increased consumption of food (hyperphagia). They are also less susceptible to diabetes and obesity. Therefore, the inhibitors of ACC2 may be used as novel anti-obesity drugs or therapeutic agents against the metabolic syndrome [45, 50].

Among currently known small potent inhibitors of mammalian ACCs, only Soraphen A binds to the BC domain (Figure 2.5). It is bound in an allosteric site, about 25 Å from the active site of the BC domain [51, 52]. Soraphen A has extensive interactions with the BC domain and most of the residues that are in contact with this natural product are highly conserved among the eukaryotic BC domains[45].

2.1.5. LYASES: Human Carbonic Anhydrase II

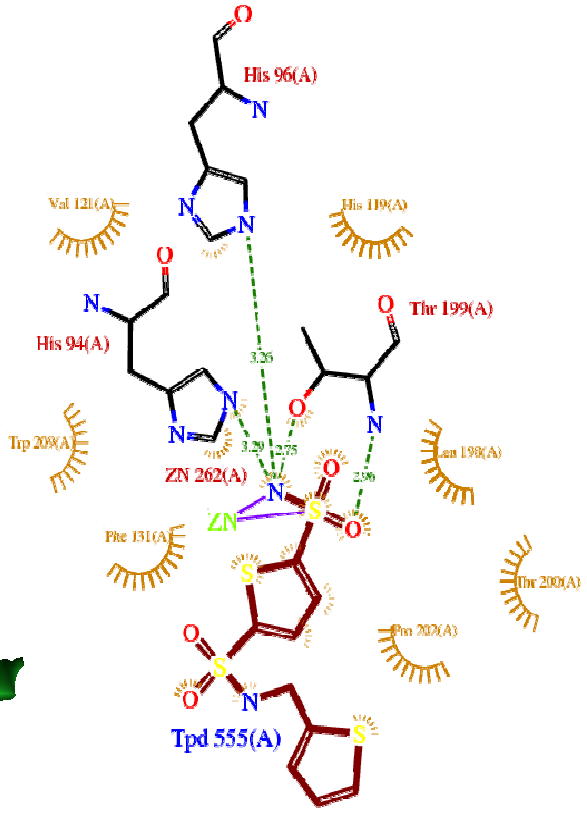
Carbonic anhydrases are found almost in all organisms, and they are used as catalysts in reversible hydration of carbon dioxides. In red blood cells, carbon dioxide reacts with water and carbonic acid, a moderately strong acid is produced. Then, it is converted into bicarbonate ion (HCO_3^-) concerted with a proton release [53].

$$CO^2 + H^2O \leftrightarrows H_2CO_3^- \leftrightarrows HCO_3^- + H^+$$

A bound zinc ion is essential for the catalytic activity of carbonic anhydrases. In humans at least seven carbonic anhydrases are present, each having its own gene, showing a substantial homology and a sequence identity. Carbonic anhydrase II, which is a major element of red blood cells, is one of the most active carbonic anhydrases and has been the most widely studied [53]. In Figure 2.7, a ribbon representation of carbonic anhydrase is given with its ligand.

The catalysis of carbon dioxide hydration by carbonic anhydrase, so the reaction rate, depends heavily on pH. Zinc ion has +2 charges in biological systems and bound to four or more ligands in carbonic anhydrases. Three coordination sites are occupied by the imidazole rings of the His residues and the forth coordination site is occupied by a water molecule or a hydroxide ion. The binding of a water molecule to the zinc ion decreases the pK_a of the water molecule to 7. At pH higher 7, water molecule with a lowered pH, loses a proton and a zinc-bound hydroxide ion (OH^-), therefore a potent nucleophile to attack carbon dioxide, is formed. As carbon dioxide is bound to the active site of carbonic anhydrase II, it reacts with the hydroxide ion and the hydroxide ion converts it into bicarbonate ion (HCO_3^-). After the release of HCO_3^-, another water molecule binds to the catalytic site. Carbonic anhydrases are mostly active at high pH values.[53]

Figure 2.6. Ligplot showing the interactions of carbonic anhydrase complexed with TPD (pdb ID: 1bnw)



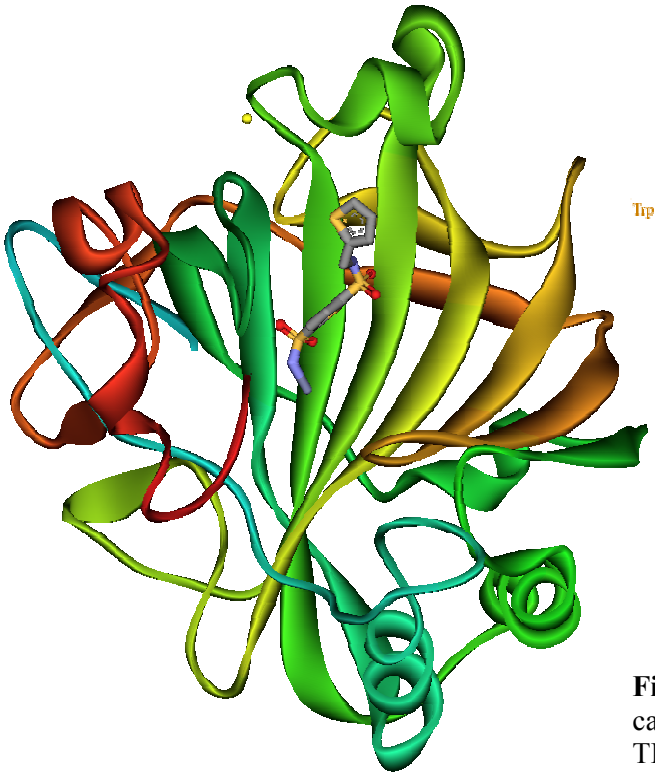


Figure 2.7. Ribbon representation of carbonic anhydrase complexed with TPD and Zn^{+2} (pdb ID: 1bnw)

Carbonic anhydrase II has evolved as a *proton shuttle* with the primary component His 64. His 64 shifts protons from the zinc-bound water molecule to the protein then to the buffer components thus regenerating an unprotonated form of the enzyme. This process has been evolved in a way of facilitating fast reproduction of the active enzyme, which is crucial for the reaction catalyzed by carbonic anhydrase II. [53] The ligplot in Figure 2.6 shows the interactions between the ligand and the carbonic anhydrase. According to this plot, the ligand is tetrahedrally coordinated by the N-atoms of His94, His96, Thr199 and the Zn^{+2} molecule.

2.1.6. Ca^{+2} BINDING S100A6

S100 proteins are small dimeric proteins which belong to the EF-hand family of calcium-binding proteins. S100 proteins are characterized by a pair of calcium-binding sites each having the helix-loop-helix structural motif (Figure 2.8). The C-terminal EF-hand (site II) contains a canonical calcium-chelation loop with flanking a-helices where calcium is ligated by the side chains of six acidic amino acids organized into a pentagonal bipyramid coordination sphere. In contrast, the N-terminal EF-hand (site I) has a typically high number of basic amino acids. In this calcium-binding loop coordination is primarily done by backbone carbonyl groups and the final ligating residue is a conserved glutamate. [54] Therefore, it has lower affinity for calcium.

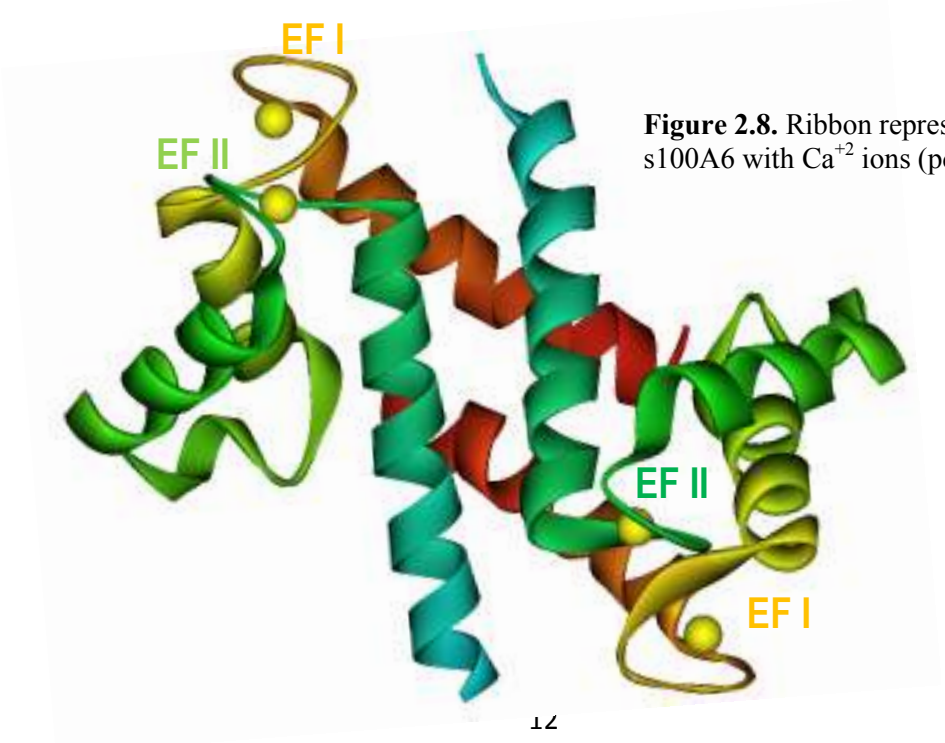


Figure 2.8. Ribbon representation of s100A6 with Ca^{+2} ions (pdb ID: 1k9p)

Upon calcium binding, S100 proteins result changes in conformation through a hand-type motion, which renders the angle between the helices of EF2 from negative to positive. [55] One of the most interesting features of the S100 proteins is their dimeric nature. For example, the 3D structure of S100B and S1000A6 exhibit a symmetric homodimeric fold which is distinctive among calcium-binding proteins. The specifity of interaction with target proteins was proposed to be on the account of the central "hinge" region with the lowest sequence homology among all S100 proteins [56, 57].

Unlike calmodulin, which is ubiquitously expressed, the expression of S100 proteins is cell and tissue-specific Most S100 genes are localized within human chromosome 1q21[58], a region which is susceptible to changes during tumor progression in transformed cells. [59] The expression of the S100A6 gene, is particularly increased in leukemia cells [22] and during the G1 phase of the cell cycle [60], which implies its role in cell cycle progression. Experiments at the protein level also show that S100A6 may be involved in cell growth, cell differentiation and motility[61-64].

We studied 30 different systems, 6 of which we selected in the present study. These are given in Table 2.1. The last two columns of Table 1 give the pdb codes of the ligand free and ligand bound structures. In all our calculations, we perform the predictions on the ligand-free structure and compared the results using the ligand-bound structure.

DEFINITION	NAME OF THE PROTEIN	PDB Code/Chain ID	
		Ligand-free state	Ligand-bound state
Oxireductases	Alcohol Dehydrogenase	1E3E/A	1E3I/A
	Heme oxygenase- 1	1NI6/B	1N3U/B
	Ispc	1ONN/A	1ONP/A
Transferases	Adenosine Kinase	2PKF/A	2PKK/A
	Glutathione S-transferase	1K3O/A	1K3Y/A
	Map Kinase P38-α	1WFC/A	1OVE/A
	Kinase Domain TRP-Ca Channel	1IAJ/A	1IAH/A
	Cyclin-Dependent Kinase	1HCL/A	1HCK/A
Lyases	Carbonic Anhydrase II	2CBE/A	1A42/A
	Ferrochelatase	1AK1/A	1C1H/A
	Hydroxynitrile Lyase	1DWO/A	1DWP/A

Hydrolases	M-phase inducer phosphatase 2 (Cdc25b)	1CWR/A	1CWS/A
	Angiogenin	1ANG/A	1GV7/A
	Carboxypeptidase A	5CPA/A	7CPA/A
	Gamma Chymotyrpsin	2GCH/B and C	1AB9/B and C
	Glyoxalase I	1FA8/A	1FA5/A
	Lysozyme	1REX/A	1REY/A
	Tyrosyl-DNA phosphodiesterase	1JY1/A	1MU7/A
	Beta-lactam synthetase	1M1Z/A	1MB9/A
	Protein Tyrosine Phosphatase (PTP1B)	2HNP/A	1BZC/A
	Vacuolar protein sorting Protein29 (VPS29)	1Z2X/A	1Z2W/A
	Phospholipase C	1PTD/A	1PTG/A
	Pancreatic α-amylase	2QMK/A	2QV4/A
Ligases	Acetyl-CoA carboxylase (ACC)	3GLK/A	3GID/A

Non-enzymes	Adipocyte Lipid binding Protein	1G7N/A	1G74/A
	Ca-Binding S100A6	1K9P/A	1K9K/A
	Copper Resistance Protein	3DSP/A	3DSO/A
	L-leucine Binding Protein	1USG/A	1USK/A
	Fibrillin	1UZQ/A	1UZJ/A
	TNF receptor associated factor (Traf 6)	1LB4/A	1LB5/A

Table 2.1 The names of each protein in our test set are given with the specific reactions they catalyze. PDB codes for both ligand-free and ligand-bound states, and chain IDs used in calculations are given in the last two columns. The first column indicates the function of the protein.

Chapter 3

METHODS

In this chapter, we provide information about the two methods applied to the test set. In the last section, we formulate the problem.

3.1. Gaussian Network Model

The Gaussian network model (GNM) is an elastic network (EN) model, first proposed in 1996 by Tirion[65] at the atomic level and reconsidered at the residue level by Bahar, Atilgan, Haliloglu and Erman[66, 67] one year later. Inspired by the work of PJ Flory on polymer networks [68] and other works that utilized normal mode analysis (NMA) and simplified harmonic potentials, the model was introduced to study, understand and characterize dynamics of a biological macromolecule[69]. It has found wide spread use from small systems, such as single domain proteins, to large macromolecular assemblies including RNA polymerase or a viral capsid.

The Gaussian network model is a coarse-grained model in which the proteins are introduced at the amino acid level, represented by nodes corresponding to their alpha carbons. The position, i.e. Cartesian coordinates, of the i^{th} C_α is denoted by R_i. A uniform spring which connects the nodes represents bonded and non-bonded interactions between the residues located within an interaction range, or cutoff distance, of r_c.

The Γ matrix of GNM is defined as

$$\Gamma_{ij} = \begin{cases} -\gamma^* & i \neq j \ \ and \ \ R_{ij} \leq r_c \\ 0 & i \neq j \ \ and \ \ R_{ij} \geq r_c \\ -\sum_k \gamma^* & i = j \neq k \end{cases} \tag{3.1.1}$$

Here, $R_{ij} = \left| R_j - R_i \right|$ is the distance between residue *i* and *j*, r_c is taken as 7 Å based on the radius of the first coordination shell around the residues observed in PDB structures. γ^* is a scaling parameter.

In the GNM, the primary interest has become determining the mean-square (ms) fluctuations of a particular residue, *i*, or the correlations between the fluctuations of two different residues, *i* and *j*. The ms fluctuations of residues are experimentally measurable (e.g., x-ray crystallographic *B*-factors, or root mean-square (rms) differences between different models from NMR), and as such, have often been used as an initial test for verifying and improving computational models and methods. Beginning with the original GNM paper [66], several applications have demonstrated that the fluctuations predicted by the GNM are in good agreement with experimental *B*-factors [70-75]. Therefore, GNM or EN models are anticipated to be useful in exploring the machinery of supramolecular structures or multi-molecular assemblies including protein–DNA, protein–protein, protein-ligand and membrane protein–lipid complexes.

In this thesis, we propose a model that extends the mechanistic description of the GNM to incorporate the role of energy fluctuations, ΔU, to determine structurally and functionally important residues in native proteins that are involved in energy exchange with a ligand and other residues along an interaction pathway. Please refer to the formulation of this problem to the last section of this chapter.

3.2. Graph theory and Cliques

One of the first results in graph theory appeared in Leonhard Euler's paper on *Seven Bridges of Königsberg*, published in 1736 [76]. It is also regarded as one of the first topological results in geometry; that is, it does not depend on any measurements. This illustrates the deep connection between graph theory and topology. The work of the physicist Gustav Kirchhoff, who published in 1845, his *Kirchhoff's circuit laws* for calculating the voltage and current in electric circuits, is such an example.

In 1852, Francis Guthrie posed the *four color problem* which asks if it is possible to color, using only four colors, any map of countries in such a way as to prevent two bordering countries from having the same color. The problem was only solved a century later by Kenneth Appel and Wolfgang Haken in 1976[77, 78], which can be considered the birth of graph theory. While trying to solve the *four color problem* mathematicians invented many fundamental graph theoretic terms and concepts.

In graph theory, graphs are represented as a set of objects that are called called "vertices" (or nodes) linked by "edges" which can be directed, i.e. assigned to a direction. There are several ways to structure the graph but the most important point is to connect correct vertices with correct edges[79, 80].

The 3D structure or the topology of a protein determines the fluctuations of its own residues. Relationships between the topology and residue fluctuations offer important clues for the function of the protein and these relationships can be conveniently understood by treating proteins as graphs of interacting residues. The contact map and the Γ matrix are used as graphs of individual proteins that can be constructed with residues as nodes and residue-residue interactions as edges. Then, those graphs can suitably be investigated by the tools of graph theory to determine important interactions in a protein.[14, 81-86]

In basic graph theory vocabulary, there is a term namely "clique" which is defined as a subset of three or more vertices, with each pair being connected by an edge, i.e. contacting (or interacting with) each other. The concept of clique comes from

Luce&Perry [15], who used complete subgraphs in social networks to model cliques of people, that is groups of people all of whom know each other. Cliques have many other applications in science and particularly in bioinformatics. In this thesis, cliques made up of residue triads are identified since triads are frequently observed as spatial forms in the active sites of the proteins. The recipe that is used to find cliques of size three will be explained in more detail in the next section.

3.3. Formulation of the Problem

The protein and its environment form a closed system with fixed energy and fixed number of molecules. But, the protein exchanges energy with the environment. The environment may contain potential ligands for the protein.

Since the total energy of the system is constant, we have

$$\Delta U_{prot} = -\Delta U_{env} \tag{3.3.1}$$

where, U_{prot} and U_{env} are the energies of the protein and its environment, respectively.

Since the protein is a small thermodynamic system, we propose the thermodynamic variables for the protein as S=Entropy, U= energy, V= Volume of protein, R= Position of the residues. These thermodynamic variables are averages and for the rest of this section they are used for the protein only, without the subscript *prot*. The instantaneous values of the energy, volume and residue positions are shown by $\hat{U}$, $\hat{V}$, $\hat{R}$, respectively. The protein exhibits energy, volume and residues fluctuations, which result from the deviations of the instantaneous extensive variables from their thermodynamic averages. These fluctuations are denoted by ΔU, ΔV and ΔR where, $\Delta U = \hat{U} - U$, $\Delta V = \hat{V} - V$ and $\Delta R = \hat{R} - R$.

In the GNM, the emphasis has been on the fluctuations ΔR, which is believed to result from coupled harmonic motions of the residues from their mean positions [3]. The present model incorporates mean-square fluctuations with energy fluctuations, ΔU.

The probability distribution $\mathrm{f}\left(\hat{\mathrm{U}}, \hat{\mathrm{V}}, \hat{\boldsymbol{R}}\right)$ of the instantaneous values, $\hat{U}$, $\hat{V}$, and $\hat{\boldsymbol{R}}$, of the energy, volume and residue positions is given by the statistical - thermodynamic expression

$$\mathrm{f}\left(\hat{\mathrm{U}}, \hat{\mathrm{V}}, \hat{\boldsymbol{R}}\right) = \exp\left\{-\mathrm{k}^{-1}\left[\mathrm{S} - \frac{\mathrm{U}}{\mathrm{T}} - \frac{\mathrm{P}}{\mathrm{T}}\mathrm{V} + \frac{\boldsymbol{F}}{\mathrm{T}} \cdot \boldsymbol{R}\right] - \mathrm{k}^{-1}\left(\frac{\hat{\mathrm{U}}}{\mathrm{T}} + \frac{\mathrm{P}}{\mathrm{T}}\hat{\mathrm{V}} - \frac{\boldsymbol{F}}{\mathrm{T}} \cdot \hat{\boldsymbol{R}}\right)\right\} \tag{3.3.2}$$

where, k is the Boltzmann constant and F is the force.

The correlation of fluctuations $\left\langle \Delta \boldsymbol{R}_{\mathrm{i}} \Delta \boldsymbol{R}_{\mathrm{j}}^{\,T} \right\rangle$ of the i^{th} and j^{th} residues are defined as

$$\left\langle \Delta \boldsymbol{R}_{\mathrm{i}} \Delta \boldsymbol{R}_{\mathrm{j}}^{\,T} \right\rangle = \sum \left(\hat{\boldsymbol{R}}_{\mathrm{i}} - \boldsymbol{R}_{\mathrm{i}}\right)\left(\hat{\boldsymbol{R}}_{\mathrm{j}} - \boldsymbol{R}_{\mathrm{j}}\right)^{T} \mathrm{f}\left(\hat{\mathrm{U}}, \hat{\mathrm{V}}, \hat{\boldsymbol{R}}\right) \tag{3.3.3}$$

Using the expression, Eq. 3.3.2, for the distribution leads to [87]

$$\left\langle \Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j^{\,T} \right\rangle = kT \left(\frac{\partial \boldsymbol{R}_i}{\partial \boldsymbol{F}_j} \right)_{T,P,F \;\; i \neq j} \tag{3.3.4}$$

In general, if Φ_k represents any of the extensive variables ΔU, ΔV, ΔR, and Ψ_k represent the conjugate variables $1/T$, $-P$, F, then, in principle, all higher moments of the extensive variables can be derived iteratively according to the rule [87]

$$\left\langle \phi \Delta \Phi_k \right\rangle = -k \frac{\partial}{\partial \Psi_k} \left\langle \phi \right\rangle - k \left\langle \frac{\partial \phi}{\partial \Psi_k} \right\rangle \tag{3.3.5}$$

where, ϕ denotes the fluctuations of the extensive variables, ΔU, ΔV, ΔR, or their product of any order. For example, letting $\phi = \Delta X_i$ and $\Delta \Phi_k = \Delta X_j^T$, since $\left\langle \Delta X_i \right\rangle = 0$, and

$$\left\langle \frac{\partial \Delta X_j}{\partial F_j} \right\rangle = \left\langle \frac{\partial \left(\hat{X}_j - X_j \right)}{\partial F_j} \right\rangle = -\left\langle \frac{\partial X_j}{\partial F_j} \right\rangle,$$

we obtain

$$\left\langle \Delta X_i \Delta X_j^T \right\rangle = kT \left(\frac{\partial X_i}{\partial F_j} \right)_{T,P,F \; i \neq j} \tag{3.3.6}$$

Higher order moments can also be obtained by a recursion relation [87].

Here, the probability function is used to derive the correlations between fluctuations of residue positions. The statistical thermodynamics interpretation of the GNM was given in full detail by Yogurtcu et al. [4], which was successfully applied to the prediction of binding sites in receptor-ligand complexes [6], of specific sites for binding[5] and the important residues along an interaction pathway[88].

In Eq. 3.3. 4, $\Delta \boldsymbol{R}_i$ presents the position vector of the C^α of the i^{th} residue and $\Delta \boldsymbol{R}_j^T$ is the transpose of the fluctuation vector of the C^α of the j^{th} residue. k is the Boltzmann constant, T is the absolute temperature. F_j is the force on the j^{th} C^α. The subscripts of the parenthesis of the right hand side indicate that the temperature, pressure and the force on each residue except the i^{th} is kept constant. Angular brackets indicate an average over all possible values of the argument. The right hand side is a thermodynamic quantity that expresses the change in the position of residues by the application of a

force. The left hand side, on the other hand denotes an average of fluctuations. Thus, this equation relates fluctuations to average quantities.

For the harmonic system, the force F_i acting on C_i^{α} is written as

$$F_i = \Gamma_{ij} R_j \tag{3.3.7}$$

where, Γ_{ij} is the force constant matrix, whose ij^{th} element is taken as a constant γ if residue *i* and residue *j* of the protein are within a cutoff distance of r_c. If the distance is larger than r_c, the ij^{th} element is zero. The diagonal element Γ_{ii} is defined as the negative sum of the i^{th} row (See Eq. 3.1.1). The correlation of fluctuations becomes

$$\left\langle \Delta R_i \Delta R_j \right\rangle = kT \left(\frac{\partial R_i}{\partial F_j} \right) = kT \left(\Gamma^{-1} \right)_{ij} \tag{3.3.8}$$

where Γ is the spring constant matrix.

The correlation matrix may be expressed in modal form as [89]

$$\left\langle \Delta R_i \Delta R_j^T \right\rangle = \sum_k \lambda_k^{-1} \left[e_k e_k^T \right]_{ij} \tag{3.3.9}$$

where, λ_k is the kth eigenvalue of the Γ matrix, $\boldsymbol{e}_k$ is the corresponding eigenvector, and $[\ ]_{ij}$ is the ij^{th} element of the enclosed matrix. . In a recent work [7], only the largest eigenvalue component of the Γ matrix is considered for a comparative study of various HLA proteins.

The mean square fluctuations of the distance between residue i and j is then written as

$$\left\langle \Delta \boldsymbol{R}^2 \right\rangle = \left\langle \left(\Delta \boldsymbol{R}_i \right)^2 \right\rangle - 2 \left\langle \Delta \boldsymbol{R}_i \ \Delta \boldsymbol{R}_j^{\ T} \right\rangle + \left\langle \left(\Delta \boldsymbol{R}_j \right)^2 \right\rangle \tag{3.3.10}$$

Eq. 3.3.8 serves as the connection between structure, represented by the Γ matrix on the right hand side and fluctuations, represented by the left hand side. This is important because a strong correlation $\left\langle \Delta U \Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j^{\,T} \right\rangle$ between the total energy uptake of the protein and the residues *i* and *j* directly indicates that the residues *i* and *j* are active in this energy transfer [5, 88]. For this, we write the energy fluctuations for the harmonic system as

$$\Delta U = F_k \Delta R_k = F_k F_j \left(\Gamma^{-1}\right)_{jk} \tag{3.3.11}$$

This expression is based on the assumption that the system is in a state of ease when the residues are at their mean positions.

Choosing $\phi = \Delta U \Delta R_i$ and using the recursion relation, Eq 3.3.5, we have

$$\left\langle \Delta U \Delta R_i \Delta R_j \right\rangle = kT \frac{\partial}{\partial F_j} \left\langle \Delta U \Delta R_i \right\rangle + kT \left\langle \frac{\partial}{\partial F_j} \left(\Delta U \Delta R_i \right) \right\rangle \tag{3.3.12}$$

We write the first term in angular brackets on the right hand side as

$$\left\langle \Delta U \Delta R_i \right\rangle = kT \frac{\partial}{\partial F_i} \left\langle \Delta U \right\rangle + kT \left\langle \frac{\partial \Delta U}{\partial F_i} \right\rangle = 0 + kT \left\langle \frac{\partial \left(F_k \Delta R_k \right)}{\partial F_i} \right\rangle \tag{3.3.13}$$

$$= kT \left\langle \delta_{ik} \Delta R_k + F_k \frac{\partial \Delta R_k}{\partial F_i} \right\rangle = 0 + kT \left\langle F_k \frac{\partial \Delta R_k}{\partial F_i} \right\rangle = kT \left\langle F_k \Gamma_{ik}^{\,-1} \right\rangle$$

$$= kT \left\langle \Delta R_i \right\rangle = 0$$

The second term is written as

$$kT\left\langle \frac{\partial}{\partial F_j}\left(\Delta U \Delta R_i\right)\right\rangle = kT\left\langle \Delta U \frac{\partial \Delta R_i}{\partial F_j}\right\rangle + kT\left\langle \Delta R_i \frac{\partial \Delta U}{\partial F_j}\right\rangle$$

$$= kT\mathbf{\Gamma}_{ij}^{-1}\left\langle \Delta U\right\rangle + kT\left\langle \Delta R_i \Delta R_j\right\rangle \tag{3.3.14}$$

Since $\left\langle \Delta U\right\rangle = 0$,

$$\left\langle \Delta U \Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j^{\,T}\right\rangle = kT\left\langle \Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j^{\,T}\right\rangle \tag{3.3.15}$$

The energy uptake of protein from the suroundings, ΔU, is coupled to the fluctuations of the residues and the residue-residue interactions according to Eq. 3.3.15. Summing both sides over the j^{th} index leads to the total coupling $C_{T,i}$ of residue *i* to its surroundings

$$C_{T,i} = \sum_j \left\langle \Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j^{\,T}\right\rangle = kT^{-1}\sum_j \left\langle \Delta U \Delta \boldsymbol{R}_i \Delta \boldsymbol{R}_j^{\,T}\right\rangle \tag{3.3.16}$$

The last term in Eq. 3.3.16 acknowledges the role of energy exchange of residue *i* with its surroundings that consist of the neighboring residues and the surroundings of the protein. Our exploratory calculations showed that there is a small dependence on the cutoff value, usually taken as 7 Å as the radius of the first coordination shell for C_α atoms. In the present study, in order to eliminate this dependence, we averaged the $C_{T,i}$ values over the interval $6.9 \le r_c \le 7.1$. The lower and upper values are selected by trial and error. If $r_c \le 6.9$, then some relevant interactions are not taken into account. If, on the other hand $r_c \ge 7.1$, then nonlocal effects that are not of interest to us are included.

In the largest eigenvalue formalism, the set of residues with finite values of $C_{T,i}$ constitute the interaction pathway. As has been shown before [88], and as will also be shown in this work, these residues are in contact with each other, in general, and constitute a path, the ends of which are exposed to the surroundings of the protein, which we termed as energy gates. Along this path lies a residue that is highly interactive with a large number of residues of the protein, and hence is referred to as the hub[88].

By its structural nature, a clique constitutes a stiff region of the protein. Considering the contact matrix A of the protein, cliques of size three are obtained according to the following recipe

$$\begin{aligned} &A_{ij} = A_{jk} = A_{ki} \quad 0 < i < j < k \le n \\ &k \ge j + c \\ &j \ge i + c \\ &4 \le c \le c_{\max} \end{aligned} \tag{3.3.17}$$

where i, j and k are residue indices, c is the residue distance (number of residues) between contacting residues, $c_{\max}$ is the upper bound of c, and n is number of residues for each protein.

We studied several different systems which are given in Table 1. The last two columns of Table 1 give the pdb codes of the ligand free and ligand bound structures. In all our calculations, we perform the predictions on the ligand-free structure and compare the results using the ligand-bound structure. In Chapter 4, we will present the results only for six selected proteins. The summary of results, related graphs and figures for all test proteins can be found in Supplementary.

Chapter 4

RESULTS and DISCUSSION

Here, we present the results for six selected proteins which are given in Table1.

4.1. Human Heme-Oxygenase-1

According to the given crystal structure (PDB Code: 1N3U), the heme binding site in the B chain of human heme oxygenase-1, contains the residues, Lys18, His25, Glu29, Gln38, Tyr134, Thr135, Gly139, Lys179, Phe207, Asn210 and Phe214. Among those, Thr21, Val24, Thr23, Thr26, Ala28 and Glu29 reside on the proximal helix while Tyr134, Thr135, Leu138, Gly139, Ser142, and Gly143 reside on the distal helix. These residues are important as they interact with heme. Phe207, Asn210 and Phe214 also lie on the proximal side of the active-site pocket. Below, we show that these specific features can be identified by applying the GNM to the apo form of the protein, i.e. 1NI6.pdb.

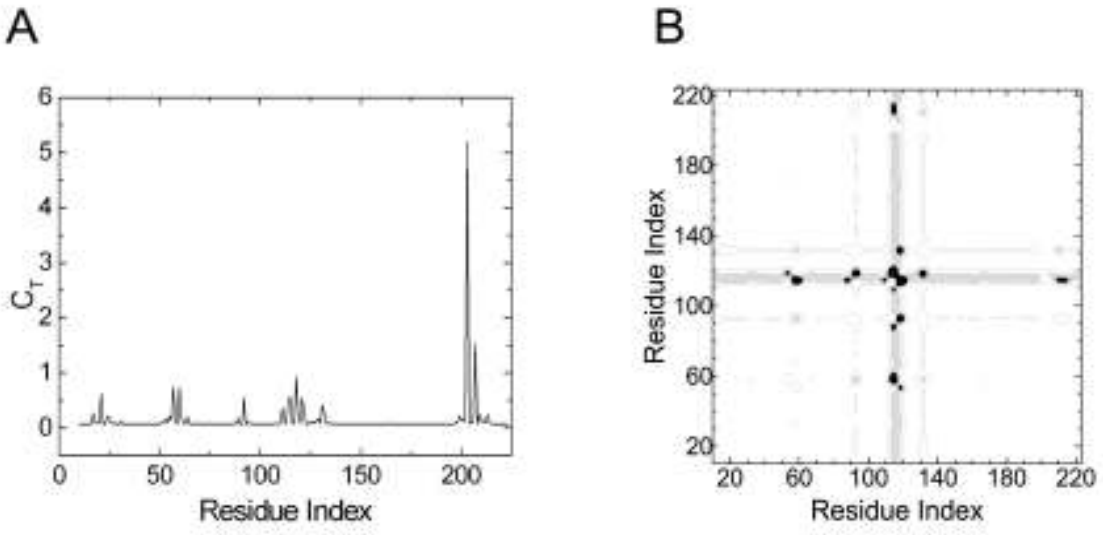


Figure 4.1 a) Total correlation C_T of residues as a function of residue indices **b)** Contour plot of distance fluctuations $\left\langle \left(\Delta R_{ij} \right)^2 \right\rangle$ of 1NI6.pdb. Highest values indicated by black.

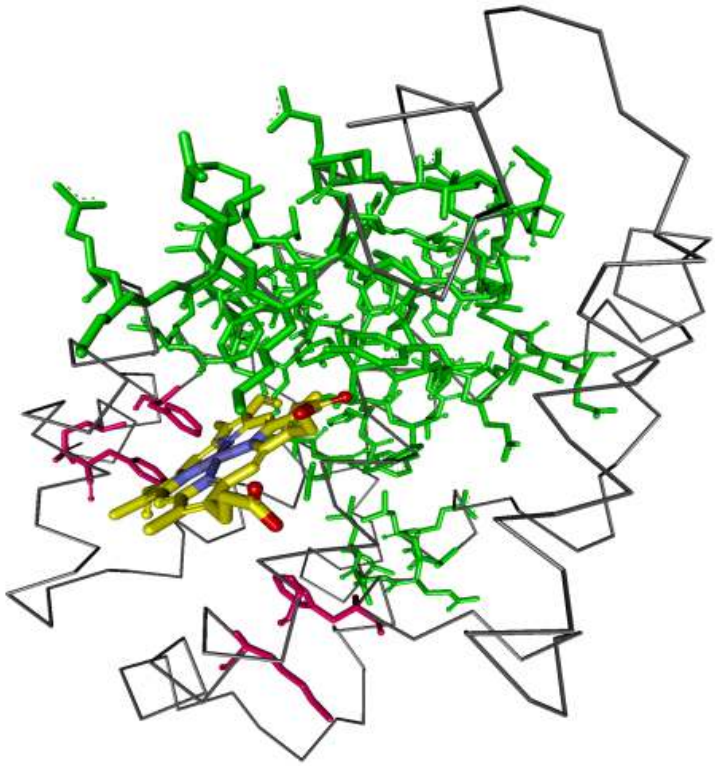


Figure 4.2 Three dimensional structure of human HO-1 B chain with Heme (yellow). Interaction path and the cliques are colored in green and pink, respectively.

Figure 4.1.a shows the total correlation, C_T, of a given residue, presented as the residue index along the abscissa, obtained by using 1NI6.pdb. In Figure 4.1b, the residues that exchange energy with the surroundings are identified with a darker hue. The heavy vertical strip shows that the residues 118-124 interact with all the residues of the protein.

In Figure 4.2.a, the ligand and the residues on the interaction path, i.e. the set of residues with finite values of C_T are shown in yellow and green, respectively. Figure 4.2.b is an enlarged version of Figure 4.2.a. Residues between 17 and 29 constituting the active site residues exhibit finite values of C_T. The path that connects the surface to the heme starts with Leu17 and Glu23 at the surface and ends at His25 that neighbors the heme. The path is colored in red and the mentioned residues are labeled in Figure 4.2.b. Residues 53-66 lie on helix H4 that contains the catalytic site Tyr58. The appearance of this region in Figure 4.1.a is mostly due to its stability, resulting from

hydrogen-bonded and electrostatic pair interactions with neighboring helix and loop structures such as Tyr58-Asp140, Glu62-Arg86, and Glu66-Tyr78[90, 91]. Similar to the Leu17-His25 path, the residues between Pro109 and Thr135 form a path, one end of which is at the surface of the protein and the other with Tyr134 and Thr135 terminates the heme.

The path that is lined by residues Pro109-Asp140 is colored in blue in Figure 2b. Finally, the largest peak corresponding to Ala203 and the second largest peak corresponding to Phe207, seen in Figure 4.1.a, are the two residues neighboring the heme. The group of residues between Val199 and Gln212 are represented as the green path, most of which neighbor the heme molecule. All of the residues observed in Figure 1a are obtained from the apo structure, indicating that the information for binding is already present in the unbound structure.

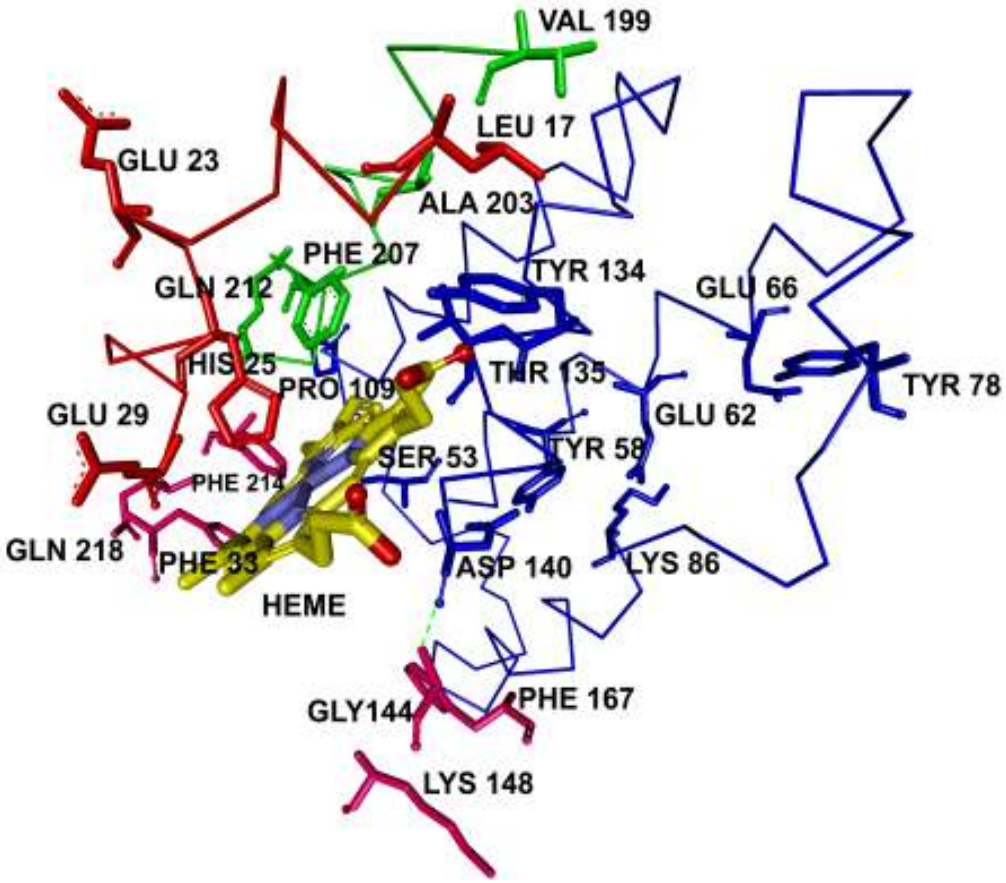


Figure 4.3 Identified interaction path residues and cliques (pink) are all labeled. Green dashed line represents the hydrogen bond between Gly144 and Asp140.

Cliques of size three are indicated in pink in this figure. These are obtained at cut-off 6.2 Å, as 33Phe-214-Phe-218Gln and Gly144-Lys148-Phe167. The first triad is located on the proximal side while the latter lies on the distal side of the Heme molecule referring to the proximal and distal helices that sandwiches the Heme molecule upon binding[92]. All of the clique residues are highly conserved. Phe214 is a binding site residue while Gly144 is a highly conserved, catalytic residue [93].

4.2. Human Glutathione Transferase A1-1

According to the given ligand-bound crystal structure 1K3Y.pdb, referred to as glutathione S-transferase (GST A1-1), S-hexyl-glutathione (GSH) binds to the site composed of the residues Tyr9-Arg45-Gln54-Val55-Pro56-Gln67-Thr68-Val111-Met208-Leu213-Phe220-Phe222 (A chain) and Ap101-Arg131 (B chain). Figure 4.3.a shows the total correlation, C_T of residues. The A chain of the unbound crystal structure, 1K3O.pdb, was used for calculations. Both chains are identical in sequence and in three dimensional structures, so are the ligands they bind. In Figure 4.3.b, the residues that exchange energy with the surroundings, i.e. energy gate residues are emphasized with a darker hue. The heavy vertical strips show that there is interaction between the all residues of the protein and the residues around the 60th residue.

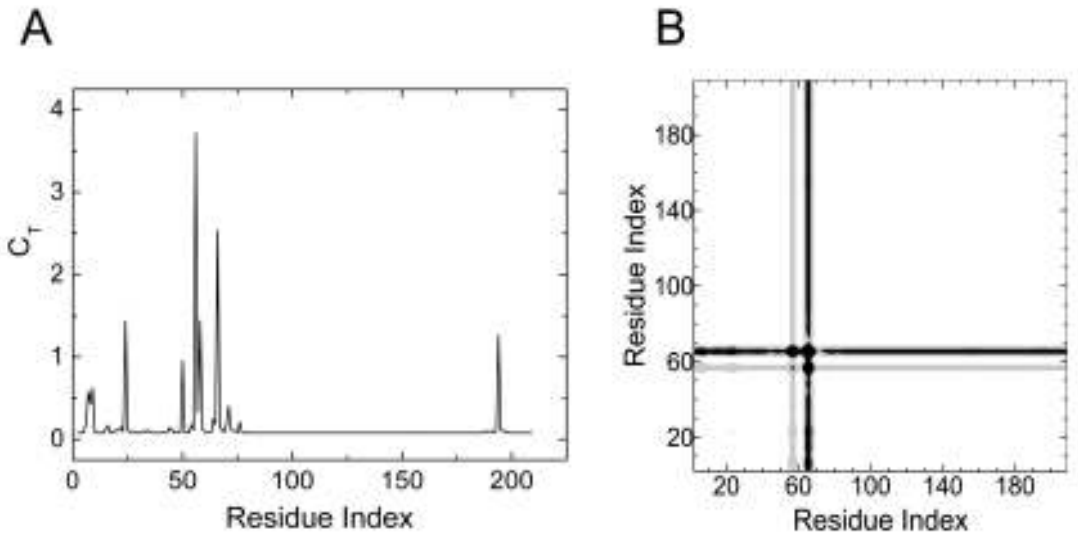


Figure 4.4 a) Total correlation C_T of residues as a function of residue indices. **b)** Contour plot of distance fluctuations $\left\langle \left(\Delta R_{ij} \right)^2 \right\rangle$ of 1K3O.pdb. Highest values indicated by black.

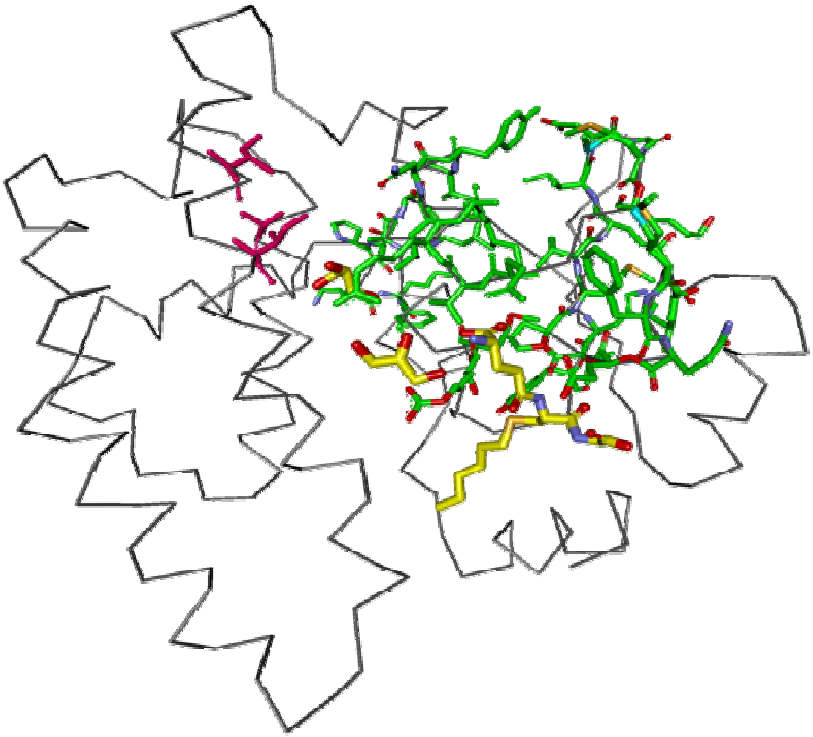


Figure 4.5. Three dimensional structure of human GST A1-1 with S-benzyl-glutathione for chain A (yellow). Interaction path and the cliques are highlighted with green and pink respectively.

Figure 4.4.a shows the ligand (yellow), the residues on the interaction path(green), i.e. the set of residues with finite values of C_T, and the clique residues (pink). In Figure 4.4.b shows all identified residues in detail. The red colored residues line a path starting with a surface residue, Lys64 and ending with the binding site residues Tyr9 and Pro56. Tyr9 is conserved among the majority of known GSTs and it is emphasized as an important catalytic residue in literature[27, 28, 30, 94]. The three-dimensional structures have shown that the hydroxyl group of Tyr9 stabilizes the thiolate of GSH through hydrogen bonding [30]. Similarly, residues Leu50-Pro56 (also shown in red) form a shorter path, which has one end at the surface and the other end at the binding site. Three highly conserved residues Gln54, Val55 and Pro56, which also interact with the ligand via hydrogen bonds, play significant roles in the stability and function of the protein[95, 96]. ARG15 and Met16 are also interacting with the residues on the Tyr9

path. Arg15 is mentioned as an important active site residue in literature as well[97]. The residues between Gln67 and Tyr74 form another path, which is presented as the green path in Figure 4b, begins at the surface and ends where Gln67 and Thr68 are positioned to participate in hydrogen bonds with the amino group and γ-glutamyl carboxyl group of glutathione, respectively[98]. It involves five conserved residues Gln67, Thr68, Ala70, Ile71 and Tyr74[93]. A member of this path, Arg69 makes three hydrogen bonds with the second glycerol molecule.

The remaining binding site residues are situated on helix 9, which is known to be highly dynamic. Since, the region is assumed to become structured and localized upon ligand binding [94, 99, 100], its electron density is unresolved for apo human GST A1-1 [101, 102].

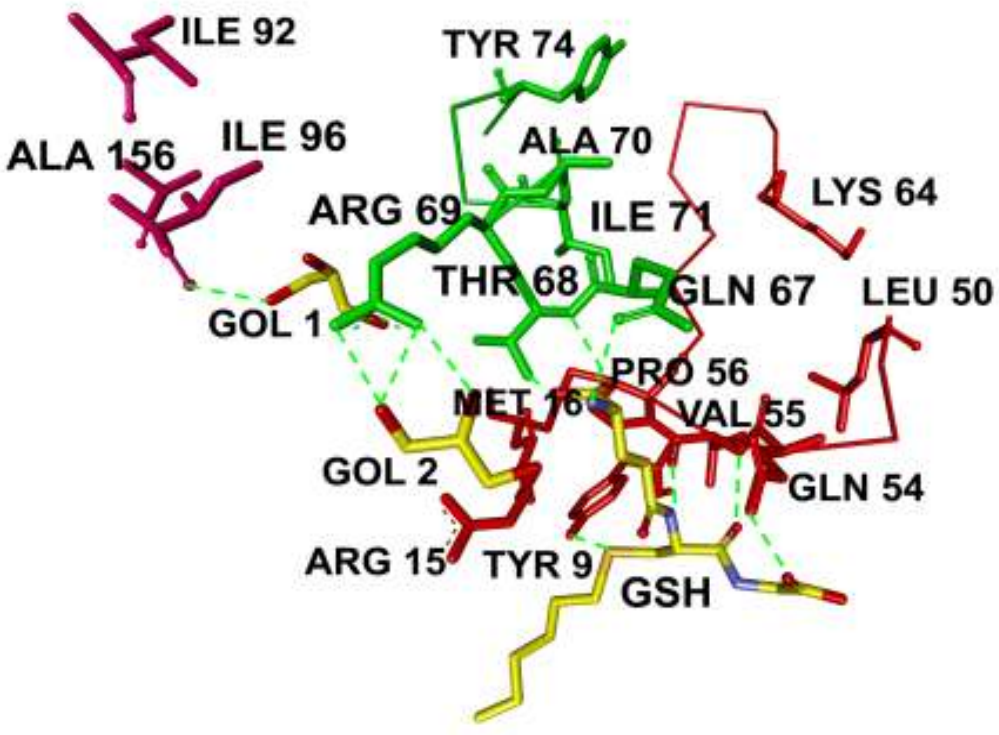


Figure 4.6 Above mentioned interaction path residues and clique residues are all labeled. Clique residues are colored in pink. Dashed lines are the hydrogen bonds.

The residues Ala24 and Val194 display relatively high total correlations. They belong to two different secondary structures and are in contact with each other. Yet, in literature there is no comment on their contribution to the structure and function of the protein. These two residues are not shown in Figure 4.4.b.

Cliques of size three, at cut-off 6.2 Å, are found as Ile92, Ile96 and Ala156 which are located near the interface of chain A and chain B. Ala156 is a highly-conserved residue[93] which is contrary to Ile92 and Ile96. Yet, Ile96 is at the glycerol binding site and hydrogen-bonded to the first glycerol. (Figure 4.4.b)

4.3. Human Protein Tyrosine phosphatase 1B

The catalytic domain of protein tyrosine phosphatase 1B (PTP1B) is composed of a single α/β domain structured around a highly twisted β-sheet which spans the entire molecule. A-well known catalytic residue Cys215 is located on the loop that stays at the edge of this β-sheet. As seen from Figure 4.5.a, residues between His214-Ser222 exhibit the highest total correlation, C_T. This group of residues are also observed in Figure 4.5.b to form a dark strip, indicating that they are correlated with rest of the residues. This His214-Ser222 region indeed corresponds to the catalytic region of the protein[103]. In PTP1B, the residues His214, Cys215 and Ser 216 have central roles in the activation of the active-site [35]. Cys215 is emphasized as an important catalytic residue in literature [33-35].

In the inset of Figure 4.5.a, small peaks around the residues Arg45, Pro51, Tyr66-Asn68, Leu83-Gln85, Met109, Lys120, Thr154- Arg156, His175 and Gln262 are observed. Arg45 and Pro51 sit in the loop where phospho-Tyr recognition occurs and Arg45, located in the binding site of PTP1B, is responsible from the electrostatic attraction of the ligand. Asn68 makes a hydrogen bond with Asn44 and it is located near a highly conserved residue, Arg257. Leu83 packs or surrounds the PTP loop (residues 213-223) where Gln85 makes a hydrogen bond with a highly buried water molecule.

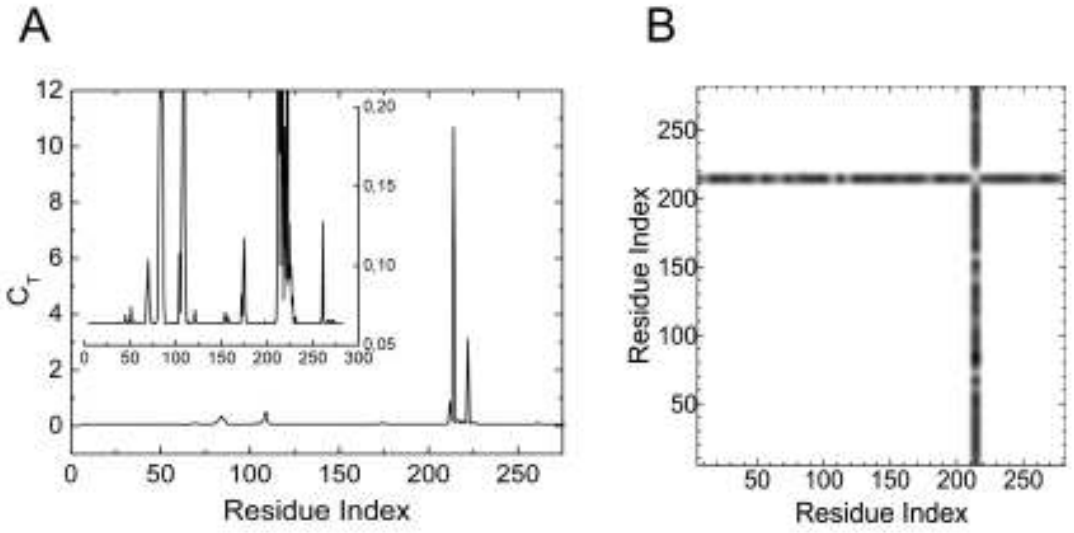


Figure 4.7 a) Total correlation C_T of residues as a function of residue indices. **b)** Contour plot of distance fluctuations $\left\langle \left(\Delta R_{ij} \right)^2 \right\rangle$ of 2HNP.pdb. Highest values indicated by black.

Residues around Met109 form the hydrophobic core structure and they are less conserved compared to the Ile82-Pro87 motif. Lys 120 is another binding site residue, which H-bonds to Ser216 and interacts with Asp181 (also not shown in Figure 4.6.b),

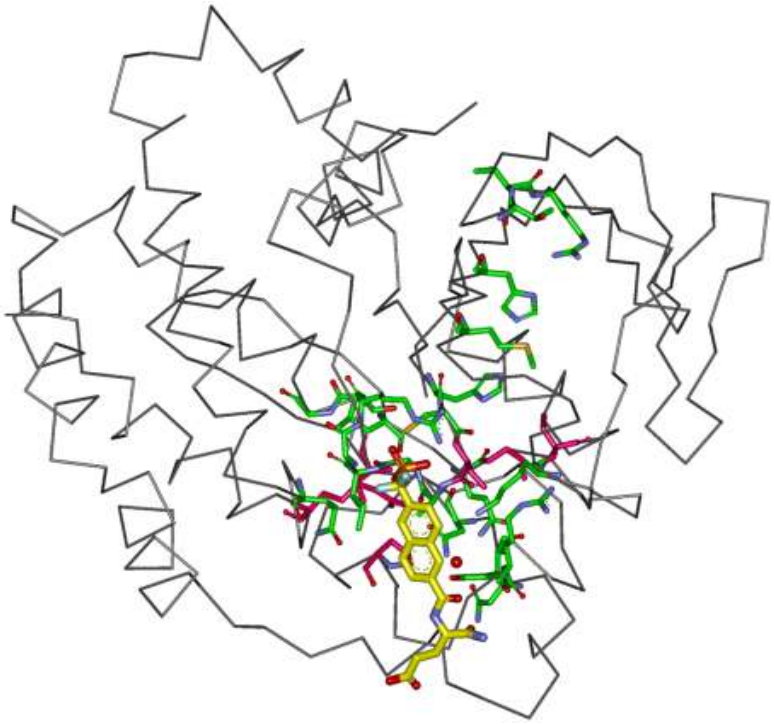


Figure 4.8 Three dimensional structure of human PTP 1B with TPI (yellow). Interaction path and the cliques are colored with green and pink, respectively.

known as a general acid catalyst among the vertebrate PTPs. Arg156 is conserved more than %80 among all vertebrate PTP domains. His175 is found in the surface exposed WPD loop (residues His 175–Val 184), where a major conformational change takes place upon binding of phosphopeptides to the PTP loop. The PTP loop then, moves several angstroms to close the active site pocket and trap the bound phosphotyrosine [104]. The WPD loop is also not shown in Figure 4.6b. Gln262 is also actively involved in ligand-binding process [35]. All residues mentioned above are colored in blue and all of them are labeled in Figure 4.6.b.

Cliques of size three are found as Pro51-Ser70-Arg257 and Gly86-Cys121-Ser216 at cut-off 6.1Å, all of which are highly conserved (Figure 6.b). The first triad is located around the active site; Pro51 is on the phosho-Tyr recognition loop and Arg257 is on the loop Leu250-Leu267 that spans the active site [105]. Arg257 makes a hydrogen bond with the PTP loop and also believed to be involved in stabilization of the

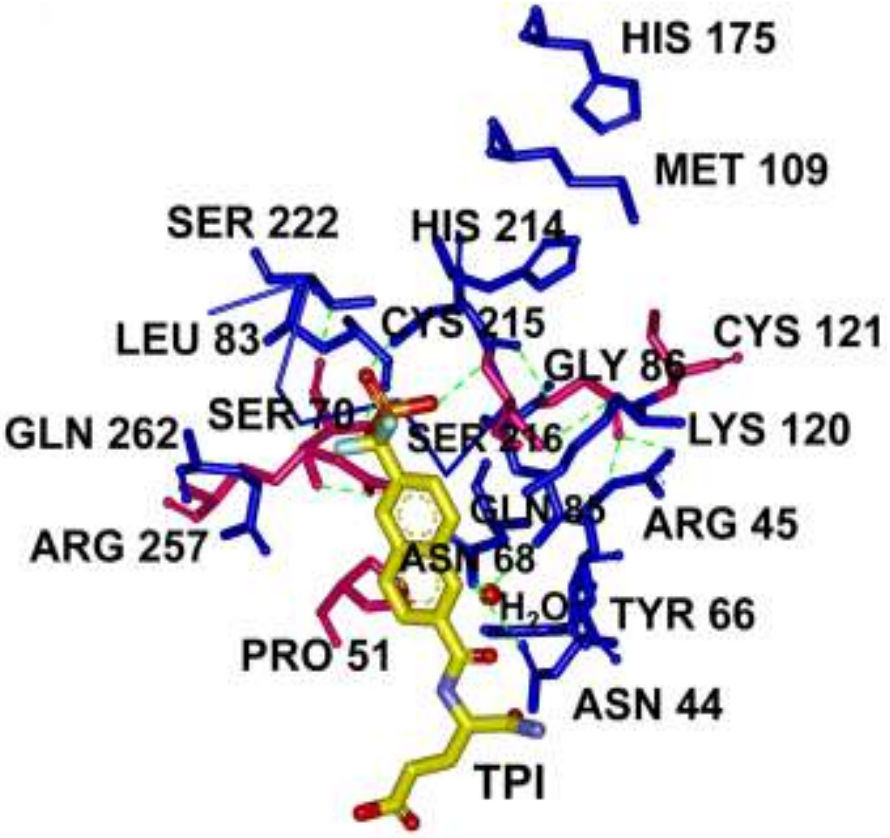


Figure 4.9 Identified interaction path residues and the clique residues are labeled. Clique residues are colored in pink. Dashed lines are the hydrogen bonds.

nucleophilic nature of the active site cysteine, Cys215[101]. Cys121, another clique residue is interacting with Cys215, as well[101]. It has been previously reported that Cys121 in PTP1B is a highly nucleophilic group accessible and ready for covalent attachment of 1,2-NQ, which is a known inhibitor of PTP1B. It causes considerable reduction in dephosphorylation activity of PTP1B. Moreover, Cys121 was reported as a non-active site cysteine residue, but it sits on an allestoric site, where it can inhibit the enzyme activity through specific mechanisms [100, 105]. There are a number of PTPs in which Cys121 (90%) is highly conserved [104]. Ser216 lies on the active site and functions in the activation of Cys215[35].

4.4. Biotin Carboxylase Domain of Acetyl-CoA Carboxylase 2

In its crystal structure, Sarophen A is bound to the human acetyl-CoA carboxylase 2 (ACC 2) (PDB code: 3GID). Binding site residues are Lys274-Ser278-Arg277-Glu593-Met594-Asn599-Asn679-Trp681-Phe704-Trp706.

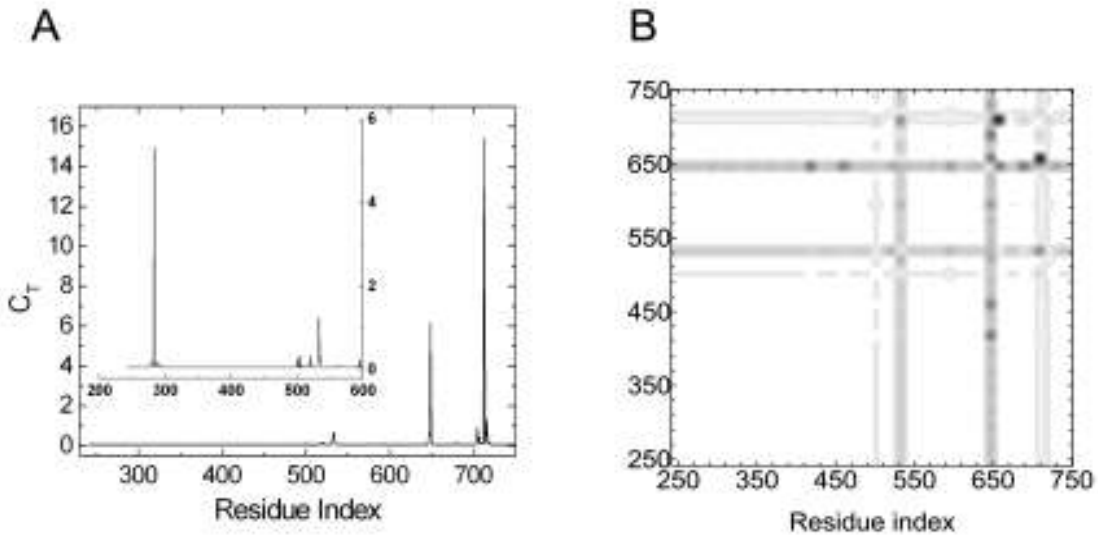


Figure 4.10 a) Total correlation C_T of residues as a function of residue indices. b) Contour plot of distance fluctuations $\left\langle \left(\Delta R_{ij} \right)^2 \right\rangle$ of 3GLK.pdb. Higher values are indicated by black.

Figure 4.10a shows total correlation, C_T as a function of residue index and the residues Val648, Glu533 and those between Phe704-Ser715, exhibit finite values of C_T. These residues also appear in Figure 4.10b as strips, signifying that they are correlated with rest of the residues.

In this thesis, we present no more than the fastest mode results for total coupling of residues. Yet, we checked the results for the second and the third fastest modes and in some proteins, including ACC2, we identified new paths of same kind which extend from surface to the ligand binding (active site) pocket. For instance, residues around Ser278 show the highest total correlation values in the fastest third mode. Lys274, Ser278 and Arg277 indeed stabilize the ligand via hydrogen bond formation[93]. These residues are not shown in Figure 4.12. Results for the second mode are presented in the inset of Figure 4.10a.

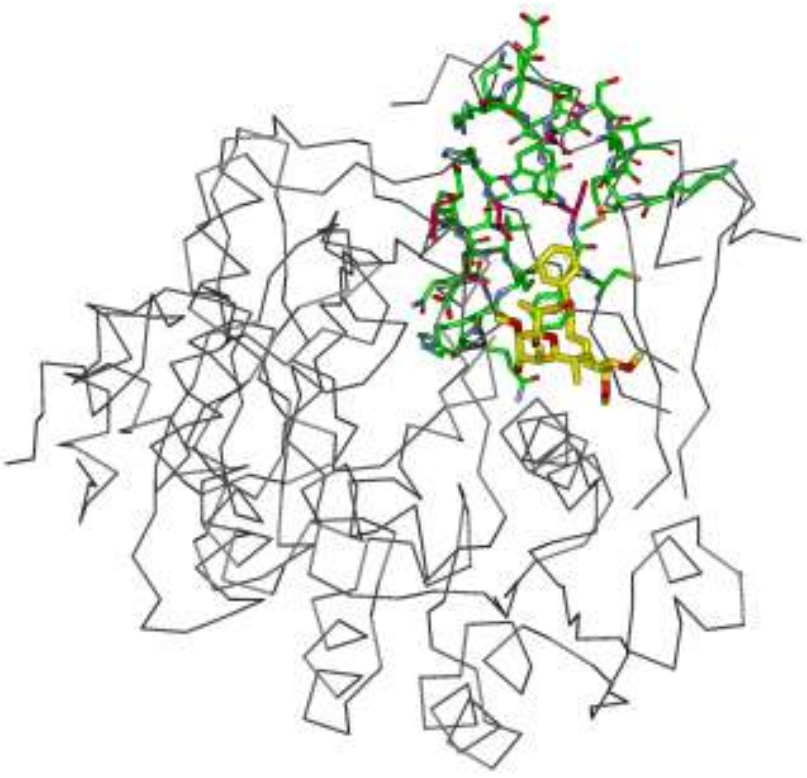


Figure 4.11 Three dimensional structure of BC domain of ACC2 with Sarophen A (yellow). Interaction path and the cliques are colored with green and pink, respectively.

According to Figure 4.12 Arg710 and Glu711 are surface-exposed residues and lie on the starting point of the green path that terminates at Phe704 and Val648. Phe704, with Ser705 and Trp706, surrounds the ligand. Ile649 and Asn679 are located around the Phe704-Ser715 path (Figure 4.12). Ile649 appears with the second highest C_T value in Figure 4.10a. Another path (colored in blue in Figure 4.12) starts with Arg710, involves Glu533, Arg519 and ends with Gln504. Glu533 is a well-conserved residue[93]. There are small peaks around the 500th and 520th residues. Gln504 and Arg519 are known as the catalytic residues in ACC2 [93, 106].

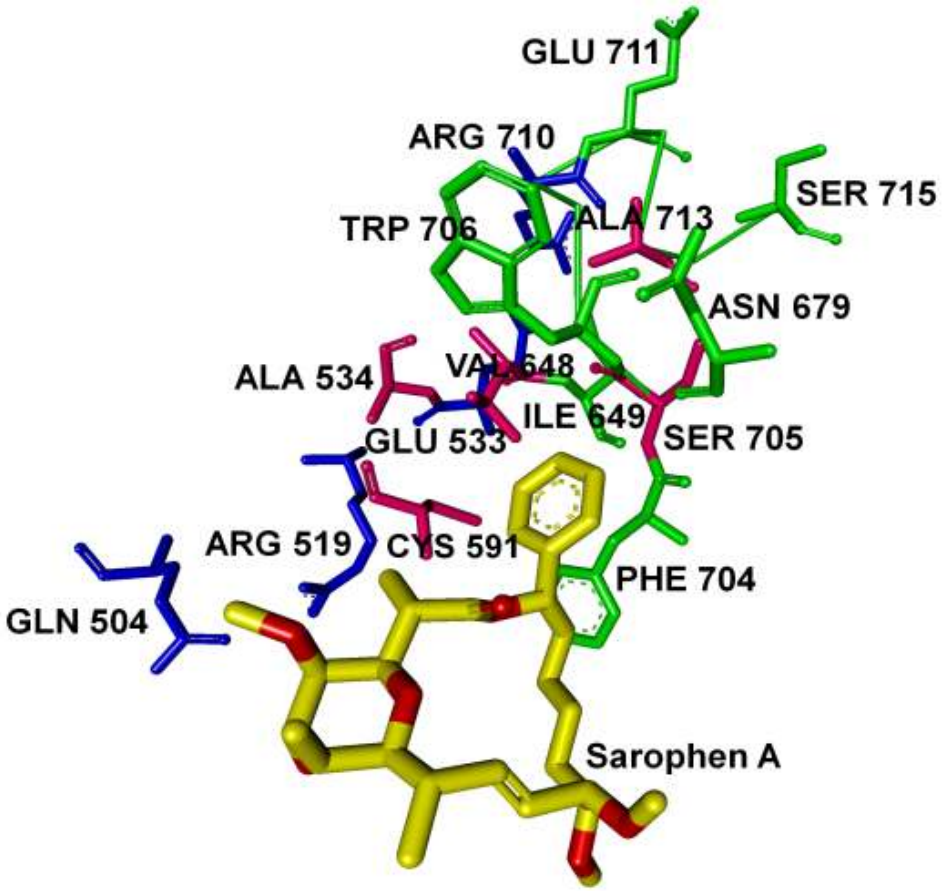


Figure 4.12 Enlarged version showing interaction path residues and cliques (pink) with their labels. Dashed lines are the hydrogen bonds.

Cliques of size three, at cut-off 6.1 Å, are found as Ala534-Cys591-Val648 and Val648-Ser705-Ala713. Cliques reside either in close proximity or within the active site

pocket, most of which fall on the interaction paths. (Figure 4.12) All clique residues are highly conserved residues[93].

4.5. Human Carbonic Anhydrase II

In its crystal structure (PDB code: 1A42), human carbonic anhydrase II is complexed with the drug used for glaucoma therapy, the sulfonamide inhibitor brinzolamide. The given binding site residues are His64-Gln92-His94-His96-His119-Val121-Phe131-Val135-Leu198-Thr199-Thr200.

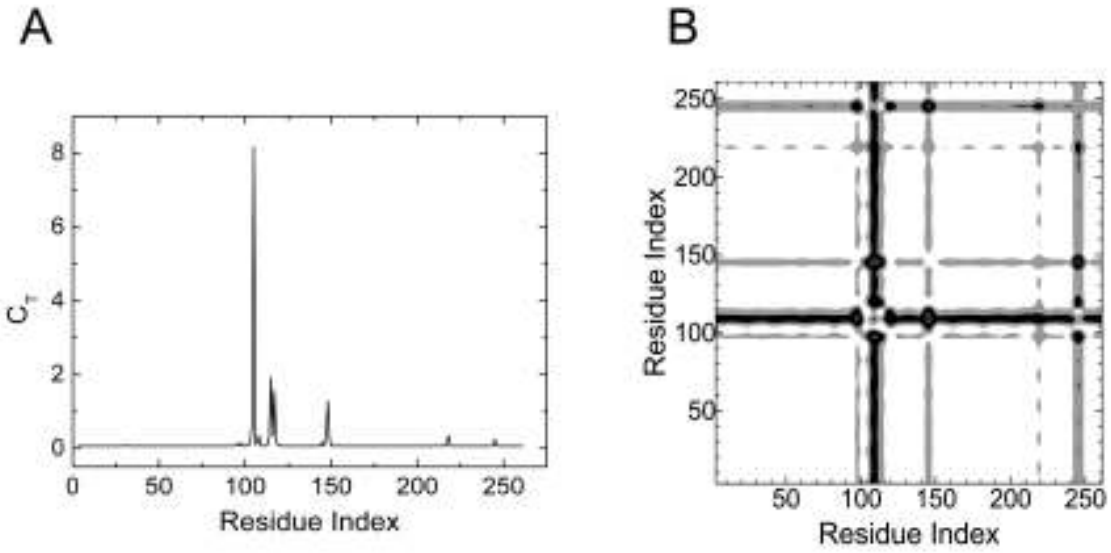


Figure 4.13 a) Total correlation C_T of residues as a function of residue indices. b) Contour plot of distance fluctuations $\left\langle \left(\Delta R_{ij} \right)^2 \right\rangle$ of 2CBE.pdb. Highest values indicated by black.

Residues between His96-His107, Tyr114-His119, Phe147-Lys149, Ser217-Val218 and Asn244-Arg246 exhibit finite values of total correlation according to Figure 4.13a. This set of residues are also form dark strips, as they correlate with the rest of the protein.

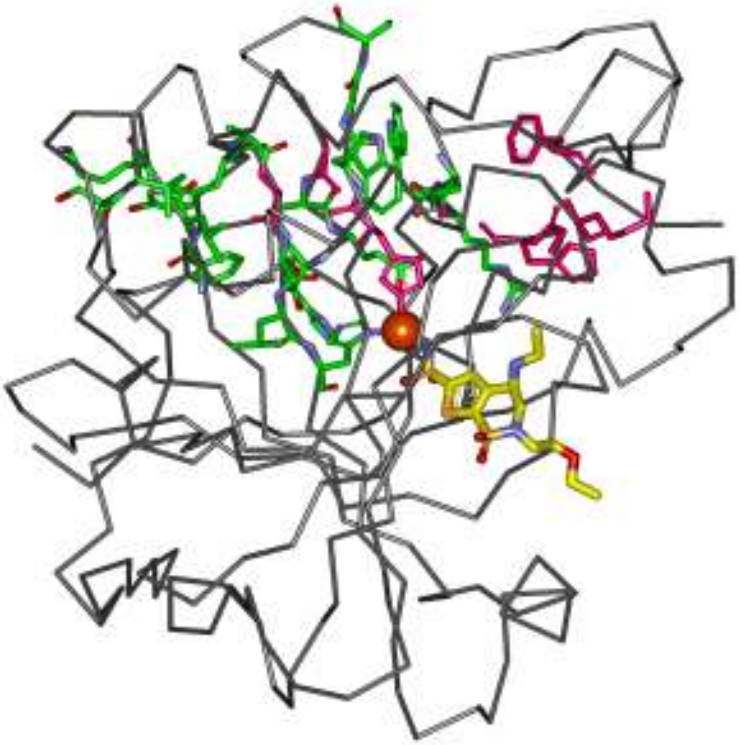


Figure 4.14 Three dimensional structure of Carbonic anhydrase II with Brinzolamide (yellow) and Zn+2(orange), interaction path (green) and cliques (pink)

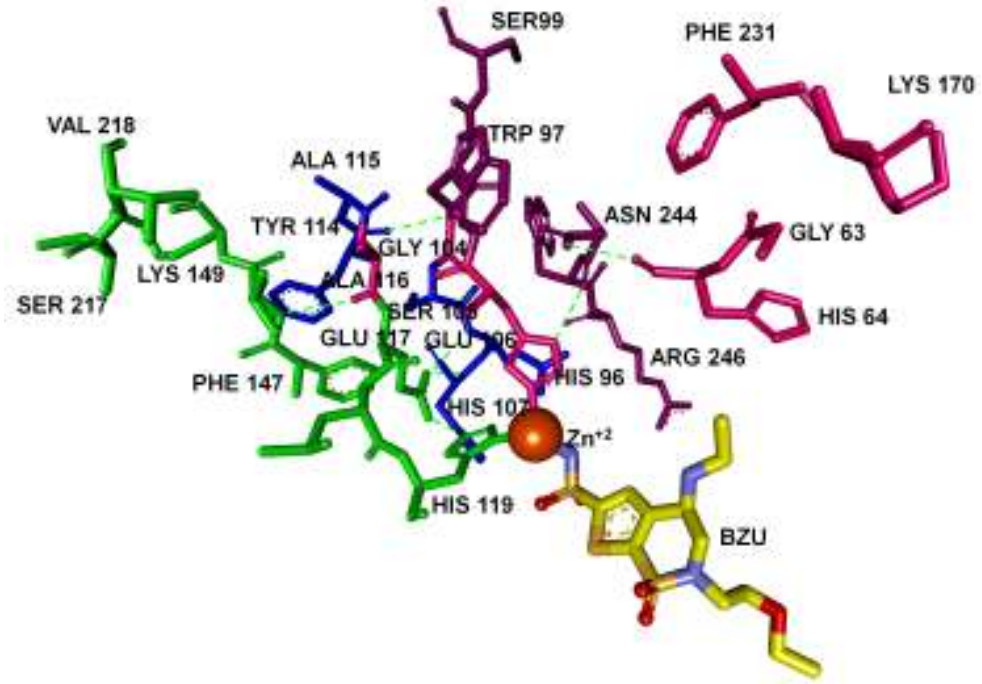


Figure 4.15 Enlarged version showing interaction path residues and cliques (pink) with their labels. Dashed lines are the hydrogen bonds.

The first path, which is colored in green in Figure 4.15, has one end at Ser217-Val218 and Lys149, and the other end at His119. The blue path starts with Ala115 and ends where the two paths are merged by the H-bonds Glu106 and His107 make with Glu117. Through the path Ala115 also interacts with Gly104 via hydrogen bonding. Ser105 links Gly104 with Glu106 and His107. The purple path has surface exposed Ser99 on one end and terminates at His96, which interacts with the Zn^{+2} ion that is bound to the ligand. (Figure 4.15) Indeed, the active site cleft is characterized by this Zn^{+2} ion which is tetrahedrally coordinated by N atoms of three histidine residues His94(not shown in Figure 4.15), His96 and His119 and a water/hydroxide molecule [107]. Ser105 and Glu117 are within the 10 residues that are completely invariant among the whole family of α-CAs and α-CA-related proteins. Ser105 is involved in stabilizing the protein structure, while Glu117 function as an indirect ligand in the active enzyme [108]. Asn244 and Arg246 are two conserved residues, (colored in purple in Figure 4.15) which also neighbor the ligand[93].

Cliques of size three are calculated at cut-off 6.1Å. The residue triads His96-Gly104-Ala116 and Gly63-Lys170-Phe231 appear around the catalytic site of the protein (Figure 4.15). His96 is an important residue which interacts with the Zn^{+2} ion during the catalysis. Gly104 and Ala116 are located in a conserved region, which involves Ser105 and Glu117 [108]. Gly63 is next to His64 which acts as a protein shuttle during catalysis[53]. The side chain of Lys170, the closest of all other residues to the pathway for protein transfer with His64 in the outward orientation [109]. It is believed that one function of Lys170 is to maintain an environment of His64 that maximizes protein transfer and catalysis of the hydration of CO_2 and dehydration of bicarbonate, by keeping it in its outward orientation [110]. In the outward conformation, the imidazole ring of His64 heads out of the active site cavity and the hydrophobic residue Phe231 is located near that cavity.

4.6. S100A6

In the crystal structure of human S100A6 (PDB code: 1K9K), binding sites for Ca^{+2} ions are given as, Ser20-Glu23-Asp25-Thr28-Glu33 and Asp61-Asn63-Asp65-Glu67-Glu72. In human S100A6, secondary structure elements are arranged into two calcium binding motifs, which compromise Ca^{+2} binding site I and site II. For site II (S100-hand motif), the most noticeable difference, upon Ca^{+2} binding is the movement of Glu33. In contrast, the coordination of the Ca^{+2} in site I (EF-hand motif), is largely mediated by main chain carbonyl of Glu67 and the side chains of Asp61, Asn63, Asp65 and Glu72. [111]

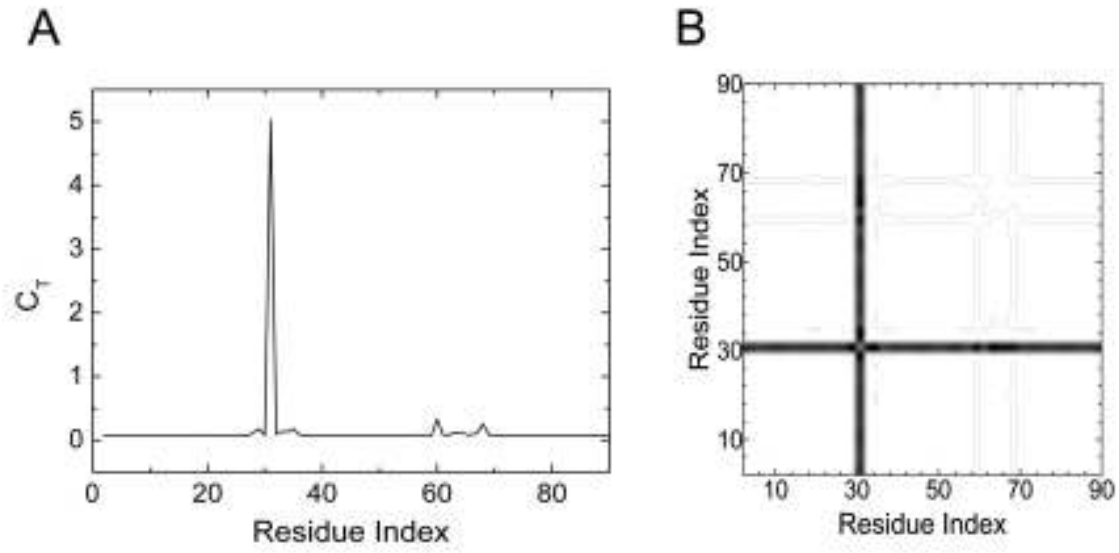


Figure 4.16 a) Total correlation C_T of residues as a function of residue indices. **b)** Contour plot of distance fluctuations $\left\langle \left(\Delta R_{ij} \right)^2 \right\rangle$ of 1K9P.pdb. Highest values indicated by black.

Residues between Thr28-Lys35, are observed with finite C_T values in Figure 4.16a. In Figure 4.16b, the residues that exchange energy with the surroundings are recognized with a darker tone. The heavy vertical strip points that the residues 28-35 interact with all the residues of the protein.

This group of residues form a path which begins with hydrogen bonded residues Lys35 and Lys31, and ends with two binding site residues Thr28 and Glu33. Groups of residues around Asp61 and Glu67 exhibit finite total correlation values, as well, referring to Figure 4.16a. The path that surrounds site I is shorter and involves Asp61, Asn63 and Glu67. The path begins with Leu60 which is a well-conserved surface-exposed residue[93].

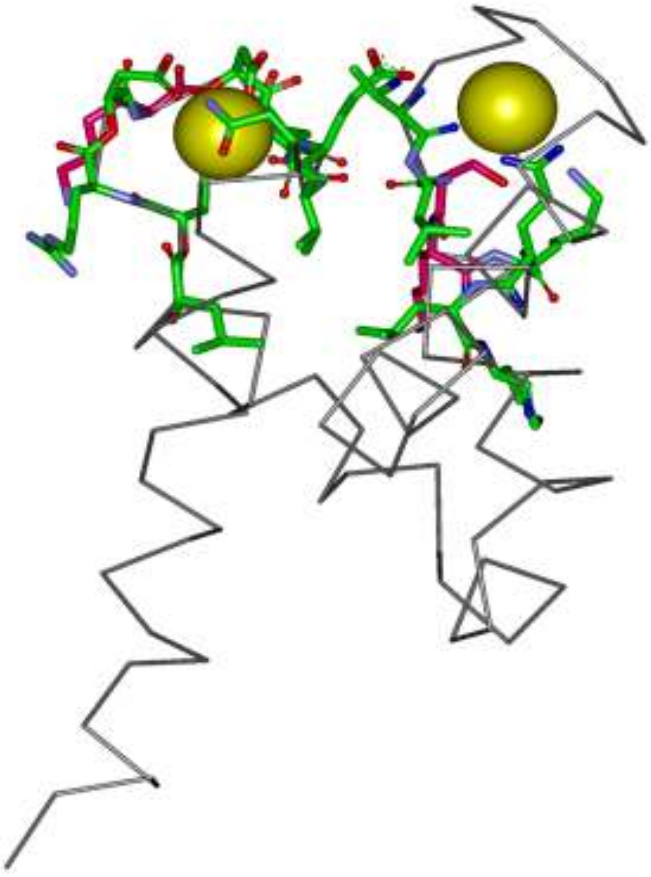


Figure 4.17 Three dimensional structure of one chain of S100A6 with bound Ca^{+2} ions (yellow). interaction path (green) and cliques (pink)

According to our results, residue pairs with the highest total correlation appear around the residues Lys31 and Leu60. Interaction path residues are mostly the binding site residues. Other residues line a network through the protein between the two Ca^{+2}

binding sites. (Figure18) Cliques are calculated at cut-off 6.1Å. The triad Lys31-Leu30-Lys64 also appears around the catalytic site of the protein (Figure 18).

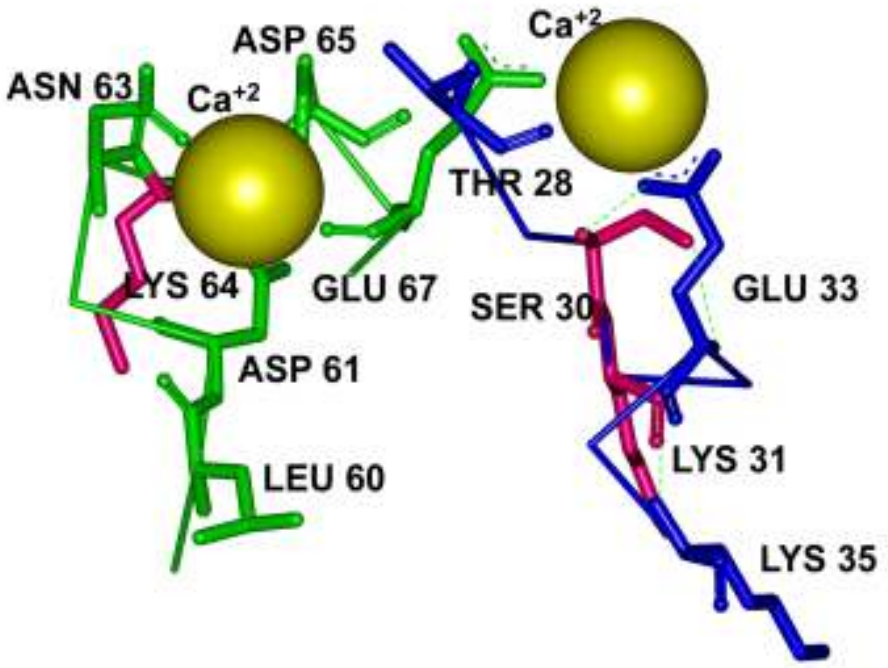


Figure 4.18 Enlarged version showing interaction path residues and cliques (pink) with their labels. Dashed lines are the hydrogen bonds.

Chapter 5

CONCLUSION

We have presented a collection of computational techniques to study the relationship between the 3-dimensional structure and the dynamics of protein. The two methods that we used in this thesis, relate protein structure with protein function and protein dynamics in terms of ligand binding. Contact map of a protein can be investigated by the tools of graph theory and provides information about the stiff and conserved, therefore functionally important regions. We introduce these certain regions as "cliques" which are made up of residue triads and typically reside either along the catalytic region, if it is an enzyme, or along the ligand binding pocket. This kind of approach establishes the structure-function correlations in proteins. Gaussian Network Model (GNM), on the other hand, correlates the fluctuations of residues, i.e. dynamics of the protein with the three dimensional structure of the protein. Based on the GNM, structural and thermodynamic features of the bound state are predicted by using the unbound structures. This was also observed in a recent work. [5]

We applied the two computational methods to the crystal structures of known systems. We used ligand-free structures to find cliques which are conserved throughout the evolution, and interaction pathways through which the energy is transferred to the whole system. Then, we used ligand-bound systems as positive controls. Findings of these two methods show that the binding information is already present in the unbound structure.

We conducted this study in a diverse set of 30 proteins, each having a distinct function. Among those we obtained successful results in 29 systems. Residues with finite total correlation (C_T) values appear along a path with one end located at the surface and the other end exposed to the ligand binding pocket (site). These residue interaction networks indicate the existence of the interaction path which is directly related with ligand binding and highly dependent on the topology of the protein. In this thesis, we present no more than the fastest mode results for total coupling of residues. Yet, we checked the results for the second fastest mode and identified new pathways of same kind which extend from different energy gate residues (to the ligand binding pocket). The investigation of these results will be the future direction of our work.

Cliques made up of residue triads always appeared as highly conserved residues therefore, they are mostly related with function. Most of the catalytic residues are predicted, those which are emphasized in the literature. They are located around the ligand binding pocket, usually interacting with the ligand. Clique results also show consistency with the GNM results for most of our test set proteins. In order to find cliques of size three, we used different cut off values and the best results come in 6.1-6.2 Å. Cut off values that offer the best results as clique residues, are provided in Supplementary Data. (Table S1)

Our approach exhibits a high predictive capability. The table which involves the data set and the summary of results for the remaining proteins are presented in *Supplementary*. We have shown this approach to be successful in the identification of interaction pathways and conserved regions in a diverse set of protein-ligand systems.

BIBLIOGRAPHY

1. Jorgensen, W.L., *Rusting of the Lock and Key Model for Protein-Ligand Binding.* Science, 1991. **254**(5034): p. 954-955.
2. Mizoue, L.S. and W.J. Chazin, *Engineering and design of ligand-induced conformational change in proteins.* Current Opinion in Structural Biology, 2002. **12**(4): p. 459-463.
3. Bahar, I., A.R. Atilgan, and B. Erman, *Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential.* Folding & Design, 1997. **2**(3): p. 173-181.
4. Yogurtcu, O.N., M. Gur, and B. Erman, *Statistical thermodynamics of residue fluctuations in native proteins.* Journal of Chemical Physics, 2009. **130**(9): p. -.
5. Haliloglu, T. and B. Erman, *Analysis of Correlations between Energy and Residue Fluctuations in Native Proteins and Determination of Specific Sites for Binding.* Physical Review Letters, 2009. **102**(8): p. -.
6. Haliloglu, T., E. Seyrek, and B. Erman, *Prediction of binding sites in receptor-ligand complexes with the Gaussian Network Model.* Physical Review Letters, 2008. **100**(22): p. -.
7. Haliloglu, T., A. Gul, and B. Erman, *Predicting Important Residues and Interaction Pathways in Proteins Using Gaussian Network Model: Binding and Stability of HLA Proteins* PLOS Computational Biology, 2010. **6**(7): p. 1-11.
8. Cvetkovic D, R.P., Simic S *Eigenspaces of graphs*. 1997, London: Cambridge University Press.
9. Lockless, S.W. and R. Ranganathan, *Evolutionarily conserved pathways of energetic connectivity in protein families.* Science, 1999. **286**(5438): p. 295-299.
10. Nelson, M.R., et al., *The EF-hand domain: A globally cooperative structural unit.* Protein Science, 2002. **11**(2): p. 198-205.

11. Pan, H., J.C. Lee, and V.J. Hilser, *Binding sites in Escherichia coli dihydrofolate reductase communicate by modulating the conformational ensemble.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(22): p. 12020-12025.
12. Llorente, M.A., A.M. Rubio, and J.J. Freire, *Moments and Distribution-Functions of the End-to-End Distance of Short Poly(Dimethylsiloxane) and Poly(Oxyethylene) Chains - Application to the Study of Elasticity in Model Networks.* Macromolecules, 1984. **17**(11): p. 2307-2315.
13. Menduina, C., et al., *Correctly Averaged Non-Gaussian Theory of Rubber-Like Elasticity - Application to the Description of the Behavior of Poly(Dimethylsiloxane) Bimodal Networks.* Macromolecules, 1986. **19**(4): p. 1212-1217.
14. Amitai, G., et al., *Network analysis of protein structures identifies functional residues.* Journal of Molecular Biology, 2004. **344**(4): p. 1135-1146.
15. Perry, R.D.L.a.A.D., *A method of matrix analysis of group structure.* Psychometrika, 1949. **14**(2): p. 95-116.
16. RSCB, *Protein Data Bank*.
17. Tenhunen, R., Marver, H.S. & Schmid, R., *Microsomal heme oxygenase.Characterization of the enzyme.* J.Biol.Chem., 1969. **244**: p. 6388-6394.
18. Dore, S.*e.a.*, *Bilirubin, formed by activation of heme oxygenase-2, protects neurons against oxidative stress injury.* . Proc. Natl. Acad. Sci., 1999. **96**: p. 2445–2450.
19. Maines, M.D., *The heme oxygenase system: a regulator of second messenger gases.* Annu. Rev. Pharmacol. Toxicol., 1997. **37**: p. 517–554.
20. Poss, K.D.T., S., *Heme oxygenase 1 is required for mammalian iron reutilization.* Proc. Natl. Acad. Sci. , 1997. **94**: p. 10919–10924.
21. Yachie, A.*e.a.*, *Oxidative stress causes enhanced endothelial injury in human heme oxygenase-1 deficiency.* J. Clin. Invest. , 1998. **103** p. 129–135.

22. Calabretta, B., et al., *Cell-cycle-specific genes differentially expressed in human leukemias.* Proc Natl Acad Sci U S A, 1985. **82**(13): p. 4463-7.
23. Lad, L., et al., *Comparison of the heme-free and -bound crystal structures of human heme oxygenase-1.* J Biol Chem, 2003. **278**(10): p. 7834-43.
24. La Mar, G.N., et al., *Solution 1H NMR of the active site of substrate-bound, cyanide-inhibited human heme oxygenase. comparison to the crystal structure of the water-ligated form.* J Biol Chem, 2001. **276**(19): p. 15676-87.
25. Atkins, W.M., et al., *The Catalytic Mechanism of Glutathione-S-Transferase (Gst) - Spectroscopic Determination of the Pk(a) of Tyr-9 in Rat Alpha-1-1 Gst.* Journal of Biological Chemistry, 1993. **268**(26): p. 19188-19191.
26. Hayes, J.D. and D.J. Pulford, *The glutathione S-Transferase supergene family: Regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance.* Critical Reviews in Biochemistry and Molecular Biology, 1995. **30**(6): p. 445-600.
27. Allardyce, C.S., et al., *The role of tyrosine-9 and the C-terminal helix in the catalytic mechanism of Alpha-class glutathione S-transferases.* Biochemical Journal, 1999. **343**: p. 525-531.
28. Armstrong, R.N., *Structure, catalytic mechanism, and evolution of the glutathione transferases.* Chemical Research in Toxicology, 1997. **10**(1): p. 2-18.
29. Sheehan, D., et al., *Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily.* Biochemical Journal, 2001. **360**: p. 1-16.
30. Thorson, J.S., et al., *Analysis of the role of the active site tyrosine in human glutathione transferase A1-1 by unnatural amino acid mutagenesis.* Journal of the American Chemical Society, 1998. **120**(2): p. 451-452.
31. Fantl, W.J., D.E. Johnson, and L.T. Williams, *Signaling by Receptor Tyrosine Kinases.* Annual Review of Biochemistry, 1993. **62**: p. 453-481.

32. Schlessinger, J. and A. Ullrich, *Growth-Factor Signaling by Receptor Tyrosine Kinases.* Neuron, 1992. **9**(3): p. 383-391.
33. Streuli, M., et al., *Distinct Functional Roles of the 2 Intracellular Phosphatase Like Domains of the Receptor-Linked Protein Tyrosine Phosphatases Lca and Lar.* Embo Journal, 1990. **9**(8): p. 2399-2407.
34. Streuli, M., et al., *A Family of Receptor-Linked Protein Tyrosine Phosphatases in Humans and Drosophila.* Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(22): p. 8698-8702.
35. Tonks, N.K., *Protein tyrosine phosphatases: from genes, to function, to disease.* Nature Reviews Molecular Cell Biology, 2006. **7**(11): p. 833-846.
36. Blume-Jensen, P. and T. Hunter, *Oncogenic kinase signalling.* Nature, 2001. **411**(6835): p. 355-65.
37. Hunter, T., *The Croonian Lecture 1997. The phosphorylation of proteins on tyrosine: its role in cell growth and disease.* Philos Trans R Soc Lond B Biol Sci, 1998. **353**(1368): p. 583-605.
38. Ventura, J.J. and A.R. Nebreda, *Protein kinases and phosphatases as therapeutic targets in cancer.* Clin Transl Oncol, 2006. **8**(3): p. 153-60.
39. Andersen, J.N.T., N. K., *Protein Tyrosine Phosphatase-Based Therapeutics: Lessons from PTP1B.* Spring-Verlag, 2004.
40. Wiesmann, C., et al., *Allosteric inhibition of protein tyrosine phosphatase 1B.* Nat Struct Mol Biol, 2004. **11**(8): p. 730-7.
41. Zhang, S.Z.a.Z.-Y., *PTP1B as a drug target: recent developments in PTP1B inhibitor discovery.* Elsevier, 2007. **12**(9/10): p. 373-381.
42. Friedman, J.M., *A war on obesity, not the obese.* Science, 2003. **299**(5608): p. 856-858.
43. Flier, J.S., *Obesity wars: Molecular progress confronts an expanding epidemic.* Cell, 2004. **116**(2): p. 337-350.
44. Lazar, M.A., *How obesity causes diabetes: Not a tall tale.* Science, 2005. **307**(5708): p. 373-375.

45. Tong, L., *Acetyl-coenzyme A carboxylase: crucial metabolic enzyme and attractive target for drug discovery.* Cellular and Molecular Life Sciences, 2005. **62**(16): p. 1784.
46. Munday, M.R., *Regulation of mammalian acetyl-CoA carboxylase.* Biochemical Society Transactions, 2002. **30**: p. 1059-1064.
47. Barber, M.C., N.T. Price, and M.T. Travers, *Structure and regulation of acetyl-CoA carboxylase genes of metazoa.* Biochimica Et Biophysica Acta-Molecular and Cell Biology of Lipids, 2005. **1733**(1): p. 1-28.
48. Wakil, S.J., *The Relationship between Structure and Function for and the Regulation of the Enzymes of Fatty-Acid Synthesis.* Annals of the New York Academy of Sciences, 1986. **478**: p. 203-219.
49. Castle, J.C., et al., *ACC2 Is Expressed at High Levels Human White Adipose and Has an Isoform with a Novel N-Terminus.* Plos One, 2009. **4**(2): p. -.
50. Abu-Elheiga, L., et al., *Continuous fatty acid oxidation and reduced fat storage in mice lacking acetyl-CoA carboxylase 2.* Science, 2001. **291**(5513): p. 2613-2616.
51. Weatherly, S.C., S.L. Volrath, and T.D. Elich, *Expression and characterization of recombinant fungal acetyl-CoA carboxylase and isolation of a soraphen-binding domain.* Biochemical Journal, 2004. **380**: p. 105-110.
52. Shen, Y., et al., *A mechanism for the potent inhibition of eukaryotic acetyl-coenzyme a carboxylase by soraphen A, a macrocyclic polyketide natural product.* Molecular Cell, 2004. **16**(6): p. 881-891.
53. Jeremy M. Berg, J.L.T., Lubert Stryer, *Biochemistry*. 6 ed. 2007, New York: W.H. Freeman and Company. 254-259.
54. Smith, S.P. and G.S. Shaw, *A change-in-hand mechanism for S100 signalling.* Biochem Cell Biol, 1998. **76**(2-3): p. 324-33.
55. Gifford, J.L., M.P. Walsh, and H.J. Vogel, *Structures and metal-ion-binding properties of the Ca2+-binding helix-loop-helix EF-hand motifs.* Biochem J, 2007. **405**(2): p. 199-221.

56. Kligman, D. and D.C. Hilt, *The S100 protein family.* Trends Biochem Sci, 1988. **13**(11): p. 437-43.
57. Lackmann, M., et al., *Identification of a chemotactic domain of the pro-inflammatory S100 protein CP-10.* J Immunol, 1993. **150**(7): p. 2981-91.
58. Engelkamp, D., et al., *Six S100 genes are clustered on human chromosome 1q21: identification of two genes coding for the two previously unreported calcium-binding proteins S100D and S100E.* Proc Natl Acad Sci U S A, 1993. **90**(14): p. 6547-51.
59. Schafer, B.W. and C.W. Heizmann, *The S100 family of EF-hand calcium-binding proteins: functions and pathology.* Trends Biochem Sci, 1996. **21**(4): p. 134-40.
60. Calabretta, B., et al., *Altered expression of G1-specific genes in human malignant myeloid cells.* Proc Natl Acad Sci U S A, 1986. **83**(5): p. 1495-8.
61. Komatsu, K., et al., *Increased expression of S100A6 (Calcyclin), a calcium-binding protein of the S100 family, in human colorectal adenocarcinomas.* Clin Cancer Res, 2000. **6**(1): p. 172-7.
62. Mani, R.S., W.D. McCubbin, and C.M. Kay, *Calcium-dependent regulation of caldesmon by an 11-kDa smooth muscle calcium-binding protein, caltropin.* Biochemistry, 1992. **31**(47): p. 11896-901.
63. Murphy, L.C., et al., *Cloning and characterization of a cDNA encoding a highly conserved, putative calcium binding protein, identified by an anti-prolactin receptor antiserum.* J Biol Chem, 1988. **263**(5): p. 2397-401.
64. Sudo, T. and H. Hidaka, *Characterization of the calcyclin (S100A6) binding site of annexin XI-A by site-directed mutagenesis.* FEBS Lett, 1999. **444**(1): p. 11-4.
65. Tirion, M.M., *Large amplitude elastic motions in proteins from a single-parameter, atomic analysis,.* Phys. Rev. Lett., 1996. **77**: p. 1905.
66. Bahar, I., A.R. Atilgan, and B. Erman, *Direct evaluation of thermal fluctuations in protein using a single parameter harmonic potential.* Folding & Design 1997. **2**: p. 173-181.

67. Haliloglu, T., I. Bahar, and B. Erman, *Gaussian dynamics of folded proteins.* Phys. Rev. Lett., 1997. **79**: p. 3090-3093.
68. Flory, P.J., *Statistical thermodynamics of random networks*. Vol. 351. 1976, Lond.: Proc. Roy. Soc.
69. Go, N., T. Noguti, and T. Nishikawa, *Dynamics of a small globular protein in terms of low-frequency vibrational modes.* Proc. Natl. Acad. Sci. USA 1983. **80**: p. 3696.
70. Atilgan, A.R., et al., *Anisotropy of fluctuation dynamics of proteins with an elastic network model.* Biophys J, 2001. **80**(1): p. 505-15.
71. Bahar, I., *Dynamics of proteins and biomolecular complexes: Inferring functional motions from structure.* Reviews in Chemical Engineering, 1999. **15**(4): p. 319-347.
72. Bahar, I. and R.L. Jernigan, *Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation.* Journal of Molecular Biology, 1997. **266**(1): p. 195-214.
73. Bahar, I. and R.L. Jernigan, *Vibrational dynamics of transfer RNAs: Comparison of the free and synthetase-bound forms.* Journal of Molecular Biology, 1998. **281**(5): p. 871-884.
74. Keskin, O., R.L. Jernigan, and I. Bahar, *Proteins with similar architecture exhibit similar large-scale dynamic behavior.* Biophysical Journal, 2000. **78**(4): p. 2093-2106.
75. Kundu, S., et al., *Dynamics of proteins in crystals: comparison of experiment with simple models.* Biophys J, 2002. **83**(2): p. 723-32.
76. Biggs, N.L., E. and Wilson, R. , *Graph Theory, 1736-1936*. 1986: Oxford University Press.
77. Appel, K.a.H., W. , *Every planar map is four colorable. Part I. Discharging.* 1977. **21**: p. 429–490.
78. Appel, K.a.H., W. , *Every planar map is four colorable. Part II. Reducibility.* Illinois J. Math, 1977. **21**(491–567).

79. Diestel, R., *Graph Theory*. 4th ed. Graduate Texts in Mathematics. Vol. 173. 2010: Springer. 451.
80. Murty, J.A.B.a.U.S.R., *Graph Theory with Applications*. 1976.
81. Vendruscolo, M., et al., *Small-world view of the amino acids that play a key role in protein folding.* Physical Review E, 2002. **65**(6): p. -.
82. Bagler, G. and S. Sinha, *Network properties of protein structures.* Physica a-Statistical Mechanics and Its Applications, 2005. **346**(1-2): p. 27-33.
83. Greene, L.H. and V.A. Higman, *Uncovering network systems within protein structures.* Journal of Molecular Biology, 2003. **334**(4): p. 781-791.
84. Atilgan, A.R., P. Akan, and C. Baysal, *Small-world communication of residues and significance for protein dynamics.* Biophysical Journal, 2004. **86**(1): p. 85-91.
85. del Sol, A., et al., *Residue centrality, functionally important residues, and active site shape: Analysis of enzyme and non-enzyme families.* Protein Science, 2006. **15**(9): p. 2120-2128.
86. del Sol, A., et al., *Residues crucial for maintaining short paths in network communication mediate signaling in proteins.* Molecular Systems Biology, 2006: p. -.
87. Callen, H.B., *Thermodynamics and an introduction to thermostatistics*. Second ed. 1985: Wiley.
88. Haliloglu, T., A. Gul, and B. Erman, *Predicting important residues and interaction pathways in proteins using Gaussian Network Model: binding and stability of HLA proteins.* PLoS Comput Biol. **6**(7): p. e1000845.
89. Bahar, I., et al., *Vibrational dynamics of folded proteins: Significance of slow and fast motions in relation to function and stability.* Phys. Rev. Lett., 1998. **80**: p. 2733-2736.
90. Li, Y.M., et al., *H-1 NMR investigation of the solution structure of substrate-free human heme oxygenase - Comparison to the cyanide-inhibited, substrate-*

bound complex. Journal of Biological Chemistry, 2004. **279**(11): p. 10195-10205.

91. Sugishima, M., et al., *Crystal structure of rat apo-heme oxygenase-1 (HO-1): Mechanism of heme binding in HO-1 inferred from structural comparison of the apo and heme complex forms.* Biochemistry, 2002. **41**(23): p. 7293-7300.
92. Schuller, D.J., et al., *Crystal structure of human heme oxygenase-l.* Nature Structural Biology, 1999. **6**(9): p. 860-867.
93. EMBL-EBI, *PDBsum*.
94. Cameron, A.D., et al., *Structural-Analysis of Human Alpha-Class Glutathione Transferase a1-1 in the Apo-Form and in Complexes with Ethacrynic-Acid and Its Glutathione Conjugate.* Structure, 1995. **3**(7): p. 717-727.
95. Kuhnert, D.C., et al., *Tertiary interactions stabilise the C-terminal region of human glutathione transferase A1-1: A crystallographic and calorimetric study.* Journal of Molecular Biology, 2005. **349**(4): p. 825-838.
96. Nathaniel, C., et al., *The role of an evolutionarily conserved cis-proline in the thioredoxin-like domain of human class Alpha glutathione transferase A1-1.* Biochemical Journal, 2003. **372**: p. 241-246.
97. Bjornestedt, R., et al., *Functional-Significance of Arginine-15 in the Active-Site of Human Class-Alpha Glutathione Transferase a1-1.* Journal of Molecular Biology, 1995. **247**(4): p. 765-773.
98. Le Trong, I., et al., *1.3-Å resolution structure of human glutathione S-transferase with S-hexyl glutathione bound reveals possible extended ligandin binding site* Proteins, 2002. **48**: p. 618– 627.
99. Bruns, C.M., et al., *Human glutathione transferase A4-4 crystal structures and mutagenesis reveal the basis of high catalytic efficiency with toxic lipid peroxidation products.* Journal of Molecular Biology, 1999. **288**(3): p. 427-439.
100. Sinning, I., et al., *Structure Determination and Refinement of Human-Alpha Class Glutathione Transferase-a1-1, and a Comparison with the Mu-Class and Pi-Class Enzymes.* Journal of Molecular Biology, 1993. **232**(1): p. 192-212.

101. Hansen, S.K., et al., *Allosteric inhibition of PTP1B activity by selective modification of a non-active site cysteine residue.* Biochemistry, 2005. **44**(21): p. 7704-7712.

102. Peters, G.H., T.M. Frimurer, and O.H. Olsen, *Electrostatic evaluation of the signature motif (H/V)CX5R(S/T) in protein-tyrosine phosphatases.* Biochemistry, 1998. **37**(16): p. 5383-5393.

103. Scapin, G., et al., *The structural basis for the selectivity of benzotriazole inhibitors of PTP1B.* Biochemistry, 2003. **42**(39): p. 11451-11459.

104. Andersen, J.N., et al., *Structural and evolutionary relationships among protein tyrosine phosphatase domains.* Molecular and Cellular Biology, 2001. **21**(21): p. 7117-7136.

105. Scapin, G., et al., *The structure of apo protein-tyrosine phosphatase 1B C215S mutant: more than just an S --> O change.* Protein Sci, 2001. **10**(8): p. 1596-605.

106. Craig Porter, G.B., James Torrance, Nicholas Furnham and Janet Thornton, *Catalytic Site Atlas*. 2010, EMBL-EBI.

107. Eriksson, A.E., T.A. Jones, and A. Liljas, *Refined structure of human carbonic anhydrase II at 2.0 Å resolution.* Proteins: Structure, Function, and Bioinformatics, 1988. **4**(4): p. 274-282.

108. Lindskog, S., *Structure and mechanism of carbonic anhydrase.* Pharmacology & Therapeutics, 1997. **74**(1): p. 1-20.

109. Maupin, C.M., et al., *Elucidation of the Proton Transport Mechanism in Human Carbonic Anhydrase II.* Journal of the American Chemical Society, 2009. **131**(22): p. 7598-7608.

110. Domsic, J.F., et al., *Structural and Kinetic Study of the Extended Active Site for ProtonTransfer in Human Carbonic Anhydrase II.* Biochemistry, 2010.

111. Otterbein, L.R., et al., *Crystal structures of S100A6 in the Ca2+-free and Ca2+-bound states: The calcium sensor mechanism of S100 proteins revealed at atomic resolution.* Structure, 2002. **10**(4): p. 557-567.

Table S1 – Summary of results for the whole data set

Name of the Protein	Prediction of Binding Sites	Clique cutoff(Å)
Alcohol Dehydrogenase	✓	6.0
Heme oxygenase- 1	✓	6.2
Ispc	✓	6.0
Adenosine Kinase	✓	6.0
Glutathione S-transferase A1-1	✓	6.2
Map Kinase P38-α	✗	6.1
Kinase Domain TRP-Ca Channel	✓	6.1
Cyclin-Dependent Kinase	✓	6.2
M-phase inducer phosphatase 2 (Cdc25b)	✓	6.1
Angiogenin	✓	6.0
Carboxypeptidase A	✓	6.1
Gamma Chymotyrpsin	✓	6.0
Glyoxalase I	✓	6.1
Lysozyme	✓	6.1
Tyrosyl-DNA phosphodiesterase	✓	6.1

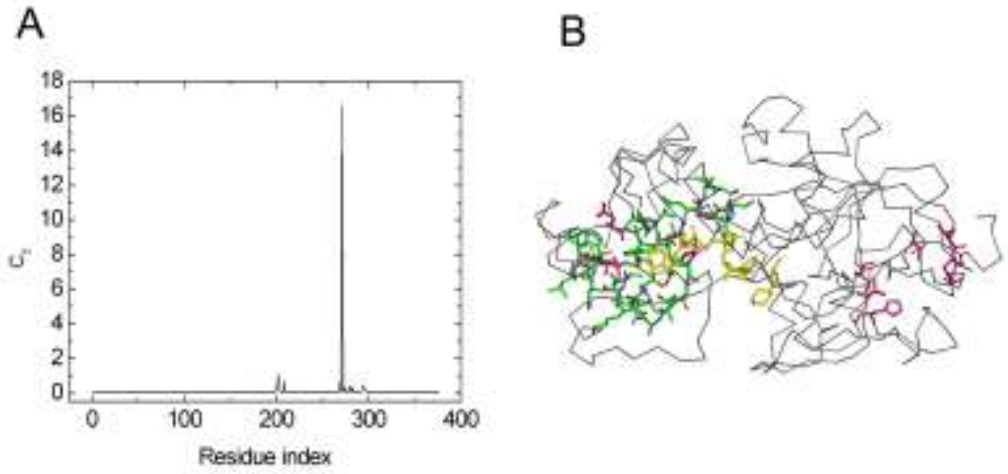
Beta-lactam synthetase	✓	6.1
Protein Tyrosine Phosphatase 1B (PTP1B)	✓	6.1
Vacuolar protein sorting Protein29 (VPS29)	✓	6.1
Phospholipase C	✓	6.4
Pancreatic α-amylase	✓	6.0
Acetyl-CoA carboxylase2 (ACC2)	✓	6.1
Carbonic Anhydrase II	✓	6.1
Ferrochelatase	✓	6.4
Hydroxynitrile Lyase	✓	6.1
Adipocyte Lipid binding Protein	✓	6.8
Ca-Binding S100A6	✓	6.1
Copper Resistance Protein	✓	No clique!
L-leucine Binding Protein	✓	6.0
Fibrillin	✓	6.2
TNF receptor associated factor (Traf 6)	✓	6.5

Table S1. 30 proteins in our test set are presented with their names and functions. In the second column, there is a check if our GNM predictions are consistent with the experimental results. Third column presents the best cut off value that predicts highly conserved residues. In 29 proteins out of 30, ligand binding sites are predicted succesfully. For cliques the best results are obtained in cut off values 6.1-6.2Å.

Total Correlation Results And Related Figures For Other Test Proteins

A. Oxireductases

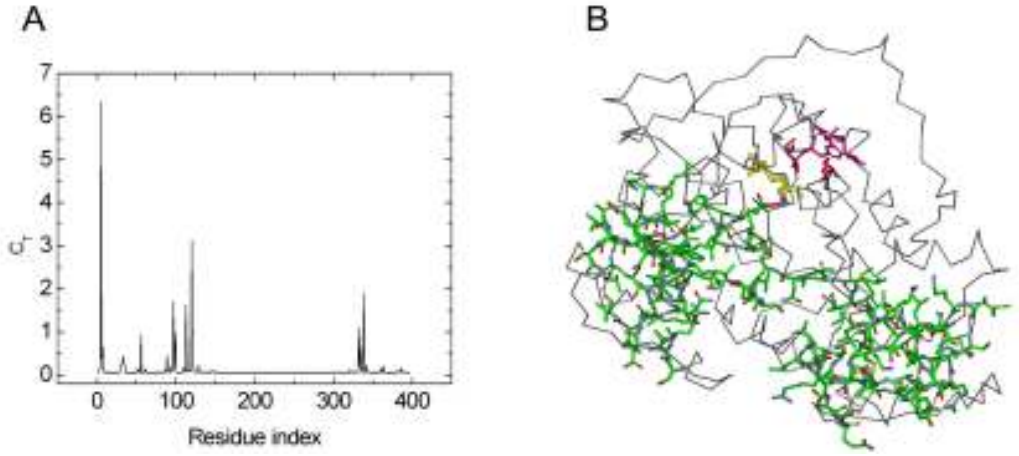
Figure S1 Alcohol Dehydrogenase



a) Total correlation C_T of residues as a function of residue indices

b) Three dimensional structure of mouse alcohol dehydrogenase chain A with NADH and inhibitor (yellow). Interaction path and the cliques are represented in green and pink, respectively.

Figure S2 Ispc

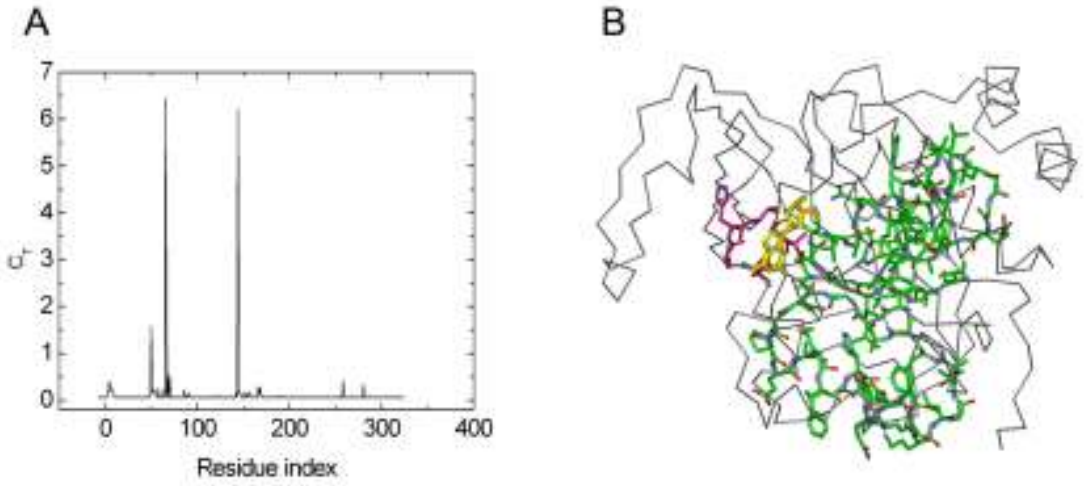


a) Total correlation C_T of residues as a function of residue indices

b) Three dimensional structure of Ispc chain A with Mn^{+2} and Fosmidomycin (yellow). Interaction path and the cliques are represented in green and pink, respectively.

B. Transferases

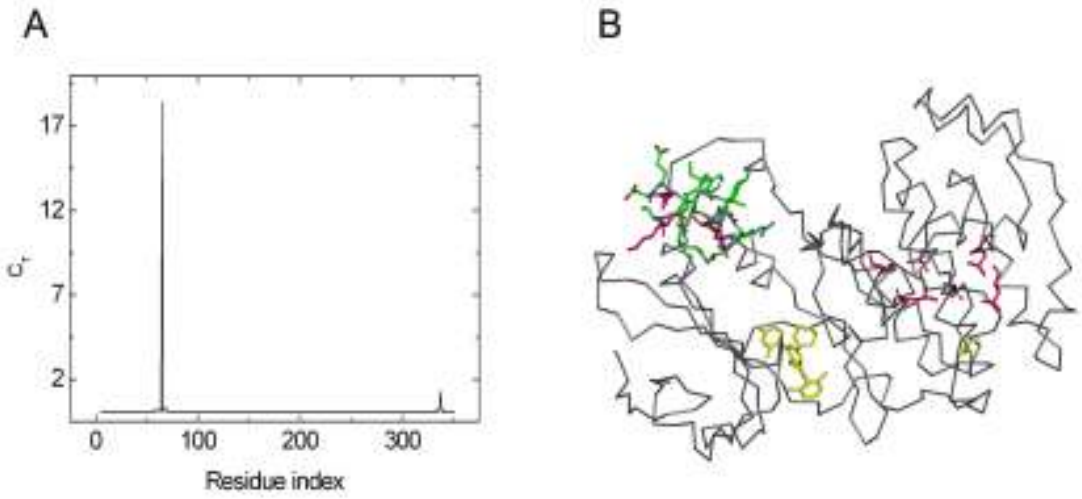
Figure S3 Adenosine Kinase



a) Total correlation C_T of residues as a function of residue indices

b) Three dimensional structure of adenosine kinase chain A with 2-Flouro adenosine (yellow). Interaction path and the cliques are represented in green and pink, respectively.

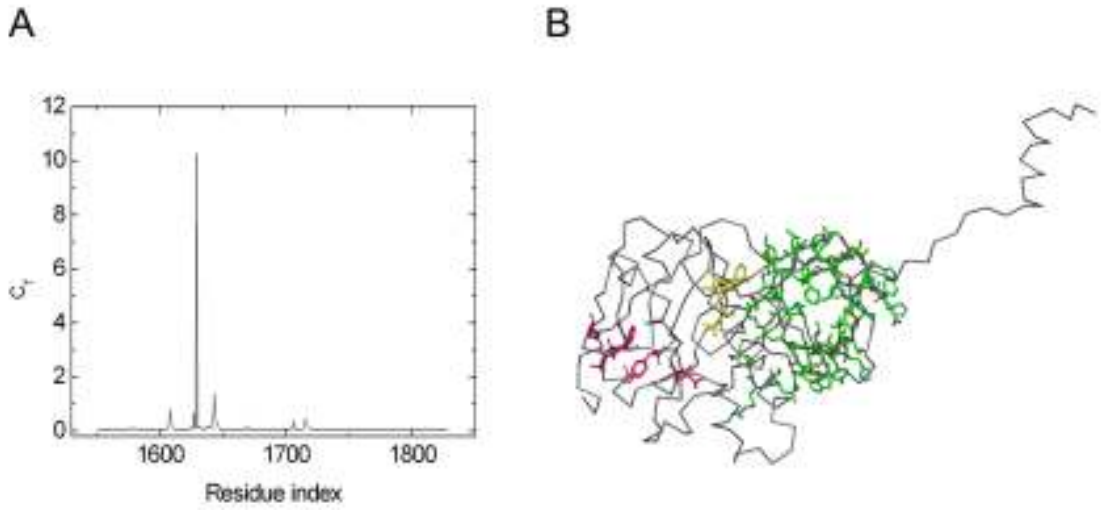
Figure S4 Map Kinase P38-α



a) Total correlation C_T of residues as a function of residue indices

b) Three dimensional structure of human map kinaseP38-α chain A with Dihydroquinolinone (yellow). Interaction path and the cliques are represented in green and pink, respectively.

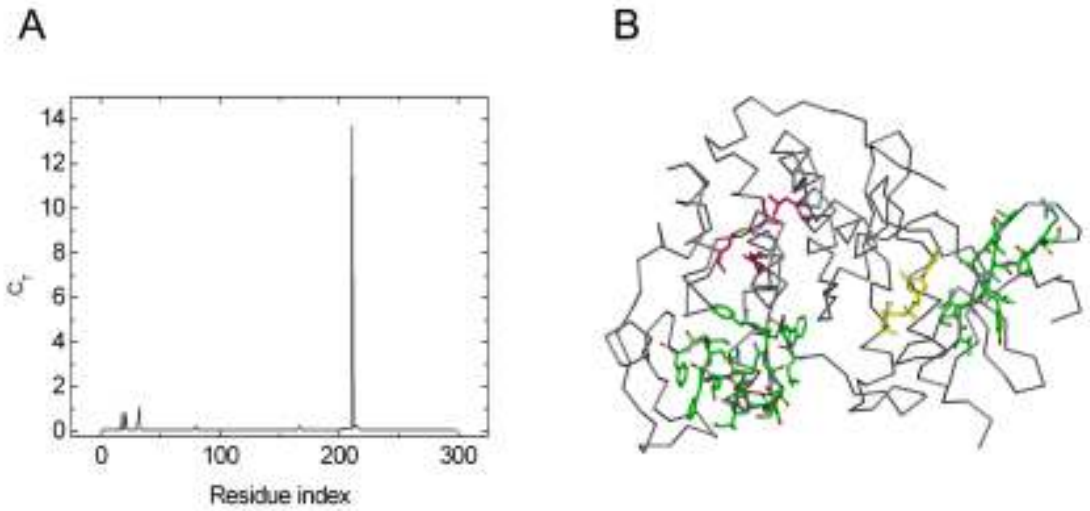
Figure S5 Kinase Domain TRP- Ca^{+2} Channel



a) Total correlation C_T of residues as a function of residue indices

b) Three dimensional structure of mouse kinase domain of Trp-Ca^{+2} channel with ADP-Mg^{+2} complex (yellow). Interaction path and the cliques are represented in green and pink, respectively.

Figure S6 Cyclin-Dependent Kinase

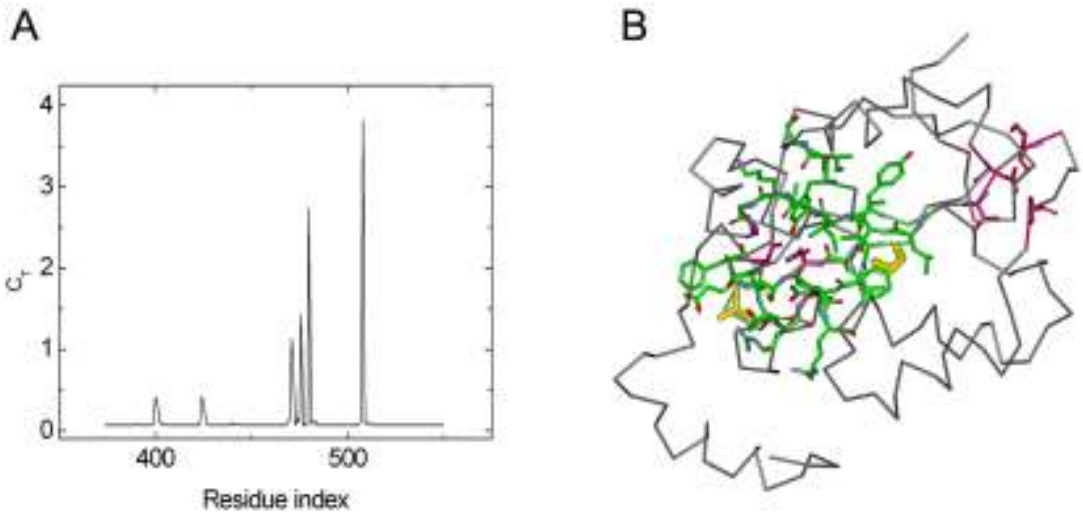


a) Total correlation C_T of residues as a function of residue indices

b) Three dimensional structure of human cyclin-dependent kinase chain A with Mg^{+2} and ATP(yellow). Interaction path and the cliques are represented in green and pink, respectively.

C. Hydrolases

Figure S7 M-phase inducer phosphatase 2 (Cdc25b)

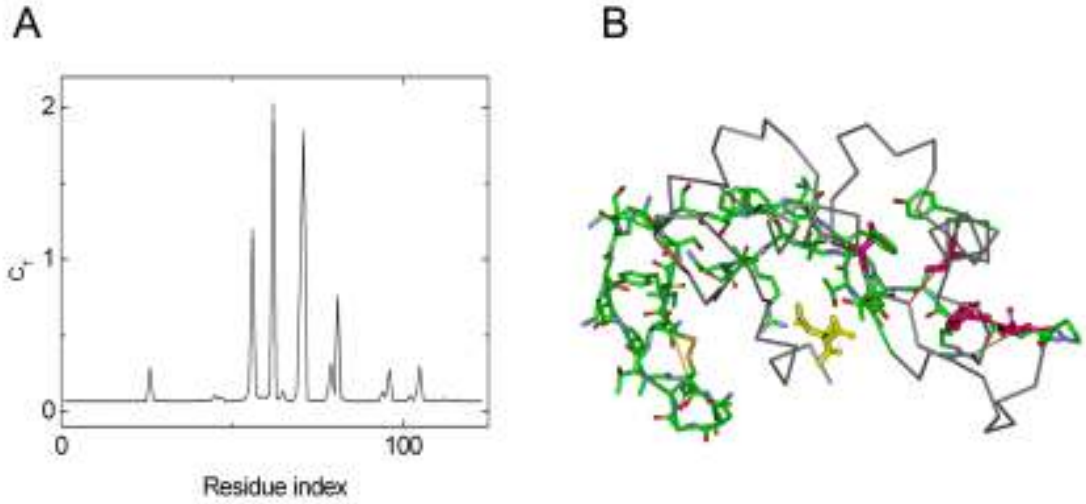


a) Total correlation C_T of residues as a function of residue indices

b) Three dimensional structure of human Cdc25b chain A with Tungstate(yellow). Interaction path and the cliques are represented in green and pink, respectively.

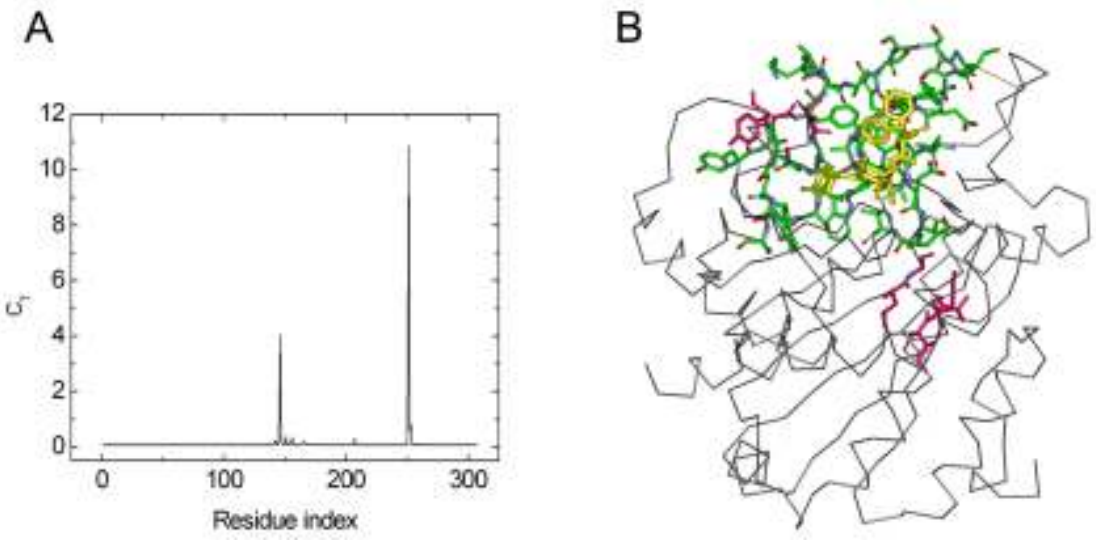
Figure S8 Angiogenin

a) Total correlation C_T of residues as a function of residue indices



b) Three dimensional structure of human angiogenin chain A with citric acid (yellow). Interaction path and the cliques are represented in green and pink, respectively.

Figure S9 Carboxypeptidase A

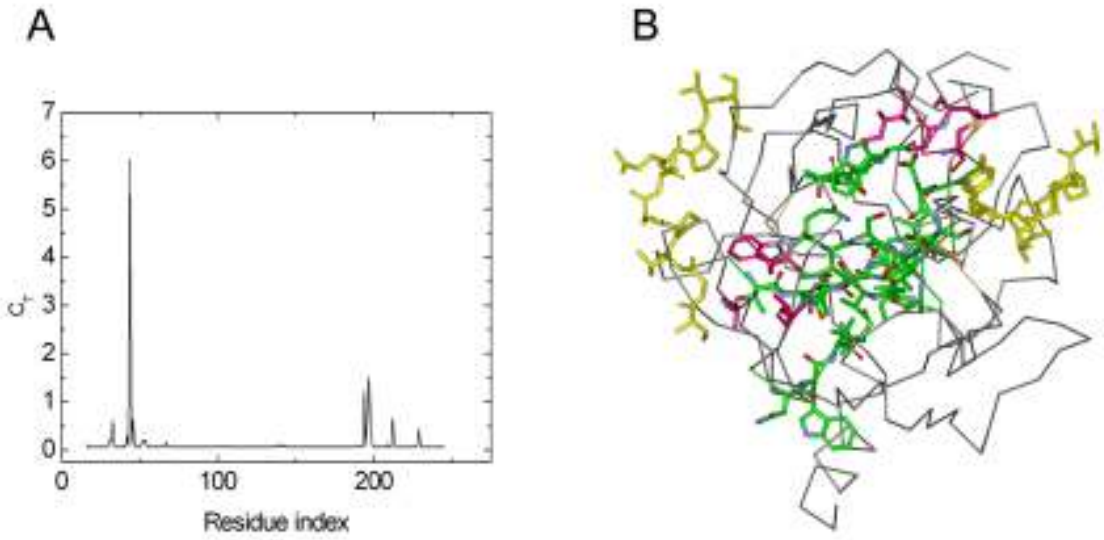


a) Total correlation C_T of residues as a function of residue indices

b) Three dimensional structure of bovine carboxypeptidase A chain A with a phosphonate (yellow). Interaction path and the cliques are represented in green and pink, respectively.

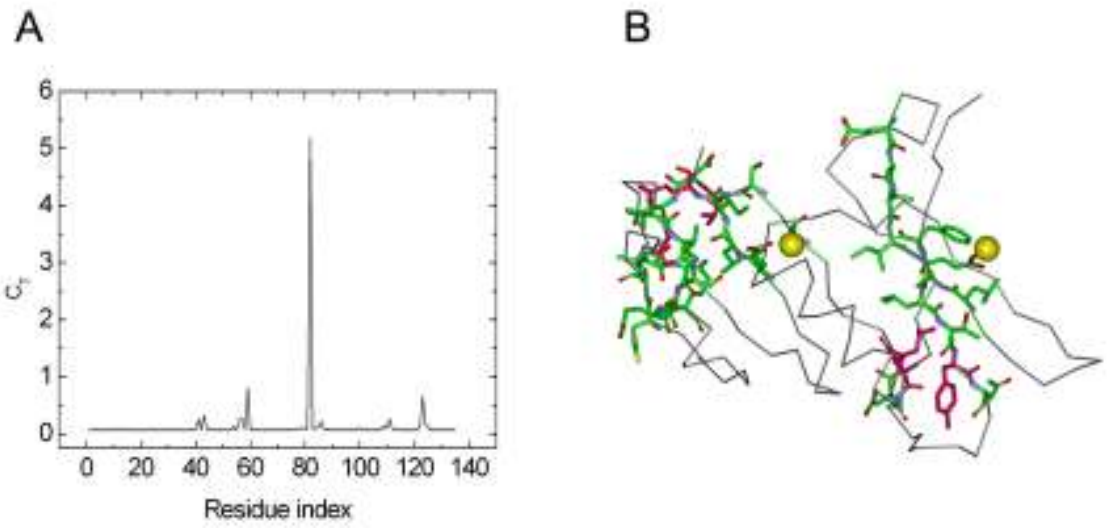
Figure S10 Gamma Chymotyrpsin

a) Total correlation C_T of residues as a function of residue indices



b) Three dimensional structure of bovine gamma chymotyrpsin chain B and C with two peptides (yellow). Interaction path and the cliques are represented in green and pink, respectively.

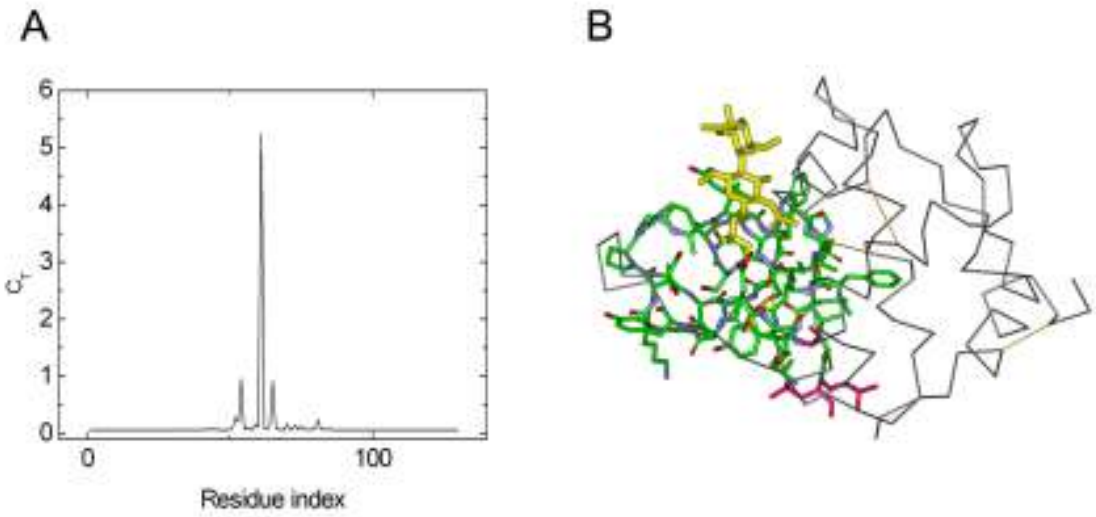
Figure S11 Glyoxalase I



a) Total correlation C_T of residues as a function of residue indices

b) Three dimensional structure of E.coli glyoxalase chain A with Zn^{+2} (yellow). Interaction path and the cliques are represented in green and pink, respectively.

Figure S12 Lysozyme

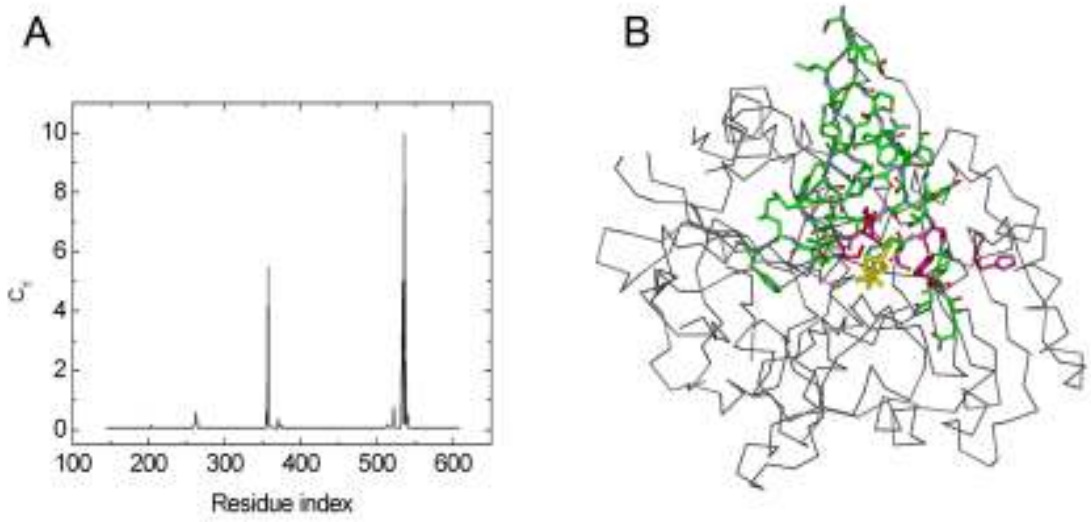


a) Total correlation C_T of residues as a function of residue indices

b) Three dimensional structure of human lysozyme with N,N'-Diacetylchitobiose (yellow). Interaction path and the cliques are represented in green and pink, respectively.

Figure S13 Tyrosyl-DNA phosphodiesterase

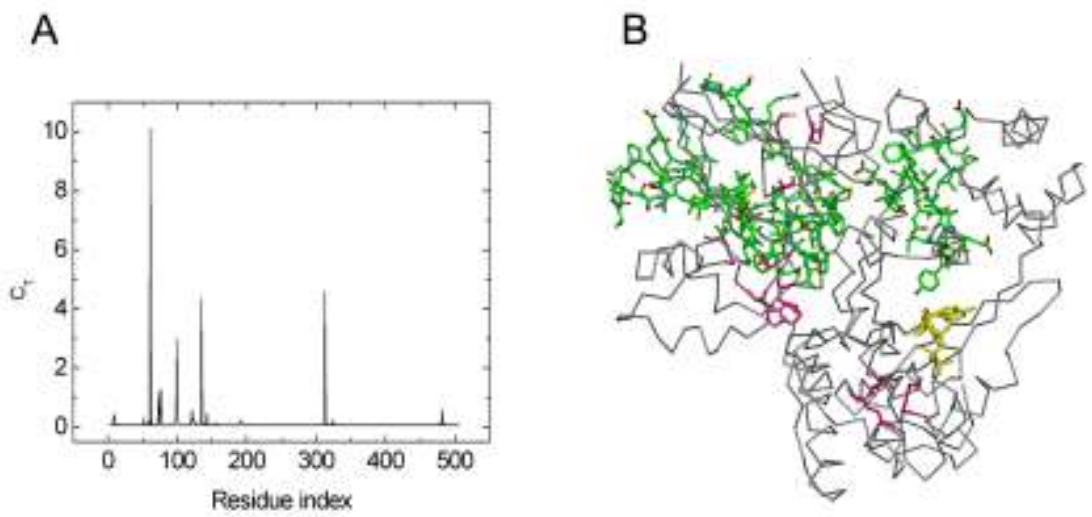
a) Total correlation C_T of residues as a function of residue indices



b) Three dimensional structure of human tyrosyl-DNA phosphodiesterase chain A with tungstate (yellow). Interaction path and the cliques are represented in green and pink, respectively.

Figure S14 Beta-lactam synthetase

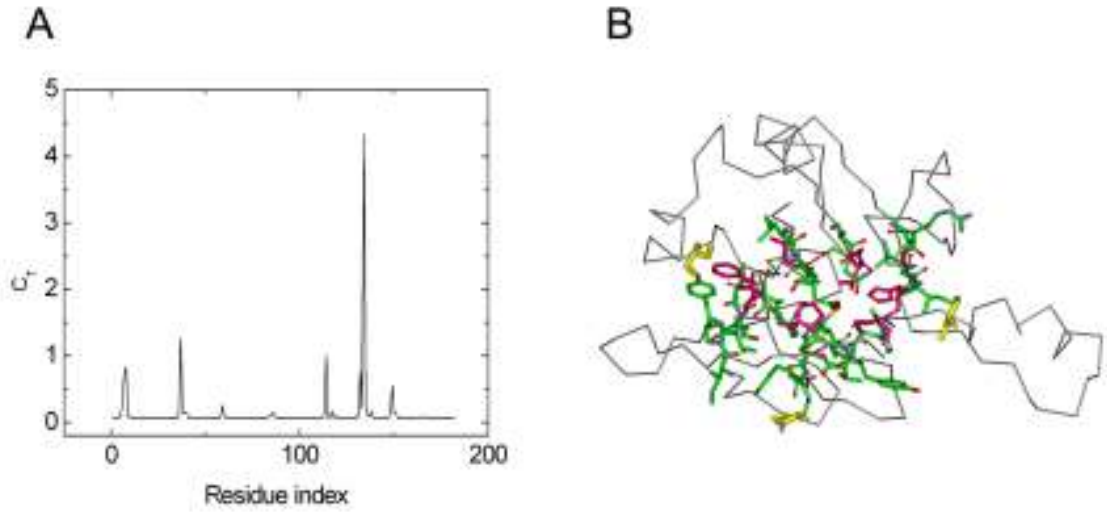
a) Total correlation C_T of residues as a function of residue indices



b) Three dimensional structure of streptomyces beta-lactam synthetase chain A with ATP(yellow). Interaction path and the cliques are represented in green and pink, respectively.

Figure S15 Vacuolar protein sorting Protein29 (VPS29)

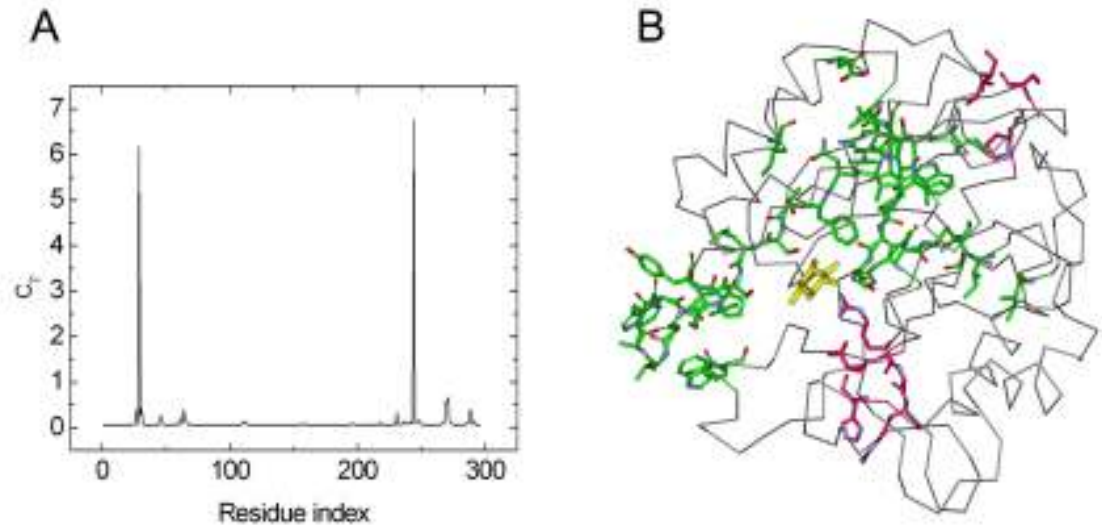
a) Total correlation C_T of residues as a function of residue indices



b) Three dimensional structure of mouse VPS29 chain A with Mn^{+2} and Glycerol (yellow). Interaction path and the cliques are represented in green and pink, respectively.

Figure S16 Phospholipase C

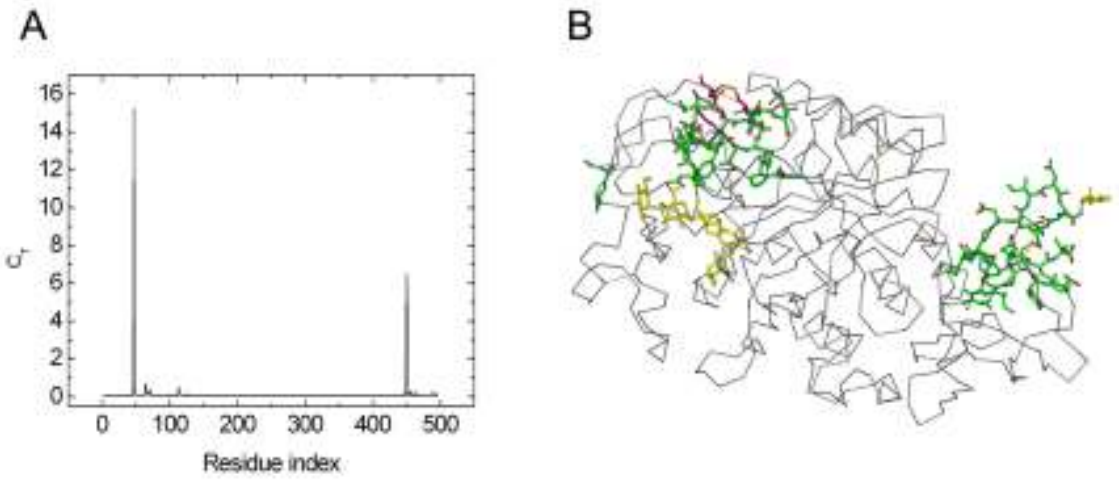
a) Total correlation C_T of residues as a function of residue indices



b) Three dimensional structure of phopholipase C chain A with Myo-Inositol (yellow). Interaction path and the cliques are represented in green and pink, respectively.

Figure S17 Pancreatic α-amylase

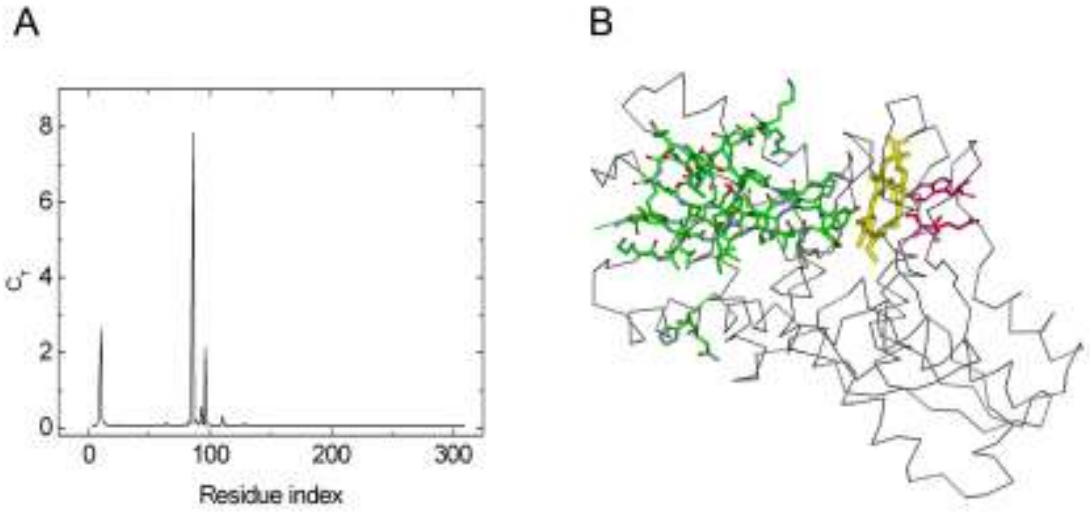
a) Total correlation C_T of residues as a function of residue indices



b) Three dimensional structure of human pancreatic α-amylase with acarbose (yellow). Interaction path and the cliques are represented in green and pink, respectively.

D. Lyases

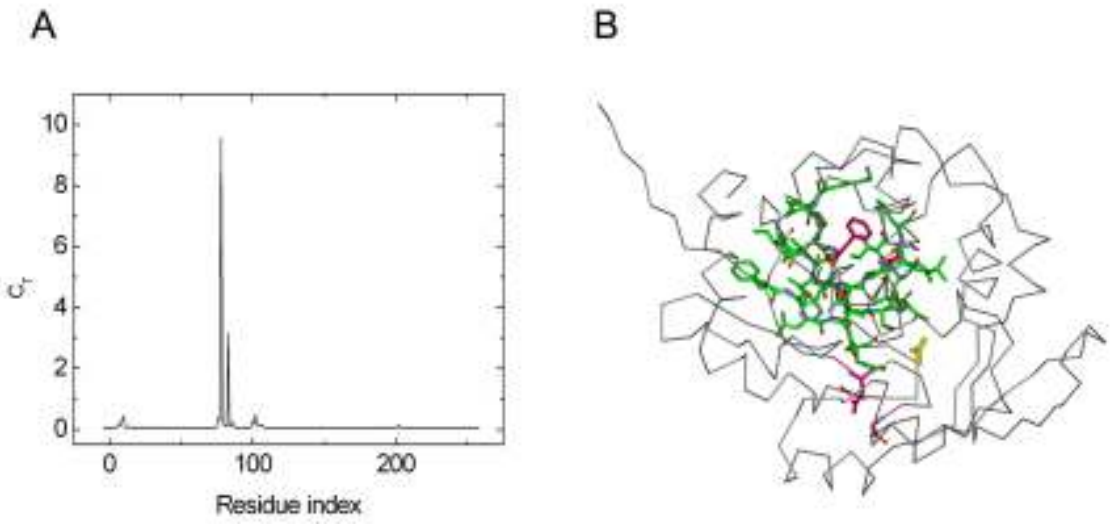
Figure S18 Ferrochelatase



a) Total correlation C_T of residues as a function of residue indices

b) Three dimensional structure of ferrochelatase chain A with N-Methyl mesoporphyrin (yellow). Interaction path and the cliques are represented in green and pink, respectively.

Figure S19 Hydroxynitrile Lyase

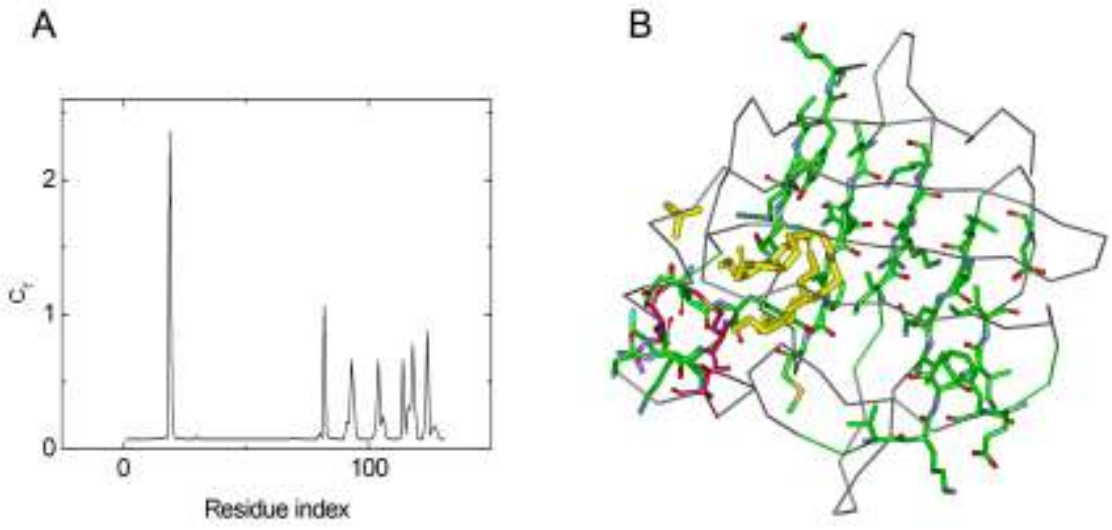


a) Total correlation C_T of residues as a function of residue indices

b) Three dimensional structure of hydroxynitrile lyase chain A with acetate(yellow). Interaction path and the cliques are represented in green and pink, respectively.

E. Non-enzymes

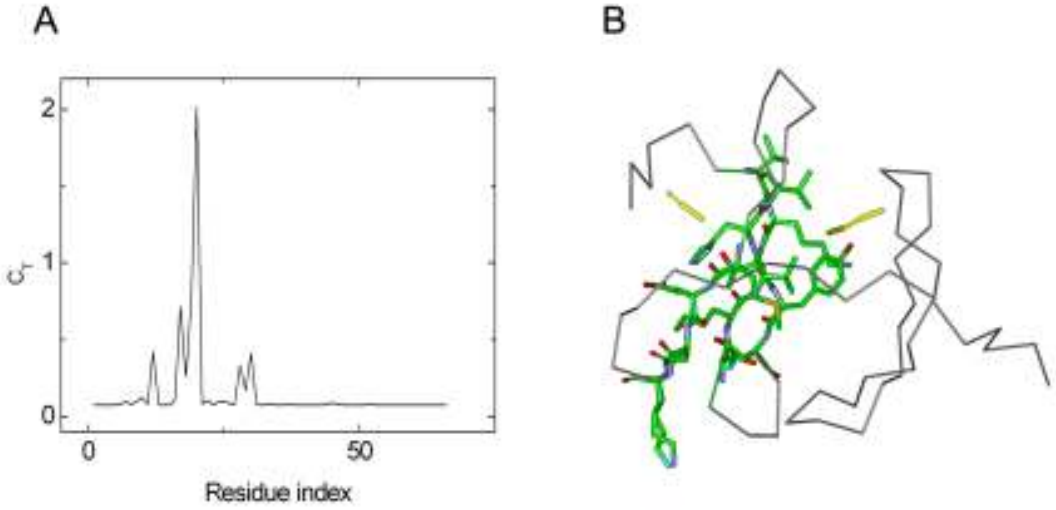
Figure S20 Adipocyte Lipid binding Protein



a) Total correlation C_T of residues as a function of residue indices

b) Three dimensional structure of mouse adipocyte lipid binding protein with oleic acid (yellow). Interaction path and the cliques are represented in green and pink, respectively.

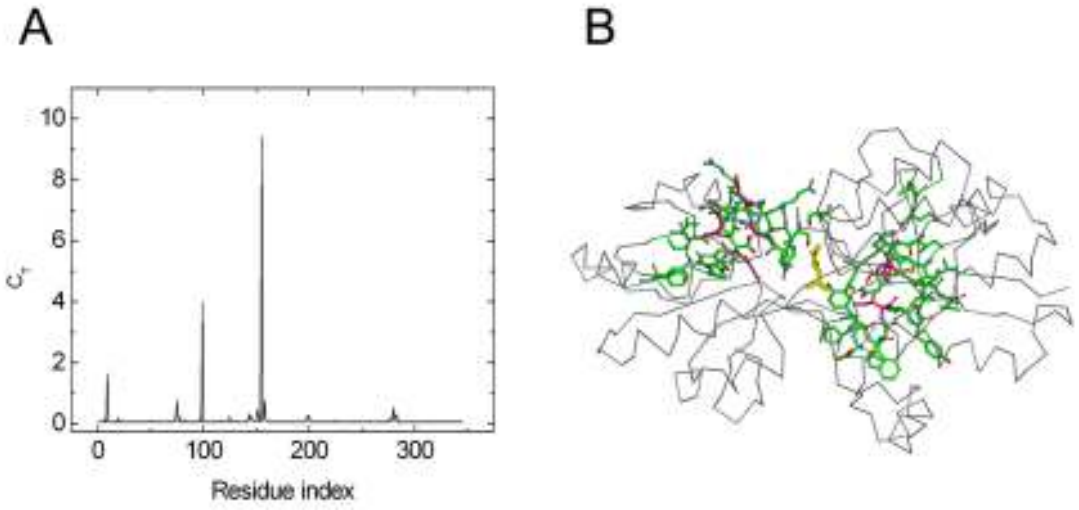
Figure S21 Copper Resistance Protein



a) Total correlation C_T of residues as a function of residue indices

b) Three dimensional structure of copper resistance protein chain A with Cu^{+2} (yellow). Interaction path is represented in green, respectively.

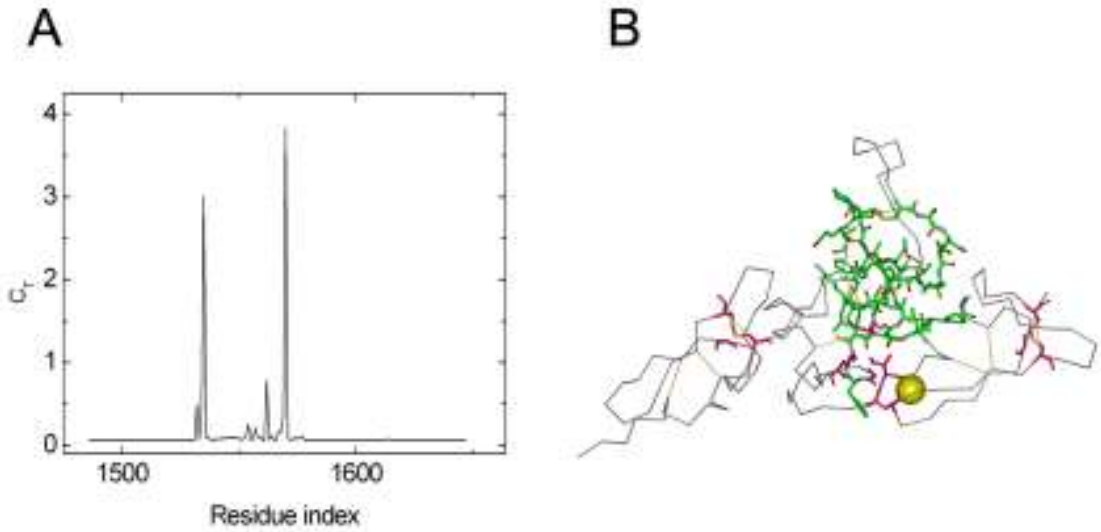
Figure S22 L-leucine Binding Protein



a) Total correlation C_T of residues as a function of residue indices

b) Three dimensional structure of L-leucine binding protein chain A with leucine (yellow). Interaction path and the cliques are represented in green and pink, respectively.

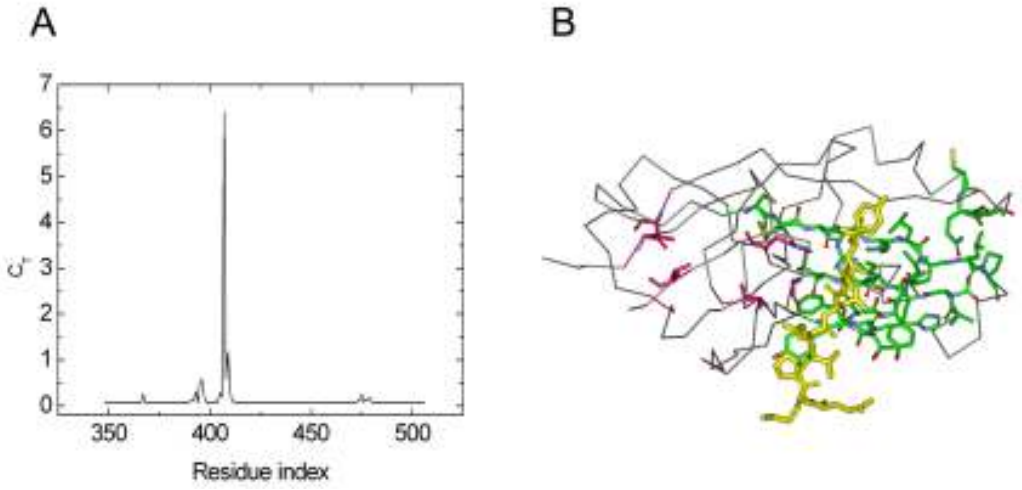
Figure S23 Fibrillin-1



a) Total correlation C_T of residues as a function of residue indices

b) Three dimensional structure of fibrillin-1 chain A with Ca^{+2} (yellow). Interaction path and the cliques are represented in green and pink, respectively.

Figure S24 TNF receptor associated factor (Traf 6)



a) Total correlation C_T of residues as a function of residue indices

b) Three dimensional structure of Traf6 with peptide(yellow). Interaction path and the cliques are represented in green and pink, respectively.

VITA

Ceren Tüzmen was born in Ankara, Turkey on September 25, 1985. She is an alumnus of Gazi Anadolu Lisesi, Ankara. She received her Bachelor of Science degree in molecular biology and genetics from Boğaziçi University, Istanbul, in June 2008.

From 2008 to 2010, she was a research and teaching assistant in the Computational Science and Engineering Department of Koç University. Her research includes the identification of binding sites in proteins using the Gaussian Network Model.