

KADİR HAS UNIVERSITY  
SCHOOL OF GRADUATE SCHOOLS  
PROGRAM OF COMPUTATIONAL BIOLOGY AND BIOINFORMATICS

**INVESTIGATION OF SPECIES-SPECIFIC ALLOSTERIC  
BINDING SITES IN GLYCOLYTIC ENZYMES via  
AlloSigMA and MOLECULAR DYNAMICS  
SIMULATIONS**

METEHAN ÇELEBİ

MASTER'S THESIS

ISTANBUL, JUNE, 2021

Metehan Celebi

M. S. Thesis

2021



**INVESTIGATION OF SPECIES-SPECIFIC ALLOSTERIC  
BINDING SITES IN GLYCOLYTIC ENZYMES via  
AlloSigMA and MOLECULAR DYNAMICS  
SIMULATIONS**

METEHAN ÇELEBİ

MASTER'S THESIS

Submitted to the School of Graduate Studies of Kadir Has University in partial fulfillment of the requirements for the degree of Master's in the Program of Computational Biology and Bioinformatics

ISTANBUL, JUNE, 2021

DECLARATION OF RESEARCH ETHICS  
METHODS OF DISSEMINATION

I, METEHAN ÇELEBİ, hereby declare that;

- This Master's Thesis is my own original work and that due references have been appropriately provided on all supporting literature and resources;
- This Master's Thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institute;
- I have followed "Kadir Has University Academic Ethics Principles" prepared in accordance with "The Council of Higher Education's Ethical Conduct Principles".

In addition, I understand that any false claim in respect of this work will result in disciplinary action in accordance with university regulations.

Furthermore, both printed and electronic copies of my work will be kept in Kadir Has University Information Center under the following condition as indicated below:

- The full content of my thesis/project will be accessible from everywhere by all means.

METEHAN ÇELEBİ

---

KADİR HAS UNIVERSITY  
SCHOOL OF GRADUTE STUDIES

**ACCEPTANCE AND APPROVAL**

This work entitled **INVESTIGATION OF SPECIES-SPECIFIC ALLOSTERIC BINDING SITES IN GLYCOLYTIC ENZYMES via AlloSigMA and MOLECULAR DYNAMICS SIMULATIONS** has been judged to be successful at the defense exam held on ----- and accepted by our jury as **MASTER THESIS**.

APPROVAED BY:

Prof. Dr. E. Demet Akdoğan (Advisor)

Kadir Has University

\_\_\_\_\_

Prof. Dr. Kemal Yelekçi

Kadir Has University

\_\_\_\_\_

Asst. Prof. Özge Kürkçüoğlu Levitaş

Istanbul Technical University

\_\_\_\_\_

I certify that the above signatures belong to the faculty members named above.

\_\_\_\_\_  
(Title, Name and Surname)

Dean of School of Graduate Studies

DATE OF APPROVAL: (Day/Month/Year)



## TABLE OF CONTENTS

<b>ABSTRACT</b> .....	<b>i</b>
<b>ÖZET</b> .....	<b>iii</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>v</b>
<b>LIST OF TABLES</b> .....	<b>vii</b>
<b>LIST OF FIGURES</b> .....	<b>viii</b>
<b>LIST OF SYMBOLS/ABBREVIATIONS</b> .....	<b>xii</b>
<b>1. INTRODUCTION</b> .....	<b>1</b>
1.1 Glycolysis .....	1
1.2 Species – Specific Allosteric Drugs .....	3
1.3 Structure and Function of Phosphofructokinase .....	4
1.4 Allosteric Regulation of Phosphofructokinase .....	5
<b>2. MATERIALS AND METHODS</b> .....	<b>8</b>
<b>2.1 AlloSigma Experiments</b> .....	<b>8</b>
2.1.1 System Preparation .....	8
<b>2.2 Molecular Dynamics Simulations of Phosphofructokinase</b> .....	<b>13</b>
2.2.1 Completing Missing Residues on Phosphofructokinase .....	15
2.2.2 Molecular Dynamics Simulation System and Configuration Preparation .....	16
2.2.3 System Preparation for Constrained Molecular Dynamics Simulations .....	18
<b>3. RESULTS AND DISCUSSION</b> .....	<b>21</b>
<b>3.1 AlloSigma Experiments</b> .....	<b>21</b>
3.1.1 SaPFK AlloSigma Experiments .....	21
3.1.2 TbPFK AlloSigma Experiments .....	25
3.1.3 hPFK AlloSigma Experiments .....	28
3.1.4 GADPH AlloSigma Experiments .....	30
3.1.5 PK AlloSigma Experiments .....	33
<b>3.2 Molecular Dynamics Simulations of Phosphofructokinase</b> .....	<b>40</b>
3.2.1 Change in Constrained Distances and System’s Energy for MD Runs .....	40
3.2.2 RMSD and RMSF Calculations .....	41
3.2.3 Principal Component Analysis (PCA) .....	50

3.2.4 Determination of Three Orthogonal Principal Axes for Chains and Domains in PFK .....	67
3.2.5 The Mean Square Distance Fluctuation Analysis .....	76
<b>4. CONCLUSION AND FUTURE WORK .....</b>	<b>79</b>
<b>REFERENCES .....</b>	<b>83</b>
<b>CURRICULUM VITAE .....</b>	<b>88</b>
<b>APPENDIX A .....</b>	<b>89</b>
<b>A.1 Prepared Configuration Files for Apo and Constrained MD Simulations.....</b>	<b>89</b>
<b>A.2 AlloSigma Experiments on other proposed allosteric sites for TbPFK. ....</b>	<b>92</b>
<b>A.3 RMSD Profiles for 2<sup>nd</sup> and 3<sup>rd</sup> Runs of Apo and Constrained Molecular Dynamics Simulations.....</b>	<b>98</b>
<b>A.3 Used Commands for Principal Component Analysis (PCA) Calculations. ....</b>	<b>102</b>



# INVESTIGATION OF SPECIES-SPECIFIC ALLOSTERIC BINDING SITES IN GLYCOLYTIC ENZYMES via AlloSigMA and MOLECULAR DYNAMICS SIMULATIONS

## ABSTRACT

In previous studies of our research group, allosteric sites have been proposed to be used as drug targets in species-specific drug design studies for phosphofructokinase (PFK), glyceraldehyde-3 phosphate dehydrogenase (GADPH) and pyruvate kinase (PK) that belong to three species bacteria, parasite, and human and are essential enzymes in the glycolytic pathway. In this thesis, they were further investigated by various tools such as AlloSigMA and MD simulations. In addition to proposed allosteric sites, known allosteric sites reported by experimental studies for *S. aureus* PFK and PK enzymes were also investigated. In the first part, AlloSigMA was used to perturb the residues at the proposed and/or known allosteric sites in order to evaluate their allosteric capacities and their effects on protein dynamics. Accordingly, a reduced dynamics in the catalytic sites indicating allosteric inhibition was observed for most of the proposed allosteric sites whereas either an opposite or no effect was observed for known allosteric sites. In addition, partial allosteric inhibition was observed for some of the proposed allosteric sites in human species.

In the second part of this thesis, Molecular Dynamics simulations of a total of nine runs, each 100 ns long, were performed for *S. aureus* phosphofructokinase enzyme in apo and constrained states which incorporated bond restraints at the proposed and known allosteric sites. Here, the goal was to investigate the effect of restraints on the protein's global dynamics. RMSD/RMSF, principal component analysis, the change in orthogonal principal axes, and the mean square distance fluctuation between each pair of residues were determined. According to PCA analysis, increase in the correlation of positional fluctuations between each pair of residues in the chains and domains were observed. Based on the mean square distance fluctuation between residue pairs, each dimer started to communicate more within itself when switched to constrained state.

**Keywords:** Allosteric Sites, Phosphofructokinase, glyceraldehyde-3-phosphate dehydrogenase, Pyruvate Kinase, Molecular Dynamics Simulations, Drug Design, AlloSigMA



# GLİKOLİTİK ENZİMLERDE TÜRE ÖZGÜ ALLOSTERİK BAĞLANMA BÖLGELERİNİN AlloSigMA ARACI VE MOLEKÜLER DİNAMİK SİMÜLASYONLARI İLE İNCELENMESİ

## ÖZET

Araştırma grubumuzun önceki çalışmalarında, türe özgü ilaç tasarım çalışmalarında hedef ilaç bölgesi olarak kullanılması amacıyla bakteri, parazit ve insan türlerine ait glikolitik yolaktaki temel enzimler olan fosfofruktokinaz (PFK), gliseraldehit 3-fosfat dehidrogenaz (GADPH) ve pürivat kinaz (PK) enzimleri için allosterik bölgeler önerilmiştir. Bu tezde, AlloSigMA aracı ve MD simülasyonları gibi çeşitli araçlarla daha detaylı araştırıldılar. Önerilen allosterik bölgelere ek olarak, *S. aureus* PFK ve PK enzimleri için deneysel çalışmalarla bildirilen bilinen allosterik bölgeler de araştırıldı. İlk bölümde, allosterik kapasitelerini ve protein dinamiği üzerindeki etkilerini değerlendirmek için önerilen ve/veya bilinen allosterik bölgelerdeki rezidülerin dinamiğini bozmak için AlloSigMA aracı kullanıldı. Buna göre, önerilen allosterik bölgelerin çoğu için katalitik bölgelerde allosterik inhibisyonu gösteren dinamikte bir azalma gözlemlenirken, bilinen allosterik bölgeler için bir ters etki veya hiç etki gözlenmedi. Ek olarak, insan türlerinde önerilen allosterik bölgelerin bazıları için kısmi allosterik inhibisyon gözlemlendi.

Bu tezin ikinci bölümünde, *S. aureus* fosfofruktokinaz enzimi için önerilen ve bilinen allosterik bölgelerde bağ kısıtlamalarını içeren apo ve kısıtlı durumlarda her biri 100 ns uzunluğunda toplam dokuz koşu Moleküler Dinamik simülasyonları gerçekleştirilmiştir. Burada amaç, kısıtlamaların proteinin global dinamikleri üzerindeki etkisini araştırmaktır. RMSD/RMSF, temel bileşen analizi, ortogonal ana eksenlerdeki değişiklik ve her bir rezidü çifti arasındaki ortalama mesafe dalgalanması belirlendi. PCA analizine göre, zincirlerdeki ve domainlerdeki her bir rezidü çifti arasındaki konumsal dalgalanmaların korelasyonunda artış gözlemlendi. Rezidü çiftleri arasındaki ortalama kare mesafe dalgalanmasına bağlı olarak, her dimer kısıtlı duruma geçtiğinde kendi içinde daha fazla iletişim kurmaya başladı.

**Anahtar Sözcükler:** Allosterik bölgeler, Fosfofruktokinaz, Gliseraldehit-3-fosfat dehidrogenaz, Pirüvat Kinaz, Moleküler Dinamik Simülasyonları, İlaç Tasarımı, AlloSigMA



## ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere appreciation to my thesis adviser, Prof. Demet AKDOĞAN for her guidance, especially for her support and patience during COVID-19 pandemic. The goal of this thesis would not have been possible during the COVID-19 pandemic without her encouragement.

I am grateful to my colleague dear Reyhan Metin who was always there to help and support me. I would like to thank our research group members for their help. I am grateful to my friends and colleagues Merve Ayyıldız, Serkan Çeliker, Mehmet Fatih Özhelvacı. I would also like to express my special thanks to my dear friend Ezgi Mehterođlu, who is always there for me with her endless love, support and motivation.

Last but not least, I am grateful to my parents for their endless love and support in all circumstances throughout my life.

I would like to acknowledge Kadir Has University for my scholarship.

This work has been supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under the grant No: 218M320.



To my parents...

## LIST OF TABLES

Table 2.1: PDB ids of X-ray crystal structures used in AlloSigma experiments.....	8
Table 2.2: Experiment numbers and corresponding residues for the perturbation sites in PFK. ....	10
Table 2.3: Experiment numbers and corresponding residues for the perturbation sites in GADPH.....	11
Table 2.4: Experiment numbers and corresponding residues for the perturbation sites in PK.....	12
Table 2.5: Selected residue pairs for constrained MD simulations.....	20
Table 3.1: The percentage explanation of the total motion of the modes obtained from the principal component analysis for three different simulation runs in the phosphofructokinase enzyme.....	52

## LIST OF FIGURES

Figure 1.1: Enzymes in glycolytic pathway. The figure is adapted from Ayyildiz Master's thesis.....	2
Figure 1.2: Allosteric drug mechanism. The figure was adapted from (Cheng and Jiang 2019). .....	4
Figure 1.3: X-Ray crystallographic structure of <i>Sa</i> PFK (PDB ID: 5XZ7). (A) <i>Sa</i> PFK's two dimers as AB and CD. (B) Tetrameric structure of <i>Sa</i> PFK. Larger domains and smaller domains in all chains are colored purple, blue, and light orange, respectively....	5
Figure 2.1: A chart showing how a molecular dynamics simulation is performed.....	14
Figure 2.2: CHARMM Potential function (Brooks et al., 1983) used to approximate the atomic forces that govern molecular movement. The figure adapted from (Durrant & McCammon, 2011). .....	14
Figure 2.3: Tetrameric structure of PFK. (a) Missing residues are shown inside of the circle. (b) Modeled structure. The residues between 303rd and 308th are modeled for all chains.....	16
Figure 2.4: Phosphofructokinase enzyme surrounded by water molecules in a periodic box prepared by VMD visualization tool (Humphrey et al., 1996). .....	17
Figure 2.5: Selected residue pairs for constrained runs of (a) proposed allosteric site (b) known allosteric site. At the allosteric sites, green sticks representing Ravicti drug molecule and PEP, respectively. PFK colored by chain (chain A: wheat, chain B: pale green, chain C: light blue, chain D: light pink). See Table 3 for selected residue pairs. .	19
Figure 3.1: Perturbed residues colored in dark blue for (a) experiment #1 (proposed allosteric site) and (b) experiment #2 (known allosteric site). At the catalytic sites of <i>S. aureus</i> PFK, ATP and F6P represented as green and yellow sticks, respectively. PFK colored based on $\Delta G$ values of each residue and maximum and minimum $\Delta G$ values indicated on the spectrum bar below. The figure was adapted from submitted paper (Çelebi et al., 2021). .....	24
Figure 3.2: Perturbed residues colored in dark blue in (a). At the catalytic sites of <i>T. brucei</i> PFK, ATP and F6P represented as green and yellow sticks, respectively. (c) $\Delta G$ profiles of all chains, and (d) $\langle \Delta G \rangle$ values in the catalytic sites. The figure was adapted from submitted paper (Çelebi et al., 2021). .....	27



Figure 3.3: Perturbed residues colored in dark blue in (a). At the catalytic sites of human PFK, ATP and F6P represented as green and yellow sticks, respectively. (b) $\Delta G$ profiles of all chains, and (c) $\langle \Delta G \rangle$ values in the catalytic sites. The figure is adapted from submitted paper (Çelebi et al., 2021). .....	29
Figure 3.4: Perturbed residues are colored in dark blue in (a), (b), and (c). At the catalytic sites of all species, GADPH, G3P, and NAD represented as green and yellow sticks, respectively. (d) $\Delta G$ profiles of all chains, and (e) $\langle \Delta G \rangle$ values in the catalytic sites. The figure is adapted from submitted paper (Çelebi et al., 2021). .....	33
Figure 3.5: Perturbation sites in PK of all three species. ....	36
Figure 3.6: $\Delta G$ energy profiles of binding, and $\langle \Delta G \rangle$ values in the catalytic sites for (a) <i>S. aureus</i> PK, (b) <i>L. mexicana</i> PK, and (c) <i>H. sapiens</i> PK. The figure is adapted from submitted paper (Çelebi et al., 2021). .....	39
Figure 3.7: The distance profiles of the constrained inter-atomic distances (a) 10 distances for the proposed allosteric site (constr-1) (b) 20 distances for the known allosteric site (constr-2). .....	40
Figure 3.8: Potential and kinetic energy profiles of nine simulations. Energy values were shown every 40 ps in the energy profile. See top legend for color code. ....	41
Figure 3.9: RMSD plots for three different simulation runs in phosphofructokinase enzyme (a) apo form, (b) constr-1 (proposed allosteric site) and (c) constr-2 (known allosteric site). .....	43
Figure 3.10: Two different states aligned RMSD plots of (a) apo form Run #1, (b) proposed allosteric site (constr-1, Run #1), and (c) known allosteric site (constr-2, Run #1). The red lines represent tetramer-aligned RMSD, while the blue lines represent chain-aligned RMSD. ....	45
Figure 3.11: RMSD profiles of the chain and tetramers in each run of three state (apo, constr-1 and constr-2) simulations are given. The blue, red, green, orange, and magenta colored lines represents the RMSD profiles of Chain A, Chain B, Chain C, Chain D, and tetramer, respectively. ....	46
Figure 3.12: RMSF plots for three different simulations run in phosphofructokinase enzyme (a) apo form, (b) proposed allosteric site (constr-1), and (c) known allosteric site (constr-2). and (d) sum of RMSF values of all runs for each simulation state. ....	48

Figure 3.13: The sum of RMSF values for three runs in phosphofructokinase enzyme and their difference from apo (a) constr-1 (proposed), (b) constr-2 (known). The red and blue circles represent the residues nearby the substrate (F6P) and cofactor (ATP), respectively. (c) *Sa*PFK is colored according to the sum of the RMSF values for the Apo simulations, the difference of the sums of the RMSF values for the constr1 - apo simulations, and the constr2 - apo simulations, respectively (increasing from blue to white and red)..... 49

Figure 3.14: Principal component analysis for three different simulation runs in the phosphofructokinase enzyme. The explanation percentage of the total motion of the modes is given cumulatively..... 51

Figure 3.15: Projection of 5000 different conformations on the first three principal components from principal component analysis for apo, constr-1 and constr-2 runs which are colored as blue and shades, green and shades, orange, and shades, respectively. .... 54

Figure 3.16: Correlation between fluctuating vectors of alpha carbon atoms calculated using components that explain 70% of the motion resulting from principal component analysis. (a) apo simulations (b) constr-1 simulations (proposed) and (c) constr-2 simulations. Values between -0.4 and 0.4 are not shown on all maps..... 55

Figure 3.17: All three runs of apo and constr-1 simulations were summed up and their difference was taken. PCA Maps of (a) apo simulation, (b) constr-1 simulation, which does not show correlations in between -1.25 and +1.25, (c) differences of constr-1 and apo simulation, which does not show correlations in between -1.0 and +1.0. The green and cyan circles represent the residues nearby the substrate (F6P) and cofactor (ATP), respectively. .... 58

Figure 3.18: Evaluation chart showing the biggest changes of total correlation differences shown in Figure 3.17c for the constr-1 simulations. (a) Inter-chain correlation maps, (b) correlation table with significant changes, (c) showing the correlations between domains from the top and bottom of phosphofructokinase with arrows. Values between -1 and 1 are not shown on the maps. .... 60

Figure 3.19: All three runs of apo and constr-2 simulations were summed up and their difference was taken. PCA Maps of (a) apo simulation, (b) constr-2 simulation, which does not show correlations in between -1.25 and +1.25, (c) differences of constr-2 and apo simulation, which does not show correlations in between -1.0 and +1.0. The green and

cyan circles represent the residues nearby the substrate (F6P) and cofactor (ATP), respectively. ....	63
Figure 3.20: Evaluation chart showing the biggest changes of total correlation differences shown in Figure 3.19c for constr-2. (a) Inter-chain correlation maps, (b) correlation table with significant changes, (c) showing the correlations between domains from the top and bottom of phosphofructokinase with arrows. Values between -1 and 1 are not shown on the maps. ....	64
Figure 3.21: Normal mode analysis of each run of three state simulations according to first principle modes. The normal mode analysis results of apo, constr-1, and constr-2 simulations were given in the first, second, and third row, respectively. For each simulation, the percentage of the explained motion in the first principle mode was shown in parentheses. ....	66
Figure 3.22: Principal axes directions of PFK aligned to (a) A/B dimer, and (b) C/D dimer, (c) domain-1, and (d) domain-2 in Chain A. The snapshots were taken from initial frame of apo run-1 trajectory. ....	68
Figure 3.23: T-and R-state of PFK (created by the author Çelebi with Keynote tool)..	69
Figure 3.24: Principal axes directions of PFK aligned to (a) initial frame of domain-1 (b) initial frame of domain-2, (c) last frame of domain-1, and (d) last frame of domain-2 in Chain A. All snapshots were taken from apo run-1 trajectory. ....	70
Figure 3.25: The variations in angles of (a) A/B – C/D dimers (b) domains in Chain A (c) domains in Chain B (d) domains in Chain C, and (e) domains Chain D. ....	75
Figure 3.26: The distance fluctuation maps of (a) apo, (b) constr-1, and (c) constr-2 simulations All maps are showing values between 0 and 0.5 Å. The red points represent the contact map. ....	78
Figure A.1.1: Configuration file script for apo simulations of SaPFK. ....	89
Figure A.1.2: (a) Configuration file script for proposed allosteric site constrained simulations, (b) restrained bonds. ....	90
Figure A.1.3: (a) Configuration file script for known allosteric site constrained simulations, (b) restrained bonds. ....	91

## LIST OF SYMBOLS/ABBREVIATIONS

Å	Angstrom
$\alpha$	Alpha
$\beta$	Beta
$\Delta$	Delta
$\Delta G$	Gibbs Free Energy
$\lambda$	Lambda
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
C $\alpha$	Carbon alpha
F6P	Fructose 6-phosphate
FDA	US Food and Drug Administration
G3P	Glyceraldehyde 3-phosphate
GADPH	Glyceraldehyde 3-phosphate dehydrogenase
GABA	Gamma-Aminobutyric acid
HIV	Human immunodeficiency virus
MD	Molecular Dynamics
NAD	Nicotinamide Adenine Dinucleotide
NAMD	Nanoscale Molecular Dynamics
NADPH	Nicotinamide Adenine dinucleotide phosphate
PA	Principal Axes
PCA	Principal Component Analysis

PDB	Protein Data Bank
PFK	Phosphofructokinase
PEP	Phosphoenolpyruvate
PK	Pyruvate Kinase
ProDy	Protein Dynamics & Sequence Analysis
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
VMD	Visual Molecular Dynamics



# 1. INTRODUCTION

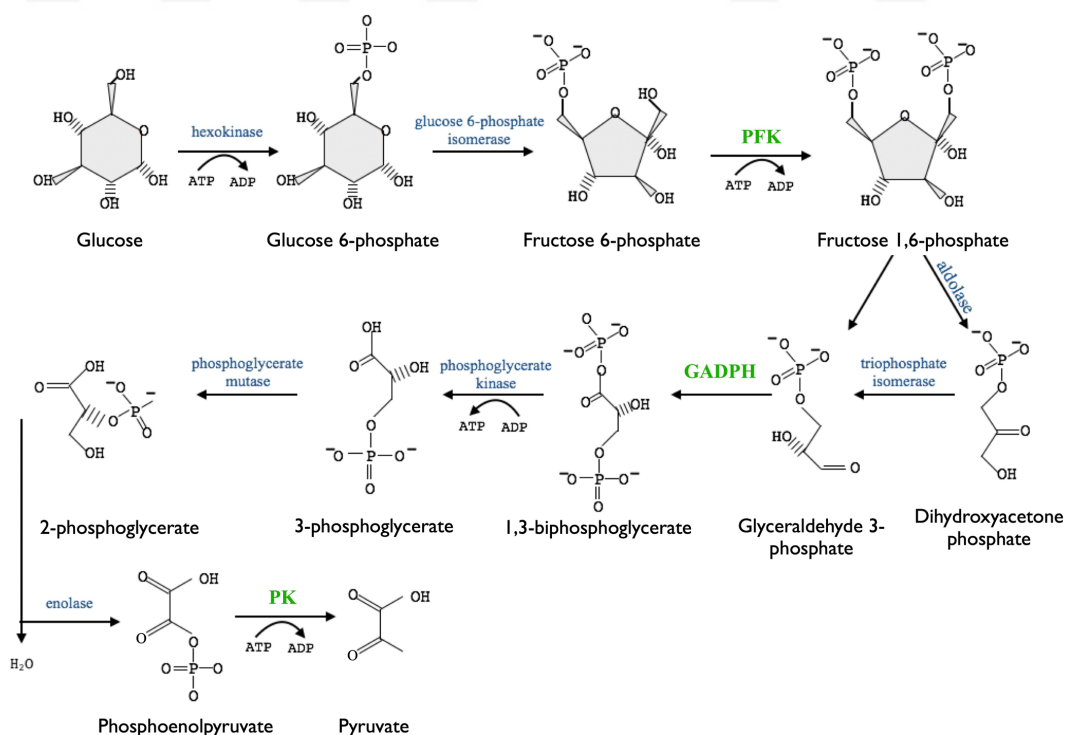
## 1.1 Glycolysis

In our planet, all living organisms require energy to be alive. For this survival requirement, all living cells produce and consume energy. There exist two main categories of organisms according to their energy production mechanism: aerobic and anaerobic organisms. Glycolysis is the most essential biochemical reaction pathway in all living cells, and it has a vital role in the survival of organisms. Also, glycolysis is the first step of aerobic respiration. In this pathway, the six-carbon glucose molecule is broken down into three-carbon pyruvate molecules by the reaction of several enzymes (Romano & Conway, 1996). The glycolysis pathway can be separated into two main parts. The first part is the use of ATP molecule to provide phosphate molecule. In the second part, four ATP molecules are produced using one molecule of NADH (Barnett, 2003). Discovered nearly seven decades ago (Campa et al., 1990), the glycolysis pathway is composed of ten different reactions catalyzed by ten different enzymes, which are hexokinase, phosphoglucose isomerase, phosphofructokinase, aldolase, triphosphate isomerase, glyceraldehyde 3-phosphate dehydrogenase, phosphoglycerate kinase, phosphoglycerate mutase, enolase, and pyruvate kinase. Three of these enzymes are known as allosteric enzymes that are phosphofructokinase (PFK) which catalyzes the phosphorylation of fructose-6-phosphate to fructose-1,6-biphosphate, glyceraldehyde 3-phosphate dehydrogenase (GADPH) that catalyzes the oxidative phosphorylation of glyceraldehyde-3-phosphate into 1,3-biphosphoglycerate which is a high energy compound with cofactors as NAD or NADPH in plant cells, and pyruvate kinase (PK) which catalyzes the conversion of phosphoenolpyruvate (PEP) into pyruvate and yields an ATP molecule. Phosphofructokinase has a vital role in glycolysis, as PFK catalyzed reaction is a rate-limiting step in the glycolytic pathway (Boscá & Corredor, 1984). As a result of a total of 10 reactions in the glycolytic pathway, 4 ATP energy molecules, 2

NADH, and 2 Pyruvate molecules are produced from 1 glucose molecule and 2 ATP molecules as shown below.



Since glycolysis is both an essential biochemical pathway for energy production and the most critical step in the organism's survival, it has made glycolysis a vital drug target in species-specific drug design studies, which has a significant place among drug design strategies today. In this technique, the drug is targeting the disease-related organism without harming human cells. The targeted enzymes in the species-specific drug design technique should play a vital role in the survival of disease-causing organisms. Since active sites of the same biological targets in different species are more conserved in the evolutionary process than other regions, inhibitors targeting the allosteric region have high specificity and low side effects (Conn et al., 2009). There is evolutionary pressure on residues in the active site to be conserved. However, since this evolutionary pressure is weaker on allosteric regions, the residues are more susceptible to mutations (Henry, 2004). As a result, differences in amino acid sequence and structure level makes these allosteric sites ideal for species-specific drug design.



**Figure 1.1:** Enzymes in glycolytic pathway. The figure is adapted from Ayyildiz Master's thesis.

## 1.2 Species – Specific Allosteric Drugs

The binding of an allosteric drug to a binding site that is distant from the active site in a protein causes changes in the protein conformation (Daniel E. Koshland & Hamadani, 2002; Jacques Monod et al., 1963). Allosteric drug molecules affect the protein's active site from the remote binding site as if it was a remote controller. This remote control triggers a conformational change in the protein and ultimately works by either increasing (also known as positive modulation) or decreasing (negative modulation) the binding affinity of small molecules in the active site (May & Christopoulos, 2003). Thus, allosteric drugs affect the active site on the protein from a distant binding site. A century ago, Niels Bohr discovered the first allosteric effect; the binding of two different molecules changes the conformation of hemoglobin structure (Bohr et al., 1904). Also, nearly eight decades ago, two allosteric models by MWC (Monod – Wyman – Changeux) (Jacob & Monod, 1961; Jacques Monod et al., 1965) where the oligomeric protein interconverts between two conformations as tense and relaxed conformations and sequential KNF (Koshland – Nemethy – Filmer) models (D. E. Koshland et al., 1966) that monomeric units change conformations, one at a time were proposed. In recent years, Gunasekaran showed that non-allosteric proteins might become allosteric as a result of some critical mutations or binding of a ligand to a specific site. In this way, all proteins have the potential to be allosteric (Gunasekaran et al., 2004).

Many drugs developed to date have targeted the orthosteric site of proteins. In recent years, a new strategy targeting allosteric sites has been established. The most important reason for this is that inhibitors targeting the allosteric site have less side effects and more specificity compared to inhibitors targeting the orthosteric site. The active site of a protein in an organism has performed the same function in order to ensure the continuation of the generation in the organisms before it. Therefore, there is an evolutionary pressure on the active site, and thus residues in this region are more conserved. In the allosteric region, evolutionary pressure is weaker, so among species, this region is more susceptible to mutations. There exist several allosteric drugs on the market. Flurazepam is an allosteric drug that is used to treat insomnia disorder. Once it binds to the GABA-A receptor, it causes conformational changes in the orthosteric site (Möhler et al., 2002; Spurny et al.,



2012). Maraviroc is another allosteric FDA approved drug on the market which is used for HIV treatment as an allosteric modulator (Abel et al., 2008).

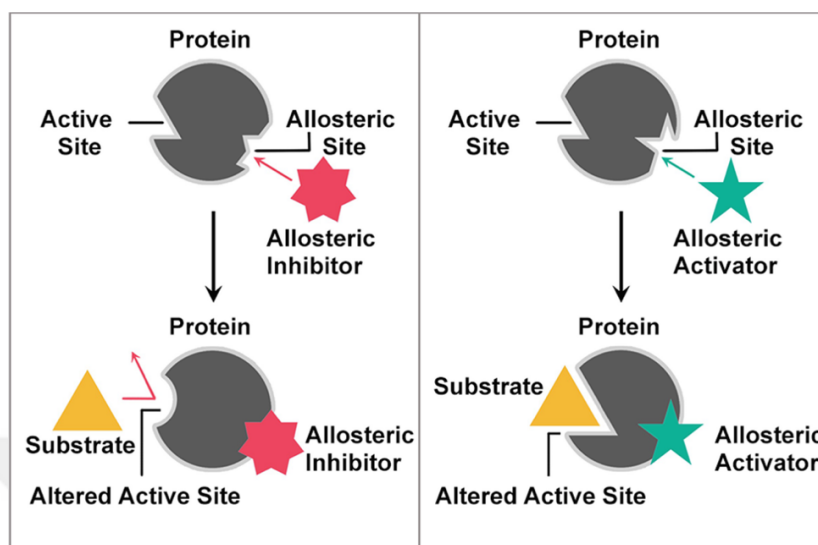
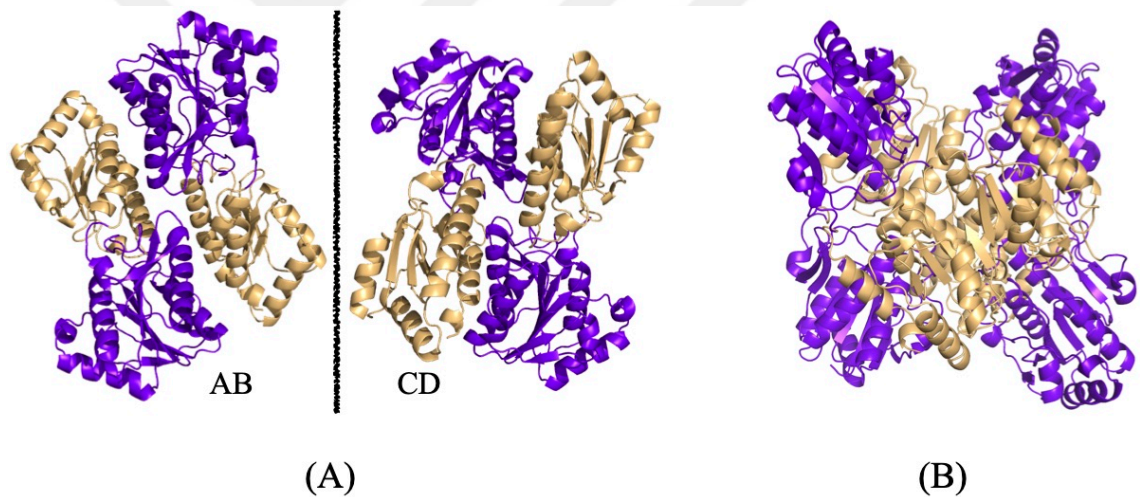


Figure 1.2: Allosteric drug mechanism. The figure was adapted from (Cheng and Jiang 2019).

### 1.3 Structure and Function of Phosphofructokinase

Phosphofructokinase is a key enzyme in the glycolytic pathway because PFK catalyzes the rate-limiting step of phosphorylation of fructose-6-phosphate (F6P) to fructose-1,6-bisphosphate and allosterically regulates the glycolytic pathway depending on the amount of glucose in the environment according to the energy needs of the organism. In the first part of this thesis, three PFKs, each from different species have been studied. These belong to *Staphylococcus aureus* (*Sa*PFK), *Trypanosoma brucei* (*Tb*PFK), and *Homo sapiens* (*h*PFK). In the second part, only bacterial PFK (*Sa*PFK) was studied in detail via Molecular Dynamics (MD) simulations. *Sa*PFK is a homotetramer which consists of four identical chains as illustrated in Figure 1.3(a) Each chain consists of 326 amino acid residues which fold into two distinct domains highlighted in magenta and yellow in Figure 1.3. Domain 1 (magenta) covers residues from 1 to 138 and 256 to 305, and is the largest domain which binds the cofactor ATP (Schirmer & Evans, 1990). On the other hand, the smaller domain 2 is composed of residues from 139 to 255 and from 306 to 326, and binds the substrate F6P (Schirmer & Evans, 1990) that plays a critical role in dimer-tetramer conversion (Tian et al., 2018). Both larger and smaller domains are 3-layered  $\alpha\beta\alpha$  structures and domains are illustrated in Figure 1.3. The A and B chains come together to form the AB dimer, while the C and D chains form the CD domain via interface I, which

consists of an effector binding site (Tian et al. 2018) (see Figure 1.3b). ATP and F6P catalytic site and an effector site are found in each dimer (Shirakihara and Evans 1988). *Sa*PFK forms a "dimer of dimers" tetrameric assembly (see Figure 1.3c). Here, AB and CD dimers form tetrameric structures via Interface – II (dimer-dimer interface), consisting of substrate binding sites (Tian et al. 2018). PFK shows enzymatic activity in tetrameric form. The two dimer structures of PFK come together to form the tetramer structure and this provides a transition from the inactive structure to the active structure. But in human, the tetrameric form of *h*PFK is different than *Sa*PFK and *Tb*PFK tetramer structures. In the human genome, the gene responsible for the expression of PFK has undergone gene duplication during the evolutionary process. Due to gene duplication, human PFK is synthesized two times. Therefore, each monomer of human PFK corresponds to two monomers in *Sa*PFK and *Tb*PFK (Brüser et al. 2012). For this reason, the dimeric structure of *h*PFK is composed of four identical subunits (Kloos et al. 2015).



**Figure 1.3:** X-Ray crystallographic structure of *Sa*PFK (PDB ID: 5XZ7). (A) *Sa*PFK's two dimers as AB and CD. (B) Tetrameric structure of *Sa*PFK. Larger domains and smaller domains in all chains are colored purple, blue, and light orange, respectively.

#### 1.4 Allosteric Regulation of Phosphofructokinase

According to the classical allosteric model (mechanistic), allosteric communication between two distant regions is provided by a series of conformational changes in the protein conformation one residue after the other, just like domino stones. It shows an allosteric effect at the distant region of protein by changing the affinity of the protein toward its substrate either increase or decrease. These conformational changes not only affect the mean conformation (enthalpic communication) of the protein but also affect

the dynamic fluctuations (entropic communication) of the protein. These dynamic fluctuations also play an important role in allosteric communication. But most studies in the literature tend to explain the allosteric mechanism in terms of structural terms and neglecting the protein dynamics. In a study, explained the allostery upon ligand-induced changes in protein dynamics could play a role in allosteric communication (Cooper & Dryden, 1984). This is the allosteric model that we are trying to reveal in PFK Molecular Dynamics simulations. The changing protein dynamics upon ligand binding to the allosteric site proposed by Ayyildiz and her coworkers (Ayyildiz et al., 2020), and reported allosteric site in the literature for *S. aureus* PFK had been studied computationally.

*Sa*PFK shows enzymatic activity in the tetrameric structure, so the dimeric form is the inactive state, tetrameric form is the active state in equilibrium in the cell. Here, ATP and F6P play a role in the conversion of inactive state (T-state) to active state (R-state) by dimer-tetramer conversion. In a study conducted by (Tian et al., 2018), they discovered F6P binding plays a vital role in dimer-tetramer conversion by mutating the F6P binding residues. Upon binding F6P to four binding pockets, the interactions between AB and CD dimers are stabilized, and the tetramer structure is formed. Arg164 and Arg245 residues in all four F6P binding pockets interact with F6P and play an allosteric role in stabilizing dimer-dimer interface interaction. As a result of mutation of these arginine residues with alanine, it was observed that F6P could not bind to the dimer structure and thus remained as a dimeric form in solution (Tian et al., 2018). Also, it was observed that in the F6P bound structure, Arg164 side chain flips to interact with F6P, and the Arg245 side chain is stabilized. This stabilization is resulting in the extension of the strand  $\beta 9$ . Additionally, upon binding of F6P, it has been observed that it performs an elongation from the C-terminus at the  $\alpha 7$  regions and an inward folding motion from residues Gly150 and Leu151. This inward folding motion causes the  $\alpha 7$  regions in one dimer to interact with the  $\alpha 11$  and loop  $\alpha 2 - \alpha 3$  region in the other dimer, and it facilitates tetramer formation (Tian et al. 2018). On the other hand, ATP has two roles as substrate and inhibitor; it has two binding regions for ATP. When the ATP molecule is bound to the inhibitory site, the PFK structure stays in the inactive state. On the other hand, when the ATP molecule is bound to the ligand-binding site, PFK remains in the

active state. As the ATP production increases in the cell, the reaction stops by binding the ATP molecule to the inhibitory site. So that's why PFK catalyzes the rate-limiting step of the glycolytic pathway. Each dimer of PFK in tetramer structure makes  $7^{\circ}$ - $8^{\circ}$  relative to each other (Schirmer and Evans 1990). The stabilization of the dimer-dimer interface between these two dimers is stabilized by the binding of F6P. However, there exist two molecules which are citrate and ADP. While citrate plays a role as an allosteric inhibitor, ADP plays a role as an allosteric activator.



## 2. MATERIALS AND METHODS

### 2.1 AlloSigMA Experiments

#### 2.1.1 System Preparation

To further investigate the allosteric sites proposed by (Ayyildiz et al., 2020) for PFK, GAPDH, and PK enzymes, several experiments were conducted via the AlloSigMA server (Guarnera et al., 2017; Guarnera & Berezovsky, 2016). AlloSigMA is an abbreviation for Allosteric Signaling and Mutation Analysis. X-ray crystallographic structures of PFK, GAPDH, and PK enzymes were extracted from Protein Data Bank, and corresponding PDB IDs were listed in Table 1. While the same bacterial species *S. aureus* was used for each enzyme, different parasitic species were used; *T. brucei* for PFK, *T. cruzi* for GAPDH and *L. mexicana* for PK.

**Table 2.1:** PDB ids of X-ray crystal structures used in AlloSigma experiments.

Enzyme	Bacterium	Parasite	Human
PFK	5XZ7*	3F5M <sup>§</sup> ( <i>T. brucei</i> )	4RH3 <sup>§</sup>
GAPDH	3HQ4 <sup>§</sup>	3DMT <sup>§</sup> ( <i>T. cruzi</i> )	4WNI <sup>§</sup>
PK	3T0T <sup>§</sup>	1PKL <sup>§</sup> ( <i>L. mexicana</i> )	4G1N <sup>§</sup>

\*Used in both AlloSigMA experiments and Molecular Dynamics Simulations.

<sup>§</sup> Only used in AlloSigMA experiments.

Here, all proposed allosteric sites were evaluated using the AlloSigMA, which uses normal mode analysis and quantifies the effect of perturbation caused by a ligand binding at the proposed site via the changes in the normal modes. It calculates the free energy ( $\Delta G$ ) of each residue upon the perturbation caused by a ligand binding, or a mutation. AlloSigMA uses the following method: Firstly, the AlloSigMA system characterizes the ligand-free  $P$  and ligand bound  $AP$  protein structure systems using the  $C_a$  harmonic model constructed

based on the structures. Here, the AlloSigma obtains the ligand-bound  $AP$  system from the ligand-free system by harmonic restraining of all the selected residue pairs as allosteric binding site  $A$ , which stabilizes the allosteric binding site. The aim is to influence the local dynamics of the protein by modeling the presence of the actual bound ligand in the allosteric site  $A$ . Then, to explore the protein dynamics, two sets of normal modes are obtained from the harmonic protein model as the ligand-free and ligand-bound. Secondly, the AlloSigma system uses these two normal modes to calculate the allosteric potential which used to evaluate the energy reflecting structural differences caused by dynamics of ligand-free and ligand-bound structures. The AlloSigma system evaluates the allosteric potentials of all residues. Third and lastly, the AlloSigma system compares the configurational ensembles of the ligand-free and ligand-bound structures to calculate the per residue allosteric free energy  $\Delta g_i (P \rightarrow AP)$  which is a statistical mechanical approach. In this way, the free energy difference between the  $n$  bound ligands  $A^{(n)}P$  states and the ligand-free state  $P$  is calculated by the Equations 2.1. This free energy difference between bound-ligand states and ligand-free states indicates the changes in global dynamics.

$$\Delta G (P \rightarrow A^{(n)}P) = \frac{1}{2} k_B T \sum_{\mu} \ln \frac{\lambda_{\mu}^{(n)}}{\lambda_{\mu}^{(0)}} \quad (2.1)$$

Where  $\lambda_{\mu}^{(n)}$  and  $\lambda_{\mu}^{(0)}$  are the eigenvalues of the Hessian matrix of the model for the ligand-bound, ligand-free states, respectively. According to  $\Delta G$  calculations, a positive  $\langle \Delta G \rangle$  indicates an increase in dynamics (destabilization), a negative  $\langle \Delta G \rangle$  indicates a decrease in dynamics (stabilization). In our AlloSigma experiments, we have quantified the allosteric effect of proposed allosteric sites upon ligand binding and the perturbation effect imposed on catalytic sites has been evaluated by determining  $\langle \Delta G \rangle$  values of all residues that belong to the catalytic sites. A favorable allosteric effect would cause a negative  $\langle \Delta G \rangle$  on the catalytic region. The residue list and experiment numbers we used in the AlloSigma experiments were shown in Table 2.2, 2.3, and 2.4. In addition to our experiments on the proposed sites, we also used the known allosteric sites reported in the literature for PFK and PK enzymes (Schirmer & Evans, 1990). Our aim here was to make a comparison of the allosteric effect between proposed and known allosteric sites. No reported allosteric sites were reported for the GADPH enzyme so far.

**Table 2.2:** Experiment numbers and corresponding residues for the perturbation sites in PFK.

Enzyme	Species	Experiment	Binding Sites*
PFK	<i>S. aureus</i> (PDB id: 5XZ7)	#1 Proposed Site(Top Druggable)	<b>A:</b> N130,D136-T143,L145,N146,W181,T259-D262,V264-A266,R268, I288-N291 <b>B:</b> F137-T143,L145,N146,W181,T259-D262,V264-A266,R268, I288, N291 <b>C:</b> N130,F137-T143,L145,N146,W181,T259,G260-D262,V264,A266, R268,I288,N291 <b>D:</b> N130,F137-T143,W181,T259-D262,V264-A266,R268,I288, N290, N291
		#2 Known Allosteric Site	<b>A:</b> R21,R25,D59,R156,V186-E189,K215,I322 <b>B:</b> R21,R25,D59,R156,V186-E189,K215,I322 <b>C:</b> R21,R25,D59,R156,V186-E189,K215,I322 <b>D:</b> R21,R25,D59,R156,V186-E189,K215,I322
	<i>T. brucei</i> (PDB id: 3F5M)	#1 Proposed Site (Top Druggable)	<b>A:</b> L84,A85,R237,Q242,N390,L393,C395,T397,L398,L401,M417, N420-Y422 <b>B:</b> R8-S11,N230,H236-T238,F241,Q242,Q282,V285,R435,L437,Q442,L443,Q446 <b>C:</b> L84-A86,D237,T390,L393,L401,N420-I423 <b>D:</b> L8-S11,N230,H236,T238,F241,Q242,F278,Q282,S432,R435, Q442,L443
		#5 Proposed Site	<b>A:</b> K247,Y375-I382 <b>B:</b> K247,Q250,A254,Y375-P378,Y380,M381 <b>C:</b> Y375-M381 <b>D:</b> Q250,Y375-I382
	<i>H. sapiens</i> (PDB id: 4RH3)	Proposed Site (Top Druggable)	<b>A:</b> M183,T187,D188,L191,Y223,V227,S315,F317,N390,F548,D553, P680- G686,K688, I722,V727 <b>B:</b> D182,M183,A316-D318,I320-A322,R324,L347-H351,N541,F548-D553,L556,N557,Y589,M593,S679,F681,D682,F685,I722-N726, W750 <b>C:</b> D182,M183,A316-D318,I320-A322,R324,L347-H351,N541,F548-D553,L556,N557,Y589,M593,S679,F681,D682,F685, I722-N726, W750 <b>D:</b> M183,T187,D188,L191,Y223,V227,S315,F317,N390,F548,D553, P680-G686,K688,I722,V727

\* Binding site residues are taken from (Ayyildiz et al., 2020).

**Table 2.3:** Experiment numbers and corresponding residues for the perturbation sites in GADPH.

<b>Enzyme</b>	<b>Species</b>	<b>Experiment</b>	<b>Binding Sites*</b>	
<b>GADPH</b>	<i>S. aureus</i> (PDB id: 3HQ4)	Proposed Site	<b>A:</b> D48-R53,Y180,A203-N205,P236-A238 <b>B:</b> D48-M50,A203-N205,P236,T239,S281,D282,V284 <b>C:</b> A203,E204,P236 <b>D:</b> D48-G52,A203,E204,P236,S281-V284	
			Proposed Site	<b>A:</b> T54,V55,S195,Y196,A218-I221,P251-V255,S298-I302 <b>B:</b> T54-H56,S195,Y196,A218-I221,P251-S256,S298, A299,I302 <b>C:</b> T54,V55,S195,Y196,A218-I221,P251-V255,S298, A299,I302 <b>D:</b> T54-H56,S195,Y196,A218-I221,P251-V255,S298, A299,I302
				Proposed Site
<i>H. sapiens</i> (PDB id: 4WNI)				

\* Binding site residues are taken from (Ayyildiz et al., 2020).



**Table 2.4:** Experiment numbers and corresponding residues for the perturbation sites in PK.

Enzyme	Species	Experiment	Binding Sites*
<b>PK</b>	<i>S. aureus</i> (PDB id: 3T0T)	#1 Proposed Site	<b>A:</b> K260,R264,N267,N299,Y302-G304,A337- Y340,K342,L343,D346,R347 <b>B:</b> K260,N299,Y302,D303,A337- Y340,K342,L343,D346 <b>C:</b> K260,R264,N267,N299,Y302,D303,D346,R347 <b>D:</b> Y302,Q338-Y340,K342,L343
		#2 Known Allosteric Site	<b>A:</b> T353,A358,I361,S362,H365,T366,N369 <b>B:</b> T353,A358,I361,S362,H365,T366,N369,L370 <b>C:</b> T353,A358,I361,S362,H365,T366,N369 <b>D:</b> A358,I361,S362,H365,T366,N369,L370
		#3 Proposed Site	<b>A:</b> K271,E324 <b>B:</b> G481,G483,R484,A572-Q574 <b>C:</b> - <b>D:</b> -
	<i>L. mexicana</i> (PDB id: 1PKL)	Proposed Site (Top Druggable)	<b>A:</b> - <b>B:</b> - <b>C:</b> R19,R22,I23,L40-S46,V76-I78,C420,T427- T434,V437-S439 <b>D:</b> -
	<i>H. sapiens</i> (PDB id: 4GIN)	Proposed Site (Top Druggable)	<b>A:</b> F26,L27,H29,M30,L33,K311,C326,N350,V352- G355,A388-H391 ,Q393-F395,E397 <b>B:</b> F26,L27,M30,K311,N350,L353,D354,A388- L394,E397 <b>C:</b> - <b>D:</b> -

\* Binding site residues are taken from (Ayyildiz et al., 2020).

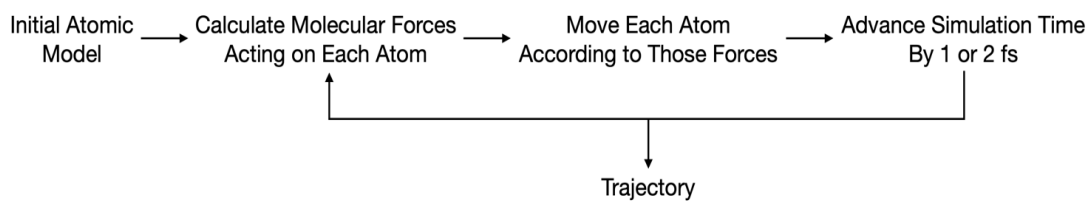
## 2.2 Molecular Dynamics Simulations of Phosphofructokinase

Molecular Dynamics (MD) simulations is a computational methodology developed to analyze the physical movements of atoms and molecules in a certain time period. MD simulations calculate the motion of molecular system with respect to time. It provides detailed information on the conformational changes and dynamics of the molecules. Nearly 70 years ago, the first MD was introduced to study the interactions of hard spheres (Alder & Wainwright, 1957, 1959). In 1974, Rahman carried out the first MD simulation of liquid water (Stillinger & Rahman, 1974). The first MD simulation of a biological system was done on bovine pancreatic trypsin inhibitor (BPTI) in 1977 (McCammon et al., 1977). MD is commonly used in computational biology to investigate the structure and dynamics of proteins and nucleic acids and their complexes, such as protein-DNA interaction or protein-ligand interaction. The simulation is based on Newton's second law of motion,  $F = m \times a$ , where  $F$  is the force applied on a particle,  $m$  is the mass, and  $a$  is the acceleration of that particle. Calculating the force on each atom in the system, it is possible to determine the acceleration of each atom by Newton's second law of motion equation. For this purpose, it is important to use the Verlet Algorithm (Verlet, 1967), a numerical integration algorithm that integrates Newton's second law of motion as follows,

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + a(t)\delta t^2 \quad (2.2)$$

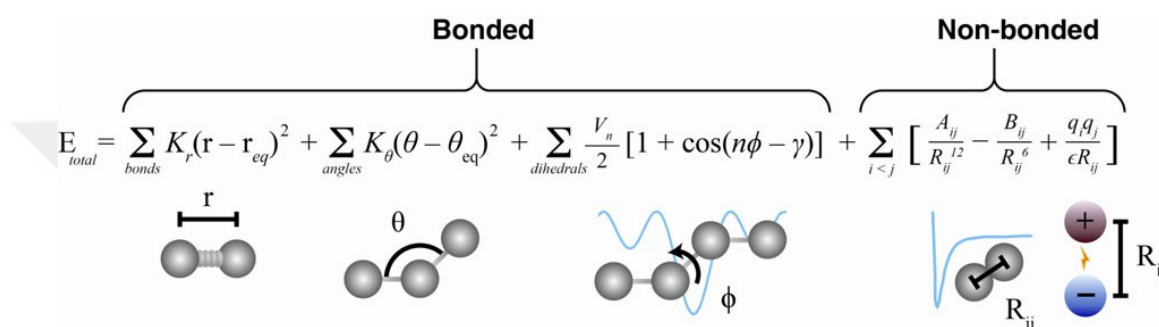
$$v(t) = [r(t + \delta t) - r(t - \delta t)]/2\delta t \quad (2.3)$$

Solution of Equations 2.2 and 2.3 yields a trajectory that includes the velocities ( $v$ ), accelerations ( $a$ ), and positions ( $r$ ) of atoms varying by time. Using this trajectory file, the new position of each atom in the system can be predicted. In the first years of MD method, simulations could be carried out on a few atoms because it requires not only lots of time but also requires high power computer power. However, today, thanks to huge improvements in computer technology, MD simulations have advanced from simulating several atoms to systems up to 100,000 – 500,000 atoms. The general algorithm used in all MD simulations is shown in Figure 2.1.



**Figure 2.1:** A chart showing how a molecular dynamics simulation is performed.

The first step in MD algorithm is the calculation of the force acting on each atom by using the forcefield equation as basically shown in Figure 2.2.



**Figure 2.2:** CHARMM Potential function (Brooks et al., 1983) used to approximate the atomic forces that govern molecular movement. The figure adapted from (Durrant & McCammon, 2011).

The calculation of this force is essential to reproduce the actual behavior of simulating molecules as real molecules in motions. The energies described in Figure 2.2 were parametrized to reproduce experimental data such as spectroscopic data. Here, force field refers to all these parameter sets used to calculate the potential energy of atoms or groups of atoms in a system. For this purpose, several force field parameters were used in MD simulations such as AMBER (Assisted Model Building with Energy Refinement) (Cornell et al., 1995; Wang et al., 2004), CHARMM (Chemistry at HARvard Macromolecular Mechanics) (Brooks et al., 1983), and GROMOS (GRoningen Molecular Simulation) (Oostenbrink et al., 2004). The difference between these force fields is in the energy components and the way they were parameterized. Basically, all force fields calculate the total energy by summing bonded and non-bonded potentials. The second step was the movement of each atom according to those forces. When the acting forces on each atom have been calculated, the acceleration of atoms can be calculated via Newton's law of motion, and with the Verlet Algorithm,

$$r(t + \delta t) = r(t) + v(t)\delta + \frac{1}{2}a(t)\delta^2 \quad (2.4)$$

$$r(t - \delta t) = r(t) - v(t)\delta + \frac{1}{2}a(t)\delta^2 \quad (2.5)$$

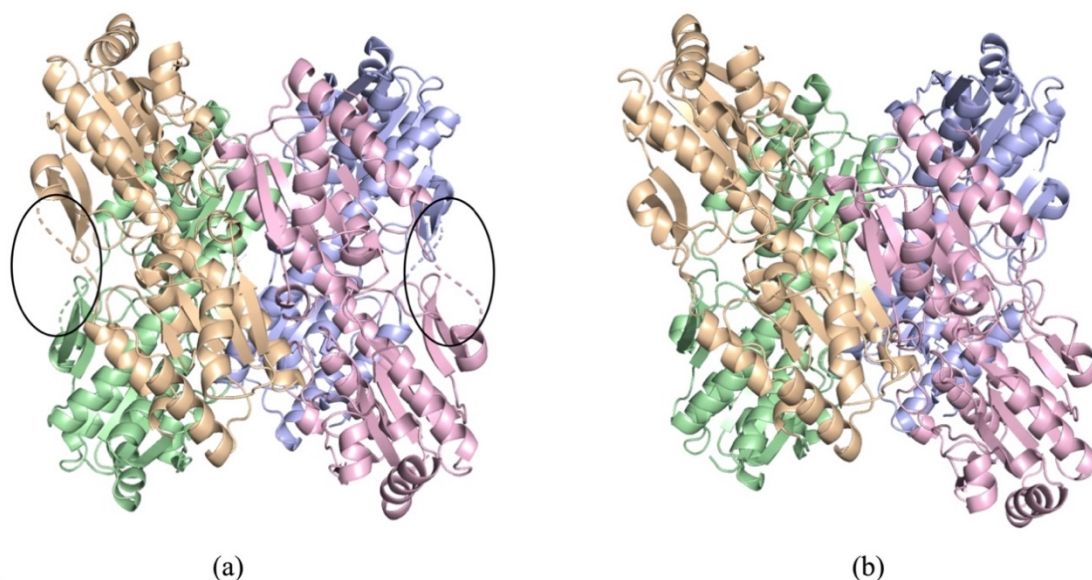
When Equations 2.4 and 2.5 are summed, the new position of the atoms in the system are obtained from the current position, previous position, and the acceleration of atoms:

$$r(t + \delta t) = 2r(t) + r(t - \delta t) + a(t)\delta^2 \quad (2.6)$$

In the third step, those positions are saved to the trajectory. Then, these steps are repeated over time until the system generates a trajectory of molecular motion. Each time step in simulation corresponds to 2 femtoseconds because the smallest atomic vibration frequency is in the femtosecond order. Thus, 1 ns simulation trajectory corresponds to a million femtoseconds, which consist of 500,000-time step.

### 2.2.1 Completing Missing Residues on Phosphofructokinase

For MD simulations of *S. aureus* PFK, the crystal structure with PDB id 5XZ7 was extracted from Protein Data Bank at 1.60 Å resolution. In this structure, as shown in Figure 2.3, some of the loop regions in all four chains were missing. Thus, during system preparation, these missing residues were completed with the MODELLER software (Eswar et al., 2006; Martí-Renom et al., 2000) which is a prediction tool for modeling protein 3D structures using homology or comparative modeling methods. A total of 16 residues were missing, four in each of the four chains of the protein. Completed residues with the MODELLER software were 304GLY, 305LYS, 306ASP, and 307HIS.

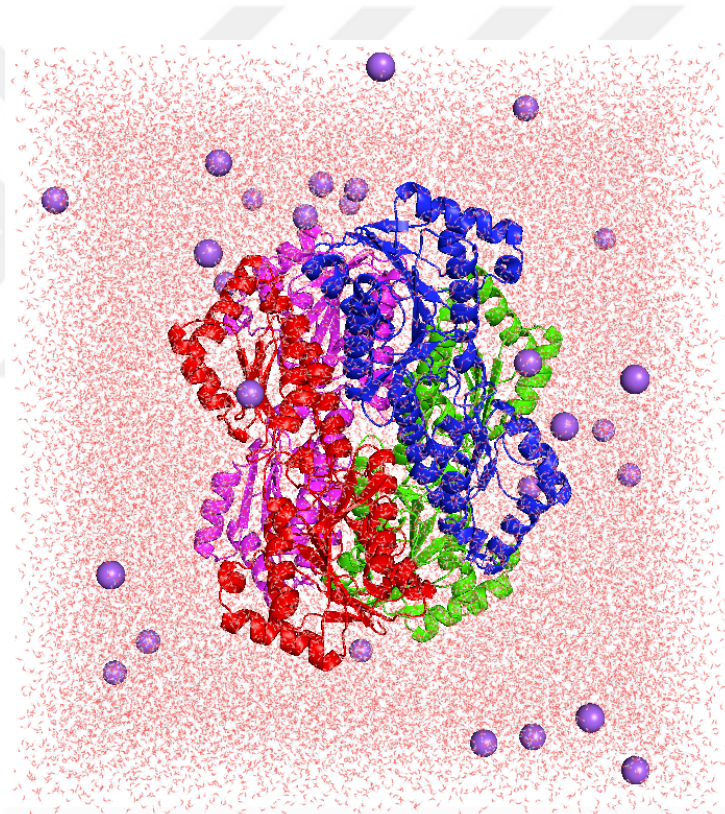


**Figure 2.3:** Tetrameric structure of PFK. (a) Missing residues are shown inside of the circle. (b) Modeled structure. The residues between 303<sup>rd</sup> and 308<sup>th</sup> are modeled for all chains.

### 2.2.2 Molecular Dynamics Simulation System and Configuration Preparation

For Molecular Dynamics simulations, nine independent 100-ns-long runs were carried out for apo form, bond restraints at the proposed allosteric site, and known allosteric site (three of each) under CHARMM36 (Huang et al., 2017) force field parameters and 1 atm pressure, 310 K temperature for each run. Each simulation consists of 5 ns equilibration and 95 ns production phases. For the solvation process, the protein was solvated in a water box with benefiting from periodic boundary conditions. In an environment with 10 grams of water molecules, there are approximately  $3 \times 10^{23}$  water molecules, and this amount is quite high to calculate with molecular simulations. Instead of creating and calculating the whole system here, an infinite system can be created based on a central cell of sufficient size and placing the image copies of this cell around it. Therefore, periodic boundary conditions allow to create a computable environment by shrinking an infinite system into a small unit cell. Since the cell cytoplasm is an aqueous environment, the protein was placed in a water box during Molecular Dynamics simulation system preparation. However, in Molecular Dynamics simulations with periodic boundary conditions that do not have sufficient water box size, it is inevitable that the macromolecule interacts with its image in the other cell. This interaction is functionally equivalent to the interaction of the head part of a protein with the tail part of itself. This is a highly undesirable issue in terms of calculating protein dynamics. The first MD simulation experiment for

phosphofructokinase of this study encountered a problem caused by the small water box size. During the analysis of the trajectory file, it was observed that Chain A interacted with Chain D of its image in the other cell. So, a new MD system was prepared and created a water layer of 17.5 Angstroms in the x, y, and z directions from the atom with the largest coordinates in this direction (see Figure 2.4). After the solvation step, to neutralize the system's charge the ionization process was performed. For the ionization step, 28 Na<sup>+</sup> ions were added using VMD visualization tool (Humphrey et al., 1996) so that the total electric charge was zero. As a result of the solvation and ionization processes, the total number of atoms in the system reached 151,342. The periodic box dimensions prepared for the simulation were 110.1 Å, 118.7 Å, and 121.8 Å, respectively, in the x,y,z-direction.



**Figure 2.4:** Phosphofructokinase enzyme surrounded by water molecules in a periodic box prepared by VMD visualization tool (Humphrey et al., 1996).

### 2.2.3 System Preparation for Constrained Molecular Dynamics Simulations

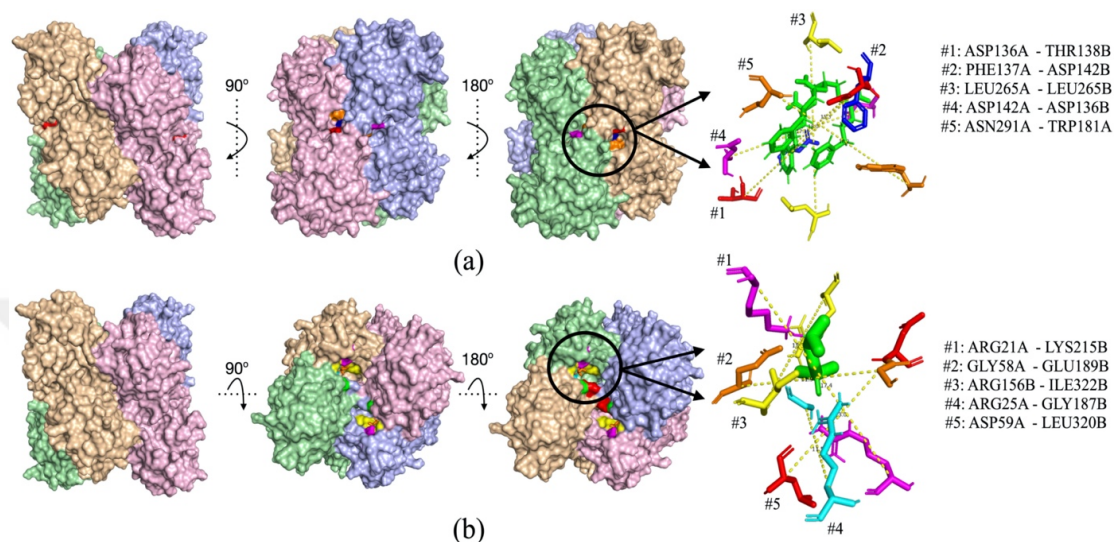
In this study, besides the apo form of PFK MD simulations, constrained Molecular Dynamics simulations for proposed and known allosteric sites were performed. The purpose of the constrained Molecular Dynamics simulations was to compare and observe the drug binding effect upon both known allosteric sites and proposed allosteric sites for *S. aureus* PFK. In these constrained runs, the extra bond property of NAMD (Phillips et al., 2020) software used instead of preparing the parameter files of the ligand. The extra bond property of NAMD is a beneficial technique for observing protein dynamics as if a ligand was bound by keeping the distance between the two or more selected atom pairs constant throughout the simulations. This property applies a harmonic restraining force so that the distance between the selected atom pairs is constant. The harmonic potentials were calculated based on Equation 2.4.,

$$U(x) = k(x - x_{ref})^2 \quad (2.4)$$

where  $x_{ref}$  is the reference distance in Å and  $k$  is the spring constant. The choice of the atom pairs is crucial in terms of the accuracy of the simulation and the behavior of the protein, just like in the biological system. While choosing these pairs of atoms in both the proposed allosteric site and the known allosteric site, the selected pairs of atoms must interact with the ligand which was taken as Ravicti (Drugs@FDA, 2018) drug molecule with its best docking pose for proposed allosteric site (Ayyildiz et al., 2020). As illustrated in Figure 2.5(a), the proposed allosteric region is located on the front and rear fronts of the enzyme, at the interface connecting the A/B and C/D chains. A total of 10 distance restrictions have been implemented, with 5 residue pairs on each facade. The region in Figure 2.5(b) is an allosteric region determined as a result of experimental measurements to which PEP (phosphoenol pyruvate) inhibitor was bound (Schirmer & Evans, 1990). It is located in the upper and lower parts of the enzyme as two regions at a short distance from each other. Therefore, for a total of four different PEP binding sites, 20 restricted inter-atomic distances were defined as shown in Figure 2.5(b). The distance between the alpha carbon atoms of these selected residue pairs was measured via VMD (Humphrey et al., 1996) program and provided with the configuration files of the constrained simulations together with their atomic indexes. Details of configuration files were given in Appendix

A1. The selected residue list and the distance between them were given in Table 2.5 and the positions of the residues on the protein in Figure 2.5.

It should be noted that the proposed allosteric site constrained MD simulations will be referred as "Constrained -1" and the known allosteric site constrained MD simulations as "Constrained -2" throughout the thesis.



**Figure 2.5:** Selected residue pairs for constrained runs of (a) proposed allosteric site (b) known allosteric site. At the allosteric sites, green sticks representing Ravicti drug molecule and PEP, respectively. PFK colored by chain (chain A: wheat, chain B: pale green, chain C: light blue, chain D: light pink). See Table 2.5 for selected residue pairs.



**Table 2.5:** Selected residue pairs for constrained MD simulations.

<b>Constrained Allosteric Site</b>	<b>Selected Residue Pairs</b>	<b>Distance (Å)</b>
<i>Proposed Allosteric Site (Constrained -1)</i>	ASP136A* – THR138B	19.1
	PHE137A – ASP142B	15.3
	LEU265A – LEU265B	18.0
	ASP142A – ASP136B	17.3
	ASN291A – TRP181A	18.2
	ASP136D – THR138C	19.1
	PHE137D – ASP142C	15.3
	LEU265D – LEU265C	18.0
	ASP142D – ASP136C	17.3
	ASN291D – TRP181D	18.2
<i>Allosteric Site (Constrained -2)</i>	ARG21A – LYS215B	17.4
	GLY58A – GLU189B	11.9
	ARG156B – ILE322B	15.3
	ARG25A – GLY187B	11.5
	ASP59A – LEU320B	13.0
	ARG21B – LYS215A	17.4
	GLY58B – GLU189A	11.9
	ARG156A – ILE322A	15.3
	ARG25B – GLY187A	11.5
	ALA188A – ASP59B	11.8
	ARG21C – LYS215D	17.4
	GLY58C – GLU189D	11.9
	ARG156D – ILE322D	15.3
	ARG25C – GLY187D	11.5
ASP59C – LEU320D	13.0	
ARG21D – LYS215C	17.4	
GLY58D – GLU189C	11.9	
ARG156C – ILE322C	15.3	
ARG25D – GLY187C	11.5	
ALA188C – ASP59D	11.8	

\* The letter next to each residue ID indicates the chain ID, e.g., ASP136 in chain A.

### 3. RESULTS AND DISCUSSION

This thesis consists of two parts. The first part consists of the investigation of allosteric sites previously proposed in our research group and reported in a paper (Ayyildiz et al., 2020 ) for each species of PFK, GADPH, and PK with AlloSigMA (Guarnera et al., 2017; Guarnera & Berezovsky, 2016). Besides proposed allosteric sites, reported known allosteric sites in the literature for *S. aureus* PFK and PK were also investigated (Axerio-Cilies et al., 2012; Schirmer & Evans, 1990). Tables 2.2 through 2.4 list the experiment details. In the second part of this thesis, Molecular Dynamic simulations were performed for *S. aureus* PFK for three simulation systems: one apo and two restrained states. Restraints were employed on the proposed and known allosteric sites. For each simulation state, three independent runs were performed to increase the sampling. To reveal the difference between these simulations and the effects of restraints on protein dynamics, various computational methods were used.

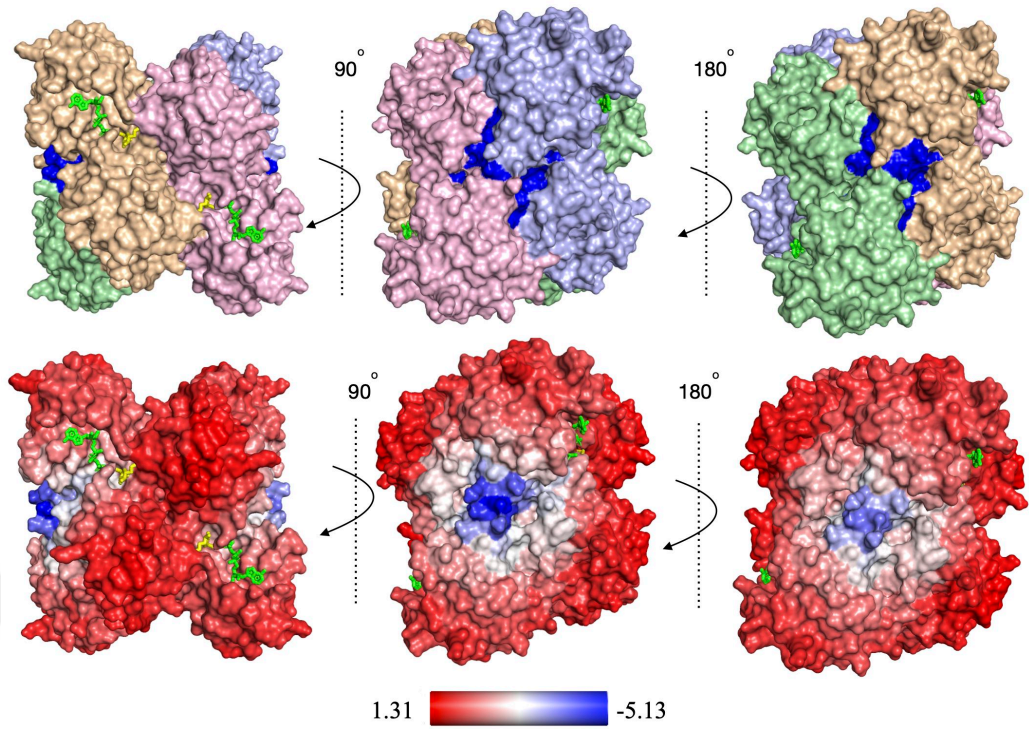
#### 3.1 AlloSigMA Experiments

##### 3.1.1 *Sa*PFK AlloSigma Experiments

In *Sa*PFK experiments, two allosteric sites were investigated for their allosteric effect on the catalytic sites of PFK. The first experiment was the proposed allosteric site identified as a top druggable site by Ayyildiz and her coworkers (Ayyildiz et al., 2020). Due to the symmetric structure of PFK, the proposed allosteric site is located on both A/B and C/D chains. Corresponding residues were highlighted with dark blue in Figure 3.1(a) (Exp #1). The second experiment was the reported allosteric site of *Sa*PFK occupied by the inhibitor phosphoenolpyruvate (PEP), retrieved from *G. Stearothermophilus* (PDB ID: 6PFK) (Schirmer & Evans, 1990). It also has symmetric counterparts on all four chains.

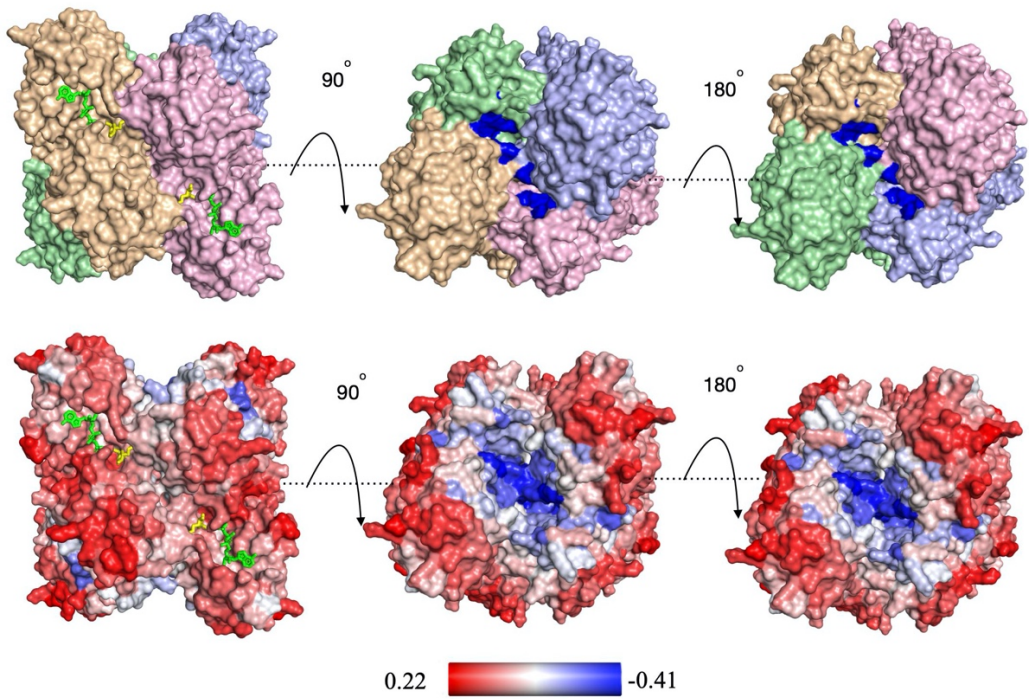
Corresponding residues were highlighted with dark blue in Figure 3.1(b) (Exp #2). The list of the residues used as perturbation sites were given in Table 2.2 in Methods section. AlloSigMA measures the allosteric effect upon ligand binding on selected residues according to change in free energy,  $\Delta G$ , of each residue. A minimum and a maximum  $\Delta G$  values were determined, and a color gradient was used to display the change in the local dynamics on the structure. The increase in free energy indicates a destabilization (increasing dynamics). The decrease in free energy indicates a stabilization (decreasing dynamics). Also, a mean allosteric free energy  $\langle \Delta G \rangle$  was determined for residues close to F6P and ATP up to 4.5 Å distance at the catalytic sites. Accordingly, perturbation of proposed sites in *Sa*PFK caused a significant decrease in  $\langle \Delta G \rangle$  (from 0.18 to -0.78 kcal/mol) in F6P and ATP binding sites in all chains. On the other hand, known allosteric sites manifested differently low positive values, from 0.04 to 0.05 kcal/mol, which indicates a nearly null effect. Also, the degree of stabilization at the proposed allosteric site was -5.13 kcal/mol, while the degree of stabilization at the known allosteric site was -0.41 kcal/mol. Additionally, as illustrated in Figure 3.1(c)  $\Delta G$  profiles for all residues in all four chains, the known allosteric site displaying minor fluctuations in  $\Delta G$ , which is nearly zero, while the proposed site significantly disrupted the dynamics of the whole structure. Eventually, these experiments indicated that the proposed allosteric site was more restrictive in the dynamics of the catalytic region than the known allosteric site.

*S.aureus* PFK Exp #1 Proposed Site

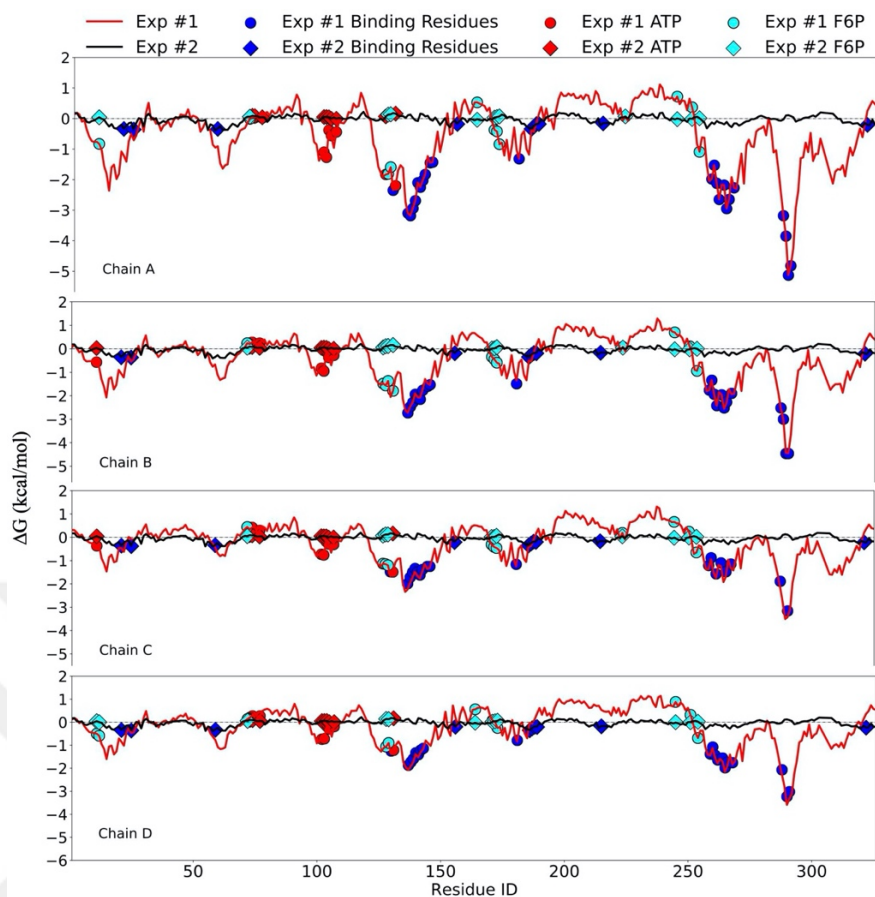


(a)

*S.aureus* PFK Exp #2 Known Allosteric Site



(b)



(c)

***S. aureus* PFK Exp #1 Proposed Site**

$\langle \Delta G \rangle$ (kcal/mol)		
Chain ID	Substrate (F6P)	Cofactor (ATP)
A	$-0.42 \pm 0.81$	$-0.78 \pm 0.72$
B	$-0.62 \pm 0.77$	$-0.51 \pm 0.62$
C	$-0.34 \pm 0.64$	$-0.33 \pm 0.59$
D	$-0.18 \pm 0.55$	$-0.32 \pm 0.46$

***S. aureus* PFK Exp #2 Known Site**

$\langle \Delta G \rangle$ (kcal/mol)		
Chain ID	Substrate (F6P)	Cofactor (ATP)
A	$0.05 \pm 0.05$	$0.06 \pm 0.05$
B	$0.05 \pm 0.06$	$0.06 \pm 0.05$
C	$0.04 \pm 0.05$	$0.06 \pm 0.05$
D	$0.04 \pm 0.05$	$0.07 \pm 0.05$

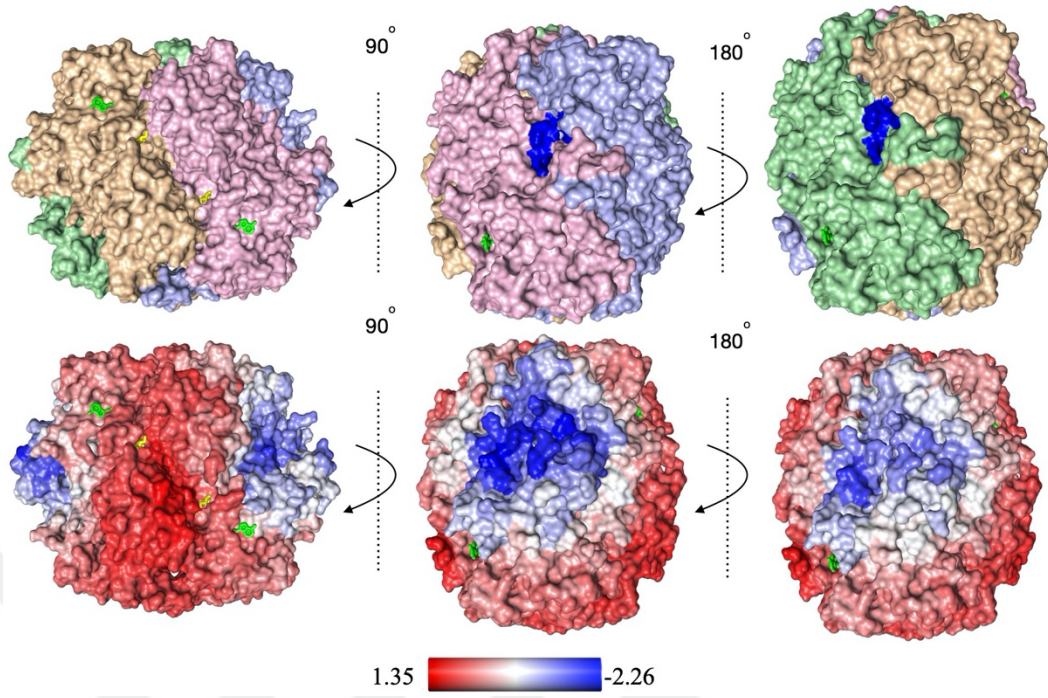
(d)

**Figure 3.1:** Perturbed residues colored in dark blue for (a) experiment #1 (proposed allosteric site) and (b) experiment #2 (known allosteric site). At the catalytic sites of *S. aureus* PFK, ATP and F6P represented as green and yellow sticks, respectively. PFK colored based on  $\Delta G$  values of each residue and maximum and minimum  $\Delta G$  values indicated on the spectrum bar below. The figure was adapted from submitted paper (Çelebi et al., 2021).

### 3.1.2 *Tb*PFK AlloSigma Experiments

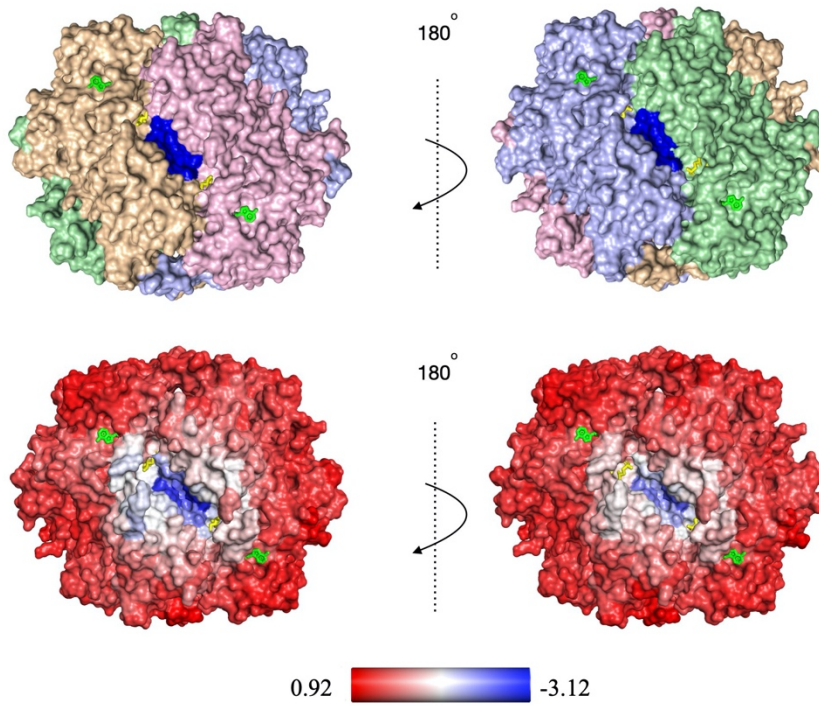
For *Tb*PFK five experiments were performed. (For 2nd, 3rd, and 4th experiments the list of perturbed residues and the corresponding results were given in Appendix A2). The residues used for the first experiment have the exact location as the proposed allosteric site in *Sa*PFK. But despite their exact locations, perturbation of these residues caused destabilization on all catalytic regions of *Tb*PFK by increasing  $\Delta G$  values as 0.08 and 0.63 kcal/mol. This destabilization indicates that the dynamics of catalytic sites have been increased, which wouldn't probably inhibit the catalytic site. In the same way, no allosteric effect was observed in experiments 2, 3, and 4 (see Appendix A2). Unlike other experiments in Experiment 4, the perturbed residues covered an area between two catalytic sites. As they were close to the catalytic site, this restriction was not allosteric. In experiment 5, the perturbed area was again in the region connecting the two catalytic sites, but this time was at an acceptable distance from the catalytic site, as shown in Figure 3.2(b). According to the results of the 5<sup>th</sup> experiment, it was observed that a significant allosteric effect occurred in all catalytic sites. As can be observed in Figure 3.2(d), negative  $\langle \Delta G \rangle$  values were observed in the catalytic region where F6P was bound (-0.84 kcal/mol). On the other hand, this region perturbs the interface between the two chains A-D and B-C. Therefore, perturbation of this region significantly affects the regulation of the global dynamics of the tetramer structure as well as the local dynamics.

*T.brucei* PFK Exp #1 Proposed Site

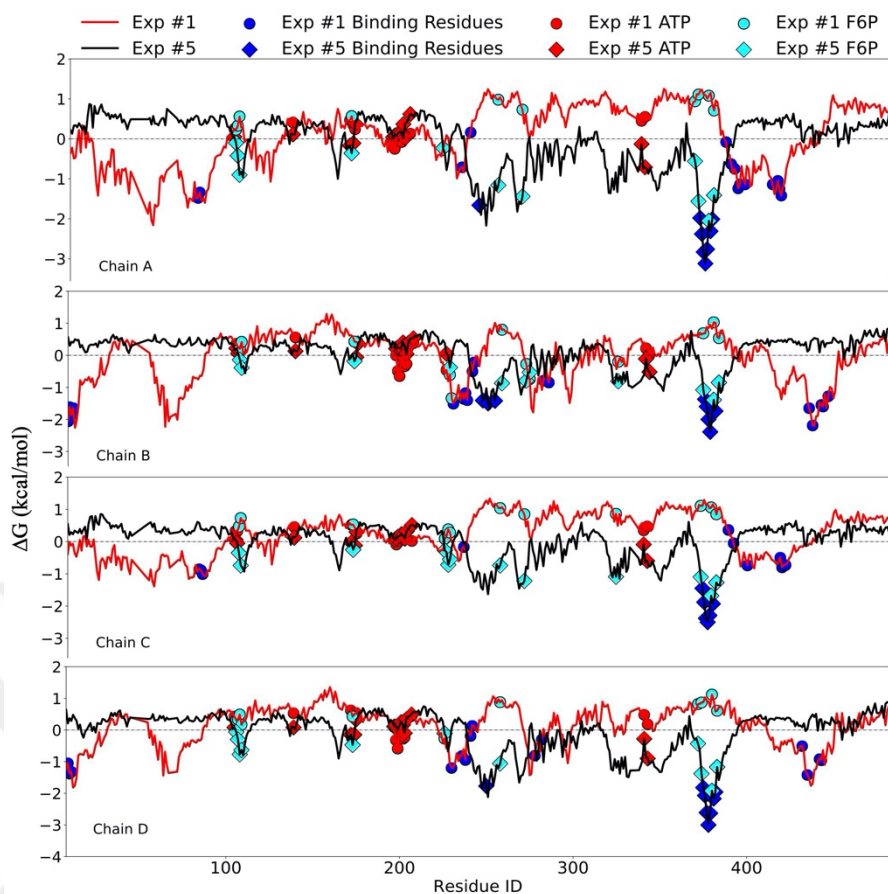


(a)

*T.brucei* PFK Exp #5 Proposed Site



(b)



(c)

***T. brucei* PFK Exp #1 Proposed Site**

$\langle \Delta G \rangle$ (kcal/mol)		
Chain ID	Substrate (F6P)	Cofactor (ATP)
A	$0.60 \pm 0.35$	$0.21 \pm 0.26$
B	$0.14 \pm 0.75$	$0.08 \pm 0.36$
C	$0.63 \pm 0.27$	$0.27 \pm 0.23$
D	$0.57 \pm 0.42$	$0.20 \pm 0.35$

***T. brucei* PFK Exp #5 Proposed Site**

$\langle \Delta G \rangle$ (kcal/mol)		
Chain ID	Substrate (F6P)	Cofactor (ATP)
A	$0.89 \pm 0.60$	$-0.07 \pm 0.42$
B	$-0.61 \pm 0.39$	$0.08 \pm 0.32$
C	$-0.84 \pm 0.43$	$-0.03 \pm 0.37$
D	$-0.79 \pm 0.57$	$-0.08 \pm 0.39$

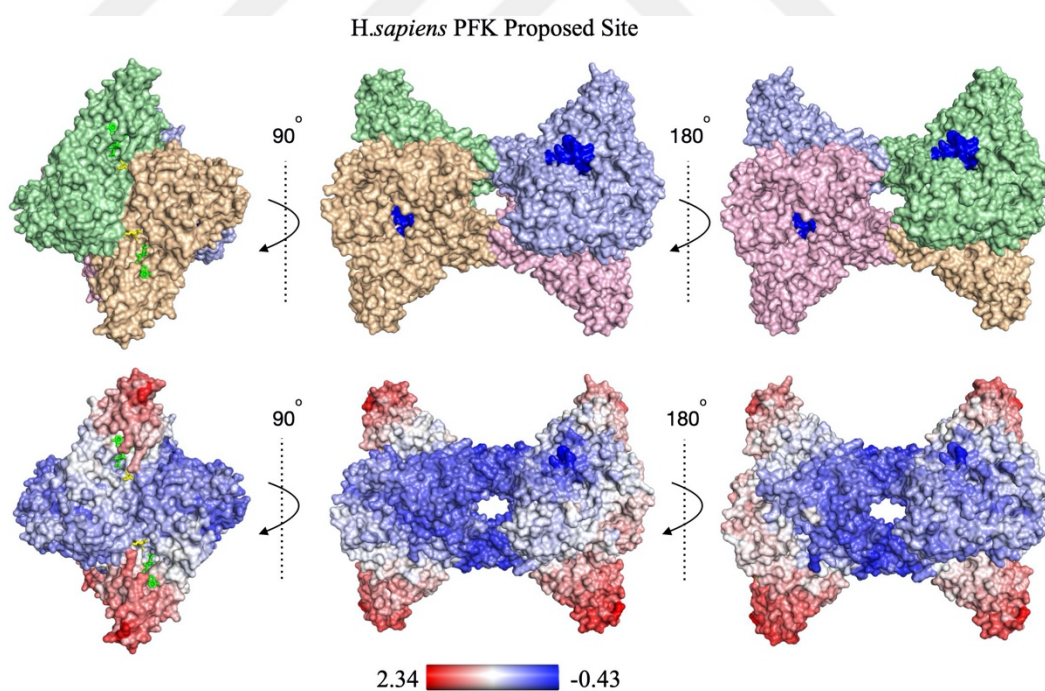
(d)

Figure 3.2: Perturbed residues colored in dark blue in (a). At the catalytic sites of *T. brucei* PFK, ATP and F6P represented as green and yellow sticks, respectively. (c)  $\Delta G$  profiles of all chains, and (d)  $\langle \Delta G \rangle$  values in the catalytic sites. The figure was adapted from submitted paper (Çelebi et al., 2021).

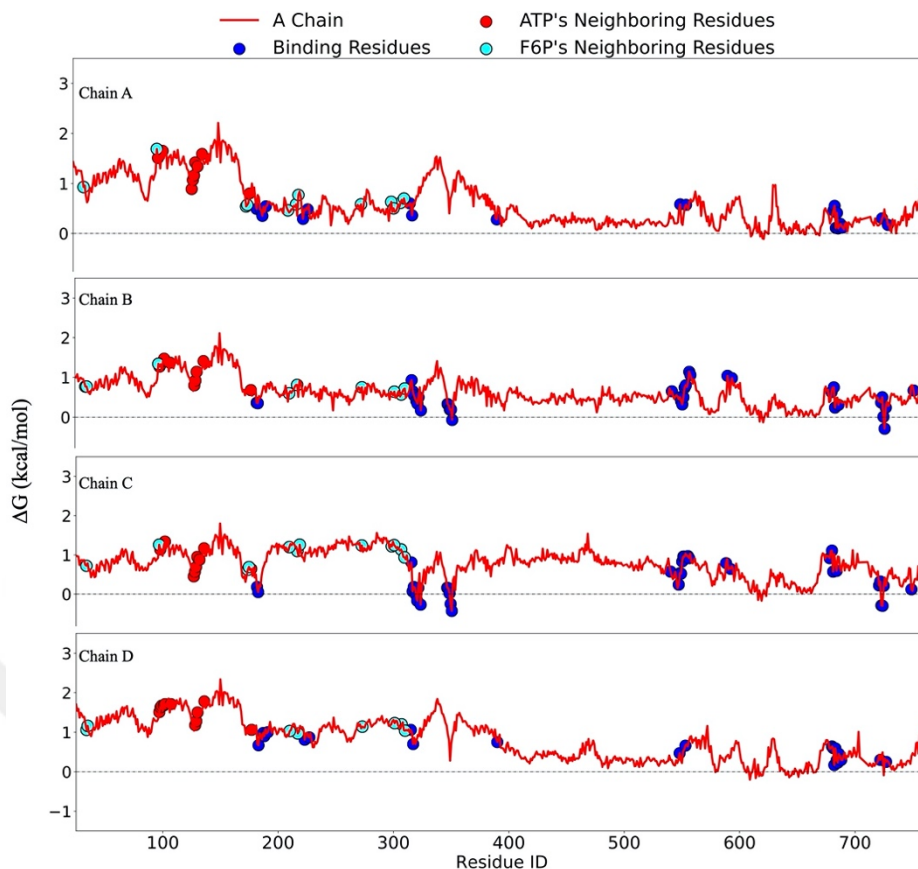


### 3.1.3 hPFK AlloSigma Experiments

Another investigated organism for PFK in our AlloSigma experiments was the host organism human. Since the basic strategy here is to propose a drug molecule that will provide complete inhibition of the disease-causing organism, it is important to investigate the allosteric effect in host organism human. The presence of an allosteric inhibition will indicate an undesirable side effect for the infected host. The residues used as binding sites were given in Table 2.2 and shown in dark blue in Figure 3.3(a). According to AlloSigma results obtained from the experiment, high  $\langle \Delta G \rangle$  values were observed in all catalytic sites in human PFK. See Figure 3.3(c). This indicates an increase in dynamics, and therefore, if the drug molecule was bound to this proposed site, it may cause activation in the protein. In the species-specific drug design strategy, a drug molecule that will effectively inhibit the disease-causing organism was designed. Therefore, *h*PFK AlloSigma experiment results showed that in the case of drug binding to the proposed allosteric sites, disease-causing organisms were inhibited, while there was no inhibition in human PFK.



(a)



(b)

***H. sapiens* PFK Exp #1 Proposed Site**

Chain ID	$\langle \Delta G \rangle$ (kcal/mol)	
	Substrate (F6P)	Cofactor (ATP)
A	$0.75 \pm 0.32$	$1.27 \pm 0.29$
B	$0.73 \pm 0.23$	$1.13 \pm 0.29$
C	$1.02 \pm 0.23$	$0.87 \pm 0.29$
D	$1.14 \pm 0.09$	$1.45 \pm 0.27$

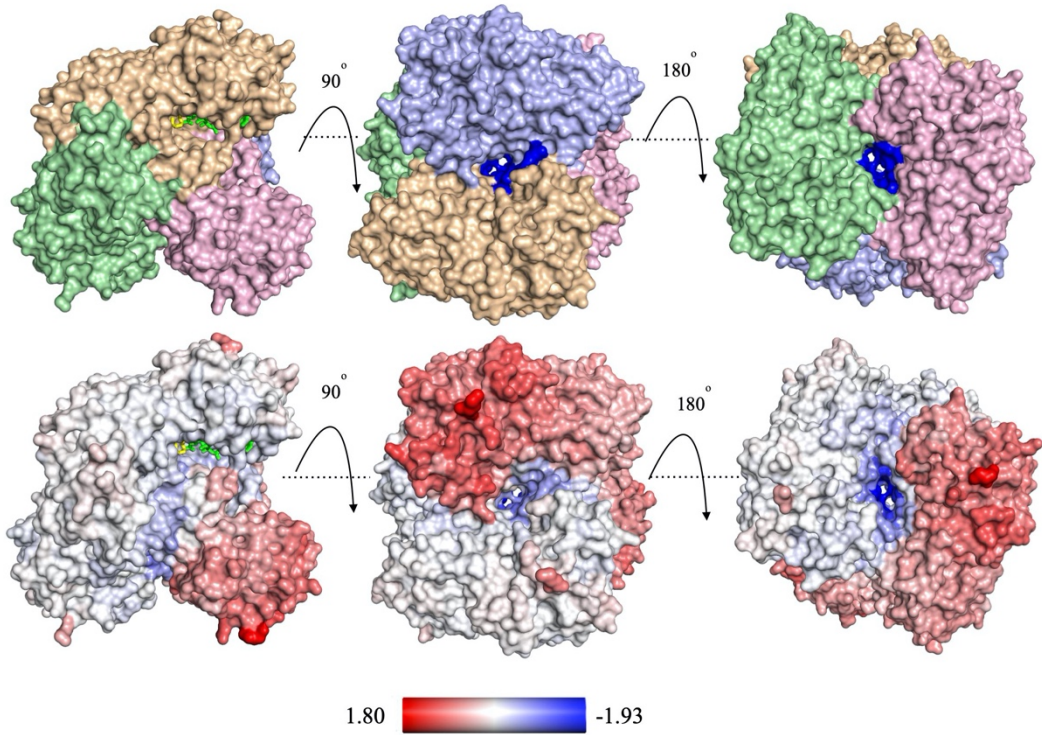
(c)

**Figure 3.3:** Perturbed residues colored in dark blue in (a). At the catalytic sites of human PFK, ATP and F6P represented as green and yellow sticks, respectively. (b)  $\Delta G$  profiles of all chains, and (c)  $\langle \Delta G \rangle$  values in the catalytic sites. The figure is adapted from submitted paper (Çelebi et al., 2021).

### 3.1.4 GADPH AlloSigma Experiments

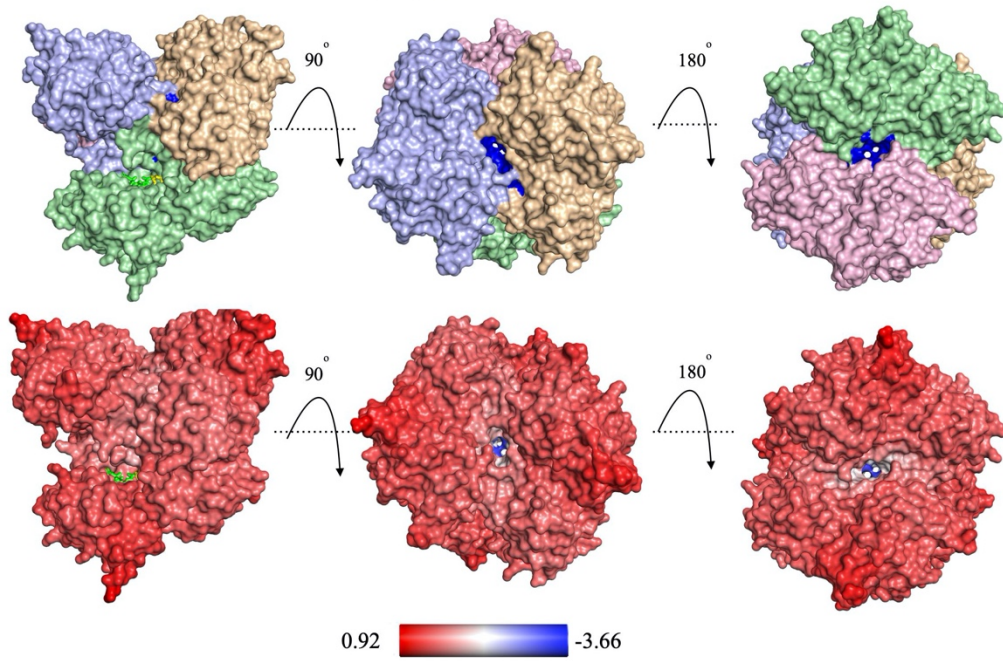
For GADPH, a total of three experiments were conducted, one experiment for each species. While searching the allosteric site for GAPDH, no allosteric site has been found in the literature. Yet, in a study conducted with five cancer cell lines, an inhibitor DC-5163 was identified, which targeted the active site of GADPH and effectively inhibited the activity of GADPH (Li et al., 2020). The residues that were perturbed in three species' structures have been illustrated in Figure 3.4(a), (b), and (c) with dark blue. In *S. aureus* GADPH, due to perturbation of the allosteric site, an increase in the dynamics of C and D chains was induced. On the other hand, in A and B chains, such a dynamic increase was not induced. Also, in the perturbation site of *S. aureus* GADPH, negative  $\Delta G$  values were observed as a minimum value of -1.93 kcal/mol, beside the binding site of substrate G3P and cofactor NAD<sup>+</sup> in A and B chains. According to *T. cruzi* GADPH experiments results, a dramatic restriction in protein dynamics was observed in all chains, especially was intensified in the G3P (substrate) binding sites. This restriction is at the level of -3.66 kcal/mol in terms of  $\langle \Delta G \rangle$  value. This showed that the perturbation of *T. cruzi* GADPH proposed site causes more allosteric restriction than the perturbation of the proposed site in *S. aureus* GADPH. In the *H. sapiens* GADPH, almost the same level of allosteric effect was observed with a minimum  $\langle \Delta G \rangle$  value of -3.45 kcal/mol. However, this restriction in dynamics was only observed on the catalytic sites in A and C chains. Based on the data obtained from GADPH AlloSigma experiments, it is clear that the strongest allosteric effect was observed in *T. cruzi* GADPH experiment, thus it can be proposed as a potential target site for the design of a species-specific drug.

*S.aureus* GADPH



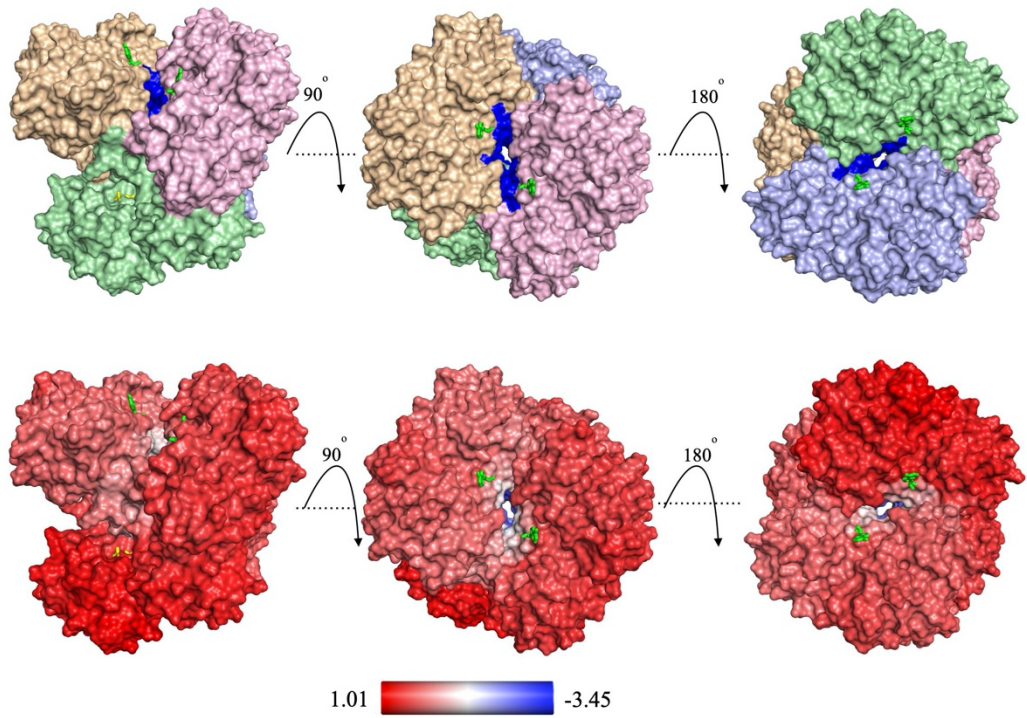
(a)

*T.cruzi* GADPH

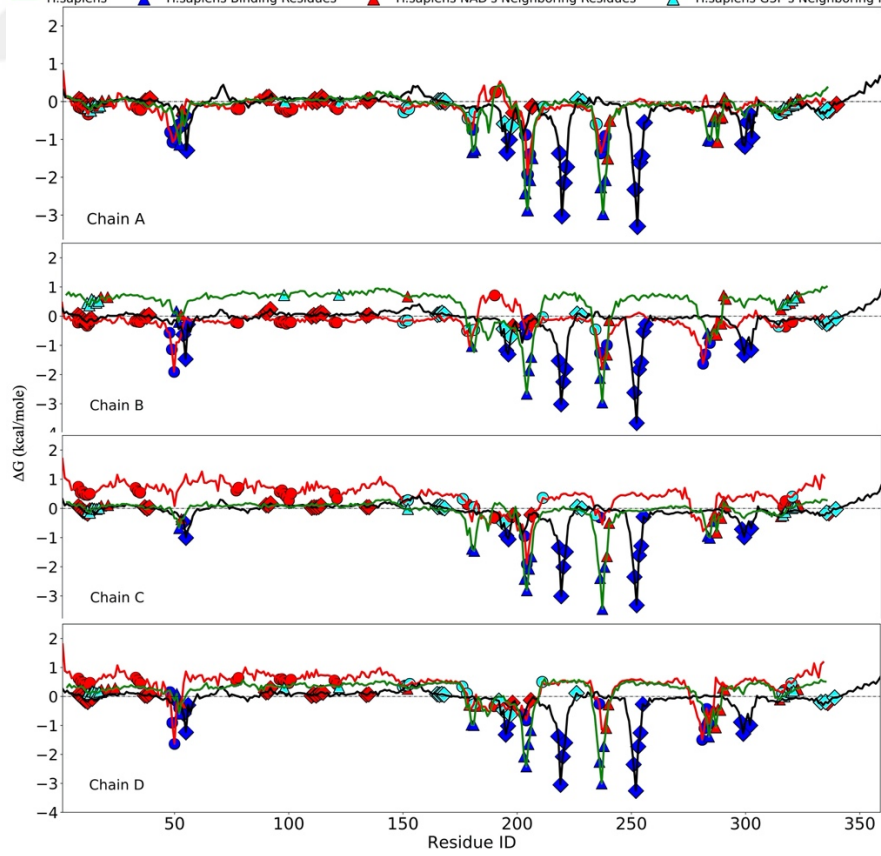
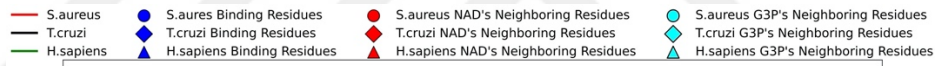


(b)

*H.sapiens* GADPH



(c)



(d)

<b>GADPH</b>					
$\langle \Delta G \rangle$ (kcal/mol)					
Species		Chain A	Chain B	Chain C	Chain D
<i>S. aureus</i>	G3P	$-0.30 \pm 0.14$	$-0.29 \pm 0.16$	$0.19 \pm 0.18$	$0.28 \pm 0.22$
	NAD	$-0.21 \pm 0.07$	$-0.25 \pm 0.09$	$0.46 \pm 0.19$	$0.42 \pm 0.20$
<i>T. cruzi</i>	G3P	$-0.17 \pm 0.25$	$-0.12 \pm 0.25$	$-0.11 \pm 0.23$	$-0.12 \pm 0.23$
	NAD	$-0.06 \pm 0.16$	$-0.06 \pm 0.18$	$-0.02 \pm 0.10$	$-0.02 \pm 0.17$
<i>H. sapiens</i>	G3P	$-0.11 \pm 0.07$	$0.53 \pm 0.12$	$-0.05 \pm 0.08$	$0.15 \pm 0.09$
	NAD	$-0.23 \pm 0.36$	$0.26 \pm 0.52$	$-0.17 \pm 0.39$	$-0.06 \pm 0.38$

(e)

**Figure 3.4:** Perturbed residues are colored in dark blue in (a), (b), and (c). At the catalytic sites of all species, GADPH, G3P, and NAD represented as green and yellow sticks, respectively. (d)  $\Delta G$  profiles of all chains, and (e)  $\langle \Delta G \rangle$  values in the catalytic sites. The figure is adapted from submitted paper (Çelebi et al., 2021).

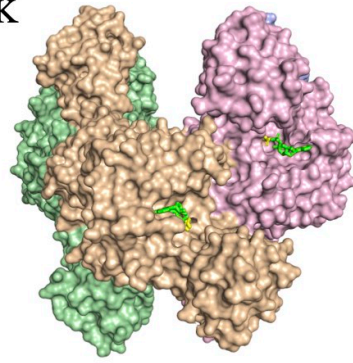
### 3.1.5 PK AlloSigma Experiments

For *S. aureus* PK enzyme two experiments were conducted in two different sites, proposed allosteric site (Ayyildiz et al., 2020) and the known allosteric site where the inhibitor IS-130 was bound, which was first reported in 2012 for methicillin-resistant SaPK by ligand-based cheminformatic studies (Axerio-Cilies et al., 2012). Using the crystallographic structure to which the IS-130 inhibitor was bound (PDB ID: 3T0T), the residues interacting with the inhibitor were selected as perturbation sites in AlloSigMA experiments. As illustrated with dark blue spots in Figure 3.5, both proposed allosteric site and known allosteric site were located at the central region of SaPK. According to results of these experiments, the allosteric effects of the proposed site and the known site were compared. As illustrated in Figure 3.6(a), the perturbation on the proposed allosteric site exhibited a significant restriction on the dynamics of the catalytic sites in all chains, while the perturbation of the known allosteric site exhibited an increase in the dynamics of the catalytic sites. Additionally, AlloSigMA experiments for *H. sapiens* and *L. mexicana* PKs were performed. Here, one experiment was conducted for each species. The perturbed region for *H. sapiens* PK was the top druggable allosteric site and it has the same location with the proposed site for *S. aureus* PK. On the other hand, the perturbed region for *L. mexicana* PK was the proposed allosteric site and its location far away from the central opening. According to the AlloSigMA results for parasitic PK, it was observed a negative  $\langle \Delta G \rangle$  value in the catalytic regions of A, C, and D chains. But,

in the B chain of *L. mexicana* PK, a small positive  $\langle \Delta G \rangle$  value was observed. Therefore, while there is an inhibition on three chains, there is an increase in dynamics in one chain. In other words, there cannot be complete inhibition here. The same result was observed in the results of *H. sapiens* PK experiments. While a negative  $\langle \Delta G \rangle$  value was observed in the catalytic regions in the A, B, and D chains, an increase in the dynamics of the catalytic regions in the C chain was observed. Thus, the common result in both *L. mexicana* and *H. sapiens* PK experiments is partial inhibition, not complete inhibition. According to AlloSigMA experiments conducted on PK, a complete inhibition was observed upon *S. aureus* PK proposed top druggable allosteric site perturbation, whereas in *L. mexicana* and *H. sapiens*, partial inhibition was observed.



*S. aureus* PK

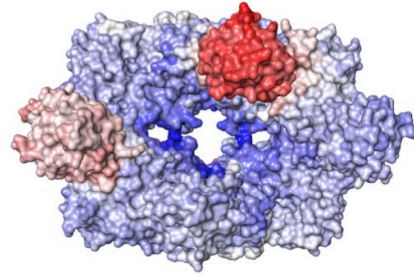
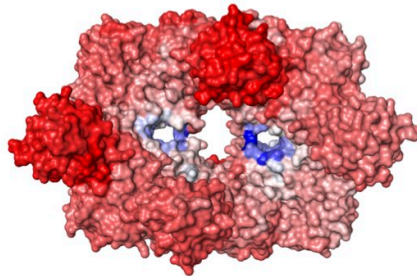
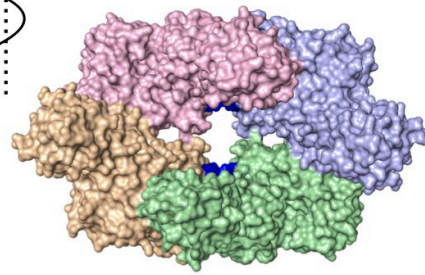
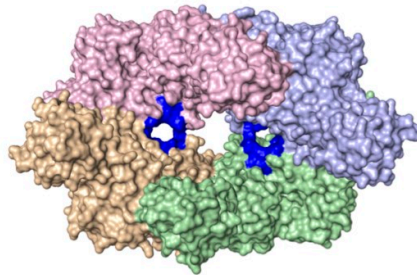



Proposed Site  
(Top Druggable)

Known  
Allosteric Site



90°  
.....  
.....  
.....



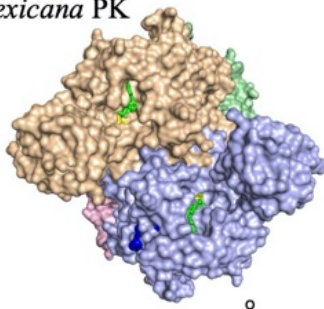
0.83  -2.51

1.27  -0.59

(a)

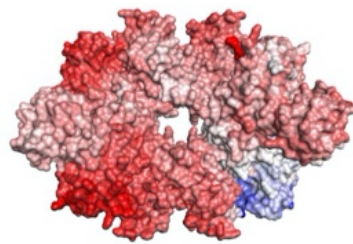
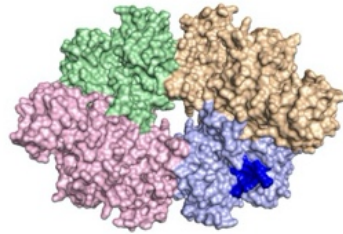



*L. mexicana* PK



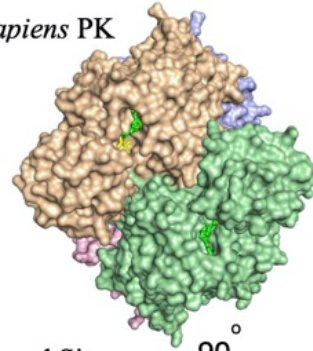
Proposed Site  
(Top Druggable)

90°



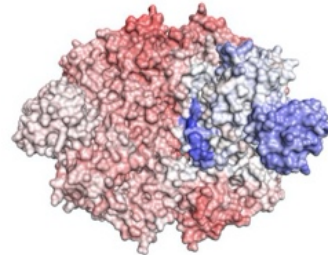
0.25  -1.21


*H. sapiens* PK



Proposed Site  
(Top Druggable)

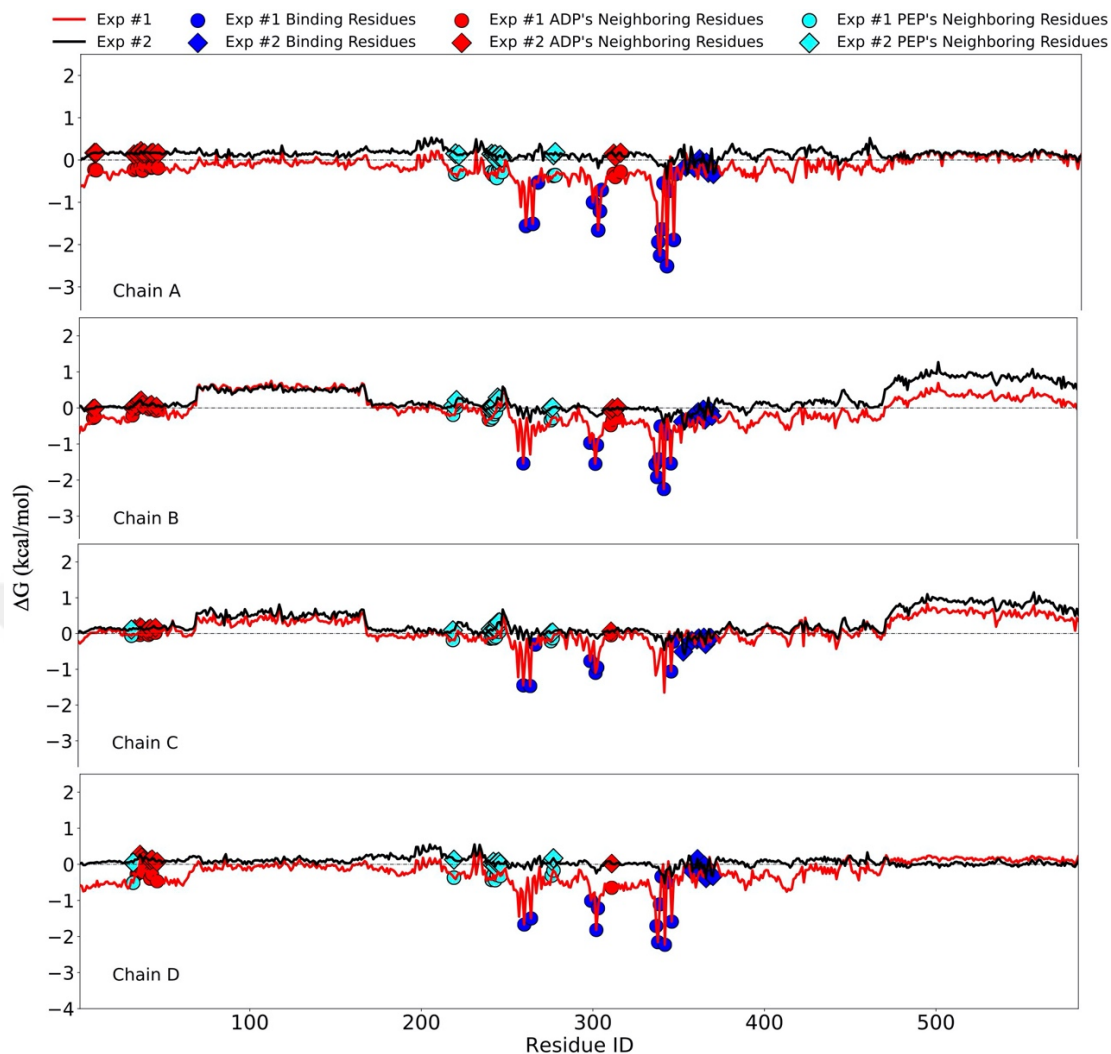
90°



0.92  -1.51

(b)

Figure 3.5: Perturbation sites in PK of all three species.



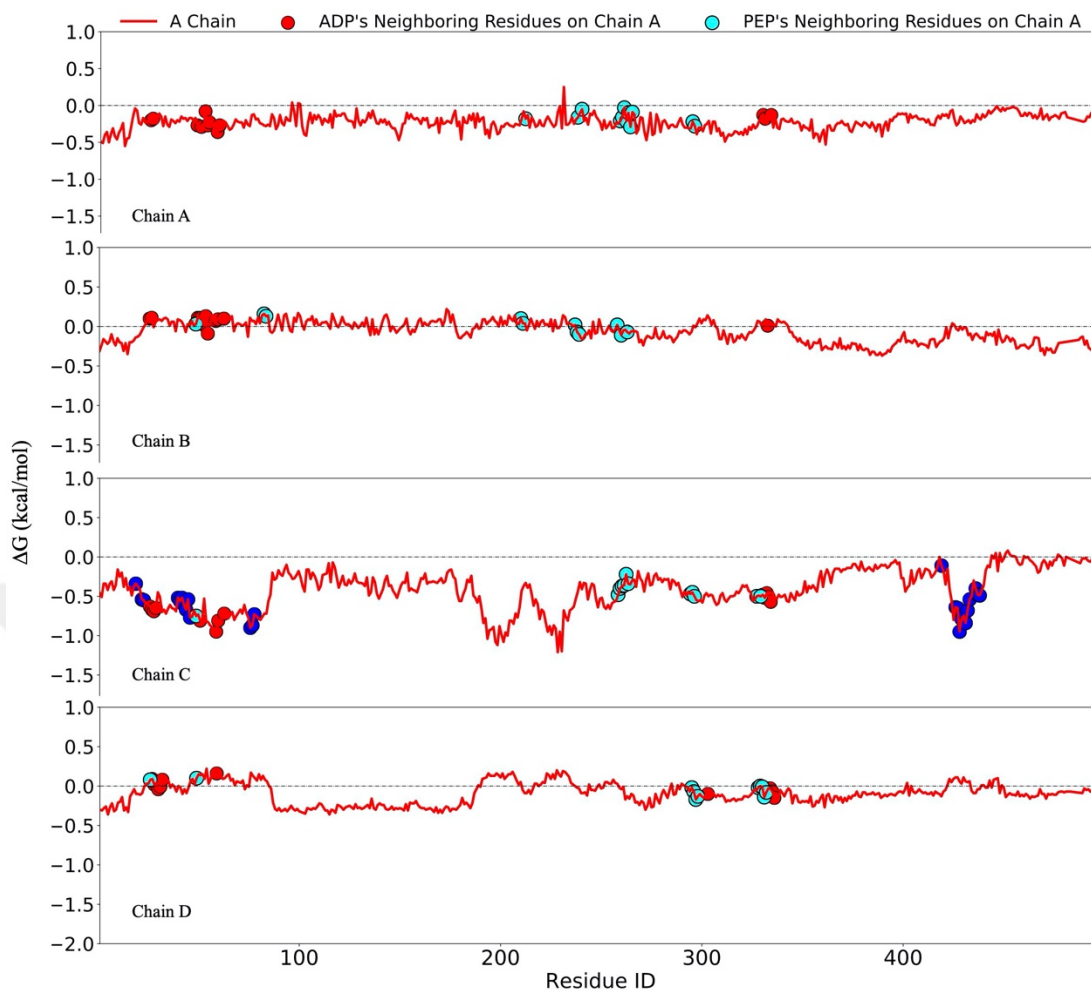
***S. aureus* PK Exp #1 Proposed Site**

$\langle \Delta G \rangle$ (kcal/mol)		
Chain ID	Substrate (PEP)	Cofactor (ADP)
A	$-0.31 \pm 0.06$	$-0.23 \pm 0.07$
B	$-0.16 \pm 0.17$	$-0.13 \pm 0.18$
C	$-0.10 \pm 0.09$	$-0.01 \pm 0.04$
D	$-0.33 \pm 0.10$	$-0.30 \pm 0.20$

***S. aureus* PK Exp #2 Known Site**

$\langle \Delta G \rangle$ (kcal/mol)		
Chain ID	Substrate (PEP)	Cofactor (ADP)
A	$0.11 \pm 0.04$	$0.17 \pm 0.06$
B	$0.09 \pm 0.10$	$0.06 \pm 0.06$
C	$0.13 \pm 0.10$	$0.16 \pm 0.07$
D	$0.06 \pm 0.06$	$0.18 \pm 0.06$

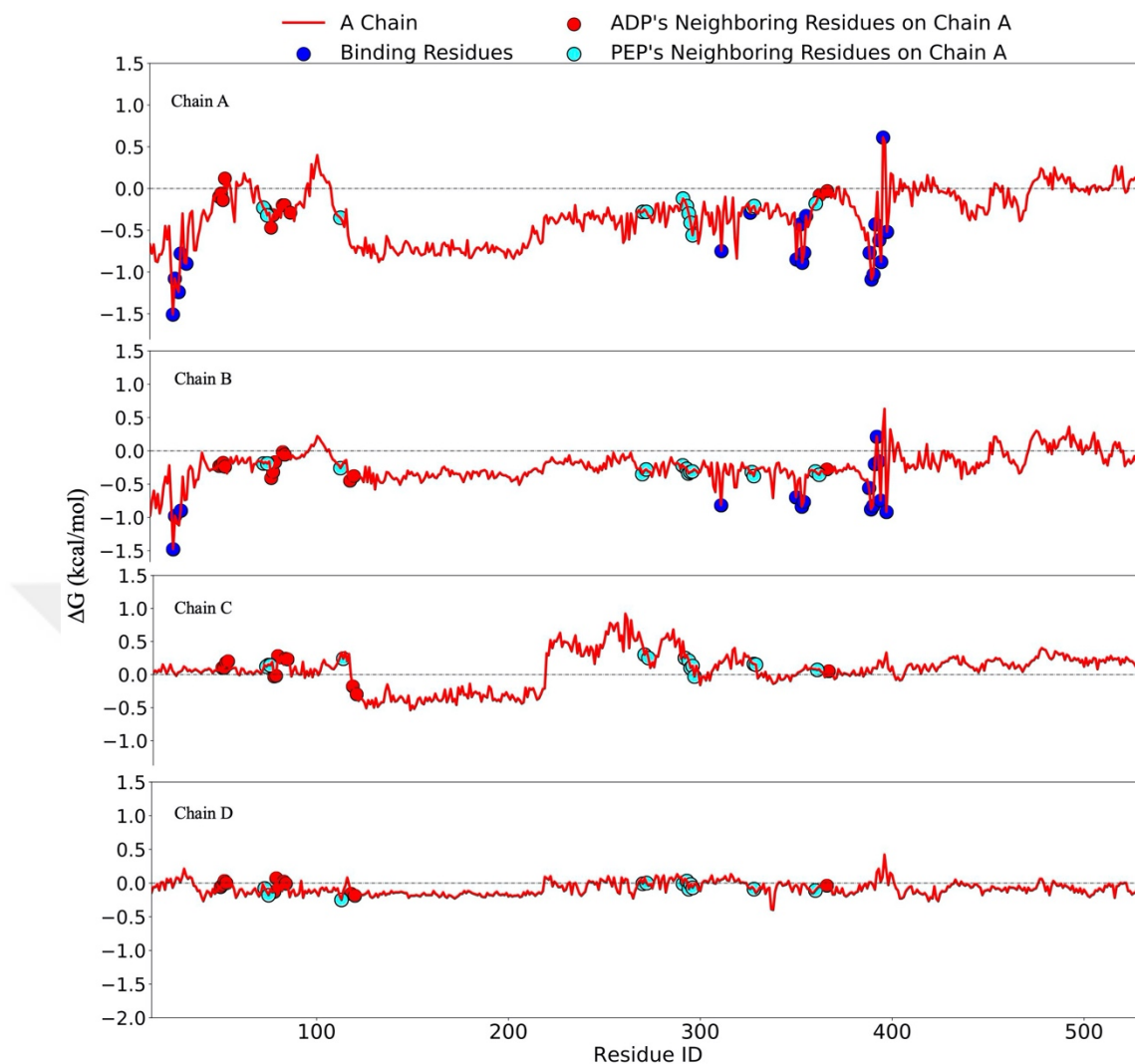
(a)



***L. mexicana* PK Proposed Site**

$\langle \Delta G \rangle$ (kcal/mol)		
Chain ID	Substrate (PEP)	Cofactor (ADP)
A	$-0.17 \pm 0.08$	$-0.22 \pm 0.08$
B	$0.01 \pm 0.09$	$0.08 \pm 0.06$
C	$-0.45 \pm 0.12$	$-0.65 \pm 0.14$
D	$-0.04 \pm 0.08$	$-0.01 \pm 0.08$

(b)



***H. sapiens* PK Proposed Site**

Chain ID	$\langle \Delta G \rangle$ (kcal/mol)	
	Substrate (PEP)	Cofactor (ADP)
A	$-0.28 \pm 0.11$	$-0.18 \pm 0.14$
B	$-0.29 \pm 0.06$	$-0.25 \pm 0.12$
C	$0.16 \pm 0.09$	$0.08 \pm 0.15$
D	$-0.07 \pm 0.08$	$-0.06 \pm 0.08$

(c)

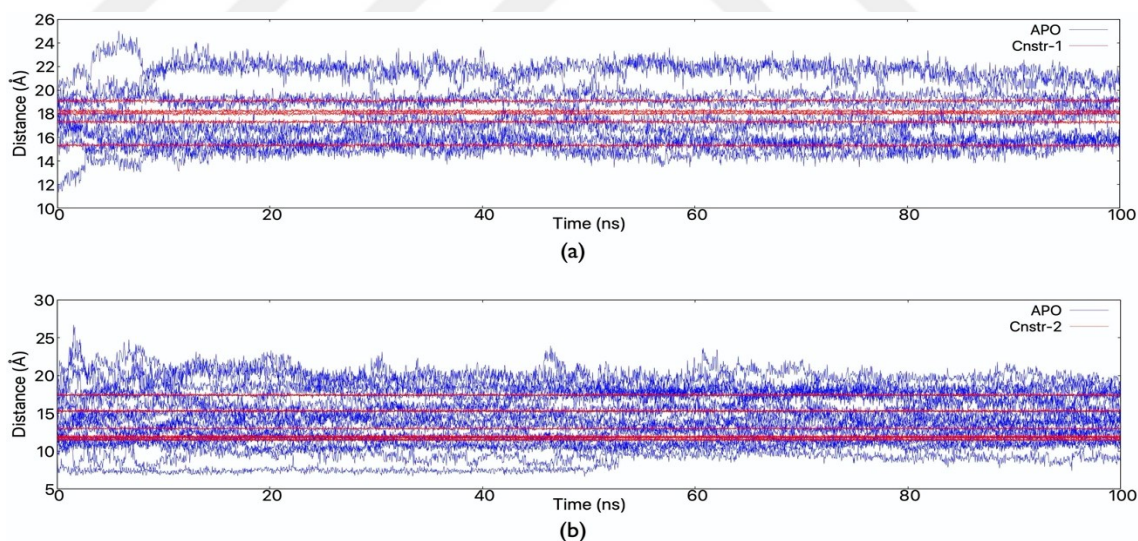
**Figure 3.6:**  $\Delta G$  energy profiles of binding, and  $\langle \Delta G \rangle$  values in the catalytic sites for (a) *S. aureus* PK, (b) *L. mexicana* PK, and (c) *H. sapiens* PK. The figure is adapted from submitted paper (Çelebi et al., 2021).

## 3.2 Molecular Dynamics Simulations of Phosphofructokinase

Using the nine trajectory files that were obtained as a result of performed Molecular Dynamics Simulations, the trajectory files were analyzed with computational methods to reveal the differences and restriction effect on protein dynamics in apo simulations and constrained simulations as mentioned in Methods section 2.2.

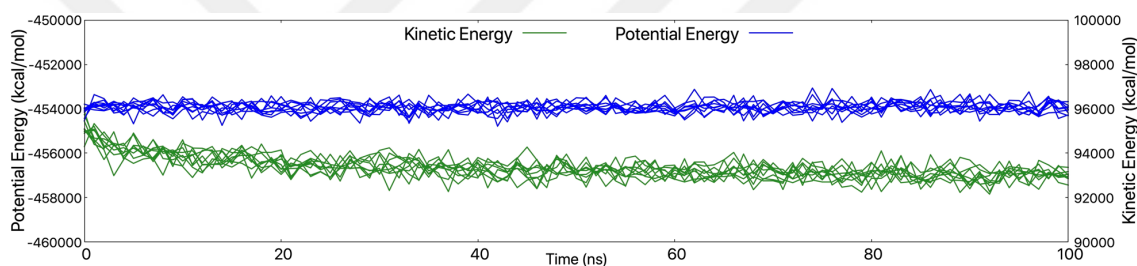
### 3.2.1 Change in Constrained Distances and System's Energy for MD Runs

To compare the change in restrained distances in constrained runs with apo runs, 10 distances for the proposed site (constrained-1) and 20 distances for the known allosteric site (constrained-2) were analyzed throughout the trajectory. According to Figure 3.7, in constrained runs the distances between selected residue pairs remained almost constant throughout the simulations (red lines), whereas in apo runs, they fluctuated about 1-2 Å as illustrated with blue lines (see Table 2.5 to selected residue pairs). Accordingly, protein's structure and dynamics were restrained as if a ligand was bound on the proposed and known allosteric sites.



**Figure 3.7:** The distance profiles of the constrained inter-atomic distances (a) 10 distances for the proposed allosteric site (constr-1) (b) 20 distances for the known allosteric site (constr-2).

Before discussing the MD simulation analysis, all trajectory files were checked whether the simulations were completed as desired parameters such as constant temperature and pressure or whether the protein interfered with its image in another cell during the simulation. The energy profiles of each 100 ns long simulations started with different initial velocities were provided in Figure 3.8. Nine independent runs were performed; three for apo state, three for the first constrained state (proposed allosteric site, labeled as “constr-1”) and the second constrained state (known allosteric site, labeled as “constr-2”). In the analysis of MD simulations, the energy values reaching a plateau indicate that the system has reached equilibrium. Therefore, it was important to plot the energy values to check whether the system has reached equilibrium. In Figure 3.8, kinetic and potential energy profiles recorded from a total of nine trajectory files were given. Accordingly, it was observed that the system reached equilibrium.



**Figure 3.8:** Potential and kinetic energy profiles of nine simulations. Energy values were shown every 40 ps in the energy profile. See top legend for color code.

### 3.2.2 RMSD and RMSF Calculations

A common linear analysis method for MD trajectory analysis is the calculation of Root-Mean-Square-Deviation (RMSD) and Root-Mean-Square-Fluctuation (RMSF) (Humphrey et al., 1996). An MD trajectory file consisting of thousands of frames, each frame represents a distinct conformation of the protein. As moving throughout the frames in the trajectory file, the conformation of the protein changes. The atoms or group of atoms in the protein constantly change their position throughout the simulation relative to the first frame of simulation or the initial 3D structure. Here, RMSD is the average displacements of the atoms. Via RMSD calculations, one can analyze time-dependent motions of the structure. RMSD value was calculated by the following equation:

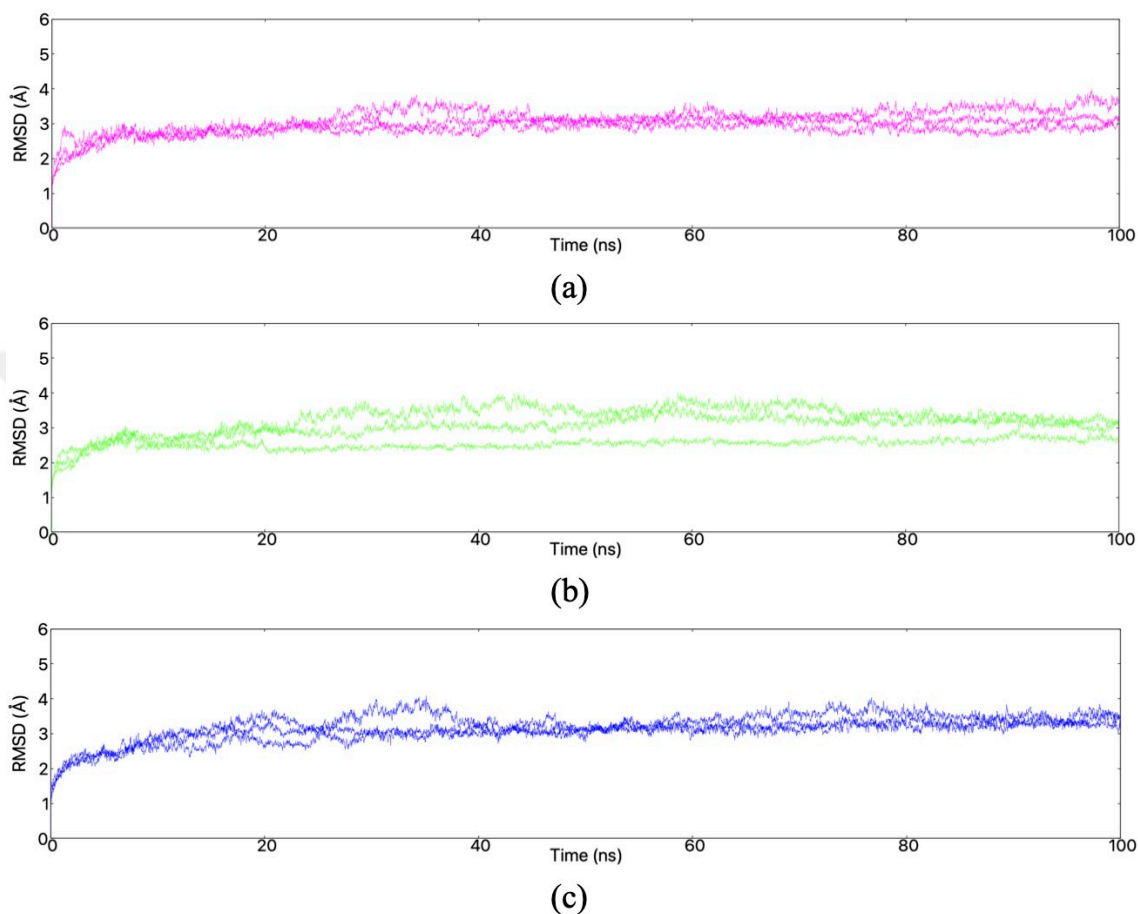
$$\begin{aligned}
RMSD (r_i) &= \sqrt{\frac{1}{n} \sum_{i=1}^n \|r_i(t) - r_i(0)\|^2} \\
&= \sqrt{\frac{1}{n} \sum_{i=1}^n (r_{ix}(t) - r_{ix}(0))^2 + (r_{iy}(t) - r_{iy}(0))^2 + (r_{iz}(t) - r_{iz}(0))^2}
\end{aligned} \tag{3.1}$$

where  $n$  is the total number of conformations in the trajectory file, and  $r_i(t)$  is the position of  $i^{\text{th}}$  atom at time  $t$ ,  $r_i(0)$  is the initial position of  $i^{\text{th}}$  atom in the initial frame ( $t = 0$ ). At the beginning of MD simulations, the RMSD values suddenly increase. These increases continue until the system reaches equilibrium which means the structure is fluctuating around a stable conformation. As an alternative linear analysis method for MD trajectory, it is helpful to calculate fluctuations of atoms or group of atoms of the protein relative to well-defined average position. RMSF is the amount of average fluctuation from the defined average position of the residues. RMSF value was calculated by the following equation:

$$RMSF (x) = \sqrt{\frac{1}{T} \sum_{t=0}^T (r_i(t) - \langle r_i \rangle)^2} \tag{3.2}$$

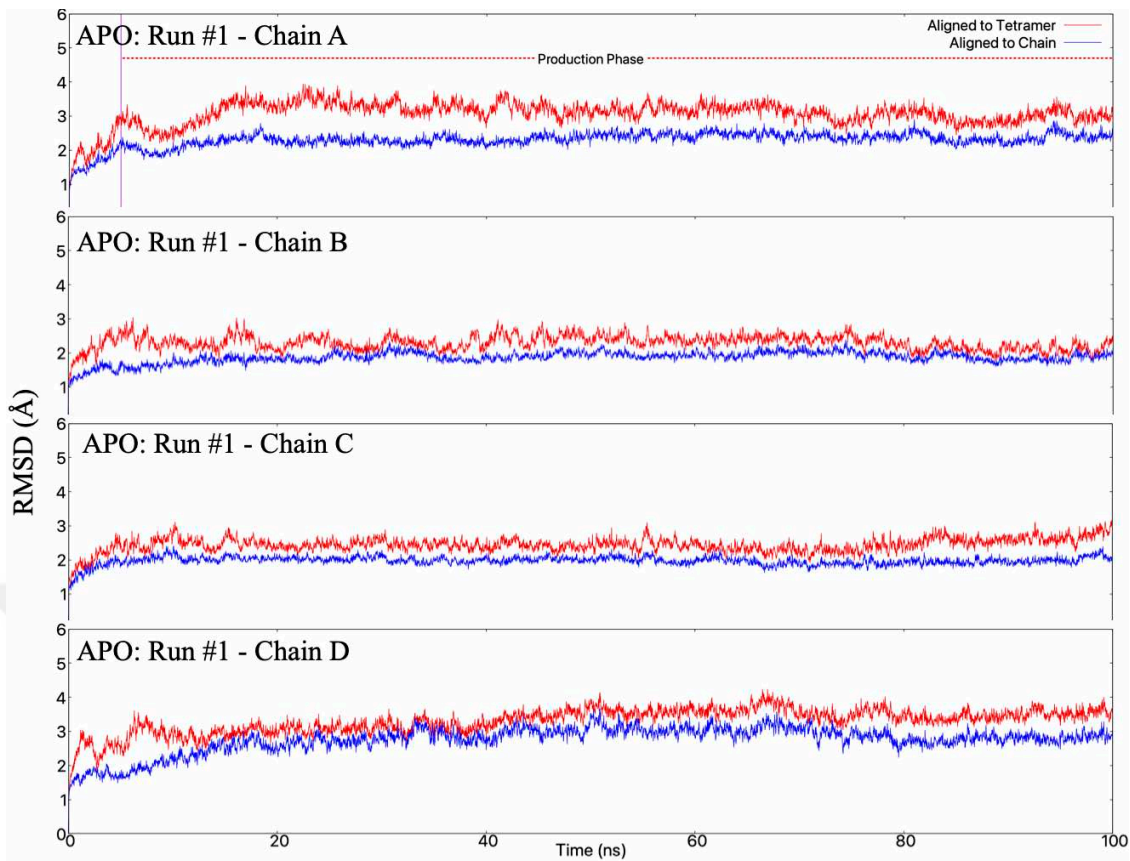
Where  $r_i(t)$  is the instantaneous position of residue  $i$ , and  $\langle r_i \rangle$  is the average position. As shown in the RMSD profiles in Figure 3.19, the data collected for each simulation comes from the next 95 ns production phase, since the first 5 ns is the time period in which the change occurs the fastest and represents the equilibrium phase. The system has reached equilibrium, which is approximately 3-3.5 Å different from its initial state. When each chain in the tetramer was aligned according to itself, and the RMSD was calculated, lower RMSD values were found. In the RMSD profiles shown in Figure 3.10 for the first runs of each of the apo, constr-1 (proposed site) and constr-2 (known site) simulations, the red lines show the RMSD change when aligned with the tetramer structure, which is 1-1.5 Å higher than the RMSD change when aligned with the single chain (blue lines). The reason behind this difference was that the RMSD with tetramer alignment also includes the displacement of the chains relative to each other. The highest difference was observed in constr-1 Chain-A, which indicated the displacement of Chain A with respect to others. On the other hand, there is no difference in the first apo runs in general (Figure 3.10 (a)). However, RMSD

profiles with high differences were observed in the second and third apo runs (see Appendix A3). Therefore, it was not possible to derive a distinction between apo and constrained simulations based on the differences in RMSD values when aligned with respect to chain or tetramer.

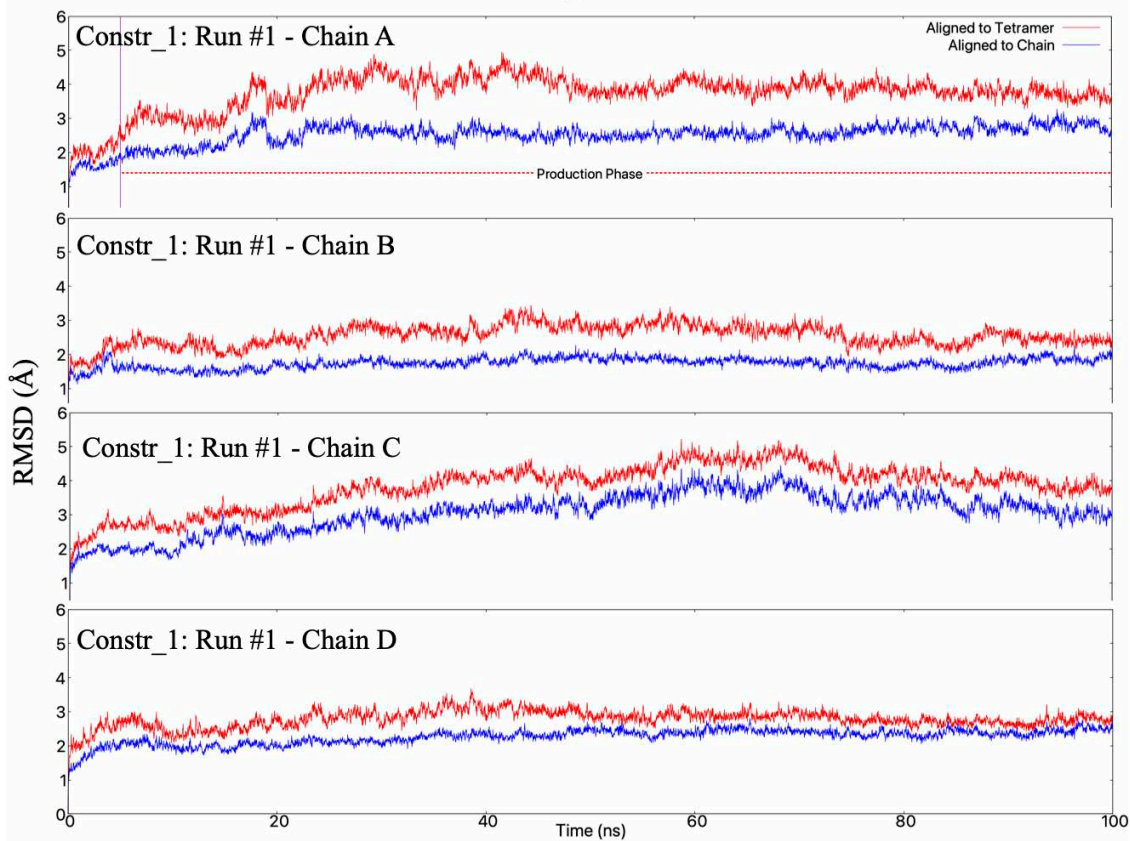


**Figure 3.9:** RMSD plots for three different simulation runs in phosphofructokinase enzyme (a) apo form, (b) constr-1 (proposed allosteric site) and (c) constr-2 (known allosteric site).

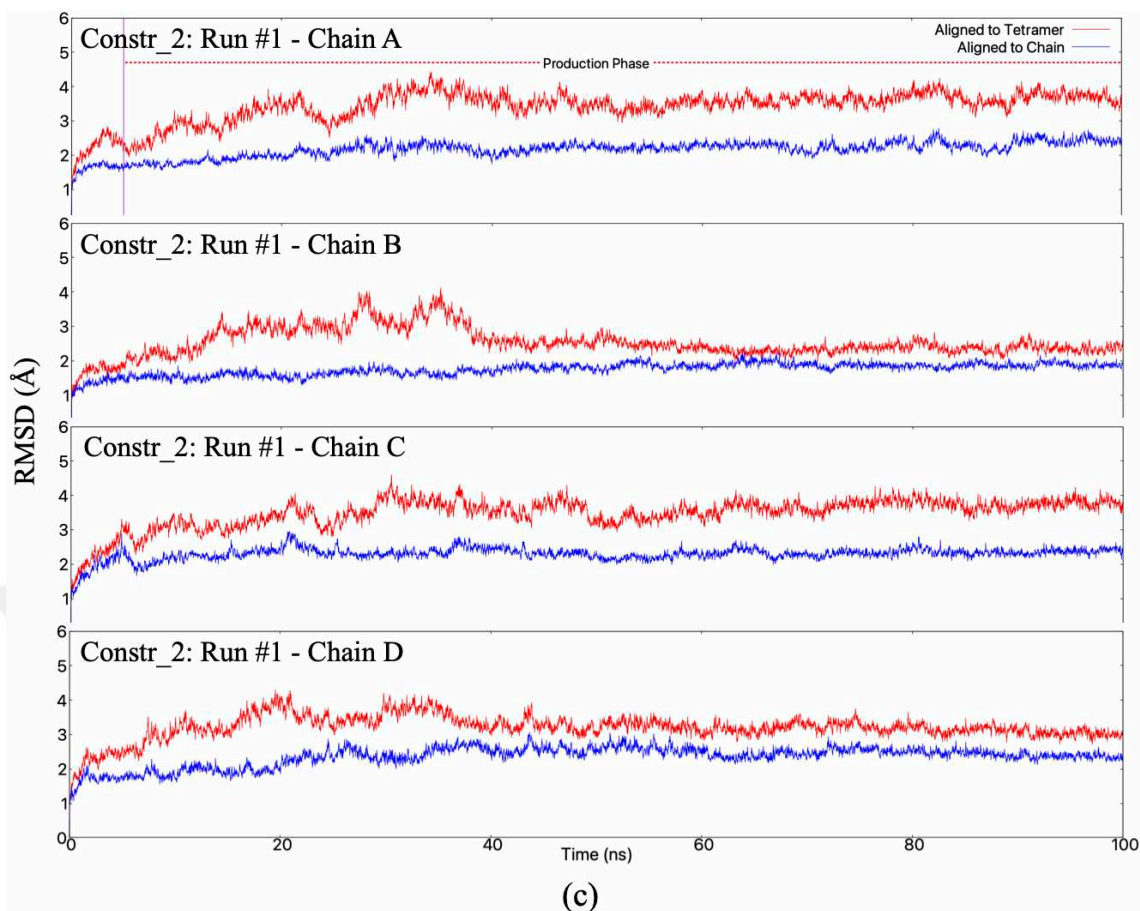




(a)

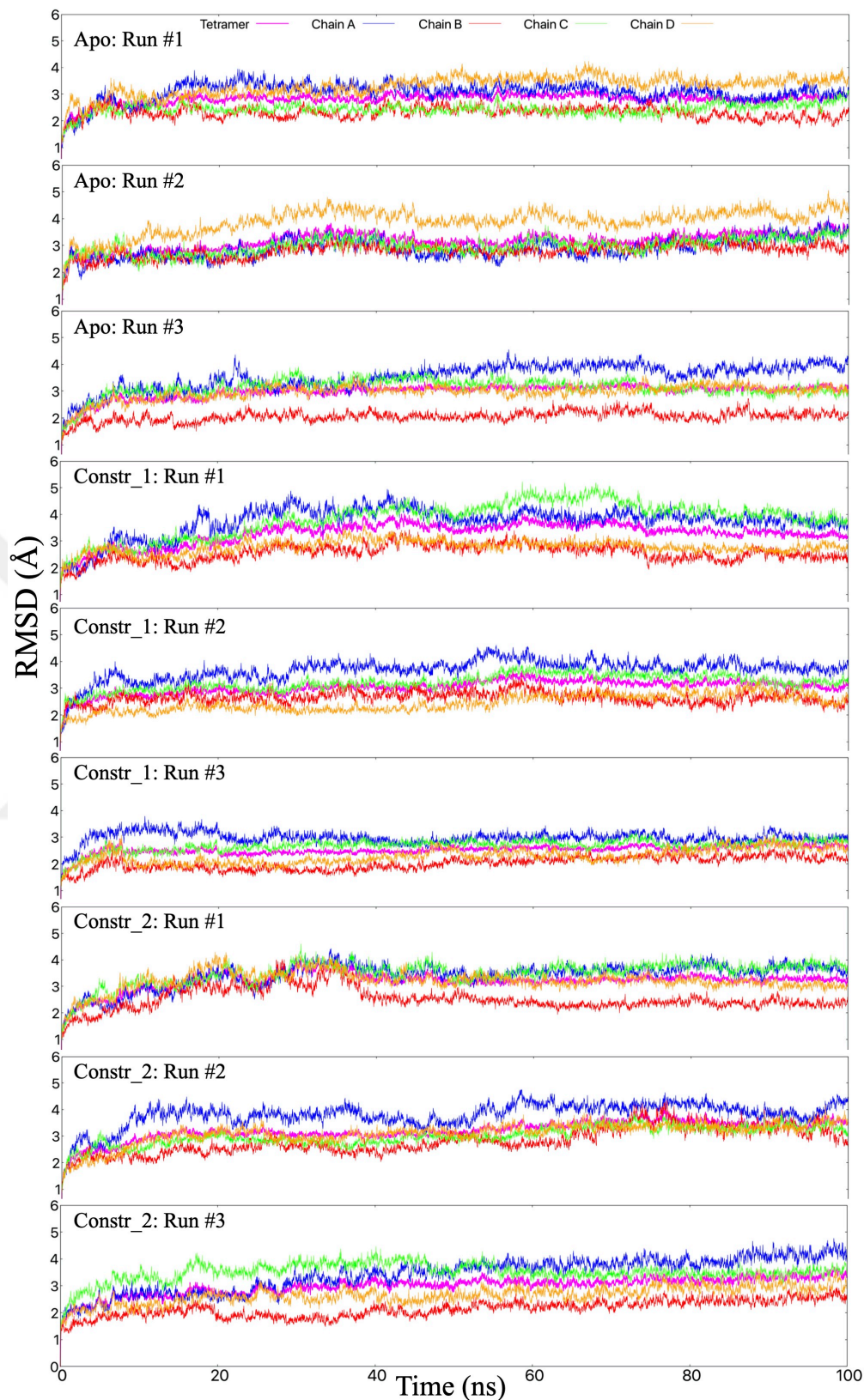


(b)



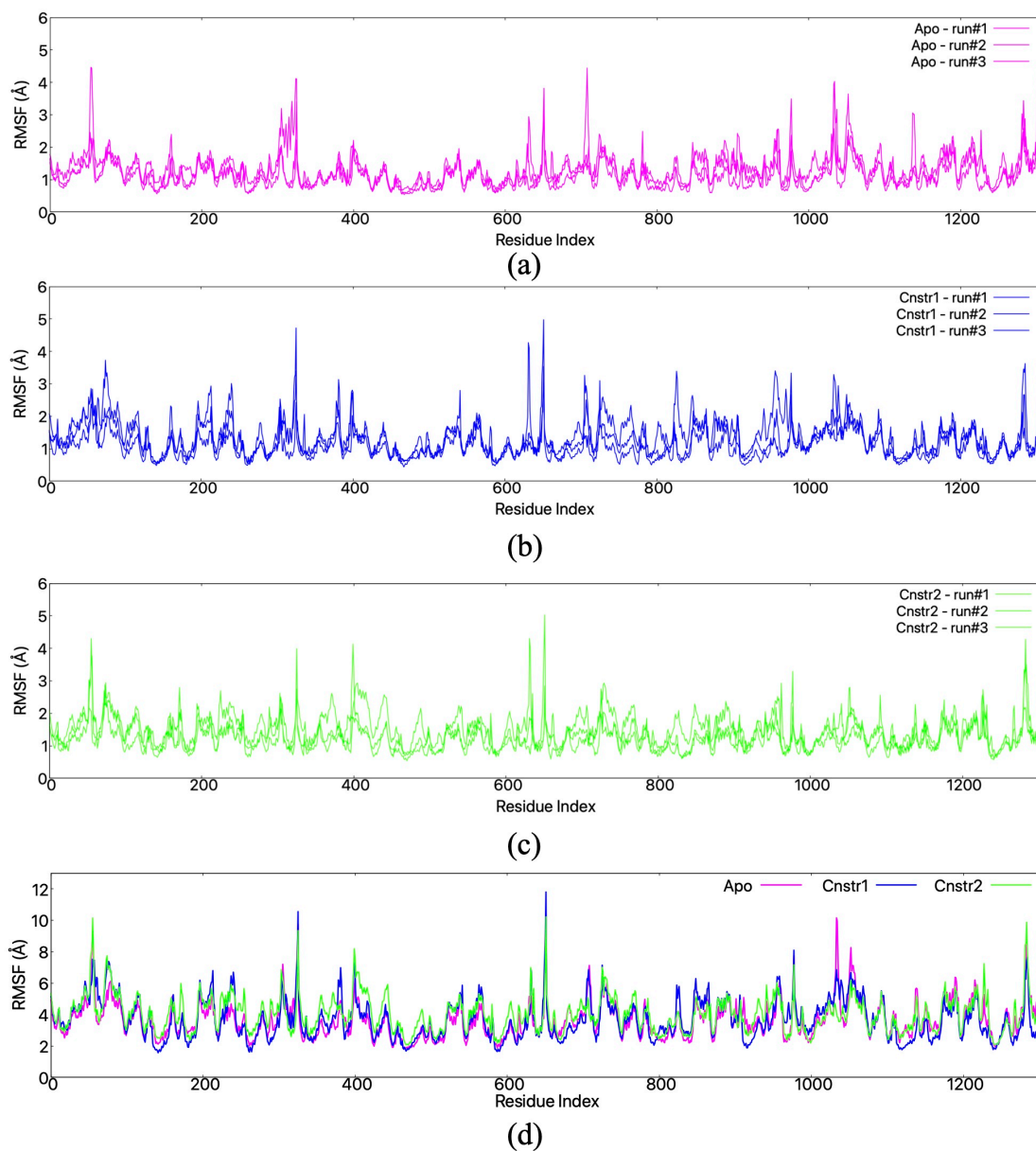
**Figure 3.10:** Two different states aligned RMSD plots of (a) apo form Run #1, (b) proposed allosteric site (constr-1, Run #1), and (c) known allosteric site (constr-2, Run #1. The red lines represent tetramer-aligned RMSD, while the blue lines represent chain-aligned RMSD.

To better reveal the displacements of chain and tetramer structure, chain RMSD and tetramer RMSD profiles for each simulation were calculated by aligning tetramer structure, then RMSD profiles were given in the same profile as shown in Figure 3.11. It has been observed that the lowest RMSD change in these chain RMSD profiles is 1.5 - 2 Å, and the highest RMSD is around 4 Å. In addition, it was observed that the change in tetramer RMSD profiles was around 3 Å. According to Figure 3.11, no significant divergence was observed in the RMSD profiles of the chains and the tetramer structure of PFK in both apo simulations and constrained simulations which indicate that the system has reached an equilibrium in each run.

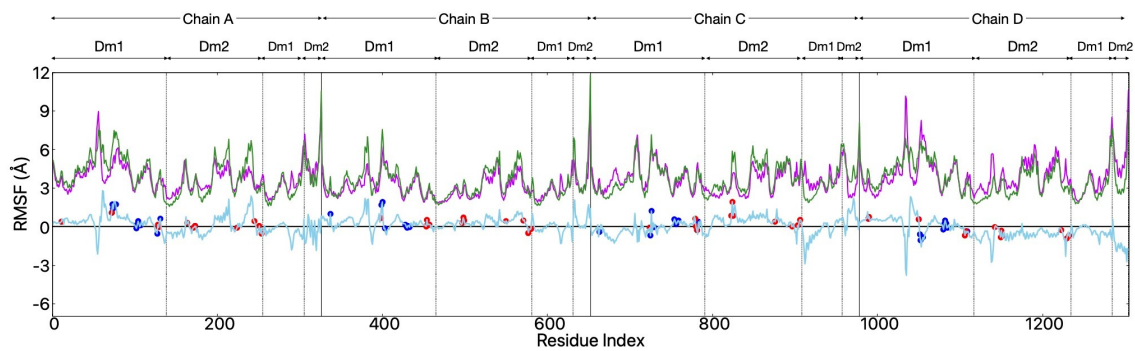


**Figure 3.11:** RMSD profiles of the chain and tetramers in each run of three state (apo, constr-1 and constr-2) simulations are given. The blue, red, green, orange, and magenta colored lines represents the RMSD profiles of Chain A, Chain B, Chain C, Chain D, and tetramer, respectively.

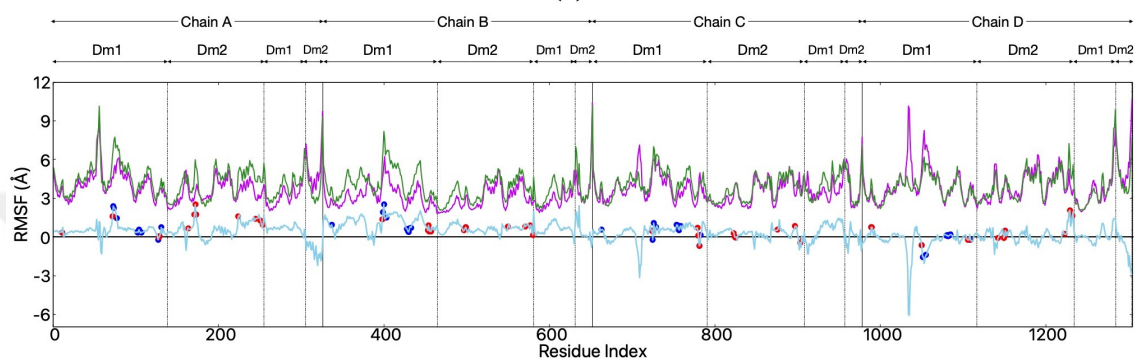
During trajectory analysis, RMSF calculations were also performed to reveal mobile regions in the protein in each run of three simulation states. As shown in Figure 3.12, the RMSF values obtained from three separate simulation runs for each different state were plotted together on the same figure. Also, the summation of RMSF profiles of each run for three state simulations were illustrated in Figure 3.12(d), and accordingly the difference between the RMSF values of apo and constrained simulations is nearly the same which indicate that the residues of the PFK were less fluctuated from the average positions. On the other hand, in order to reveal the RMSF difference between apo and constrained simulations more clearly, the difference between constrained and apo sums was taken as shown in Figure 3.13. It was observed that the effect of the restrictions applied to the proposed and known allosteric sites on protein dynamics was not reflected in the RMSF values. Additionally, there are some increases in both proposed and known allosteric site constrained simulations, but they are negligible. On the other hand, there is sharp decrease in RMSF for domain-1 region of chain D in both constrained simulations. This corresponds to a loop region in apo form which later transformed into a helical state in constrained simulations. In fact, it was observed that the alpha-helix structure in this region consisted of 4 residues (Leu54 – Val57) in the constr-1 simulations, while it was observed that it consisted of 8 residues (Leu54 – Ile61) in the constr-2 simulations. Additionally, the dynamics of the catalytic region to which the residues that are interacted with substrate and cofactor have not changed much as a result of the restrictions. The interacted residues with the substrate and cofactor were indicated in Figure 3.13(a), and (b) as filled circles in red and blue, respectively. In addition, as illustrated in Figure 3.13 (c), the differences of the sum of RMSF values of the apo and constrained simulations were illustrated on *Sa*PFK snapshot (blue: negative, white: zero, red: positive). It was aimed to reveal the regions where the difference increased the most.



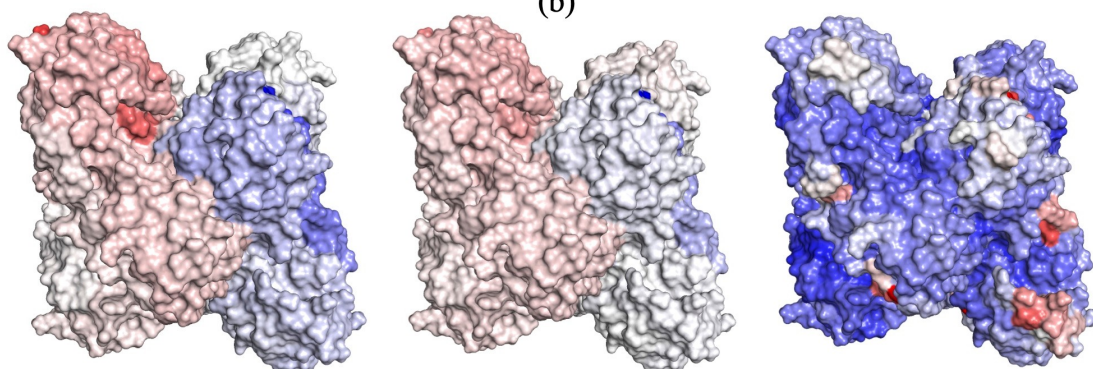
**Figure 3.12:** RMSF plots for three different simulations run in phosphofructokinase enzyme (a) apo form, (b) proposed allosteric site (constr-1), and (c) known allosteric site (constr-2). and (d) sum of RMSF values of all runs for each simulation state.



(a)



(b)



(c)

**Figure 3.13:** The sum of RMSF values for three runs in phosphofructokinase enzyme and their difference from apo (a) constr-1 (proposed), (b) constr-2 (known). The red and blue circles represent the residues nearby the substrate (F6P) and cofactor (ATP), respectively. (c) *Sa*PFK is colored according to the sum of the RMSF values for the Apo simulations, the difference of the sums of the RMSF values for the constr1 - apo simulations, and the constr2 - apo simulations, respectively (increasing from blue to white and red).

### 3.2.3 Principal Component Analysis (PCA)

In a biological system, a protein is constantly fluctuating because it is not a rigid structure, and a protein molecule may have both low- and high-frequency motions which can be observed during MD simulations. The dynamics of a protein can be divided into a small number of collective variables in order to describe these large-scale motions (Amadei et al., 1993; García, 1992; S. Hayward et al., 1993; Kitao et al., 1991). Altogether these collective variables define the subspaces, which are essential to explain dynamics – fluctuation relationship in a protein (Steven Hayward et al., 1995). These collective variables can be calculated via Principal Component Analysis (PCA) of MD trajectory. PCA method was first introduced under the name “quasi-harmonic analysis” to calculate configurative entropies of proteins (Ichiye & Karplus, 1991; Karplus & Kushick, 1981; Perahia et al., 1990).

The motions of the atoms that are frequently fluctuating in the water-box can be characterized by the PCA method to calculate collective variables, which are called “principal modes.” These principal modes are ordered according to their contributions to the dynamics of the protein. The first mode is the one that made the biggest contribution to explaining protein dynamics. The second mode is the one that made the second-biggest contribution to explain the protein dynamics. In general, the modes that contribute to explaining 70% of total motion and 90% of total motion are calculated in PCA analysis.

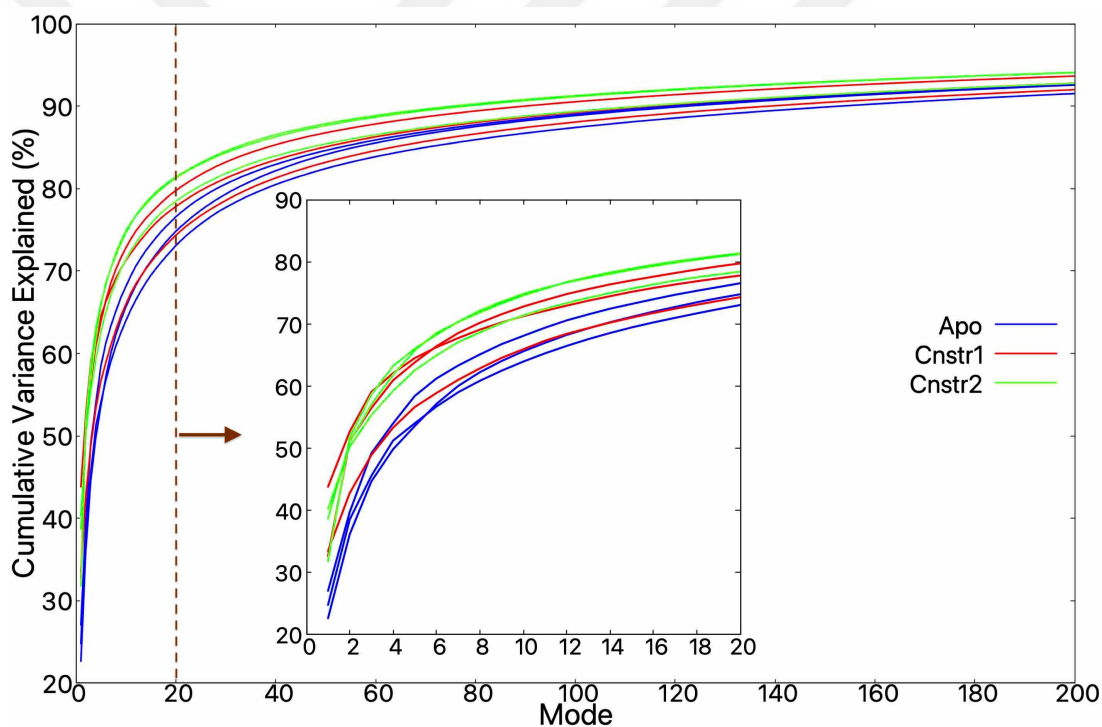
In this thesis, PCA was performed for all three runs of all three simulation states. Using the data obtained from PCA analysis, the regions moving in correlation with each other were revealed. Equation 3.3 was used to calculate the correlations between the positional fluctuations of two residues  $i$  and  $j$  with respect to their average position.

$$C_{i,j} = C(\Delta R_i, \Delta R_j) = \frac{\langle \Delta R_i(t) \cdot \Delta R_j(t) \rangle}{\sqrt{\langle (\Delta R_i)^2 \rangle \langle (\Delta R_j)^2 \rangle}} \quad (3.3)$$

where  $\Delta R_i(t)$ , and  $\Delta R_j(t)$  is the time dependent position of residue  $i$ , and  $j$ , respectively. The average in the numerator was taken over the entire trajectory and was normalized. Here, when  $C_{i,j}$  is equal to +1, it means that residues  $i$  and  $j$  are fluctuating in the same

direction. If  $C_{i,j}$  is equal to -1, the residues  $i$  and  $j$  are fluctuating in opposite directions. If  $C_{i,j}$  is equal to 0, the residues  $i$  and  $j$  are fluctuating independently.

PCA was performed via ProDy software tool (Bakan et al., 2011). For PCA calculation commands see Appendix A4. Accordingly, the contributions of principal modes were investigated. As illustrated in Figure 3.14, while the first 20 of the principal components in apo state represent 70-75% of the total motion whereas in constrained simulations, 80% of the motion can be explained with the first 20 modes. Percentage values corresponding to the modes used in the graphs are also detailed in Table 3.1 According to contribution percentages of principal modes, more components are needed to explain the total motion of apo state simulations.



**Figure 3.14:** Principal component analysis for three different simulation runs in the phosphofructokinase enzyme. The explanation percentage of the total motion of the modes is given cumulatively.



**Table 3.1:** The percentage explanation of the total motion of the modes obtained from the principal component analysis for three different simulation runs in the phosphofructokinase enzyme.

APO								
Run #1			Run #2			Run #3		
Mode	% Explained	Cumulative % Explained	Mode	% Explained	Cumulative % Explained	Mode	% Explained	Cumulative % Explained
1	24.8	24.8	1	27.05	27.05	1	22.7	22.7
2	13.8	38.6	2	12.7	39.8	2	13.5	36.1
3	7.01	45.6	3	9.5	49.2	3	8.5	44.7
13	1.02	67.6	12	1.2	70.6	10	1.6	65.6
16	0.8	70.3	13	1.0	71.6	14	0.98	70.3
25	5	75.7	19	0.6	75.98	22	0.5	75.9
38	0.25	80.0	29	0.3	80.3	33	0.3	80.3
153	0.04	90.0	119	0.05	90.00	123	0.05	90.03

Cnstr-1 (Proposed Allosteric Site Constrained)								
Run #1			Run #2			Run #3		
Mode	% Explained	Cumulative % Explained	Mode	% Explained	Cumulative % Explained	Mode	% Explained	Cumulative % Explained
1	32.6	32.6	1	43.8	43.9	1	33.5	33.5
2	18.6	51.2	2	8.9	52.7	2	9.3	42.8
3	5.3	56.5	3	6.4	60.0	3	6.1	48.9
4	4.5	60.1	4	3.03	62.1	4	4.4	53.3
5	2.9	63.9	5	2.4	64.5	5	3.3	56.7
8	1.7	70.2	9	1.15	70.3	14	0.9	70.2
21	0.43	80.2	26	0.3	80.1	35	0.3	80.1
90	0.06	90.01	115	0.05	90.02	139	0.04	90.0

Cnstr-2 (Known Allosteric Site Constrained)								
Run #1			Run #2			Run #3		
Mode	% Explained	Cumulative % Explained	Mode	% Explained	Cumulative % Explained	Mode	% Explained	Cumulative % Explained
1	31.8	31.8	1	38.7	38.7	1	40.3	40.3
2	20.0	51.8	2	12.3	51.0	2	10.0	50.3
3	6.9	58.7	3	6.0	57.0	3	5.1	55.4
4	4.5	60.1	4	4.8	61.9	4	3.9	59.3
5	2.8	66.0	5	3.9	65.7	5	3.3	62.6
7	2.1	70.4	7	1.7	70.3	9	1.5	70.2
18	0.5	80.4	17	0.5	80.1	24	0.4	80.0
79	0.06	90.1	76	0.06	90.05	114	0.05	90.04

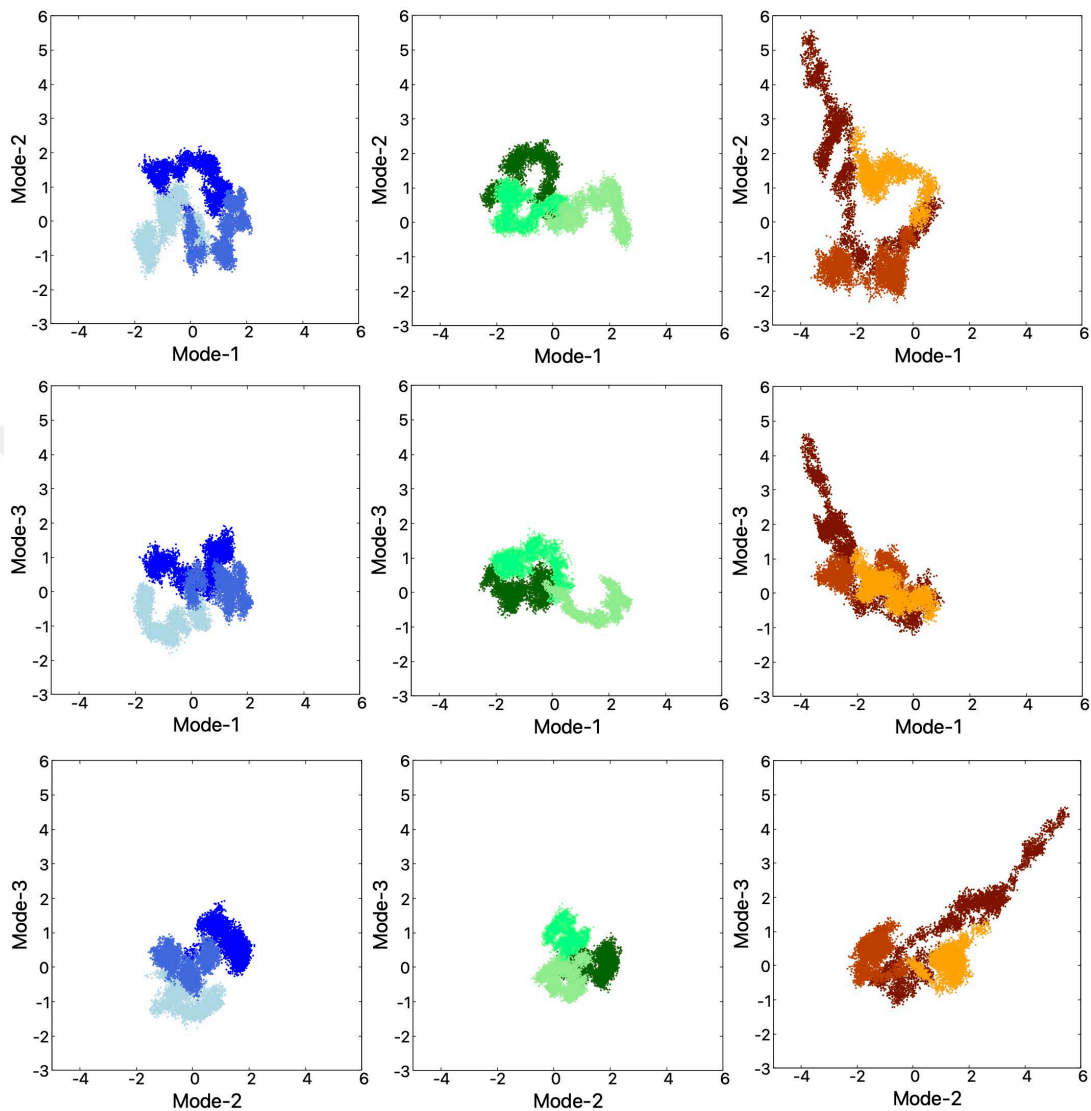
Also, an important result from PCA was extracted from the projection of a total of 5000 conformation (or snapshots) of the simulation trajectory on the first three principal components that represent the greatest variance. The projection of each snapshot in a trajectory is calculated by taking the dot product of the fluctuation vector of each atom from its average position with the eigenvector in that mode. The projection of each snapshot is calculated by the Equation 3.4:

$$\mu_i = v_i \cdot (q - \langle q \rangle) \quad (3.4)$$

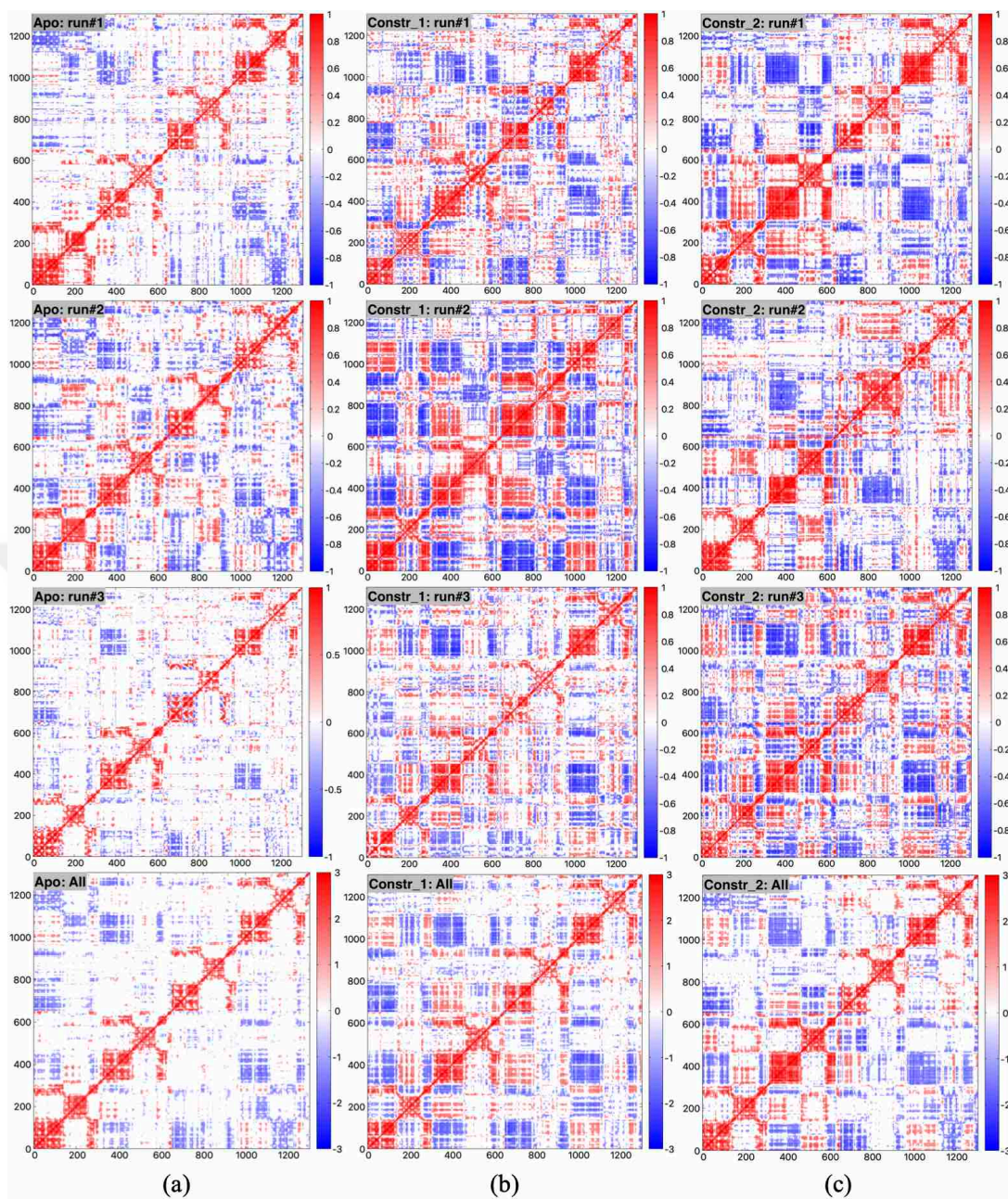
Where  $v_i$  is the eigenvector of  $i^{\text{th}}$  mode,  $q$  represent each snapshot's coordinates, and  $\langle q \rangle$  is the averaged coordinates of these snapshots. These projections were shown for mod1 versus mod2, mod1 versus mod3, and mod2 versus mod3. Each point in Figure 3.15 corresponds to one snapshot. Accordingly, the greatest variance was observed in restricted simulations. As illustrated in Figure 3.15, in apo simulations, the range of projections varied between -2 and 2 for all three modes. In contrast, in the constr-1 simulations, the ranges varied between -2 and 3 for mode1, -0.5 and 2.5 for mode2, and -1 and 2 for mode3. Also, in constr-2 simulations, the first mode was observed between -4 and 1, the second mode between -3 and 6, and the third mode between -1 and 5. It was observed that the intervals were significantly larger in the constr-2 simulations compared to constr-1 and apo simulations. The widening of the intervals indicates that the protein samples a wider range of conformations. This clearly shows that the performed restrictions changed the dynamic character of the protein, especially for constr-2.

Next, the simulation trajectory files were projected on all the components that explain 70% of the total dynamics. Then, the correlation values between the inter-residual fluctuation vectors were calculated. As shown in Figure 3.16, correlation maps in the first column were obtained for three different simulations in apo form. Similarly, the maps of the constrained simulations were given in the second (constr-1) and third columns (constr-2). The map in the last row is a summation of correlation values of all three runs. Accordingly, it was observed that there is a significant increase in both positive and negative correlations in constrained simulations. In order to more clearly reveal these increases and decreases in correlations, the differences of the sums of the simulations were taken. By taking the difference between the sums, the change was illustrated in

Figure 3.17(c), and 3.19(c) for two restricted simulations. These differences were investigated in more detail in the following sections.



**Figure 3.15:** Projection of 5000 different conformations on the first three principal components from principal component analysis for apo, constr-1 and constr-2 runs which are colored as blue and shades, green and shades, orange, and shades, respectively.



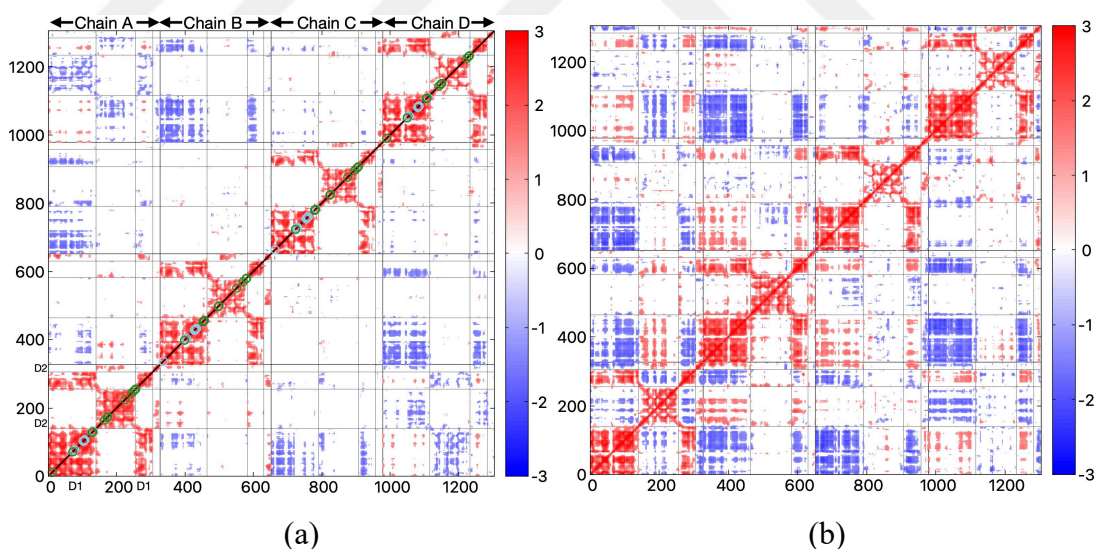
**Figure 3.16:** Correlation between fluctuating vectors of alpha carbon atoms calculated using components that explain 70% of the motion resulting from principal component analysis. (a) apo simulations (b) constr-1 simulations (proposed) and (c) constr-2 simulations. Values between -0.4 and 0.4 are not shown on all maps.

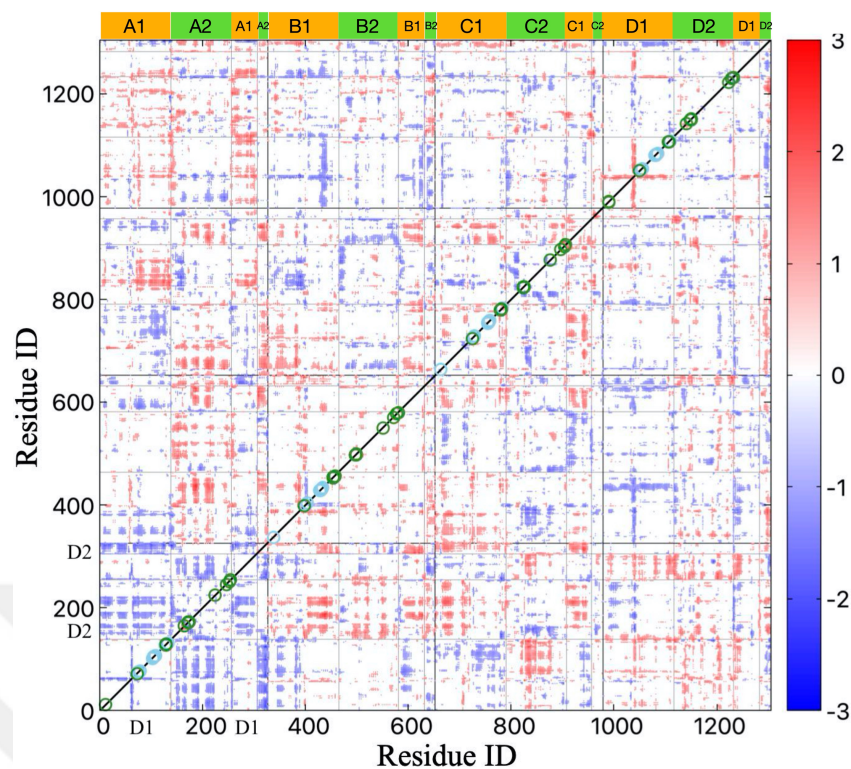
To better highlight the difference between apo and constrained simulations, the sum of correlations for apo and constrained sums were illustrated in the same figure with the difference map (see Figure 3.17 and Figure 3.19). Since the binding of the substrate F6P plays a critical role in the dimer-tetramer conversion of PFK enzyme (Tian et al., 2018), it was important to focus on the correlation changes of the residues nearby the F6P in correlation maps. Hence, in Figure 3.17 and 3.19, the residues nearby the substrate (F6P) and cofactor (ATP) were showed on the diagonal line. Accordingly, the cross correlations of residues near F6P in all chains were examined and a negative increase in cross correlations of residues near F6P in the A chain of both restricted simulations was observed. But it was seen that this increase is more clear in the constr-1 simulations. On the other hand, cross correlations of residues near F6P in the other three chains were observed to increase positively in both restricted simulations.

The boundaries of both chains and domains of PFK were specified in each map. According to Figure 3.17, it was observed that highly significant increases in correlations were observed between domains and chains in constr-1 simulations. In Figure 3.17(a), it was seen that there is no correlation between domain-1 (D1) and domain-2 (D2) within the same chains in apo simulations. The correlations of domains are seen as inter-chain. Interchain domains with a correlation value of 0 or with a certain correlation in apo simulations increased both positively and negatively in the constr-1 simulations (see Figure 3.17(b)). While a positive correlation increase was observed between Chain A - Domain 2 (A2) and Chain B - Domain 1 (B1), a negative correlation increase was observed between A1 and B1. Also, the negative correlation between some interchain domains observed in apo simulations was observed to increase further in the same direction in constr-1 simulations. This change is clearly observed with the increase in the negative correlation between A1 and C1 and between B1 and D1 in constr-1 simulations. Another important point is that some of the negative correlations between domains in apo simulations were observed to move in positive correlations in constr-1 simulations. This correlation change was clearly observed between C2 and A1. In addition to the inter-domain correlations in different chains, significant changes were observed in the inter-domain correlations in the same chain. While the correlation between domains in the same chain was zero in apo simulations, both negative and positive correlations were observed

in constr-1 simulations. This change was observed in all chains along the diagonal line except for Chain A. As seen in the map where the difference between constr-1 and apo was given in Figure 3.17(c), correlations in Chain A decreased. In addition, positive correlations observed in apo simulations along the diagonal line increased in constr-1 simulations. These correlation changes indicate the performed restrictions on proposed allosteric site had an impact on the dynamics of the protein.

In order to provide a better understanding of the correlation differences, the difference maps given in Figure 3.17(c), and 3.19(c) were divided into regions (chains), and the changes in the correlation between each chain (A, B, C, and D) in the tetramer structure were investigated, and the regions where the changes in the correlations were numbered. Then, the increase or decrease in correlations between dimers were indicated at the top (blue) and bottom (yellow) view of the cartoon illustration of the protein in Figure 3.18(c) and 3.20(c).





(c)

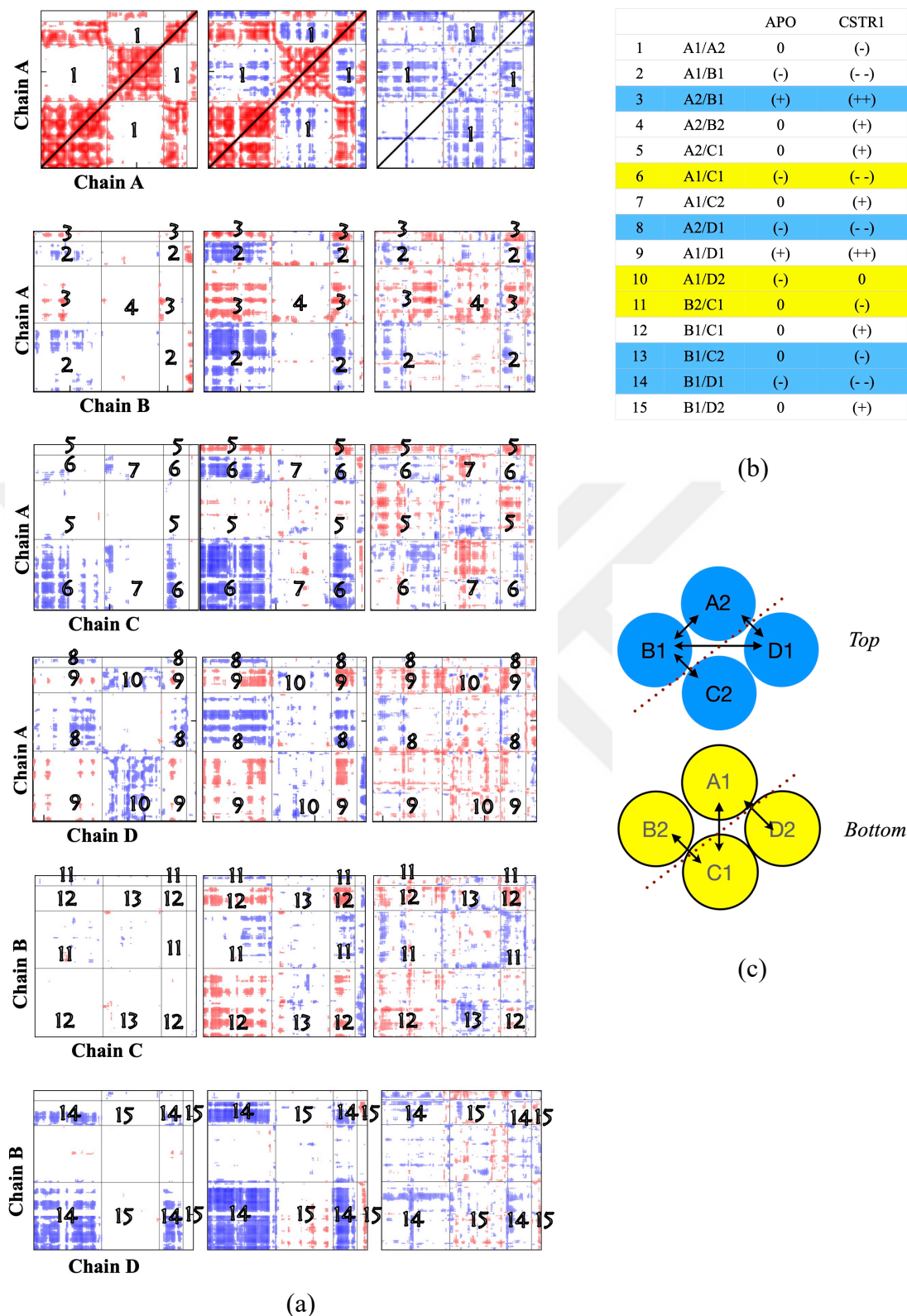
**Figure 3.17:** All three runs of apo and constr-1 simulations were summed up and their difference was taken. PCA Maps of (a) apo simulation, (b) constr-1 simulation, which does not show correlations in between -1.25 and +1.25, (c) differences of constr-1 and apo simulation, which does not show correlations in between -1.0 and +1.0. The green and cyan circles represent the residues nearby the substrate (F6P) and cofactor (ATP), respectively.

Figure 3.18(a) shows the sum of the apo simulations, the sum of the constr-1 simulations, and the correlation maps of the differences of these sums. According to these maps, regions with negative correlation in apo state became more negative in constr-1 and similarly regions with positive correlation became more positive. Also, regions that changed from neutral to positive correlation and from neutral to negative correlation were observed. In Figure 3.18(b), inter-regional correlation changes were given. According to these changes, three regions (A1/B1, A1/C1, B1/D1) that showed negative correlation in apo simulations became more negative in constr-1 simulations. Two regions (A2/B1, A1D1) that showed positive correlation in apo simulations became more positive in constr-1 simulations. Three regions (A1/A2, B2/C1, B1/C2) that showed neutral correlation in apo simulations became negatively correlated in constr-1 simulations. In the apo simulations, 5 regions (A2/B2, A2/C1, A1/C2, B1/C1, B1/D2), which showed neutral correlation, became positively correlated. In addition, a single region (A1/D2) was observed, which became correlated in apo simulations and became neutral in constr-1

simulations. This means that the amount of correlations increases rather than decreasing, both negatively and positively. In the phosphofructokinase enzyme, the Chain A and B form the AB dimer, and the Chain C and D form the CD dimer. These dimers come together to form the tetramer structure. Correlations between dimers were observed by detailed analysis of the correlation maps given in Figure 3.18(a). In Figure 3.18(c), correlations between dimers in the PFK enzyme were shown both from top (blue) and bottom (yellow). Correlation between dimers was observed in both the upper and lower portions of the protein in the constr-1 simulations.







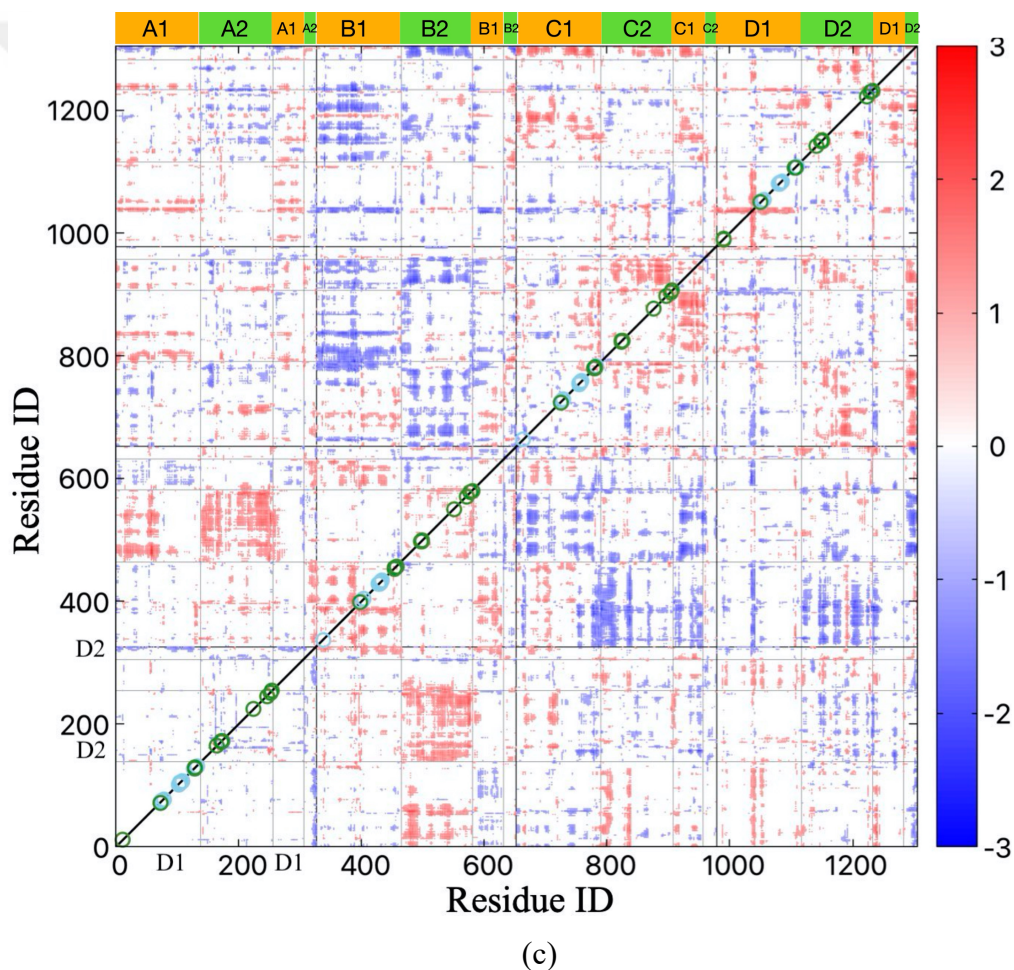
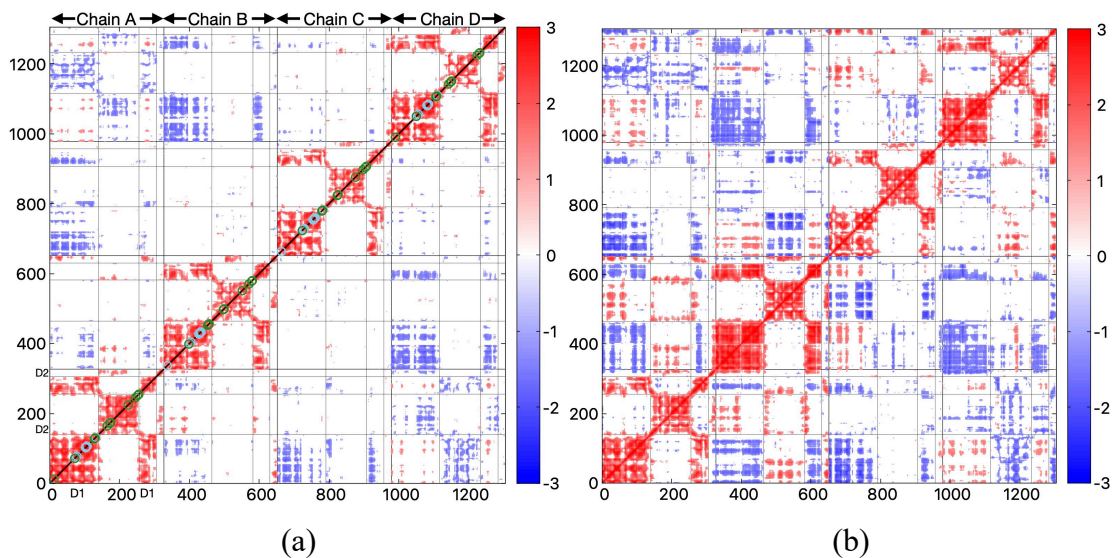
**Figure 3.18:** Evaluation chart showing the biggest changes of total correlation differences shown in Figure 3.17c for the constr-1 simulations. (a) Inter-chain correlation maps, (b) correlation table with significant changes, (c) showing the correlations between domains from the top and bottom of phosphofructokinase with arrows. Values between -1 and 1 are not shown on the maps.

In Figure 3.19(a), as with the correlation changes observed in constr-1 simulations, no correlation was observed between domain-1 and domain-2 in the same chain in apo simulations, while both negative and positive changes were observed in constr-2 simulations. However, it is clearly seen that this change is more especially in constr-1 simulations. Again, as observed in the constr-1 simulations, the regions with 0 or a certain correlation in the apo simulations were observed to be correlated both positively and negatively in the constr-2 simulations (see Figure 3.19(b)). While the negative correlation between A1 and B1 became more negative, the positive correlation between A1 and B2 became more positive in constr-2 simulations. The significant increase in negative correlation observed between Constr-1 A1 and C1, and B1 and D1 regions was also observed in constr-2 simulations. However, it was clearly seen that this increase is much higher in constr-1.

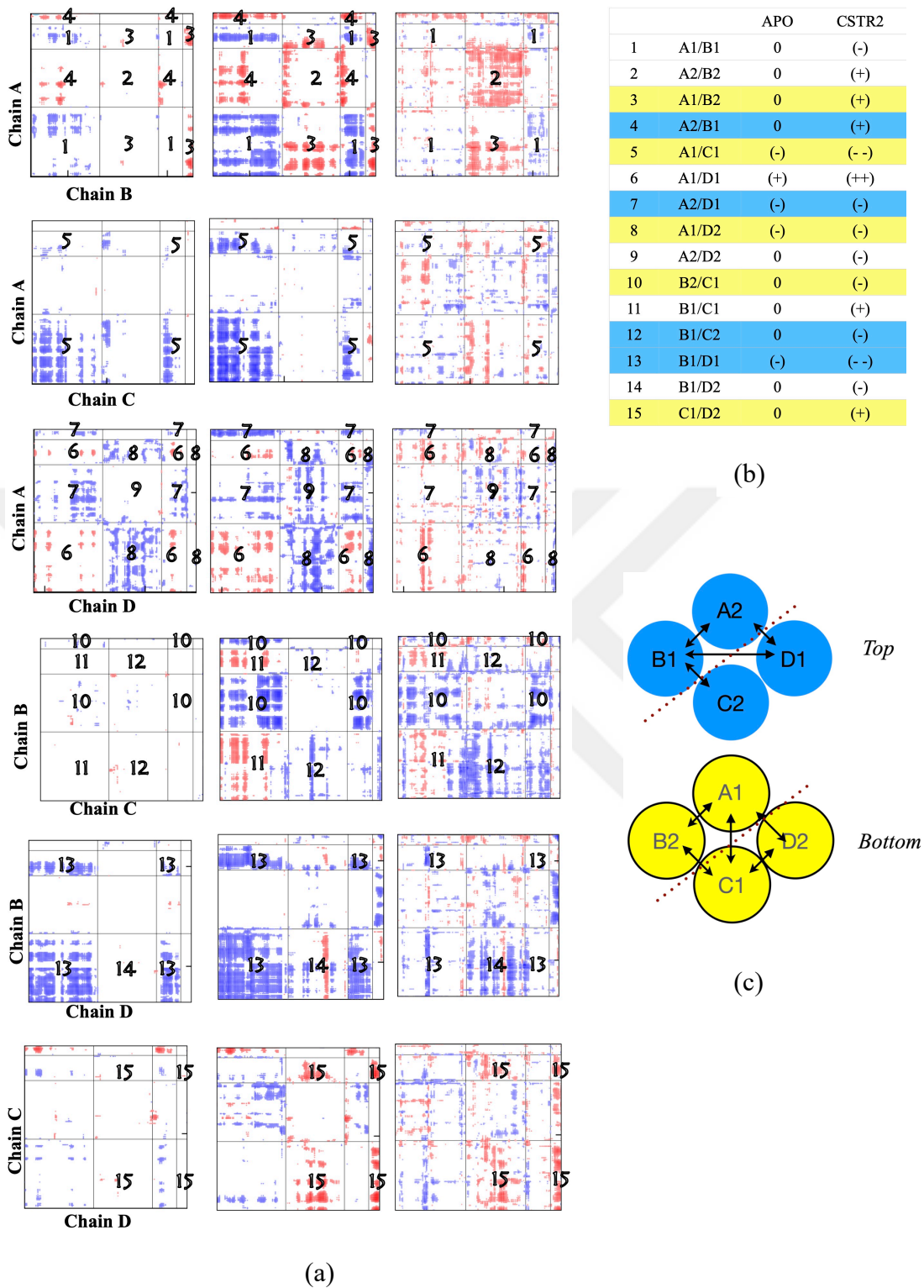
In addition to the inter-domain correlations in different chains, significant changes were observed in the inter-domain correlations in the same chain. While the correlation between domains in the same chain was zero in apo simulations, both negative and positive correlations were observed in constr-2 simulations. The decrease in correlation observed in the Chain A in the constr-1 simulation (see Figure 3.17(c)) was also observed in the constr-2 simulation. However, it was observed in Figure 3.19(c) that this decrease is less in constr-2 simulations. In addition, positive correlations observed in apo simulations along the diagonal line increased in constr-2 simulations. However, it was observed that the negative correlation increases seen in the diagonal line in the constr-1 simulations were less in the constr-2 simulations. It was observed that the negative correlation increases seen in the diagonal line in the constr-1 simulations were less in the constr-2 simulations. Shortly, in both constrained simulations, there are significant increases in both negative and positive correlations between chain and domains compared to apo simulations. However, it was clear that these increases are higher in constr-1 simulations. These changes in correlations indicates the performed restrictions have affected the dynamics of the protein.

In order to provide a better understanding of the correlation differences, the difference maps given in Figure 3.19(c) was divided into regions and investigated. Figure 3.20(a) shows the sum of the apo simulations, the sum of the constr-2 simulations, and the correlation maps of the differences of these sums. According to these maps, it was observed that some regions that are in correlation have increased further in the same direction. In addition, the regions changed from neutral to positive correlation and from neutral to negative correlation were observed. In Figure 3.20(b), inter-regional correlation changes were given. According to these changes, two regions (A1/C1, B1/D1) that showed negative correlation in apo simulations became more negative in constr-2 simulations. One region (A1/D1) that showed positive correlation in apo simulations became more positive in constr-2 simulations. Five regions (A1/B1, A2/D2, B2/C1, B1/C2, B1/D2) that showed neutral correlation in apo simulations became negatively correlated in constr-2 simulations. In the apo simulations, five regions (A2/B2, A1/B2, A2/B1, B1/C1, C1/D2), which showed neutral correlation, became positively correlated. In addition, it was observed that the regions that were in correlation in the apo simulations were also in correlation in the constr-2 simulations. It means that the amount of correlations increases rather than decreasing, both negatively and positively.

Correlations between AB and CD dimers were investigated by detailed analysis of the correlation maps given in Figure 3.20(a). In Figure 3.20(c), correlations between dimers and chains in the PFK enzyme were shown both from top (blue) and bottom (yellow). Here, inter-chain correlation in addition to inter-dimer correlation was observed at the bottom of the protein in constr-2 simulations. However, in both restricted simulations, correlation between dimers was observed at the top and bottom of the protein rather than between monomers in the same dimer.



**Figure 3.19:** All three runs of apo and constr-2 simulations were summed up and their difference was taken. PCA Maps of (a) apo simulation, (b) constr-2 simulation, which does not show correlations in between  $-1.25$  and  $+1.25$ , (c) differences of constr-2 and apo simulation, which does not show correlations in between  $-1.0$  and  $+1.0$ . The green and cyan circles represent the residues nearby the substrate (F6P) and cofactor (ATP), respectively.



**Figure 3.20:** Evaluation chart showing the biggest changes of total correlation differences shown in Figure 3.19c for constr-2. (a) Inter-chain correlation maps, (b) correlation table with significant changes, (c) showing the correlations between domains from the top and bottom of phosphofructokinase with arrows. Values between -1 and 1 are not shown on the maps.

As a result of this detailed analysis, it was observed that the correlation between domains and, therefore, between chains increased significantly due to the restriction of phosphofructokinase. It was observed that this increase especially was much higher in constr-1 simulations. The drawings showing the phosphofructokinase from top and bottom, especially the amount and degree of correlation between A/B and C/D dimers increased (see Figure 3.18(c), and 3.20(c)). Phosphofructokinase is inactive in its dimer state and can only be activated when tetramer formation occurs. The increase in correlations, together with minor increases in RMSF values, may trigger the separation of dimers from each other.

To display the directions of the eigenvectors in apo form and the restricted simulations, VMD software Normal Mode Wizard (Humphrey et al., 1996) plugin was used. Here, only the first principle modes that contributed the highest to explaining the protein dynamics were used. As illustrated in Figure 3.21, in the analysis of the first principle mode, which explains about a quarter of the entire motion in apo simulations, it was observed that the chains either rotate around their own axes or move together in the same direction. On the other hand, in constrained simulations, it was observed that the chains in one dimer were more inclined to move in the opposite directions with the chains in the other dimer, as if they were ready to separate. This might indicate that the global dynamics in constrained simulations may trigger the protein to dissociate from its tetrameric state into its dimer form.

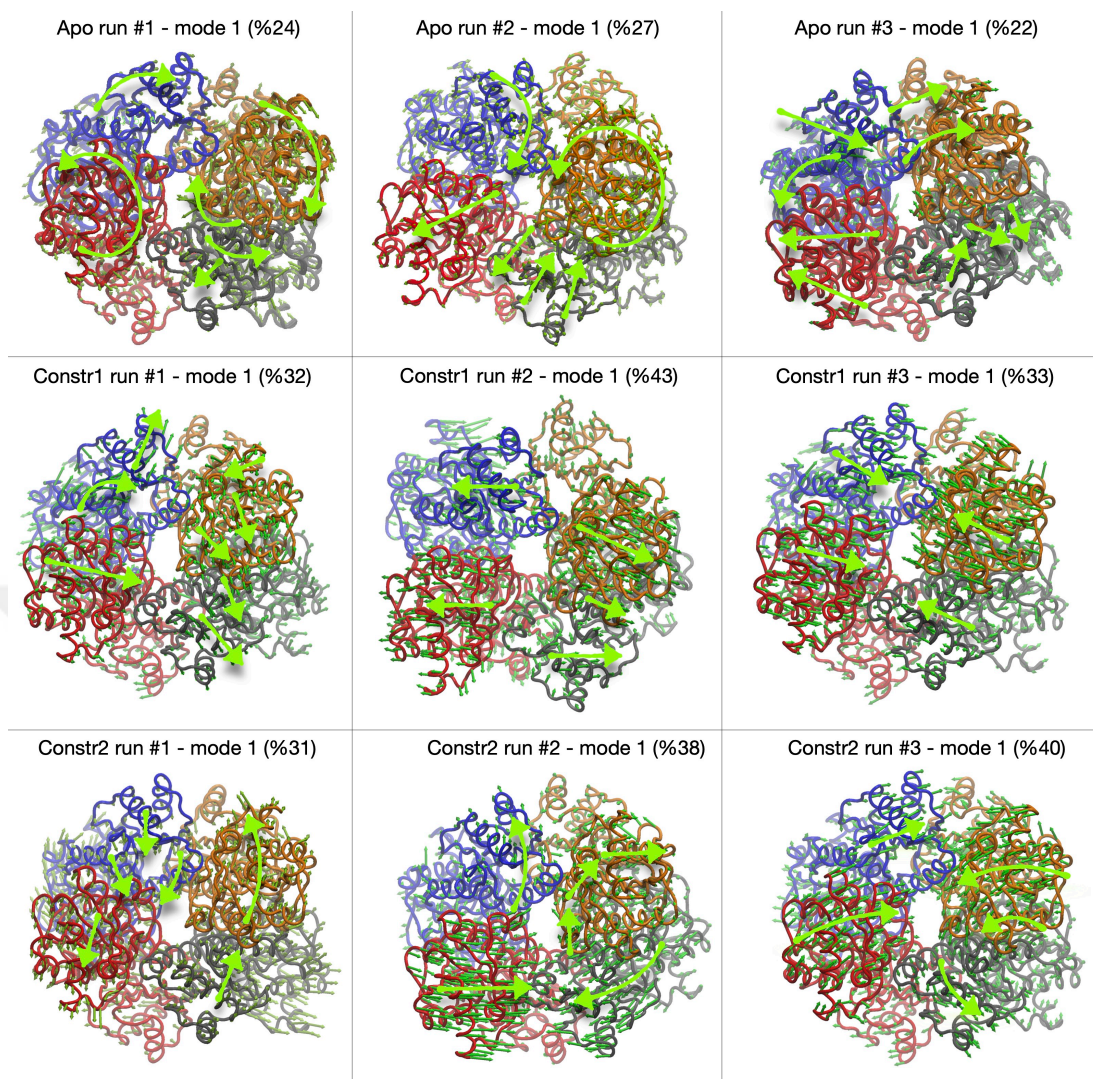


Figure 3.21: Normal mode analysis of each run of three state simulations according to first principle modes. The normal mode analysis results of apo, constr-1, and constr-2 simulations were given in the first, second, and third row, respectively. For each simulation, the percentage of the explained motion in the first principle mode was shown in parentheses.

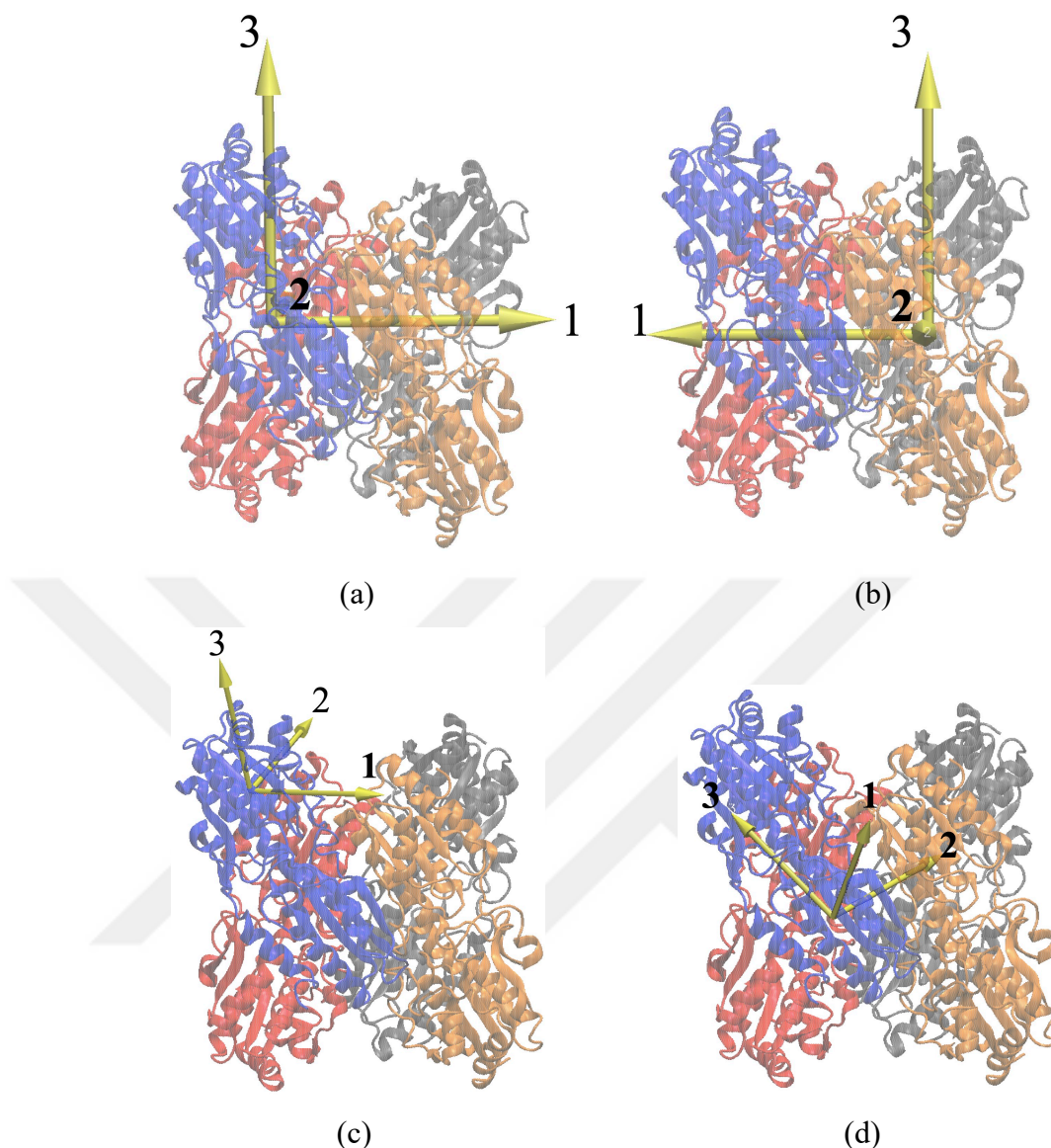
### 3.2.4 Determination of Three Orthogonal Principal Axes for Chains and Domains in PFK

Three orthogonal principal axes of rotation were determined for each chain and domain in PFK using *orient* script tool of VMD (Humphrey et al., 1996). The principal axes represent the non-zero inertia tensor elements. This analysis was performed for all frames in a trajectory file. Thus, variations in the angles which represent chain and domain motions can be observed throughout the entire trajectory. As shown in Figure 3.22, three axes of symmetry were illustrated for two dimers (A/B and C/D), and two domains in a chain. In this study, the orientations of each dimer with respect to each other was determined by calculating the angle between each component of the principal axes which is equal to the dot product of two axes divided by their magnitudes (Equation 3.5).

$$\cos(\theta) = \frac{(\vec{x}_1 \cdot \vec{x}_2)}{|\vec{x}_1| |\vec{x}_2|} \quad (3.5)$$

where  $\mathbf{x}_1$  is the first principal axis of one dimer and  $\mathbf{x}_2$  is the first principal axis of the second dimer. The PFK enzyme consists of two domains in each chain, as domain-1, and domain-2. Domain-1 in all chains contains a cofactor (ATP) binding site. Domain -2 in all chains contains a substrate (F6P) binding site, and F6P binding plays a critical role in dimer-tetramer conversion of PFK. Therefore, in the principal axis calculations, it was important to investigate the variations in the angles of the domains by considering the domains in each chain within themselves. Here, the orientations of domain-1 and domain-2 in each chain with respect to each other were also determined by the Equation 3.5, where  $x_1$  is the first principal axis of domain-1 in a chain and  $x_2$  is the first principal axis of domain-2 in the same chain. Then, the angle was calculated by taking arccosine of  $\cos(\theta)$ .

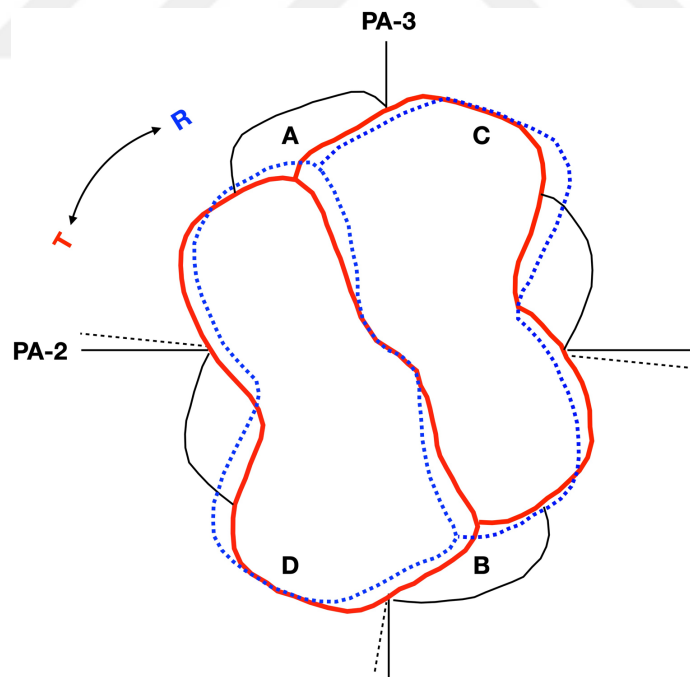




**Figure 3.22:** Principal axes directions of PFK aligned to (a) A/B dimer, and (b) C/D dimer, (c) domain-1, and (d) domain-2 in Chain A. The snapshots were taken from initial frame of apo run-1 trajectory.

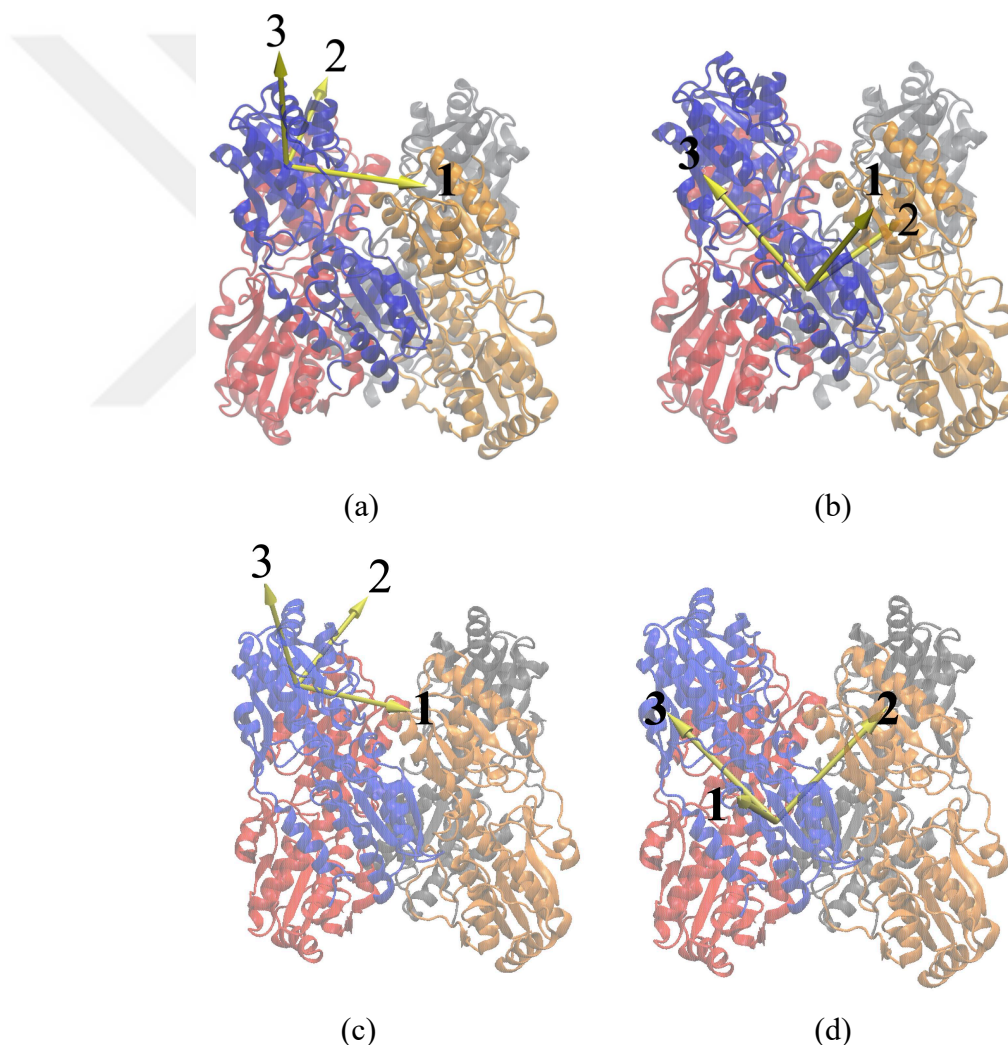
In PFK enzyme, A and B chains form A/B dimers, C and D chains form C/D dimers, and these two dimers form the active tetramer structure. The third principal axes of A/B and C/D dimers in tetramer structure makes  $7^{\circ}$ - $8^{\circ}$  relative to each other in T-state (inactive state), and as the tetramer structure converts to the R-state (active state), this angle decreases to  $0^{\circ}$  (Schirmer & Evans, 1990). Therefore, investigating the variations in the angles between principal axes of dimers was important to reveal the effects of restraints on the orientations of dimers. In Figure 3.23, the T- and R- state of PFK were illustrated. In Figure 3.24(a), variations in the angles in principal axes-2 (PA2) and principal axes-3 (PA3) directions of dimers were observed in both apo simulations and constrained

simulations. However, these variations in angles were higher in apo simulations than those in constrained simulations. Indeed, these variations in the angles were almost fixed at  $7^{\circ}$ - $8^{\circ}$  especially in all three runs of constr-1 simulations. In apo simulations, angle values varied from  $7^{\circ}$ - $8^{\circ}$  up to  $20^{\circ}$ . On the other hand, in constr-2 simulations, it was observed that the angles in PA2 and PA3 directions decreased to almost  $0^{\circ}$  in two runs, indicating that the A/B and C/D dimers stood parallel to each other and switched to the R-state. Because when the PFK is in R-state (active state), the angle between the dimers in PA3 direction is  $0^{\circ}$  (Schirmer & Evans, 1990). Moreover, less variation in angle between dimers was observed as a result of restraints on protein dynamics. In addition, it was observed that the angle variations in the PA1 direction had the least change in both apo and constrained simulations, because the dimers rotated mostly around PA1 axis. This is why both PA2 and PA3 directions were changing more than PA1 directions (see Figure 3.23). The initial and last frames of the trajectory along with their principal axes of domains in Chain A were illustrated in Figure 3.24. While the first domain preserved its initial orientation (See Figs 3.24a and 3.24c), the second dimer rotated around its third principal axes by about  $60^{\circ}$  (See Figs 3.24b and 3.24d).

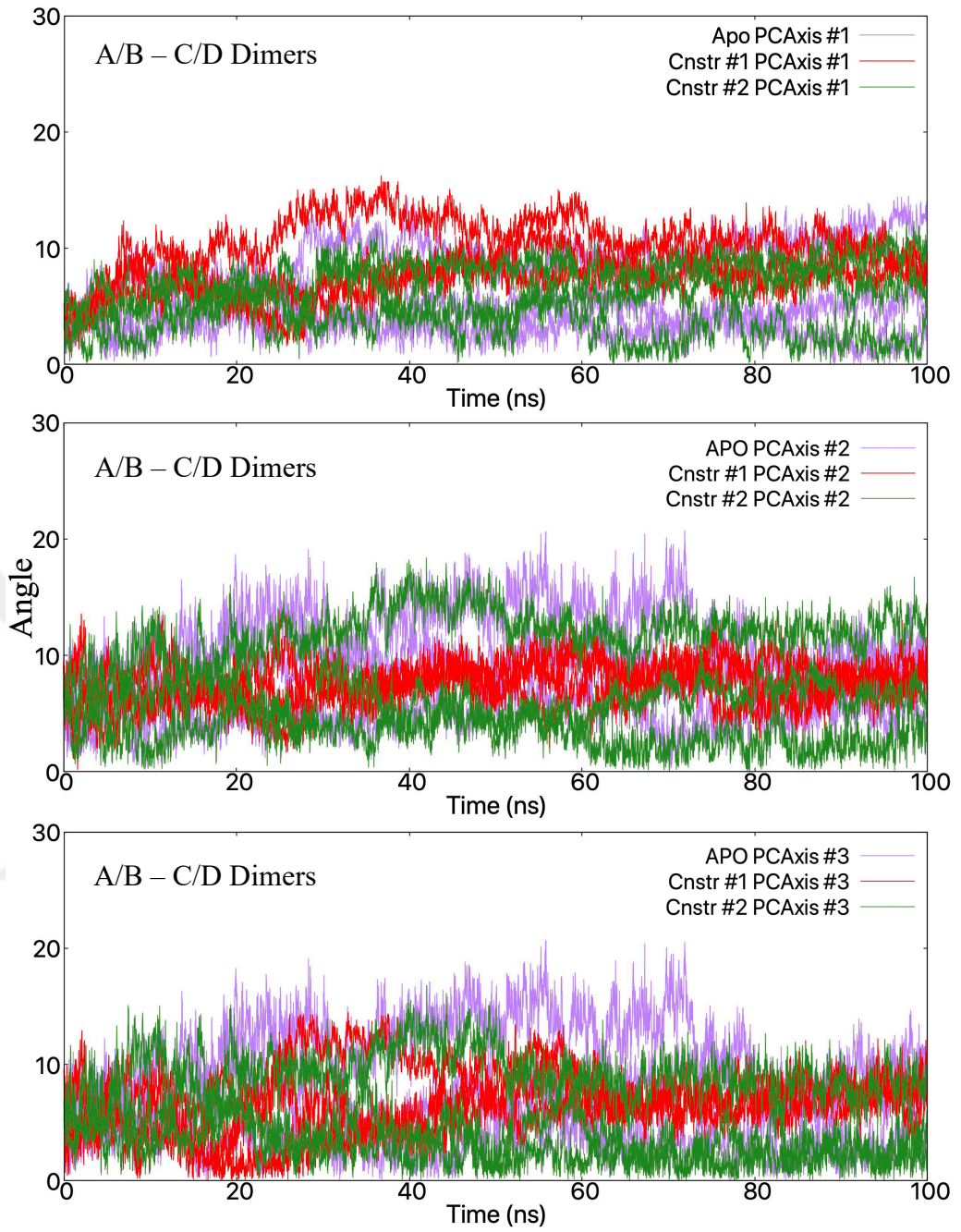


**Figure 3.23:** T-and R-state of PFK (created by the author Çelebi with Keynote tool).

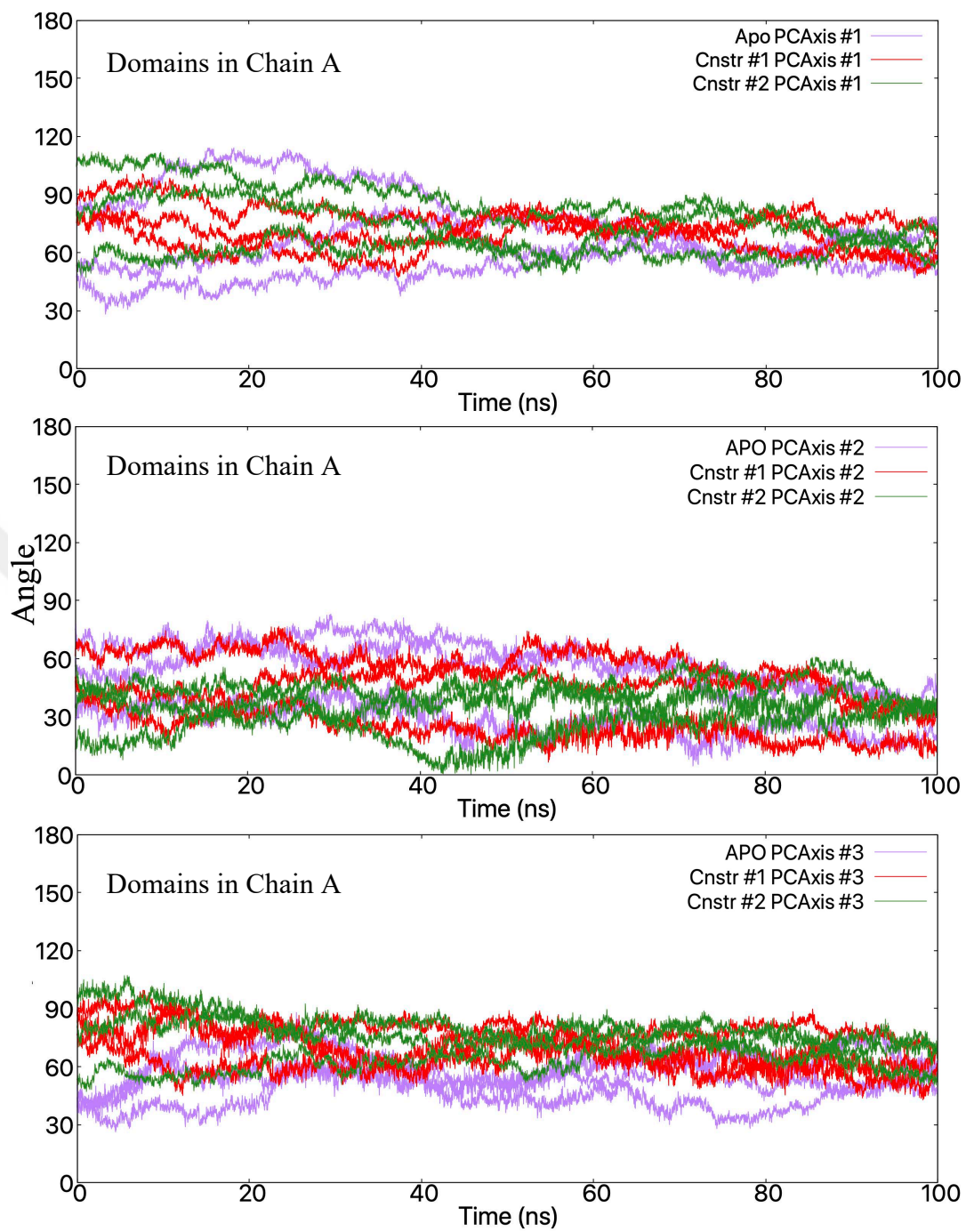
In Figure 3.25(b), (c), (d), and (e) the variations in the angles between the principal axes of domains in each chain were illustrated. Accordingly, it was observed that the variations in the angles between principal axes of two domains in chain A were almost the same in both apo simulations and constrained simulations. In general, it was observed that the angle variation changes in the PA1 and PA3 directions between domain-1 and domain-2 in all chains converge at  $60^\circ$  starting from  $20^\circ$ -  $110^\circ$  for PA1 direction,  $30^\circ$ -  $100^\circ$  for PA3 direction. It has been observed that the angle variation changes in the PA2 direction converge to  $30^\circ$ . In addition, it was observed that although the angle values started from different values at the beginning of the simulations, they converged to the same values in the common direction at the end of the simulation.



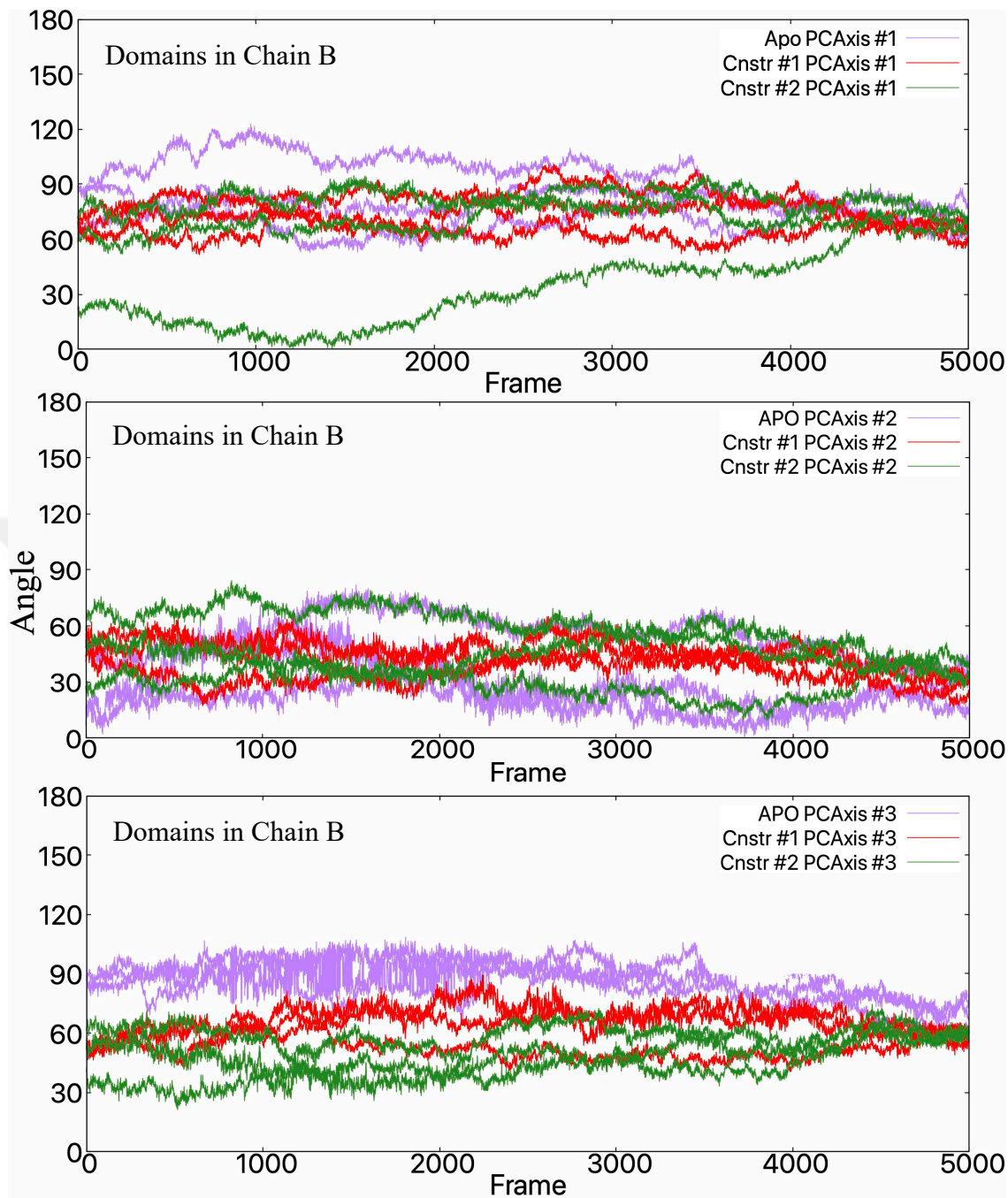
**Figure 3.24:** Principal axes directions of PFK aligned to (a) initial frame of domain-1 (b) initial frame of domain-2, (c) last frame of domain-1, and (d) last frame of domain-2 in Chain A. All snapshots were taken from apo run-1 trajectory.



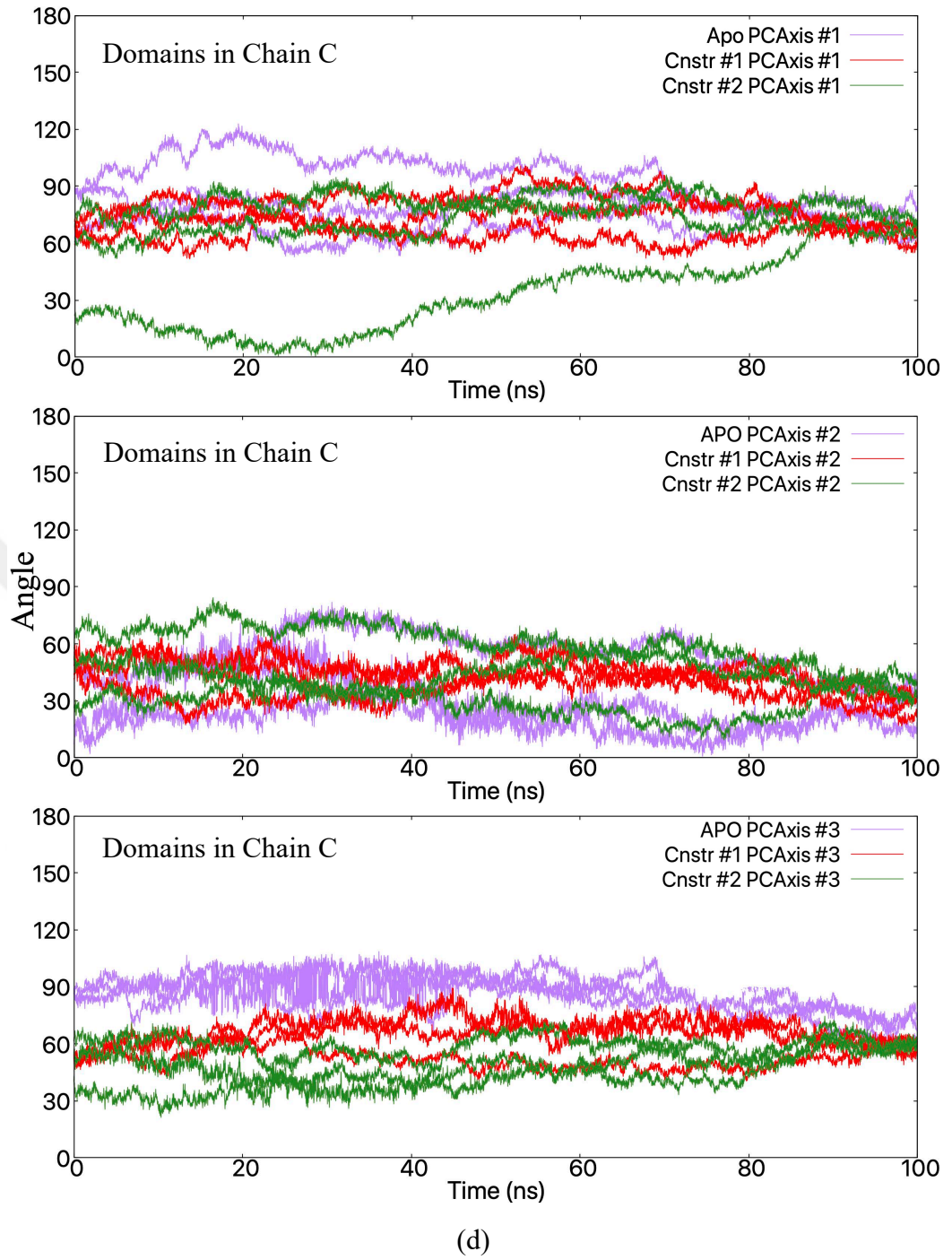
(a)

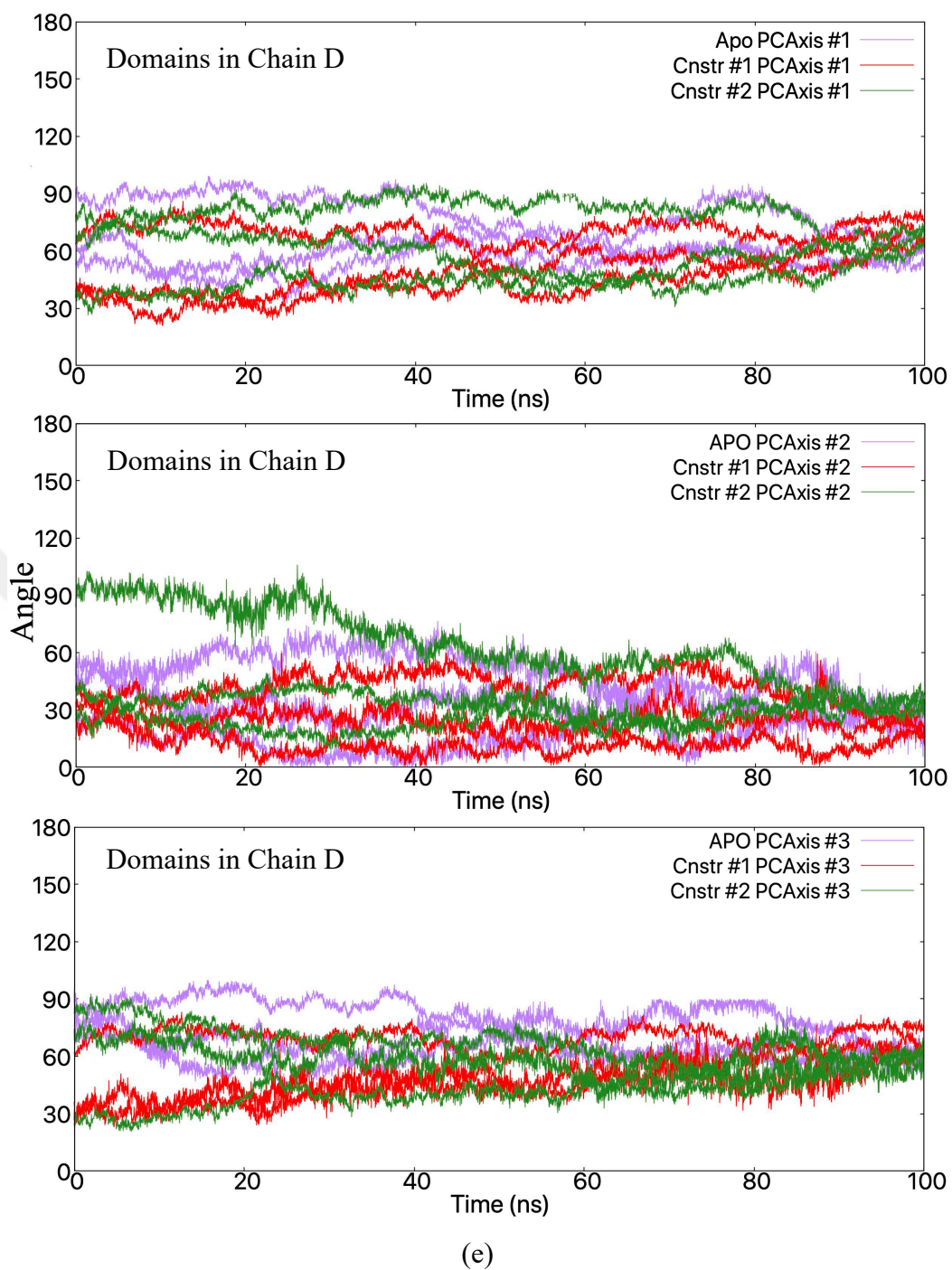


(b)



(c)





**Figure 3.25:** The variations in angles of (a) A/B – C/D dimers (b) domains in Chain A (c) domains in Chain B (d) domains in Chain C, and (e) domains Chain D.



### 3.2.5 The Mean Square Distance Fluctuation Analysis

The A/B and C/D dimers of the phosphofructokinase enzyme come together to form the active tetramer structure where interaction between A/B and C/D dimers plays a role in keeping the protein together in R (active) and T (inactive) states (Schirmer & Evans, 1990). In this study, the mean square distance fluctuations were calculated to reveal which regions of the PFK are in communication in the dimer-tetramer conversion. The mean square distance fluctuation is a computational method used to reveal residues with high communication potential in a protein. In this method, signal transmission potentials can be determined according to the numerical values of the distance fluctuations between the residues (Chennubhotla & Bahar, 2007). Here, the distance fluctuation is the commute-time, and the distance fluctuations ( $D_{ij}$ ) between each amino acids  $i$  and  $j$  in the system were calculated throughout the trajectory with the following Equation 3.6.

$$D_{ij} = \langle (d_{ij} - \langle d_{ij} \rangle)^2 \rangle^{1/2} \quad (3.6)$$

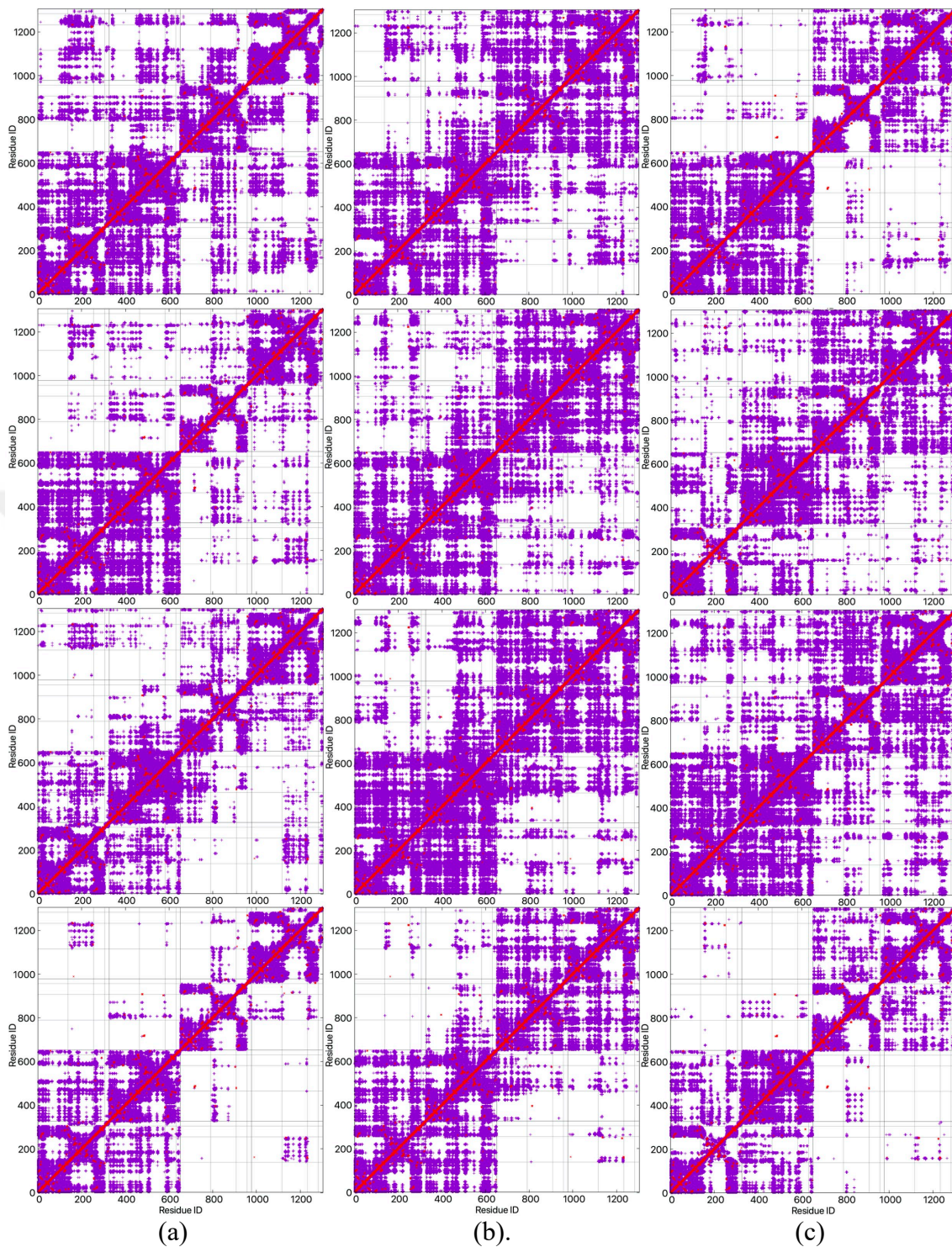
Where  $d_{ij}$  is the distance between  $i$  and  $j$  amino acids  $C_\alpha$  atoms, and the brackets represent the time-averaged over the entire trajectory file. According to  $D_{ij}$  value, the signal transmission potential between the two residues can be determined. A low  $D_{ij}$  value between two residues means low 'commute-time' and indicates high signal exchange potential between each other. In this way, the effect of restrictions on protein dynamics and communication in between residues were investigated Also, the contact map was calculated from the initial frame of the trajectory with the following Equation 3.7.

$$C_{ij} = \begin{cases} 1, & \text{if } \delta_{ij} < R_c \\ 0, & \text{Otherwise} \end{cases} \quad (3.7)$$

Where  $R_c$  is the threshold distance for  $C_\alpha$  atoms of residues which was taken as 6 Å to be considered in contact. The data that were obtained from the mean square distance fluctuation calculations of three runs of three simulation states and the contact map of the simulations were illustrated together in Figure 3.26. The aim here was to investigate the regions that have low commute-time even when they are far from each other. As shown in Figure 3.26, the distance fluctuation maps and contact map of apo simulations (first column), constr-1 (second column), and constr-2 (third column) were illustrated together

to reveal the regions that have low commute-time, and all maps are showing values between 0 and 0.5 Å. According to Figure 3.26, a decrease in the amount of distance fluctuation between two dimers, A/B and C/D, was observed in constr-1 runs. But this decrease was mostly restricted to chains B and C only. In addition, significant changes in distance fluctuations between chains were also observed; in apo runs, the distance fluctuation between chains C and D was high, in other words, the communication was low, however a significant decrease in distance fluctuation was observed in the same regions in both constrained simulations, indicating that the signal transmission between chains C and D increased. And it was clearly observed that this increase was more significant in constr-1 than in constr-2.

Also, in apo runs, a low distance fluctuation value was observed between A2 and D2 domains which increased in both constrained simulations, in other words, the communication between these domains almost disappeared. As a result, each dimer started to communicate more within itself when switched to a constrained state. Indeed, it is clearly seen that this increased communication between monomers of the dimers was significant especially in constr-1 simulations. This may also indicate the state of two dimers that break up their communication just before they are about to separate.



**Figure 3.26:** The distance fluctuation maps of (a) apo, (b) constr-1, and (c) constr-2 simulations. All maps are showing values between 0 and 0.5 Å. The red points represent the contact map.

## 4. CONCLUSION AND FUTURE WORK

In previous studies of our research group, allosteric sites were discovered and proposed for use in species-specific drug design studies for allosteric enzymes in the glycolytic pathway of bacteria, parasites, and human species. In this thesis, the allosteric effects of proposed allosteric sites were further investigated by two different methods. The first method was AlloSigMA which is based on low frequency normal modes, and the second method was Molecular Dynamics simulations.

In the first part of this thesis, allosteric effects of the proposed sites for PFK, GADPH, and PK enzymes belonging to *S. aureus* and parasites *T. brucei*, *T. cruzi* and *L. mexicana* were investigated with AlloSigMA prediction tool. In addition to the proposed sites, the allosteric effects of the known allosteric sites experimentally reported in the literature for *S. aureus* PFK, and *S. aureus* PK enzymes were also studied with this approach and compared with the proposed sites. The result of the experiment on the proposed allosteric site for *S. aureus* PFK showed a significant decrease in protein dynamics. In contrast, the result of our experiment on the known allosteric site showed that it had almost no effect on protein dynamics. It was observed that the top druggable allosteric site in the *T. brucei* PFK experiment, which had the same location as *S. aureus* PFK, caused an increase in protein dynamics. Additionally, the proposed allosteric sites were perturbed which were found in between catalytic sites at the interface of *T. brucei* PFK. Then, a significant decrease in protein dynamics were observed. In the experiment conducted for human PFK, it was observed that perturbation of the proposed allosteric site did not restrict the protein dynamics. This indicates that in the case of drug binding to the proposed allosteric sites, disease-causing organisms would be inhibited, while there will no inhibition in human PFK.

For GADPH enzyme, the proposed allosteric sites found in the tunnel region of GADPH were used as perturbation sites for all three species, *S. aureus*, *T. cruzi*, and *H. sapiens*. As a result of these experiments, it was observed that the most effective allosteric inhibition occurred in all catalytic regions of *T. cruzi* GADPH enzyme. In contrast, partial allosteric inhibition effect was observed in *S. aureus* (Chains A and B) and *H. sapiens* (Chains B and D) GADPH enzymes.

Similar to *S. aureus* PFK, the allosteric inhibition capacities of both the proposed allosteric site and the known allosteric site experimentally reported in the literature for the bacterial PK enzyme were investigated. As a result of these experiments, it has been observed that the perturbation of the proposed allosteric region decreased the dynamics of the catalytic regions in all chains while perturbing the known allosteric region led to an increase in the dynamics of the catalytic regions. For *L. mexicana* parasitic PK, it was observed that allosteric inhibition occurred in only three catalytic sites in the perturbation of the allosteric region located far from the central opening. Similar to *L. mexicana* parasitic PK, allosteric inhibition occurred in only two catalytic sites in the perturbation of proposed allosteric region for the *H. sapiens* PK enzyme located at the A-domain interface of the central region.

In the second part of this thesis, the effect of restrictions imposed on allosteric sites on protein dynamics was investigated by Molecular Dynamics Simulations. PFK systems were created to mimic the drug binding effect by applying bond restraints in between the residues in the proposed and known allosteric sites for *S. aureus* PFK enzyme. The purpose of applying bond restraints instead of incorporating bound ligand molecules is due to low accuracy encountered in ligand parameterizations. To observe the allosteric effects on protein dynamics both apo form, bond restrained proposed allosteric site and known allosteric site molecular dynamics simulations were performed. For each simulation state, three run simulations were performed. The aim here was to observe whether the bond restraints applied to the residues in the allosteric sites had a similar effect in each run. Then, to see the differences between apo and restrained simulations various computational methods for both apo and restrained MD simulation trajectory files were performed. RMSD values were checked to observe whether the system reached equilibrium and it was observed that the system reached equilibrium. RMSF calculations

were also performed to reveal mobile regions in the protein. Accordingly, the differences between the RMSF values of apo, and constrained simulations is nearly the same which indicated that the residues of the PFK were less fluctuated from the average positions. But a sharp decrease in RMSF values for domain-1 region of chain D in both constrained simulations were observed. This corresponds to a loop region in apo form which later transformed into a helical state in constrained simulations. PCA analysis was applied to reveal more clearly and dramatically the effect of constraints on protein dynamics. Here, the correlation of positional fluctuations between each pair of residues in the chains and domains in apo and constrained simulations were investigated. Here, increase in both positive and negative correlations was observed in restricted simulations between regions that were not correlated with each other or were slightly correlated in apo simulations. In addition, a significant increase in correlation was observed in restricted simulations in regions that were in correlation with each other in apo simulations. Also, while there was no correlation between domains in the same chain in apo simulations, both negative and positive correlations were observed between domains in the same chain in restricted simulations. As a result of PCA analysis, it was observed that the amount of correlation between domains and, therefore, between chains increased significantly in both positive and negative directions as a result of the restriction imposed on PFK. It was observed that this increase especially was much higher in constr-1 runs which represented the restricted state on proposed sites.

Then, the directions of the eigenvectors for the first low-frequency normal modes in apo and restricted simulations were investigated. Accordingly, in apo simulations, it was observed that the chains either rotated around their own axes or moved together in the same direction. On the other hand, in constrained simulations, it was observed that the chains in one dimer were more inclined to move in the opposite directions with the chains in the other dimer, as if they were ready to separate.

To investigate the variance in angles of dimers and domains, three orthogonal principal axes of rotation were determined. In this analysis, while variations in angles between dimers and domains were observed in apo simulations, it was observed that this variance in angles is less in restricted simulations. As a final analysis, the mean square distance

fluctuation between each pair of residues in the system was calculated for apo and constrained simulations. In this analysis, it was aimed to investigate the regions that have low commute-time even when they are far from each other. In this analysis, a significant increase in communication was observed in restricted simulations in regions where there is no or very low inter-chain communication in apo simulations. As a result of this analysis, it was observed that each dimer of PFK in apo simulations started to communicate more within itself when switched to a constrained state. Indeed, it was clearly observed that this increased communication between monomers of the dimers was significant especially in proposed site restricted simulations. To sum up, it has been revealed that by the analysis of nine independent simulations, restrictions on both the proposed and known allosteric region affected the protein dynamics.

In this study, the allosteric effects of known allosteric regions in the literature, and the allosteric regions proposed by our research group in previous studies were investigated by short MD simulations. Despite its short duration, it was possible to observe the effect of imposing restraints on the protein's global dynamics. However, the simulations with 100 ns long were not sufficient, and its extension in future studies would be beneficial in order to observe a more noticeable allosteric effect.

## REFERENCES

- Abel, S., Jenkins, T. M., Whitlock, L. A., Ridgway, C. E., & Muirhead, G. J. (2008). Effects of CYP3A4 inducers with and without CYP3A4 inhibitors on the pharmacokinetics of maraviroc in healthy volunteers. *British Journal of Clinical Pharmacology*, *65 Suppl 1*, 38–46. <https://doi.org/10.1111/j.1365-2125.2008.03134.x>
- Amadei, A., Linssen, A. B. M., & Berendsen, H. J. C. (1993). Essential dynamics of proteins. *Proteins: Structure, Function, and Genetics*, *17*(4), 412–425. <https://doi.org/10.1002/prot.340170408>
- Axerio-Cilies, P., See, R. H., Zoraghi, R., Worrall, L., Lian, T., Stoyanov, N., Jiang, J., Kaur, S., Jackson, L., Gong, H., Swayze, R., Amandoron, E., Kumar, N. S., Moreau, A., Hsing, M., Strynadka, N. C., McMaster, W. R., Finlay, B. B., Foster, L. J., ... Cherkasov, A. (2012). Cheminformatics-Driven Discovery of Selective, Nanomolar Inhibitors for Staphylococcal Pyruvate Kinase. *ACS Chemical Biology*, *7*(2), 350–359. <https://doi.org/10.1021/cb2003576>
- Ayyildiz, M., Celiker, S., Ozhelvaci, F., & Akten, E. D. (2020). Identification of Alternative Allosteric Sites in Glycolytic Enzymes for Potential Use as Species-Specific Drug Targets. *Frontiers in Molecular Biosciences*, *7*. <https://doi.org/10.3389/fmolb.2020.00088>
- Bakan, A., Meireles, L. M., & Bahar, I. (2011). ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics*, *27*(11), 1575–1577. <https://doi.org/10.1093/bioinformatics/btr168>
- Barnett, J. A. (2003). A history of research on yeasts 5: The fermentation pathway. *Yeast*, *20*(6), 509–543. <https://doi.org/10.1002/yea.986>
- Bohr, C., Hasselbalch, K., & Krogh, A. (1904). Ueber einen in biologischer Beziehung wichtigen Einfluss, den die Kohlensäurespannung des Blutes auf dessen Sauerstoffbindung übt. *Skandinavisches Archiv Für Physiologie*, *16*(2), 402–412. <https://doi.org/10.1111/j.1748-1716.1904.tb01382.x>
- Boscá, L., & Corredor, C. (1984). Is phosphofructokinase the rate-limiting step of glycolysis? *Trends in Biochemical Sciences*, *9*(9), 372–373. [https://doi.org/10.1016/0968-0004\(84\)90214-7](https://doi.org/10.1016/0968-0004(84)90214-7)
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, *4*(2), 187–217. <https://doi.org/10.1002/jcc.540040211>



- Campa, J. S., McAnulty, R. J., & Laurent, G. J. (1990). Application of high-pressure liquid chromatography to studies of collagen production by isolated cells in culture. *Analytical Biochemistry*, 186(2), 257–263. [https://doi.org/10.1016/0003-2697\(90\)90076-L](https://doi.org/10.1016/0003-2697(90)90076-L)
- Çelebi, M., İnan, T., Kürkçüoğlu, Ö., & Akten, E. D. (2021). Potential Allosteric Sites Captured in Glycolytic Enzymes via Residue-Based Network Models: Phosphofructokinase, Glyceraldehyde-3-Phosphate Dehydrogenase and Pyruvate Kinase. *Proteins: Structure, Function, and Bioinformatics*, Submitted (To the Journal), On May 2021.
- Chennubhotla, C., & Bahar, I. (2007). Signal propagation in proteins and relation to equilibrium fluctuations. *PLoS Computational Biology*, 3(9), 1716–1726. <https://doi.org/10.1371/journal.pcbi.0030172>
- Conn, P. J., Christopoulos, A., & Lindsley, C. W. (2009). Allosteric modulators of GPCRs: a novel approach for the Nat Rev Drug Discov . 2009 January ; 8(1): 41–54. doi:10.1038/nrd2760. treatment of CNS disorders. *Nat Rev Drug Discov.*, 8(1), 41–54. <https://doi.org/10.1038/nrd2760>. Allosteric
- Cooper, A., & Dryden, D. T. F. (1984). Allostery without conformational change - A plausible model. *European Biophysics Journal*, 11(2), 103–109. <https://doi.org/10.1007/BF00276625>
- Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., & Kollman, P. A. (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*, 117(19), 5179–5197. <https://doi.org/10.1021/ja00124a002>
- Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M., Pieper, U., & Sali, A. (2006). Comparative Protein Structure Modeling Using Modeller. *Current Protocols in Bioinformatics*, 15(1). <https://doi.org/10.1002/0471250953.bi0506s15>
- García, A. E. (1992). Large-amplitude nonlinear motions in proteins. *Physical Review Letters*, 68(17), 2696–2699. <https://doi.org/10.1103/PhysRevLett.68.2696>
- Guarnera, E., & Berezovsky, I. N. (2016). Structure-Based Statistical Mechanical Model Accounts for the Causality and Energetics of Allosteric Communication. *PLoS Computational Biology*, 12(3), 1–27. <https://doi.org/10.1371/journal.pcbi.1004678>
- Guarnera, E., Tan, Z. W., Zheng, Z., & Berezovsky, I. N. (2017). AlloSigma: Allosteric signaling and mutation analysis server. *Bioinformatics*, 33(24), 3996–

3998. <https://doi.org/10.1093/bioinformatics/btx430>

Gunasekaran, K., Ma, B., & Nussinov, R. (2004). Is allostery an intrinsic property of all dynamic proteins? *Proteins: Structure, Function and Genetics*, 57(3), 433–443. <https://doi.org/10.1002/prot.20232>

Hayward, S., Kitao, A., Hirata, F., & Gō, N. (1993). Effect of Solvent on Collective Motions in Globular Protein. *Journal of Molecular Biology*, 234(4), 1207–1217. <https://doi.org/10.1006/jmbi.1993.1671>

Hayward, Steven, Kitao, A., & Gō, N. (1995). Harmonicity and anharmonicity in protein dynamics: A normal mode analysis and principal component analysis. *Proteins: Structure, Function, and Bioinformatics*, 23(2), 177–186. <https://doi.org/10.1002/prot.340230207>

Henry, B. (2004). GPCR allostery and accessory proteins: new insights into drug discovery, 17 July 2004, Glasgow, Scotland. *IDrugs : The Investigational Drugs Journal*, 7(9), 819–821. <http://www.ncbi.nlm.nih.gov/pubmed/15470594>

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., Grubmüller, H., & MacKerell, A. D. (2017). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods*, 14(1), 71–73. <https://doi.org/10.1038/nmeth.4067>

Humphrey, W., Dalke, A., & Schulten, K. (1996). Sartorius products. *Journal of Molecular Graphics*, 14(October 1995), 33–38. <https://www.tapbiosystems.com/tap/products/index.htm>

Ichiye, T., & Karplus, M. (1991). Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins: Structure, Function, and Genetics*, 11(3), 205–217. <https://doi.org/10.1002/prot.340110305>

Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3(3), 318–356. [https://doi.org/10.1016/S0022-2836\(61\)80072-7](https://doi.org/10.1016/S0022-2836(61)80072-7)

Karplus, M., & Kushick, J. N. (1981). Method for estimating the configurational entropy of macromolecules. *Macromolecules*, 14(2), 325–332. <https://doi.org/10.1021/ma50003a019>

Kitao, A., Hirata, F., & Gō, N. (1991). The effects of solvent on the conformation and the collective motions of protein: Normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. *Chemical Physics*, 158(2–3), 447–472. [https://doi.org/10.1016/0301-0104\(91\)87082-7](https://doi.org/10.1016/0301-0104(91)87082-7)

- Koshland, D. E., Nemethy, J. G., & Filmer, D. (1966). Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry*, 5(1), 365–385. <https://doi.org/10.1021/bi00865a047>
- Koshland, Daniel E., & Hamadani, K. (2002). Proteomics and models for enzyme cooperativity. *Journal of Biological Chemistry*, 277(49), 46841–46844. <https://doi.org/10.1074/jbc.R200014200>
- Li, T., Tan, X., Yang, R., Miao, Y., Zhang, M., Xi, Y., Guo, R., Zheng, M., & Li, B. (2020). Discovery of novel glyceraldehyde-3-phosphate dehydrogenase inhibitor via docking-based virtual screening. *Bioorganic Chemistry*, 96, 103620. <https://doi.org/10.1016/j.bioorg.2020.103620>
- Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., & Šali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annual Review of Biophysics and Biomolecular Structure*, 29(February), 291–325. <https://doi.org/10.1146/annurev.biophys.29.1.291>
- May, L. T., & Christopoulos, A. (2003). Allosteric modulators of G-protein-coupled receptors. *Current Opinion in Pharmacology*, 3(5), 551–556. [https://doi.org/10.1016/S1471-4892\(03\)00107-3](https://doi.org/10.1016/S1471-4892(03)00107-3)
- Möhler, H., Fritschy, J. M., & Rudolph, U. (2002). A New Benzodiazepine Pharmacology. *Journal of Pharmacology and Experimental Therapeutics*, 300(1), 2–8. <https://doi.org/10.1124/jpet.300.1.2>
- Monod, Jacques, Wyman, J., & Changeux, J. P. (1965). On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology*, 12(1), 88–118. [https://doi.org/10.1016/S0022-2836\(65\)80285-6](https://doi.org/10.1016/S0022-2836(65)80285-6)
- Monod, Jacques, Changeux, J. P., & Jacob, F. (1963). Allosteric proteins and cellular control systems. *Journal of Molecular Biology*, 6(4), 306–329. [https://doi.org/10.1016/S0022-2836\(63\)80091-1](https://doi.org/10.1016/S0022-2836(63)80091-1)
- Oostenbrink, C., Villa, A., Mark, A. E., & van Gunsteren, W. F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry*, 25(13), 1656–1676. <https://doi.org/10.1002/jcc.20090>
- Perahia, D., Levy, R. M., & Karplus, M. (1990). Motions of an  $\alpha$ -helical polypeptide: Comparison of molecular and harmonic dynamics. *Biopolymers*, 29(4–5), 645–677. <https://doi.org/10.1002/bip.360290402>
- Phillips, J. C., Hardy, D. J., Maia, J. D. C., Stone, J. E., Ribeiro, J. V., Bernardi, R. C., Buch, R., Fiorin, G., Hénin, J., Jiang, W., McGreevy, R., Melo, M. C. R., Radak,

B. K., Skeel, R. D., Singharoy, A., Wang, Y., Roux, B., Aksimentiev, A., Luthey-Schulten, Z., ... Tajkhorshid, E. (2020). Scalable molecular dynamics on CPU and GPU architectures with NAMD. *Journal of Chemical Physics*, 153(4).  
<https://doi.org/10.1063/5.0014475>

Romano, A. H., & Conway, T. (1996). Evolution of carbohydrate metabolic pathways. *Research in Microbiology*, 147(6–7), 448–455. [https://doi.org/10.1016/0923-2508\(96\)83998-2](https://doi.org/10.1016/0923-2508(96)83998-2)

Schirmer, T., & Evans, P. R. (1990). Structural basis of the allosteric behaviour of phosphofructokinase. *Nature*, 343(6254), 140–145.  
<https://doi.org/10.1038/343140a0>

Spurny, R., Ramerstorfer, J., Price, K., Brams, M., Ernst, M., Nury, H., Verheij, M., Legrand, P., Bertrand, D., Bertrand, S., Dougherty, D. A., De Esch, I. J. P., Corringer, P. J., Sieghart, W., Lummis, S. C. R., & Ulens, C. (2012). Pentameric ligand-gated ion channel ELIC is activated by GABA and modulated by benzodiazepines. *Proceedings of the National Academy of Sciences of the United States of America*, 109(44), 3028–3034. <https://doi.org/10.1073/pnas.1208208109>

Tian, T., Wang, C., Wu, M., Zhang, X., & Zang, J. (2018). Structural Insights into the Regulation of Staphylococcus aureus Phosphofructokinase by Tetramer-Dimer Conversion. *Biochemistry*, 57(29), 4252–4262.  
<https://doi.org/10.1021/acs.biochem.8b00028>

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). 20035\_Ftp. *Journal of Computational Chemistry*, 25(12), 1157–1174.

Williams, T., & Kelly, C. (2011). *Gnuplot*.

## CURRICULUM VITAE

### Personal Information:

Name Surname : Metehan ÇELEBİ

### Education:

Undergraduate Education : Inonu University – Molecular Biology and Genetics

Foreign Language Skills : English (Professional Working)

## APPENDIX A

### A.1 Prepared Configuration Files for Apo and Constrained MD Simulations.

```
##                                JOB DESCRIPTION                                ##
# Minimization and Equilibration of
# PFK in a Water Box
##                                ADJUSTABLE PARAMETERS                        ##
structure      ionized.psf
coordinates    ionized.pdb

set temperature 310
set outputname pfk_run1

firsttimestep 0
##                                SIMULATION PARAMETERS                        ##
# Input
paraTypeCharmm on
parameters    par_all36m_prot.prm
parameters    toppar_water_ions_namd.str
temperature   $temperature
# Force-Field Parameters
exclude       scaled1-4
1-4scaling    1.0
cutoff        12.0
switching     on
switchdist    10.0
pairlistdist  14.0
# Integrator Parameters
timestep      2.0 ;# 2fs/step
rigidBonds    all ;# needed for 2fs steps
nonbondedFreq 1
fullElectFrequency 2
stepscycle    10
# Constant Temperature Control
langevin      on ;# do langevin dynamics
langevinDamping 1 ;# damping coefficient (gamma) of 1/ps
langevinTemp  $temperature
langevinHydrogen off ;# don't couple langevin bath to hydrogens
# Periodic Boundary Conditions
cellBasisVector1 110.123 0. 0.0
cellBasisVector2 0.0 118.7 0.0
cellBasisVector3 0.0 0 121.841
cellOrigin       37.472 39.648 2.269
wrapAll          on
# PME (for full-system periodic electrostatics)
PME             yes
PMEGridSpacing  1.0
#manual grid definition
#PMEGridSizeX   45
#PMEGridSizeY   45
#PMEGridSizeZ   48
# Constant Pressure Control (variable volume)
useGroupPressure yes ;# needed for rigidBonds
useFlexibleCell  no
useConstantArea  no
langevinPiston  on
langevinPistonTarget 1.01325 ;# in bar -> 1 atm
langevinPistonPeriod 100.0
langevinPistonDecay 50.0
langevinPistonTemp  $temperature
# Output
outputName      $outputname
restartfreq     10000 ;# 10000steps = every 20 ps
dcdfreq        10000
xstfreq        10000
outputEnergies 10000
outputPressure 10000

##                                EXTRA PARAMETERS                            ##
##                                EXECUTION SCRIPT                            ##
# Minimization
minimize        1000
reinitvels     $temperature
run 50000000 ;# 100 ns
```

Figure A.1.1: Configuration file script for apo simulations of SaPFK.

```

##                                JOB DESCRIPTION                                ##
# Minimization and Equilibration of
# PFK in a Water Box
##                                ADJUSTABLE PARAMETERS                        ##
structure      ionized.psf
coordinates    ionized.pdb

set temperature 310
set outputname pfkctr_run1

firsttimestep 0
##                                SIMULATION PARAMETERS                      ##
# Input
paraTypeCharmm on
parameters    par_all36m_prot.prm
parameters    toppar_water_ions_namd.str
temperature   $temperature
# Force-Field Parameters
exclude       scaled1-4
1-4scaling    1.0
cutoff        12.0
switching     on
switchdist    10.0
pairlistdist  14.0
# Integrator Parameters
timestep      2.0 ;# 2fs/step
rigidBonds    all ;# needed for 2fs steps
nonbondedFreq 1
fullElectFrequency 2
stepspercycle 10
# Constant Temperature Control
langevin      on ;# do langevin dynamics
langevinDamping 1 ;# damping coefficient (gamma) of 1/ps
langevinTemp  $temperature
langevinHydrogen off ;# don't couple langevin bath to hydrogens
# Periodic Boundary Conditions
cellBasisVector1 110.123 0.0 0.0
cellBasisVector2 0.0 118.70 0.0
cellBasisVector3 0.0 0.0 121.841
cellOrigin       37.472 39.649 2.266
wrapAll          on
# PME (for full-system periodic electrostatics)
PME              yes
PMEGridSpacing  1.0
#manual grid definition
#PMEGridSizeX   45
#PMEGridSizeY   45
#PMEGridSizeZ   48
# Constant Pressure Control (variable volume)
useGroupPressure yes ;# needed for rigidBonds
useFlexibleCell  no
useConstantArea  no
langevinPiston   on
langevinPistonTarget 1.01325 ;# in bar -> 1 atm
langevinPistonPeriod 100.0
langevinPistonDecay 50.0
langevinPistonTemp  $temperature
# Output
outputName       $outputname
restartfreq      10000 ;# 10000 steps = every 20 ps
dcdfreq         10000
xstFreq         10000
outputEnergies  10000
outputPressure  10000
##                                EXTRA PARAMETERS                          ##
##                                EXECUTION SCRIPT                          ##
##                                CONSTRAINTS DURING MINIMIZATION          ##
extraBonds on
extraBondsFile constraint.txt

# Minimization
minimize        1000
reinitvels     $temperature

run 50000000 ;# 100 ns
bond 2061 7073 50 19.1
bond 2073 7133 50 15.3
bond 4015 8995 50 18.0
bond 2153 7041 50 17.3
bond 4396 2737 50 18.2
bond 17001 12053 50 19.1
bond 17013 12113 50 15.3
bond 18955 13975 50 18.0
bond 17093 12021 50 17.3
bond 19336 17677 50 18.2

```

(a)

(b)

**Figure A.1.2:** (a) Configuration file script for proposed allosteric site constrained simulations, (b) restrained bonds.

```

##                                JOB DESCRIPTION                                ##
# Minimization and Equilibration of
# PFK in a Water Box
##                                ADJUSTABLE PARAMETERS                        ##
structure      ionized.psf
coordinates    ionized.pdb

set temperature 310
set outputname  pfkctr_run1

firsttimestep  0
##                                SIMULATION PARAMETERS                        ##
# Input
paraTypeCharmm      on
parameters          par_all36m_prot.prm
parameters          toppar_water_ions_namd.str
temperature         $temperature
# Force-Field Parameters
exclude            scaled1-4
1-4scaling         1.0
cutoff             12.0
switching          on
switchdist         10.0
pairlistdist       14.0
# Integrator Parameters
timestep           2.0 ;# 2fs/step
rigidBonds         all ;# needed for 2fs steps
nonbondedFreq     1
fullElectFrequency 2
stepspercycle     10
# Constant Temperature Control
langevin           on ;# do langevin dynamics
langevinDamping   1 ;# damping coefficient (gamma) of 1/ps
langevinTemp      $temperature
langevinHydrogen  off ;# don't couple langevin bath to hydrogens
# Periodic Boundary Conditions
cellBasisVector1  110.123  0.0  0.0
cellBasisVector2  0.0  118.70  0.0
cellBasisVector3  0.0  0.0  121.841
cellOrigin        37.472  39.649  2.266
wrapAll           on
# PME (for full-system periodic electrostatics)
PME               yes
PMEGridSpacing    1.0
#manual grid definition
#PMEGridSizeX     45
#PMEGridSizeY     45
#PMEGridSizeZ     48
# Constant Pressure Control (variable volume)
useGroupPressure  yes ;# needed for rigidBonds
useFlexibleCell   no
useConstantArea   no
langevinPiston    on
langevinPistonTarget 1.01325 ;# in bar -> 1 atm
langevinPistonPeriod 100.0
langevinPistonDecay 50.0
langevinPistonTemp $temperature
# Output
outputName        $outputname
restartfreq       10000 ;# 10000 steps = every 20 ps
dcdfreq          10000
xstFreq          10000
outputEnergies   10000
outputPressure   10000
##                                EXTRA PARAMETERS                            ##
##                                EXECUTION SCRIPT                            ##
##                                CONSTRAINTS DURING MINIMIZATION            ##
extraBonds on
extraBondsFile constraint.txt

# Minimization
minimize         1000
reinitvels      $temperature

run 50000000 ;# 100 ns
bond 279      8243 50 17.4
bond 867      7821 50 11.9
bond 7349     9874 50 15.3
bond 345      7804 50 11.5
bond 874      9844 50 13.0
bond 5259     3263 50 17.4
bond 5847     2841 50 11.9
bond 2369     4894 50 15.3
bond 5325     2824 50 11.5
bond 2831     5854 50 11.8
bond 10329    18203 50 17.4
bond 10827    17781 50 11.9
bond 17309    19834 50 15.3
bond 10305    17764 50 11.5
bond 10834    19804 50 13.0

```

(a)

(b)

**Figure A.1.3:** (a) Configuration file script for known allosteric site constrained simulations, (b) restrained bonds.

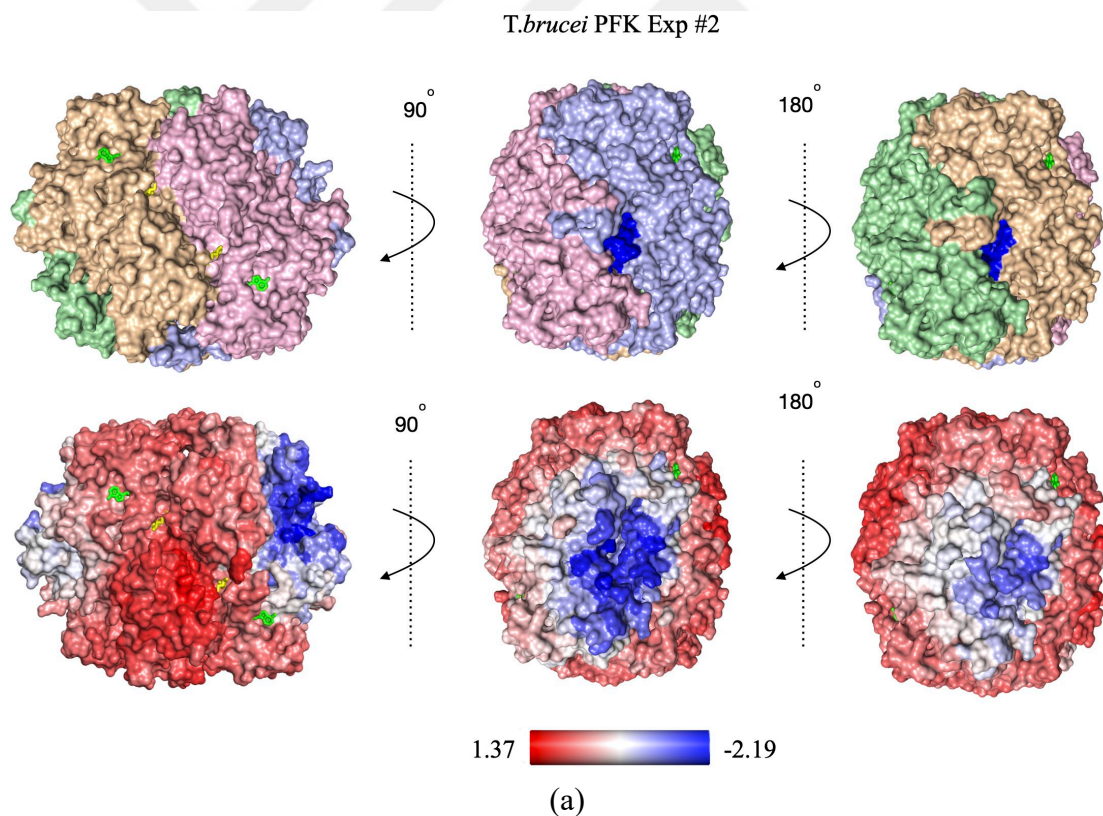


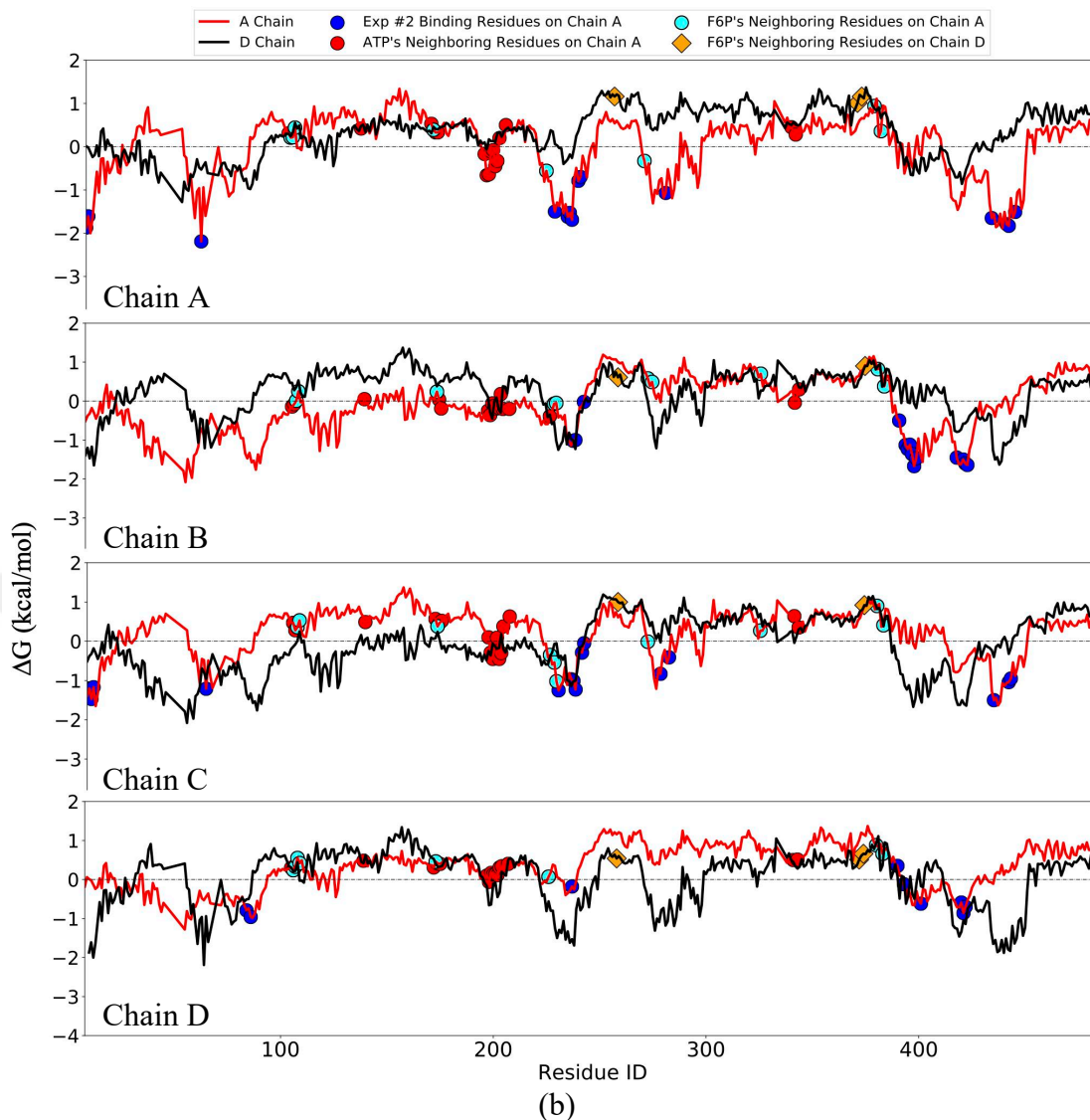
## A.2 AlloSigma Experiments on other proposed allosteric sites for *Tb*PFK.

**Table A.2:** Experiment numbers and corresponding residues for the perturbation sites in TbPFK.

Enzyme	Experiment	Binding Sites*
PFK <i>T. brucei</i>	#2 Proposed Site	A: T10, S11, R64, N230, H236-T238, F241, Q242, Q282, R435, Q442, L443, Q446 B: R237, T238, Q242, N390, L393-L398, M417, N420-Y422 C: R8-S11, R64, N230, H236, R237, T238, F241, Q242, F278, Q282, R435, Q442, L443 D: L84, A86, R237, N390, L393, L401, N420-Y422
	#3 Proposed Site	A: - B: K270-M272, R274, E325, K345-I349, D377-S379, Y380, R383 C: - D: K270-M272, R274, E325, K345-I349, D377, S379, Y380, R383
	#4 Proposed Site	A: R173, K270-M272, R274, E325, D339-G350, D377, S379, Y380, R383 B: K270-M272, R274, E325, K345-I349, D377, S379, Y380, R383 C: M272, E325, Q329, L346-G350, D377, S379, Y380, R383 D: K270-M272, R274, E325, K345-I349, D377, S379, Y380, R383

\* Proposed residues are taken from (Ayyildiz et al., 2020).





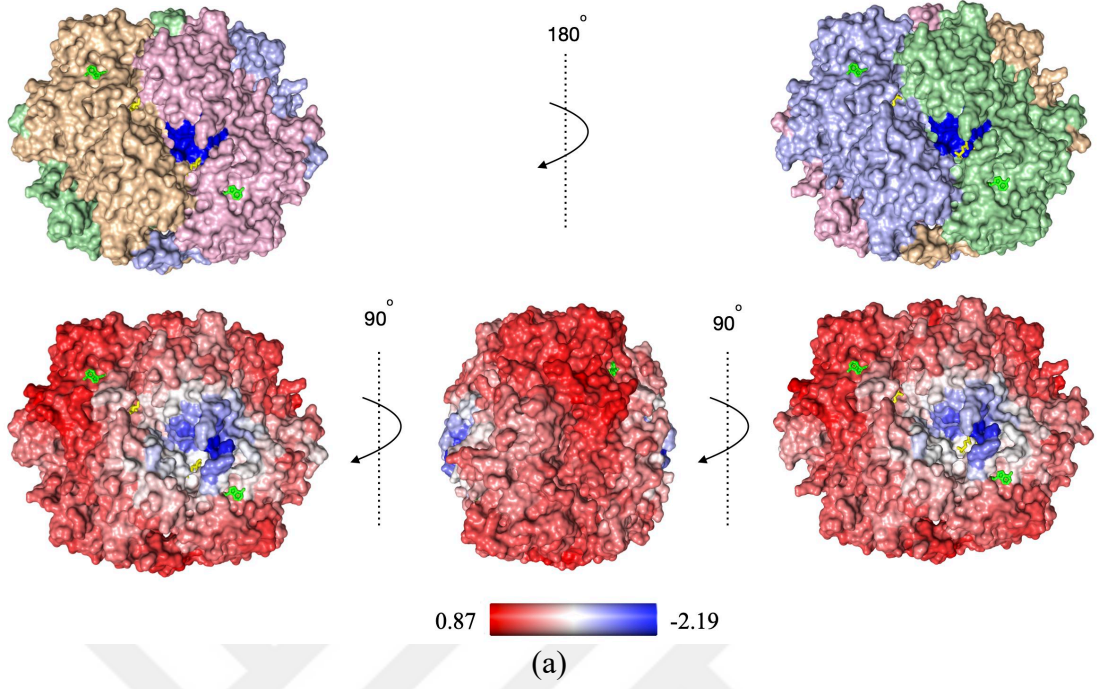
***T. brucei* PFK Exp #2**

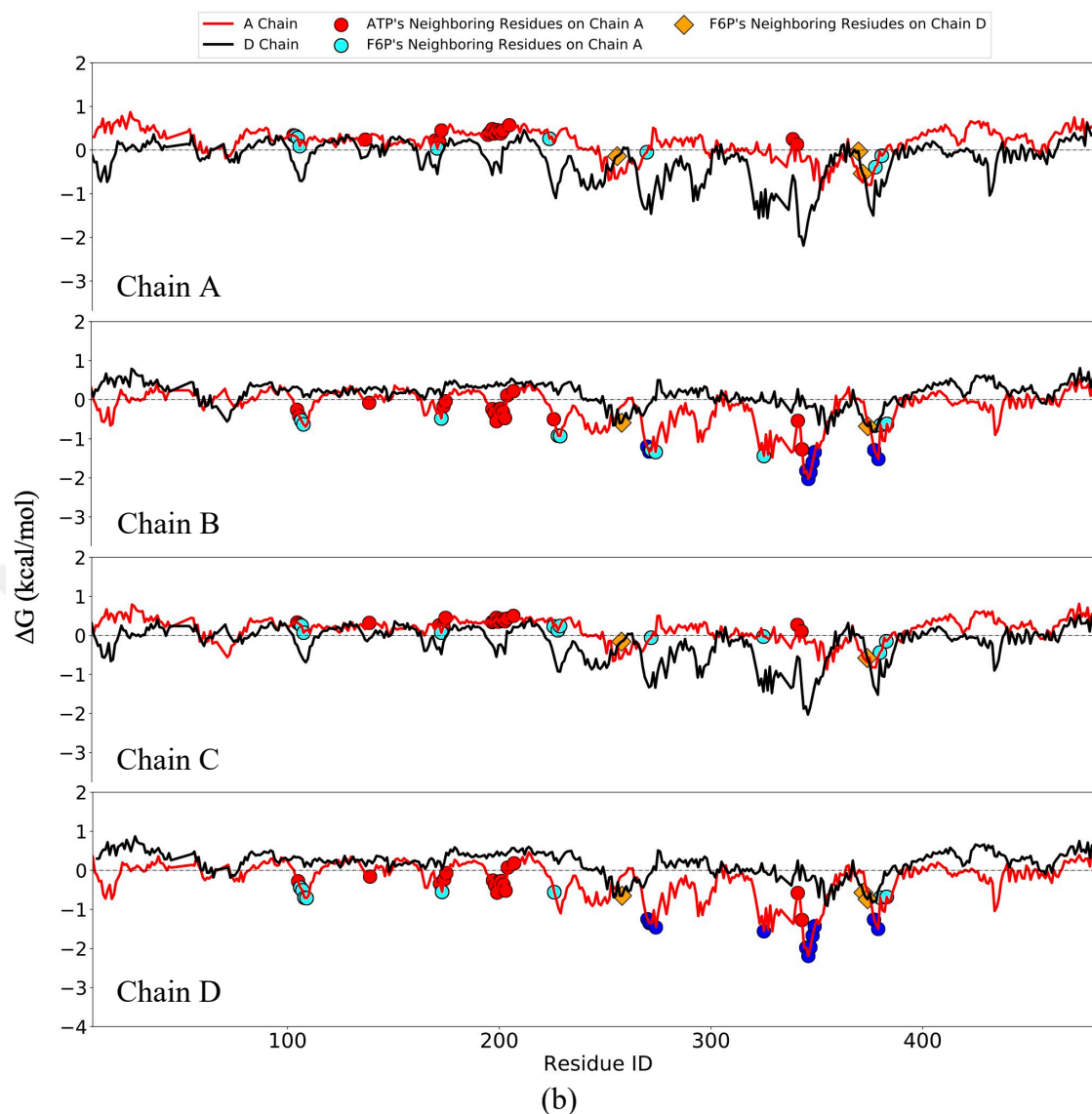
$\langle \Delta G \rangle$ (kcal/mol)		
Chain ID	Substrate (F6P)	Cofactor (ATP)
A	$0.46 \pm 0.55$	$0.07 \pm 0.39$
B	$0.41 \pm 0.32$	$-0.06 \pm 0.20$
C	$0.24 \pm 0.59$	$0.22 \pm 0.34$
D	$0.48 \pm 0.21$	$0.30 \pm 0.18$

(c)

**Figure A.2.1:** Perturbed residues for Exp #2 colored in dark blue in (a). At the catalytic sites of *T. brucei* PFK, ATP and F6P represented as green and yellow sticks, respectively. (b)  $\Delta G$  profiles of all chains, and (c)  $\langle \Delta G \rangle$  values in the catalytic sites. The figure was adapted from submitted paper (Çelebi et al., 2021)

*T.brucei* PFK Exp #3





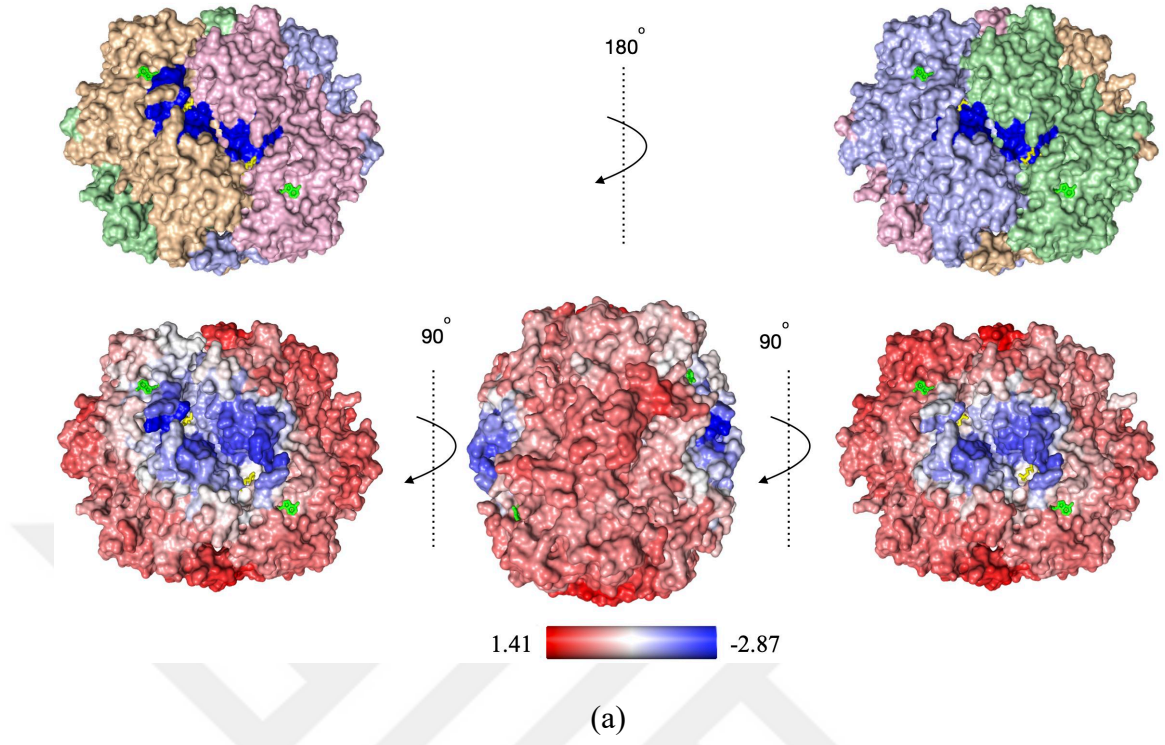
### *T. brucei* PFK Exp #3

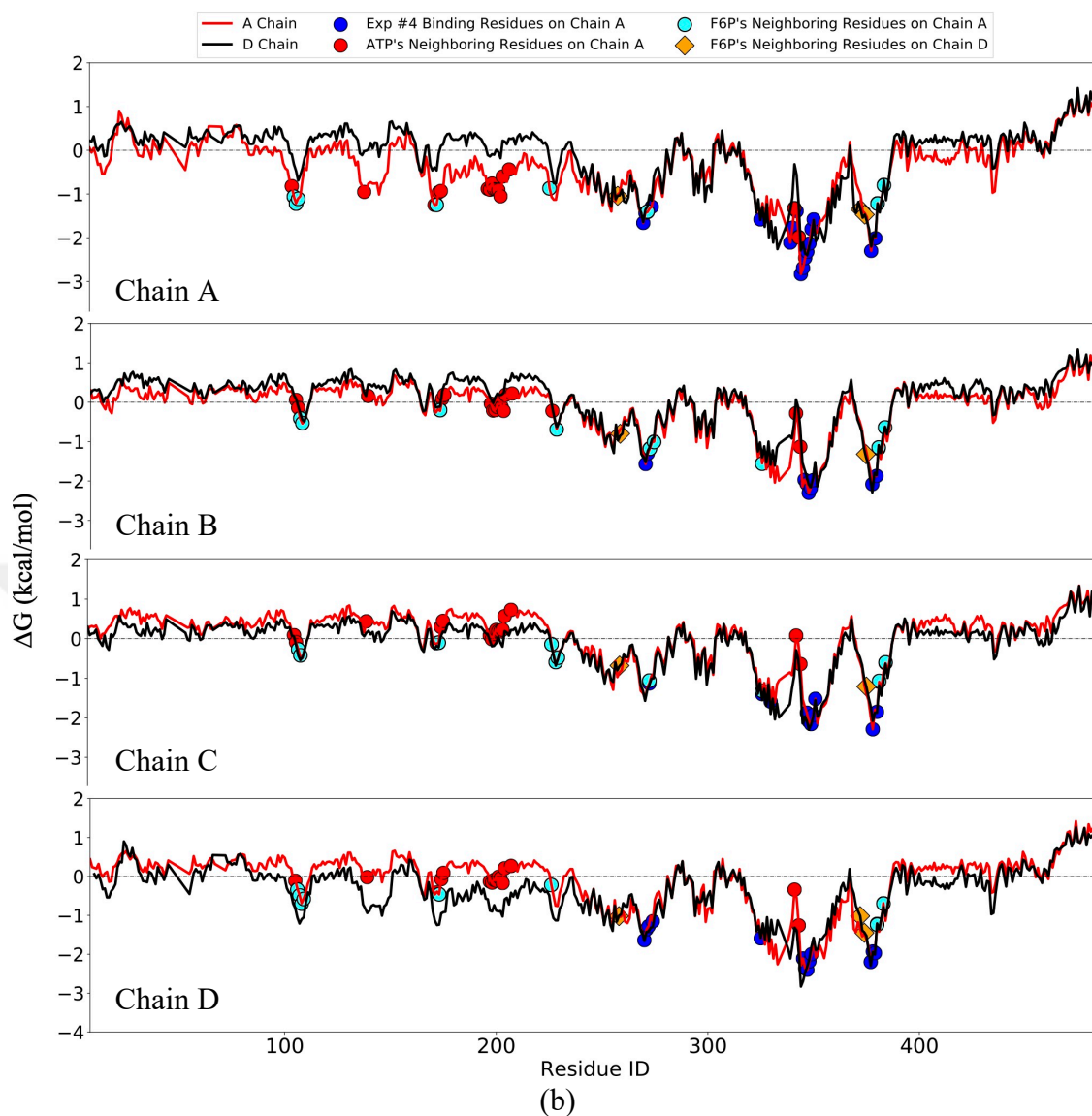
$\langle \Delta G \rangle$ (kcal/mol)		
Chain ID	Substrate (F6P)	Cofactor (ATP)
A	$-0.03 \pm 0.26$	$0.31 \pm 0.16$
B	$-0.84 \pm 0.33$	$-0.35 \pm 0.31$
C	$-0.04 \pm 0.25$	$0.30 \pm 0.15$
D	$-0.62 \pm 0.09$	$-0.39 \pm 0.29$

(c)

**Figure A.2.2:** Perturbed residues for Exp #3 colored in dark blue in (a). At the catalytic sites of *T. brucei* PFK, ATP and F6P represented as green and yellow sticks, respectively. (b)  $\Delta G$  profiles of all chains, and (c)  $\langle \Delta G \rangle$  values in the catalytic sites. The figure was adapted from submitted paper (Çelebi et al., 2021).

*T.brucei* PFK Exp #4





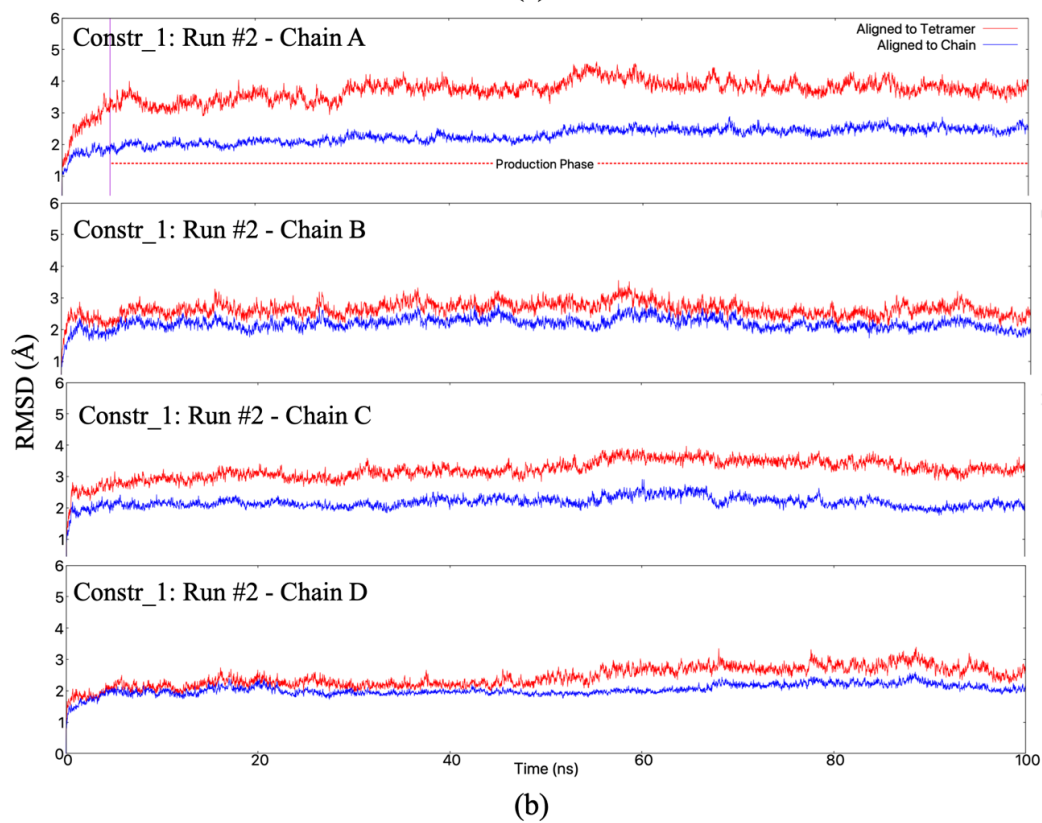
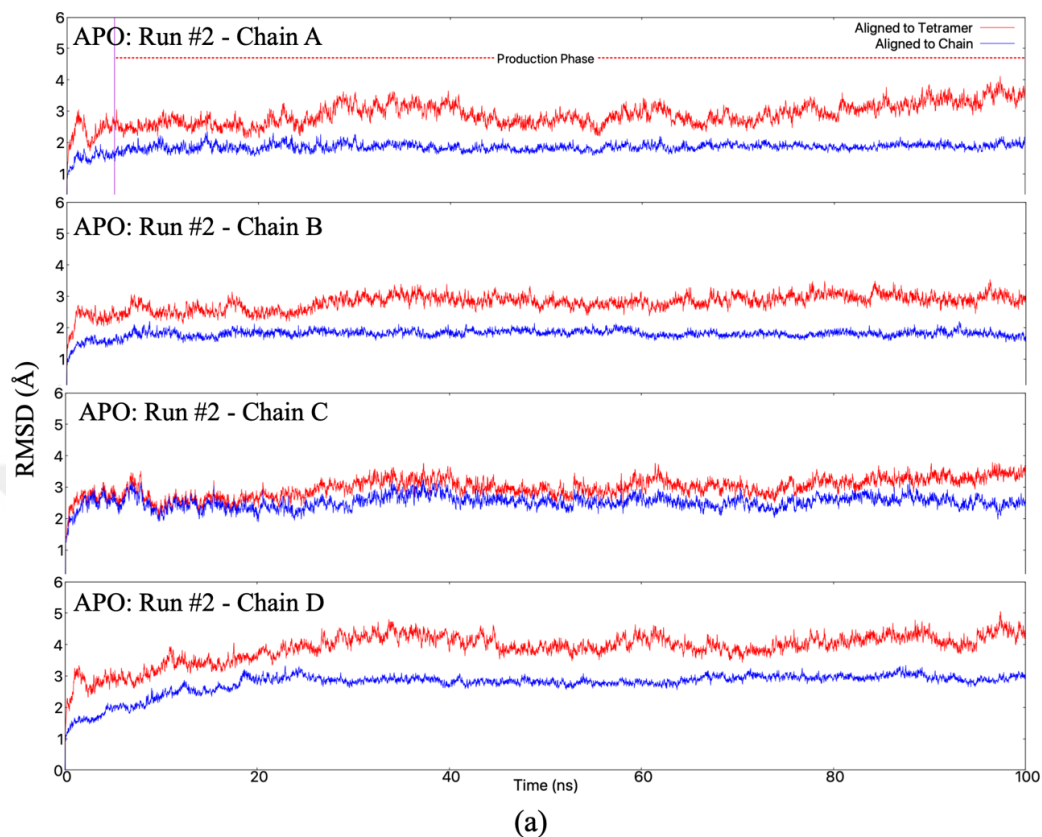
***T. brucei* PFK Exp #4**

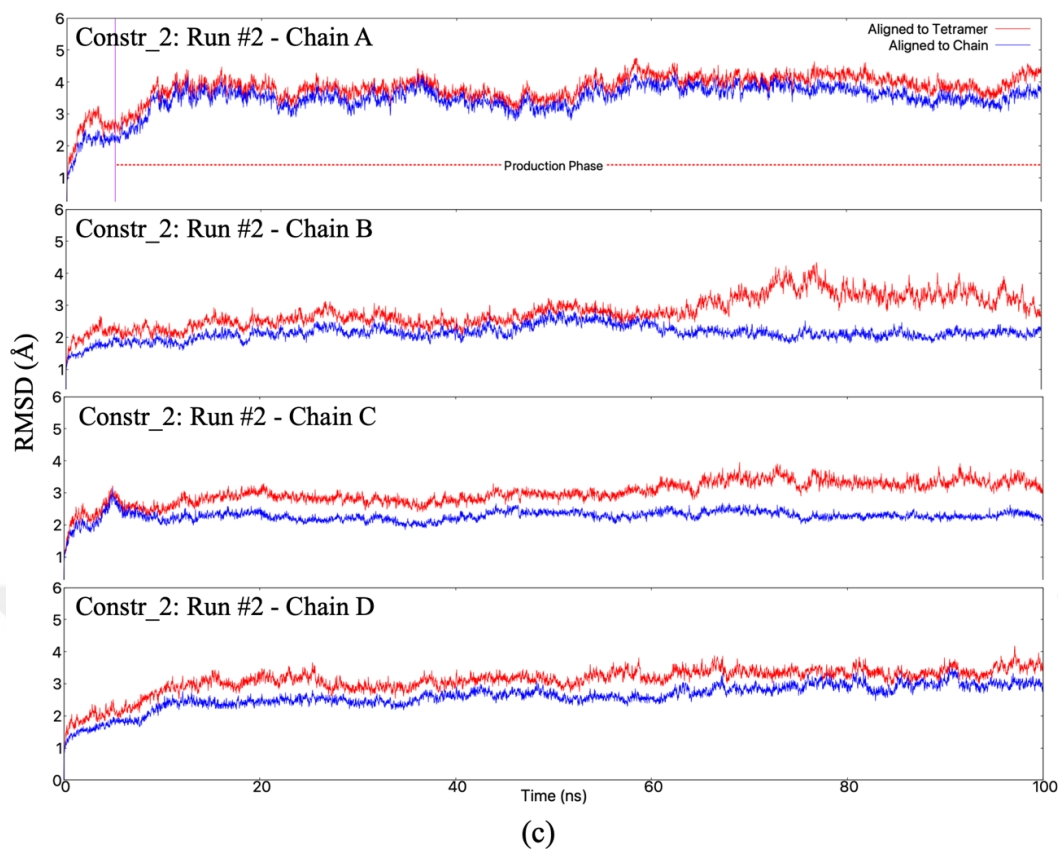
$\langle \Delta G \rangle$ (kcal/mol)		
Chain ID	Substrate (F6P)	Cofactor (ATP)
A	$-1.16 \pm 0.20$	$-0.99 \pm 0.31$
B	$-0.84 \pm 0.39$	$-0.15 \pm 0.31$
C	$-0.66 \pm 0.40$	$0.08 \pm 0.33$
D	$-0.75 \pm 0.37$	$-0.22 \pm 0.33$

(c)

**Figure A.2.3:** Perturbed residues for Exp #4 colored in dark blue in (a). At the catalytic sites of *T. brucei* PFK, ATP and F6P represented as green and yellow sticks, respectively. (b)  $\Delta G$  profiles of all chains, and (c)  $\langle \Delta G \rangle$  values in the catalytic sites. The figure was adapted from submitted paper (Çelebi et al., 2021).

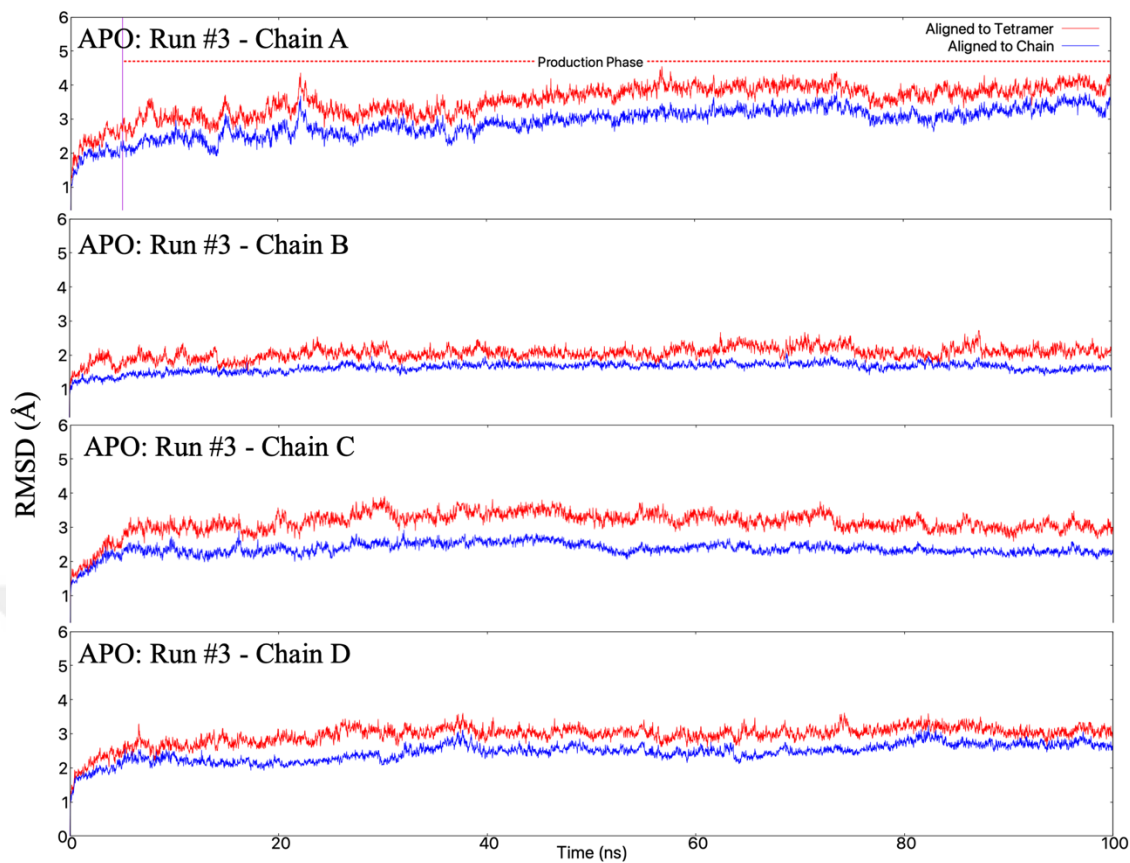
### A.3 RMSD Profiles for 2<sup>nd</sup> and 3<sup>rd</sup> Runs of Apo and Constrained Molecular Dynamics Simulations.



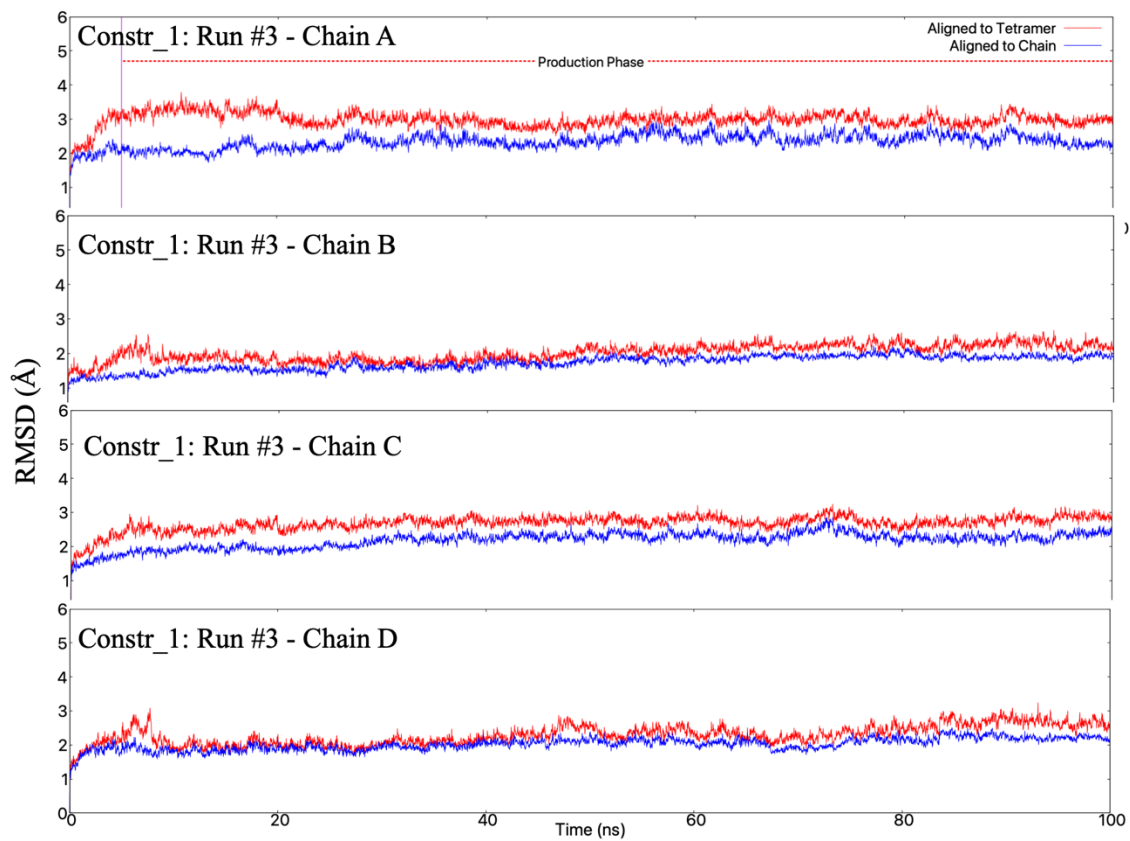


**Figure A.3.1:** Two different states aligned RMSD plots of (a) Apo form Run #2, (b) proposed allosteric site constrained Run #1, and (c) known allosteric site constrained Run #2. The red lines represent tetramer-aligned RMSD while the blue lines represent chain-aligned RMSD.

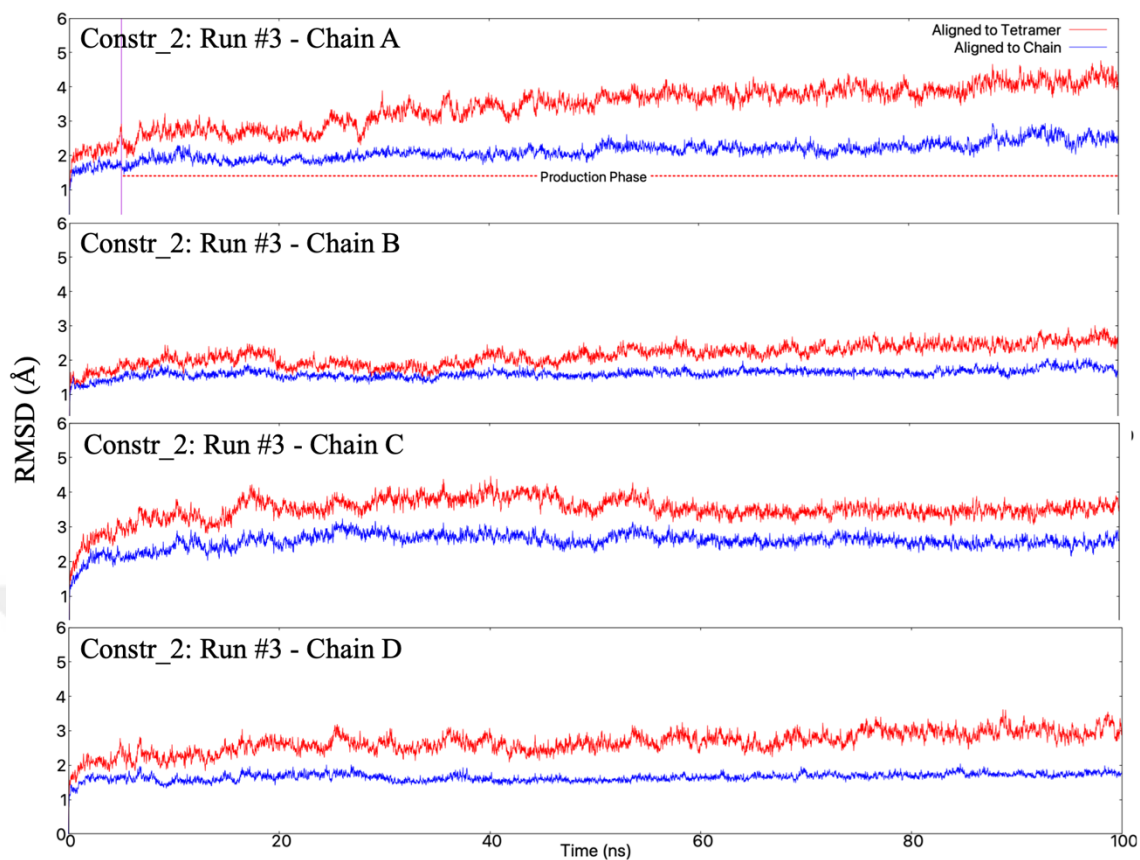




(a)



(b)



(c)

**Figure A.3.2:** Two different states aligned RMSD plots of (a) Apo form Run #3, (b) proposed allosteric site constrained Run #3, and (c) known allosteric site constrained Run #3. The red lines represent tetramer-aligned RMSD while the blue lines represent chain-aligned RMSD.

## A.4 Used Commands for Principal Component Analysis (PCA) Calculations.

### Step 1: Installation of ProDy with PyLab on MacOS Terminal:

```
pip install -U ProDy
```

### Step 2: Running on the terminal:

```
prody pca -a -A --select calpha --pdb ionized.pdb

# In[1]:
from prody import *
from pylab import *
ion()
structure = parsePDB('ionized.pdb')
structure
dcd = DCDFile('pfk_cat_run1.dcd')
dcd.link(structure)
dcd.setAtoms(structure.calpha)
dcd
# In[2]:
pca = PCA('Apo Run-1')
pca.buildCovariance( dcd )
#In [3]:
pca.calcModes('all')
pca
# In[4]:
saveModel(pca)
# In[5]:
for mode in pca[:400]:
    print (calcFractVariance(mode).round(5))
```

**At the end of Step-2, the cross-correlation values between residues were calculated and written in an txt file.**

**Step 3, The calculated cross correlation values were written in a txt file as one column with the following MATLAB commands.**

```
cor=readtable('pfk_apo_run1_pca_cross-correlations.txt')
A = table2array(cor)
a=reshape(A',[],1)
fileID = fopen('corr.txt','w');
fprintf(fileID,'%6.4f\n',a);
```

**Step 4, The cross-correlation maps were then plotted with Gnuplot v5.4 (Williams & Kelly, 2011).**