KADIR HAS UNIVERSITY

SCHOOL OF GRADUATE STUDIES

DEPARTMENT OF COMPUTATIONAL BIOLOGY AND BIOINFORMATICS

INVESTIGATION OF ALLOSTERIC INHIBITION EFFECT IN PYRUVATE KINASE BY CONSTRAINED MOLECULAR DYNAMICS SIMULATION METHOD

REYHAN AKKAYA

PROF. DR. E. DEMET AKDOĞAN

MASTER'S DEGREE THESIS

ISTANBUL, JUNE, 2021

REYHAN AKKAYA

Master's Degree Thesis

2021

INVESTIGATION OF ALLOSTERIC INHIBITION EFFECT IN PYRUVATE KINASE BY CONSTRAINED MOLECULAR DYNAMICS SIMULATION METHOD

REYHAN AKKAYA

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 of Science in Computational Biology and Bioinformatics

İSTANBUL, JUNE, 2021

DECLARATION OF RESEARCH ETHICS / METHODS OF DISSEMINATION

I, REYHAN AKKAYA, 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 institution;
- 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 Information Center under the following condition as indicated below:

☐ The full content of my thesis will be accessible from everywhere by all means.

REYHAN AKKAYA

30/07/2021

KADİR HAS UNIVERSITY

SCHOOL OF GRADUATE STUDIES

ACCEPTANCE AND APPROVAL

This work entitled INVESTIGATION OF ALLOSTERIC INHIBITION EFFECT IN PYRUVATE KINASE BY CONSTRAINED MOLECULAR DYNAMICS SIMULATION METHOD prepared by REYHAN AKKAYA has been judged to be successful at the defense exam on 30/07/2021 and accepted by our jury as master's thesis.

APPROVED BY:

E. Demet AKDOĞAN (Advisor)   . . . .
Kadir Has University

Prof. Dr. Kemal YELEKÇİ   . . . .
Kadir Has University

Assoc. Prof. Ayşe Özge Kürkçüoğlu LEVITAS   . . . .
İstanbul 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: 30/07/2021

TABLE OF CONTENTS

INVESTIGATION OF ALLOSTERIC INHIBITION EFFECT IN PYRUVATE KINASE BY CONSTRAINED MOLECULAR DYNAMICS SIMULATION METHOD

ABSTRACT

In previous studies, our research group identified species-specific allosteric sites in glycolytic enzymes from different organisms and identified candidate inhibitory molecules that strongly interact with the residues at these sites. In this study, a Molecular Dynamics simulation study was performed on one of the glycolytic enzymes, bacterial pyruvate kinase (*S. aureus* PK), by employing bond restraints between selected pairs of residues at the suggested allosteric region in order to mimic the presence of a drug molecule. At the same time, interacting residues in an experimentally identified allosteric region were also restricted and compared with the proposed area. Three 100 ns long independent runs were conducted for each of three different states of the receptor; apo state (no restraints), constr-1 (restriction on proposed allosteric site) and constr-2 (restriction on known allosteric site) which amount to a total of nine runs, e.g., 900 ns.

Several analytical methods were used to elucidate the effect of restricted regions on protein dynamics and the allosteric character of pyruvate kinase. While structural changes were examined with RMSD-RMSF analysis, correlations between global movements and structural components were analysed with principal component analysis. From PCA results, it was observed that both restricted allosteric regions led to decreased correlations of positional fluctuations in comparison to apo state. Investigation of changes in the secondary structure at the catalytic site showed that the $\alpha'6$ helix, which has an essential role in the stabilization of the active structure, shifted to a coil-turn structure in the apo state more frequently than in the constrained states. Additionally, domain rotations identified via principal axes analysis, showed that the constrained state disrupted the domain rotations more often. Fi-

nally, the distance fluctuation analysis was performed to observe the effect of the restricted residues in the allosteric signal transduction. The communications between residues increased in the constr-1 state while there was slight decrease in the communications of the constr-2 state.

Keywords: Pyruvate kinase, Allosteric inhibition, Molecular Dynamics Simulation, Constrained residues

PİRUVAT KİNAZIN ALLOSTERİK OLARAK İNHİBE EDİLMESİNİN KISITLANMIŞ MOLEKÜLER DİNAMİK SİMİLASYON METHODU KULLANILARAK İNCELENMESİİ

ÖZET

Önceki çalışmalarda, araştırma grubumuz farklı organizmalarda bulunan glikolitik enzimlerdeki türe özgü allosterik bölgeleri tanımladı ve bu bölgelerdeki kalıntılarla güçlü bir şekilde etkileşime giren aday inhibitör molekülleri belirledi. Bu çalışmada ise, glikolitik enzimlerden biri olan bakteriyel piruvat kinaz enzimi (*S. aureus* PK) üzerinde bir Moleküler Dinamik simülasyon çalışması gerçekleştirildi ve önerilen allosterik bölgede seçilen kalıntı çiftleri arasındaki bağlar kısıtlanarak bir ilaç molekülünün varlığı taklit edildi. Aynı zamanda, deneysel olarak tanımlanmış bir allosterik bölgeye bağlanan ligand ile etkileşime giren kalıntılar da kısıtlandı ve önerilen alanla karşılaştırıldı. Reseptörün üç farklı durumunun her biri için üç adet 100 ns uzunluğunda bağımsız çalışma yapılmıştır; apo durumu (kısıtlama yok), constr-1 (önerilen allosterik bölgede kısıtlama) ve constr-2 (bilinen allosterik bölgede kısıtlama), toplam dokuz çalıştırma, örneğin 900 ns.

Kısıtlı bölgelerin protein dinamikleri üzerindeki etkisini ve piruvat kinazın allosterik karakterini aydınlatmak için çeşitli analitik yöntemler kullanıldı. RMSD-RMSF analizi ile yapısal değişimler incelenirken, temel bileşen analizi ile küresel hareketler ve yapısal bileşenler arasındaki korelasyonlar analiz edilmiştir. PCA sonuçlarından, her iki kısıtlı allosterik bölgenin, apo durumuna kıyasla konumsal dalgalanma korelasyonlarının azalmasına yol açtığı gözlemlendi. Katalitik bölgedeki ikincil yapıdaki değişikliklerin araştırılması, aktif yapının stabilizasyonunda önemli bir role sahip olan $\alpha'6$ sarmalının, apo durumunda, kısıtlı durumlara göre daha sık bir bobin-dönüş yapısına kaydığını göstermiştir. Ek olarak, temel eksen analizi ile tanımlanan alan rotasyonlarının, kısıtlı durumda çok daha fazla bozulduğunu gösterdi. Son olarak, allosterik sinyal iletimi açısından kısıtlanmış kalıntıların etkisini gözlemlemek için

mesafe dalgalanma analizi yapıldı. Kalıntılar arasındaki iletişim, constr-1 durumunda artarken, constr-2 durumunun iletişiminde çok hafif bir düşüş görüldü.

Anahtar Sözcükler: Piruvat Kinaz, Allosterik inhibisyon, Moleküler Dinamik Similasyon, Kısıtlanmış kalıntılar

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deep and sincere gratitude to my thesis supervisor Prof. E. Demet Akten Akdoğan for the sensitivity, the motivation, and the support that she has shown me since the beginning of my thesis studies.

I am especially grateful to my colleague Metehan Çelebi who were always there to help and support me in any difficulties. I would also like to thank our research group members Merve Ayyıldız, Serkan Çeliker and Mehmet Fatih Özhelvacı for their assistance. Furthermore, a special thanks goes to my dearest friend Çağrı Akkaya for his endless love, friendship, and psychological support to me in my education life.

Finally, I am grateful to my family and friends for their support in all circumstances throughout my life.

I would like to thank Kadir Has University for giving me scholarship.

To My Dearest Family...

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS/ABBREVIATIONS

α	Alpha
Å	Angstrom
β	Beta
Δ	Delta
ΔG	Gibbs Free Energy
ADP	Adenosine diphosphate
AMP	Adenosine 5'-monophosphate
ATP	Adenosine triphosphate
λ	Lambda
Cα	Carbon Alpha
FDA	US Food and Drug Administration
FBP	Fructose 1,6-biphosphate
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
PEP	Phosphoenolpyruvate
PK	Pyruvate Kinase
ProDy	Protein Dynamics & Sequence Analysis
RMSD	Root Mean Square Deviation
RMSDF	Root mean square distance fluctuations
RMSF	Root Mean Square Fluctuation
VMD	Visual Molecular Dynamics

1. INTRODUCTION

1.1 Role of Pyruvate Kinase In Glycolysis
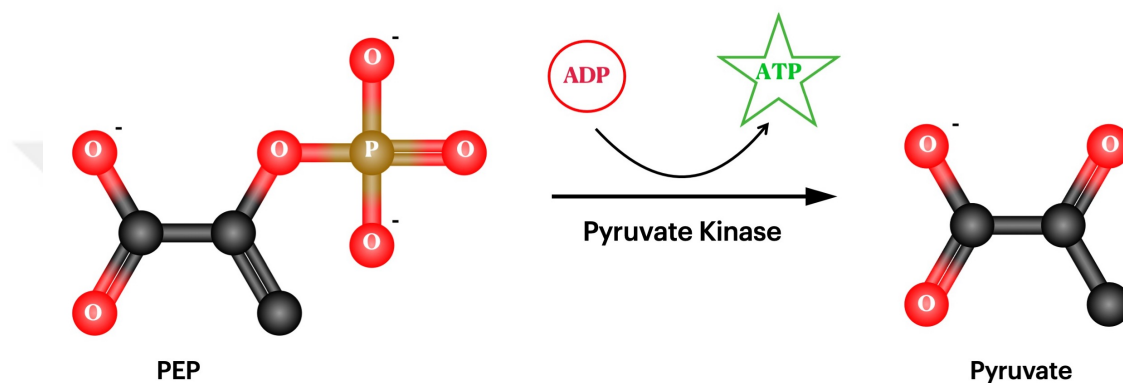
Figure 1.1 Pyruvate kinase reaction in the glycolysis

Pyruvate kinase (PK) enzyme plays a vital role in the energy production mechanism of living things by catalysing the transformation of phosphoenolpyruvate into pyruvate in the last step of glycolysis. As illustrated in Figure 1.1, a phosphate group from phosphoenolpyruvate joins ADP to produce ATP, while on the other hand, the pyruvate necessary for the cell is made (Zoraghi et al., 2010). It is converted to acetyl-CoA in aerobic conditions, producing more energy or fatty acids in the mitochondria while converted into lactic acid under oxygen-free conditions. Also, pyruvate can be used to synthase glucose, alanine or oxaloacetate according to cell needs. The regulation of pyruvate transformation is firmly regulated and controlled by hexokinase, phosphofructokinase, and pyruvate kinase. ATP generation and metabolic flux are mainly regulated by pyruvate kinase; thus it becomes an important drug target. The fact that the amino acid sequence and composition of allosteric regions differ from species to species is another reason they are preferred for species-specific drug design (Schormann et al., 2019). *S. aureus* PK

is homeotropically activated by its own substrate which is phosphoenolpyruvate (PEP) and not by adenosine-diphosphate (ADP). Unlike other species PKs, *S. aureus* PK is heterotrophically activated by ribose 5-phosphate (R5P) and adenosine 5'-monophosphate (AMP) effectors, e.g. activated/inhibited by an effector not a substrate. Differently from other known bacterial PK enzymes, *S. aureus* PK inhibited by FBP at physiological concentrations. Because it has a glutamate residue in its effector binding pocket instead of a conserved lysine residue in that position in the PKs activated by fructose 1,6-biphosphate (FBP) (Zoraghi et al., 2010). Additionally, it is the only bacterial PK that is showing the dual effect of ATP. While low concentrations of ATP activate the enzyme, high ATP concentrations (> 2.5 mM) are enough to completely inhibit it.

Due to rapid consumption of ATP during the growth of bacteria, PK is highly activated with the low concentration of PEP and the presence of R5P and AMP activators. In the stationary growth phase, deficiency of existing sugar causes a decrease in the intracellular R5P by starting a reduction in the pentose phosphate pathway action. The decreased demand for ATP stimulates the inhibition of PK activity. This leads to an increase in the PEP concentration, which means a continuous ATP production mechanism (Zoraghi et al., 2010).

1.2 Structure of *S. aureus* Pk

Many PK structures form various organisms, for instance, *Escherichia coli*, *Leishmania Mexicana*, *Bacillus stearothermophilus*, rabbit muscle, human erythrocyte, and yeast shows the high level of sequence conservation and structural homology (Zoraghi et al., 2010). As illustrated in Figure 1.2, the structure of PK exists as a homo-tetramer. Each unit consists of a combination of four domains, A,B,C domains are the shared domains in all organisms and the N- terminal, C -terminal domains are the varying additional domains in different species. The most conserved one is the α/β barrel A domain located between B and C domains. The active site of PK is positioned between A and B domains. The sequentially most variable B domain

is a small unit that formed α/β-barrel structure (Mattevi et al., 1995). The binding pocket of the allosteric effector is found in the α/β structured C domain. While prokaryotic pyruvate kinases do not carry helical N-terminal domain, CT domain found in *S. aureus* and *G. stearothermophilis* (Morgan et al., 2014). It is suggested that the CT domain is derived from the PEP synthase in the late evolutionary phase for *B. stearothermophilus* (Pocalyko et al., 1990). However, its biological significance and function are not fully understood because of its absence in other bacterial sequences.

The tetrameric form is strengthened by interactions between chains and domains . Neighbouring A/A (large interface), C/C (small interface) and A/CT domains interactions provide the rigidity and stability of the overall structure. The communication between domains and chains is crucial for the allosteric activation or inhibition signalling (Zoraghi et al., 2010).
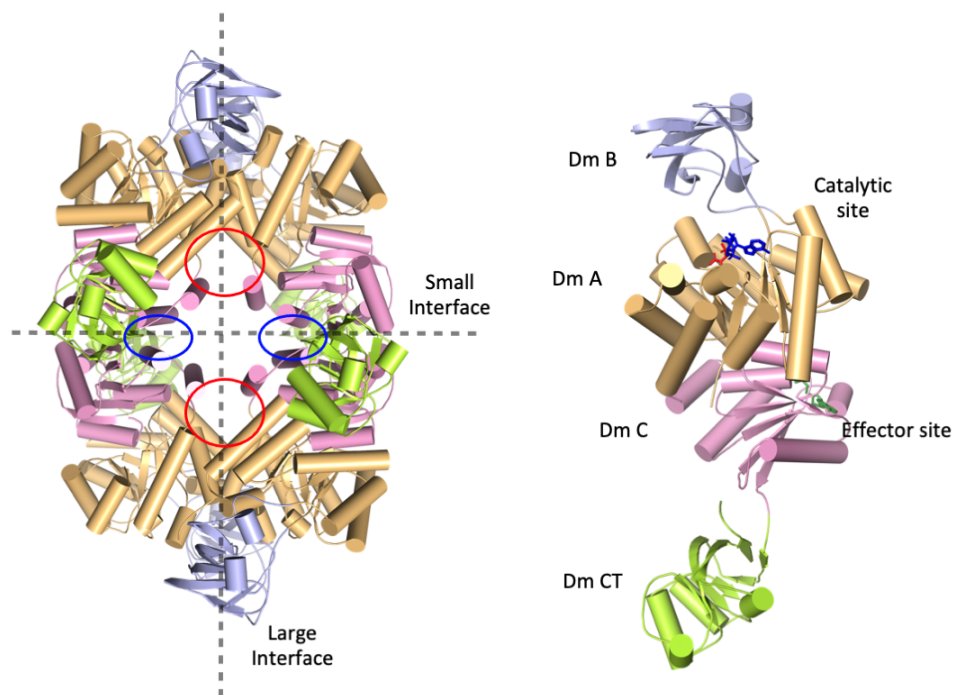
Figure 1.2 Structure of *S. aureus* PK coloured based on four different domains (PDB id: 3t0t (Axerio-Cilies et al., 2012))

1.3 Mechanism of Allosteric Regulation

As pyruvate kinase adapts an active R-state, the stability and rigidity of the tetramer structure increases. Interactions and conformational changes between chains and subunits cause transitions between active and inactive states. It has been examined that there are differences in allosteric regulation mechanisms between species and even between isomeric PKs found in one species. Therefore, almost every PK is unique to a specific allosteric regulation mechanism (Schormann et al., 2019). Structural differences of PK have a significant effect on the evolution of diverse allosteric mechanism strategies which is shown in Figure 1.3. From primary bacteria to the complex mammalians, even if the strategy is changed, all aimed to provide rigidity and stability of the tetramer in the active form by increasing efficiency of the reaction (Morgan et al., 2014).
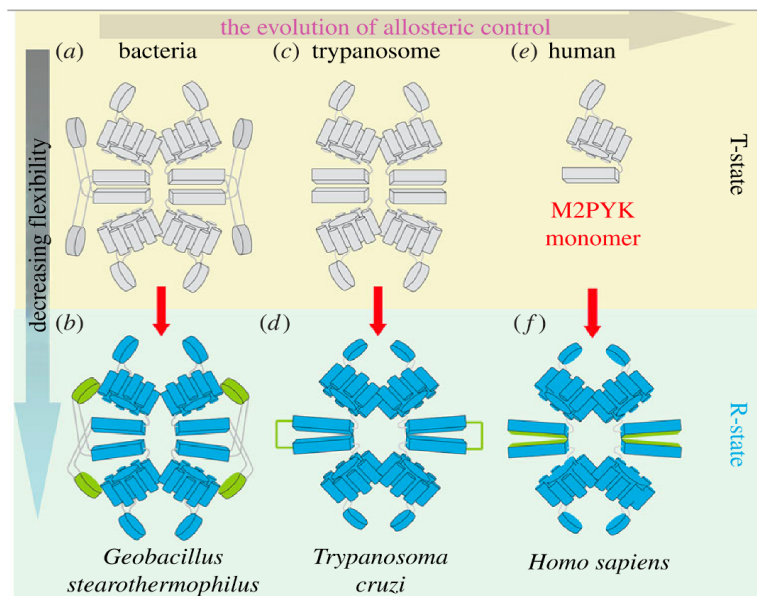
Figure 1.3 Two proposed allosteric activation model shown for E.coli PK
(Donovan et al., 2016)

Human PK-m2 has a unique allosteric activation mechanism that interconverts between inactive and active tetramer (Ashizawa et al., 1991). The binding of the FBP activator starts a rigid body rotation of monomers that disturbs the inactive T state, which is provided by hydrogen bonding networks (Morgan et al., 2013). Parasitic *L. mexicana* PK uses symmetrical rigid body rocking motions of A and C

domains of four monomers on the tetramer, which is called "rock and lock". The stability of the active R state is provided by eight salt bridges between the C-C small interface. Besides the rigid body movement, binding of the effector provides faster activation(Morgan et al., 2010; Naithani et al., 2015).

For *M. tuberculosis* PK, similar to *S. aureus* PK activated by AMP, allosteric activation carried on the rigid body rotations of the A and C domains on each monomer. This rotation ruins the hydrogen bonds interactions between AMP binding and C-tail loops on the C-C interface and neighbouring C domains. Allosteric activator signals cause the rotation of the neighbouring Aα6′ helix on the large A-A interface besides the catalytic region. Also, adjacent Aα7′ helixes make new hydrogens that stabilize the active states. Each monomer and allosteric ligand create new hydrogen bonds that provide stability and rigidity for the active form (Zhong et al., 2017).

The study of Donovan and his coworkers shows that bacterial *E. coli* PK is allosterically activated by fructose 1,6-biphosphate, leading to destabilisation on the α/β barrels between the catalytic and allosteric site of each monomeric units. The destabilisation and increase in the conformational flexibility provide the substrate binding. They showed that this mechanism is compatible with the domain rotation model which is illustrated in Figure 1.4. Increased flexibility at the interdomain interfaces is conflicting with rigid body reorientation model which suggest that active site rigidification is provided only by reorientation of the monomers. (Donovan et al., 2016).

The close relative of *S. aureus* PK, *G. stearothermophilis* pyruvate kinase, is also activated by R5P and AMP, inhibited by FBP, and has an extra CTminal domain. The rotation of the CTminal domain is provided to transform into a stable structure by binding the effector. When the *G. stearothermophilis* PK and *T. cruzi* PK structures are compared (Figure 1.4), it is thought that the CT domain takes the role of the effector loop fixing the C-C interface (Suzuki et al., 2008). On the other hand, CT domain deletion on mutant *S. aureus* PK and *B. stearothermophilus*
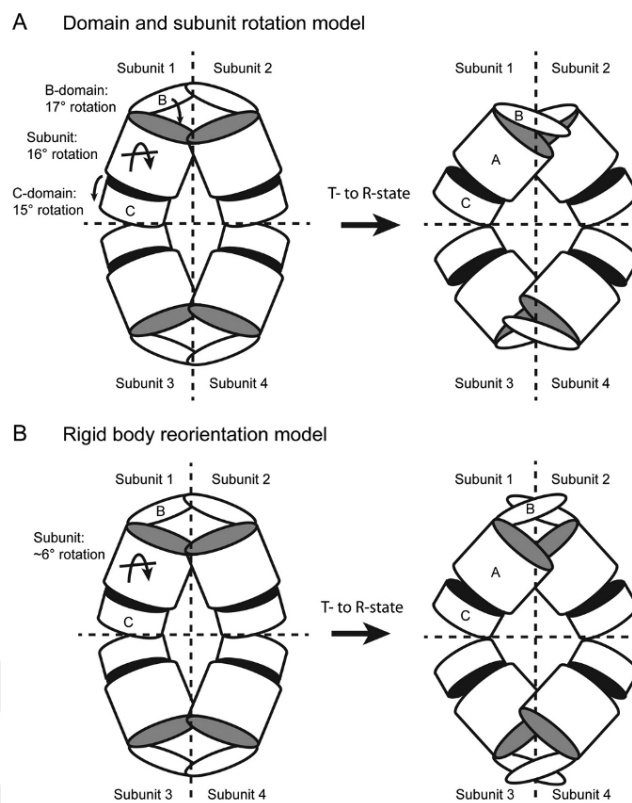
Figure 1.4 The evolution of allosteric regulation (Morgan et al., 2014)

PK has decreased catalytic activity while not leading to any significant change in homo-tetramerization (Sakai, 2004; Zoraghi et al., 2010). Inosine monophosphate (IMP) inhibits *T. saccharolyticum* PK, which has an extra CT domain without CT domain in the mutant protein (Fenton et al., 2021). For the efficient catalytic activity, structural rigidity and stability of the tetramer are provided by CT domain interactions (Zoraghi et al., 2010).

On the other hand, it is thought that allosteric inhibition of the PK enzyme works by disrupting the stability and rigidity of the tetramer structure. Although it is unknown how *S. aureus* PK provides inhibition by binding to the effector site by FBP, there are inhibitor drugs that bind to other allosteric sites. IS-130, one of these allosteric drugs, connects to the area shown in blue on the small interface illustrated in the figure. This drug is thought to inhibit the enzyme by strengthening the bonds at the C-C interface (Axerio-Cilies et al., 2012). Another allosteric inhibitor bound to this region cause conformational rearrangement in the tetramer by compressing

and elongating it (Zoraghi et al., 2011).

2. MATERIALS AND METHODS

2.1 Theory of Molecular Dynamics Simulations

Molecular Dynamics simulation is a widely used method in computational studies. It allows us to investigate protein behaviours with small- and large-scale motions and calculates system's potential energy by interatomic potentials and applies force on particles in each iteration. Potential energy is the summation of the bonded and nonbonded energy terms represented as in Figure 2.1. Energy components describe these bonded and nonbonded interaction types. Forcefield equation mathematically express the dependence of the energy of a system on the coordinates of its particles. Parameters can be obtained from quantum mechanics calculations to recover experimental data. Commonly used forcefields are AMBER (Assisted Model Building with Energy Refinement) (Salomon-Ferrer et al., 2013), CHARMM (Chemistry at HARvard Macromolecular Mechanics) (Brooks et al., 1983), OPLS (Optimized Potentials for Liquid Simulations)(Sambasivarao & Acevedo, 2009) , GROMOS (GROningen Molecular Simulation) (Christen et al., 2005) (González, 2011). Describing quantum mechanics facts in classical terms is the aim of a good forcefield. Unfortunately, incorporating the details of electron's behaviour of proteins contains large number of atoms is time consuming and challenging for QM calculations. Therefore, it is necessary to use notable physical approximations to identify intermolecular interactions even if they restrict their accuracy. These lead to the development of empirical forcefield potentials (González, 2011).

Newton's second law of motion gives the equations to calculate movements of the system with 3N particles in x, y and z directions.
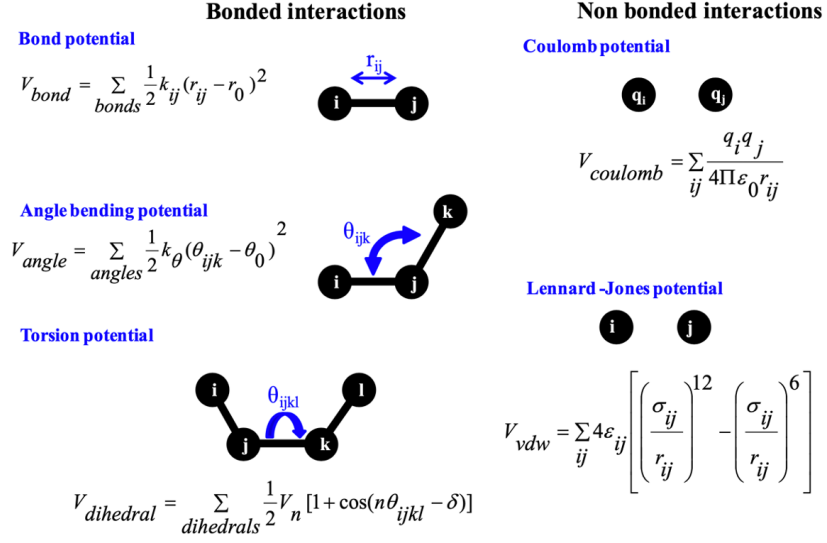
Figure 2.1 Bonded and nonbonded interactions representations and potential energy force field (adopted from (Sciences & Dantu, 2012).

$$\vec{F}_i = m_i \vec{a}_i \tag{2.1}$$

$$\vec{F}_i = -\Delta V_i = -\frac{dV\left(r_1, \ldots, r_N\right)}{dr_i} = m_i \frac{d^2 r_i}{dt^2} = -\frac{dV}{dr_i} \tag{2.2}$$

Equation 2.1 shows the applied force $\vec{F}_i$ which is the product of mass m_i and acceleration $\vec{a}_i$ of the particle. Taking the derivative of potential energy V of the system formulated in Equation 2.2 gives the applied force $\vec{F}_i$ where r_i represent the coordinates changing with respect to time t. The expression of potential energies in the equations $V\left(r_1, \ldots, r_N\right)$ shows that potential energy depends on the positions of N particles. Because of this system is a second-degree non-linear differential equation, requiring numerical solvation step by step with a suitable integration algorithm (González, 2011). For every time step of an MD simulation, momentum and new positions were calculated numerically using integration algorithms (M.P. Allen & Tildesley, 1989). The first derivative of the position with respect to time is the velocity v, and the second derivative is the acceleration a.

$$r(t + \Delta t) = r(t) + \Delta v(t) + \frac{1}{2}a(\Delta t)^2 t + \dots \qquad (2.3)$$

$$r(t - \Delta t) = r(t) - \Delta v(t) + \frac{1}{2}a(\Delta t)^2 t - \dots \qquad (2.4)$$

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + a(\Delta t)^2 t \qquad (2.5)$$

Nearly all integrators suppose that Taylor's series expansions in equations 1.3 and 1.4 approximate the velocity and acceleration by ignoring the rest of the terms after third term (Zheng et al., 2018). The summation of these equations gives the Verlet algorithm equation 1.5 computes the next step positions $r(t + \Delta t)$ (Verlet, 1967).

A microscopic level of information likes atomic positions and velocities are produced as a result of computer simulation. In order to translate this information into a macroscopic level like internal energy and pressure, statistical mechanics calculations are used (McQuarrie, 1975). Simulation produces possible ensemble of frames that represent the same macroscopic state, but it does not necessarily have to be the same microstates. A time-dependent ensemble called trajectory represents the evolution of a system with respect to time. The average of many replicas is calculated as an ensemble average that is equal to the time average in the system (Leach Andrew R., 2001). There are different types of statistical ensembles used in Molecular Dynamics simulations. Canonical ensemble (NVT) is the system with the fixed number of atoms N, volume V, and temperature T. Microcanonical ensemble (NVE) does not allow to change the number of particles N, volume V and the energy E of the system. In the isobaric-isothermal system (NPT), the number of atoms N, pressure P and temperature T are fixed (Zheng et al., 2018)

The biological system to be simulated should be first solvated and ionized. This provides biomolecules to experience their natural environment. The periodic boundary conditions are used to preserve the total number of particles in the system. A small part of the system, which is called a unit cell, is used to approximate an infinite

system. At the same time, use of periodic boundary conditions provides to neglect the solvent surface tension effect. The unit cell has an imaginary box in its environment represented in Figure 2.2, so the particle in the unit cell will not be lost if it goes out of the box; it will appear in the neighbour box (Leach Andrew R., 2001).
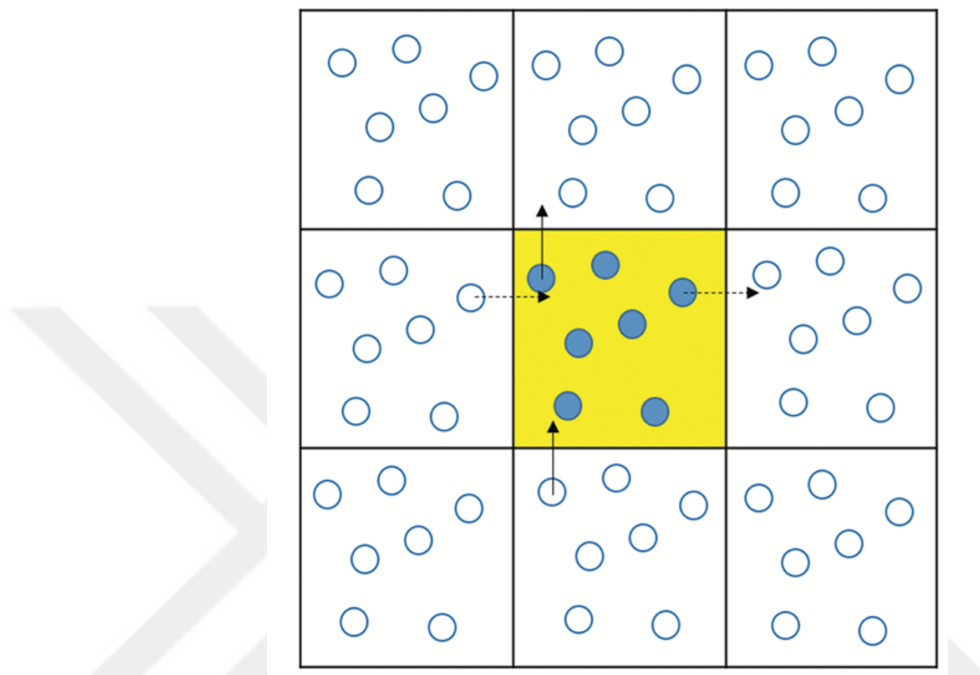
Figure 2.2 Periodic boundary Conditions (Katiyar & Jha, 2018)

In some conditions, intramolecular bonds are not represented in terms of potential energy functions; instead, they are treated like fixed by length. Restricting high vibration bonds can be used where it is correct to perform highly complex quantum calculations or to excite as if a ligand is attached to the protein, as in our case. The addition of some constraints requires control of the distance between atoms during the simulation. Their new positions and velocities are calculated after all the other atoms position and velocities of the system. Lagrangian or Hamiltonian formalisms specify constraints in classical mechanics. Instead of Verlet or Velocity Verlet algorithms, their new versions, which are suitable for constraints, are used to calculate new positions and velocities with respect to time; RATTLE and SHAKE (Michael P. Allen, 2004). In a constraint simulation on NAMD, the harmonic potentials $U = k\left(x - x_{\text{ref}}\right)^2$ are applied to hold specified molecules immobile. The computational cycle consists of calculating forces and acceleration from the position

at time t, then producing and updating a new position by applying Verlet algorithm, finally recording energy values to a trajectory file from statistical mechanics. Visualizing the atomic vibrations requires to use small time step size, commonly used as 2 femtoseconds. Increasing the time step size may result in missing events occurring on a small scale while using small time steps will increase computational effort.

2.2 System Preparation

For MD simulations, the crystal structure of *S. aureus* pyruvate kinase was extracted from Protein Data Bank (Berman et al., 2002) with PDB id 3T0T. It is bound to the allosteric inhibitor IS-130 molecule, and its protein structure used for all nine simulations (Axerio-Cilies et al., 2012). It has been run for 3 different systems for apo, constrained-1 for proposed region and constrained-2 for known region, and each of them consists of 3 sets in itself. In the continuation of this study, Constrained-1 will be referred to as Constr-1 and Constrained-2 will be mentioned with the abbreviation Constr-2.

During system preparation, the crystal structure of *S. aureus* PK was cleaned from its crystallographic water and ligand molecules and was solvated in a water box with 12.5 Å x 12.5 Å x12.5 Å dimensions consisting of TIP3P water and then neutralized by adding Na+ and Cl- ions as illustrated in Figure 2.3. This system preparation and visualisation processes were performed by Psfgen plugin of VMD (Visual Molecular dynamics) graphics visualization tool (Humphrey et al., 1996).

Ayyıldız et al. worked on possible allosteric regions of glycolytic enzymes including pyruvate kinase on different organisms. They propose novel allosteric binding regions and found specific residues important for allosteric signalling in those regions (Ayyildiz et al., 2020). In their research, which is also a species-specific drug design study, the locations of potential allosteric sites were determined using solvent mapping and elastic network analysis method which is represented in Figure 2.4. Using AlloSigMA tool (Guarnera et al., 2017) was revealed how perturbation formed by
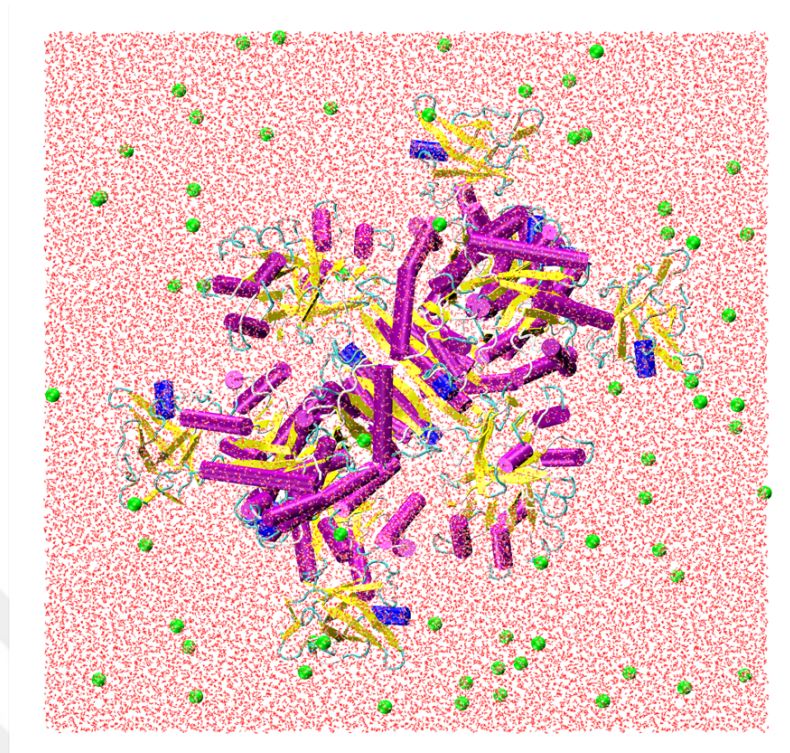
Figure 2.3 Pyruvate kinase in a water box

ligand binding to the allosteric region effected the dynamics of the catalytic region of the protein. The differences caused by the proposed allosteric region and the known allosteric region where the known allosteric inhibitor IS-130 was bound were compared. As illustrated in Figure 2.5, known allosteric region were placed near the small interface. Accordingly, when a ligand binds to the proposed site, there was a decrease in the dynamics of the four sites to which PEP would bind, while an increase in dynamics was observed for ADP. The position of the proposed allosteric site and the variation of the dynamics in the binding site of the substrates are shown in the figure. They also demonstrated that this proposed region has much stronger allosteric properties than the known region, by observing an increase of 0.78 ± 0.21 in ΔG (Ayyildiz et al., 2020).

The first constrained simulation (constr-1) was for the proposed allosteric region. Ligand molecules were detected by a docking study of created library from ZINC15 (Sterling & Irwin, 2015) database (Özhelvacı, 2020). Firstly, all residues interacting with candidate drug molecule were determined as the most critical residues for the
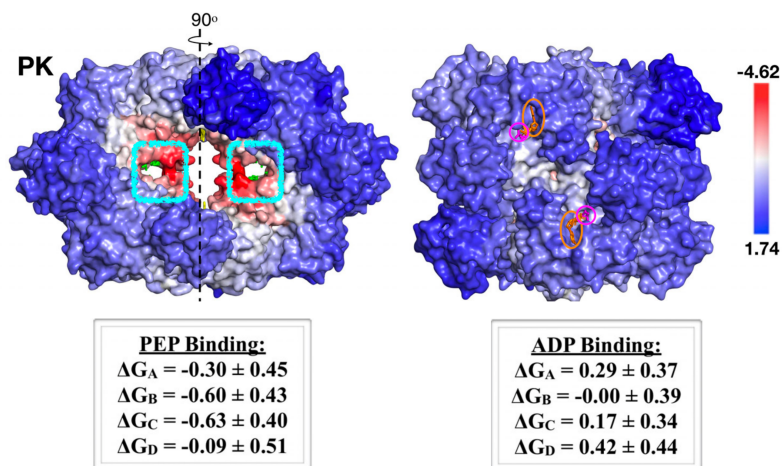
Figure 2.4 Proposed allosteric region and catalytic regions, red and blue colours shows decreased and increased dynamics (Ayyildiz et al., 2020).

protein-ligand interaction. Among these residues, a couple of them were selected to be constrained in order to stabilize the binding pocket like a cage. On the other hand, for the known allosteric site, residues interacting with IS-130 were determined and selected for constraints. The distance between selected residues was measured and used as constrained values. These distances will slightly fluctuate by 0.1 Å around their constrained values during the simulation. The pair of residues were listed in Table 2.1, and the localisation is represented in Figure 2.5.
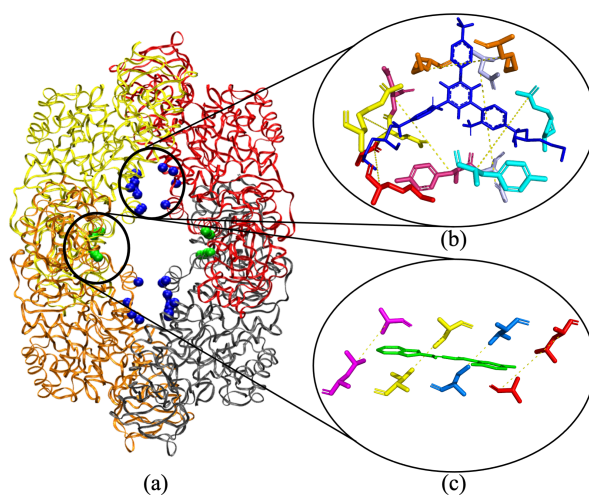
Figure 2.5 a) Constrained allosteric regions on pyruvate kinase structure. Blue and green regions are proposed and known allosteric regions, respectively. Picked residues shown on the binding pocket for b) proposed region and a ligand molecule (ZINC11677890) represented with blue color , and c) known allosteric regions and IS-130 is represented with green color. Coupled residues shown with same colour.

Table 2.1 Restricted residue pairs and their distances for constrained MD simulations

Constrained Allosteric Site	Selected Residue Pairs	Distance (Å)
Proposed Allosteric Site (*Constrained -1*)	LYS260A – ALA337D	9.80
	ARG264A – GLN338D	12.08
	ILE302A – ARG264D	15.06
	LYS342A – LYS342D	11.0
	ILE303D – LYS346D	11.08
	ILE302D – LYS339D	10.04
	LYS260C –ALA337B	9.70
	ARG264C– GLN338B	12.06
	ILE302C – ARG264B	15.07
	LYS342C – LYS342B	10.09
	ILE303B – LYS346B	11.80
	ILE302B – LYS339B	10.04
Known Allosteric Site (*Constrained -2*)	ALA358A – LEU370B	8.22
	SER362A – THR366B	7.93
	THR366A – SER362B	7.91
	LEU370A – ALA358B	8.13
	ALA358C – LEU370D	8.18
	SER362C – THR366D	7.91
	THR366C – SER362D	7.92
	LEU370C – ALA358D	8.14

After the system preparation, the simulation conditions were specified on the configuration file. MD simulations were performed on NAMD (Nelson et al., 1995) with the CHARMM36m forcefield (Huang & Mackerell, 2013). The NPT ensemble conditions are used, which are the constant number of particles (N), constant pressure (P) and constant temperature (T). Langevin Dynamics was used to stabilize the pressure at 1 atm and temperature at 310 K. Periodic boundary conditions are used in each simulation, with the box size 163, 163,120 Å in each direction. All the simulations are minimized 1000 steps with conjugate gradient and line search algorithm. 2 fs is used as a time step, and the simulations length was 100 ns. Energy values and positions were collected at every 10000 steps which produced a total of 5000 frames.

3. RESULTS AND DISCUSSION

3.1 Energy Stabilization and Constrained Residues During Simulation

A simulation's energy profile helps to figure out if the system has reached the equilibrium state. Stable energy profile with minor fluctuations at the end of the simulations show that the equilibration is sufficient for the system. The potential and the kinetic energy profile of nine distinct simulations were plotted in Figure 3.1. Potential energy values were represented with blue colour, while kinetic energy represented with red. The system stabilized within 5 ns at the very early stages of the simulation.
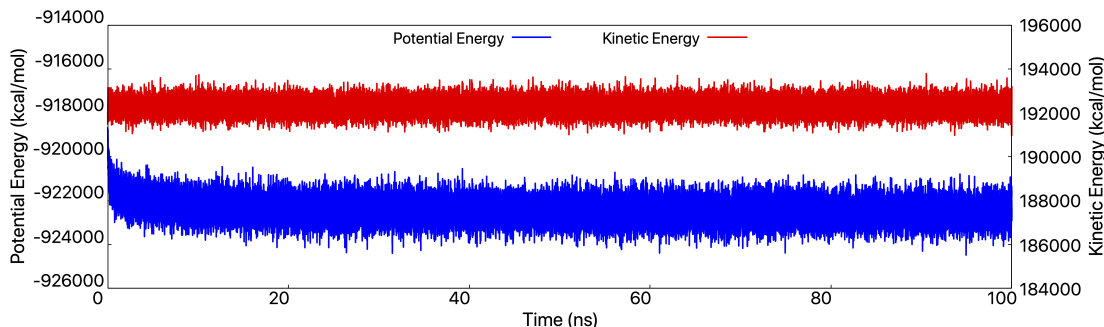
Figure 3.1 Potential energy and Kinetic energy change of 9 different simulations (5000 energy points for each simulation).

Distances between residues were defined as fixed before the simulation for the constrained sets of simulation. As illustrated in Figure 3.2, distance values in 12 pairs of residues for constr-1 and 10 pair of residues for constr-2 compared to apo runs were monitored throughout the trajectory. Residue distances in the apo states fluctuated approximately 1-2 Å and only 0.1 Å in the constrained states, as we expected.
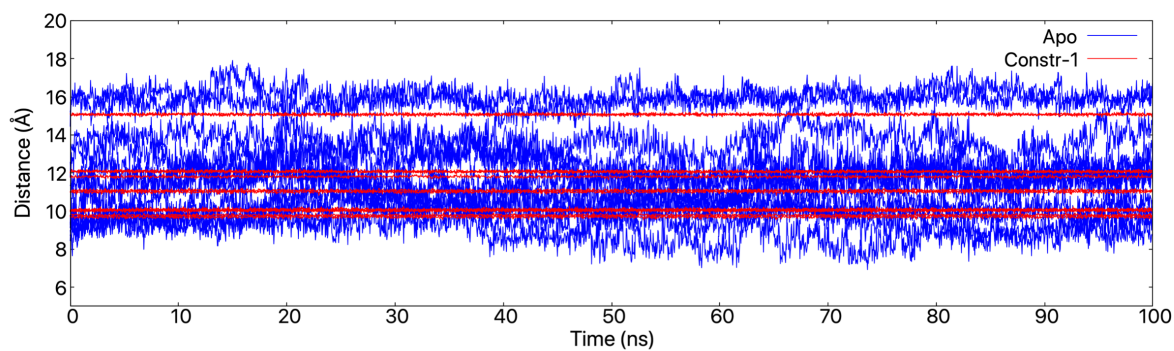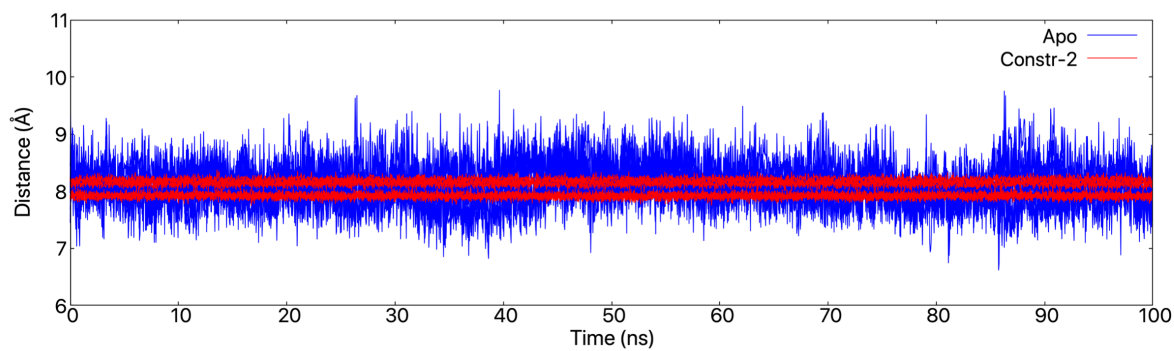
Figure 3.2 Fluctuations of the distance between pairs of residues in apo and (a) constr-1 and (b) constr-2 runs (see text for details)

3.2 RMSD and RMSF

Root mean the square root of deviation (RMSD) is used to measure the similarity between two imposed atomic coordinates. In this study, it was used to monitor the deviation of the conformation with respect to its initial state. RMSD of a molecule is calculated by equation 3.1 for n pairs of particles where $r_i(t)$ is the position at time t, and $r_i(0)$ is the initial position of the ith particle.

$$RMSD\,(r_i) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}||r_i(t) - r_i(0)||^2} \tag{3.1}$$

$$= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(r_{ix}(t) - r_{ix}(0)\right)^2 + \left(r_{iy}(t) - r_{iy}(0)\right)^2 + \left(r_{ix}(t) - r_{ix}(0)\right)^2}$$

RMSD calculation, which is frequently used in the analysis of MD simulations, allows us to understand how much the protein structure has changed during the simulation and whether it is equilibrated. It also provides to inspect changes only in a desired region of the system, e.g., it is possible to see the difference of each residue or a specific chain or domain in the protein throughout the trajectory and compare them with each other(Kufareva & Abagyan, 2012).

All RMSD profiles of each set of three runs were plotted as in Figure 3.3. At the end of the simulation, all curves reached a plateau at around 4-6 RMSD values. Only in one apo run (after 20 ns) and one constr-1 run (after 75 ns), there exist slight increases of about 1-2 which was due to highly mobile B and CT domains. On the other hand, both RMSD and RMSF results indicate that A and C domains were more stable. Integration of highly mobile domains to RMSD calculation for whole tetramer increases the RMSD values of less mobile domains. Therefore the tetramer aligned to the more stable regions which are A and C domain.
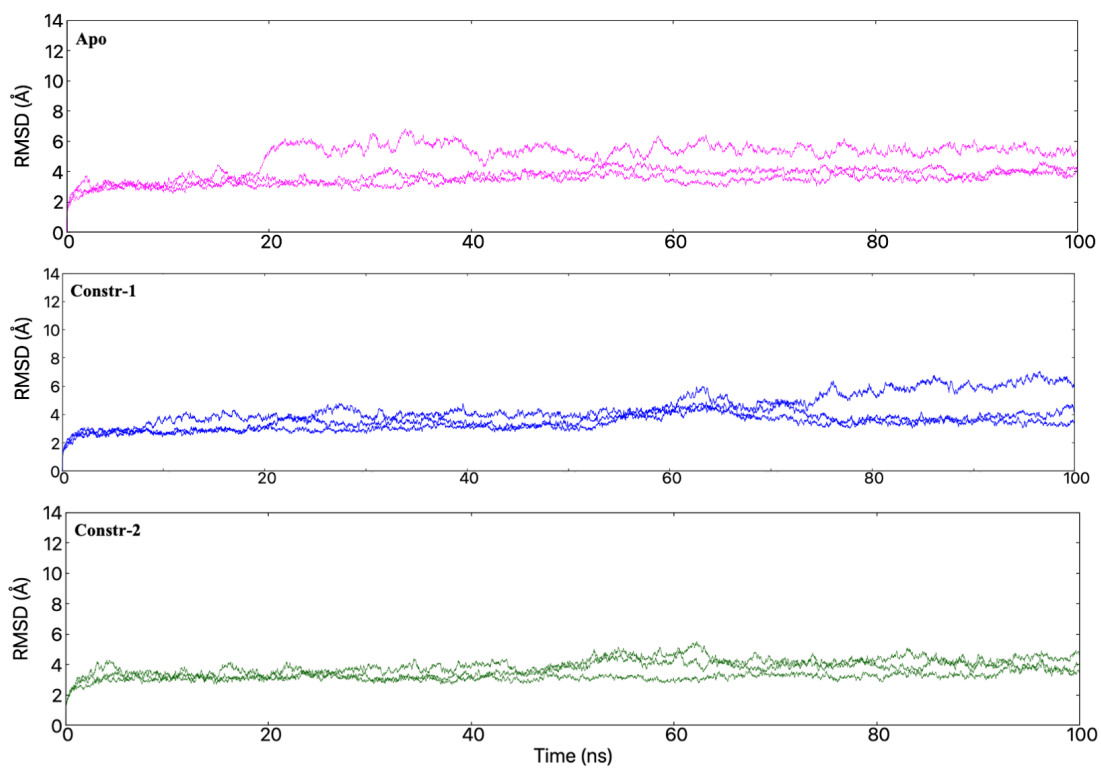
Figure 3.3 RMSD plots of apo, constr-1 and constr-2 simulations of three-set runs represented with pink, blue, and green colours.

In Figure 3.4, each chain was aligned to its less mobile AC domain. The entire tetramer structure was also aligned to all four AC domains. In these graphs, each chain and its tetramer structure were shown in different colours. In this way, one can see which chain increases the RMSD value of the tetramer. For example, in apo-run#2 (Figure 3.4a), fluctuations in D chain increased the RMSD values of all tetramers. Again, the D chain was responsible for increasing the RMSD values of the tetramer in the first run of the constrained-1 simulation. The RMSD values of the majority of the chains vary between 3-6 Å.

Slightly higher RMSD values for each chain were obtained when the structural alignment was made with respect to all four AC domains as opposed to alignment to its chain as illustrated in Figures 3.5, 3.6 and 3.7. The slightly higher values clearly indicate the slight displacement of the chains with respect to each other in the tetrameric structure. The RMSD values calculated with alignment to all four AC domains are expected to be higher because they also include the displacements of the chains relative to each other. As illustrated in Figure 3.5 a, displacement of chain A in the tetramer responsible for the insignificantly increased RMSD of tetramer (3.4 a). In the second run of apo state, displacement of the domain D caused the rise in the tetramer RMSD. Regarding Figure 3.5b, the difference of two RMSD from two alignment state showed that the chain D changing its position in the tetramer. For constr-1-run#1, fluctuation of chain D increased the RMSD of the tetramer similarly because it was independently displaced from the tetramer after 50 ns (3.6a). In the second run of the constr-1 state, the RMSD of chain A inappreciably higher than other chains and tetramer. As illustrated in Figure 3.6b, there was not too much difference between the two-alignment state. The movement of chain C generated higher fluctuations than other chains and tetramer in the constr-2-run#2 (3.4c). According to Figure 3.7b, the displacement of chain C was independent of its tetramer. These independent motions from tetramer might occur from chains independence or the mobile regions inside them.
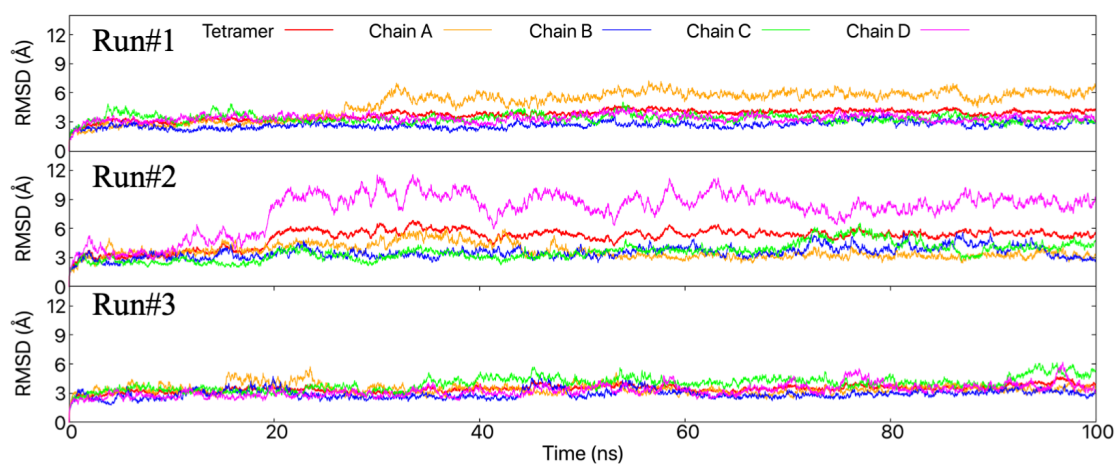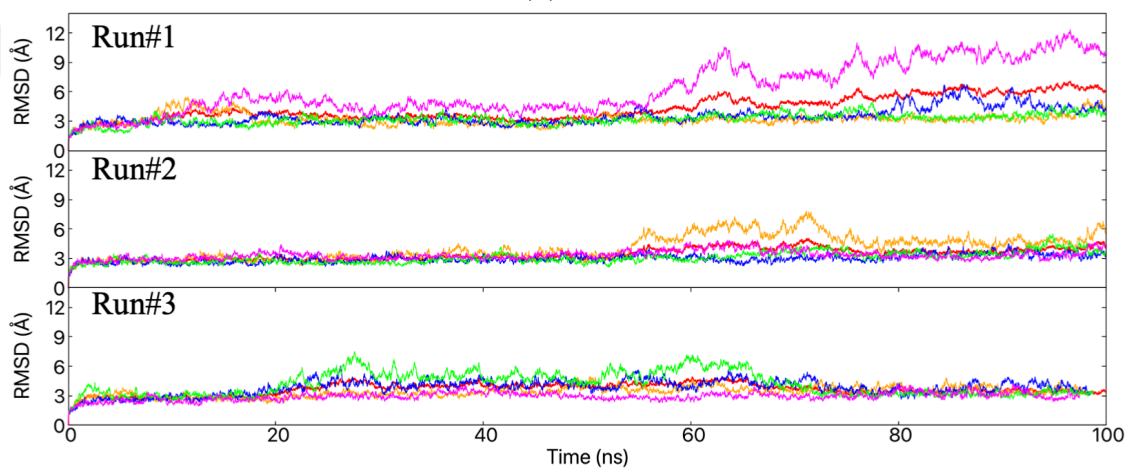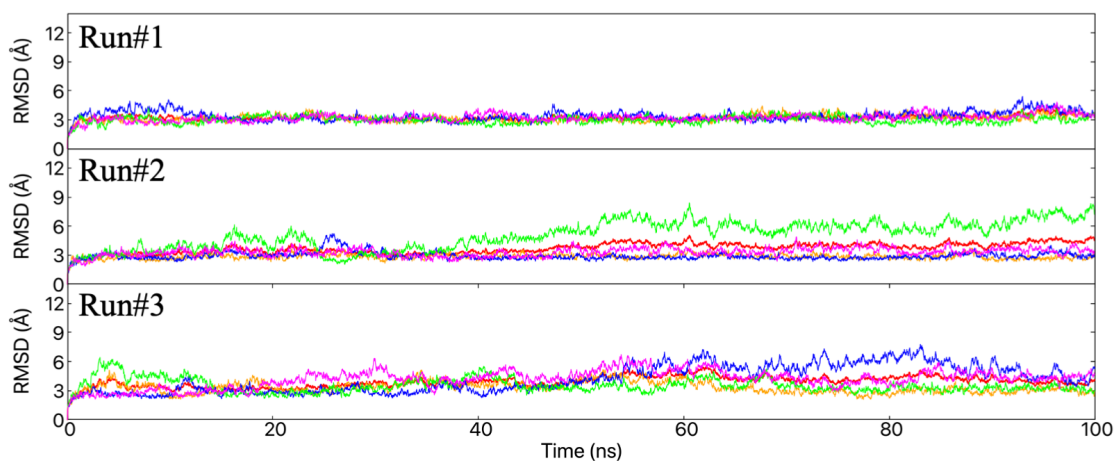
Figure 3.4 RMSD plot of individual chains for each simulation set represented, a) apo b) constr-1 c) constr-1. Each chain and tetramer represented with a colour in the legend box of the plot.

(a)

(b)

(c)
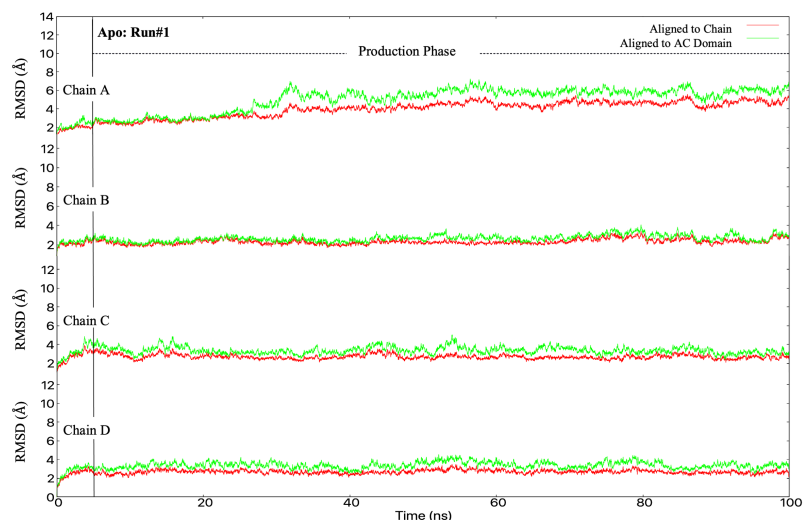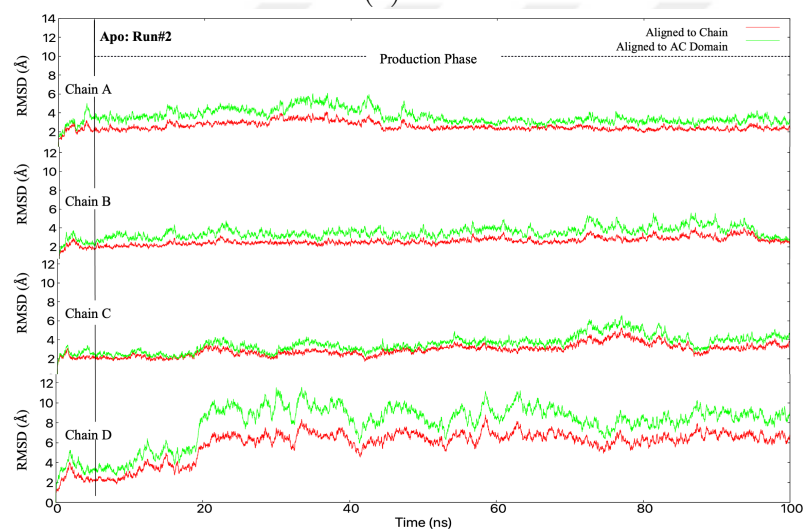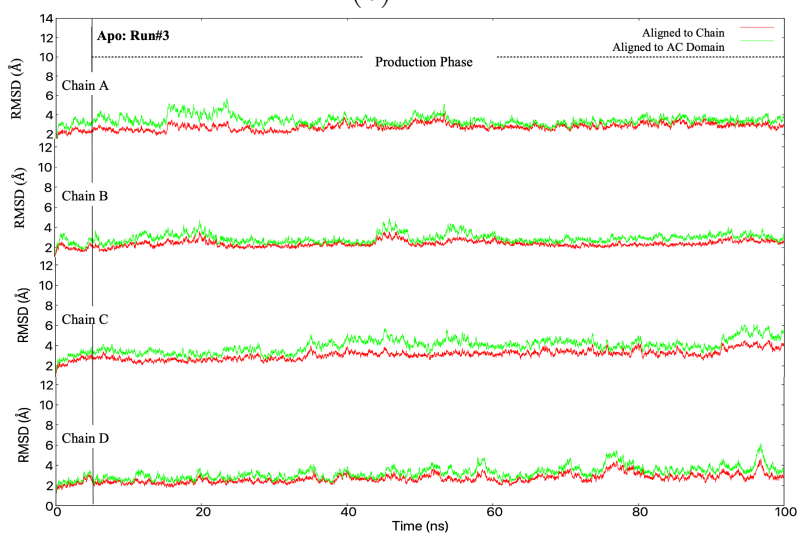
Figure 3.5 RMDS profile of Apo simulations according to two different alignment states (See text for details).

(a)

(b)

(c)
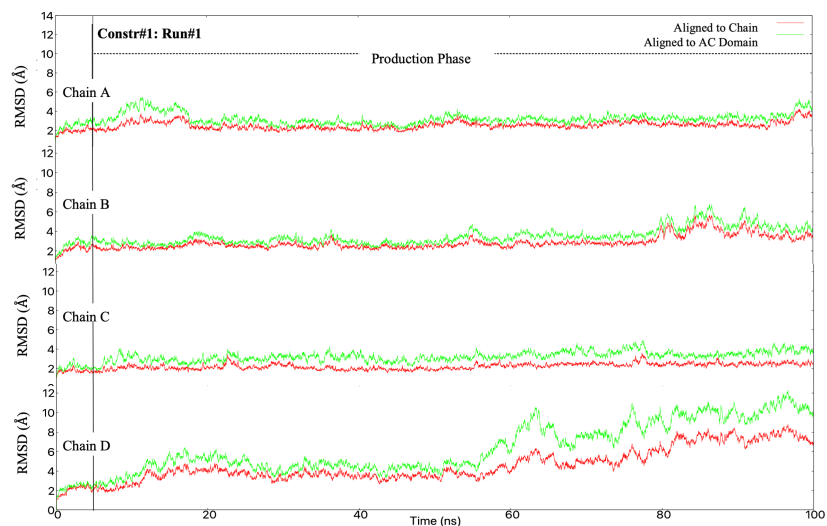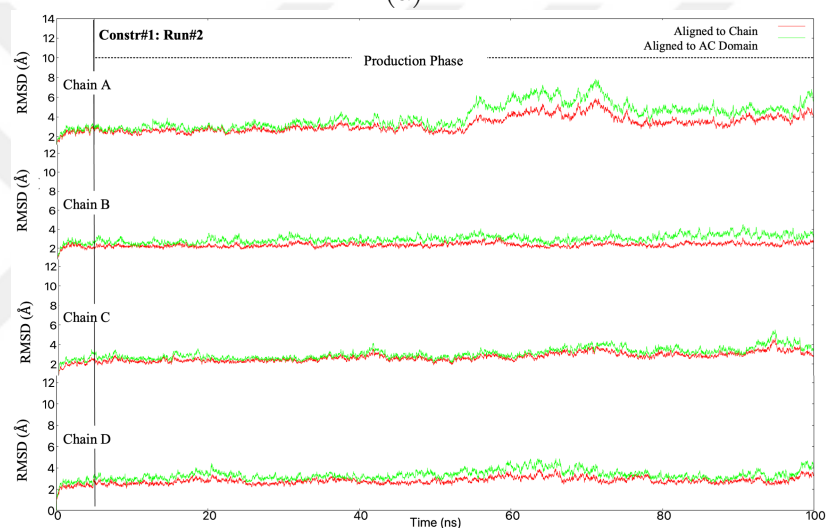
Figure 3.6 RMDS profile of constr-1 simulations according to two different alignment states (See text for details).

(a)

(b)

(c)
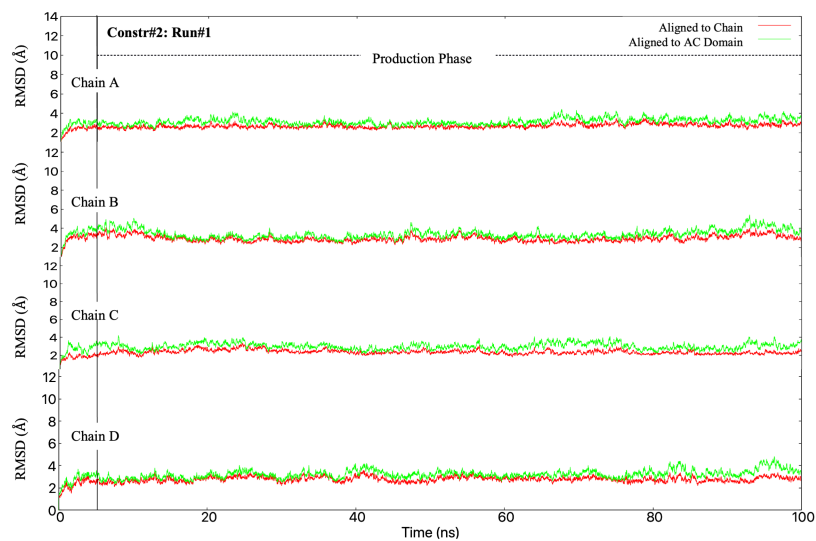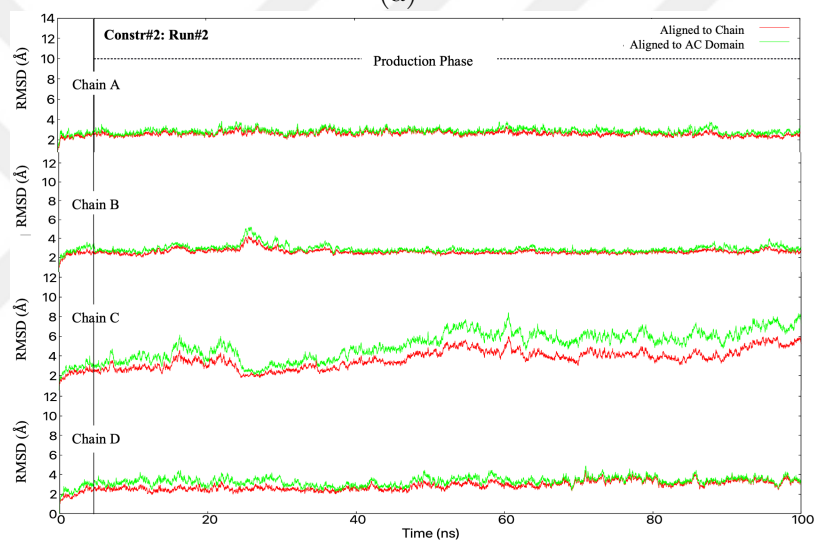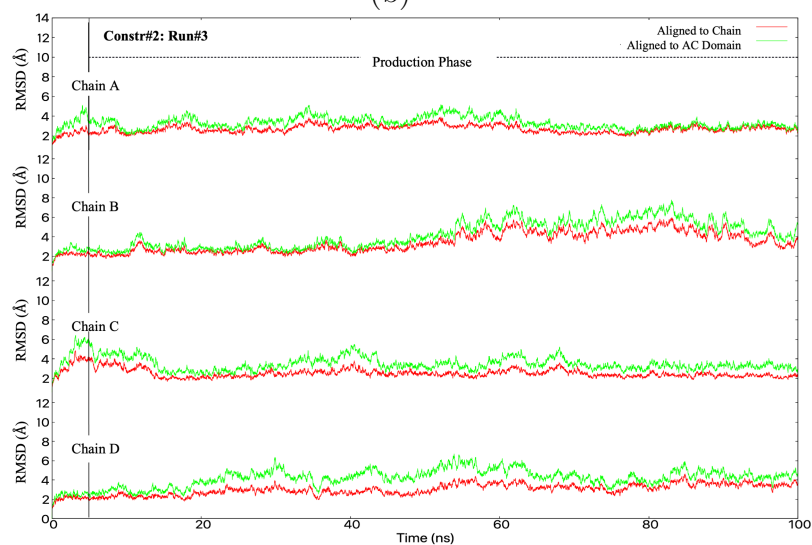
Figure 3.7 RMDS profile of constr-2 simulations according to two different alignment states (See text for details).

Mean Square Fluctuation gives the deviation of the current position $r_i(t)$ of the particle at time t from its average position $\bar{r}_i$ and is defined by equation 3.2. It gives information about how its flexibility changes regionally in MD simulation.

$$RMSF_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}|r_i(t) - \bar{r}_i|^2}\tag{3.2}$$

RMSF profiles of all simulations were calculated to detect the mobile regions by aligning the tetramer to AC Domains. Most mobiles parts indicated the B and C domains in almost all chains in Figure 3.8. Clearly, the mobility of these parts led to an increase in the majority of RMSD plots. In apo-run#2 simulation, CT domain of the D chain shows high RMSF values.
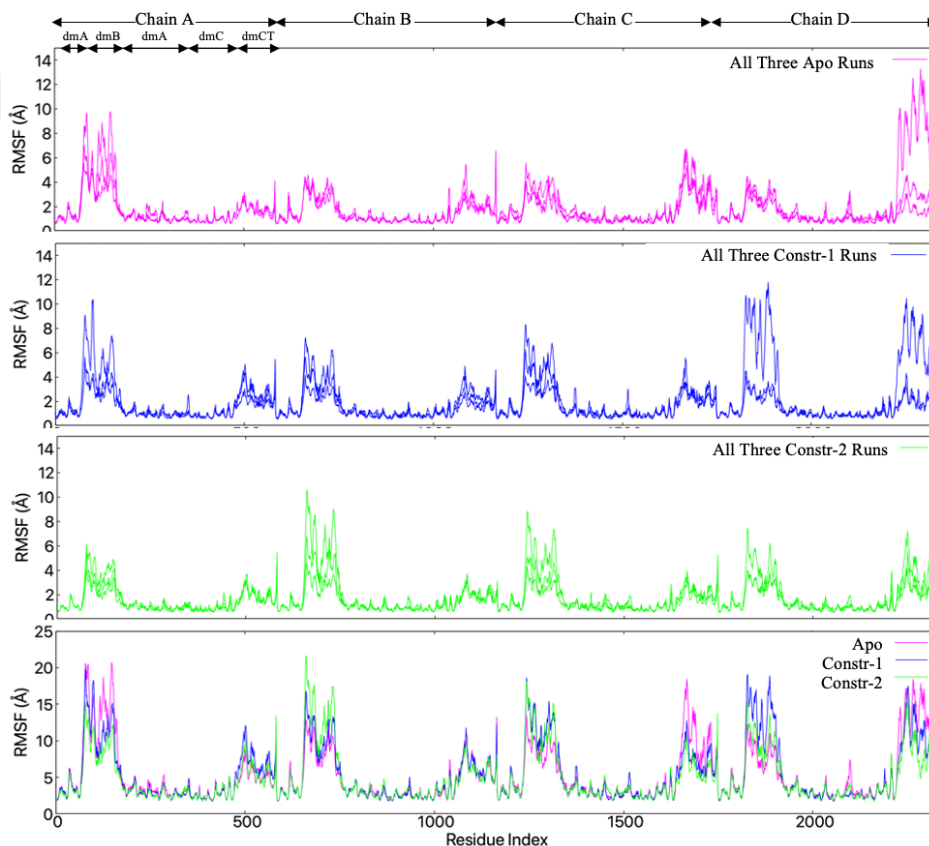
Figure 3.8 RMSF profile for three sets of simulations and sum of all three runs in each state

RMSF values of three runs for each set of simulations were summed. Then these sum of RMSF values for constrained simulations were subtracted from sum of apo RMSF values to understand their difference in the figure 3.9. The location of chains and domains represented on the plot while substrate interacted residues highlighted with red and blue dots on a and b. According to the figure, the RMSF values in the apo state of B and CT domains increased while in the constrained state. The most affected parts were B and CT domains from the restricted residues. RMSF differences decreased in the B domain of chain A and CT domain of chain C and D. This means that the RMSF values of these regions decreased as their mobility decreased when they were constrained. On the other hand, there were increases in the CT domain of Chain A and B domain of B, C, D chains which means that the mobility of these domains increased in the constrained state. These differences were observed in the same regions for both comparison and also represented in the protein structure. Red regions represent increased RMSF differences and blue regions for decreased differences. Among the 4 B domains, 3 were becoming more mobile with the effects of the constrained residue. It is known that the B domain loses its flexibility by closing to the A domain catalytic site in the active R state of the protein. Because those with constrained states showed higher RMSF results than those with apo states, they showed that the pyruvate kinase protein was far from the active R conformation and was closer to inactive structure dynamics in the constrained simulations.

RMSF difference of constrained residues between constr-1 and apo, constr-2 and apo were represented in Figure 3.10. RMSF values were lower than zero in both comparison, which means the constrained residues were more mobile in the apo states than in constrained states as expected.

27
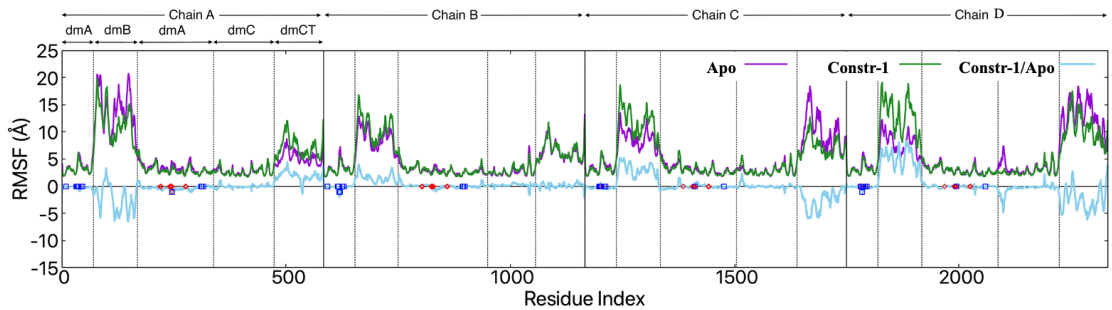
(a)
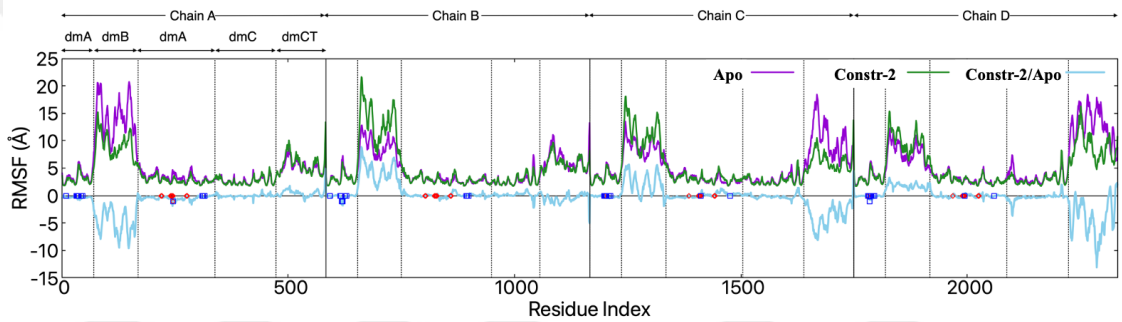
(b)
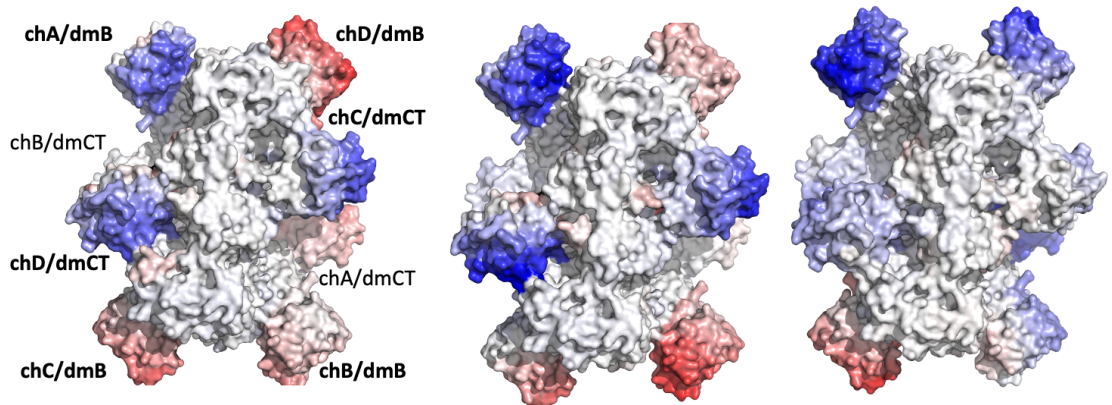
(c)

Figure 3.9 Summation of RMSF values for three runs and their differences between a) constr-1 and apo , b) constr-2 and apo. c) Colourful representation of these differences on pyruvate kinase constr-1and apo, constr-2 and apo, constr2 and constr1 respectively
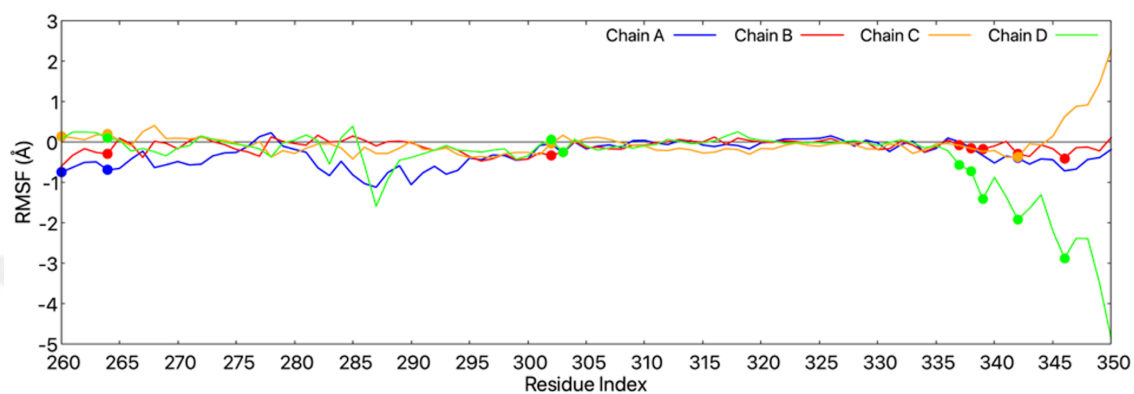
(a)
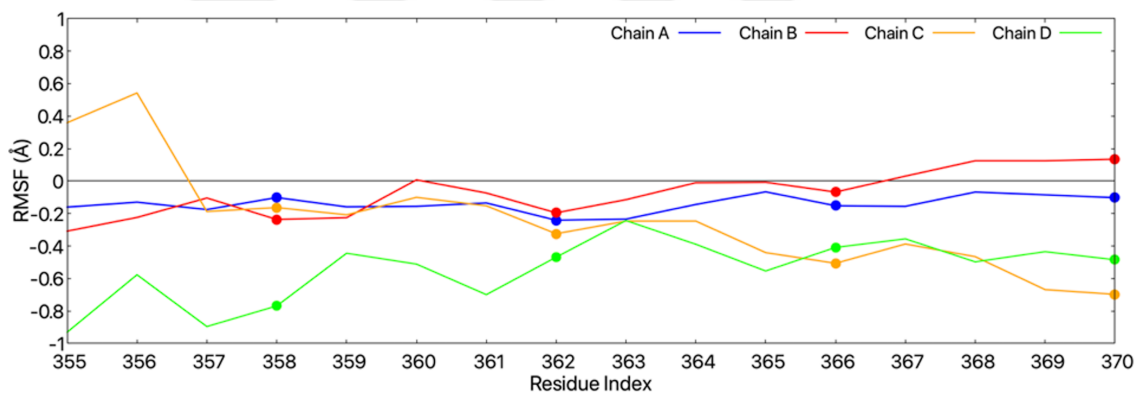
(b)

Figure 3.10 Summation of RMSF values for three runs and their differences between a) constr-1 and apo , b) constr-2 and apo. c) Colourful representation of these differences on pyruvate kinase constr-1and apo, constr-2 and apo, constr2 and constr1 respectively

29

3.3 Principal Component Analysis (PCA)

Principal Component Analysis is a way to reveal the dominant motion of trajectories in a Molecular Dynamics simulation. It explains essential structural transitions with a small number of variables by reducing phase space in long simulations (Amadei et al., 1993). A translation to the average geometrical center of the molecule and a least-squares fit alignment onto a reference structure eliminates the trajectory's overall translational and rotational motions (Grubmuller, 1995). Linear transformation in Cartesian coordinate space that generates a 3Nx3N covariance matrix provides reconstructing the configuration space (Kamberaj, 2016).

The direction of the motion is indicated by the vectorial description of each component of motion generated from diagonalization of the covariance matrix. The energetic contribution of a specific component of the motion is represented by the eigenvalue that made the eigenvector as a motion describer. Time-dependent motions that the components perform in the particular vibrational mode are highlighted by the projection of the trajectory on a certain eigenvector(Haider et al., 2008).

In this study, while performing PCA, the Prody Tool (Bakan et al., 2011) was used to calculate the fraction of the modes and their cumulative contributions that explain a certain percentage of the motion were computed. Trajectories of pyruvate kinase structure, without equilibration phase was aligned to the more stable AC domains for this analysis. Figure 3.11 represents the cumulative variance in terms of percentage with respect to modes for three different states. According to the Figure 3.11, the first 20 modes of the apo state explained 85-90 % of the total motion. On the other hand, the first 20 modes contributed 80-90 % of the total motion in the constrained states.

In Table 3.1, cumulative percentage of explained motion for each mode of all simulation states clarified. 90 % of the total motion was explained with the first 38, 19 and 38 modes in three apo simulations, the first 19, 39, 39 modes in three constr-1 simulations and the first 60, 43, 24 modes in three constr-2 runs. Because the

differences of the apo and constrained states did not clear too much, the average mode number explains 90 %of the total motion was compared. The average mode number that explained the 90 % of the entire motion for three apo runs was 31. This average mode number increased to 32 and 42 for the three runs of constr-1 and constr-2 simulations
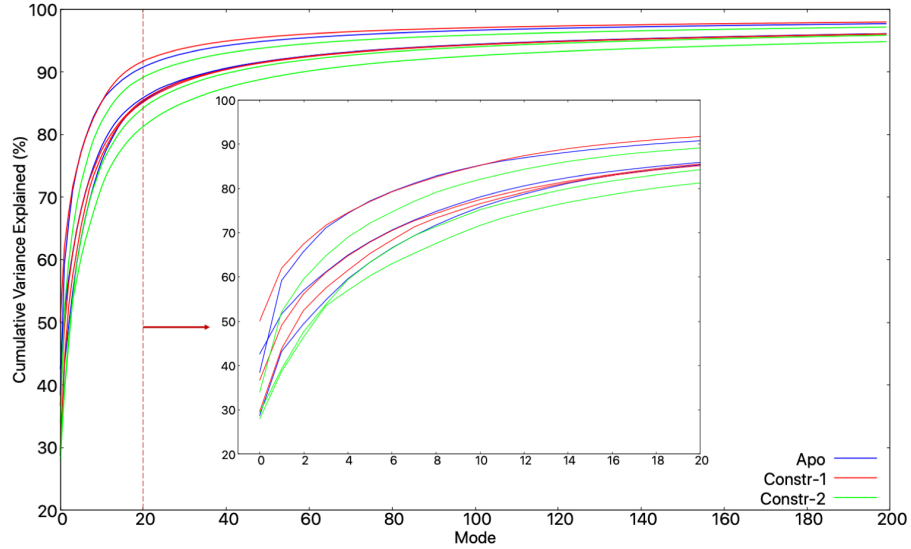
Figure 3.11 Principal component analysis that explain the percentage contributions to the total motion for three sets of simulation

In Figure 3.12 the first three mode projections were illustrated for 5000 frames. For this analysis Normal mode wizard package (Bakan & Bahar, 2009) in VMD tool was used. Each frame on trajectory was represented as a dot on the graph. Intervals of the apo runs were PCA#1:[-1:6], PCA#2:[-6:2], PCA#3:[-2:2]. Constr-1 runs mode intervals were PCA#1:[-6:1], PCA#2:[-3:4], PCA#3:[-3:3] while the points of the constr-2 runs fell PCA#1:[-4:4], PCA#2:[-4:1], PCA#3:[-2:3]. Comparing PCA#1, PCA#2 and PCA#3 of apo and constr-2 states were close to each other while constr-1 differs from them by existing in different intervals. In the 1st and 2nd modes, run2 (colored in light blue) showed a wider spread in the apo simulations, unlike the other two simulations. Similarly, run1 simulation (colored in brown) of constr-2 showed a different spread than the other two runs. On the other hand, three runs of constr-1 showed similar distributions.

In Figure 3.13, representation of the first principal component from one run of each

Table 3.1 Individual and cumulative % contributions of modes for each simulation

APO								
Run #1			Run #2			Run #3		
Mode	% Explained	Cumulative % Explained	Mode	% Explained	Cumulative % Explained	Mode	% Explained	Cumulative % Explained
1	42.6	42.6	1	38.4	38.4	1	28.7	28.7
2	9.0	51.6	2	20.7	59.2	2	14.5	43.2
3	5.3	56.9	3	6.5	65.8	3	6.2	49.4
4	4.1	60.1	4	5.3	71.1	4	5.3	54.8
5	3.8	64.0	5	3.3	74.4	5	4.8	59.6
7	2.6	70.6	8	1.8	80.1	9	2.4	71.7
13	1.1	80.6	19	1.8	90.1	15	1.2	81.2
38	0.2	90.1				38	0.2	90.1
CONSTR-1								
Run #1			Run #2			Run #3		
Mode	% Explained	Cumulative % Explained	Mode	% Explained	Cumulative % Explained	Mode	% Explained	Cumulative % Explained
1	50.0	50.0	1	36.7	36.7	1	29.6	29.6
2	11.9	61.9	2	12.3	49.1	2	14.1	43.8
3	5.5	67.5	3	7.1	56.2	3	8.6	52.5
4	4.1	71.7	4	4.7	61.0	4	5.0	57.5
5	2.8	74.5	5	3.7	64.7	5	3.9	61.5
8	1.7	80.9	7	2.6	70.5	8	2.9	71.2
19	0.5	90.1	14	0.9	80.7	14	1.1	80.3
			39	0.1	90.1	39	0.1	90.0
CONSTR-2								
Run #1			Run #2			Run #3		
Mode	% Explained	Cumulative % Explained	Mode	% Explained	Cumulative % Explained	Mode	% Explained	Cumulative % Explained
1	27.9	27.9	1	29.3	29.3	1	34.0	34.0
2	10.6	38.6	2	9.8	39.3	2	18.1	52.1
3	7.8	46.4	3	8.4	47.6	3	7.3	59.5
4	6.9	53.4	4	6.1	53.7	4	5.2	64.8
5	3.5	56.9	5	5.6	59.4	5	4.1	68.9
11	1.9	71.6	9	2.0	71.3	6	3.1	72.1
19	0.7	80.1	16	0.9	80.9	10	1.5	80.7
60	0.1	90.0	43	0.1	90.0	24	0.2	90.0
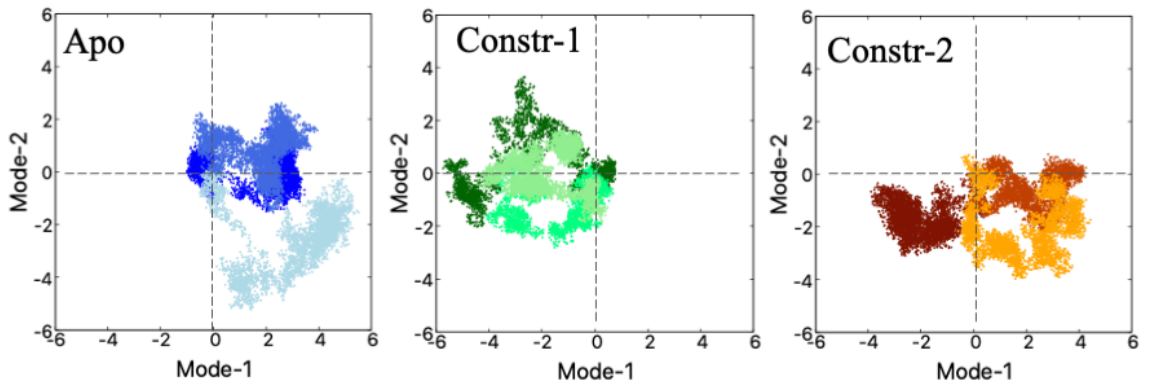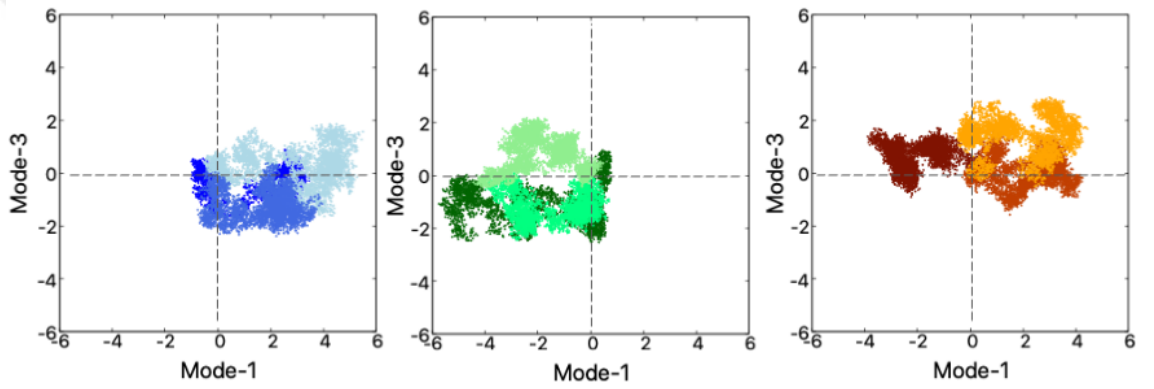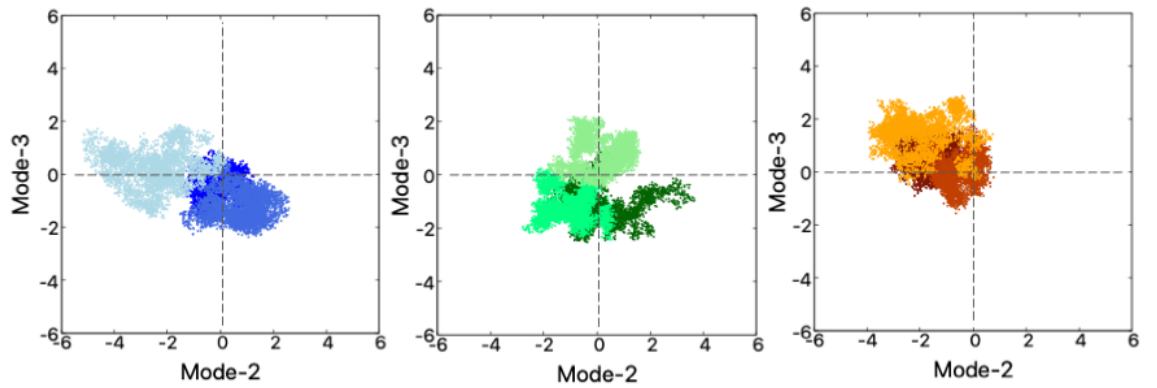
Figure 3.12 Principal component analysis. Projections of (a) mode 1-2, (b) mode 1-3, (c) mode 2-3 of apo, constr-1 and constr-2 runs which are colored as blue and shades, green and shades, orange and shades, respectively.
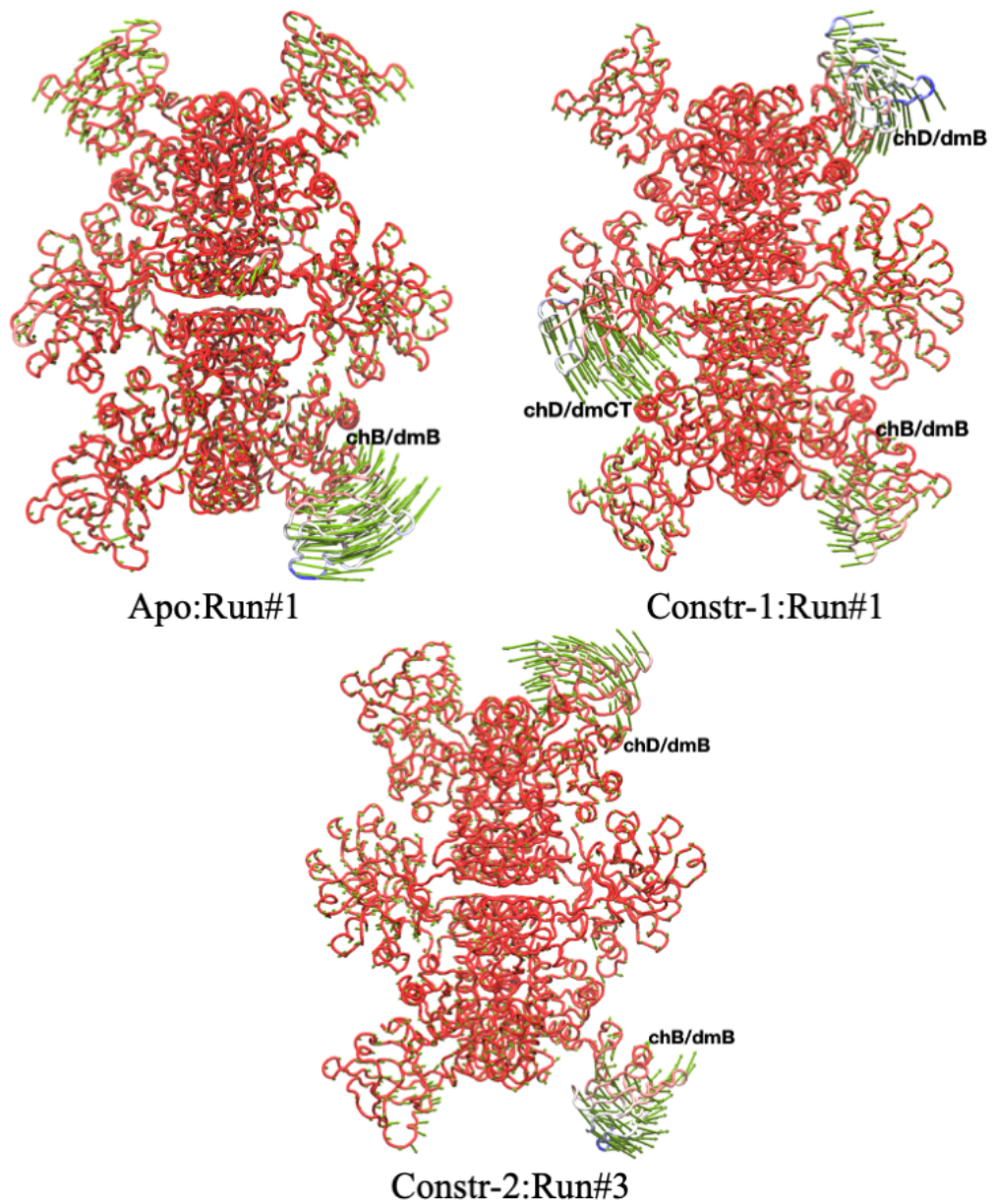
Figure 3.13 Directions of the first principal component that contributes the most to the total motion for three different sets of simulation

set of simulation was illustrated. In these modes, most mobile parts were identified as domain B of all simulations. For the apo state, domain B of chain B (abbreviated as chB/dmB) was the most mobile part. Also, all neighbouring B domains were fluctuating to opposite directions in the first mode. B domain of chain C and D (chD/dmB and chB/dmB) and CT domain of chain D (chD/dmCT) made large oscillation in constr-1 state. Neighbouring B domains on top of the structure were oscillating in the opposite directions while domains in the bottom fluctuates in the same direction. At the same time, all of the neighboring CT domains were fluctuating in the opposite directions. In the situation of constr-2, B domains of chain C and D (chD/dmB and chB/dmB) made large fluctuations. Upper B domains vibrated in the opposite direction while domains in the bottom fluctuated in the same direction. Projection of the modes and oscillation profiles shows that, constr-2 and apo simulation were similar to each other than constr-1 and apo.

As a result of principal component analysis, the simulation trajectory was projected by using components that would explain 70 % of the motion, and then the correlation values between the positional fluctuations of every pair of residues were calculated over this trajectory. Equation 3.3 was used to calculate the cross-correlation values for atoms i and j. Time averaged values represented with brace-enclosed quantities and positional vectors of atoms were represented as r_i and r_j.

$$C_{i,j} = \frac{\langle (r_i - \langle r_i \rangle) \cdot (r_j - \langle r_j \rangle) \rangle}{\sqrt{\left(\langle r_i^2 \rangle - \langle r_i \rangle^2 \right) \left(\langle r_j^2 \rangle - \langle r_j \rangle^2 \right)}} \tag{3.3}$$

Accordingly, these values varying between -1 and 1 indicate that the fluctuations are in the same (correlated) or opposite direction (anticorrelated), respectively, and values close to zero indicate that there is no correlation between them (Bowerman & Wereszczynski, 2016). As shown in Figure 3.14, the correlation maps in the first column were obtained for three different simulations in apo form. The map in the last row is the sum of all correlation values from three runs. Similarly, the maps of

the constrained-1 and constrained-2 simulations were given in the second and third columns. Accordingly, there was a significant decrease in both positive and negative correlations on constrained simulations. By taking the difference between the sum of correlations, the variation was shown for two different constrained state from apo state as in Figure 3.15

Difference of correlation differences between apo and constrained simulations were investigated from closer window. The regions of the chains in Figure 3.14 were separated as in Figures 3.16, 3.17, 3.18 and 3.19.

According to Figure 3.16, the positive correlation between domains of the same chain mostly decreased on the proposed region while correlations between A and B domain of A chain and C and CT domain of chain C increased. The positive correlations between A and B domains of chain A became stronger in the constr-1 state. Similarly, fluctuation rate of C and CT domains has increased.

Comparison of the correlations between chains were illustrated on Figure 3.17 indicates that proposed region restriction caused decrease in most of the correlations. Among 18 correlations (either positive or negative) in the apo form, 10 of them were turned in to zero correlation in the constr-1 simulations. On the other hand, the negative correlations between A domain of chain A and C domain of chain B strengthened in the constrained simulation of proposed region. In addition to this, same domains in the C and D chains increased their fluctuation rate in the opposite direction. At the same time, there were two positive correlation and two negative correlations created between the A domain and CT domain of different chains appeared in constr-1 state. Four out of five newly created correlations incorporated Ct domain.

Similar to constr-1 correlation results, intrachain correlations of known region mostly decreased except chain A, as illustrated in Figure 3.18. There was no significant correlation difference within D chain. As an exception, A and B domains of chain A started to fluctuate more in the same direction in constr-2 runs.
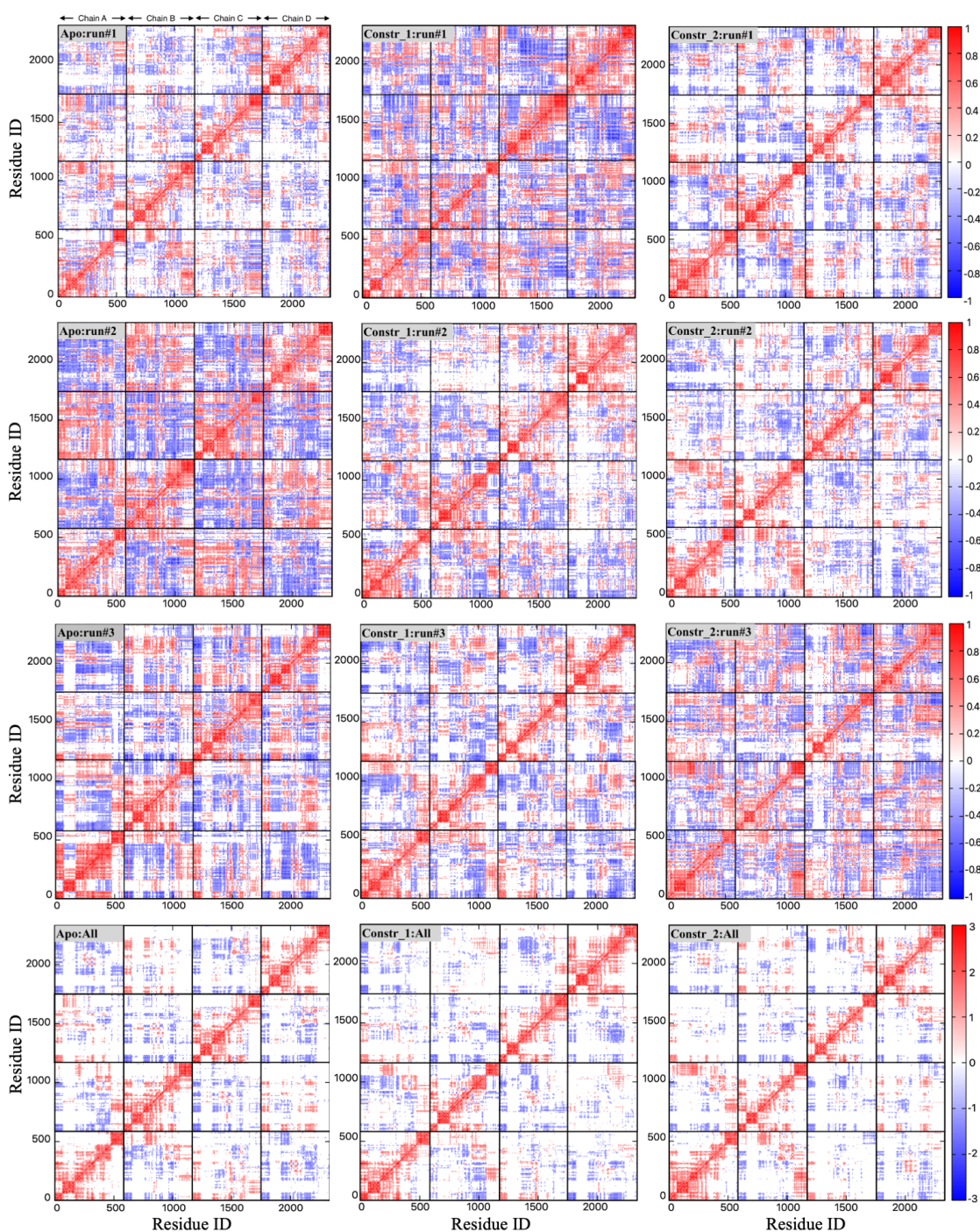
Figure 3.14 Correlations between motion vectors of alpha carbon atoms calculated using components that explain 70% of the motion as a result of principal component analysis in pyruvate kinase enzyme. Values between -0.4 and 0.4 are not shown on all maps.
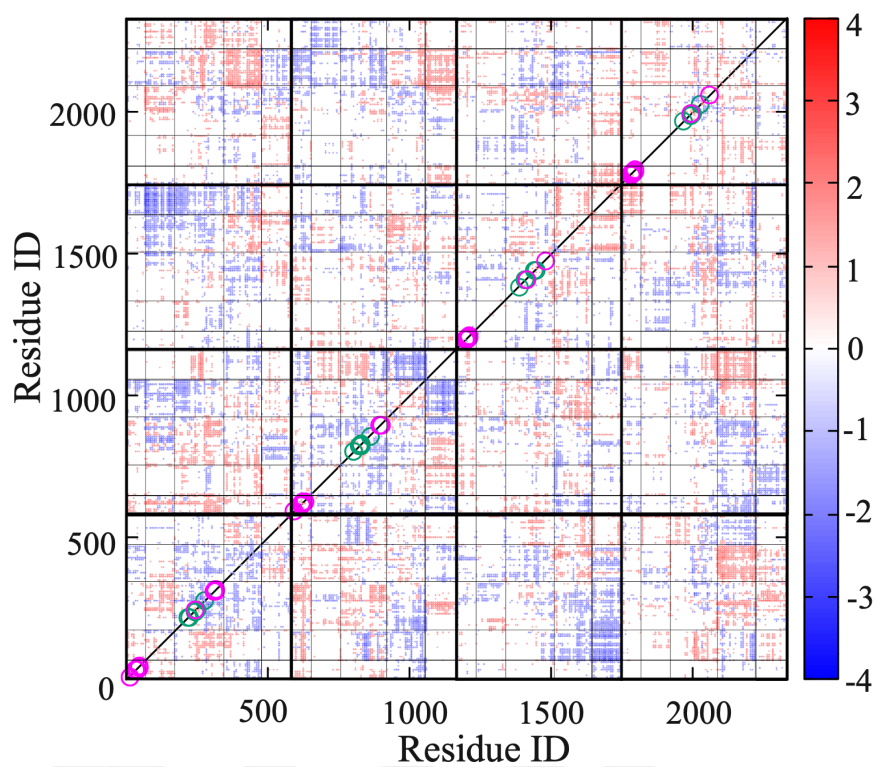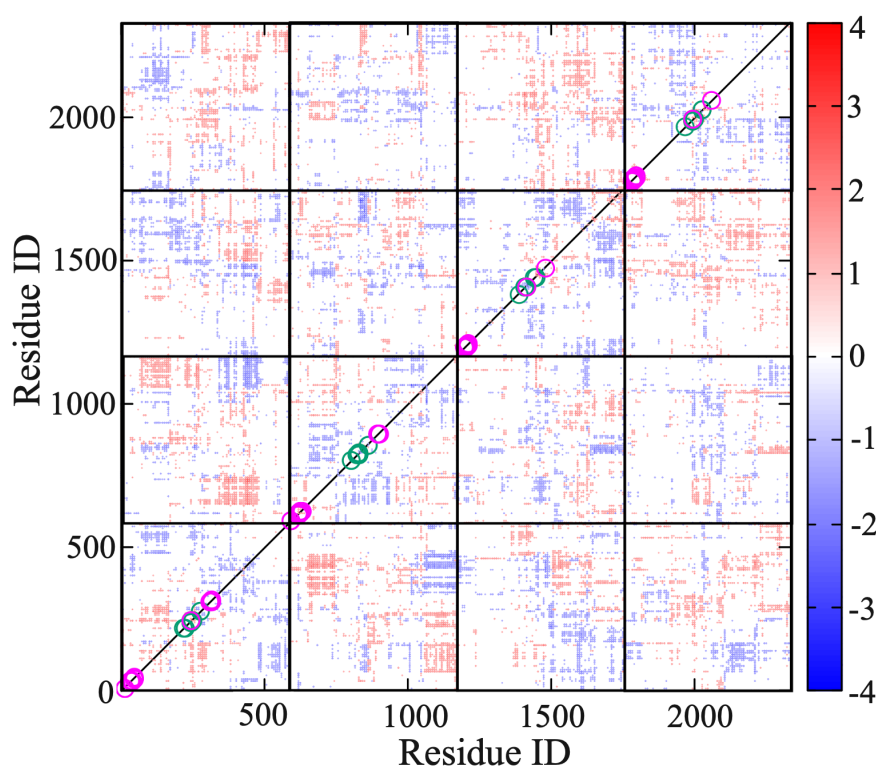
Figure 3.15 Difference in correlation of positional fluctuations between a) Constr-1 and apo, b) Constr-2 and apo.

		APO	CONSTR-1
1	Aa/Ac	(+)	0
2	Aa/Ab	(+)	(++)
3	Ba/Bb	(++)	(+)
4	Bc/Bct	(+)	0
5	Ca/Cb	(+)	0
6	Cc/Cct	(+)	(++)
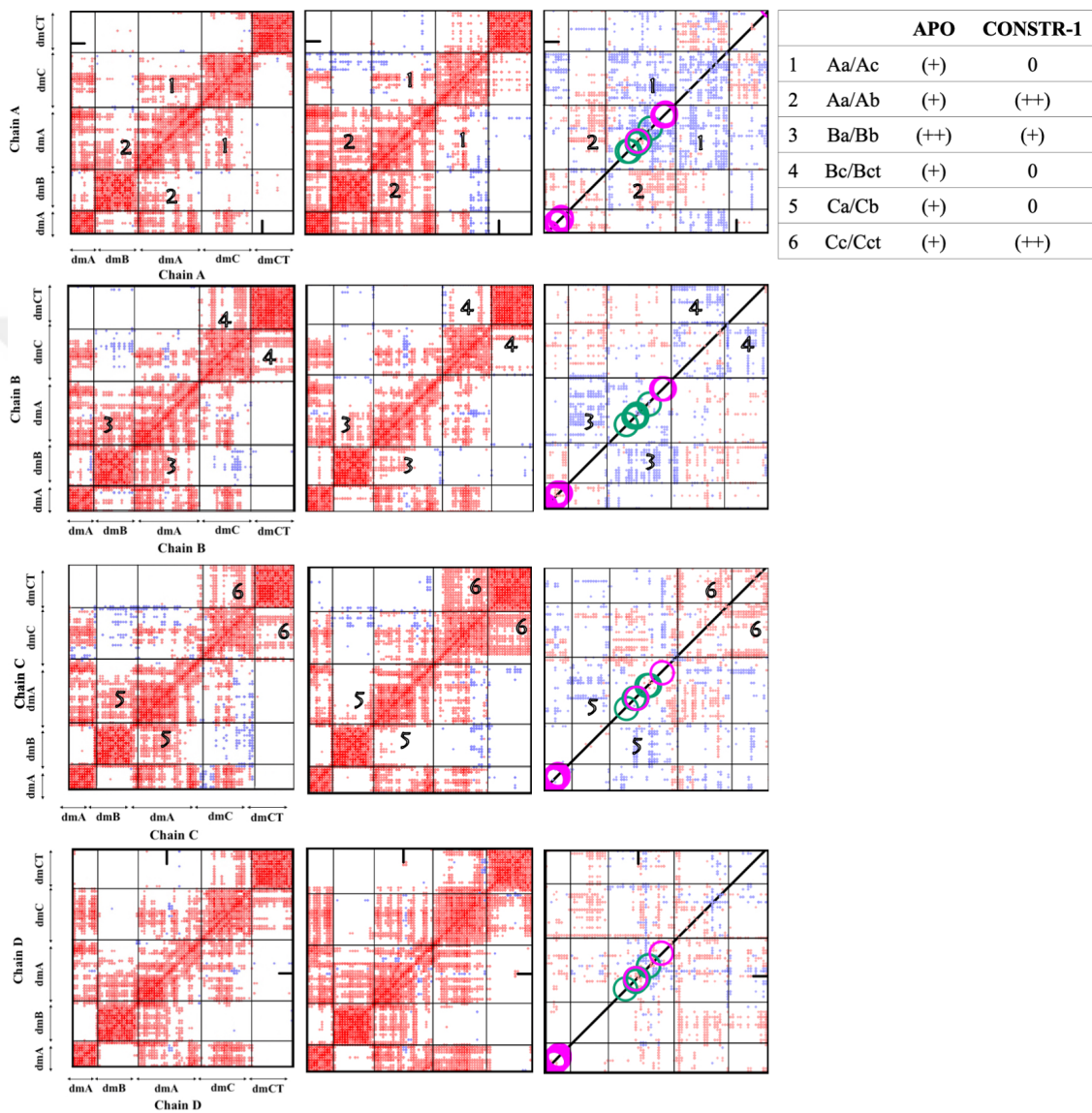
Figure 3.16 Apo correlation sum, contsr-1 correlation sum and their difference maps in row accordingly. Any significant correlation and anticorrelation highlighted with number on map and represented as (+), (-) , (0) in the table.
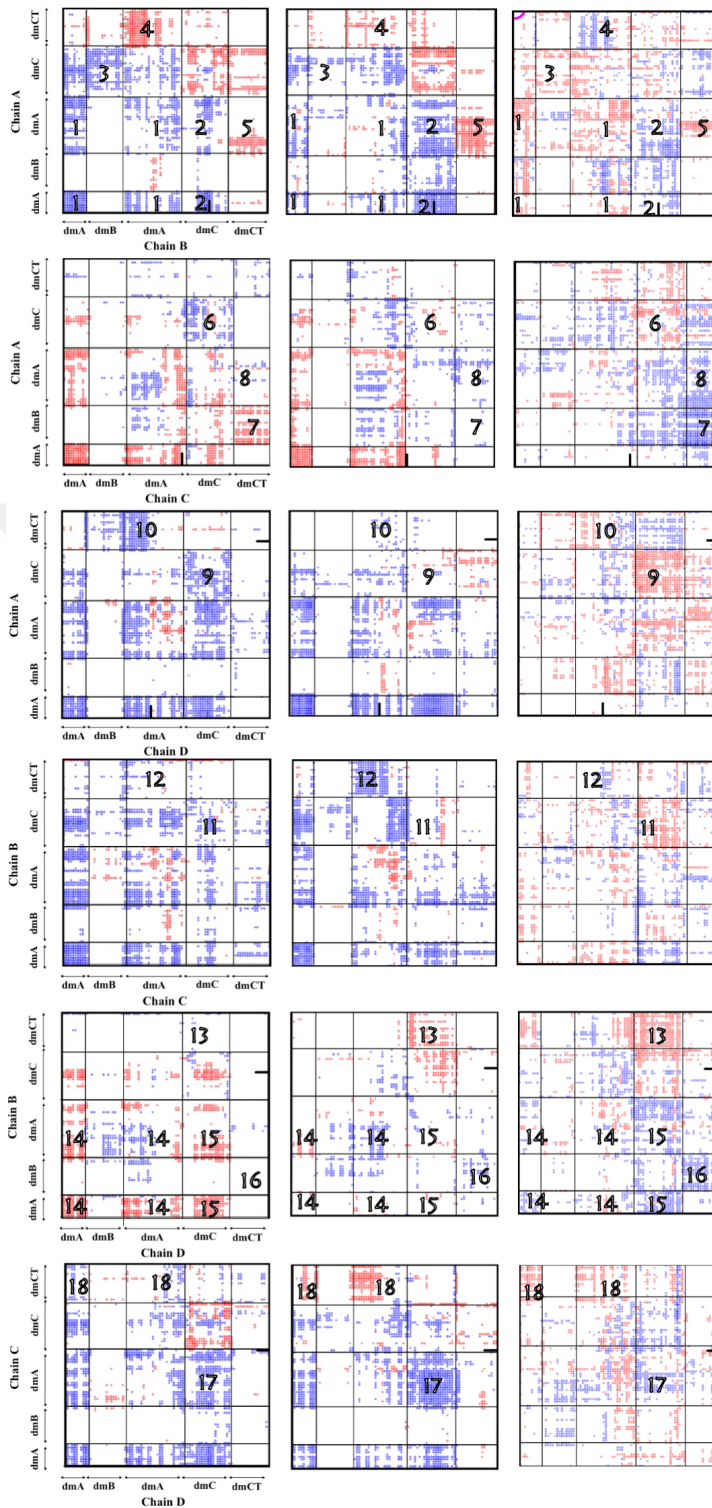
		APO	CONSTR-1
1	Aa/Ba	(-)	0
2	Aa/Bc	(-)	(- -)
3	Ac/Bb	(-)	0
4	Act/Ba	(+)	0
5	Aa/Bct	(+)	(++)
6	Ac/Cc	(-)	0
7	Ab/Cct	(+)	0
8	Aa/Cct	0	(-)
9	Ac/Dc	(-)	0
10	Act/Da	(-)	0
11	Bc/Cc	(-)	0
12	Bct/Ca	0	(-)
13	Bct/Dc	0	(+)
14	Ba/Da	(+)	0
15	Ba/Dc	(+)	0
16	Bb/Dct	0	(-)
17	Ca/Dc	(-)	(- -)
18	Cct/Da	0	(+)

Figure 3.17 Apo correlation sum, contsr-1 correlation sum and their difference maps in row accordingly. Any significant correlation and anticorrelation highlighted with number on map and represented as +, - ,0 in the table

In Figure 3.19, among 15 interchain correlation differences, 9 of them represented a lost in correlation between interchain domains in constr-2 simulation. On the other hand, There were four newly created correlations and two of them belonged to CT domain. CT of chain B and B domain of chain A started to fluctuate in the same directions. Interestingly, the CT domains of the B and D chains, which were far from each other, started to move in opposite directions. A negative correlation occurred between B and C domain of chains A and D. Finally, the degree of positive correlation between neighboring C domains in the C and D chains became intensified.

		APO	CONSTR-2
1	Aa/Ab	(+)	(++)
2	Ba/Bb	(+)	0
3	Ca/Cb	(+)	0
4	Cct/Cc	(+)	0
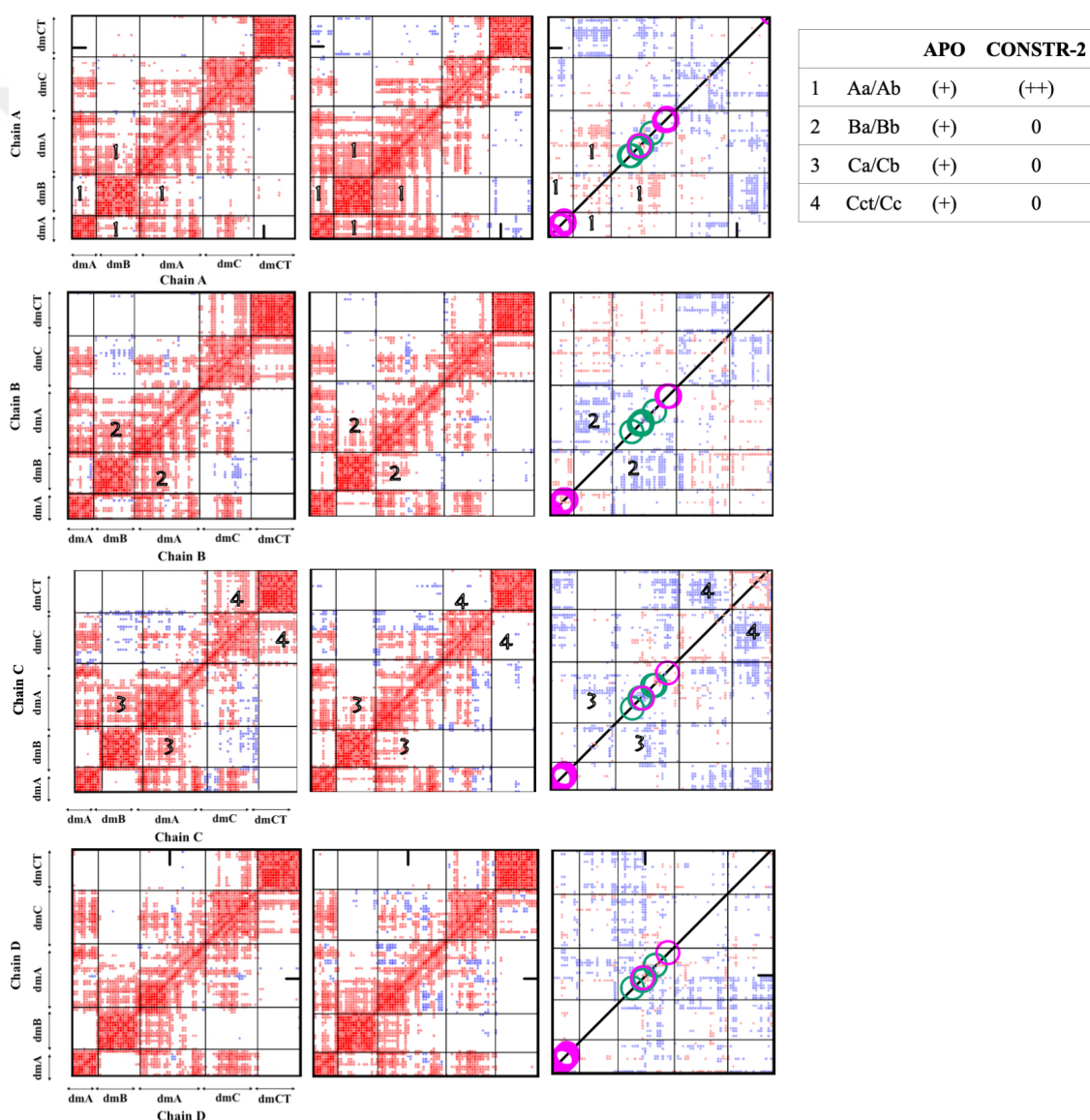
Figure 3.18 Apo correlation sum, contsr-2 correlation sum and their difference maps in row accordingly. Any significant correlation and anticorrelation highlighted with number on map and represented as +, - ,0 in the table
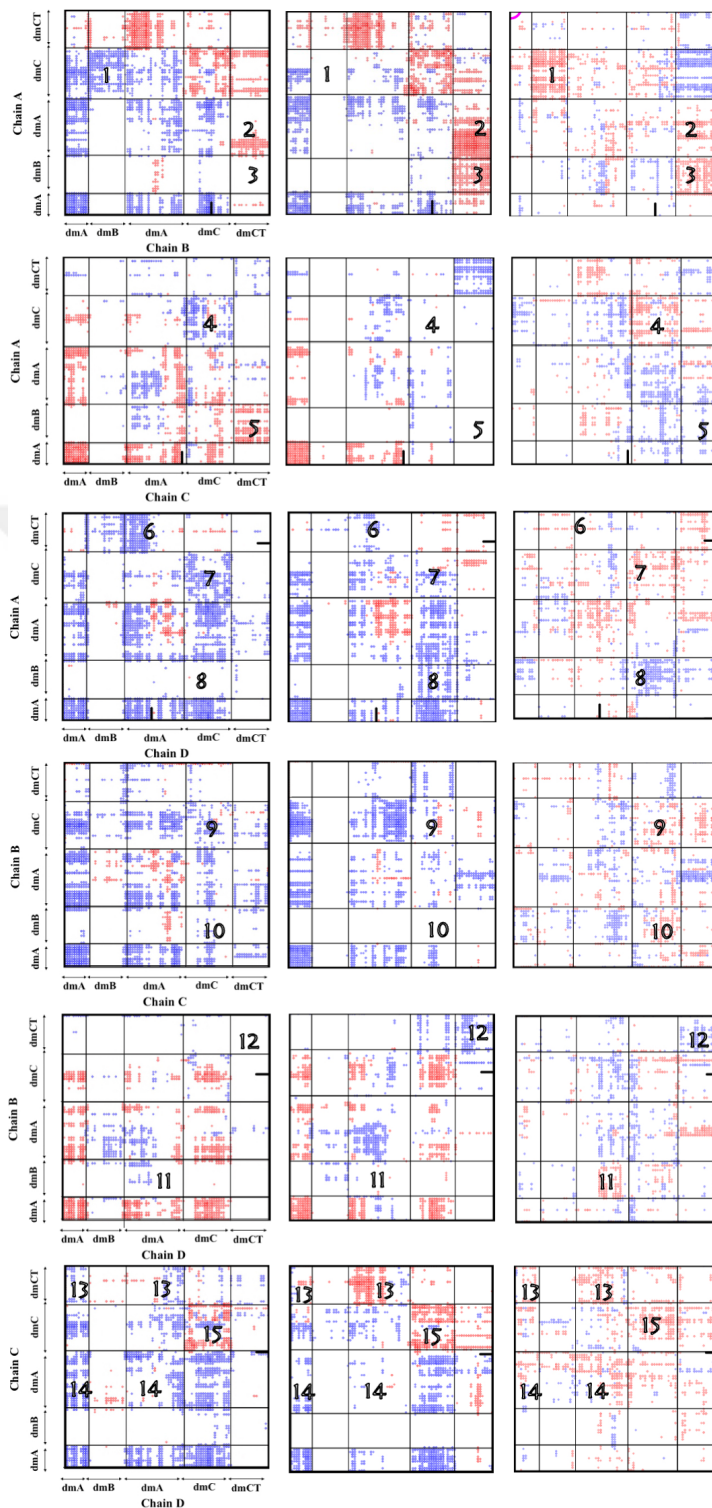
		APO	CONSTR-2
1	Ac/Bb	(-)	0
2	Aa/Bct	(+)	(++)
3	Ab/Bct	0	(+)
4	Ac/Cc	(-)	0
5	Ab/Cct	(+)	0
6	Act/Da	(-)	0
7	Ac/Dc	(-)	0
8	Ab/Dc	0	(-)
9	Bc/Cc	(-)	0
10	Bb/Cc	(-)	0
11	Bb/Da	(-)	0
12	Bct/Dct	0	(-)
13	Cct/Da	(-)	(+)
14	Ca/Da	(-)	0
15	Cc/Dc	(+)	(++)

Figure 3.19 Apo correlation sum, contsr-2 correlation sum and their difference maps in row accordingly. Any significant correlation and anticorrelation highlighted with number on map and represented as +, - ,0 in the table

42

A-A domains, which are essential for the protein structure stability, were examined by looking at their correlation graphs. Accordingly, in the case of apo, the neighboring A domains in all A-D and B-C chains moved in opposite directions to each other (Figure 3.17 chA/D). While there was no change for the A-D chains in Constr-1, the negative correlations in the B-C chains were slightly decreased (3.17 map chB/C). On the other hand, although it is not very clear in constr-2, there was a decrease in the strength of negative correlation between these two domains in both A-D and B-C chains (Figure 3.19 map chA/D, chB/C). The effect of the residues that were restricted was noticeable on correlation maps. When four A-A domain interactions were examined, it was observed that there were decreases in correlations in three regions for constrained states.

The C-C domains adjacent to each other are significant for the rigidity of the protein and the function of the effector region. Accordingly, in the case of the apo, the domains in both regions fluctuated in the same direction. When one looked at the constrained cases, it was seen that the correlation of the adjacent fields also increased partially which is prominently seen in #15 on Figure 3.19.

CT domain provides the structural stability by interacting the A domain in the active R state of pyruvate kinase of some organisms (Lovell et al., 1998; Sakai, 2004; Suzuki et al., 2008; Tanaka et al., 1995). When the A and CT domains were examined, different correlations from apo were observed in both constrained cases. There was a positive correlation between the Aa/Bct in the Apo state (Figure 3.17 #5). When the maps of the constrained state examined, it was seen that these correlations were strengthened in a positive direction (3.17 #5 and 3.19 #2). Again, there was a positive correlation between the Ba/Act for apo state (3.17 #4), but in the case of constrained, a slight decrease in the strength of positive correlations occurred. On the other hand, when looking at the A-CT correlations of both D-C and C-D chains in apo state, it was observed that there was a positive correlation between the Da/Cct (#18) and no correlation for Ca-Dct. To sum up, there were increased correlations between CT domains and A and also with other domains in

the constrained state (from table 3.17 #8, #13, #16,#18 and 3.19 #3, #12). This was one of the apparent differences between apo and constrained states.

A positive correlation of the A and CT domains shows that they fluctuated in the same directions. While they fluctuated in the same direction, their distance might increase or decrease. Structural differences between the A-CT domain regions in the apo and constrained structures were analyzed using distance distribution. Interatomic distances between pair of selected residues (GLU208- LEU536) were calculated for each snapshot in the trajectory. Distance results of three runs were combined for each set of MD simulation. Then the frequency distribution of the distances was plotted for apo, constr-1 and constr-2 states. Distance between these domains was approximately 10 Å in the crystal structure. Distance values lower than 10 Å indicates that domains were getting close to each other.

According to Figure 3.20, bimodal and skewed distributions was mainly formed for four chains. The distance between A-B chains in the apo state was mostly around 8 Å. In the constr-1 distribution, the distance was peaked at around 10.5 Å. The distribution of the constr-2 state was grouped in two peaks of distance at approximately 8 and 10.5 Å in the Figure 3.20 (a). The plot of the Ba-Act domains for apo and constr-1 shows bimodal distribution with two peaks. In this region, domains were far away from each other in the apo state. It was observed that the distances both increased and decreased in constr-1. On the other hand, the distances in constr-2 increased (see Figure 3.20 (b)). Interestingly, distances between Ca/Dct which is represented in Figure 3.20 (c), increased up to 35 Å - 45 Å in apo state. However, two different peaks were formed in three simulations. The most frequent distances were 10 Å. Finally, bimodal distributions were observed for Da/Cct domains in the Figure 3.20 (d). The distance frequency of the two peaks, which were at 8 Å and 10 Å, were very close to each other in the apo simulation. On the other hand, peaks of the constrained states were higher than 10 Å and apo states were lower than 10 Å.
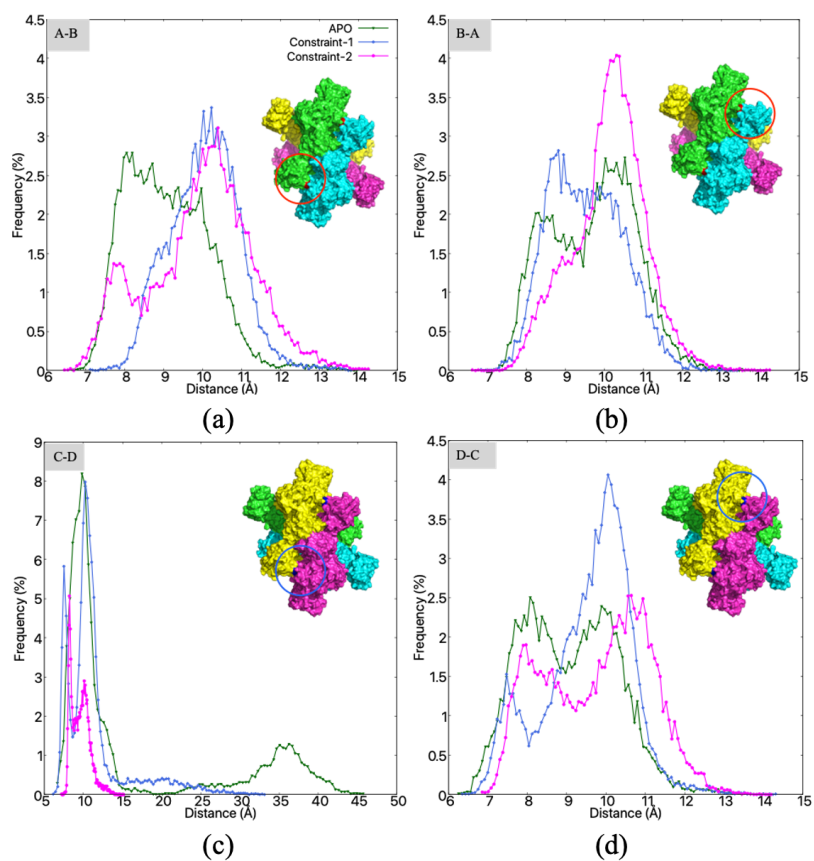
Figure 3.20 Distribution of the distance between GLU208 of A domain and LEU536 CT domains. Frequency of distances between a) A domain of A chain and CT domain of B chain b) A domain of B chain and CT domain of A chain c) A domain of C chain and CT domain of D chain d) A domain of D chain and CT domain of C chain. Circle on the PK represents the domain regions.

Optimum stability and structural rigidity for the efficient catalytic activity was provided by CT-A domain interactions in *S. aureus* PK (Zoraghi et al., 2010) Although it is not possible to understand which structural state was more stable by looking at the distance frequencies, apo and constrained showed bimodal distribution in their plots. The frequency of apo and constrained states at two different distances indicates that they sampled two separate conformational states. On the other hand, the distance between them was not close enough to interact in both apo and constraints. This indicates that they were far from the active state structure.

3.4 Clustering of Catalytic Region And Change in The Secondary Structure for $\alpha'6$ Helix

Any activator or inhibitor binding to the effector site affects the catalytic region of pyruvate kinase. The residues that was restricted in the known and proposed region might stimulate the catalytic region and the global dynamics of the protein. It is not easy to understand this effect by tracking MD trajectories, which are very long and represent the different conformational states in each frame. Reducing large numbers of frames into characteristic states can be performed via RMSD clustering method.

For this analysis, trajectories of all nine simulations which represent three independent simulations for three states (apo, constr-1 and constr-2) were combined. Structure of the 5000 frames aligned to the least mobile AC domain were clustered based on an RMSD value of 4.5 Å. As a result, 4 clusters were determined, and a conformation from each cluster was selected as a representative. This chosen conformation represented the average property of that cluster which is closest to the cluster centroid (i.e. average structure).

As a result, four frames emerged. One of them incorporated all three states, the second only apo, the third one only constr-1 and the last one only constr-2. These frames were aligned to each other and combined with the substrates in the active region as illustrated in Figure 3.21 It is known that in the active structure *L. mexicana*

PK B domain converges towards the A domain and remains more stable (Naithani et al., 2015). Accordingly, we expect the B domain to move away from the A domain or uncertainly fluctuate in constrained simulations. Although the RMSD results showed that the B domains were extremely active and unstable, this difference was clearly seen in the cluster results. However, when the region to which the substrate was attached was investigated, it was observed that some mobile loops (structures shown in pink and green color) approached slightly inside the catalytic region in the constrained states. At the same time some alpha helices transitioned into coil structure as illustrated in the Figure 3.21 This showed that the allosteric signals from the restricted site reached the catalytic site.

The helix that turned into a coil structure was the , $\alpha'6$ helix, which interacts with PEP in the active state of the *L. mexicana*. Close views of these differences between ,$\alpha'6$ helix cluster were illustrated chain by chain in Figure 3.22 In chain A, two constrained states (magenta and green) moved away from the PEP and shifted to the left. Clusters representing the apo and all structures in chain B, turned into coil while the constrained ones were still helix. In the C chain, the helix of the constr-1 shifted to the left and the constr-2 structure oriented upwards. On the other hand, helices from both constrained states moved up, and the helix of the cluster representing all states (grey) turned into the coil. In addition to these, different states were observed in the four different chains. While one of them remained in the turn or coil, those in the other chain persistently remained in the helix state.

In a study conducted by Morgan and his coworkers, it was shown that Arg310 in the catalytic region makes hydrogen bonds with Arg and Gly in the neighbouring $\alpha'6$ helix (243-249), while *L. mexicana* pyruvate kinase transitions into active structure followed by the rotation of AC domains (Morgan et al., 2010). Likewise, based on another study, it was shown that this helix had a more regular and stable structure when it was in the active state for *L. mexicana* PK (Naithani et al., 2015). Additionally, $\alpha'6$ helix structure closed onto the PEP molecule in the E.coli pyruvate kinase type I (Mattevi et al., 1995). Although no comparison was made with the
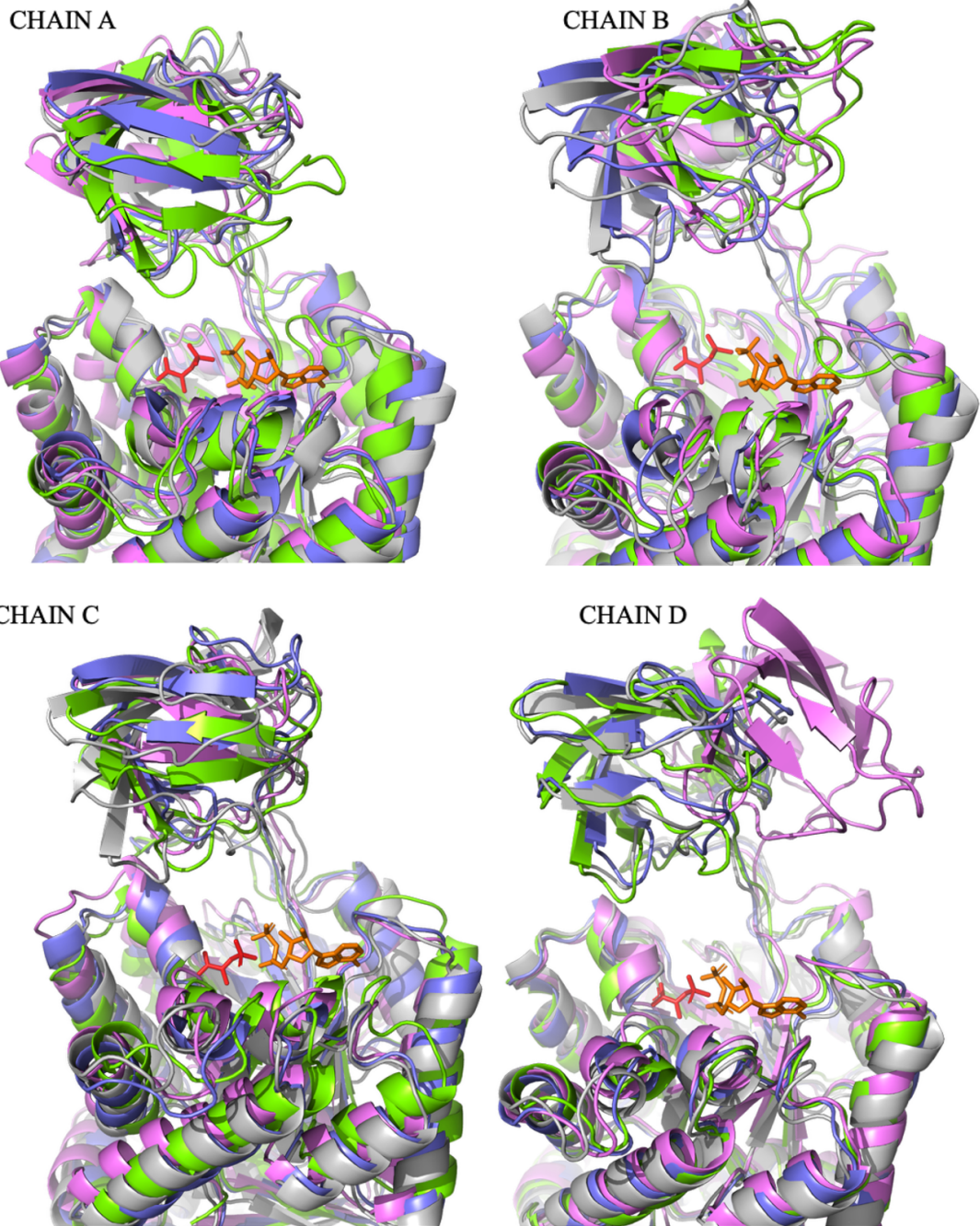
Figure 3.21 Catalytic region clusters of chains. Grey: all, Blue: apo, Magenta: Constr-1, Green:Constr-2, Red: PEP and Orange: ADP. Significant variations on the structure marked with arrow.
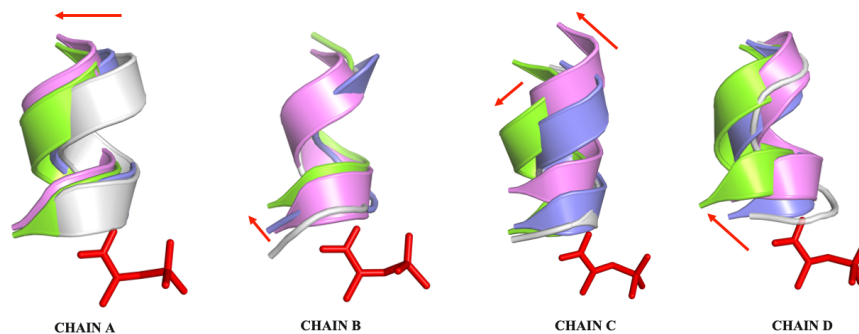
Figure 3.22 Structural comparison of $\alpha'6$ helix from four different clusters. Color representation; Grey:all, Blue:apo, Magenta: Constr-1, Green:Constr-2, Red:PEP

active structure, it was seen that there were some structural differences between the apo and constrained states which have to be in the form of inactive structure. Coil formation was one of them and it was further investigated by the analysis of the secondary structure change.

The variations in the secondary structure of $\alpha'6$ helix was monitored throughout the simulation time as illustrated in Figure 3.23 For this purpose, secondary structure type of each of seven residues on $\alpha'6$ helix was computed using VMD software tool. In each simulation consisting of 5000 frames, 35000 secondary structure information was obtained for seven residues in one chain. For four chains, this corresponds to 140000 data points indicated by colours according to the type of secondary structure.

Accordingly, it can be deduced that yellow and orange colors corresponding to turn and coil types were quite intense in some chains of apo simulations than constrained states. Although coil and turn conversions were more frequent in the apo structure, occasional coil conversions were also seen in the constrained states. In this respect, they are both similar to each other. In Table 3.2, the occurrence of the secondary structure types was shown numerically for four chains. When the total number of helices were subtracted from the total number of coils, it was seen that the coils in the apo structure were in the majority, while the helices were in the majority in constrained simulations.
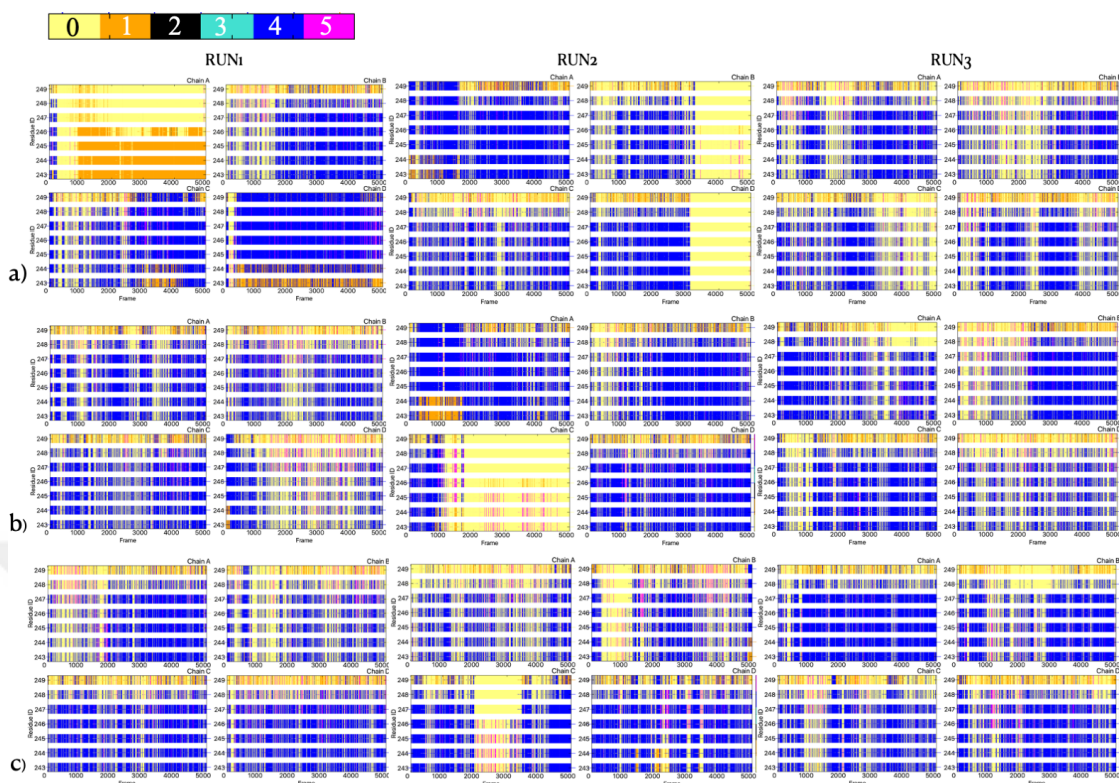
Figure 3.23 $\alpha'6$ Secondary structure alteration for a) apo, b) constr-1 c) constr-2 simulations. Chain A, Chain B, Chain C, Chain D plots represented for each simulation. The numbers and colors on the scale represents type of secondary structure; 0:yellow: Turn, 1:orange:coil, 2:black:isolated bridge, 3:cyan:beta sheet, 4:blue:alpha helix, 5:magenta:3-10 helix

Table 3.2 Secondary structure change in the three states of simulation

Simulation	Set #	0 (turn)+1 (coil)	4(a-helix)+5(3^{10} helix)	Difference (4+5) - (0+1)
Apo	Run1	70249	69720	-529
	Run2	68031	71969	3938
	Run3	72572	67428	-5144
Constr-1	Run1	67337	72663	5326
	Run2	61247	78753	17506
	Run3	67802	72193	4391
Constr-2	Run1	63215	76785	13570
	Run2	68811	71217	2406
	Run3	58842	81186	22344

Although the results seem to be the opposite of what should be expected for inactive states, it is unknown whether the allosteric interaction will affect the catalytic site of CT PK in the same way as *L. mexicana* PK which was activated and inhibited with different molecules. All these results indicated that alpha helices in the constrained states were more similar to the active state, but not close to the active conformation. They may simply represent an intermediate inactive state. The $\alpha'6$ helix structures in the constrained states moved away from PEP, but in the apo state it has turned into a coil or a turn or has been found closer to the PEP. On the other hand, maybe this could be a sign that the helix tends to become active. Making new stronger bonds can happen by breaking down the old ones, the bonds that give the helical shape.

3.5 Profiles of Principal Axes of Chains and Domains

Studies show that in pyruvate kinase transition to the active R state from inactive T state was possible by domain rotations. According to domain and subunit rotation model, the subunit consisting of A, B, C domains rotates 16°,17° and 15° respectively around the center point of the domains of the *E. coli* PK (Donovan et al., 2016). Rotation of the subunits and domains for *E. coli* PK is schematically represented in the Figure 3.24 a. On the other hand, rotation of the domains as a rigid body is another model of orientation for pyruvate kinase which is illustrated on the Figure 3.24 b. For *T. cruzi* PK, rotation degree is 8° around the pivot point of A/C core domains (Morgan et al., 2014). A/C domains of *L. mexicana* pyruvate kinase rotates 6° in all subunits (Morgan et al., 2010). These rotations stabilizes the structure in the active state by salt bridges on the C-C interface region.

Each domain rotation for 5000 frames of our MD trajectory for each of the nine independent runs was investigated via principal axes analysis using Orient package of VMD software tool (Humphrey et al., 1996). The main axis of rotation is an eigenvector of the moment of mass of the tensor of inertia, and it is defined with respect to the center of mass point. Since the tensor moment of inertia is defined with
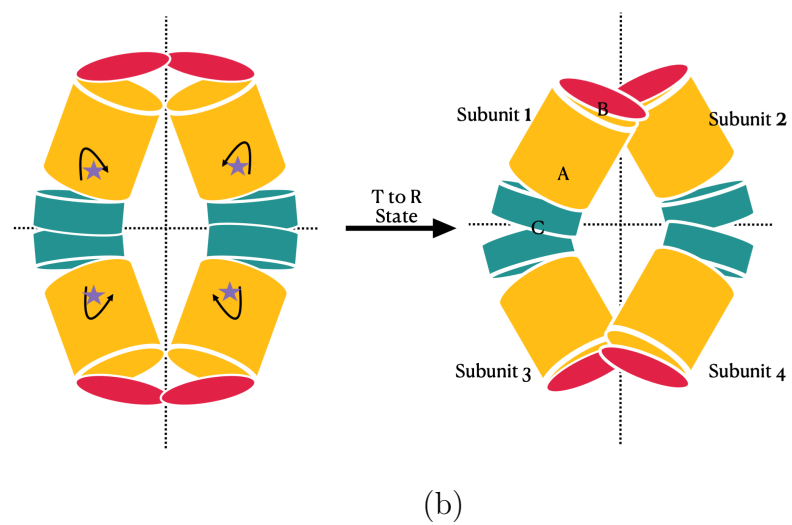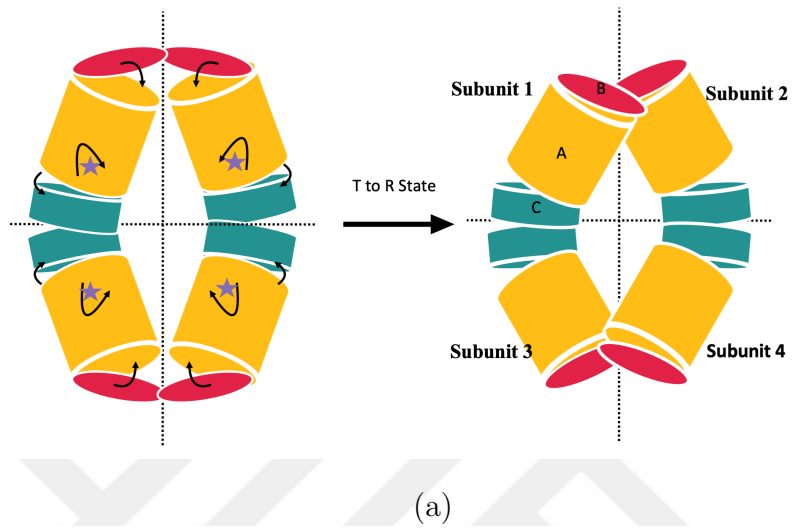
Figure 3.24 Schematic representation of allosteric transition models for a)domain and subunit rotation and b) Rigid body orientation (illustrated by Metin with Keynote tool)

respect to the zero point in space, all the principal axes pass through this center of mass. Orient package determines the principal axes for two selected segments. Using a TCL script developed in our group, the angle changes between PCA1 PCA2 and PCA3 axes were calculated for each frame in the trajectory using the dot product of each normalized direction vector as defined in equation 3.4,

$$\cos(\theta) = \frac{\langle \overline{x_1} \cdot \overline{x_2} \rangle}{|\overline{x_1}||\overline{x_2}|} \tag{3.4}$$

where $\overline{x_1}$ represents the first principal axis of A/C domain and $\overline{x_2}$ the first principal axis of all tetramers. Figures 3.25 and 3.26 represent the directions of principal axes of tetramer and A/C domains respectively, for the initial frame of apo-run#1 simulation. The angle between directions of A/C domains and all protein were compared in Figures 3.25, 3.26.
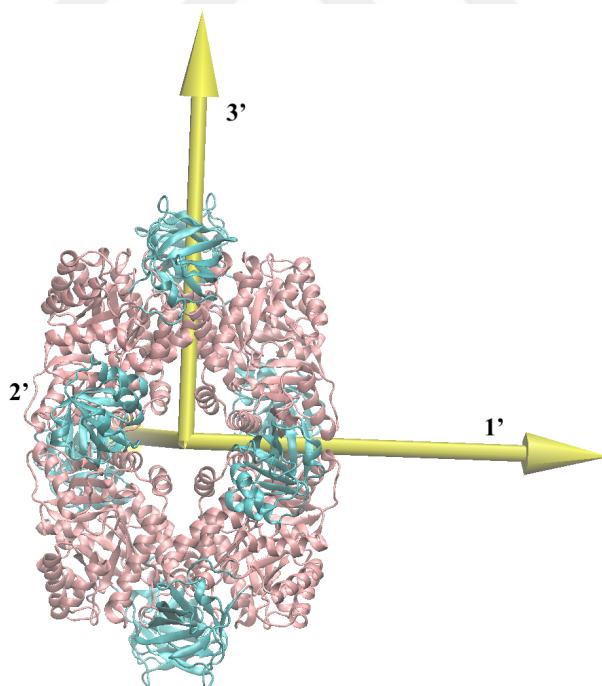
Figure 3.25 Principal axes of tetramer. The initial frame of apo-run#1 was used for illustration. A/C and B/CT domains represented in pink and cyan respectively.

Angles between PCA1 of apo state fluctuated less than constrained states for all chains. Similarly, in the PCA2 direction, A/C domains in apo state fluctuated more
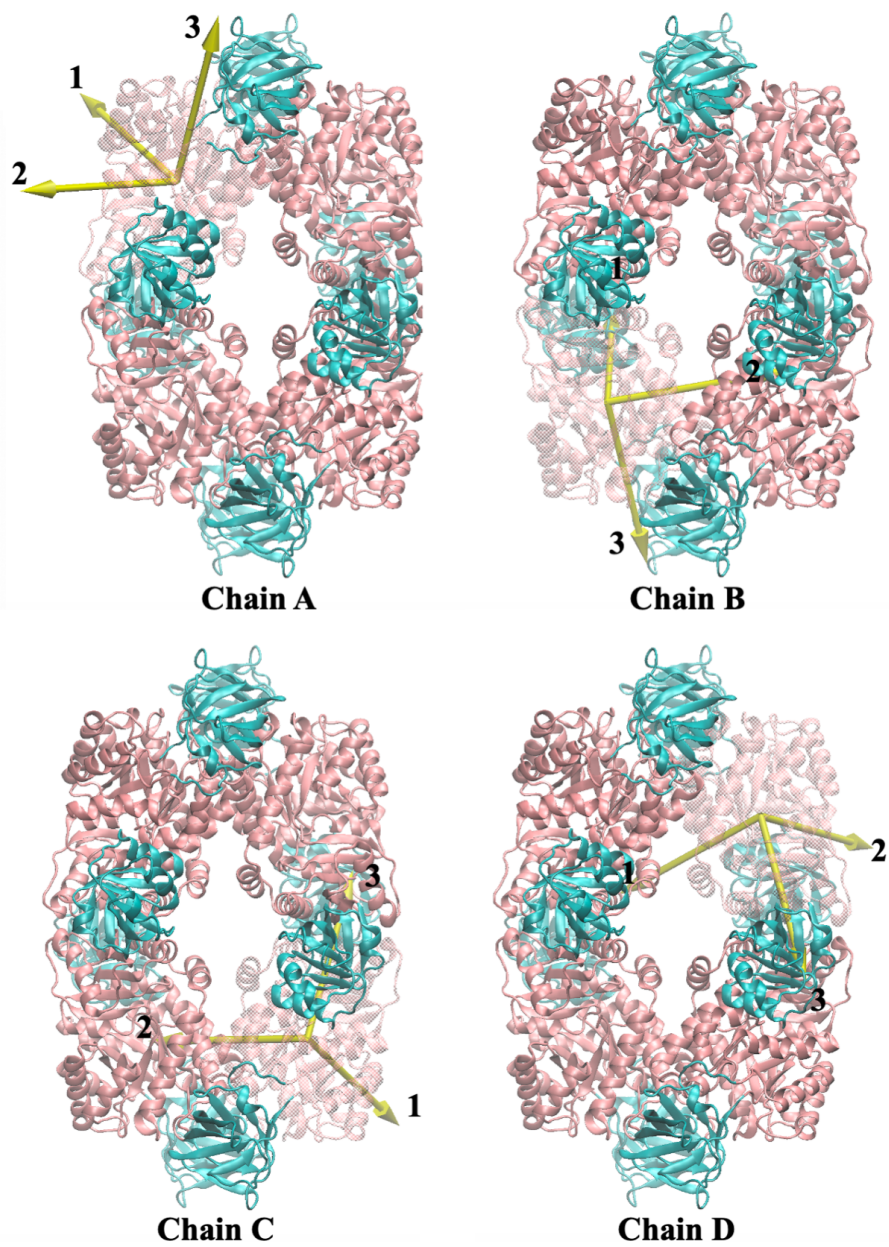
Figure 3.26 Principal axes of A/C domains for each chain in the tetramer. The initial frame of apo-run#1 was used for illustration. A/C and B/CT domains represented in pink and cyan respectively

regularly than constrained states. Fluctuations of angles between A/C domains and the tetramer for the apo states were decreased towards the end of the simulation. As illustrated in Figure 3.27, there were no significant variations in the angles for this direction. From this point of view, it turns out that A/C domains rotate around direction 3 of the principal axes for all chains. In Figure 3.30, rotation around the PCA3 axes was illustrated by using the initial and 276th frames of run#1 simulation of apo state. The difference in the angles between apo and constrained states was noticeable. Constrained residues affected the domain rotations, which are very crucial for protein function and activity. The range of rotation angles increased, and fluctuations became more irregular. Any disruption on rotations of the domains will damage the stability of the protein structure.

Investigation of principal axes of individual A and C domains were illustrated as in Figures B1-B6 provided in Appendix B. According to Figure B.2, domain A rotated on PCA2 in all chains more than the other axes. Schematical representation of the variations in angle between different frames (frame #1 and frame#3450) of chain A in run#1 of apo simulation was illustrated in Figure 3.31. The difference between apo and constrained states was evident in PCA2 angles. Almost in all chains, rotation angles in constrained (red and green) states fluctuated more than the apo state (purple). On the other hand, there were no significant differences between apo and constrained states in PCA1 and PCA3 (Figures B.1 and B.3). Individual examination of rotations in domain C showed similar angles with domain A as illustrated in Figures B.4, B.5, B.6. Rotation angles in PCA2 axes (Figure B.5) differentiated between constrained and apo simulations. Rotation angles of constrained states fluctuated in a distinct range of 110°-170°. Different set of simulations showed separate rotation angles from each other for the constrained states. In contrast to constrained runs, apo states in the direction of PCA2 for three simulations showed similar rotation angles.

For the rotation of individual domains, the angle difference between apo and constrained states was significant in PCA2. On the other hand, fluctuations that were
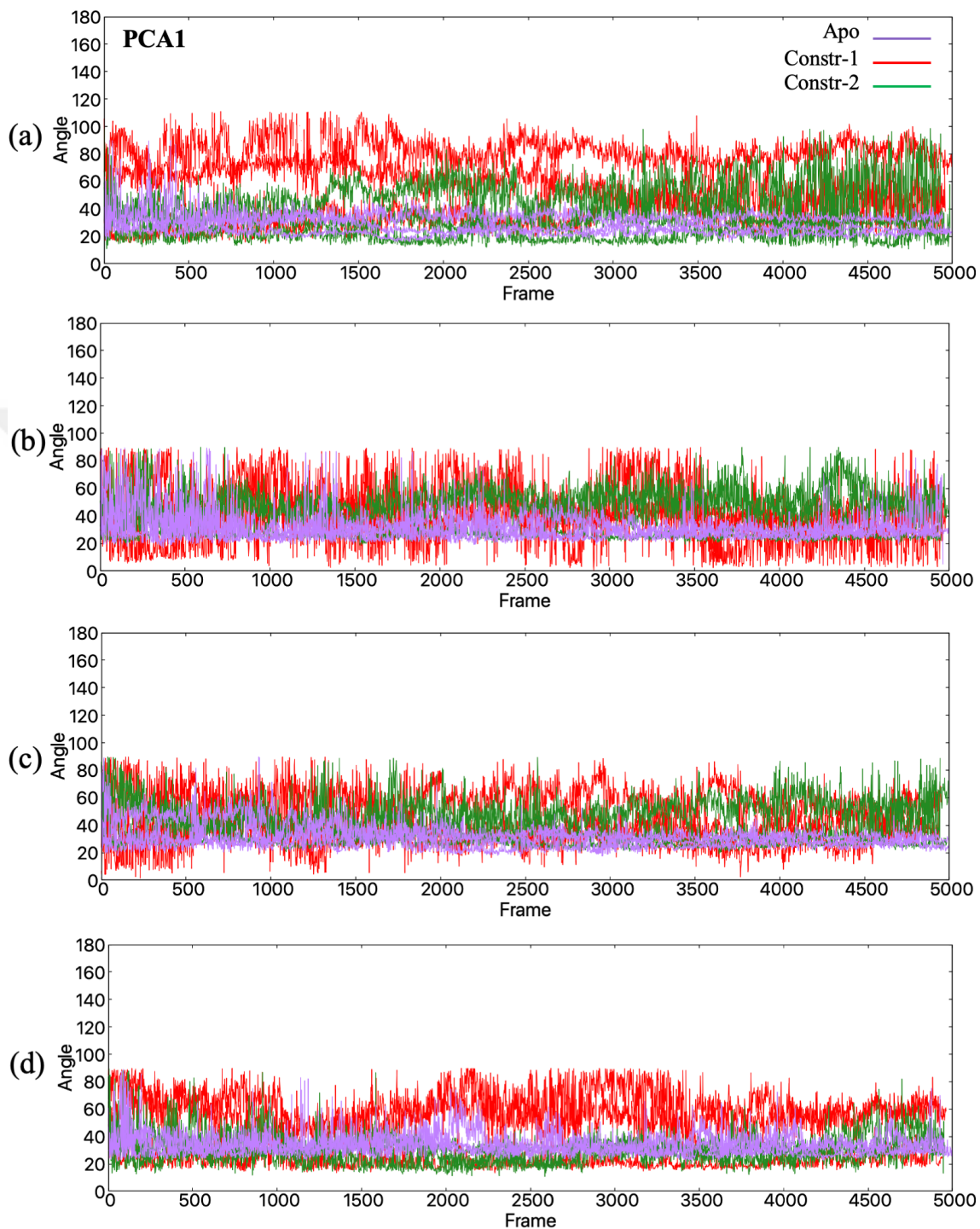
Figure 3.27 Rotation angles alterations of domain A/C through the direction 1 for a) chain A, b) chain B, c) chain C d) chain D. Apo, constr-1 and constr-2 represented with purple, red, green respectively
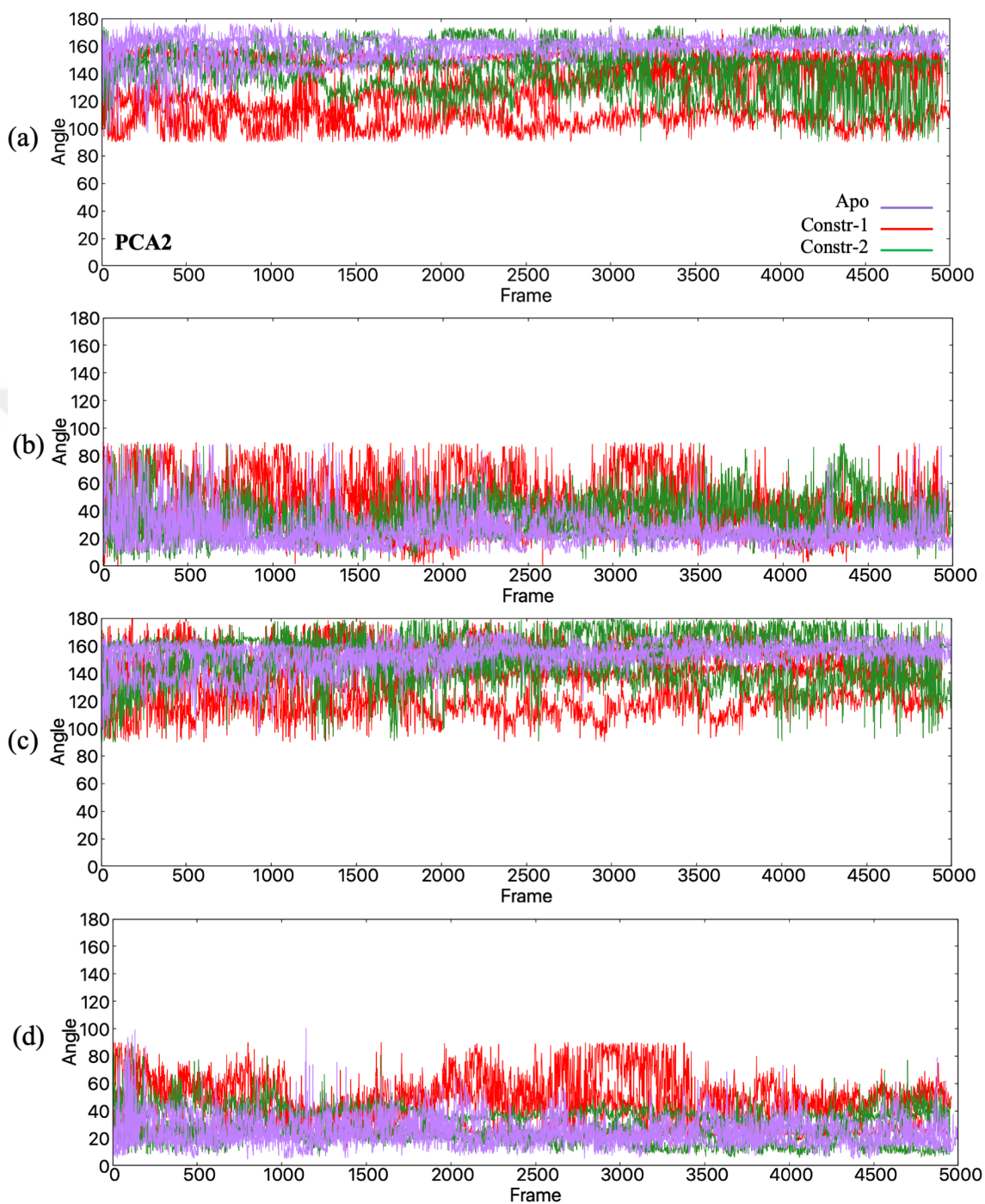
Figure 3.28 Rotation angles alterations of domain A/C through the direction 2 for a)chain A b)chain B, c)chain C d) chain D. Apo, constr-1 and constr-2 represented with purple, red, green respectively
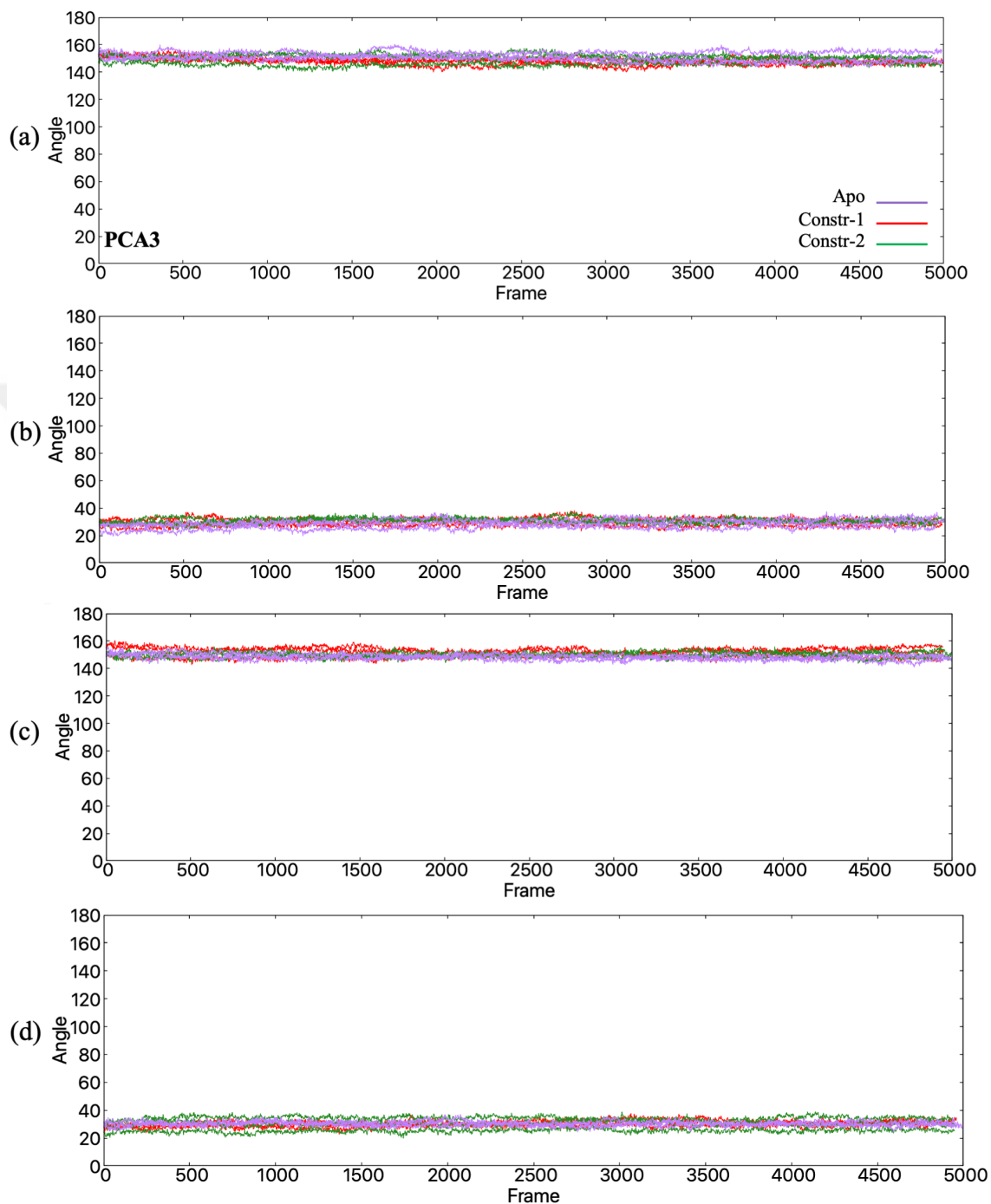
Figure 3.29 Rotation angles alterations of domain A/C through the direction 3 for a)chain A b)chain B, c)chain C d) chain D. Apo, constr-1 and constr-2 represented with purple, red, green respectively
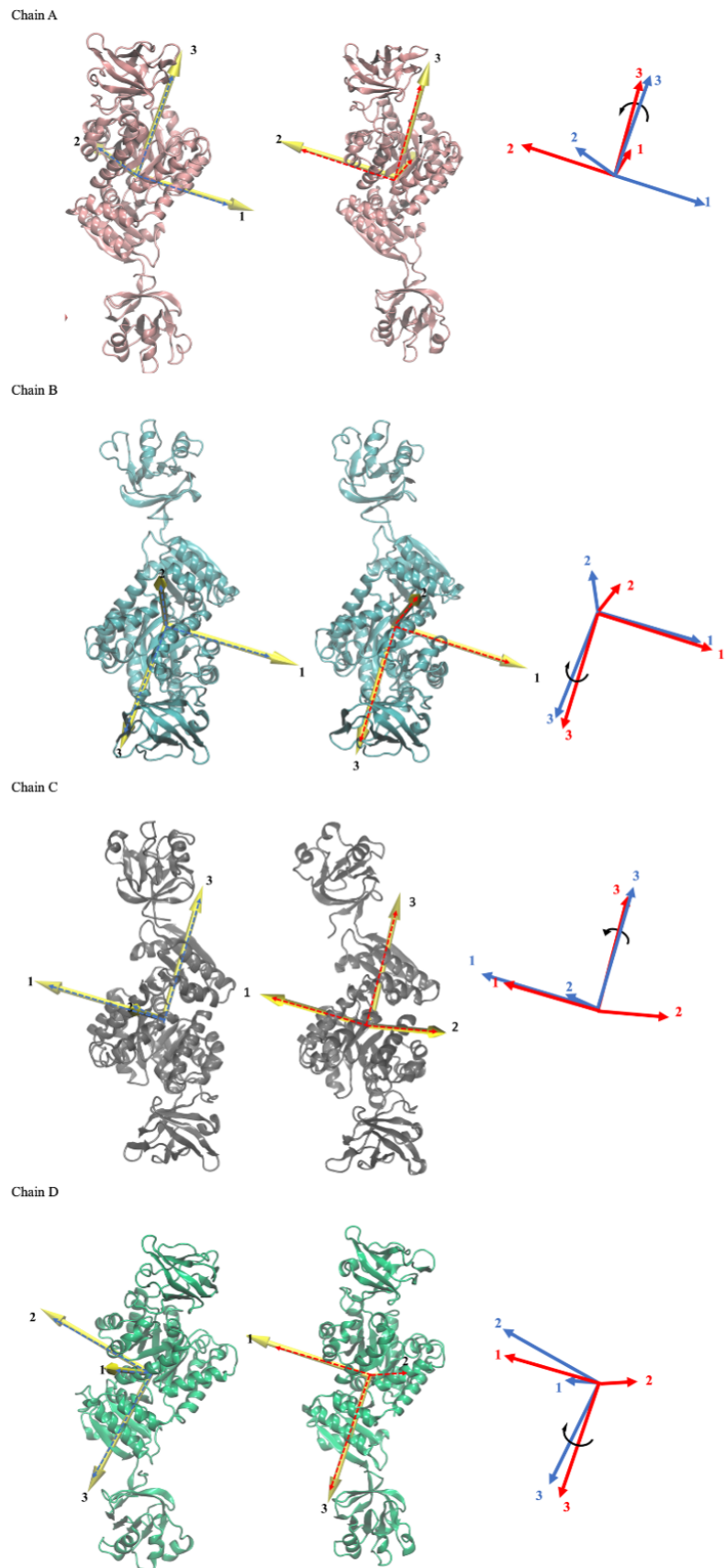
Figure 3.30 A/C domain rotation around PCA3 axes for each chain
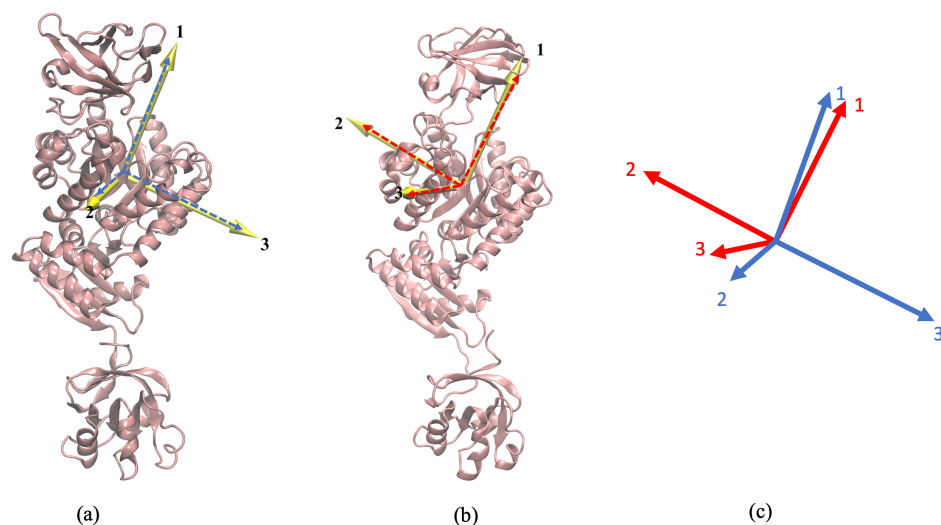
Figure 3.31 Comparison of the principal axes of a) frame#1 (PCA1=106°, PCA2=25°, PCA3=73°) and b) frame#3450 (PCA1=106°, PCA2=104°, PCA3=65°) for domain A of PK from run1 simulation of apo state

colored in red for all angles were generally higher than the green ones. For example, in Figure 3.27, for constr-1 state illustrated in red, domains fluctuated more than the green ones which corresponded to constr-2 state. Accordingly, the restriction of the proposed regions disrupted the rotations more than the restriction of known allosteric region. In the study of Ayyıldız, it was suggested that the proposed region had stronger allosteric features than the known allosteric regions (Ayyildiz et al.2020).

3.6 Distance Fluctuation Analysis

Communication between the catalytic region and the effector region takes place through chains and domains. With the principal component analysis, the correlated motions between domains and chains were examined. However, the residues and domains in communication can be revealed through the distance fluctuation analysis. In this study, the regions in communication in the apo state and the regions in the constrained state were compared. Root mean square distance fluctuations (RMSDF) is a mathematical expression to detect the possible signaling between

residues. Distance fluctuations between residues were calculated with the Equation 3.5

$$RMSDF_{ij} = \left\langle r_{ij} - \langle r_{ij} \rangle^2 \right\rangle^{1/2} \tag{3.5}$$

Where $\langle \ \rangle$ is the time average over the all trajectory and r_{ij} represent the distances between the C_a atoms of i and j. Higher mean square displacement means low communication between atoms, while low RMSDF value indicates a strong communication. If there are low distance fluctuation values between two distant residues, it indicates that these two distant residues communicate with each other, and these regions are often key regions in signal transmission. Since the distance fluctuation values of the residues close to each other might be low, the contact map of the protein was created to see the difference between the close ones and the far ones. From the Equation 3.6, contact map of the protein was calculated by using the initial frame of the trajectory .

$$C_{ij} = \begin{cases} 1, if \delta_{ij} < R_c \\ 0, Otherwise \end{cases} \tag{3.6}$$

In this case, the atoms relative to each other by 6 Åas a threshold R_c were examined with the contact map. In this way, it will be understood whether there was a communication between them by looking at the distance fluctuations, apart from proximity. The distance fluctuations and contact degree for each pair of particles were illustrated on Figure 3.32. The maps of each set of simulation represented on columns for apo, constr-1 and const-2 states respectively. The maps on the last row showed the average of distance fluctuation for the three runs of simulations.

In the apo state, there were high communications observed between A-B and C-D chains which was illustrated in Figure 3.33 The minor communication was seen

between A-C and B-D with no contact, while the B-C and A-D chains showed slightly more communication through the (A-A domain interactions) large interface. According to the maps in the row of constr-1 state, while the communication between the A-B and C-D chains continues, the signaling between the A-D and C-B chains further increased. Also, there was a slight increase in communication between C-A and B-D chains. It can be revealed that this increased communication was seen since the restricted residues were found at the interface of A/D chains and the interface of B/C chains. When the maps of the constr-2 states were examined, it was seen that there was no significant difference between them (565352 data points) and apo (551540 data points) states. The restriction of the known region between the C-C domains of A-B and C-D chains localized in the small interface region did not lead to major changes in inter-chain communication.

Figure 3.32 Distance difference fluctuations of (a) Apo, (b) Constr-1 and (c) Constr-2 simulations and their average values shown on maps, respectively. Distance fluctuation values were shown in purple while the residues contacts shown in red colors and constrained residues represented with yellow color. (Distance fluctuation values between 0-0.5Åshown on the map)
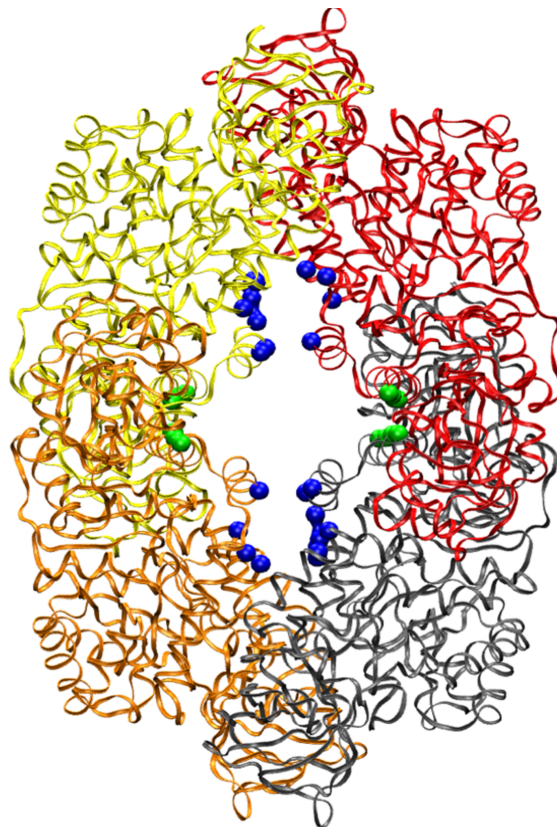
Figure 3.33 Structural representation of constrained regions on *S. aureus* pyruvate kinase. Constrained residues. for Constr-1 state were represented with blue points while constr-2 state residues were coloured with green points

4. CONCLUSIONS

In this study, the allosteric effect of the proposed and the known allosteric regions in *S. aureus* PK was investigated via Molecular Dynamics simulations. The proposed allosteric site is of great importance in being used as a drug target in species-specific drug studies due to high rate of mutation compared to highly conserved catalytic regions. Since the parameterization of the ligand during the system preparation had a low degree of accuracy, distances between selected residues interacting with the ligand were restrained such that the distance remains constant during simulation. The imposed distance restraints at the allosteric site allows to mimic the state of a bound ligand which would similarly restrict the interacting residues. Three independent sets of 100 ns long MD simulations were conducted for apo state where there is no restriction, for the proposed site with restrictions (constr-1) and the known allosteric site with restrictions (constr-2).

Several analysis methods were used to see the differences between apo and constrained states and the effects of restricted residues on the global dynamics of the protein. First, RMSD values of tetramer structures with respect to the initial state monitored for all nine simulations indicated that the systems reached equilibrium. In addition, RMSD value of each chain was also monitored, and the chains causing the highest RMSD deviation were determined. In apo simulations (within 3 sets of runs), the highest RMSD values were observed for A and D chains. In constr-1 simulations where the proposed regions (between A-D, B-C chains) were restricted, the highest RMSD values belonged to A and D chains. On the other hand B and C chains had the highest RMSD values when the residues in the known allosteric region were restricted (constr-2 state). According to RMSD values of chains, apo and constr-1 states were similar while constr-2 were different from them. Then, the

mobile regions in the tetramer were revealed by RMSF analysis. It was determined that the B domain and CT domains were the most mobile domains in apo and constrained cases. On the other hand, the A and C domains remained more stable. When the RMSF values of apo and constrained ones were compared, B domains were more mobile in constrained states. This suggests that the protein dynamics of the restrictions were closer to allosteric inhibition because it was known that this region becomes more stable and closes on the A domain when it becomes active. Next, principal component analysis was employed, and the large-scale essential motions in the simulations and their contributions to the total dynamics were examined. It has been seen that higher number of principal modes (first 32 for constr-1 state and 42 modes for constr-2 state) were required in constrained states than apo states (first 31 modes) to explain 90% of the total dynamics. On the other hand, when each snapshot in the trajectory was projected on the first three modes, it was seen that apo and constr-2 were similar to each other. Also, projection ranges of constr-1 and apo states were different from each other. Analysis of the correlation graphs for residue fluctuations showed less correlation in constrained cases than apo. When comparing constrained and apo states, it was seen that positive correlations in the same chains disappeared and became zero. At the same time, the correlations in the inter-chain regions were significantly reduced. Considering the A-A domain interface, an essential region for protein function, the restricted residues led to a decrease in correlations between these neighboring regions. In another critical area, the C-C interface, positive correlations between adjacent domains were observed to be stronger. In the constrained state, the fluctuation of these domains in the same direction was intensified. On the other hand, new correlations, albeit small, emerged in constrained states. Accordingly, while communications between some regions ended, new inter-regional correlations appeared. New or strengthened positive correlations were mainly formed between A and CT domains. These domains were known to interact with each other. The interaction of the A-CT domains was considered to provide the optimum stability necessary for the catalytic activity of the protein. Distance analysis for A-CT domains through the trajectory was performed to see how the distance between domains changed as a result of movements

in the same direction. Accordingly, there was not enough closeness for interaction between domains in both apo and constrained states. Both states seem quite far from the active structure and unstable.

Clustering analysis based on RMSD value was performed to examine the effects of restricted residues on the catalytic region. A representative snapshot was determined for representing each state, apo, constr-1, constr-2. In this way, it was possible to compare apo and constrained forms. Noticeable differences in the catalytic region were obtained as transformations from helix to coil/turn structure. The most prominent of these was the $\alpha'6$helix, which plays an essential role in domain rotation. It is interacting with PEP, also takes part in establishing the necessary bonds to stabilize the active structure. The $\alpha'6$helix in apo structure was more frequently transformed into coil or turn than in both constrained states and was found closer to PEP. In constrained states, on the other hand, the helix structure moved away from PEP or rotated slightly. Here, apo and constrained states were also far away from the active form. Still, since the helical structures in the constrained state did not undergo any secondary structural change, they were more stable and closer to the active conformation. However, this may have occurred due to a signal transmitted allosterically by the restricted residues. In this case, it may have remained stable or moved away from PEP molecule to form new bonds that would lead to inhibition.

A principal axis analysis was performed to examine the rotational movements of the domains. With the results from here, it was examined how the domains rotate according to the tetramer structure and compared apo state with the constrained states. . The angles of rotation with respect to the tetramer were investigated with the AC domains as a single piece and the A, C domains alone. Accordingly, while the rotation angle fluctuated less in apo simulations, these angles began to fluctuate more in constrained states.

Finally, distance fluctuation analysis was performed to determine the effect of restrictions on allosteric regions on the communication of distant residues. The obtained

results showed that there was a strong communication between A-B and C-D chains in apo simulation. However, according to the result of constr-2 state, it was seen that the communication between the other chains has increased in addition to the interactions between chains in the apo state. This increase was particularly evident in interactions between the A-D and B-C chains. On the other hand, there was no significant difference in the results of constr-2 state under the effect of restricted residues in the known region compared to apo state.

In this study, we had the opportunity to observe the allosteric effects of the proposed allosteric sites on the *S. aureus* pyruvate kinase protein, based on the previous studies of our group, via short Molecular Dynamics simulations. The simulations we made without including the inhibitor molecule in the system showed their differences from apo states and showed that our method was effective and practical. At the same time, the two constrained states showing similar results to each other revealed that the proposed region had a robust allosteric effect based on the known allosteric region. The most challenging part was the absence of an active *S. aureus* pyruvate kinase structure that would be taken as a reference. In future studies, the active form of *S. aureus* can be simulated, and its allosteric mechanism can be clarified. In addition, a more stable state of such a large protein can be examined and large-scale movements can be captured easily by performing longer Molecular Dynamics simulation studies.

APPENDIX A: Configuration Files of Apo and Constrained Simulations

The appendices start here.

```
## JOB DESCRIPTION                    ##
# Minimization and Equilibration of
# Pyruvate Kinase in a Water Box
## ADJUSTABLE PARAMETERS
structure        ./ionized.psf
coordinates      ./ionized.pdb
set temperature   310
set outputname    ionized_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  ;#nonbonded exclusion
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   163.262  0.0  0.0
cellBasisVector2   0.0  163.367  0.0
cellBasisVector3   0.0  0.0  120.129
cellOrigin        27.355 0.190 0.122
wrapAll           on
# PME (for full-system periodic electrostatics)
PME              yes
PMEGridSpacing    1.0
#manual grid definition
#PMEGridSizeX      56
#PMEGridSizeY      74
#PMEGridSizeZ      70
# 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   ;# 500steps = every 1ps
dcdfreq           10000   ;# at every 40 ps
xstFreq           10000
outputEnergies    10000
outputPressure    10000
# Minimization
minimize          1000
reinitvels        $temperature
run 50000000 ;# 100ns
```

Figure A.1 Configuration file script for apo simulations of *S. aureus* PK

```
## JOB DESCRIPTION                          ##
# Minimization and Equilibration of
# Pyruvate Kinase in a Water Box
## ADJUSTABLE PARAMETERS                     ##
structure        ./ionized.psf
coordinates      ./ionized.pdb
set temperature   310
set outputname    constraint1_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   ;#nonbonded exclusion
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
langevinDamping   1
langevinTemp      $temperature
langevinHydrogen  off
# Periodic Boundary Conditions
cellBasisVector1   163.262  0.0   0.0
cellBasisVector2   0.0   163.367  0.0
cellBasisVector3   0.0   0.0   120.129
cellOrigin        27.355 0.190 0.122
wrapAll           on
# PME (for full-system periodic electrostatics)
PME              yes
PMEGridSpacing    1.0
#manual grid definition
#PMEGridSizeX      56
#PMEGridSizeY      74
#PMEGridSizeZ      70
# 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   ;# 500steps = every 1ps
dcdfreq            10000   ;# at every 40 ps
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 ;# 100ns


#_____CHAIN C-B_____
#C260-B337
bond 21985 14194  150 9.7
#C264-B338
bond 22057 14204 150 12.06
#C302-B264
bond 22641 13116 150 15.07
#C342-B342
bond 23217 14276 150 10.09
#B303-B346
bond 13721 14347 150 11.8
#B302-B339
bond 13700 14221 150 10.04


#_____CHAIN A-D_____
#A260-D337
bond 4103 32076 150 9.8
#A264-D338
bond 4175 32086 150 12.08
#A302-D264
bond 4759 30998  150 15.06
#A342-D342
bond 5335 32158 150 11.0
#D303-D346
bond 31603 32229 150 11.08
#D302-D339
bond 31582 32103 150 10.04
```

Figure A.2 Configuration file script for constr-1 simulations and content of constrained.txt file for *S. aureus* PK

70

```
## JOB DESCRIPTION                              ##
# Minimization and Equilibration of
# Ubiquitin in a Water Box
## ADJUSTABLE PARAMETERS                        ##
structure         ./ionized.psf
coordinates       ./ionized.pdb
set temperature   310
set outputname    ionized_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  ;#nonbonded exclusion
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   163.262  0.0   0.0
cellBasisVector2   0.0   163.367  0.0
cellBasisVector3   0.0   0.0   120.129
cellOrigin        27.355 0.190 0.122
wrapAll           on
# PME (for full-system periodic electrostatics)
PME               yes
PMEGridSpacing    1.0
#manual grid definition
#PMEGridSizeX      56
#PMEGridSizeY      74
#PMEGridSizeZ      70
# Constant Pressure Control (variable volume)
useGroupPressure  yes ;# needed for rigidBonds
```

```
useFlexibleCell     no
useConstantArea     no
langevinPiston      on
angevinPistonTarget 1.01325 ;#  in bar -> 1 atm
langevinPistonPeriod 100.0
langevinPistonDecay 50.0
langevinPistonTemp   $temperature
# Output
outputName        $outputname
restartfreq       10000  ;# 500steps = every 1ps
dcdfreq           10000 ;# at every 40 ps
xstFreq           10000
outputEnergies    10000
outputPressure    10000
## EXTRA PARAMETERS                             ##
## EXECUTION SCRIPT                             ##
## CONSTRAINTS DURING MINIMIZATION
extraBonds on
extraBondsFile constraint.txt
#constraints          on
#consexp          2
#consref          restraint.ref
#conskfile        restraint.ref
#conskcol         B
#constraintScaling     1.0
# Minimization
minimize          1000
reinitvels        $temperature
run 50000000 ;# 100ns

#_____CHAIN A-B_____
#ALA358-LEU370
bond 5602 14709 150 8.22
#SER362-THR366
bond 5657 14652 150 7.93
#THR366-SER362
bond 5711 14598 150 7.91
#LEU370-ALA358
bond 5768 14543 150 8.13


#_____CHAIN C-D_____

#ALA358-LEU370
bond 23484 32591 150 8.18
#SER362-THR366
bond 23539 32534 150 7.91
#THR366-SER362
bond 23593 32480 150 7.92
#LEU370-ALA358
bond 23650 32425 150 8.14
```

Figure A.3 Configuration file script for constr-2 simulations and content of constrained.txt file for *S. aureus* PK.

Figure B.1 Rotation angles alterations of domain A through the direction 1 for a) chain A,b)chain B, c)chain C d) chain D. Apo, constr-1 and constr-2 represented with purple, red, green respectively.
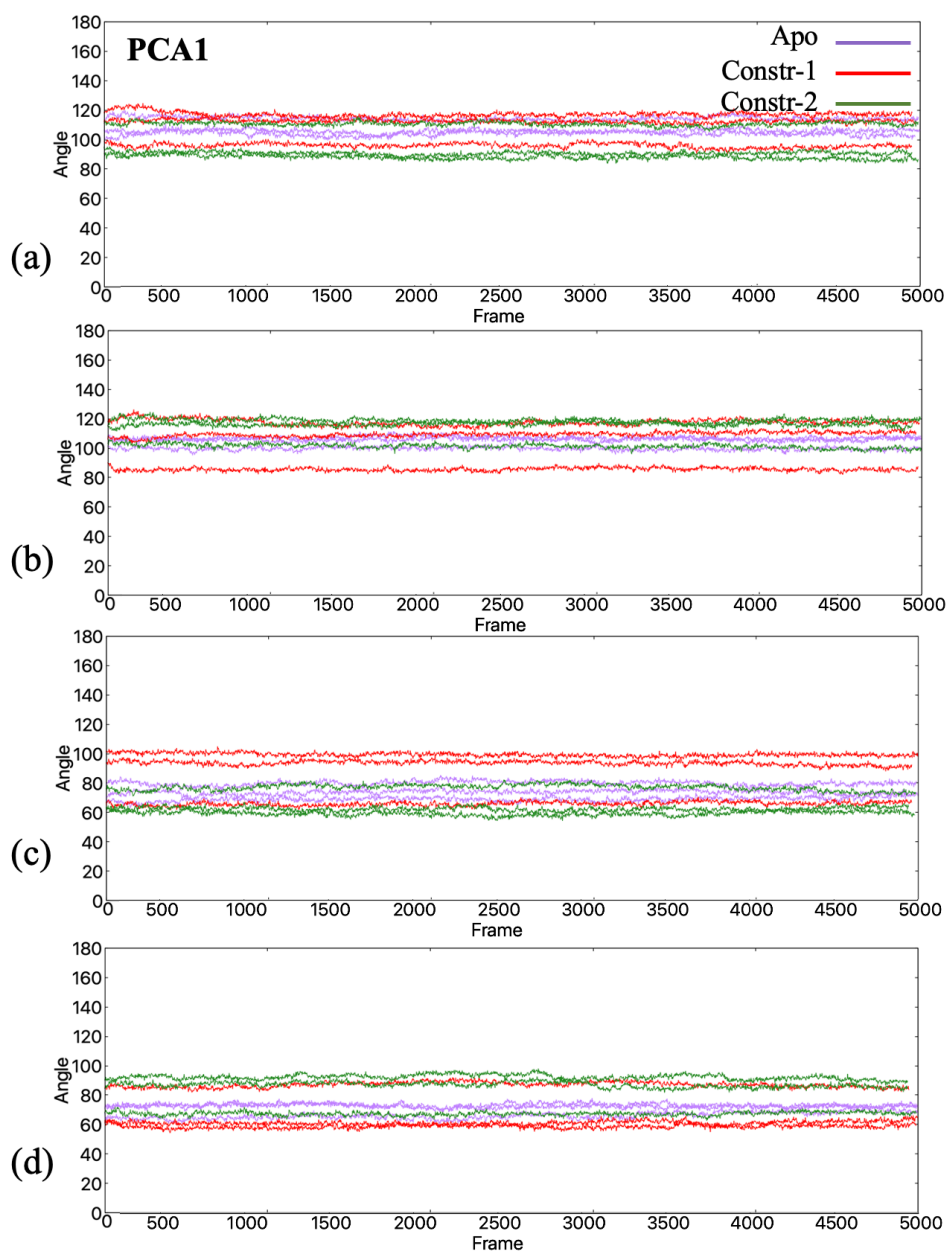
Figure B.2 Rotation angles alterations of domain A through the direction 2 for a) chain A,b)chain B, c)chain C d) chain D. Apo, constr-1 and constr-2 represented with purple, red, green respectively.
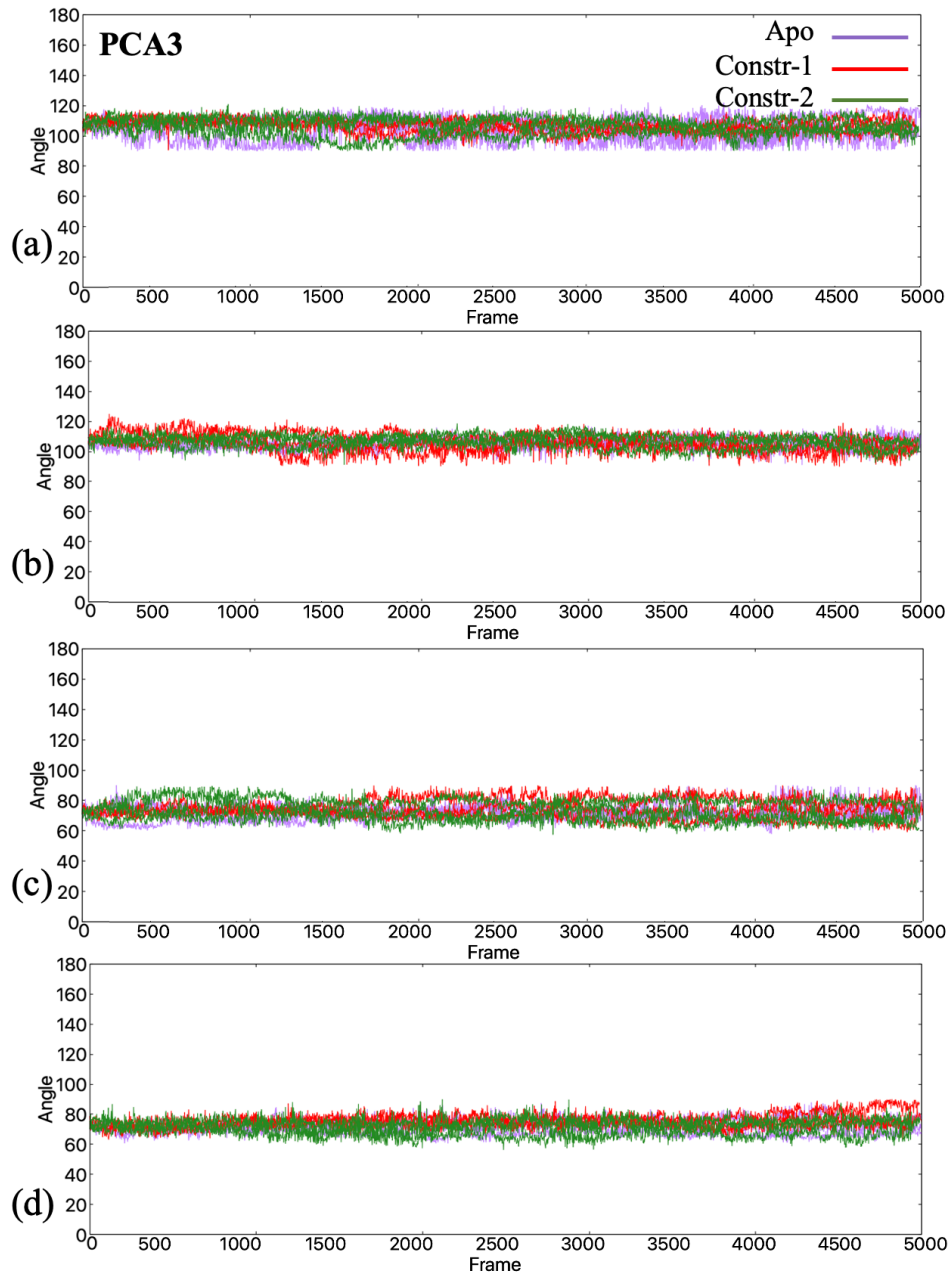
Figure B.3 Rotation angles alterations of domain A through the direction 3 for a) chain A,b)chain B, c)chain C d) chain D. Apo, constr-1 and constr-2 represented with purple, red, green respectively.
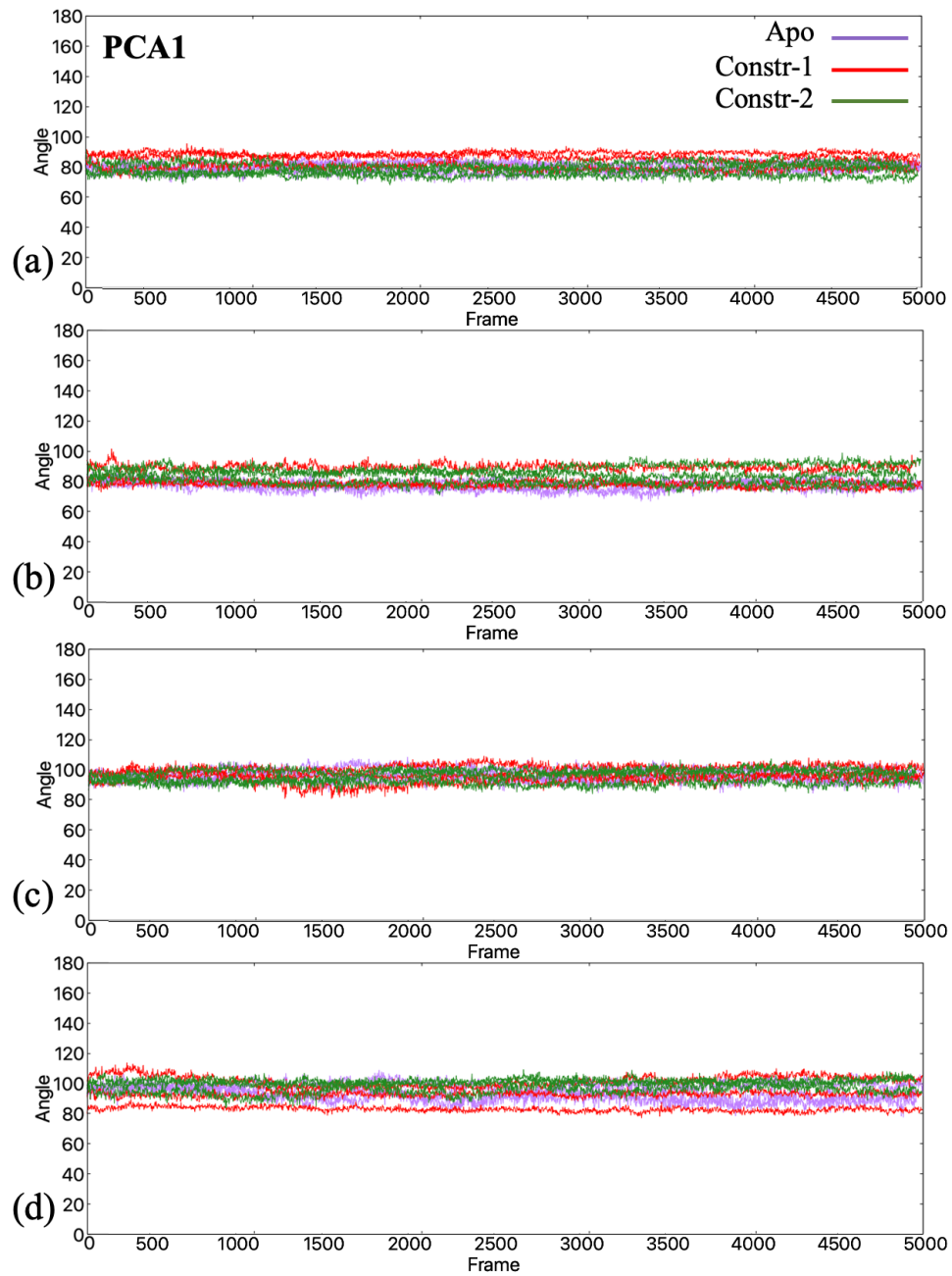
Figure B.4 Rotation angles alterations of domain C through the direction 1 for a) chain A,b)chain B, c)chain C d) chain D. Apo, constr-1 and constr-2 represented with purple, red, green respectively.
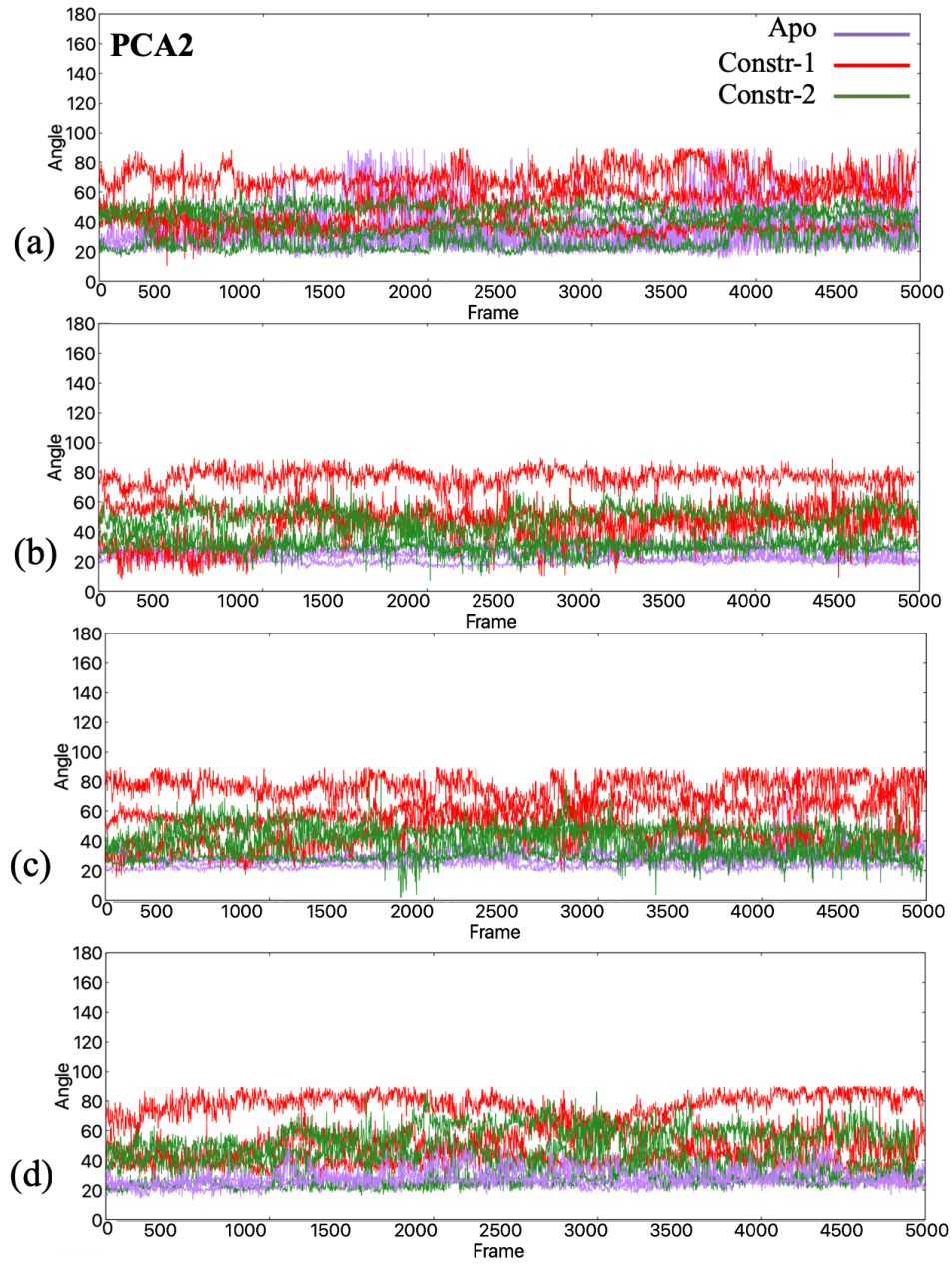
Figure B.5 Rotation angles alterations of domain C through the direction 2 for a) chain A,b)chain B, c)chain C d) chain D. Apo, constr-1 and constr-2 represented with purple, red, green respectively.
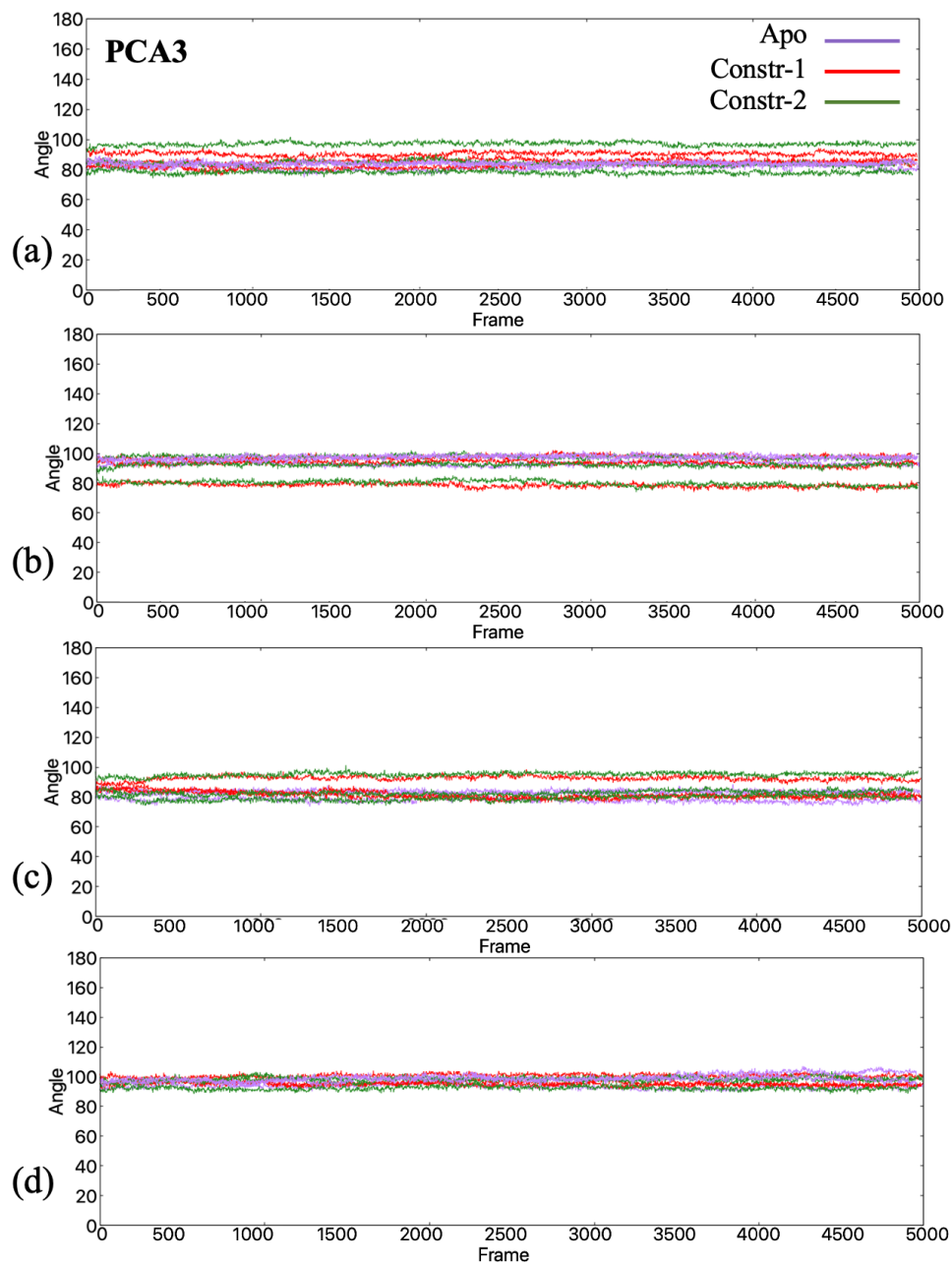
Figure B.6 Rotation angles alterations of domain C through the direction 3 for a) chain A,b)chain B, c)chain C d) chain D. Apo, constr-1 and constr-2 represented with purple, red, green respectively.
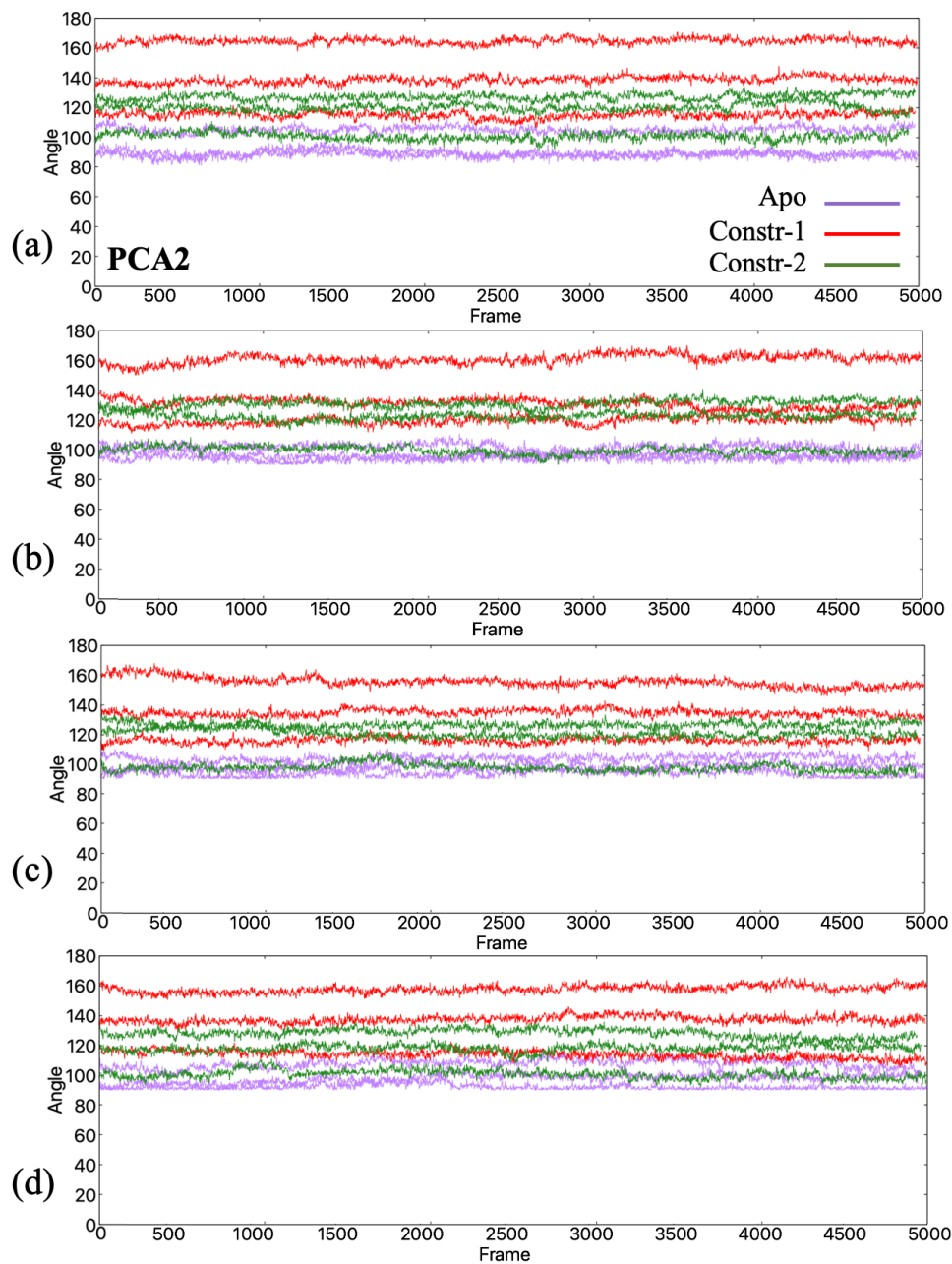
APPENDIX C: Used Commands for Principal Component Analysis

Step 1: Installation of ProDy on MacOS Terminal:

pip install -U ProDy

Step 2: Running Prody on the terminal:

prody pca -a -A –apo_run1.dcd calpha -pdb ionized.pdb

As a result of step 2, the cross-correlation values between residues were calculated and written in an txt file.

Step 3: Running python scripts to calculate fraction of variance for each modes:
In[1]: from prody import *
from pylab import *
ion()
structure = parsePDB('ionized.pdb')
structure
dcd = DCDFile('pk_run1.dcd')
dcd.link(structure)
dcd.setAtoms(structure.calpha)
dcd
In[2]:
pca = PCA('Apo_Run1')
pca.buildCovariance(dcd)

In [3]:

pca.calcModes('all')

pca

In[4]:

saveModel(pca)

In[5]:

for mode in pca[:200]:

print (calcFractVariance(mode).round(5))

Cumulative percentage values for the fraction of variance of each modes calculated on Excell and plotted with Gnuplot (Williams and Kelly, 2011).

Step 4, The calculated cross correlation values were written in a txt file as one column with the following MATLAB commands.

cor=readtable('pk_apo_Run1_pca_cross-correlations.txt')

A = table2array(cor)

a=reshape(A',[] ,1)

fileID = fopen('corr.txt','w');

fprintf(fileID,'%6.4f',a);

Step 5: The cross-correlation maps were plotted with Gnuplot v5.4 (Williams and Kelly, 2011).

REFERENCES

Allen, M. P. (2004) "Introduction to Molecular Dynamics Simulation," *Computational Soft Matter: From Synthetic Polymers to Proteins, Lecture Notes*, 23, pp. 1–28. doi: 10.1016/B978-0-12-816954-4.00001-2.

Allen, M. P. and Tildesley, D. J. (1989) *Computer Simuation of Liquids, Journal of Solution Chemistry.* doi: 10.1063/1.3656072.

Amadei, A., Linssen, A. B. M. and Berendsen, H. J. C. (1993) "Essential dynamics of proteins Opportunities Connect with Wiley," *Proteins: Structure, Function, and Genetics*, 17(4), pp. 412–425.

Axerio-Cilies, P. et al. (2012) "Cheminformatics-driven discovery of selective, nanomolar inhibitors for staphylococcal pyruvate kinase," *ACS Chemical Biology*, 7(2), pp. 350–359. doi: 10.1021/cb2003576.

Ayyildiz, M. et al. (2020) "Identification of Alternative Allosteric Sites in Glycolytic Enzymes for Potential Use as Species-Specific Drug Targets," *Frontiers in Molecular Biosciences*, 7(May), pp. 1–19. doi: 10.3389/fmolb.2020.00088.

Bakan, A. and Bahar, I. (2009) "The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding," *Proceedings of the National Academy of Sciences of the United States of America*, 106(34), pp. 14349–14354. doi: 10.1073/pnas.0904214106.

Bakan, A., Meireles, L. M. and Bahar, I. (2011) "ProDy: Protein dynamics inferred from theory and experiments," *Bioinformatics*, 27(11), pp. 1575–1577. doi: 10.1093/bioinformatics/btr168.

Berman, H. M. et al. (2002) "The protein data bank," *Acta Crystallographica Section D: Biological Crystallography*, 58(6 I), pp. 899–907. doi: 10.1107/S0907444902003451.

Bowerman, S. and Wereszczynski, J. (2016) *Detecting Allosteric Networks Using Molecular Dynamics Simulation.* 1st edn, *Methods in Enzymology.* 1st edn. Elsevier Inc. doi: 10.1016/bs.mie.2016.05.027.

Donovan, K. A. et al. (2016) "Conformational dynamics and allostery in pyruvate kinase," *Journal of Biological Chemistry*, 291(17), pp. 9244–9256. doi: 10.1074/jbc.M115.676270.

González, M. A. (2011) "Force fields and molecular dynamics simulations," *EDP Sciences*, 12, pp. 169–200. doi: DOI: 10.1051/sfn/201112009.

Guarnera, E. et al. (2017) "AlloSigMA: Allosteric signaling and mutation analysis server," Bioinformatics, 33(24), pp. 3996–3998. doi: 10.1093/bioinformatics/btx430.

Haider, S., Parkinson, G. N. and Neidle, S. (2008) "Molecular dynamics and principal components analysis of human telomeric quadruplex multimers," *Biophysical Journal*, 95(1), pp. 296–311. doi: 10.1529/biophysj.107.120501.

Huang, J. and Mackerell, A. D. (2013) "CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data," *Journal of Computational Chemistry*, 34(25), pp. 2135–2145. doi: 10.1002/jcc.23354.

Humphrey, W., Dalke, A. and Schulten, and K. (1996) "VMD: Visual Molecular Dynamics," Journal of Molecular Graphics, (14), pp. 33–38. doi: 10.1016/j.carbon.2017.07.012.

Kamberaj, H. (2016) *Molecular Dynamics Simulations in Statistical Physics. Theory and Applications., Advanced molecular dynamics techniques.*

Kufareva, I. and Abagyan, R. (2012) "Methods of protein structure comparison," *Methods in Molecular Biology*, 857, pp. 231–257. doi: 10.1007/978-1-61779-588-6.

Leach Andrew R. (2001) *Molecular Modeling: Principles and Applications, Pearson Education Limited.* doi: 10.1021/ci9804241.

Lovell, S. C., Mullick, A. H. and Muirhead, H. (1998) "Cooperativity in Bacillus stearothermophilus pyruvate kinase," *Journal of Molecular Biology*, 276(4), pp. 839–851. doi: 10.1006/jmbi.1997.1569.

Mattevi, A. et al. (1995) "Crystal structure of Escherichia coli pyruvate kinase type I: molecular basis of the allosteric transition," *Structure*, 3(7), pp. 729–741. doi: 10.1016/S0969-2126(01)00207-6.

McQuarrie, D. A. (1975) *Statistical mechanics.*

Morgan, H. P. et al. (2010) "Allosteric mechanism of pyruvate kinase from Leishmania mexicana uses a rock and lock model," *Journal of Biological Chemistry*, 285(17), pp. 12892–12898. doi: 10.1074/jbc.M109.079905.

Morgan, H. P. et al. (2014) "Structures of pyruvate kinases display evolutionarily divergent allosteric strategies," *Royal Society Open Science*, 1(1). doi: 10.1098/rsos.140120.

Naithani, A. et al. (2015) "A Molecular Dynamics Study of Allosteric Transitions in Leishmania mexicana Pyruvate Kinase," *Biophysical Journal*, 109(6), pp. 1149–1156. doi: 10.1016/j.bpj.2015.05.040.

Nelson, M. et al. (1995) "MDScope - a visual computing environment for structural biology," *Computer Physics Communications*, 91(1–3), pp. 111–133. doi: 10.1016/0010-4655(95)00045-H.

Özhelvacı, M. F. (2020) *Determination of Species-Specific Allosteric Binding Sites In Pyruvate Kinase and its use inDrug Design Studies.*

Sakai, H. (2004) "Possible structure and function of the extra C-terminal sequence of pyruvate kinase from Bacillus stearothermophilus," *Journal of Biochemistry*, 136(4), pp. 471–476. doi: 10.1093/jb/mvh152.

Sterling, T. and Irwin, J. J. (2015) "ZINC 15 - Ligand Discovery for Everyone," *Journal of Chemical Information and Modeling*, 55(11), pp. 2324–2337. doi: 10.1021/acs.jcim.5b00559.

Suzuki, K. et al. (2008) "Crystal structure of pyruvate kinase from Geobacillus stearothermophilus," *Journal of Biochemistry*, 144(3), pp. 305–312. doi: 10.1093/jb/mvn069.

Tanaka, K. et al. (1995) "Molecular Cloning of the Genes for Pyruvate Kinase of Two Bacilli, Bacillus psychrophilus and Bacillus licheniformis, and Comparison of the Properties of the Enzymes Produced in Escherichia coli," *Bioscience, Biotechnology, and Biochemistry*, 59(8), pp. 1536–1542. doi: 10.1271/bbb.59.1536.

Verlet, L. (1967) "Computer 'Experiments' on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules," *Journal of Physics D: Applied Physics*, 159(1), pp. 183–195. doi: 10.1088/0022-3727/9/2/008.

Williams, T. and Kelly, C. (2011) "Gnuplot."

Zheng, L. et al. (2018) "Molecular dynamics and simulation," *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 1–3(V), pp. 550–566. doi: 10.1016/B978-0-12-809633-8.20284-7.

Zoraghi, R. et al. (2010) "Functional analysis, overexpression, and kinetic characterization of pyruvate kinase from methicillin-resistant staphylococcus aureus," *Biochemistry*, 49(35), pp. 7733–7747. doi: 10.1021/bi100780t.