APPLICATION OF MIXED COARSE-GRAINED ELASTIC NETWORK MODEL TO PROTEIN-LIGAND INTERACTIONS

by

Nurver TEZCAN

B.S., Chemical Engineering, Boğaziçi University, 2012

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

Graduate Program in Computational Science and Engineering
Boğaziçi University

2014

ACKNOWLEDGEMENTS

I have been inspired and encouraged by several people who have helped me anyways to complete this thesis. I would like to express my gratitude to my thesis advisor Prof. Pemra Doruker Turgut for her patience and support throughout my study. It was honorary to be a part of the Polymer Research Center in square bloc and I owe this honor to her.

It was a pleasure to thank to my thesis committee members: Assoc. Prof. Elif Özkırımlı Ölmez and Assist. Prof. Özge Kürkçüoğlu Levitas for their valuable time they spend for evaluating and commenting on my thesis.

I have to mention about TÜBİTAK project (113M237) and BAP 5714 (10A05P8) project that support this thesis.

Many thanks to Zeynep Kürkçüoğlu Soner mentor of me and Arzu Uyar a master of me with her endless support in anytime anywhere and all people smiling at me in Polymer Research Center.

I am very grateful to my mum Nurhayat Tezcan and my dad Enver Tezcan for their absolute support, love and fun throughout my life, my dearest grandmother Fatma Tezel for her care in my childhood, and my beloved Aykut Özer for his easygoing personality and support. This thesis is indeed dedicated to them.

ABSTRACT

APPLICATION OF MIXED COARSE-GRAINED ELASTIC NETWORK MODEL TO PROTEIN-LIGAND INTERACTIONS

Ligand binding to an active or allosteric regulatory site of enzyme may lead conformational changes or changes in protein dynamics. The aim of this thesis is to explore the effect of ligand binding on vibrational dynamics of proteins, specifically the collective modes that are biologically relevant. For this aim, the mixed coarse-grained version of anisotropic network model, named MCG_ANM, was applied to more than 30 crystal structures including different monomeric and multimeric enzymes. In MCG_ANM, the ligands are considered at atomic (high) resolution, whereas the protein is at low resolution with only alpha-carbons considered. Solvents, catalytic ligands, catalytic loop conformational changes, allosteric ligands and coenzymes were considered in terms of their effects on vibrational frequencies of the enzymes. The frequencies of slow modes mostly shifted to higher values due to binding. Swapping of modes and changes in mode character were also observed. Solvents bound to specific regions, mainly catalytic and allosteric sites, were more effective in shifting frequencies. For the enzymes with catalytic loops that close over the active site, the frequency shifts due to the conformational change from open to closed loop were found to dominate the ligand binding effect. Among 17 allosteric protein structures considered, shifts in vibrational frequencies were higher for of three (out of four) monomers compared to dimers and tetramers based on cutoff distance of 7 Å. Moreover, mode swapping and character change was observed dominantly for these monomers. Specifically, allosteric ligand binding to monomeric, dimeric and tetrameric enzymes lead to 3.2%, 0.8% and 1.6% frequency changes, respectively, averaged over first five modes. The binding of allosteric ligands to monomers and dimers also lead to significant changes in the orientational correlations of catalytic residues. Finally, coenzymes bound to tetramers lead to higher frequency shifts compared to allosteric ligands and cofactors.

ÖZET

KARIŞIK KABA-ÖLÇEKLİ ELASTİK AĞ YAPI MODELİNİN PROTEİN-LİGAND İLİŞKİLERİNE UYGULANMASI

Enzimin aktif veya allosterik düzenleyici bölgesine ligant bağlanması proteinde konformasyonel değişikliklere veya protein dinamiğinde değişikliklere neden olur. Bu tezin amacı ligant bağlanmasının özellikle biyolojik olarak ilgili kollektif modlarda proteinlerin titreşimsel dinamiğine etkisini araştırmaktır. Bu amaçla, anizotropik ağ yapı modelinin karışık kaba ölçekli (MCG_ANM) sürümü, çeşitli tek ve mültimerik enzimlerin toplam 30'dan fazla kristal yapısına uygulandı. MCG_ANM'de, ligantlar atomik yani yüksek çözünürlükte proteinler ise alfa-karbon atomları hesaba katılarak düşük çözünürlükte hesaplanmıştır. Solventler, katalitik ligantlar, konformasyonel değişiklik geçiren katalitik ilmekler, allosterik ligantlar ve koenzimler, enzimlerin titreşimsel frekanslarına olan etkileri açısından değerlendirilmiştir. Yavaş modların frekansları ligant bağlanmasından dolayı çoğunlukla daha yüksek değerlere kaymıştır. Modların yer değiştirmesi ve mod karakterindeki değişimler de gözlenmiştir. Katalitik ve allosterik bölgelere bağlı solventler frekans değişiminde daha etkindir. Katalitik ilmiklerin enzimin aktif bölgesi üzerine kapanması ile ilgili konformasyonel değişimden kaynaklanan frekans değişimleri, ligant bağlanmasına oranla daha yüksektir. 7 Å etkileşim yarıçapı kullanılarak 17 allosterik protein yapısı değerlendirildiğinde, titreşimsel frekanslardaki kaymalar üç monomerik yapı için dimerler ve tetramerlere göre daha yüksektir. Mod yer değişimi ve karakter değişiminin de bu monomerler için daha yüksek olduğu gözlenmiştir. Özellikle, monomerik, dimerik ve tetramerik enzimlere allosterik ligant bağlanması, ilk beş modun ortalaması alınarak sırasıyla % 3.2, % 0.8 ve % 1.6 frekans değişimine neden olmuştur. Allosterik ligantların monomer ve dimerlere bağlanması katalitik rezidülerin yönelimsel korelasyonlarında önemli değişikliklere sebep olmaktadır. Son olarak, tetramerlere bağlı koenzimler allosterik ligant ve kofaktörlerle kıyaslandığında, daha yüksek frekans kaymalarına neden olmuştur.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

Å	Angstrom
C_α	Carbon Alpha
H	Hessian matrix
k	Number of interaction between atom and coarse-grained residues
K	Kelvin
m	Number of node-node interaction
N	Total number of residues
r_c	Cutoff distance
$R_{c,a}$	Atomistic cutoff
$\mathbf{R}_i$	Instantaneous position vector of residue *i*
$\mathbf{R}_i^o$	Equilibrium position vector of residue *i*
$\mathbf{R}_{ij}$	Distance between site i and j
U	Eigenvector of the connectivity matrix
v	Eigenvector of overlap
V	Potential energy
ω	Eigenvector of overlap
$\Delta\mathbf{R}_i$	Fluctuation vector of residue *i*
$\langle\Delta\mathbf{R}_{ij}^2\rangle$	Relative mean square fluctuation of residue *i* and *j*
λ	Eigenvalue of the connectivity matrix
γ	Force constant
Γ_{ij}	Kirchhoff matrix element of residues *i* and *j*

LIST OF ACRONYMS/ABBREVIATIONS

ANM	Anisotropic Network Model
ASD	Allosteric Database
BLZPACK	Block Lanczos Package
CASTp	Computed Atlas of Surface Topography of proteins
cg	Coarse-grained
ENM	Elastic Network Model
GNM	Gaussian Network Model
HEX	Hexane
KNF	Koshland-Nemethy-Filmer model
LIG_ANM	Ligand binded Anisotropic Network Model
MCG_ANM	Mixed coarse-grained Anisotropic Network Model
MWC	Monod-Wyman-Changeux model
NMA	Normal Mode Analysis
PDB	Protein Data Bank
R	Relaxed State
RMSD	Root Mean Square Deviation
T	Tense State
TIM	Triosephosphate Isomerase

1. INTRODUCTION

Ligands bind to proteins at specific regions, such as an enzyme's active site or allosteric regulatory sites. Different models have been used to explain ligand binding, namely lock and key, induced fit and population shift models. Induced fit model have changed the original lock and key perspective of fitted shapes into unfitted ones with conformational rearrangements leading to the active state (Koshland, 1958). Besides that, the population shift model, being the more recent one, indicates that populations of available conformational states change upon ligand binding (Kumar *et al.*, 2000).

The conformational changes observed or stabilized upon ligand binding are crucial for protein function. However, ligand binding does not always lead to conformational changes, but may cause changes in protein dynamics. It is currently not possible to gain full understanding conformational transitions in proteins by experimental techniques, such as x-ray crystallography or electron microscopy. Computer simulations may aid in this respect, which can provide specific snapshots along the transition, i.e. describe pathways. Specifically, coarse-grained models are important in terms of decreasing computing time (Tozzini, 2005; Chu and Voth, 2007) for the study of ligand binding effect on protein conformational dynamics.

Elastic network model (ENM) is a coarse-grained normal mode analysis that is computationally efficient. In ENM, vibrational analysis of the system is performed by forming a network of interconnected springs using C_α atoms as nodes. Potential energy is a summation over all harmonic interactions between close-neighboring node pairs (Doruker *et al.*, 2000; Atilgan *et al.*, 2001). With these coarse-grained models, slowest or low-frequency modes are investigated because they describe the functionally important, collective motions of the system (Bahar and Jernigan, 1998). ENM is further developed using mixed coarse-graining (MCG) (Kurkcuoglu *et al.*, 2004; Kurkcuoglu *et al.*, 2009), where the ligand can be described at atomistic resolution and the protein at coarse-grained level.

The objective of this thesis is to observe the extent to which different type of ligands affect the vibrational dynamics of proteins, specifically the collective modes that are biologically relevant. Literature review and methodology are discussed in Chapters 2 and 3, respectively. Results and discussion comprises Chapter 4, which is formed of three subsections. In first section of results, effects of solvent binding to three different enzymes are analyzed with standard ANM (Doruker *et al.*, 2000; Atilgan *et al.*, 2001) and mixed coarse-grained ANM (Kurkcuoglu *et al.*, 2004; Kurkcuoglu *et al.*, 2009). Then, seven enzymes with open and closed conformations of a catalytic loop are investigated with MCG_ANM method. The conformational change from open state to closed state and ligand binding to the closed state of the enzymes are studied to determine changes in vibrational dynamics. In the third section, 17 enzymes from the allosteric database (ASD) (Huang *et al.*, 2011; Huang *et al.*, 2014) are chosen to investigate the effects of mainly allosteric ligand binding on collective modes. The last chapter of the thesis comprises conclusion and recommendations.

2. LITERATURE SURVEY

Proteins are dynamic structures built from sequences of amino acids. Twenty amino acids have various properties. Each amino acid has a backbone, which is the main chain formed by peptide linkages. Each monomer or amino acid residue consists of a C_α atom bound to the side chain, carbonyl and amino groups. Since each residue has one C_α atom, computational models may focus on only C_α atoms for representing coarse-grained residues in order to observe the conformational dynamics of structures. Proteins sustain collective domain motions like hinge bending and shear motion, which may be related to their biological activity. Proteins may also undergo conformational transitions that are crucial for their function. These transitions, such as those between open-closed or on-off states, may be triggered and stabilized upon ligand binding.

2.1. Protein-Ligand Interactions and Allostery

Enzyme ligand interaction was firstly explained with lock-and-key model by Fischer at the end of 19th Century. In this model, the lock-key demonstrates enzyme-ligand interaction and in that interaction only the correctly sized key (ligand) can bind to the key hole (active site) of the lock (enzyme). In the middle of 20th Century, the induced fit model was purposed stating that ligands can bind to unfitted regions and induce conformational changes in the active site that result in a new active conformation of the enzyme (Koshland, 1958). In other words, conformational changes occur during substrate binding.

Ligand binding processes may be categorized. For instance, in enzyme-substrate binding, substrate is converted into the product in a chemical reaction with the aid of the enzyme. On the other hand, in receptor-ligand binding, the ligand causes some conformational changes in the protein, thereby affecting its function related to signal transduction.

There are ligands binding to sites that are different and distant from the active site of an enzyme and regulate its activity. This type of ligand binding, which can induce protein's function by causing changes in protein's structure, is called "allostery". Allostery induces regulating and signaling effect on biological systems. The proteins with allosteric

behaviors are admitted as major component of biological systems, since they react instantly and irreversibly to other factors of the system (Monod *et al.*, 1963). Monod, Wyman, Changeux (MWC) model notifies that allosteric systems exist at least in two conformational states R (relaxed) and T (tense or taut) with homotropic (identical ligands) and heterotropic (non-identical) effects by preexisting equilibrium mechanism. MWC model proposed that before substrate binding to protein, all subunits in T conformation have low affinity, but after binding they turn to high affinity R state showing positive cooperativity (Monod *et al.*, 1965). In Figure 2.1, subunits of protein change shape to protect the symmetry of the molecule while changing from T to R conformation due to ligand binding.

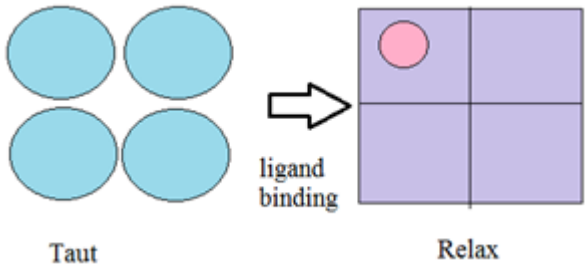


Figure 2.1. MWC model demonstration.

In KNF (Koshland Nemethy Filmer) "sequential" model (Figure 2.2), subunits forming the protein structure change conformation, individually, where ligand binding to each subunit may lead to some changes in other subunits (Koshland *et al.*, 1966; Koshland and Hamadani, 2002). Through KNF model, negative cooperativity between subunits could also be explained, which occurs if ligand binding to a subunit hinders binding to other subunits (Levitzki and Koshland, 1969). When comparing KNF and MWC, KNF model invalidates the assumption of symmetry but it assumes there is a move from T to ligand bound R state with sequential model.

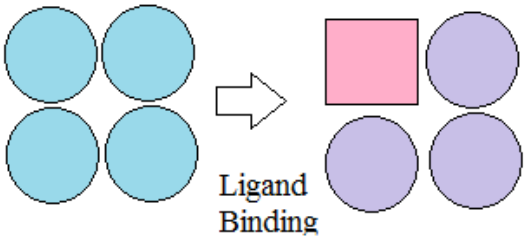


Figure 2.2. KNF model demonstration.

In addition to two allosteric regulation models, Weber asserted the “population shift” idea which indicates ligand binding shifts the set of conformational states (Weber, 1972). This idea is consistent with recent ones, which propose that allostery can take place in systems even in the absence of a conformational change (Tsai *et al.*, 2008) and that all proteins may show allosteric behaviors even if they are known as non-allosteric (Gunasekaran *et al.*, 2004). The ensemble allosteric model consists both MWC and KNF models that asserts different domains can have different conformations with and without the relation between these different domains (Williams and Hall, 2014).

At the same time, ligand binding to the allosteric site tends to cause reversible conformational transition that also triggers active site effectiveness. Physical changes that are not directly related to the binding properties of the proteins is another focal point, which indicates that the interaction between the subunits especially in original quaternary structures express allosteric relationship such as ATCase (Changeux *et al.*, 1968). Morpheein model considers monomeric conformational change that affects oligomerization states in the structure (Jaffe, 2005; Jaffe and Lawrence, 2012).

Recently, the structural data available on allosteric proteins have been collected in the so-called Allosteric Database (ASD), which currently contains structures for more than 1000 allosteric proteins combined with about 22000 allosteric ligands (Huang *et al.*, 2011; Huang *et al.*, 2014). ASD supplies a resource for studies concentrated on allosteric macromolecules. Besides that, there are servers for predicting possible binding sites of allosteric proteins. PARS (Protein Allosteric and Regulatory Sites) server (Panjkovich and Daura, 2014) and SPACER (Server for Predicting Allosteric Communication and Effects of Regulation) server (Goncearenco *et al.*, 2013) predict possible allosteric ligand binding pockets and display allosteric interaction between sites.

2.2. Ligand Binding Effect on Protein Dynamics

Ligands that bind to enzymes are substrates, coenzymes and cofactors i.e. structural supplements of proteins. Protein ligand interaction is presumed to originated from hydrogen bonds, van der Waals interactions, hydrophobic and hydrophilic properties of its location, temperature, pressure, pH, entropy, but (Balog *et al.*, 2004). Physically, ligand

binding to proteins causes extra energy to the systems (Eren and Alakent, 2013). Signals arise from local interaction of ligand can transfer to distant sites of the proteins which is known as allostery (Cui and Karplus, 2008).

The energy landscape of a protein is complicated and consist many local minima. These local minima demonstrate different conformational states of proteins (Thielges and Fayer, 2012). Change in energy landscape of proteins is a kind of signal transmission in molecular level. Possible ligand binding to the systems shifts the equilibrium population on a rugged energy landscape. This is also consistent with population shift model on allosteric systems which supports ensembles of conformational changes before and after ligand binding affects their dynamics (Smock and Gierasch, 2009), as shown for a hypothetical two-state system in Figure 2.3.

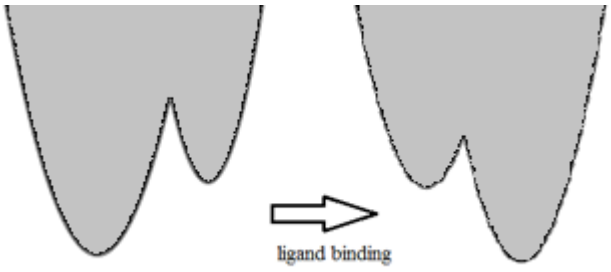


Figure 2.3. Energy landscapes demonstration of conformational change.

Landscape theories are also used for predictions of protein folding process but, since it is hard to measure free energy profiles of proteins, the landscapes are calculated from simulations (Yu *et al.*, 2012).

Ligand binding is crucial for enzymatic reactions, signal transduction, and allosteric regulation. It causes enthalpic and entropic changes due to hydrophobic effects, intramolecular vibrations, conformational entropy, electrostatic interactions and hydrogen bonds (Sturtevant, 1977). Binding may leads to decrease in entropy by either shifting of the vibrational frequencies to higher values or decrease in iso-energetic conformational substrate. Even binding of a small molecule without changing the structure can be effective on the binding free energy and as some studies indicate that without ligand binding, proteins are more flexible (Alakent *et al.*, 2011). On the other hand, NMR relaxation

experiments on mouse major urinary protein has indicated increased in backbone flexibility due to ligand binding i.e. increase in vibrational entropy (Zidek *et al.,* 1999). Normal mode analysis of unbound and water bound structures of pancreatic trypsin inhibitor has also shown decreasing vibrational frequencies with increasing entropy and flexibility (Fischer and Verma, 1999). Through neutron scattering investigations of ligand bound and unbound states at 120 K have displayed shifting of vibrational frequencies to low values (Balog *et al.*, 2004). However, there has been increase in normal modes of vibrational frequency with ligand binding on diguanylate cyclase PleD (Schmid and Meuwly, 2007).

Biological signal transferring happens in a wide range of time; especially conformational arrangements take place slowly. The functional mechanism of proteins may be explained within the information of molecular conformations and dynamics. There are various internal motions in globular proteins. Specific structural transitions may last more than 10^2 seconds but localized motions last around 0.1 ps (Careri *et al.*, 1975). Hinge-type motions of the protein can be observed with molecular dynamic simulation, such as the opening/closing of domains in the structure. However, these movements take long time especially for large proteins, which makes it impossible to get information about the ligand binding effects on the dynamic character of the protein. To increase the efficiency of computations for large system sizes, proteins can be modeled using the elastic network approach. ENM (elastic network model) is a coarse-grained normal mode analysis to get information about the conformational dynamics of protein structures (Tirion, 1996; Bahar *et al.*, 1997; Atilgan *et al.*, 2001; Doruker *et al.*, 2002; Tama *et al.*, 2002). Alpha-carbon atoms of residues are taken into consideration as nodes of the network, which are interconnected by springs. There is correlation between ENM low frequency modes and the conformational changes observed upon ligand binding (Tama and Sanejouand, 2001). To sum up, internal motions of proteins are important for protein function and structural transitions. And the low frequency modes are related to functional properties of proteins and conformational changes.

3. METHODS

3.1. Anisotropic Network Model (ANM)

Anisotropic network model refers to the three-dimensional (3d) version of elastic network model, where anisotropic residue fluctuations could be obtained through vibrational analysis of biological systems (Doruker *et al.*, 2000; Atilgan *et al.*, 2001). In the network, alpha-carbon atom of each residue is used as a node with unit mass. And the total potential energy of the system is a summation of over harmonic interactions between node pairs that fall within a predefined cutoff radius.

$$V=\frac{\gamma}{2}\sum_{i}^{N}\sum_{\substack{j>i \\ R_{ij}^{0}\le R_c}}^{N}\left(\mathbf{R}_{ij}-\mathbf{R}_{ij}^{0}\right)^2 \tag{3.1}$$

Here $\mathbf{R}_{ij}$ and $\mathbf{R}_{ij}^{0}$ are instantaneous and equilibrium distances between nodes i and j, respectively. N represents the total number of coarse-grained residues or nodes in the protein. The equilibrium distances ($\mathbf{R}_{ij}^{0}$) are taken as those in the crystal structure, thus energy minimization of the system are not required prior to normal mode analysis. γ is the force constant that assigns a uniform strength for all residue-residue interactions, whether the pairs are bonded or non-bonded to each other in the protein structure. If all the modes of the system are calculated, γ is adjusted so as to match the average mean square fluctuations of the residues from theory with that from experiments. If only the slow modes are calculated, then γ is taken as one.

Overall potential energy of the system with N residues is expressed in matrix notation as:

$$V=\frac{1}{2}\Delta R^{T}\boldsymbol{\mathcal{H}}\Delta R \tag{3.2}$$

$\boldsymbol{\mathcal{H}}$ is the (3Nx3N) Hessian matrix that composed of the second derivatives of potential energy for three coordinates with combination of three by two dimensions that

equals three, and for each N residues $\Delta\mathbf{R}$ is a 3N dimensional vector of the positional vector fluctuations, ΔR_i contains all sites, $\Delta\mathbf{R}^T$ is transpose of $\Delta\mathbf{R}$.

As in normal mode analysis, force constants of the system are used with the help of Hessian matrix:

$$\boldsymbol{\mathcal{H}}=\begin{bmatrix} \boldsymbol{H}_{11}\,\boldsymbol{H}_{12}\ldots\boldsymbol{H}_{1N} \\ \boldsymbol{H}_{21}\,\boldsymbol{H}_{22}\ldots\boldsymbol{H}_{2N} \\ . \quad\quad . \\ . \quad\quad . \\ . \quad\quad . \\ \boldsymbol{H}_{N1}\boldsymbol{H}_{N2}\ldots\boldsymbol{H}_{NN} \end{bmatrix} \tag{3.3}$$

For $i \neq j$ i.e. off diagonal elements of the Hessian matrix:

$$\boldsymbol{H}_{ij} = \begin{bmatrix} \frac{\partial^2 V}{\partial X_i \partial X_j} & \frac{\partial^2 V}{\partial X_i \partial Y_j} & \frac{\partial^2 V}{\partial X_i \partial Z_j} \\ \frac{\partial^2 V}{\partial Y_i \partial X_j} & \frac{\partial^2 V}{\partial Y_I \partial Y_j} & \frac{\partial^2 V}{\partial Y_i \partial Z_j} \\ \frac{\partial^2 V}{\partial Z_i \partial X_j} & \frac{\partial^2 V}{\partial Z_i \partial Y_j} & \frac{\partial^2 V}{\partial Z_i \partial Z_j} \end{bmatrix} \tag{3.4}$$

Using singular value decomposition technique 3N eigenvalues and eigenvectors of the system can be calculated to get information about the vibrational dynamics of the structure (Alter *et al.*, 2000). Eigenvalues correspond to frequencies, and eigenvectors give shapes of individual modes. Especially the low-frequency end of the spectrum consists of the so-called collective and functional modes that are critical for the conformational changes and dynamics and can be linked to protein's specific function (Gniewek *et al.*, 2012).

Alternatively, if one is only interested in the slow modes that form the low-frequency end of the spectrum, the computationally efficient BLZPACK (Grimes *et al.*, 1994; Marques and Sanejouand, 1995) can be used for calculating eigenvalues and eigenvectors.

3.2. Treatment of Protein-Ligand Interactions by ANM

The effect of small ligand binding on protein vibrational dynamics was considered by two alternative ANM methodologies in this thesis. When binding of a ligand to a protein introduces additional interactions to the system, both approaches treat these interactions as being uniform and non-directional as those within the protein itself by ANM.

The first approach, called LIG_ANM, is a straightforward extension of ANM discussed in previous section. Here, additional nodes with unit mass are placed either at all or selected heavy atom(s) of the ligand. As a result, the total number of nodes and interactions in the system increases, but all pairwise harmonic interactions (residue-ligand, residue-residue, ligand-ligand) are treated with the same γ. However, placement of the ligand nodes seems arbitrary considering the diversity of the ligands available in crystal structures, such as solvents, cofactors, inhibitors and allosteric ligands. Moreover, treating each ligand atom as having unit mass enhances the effect of ligand since the coarse-grained residue masses in ANM are also considered to be unity. A uniform cutoff distance (R_c) of 15 Å is used for this purpose (Eyal *et al.*, 2006).

Thus, it is clear that a more refined approach is necessary to take care of the non-homogeneity in terms of masses and spring constants between ligand atoms and coarse-grained residue nodes. Such an approach, named mixed coarse-grained ANM (MCG_ANM) (Kurkcuoglu *et al.*, 2009) has been developed recently to analyze vibrational dynamics of protein structures composed of regions with different resolutions, namely high and low resolution. The high resolution region of the structure considers all heavy atoms except hydrogen atoms of the residues or ligands of interest, whereas the low resolution region considers just the center of mass of alpha carbon atoms of residues involved (Kurkcuoglu *et al.*, 2009). Such an approach has proven effective for analyzing the functional sites of supra molecular assemblages, such as the ribosome, at atomistic detail by reducing computational time significantly due to the low-resolution parts involved.

In MCG_ANM, the molecular weight of any heavy atom is counted as one without distinguishing between atom types. Similarly, the mass of a coarse-grained (cg) node or residue is equal to the total number of heavy atoms it comprises. Thus, the Hessian matrix should be mass-weighted to take care of the non-uniform masses into account in this approach. Moreover, the pair-wise interactions are also scaled by using a predetermined atomistic cutoff ($R_{c,a}$) for this aim. All atom-atom pair wise interactions within $R_{c,a}$ are assigned a uniform force constant of γ, which can be taken as unity. For interactions between two cg (low-resolution) nodes/residues, if the total number of atom pairs that belong to different nodes and are closer than $R_{c,a}$ is m, then the force constant between the nodes is $m\gamma$. Similarly, the force constant is $k\gamma$ between an atom and a cg residue, where k equals the total number of pairwise interactions the specific atom makes with the cg residue based on $R_{c,a}$. In this thesis, two alternative atomistic cutoff values are used, namely $R_{c,a}$ = 7 and 10 Å. The ligands are considered in atomistic detail, i.e. at high-resolution, whereas the cg protein residues are low-resolution.

In Figure 3.1, the circles are low resolution node node interaction which is $m\gamma$, gray lines show ligand coarse-grain interactions which are $k\gamma$ and γ shows atom-atom interaction.

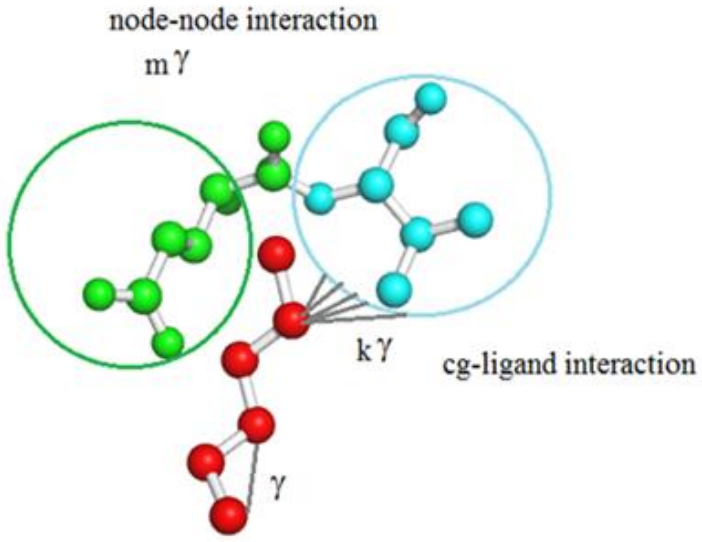


Figure 3.1. MCG_ANM presentation.

3.3. Analysis of Data

3.3.1. Overlap

The overlap value gives the overall correlation between two eigenvector subspaces that may be resulting from independent ANM calculations on the same protein. In the following overlap definition, the inner dot product between 3N-dimensional eigenvectors, υ_i and ω_i, is calculated and averaged over all possible (i,j) combinations based on k modes forming each subspace.

$$\text{Overlap}=\left(\frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{k}(\upsilon_i\omega_j)^2\right)^{1/2} \tag{3.5}$$

The resulting overlap value is between 0 and 1. If it is close to 1, there is high correspondence between the two subspaces. In this thesis, the overlap is used to determine the correspondence between ANM calculations that are either carried on apo vs. bound conformers or closed conformers in the presence vs. absence of ligands.

To determine whether there is one-to-one correspondence between the subspaces with no mode swapping, the diagonal overlap is also calculated using the following equation;

$$\text{Overlap (diagonal)} = \left(\frac{1}{k}\sum_{i=1}^{k}(\upsilon_i\omega_i)^2\right)^{1/2} \tag{3.6}$$

3.3.2. Pocket Size

Pocket sizes of the protein-ligand binding regions are calculated with CASTp: computed atlas of surface topography (Dundas *et al.*, 2006). The server focuses on functional residues for determination of surface pockets of structures and gives surface area and volume of the pockets. CASTp server applies Delaunay triangulation and the alpha complex for shape measurements of the area and volume of each pocket and cavity.

It provides identification and measurements of surface accessible pockets as well as interior inaccessible cavities, for proteins and other molecules.

3.3.3. Hinge Residues

Hinge residues are found with the help of HingeProt (Emekli *et al.*, 2008). HingeProt predicts the flexible hinge regions connecting them in the native topology of protein chains by using elastic network models. HingeProt uses both Gaussian Network Model (GNM) (Bahar *et al.*, 1997) and Anisotropic Network models (ANM) (Atilgan *et al.*, 2001) for calculations.

4. RESULTS AND DISCUSSION

The original ANM (Atilgan *et al.*, 2001) with residue-based cg is applied by including all heavy atoms of the ligand. This approach is named as LIG_ANM in this thesis. The second method, called MCG_ANM, is an extension of ANM that includes high-resolution ligand and low-resolution protein residues, i.e. the overall system is mixed cg (Kurkcuoglu *et al.*, 2004; Kurkcuoglu *et al.*, 2009). In the next section, both methods are considered in order to determine, which one is more proper for incorporating ligand effect on vibrational dynamics. Later using two different cutoff distances in MCG_ANM, the effect of orthosteric or allosteric ligands bound to different proteins, mostly enzymes, are investigated.

4.1. Frequency Shift due to Organic Solvent Binding via LIG_ANM and MCG_ANM

LIG_ANM (R_c = 15 Å) and MCG_ANM ($R_{c,a}$ = 7 Å) are applied to three different enzymes, which carry either bound hexane or hex-5-ene-1, 2 diol molecules as ligand. These specific cutoff values fall in the ranges of suitable cutoff values reported in previous studies (Atilgan *et al.*, 2001; Eyal *et al.*, 2006; Kurkcuoglu *et al.*, 2009). The enzymes are triose phosphate isomerase (TIM) (PDB ID: 1CI1), elastase (PDB ID: 2FOE), and gamma-chymotrypsin (PDB ID: 1GMD). In the case of LIG_ANM, all modes are calculated covering the whole spectrum from the slowest (lowest frequency), global modes to the high-frequency localized, fast modes. In contrast, the MCG_ANM version of the algorithm provides only a limited number of global modes due to use of BLZPACK (Grimes *et al.*, 1994) algorithm for Hessian matrix inversion.

4.1.1. Triose phosphate isomerase

Triose phosphate isomerase (TIM) is a homo-dimeric enzyme. The specific crystal structure of the enzyme from *Trypanosome cruzi* (PDB ID: 1CI1) contains a total of 497 residues. This crystal structure can be viewed in Figure 4.1 with three hexane molecules that are numbered 261, 262 and 263. Two of them, HEX-262 and HEX-263, are located at the interface between chains A (grey) and B (wheat), whereas HEX-261 is bound to an

outer region of chain B. HEX-262 and HEX-263 that are mutually in contact are explicitly used in ANM calculations discussed next. This specific interface region named as the tunnel is known to have functional significance in inhibition (Tellez-Valencia *et al.*, 2004) of TIM's catalytic activity. The catalytic residues (K13, H95, and E167) of TIM, shown in magenta color on the surface representation, are distant from the solvent binding pockets.

a)

b)

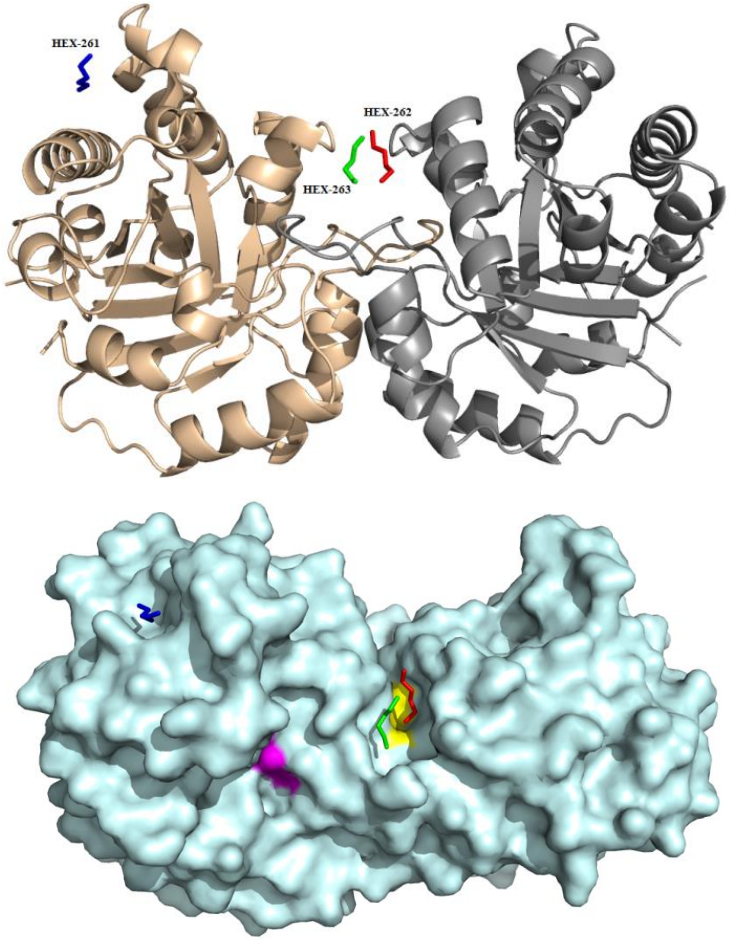


Figure 4.1. Triose phosphate isomerase from *Trypanosome cruzi* (a) cartoon and (b) surface representations showing binding pockets for hexane molecules and catalytic residues on one monomer.

A total number of 1485 modes (excluding the first six modes with zero eigenvalues related to translation and rotation) are calculated by LIG_ANM for the protein only, i.e. after removal of all ligands. In Figure 4.2, the percentage change in the frequency of each mode due to ligand binding is plotted as the difference between frequencies with and without ligand divided by the frequency without ligand for the same mode index. The

percentage change indicates pronounced frequency shifts to higher values in the slowest modes, i.e. at the left-hand side of the spectrum. In contrast, the frequencies on the right-hand side of the spectrum are lowered in the presence of ligands. Of course total number of modes calculated in the presence of ligands exceeds the unbound case by 36 due to the presence of two hexane molecules that are represented by 12 nodes. And in the construction of Figure 4.2, the last 36 modes for the ligand bound case are not considered in order to have matching number of modes. In Figure 4.3a and b percentage changes in the frequencies of first 100 modes are shown for LIG_ANM and MCG_ANM, respectively. In both cases, pronounced frequency shifts are observed in the first 8-10 modes indicating that the specific binding site at the interface affects the collective modes significantly.

Figure 4.2. Percentage change in frequencies of TIM due to hexanes at interface from LIG_ANM showing all modes.

Of course it is not guaranteed that matching mode index corresponds to the same mode in the presence and absence of ligands, because there might be swapping between modes or changes in mode character due to binding.

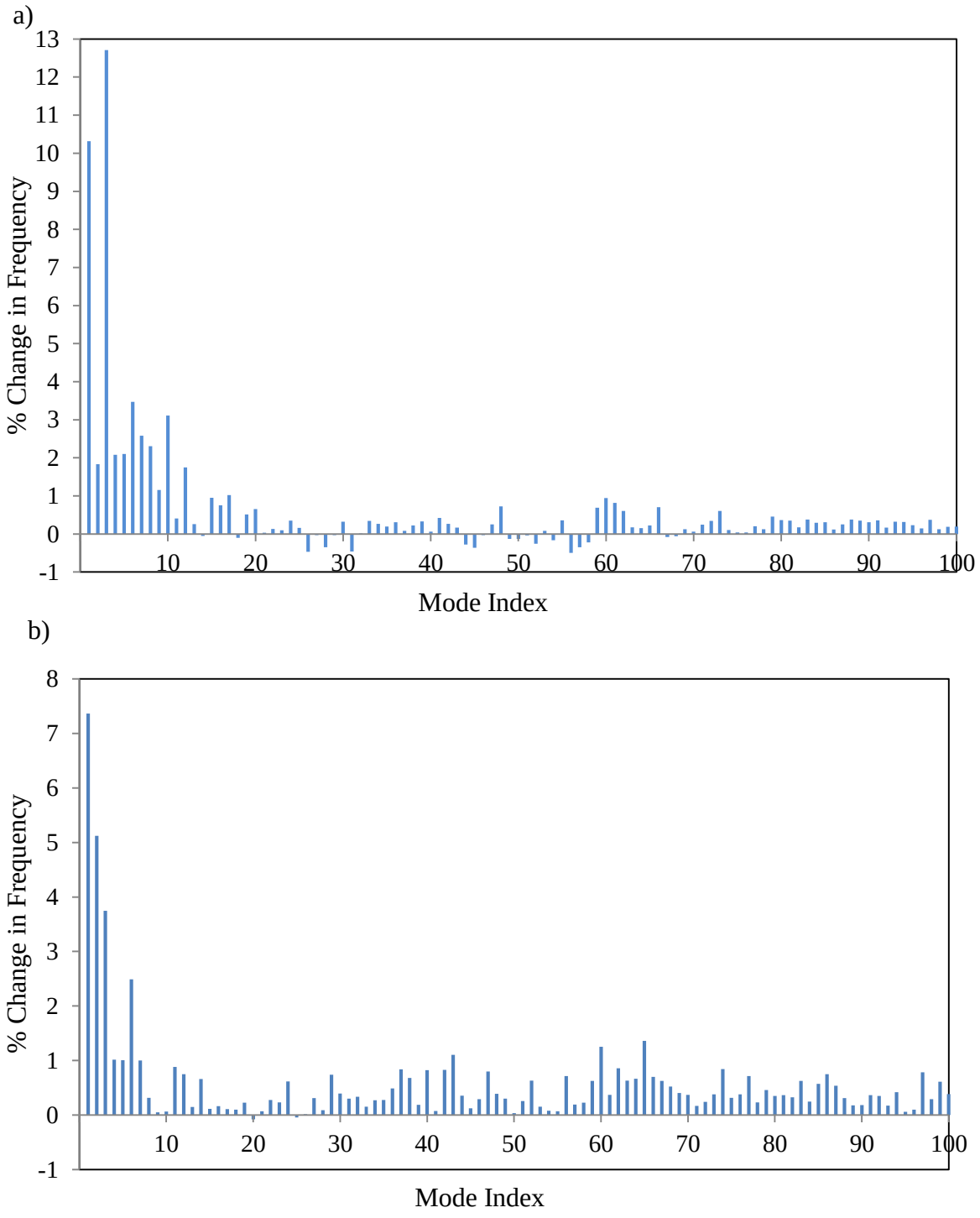


Figure 4.3. Percentage change in frequencies of TIM due to hexanes at interface from (a) LIG_ANM (b) MCG_ANM.

This is checked by calculating the overlap matrix, which gives the inner dot product between the eigenvectors from the different subspaces, i.e. with and without ligand. This matrix is shown in Figure 4.4 (a) and (b) for the first 100 modes of LIG_ANM and

MCG_ANM. MCG_ANM matrix shows more overlapping values than LIG_ANM matrix. In fact, the collective modes overlap to a large extent for bound and unbound states with minimal mode swapping and changes in character. Moreover, the fact that collective modes are also the functional ones (Tama and Sanejouand, 2001) leads to the consideration of ligand effect only in the low-frequency end of the spectrum in the following sections of this thesis.

Finally, the effect of each hexane molecule is considered separately using the LIG_ANM and MCG_ANM in Figure 4.5 (a) and (b), respectively. The effect of HEX-262 and HEX-263 at the interface seem to be predominant in MCG_ANM approach compared to HEX-261, which is bound to an outer pocket with no known effect on activity. Using the CASTp server (Dundas *et al.*, 2006), area and volume of the pockets on the surface of proteins can be calculated. The pockets, where the hexane molecules are bound to, are recorded using CASTp server and the residues forming each pocket are colored yellow in Figure 4.1b. HEX-262 and HEX-263 bind to the fourth biggest pocket with area 214.4$Å^2$ and volume 192.8$Å^3$. In contrast, HEX-261 binds to a smaller pocket with area 71.1$Å^2$ and volume 48.2 $Å^3$ with no reported functional significance. Consequently, ligand binding to this specific site is not expected to significantly affect the collective modes. This is observed in MCG_ANM results, whereas the opposite situation exists in LIG_ANM, possibly due to the bulk of ligand nodes. Thus, MCG_ANM may be more appropriate for studying protein-ligand interactions, at least for the current case of TIM. Finally, the catalytic residues of TIM (K13, H95, and E167) are also shown in Figure 4.1b with magenta. The catalytic sites' pocket sizes are reported as 198$Å^2$ and 238.2 $Å^3$ on chain A and as 215.4 $Å^2$ and 249.1$Å^3$ on chain B. The effect of ligand binding at this site will be further analyzed in Section 4.2.

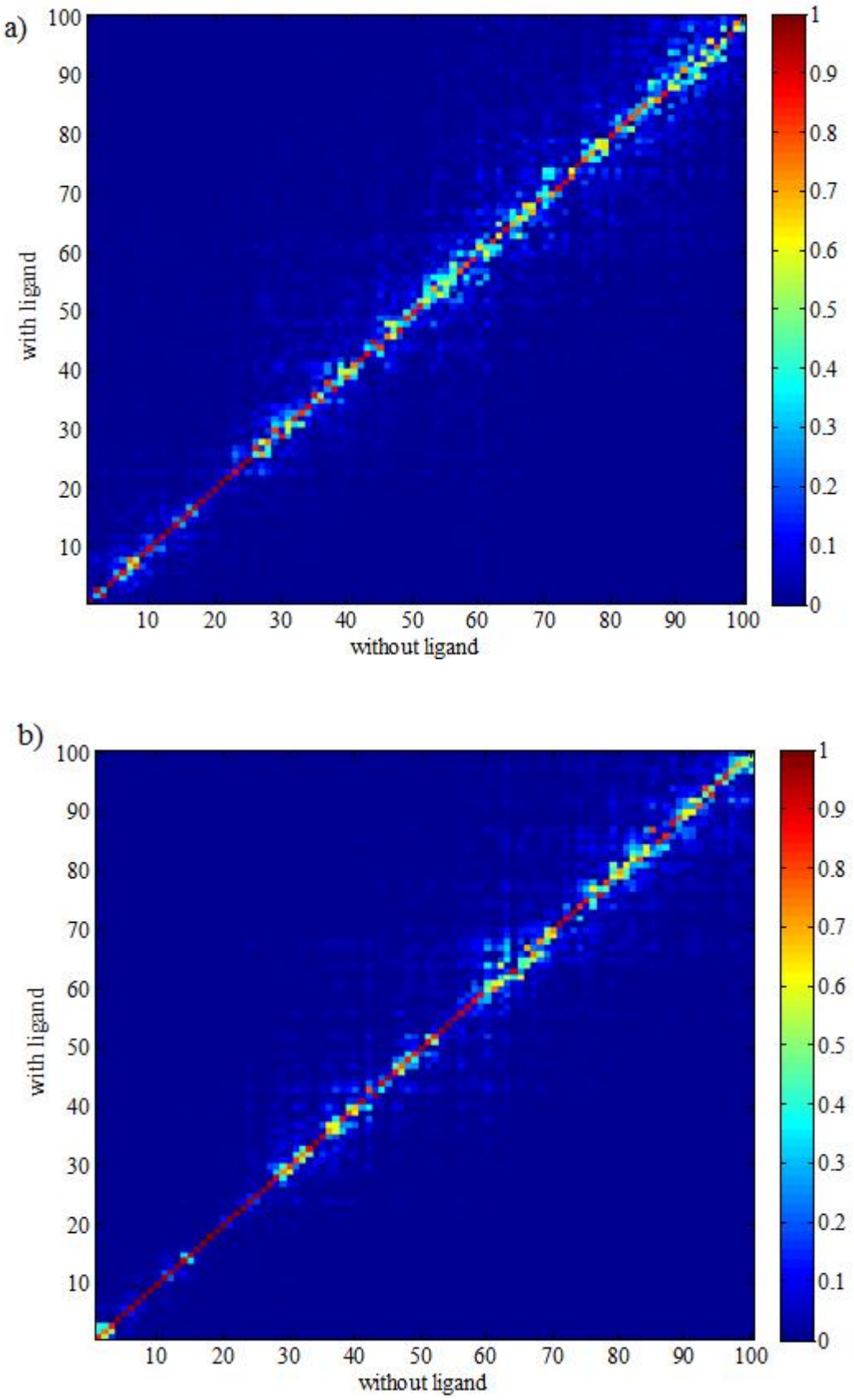


Figure 4.4. Overlap between modes calculated for TIM in the presence and absence of hexanes at interface with (a) LIG_ANM (b) MCG_ANM.

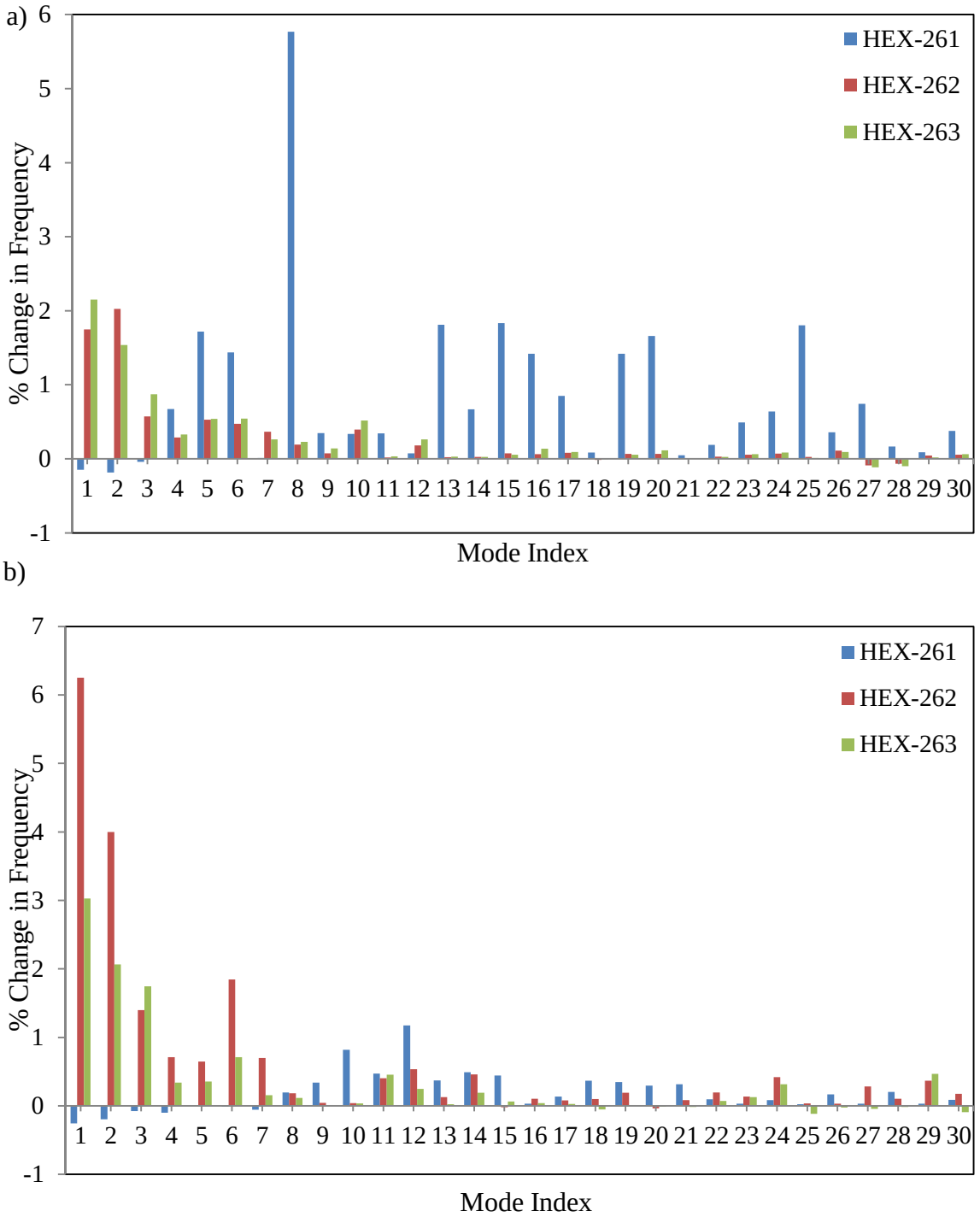


Figure 4.5. Percentage change in frequencies of TIM due to bound hexane molecules based on (a) LIG_ANM (b) MCG_ANM.

4.1.2. Elastase

Elastase is a monomeric enzyme, shown in Figure 4.6. The specific crystal structure (PDB ID: 2FOE) consists of 240 residues and two bound hex-5-ene-1, 2 diol ligands. One of bound solvents, HEX-1001 (blue), is located in the active site and in contact with catalytic residues H60, D108, G201 and S203 (Mattos *et al.*, 2006), which are shown in green in Figure 4.6b. The other one, HEX-1004 (red), is bound to a relatively outer binding pocket of the chain. In the following calculations, HEX-1001 and 1004 are both present. Therefore, it is expected that the collective modes should be affected due to the presence of at least HEX-1001. In Figure 4.7, the percentage change in frequencies based on LIG_ANM indicates pronounced solvent effect in low-frequency end of the spectrum, which is consistent with TIM results. Figure 4.8 based on slowest 100 modes, indicates higher changes for LIG_ANM than MCG_ANM.

The overlap matrix between the modes in the presence and absence of ligands presents a more scattered picture than TIM for LIG_ANM. Still there is some correspondence among the first 10 modes. Comparatively MCG_ANM overlap matrix exhibits a less scattered behavior over the whole range of modes.

In Figure 4.10, the effects of HEX-1001 and 1004 are presented separately. HEX-1001 binding pocket has 80.6$Å^2$ surface area with volume 74.9 $Å^3$, whereas HEX-1004 binding pocket has 82.2 $Å^2$ surface area with volume 111.4 $Å^3$ (Dundas *et al.*, 2006). Thus, the two pockets have similar sizes. In MCG_ANM, both molecules seem to cause similar changes in frequencies, whereas the effect of HEX-1004 is much more pronounced in LIG_ANM. Again, MCG_ANM seems appropriate for studying protein-ligand interactions as discussed for TIM.

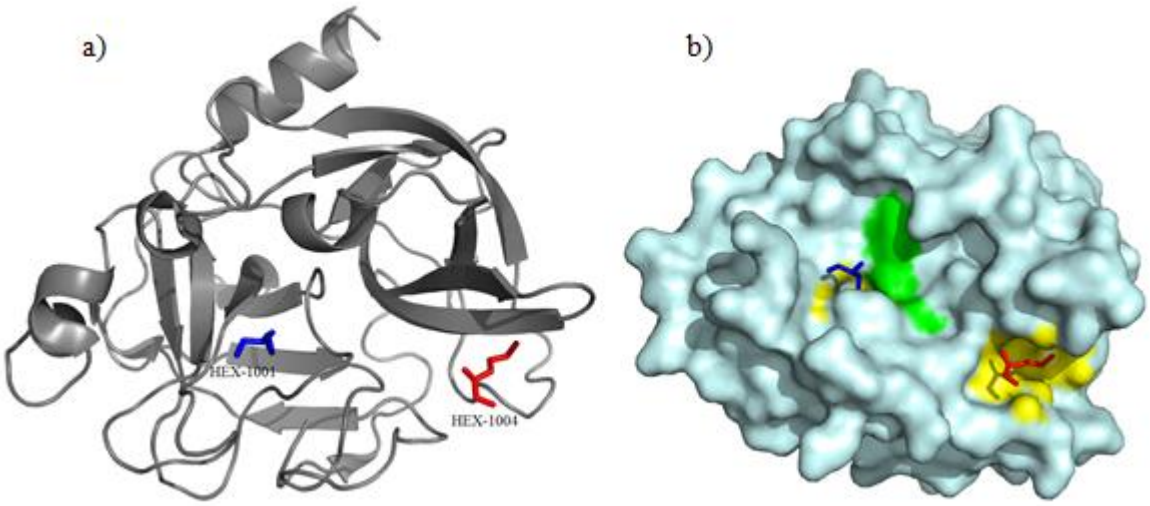


Figure 4.6. Structure of porcine pancreatic elastase in 80% hexane (a) cartoon demonstration (b) surface demonstration with pockets.

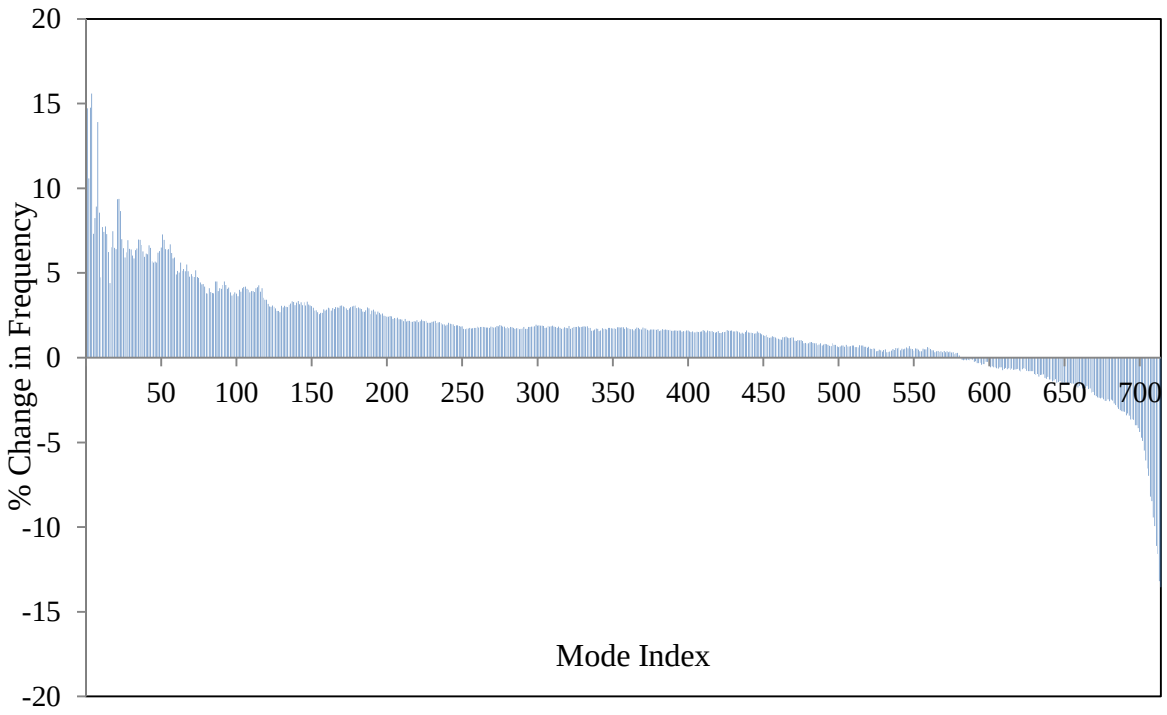


Figure 4.7. Percentage change in frequencies of elastase due to hexanes from LIG_ANM showing all modes.

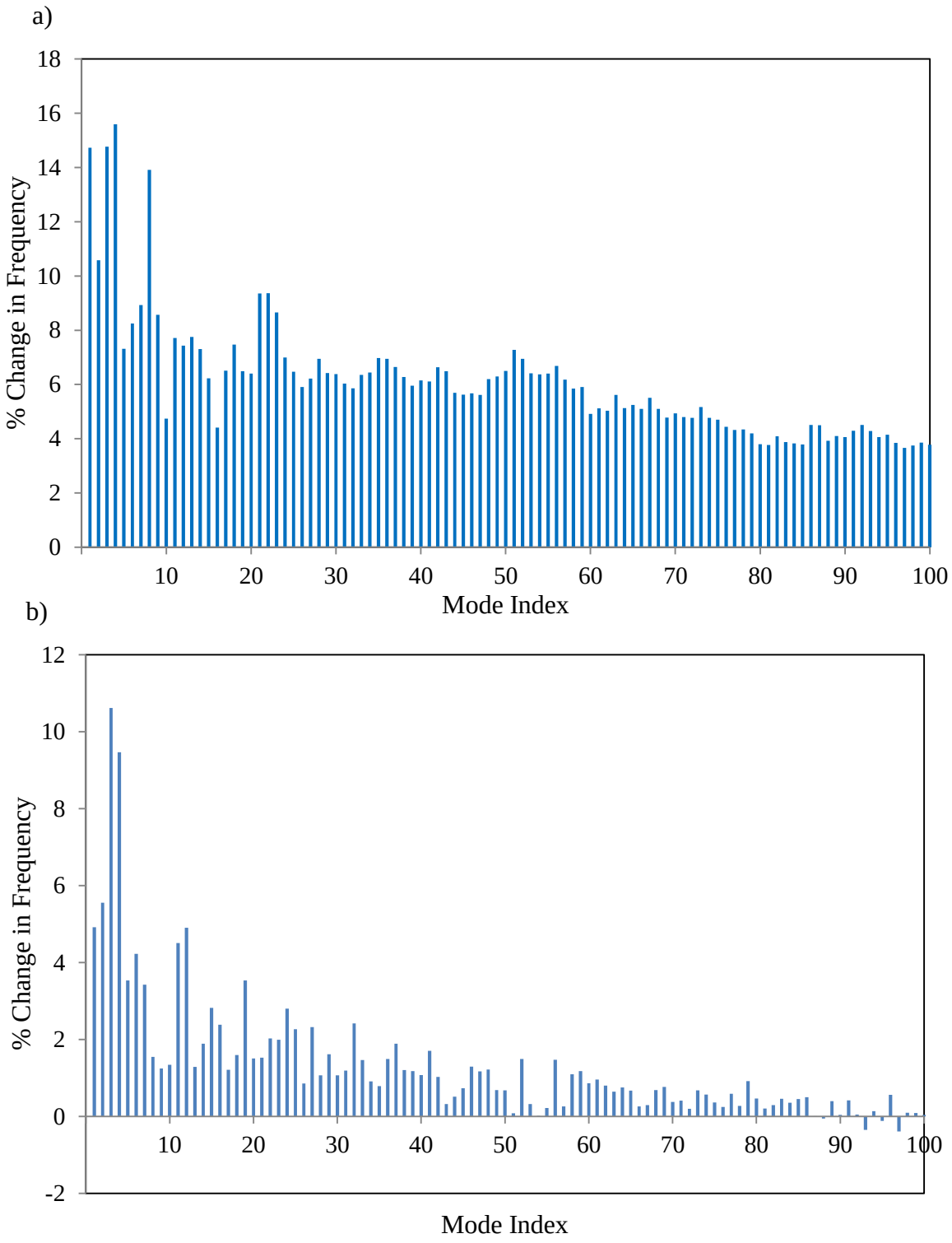


Figure 4.8. Percentage change in frequencies of elastase due to hexanes from (a) LIG_ANM (b) MCG_ANM.

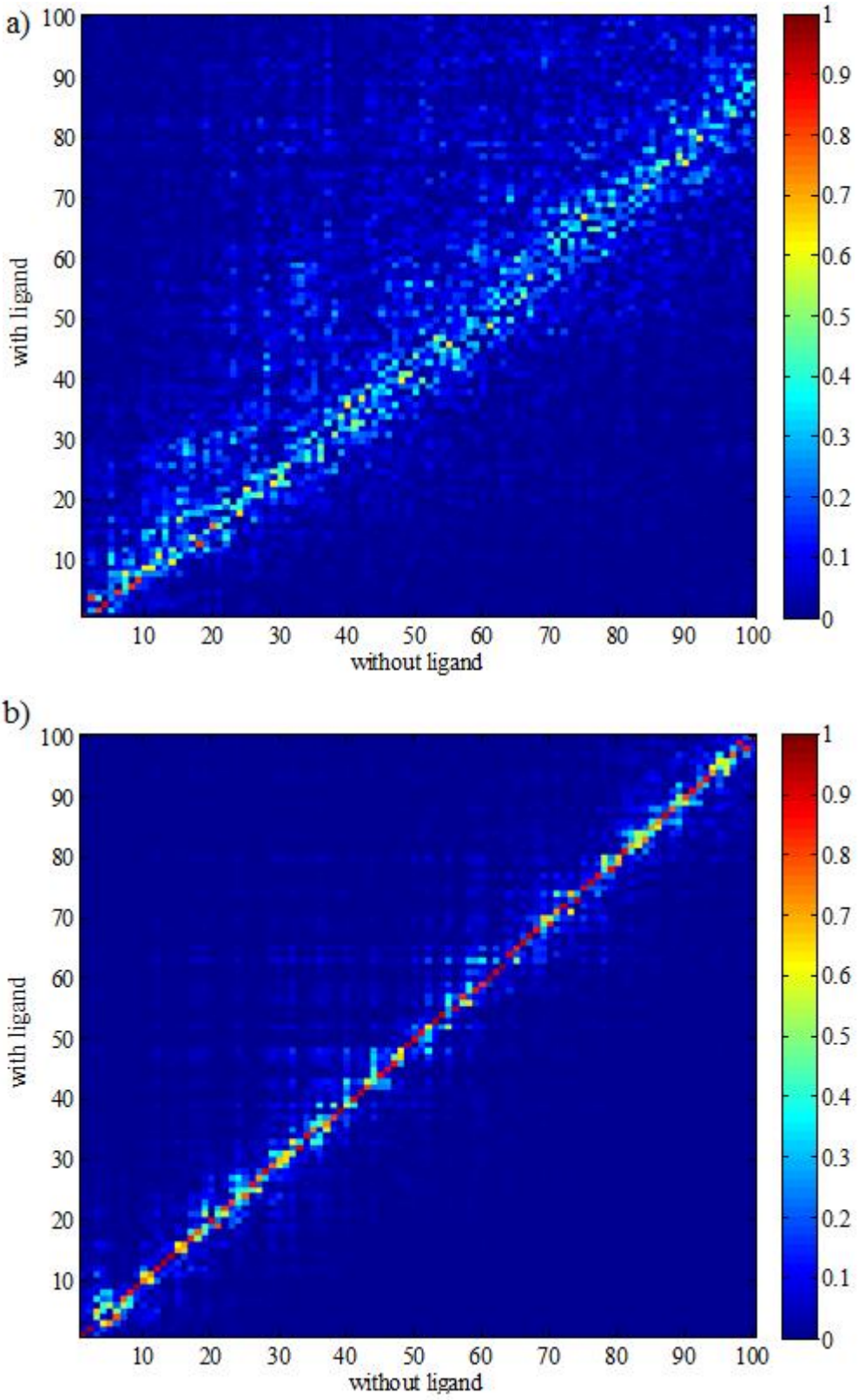


Figure 4.9. Overlap between modes calculated for elastase in the presence and absence of hexanes with (a) LIG_ANM (b) MCG_ANM.

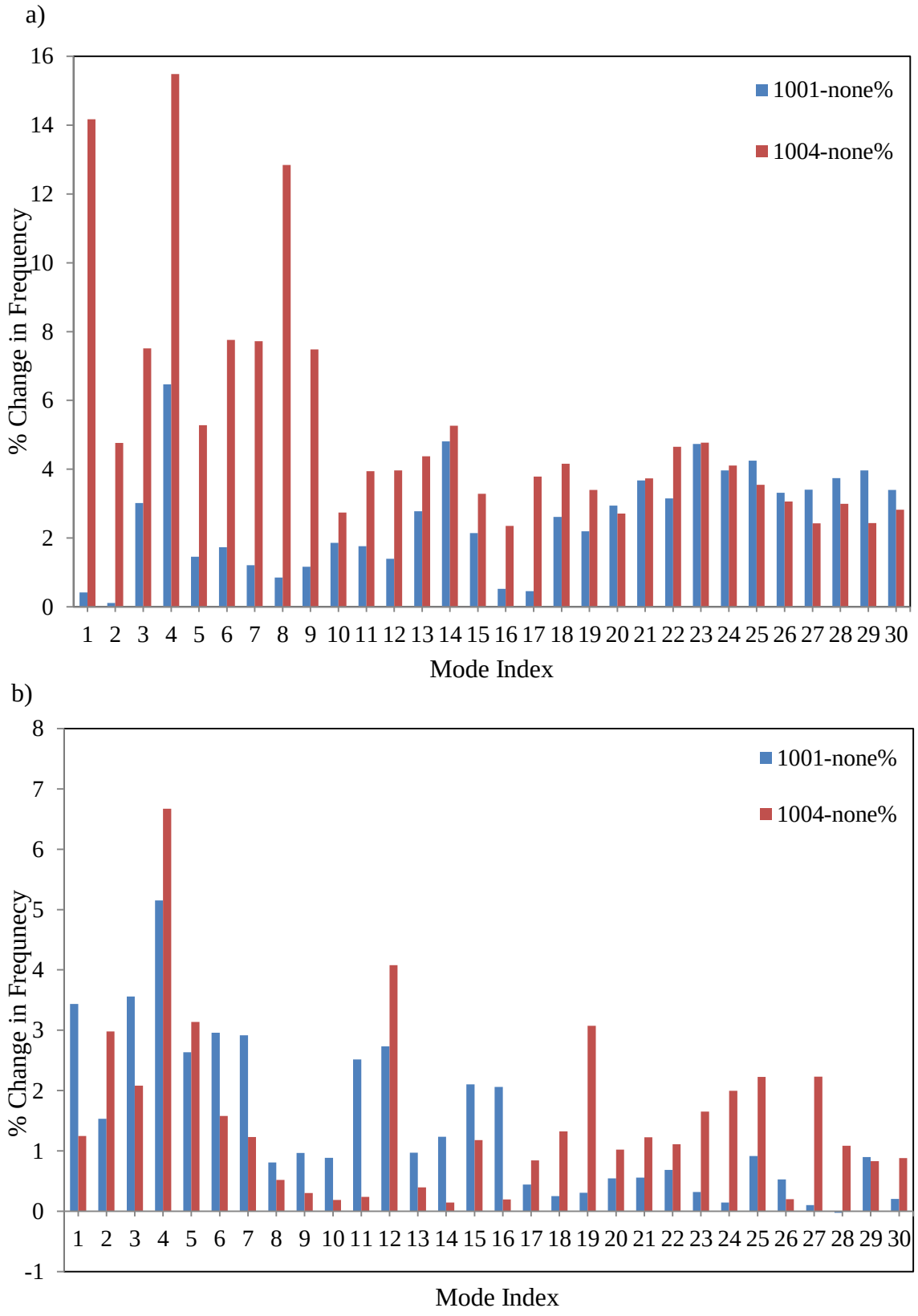


Figure 4.10. Percentage change in frequencies of elastase due to bound hexane molecules based on (a) LIG_ANM (b) MCG_ANM.

4.1.3. Gamma-Chymotrypsin

Gamma-chymotrypsin is hetero-tetrameric enzyme with its different chains shown in pink, yellow, wheat and gray in Figure 4.11 (a). The specific crystal structure (PDB ID: 1GMD) contains 242 residues and seven hexane molecules. HEX-403 and HEX-406 are located near the active site and the rest are on other parts of the surface. The active site residues are S195, H57, D102, colored green in Figure 4.11b (Yennawar *et al.*, 1994). Initially, LIG_ANM and MCG_ANM are performed with the two hexanes (403 and 406), which are close or in contact with the active site.

In Figure 4.12, frequency shifts for LIG_ANM follow a similar trend with the other two enzymes, but the percentage changes in collective modes are much lower. Also, more high frequency modes indicate shifts to the left. These differences may result from the fact that gamma-chymotrypsin is a small protein made up of four chains with 5, 11, 95 and 131 residues. Especially due the presence of two very short chains, ligands may affect global motions and local flexibility of the structure in different way. In Figure 4.13a and b, negative values are observed both in LIG_ANM and MCG_ANM, which means binding of two ligands, decreases the frequency of some collective modes. Thus, the ligands do not act as constraints that increase the vibrational frequencies, rather they move in coordination with parts of the system, thereby decreasing the frequencies.

The overlap matrix in presence and absence of ligands presents a high correlation between the mode subspaces, especially for MCG_ANM matrix. The effects of hexanes are considered separately in Figures 4.15 and 4.16 using LIG_ANM and MCG_ANM, respectively. HEX-406 that is contact with the active site residues seems more effective than other hexanes in terms of changing frequencies and acting as a constraint. The closest binding pocket to HEX-406 is shown in yellow in Figure 11b and has 78.6 $Å^2$ area and 63.8 $Å^3$ volume. HEX-405 also causes some change although it is not close to the active site. HEX-405 binding pocket is in fact reported as the biggest pocket in gamma-chymotrypsin with 317.2 $Å^2$ area and 236.4 $Å^3$ volume. Frequency changes due to HEX-403, 404 and 407 are negative, as can be seen in Figure 4.16b and c. It is observed from surface presentation of the gamma-chymotrypsin that these three hexanes do not reside in a pocket, but rather lie on the surface. HEX-401 is bound to a small pocket and HEX-402 is

rather buried inside. These two hexanes also cause marginal positive changes in frequencies.

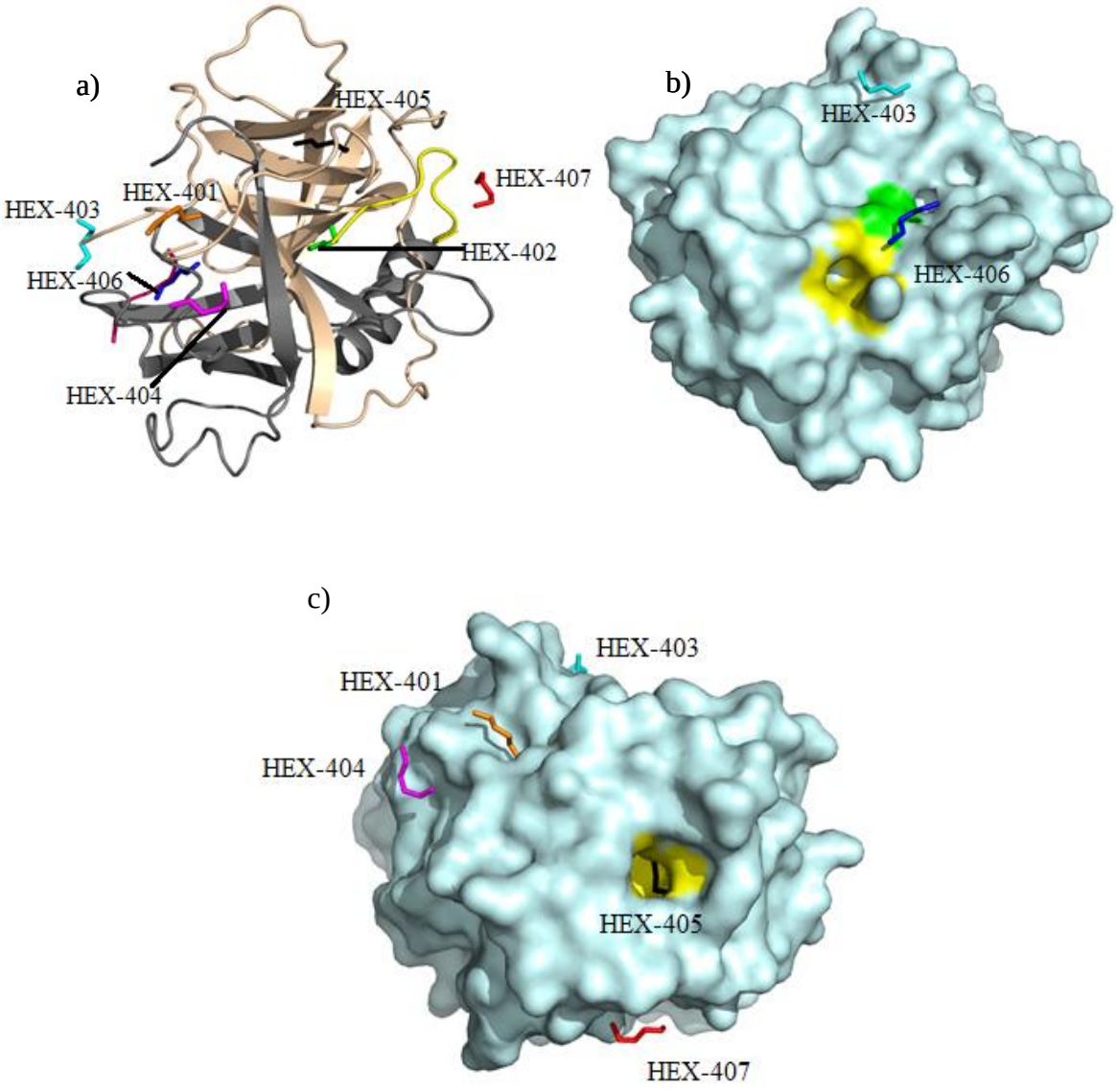


Figure 4.11. Gamma-chymotrypsin (a) cartoon demonstration, surface representation (b) showing binding pockets for hexane molecules (c) showing other hexane molecules.

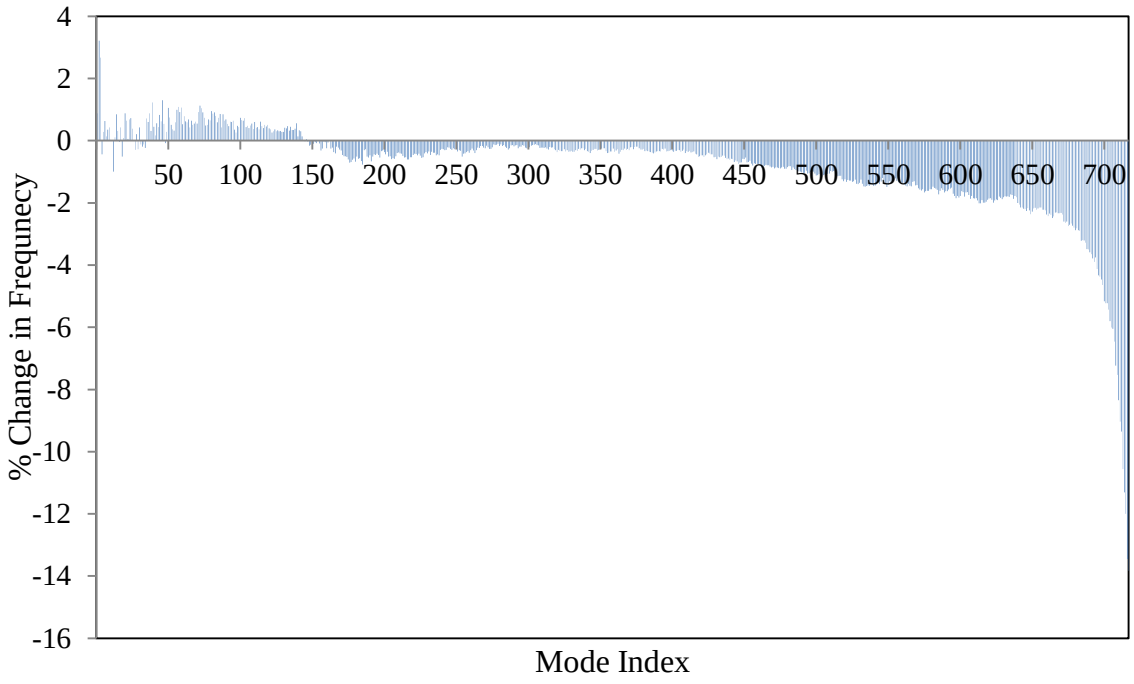


Figure 4.12. Percentage change in frequencies of gamma-chymotrypsin due to hexanes at active site from LIG_ANM showing all modes.

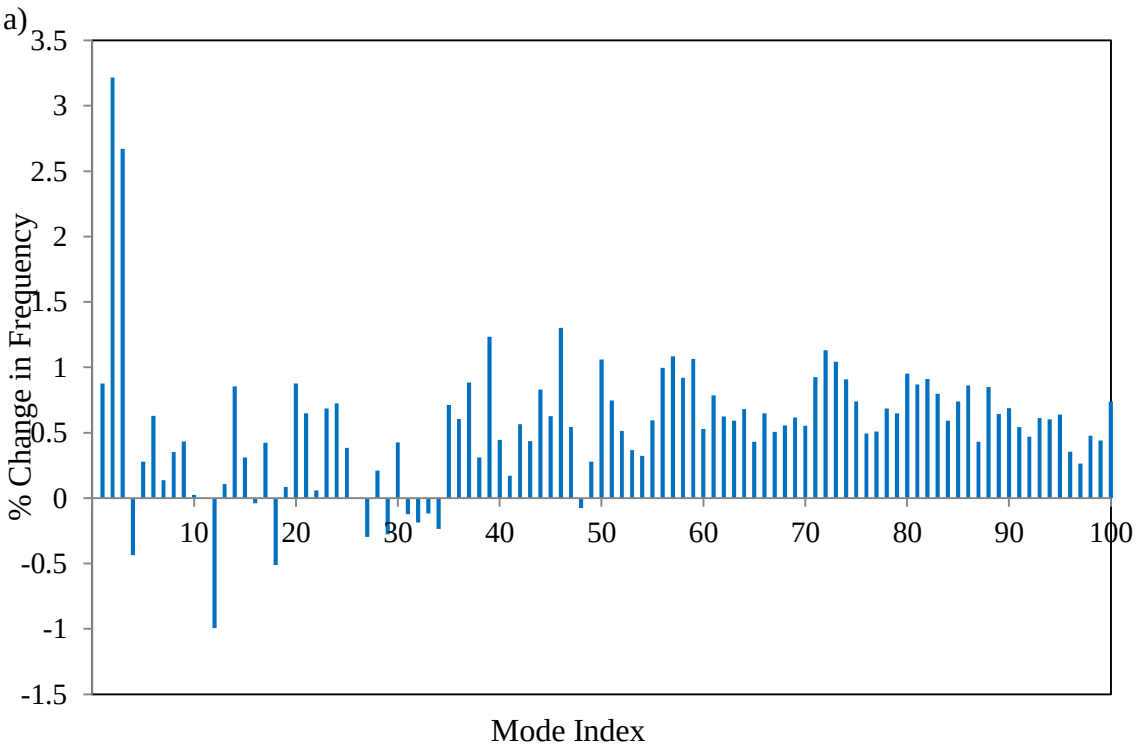


Figure 4.13. Percentage change in frequencies of gamma-chymotrypsin due to hexanes at active site from (a) LIG_ANM (b) MCG_ANM.

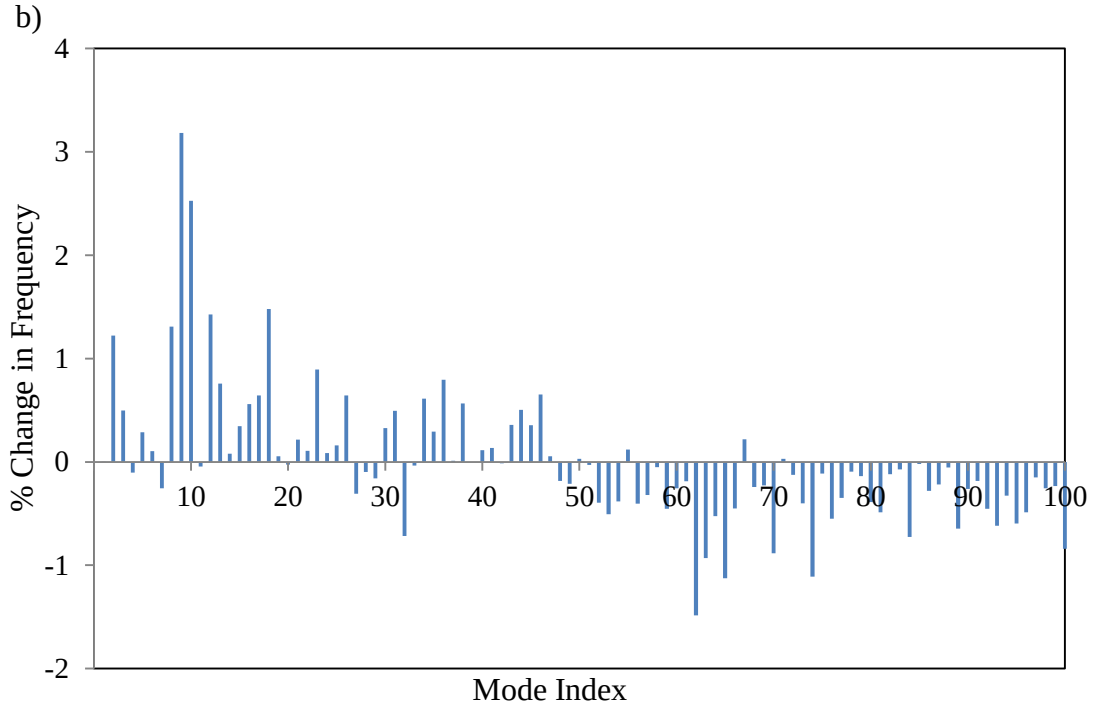


Figure 4.13. Percentage change in frequencies of gamma-chymotrypsin due to hexanes at active site from (a) LIG_ANM (b) MCG_ANM cont.

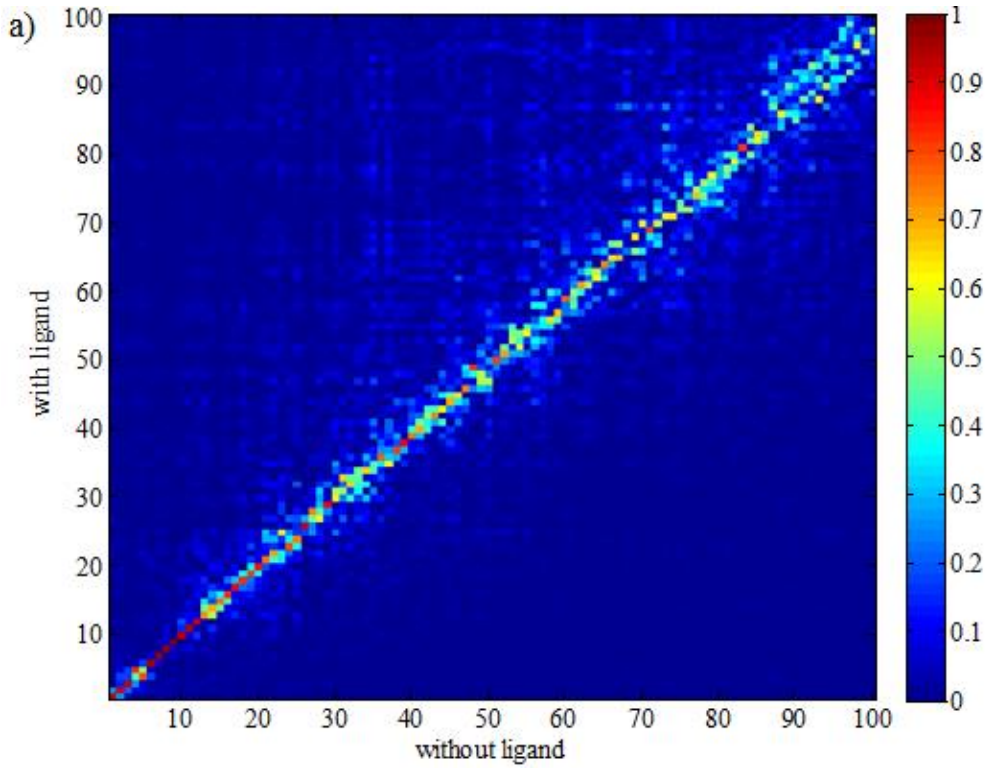


Figure 4.14. Overlap between modes calculated for gamma-chymotrypsin in the presence and absence of hexanes with (a) LIG_ANM (b) MCG_ANM.

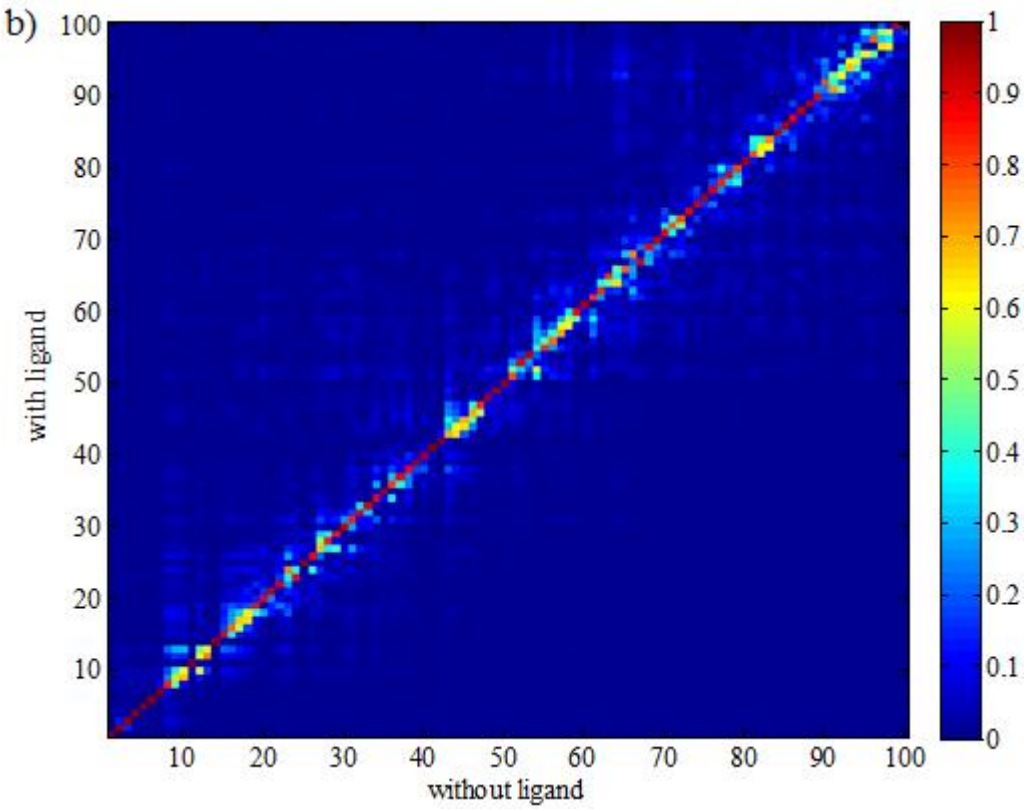


Figure 4.14. Overlap between modes calculated for gamma-chymotrypsin in the presence and absence of hexanes with (a) LIG_ANM (b) MCG_ANM cont.

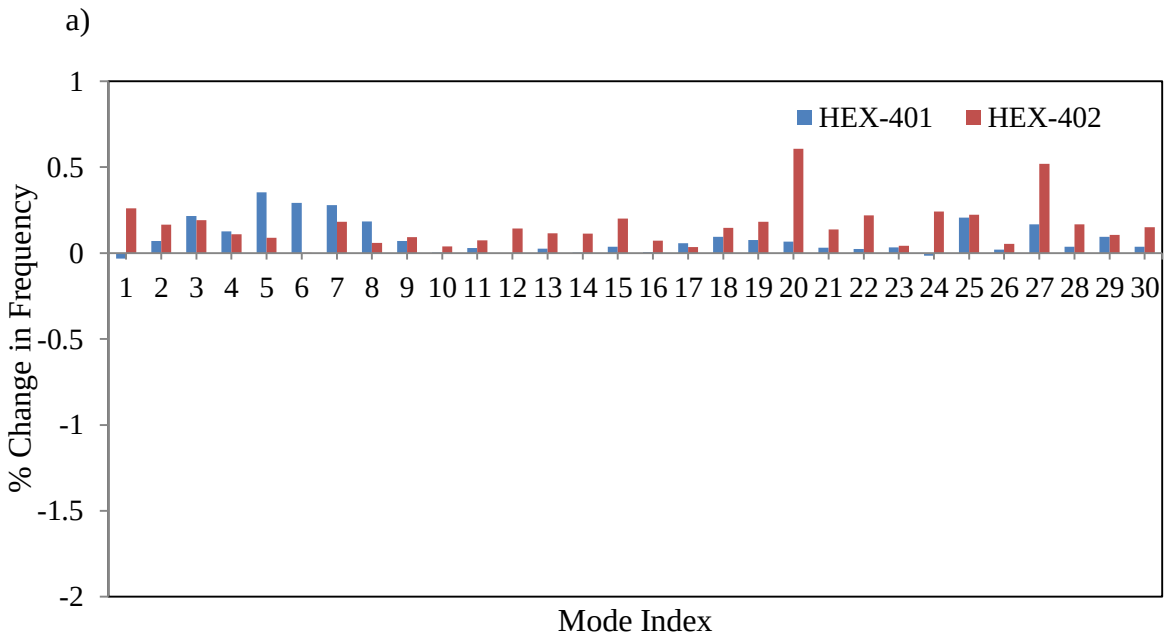


Figure 4.15. Percentage change in frequencies of gamma-chymotrypsin due to bound hexane molecules based on LIG_ANM for all hexanes.

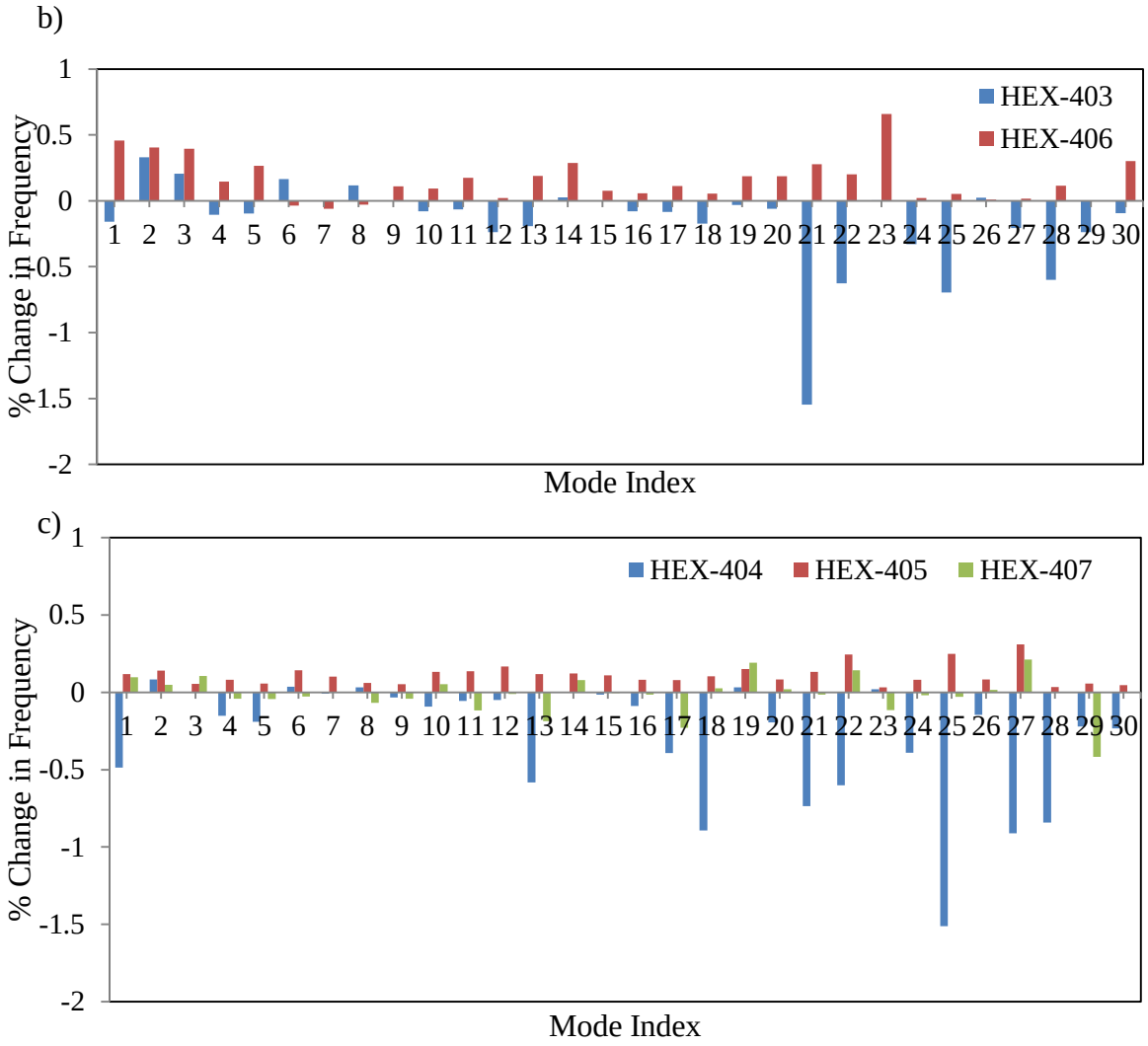


Figure 4.15. Percentage change in frequencies of gamma-chymotrypsin due to bound hexane molecules based on LIG_ANM for all hexanes cont.

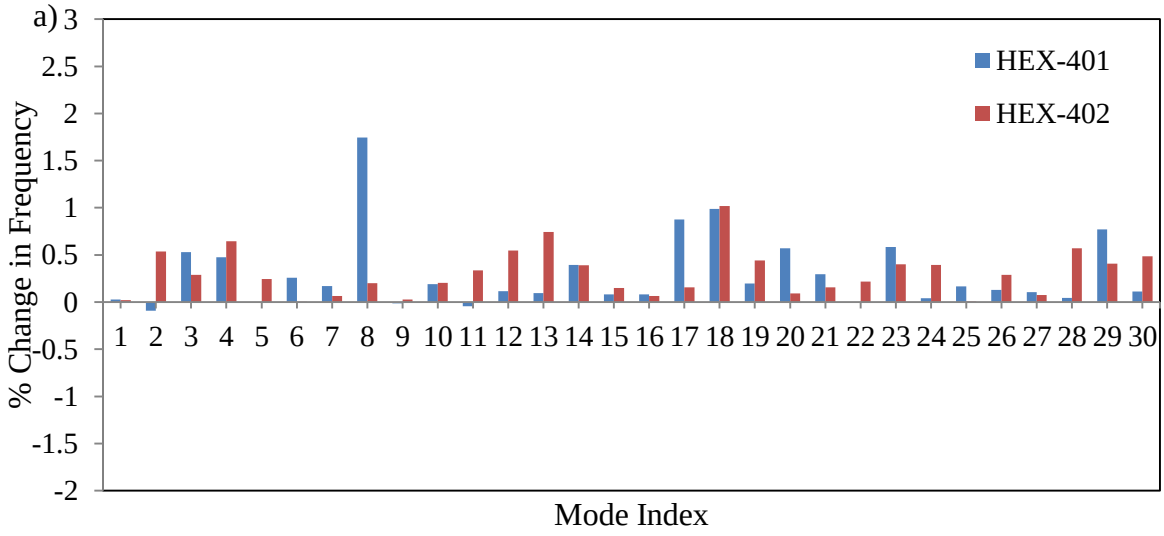


Figure 4.16. Percentage change in frequencies of gamma-chymotrypsin bound hexane molecules based on MCG_ANM for all hexanes.

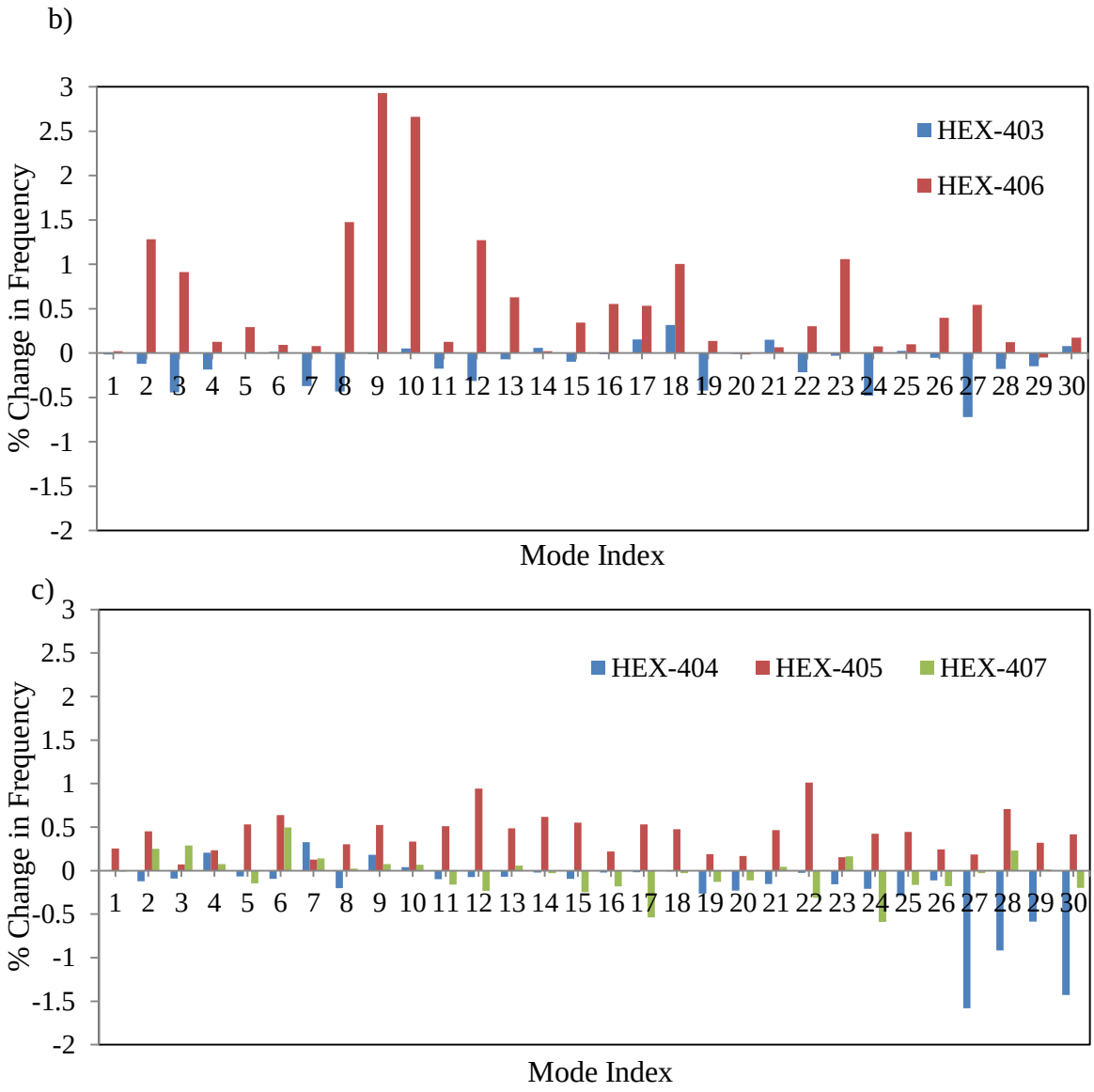


Figure 4.16. Percentage change in frequencies of gamma-chymotrypsin bound hexane molecules based on MCG_ANM for all hexanes cont.

To summarize results for the three enzymes, Figure 4.17 shows the average frequency changes caused by each solvent considering different mode ranges from MCG_ANM. Concentrating on the first five modes, HEX-1001 and HEX-1004 binding to elastase has higher effect than all other cases. HEX-1001 binds to the active site and HEX-1004 may indicate a potential binding site. HEX-262 and 263 are effective in TIM that binds to the interface region, where some reported inhibitors bind (Tellez-Valencia *et al.*, 2004). In gamma-chymotrypsin, HEX-406 in contact with the active site leads to relatively high changes but not as much as other two enzymes. Moreover, positive values in gamma-chymotrypsin correspond to hexanes within pockets, whereas negative value to outer

positions. These findings indicate that the effect of solvent binding on frequencies can give insight about functional sites on the protein surface.

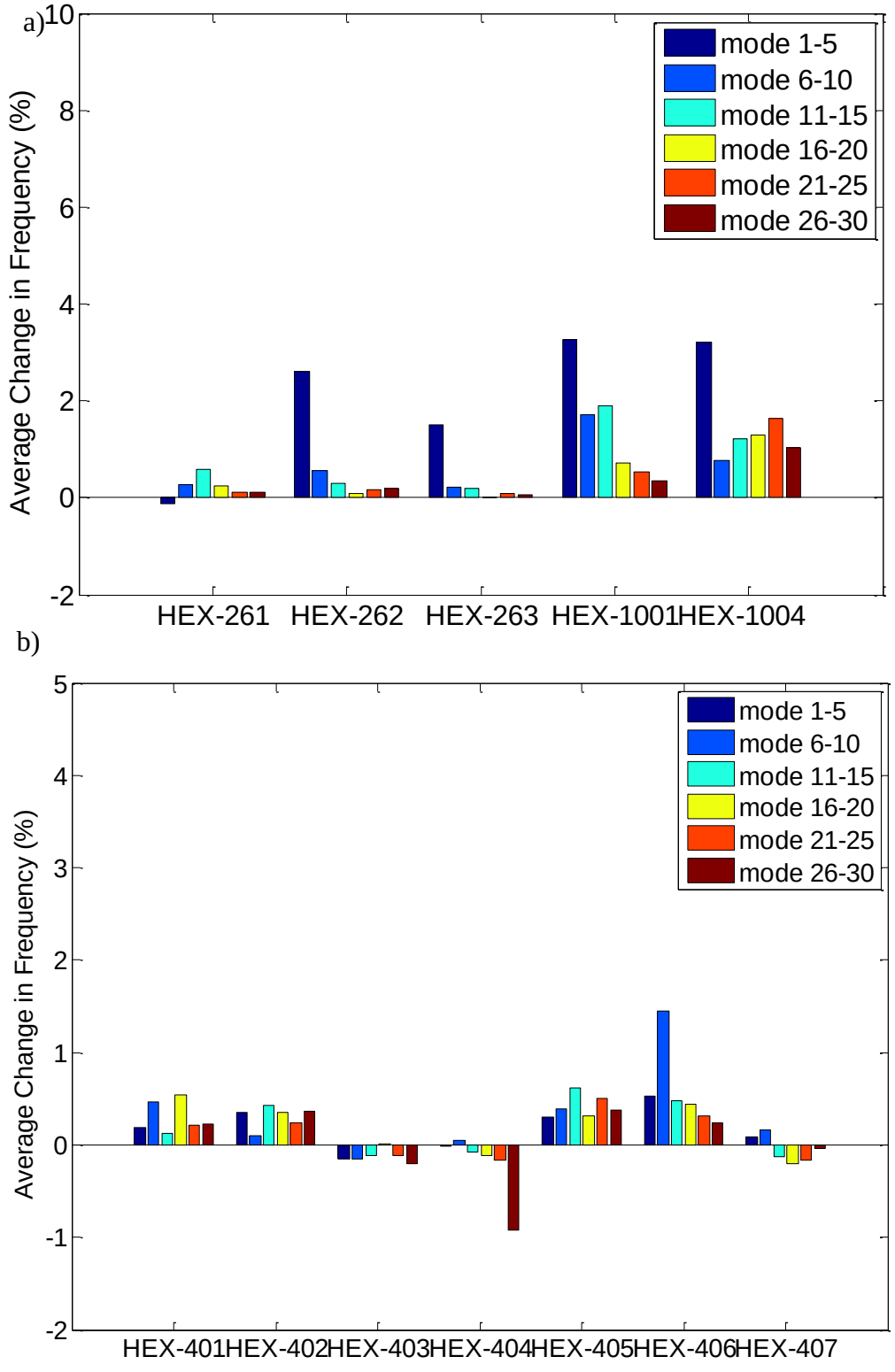


Figure 4.17. Average Frequency % distribution due to each ligand binding based on MCG_ANM (a) TIM and elastase (b) gamma-chymotrypsin.

4.2. MCG_ANM Applications to Enzymes with Catalytic Loops

The effect of ligand binding to enzymes with catalytic loops is investigated in this section using MCG_ANM. The ligand is an inhibitor and/or substrate analog that binds to the catalytic site of the enzyme. This is accompanied by a conformational change, where the catalytic loop closes over the ligand in order to protect the site from solvent during catalysis. A dataset of such enzymes have been formed previously (Kurkcuoglu *et al.*, 2012) for analyzing the correlation between collective modes and the experimental loop closure direction. Standard ANM calculations and MD simulations performed on apo structures (Kurkcuoglu *et al.*, 2012) have indicated several collective modes that can drive closure of the catalytic loop. In this thesis, two monomers, four dimers and one tetramer from the same dataset are considered for analysis of ligand binding effect.

For these enzymes, the apo structure with the loop in open position (denoted by 'O' in Table 4.1) and the ligand-bound complex with the loop in closed position (denoted by 'C') are both considered in MCG_ANM calculations. Figures 4.18 and 4.19 show the conformational change of the protein from open to closed states by vector representation, which are given on the left panels. The arrows correspond mainly to the loop region that presents the main conformational change in each case. In the case of multimeric enzymes, the subunits are identical so that there are either two or four loops closing at the same time. On the right panels of the same figures, surface representations of the enzymes are given with loops in closed position and colored yellow and the ligands shown as red sticks.

Comparison of the modes from open and closed states gives the change in frequencies due to the conformational change. Moreover, independent ANM calculations performed on the closed structure with ligand and after removing the ligand (named as 'without ligand') are compared in order to reveal the effect of ligand only.

Cutoff distances in the range of 6-9 Å have been reported suitable for fully atomistic ANM calculations (Kurkcuoglu *et al.*, 2006). In order to determine an appropriate cutoff for MCG_ANM calculations, three values ($R_{c,a}$ = 5, 7, 10 Å) are considered for the dataset of enzymes. For the case of $R_{c,a}$ = 5, more than six zero eigenvalues are found for all

enzymes, indicating that it is not a suitable value for forming an intact network. Therefore, the results for $R_{c,a}$ = 7 and 10 Å will be reported in this section and in Appendix A.

Table 4.1. Dataset of Catalytic Enzymes and Ligands.

Enzyme	PDB ID	Total # of Residues (C)	Ligand	% of ligand atoms per residue
Protein Tyr Phosphatase (PTP)	1ypt(O) 3blt(C)	278	$C_8H_8O_3S$	4.32
β 1,4-Galactosyltransferase (bGT)	1fgx(O) 1nkh(C)	272	$C_9H_{14}N_2O_{12}P_2$	8.99
Biphosphate Aldoase (BPA) (x2)	3c4u(O) 3c52(C)	582	$C_2H_6NO_6P$(x2)	3.44
Enolase (Enl) (x2)	3enl(O) 7enl(C)	872	$C_3H_7O_7P$(x2)	2.52
Triosephosphate Isomerase (TIM) (x2)	8tim(O) 1tph(C)	490	$C_2H_6NO_6P$(x2)	4.08
3-dehydroquinase (3dhq) (x2)	1gqn(O) 1l9w(C)	504	$C_7H_{10}O_4$(x2)	4.37
Lactate Dehydrogenase (Lldh) (x4)	3d0o(O) 3d4p(C)	1234	$C_3H_4O_3$(x4)	1.94

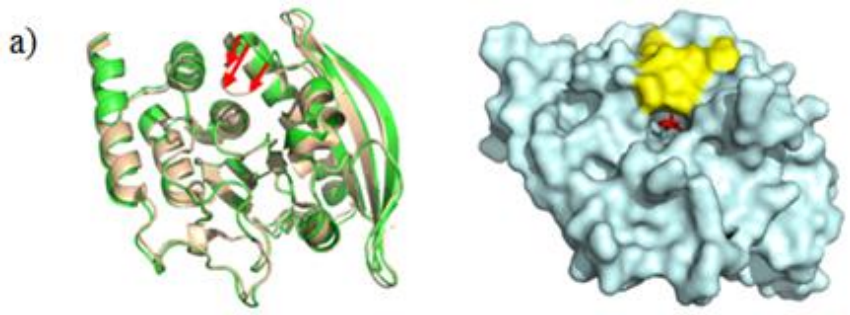


Figure 4.18. Cartoon and surface representations of (a) PTP (monomer) (b) bGT (monomer) (c) Lldh (tetramer).

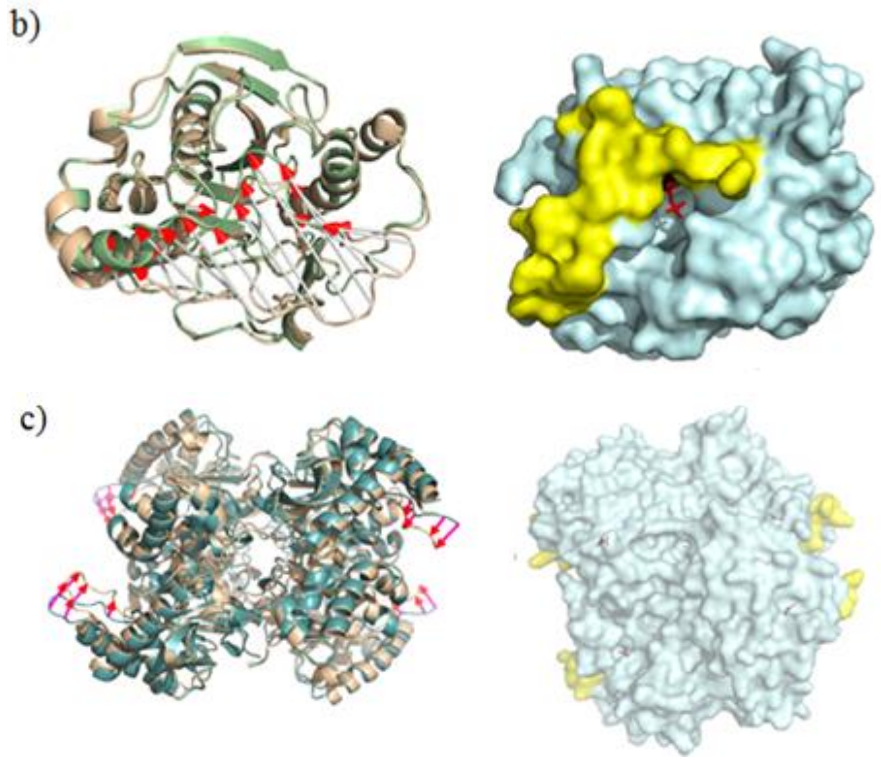


Figure 4.18. Cartoon and surface representations of (a) PTP (monomer) (b) bGT (monomer) (c) Lldh (tetramer) cont.

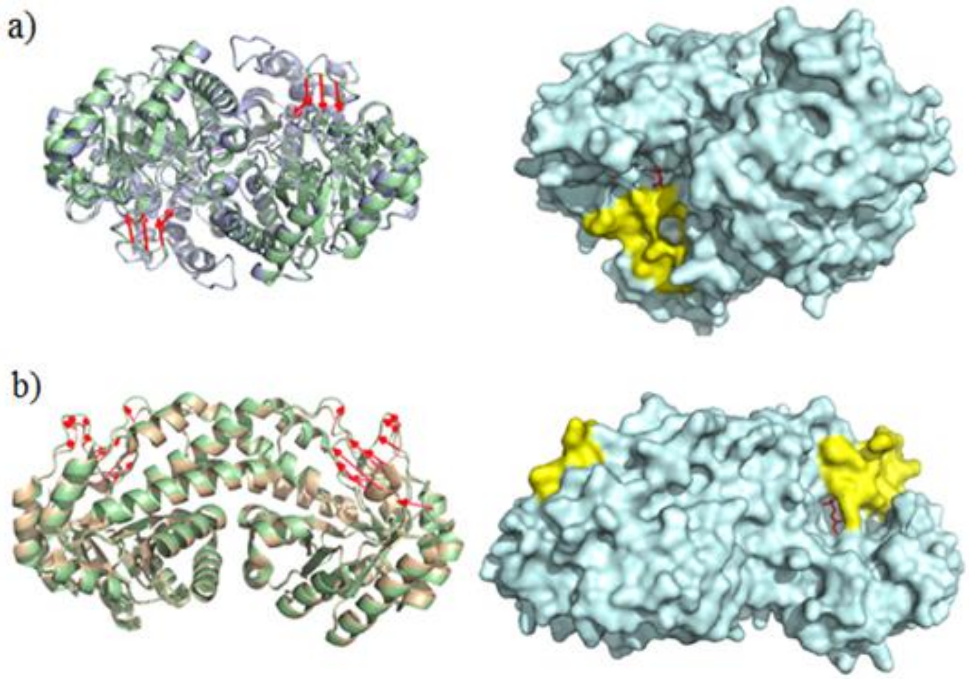


Figure 4.19. Cartoon and surface representations of (a) Enl (dimer) (b) BPA (dimer) (c) TIM (dimer) (d) 3dhq (dimer).

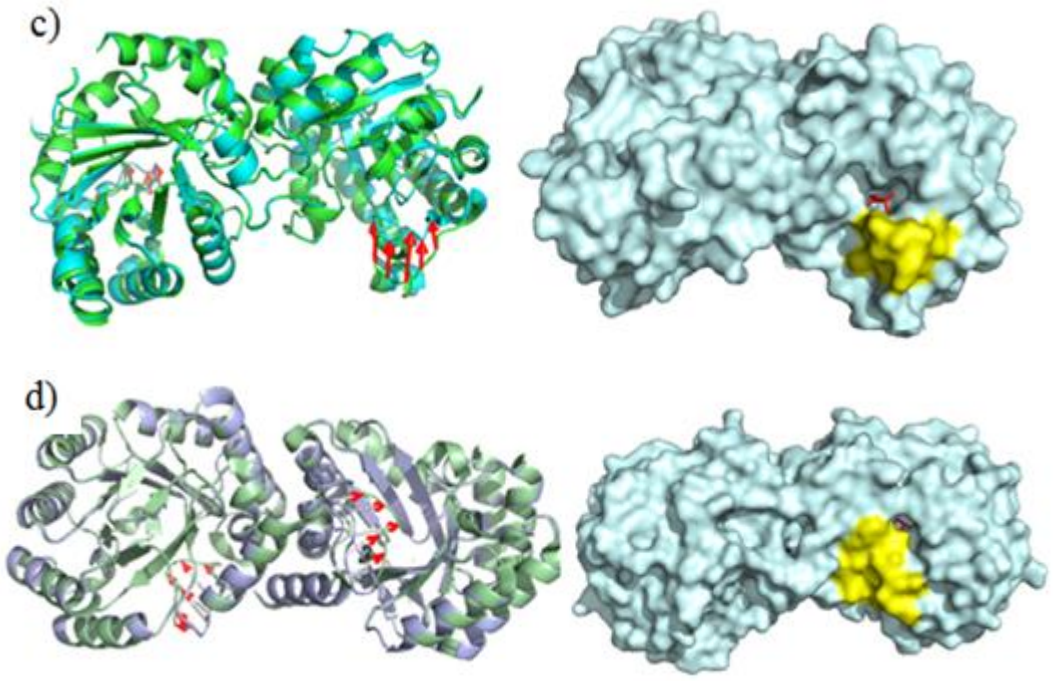


Figure 4.19. Cartoon and surface representations of (a) Enl (dimer) (b) BPA (dimer) (c) TIM (dimer) (d) 3dhq (dimer) cont.

Figures 4.20 and 4.21 give the overlap matrices between eigenvector subspaces for $R_{c,a}$ = 7 Å. The left panels compare modes from open and closed states, whereas the right panels give the correlation between modes of the closed state with and without ligand. Left panels indicate some overlap for the first 5-10 modes, but higher modes seem more dispersed. These significant changes in mode character are due to the large conformational changes due to loop closure. Specifically, modes for the closed state are calculated after removal of the ligand. In contrast, the right panels indicate an almost one-to-one correspondence between the modes observed in the presence and absence of ligand for the closed state. Thus, removal of the ligand from the active site, provided that the loop stays closed in its closed position, does not change the character of the modes. The overlap matrices for 10 Å cutoff provided in Appendix A also present a consistent picture.

Overlap values based on Equation 3.5 are listed in Table 4.2 using the first 30 modes for 7 Å cutoff. An analogous table for 10 Å cutoff is provided in Table A.1. Even though overlap values are quite high, there are quite low values reported for diagonal overlap. Low diagonal overlap values indicate that there is not a one-to-one overlap between the modes. The dispersion of modes is most dominant in bGT, which is consistent with the lowest

overlap values reported in the Table 4.2. At the same time, highest RMSD values between open and closed states are observed for bGT, both for the whole chain (3.4 Å) and the loop region (11.6 Å).

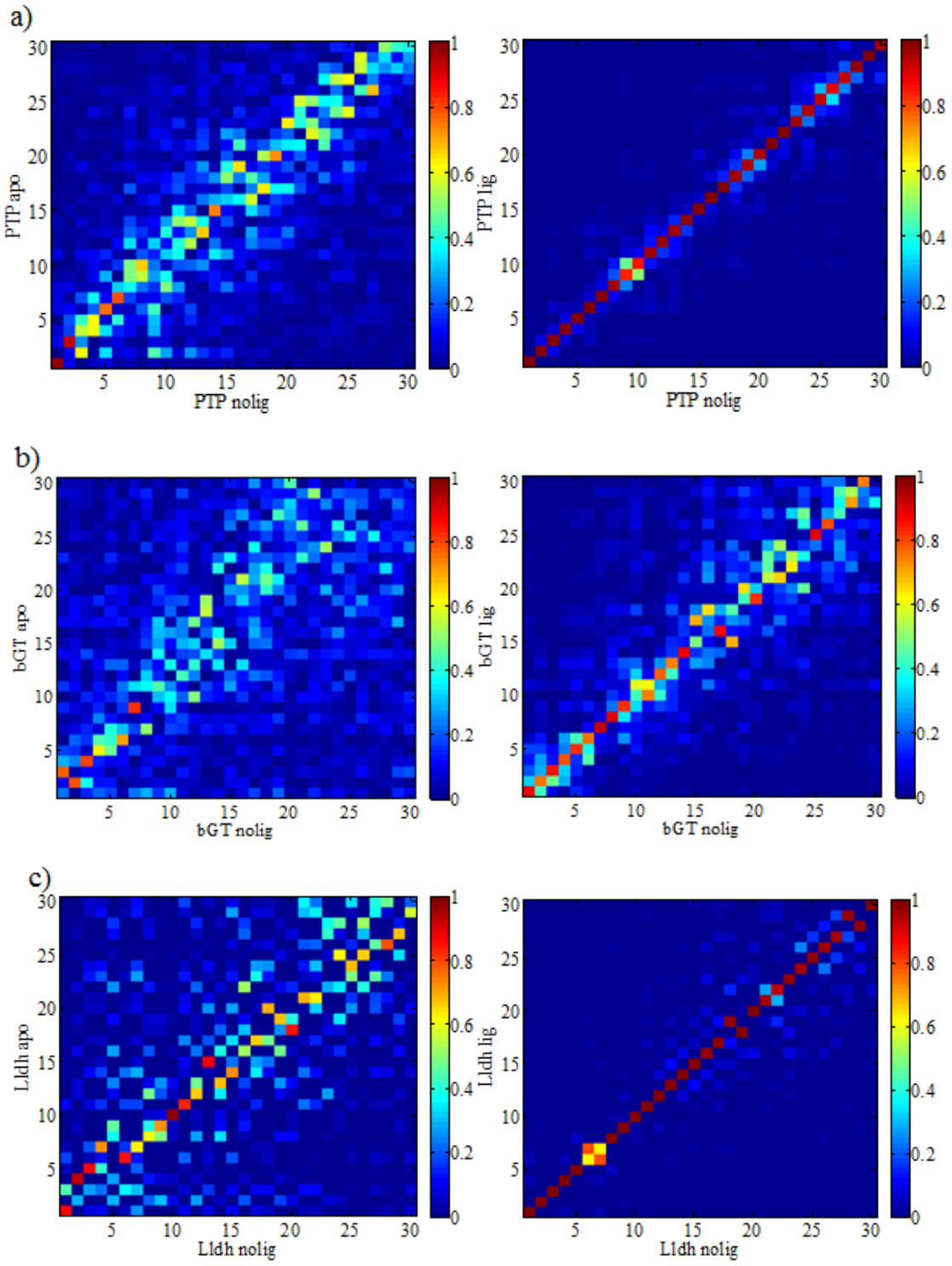


Figure 4.20. Overlap matrices for (a) PTP (monomer) (b) bGT (monomer) (c) Lldh (tetramer) with 7 Å cutoff.

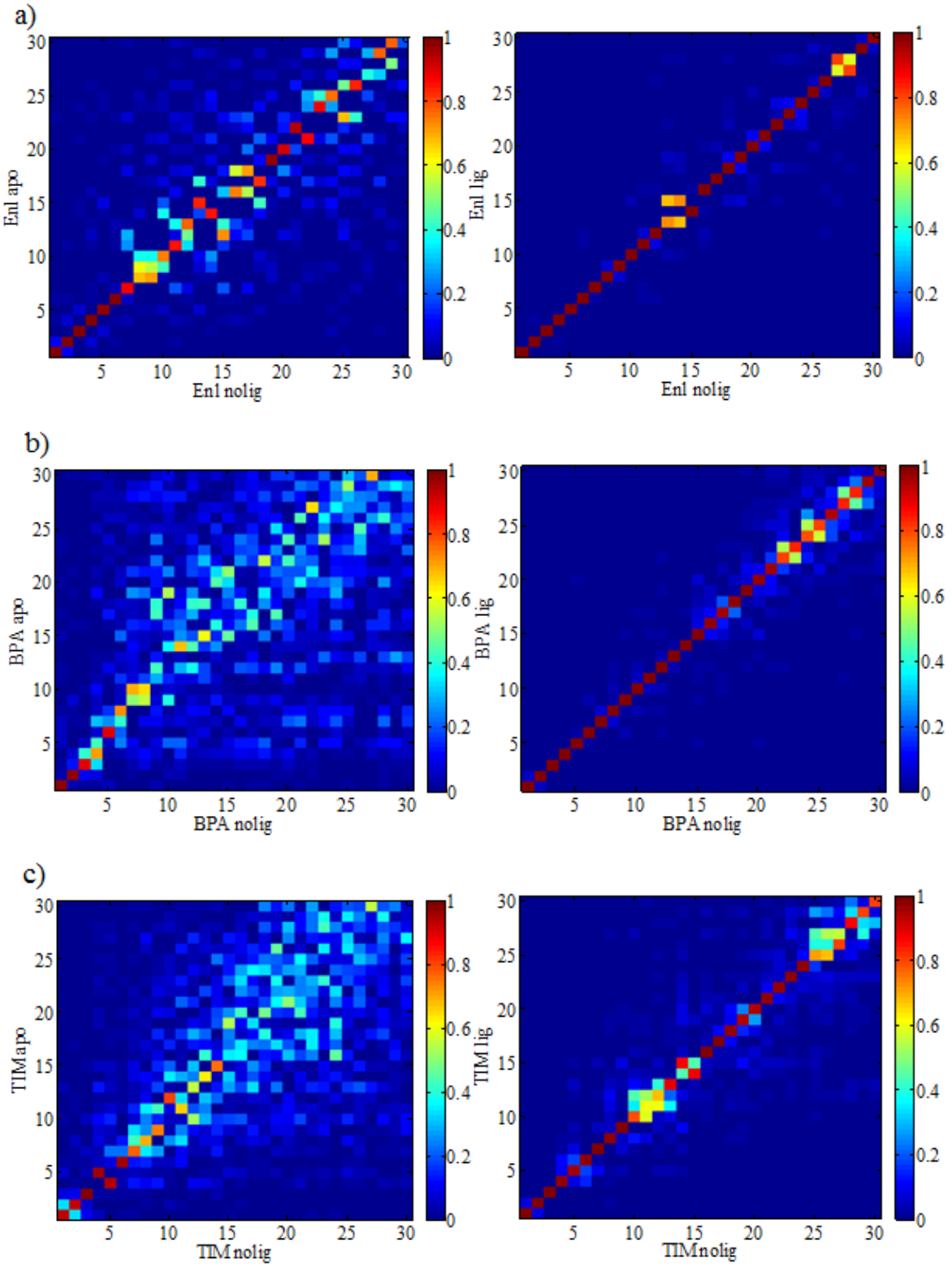


Figure 4.21. Overlap matrices for (a) Enl (dimer) (b) BPA (dimer) (c) TIM (dimer) (d) 3dhq (dimer) with 7 Å cutoff.

Table 4.2. Overlap values of enzymes with 7 Å cutoff distances.

Enzyme	Overlap		Overlap- diagonal only	
	Open and closed (no ligand)	Closed with and without ligand	Open and closed (no ligand)	Closed with and without ligand
PTP	0.95	0.99	0.36	0.96
bGT	0.85	0.96	0.27	0.64
BPA	0.88	0.99	0.38	0.95
Enl	0.94	0.99	0.66	0.93
TIM	0.90	0.99	0.47	0.87
3dhq	0.91	0.99	0.40	0.85
Lldh	0.94	0.99	0.48	0.90

Figures 4.22, 4.23 and 4.24 represent the percentage change in frequencies for the monomer bGT, the dimer Enl and the tetramer Lldh, respectively. Figures for the rest of the proteins are given in Appendix A. The effect of conformational change and presence of ligand at the catalytic site are separately reported for two cutoff distances (7 and 10 Å). In general, the conformational change of the loop shifts the slow mode frequencies to significantly higher values. In contrast, binding of the ligand to the closed conformer induces relatively lower shifts, again towards higher values.

The monomer bGT carries two identical ligands (UDP), one of them located at the catalytic site (UDP_cat) and the other at another site (UDP). Percentage change in frequency, overlap matrices and overlap value calculations for bGT are done by considering two ligands. However, the ligand at the catalytic site is mainly effective on frequencies of slow modes, which can be seen in Figure 4.22c. Specifically, the effect of UDP_cat almost equals that of two UDPs shown in green for comparison.

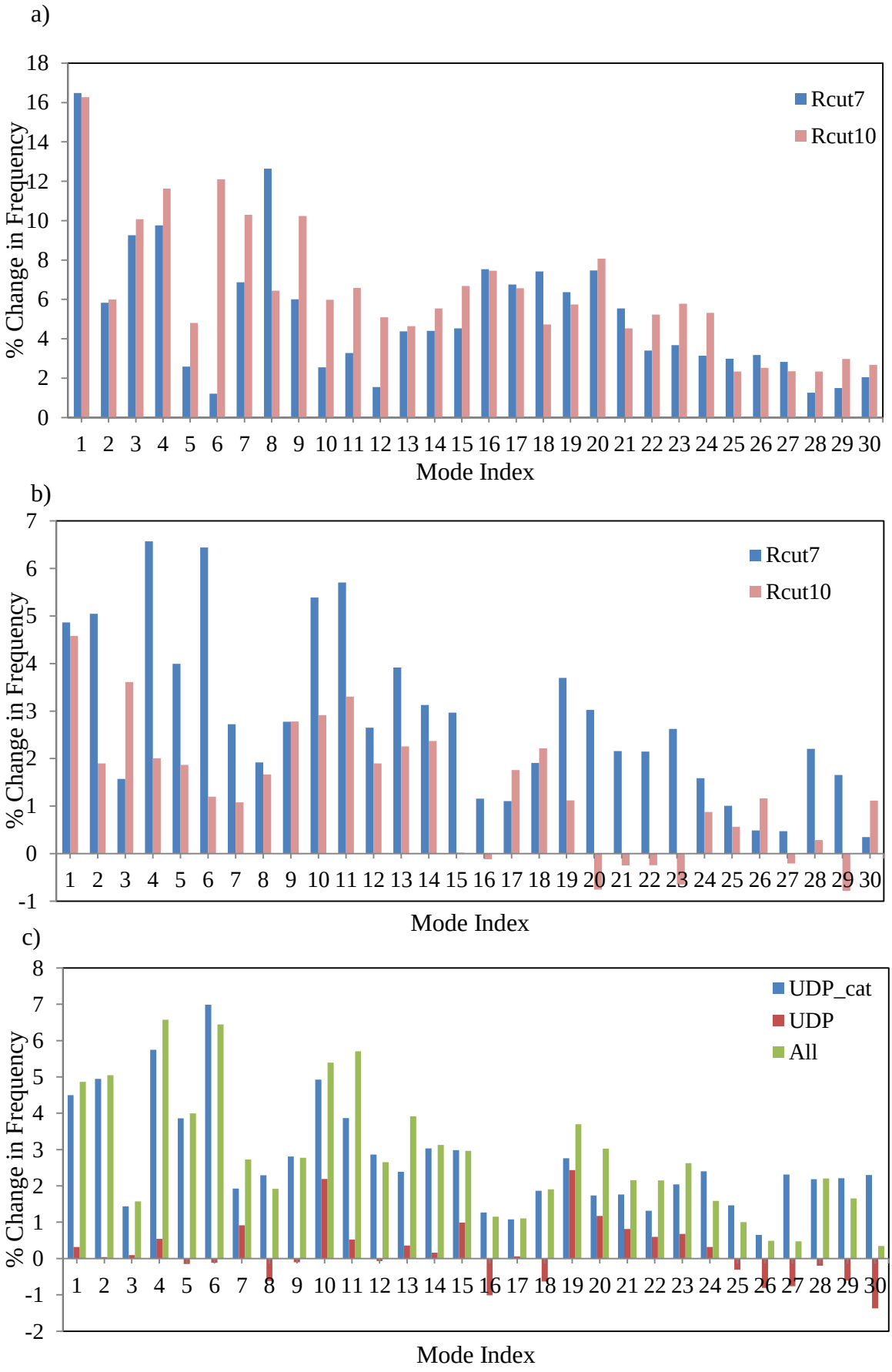


Figure 4.22. Percentage change in mode frequencies due to (a) conformational change (b) ligand binding (c) each ligand of bGT

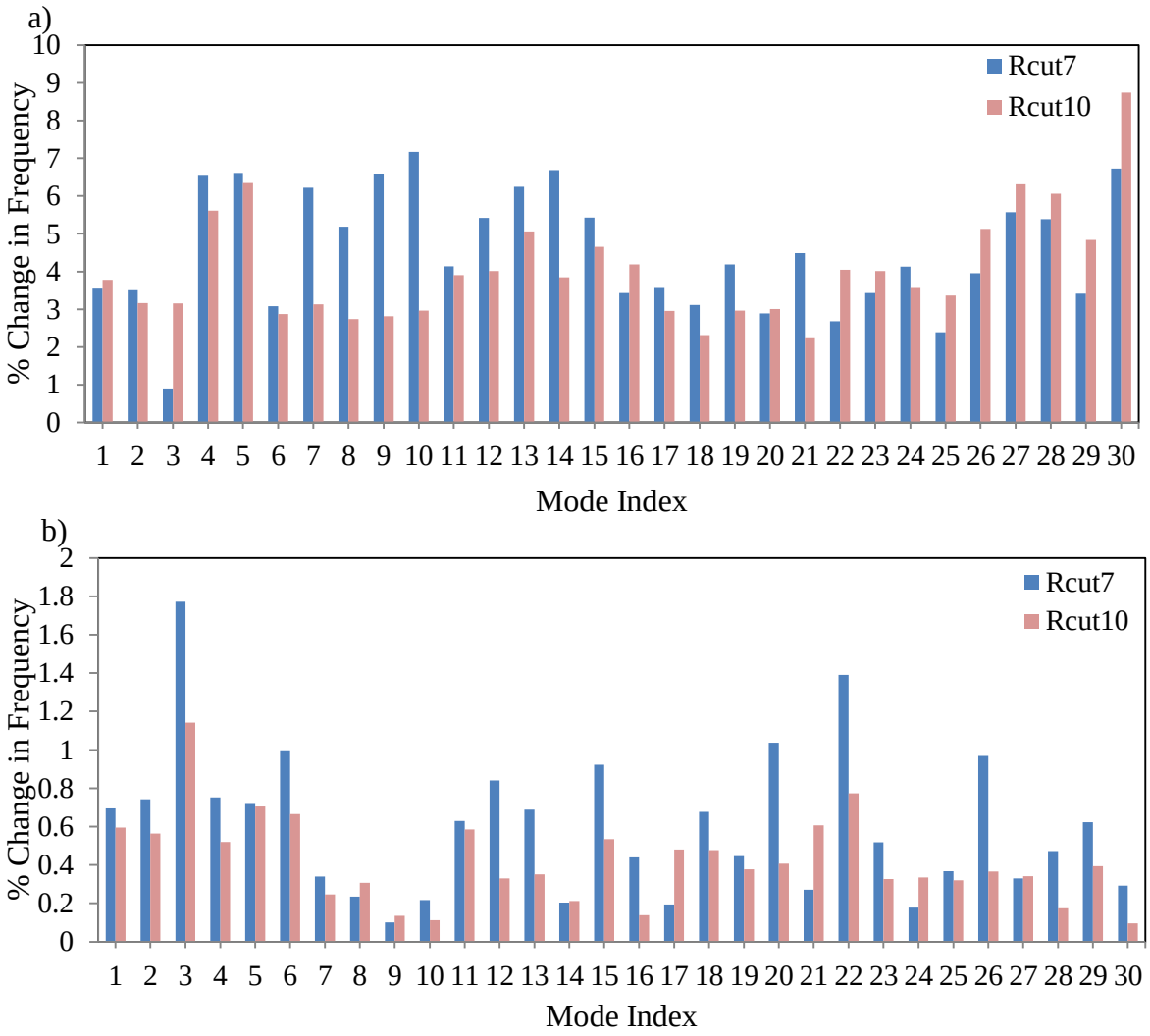


Figure 4.23. Percentage change in mode frequencies due to (a) conformational change (b) ligand binding of Enl

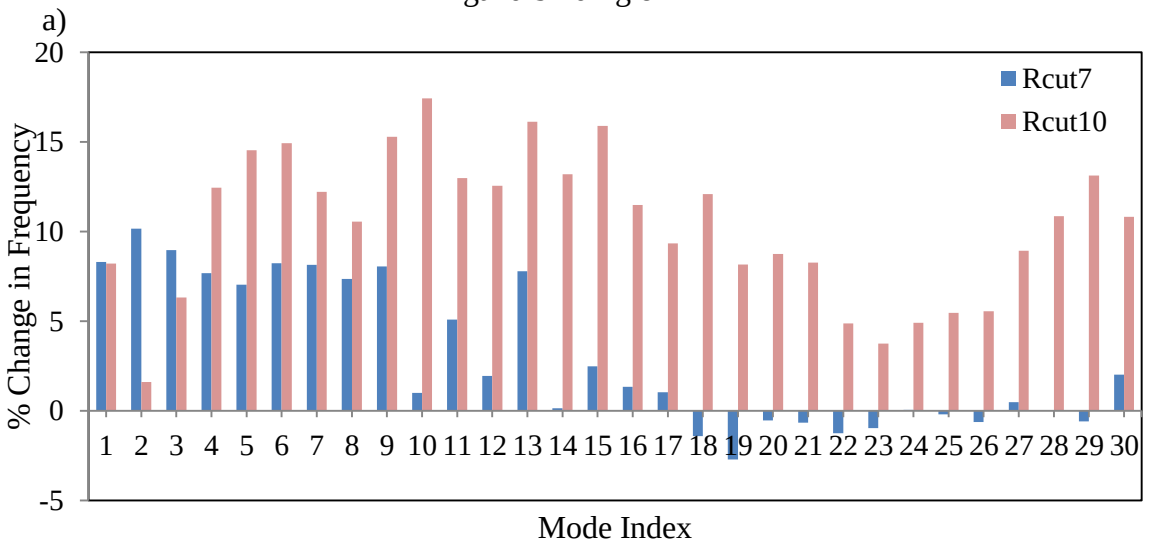


Figure 4.24. Percentage change in mode frequencies due to (a) conformational change (b) ligand binding of Lldh

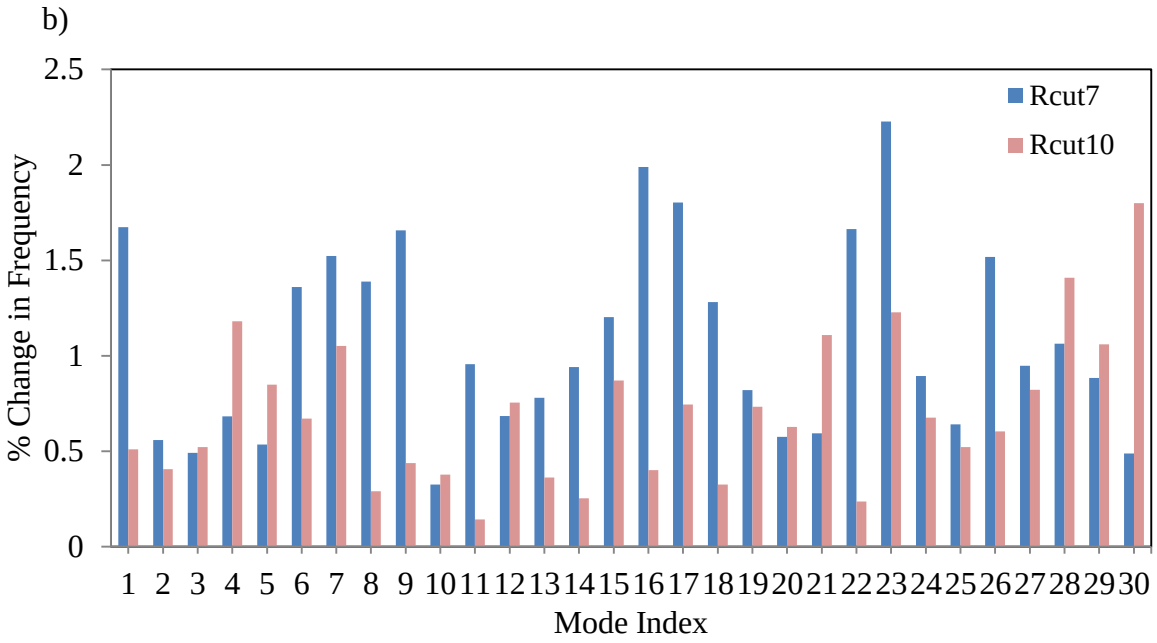


Figure 4.24. Percentage change in mode frequencies due to (a) conformational change (b) ligand binding of Lldh cont.

Conformational change and ligand binding effects on all enzymes are summarized in Figures 4.25 and 4.26, respectively. Average values of frequency change are reported for modes 1-5, 6-10, 11-15, 16-20, 21-25 and 26-30 (panel a) and for 1-5 and 1-30 modes (panel b) for cutoff 7 Å. Similar figures for cutoff 10 Å are given in Appendix A. Consistent results are observed for the two cutoffs, but the percentage changes are lower for cutoff 10 Å. Changes in the trends are observed for the tetrameric enzyme and PTP between the two cutoffs.

Conformational changes lead largest changes in frequencies. The ligand binding effect causes the highest change in bGT, whose ligand (UDP_cat) is the largest with at least twice the number of heavy atoms as compared to other ligands. It is also bound to the largest pocket among the monomeric and dimeric enzymes, as reported by CASTp (Dundas *et al.*, 2006) in Table 4.3.

Additionally, hinge residues are calculated using HINGE prot (Emekli *et al.*, 2008) server based on a single chain (PDB chain index given in parenthesis) for each enzyme. If any hinge residue (from first or second modes) is in contact with the ligand based on PDBsum server, that residue is underlined in the Table 4.3. In four of the enzymes (monomer PTP, dimers BPA and Enl and tetramer Lldh), the ligand interacts with one of

the hinge residues. If the hinges are calculated based on dimeric and tetrameric structures of the multimeric enzymes, hinges are observed in contact with ligand for BPA and Enl.

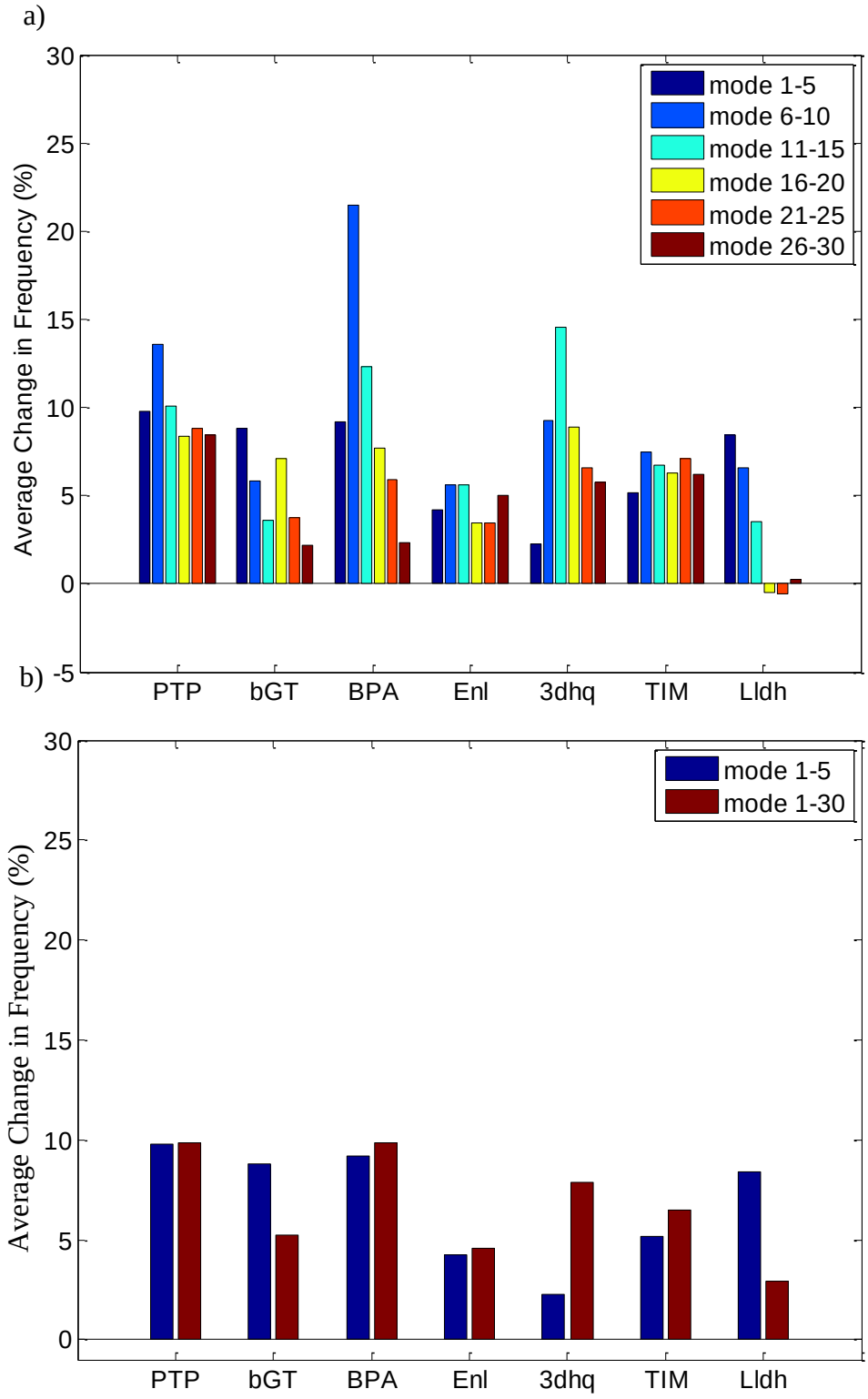


Figure 4.25. Average frequency percentage distribution due to conformational change for cutoff 7 Å for (a) each 5 (b) 5 and 30 modes.

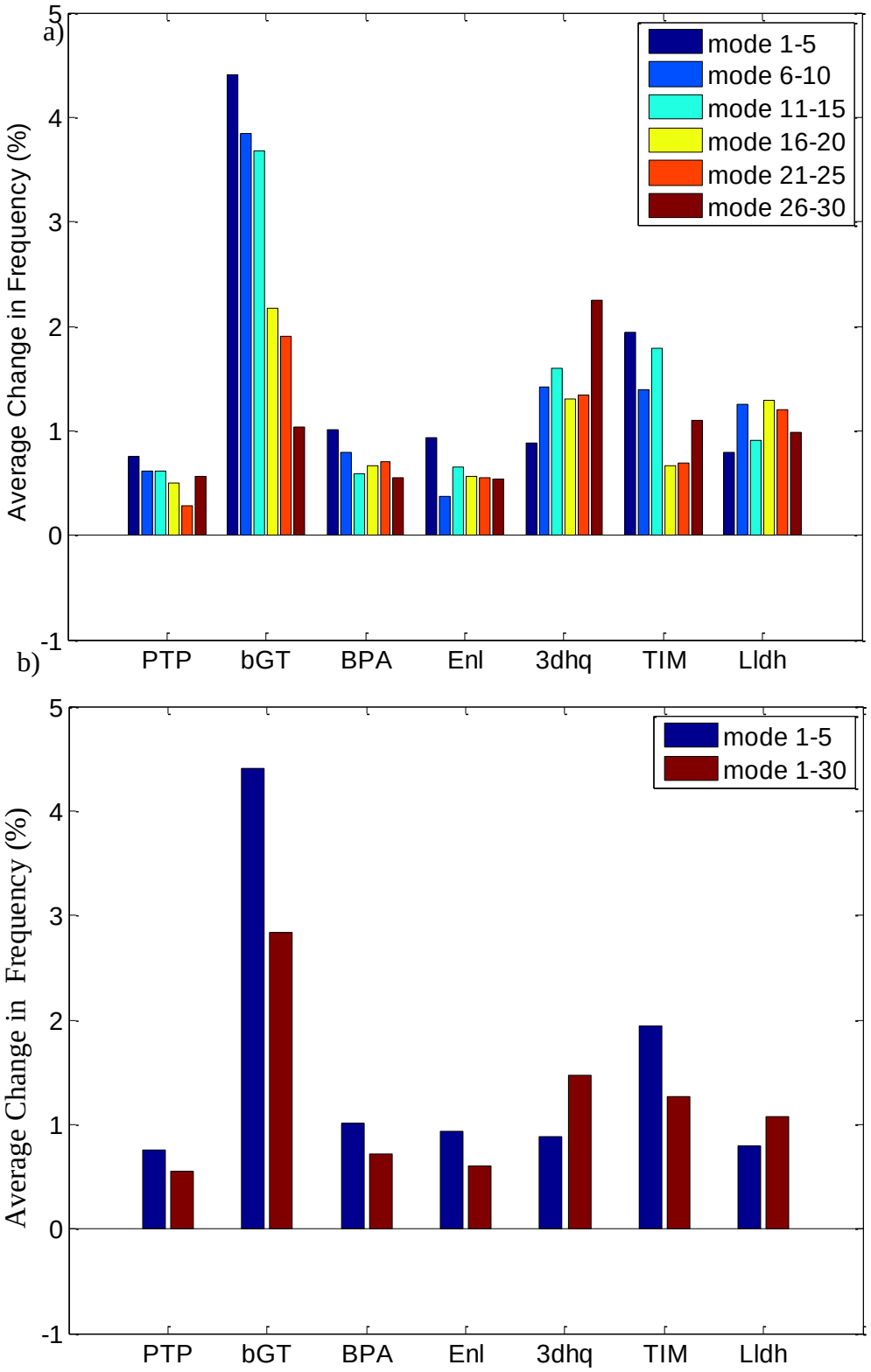


Figure 4.26. Average frequency percentage distribution due to ligand binding for cutoff 7 Å for (a) each 5 (b) 5 and 30 modes.

Table 4.3. Conformational changes, catalytic pocket sizes and hinge residues of enzymes.

Enzyme	Overall RMSD	Loop RMSD	Pocket Size Area (Å^2) Volume (Å^3)	Hinge Residues	
				Slowest mode1	Slowest mode 2
PTP	0.9	3.5	183 195.5	239, 404 (A)	199, 313
bGT	3.4	11.6	772.5 1022.5	216, 241, 295 (D)	249, 296, 315, 369
BPA	1.8	7.2	153.4 178.4	211, 299 (A)	103, 261
Enl	0.9	3.7	498.3 569.8	141,343, 391 (A)	117, 244, 355
TIM	1.0	4.5	187.9 209.3	91, 227 (1)	40, 126, 144, 161
3dhq	1.2	5.1	238.8 283.9	18,139 (A)	78, 205
Lldh	0.9	4.3	1297.5 1866.1	147, 232, 260 (A)	81, 153, 249

Loop RMSD and overall RMSD of the conformational change between open and closed forms are also reported in Table 4.3. To observe if there is any correlation between the extent of conformational changes, either loop or overall RMSDs and the change in frequencies, Figure 4.27a and b are plotted. In these figures, monomeric, dimeric and tertameric enzymes are demonstrated in red, blue and purple, respectively. The average effects of 30 modes are higher than 5 modes for the dimers, which mean conformational changes of loops, are more effective in the first 5 modes. Even though some correlation is observed in these figures when only the dimers are considered, there is no such correlation considering all enzymes.

In Figure 4.28, average changes in frequencies due to ligand binding are plotted against percentage of ligand atoms per residue. The variables seem to be correlated to an extent considering the limited number of enzymes.

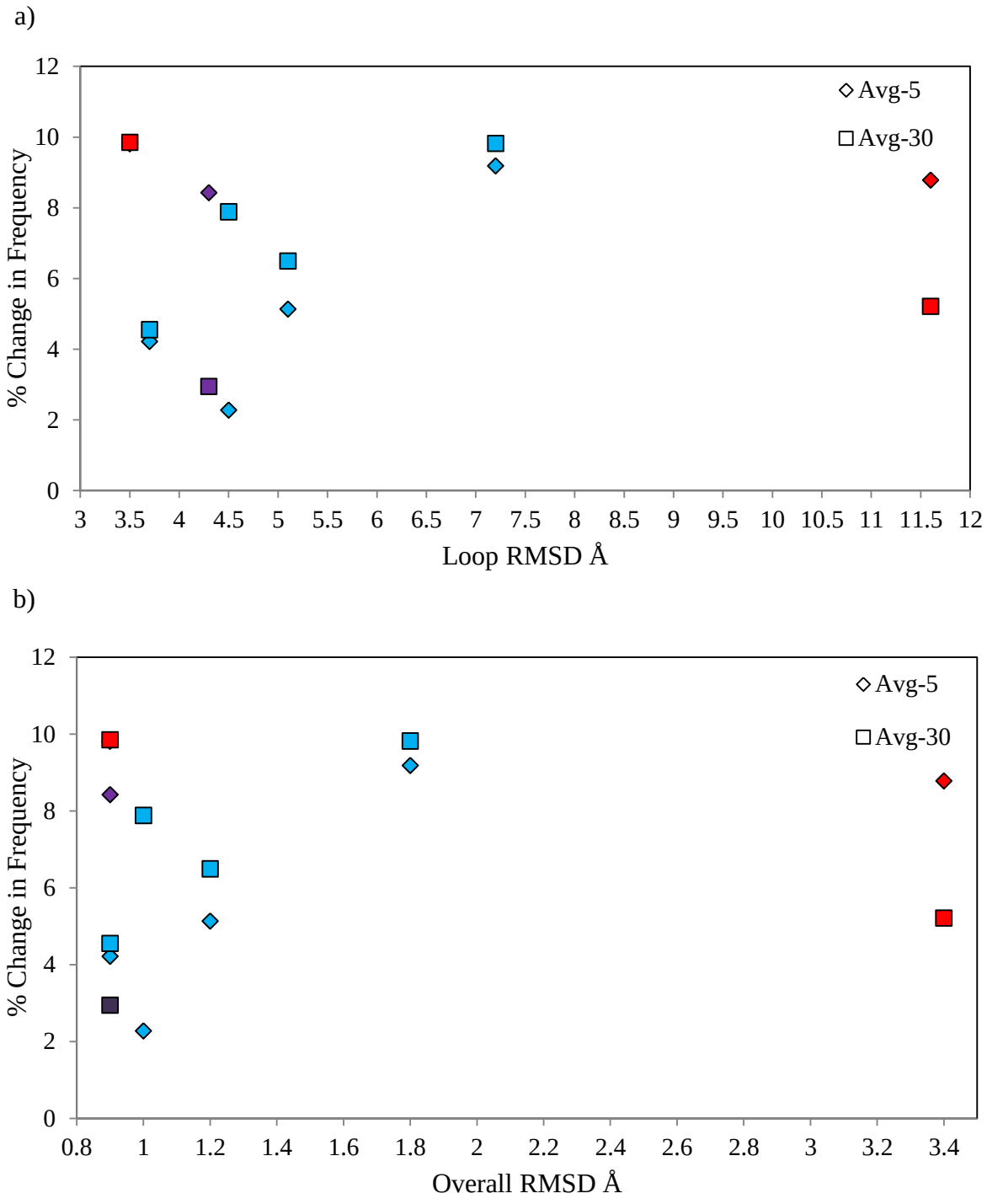


Figure 4.27. Average change in frequencies based on first 5 or 30 modes due to conformational change versus (a) loop RMSD (b) overall RMSD.

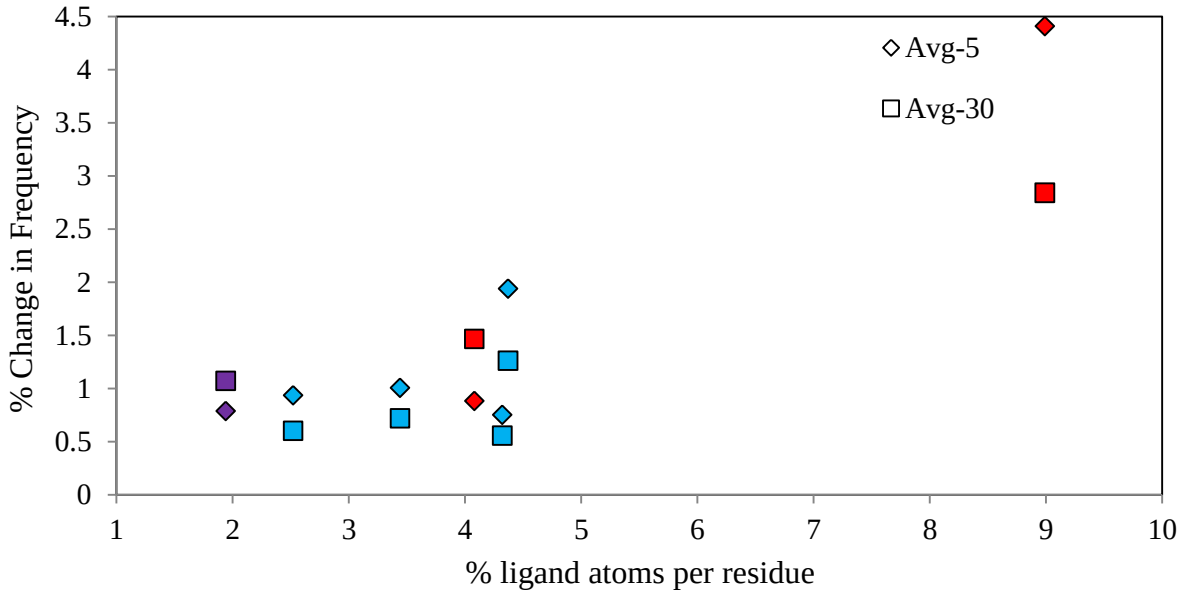


Figure 4.28. Average change in frequencies based first 5 or 30 modes due to ligand binding versus percentage of ligand atoms per residue.

A summary of frequency effect based on the first 5 and 30 modes is given in Table 4.4 for the two cutoffs. Here the changes in monomeric enzymes seem to be more dominant for both conformational change and ligand binding cases.

Table 4.4. Percentage change in frequencies for catalytic enzymes

	Cutoff 7Å				Cutoff 10 Å			
	Conformation Change		Ligand Binding		Conformation Change		Ligand Binding	
	1-5 modes	1-30 modes	1-5 modes	1-30 modes	1-5 modes	1-30 modes	1-5 modes	1-30 modes
Monomers	9.30	7.53	2.58	1.70	5.95	5.29	1.70	0.84
Dimers	5.20	7.19	1.19	1.01	4.66	5.86	0.80	0.66
Tetramers	8.42	2.94	0.79	1.07	8.62	10.35	0.69	0.70
All enzymes	6.83	6.68	1.53	1.21	5.59	6.34	1.04	0.72

4.3. MCG_ANM applications to allosteric enzymes

In this part of the thesis, MCG_ANM ($R_{c,a}$ = 7 and 10 Å) is used to study 12 different enzymes from the ASD (Allosteric Database), for which the allosteric sites and modulators are reported (Huang *et al.*, 2011; Huang *et al.*, 2014). Some of the enzymes also carry coenzymes or cofactors besides allosteric ligands. These enzymes form a diverse dataset consisting of four monomers, four dimers and four tetramers. Multiple crystal structures are used for some enzymes in dataset to analyze the effect of different ligands or tense/relaxed states. Table 4.5 summarizes properties of 17 crystal structures used in this section. In the ligand column of the Table 4.5, A and I indicate activator and inhibitor ligands, respectively. The star designates a ligand that is on the interface between the subunits and interacts with both chains. Number of ligand heavy atoms divided by total residue number in the enzyme (reported as %) gives idea about the relative size of the ligand. Last column gives the size of the pocket, where the allosteric ligand is bound. The pocket size is calculated using CASTp (Computed Atlas of Surface Topography of proteins) (Dundas *et al.*, 2006). The intact multimeric structures are provided to the CASTp server for pocket size calculations. In this section, frequency changes and overlap values are reported mainly for 7 Å cutoff and corresponding results for 10 Å are provided in Appendix B.

Table 4.5. Dataset of Allosteric Enzymes and Ligands.

Enzyme	PDB ID	Total # of Residues	Allosteric ligand	% of ligand atoms per residue	Pocket Area (Å^2), Volume (Å^3)
Checkpoint Kinase 1 (Chk1)	3jvr	258	$C_{16}H_{13}Cl_2N_3O_2$ (I)	8.91	-
3-Phosphoinoisitide-dependent Protein Kinase (Pdk1)	3hrf	284	$C_{17}H_{15}ClO_2$ (A)	7.04	151.1 206.2

Table 4.5. Dataset of Allosteric Enzymes and Ligands cont.

Focal Adhesion Kinase (FAK)	4ebw	259	$C_{20}H_{21}N_3O_2S$ (I)	10.04	1802.9 2999
ACK1 Kinase (ACK1)	1u4d	258	$C_{11}H_{11}N_5O_2$ (A)	6.98	497.9 757.5
Chorismate Mutase (CM)	1csm (x2)	504	$C_{11}H_{12}N_2O_2$(x2) (A)*	5.95	2852.1 3548.2
Acetylcholinesterase (AChE)	2ha5 (x2)	1068	$C_7H_{16}NOS$(x5) (A)	4.68	706 904.6
Glycogen Phosphorylase (GP)	3cej (x2)	1580	$C_{20}H_{17}ClF_3N_3O_4$(x2) (A)*	3.92	3508.9 5310.5
Reverse Transcriptase (RT_1)	3i0r (x2)	950	$C_{19}H_{20}ClN_3O_4S_2$ (I)	3.06	2425.3 3767.6
Reverse Transcriptase (RT_2)	4ifv (x2)	968	$C_{22}H_{18}N_6$ (I)*	2.89	4466.9 11327
Pyruvate Carboxylase (PC)	3ho8 (x4)	3888	$C_{21}H_{36}N_7O_{16}P_3S$(x4) (A)*	4.94	1079 1778.7
Pyruvate Kinase (PK_1)	1a3w (x4)	1962	$C_6H_{14}O_{12}P_2$(x4) (A)	4.08	1576.3 2240.1
Pyruvate Kinase (PK_2)	1t5a (x4)	2076	$C_6H_{14}O_{12}P_2$(x4) (A)	3.85	881.3 1533
Glyceraldehyde-3-Phosphate Dehydrogenase ($GAPN_1$)	1uxp (x4)	1996	$C_{10}H_{14}N_5O_7P$(x4) (A)	4.61	2486 3611.2
Glyceraldehyde-3-Phosphate Dehydrogenase ($GAPN_2$)	1uxq (x4)	1996	$C_6H_{13}O_9P$(x4) (A)	3.21	2568.3 3406.1

Table 4.5. Dataset of Allosteric Enzymes and Ligands cont.

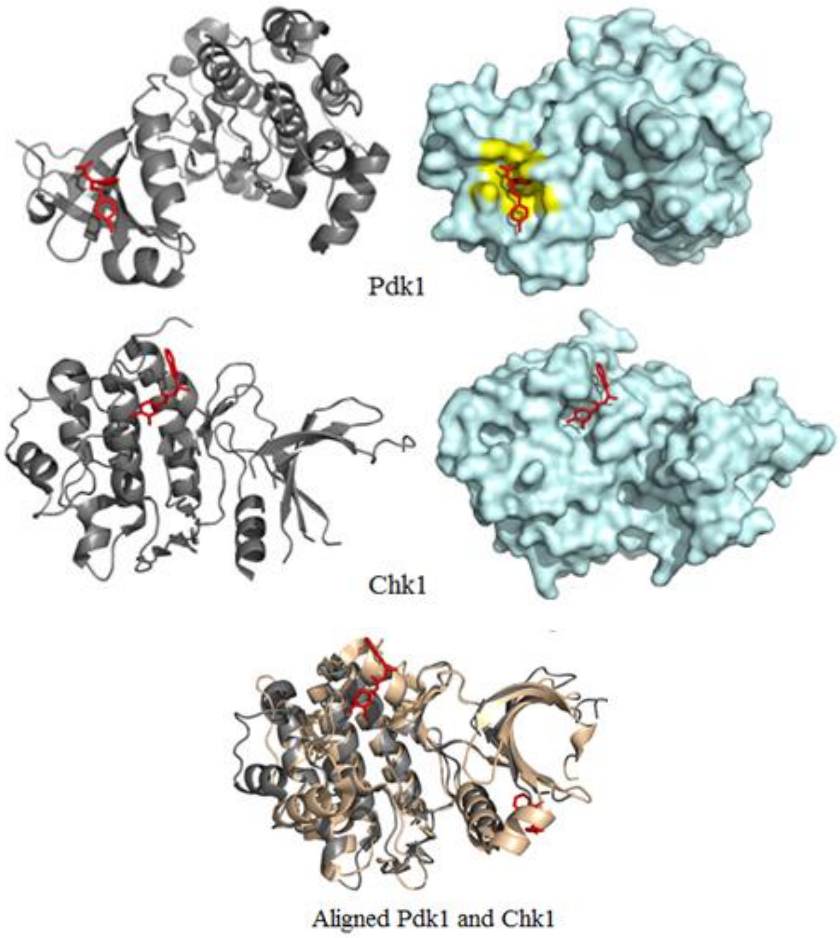
Glyceraldehyde-3-Phosphate Dehydrogenase ($GAPN_3$)	1uxr (x4)	1996	$C_6H_{13}O_9P$(x4) (A)	3.21	2507.1 3515.5
L-Lactate Dehydrogenase (Lldh-R)	1lth (x4)	1252	$C_6H_{14}O_{12}P_2$(x4) (A)	6.39	6535.7 11008
L-Lactate Dehydrogenase (Lldh-T)	1lth (x4)	1252	$C_6H_{14}O_{12}P_2$(x4) (A)	6.39	7221.4 15331

4.3.1. Allosteric Ligands Bound to Monomeric Enzymes

Four different monomeric enzymes are selected from the kinase family. Chk1 and FAK are bound to allosteric inhibitors, whereas Pdk1 and ACK1 are bound to allosteric activators. Cartoon representations of monomeric enzymes can be viewed in Figure 4.29 together with the bound ligands, shown in red sticks. Among four monomers, Pdk1 and Chk1 are structurally similar, even though their sequence identity is extremely low, namely 3%, based on Dali Server (Holm and Rosenstrom, 2010). Their ligands are bound to different regions, as can be observed in the alignment provided in the same figure.

Percentage changes in frequency are compared in Figure 4.30 for monomers. Inhibitor effect on Chk1 frequencies is extremely low, whereas activator effect on ACK1 values is highest among all monomers with the first two modes causing above 10% change. Looking at the binding pocket sizes, Chk1 is bound to the smallest pocket reported in Table 4.5, whereas ACK1 is bound to one of the largest pockets. This distinction between the inhibitor and activator effect is not observed for the other two enzymes, which present an intermediate level of change in frequencies. Interestingly, FAK is bound to an extremely large pocket, which in fact has a tunnel-like shape passing through the enzyme. These pockets are presented in Figure 4.29.

a)



b)

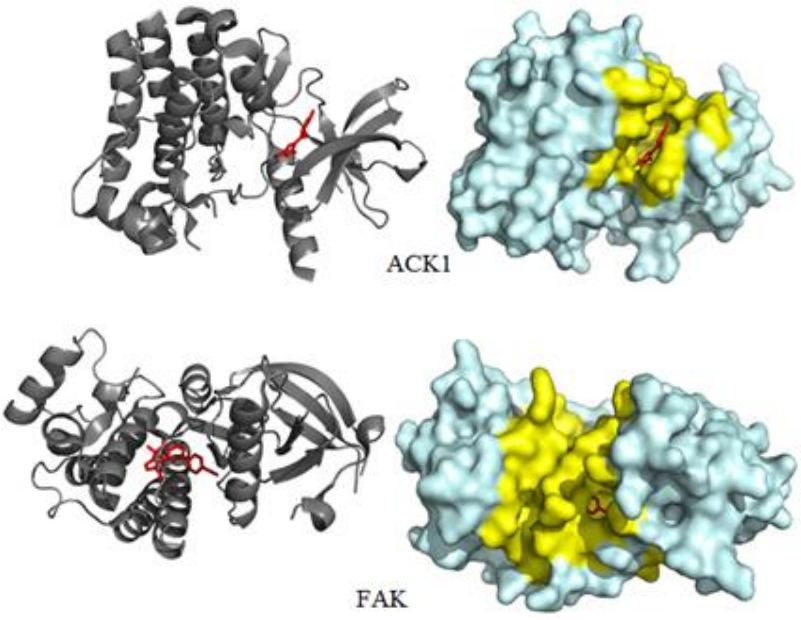


Figure 4.29. Cartoon and surface representations of monomeric enzymes (a) Chk1 and Pdk1 (b) ACK1 and FAK enzymes with allosteric ligands and pockets.

The overlap matrices in the presence and absence of allosteric ligands are given in Figure 4.31. An almost one-to-one correspondence is observed for Chk1, which is also least effective on frequency shifting. For the other three enzymes, mode swapping and shifting in mode character are observed to a certain extent after the first 10 modes or so. In summary, a characteristic difference cannot be reported between inhibitor and activator effects based on the monomeric enzymes studied.

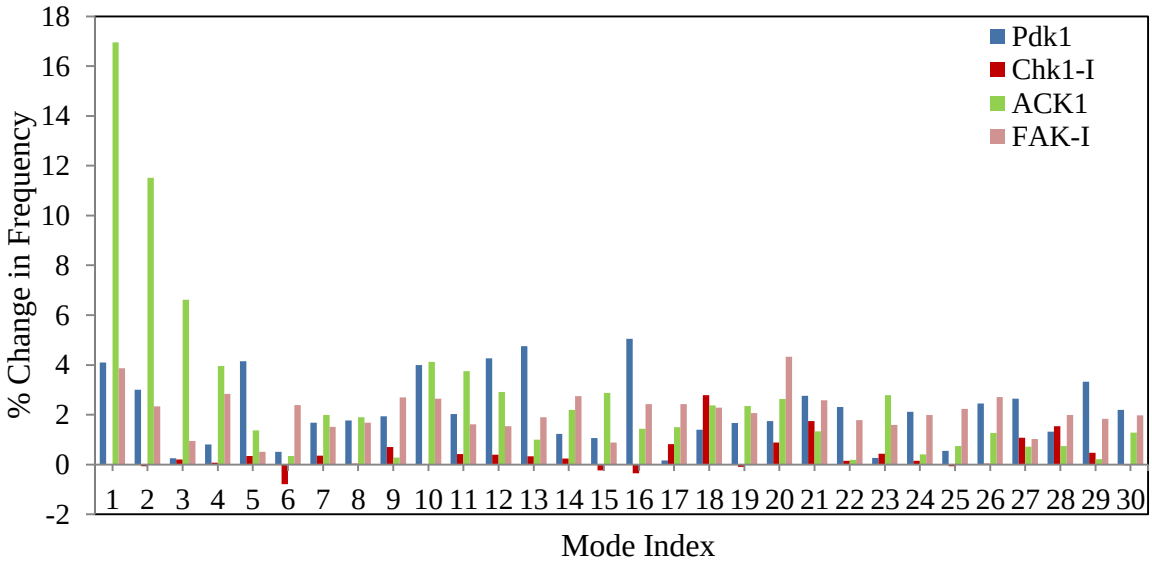


Figure 4.30. Percentage change in mode frequencies due to allosteric ligand binding to monomeric enzymes with 7 Å cutoff.

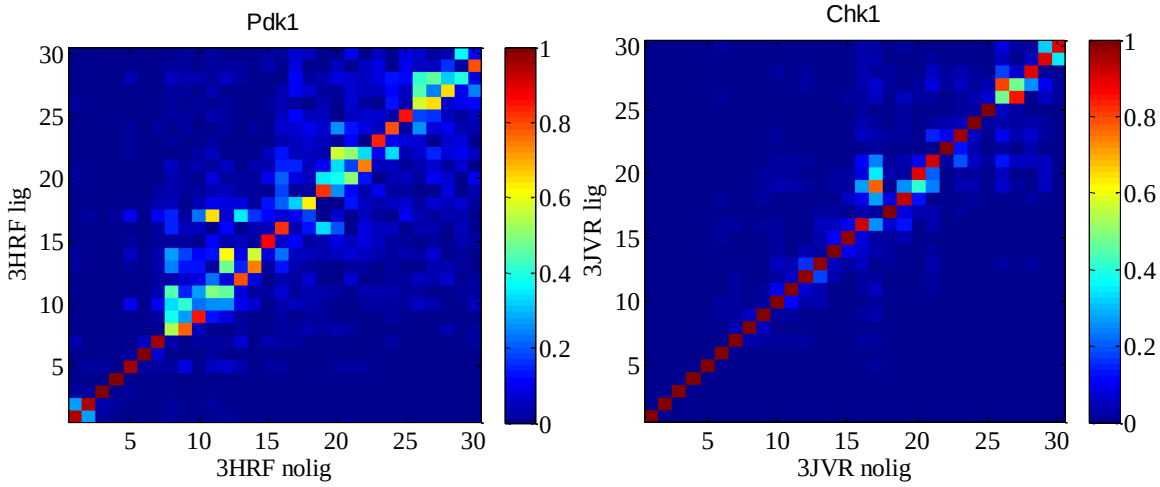


Figure 4.31. Overlap matrices due to allosteric ligand binding to monomeric enzymes with 7 Å cutoff.

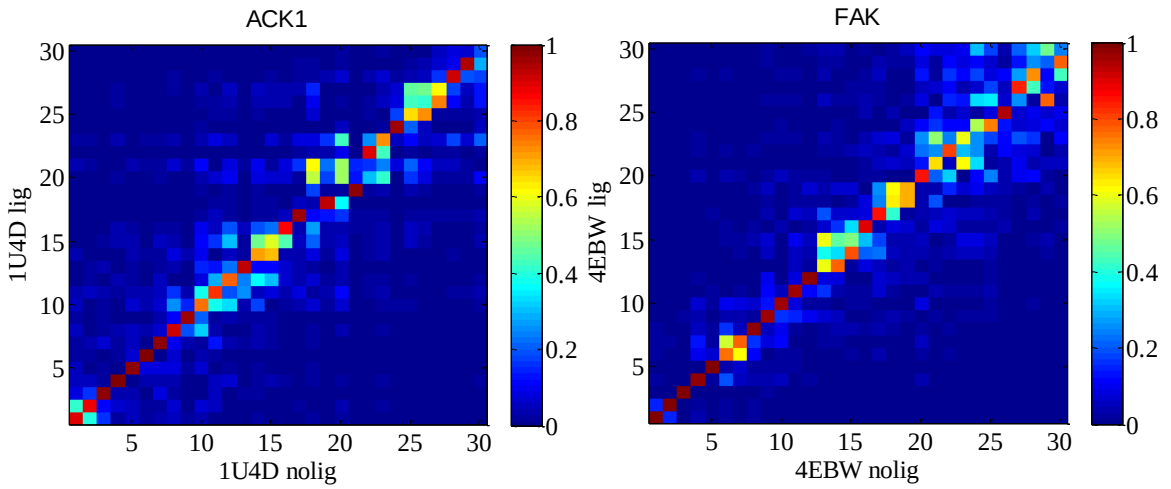


Figure 4.31. Overlap matrices due to allosteric ligand binding to monomeric enzymes with 7 Å cutoff cont.

4.3.2. Allosteric Ligands Bound to Dimeric Enzymes

Among five dimeric crystal structures, two belong to reverse transcriptase bound to different inhibitors, distinguished as RT_1 and RT_2. Crystal structures of dimers and the specific pockets, where the allosteric ligands are bound, are presented in Figure 4.32a and b, where the residues forming the cavity surface are colored yellow and the bound ligands are shown as red sticks. The sizes of these pockets reported in Table 4.5 for the dimers are extremely large. In fact, each pocket represents a large connected tunnel-like cavity passing all the way through the interface in CM and GP. In RT, the tunnel-like cavity starts at the interface and goes through one of the monomers. In AChE, a deep cavity is located on each subunit. In subunit A there are two and in subunit B there are three allosteric ligands which are closer each other.

Percentage changes in frequency are plotted in Figure 4.33 for dimers (only RT_1 shown). Ligand effect on the first three modes of GP is comparatively high, possibly due to the binding of ligands to the interface region. The overlap matrices of dimer structures given in Figure 4.34 show less scattering than monomer structures. This indicates that mode shifting effect of ligands decrease as system size and number of subunits increase.

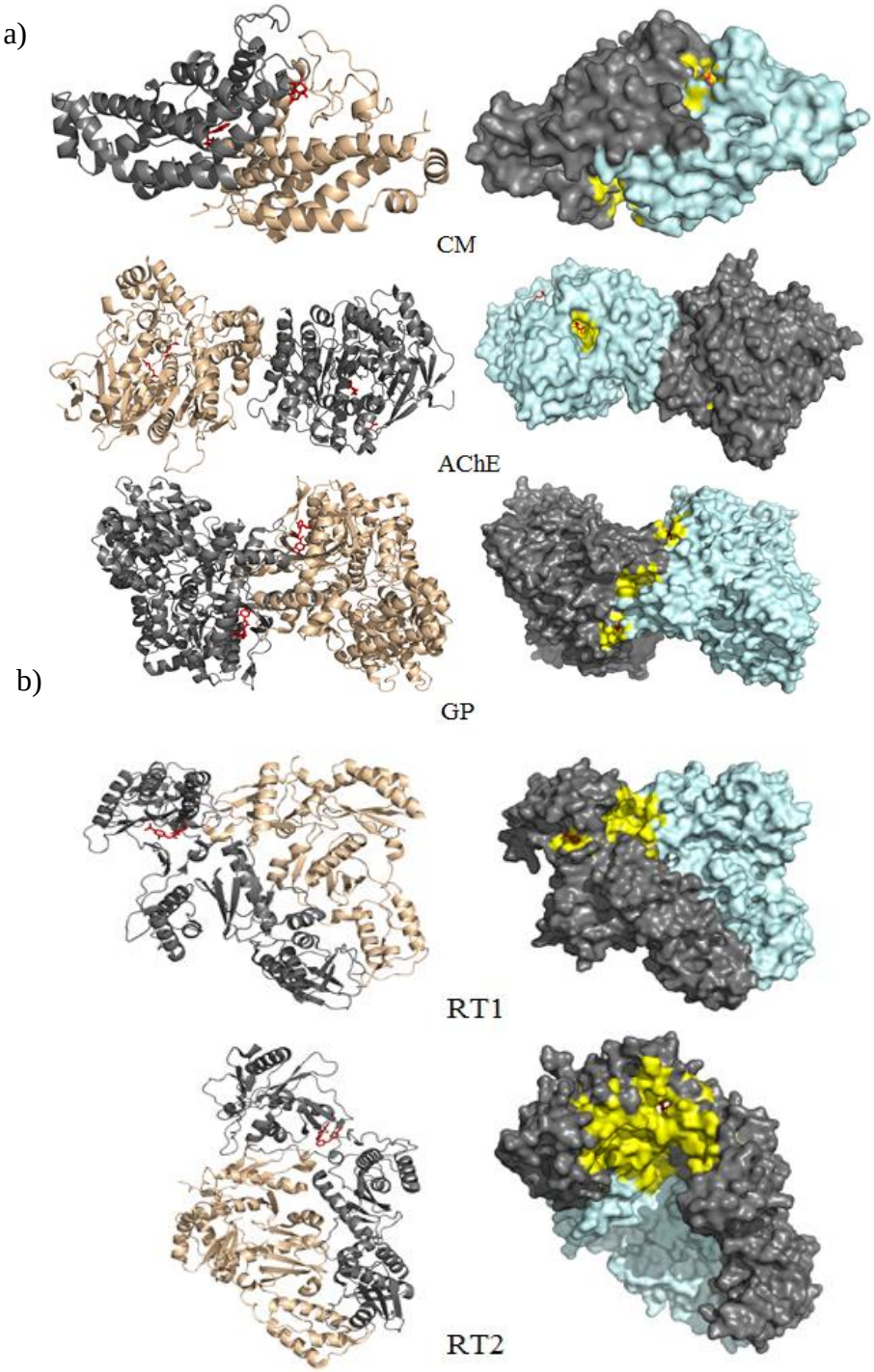


Figure 4.32.Cartoon and surface representations of dimer enzymes (a) CM, AChE and GP (b) RT1 and RT2 enzymes with allosteric ligands and pockets.

Figure 4.32. Cartoon and surface representations of dimer enzymes (a) CM, AChE and GP (b) RT1 and RT2 enzymes with allosteric ligands and pockets cont.

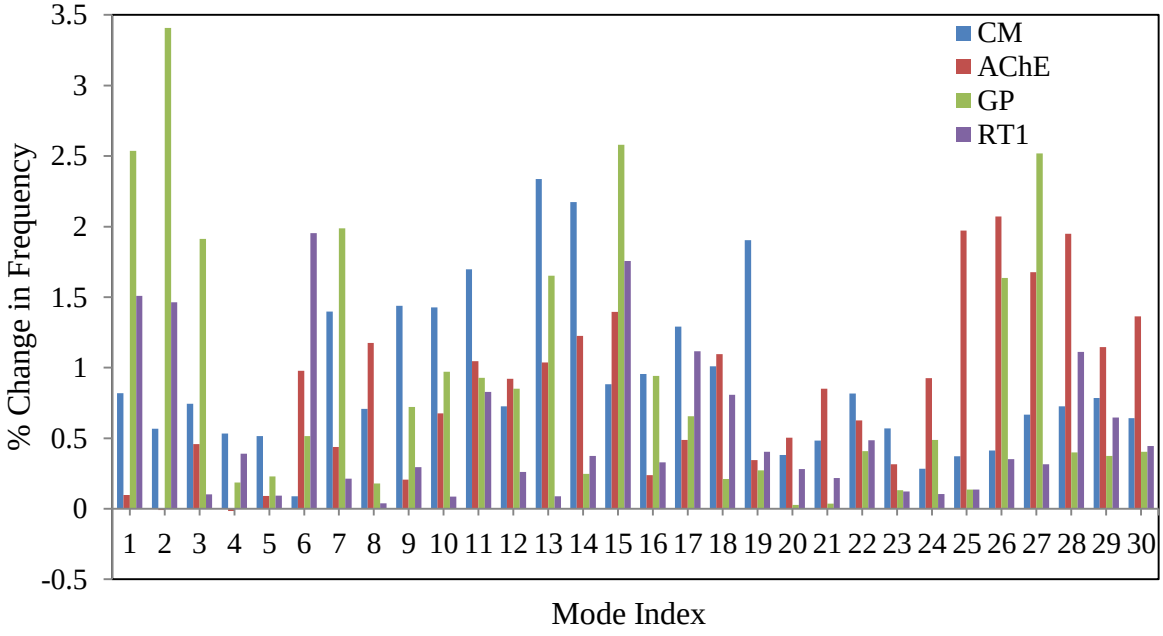


Figure 4.33.Percentage change in mode frequencies due to allosteric ligand binding to dimeric enzymes with 7 Å cutoff.

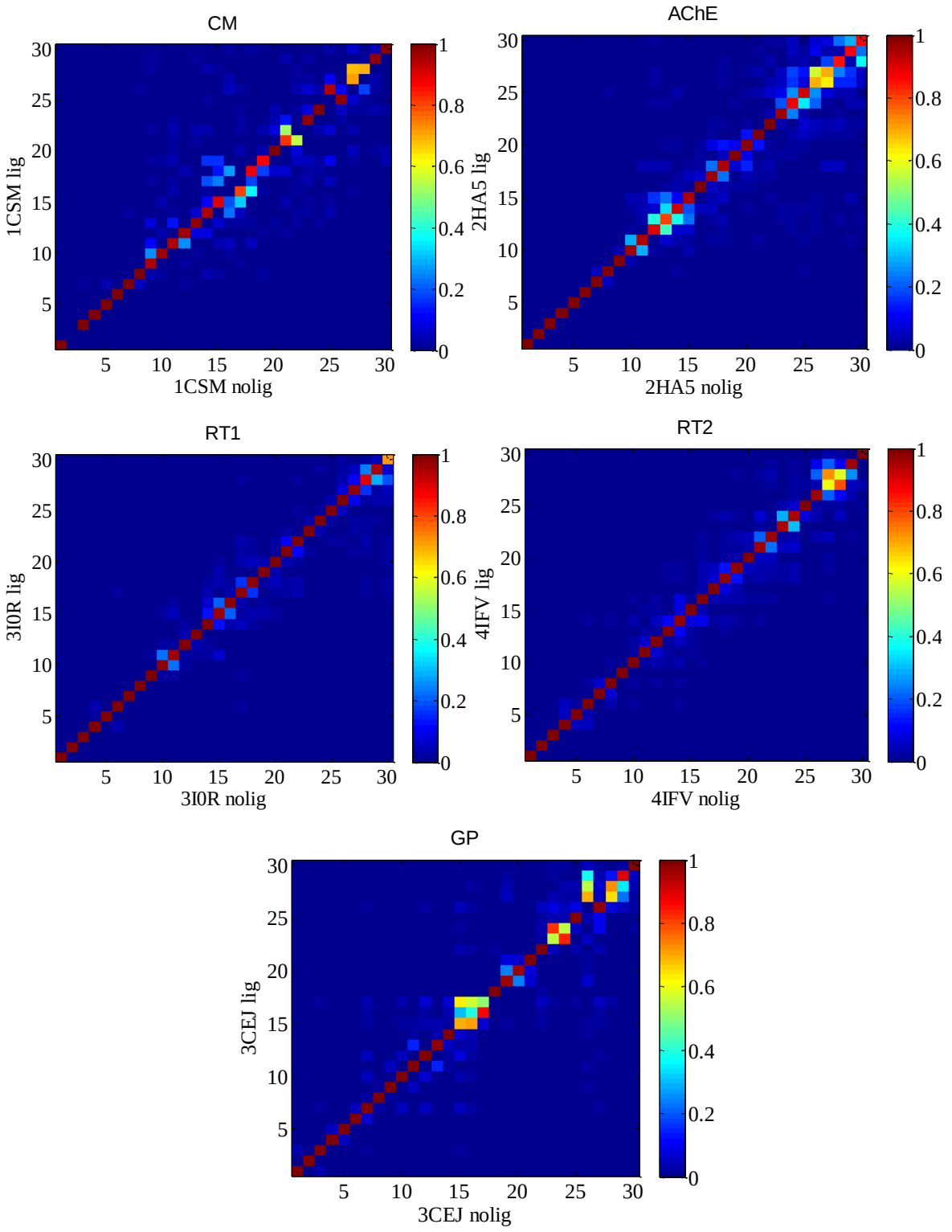


Figure 4.34. Overlap matrices due to allosteric ligand binding to dimer enzymes with 7 Å cutoff.

There are two crystal structures (RT_1 and RT_2) for reverse transcriptase, each bound to a different inhibitor. The RMSD between the two structures is 1.36 Å. RT is the only

hetero-dimer in the dataset and each inhibitor is bound to approximately the same site in one of the subunits. Even though the inhibitor in RT_2 is larger than the one RT_1, there is no significant difference in terms of frequency changes between the structures shown in Figure 4.35.

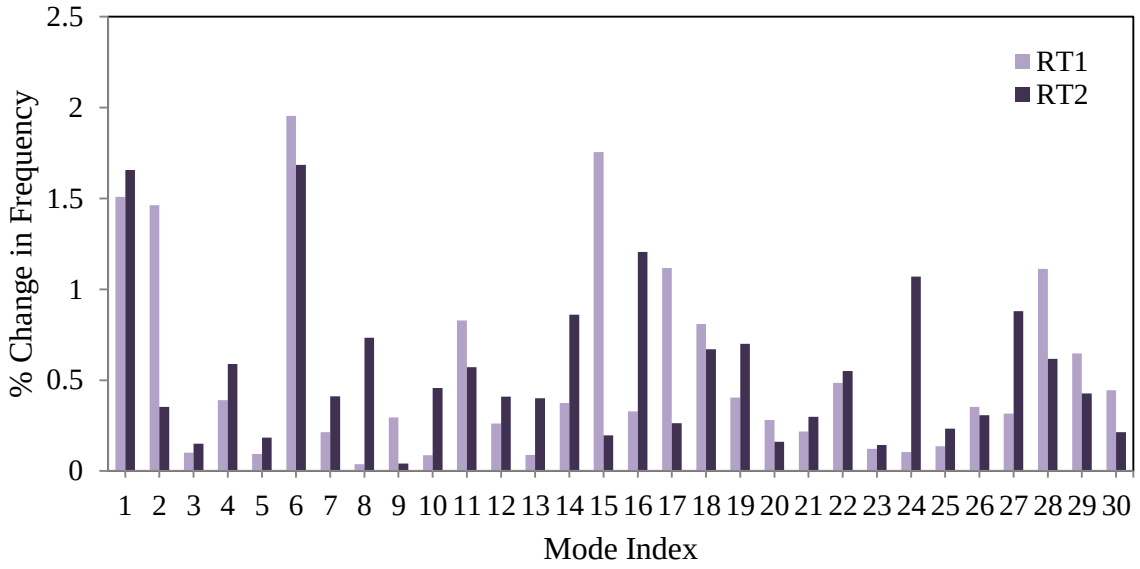


Figure 4.35. Percentage change in mode frequencies due to allosteric ligand binding to RT1 and RT2 dimeric samples with 7 Å cutoff.

4.3.3. Allosteric Ligands Bound to Tetrameric Enzymes

Mainly four different tetrameric enzymes are studied in this section, but multiple crystal structures for some of them are reported, which will be discussed in next sections. Crystal structures of tetramers can be viewed in Figure 4.36.

Percentage change in frequency for tetramers is plotted in Figure 4.37 using a single structure for each enzyme, which is the first structure reported in italic in Table 4.5. Binding to PC causes higher frequency changes in the slowest modes. Its binding pocket is small and ligand size is large among the tetramers. The allosteric ligand of PC is located between the two domains forming its monomeric structure, i.e. a hinge region. Overlap matrices of tetramer structures are shown in Figure 4.38. Overlap matrices for modes with and without ligand display a neat one-to-one correspondence, possibly due larger sizes of the tetrameric systems.

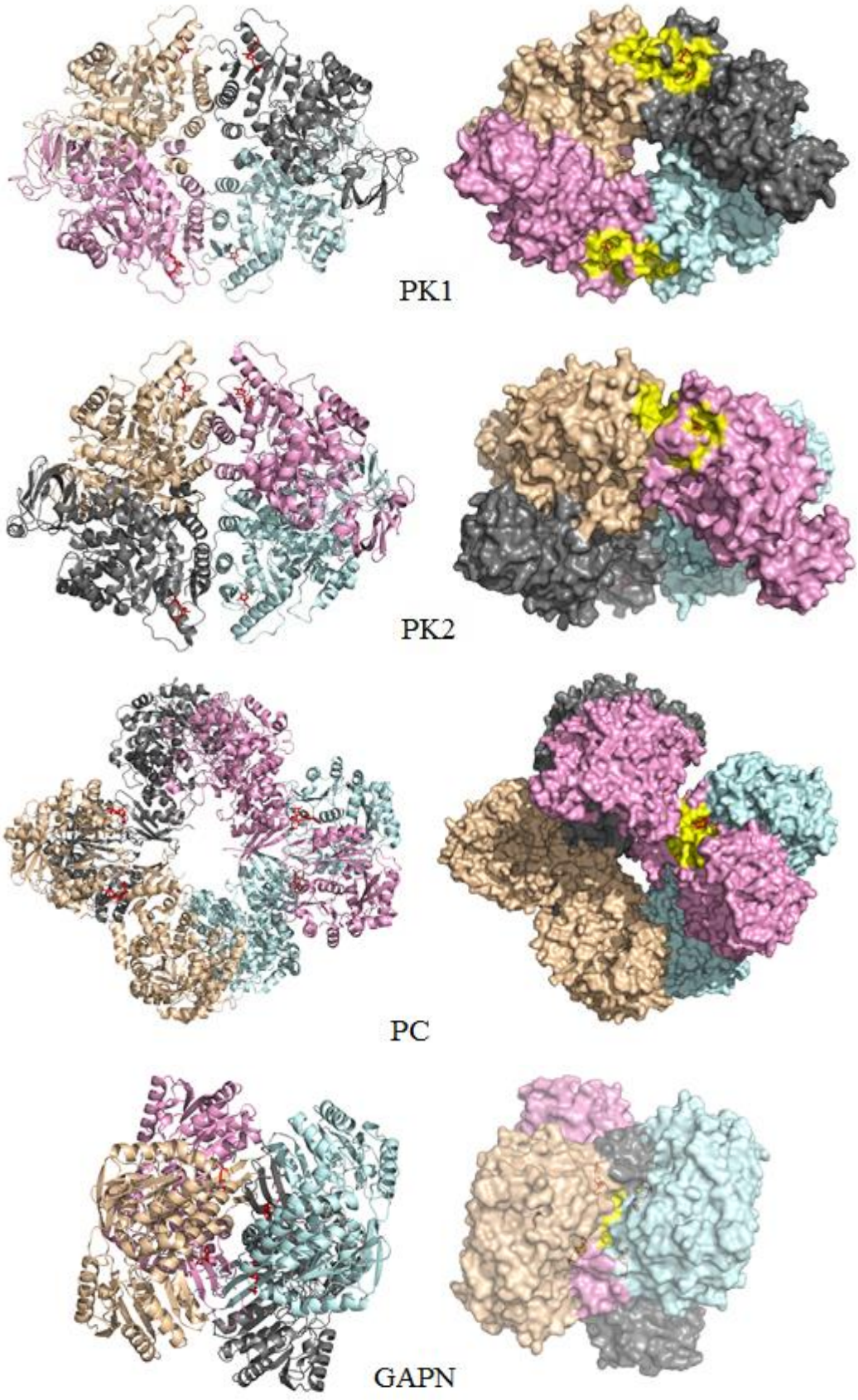


Figure 4.36. Cartoon and surface representations of tetramer enzymes with allosteric ligands and pockets.

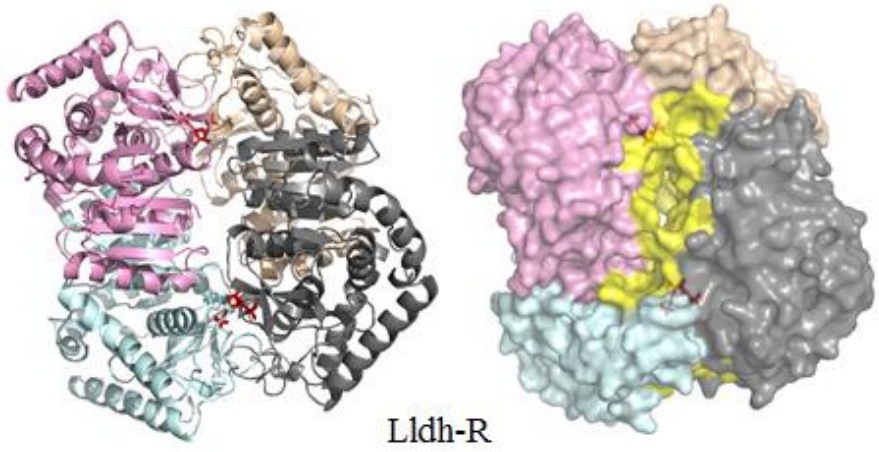


Figure 4.36. Cartoon and surface representations of tetramer enzymes with allosteric ligands and pockets cont.

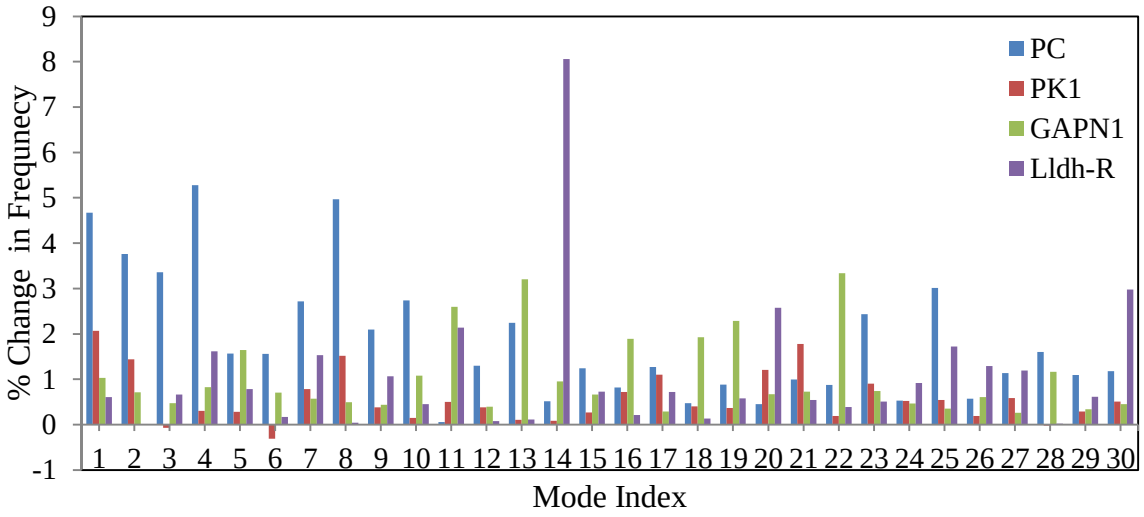


Figure 4.37. Percentage change in mode frequencies due to allosteric ligand binding to tetrameric enzymes with 7 Å cutoff.

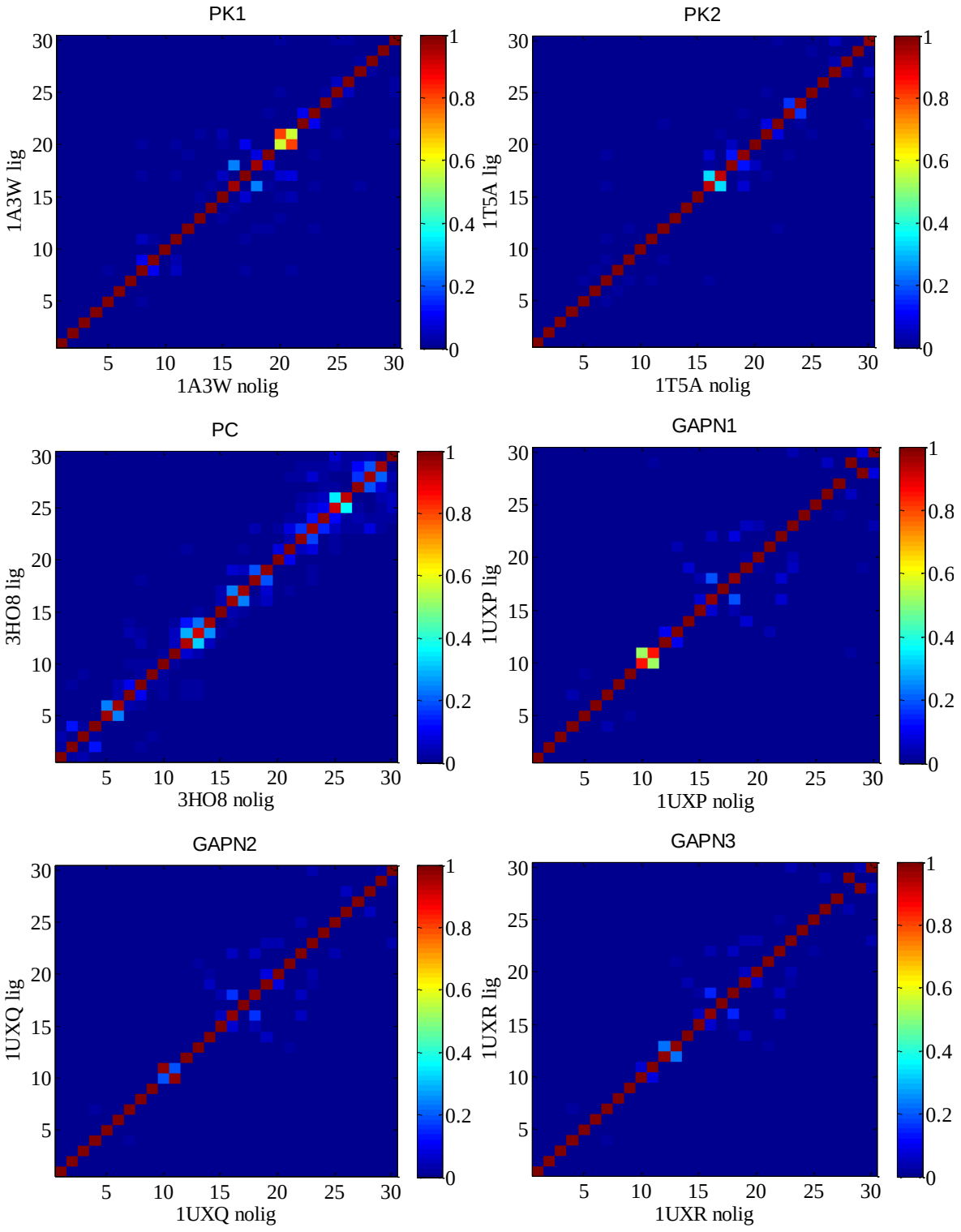


Figure 4.38. Overlap matrices due to allosteric ligand binding to tetrameric enzymes (PKs, PC, GAPNs, Lldh-T and Lldh-R with 7 Å cutoff.

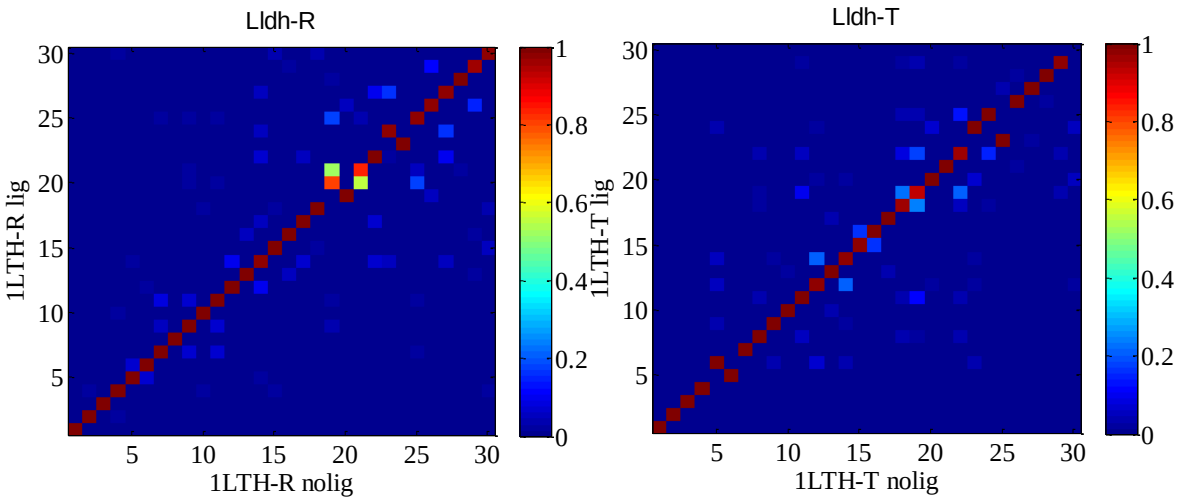


Figure 4.38. Overlap matrices due to allosteric ligand binding to tetrameric enzymes (PKs, PC, GAPNs, Lldh-T and Lldh-R with 7 Å cutoff cont.

4.3.3.1. Different Ligands Bound to the Same Tetramer. Multiple crystal structures of pyruvate kinase and glyceraldehyde-3-phosphate dehydrogenase are considered with different ligands binding to the same site. The effects of these different ligands are compared according to their sizes. Ligand bound to GAPN1 has 92 heavy atoms, whereas the same ligand with 64 atoms is bound to GAPN2 and GAPN3, their percentage per residue values are in Table 4.5. RMSD value of GAPN1 and GAPN2 equals 0.27 Å, GAPN1 and GAPN3 equals 0.20 Å and GAPN2 and GAPN3 equals 0.30 Å. In all three structures, the same modes are affected by the ligand, as shown in Figure 4.39. The larger ligand in GAPN1 has more pronounced effect on some of the modes.

PK_1 and PK_2 are different crystal structures of same enzyme with same ligand. As shown in Figure 4.40, PK_1 values are higher than PK_2. Binding pocket size of PK_1 is bigger than PK_2 and percentage of ligand atoms over residue numbers of PK_1 is higher than PK_2. RMSD of tetrameric PKs are 2.96 Å.

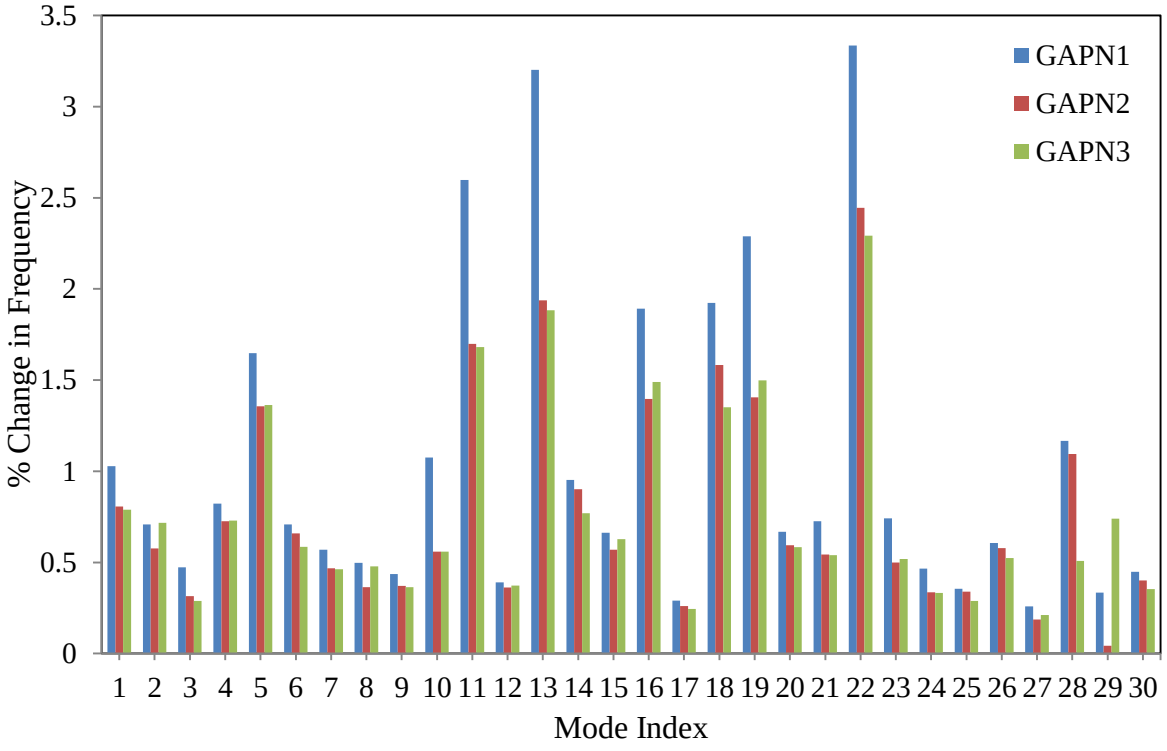


Figure 4.39. Percentage change in mode frequencies due to allosteric ligand binding to GAPN tetrameric enzymes with 7 Å cutoff.

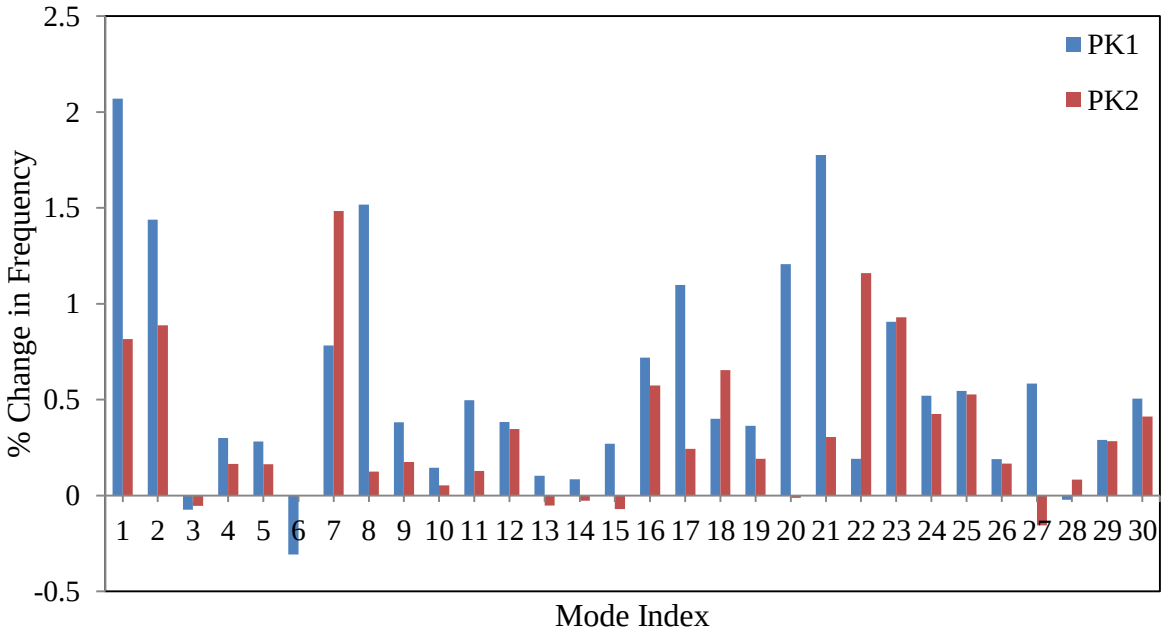


Figure 4.40. Percentage change in mode frequencies due to allosteric ligand binding to PK tetrameric enzymes with 7 Å cutoff.

4.3.3.2. Comparison of Tense and Relaxed States of a Tetramer. There are two conformations for L-lactate dehydrogenase, namely tense and relaxed states. Tense state has lower affinity for substrate binding and has lower catalytic activity than relaxed state (Stefan *et al.*, 2008). To determine the binding effect on the tense and relaxed conformations of the same enzyme bound to the same ligand, MCG_ANM is applied to both states of Lldh. Figure 4.41 compares the effect on frequencies. For the tense state, ligand binding affects the first 5 modes more than the relaxed state, with average changes of 1.32 and 0.73, respectively. In contrast, the average change in first 30 modes is close to each other (0.90 for tense and 1.08 for relaxed). RMSD value between tense and relaxed monomers of Lldh is 0.91 Å, but based on tetrameric structures RMSD equals to 3.028 Å.

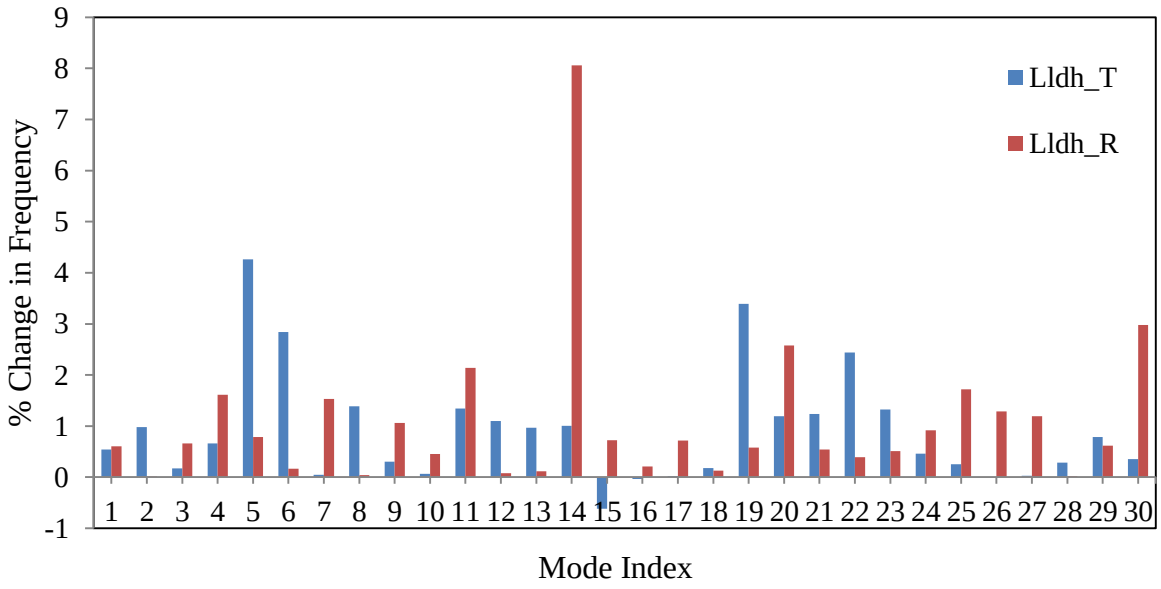


Figure 4.41. Percentage change in mode frequencies due to allosteric ligand binding of Lldh tetrameric sample of tense and relaxed form with 7 Å cutoff.

4.3.4. Overview of Allosteric Ligand Effect on Enzyme Collective Dynamics

Figure 4.42 summarizes the effect of allosteric ligands on collective modes. Higher percentage change is observed in three of the monomeric enzymes (Pdk1, ACK1 and FAK) and one tetramer (PC). Table 4.6 summarizes overlap values, both averaging the whole matrix and its diagonal for different cutoffs. Relatively lower overlap values, indicating mode swapping and/or character change, are observed for monomers. Highest ratio of

ligand atoms per residue exist for monomeric enzymes. Specifically, highest change in first 5 modes due to ligand binding is observed for monomer ACK1 and tetramer PC.

In Table 4.7, hinge residues are reported using HingeProt server (Emekli *et al.*, 2008). The residues are calculated based on only a single chain (chain A) for each enzyme. If any of the reported hinge residues interact with the allosteric ligand based on PDBSUM server (Laskowski *et al.*, 1997) that specific residue is underlined in the table. At the same time, hinge residues of intact multimeric structures are also calculated and given in Table B.1. In Table 4.7 based on monomeric forms, one of the hinge residues in either the 1st or 2nd modes interacts with the ligand for about half of the enzymes considered. Specifically, ACK1 and PC with dominant changes indicate such hinge residues. In contrast, this situation does not exist when multimeric forms of the same enzymes are considered as listed in Table B.1.

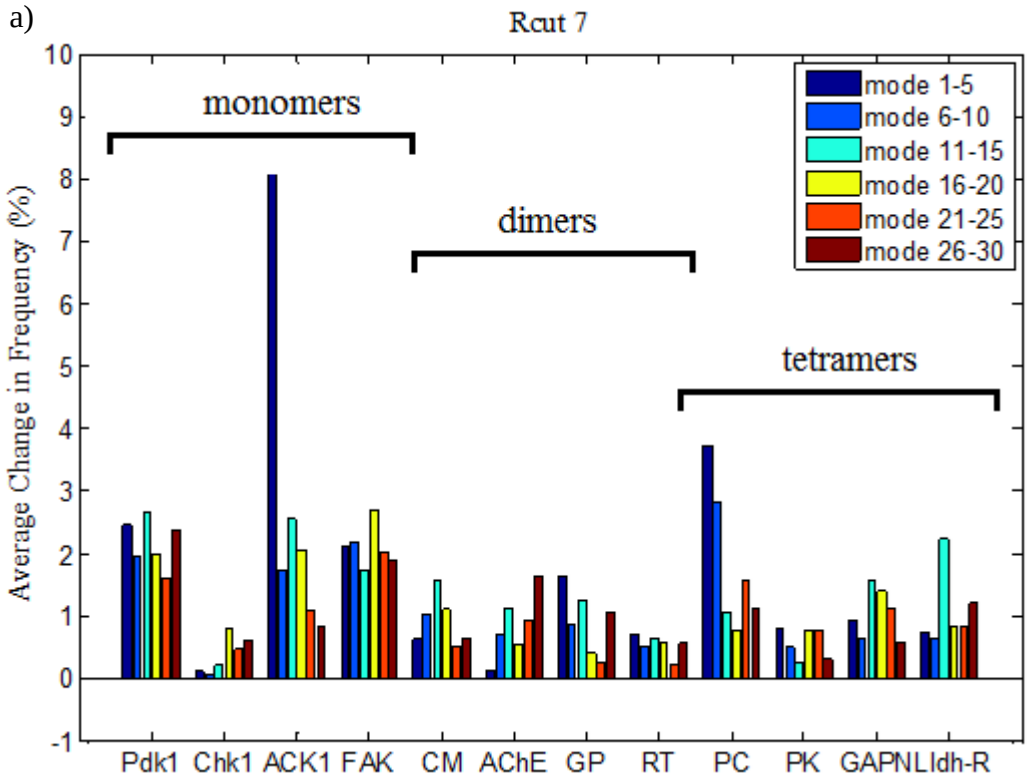


Figure 4.42. Average frequency percentage distribution due to allosteric ligand binding for cutoffs 7 Å (a) all modes (b) first five and 30 modes.

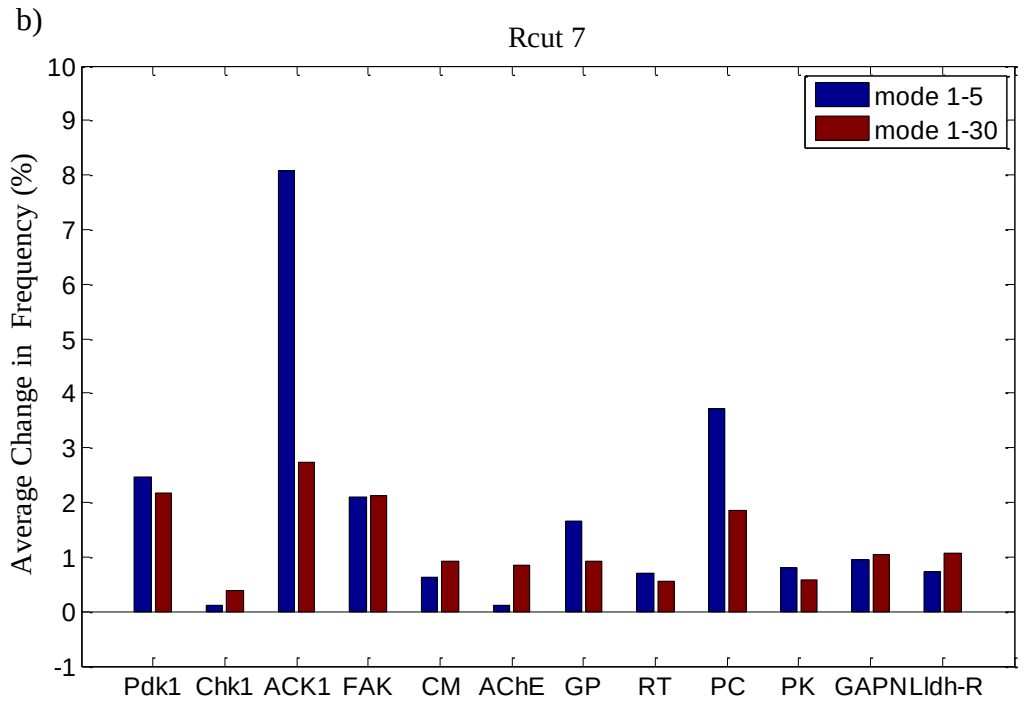


Figure 4.42. Average frequency percentage distribution due to allosteric ligand binding for cutoffs 7 Å (a) all modes (b) first five and 30 modes cont.

Table 4.6. Overlap Values of all Enzymes with 7 and 10 Å cutoffs.

Enzyme	Overlap Value		Diagonal-Overlap Value	
	Ligand bound-unbound R_{cut}=7	Ligand bound-unbound R_{cut}=10	Ligand bound-unbound R_{cut}=7	Ligand bound-unbound R_{cut}=10
Chk1	0.990	0.985	0.895	0.959
Pdk1	0.904	0.909	0.657	0.687
FAK	0.982	0.939	0.767	0.552
ACK1	0.978	0.994	0.790	0.882
CM	0.996	0.998	0.881	0.963
AChE	0.991	0.996	0.938	0.948

Table 4.6. Overlap Values of all Enzymes with 7 and 10 Å cutoffs cont.

GP	0.997	0.998	0.885	0.982
RT_1	0.989	0.999	0.978	0.996
RT_2	0.999	0.984	0.966	0.981
PC	0.998	0.998	0.974	0.960
PK_1	0.998	0.981	0.971	0.946
PK_2	0.997	0.994	0.990	0.960
$GAPN_1$	0.999	0.999	0.952	0.808
$GAPN_2$	0.999	0.999	0.965	0.958
$GAPN_3$	0.999	0.999	0.962	0.891
Lldh-T	0.980	0.997	0.883	0.831
Lldh-R	0.995	0.996	0.917	0.916

Table 4.7. Hinge residues of all enzymes considering monomeric structures.

Enzyme	Slowest mode1 Hinge residues	Slowest mode2 Hinge residues
Chk1	91	12, 92, 112,190
Pdk1	165	111, 157, 222, 332
FAK	506, 550, 567	499, 567, 642
ACK1	212	204, 270, 350
CM	22, 71, 114, 154, 199, 237	15, 38, 118, 134, 167
AChE	312, 450, 497	35, 99, 148, 179, 202, 315, 337, 398

Table 4.7. Hinge residues of all enzymes considering monomeric structures.

GP	482, 812	193, 225, 243, 484, 545, 797
RT_1	229	90, 111, 167, 424
RT_2	237	89, 114, 163, 424
PC	493	375, 547, 565, 642, 824, 852, 984
PK_1	84, 195, 239, 279	20, 89, 184, 356
PK_2	113, 227, 269, 309	44, 119, 211, 382
$GAPN_1$	125, 161, 474	135, 152, **263**, 470, 487
$GAPN_2$	125, 161, 474	135, 152, **263**, 470, 487
$GAPN_3$	63, 79, 125, 161, 474	135, 152, **263**, 470, 487
Lldh-T	148, 230, 264, 310	**82**, 154, 174, 190, 255
Lldh-R	149, 235, 290, 309	83,154,220,236,253

Finally, Table 4.8 summarizes the average percentage change in mode frequencies due to ligand binding (catalytic, allosteric or coenzyme) for the monomeric, dimeric and tetrameric enzymes with 7 Å. When catalytic and allosteric ligands are considered, first 5 modes are highly affected in monomeric enzymes. For dimer and tetrameric enzymes, the effect is less pronounced and very similar when either 5 or 30 modes are considered. Frequency change due to coenzyme binding seems highly effective in tetramers, which will be explained in detail in Section 4.4.

Table 4.8. Average percentage change in mode frequencies due to ligand binding with 7 Å.

	Catalytic ligands		Allosteric ligands		Coenzymes	
	1-5 modes	1-30 modes	1-5 modes	1-30 modes	1-5 modes	1-30 modes
Monomers	2.58	1.70	3.19	1.85	-	-
Dimers	1.19	1.01	0.78	0.80	0.22	0.37
Tetramers	0.79	1.07	1.55	1.13	2.85	3.40
All	1.53	1.21	1.84	1.26	2.41	2.90

Table 4.9. Average percentage change in mode frequencies due to ligand binding with 10 Å.

	Catalytic ligands		Allosteric ligands		Coenzymes	
	1-5 modes	1-30 modes	1-5 modes	1-30 modes	1-5 modes	1-30 modes
Monomers	1.70	0.84	0.74	0.77	-	-
Dimers	0.80	0.66	1.02	0.66	0.12	0.17
Tetramers	0.69	0.70	0.52	0.61	1.69	2.16
All	1.04	0.72	0.76	0.68	1.43	1.83

4.3.5. Effect of Allosteric Ligands on Catalytic Residue Dynamics

It is further analyzed whether the allosteric ligands cause any change in the fluctuation of catalytic residues. The catalytic residues of six enzymes in the dataset could be determined from the catalytic atlas (Porter *et al.*, 2004). The dot product between fluctuation vectors calculated in the presence and absence of the allosteric ligand is calculated for each residue in each collective mode. The vectors are normalized so that orientational correlation is reported in terms of cosine of the angle between the unit vectors for each residue. For each mode, an average orientational correlation over all residues is calculated to observe the overall change in the fluctuation of residues. If mean orientational correlation values over all residues for a specific mode is close to either 1 or -1, this represents that the harmonic mode shape has not changed. If orientational correlation of a catalytic residue is different from 1 or -1, it indicates that the catalytic residue's orientation has changed due to interactions with allosteric ligand.

Orientational correlation versus mode index is plotted in Figures 4.43-4.48 for catalytic residues and mean of all residues. Monomeric enzyme Pdk1 exhibits higher effect of allosteric ligand on its catalytic residues even in slowest modes in including 4^{th} and higher indexed ones. The other monomer Chk1 and the dimers indicate changes in modes with indices above 10. The least amount of change is observed in the tetramers.

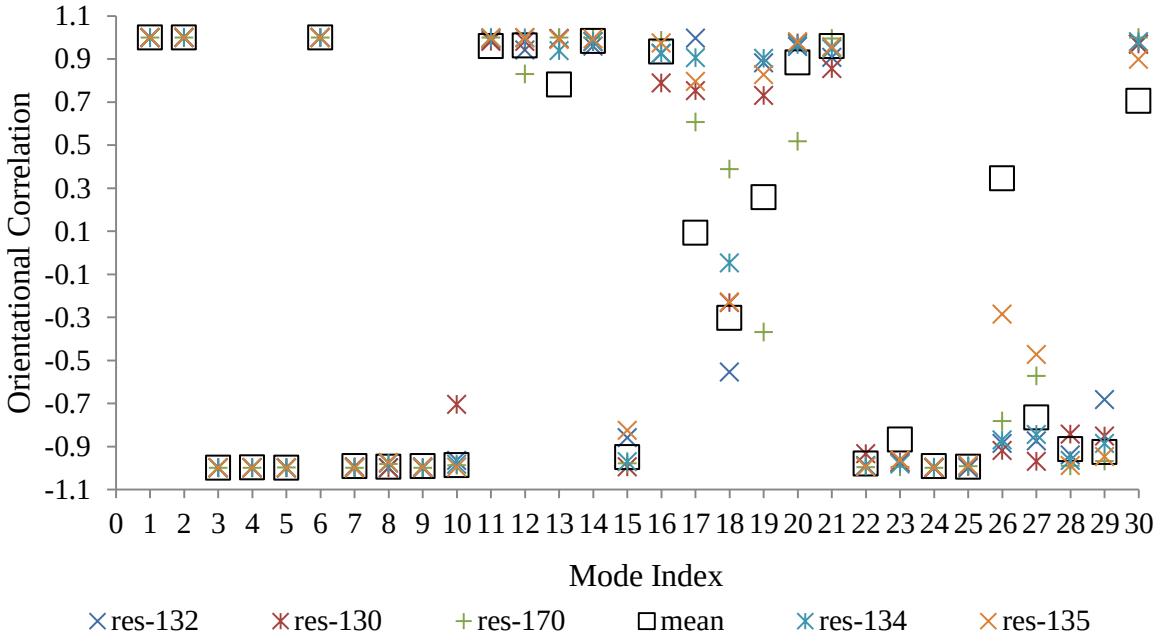


Figure 4.43. Orientational correlation of residues upon ligand binding for Chk1 monomer.

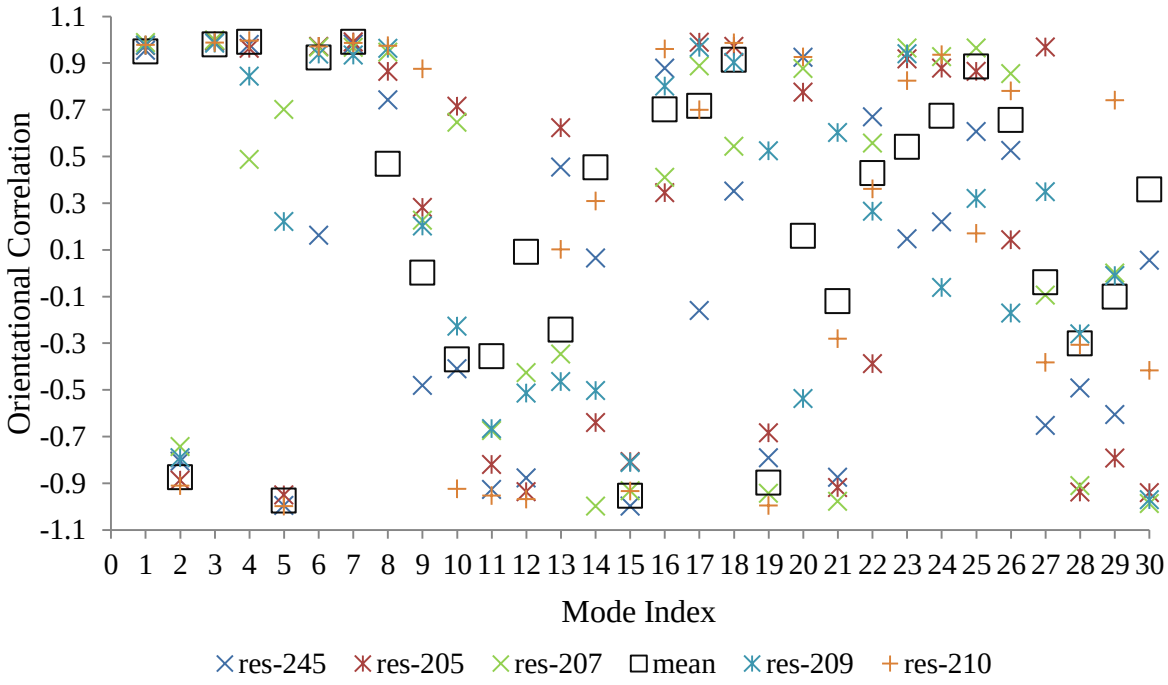


Figure 4.44. Orientational correlation of residues upon ligand binding for Pdk1 monomer.

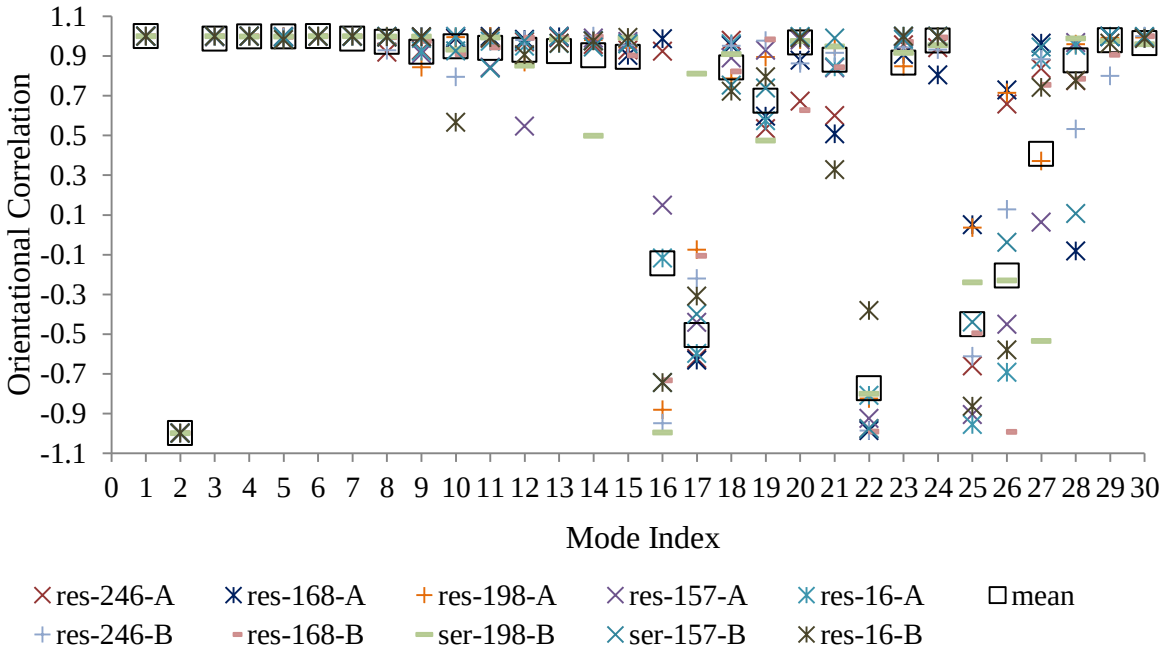


Figure 4.45. Orientational correlation of residues upon ligand binding for CM dimer.

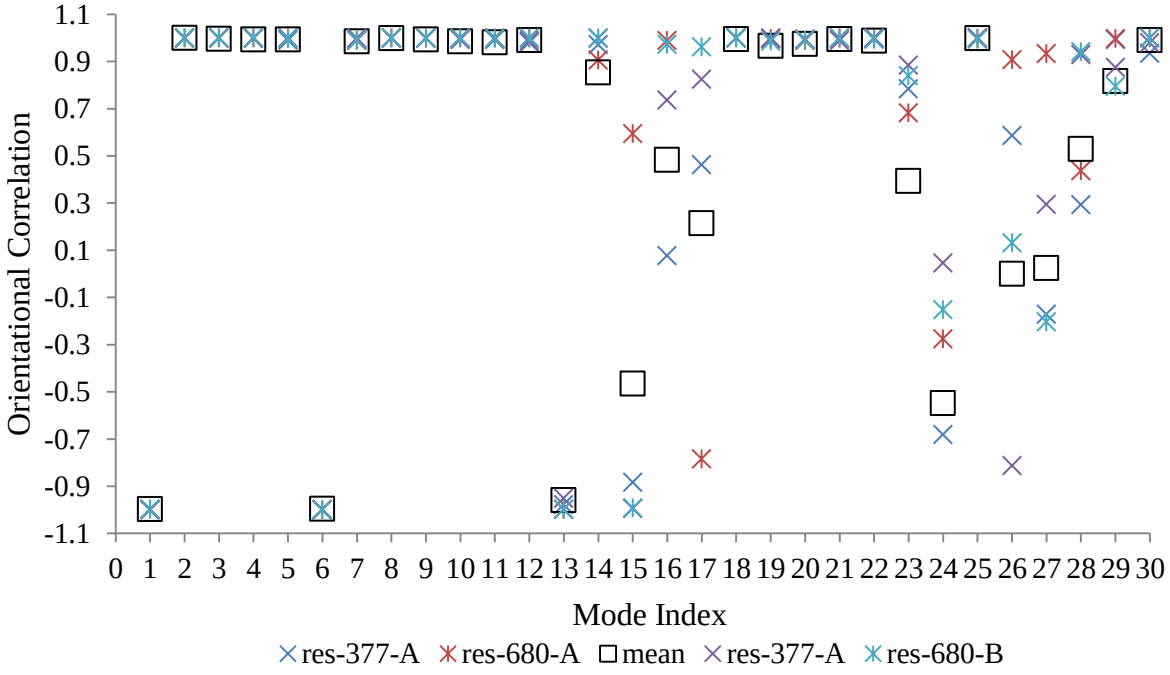


Figure 4.46. Orientational correlation of residues upon ligand binding for GP dimer.

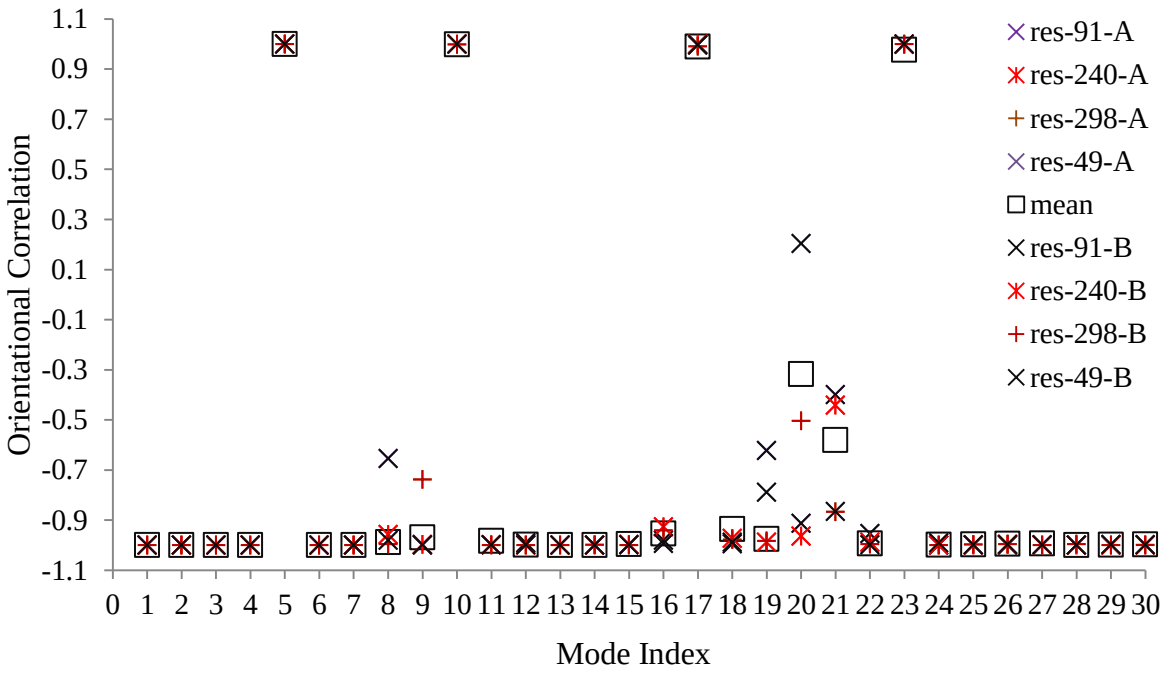


Figure 4.47. Orientational correlation of residues upon ligand binding for PK_1 tetramer.

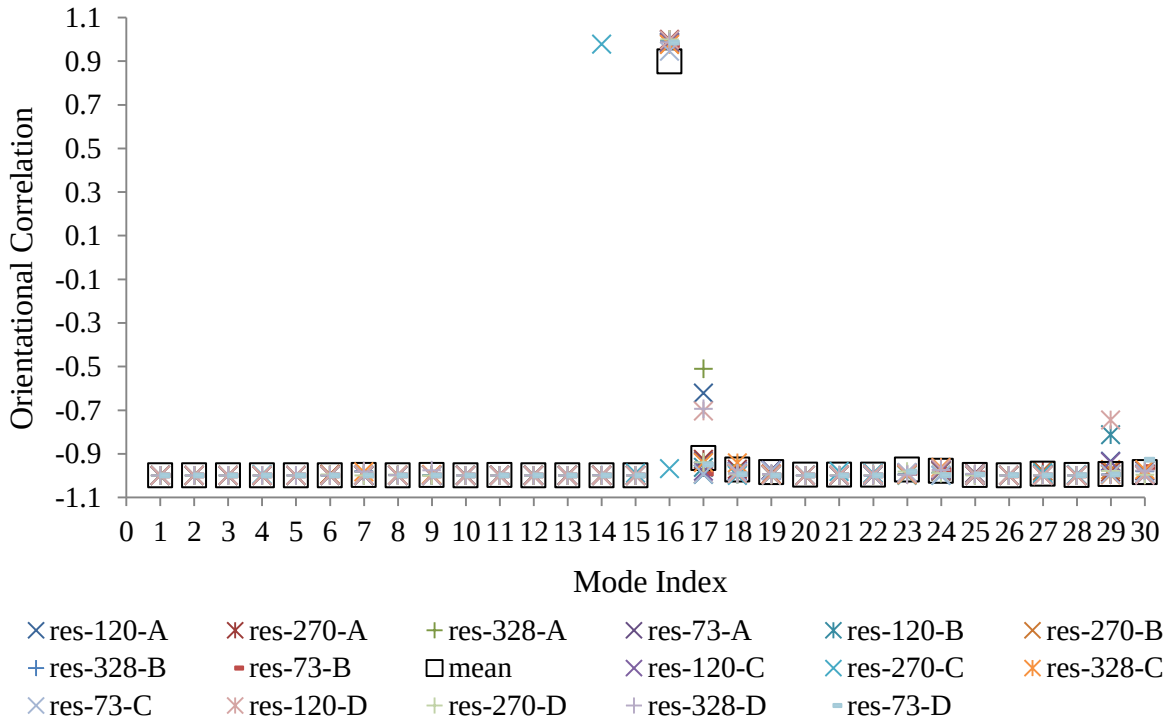


Figure 4.48. Orientational correlation of residues upon ligand binding for PK_2 tetramer.

4.4. Effects of Coenzymes and Cofactors on Enzyme Dynamics

Coenzymes and cofactors are organic and inorganic substances that are needed for the catalysis of enzymatic reactions. In this dataset, the enzymes that have either cofactors or coenzymes are listed in Table 4.10. Figure 4.49 indicate the effect of cofactors on frequencies, whereas Figures 4.52 and 4.53 give the effect of coenzymes. Cofactors, either Mn^{2+} or Mg^{2+}, lead to negligible changes in frequencies. In contrast, coenzymes NAP and NAD cause significant shifts in the frequencies to higher values. NAP and NAD are larger ligands compared to the third coenzyme PLP that has a relatively insignificant effect. Moreover, NAP and NAD cause even larger changes than the corresponding allosteric ligands. One of the 2^{nd} slowest mode hinge residues are in contact with the coenzyme for GAPN and Lldh, which are shown in bold in Table 4.7. They also seem to have a quite uniform effect in the range 1-30 modes, when the averages are considered in Figure 4.52.

Table 4.10. Coenzymes and cofactors in the dataset

Enzyme	Cofactor	Coenzyme	% coenyzme atoms per residue	Pocket Area (Å^2), Volume (Å^3)
GP	-	$C_8H_{15}NO_6$ (x2) (PLP)	1.90	1292.4 1473.7
RT_2	Mg^{2+}	-	-	-
PC	Mn^{2+}	-	-	-
PK_1	Mn^{2+}	-	-	-
PK_2	Mg^{2+}	-	-	-
$GAPN_1$, $GAPN_2$, $GAPN_3$	-	$C_{21}H_{28}N_7O_{17}P_3$ (x4) (NAP)	9.62	4859.9 7429.2
Lldh-R	-	$C_{21}H_{27}N_7O_{14}P_2$(x4) (NAD)	14.06	553.4 718.7
Lldh-T	-	$C_{21}H_{27}N_7O_{14}P_2$(x4) (NAD)	14.06	3513.1 5352.6

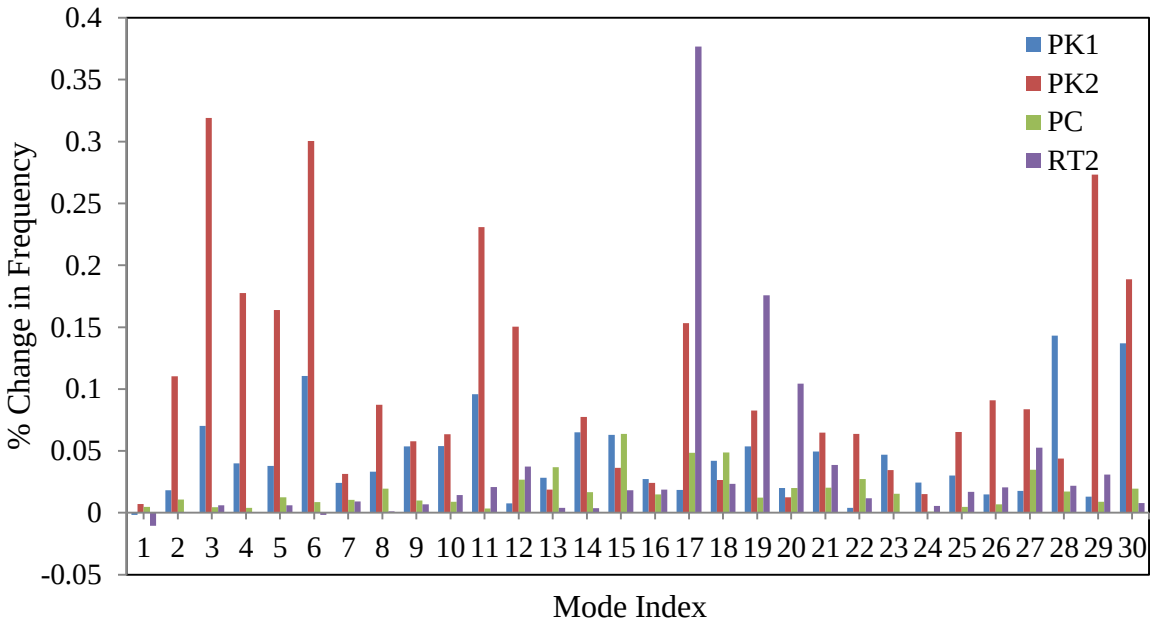


Figure 4.49. Percentage change in mode frequencies due to cofactor binding with 7 Å cutoff.

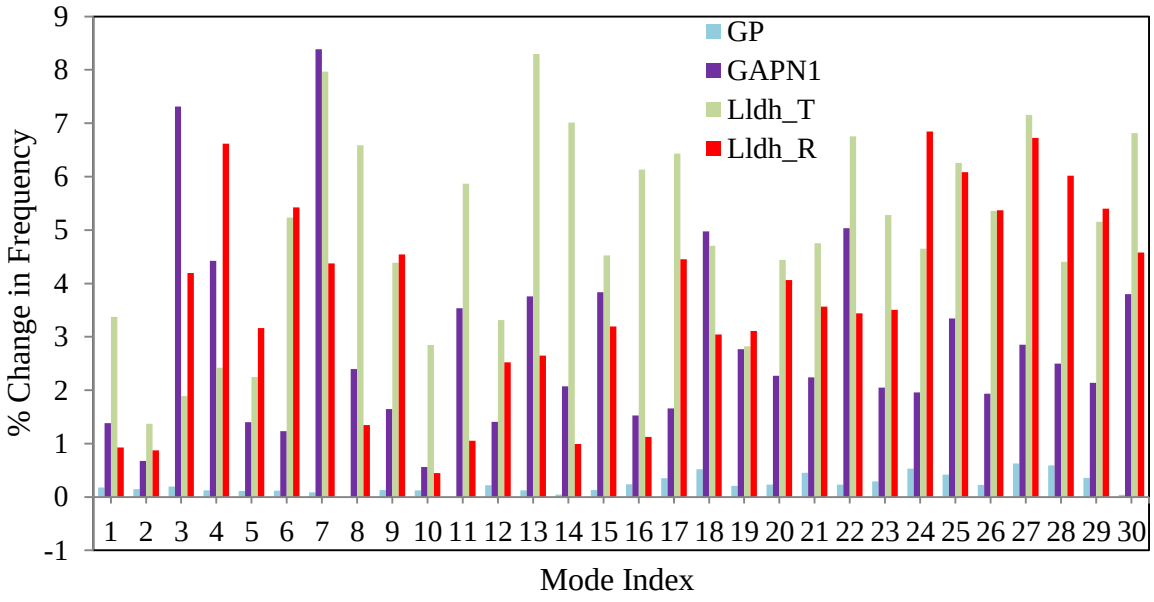


Figure 4.50. Percentage change in mode frequencies due to coenzyme binding with 7 Å cutoff.

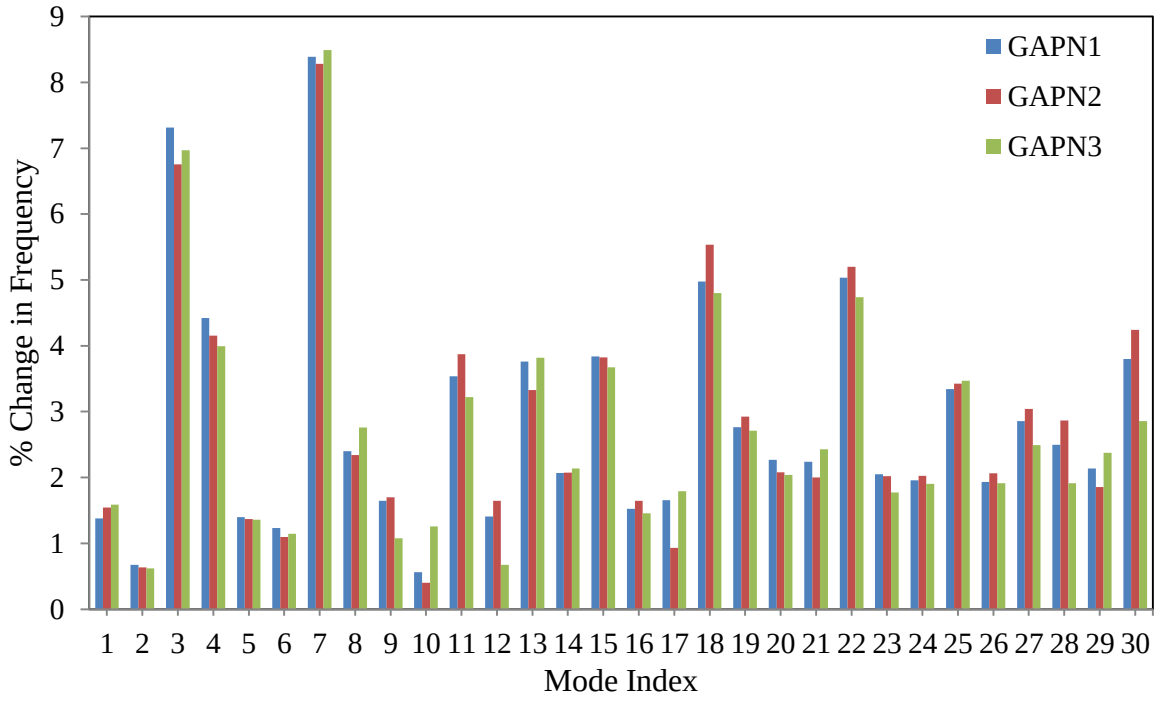


Figure 4.51. Percentage change in mode frequencies due to coenzyme binding to GAPNs with 7 Å cutoff.

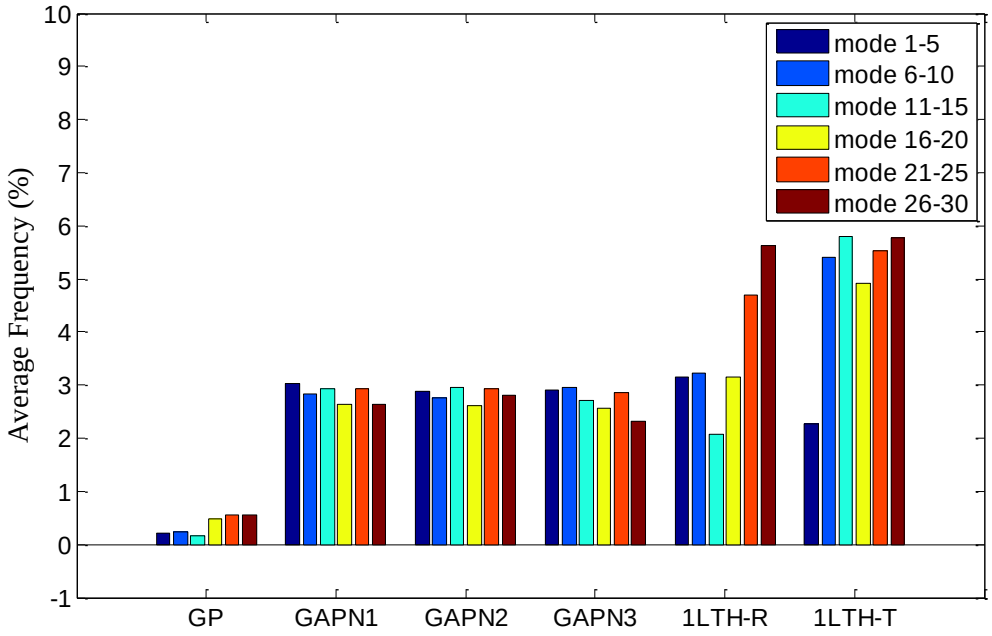


Figure 4.52. Average frequency percentage distribution due to coenzyme binding with cutoff 7 Å.

5. CONCLUSION

Elastic network model with mixed coarse-graining approach, named as MCG_ANM (Kurkcuoglu *et al.*, 2009), is applied to more than 30 crystal structures including different monomeric and multimeric enzymes. The ligands are considered at atomic detail, i.e. at high resolution, whereas the protein is at low resolution with only alpha-carbons considered. The effect of solvents, catalytic ligands, catalytic loop conformational changes, allosteric ligands and coenzymes are discussed in terms of their effect on vibrational dynamics of mainly the first 30 collective modes of the enzymes. The frequencies of slow modes are found to shift to higher values in most of the cases studied.

Three of the enzymes studied have solvent-bound crystal structures. For TIM and elastase, first five modes are highly affected due to solvent binding than the remaining slow modes. Solvents bound to catalytic sites and the functional interface region in TIM is found more effective in shifting frequencies. In literature, methodologies based solvents scanning are used to determine possible binding sites of a protein either experimentally or computationally (Mattos *et al.*, 2006). The extent of frequency shifts may be used to determine among the different poses of the solvent the functionally relevant ones.

There are seven enzymes with catalytic loop motions taken from a previous study (Kurkcuoglu *et al.*, 2012). The effect of conformational change from open to closed loop dominates the frequency shifts in collective modes compared to the ligand binding effect on the closed conformer. An average change of 6.8% in frequencies of the first five modes is observed due to loop closure, whereas the change is 1.5% due to ligand binding to the closed state. Conformational changes also lead to swapping of modes and changes in mode character to a large extent. The monomeric enzyme bGT exhibits the highest change in frequencies due to ligand binding, which can be explained by its relatively large ligand and pocket sizes.

All calculations are performed with atomistic cutoff distances of 7 and 10 Å. As a result of extensive calculations on catalytic and allosteric ligands, both cutoffs seem suitable for MCG_ANM calculations, but the relatively smaller cutoff 7 Å leads to larger

percentages in frequency shifts in general. In fact 7 Å seems a reasonable value, when considered as an atomistic cutoff in MCG_ANM calculations.

To investigate the effect of allosteric ligands on enzyme collective dynamics, MCG_ANM is applied on 17 allosteric proteins taken from ASD (Huang *et al.*, 2014). Results show that vibrational frequencies of monomeric enzymes excluding Chk1 are more affected than dimers and tetramers for cutoff 7 Å. Moreover, mode swapping and character change is observed dominantly for these monomers. Specifically, allosteric ligand binding to monomeric, dimeric and tetrameric enzymes leads to 3.2%, 0.8% and 1.6% changes, respectively, averaged over first five modes. The fact that ratio of ligand atoms per residue number is highest for monomers may be a reason for these observations. Moreover, for these monomeric enzymes and the tetrameric PC, which also exhibits a large change in first five mode frequencies, the allosteric ligand is in contact with a hinge residue. The binding of allosteric ligands to monomers and dimers also lead to significant changes in the orientational correlations of catalytic residues. Finally, coenzymes bound to tetramers lead to higher frequency shifts compared to allosteric ligands and cofactors, which may similarly be associated with their large ligands that are in contact with hinges.

APPENDIX A: Graphs of Catalytic Enzymes

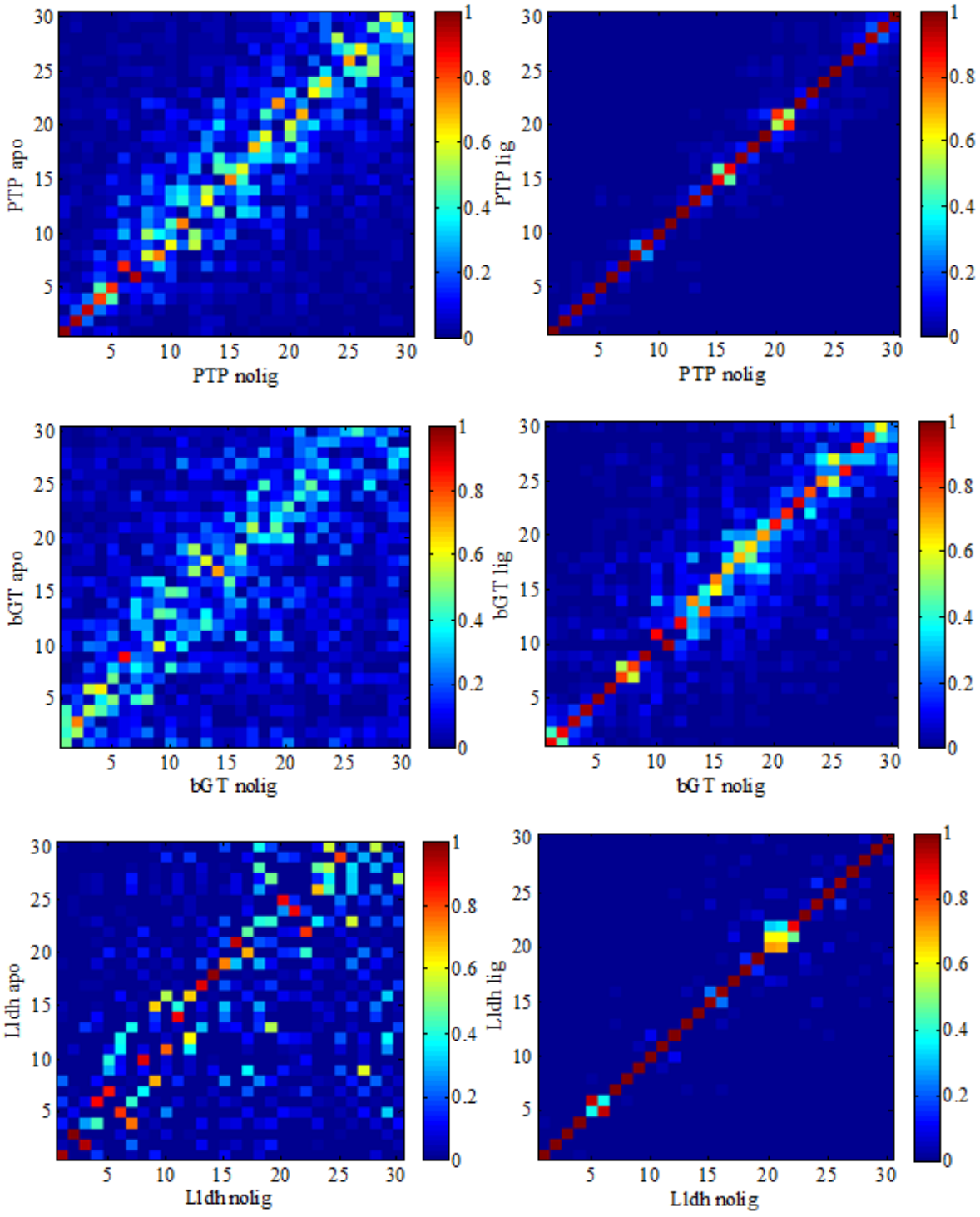


Figure A.1. Overlap matrices for all dataset with 10 Å cutoff.

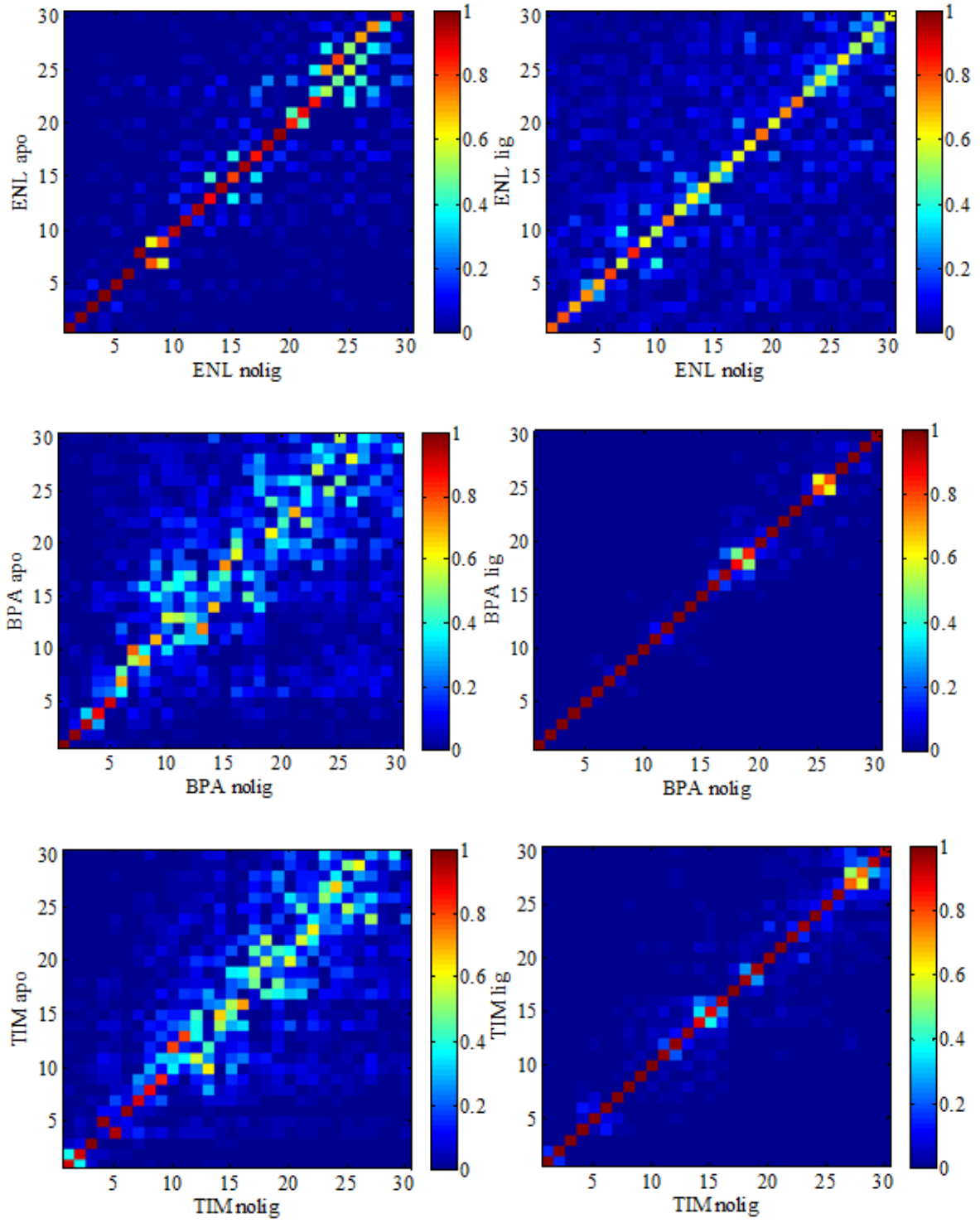


Figure A.2. Overlap matrices for all dataset with 10 Å cutoff.

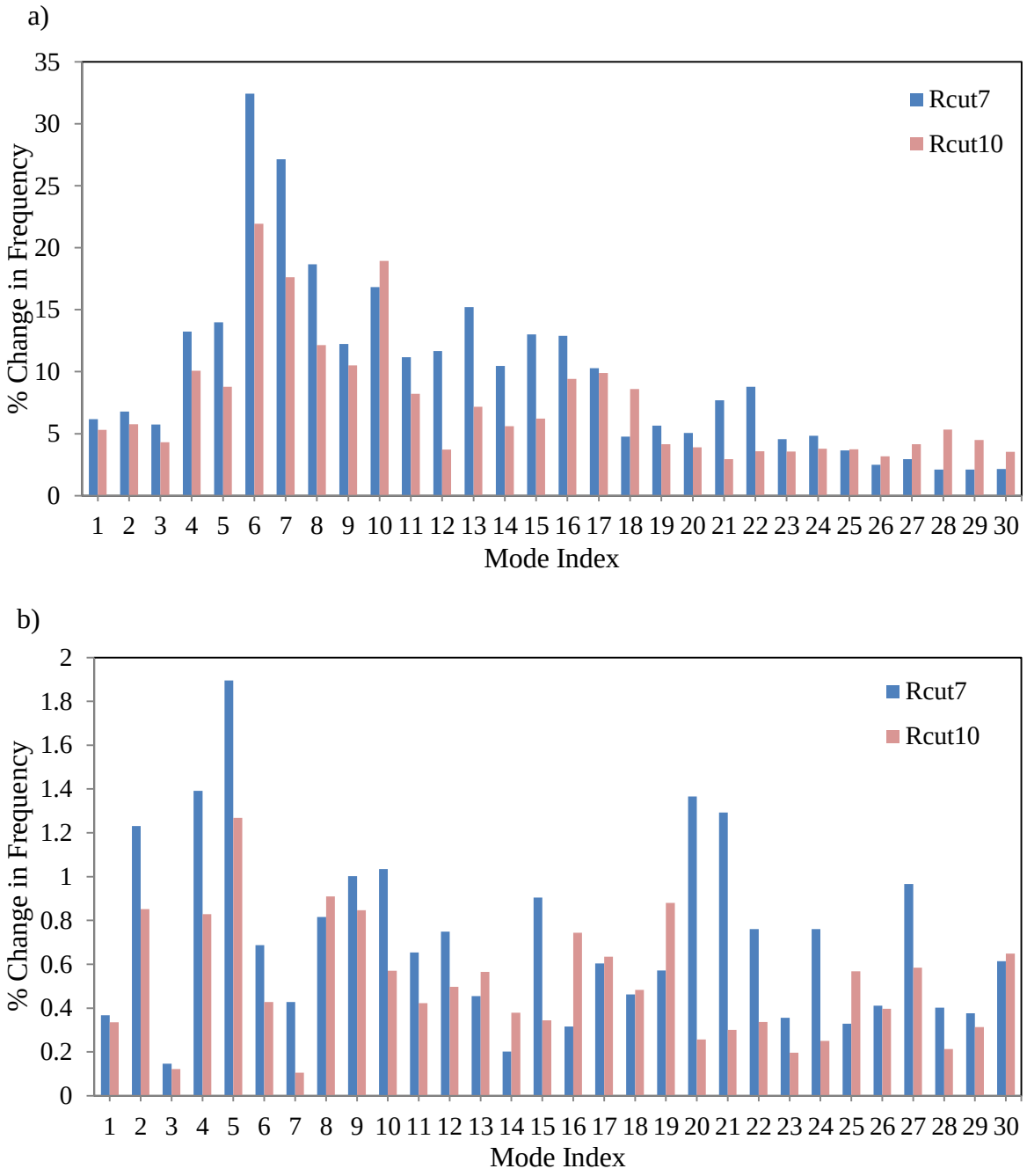


Figure A.3. Percentage change in mode frequencies due to (a) conformational change (b) ligand binding to BPA in bound conformation.

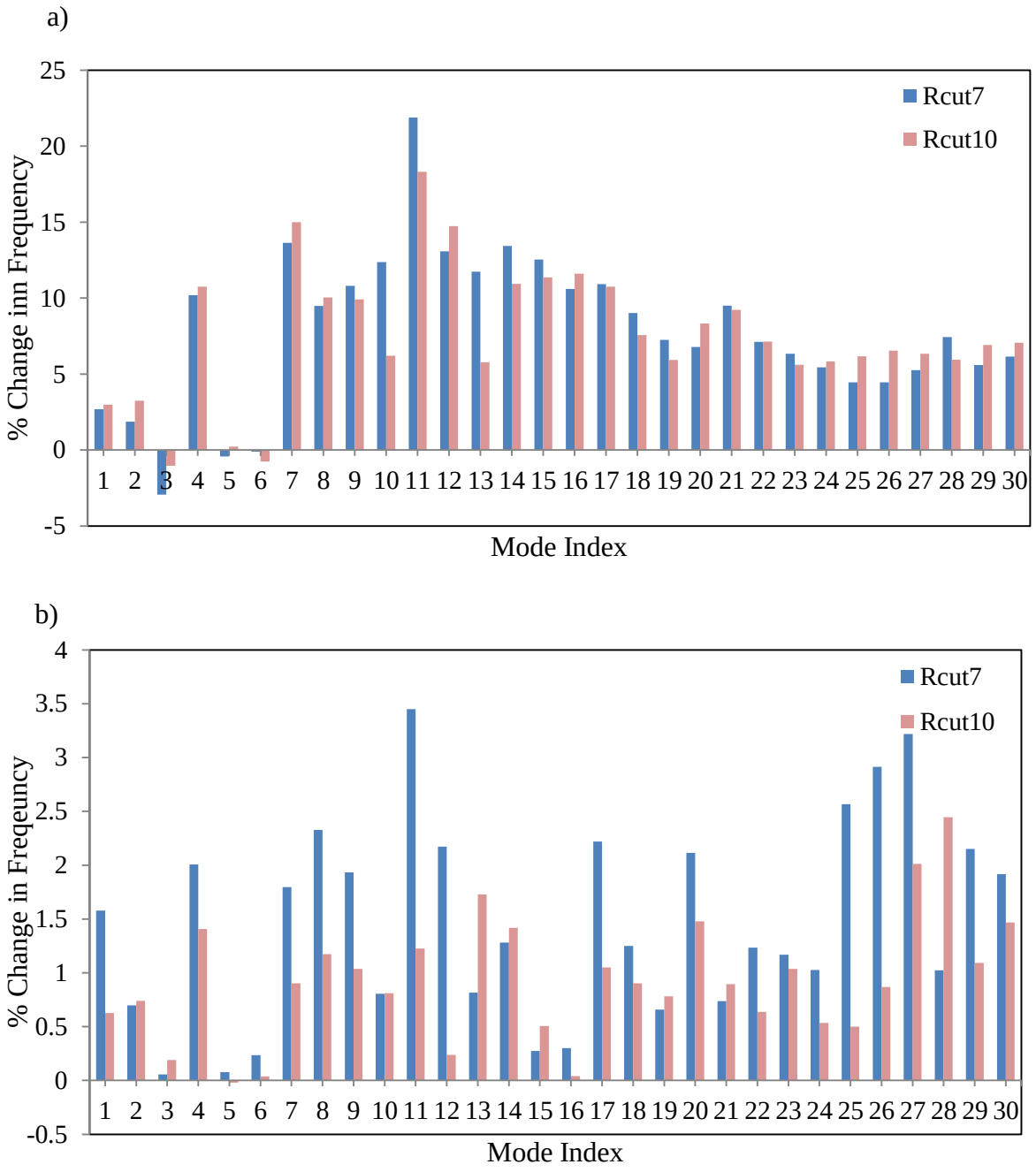


Figure A.4. Percentage change in mode frequencies due to (a) conformational change (b) ligand binding to 3dhq in bound conformation.

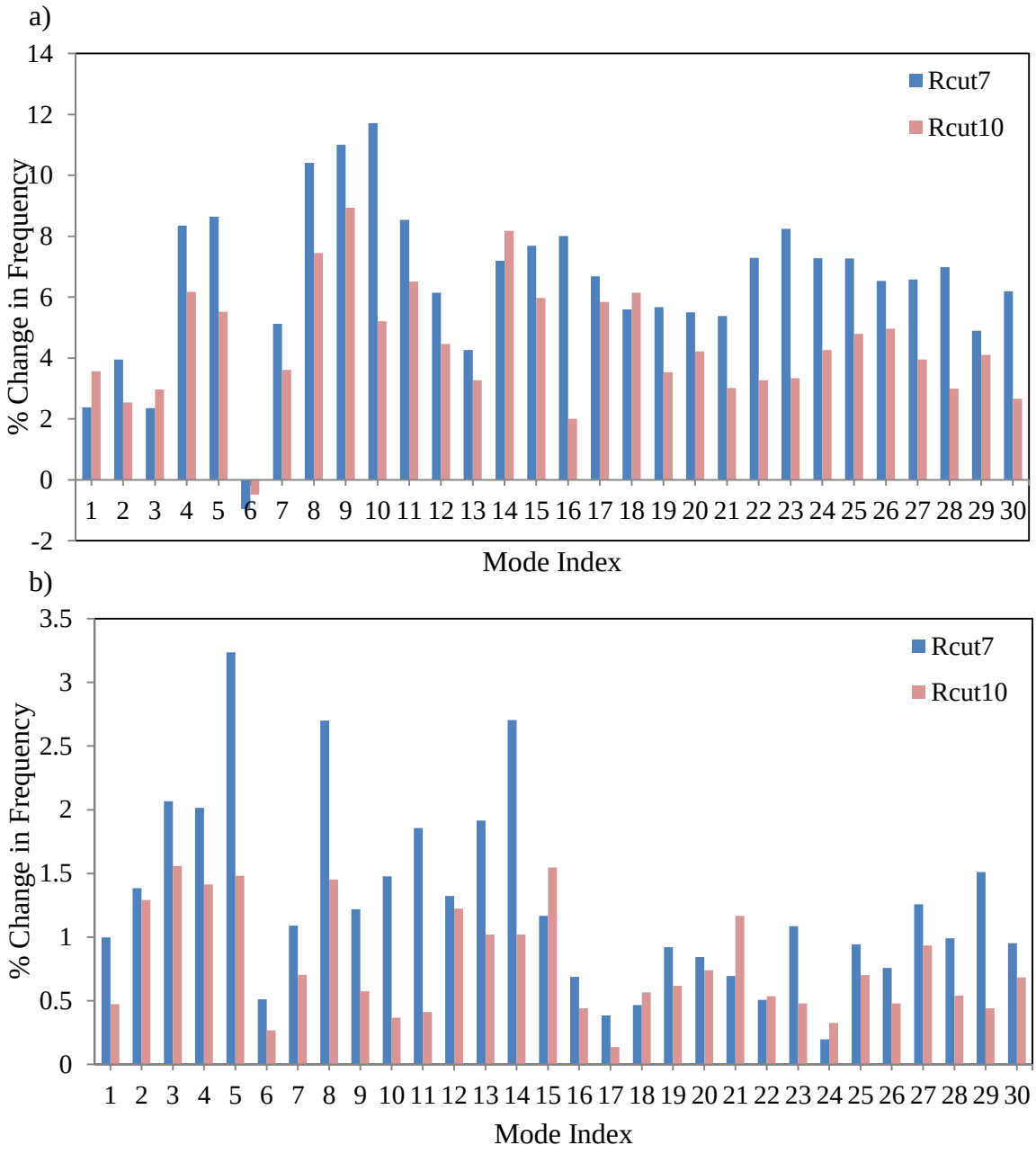


Figure A.5. Percentage change in mode frequencies due to (a) conformational change (b) ligand binding to TIM in bound conformation.

Table A.1. Overlap values of enzymes with 10 Å cutoff distances.

Enzyme	Overlap Value		Overlap- diagonal only	
	Apo-ligand removed	Ligand bound-unbound	Apo-ligand removed	Ligand bound-unbound
PTP	0.96	0.99	0.53	0.96

Table A.1. Overlap values of enzymes with 10 Å cutoff distances cont.

bGT	0.86	0.96	0.25	0.60
BPA	0.88	0.99	0.46	0.97
Enl	0.76	0.95	0.78	0.65
TIM	0.92	0.99	0.52	0.96
3dhq	0.92	0.99	0.52	0.85
Lldh	0.95	0.99	0.31	0.94

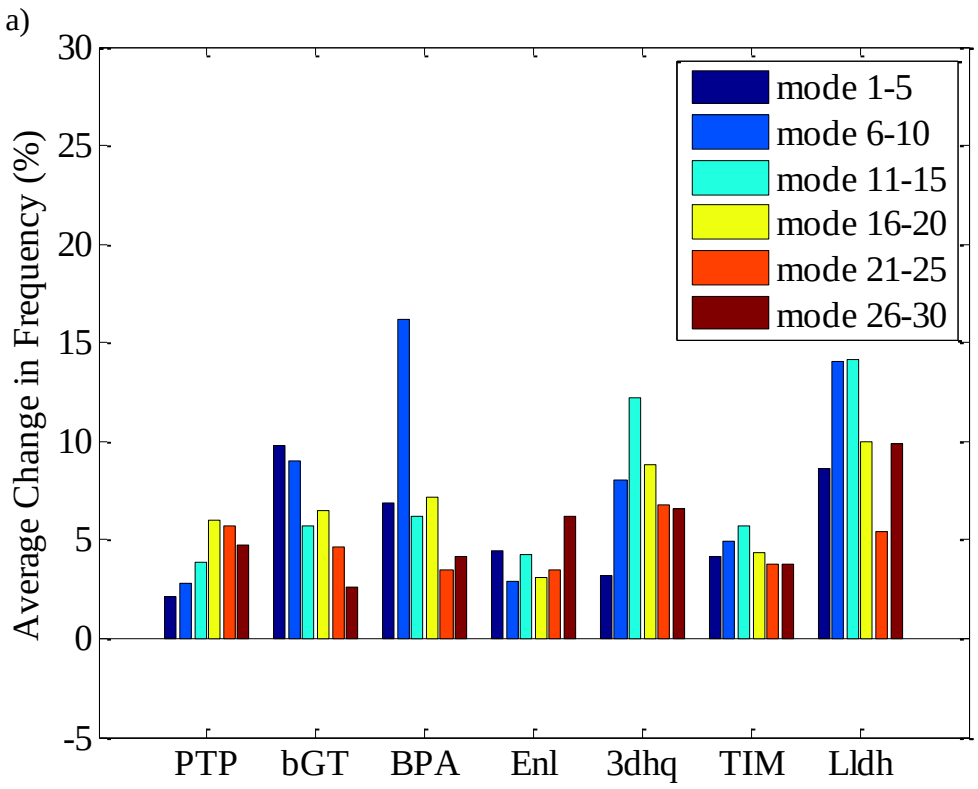


Figure A.6. Average frequency percentage distribution due to conformational change for cutoff for 10 Å (c) each 5 (d) 5 and 30 modes.

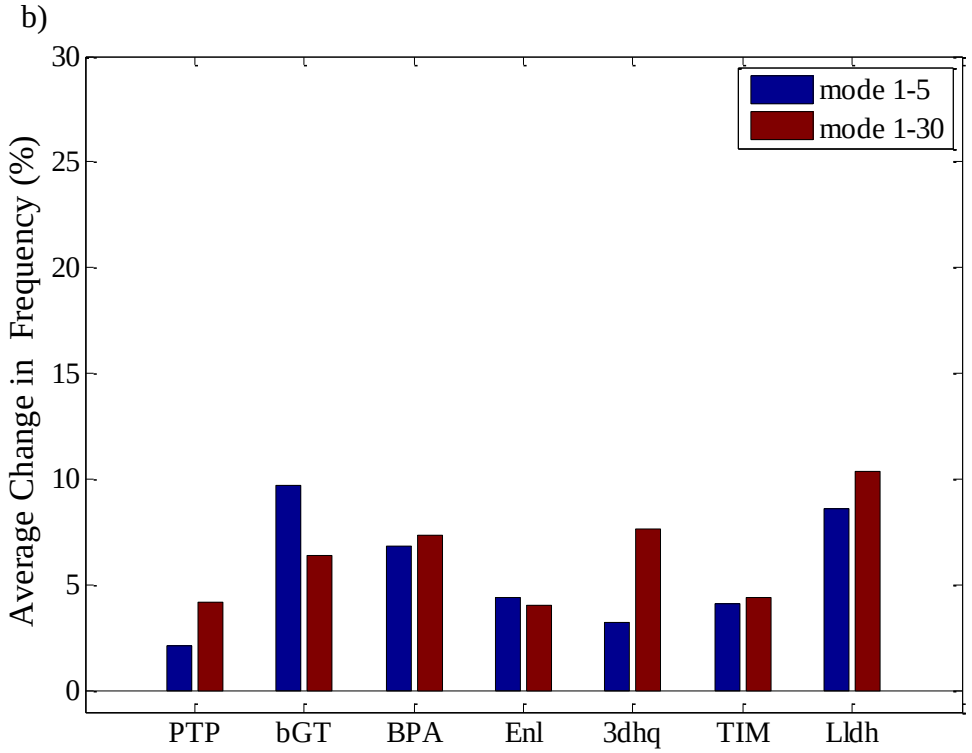


Figure A.6. Average frequency percentage distribution due to conformational change for cutoff for 10 Å (c) each 5 (d) 5 and 30 modes.

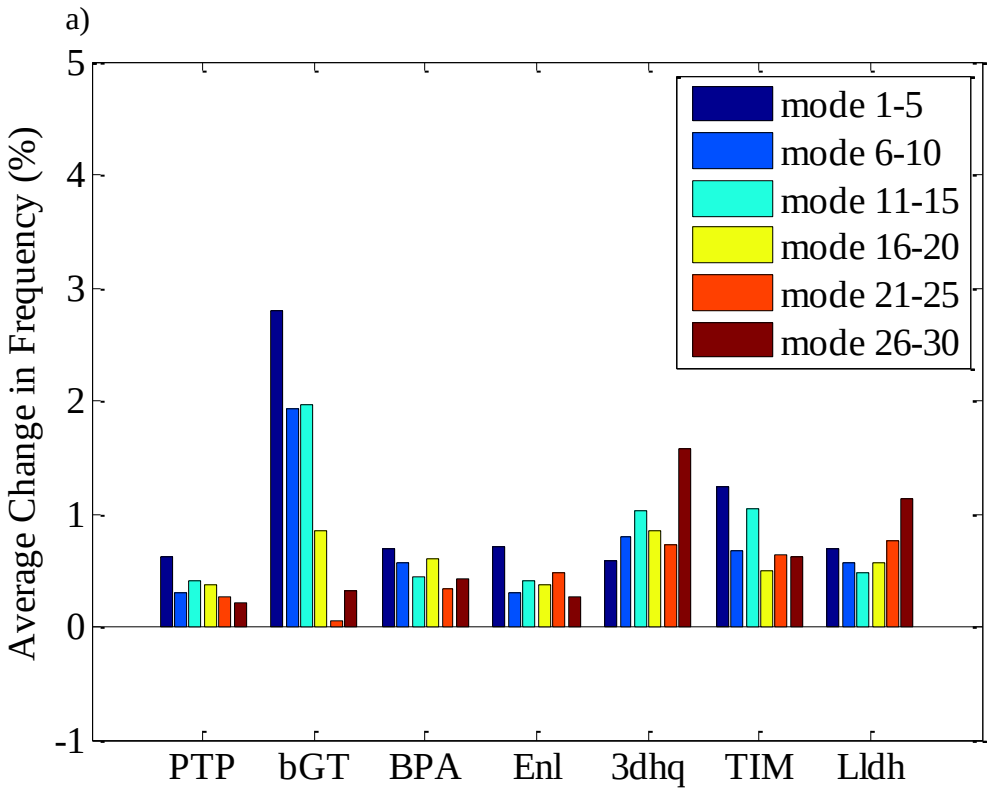


Figure A.7. Average frequency percentage distribution due to ligand binding for cutoff 10 Å for (a) each 5 (b) 5 and 30 modes.

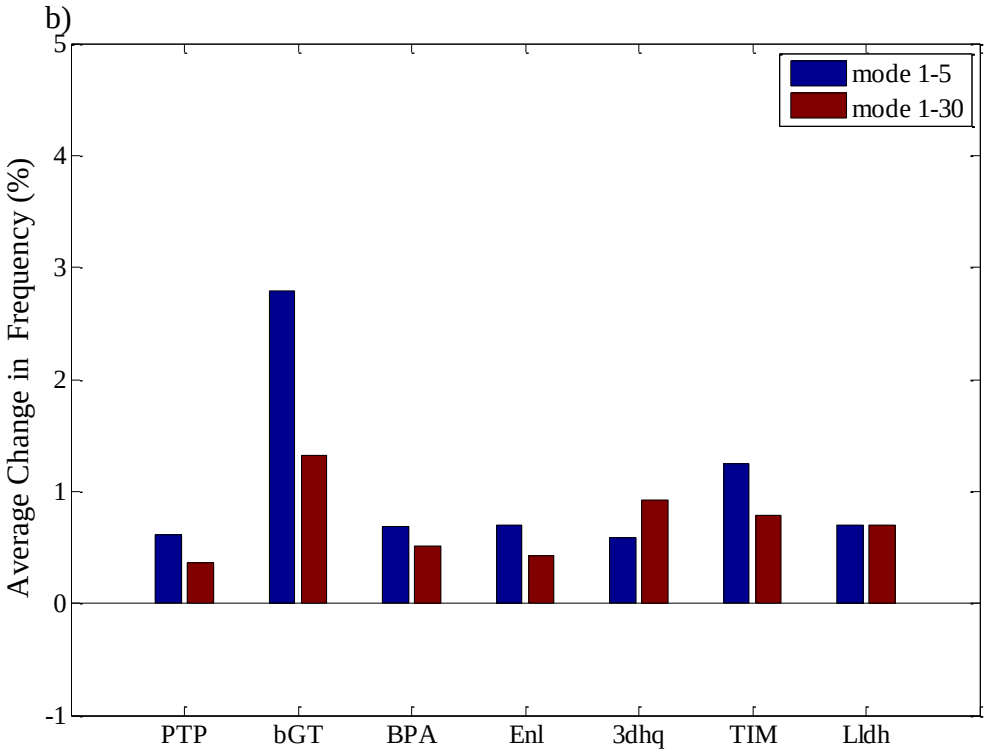


Figure A.7. Average frequency percentage distribution due to ligand binding for cutoff 10 Å for (a) each 5 (b) 5 and 30 modes cont.

Table A.2. Hinge residues of enzymes in dimeric and tetrameric forms.

Enzyme	Hinge Residues	
	Slowest mode1	Slowest mode 2
BPA	307 (A)	21, 61, 210, 307 (A) 21, 42, 210, 300 (B)
Enl	436 (A)	147, 343 (A) 148, 343 (B)
TIM	248 (1)	93, 229, 248 (1) 7, 93, 229 (2)
3dhq	251 (A)	18, 139 (A) 18, 139 (B)
Lldh	41, 65, 312(A) 6,41,64 (B) 7, 41, 64, 312 (C) 6, 41, 64 (D)	4 (B,C,D)

APPENDIX B: Graphs of Allosterics

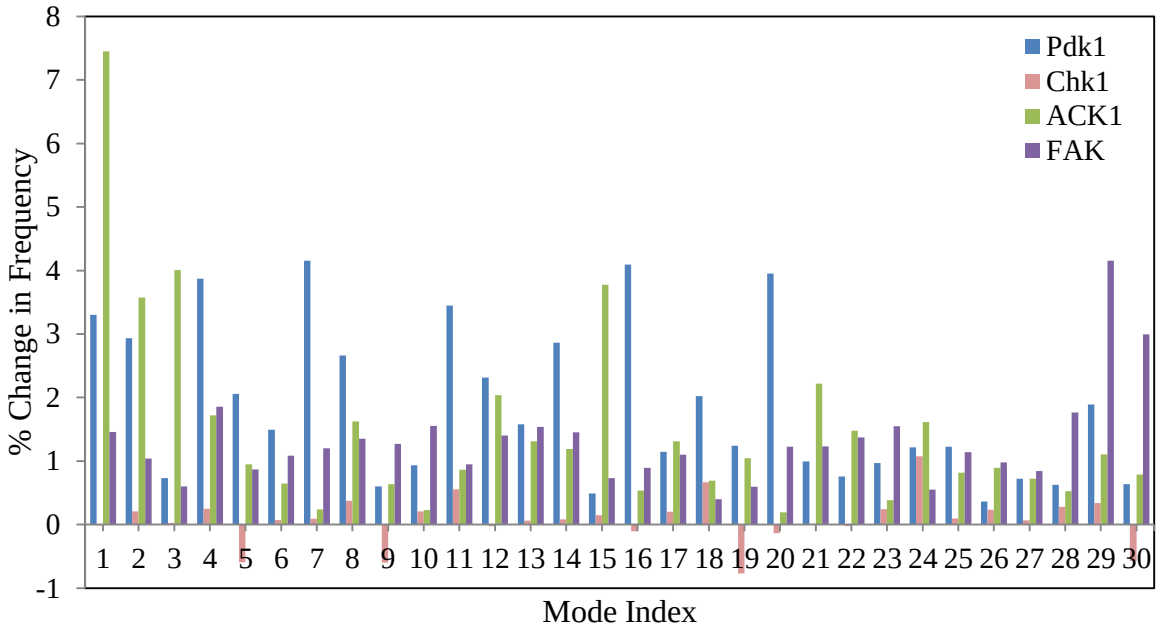


Figure B.1. Percentage change in mode frequencies due to allosteric ligand binding to monomeric enzymes with 10 Å cutoff.

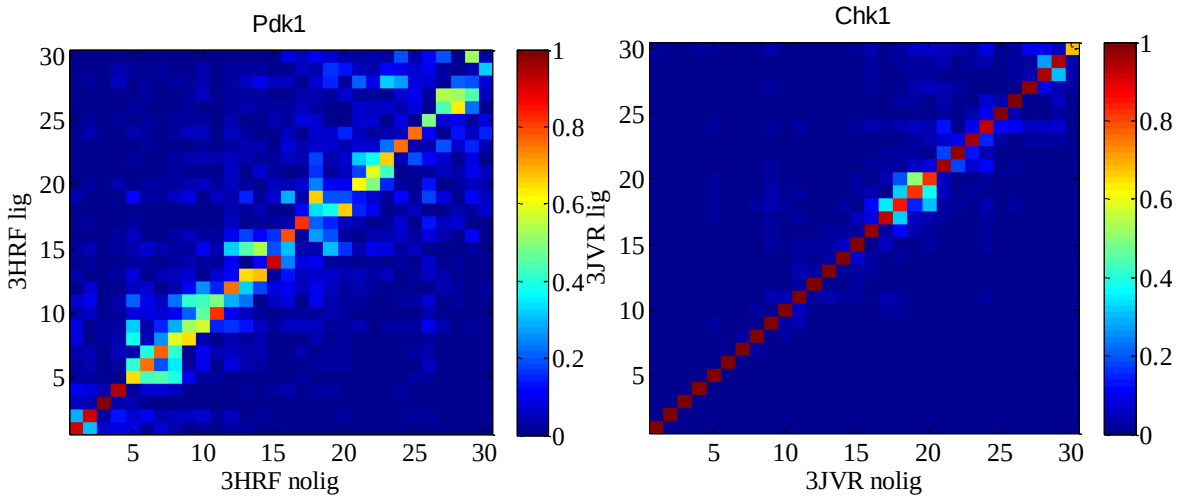


Figure B.2. Overlap matrices of monomer samples with 10 Å cutoff.

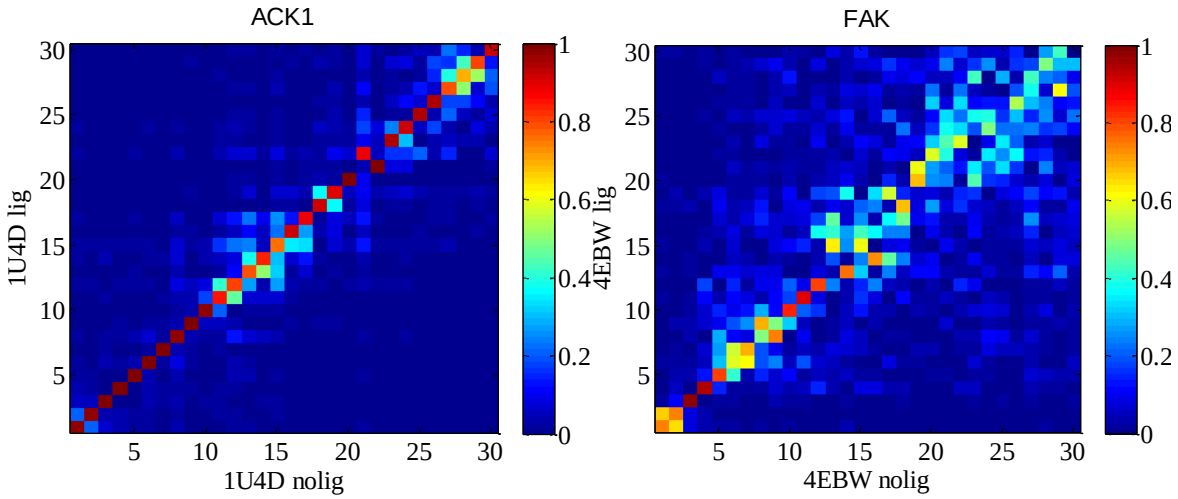


Figure B.2. Overlap matrices of monomer samples with 10 Å cutoff cont.

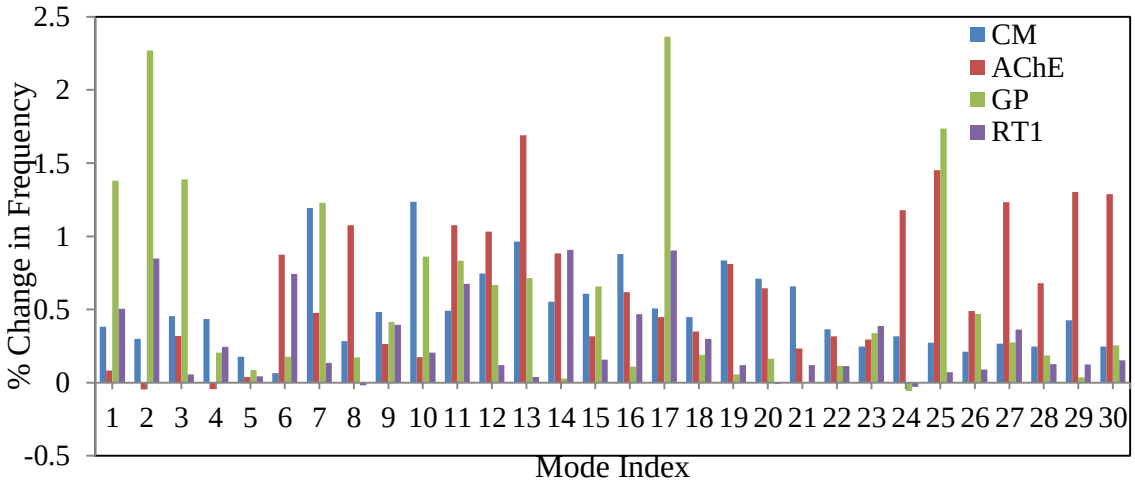


Figure B.3. Percentage change in mode frequencies due to allosteric ligand binding to dimeric enzymes with 10 Å cutoff.

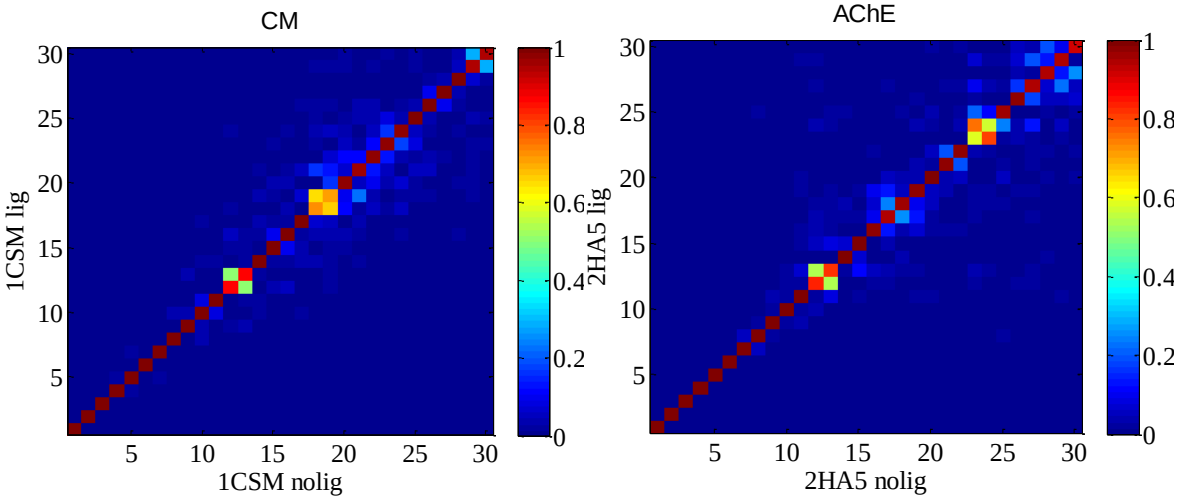


Figure B.4. Overlap matrices of dimer samples with 10 Å cutoff cont.

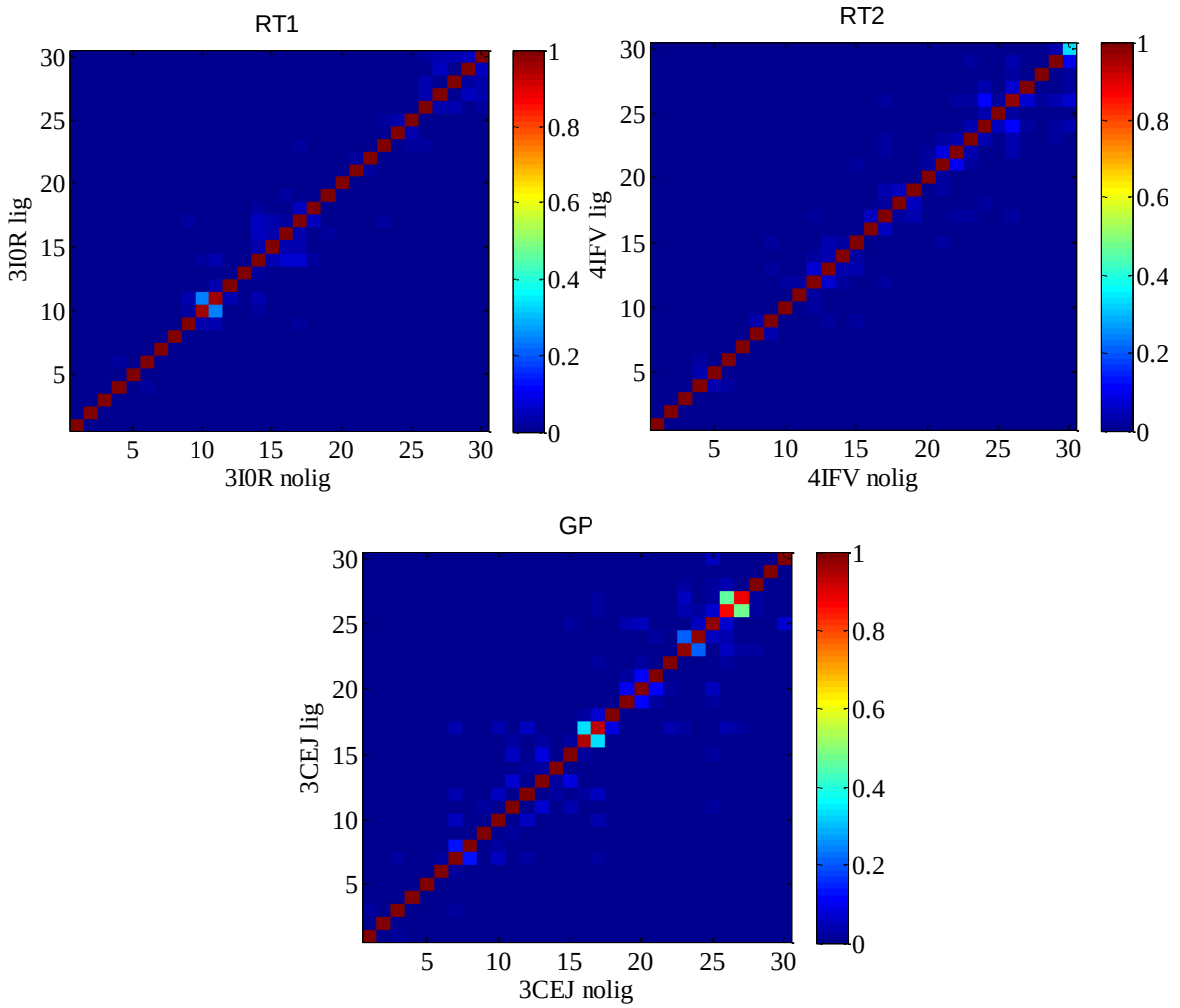


Figure B.4. Overlap matrices of dimer samples with 10 Å cutoff cont.

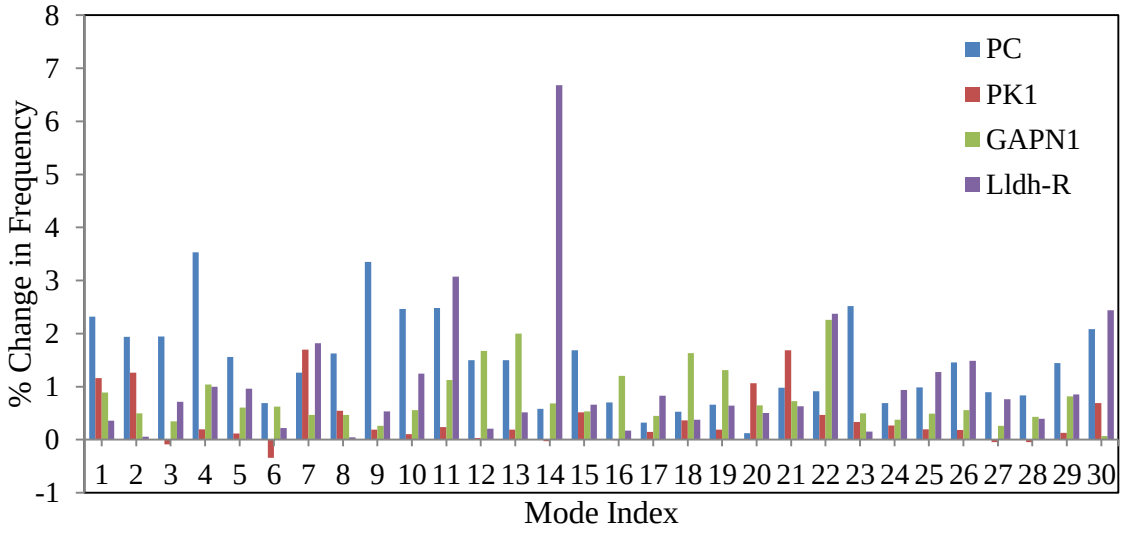


Figure B.5. Percentage change in mode frequencies due to allosteric ligand binding to tetrameric enzymes with 10 Å cutoff.

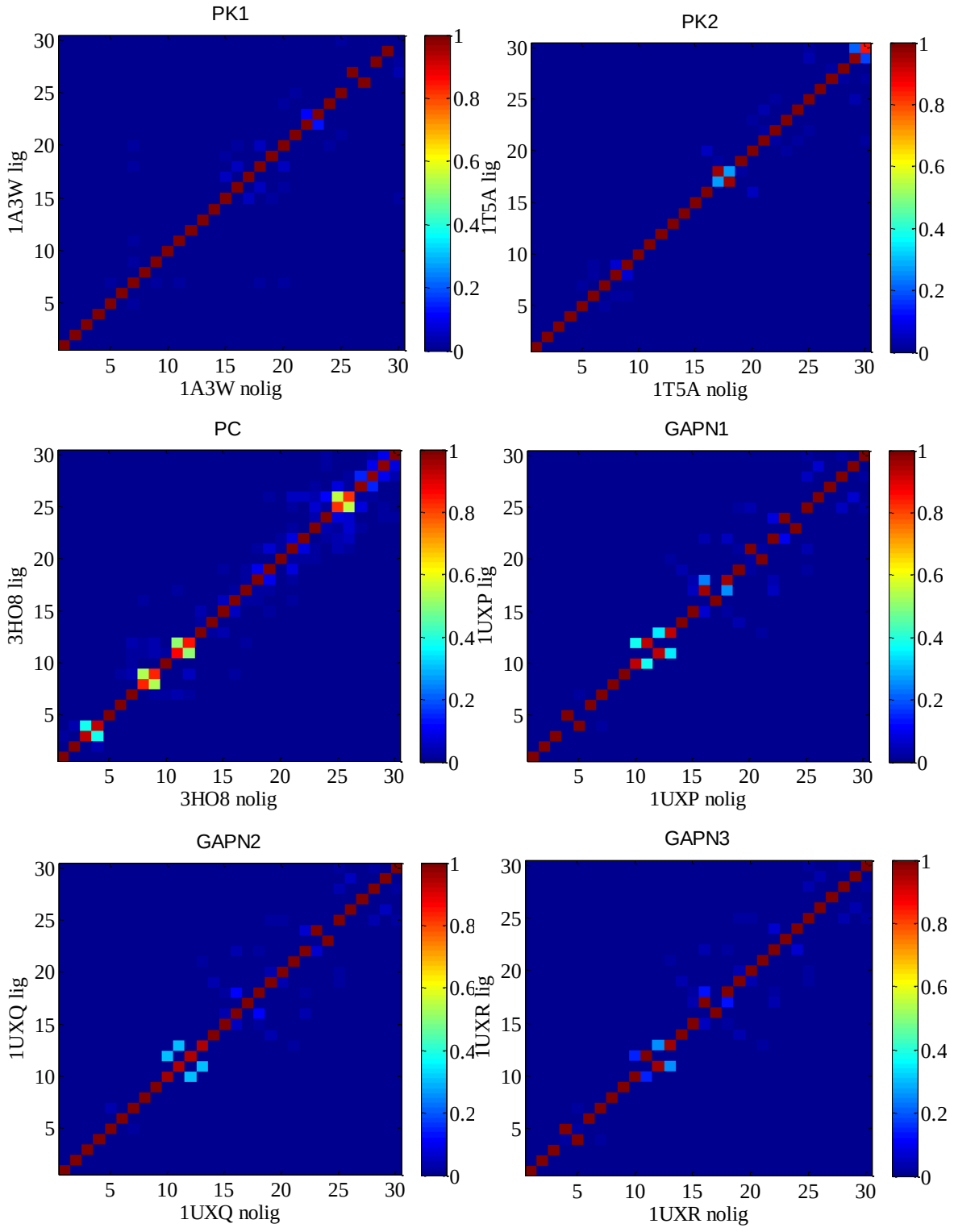


Figure B.6. Overlap matrices of tetramer samples with 10 Å cutoff.

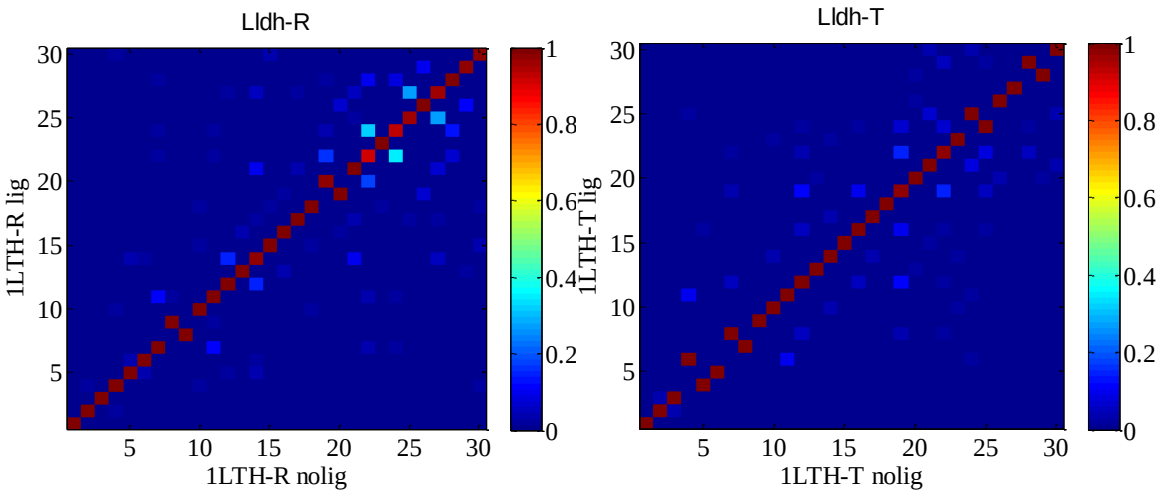


Figure B.6. Overlap matrices of tetramer samples with 10 Å cutoff cont.

a)

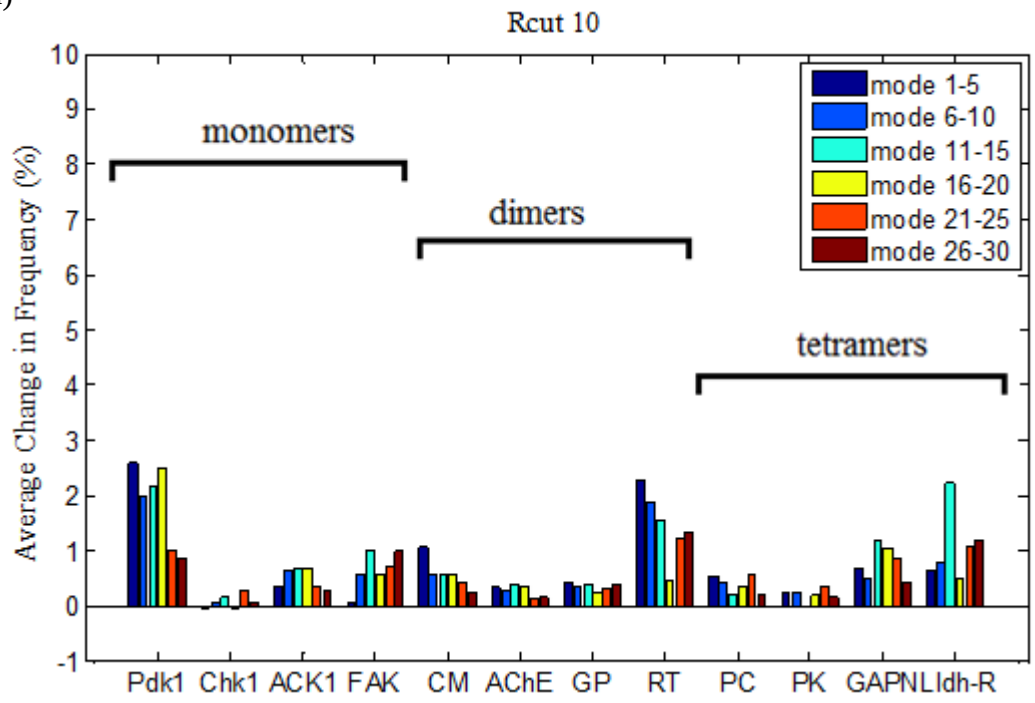


Figure B.7. Average Frequency % distribution due to allosteric ligand binding for cutoff 10 Å (a) each 5 (b) 5 and 30 modes.

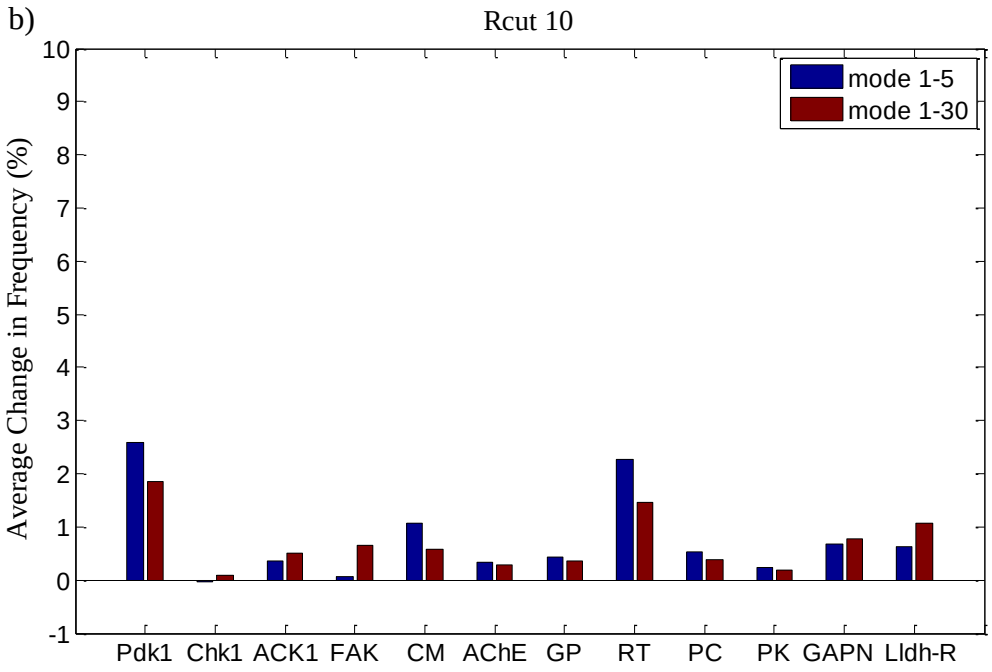


Figure B.7. Average Frequency % distribution due to allosteric ligand binding for cutoff 10 Å (a) each 5 (b) 5 and 30 modes cont.

Table B.1. Hinge residues of all samples with their original structures.

Enzyme	Slowest mode1 Hinge residues	Slowest mode2 Hinge residues
Chk1	91A	12, 92, 112,190
Pdk1	165	111, 157, 222, 332
FAK	506, 550, 567	499, 567, 642
ACK1	212	204, 270, 350
CM	256A	22, 155, 203, 233A 20, 155, 203, 233 B
AChE	542A	311,450,489,542 A 311, 450, 492 B
GP	249, 268, 831 A 249, 268 B	482, 812 A 482, 812 B
RT_1	249, 308, 325, 544 A 92, 114, 160 B	92, 114, 163, 430, 504, 531 A

Table B.1. Hinge residues of all samples with their original structures cont.

RT_2	245, 309, 325, 554 A 90, 112, 160 B	89, 106, 223, 392, 554 A 242, 361, 418 B
PC	1093A 771, 789, 807, 824, 853, 892, 1093 C 869, 890, 1093 D	377, 404, 436A 490 B 262, 284, 368, 1093 C
PK_1	500 B	10 A,B,C,D
PK_2	531 A,B,C	38, 531A 37 B 37, 531 C 37D
$GAPN_1$	493 B, D	132, 156, 483 A 132, 156, 483 B 132, 156, 483 C 132, 156, 483 D
$GAPN_2$	492 B, D	132, 156, 483 A 132, 156, 483 B 132, 156, 483 C 132, 156, 483 D
$GAPN_3$	493 B 493 D	132, 156, 483 A 132, 156, 483 B 132, 156, 483 C 132, 156, 483 D
Lldh-T	319 A, B, C	319 A, C
Lldh-R	319 A, B, C	46, 63 A 46, 63, 319 B 46, 63 C 46, 63 D

REFERENCES

Alakent B., S. Baskan, P. Doruker, 2011, "Effect of Ligand Binding on the Intraminimum Dynamics of Proteins", *Journal of Computational Chemistry,* Vol. 32, No. 3, pp. 483-496.

Alter O., P.O. Brown, D. Botstein, 2000, "Singular Value Decomposition for Genome-Wide Expression Data Processing and Modeling", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 97, No. 18, pp. 10101-10106.

Atilgan A.R., S.R. Durell, R.L. Jernigan *et al*, 2001, "Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model", *Biophysical Journal*, Vol. 80, No. 1, pp. 505-515.

Bahar I., A.R. Atilgan, B. Erman, 1997, "Direct Evaluation of Thermal Fluctuations in Proteins Using a Single-Parameter Harmonic Potential", *Fold Des*, Vol. 2, No. 3, pp. 173-181.

Bahar I., B. Erman, T. Haliloglu, R.L. Jernigan, 1997, "Efficient Characterization of Collective Motions and Interresidue Correlations in Proteins by Low-Resolution Simulations", *Biochemistry,* Vol. 36, No. 44, pp. 13512-13523.

Bahar I., R.L. Jernigan, 1998, "Vibrational Dynamics of Transfer Rnas: Comparison of the Free and Synthetase-Bound Forms", *Journal of Molecular Biology*, Vol. 281, No. 5, pp. 871-884.

Balog E., T. Becker, M. Oettl *et al*, 2004, "Direct Determination of Vibrational Density of States Change on Ligand Binding to a Protein", *Physical Review Letters*, Vol. 93, No. 2, pp. 028103.

Careri G., P. Fasella, E. Gratton, 1975, "Statistical Time Events in Enzymes: A Physical Assessment", *CRC Crit Rev Biochem*, Vol. 3, No. 2, pp. 141-164.

Changeux J.P., J.C. Gerhart, H.K. Schachman, 1968, "Allosteric Interactions in Aspartate Transcarbamylase. I. Binding of Specific Ligands to the Native Enzyme and Its Isolated Subunits", *Biochemistry,* Vol. 7, No. 2, pp. 531-538.

Chu J.W., G.A. Voth, 2007, "Coarse-Grained Free Energy Functions for Studying Protein Conformational Changes: A Double-Well Network Model", *Biophysical Journal,* Vol. 93, No. 11, pp. 3860-3871.

Cui Q., M. Karplus, 2008, "Allostery and Cooperativity Revisited", *Protein science : a publication of the Protein Society,* Vol. 17, No. 8, pp. 1295-1307.

Doruker P., A.R. Atilgan, I. Bahar, 2000, "Dynamics of Proteins Predicted by Molecular Dynamics Simulations and Analytical Approaches: Application to Alpha-Amylase Inhibitor", *Proteins,* Vol. 40, No. 3, pp. 512-524.

Doruker P., R.L. Jernigan, I. Bahar, 2002, "Dynamics of Large Proteins through Hierarchical Levels of Coarse-Grained Structures", *Journal of Computational Chemistry,* Vol. 23, No. 1, pp. 119-127.

Dundas J., Z. Ouyang, J. Tseng *et al*, 2006, "Castp: Computed Atlas of Surface Topography of Proteins with Structural and Topographical Mapping of Functionally Annotated Residues", *Nucleic Acids Research,* Vol. 34, No. Web Server issue, pp. W116-118.

Emekli U., D. Schneidman-Duhovny, H.J. Wolfson, R. Nussinov, T. Haliloglu, 2008, "Hingeprot: Automated Prediction of Hinges in Protein Structures", *Proteins,* Vol. 70, No. 4, pp. 1219-1227.

Eren D., B. Alakent, 2013, "Frequency Response of a Protein to Local Conformational Perturbations", *PLoS Computational Biology,* Vol. 9, No. 9, pp. e1003238.

Eyal E., L.W. Yang, I. Bahar, 2006, "Anisotropic Network Model: Systematic Evaluation and a New Web Interface", *Bioinformatics,* Vol. 22, No. 21, pp. 2619-2627.

Fischer S., C.S. Verma, 1999, "Binding of Buried Structural Water Increases the Flexibility of Proteins", *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 96, No. 17, pp. 9613-9615.

Gniewek P., A. Kolinski, R.L. Jernigan, A. Kloczkowski, 2012, "Elastic Network Normal Modes Provide a Basis for Protein Structure Refinement", *The Journal of Chemical Physics,* Vol. 136, No. 19, pp. 195101.

Goncearenco A., S. Mitternacht, T. Yong *et al*, 2013, "Spacer: Server for Predicting Allosteric Communication and Effects of Regulation", *Nucleic Acids Research,* Vol. 41, No. Web Server issue, pp. W266-272.

Grimes R.G., J.G. Lewis, H.D. Simon, 1994, "A Shifted Block Lanczos-Algorithm for Solving Sparse Symmetrical Generalized Eigenproblems", *Siam J Matrix Anal A,* Vol. 15, No. 1, pp. 228-272.

Gunasekaran K., B. Ma, R. Nussinov, 2004, "Is Allostery an Intrinsic Property of All Dynamic Proteins?", *Proteins,* Vol. 57, No. 3, pp. 433-443.

Holm L., P. Rosenstrom, 2010, "Dali Server: Conservation Mapping in 3d", *Nucleic Acids Research,* Vol. 38, No. Web Server issue, pp. W545-549.

Huang Z., L. Mou, Q. Shen *et al*, 2014, "Asd V2.0: Updated Content and Novel Features Focusing on Allosteric Regulation", *Nucleic Acids Research,* Vol. 42, No. Database issue, pp. D510-516.

Huang Z., L. Zhu, Y. Cao *et al*, 2011, "Asd: A Comprehensive Database of Allosteric Proteins and Modulators", *Nucleic Acids Research,* Vol. 39, No. Database issue, pp. D663-669.

Jaffe E.K., 2005, "Morpheeins--a New Structural Paradigm for Allosteric Regulation", *Trends in Biochemical Sciences,* Vol. 30, No. 9, pp. 490-497.

Jaffe E.K., S.H. Lawrence, 2012, "The Morpheein Model of Allostery: Evaluating Proteins as Potential Morpheeins", *Methods in Molecular Biology,* Vol. 796, No. pp. 217-231.

Koshland D.E., 1958, "Application of a Theory of Enzyme Specificity to Protein Synthesis", *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 44, No. 2, pp. 98-104.

Koshland D.E., Jr., K. Hamadani, 2002, "Proteomics and Models for Enzyme Cooperativity", *The Journal of Biological Chemistry,* Vol. 277, No. 49, pp. 46841-46844.

Koshland D.E., Jr., G. Nemethy, D. Filmer, 1966, "Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits", *Biochemistry,* Vol. 5, No. 1, pp. 365-385.

Kumar S., B. Ma, C.J. Tsai, N. Sinha, R. Nussinov, 2000, "Folding and Binding Cascades: Dynamic Landscapes and Population Shifts", *Protein Science : a publication of the Protein Society,* Vol. 9, No. 1, pp. 10-19.

Kurkcuoglu O., R.L. Jernigan, P. Doruker, 2004, "Mixed Levels of Coarse-Graining of Large Proteins Using Elastic Network Model Succeeds in Extracting the Slowest Motions", *Polymer,* Vol. 45, No. 2, pp. 649-657.

Kurkcuoglu O., R.L. Jernigan, P. Doruker, 2006, "Loop Motions of Triosephosphate Isomerase Observed with Elastic Networks", *Biochemistry,* Vol. 45, No. 4, pp. 1173-1182.

Kurkcuoglu O., O.T. Turgut, S. Cansu, R.L. Jernigan, P. Doruker, 2009, "Focused Functional Dynamics of Supramolecules by Use of a Mixed-Resolution Elastic Network Model", *Biophysical Journal,* Vol. 97, No. 4, pp. 1178-1187.

Kurkcuoglu Z., A. Bakan, D. Kocaman, I. Bahar, P. Doruker, 2012, "Coupling between Catalytic Loop Motions and Enzyme Global Dynamics", *PLoS Computational Biology*, Vol. 8, No. 9, pp. e1002705.

Laskowski R.A., E.G. Hutchinson, A.D. Michie *et al*, 1997, "Pdbsum: A Web-Based Database of Summaries and Analyses of All Pdb Structures", *Trends in Biochemical Sciences,* Vol. 22, No. 12, pp. 488-490.

Levitzki A., D.E. Koshland, Jr., 1969, "Negative Cooperativity in Regulatory Enzymes", *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 62, No. 4, pp. 1121-1128.

Marques O., Y.H. Sanejouand, 1995, "Hinge-Bending Motion in Citrate Synthase Arising from Normal Mode Calculations", *Proteins,* Vol. 23, No. 4, pp. 557-560.

Mattos C., C.R. Bellamacina, E. Peisach *et al*, 2006, "Multiple Solvent Crystal Structures: Probing Binding Sites, Plasticity and Hydration", *Journal of Molecular Biology,* Vol. 357, No. 5, pp. 1471-1482.

Monod J., J.P. Changeux, F. Jacob, 1963, "Allosteric Proteins and Cellular Control Systems", *Journal of Molecular Biology,* Vol. 6, No. pp. 306-329.

Monod J., J. Wyman, J.P. Changeux, 1965, "On the Nature of Allosteric Transitions: A Plausible Model", *Journal of Molecular Biology,* Vol. 12, No. pp. 88-118.

Panjkovich A., X. Daura, 2014, "Pars: A Web Server for the Prediction of Protein Allosteric and Regulatory Sites", *Bioinformatics,* Vol. 30, No. 9, pp. 1314-1315.

Porter C.T., G.J. Bartlett, J.M. Thornton, 2004, "The Catalytic Site Atlas: A Resource of Catalytic Sites and Residues Identified in Enzymes Using Structural Data", *Nucleic Acids Research,* Vol. 32, No. Database issue, pp. D129-133.

Schmid F.F., M. Meuwly, 2007, "All-Atom Simulations of Structures and Energetics of C-Di-Gmp-Bound and Free Pled", *Journal of Molecular Biology,* Vol. 374, No. 5, pp. 1270-1285.

Smock R.G., L.M. Gierasch, 2009, "Sending Signals Dynamically", *Science,* Vol. 324, No. 5924, pp. 198-203.

Stefan M.I., S.J. Edelstein, N. Le Novere, 2008, "An Allosteric Model of Calmodulin Explains Differential Activation of Pp2b and Camkii", *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 105, No. 31, pp. 10768-10773.

Sturtevant J.M., 1977, "Heat Capacity and Entropy Changes in Processes Involving Proteins", *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 74, No. 6, pp. 2236-2240.

Tama F., Y.H. Sanejouand, 2001, "Conformational Change of Proteins Arising from Normal Mode Calculations", *Protein Eng,* Vol. 14, No. 1, pp. 1-6.

Tama F., W. Wriggers, C.L. Brooks, 3rd, 2002, "Exploring Global Distortions of Biological Macromolecules and Assemblies from Low-Resolution Structural Information and Elastic Network Theory", *Journal of Molecular Biology,* Vol. 321, No. 2, pp. 297-305.

Tellez-Valencia A., V. Olivares-Illana, A. Hernandez-Santoyo *et al*, 2004, "Inactivation of Triosephosphate Isomerase from Trypanosoma Cruzi by an Agent That Perturbs Its Dimer Interface", *Journal of Molecular Biology,* Vol. 341, No. 5, pp. 1355-1365.

Thielges M.C., M.D. Fayer, 2012, "Protein Dynamics Studied with Ultrafast Two-Dimensional Infrared Vibrational Echo Spectroscopy", *Accounts of Chemical Research,* Vol. 45, No. 11, pp. 1866-1874.

Tirion M.M., 1996, "Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis", *Physical Review Letters,* Vol. 77, No. 9, pp. 1905-1908.

Tozzini V., 2005, "Coarse-Grained Models for Proteins", *Current Opinion in Structural Biology,* Vol. 15, No. 2, pp. 144-150.

Tsai C.J., A. del Sol, R. Nussinov, 2008, "Allostery: Absence of a Change in Shape Does Not Imply That Allostery Is Not at Play", *Journal of Molecular Biology,* Vol. 378, No. 1, pp. 1-11.

Weber G., 1972, "Ligand Binding and Internal Equilibria in Proteins", *Biochemistry,* Vol. 11, No. 5, pp. 864-878.

Williams S.G., K.B. Hall, 2014, "Linkage and Allostery in Snrnp Protein/Rna Complexes", *Biochemistry,* Vol. 53, No. 22, pp. 3529-3539.

Yennawar N.H., H.P. Yennawar, G.K. Farber, 1994, "X-Ray Crystal Structure of Gamma-Chymotrypsin in Hexane", *Biochemistry,* Vol. 33, No. 23, pp. 7326-7336.

Yu H., A.N. Gupta, X. Liu *et al*, 2012, "Energy Landscape Analysis of Native Folding of the Prion Protein Yields the Diffusion Constant, Transition Path Time, and Rates", *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 109, No. 36, pp. 14452-14457.

Zidek L., M.V. Novotny, M.J. Stone, 1999, "Increased Protein Backbone Conformational Entropy Upon Hydrophobic Ligand Binding", *Nat Struct Biol,* Vol. 6, No. 12, pp. 1118-1121.