# SABANCI UNIVERSITY

# GRADUATE SCHOOL OF ENGINEERING AND NATURAL SCIENCES

Sabancı
Universitesi

# COMPUTATIONAL ASSESSMENT OF THE EFFECT OF ALLOSTERIC MUTATIONS ON THE DYNAMICS OF PDZ DOMAINS

**M.Sc. THESIS**

**Nazlı KOCATUĞ**

**Department of Molecular Biology-Genetics and Bioengineering**

**Thesis Advisor: Prof. Dr. Canan Atılgan**

**JULY 2018**

# COMPUTATIONAL ASSESSMENT OF THE EFFECT OF ALLOSTERIC MUTATIONS ON THE DYNAMICS OF PDZ DOMAINS

APPROVED BY:

Prof. Dr. Canan Atılgan (Thesis Advisor) .................................................

Prof. Dr. Ali Rana Atılgan .................................................

Assist. Prof. Mehdi Sahihi .................................................

DATE OF APPROVAL: 13/07/2018

*To all the sparks that have shown courage to turn into flames*

# ABSTRACT


**Keywords:** PDZ domains, allostery, molecular dynamics simulations, free energy calculations



PDZ domain-containing proteins are involved in intercellular interactions such as trafficking, signaling, cell to cell communication and organization of signaling complexes. PDZ domains are themselves small proteins which typically consist of 90 to 100 amino acids. However, the extra α helix structure at the carboxyl terminus introduces a selective structural feature to the third PDZ domain of PSD-95 which has a stabilizing effect and participates in allosteric communication. PDZ domains are the most commonly studied models to understand single domain allostery without resulting in significant structural changes. One change triggers another change at distal site, and the source of the 'changes' are localized perturbations such as a binding event, posttranslational modification, a mutation or light absorption. Mutations can alter the stabilization of the protein and result ON or OFF state for ligand binding. They can also cause a change in the active site and affect the ligand preference. Here we investigate the reasons leading to the allosteric regulation of mutations and their effect on the ligand preferences.

By using third PDZ domain of postsynaptic density 95 (PSD-95) as a model system H372 directly connected to the binding site and G330 with a somewhat removed position were selected to assess the effect of allosteric mutations on the dynamics. In the literature, it was observed that the H372A and G330T/H372A mutations change ligand preferences from class I (T/S amino acid preference at position 2 of the ligand) to class II (hydrophobic amino acid preference at position 2 of the ligand). On the other hand, the G330T mutation leads to the recognition of both class I and class II types of ligands. Therefore, H372A is a 'switching mutation' while G330T mutation is 'class bridging'. We have performed 200 ns molecular dynamics simulations for wild-type, H372A, G330T single mutants and a double mutant of third PDZ domain in the absence and presence of both types of ligands. The comparative study helps to identify the changes in the dynamics that are effective in the onset and prevention of allosteric communication. With the combination of free energy difference calculations and a detailed analysis of MD trajectories, the behavior of the PDZ domain under the mutations, which are 'class bridging'(G330T) and 'class changing'(H372A), and their effects on the ligand preferences and binding affinities are explained. We show that the ensemble view of allostery provides a better description of site-to-site coupling rather than a pathway view that assumes a direct connection between the effector and binding sites.

# ÖZET

**Anahtar kelimeler:** PDZ bölgeleri, allosteri, moleküler dinamik simülasyonları, serbest enerji hesaplamaları

PDZ bölgeleri içeren proteinler sinyalleşme, hücre-hücre iletişimi, sinyalleşme komplekslerinin organizasyonu gibi hücreler arası etkileşimlerde yer alırlar. PDZ bölgeleri 90 ila 100 aminoasit içeren küçük proteinlerdir. Bununla birlikte, karboksil terminalindeki ekstra alfa sarmal yapısı, PSD-95'in üçüncü PDZ bölgesine stabilize edici ve allosterik iletişime katılan seçici bir yapısal özellik kazandırır. PDZ bölgeleri, önemli yapısal değişikliklere yol açmayan allosteriyi anlamak için en çok çalışılan modeller moleküllerdir. Bir değişiklik, uzak bir bölgede başka bir değişikliği tetikler ve bu "değişikliklerin" kaynağı, bir molekül bağlama olayı, genetik çeviri sonrası modifikasyon, mutasyon veya ışık soğurumu gibi bölgesel sarsımlardır. Mutasyonlar proteinin stabilizasyonunu değiştirebilir ve proteini ligand bağlanması için açık veya kapalı duruma getirebilir. Ayrıca aktif bölgede değişikliğe sebep olarak ligand tercihini etkileyebilirler. Bu çalışmada, mutasyonların allosterik regülasyonlara yol açan sebepleri ve ligand tercihleri üzerindeki etkileri araştırılmıştır.

Postsinaptik yoğunluk 95 (PSD-95) proteininin üçüncü PDZ bölgesi model sistem olarak kullanılarak; doğrudan bağlanma bölgesiyle ilişkili H372 ve bağlanma bölgesinden biraz uzaktaki G330 amino asitleri, allosterik mutasyonların protein dinamiği üzerindeki etkisini anlamak için seçilmiştir. Literatürde H372A ve G330T / H372A mutasyonlarının ligand tercihlerini sınıf I'den (ligandın 2. pozisyonunda T / S kalıntı tercihi) sınıf II'ye (ligandın 2. pozisyonunda hidrofobik amino asit tercihi) değiştirdiği gözlenmiştir. Öte yandan G330T mutasyonu hem sınıf I hem de sınıf II tip ligandların kabul edilmesine yol açar. Bu sebeple, H372A 'sınıf değiştirme mutasyonu', G330T mutasyonu 'sınıf köprüleyen mutasyon' olarak adlandırılır. Üçüncü PDZ bölgesinin yabanıl tip, H372A ve G330T tek mutasyonlu ve her iki mutasyonu da barındıran durumları için, her iki tip ligandın varlığında ve yokluğunda 200 ns'lık moleküler dinamik simülasyonlarını ikişer defa gerçekleştirdik. Karşılaştırmalı çalışma, allosterik iletişimin tetiklenmesinde ve engellenmesinde etkili olan dinamiklerdeki değişimlerin belirlenmesine yardımcı olmaktadır. Serbest enerji değişimleri hesaplamaları ve simülasyon süreçlerinin detaylı analizleri ile, 'sınıf birleştiren' (G330T) ve 'sınıf değiştiren' (H372A) mutasyonların; PDZ bölgelerinin davranışı, ligand tercihleri ve bağlanma afiniteleri üzerindeki etkileri açıklanmıştır. Çalışmamız, bu proteinde gözlemlenen allosterinin "genel uyum" görüşünün, bölgeden bölgeye bağlantıyı tanımlamak için, efektör ve bağlanma alanları arasında doğrudan bir bağlantı olduğunu kabul eden "yolak anlayışı"na göre daha iyi bir tanım olduğunu göstermektedir.

# ACKNOWLEDGEMENTS

**Table of Contents**

# List of Figures

# List of Tables

# Chapter 1

## Introduction and Motivation

### 1.1 PDZ Domain Family

PDZ domain-containing proteins were first recognized in the early 1990s. Such proteins involve intercellular interactions such as trafficking, signaling, cell to cell communication and organization of signaling complexes [1, 2]. They are found in a wide variety of organisms from bacteria to vertebrates. For instance, the human proteome has over 250 different variations of PDZ domain-containing proteins [1]. The domain's abbreviation stems from the very first discovered PDZ-containing proteins – PSD95 (postsynaptic density protein 95) *P*, the Drosophila tumor suppressor protein Dlg-1 (discs large) *D*, and ZO-1 (zonula occludens protein) *Z* [3] (Fig 1.1). PSD-95 and Dlg-1 are homologous – mostly similar with respect to their sequences – insofar as they both have three PDZ domains with Src Homology 3 (SH3) and guanylate kinase (GK) domains at their C-termini [4]. SH3 domains are specific motifs, target proteins which have a role in signaling pathways [5]. GK domain is one of the members of Membrane-associated guanlylate kinases (MAGUKs) which are responsible for cell to cell interaction and cell polarity control [6]. However, in spite of extensive knowledge on MAGUK family, the information in the GK domain is very limited [6]. Unlike SH3 and GK domains, ZU5 domain is found only in Zonula Occludens Protein (ZO-1), and it contributes to the stabilization of ZO-1 protein [7].
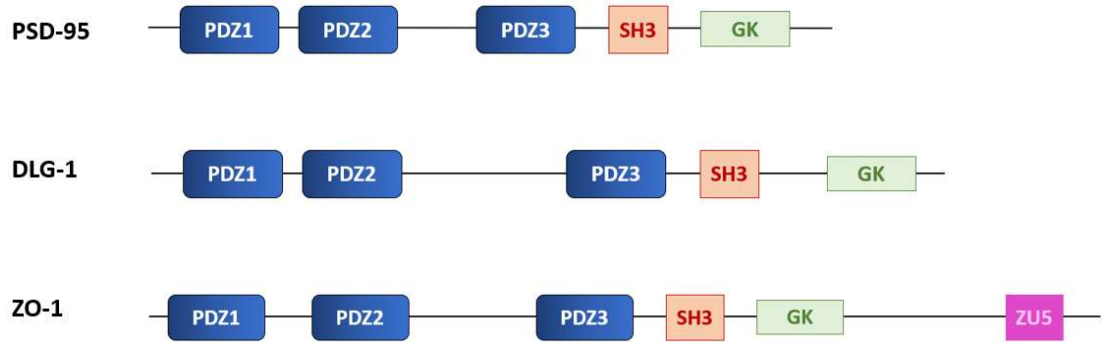
**Figure 1.1. Schematic representation of PDS-95, DLG-1, and ZO-1.**

PDZ domain-containing proteins consist of varying numbers of PDZ domains, demonstrating a great variety among species which account for the abundance of the PDZ domains [8]. However, it is yet to be determined if the number of domains has a significant effect on the function of PDZ domain-containing proteins.

PDZ domains are themselves small proteins which typically consist of 90 to 100 amino acids and have 6-7 $\beta$-strands and 2 $\alpha$-helical structures as characteristic features [1-5]. Common structural motifs are consistent in the folding respect but differ in length amongst the PDZ domain family [9] (Fig. 1.3). In addition to its secondary structure, the loop at the binding cleft, referred to as the carboxylate-binding loop, is one of the hallmarks of the PDZ domain [10]. This loop consists of 'GLGF' amino acids, and, therefore, PDZ domains are also known as 'GLGF repeats' [8, 11]. However, there is a selective structural feature of the third PDZ domain of PSD-95 which makes it more interesting to work on; it is the extra α helix structure at the carboxyl terminus [12] (Fig 1.2). Studies reveal the significance of the unusual α-helix, which has a stabilizing effect and participates in allosteric communication [12, 13].
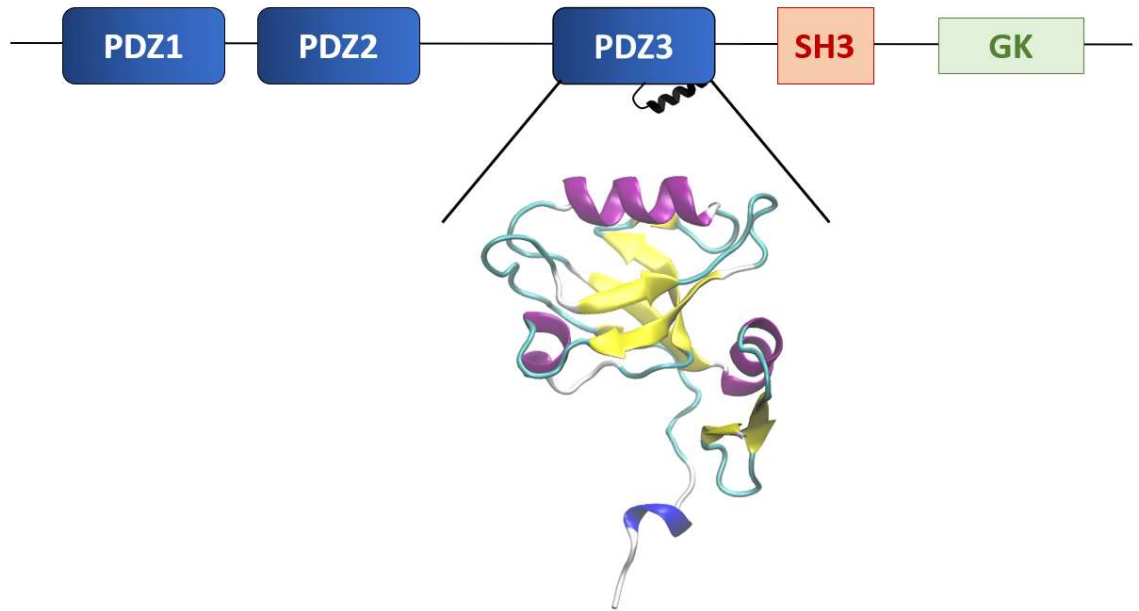
**Figure 1.2. The third PDZ domain of PSD-95.**

Members of the PDZ domain family bind selectively to short amino acid patterns at the C-termini of binding proteins [14]. Therefore, the pattern of the target protein is the definitive factor for the classification of PDZ domains [1]. PDZ domain-binding protein interactions are categorized based on the amino acid at the second position of the binding protein downstream from the C-Terminus [10, 15]. - *Class-I* for Thr/Ser, *Class-II* for hydrophobic amino acids, and *Class-III* for Asp/Glu [10, 16, 17].
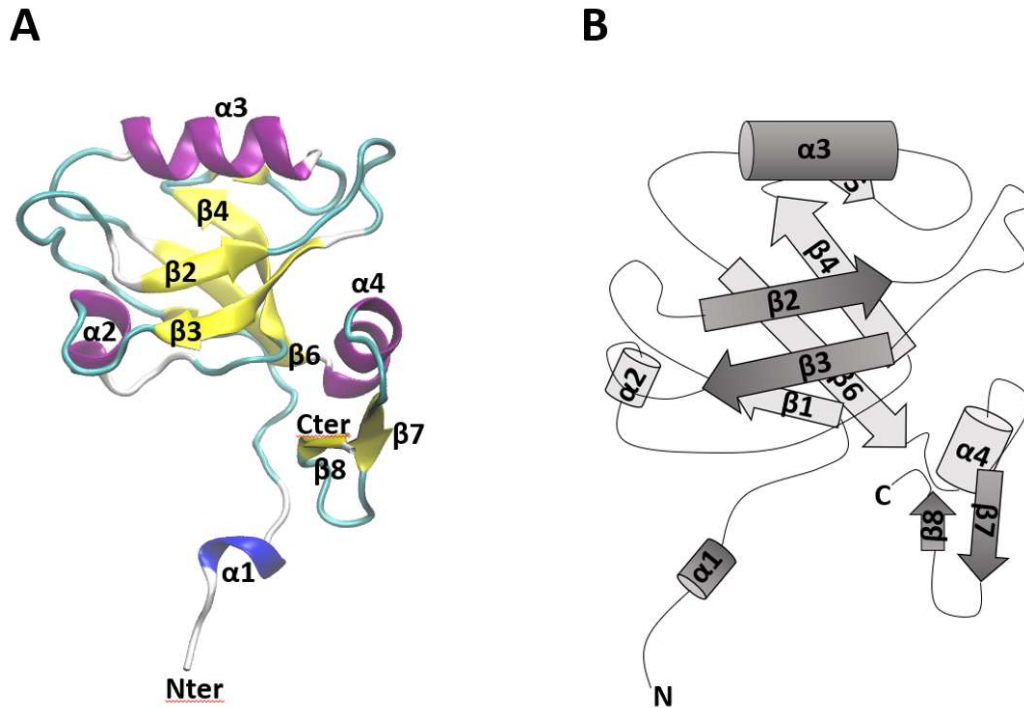
**Figure 1.3. Schematic of the Secondary structure of third PDZ domain.** Secondary structure, purple represents α helix, yellow shows β strands (A). Schematic representation of the secondary structure (B).

For this study, the third PDZ domain of PSD-95 is used as a model protein. For this protein, *Class I* and *Class II* interactions are possible (Fig 1.4A). There are two determinant residues affecting ligand preference – His372 and Gly330 [18]. His372 is at the binding pocket of the protein and has direct contact with the ligand. On the other hand, Gly330 does not have direct contact with ligands, but it is located on the right loop of the binding pocket [18]. It has been proven that mutations on these two residues alter the protein's affinity to different classes of ligands [18]. The wild-type (WT) third PDZ domain prefers to bind ligands of *Class I* (Fig. 1.4B). On the contrary, if His372 is mutated into Alanine, the binding protein can bind to *Class II* instead of *Class I*. Due to the resulting change in ligand preference, H372A mutations are called *class changing mutations* [18] (Fig. 1.4B). While the H372A mutation has the inhibitory effect on *Class I*, a Gly to Thr mutation at the 330th residue allows the binding protein to bind both *Classes I* and *II* ligands. Mutations such as G330T mutations are called *class bridging mutations* [18] (Fig. 1.4B). If the binding protein undergoes both H372A and G330T

mutations simultaneously as a double mutant (DM), the protein binds *Class-II* ligands only; i.e. the double mutation is also class changing [18] (Fig. 1.4B).
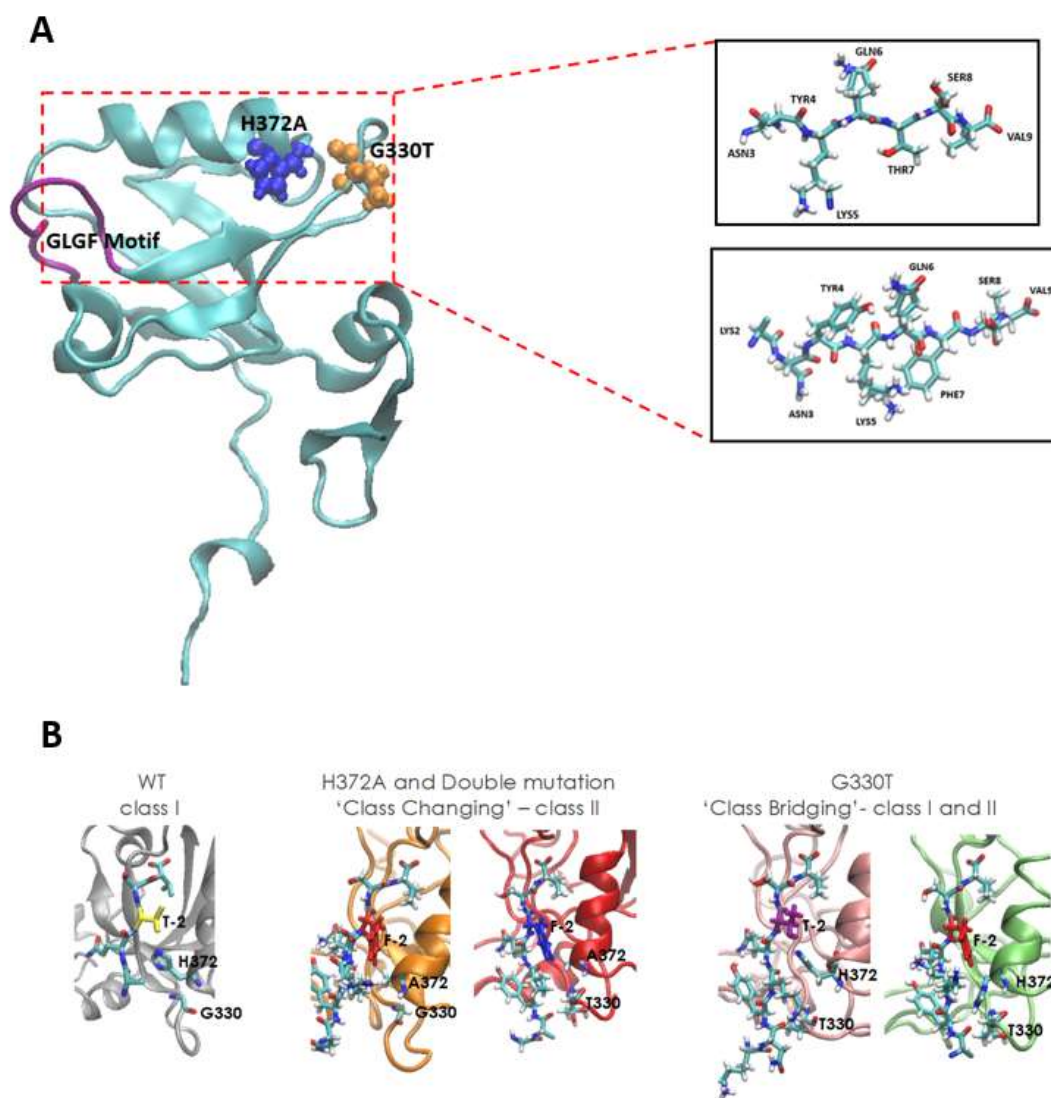


**Figure 1.4. Binding site and ligand preferences.** Schematic of the binding pocket and ligand classes (A). The change of ligand preferences due to mutation types accordingly (B).

## 1.2 PDZ Domains as Models of Allostery

Allostery is originated from Greek word, meaning "other sites", where perturbations at one site of the protein, can influence changes at a distal site, complicating deducing structure-function relationship in proteins, the concept has changed in the years [19, 20]. In 1904 the 'Bohr effect' concept which was describing the effect of $CO_2$ on the

binding affinity of $O_2$ to hemoglobin was proposed [21]. With the description of this new concept, the era leading to the naming of 'allostery' begun [21].

The word 'allostery' first appeared in 1961, and in 1963 the concept of 'allosteric sites' was used instead of 'regulatory sites' by Jacques Monod [22]. In the 1960s two significant concepts were described: the first one is the concerted model of Monod, Wyman, and Changeux (MWC) model and the second one is the sequential model which was proposed by Koshland, Némethy, and Filmer (KNF) [23]. Both models take cooperativity as the key regulatory. The conformational changes were taken as a signature of allostery until 1984 [21, 24]. In 1984, the 'dynamic allostery' concept was described by Cooper and Dryden which describes allostery without the need of conformational change by introducing the entropy contribution [21, 25]. With the introduction of non-conformational change allostery, in addition to multi-domain proteins, single domain ones started to be investigated for a better understanding of allostery. PDZ domains are the most commonly studied models to understand single domain allostery without resulting in significant structural changes [26].

The required localized perturbation for allosteric regulation can cause by a binding event, post-transitional modifications or mutations [27]. Mutations attributes in two significant regulations. Firstly, it can alter the stabilization of the protein and result ON or OFF state for ligand binding [28]. Secondly, it can cause a change in the active site and affect the ligand preference [28]. In this study, the reason for the allosteric regulation is mutation and its effect on the ligand preferences.

## 1.3 The Scope of the Thesis

PDZ domains themselves are small proteins, typically composed of 90 to 100 amino acids [1-5]. The extra α-helix structure at the carboxyl terminal is the specific structural property of the third PDZ domain of PSD-95, which is why it is widely studied [12]. To work on the third PDZ field of PSD-95 is appealing not only because of the additional α-helical structure but also due to its significance and abundance in the understanding of single domain allostery. In 1999, with the introduction of the allosteric networks of PDZ [29], PDZ domains became the preferred model to understand the single domain allostery

We performed 200 ns molecular dynamics simulations for wild-type, H372A, G330T single mutants and a double mutant of third PDZ domain in the absence and presence of both types of ligands described in figure 1.4. The comparative study helps to identify the changes in the dynamics that are effective in the onset and prevention of allosteric communication.

The first step in this work has been to use 50 ns apart snapshots from the MD trajectories to carry out free energy perturbation calculations. As a result, it is possible to make a thermodynamic assessment of ligand binding which is the key to understanding the function. With the combination of Free Energy Perturbation (FEP) results and a detailed analysis of MD trajectories, the behavior of the PDZ domain under the mutations, which are 'class bridging'(G330T) and 'class changing'(H372A), and their effects on the ligand preferences and binding affinities are explained. We conclude with a simple model for explaining PDZ domain allostery by considering entropy-enthalpy compensation.

# Chapter 2

## Theory and Method

### 2.1 Molecular Dynamics (MD) Simulations

MD simulations were carried out with *Nanoscale Molecular Dynamics* (NAMD) program using CHARMM36 force field parameters. The simulations were visualized using the molecular graphics program *VMD* [30].

The very first concept and keywords of the MD simulations to understand protein folding were introduced in 1975 [31]. MD is a Newtonian physics-based approach used for the understanding of molecular conformations, interactions and rearrangements [32][33].

The first step in an MD simulation of a protein is to obtain its structural information (coordinates) from the Protein Data Bank (PDB). The protein is placed into a water box whereby a distance (10 Å) is maintained between each atom of the protein and the nearest edge of the water box by using the solvent plug-in in VMD 1.9.3. By adding a sufficient number of potassium chloride (KCl) to the system, the ionic strength is adjusted to 150 mM after achieving charge neutrality. Long-range electrostatics are calculated by the particle mesh Ewald method [34], with a cut-off distance of 12 Å. Temperature control is maintained by Langevin Dynamics. The system is run under 1 atm and 310 K in the NPT ensemble. The complete system is minimized for 10,000 steps and equilibrated for 100,000,000 steps. Each step is 2 fs.

This procedure is repeated twice for 12 proteins. Their PDB codes are presented in Table2.1.

**Table 2.1. PDB codes of model proteins.**

| | |
|---|---|
| 5HDY | The third PDZ domain from the synaptic protein PSD-95 |
| 5HET | The third PDZ domain from the synaptic protein PSD-95 (G330T mutant) |
| 5HF4 | The third PDZ domain from the synaptic protein PSD-95 (H372A mutant) |
| 5HFD | The third PDZ domain from the synaptic protein PSD-95 (G330T, H372A double mutant) |
| 5HEB | The third PDZ domain from the synaptic protein PSD-95 in complex with a C-terminal peptide derived from CRIPT |
| 5HEY | The third PDZ domain from the synaptic protein PSD-95 (G330T mutant) in complex with a C-terminal peptide derived from CRIPT |
| 5HFB | The third PDZ domain from the synaptic protein PSD-95 (H372A mutant) in complex with a C-terminal peptide derived from CRIPT |
| 5HFE | The third PDZ domain from the synaptic protein PSD-95 (G330T, H372A double mutant) in complex with a C-terminal peptide derived from CRIPT |
| 5HED | The third PDZ domain from the synaptic protein PSD-95 in complex with a mutant C-terminal peptide derived from CRIPT (T-2F) |
| 5HF1 | The third PDZ domain from the synaptic protein PSD-95 (G330T mutant) in complex with a mutant C-terminal peptide derived from CRIPT (T-2F) |
| 5HFC | The third PDZ domain from the synaptic protein PSD-95 (H372A mutant) in complex with a mutant C-terminal peptide derived from CRIPT (T-2F) |
| 5HFF | The third PDZ domain from the synaptic protein PSD-95 (G330T, H372A double mutant) in complex with a mutant C-terminal peptide derived from CRIPT (T-2F) |

## 2.1.1 Trajectory Analysis

Trajectory analysis was carried out considering two sets of simulations. For RMSF, cross-correlation, and hydrogen-bond occupancy analyses, the first 80 ns of the MD simulation (200 ns) were excluded for the molecule to get equilibrated. The remaining 120 ns was divided into three equal 40 ns chunks. Their average is used for RMSF and cross-correlation calculations.

## 2.1.1.1 Root Mean Square Deviation (RMSD)

For understanding the similarities and differences between two three-dimensional structures, calculating the root mean square deviation (RMSD) is one of the most widely used methods [1]. RMSD defined as:

$$RMSD\ (t_2, t_1) = \sqrt{\frac{1}{N_\alpha}\sum_{\alpha=1}^{N_\alpha}\left(\langle\vec{R}_\alpha t_2\rangle - \langle\vec{R}_\alpha t_1\rangle\right)^2} \qquad (1)$$

where $N_\alpha$ is the total number of $C_\alpha$ atoms in the molecule, $\vec{R}_\alpha$ is the atomic position of the α atom in the predetermined time interval. By taking the first time-step as a reference point, the average atomic position is calculated throughout whole the trajectory.

### 2.1.1.2 Root Mean Square Fluctuation (RMSF)

Root mean square fluctuation (RMSF) gives the information about the local flexibility of the protein. RMSF is defined as:

$$RMSF_i = \sqrt{\frac{1}{T}\sum_t^T(\langle r_i(t) - \overline{r_i}\rangle)^2} \qquad (2)$$

where $T$ is the total frame number, $r_i(t)$ is the position of the $i^{\text{th}}$ atom in $t^{\text{th}}$ frame, and $\overline{r_i}$ is the average position of the $i^{\text{th}}$ atom according to the frame. RMSF takes mean structure over the trajectory as a reference [36].

### 2.1.1.3 Principal Component Analysis (PCA)

Molecular Dynamics simulations can provide a considerable amount of data which needs to be processed. At this point, reducing produced data into smaller data without losing any crucial information via clustering methods is an essential step.

Principle component analysis is a mathematical method based on the covariance matrix, useful for finding correlated motions by reducing the dimensionality of datasets [37, 38]. By diagonalizing the covariance matrix, the essential motions -principal components- are extracted.

The first step is superposing all the conformations through the trajectory. With the gathering of the displacement information from the first step, constructing a covariance matrix is the second step. For a system consists of $N$ residues, which is the total number of atoms in the molecule, the covariance matrix has the size of $3N$x$3N$ [39]:

$$C_{ij} = \langle x_i - \langle x_i \rangle \rangle \cdot \left\langle x_j - \langle x_j \rangle \right\rangle \tag{3}$$

Where, $C$ is the covariance matrix and $x_i$ and $x_j$ are the displacement of $i^{th}$ and $j^{th}$ atom respectively.

Decomposition of the covariance matrix needs to be done to extract the information of direction and magnitude of the displacement.

$$C = Q\Lambda Q^T \tag{4}$$

While is an orthonormal matrix whose columns are eigenvectors-also called as modes and is a diagonal matrix which has the eigenvalues. Eigenvectors($Q$) have the information of the direction of the motion. While eigenvalues($\lambda$) give the insight of the magnitude of the motion. In this work, we display the contribution of the slowest mode of motion (1$^{st}$ PCA) by projecting the eigenvectors on the protein structure.

### 2.1.1.4 Cross-Correlation Analysis

Cross-correlation analysis is used for identifying correlated motions of a protein in its equilibrated state [37]. Cross-correlation is defined as follows:

$$c_{i,j} = \frac{\left\langle \Delta R_i \cdot \Delta R_j \right\rangle}{\sqrt{\langle \Delta R_i \cdot \Delta R_i \rangle \left\langle \Delta R_j \cdot \Delta R_j \right\rangle}} \tag{5}$$

While $\Delta R_i$ is the displacement vector of residue $i$ and $\Delta R_j$ represents the displacement vector of residue $j$. The cross-correlation value varies between -1, total anticorrelation, and 1, total correlation, and represents the level of variation between residues. In the representation of the cross-correlation maps, red and blue represents positive and negative correlation, respectively. The cross-correlation analysis is performed by considering only $C\alpha$ atoms of the amino acids in the protein.

## 2.2 Free Energy Perturbation (FEP) Calculations

Paramount in understanding the relationship between structure and function of a biological process are thermodynamic features, more specifically free energy change [38]. The thermodynamic features of a system may become altered due to ligand binding or protein side-chain mutations [39].
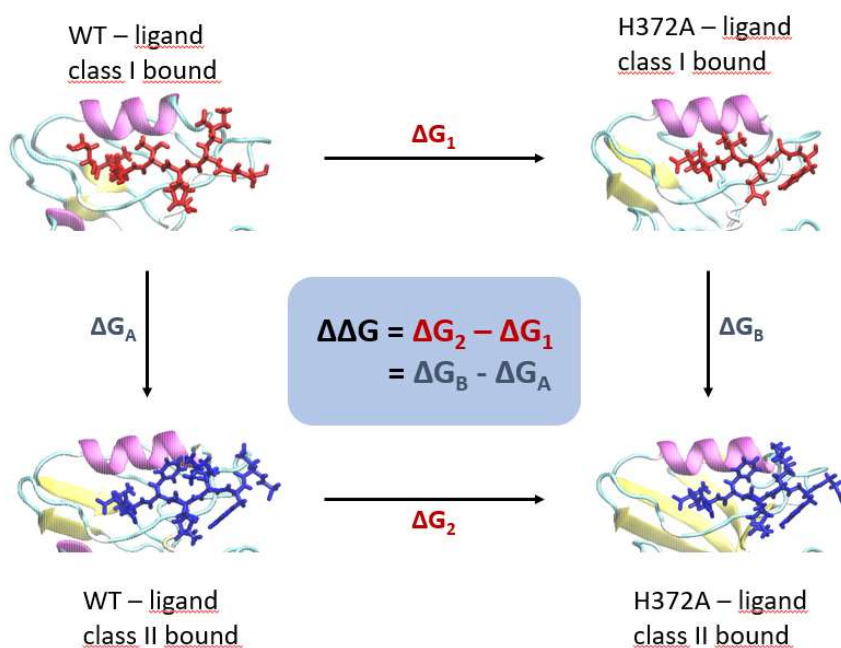


**Figure 2.1. Thermodynamic cycle of H372A mutation.**

In 1954, Zwanzig revealed the relationship between the free energy of the system and the average of the potential energy differences [40].

$$e^{(-\beta \Delta G)} = \left\langle e^{(-\beta \Delta U)} \right\rangle_0^{eq} \tag{6}$$

The free energy difference is calculated between reference and target systems, which has the coupling parameter, $\lambda$, between 0 and 1, where 0 and 1 represent the initial and final states, respectively [41]. The potential energy difference between two states are calculated via the alchemical transformation equation as follows:

$$U(\lambda_i) = \lambda_i U_0 + (1 - \lambda_i)U_1 = U_0 + \lambda_i \Delta U \tag{7}$$

If there are $n$ number of intermediate states, windows, between the reference and target system, the partition function equals as follows:

$$\Delta G = -\frac{1}{\beta} ln\left(\frac{Q_1}{Q_2}\right) = -\frac{1}{\beta} ln\left(\frac{Q_M}{Q_0}\frac{Q_1}{Q_M}\right) = \Delta A_{0,M} + \Delta A_{M,1} \tag{8}$$

Here $\Delta G_{0,M}$ and $\Delta G_{M,1}$ represent the free energy differences between 0 and $M$, and $M$ and 1 intermediate states [42]. $N$ number of windows allows calculating $\Delta\Delta G$.

While calculating free energy changes, the system space is sampled in both ways, forward and backward. The ratio of the probability distribution equals the difference between free energy and potential energy. To increase the overlap of a probability distribution, sampling carries great significance [42]. Thereby, the window number is set as 32 between WT and mutated states in our simulations (Fig. 2.2). Each window is 200 ps long with 50ps equilibration and 150 ps data generation.

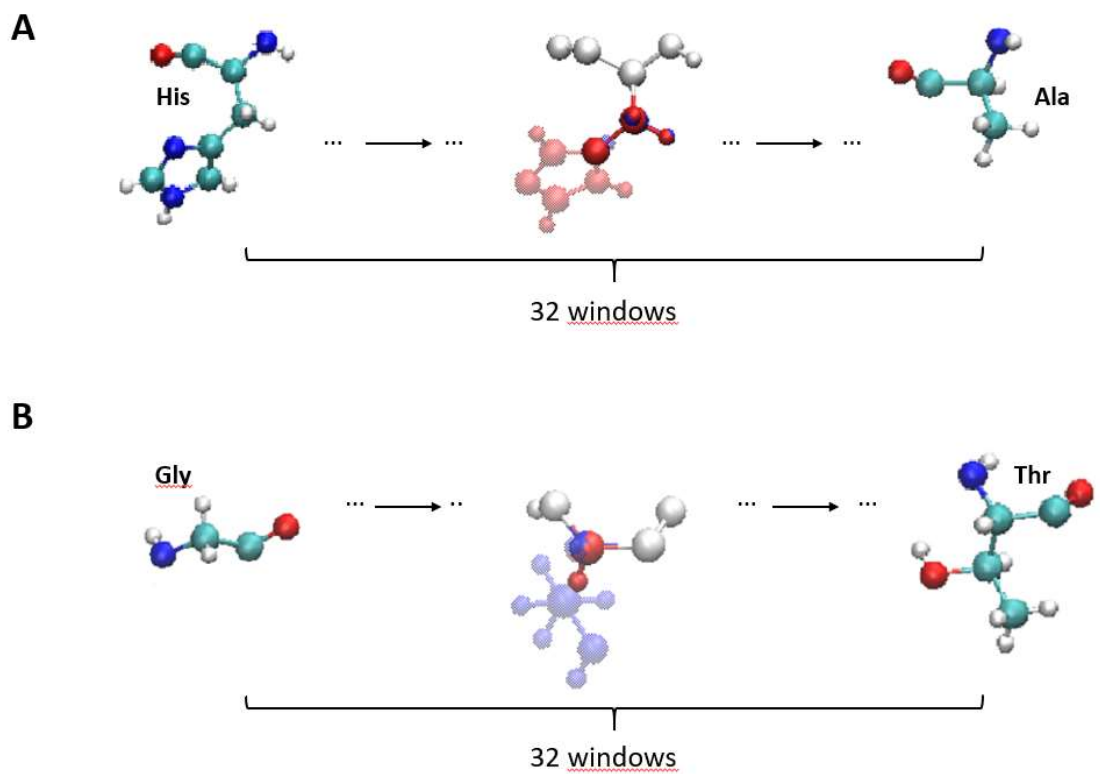**Figure 2.2. Intermediate steps between the initial and final steps.** Windows for H372A(A), intermediate states in the case of G330T(B). For intermediate steps, red color represents vanishing atoms and blue color represents appearing atoms.

# Chapter 3

## Structural and thermodynamic information: Free Energy Perturbation Calculations

To determine which states of the PDZ domain would be proper for free energy perturbation calculations, we first carry out 200 ns MD simulations on apo, and either ligand bound forms of PDZ domains for the wild type, G330T, H372A, and the double mutant forms of the protein. The initial structures for FEP calculations are taken from the 50ns, 100ns, 150ns and 200ns time points of the samples. To decide whether the simulations and sampling are reliable, free energy perturbation calculations are compared with an experimental study reporting $K_d$ values which shows the binding affinity of PDZ domain in the presence of ligand class I and II [18].

## 3.1 How the RMSD differed due to mutation types and ligand-bound/unbound states?

The wild-type, H372A, G330T single mutants and double mutant cases were simulated to understand the dynamics of the PDZ domain in the presence and absence of both ligand types. The simulations were carried out for 200 ns and it is observed that the equilibration time varies considerably among different states.

By comparing the RMSD results, it is aimed to understand whether the behavior of the protein has changed between the two simulations. For each case, ligand unbound, ligand class-I bound, and ligand class-II bound conditions were compared.

For WT in the absence of ligand, the initial simulation has an RMSD range between 2 and 3 Å, while the repeated simulation has a range of 3 to 4 Å. In Figure 3.1A, the green shaded area shows a distinct state sampled between 80-120 ns. The reason is that, the N terminus is only twisted on *α2* (regions are as defined in Fig. 1.3B) within this

interval (a sample snapshot is shown in Fig. 3.1B). Outside this range, the N terminus moves freely. Although the molecules have different states throughout the trajectory, the two simulations seem to have an almost identical endpoint (Figure 3.1A).

When the results are compared to the ligand-bound and unbound conditions of the wild type protein, it has been observed that ligand binding increases the RMSD values of the protein. RMSD was increased to 3.5 and 4 Å by ligand binding, in both simulations (Fig. 3.1C, 3.1D). However, the movement of the N terminus is different from the case where the ligand is absent. In the presence of ligand class II, the N terminus is packed on the C terminus in the range of 130-200 ns in one simulation (Fig. 3.1E, dark blue). Apart from these observations on RMSD, there is no significant difference in terms of ligand class I and ligand class-II bound status.
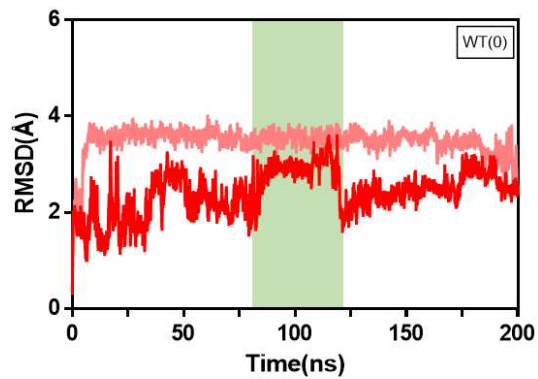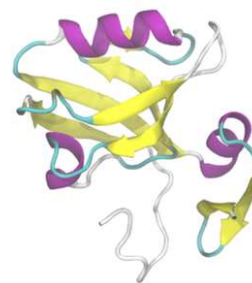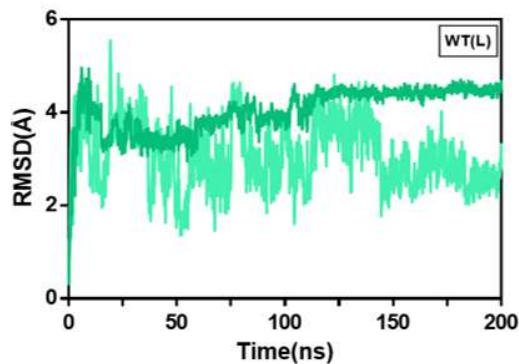
**Figure 3.1. RMSD results and sample snapshots for WT in the absence of ligand and in the presence of both ligand types.** RMSD result in the absence of the ligand(A). The state of the protein in the dark green shaded interval in the previous figure (B). RMSD results in the presence of ligand class I (C). RMSD results in the presence of ligand class II (D). The conformation of the protein in the dark blue shaded area in the previous figure(E). L stands for the ligand class I and L* represents ligand class II.

17

As it has shown in the figure 3.2, the effect of the G330T mutation on the dynamics of the protein differs in the absence and presence of ligand.RMSD results show that the protein in Ligand class-I bound state has two different states, 1-120 ns and 120-200 ns (Fig. 3.2B). This difference is due to N terminus as it is in the case of WT. Until 120 ns, the N terminus is curled over the $\alpha2$ region and after 120ns the N terminus moves freely and is not fixed to any region of the protein (Fig. 3.2 C). In the presence of ligand class II, due to N terminus movements, protein has a distinct equilibrated state between 75 and 120ns (Fig. 3.2D). In this region, N terminus is packaged into the $\alpha2$ region of the protein (Fig. 3.2E).

For the G330T mutation, in the absence of ligand (Fig. 3.2A) and in the presence of ligand II, the average value of RMSD is between 3 and 4 Å. This result is also valid for repeated simulations. Taking these results into consideration, there is no significant difference between the RMSD averages, in the absence of ligands and their presence.
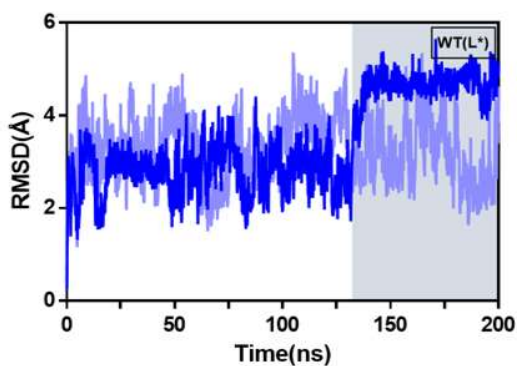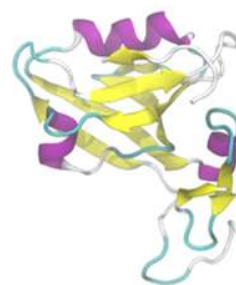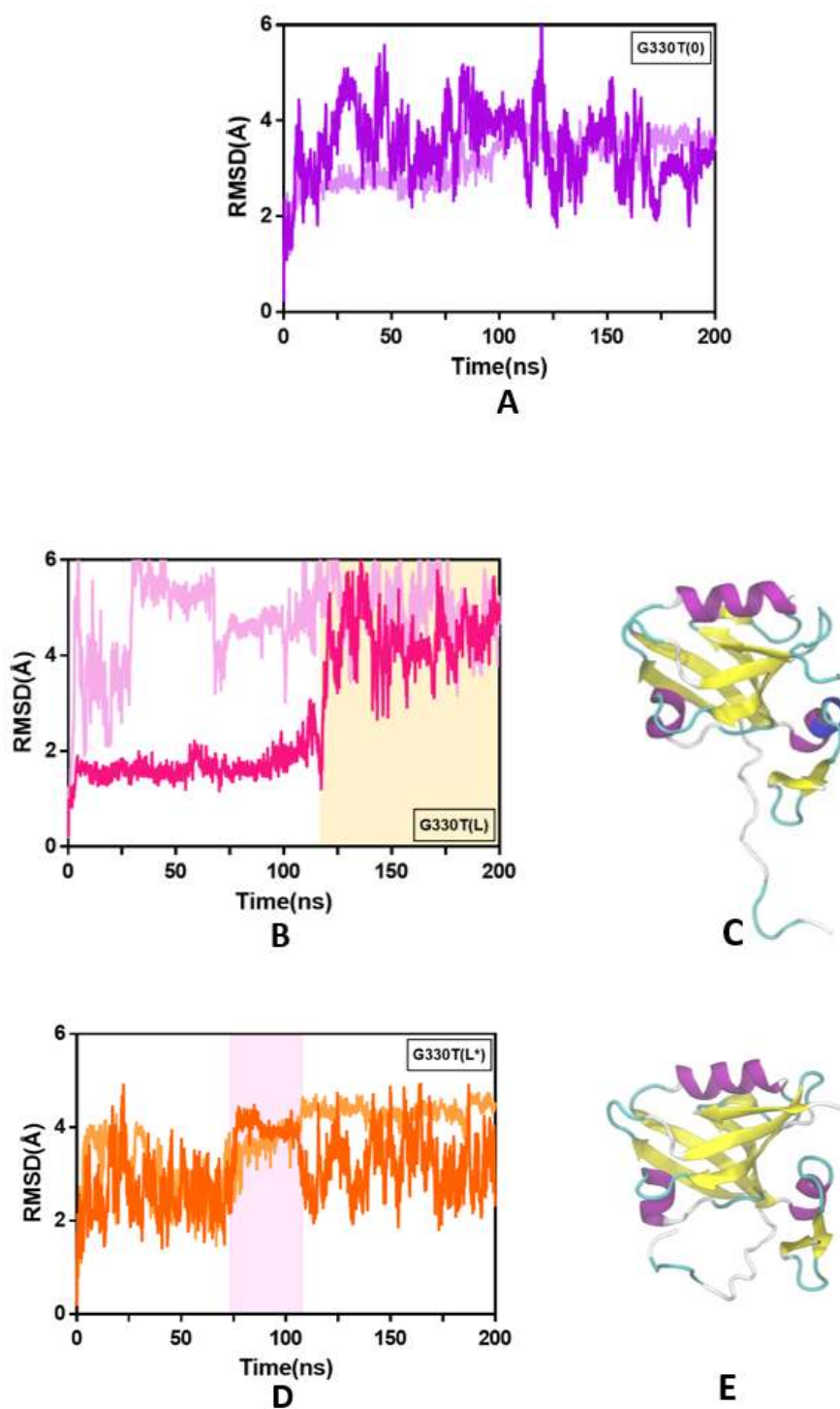
**Figure 3.2. RMSD results and sample snapshots for G330T in the absence of ligand and in the presence of both ligand types.** RMSD results in the absence of ligand (A). RMSD in the presence of ligand class I (B). The conformational state between 120-150 ns in the dark pink shaded region ligand class-I bound state (C). Ligand class-II bound condition RMSDs (D). Representation of N terminus position in the dark orange shaded region (E). L stands for the ligand class I and L* represents ligand class II.

19

For the H372A mutation case, in the absence of ligand, the repeated simulation has a slightly lower RMSD when it is compared to the first simulation, but the variation pattern of the protein is similar (Fig. 3.3A). In the case of ligand class-I bound case, RMSD results are nearly identical and have the most equilibrated condition (Fig. 3.3B). In terms of ligand class-II bound instance, even though the fluctuations of the RMSD is high, its average is between 2 and 3 Å (Fig. 3.3C).
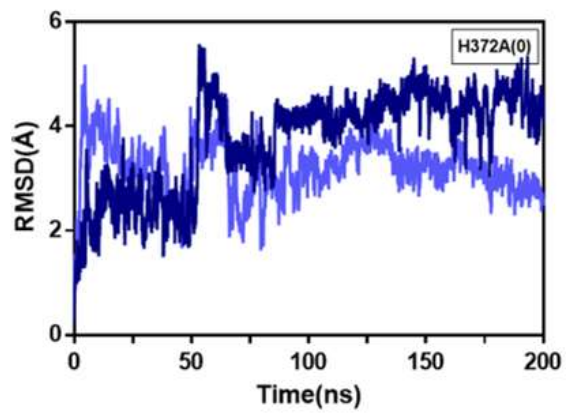
**A**
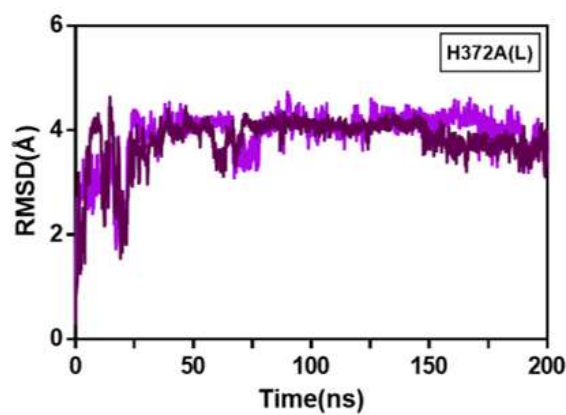


**B**
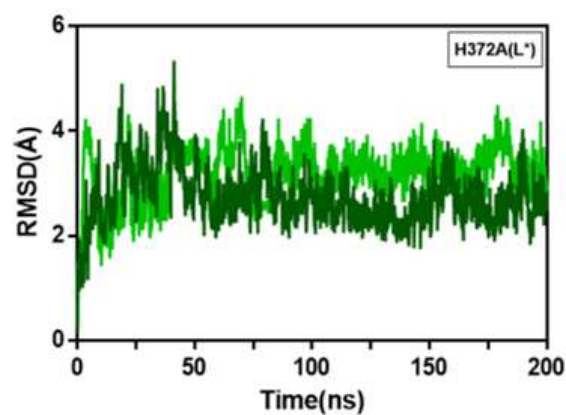


**C**

**Figure 3.3. RMSD results for H372A in the absence of ligand (A) and in the presence of ligand class-I (B) and II (C).** L stands for the ligand class I and L* represents ligand class II.

As demonstrated in the figure 3.4, the effect of the DM mutation on the dynamics of the protein differs in the absence and presence of ligand. For the double mutant (DM) cases, when there is no ligand attached to the protein, protein has shown frequent fluctuations during the whole trajectory, that situation is also true for the repeat simulation (Fig. 3.4A). When the average RMSD value is taken into consideration, it varies between 2.7 and 3 Å. Nevertheless, at ligand class I bound state, protein equilibrated a short while after simulation started and was fixed at 4 Å for both replica simulations (Fig. 3.4B). During the whole trajectory of ligand class-I bound state, N terminus is curled towards $\alpha2$ region (Fig. 3.4C) In the case of ligand class-II bound state, simulations have different states, but have a similar endpoint (Fig. 3.4D). It is observed that N terminus curled on itself between 70 and 90 ns in the presence of ligand class II (Fig. 3.4D). As a second conformation, N terminus is close to C terminus between 90 and 200 ns (Fig. 3.4E).
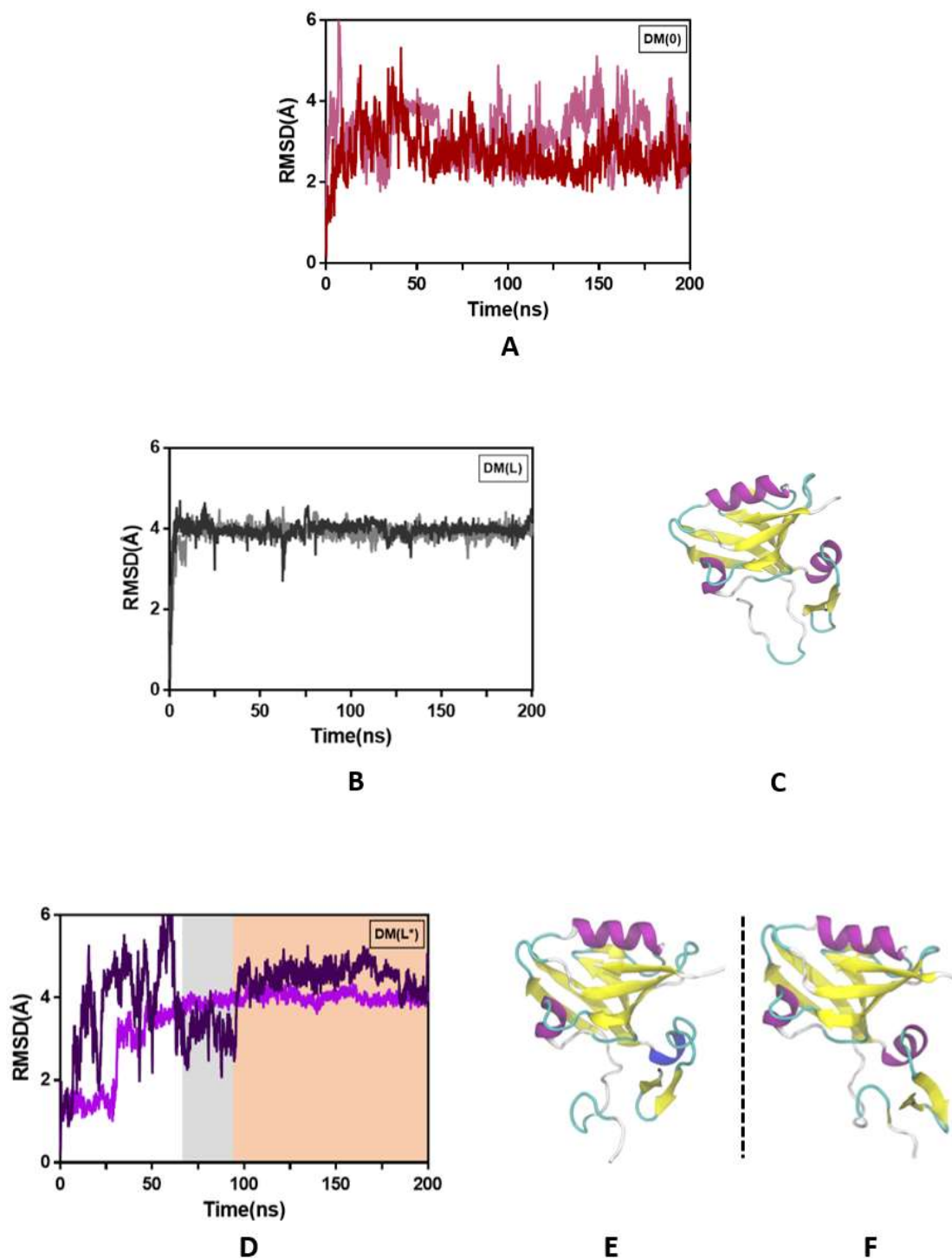
**Figure 3.4. RMSD results and sample snapshots for the G330T-H372A double mutation in the absence of ligand and in the presence of both ligand types.** RMSD results when ligands are absent (A). Ligand class-I presence effect on the RMSD (B). N terminus location in the presence of ligand class I (C). RMSD results in the presence of ligand class-II (D). Different conformations of the protein in ligand class-II bound state (E and F). L stands for the ligand class I and L* represents ligand class II.

## 3.2 Free Energy Perturbation Calculations

According to the results of RMSD, it was decided to take samples from 50ns, 100ns, 150ns and 200 ns for the H372A, G330T and DM cases from the first set of simulations (the ones shown in dark colors in Figure 3.2-3.4). To understand how reliable the results obtained are, the calculations were compared with the experimental results and were observed to be consistent.

### 3.2.1 G330T and DM cases

By taking the average of four FEP simulations, $\Delta\Delta G$ values are calculated in the presence of ligand class I and II for G330T and DM cases. FEP calculation results are then compared to experimental results.



**Figure 3.5 $K_d$ results in the presence of ligand class-I and II with G330T and DM.** L stands for the ligand class I and L* represents ligand class II.

The $K_d$ values that are obtained from the binding affinity experiments [18] (Fig. 3.7) were again calculated via *eq.7* and *eq.8*.
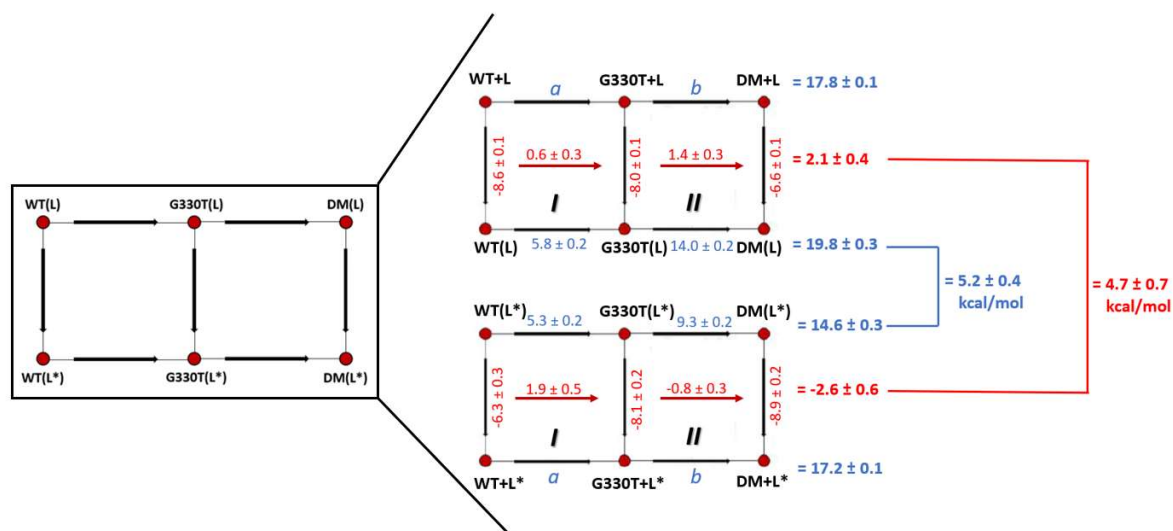
**Figure 3.6.** *ΔG* **and** *ΔΔG* **results obtained from FEP simulations(blue) and experiments(red).** Calculations are done by eq.7 and eq.8. L stands for the ligand class I and L* represents ligand class II.

More negative results are obtained when it is compared to H372A mutant conditions since class bridging mutant, G330T, allows binding of both ligand types [18]. Ligand class II has a more negative *ΔG* value which shows that in the presence of the G330T mutation, ligand class-II binding is more favorable.

While the experimental *ΔΔG* is -4.7 kcal/mol, the FEP results is -5.2 kcal/mol. The fact that the FEP and the experimental results from G330T the mutation case are similar to each other shows the sampling for FEP and MD simulations are reliable.

In the Figure 3.6, cycles of G330T and DM is represented in the absence and presence of ligand classes. a value in the absence of ligand class I cycle is 5.2 kcal/mol and b value in the absence of ligand class I cycle is 12.6 kcal/mol. In the ligand class II cycle, a is 7.2 kcal/mol and b is 10.0 kcal/mol. Due to the sum of a and b values are almost equal in the condition of absence of ligands cases, they cancel out each other.

### 3.2.2 H372A and DM cases

By taking the average of four FEP simulations, the *ΔΔG* values are calculated in the presence of ligand class I and II for H372A and DM cases and compared with the experimental results ($K_d$ are listed in Figure 3.5) [18].
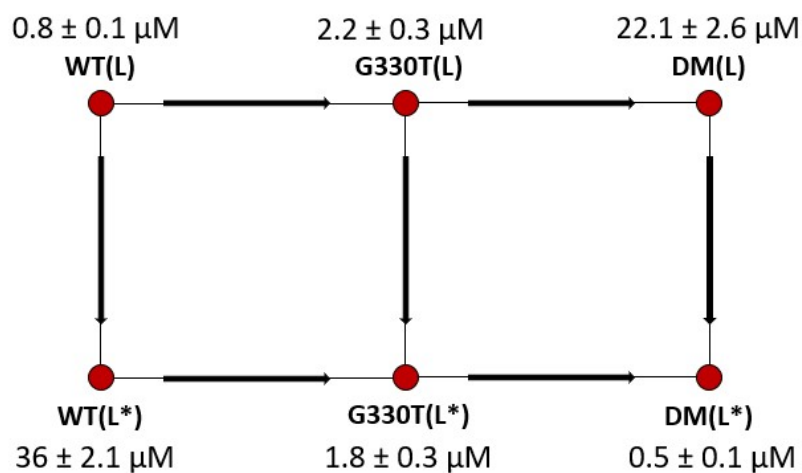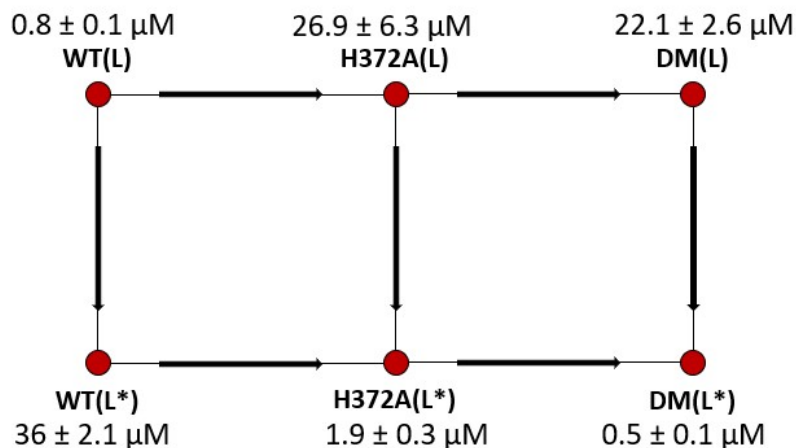
**Figure 3.7 $K_d$ results in the presence of ligand class-I and II with H372A and DM.**
L stands for the ligand class I and L* represents ligand class II.

From the $K_d$ values that are obtained from the binding affinity experiments [18] (Fig. 3.5) we calculate the relative free energy differences using the following relations:

$$\Delta G = RT \, ln(K_d) \tag{7}$$

$$\Delta\Delta G = RT \, ln\left(\frac{K_{d_2}}{K_{d_1}}\right) \tag{8}$$

The more negative $\Delta G$ is, the better the ligand binds. The most negative results were calculated in cases where ligand class II was bound under H372A and DM conditions, indicating that ligand class-II is preferred to ligand class-I. It has also been shown by FEP calculations that the H372A mutation does not allow class I binding.

While $\Delta\Delta G$ of experimental results is -4.7 kcal/mol, FEP results is -8.1 kcal/mol. The fact that the FEP and the experimental results display non-contradicting trends suggests that the sampling for FEP simulations and MD simulations are reliable.

In the Figure 3.8, cycles of H372A and DM is represented in the absence and presence of ligand classes. a value in the absence of ligand class I cycle is 13.0 kcal/mol and b value in the absence of ligand class I cycle is 4.9 kcal/mol. In the ligand class II cycle, a is 12.1 kcal/mol and b is 2.2 kcal/mol. Because of the close values of the sum of a and b in the thermodynamic cycle, they cancel out each other.
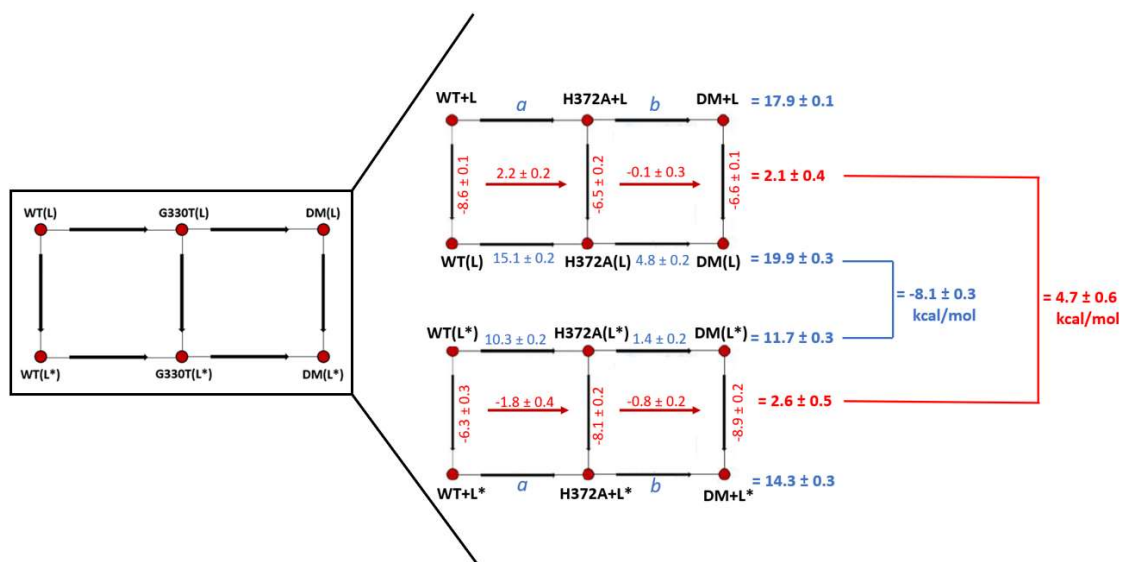


**Figure 3.8. *ΔG* results obtained from FEP simulations(blue) and experiments(red).** Calculations are done by eq.7 and eq.8. L stands for the ligand class I and L* represents ligand class II.

The fact that the results that are obtained from both the G330T and H372A mutations are close to the experimental results indicate that the examination of MD simulations is reasonable for determining the PDZ domain functions.

The values *a* and *b* for the free energy cost of the mutation in ligand free protein may be calculated. However, we have not gone forward with this calculation as they cancel out while we compare the binding free energy differences.

# Chapter 4

## Understanding the Effects of Mutations on the Dynamics of the Protein: Molecular Dynamics Simulations

## 4.1 Results

### 4.1.1 Entropic cost of mutations and ligand binding: RMSF Results

RMSF results show that there are certain regions of the protein that are affected due to ligand binding or mutations. Therefore, in the name of understanding the fluctuations completely, the cost of mutations and ligand binding are analyzed. We note that all the following results are averaged over 40 ns apart regions following equilibration of the both replica. Thus, the 80-120, 120-160 and 160-200 ns portions of the trajectories are utilized for each replica, resulting in a sample size of 6 sets. The displayed error bars are calculated accordingly.

### 4.1.1.1 Cost of Mutations

In this subsection, we analyze the effect of adding mutations to the wild type protein in terms of the fluctuation patterns of the residues.

As demonstrated in Figure 4.1, the absence of ligand case, DM mutation increased the fluctuations in most of the regions. However, it loses its domination, especially on the N and C terminus sites. On the other hand, H372A mutation does not affect the WT as much as the other mutations. On the N terminal region, the H372A mutation leads to an appreciable decrease when it is compared to other mutations. In contrast to the N terminus, the C terminus results indicate that all mutations have a fluctuation-decreasing effect.
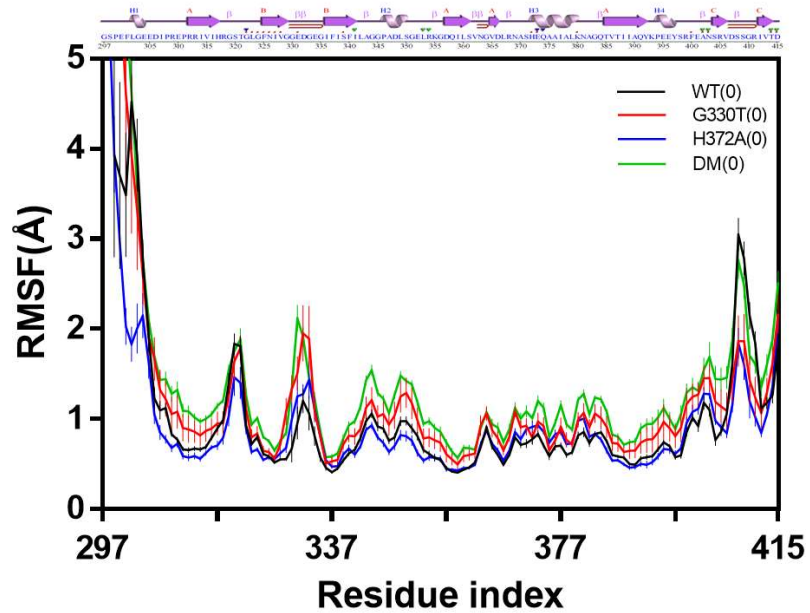
**Figure 4.1. RMSF results in the absence of ligand.**

Figure 4.2 represents the RMSF change regarding ligand class I presence. In terms of ligand class I binding, G330T has an additive effect of fluctuations on all the regions of the protein. Even though DM and H372A mutation inhibit the class-I type ligand binding, there is no significant difference from the WT in terms of residue fluctuations which bindins to the class-I ligand. Especially on the N terminus site, where the WT has been taken as a reference point, it can be concluded that G330T mutation results in an increase, but H372A and DM cause a decrease. Corresponding outcomes of H372A and DM mutation effects on the dynamics suggest that in the presence of class I ligand, H372A mutation is much more effective than the G330T mutation (Fig. 4.2).
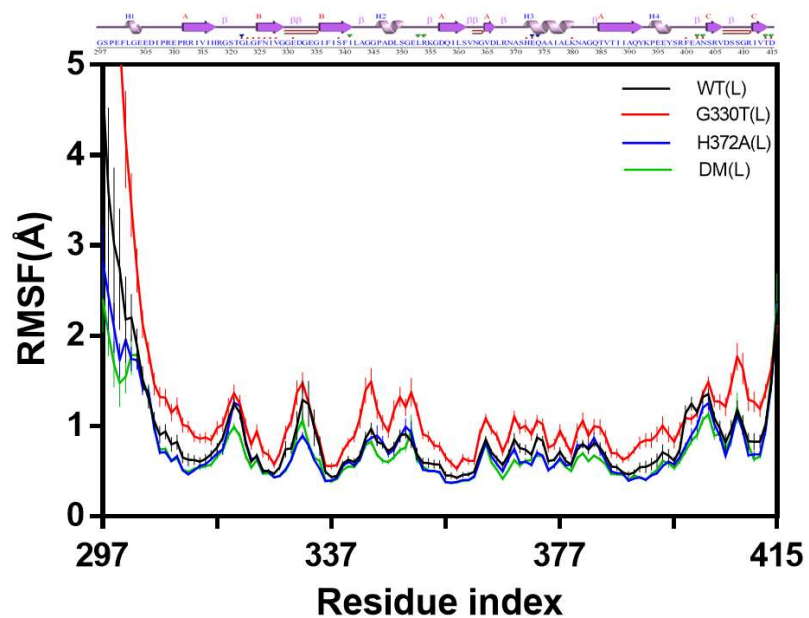
**Figure 4.2. RMSF results in the presence of ligand class I.**

As it has shown in Figure 4.3, ligand class II can bind to all mutation types except WT, but this did not cause a significant difference in RMSF results. WT has shown only a slight increase, but there are no significant differences. On the N terminus site, only the DM case has a decreasing effect on the fluctuations (Fig. 4.3).
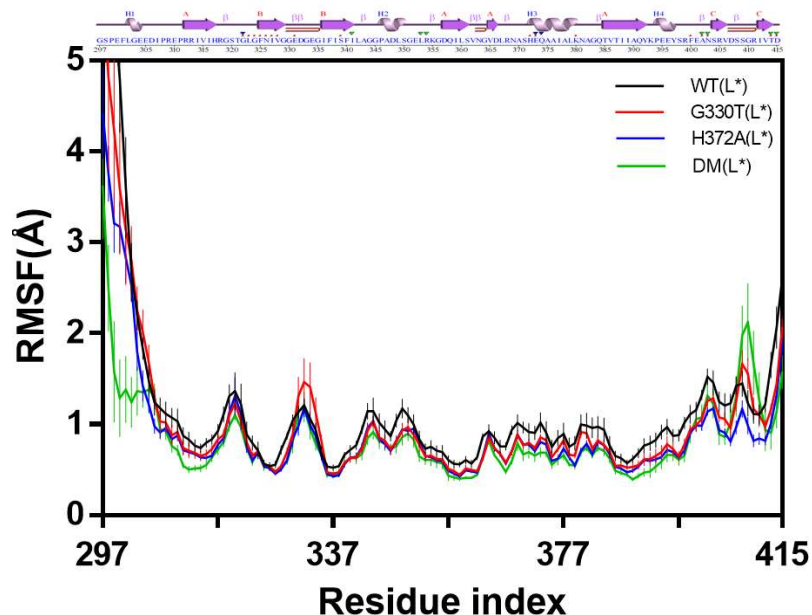
**Figure 4.3. RMSF results in the presence of ligand class II.**

## 4.1.1.2 Cost of ligand preference and ligand binding

In this subsection, we repolt the data of the previous one, but by grouping them in terms of the changes caused by ligand binding.

For WT, it is known that class I is preferable, but it does not bind to class II type ligand. As shown in figure 4.4, apart from the C terminus, the presence of ligand class II increases the fluctuations of the protein in all residues, but it is not a dramatic difference. On the C terminus site, the presence of both ligand reduce the fluctuation dramatically. It can be deduced that ligand presence decreases the variation and the presence of ligand class II slightly raises the fluctuations as a whole. Conversely, ligand class I binding does not cause any noticeable effect compared to the lack of ligand (Fig 4.4).
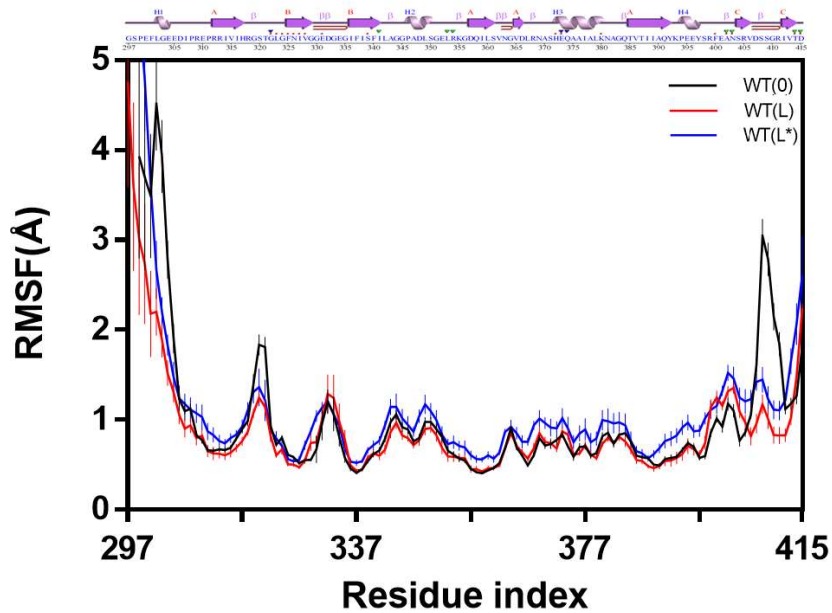
**Figure 4.4. RMSF results for WT case.**

G330T mutation allows to bind both classes of ligands, the only difference is their binding affinities. Regarding figure 4.5, If the absence of a ligand case is taken as a reference, ligand class I presence results in an insignificant rise. On the contrary, ligand class II binding, decreases the fluctuation in all residues, albeit in a minimal amount (Fig 4.5).
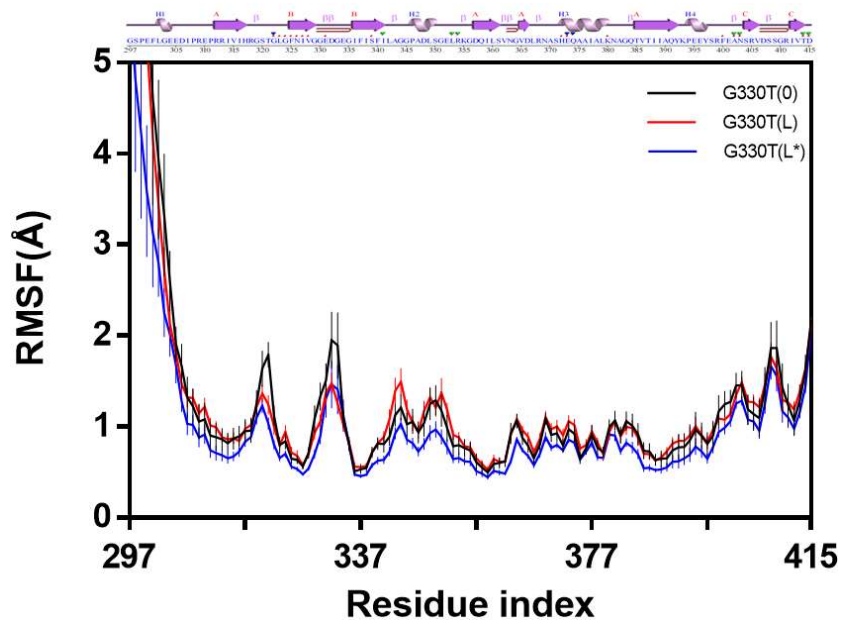
**Figure 4.5. RMSF results for the G330T case.**

H372A mutation inhibits the ligand class I binding, and only allows class II type ligand binding. However, as it has shown in the Figure 4.6, besides the N terminus, there is no consequential variance between the absence and the presence of the ligand. The presence of Ligand class I pacifies the N terminus when it is compared with the absence of ligand and presence of ligand class-II cases.
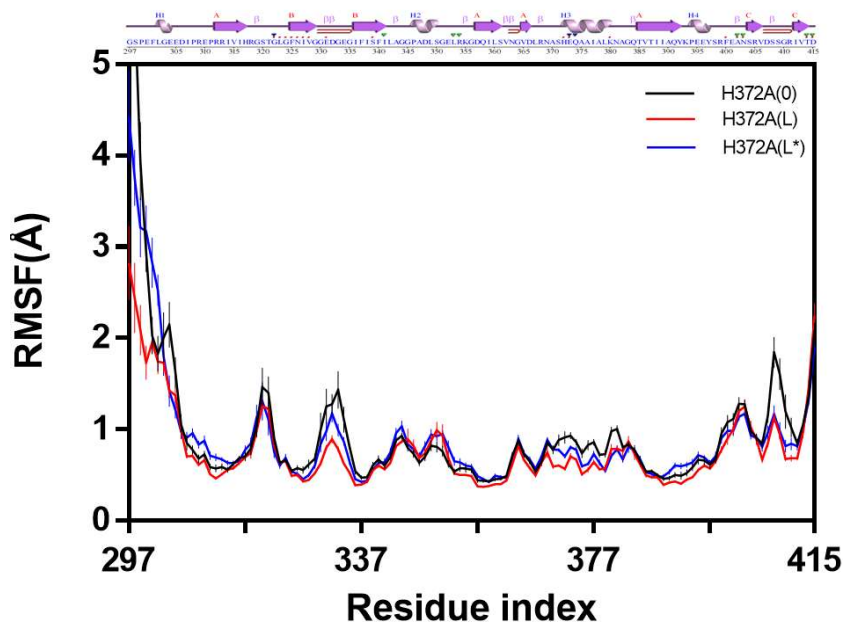


**Figure 4.6. RMSF results for H372A case.**

DM approves only ligand class-II binding. However, as the results have shown in Figure 4.7, that there is an equivalent effect of both classes I and II type ligand presence on the fluctuation patterns. Apparently, the presence of a ligand decreases the fluctuation of overall protein, especially at the N terminus. It can be concluded that, for the DM case, ligand presence ease the fluctuation for all residues, in contrast to ligand absence (Fig. 4.7).
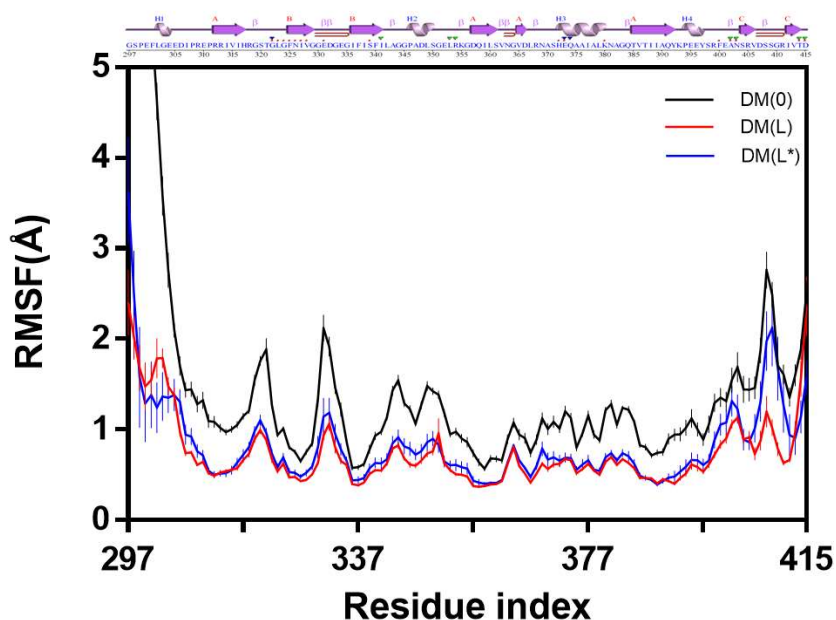


**Figure 4.7. RMSF results for DM case.**

**4.1.2 Deeper Understanding of Fluctuations: Cross-Correlation and PCA Results**

**4.1.2.1 Absence of Ligand**

To have the better understanding of mutations on the dynamics of the protein, WT is taken as a reference structure. G330T mutation increases both negative and positive correlation for the whole protein. H372A mutation is more similar to WT, but between residues 330-350 this mutation decreases the positive correlations. DM has equivalent results with the G330T mutant condition. When all results are taken into consideration, in the absence of a ligand G330T is more effective than H372A when they are present together as DM (Fig 4.8).

**Figure 4.8. Cross-correlation maps in the absence of ligand.** (For Cross-correlation matrix calculations, first 80 ns parts of the simulations are excluded. Calculations belong only to the first set of simulations.)

### 4.1.2.2 Presence of Ligand Class I

In the presence of ligand class-I, only the G330T mutation intensifies both positive and negative correlations. Both H372A mutation and DM increases the negative correlation in N terminus region and slightly increases the positive correlation, but both have comparable results with the WT condition. The resemblance in H372A and DM maps suggest that H372A mutation is more impressive than G330T mutation in the presence of a ligand which explains the inhibitory effect of H372A mutation towards the ligand class I binding (Fig. 4.9).



**Figure 4.9. Cross-correlation maps in the presence of ligand class-I.** (For Cross-correlation matrix calculations, first 80 ns parts of the simulations are excluded. Calculations belong only to the first set of simulations.)

The effectiveness of the H372A mutation in the presence of ligand class I is also observed via PCA results. We first note that the major mode of motion is always on the

N-terminus region. It is observed that H372A mutation introduces more fluctuation in the binding pocket which leads a distortion in the binding pocket and prevents the ligand binding (Fig. 4.10C). Because the H372A mutant is more effective than G330T mutant in the presence of ligand class I, binding inhibition is also effective in the DM condition (Fig. 4.10D). On the other hand, like the WT case, the binding pocket is stable in the condition of G330T mutant, so that ligand class I can bind to the protein (Fig. 4.10B).



**Figure 4.10. PCA results in the presence of ligand class-I.** (For PCA calculations, first 80 ns parts of the simulations are excluded. Only first modes are considered)

### 4.1.2.3 Presence of Ligand Class II

In ligand class II circumstance, G330T mutation increased both negative and positive correlation in the whole protein. Although not as much as the G330T mutation, the H372A mutation also increased positive and negative correlations. On the other hand, the DM condition shows the effects of both G330T and H372A mutation at the same time (Fig. 4.11).
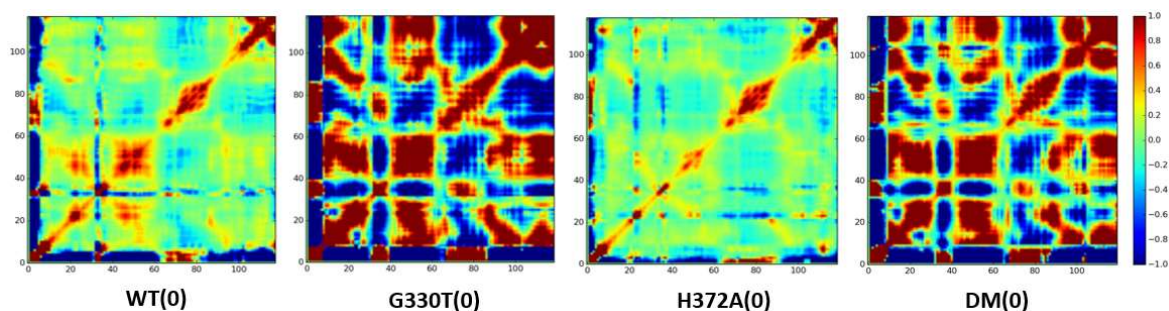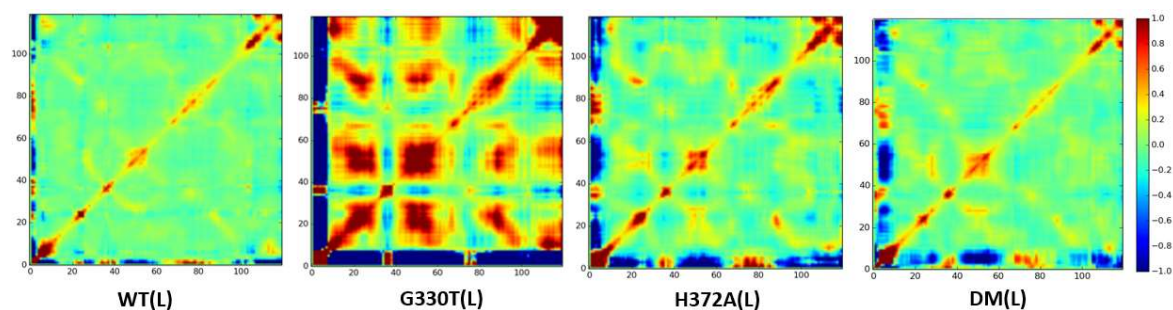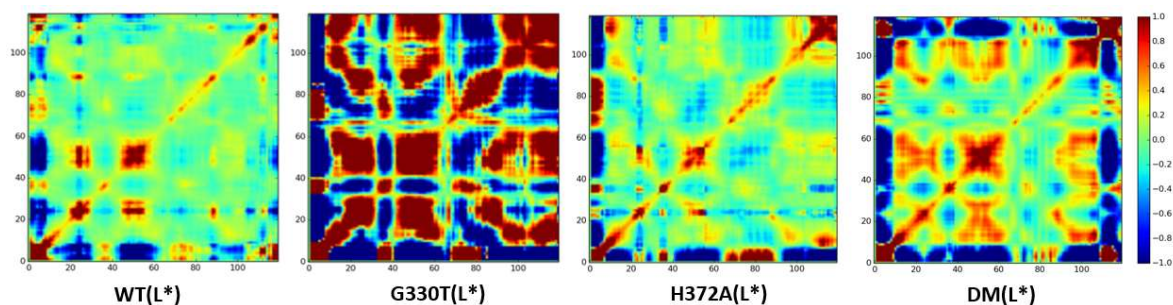
**Figure 4.11. Cross-correlation maps in the presence of ligand class-II.** (For Cross-correlation matrix calculations, first 80 ns parts of the simulations are excluded. Calculations belong only to the first set of simulations.)

The effects of mutants on the dynamics of the protein are observed in the 1$^{st}$ PCA. In the presence of ligand class-II, an intensive fluctuation on the binding site, especially at GLGF motif site, is observed for the WT condition (Fig. 4.12A). However, the disturbing fluctuation at the GLGF motif region is eased down with H372A and G330T mutation. Therefore, G330T, H372A and DM mutants all allow ligand class-II binding (Fig. 4.12B, C, D).
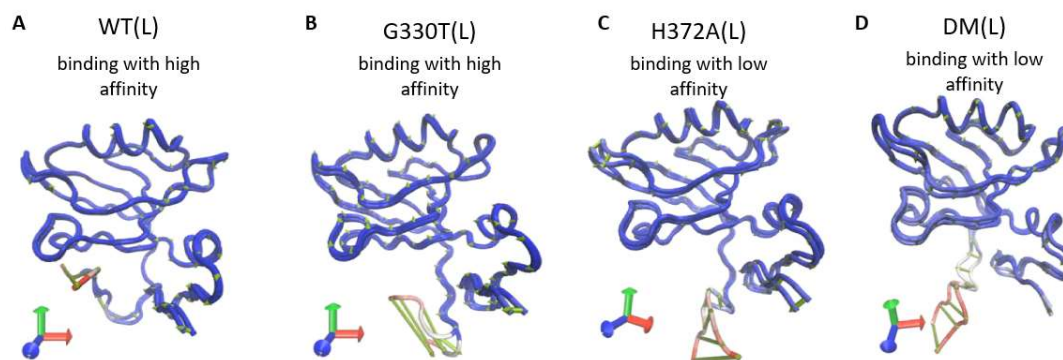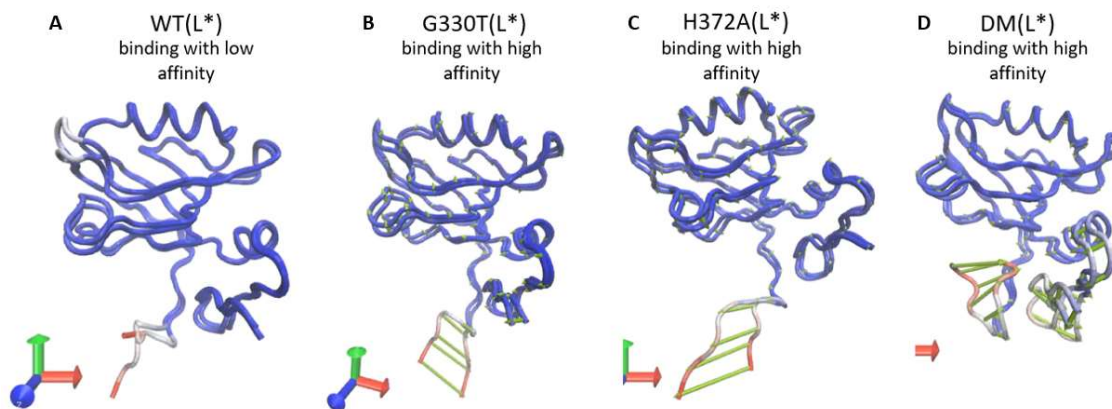


**Figure 4.12. PCA results in the presence of ligand class-II.** (For PCA calculations, first 80 ns parts of the simulations are excluded. Only first modes are considered)

**4.1.3 How stable are the ligands during the simulation? Hydrogen Bond Occupancies**

It has observed that, under different mutation presences the dynamics of the protein altered. Therefore, we also wanted to see the effect of the mutations on the stability of ligands.

Hydrogen bonds between the ligand and the bonding pocket were observed for 200 ns and the hydrogen bond occupancies were calculated as a percentage, with the length of the hydrogen bond being limited to a maximum of 3 Å.

Hydrogen bond occupations have changed over mutation and ligand type. In the presence of ligand class I, occupations of residues 327, 329 and 339 appear to be strengthened (Figure 4.13 B, C, D). However, occupations of hydrogen bonds of 323, 325 and 326 residues were significantly reduced compared to WT (Fig. 4.13 B, C, D). Given that only the WT and G330T mutants can bind ligand class I and their hydrogen bond occupancy results are equivalent, it is important to note that not only the amount of increase but also the residues in which the increase occurs matters.



**Figure 4.13. Hydrogen bond occupancy results under mutations and ligand class-I presence.** H-bond occupancies for WT (A). In the presence of G330T mutation with ligand class I (B). H372A mutation effect on the hydrogen bond occupancies (C). The effect of double mutation on the occupancies (D). The representation of the interactions between binding pocket and ligand class I (E).

In the case of ligand class-II, it is known that it does not bind to WT, but all introduced mutations allow the binding to this ligand [18]. From FEP calculations and ligand affinity assay results, it is known that the DM binds better to the class-II ligand. Hydrogen bond occupancy results also indicate that ligand class-II binds better to DM than others (Figure 4.14D). The hydrogen bond in WT shows weaker bonds at residue 329 when compared to other conditions (Figure 4.14A). This may indicate the importance of the location of hydrogen bonds.
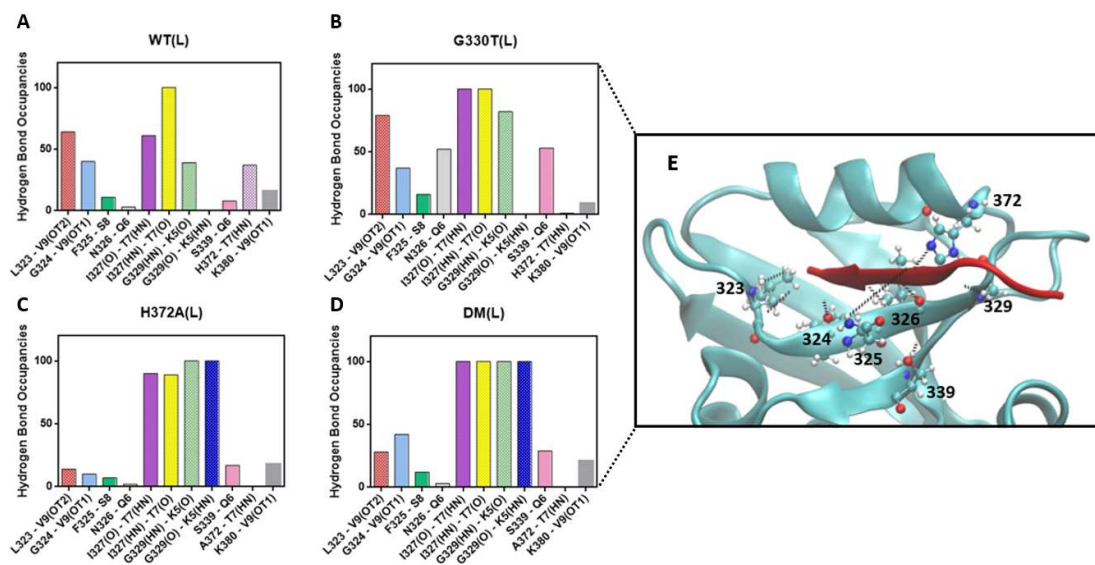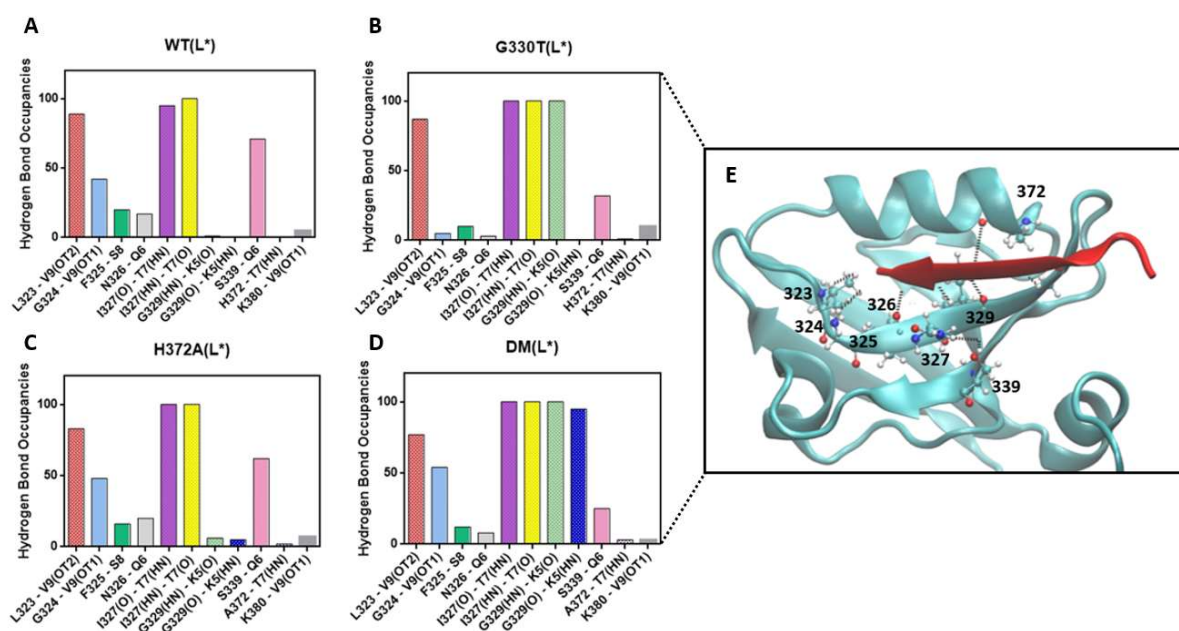


**Figure 4.14. Hydrogen bond occupancy results under mutations and ligand class-II presence.** H-bond occupancies for WT (A). In the presence of G330T mutation with ligand class II (B). H372A mutation effect on the hydrogen bond occupancies (C). The effect of double mutation on the occupancies (D). The representation of the interactions between binding pocket and ligand class II (E).

# Chapter 5

## A Simple Model to Explain Entropy-Enthalpy Compensation in PDZ Domain

PDZ domains are small proteins that are widely used for understanding the single domain allostery. Following the discovery that conformational change is not always necessary for allostery to occur [21][25], and with the first study of the allosteric pathways of the PDZ domain in 1999 [29], it became the favored model protein to understand single domain allostery.

Mutation may be thought of as a local perturbation, which may lead to a change in the ligand preference due to the alterations in the active site [28]. In this study, the third PDZ domain of the PSD-95 is used as a model molecule. Third PDZ domain has two possible ligand choices whose preferences may be modulated by perturbing two important residues, H372 and G330. It is observed that the H372A mutation, which is in the binding pocket and has a direct effect on the ligand binding, has an inhibitory effect on ligand class I [18] (Fig. 1.4). Therefore, it is known to be a 'class changing' mutation. On the other hand, G330T mutation allows the protein to bind both ligand class-I and II. Even though G330 is not located in the binding pocket, due to the indirect but significant effect of the mutation on this residue, it is considered as an 'allosteric effect'[18] (Fig. 1.4). Because G330T mutant allows binding both classes, it is called a 'class bridging' mutation[18].

FEP simulations and experimental results [18] show the relationship between mutations and ligand binding affinities (Fig. 3.6, 3.8). As it is excepted, H372A and DM allow the binding of ligand class-II but not ligand class-I. In addition, PDZ domains which have G330T mutant can bind both ligand classes.

MD analysis showed the dynamic effects of the mutations. The new movements that are introduced in the binding pocket, especially at the region of GLGF motif, due to H372A and G330T mutation are the determinants of binding. Both PCA and cross-correlation results are compatible with ligand binding affinity results.

We find that the binding affinities are intimately related to entropy enthalpy compensation. Recalling $\Delta G = \Delta H - T\Delta S$, we first crudely associate $\Delta H$ of binding by the forming and disappearing H bonds between ligand and protein in the binding pocket (Section 4.1.3) as these bonds are expected to make the largest change in the energetic contributions. On the other hand, the fluctuations may be translated into conformational entropy change through the sum of the eigenvalues of the matrices visualized in figures 4.9 and 4.10 for L and L* bound cases, respectively. $\Delta G$ has already been calculated, both experimentally and computationally (Section 3.2). These are summarized in Table 5.1.

**Table 5.1. $\Delta$H, $\Delta$S, $\Delta$G values from FEP calculations and MD trajectory analysis.**

|  | $K_d$ (μM) | $\Delta$G (kcal/mol) | Enthalpic contributions[a] | Entropic contributions[b] |
|---|---|---|---|---|
| WT(L) | 0.8 | -8.6 ± 0.1 | 48% | 83 |
| WT(L*) | 36 | -6.3 ± 0.1 | 65% | 111 |
| G330T(L) | 2.2 | -7.9 ± 0.2 | 85% | 379 |
| G330T(L*) | 1.8 | -8.1 ± 0.2 | 76% | 491 |
| H372A(L) | 26.9 | -6.5 ± 0.2 | 62% | 97 |
| H372A(L*) | 1.9 | -8.1 ± 0.2 | 73% | 186 |
| DM(L) | 22.1 | -6.6 ± 0.1 | 74% | 95 |
| DM(L*) | 0.5 | -8.9 ± 0.1 | 80% | 188 |

a) quantified by hydrogen bond occupancies
b) quantified by sum of the eigenvalues of correlation matrices

As it is mentioned in discussing the free energy perturbation calculations, the more negative $\Delta G$, the better the ligand binds. For the H372A mutation, L* binding is significantly better than L. In fact, this is directly associated with the increase in the number of hydrogen bond occurrences in the binding pocket (negative enthalpic contribution) and the increase in the sum of the eigenvalues (positive entropic contribution). The same is true for the DM. On the other hand, the case of the G330T mutation which allows both ligands to bind is subtler. While the change in the hydrogen bonding occupancies with respect to the WT is larger for the class I ligand than that of class II ligand, the conformational entropy change for the latter is dominant. Thus, the G330T mutant prefers to bind L due mainly to enthalpic gain, it bind L* due to the

entropic gain. We note that the N terminus has a great contribution in the allosteric modulation of ligand binding in the PDZ domains. Thus, rather than a pathway view that analyses the changes between the effector and binding sites in allosteric models, we suggest that the ensemble view of allostery provides a better description of site-to-site coupling [43].

# References

[1]     V. K. Subbaiah, C. Kranjec, M. Thomas, and L. Banks, "PDZ domains: the building blocks regulating tumorigenesis," *Biochem. J.*, vol. 439, no. 2, pp. 195–205, 2011.

[2]     F. Ye and M. Zhang, "Structures and target recognition modes of PDZ domains: recurring themes and emerging pictures," *Biochem. J.*, vol. 455, no. 1, pp. 1–14, 2013.

[3]     R. Tonikian *et al.*, "A specificity map for the PDZ domain family," *PLoS Biol.*, vol. 6, no. 9, pp. 2043–2059, 2008.

[4]     K. O. Cho, C. A. Hunt, and M. B. Kennedy, "The rat brain postsynaptic density fraction contains a homolog of the drosophila discs-large tumor suppressor protein," *Neuron*, vol. 9, no. 5, pp. 929–942, 1992.

[5]     S. J. Broderick, M.J.F.;Winder, "Spectrin, α-Actinin, and Dystrophin," vol. 70, pp. 203–246, 2005.

[6]     J. Zhu, Y. Shang, C. Xia, W. Wang, W. Wen, and M. Zhang, "Guanylate kinase domains of the MAGUK family scaffold proteins as specific phospho-protein-binding modules," *EMBO J.*, vol. 30, no. 24, pp. 4986–4997, 2011.

[7]     J. R. King, Jonathan M; Tan, Christine Joy P ; Thomason, Nicole C; White, Addison R; Shen, Le ; Turner, "Zonula occludens-1 ZU5 domain contributes essential stabilizing interactions at the tight junction," *Faseb J.*, vol. 30, 2016.

[8]     M. Zhang and W. Wang, "Organization of signaling complexes by PDZ-domain scaffold proteins," *Acc. Chem. Res.*, vol. 36, no. 7, pp. 530–538, 2003.

[9]     H.-J. Lee and J. J. Zheng, "PDZ domains and their binding partners: structure, specificity, and modification," *Cell Commun. Signal.*, vol. 8, no. 1, pp. 8, 2010.

[10]    G. P. Manjunath, P. L. Ramanujam, and S. Galande, "Structure function relations in PDZ-domain-containing proteins: Implications for protein networks in cellular signalling," *J. Biosci.*, vol. 43, no. 1, pp. 155–171, 2017.

[11]    D. A. Doyle, A. Lee, J. Lewis, E. Kim, M. Sheng, and R. MacKinnon, "Crystal structures of a complexed and peptide-free membrane protein- binding domain: Molecular basis of peptide recognition by PDZ," *Cell*, vol. 85, no. 7, pp. 1067–1076, 1996.

[12]    C. M. Petit, J. Zhang, P. J. Sapienza, E. J. Fuentes, and A. L. Lee, "Hidden dynamic allostery in a PDZ domain," *Proc. Natl. Acad. Sci.*, vol. 106, no. 43, pp. 18249–18254, 2009.

[13]    S. Mostarda, D. Gfeller, and F. Rao, "Beyond the binding site: The role of the β2 - β3 loop and extra-domain structures in PDZ domains," *PLoS Comput. Biol.*, vol. 8, no. 3, 2012.

[14]    H. J. Lee and J. J. Zheng, "PDZ domains and their binding partners: Structure, specificity, and modification," *Cell Commun. Signal.*, vol. 8, pp. 1–18, 2010.

[15]    B. Z. Harris and W. a Lim, "Mechanism and role of PDZ domains in signaling complex assembly.," *J. Cell Sci.*, vol. 114, no. Pt 18, pp. 3219–3231, 2001.

[16]  T. T. Tani and A. M. Mercurio, "PDZ Interaction Sites in Integrin α Subunits. TIP-2/GIPC binds to a type I recognition sequence in $\alpha_{6A}/ \alpha_5$ and a novel sequence in $\alpha_{6B}$," *J. Biol. Chem.*, vol. 276, no. 39, pp. 36535–36542, 2001.

[17]  R. Ranganathan and E. Ross, "PDZ domain proteins: scaffolds for signaling complexes," *Curr Biol*, vol. 7, no. 12, pp. 770-773, 1997.

[18]  A. S. Raman, K. I. White, and R. Ranganathan, "Origins of Allostery and Evolvability in Proteins: A Case Study," *Cell*, vol. 166, no. 2, pp. 468–481, 2016.

[19]  D. Kern and E. R. P. Zuiderweg, "The role of dynamics in allosteric regulation," *Curr. Opin. Struct. Biol.*, vol. 13, no. 6, pp. 748–757, 2003.

[20]  H. N. Motlagh, J. O. Wrabl, J. Li, and V. J. Hilser, "The ensemble nature of allostery," *Nature*, vol. 508, no. 7496, pp. 331–339, 2014.

[21]  J. Liu and R. Nussinov, "Allostery: An Overview of Its History, Concepts, Methods, and Applications," *PLoS Comput. Biol.*, vol. 12, no. 6, pp. 3–7, 2016.

[22]  N. M. Goodey and S. J. Benkovic, "Allosteric regulation and catalysis emerge via a common route," vol. 4, no. 8, pp. 474–482, 2008.

[23]  A. A. S. T. Ribeiro and V. Ortiz, "A Chemical Perspective on Allostery," *Chem. Rev.*, vol. 116, no. 11, pp. 6488–6502, 2016.

[24]  R. Nussinov, "Introduction to Protein Ensembles and Allostery," *Chem. Rev.*, vol. 116, no. 11, pp. 6263–6266, 2016.

[25]  A. Cooper and D. T. F. Dryden, "Allostery without conformational change: A plausible m o d e l," *European Biophysics*, pp. 103–109, 1984.

[26]  A. Kumawat and S. Chakrabarty, "Hidden electrostatic basis of dynamic allostery in a PDZ domain," *Proc. Natl. Acad. Sci.*, vol. 114, no. 29, pp. E5825–E5834, 2017.

[27]  A. Weinkama, Patrick; Chenb, Yao Chi; Ponsc, Jaume; Sali, "Impact of mutations on the allosteric conformational equilibrium," *J Mol Biol.*, vol. 425, no. 3, pp. 647–661, 2013.

[28]  R. Nussinov and C. Tsai, "Review Allostery in Disease and in Drug Discovery," *Cell*, vol. 153, no. 2, pp. 293–305, 2013.

[29]  S. W. Lockless, "Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families," *Science*, vol. 286, no. 5438, pp. 295–299, 2009.

[30]  K. Humphrey, William ; Dalke, Andrew; Schulten, "VMD: Visual molecular dynamics," *J Mol Graph*, vol. 14, no. 1, pp. 33–38, 1996.

[31]  A. Levitt, Michael ; Warshel, "Computer Simulation of Protein Folding," *Nature*, vol. 253, no. 1, 1975.

[32]  S. K. D.Jefferies, "Molecular Simulations of Complex Membrane Models," in *Modeling of Microscale Transport in Biological Processes*, 2017, pp. 1–18.

[33]  D. Vepa, Vidana; Winkler, "Computational Approaches," in *Adverse Effects of Engineered Nanomaterials*, 2012, pp. 85–96.

[34]  T. Darden, L. Perera, L. Li, and P. Lee, "New tricks for modelers from the crystallography toolkit: The particle mesh Ewald algorithm and its use in nucleic acid simulations," *Structure*, vol. 7, no. 3, pp. 55–60, 1999.

[35]  W. Schreiner, R. Karch, B. Knapp, and N. Ilieva, "Relaxation Estimation of RMSD in Molecular Dynamics Immunosimulations," vol. 2012, 2012.

[36]  V. Benson,Noah C. and Daggett, "A Comparison of Multiscale Methods for the Analysis of Molecular Dynamics Simulations," *J Phys Chem B.*, vol. 116, no. 29, pp. 8722–8731, 2013.

[37]  N. Ota and D. A. Agard, "Intramolecular Signaling Pathways Revealed by Modeling Anisotropic Thermal Diffusion," *J Mol Biol.,* vol.2*,* pp. 345–354, 2005.

[38]  I. Feierberg, I. Feierberg, V. B. Luzhkov, and V. B. Luzhkov, "Free Energy Calculations and Ligand Binding," *Adv. Protein Chem.*, pp. 123–158, 2003.

[39]  T. Steinbrecher, R. Abel, A. Clark, and R. Friesner, "Free Energy Perturbation Calculations of the Thermodynamics of Protein Side-Chain Mutations," *J. Mol. Biol.*, vol. 429, no. 7, pp. 923–929, 2017.

[40]  R. W. Zwanzig, "High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases*," *J. Chem. Phys.*, vol. 22, pp. 1420-1426, 1954.

[41]  N. Lu, D. A. Kofke, and T. B. Woolf, "Improving the Efficiency and Reliability of Free Energy Perturbation Calculations Using Overlap Sampling Methods," vol. 14677, 2003.

[42]  P. Liu, F. Dehez, W. Cai, and C. Chipot, "A toolkit for the analysis of free-energy perturbation calculations," *J. Chem. Theory Comput.*, vol. 8, no. 8, pp. 2606–2616, 2012.

[43]  V. J. Hilser, J. O. Wrabl, and H. N. Motlagh, "Structural and Energetic Basis of Allostery." *Annu. Rev. Biophys.*, vol.41, pp.585-609, 2012.